Study of an OMT-like genomic DNA sequence in *Chrysosplenium americanum*

Dina Al-Khairy

A Thesis

in

The Department

of

Biology

Presented in Partial fulfillment of the requirements for the Degree

of Master of Science at Concordia University

Montreal, Quebec, Canada

April 2003

ABSTRACT

Study of an OMT-like genomic
DNA sequence in
Chrysosplenium americanum

Dina Al-Khairy

Chrysosplenium americanum has a number of flavonoid *O*-methyltransferases (OMT's) that methylate substrates in a substrate and position specific fashion. These OMT's have been characterized biochemically, but only one such enzyme has been cloned to date, namely a 3'5' flavonoid OMT. A genomic DNA clone containing OMT like sequences which was previously isolated was characterized in this work in order to investigate the possibility that these sequences might be derived from genes that had novel substrate and position specificity. These clones also gave insight into gene duplication and rearrangement in the gene family of flavonoid OMTs. Analysis indicated that the OMT-like sequences in the genomic clone are likely pseudogenes. They are likely the product of gene duplication, multiple segmental rearrangements and frame shift mutations. A PCR based strategy was used to isolate partial length cDNA clones that were nearly identical to the genomic sequences. This indicated that highly related genes are expressed, however there is no indication that the cDNA encode active enzymes. The screening of the *C. americanum* cDNA library using the partial cDNA clone isolated by PCR as a radioactive probe led to the recovery of 16 cDNA clones that were nearly identical to previously characterized 3'5' flavonoid OMT. These cDNAs contained polymorphims for dinucleotide and trinucleotide repeats in their 5' UTR region.

ACKNOWLEDGEMENTS

For my parents, who gave me extreme love and full support, not only through this study, but also through-out my life, I owe you for my success. Your support and your hard work inspired me through the years.

I would like to thank my husband Asaad for the encouragement, love and support he gave me since day one.

I would also like to thank my brothers Isaam and Jawad. Both of them have been always there for me. Thank you.

I would like to truly thank my supervisor Dr. Gulick for all the support and help he gave me, not only through guidance but also thorough working with me on my bench top, side by side. I will always remember the help and treasure the friendship.

I would also like to thank Dr. Ibrahim and Dr. Varin for helpful discussions and continuous guidance.

I would like to pass my gratitude to Dr. Musac, Dr. Pitsikas, Demetra Elias, Dominique Anzelloti, and Mohammed Atari for sincere help.

Special thanks go to Dr. Robin Ingalls who gave me the chance to work on my thesis while working in her lab.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVIATIONS

PCR: Polymerase chain reaction

UV: Ultraviolet light

DCTP: 2'-Deoxycytidine 5'-triphosphate

Tris: Tris (hydroxymethyl) aminomethane

OMT: *O*-Methyltransferase

CHS : Chalcone synthase

FLS : Flavonoid synthase

PAL : L-phenylalanine ammonia lyase

SAM : *S*-adenosyl-L-methionine

SAH : *S*-adenosyl-Homocysteine

AdoMet: *S*-adenosyl-L-methionine

CHAPTER 1. LITERATURE REVIEW

1.1. Secondary metabolites in plants

Plants elaborate an astonishing array of secondary metabolites. Such compounds are usually synthesized from intermediates of primary metabolism via complex biosynthetic pathways. There has been considerable interest in understanding how these compounds are synthesized and how these pathways are regulated, since it has been generally viewed that such compounds play an important role in plant's interaction with the environment (Dixon *et al.*, 1996). Although our knowledge about the extended roles of these metabolites is limited, there has been a fast growing appreciation of the role of secondary metabolites in improving the plant's vigor and survival in the last two decades.

Secondary metabolites are considered as important compounds with a restricted occurrence in certain taxonomic groups. These compounds are believed to ensure a better survival of the organism in its ecosystem by playing a role in its interaction with the environment (Verpoorte *et al.*, 2000). In 1998, the number of known plant secondary metabolites were listed to be around 85,000 compounds (NAPRALERT Database, Dictionary of Natural products) in which the largest group was terpenoid compounds for which 27,463 forms are known. Flavonoids include at least 8128 different compounds (Chapman and Hall, 1998).

Secondary metabolites are subject to further modification in each plant species. Such modifications include addition of new functional groups, isomerization, as well as coupling of different secondary metabolites (Barron *et al.*, 1996). Some of these modifications are believed to play a role in disease resistance (Verpoorte *et al.*, 2000). A

well-documented example is that 7,3'-dimethoxy-4'hydroxyflavan showed the maximum activity of inhibition of growth of *Candida albicans* as compared to the closely related flavans, 5,4'-dihydroxy-7,3'-dimethoxy-flavan, and 3',4'-dihydroxy-5,7-dimethoxyl flavan (Garo *et al.*, 1996). 4',5,7-Trihydroxyflavanone (Naringenin) and quercetin-3-β-D-rutinoside (Rutin), showed differential activity as UV protectants. The former was more effective in preventing the accumulation of UVB-induced DNA damage (Kootstra, 1994).

1.2. The general phenylpropanoid pathway

The general phenylpropanoid pathway (Figure 1) leads to the synthesis of phenylpropanoids as well as flavonoids in plants (Verpoorte *et al.*, 2000). Various plant phenolic compounds originate from the deamination of L-phenylalanine by L-phenylalanine ammonia lyase (PAL), giving rise to trans-cinnamic acid, which is further hydroxylated into para-coumaric acid by the action of cinnamate-4-hydroxylase. The latter is then esterified by 4-coumarate CoA ligase producing 4-coumaroyl-CoA.

The most important steps leading to flavonoid synthesis are the coupling of two precursors, one molecule of 4-coumaroyl-CoA with three molecules of malonyl CoA. The reaction which is the first committed step towards flavonoid synthesis, is catalyzed by chalcone synthase CHS. Three molecules of malonyl-CoA contribute to ring A formation, whereas the 4-hydroxycinnamte forms ring B and the 3-C side chain. The resulting chalcone (4,2',4',6'-tetrahydroxychalcone) is a C15 intermediate which undergoes further steriospecific rearrangement and C-ring closure by chalcone isomerase

(CHI) resulting in naringenin, the first flavonoid in this pathway (Heller and Forkmann, 1988).

1.3. Methyltransferases

Methyltransferases are members of an important group of enzymes that play an essential role in the maintenance of biological activities of all organisms. These enzymes catalyze the transfer of a methyl group of S-adenosyl-L-methionine AdoMet to an acceptor molecule, with the concomitant formation of S-adenosylhomosystein (AdoHcy) and the corresponding methylated derivatives as products. The enzymatic methylation of plants natural products by S-adenosyl-L-methionine (AdoMet) dependent–methyltransferases (MTs) is believed to inactivate the reactive hydroxyl and/or carboxyl groups in the acceptor molecule, and thus alters their solubility and their intracellular compartmentation (Ibrahim and Musac, 2000).

MTs are divided into four distinct families. These are O, C, N, and S MTs, in which the different families are divided according to the target atom in the methyl acceptor molecule. *O*-methyltransferases are considered the most important family of OMT's. Several of these genes have been cloned and characterized in various plants. Some phenylpropanoid OMT's are involved in lignin biosynthesis such as caffeic/5-hydroxyferulic acid (Gauthier *et al.*, 1998). Other OMT's are involved in secondary metabolism involved in pathogen defense. Stilbenes are believed to function as phytoalexins. A stilbene OMT recently isolated from ozone-induced Scots pine needles was found to catalyze the methylation of pinosylvin, as well as quercetin, catechol, and HCAs. Chalcone cDNAs of licorice (Haga *et al.*, 1997) and of alfalfa (Maxwell *et al.*,

1993) were found to encode OMTs for the 2'-methylation of the licodione and 5-deoxy chalcones isoliquiritigenin, respectively. A flavonoid 7-OMT with highest affinity to apigenin was found to be expressed in barley in response to pathogen infection (Christensen *et al.*, 1998). One of three constitutively expressed flavonol cDNA clones isolated is from *C. americanum* which was found to encode 3'/ 5'- methylation of 3,7,4'-trimethylquercetin (Gauthier *et al.*, 1996). A recently isolated cDNA clone from *Arabidopsis thaliana* has been identified as a quercetin 3'-OMT (Muzac *et al.*, 1999). The third cDNA clone isolated from *C. americanum* encodes the 3'-methylation of luteolin and quercetin, in addition to the 3 / 5-methylation of caffeic and 5-hydroxyferulic acids respectively (Gauthier *et al.*, 1998).

Another stress-induced isoflavone OMT cDNA clone is the isoflavone 7-*O*-methyltransferase from alfalfa, which mediates the 7-*O*-methylation of daidzein (He *et al.*, 1998). This OMT shares 50% sequence similarity with fungal-induced pterocarpan 3-OMT of pea, another isolated OMT that catalyzes the 3-*O*-methylation of 6a-hydroxymaackiain, the terminal step in the biosynthesis of pisatin, a known phytoalexin in pea (Wu *et al.*, 1997).

1.4. Methylation in *Chrysosplenium americanum*

Chrysosplenium americanum (saxifragaceae), is a small semiaquatic freshwater weed that is widely distributed in the Northern hemisphere. It synthesizes and accumulates wide variety of tetra- and penta-*O*-methylated flavonol glucosides identified as 2' and 5'-*O*-glucosides (Bajaj *et al.*, 1983). However, this plant does not accumulate intermediate compounds with lower level of methylation, nor does it produce significant amounts of

lignin. However, it was possible to demonstrate the enzymatic synthesis of flavonoids with lower degree of methylation *in vitro* using cell free extracts from this species and quercetin as substrate (De Luca and Ibrahim, 1982)

O-methylation of flavonoids involves the transfer of a methyl group of S-adenosyl-L-methionine (AdoMet) to OH group on a flavonoid acceptor molecule, and is catalyzed by *O*-methyltransferase (OMT) (De Luca and Ibahim I, 1985).

The occurrence of polymethylated flavonoid compounds in plants such as *C. americanum* raised the question as to whether multiple *O*-methylation reactions with a single type of flavonoid substrate are catalyzed by one or several related, position - specific OMT's. Stepwise *O*-methylation of flavonoids such as quercetin was reported for the first time using cell-free extracts of Calamondin orange (*Citris mitis*) (Brunet and Ibrahim., 1980). A stepwise methylation in tobacco plants (De Luca *et al.*, 1982), as well as in apple cell culture (Macheix and Ibrahim, 1984) has been reported. In *C. americanum*, the biosynthesis of such polymethylated flavonoids has been shown to be catalyzed in a stepwise manner, by a family of five position-specific OMT's as well as two *O*-glucosyltransferases (Bajaj *et al.*, 1983; Ibrahim *et al.*, 1987).

Figure 2 shows the proposed pathway for the stepwise *O*-methylation of flavonoids with quercetin as the main substrate in *C. americanum* for producing polymethylated flavonoids (Ibrahim *et al.*, 1987). The first step in this pathway is the 3-methylation of quercetin, which is catalyzed by the 3-OMT (De Luca and Ibrahim, I, 1985), giving rise to 3-methylquercetin. The second enzyme in this pathway, the 7-OMT, which showed strict position specificity towards its substrate, 3-methylquercetin, gives rise to 3,7-dimethylquercetin. The following two enzymes in this pathway, the 4'- and 6-

OMT, accept partially methylated flavonol intermediates. For example, 4'-OMT converted both 3,7,dimethylquercetin and its 6-hydroxy analog equally well (De Luca and Ibrahim I, 1985). On the other hand, the 6-OMT was found to accept 3,7-dimethylquercetagetin and 3,7,3'-trimethylquercetagetin, producing their 6-methyl derivatives. The fact that the 6-OMT accepted trimethyl derivatives suggested that methylation in this plant occurs in the order 3,7,4',6. Interestingly, the 6-OMT was inhibited by its substrate, as compared with either the 3- or the 4'-OMT's which were less affected by their respective substrates (De Luca and Ibrahim II, 1985), suggesting that the latter enzymes are involved in earlier steps of methylation and are possibly involved in regulation of synthesis rate of the final products in the proposed pathway.

Two terminal methylation steps take place at the level of glycosylated substrates (Ibrahim *et al.*, 1987). A glucosyltransferase involved in this pathway catalyses the transfer of glucose from UDP-glucose to positions 2' or 5' of partially methylated flavonols prior to the final methylations steps (Bajaj *et al.*, 1983). The remaining methylation steps in this pathway are the 2'- and 5'-*O*-methylations. It was first proposed that 2'- and 5'-O-methylations were catalyzed by one enzyme, which was believed to possess the ability to methylate both positions (Khouri *et al.* 1986). Only recently, the gene coding for 3'/5' OMT in this plant was isolated, and its gene product has been characterized to have specificity for methylation of the 3' positions of 2'- glucosylated flavonols (Gauthier *et al.*, 1996).

1.5. Functional properties of flavonoids

1.5.1. Functional properties of flavonoids

Flavonoid compounds play an important role in plant growth and development. Well-documented examples about the roles of flavonoids as UV protectants have been reported (Booij-James *et al.*, 2000). In addition to their frequently ascribed function as UV sunscreens, flavonoids' role in disease resistance is an area of interest (Malhotra *et al.* 1996). Flavonoids act as auxin transport regulators (Jacobs *et al.*, 1988), as well as molecular signals synthesized by the host plant to activate symbiosis. Flavonoid compounds such as luteolin (3',4',5,7-tetrahydroxy flavone) as well as naringenin (4',5,7-trihydroxy flavanone) and daidzein (4',7-dihydroxyisoflavone) have been reported to be Nod inducers in plant symbiotic interaction (Long, 1989). Studies carried out on the pathogen *Fusarium solani* f.sp.pisi, revealed that flavonoids such as pisatin, naringenin and apigenin act as stimulatory signals that stimulate the germination of *Fusarium solani* f.sp.pisi macroconidia (Ruan *et al.*, 1995).

A more pronounced role of flavonoids in plant growth and development was seen in the involvement of flavonoids such as quercetin, kaempferol and myricetin in pollen germination and pollen tube growth elongation (Ylstra *et al.*, 1992). Pollen tubes of quercetin-treated tobacco (*Nicotiana tabacum L.*) pollen germinated faster and were longer than the untreated pollen. Flavonoid-treated pollen showed 59.9% germination after 72 h, compared to 25.5 % of the pollen germinated in the control medium. Other studies which monitored gene expression in potato for genes encoding enzymes involved in flavonoid biosynthesis during pollen germination showed that levels of flavonol

synthase (FLS) and chalcone synthase (CHS) transcripts have increased markedly in both anthers and pistils (Eldik *et al.* 1996).

There are numerous reports, which indicate a major role of flavonoid compounds during the acclimation of plant species to environmental stress, in response to pathogen challenge, and during plant development, (Long, 1989; Malhotra *et al.*, 1996;Ylstra *et al.*, 1992).

1.5.1.1 Flavonoids as UV protectants

Plants intercept sunlight in order to maintain their physiological and functional integrity, indeed, exposure to sunlight is essential for survival and growth. However, in the course of using sunlight for photosynthesis, plants are unavoidably exposed to the damaging effects of UV-B radiation. Exposure of plants to high levels of UV-B irradiation causes DNA mutagenesis as well as other cellular injuries and ultimately plant death (Booij-James *et al.*, 2000).

Plants have developed a range of strategies to protect themselves from deleterious effects of such UV-B radiation. Plants have the ability to filter sunlight by the production of UV absorbing compounds including flavonoids, which can be deposited on the leaf surface or contained in epidermal cell vacuoles (Cuadra *et al.*, 1996). It is believed that UV absorbing compounds can act as a shield that block UV before it reaches sensitive targets, rendering plants with high amounts of those compounds less susceptible to damage by UV B exposure (Booij-James *et al.*, 2000). In *Arabidopsis thaliana*, a mutant with a constitutively elevated accumulation of flavonoids in the leaf epidermal cells

showed a remarkable increase in tolerance to UVB damage at doses that were lethal to wild-type plants (Bieza *et al.*, 2001).

