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}

¹Natural Sciences Research Institute, College of Science, University of the Philippines Diliman, Diliman, Quezon City 1101 the Philippines ²Institute 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 neighborjoining (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)	CAF2-STOP ^a (ATTCACGTCTGACTTCTCTGGCT)	Clones labeled C (ge- nomic)
	CAF1-MIDR ^b (CCCAATTCAATCACCAAACCGC)		
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)

^aPrimers from Perrois et al. (2015)

^bAdditional 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- μ L reactions as follows: 32.5 μ L of nuclease-free water, 10 μ L of 5x MyTaq Reaction Buffer (Bioline, Cat No.: BIO-21106), 1 μ L of MyTaq DNA polymerase (Bioline, Cat No.: BIO-21106), 2 μ L each of the forward and reverse primers (10 μ M), and 2.5 μ L 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 15 s, 60 °C for 30 s, 72 °C for 30 s, 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 3^{rd} 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.

			1 1 0		11 0	
	1 st intron	2 nd exon	2 nd intron	3 rd exon	3 rd intron	4 th 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
МХМТ			759A>C 761A>T 801C>A 819T>C 830T>G	936T>C 1156A>G	1201T>C Multiple indels	1 7 2 7 T > A 1 7 6 3 A > 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	1 7 1 7 G > C 1 7 3 8 A > G 1773T>C

 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.

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 2^{nd} intron, 3^{rd} exon, and 4^{th} exon, the largest amount of variation between the *Coffea* species can be found in the 3^{rd} 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).

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



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.

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 3^{rd} 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 3^{rd} 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 β 5- α 6 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* (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-Nmethyltransferase; [XMT] xanthosine methyltransferase; [TrigS] trigonelline synthase.

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,



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.

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

Mondragon *et al.*: Characterization of *NMT* Genes in Liberica Coffee

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



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.

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.

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, MU-RAT 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, MAZ, 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, WINCK-ER 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, CROUZ-ILLAT 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, GROS-SI-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, JA-YARAMA, 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.
- DROZDETSKIYA, 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, RAKOTO-MALALA 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-N⁷-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-knowabout-liberica-the-unique-philippine-coffee/
- PERROIS C, STRICKLER SR, MATHIEU G, LEP-ELLEY 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.
- RAHARIMALALAN, ROMBAUTS S, MCCARTHYA, GARAVITO A, OROZCO-ARIAS S, BELLANGER

L, MORALES-CORREAAY, FROGER S, MICHAUX S, BERRY V, S METAIRON S, FOURNIER C, LEP-ELLEY 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).

	1	10	20	30	40	50	60	7	0	80	90
	15				15		51	-51	5		[
E1/11											
CGXMT1	ADGGAGO	CCAAGAAGE	CUIGCAGAIG	ARTGGRGGCG	RAGGUGA	ACARGU	ACGCCARGAR	LCALCU	ACAALGE	Tel Cresci	
CAXMI1	AL GGAGO	CCAAGAAGI	CCIGCAGAIG	RAIGGAGGUG	ARCOLAR	ACAAGU	ACCCCCARGAA	CABCO	ACAALSI		
CAMPIZ	Read Alar	CCRAGRAGE				ACARGU	ACCCCARCAR	TOR DOC	ACAA SOL		
CIO AVMT						ACARSO	ACCCARGAR	TO A DOOR	ACAA DOL		
C15 *XMT				-AAGGAGGCG	ACCCCCAT	ACAACC	ACCCCAACAAT	TCACCC	ACARTON	TETCTETC	c c
CcMXMT1	ATGGAGC	CCAAGAACT	CC <mark>IGCAIAIG</mark>	AATGAAGGTGI	AAGGCGAT	ACAA <mark>C</mark> C	ACCCCAACAAT	GCATCO	ACAA <mark>TGT</mark> O	CIGICICICI	CC
CaMXMT1	ATCGAGC	ICCAAGAAGT	CCIGCAIAIG	AATGAAGGTG	AAGGCGAT	ACAAGC	ACOCCARGAAT	GCATCCI	ACAATGT	Telcicici	CTC
CaMXMT2	ATGGAGC	TCCAAGAAGT	CCIGCAIAIG	AATGGAGGTG	AA <mark>go</mark> cga t	ACAAGCI	ACCCCARGAAT	TCATCCI	ACAATGI	TGICIGIC	C C
C1 *MXMT				GAAGGAGGTG	AAGGCGAT	ACAAGC	ACCCCARCAAT	TCATCCI	ACAAIGI	Teicieic	CC
C2 *MXMT			ACCEGRA	GAAGGAGGCG.	AGGGCGAI	ACAAGC	ACCCCARGAAL	ICACCCI	ACAAIGI	TGICIGICI	CTC .
C4 *MXMT					GCGA	ACAAGC	ACCCCAAGAAI	TCATCCI	ACAAIGI	CIGICIGEC	CCC
C5 *MXMT				GAAGGAGGTG	AAGGCGAT	ACAAGC	ACOCCAAGAAT	ICAICCI	ACAATGT	TGICICICI	CEC
C6 *MXMT					GGCGA	ACAACC	ACCCCAAGAA	TCATCCI	ACAAIGT		CTC .
C9 *MXMT					BCGA	ACAAGC	ACCCCAAGAAT	TCALCCI	ACAATGE	TGICIGICI	CEC
C10 *MXMT				GAGGTG	AAGGCGA	ACAAGC	RCGCCARGAAT	TCATCCI	ACAAIGI	TGICIGICI	CTC
C11 *MXMT					AAGGCGA	ACAAGC	ACOCCAAGAAT	ICAI COL	ACAATGI	CIGICICIC)	C C
C13 *MXMT				<mark>G</mark>	AAGGCGA	ACAAGC	ACCCCARGAAT	TCATCCI	ACAATGI	TGICIGICI	CCC
C14 *MXMT			-CIGCGIAIG	AATGGAGGTG	AAGGCGAT	ACAAGC	ACOCCAAGAAT	TCATCCI	ACAATGT	TGICIGICI	CTC
C16 *MXMT				AATGGAGGTG	AAGGCGA	ACCAGC	ACOCCAAGAAI	ICAICCI	ACAATGI	TGICIGICI	CTC .
C17 *MXMT					-AGGCGAI	ACAAGC	ACOCCARGAAT	TCATCCI	ACAAIGI	TGICIGICI	C.C.
C18 *MXMT				G	AAGGCGA	ACAAGC	ACCCCAAGAAT	TCATCCI	ACAAIGI	Telcieici	C
C19 *MXMT		CCA	ICIGCCIAIC	AA <mark>IGGA</mark> GGIG	AAGGCGA	ACAAGC	ACCCCARGAA		ACAAIGI		
CoDXMT	ATCGAGC	CCAAGAACT	COTOCATATO	AATGGACGCG	AACCCCA	ACAACC	ACCCAACAAC		ACAATGT		CIC
CaDXMT2	ADGGAGC	TCCAAGAAGT	CCTGCATATG	AATGGAGGCG	AA <mark>gg</mark> c <mark>ga</mark> t	ACAACC	ACCCCAACAAC	TCATCCI	ACAATGT	TGICIGICI	CIC
CaDXMT1	ATCGAC	TCCAAGAAGT	CCTGCATATG	AATGGACCC	AAGGCGAT	ACAACC	ACCCCAACAAC	ICATICI	ACAATGE	Tercierci	CCC
R1 *DXMT			CCTGCCTATE	AATGGAGGCG	AAGGCGAT	ACAACC	ACCCAACAAC		ACAATCT	TGTCTCTC	ctc
R3 *DXMT		CCCAGAAG	CCTGCGTATC	AATGGAGGCG	AAGCCGAT	ACAACC	ACCCCAACAAC	TCATCCI	ACAATCE	TETCTETC	CIC
R4 *DXMT	GGAGC	TCCCAGAAGT	CCIGCCIAIC	AATGGAGGCG	AAGGCGAT	ACAAGC	ACCCAACAAC	CALCCI	ACAATGE	Tercretc	C C

	91 15-	100	110	120	130	140	150	160	170	180
I1										
COMMIT			ACACACACACA	ACACA		TAGEAAA UA			AAC LAGER	
CaXMT2			ACACACACACACA	CACACA		TACTARAC	TOCT MODEL			
C3 *XMT			ACACACACACA	CACACACACA		TACTAAATCA	TOCTATCATA	CICCADO	TAACTTACCA	CTCT
C12 *XMT	ACCO		ACACACACACA	CACACACACA	CAGACTA	TAGTAAATCA	TOCTACCATA	CETCEADCE	TAACI TAGCA	Terer
C15 *XMT	TATCE	TCTTTA	ACACACACAC	CACACACACA	CAGAGTAC	TA <mark>gt</mark> aaa <mark>t</mark> ca	IGCTATGATA	CERCEADCE	C <mark>TAACTT</mark> AGCA	TGCCT
CoMXMT1		101 11 1	SCECECECECE							
C=MXMT1			ACACACACACA	ACACACACA	TAGAGAATA	TOCTANATCA	TOCTATCATA	CHCCARCE	AACTT	
CaMXMT2	ACCO		ACACACACACA	A	GACTA	TGGTAAATCA	TOCTATCATA	CICCAICI	TAACTTACTA	TGECT
C1 *MXMT	TATCTC		ACACACACACA		<mark>GACT</mark> AC	TGGTAAATCA	ICCTATCATA	CETCEATCE	C <mark>TAACITAGI</mark> A	TGTCT
C2 *MXMT	TATCTO		ACACACACACA	CACACACACA	CAGACTA	TAGTAAATCA	TECTATETTA	CGICGAICT	CTAACTTAGCA	TCTCT
C4 *MXMT	TATCE		ACACACACACA		GAGTAC	TOGTAAATCA	ICCTATCATA	CGECGAECE	CTAACTIAGIA	IGICI

