CLONING AND CHARACTERIZATION OF *N*-METHYLTRANSFERASE GENE ASSOCIATED WITH CAFFEINE BIOSYNTHESIS IN COFFEE

A Thesis
Submitted to the

University of Mysore

for the award of the degree of

Doctor of Philosophy
in
BIOTECHNOLOGY

by

K.V. Satyanarayana, M.Sc (Ag)

cftri

PLANT CELL BIOTECHNOLOGY DEPARTMENT
CENTRAL FOOD TECHNOLOGICAL RESEARCH INSTITUTE
MYSORE - 570 020, INDIA

Dedicated

To my parents

who gave me their genes

K V Satyanarayana
Research Fellow
Department of Plant Cell Biotechnology
CFTRI, Mysore-570020

DECLARATION

I hereby declare that the thesis entitled "**CLONING AND CHARACTERIZATION OF *N*-METHYLTRANSFERASE GENE ASSOCIATED WITH CAFFEINE BIOSYNTHESIS IN COFFEE"** submitted to the **University of Mysore,** for the award of the degree of **DOCTOR OF PHILOSOPHY** in **BIOTECHNOLOGY,** is the result of the research work carried out by me under the guidance of **Dr. ARUN CHADRASHEKAR**, Scientist, Central Food Technological Research Institute, Mysore- 570 020, India during the period 2001-2005.

I further declare that the results presented in this thesis have not been submitted for the award of any other degree.

Date: 26-12-5

Place: Mysore

K V Satyanarayana
(Candidate)

Dr. ARUN CHANDRASHEKAR
Scientist
Department of Plant Cell Biotechnology

December 26, 2005

CERTIFICATE

This is to certify that the thesis entitled "**CLONING AND CHARACTERIZATION OF *N*-METHYLTRANSFERASE GENE ASSOCIATED WITH CAFFEINE BIOSYNTHESIS IN COFFEE"** submitted by **Mr. K. V. Satyanarayana** for the award of the degree of **DOCTOR OF PHILOSOPHY** in **BIOTECHNOLOGY** to the **UNIVERSITY OF MYSORE** is the result of research work carried out by him under my guidance in the Department of Plant Cell Biotechnology, CFTRI during the period 2001-2005.

ARUN CHANDRASHEKAR
(Research Guide)

Acknowledgements

I take this opportunity to acknowledge my guide Dr. Arun Chandrashekar, Scientist, Plant Cell Biotechnology Department, CFTRI, Mysore for his valuable guidance, constant supervision and co-operation. I am truly indebted to him for his helpful attitude and for the freedom to pursue my specific goals in his laboratory. For without his efforts, it is unlikely that the thesis would have been completed.

I wish to express my heartfelt gratitude to Dr. V. Prakash, Director, CFTRI, Mysore, for granting me the opportunity to utilize the excellent facilities available at CFTRI and submit the results of my work in the form of a thesis.

I am ever grateful to Dr. G. A. Ravishankar, Head, Department of Plant Cell Biotechnology, CFTRI, for his constant encouragement and support during the pursuit of my research work. I sincerely thank the former Head of Food Microbiology Department, Dr. M.S. Prasad and the present Head, Dr. S. Umesh Kumar for their constant encouragement and co-operation during my research work. I sincerely acknowledge the support of all the staff members of Departments of Plant Cell Biotechnology and Food Microbiology during my research work.

I thank my seniors Dr. A. Ramesh and Dr. R.T. Venkatesha, who as senior colleagues helped me when I joined CFTRI. I have greatly benefited from their experience and enormous skills. I thankfully acknowledge the invaluable friendship and support given to me by Dr. Prakash M. Halami, Scientist, Food Microbiology Department. I thank my past and present colleagues in the project Obul, Indu, Vinod, Disha, Sowmya, Ramakrishna, and Sarala for their support in my work. It was pleasure working in the environment constituting of research students like Rajeshwari, Vageeshwari, Krupa, Priya, Anil, Deepak, Lincy, Reeta, Lokesh, Shaan, Kiran, Najma, Uma and many others. I would thank them all for creating friendly

surroundings during my stay in this lab. I thank all other research fellows in the Departments of Plant Cell Biotechnology and Food Microbiology for their affection and encouragement.

I greatly acknowledge Dr. A. Santa Ram, Central Coffee Research Institute, Balehannur for providing me coffee plant samples for my work. Material support in the form of primers and DNA cassettes from George Thomas, W.V. Baird and M. Kilstrup is gratefully acknowledged.

I sincerely acknowledge the Council of Scientific and Industrial Research (CSIR), New Delhi for a research fellowship, which enabled me to undertake the research project.

I reverentially express my gratitude towards my parents, wife and all other family members for their unwavering support and understanding. But for their moral support, it would have been difficult to accomplish the task.

K V Satyanarayana

Table of contents

List of Abbreviations

AA	Amino acids
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BSA	Bovine Serum Albumin
DIG	Digoxigenin
DMSO	Dimethyl sulfoxide
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra acetic acid
GUS	β-glucuronidase
IPTG	Isopropyl-β–D–thiogalactopyranoside
kb	Kilobase
kDa	Kilodalton
LB	Luria- Bertani (medium)
MES	2-(N-morpholino) ethanesulfonic acid
mM	millimole(s)
mRNA	Messenger RNA
NBT	Nitroblue tetrazolium
NMT	*N*-Methyltransferase
OD	Optical density
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SDS	Sodium dodecyl sulphate
SSC	Saline sodium citrate
TAE	Tris-acetate-EDTA buffer
TAPS	(N-Tris-[hydroxymethyl] methyl -3- aminopropane sulfonic acid
TE	Tris-EDTA buffer
Tris	Tris (hydroxymethyl) amino methane
X-GAL	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
X-Phosphate	5-bromo-4-chloro-3-indolyl phosphate
β	Beta
µl	Micro liter

List of Tables

List of Figures

Synopsis

Caffeine (1,3,7-trimethyl xanthine) is found in seeds and leaves of coffee, cola, mate and tea. Along with other methylxanthines, including theobromine, paraxanthine and methyluric acids, caffeine is a member of a group of compounds known collectively as purine alkaloids. The biosynthetic pathway of theobromine and caffeine has been the subject of much study over the years. Caffeine is synthesized through three sequential steps involving methylation of xanthine derivatives at positions 7-*N*, 3-*N* and 1-*N* (xanthosine $\rightarrow$ 7-Methylxanthosine $\rightarrow$ 7-Methylxanthine $\rightarrow$ Theobromine $\rightarrow$ Caffeine), catalyzed by *N* -Methyltransferases that use S-Adenosyl-L-Methionine as methyl donor.

A survey of literature suggests that the *N*-methyltransferases (NMTs) that are involved in the caffeine biosynthetic pathway in coffee comprise a multigene family. Several cDNA clones for NMTs from coffee have been reported. However, there are no reports on the genomic sequences for the NMT genes. The studies on NMT gene structure and organisation are lacking. Therefore, it was envisaged that PCR based genomic cloning of the NMT genes could provide valuable information regarding the number and size of introns and exons present in NMT genes. The study of introns and exons could possibly shed light on the mechanisms for origin of NMT multigene family in coffee and possible functional divergence among different members of the gene.

The survey of literature and patents indicated a distinct lacuna in the studies related to NMT gene regulation. The isolation of promoter for the NMT genes has not been hitherto reported and hence it became one of the major objectives of this investigation. The promoter isolation could help in understanding the NMT gene regulation and also open up the possibilities of regulating caffeine content in coffee plants.

There are wide variations in the caffeine contents of different coffee species, the reasons for which are unknown. Very little is known about the NMT genes in *Coffea* species other than *C. arabica*. It may be speculated that differential transcription of various NMT genes could be one of the reasons for such variations in the caffeine content. Study of possible

polymorphisms for the NMT genes and their corresponding promoters in different *Coffea* species could be vital in understanding differential transcription of various NMT genes.

The production of transgenic caffeine deficient coffee plants by RNAi using specific region of the theobromine synthase gene has been reported earlier, wherein spreading of silencing to other NMT genes was observed. The use of a conserved region common to all the NMT genes should theoretically result in silencing of all the NMT genes. Moreover, other posttranscriptional gene silencing strategies were not explored for caffeine down regulation. The use of conserved region of NMT genes for different posttranscriptional gene silencing strategies could be a viable option for obtaining phenotypes with desirable purine alkaloid profiles.

With this background knowledge, the present research work was undertaken. The research work carried out to meet the proposed objectives is presented in five chapters in the thesis. The references are cited at the end in alphabetical order. The title and summary of the work carried out under each chapter are presented as follows:

Chapter I

Introduction and review of literature

This chapter provides an overview of the current literature regarding the purine alkaloid biosynthesis in coffee. The taxonomy of coffee, distribution of various purine alkaloids and their biosynthesis in different *Coffea* species are highlighted. The properties of *N*-methyltransferases that catalyze the conversion of xanthosine to caffeine and the genes encoding them have been described in detail. The comparisons of the NMT sequences in tea and coffee with related methyltransferases are also presented. The pathway for caffeine catabolism in plants is discussed. The uses of biotechnological approaches for regulating the caffeine content in plants are discussed at the end of the chapter. This chapter also highlights the unexplored areas in coffee NMT research, which forms the basis for the present investigation.

Chapter II

Molecular Cloning of the *N* - Methyltransferase Gene from *Coffea canephora*

Primers for PCR amplification of *N*-methyltransferase (NMT) genes from *Coffea canephora* were designed based on the available cDNA sequences for NMT genes of *Coffea arabica*. Different sets of primers were used to obtain amplicons for partial and full-length NMT gene. Various PCR products were cloned and sequenced. Variations were observed with respect to restriction patterns among the clones for partial NMT gene. Sequence analysis of these clones revealed the presence of introns in the gene. High degree of similarity in the nucleotide sequences of exon regions and differences in the length and sequence of introns were observed, suggesting that these clones may belong to different members of the NMT multigene family. Using primers based on 5′ and 3′ untranslated regions specific for theobromine synthase, involved in the second methylation step, amplicons for full-length theobromine synthase gene were obtained. Sequencing of the clones for full-length gene revealed the presence of four exons interrupted by three introns. Based on the insertions and deletions present in the cDNA sequences of various NMT genes reported so far and the sizes of their open reading frames, the genomic clone corresponded to the theobromine synthase gene encoding for 378 amino acids. Variations with respect to the size and sequence of introns in two of the full-length clones for theobromine synthase were observed, though one of them contained one base nucleotide deletion in fourth exon. From this study, it appears that the coffee plants possess multiple alleles and/or multiple copies of the same gene; some of which may have undergone pseudogenisation.

Chapter III

Cloning of Promoter for an *N*-Methyltransferase Gene

A PCR based genome walking method was adopted for isolation and cloning of the promoter for one of the coffee NMT genes. Inspection of the promoter sequence revealed the presence of several motifs important for the regulation of the gene expression. The isolated promoter fragment was fused to the β-glucuronidase (GUS) reporter gene and used

in *Agrobacterium tumefaciens* mediated transformation of *Nicotiana tabacum* and also for electroporation of coffee endosperms. GUS assays proved that the isolated promoter was able to direct the expression of the reporter gene in transgenic tobacco. The cellular localization of GUS activity in transgenic tobacco confirms that coffee NMT gene is a cytosolic enzyme. Transient assays demonstrated the ability of the isolated promoter fragment to drive gene expression in endosperms. The cloning of promoter for a gene involved in caffeine biosynthetic pathway opens up the possibility of studying the molecular mechanisms that regulate the production of caffeine. Based on the promoter sequence, primer was designed and the genomic fragment comprising the promoter and its corresponding gene was amplified and cloned. Sequencing of one of the genomic clones revealed the presence of four exons and three introns in NMT gene. Sequence comparisons were made between coffee NMT gene and *Arabidopsis* Jasmonic acid carboxyl methyltransferase, which belongs to coffee *N*-methyltransferase related family. Though the gene structure was highly conserved in terms of number and position of introns, AtJMT revealed great variation for coding regions, introns, and the promoter region sequences, suggesting that the two genes have diverged greatly after their origin from common ancestor.

Chapter IV
Polymorphism for Coffee NMT Genes

Based on the promoter sequence for theobromine synthase gene-1, primers were designed and the genomic fragment comprising the promoter and its corresponding gene was amplified and cloned. Differences in the restriction pattern among the genomic clones of theobromine synthase gene-1 were studied by performing PCR-RFLP. Though differences were observed in restriction pattern for the clones, Southern blotting of the restricted PCR products from these clones with probes for promoter as well as the gene suggested that genomic clones are highly homologous and possibly belonged to different alleles/ members of NMT gene family. Sequencing of some of these clones from *Coffea canephora* revealed differences in intron length and sequences, but the promoter regions and the coding regions were highly conserved. It may be concluded that *C. canephora*

possesses several alleles/ copies of theobromine synthase gene-1. NMT gene fragment from Indian wild species *Psilanthus bengalensis* was also cloned and partially sequenced. The conservation of NMT gene structure was studied by comparing the introns from various genomic clones. To assess relationships between coffee NMT family members within and across species, their exon-intron organization patterns (e.g. number, positions, and phases of introns) were determined. Intron phase is strictly conserved in all genomic NMT clones available. All the three introns were located at conserved positions, being inserted between two codons (phase 0). Intron positions are absolutely conserved in the NMT genes from various species. Intron sequences have the consensus dinucleotide GT at the 5′-end and AG at the 3′-end. Although the splicing sites are conserved, the sizes of introns varied for different NMT genes. Two of the genomic clones obtained in this study are probably pseudogenes, due to a single nucleotide deletion in fourth and third exon respectively. Gene duplication followed by sequence divergence has been proposed as a model to explain how novel biochemical functions may have arisen for different duplicated genes of Coffee NMTs. The promoter regions were PCR amplified from different *Coffea* species. The sequence analyses of promoter regions from different species revealed a very high degree of conservation. The recent origin of coffee could be the reason for this high level of conservation of promoter sequences within the species and across the species.

Chapter V

Posttranscriptional Gene Silencing of Coffee *N*-Methyl transferases

A highly conserved region among coffee *N*-Methyltransferase genes involved in caffeine biosynthetic pathway was used in making sense, antisense and RNA interference constructs. Sonication Assisted *Agrobacterium tumefaciens* (SAAT) mediated transformation was used to mobilize these constructs into *Coffea canephora*. Somatic embryos were used for the transformation experiments. The transformed coffee plantlets, selected on hygromycin containing selection media, were used for estimating the purine alkaloids content by High Pressure Liquid Chromatography (HPLC). The purine alkaloids viz: 7-methylxanthine, theobromine, caffeine and theophylline were estimated from the transgenic plantlets. The three constructs differed in their efficiencies and specificity of

silencing. RNAi was found to be more efficient in reducing caffeine content in transformed coffee plantlets, when compared to sense and antisense constructs. Though the constructs were not specific for a single gene, sense and RNAi transformants were obtained that were mainly affected in one of the *N*-Methyltransferases. The results suggest that the use of highly homologous coding region is highly effective in down regulating caffeine biosynthesis, encoded by *N*-methyltransferase multigene family. Increased transcription of CaMTL1 gene, substrate specificity of which is not known, was observed indicating possible activation of parallel or bypass pathways.

Chapter VI

Conclusions and Future perspectives

The major findings of this investigation and the future prospects are briefly highlighted in the chapter.

CHAPTER I

INTRODUCTION AND REVIEW OF LITERATURE

1. REVIEW OF LITERATURE

1.1 Introduction

Coffee trees belong to the tribe *Coffeae* and the genus *Coffea* in the family *Rubiaceae.* While more than 100 distinct taxa have been identified, commercial coffee production relies mainly on two species: *Coffea* arabica L. and *Coffea canephora* Pierre. *Coffea arabica* (Arabica coffee) is cultivated extensively and represents ~ 70% of the market. The remaining 30% consists mainly of *Coffea canephora* (Robusta coffee). Some other species such as *C. liberica, C. dewevrei* and *C. racemosa* are produced only to satisfy local consumption. In India, traditionally Arabica and Robusta coffee are the two important commercial species of coffee cultivated in the hills of Western Ghats of India. In India, coffee is cultivated in 3,54,840 hectares with an average production of 2,75,500 metric tonnes (Source: Coffee Board of India, http://www.indiacoffee.org/coffeeindia). The exports from India for the year 2004 amounted to 2,18,737 metric tonnes valuing for Rupees 1,13,421 lakhs. The world coffee production for 2005-06 is expected to reach 6.3 metric tonnes (Source: Food Outlook, June 2005).

1.2 Taxonomy of coffee:

While the international coffee trade is concerned with only two coffee species- *Coffea arabica* and *C. canephora*- botanists regard as coffee trees of all tropical plants of the Rubiaceae family, which produce seed resembling coffee beans (Charrier and Berthaud, 1985). Chevalier (1947) and Leroy (1980) studied in detail the taxonomy of coffee.

1.2.1 Classification of coffee according to Chevalier (1947)

Chevalier (1947) has grouped the valid coffee species into the following sections:

(1) *Argocoffea*

(2) *Paracoffea*

(3) *Mascarocoffea*

(4) *Eucoffea*

The Coffee species belonging to *Mascarocoffea* section all have one characteristic in common: the absence of caffeine. The *Eucoffea* (now named Coffea) has been again divided into five subsections according to diverse criteria: tree height (Nanocoffea), leaf thickness (Pachycoffea), fruit colour (Erythrocoffea, Melanocoffea) and geographical distribution (Mozambicoffea) (Table 1.1).

Table 1.1. The grouping of the species in the subsection *Eucoffea* according to Chevalier (1947)

Section	Subsection	Species
Eucoffea	Erythrocoffea	*C. arabica*
		C. canephora
		C. congensis
	Pachycoffea	*C. abeokutae*
		C. liberica
		C. klainii
		C. oyemensis
		C. dewerei
	Melanocoffea	*C. stenophylla*
		C. carissoi
		C. mayombensis
	Nanocoffea	*C. humilis*
		C. brevipes
		C. togoensis
	Mozambicoffea	*C. schumanniana*
		C. eugenioides
		C. kivuensis
		C. mufindiensis
		C. zanguebariae
		C. racemosa
		C. ligustroides
		C. salvatrix

Table 1.2. Classification system for the genera *Coffea* and *Psilanthus* according to Leroy (1980) with indication of geographical distribution

Genus	Subgenus	Localization
Coffea L.	Coffea	Africa, Madagascar
	Psilanthopsis	Angola
	Baracoffea	Africa, Madagascar
Psilanthus (Hook. f)	Paracoffea	Africa, Asia, New Guinea
	Psilanthus	Africa

Table 1.3. The criteria used by Leroy (1980) to differentiate Genera and Subgenera

		Criterion 1	
		- Long corolla tube - Anthers not exserted - Short style	- Short corolla tube - Anthers exserted - Long style
		Genus: Psilanthus	**Genus: Coffea**
Criterion 2	- Axillary flowers - Monopodial development	P. Subgenus Psilanthus	C. Subgenus Coffea
	- Terminal flowers - Predominantly sympodial development	P. Subgenus Paracoffea	C. Subgenus Baracoffea

1.3 Commercial importance of caffeine

Since its isolation from coffee seeds by Runge in 1820, caffeine has been exhaustively studied. Because of its stimulatory effect on the central nervous system, a great number of reports have been published on its content in beverages and foodstuffs. Caffeine is in demand for medicines, and large amounts of caffeine are being used to supplement soft drinks. Recently, caffeine and related methylxanthines have been reported to inhibit HIV-1 replication (Nunnari et al., 2005). The caffeine required for these products is obtained either by large-scale chemical synthesis or as a byproduct of procedures used to decaffeinate tea and coffee.

1.4 Composition of coffee beans

Navellier (1961) gave a mean composition of green coffee as: glucides (58%), lignin (2%), lipids (13%), proteins (13%), ash (4%), non-volatile acids (8%), trigonelline (1%) and caffeine (1%). The caffeine content of green coffee is relatively limited (1-2.5%) of dry matter and changes little with seed development (Clifford and Kazi, 1987). In coffee seeds, the concentration of trigonelline is approximately 2% of dry weight (Clifford, 1985; Mazzafera, 1991). Mazzafera (1999) found a higher protein content in the mature beans than in the immature beans but a lower content of free amino acids, with asparagine as the main component. Recently, Flament (2002) estimated the composition of coffee beans as follows:

Proteins	: 10.0%	(dry basis of green coffee)
Carbohydrates	: 50.0%	-do-
Lipids	: 11.7-14.0 %	(Arabicas)
	: 7.6- 9.5 %	(Robustas)
Chlorogenic acids	: 6.5%	(Arabicas)
	: 9.0%	(Robustas)

1.5 Role of purine alkaloids

Among more than 50,000 secondary metabolites of plants, 12,000 are alkaloids (Ogawa et al., 2001). Alkaloids are defined as secondary metabolites that contain secondary, tertiary, or quaternary nitrogen atoms in their molecules. Many of the alkaloids for which structures have been described are known to be involved in the defence of plants against herbivores and pathogens; and some have been used by man as stimulants, drugs, narcotics and poisons (Hashimoto and Yamada, 1994). The physiological role of endogenous purine alkaloids and related compounds in higher plants remains undetermined. There are two hypotheses about the role of the high concentrations of caffeine that accumulate in tea, coffee and a few other plant species. The 'allelopathic theory' proposes that caffeine in seed coats is released into the soil and inhibits the germination of seeds. The 'chemical defence theory' proposes that caffeine in young leaves, fruits and flower buds acts to protect soft tissues from insect pests (Ashihara and Crozier, 2001). Several observations suggest that caffeine is generally effective as a pesticide (Hollingsworth et al., 2003; Mathavan et al., 1985; Nathanson, 1984). Transgenic tobacco plants expressing genes of the caffeine biosynthetic pathway produced up to 5-µg caffeine per g fresh weight of leaf tissues. Tobacco cutworms subjected to choice test avoided the transgenic leaves, indicating caffeine to have a clear repelling effect on pest insects. However, this amount was not high enough to confer a lethal effect (Uefuji et al., 2005).

1.6 Structures of purine alkaloids

Purine alkaloids consist of methylxanthines and methyluric acids. The most abundant purine alkaloid in nature is caffeine (1,3,7-trimethylxanthine), followed by theobromine (3,7-dimethylxanthine). Theophylline (1,3-dimethylxanthine), paraxanthine (1,7-dimethylxanthine), 1-methylxanthine, 3-methylxanthine, 7-methylxanthine and 7-methylxanthosine are found in small amounts. These minor compounds may be intermediates of caffeine biosynthesis and/or catabolism (Ashihara and Suzuki, 2004). The methyluric acids include theacrine (1,3,7,9-tetra methyluric acid), liberine [O(2),1,9-trimethyluric acid], and methyl liberine[O(2),1,7,9-tetramethyluric acid]. The structures of various purine alkaloids are presented in Figure. 1.1.

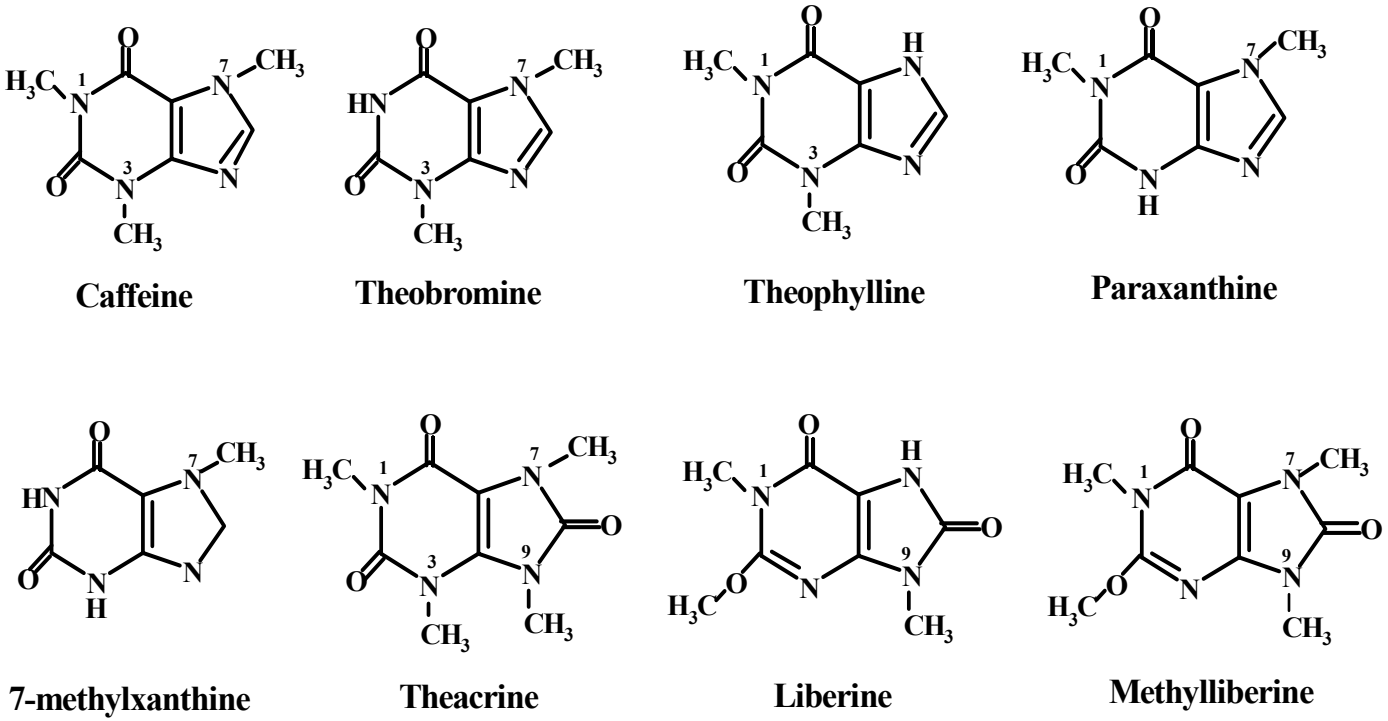


Figure 1.1 Structures of the methylxanthines and the methyluric acids

1.7 Distribution of caffeine in plants

Compared with other alkaloids, such as nicotine, morphine and strychnine, purine alkaloids are distributed widely throughout the plant kingdom. Caffeine has been found in 13 orders of the plant kingdom (Ashihara and Suzuki, 2004). In some species, the main purine alkaloid is theobromine or methyl uric acids such as theacrine, liberine and methylliberine, rather than caffeine (Ashihara and Crozier, 1999a; 2001) (Table 1.4).

Table 1.4 Distribution of purine alkaloids in plants

S.No	Plant species	Common name	Major alkaloid	Plant parts containing alkaloids
1	*Coffea arabica*	Arabica coffee	Caffeine	Leaves, Seeds
2	*Coffea canephora*	Robusta coffee	Caffeine	Leaves, Seeds
3	*Coffea liberica*		Caffeine Theacrine, Liberine	Seeds Mature leaves
4	*Coffea dewevrei*		Caffeine Theacrine, Liberine	Seeds Mature leaves
5	*Camellia sinensis*	Tea	Caffeine	Leaves
6	*Camellia assamica*	Assam tea	Caffeine	Leaves
7	*Camellia assamica var kucha*	Kucha	Theacrine, Caffeine	Leaves
8	*Camellia irrawadensis*		Theobromine	Leaves
9	*Camellia ptilophylla*	Cocoa tea	Theobromine	Leaves
10	*Theobroma cacao*	Cocoa	Theobromine	Seeds
11	*Paullinia cupuna*	Guarana	Caffeine	Seeds
12	*Cola nitida*		Caffeine	Seeds
13	*Citrus sp.*		Caffeine	Pollen
14	*Ilex paraguariensis*	Mate	Caffeine	Leaves

Adapted from Ashihara and Crozier (1999a; 2001)

Variations in caffeine content are found within the same tree, with the highest levels in seeds, flowers, and leaves (Raju and Gopal, 1979). Caffeine synthesis is restricted to very young parts of the leaves and stem. Zheng and Ashihara (2004) determined the amounts of purine alkaloids in various parts of 6-month-old seedlings. In seedlings, caffeine was distributed mainly in leaves and cotyledons at concentrations varying from 43 to 104 µmol $g^{-}1$ dry weight (Table 1.5) (Zheng and Ashihara, 2004). Theobromine was either not detected or was found in only trace amounts. The highest caffeine content was observed in cotyledons (104 µmol $g^{-}1$ dry weight). In contrast to cotyledons and leaves, stems and roots contained only very low concentrations of purine alkaloids. Caffeine was virtually absent in roots. They concluded that caffeine accumulation is specific to above ground parts (leaves, cotyledons and shoots) of the seedlings and that biosynthesis is performed only in very young tissues.

Table 1.5 Contents of caffeine in 6-month old *C. arabica* seedlings

Parts of coffee seedling	Caffeine (µmol $g^{-}1$dry weight)
Leaves	
Upper leaves	55.8 ± 0.3
Middle leaves	47.2 ± 6.9
Lower leaves	42.6 ± 2.8
Cotyledons	103.6 ± 8.5
Stem	
Upper part (including a bud)	6.2 ± 0.7
Lower part	0.9 ± 0.4
Roots	1.1 ± 0.3

Adapted from Zheng and Ashihara (2004)

The caffeine content in seeds of different *Coffea* species varies from 0.4 to 1.8 % on dry weight basis (Mazzafera et al., 1997) (Table1.6). Young expanding leaves of *C. arabica* and *C. canephora* plants also contain theobromine and other alkaloids, though in lower levels than caffeine.

Table 1.6 Caffeine content in seeds of *Coffea* and *Paracoffea* species

S. No	Coffea sp.	Caffeine (% dry matter in seeds) [A]		Caffeine (% dry matter in seeds) [B]	
		Mean	Range	Mean	Range
1	*Coffea arabica*	1.20	0.46 - 2.82	-	-
2	*Coffea canephora*	1.80	1.13 - 2.88	2.64	1.51 - 3.33
3	*Coffea dewevrei*	0.80	0.56 - 0.94	0.94	0.81 - 1.10
4	*Coffea congensis*	1.60	0.94 - 2.21	1.47	1.08 - 1.83
5	*Coffea brevipes*	0.57	0.48 - 0.66	2.54	2.36 - 2.96
6	*Coffea kapakata*	0.66	0.53 - 0.83	1.20	1.04 - 1.39
7	*Coffea liberica*	1.11	0.73 - 1.43	1.24	1.12 - 1.39
8	*Coffea stenophylla*	1.12	0.64 - 1.12	2.27	2.05 - 2.43
9	*Coffea eugenioides*	0.43	0.29 - 0.68	0.51	0.44 - 0.60
10	*Coffea salvatrix*	0.52	0.36 - 0.84	0.03	0.01 - 0.06
11	*Coffea racemosa*	0.75	0.48 - 1.11	1.06	0.86 - 1.25
12	*Psilanthus bengalensis*	0.05	0.03 - 0.07	-	-
13	*Psilanthus ebracteolata*	0.32	0.12 - 0.66	-	-

[A]Adapted from Mazzafera et al., (1997); Silvarolla et al., (2000)

[B] Adapted from Campa et al., (2005)

The marked differences observed in the values of caffeine estimated by two groups are due to the different extraction methods used. Campa et al., (2005) used a simplified extraction method, which limits yield loss and permits detection of lower caffeine concentrations. The caffeine contents in leaves, immature and mature fruits in various *Coffea* species are presented in Table 1.7.

Table 1.7 Caffeine content in leaves and fruits of species of *Coffea* and *Paracoffea*

Species		*Caffeine (% dry weight)*		
		Leaves	Immature fruits	Mature fruits
C. arabica	var. Catuai	0.929	--	1.343
	var. Laurina	0.721	0.753	0.618
	var. Mundo Novo	0.978	1.186	1.110
C. canephora	var. Robusta	0.456	--	1.710
	var. Guarini	0.248	--	2.359
C. congensis	var. Uganda	0.145	--	1.930
C. dewevrei	var. Excelsa	0.021	0.078	1.205
C. eugenioides		ND	0.509	0.401
C. kapakata		ND	--	0.715
C. liberica		0.021	--	1.360
C. racemosa		0.031	1.348	0.829
C. salvatrix		0.014	0.509	0.715
C. stenophylla		0.030	1.710	1.650
P. bengalensis		0.152	0.142	0.039
P. travancorensis		0.056	0.104	0.053
P. ebracteolata		0.059	--	0.022

Adapted from Mazzafera and Magalhaes (1991)
"--" Not analyzed; "ND" Not determined

Mazzafera and Magalhaes (1991) observed great variability of caffeine content both in leaves and seeds of mature fruits and only *C. arabica* var. Laurina, *P. ebracteolata* and *P. bengalensis* had higher caffeine levels in leaves than in seeds. Only *C. dewevrei* var. Excelsa and *C. salvatrix* had less caffeine in seeds of immature fruits than those of mature fruits.

1.8 Biosynthesis of purine alkaloids

Caffeine is a trimethylxanthine whose xanthine skeleton is derived from purine nucleotides that are converted to xanthosine, the first committed intermediate in the caffeine biosynthesis pathway. The available data indicate that the major pathway from xanthosine to caffeine is xanthosine → 7- methylxanthosine → 7-methyl xanthine → theobromine → caffeine (Figure 1.2); the first, third and fourth steps being catalyzed by *N*-methyltransferases (NMTs) that use S- adenosyl-L – methionine (SAM) as the methyl donor (Ashihara and Crozier, 2001). Besides this core pathway, some minor routes such as 7-methylxanthine → paraxanthine → caffeine, and xanthine → 3-methylxanthine → theobromine → caffeine, also function (Figure 1.2)(Ashihara et al., 1996a; Kato et al., 1996). These minor pathways may be result of the broad methyl acceptor specificity of the *N*-methyltransferases. There is one report of an alternative entry in the caffeine biosynthetic pathway in coffee that involves conversion of xanthosine 5′ monophosphate (XMP) → 7-methyl XMP → 7- methylxanthosine (Schulthess et al., 1996).

1.8.1 Production of 7-methylxanthosine

The activity of first methylation enzyme, xanthosine *N*-methyltransferase, to catalyze the formation of 7- methylxanthosine from xanthosine was demonstrated in tea (Fujimori et al., 1991; Kato et al., 1996) and coffee (Negishi et al., 1985b). However, Baumann and his group speculated that 7- methylxanthosine was not an intermediate of caffeine biosynthesis in coffee, because it was not detected in extracts from cultured coffee cells even when caffeine biosynthesis was stimulated (Schulthess and Baumann, 1995). Based on further observations, they proposed a new pathway: xanthosine monophosphate (XMP) → 7-methyl XMP → 7- methyl xanthosine → 7-methylxanthine (Schulthess et al., 1996). Mizuno et al., (2003a) identified and characterized the gene encoding the first *N*-methyltransferase, 7- methylxanthosine synthase, from coffee. The recombinant protein used by them exhibited xanthosine *N*-methyltransferase activity, but XMP could not be used as a substrate. Uefuji et al., (2003) also cloned a 7- methylxanthosine synthase from coffee similar to that reported by Mizuno et al., (2003a). The recombinant protein showed

xanthosine specific activity. The available data therefore suggest that xanthosine, and not XMP, is the primary methyl acceptor in coffee plants.

1.8.2 Hydrolysis of 7- methylxanthosine to 7- methylxanthine

The conversion of 7- methylxanthosine to 7- methylxanthine appears to be catalyzed by a specific *N*- methylnucleosidase (step 2 in Fig 1.2). This enzyme has been partially purified from tea leaves by Negishi et al., (1988). However, Uefuji et al., (2003) have demonstrated the *in vitro* conversion of 7- methylxanthosine to 7- methylxanthine using nonspecific purine nucleosidase phosphorylase derived from *E. coli*.

1.8.3 Production of caffeine from 7-methylxanthine via theobromine

The conversion of 7- methylxanthine to caffeine is catalyzed in two *N*-methylation steps (steps 3 and 4 in Fig 1.2). Suzuki and Takahashi (1975) have first demonstrated the activities of 7- methylxanthine *N*- methyltransferase and theobromine *N*- methyl transferase, which catalyze the second and third methylation steps in the core pathway in tea. *N*- methyltransferase activity has been detected in cell-free extracts prepared from immature fruits, callus and cell suspension cultures of *Coffea arabica* (Baumann et al., 1983; Roberts and Waller, 1979; Waller et al., 1983). Fujimori et al., (1991) confirmed the activities of the three *N*- methyltransferases in tea leaf extracts and found that they were present in high levels in very young developing leaves, but absent in fully developed leaves.

1.9 Sources for Xanthosine

A number of possible sources for xanthosine have been postulated. These are

(i) Inosine 5′-monophosphate (IMP) synthesized by *de novo* purine nucleotide biosynthesis.

(ii) Intracellular pools of adenine and guanine nucleotides.

(iii) Purine nucleotides and nucleosides produced by degradation of RNA.

(iv) Pathway in which adenosine, released from SAH, is converted to xanthosine via adenine, adenosine 5′ –monophosphate, inosine 5′ –monophosphate and xanthosine 5′ –monophosphate.

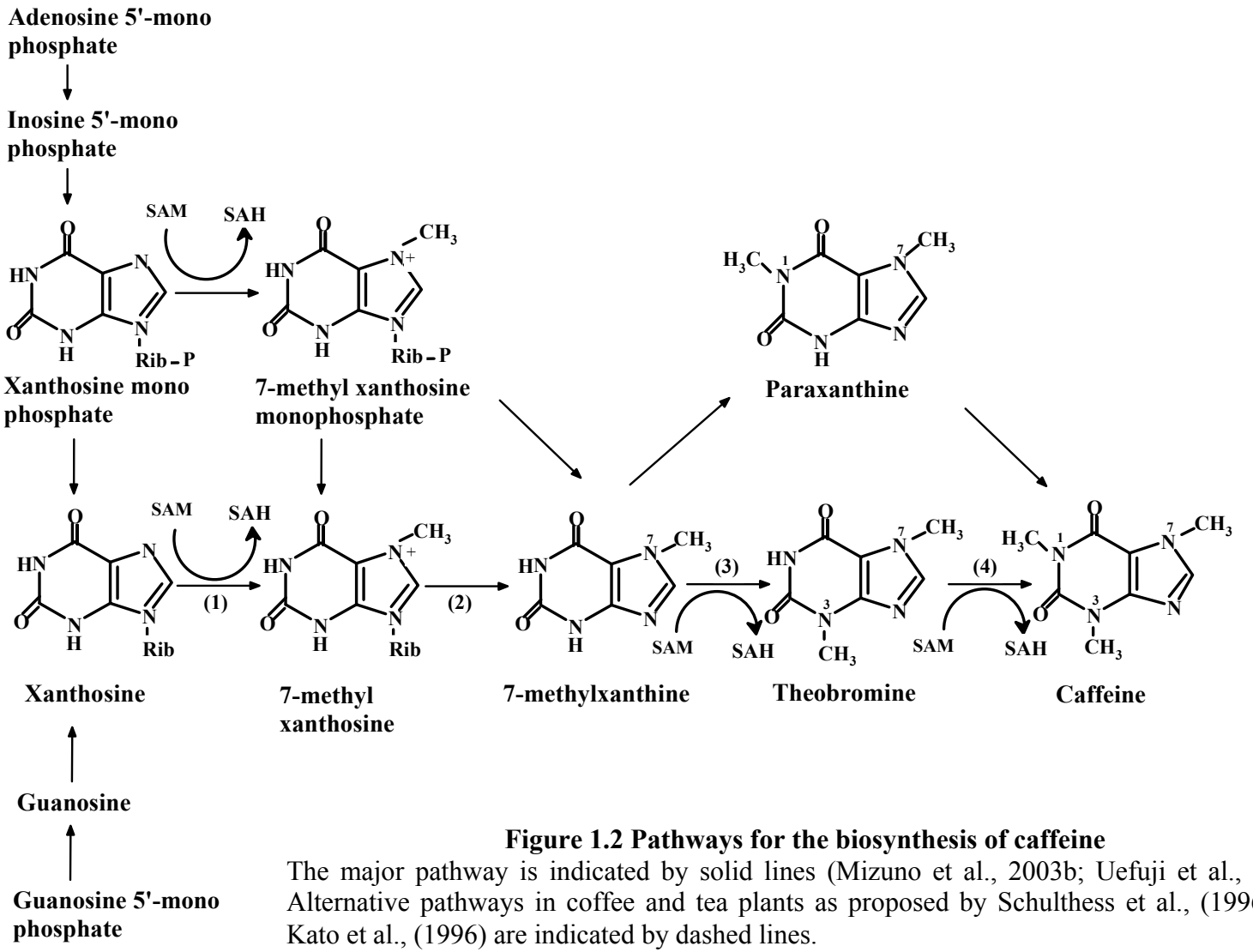


Figure 1.2 Pathways for the biosynthesis of caffeine

The major pathway is indicated by solid lines (Mizuno et al., 2003b; Uefuji et al., 2003). Alternative pathways in coffee and tea plants as proposed by Schulthess et al., (1996) and Kato et al., (1996) are indicated by dashed lines.

Koshiishi et al., (2001) have indicated that the conversion of SAH to xanthosine is such that the purine ring of caffeine can be produced exclusively by this route in young tea leaves. The formation of caffeine by this pathway is closely associated with the SAM cycle (also known as activated-methyl cycle) because the three-methylation steps in the caffeine biosynthesis pathway use SAM as the methyl donor. During this process, SAM is converted to SAH, which in turn is hydrolysed to L-homocysteine and adenosine (Figure 1.3). The adenosine is used to synthesize the purine ring of caffeine and the L-homocysteine is recycled to replenish SAM levels. Because three moles of SAH are produced via the SAM cycle for each mole of caffeine that is synthesized, this pathway has the capacity to be the sole source of both the purine skeleton and the methyl groups required for caffeine biosynthesis in young tea leaves.

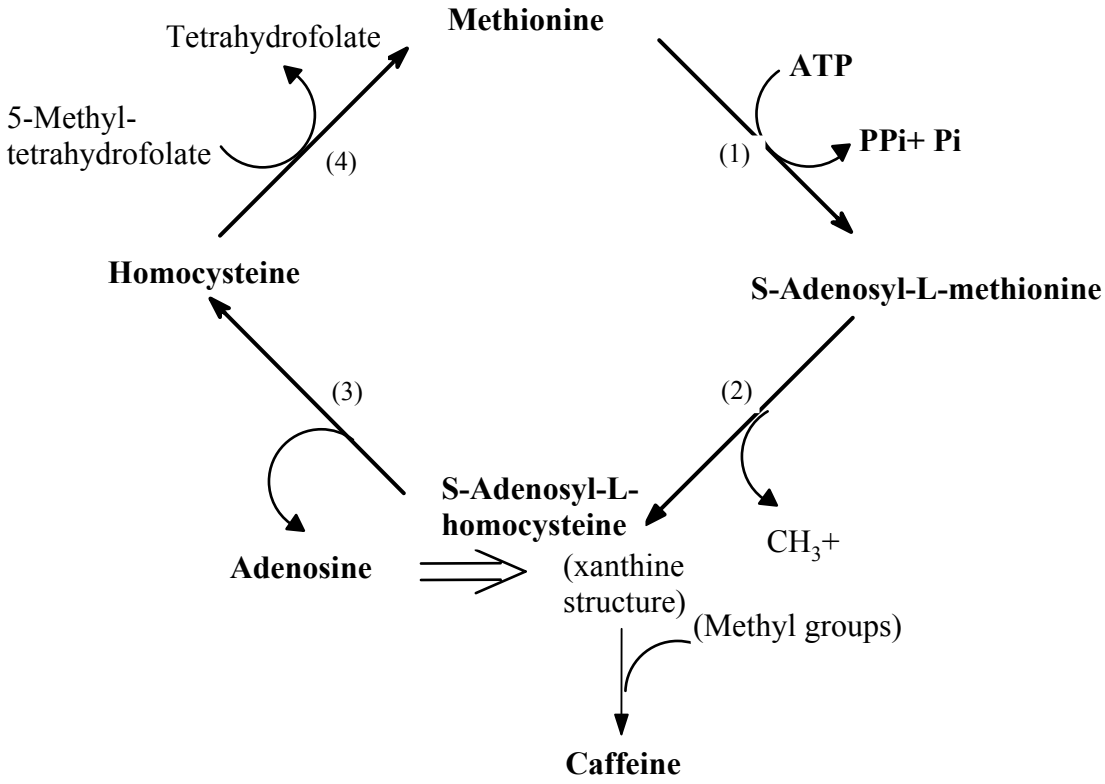


Figure 1.3 The SAM cycle (activated methyl cycle) in plants
(Ashihara and Suzuki, 2004)
Enzymes: (1) SAM synthetase, (2) SAM-dependent *N*-methyltransferases, (3) S-adenosyl-homocysteine (SAH) hydrolase, (4) Methionine synthase. Adenosine released from the cycle is salvaged to adenine nucleotides and utilized both for purine structure of caffeine via xanthosine and for re-synthesis of SAM via ATP.

1.10 Purification of *N*-methyltransferases

Several investigators have attempted the purification of the *N*- methyltransferase(s). Mazzafera et al., (1994a) first reported the purification of an *N*- methyltransferase from the endosperm and leaves of coffee and showed the presence of 7-methylxanthine and theobromine *N*- methyltransferase activity. However, these activities were extremely labile and the specific activity of the enzyme diminished markedly with purification. The specific activity of the final preparation was less than 1 fkat $mg^{-}1$ protein. Gillies et al., (1995) partially purified *N*- methyl transferase from liquid endosperm of coffee by ion-exchange chromatography, with final specific activity of 420 fkat mg^{-1}. Kato et al., (1996) partially purified *N*- methyltransferase from tea leaves by ion exchange and gel-infiltration chromatography with final specific activity of 100 pkat mg^{-1} with paraxanthine. Mösli Waldhauser et al., (1997a, b) purified *N*- methyltransferases from coffee leaves up to 39-fold using ion-exchange chromatography and chromatofocussing, showing that XMP-*N*-methyltransferase was different from other two *N*- methyltransferases.

N- methyltransferase activities involved in caffeine biosynthesis are extremely labile in cell free preparations and for this reason obtaining a purified preparation with high specific activity proved elusive. An *N*- methyltransferase from young tea leaves was purified 520 fold to apparent homogeneity (Kato et al., 1999) and a final specific activity of 5700 pkat mg^{-1} protein was obtained with several purification steps. The estimated molecular mass of native enzyme was 61kD by gel filtration chromatography and 41 kD by SDS-PAGE. The final preparation exhibited 3- and 1- *N*- methyltransferase activity with broad substrate specificity, showing high activity towards paraxanthine, 7-methylxanthine and theobromine, and low activity with 3-methylxanthine and 1- methylxanthine. However, the enzyme had no 7-*N*- methyltransferase activity towards xanthosine and XMP. Therefore, the single *N*- methyltransferase obtained was referred to as caffeine synthase (CS). It is likely that xanthosine *N*- methyltransferase protein is different from CS, and hence atleast two different enzymes catalyze the three methylation steps in the caffeine biosynthetic pathway (Kato et al., 1999). The K_m value for paraxanthine was lowest; hence, paraxanthine is the best substrate for CS. However, as there is limited synthesis of

paraxanthine from 7-methylxanthine, it is not an important methyl acceptor *in vivo* (Kato et al., 1996). The K_m value for theobromine was high, which may explain the transient accumulation of theobromine in young tea leaves (Ashihara and Kubota, 1986).

1.11 Molecular cloning of caffeine synthase gene from tea leaves

The determination of the amino- terminal sequence of caffeine synthase from tea (Kato et al., 1999) led to the cloning of the gene encoding caffeine synthase (TCS). Kato et al., (2000) obtained a cDNA sequence consisting of 1,438 base pairs, encoding a protein of 369 amino acids (TCS1, GenBank accession no. AB031280). They expressed the full-length coding region for TCS1 protein in *Escherichia coli* and demonstrated the CS activity of recombinant protein *in vitro*. The substrate specificity of recombinant CS was very similar to that of the native enzyme purified from young tea leaves. The recombinant enzyme mainly catalyzed *N*-1- and *N*-3 methylation of mono and dimethylxanthines. No 7-*N*- methylation activity was observed when xanthosine was used as the methyl acceptor. These results provided convincing evidence that TCS1 encodes CS.

1.12 Molecular cloning of *N*-methyltransferase genes from Coffee

The conserved amino acid regions of tea CS, O-methyltransferases derived from plant origin and Arabidopsis hypothetical proteins were made use of in designing primers for RT-PCR and RACE technique. Independent groups have isolated *N*-methyltransferase (NMT) genes from coffee (Table 1.8), which have been classified based on the substrate specificity of the recombinant proteins. Mizuno et al., (2003a) first reported the identification of 7-methylxanthosine synthase gene. They obtained seven homologous genes from coffee and referred them to CtCS (Coffee tentative caffeine synthase) series. They demonstrated that CtCS1 catalysed the 7-*N*-methylation of xanthosine and named it as CmXRS1. The full-length cDNA of CmXRS1 had 1,307 base pairs and encoded 372 amino acids. The recombinant CmXRS1 was specific for xanthosine and XMP could not be used as a substrate. This confirms the finding of Negishi et al., (1985a) that xanthosine is the first committed precursor of the caffeine biosynthetic pathway, although the possibility that unidentified CtCSs may be related to XMP pathway cannot be ruled out. CmXRS1 seems to be a key enzyme in the caffeine biosynthetic pathway, because it is

located at the branching point for caffeine biosynthesis from the general purine nucleoside and nucleotide metabolism (Mizuno et al., 2003a). Another group also isolated the gene for 7-*N*- methylxanthosine synthase gene (Uefuji et al., 2003).

Theobromine synthase catalyses the second methylation step, which is the conversion of 7-methylxanthine to theobromine. Coffee plants possess multiple enzymes, or isoforms to catalyze the second step of methylation, or theobromine synthesis. The cDNAs encoding theobromine synthase from coffee were denoted as CTS1, CTS2 (Mizuno et al., 2001) and CaMXMT1, CaMXMT2 (Ogawa et al., 2001; Uefuji et al., 2003) (Table 1.8). The cDNAs encoding for protein with 1-*N*-methyltransferase activities were isolated from coffee endosperm (Mizuno et al., 2003b; Uefuji et al., 2003). The recombinant proteins (CCS1, CaDXMT1) had both 1-*N*- and 3-*N*- methyltransferase activity, and were therefore designated as caffeine synthases. However, cDNA encoding for a protein with only 1-*N*-specific methyltransferase activity has not been isolated. Uefuji et al., (2003) expressed three cDNAs encoding *N* -methyltransferases from immature fruits of *Coffea arabica* plants, designated as CaXMT1, CaMXMT2, and CaDXMT1 respectively in *E. coli*. Combining the three recombinant proteins and *E. coli* crude extract, they demonstrated *in vitro* conversion of xanthosine into caffeine. The same group simultaneously expressed these coffee genes in tobacco plants, using a multiple-gene transfer method, and confirmed successful caffeine production up to 5 μg g^{-1} fresh weight in leaves of transgenic tobacco plants (Uefuji et al., 2005). These studies substantiated the proposed caffeine biosynthetic pathway: xanthosine → 7-methylxanthosine → 7-methylxanthine → theobromine → caffeine. Bacterially expressed encoded proteins were used for the characterization of their catalytic properties. Purified CaXMT1 solely catalyzed the conversion of xanthosine to 7-methylxanthosine. In contrast, crude CaXMT1 catalyzed the conversion of xanthosine to 7-methylxanthine, suggesting that conversion of xanthosine to 7-methylxanthosine and that of 7-methylxanthosine to 7-methylxanthine occurred successively. The former reaction was catalyzed by CaXMT-1 and the latter was catalyzed by a nonspecific purine-nucleoside phosphorylase derived from *E.coli*. CaMXMT-2 catalyzed conversion of 7-methylxanthine to theobromine. In addition, it catalyzed conversion of paraxanthine to caffeine with low activity, consistent with the properties of CaMXMT-1.

Table 1.8 NMT gene family from coffee

Gene/function	Name	Accession No.	Source	Technique used	Reference
7-methylxanthosine synthase	CaXMT1	AB048793	Immature fruit cDNA	Screening	Uefuji et al., 2003
	CmXRS1[a]	AB034699	Leaf cDNA library	RACE	Mizuno et al., 2003a
Theobromine synthase	CaMXMT1	AB048794	Leaf cDNA library	Screening	Ogawa et al., 2001
	CaMXMT2	AB084126	Immature fruit cDNA	Screening	Uefuji et al., 2003
	CTS1[b]	AB034700	Leaf cDNA library	Screening	Mizuno et al., 2001
	CTS2	AB054841	Leaf cDNA library	Screening	Mizuno et al., 2001
Caffeine synthase	CaDXMT1	AB084125	Immature fruit cDNA	Screening	Uefuji et al., 2003
	CCS1	AB086414	Immature fruit cDNA	RACE	Mizuno et al., 2003b
	CtCS7				

[a] The amino acid sequence is same as that of CaXMT1
[b] The amino acid sequence has only one substitution of that of CaMXMT1

CaMXMT-1 and CaMXMT-2 share the same reaction, being high and low- affinity enzymes, respectively. CaDXMT1 catalyzed conversion of 7-methylxanthine → theobromine → caffeine, and that of paraxanthine to caffeine. Paraxanthine was found to be the most preferred substrate, similar to tea caffeine synthase (TCS1). CaDXMT1 had an extremely high K_m value for three acceptable substrates (7-methylxanthine, theobromine, and paraxanthine), indicating that caffeine may not be synthesized before sufficient amounts of precursors have accumulated. Moreover, the K_m for Ado-Met was low (around 10 μM) for CaXMT-1, CaMXMT-1, and CaMXMT-1, whereas it was 150μM for CaDXMT1. Since Ado-Met is the common methyl donor in biological methylation and, therefore, is always scarce in cells, a low affinity for CaDXMT1 may ensure that caffeine synthesis takes place only upon accumulation of sufficient substrates. The substrate affinity of the concerned enzyme proportionally decreased toward the end point, indicating that further down the pathway, more amounts of substrate compound is required, making the reaction proceed irreversible and stepwise. The K_m values for enzymes involved in caffeine biosynthesis vary (Mizuno et al., 2003b) depending upon protein species. The caffeine synthase activity borne by CaDXMT-1 may be low, since the enzyme demonstrated a relatively low K_m towards theobromine (Uefuji et al., 2003), when compared with other caffeine synthases from coffee (Mizuno et al., 2003b) and tea (Kato et al., 2000). These variations suggest that coffee plants possess multiple enzymes with distinct affinities toward substrates, thereby catalyzing caffeine biosynthesis at a broad range of available substrate concentrations (Uefuji et al., 2003).

When dimethylxanthines were used as substrates *in vitro*, paraxanthine was the best methyl acceptor followed by theobromine. Similar observation was made by Roberts and Waller (1979), but formation of paraxanthine from 1-methylxanthine or 7-methylxanthine was not detected; leading to a conclusion that caffeine is not synthesized from paraxanthine in coffee plants. Although CaDXMT1, like TCS1, showed clear potential for conversion of paraxanthine into caffeine, Uefuji et al., (2003) could not identify a cDNA encoding a paraxanthine synthase from their coffee cDNA pool. This leaves the question open as to whether or not paraxanthine is a substrate for caffeine synthesis *in planta*.

As in the case of theobromine synthase, multiple enzymes or isoforms may exist for first and last methylation steps, as judged from the presence of many closely related genes, and from the expression pattern of the identified genes. For example, CaDXMT1 is almost exclusively expressed in immature fruits, whereas CaXMT1, CaMXMT1 and CaMXMT2 are found in leaves, floral buds and immature fruits (Uefuji et al., 2003). Because young leaves contain a high level of caffeine and DXMT activity (Ashihara et al., 1996a; Mösli Waldhauser et al., 1997a), these observations suggest that another form of DXMT is present in young leaves. Thus it can be said that coffee plants are equipped with multiple sets of enzymes, which may be necessary for constitutive production of caffeine in relevant tissues. Similarly, another group Mizuno et al., (2003a) observed that different caffeine synthetic pathways are involved in the differential expression of CmXRS1 and CTS-2 in the young leaves and endosperms of the coffee plant.

1.13 Regulation of the core caffeine biosynthetic pathway

Transcriptional regulation is the major controlling step for secondary metabolite pathways studied in detail to date (Memelink et al., 2001). The rate of caffeine biosynthesis too appears to be transcriptionally regulated. There appears to be spatial and temporal regulation of *N*- methyltransferase(s) genes as evident from the differential transcript accumulation of various biosynthetic genes in different tissues (Mizuno et al., 2003a; Uefuji et al., 2003). However, little is known about the regulatory elements involved in the differential expression of various genes involved in caffeine biosynthetic pathway.

The regulation of core pathway may also be possible by regulation of *N*-methyltransferase(s) enzyme activities by the control of SAM/SAH ratio (fine control) (Kato et al., 1999). Although SAM is used in numerous reactions, it is not thought to be a limiting resource in plant cells, as along as the activated methyl cycle and methionine salvage pathway remain active (Moffatt and Weretilnyk, 2001). In addition to its involvement in caffeine biosynthesis, SAM is utilized in the biosynthesis of polyamines and ethylene in plants with the carboxyl group of SAM being removed as CO_2 in reactions

catalysed by SAM decarboxylase and 1-aminocyclopropane-1-carboxylic acid oxidase, respectively. Using radiolabelled SAM, Koshiishi et al., (2001) have shown that these pathways also operate in tea leaves, but that the proportion of SAM utilized is lower than that involved in caffeine biosynthesis. Upon donation of the methyl group to acceptor molecules, S- adenosyl-L-methionine (SAM) is converted to S- adenosine-L-homocysteine (SAH). SAH is a potent inhibitor of methyltransferases, including caffeine synthase (Kato et al., 1999). Hence, removal of SAH by SAH hydrolase is essential for the continued operation of the *N*- methyltransferases of caffeine biosynthetic pathway. The SAH hydrolase reaction is reversible and the equilibrium lies heavily in favour of SAH synthesis from adenosine and homocysteine. Removal of adenosine and homocysteine is, therefore essential for the effective hydrolysis of SAH and this is achieved by metabolism of homocysteine to methionine, which is further converted to SAM thereby completing the SAM regeneration cycle (Figure 1.3). The effects of the concentration of SAM and several methyl acceptors on the activity of caffeine synthase show typical Michaelis-Menten kinetics, and there is no feed back inhibition by caffeine. Caffeine synthesis takes place only upon accumulation of sufficient substrates. The substrate affinity of the concerned enzyme proportionally decreases toward the end point, indicating that further down the pathway, more amount of substrate compound is required, making the reaction proceed irreversible and stepwise (Uefuji et al., 2003). Coffee plants are equipped with multiple sets of enzymes, which may be necessary for constitutive production of caffeine in relevant tissues. The control of caffeine levels in plants is a function of the balance between rates of synthesis and degradation; and this balance seems to vary depending on the plant species and the tissue developmental stage (Mazzafera, 2004). It is suggested that a change in balance between biosynthesis and degradation occurs such that when degradation is lower than synthesis, caffeine levels increase.

Further, the availability of purine precursors originating from *de novo* purine biosynthesis and purine salvage pathways may regulate the biosynthesis of caffeine. It has been shown that *de novo* purine biosynthesis pathway is involved in the caffeine biosynthesis, atleast in young growing tea leaves in which a rapid net accumulation of purine alkaloids is observed (Ito and Ashihara, 1999). There is not much information

available on the regulation of *de novo* purine biosynthesis pathways in plants. The available literature suggests developmental and hormonal control of purine biosynthesis, implying transcriptional control in regulating the purine biosynthesis (Boldt and Zrenner, 2003). The effects of auxins on caffeine content in coffee cell suspension cultures were reported by Furuya et al., (1989). They reported an increase in caffeine production under the influence of auxins, but observed no influence on theobromine. Light irradiation may play an important role in enhancing purine alkaloid production as an inducer of *N*-methyl transferases (Kurata et al., 1997).

Schulthess and Baumann (1995) concluded that caffeine formation is dependent on chlorogenic acid accumulation. If the latter is impaired, deficient caffeine complexation results in the inhibition of the purine alkaloid biosynthesis. Mösli Waldhauser and Baumann (1996) concluded that there is a regulatory connection between the caffeine and chlorogenic acid complex partners, possibly guided by the cytoplasmic caffeine concentration.

1.14 Comparison of the amino acid sequences of caffeine synthase family in tea and coffee with related methyltransferases

The predicted amino acid (AA) sequences of caffeine biosynthetic pathway genes reported so far and others in CtCS series derived from *C. arabica* share more than 80% homology amongst themselves, but share only 40% homology with TCS1 derived from *C. sinensis*. This indicates that the caffeine biosynthetic pathways in coffee and tea might have evolved independently, consistent with different catalytic properties of the enzymes involved (Mizuno et al., 2003b). Three conserved motifs (A, B and C) of the binding site of the methyl donor SAM have been reported in most plant SAM- dependent *O*-methyltransferases (Figure 1.4)(Joshi and Chiang, 1998). The predicted amino acid sequences of caffeine biosynthetic pathway genes share 40% homology with the amino acid sequences of salicylic carboxyl methyltransferase (*Clarkia breweri*), benzoic acid carboxyl methyltransferase (*Antirrhinum majus*) and jasmonic acid carboxyl methyltransferase (*Arabidopsis thaliana*), which belong to a family of motif B' instead of motif B as the conserved region (Figure 1.5).

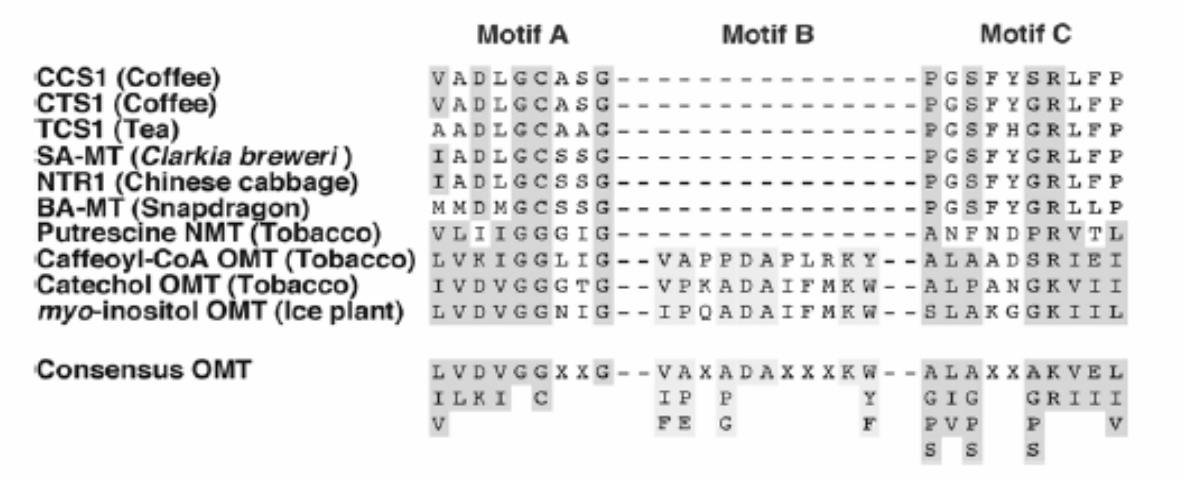


Figure 1.4 Sequence similarity and degree of relatedness among Caffeine synthase family and the related methyltransferases. Adapted from Mizuno et al., (2001)

The caffeine synthase genes share four highly conserved regions, motifs A, B', C and YEFF region, but lack motif B. The consensus sequence of motif B' region, which is located between motifs A and C, is 'LNDLF/P XNDFN' containing a large number of hydrophilic amino acids like asparagine and aspartic acid. The consensus sequence of YEFF region, which is located downstream of motif C, is 'AYXXQFXXDFXXFL' containing bulky amino acids (Figure 1.6) (Mizuno et al., 2003b).

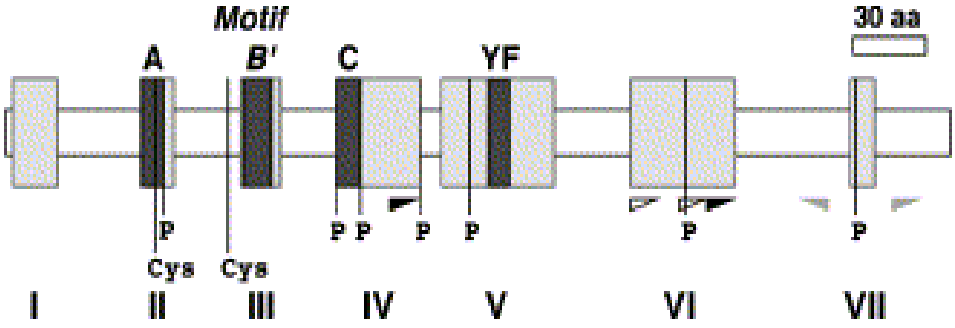


Figure 1.5 Schematic diagram of homologous regions of motif B' methyltransferase family. Adapted from Mizuno et al., (2003b)

The proposed SAM-binding motifs (A, B' and C) and the conserved region nominated as 'YFFF region' (YF) shown by closed boxes. The conserved sequences in motif B' methyltransferase family are indicated by open boxes (I, II, III, IV, V, VI, and VII). Conserved amino acid residues, especially proline and cysteine, are indicated by P and Cys, respectively.

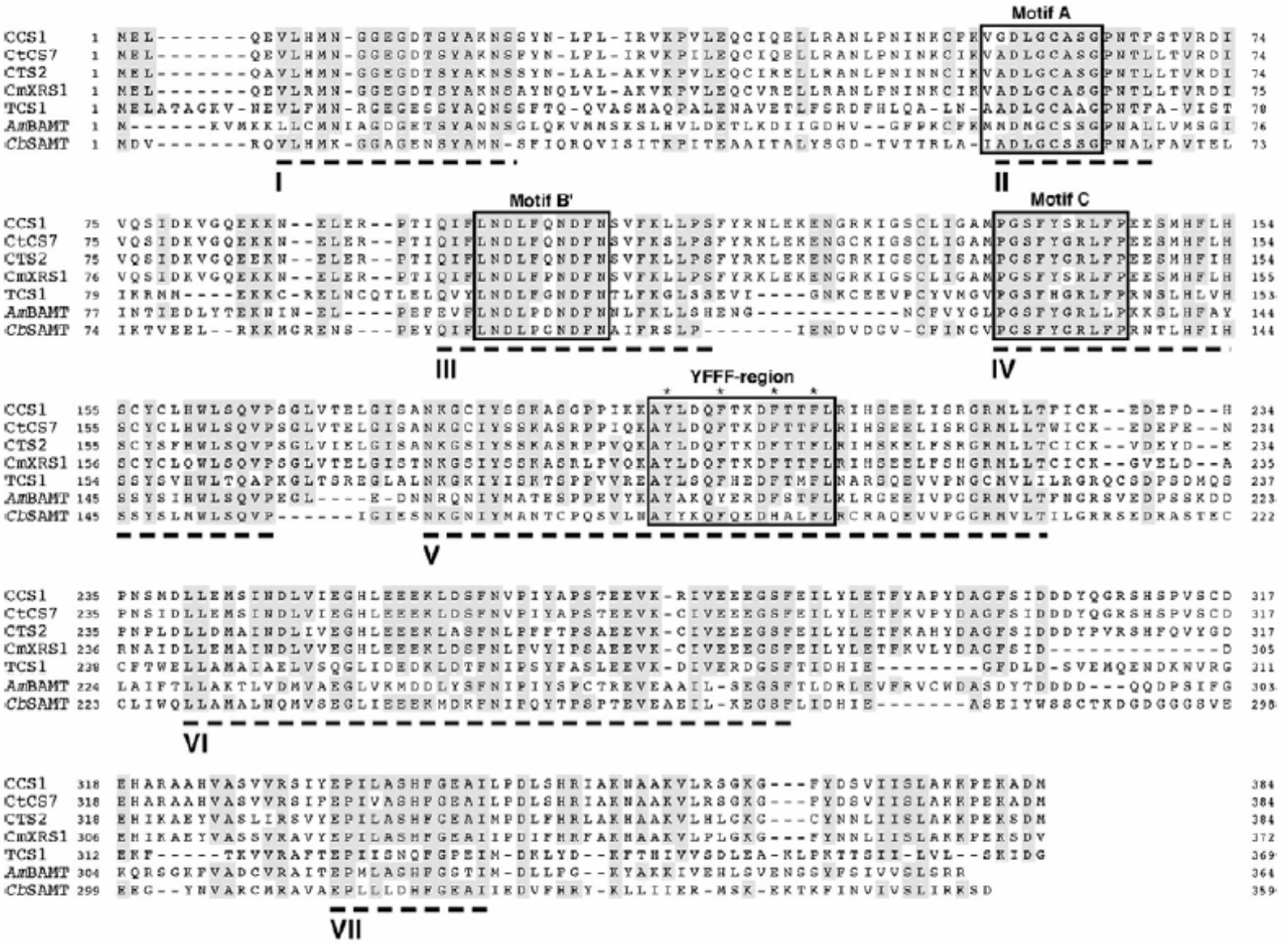

Motif A
Motif B'
Motif C
YFFF-region
I
II
III
IV
V
VI
VII
CCS1
CtCS7
CTS2
CmXRS1
TCS1
AmBAMT
CbSAMT

Figure 1.6 Comparison of the amino acid sequences of NMT proteins with related motif B' methyltransferases (Mizuno et al., 2003b)

Alignment of the amino acid sequences for CCS1, CtCS7, CTS2 and CmXRS1 (accession numbers AB086414, AB086415, AB054841 and AB034699, respectively) from *C. arabica*, TCS1 from tea (accession number AB031280), BAMT from *Antirrhinum majus* (accession number AF198492), and SAMT from *Clarkia breweri* (accession number AF133053). Shaded boxes represent conserved amino acid residues, and dashes represent gaps that have been inserted for optimal alignment. The proposed SAM-binding motifs (A, B' and C) and the conserved region nominated as ’ YFFF region’ are shown by open boxes. The conserved sequences in the motif B' methyltransferase family are indicated by dotted lines (I, II, III, IV, V, VI, and VII).