Several studies have provided information supporting the importance of flavonoid glucosides in plants. Acylated kaempferol glucosides were found to be the main flavonoid compound found in leaf hairs in *Quercus ilex* (Skaltsa *et al.* 1994). Interestingly, UVB was found to cause a considerable reduction in photosystem II photochemical efficiency only in dehaired leaves, emphasizing the role of flavonoids in protection against damage caused by UV. Kaempferol and quercetin were also found to cause protection against UVB light in both Clark and Harosoy cultivars of soybean (*Glycine max)* (Middleton and Teramura, 1993). Recent studies on *Brassica napus* have also emphasized the role of quercetin and kaempferol glucosides as UVB protectants (Olsson *et al.*, 1998). There was an increase of 150% of those two compounds in cultivars Paroll, and of 70% in cultivar Stallion when plants were irradiated with UVB as compared to non-irradiated controls.

Phenolic compounds which also absorb UV radiation are also thought to be key players in the defense mechanism against UVB damage in plants. In the Arabidopsis ecotype Landsberg erecta, sinapic acid esters such as *O*-sinapoyl-L-malate and 1-*O*-sinapoyl-β-D-glucose were found to be elevated in response to UVB radiation (Li *et al.*, 1993). An Arabidopsis mutant for ferulic acid hydroxylase, which lacks sinapate esters, was found to have enhanced UVB sensitivity, demonstrating again the importance of hydroxycinnamate esters as UVB protectants together with flavonoids (Landry *et al.* 1995). The greater sensitivity of ferulic acid hydroxylase mutant than the chalcone

synthase and the chalcone isomerase mutants in *A. thaliana* suggests that sinapate esters are the major UVB sunscreens in *A. thaliana*.

1.5.1.2. Role of flavonoids in disease resistance

Plants have evolved several defense mechanisms against disease, one of which is the synthesis of natural products that inhibit pathogen growth. Phytoalexins are classically defined as antimicrobial compounds that are produced after pathogen challenge. In addition, preformed phenylpropanoids have also been found to be important in decreasing disease susceptibility. PAL suppressed transgenic tobacco leaves have more severe lesions caused by the virulent fungus *Cercospora nicotianae* (Maher *et al.*, 1994) than wild type plants; providing direct evidence of the importance of the products of the phenylpropanoid pathway in disease limitation.

Changes in expression of genes coding for enzymes that are involved in the flavonoid pathway have been observed in response to pathogen attack. In response to infection by *Blumeria graminis*, barley leaves showed marked transcript accumulation for a flavonoids 7-*O*-methyltransferase but not for caffeic acid OMT suggesting possible role of flavonoids such as apigenin (4',5,7-trimethoxyflavone) in response to pathogen challenge (Christensen *et al.*, 1998).

Mariscus psilostachys, which belongs to the family Cyperaceae which includes several herbs commonly used by West African healers to treat infections, was used to isolate several flavans, including (2S)-4'-hydroxy-5,7,3'-trimethoxyflavan, 5,4'-dihydroxy-7,3'-dimethoxyflavan and 3',4'-dihydroxy-5,7-dimethoxyflavan (Garo *et al.*,

1996). These were all found to inhibit growth of *Candida albicans*, and *Cladosporium cucumerinum* and thus were showed to act as antifungal compounds (Garo *et al.*, 1996).

Flavonoids have also been shown to have potent activity against plant viruses (Malhotra *et al.*, 1996). Quercetin and its methylated derivatives, such as quercetin 7,4'-dimethyl ether, quercetin 3,7,4'-trimethyl ether were found to inhibit Tomato Ringspot Nepovirus (TomRSV) replication. Quercetin was found to inhibit TomRSV infectivity up to 70% when applied in a mixed inoculum with TomRSV. Flavonoids may interfere with an early event in the virus replication cycle, such as binding to the viral coat protein and/or interfering with uncoating of the virus (Malhotra *et al.*, 1996).

1.6. Signature motifs in plant OMT

S-Adenosyl-L-methionine (AdoMet)-dependent *O*-methyltransferases (OMT's) are widely distributed in plants. They catalyze the transfer of a methyl group from S-adenosyl-L-methionine to an accepter molecule, with the concomitant formation of the corresponding methylated derivative and S-adenosyl-L-homocysteine (AdoHcy) as products. OMT's are considered as important plant enzymes that play a variety of roles in plant metabolism.

Plant OMT's share sequence similarity (Ibrahim, 1997), and certain conserved regions are believed to be functional domains of these OMT's. Multiple alignment of methyltransferases from various organisms revealed three semi-conserved motifs, I, II and III, that occur in S-adenosyl-L-mthionine-dependent methyltransferases (Kagan and Clark, 1994) and suggested that those conserved motifs contribute to the binding of S-adenosyl-L-methionine and/or the product S-adenosyl-L-homocysteine. Comparison of

27 different plant OMT's belonging to several plant families (Ibrahim *et al.*, 1997), showed five regions (I-V) in the carboxyl half of the enzyme including 36 amino acid residues that were highly conserved among the majority of the plant OMT's that were used in the study (Figure 3). Region I, which also showed similarity to motif I of Kagan and Clark (1994) was proposed to be involved in AdoMet binding, and that region IV is involved in the binding of Mg^{2+} (Ibrahim *et al.*, 1997).

Since the previous study by Kagan and Clark included only six plant AdoMet-MTases out of a total of eighty-four surveyed methyltransferases, a more recent study aimed to include only plant AdoMet-dependent methyltransferases (Joshi *et al.*, 1998). Although they were not successful in finding the motifs that were proposed by Kagan and Clark except for some similarity with motif I proposed by the latter authors, they were successful in proposing three conserved motifs, (A, B and C), which were highly conserved in 56 different plants methyltransferases, representing distinct groups of AdoMet-MT's such as *C*-methyltransferases, *N*-transferases. Among AdoMet-MT's, there is 100%, 98%, and 56% amino acid conservation for motifs A, B, and C respectively (Joshi *et al.*, 1998). That study showed that there is a close relation and a high degree of similarity between various groups of MT's, including OMT's, *N*-methyltransferases, and *C*-methyltransferases. This similarity is likely related to the fact that all these enzymes utilize AdoMet as the methyl donor, and thus have to share a common ability to bind this molecule. Such conserved signature motifs in plant methyltransferases, and of plant OMT's in particular should be useful for the identification of genes of unknown function in the public domain DNA sequence databases. In addition, those conserved motifs could be used to design primers for

isolating new members of AdoMet-dependent methyltransferases from plants (Gauthier *et al.*, 1996).

Although several enzymes in the proposed pathway for methylation in C. *americanum* (Figure 2) were characterized at the biochemical level (Ibrahim *et al.*, 1987), only two genes, both of which code for 3'/5' OMT's, have been isolated in this plant as cDNA clones (Gauthier *et al.*, 1996). In this study, genomic DNA sequences of *C. americanum* that contain OMT-like genes were obtained, and the sequence was used to isolate cDNA clones by PCR-based screening of a *C. americanum* cDNA library. A second screening by plaque hybridization identified additional cDNA clones which appear to be 3'/5' OMT's.

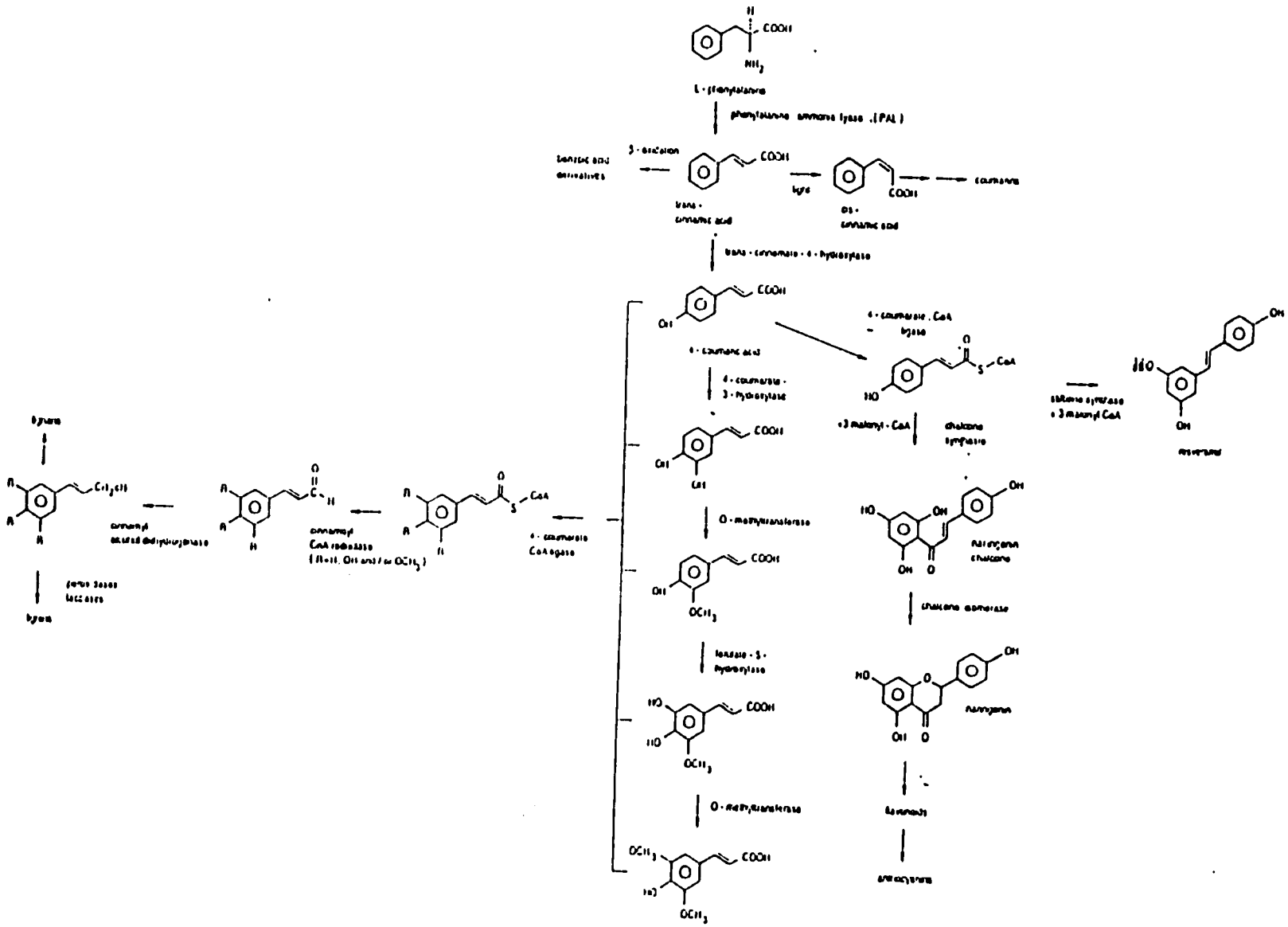


Figure 1: The phenylpropanoid pathways (Verpoorte *et al.*, 2000).

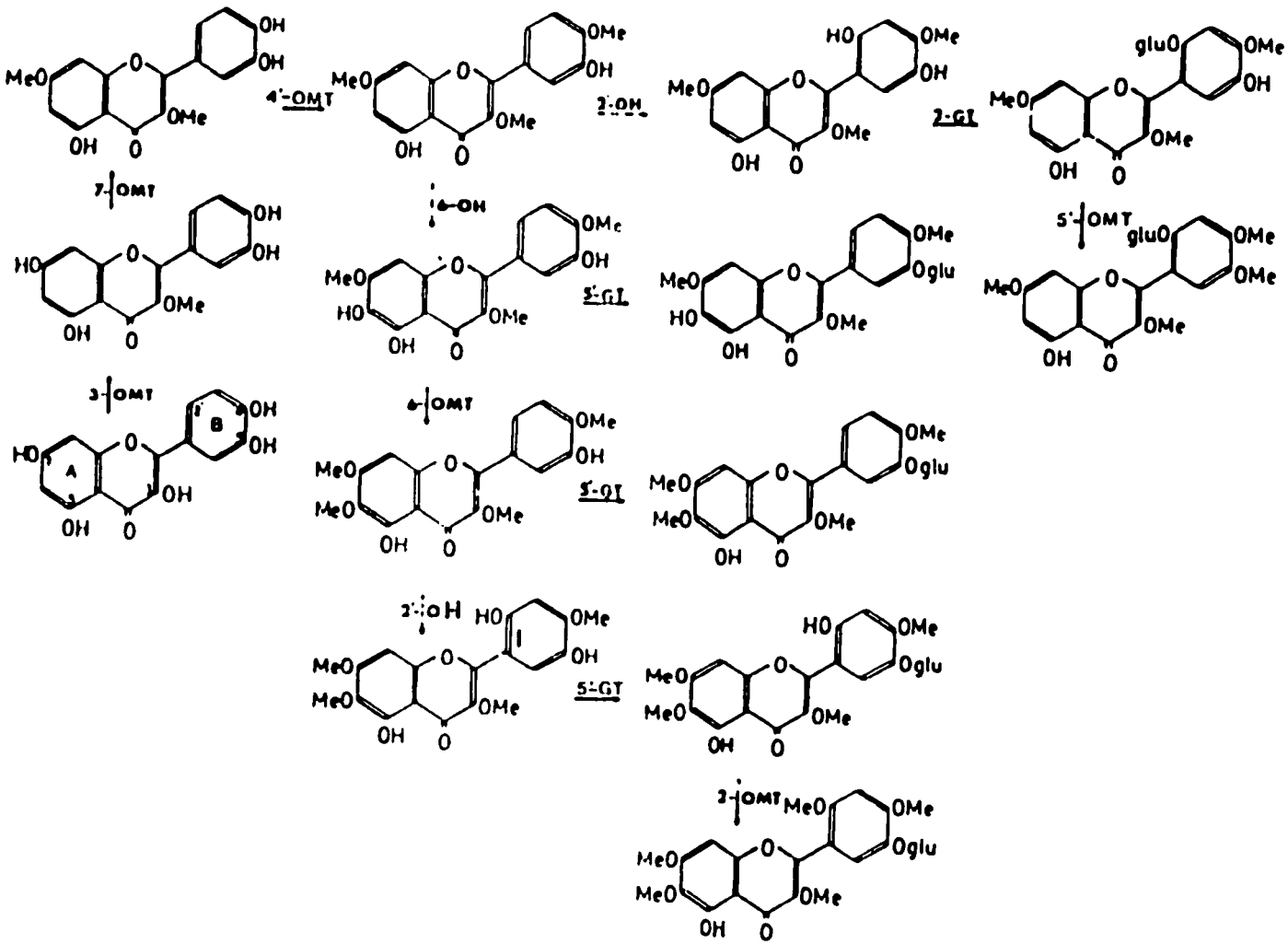


Figure 2: Proposed pathway for the enzymatic synthesis of polymethylated flavonol glucosides in *C. americanum*. OH, hydroxylase; OMT, *O*-methyltransferase; GT, *O*-glucosyltransferase. Dashed arrows are hypothetical steps (Khouri *et al.*, 1988).

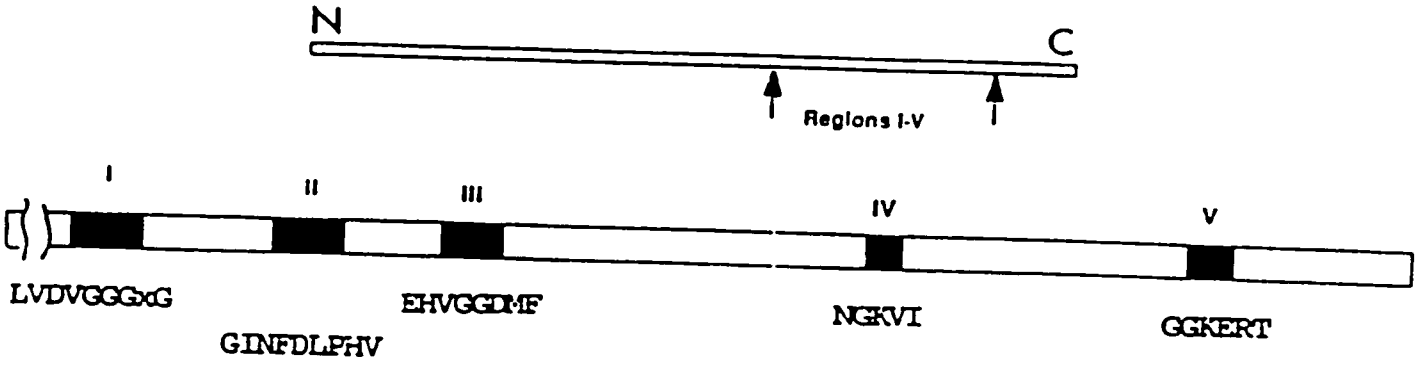


Figure 3: Diagram showing the relative positions of the *O*-methyltransferase consensus regions (I-V) within the polypeptide, and the sequences of the regionns. Regions I and IV are involved in *S*-adenosyl-L-methionine and metal binding, respectively (Ibrahim, 1997).