C5 *MKMT C6 *MKMT C9 *MKMT C10 *MKMT C11 *MKMT C13 *MKMT C14 *MKMT C16 *MKMT C17 *MKMT C18 *MKMT C19 *MKMT		HACACACACACA HACACACACACA HACACACACACA HACACACAC		CACTA CACTA CACTA CACTA CACTA CACTA CACTA CACTA CACTA CACTA CACTA CACTA CACTA	GT GG TAAN GT GG TAAN	Caigo a ga Caigo a caigo Caigo a caiga Caigo a caiga		CIAACIACIA CIAACIACIA CIAACIACIA CIAACIACIA CIAACIACIA CIAACIACIA CIAACIACIA CIAACIACIA CIAACIACIA CIAACIACIA CIAACIACIA	61 CI 61 CI
CcDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT		IACACACACACA IACACACACACA IACACACACACA IACACACAC	CACACACAC CACACACACA CACACACACA CACACACACA CACACACACACA CACACACACACA	GACTA GACTA GACTA GACTA GACTA 	GTAG AAA GTAG AAA GTAG AAA GTAG AAA GTAG AAA GTAG AAA GTAG AAAT	CAIGCIAIGA CAIGCIAIGA CAIGCIAIGA CAIGCIAIGA CAIGCIAIGA CAIGCIAIGA CAIGCIAIGA	AGC 002 0 AGC 002 0 AGC 002 0 AGC 002 0 AGC 002 0 AGC 002 0 AGC 002 0	C AAC TACTA C AAC TACTA C AAC TACTA C AACCTACTA C AACCTACTA C AACCTACTA C AACCTACTA	
I1/E2 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT	181 I5 TTTTTCC TTTTTCC TTTTTCC TTTTTCC	200 -5I5 AACAPITGEATI AACAPITGEATI AACAPITGEATI AACAPITGEATI	210 	220 5	230 SI AACTGGTTC AACTGGTTC ATCTGGTTC ATCTGGTTC ATCTGGTTC	240 -SI COCCAACCI CCCCAACCI CCCCAACCI CCCCAACCI TAACAACCI TAACAACCI TAACAACCI TAACAACCI	250 -5	260 5 5- TICAACAALGCC TICAACAALGCC TICAACAALGCA TICAACAALGCA	270 I TACG TACG TACG TACG
C15 *XMT CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C3 *MXMT C4 *MXMT C6 *MXMT C1 *MXMT C11 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT C15 *MXMT C18 *MXMT			TICGCCACTGC TICGACTGC TICGACTGC TICGACTGC TICGACTGC TICGACTGC TICGACTGC TICGACTGC TICGACTGC TICGACTGC TICGACTGC TICGACTGC TICGACTGC TICGACTGC	AT ST STARS AT ST STARS	ATC SCORE C ACC SCORE ACC SCORE C ACC SCORE C	TTACCAACCI TTGCCAACCI	AAACCICICIC AAACCICICIC AAACCICICIC AAACCICICIC AAACCICICIC AAACCICICICI	TICAACAAICCA TICAACAAICCA TICAACAAICCA TICAACAAICCA TICAACAAICCA TICAACAAICCA TICAACAAICCA TICAACAAICCA TICAACAAICCA TICAACAAICCA TICAACAAICCA TICAACAAICCA TICAACAAICCA	ACG IACG IACG IACG IACG IACG IACG IACG I
C19 *MXMT CcDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT		ACATTGIAT ACATTGIAT ACATTGIAT ACATTGIAT ACATTGIAT ACATTGIAT	I GGAC 96 I GGAGIGG I GGAGIGG I GGAGIGG I GGAGIGG I GGAGIGG I GGAGIGG	AIGICIAAC AIGICIGO AIGICIGO AIGICIGO AIGICIGO AIGICIGO AIGICIGO AIGICIGO	ABCIGUE ABCIGUE ABCIGUE ABCIGUE ABCIGUE ABCIGUE ABCIGUE ABCIGUE	CALCAAGE IS CALCAGGE IG CALCAGGE IG CALCAGGE IG CALCAGGE IG CALCAGGE IG CALCAGGE IG	AAACO C CC AAACO C CC AAACO C C CC	CAACAAI GCA TI CAACAAI GCA TI CAACAAI GCA TI CAACAAI GCA TI CAACAAI GCA TI CAACAAI GCA	TACA TACA TACA TACA TACA TACA TACA

	271	280	290	300	310	320	330	340	350	360
E2 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	GGAAT I GGAAT I AGAAT I AGAAT I AGAAT I AGAAT I	GITGCGG GITGCGG GITGCGG GITGTGC GITGTTGC GITGITGC	CCAACTIGC CCAACTIGC CCAACTIGC CCAACTIGC CCAACTIGC	CCAACATCAACA CCAACATCAACA CCAACATCAACA CCAACATCAACA CCACATCAACA CCACATCAACA CCACATCAACA	AGIGCALI AGIGCALI AGIGCALI AGIGCALI AGIGCALI AGIGCALI	AAA CTT CC C AAA CTT CC CC AAA CTT CC CC AAA CTT CC CC AAA CTT CC CC AAA CTT CC CC	ATTIGGATG ATTIGGATG ATTIGGATG ATTIGGATG ATTIGGATG ATTIGGATG		ACCAACACA ACCAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA	CTTTAAC CTTTAAC CTTTAAC CTTTAAC CTTTAAC CTTTAAC
CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C5 *MXMT C6 *MXMT C1 *MXMT C11 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT C16 *MXMT C17 *MXMT C18 *MXMT C19 *MXMT	A (A A A (A A)) A (A A) A (A) A (A) (A) A (A) (A) (A) (A) (A) (A) (A) (A) (A) (A		CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING CCAACING	CCAACA CCAACA<	AG (CA) AG (CA) AC (CA) AC (CA) AG (CA) AC (CA)	ARAG I COS ARAG I COS ARAG I COS ARAG I COS ARAC I COS	Laiti GGA d Alti GGA G Alti GGA G		ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA	CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC CITTARC
CcDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT	agaal 1 Agaal 1 Agaal 1 Agaal 1 Agaal 1 Agaal 1 Agaal 1 Agaal 1	GI IGCEG GIIGCEG GIIGCEG GIIGCEG GIIGCEG GIIGCEG	CCAACITEC CCAACITEC CCAACITEC CCAACITEC CCAACITEC CCAACITEC	CCAACA CAACA CCAACA CAACA CCAACA CAACA CCAACA CAACA CCAACA CAACA CCAACA CAACA	AGEGCETT AGEGCETT AGEGCATT AGEGCATT AGEGCATT AGEGCATT	AAAGTIGGG AAAGTIGGG AAAGTIGGG AAAGTIGGG AAAGTIGGG AAAGTIGGGG	A PTTGGCATG ATTTGGCATG ATTTGGCATG ATTTGGCATG ATTTGGCATG ATTTGGCATG		ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ACCAAACACA ATCAAACACA	TTTCAAC TTTTCAAC TTTTAAC TTTCAAC TTTCAAC TTTCAAC CTTTCAAC
E2 CcXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	361 5- Act cl Act cl Act cl Act cl Act cl Act cl Act cl Act cl Act cl	370 	380 -5 GICCAAAGIA GICCAAAGIA GICCAAAGIA GICCAAAGIA GICCAAAGIA	390 -5	400 	410 S	420 5 PAL BGAACC RAT BGAACC RAT BGAACC RAT BGAACC RAT BGAACC RAT BGAACC RAT BGAACC	430 51- CCCACCA CCCACCA TCCCACCA TCCCACCA TCCCACCA CCCCACCA	440 	450 51 CTCAALCA CTCAALCA CTCAALCA CTCAALCA CTCAALCA CTCAALCA CTCAALCA
CcMXMT1 CaMXMT1 CaMXMT2 C1 *MKMT C2 *MKMT C4 *MKMT C5 *MKMT C6 *MKMT C10 *MKMT C11 *MKMT C13 *MKMT C14 *MKMT C14 *MKMT C16 *MKMT C19 *MKMT	AGTGCI AGTIC AGTIC AGTIC AGTIC AGTIC AGTIC AGTIC AGTIC AGTIC AGTIC AGTIC AGTIC AGTIC	PGGRCATT PGGRCATT PGGRCATT PGGRCATT PGGRCATT PGGRCATT PGGRCATT PGGRCATT PGGRCATT PGGRCATT PGGRCATT PGGRCATT		I I GACAAAG I I G II GACAAAG I G II GACAAAG I G II GACAAAG I G II GACAAAG I G II GACAAAC I G		AGAGAAGAA AGAGAAGAA AGAGAAGAA AGAGAAGAA AGAGAAGA	IAI DALACI IAITAAACI IAITAAACI IAITAAACI IAITAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI IAITAAAACI			CTORATOR CTORATOR CTORATOR CTORATOR CTORATOR CTORATOR CTORATOR CTORATOR CTORATOR CTORATOR CTORATOR CTORATOR CTORATOR CTORATOR CTORATOR
CcDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT	AGTIC AGTIC AGTIC AGTIC AGTIC	SGGACATT SGGACATT SGGACATT SGGACATT SGGACATT	CTACAAAGTA GTACAAAGTA GTACAAAGTA CTACAAAGTA GTACAAAGTA GTACAAAGTA	CACRAROUS CACRAROUS CACRAROUS CACRAROUS CACRAROUS CACRAROUS	CCAGGA CCAGGA CCAGGA CCAGGA CCAGGA CCAGGA	AAAGAAGAA AAAGAAGAA AAAGAAGAA AAAGAAGAA AAAGAAG	AATTA AAG AATTA AAG AATTA AAG AATTA AAG AATTA AAG AATTA AAG	CCCACCA TCCCACCA TCCCACCA TCCCACCA TCCCACCA TCCCACCA	CAGAIITII TCAGAIITII CAGAIITII CAGGIIIII CAGGIIIII CAGGIIIII	CIGAAIGA CIGAAIGA CIGAAIGA CIGAAIGA CIGAAIGA CIGAAIGA