1.15 Structural basis for substrate recognition in caffeine synthase gene family

The apparent molecular mass of the native caffeine synthase that was purified from young tea leaves was 61 kD, as estimated by gel-filtration chromatography, and 41 kD, as estimated by SDS-PAGE (Kato et al., 1999). The molecular mass estimated from the predicted amino acid sequence of TCS1 is 41kD and it agrees well with the size of CtCS series (Kato and Mizuno, 2004). Most of motif B' methyltransferases were thought to function as the dimeric form *in vivo* by Zubieta et al., (2003). They clarified the 3.0-Å crystal structure of *Clarkia breweri* SAMT in complex with the substrate salicylic acid, and the demethylated product *S*-adenosyl-L-homocysteine has a protein structure that possesses a helical active site capping domain and a unique dimerization interface and proposed the structural model of TCS1 and 7-MXMT (Figure 1.7). The positioning of the 7-methylxanthine and theobromine molecules in the modeled active sites was accomplished by preserving the hydrogen-bonding pattern that brackets the reactive carboxyl group in motif B' *O*-methyltransferases. This functional dimeric model of caffeine synthase needs to be investigated in detail.

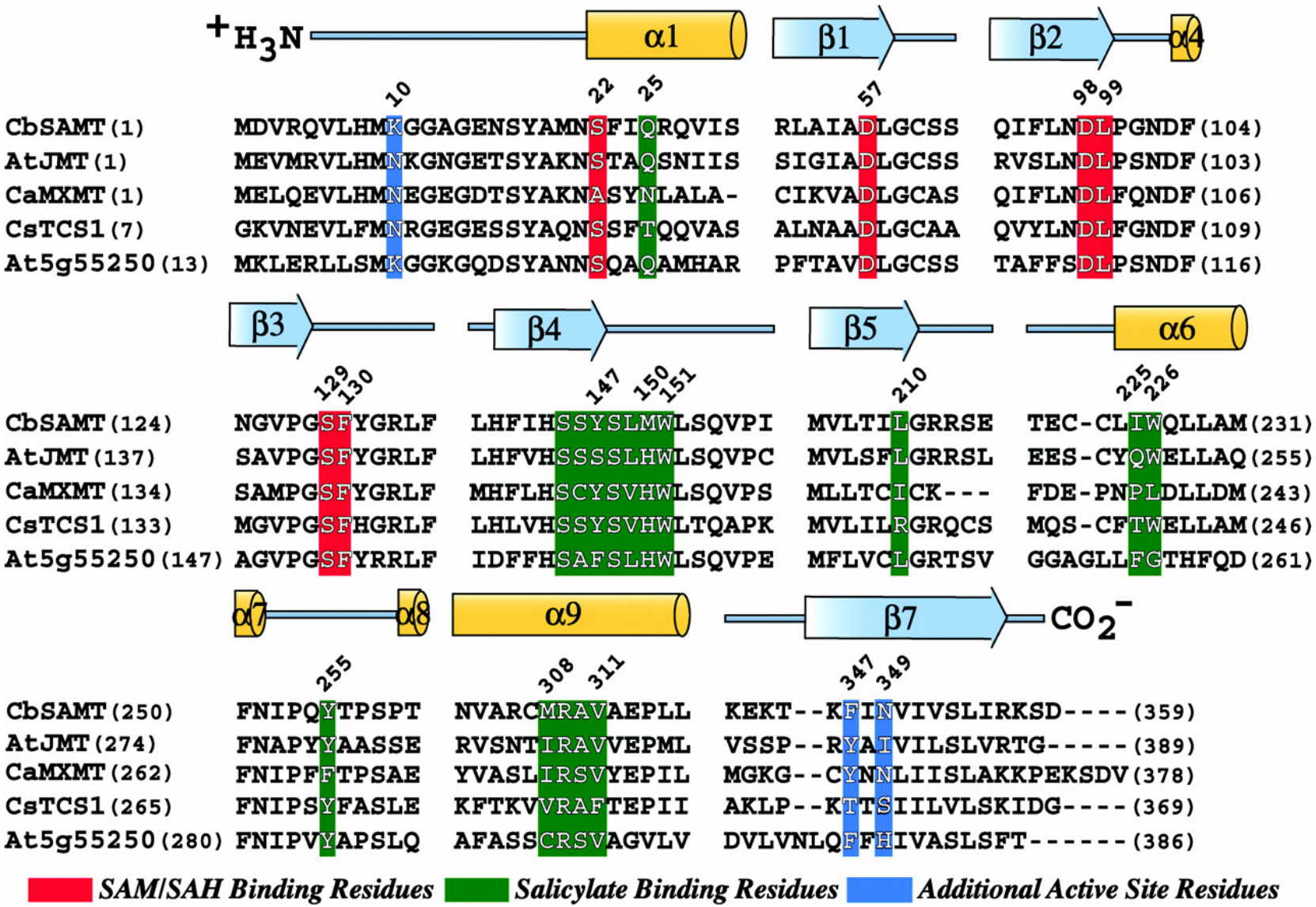

+H3N
α1
β1
β2
α4
10
22
25
57
98
99
CbSAMT (1) MDVRQVLHMKGGAGENSYAMNSFIQRQVIS RLAIADLGCSS QIFLNDLPGNDF (104)
AtJMT (1) MEVMRVLHMNKGNGETSYAKNSTAQSNIIS SIGIADLGCSS RVSLNDLPSNDF (103)
CaMXMT (1) MELQEVLHMNEGEGDTSYAKNASYNLALA- CIKVADLGCAS QIFLNDLFQNDF (106)
CsTCS1 (7) GKVNEVLFMNRGEGESSYAQNSSFTQQVAS ALNAADLGCAA QVYLNDLFGNDF (109)
At5g55250 (13) MKLERLLSMKGGKGQDSYANNSQAQAMHAR PFTAVDLGCSS TAFFSDLPSNDF (116)
β3
β4
β5
α6
129
130
147
150
151
210
225
226
CbSAMT (124) NGVPGSFYGRLF LHFIHSSYSLMWLSQVPI MVLTILGRRSE TEC-CLIWQLLAM (231)
AtJMT (137) SAVPGSFYGRLF LHFVHSSSSLHWLSQVPC MVLSFLGRRSL EES-CYQWELLAQ (255)
CaMXMT (134) SAMPGSFYGRLF MHFLHSCYSVHWLSQVPS MLLTCICK--- FDE-PNPLDLLDM (243)
CsTCS1 (133) MGVPGSFHGRLF LHLVHSSYSVHWLTQAPK MVLILRGRQCS MQS-CFTWELLAM (246)
At5g55250 (147) AGVPGSFYRRLF IDFFHSAFSLHWLSQVPE MFLVCLGRTSV GGAGLLFGTHFQD (261)
α7
α8
α9
β7
CO2−
255
308
311
347
349
CbSAMT (250) FNIPQYTPSPT NVARCMRAVAEPLL KEKT--KFINVIVSLIRKSD---- (359)
AtJMT (274) FNAPYYAASSE RVSNTIRAVVEPML VSSP--RYAIVILSLVRTG----- (389)
CaMXMT (262) FNIPFFTPSAE YVASLIRSVYEPIL MGKG--CYNNLIISLAKKPEKSDV (378)
CsTCS1 (265) FNIPSYFASLE KFTKVVRAFTEPII AKLP--KTTSIILVLSKIDG---- (369)
At5g55250 (280) FNIPVYAPSLQ AFASSCRSVAGVLV DVLVNLQFFHIVASLSFT------ (386)
SAM/SAH Binding Residues
Salicylate Binding Residues
Additional Active Site Residues

Figure 1.7 Partial sequence alignments of representative SAMT-like methyltransferases. Adapted from Zubieta et al., (2003)

Regions of secondary structure shown as labeled gold cylinders (helices) and blue arrows (-strands) are those of *Clarkia* SAMT, whose structure has been experimentally determined (Zubieta et al., 2003). Residues highlighted by red boxes are SAM/SAH binding residues in SAMT, residues indicated with green boxes are amino acid side chains in contact with SA in SAMT and substrates in other members of this methyltransferase family, and residues colored with blue boxes are additional side chains lining the accessible active-site pocket of SAMT. *At5g55250* encodes an *Arabidopsis* IAMT (Zubieta et al., 2003).

1.16 Subcellular distribution and transport of purine alkaloids

Subcellular compartmentation of purine alkaloids has not been studied in detail. However, it is in keeping with many other secondary metabolites that caffeine and other purine alkaloids are stored in vacuoles (Wink, 1997). It is suggested that vacuolar compartmentation of purine alkaloids depends exclusively on the formation of complexes with chlorogenic acids (Baumann and Rohring, 1989; Mösli Waldhauser and Baumann, 1996). The concerted formation of caffeine and chlorogenic acids (mainly 5-caffeoylquinic acid; 5-CQA) has a physiological significance. Caffeine, which easily permeates through all kinds of biological barriers, is physico-chemically complexed by 5-CQA and thus, compartmented in the cell vacuole in order to avoid autotoxicity (Mösli Waldhauser and Baumann, 1996). They concluded that there is a regulatory connection between the complex partners, possibly guided by the cytoplasmic caffeine concentration.

Compartmentation of enzymes of caffeine biosynthesis and SAM cycle in tea has been examined (Kato et al., 1998; Koshiishi et al., 2001). It was observed that caffeine synthase in tea is located in chloroplasts (Kato et al., 1998). Koshiishi et al., (2001) proposed a model in which conversion of methionine to SAM takes place in the cytosol of young tea leaves, and SAM returns to chloroplasts, where it serves as a methyl donor in caffeine biosynthesis; in the process SAM is converted to SAH. Tea SAH hydrolase is more likely to be a chloroplastic enzyme (Koshiishi et al., 2001) and the homocysteine produced in chloroplasts is transported to cytosol.

Examination of the subcellular localization of coffee theobromine synthase (CaMXMT) using the fusion protein of CaMXMT and GFP by contrast in onion epidermal cells revealed that it is present predominantly in the cytoplasm (Ogawa et al., 2001). The PSORT program running with the deduced amino acid sequence also predicted a high possibility of cytosolic localization for CaMXMT. Based on these results (Ogawa et al., 2001) suggested that caffeine biosynthesis occurs in the cytoplasm (cytosol) of cells in buds and young leaves of coffee, in contrast to the localization in chloroplast in tea.

1.17 Caffeine catabolism in plants

Kalberer (1965) was the first to show caffeine catabolism in *Coffea arabica* using ring- and methyl-labeled caffeine to feed leaves. Although degradation was low, he was able to show accumulation of radioactivity in the ureides allantoin and allantoic acid and in three other unknown compounds. Because of the low degradation rate of this species, further feeding experiments were carried out with labeled caffeine and other methylxanthines and xanthine to investigate caffeine catabolic route in coffee. Suzuki and Waller (1984) studied caffeine degradation in the fruits of *Coffea arabica.* They fed the fruits with labeled caffeine, and observed that most of the radioactivity was recovered in caffeine. However, with labeled theophylline significant radioactivity was found in ureides, indicating that first demethylating reaction would be a limiting step in degradation of the alkaloid. They proposed caffeine degradation through theophylline, theobromine, 3-methylxanthine, 7-methylxanthine, xanthine, uric acid, allantoin, allantoic acid, glyoxylic acid, NH3 and CO_2, though radioactivity was not detected in some of these substances. They concluded that in immature fruits both biosynthetic and degradation reactions rates are more rapid than for mature fruits. However, caffeine content in immature coffee fruits was a consequence of higher biosynthetic rate relative to biodegradation, despite degradation being highest in these fruits.

Mazzafera et al., (1991) however, showed that the same was not observed with *C. dewevrei* and *C. salvatrix*, wherein higher caffeine content was observed in mature fruits than in immature fruits. Using labeled caffeine, they observed higher rates of caffeine degradation in immature fruits of *C. eugenioides*, *C. salvatrix*, *C. bengalensis* and *C.*

dewevrei when compared to *C. arabica* and *C. canephora*. In *C. arabica* and *C. canephora*, 90% of radioactivity was recovered in caffeine; and no radioactivity was detected in theobromine and theophylline. However, in *C. eugenioides, C. stenophylla*, *C. salvatrix* and *C. bengalensis* radioactivity was observed in theobromine, theophylline and 3-methylxanthine. They also proved beyond doubt that 7-methylxanthine did not appear in caffeine degradation pathway in coffee fruits. They suggested that *C. dewevrei* has a high turnover of caffeine in immature fruits, decreasing with maturation.

Ashihara et al., (1996b) carried out detailed study with leaves of *C. arabica* at different developmental stages. Feeding experiments with labeled caffeine, theobromine, theophylline and xanthine confirmed the earlier reports that caffeine is degraded mainly through theophylline, 3-methylxanthine, xanthine, uric acid, allantoin, allantoic acid, urea, NH3 and CO_2. Despite the variations observed for 7-methylxanthine in leaves and fruits of *C. arabica* and *C. dewevrei*, in both species 7-methylxanthine is not an intermediate in the demethylating steps occurring between caffeine and xanthine. Instead, it is salvaged by xanthine methylation and re-enters caffeine biosynthesis.

Mazzafera et al., (1994b) investigated the control of caffeine metabolism in *C. dewevrei* and *C. arabica* by measuring N-methyltransferase activities, as well as by feeding experiments with labeled caffeine. Lower caffeine content in immature fruits of *C. dewevrei* was suggested to be due to a low rate of biosynthesis associated with a high rate of degradation. However, mature fruits of this species contain higher caffeine than immature fruits. It is suggested that a change in balance between biosynthesis and degradation occurs such that when degradation is lower than synthesis, caffeine levels increase. The caffeine contents in endosperms from immature and mature fruits of *C. arabica* var. Laurina were 0.8% and 0.6%, respectively, although showing the same degradation rate as *C. arabica* var. Mundo Novo having 1.3% caffeine (Mazzafera et al., 1991). Despite the similar 1-*N*- and 3-*N*- methyltransferase activities in endosperms of *C. arabica* var. Laurina and Mundo Novo as observed by Mazzafera et al., (1994b), the differences in caffeine content between these two *C. arabica* varieties needs to be investigated.

Ashihara and Crozier (1999b) expanded the work of Mazzafera et al., (1994b) to other species containing low caffeine in leaves. They observed that *C. salvatrix*, C. *eugenioides* and *C. bengalensis* synthesized caffeine in similar way to *C. arabica*, but at much lower rates, especially in *C. bengalensis*, partially explaining its low alkaloid content. Radiolabel feeding experiments showed that the low caffeine content in leaves of C. *eugenioides* was due to a strong catabolic activity. The earlier proposed degradation pathway from caffeine → theophylline → 3-methylxanthine→ xanthine→ was observed. In contrast to previous studies with *C. arabica*, 7-methylxanthine was not formed from theophylline in *C. salvatrix*, C. *eugenioides* and *C. bengalensis* leaves. Association of low biosynthesis with rapid degradation, as in immature fruits of *C. dewevrei* seems to be the reason for low caffeine content in young leaves of C. *eugenioides.* In old leaves of C. *eugenioides*, the low caffeine is as a result of still lowered biosynthesis coupled with active degradation.

In most of the studies on caffeine degradation using labeled compounds, radioactivity is usually associated with ureides but uric acid is not detected. In addition, the low recovery of radioactivity in xanthine has been attributed to the fast rate of degradation imposed by xanthine oxidase/dehydrogenase. The control of these metabolic steps was studied in *C. arabica* and *C. dewevrei* (Vitória and Mazzafera, 1999). Leaves of both species accumulate 0.2 - 0.4 mg/g of ureides and fruits approximately 3 times more. Incubations with labeled xanthine showed that young leaves degrade ureides less efficiently, since they retained more radioactivity in these compounds. No uric acid was detected in leaves. However *C. dewevrei* fruits were less efficient in xanthine degradation, since proportionally more radioactivity was recovered in ureides and uric acid compared with fruits of *C. arabica*.

These data were in agreement with the observation that leaves had the highest activity of xanthine dehydrogenase while this activity was very low or not detected in fruits of *C. dewevrei*. Similar results were observed with uricase and urease. Instead of an oxidase, a dehydrogenase activity was characterized as the enzyme responsible for xanthine conversion to uric acid. Activity determinations showed that uricase was very active which

explains why uric acid is usually not detected in incubations with labeled compounds. The activity of uricase in leaves varied from 1,500 to 5,300 ng of allantoin formed per mg protein per hour while xanthine dehydrogenase activity was 18 to 399 ng of uric acid formed per mg protein per hour. It appears that the radioactivity recovered in ureides is related to the levels of allantoinase and allantoate amidohydrolase.

Based on the studies so far, the consensus main catabolic pathway in coffee plants is: caffeine → theophylline → 3-methylxanthine → xanthine → uric acid → allantoin → allantoic acid → glyoxylic acid + urea → NH_3 +CO_2 (Mazzafera, 2004). Depending on the plant species, other minor routes may operate with the formation of theobromine and 7-methylxanthine, which are salvaged for caffeine formation since they also appear in the biosynthetic pathway. Conversion of caffeine to theophylline is the rate-limiting step in caffeine biodegradation in coffee plants.

In humans and other animals, caffeine is first degraded either to paraxanthine, theobromine or theophylline by cytochrome p450 (Figure 1.8). In plants, limited success has been achieved in detecting demethylase activity. Strong indication that a cytochrome p450 would be responsible for nicotine demethylation to nornicotine was suggested by Chelvarajan et al., (1993). Maximum activity was recovered from microssomes, together with indirect evidence obtained with cofactors and inhibitors that indicated nicotine breakdown was carried out by a cytochrome p450. Initial attempts to detect this demethylating activity in coffee were unsuccessful (Mazzafera, 2004). Cesarino and Mazzafera (2005) recently obtained evidence for involvement of cytochrome p450 mechanism in caffeine demethylation. Using p450 inhibitors, they were able to arrest caffeine degradation in leaves of *C. dewevrei*, a species known to have fast caffeine catabolism. Leaf discs were infiltrated with caffeine with or without the inhibitors and after different incubation periods, the alkaloid was extracted and analyzed. Caffeine degradation was reduced with two inhibitors, indicating that caffeine demethylase is indeed a P450 enzyme and that it belongs to a specific class of this large oxygenase group, as has been observed for animals and microorganisms. A search in the Brazilian EST Coffee Genome Project revealed the presence of the P450 enzyme (Cesarino and Mazzafera, 2005).

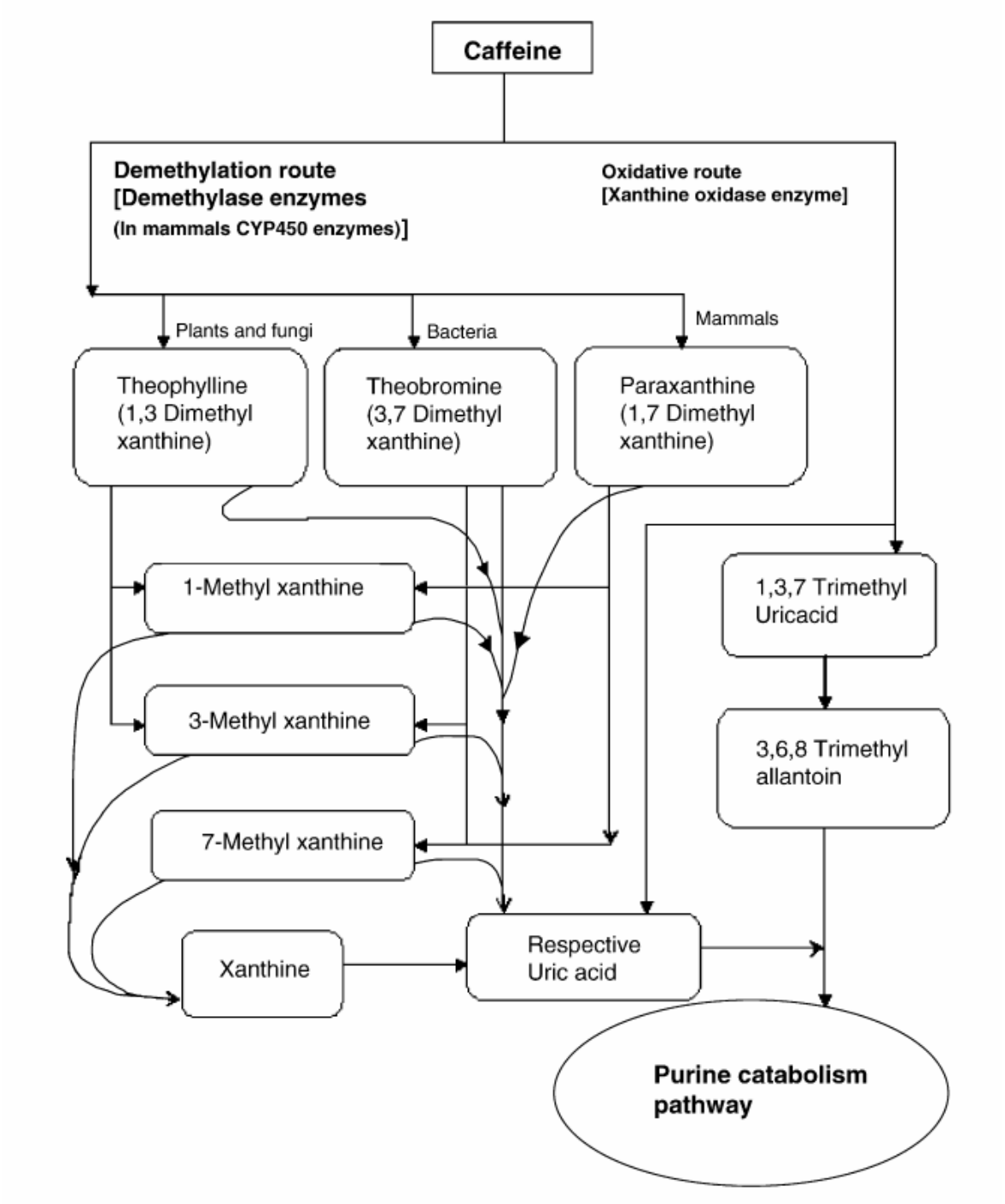


Figure 1.8 Overall view of caffeine degradation in prokaryotes and eukaryotes
Adapted from Gokulakrishnan et al., (2005)

The cloning of this gene will help the researchers to study the control of catabolism in greater detail in particular, and the overall caffeine content in plants in general. It is widely believed that the control of caffeine levels in plants is a function of the balance between rates of synthesis and degradation; and this balance seems to vary depending on the plant species and the tissue developmental stage (Mazzafera, 2004).

1.18 Biotechnology of purine alkaloids

A growing belief that the ingestion of caffeine can have adverse effects on health has resulted in an increased demand for decaffeinated beverages (Ashihara and Crozier, 2001). The market for decaffeinated coffee has increased upto 10% of coffee consumption worldwide, despite the loss of key flavour compounds in the industrial decaffeinating process (Silvarolla et al., 2004). In the long term, the increasing demand for decaffeinated coffee would be better met by the use of *Coffea* species with beans that contain significantly lower levels of caffeine than *C. arabica* and *C. canephora*.

Caffeine content is a quantitative character with high heritability (Montagnon et al., 1998). Barre et al., (1998) deduced from interspecific crosses between the cultivated coffee species *Coffea liberica* var *dewevrei* and caffeine free *C. pseudozanguebariae* that presence or absence of caffeine is under control of one major gene with double recessive genotype conditioning absence. Although a breeding programme to obtain low caffeine producing plants is feasible, it would take 25 years or more to establish and stabilize the desired traits. Unfortunately, absence of caffeine is linked to the presence of a heteroside diterpene causing bitterness, which is also under the control of one major (codominant) gene. Molecular markers for these characteristics could accelerate the search for a genotype lacking both caffeine and the bitter taste. Since *C. arabica* is tetraploid and most other species are diploid (n=2), there are genetic barriers also that prevent effective crossings between *C. arabica* and other *Coffea* species such as *C. salvatrix, C. bengalensis*, that contain much lower levels of caffeine. In these circumstances, the use of genetic engineering to produce transgenic caffeine-deficient tea and coffee plants might be a more practical proposition (Ashihara and Crozier, 2001).

Very recently, Silvarolla et al., (2004) reported a wild accession of *Coffea arabica* plants that accumulated theobromine (about 6.1 mg g^{-1} dry weight), the immediate precursor of caffeine. Using radiolabelled precursors, they indicated that the naturally low caffeine content in these plants relative to commercial variety was not due to enhanced degradation of caffeine, but due to lower biosynthesis. Since theobromine accumulated in these plants, and no caffeine synthase activity was detectable in leaves of these plants, they concluded that the caffeine synthase gene is mutated in these plants. Using a conventional breeding approach, this trait could be successfully transferred to highly productive cultivars of *C. arabica*, because hybridizations will occur inside the same autogamous species.

The cloning of the *N*-methyltransferase gene is an important step towards the production of transgenic caffeine-deficient tea and coffee plants through gene silencing with antisense mRNA or RNA interference. The genes encoding *N*-methyltransferases for caffeine biosynthesis have been cloned and transgenic caffeine-deficient *C. arabica* and *C. canephora* plants have been produced (Ogita et al., 2003; 2004) An alternative way to produce transgenic caffeine-deficient tea and coffee plants would be to over express a *N*-7-demethylase encoding gene associated with caffeine degradation from *C. eugenioides* (Ashihara and Crozier, 2001). Caffeine content can also be lowered by changing the expression of the chlorogenic acid pathway using genetic engineering (Mösli Waldhauser and Baumann (1996).

The *N*-methyltransferases activity can be also regulated by the control of SAM/SAH ratio (fine control) (Kato et al., 1999). The reduction in SAH hydrolase activity may lead to reduction in the activity of *N*-methyltransferases. However, this reduction in SAH hydrolase activity may affect other physiological processes also that involve SAM. Keya et al., 2003 proposed the suppression of inosine-5' monophosphate dehydrogenase (IMPDH) gene expression as an alternative way to produce caffeine-deficient transgenic tea and coffee plants. Transgenic tea plants with reduced IMPDH activity offer the prospect of a beverage with low caffeine content coupled with enhanced flavour, owing to the accumulation of IMP.

Scope of the present investigation

The literature survey suggests that the *N*-methyltransferases (NMTs) that are involved in the caffeine biosynthetic pathway in coffee comprise a multigene family. Several cDNA clones for NMTs from coffee were reported. However, there were no reports for the genomic sequences for the NMT genes. The studies on NMT gene structure and organisation were lacking. Therefore, it was envisaged that PCR based genomic cloning of the NMT genes could provide valuable information regarding the number and size of introns and exons present in NMT genes. The study of introns and exons could possibly shed light on the mechanisms for origin of NMT multigene family in coffee and possible functional divergence among different members of the gene.

The survey of literature and patents indicated a distinct lacuna in the studies related to NMT gene regulation. The isolation of promoter for the NMT genes has not been hitherto reported and hence it became one of the major objectives of this investigation. The promoter isolation could help in understanding the NMT gene regulation and also open up the possibilities of regulating caffeine content in coffee plants.

There are wide variations in the caffeine contents of different coffee species, the reasons for which are unknown. Very little is known about the NMT genes in *Coffea* species other than *C. arabica*. It can be speculated that differential transcription of various NMT genes could be one of the reasons for such variations in the caffeine content. Study of possible polymorphisms for the NMT genes and their corresponding promoters in different *Coffea* species could be vital in understanding differential transcription of various NMT genes.

The production of transgenic caffeine deficient coffee plants by RNAi using specific region for theobromine synthase gene was reported earlier, wherein spreading of silencing to non-specific gene was observed. The use of conserved region that is common for all the NMT genes should theoretically result in silencing of all the NMT genes. However, other posttranscriptional gene silencing strategies were not explored for caffeine down regulation. The use of conserved region of NMT genes for different posttranscriptional

gene silencing strategies could be a viable option for obtaining phenotypes with desirable purine alkaloid profiles.

With this background knowledge, the objectives of the present research work were laid as follows:

1. To clone the *N* - Methyltransferase gene from *Coffea canephora* and study the gene organization and polymorphism of genes associated with caffeine biosynthetic pathway in different species.

2. To clone the promoter for NMT gene from coffee and analyze them using vectors with reporter gene.

3. To use the posttranscriptional gene silencing constructs for *N* - Methyltransferase gene in coffee transformation.

The subsequent chapter deals with the molecular cloning of *N*-Methyltransferase gene from *Coffea canephora* using PCR based genomic cloning approach and preliminary characterization of the gene by nucleic acid sequencing and analyses.

CHAPTER II

MOLECULAR CLONING OF THE N - METHYLTRANSFERASE GENE FROM COFFEA CANEPHORA

2.0 ABSTRACT

Primers for PCR amplification of *N*-methyltransferase (NMT) genes from *Coffea canephora* were designed based on the available cDNA sequences for NMT genes of *Coffea arabica*. Different sets of primers were used to obtain amplicons for partial and full-length NMT gene. Various PCR products were cloned and sequenced. Variations were observed with respect to restriction patterns among the clones for NMT gene fragment. Sequence analysis of these clones revealed the presence of introns in the gene. High degree of similarity in the nucleotide sequences of exon regions and differences in the length and sequence of introns were observed, suggesting that these clones may belong to different members of the NMT multigene family. Using primers based on 5′ and 3′ untranslated regions specific for theobromine synthase, involved in the second methylation step, amplicons for full-length theobromine synthase gene were obtained. Sequencing of the clones for full-length gene revealed the presence of four exons interrupted by three introns. Based on the insertions and deletions present in the cDNA sequences of various NMT genes reported so far and the length of their open reading frames, it was concluded that genomic clone obtained corresponded to the theobromine synthase gene encoding for 378 amino acids. Variations with respect to the size and sequence of introns in two of the full-length clones for theobromine synthase were observed, though one of them contained one base nucleotide deletion in the fourth exon. From this study, it appears that the coffee plants possess multiple alleles and/or multiple copies of the same gene; some of which may have undergone pseudogenisation.

2.1 INTRODUCTION

The general structure and regulatory processes of genes in plants and animals are similar. A generalized structure of plant gene is depicted in Figure 2.1.1, wherein three major regions can be identified. These are:

(i) 5′ sequences that are not transcribed but regulate the expression of the genes.

(ii) The transcribed sequences which includes a 5′untranslated region and protein-coding sequence (which may be intervened by introns). The core of the gene is the coding region, which contains the nucleotide sequence that is eventually translated into the sequence of amino acids in the protein.

(iii) 3′ sequences that are transcribed, but not translated.

Most of the plant genes have similar general architecture as described above, however there are few exceptions. For instance, the histone genes have non- polyadenylated mRNAs, while maize zein genes lack introns. On average, 5′ and 3′ UTRs are less conserved across species than protein coding sequences, but are more conserved than untranscribed sequences (Pesole et al., 2001).

The 5′ untranscribed region includes promoter elements upstream to the coding region. Promoter elements are classified into two categories: common core promoter elements that are needed for basal transcription initiation; and gene specific regulatory cis-acting elements located upstream (Roeder, 1996). The structural and functional features of various promoter elements are dealt in detail in chapter IV.

The 5′-untranslated region (5′ UTR) in the transcribed region plays an important role in regulation of gene expression. It can influence the efficiency of translation by its secondary structure, the context of the AUG codon, and the existence of upstream AUG or upstream open reading frames (Futterer and Hohn, 1996). The average length of 5′ UTRs ranges between 100 and 200 nucleotides. Evidences indicate that 5′ UTR can also act in transcriptional and post-transcriptional regulation (Bolle et al., 1994; Hua et al., 2001). Messenger RNAs (mRNAs) encoding proteins involved in developmental processes, such

as growth factors, transcription factors or proto-oncogenes, all of which need to be strongly and finely regulated, often have 5′ UTRs that are longer than average (Kozak, 1987), with upstream initiation codons or open reading frames (ORFs) and stable secondary structures that hamper translation efficiency. Other specific motifs and secondary structures in the 5′ UTR can also modulate translation efficiency.

According to the scanning model of translation initiation, the eukaryotic ribosome binds to 5′- terminal cap and starts scanning the mRNA, typically until detecting the first AUG where it initiates translation. However, the context around an AUG codon is critical for this process and determines whether the ribosome initiates at the 5′ ultimate AUG codon or proceeds scanning to the next AUG (Kozak, 1999). Higher plants have an AC-rich consensus sequence, caA(A/C)a**AUG**GCg as a context of translation initiator codon. The context of the AUG initiator codon in mRNAs of dicotyledonous plants is aaA(A/C)a**AUG**GCu which is similar to the higher-plant consensus but monocot mRNAs have c(a/c)(A/G)(A/C)c**AUG**GCG as a consensus which exhibits an overall similarity with the vertebrate consensus. Purines at the critical –3 and G at +4 positions are known to be required for the fidelity of translation initiation (Joshi et al., 1997). The presence of an upstream AUG correlates with a long 5′ UTR and with a ‘weak’ start codon context of the AUG that is usually used. Transcripts with an optimal start codon context have short 5′ UTRs without upstream AUGs, suggesting that upstream AUGs may have a role in keeping the basal translational level of a gene low (Mignone et al., 2002). The genomic region corresponding to the 5′ UTRs of an mRNA may also contain introns (Fu et al., 1995). Atleast 9% of *Arabidopsis* genes have introns in their UTRs (Haas et al., 2002; Zhu et al., 2003).

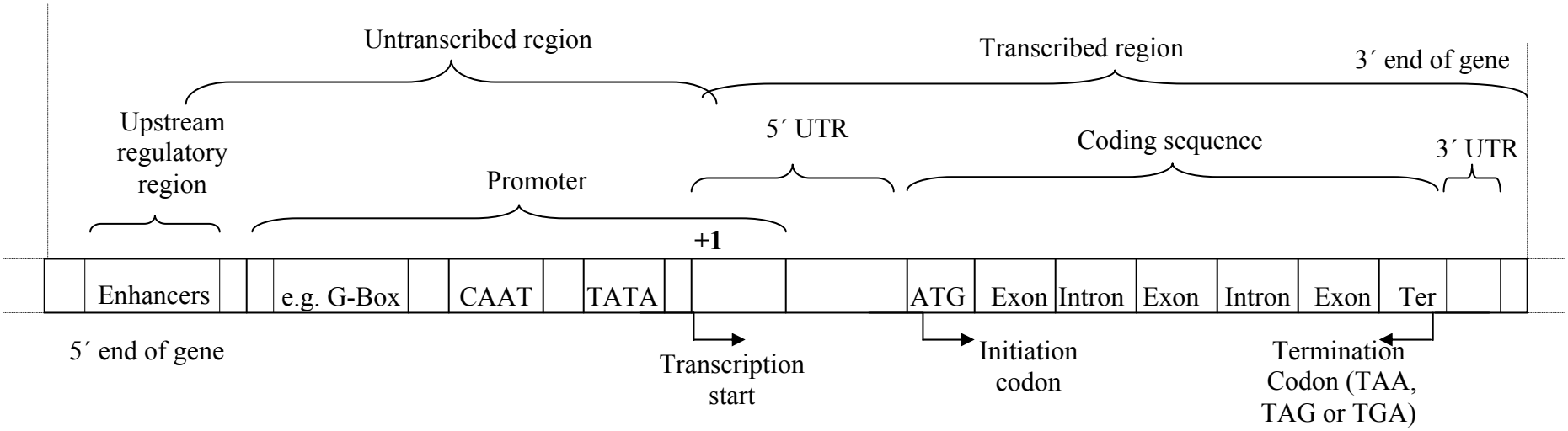


Figure 2.1.1 Generalized structure and organization of a plant gene

The presence of intervening sequences (introns) is an outstanding feature of eukaryotic genes, including plant genes. Introns present in primary gene transcripts are removed and exons are ligated to produce translationally competent mRNAs, which is known as pre-mRNA splicing. Pre-mRNA splicing is one of the fundamental processes in constitutive and regulated gene expression in eukaryotes. The intron and exon organisation of higher plant genes is similar to that of vertebrates (Brown and Simpson, 1998). Most plant genes (80-85%) are interrupted by introns, ranging in size from ~60 to 10,000 nt. About two-thirds of plant introns are < 150 nt long (Lorkovic et al., 2000). The consensus sequences of the 5′ splice site and the 3′ splice site are AG/GTAAGT and TGCAG/G respectively. The GT dinucleotide at the 5′ end and AG dinucleotide at the 3′ end of the introns are consensus intron/exon borders of plant genes (Brown et al., 1996).

T or AT richness is a general feature of plant introns. On an average, in both dicots and monocots, introns are ~ 15% more AT-rich than exons, the T residues being largely responsible for this difference. T- rich sequences, typically distributed throughout the entire length of plant introns, are required for efficient intron processing and splice-site selection (Brown and Simpson, 1998). The presence of functional introns in transcription units can dramatically enhance mRNA accumulation, especially when the intron is placed close to the 5′ end of the gene. The extent of this intron-mediated enhancement depends on the origin of intron, the flanking exonic sequences, strength of the promoter used to drive RNA expression, and on the type and physiological state of the cells (Lorkovic et al., 2000). Primary transcripts of many genes are alternatively spliced, producing different mRNA forms that encode proteins with functional differences.

The region following the termination codon marking the end of the open reading frame of the coding region constitutes the 3′-untranslated region (3′-UTR). The context of the stop codon, in particular, the immediate downstream (+1) base, is important for termination. Additional complexity is introduced into the termination mechanism by the fact that there are three stop codons, for which the context effects seem to differ substantially. However, beyond the +1 position, the 3′-UTRs have not been reported to have a distinct consensus analogous to the Kozak consensus in the 5′-UTR (Ozawa et al., 2002). A poly (A) tail is found at the end of nearly every fully processed eukaryotic

mRNA and its proposed functions include conferring mRNA stability, promoting mRNA's translational efficiency, and having a role in transport of processed mRNA from the nucleus to the cytoplasm (Colgan and Manley, 1997). The poly (A) tail is added at the signal generated by endonucleage cleavage. Nearly every known mRNA contains a polyadenylation signal sequence, the hexanucleotide AAUAAA, 10–30 bases upstream of the cleavage/polyadenylation site. AAUAAA is found in 90% of all sequenced polyadenylation elements and is one of the most highly conserved sequence elements known. Most plant genes have more than one of the consensus sequences for the polyadenylation signal.

The results of genome-wide comparison of orthologous mRNAs in mammals and fungi (Shabalina et al., 2004) support the evolutionary conservation of the context sequence in the immediate vicinity of the principal signals, namely, the start and stop codons, and the polyadenylation site. These well recognized context elements include the Kozak consensus around the start codon, the +1 base following the stop codon, which is critical for the efficiency of termination, and the GU elements downstream of the polyadenylation AAUAAA motif.

Virtually all plant genes are represented within most plant genomes as small gene families that originate through duplication. A gene family can be defined as 'a set of genes coding for diverse proteins which, by virtue of their high degree of sequence similarity, are believed to have evolved from a single ancestral gene' (Lackie and Dow, 1999). Many of these families are involved in plant defence, secondary metabolism and signaling. It is possible that genes required to undertake quick, large transcription responses are prone to becoming large families. Such a development may help provide a readily tuneable on/off switch for a variety of specialized functions in response to a host of stimuli. In contrast, metabolic functions, house keeping and structural proteins those are more constitutively expressed, tend to be encoded in smaller gene families and more significantly these classes contain many single genes in plants (Sappl et al., 2004).

In plants, gene duplication seems to be the most predominant way of increasing transcript diversity and gene products rather than other mechanisms like alternative splicing and multiple transcription starts (European Union Chromosome 3 Arabidopsis Sequencing Consortium, 2000). Tandem gene duplications and segmental duplications

are the major driving forces or atleast the mechanistic processes in the explosion of many of the gene families (Arabidopsis Genome Initiative, 2000). To persist over long periods of evolutionary time, new gene duplications must be positively selected; otherwise the relentless accumulation of mutations would ultimately convert the duplicate copy to a pseudogene. There are three ways that newly duplicated genes may be of adaptive value to the plant: (1) genetic redundancy may be favored because it increases the production of a limiting product; (2) the protein may take on a new or modified enzymatic function; and (3) there may be a specialization in developmental expression that also permits subtle shifts in catalytic properties so as to optimize pathway throughput in different tissues and at different points in the life cycle of the plant (Durbin et al., 2000). Gene duplication may also result in the development of metabolic pathways governed by multigene families.

In this study, attempts were made to obtain the genomic clones for one of the *N*-methyltransferase genes involved in caffeine biosynthetic pathway in *Coffea canephora* and to study the gene structure in terms of number and sizes of introns and exons.

2.2 MATERIALS AND METHODS

All laboratory reagents were of molecular biology or A.R. grade or higher wherever necessary. Stock solutions and media were sterilized by autoclaving at 121^0C for 20 min where necessary. Reagents and buffers not suitable for autoclaving were sterilized by filter sterilization using Millipore disposable filters (0.4 μm). Standard procedures described by (Sambrook and Russell, 2001) or manufacturer's instructions in the case of commercial kits were followed for all the experiments, or unless specified otherwise.

2.2.1 Isolation of chromosomal DNA from Coffee leaves

Certified seeds of coffee were obtained from Central Coffee Research Institute (CCRI), Chikamaglur district, Karnataka, India and seedlings were obtained after seed germination. Isolation of genomic DNA from fully developed expanded coffee leaves of *Coffea canephora* var. S-274 was performed by slight modification of the protocol described by Paillard et al., (1996). Addition of β-Mercaptoethanol helped to inhibit the polyphenol oxidase activity responsible for coloration of the precipitate.

2.2.1.1 Materials

A. Extraction buffer (50 ml):

d-Sorbitol	0.35M
Sodium metabisulphite	0.5%
Tris-HCl (pH 8.0)	0.1M
EDTA (pH 8.0)	0.005M

3.188g of sorbitol and 0.25g of sodium metabisulphite were added to 35 ml of distilled water and dissolved. 5ml each of 0.5M Tris-HCl solution (pH 8.0) and 0.5M EDTA solution (pH 8.0) were added and the volume was made up to 50 ml with distilled water. The solution was autoclaved and stored.

B. Lysis Buffer (10ml):

CTAB	2%
NaCl	2M
Tris- HCl (pH 8.0)	0.2M
EDTA (pH 8.0)	0.05M

0.2g of CTAB, 4ml of 5M NaCl, 4ml of 0.5M Tris-HCl (pH 8.0) and 1 ml of 0.5M EDTA (pH 8.0) were mixed. The volume was made up to 10 ml with distilled water, and the solution was autoclaved and stored.

C. 0.1X T.E (20 ml):
40µl of 0.5M Tris-HCl (pH8.0) and 4.0µl of 0.5M EDTA (pH 8.0) were added and the volume was made up to 20 ml with distilled water, autoclaved and stored.

2.2.1.2 Protocol

Approximately, 500mg of young coffee leaves were ground to fine powder with liquid nitrogen at room temperature. The frozen powder was suspended in 4ml of extraction buffer taken in a 50 ml Oakridge tube and 100mM β-Mercaptoethanol was added to it. The sample was filtered using muslin cloth. The tubes were centrifuged at 10,000 rpm for 10 min at 4^0C. The pellet was suspended in 100 µl of cold extraction buffer. 140 µl of lysis buffer and 50 µl of sodium lauryl sarcosine (5%) were added to the mixture. The mixture was incubated at 65 ^{0}C for 30 minutes. 200 µl of Chloroform: isoamylalcohol (24:1) was added and the tubes were centrifuged at 10,000 rpm for 10 minutes at 4^0C. The supernatant was transferred to an Eppendorf tube and 9 µl of 3M sodium acetate and 200 µl of cold ethanol were added. The tubes were centrifuged at 10,000 rpm at 4^0C. The pellet was washed again with 70% ethanol and finally suspended in 100 µl of 0.1X TE.

2.2.2. Designing of oligonucleotides

Several sets of primers were synthesized based on the cDNA sequences available in the GenBank (accession nos. AB039725, AB048794). The nucleotide sequence of the primers designed is presented in Table 2.2.1.

Table 2.2.1 The list of primers designed for amplification of coffee NMT genes

S. No	Oligo name	5′-sequence-3′	Length	Location on gene sequence
1	CSF1	5′-CTTCCGGACCAAACACACTT-3′	20	201-220
2	CSR1	5′-CGATCCTATTTTGCGTCCAT-3′	20	387-406
3	CSF2	5′-CGAGGAGTCCATGCATTTTT-3′	20	454-473
4	CSR2	5′-CCTCCTCAACCATGCACTTT-3′	20	838-857
5	CSF3	5′-TTTAAGGATGCGTTCGGAAG-3′	20	637-656
6	CSR3	5′-CACTGAATCCTTTCCCCGTA-3′	20	1216-1235
7	CSF4	5′-TCTTTCCACGGCAGACTCTT-3′	20	440-460
8	CSR4	5′-CACTGAATCCTTTCCCCGTA-3′	20	1216-1235
9	CSF5	5′-AGCTCCAAGAAGTCCTGCAT-3′	20	18-37
10	CSR5	5′-ACTCGGATAGAACCCCAACC-3′	20	1237-1256
11	XMTF1	5′-AGCAGTCGCAATTCGATTGTC-3′	21	1-21
12	XMTR1	5′-ACGACAATACCCGAAAGACC-3′	20	1203-1222

Oligos 1-10 were based on GenBank accession no. AB039725, and 11-12 were based on AB048794

2.2.3 PCR amplification of partial *N*-methyltransferase gene

2.2.3.1 Methodology for PCR amplification

Various combinations of primers were tried to get the largest possible amplicons for NMT gene. Primer combination of CSF5 and CSR2 resulted in amplification of a product approximately 1.7 kb in size. The PCR reaction was carried out by combining the following reaction components in 50µl reaction volume:

Components	**Volume (µl)**	**Final concentration**
Nuclease-free water	37.5	
10 X Reaction Buffer	5.0	1 X
dNTP mix (10 mM)	1.0	0.2 mM
XT-Taq system (3U/ µl)	0.5	0.03U/µl
Primer CSF5 (Forward)	2.0	0.2 µM
Primer CSR2 (Reverse)	2.0	0.2 µM
Template (~100 ng)	2.0	

10X reaction buffer for XT-Taq system (Bangalore genie, India) contains 15 mM $MgCl_2$ and 0.1% gelatin. The contents of the tube were mixed by a brief spin in a micro centrifuge. The reaction was carried out in a thermocycler GeneAmp PCR System 9700 (Perkin-Elmer, USA) and the reaction parameters were as follows:

a. Initial Denaturation: 94^0C for 4 min
b. Denaturation: 94^0C for 30 sec
c. Annealing: 60^0C for 45 sec
d. Extension: 72^0C for 2min

(b–d) 30 cycles

e. Final extension: 72^0C for 10 min

2.2.3.2 Analysis of PCR product by agarose gel electrophoresis

A 10 µl aliquot of the PCR product was analyzed by agarose gel (1.2%) electrophoresis as described below. The size of the partial NMT gene amplicon was checked by comparing with a 3kb DNA Marker (MBI Fermentas, Lithuania).

2.2.3.2.1 Materials and solutions

1. Agarose (SRL, Mumbai, India).
2. TAE 50 X buffer:
   24.2 g Tris base, 5.71 ml of glacial acetic acid and 10 ml of 0.5 M EDTA (pH 8.0) were added to 80 ml of distilled water. The pH was adjusted to 7.2 and the final volume was brought to 100 ml with distilled water. The buffer was sterilized by autoclaving and stored at room temperature.
3. 1kb DNA marker (MBI Fermentas, Lithuania)
4. Gel casting boat
5. Mini gel apparatus and power supply (Bangalore Genie, India).
6. Ethidium bromide stock solution (10 mg ml^{-1}):
   10 mg of ethidium bromide (Sigma, MO, USA) was dissolved in 1 ml of distilled water. The solution was stored in a micro centrifuge tube wrapped in aluminium foil at 4^0C.

2.2.3.2.2 Methodology

The boat was sealed with an adhesive tape and the comb was placed for the wells. 1.2 g of agarose was added to 100 ml of 1 X TAE buffer. The mixture was heated on a hot plate to allow the agarose to dissolve. The solution was cooled to 50^0C and poured into

the sealed boat. The gel was allowed to polymerize. The comb and the adhesive tape were removed and the gel was placed in the electrophoresis tank with sufficient volume of 1 X TAE buffer to cover the surface of the gel. The PCR reaction sample and the standard DNA size marker were loaded in the wells. Electrophoresis was carried out at 50 volts till the dye reached 3/4th of the gel. The gel was removed from the tank and stained by soaking in a solution of 0.5 μg ml^{-1} ethidium bromide for 30 min at room temperature. The gel was destained in distilled water for 10 min, examined on a UV transilluminator and documented using Gel Documentation system (Herolab, Germany).

2.2.3.3 Purification of coffee NMT amplicon

The amplicon for coffee partial NMT gene was purified using GenElute PCR Clean - Up Kit (Sigma, USA) following manufacturers instructions.

2.2.4 A-tailing of PCR products using *Taq* DNA polymerase

A-tailing of the purified coffee NMT amplicon was carried out by the method described by (Kobs, 1997). To 5 μl of purified PCR fragment, 1 μl of *Taq* DNA Polymerase reaction buffer (1X) and 1 μl of 25 mM $MgCl_2$ were added. dATP to final concentration of 0.2 mM and 5 Units (2 μl) of *Taq* DNA polymerase (Bangalore Genie, India) were added to the reaction. The reaction was incubated at 70^{0}C for 20 to 30 min. To remove the residual dATP present in the reaction mixture, the PCR product was purified using GenElute PCR Clean - Up Kit (Sigma, USA). The A-tailed PCR product was ligated to T-tailed vector.

2.2.5 T/A cloning of partial NMT gene amplicon for coffee

2.2.5.1 Ligation of A-tailed PCR product to T-tailed vector

The A-tailed purified PCR product for partial NMT gene was T/A cloned by ligating to pTZ57R/T vector using InsT/A Clone PCR Product Cloning Kit (MBI Fermentas, Lithuania) and transforming competent cells of *Escherichia coli* strain DH5α.

2.2.5.1.1 Methodology

The following components were added in a thin-walled 0.2 ml PCR reaction tube:

Plasmid vector pTZ57R/T DNA	:	2.0 µl
Purified PCR fragment	:	10.0 µl
10X Ligase Buffer	:	3.0 µl
PEG 4000 solution	:	3.0 µl
T4 DNA Ligase, 5U/µl	:	1.0 µl
Deionized water (to make upto 30.0 µl)	:	11.0 µl

The reaction components were mixed by pipetting or brief spin. The reaction was incubated at 22^0C for four hours. The enzyme was inactivated by heating the reaction to 65^0C for 10 min.

2.2.5.1.2 Solutions

10 x Ligation Buffer*

400 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT, 5.0 mM ATP (pH 7.8)

PEG 4000*

10 x (50% w/v) PEG 4000 solution

T4 DNA Ligase, 5U/µl*

Prepared in 20 mM Tris-HCl (pH 7.5), 1 mm DTT, 50 mM KCl, 0.1 mM EDTA and 50% glycerol.

* Supplied with the kit

2.2.5.2 Transformation of *E. coli* using the ligation reaction mix

2.2.5.2.1 Preparation of competent cells using $CaCl_2$

A single colony of *E. coli* (DH5α strain) from a plate, freshly grown for 16-20 h at 37^0C was picked and transferred into 50 ml of LB broth in a 250 ml conical flask. The culture was incubated at 37^0C with rigorous shaking. The OD_{600} of the culture was determined periodically to monitor cell growth. When the OD_{600} reached 0.40-0.50, the cells were transferred aseptically to 50 ml sterile polypropylene tube. The culture was cooled by storing the tube on ice for 10 minutes. The cells were recovered by centrifugation at 4000 rpm for 10 minutes at 4^0C. The media was decanted from the cell pellet. The cell pellet was resuspended in 10 ml of ice-cold 0.1 M $CaCl_2$ and stored on

ice. The cells were recovered by centrifugation at 4000 rpm for 10 minutes at 4^0C. The fluid from the cell pellet was decanted and the tubes were kept in an inverted position for 1 minute to allow the last traces of fluid to drain away. The cell pellet was resuspended in 2.0 ml of ice-cold 0.1 M $CaCl_2$ and cells were stored at 4^0C overnight.

2.2.5.2.2 Transformation of competent cells

About 200 µl suspensions of competent cells were added to sterile micro-centrifuge tubes. Plasmid DNA (~50 ηg) or 2 to 5 µl of ligation mixture was added to each tube. The contents of the tubes were mixed by swirling gently and the tubes were stored on ice for 30 minutes. Control samples were included as following: (a) competent cells that received standard supercoiled plasmid DNA and (b) competent cells that received no plasmid DNA. The tubes were transferred to water bath set at 42^0C for 90 seconds to subject the cells to heat shock. The tubes were rapidly transferred to ice and the cells were allowed to chill for 1-2 minutes. 800 µl of SOC medium was added to each tube and the cultures were incubated for 45 minutes at 37^0C in a shaker incubator set at 150 rpm.

2.2.5.2.3 Selection of transformants/recombinants

About 100 µl of transformation mix was plated onto LB agar plates containing 100 µg ml^{-1} ampicillin, 0.5 mM IPTG and 80 µg ml^{-1} X-Gal. The plates were incubated at 37^0C overnight for the colonies to grow.

2.2.5.2.4 Analysis of transformants/recombinants

The transformants were checked for the presence of plasmid by isolating plasmid DNA and performing 0.8% agarose gel electrophoresis. Restriction digestion experiments were set up to check for insert release from recombinants.

2.2.5.2.5 Solutions & Media

Luria-Bertani broth (LB)	**(per liter)**
Bacto-tryptone	10.0 g
Bacto-Yeast extract	5.0 g
Sodium chloride	10.0 g

The pH was adjusted to 7.0 with 2N NaOH and the total volume was made up to 1 liter with deionized water.

SOB (per liter)

Bacto-tryptone	20.00 g
Bacto-Yeast extract	5.00 g
Sodium chloride	0.60 g
Potassium chloride	0.19 g
Magnesium sulphate	10.0 mM (added from 1.0 M stock)
Magnesium chloride	10.0 mM (added from 1.0 M stock)

The first four components and the magnesium salt were autoclaved separately and then mixed to constitute the SOB medium.

SOC (per 100 ml)

To 1.0 ml of SOB, 7 μl of filter-sterilized (Millipore, 0.4 μm) glucose solution (50%w/v) was added.

0.1 M $CaCl_2$ stock solution

1.47 g of $CaCl_2$ was dissolved in 100 ml of deionized water. The solution was sterilized by filtration and stored as 20 ml aliquots at -20^0C.

Ampicillin stock solution

100 mg ampicillin (Sigma, USA) was dissolved in 1.0 ml of deionized water. The solution was sterilized by filtration and stored at -20^0C and used at a working concentration of 100 μg ml^{-1}

0.1 M IPTG stock solution

0.12 g of IPTG was dissolved in 5.0 ml of deionized water. The solution was filter-sterilized and stored as aliquots at -20^0C.

X-Gal stock solution

100 mg of X-Gal was dissolved in 2.0 ml of N, N'-dimethylformamide (DMF). The solution was stored in micro centrifuge tube, wrapped in aluminium foil at -20^0C.

2.2.6 Isolation of plasmid DNA from the transformed colonies (Birnboim and Doly, 1979)

2.2.6.1 Methodology

Single colonies of appropriate strain were inoculated in 2 ml of LB broth containing required antibiotic and grown overnight in a shaker incubator at 37^0C and 180 rpm. 1.5 ml of the overnight culture was transferred to a 1.5 ml micro centrifuge tube and the cells were harvested by centrifugation at 10,000 rpm for 2 min. After discarding the supernatant, 100 µl of solution I was added and vortexed vigorously until no visible clumps of cells were observed. The samples were kept on ice for 5 min. About 200 µl of freshly prepared alkaline solution (solution II) was added to the tube and mixed gently by inverting the tubes several times until the lysed cell suspension became clear. The samples were kept on ice for 5 min, 150 µl of ice-cold potassium acetate solution (solution III) was added, and tubes were inverted gently. The tubes were centrifuged at 10,000 rpm for 10 min. The supernatant was transferred to a fresh tube and equal volume of phenol- chloroform was added and vortexed thoroughly. Centrifugation at 10,000 rpm for 10 min separated the two phases. The upper aqueous phase was transferred to a fresh tube and equal volume of chloroform was added. The tubes were centrifuged at 10,000 rpm for 10 min. The upper aqueous phase was transferred to a fresh tube and 2 volumes of absolute ethanol were added. The tubes were kept at -20^0C for 1 hr to overnight for precipitation. The tubes were centrifuged at 10,000 rpm for 10 min and the supernatant was discarded. The pellet was washed with 300 µl of 70% ethanol and the air-dried pellet was dissolved in 20 µl of TE buffer. Samples were tested by carrying out agarose gel (0.8%) electrophoresis along with control plasmid.

2.2.6.2 Solutions and Reagents

LB Medium	**(g/liter)**
Tryptone	10.0
Yeast extract	5.0
Sodium chloride	10.0

The pH was adjusted to 7.0 with 2N NaOH and the total volume was made up to 1 liter with deionized water.

Solution I

50 mM glucose

25 mM Tris-Cl (pH 8.0)

10 mM EDTA (pH 8.0)

Solution II

0.2 N NaOH (freshly prepared from 10 N NaOH)

1.0% SDS. Prepared freshly before use.

Solution III

5.0 M Potassium acetate	60.0ml
Glacial acetic acid	11.5 ml
Distilled water	28.5 ml

The resulting solution is 3.0 M and 5.0 M with respect to potassium and acetate, respectively.

2.2.7 Double restriction digestion of recombinant plasmids for insert release

2.2.7.1 Materials

1. Restriction enzymes: E*co*RI, B*am*HI or H*in*dIII (MBI Fermentas, Lithuania)
2. 10 X restriction enzyme buffer (MBI Fermentas, Lithuania)
3. Nuclease-free deionized water.

2.2.7.2 Methodology

1. The following constituents were added in a micro centrifuge tube in the order stated:

Constituents	Volume (µl)
Nuclease-free water	13.0
Restriction enzyme 10 x buffer	2.0
Plasmid DNA	3.0
Restriction enzyme 1	1.0
Restriction enzyme 2	1.0
Final volume	20.0

The contents of the tube were mixed gently by pipetting and the tube was centrifuged briefly at 10, 000 x g to collect the contents at the bottom of the tube. The reactions

were incubated at 37^0C for 4 - 8 hrs. The samples were analyzed by agarose gel electrophoresis.

2.2.8 Sequencing of the clones

DNA sequencing was carried out using M13F and M13R primers by dideoxy chain termination method (Sanger et al., 1977). The reaction was carried out in an automatic DNA sequencer (ABI prism, Applied Biosystems, USA) with fluorescent dideoxy chain terminators at the Dept. of Biochemistry, University of Delhi, South Campus, New Delhi.

2.2.9 PCR amplification of full length *N*-Methyltransferase gene

2.2.9.1 PCR conditions

The primers XMTF1 and XMTR1 (Table 2.2.1) based on the 5′ and 3′ UTRs of the cDNA sequence of theobromine synthase (GenBank Accession no: AB048794) were used for the specific amplification of theobromine synthase gene. The two cycle parameters for 'Touchdown PCR' used on PE Biosystems GeneAmp PCR System 9700 were as follows: for initial 7 cycles, denaturation at 94^0C for 2 sec, annealing/extension at 67^0C for 4 min; for next 35 cycles, denaturation at 94^0C for 2 sec, annealing/extension at 62^0C for 4 min followed by a final extension step at 62^0C for 4 min. The PCR amplifications were performed using AccuTaq Long PCR DNA polymerase (Sigma, MO, USA). The PCR reactions were conducted in 50 µl volumes containing 50 ng of coffee genomic DNA, 10 mM of dNTP mix, 10 µM each of forward and reverse primer and 1x Reaction Buffer (400 mM Tricine-KOH, pH 9.2 at 25^0C, 150 mM KOAc, 35mM Mg $(OAc)_2$, 750 µg/ml bovine serum albumin).

2.2.9.2 Analysis of PCR product

After the reaction, 5 µl of the reaction was run on 1% agarose gel electrophoresis as described earlier. The size of the coffee NMT gene amplicon was checked in comparison with a 3kb DNA ladder (MBI Fermentas, Lithuania).

2.2.10 Blunt-end cloning of the full length *N*-Methyltransferase gene from *C. canephora*

The PCR product for full length NMT gene was purified using GenElute PCR Clean - Up Kit (Sigma, USA). The purified 2 kb product was blunt end ligated to pETBlue-2

blunt vector using Perfectly Blunt Cloning Kit (Novagen, USA) and following the manufacturer's instructions. End conversion reaction was performed as follows:

Purified PCR product:	2µl
Sterile water:	3µl
End conversion mix∗:	5µl
Total reaction:	10µl

∗ Provided with the kit.

The reaction was incubated at 22^0C for 15 min, and the kinase in the end conversion mix was heat inactivated at 75^0C for 5 min. To the cooled end conversion reaction, 1µl (50 ng) of pETBlue-2 blunt vector and 1µl (4U) of T4 DNA Ligase were added. The reaction was incubated at 22^0C for 15 min. 2.5 µl of the ligation mix was used for transforming DH5α strain of *E. coli* as described earlier. Several plasmids were isolated and checked for the presence of insert by comparing with control vector. The insert release from the clones was checked by restriction digestion with B*gl* II / P*vu* II. The positive recombinant plasmids were sequenced at sequencing facility at the Dept. of Biochemistry, University of Delhi, South Campus, New Delhi.

2.3 RESULTS AND DISCUSSION

2.3.1 PCR amplification and cloning of partial *N*-methyltransferase gene(s)

A fragment of *N*-methyltransferase (NMT) gene was PCR amplified from *Coffea canephora* var. S-274 using CCSF5 and CCSR2 primers as forward and reverse primers, respectively. The primers were designed based on the published sequence of *N*-methyltransferase gene of *Coffea arabica* (accession nos. AB048794, AB039725). The samples were run on 1.2 % agarose gel and the size of the amplicon was determined based on the 3.0 Kb DNA marker run along with the sample. The 1.8 kb size (Figure 2.3.1A) of genomic PCR fragment in comparison to expected 1.0 kb cDNA fragment suggested the presence of introns in the gene. The purified NMT gene amplicon was cloned using the pTZ57R/T vector in DH5α strain of *E.coli*. Recombinants were selected on the basis of blue/white selection on IPTG/X-Gal plates. The white colonies were taken for plasmid DNA isolation and restriction digestion was performed to check for insert release. Clones 2A and 2B were observed to be differing in restriction patterns when digested with the two sets of enzymes used for the release of insert (Figure 2.3.1B).

Insert release from clone 2A with E*co*RI/H*in*dIII enzymes resulted in the release of two fragments of 1.2 kb and 0.7 kb; whereas with clone 2B, 1.4 kb and 0.5 kb fragments were obtained. Similarly, E*co*RI/B*am*HI digestion of clone 2A resulted in three fragments of 0.5 kb, 0.6 kb and 0.75 kb in size, whereas the digestion of clone 2B resulted in a single fragment of 1.9 kb size. The restriction analyses clearly show that the two clones differ in restriction pattern and the two clones might belong to different NMT genes.

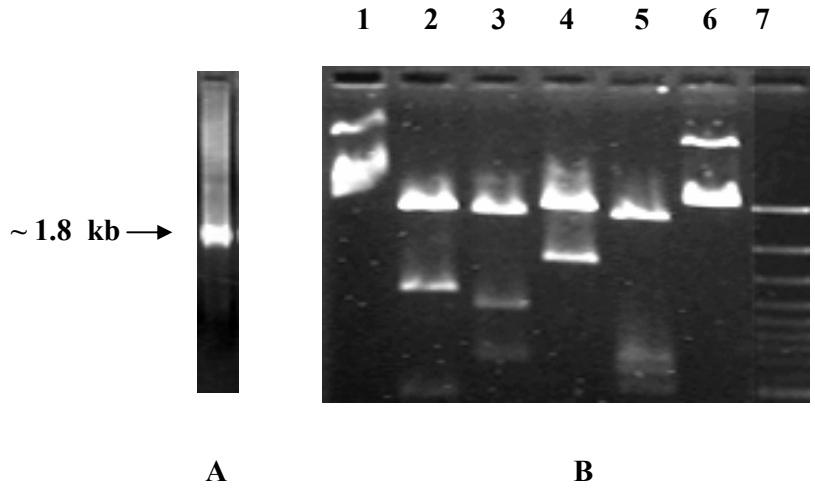


Figure 2.3.1 PCR amplification of coffee NMT gene and insert release from the clones

Figure 2.3.1A: Amplification of NMT gene fragment from *Coffea canephora*.

Figure 2.3.1B: Restriction analyses of the clones 2Aand 2B

Lane 1 Clone 2B undigested

Lane 2 Clone 2B - E*co*RI/H*in*dIII

Lane 3 Clone 2A - E*co*RI/H*in*dIII

Lane 4 Clone 2B - E*co*RI/B*am*HI

Lane 5 Clone 2A - E*co*RI/B*am*HI

Lane 6 Clone 2A undigested

Lane 7 3kb DNA marker

2.3.2 Nucleic acid sequence analysis of partial clones for NMT gene(s)

To see the possible differences existing within the clones 2A and 2B, both the clones were sequenced. Sequencing of Clones 2A and 2B revealed the size of the inserts i.e. amplified products to be 1729 bp and 1710 bp, respectively. The nucleotide sequences of the clones were analyzed using CLONE MANAGER version 5 software. The restriction maps of the clones 2A and 2B reveal differences in some of the restriction sites (Figure 2.3.2).

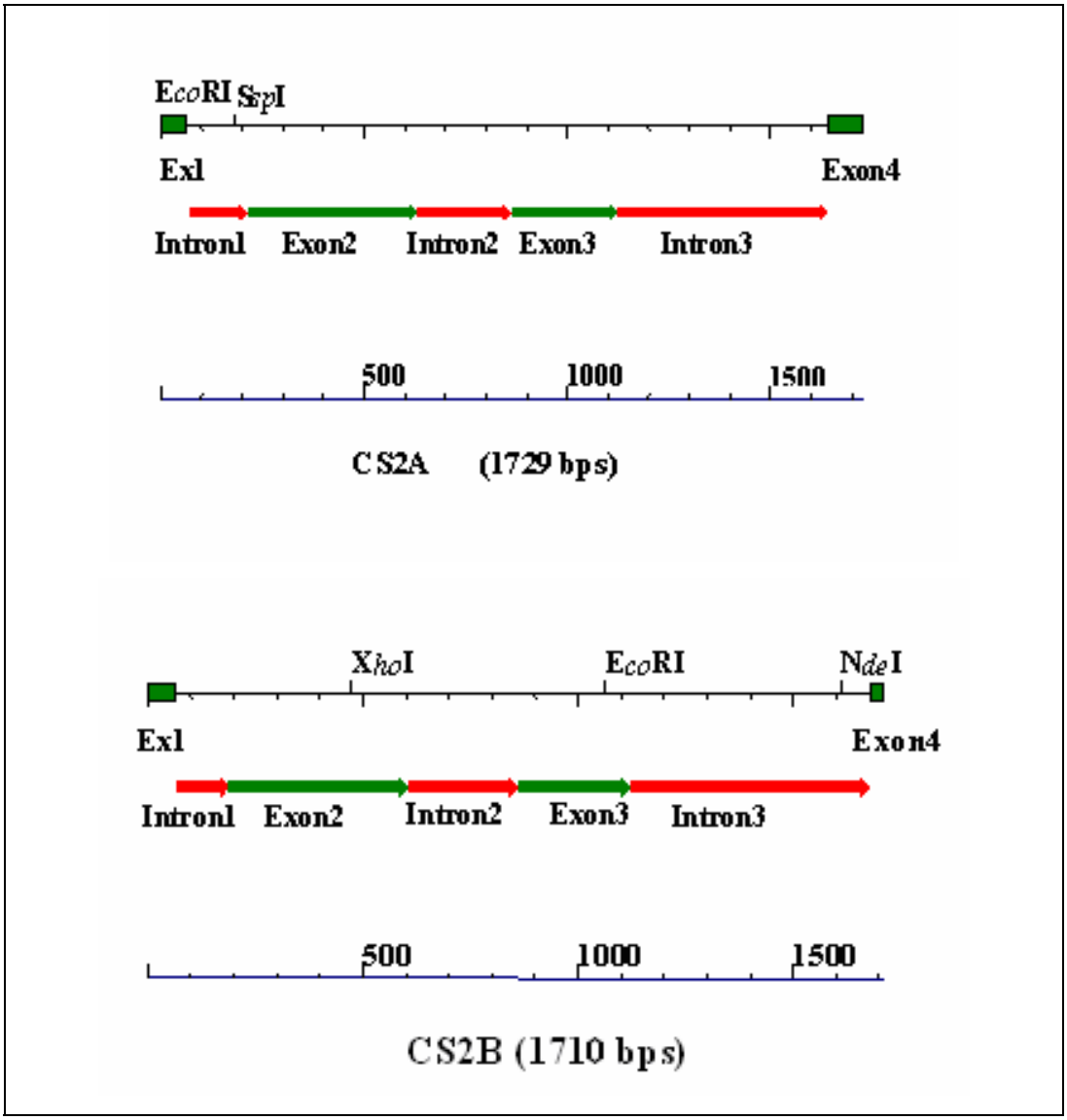


Figure 2.3.2. Restriction maps of clones 2A and 2B showing differences in the restriction sites

Sequence analysis by nucleotide BLAST (Altschul et al., 1997) revealed significant homology to NMT genes from several *Coffea* species. Highest scores for clone CS2A were observed with respect to NMT genomic clones from *Coffea liberica* var. *dewevrei* (accession no. AY362825), followed by *Coffea canephora* (AY273814) (Figure 2.3.3). This probably could be due to the presence of intron sequences. The other accessions that showed significant hits were cDNA sequences. However, there was drastic difference in the scores for these two genomic accessions with respect to clone CS2B, suggesting that clone CS2B is closely related to accession from *Coffea canephora* (Figure 2.3.3).