CHAPTER 2. MATERIALS AND METHODS

2.1.Study materials

A genomic DNA clone (11 KB) with sequence similarity to flavonoid OMT's had been previously isolated from *Chrysosplenium americanum* and had been partially sequenced. Further sequencing was done to complete 8 kb of contiguous genomic DNA that includes and flanks areas of similarity to OMT's. Primers based on these genomic sequences were designed and used for PCR based screening of a *C. americanum* cDNA library. cDNA fragments isolated by PCR were used as probes to screen the *C. americanum* cDNA library for OMT clones.

2.1.1. cDNA library

A cDNA library constructed from the plant *Chrysosplenium americanum* arial parts in λ Unizap vector from Stratagene. (Gauthier *et al.*, 1996) was available for this work.

2.1.2. Sequencing of genomic clones

A series of sub clones were previously prepared from the original genomic clone and sub-clones are mapped on a 11 KB genomic DNA fragment as shown in (Figure 4) (Gauthier, unpublished). Primers were designed for the ends of each known sequence fragment to sequence across the gaps. The sequencing reactions were done with the CEQ 2000 Dye Terminator Cycle Sequencing kit and were run on the Beckman CEQ2000 automated sequencer at the Center for Structural and Functional Genomics (Concordia

University). In addition, some of the sequencing reactions were done by the Center for Applied Genomics in the Sick Children Hospital, Toronto.

2.2. Compiling sequences, and producing the sequence of the full length genomic DNA fragment

The sequence of the 11 KB of genomic DNA was compiled from overlapping sequences using Blastn (Basic Local Alignment Search Tool) computer program at the NCBI web site (http://www.ncbi.nlm.nih.gov/BLAST/).

2.3. Primers for *C. americanum* cDNA Library screening

Primers were designed to be unique to each OMT like genomic DNA sequence. Specific primers for each of the two OMT-like sequences that were present in the cloned genomic DNA of *C. americanum* were chosen for the regions that showed the least similarity to other flavonoid OMT's amino acid sequence known in this species in order to amplify related clones from the cDNA library. Gene 1-B-F and Gene 1-B-R specific primers were used in PCR reaction with the cDNA library and with genomic DNA isolated from the plant to amplify the region corresponding to the genomic clone and verify that those OMT like clones are not an artifact of rearrangements during cloning. Primers sequences are shown in Table 1.

Gene 1	TM	Gene 2	TM °C
Gene1F 5'GATCTCGATGTCCTCAACTTGACGCTAAGT 3'	61.5	Gene2F 5' CTTGGTCATAGATTAAGGATC 3'	48
Gene1R 5'CTTAGCGTCAAGTTGAGGACATCGAGATC 3'	61.5	Gene2R 5' GATCCTTAATCTATGACCAAG 3'	48
Gene1-B-F 5' GATCATAGATGCTCAAAA TACGTACAT 3'	53	Genomic 2F 5' TTGGCAAGTGCATCCGTG	55
Gene1-B-R 5' AATCCCTGATCTATGCCAAG 3'	49.6	Gene2-A-R 5'CTGGGTTGGTTGAAGGCAAATAAGAAG 3'	58

Table 1: Oligonucleotide primers used in PCR reactions.

2.4. *C. americanum* cDNA library screen by PCR

PCR reaction conditions with Taq DNA polymerase (Takara) were set according to manufacturer's protocol. 5µL aliquots of phage from *C. americanum* Lambda Zap cDNA library with a concentration of 4.5 X 10^9 pfu/ml were lysed by heating for 5 minutes at 70 C°, chilled on ice for an additional 5 minutes, then used as templates for PCR. The amplification mixture consisted of 100 picomoles of each primer, 200µM of each dNTP, PCR buffer (75mM Tris-HCl (pH 8.8 at 25C°), 20mM $(NH_4)_2SO_4$, 0.01% Tween 20), 1.5 mM $MgCl_2$ and 2.5 units of Taq polymerase in a total volume of 100 µL. The PCR amplification program consisted of 30 cycles of denaturation (94 C° 1.5 min),

annealing (different temperature was used for different primers), 1 min), and primer extension (72 C°, 1 min).

The products of the PCR reaction were separated by electrophoresis on 0.8% agarose gels and verified by southern analysis using probes produced by PCR amplification of the genomic clone. Positive DNA bands were isolated from the gel using Qiagen II Agarose Gel Extraction Kit.

2.5. Cloning of isolated cDNA fragments

cDNA fragments that showed positive hybridization signal in the southern blot were cloned in the pGEM-T vector system (Promega, cat. #: A3600). Ligation reaction was carried out overnight according to the manufacturer protocol using T4 DNA ligase. Potential positive transformants were selected as white colonies among negative blue colonies.

2.6. Southern blot of cDNA fragments (PCR) products

Clones were then further tested by restriction enzyme digestion using Nco I and Sal I. Clones with inserts of the expected sizes were confirmed by southern hybridization, using a caffeic acid OMT (OMT-N) cDNA clone as a probe.

Southern blot analysis was done using 10-50 ng of DNA from a PCR reaction. DNA samples along with positive and negative controls were separated by electrophoresis on a 1% agarose gel. Electrophoresis was carried between 60 and 100V for 30 min, until the dye front migrated half way through the gel. The agarose gel was then soaked for 45 minutes in denaturing solution (1.5 M NaCl, 0.5 N NaOH) with gentle

agitation, the gel was rinsed briefly with de-ionized water, neutralized by soaking in 1 M Tris (pH 7.4), 1.5 M NaCl for 30 minutes at room temperature, and neutralization was repeated for 15 minutes with fresh solution. DNA was transferred to a Hybond-N nylon filter from Amersham by capillary transfer for 16 hours using 20X SSC as transfer buffer.

The probe used for southern blot hybridization was derived from a full length cDNA clone of caffeic acid OMT (OMT N, accession: U16793) from *C. americanum*. The clone was digested with Nco I and Sac I restriction enzymes. The insert was separated on 0.8 % agarose gel and purified using Qiagen II Gel Extraction Kit. The probe was labeled according to the supplier's protocol using a multiprime DNA labeling kit (Amersham) with P^{32} dCTP (3000 Ci/mmole, ICN). The probe was separated from unincorporated radioactive nucleotides with a filtered spin column.

Prehybridization of the southern blots membrane was done for 2 hours at 55 C° in hybridization solution containing 5 X SSC (75 mM sodium citrate, 0.75 M NaCl), 5 X Denhardt's solution (0.1 % bovine albumin, 0.1 % polyvinylpyrrolidone, 0.1 %Ficoll), 20 µg/mL sonicated salmon sperm DNA (denatured at 95 C°), 0.02 M Tris (pH 7.6), 0.5% SDS. Prehybridization solution was discarded. The radioactive probe was denatured by heating to 95 C° for 10 minutes, chilled on ice, and then added to the hybridization solution. The membrane was hybridized for 16 hours at 65 C°. Blots were then washed twice in 0.1 % SDS and 1 X SSC at 55 C° for 15 minutes. Nylon membrane containing DNA was exposed to X ray film for 16 hours at - 80 C°. Successfully isolated cDNA clones are listed in Table 2.

2.7. *C. americanum* cDNA Library screening by plaque hybridization

The purpose of cDNA library screening was to isolate novel flavonoid OMT cDNA clones that might have new position specificity. The cDNA fragment (7-1) that was isolated by PCR screening of the *Chrysosplenium americanum* cDNA library, and showed 99% identity to gene2 OMT-like genomic sequence was used as a probe. Positive plaques which hybridized to the 7-1G2 probe were selected for further study.

2.7.1. *C. americanum* cDNA Library titering

The cDNA library titer was determined according to protocols in the instruction manual of the Uni-ZAP XR Library (Stratagene). An overnight culture of XL1-BlueMRF- bacterial strain was grown in LB broth supplemented with 0.2 % (w/v) maltose and 10 mM $MgSO_4$. Cells were centrifuged at 2,000 rpm for 10 minutes, and resuspended in 10 mM Mg SO_4 at $OD^{600} = 0.5$. Serial dilutions of the *C. americanum* cDNA library was prepared in SM buffer (0.01 % gelatin, 0.05 M Tris-pH 7.5, 0.1 M NaCl, 8 mM $MgSO_4$). Aliquots of 2 µl were added to 200 µl of host cells. Phage and cells were incubated for 15 minutes at 37 C°, 3 ml of NZY top agar (1 % N-Z-amine casein hydrolysate, 0.1 M NaCl, 0.5 % yeast extract, 8 mM $MgSO_4$, 0.7% (w/v) agarose) was added to the cells, mixed immediately and plated on 100 mm NZY plates containing 1.5% agar. Plates were incubated at 30 C° overnight. The plaques were counted and the cDNA library titer was determined to be 1.75×10^4 pfu/ µl.

2.7.2. cDNA Library screen by plaque hybridization

A λ Zap cDNA library of *C. americanum* was plated on 150 mm petri dishes containing NZY solid media at the density of 50,000 plaques/plate. Each plate was lifted in duplicate using HybondN+ nylon membranes (Amersham). Membranes were treated in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 2 minutes, neutralized for 5 minutes in neutralizing solution (1.5 M NaCl, 0.5 M Tris-HCl (pH 8.0)), and rinsed for no longer than 30 minutes in a 0.2 M Tris-HCl (pH 7.5), 2 X SSC buffer solution. They were then air dried on Whatman® 3MM paper. DNA was crosslinked to the membranes using the autocross link setting on the Stratalinker® UV crosslinker (120,000 μJ) for 30 seconds.

The membranes were hybridized with a probe derived from cDNA clone 7-1G2 which has 100% sequence identity with the gene 2 genomic clones. To make the probe, the insert of clone 7-1G2 was gel purified using Qiagen II Gel Extraction Kit, and 100ng of DNA was labeled with P^{32} using a multiprime DNA labeling kit (Amersham) according to the supplier's protocol.

Positive plaques were selected and purified by two additional rounds of plating and hybridization, and then subjected to *in vitro* excision according to the protocol of Stratagene for λ-Zap clones. Plaque cores were transferred from agar plates and mixed with 500 μl SM buffer containing 4% chloroform, vortexed and incubated overnight at 4 C°. The phage stock was combined with an overnight culture of XL1-Blue MRF' cells at an OD 600 of 1.0, with the ExAssist helper phage. The culture was incubated in a Falcon 15 ml polypropylene tube at 37 C° for 15 minutes. Three ml of LB broth was added, followed with shaking for 3 hours at 37 C°. The culture was then heated for 20

minutes at 65 C°, centrifuged for 15 minutes at 1000 X g. The excised pBluescript phagemid was recovered in the supernatant. The phagmid was incubated with freshly grown SOLR cells (OD 600 = 1.0) for 15 minutes at 37 C°. Cells were then plated on LB-ampicillin (50 μg/ml) agar plates. The colonies were purified by re-streaking on fresh LB-ampicillin plates. Nine different colonies were streaked for each isolated cDNA clone. Colonies were regrown and used for colony hybridization to verify the identity of the clones. Colony lifts were done according to Maniatis *et al.*, 1999. Nylon membranes were lifted from the colonies' plates, and placed side up on 3 MM paper saturated with 10% SDS for 3 minutes, and transferred into denaturing solution (0.5N NaOH, 1.5 M NaCl) for 5 minutes. Filters were then transferred to 3 MM paper saturated with neutralizing solution (1.5 M NaCl, 0.5 M Tris.Cl (pH 7.4)) for 5 minutes, and then transferred to 2X SSC for 5 minutes. Filters were left to air dry for 30 minutes colony side up.

Hybridization of the colony lifts was done with the (7-1G2) probe following the same protocol described earlier for the cDNA library screening. Identity of positive clones was determined by single pass 5' end sequencing as designated above.

Forward primer	Reverse primer	Successful DNA fragment isolated from cDNA library
Gene1F 5'GATCTCGATGTCCTCAACTT GACGCTAAGT 3'	T7-vector primer 5'GTAATACGACTCACTATAGGGC 3'	none
T3-vector primer 3'ATTAACCCTCACTAAAG 5'	Gene1R 5'CTTAGCGTCAAGTTGAGGACATCG AGATC 3'	none
Gene1-B-F 5' GATCATAGATGCTCAAAA TACGTACAT 3'	Gene1-B-R 5' AATCCCTGATCTATGCCAAG 3'	cDNA fragment 445 bp clone 4-A-4G1
T3-vector primer 3'ATTAACCCTCACTAAAG 5'	Gene2R 5' GATCCTTAATCTATGACCAAG 3'	cDNA fragment 431 bp clone 7-1G2
Gene2F 5' CTTGGTCATAGATTAAGG ATC 3'	T7-vector primer 5'-GTAATACGACTCACTATAGGGC-3'	none
Genomic 2F 5' TTGGCAAGTGCATCCGTG 3'	Gene2-A-R 5'CTGGGTTGGTTGAAGGCAAATAAG AAG 3'	cDNA fragment 138 bp

Table 2: **Nucleotide base sequence of primers used in PCR reactions from cDNA library.** Primers that have the letter F are forward primers. Primers that have the letter R are reverse primers.

CHAPTER 3. RESULTS

3.1. Sequencing results of Genomic clones

A genomic clone containing a OMT-like gene sequence had been partially sequenced prior to the present work (Gauthier, unpublished data). In the work presented here, further sequencing was done using gene specific primers designed to fill the gaps in the sequence. Four gaps in the contiguous sequence were filled by the strategy shown in figure 4. The full-length sequence of the genomic DNA is 8021 bp (Figure 5). Two regions show similarity to OMT encoding genes previously characterized from the same plant (Table 3). The DNA sequence of 8021 bp of *C. americanum* genomic DNA is shown in figure 5, and the putative coding regions are indicated.

3.2. Sequence analysis of Gene1 OMT-Like genomic sequence

The translation of the genomic DNA sequence showed it contained two adjacent OMT-like sequences, which showed a high degree of similarity to other OMT's in the GenBank database. The translation of the first region, Gene1, from nucleotide 819 to 2885 (Figure 5) showed highest similarity to the caffeic acid OMT from *C. americanum* (Accession: U16793) with 86.9 % amino acid similarity (Figure 6), and to a significant but lesser degree of similarity with the 3'5' flavonoid OMT from *C. americanum* with 77 % amino acid (Accession: U16794). However, Gene1 shows several rearrangements relative to the known cDNA sequence of the caffeic acid OMT. A duplication of the region at position 1152 to 1175 which showed similarity to amino acids 112 to 119 of the caffeic acid OMT (Figure 6) was found at position 1235 to 1258. The coding region

corresponding to amino acids 156 to 219 in the caffeic acid OMT was absent from Gene1 sequence. There is a deletion at position 2714 in which there is no corresponding sequence to code for nine amino acids that are present in caffeic acid OMT from amino acid 265 to 274 (Figure 6). There are 1 codons missing at the 3' end of the sequence relative to the caffeic acid OMT cDNA coding for amino acid 245. The last block of coding sequence ends with an in-frame stop codon, but Gene1 has four blocks of coding sequence with three intervening regions that did not show similarity to known OMT cDNA sequences. However, the non-coding regions are not readily classified as introns. They are not boarded by classically conserved intron junction sequences. In most cases the ends of the coding regions correspond to aberrations such as deletions or duplications, indicating that of blocks of coding sequence may have been generated by chromosomal rearrangements.

3.3. Sequence analysis of Gene2 OMT-Like genomic sequence

The translation of the sequence for Gene 2 from nucleotides 3125 to 3966 (Figure 5) showed highest similarity to *C. americanum* caffeic acid OMT (Accession: U16793) (Figure 7) with 89.6 % amino acids similarity, and to a slightly lesser degree of similarity to *C. americanum* 3'5' flavonoid OMT (Accession: U16794) with 86.7 % similarity. However, Gene2 also showed several mutations. There was no start codon found for the Gene 2 sequence, but a stop codon was present (Figure 5). A frame shift mutation in the first block of coding sequence (Figure7) at position 3225 was found, in which there was a missing adenine at this position between two guanines. This base was present in the caffeic acid OMT cDNA nucleic acid sequence and was not found in the genomic DNA

sequence. The deleted base caused a frame shift mutation at this position and led to the loss of a codon that codes for amino acid 40 of the caffeic acid OMT. A second insertion and frame shift mutation was found at position 3664, and a third frame shift mutation and deletion was found at base 3880 corresponding to the region between amino acids 220 to 311 of *C. americanum* caffeic acid OMT (Figure7).