	451	460	470	480	490	500	510	520	530	540
E2 CcXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT		CCCAAAT GA TCCCAAAT GA TCCCAAAT GA TCCCAAAT GA TCCCAAAAT GA TCCCAAAAT GA	ITTCAATICGG ITTCAATICGG ITTCAATICGG ITTCAATICGG ITTCAATICGG ITTCAATICGG	TTTCAAGT TTTCAAGT TTTCAAGT TTTCAAGT TTTCAAGT	ICCI CCAACC ICCI CCAACC ICCI CCAACC ICCI CCAACC ICCI CCAACC ICCI CCAACC	C ACC CA C ACC CA C ACC CA C ACC CA C ACC CA C ACC CA C ACC CA	AACTIGAGA AACTIGAGA AACTIGAGA AACTIGAGA AACTIGAGA AACTIGAGA	AACAAAAICC AACAAAAICC AACAAAAICC AACAAAAICC AACAAAAICC AACAAAAICC AACAAAAICC	AC SCARARIA AC SCARARIA AC SCARARIA AC SCARARIA AC SCARARIA AC SCARARIA AC SCARARIA	GGATC GGATC GGATC GGATC GGATC GGATC
CoMXMT1 CaMXMT2 CaMXMT2 C1 *MXMT C2 *MXMT C3 *MXMT C6 *MXMT C6 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT C16 *MXMT C17 *MXMT C19 *MXMT			III CAAIICGC III CAAIICGC		CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC CCICCAACC			AACAAAA CG AACAAAA CG	ICOCRAGATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA ACOCARATA	GGA C GGA C
CoDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT		CCAAAA GA CCAAAA GA CCAAAA GA CCAAAA GA CCAAAA GA CCAAAA GA	I CAATICGG ITTCAATICGG ITTCAATICGG ITTCAATICGG ITTCAATICGG ITT <mark>CAATIC</mark> GG	CAAS CAAS CAAS CAAS CAAS CAAS	CICCAAC CICCCAAC CICCCAAC CICCCAAC CICCCAAC CICCAAC	C ACC -CA C ACC -CA ACC -CA ACC -CA ACC -CA ACC -CA	APCTICACA AICTICACA AACTICACA AACTICACA AACTICACA AACTICACA	AAGAAAA GO AAGAAAA GO AAGAAAA GO AAGAAAA GO AAGAAAA GO AAGAAAA GO	ACICAAAA IA ACICAAAAIA ACICAAAAIA ACICAAAAIA ACICAAAAIA ACICAAAAIA ACICAAAAIA	GGATC GGATC GGATC GGATC GGATC GGATC
E2 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	541 I5 CTGCC CTGCC CTGCC CTGCC CTGCC CTGCC	550 Sana Cocco Tantaccocc Tantaccocc Tantaccocc Tantaccocc Tantaccocc Tantaccocc Tantaccocc Tantaccocc	560 	570 	580 5	590 	600 CCATGCAIT CCATGCAIT CCATGCAIT CCATGCAIT CCATGCAIT CCATGCAIT CCATGCAIT	610 I S TTTTACACTC TTTTACACTC TTTTACACTC TTTTACACTC TTTTACACTC	620 	630 CTICA CTICA CTICA CTICA CTICA CTICA CTICA
CoMXMT1 CaMXMT1 CaMXMT2 C1 *NXMT C2 *NXMT C4 *NXMT C5 *NXMT C6 *NXMT C10 *MXMT C11 *MXMT C13 *MXMT C13 *MXMT C13 *MXMT C16 *MXMT C16 *MXMT C17 *MXMT C19 *MXMT C19 *MXMT	CLOCK CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC	TAN AR COC TANTAR COC	AA GCCI CGCI AA GCCI CGCI				CCALCATT CCALCATT CCALCATT CCALCATT CCALCCATT CCALCCATT CCALCCATT CCALCCATT CCALCCATT CCALCCATT CCALCCATT CCALCCATT CCALCCATT CCALCCATT CCALCCATT CCALCCATT	IIIIGCAGIC ITTIGCACIC ITTIGCACIC ITTIACACIC ITTIACACIC ITTIACACIC ITTIACACIC ITTIACACIC ITTIACACIC ITTIACACIC ITTIACACIC ITTIACACIC ITTIACACIC ITTIACACIC ITTIACACIC	I GILACASI TGTIACASI TGTIACASI TGTIACASI TGTIACASI TGTIACASI TGTIACASI TGTIACASI TGTIACASI TGTIACASI TGTIACASI TGTIACASI TGGIACASI TGGIACASI TGGIACASI	GIICA GIICA AIICA CIICA AIICA AIICA AIICA AIICA AIICA AIICA AIICA AIICA AIICA AIICA AIICA
CoDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT	GIGCO GIGCO AIGCO GIGCO GIGCO	TGATAGGCGCI TGATAGGCGCI TGATAGGCGCI TGATAGGCGCI TGATAGGCGCI TGATAGGCGCI	AALGCCCGGCCC AALGCCCCGGCCC AALGCCTGGCCT AACGCCTGGCCT AALGCCTGGCCT	CITCIACA CITICIACA CITICIACG CITICIACG CITICIACG CITICIACG CITICIACG	CA SACICUICO CA SACICUICO CA SACICUICO CA SACICUICO CA SACICUICO CA SACICUICO CA SACICUICO	CCCACCACACI CCCCACCACI CCCCACCACI CCCCACCACI CCCACCACI CCCACCACI	CCAIGCAIT CCAIGCAIT CCAIGCAIT CCAIGCAIT CCAIGCAIT CCAIGCAIT	ITTACACIC ITTACACIC ITTACACIC ITTACACIC ITTACACIC ITTACACIC	IIGIIACIGI IIGIIACIGI IIGIIACIGI IIGIIACIGI IIGIIACIGI IIGIIACIGI	TIGCA TIGCA TIGCA TIGCA TIGCA TIGCA

F2/T2	631 5	640	650 5	660 15-	670 15-	680 !!	690 515	700 5-	710 [5
CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C12 *XMT C15 *XMT	ATGGTT ATGGTT TTGGTT TTGGTT TTGGTT TTGGTT	ATCTCAGGT ATCTCAGGT ATCTCAGGT ATCTCAGGT ATCTCAGGT ATCTCAGGT ATCTCAGGT	TTTGAGTTA TTTGAGTTA TTTGAGTTA TTTGAGTTA TTTGAGTTA TTTGAGTTA	ALCCCTTA ALCCCTTA ALCCCTTA ALCCCTTA ALCCCTTA ALCCCTTA ALCCCTTA	CTT - TTAA CTT - TTAA CTT - TTAA CTT - TTAC CTT - TTAC CTT - TTAC CTT - TTAC	TITICIIG ITTICIIG ITTICIIG ITTICIIG ITTICIIG ITTICIIG ITTICIIG	ACCAAAAATA ACCAAAAATA ACCAAAAATA ACCAAAAATA ACCAAAAATA ACCAAAAATA ACCAAAAATA	- CAICAICA - CTICAICAICA - CTICAICAICA - CTICAICAICA - CTICAICAICA	CA TC TCA TC
CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C5 *MXMT C6 *MXMT C10 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C14 *MXMT C16 *MXMT C17 *MXMT C18 *MXMT C19 *MXMT					C 19 - 200 AC C 19 -		а салалта а салалта	- GTICAIGAT - CTICAIGAT - CTICAIGAT - GTICAIGAT - GTICAIGAT	
CoDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT	TTGGTT TTGGTT TTGGTT TTGGTT TTGGTT TTGGTT TTGGTT	ADCICAGGI ACCICAGGI ACCICAGGI ACCICAGGI ACCICAGGI ACCICAGGI ACCICAGGI	TTIGAGTIA TTIGAGTIA TTIGAGTIA TTIGAGTIA TTIGAGTIA	A CCC C A ACCCC C A A CCC C A		IIIICIIG IIIICIIG IIIICIIG IIIICIIG IIIICIIG IIIICIIG IIIICIIG	ACCAAAAATA Recaaaaata Recaaaaata Recaaaaata Recaaaaata Recaaaaata Recaaaaata	AGTICACGATI AGTICACGATI - GTICACGATI - GTICACGATI - GTICACGATI - GTICACGATI	177817C 1778877C 1778877C 1778877C 1778877C 1778877C
I2 CoXMT1 CaXMT1 CAXMT2 C3 *XMT C12 *XMT C15 *XMT	721 I5 CATTAC CATTAC CATTAC CATTAC CATTAC	730 AACIAL GO AACIAL GO AACIAL GO AACIAL GO AACIAL GO AACIAL GO AACIAL GO	740 I5 ACSGAAATIT ACSGAAATIT ATGGAAATIT ATGGAAATIT ATGGAAATIT ATGGAAATIT	750 	760 	770 !! CCACAGGAI CCACAGGAI CCACAGGAI CCACAGGAI CCACAGGAI	780 55 AL CCAAGAAAG AL CCAAGAAAG AL CCAAGAAAG AL CCAAGAAAG AL CCAAGAAAG CCAAGAAAG CCAAGAAAG	790 	800
CoMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C3 *MXMT C4 *MXMT C5 *MXMT C6 *MXMT C10 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C14 *MXMT C16 *MXMT C16 *MXMT C16 *MXMT C18 *MXMT C19 *MXMT	CATTAG CATTAG CATTAG CATTAG CATTAG CATTAG CATTAG CATTAG CATTAG CATTAG CATTAG CATTAG CATTAG	AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC AACTATICC	A GGAAATTT A GGAAATTT		C 3.3.3.4.0.3.1.2.1 C C.3.1.0.3.1.2.1 C C.3.1.0.3.1.3.1	CCACAGA CCACAGA CCACAGA CCACAGA CCACAGA CCACAGA CCACAGA CCACAGA CCACAGA CCACAGA CCACAGA CCACAGA CCACAGA	A CCAA CAAAC A CCAA CAAAAC A CCAA CAAAAC A CCAA CAAAAC A CCAA CAAAAC A CCAA CAAAAC A CCAA CAAAAC A CCAA CAAAAAAA A CCAA CAAAAAAAA	AGATTICIGA AGATTICIGA AGATTICIGA AGATTICIGA AGATTICIGA AGATTICIGA AGATTICIGA AGATTICIGA AGATTICIGA AGATTICIGA AGATTICIGA AGATTICIGA AGATTICIGA	ACAAACTAA IA CAAACTAA IA CAAACTAA IA ACAAAATAA IA
CoDXMT CaDXMT2 CaDXMT1 P1 ADVMT	CATTAG CATTAG	AACIATOCI AACIATOCI		CTTTAATAATI CTTTAATAATI CTTTAATAATI	CTCAAGATAT CTCAAGATAT	CCACAGGA CCACAGGA CCACAGGA	ATCCAAGAAAG ATCCAAGAAAG ATCCAAGAAAG	AGATTICEGAC AGATTICEGAC AGATTICEGA	CCAAACTAATA CCAAACTAATA ACAAACTAATA

	811	820 15	830 ; !	840 5	850 -515-	860 15-	870 	880 51	890 515	900
I2 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	AITTIA AITTIA AITTIA AITTIA AITTIA AITTIA AITTIA	II CAAGICG II CAAGICG II CAAGICG II CAAGICG II CAAGICG	H SCHCALS H GCHCALS H GCHCALS H GCHCALS H GCCHCALS H GCCHCALS H GCCHCALS	AITTATATI AITTATATI AITTATATI AITTATATI AITTATATI AITTATATI	CCCACATGCAACA CCCACATGCAACA CCCACATGCAACA CCCACATGCAACA CCCACATGCAACA CCCACATGCAACA	СТААСАААА СТААСАААА СТААСАААА СТААСАААА СТААСАААА СТААСАААА СТААСАААА	GALCCAAC GALCCAAC GALCCAAC GALCCAAC GALCCAAC GALCCAAC	TATATAAGTI TATATAAGTI TATATAAGTI TATATAAGTI TATATAAGTI TATATAAGTI	ACCASCICIOS ACCASCICIOS ACCASCICIOS ACCASCICIOS ACCASCICIOS ACCASCICIOS	ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC
CoMXMT1 CaMXMT1 CaMXMT2 C1 *MKMT C2 *MKMT C2 *MKMT C5 *MKMT C6 *MKMT C6 *MKMT C10 *MKMT C11 *MKMT C13 *MKMT C14 *MKMT C14 *MKMT C16 *MKMT C18 *MKMT C18 *MKMT C19 *MKMT	A LITTA A LITTA	E JAAC C JAAC C C CAAC C C	T GGC CAIG T GGC CAIG		CCCACH CANCA CCCACH CAACA CCCACH CAACA	C TARCAAAA C TARCAAAA	CA CCAAC CA CCAAC	IA A AAC I IAAA AAC I	ACCACITIC AC ACCACITIC AC	ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC ACCIC
CcDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT	ATTTA ATTTA ATTTA ATTTA ATTTA ATTTA ATTTA	TTTAATIO TTTAATIO TTTAATIO TTTAATIO TTTAATIO TTTAATIO	CIGCICATO CIGCICATO CIGCICATO CIGCICATO CIGCICATO CIGCICATO CIGCICATO	ATTIATATI ATTIATATI ATTIATATI ATTIATATI ATTIATATI ATTIATATI	CCCACAL SCAACA CCCACAL GCAACA CCCACAC CCAACA CCCACAC CCAACA CCCACAC CCAACA CCCACAC CCAACA	C AACAAAA C AACAAAA C BACABAA C BACABAA C BACABAA C BACABAA C BACABAA	GAL CCAAC GAL CCAAC GAL CCAAC GAL CCAAC GAL CCAAC GAL CCAAC	A A AAC A A AAC A A AAC A A AAC A A AAC A A AAC A A AAC	ACCACCICIAC ACCACCICIA ACCACCICIA ACCACCICIA ACCACCICIA ACCACCICIA	ACGTG ACGTG ACGTG ACGTG ACGTG ACGTG ACGTG
	901	10000								
	15	910 !5	920	930	940	950	960	970	980	990
12/E3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	I5- CASST CASST CASST CASST CASST CASST CASST CASST CASST CASST	910 	920 ; ; TTGET GACT TTGET GACT TTGET GACT TTGET GACT TTGET GACT	930 5 IAATCGGGG IAATTGGGG IAATTGGGG IAATTGGGG IAATTGGGG IAATTGGGG	940 -5I5- COCOCOCOCOCO ALCACIACAACA ALCACIACAACA ALCACIACAACA ALCACIACAACA ALCACIACGAACA ALCACIACGAACA	950 AAGCGAGCA AAGCGAGCA AAGCGACCA AAGCGACCA AAGCGACCA AAGCGACCA	960 ILACICII ILACICII ILACICII ILACICII ILACICII ILACICII ILACICII	970 S I CCAAAG CAAG CCAAAG CAAG CCAAAG CAAG CCAAAG CAAG CCAAAG CAAG	980 5 5 10000000000 10010100000 1001010000 1001010000 1001010000 1001010000 1001010000	990 i TCCAG TCCAG TCCAG TCCAG TCCAG TCCAG
12/E3 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT C15 *XMT CaMXMT1 CaMXMT1 CaMXMT1 C4 *MXMT C4 *MXMT C4 *MXMT C4 *MXMT C1 *MXMT C10 *MXMT C10 *MXMT C11 *MXMT C14 *MXMT C16 *MXMT C16 *MXMT C16 *MXMT C16 *MXMT C16 *MXMT C19 *MXMT	CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID CASCID	910 910 910 910 910 910 910 910	920 	930 5	940 -5	950 		970 S		990