Sequences producing significant alignments for clone CS2A	Score (bits)	E Value
gi|37596300|gb|AY362825.1| Coffea liberica var. dewevrei puta...	**1096**	**0.0**
gi|33391745|gb|AY273814.1| Coffea canephora putative caffeine sy	**995**	**0.0**
gi|20271027|gb|AF494416.1| Coffea canephora clone 640.5 N-met...	759	0.0
gi|20271019|gb|AF494412.1| Coffea arabica clone 629.7 N-methy...	759	0.0
gi|26453372|dbj|AB034699.1|Coffea arabica CmXRS1 mRNA for 7-...	759	0.0
gi|13365750|dbj|AB048793.1|Coffea arabica CaXMT1 mRNA for xa...	759	0.0
gi|20271029|gb|AF494417.1| Coffea liberica clone 649.1 N-meth...	724	0.0
gi|20271021|gb|AF494413.1| Coffea arabica clone 629.10 N-meth...	724	0.0
gi|26453376|dbj|AB054841.1|Coffea arabica CTS2 mRNA for theo...	724	0.0
gi|20271023|gb|AF494414.1| Coffea canephora clone 640.2 N-met...	708	0.0
gi|20271025|gb|AF494415.1| Coffea canephora clone 640.3 N-met...	700	0.0
gi|20271017|gb|AF494411.1| Coffea arabica clone 629.2 N-methy...	700	0.0
gi|30023551|dbj|AB084126.1| Coffea arabica CaMXMT2 mRNA for 7...	700	0.0
gi|13365752|dbj|AB048794.1| Coffea arabica CaMXMT1 mRNA for 7...	700	0.0
gi|26453374|dbj|AB034700.1| Coffea arabica CTS1 mRNA for theo...	700	0.0
gi|30023549|dbj|AB084125.1| Coffea arabica CaDXMT1 mRNA for 3...	680	0.0
gi|26453392|dbj|AB086414.1| Coffea arabica CCS1 mRNA for caffein	680	0.0
gi|26453394|dbj|AB086415.1| Coffea arabica CtCS7 mRNA for ten...	672	0.0
gi|26453378|dbj|AB054842.1| Coffea arabica CtCS3 mRNA for ten...	593	7e-166
gi|13365748|dbj|AB048792.1| Coffea arabica CaMTL2 mRNA for th...	593	7e-166
gi|26453380|dbj|AB054843.1| Coffea arabica CtCS4 mRNA for ten...	579	1e-161
gi|13365693|dbj|AB039725.1| Coffea arabica CaMTL1 mRNA for caffe	571	2e-159

Sequences producing significant alignments for clone CS2B	Score (bits)	E Value
gi|33391745|gb|AY273814.1| Coffea canephora putative caffeine sy	**1945**	**0.0**
gi|37596300|gb|AY362825.1| Coffea liberica var. dewevrei puta...	**1229**	**0.0**
gi|20271017|gb|AF494411.1| Coffea arabica clone 629.2 N-methy...	815	0.0
gi|30023551|dbj|AB084126.1| Coffea arabica CaMXMT2 mRNA for 7...	815	0.0
gi|13365752|dbj|AB048794.1| Coffea arabica CaMXMT1 mRNA for 7...	815	0.0
gi|26453374|dbj|AB034700.1| Coffea arabica CTS1 mRNA for theo...	815	0.0
gi|20271023|gb|AF494414.1| Coffea canephora clone 640.2 N-met...	807	0.0
gi|20271025|gb|AF494415.1| Coffea canephora clone 640.3 N-met...	799	0.0
gi|20271021|gb|AF494413.1| Coffea arabica clone 629.10 N-meth...	751	0.0
gi|26453376|dbj|AB054841.1| Coffea arabica CTS2 mRNA for theo...	751	0.0
gi|20271029|gb|AF494417.1| Coffea liberica clone 649.1 N-meth...	728	0.0
gi|20271027|gb|AF494416.1| Coffea canephora clone 640.5 N-met...	680	0.0
gi|20271019|gb|AF494412.1| Coffea arabica clone 629.7 N-methy...	680	0.0
gi|26453372|dbj|AB034699.1| Coffea arabica CmXRS1 mRNA for 7-...	680	0.0
gi|13365750|dbj|AB048793.1| Coffea arabica CaXMT1 mRNA for xa...	680	0.0
gi|30023549|dbj|AB084125.1| Coffea arabica CaDXMT1 mRNA for 3...	668	0.0
gi|26453394|dbj|AB086415.1| Coffea arabica CtCS7 mRNA for ten...	660	0.0
gi|20271031|gb|AF494418.1| Coffea liberica clone 649.3 N-meth...	652	0.0
gi|37596302|gb|AY362826.1| Coffea canephora putative N-methyl...	636	5e-179
gi|20271033|gb|AF494419.1| Coffea liberica clone 649.9 N-meth...	636	5e-179
gi|26453392|dbj|AB086414.1| Coffea arabica CCS1 mRNA for caffein	628	1e-176
gi|26453378|dbj|AB054842.1| Coffea arabica CtCS3 mRNA for ten...	595	2e-166
gi|13365748|dbj|AB048792.1| Coffea arabica CaMTL2 mRNA for th...	595	2e-166
gi|26453380|dbj|AB054843.1| Coffea arabica CtCS4 mRNA for ten...	571	2e-159
gi|13365693|dbj|AB039725.1| Coffea arabica CaMTL1 mRNA for caffe	563	6e-157

Figure 2.3.3 Gene sequences showing significant homology to clones CS2A and CS2B

It was interesting to note that the two genomic accessions from *Coffea canephora* and *Coffea liberica* var. *dewevrei* encode for 378 and 384 amino acids, respectively corresponding to two theobromine synthase genes MXMT-1 and MXMT-2 (Uefuji et al., 2003). This would suggest that the clones CS2A and CS2B probably belong to the two different NMT genes, i.e. theobromine synthase genes.

Multiple sequence alignment with cDNA sequence for theobromine synthase revealed the presence of introns in coffee NMT genes (Figure 2.3.6). Though the two clones 2A and 2B shared high percentage of similarity in the exon regions, they differed in the sequences and lengths of introns (discussed in detail in chapter IV). The differences in the clones obtained might be due to amplification of multiple PCR products, which co-migrate during the agarose gel electrophoresis. Since the coffee NMT genes belong to a multigene family whose members share over 80% homology at the nucleotide level, multiple PCR products obtained may belong to different genes of the family or different alleles of the same gene.

2.3.3 PCR amplification and cloning of full length *N*-methyltransferase gene(s)

Since the members of coffee NMT gene family share high homology at the nucleotide as well as the deduced amino acid levels in the coding region, primers were designed based on the untranslated regions of the theobromine synthase gene (accession no. AB048794). The nucleotide sequence for only the gene encoding theobromine synthase was known in the literature at that time (Ogawa et al., 2001). The theobromine synthase is responsible for the second methylation step in the caffeine biosynthetic pathway.

Attempts to obtain the full-length amplicon for NMT gene by widely used three cycled PCR parameters proved futile. Adoption of two cycle parameters with drastic reduction in denaturation time resulted in successful amplification of longer DNA fragments. The two cycle parameters (section 2.2.9.1) were followed for most of the PCR experiments in this study. The extremely short incubation, i.e. for 2 sec, at 94^0C could have been helpful in preserving the integrity of the larger genomic DNA templates required for long PCR. The full-length gene for theobromine synthase was PCR amplified from genomic DNA of *Coffea canephora* var. S-274 using XMTF1 and XMTR1 primers. The PCR products were run on 1.0 % agarose gel and the size of the amplicon was determined based on the 3.0 kb DNA marker run along with the sample.

A PCR product of ~ 2.0 kb (Figure 2.3.4 A) was obtained and the purified amplicon for NMT gene was cloned in DH5α strain of *E.coli* using the pETBlue-2 blunt vector. Recombinants were selected on the basis of blue/white selection on IPTG/X-Gal plates. Plasmid DNA was isolated and the recombinants were selected after comparing with the control vector; and restriction digestion was performed with B*gl*II / P*vu*II to check for insert release (Figure 2.3.4 B and C).

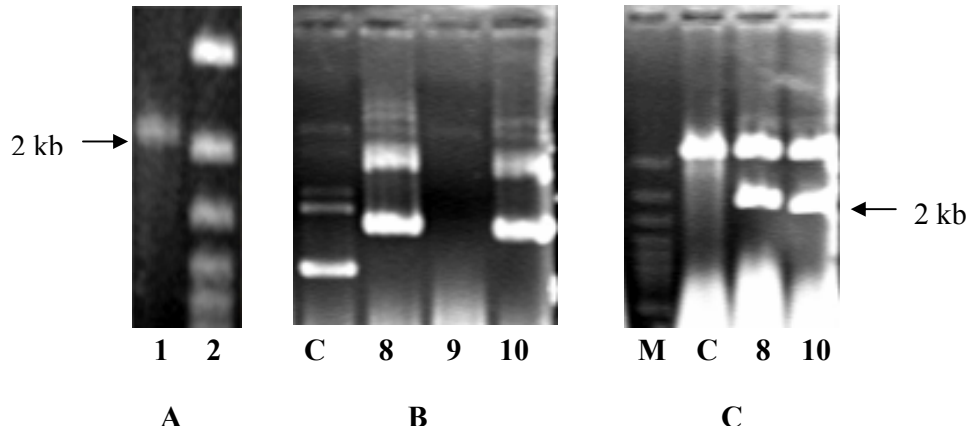


Figure 2.3.4 PCR amplification and cloning of full-length NMT gene from *Coffea canephora*

Figure 2.3.4 A. PCR product for full-length NMT gene
Figure 2.3.4 B. '8' and '10' are clones for full-length NMT gene; 'C' is control vector.
Figure 2.3.4 C. B*gl*II / P*vu*II double digestion of clones 8 and 10 for insert release, 'C' is linearised control vector.

2.3.4 Nucleic acid sequence analysis of full-length clones for NMT gene(s)

One of the recombinants i.e. clone CX8 was sequenced, but it was found to have one nucleotide deletion in fourth exon. Hence, another recombinant i.e. clone10 was also sequenced. The inset sizes in the clones CX8 and CX10 were 2003 and 1943 bps, respectively, suggesting that differences existed between them. Comparison of the sequences of these two clones with those available in the GenBank using nucleotide BLAST revealed significant homology to NMT genes from several *Coffea* species (Figure 2.3.5). The highest scores were observed with *C. canephora* genomic clone (378 AA) when compared to *C. liberica* var. *dewevrei* (384 AA) for both the genomic clones CX8 and CX8, indicating their close relationship.

Sequences producing significant alignments for clone CX8	Score (bits)	E Value
gi|33391745|gb|AY273814.1| Coffea canephora putative caffeine sy	**1937**	**0.0**
gi|37596300|gb|AY362825.1| Coffea liberica var. dewevrei puta...	**1261**	**0.0**
gi|13365752|dbj|AB048794.1| Coffea arabica CaMXMT1 mRNA for 7...	856	0.0
gi|26453374|dbj|AB034700.1| Coffea arabica CTS1 mRNA for theo...	856	0.0
gi|20271023|gb|AF494414.1| Coffea canephora clone 640.2 N-met...	823	0.0
gi|20271025|gb|AF494415.1| Coffea canephora clone 640.3 N-met...	815	0.0
gi|20271017|gb|AF494411.1| Coffea arabica clone 629.2 N-methy...	815	0.0
gi|30023551|dbj|AB084126.1| Coffea arabica CaMXMT2 mRNA for 7...	815	0.0
gi|20271021|gb|AF494413.1| Coffea arabica clone 629.10 N-meth...	783	0.0
gi|26453376|dbj|AB054841.1| Coffea arabica CTS2 mRNA for theo...	783	0.0
gi|20271029|gb|AF494417.1| Coffea liberica clone 649.1 N-meth...	759	0.0
gi|30023549|dbj|AB084125.1| Coffea arabica CaDXMT1 mRNA for 3...	700	0.0
gi|20271027|gb|AF494416.1| Coffea canephora clone 640.5 N-met...	696	0.0
gi|20271019|gb|AF494412.1| Coffea arabica clone 629.7 N-methy...	696	0.0
gi|26453372|dbj|AB034699.1| Coffea arabica CmXRS1 mRNA for 7-...	696	0.0
gi|13365750|dbj|AB048793.1| Coffea arabica CaXMT1 mRNA for xa...	696	0.0
gi|26453394|dbj|AB086415.1| Coffea arabica CtCS7 mRNA for ten...	692	0.0
gi|33355460|gb|AY273813.1| Coffea canephora putative caffeine sy	690	0.0
gi|20271035|gb|AF494420.1| Coffea liberica clone 649.6 N-meth...	684	0.0
gi|20271031|gb|AF494418.1| Coffea liberica clone 649.3 N-meth...	684	0.0
gi|37596302|gb|AY362826.1| Coffea canephora putative N-methyl...	668	0.0
gi|20271033|gb|AF494419.1| Coffea liberica clone 649.9 N-meth...	668	0.0
gi|26453392|dbj|AB086414.1| Coffea arabica CCS1 mRNA for caffein	660	0.0
gi|26453378|dbj|AB054842.1| Coffea arabica CtCS3 mRNA for ten...	595	2e-166
gi|13365748|dbj|AB048792.1| Coffea arabica CaMTL2 mRNA for th...	595	2e-166
gi|26453380|dbj|AB054843.1| Coffea arabica CtCS4 mRNA for ten...	587	5e-164
gi|13365693|dbj|AB039725.1| Coffea arabica CaMTL1 mRNA for caffe	579	1e-161

Sequences producing significant alignments for clone CX10	Score (bits)	E Value
gi|33391745|gb|AY273814.1| Coffea canephora putative caffeine sy	**1953**	**0.0**
gi|37596300|gb|AY362825.1| Coffea liberica var. dewevrei puta...	**1269**	**0.0**
gi|13365752|dbj|AB048794.1| Coffea arabica CaMXMT1 mRNA for 7...	856	0.0
gi|26453374|dbj|AB034700.1| Coffea arabica CTS1 mRNA for theo...	856	0.0
gi|20271023|gb|AF494414.1| Coffea canephora clone 640.2 N-met...	823	0.0
gi|20271025|gb|AF494415.1| Coffea canephora clone 640.3 N-met...	815	0.0
gi|20271017|gb|AF494411.1| Coffea arabica clone 629.2 N-methy...	815	0.0
gi|30023551|dbj|AB084126.1| Coffea arabica CaMXMT2 mRNA for 7...	815	0.0
gi|20271021|gb|AF494413.1| Coffea arabica clone 629.10 N-meth...	783	0.0
gi|26453376|dbj|AB054841.1| Coffea arabica CTS2 mRNA for theo...	783	0.0
gi|20271029|gb|AF494417.1| Coffea liberica clone 649.1 N-meth...	759	0.0
gi|30023549|dbj|AB084125.1| Coffea arabica CaDXMT1 mRNA for 3...	700	0.0
gi|20271027|gb|AF494416.1| Coffea canephora clone 640.5 N-met...	696	0.0
gi|20271019|gb|AF494412.1| Coffea arabica clone 629.7 N-methy...	696	0.0
gi|26453372|dbj|AB034699.1| Coffea arabica CmXRS1 mRNA for 7-...	696	0.0
gi|30023553|dbj|AB084127.1| Coffea arabica CaXMT2 mRNA for Xa...	696	0.0
gi|13365750|dbj|AB048793.1| Coffea arabica CaXMT1 mRNA for xa...	696	0.0
gi|26453394|dbj|AB086415.1| Coffea arabica CtCS7 mRNA for ten...	692	0.0
gi|33355460|gb|AY273813.1| Coffea canephora putative caffeine sy	690	0.0
gi|20271035|gb|AF494420.1| Coffea liberica clone 649.6 N-meth...	684	0.0
gi|20271031|gb|AF494418.1| Coffea liberica clone 649.3 N-meth...	684	0.0
gi|37596302|gb|AY362826.1| Coffea canephora putative N-methyl...	668	0.0

gi|20271033|gb|AF494419.1|	Coffea liberica clone 649.9 N-meth...	668	0.0
gi|26453392|dbj|AB086414.1|	Coffea arabica CCS1 mRNA for caffein	660	0.0
gi|26453378|dbj|AB054842.1|	Coffea arabica CtCS3 mRNA for ten...	595	2e-166
gi|13365748|dbj|AB048792.1|	Coffea arabica CaMTL2 mRNA for th...	595	2e-166
gi|26453380|dbj|AB054843.1|	Coffea arabica CtCS4 mRNA for ten...	587	5e-164
gi|13365693|dbj|AB039725.1|	Coffea arabica CaMTL1 mRNA for caffe	579	1e-161

Figure 2.3.5 Gene sequences showing significant homology to clones CX8 and CX10

The significant hits obtained in nucleotide BLAST search included mainly *N*-methyltransferases from different *Coffea* species. The results indicate high level of sequence conservation across different species of coffee namely, *C. arabica, C. canephora* and *C. liberica*. However, significant hits were not obtained for any other methyltransferases, indicating that coffee NMT genes are distinct from other SAM dependent methyltransferases.

Multiple alignments of the nucleotide sequences of four genomic clones for NMT genes with the cDNA sequence for one of the *Coffea arabica* theobromine synthase-1 (CaMXMT-1) revealed the presence of three introns and four exons in the NMT gene (Figure 2.3.6). However, the clone CX8 was observed to have one nucleotide deletion in the fourth exon at nucleotide position 1727, which results in stop codon in the predicted amino acid sequence. Since the PCR reactions were performed using proof reading AccuTaq Long PCR DNA polymerase (Sigma, USA), one nucleotide deletion in one of the clones could have been due to error in sequencing. To check this possibility, the sequencing was repeated. The repeated sequencing confirmed one base deletion in the fourth exon. Since the two clones displayed sequence variations in the length of the introns (Figure 2.3.6), it was assumed that these two clones represent two different genes of the family or two copies/ alleles of the same gene, and that one of the genes (clone CX8) has lost its function and represents a pseudogene. It is known that pseudogenization is one of the evolutionary fates of duplicated genes (Lawton-Rauh, 2003) and existence of pseudogenes in multigene families is a common feature (discussed in Chapter IV also).

```
CaMXMT1 AGCAGTCGCAATTCGATTGTCCTGCATATGAATGGAGCTCCAAGAAGTCCTGCATATGAA 60
CX10    AGCAGTCGCAATTCGATTGTCCTGCATATGAATGGAGCTCCAAGAAGTCCTGCATATGAA 60
CX8     AGCAGTCGCAATTCGATTGTCCTGCATATGAATGGAGCTCCAAGAAGTCCTGCATATGAA 60
CS2A    ----------------------------------AGCTCCAAGAAGTCCTGCATATGAA 25
CS2B    ----------------------------------AGCTCCAAGAAGTCCTGCATATGAA 25
                                           ************************

CaMXMT1 TGAAGGTGAAGGCGATACAAGCTACGCCAAGAATGCATCCTACAAT-------------- 106
CX10    TGAAGGTGAAGGCGATACAAGCTACGCCAAGAATGCATCCTACAATGTCTGTCTGTCTCT 120
CX8     TGAAGGTGAAGGCGATACAAGCTACGCCAAGAATGCATCCTACAATGTCTGTCTGTCTCT 120
CS2A    TGGAGGCGAAGGCGATACAAGCTACGCCAAGAATTCATCGTACAATGTCTGTCTGTCTCT 85
CS2B    TGAAGGTGAAGGCGATACAAGCTACGCCAAGAATGCATCCTACAATGTCTGTCTGTCTCT 85
        ** *** ************************** **** ******---Intron1--

CaMXMT1 ------------------------------------------------------------
CX10    CTATCTCTCTTTAACACACACACACA----GAGAATAGTGGTAAATCATACTATGATA 176
CX8     CTATCTCTCTTTAACACACACACACACACAGAGAATAGTGGTAAATCATGCCATGATA 180
CS2A    CCGTCTCTCTCTTTAACACACACACACACACAGAGTAGTAGTAAATCATGCTATGATA 145
CS2B    CTATCTCTCTTT----AACACACACACACACAGAGAATAGTGGTAAATCATGCTATGATA 141
        ----------------------------Intron1-------------------------

CaMXMT1 ------------------------------------------------------------
CX10    CGTCGATCTCTAACTTC----------------------A-CATTTGTATTTTGGACTGG 213
CX8     CGTCGATCTCTAACTTC----------------------A-CATTTGTACTTTGGACTGG 217
CS2A    CGTTGATCTCTGACTTAGTATGTCTTTTTTCCCACCTTAATATTTGTATTTTGGAGTGGT 205
CS2B    CGTCGATCTCTAACTTC----------------------ACATTTGTATTTTGGACTGGT 179
        ----------------------------Intron1-------------------------

CaMXMT1 ------------CTGGCTCTTGCCAAGGTGAAACCTTTCCTTGAACAATGCATACGAGAA 154
CX10    TATGTGTAACAGCTGGCTCTTGCCAAGGTGAAACCTTTCCTTGAACAATGCATACGAGAA 273
CX8     TATGTGTAACAGCTGGCTCTTGCCAAGGTGAAACCTTTCCTTGAACAATGCATACGAGAA 277
CS2A    TATGTGTAGCAGCTGGTTCTCACCAAGGTGAAACCTGTCCTTGAACAATGCATACGAGAA 265
CS2B    -ATGTGTAACAGCTGGCTCTTGCCAAGGTGAAACCTTTCCTTGAACAATGCATACGGGAA 238
                  **** ***  ************** ****************** ***
        -Intron1-

CaMXMT1 TTGTTGCGGGCCAACTTGCCCAACATCAACAAGTGCATTAAAGTTGCGGATTTGGGATGC 214
CX10    TTGTTGCGGGCCAACTTGCCCAACATCAACAAGTGCATTAAAGTTGCGGATTTGGGATGC 333
CX8     TTGTTGCGGGCCAACTTGCCCAACATCAACAAGTGCATTAAAGTTGCGGATTTGGGATGC 337
CS2A    TTGTTGCGGGCCAACTTGCCCAACATCAACAAGTGCATTAAAGTTGCGGATTTGGGATGC 325
CS2B    TTGTTGCGGGCCAACTTGCCCAACATCAACAAGTGCATTAAAGTTGCGGATTTGGGATGC 298
        ************************************************************

CaMXMT1 GCTTCTGGACCAAACACACTTTTAACAGTGCGGGACATTGTGCAAAGTATTGACAAAGTT 274
CX10    GCTTCTGGACCAAACACACTTTTAACAGTTCGGGACATTGTGCAAAGTATTGACAAAGTT 393
CX8     GCTTCTGGACCAAACACACTTTTAACAGTTCGGGACATTGTGCAAAGTATTGACAAAGTT 397
CS2A    GCTTCTGGACCAAACACACTTTTAACAGTTCGGGACATTGTGCAAAGTATTGACAAAGTT 385
CS2B    GCTTCTGGACCAAACACACTTTTAACAGTGCGGGACATTGTGCAGAGTATTGACAAAGTT 358
        ***************************** ************* ***************
```

```
CaMXMT1 GGCCAGGAAGAGAAGAATGAATTAGAACGTCCCACCATTCAGATTTTTCTGAATGATCTT 334
CX10    GGCCAGGAAGAGAAGAATGAATTAGAACGTCCCACCATTCAGATTTTTCTGAATGATCTT 453
CX8     GGCCAGGAAGAGAAGAATGAATTAGAACGTCCCACCATTCAGATTTTTCTGAATGATCTT 457
CS2A    GGCCAGGAAGAGAAGAATGAATTAGAACATCCCACCATTCAAATTTTTCTGAATGATCTT 445
CS2B    GGCCAGGAAGAGAAGAATGAATTAGAACGTCCCACCATTCAGATTTTTCTGAATGATCTT 418
        *************************** ************ ******************

CaMXMT1 TTCCAAAATGATTTCAATTCGGTTTTCAAGTTGCTGCCAAGCTTCTACCGCAAACTCGAG 394
CX10    TTCCAAAATGATTTCAATTCGGTTTTCAAGTTGCTGCCAAGCTTCTACCGCAAACTCGAG 513
CX8     TTCCAAAATGATTTCAATTCGGTTTTCAAGTTGCTGCCAAGCTTCTACCGCAAACTCGAG 517
CS2A    TTCCAAAATGATTTCAATTCGGTTTTCAAGTTGCTGCCAAGCTTCTACCGCAAACTTGAG 505
CS2B    TTCCAAAATGATTTCAATTCGGTTTTCAAGTTGCTGCCAAGCTTCTACCGCAAACTCGAG 478
        ********************************************************* ***

CaMXMT1 AAAGAAAATGGACGCAAGATAGGATCGTGCCTAATAAGCGCAATGCCTGGCTCTTTCTAC 454
CX10    AAAGAAAATGGACGCAAAATAGGATCGTGCCTAATAAGCGCAATGCCTGGCTCTTTCTAC 573
CX8     AAAGAAAATGGACGCAAAATAGGATCGTGCCTAATAAGCGCAATGCCTGGCTCTTTCTAC 577
CS2A    AAAGAAAATGGACGCAAAATAGGATCGTGCCTAATATGGGCAATGCCCGGCTCTTTCTAC 565
CS2B    AAAGAAAATGGACGCAAGATAGGATCGTGCCTAATAAGCGCAATGCCTGGCTCTTTCTAC 538
        ***************** ****************** * ******** ************

CaMXMT1 GGCAGACTCTTCCCCGAGGAGTCCATGCATTTTTTGCACTCTTGTTACAGTGTTCATTGG 514
CX10    GGCAGACTCTTCCCCGAGGAGTCCATGCATTTTTTGCACTCTTGTTACAGTGTTCATTGG 633
CX8     GGCAGACTCTTCCCCGAGGAGTCCATGCATTTTTTGCACTCTTGTTACAGTGTTCATTGG 637
CS2A    AGCAGACTCTTCCCCGAGGAGTCCATGCATTTTTTACACTCTTGTTACTGTCTTCAATGG 625
CS2B    GGCAGACTCTTCCCCGAGGAGTCCATGCATTTTTTGCACTCTTGTTACAGTGTTCATTGG 598
         ********************************** ************ ** **** ***

CaMXMT1 TTATCTCAG---------------------------------------------------- 523
CX10    TTATCTCAGGTCTTTGAGTTAATCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATG 693
CX8     TTATCTCAGGTCTTTGAGTTAATCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATG 697
CS2A    TTATCTCAGGTCTTTGAGTTAATCCCTTTTATCTTTTTAATTTTTCTTGTAGCAAAAATA 685
CS2B    TTATCTCAGGTCTTTGAGTTAATCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATG 658
        *********

                 ------------------Intron2-------------------------
CaMXMT1 ------------------------------------------------------------
CX10    GTTCATGATTTTCATTCAACACATTAGTAACTATGCATGGAAATTTCTTTAATAATTCTA 753
CX8     GTTCATGATTTTCATTCAACACATTAGTAACTATGCATGGAAATTTCTTTAATAATTCTA 757
CS2A    GTTCATGATTTTCATTCAACACATTGGTAACTATGCACGGAAATTTCTTTAGCAATTCTA 745
CS2B    GTTCGTGATTTTCATTCAACACATTAGTAACTATGCATGGAAATTTCTTTAATAATTCTA 718
        ----------------------------Intron2-------------------------

CaMXMT1 ------------------------------------------------------------
CX10    AAGATATCCACAGGAATCCAAGAAAGAGATTTCTGAAGAAACTAATAACATATTTTATCT 813
CX8     AAGATATCCACAGGAATCCAAGAAAGAGATTTCTGAAGAAACTAATAACATATTTTATCT 817
CS2A    AAGATATCCACAGGAATCCAAGAAAGAGATTTCTGAAGAAACTAATAACATATTT----- 800
CS2B    AAGATATCCACAGGAATCCAAGAAAGAGATTTCTGAAGAAACTAATAACATATTTTATTT 778
        ----------------------------Intron2-------------------------
```

```
CaMXMT1 ------------------------------------------------------------
CX10    AAGTCGTGGCTCATGATTTACATTCCCACATGCAACACTAACAAAATGATCCAACTATAT 873
CX8     AAGTCGTGGCTCATGATTTACATTCCCACATGCAACACTAACAAAATGATCCAACTATAT 877
CS2A    -------------------ATATTCCCACATGCAACACTAACAAAATGATCCAACTATAT 841
CS2B    AAGTCGTGGCTCATGATTTACATTCCCACATGCAACACTAACAAAATGATCCGACTATAT 838
        ----------------------------Intron2-------------------------

CaMXMT1 ------------------------GTTCCCAGCGGTTTGGTGATTGAATTGGGGATTGGT 559
CX10    AAGTTACCAGTTCTGGACGTGCAGGTTCCCAGCGGTTTGGTGATTGAATTGGGGATTGGT 933
CX8     AAGTTACCAGTTCTGGACGTGCAGGTTCCCAGCGGTTTGGTGATTGAATTGGGGATTGGT 937
CS2A    AAGTTACCAGTTCTAGACGTGCAGGTTCCCAGCGGATTGGTGACTGAACTGGGGATCAGT 901
CS2B    AAGTTACCAGTTCTAGACGTGCAGGTTCCCAGCGGTTTGGTGATTGAATTGGGGATTGGT 898
                               *********** ******* **** *******  **
        -----Intron2------------
CaMXMT1 GCAAACAAAGGGAGTATTTACTCTTCCAAAGGATGTCGTCCGCCCGTCCAGAAGGCATAT 619
CX10    GCAAACAAAGGGAGTATTTACTCTTCCAAAGGATGTCGTCCGCCCGTCCAGAAGGCATAT 993
CX8     GCAAACAAAGGGAGTATTTACTCTTCCAAAGGATGTCGTCCGCCCGTCCAGAAGGCATAT 997
CS2A    GCGAACAAAGGGATCATTTACTCTTCCAAAGCAAGTCCTCCGCCCGTCCAGAAGGCATAT 961
CS2B    GCAAACAAAGGGAGTATTTACTCTTCCAAAGGATGTCGTCCGCCCGTCCAAAAGGCATAT 958
        ** **********  *************** * *** ************ *********

CaMXMT1 TTGGATCAATTTACGAAAGATTTTACCACATTTCTAAGGATTCATTCGAAAGAGTTGTTT 679
CX10    TTGGATCAATTTACGAAAGATTTTACCACATTTCTAAGGATTCATTCGAAAGAGTTGTTT 1053
CX8     TTGGATCAATTTACGAAAGATTTTACCACATTTCTAAGGATTCATTCGAAAGAGTTGTTT 1057
CS2A    TTGGACCAATTTACAAAAGATTTTACCACATTTCTGAGGATTCATTCGGAAGAATTGCTT 1021
CS2B    TTGGATCAATTTACGAAAGATTTTACCACATTTCTAAGGATTCATTCGAAAGAGTTGTTT 1018
        ***** ******** ******************* ************ **** *** **

CaMXMT1 TCACGTGGCCGAATGCTCCTTACCTGCATTTGTAAAGTAGATGAATTCGACGAACCGAAT 739
CX10    TCACGTGGCCGAATGCTCCTTACCTGCATTTGTAAAGTAGATGAATTCGACGAACCGAAT 1113
CX8     TCACGTGGCCGAATGCTCCTTACTTGCATTTGGAAAGTAGATGAATTCGACGAACCGAAT 1117
CS2A    TCACGTGGCCGAATGCTCCTTACTTGCATTTGTAAAGGAGATGAATCCGATGGCCTGAAT 1081
CS2B    TCACGTGGCCGAATGCTCCTTACTTGCATTTGTAAAGTAGATGAATTCGACGAACCGAAT 1078
        ********************** ******** **** ******** *** *  * ****

CaMXMT1 CCCCTAGACTTACTTGACATGGCAATAAACGACTTGATTGTTGAG--------------- 784
CX10    CCCCTAGACTTACTTGACATGGCAATAAACGACTTGATTGTTGAGGTTATCATTTCTCTC 1173
CX8     CCCCTAGACTTACTTGACATGGCAATAAACGACTTGATTGCTGAGGTTATCATTTCTCTC 1177
CS2A    ACCATAGACTTACTTGAGAGAGCAATAAACGACTTGGTTGTTGAGGTTATCATTTCTCTG 1141
CS2B    CCCCTAGACTTACTTGACATGGCAATAAACGACTTGATTGTTGAGGTTATCATTTCTCTC 1138
         ** ************* *  *************** *** ****
                                                   ---Intron3--
CaMXMT1 ------------------------------------------------------------
CX10    TCTCTCTTTGATGATCAGATGTTCATTGCTCGTTATCTGAAATAAACTAGATAGCTACCT 1233
CX8     TCTCTCTTTGATGATCAGATGTTCATTGCTCGTTATCTGAAATAAACTAGATAGCTACCT 1237
CS2A    TCTCT--TTGATGATCAGATGCTCATTGCTTGTTATCTGAAATAAACTAGATAGCTAGCT 1199
CS2B    TCTCTCTTTGATGATCAGATGTTCATTGCTTGTTATCTGAAATAAACTAGATAGCTACCT 1198
        ----------------------------Intron3-------------------------
```

```
CaMXMT1 ------------------------------------------------------------
CX10    TAGCTTATTCAGGGTTCCTACCTTACCT------------------------------- 1261
CX8     TAGCTGATTCAGGGTTCCTACCTTACCT------------------------------- 1265
CS2A    -AGCTTATTCAGGGTTCCTACCTTAGCTGAAAAATGGATGCAATTTTATTCTTTGTATCC 1258
CS2B    TAGCTGATTCAGGGTTCCTACCTTACCT------------------------------- 1226
        ----------------------------Intron3-------------------------

CaMXMT1 ------------------------------------------------------------
CX10    -----------AACTTTGTGTAGACCAAGTTCACCGTAAATCAGATGGTAGTAACCTTT- 1309
CX8     -----------AACTTTGTGTAGACCAAGTTCACCGTAAATCAGATGGTAGTAACCTTTA 1314
CS2A    TTGATAGAATAAACTTTGTGTAGACGAAGTTCACCGTAACTCAGATGGCAGTAACCTTTA 1318
CS2B    -----------AACTTTGTGTAGACCAAGTTCACCGTAAATCAGATGGTAGTAACCTTTA 1275
        ----------------------------Intron3-------------------------

CaMXMT1 ------------------------------------------------------------
CX10    ------------------------------------------------------------
CX8     ATTAGAAGTTAGGTCACACAACTCCAATGCAGTTAAGATAATTCTTTC------------ 1362
CS2A    ATTAGGAGTTAGGTCACACAACTCCAATACATTTAATATAATTCTTTCCTTTCCAATAAT 1378
CS2B    ATTAGGAGTTAGGCGTCACAACTCCAATACAGTTAAGATAATTCTTTCCCTTCCAATAAT 1335
        ----------------------------Intron3-------------------------

CaMXMT1 ------------------------------------------------------------
CX10    ------------------------------------------------------------T 1310
CX8     -------AAGTTAGAAATAT---------------GTCGGCTATCTTTTTTTCCCCCTTT 1400
CS2A    TTGCTGGAAGTTAGAAATATAGGGGAATTAAATCTGTTGGCTATCATTTTTTCCCTTTTT 1438
CS2B    TTGTTGGAAGTTAGAAATATGGGGGGATTTAATTTGTCGGCTATCTTTTTTCCCCCCTTT 1395
        ----------------------------Intron3-------------------------

CaMXMT1 ------------------------------------------------------------
CX10    TTGGTTAAAAAGGTAC-----------------------------------TGGAAAAA 1334
CX8     TTGGTTAAAAAGGTAC-----------------------------------TGGAAAAA 1424
CS2A    TGGTT----------------------AAATAGTTCATTATTTAATCTTTTTTTGGCAAA 1476
CS2B    TTGGTTAAAAAGGTATTGGAAAGGAAGAAATAGTTCATTATTTAATCGTTTTT-GGCAAA 1454
        ----------------------------Intron3-------------------------

CaMXMT1 ------------------------------------------------------------
CX10    ATAATAGTATGGACTAAAAGCACAAGTAATAATAATATATTCATTTTGAAATTAAGG-AA 1393
CX8     ATAATAATATGGACTAAAAGCACAAGTAATAATAATATATTCATTTTGAAATTAAGGAAA 1484
CS2A    TTTTTGATATGAACTTGAAGCACAAGTAATAATAATATATTCATATTGAAATTAAGGAAA 1536
CS2B    TTTTTGATATGAACTAAAAGCCCAAGTAATAATCATGTATTCATATTGAAATTAAGGAAA 1514
        ----------------------------Intron3-------------------------

CaMXMT1 ------------------------------------------------------------
CX10    AAAAACTGTACCTTTTTTGTCCTAGAGTTGACGTTTTAGCAAATTGGCTAATTTCCAACT 1453
CX8     AAATACTGTACCTTTTTTGTCCTAGAGTTGACGTTTTAGCAAATTGGCTAATTTCCAACT 1544
CS2A    AAAAACGT----TTTTTTTTTCTAGAGTTGACTTTTTAGCAACTTGGCTAATTTCCAAGT 1592
CS2B    AAAGC---TGCACGTTTTATTCTAGAGTTGACGTTTTAGCAAATTGGCTGATTTCCAAGT 1571
        ----------------------------Intron3-------------------------
```

```
CaMXMT1 ------------------------------------------------------------
CX10    CTCGTGCTTCAAA--------------------------TAATTTCAATTGAATTGATG 1486
CX8     CTCGTGCTTCAAA---------------------------------------------- 1557
CS2A    CTTGTGCTTCAAATAATTTGAATTGAATTGATGATTGGCATA----------------- 1634
CS2B    CTTGTACGTCCAGTAATTCGAATTGACTTGATGATTGGCATATGAA-------------- 1617
        ----------------------------Intron3-------------------------

CaMXMT1 -------------------GGACTTCTGGAGGAAGAAAAATTGGATAGTTTCAATATTCC 825
CX10    ATGATTGGCATATCTGCAGGACTTCTGGAGGAAGAAAAATTGGATAGTTTCAATATTCC 1546
CX8     ----------TATCTGCAGGACTTCTGGAGGAAGAAAAATTGGATAGTTTCAATATTCC 1607
CS2A    ------------TCTTCAGGACTTCTGGAGGAAGAAAAATTGGATAGTTTCAATCTTCC 1682
CS2B    --------------TACAGGACGTCTGGGGGAAGAAAAATTGGACAGTTTCAATGTTCC 1663
                           **** ***** ************** ********* ****

         ---Intron3-----

CaMXMT1 ATTCTTTACACCTTCAGCAGAAGAAGTAAAGTGCATAGTTGAGGAGGAAGGTTCTTGCGA 885
CX10    ATTCTCTACACCTTCAGCAGAAGAAGTAAAGTGCATAGTTGAGGAGGAAGGTTCTTGCGA 1606
CX8     ATTCTTTACACCTTCAGCAGAAGAAGTAAAGTGCATAGTTGAGGAGGAAGGTTCTTGCGA 1667
CS2A    ACTCTATACACCTTCACTAGAAGTAGTAAAGTGCATGGTTGAGGAGG               1729
CS2B    AATCTATACAGCTTCAGTAGAAGAAGTAAAGTGCATGGTTGAGGAGG               1710
        * *** **** *****  ***** ************ ***********************

CaMXMT1 AATTTTATATCTGGAGACTTTTAAGGCCCATTATGATGCTGCCTTCTCTATTGATGATGA 945
CX10    AATTTTATATCTGGAGACTTTTAAGGCCCATTATGATGCTGCCTTCTCTGTTGATGATGA 1666
CX8     AATTTTATATCTGGAGACTTTTAAGGCCCATTATGATGCTGCCTTCTCTATTGATGATGA 1727
        ************************************************ **********

CaMXMT1 TTACCCAGTAAGATCCCATGAACAAATTAAAGCAGAGTATGTGGCATCATTAATTAGATC 1005
CX10    TTACCCAGTAAGATCCCATGAACAAATTAAAGCAGAGTATGTGGCATCATTAATTAGATC 1726
CX8     TTACC-AGTAAGATCCCATGAACAAATTAAAGCAGAGTATGTGGCATCATTAATTAGATC 1786
        ***** ******************************************************

CaMXMT1 AGTTTACGAACCCATCCTCGCAAGTCATTTTGGAGAAGCTATTATGCCTGACTTATTCCA 1065
CX10    AGTTTACGAACCCATCCTCGCAAGTCATTTTGGAGAAGCTATTATGCCTGACTTATTCCA 1786
CX8     AGTTTACGAACCCATCCTCGCAAGTCATTTTGGAGAAGCTATTATGCCTGACTTATTCCA 1846
        ************************************************************

CaMXMT1 CAGGCTTGCGAAGCATGCAGCAAAGGTTCTCCACATGGGCAAAGGCTGCTATAATAATCT 1125
CX10    CAGGCTTGCGAAGCATGCAGCAAAGGTTCTCCACATGGGCAAAGGCTGCTATAATAATCT 1846
CX8     CAGGCTTGCGAAGCATGCAGCAAAGGTTCTCCACATGGGCAAAGGCTGCTATAATAATCT 1906
        ************************************************************

CaMXMT1 TATCATTTCTCTCGCCAAAAAGCCAGAGAAGTCAGACGTGTAAAAGTTTGTTTTTAGTTG 1185
CX10    TATCATTTCTCTCGCCAAAAAGCCAGAGAAGTCAGACGTGTAAAAGTTTGTTTTTAGTTG 1906
CX8     TATCATTTCTCTCGCCAAAAAGCCAGAGAAGTCAGACGTGTAAAAGTTTGTTTTTAGTTG 1966
        ************************************************************
```

```
CaMXMT1 GTTTTTGTGCCGTTGGGGGTCTTTCGGGTATTGTCGTTTTGTATTCGTAATAAAAGTGAT 1245
CX10    GTTTTTGTGCCGTTGGGGGTCTTTCGGGTATTGTCGT                        1943
CX8     GTTTTTGTGCCGTTGGGGGTCTTTCGGGTATTGTCGT                        2003
        ************************************

CaMXMT1 GTGCAAGAATAAGATATTTAGTACAATATTTTCATAAAAAAAAAAAAAAAAAA        1298
CX10    ----------------------------------------------------        1943
CX8     ----------------------------------------------------        2003
```

Figure 2.3.6 Multiple alignments of the nucleotide sequences of the *Coffea canephora* NMT genomic clones using Dialign 2 (Morgenstern, 1999)

Note: CaMXMT1 is the cDNA sequence for theobromine synthase (accession no. AB048794). The start codon and consensus intron/exon borders of genes are indicated in bold. The AUG triplet upstream to start codon is underlined.

Splice sites were inferred through direct comparisons of genomic sequences with the orthologous coffee cDNA from which the original PCR amplification primers were designed. The sequence comparisons of clones CX8 and CX10 revealed that the two clones were indeed different, as there were differences in the size and sequences of introns, as in case of clones CS2A and CS2B. The comparative sizes of the introns in the four genomic clones for NMT genes obtained in this study are presented in Table 2.3.1.

Table 2.3.1 Variations in the intron size in genomic clones for NMT genes

Clones	Intron 1 (bp)	Intron 2 (bp)	Intron 3 (bp)
CX10	119	255	347
CX8	123	255	404
CS2A	146	231	515
CS2B	119	255	499

The coding region of clone CX10 comprises 1134 base pairs and encodes 378 amino acids. Based on the insertions and deletions present in the cDNA sequences of various NMT genes reported so far (Figure 2.3.7), it could be assumed that the genomic clone corresponds to CaMXMT-1 (378 amino acids) and possibly encodes a theobromine synthase-1. The coffee theobromine synthase-1 gene is interrupted by three introns containing A+T rich sequences (Table 2.3.2), whose junctions agree with the consensus intron/exon borders of plant genes.

Table 2.3.2 Size and position of introns present in the clone CX10, encoding for theobromine synthase -1 gene

S.No	Size (bp)	Position from start codon (bp)	5´ and 3´ intron/ exon border	AT%
Intron 1	119	76-194	GT…..AG	62.2%
Intron 2	255	612-866	GT…..AG	68.6%
Intron 3	347	1128-1474	GT…..AG	67.4%

The cloned coffee theobromine synthase-1 gene has an AUG immediately upstream to the start codon (AUGA**AUG**) (Figure 2.3.6). The preceding AUG is in an inefficient context probably resulting in very poor recognition and ribosome proceeding to next AUG for translation initiation. However, the context of start AUG for coffee NMT gene is taUGa**AUG**GAg and does not agree fully with the consensus context of caA(A/C)a**AUG**GCg for higher plants/dicots. It is believed that AUG context effects might play a less significant role in plants than in animals.

The sequence comparisons of four genomic clones CX10, CX8, CS2A and CS2B indicate that the coding regions share high % of similarity but differences exist mainly in the size and sequence of introns (discussed in detail in chapter IV). This suggests that the four clones belong to different alleles/genes of the NMT family. Evidences obtained earlier also suggest existence of multiple enzymes or isoforms for *N*-methyltransferases in coffee plants, as indicated from the presence of many closely related genes and from the expression pattern of the identified genes. Two genes for theobromine synthase were reported by two independent groups (Mizuno et al., 2001; Uefuji et al., 2003). CaDXMT1 gene is almost exclusively expressed in immature fruits, whereas CaXMT1, CaMXMT1 and CaMXMT2 are found in leaves, floral buds and immature fruits (Uefuji et al., 2003). Because young leaves contain a high level of caffeine and DXMT activity, another form of DXMT may be present in young leaves. Thus it can be said that coffee plants are equipped with multiple sets of enzymes, which may be necessary for constitutive production of caffeine in relevant tissues.

```
CaXMT1    ------MELQEVLRMNGGEGDTSYAKNSAYNQLVLAKVKPVLEQCVRELLRANLPNINKC 54
CaMXMT1   ------MELQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKC 53
CaMXMT2   ------MELQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKC 53
CaDXMT1   ------MELQEVLHMNGGEGDTSYAKNSFYN-LFLIRVKPILEQCIQELLRANLPNINKC 53
CX10      ------MELQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKC 53
CS2A      --------LQEVLHMNGGEGDTSYAKNSSYN-LVLTKVKPVLEQCIRELLRANLPNINKC 51
CS2B      --------LQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKC 51
TCS1      MELATAGKVNEVLFMNRGEGESSYAQNSSFTQQVASMAQPALENAVETLFSRDFH--LQA 58

CaXMT1    IKVADLGCASGPNTLLTVRDIVQSIDKVGQEKKNELERPTIQIFLNDLFPNDFNSVFKLL 114
CaMXMT1   IKVADLGCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLL 113
CaMXMT2   IKVADLGCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLL 113
CaDXMT1   IKVADLGCASGPNTLLTVRDIVQSIDKVGQEKKNELERPTIQIFLNDLFQNDFNSVFKSL 113
CX10      IKVADLGCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLL 113
CS2A      IKVADLGCASGPNTLLTVRDIVQSIDKVGQEEKNELEHPTIQIFLNDLFQNDFNSVFKLL 111
CS2B      IKVADLGCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLL 111
TCS1      LNAADLGCAAGPNTFAVISTIKRMMEKKCRELN--CQTLELQVYLNDLFGNDFNTLFKGL 116

CaXMT1    PSFYRKLEKENGRKIGSCLIGAMPGSFYSRLFPEESMHFLHSCYCLQWLSQVPSGLVTEL 174
CaMXMT1   PSFYRKLEKENGRKIGSCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIEL 173
CaMXMT2   PSFYRKLEKENGRKIGSCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIEL 173
CaDXMT1   PSFYRKLEKENGRKIGSCLIGAMPGSFYGRLFPEESMHFLHSCYCLHWLSQVPSGLVTEL 173
CX10      PSFYRKLEKENGRKIGSCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIEL 173
CS2A      PSFYRKLEKENGRKIGSCLIWAMPGSFYSRLFPEESMHFLHSCYCLQWLSQVPSGLVTEL 171
CS2B      PSFYRKLEKENGRKIGSCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIEL 171
TCS1      SSEVIGNKCEEVP----CYVMGVPGSFHGRLFPRNSLHLVHSSYSVHWLTQAPKGLTSRE 172

CaXMT1    GISTNKGSIYSSKASRLPVQKAYLDQFTKDFTTFLRIHSEELFSHGRMLLTCICK----G 230
CaMXMT1   GIGANKGSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK----V 229
CaMXMT2   GIGANKGSIYSSKASRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK----V 229
CaDXMT1   GISANKGCIYSSKASRPPIQKAYLDQFTKDFTTFLRIHSEELISRGRMLLTWICK----E 229
CX10      GIGANKGSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK----V 229
CS2A      GISANKGIIYSSKASPPPVQKAYLDQFTKDFTTFLRIHSEELLSRGRMLLTCICK----G 227
CS2B      GIGANKGSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK----V 227
TCS1      GLALNKGKIYISKTSPPVVREAYLSQFHEDFTMFLNARSQEVVPNGCMVLILRGRQCSDP 232

CaXMT1    VELDARNAIDLLEMAINDLVVEGHLEEEKLDSFNLPVYIPSAEEVKCIVEEEGSFEILYL 290
CaMXMT1   DEFDEPNPLDLLDMAINDLIVEGLLEEEKLDSFNIPFFTPSAEEVKCIVEEEGSCEILYL 289
CaMXMT2   DEYDEPNPLDLLDMAINDLIVEGHLEEEKLASFNLPFFTPSAEEVKCIVEEEGSFEILYL 289
CaDXMT1   DEFENPNSIDLLEMSINDLVIEGHLEEEKLDSFNVPIYAPSTEEVKCIVEEEGSFEILYL 289
CX10      DEFDEPNPLDLLDMAINDLIVEGLLEEEKLDSFNIPFSTPSAEEVKCIVEEEGSCEILYL 289
CS2A      DESDGLNTIDLLERAINDLVVEGLLEEEKLDSFNLPLYTPSLEVVKCMVEE--------- 278
CS2B      DEFDEPNPLDLLDMAINDLIVEGRLGEEKLDSFNVPIYTASVEEVKCMVEE--------- 278
TCS1      SDMQSCFTWELLAMAIAELVSQGLIDEDKLDTFNIPSYFASLEEVKDIVERDGSFTIDHI 292

CaXMT1    ETFKVLYDAGFSIDD-------------EHIKAEYVASSVRAVYEPILASHFGEAIIPDI 337
CaMXMT1   ETFKAHYDAAFSIDDDYPVRSH------EQIKAEYVASLIRSVYEPILASHFGEAIMPDL 343
CaMXMT2   ETFKAHYDAGFSIDDDYPVRSHFQVYGDEHIKAEYVASLIRSVYEPILASHFGEAIMPDL 349
CaDXMT1   ETFKVPYDAGFSIDDDYQGRSHSPVSCDEHARAAHVASVVRSIFEPIVASHFGEAIMPDL 349
CX10      ETFKAHYDAAFSVDDDYPVRSH------EQIKAEYVASLIRSVYEPILASHFGEAIMPDL 343
TCS1      EGFDLDSVEMQENDK-------------WVRGEKFTKVVRAFTEPIISNQFGPEIMDKL 338

CaXMT1    FHRFAKHAAKVLPLGKGFYNNLIISLAKKPEKSDV 372
CaMXMT1   FHRLAKHAAKVLHMGKGCYNNLIISLAKKPEKSDV 378
CaMXMT2   FHRLAKHAAKVLHLGKGCYNNLIISLAKKPEKSDV 384
CaDXMT1   SHRIAKNAAKVLRSGKGFYDSLIISLAKKPEKSDV 384
CX10      FHRLAKHAAKVLHMGKGCYNNLIISLAKKPEKSDV 378
TCS1      YDKFTHIVVSDLEAKLPKTTSIILVLSKIDG---- 369
```

Figure 2.3.7 Multiple sequence alignments of the predicted amino acids for the genomic clones using Clustal W (Thompson et al., 1994)

The multiple sequence alignments of the predicted amino acid sequences for the genomic clones obtained in this study were performed using Clustal W (Thompson et al., 1994). The amino acid sequence alignments also revealed a high degree of similarity amongst them (Figure 2.3.7). Clone CX8 was not used for alignment as it contained one nucleotide deletion in the fourth exon, which results in pre- mature termination of the predicted amino acid sequence.

2.3.5 Summary and Conclusions

Using PCR based genomic cloning approach, partial and full-length genomic clones for NMT gene from *C. canephora* were obtained. The position of the introns was ascertained by comparison with cDNA sequence for theobromine synthase-1 gene. The coding region of clone CX10 comprises 1134 base pairs and encodes 378 amino acids. Based on the insertions and deletions present in the cDNA sequences of various NMT genes reported so far, it was concluded that the genomic clone corresponds to CaMXMT-1 (378 amino acids) and possibly encodes a theobromine synthase-1. The coffee theobromine synthase-1 gene comprises of four exons interrupted by three introns. The sequence comparisons of four genomic clones CX10, CX8, CS2A and CS2B indicate that the coding regions share high % of similarity but differences exist mainly in the size and sequence of introns. This suggests that the four clones belong to different alleles/genes of the NMT family.

The strategy of genomic cloning of NMT genes adopted in this study has proved advantageous in isolation of different clones that differed in their intron sequences. Use of cDNA approach would have resulted in ignoring the important unit of gene structure i.e. introns. It is well known that when two genes are related, the relationship between their exons is closer than the relationship between the introns. In extreme cases, the exons of two genes may code for the same protein sequence, but the introns may be different (Lewin, 2004). This situation was also observed for coffee NMT genes cloned in this study, wherein the different clones had identical exonic sequences, but the introns differed in sequence and more strikingly in their lengths. In introns, the pattern of divergence involves both changes in size (due to deletions and insertions) and base substitutions. It is known that introns evolve much more rapidly than exons. Introns may participate in the creation of divergence between paralogs, similarly to promoters

or other regulatory elements (Lynch and Conery, 2000) and the possibility of this in the evolution of coffee NMT gene family needs to be explored.

In the subsequent investigation, experiments were undertaken to isolate and characterize the promoter for an NMT gene from coffee. Attempts were also made to isolate NMT genomic clones containing the promoter and the corresponding gene.

Chapter III

Cloning of Promoter for an N-Methyltransferase gene

3.0 ABSTRACT

In this study, a PCR based genome walking method was adopted to isolate and clone the promoter for one of the coffee NMT genes. Inspection of the promoter sequence revealed the presence of several motifs important for the regulation of the gene expression. The isolated promoter fragment was fused to the β-glucuronidase (GUS) reporter gene and used in *Agrobacterium tumefaciens* mediated transformation of *Nicotiana tabacum* and also for electroporation of coffee endosperms. GUS assays proved that the isolated promoter was able to direct the expression of the reporter gene in transgenic tobacco. The cellular localization of GUS activity in transgenic tobacco confirms that coffee NMT gene is cytosolic enzyme. Transient assays demonstrated the ability of the isolated promoter fragment to drive gene expression in endosperms. The cloning of promoter for a gene involved in caffeine biosynthetic pathway opens up the possibility of studying the molecular mechanisms that regulate the production of caffeine. Based on the promoter sequence, primer was designed and the genomic fragment comprising the promoter and its corresponding gene was amplified and cloned. Sequencing of one of the genomic clones revealed the presence of four exons and three introns in NMT gene. Sequence comparisons were made between coffee NMT gene and *Arabidopsis* Jasmonic acid carboxyl methyltransferase, which belongs to coffee *N*-methyltransferase related family. Though the gene structure was highly conserved in terms of number and position of introns, AtJMT revealed great variation for coding regions, introns, and the promoter region sequences, suggesting that the two genes have diverged greatly after their origin from a common ancestor.

3.1 INTRODUCTION

Eukaryotes employ diverse mechanisms to regulate gene expression, including chromatin condensation, DNA methylation, transcriptional initiation, alternative splicing of RNA, mRNA stability, translational controls, several forms of post-translational modification, intracellular trafficking, and protein degradation (Lewin, 2004). A major level at which gene expression is regulated is the initiation of transcription, and this is reflected in the percentage of the genome dedicated to transcription factors in plants and other eukaryotes (Singh, 1998). Approximately, 5% of the genes of Arabidopsis encode for transcription-related proteins comprising more than 1500 transcriptional factors, 45% of which are from families specific to plants (Riechmann et al., 2000).

Only some of the genes in a eukaryotic cell are expressed at any given moment. The proportion and composition of transcribed genes changes considerably during the life cycle, among cell types and in response to fluctuating physiological and environmental conditions. Plants, by and large, also seem to follow typical eukaryotic norms for gene expression and regulation. It is estimated that plants may contain approximately 25,000 to 50,000 genes. One third of these genes are estimated to express in all the organs, though at various levels. Another one-third may have their products present in a few, but not all, organs and the rest may belong to unique organ as far as expression is concerned. This variability is reflected in the organization of promoters and regulatory elements as well as in the genes characterized for the regulating promoters (Tyagi, 2001).

Eukaryotic genomes are complex and organized within compact nucleoprotein (chromatin) structures. The DNA of eukaryotes is generally refractory to transcription because of its organisation in nucleosomes (basic unit of chromatin organisation), wherein approximately 150 bp of DNA is wrapped around a histone octamer. Nucleosomes prevent the initiation of transcription by blocking access of RNA polymerase II (Pol II) apparatus to promoter DNA. Thus nucleosomes serve as general gene repressors and they have to be specifically removed for activation of transcription, which is a prerequisite for assembly of

the giant Pol II transcription initiation complex (TIC)/ preinitiation complex (PIC). Extensive post-translational modifications of the histones (acetylation, methylation, phosphorylation and ubiquitylation) determine whether particular nucleosomes remain bound and hence repressive or else whether they are destined for removal on the pathway of gene activation (Klug, 2005). Chromatin-remodeling complexes such as SWI/SNF and RSC complex accomplish the removal of nucleosomes in an ATP-dependent manner. Due to the loss of nucleosomes from transcriptionally active promoters, the RNA polymerase II TIC is assembled on a naked DNA molecule *in vivo*.

Promoter Structure

The organization of promoters is much less regular than that of coding sequences and lacks an equivalent of the genetic code or other sequence features that provide a consistent relationship to function (Table 3.1.1). The organisation of plant promoters follows the general structure common for other eukaryotes: a core promoter adjacent to the transcriptional start site, followed by an upstream region of about 1 kb, containing the proximal and upstream *cis*-acting elements and the outside enhancer like sequences. Although promoters lack universal structural features, these functional features are always present. Plant gene promoters are compact, in which less than 1000 bp of promoter sequence is sufficient to drive proper regulated patterns of transcription and most regulatory elements tend to be clustered near the transcriptional start site (Guo and Moose, 2003).

(i) Basal promoter/ Core Promoter:

Core promoter elements are defined as minimal DNA elements that are necessary and sufficient for accurate initiation of transcription by RNA polymerase II in a reconstituted cell-free system (Komarnytsky and Borisjuk, 2003). The transcriptional machinery assembles on the basal promoter, ~ 100 bp region whose function is to provide a docking site for the transcription complex and to position the start of the transcription relative to coding sequences. Basal promoter sequences vary among genes. For many genes, the critical binding site is TATA box, usually located 25-30 bp 5' of the transcription start site (TSS). However, many eukaryotic genes lack a TATA box and instead contain an initiator element spanning the TSS.

Table 3.1.1. Structural and functional differences in coding and promoter sequences for a protein-coding locus (Adapted from Wray et al., 2003)

	Coding	Promoter
Physical boundaries	*Defined by sequence*	*Not defined by sequence*
Start	ATG	None
End	TAA, TAG, or TGA	None
Internal[1]	[C/A]AG|GU[A/G]AGU $(U/C)_n$ NAG| G/A	None
Physical organization	*Discontinuous, colinear*	*Discontinuous, nonlinear*
Physical unit	Exon	Module[2]
Typical unit size	~20-2000 bp	~200-2000 bp
Number of units	1-10, rarely more	1-10, rarely more
Organization and function colinear	Yes	No
Relative order of units	Consistent	Not consistent
Modules correspond to functions	Sometimes	Often
Functional Organization	*Direct, local*	*Indirect, distributed*
Direct functional output	Protein sequence	Transcriptional profile
Unit of information	Codon	Binding site
Number of units	~150-1000, rarely more	~6-60
Information content	0.3-2.0 kb	0.08-0.8 kb[3]
Spacing between units	Doesn't matter	Sometimes matters
Mapping	Precise (1 amino acid)	Imprecise (>1Transcription factor)
Degeneracy	Precise (same AA)	Imprecise (different transcription factor)
Consequence	Qualitative (1 codon:1 AA)	Quantitative (Level of transcription)
Order of units	Usually matters	Sometimes matters
Genetic basis	*Cis* only	*Cis* and *trans* required

[1]Type I introns; other splice junction sequences also exist
[2] Cluster of transcription factor binding sites
[3]*cis*-regulatory sequences only,

Knowledge of TATA-less promoters in plants is limited, but it was recently found that the majority of the TATA-less promoters were common to genes associated with photosynthesis (Nakamura et al., 2002). The functional consequences of differences in basal promoter structure are not well understood, although genes with TATA-less basal promoters may generally be transcribed constitutively at relatively low levels. A key early step in the transcriptional initiation is attachment of TATA-binding protein (TBP) to DNA. In promoters lacking TATA boxes, proteins that associate with other basal promoter motifs facilitate TBP-association with DNA in a sequence-independent manner. Once TBP binds, several TBP-associated factors (TAFs) guide the RNA polymerase II holoenzyme complex onto the DNA (Figure 3.1.1). This step, which can be positively or negatively modulated by transcription factors bound at other sites, is one of the most important points of transcriptional regulation (Lemon and Tjian, 2000).

Two more proximal regulatory elements, not universally required for the activity of the core eukaryotic promoters, when present significantly affect their initiation frequency. These sequences are located close to TSS (around –100 bp) and appear to assist in the recruitment of basic transcription factors to the TSS sites of housekeeping and TATA-less genes (Komarnytsky and Borisjuk, 2003). The GC-box, GGGCGG plays an important role in the core promoter regions of mammalian genes. Another well-conserved element, the CCAAT-box, is often present at –80 to –150 bp upstream of TSS, and may operate cooperatively with other putative conserved motifs. While many plant core promoters share combinations of well-defined elements such as TATA-box, initiator and CCAAT-box, their consensus sequences and spacing may vary significantly, such that no two core promoters are identical. Therefore, the composition of the TIC that binds to the core promoter is variable, and this offers additional opportunities for the regulation of basic transcription.

The TSS, unlike the start site of translation, varies in both sequence and position. Spacing between the start sites of transcription and translation differs considerably among genes, ranging from ~10^1 to 10^4 bp; the 5' untranslated region (UTR) can also contain introns that alter its length post-transcriptionally. The promoter sequence is usually located

upstream from the TSS, but regulatory elements may also be located downstream, for example in the intron of the gene itself, or at the 3' end of the gene.

By itself, a basal promoter initiates transcription at a very low rate, even when the local chromatin is suitably decondensed. Furthermore, most of the proteins that bind to basal promoter motifs are ubiquitously expressed and therefore provide little regulatory specificity (Lemon and Tjian, 2000). These proteins are known as general transcription factors. The major level of transcriptional control is mediated by *trans*-acting factors binding to the upstream regulatory elements. At any given time, a distinct set of transcription factors is available to form the higher-order nucleoprotein complexes at the active promoter. The composition of individual trans-acting factors within these complexes may often change when a specific signal is perceived, so that a given transcription factor can play multiple roles, and affect multiple gene sets, depending on its local concentration and availability (Komarnytsky and Borisjuk, 2003).

(ii) Transcriptional Factor Binding Sites/ upstream cis-acting elements:

The transcriptional factors bind to DNA on specific cis-acting regulatory elements or transcription factor-binding sites. The composition and organization of these transcription factor-binding sites varies enormously among genes. The changing array of transcription factors provides nearly all of the control over when, where, at what level, and under what circumstances a particular gene is transcribed. Thus, the genetic basis for the expression profile of each gene resides in part within its promoter and in part within the many other segments of the genome that encode specific transcription factors that bind to the promoter.

The salient features of transcription factor-binding sites are as follows:

(i) Numerous transcription factor-binding sites are present in promoters.
(ii) Transcription factor-binding sites are distributed sparsely and unevenly.
(iii) Transcription factor-binding sites are short and imprecise.
(iv) Many potential binding sites are nonfunctional.
(v) Variants within a binding site matrix can differ functionally.
(vi) Transcription factor-binding sites occupy a wide range of positions relative to the transcription unit.

(vii) Specific sequences limit the regulatory influence of binding sites.

(viii) Some binding sites affect transcription at more than one locus.

Interaction of the upstream regulatory elements with sequence-specific transcription factors determines the time, place and level of activity of all genes within a set of highly coordinated expression networks. Functional studies of plant promoters have given rise to a general concept that upstream regulatory elements are composite and not individual, where each *cis*-acting element contributes to the overall activity of the module through synergistic interactions between cognate transcription factors. The well-known cis-acting elements nested in the plant promoters are briefly summarized in Table 3.1.2

(iii) Enhancers:

An enhancer is defined as cis-acting module capable of stimulating gene expression when placed, in either orientation, upstream or downstream of the gene. While many enhancer-like sequences often direct tissue-specific or regulated expression, numerous plant genes have been reported to include non-tissue-specific upstream regulatory elements with quantitative enhancer-like qualities. A/T-rich sequences have been observed to direct quantitative expression enhancement of some of the plant genes (Bustos et al., 1989, Dean et al., 1989).

Table 3.1.2 Common *cis*-acting elements involved in temporal and/or spatial regulation of gene expression in plants

Responses		*Cis*-elements/ Consensus	*Trans*-acting Factors
Light	Light	GT-1 box, GGTTAA I-Box, GATAAAGR G-box, CACGTG H-box, ACCTA(A/C)C(A/C)	HY5
Metabolic Regulation	Auxin	TGTCTC motif TGTCCCAT box osc/as-1 element	ARF1
	Gibberellin	TAACA(A/G)A element TATCCAC element (C/T)CTTTT(C/T) element	GAmyb
	Abscisic acid	G-box, CCACGTGG	
	Ethylene	A(T/A)TTCAAA element	
	Sugars	TATCCA element GC-box, GCC(G/C)C G-box, CACGTG SURE, (AA)TACTA(A/T)T W-box, (T)TGAC(C/T)	
Environmental Stress	Heat	GAATTC element	HSF
	Oxidation	G-box, CACGTG H-box, ACCTA(A/C)C(A/C) W-box, (T)TGAC(C/T) GCC element	
	Cold, Drought	C- repeat, CCGAC G-box, CACGTG	
	Hypoxia	GT motif, (T/C)GGTTT GC motif, GCC(G/C)C	
	Pathogen, Wounding, Ethylene, JA, SA	G-box, CACGTG C- repeat, CCGAC GCC element, AGCCGCC W-box, (T)TGAC(C/T)	ERF1 WRKY

Development	Seed	RY motif, CATGCATG G-box, CACGTG E-box, CACCCTG AACA motif GNC4 motif, TGAGTCA P-box, TGTAAAG	Opaque2 PBF
	Fruit	TGTCACA motif	
	Pollen	32 bp motif	

Only transcription factors, for which there is strong evidence suggesting their importance in gene regulation, are included in this table (Adapted from Komarnytsky and Borisjuk, 2003).

Promoter function

At its most fundamental level, the function of a promoter is to integrate information about the status of the cell in which it resides, and to alter the rate of transcriptional initiation of a single gene accordingly (Wray et al., 2003). The promoters of genes encoding housekeeping proteins are constitutively active, but they can shut down in response to specific conditions, such as heat shock or starvation. Other promoters are off by default, but they can be activated in response to specific hormonal, physiological, or environmental cues. These diverse inputs eventually reach promoters in the form of transcription factors, proteins that bind in a sequence-specific manner to the DNA near a gene, altering rates of transcriptional initiation.

The mechanisms by which individual genes are activated are of intense interest and physiological importance, and studies over the past several years have revealed several levels of control (Roeder, 2005). First, eukaryotes contain three functionally distinct classes of nuclear RNA polymerases that selectively transcribe large ribosomal RNA genes (RNA polymerase I / Pol I), protein coding and some small structural RNA genes (RNA polymerase II / Pol II) and tRNA, 5S RNA and other small structural RNA genes (RNA polymerase III / Pol III). These specificities are reflected in the structurally distinct subunit compositions of the three RNA polymerases.

Second, eukaryotic cells contain RNA polymerase-specific general initiation factors that, despite the structural complexity of the enzymes (14, 12 and 17 subunits in RNA polymerases I, II and III, respectively), are necessary for accurate transcription initiation

on corresponding core promoter elements by purified RNA polymerases. These factors are now known to include TFIIIC and TFIIIB for RNA polymerase III; TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH for RNA polymerase II; and several factors for RNA polymerase I.

Third, eukaryotes contain diverse sequence-specific DNA binding transcriptional regulatory factors that facilitate RNA polymerase function on corresponding target genes. The vast majority of these are involved in the regulation of the large group of protein-coding genes transcribed by RNA polymerase II.

Fourth, the DNA-binding factors that regulate the transcription of protein coding genes act in conjunction with an expanding group of cofactors that act either through modifications of chromatin structure or, more directly, to regulate formation or function (transcription initiation or elongation) of the preinitiation complex (Figure 3.1.1) (Roeder, 2005).

Regulation of transcription initiation by RNA polymerase II is complicated by the requirement for a very large, multisubunit ‘‘adaptor’’ that bridges RNA polymerase II and its myriad DNA binding regulatory proteins and transduces both positive and negative signals that turn on and off messenger RNA synthesis in response to the ever changing microenvironment of the cell. This adaptor is composed of more than twenty subunits and has been named Mediator (Med) for its role in mediating transcriptional signals from DNA binding transcription factors bound at upstream promoter elements and enhancers to RNA polymerase II and the general initiation factors bound at the core promoter surrounding the transcriptional start site (Conaway et al., 2005). Mediator is required for transcription of all genes, whereas coactivators are specific to one or a small number of genes. Mediator, like pol II and the general transcription factors, is universal (Klug, 2005).