3.4. Partial cDNA clones from *Chrysosplenuim americanum* cDNA library isolated by PCR

A partial length cDNA clone for an OMT, 7-1G2, was isolated from *C. americanum* cDNA library by PCR. A gene 2 specific primer (OMT-Like genomic sequence) and a vector specific primer (T3) were used (Figure 9). This 431 bp cDNA fragment contains an open reading frame of 111 amino acids (Figure 10), and shows 99% nucleotide base identity to genomic Gene 2 coding region, with a single mismatch at base 188 (possibly a sequencing error) (Figure 11). The first 98 bp of the fragment correspond to the expected vector sequence derived from the cDNA library cloning vector (Figure 9). It is missing an intron-like gap that existed in the genomic sequence (Figure 11). The open reading frame of this cDNA fragment showed 84 % identity, and 90 % similarity to the caffeic OMT from *C. americanum*. It also showed 80 % identity and 85 % similarity to the *C. americanum* 3'5' flavonoid OMT.

A second partial length cDNA clone of 445 bp, 4-A-4G1, was isolated using gene 1 specific primers (OMT-Like genomic sequence) (Figure 12). Although the cDNA sequence showed 99% nucleotide identity to the genomic sequence (Figure 13), it

included a region corresponding to a non coding, putative intron in the genomic clone. This region extends from base 277 to base 335 of the cDNA clone (Figure 12).

A fragment was amplified from *C. amreicanum* genomic DNA by PCR using Gene 1-B-F and Gene 1-B-R specific primers. This regenerated a part of the existing genomic clone corresponding to a part of Gene1. This verified the OMT like clone is not an artifact of rearrangements that occurred during cloning. A genomic clone that showed 100 % identity to the genomic sequence was obtained.

3.5. cDNA library screen using a radioactive probe:

Sixteen OMT cDNA clones with DNA sequences identical to the flavonoid OMT's were isolated from *C. americanum* cDNA library using cDNA clone 7-G2 as a probe.

Multiple sequence alignment between the cDNA clones isolated in this work and previously isolated *C. americanum* OMT cDNA clones (Gauthier et al., 1996 . Gauthier et al., 1998 and Gauthier, unpublished data) was carried out. The alignment included 31 cDNA clones and showed that all OMT clones isolated in this work had very high sequence similarity to the *C. americanum* 3'5'flavonoid OMT (see Appendix 1). Sequence differences between caffeic acid OMT clones and the rest of the 3'5' flavonoid OMT clones are indicated in Appendix 1. Such differences in sequence are unique to each type of OMT, and are considered important for substrate specificity of the enzymes encoded by these genes.

Although cDNA clones isolated in this work appeared to be very similar to 3'5' flavonoid OMT clones previously isolated (Gauthier et al., 1996 and Gauthier,

unpublished Data) which are Omta, Omtl, Omtb, Omte, Omtp, Omtj, Omtf, Omtg, Omth and chrysoc, some variations in the 5' untranslated region (5'UTR) of these cDNA clones and the previously isolated cDNA clones were found. There are four variants observed for the 5' UTR region. The region has GAA trinucleotide and few GA dinucleotide repeats. The four variants differ in the presence and number of repeats. Polymorphism in the repeated region and sequence difference upstream of these repeats present in these four OMT clones may indicate allelic variation within the species or may be derived from divergence among duplicated genes.

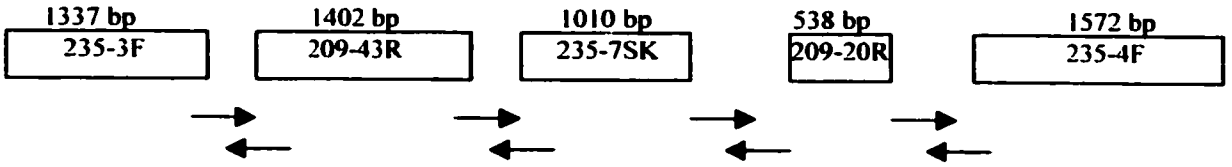
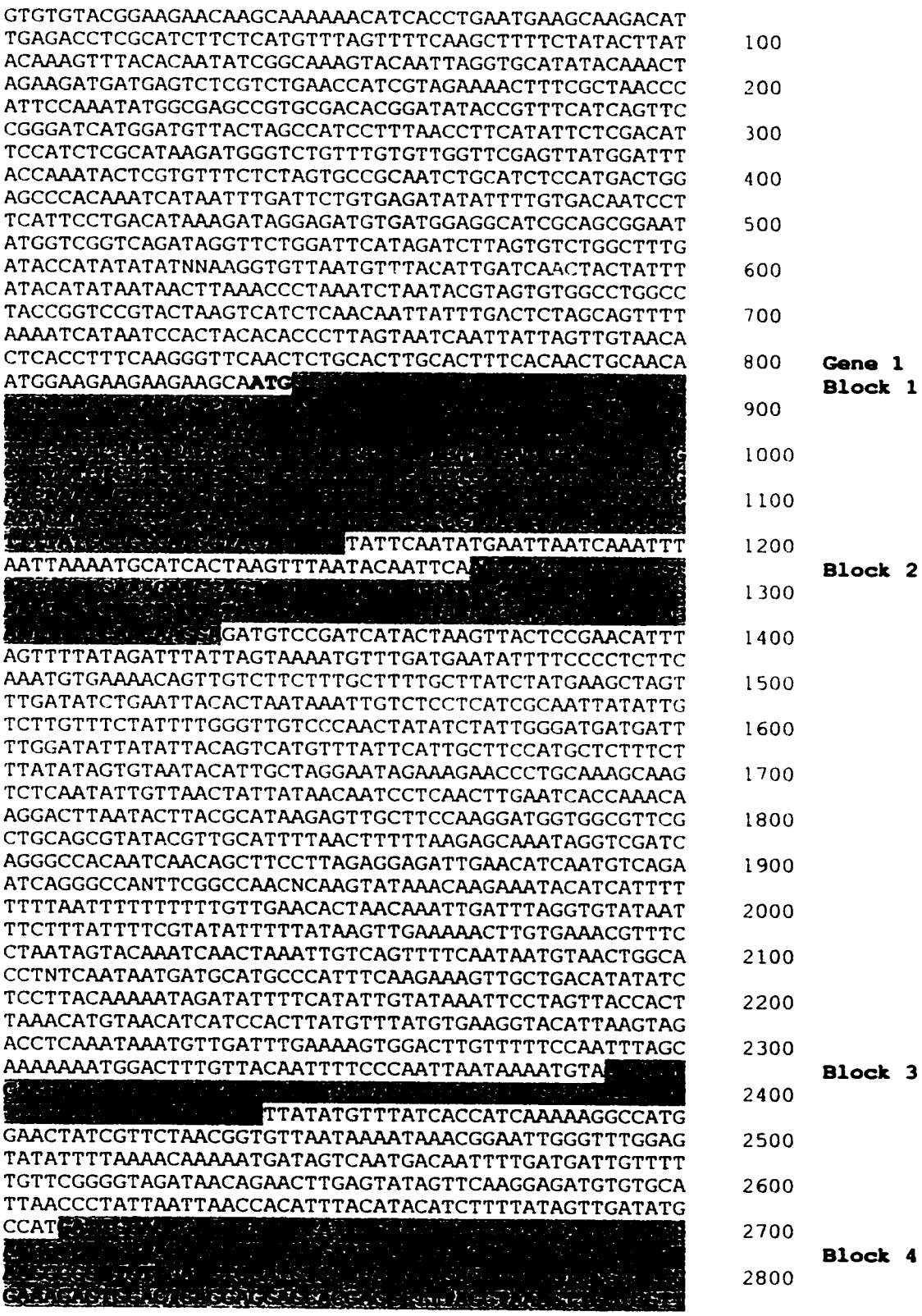


Figure 4: **Sequencing of *C. americanum* genomic DNA.** Boxes show regions of previously sequenced genomic DNA. Arrows show the direction of primers used for sequencing the gaps between previously sequenced blocks. Overlapping sequences in opposite directions were aligned and a 8 kb genomic sequence resulted from combining blocks and gaps.

Genomic DNA sequence containing OMT-like coding regions



[redacted]TTATTCTTGTTCAATGTTTTCATTAAA**TAA**AGACAAATATACATT 2900
ATTTCAAATTATTTTAAGCGACAAAACCAGAAGGTTGGCTGTGGTTGCAT
TATACTGAGTGTGCCTGCCATTATTTCCATTTTATTAAGCTATTTCAGCA 3000
TGATTGGAATGTACTTAAAACTAGATATCTTAAAAAAATAGATATTGTAT
CGAATGTATTAAAAACTTGCCTAATAAGCCACGAGAGGGCGACTCTATCC 3100 **Gene 2**
AATTATTATAATAACATTATCCTA[redacted] **Block 1**
[redacted]GG 3200
3300 **Block 2**
3400
[redacted]TACTACTATACTTACTTTTTTTATTTACTACTGCATCAAAGTATGAC 3500
TTAGTAATTTAATTTACTAGTGCATCAAAGTATGACTTAGTTTAATTAAT
ACCATTATTATTGCAGG[redacted] 3600 **Block 3**
[redacted]C 3700
3800 **Block 4**
3900 **Block 5**
[redacted]AACACT**TAG**TGTCTAGCTTCAAACTACTGGTGTT 4000
CTGTTGTAAATTCTATGTATATGTGGTCATTCTTCTAGTTTGTATATGCA
CCTAATTGTACTTTGCCGATATTGTGTAACCTTTGTATAAGTATAGAAAA 4100
CTTGAAACTAAACATGAGAAGATGCGAGTCTCAATGTCTTGACTTAATTC
GGGTGAGTTTCTCATCCGTCAAACATGTAGTGCCATGTCCAAACTCATGG 4200
TGTATAAATAATTTTCCATCTCTTGTTTTTGTTTTTAAGCCTATTCTTCT
ATAAAGCTCATATTCTATCAAATTATGATAGAACTAATCCTTGTAATAGT 4300
TGTTAGTTTGTTCAATATGTGCATGATTACCGTTCACCATACAAGAAAAA
GGGAAGATAGCGTTTTAATCTTGTTTGACATAGTTATTGTACTCTCATGT 4400
TTTAACAACAGATACTCTTTTATAGTCATGTTGTGAAACAGAAACCGTAA
GCAATGTGATGCGATAAATGAATATGCGATGACGCCGCGTGGATGAAAGG 4500
CATTGGAGGCAAGAGGATGGAGAGGTGGGACTCCCTTTTGGATGGAAATC
TATCCATTGTGTAGATATGATTATCCATCCTAGTTGGGTCAATATAAACC 4600
AAACCACAAGAGAGAAATAGAAGAAGAGAGTACTAGTAGTAGTAGTGCTG
AATTGATTAGATATAATGTCTTGTTCAACATCTGCAGCAGTGTTGTATGG 4700
ATCTTCCTTGTTATGTGGTGGAAGTAACAAGGCATTATTATTATTTTCAG
CTAGAGCTGTTAATACTCTTACTGCTACTCAGAAGAAGAAGTTTACTGTC 4800
ACTGCTAAGTCGAGGAAGTCTTGGATCCCTGCTGTTAAGGTCTCCGGCAA
CACCTTAGTCGACCCCATGGCGGAAATCATCGAGCACATGGTCCTTTTGA 4900
AGGTGAAAGGCGATACCGATCCTTTAACAATCGAAAGCATGATGAATGGT
CTCAAAAGCTTGAGCTCCTCCAAAGTTCTGTACCTCAACGCAGGCCCAAT 5000
TCACAAAAACAAGTCCTCGTCACCATTACGTTTTACCCACATTTTCCATG
CTAGATTCAACTCCAAACAAGACCTCGCAAATTACTTAGCACACATTAGT 5100
GAATTCCATGATAATAACATATCTCCCATTTGTGATGACATGATAGTCAT
GGACTGGGTCAATCCGGATGTAAATGGGCCGGTTGTCCTTAAACCCGGTT 5200
CGGCAATGAGGGTGACTTTCATGAAGTTGAAGGAAGGTTTGAGTGATGAG
AAGAAACACGAGGTTTATATGGTGGCGAAAGGGCTGAAAGGGTATTTCAA 5300
ATTCATTGAGCAAATGTAGTGTGGGGTGAGTTTCACAGTTGAGAGTAGCA
AAGGGTTTAGTATTGTGGCCGTGTCAATAAGCCCGGACATCGGTGAGTTG 5400
GAGGCCGCTGATTCGGGGGAGTATTTCAAGACATGCAAAGCTAACCCGTT
GCAGTATTTGGTATGCGATGTTGGTGTTTATTGCGTCATCAATTAATATT 5500
AATGTGTTGTTTCGTCTGAATAAGTAATGAAATAATGGTGTATGTATGTC
AGTATTTGTATCCTTATCCAGTTATCCCTATCAGGGTGAATAAATAAATA 5600
TTTATTTCATGCGGAAGCTTTGGCCATGGATATGAGAAAGCACAAGCAAA
GAAAAAAGATGCCATTTTTTTTCTACAAGGCTCATATTCTAGCAACAATG 5700
CTAGAACAAATCCTTGTATTCTTTTATAATTTCAACTTAATAATACTAAT
ATGTGATCACCATTTACCCACTTAAAAAATAAACGCATAATCGAGAAAAC 5800
AGGAAAAAACGAGTCACGGTCACCCACATGTGGAGTTGTTTGATGAACTG
TTCCTATTGCAACAAGTCTATTGTCGCAATAAGTCTAAGTGATAAAATAA 5900

TAAATCATTCAAAGTGTCTAATAAAATAATAATAGTAGTAGATAAATAAT
ATTCATGATGACATCTTGTAGATAATCCATTCTCTGCGATTATCATCTGG 6000
TTAATTCTCATCTGGTTACTATTTAAATATTCACCGAATTGAAATACCCT
TTCACCCCAACACTAGACTGGAAGTAACCAGGTGAATATTTAATTCTCAT 6100
CTGGTTACTTCCAGTCTTCCACTACATTATCATACTTCTTCATAGAAACA
GAATAATCATCTTCAAAAATTGATTGCAGGGAAATGCAAATATCAGCTTC 6200
AACAACGTCTTTCGATCATCATCTCGTCGGTGAGTTACTGGCAACACCGA
TGAAGGAATCGGTGTCTTCTGCTGAATCGCGATTTCTCTCTGCATCGACA 6300
ACCTCTTTCCACAATTTNNNNTTAGATCANCATCTTCAACTCGAAACCCT
AGATCTAAAGTATTAATTTATGTTTTAGATCCAATTTATNAAATTAGATC 6400
ATCATCTTCAACTCGAAACGCGATTTCTAAATATAAGTTTTAGATCTAAT
TTTTAATTTATGTTTGTGTGTATTGAGGACTCTGAGACTTGCTAGAGTCG 6500
ACTATGAGAAAAGATTTTAATTTGGTTCCTTTTGAATCGATTTTAATAAT
TGATTTATTTCCCAAATAATTTTCATTCATGAAATATTTAATTCAATCCA 6600
TTCAATAATCGATTTAGTTGTATTTCAGACTCCGAAATTATGTTTTACAT
ACTAATCCCAAATAGTGAAACGACTCATGCGTTTCATATCAAACACATAG 6700
AAATTTTCATATACGCATAACATCAAAACATAAAGAGTCAAGAGAATAGA
CATAGAAGAACTATACATTTATTTCTAGTCTAACATAATGCATGGCAAAA 6800
AAAAAAAAAAAAAAAAAAACTTAGTTTCCTGAACTGTCGAATTCTGAGGAA
TAGTTCTTTTACGAACCCATAATTTTCTCATGCGCTTTCCATTTCTTAGA 6900
TAGATTTTCTTGAAAGAGTAATTATGTACATCTGTTTACAGTTCTCTGCA
TTCGTTGCAGACTTGTATGACCTTCTTTTAGGAAGCAGGAAGATTGTAGG 7000
CCTTTTTGTTTGCCTTATATCCAATTCCAGTCTTTATGTGTGAGTTTCTT
TTAGCTTTCAAAATAGAGTTTAGTTTTTGAGTGCTCAGTTAGAATTTATT 7100
CAAATTTGACCGAGAGATGTTGGTCTTCTTGCTGGTAGGAATGACAATGG
ACCGGGTCGGGGAGATCCCATACACGGCCCGGCCGGGTTTTTTTTATCCC 7200
AGACCCGTACCTGTACCTGGGCGGGTCTAAAAAAAAATAAACCATACCCAG
CCCGTTAAATTACCCAGGGGTTCGGTTAGGGCCGGCCTGAAAAATAATAT 7300
TTTTTTTATCTAAAATTACAAACTCTTATATAAAATTAAAAAGTTTTTTT
TGGATAAAAAGATGGATTATATACAAACTTTTTTTTTTTGGATAAAATGA 7400
TGATTATATACAACCTATTTTTTGGATAAAAAGATGATTAAAAAATTATA
TAAAATGACAATTTNNNNTTCTAGGGCGGGCTGGGACGGGCCGATGATCG 7500
AATACCAGACCCGTACCTGAAACGGGTCTTCATATTTCATACCAAGTCCG
ACCCATACCCGATTTTTTTTTGGCGCCAACCCAGACCTGTCGGGTCCGTTC 7600
CAAATTGCCATCTCTACTTGCTGGGTTTCCCCTCTTTATTGTTGACAGAC
GTACGATCGCGTGTAAGAATTTTAGTACAACATTTTGATGACTTGTATTC 7700
ATCTGAGCTTTCTTCTTTTGAATTCGAAGTTAGACAAATGCTTTCATTCT
CGCGTGACTTCTGATCAGAAAGCTTGTCACTTAAGGATTTACAATACACC 7800
AACACAGCAAATTTGTGACGGCAGGCGCGTCACGTTTCTGACGGCTAAGT
GTCGTCACAAATATCTGGGCTGTCACAAATTGATTTGTGACGGTTTTAGC 7900
CGTCATAAATCTATCCAGCTTTTGATGGGTTTAGAAATAAAATTTCTGAC
TGCTGTTCACCGGCACAAATTGTAGATATAATTTCCCCCAACGTTCTTTC 8000
TTCTTTCCGGTCTTTTCTAGA 8021