	991	1000	1010	1020	1030	1040	1050	1060	1070	1080
-	15-	5	515	15		5	-5	5!!	5	5
E 3										
CaXMT1	110001	TTTTCCT	CARTTACCA	A A PATTTAC	CACATTE	TARCATTCA	TTCCCARCACT	TOTTTCACE	COCC31 C	
CaXMT2	AAGCC	TATTICCA	CAATTTACCA	AAGATTTTAC	CACATTT	TAACGATTCA	TTCGGAACAGT	TGTTTTCACA	IGCCCAATE	CICCII
C3 *XMT	AAGGCZ	TIGGGA	CAATTTACCA	aa <mark>ga</mark> ttti <mark>a</mark> c	CACATTE	TAGGGATTCA	TT <mark>C</mark> GG <mark>AA</mark> G <mark>A</mark> GT	TGTTTTCAC	IGGCC GAATG	CICCII
C12 *XMT	AA <mark>GG</mark> CF	ATTTGGA	CARTTTACCA	aa <mark>ga</mark> ttti <mark>a</mark> c	CACATTT	TAAGGAITCT	TT <mark>CGGAA</mark> GAGT	TGTTTT <mark>CACA</mark>	IGG <mark>CC</mark> GAA <mark>T</mark> G	CICCII
C15 *XMT	AAGGCZ	ATTGGGA	CARTTEACGA	AAGATTTTAC	CACATIT	TAAGGAITCT	TT <mark>CGGAA</mark> GAGI	TGITTICACA	IGGCCGAAIG	CICCII
0.100.004	-									
CAMXMT1	AACCC	TTTCCL	CANTTACCA	AACATTTAC	CACATTE	TARCATTCA	TTCCAABCACT	TOTTTTCZCO	TCCCCCAATC	
CaMXMT2	AAGGC	TATTICCA	CARTTTACCA	AAGATTTTAC	CACATTE	TAAGGATTCA	TTCGAAAGAGT	TETTTECACE	IGCCGAATG	CICCIT
C1 *MXMT	AAGGC	ATTTCCAT	CAATTEACCA	AAGATTTTAC	CACATTE	TAAGGATTCA	TTCGAAAGAGT	TGTTTT <mark>CAC</mark> G:	FGCCGAATG	CICCIT
C2 *MXMT	AA <mark>GG</mark> CF	ATTTGGAT	CARTTEACGA	aa <mark>ga</mark> tttt <mark>a</mark> c	CACATT	C <mark>TAAGGAIT</mark> CA	TT <mark>CGAAAG</mark> AGT	TGTTTT <mark>CAC</mark> G	IGG <mark>CC</mark> GAA <mark>T</mark> G	CICCII
C4 *MXMT	AAGGCF	ATTIGGAT	CAATTTACGA	AAGATTTTAC	CACATT	TAAGGAITCA	TTCGAAAGAGI	TGTTTT <mark>CAC</mark> G	IGGCCGAATG	CICCII
C5 *MXMT	AAGGCF	TATTICCA.	CAATTEACEA	AAGATITIAC	CACATTE	TAAGGAITCA	TICGAAAGACI	TGITITCACC:	IGECCEAAIG	CICCII
C6 *MXMI	ARGGUP		CARLINAC SA	A A CALLER AC	CACATIN	ARGORISCA	TROCALACT	TOTITICACO.	I GGCCGAALG	
C10 *MXMT	AACCC	TATTCCAT	CARTTRACCA	AACATTTAC	CACATTT	TAACCATTCA	TTCCARACACT	TETTTTCACE	TEGCCGAATE	C C C C C C C C C C C C C C C C C C C
C11 *MXMT	AAGGC	TATTTCCAT	CAATTTACCA	AAGATTTTAC	CACATIT	TAACGATTCA	TTC GAAAGAGI	TGTTTTCACG	FGGCCGAATG	CICCII
C13 *MXMT	AA <mark>GG</mark> C7	TATTECA	CAATTTACGA	AA <mark>GA</mark> TTTT <mark>A</mark> C	CACATTI	TAAGGATTCA	TT <mark>CGAAA</mark> GAGT	TGTTTT <mark>CAC</mark> GI	IGCCCAAIG	CICCII
C14 *MXMT	AAGGTA	ATTTGGAT	CAATTTAC CA	AA <mark>GA</mark> TTTT <mark>A</mark> C	CACATTT	C <mark>TAAGGATT</mark> CA	TT <mark>CGAAA</mark> GAGI	TGITII <mark>CAC</mark> GI	IGGCCGAATG	CCCCTT
C16 *MXMT	AAGGCF	TATTICCA	CAATTEACEA	AAAATTTTAC	CACATT	TAAGGAITCA	TICGAAAGAGI	TGITITCACC	IGECCEAATE	CICCII
C10 +MXMT	AAGGCA		CARLES ACGA	AAGALIIIAC	CACATIT	ARCORT CA	TICGAAAGAGI	TGITTTCACS.	IGGCCGAAIG	
C19 *MXMT	22CCC	TATTCCAT	CARTTRACCA	AACATTTAC	CACATTT	TAACCATTCA	TTCCASACACT	TOTTTTCLCS	ISCCCARTS	
CCDXMT	AAGGCA	TATTTCCA	CAATTTACCA	AAGATTTTAC	CACATT	IGAGGATICA	TT <mark>CGGAAGA</mark> GI	TGATTTCACC	ICCCCCAA IC	CECCE
CaDXMT2	AA <mark>GG</mark> CF	TATTEGAI	CAATTTAC CA	AA <mark>GA</mark> TTTTAC	CACATTT	C <mark>TGAGGATT</mark> CA	TT <mark>CGGAA</mark> GAGI	TG <mark>A</mark> TTTCAC6	IGCCCCAATC	CICCII
CaDXMT1	AAGGCI	ATTICCAT	CAATTEACGA	AAGATTTTAC	CACATTE	TEAGGATECA	TICGGAAGAGI	IGATIICACG	IGGCCGAAIG	CICCII
R1 *DXMT	AAGGCH		CARITERCER	AAGAITTIAC	COCATTIC	TRACCEL CA	TIC GGAAGAGI	TORITICAC SI	IGGCCGAAIG	
R4 +DXMT	110001	TETTOSET	CARTTRACTA	A ACA TTTAC	CACATTE	TTACCETTCA	TTCCCARCACT	TOATTTCLCS	TCCCC22 TC	
	1081	1090	1100	1110	1120	1130	1140	1150	1160	1170
F2/T2	1081	1090	1100 5 5	1110 5	1120	1130 5	1140 -5I	1150 5!!	1160 5	1170 51
E3/I3	1081 5- 000000	1090 5	1100 5 5	1110 5 000000000		1130 5	1140 -5	1150 5 0000000000	1160 5	1170 5) 0
E3/I3 CoXMT1	1081 5- DDCDDC	1090 	1100 5 5 10000000000	1110 5	1120	1130 5 00000000000000000000000000	1140 -5	1150 5!!	1160 5	1170 51 0
E3/I3 CoXMT1 CaXMT1		1090 5 5 5 5 	1100 5I5 CGACTICAAT				1140 -5 GAGA COCAR GAGA COCAR	1150 5	1160 5 1000000000 CETTCITCAC	1170 51 0
E3/I3 CcXMT1 CaXMT1 CaXMT2	1081 J5- DCCDCC ACTIGO ACTIGO	1090 9 9 9 	1100 SI5 DECODODODO DECODODODO DECODODODO DECODODODO DECODODODO DECODODODO DECODODODO DECODODODO DECODODODO DECODODODO DECODODO DECODODO DECODODO DECODODO DECODODO DECODODO DECODODO DECODODO DECODE		1120 	1130 5 MAGACTIACT AGACTIACT	1140 -5 ICAGALGOCAA ICAGALGOCAA ICAGALGOCAA	1150 5 ! TARACCACTT TARACCACTTC TARACCACTTC	1160 5 5 5	1170 51 0 GTTAIC GTTAIC
E3/I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT		1090 	1100 S	1110 J5 TACACCCCCC TACACCCCCC TCCACCCCCCCCCCCC	1120 	1130 	1140 -SI TGAGALGGCAA TGAGALGGCAA TGAGALGGCAA TGAGALGGCAA	1150 5	1160 5 SCITCIICAC SCITCIICAC SCITCIICAC SCITCIICAC	1170 51 D GTIAIC GTIAIC GTIAIC GTIAIC
E3/I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	1081 J5- DDDDD ACTIGO ACTIGO ACTIGO ACTIGO	1090 	1100 			1130 	1140 -SI TGAGALGCCAA TGAGALGCCAA TGAGALGCCAA TGAGALGCCAA TGAGALGCCAA	1150 5		1170 5I GTIAIC GTIAIC GTIAIC GTIAIC GTIAIC GTIAIC
E3/I3 CcXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	1081 J5- DCCCC ACTIC ACTIC ACTIC ACTIC	1090 	1100 	1110 5 TAGACCCCCC TAGACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		1130 	1140 -5	1150 5	1160 5 SCTTGTTGAC SCTTGTTGAC SCTTGTTGAC SCTTGTTGAC SCTTGTTGAC	1170 51 CTTAIC CTTAIC CTTAIC CTTAIC CTTAIC CTTAIC CTTAIC CTTAIC
E3/I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CaMXMT1	1081 I5- DODDDC ACTICO ACTICO ACTICO ACTICO ACTICO ACTICO	1090 IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	1100 			1130 SI A A SAC TI ACT A A SAC TI ACT A SAC TI ACT	1140 -5i CAGALGGCAA TGAGALGGCAA IGAGALGGCAA IGAGALGGCAA IGAGALGGCAA		1160 5 CONCOLOGIC	1170 5 GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC
E3/I3 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1		1090 ATTICIAAR CATTICIAAR CATTICIAAR CATTICIAAR CATTICIAAR CATTICIAAR CATTICIAAR	1100 SI5 CCACIT CAAT CCACIT CAAT CCACIT CAAT CCACIT CAAT CCACIT CAAT CCACIT CAAT	1110 TATACCCCC TATACCCCC CAACCCCC CAACCCCC CAACCCCC CAACCCCC CAACCCCC CAACCCCCC	CAAT CCC CAAT CCC CAAT CCC CAAT CCC CAAT CCC CAAT CCC CAAT CCC CAAT CCCC	1130 SI AAAACTIACT AAAACTIACT AAAACTIACT AAAACTIACT AAAACTIACT AAAACTIACT	1140 -5 ICAGALGOCAA ICAGALGOCAA ICAGALGOCAA ICAGALGOCAA ICAGALGOCAA ICACALGOCAA	1150 5	1160 SETCTICA SETCTICA SETTCTICA SETTCTICA SETTCTICA SETTCTICA SETTCTICA SETTCTICA SETTCTICA	1170 5 CTAC CTAC CTAC CTAC CTAC GTAC GTAC GTAC
E3/I3 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT2	ACTICO ACTICO ACTICO ACTICO ACTICO ACTICO ACTICO ACTICO ACTICO	1090 ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA	1100 	1110 AGACGCCCC AGACGCCCCCCCCCCCCCCCCCCCCCC	1120 	1130 5I A A SACTACT A A SACTACT A A SACTACT A SACTACT A SACTACT A SACTACT A SACTACT A SACTACT A SACTACT	1140 -5 ICACALCCCAP ICACALCCCAP ICACALCCCAP ICACALCCCAP ICACALCCCAP ICACALCCCAP ICACALCCCAP ICACALCCCAP	1150 5	1160 SETCII dag SETCII dag SETICII dag SETICII dag SETICII dag SETICII dag SETICII dag SETICII dag SATICII dag	1170 5 CTIAC CTIAC CTIAC CTIAC CTIAC CTIAC CTIAC GTIAC GTIAC GTIAC GTIAC
E3/I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT	1081 5- DODDC ACTIC ACTIC ACTIC ACTIC ACTIC ACTIC ACTIC	1090 	1100 LOCACTICAAT CCACTICAAT CCACTICAAT CCACTICAAT CCACTICAAT CCACTICAAT CCACTICAAT CCACTICAAT CCACTICAAT CCACTICAAT	1110 AGACSCCCC CGACSCCCC CGACSCCCC CGACSCCCC CGACSCCCC CGACSCCCC CGACSCACCCC CGACSCACCCC CGACSCACCCCC CGACSCACCCCC CGACSCACCCCCCCCCC	1120 CAAT GCC CAAT GCC		1140 -5 I CACAL COCAR I CACAL COCAR	1150 5	1160 5 CONCONCONCONC CONTICUE CONTICUE CONTICUE CONTICUE CANTICUE CANTICUE CANTICUE CANTICUE CANTICUE CANTICUE CANTICUE CONTICUE CANTICUE CONTI	1170 5 GTIAC GTIAC GTIAC CTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC
E3/I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT	1081 I5- DODDDD ACTICO ACTICO ACTICO ACTICO ACTICO ACTICO ACTICO ACTICO ACTICO	1090 	1100 LICO CACTICAA CCACTICAA CCACTICAA CCACTICAA CCACTICAA CCACTICAA CCACTICAA CCACTICAA CCACTICAA CCACATICAA CCACATICAA CCACATICAA CCACATICAA CCACATICAA CCACATICAA		1120 CAALGCCI CAALGCCI CAALGCCI CAALGCCI CAALGCCI CAALGCCI CAALGCCI CAALGCCI CAALGCCI CAALGCCI CAALGCCI CAALGCCI CAALGCCI CAALGCCI CAALGCCI	1130 	1140 -5		1160 CCTISITICAC SCTISITICAC SCTISITICAC SCTISITICAC SCTISITICAC SCTISITICAC SCTISITICAC SCTISITICAC SCTISITICAC SCTISITICAC	1170 5 GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC
E3/I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 C1 *MXMT C1 *MXMT C2 *MXMT C4 *MXMT		1090 ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA	1100 	1110 TAGACCCCC AGACCCCCC CGACCCCCCCCCCCCCCCC	1120 CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC CAALGCC		1140 -5		1160 SCTICTICAC SCTICTICAC SCTICTICAC SCTICTICAC SCTICTICAC SCTICTICAC SCTICTICAC SCTICTICAC SCTICTICAC SCTICTICAC SCTICTICAC	1170 51 GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC
E3/I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C5 *MXMT	1081 I5 ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC	1090 ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA	1100 	1110 TAGACCCCC TAGACCCCCC CGACCCCCCCCCCCCCCCCCCCCCCCCCC	1120 CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI CAATGCCI		1140 -5	1150 5	1160 SCTICITCAC SCTICITCAC SCTICITCAC SCTICITCAC SCTICITCAC SCTICITCAC SCTICITCAC SCTICITCAC SCTICITCAC SCTICITCAC SCTICITCAC SCTICITCAC	1170 51 GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC
E3/I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CaMXMT1 CaMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C2 *MXMT C5 *MXMT C6 *MXMT C9 *MXMT	1081 I5- DOCODE ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC ACTICC	1090 CATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA ATTICIAA	1100 	1110 IAAACCCCC IAAACCCCC ICAACCCCC ICAACCCCC ICAACCCCC ICAACCCCC ICAACCCCC ICAACCAAC	1120 CAAT CCC CAAT CCC		1140 -5	1150 5	1160 SCTCTTCAC SCTCTTCAC SCTCTTCAC SCTCTTCAC SCTCTTCAC SCTCTTCAC SCTCTTCAC SCTCTTCAC SCTCTTCAC SCTCTTCAC SCTCTTCAC SCTCTTCAC	1170 51 CTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC
E3/I3 CcXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CaXXMT1 CaMXMT1 CaMXMT1 CaMXMT1 C2 *MXMT C4 *MXMT C5 *MXMT C6 *MXMT C9 *MXMT		1090 	1100 	1110 TARACCCCC TARACCCCCC TARACCCCCC TCRACCCCCC TCRACCCCCC TCRACCCCCCC TCRACCACCCC TCRACCACCCC TCRACCAACCACCACC TCRACCAACCACCACCACCACCACCACCACCACCACCACCA	1120 CAAT CCC CAAT CCCC CAAT CCCC CCCC	1130 	1140 -5	1150 5	1160 SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC SCITCICAC	1170 5 GTIAC GTIAC CTIAC CTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC
E3/I3 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT1 C4 *MXMT C5 *NXMT C6 *MXMT C6 *MXMT C1 *MXMT C1 *MXMT C1 *MXMT	1081 I5- DDDDDD ACTIC ACTIC ACTIC ACTIC ACTIC ACTIC ACTIC ACTIC ACTIC ACTIC ACTIC ACTIC	1090 CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA CATTICIAA	1100 	1110 TAGACSCCCC TAGACSCCCCC TAGACSCCCCC TCGACSCCCCC TCGACSCCCCCCCCCC TCGACSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	1120 CAAL CCC CAAL CCC CAAL CCC CAAL CCC CAAL CCC CAAL CCC CAAL CCC CAAL CCC CAA CCC CCC	1130 	1140 -5 ICACAR GGCAR ICACAR GGCAR	1150 5	1160	1170 5 GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC
E3/I3 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT C15 *XMT CaMXMT1 CaMXMT1 CaMXMT1 C4 *MXMT C4 *MXMT C5 *MXMT C5 *MXMT C6 *MXMT C1 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C13 *MXMT	1081 I5- I5- ACTIONAC	1090 – J – – – 5 ATTICIAAP CATTICIAAP CATTICIAAP CATTICIAAP CATTICIAAP CATTICIAAP CATTICIAAP CATTICIAAP CATTICIAAP CATTICIAAP CATTICIAAP CATTICIAAP CATTICIAAP CATTICIAAP	1100 GACTICAA GCACTICAA GCACTICAA GCACTICAA GCACTICAA GCACTICAA GCACTICAA GCACTICAA GCACTICAA GCACATICAA GCACATICAA GCACATICAA GCACATICAA GCACATICAA GCACATICAA GCACATICAA GCACATICAA GCACATICAA GCACATICAA GCACATICAA	1110 AGACSCCCC ICACSCCCC ICACSCCCC ICACSCCCC ICACSCCCC ICACSCCCC ICACSCCCC ICACSCCCC ICACSACC ICACSACC ICACSACC ICACSACC ICACSACC ICACSACC ICACSACC ICACSACC ICACSACC ICACSACC ICACSACC	1120 CARL GCC CARL GCC CARL GCC CARL GCC CARL GCC CARL GCC CARL GCC GARL GCC GARL GCC GARL GCC GARL CCC GARL CCC GARL CCC GARL CCC GARL CCC GARL CCC GARL CCC GARL CCC GARL CCC	1130 AACACITACI	1140 -S			1170 5 CTIAC
E3/I3 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C1 *MXMT C4 *MXMT C6 *MXMT C6 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT C14 *MXMT C14 *MXMT		1090 	1100 		1120 CRAILGCC CRAILGC	1130 			1160 CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG CTTGTTCAG	1170 51 GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC GTTAC
E3/I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C1 *MXMT C4 *MXMT C6 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT C14 *MXMT C15 *MXMT C14 *MXMT		1090 ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA ATTTCIAA	1100 		1120 CAAL CCC CAAL CCC				1160 CTTCTTCAC CTTCTTCAC CTTCTTCAC CTTCTTCAC CTTCTTCAC CTTCTTCAC CTTCTTCAC CTTCTTCAC CTTCTTCAC CTTCTTCAC CTTCTTCAC CCTTCTTCAC CCTTCTTCAC CCTTCTTCAC CCTTCTTCAC CCTTCTTCAC CCTTCTTCAC CCTTCTTCAC	1170 51 GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC GTIAC
E3/I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C15 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT1 C1 *MXMT C1 *MXMT C1 *MXMT C1 *MXMT C1 *MXMT C1 *MXMT C13 *MXMT C14 *MXMT C16 *MXMT C16 *MXMT C18 *MXMT		1090 ATTICIAA	1100 		1120 CAAT CCC CAAT CCCC CAAT CCC CAAT CCC				1160 SCTICTICAC	1170 51 GTIAC
E3/I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C4 *MXMT C10 *MXMT C10 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT C16 *MXMT C16 *MXMT C18 *MXMT C19 *MXMT C19 *MXMT		1090 ATTICIAA	1100 		1120 CAATGCC CCC CAATGCC CAATGCC CAATGCC CAATGCC CAATGCC CCC					1170 5I GTIALC
E3/I3 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C4 *MXMT C10 *MXMT C10 *MXMT C13 *MXMT C14		1090 ATTICIAA	1100 		1120 CAAT CCC CAAT CCC					1170 51 CTIAC CTIAC GTIAC
E3/I3 CoXMT1 CaXMT2 CaXMT2 C3 *XMT C12 *XMT C12 *XMT C12 *XMT C3 *XMT C3 *XMT C3 *XMT C4 *MXMT C4 *MXMT C4 *MXMT C4 *MXMT C1 *MXMT	1081 I5- ICIDO ACTIONA	1090 ATTICIAN	1100 GACIGAI GACIGAI GACIGAI GACIGAI GACIGAA GACIGAA GACIGAA GACIGAA GACIGAA GACIGAA GACIGAA GAA GAA GAA GAA GAA GAA GAA GAA GAA		1120 CAAT CCC CAAT CCC	1130 			1160 SCITCIICAC	1170 5 GTIAC
E3/I3 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT C15 *XMT C3 *XMT1 CaMXMT1 CaMXMT1 CaMXMT1 C4 *MXMT C2 *MXMT C5 *MXMT C1 *MXMT C1 *MXMT C13 *MXMT C13 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT C15 *MXMT C15 *MXMT C15 *MXMT C16 *MXMT C16 *MXMT C17 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT CaDXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 CADXMT2 C1 *MXMT C1 C1 *MXMT C1 *	1081 I5- ICIDOD ACTICC	1090 - J 5 5 5 5 	1100 GACTICAA GACTICAA GACTICAA GACTICAA GACTICAA GACTICAA GACTICAA GACTICAA GACTICAA GACTICAA GACTICAA GACTICAA GACTICAA GACACAA GACACAA GACACAA GACACAA GACACAA GACACAA GACACAA GACACAA GACACAA GACACAA GACACAA GACACAA GACACAA GACAA		1120 CAAL CCC CAAL CCC					1170 5
E3/I3 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT C15 *XMT C3 *XMT1 CaMXMT1 CaMXMT1 C4 *NXMT C2 *NXMT C4 *NXMT C5 *NXMT C1 *NXMT C10 *NXMT C13 *NXMT C13 *NXMT C14 *NXMT C13 *NXMT C14 *NXMT C13 *NXMT C19 *NXMT C19 *NXMT C19 *NXMT CaDXMT2 CaDXMT2 CaDXMT1 CaDXMT2 CaDXMT1 CADXMT1 CADXM	1081 I5- ACTICC	1090 	1100 		1120 CAALGCCI				1160 CTTSTTCAC	1170 5 GTIAC
E3/I3 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT C15 *XMT CoMXMT1 CaMXMT1 C4 *MXMT C1 *MXMT C1 *MXMT C1 *MXMT C1 *MXMT C14 *MXMT C14 *MXMT C14 *MXMT C13 *MXMT C15 *MXMT C19 *MXMT C2DXMT2 CaDXMT1 R1 *DXMT R1 *DXMT		1090 ATTICIAA	1100 		1120 CAAL GCC CAAL CCC CAAL CCC				1160 SCTICTICAC	1170 51 GTIAC