The general transcription factors and Mediator are responsible for the distinctive and most important characteristic of eukaryotic transcription, namely the recognition of 10 times as many different promoters as in bacteria, and the capacity to respond to regulatory inputs from dozens of upstream DNA sequence response elements, delivered through

specific transcription factors binding to each element (Klug, 2005). These specific transcription factors have a similar organisation to that found in bacteria, with a DNA binding domain and an interaction domain, but the key difference is that in eukaryotes their interaction is with Mediator, not directly with the polymerase.

Each polymerase interacts with a distinct set of factors that is unique to each class of genes, with the notable exception of TATA-binding protein, which is common to all three sets of factors indicating that there can be coordinate control of transcription from all three classes of genes. The resulting transcriptional complex for each polymerase recognizes a DNA region near the transcription start site (core promoter) in the genes of the specific class. The RNA polymerase II (pol II) transcription machinery is made up of three components:

- A 12-subunit polymerase, capable of synthesizing RNA and proofreading the nascent transcript,
- A set of five general transcription factors, denoted TFII B, -D, -E, -F, and –H, responsible for promoter recognition and for unwinding the promoter DNA,
- And a 20-subunit Mediator.

Assembly of a preinitiation complex (PIC) containing RNA polymerase II and general initiation factors is nucleated by binding of TFIID to the TATA element of the core promoter. A model for the regulation of PIC assembly and function involves, sequentially: (i) binding of regulatory factors to distal control elements; (ii) regulatory factor interactions with cofactors that modify chromatin structure to facilitate additional factor interactions; and (iii) regulatory factor interactions with cofactors that act after chromatin remodeling to facilitate, through direct interactions, recruitment or function of the general transcription machinery (Figure 3.1.1). The steps that follow this ‘preinitiation’ complex formation include initiation, promoter clearance, elongation and termination.

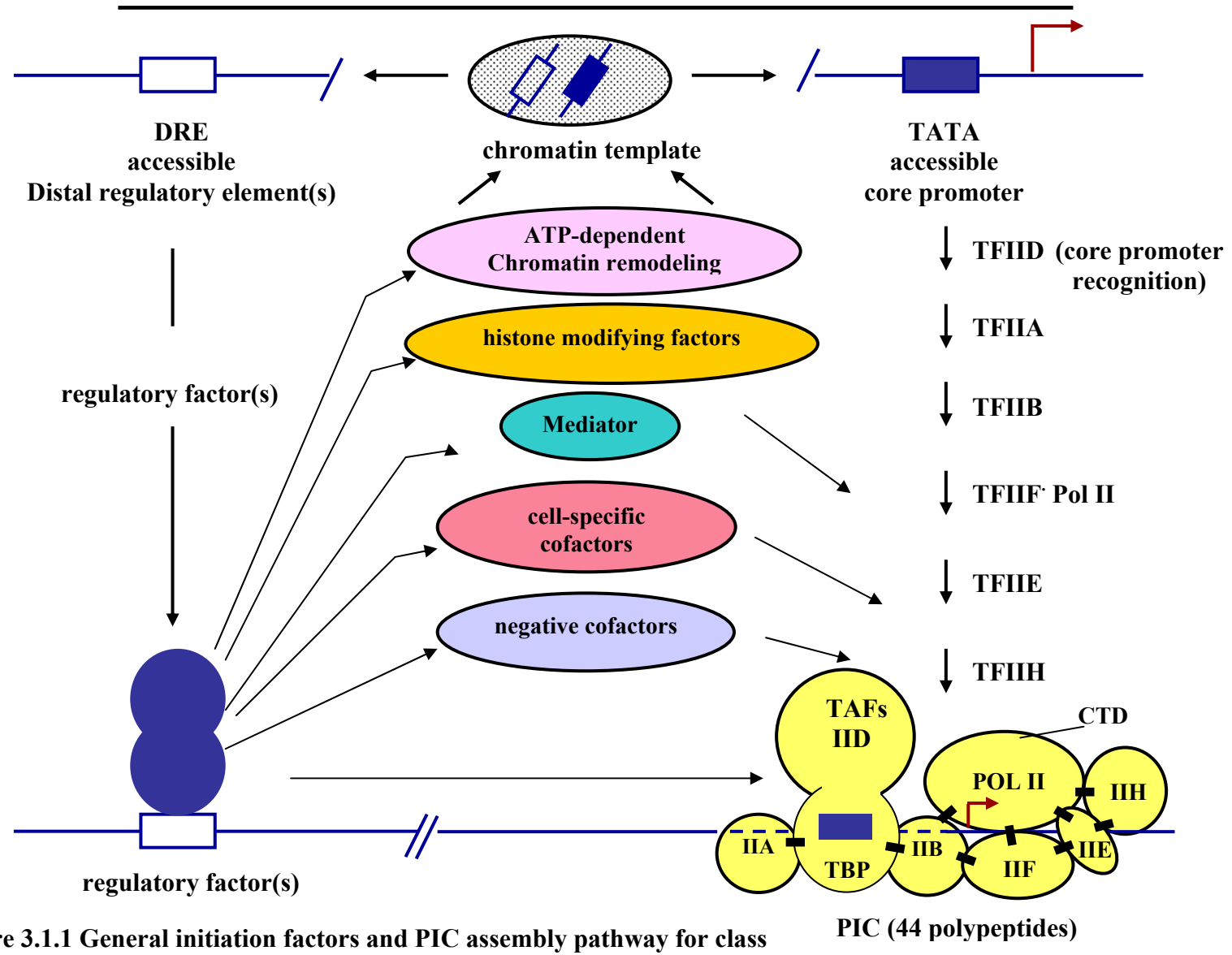


Figure 3.1.1 General initiation factors and PIC assembly pathway for class II genes with a TATA-containing core promoter, and regulation by gene specific factors and interacting cofactors (Roeder, 2005)

Transcription Factors

The transcription of every gene is regulated by transcription factors and cofactors those interact with its promoter. Transcription factors play important and diverse roles in gene expression, including chromatin remodeling and recruitment/ stabilization of the Pol II transcription initiation complex. Transcriptional factors, which come in many shapes and sizes, can be divided into a number of functional classes with some proteins belonging to more than one class (Singh, 1998):

(i) General transcription factors: These are important components of Pol II transcription initiation complex.

(ii) Architectural transcription factors: These are involved in remodeling DNA, e.g. by inducing bends that facilitate the binding of other proteins to the promoter.

(iii) Activators and repressors: These proteins bind to specific DNA sequences found only in certain promoters and are instrumental in giving rise to gene-specific regulation.

(iv) Co-activators or corepressors: These proteins mediate the transcriptional effects of specific activators/repressors, in some cases by remodeling chromatin. This group of transcription factors does not bind to DNA on their own; they can still be promoter specific as a result of protein-protein interactions with specific activators and repressors.

A typical plant transcription factor contains, with a few exceptions, a DNA-binding region, an oligomerization site, a transcription regulation domain, and a nuclear localization signal (Liu et al., 1999). Plant transcription factors have been classified on the basis of the structural features of conserved domains, which also show similarity to conserved domains of other organisms. Most known transcription factors can be grouped into families according to their DNA binding domain. There are many transcription factor families that are found only in plants, some of which have been greatly expanded. These include AP2/EREBP, NAC and WRKY families; the trihelix DNA binding proteins, the auxin response factors ARFs), the Aux/IAA proteins (which do not bind to DNA by themselves, but interact with ARF proteins), Dof proteins and other smaller families (Riechmann et al., 2000). The transcriptional output of a promoter is not a simple function

of which binding sites are present. The relative position, orientation, and nucleotide sequences of these binding sites, as well as the expression profiles of their cognate transcription factors and cofactors, all interact to produce the transcription profile of a gene (Wray et al., 2003).

Promoter Cloning

Promoters of genes have been isolated conventionally by screening genomic libraries using cDNA probes. Although efficient, this method consumes considerable amount of time, effort, and expenses and normally requires the use of radioactivity. Since the introduction of polymerase chain reaction (Saiki et al., 1988) several PCR-based methods to isolate flanking regions have been developed as alternative. The simple/complex nature of these methods and the effectiveness of these protocols are variable, and in many cases do not always result in successful isolation of flanking regions; with each method having its own advantages and disadvantages. These methods can be broadly grouped into restriction digestion dependent and restriction independent.

Restriction digestion based methods can be again grouped as ligation dependent and ligation independent. Inverse PCR (Ochman et al., 1988) and ligation-mediated PCR (Rosenthal and Jones, 1990; Mueller and Wold, 1991; Devon et al., 1995; Siebert et al., 1995; Kilstrup and Kristiansen, 2000) are some of the methods that rely on the presence of restriction sites in or near the target region. The restricted DNA fragments are self ligated (intra-molecular ligation) to form circular molecules (inverse PCR) or ligated to DNA cassette/adapters (ligation-mediated PCR) prior to PCR. Since prior information on restriction sites in the genomic DNA is not available, several enzymes need to be tried. Furthermore, achieving intra-molecular ligation at low DNA concentration to form circles can be a limiting factor in inverse PCR. In the case of DNA cassette/adapters mediated PCR, non-specific amplification products that are flanked by the DNA cassette/adapter sequence at both ends may be obtained. Although several modifications of the cassette/ adapters have been made (Lagerstrom et al., 1991; Siebert et al., 1995; Padegimas and Reichert, 1998), complete elimination of non-specific amplification is difficult. Use of biotin-labelled gene specific primers and streptavidin-coated magnetic beads (Rosenthal and Jones, 1990) can alleviate this problem, but this procedure requires biotin labelling and

separation of primers. Moreover, this increases the costs and time taken, and reduces the yield of desired target regions.

Ligation independent methods are also available wherein the restricted DNA fragments after denaturing are polyadenylated (Cormack and Somssich, 1997) or homo-polymer tailed (Liu and Baird, 2001) with terminal transferase. Techniques that are restriction and ligation independent include TAIL-PCR (Liu and Whittier, 1995) and Uneven PCR (Chen and Ray, 1997).

Use of reporter genes for promoter studies

The most widely used reporter gene for showing promoter activity by gene fusion experiments has been *uid*A, a gene from *Escherichia coli* that encodes the β-glucuronidase protein (E.C.3.2.1.31; Jefferson et al., 1987) also known as GUS. *E.coli* GUS has a monomeric molecular weight of 68.2 kDa, a pH optimum of 7-8, and functions as a homotetramer. In order to evaluate GUS activity several substrates are available. The most commonly used one is 5-Bromo-4-chloro-3-indolyl-β-D-glucuronide (X-Gluc). The reaction taking place in the histochemical GUS assay is presented in Figure 3.1.2. The reaction with X-Gluc generates a blue colour that is useful in histochemical detection of the gene activity. In plants, *E. coli* GUS works as fusion gene, where a promoter from a different organism directs the transcription of the *uid*A coding sequence, specifically regulating gene expression in time, quantity, and cell or tissue localization. Chimeric GUS fusion protein carrying a foreign peptide at its amino terminus retains enzymatic activity. There are two ways to fuse a promoter to the reporter gene (An, 1987). One way is to fuse genes such that the start codon of the natural reporter gene is the first AUG at the 5′ end of the recombinant primary transcript (transcriptional fusion). Translation initiation from internal AUG codons is found to be much less efficient in animals and plants. Therefore, the inserted promoter fragment should not introduce any extra start codons. The second way to fuse the genes is to make an in-reading-frame translational fusion so that the start codon belonging to the foreign fragment becomes the first codon of a reading frame that encodes a hybrid reporter fusion protein. The two start codons, one belonging to foreign fragment and the other to reporter gene, are in frame avoiding bicistronic messages.

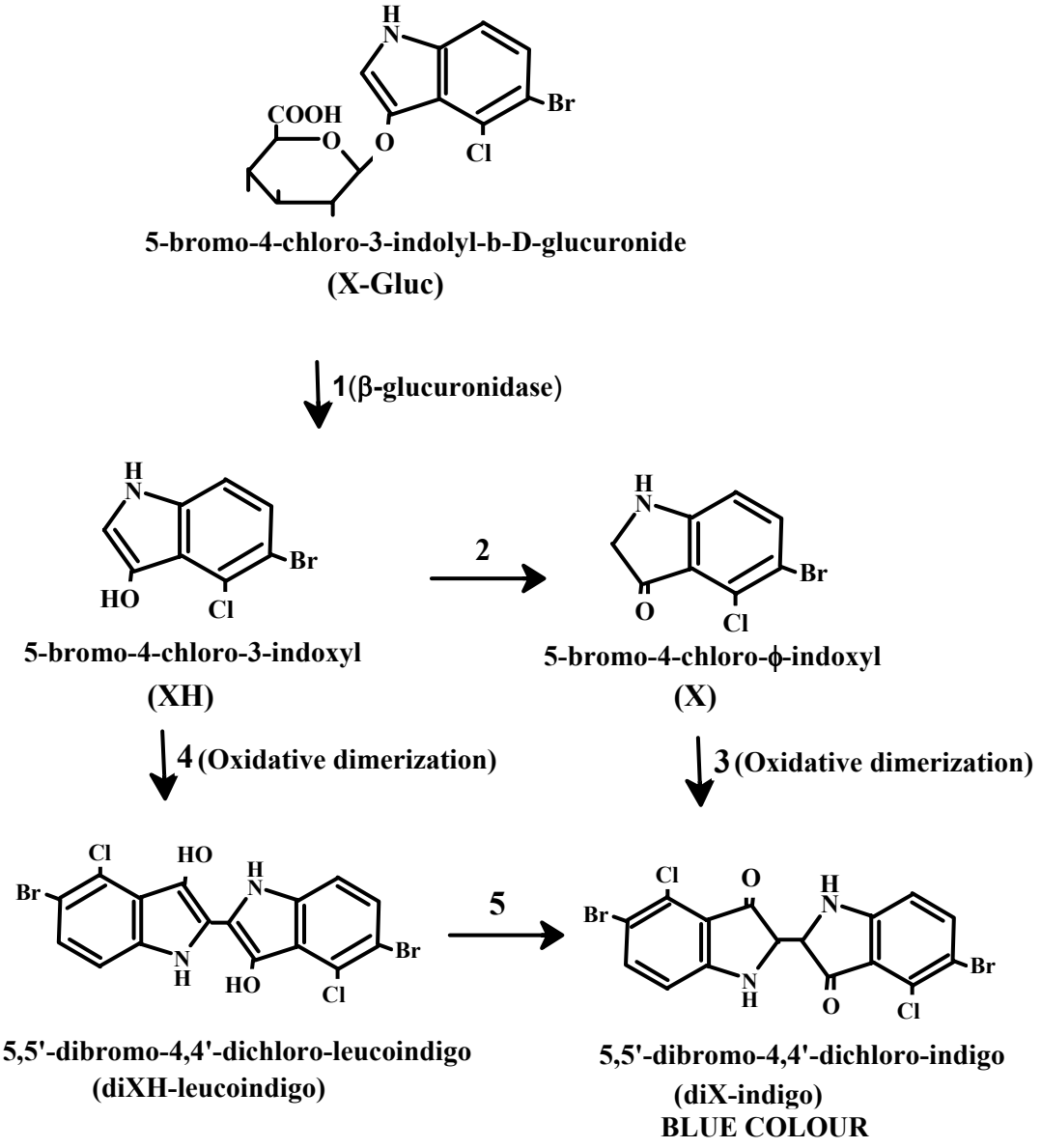


Figure 3.1.2 Reaction taking place in the histochemical GUS assay (Guivarc'h et al., 1996)

5-Bromo-4-chloro-3-indolyl-β-D-glucuronide (X-Gluc) is hydrolyzed by GUS to produce glucuronic acid and XH (5-Bromo-4-chloro-3-indoxyl), which reaches equilibrium with its tautomer X (5-Bromo-4-chloro-ϕ-indoxyl). Both XH and X are colourless. After oxidative dimerization of XH, another colourless molecule is formed, diXH-leucoindigo (5,5'-dibromo- 4,4'-dichloro-leucoindigo). Dimerization of X and oxidation of diXH-leucoindigo results in diX-indigo (5,5'-dibromo- 4,4'-dichloro-indigo) that has a blue colour and precipitates at the site of enzyme activity.

Transcriptional fusions monitor the transcriptional activity of the target gene, whereas with translational fusions, in addition, protein trafficking can be studied. However, caution should be observed while interpreting the results from promoter fusion analysis. It is becoming increasingly clear that promoter-reporter gene fusions are prone to artifactual expression that does not accurately reflect the *in vivo* regulation of the gene of interest (Taylor, 1997).

3.2 MATERIALS AND METHODS

3.2.1 Isolation of promoter for NMT gene from *Coffea canephora*

3.2.1.1 Genomic DNA Isolation

The genomic DNA from *Coffea canephora* var S-274 was isolated using the GenElute Plant Genomic DNA kit (Sigma, St. Louis, USA) following the manufacturer's instructions.

3.2.1.2 Design of oligonucleotides

For isolation of specific promoter regions by PCR based genomic walking method, non-overlapping gene-specific primers were designed based on the sequence of genomic clone for coffee *N*-methyltransferase gene (Clone CX10, Chapter 2). The primers were designed using the software PRIMER-3 and were synthesized at Genosys (Sigma, USA). The primers were used to amplify the region upstream of the coding sequence i.e. the 5′ untranslated region and the promoter region of the chosen gene. One of the nested gene specific primer was designed spanning the first intron /exon junction. Two adaptor specific primers were designed as described by Siebert et al., (1995).

Gene specific primers

2BEXR (5′-TCCGAATTCCGTAGAAAGAGCCAG-3′)

GSPin (5′-AGACAGACATTGTCGGATGC-3′)

GSP3 (5′-CAATGTCCCGCACTGTTAAA-3′)

Adapter specific primers

ASP1 (5′- GGATCCTAATACGACTCACTATAGGGC-3′)

ASP2 (5′- AATAGGGCTCGAGCGGC-3′)

3.2.1.3 Genomic walking as described by Siebert et al., 1995

The procedure for promoter isolation was the same as described by (Siebert et al., 1995) which combines the 'vectorette PCR' (Lagerstrom et al., 1991) with 'suppression PCR' (Lukyanov et al., 1994). The 'vectorette' feature of the adaptor is the presence of an amine group on the 3´- end of the lower strand. This blocks polymerase-catalyzed extension of the lower adaptor strand, preventing the generation of the primer-binding site unless a defined, distal, gene-specific primer extends a DNA strand opposite the upper strand of the adaptor. In rare cases, the 3´ end of the adaptor gets extended, due to incomplete amine modification during oligonucleotides synthesis or incomplete adaptor ligation. This creates a molecule that has the full-length adaptor sequence on both ends and can serve as a template for end-to-end amplification.

Without suppression effect, these rare events would lead to unacceptable backgrounds due to the exponential nature of PCR amplification. However, in 'suppression PCR' the adaptor primer is much shorter in length than the adaptor itself and is capable of hybridizing to the outer primer-binding site. During subsequent thermal cycling, nearly all the DNA strands will form the 'panhandle' structure (following every denaturation step due to the presence of inverted terminal repeats), which cannot be extended (Figure 3.2.1). At the appropriate annealing/ extension temperature, this intramolecular annealing is strongly favoured over (and more stable than) the intermolecular annealing of the much shorter adaptor primer to the adaptor. The suppression PCR effect is lost or reduced if a primer with annealing temperature lower than 60-65^{0}C is used. However, when a distal gene specific primer extends a DNA strand through the adaptor, the extension product will contain the adaptor sequence only on one end and thus cannot form the 'panhandle' structure. PCR amplification can then proceed normally.

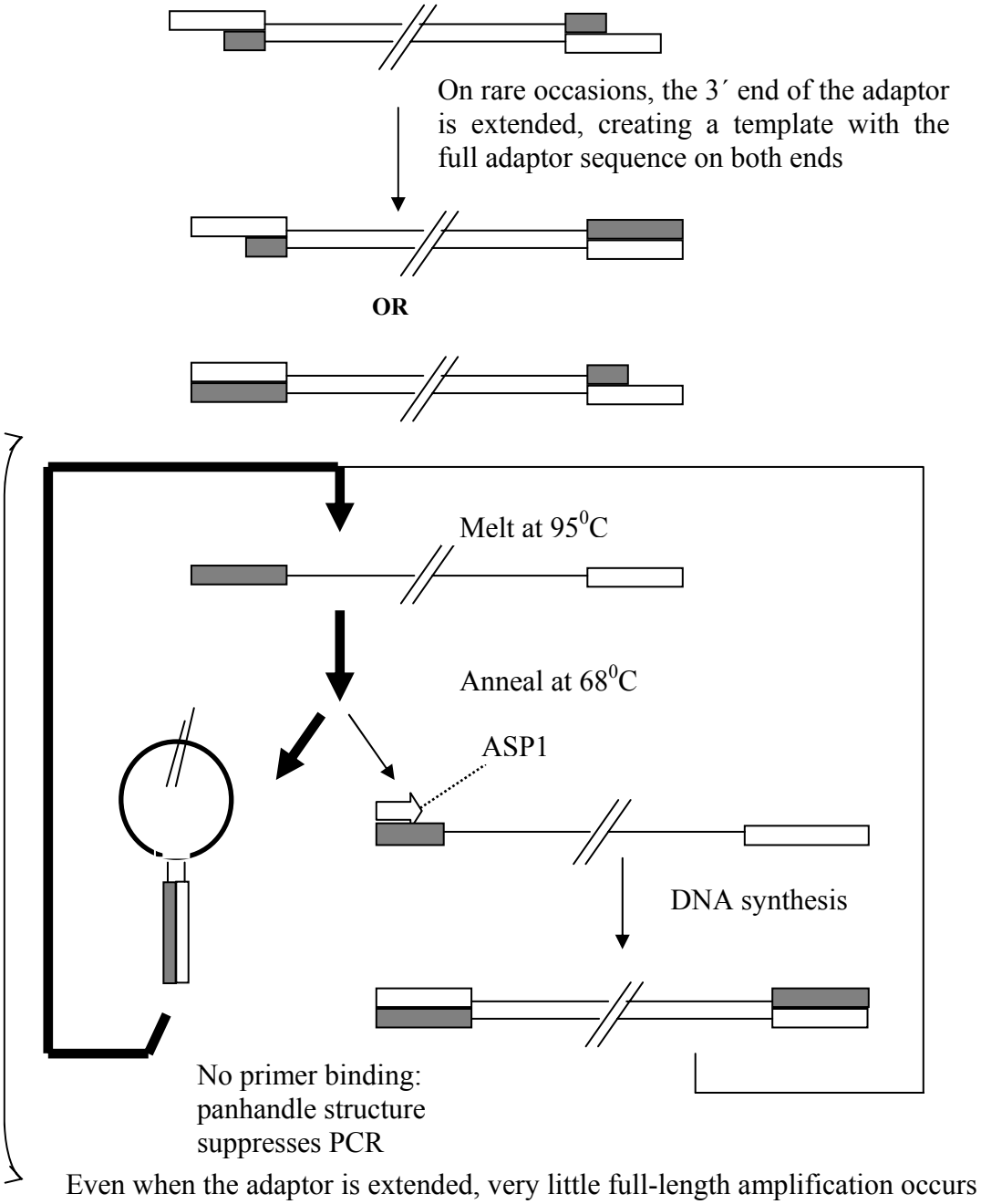


Figure 3.2.1 The suppression PCR effect

3.2.1.3.1 Adaptor annealing

Adapter Sequences

Adapter long arm-

5′- CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT- 3′

Adapter short arm -

5′- ACCTGCCC -NH_2-3′

The two adapters were dissolved in 25 µl of sterile deionized water and mixed together. The mixture was kept in a boiling water bath for 5 min and then the mixture was kept at room temperature for it to anneal slowly. The volume was made up to 100 µl to obtain a 10X solution (50 µM). The double stranded adapter was used at 1X concentration.

3.2.1.3.2 Construction of Adapter ligated DNA libraries

Coffee genomic DNA libraries were constructed by digesting the *Coffea canephora* var. S-274 genomic DNA (3µg) with 50 units of several blunt cutting restriction endonucleases. The DNA fragments were purified by phenol: chloroform: Isoamyl alcohol (25: 24: 1) extraction, followed by a chloroform extraction and precipitated in 95% ethanol in the presence of 3M Sodium acetate. The precipitates were washed with 70% ethanol and air-dried. The DNA fragments were dissolved in 10 µl of TE buffer (pH 8.0). Ten µl of DNA fragments were ligated to adapters at 16^0C under the following conditions: 40mM Tris-HCl, pH 7.8, 10mM DTT, 0.5mM ATP and 10 units of T4 DNA Ligase (MBI Fermentas) in a total volume of 20 µl. The ligation reaction was terminated by incubating the tubes at 65^0C for 10 min. The ligation mix was diluted 1:10.

3.2.1.3.3 PCR with Adapter ligated DNA libraries

Primary PCR was carried out using D*ra*I, E*co*RV, H*inc*II, S*ca*I and S*sp*I adaptor ligated DNA libraries as templates with specific primer ASP1 and gene specific primer 2BEXR (section 3.2.1.2). Primary PCR reactions were carried out in 50 µl volumes containing 1µl of ligated and diluted DNA, 10 mM TAPS (pH 8.8), 50mM KCl, 1.75mM MgCl2, 0.01% gelatin, 100 pm each of ASP1 and 2BEXR primers, and 1 unit of XT-Taq PCR system (Bangalore Genie, India) having proof reading activity. PCR was performed in a GenAmp

PCR System 9700 (PE Applied Biosystems). The 'touchdown' PCR parameters used during primary PCR reactions were as follows: 94^0C, 2 s; 70^0C, 3 min for seven cycles; 94^0C, 2 s; 65^0C, 3 min for 32 cycles; and a final hold at 65^0C for 7 min. Touchdown PCR involves using an annealing/extension temperature that is several degrees higher than the T_m of the primers during the initial PCR cycles. Although primer annealing (and amplification) is less efficient at this higher temperature, it is also much more specific. The higher temperature also enhances the suppression PCR effect with ASP1. This allows a critical amount of gene-specific product to accumulate. The annealing/extension temperature is then reduced to slightly below the primer T_m for the remaining PCR cycles, permitting efficient, exponential amplification of the gene-specific template. The extremely short incubation at 94^0C is necessary to preserve the integrity of the larger genomic DNA templates required for long distance PCR in the genomic walking protocol. The conditions for secondary nested PCR were as follows: 94^0C, 2 s; 67^0C, 3 min for seven cycles; 94^0C, 2 s; 62^0C, 3 min for 32 cycles; and a final hold at 62^0C for 7 min. Secondary PCR was conducted with 1µl of a 50-fold dilution of the primary PCR product using adapter specific primer ASP2 and the nested gene specific primer GSPin.

Another round of PCR based walking was performed using a different gene specific primer for primary PCR. Primary PCR was carried out using D*ra*I, E*co*RV, H*inc*II, P*vu*II, E*co*721, S*sp*I and S*ma*I adaptor ligated DNA libraries as template with the adaptor specific primer ASP1 and gene specific primer GSP3. Primary PCR reactions were carried out in 50 µl volumes containing 1µl of ligated and diluted DNA, 100 pm each of ASP1 and GSP3 primers, and 1 unit of AccuTaq Long PCR DNA polymerase (Sigma, MO, USA) having proof reading activity. Secondary PCR was conducted with 1µl of a 50-fold dilution of the primary PCR product using adapter specific primer ASP2 and the nested gene specific primer GSPin. The PCR parameters used for primary and secondary PCR were same as described earlier, except for a 4 min annealing / extension step.

3.2.1.4 Cloning of walking products

Approximately, ten µl of PCR products were size fractionated on 1% agarose gel. PCR product from E*co*RV adapter ligated DNA library was eluted from the gel, purified using

QIAQuick Gel Extraction Kit (QIAGEN, GmbH, Hilden Germany) and T/A cloned as described earlier using InsT/Aclone PCR Product Kit (MBI Fermentas, Lithuania). The PCR product from S*sp*I DNA library obtained in another round of PCR based walking was purified using QIAGEN MinElute PCR purification kit (QIAGEN, GmbH, Hilden Germany) and blunt end ligated to pETBlue-2 blunt vector using Perfectly Blunt Cloning Kit (Novagen, USA). The cloned PCR products were sequenced using the dideoxy chain-termination method at sequencing facility of Delhi University, South campus.

3.2.1.5 *In Silico* identification of promoter *cis*-elements

The sequences obtained for walking products were analyzed against the three databases that identify transcription factor binding sites or *cis*-acting sequences in plant promoters. PLACE (http://www.dna.affrc.go.jp/htdocs/PLACE/) is a database of 319 *cis*-acting regulatory DNA elements that were collected from previously published reports (Higo et al., 1999). PlantCARE http://sphinx/rug/ac.be: 8080/Plant CARE/index.htm) is a referential database with 417 different names of plant transcription sites describing more than 159 plant promoters (Lescot et al., 2002). TRANSFAC (http://www.gene-regulation.de/), the first transcription factor database of eukaryotic *cis*-acting regulatory elements and *trans*-acting factors, covers transcription factors from yeast to humans (Wingender et al., 2001). There are several programs on the home page like Patch, Match and SignalScan that are relevant for identifying *cis*-acting sequences and binding factors in a given DNA sequence. The sequence analyses for the promoter were also performed using Databases of Orthologous Promoters (DoOP; http://doop.abc.hu)(Barta et al., 2005). DoOP is a database of eukaryotic promoter sequences aiming to facilitate the recognition of regulatory sites conserved between species. The plant database contains 7548 clusters that have sequences from at least two species.

3.2.2 Creation of promoter: GUS translational fusion constructs

The 774 bp PCR walking product obtained from E*co*RV adapter ligated DNA library was cloned into pTZ57R/T vector and sequenced. The resulting construct was designated as pTZ774. An 815 bp E*co*R1/B*am*H1 fragment spanning the - 691 bp to + 83 bp region (Figure 3.2.2)(numbering related to ATG of the theobromine synthase gene from a cDNA clone, accession no. AB048794) was released from pTZ774. The +83 bp region comprised

of 75 bases of first exon and 8 bases of the first intron. The binary vector pCAMBIA1381 with a cassette containing the *GUS* gene and the NOS terminator was selected for insertion of the promoter next to *GUS* reporter gene and for subsequent transformation into tobacco. This binary vector had Kanamycin as bacterial and Hygromycin as plant selection marker gene. The binary vector was digested with E*co*R1/B*am*H1 restriction enzymes for 8 h at 37^0C. After heat denaturing the enzymes for 15 min at 70^0C, the mixture was extracted with chloroform, and DNA was precipitated with one tenth volume 3 M sodium acetate and 2 volumes of absolute ethanol. The pellet was resuspended in TE buffer. The E*co*R1/B*am*H1 promoter fragment released from pTZ774 was fused in frame with a cassette containing the GUS gene and the NOS terminator in pCAMBIA 1381 digested with E*co*R1/B*am*H1. Digested promoter was mixed with digested vector in a 2:1 ratio and ligation buffer was added to 1X. Five units of T4 DNA Ligase (MBI Fermentas, Lithuania) were added and the mixture was incubated at 22^0C for 8 h. The ligation reaction was used to transform 100 µl of *E. coli* competent cells (DH5α), which were spread on LB plates containing 50µg/ml kanamycin (kan^+) (Sigma, USA). Positive clones, which grew on kan^+ plates, were screened by plasmid isolation, followed by restriction digests. The resulting translational fusion construct was designated as pPCTS774.

The 938 bp PCR walking product obtained from SSPI adapted library was cloned into pETBlue-2 vector and sequenced. The resulting construct was designated as pETSSPI. Another promoter: GUS fusion construct was made with the 938 bp PCR product comprising the 29 bases of adapter and 909 bases of the coffee genomic fragment from the clone pETSSPI. The 938 bp PCR product was reamplified from the clone pETSSPI using the primers ASP2 and GSPin. The PCR product was cloned into pTZ57R/T vector and sequenced for checking the orientation of the insert. The resulting construct was designated as pTZ938. A 979 bp E*co*R1/B*am*H1 fragment spanning the - 896 bp to + 83 bp region (Figure 3.2.2)(numbering related to ATG of the theobromine synthase gene, accession no. AB048794) was released from pTZ938. The +83 bp region comprised of 75 bases of first exon and 8 bases of the first intron. The E*co*R1/B*am*H1 promoter fragment released from pTZ938 was fused in frame with a cassette containing the *uidA* gene and the NOS

terminator in pCAMBIA 1381 digested with E*co*R1/B*am*H1. The resulting translational fusion construct was designated as pPCTS938.

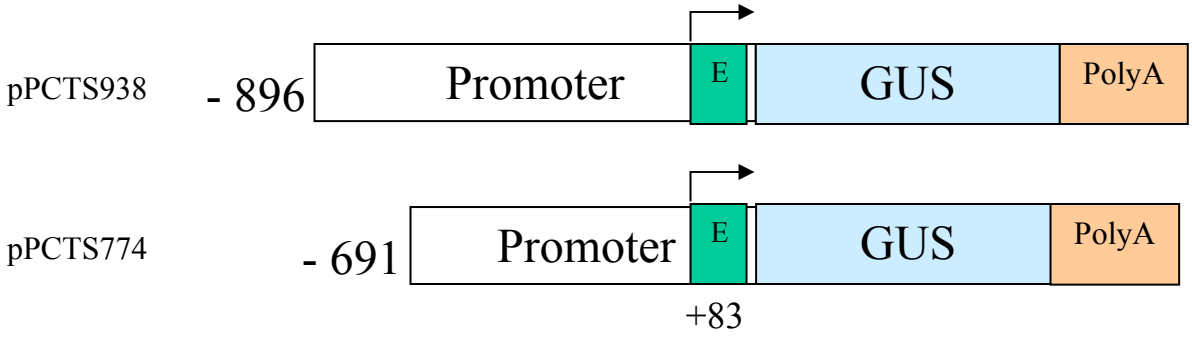


Figure 3.2.2 Schematic representation of the two promoter: reporter gene fusion constructs

'E' represents the first exon; arrow indicates the translational start codon.

3.2.3 *Agrobacterium* mediated transformation of tobacco

Agrobacterium tumefaciens strain EHA 101 was used in the transformation experiments. Binary vectors pCAMBIA 1301 as positive control, pCAMBIA 1381 as negative control and pPCTS774 were introduced to *Agrobacterium tumefaciens* using the freeze-thaw method (An et al., 1988) in presence of 20mM $CaCl_2$. *Nicotiana tabacum* var. Anand 119 was used for plant transformation experiments. The seeds were surface sterilized with 0.2% $HgCl_2$ and germinated *in vitro* in Murashige and Skoog (MS) medium. The leaf segments of 2 x 2cm size were used as explants for the transformation experiments. The leaf segments were co-cultivated with *Agrobacterium tumefaciens* strain EHA101 harbouring various expression cassettes for infection. For co- cultivation, wounded leaf segments were immersed in the bacterial culture, resuspended in MS medium and cultivated on co-cultivation medium containing 100µM acetosyringone for 2 days. Subsequently, the explants were treated with 500 mgl^{-1} cephotaxim in sterile water and cultured on callus induction medium comprising of 2 mgl^{-1} IAA and 0.5 mgl^{-1} Kinetin, 5 mgl^{-1} hygromycin, 500 mgl^{-1} cefotaxim and 300 mgl^{-1} potassium clavaluanate. After one month, the cultures were selected on 10mg/l hygromycin (Duchefa, The Netherlands). The

selected callus lines were subjected to GUS assay to confirm the expression of the reporter gene.

3.2.3.1 Murashige and Skoog (MS) medium

The composition of MS medium is given in Table 3.2.1

Table 3.2.1 Composition of Murashige and Skoog (1962) media

Nutrients	Quantity mg l^{-1}
Macroelements:	
$CaCl_2.7H_2O$	440
KH_2PO_4	170
KNO3	1900
$MgSO_4.7H_2O$	370
NH_4NO_3	1650
Microelements:	
$CoCl_2.6H_2O$	0.025
$CuSO_4.5H_2O$	0.025
$FeSO_4.7H_2O$	27.85
Na_2 EDTA	37.35
H_3 BO_3	6.20
KI	0.83
$MnSO_4.4H_2O$	22.3
$Na_2MoO_4.2H_2O$	0.25
$ZnSO_4.7H_2O$	8.60
Myo-inositol	100.00
Vitamins:	
Glycine	2.00
Nicotinic acid	0.50
Pyridoxine HCl	0.50
Thiamine HCl	0.10

3.2.4 Regeneration of transformed plants

Plantlets were regenerated from the transgenic callus under the influence of 4 mgl^{-1} benzyl amino purine and 0.5 mgl^{-1} Indole acetic acid along with 5 mgl^{-1} hygromycin, 500 mgl^{-1} cefotaxim and 300 mgl^{-1} potassium clavaluanate. These plants were also subjected to GUS assay to analyze the tissue localization of the GUS reporter gene expression driven by the promoter for coffee *N*-Methyltransferase gene.

3.2.5 Electroporation of coffee endosperm

Green unripe fruits of *C. canephora* cv. S-274 were collected and the endosperm tissues were used in electroporation experiments. Electroporation was carried out according to earlier report (Fernandez and Menendez, 2003). The endosperm tissues were taken with 100µg/ml of pPCTS938 plasmid DNA in 350 µl electroporation buffer (pH 5.6) consisting 70 mM aspartic acid, 5 mM calcium gluconate, 5 mM MES and 0.5 mM mannitol. The tissues were incubated one hour at 4^0C prior to electroporation. Electroporation was carried out using a BioRad Gene Pulser with single electric pulse of 500 V/cm (Fernandez and Menendez, 2003), discharged from a 900-µF capacitor. Electroporation of endosperm without plasmid DNA served as a negative control. The tissues were incubated for an additional hour at 4^0C after discharge.

3.2.6 Histochemical staining of GUS activity

Histochemical staining of GUS activity with 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc) was performed using a procedure described by Jefferson et al., (1987). The transformed tissues and electroporated endosperm were immersed in 500 µl of X-Gluc (Sigma, St. Louis, USA) staining solution (2 mM X-Gluc, 50 mM sodium phosphate buffer pH 7.0, 10 mM EDTA, 0.5 mM potassium ferricyanide, 0.5 mM potassium ferrocyanide, 0.1% Triton X-100). Methanol was added to a final concentration of 20% to reduce the background caused by endogenous activities in the plant material. The tissues were incubated in dark at 37^0C overnight. In the case of plantlets, the samples were incubated in 70% alcohol to remove the chlorophyll. Squash preparations of transformed tissues were performed. The specimens were observed under microscope and photographed.

3.2.7 PCR amplification and cloning of genomic fragments containing the promoter and the gene

An upstream primer CTSF1 5′-GTTCCCCATTTTCCTCCTTTCAAGTAG-3′ was designed based on the additional sequence obtained through genome walking. This primer along with the reverse primer XMTR1 5′-ACGACAATACCC GAAAGACC-3′ located on unique sequence of 3′ UTR of CaMXMT-1 gene (accession no: AB048794) was used in amplification of genomic fragment containing the promoter and the coding region. The

parameters used during the PCR reactions were as follows- seven cycles: 94^0C, 2 s; 67^0C, 4 min; 30 cycles: 94^0C, 2 s; 62^0C, 4 min, in the last cycle additional 7 min at 62^0C. JumpStart KlenTaq LA DNA Polymerase (Sigma, MO, USA) was used in PCR for increased fidelity.

PCR products were size fractionated on 1% agarose gel, eluted from the gel and purified using QIAquick Gel Extraction Kit (QIAGEN, Germany). The purified PCR product was A-tailed and T/A cloned using InsT/Aclone PCR Product Kit (MBI Fermentas, Lithuania) as described earlier. Plasmids from several transformed colonies were isolated and recombinant nature of plasmids was verified by insert release.

3.2.8 Sequencing and analysis of the genomic clones of the clones

One of the positive recombinant plasmid PG-5 was sequenced at the Dept. of Biochemistry, University of Delhi South Campus, New Delhi.

3.3 RESULTS AND DISCUSSION

3.3.1 Isolation and cloning of promoter for *N*-Methyltransferase gene

The promoter was isolated by PCR based genomic walking procedure as described by Siebert et al., 1995. Based on the genomic sequence of theobromine synthase gene (Clone CX10, Chapter II), two non-overlapping nested gene-specific primers were designed to amplify the region upstream of the coding sequence. The primary PCR was carried out using ASP1 and 2BEXR primers with five adaptor-DNA ligated libraries as templates; followed by a second PCR with nested ASP2 and GSPin primers. Ten µl of PCR products were size fractionated on 1% agarose gel. PCR products were obtained from two adapter libraries (Figure 3.3.1). Approximately, 800 bp PCR product from E*co*RV adaptor ligated DNA library was eluted from the gel and purified using QIAQuick Gel Extraction Kit (QIAGEN, GmbH, Hilden Germany). The purified PCR product was cloned using InsT/Aclone PCR Product Kit (MBI Fermentas, Lithuania) and the cloned product was sequenced.

Similarly, another round of walking experiment was carried out using seven adaptor ligated DNA libraries as templates. The primary PCR was carried out using ASP1 and GSP3 primers; followed by a second PCR with nested ASP2 and GSPin primers. Ten µl of PCR products were size fractionated on 1% agarose gel. Successful amplification was obtained only with the S*sp*I library (Figure 3.3.2), which resulted in ~950 bp PCR product. The PCR product was blunt end ligated to pETBlue-2 blunt vector and the cloned PCR product was sequenced. The 774 bp sequence obtained from E*co*RV library comprised of 29 bases of the adapter sequence. Similarly, the 938 bp sequence obtained from S*sp*I library contained 29 bases of the adapter sequence.

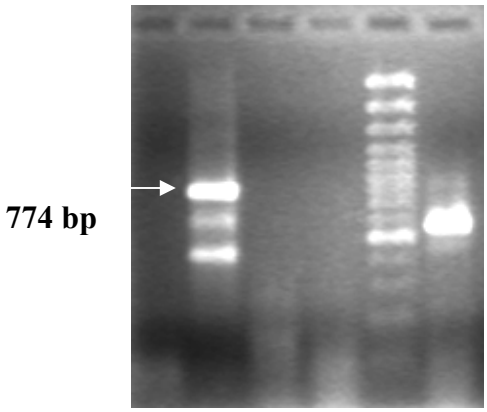


Figure 3.3.1 Agarose gel electrophoresis of PCR walking products
Lane 1: D*ra*I library, Lane 2: E*co*RV library, Lane 3: H*inc*II library,
Lane 4: S*ca*I library, Lane M: 100 bp DNA ladder, Lane 5: S*sp* I library

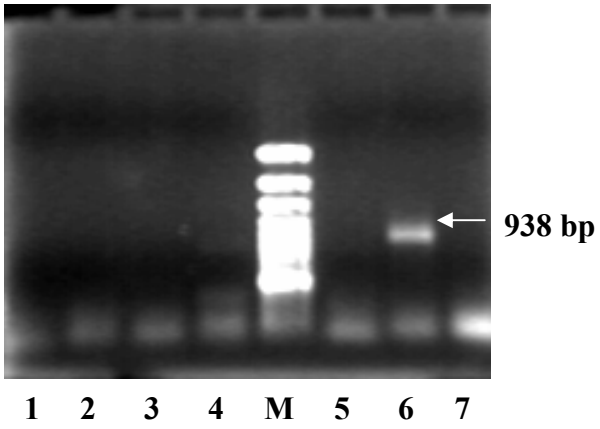


Figure 3.3.2 Agarose gel electrophoresis of PCR walking products
Lane 1: D*ra*I library, Lane 2: E*co*RV library, Lane 3: H*inc*II library,
Lane 4: P*vu*II library, Lane M: 100 bp DNA ladder, Lane 5: E*co*721 library,
Lane 6: S*sp* I library, Lane 7: S*ma* I library

The DNA alignments (Figure 3.3.3) of the sequences for two walking products revealed that the two walking products were identical i.e. the 745 bp fragment was common to the 909 bp fragment; and possibly originated from the same genomic region.

```
SSPI            ATTATGAGAAATTCACCGAATAAATAGGCAAAGGTAGAACCTCACTTCTTAAATGAAATT 60

SSPI            ACCTAATATCCTTTCCATTTAACCAAAAGTTCCCCATTTTCCTCCTTTCAAGTAGAATTA 120

SSPI            TCATGTGGGCAAGTAAATTCTGAGCAAGAAATAATAAAAAAAGGAGTTGATTTGACATTG 180
EcoRV           ----------------------------------------------AGTTGATTTGACATTG 16

SSPI            AATAGACTCCAGCTTCTTCCTTTCATTTGTTGAGAGGAAAAAGAAGCAGACGGGAAAGAA 240
EcoRV           AATAGACTCCAGCTTCTTCCTTTCATTTGTTGAGAGGAAAAAGAAaCAGACGGGAAAGAA 76

SSPI            ATAGGAATGAAGCAGAAAAATGGTGGGCTAAAAGTCAAAAGAAGCGGAAAGTGATAGTGT 300
EcoRV           ATAGGAATGAAGCAGAAAAATGGTGGGCTAAAAGTCAAAAGAAGCGGAAAGTGATAGTGT 136

SSPI            AGAGGAAGAAGGGAAAGAGGCGGAAGAACAGGAGGAAAGGTACAGGTGAAAAATCAAAAT 360
EcoRV           AGAGGAAGAAGGGAAAGAGGCGGAAGAACAGGAGGAAAGGTACAGGTGAAAAATCAAAAT 196

SSPI            TCAGCCCAAATAGTAAGCTACTATTCTGAAAGCCTGTTATGTCTAGAAGGACAAACACGT 420
EcoRV           TCtGCCCAAATAGTAAGCTACTATTCTGAAAGCCTGTTATGTCTAGAAGGACAAACACGT 256

SSPI            ACTTTTAAGGAAAAGGACAAAAATGCCCTAGAATAGTTATACACTGTTTAGACCAGGGTA 480
EcoRV           ACTTTTAAGGAAAAGGACAAAAATGCCCTAGAATAGTTATACACTGTTTAGACCAGGGTA 316

SSPI            CACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTGGCTGCTTCTCTCCCATCTTT 540
EcoRV           CACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTGGCTGCTTCTCTCCCATCTTT 376

SSPI            TGCAACGTGGGAGAAGAAGCCTTAGGTTTTTATATATATATGTAAGATGTAAGATAAGAT 600
EcoRV           TGCAACGTGGGAGAAGAAGCCTTAGGTTTTTATATATATATGTAAGATGTAAGATAAGAT 436

SSPI            AAGATATTTCATCCCCGCATCTCAACTTCTGATTTTATCATTCGTGTCTGGTTCCCATTG 660
EcoRV           AAGATATTTCATCCCCGCATCTCAACTTCTGATTTTATCATTCGTGTCTGGTTCCCgTTG 496

SSPI            GCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGCTGGTTA 720
EcoRV           GCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGCTGGTTA 556

SSPI            GCATACGCATCTATGAAATTTTCGCTACTTAAGCCCGAAATTTTGCACAATTAATCATTA 780
EcoRV           GCATACGCATCTATGAAATTTTCGCTACTTAAGCCCGAAATTTTGCACAATTAATCATTA 616


SSPI            ACAGACACCTTCTTTAGCAGTCGCAATTCGATTGTCCTGCATATGAATGGAGCTCCAAGA 840
EcoRV           ACAGACACCTTCTTTAGCAGTCGCAATTCGATTGTCCTGCATATGAATGGAGCTCCAAGA 676


SSPI            AGTCCTGCATATGAATGAAGGTGAAGGCGATACAAGCTACGCCAAGAATGCATCCGACAA 900
EcoRV           AGTCCTGCATATGAATGAAGGTGAAGGCGATACAAGCTACGCCAAGAATGCATCCGACAA 736

SSPI            TGTCTGTCT                                                     909
EcoRV           TGTCTGTCT                                                     745
                 Intron
```

Figure 3.3.3 Nucleotide alignments for two genomic walking products

Nucleotide BLAST searches of sequences upstream of start codon did not reveal any significant homology to any known nucleotide sequence in the GenBank database and the Eukaryotic Promoter Database. However, the first exon of the gene showed significant

homology to the theobromine synthase gene sequences as shown by the BLAST alignment (Figure 3.3.4).

gi|26453375|dbj|BAC43756.1| **theobromine synthase 1** [Coffea arabica]
Length = 378

Score = 53.9 bits (128), Expect = 7e-07
Identities = 24/26 (92%), Positives = 25/26 (96%)

Query: 1 MELQEVLHMNEGEGDTSYAKNASDNV 26
MELQEVLHMNEGEGDTSYAKNAS N+
Sbjct: 1 MELQEVLHMNEGEGDTSYAKNASYNL 26

gi|13365753|dbj|BAB39216.1| **7-methylxanthine N-methyltransferase** [Coffea arabica]
gi|20271018|gb|AAM18501.1|AF494411 N-methyltransferase [Coffea arabica]
gi|20271024|gb|AAM18504.1|AF494414 N-methyltransferase [Coffea canephora]
Length = 378

Score = 53.5 bits (127), Expect = 8e-07
Identities = 24/26 (92%), Positives = 25/26 (96%)

Query: 1 MELQEVLHMNEGEGDTSYAKNASDNV 26
MELQEVLHMNEGEGDTSYAKNAS N+
Sbjct: 1 MELQEVLHMNEGEGDTSYAKNASYNL 26

Figure 3.3.4 Nucleotide alignments performed using BLASTn programme

The 5′ untranslated region (UTR) of the walked DNA fragments also correspond to the cDNA sequences of coffee NMT gene i.e. theobromine synthase-1 gene as observed from the multiple sequence alignment (Figure 3.3.5).

```
TS745              1   aatagggctc gagcggccgc ccggcagtga gttgatttga cattgaatag
                                Adapter sequence
TS745             51   actccagctt cttcctttca tttgttgaga ggaaaaagaa acagacggga

TS745            101   aagaaatagg aatgaagcag aaaaatggtg ggctaaaagt caaaagaagc

TS745            151   ggaaagtgat agtgtagagg aagaagggaa agaggcggaa gaacaggagg

TS745            201   aaaggtacag gtgaaaaatc aaaattctgc ccaaatagta agctactatt

TS745            251   ctgaaagcct gttatgtcta gaaggacaaa cacgtacttt taaggaaaag

TS745            301   gacaaaaatg ccctagaata gttatacact gtttagacca gggtacactt

TS745            351   tatgaggtta ttgctggaaa aaacgaaaca tagtggctgc ttctctccca

TS745            401   tcttttgcaa cgtgggagaa gaagccttag gtttttatat atatatgtaa

TS745            451   gatgtaagat aagataagat atttcatccc cgcatctcaa cttctgattt

TS745            501   tatcattcgt gtctggttcc cgttggctgt gcgctttctt tctgacgaat

TS745            551   tgacagactt ttctacgcac ggaggtagct ggttagcata cgcatctatg

CTS1               1   ---------- ---------- ---------- ---------tt gATTAACAGA
TS745            601   aaattttcgc tacttaagcc cgaaattttg cacaattaat cATTAACAGA

CTS1              13   CACCTTCTTT AGCAGTCGCA ATTCGATTGT CCTGCATATG AATGGAGCTC
CaMXMT1            1   ---------- AGCAGTCGCA ATTCGATTGT CCTGCATATG AATGGAGCTC
TS745            651   CACCTTCTTT AGCAGTCGCA ATTCGATTGT CCTGCATATG AATGGAGCTC

CTS1              63   CAAGAAGTCC TGCATATGAA TGAAGGTGAA GGCGATACAA GCTACGCCAA
CaMXMT1           41   CAAGAAGTCC TGCATATGAA TGAAGGTGAA GGCGATACAA GCTACGCCAA
TS745            701   CAAGAAGTCC TGCATATGAA TGAAGGTGAA GGCGATACAA GCTACGCCAA

CTS1             113   GAATGCATCC TACAATCTGG CTCTTGCCAA GGTGAAACCT TTCCTTGAAC
CaMXMT1           91   GAATGCATCC TACAATCTGG CTCTTGCCAA GGTGAAACCT TTCCTTGAAC
TS745            751   GAATGCATCC GACAATgtct gtct------ ---------- ----------

CTS1             163   AATGCATACG AGAATTGTTG CGGGCCAACT TGCCCAACAT CAACAAGTGC
CaMXMT1          141   AATGCATACG AGAATTGTTG CGGGCCAACT TGCCCAACAT CAACAAGTGC
TS745            775   ---------- ---------- ---------- ---------- ----------
```

Figure 3.3.5 Multiple alignments of sequences for genomic walking product and theobromine synthase-1 cDNA

3.3.2 *In Silico* analysis of the 5′ upstream region

The putative transcription start site (TSS) was found 216 nucleotides upstream to the start codon (Figure 3.3.6). The adenine flanked by pyrimidine bases present in the consensus sequence of transcription start site, TSS (Joshi, 1987) was numbered as +1. The TATA box sequence was found 30 bases upstream to the indicated TSS. Two putative CAAT boxes were found upstream to predicted TSS at –79 and –111 nucleotides. The promoter sequence for pETSSPI was analyzed for potential cis-elements using PLACE and

PlantCARE programs. Besides the TATA-box and CAAT-box, several motifs have been found (Table 3.3.1).

Table 3.3.1. Putative *cis*-elements present in the upstream sequence of coffee NMT gene

Putative cis-element/ consensus	Function/ Response	Location
Dof protein binding sites AAAG	Signal responsive and / or tissue specific gene expression	-179, -222, -275, -297, -323, -333, -340, -376, -391, -451, -525, -581,
GT-1 Box, GRWAAW	Light-responsive element	-110, -182, -263, -356, -395
GATA motif	Light-responsive element	-318
TATTCT, -10 promoter element	Light responsive element	-229
TTGAC, W-box	Biotic stress responsive element	-142, -338, -439,
CANNTG, E-box	Seed specific expression/ Plant pathogen interaction	-268, -407, -489
(CA)n element	Seed specific activator and repressor	-199
TGHAAARK, -300 element	Endosperm specific enhancer element	-264
SEF4 motif, RTTTTTR	Seed specific enhancer	-45
GCAACG	Conserved in seed specific promoters	-69
ACGT core	bZIP protein binding site	-66, -194
AACGTG, T/G box	Light, ABA, or Methyl Jasmonate induction	-67
AACCAA, REALPHALGLHCB21	Phytochrome	-530
TACTATT, SPF-1 binding site		-232, -242

Sequence analysis of DNA fragment located upstream to first ATG of pETSSPI, using PLACE program (Higo et al., 1999) based on TRANSFAC database, revealed the presence of several putative motifs known to be of great importance in the promoter function. The most frequently reported domains in the promoter sequence are those recognized by Dof proteins. Twelve domains with AAAG core sequence at -179, -222, -275, -297, -323, -333, -340, -376, -391, -451, -519 and -581 nt were found (Figure 3.3.6). Dof proteins bind to the AAAAGG core sequence motif found in many plant promoters. Dof1 seems to be a key factor for the response to light in leaves. Dof domains act as transcription activation domains *in vivo* (Yanagisawa, 1997). GT-1 *cis*- element (GRWAAW) was found at five positions -110, -182, -263, -356 and –395, whereas GATA motif was present at –318. The presence of several light regulatory elements including Dof suggests that light may play an important role in regulation of *N*-methyltransferases. It has been demonstrated in coffee cell suspension cultures that light irradiation may play an important role in enhancing purine alkaloid production as an inducer of *N*-methyltransferases (Kurata et al., 1997).

Analysis of the promoter sequence also revealed the presence of binding sites for W-box motif (TTGAC at -338, -439 and -142 nt) for the plant transcription factor family WRKY. The WRKY transcription factors have been shown to induce gene expression in response to wounding (Eulgem et al., 1999) and the W-box was present in 80% of genes identified as pathogen inducible (Chen et al., 2002). WRKY factors, being transcriptional regulators, can act by directing the temporal and spatial expression of specific genes, thereby ensuring proper cellular responses to internal and external stimuli (Ulker and Somssich, 2004). WRKY factors show high binding affinity to a DNA sequence designated as W box (C/T)TGAC(T/C). These elements invariably contain the TGAC core, which is essential for function and WRKY binding. The transcription of WRKY genes is strongly and rapidly upregulated in response to wounding, pathogen infection or abiotic stresses in numerous plant species (Eulgem et al., 2000). WRKY proteins are also involved in other plant-specific processes such as trichome development, senescence and the biosynthesis of secondary metabolites (Eulgem et al., 2000).

Among other *cis*- elements that are recognized are E box CANNTG (-268, -407 and -489 nt), (CA)n element (-199); SEF4 motif (RTTTTTR, -45); -300 element (TGHAAARK, -

264). E box CANNTG-motifs are responsible for temporally and spatially specific gene-expression (Kawagoe and Murai, 1992). The E-box is also recognized as a Myc-binding site that participates in ABA induction and binds a drought inducible Myc homologue (Busk and Pages, 1998). ACGT is the core sequence of DNA motifs that bind bZIP proteins. Plant myb-elements function in controlling secondary metabolism, regulating cellular morphogenesis and serving in the signal transduction pathways responding to plant growth regulators. CACTTC motif was observed at -563 position, which is conserved in the promoter region of many seed protein genes, including *HaG3-D*, *LegK*, *Arcelin*, *Del2*, *Zein Zc2* and B-*conglycinin* (Li and Hall, 1999).

The –300 element (–264, TGHAAARK) was identified as an enhancer element in hordein-, gliadin- and glutenin-encoding genes (Szopa et al., 2003). There are several sequences that are required for high level, light-regulated and tissue-specific expression, like GATA boxes, I-box and GT1 boxes. The latter can stabilize the TFIIA-TBP-DNA complex. The presence of an SEF4-binding site suggests that these elements operate as an enhancer, similar to the soyabean gene encoding storage proteins (Lessard et al., 1991). Although the CCAAT box generally acts as a general transcription activating sequence in eukaryotes, it can also contribute to tissue-specific gene regulation. For example, *Leafy cotyledon 1* (*LEC1*), a homologue of CCAAT box-binding factor HAP3 subunit, is involved in seed-specific regulation of 2S storage protein genes and the *oleosin* gene (Lotan et al., 1998).

Many plant nuclear genes contain enhancer elements responsible for regulated gene expression located upstream of the transcription start sequences (Sandhu et al., 1988). A/T rich regions were observed in the position –523 to –565 and -432 to –505. Quantitative enhancement has been observed with some A/T-rich enhancer elements (Bustos et al., 1989; Czarnecka et al., 1992). A/T-rich sequences are able to act as quantitative, non-tissue-specific enhancer elements in higher plants (Sandhu et al., 1988) or as tissue-specific enhancer (Pwee and Gray, 1993). A/T-rich enhancer elements, acts by remodeling local chromatin structure from scaffold-associated regions (SARS), thus influencing expression from adjacent promoters (Sandhu et al., 1988).

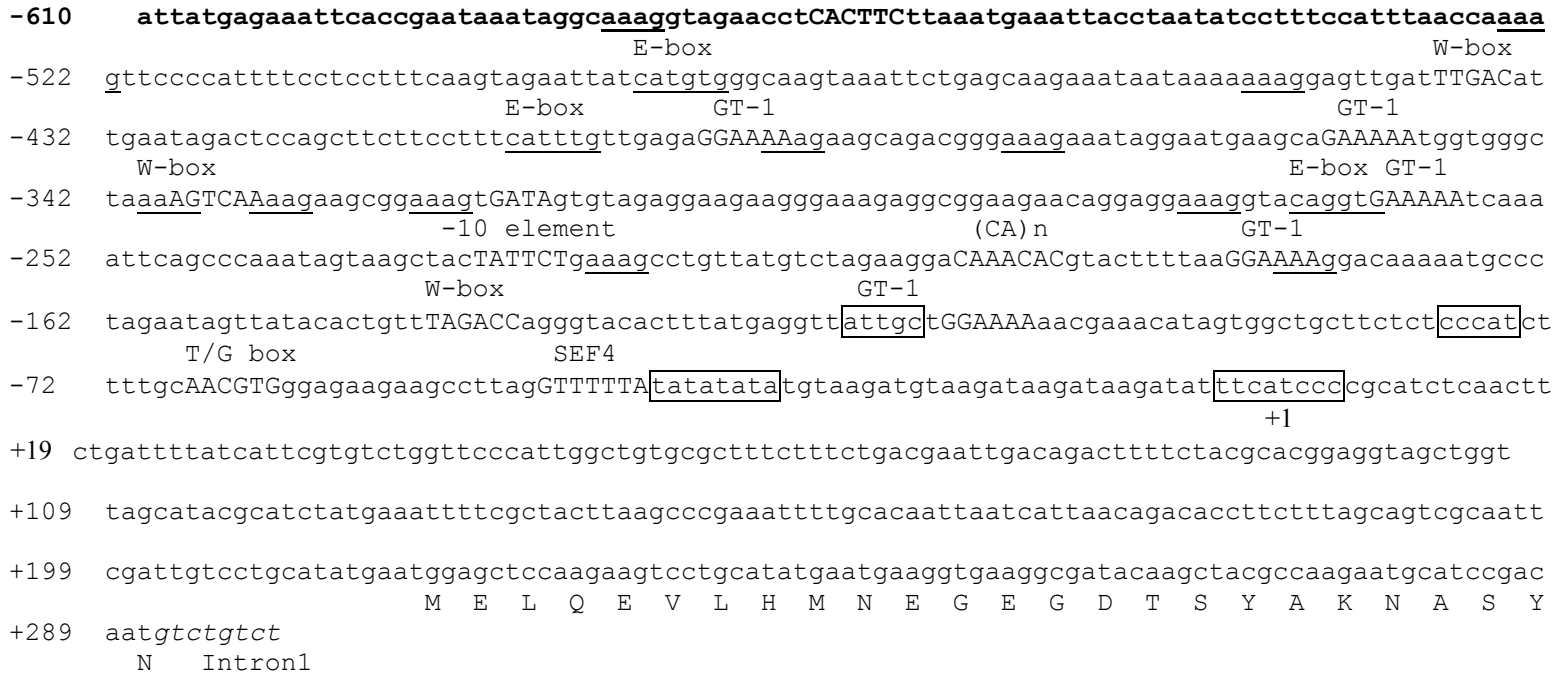


Figure 3.3.6 Putative *cis*-elements located in the promoter sequence of the coffee NMT gene

The TSS and basal promoter elements TATA and CAAT box are boxed. Some of the cis-elements are indicated in capital letters or underlined. The sequence of the introns is presented in italics. The deduced AA sequence is shown below the nucleotide sequence as single letter code. The 'A' nucleotide of the transcriptional start site is assigned as position 1 in the nucleotide sequence, and the nucleotide positions upstream of position 1 are presented with minus numbers.

3.3.3 Histochemical GUS staining

The 774 bp PCR walking product obtained with E*co*RV library contained 745 bp coffee genomic fragment and 29 bases of adapter. The 745 bp fragment spanning the 5′ upstream region, 75 bp of the first exon and 8 bases of the first intron for the NMT gene was used in making promoter-GUS fusion construct pPCTS774 in binary vector pCAMBIA 1381 and introduced into tobacco through *Agrobacterium tumefaciens* mediated transformation. Histochemical staining of GUS activity was performed as described by Jefferson et al., (1987). GUS activity was not detected in the transgenic calli transformed with the vector pCAMBIA 1381(negative control). However, GUS activity was detected in transgenic calli transformed with the vector pCAMBIA 1301, confirming the expression of CaMV 35S promoter. Similarly, GUS activity was detected in transgenic calli transformed with the NMT promoter- GUS fusion construct pPCTS774, confirming the ability of the isolated promoter to drive gene expression (Figure 3.3.7). Plantlets were regenerated from the transgenic callus and the plant tissues were subjected to GUS assay. Gus activity was observed with construct pPCTS774 in the leaves of the transformed tobacco plant (Figure 3.3.8a). GUS activity was also observed in transformed plant with vector pCAMBIA1301 containing CaMV 35S promoter (Figure 3.3.8b). GUS activity was localized in the cytoplasm of tobacco cells (Figure 3.3.8c). This is consistent with the observation of Ogawa et al., (2001) who reported that coffee theobromine synthase (CaMXMT) is present predominantly in the cytoplasm. The sub-cellular localization of coffee theobromine synthase was examined using the fusion protein of CaMXMT and GFP by contrast in onion epidermal cells (Ogawa et al., 2001). The *N*-terminal sequences lack the typical features of transit peptide signals, suggesting that the gene encodes the cytosol form of NMT.

To test the ability of the isolated promoter to drive gene expression in coffee endosperm, the construct pCTS938 was used in transformation of coffee endosperm by electroporation. Transient GUS assays were performed with electroporated coffee endosperm. The GUS assays demonstrated the ability of the NMT promoter fragment to drive gene expression in coffee endosperm (Figure 3.3.9), where highest caffeine content in coffee plants is observed.

Figure 3.3.7 Detection of GUS activity in transformed tobacco calli
(a) pCAMBIA1381, (b) pCAMBIA1301 (c) pPCTS774

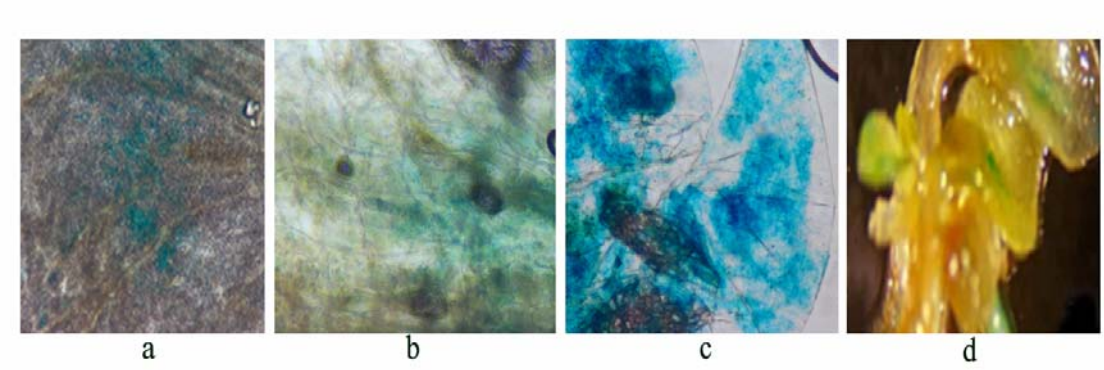


Figure 3.3.8 GUS activity in regenerated transgenic tobacco plants

GUS activity in leaves of tobacco plant transformed with (a) construct pPCTS745, (b) CaMV35S promoter driven cassette, (c) localization of GUS activity in cytoplasm of tobacco cells and (d) regenerated plantlet with construct pPCTS745.

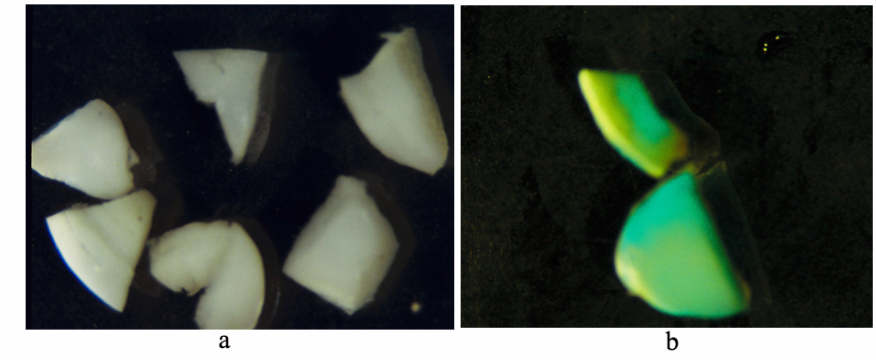


Figure 3.3.9 Transient GUS activity in the electroporated coffee endosperm
(a) Endosperm electroporated with no construct, (b) Construct pPCTS938

However, no GUS activity was observed in the case of electroporated coffee when no construct was used during electroporation. The GUS assays reveal that the isolated promoter could drive gene expression in leaves and endosperm, the two most important sites where very high caffeine biosynthesis is observed in coffee plants.

The cloning of the promoter for the gene involved in caffeine biosynthetic pathway opens up the possibility of studying the molecular mechanisms that regulate the production of caffeine. Grotewold et al., (1998) demonstrated in maize the possibility of using regulatory factors to engineer expression of entire pathways, an approach that requires a high level of knowledge about regulatory mechanisms in the species of interest. The manipulation of trans-acting factors specifically involved in alkaloid formation is an attractive idea, but much remains to be learned before we can control native gene expression patterns in alkaloid biosynthesis. Recently, naturally decaffeinated arabica coffee plants have been reported, where the low caffeine content observed was not due to enhanced degradation of caffeine, but more likely due to possible mutation in caffeine synthase gene (Silvarolla et al., 2004). The regulation for this naturally decaffeinated coffee might be promoter controlled, instead of a mutation in the coding region of the caffeine synthase gene as suggested. Recently, Ogita et al., (2003, 2004) have reported the simultaneous down-regulation of three distinct methylation steps of the caffeine biosynthetic pathway by RNAi, even though 3′ UTR and coding sequence specific for one key enzyme was used. The use of promoter invert repeats may be useful in specific silencing of individual members of NMT multiple gene family in coffee by RNA directed DNA Methylation (Mette et al., 2000), whose promoters may be more divergent than their coding regions.