Figure 5: DNA sequence of *C. americanum* genomic clone. Two OMT-like coding sequences are highlighted. ATG start codon and TAA stop codon of gene 1 and the TAG stop codon of gene 2 are in bold. The blocks of coding regions are defined by intervening non-coding sequences or by frame shift mutation.

Block 1

```
Gene1:   819 MLFAMQLAGASVLPMVLKSAIELDLLEIIDAQNTYMSPTEIASHLPTTNPDAATMIDRILR
             MLFAMQLA ASVLPMVLKSAIELDLLEII  Q+T MSPTEIASHLPTTNPDA  M+DRILR
Caffeic: 1   MLFAMQLACASVLPMVLKSAIELDLLEIIRGQDTCMSPTEIASHLPTTNPDAPAMVDRILR
acid OMT
```

```
Gene1:       LLSCYSXVTCSVRSDDDQRVYGLAPVCQYLTKNKQGVSIGALCLVIQDKV[redacted] 1175
             LLSCYS VTCSVRS DDQRVYGLAPVC+YLTKN+ GVSI ALCL+ QDKV ME WY L
Caffeic:     LLSCYSVVTCSVRSVDDQRVYGLAPVCKYLTKNQDGVSIAALCLMNQDKVLMESWYHL 119
acid OMT
```

Block 2

```
Gene1: 1235 [redacted]KDAVLNGGIPCNKEKGISXLDYLGIDPRFNRIFQQG 1366
            +MES YHLKDAVL+GGIP NK  G+S  +Y G DPRFN++F +G
Caffeic:112 LMESWYHLKDAVLDGGIPFNKAYGMSSFEYHGTDPRFNKVFNRG 155
acid OMT
```

Block 3

```
Gene1: 2345 GIEHVGGDMFVSVPKGDAIFMKVIC 2419
            GIEHVGGDMFVSVPKGDAIFMK IC
Caffeic:220 GIEHVGGDMFVSVPKGDAIFMKWIC 244
acid OMT
```

Block 4

```
Gene1: 2655 DWSDKHYLKLLKNCYDALPS---------ILPVVRDSSLATKGVVDLDVLKLTLSPVGKE 2807
            DWSD+H LKLLKNCYDALP+         ILP V DSSLATKGVV +DV+ +  +P GKE
Caffeic:246 DWSDEHCLKLLKNCYDALPN[redacted]ILPEVPDSSLATKGVVHIDVITVAHNPGGKE 305
acid OMT

Gene1: 2808 WTEEEHEALGKLAGFQ 2855
             TE+E EAL K AGFQ
Caffeic:306 RTEKEFEALAKAAGFQ 321
acid OMT
```

Figure 6: Alignment of Gene1 OMT-like sequence to the caffeic acid cDNA by blastx. The ORF of gene 1 OMT-like sequence aligned with the amino acid sequence of *C. americanum* caffeic acid OMT (accession: U16793). Sequence numbering for gene 1 corresponds to nucleic acid numbers in genomic DNA sequence of Figure 5. Numbers for caffeic acid OMT correspond to amino acids numbers of Genbank accession : U16793. Highlighted amino acids in block 1 and 2 is a duplicated sequence. Highlighted amino acids in block 4 is a deleted sequence in Gene1. Gene 1 sequence is 86.9 % similar to *C. americanum* caffeic acid OMT, regions of duplications are not included in full length percent similarity calculation. Blocks represent regions with high sequence similarity to OMT's separated by regions of no similarity.

Block 1

```
Gene2: 3125  LASASVLPMVLKSAIELDLLEIIGSQDACMSST 3223
             LA ASVLPMVLKSAIELDLLEII  QD CMS T
Caffeic: 7   LACASVLPMVLKSAIELDLLEIIRGQDTCMSPT 39
acid OMT
```

Block 2

```
Gene2: 3226  ISSYLPSTNPDAPAMIDRILCLLSCYSVVTCSVQSVDDQRVYGLAPVCKYLTKNQDGVCI 3405
             I+S+LP+TNPDAPAM+DRIL LLSCYSVVTCSV+SVDDQRVYGLAPVCKYLTKNQDGV I
Caffeic: 41  IASHLPTTNPDAPAMVDRILRLLSCYSVVTCSVRSVDDQRVYGLAPVCKYLTKNQDGVSI 100
acid OMT

Gene2: 3406  AALCLAVHDKVFMESW  3453
             AALCL   DKV MESW
Caffeic:101  AALCLMNQDKVLMESW  116
acid OMT
```

Block 3

```
Gene2: 3568  YHMKDAVLDGGIPFNKAYGMPIFDYLGHRLRI 3663
             YH+KDAVLDGGIPFNKAYGM  F+Y G   R
Caffeic:117  YHLKDAVLDGGIPFNKAYGMSSFEYHGTDPRF 148
acid OMT
```

Block 4

```
Gene2: 3665  NKVFNKAMSDHSTIIMKKVLETYKGFQGLRSVVDVGGGTGATLTMILSKYPTIQCINFDL 3844
             NKVFN+ MSDHSTI MKKV +TY+GFQGL S+VDVGGGTGATLTMILSKYPTI+CINFDL
Caffeic:149  NKVFNRGMSDHSTITMKKVFQTYQGFQGLTSLVDVGGGTGATLTMILSKYPTIRCINFDL 208
acid OMT

Gene2: 3845  PHVIQDAPEYPG 3880
             PHVI+DAPEYPG
Caffeic:209  PHVIEDAPEYPG 220
acid OMT
```

Block 5

```
Gene2: 3880  YEVLSKAAGFQAFQVCCNAFNTYIMEFSK 3966
             +E L+KAAGFQ FQV CNAFNTYI+EFSK
Caffeic: 311 FEALAKAAGFQGFQVFCNAFNTYIIEFSK 339
acid OMT
```

Figure 7: Alignment of Gene 2 OMT-like sequence to the caffeic acid cDNA by blastx. The ORF of gene 2 OMT-like sequence aligned with the amino acid sequence of *C. americanum* caffeic acid OMT (accession: U16793). Sequence numbering for gene 2 corresponds to nucleic acid numbers in genomic DNA sequence of Figure 5. Numbers for caffeic acid OMT correspond to amino acids numbers of Genbank accession : U16793. Blocks represent coding regions separated by non-coding regions, a frame shift mutation or duplication in the genomic DNA. Translation of gene 2 DNA sequence is 89.6 % similar to caffeic acid OMT, regions of duplications are not included in percent similarity calculation.

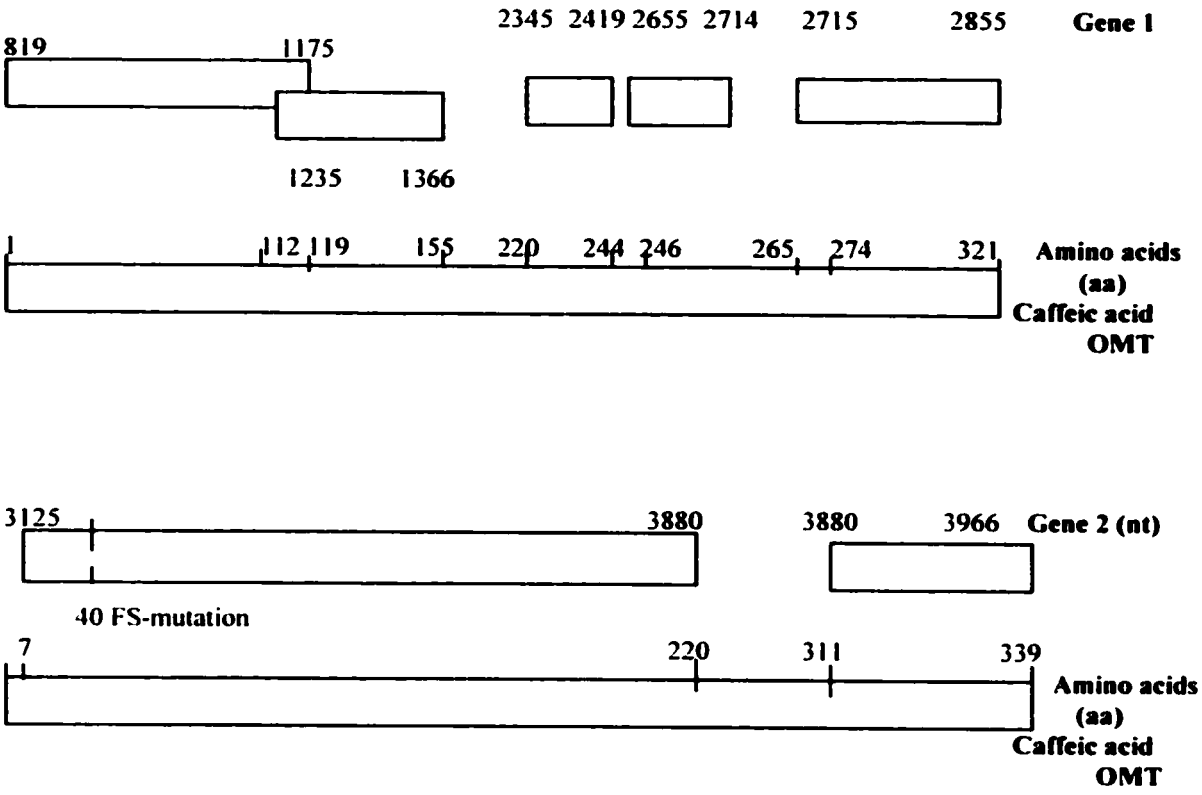


Figure 8: Alignment of Gene1, 2-genomic clones and caffeic acid OMT cDNA. The diagram shows the alignment of the genomic sequence with the caffeic acid OMT and represents deletions, duplications and location of frame shift mutations. Numbers for the genomic regions correspond to nucleotide positions in Figure 5. Numbers for the cDNA correspond to amino acid positions in Accession U16793.

	Gene 1	Gene2	3'5' Flavonoid OMT	Caffeic acid OMT
Gene 1	100%	85%	89.4%	79%
Gene2	85%	100%	91%	89.9%
3'5' Flavonoid OMT	89.4%	91%	100%	89%
Caffeic acid OMT	79%	89.9%	89%	100%

Table 3: Percent nucleotide sequence identity between Gene 1 and Gene 2 OMT-like sequences and caffeic acid and flavonoid OMT's nucleotide sequence. Blast two sequences computer program was used for comparison between gene 1 and gene 2 OMT-like genomic DNA sequences. Blast n to the database was used to compare each of the genomic DNA sequences to 3'5' flavonoid OMT and caffeic acid OMT.

	Gene 1	Gene2	3'5' Flavonoid OMT	Caffeic acid OMT
Gene 1	100%	74%	77%	86.9%
Gene2	74%	100%	86.7%	89.6%
3'5' Flavonoid OMT	77%	86.7%	100%	88%
Caffeic acid OMT	86.9%	89.6%	88%	100%

Table 4: Percent amino acid sequence identity between Gene 1 and Gene 2 OMT-like amino acid sequences and caffeic acid and flavonoid OMT's amino acid sequences. Blast two sequences computer program was used for comparison between gene 1 and gene 2 OMT-like genomic DNA sequences. Blastx to the database was used to compare each of the genomic DNA sequences to 3'5' flavonoid OMT and caffeic acid OMT amino acid sequence.

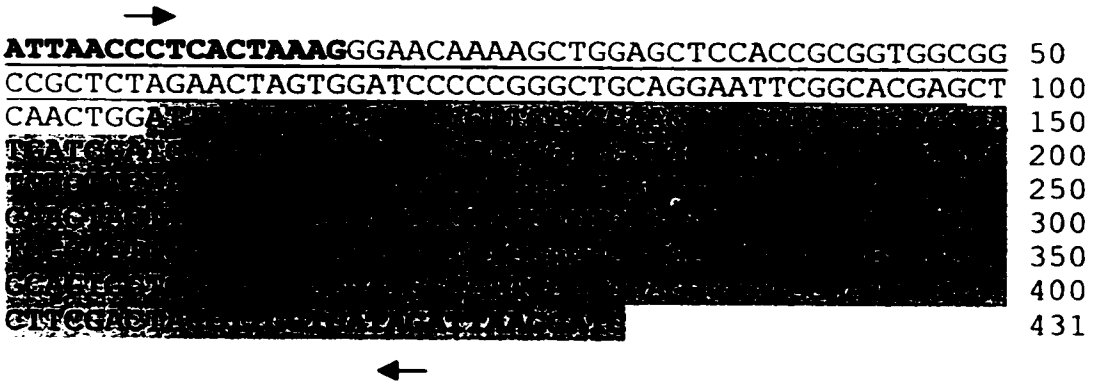


Figure 9: cDNA clone 7-1G2. T3 vector specific primer sequence and Gene 2 specific primer regions are in bold. Direction of both primers used in PCR amplification reaction from a cDNA library are indicated by arrows. cDNA fragment has an open reading frame of 111 amino acids, and shows 99% similarity to the genomic Gene 2 sequence. The underlined region corresponds to vector sequence from the cDNA library vector. Highlighted regions are regions with similarity to caffeic acid OMT.

L N W I S S Y L P S T N P D A P A **Met** I D R I L C L L S C Y S V V T C S V Q S V D
D Q R V Y G L A P V C K Y L T K N Q D G V C I A A L C L A V H D K V F **Met** E S
W Y H **Met** K D A V L D G G I P F N K A Y G **Met** P I F D Y L G H R L R I

Figure 10: Open reading frame of cDNA clone 7-1G2. The open reading frame is 111 amino acid long. It showed 84 % identity, and 90 % similarity to *C. americanum* caffeic acid OMT.