	1171 5-	1180 5-	1190 5-	1200 5-	1210 	1220 5-	1230	1240	1250	1260
I3 CcXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	aitict Aitict Aitict Aitict Aitict Aitict Aitict		III CAICAICI III CACCAICI III CAICAICI CI CAICAICI III CAICAICI III CAICAICI III CAICAICI	AGAIGCICATI AGAIGCICATI AGAIGCICATI AGAIGCICATI AGAIGCICATI AGAIGCICATI	SCITCITA CTICITATC CTICITATC CTICITATC CTICITATC CTICITATC CTICITA	TGAAAT GAAC TGAAAT GAAC TGAAATAAAC TGAAATAAAC TGAAATAAAC TGAAATAAAC	TACATACCIP TACATACCIP TACATACCIP CACATACCIP TACATACCIP TACATACCIP	.GC1-AGC .GC1-AGC .GC1-AGC .GC1-AGC	- TIATICA () - TIATICA () - TIATICA () - TIATICA () - TIATICA () - TIATICA () - TIATICA ()	GTIC GTIC GTIC GTIC GTIC GTIC
CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C4 *MXMT C5 *MXMT C6 *MXMT C10 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C16 *MXMT C16 *MXMT C18 *MXMT C19 *MXMT	ATTICI ATTICI ATTICI ATTICI ATTICI ATTICI ATTICI ATTICI ATTICI ATTICI ATTICI ATTICI ATTICI		T CAICA CAICI T CAICA CAICI T CAICAICA T CAICAICA T CAICAICA T CAICAICA T CAICAICA T CAICAICA T CAICAICA T CAICAICA T CAICAICA T CAICAICA	AGAIGIICAIT AGAIGICAIT AGAIGICAIT AGAIGICAIT AGAIGICAIT AGAIGICAIT AGAIGICAIT AGAIGICAIT AGAIGICAIT AGAIGICAIT AGAIGICAIT AGAIGICAIT AGAIGICAIT		ICAAA TAAAC ICAAA TAAAC	TACATACC 2 TACATACC 2 TACATACC 1 TACATACC 1		- DCATTCAC - TCALTCAC - TCALTCAC - TTATTCAA	
CoDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT	ALTICI ATTICI ATTICI ATTICI ATTICI ATTICI ATTICI 1171	CTETCTAT CTETCTAT CTETCTAT CTETCTAT CTETCTAT CTETCTAT T1180	I CAIGAICI TIGAIGAICI TIGAIGAICI TIGAIGAICI TIGAIGAICI 1190	AGATGCTCATT AGATGCTCATT AGATGCTCATT AGATGCTCTTT AGATGCTCTTT AGATGCTCTTT 1200		IGRANARAAC IGRANARAAC IGRANARAAC IGRANARAAC IGRANARAAC IGRANARAAC 1220	AGATAGCE AGATAGCE AGATAGCE AGATAGCE AGATAGCE AGATAGCE 1230	1240	-TRAILCAS -TRAILCAS -TRAILCAS CTRAILCAS CTRAILCAS CTRAILCAS 1250	2011C 2011C 2011C 2011C 2011C 3011C 3011C
I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	ATTTCT ATTTCT ATTTCT ATTTCT ATTTCT ATTTCT		CA CA C C CA CA C C CA CA C C CA CA CA C CA CA C C CA CA C C CA CA C	ACATECT CATT ACATECT CATT ACATECT CATT ACATECT CATT ACATECT CATT ACATECT CATT		IGAAA GAAC IGAAA GAAC IGAAA AAAC IGAAA AAAC IGAAA AAAC IGAAA AAAC	A GA DA GC A GA DA GC	GC - AGC GC - AGC GC - AGC GC - AGC GC - AGC	- 12 1 CA 3 - 12 A 1 CA 3 - 12 A 1 CA 3	GTIC GTIC GTIC GTIC GTIC GTIC
CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C5 *MXMT C6 *MXMT C10 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT C17 *MXMT C18 *MXMT C19 *MXMT	ATTIC ATTIC ATTIC ATTIC ATTIC ATTIC ATTIC ATTIC ATTIC ATTIC ATTIC ATTIC ATTIC ATTIC	CI C	TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA TTCATCATCA	ACAT CTTCATT ACAT CTTCATT ACAT CTTCATT ACAT CCTCATT ACAT CCTCATT		TGAARTAAAC TGAARTAAAC TGAARTAAAC TGAARTAAAC TGAARTAAAC TGAARTAAAC TGAARTAAAC TGAARTAAAC TGAARTAAAC TGAARTAAAC TGAARTAAAC TGAARTAAAC TGAARTAAAC	TACA ACC P TACA ACC P TACA ACC P TACA ACC TACA ACC		- TCATTOAC - TCATTCAC - TTATTCAC - TTATTCAC	
CoDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT	ATTTCT ATTTCT ATTTCT ATTTCT ATTTCT	CT GTCTAT CT GTCTAT CT GTCTAT CT GTCTAT CT GTCTAT CT GTCTAT	TT GAT GAT CI TT GAT GAT CI	AGATGCICATI AGATGCICATI AGATGCICATI AGATGCICITI AGATGCICITI	SCITCICATC SCITCICATC SCITCITATC SCITCITAT SCITCITAT	I GRAATAAAC I GRAATAAAC I GRAATAAAC I GRAATATAC I GRAATATAC	TAGATAGCIA TAGATAGCIA TAGATAGCIA TAGATAGCIA TAGATAGCIA	.GC .GC .GCT-AGC .GCT-AGCTEG .GCT-AGCTEG	-TTATTCAGO -TTATTCAGO -TTATTCAGO CTTATTCAGO CTTATTCAGO	GTIC GTIC GTIC GTIC GTIC