3.3.4. Amplification of genomic fragment comprising promoter and the gene for NMT

An upstream primer CTSF1 5′- GTTCCCCATTTTCCTCCTTTCAAGTAG -3′ was designed based on the additional sequence obtained through genome walking. The use of this primer along with the reverse primer XMTR1 5′-ACGACAATACCCGAAAGACC-3′ located on unique sequence of 3′ UTR of CaMXMT-1 gene (accession no: AB048794) resulted in amplification of ~2.7 kb genomic fragment containing the promoter and the

coding region. JumpStart KlenTaq LA DNA Polymerase (Sigma, MO, USA) was used in PCR for increased fidelity. However, no amplifications were obtained when CTSF1 primer was used in combination with a primer located on 3′ UTR of CmXRS1 or CaMTL1 (accession nos: AB034699 and AB039725, respectively). This indicates that the amplification of genomic fragment was specific for CaMXMT-1 gene. This also suggests that the promoter regions of the CmXRS1 or CaMTL1 genes may be varying and CTSF1 primer binding site is probably not present or variations exist in that region. The PCR products were cloned and several recombinant plasmids obtained. One of the clones was sequenced and the gene was found to comprise of four exons interrupted by three introns containing A+T rich sequences (Figure 3.3.10), whose junctions agree with the consensus intron/exon borders of plant genes (Brown et al., 1996). The coding region of the clone PG-5 (accession no. DQ010011) comprises 1134 bp and encodes 378 amino acids. Based on the extent of homology of the deduced amino acid sequence and the insertions and deletions present in the cDNA sequences of various NMT genes reported so far, it can be assumed that the genomic clone corresponds to the CaMXMT1 (378 AAs) and possibly encodes putative theobromine synthase-1.

Two genes for theobromine synthase (CaMXMT-1, CaMXMT-2) have high sequence homology (>95%), but there is a deletion of six amino acids in the CaMXMT1, when compared to that of CaMXMT2. The deletion has been used as a signature for assigning the specific name for theobromine synthase gene. The four genes for which the functional activity has been shown vary in the number of amino acids present. The two theobromine synthase genes CaMXMT1 and CaMXMT2 have 378 and 384 AAs respectively, while CaXMT1 and CaDXMT1 have 372 and 384 AAs respectively. The genomic clone described in this study encodes for 378 AAs and not 384 AAs. Based on the insertions and deletions present in the cDNA sequences of various genes published so far, rather than just sequence homology, the specific activity of the enzyme coded by clone PG-5 was assigned as theobromine synthase-1. The promoter region of the clone PG-5 is identical to the sequences of promoter fragments isolated by genome walking. Hence, it can be concluded that these two walking products also correspond to the promoter for theobromine synthase gene-1.

```
-610    ATTATGAGAAATTCACCGAATAAATAGGCAAAGGTAGAACCTCACTTCTTAAATGAAATTACCTAATATCCTTTCCATTTAACCAAAA  pSSPI
-522  GTTCCCCATTTTCCTCCTTTCAAGTAGAATTATCATGTGGGCAAGTAAATTCTGAGCAAGAAATAATAAAAAAAgGAGTTGATTTGACAT  pSSPI
-522  GTTCCCCATTTTCCTCCTTTCAAGTAGAATTATCATGTGGGCAAGTAAATTCTGAGCAAGAAATAATAAAAAAAaGAGTTGATTTGACAT  PG-5

-432  TGAATAGACTCCAGCTTCTTCCTTTCATTTGTTGAGAGGAAAAAGAAgcAGACGGGAAAGAAATAGGAATGAAGCAGAAAAATGGTGGGC  pSSPI
-432  TGAATAGACTCCAGCTTCTTCCTTTCATTTGTTGAGAGGAAAAAGAAatAGACGGGAAAGAAATAGGAATGAAGCAGAAAAATGGTGGGC  PG-5

-342  TAAAAGTCAAAAGAAGCGGAAAGTGATAGTGtAGAGGAAGAAGGGAAAGAGGCGGAAGAACAGGAGGAAAGGTACAGGTGAAAAATCAAA  pSSPI
-342  TAAAAGTCAAAAGAAGCGGAAAGTGATAGTGCAGAGGAAGAAGGGAAAGAGGCGGAAGAACAGGAGGAAAGGTGCAGGTGAAAAATCAAA  PG-5

-252  ATTCAGCCCAAATAGTAAGCTACTATTCTGAAAGCCTGTTATGTCTAGAAGGACAAACACGTACTTTTAAGGAAAAGGACAAAAATGCCC  pSSPI
-252  ATTCAGCCCAAATAGTAAGCTACTATTCTGAAAGCCTGTTATGTCTAGAAGGACAAACACGTACTTTTAAGGAAAAGGACAAAAATGCCC  PG-5

-162  TAGAATAGTTATACACTGTTTAGACCAGGGTACACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTgGCTGCTTCTCTCCCATCT  pSSPI
-162  TAGAATAGTTATACACTGTTTAGACCAGGGTACACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTaGCTGCTTCTCTCCCATCT  PG-5

-72   TTTGCAACGTGGGAGAAGAAGCCTTAGGTTTTTATATATATATGTAAGATGTAAGATAAGATAAGATATTTCATCCCCGCATCTCAACTT  pSSPI
-72   TTTGCAACGTGGGAGAAGAAGCCTTAGGTTTTTATATATATATGTAAGATGTAAGATAAGATAAGATATTTCATCCCCGCATCTCAACTT  PG-5
                                                                                +1
+19   CTGATTTTATCATTCGTGTCTGGTTCCCATTGGCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGCTGGt  pSSPI
+19   CTGATTTTATCATTCGTGTCTGGTTCCCATTGGCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGCTGGC  PG-5

+109  TAGCATACGCATCTATGAAATTTTCGCTACTTAAGCCCGAAATTTTGCACAATTAATCATTAACAGACACCTTCTTTAGCAGTCGCAATT  pSSPI
+109  TAGCATACGCATCTATGAAATTTTCGCTACTTAAGCCCGAAATTTTGCACAATTAATCATTAACAGACACCTTCTTTAGCAGTCGCAATT  PG-5

+199  CGATTGTCCTGCATATGAATGGAGCTCCAAGAAGTCCTGCATATGAATGAAGGTGAAGGCGATACAAGCTACGCCAAGAATGCATCCGAC  pSSPI
+199  CGATTGTCCTGCATATGAATGGAGCTCCAAGAAGTCCTGCATATGAATGAAGGTGAAGGCGATACAAGCTACGCCAAGAATGCATCCTAC  PG-5
                       m  e  l  q  e  v  l  h  m  n  e  g  e  g  d  t  s  y  a  k  n  a  s  y
+289  AATgtctgtct                                                                                 pSSPI
+289  AATgtctgtctgtctctctatctctctttaacacacacacacacagagaatagtggtaaatcatgctatgatacgtcgatctctaacttc  PG-5
       n
+379  acatttgtattttggactggtatgtgtaacagCTGGCTCTTGCCAAGGTGAAACCTTTCCTTGAACAATGCATACGAGAATTGTTGCGGG  PG-5
                                     l  a  l  a  k  v  k  p  f  l  e  q  c  i  r  e  l  l  r
+469  CCAACTTGCCCAACATCAACAAGTGCATTAAAGTTGCGGATTTGGGATGCGCTTCTGGACCAAACACACTTTTAACAGTGCGGGACATTG  PG-5
      a  n  l  p  n  i  n  k  c  i  k  v  a  d  l  g  c  a  s  g  p  n  t  l  l  t  v  r  d  i
+559  TGCAAAGTATTGACAAAGTTGGCCAGGAAGAGAAGAATGAATTAGAACGTCCCACCATTCAGATTTTTCTGAATGATCTTTTCCAAAATG  PG-5
      v  q  s  i  d  k  v  g  q  e  e  k  n  e  l  e  r  p  t  i  q  i  f  l  n  d  l  f  q  n
```

```
+649   ATTTCAATTCGGTTTTCAAGTTGCTGCCAAGCTTCTACCGCAAACTCGAGAAAGAAAATGGACGCAAGATAGGATCGTGCCTAATAAGCG  PG-5
         d  f  n  s  v  f  k  l  l  p  s  f  y  r  k  l  e  k  e  n  g  r  k  i  g  s  c  l  i  s
+739   CAATGCCTGGCTCTTTCTACGGCAGACTCTTCCCCGAGGAGTCCATGCATTTTTTGCACTCTTGTTACAGTGTTCATTGGTTATCTCAGg
        a  m  p  g  s  f  y  g  r  l  f  p  e  e  s  m  h  f  l  h  s  c  y  s  v  h  w  l  s  q
+829   tctttgagttaatccctttatcttttacttttcttgtagcaaaaatggttcgtgattttcattcaacacattagtaactatgcatgg
+919   aaatttctttaataattctaaagatatccacaggaatccaagaaagagatttctgaagaaactaataacatattttatctaagtcgtggc
+1009  tcatgatttacattcccacatgcaacactaacaaaatgatccaactatataagttaccagttctggacgtgcagGTTCCCAGCGGTTTGG
                                                                                  v  p  s  g  l
+1099  TGATTGAATTGGGGATTGGTGCAAACAAAGGGAGTATTTACTCTTCCAAAGGATGTCGTCCGCCCGTCCAGAAGGCATATTTGGATCAAT
        v  i  e  l  g  i  g  a  n  k  g  s  i  y  s  s  k  g  c  r  p  p  v  q  k  a  y  l  d  q
+1189  TTACGAAAGATTTTACCACATTTCTAAGGATTCATTCGAAAGAGTTGTTTTCACGTGGCCGAATGCTCCTTACTTGCATTTGTAAAGTAG
        f  t  k  d  f  t  t  f  l  r  i  h  s  k  e  l  f  s  r  g  r  m  l  l  t  c  i  c  k  v
+1279  ATGAATTCGACGAACCGAATCCCCTAGACTTACTTGACATGGCAATAAACGACTTGATTGTTGAGgttatcatttctctctctctctttg
        d  e  f  d  e  p  n  p  l  d  l  l  d  m  a  i  n  d  l  i  v  e
+1369  atgatcagatgttcattgcttgttatctgaaataaactagatagctaccttagctgattcagggttcctaccttacctaactttgtgtag
+1459  accaagttcaccgtaaatcagatggtagtaacctttttggttaaaaaggtactggaaaaaataatagtatggactaaaagcacaagtaa
+1549  taataatatattcattttgaaattaaggaaaaaatactgtacctttttgtcctagagttgacgttttagcaaattggctaatttccaac
+1639  tctcgtgcttcaaatatctgcagGGACTTCTGGAGGAAGAAAAATTGGATAGTTTCAATATTCCATTCTTTACACCTTCAGCAGAAGAAG
                              g  l  l  e  e  e  k  l  d  s  f  n  i  p  f  f  t  p  s  a  e  e
+1729  TAAAGTGCATAGTTGAGGAGGAAGGTTCTTGCGAAATTTTATATCTGGAGACTTTTAAGGCCCATTATGATGCTGCCTTCTCTATTGATG
        v  k  c  i  v  e  e  e  g  s  c  e  i  l  y  l  e  t  f  k  a  h  y  d  a  a  f  s  i  d
+1819  ATGATTACCCAGTAACATCCCATGAACAAATTAAAGCAGAGTATGTGGCATCATTAATTAGATCAGTTTACGAACCCATCCTCGCAAGTC
        d  d  y  p  v  t  s  h  e  q  i  k  a  e  y  v  a  s  l  i  r  s  v  y  e  p  i  l  a  s
+1909  ATTTTGGAGAAGCTATTATGCCTGACTTATTCCACAGGCTTGCGAAGCATGCAGCAAAGGTTCTCCACATGGGCAAAGGCTGCTATAATA
        h  f  g  e  a  i  m  p  d  l  f  h  r  l  a  k  h  a  a  k  v  l  h  m  g  k  g  c  y  n
+1999  ATCTTATCATTTCTCTCGCCAAAAAGCCAGAGAAGTCAGACGTGTAAAagtttgtttttagttggttttgtgccgttgggggtctttcg  PG-5
        n  l  i  i  s  l  a  k  k  p  e  k  s  d  v  *
+2089  ggtattgtcgt
```

Figure 3.3.10 Nucleotide and deduced amino acid sequences for coffee theobromine synthase-1 gene

PG-5 and pETSSPI represent the nucleotide sequences of the genomic clone for NMT gene and walking product for promoter, respectively. The TSS and basal promoter elements TATA and CAAT box are underlined. The sequence of the introns is presented in lowercase letters. The deduced AA sequence is shown below the nucleotide sequence as single letter code. The translational stop codon is marked with an asterisk. The 'A' nucleotide of the transcriptional start site is assigned as position 1 in the nucleotide sequence, and the nucleotide positions upstream of position 1 are presented with minus numbers.

Multiple sequence alignments of the deduced amino acid sequences for the various NMT genes reported and clone PG-5 is presented in Figure 3.3.11.

```
CaXMT1          MELQEVLRMNGGEGDTSYAKNSAYNQLVLAKVKPVLEQCVRELLRANLPNINKCIKVADL 60
CaMXMT1         MELQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKCIKVADL 59
PG-5            MELQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKCIKVADL 59
CaMXMT2         MELQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKCIKVADL 59
CaDXMT1         MELQEVLHMNGGEGDTSYAKNSFYN-LFLIRVKPILEQCIQELLRANLPNINKCIKVADL 59

CaXMT1          GCASGPNTLLTVRDIVQSIDKVGQEKKNELERPTIQIFLNDLFPNDFNSVFKLLPSFYRK 120
CaMXMT1         GCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRK 119
PG-5            GCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRK 119
CaMXMT2         GCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRK 119
CaDXMT1         GCASGPNTLLTVRDIVQSIDKVGQEKKNELERPTIQIFLNDLFQNDFNSVFKSLPSFYRK 119

CaXMT1          LEKENGRKIGSCLIGAMPGSFYSRLFPEESMHFLHSCYCLQWLSQVPSGLVTELGISTNK 180
CaMXMT1         LEKENGRKIGSCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIELGIGANK 179
PG-5            LEKENGRKIGSCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIELGIGANK 179
CaMXMT2         LEKENGRKIGSCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIELGIGANK 179
CaDXMT1         LEKENGRKIGSCLIGAMPGSFYGRLFPEESMHFLHSCYCLHWLSQVPSGLVTELGISANK 179

CaXMT1          GSIYSSKASRLPVQKAYLDQFTKDFTTFLRIHSEELFSHGRMLLTCICKGVELDARNAID 240
CaMXMT1         GSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICKVDEFDEPNPLD 239
PG-5            GSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICKVDEFDEPNPLD 239
CaMXMT2         GSIYSSKASRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICKVDEYDEPNPLD 239
CaDXMT1         GCIYSSKASRPPIQKAYLDQFTKDFTTFLRIHSEELISRGRMLLTWICKEDEFENPNSID 239

CaXMT1          LLEMAINDLVVEGHLEEEKLDSFNLPVYIPSAEEVKCIVEEEGSFEILYLETFKVLYDAG 300
CaMXMT1         LLDMAINDLIVEGLLEEEKLDSFNIPFFTPSAEEVKCIVEEEGSCEILYLETFKAHYDAA 299
PG-5            LLDMAINDLIVEGLLEEEKLDSFNIPFFTPSAEEVKCIVEEEGSCEILYLETFKAHYDAA 299
CaMXMT2         LLDMAINDLIVEGHLEEEKLASFNLPFFTPSAEEVKCIVEEEGSFEILYLETFKAHYDAG 299
CaDXMT1         LLEMSINDLVIEGHLEEEKLDSFNVPIYAPSTEEVKCIVEEEGSFEILYLETFKVPYDAG 299

CaXMT1          FSIDD------------EHIKAEYVASSVRAVYEPILASHFGEAIIPDIFHRFAKHAAK 347
CaMXMT1         FSIDDDYPVRSH------EQIKAEYVASLIRSVYEPILASHFGEAIMPDLFHRLAKHAAK 353
PG-5            FSIDDDYPVTSH------EQIKAEYVASLIRSVYEPILASHFGEAIMPDLFHRLAKHAAK 353
CaMXMT2         FSIDDDYPVRSHFQVYGDEHIKAEYVASLIRSVYEPILASHFGEAIMPDLFHRLAKHAAK 359
CaDXMT1         FSIDDDYQGRSHSPVSCDEHARAAHVASVVRSIFEPIVASHFGEAIMPDLSHRIAKNAAK 359

CaXMT1          VLPLGKGFYNNLIISLAKKPEKSDV 372
CaMXMT1         VLHMGKGCYNNLIISLAKKPEKSDV 378
PG-5            VLHMGKGCYNNLIISLAKKPEKSDV 378
CaMXMT2         VLHLGKGCYNNLIISLAKKPEKSDV 384
CaDXMT1         VLRSGKGFYDSLIISLAKKPEKSDV 384
```

Figure 3.3.11 Multiple sequence alignments of the deduced amino acid sequences for the various NMT genes

3.3.5 NMT Gene structure similarities with other SAM dependent methyltransferases

BLAST search against the GenBank nucleotide database was performed with the 5′-UTR sequences for the clone PG-5. Sequences producing significant alignments belonged to only *N*-Methyltransferases from *Coffea* species indicating that 5′-UTR sequences are unique to coffee NMTs (Figure 3.3.12).

Sequences producing significant alignments for 5′UTR:		Score (Bits)	E Value
gb|DQ010010.1|	Coffea canephora clone pETSSPI N-m...	428	3e-117
gb|DQ010011.1|	Coffea canephora clone PG-5 N-meth...	420	7e-115
gb|AY907697.1|	Coffea canephora N-methyltransfera...	420	7e-115
dbj|AB034700.1|	Coffea arabica CTS1 mRNA for theo...	99.6	3e-18
dbj|AB054841.1|	Coffea arabica CTS2 mRNA for theo...	71.9	7e-10
gb|AY918125.1|	Coffea canephora clone CX8 putativ...	61.9	7e-07
gb|AY918124.1|	Coffea canephora clone CX10 N-meth...	61.9	7e-07
dbj|AB048794.1|	Coffea arabica CaMXMT1 mRNA for 7...	61.9	7e-07
gb|AY273813.1|	Coffea canephora putative caffeine sy	58.0	1e-05
dbj|AB034699.1|	Coffea arabica CmXRS1 mRNA for 7-...	54.0	2e-04

Figure 3.3.12 Sequences producing significant alignments for UTR sequence of clone PG-5

When the sequence corresponding to *N*- terminal of NMT gene (first 44 amino acids from start codon) was used for protein BLAST, sequences producing significant alignments included several S-adenosyl-L-methionine dependent methyltransferases (Figure 3.3.13).

Sequences producing significant alignments for *N*-terminal region:		Score (Bits)	E Value
gb|AAY56107.1|	N-methyltransferase [Coffea canephora]	90.9	1e-17
gb|AAX07284.1|	N-methyltransferase [Coffea canephora]	90.9	1e-17
gb|AAM18505.1|	N-methyltransferase [Coffea canephora]	90.9	1e-17
dbj|BAC75664.1|	7-methylxanthine N-methyltransferase [Coffea ar	90.9	1e-17
gb|AAQ16155.1|	putative caffeine synthase [Coffea canephora]	90.9	1e-17
dbj|BAC43756.1|	theobromine synthase 1 [Coffea arabica]	90.9	1e-17
gb|AAX07285.1|	putative N-methyltransferase [Coffea canephora]	87.0	2e-16
gb|AAQ94895.1|	putative N-methyltransferase [Coffea liberica var	84.3	1e-15
gb|AAM18503.1|	N-methyltransferase [Coffea arabica]	84.3	1e-15
dbj|BAC43757.1|	theobromine synthse 2 [Coffea arabica]	82.0	5e-15
gb|AAM18507.1|	N-methyltransferase [Coffea liberica]	81.3	9e-15
gb|AAX07286.1|	putative N-methyltransferase [Coffea canephora]	77.4	1e-13
dbj|BAC43760.1|	caffeine synthase 1 [Coffea arabica]	77.4	1e-13
dbj|BAC75663.1|	3,7-dimethylxanthine N-methyltransferase [Coffea	75.5	5e-13
dbj|BAC43761.1|	tentative caffeine synthase 7 [Coffea arabica]	75.5	5e-13
gb|AAQ16154.1|	putative caffeine synthase [Coffea canephora] ...	73.6	2e-12
dbj|BAC75665.1|	Xanthosine N-methyltransferase [Coffea arabica]	73.6	2e-12
gb|AAM18502.1|	N-methyltransferase [Coffea arabica]	73.6	2e-12
gb|AAQ94896.1|	putative N-methyltransferase [Coffea canephora]	73.2	2e-12
dbj|BAC43759.1|	tentative caffeine synthase 4 [Coffea arabica]	72.8	3e-12

dbj|BAB39213.1| caffeine synthase [Coffea arabica]	72.8	3e-12
gb|AAM18510.1| N-methyltransferase [Coffea liberica]	71.6	7e-12
gb|AAM18509.1| N-methyltransferase [Coffea liberica]	71.6	7e-12
gb|AAM18508.1| N-methyltransferase [Coffea liberica]	71.6	7e-12
dbj|BAC43758.1| tentative caffeine synthase 3 [Coffea arabica]	71.6	7e-12
dbj|BAB39214.1| theobromine synthase [Coffea arabica]	71.6	7e-12
gb|AAY56106.1| N-methyltransferase [Coffea canephora]	51.2	1e-05
emb|CAC33768.1| S-adenosyl-L-methionine:salicylic acid carbox...	39.3	0.038
emb|CAI05934.1| S-adenosyl-L-methionine:salicylic acid carbox...	38.9	0.049
gb|AAW66850.1| SAMT [Nicotiana tabacum]	38.9	0.049
gb|AAO45013.1| S-adenosyl-L-methionine:benzoic acid/salicylic...	38.5	0.065
gb|AAO45012.1| S-adenosyl-L-methionine:benzoic acid/salicylic...	38.5	0.065
gb|AAS83524.1| caffeine synthase 1 [Camellia sinensis var. sinen	38.1	0.084
dbj|BAB12278.1| caffeine synthase [Camellia sinensis]	38.1	0.084
dbj|BAB39396.1| S-adenosyl-L-methionine:salicylic acid carbox...	38.1	0.084
gb|ABB02661.1| jasmonic acid carboxyl methyltransferase [Capsicu	37.7	0.11
gb|AAF98406.1| Hypothetical protein [Arabidopsis thaliana]	37.7	0.11
gb|AAG23344.1| S-adenosyl-L-methionine:jasmonic acid carbox	**37.7**	**0.11**
ref|NP_173394.1| S-adenosyl-L-methionine:jasmonic acid carbox...	37.7	0.11
sp|Q9AR07|JMT_ARATH Jasmonate O-methyltransferase	37.7	0.11
dbj|BAD42854.1| caffeine synthase [Camellia sinensis]	37.4	0.14
dbj|BAD33074.1| putative S-adenosyl-L-methionine:jasmonic aci...	37.4	0.14
gb|AAW88351.1| caffeine synthase [Camellia sinensis]	37.4	0.14
gb|AAV52268.1| methyl transferase [Brassica juncea]	37.0	0.19
ref|XP_467504.1| putative S-adenosyl-L-methionine:salicylic a...	37.0	0.19
gb|AAF22289.1| floral nectary-specific protein [Brassica rapa...	37.0	0.19
gb|AAX96678.1| expressed protein [Oryza sativa (japonica cultiva	36.6	0.25
gb|AAX96677.1| expressed protein [Oryza sativa (japonica cultiva	36.6	0.25
gb|ABA92537.1| SAM dependent carboxyl methyltransferase [Oryz.	36.6	0.25
emb|CAF31508.1| S-adenosyl-L-methionine:benzoic acid/salicyli...	35.4	0.55
ref|NP_917090.1| putative S-adenosyl-L-methionine:salicylic a...	35.0	0.71
gb|AAN40745.1| S-adenosyl-L-methionine:salicylic acid methylt...	34.7	0.93
dbj|BAD62438.1| putative S-adenosyl-L-methionine [Oryza sativ...	34.3	1.2
dbj|BAD82223.1| S-adenosyl-L-methionine:jasmonic acid carboxy...	33.9	1.6
emb|CAD70566.1| Carboxyl methyltransferase [Crocus sativus]	33.1	2.7
dbj|BAD45797.1| putative benzothiadiazole-induced S-adenosyl-...	32.7	3.5
ref|NP_917101.1| putative S-adenosyl-L-methionine:salicylic a...	32.3	4.6
gb|ABA92507.1| SAM dependent carboxyl methyltransferase [Ory..	31.6	7.9
gb|AAS18419.1| benzothiadiazole-induced S-adenosyl-L-methioni...	31.6	7.9
dbj|BAD37845.1| putative benzothiadiazole-induced S-adenosyl-...	31.6	7.9
gb|AAF00108.1| S-adenosyl-L-methionine:salicylic acid carboxy...	31.6	7.9

Figure 3.3.13 Sequences producing significant alignments for N- terminal end (44 amino acids from start codon) of clone PG-5

Among the sequences that produced significant alignments with *N* -terminal end of coffee NMT gene, most of the sequences belonged to SAM-dependent Methyltransferases. Several plant *O*-methyltransferases from different species have relatively high sequence homology ranging from 32 to 71% (Joshi and Chiang, 1998). However, the enzymes SAMT (SAM: salicylic acid carboxyl methyltransferase of *Clarkia breweri*), BAMT (SAM: benzoic acid carboxyl methyltransferase of *Antirrhinum majus*), JMT (SAM: jasmonic acid carboxyl methyltransferase of *Arabidopsis thaliana*), NMT (caffeine related SAM: *N* - methyltransferases of *Coffea* species) and IAMT (IAA carboxyl methyltransferase of *Arabidopsis*) all belong to a novel family of methyltransferases identified recently in plants (Zubieta et al., 2003). This family contains motif B' instead of motif B as the conserved region (Mizuno et al., 2003b).

Among the sequences that produced significant alignments with *N* -terminal end of coffee NMT gene, hits that had gene structure similar to that of coffee NMT gene were looked for. However, this was limited only to the hits for which the genomic sequences were available. The accession closest to Coffee NMT gene in relation to the gene structure was that of *Arabidopsis thaliana* S-adenosyl-L-methionine: jasmonic acid carboxyl methyltransferase (JMT) gene (accession no. AY008435, Seo et al., 2001). The location of the first intron was similar to that of coffee NMT gene i.e. after 25 amino acids. The location of the first intron was not conserved in other S-adenosyl-L-methionine dependent methyltransferases for which genomic sequences are available. There was great diversity for the three intron sequences and lengths between coffee NMT and AtJMT genes (Table 3.3.2). However, all the three introns have the consensus dinucleotide GT at the 5´-end and AG at the 3´-end.

Table 3.3.2 Variations in gene structure of NMT and JMT genes

Feature	Coffee NMT (Clone PG-5)	At JMT Seo et al., (2001)	% Sequence similarity
Intron Length			
Intron 1	119	140	21%
Intron 2	255	463	16%
Intron 3	318	738	5%
Intron/Exon splice joint			**Splice joint**
Intron 1	TACAAT**gtc**tgt…taa**cag**CTGGCT	GCTCAG**gtt**gat…tcg**cag**AGCAAC	Different
Intron 2	TCTCAG**gtc**ttt…gtg**cag**GTTCCC	TCTCAG**gtg**tgt…ttt**aag**GTTCCA	Conserved
Intron 3	GTTGAG**gtt**atc…ctg**cag**GGACTT	AAAGAG**gta**tag…ctt**aag**CTTAAG	Conserved

The location of the first intron in JMT was after 25 AAs as in the case of coffee NMT but the adjacent codons, in between which the intron was located, differed in JMT gene (Table 3.3.2). Despite the low intron sequence similarities between the two genes, the intron/ exon splice sites were conserved for the second and third introns (Figure 3.3.14). The exon and open reading frame lengths also varied for the two genes due to several insertions and deletions (indels) within the coding regions (Figure 3.3.14). The variations in intron and exon sequences suggest that the coffee NMT genes have evolved independently or diverged greatly from the related methyltransferases especially from Arabidopsis. Coffee NMT genes may represent the genes that have been evolving at such a rapid rate that they no longer bear significant homologies with proteins from Arabidopsis. Based on the high diversity of *N*-methyltransferases involved in alkaloid biosynthesis in different species, Choi et al., (2002) suggested that each NMT might have evolved from an independent origin.

```
PG-5            MELQEVLHMNEGEGDTSYAKNASYNLALAKV-KPFLEQCIRELLRANLPNINKCIKVADL 59
JMT             MEVMRVLHMNKGNGETSYAKNSTAQSNIISLGRRVMDEALKKLMMSNSE--ISSIGIADL 58
                **: .*****:*:*:******:: :  : .: : .:::.:::*: :*  .  ..* :***

PG-5            GCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRK 119
JMT             GCSSGPNSLLSISNIVDTIHNLCPDLDR--PVPELRVSLNDLPSNDFNYICASLPEFYDR 116
                **:****:**:: :**::*.::  : ...   * ::: **** .**** :   **.** :

PG-5            LEKEN-----GRKIG-SCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIEL 173
JMT             VNNNKEGLGFGRGGGESCFVSAVPGSFYGRLFPRRSLHFVHSSSSLHWLSQVPCREAEKE 176
                :::::     **  * **::**:**********..*:**:**. *:*******.  . :

PG-5            G--IGA---NKGSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK 228
JMT             DRTITADLENMGKIYISKTSPKSAHKAYALQFQTDFLVFLRSRSEELVPGGRMVLSFLGR 236
                .  * *   * *.** ** .  ..:***  ** .** .*** :*:**.. ***:*: : :

PG-5            --VDEFDEP--NPLDLLDMAINDLIVEGLLEEEKLDSFNIPFFTPSAEEVKCIVEEEGSC 284
JMT             RSLDPTTEESCYQWELLAQALMSMAKEGIIEEEKIDAFNAPYYAASSEELKMVIEKEGSF 296
                  :*   *      :**  *: .:  **::****:*:** *::::*:**:* ::*:***

PG-5            EILYLETFKAHYDAAFSIDDDYPVTSHEQIKAEY----VASLIRSVYEPILASHFGEAIM 340
JMT             SIDRLEISPIDWEGGSISEESYDLAIRSKPEALASGRRVSNTIRAVVEPMLEPTFGENVM 356
                .*  **    .::..   ::.* :: :.: :*      *:. **:* **:* . *** :*

PG-5            PDLFHRLAKHAAKVLHMGKGCYNNLIISLAKKPEKSDV 378
JMT             DELFERYAKIVGEYFYVSSPRYAIVILSLVRTG----- 389
                 :**.* ** ..: ::::.  *  :*:**.:. ..:.
```

Figure 3.3.14 Alignments of predicted amino acid sequences for coffee NMT and *Arabidopsis thaliana* JMT genes
Codons in between which the introns are located are highlighted

Seo et al., (2001) suggested that JMT is a cytoplasmic enzyme based on the absence of apparent organ-specific transit signal peptide and hydrophobic regions long enough to be integrated in membranes, and also on the observation that its ortholog NTR1 was located in the cytoplasm of *Brassica* cells by immuno-localization experiments (Song et al., 2000). This observation and the fact that the coffee NMT gene shares significant identity to JMT gene at *N*-terminal end (Figure 3.3.14) strengthen the results obtained in this study regarding the cytoplasmic localization of GUS activity using the promoter for coffee theobromine synthase-1 gene.

3.3.6 Search for orthologous upstream regions

The sequence alignment of promoter sequences for coffee NMT and Arabidopsis JMT did not produce significant alignments, suggesting great variations in the two sequences. However, the function of promoters is by virtue of its modular nature i.e. motifs

arrangement rather than sequence *per se*. Hence, attempts were made to analyze the conservation of motifs in coffee NMT gene with respect to orthologous promoters. Database of Orthologous Promoters (DoOP; http://doop.abc.hu/) (Barta et al., 2005) provides an easy way to obtain clusters of orthologous promoter sequences. After selecting the plant database section, several clusters i.e. set of orthologous promoter sequences were retrieved from the database (Release 1.3) using the keyword S-adenosyl -L- methionine for search. Since coffee NMT genes show similarity to SAM dependent carboxyl methyltransferases, clusters for these methyltransferases were selected to see if the promoter for coffee NMT gene had any orthologous counterpart in terms of arrangements of various conserved motifs. The conserved motifs from each cluster were retrieved and the presence or absence of each motif in promoter sequence for coffee NMT gene was surveyed manually. Of the several conserved motifs present in various clusters, few motifs were present in coffee NMT gene promoter (Table 3.3.3). For most of the motifs that were conserved, one or two nucleotide variations were observed in the coffee NMT promoter, and many of them were present in the opposite strand. However, the functional significance of these motifs remains to be evaluated.

However, it was significant to note that the positions of the various motifs in the AtJMT and coffee NMT had no pattern (Table 3.3.3). This is consistent with the poor nucleotide sequence alignments for promoter regions for the two genes. The results reveal that the modular structure of coffee NMT gene promoter is different from that of related SAM dependent carboxyl methyltransferases. The frequency of detectable conserved elements in the non-coding regions of orthologous genes decreases as species separated by increasing evolutionary distances are compared.

Table 3.3.3 Results for motif search in orthologous promoter sequences using DoOP

S. No	Cluster ID	Species in cluster	Gene / Product	Motifs in cluster of 500 bp (Position in Arabidopsis)	Presence in coffee NMT gene (position in coffee)
1	81019640	*Arabidopsis thaliana*	S-adenosyl-L-methionine: jasmonic acid carboxyl methyltransferase (JMT)	AGAGAGAG, -211	ND
		Populus trichocarpa		TTCATCA, -421	ND
				TTTGTTTTT, -460	TagGTTTTT, -49 TTTcGTTTTT (-), -97
2	81066690	*Arabidopsis thaliana*	S-adenosyl-L-methionine: carboxyl methyl transferase family protein	AGAGATAT, -3	taAGATAT, -12
		Brassica napus		AAACAA, -40	cAACAA (-), -399
		Brassica rapa		CTTTGAA, -498	aTTTGAc, -441 tTTTGAc (-), -331 tTTTGAt (-), -252
3	81066720	*Arabidopsis thaliana*	S-adenosyl-L-methionine: carboxyl methyltransferase family protein	GAGAAAAAA, -24	cAGAAAAAt, -358
		Brassica oleracea		AAGAAAAG, -33	CAGAAAAa, -358
		Brassica rapa		TATCAAAC, -46	AATCAAAa, -259
		Brassica napus		TCATCACA, -74	ND
				CGTCCATAG, -84	ND
				GTTTCA, -188,	GTTTCg (-), -97 aTTTCA (-), -551 cTTTCA, -411, -506

				AAACAT, -245 AAATAATA, -319	AAACAT, -100 AAATAATA, -463
4	81066700	*Arabidopsis thaliana* *Brassica napus* *Brassica oleracea* *Brassica rapa*	S-adenosyl-L-methionine: carboxyl methyl transferase family protein	GAAAAAA, -20 AGATTTTtg, -30 ATCAAACAA, -43 ACACGGAATA, -114	GAAAAAA, -109 gcATTTTTG (-), -165 ND ND
5	81015125	*Arabidopsis thaliana* *Brassica oleracea*	S-adenosyl-L-methionine: carboxyl methyl transferase family protein	CATACAATAC, -21 TTCTCCCAT, -70 ATGTAG, -77, CACTCTACA, -92 TTAGACTGAATC, -114 CACATG, -237 AGGTTGAA, -262	ND cTCTCCCAT, -83 GTGTAG, -326 tcCTCTACA (-), -333 ND CACATG (-), -489 AGGTTctA (-), -569
6	85004380	*Arabidopsis thaliana* *Brassica oleracea*	S-adenosyl-L-methionine: carboxyl methyl transferase family protein	ATTCATC, -51 CACCAATCCTT, -86 AATTTAT, -130 AAGTTTCC, -196 AGATAT, -311	ND ND AATTTAc (-), -472 ND AGATAT, -10

				AAAAAGA, -318	AAAAAGA, -392
7	85066430	*Arabidopsis thaliana* *Brassica rapa*	S-adenosyl-L-methionine: carboxyl methyl transferase family protein	TAAATA, -257	TAAATA, -590
				TTCATCC, -172	
				ATGTTG, -90	ATGTTt, -95
				TGTATATT, -72	TGTATAac (-), -148
				ATGCAG, -60	ND
				GAGAGAAGTAGT, -38	ND
				TTTTGTGA, -18	TTTTGTcc (-), -169
				TGCTCT, -9	TGCTCa (-), -465

Lower case letters in the motif sequence suggest nucleotide variation in sequence
(-) Minus strand, ND- Not detected

It can be concluded that the coffee NMT genes may have evolved independently or diverged greatly from a common ancestor and as a result have no similarities in promoter modular structure and gene structure to related methyltransferases especially from Arabidopsis. Coffee and Arabidopsis belong to different plant families (Rubiaceae and Brassicaceae, respectively) which are distantly related phylogenetically and which diverged from their last common ancestor approximately 94 million years ago (Lin et al., 2005). Coffee NMT genes may represent the genes that have been evolving at such a rapid rate that they no longer bear significant homologies with proteins from Arabidopsis.

3.3.7 Summary and Conclusions

The promoter for one of the coffee NMTs was isolated by PCR based genome-walking method. The sequence analyses revealed that the isolated genomic fragment contained several *cis* - elements that have roles in regulating gene expression. The promoter fragment isolated in this study was capable of gene expression in leaves and endosperms as demonstrated by GUS assays with transformed tissues. The cellular localization of the GUS activity suggests that coffee *N*-Methyltransferases are cytosolic enzymes. The genomic clone obtained in this study possibly encodes for theobromine synthase-1 gene. This gene comprises of four exons interrupted by three introns. Based on the comparisons of the nucleotide sequences and gene structure, it was concluded that coffee NMT gene structure is unique when compared to that of related methyltransferases. Though few conserved motifs were present, there was no sequence similarity in the NMT promoter sequence when compared to that of JMT from Arabidopsis. The coding regions shared similarities but the non-coding regions showed great diversity among NMT and JMT genes. This suggests that coffee NMT genes might have diverged rapidly from the ancestral gene after the Arabidopsis and Coffee families (i.e. Brassicaceae and Rubiaceae) separated in evolutionary lineage. Similar results were also obtained by Gaitan (1998), who concluded that in coffee the exons are similar to those in other plants, but introns, 5′ UTRs and promoters of the genes do not share significant similarities.

In the subsequent chapter polymorphisms for coffee NMT genes are discussed. Intra and inter species conservation of non-coding regions (i.e. introns and promoter regions) and gene structure are dealt in detail.

Chapter IV

Polymorphism for coffee NMT genes

4.0 ABSTRACT

Based on the promoter sequence for theobromine synthase gene-1, primer was designed and the genomic fragment comprising the promoter and its corresponding gene was amplified and cloned. The differences in the restriction pattern among the genomic clones for theobromine synthase gene-1 were studied using PCR-RFLP. Though differences were observed in restriction pattern for the clones, Southern blotting of the restricted PCR products from these clones with probes for promoter as well as the gene suggested that genomic clones are highly homologous and possibly belonged to different alleles/ members of NMT gene family. Sequencing of some of these clones from *Coffea canephora* revealed polymorphism in intron length and sequences, but the promoter regions and the coding regions were highly conserved. It can be concluded that *C. canephora* possesses several alleles/ copies of theobromine synthase gene-1. A fragment of NMT gene from Indian wild species *Psilanthus bengalensis* was also cloned and partially sequenced. The conservation of NMT gene structure was studied by comparing the introns from various genomic clones. The intron/ exon splice joints were highly conserved for NMT genes across the species. The promoter regions were PCR amplified from different *Coffea* species. The sequence analyses of promoter regions from different species revealed a very high degree of conservation. The possible reasons for this high level conservation of promoter sequences within the species and across the species are dealt in this chapter.

4.1 INTRODUCTION

Most genes are members of multigene families, indicative of the prevalence of gene duplication in the origin and formation of diverse gene functions. In *Arabidopsis thaliana*, 65% of all genes belong to gene families, 37.4% of which are members of families composed of at least five genes (Yuan et al., 2002). Multigene families are believed to have arisen from a single ancestral gene by gene duplication via either unequal crossing-over or intra/ inter-chromosomal transposition after chromosome breaks. In addition to these processes, duplication events involving large segments of chromosomes are thought to have further increased the number of genes in what were initially small multigene gene families (Vision et al., 2000).

Homologous genes in distinct species are called orthologs, whereas paralogs refer to homologous genes that are found in the same genome/ gene family; and have been created through gene duplication. True orthologs have usually retained very similar functions in distinct species, whereas this is not necessarily true for paralogs. In many cases, paralogs have acquired different or complementary functions. Duplicated genes can have different expression domains (i.e. the tissues in which both genes are expressed might have changed as well at the time of expression) because of changes in their regulatory elements in the promoter region (Force et al., 1999). Therefore, promoter regions of true orthologs probably contain similar regulatory motifs, which may no longer be true for paralogs. Paralogs within gene families are typically regulated independently and often have quite different expression profiles (Wray et al., 2003). Within gene families some members are often highly expressed, perhaps providing activity at a constitutive level, while other members are lowly expressed, possibly expressed only in specific tissues or under more specific conditions. It is speculated that coffee plants possess multiple sets of enzymes, which are necessary for constitutive production of caffeine in relevant tissues. The multiple NMT enzymes may have distinct affinities toward substrates, thereby catalyzing caffeine biosynthesis at a broad range of available substrate concentrations (Uefuji et al., 2003).

Variations in caffeine content are found within the same tree, with the highest levels in seeds, flowers, and leaves. Variations in purine alkaloids content are also observed in various species. The reasons for this differential accumulation of caffeine in various tissues are not known. Differences in the genes and/ or promoter sequences for *N*-Methyltransferases could be one of the reasons. In this study, attempts were made to look for polymorphisms between coffee species for NMT genes. To study the conservation or polymorphism in regulatory regions for orthologs for *N*-methyltransferase gene, attempts were made to clone the promoter regions of theobromine synthase-1 gene from different species of *Coffea*. The promoter regions were also compared among copies/ alleles for theobromine synthase-1 gene in *C. canephora*.

4.2 MATERIALS AND METHODS

4.2.1 PCR-Restriction Fragment Length Polymorphism (RFLP) of the various clones

PCR amplification and cloning of genomic fragments containing the promoter and the gene was described in section 3.2.2 of chapter III. The insert from eight randomly selected clones was reamplified by PCR using the conditions described in section 2.2.9. XT-Taq PCR system (Bangalore Genie, India) was used in PCR for increased fidelity. Restriction digestion of the PCR products was performed with several enzymes: E*co*RI, X*ho*I, H*in*dIII, E*co*RV, M*bo*I and X*ba*I. Restriction digestion was not performed with the clones since the use of clones for studying polymorphism could create difficulties in analyzing restriction pattern due to vector background.

4.2.2 Southern blot of restricted PCR products

The restricted fragments were transferred to nylon membrane and hybridized to different probes. The probes were prepared for the coding region as well as for the promoter region. The probe corresponding to coding region of partial NMT gene was prepared by labeling the reamplified PCR product from the clone CS2B. The promoter specific probe was prepared by labeling the PCR product obtained using primers CTSF2 (5′-GGAATGAAGCAGAAAAATGGTGGGC-3′) and CTSR3 (5′-GCAAAAGATGGGAGAGAAGCAGCCA-3′), and clone pPCTS774 as template. These two probes were prepared by labeling the PCR products with Psoralen label. The PCR product from clone CX10 corresponding to the full-length gene was labeled with DIG-label. The PCR conditions used were the same as described in Chapter II for coding region and in Chapter III for promoter region.

4.2.2.1 Preparation of probe by labeling PCR product for NMT gene using DIG DNA Labeling and Detection Kit (Boehringer Mannheim, Germany)

4.2.2.1.1 Materials

*Hexanucleotide mix

*dNTP mix

*Klenow enzyme

0.2 M EDTA

4 M LiCl

Ethanol, (absolute and 70%)

TE buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0.

* Supplied along with the kit

4.2.2.1.2 Labeling methodology

A 15 µl aliquot of the purified PCR product was taken into a micro centrifuge tube and denatured by heating for 10 min in boiling water, followed by snap-cooling on ice.

1. The tube containing denatured DNA was centrifuged briefly and the following were added to the tube kept on ice:
   2.0 µl hexanucleotide mix
   2.0 µl dNTP mixture
   1.0 µl Klenow enzyme
2. The contents were mixed, centrifuged briefly and incubated at 37^0C for 20 h.
3. The reaction was stopped by adding 2.0 µl of 0.2 M EDTA.
4. DNA was precipitated by adding 2.5 µl of 4 M LiCl and 75 µl of prechilled ethanol. The sample was kept at -20^0C overnight.
5. The sample was centrifuged at 10, 000 x g for 15 min at 4^0C.
6. The pellet was washed with 100 µl of 70% ethanol and dried.
7. The dried pellet was dissolved in 50 µl of TE buffer and stored at -20^0C till further use.

4.2.2.2. Preparation of Probe by labeling PCR Product using BrightStar Psoralen-Biotin Nonisotopic Labeling Kit (Ambion, USA)

4.2.2.2.1 Labeling methodology

A clean untreated 96-well plate was placed on ice bath. The DNA sample (PCR product) was denatured by heating to 100^0C for 10 min, and then rapidly cooled the mixture on ice. The vial containing the lyophilized Psoralen-Biotin was centrifuged for ~15 seconds at 10,000 rpm. Psoralen-Biotin was reconstituted in 33µl DMF and mixed by pipetting up and down a few times to get the Psoralen-Biotin into solution. One µl of the Psoralen-

Biotin was added to 10 µl of the DNA solution in a microfuge tube, mixed and transferred the sample to a well in a 96 well plate. Alternatively, Psoralen-Biotin can be added to nucleic acid already in the 96 well plate and mixed up with a pipet tip. A 365 nm UV light source was placed directly over the sample on the plate and the sample was irradiated for 45 minutes. The sample was diluted to 100 µl by adding 89 µl of TE Buffer and the mixture was transferred to a clean microfuge tube. 200 µl of water-saturated n-Butanol was added, vortexed and centrifuged for 1 minute 10,000 rpm. The top layer of n-Butanol was pipetted off. This step was repeated and the biotin-labeled nucleic acid was stored at -20^{0}C.

4.2.2.2.2 Materials

Reagents (provided with the Kit):

1. Psoralen-Biotoin
2. Dimethylformamide (DMF)
3. Water- saturated n-Butanol
4. TE buffer

4.2.2.3 Capillary transfer of DNA from agarose gel onto nylon membrane

4.2.2.3.1 Materials

- Transfer apparatus
- 3 MM Whatmann filter paper
- Nylon membrane (Hybond+, Amersham Pharmacia, U.K)
- Hand made filter paper
- Parafilm strips
- Glass rod
- Depurination solution: 0.2 N HCl
- Denaturation solution: 1.5 M NaCl, 0.5 M NaOH
- Neutralization solution: 1.5 M NaCl, 0.5 M Tris-HCl pH 7.0
- 20 x SSC: 3M NaCl, 0.3 M sodium citrate pH 7.2
- Maleic acid buffer: 0.1 M maleic acid, 0.15 M NaCl, adjusted to pH 7.5
- Blocking stock solution (10 X concentration): Blocking reagent 10% (w/v) was dissolved in maleic acid buffer by constantly stirring on a heating block (65^{0}C). The solution was autoclaved and stored at 4^{0}C.

- Blocking solution: The 1 X working solution was prepared by diluting the stock solution 1:10 in maleic acid buffer.
- Hybridization buffer: 5X SSC, 0.1% (w/v) N-lauroylsarcosine, 0.02% (w/v) SDS.
- Post-hybridization washing buffer I: 2XSSC, 0.1% SDS
- Post-hybridization washing buffer II: 0.1% SSC, 0.1% SDS

4.2.2.3.2 Transfer methodology

1. The marker lane was cut from the gel after Ethidium Bromide staining and documentation.
2. The gel was rinsed in deionized water and placed in a glass dish containing 0.2 N HCl solution. The gel was incubated in this solution with mild agitation till the dyes changed their colors.
3. The gel was rinsed with deionized water and incubated in 100 ml of denaturation solution at room temperature for 20 min with mild shaking. The solution was poured off and replaced with fresh denaturation solution and the gel was incubated for further 20 min with mild agitation.
4. The denaturation solution was removed and the gel was rinsed with deionized water and incubated in 100 ml of neutralizing solution at room temperature for 20 min with mild shaking. The solution was poured off and replaced with fresh neutralizing solution and the gel was incubated for further 20 min with mild agitation.
5. The two troughs in the transfer tank were filled with 10 X SSC each. A Whatmann 3 MM filter paper was placed on the platform of the transfer tank. The sides of the filter paper were folded and dipped into the buffer. Glass rod was rolled over the filter paper to remove air bubbles.
6. Six Whatmann 3 MM sheets and nylon membrane were cut to the exact size of the gel. The nylon membrane was rinsed in deionized water and then incubated in 10 X SSC solution for 5 min. Two Whatmann 3 MM sheets were rinsed in 10 X SSC solution and placed in the middle of the platform. A glass rod was rolled on them to remove the air bubbles.

7. The gel was placed on top of the filter paper in an inverted fashion. A glass rod was rolled over the gel to remove air bubbles. The right side of the gel was nicked to serve as an identification mark. Parafilm strips were placed all around the gel to avoid short circuit of buffer during transfer.
8. The nylon membrane with its right side cut was placed over the gel such that the cut side matched with that of the gel. A glass rod was rolled over the membrane to remove air bubbles. The remaining four Whatmann 3 MM filter sheets were soaked in 10 X SSC solution and placed over the membrane. The glass rod was rolled over the filter paper to remove air bubbles.
9. Stacks of hand made filter paper were placed over the Whatmann 3 MM filter paper and a weight of about 500 g was applied over the entire assembly.
10. The transfer was allowed to take place for 10-24 h with intermittent changes of paper towels and transfer buffer.
11. The weight, paper towels and Whatmann 3 MM filter papers were removed. The position of the wells was marked on the membrane with a pencil. The membrane was carefully peeled off the gel and submerged in a solution of 6 X SSC. The membrane was incubated in this solution for 5 min and then laid on a clean Whatmann 3 MM filter and allowed to air-dry.
12. The gel was restained in Ethidium Bromide to check the efficiency of transfer.
13. After the membrane was allowed to dry, it was placed inside a polythene bag. The bag was placed over a UV transilluminator (Photodyne, USA) for 2 min to allow cross-linking of DNA to the membrane.
14. The membrane was stored at 4^0C prior to hybridization.

4.2.2.4 Hybridization and processing of nylon membrane bearing transferred DNA

4.2.2.4.1 Prehybridization

1. The membrane was put in a polythene bag and 15 ml pre-warmed (65^0C) hybridization buffer was added to it.
2. The bag was sealed and incubated overnight at 65^0C with mild agitation.

4.2.2.4.2 Hybridization

1. A 5 µl aliquot of probe for partial NMT gene of coffee was heat-denatured by incubating in boiling water for 5 min, followed by snap cooling on ice.
2. A 2 µl aliquot of denatured probe was added to 2 ml of prewarmed (65^0C) hybridization buffer.
3. The hybridization buffer containing the denatured probe was added to the polythene bag containing the membrane.
4. The polythene bag was sealed and incubated at 65^0C for 6 h with mild agitation in a hybridization oven (GATC-HYBE, Germany).

4.2.2.5 Processing of membrane for colour development

4.2.2.5.1 Post-hybridization washes for membrane treated with DIG-labeled probe

1. The hybridization buffer was discarded and the membrane was washed twice in 50 ml of post hybridization washing buffer I for 5 min at room temperature under mild agitation.
2. The membrane was washed twice in 50 ml of post hybridization washing buffer II for 15 min at 65^0C under mild agitation.

4.2.2.5.1.1 Detection procedure for membrane treated with DIG-labeled probe

1. The membrane was rinsed briefly at room temperature in maleic acid buffer and incubated in 50 ml 1X blocking solution for 30 min.
2. The blocking solution was discarded and the membrane was incubated in 10 ml of antibody solution (1: 5000 Anti-DIG-AP conjugate in 1X blocking solution) at room temperature for 45 min under mild agitation.
3. The membrane was washed twice for 15 min in 50 ml maleic acid buffer and incubated in 20 ml detection buffer for 5 min.
4. The detection buffer was discarded and the membrane was incubated overnight in 10 ml freshly prepared color solution (45 µl NBT and 35 µl X-phosphate in 10 ml detection buffer).
5. The membrane was kept in the dark for color development.
6. The reaction was stopped by washing the membrane for 5 min with 50 ml of deionized water.

7. The results were documented by photography of the wet membrane following which the membrane was air-dried and stored in between filter papers at room temperature.

4.2.2.5.1.2 Solutions

- Anti-DIG antibody-alkaline phosphatase conjugate solution*
- Detection buffer: 0.1 M Tris-HCl, 0.1 M NaCl, 50 mM $MgCl_2$, pH 9.5
- Color-substrate solution: Prepared freshly by adding 45 µl NBT* solution and 35 µl X-phosphate* solution in 10 ml of detection buffer.

  * Supplied with the kit

4.2.2.5.2 Processing of membrane and Detection procedure for membrane treated with Psoralen-Biotin label using Brightstar Biodetect Kit (Ambion, USA)

The membrane was washed for 2 X 5 min at room temperature in approximately 1ml Ambion wash buffer/cm^2/wash, followed by washing for 2 X 5 min in approximately 0.5ml blocking solution / cm^2 membrane/ incubation. Then the membrane was incubated for 30 min in approximately 1ml blocking buffer/ cm^2 membrane. The diluted streptavidin-alkaline phosphate (Strep-AP) was prepared by gently and thoroughly mixing together 10ml blocking buffer and 1µl Strep-AP for 100cm^2-size membrane. The membrane was incubated for 10 -15 min in blocking buffer, using 0.5ml blocking buffer / cm^2 membrane. It was followed by 3 X 5 min washes in 1ml wash buffer/ cm^2 membrane /incubation. The membrane was given 2 X 2 min incubations with 0.5 ml 1X assay buffer/ cm^2 membrane /incubation. 10 ml freshly prepared color solution (45 µl NBT and 35 µl X-phosphate in 10ml detection buffer) was added to the membrane. The membrane was kept in the dark without shaking overnight, for the development of color. The reaction was stopped by washing the membrane for 5 min with 50 ml of deionized water. The results were documented by photography of the wet membrane, following which the membrane was air-dried and stored in between filter papers at room temperature.

4.2.2.5.2.1 Solutions and Reagents

* Blocking reagent

* Alternatively, Ambion wash buffer was used

* Strep Alkaline Phosphatase

Detection buffer: 0.1 M Tris-HCl, 0.1 M NaCl, 50 mM MgCl2, pH 9.5

Color-substrate solution: Prepared freshly by adding 45 µl NBT-solution and 35 µl X-phosphate solution in 10 ml of detection buffer.

* Supplied with the kit

4.2.3 Sequencing and analysis of the genomic clones of the clones

Three positive recombinant plasmids PG-1, 4 and 5 were sequenced at the Dept. of Biochemistry, University of Delhi South Campus, New Delhi and Bangalore Genie, Bangalore. Sequence analysis of the nucleotides was performed using Clustal W.

4.2.4 PCR amplification of theobromine synthase gene from different *Coffea* species

4.2.4.1 Isolation of genomic DNA

The genomic DNA from different species of *Coffea* was isolated using the GenElute Plant Genomic DNA kit (Sigma, St. Louis, USA) following the manufacturer's instructions.

4.2.4.2 PCR amplification

PCR amplification of partial NMT gene from *Psilanthus bengalensis* was carried out as described in the section 2.2.3.1 in Chapter II, using the same primers and the PCR parameters. PCR amplification of genomic fragments comprising the full theobromine synthase-1 gene from various *Coffea* species was performed as described in section 2.2.9 of chapter II. The PCR products were analyzed by 1.2% agarose gel electrophoresis.

4.2.5 PCR amplification and cloning of promoter for NMT gene from different *Coffea* species

4.2.5.1 Designing of oligonucleotides

Based on the sequence of clone pETSSPI (chapter III), following primers were designed to amplify promoter regions from different species of *Coffea*:

CTSF1 (5′-GTTCCCCATTTTCCTCCTTTCAAGTAG-3′)

CTSF2 (5′-GGAATGAAGCAGAAAAATGGTGGGC-3′)

GSP2 (5′-ATGCAGGACTTCTTGGAGCT-3′)

4.2.5.2 PCR amplification and cloning

CTS1/CTSF2 primer was used in combination with GSP2 primer. The conditions for PCR were as follows: 94^0C, 4 min; 94^0C, 30s; 55^0C, 45s; 72^0C, 1 min 15s for 30 cycles. XT-Taq PCR system (Bangalore Genie, India) was used in PCR for increased fidelity. The PCR products were purified using QIAGEN MinElute PCR purification kit (QIAGEN, GmbH, Germany) and the purified PCR products were A-tailed, T/A cloned using InsT/Aclone PCR Product Kit (MBI Fermentas, Lithuania) as described earlier. The recombinant plasmids were selected and sequenced.

4.2.5.3 Sequencing and analysis of promoter sequences from different species

The sequencing of the clones was carried out at Bangalore Genie, Bangalore and the sequences were analyzed by using Clustal W.

4.3 RESULTS AND DISCUSSION

4.3.1 PCR-Restriction Fragment Length Polymorphism (RFLP) of the various NMT clones

PCR amplification using primers located in the coding region as well as untranslated regions of CaMXMT-1 gene gave multiple PCR products of almost similar size, which co-migrate in agarose gel as a single band. The sequence comparisons of four such clones described in chapter II (clones CS2A, CS2B, CX8 and CX10) indicate that the coding regions are similar but differences exist mainly in the size and sequence of introns (Chapter II), which suggests that the four clones belong to different alleles/genes of the NMT family.

To examine whether multiple PCR products were obtained even with a primer located on promoter region, PCR-RFLP (Restriction Fragment Length Polymorphism) was performed on few randomly selected genomic clones comprising the promoter and the gene (cloning was described in chapter III). For increased fidelity of PCR products, XT-5 PCR system (Bangalore Genie, India) with proof reading activity was used. The restriction digestions of the PCR products from the clones indicate differences among them (Figure 4.3.1). Formation of multiple PCR products using a primer set designed on the basis of the CaMXMT-1 sequence has been reported earlier (Uefuji et al., 2003), wherein cDNAs encoding four different proteins were screened among several cDNA PCR products. However, the four cDNAs were screened after sequencing 56 randomly selected cDNA clones. The PCR - RFLP method used in this study offers a simple and rapid means of screening different clones of multiple PCR products, where large-scale sequencing is not feasible.

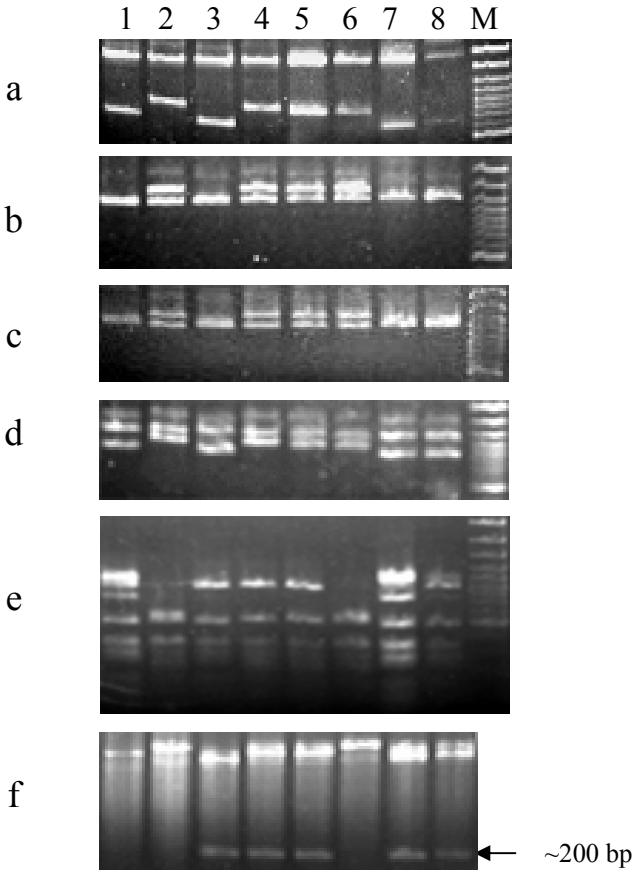


Figure 4.3.1 PCR-RFLP of the genomic clones comprising of promoter and the gene

Based on the restriction map of clone CX10, the following restriction enzymes were used: (a) E*co*RI, (b) X*ho*I, (c) H*in*dIII, (d) E*co*RV, (e) M*bo*I and (f) X*ba*I. Lanes 1-8 represent the randomly selected clones, Lane M contains DNA ladder.

Despite the differences in the restriction pattern among the clones, Southern blotting of the restricted PCR products from these clones with probes for promoter as well as the coding region (Figure 4.3.2), suggests that genomic clones are highly homologous and possibly belong to different alleles/ members of NMT gene family.

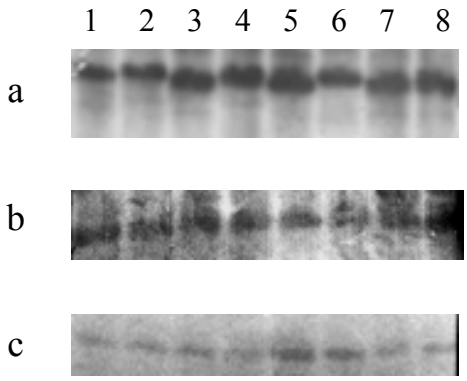


Figure 4.3.2 Southern blots of the restriction digested PCR products from the genomic clones comprising of promoter and the gene

Lanes 1-8 represent the randomly selected clones. (a) X*ba*I digested PCR products from 8 clones were detected with clone CX10 as probe. However, 200bp fragment did not react with probe, confirming that X*ba*I site is present in the upstream region. In (b) and (c), the 5´ upstream region was separated from the coding region by digesting PCR products with S*ac*I and detected with probes. (b) The whole walking product (promoter + first exon) was used as probe. (c) Probe specific for promoter region was used. In all the cases, hybridization was carried out under stringent conditions.