Block 1

```
7-1G2:     99  ctcaactggatttcttcttatttgccttcaaccaacccagatgcaccagccatgatcgat 158
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Genomic: 3217 ctcaactggatttcttcttatttgccttcaaccaacccagatgcaccagccatgatcgat 3276
DNA

7-1G2:    159  cgcattttgtgcctcctcttcttgctactc gtcgttacatgctctgttcaaagtgtcgat 218
               |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Genomic: 3277 cgcattttgtgcctcctcttcttgctactc gtcgttacatgctctgttcaaagtgtcgat 3336
DNA

7-1G2:    219  gatcaaagagtttacggtcttgcgcccgtttgcaagtacttgaccaaaaatcaagatggt 278
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Genomic: 3337 gatcaaagagtttacggtcttgcgcccgtttgcaagtacttgaccaaaaatcaagatggt 3396
DNA

7-1G2:    279  gtctgcattgctgctctttgtctcgctgttcatgataaagtctttatggagagctggtac 338
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Genomic: 3397 gtctgcattgctgctctttgtctcgctgttcatgataaagtctttatggagagctggtac 3456
DNA
```

Block 2

```
7-1G2:    334  ggtaccacatgaaagacgcagtgctcgatggtggcatcccattcaacaaagcctatggaa 393
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Genomic: 3566 ggtaccacatgaaagacgcagtgctcgatggtggcatcccattcaacaaagcctatggaa 3625
DNA

7-1G2:    394  tgcccatcttcgactaccttggtcatagattaaggatc 431
               |||||||||||||||||||||||||||||||||||||
Genomic: 3626 tgcccatcttcgactaccttggtcatagattaaggatc 3663
DNA
```

Figure 11: Sequence comparison of cDNA clone 7-1G2 with Gene 2 genomic DNA sequence. cDNA sequence showed 99% identity from base 99 to 338, and 100% from base 334 to 431 with gene 2. There was a region that aligned twice, from base 334 to 338. It is repeated in the genomic sequence at the boarders of two blocks of coding sequence in Gene1, but is a single repeat (ggtac) in the cDNA clone. The cDNA sequence is a continuous open reading frame, but when aligned with genomic DNA sequence, there appeared to be an intron region in the genomic DNA which did not align with the cDNA. The intron is from base 3457 to base 3570 or from base 3452 to 3565 of genomic DNA sequence. Mismatched bases between cDNA nucleotide sequence and genomic DNA are highlighted. Numbering of cDNA clone corresponds to numbering in figure 9, and numbering of the genomic sequence corresponds to numbering in figure 5.

GATCATAGATGCTCAAAATACGTACATGTCCCCAACTGAGATAGCTTCTC 50
ATTTGCCTACAACCAACCCTGATGCAGCAACCATGATCGATCGCATTTTG 100
CGCCTCCTCTCATGTTACTCTGCGTTACATGCTCTGTTCGAAGTGACGAT 150
GATCAAAGAGTTTACGGTCTTGCCCCTGTTTGCCAGTACTTGACCAAAAA 200
CAAACAAGGTGTCTCTATTGGTGCTCTTTGTCTCGTTATTCAAGACAAAG 250
TTTTTATGGAGGGATGGTACTCTCTTTATTCAATATGAATTAATCAAATT 300
TAATTAAAATGCATCACTAAGTTTAATACAATTCAATTATGGAAAGCAGG 350
TACCACTTGAAAGATGCAGTGCTAAATGGTGGCATCCCATGCAACAAAGA 400
GAAAGGAATTTCCGCCTTCGACTAC**CTTGGCATAGATCAGGCATT** 445

Figure 12: cDNA 4-A-4 G1 nucleotide sequence. A PCR fragment was amplified using primer gene 1-B-R and primer gene 1-B- F, (primers positions are in bold).

```
4-A-4 G1:  1     gatcatagatgctcaaaatacgtacatgtccccaactgagatagcttctcatttgcctac 60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
genomic: 899     gatcatagatgctcaaaatacgtacatgtccccaactgagatagcttctcatttgcctac 958
DNA

4-A-4 G1: 61     aaccaaccctgatgcagcaaccatgatcgatcgcattttgcgcctcctctcatgttactc 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
genomic: 959     aaccaaccctgatgcagcaaccatgatcgatcgcattttgcgcctcctctcatgttactc 1018
DNA

4-A-4 G1: 121    t-gcgttacatgctctgttcgaagtgacgatgatcaaagagtttacggtcttgcccctgt 179
                 | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
genomic: 1019    tngcgttacatgctctgttcgaagtgacgatgatcaaagagtttacggtcttgcccctgt 1078
DNA

4-A-4 G1: 180    ttgccagtacttgaccaaaaacaaacaaggtgtctctattggtgctctttgtctcgttat 239
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
genomic: 1079    ttgccagtacttgaccaaaaacaaacaaggtgtctctattggtgctctttgtctcgttat 1138
DNA

4-A-4 G1: 240    tcaagacaaagtttttatggagggatggtactctctttattcaatatgaattaatcaaat 299
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
genomic: 1139    tcaagacaaagtttttatggagggatggtactctctttattcaatatgaattaatcaaat 1198
DNA

4-A-4 G1: 300    ttaattaaaatgcatcactaagtttaatacaattcaattatggaaagcaggtaccacttg 359
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
genomic: 1199    ttaattaaaatgcatcactaagtttaatacaattcaattatggaaagcaggtaccacttg 1258
DNA

4-A-4 G1: 360    aaagatgcagtgctaaatggtggcatcccatgcaacaaagagaaaggaatttccgccttc 419
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||  | ||
genomic: 1259    aaagatgcagtgctaaatggtggcatcccatgcaacaaagagaaaggaatttccngcctc 1318
DNA

4-A-4 G1: 420    gactaccttggcatagatc 438
                 ||||||||||||||||||
genomic: 1319    gactaccttggcatagatc 1337
DNA
```

Figure 13: Sequence comparison of cDNA clone 4-A-4 G1 with Gene 1 genomic DNA sequence. cDNA sequence is 99% identical to genomic DNA. Numbering of 4-A-4 cDNA clone corresponds to numbering in figure 12, and numbering of genomic sequence corresponds to numbering in figure 5.

CHAPTER 4. DISCUSSION

Plant *O*-methyltransferases are believed to play key roles in stress signaling pathways, pathogen defense, growth and development (Booij-James *et al.*, 2000), (Malhotra *et al.*, 1996), (Ylstra *et al.*, 1992). Many *O*-methyltransferases have been characterized, and their genes have been cloned in various plants such as *Arabidopsis thaliana* (Muzac *et al.*, 2000) and *C. americanum* (Gauthier et al., 1996). The vast majority of OMT sequences that can be found in Genebank data base by sequence similarity search are yet to be characterized at the biochemical level. The study of *O*-methyltransferases function and specificity will provide a better understanding of the mechanisms involved in plant defense and development. In this study, cDNA clones of 3'/5' flavonoid OMT from *C. americanum* were isolated, their sequence identity and similarity was investigated.

Sequencing of a genomic DNA clone from *C. americanum* revealed some regions with similarity to the 3'5' flavonoid OMT and the caffeic acid OMT previously characterized from the same plant (Gauthier *et al.*, 1996, 1998). I used the blastn and blastx computer programs (NCBI) to verify the identity of these genomic sequences as well as to find out to which extent they differ or resemble the already characterized OMT's in the database. I showed that these sequences have open reading frames (Figure 6 and 7) that were similar but not identical to OMT's in the database. The alignment results showed that there were two different OMT-like sequences located in the 8 kb sequenced genomic DNA. The two OMT-like sequences Gene1 and Gene2 are separated by 269 base pairs (Figure 5).

Although these sequences had open reading frames, they also contained frame shift mutations, duplications and deletions. It was evident from the long deletions and the frame shift mutations that those gene-like sequences were unlikely to be translated in their present form, and are likely pseudogene or degenerate copies of genes that were encoded active OMT's.

It has been found that plant OMT's in particular have several conserved amino acid motifs (Ibrahim, 1997). Although these enzymes could have different specificity and function, plant OMT's conserve motifs important for AdoMet binding or magnesium binding, and those conserved amino acid sequences can be found in most if not all plant OMT's. In this study, some signature motifs were found in our genomic sequence. They were interestingly conserved in the genomic sequence although a number of mutations and deletions were found in other regions. Of the five signature motifs reported by Ibrahim 1997, Gene2 OMT-like sequence had the first two signatures while Gene1 OMT-like sequence has signatures three and five. This conservation supports the identity of these sequences as being derived from ancestral OMT's. I hypothesize that the ancestral OMT's could still be present and functional in the plant yet to be isolated.

While the 3'/5' flavonoid OMT from *C. americanum* and the caffeic acid OMT from the same plant share 88 % identity at the nucleotide level (Table 3), they have distinct substrate specificity and methylation regiospecificity (Gauthier *et al.*, 1996; Gauthier *et al.*, 1998). The fact that two enzymes with few amino acids differences could have distinct function provides strong evidence that *O*-methyltransferase substrate preference could be determined by a few amino acid residues and that new OMTs with different substrate specificity could begin to evolve from an existing OMT by mutation of

a few amino acids (Wang and Pichersky, 1999). In this study, the percentage identity between gene1 and gene2 OMT-like sequences as compared to *C. americanum* caffeic acid OMT was found to be 79 % and 89.9 % respectively (Table 3). We also found that gene1 and gene2 genomic sequences were 89.4 % and 91 % identical to *C. americanum* 3'/5' flavonoid OMT respectively. Gene1 and Gene2 are no more similar to each other than they are to the 3'/5' flavonoid OMT and the caffeic acid OMT. This indicates that they may have derived from different ancestral functional genes. Their tandem proximity to each other suggests that they are a duplicated and rearranged copy of ancestral genes that were also adjacent to each other and that those genes, in turn arose by tandem duplication. The Gene1 and Gene2 OMT-like sequences could be derived from OMT's that have a different substrate and position specificity than 3'/5' flavonoid OMT that was previously identified (Gautier *et al.*, 1996). I believe that the degree of similarity between these OMT-like sequences and the already characterized OMT's from this plant is enough to allow us to hypothesize that these OMT-like sequences are likely derived from unique OMT's with specificities different from those of the known OMT's of *C. americanum*, namely the flavonoid 3'/5' OMT and the caffeic acid OMT.

PCR analysis was used to screen a *C. americanum* cDNA library made from arial parts of the plant to identify cDNA clones representing our genomic sequences. Screening by PCR has been used in *C. americanum* for the successful isolation of a number of cDNA clones coding for flavonoid OMT's using degenerate oligonucleotides designed from conserved amino acid regions of various OMT's (Gauthier *et al.*, 1998). Since plant OMT's have been shown to have a high degree of sequence similarity (Ibrahim, 1997), we designed the primers used in the PCR reactions in regions that

showed very low sequence similarity to the already existing OMT's in order to isolate new OMT cDNA clones from *C. americanum*. We were successful in isolating a partial cDNA clone 7-1G2 where a specific primer (Gene2F) from gene2 and a T3 vector primer were used (Table 2). The isolated cDNA clone had an open reading frame of 111 amino acids and showed 99 % nucleotide identity to gene 2 OMT-like sequence. The amino acid sequence of cDNA 7-1G2 showed 90 % similarity to caffeic acid OMT and 85 % similarity to flavonoid 3'5' OMT from C. americanum. An intron region present in gene2 genomic sequence was spliced out in the cDNA clone (Figure 11). These findings suggested the presence of a unique OMT gene sequence in the *C. americanum*, in which gene2 OMT-like sequence was derived from.

A specific primer was designed in the same region of the primer Gene2R, but in the forward direction (Gene2F) (Table 2). This primer was used in a PCR reaction with a T7 vector primer to isolate the 3' end of gene2. I was not able to isolate the rest of the gene by this technique. It is possible that there is a mutation in the 3' end of the primer region, in which a product in the reverse direction was possible to obtain, but is difficult for the reaction to occur in the forward direction.

A second cDNA clone 4-A-4G1 was isolated from the *C. americanum* cDNA library. Two specific primers designed from genomic sequence gene1 OMT-like were used (Table 2) to obtain an internal cDNA fragment. cDNA clone 4-A-4G1 contained a frame shift mutation and an intron region, suggesting that a cDNA clone of a nonfunctional gene copy was isolated. It is possible that this region of the cDNA clone represents a pseudogene.

cDNA clone 7-1G2 isolated in this study was further used as a radioactive probe to screen a *C. americanum* cDNA library to obtain more OMT clones. Sixteen flavonoid OMT cDNA clones were isolated. Multiple sequence alignment including the sixteen new clones and fifteen previously identified cDNA clones (see Appendix) showed that OMT clones isolated in this work possess high sequence similarity to *C. americanum* 3'/5' flavonoid OMT. Although the new cDNA clones resembled the 3'/5' flavonoid OMT throughout the coding region, there existed some variations in the 5'UTR. These regions are used as a sign of the gene's origin. In this work, we showed that there were four gene variations between the OMT cDNA clones. Those variations indicate multiple clones derived from multiple gene family, or they could be allelic variations in the gene. The data presented in this work, cannot determine the reason for these variations since cDNA library was constructed from a pool of different *C. americanum* plants.

OMT's with different substrate specificity can be highly similar with only few amino acids determining their substrate specificity (Wang J. Pichersky E.,1999). Hybrid proteins having only seven amino acids replaced from another OMT developed a new substrate specificity and resembled the donor OMT (Wang J. Pichersky E., 1999). To further investigate the nature and identity of our genomic sequence, we searched for those seven amino acids in our sequence. The DNA sequence coding for those seven amino acids was found in Gene1 and some in Gene2. These regions had a unique amino acid sequence and could be a potential target for producing a hybrid protein that allows us to determine the specificity of our new gene-like sequences in future work.

The study of plant genes coding for enzymes such as flavonoid OMT's from *C. americanum* will help to reveal the mechanism used by plants to produce polymethylated

compounds. Changing the solubility of flavonoid compounds by addition of methyl groups leads to change in their compartmentation and their effectiveness as anti microbial compounds. The isolation and characterization of all genes coding for enzymes involved in the stepwise methylation pathway of flavonoids in *C. americanum* will lead to a better understanding of the mechanisms used in this plant to produce such important secondary metabolites. Manipulation of certain enzymes in the pathway upon characterizing of the genes involved will enable us to decrease or increase the production of a certain flavonoid depending on the benefits of such products.

LITERATURE CITED

BAJAJ, K. L., DE LUCA, V., KHOURI, H. IBRAHIM, R. K., 1983. Purification and properties of flavonoid-ring B glucosyltransferase from *Chrysosplenium americanum*. *Plant Physiol.* **72**: 891-896

BARRON, D. AND IBRAHIM R.K. 1996 Isoprenylated flavonoids- a survey. *Phytochem.*; **43:** 921-982

BIEZA, K., LOIS, R. 2001. An Arabidopsis mutant tolerant to lethal ultraviolet B levels shows constitutively accumulation of flavonoids and other phenolics. *Plant Physiol.* **126**: 1105-1115

BOOIJ-JAMES, S., DUBE, S.K., JANSEN, M.A.K., EDELMAN, M., MATTO, A.K. 2000. Ultraviolet-B radiation impacts light-mediated turnover of the photo system II reaction center heterodimer in Arabidopsis Mutants altered in phenolic metabolism. *Plant Physiol.* **124**: 1275-1283

BRUNET, G., IBRAHIM, R. K. 1980. O-Methylation of flavonoids by cell-free extracts of Calamondin orange. *Phytochemisrty* 741-746

CAUDRA, P., HARBORNE, J. B., WATERMAN, P. G. 1997. Increasing surface flavonols and photosynthetic pigments in Gnaphalium luteoalbum in response to UV-B radiation. *Phytochemistry* **45**: 1377-1383

CHAPMAN AND HALL, 1998. Dictionary Natural products on CD-Rom, London.

CHRISTENSSEN, A.B., GREGERSEN, P.L., OLSEN, C.E., COLLINGE, D.B. 1998. Flavonoid 7-*O*-methyltransferase is expresses in barley leaves in response to pathogen attack. *Plant Mol. Biol.* **36**: 219-227.

DE LUCA, V, IBRAHIM, R. K., 1985. Enzymatic synthesis of polymethylated flavonols in *Chrysosplenium americanum*. - I- Partial purification and some properties of *S*-Adenosyl-L-methionine flavonol 3-, 6-, 7-, and 4'-*O*-methyltransferases. *Arch. Biochem. Biophys.* **238**, 2: 596-605.

DE LUCA, V., IBRAHIM , R. K., 1982. Characterization of three distinct flavonol O-methyltransferases from *Chrysosplenium americanum*. *Phytochem.* **21**, 7: 1537-1540.