	1351	1360	1370 515-	1380	1390	1400	1410	1420	1430 515-	1440
I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	GAT GOT GAT GOT GAT GOT GAT GOT GAT GOT	ACTAACC ACTAACC ACTAACC ACTAACC ACTAACC ACTAACC	- ACT AGAAG I-IAATTAGAAG I-TAATTCGAAG - TAATTGGAAG I-TAATTAGAAG I-TAATTAGAAG	T AGG CAC T AGG CAC T AGG CAC T AGG CAC T AGG CCC	CAAC COAA ICAAC COAA ICAAC COAA ICAAC COAA ICAAC COAA ICAAC COAA	ACAGITAAA ACAGITAAA ACAGITAA ACAGITAA ACAGITAA ACAGITAA	2A 2A 2A 2A 2A 2A 2A 2A	AATTCTTT AATTCTTT AATTCTTT AATTCTTT AATTCTTT AATTCTTT	CAAC II ACAAA CAACIIAGAAA CAACIIAGAAA CAACIIAGAAA CAACIIAGAAA CAACIIAGAAA CAACIIAGAAA	ACAG ACAG ATAG ATAG ATAG ATAG
CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C5 *MXMT C6 *MXMT C6 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT C17 *MXMT C18 *MXMT C19 *MXMT	CALSE CALSE CALSE CACE CACE CACE CACE CACE CACE CACE CA				CAAC CCAA				CAAC TAGAAA CAAC TAGAAA	
CcDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT	AATACA AATACA GATGC GATGC GATGC	CTEACAE CTEACAE ACEACCE ACEACCE ACEACCE ACEACCE ACEACCE	AAT CTTTCAAC AAT CTTTCAAC - TAATTCGAAC I - TAATTACAAC I - TAATTACAAC I - T <mark>AATTACAA</mark> C	T AGAA - A D T AGAA - A D AGGI CAC T AGGI CAC T AGGI CAC T AGGI CAC T AGGI CAC	A ACCCCA A ACCCCA CAA CAA CAA CAA CAA	1 A A I I I G TA A I I I G A C A C I I A A A A A C A C I I A A A A	2	ACTACA	AR ARR 200000	
I3 CcXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	1441 5- GCCGA GCCGA GCCGA GCCGA	1450 				1490 I5 IAAAA-GCIAT IAAAA-GCIATI IAAAA-GCIATI IAAAA-GCIATI IAAAA-GCIATI	1500 GGAAA-AAA GGAAA-AAA GGAAA-AAA GGAAA-AAA GGAAA-AAA GGAAA-AAA GGAAA-AAA	1510 AATAATA AATAATA AATAATA AATAATA AATAATA	1520 SIS- IGRACIAAAAG IGRACIAAAAG IGRACIAAAAG IGRACIAAAAG IGRACIAAAAG	1530 I IACAA IACAA IACAA IACAA IACAA IACAA IACAA
CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C5 *MXMT C5 *MXMT C10 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C13 *MXMT C16 *MXMT C16 *MXMT C18 *MXMT C19 *MXMT						AAAAA - GCTAC AAAAA - CCTAT GCTAT CCTAT CCTAT CCTAT CCTAT CCTAT CCTAT CCTAT CCTAT CCTAT CCTAT CCTAT	CGARA - ARA CGARA - ARA	120 120 120 120 120 120	GAC AAAAG	DA CAA DA CAA DA CAA
CcDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT	ATTIC			CCCCCCCCA	AATTISGGTI AATTISGGTI CATIIGGGII	AAAAA - CCTAT AAAAA - CCTAT AAAAA T CTATT	CCAAA-AAA CCAAA-AAA CCAAAAAAAAA	TAATAATA TAATAATA TAATAATA	GAAC AAAAC GAAC AAAAC I GAAC AAAAC	CACAA CACAA CACAA A A