4.3.2 Nucleotide sequence analyses of the various NMT genomic clones

Three genomic clones PG-1, PG-4 and PG-5 were screened based on the PCR-RFLP. Nucleotide sequencing of these genomic clones revealed high degree of similarities in the coding regions as well as non-coding regions, though insertions and deletions (indels) were observed in the non-coding regions (Figure 4.3.3). However, clone PG-4 had one base deletion in the second exon, resulting in premature termination of the coding region (possibly pseudogene). The two clones PG-1 and PG-5 encode for proteins containing 378 amino acid residues, identical to theobromine synthase-1 gene. Therefore, it can be concluded that genomic clones PG-1 and PG-5 isolated from *Coffea canephora* are multiple copies of CaMXMT-1 gene, rather than other members of the NMT gene family. CaMXMT-1 differs from other known members of the NMT gene family primarily in terms of the number of amino acids comprising the open reading frame.

```
-522  GTTCCCCATTTTCCTCCTTTCAAGTAGAATTATCATGTGGGCAAGTAAATTCTGAGCAAGAAATAATAAA------AAAAAGAGTTGATTTGACAT PG-5
-522  GTTCCCCATTTTCCTCCTTTCAAGTAGAATTATCATGTGGGCAAGTAAATTCTGAGCAAGAAATAATAAA------AAAAAGAGTTGATTTGACAT PG-4
-528  GTTCCCCATTTTCCTCCTTTCAAGTAGAATTATCATGTGGGCAAGTAAATTCTGAGCAAGAAATAAAAAATAAAAAAAAAAGAGTTGATTTGACAT PG-1

-432  TGAATAGACTCCAGCTTCTTCCTTTCATTTGTTGAGAGGAAAAAGAAATAGACGGGAAAGAAATAGGAATGAAGCAGAAAAATGGTGGGCTAAAAG PG-5
-432  TGAATAGACTCCAGCTTCTTCCTTTCATTTGTTGAGAGGAAAAAGAAATAGACGGGAAAGAAATAGGAATGAAGCAGAAAAATGGTGGGCTAAAAG PG-4
-432  TGAATAGACTCCAGCTTCTTCCTTTCATTTGTTGAGAGGAAAAGGAAATAGACGGGAAAGAAATAGGAATGAAGCAGAAAAATGGTGGGCTAAAAG PG-1

-336  TCAAAAGAAGCGGAAAGTGATAGTGCAGAGGAAGAAGGGAAAGAGGCGGAAGAACAGGAGGAAAGGTGCAGGTGAAAAATCAAAATTCAGCCCAAA PG-5
-336  TCAAAAGAAGCGGAAAGTGATAGTGCAGAGGAAGAAGGGAAAGAGGCGGAAGAACAGGAGGAAAGGTGCAGGTGAAAAATCAAAATTCAGCCCAAA PG-4
-336  TCAAAAGAAGCGGAAAGTGATAGTGTAGAGGAAGAAGGGAAGGAGGCGGAAGAACAGGAGGAAAGGTGCAGGTGAAAAATCAAAATTCAGCCCAAA PG-1

-240  TAGTAAGCTACTATTCTGAAAGCCTGTTATGTCTAGAAGGACAAACACGTACTTTTAAGGAAAAGGACAAAAATGCCCTAGAATAGTTATACACTG PG-5
-240  TAGTAAGCTACTATTCTGAAAGCCTGTTATGTCTAGAAGGACAAACACGTACTTTTAAGGAAAAGGACAAAAATGCCCTAGAATAGTTATACACTG PG-4
-240  TAGTAAGCTACTATTCTGAAAGCCTGTTATGTCCAGAAGGACAAACACGTACTTTTAAGGAAAAGGACAAAAATGCCCTAGAATAGTTATACACTG PG-1
                                     XbaI
-144  TTTAGACCAGGGTACACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTaGCTGCTTCTCTCCCATCTTTTGCAACGTGGGAGAAGAAGCCT PG-5
-144  TTTAGACCAGGGTACACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTAGCTGCTTCTCTCCCATCTTTTGCAACGTGGGAGAAGAAGCCT PG-4
-144  TTTAGACCAGGGTACACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTAGCTGCTTCTCTCCCATCTTTTGCAACGTGGGGGAAGAAGCCT PG-1

-48   TAGGTTTTTATATATATATGTAAGATGTAAGATAAGATAAGATATTTCATCCCCGCATCTCAACTTCTGATTTTATCATTCGTGTCTGGTTCCCAT PG-5
-48   TAGGTTTTTATATATATATGTAAGATGTAAGATAAGATAAGATATTTCATCCCCGCATCTCAACTTCTGATTTTATCATTCGTGTCTGGTTCCCAT PG-4
-48   TAGGTTTTTATATATATATGTAAGATGTAAGATAAGATAAGATATTTCATCCCCGCATCTCAACTTCTGATTTTATCATTCGTGTCTGGTTCCCAT PG-1
                                                          +1
+49   TGGCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGCTGGCTAGCATACGCATCTATGAAATTTTCGCTACTTAAGC PG-5
+49   TGGCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGCTGGCTAGCATACGCATCTATGAAATTTTCGCTACTTAAGC PG-4
+49   TGGCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGCTGGCTAGCATACGCATCTATGAAATTTTCGCTACTTAAGC PG-1

+144  CCGAAATTTTGCACAATTAATCATTAACAGACACCTTCTTTAGCAGTCGCAATTCGATTGTCCTGCATATGAATGGAGCTCCAAGAAGTCCTGCAT PG-5
+144  CCGAAATTTTGCACAATTAATCATTAACAGACACCTTCTTTAGCAGTCGCAATTCGATTGTCCTGCATATGAATGGAGCTCCAAGAAGTCCTGCAT PG-4
+144  CCGAAATTTTGCACAATTAATCATTAACAGACACCTTCTTTAGCAGTCGCAATTCGATTGTCCTGCATATGAATGGAGCTCCGAGAAGTCCTGCAT PG-1

+240  ATGAATGAAGGTGAAGGCGATACAAGCTACGCCAAGAATGCATCCTACAATgtctgtctgtctctctatctctctttaacacacacacacacagag PG-5
+240  ATGAATGAAGGTGAAGGCGATACAAGCTACGCCAAGAATGCATCCTACAATgtctgtctgtctctctatctctctttaacacacacacacacagag PG-4
+240  ATGAATGAAGGTGAAGGCGATACAAGCTACGCCAAGAATGCATCCTACAATgtctgtctgtctctctatctctctttaacacacacacacacagag PG-1
```

```
+336  aatagtggtaaatcatgctatgatacgtcgatctctaacttcacatttgtattttggactggtatgtgtaacagCTGGCTCTTGCCAAGGTGAAAC PG-5
+336  aatagtggtaaatcatgctatgatacgtcgatctctaacttcacatttgtattttggactggtatgtgtaacagCTGGCTCTTGCCAAGGTGAAAC PG-4
+144  aatagtggtaaatcatgctatgatacgtcgatctctaacttcacatttgtattttggactggtatgtgtaacagCTGGCTCTTGCCAAGGTGAAAC PG-1

+432  CTTTCCTTGAACAATGCATACGAGAATTGTTGCGGGCCAACTTGCCCAACATCAACAAGTGCATTAAAGTTGCGGATTTGGGATGCGCTTCTGGAC PG-5
+432  CTTTCCTTGAACAATGCATACGAGAATTGTTGCGGGCCAACTTGCCCAACATCAACAAGTGCATTAAAGTTGCGGATTTGGGATGCGCTTCTGGAC PG-4
+432  CTTTCCTTGAACAATGCATACGAGAATTGTTGCGGGCCAACTTGCCCAACATCAACAAGTGCATTAAAGTTGCGGATTTGGGATGCGCTTCTGGAC PG-1

+528  CAAACACACTTTTAACAGTGCGGGACATTGTGCAAAGTATTGACAAAGTTGGCCAGGAAGAGAAGAATGAATTAGAACGTCCCACCATTCAGATTT PG-5
+528  CAAACACACTTTTAACAGTTCGGGACATTGTGCAAAGTATTGACAAAGTTGGCCAGGAAGAGAAGAATGAATTAGAACGTCCCACCATTCAGATTT PG-4
+528  CAAACACACTTTTAACAGTTCGGGACATTGTGCAAAGTATTGACAAAGTTGGCCAGGAAGAGAAGAATGAATTAGAACGTCCCACCATTCAGATTT PG-1

+624  TTCTGAATGATCTTTTCCAAAATGATTTCAATTCGGTTTTCAAGTTGCTGCCAAGCTTCTACCGCAAACTCGAGAAAGAAAATGGACGCAAGATAG PG-5
+624  TTCTGAATGATCTTTTCCAAAATGATTTCAATTCGGTTTTCAAGTTGCTGCCAAGCTTCTACCGCAAACTCGAGAAAGAAAATGGACGCAAAATAG PG-4
+624  TTCTGAATGATCTTTTCCAAAATGATTTCAATTCGGTTTTCAAGTTGCTGCCAAGCTTCTACCGCAAACTCGAGAAAGAAAATGGACGCAAGATAG PG-1
                                                  HindIII          XhoI
+720  GATCGTGCCTAATAAGCGCAATGCCTGGCTCTTTCTACGGCAGACTCTTCCCCGAGGAGTCCATGCATTTTTTGCACTCTTGTTACAGTGTTCATT PG-5
+720  GATCGTGCCTAATAAGCGCAATGCCTGGCTCTTTCTACGGCAGACTCTTCCCCGAGGAGTCCATGCATTTTT-GCACTCTTGTTACAGTGTTCATT PG-4
+720  GATCGTGCCTAATAAGCGCAATGCCTGGCTCTTTCTACGGCAGACCCTTCCCCGAGGAGTCCATGCATTTTTTGCACTCTTGTTACAGTGTTCATT PG-1
                                                                    Deletion
+816  GGTTATCTCAGgtctttgagttaatccctttttatcttttacttttttcttgtagcaaaaatggttcgtgattttcattcaacacattagtaactat PG-5
+815  GGTTATCTCAGgtctttgagttaatccctttttatcttttacttttttcttgtagcaaaaatggttcatgattttcattcaacacattagtaactat PG-4
+816  GGTTATCTCAGgtctttgagttaatccctttttatcttttacttttttcttgtagcaaaaatggttcgtgattttcattcaacacattagtaactat PG-1

+912  gcatggaaatttctttaataattctaaagatatccacaggaatccaagaaagagatttctgaagaaactaataacatattttatctaagtcgtggc PG-5
+911  gcatggaaatttctttaataattctaaagatatccacaggaatccaagaaagagatttctgaagaaactaataacatattttatctaagtcgtggc PG-4
+912  gcatggaaatttctttaataattctaaagatatccacaggaatccaagaaagagatttctgaagaaactaataacatattttatccaagtcgtggc PG-1
                                  EcoRV
+1008 tcatgatttacattcccacatgcaacactaacaaaatgatccaactatataagttaccagttctggacgtgcagGTTCCCAGCGGTTTGGTGATTG PG-5
+1007 tcatgatttacattcccacatgcaacactaacaaaatgatccaactatataagttaccagttctggacgtgcagGTTCCCAGCGGTTTGGTGATTG PG-4
+1008 tcatgatttacattcccacatgcaacactaacaaaatgatccaactatataagttaccagttctggacgtgcagGTTCCCAGCGGTTTGGTGATTG PG-1

+1104 AATTGGGGATTGGTGCAAACAAAGGGAGTATTTACTCTTCCAAAGGATGTCGTCCGCCCGTCCAGAAGGCATATTTGGATCAATTTACGAAAGATT PG-5
+1103 AATTGGGGATTGGTGCAAACAAAGGGAGTATTTACTCTTCCAAAGGATGTCGTCCGCCCGTCCAGAAGGCATATTTGGATCAATTTACGAAAGATT PG-4
+1104 AATTGGGGATTGGTGCAAACAAAGGGAGTATTTACTCTTCCAAAGGATGTCGTCCGCCCGTCCAGAAGGCATATTTGGATCAATTTACGAAAGATT PG-1
```

```
+1200 TTACCACATTTCTAAGGATTCATTCGAAAGAGTTGTTTTCACGTGGCCGAATGCTCCTTACTTGCATTTGTAAAGTAGATGAATTCGACGAACCGA PG-5
+1199 TTACCACATTTCTAAGGATTCATTCGAAAGAGTTGTTTTCACGTGGCCGAATGCTCCTTACCTGCATTTGTAAAGTAGATGAATTCGACGAACCGA PG-4
+1200 TTACCACATTTCTAAGGATTCATTCGAAAGAGTTGTTTTCACGTGGCCGAATGCTCCTTACTTGCATTTGTAAAGTAGATGAATTCGACGAACCGA PG-1
                                                                                    EcoRI
+1296 ATCCCCTAGACTTACTTGACATGGCAATAAACGACTTGATTGTTGAGgttatcatttctctctctctttgatgatcagatgttcattgcttgtt PG-5
+1295 ATCCCCTAGACTTACTTGACATGGCAATAAACGACTTGATTGTTGAGgttatcatttccctctctctctttgatgatcagatgttcattgctcgtt PG-4
+1296 ATCCCCTAGACTTACTTGACATGGCAATAAACGACTTGATTGTTGAG---------------------------------------------- PG-1

+1392 atctgaaataaactagatagctaccttagctgattcagggttcctaccttacctaactttgtgtagaccaagttcaccgtaaatcagatggtagta PG-5
+1391 atctgaaataaactagatagctaccttagcttattcagggttcctaccttacctaactttgtgtagaccaagttcaccgtaaatcagatggtagta PG-4
+1343 ---------------------------------------------------------------------------------------------- PG-1

+1488 acctttttggttaaaaaggtactggaaaaaataatagtatggactaaaagcacaagtaataataatatattcattttgaaattaaggaaaaaata PG-5
+1487 acctttttggttaaaaaggtactggaaaaaataatagtatggactaaaagcacaagtaataataatatattcattttgaaattaaggaaaaaa-a PG-4
+1343 ----------------------------------------------------------gtaataatatattcattttgaaattaaggaaaaaata PG-1

+1584 ctgtacctttttgtcctagagttgacgttttagcaaattggctaatttccaactctcgtgcttcaaa---------------------------- PG-5
+1582 ctgtacctttttgtcctagagttgacgttttagcaaattggctaatttccaactctcgtgcttcaaataatttcaattgaattgatgatgattgg PG-4
+1383 ctgtacctttttgtcctagagttgaggttttagcaaattggctaatttccaagtctcgtgctttgaataatttgaattgaattgatga---ttgg PG-1

+1652 --tatctgcagGGACTTCTGGAGGAAGAAAAATTGGATAGTTTCAATATTCCATTCTTTACACCTTCAGCAGAAGAAGTAAAGTGCATAGTTGAGG PG-5
+1678 catatctgcagGGACTTCTGGAGGAAGAAAAATTGGATAGTTTCAATATTCCATTCTTTACACCTTCAGCAGAAGAAGTAAAGTGCATAGTTGAGG PG-4
+1476 catatctgcagGGACTTCTGGAGGAAGAAAAATTGGATAGTTTCAATATTCCATTCTTTACACCTTCAGCAGAAGAAGTAAAGTGCATAGTTGAGG PG-1

+1746 AGGAAGGTTCTTGCGAAATTTTATATCTGGAGACTTTTAAGGCCCATTATGATGCTGCCTTCTCTATTGATGATGATTACCCAGTAACATCCCATG PG-5
+1774 AGGAAGGTTCTTGCGAAATTTTATATCTGGAGACTTTTAAGGCCCATTATGATGCTGCCTTCTCTGTTGATGATGATTACCCAGTAAGATCCCATG PG-4
+1572 AGGAAGGTTCTTGCGAAATTTTAAATCTGGAGACTTTTAAGGCCCATTATGATGCTGCCTTCTCTATTGATGATGATTACCCAGTAAGATCCCATG PG-1

+1842 AACAAATTAAAGCAGAGTATGTGGCATCATTAATTAGATCAGTTTACGAACCCATCCTCGCAAGTCATTTTGGAGAAGCTATTATGCCTGACTTAT PG-5
+1870 AACAAATTAAAGCAGAGTATGTGGCATCATTAATTAGATCAGTTTACGAACCCATCCTCGCAAGTCATTTTGGAGAAGCTATTATGCCTGACTTAT PG-4
+1668 AACAAATTAAAGCAGAGTATGTGGCATCATTAATTAGATCAGTTTACGAACCCATCCTCGCAAGTCATTTTGGAGAAGCTATTATGCCTGACCTAT PG-1

+1938 TCCACAGGCTTGCGAAGCATGCAGCAAAGGTTCTCCACATGGGCAAAGGCTGCTATAATAATCTTATCATTTCTCTCGCCAAAAAGCCAGAGAAGT PG-5
+1966 TCCACAGGCTTGCGAAGCATGCAGCAAAGGTCCTCCACATGGGCAAAGGCTGCTATAATAATCTTATCATTTCTCTCGCCAAAAAGCCAGAGAAGT PG-4
+1764 TCCACAGGCTTGCGAAGCATGCAGCAAAGGTTCTCCACATGGGCAAAGGCTGCTATAATAATCTTATCATTTCTCTCGCCAAAAAGCCAGAGAAGT PG-1
```

```
+2034 CAGACGTGTAAaagtttgtttttagttggtttttgtgccgttgggggtctttcgggtattgtcgt   PG-5
+2062 CAGACGTGTAAaagtttgtttttagttggtttttgtgccgttgggggtctttcgggtattgtcgt   PG-4
+1860 CAGACGTGTAAaagtttgtttttagttggtttttgtgccgttgggggtctttcgggtattgtcgt   PG-1
```

Figure 4.3.3 Nucleotide sequence alignments for the coffee NMT genomic clones

PG-1, 4 and 5 represent the nucleotide sequences of the genomic clones for NMT gene. The TSS and basal promoter elements TATA and CAAT box are underlined. The sequence of the introns is presented in lowercase letters. The translational start and stop codons are underlined. The 'A' nucleotide of the transcriptional start site is assigned as position 1 in the nucleotide sequence, and the nucleotide positions upstream of position 1 are presented with minus numbers. Deletion is represented with '-' gap. Restriction sites that were used for observing polymorphism are also indicated. The X*ba*I site was abolished in clone PG-1 by a single base substitution.

Sequencing revealed that the clone PG-1 indeed differed from PG-4 and PG-5 in restriction site for X*ba*I. X*ba*I digestion for PG-1, 4 and 5 had resulted in different restriction pattern for clone PG-1 when compared to PG-4 and PG-5 (Figure 4.3.1f). The differences in the banding pattern for other enzymes were mainly due to the differences in length of the introns in various clones. The nucleotide sequence alignments revealed high % sequence similarity among the three genomic clones (Figure 4.3.3). However, these clones showed some differences in their intron sequences. This further substantiates the use of PCR- RFLP in this study for screening of genomic clones.

Sequence comparison for exons:

The coding regions (i.e. exon sequences) for the three clones exhibited very high similarity (98%) at the nucleotide level. The predicted amino acid sequences for clones PG-1 and PG-5 also exhibited high similarity (98%) (Figure 4.3.4).

```
PG-5 MELQEVLHMNEGEGDTSYAKNASYNLALAKVKPFLEQCIRELLRANLPNINKCIKVADLG
PG-1 MELREVLHMNEGEGDTSYAKNASYNLALAKVKPFLEQCIRELLRANLPNINKCIKVADLG
     ***:**********************************************************

PG-5 CASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRKL
PG-1 CASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRKL
     ************************************************************

PG-5 EKENGRKIGSCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIELGIGANKG
PG-1 EKENGRKIGSCLISAMPGSFYGRPFPEESMHFLHSCYSVHWLSQVPSGLVIELGIGANKG
     ********************** ***********************************

PG-5 SIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICKVDEFDEPNPLDL
PG-1 SIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICKVDEFDEPNPLDL
     ************************************************************

PG-5 LDMAINDLIVEGLLEEEKLDSFNIPFFTPSAEEVKCIVEEEGSCEILYLETFKAHYDAAF
PG-1 LDMAINDLIVEGLLEEEKLDSFNIPFFTPSAEEVKCIVEEEGSCEILNLETFKAHYDAAF
     ********************************************** ************

PG-5 SIDDDYPVTSHEQIKAEYVASLIRSVYEPILASHFGEAIMPDLFHRLAKHAAKVLHMGKG
PG-1 SIDDDYPVRSHEQIKAEYVASLIRSVYEPILASHFGEAIMPDLFHRLAKHAAKVLHMGKG
     *:****** ****************************************************

PG-5 CYNNLIISLAKKPEKSDV
PG-1 CYNNLIISLAKKPEKSDV
     ******************
```

Figure 4.3.4 Multiple alignments of the predicted amino acid sequences for two genomic clones

Considering the high % of similarities for different *N*- Methyltransferase genes, both at the nucleotide (98%) as well as at predicted amino acid levels (98%), the three genomic clones may be allelic for the theobromine synthase-1 gene of *Coffea canephora*. Since, the coffee NMT genes share very high sequence similarity (>90%) in general, these clones may also represent different copies of the theobromine synthase-1 gene or other members of the gene family. However, due to non-availability of genetic maps for NMT genes, it is difficult to prove the allelic nature of these clones.

Sequence comparison for intron / exon structure:

A very useful feature of gene structure that is often conserved is the intron/exon border region. The conservation of intron/exon splice joint, in addition to similarity of exon sequences, provides a useful guide to genes that share a common evolutionary lineage. The genomic structure of coffee *N*-Methyltransferase genes appears to be highly conserved (Figure 4.3.3 and Table 4.3.2). The position of the three introns within the coding region of the various NMT genes in *Coffea canephora* is identical and appears to be conserved among other *Coffea* species as well i.e. *Coffea liberica* (accession no. AY362825) and *Psilanthus bengalensis*. The cloning and sequencing of partial NMT gene from *P. bengalensis* is presented in section 4.3.3.1 of this chapter.

The nucleotide sequences that comprise intron/exon splice junctions are also conserved in the various NMT genes in *Coffea canephora*, *Coffea liberica* and *P. bengalensis* (Table 4.3.2). All the intron sequences have the consensus dinucleotide GT at the 5′-end and AG at the 3′-end. The introns present in the coffee NMT genes are of phase 0 in nature (defined as the position of intron between codons). It is believed that phase 0 introns are located in more highly conserved portions of genes than phase 1 and 2 introns (position of intron within codons) (Rogozin et al., 2005).

Sequence comparison for introns:

The multiple alignments of the nucleotide sequences for the three genomic clones reveals high similarity in the intron regions also (Figure 4.3.3). The observed conservation in the intron and exon sequences of the three clones suggests that they might have originated as gene duplication event from a common ancestry. The intron sizes for all the NMT gene clones obtained in this study (Table 4.3.1) and those available in NCBI GenBank database were compared and presented in Table 4.3.2.

The introns of the *Arabidopsis* JMT gene were also analyzed. Although there was no sequence homology in the introns between coffee NMT and Arabidopsis JMT (discussed in Chapter III), all the three introns start with dinucleotide GT and end with AG (Table 4.3.2), as is normal for splice sites in plants. The multiple sequence alignments for the three intronic regions for all the genomic sequences were made separately (Figures 4.3.5 - 4.3.7), which revealed an overall high degree of conservation among all the genes within and across species.

Table 4.3.1. Description of NMT genomic clones whose sequences were analyzed in this study

S. No	Clones/ NMT Gene Accession no.	Description of clone	Species	ORF length (Amino acids no.)	Putative function	Reference
1.	CS2A	Partial coding region	*C. canephora*	Partial gene	?	This study
2.	CS2B	Partial coding region	*C. canephora*	Partial gene	?	This study
3.	CX10	Complete coding region	*C. canephora*	378	Theobromine synthase-1	This study
4.	CX8	Complete coding region	*C. canephora*	Pseudogene?		This study
5.	PG-5	Promoter + coding region	*C. canephora*	378	Theobromine synthase-1	This study
6.	PG-1	Promoter + coding region	*C. canephora*	378	Theobromine synthase-1	This study
7.	PG-4	Promoter + coding region	*C. canephora*	Pseudogene?		This study
8.	NMT (AY273814)	Complete coding region	*C. canephora*	378	Theobromine synthase-1	de Kochko, A et al., unpublished
9.	NMT (AY362825)	Complete coding region	*Coffea liberica var. dewevrei*	384	Theobromine synthase-2	de Kochko, A et al., unpublished
10.	JMT (AY008435)		*A. thaliana*	389	Jasmonic acid carboxyl methyltransferase	(Seo et al., 2001)

Table 4.3.2 Intron size and intron / exon border regions in NMT genes

Clone/ accession	Intron 1		Intron 2		Intron 3	
	Intron / exon splice joint	size (bp)	Intron / exon splice joint	size (bp)	Intron/exon splice joint	size (bp)
PG-5	TACAAT**gtc**tgt…taa**cag**CTGGCT	119	TCTCAG**gtc**ttt…g.tg**cag**GTTCCC	255	GTTGAG**gtt**atc…ctg**cag**GGACTT	318
PG-4	TACAAT**gtc**tgt…taa**cag**CTGGCT	119	TCTCAG**gtc**ttt…g.tg**cag**GTTCCC	255	GTTGAG**gtt**atc…ctg**cag**GGACTT	347
PG-1	TACAAT**gtc**tgt…taa**cag**CTGGCT	119	TCTCAG**gtc**ttt…gtg**cag**GTTCCC	255	GTTGAG**gta**ata...ctg**cag**GGACTT	143
CX10	TACAAT**gtc**tgt…taa**cag**CTGGCT	119	TCTCAG**gtc**ttt…g tg**cag**GTTCCC	255	GTTGAG**gtt**atc…ctg**cag**GGACTT	347
CX8	TACAAT**gtc**tgt…taa**cag**CTGGCT	123	TCTCAG**gtc**ttt…g tg**cag**GTTCCC	255	GCTGAG**gtt**atc…ctg**cag**GGACTT	404
CS2B	TACAAT**gtc**tgt…taa**cag**CTGGCT	119	TCTCAG**gtc**ttt…gtg**cag**GTTCCC	255	GTTGAG**gtt**atc…ata**cag**GGACGT	499
CS2A	TACAAT**gtc**tgt…tag**cag**CTGGTT	146	TCTCAG**gtc**ttt…gtg**cag**GTTCCC	231	GTTGAG**gtt**atc…ctt**cag**GGACTT	515
C. canephora AY273814	TACAAT**gtc**tgt…taa**cag**CTGGCT	129	TCTCAG**gtc**ttt…gtg**cag**GTTCCC	256	GTTGAG**gta**ata…ctg**cag**GGACTT	146
C. liberica AY362825	TACAAT**gtc**tgt…taa**cag**CTGGCT	147	TCTCAG**gtc**ttt…ctg**cag**GTTCCC	244	GTTGAG**gtt**atc…ctg**cag**GGACAT	184
P. bengalensis	TACAAT**gtc**tgt…tag**cag**CTGGTT	133	TCTCAG**gtc**ttt…ctg**cag**GTTCCC	170	XXXXXG**gtt**atc…ata**cag**GGACAT	459
A. thaliana JMT AY008435	GCTCAG**gtt**gat…tcg**cag**AGCAAC	140	TCTCAG**gtg**tgt**….**ttt**aag**GTTCCA	463	AAAGAG**gta**tag**…**ctt**aag**CTTAAG	738

XXXXX: Complete sequence not known for *P. bengalensis*

Intron 1

```
CX8             GTCTGTCTGTCTCTCTATCTCTCT--TTAACACACACACACACACACA------GAGAAT 52
CANE            GTCTGTCTGTCTCTCTATCTCTCT--TTAACACACACACACACACACACACACAGAGAAT 58
CX10            GTCTGTCTGTCTCTCTATCTCTCT--TTAACACACACACACACA----------GAGAAT 48
PG-5            GTCTGTCTGTCTCTCTATCTCTCT--TTAACACACACACACACA----------GAGAAT 48
PG-4            GTCTGTCTGTCTCTCTATCTCTCT--TTAACACACACACACACA----------GAGAAT 48
PG-1            GTCTGTCTGTCTCTCTATCTCTCT--TTAACACACACACACACA----------GAGAAT 48
CS2B            GTCTGTCTGTCTCTCTATCTCTCT--TTAACACACACACACACA----------GAGAAT 48
LIB             GTCTGTCTGTCTCTCTATCTCTCT--TTAACACACACACACACACACACA----CAGAGT 54
CS2A            GTCTGTCTGTCTCTCCGTCTCTCTCTTTAACACACACACACACACA--------CAGAGT 52
BENG            GTCTGTCTCTCTCTCTATCTCTCTCTTTAACACACAAAAACACAC--------------- 45
                ******** ******  *******  ********** * *****

CX8             AGTGGTAAATCATGCCATGATACGTCGATCTCTAACTT---------------------- 90
CANE            AGTGGTAAATCATGCTATGATACGTCGATCTCTAACTT---------------------- 96
CX10            AGTGGTAAATCATACTATGATACGTCGATCTCTAACTT---------------------- 86
PG-5            AGTGGTAAATCATGCTATGATACGTCGATCTCTAACTT---------------------- 86
PG-4            AGTGGTAAATCATGCTATGATACGTCGATCTCTAACTT---------------------- 86
PG-1            AGTGGTAAATCATGCTATGATACGTCGATCTCTAACTT---------------------- 86
CS2B            AGTGGTAAATCATGCTATGATACGTCGATCTCTAACTT---------------------- 86
LIB             AGTGGTAAATCATGCTATGATACGTCGATCTCTAACTTAGTATGTCTTTTTTCCCCCCTT 114
CS2A            AGTAGTAAATCATGCTATGATACGTTGATCTCTGACTTAGTATGTCTTTTTTCCCACCTT 112
BENG            ----GTAAATCATGCAATGATACGTCGATCTCTAACTTAGTGAGTCTTTTTACCC--CAT 99
                    ********* * ********* ******* ****

CX8             CACATTTGTACTTTGGACTGGT-ATGTGTAACAG 123
CANE            CACATTTGTATTTTGGACTGGT-ATGTGTAACAG 129
CX10            CACATTTGTATTTTGGACTGGT-ATGTGTAACAG 119
PG-5            CACATTTGTATTTTGGACTGGT-ATGTGTAACAG 119
PG-4            CACATTTGTATTTTGGACTGGT-ATGTGTAACAG 119
PG-1            CACATTTGTATTTTGGACTGGT-ATGTGTAACAG 119
CS2B            CACATTTGTATTTTGGACTGGT-ATGTGTAACAG 119
LIB             CACATTTGTATTTTGGACTGGT-ATGTGTAACAG 147
CS2A            AATATTTGTATTTTGGAGTGGTTATGTGTAGCAG 146
BENG            AACATTTGTTTTTTGGAGTGGTTATCTGTAGCAG 133
                 * ******  ****** **** ** **** ***
```

Figure 4.3.5 Multiple sequence alignments for Intron 1 (CLUSTAL W)
CANE and LIB represent sequences for *C. canephora* (AY273814) and *C. liberica* (AY362825), respectively. All the remaining sequences except BENG are that of various clones obtained in this study from *C. canephora* cv. S-274. BENG represents NMT gene of *Psilanthus bengalensis*.

The size of the first intron in various clones/ genes ranged from 119 to 147 nucleotides. Based on nucleotide alignments (Figure 4.3.5), it can be concluded that intron 1 sequences are highly conserved among the NMT genes. Scores for pairwise comparisons of intron 1 sequences from all genes and clones ranged from 76 to 100% (Table 4.3.3). It can be concluded that intron nucleotide sequence for *Psilanthus bengalensis* varied the most when compared to other clones, as the pairwise % scores for it were the least. However, intron 1 sequence of clone CS2A shares greater similarity with that of *C. liberica* var. *dewevrei* and

Psilanthus bengalensis; having an identical region of insertion indicating possible functional relationship between these three genes.

Table 4.3.3 Percentage sequence identity for Intron 1 among NMT clones (Clustal W)

	PG-5	PG-4	PG-1	CX10	CX8	CS2B	CS2A	CANE	LIB	BENG
PG-5	**100**	100	100	99	98	100	93	100	98	84
PG-4		**100**	100	99	98	100	93	100	98	84
PG-1			**100**	99	98	100	93	100	98	84
CX10				**100**	97	99	92	99	97	83
CX8					**100**	98	91	98	96	79
CS2B						**100**	93	100	98	84
CS2A							**100**	88	91	88
CANE								**100**	96	76
LIB									**100**	86
BENG										**100**

The nucleotide sequence for NMT gene of *Coffea liberica* (accession no. AY362825) encodes for 384 amino acids, possibly corresponding to the theobromine synthase-2 gene (CaMXMT-2, Uefuji et al., 2003). The clone CS2A contains an insertion identical to that of LIB accession. Therefore, it is likely that the clone CS2A corresponds to the theobromine synthase-2 (CaMXMT-2) gene in *C. canephora*. The 22-nucleotide insertion was not observed in other genomic clones from *C. canephora* that encode 378 AA corresponding to theobromine synthase-1 (CaMXMT-1) gene (Figure 4.3.5).

Intron 2

The size of the second intron in various clones / genes ranged from 231 to 256 nucleotides. Scores for pairwise comparisons of intron 2 sequences from all the genes and clones ranged from 81 to 100%, with lower scores associated with *P. bengalensis* (Table 4.3.4). The second intronic sequences showed the highest conservation of all three introns (Figure 4.3.6). Great similarity in the intron sequences of paralogues has been reported previously. Pujade-Renauda et al., (2005) observed 100% sequence conservation in the introns from four hevein genes. The only discernible difference for second intron was 24 and 86 nucleotide deletion in clone CS2A and BENG, respectively.

```
PG-5            GTCTTTGAGTTAATCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATGGTTCGTGAT 60
CANE            GTCTTTGAGTTAATCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATGGTTCGTGAT 60
PG-1            GTCTTTGAGTTAATCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATGGTTCGTGAT 60
CX10            GTCTTTGAGTTAATCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATGGTTCATGAT 60
Cx8             GTCTTTGAGTTAATCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATGGTTCATGAT 60
PG-4            GTCTTTGAGTTAATCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATGGTTCATGAT 60
CS2A            GTCTTTGAGTTAATCCCTTTTATCTTTTTAATTTTTCTTGTAGCAAAAATAGTTCATGAT 60
CS2B            GTCTTTGAGTTAATCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATGGTTCGTGAT 60
LIB             GTCTTTGAGTTAATCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATAGTTCATGAT 60
BENG            GTCTTTGAATTACTCCCTTTTATCTTTTTACTTTTTCTTGTAGCAAAAATAGTTCATGAT 60
                ******** *** ***************** ******************* **** ****

PG-5            TTTCATTCAACACATTAGTAACTATGCATGGAAATTTCTTTAATAATTCTAAAGATATCC 120
CANE            TTTCATTCAACACATTAGTAACTATGCATGGAAATTTCTTTAATAATTCTAAAGATATCC 120
PG-1            TTTCATTCAACACATTAGTAACTATGCATGGAAATTTCTTTAATAATTCTAAAGATATCC 120
CX10            TTTCATTCAACACATTAGTAACTATGCATGGAAATTTCTTTAATAATTCTAAAGATATCC 120
Cx8             TTTCATTCAACACATTAGTAACTATGCATGGAAATTTCTTTAATAATTCTAAAGATATCC 120
PG-4            TTTCATTCAACACATTAGTAACTATGCATGGAAATTTCTTTAATAATTCTAAAGATATCC 120
CS2A            TTTCATTCAACACATTGGTAACTATGCACGGAAATTTCTTTAGCAATTCTAAAGATATCC 120
CS2B            TTTCATTCAACACATTAGTAACTATGCATGGAAATTTCTTTAATAATTCTAAAGATATCC 120
LIB             TTTCATTCAACACATTAGTAACTATGCATGGAAATTTCTTTAATAATTCTAAAGATATCC 120
BENG            TTTCATTCAACACATCAGTAACTATGCATGGAAATTTCTTTAATAATTCTCAAGATATCC 120
                *************** *********** ************* ****** *********

PG-5            ACAGGAATCCAAGAAAGAGATTTCTGAAGA-AACTAATAACATATTTTATCTAAGTCGTG 179
CANE            ACAGGAATCCAAGAAAGAGATTTCTGAAGA-AACTAATAACATATTTTATCTAAGTCGTG 179
PG-1            ACAGGAATCCAAGAAAGAGATTTCTGAAGA-AACTAATAACATATTTTATCCAAGTCGTG 179
CX10            ACAGGAATCCAAGAAAGAGATTTCTGAAGA-AACTAATAACATATTTTATCTAAGTCGTG 179
Cx8             ACAGGAATCCAAGAAAGAGATTTCTGAAGA-AACTAATAACATATTTTATCTAAGTCGTG 179
PG-4            ACAGGAATCCAAGAAAGAGATTTCTGAAGA-AACTAATAACATATTTTATCTAAGTCGTG 179
CS2A            ACAGGAATCCAAGAAAGAGATTTCTGAAGA-AACTAATAACATATTT------------- 166
CS2B            ACAGGAATCCAAGAAAGAGATTTCTGAAGA-AACTAATAACATATTTTATTTAAGTCGTG 179
LIB             ACAGGAATTCAAGAAAGAGATTTCTGAAGA-AACTAATAACATATTTTATTTAAGTCGTG 179
BENG            ACAGGAATCCAAGAAAGAGATTTCTAAAGGGAAC-------------------------- 154
                ******** *************** ***  ***

PG-5            GCTCATGATTTACATTCCCACATGCAACACTAACAAAATGATCCAACTATATAAGT-TAC 238
CANE            GCTCATGATTTACATTCCCACATGCAACACTAACAAAATGATCCAACTATATAAGTCTAC 239
PG-1            GCTCATGATTTACATTCCCACATGCAACACTAACAAAATGATCCAACTATATAAGT-TAC 238
CX10            GCTCATGATTTACATTCCCACATGCAACACTAACAAAATGATCCAACTATATAAGT-TAC 238
Cx8             GCTCATGATTTACATTCCCACATGCAACACTAACAAAATGATCCAACTATATAAGT-TAC 238
PG-4            GCTCATGATTTACATTCCCACATGCAACACTAACAAAATGATCCAACTATATAAGT-TAC 238
CS2A            -----------ATATTCCCACATGCAACACTAACAAAATGATCCAACTATATAAGT-TAC 214
CS2B            GCTCATGATTTACATTCCCACATGCAACACTAACAAAATGATCCGACTATATAAGT-TAC 238
LIB             GCTCATGATTTATATTCCCACATGCAACATTAACAAAATGATCCAACTATATAAGT-TAC 238
BENG            ------------------------------------------------------------

PG-5            CAGTTCTGGACGTGCAG 255
CANE            CAGTTCTGGACGTGCAG 256
PG-1            CAGTTCTGGACGTGCAG 255
CX10            CAGTTCTGGACGTGCAG 255
Cx8             CAGTTCTGGACGTGCAG 255
PG-4            CAGTTCTGGACGTGCAG 255
CS2A            CAGTTCTAGACGTGCAG 231
CS2B            CAGTTCTAGACGTGCAG 255
LIB             -----------CTGCAG 244
BENG            CAGCTTTAGAC-TGCAG 170
```

Figure 4.3.6 Multiple sequence alignments for Intron 2 (CLUSTAL W)

Table 4.3.4 Percentage sequence identity for Intron 2 among NMT clones (Clustal W)

	PG-5	PG-4	PG-1	CX10	CX8	CS2B	CS2A	CANE	LIB	BENG
PG-5	**100**	99	99	99	99	98	96	100	95	83
PG-4		**100**	99	100	100	98	96	99	96	84
PG-1			**100**	99	99	98	96	99	95	83
CX10				**100**	100	98	96	99	96	84
CX8					**100**	98	96	99	96	84
CS2B						**100**	96	98	95	83
CS2A							**100**	96	90	81
CANE								**100**	95	83
LIB									**100**	84
BENG										**100**

Intron 3

The third intron sequences in the NMT genes exhibited the highest polymorphism in sequence and length (ranging from 143 to 515 bases) with several indels spread across the intron (Figure 4.3.7). The most notable observation was 201 nucleotides identical deletion for the clone PG-1 and CANE (accession no: AY273814 belonging to *C. canephora*). As a result of this major deletion and few smaller stretches of deletions, these two genes possessed the shortest introns of 143 and 146 nucleotides, respectively. The two genes code for 378 and 384 AA, respectively and belong to two different paralogs of the coffee NMT gene family. This suggests that in coffee NMT gene family, identical changes in intron sequences are prevalent across the paralogs of gene family. The mechanisms governing the creation of such indels in coffee introns need to be evaluated. The introns of other genes/ clones also differed in size, primarily due to shorter indels. Scores for pairwise comparisons of intron 3 sequences from all genes and clones are presented in Table 4.3.5. The scores varied from 30 to 99%, with lowest values for LIB and BENG. The comparatively lower scores suggest that NMT genes are highly variable in sequence in third intron in particular and towards the 3′ end or 'C' terminal of the gene, in general. This is also reflected in the amino acid sequence variations towards the 3′ end or 'C' terminal of the NMT genes, wherein several indels are present, which account for the differences in the number of AA residues in different NMT genes reported from *Coffea arabica*. It is possible that these variations towards the end of the gene may reflect in the

different substrate specificities exhibited by various members of the gene family, especially when the genes are highly conserved in sequence towards the N-terminal end.

```
CS2B            GTTATCATTTCTCTCTCTCTCTTTGATGATCAGATGTTCATTGCTTGTTATCTGAAATAA 60
BENG            GTTATCATTTCTCTGTCT--CTCTGATAATCAGATGCTCATTGCTTGTTAT--------- 49
CS2A            GTTATCATTTCTCTGTCT--CTTTGATGATCAGATGCTCATTGCTTGTTATCTGAAATAA 58
PG-5            GTTATCATTTCTCTCTCTCTCTTTGATGATCAGATGTTCATTGCTTGTTATCTGAAATAA 60
CX8             GTTATCATTTCTCTCTCTCTCTTTGATGATCAGATGTTCATTGCTCGTTATCTGAAATAA 60
PG-1            ------------------------------------------------------------
CANE            ------------------------------------------------------------
CX10            GTTATCATTTCTCTCTCTCTCTTTGATGATCAGATGTTCATTGCTCGTTATCTGAAATAA 60
PG-4            GTTATCATTTCCCTCTCTCTCTTTGATGATCAGATGTTCATTGCTCGTTATCTGAAATAA 60
LIB             GTTATCATTTCTCTGTCT--CTTTGATGATCAGATGCTCATTGCTTGTTATCTGAAATAA 58

CS2B            ACTAGATAGCTACCTTAGCTGATTCAGGGTTCCTACCTTACCT----------------- 103
BENG            ----------------------TCAGGGTTCCTACCTTAGCTGAAAAATGGATGCAATT 86
CS2A            ACTAGATAGCTAGCT-AGCTTATTCAGGGTTCCTACCTTAGCTGAAAAATGGATGCAATT 117
PG-5            ACTAGATAGCTACCTTAGCTGATTCAGGGTTCCTACCTTACCT----------------- 103
CX8             ACTAGATAGCTACCTTAGCTGATTCAGGGTTCCTACCTTACCT----------------- 103
PG-1            ------------------------------------------------------------
CANE            ------------------------------------------------------------
CX10            ACTAGATAGCTACCTTAGCTTATTCAGGGTTCCTACCTTACCT----------------- 103
PG-4            ACTAGATAGCTACCTTAGCTTATTCAGGGTTCCTACCTTACCT----------------- 103
LIB             ACTAGATAGCTCTTTCATGATCAGACGTTCCTTGGTATCTGAAATAAA------------ 106

CS2B            --------------------------AACTTTGTGTAGACCAAGTTCACCGTAAATCAGA 137
BENG            TTATTCTTTGTAACCTTGATAGACTAAACTTTGTGTAGACGAAGTTCACCGTAAATCAGA 146
CS2A            TTATTCTTTGTATCCTTGATAGAATAAACTTTGTGTAGACGAAGTTCACCGTAACTCAGA 177
PG-5            --------------------------AACTTTGTGTAGACCAAGTTCACCGTAAATCAGA 137
CX8             --------------------------AACTTTGTGTAGACCAAGTTCACCGTAAATCAGA 137
PG-1            ------------------------------------------------------------
CANE            ------------------------------------------------------------
CX10            --------------------------AACTTTGTGTAGACCAAGTTCACCGTAAATCAGA 137
PG-4            --------------------------AACTTTGTGTAGACCAAGTTCACCGTAAATCAGA 137
LIB             ------------------------------------------------------------ 106

CS2B            TGGTAGTAACCTTTAATTAGGAGTTAGGCGTCACAACTCCAATACAGTTAAGATAATTCT 197
BENG            TGGTAGTAACCTTTAATTAGGAGTTAGGACTCACAACTCCAATACAGTTAAGATAATTCT 207
CS2A            TGGCAGTAACCTTTAATTAGGAGTTAGGTCACACAACTCCAATACATTTAATATAATTCT 237
PG-5            TGGTAGTAACCTTT-------------------------------------------- 151
CX8             TGGTAGTAACCTTTAATTAGAAGTTAGGTCACACAACTCCAATGCAGTTAAGATAATTCT 197
PG-1            ------------------------------------------------------------
CANE            ------------------------------------------------------------
CX10            TGGTAGTAACCTTT-------------------------------------------- 204
PG-4            TGGTAGTAACCTTT-------------------------------------------- 204
LIB             ------------------------------------------------------------ 106

CS2B            TTCCCTTCCAATAATTTGTTGGAAGTTAGAAATATGGGGGGATTTAATTTGTCGGCTATC 257
BENG            TTCCCTTTCAATAATTTGTTGGAAGTTAGAAGTGTAGGGGGATTTAATTCGTCGGCTCTC 267
CS2A            TTCCTTTCCAATAATTTGCTGGAAGTTAGAAATATAGGGGAATTAAATCTGTTGGCTATC 297
PG-5            ----------------------------------------TTTGGTTAAAAAGGTACT 169
CX8             TTCAAGTTAGAAATATGTCGGCTATCTTTTTTTCCCC--CTTTTTGGTTAAAAAGGTACT 277
PG-1            ------------------------------------------------------------
CANE            ------------------------------------------------------------
CX10            ---------------------------------------------TTTGGTTAAAAAGGTACT 213
PG-4            ---------------------------------------------TTTGGTTAAAAAGGTACT 213
LIB             ------------------------------------------------------------ 106

CS2B            TTTTTTCCC---CCCTTTTTGGTTAAAAAGGTATTGGAAAGGAAGAAATAGTTCATTATT 314
BENG            TCTTTCTTTTTGCCCTTTTTGGTTAAAAAGGTATTGGAAAGGAAGAAATAGTTCATTATT 327
```

```
CS2A            ATTTTTTC----CCTTTTTTGGTTAAA---------------------TAGTTCATTATT 333
PG-5            ------------GGAAAAAATAATAGTATGGACTAAAAGCACAAGTAATAATAATATATT 218
CX8             ------------GGAAAAAATAATAATATGGACTAAAAGCA------------------- 304
PG-1            ---------------------------------------------GTAATAATAATATATT 17
CANE            ---------------------------------------------GTAATAATAATATATT 17
CX10            ------------GGAAAAAATAATAGTATGGACTAAAAGCACAAGTAATAATAATATATT 218
PG-4            ------------GGAAAAAATAATAGTATGGACTAAAAGCACAAGTAATAATAATATATT 218
LIB             ------------------------------------------------------------ 106

CS2B            TAATCGTTTTT-GGCAAATTTTTGATATGAACTAAAAGCCCAAGTAATAATCATGTATTC 374
BENG            TGATTTTTTTTTGGGGAAATTTTGATATGTAC---------------------------- 359
CS2A            TAATCTTTTTTTGGCAAATTTTTGATATGAACTTGAAGCACAAGTAATAATAATATATTC 393
PG-5            CATTTTGAAATTAAGGAAAAAATACT--GTAC---------------------------- 248
CX8             -ATTTTGAAATTAAGGAAAAAATACT--GTAC---------------------------- 334
PG-1            CATTTTGAAATTAAGGAAAAAATACT--GTAC---------------------------- 47
CANE            CATTTTGAAATTAAGGAAAAAA-ACT--GTAC---------------------------- 46
CX10            CATTTTGAAATTAAGGAAAAAA-ACT--GTAC---------------------------- 247
PG-4            CATTTTGAAATTAAGGAAAAAA-ACT--GTAC---------------------------- 247
LIB             ------------------------------------------------------------ 106

CS2B            ATATTGAAATTAAGGAAAAAAGCTGCACGTTTTATTCTAGAGTTGACGTTTTAGCAAATTG 434
BENG            GTTTTA-------------------------TTTTTCCAGAGTTGACGTTTTAGCAAATTG 394
CS2A            ATATTGAAATTAAGGAAAAAAAACGT-TTTTTTTTTCTAGAGTTGACTTTTTAGCAACTTG 452
PG-5            CTTTTT-------------------------T-GTCCTAGAGTTGACGTTTTAGCAAATTG 282
CX8             CTTTTT-------------------------T-GTCCTAGAGTTGACGTTTTAGCAAATTG 368
PG-1            CTTTTT---------------------------GTCCTAGAGTTGAGGTTTTAGCAAATTG 80
CANE            CTTTTT-------------------------T-GTCCTAGAGTTGACGTTTTAGCAAATTG 80
CX10            CTTTTT-------------------------T-GTCCTAGAGTTGACGTTTTAGCAAATTG 281
PG-4            CTTTTT-------------------------T-GTCCTAGAGTTGACGTTTTAGCAAATTG 281
LIB             ---------------------------------------------CTAGATAGCAAATTG 121
                *                                   *   * *  *     ****** ***

CS2B            GCTGATTTCCAAGTCTTGTACGTCCAGTAATTCGAATTGACTTGATGAT---TGGCATAT 491
BENG            GCTAATTTCCAAGTTTTGTTCGTCCGGTAATTCGAATTGACTTGATGAT---TGGCATAC 451
CS2A            GCTAATTTCCAAGTCTTGTGCTTCAAATAATTTGAATTGAATTGATGAT---TGGCATAT 509
PG-5            GCTAATTTCCAACTCTCGTGCTTCAAATAT------------------------------ 312
CX8             GCTAATTTCCAACTCTCGTGCTTCAAATAT------------------------------ 398
PG-1            GCTAATTTCCAAGTCTCGTGCTTTGAATAATTTGAATTGAATTGATGAT---TGGCATAT 137
CANE            GCTAATTTCCAAGTCTCGTGCTTCAAATAATTTCAATTGAATTGATGATGATTGGCATAT 140
CX10            GCTAATTTCCAACTCTCGTGCTTCAAATAATTTCAATTGAATTGATGATGATTGGCATAT 341
PG-4            GCTAATTTCCAACTCTCGTGCTTCAAATAATTTCAATTGAATTGATGATGATTGGCATAT 341
LIB             GCTAATTTCCAAGTCTCGTGCTTCAAATAATTTGAATTGAATTGATGAT---TGGCATAT 178
                *** ******** * * ** * *    **

CS2B            GAATACAG 499
BENG            GGATACAG 459
CS2A            C--TTCAG 515
PG-5            --CTGCAG 318
CX8             --CTGCAG 404
PG-1            --CTGCAG 143
CANE            --CTGCAG 146
CX10            --CTGCAG 347
PG-4            --CTGCAG 347
LIB             --CTGCAG 184
```

Figure 4.3.7 Multiple sequence alignments for Intron 3 (CLUSTAL W)

Table 4.3.5 Percentage sequence identity for Intron 3 among NMT clones (Clustal W)

	PG-5	PG-4	PG-1	CX10	CX8	CS2B	CS2A	CANE	LIB	BENG
PG-5	**100**	97	73	97	99	47	87	76	43	38
PG-4		**100**	95	99	89	89	89	99	38	38
PG-1			**100**	95	73	81	88	96	49	55
CX10				**100**	89	89	89	99	38	34
CX8					**100**	90	86	76	42	74
CS2B						**100**	84	80	43	84
CS2A							**100**	89	50	83
CANE								**100**	49	56
LIB									**100**	30
BENG										**100**

Sequence comparisons for coding regions among all the NMT genes

The coding regions for all the available NMT genes from coffee species and *Arabidopsis* JMT were compared by performing multiple alignments for predicted amino acid sequences (Figure 4.3.8). The percentage scores for pairwise alignments for the various sequences are presented in Table 4.3.6. Based on the alignments, it can be concluded that members of NMT gene family share a high degree of sequence similarity, with few indels at the 3′ end of the genes. The scores for pairwise alignments were lowest with AtJMT, suggesting that JMT sequences are highly diversed when compared to NMT. The most notable difference towards *N*-terminal portion was the insertion of single codon after 25 AAs, before or after the first intron (i.e. at the intron/ exon splice site (Figure 4.3.8). Additional codon at this position was observed only in AtJMT, though the codon was different. This conservation in gene structure between XMT-1 (enzyme catalyzing first methylation) and AtJMT suggests that these two genes are evolutionarily related. It is logical to conclude that these two genes had a common ancestor before they diverged in sequence and function. XMT-1 or CmXRS-1 seems to be a key enzyme in the caffeine biosynthetic pathway, because it is located at the branching point for caffeine biosynthesis from the general purine nucleoside and nucleotide metabolism (Mizuno et al., 2003a) and hence, this gene must have evolved first among the coffee NMTs. Other NMT genes must have arisen from the gene duplications for this particular gene followed by functional divergence.

```
Cane            MELQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKCIKVADL 59
MXMT1           MELQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKCIKVADL 59
PG-1            MELREVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKCIKVADL 59
PG-5            MELQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKCIKVADL 59
CX10            MELQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKCIKVADL 59
CS2B            --LQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKCIKVADL 57
MXMT2           MELQEVLHMNEGEGDTSYAKNASYN-LALAKVKPFLEQCIRELLRANLPNINKCIKVADL 59
Libe            MELQEVLHMNGGEGDTSYAKNSSYN-LALAKVKPVLEQCIRELLRANLPNINNCIKVADL 59
XMT1            MELQEVLRMNGGEGDTSYAKNSAYNQLVLAKVKPVLEQCVRELLRANLPNINKCIKVADL 60
CS2A            --LQEVLHMNGGEGDTSYAKNSSYN-LVLTKVKPVLEQCIRELLRANLPNINKCIKVADL 57
DXMT1           MELQEVLHMNGGEGDTSYAKNSFYN-LFLIRVKPILEQCIQELLRANLPNINKCIKVADL 59
AtJMT           MEVMRVLHMNKGNGETSYAKNSTAQSNIISLGRRVMDEALKKLMMSNSE--ISSIGIADL 58
                 .: .**:** *:*:******:  :   :   : .:::.:::*: :*  .  ..* :***

Cane            GCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRK 119
MXMT1           GCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRK 119
PG-1            GCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRK 119
PG-5            GCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRK 119
CX10            GCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRK 119
CS2B            GCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRK 117
MXMT2           GCASGPNTLLTVRDIVQSIDKVGQEEKNELERPTIQIFLNDLFQNDFNSVFKLLPSFYRK 119
Libe            GCASGPNTLLTVRDIVQSIDKVGLEEKNELERPTVQIFLNDLFQNDFNSVFKLLPSFYRK 119
XMT1            GCASGPNTLLTVRDIVQSIDKVGQEKKNELERPTIQIFLNDLFPNDFNSVFKLLPSFYRK 120
CS2A            GCASGPNTLLTVRDIVQSIDKVGQEEKNELEHPTIQIFLNDLFQNDFNSVFKLLPSFYRK 117
DXMT1           GCASGPNTLLTVRDIVQSIDKVGQEKKNELERPTIQIFLNDLFQNDFNSVFKSLPSFYRK 119
AtJMT           GCSSGPNSLLSISNIVDTIHNLCPDLDR--PVPELRVSLNDLPSNDFNYICASLPEFYDR 116
                **:****:**:: :**::*.::  : ...   * ::: ****  **** :   **.** :

Cane            LEKEN-----GRKIG-SCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIEL 173
MXMT1           LEKEN-----GRKIG-SCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIEL 173
PG-1            LEKEN-----GRKIG-SCLISAMPGSFYGRPFPEESMHFLHSCYSVHWLSQVPSGLVIEL 173
PG-5            LEKEN-----GRKIG-SCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIEL 173
CX10            LEKEN-----GRKIG-SCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIEL 173
CS2B            LEKEN-----GRKIG-SCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIEL 171
MXMT2           LEKEN-----GRKIG-SCLISAMPGSFYGRLFPEESMHFLHSCYSVHWLSQVPSGLVIEL 173
Libe            LEKEN-----GRKIG-SCLISAMPGSFHGRLFPEESMHFLHSCYSIHWLSQVPSGLVIEL 173
XMT1            LEKEN-----GRKIG-SCLIGAMPGSFYSRLFPEESMHFLHSCYCLQWLSQVPSGLVTEL 174
CS2A            LEKEN-----GRKIG-SCLIWAMPGSFYSRLFPEESMHFLHSCYCLQWLSQVPSGLVTEL 171
DXMT1           LEKEN-----GRKIG-SCLIGAMPGSFYGRLFPEESMHFLHSCYCLHWLSQVPSGLVTEL 173
AtJMT           VNNNKEGLGFGRGGGESCFVSAVPGSFYGRLFPRRSLHFVHSSSSLHWLSQVPCREAEKE 176
                :::::     **  * **:: *:****:.* **..*:**:**. .::******.  . :

Cane            G--IGA---NKGSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK 228
MXMT1           G--IGA---NKGSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK 228
PG-1            G--IGA---NKGSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK 228
PG-5            G--IGA---NKGSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK 228
CX10            G--IGA---NKGSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK 228
CS2B            G--IGA---NKGSIYSSKGCRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK 226
MXMT2           G--IGA---NKGSIYSSKASRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK 228
Libe            G--ISA---NKGSIYSSKASRPPVQKAYLDQFTKDFTTFLRIHSKELFSRGRMLLTCICK 228
XMT1            G--IST---NKGSIYSSKASRLPVQKAYLDQFTKDFTTFLRIHSEELFSHGRMLLTCICK 229
CS2A            G--ISA---NKGIIYSSKASPPPVQKAYLDQFTKDFTTFLRIHSEELLSRGRMLLTCICK 226
DXMT1           G--ISA---NKGCIYSSKASRPPIQKAYLDQFTKDFTTFLRIHSEELISRGRMLLTWICK 228
AtJMT           DRTITADLENMGKIYISKTSPKSAHKAYALQFQTDFLVFLRSRSEELVPGGRMVLSFLGR 236
                .  * :   * * ** ** .  . :***  ** .** .*** :*:**.. ***:*: : :
```

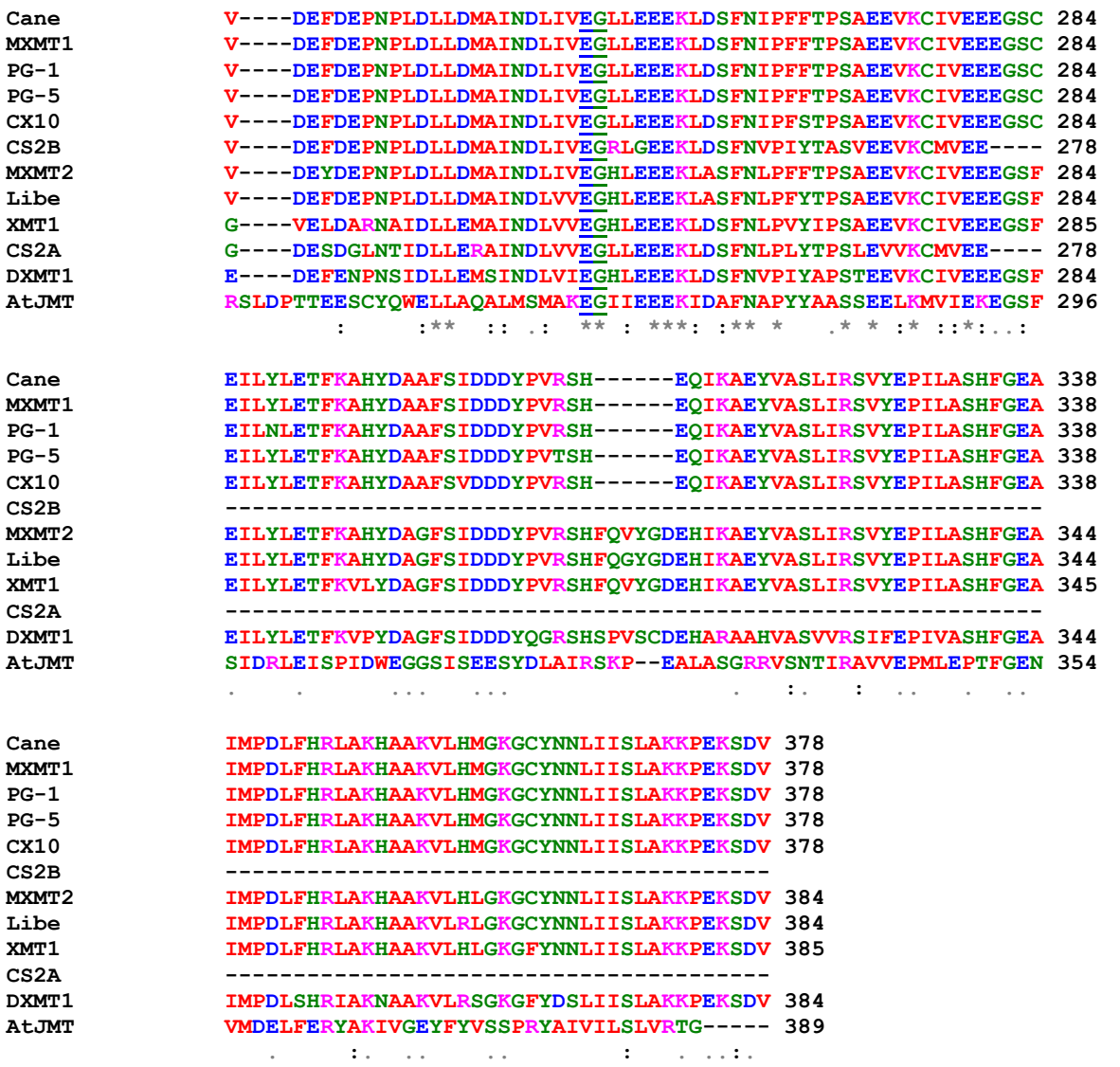


Figure 4.3.8 Multiple sequence alignments of predicted amino acids (CLUSTAL W)
The codons in between which the introns are positioned are underlined

The sequences of two genomic clones CX8 and PG-4 were not used for alignments in Figure 4.3.8, since they contained one base deletion resulting in the premature termination of protein sequences. CX8 and PG-4 may possibly represent pseudogenes. The presence of pseudogenes in multigene families is a common feature. Aubourg et al., reported eight pseudogenes in a family of 40 terpenoid synthase genes in Arabidopsis.

Table 4.3.6 Percentage sequence identity for predicted amino acid sequences among NMT genes (Clustal W)

	XMT-1	MXMT-1	MXMT-2	DXMT-1	PG-5	PG-1	CX10	CS2B	CS2A	CANE	LIB	JMT
XMT-1	**100**	88	90	85	88	87	88	85	90	88	90	37
MXMT-1		**100**	97	83	99	99	99	97	87	100	94	37
MXMT-2			**100**	84	97	96	96	95	88	97	96	36
DXMT-1				**100**	83	82	83	86	88	83	84	38
PG-5					**100**	98	99	97	87	99	94	37
PG-1						**100**	98	96	87	99	93	36
CX10							**100**	97	87	99	94	37
CS2B								**100**	87	97	92	40
CS2A									**100**	87	88	40
CANE										**100**	94	37
LIB											**100**	37
AtJMT												**100**

Sequence comparison for 5′ UTRs:

The 5′-untranslated regions of the three genomic clones were highly conserved (100% sequence identity). However, for other genomic clones, UTR sequences are not available for comparison (Figure 4.3.9).

Sequence comparison for 5′ upstream regulatory regions:

The three genomic clones were remarkably identical to each other in their *5′* upstream regulatory / promoter regions. The only notable difference was a 6 bp addition in the A/T rich region of the promoter in the clone PG-1 at –453 to position (Figure 4.3.9). However, this region could also correspond to a deletion in the clones PG-4 and 5. The significance of this deletion or addition in A/T rich sequence, if any, needs to be evaluated. The search for motifs in promoter regions of these three genomic clones using PLACE and PlantCARE did not reveal any major changes and the motifs were the same in all three clones as described earlier.

```
PG-5    -522    GTTCCCCATTTTCCTCCTTTCAAGTAGAATTATCATGTGGGCAAGTAAATTCTGAGCAAG -462
PG-4    -522    GTTCCCCATTTTCCTCCTTTCAAGTAGAATTATCATGTGGGCAAGTAAATTCTGAGCAAG -462
PG-1    -528    GTTCCCCATTTTCCTCCTTTCAAGTAGAATTATCATGTGGGCAAGTAAATTCTGAGCAAG -468
                ************************************************************

PG-5            AAATAATAAA------AAAAAGAGTTGATTTGACATTGAATAGACTCCAGCTTCTTCCTT -408
PG-4            AAATAATAAA------AAAAAGAGTTGATTTGACATTGAATAGACTCCAGCTTCTTCCTT -408
PG-1            AAATAAAAAATAAAAAAAAAAGAGTTGATTTGACATTGAATAGACTCCAGCTTCTTCCTT -408
                ****** ***      ****************************************

PG-5            TCATTTGTTGAGAGGAAAAAGAAATAGACGGGAAAGAAATAGGAATGAAGCAGAAAAATG -348
PG-4            TCATTTGTTGAGAGGAAAAAGAAATAGACGGGAAAGAAATAGGAATGAAGCAGAAAAATG -348
PG-1            TCATTTGTTGAGAGGAAAAGGAAATAGACGGGAAAGAAATAGGAATGAAGCAGAAAAATG -348
                ****************** ************************************

PG-5            GTGGGCTAAAAGTCAAAAGAAGCGGAAAGTGATAGTGCAGAGGAAGAAGGGAAAGAGGCG -288
PG-4            GTGGGCTAAAAGTCAAAAGAAGCGGAAAGTGATAGTGCAGAGGAAGAAGGGAAAGAGGCG -288
PG-1            GTGGGCTAAAAGTCAAAAGAAGCGGAAAGTGATAGTGTAGAGGAAGAAGGGAAGGAGGCG -288
                ********************************** ************** ******

PG-5            GAAGAACAGGAGGAAAGGTGCAGGTGAAAAATCAAAATTCAGCCCAAATAGTAAGCTACT -228
PG-4            GAAGAACAGGAGGAAAGGTGCAGGTGAAAAATCAAAATTCAGCCCAAATAGTAAGCTACT -228
PG-1            GAAGAACAGGAGGAAAGGTGCAGGTGAAAAATCAAAATTCAGCCCAAATAGTAAGCTACT -228
                ************************************************************

PG-5            ATTCTGAAAGCCTGTTATGTCTAGAAGGACAAACACGTACTTTTAAGGAAAAGGACAAAA -168
PG-4            ATTCTGAAAGCCTGTTATGTCTAGAAGGACAAACACGTACTTTTAAGGAAAAGGACAAAA -168
PG-1            ATTCTGAAAGCCTGTTATGTCCAGAAGGACAAACACGTACTTTTAAGGAAAAGGACAAAA -168
                ******************** ************************************
```

```
PG-5            ATGCCCTAGAATAGTTATACACTGTTTAGACCAGGGTACACTTTATGAGGTTATTGCTGG -108
PG-4            ATGCCCTAGAATAGTTATACACTGTTTAGACCAGGGTACACTTTATGAGGTTATTGCTGG -108
PG-1            ATGCCCTAGAATAGTTATACACTGTTTAGACCAGGGTACACTTTATGAGGTTATTGCTGG -108
                ************************************************************

PG-5            AAAAAACGAAACATAGTAGCTGCTTCTCTCCCATCTTTTGCAACGTGGGAGAAGAAGCCT -48
PG-4            AAAAAACGAAACATAGTAGCTGCTTCTCTCCCATCTTTTGCAACGTGGGAGAAGAAGCCT -48
PG-1            AAAAAACGAAACATAGTAGCTGCTTCTCTCCCATCTTTTGCAACGTGGGGGAAGAAGCCT -48
                ********************************************** **********
                                                             +1
PG-5            TAGGTTTTTATATATATATGTAAGATGTAAGATAAGATAAGATATTTCATCCCCGCATCT +12
PG-4            TAGGTTTTTATATATATATGTAAGATGTAAGATAAGATAAGATATTTCATCCCCGCATCT +12
PG-1            TAGGTTTTTATATATATATGTAAGATGTAAGATAAGATAAGATATTTCATCCCCGCATCT +12
                ************************************************************

PG-5            CAACTTCTGATTTTATCATTCGTGTCTGGTTCCCATTGGCTGTGCGCTTTCTTTCTGACG +72
PG-4            CAACTTCTGATTTTATCATTCGTGTCTGGTTCCCATTGGCTGTGCGCTTTCTTTCTGACG +72
PG-1            CAACTTCTGATTTTATCATTCGTGTCTGGTTCCCATTGGCTGTGCGCTTTCTTTCTGACG +72
                ************************************************************

PG-5            AATTGACAGACTTTTCTACGCACGGAGGTAGCTGGCTAGCATACGCATCTATGAAATTTT +132
PG-4            AATTGACAGACTTTTCTACGCACGGAGGTAGCTGGCTAGCATACGCATCTATGAAATTTT +132
PG-1            AATTGACAGACTTTTCTACGCACGGAGGTAGCTGGCTAGCATACGCATCTATGAAATTTT +132
                ************************************************************

PG-5            CGCTACTTAAGCCCGAAATTTTGCACAATTAATCATTAACAGACACCTTCTTTAGCAGTC +192
PG-4            CGCTACTTAAGCCCGAAATTTTGCACAATTAATCATTAACAGACACCTTCTTTAGCAGTC +192
PG-1            CGCTACTTAAGCCCGAAATTTTGCACAATTAATCATTAACAGACACCTTCTTTAGCAGTC +192
                ************************************************************

PG-5            GCAATTCGATTGTCCTGCATATGA +216
PG-4            GCAATTCGATTGTCCTGCATATGA +216
PG-1            GCAATTCGATTGTCCTGCATATGA +216
                ***********************
```

Figure 4.3.9 Multiple sequence alignments of 5′ upstream regions (CLUSTAL W)
The nucleotide sequences upstream of transcriptional start site (+1) are presented with minus numbers.

It may be concluded that gene duplication most likely is not restricted to the coding or transcribed portion of the genes but also comprises the respective promoters. Haberer et al., (2004) demonstrated that duplicated genes could share *cis*- regulatory elements. Gene divergence within a gene family typically occurs at a much slower rate than gene duplication (Tavares et al., 2000). Recent duplications within a gene family can be recognized by nucleotide comparisons. In non-coding regions of the genes, the absence of selective pressure allows rapid and free sequence divergence after duplication. High identity values are particularly significant between the non-coding regions of the genes in recently duplicated genes. In this study, high similarity in the nucleotide sequences of non-coding regions was observed. High similarities in non-coding regions were also observed

by Riechers and Timko (1999) and Tavares et al., (2000), which were attributed to recent gene duplication events. The high sequence conservation in coding as well as in non-coding regions of the three genomic clones for NMT genes suggest recent gene duplication event in *C. canephora.* It may be concluded that NMT gene family might have originated due to recent gene duplication events. Lashermes et al., (1999) and Campa et al., (2005) have also proposed the recent origin of genus *Coffea*. It has been shown that recently duplicated promoter pairs share a significant similarity within their regulatory regions (Haberer et al., 2004). The degeneration within regulatory regions is a continuous and ongoing process in contrast to degeneration within coding regions. Gene duplication followed by sequence divergence seems to be a model to explain how novel biochemical functions may have arisen for different duplicated genes of Coffee NMTs. It would be interesting to know the nature of duplication of NMT genes in coffee i.e. tandem or segmental duplications. The physical mapping of coffee NMT genes would be useful in determining this and also exploring the possibility of co-regulation of the different NMT genes.

At the gene level, there are several potential functional fates of duplicated genes: one gene copy can substitute for another (functional redundancy), combined activities of both genes may be required for the original function (subfunctionalization), or gain of novel function (neofunctionalization) or loss of function (pseudogene formation) (Lawton-Rauh, 2003). Evolutionary theory predicts that the majority of newly duplicated functional gene copies should lose function quickly, whereas a smaller proportion should retain the same function or acquire a new function through sequence divergence (Mitchell-Olds and Clauss, 2002). It is possible that clones CX8 and PG-4 probably represent pseudogenes that have lost function due to a single nucleotide deletion in fourth and third exon respectively. Similarly, Aubourg et al., 2002 reported the presence of eight pseudogenes in a family of 40 terpenoid synthase genes in *Arabidopsis thaliana*. In view of the high similarity in the coding regions as well as the non-coding regions, it may be concluded that coffee NMT gene family arose from recent gene duplications. It is difficult to distinguish between polymorphic alleles and highly similar paralogs. *C. canephora* is known to be a highly heterozygous and polymorphic species. This high diversity could be explained by

its reproductive allogamy and the absence of strong genetic bottlenecks (Cros et al., 1995). High polymorphism in *Coffea canephora* has been observed in several studies. 77% of RFLP loci studied by Dussert et al ., (1999) were polymorphic with five alleles per locus on average. Recently, high polymorphism in *C. canephora* was also observed by Baruah et al., (2003); Poncet et al., (2004). There is also high polymorphism in biochemical compounds (Ky et al., 2001). Therefore the three genomic clones PG1, 4 and 5 described in this study could also be allelic. Lin et al., (2005) too suggested the possibility that some of the *C. canephora* EST-derived unigenes might actually represent different allelic forms of the same gene. The authors proposed a threshold % sequence identity for distinguishing alleles from true orthologs: higher identity (over 99%) to distinguish allelic ESTs and lower identity (over 91%) to distinguish true paralogs. However, validity of these thresholds for genomic sequences needs to be evaluated, especially in the case of clone PG-1, which has 201 bases deletion in intron. The inclusion or exclusion of this deletion in sequence comparison alters the % identities. However, the allelic nature of these theobromine synthase-1 clones needs to be confirmed by mapping them.

Recently, allele-specific expression has been described in plants. Guo et al., (2003) found that the level of allelic expression in the endosperm, which is a triploid tissue (2n maternal + 1n paternal), was dosage dependent. The possibility of this dosage dependent allelic expression cannot be ignored in coffee as the coffee endosperm contains highest caffeine content among plant tissues. In hybrid maize seedling, of the 15 genes analyzed, almost 75% showed differential allelic expression (Guo et al., 2004). Pyrosequencing for expression analysis at the allele level is a powerful combination for quantitative analysis of allele-specific gene expression. Recently, Schaart et al., (2005) used pyrosequencing to analyze allele-specific expression in octoploid strawberry (Fragaria x ananassa). They reported different alleles of a single gene, which expressed differentially in different tissues. The different alleles therefore appear to be controlled by different regulatory mechanisms in different tissues. The method described (Schaart et al., 2005) is generally applicable for determination of frequencies of any polymorphism that discriminates allelic sequences and is particularly suitable for expression analysis of duplicated gene sequences like paralogous loci and homoeologs in hybrid polyploids. Pyrosequencing technique can

be a viable option for determining the possible differential expression of alleles for coffee NMT gene.

The fact that multiple copies exist for theobromine synthase-1 suggests that it is tightly regulated and that this regulation may be in an organ or stimulus dependent manner. It is known that different genes encoding similar enzymes may be regulated in an organ or tissue specific manner and may respond to different external signals. It is anticipated that, although members of NMT subfamily may undertake the same biochemical function, they may be expressed in different cells, tissues, or organs. It is also reasonable to expect, especially with multiple copies of recently duplicated genes, that genes could exhibit true functional redundancy.

Uefuji et al., (2003) reported that CaDXMT1 had an extremely high K_m value for three acceptable substrates (7-methylxanthine, theobromine, and paraxanthine), indicating that caffeine may not be synthesized before sufficient amounts of precursors have accumulated. The substrate affinity of the concerned enzyme proportionally decreased toward the end point, indicating that further down the pathway, more amount of substrate compound is required, making the reaction proceed irreversibly and stepwise. The presence of multiple copies of theobromine synthase-1 gene in coffee plant may be speculated to result in increased production of theobromine as substrate, which will make the reaction proceed irreversibly. Moreover, the K_m values for enzymes involved in caffeine biosynthesis varies (Mizuno et al., 2003b) depending upon protein species, and these variations suggest that coffee plants possess multiple enzymes with distinct affinities toward substrates, thereby catalyzing caffeine biosynthesis at a broad range of available substrate concentrations (Uefuji et al., 2003). For example, if a group of cells or tissues have low concentrations of the AdoMet or any substrate, then these cells express a high affinity (low Km) NMT isozyme. Accordingly, the distinct biological function of each isozyme is defined by its biochemical properties, which in turn define its tissue specific expression. Such a concept enhances the physiological fine-tuning of the cell and demands that the enzymatic properties of each isozyme be distinct. Tsuchisaka and Theolog (2004) reported that isozyme diversity of the Arabidopsis 1-Amino-Cyclopropane-1-Carboxylate synthase gene family provides physiological versatility by being able to operate in a broad gradient of

AdoMet concentration in various cells/tissues during plant growth and development. Since AdoMet or SAM is the methyl donor in caffeine biosynthetic pathway, this concept may be applicable to coffee NMTs as well.

It is also possible that the multiple copies of the genes may have different affinities towards substrate specificities. However, the motifs determining the substrate specificities for coffee *N*-Methyltransferases are yet to be determined.