DE LUCA, V., IBRAHIM, R. K., 1985. Enzymatic synthesis of polymethylated flavonols in *Chrysosplenium americanum*. – II-. Substrate interaction and product inhibition studies of flavonols 3-, 6-, and 4'-*O*-methylatransferases. *Arch. Biochem. Biophys.* **238**, 2: 606-618.

ELDIK, G. J. V., REIJNEN, W. H., RUITER, R. K., HERPEN, M., SCHRAUWEN, J. A. M.,WULLEMS, G. J. 1997. Regulation of flavonol biosynthesis during anther and pistil development, and during pollen tube growth in *Solantum tuberosum*. *Plant Journal*, **111**: 105-113

GARO, E., MAILLARD, M., ANTUS, S., MAVI, S., HOSTETTMANN, K. 1996. Five flavans from Mariscus psilostachys. *Phytochemistry*. **43**, 1265-1269

GAUTHIER, A., GULICK, P.J., IBRAHIM, R.K. 1998. Characterization of two cDNA clones which encode *O*-methyltransferases for the methylation of both flavonoid and phenylpropanoid compounds. *Arch. Biochem. Biophys*. 2, **351**: 243-249

GAUTHIER, A., GULICK, P.J., IBRAHIM, R.K., 1996. cDNA cloning and characterization of a 3'/5' -*O*-methyltransferase for partially methylated flavonols from *Chrysosplenium americanum*. *Plant Mol. Biol.* **32**:1163-1169.

HAGA, M., AKASHI, T., AOKI, T., AYABE, S. 1997. A cDNA for *S*-adenosyl-*L*-methionine:isoliquiritigenin/licodione 2'-*O*-methyltransferase from cultured licorice (*Glycyrrhiza echinata L.)* cells. *Plant Physiol.* 113:663.

HE, X.-E., REDDY, J.T., DIXON, R.A., 1998. Stress responses in alfalfa (*Medicago sativa L.)*. XXII. CDNA cloning and characterization of an elicitor-inducible isoflavone 7- *O*-methyltransferase. *Plant Mol. Biol.* **36**:43-54.

IBRAHIM, R. K. DE LUCA, V. KHOURY, H., LATCHINIAN, L., BRISSON, L., CHAREST, P. M., 1987. Enzymology and compartmentation of polymethylated flavonol glucosides in *Chrysosplenium americanum*. *Phytochem.* **26** No. 5: 1237-1245.

IBRAHIM, RAGAI. K., BRUNEAU, ANNE., BANTIGNIES, BRIGITTE. 1998. Plant *O*-Methyltransferases: molecular analysis, common signature and classification. *Plant Mol Biol* **36**: 1-10

IBRAHIM,R.K. 1997. Plant *O*-Methyltransferase signatures. *Trends Plant Sci.* **2**: 249-250

Ibrahim, R. K. and Musac, I., The methyltransferase gene superfamily: A tree with multiple branches. In: Evolution of metabolic pathways, recent advances in phytochemistry volume 34. J. T. Romeo, R. K. Ibrahim, L. Varin and V. de Luca (Eds), Pergamon Press, Amesterdam, 2000.**480**: 349-384

JOSHI, P. CHANDRASHEKHAR., CHIANG, VINCENCENT. L. 1998. Conserved Sequence Motifs in Plant *S*-adenosylmethionine-dependent Methyltransferases. *Plant Mol. Biol.*. **37**: 663-674

KAGAN, RON M., CLARK, STEVE. 1994. Widspread Occurrence of Three Sequence Motifs in Diverse *S*-adenosylmethionine-Dependent Methyltransferases Suggests a Common Structure for these Enzymes.Arch.Biochem. *Biophys.* **310**: No. 2, 417-427

KHOURI, H. E., ISHIKURA, N., IBRAHIM, R. K., 1986. Fast protein chromatographic purification and some properties of a partially *O*-methylated flavonol glucoside 2'-/5'-*O*-methyltransferases. *Phytochem.* **25** No. 11: 2475-2479.

KOOTESTRA, A. 1994. Protection from UV-B induced DNA damage by flavonoids. *Plant Mol. Biol.* **26**: 771-774

LAUNDRY, L.G, CHAPPLE, C.C.S., LAST, R. L. 1995. Arabidopsis mutants lacking phenolic sunscreens exibit enhanced ultraviolet-B injury and oxidative Damage. *Plant physiol.* **109**: 1159-1166

LI., J., QU-LEE, T.M., RABA, R., AMUNDSON, R.G., LAST, L. R. 1993. Arabidopsis flavonoid mutants are hypersensitive to UV-B irradiation. *The Plant cell* **5**: 171-179

LONG, SHARON R. 1989. Rhizobium-Legume Nodulation: Life together in the underground. *Cell* **56:** 203-214

MAHER, E.A., BATE, N.J., NI, W., ELKIND, Y., DIXON, R. A. 1994. Increased disease susceptibility of transgenic tobacco plants with suppressed levels of preformed phenyl propanoid products. *Proc. Natl. Acad. Sci. USA* **91**: 7802-7806

MALHOTRA, B. ONYLAGHA, J.C., BOTT M.B.A., TOWERS, G.H.N., JAMES, D., HARBORNE, J. B.,FRNCH, C. J., 1996. Inhibition of Tomato ringspot virus by flavonoids. *Phytochemistry.* **43**: 1271-1276. No.6

MAXWELL, C.A., HARRISON, M.J., DIXON, R.A. 1993. Molecular characterization and expression of alfalfa isoliquiritigenin 2'-*O*- methyltransferase, an enzyme specifically involved in the biosynthesis of an inducer of *Rhizobium meliloti* nodulation genes. *Plant J.* **4**: 971-981.

MIDDLETON, E., TERAMURA, A. H. 1993. The role of flavonol glycosides and carotenoids in protecting soybean from Ultraviolet-B damage. *Plant Physiol.* **103**: 741-752

MUZAC, I., NOCE, A., ZHANG, H., IBRAHIM, R.K., 1999. Purification and characterization of a novel flavonol 3'-*O*-methyltransferase cDNA clone from *Arabidopsis thaliana.* PSNA conf. Proc., Poster Abstract No. 29, p.46.

OLSSON, L. C. VEIT, M., W EISSENBOCK., G., BORNMAN, J. F. 1998. Flavonoid response to UV-B radiation in *Brassica napus*. *Phytochemistry* **49**, 1021-1028

RUAN, Y., KOTRAIAH V., STRANEY, D. C. 1995. Flavonoids simulate spore germination in *Fusarium solani* pathogenic on legumes in a manner sensitive to inhibitors of cAMP-Dependent Protein Kinase. *MPMI* **8**: 929-938 No. 6

SKALTSA, H., VERYKOKIDOU, E., HARVALA, C., KARABOURNIOTIS, G., MANETAS, Y. 1994. UV-B protective potential and flavonoid content of leaf hairs in *Quercus ilex*. *Phytochemistry* **37**, 987-990

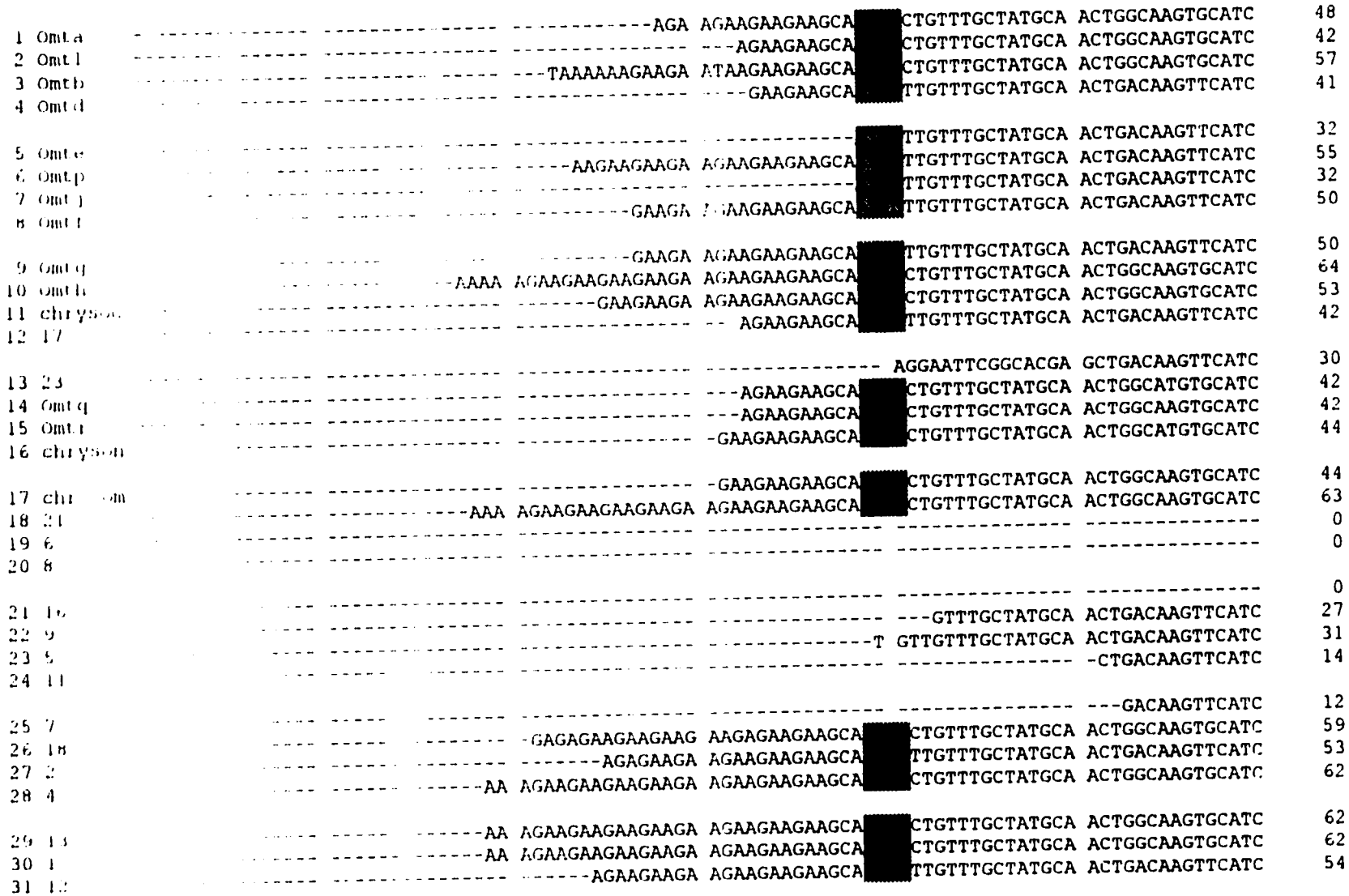
VERPOORTE, R., ALFERMANN, A.W., 2000. Metabolic engineering of plant secondary metabolism, 1-29

WANG, J., PICHERSKY, E., 1999. Identification of specific residues involved in substrate discrimination in two plant *O*-methyltransferases. *Arch Biochem. Biophys.* 1; **386** (1): 172-180.

WU, Q., PREISING, C.L., VANETTEN, H.D. 1997. Isolation of the cDNAs encoding (+) 6a-hydroxymaackiain 3-*O*-methyltransferase, the terminal step for the biosynthesis of the phytoalexin pisatin in *pisim sativum L. Plant Mol.* Biol. **35**:551-560.

YLSTRA, B., TOURAEV, A., MORENO. R.M.B, STROGER, E., TUNEM, A. J. V., VINCENTE, O., MOL, J. N. M., HEBERTE-BORS, E. 1992. Flavonols stimulate development, germination, and the tube growth of Tobacco pollen. *Plant Physiol.* **100**: 902-907

APPENDIX



```
 1 Omt a    CGTGTTACCAATGGT ACTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   135
 2 Omt l    CGTGTTACCAATGGT ACTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   129
 3 Omt b    CGTGTTACCAATGGT ACTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   144
 4 Omt d    CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   128

 5 Omt e    CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   119
 6 Omt p    CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   142
 7 Omt j    CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   119
 8 Omt f    CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   137

 9 Omt g    CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   137
10 Omt h    CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAAGAGG TCAAGATACGTG CATGTCCCCAACTGA   151
11 chryso   CGTGTTACCAATGGT ACTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   140
12 17       CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   129

13 23       CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   117
14 Omt q    CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAAGAGG TCAAGATACGTG CATGTCCCCAACTGA   129
15 Omt r    CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAAGAGG TCAAGATACGTG CATGTCCCCAACTGA   129
16 chryson  CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAAGAGG TCAAGATACGTG CATGTCCCCAACTGA   131

17 chrysom  CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAAGAGG TCAAGATACGTG CATGTCCCCAACTGA   131
18 21       CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAAGAGG TCAAGATACGTG CATGTTCCCAACTGA   150
19 6        ---- --------- ---------- ---------- ----------AG GCAAGATACGTG CATGTCCCCAACTGA    29
20 8        ---- --------- ---------- AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA    57

21 16       ---- --------- ---------- AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA    57
22 9        CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   114
23 5        CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   118
24 11       CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   101

25 7        CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA    99
26 18       CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAAGAGG TCAAGATACGTG CATGTCCCCAACTGA   146
27 2        CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   140
28 4        CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAAGAGG TCAAGATACGTG CATGTCCCCAACTGA   149

29 15       CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAAGAGG TCAAGATACGTG CATGTCCCCAACTGA   149
30 1        CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA   141
31 12       CGTGTTACCAATGGT CCTCAAATCGGCCAT AGAACTAGACCTGTT GGAGATCATAGCTAG TCAAGATACGTG CATGTCCCCAACTGA
```