	1531 5-	1540 5·	1550 5-	1560	1570 5 9	1580 5 5	1590 5-	1600 !	1610 515	1620
I3 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	CIAAIA CIAAIA CIACIA CIAAIA CIAAIA CIAAIA	ATAATATET ATAATATET ATAATATOT ATTAATATOT ATTAATATOT ATAATATO	ICATITI GAAP ICATITI GAAP ICATATIGAAP ICATATIGAAP ICATATIGAAP ICATATIGAAP	TAAGGAA TAAGGAA TAAGGAA TAACGAA TAACGAA	AAAAAACIGI RAAA-ACIGI RAAA-ACIGI RAAA-ACIGI RAAA-ACIGI RAAA-ACIGI RAAA-ACIGI		CC AGAG CC AGAG CC AGAG CC AGAG CC AGAG	TTGACGTT TTGACGTT TTGACGTT TTAACGTT TTAACGTT TT <mark>AACGTT</mark>	ITAGCAAATIG ITAGCAAATIG IGAGCAAATIG ITAGCAAATIG ITAGCAAATIG ITAGCAAATIG	CCIAA CCIAA CCIAA CCIAA CCIAA GCIAA
CoMKMT1 CaMKMT1 CaMKMT2 C1 *MKMT C2 *MKMT C3 *MKMT C5 *MKMT C9 *MKMT C10 *MKMT C11 *MKMT C11 *MKMT C13 *MKMT C14 *MKMT C14 *MKMT C16 *MKMT C18 *MKMT C19 *MKMT			CATTITGAN CATTITGAN CATTITGAN CATTITGAN						NACCAAATIC TACCAAATIC - ACCAAATIC - ACCAAATIC	SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA SCIAA
CcDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT	CTARTA CTARTA CTARTA CTARTA CTARTA CTARTA	a laatatata alaatatata alaatatata attatatat attatatat alaatatat alaatatat	ACATATI DAAA CATATI DAAA CATATI DAAA CATATI DAAA CATATI DAAA CATATI DAAA	CTAAGGAA CTAAGGAA CTAAGGAA TTAAGGAA TTAAGGAA	AAAA - CCTGT AAAA - CCTGT AAAA - ACTGT AAAA - ACTGT AAAA - ACTGT AAAA - ACTGT		CC ACAC CC ACAC CC ACAC CC ACAC CC ACAC CC ACAC CC ACAC	TTGACGIT TTGACGIT TTGACGIT TTGACGIT TTGACGIT TTG <mark>AC</mark> GIT	ITA CAAATIG TA CCAAATIG ITA CCAAATIG TA CCAAATIG TA CCAAATIG TA CCAAATIG	SCTGA SCTGA SCTGA SCTGA SCTGA SCTGA
13/E4 CcXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	1621 I5- PTICCA TTICCA TTICCA TTICCA	1630 !5 Actorcete Actorcete Actorcete Actorcete Actorcete Actorcete	1640 	1650 TTCAATTC TTCAATTC TTCAATTC TTCAATTC TTCAATTC	1660 5 AATIGATGAT AATIGATGAT AATIGATGAT AATIGATGAT AATIGATGAT	1670 5 5 5 	1680 IIIIII ICCACCACA ICCACCACA ICCACCACA ICCACCACA ICCACCACA ICCACCACA ICCACCACA	1690 	1700 5	1710 I AFAGI AFAGI AFAGI AFAGI ATAGI ATAGI
CoMXMT1 CaMXMT2 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C4 *MXMT C5 *MXMT C6 *MXMT C10 *MXMT C10 *MXMT C13 *MXMT C13 *MXMT C14 *MXMT C16 *MXMT C18 *MXMT C19 *MXMT	CONTENT CONTEN	ACICICAIC ACICICAICA ACICICCCIC ACICICCCIC ACICICCCIC ACICICCIC ACICICCIC ACICICCIC ACICICCIC ACICICCIC ACICICCIC ACICICCIC ACICICCIC ACICICCIC ACICICCIC	C C CRAADA- C C C CRAADAA C C C CRAADAA	I GAAIIG TI GAAIIG TI GAAIIG TI GAAIIG TI GAAIIG TI GAAIIG TI GAAIIG TI GAAIIG C GAAIIG G GAAIIG T GAAIIG	AATICA CALAN AATICA CALAN			CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA CTCCACCA	AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO AGAAAANTCO	ALAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI CIAGI
	TTECCA	ACTCICCIC	CAAALAA	TT GAN IT G					ion in the observed of the obs	

	1711	1720	1730	1740	1750	1760	1770	1780	1790	1800
E4	000000									
CeXMT1 CeXMT1	TTCAAT	CTTCCAGE CTTCCAGE	TATATACCI TATATACCI	I CAGCAGAAGAA I CAGCAGAAGAA	GTAAAGIG GT <mark>AAA</mark> GIG	CATAGITGAGGA Cat <mark>agitga</mark> gga	GG <mark>AA</mark> GGTICI GG <mark>AA</mark> GGTICI	TTTGAAATT TTTG <mark>AAA</mark> TT	TIATACCIGO TIATACCIGO	AGACT
CaXMT2 C3 *XMT	CARD	CTECCAGE CTECCAGE	TATATACCT TATATECCT	I CAGCAGAAGAA I CAACAGAAGAA	GTAAAGTG GTAAAGTG	CATAGTTGAGGA CATAGTTGAGGA	COAACCTTCT CCAACCTTCT	TTTGAAATT	TEATACCEGO TEATACCEGO	ACAC ACAC
C12 *XMT	TE CAA	CTECCAGE	TATATECCT	CAACAGAAGAA	CTAAA GTO	CATAGTTGAGGA	CCAA COTTC	TTTCAAATT	TIATACCEGO	ACAC
CIS "AMI	CAR	CI DCCADE		CARCAGRAGAA	a sur a su	-ALAGII JAGA		III GAAALI		
CcMXMT1 CaMXMT1		ATTCCATT	TTTACACCI TTTACACCI	I CAGCAGAAGAA I CAGCAGAAGAA	CTAAACTO CTAAACTO	CATAGTTGAGGA CATAGTTGAGGA	GGAAGGTTCI GGAAGGTTCI	TGCGAAATT TGCGAAATT	TTATATCTCC TTATATCTCC	ACAC
CaMXMT2	TT CAR	CTTCCATTO	TTTACACCT	CACCACAACAA	GTAAAGTG	CATAGTICAGGA	Ge <mark>rr</mark> gett C	TTTGAAATT	TTATACCICO	ACAC
C2 *MXMT		CTTCCATT	TATACACCI	ICAGCAGAAGAA	GTAAAGTG	CATAGTTGGGGA	GG <mark>AA</mark> GGTICI	TTTG <mark>AAA</mark> TT	TIATATCIC	AGAC
C4 *MXMT C5 *MXMT		CTECCATE CTECCATE	TATACACCI TATACACCI	CACCACAACAA CACCACAACAA	CTAAACTO CTAAACTO	CATACTIGGGGA CATACTIGGGGA	CGAAGGTICI CCAAGGTICI	TTTCAAATT TTTCAAATT	TTATATCICO TTATATCICO	ACACT
C6 *MXMT	TT CAR	CTICCATT	TATACACCT	CACCAGAAGAA	GTAAA GTG	CATACTTCCCCA	CO <mark>AA</mark> COTTC	TTTCAAATT	TTATATCEC	ACAC
C10 *MXMT	TICAAT	CTECCATE	TATACACC	ICAGCAGAAGAA	GTAAAGTG	CATAGTIGGGGA	GG <mark>AA</mark> GGTTCT	TTTCAAATT	TIATATCIC	ACAC
C11 *MXMT C13 *MXMT			TATACACCI	CACCACAACAA CACCACAACAA	CTAAACIC CTAAACIC	CATACTICCCCA CATACTICCCCA	GGAAGGTICI GGAAGGTICI	TTTCAAATT'	TIATATCIGO TIATATCIGO	ACAC
C14 *MXMT	TTCART	CTICCATT	TATACACC	CACCACAACAA	GTAAAGTG	CATACTTCCCCA	GGAAGGTIC	TTTGAAATT	TTATATCTCC	ACAC
C16 *MXMI C17 *MXMI	TICAA	CTECCATE	TATACACCI	ICAGCAGAAGAA	GTAAAGIG	CATAGIT GGGGA	GG <mark>AR</mark> GGTICI	TTTGAAATT	TIATATCIC	ACACT
C18 *MXMT C19 *MXMT		CTTCCATT CTTCCATT	TATACACCT TATACACCT	I CACCACAACAA I CACCACAACAA	CTAAACTO	CATACTTCCCCA CATACTTCCCCA	GG <mark>AA</mark> GGTTCT CGAAGGTTCT	TTTCAAATT	TTATATCIGG TTATATCIGG	ACACT
0-D10/7										
CaDXMT2	TTCART	GTECCAAT	TATCCACCT	I CAACAGAAGAA	GTAAAGCG	CATAGTTGAGGA	GG <mark>AA</mark> GGTICI	TTTGAAATT	TTATACCTC	ACAC
CaDXMT1 R1 *DXMT		GTECCAAT(TATOCACCI TATOCACCI	I CAACAGAAGAA	CTAAACTO	CATACTICACCA CATACTICACCA	CCAACCTTC CCAACCTCTC	TTTGAAATT	TEATACCIGO	ACAC
R3 *DXMT	TTCAAD	CTECCAAT	TATECACC	I CACCAGAACAA	GTAAAGTG	CATAGTT GAGGA	GGAAGGCTCI	TTTCAAATT	TEATACCIEG	AGACT
R4 *DXMI	LICAAL		ALGUACUL			-Albitic Light	I SI S A SI			
	NUMBER							and an		
	1801	1810	1820	1830 5 5-	1840	1850 5 5-	1860	1870	1880 8	1890
E4 CoXMT1	1801 5-			1830 5 5-	1840 !	1850 5 5-	1860 5-	1870 5		1890
E4 CoXMT1 CaXMT1	1801 5- 	1810 !! GICCTIIA GICCTIIA	1820 S CATCCIGGC CATCCIGGC	1830 5 5- IICICIATICA- IICICIATICA-		1850 5 5- 300000000000	1860 5- 000000000000	1870 5 	1880 	1890
E4 CoXMT1 CaXMT1 CaXMT2 C3 *XMT	1801 5- IITAAG IITAAG IITAAG IITAAG			1830 5 5- 1000000000000000000000000000000000000	1840	1850 5 5- 200000000000 CCCACTAACAT CCCATTAACAT	1860	1870	1880 	1890]]]]
E4 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	1801 I 5 DODDDD I TTAAG I TTAAG I TTAAG I TTAAG		1820		1840 	1850 5		1870	1880 J S CAT CAACATA CAT CAACATA CAT CAACATA CAT CAACATA CAT CAACATA CAT CAACATA CAT CAACATA CAT CAACATA	1890
E4 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	1801 I5- IIIAAG IIIAAG IIIAAG IIIAAG IIIAAG				1840 000000000 GATGATGA GATGATGA GATGATGA GATGATGA	1850 5		1870	1880	1890
E4 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1	1801 I5- IIITAAG IIITAAG IIITAAG IIITAAG IIITAAG IIITAAG				1840 	1850 5	1860 5- 19000000000000000000000000000000000000	1870	1980 	1890
E4 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT1 CaMXMT2 C1 *MWT	1801 I5- ITTAAG ITTAAG ITTAAG ITTAAG ITTAAG ITTAAG ITTAAG				1840 	1850 5			1980 A CAACAD CA CAACAD	1890
E4 CcXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT					1840 	1850 5	1860 	1870	1880 	1890
E4 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT			1820 1820 18416C166C 18416C 18416C166C 18416C166C 18416C166C 18416C166C		1840 CONCOLOGIO CATAATIA CATAATIA CATAATIA CATAATIA CATCATIA CATCATIA CATCATIA CATCATIA CATCATIA CATCATIA CATCATIA CATCATIA	1850 5		1870	1880 	1890 ITDAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA
E4 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT1 C2 *MXMT C4 *MXMT C5 *MXMT C6 *MXMT C6 *MXMT C6 *MXMT			1820 		1840 	1850 5			1980 	1890 1 100000 ITAAA ITAAA ITAAA ITAAA ITAAA ITAAA ITAAA ITAAA ITAAA ITAAA ITAAA ITAAA ITAAA
E4 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C3 *MXMT C5 *MXMT C6 *MXMT C9 *MXMT C10 *MXMT	1801 I5- IIIAAG IIIAAG IIIAAG IIIAAG IIIAAG IIIAAG IIIAAG IIIAAG IIIAAG IIIAAG IIIAAG IIIAAG				1840 	1850 5			1980 	1890 1 100000 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA
E4 CcXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CaMXMT1 CaMXMT1 CaMXMT1 C1 *MXMT C1 *MXMT C4 *MXMT C5 *MXMT C6 *MXMT C1 *MXMT C1 *MXMT C1 *MXMT C13 *MXMT					1840 CATGATIA GATGATIA GATGATIA GATGATIA GATGATIA GATGATIA GATGATIA GATGATIA GATGATIA GATGATIA GATGATIA GATGATIA GATGATIA GATGATIA	1850 5			1980 	1890 IDDOD ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA ITTAAA
E4 CcXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CaMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C1 *MXMT C2 *MXMT C4 *MXMT C6 *MXMT C10 *MXMT C10 *MXMT C13 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT					1840 CONCONCIONAL CATAATTA CATAATTA CATAATTA CATGATA CATGATA CATGATA CATGATA CATGATA CATGATA CATGATA CATGATA CATGATA CATGATA CATGATA CATGATA CATGATA CATGATA CATGATA	1850 5		1870	1880 	1890 1 100000 100400 1114AA 1114AA 1114AA 1114AA 1114AA 1114AA 1114AA 1114AA 1114AA 1114AA 1114AA 1114AA 1114AA
E4 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C4 *MXMT C4 *MXMT C5 *MXMT C1 *MXMT C10 *MXMT C11 *MXMT C14 *MXMT C16 *MXMT C16 *MXMT C16 *MXMT C16 *MXMT			1920 		1840 	1850 5			1980 	1890 1 10000 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA 174AA
E4 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT C15 *XMT CaMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C5 *MXMT C10 *MXMT C10 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT C14 *MXMT C14 *MXMT C16 *MXMT C18 *MXMT C19 *MXMT C19 *MXMT	1801 I				1840 	1850 5			1980 	1890 10000 174AA
E4 CcXMT1 CaXMT2 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CcMXMT1 CcMXMT1 CcMXMT1 CcMXMT2 C1 *MXMT C1 *MXMT C1 *MXMT C1 *MXMT C1 *MXMT C1 *MXMT C1 *MXMT C14 *MXMT C14 *MXMT C14 *MXMT C14 *MXMT C15 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT					1840 	1850 5			1980 A CAACAD CA CAACAD	1890 1890 10000 104AA 117A
E4 CoXMT1 CaXMT2 C3 *XMT C12 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C1 *MXMT C1 *MXMT C3 *MXMT C6 *MXMT C10 *MXMT C10 *MXMT C13 *MXMT C13 *MXMT C14 *MXMT C14 *MXMT C17 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT					1840 	1850 5			1980 	1890 1 IDAAA
E4 CcXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT CoMXMT1 CaMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C3 *MXMT C4 *MXMT C10 *MXMT C10 *MXMT C11 *MXMT C14 *MXMT C15 *MXMT C19 *MXMT C19 *MXMT C19 *MXMT					1840 	1850 CCACIACAACA CCCACIAACAAC CCCACIAACAAC CCCACIAACAAC CCCACIAACAAC CCCACIACAACAC			1980 	1890 1890 1990