4.3.3 Polymorphism for coffee NMT (theobromine synthase-1) gene between *Coffea* species

4.3.3.1 Amplification and cloning of partial gene from Indian wild coffee species

Indian wild coffee species *Psilanthus bengalensis* and *P. travencorensis* are considered to be caffeine free or low-caffeine species. Begum et al., (2003) showed that *P. bengalensis* is, typical of coffee species, able to synthesize caffeine, but at low levels. The reasons for low caffeine biosynthesis in these species are not known. To look for possible polymorphisms in the NMT gene structure or sequence, NMT gene amplification from wild species were attempted. Approximately, 1.7 kb PCR product was obtained for NMT gene(s) from two Indian wild species of coffee: *P. bengalensis* and *P. travencorensis* (Figure 4.3.10a). The PCR products were transferred to nylon membrane and probed with labelled clone CS2B from *Coffea canephora* as described in section 4.2.2. The PCR product from two wild Indian coffee species reacted with the probe, confirming the presence of the NMT gene and their homology to *Coffea canephora* NMT gene (Figure 4.3.10b).

The PCR product from *Psilanthus bengalensis* was T/A cloned as described in section 2.2.5 in chapter II. Two clones that differed in the size of the inserts were sequenced and named as BENG-B and BENG-S (big and small). On sequencing it was observed that one of the amplicons (clone BENG-S) was obtained due to amplification by mis-priming of the forward primer in PCR. The primer annealed to a region in second exon instead of the first exon, which resulted in a shorter fragment. The two clones were partially sequenced from the two ends using M13 forward and reverse primer (Figure 4.3.11 a & b).

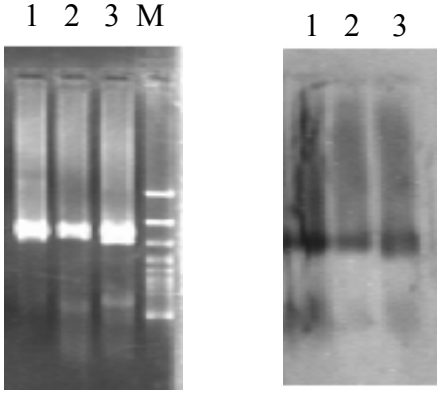
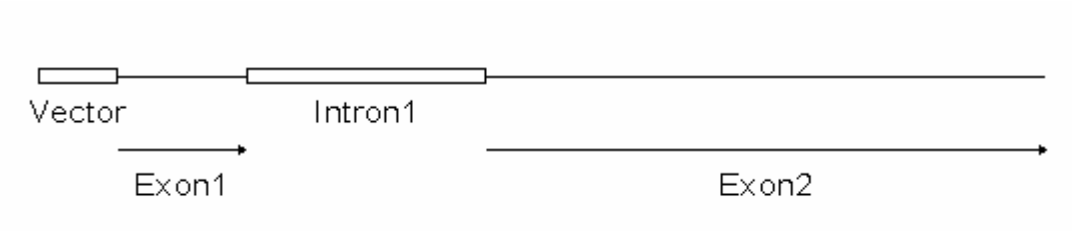


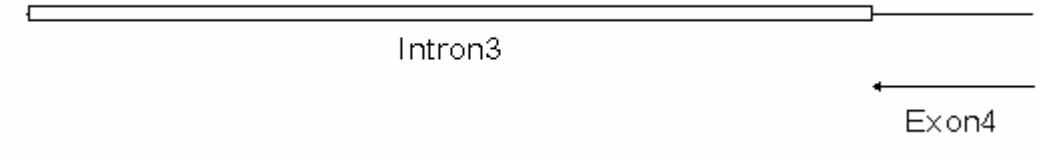
Figure 4.3.10 Southern blot of PCR amplicons from wild Indian coffee species

Figure 4.3.10a PCR amplicons for NMT genes from (1) *Coffea canephora*, (2) *P. bengalensis*, and (3) *P. travencorensis*. 'M' is a 3 kB DNA marker. Figure 4.3.10b Southern blot of PCR amplicons using labeled CS2B clone from *C. canephora* as a probe

The partial sequence of *P. bengalensis* NMT gene fragment was significantly similar to NMT genes of *C. canephora* (Figure 4.3.12) and it revealed identical gene structure. It is difficult to predict whether the two clones i.e. BENG-B and BENG-S belong to same gene. However, the three introns from *P. bengalensis* were taken as representative of this species for comparative study of gene structure (Table 4.3.2). Based on the nucleotide sequences and intron/exon joints in NMT genes from *C. canephora*, *C. liberica* and *P. bengalensis*, it can be concluded that the gene structure in terms of number of introns and their position is conserved in different *Coffea* species. Though intron sequences were also conserved, they exhibited relatively greater diversity/ polymorphism than coding regions. It is known that introns evolve much more rapidly than exons. This study also reveals that *P. bengalensis* possesses NMT genes, despite its near zero caffeine levels. The reasons for low levels of caffeine biosynthesis in this species need to be studied. Caillieux et al., (2004) suggested that mis-splicing or non-splicing of introns from deficient spliceosome generates truncated proteins in *C. pseudozanguebariae*. These wrongly processed transcripts may interfere severely with the caffeine biosynthesis leading to its inactivation in *C. pseudozanguebariae*. This possibility needs to be studied in *P. bengalensis*.

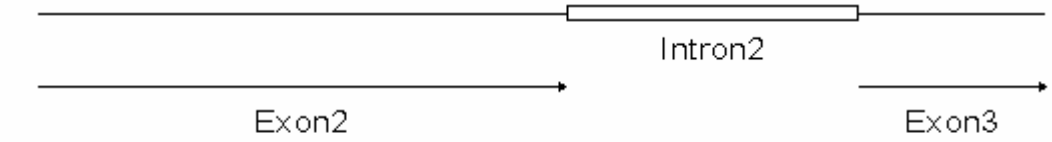



```
  1  ggccagtgaa ttcgagctcg gtacctcgcg aatgcatcta gattagctcc aagaagtcct
                                                      >>....Exon1.....>
     >>...................Vector..................>>
 61  gcatatgaat ggaggcgaag gcgatacaag ctacgccaag aattcatcgt acaatgtctg
     >.........................Exon1.........................>>
121  tctctctctc tatctctctc tttaacacac aaaaacacac gtaaatcatg caatgatacg
181  tcgatctcta acttagtgag tctttttacc ccataacatt tgttttttgg agtggttatc
241  tgtagcagct ggttctcacc aaggtgaaac ctgtccttga acaatgcata cgagaattgt
            >>.......................Exon2.......................>
301  tgcgggccaa cttgtccaac atcaacaagt gcattaaagt tgcggatttg ggatgcgctt
     >.............................Exon2.............................>
361  ctggaccaaa cacactttta acagttgggg acattgtaca gagtatcgac aaagttaggc
     >.............................Exon2.............................>
421  aagaaatgaa gaatgaatta gaacgtccca ccattcaggt ttttctgact gatcttttcc
     >.............................Exon2.............................>
481  aaaatgattt caattcggtt ttcaagttgc tgccaagctt ctacagcaaa cttgagaaag
     >.............................Exon2.............................>
541  aaaaatggac gcaaaatt
     >.....Exon2......>>
```




```
  1  ggttatcatt tctctgtctc tctgataatc agatgctcat tgcttgttat tcagggttcc
 61  taccttagct gaaaaatgga tgcaatttta ttctttgtaa ccttgataga ctaaactttg
121  tgtagacgaa gttcaccgta aatcagatgg tagtaacctt taattaggag ttaggactca
181  caactccaat acagttaaga taattctttc cctttcaata atttgttgga agttagaagt
241  gtagggggat ttaattcgtc ggctctctct ttctttttgc ccttttttggt taaaaaggta
301  ttggaaagga agaaatagtt cattatttga tttttttttg gggaaatttt gatatgtacg
361  ttttattttt ccagagttga cgttttagca aattggctaa tttccaagtt ttgttcgtcc
421  ggtaattcga attgacttga tgattggcat acggatacag ggacatcggg aggaagaaaa
                                                 <<......Exon4.......<
481  attggacagt ttcaatgttc caatctatac accttcagta gaagaagtaa agtgcatggt
     <.............................Exon4.............................<
541  tgaggagg
     <.....<<
```

Figure 4.3.11a Partial sequencing of NMT gene fragment from *Psilanthus bengalensis* (clone BENG-B)




```
  1  cctcctcaac catgcacttt taacagttgg ggacattgta cagagtatcg acaaagttag
     >>.............................Exon2.............................>
 61  gcaagaaatg aagaatgaat tagaacgtcc caccattcag gtttttctga ctgatctttt
     >.............................Exon2.............................>
121  ccaaaatgat ttcaattcgg ttttcaagtt gctgccaagc ttctacagca aacttgagaa
     >.............................Exon2.............................>
181  agaaaatgga cgcaaaatag gatcgtgcct aatagccgca atgcctggat ctttctacgg
     >.............................Exon2.............................>
241  cagactcttc cccgaggagt ccatgcattt tttacactct tcttacagtg ttcatttttt
     >.............................Exon2.............................>
301  atctcaggtc tttgaattac tccctttttat ctttttactt tttcttgtag caaaaatagt
     >....>>
361  tcatgatttt cattcaacac atcagtaact atgcatggaa atttctttaa taattctcaa
421  gatatccaca ggaatccaag aaagagattt ctaaagggaa ccagctttag actgcaggtt
                                                                    >>>
481  cccagcggtt tggtggctga attggggatc agtgcgaaca aaaggagcat ttactcttcc
     >.............................Exon3.............................>
541  aaagcaagtc ctccgctcgt ccagaaggca tatttggatc aattta
     >......................Exon3.....................>>
```

Figure 4.3.11b Partial sequencing of NMT gene fragment from *Psilanthus bengalensis* (clone BENG-S)

```
CaMXMT        1            C AGTCGCAATT CGATTGTCCT GCATATGAAT GGAGCTCCAA
CS2A          1            - ---------- ---------- ---------- --AGCTCCAA
CS2b          1            - ---------- ---------- ---------- --AGCTCCAA
BengB         1            - ---------- ---------- ---------- --AGCTCCAA

CaMXMT       43   GAAGTCCTGC ATATGAATGA AGGTGAAGGC GATACAAGCT ACGCCAAGAA
CS2A          9   GAAGTCCTGC ATATGAATGG AGGCGAAGGC GATACAAGCT ACGCCAAGAA
CS2b          9   GAAGTCCTGC ATATGAATGA AGGTGAAGGC GATACAAGCT ACGCCAAGAA
BengB         8   GAAGTCCTGC ATATGAATGG AGGCGAAGGC GATACAAGCT ACGCCAAGAA

CaMXMT       93   TGCATCCTAC AAT------- ---------- ---------- ----------
CS2A         59   TTCATCGTAC AATGTCTGTC TG--TCTCTC cgTCTCTCTC TTTAACACAC
CS2b         59   TGCATCCTAC AATGTCTGTC TG--TCTCTC --TATCTCTC TTTAACACAC
BengB        58   TTCATCGTAC AATGTCTGTC TctcTCTCTA --TCTCTCTC TTTAACACAC

CaMXMT      107   ---------- ---------- ---------- ---------- ----------
CS2A        107   ACACACACAC acAGAGTAGT AGTAAATCAT GCTATGATAC GTTGATCTCT
CS2b        105   ACACACACAg --AGAATAGT GGTAAATCAT GCTATGATAC GTCGATCTCT
BengB       106   AAAAACACAC ---------- -GTAAATCAT GCAATGATAC GTCGATCTCT
CaMXMT      107   ---------- ---------- ---------- ---------- ----------
CS2A        157   GACTTAGTat gtcttTTTTC CCACCTTAAT ATTTGTATTT TGGAGTGGTT
```

```
CS2b        153   AACTTc---- ---------- --------AC ATTTGTATTT TGGACTGG-T
BengB       145   AACTTAGTga gtc--TTTTT ACCCCATAAC ATTTGTTTTT TGGAGTGGTT

CaMXMT      107   ---------- -CTGGCTCTT GCCAAGGTGA AACCTTTCCT TGAACAATGC
CS2A        207   ATGTGTAGCA GCTGGTTCTC ACCAAGGTGA AACCTGTCCT TGAACAATGC
CS2b        180   ATGTGTAACA GCTGGCTCTT GCCAAGGTGA AACCTTTCCT TGAACAATGC
BengB       193   ATCTGTAGCA GCTGGTTCTC ACCAAGGTGA AACCTGTCCT TGAACAATGC

CaMXMT      145   ATACGAGAAT TGTTGCGGGC CAACTTGCCC AACATCAACA AGTGCATTAA
CS2A        257   ATACGAGAAT TGTTGCGGGC CAACTTGCCC AACATCAACA AGTGCATTAA
CS2b        230   ATACGGGAAT TGTTGCGGGC CAACTTGCCC AACATCAACA AGTGCATTAA
BengB       243   ATACGAGAAT TGTTGCGGGC CAACTTGtCC AACATCAACA AGTGCATTAA

CaMXMT      195   AGTTGCGGAT TTGGGATGCG CTTCTGGACC AAACACACTT TTAACAGTGC
CS2A        307   AGTTGCGGAT TTGGGATGCG CTTCTGGACC AAACACACTT TTAACAGTTC
CS2b        280   AGTTGCGGAT TTGGGATGCG CTTCTGGACC AAACACACTT TTAACAGTGC
BengB       293   AGTTGCGGAT TTGGGATGCG CTTCTGGACC AAACACACTT TTAACAGTTg

CaMXMT      245   GGGACATTGT GCAAAGTATT GACAAAGTTG GCCAGGAAGA GAAGAATGAA
CS2A        357   GGGACATTGT GCAAAGTATT GACAAAGTTG GCCAGGAAGA GAAGAATGAA
CS2b        330   GGGACATTGT GCAGAGTATT GACAAAGTTG GCCAGGAAGA GAAGAATGAA
BengB       343   GGGACATTGT ACAGAGTATC GACAAAGTTA GGCAAGAAAt GAAGAATGAA

CaMXMT      295   TTAGAACGTC CCACCATTCA GATTTTTCTG AATGATCTTT TCCAAAATGA
CS2A        407   TTAGAACATC CCACCATTCA AATTTTTCTG AATGATCTTT TCCAAAATGA
CS2b        380   TTAGAACGTC CCACCATTCA GATTTTTCTG AATGATCTTT TCCAAAATGA
BengB       393   TTAGAACGTC CCACCATTCA GGTTTTTCTG ACTGATCTTT TCCAAAATGA

CaMXMT      345   TTTCAATTCG GTTTTCAAGT TGCTGCCAAG CTTCTACCGC AAACTCGAGA
CS2A        457   TTTCAATTCG GTTTTCAAGT TGCTGCCAAG CTTCTACCGC AAACTTGAGA
CS2b        430   TTTCAATTCG GTTTTCAAGT TGCTGCCAAG CTTCTACCGC AAACTCGAGA
BengB       443   TTTCAATTCG GTTTTCAAGT TGCTGCCAAG CTTCTACAGC AAACTTGAGA

CaMXMT      395   AAGAAAATGG ACGCAAGATA GGATCGTGC CTAATAAGCG
CS2A        507   AAGAAAATGG ACGCAAAATA GGAT----- ----------
CS2b        480   AAGAAAATGG ACGCAAGATA GGATCGTGC CTAATAAGCG
BengB       493   AAGAAAATGG ACGCAAAAT- --------- ----------
```

Figure 4.3.12 Multiple sequence alignments of *C. canephora* and *P. bengalensis* NMT genes

The NMT genes seem to be conserved in *Coffea* and *Psilanthus* genera. The present classification of coffee trees into these two genera is disputed. Molecular analyses (Cros et al., 1998; Lashermes et al., 1997) indicate a close relationship of species belonging to the two genera and prompt modifications in the adopted classification based on flowering and flower characteristics (Leroy, 1980).

4.3.3.2 Amplification of full-length theobromine synthase gene from *Coffea species*

PCR amplifications of NMT gene from various *Coffea* species performed with the specific primers located on UTR regions of theobromine synthase-1 revealed polymorphism in the form of multiple bands from several species (Figure 4.3.13). The multiple bands observed in few species suggest that multiple copies or alleles of the theobromine synthase-1 gene are present in *Coffea* species. The multiple bands observed in *Coffea canephora* confirm the earlier results of presence of multiple copies / alleles of the NMT gene (Clones PG-1, 4, 5, CX8 and CX10). However, with *Coffea eugenioides* and *Psilanthus travencorensis* only a single band was observed. The size difference between the observed amplified fragments suggest that intronic sequences differ in length as observed earlier for NMT genomic clones in *Coffea canephora*. Intra-specific variations were observed in banding pattern for *Coffea canephora* cv. S-274 and cv. C x R. C X R variety of Robusta coffee is a commercial interspecific hybrid between *C. congensis and C. canephora* and the observed differences between the two varieties could be due to their genetic background and also due to highly polymorphic nature of *C. canephora* species.

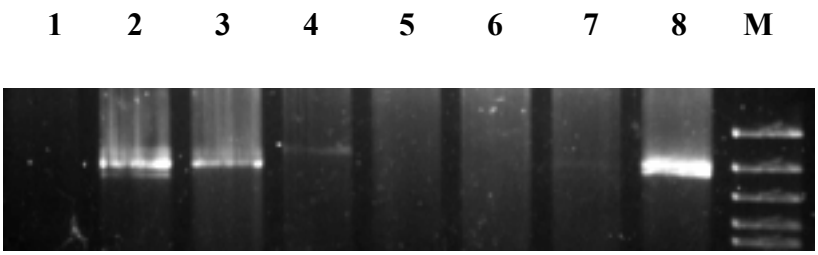


Figure 4.3.13 Polymorphism for theobromine synthase-1 gene between *Coffea* species

Lane 1 *Psilanthus bengalensis*
Lane 2 *Coffea canephora S-274*
Lane 3 *Coffea congensis*
Lane 4 *Coffea eugenioides*
Lane 5 *Coffea racemosa*
Lane 6 *Coffea salvatrix*
Lane 7 *P. travencorensis*
Lane 8 *Coffea canephora C* x *R*
M 3 kb DNA marker

The number of bands and band intensity observed for different species suggest that these species vary in the number of NMT gene copies / alleles. It also suggests that gene

duplications in coffee are recent events, occurring in only few species. Interestingly, species with low caffeine gave a single band (Figure 4.3.13). The multiple copies may also be allelic in nature, consistent with the earlier reports of high number of allelic locus in *C. canephora.*

4.3.4 Regulatory elements in orthologs for NMT gene

The NMT promoter fragments were amplified with two primer combinations from several species, except *C. eugenioides* and *P. bengalensis* (Table 4.3.7). The PCR products were in the size range of 0.8-0.9 kb (Figure 4.3.14). However, the primer combination F1/GSP2 did not yield any PCR product with *C. salvatrix*. This suggests that the promoter sequences of the corresponding NMT gene from *C. eugenioides* and *P. bengalensis* may be varying and the CTSF1 primer-binding site is probably not present or variations exist in that region in *C. salvatrix*. Another primer CTSF2 was used in successful amplification of promoter fragment from *C. salvatrix.* The amplified products were T/A cloned and sequenced. The sequences of five clones for the 5′ upstream region of the NMT gene from different *Coffea* species were compared to determine sequence variation in the promoter region of NMT gene between and within species.

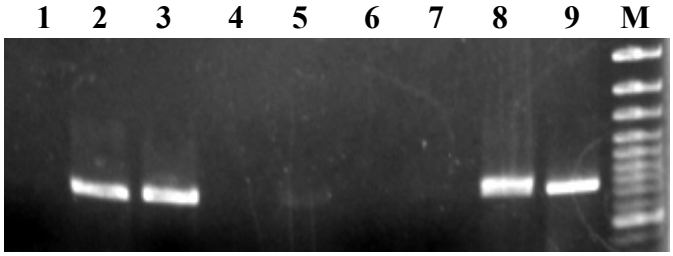


Figure 4.3.14 Amplification of promoter region for NMT gene from different *Coffea* species

Lane 1: *Psilanthus bengalensis,* Lane 2: *Coffea canephora S-274*,
Lane 3: *C. congensis,* Lane 4: *C. eugenioides*, Lane 5: *C. racemosa*,
Lane 6: *C. salvatrix*, Lane 7: *P. travencorensis*, Lane 8: *C. canephora C* x *R*,
Lane 9: *C.liberica* x *C. eugenioides*, M: 3 kb DNA marker

Table 4.3.7. Amplification results of the NMT promoter region with different primer combinations

Coffea species	F1/ GSP2 primers	F2/ GSP2 primers
C. canephora cv S-274	+	+
C. canephora cv CXR	+	+
C. congensis	+	+
C. eugenioides	−	−
C. racemosa	+	−
C. salvatrix	−	+
P. bengalensis	−	−

The NMT promoter sequences of four species viz., *C. canephora var.* S-274, *C. canephora var.* C x R, *C. racemosa*, *C. congensis*, *and C. salvatrix* were analyzed. Alignments of promoter sequences were performed using Clustal W. However, for *C. racemosa*, only 471 bases were included for alignment due to bad quality of sequences for AT rich region. The alignments of the promoter sequences of different accessions revealed no discernible differences for ~ 700 bp sequences upstream from the start codon in these species (Figure 4.3.15).

```
pSSPI           ATTATGAGAAATTCACCGAATAAATAGGCAAAGGTAGAACCTCACTTCTTAAATGAAATT 60

CXR             ----------------------------GTTCCCCATTTTCCTCCTTTCAAGTAGAATTA 32
pSSPI           ACCTAATATCCTTTCCATTTAACCAAAAGTTCCCCATTTTCCTCCTTTCAAGTAGAATTA 120
Cong            -------------------------------CCCCATTTTCCTCCTTTCAAGTAGAATTA 29

CXR             TCATGTGGGCAAGTAAATTCTGAGCAAGAAATAAAAAAAAAAAAAGAGTTGATTTGACATT 92
pSSPI           TCATGTGGGCAAGTAAATTCTGAGCAAGAAATAATAAA-AAAAGGAGTTGATTTGACATT 189
Cong            TCATGTGGGCAAGTAAATTCTGAGCAAGAAATAATAAA-AAAAAGAGTTGATTTGACATT 88

CXR             GAATAGACTCCAGCTTCTTCCTTTCATTTGTTGAGAGGAAAAAGAAATAGACGGGAAAGA 152
pSSPI           GAATAGACTCCAGCTTCTTCCTTTCATTTGTTGAGAGGAAAAAGAAGCAGACGGGAAAGA 249
Cong            GAATAGACTCCAGCTTCTTCCTTTCATTTGTTGAGAGGAAAAAGAAACGGACGGGAAAGA 148

Sal             ----GGAATGAAGCAGAAAAATGGTGGGCTAAAAGTCAAAAGAAGCGGAAAGTGA      51
CXR             AATAGGAATGAAGCAGAAAAATGGTGGGCTAAAAGTCAAAAGAAGCGGAAAGTGA      207
pSSPI           AATAGGAATGAAGCAGAAAAATGGTGGGCTAAAAGTCAAAAGAAGCGGAAAGTGA      304
Cong            AATAGGAATGAAGCAGAAAAATGGTGGGCTAAAAGTCAAAAGAAGCGGAAAGTGA      203

Sal             TAGTGTAGAGGAAGAAGGGAAGGAGGCGGAAGAACAGGAGGAAAGGTGCAGGTGAAAAAT 111
CXR             TAGTGTAGAGGAAGAAGGGAAAGAGGCGGAAGAACAGGAGGAAAGGTGCAGGTGAAAAAT 267
Race            ------------------------------------------------------------T 1
pSSPI           TAGTGTAGAGGAAGAAGGGAAAGAGGCGGAAGAACAGGAGGAAAGGTACAGGTGAAAAAT 364
Cong            TAGTGTAGAGGAAGAAGGGAAAGAGGCGGAAGAACAGGAGGAAAGGTACAGGTGAAAAAT 263

Sal             CAAAATTCAGCCCAAATAGTAAGCTACTATTCTGAAAGCCTGTTATGTCTAGAAGGACAA 171
CXR             CAAAATTCAGCCCAAATAGTAAGCTACTATTCTGAAAGCCTGTTATGTCTAGAAGGACAA 327
Race            CAAAATTCAGCCCAAATAGTAAGCTACTATTCTGAAAGCCTGTTATGTCTAGAAGGACAA 61
pSSPI           CAAAATTCAGCCCAAATAGTAAGCTACTATTCTGAAAGCCTGTTATGTCTAGAAGGACAA 424
```

```
Cong            CAAAATTCAGCCCAAATAGTAAGCTACTATTCTGAAAGCCTGTTATGTCTAGAAGGACAA 323
                ******** ************************************** **********

Sal             ACACGTACTTTTAAGGAAAAGGGCAAAAATGCCCTAGAATAGTTATACACTGTTTAGACC 231
CXR             ACACGTACTTTTAAGGAAAAGGACAAAAATGCCCTAGAATAGTTATACACTGTTTAGACC 387
Race            ACACGTACTTTTAAGGAAAAGGACAAAAATGCCCTAGAATAGTTATACACTGTTTAGACC 121
pSSPI           ACACGTACTTTTAAGGAAAAGGACAAAAATGCCCTAGAATAGTTATACACTGTTTAGACC 484
Cong            ACACGTACTTTTAAGGAAAAGGACAAAAATGCCCTAGAATAGTTATACACTGTTTAGACC 383
                ********************** ************************************

Sal             AGGGTACACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTAGCTGCTTCTCTCCC 291
CXR             AGGGTACACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTAGCTGCTTCTCTCCC 447
Race            AGGGTACACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTAGCTGCTTCTCTCCC 181
pSSPI           AGGGTACACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTGGCTGCTTCTCTCCC 544
Cong            AGGGTACACTTTATGAGGTTATTGCTGGAAAAAACGAAACATAGTGGCTGCTTCTCTCCC 443
                ******************************************** **************

Sal             ATCTTTTGCAACGTGGGAGAAGAAGCCTTAGGTTTTTATATATATATGTAAGATGTAAGA 351
CXR             ATCTTTTGCAACGTGGGAGAAGAAGCCTTAGGTTTTTATATATATATGTAAGATGTAAGA 507
Race            ATCTTTTGCAACGTGGGAGAAGAAGCCTTAGGTTTTTATATATAT--GTAAGATGTAAGA 239
pSSPI           ATCTTTTGCAACGTGGGAGAAGAAGCCTTAGGTTTTTATATATATATGTAAGATGTAAGA 604
Cong            ATCTTTTGCAACGTGGGAGAAGAAGCCTTAGGTTTTTATATATATATGTAAGATGTAAGA 503
                ***************** ************************** *************

Sal             TAAGATAAGATATTTCATCCCCGCATCTCAACTTCTGATTTTATCATTCGTGTCTGGTTC 411
CXR             TAAGATAAGATATTTCATCCCCGCATCTCAACTTCTGATTTTATCATTCGTGTCTGGTTC 567
Race            TAAGATAAGATATTTCATCCCCGCATCTCAACTTCTGATTTTATCATTCGTGTCTGGTTT 299
pSSPI           TAAGATAAGATATTTCATCCCCGCATCTCAACTTCTGATTTTATCATTCGTGTCTGGTTC 664
Cong            TAAGATAAGATATTTCATCCCCGCATCTCA-CTTCTGATTTTATCATTCGTGTCTGGTTC 562
                ***************************** ***************************

Sal             CCATTGGCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGC 471
CXR             CCATTGGCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGC 627
Race            CCATTGGCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGC 359
pSSPI           CCATTGGCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGC 724
Cong            CCATTGGCTGTGCGCTTTCTTTCTGACGAATTGACAGACTTTTCTACGCACGGAGGTAGC 622
                ** ********************************************************

Sal             TGGTTAGCATACGCATCTATGAAATTTTCGCTACTTAAGCCCGAAATTTTGCACAATTAA 531
CXR             TGGTTAGCATACGCATCTATGAAATTTTCGCTACTTAAGCCCGAAATTTTGCACAATTAA 687
Race            TGGTTAGCATACGCATCTATGAAATTTTCGCTACTTAAGCCCGAAATTTTGCACAATTAA 419
pSSPI           TGGTTAGCATACGCATCTATGAAATTTTCGCTACTTAAGCCCGAAATTTTGCACAATTAA 784
Cong            TGGTTAGCATACGCATCTATGAAATTTTCGCTACTTAAGCCCGAAATTTTGCACAATTAA 682
                *** ********************************************************

Sal             TCATTAACAGACACCTTCTTTAGCAGTCGCAATTCGATTGTCCTGCATATGA 583
CXR             TCATTAACAGACACCTTCTTTAGCAGTCGCAATTCGATTGTCCTGCATATGA 739
Race            TCATTAACAGACACCTTCTTTAGCAGTCGCAATTCGATTGTCCTGCATATGA 471
pSSPI           TCATTAACAGACACCTTCTTTAGCAGTCGCAATTCGATTGTCCTGCATATGA 836
Cong            TCATTAACAGACACCTTCTTTAGCAGTCGCAATTCGATTGTCCTGCATATGA 734
                ***************************************************
```

Figure 4.3.15 Multiple sequence alignments of NMT promoter from different species
Sal: *C. salvatrix*, CXR: *C. canephora* cv. CXR, Race: *C. racemosa*, pSSPI: *C. canephora* cv. S-274, Cong: *C. congensis*. Due to bad quality of nucleotide sequence for *C. racemosa*, only 471 bases were used for sequence analysis.

The pairwise alignment scores for the sequences very high (97 to 100%). The low level of DNA variation in the 5′-upstream regions exhibited by coffee tree species could

probably be related to the recent origin of genus *Coffea*. Lashermes et al., (1996) and Cros et al., (1998) also suggested the recent origin of *Coffea*. The results obtained in this study on coffee are in sharp contrast to the studies in cereals. The promoter sequences among cereal genomes, which diverged from a common ancestor ~50 million years ago, are evolving rapidly and show little sequence conservation even among orthologous genes (Guo and Moose, 2003).

Earlier coffee trees have been categorized into two groups according to their green bean caffeine content, with one grouping caffeine-free species, including most Mascarocoffea and some Mozambicoffea, and the other pooling species with caffeine in their green beans. Recently, Campa et al., 2005 regrouped 21 *Coffea* species into four classes based on their caffeine content (Table 4.3.8).

Table 4.3.8 Classification of *Coffea* species based on their caffeine content

Class	Caffeine content (%dmb) in the class	Species
CAF1	<0.03	*C. salvatrix*
CAF2	0.51- 0.58	*C. sp Moloundou* *C. eugenioides*
CAF3	0.92 – 1.47	*C. congensis* *C. racemosa*
CAF4	>1.93	*C. canephora*

The four species, for which promoter sequences for putative theobromine synthase gene were cloned, belonged to three different classes based on their caffeine content (described by Campa et al., 2005) with each class having different range of caffeine content. One would have expected that there is difference in the synthesis rates of caffeine in these classes as a result of differential rates of transcription of the genes involved in the caffeine biosynthesis. However, the nucleotide sequences of promoter regions in four species were highly conserved, with no significant differences among them. The promoter sequence similarities in these species suggest that these species have identical rates of transcription for the corresponding NMT gene. This leads to a speculation that varying caffeine content

in these coffee species could be a result of the balance between synthesis and degradation, and not just differential rates of transcription of biosynthetic genes. However, the promoter sequences for other genes in the pathway need to be studied. Moreover, various other cellular processes may also control caffeine biosynthesis and / or accumulation in coffee *in vivo*, such as caffeine: chlorogenic acid complexing, which need to be ascertained.

4.3.5 Summary and Conclusions

Several genomic clones, those are significantly similar in their exonic regions but varying in one of the introns, have been obtained in this study. Considering the great similarity in coding regions, it may be said that all the clones belong to NMT gene family. It is well known that when two genes are related, the relationship between their exons is closer than the relationship between the introns. In extreme cases, the exons of two genes may code for the same protein sequence, but the introns may be different (Lewin, 2004). This situation has been observed for three genomic clones PG-1, 5 and CX10, which encode for 378 AAs and share high % of identities among themselves (>98%) but significantly vary in sequence of the third intron. Among introns, the pattern of divergence involves both changes in size (due to deletions and insertions) and base substitutions. It is known that introns evolve much more rapidly than exons. Introns may participate in the creation of divergence between paralogs, similarly to promoters or other regulatory elements (Lynch and Conery, 2000) and the possibility of this in the evolution of coffee NMT gene family needs to be explored. Based on sequence comparison alone, it is difficult to assign if three genomic clones belong to different members of the gene family (i.e. paralogs) or copies of the same gene (i.e. allelic). The comparisons of promoter regions from different species revealed high sequence similarities among the species. This may be due to the shorter evolutionary distance between the studied species and may indicate an ongoing degeneration process within promoters.

Genes required to undertake quick, large transcription responses are prone to becoming large families. Such a development may help provide a readily tuneable on/off switch for a

variety of specialized functions in response to a host of stimuli. It is known that most of the enzymatic components of plant secondary metabolism are encoded by small families of genes that originated through duplication over evolutionary time (Durbin et al., 2000). Based on the available literature and results obtained in this study, it can be concluded that coffee *N*-methyltransferases belong to a large multigene family. The possible reasons for the coffee plants possessing large multigene family of *N*-methyltransferases, in terms of quick transcriptional responses and variety of specialized functions in response to a host of stimuli, (features generally associated with large multigene families) needs to be studied.

The coffee NMT genes belong to a new methyltransferase family recently identified. The closest accession to NMT gene in terms of gene structure among this family based on available evidence is JMT from Arabidopsis (discussed in chapter III). Though the location of introns was highly conserved, the intron sequences and their sizes exhibited a great diversity. Also there were no significant similarities in the promoter sequences of the two genes. This could be due to the fact that coffee and Arabidopsis belong to different plant families (Rubiaceae and Brassicaceae, respectively) that are distantly related phylogenetically and which diverged from their last common ancestor approximately 94 million years ago.

The uniqueness of coffee is its high diversity of primary and secondary compounds, which contributes to the sensory quality of brewed coffee beans (Lin et al., 2005). The Rubiaceae family in general contains some of the most diverse species with regard to secondary metabolism and is an especially rich source of alkaloids- a number of which have pharmacological and/or psychotropic properties (Facchini, 2001). This metabolic diversity was reflected in the relatively high proportion of coffee genes with putative functions related to metabolism in the coffee ESTs (Lin et al., 2005). The authors observed a number of genes expressed that are unique to coffee when compared to Arabidopsis and close relatives of coffee in Solanaceae family. They speculated that these genes might be related to chemical or morphological features unique to coffee. Considering this observation of Lin et al., (2005), the coffee NMT genes may represent coffee-specific genes or genes that have been evolving at such a rapid rate that they no longer bear any recognizable homologies with proteins from other plants like Arabidopsis, and the closely

related Solanaceous plants. Similar results were also obtained by Gaitan (1998) who concluded that in coffee the exons can be similar to those in other plants, but introns, 5´-UTRs and promoters of the genes studied do not share significant similarities.

In the subsequent chapter, posttranscriptional gene silencing of coffee *N*-Methyltransferases using a conserved region for the NMT genes is discussed.

Chapter V

Posttranscriptional Gene Silencing of Coffee N-Methyltransferases

5.0 ABSTRACT

A highly conserved region among coffee *N*-Methyltransferase genes involved in caffeine biosynthetic pathway was used in making sense, antisense and RNA interference constructs. Sonication Assisted *Agrobacterium tumefaciens* (SAAT) mediated transformation was used to mobilize these constructs into *Coffea canephora*. Somatic embryos were used for the transformation experiments. The transformed coffee plantlets, selected on hygromycin containing selection media, were used for estimating the purine alkaloids content by High Pressure Liquid Chromatography (HPLC). The purine alkaloids viz: 7-methylxanthine, theobromine, caffeine and theophylline were estimated from the transgenic plantlets. The three constructs differed in their efficiencies and specificity of silencing. RNAi was found to be the most efficient in reducing the caffeine content in transformed coffee plantlets, when compared to sense and antisense constructs. Though the constructs were not specific, sense and RNAi transformants were obtained that were mainly affected in one of the *N*-Methyltransferases. The results suggest that the use of highly homologous coding region is highly effective in down regulating caffeine biosynthesis, encoded by *N*-methyl transferase multigene family. Increased transcription of CaMTL1 gene, substrate specificity of which is not known, was observed indicating possible activation of parallel or bypass pathways.

5.1 INTRODUCTION

Regulation of eukaryotic gene expression occurs at different stages of protein biosynthesis: at the transcriptional, RNA processing and translational levels, and at the level of protein maturation/degradation. Homology dependent gene silencing (HDGS)/ RNA silencing involves two distinct mechanisms that operate at either the transcriptional or post-transcriptional level (Mette et al., 1999). Posttranscriptional gene silencing (PTGS) is a form of HDGS that requires homology in protein coding or transcribed regions. Transcriptional gene silencing (TGS) occurs in the nucleus when double stranded (dsRNAs) containing sequences identical to promoter regions can induce methylation of unlinked homologous promoters (Mette et al., 2000). Post-transcriptional gene silencing (PTGS), a sequence specific RNA degradation mechanism in the cytoplasm, can be induced in plants by transforming them with antisense, co-suppression or hairpin RNA constructs (Smith et al., 2000).

Antisense is usually considered as a mechanism for sequence- specific messenger RNA (mRNA) recognition that leads to transcript degradation. Regulation of gene expression by antisense RNA was first discovered as naturally occurring phenomena in bacteria. The antisense technology is based on blocking the information flow from DNA via RNA to protein by the introduction of an RNA strand complementary to (part of) the sequence of the target mRNA. This so-called antisense RNA base pairs to its target mRNA thereby forming double stranded RNA. Duplex formation may impair mRNA maturation and / or translation or alternatively may lead to rapid mRNA degradation (Mol et al., 1990). In eukaryotes, antisense RNAs can be produced by utilizing a DNA fragment that covers the 5´ end of the mRNA to block the capping reaction, or a DNA fragment which covers an exon-intron (and intron-exon) junction to block the splicing reaction. Alternatively, a DNA fragment which covers the poly (A) site of the mRNA can give rise to antisense transcripts that block mRNA transfer from the nucleus to the cytoplasm (Inouye, 1988).

~~The regulation of gene expression at posttranscriptional level has recently attracted much~~ attention because of the discovery of the phenomenon called RNA interference (RNAi). RNAi has advantages over antisense mediated gene silencing and co-suppression in terms of its efficiency and stability (Smith et al., 2000; Wesley et al., 2001). Post-transcriptional gene silencing (PTGS) was first reported in plants as a coordinated and reciprocal inactivation of host genes and transgenes encoding the same RNA (Napoli et al., 1990). A similar phenomenon was observed in the filamentous fungus *Neurospora crassa* and termed quelling (Romano and Macino, 1992). A major breakthrough in the silencing history was the discovery of a gene silencing response in the nematode *Caenorhabditis elegans*. This phenomenon, in which experimentally introduced double stranded RNA (dsRNA) leads to loss of expression of the corresponding cellular gene by sequence-specific RNA degradation, was called RNA interference (RNAi). The protective effect of RNA silencing in reducing virus infectivity supports the view that PTGS has evolved as a mechanism to defend plants against virus infection and also to moderate the possible deleterious, genome-restructuring (insertional) activity of virus-like mobile genetic elements e.g. retrotransposons. A growing body of biochemical and genetic data has further demonstrated that plants, animals and yeasts share related mechanisms of specific degradation of RNAs in which double-stranded forms of RNA are initiator molecules (Bernstein et al., 2001; Dalmay et al., 2000; Fagard et al., 2000).

RNAi can be divided into four stages: (i) double- stranded RNA cleavage, (ii) silencing complex formation, (iii) silencing complex activation, and (iv) mRNA degradation (Szweykowska-Kuliñska et al., 2003). The first one includes ATP-dependent, processive dsRNA cleavage into double-stranded fragments 21 to 25 nucleotides long. The generated fragments are called small interfering RNAs (siRNAs) which contain 5′ phosphate and 3′ hydroxyl termini, and two additional overhanging nucleotides on their 3′ ends (Elbashir et al., 2001; Hamilton et al., 1999) (Figure 5.1.1). SiRNAs are the direct products of dsRNA cleavage by the multidomain RNase III enzyme DICER (Bass, 2000; Bernstein et al., 2001).

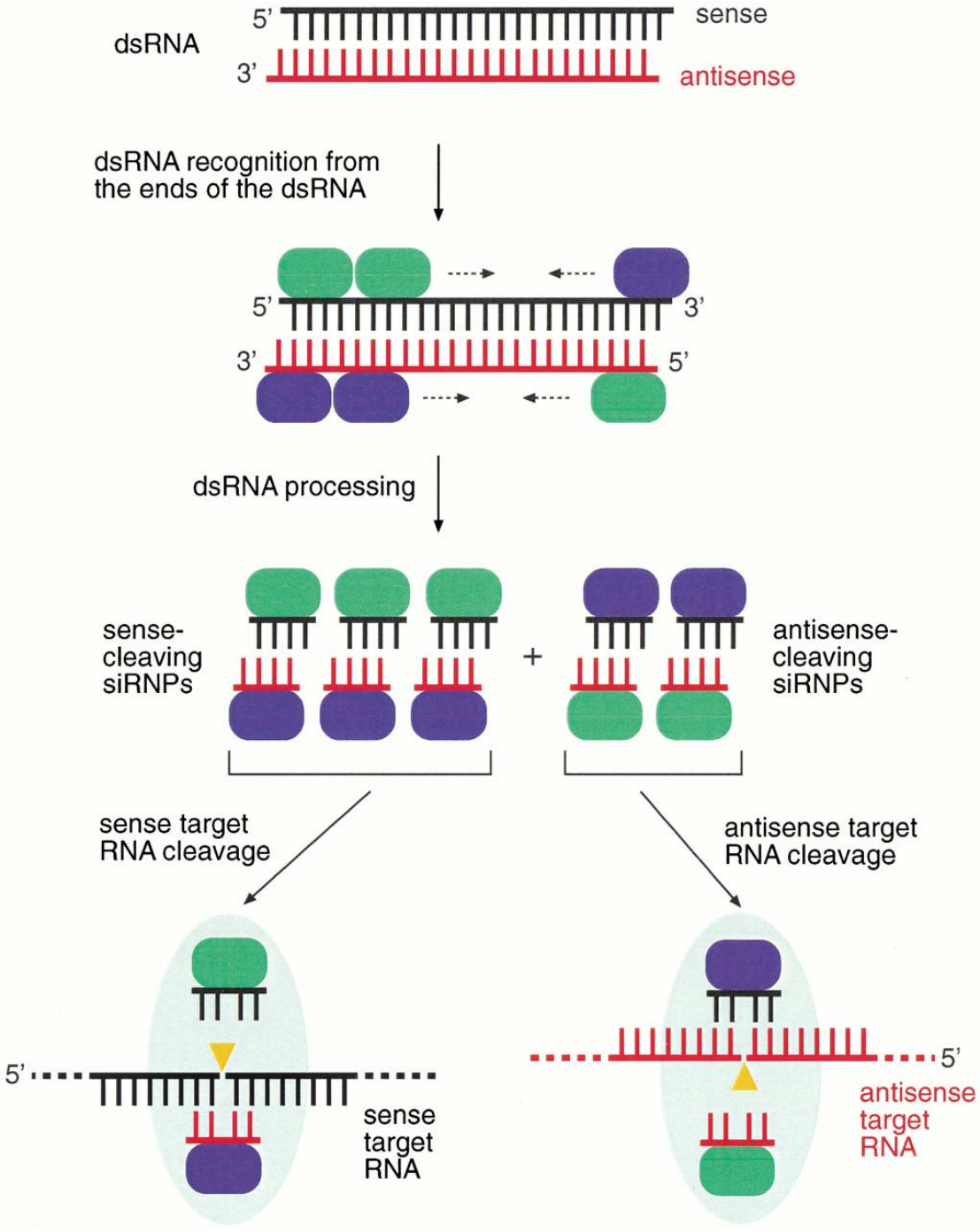


Figure 5.1.1 Mechanism of RNAi

~~The production of small (21–26 nucleotide) RNAs, which act as specificity determinants~~ that down regulate gene expression, is a unifying feature of RNA silencing across kingdoms. The Dicer enzymes, and also its family members, are large RNase III-type enzymes that distinctively contain one or two dsRNA- binding domains, putative RNA helicase, RNase III and PAZ domains. The RNase III-type enzymes cut dsRNA into 21–26 nucleotide long siRNAs molecules. Processing of dsRNA into siRNA duplexes starts from the ends of both blunt-ended dsRNAs or dsRNAs with short (1–5 nucleotide) 3′ overhangs, and proceeds in 21–26 nucleotide steps (Elbashir et al., 2001).

In the second step, siRNAs are incorporated into a protein complex called RISC (RNA-induced silencing complex), which is inactive in this form to conduct RNAi (Nykänen et al., 2001). The third step involves unwinding of the siRNA duplex and remodelling of the complex to create an active form of RISC (Nykänen *et al*. 2001). The final step includes the recognition and cleavage of mRNA complementary to the siRNA strand present in RISC (guide strand of siRNA).

In some organisms (*C. elegans*, *Arabidopsis thaliana*) an additional step in the RNAi pathway has been described involving a population of secondary siRNAs derived from the action of a cellular RNA-dependent RNA polymerase (RdRp). They are most likely generated during cyclic amplification in which RdRp is primed on the target mRNA template by existing siRNAs (Sijen et al., 2001). In contrast, transgenes that generate dsRNA and viruses that replicate through a dsRNA intermediate bypass the requirement for RdRp (Dalmay et al., 2000). A phenomenon known as transitivity increases the initial pool of siRNAs by producing new siRNAs corresponding to sequences located outside the primary targeted regions of a transcript (Sijen et al., 2001; Vaistij et al., 2002). Transitive RNA silencing is transcript-dependent i.e. some mRNAs but not all mRNAs engage in the process (Vaistij et al., 2002). In plants, transitive RNA silencing is dependent on RNA-dependent RNA polymerase activity and is bi-directional, i.e. it can travel both in 5′ → 3′ and 3′ → 5′ directions, and is responsible for spreading of silencing effect (transitive spreading).

~~In contrast to the production of siRNAs in Drosophila embryos and mammalian cells,~~ two distinct classes of siRNAs, short (21–23 bp) and long (24–26 bp) siRNAs, are produced in fungi, and also in plants during transgene silencing. Short and long siRNAs serve different functions (Hamilton et al., 2002; Nicolas et al., 2003; Tang et al., 2003). In plants, short siRNAs are involved in local RNA silencing whereas the long siRNAs appear to have an explicit role in systemic signaling and also in RNA- directed DNA methylation (Hamilton et al., 2002).

An endogenous negative gene regulatory pathway for translation repression, known as the microRNA (miRNA) pathway, also uses small RNAs of about 22 nucleotides. Unlike siRNAs, micro-RNAs are processed by Dicer-mediated cleavage of stem-loop precursor RNAs that are transcribed from genomic loci distinct from other recognized genes, and are predicted to form dsRNA hairpin structures. Plant miRNAs mediate target mRNA cleavage via a RISC-mediated RNA silencing mechanism (Tang et al., 2003). It appears that siRNAs and miRNAs, both in plants and animals, may be functionally interchangeable, i.e. some siRNAs may act like miRNAs to repress translation, and some miRNAs may act like siRNAs to silence transcription of the target gene (Tang et al., 2003).

The essence of RNAi technology is the delivery of dsRNA as a potent activator of RNA silencing into an organism, or cell, with the purpose of triggering sequence-specific degradation of homologous target RNAs. dsRNA can be delivered by stably transforming plants with transgenes that express a self-complementary RNA (Hamilton et al., 1998; Waterhouse et al., 1998). The resulting transcript hybridizes with itself to form a hairpin structure that comprises a single-stranded loop region (usually corresponding to an intron) and a base-paired stem, which mimics the dsRNA structure that induces RNAi. It has been shown that transgene constructs encoding intron-spliced RNA with hairpin structure give stable silencing with almost 100% efficiency when directed against viruses (Smith et al., 2000).

~~PTGS is a highly variable process dependent on the transgene, the host-plant species, the~~ developmental stage of the plants, and environmental factors (Meins, 2000). There appears to be strict species-specific requirements both in terms of the length of sequence and the degree of homology shared, to initiate RNA silencing; generally, more than 75% homology is needed to induce RNA silencing against a specific gene (Sijen et al., 1996). Silencing appears to be most efficient when sequences of more than 300 base pairs are used to design hpRNA constructs (Wang and Waterhouse, 2001). By choosing unique or conserved region, a single or all members of gene families, respectively can be silenced (Wesley et al., 2001). PTGS based on untranslated regions is useful for inactivating individual alleles or gene family members that have unique sequences in these regions. By contrast, PTGS based on coding sequence homology is suited for silencing all members of gene families that share high sequence homology in these regions (Kooter et al., 1999). Roughly, 60–70% identity of sequence is required for PTGS of cDNAs representing members in the same multigene family (Meins, 2000). One, some or all members of a multigene family have been silenced by RNAi by targeting sequences that are unique or conserved (Allen et al., 2004; Fukusaki et al., 2004; Ifuku et al., 2003).

Application of RNAi in crop plants

RNAi technology has also been successful in genetic modification of the fatty acid composition of oil (Liu et al., 2002). RNAi technology has also been used in several other plants to reduce alkaloids content in coffee (Ogita et al., 2003, 2004) and opium poppy (Allen et al., 2004). RNAi has been successfully used to generate a dominant high-lysine maize variant by knocking out the expression of the 22-kD maize zein storage protein, a protein that is poor in lysine content (Segal et al., 2003). Reduction of lysine catabolism specifically during seed development by an RNAi approach indeed improves seed germination (Zhu and Galili, 2004).

5.2 MATERIALS AND METHODS

5.2.1 Designing of antisense, sense and inverse repeat construct

5.2.1.1 Antisense construct:

Primers for sub cloning were designed to amplify a 339 bp region from the second exon of the genomic clone CS2B described in chapter II. The primer sequences were as follows: 2BHASF 5′-CCGATCTAGAGCCAAGGTGAAACCTTTC-3′ and 2BHASR 5′- CTGTATCGATCGTAGAAAGAGCCAGGCA -3′ with X*ba*I and C*la*I sites respectively, at their 5' ends (underlined). The conditions for PCR were as follows: Initial denaturation at 94^0C for 4 min, denaturation at 94^0C for 30 sec, annealing temperature of 56^0C for 45 sec, extension temperature of 72^0C for 1min 15 sec for 30 cycles, followed by final extension at 72^0C for 10 min. The PCR was carried out using proof reading DNA polymerase XT-Taq system (Bangalore Genie, India) in 50 µl reaction. The PCR reaction was checked on 1% agarose gel. The PCR product was purified, A-tailed by incubating the purified PCR product with 1U of Taq Polymerase (Bangalore Genie, India) at 72^0C for 30 min, and T/A cloned using InsT/A Clone PCR Product Cloning kit (MBI Fermentas). The recombinant plasmids were isolated, checked by 0.8% gel analysis, restriction digestion and PCR analysis. The orientation of the resulting construct pTZBHAS was verified by sequencing with M13 primer. X*ba* I / B*am*H I insert from pTZBHAS was gel eluted, ligated into X*ba* I / B*am*H I digested pRT100 vector. The resulting construct pRTBHAS contained the insert in antisense direction with respect to CaMV 35S promoter.

The expression cassette containing CaMV 35S promoter, antisense NMT and *Nos* terminator was released by restricting with P*st* I, gel eluted and ligated into P*st* I restricted, dephosphorylated pCAMBIA 1301 (Figure 5.2.1). The resulting antisense construct was called as pSAT202.

5.2.1.2 Sense construct

S*ac*I / B*am*HI insert was released from construct pTZBHAS (section 5.2.1.1); gel eluted and ligated to B*am*HI/ S*ac* I restricted pRT100. The resulting construct pRT201 contained the NMT gene fragment in sense orientation with respect to CaMV 35S promoter. The expression cassette containing CaMV35S promoter, sense NMT gene fragment and *Nos* terminator was released by restricting with P*st* I, gel eluted and ligated into P*st* I treated, dephosphorylated pCAMBIA 1301 (Figure 5.2.2). The resulting sense construct was called as pSAT201.

5.2.1.3 Invert Repeat construct

Using another set of primers, 2BINAF 5′-ATGGAGCTCGCCAAGGTGAAACCTTTC-3′ and 2BINAR 5′- ATAGGATCCCGTAGAAAGAGCCAGGCA-3′, 339 bp region from the second exon of the genomic clone CS2B was amplified as described in 5.2.1.1. The PCR product was T/A cloned and the orientation of the resulting clone pTZBIN(S) was checked by restriction mapping and sequencing. ~380 bp B*am*HI / S*ac*I insert from pTZBIN was ligated to B*am*HI / S*ac*I digested pBluescript KSII$^+$ to get pKSBIN. The K*p*nI / S*ac*I fragment from pKSBIN containing the sense NMT gene fragment was ligated to K*p*nI / S*ac*I treated pRTBHAS (described in section 5.2.1.1) containing the antisense NMT gene fragment with respect to CaMV 35S promoter.

The resulting construct pRT222 contained an inverse repeat of NMT gene fragments i.e. sense and antisense fragments were ligated in head to head fashion. The two arms of invert repeat structure were separated by a spacer of 90 bases comprising of multiple cloning site of the cloning vector. The expression cassette containing CaMV 35S promoter, NMT invert repeat and *Nos* terminator was released by partial digestion with P*st*I, gel eluted and ligated into P*st*I treated, dephosphorylated pCAMBIA 1301 to get pSAT222 (Figure 5.2.3).

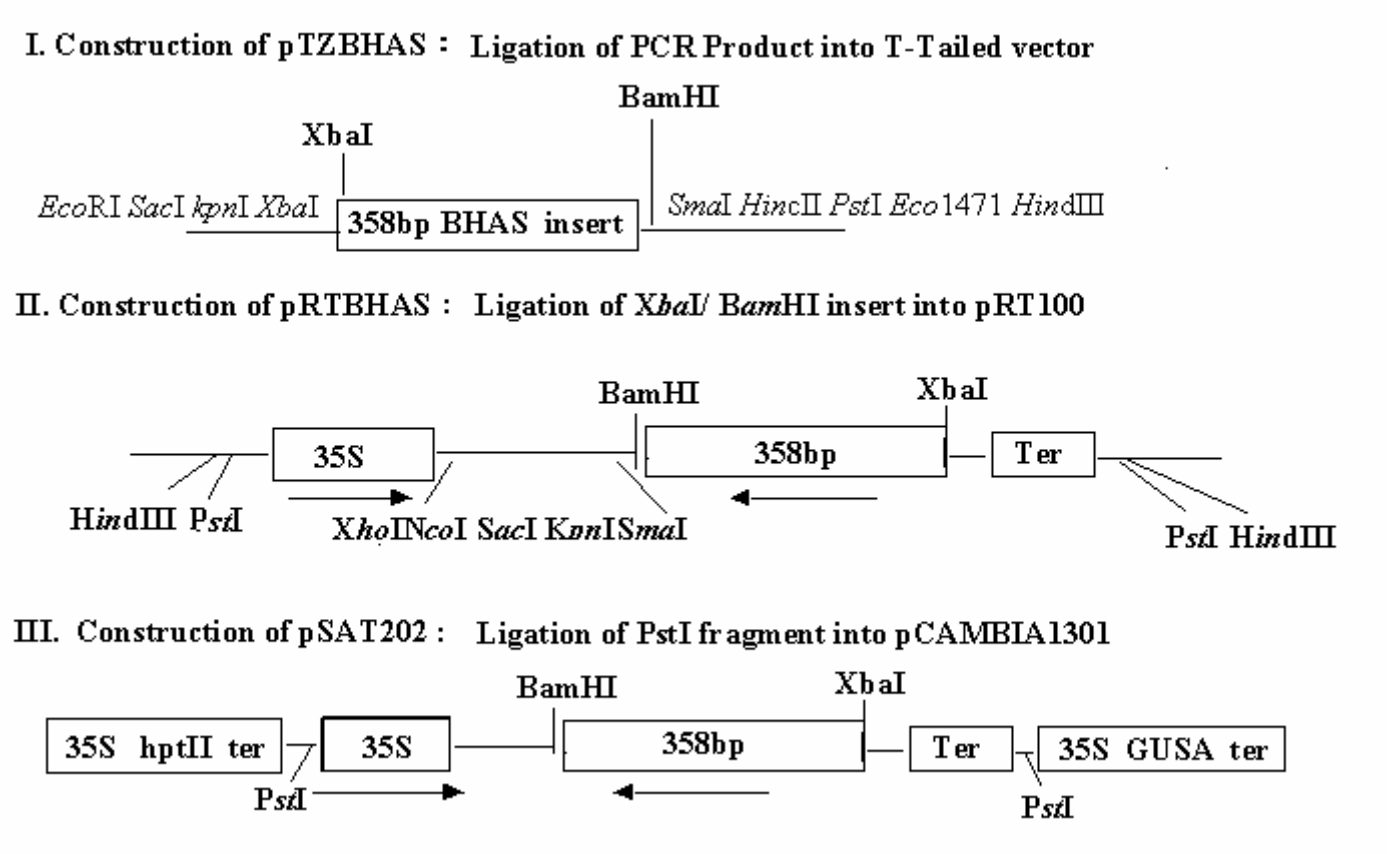


Figure 5.2.1 Designing of antisense construct pSAT202

Figure 5.2.2 Designing of co-suppression construct pSAT201

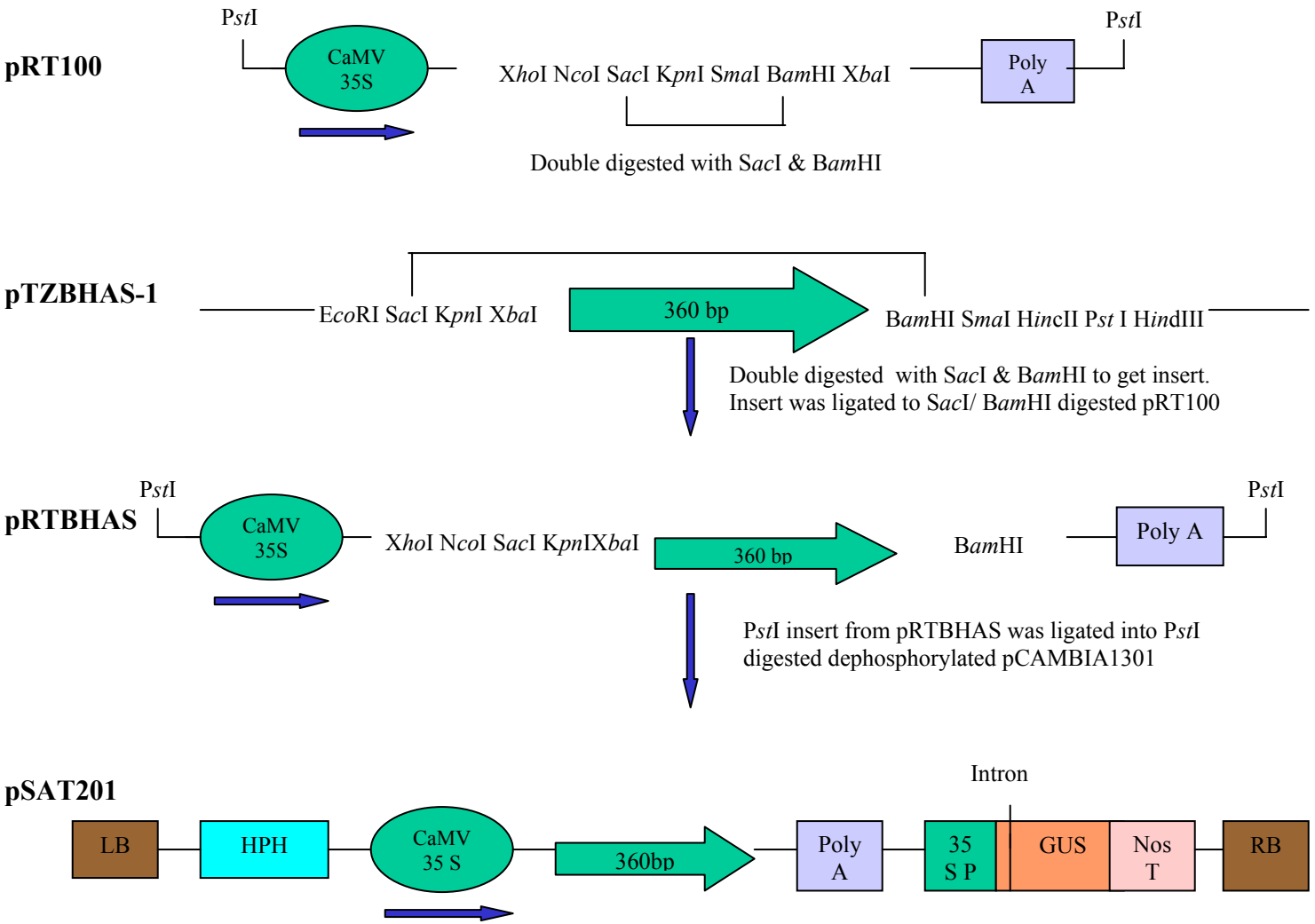
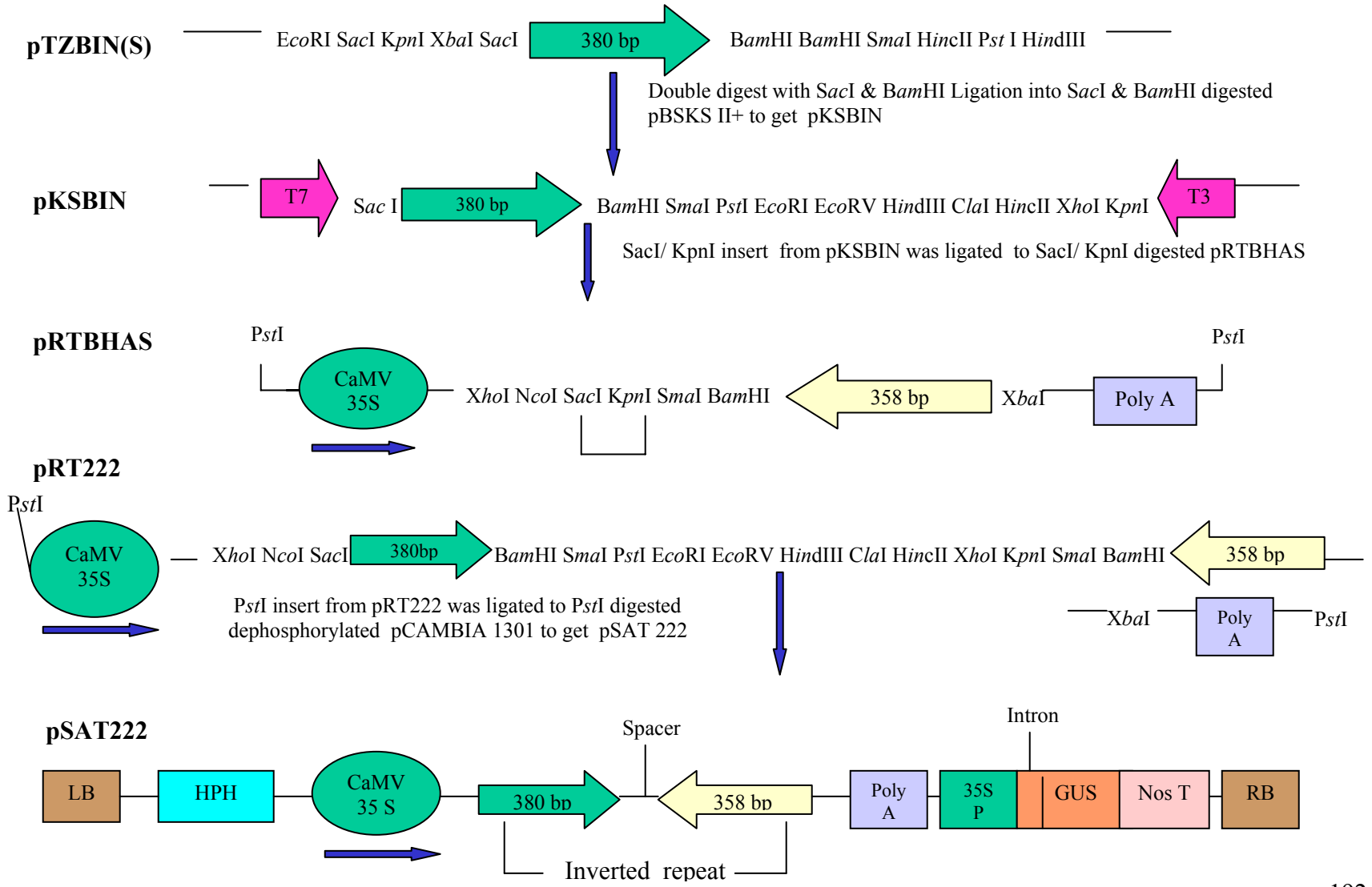


Figure 5.2.3 Designing of invert repeat (dsRNA) construct



5.2.2 Transformation

5.2.2.1 Plant material

Certified seeds of *Coffea canephora* cv S-274 and Cx R were purchased from Central Coffee Research Institute (CCRI), India. The seeds were germinated *in vitro* and hypocotyls and cotyledonary leaves were used as explants for induction of embryogenesis. The embryos thus obtained were used as explants for all transformation experiments.

5.2.2.2 Culture media and induction of somatic embryos

The seeds were surface sterilized using 1% sodium hypochlorite for 10 minutes, followed by 0.2% mercuric chloride for three minutes. The seeds were subsequently washed five times with sterile water. The seeds were kept for germination in half strength MS medium pH 5.6 (Murashige and Skoog, 1965) with 2% sucrose, 40 mgl $^{-1}$ cysteine HCl and 0.8% agar (Hi-media, India). The cultures were maintained at 25 ± 2 ^{o}C in dark for 45 days. Primary somatic embryos were obtained from leaf as well as hypocotyls from *in vitro* germinated seedlings. The infection medium, consisting of ½ MS salts and B5 vitamins, 2% sucrose and 200µM acetosyringone (Sigma, USA), was used for ultra sonication of somatic embryos. Co-cultivation medium consisted of ½ strength MS salts and B5 vitamins supplemented with 2% sucrose, 100µm acetosyringone, 0.8% agar for solidification, pH 5.6. Selection and regeneration medium consisted of different modifications of embryogenic (EG) medium (van Boxtel and Berthouly, 1996) comprising either of 0.5 mgl $^{-1}$ thidiazuron (TDZ), 60 µm AgNO3, 1 mgl^{-1} 2ip and 5µM triacontanol with hygromycin for selection. Cefotaxim was used along with potassium clavaluanate for killing *Agrobacterium*.

5.2.2.3 Agrobacterium culture

Agrobacterium tumefaciens strain EHA 101, carrying different constructs for posttranscriptional gene silencing, was used to transform *Coffea canephora.* The binary vector pCAMBIA 1301 contains the selectable marker gene hygromycin phosphotransferase and *gus* gene with a catalase intron under the control of CaMV 35S promoter and NOS terminator. All the constructs pSAT201, pSAT202, pSAT222 designed

for post-transcriptional gene silencing (PTGS) of coffee *N*-methyltransferase gene were maintained in LB medium with 100 mgl^{-1} kanamycin solidified with 1.5% Agar. For transformation experiments, the cultures were grown overnight in LB medium supplemented with 50 mgl^{-1} kanamycin at 28^0C and at 120 rpm (OD 0.5-1.00). The cells were harvested by centrifugation, resuspended in infection medium, and used for co-cultivation.

5.2.2.4 Transformation and Regeneration

Sonication assisted *Agrobacterium* mediated transformation (Trick and Finer, 1997) was modified and adopted for transformation. The somatic embryos and *Agrobacterium* strain were suspended in infection medium in 30ml Oakridge tube and ultrasonicated for 30 sec at 80% amplitude, incubated in a shaker at 120 rpm for 2 hrs. The embryos were blot dried and cultivated on co-cultivation medium for 2 days. Subsequently, the embryos were washed with sterile water three times, followed by 500 mgl^{-1} cephotaxim in sterile water. The embryos were blot dried and cultured on secondary embryogenic selection media containing modified EG medium supplemented with 3 mgl^{-1} hygromycin, 500 mgl^{-1} cefotaxim and 300 mgl^{-1} potassium clavaluanate. Every two months, the hygromycin concentration was increased from 3 mgl^{-1} to 5 mgl^{-1}, 10 mgl^{-1} and upto 20 mgl^{-1}.

5.2.2.5 Experimental Design

400-600 somatic embryos were inoculated with *A. tumefaciens* for each regeneration medium and 25 explants each were cultured as positive and negative control in selection and non-selection medium. The putative transgenic secondary embryos were selected on 20 mgl^{-1} hygromycin after a period of minimum 8 months of culture.

5.2.3 Extraction of total purine alkaloids

40-50 mg of fresh tissue (embryos / transformed plantlets) was used for alkaloid extraction. The tissue was crushed in 80% ethanol, using a mortar and pestle and the resultant slurry was homogenized using neutralized sand. The extract was centrifuged for 10 min at 8000 rpm and the supernatant collected. The pellet was rewashed with 2-3 ml of ethanol and the supernatant was collected after centrifugation. The extract was flash evaporated to dryness using rotavapour (Laborota-4000, Heidolph) and dissolved in 1ml of 80% ethanol prior to the estimation of caffeine and its metabolites by HPLC.

5.2.3.1 HPLC separation

The purine alkaloids from hygromycin resistant plantlets were extracted using 80% (v/v) ethanol as solvent. HPLC was performed on a μ Bondapak C_{18} column (5μ X 250mm) using isocratic solvent system (50mM sodium acetate: Methanol: Tetrahydrofuran in a ratio of 90: 9: 1) at a flow rate of 1ml /min. Parameters were controlled by a Shimadzu LC 10- AS liquid chromatograph equipped with a dual pump and a UV spectrophotometric detector (Model SPD –10 A) set at 270nm. The recorder, a Shimadzu C–R7A chromatopac, was set at a chart speed of 2.5 cm/min. Injection volume was 10μl, injected with Rheodyne 7125 injector. Peak identification was achieved by comparing with the retention time of standards (SIGMA, USA). HPLC analysis was repeated for each treatment atleast four times.