	1891	1900	1910	1920	1930	1940	1950	1960	1970	1980
E4 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	CCACAG CCACAG CCACAG CCACAG CCACAG CCACAG	IAT GTI GCA IAT GTI GCA IAI GTI GCA IAI GTI GCA IAT GTI GCA IAT GTI GCA			CAACCCATCO CAACCCATCO CAACCCATCO CAACCCATCO CAACCCATCO CAACCCATCO	CCCAACTCAT TCCCAACTCAT TCCCAACTCAT TCCCAACTCAT TCCCAACTCAT TCCCAACTCAT	TTI GGAGAJ ITTI GGAGAJ ITTI GGAGAJ ITTI GGAGAJ ITTI GGGCAJ ITTI GG <mark>A</mark> GAJ	JOCCICCON LCC ATTALAC ACCIATTALAC ACCIATTALAC ACCIATTALAC ACCIATTALAC	S CC ACA A CC ACA A CC ACA A CC ACA A CC ACA A CC ACA A CC ACA A	
CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C2 *MXMT C5 *MXMT C6 *MXMT C6 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C14 *MXMT C16 *MXMT C17 *MXMT C18 *MXMT C19 *MXMT	CCACAC CCACAC CCACAC CCACAC CCACAC CCACAC CCACAC CCACAC CCACAC CCACAC CCACAC CCACAC CCACAC CCACAC CCACAC CCACAC	ALCI GOCA ALCI GOCA		21 CAC SA 21 CAC SA 21 CAC SA 21 CAC SA 22 CAC SA 23 CAC SA 24 CAC SA 25 CAC SA 26 CAC SA 27 CAC SA 28 CAC SA 29 CAC SA 21 CAC SA 22 CAC SA 23 CAC SA 24 CAC SA 25 CAC SA 26 CAC SA 27 CAC SA 28 CAC SA 29 CAC SA 20 CAC SA 21 CAC SA 22 CAC SA 23 CAC SA 24 CAC SA 25 CAC SA 26 <td< td=""><td></td><td></td><td></td><td></td><td>CC GAC T A.T CC GAC T A.T CC GAC T A.T CC GAC T A.T</td><td>CCAC CCAC CCAC CCAC CCAC CCAC CCAC CCA</td></td<>					CC GAC T A.T CC GAC T A.T CC GAC T A.T CC GAC T A.T	CCAC CCAC CCAC CCAC CCAC CCAC CCAC CCA
CoDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT	CACCO CACCO CACCO	CATCT GGCA CATCT GGCA CATCT GGCA	TCTETCETTAC TCTETCETTAC	ATCAATTT ATCAATTT ATCAATTT ATCAATTT	CCAACCCATCO CCAACCCATCO CCAACCCATCO	T CCCAAC CA T CCCAAC CA T CCCAAC CA T CCCAAC CAT	TTTCGAGA TTTCGAGAA TTTCG <mark>A</mark> CAA	AGCTATUTTAC AGCTATUTTAC AGCTATOTAC	ICT GACT TAB ICT GACT TAB ICT GACT TAT	CCCAC CCCAC CCCAC
	1981	1990	2000	2010	2020	2030	2040	2050	2060	2070
E4 CoXMT1 CaXMT1 CaXMT2 C3 *XMT C12 *XMT C15 *XMT	AGGI II AGGI II AGGI II AGGI II AGGI II AGGI II AGGI II	C JAAGCA C JAAGCA C JAAGCA C JAAGCA C JAAGCA C JAAGCA C JAAGCA	CACCAAACCI CCACCAAACCI CCACCAAACCI CCACCAAACCI CCACCAAACCI CCACCAAACCI CCACCAAACCI		GCCCAAAGGC GCCCAAAGGC GCCCAAAGGC GCCCAAAGGC GGCCAAAGGC GGCCAAAGGC GGCCAAAGGC		CITALCAN CITALCAN CITALCAN		JECCECE LEALACCAC LAALACCAC LAALACCAC	AGAAG AGAAG AGAAG
CoMXMT1 CaMXMT1 CaMXMT2 C1 *MXMT C2 *MXMT C4 *MXMT C5 *MXMT C6 *MXMT C10 *MXMT C10 *MXMT C11 *MXMT C13 *MXMT C13 *MXMT C16 *MXMT C16 *MXMT C18 *MXMT C19 *MXMT	ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT ÀCOCTT	C 374 C 2 C 374	C CAS C ARACCI C CAC CARACI C CAC CARACI		GGG CAAASGC GGG CAAASGC	CO A AA AA CO A AA AA </td <td></td> <td></td> <td>AAAACCAG</td> <td></td>			AAAACCAG	
CcDXMT CaDXMT2 CaDXMT1 R1 *DXMT R3 *DXMT R4 *DXMT	AGGAT IC AGGAT TO AGGAT TO	CCAACAAT CCAACAAT CCAACAAT	CA CAAAGO CA CAAAGO CA CAAAGO CA CAAAGO		GCC CAAAC GC CCC CAAAC GC CCC CAAAC GC		SITA CASI SITA CASI SITA CASI		AAAA COCA S AAAA COCO AAAA COCA S AAAA COCA S	ICAAC ACAAC ACAAC

	2071	2080
	15-	
E4	000000	000000
CoXMT1	CAGAC	GIGIAA
CaXMT1	CAGAC	GIGIAA
CaXMT2	CAGAC	ATCTAA
C3 *XMT	8 	
C12 *XMT		
C15 *XMT		. 0.755
CcMXMT1	CAGAO	GIGIAA
CaMXMT1	CAGAC	GIGIAA
CaMXMT2	CAGAC	GTGTAA
C1 *MXMT		

Philippine Journal of Science Vol. 153 No. 3, June 2024

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	GCAGACATGTAA
CaDXMT1	TCAGACCTCTAA
D1 +DYMT	

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.