5.2.4 Molecular analysis of putative transformants

Genomic DNA was isolated from control and transgenic plants using GenElute Plant Genomic DNA isolation kit (Sigma, USA). PCR was performed for hygromycin phosphotransferase (*hptII*) gene using the following primers: forward 5′-CGGAAGTGCTTGACATTGG-3′ and reverse 5′- AGAAGAAGATGTTGGCGA-3′. PCR was performed with following conditions: initial denaturation at 94^0C for 5 min, followed by 35 cycles of 1 min denaturation at 94^0C, 1 min annealing at 50^0C and 1 min extension at 72^0C; with a final extension of 72^0C for 10 min.

5.2.5 Reverse Transcription-PCR

Total RNA from transgenic tissues was extracted using RNAqueous Kit and Plant RNA Isolation Aid (Ambion, USA). For RT-PCR, the primer binding regions chosen for CSF2 (5′-CGAGGAGTCCATGCATTTTT-3′) and CSR2 (5′-CCTCCTCAACCATGCACTTT-3′) had identical sequences for all NMT genes. Primers were also designed for the specific amplification of CmXRS1 (AB034699) and CaMTL1 (AB039725), based on the 5′ and 3′ UTRs for these genes.

CmXRS1: (F) 5′-ATGGAGCTCCAAGAAGTCCT-3′,
(R) 5′-CACTTATTCCTTTCCCCAACAC-3′
CaMTL1: (F) 5′-AGCTCCAAGAAGTCCTGCAT-3′,
(R) 5′-ACTCGGATAGAACCC CAACC-3′.

Equal quantities of total RNA from each sample was used for first strand cDNA synthesis with random decamer using enhanced AMV Reverse Transcriptase (Sigma, USA). The total RNA, dNTPS (final concentration of 1mM each) and random primer were mixed in 10µl volume and incubated at 70^0C for 10 min, followed by cooling on ice. To this 10µl volume, buffer for Reverse Transcriptase (final concentration of 1X), 1 unit of enhanced AMV Reverse Transcriptase, 1 unit of RNase inhibitor were added to a total of 20µl reaction volume. The reaction was placed at 25^0C for 15 min, followed by incubation at 50^0C for 1 hour. The conditions for PCR were as follows: Initial denaturation at 94^0C for 4 min, denaturation at 94^0C for 2 sec, annealing /extension temperature of 62^0C for 3 min for 30 cycles, followed by final extension at 62^0C for 10 min.

5.3 RESULTS AND DISCUSSION

5.3.1 Designing of antisense, sense and inverse repeat construct

5.3.1.1 Sub-cloning of NMT gene fragment

An ~ 360 bp PCR product was amplified using 2BHASF & 2BHASR primers and clone CS2B as the template for PCR. The PCR product was T/A cloned and the sense orientation of the resulting construct pTZBHAS-1 was verified by sequencing with M13 primer. The sequence of the clone is represented in Figure 5.3.1. The vector sequences are in italics and the primer regions are underlined.

Vector **Forward Primer**

*GCTCGGTACCTCGCGAATGCA**TCTAGA**TT*CCGA**TCTAGA**GCCAAGGTGAAACCTTTCCTT

X*ba*I X*ba*I

GAACAATGCATACGGGAATTGTTGCGGGCCAACTTGCCCAACATCAACAAGTGCATTAAAG
TTGCGGATTTGGGATGCGCTTCTGGACCAAACACACTTTTAACAGTGCGGGACATTGTGCA
GAGTATTGACAAAGTTGGCCAGGAAGAGAAGAATGAATTAGAACGTCCCACCATTCAGATT NMT gene fragment
TTTCTGAATGATCTTTTCCAAAATGATTTCAATTCGGTTTTCAAGTTGCTGCCAAGCTTCT
ACCGCAAACTCGAGAAAGAAAATGGACGCAAGATAGGATCGTGCCTAATAAGCGCAATGCC
TGGCTCTTTCTACG**ATCGAT**ACAG*AATC**GGATCCC**GGGCCCGTCGACTGCAGAGG*

Reverse Primer C*la*I B*am*HI **Vector**

CCTGCATGCAAGCTTTCCCTATAGTGAGTCGTATTAGAGCTTGGCGTAATCATGGTCATAG
CTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCA
TAAAGTGTAAAGCCTGGGGTG

Figure 5.3.1 Sequence of the clone pTZBHAS-1 M13F

The NMT gene fragment in the clone pTZBHAS was used in making three constructs as described in section 5.2.1.

5.3.2 Coffee Transformation

Sonication assisted *Agrobacterium* mediated transformation was adopted for coffee transformation. The three constructs pSAT201, pSAT202 and pSAT222 and control pCAMBIA 1301 were mobilized into *Agrobacterium tumefaciens* strain EHA 101. Somatic embryos and *Agrobacterium* strain were suspended in infection medium and ultra sonicated. Ultrasonicated somatic embryos were co-cultivated for 2 days, followed by culturing on secondary embryogenic selection media containing hygromycin. The putative transgenic secondary embryos were selected on 20 mgl^{-1} hygromycin after a period of minimum 8 months of culture. The transgenic secondary embryos further gave rise to transgenic plantlets (Figure 5.3.2).

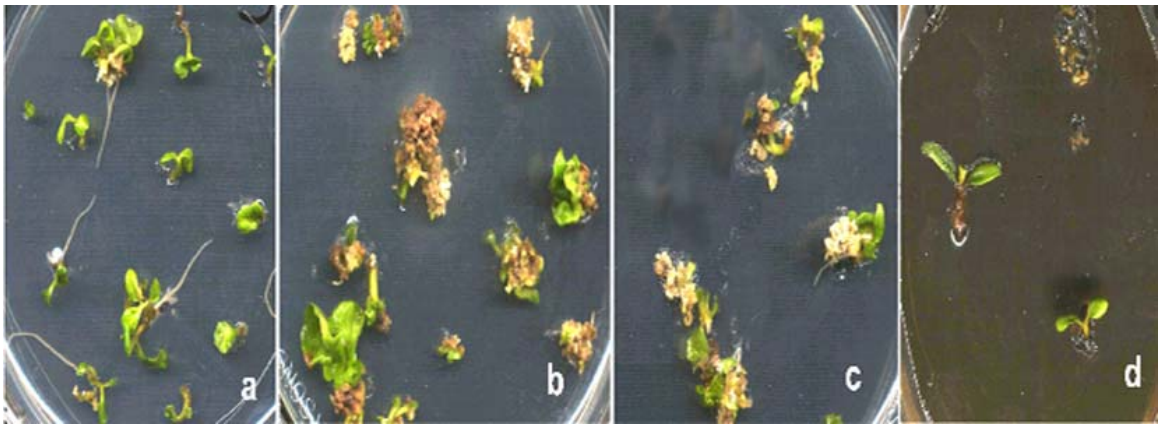


Figure 5.3.2 Transgenic plantlets containing various constructs
(a) pCAMBIA1301, (b) pSAT201, (c) pSAT202 and (d) pSAT222

5.3.3 Molecular analysis of transgenic plants

Transgenic nature of the regenerants was confirmed by PCR amplification of the 479 bp sized *hpt* II gene fragment from coffee genomic DNA isolated from some of the hygromycin resistant regenerant plants (Figure 5.3.3). A 479 bp PCR product was obtained with all the randomly selected transformants for various constructs.

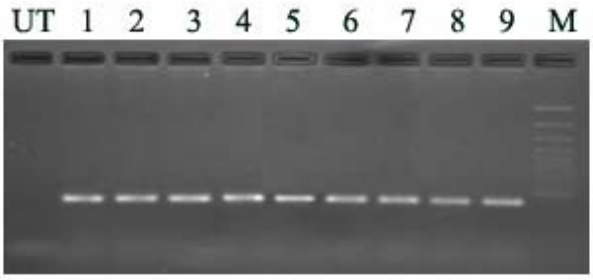


Figure 5.3.3 PCR amplification of *hpt* II gene from genomic DNA of transgenic plantlets

'UT': Genomic DNA from untransformed coffee plant used as control
1-3: DNA from control vector pCAMBIA 1301
4: DNA from sense transformants
5-7: DNA from antisense transformants
8-9: DNA from RNAi transformants
M: 3 Kb DNA Marker

5.3.4 Estimation of purine alkaloids

The purine alkaloids from hygromycin resistant embryonic tissues / transformed plantlets containing either control vector or one of the three PTGS constructs were estimated by HPLC. Various purine alkaloids viz., 7-methylxanthine, theobromine, caffeine and theophylline, were estimated. The purine alkaloids content varied in the transformants with different posttranscriptional gene silencing constructs (Table. 5.3.1). The values presented in Table are average of atleast 5 estimates ± Standard Deviation.

The sense construct resulted in 50% reduced theobromine content based on average value when compared to the control (pCAMBIA1301), but it had marginal effect (about 10% reduction) on the caffeine content in the independent transformants, suggesting that theobromine synthase was mainly affected. Despite this reduction in its precursor content, caffeine content was found to be 90% of that of control. The role of dual functional caffeine synthase in catalyzing both theobromine and caffeine synthesis has been demonstrated (Mizuno et al., 2003b; Uefuji et al., 2003). In the sense transformants, this enzyme must be active and must have substituted for main enzyme theobromine synthase in caffeine biosynthesis. Ogita et al., 2004 made similar observation for RNAi construct. The average 7-methylxanthine contents were the same for sense and control transformants.

The antisense construct resulted in reduced contents of caffeine, theobromine and 7-methylxanthine in the transformants when compared to the control. There was 80% reduction in the theobromine, 7-methylxanthine average content, and 90% reduction in caffeine content in antisense transformants when compared to control. Extremely low levels of the three purine alkaloids estimated suggest that antisense to homologous sequence shared by several members of the multigene family is highly effective in suppressing the expression of the genes. This was in contrast to the observation of Tada et al., (2003), who opined that antisense strategy is not suitable for completely suppressing the expression of genes comprising a multigene family.

Table 5.3.1 Purine alkaloid contents in transgenic plantlets of *Coffea canephora*[1]

Construct		7-methylxanthine	Theobromine	Caffeine	Theophylline
pCAMBIA1301 (Control)	Line # 1	0.032 ± 0.0034	0.228 ± 0.0082 *	0.770 ± 0.0081	0.023 ± 0.0034
	Line # 2	0.035 ± 0.0031	0.098 ± 0.0024	0.784 ± 0.0042 *	0.033 ± 0.0018 *
	Line # 3	0.053 ± 0.0836 *	0.112 ± 0.0027	0.754 ± 0.0035	0.022 ± 0.0025
	Average	**0.040 ± 0.0115**	**0.146 ± 0.017**	**0.769 ± 0.0153**	**0.026 ± 0.0063**
Sense (Co-suppression)	Line # 1	0.032 ± 0.0029	0.084 ± 0.0060 *	0.675 ± 0.0196	0.028 ± 0.0030 *
	Line # 2	0.042 ± 0.0028 *	0.074 ± 0.0035	0.705 ± 0.0055 *	0.023 ± 0.0030
	Average	**0.037 ± 0.0070**	**0.079 ± 0.0072**	**0.690 ± 0.0215**	**0.026 ± 0.0035**
Antisense	Line # 1	0.008 ± 0.0018	0.027 ± 0.0022	0.055 ± 0.0057	0.000 ± 0.0000
	Line # 2	0.010 ± 0.0005 *	0.034 ± 0.0030 *	0.087 ± 0.0027 *	0.000 ± 0.0000
	Line # 3	0.007 ± 0.0056	0.021 ± 0.0020	0.072 ± 0.0023	0.000 ± 0.0000
	Average	**0.008 ± 0.0016**	**0.027 ± 0.0067**	**0.067 ± 0.0161**	**0.0**
RNAi	Line # 1	0.042 ± 0.0027	0.182 ± 0.0033	0.000 ± 0.0000	0.000 ± 0.0000
	Line # 2	0.043 ± 0.0037	0.157 ± 0.0042	0.013 ± 0.0029	0.000 ± 0.0000
	Line # 3	0.070 ± 0.1012 *	0.196 ± 0.0025 *	0.000 ± 0.0000	0.000 ± 0.0000
	Average	**0.052 ± 0.0155**	**0.178 ± 0.0197**	**0.004**	**0.0**

[1] Results are expressed as μg 100^{-1} mg dry weight. Average values of atleast four estimates are provided
*Significant at 1% level over general mean ± SED (standard error deviation)

The RNAi (invert repeat) construct was the most efficient in reducing the caffeine content; almost 100% reduction in caffeine content was observed over control. RNAi was shown to be more effective than co-suppression and antisense strategies (Wesley et al., 2001; Wang and Wagner, 2003). However, there was approximately 20% increase (on average) in theobromine and 7-methylxanthine contents in the RNAi transformants over control (in lines #2 and #3). One would have expected this accumulation of immediate precursors if the RNAi construct were specific for one enzyme. However, with the coding sequence used in this construct being common to all NMT genes including the genes encoding for first two methylation steps i.e. 7-methylxanthosine synthase and theobromine synthase, accumulation of 7-methylxanthine and theobromine is surprising. In this study, 100% reduction of caffeine was observed with RNAi when compared to 30% to 50% reduction reported by Ogita et al., 2004. This could primarily be because of tighter loop of about 90 bp between the two arms of the invert repeat when compared to the 517-bp GUS gene fragment as spacer used by them. It is known that tightness of the hairpin loop can contribute to enhance RNAi silencing (Wesley et al., 2001). The use of homologous coding sequence could have also been one of the reasons. But the fact that accumulation of 7-methylxanthine and theobromine was observed, does not favour this proposition.

Theophylline was observed only in control and sense transformants, while it was not detected with antisense and RNAi constructs. The fact that theophylline content was detected only when considerable amounts of caffeine accumulated, suggests that the caffeine degradative pathway was active in transgenic plantlets of *Coffea canephora*. It is known that theophylline is an intermediate in the catabolism of caffeine rather than *de novo* caffeine biosynthesis (Ito et al., 1997).

5.3.5 RT-PCR experiments

RT-PCR (Figure 5.3.4) confirmed that the transcript levels decreased in the three constructs over control, with RNAi being most efficient in down regulating. The primers CSF2 and CSR2 were not specific and the accumulated transcripts might comprise of several NMT genes having the conserved primer binding sites. However, RT-PCR with

specific forward and reverse primers for CmXRS1 catalyzing the first methylation step (Mizuno et al., 2003a) revealed the degradation of transcripts for CmXRS1.

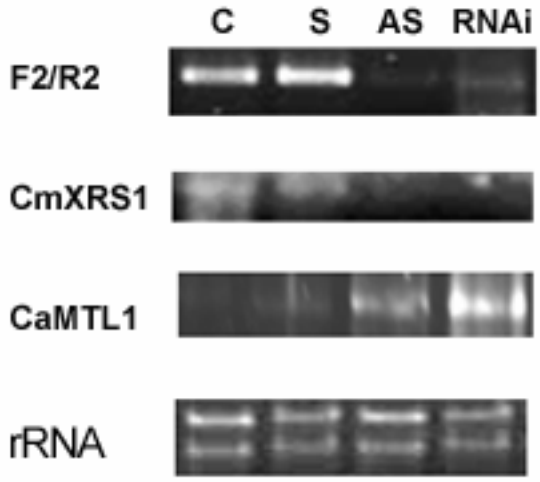


Figure 5.3.4 Transcript accumulation in transgenic plantlets with various constructs

Random decamer was used to prime the Reverse Transcription reaction with equal quantity of total RNA from each sample. Total RNA from the representative plantlets for each construct was isolated. Ethidium bromide staining of rRNA reveals RNA integrity and quantity. 'C' represents Control vector; whereas 'S', 'AS' and 'RNAi' represent transgenics with sense, antisense and RNAi construct, respectively. F2/R2: Transcripts obtained with CSF2 and CSR2 primers that are shared by all the NMT genes. CmXRS1 and CaMTL1: Transcripts obtained with specific primers for the corresponding genes.

Differences in the reduction rates of purine alkaloids content in the transgenic lines and that expected from the transcript levels for NMT genes were observed. Ogita et al., 2004 also observed similar differences and attributed these differences to the synergistic RNAi suppression of precursors or activation of bypass pathways. To test the hypothesis for possible induction of alternate pathways or activated transcription of other genes in the caffeine biosynthetic pathway, transcript accumulation of CaMTL1 gene, reported to be expressed in various coffee tissues (Uefuji et al., 2003), was analyzed. However, the

substrate specificity of the enzyme encoded by this gene is not known. Surprisingly, CaMTL1 transcript accumulation increased proportionally with the decrease in caffeine content in the transformants. The highest accumulation of CaMTL1 transcript was observed with RNAi and was least in control (Fig. 5.3.4), suggesting transcriptional activation of CaMTL1 gene upon silencing of other NMT genes and possible involvement of parallel or bypass pathways. Activation of other genes also cannot be ruled out. The primer sequences for CSF2 and CSR2 were conserved in all the genes reported so far including CaMTL1 and possibly in the unknown genes also. The detection of transcripts in the case of RNAi despite of zero caffeine might be due to the accumulation of transcripts for CaMTL1 or similar unknown genes.

5.3.6 Summary and conclusions

From this study, it can be observed that though the co-suppression and RNAi constructs were not specific for a single NMT gene, transformants were obtained that were mainly affected in only one of the three NMTs. This is in contrast to the earlier report (Ogita et al., 2004), wherein non-specific RNAi effect was observed with specific RNAi construct. They observed a concomitant reduction of theobromine and caffeine contents to a range between 30 and 50% of that of control. The non-specificity was assigned to the possible effects of RNAi spreading from the initiator region into adjacent regions of the target gene, and among genes whose sequences were closely related (transitive spreading). Despite the RNAi construct targeted for all NMT genes sharing highly conserved region, transformants that were mainly affected in third methylation step of the three NMTs were obtained. Such unexpected results are not unknown in PTGS. Allen et al., (2004) observed the accumulation of rare alkaloids several steps upstream to the target gene codeinone reductase (COR) in opium alkaloid biosynthesis pathway, which was silenced with chimeric RNAi construct; but surprisingly did not observe the accumulation of immediate precursors as expected. Chintapakorn and Hamill (2003) observed that reduced PMT activity and nicotine levels are often correlated with significantly raised anatabine levels in antisense-*PMT* transformed roots of *N. tabacum*. The differences in the efficiencies and

specificity of silencing observed in this study may be due to the different mechanisms of action involved in different PTGS phenomena.

Although the basic concept of the application of transgene-based RNAi for crop improvement has been developed, further studies are needed for its wider application (Kusaba, 2004). The 100% elimination of caffeine but accumulation of theobromine and 7-methylxanthine in spite of using highly homologous sequence for NMT gene family cannot be explained based on this studies alone. The possible transcriptional activation of other genes as observed in our studies suggest that further studies need to be carried out before PTGS can be used as tool for metabolic engineering of caffeine biosynthetic pathway.

In this study, the three PTGS silencing strategies resulted in production of transgenic plants with different purine alkaloids profiles. However, the silencing was specific for one of the coffee NMTs, in contrast to the expected simultaneous silencing of all the NMT genes. The reasons for this deviation are unknown and possibly needs to be ascertained. In view of these uncertainties, the use of other strategies for regulating caffeine content in plants needs to be explored. Naturally decaffeinated arabica coffee plants have been reported, where the low caffeine content observed was not due to enhanced degradation of caffeine, but more likely due to possible mutation in caffeine synthase gene (Silvarolla et al., 2004). These plants accumulate theobromine indicating that these plants might be deficient in caffeine synthase that acts on theobromine. Understanding the regulation for this naturally decaffeinated coffee might lead to formulation of more effective strategies for down regulating caffeine.

6. Conclusions and future perspectives

The major findings of this investigation and the future prospects are highlighted as follows:

1. Cloning and sequencing of theobromine synthase -1 gene from *Coffea canephora* revealed that NMT gene comprises of four exons interrupted by three introns. The sequencing of several clones revealed high % of sequence similarities for the coding regions, but the introns varied in length and sequence. It was concluded that coffee plant possess multiple number of NMT genes with similarities in exons, but variations in introns.

2. The promoter for theobromine synthase-1 gene was isolated and the GUS assays proved that the isolated promoter fragment could drive gene expression. Sequence analyses of promoter region revealed the presence of several *cis*-elements involved in gene expression. The role of these *cis*-elements in regulation of coffee NMT gene needs to be evaluated by performing promoter deletion analyses. Defining the combination of transcription factors and promoter elements that regulate the expression of individual members of gene families will be very useful in order to understand how diversification of regulation has been achieved.

3. Several genomic clones for NMT genes were screened for polymorphism and sequenced. The clones sequenced showed remarkable conservation in the promoter and coding regions, but exhibited some differences in the introns. The allelic nature of these clones for theobromine synthase -1 needs to be verified. The mapping of different NMT genes would also provide information regarding the nature of gene duplications i.e. tandem or segmental duplications.

4. The results from this study suggest that NMT gene family has arisen by recent gene duplications from ancestral gene, followed by their functional divergence. The gene structure in terms of intron positions, phases and intron /exon splice joints was highly conserved for all the genomic clones obtained in this study. This study also suggested the presence of pseudogenes in the NMT family. The gene structure for other members of the

NMT gene family needs to be studied. Traditional methods for tracking gene expression, such as northern hybridization, often fail to distinguish between similarly sized members of gene families and suffer from cross-hybridization between similar sequences. The questions of differential expression, localization and gene function within families can be addressed by identifying individual members of gene families and their products. Recent transcript profiling of gene families has revealed that in most families all members are differentially expressed at some time, including tissues specificities, cell specific and treatment specific induction patterns. Transcript profiling of coffee NMT gene family will reveal the spatial and temporal differential expression of the individual members of the family.

5. The promoter regions from various *Coffea* species were cloned and sequenced. The sequence analyses revealed high degree of promoter sequence conservation. The high degree of promoter sequence conservation was attributed to relatively recent origin of Coffee. Despite conservation in promoter sequences, the reasons for different rates of caffeine biosynthesis in various *Coffea* species need to be investigated.

6. The posttranscriptional gene silencing of coffee NMT genes using highly conserved sequence of coding region resulted in transgenics with silencing of the specific methylation step, in contrast to the expected simultaneous silencing of all the genes that share the targeted region. The reasons for this anomaly in results need to be studied.

Chapter VII

References

Allen, R.S., A.G. Millgate, J.A. Chitty, J. Thisleton, J.A.C. Miller, A.J. Fist, W.L. Gerlach, and P.J. Larkin. 2004. RNAi-mediated replacement of morphine with the nonnarcotic alkaloid reticuline in opium poppy. Nat. Biotech. 22:1559-1565.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J.H. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

An., G. 1987. Binary Ti vectors for plant transformation and promoter analysis. Methods Enzymol. 153:292-305.

An, G., P. Ebert, A. Mitra, and S. Ha. 1988. Binary vectors, p. A3/1-A3/19, *In* S. B. Gelvin and R. A. Schilperoot, eds. Plant Molecular Biology Manual. Kluwer Academic Publishers, Dordrecht.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815.

Ashihara, H., and A. Crozier. 1999a. Biosynthesis and metabolism of caffeine and related purine alkaloids in plants. Adv. Bot. Res. 30:118-205.

Ashihara, H., and A. Crozier. 1999b. Biosynthesis and catabolism of caffeine in low-caffeine-containing species of *Coffea*. J. Agric. Food Chem. 47:3425-3431.

Ashihara, H., and A. Crozier. 2001. Caffeine: a well-known but little mentioned compound in plant science. Trends Plant Sci. 9:407-413.

Ashihara, H., and H. Kubota. 1986. Patterns of adenine metabolism and caffeine biosynthesis in different parts of tea seedlings. Physiol. Plantarum 68:275-281.

Ashihara, H., A.M. Monterio, F.M. Gillies, and A. Crozier. 1996a. Biosynthesis of caffeine in leaves of coffee. Plant Physiol. 111:747-753.

Ashihara, H., A.M. Monterio, T. Moritz, F.M. Gillies, and A. Crozier. 1996b. Catabolism of caffeine and related purine alkaloids in leaves of *Coffea arabica* L. Planta 198:334-339.

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

Aubourg, S., A. Lecharny, and J. Bohlmann. 2002. Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana*. Mol. Gen. Genom. 267:730-745.

Barre, P., S. Akaffou, J. Louarn, A. Charrier, S. Hamon, and M. Noirot. 1998. Inheritance of caffeine and heteroside contents in an interspecific cross between a cultivated

coffee species *Coffea liberica* var *dewevrei* and a wild species caffeine free *C. pseudozanguebariae*. Theor. App. Genet. 96:306-311.

Barta, E., E. Sebestyen, T.B. Palfy, G. Toth, C.P. Ortutay, and L. Patthy. 2005. DoOP: Databases of Orthologous Promoters, collections of clusters of orthologous upstream sequences from chordates and plants. Nucleic Acids Res. 33:D86–D90.

Baruah, A., V. Naik, P.S. Hendre, R. Rajkumar, P. Rajendrakumar, and R.K. Aggarwal. 2003. Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications. Mol. Ecol. Notes 3:647–650.

Bass, B.L. 2000. Double-stranded RNA as a template for gene silencing. Cell 101:235–238.

Baumann, T.W., R. Koetz, and P. Morath. 1983. N-methyltransferase activities in suspension cultures of *Coffea arabica* L. Plant Cell Rep. 2:33-35.

Baumann, T.W., and L. Rohring. 1989. Formation and intracellular accumulation of caffeine and chlorogenic acid in suspension cultures of *Coffea arabica*. Phytochemistry 28:2667-2669.

Begum, B., C.M. Hasan, and M.A. Rashid. 2003. Caffeine from the mature leaves of *Coffea bengalensis*. Biochem. Syst. Ecol. 31:1219-1220.

Bernstein, E., A.A. Caudy, S.M. Hammond, and G.J. Hannon. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409:363–366.

Birnboim, H.C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513-1520.

Boldt, R., and R. Zrenner. 2003. Purine and pyrimidine biosynthesis in higher plants. Physiol. Plantarum 117:297-304.

Bolle, C., S. Sopory, T. Lubberstedt, R.G. Herrmann, and R. Oelmuller. 1994. Segments encoding 5'- untranslated leaders of genes for thylakoid proteins contain *cis*-elements essential for transcription. Plant J. 6:513-523.

Brown, J.W.S., and C.G. Simpson. 1998. Splice site selection in plant pre-mRNA splicing. Ann. Rev. Plant Physiol. Plant Mol. Biol. 49:77-95.

Brown, J.W., P. Smith, and C.G. Simpson. 1996. Arabidopsis consensus introns sequences. Plant Mol. Biol. 32:531-535.

Busk, P.K., and M. Pages. 1998. Regulation of abscisic-induced transcription. Plant Mol. Biol. 37:425–435.

Bustos, M.M., M.J. Guiltinan, J. Jordano, D. Begum, F.A. Kalkan, and T.C. Hall. 1989. Regulation of β-glucuronidase expression in transgenic tobacco plants by an A/T-rich, cis-acting sequence found upstream of a French bean β-phaseolin gene. Plant Cell 1:839–853.

Campa, C., S. Doulbeau, S. Dussert, S. Hamon, and M. Noirot. 2005. Diversity in bean caffeine content among wild Coffea species: Evidence of a discontinuous distribution. Food Chem. 91:633-637.

Caillieux, E., E. Bullier, C. Moreau, C. Campa, M. Noirot, S. Hamon, and A. de Kochko. 2004. Some genes encoding enzymes involved in the caffeine bioynthesis of *Coffea pseudozanguebariae* are regulated by an incomplete splicing of their pre-messenger RNA, p. 688-691 In: Proceedings 20th International Scientific Colloquim on Coffee. ASIC, Bangalore, India

Cesarino, I., and P. Mazzafera P. 2005. Catabolismo de cafeina em plantas de café: Evidencias da participacao de citocromo p450. In: Proceedings Xth Brazilian Congress on Plant Physiology, 11-17 September, 2005, Recife, PE.

Charrier, A., and J. Berthaud. 1985. Botanical classification of coffee, p. 13-47, *In* Clifford, et al., eds. Coffee- Botany, biochemistry and production of beans and beverage. The Avi Publishing Company, Inc., Westport, Connecticut.

Chelvarajan, R.L., F.F. Fannin, and L.P. Bush. 1993. Study of nicotine demethylation in *Nicotiana otophora*. J. Agric. Food Chem. 41:858-862.

Chen, W., N.J. Provart, J. Glazebrook, F. Katagiri, H.-S. Chang, T. Eulgem, F. Mauch, S. Luan, and G. Zou. 2002. Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell 14:559-574.

Chen, X., and W. Ray. 1997. Direct amplification of unknown genes and fragments by Uneven polymerase chain reaction. Gene 185:195-199.

Chevalier, A. 1947. 'Les caféiers du globe III. Systématique des caféiers et faux caféiers. Maladies et insects nuisibles' Encyclopedie biologique, Vol. 28. Paul Le Chevalier, Paris.

Chintapakorn, Y., and J.D. Hamill. 2003. Antisense-mediated down-regulation of putrescine *N*-methyltransferase activity in transgenic *Nicotiana tabacum* L. can lead to elevated levels of anatabine at the expense of nicotine. Plant Mol. Biol. 53:87–105.

Choi, K.-B., T. Morishige, N. Shitan, K. Yazaki, and F. Sato. 2002. Molecular cloning and characterization of Coclaurine *N*-Methyltransferase from cultured cells of *Coptis japonica*. J Biol. Chem. 277:830–835.

Clifford, M.N. 1985. Chemical and physical aspects of green coffee and coffee products, p. 305–374, *In* M. N. Clifford and K. C. Willson, eds. Coffee: Botany, Biochemistry and Production of Beans and Beverage, Croom Helm, London.

Clifford, M.N., and T. Kazi. 1987. The influence of coffee bean maturity on the content of chlorogenic acids, caffeine and trigonelline. Food Chem. 26:59-69.

Colgan, D.F., and J.M. Manley. 1997. Mechanism and regulation of mRNA polyadenylation. Genes Dev. 11:2755-2766.

Conaway, J.W., L. Florens, S. Satoa, C. Tomomori-Sato, T.J. Parmely, T. Yao, S.K. Swanson, C.A.S. Banks, M.P. Washburn, and R.C. Conaway. 2005. The mammalian Mediator complex. FEBS Lett. 579:904–908.

Cormack, R.S., and I.E. Somssich. 1997. Rapid amplification of genomic ends (RAGE) as a simple method to clone flanking genomic DNA. Gene 194:273-276.

Cros, J., M.C. Combes, N. Chabrillange, C. Duperray, des Angles A. M., and S. Hamon. 1995. Nuclear DNA content in the subgenus Coffea (Rubiaceae): inter- and intra-specific variation in African species. Can. J. Bot. 73:14–20.

Cros, J., M.C. Combes, P. Trouslot, F. Anthony, S. Hamon, A. Charrier, and P. Lashermes. 1998. Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. Mol. Phylogenet. Evol. 9:109-117.

Czarnecka, E., J.C. Ingersoll, and W.B. Gurley. 1992. AT-rich promoter elements of soybean heat shock gene Gmhsp 17.5E bind two sets of nuclear proteins *in vitro*. Plant Mol. Biol. 19:985–1000.

Dalmay, T., A. Hamilton, S. Rudd, S. Angell, and D.C. Baulcombe. 2000. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101:543–553.

Dean, C., M. Favreau, D. Bond-Nutter, J. Bedbrook, and P. Dunsmuir. 1989. Sequences 5' to translation start regulate expression of petunia rbcS genes. Plant Cell 1:209–215.

Devon, R.S., D. Proteous, and A.J. Brookes. 1995. Splinkerettes-improved vectorettes for greater efficiency in PCR walking. Nucleic Acids Res. 23:1644-1645.

Durbin, M.L., B. McCaig, and M.T. Clegg. 2000. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol. Biol. 42:79-92.

Dussert, S., P. Lashermes, F. Anthony, C. Montagnon, P. Trouslot, M.C. Combes, J. Berthaud, M. Noirot, and S. Hamon. 1999. Le caféier, *Coffea canephora*, p. 175–194, *In* P. Hamon, et al., eds. Diversité génétique des plantes tropicales cultivées. CIRAD, Montpellier, France.

Elbashir, S.M., J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, and T. Tuschl. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411:494–498.

Eulgem, T., P.J. Rushton, E. Schmelzer, K. Hahlbrock, and I.E. Somssich. 1999. Early nuclear events in plant defence signalling: rapid gene activation by WRKY transcription factors. EMBO J.18:4689-4699.

Eulgem, T., P.J. Rushton, S. Robatzek, and I.E. Somssich. 2000. The WRKY superfamily of plant transcription factors. Trends Plant Sci. 5:199-498.

European Union Chromosome 3 Arabidopsis Sequencing Consortium; The Institute for Genomic Research; Kazusa DNA Research Institute. 2000. Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. Nature 408: 820-821.

Facchini, P.J. 2001. Alkaloid biosynthesis in plants: biochemistry, cell biology, molecular regulation, and metabolic engineering applications. Annu. Rev. Plant Phys. Plant Mol. Biol. 52:29–66.

Fagard, M., S. Boutet, J.-B. Morel, C. Bellini, and H. Vaucheret. 2000. AGO-1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc. Natl. Acad. Sci. U.S.A. 97:11650–11654.

Fernandez, R., and A. Menendez. 2003. Transient gene expression in secondary somatic embryos from coffee tissues electroporated with the genes *gus* and *bar*. Electronic J Biotechnol 6:29-38.

Flament, I. 2002. Coffee flavour chemistry. John Wiley and Sons, Chichester, England.

Force, A., M. Lynch, F.B. Pickett, A. Amores, Y.L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary,degenerative mutations. Genetics 151:1531–1545.

Fu, H., S.Y. Kim, and W.D. Park. 1995. High-level tuber expression and sucrose inducibility of a potato *Sus4* sucrose synthase gene require 5' and 3' flanking sequences and the leader intron. Plant Cell 7:1387-1394.

Fujimori, N., T. Suzuki, and A. Ashihara. 1991. Seasonal variations in biosynthetic capacity for the synthesis of caffeine in tea leaves. Phytochemistry 30:2245-2248.

Fukusaki, E., K. Kawasaki, S. Kajiyama, C. An, K. Suzuki, Y. Tanaka, and A. Kobayashi. 2004. Flower color modulations of *Torenia hybrida* by downregulation of chalcone synthase genes with RNA interference. J. Biotechnol. 111:229-240.

Furuya, T., Y. Orihara, K. Koge, and Y. Tsuda. 1989. Effects of auxins and immobilization on growth and purine alkaloids production in coffee (*Coffea arabica*) cell suspension culture. Plant Tissue Cult. Lett. 6:148-151.

Futterer, J., and T. Hohn. 1996. Translation in plants-rules and exceptions. Plant Mol. Biol. 32:159-189.

Gaitan, A.L. 1998. Cloning and evaluation of gene promoters in *Coffea* spp. Ph.D. dissertation. Cornell University.

Gillies, F.M., G.I. Jenkins, H. Ashihara, and A. Crozier. 1995. In vitro biosynthesis of caffeine: stability of N-methyltransferase activity of cell-free preparations from liquid endosperm of *Coffea arabica*. In: Proceedings of the 16th International Symposium on Coffee Science, Kyoto,ASIC, Paris:559-565.

Gokulakrishnan, S., K. Chandraraj, and N.G. Sathyanarayana. 2005. Microbial and enzymatic methods for the removal of caffeine. Enzyme Microb. Tech. 37:225–232.

Grotewold, E., M. Chamberlin, M. Snook, B. Siame, L. Butler, J. Swenson, S. Maddock, G. St. Clair, and B. Bowen. 1998. Engineering secondary metabolism in maize cells by ectopic expression of transcription factors. Plant Cell 10:721-740.

Guivarc'h, A., J.C. Caissard, A. Azmi, T. Elmayan, D. Chriqui, and M. Tepfer. 1996. *In situ* detection of expression of the *gus* reporter gene in transgenic plants: ten years of blue genes. Transgenic Res. 5:281-288.

Guo, H., and S.P. Moose. 2003. Conserved noncoding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. Plant Cell 15:1143-1158.

Guo, M., M.A. Rupe, O.N. Danilevskaya, X. Yang, and Z. Hu. 2003. Genome-wide mRNA profiling reveals heterochronic allelic variation and a new imprinted gene in hybrid maize endosperm. Plant J. 36:30–44.

Guo, M., M.A. Rupe, C. Zinselmeier, J. Habben, B.A. Bowen, and O.S. Smith. 2004. Allelic variation of gene expression in maize hybrids. Plant Cell 16:1707–1716.

Haas, B.J., N. Volfovsky, C.D. Town, M. Troukhan, N. Alexandrov, K.A. Feldmann, R.B. Flavell, O. White, and S.L. Salzberg. 2002. Full-length messenger RNA sequences greatly improve genome annotation. Genome Biol. 3:reviews0029.1–0029.12.

Haberer, G., T. Hindemitt, B.C. Meyers, and K.F.X. Mayer. 2004. Transcriptional similarities, dissimilarities, and conservation of *cis*-elements in duplicated genes of Arabidopsis. Plant Physiol. 136:3009-3022.

Hamilton, A.J., and Baulcombe, D.C. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286:950-952.

Hamilton, A.J., S. Brown, H. Yuanhai, M. Ishizuka, A. Lowe, A.G.A. Solis, and D. Grierson. 1998. A transgene with repeat DNA causes high frequency, post-transcriptional suppression of ACC-oxidase gene expression in tomato. Plant J. 15:737–746.

Hamilton, A.J., Voinnet, O., Chappell,L., and Baulcombe, D.C. 2002. Two classes of short interfering RNA in RNA silencing. EMBO J. 21:4671-4679.

Hashimoto, T., and Y. Yamada. 1994. Alkaloid biogenesis: Molecular aspects. Annu. Rev. Plant Physiol. Plant Mol. Biol. 45:257-285.

Higo, K., Y. Ugawa, M. Iwamoto, and T. Korenaga. 1999. Plant cis-acting regulatory DNA Elements (PLACE) database. Nucleic Acids Res. 27:297-300.

Hollingsworth, R.G., J.W. Armstrong, and E. Campbell. 2003. Pest control: caffeine as a repellent for slugs and snails. Nature 417:915-916.

Hua, X.J., B. van de Cotte, M.V. Montagu, and N. Verbruggen. 2001. The 5' untranslated region of the *At-P5R* gene is involved in both transcriptional and post-transcriptional regulation. Plant J. 26:157-169.

Ito, E., and H. Ashihara. 1999. Contribution of purine nucleotide biosynthesis *de novo* to the formation of caffeine in young tea *(Camellia sinensis*) leaves. J. Plant Physiol. 154:145-151.

Ito, E., A. Crozier, and H. Ashihara. 1997. Theophylline metabolism in higher plants. Biochim. Biophys. Acta 1336:323–330.

Ifuku, K., Y. Yamamoto, and F. Sato. 2003. Specific RNA interference in *psb*P genes encoded by a multigene family in *Nicotiana tabacum* with a short 3'-untranslated sequence. Biosci. Biotech. Biochem. 67:107-113.

Inouye, M. 1988. Antisense RNA: its functions and applications in gene regulation- a review. Gene 72:25-34.

Jefferson, R.A., T.A. Kavanagh, and M.W. Bevan. 1987. GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J 6:3901-3907.

Joshi, C.P. 1987. An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res. 15:6643-6653.

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

Joshi, C.P., H. Zhou, X. Huang, and V.L. Chiang. 1997. Context sequences of translation initiation codon in plants. Plant Mol. Biol. 35:993-1001.

Kalberer, P. 1965. Breakdown of caffeine in leaves of *Coffea arabica* L. Nature 205:597-598.

Kato, M., T. Kanehara, H. Shimizu, T. Suzuki, F.M. Gillies, A. Crozier, and H. Ashihara. 1996. Caffeine biosynthesis in young leaves of *Camellia sinensis*: in vitro studies on N-methyltransferase activity involved in the conversion of xanthosine to caffeine. Physiol. Plantarum 98:629-636.

Kato, A., A. Crozier, and H. Ashihara. 1998. Subcellular localization of the N-3 methyltransferase involved in caffeine biosynthesis in tea. Phytochemistry 48:777-779.

Kato, M., K. Mizuno, T. Fujimura, M. Iwama, M. Irie, A. Crozier, and H. Ashihara. 1999. Purification and characterisation of caffeine synthase from tea leaves. Plant Physiol. 120:579-586.

Kato, M., K. Mizuno, A. Crozier, T. Fuzimura, and H. Ashihara. 2000. Caffeine synthase gene from tea leaves. Nature 406:956-957.

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

Kawagoe, Y., and N. Murai. 1992. Four distinct nuclear proteins recognize in vitro the proximal promoter of the bean seed storage protein ß-phaseolin gene conferring spatial and temporal control. Plant J 2:927-936.

Keya, C.A., A. Crozier, and H. Ashihara. 2003. Inhibition of caffeine biosynthesis in tea (*Camellia sinensis*) and coffee (*Coffea arabica*) plants by ribavirin. FEBS Lett. 554:473-477.

Kilstrup, M., and K.N. Kristiansen. 2000. Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. Nucleic Acids Res. 28:e55.

Klug, A. 2005. Introduction to eukaryotic transcription and chromatin. FEBS Lett. 579: 890–891.

Kobs, G. 1997. Cloning blunt-end DNA fragments into the pGEM-T vector systems. Promega Notes 62:15-17.

Komarnytsky, S., and N. Borisjuk. 2003. Functional analysis of promoter elements in plants, p. 113-141, *In* J. K. Setlow, ed. Genetic Engineering, Vol. 25. Kluwer Academic, Plenum Publishers.

Kooter, J.M., M.A. Matzke, and P. Meyer. 1999. Listening to the silent genes:transgene silencing, gene regulation and pathogen control. Trends Plant Sci. 4:340-347.

Koshiishi, C., A. Kato, S. Yama, A. Crozier, and H. Ashihara. 2001. A new caffeine biosynthetic pathway in tea leaves: utilisation of adenosine released from the S-adenosyl-L-methionine cycle. FEBS Lett. 499:50-54.

Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15:8125-8148.

Kozak, M. 1999. Initiation of translation in prokaryotes and eukaryotes. Gene 234:187–208.

Kurata, H., S. Matsumara, and S. Furusaki. 1997. Light irradiation causes physiological and metabolic changes for purine alkaloid production by a *Coffea arabica* cell suspension culture. Plant Sci. 123:197 -203.

Kusaba, M. 2004. RNA interference in crop plants. Curr. Opin. Biotech. 15:139-143.

Ky, C.-L., J. Louarn, S. Dussert, B. Guyot, S. Hamon, and M. Noirot. 2001. Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. Food Chem. 75:223-230.

Lackie, J.M., and J.A.T. Dow. 1999. The Dictionary of Cell and Molecular Biology. Third ed. Academic Press, London.

Lagerstrom, M., J. Parik, H. Malmgren, J. Stewart, U. Pettersson, and U. Landegren. 1991. Capture PCR: efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. PCR Methods Appl. 1:111-119.

Lashermes, P., J. Cros, M.C. Combes, P. Trouslot, F. Anthony, S. Hamon, and A. Charrier. 1996. Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. Theor. Appl. Genet. 93:626-632.

Lashermes, P., M.C. Combes, P. Trouslot, and A. Charrier. 1997. Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. Theor. Appl. Genet. 94:947–955.

Lashermes, P., M.C. Combes, J. Robert, P. Trouslot, A. D'Hont, F. Anthony, and A. Charrier. 1999. Molecular characterisation and origin of the *Coffea arabica* L. genome. Mol. Gen. Genet. 261:259–266.

Lawton-Rauh, A. 2003. Evolutionary dynamics of duplicated genes in plants. Mol. Phylogenet. Evol. 29:396–409.

Lemon, B., and R. Tjian. 2000. Orchestrated response: a symphony of transcription factors for gene control. Genes Dev. 14:2551–2569.

Leroy, J.F. 1980. 'Evolution et taxogenèse chez les caféiers. Hypothese sur leur origine'. Comptes Rendusde l'Academie des Sciences 291:593-596.

Lescot, M., P. Dehais, G. Thijs, K. Marchal, Y. Moreau, Y. Van de Peer, P. Rouze, and S. Rombauts. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. Nucleic Acids Res. 30:325–327.

Lessard, P.A., R.D. Allen, F. Bernier, J.D. Crispino, T. Fujiwara, and R.N. Beachy. 1991. Multiple nuclear factors interact with upstream sequences of differentially regulated β-conglycinin genes. Plant Mol. Biol. 16:397–413.

Lewin, B. 2004. Genes. VIII ed. Pearson Prentice Hall, Pearson Education, Inc. NJ.

Li, G., and T.C. Hall. 1999. Footprinting in vivo reveals changing profiles of multiple factor interactions with the β-phaseolin promoter during embryogenesis. Plant J. 18:633-641.

Lin, C., L.A. Mueller, J. Mc Carthy, D. Crouzillat, V. Petiard, and S.D. Tanksley. 2005. Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. Theor. Appl. Genet. 112: 114-130.

Liu, Y.G., and R.F. Whittier. 1995. Thermal asymmetric interlaced PCR: Automatable amplification and sequencing of insert and fragments from P1 and YAC clones for chromosome walking. Genomics 25:674-681.

Liu, X., and W.V. Baird. 2001. Rapid Amplification of Genomic DNA Ends by NlaIII partial digestion and polynucleotide tailing. Plant Mol. Biol. Rep. 19:261-267.

Liu, L., M.J. White, and T.H. MacRae. 1999. Transcription factors and their genes in higher plants. Functional domains, evolution and regulation. Eur. J. Biochem. 262:247-257.

Liu, Q., S.P. Singh, and A.G. Green. 2002. High-stearic and high-oleic cottonseed oils produced by hairpin-RNA mediated post-transcriptional gene silencing. Plant Physiol. 129:1732-1743.

Lorkovic, Z.J., D.A. Wieczorek, M.H.L. Lambermon, and W. Filipowicz. 2000. Pre-mRNA splicing in higher plants. Trends Plant Sci. 5:160-167.

Lotan, T., M. Ohto, K.M. Yee, M.A. West, R. Lo, R.W. Kwong, K. Yamagishi, R.L. Fischer, R.B. Goldberg, and J.J. Harada. 1998. Arabidopsis LEAFY COTYLEDON is sufficient to induce embryo development in vegetative cells. Cell 93:1195-1205.

Lukyanov, S.A., N.G. Gurskaya, K.A. Lukyanov, V.S. Tarabykin, and E.D. Sverdlov. 1994. Highly efficient subtractive hybridization of cDNA. Bioorg.Chem. 20:701-704.

Lynch, M., and J.S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151-1155.

Mathavan, S., Y. Premalatha, and M.S.M. Christopher. 1985. Effects of caffeine and theophylline on the fecundity of four lepidopteran species. Exp. Biol. 44:133–138.

Mazzafera, P. 1991. Trigonelline in coffee. Phytochemistry 30:2309-2310.

Mazzafera, P. 1999. Chemical composition of defective coffee beans. Food Chem. 64:547-554.

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

Mazzafera, P., and A.C. Magalhaes. 1991. Caffeine content in leaves and seeds of species of *Coffea* and *Paracoffea*. Revta brasil. Bot. 14:157-160.

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

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

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

Mazzafera, P., M.B. Silvarolla, M.M.A. de Lima, and H.P.M. Filho. 1997. Caffeine content of diploid coffee species. J. Braz. Assoc. Adv. Sci. 49:216-218.

Meins, F., Jr. 2000. RNA degradation and models for post-transcriptional gene silencing. Plant Mol. Biol. 43:261-273.

Memelink, J., R. Verpoorte, and J.W. Kijne. 2001. ORCAnization of jasmonate-responsive gene expression in alkaloid metabolism. Trends Plant Sci. 6:212–219.

Mette, F., J. van der Winden, M.A. Matzke, and A.J.M. Matzke. 1999. Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters in trans. EMBO J. 18:241–248.

Mette, M.F., W. Aufsatz, J. van der Winden, M.A. Matzke, and A.J.M. Matzke. 2000. Transcriptional gene silencing and promoter methylation triggered by double standard RNA. EMBO J. 19:5194-5201.

Mignone, F., C. Gissi, S. Liuni, and G. Pesole. 2002. Untranslated regions of mRNAs. Genome Biol. 3:reviews0004.1–0004.10.

Mitchell-Olds, T., and M.J. Clauss. 2002. Plant evolutionary genomics. Curr. Opin. Plant Biol. 5:74–79.

Mizuno, K., H. Tanaka, M. Kato, H. Ashihara, and T. Fujimura. 2001. cDNA cloning of caffeine (theobromine) synthase from coffee (*Coffea arabica* L.). In: 19th International Scientific Colloquim on Coffee,ASIC, Paris: 815-818.

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

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

Moffatt, B.A., and E.A. Weretilnyk. 2001. Sustaining *S*-adenosyl-L-methionine-dependent methyltransferase activity in plant cells. Physiol. Plantarum 113:435-442.

Mol, J.N.M., A.R. van der Krol, A.J. van Tunen, R. van Blokland, P. de Lange, and A.R. Stuitje. 1990. Regulation of plant gene expression by antisense RNA. FEBS Lett. 268:427-430.

Montagnon, C., B. Guyat, C. Cilas, and T. Leroy. 1998. Genetic parameters of several biochemical compounds from green coffee, *Coffea canephora*. Plant Breeding 117:576-578.

Morgenstern, B. 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics 15:211 - 218.

Mösli Waldhauser, S.S., and T.W. Baumann. 1996. Compartmentation of caffeine and related purine alkaloids depends exclusively on the physical chemistry of their vacuolar complex formation with chlorogenic acids. Phytochemistry 42:985-996.

Mösli Waldhauser, S.S., J.A. Kretschmar, and T.W. Baumann. 1997a. N-methyltransferase activities in caffeine biosynthesis: biochemical characterization and time course during leaf development of Coffea arabica. Phytochemistry 44:853-859.

Mösli Waldhauser, S.S., F.M. Gillies, A. Crozier, and T.W. Baumann. 1997b. Separation of the N-methyltransferase, the key enzyme in caffeine biosynthesis. Phytochemistry 45:1407-1414.

Mueller, P.R., and B. Wold. 1991. Ligation-mediated PCR: applications to genomic footprinting. Methods 2:20-31.

Murashige, T., and F. Skoog. 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plantarum 15:73-497

Nakamura, M., T. Tsunoda, and J. Obokata. 2002. Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator. Plant J. 29:1-10.

Napoli, C., C. Lemieux, and R. Jorgensen. 1990. Introduction of a chimeric chalcone synthase gene into Petunia results in reversible co-suppression of homologous genes in *trans*. Plant Cell 2:279–289.

Nathanson, J.A. 1984. Caffeine and related methylxanthines: possible naturally occurring pesticides. Science 226:184–187.

Navellier, P. 1961. Composition moyenne du cafe´´ vert. Les cafe´iers et les cafe´s dans le monde, p. 163-220, *In* R. Coste, ed. Larose, Vol. 2.

Negishi, O., T. Ozawa, and H. Imagawa. 1985a. Methylation of xanthosine by tea leaf extracts and caffeine biosynthesis. Agric. Biol. Chem. 49:887-890.

Negishi, O., T. Ozawa, and H. Imagawa. 1985b. The role of xanthosine in the biosynthesis of caffeine in coffee plants. Agric. Biol. Chem. 49:2221-2222.

Negishi, O., T. Ozawa, and H. Imagawa. 1988. N-methylnucleosidase from tea leaves. Agric. Biol. Chem. 52:169-175.

Nicolas, F.E., S. Torres-Martinez, and R.M. Ruiz-Vazquez. 2003. Two classes of small antisense RNAs in fungal RNA silencing triggered by non-integrative transgenes. EMBO J. 22:3983–3991.

Nunnari, G., E. Argyris, J. Fang, K.E. Mehlman, R.J. Pomerantz, and R. Daniel. 2005. Inhibition of HIV-1 replication by caffeine and caffeine-related methylxanthines. Virology 335:177-184.

Nykänen, A., B. Haley, and P.D. Zamore. 2001. ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell 107:309–321.

Ochman, H., A.S. Gerber, and D.L. Hartl. 1988. Genetic application of an inverse polymerase chain reaction. Genetics 120:621-623.

Ogawa, M., Y. Herai, N. Koizumi, T. Kusano, and H. Sano. 2001. 7-Methylxanthine Methyltransferase of coffee plants. Gene isolation and enzymatic properties. The J. Biol. Chem. 276:8213-8218.

Ogita, S., H. Uefuji, Y. Yamaguchi, N. Koizumi, and H. Sano. 2003. Producing decaffeinated coffee plants. Nature 423:823.

Ogita, S., H. Uefuji, M. Morimoto, and H. Sano. 2004. Application of RNAi to confirm theobromine as the major intermediate for caffeine biosynthesis in coffee plants with potential for construction of decaffeinated varieties. Plant Mol. Biol. 54:931-941.

Ozawa, Y., S. Hanaoka, R. Saito, T. Washio, S. Nakano, A. Shinagawa, M. Itoh, K. Shibata, P. Carninci, H. Konno, J. Kawai, Y. Hayashizaki, and M. Tomita. 2002. Comprehensive sequence analysis of translation termination sites in various eukaryotes. Gene 300:79-87.

Padegimas, L.S., and N.A. Reichert. 1998. Adapter ligation-based polymerase chain reaction-mediated walking. Anal. Biochem. 260:149-153.

Paillard, M., P. Lashermes, and V. Petiard. 1996. Construction of molecular linkage map in Coffee. Theor. App. Genet.93: 41-47.

Pesole, G., F. Mignone, C. Gissi, G. Grillo, F. Licciulli, and S. Liuni. 2001. Structural and functional features of eukaryotic mRNA untranslated regions. Gene 276:73-81.

Poncet, V., P. Hamon, J. Minier, C. Carasco, S. Hamon, and M. Noirot. 2004. SSR cross-amplification and variation within coffee trees (*Coffea* spp.). Genome 47:1071–1081.

Pujade-Renauda, V., C. Saniera, L. Cambillaua, A. Pappusamyb, H. Jonese, N. Ruengsric, D. Tharreaua, H. Chrestind, P. Montoroa, and J. Narangajavana. 2005. Molecular characterization of new members of the *Hevea brasiliensis* hevein multigene family and analysis of their promoter region in rice. Biochim. Biophys. Acta 1727:151–161.

Pwee, K.-H., and J.C. Gray. 1993. The pea plastocyanin promoter directs cell-specific but not full light-regulated expression in transgenic tobacco plants. Plant J 3:437–449.

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

Riechers, D.E., and M.P. Timko. 1999. Structure and expression of the gene family encoding putrescine N-methyltransferase in *Nicotiana tabacum*: new clues to the evolutionary origin of cultivated tobacco. Plant Mol. Biol. 41:387–401.

Riechmann, J.L., J. Heard, G. Martin, L. Reuber, C.-Z. Jiang, J. Keddie, L. Adam, O. Pineda, O.J. Ratcliffe, R.R. Samaha, R. Creelman, M. Pilgrim, P. Broun, J.Z. Zhang, D. Ghandehari, B.K. Sherman, and G.-L. Yu. 2000. *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. Science 290:2105-2109.

Roberts, M.F., and G.R. Waller. 1979. N-methyltransferases and 7-methyl-N9-nucleoside hydrolase activity in *Coffea arabica* and the biosynthesis of caffeine. Phytochemistry 18:451-455.

Roeder, R.G. 1996. The role of general initiaion factors in transcription by RNA polymerase II. Trends Biochem. Sci. 21:327-335.

Roeder, R.G. 2005. Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett. 579:909–915.

Rogozin, I.B., A.V. Sverdlov, V.N. Babenko, and E.V. Koonin. 2005. Analysis of evolution of exon-intron structure of eukaryotic genes. Briefings Bioinformatics 6:118-134.

Romano, N., and G. Macino. 1992. Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. Mol. Microbiol. 22:3343–3353.

Rosenthal, A., and D.S. Jones. 1990. Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction. Nucleic Acids Res. 18:3095-3096.

Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Sambrook, J., and D.W. Russell. 2001. Molecular Cloning: a laboratory manual. Third ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sandhu, J.S., C.I. Webster, and J.C. Gray. 1988. A/T-rich sequences act as quantitative enhancers of gene expression in transgenic tobacco and potato plants. Plant Mol. Biol. 37:885–896.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A 74:5463-5467.

Sappl, P.G., L. Joshua, A. Heazlewood, and H. Millar. 2004. Untangling multi-gene families in plants by integrating proteomics into functional genomics. Phytochemistry 65:1517-1530.

Schaart, J.G., L. Mehli, and H.J. Schouten. 2005. Quantification of allele-specific expression of a gene encoding strawberry polygalacturonase-inhibiting protein (PGIP) using Pyrosequencing. Plant J. 41: 493-500.

Schulthess, B.H., and T.W. Baumann. 1995. Stimulation of caffeine biosynthesis in suspension-cultured coffee cells and the in situ existence of 7-methylxanthosine. Phytochemistry 38:1381-1386.

Schulthess, B.H., P. Morath, and T.W. Baumann. 1996. Caffeine biosynthesis starts with the metabolically chanelled formation of 7-methyl-XMP: a new hypothesis. Phytochemistry 41:169-175.

Segal, G., R. Song, and J. Messing. 2003. A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. Genetics 165:387–397.

Seo, H.S., J.T. Song, J.J. Cheong, Y.H. Lee, Y.W. Lee, I. Hwang, J.S. Lee, and Y.D. Choi. 2001. Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. Proc. Natl. Acad. Sci. U.S.A. 98:4788-4793.

Shabalina, S.A., A.Y. Ogurtsov, I.B. Rogozin, E.V. Koonin, and D.J. Lipman. 2004. Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals. Nucleic Acids Res. 32:1774-1782.

Siebert, P.D., A. Chenchik, D.E. Kellogg, K.A. Lukyanov, and S.A. Lukyanov. 1995. An improved method for walking in uncloned genomic DNA. Nucleic Acids Res. 23:1087-1088.

Sijen, T., J. Wellink, J.-B. Hiriart, and A. van Kammen. 1996. RNA-mediated virus resistance: role of repeated transgenes and delineation of targeted regions. Plant Cell 8:2277–2294.

Sijen, T., J. Fleenor, F. Simmer, K.L. Thijssen, S. Parrish, L. Timmons, R.H.A. Plasterk, and A. Fire. 2001. On the role of RNA amplification in dsRNA-triggered gene silencing. Cell 107:465–476.

Silvarolla, M.B., P. Mazzafera, and M.M.A. de Lima. 2000. Caffeine content of Ethiopian *Coffea arabica* beans. Genet. Mol. Biol. 23:213-215.

Silvarolla, M.B., P. Mazzafera, and L.C. Fazuoli. 2004. A naturally decaffeinated arabica coffee. Nature 429:826.

Singh, K.B. 1998. Transcriptional regulation in plants: The importance of combinatorial control. Plant Physiol. 118:1111-1120.

Smith, N.A., S.P. Singh, M.-B. Wang, P.A. Stoutjesdijk, A.G. Green, and P.M. Waterhouse. 2000. Total silencing by intron-spliced hairpin RNAs. Nature 407:319-320.

Song, J.T., H.S. Seo, S.I. Song, J.S. Lee, and Y.D. Choi. 2000. NTR1 encodes a floral nectary -specific gene in *Brassica campestris* L. ssp. *pekinensis*. Plant Mol. Biol. 42:647–655.

Suzuki, T., and E. Takahashi. 1975. Biosynthesis of caffeine by tea-leaf extracts enzymic formation of theobromine from 7-methylxanthine and of caffeine from theobromine. Biochem. J 146:87-96.

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

Szopa, J., M. Lukaszewicz, A.K. Anna Aksamit , and D. Kwiatkowska. 2003. Structural organisation, expression, and promoter analysis of a 16R isoform of 14-3-3 protein gene from potato. Plant Physiol. Biochem. 41:417–423.

Szweykowska-Kuliñska, Z., A. Jarmolowski, and M. Figlerowicz. 2003. RNA interference and its role in the regulation of eukaryotic gene expression. Acta Biochim. Pol. 50:217–229.

Tada, Y., H. Akagi, T. Fujimura, and T. Matsuda. 2003. Effect of an antisense sequence on rice allergen genes comprising a multigene family. Breeding Sci. 53:61-67.

Tang, G., B.J. Reinhart, D.P. Bartel, P.D. Zamore. 2003. A biochemical framework for RNA silencing in plants. Genes Dev. 17:49–63.

Tavares, R., S. Aubourg, A. Lecharny, and M. Kreis. 2000. Organization and structural evolution of four multigene families in *Arabidopsis thaliana*: AtLCAD, AtLGT, AtMYST and AtHD-GL2. Plant Mol. Biol. 42:703-717.

Taylor, C.B. 1997. Promoter fusion analysis: An insufficient measure of gene expression. Plant Cell 9:273-275.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,

position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.

Trick, H.N., and J.J. Finer. 1997. Sonication-Assisted *Agrobacterium* mediated transformation. Transgenic Res. 6:329-337.

Tsuchisaka, A., and A. Theolog. 2004. Unique and overlapping expression patterns among the Arabidopsis 1-Amino-Cyclopropane-1-Carboxylate synthase gene family members. Plant Physiol. 136:2982–3000.

Tyagi, A.K. 2001. Plant genes and their expression. Curr. Sci. 80:161-169.

Uefuji, H., S. Ogita, Y. Yamaguchi, N. Koizumi, and H. Sano. 2003. Molecular cloning and functional characterization of three distinct N-Methyltransferases involved in the caffeine biosynthetic pathway in coffee plants. Plant Physiol. 132:372-380.

Uefuji, H., Y. Tatsumi, M. Morimoto, P. Kaothien-Nakayama, S. Ogita, and H. Sano. 2005. Caffeine production in tobacco plants by simultaneous expression of three coffee *N*-methyltransferases and its poential as a pest repellant. Plant Mol. Biol. 59:221-227.

Ulker, B., and I.E. Somssich. 2004. WRKY transcription factors: from DNA binding towards biological function. Curr. Opin. Plant Biol. 7:491-498.

Vaistij, F.E., L. Jones, and D.C. Baulcombe. 2002. Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. Plant Cell 14:857–867.

van Boxtel, J., and M. Berthouly. 1996. High frequency somatic embryogenesis from coffee leaves: Factors influencing embryogenesis, and subsequent proliferation and regeneration in liquid medium. Plant Cell Tiss. Org. Cult. 44:7–17

Vision, T.J., D.G. Brown, and S.D. Tanksley. 2000. The origins of genomic duplications in Arabidopsis. Science 290:2114-2117.

Vitória, A.P., and P. Mazzafera. 1999. Xanthine degradation and related enzymes activities in leaves and fruits of two *Coffea* species differing in caffeine catabolism. J. Agric. Food Chem. 47:1851-1855.

Waller, G.R., C.D. MacVean, and T. Suzuki. 1983. High production of caffeine and related enzyme activities in callus cultures of *Coffea arabica* L. Plant Cell Rep. 2:109-112.

Wang, G., and G.J. Wagner. 2003. Elucidation of the functions of genes central to diterpene metabolism in tobacco trichomes using posttranscriptional gene silencing. Planta 216:686-691.

Wang, M.B., and P.M. Waterhouse. 2001. Application of gene silencing in plants. Curr. Opin. Plant Biol. 5:146–150.

Waterhouse, P.M., M. Gramham, and M.B. Wang. 1998. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc. Natl. Acad. Sci. U.S.A. 95:13959–13964.

Wesley, S.V., C.A. Helliwell, N.A. Smith, M. Wang, D.T. Rouse, Q. Liu, P.S. Gooding, S.P. Singh, D. Abbott, and P.A. Stoutjesdijk. 2001. Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J 27:581-590.

Wingender, E., X. Chen, E. Fricke, R. Geffers, R. Hehl, I. Liebich, M. Krull, V. Matys, H. Michael, R. Ohnhauser, M. Prub, F. Schacherer, S. Thiele and S. Urbach. 2001. The TRANSFAC system on gene expression regulation. Nucleic Acids Res. 29:281–283.

Wink, M. 1997. Compartmentation of secondary metabolites and xenobiotics in plant vacuoles. Adv. Bot. Res. 25:141-169.

Wray, G.A., M.W. Hahn, E. Abouheif, J.P. Balhoff, M. Pizer, M.V. Rockman, and L.A. Romano. 2003. The evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol. 20:1377-1419.

Yanagisawa, S. 1997. Dof DNA-binding domains of plant transcription factors contribute to multiple protein-protein interactions. Eur. J. Biochem. 250:403–410.

Yuan, Q., J. Hill, J. Hsiao, K. Moffat, S. Ouyang, Z. Cheng, J. Jiang, and C.R. Buell. 2002. Genome sequencing of a 239-kb region of rice chromosome 10L reveals a high frequency of gene duplication and a large chloroplast DNA insertion. Mol. Gen. Genom. 267:713–720.

Zheng, X.Q., and H. Ashihara. 2004. Distribution, biosynthesis and function of purine and pyridine alkaloids in *Coffea arabica* seedlings. Plant Sci. 166:807–813.

Zhu, X., and G. Galili. 2004. Lysine metabolism is concurrently regulated by synthesis and catabolism in both reproductive and vegetative tissues. Plant Physiol. 135:129–136.

Zhu, W., S.D. Schlueter, and V. Brendel. 2003. Refined annotation of the Arabidopsis genome by complete expressed sequence tag mapping. Plant Physiol. 132:469–484.

Zubieta, C., J.R. Ross, P. Koscheski, Y. Yang, E. Pichersky, and J.P. Noel. 2003. Structural basis for substrate recognition in the Salicylic Acid Carboxyl Methyltransferase family. Plant Cell. 1704–1716 15:1704–1716.

Publications

Research papers

1. **K.V. Satyanarayana,** K. Vinod, A. Chandrashekar and G.A. Ravishankar (2005). Isolation of promoter for *N*-methyltransferase gene associated with caffeine biosynthesis in *Coffea canephora*. **Journal of Biotechnology** 119:20-25

2. **K.V. Satyanarayana,** A. Chandrashekar and G.A. Ravishankar (2006). Evaluation of PCR based methods for isolating flanking regions of genes. **Molecular Biotechnology** 32 (*In press*)

3. K. Vinod, **K.V. Satyanarayana**, S. Sarala Itty, E. P. Indu, P. Giridhar, A. Chandrashekar and G.A. Ravishankar (2006). Stable transformation and direct regeneration in *Coffea canephora* P ex. Fr. by *Agrobacterium rhizogenes* mediated transformation without hairy root phenotype. **Plant Cell Reports** (*In Press*). (Available online DOI 10.1007/s00299-005-0045-x).

4. **K.V. Satyanarayana,** K. Vinod, S. Sarala Itty, A. Chandrashekar and G.A. Ravishankar. Posttranscriptional gene silencing using homologous coding sequence for coffee *N*-methyltransferase genes comprising a multigene family. (Communicated).

5. S. Sarala Itty, Vinod Kumar, **K.V. Satyanarayana**, E.P. Indu, P. Giridhar, A. Chandrashekar and G.A. Ravishankar. Growth regulator and light influence purine alkaloid formation in somatic embryos of *Coffea canephora* Pierre ex. Froehner. (Communicated).

6. Vinod Kumar, **K.V. Satyanarayana**, A., Ramakrishna, A., Chandrashekar and G.A. Ravishankar. Evidence for localization of *N*-methyltransferases of caffeine biosynthetic pathway in vacuolar surface of *Coffea canephora* endosperm elucidated through localization of GUS reporter gene driven by NMT promoter. (Communicated).

7. **K.V. Satyanarayana,** A. Chandrashekar and G.A. Ravishankar. Intra and Inter species conservation of non-coding regions of *N*-methyltransferase genes involved in caffeine biosynthetic pathway in *Coffea*. (To be communicated).

Patents

Vinod Kumar, E. P. Indu, **K.V. Satyanarayana**, P. Giridhar, A. Chandrashekar and G. A. Ravishankar. "An improved culture media for regeneration of transgenic secondary embryos of *Coffea canephora* P.ex.Fr". (Indian Patent 397/DEL/03).

Submissions to the NCBI GenBank database:
Accession nos.: AY918124, AY918125, AY918126, AY918127, AY907697, DQ010010 and DQ010011

Papers in Proceedings

1. **Satyanarayana K.V.**, Vinod Kumar, Chandrashekar A., Ravishankar G.A. Cloning and characterization of promoter for *N*-methyltransferase gene from coffee. In: Proceedings of 20th International Conference on Coffee Science (ASIC 2004, Bangalore, India).

2. Vinod Kumar, **Satyanarayana K.V.**, Sarala Itty S., Giridhar P., Chandrashekar A., Ravishankar G.A. Posttranscriptional gene silencing for down regulating caffeine biosynthesis in *Coffea canephora* Pex Fr. In: Proceedings of 20th International Conference on Coffee Science (ASIC 2004, Bangalore, India).

Posters/Abstracts in Academic Conferences/ Proceedings:

1. **Satyanarayana K.V.**, Vinod Kumar, Sarala Itty S., Chandrashekar A., Ravishankar G.A. "Cloning and Characterization of *N*- methyltransferase gene and its promoter from coffee". Presented at '73rd Annual meeting of Society of Biological Chemists (India)' held during December 2004 at G.B.P.U.A.T, Pantnagar, India.

2. **Satyanarayana K.V.**, Chandrashekar A., Ravishankar G.A. "Cloning and sequencing of N- methyltransferase gene associated with caffeine biosynthesis in coffee" Poster presented at the first International Symposium on 'Molecular Approaches for Improved Crop Productivity and Quality' held during May 22-24,2002 at T.N.A.U., Coimbatore, India.