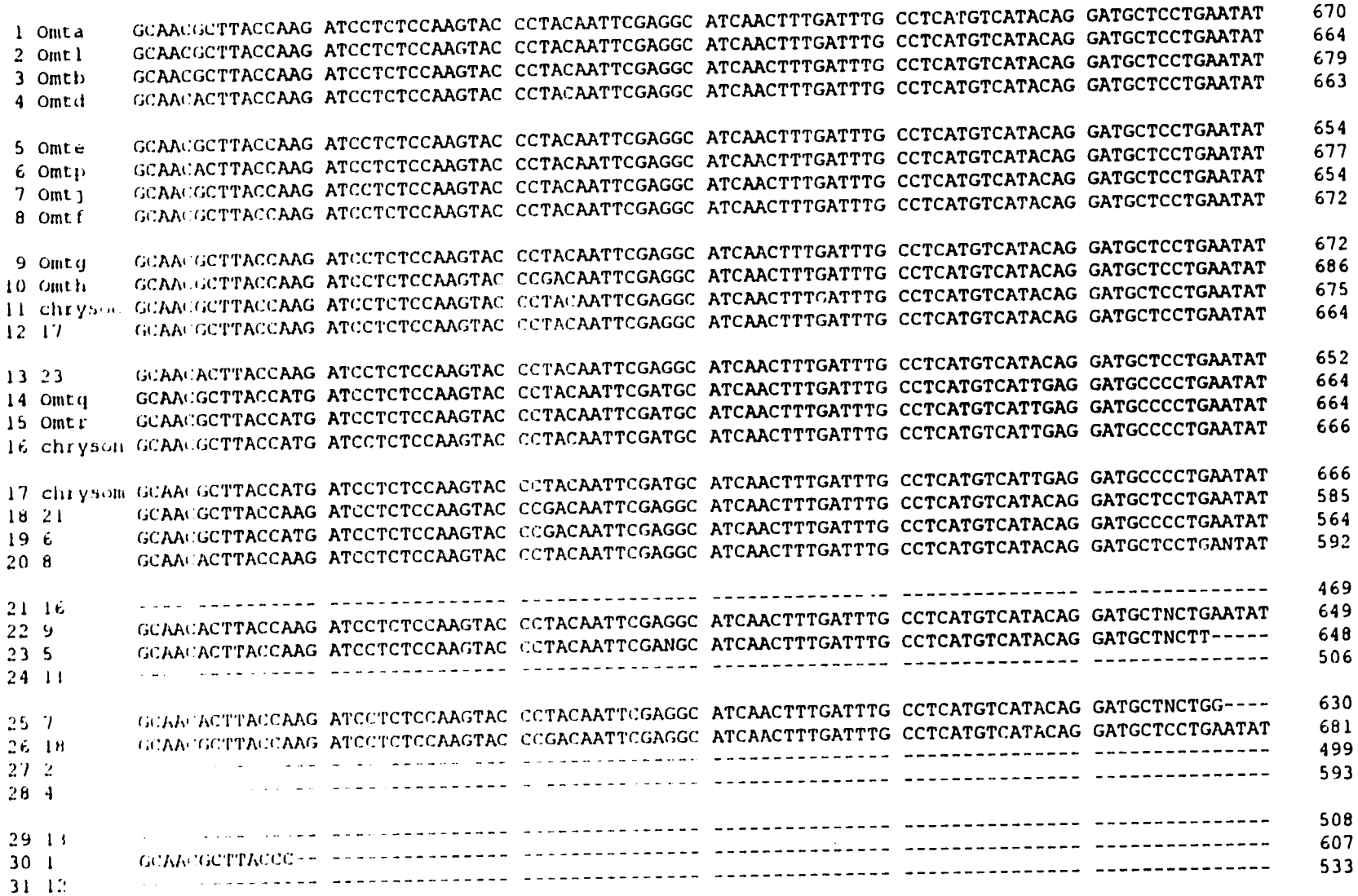
```
 1 Omta     GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   225
 2 Omtl     GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   219
 3 Omtb     GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   234
 4 Omtd     GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   218

 5 Omte     GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   209
 6 Omtp     GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   232
 7 Omtj     GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   209
 8 Omtf     GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   227

 9 Omtg     GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   227
10 Omth     GATTGCTTCTCATTT GCCTACAACCAACCC TGATGCACCAGCCAT GGTCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   241
11 chrysOc  GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   230
12 17       GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   219

13 23       GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   207
14 Omtq     GATTGCTTCTCATTT GCCTACAACCAACCC TGATGCACCAGCCAT GGTCGATCGCATTTT GCGCCTCCTCTCGTG CTACTCTGTCGTCAC   219
15 Omti     GATTGCTTCTCATTT GCCTACAACCAACCC TGATGCACCAGCCAT GGTCGATCGCATTTT GCGCCTCCTCTCGTG CTACTCTGTCGTCAC   219
16 chrysOm  GATTGCTTCTCATTT GCCTACAACCAACCC TGATGCACCAGCCAT GGTCGATCGCATTTT GCGCCTCCTCTCGTG CTACTCTGTCGTCAC   221

17 chrysOm  GATTGCTTCTCATTT GCCTACAACCAACCC TGATGCACCAGCCAT GGTCGATCGCATTTT GCGCCTCCTCTCGTG CTACTCTGTCGTCAC   221
18 21       GATTGCTTCTCATTT GCCTACAACCAACCC TGATGCACCAGCCAT GGTCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   240
19 6        GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   119
20 8        GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   147

21 16       GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   147
22 9        GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   204
23 5        GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   208
24 11       GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   291

25 7        GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   189
26 18       GATTGCTTCTCATTT GCCTACAACCAACCC TGATGCACCAGCCAT GGTCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   236
27 2        GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   230
28 4        GATTGCTTCTCATTT GCCTACAACCAACCC TGATGCACCAGCCAT GGTCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   239

29 13       GATTGCTTCTCATTT GCCTACAACCAACCC TGATGCACCAGCCAT GGTCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   239
30 1        GATTGCTTCTCATTT GCCTACAACCAACCC TGATGCACCAGCCAT GGTCGATCGCATTTT GCGCCTCCTCTCGTC CTACTCTATCGTTAC   239
31 12       GATTGCTTCTCATTT GCCTACAACCAACCC TCATGCACCAACCAT GATCGATCGCATTTT GCGCCTCCTCTCGTG TTACTCTGTTGTTAC   231
```

1	Omta	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	315
2	Omtl	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	309
3	Omtb	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	324
4	Omtd	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	308
5	Omte	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	299
6	Omtp	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	322
7	Omtj	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	299
8	Omtf	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	317
9	Omtg	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	317
10	Omth	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	331
11	chrys...	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	320
12	17	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	309
13	23	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	297
14	Omtq	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACGGTCTTGC ACCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	309
15	Omtr	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACGGTCTTGC ACCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	309
16	chrys...	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACGGTCTTGC ACCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	311
17	chrys...	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACGGTCTTGC ACCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	311
18	21	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	330
19	6	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATGA AGATGGTGTCTCCAT	209
20	8	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	237
21	16	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	237
22	9	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	294
23	5	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	298
24	11	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	281
25	7	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	279
26	18	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	326
27	2	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	320
28	4	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	329
29	13	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	329
30	1	GTGCTCTGTTCGAAG TGTCGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	321
31	12	ATGCTCAGTTCTAAG TGACGATGATCAAAG AGTTTACAGTCCTGC GCCAGTTTGCAAGTA CCTGACCAAAAATCA AGATGGTGTCTCCAT	

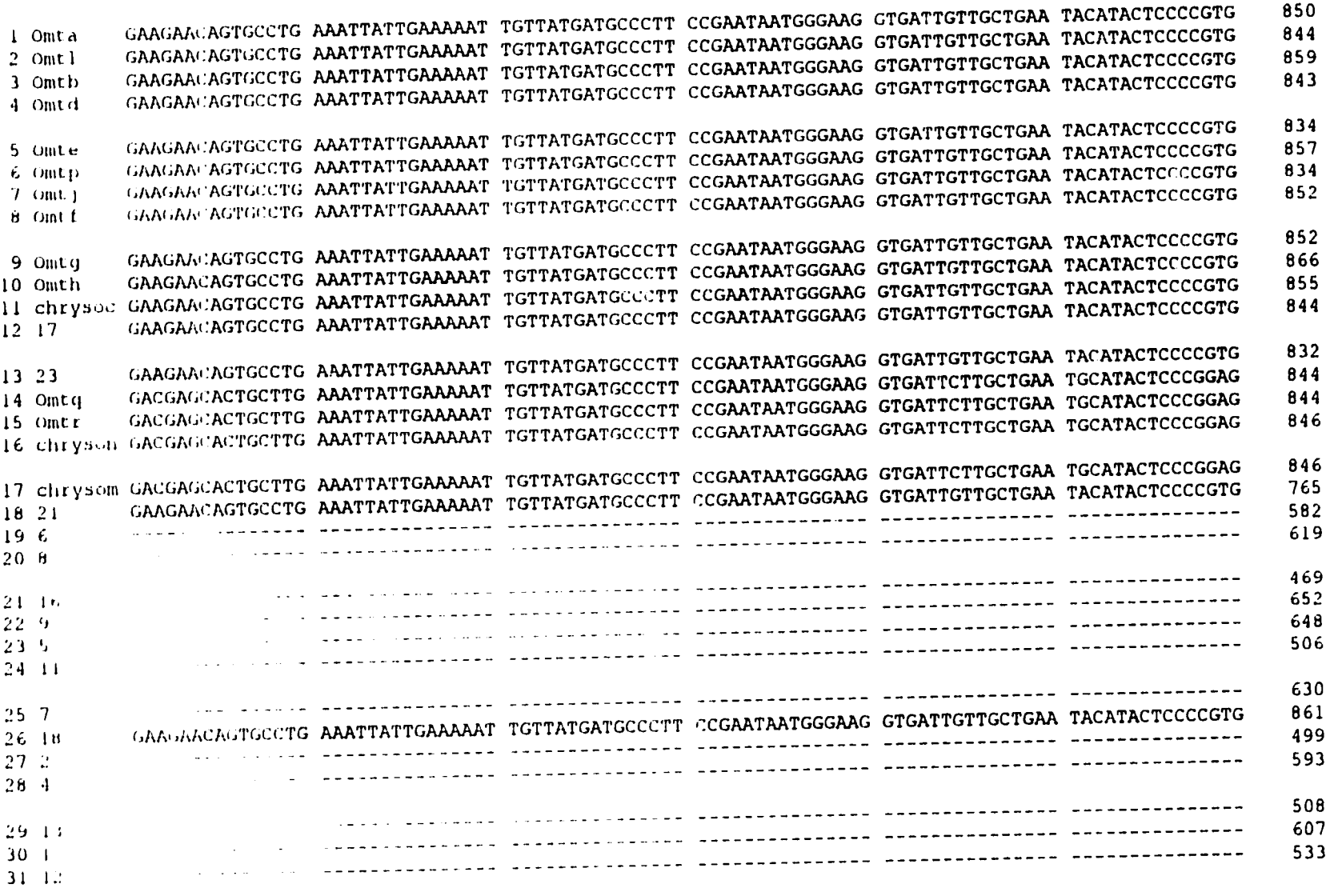
```
 1 Omt a    TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   404
 2 Omt l    TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   398
 3 Omt b    TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   413
 4 Omt d    TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   397

 5 Omt e    TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   388
 6 Omt p    TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   411
 7 Omt j    TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   388
 8 Omt f    TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   406

 9 Omt q    TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   406
10 Omt h    TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   420
11 chryso   TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   409
12 17       TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCATGGCTCG ATGGTGGCATCCCAT   389

13 23       TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCATGGCTCG ATGGTGGCATCCCAT   386
14 Omt q    TGCTGCTCTCTGTCT CATGAATCAAGATAA AGTCCTTATGGAGA GCTGGTACCACTTGA AAGATGCAGTGCTTG ATGGTGGAATTCCAT   398
15 Omt r    TGCTGCTCTCTGTCT CATGAATCAAGATAA AGTCCTTATGGAGA GCTGGTACCACTTGA AAGATGCAGTGCTTG ATGGTGGAATTCCAT   398
16 chryso   TGCTGCTCTCTGTCT CATGAATCAAGATAA AGTCCTTATGGAGA GCTGGTACCACTTGA AAGATGCAGTGCTTG ATGGTGGAATTCCAT   400

17 chryso   TGCTGCTCTCTGTCT CATGAATCAAGATAA AGTCCTTATGGAGA GCTGGTACCACTTGA AAGATGCAGTGCTTG ATGGTGGAATTCCAT   400
18 21       TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCATGGCTCG ATGGTGGCATCCCAT   419
19 6        TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   298
20 8        TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   326

21 16       TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   326
22 9        TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   383
23 5        TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   387
24 11       TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   370

25 7        TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   368
26 18       TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATTGAGT GCTGGTACCACATGA AAGACGCATGGCTCG ATGGTGGCATCCCAT   415
27 2        TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   409
28 4        TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   418

29 15       TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   418
30 1        TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   418
31 12       TGCTGCTCTTTGTGT CGCTGCTCAAGATAA AGTCTTAATGGAGT GCTGGTACCACATGA AAGACGCAGTGCTCG ATGGTGGCATCCCAT   410
```

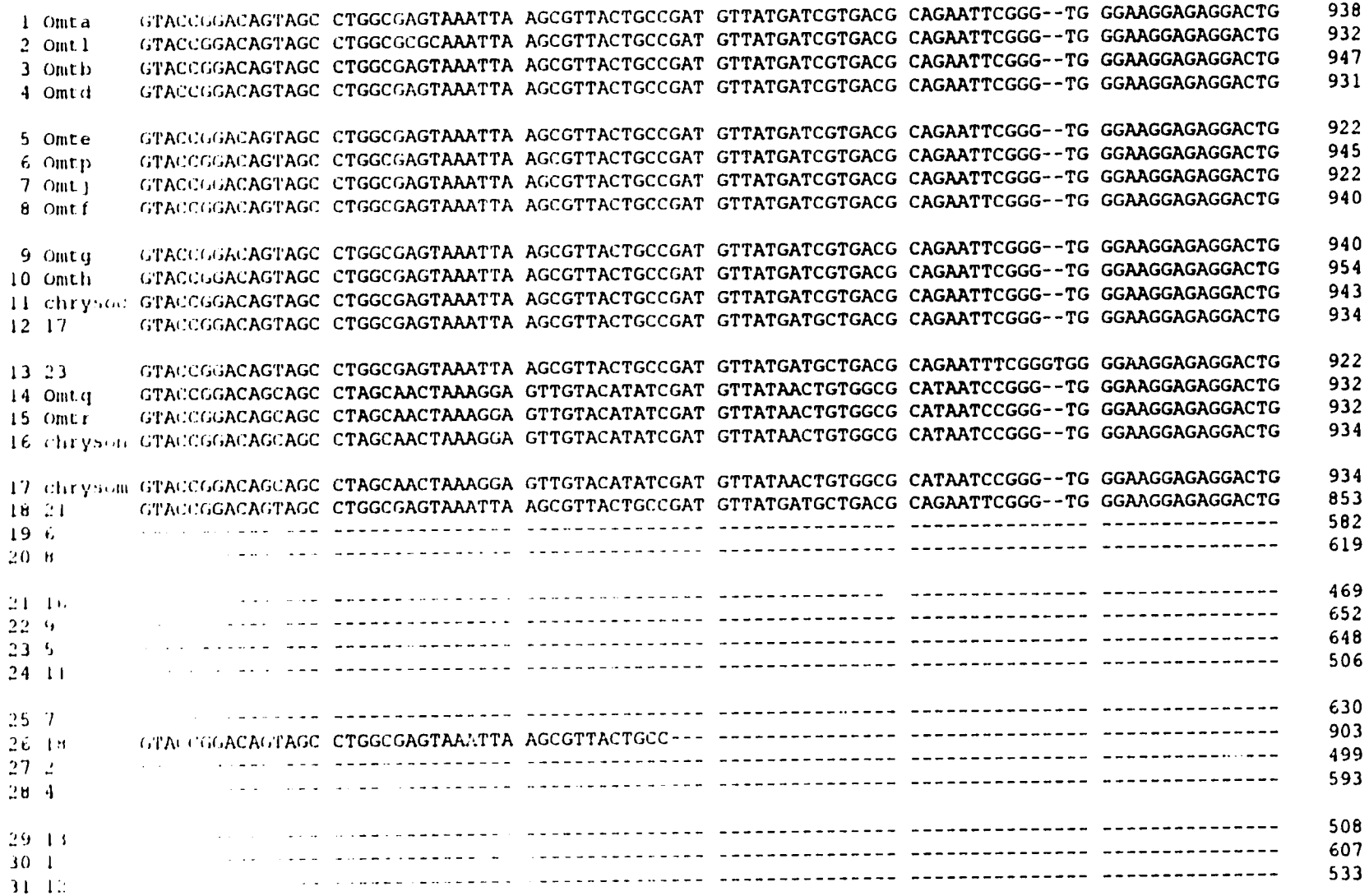
```
 1 Omta     TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   491
 2 Omtl     TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   485
 3 Omtb     TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   500
 4 Omtd     TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   484

 5 Omte     TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   475
 6 Omtp     TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   498
 7 Omtj     TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   475
 8 Omtf     TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   493

 9 Omtg     TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   493
10 Omth     TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   507
11 chrysoc  TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   496
12 17       TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   485

13 23       TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   473
14 Omtq     TCAACAAAGCCTAC GGAATGTCTTCTTT CGAGTACCACGGCA CTGATCCAAGATTCA ACAAAGTTTTCAATA GGGGAATGTCGGATC   485
15 Omtr     TCAACAAAGCCTAC GGAATGTCTTCTTT CGAGTACCACGGCA CTGATCCAAGATTCA ACAAAGTTTTCAATA GGGGAATGTCGGATC   485
16 chrysom  TCAACAAAGCCTAC GGAATGTCTTCTTT CGAGTACCACGGCA CTGATCCAAGATTCA ACAAAGTTTTCAATA GGGGAATGTCGGATC   487

17 chrysom  TCAACAAAGCCTAC GGAATGTCTTCTTT CGAGTACCACGGCA CTGATCCAAGATTCA ACAAAGTTTTCAATA GGGGAATGTCGGATC   487
18 21       TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   506
19 6        TCAACAAAACCTAT GGAATGCCCGTCTT CGACTACCTTGCCA TAGATTTAAGATCCA ACAAACTTTTCAACA AGGCCATGTCGGATC   385
20 8        TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   413

21 16       TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   413
22 9        TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   470
23 5        TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   474
24 11       TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATN   457

25 7        TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   455
26 18       TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   502
27 2        TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA NCAAAGTTTNNAACA AGGGCATGTCGGATT   496
28 4        TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATATAGGATCCA ACAAAGTGTTCATCA ANGGCATGTCGGATT   505

29 14       TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   505
30 1        TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   505
31 12       TCAACAAAGCCTAT GGAATGCCCATCTT CGACTACTTTGCCA AAGATTTAGGATCCA ACAAAGTTTTCAACA AGGGCATGTCGGATT   497
```

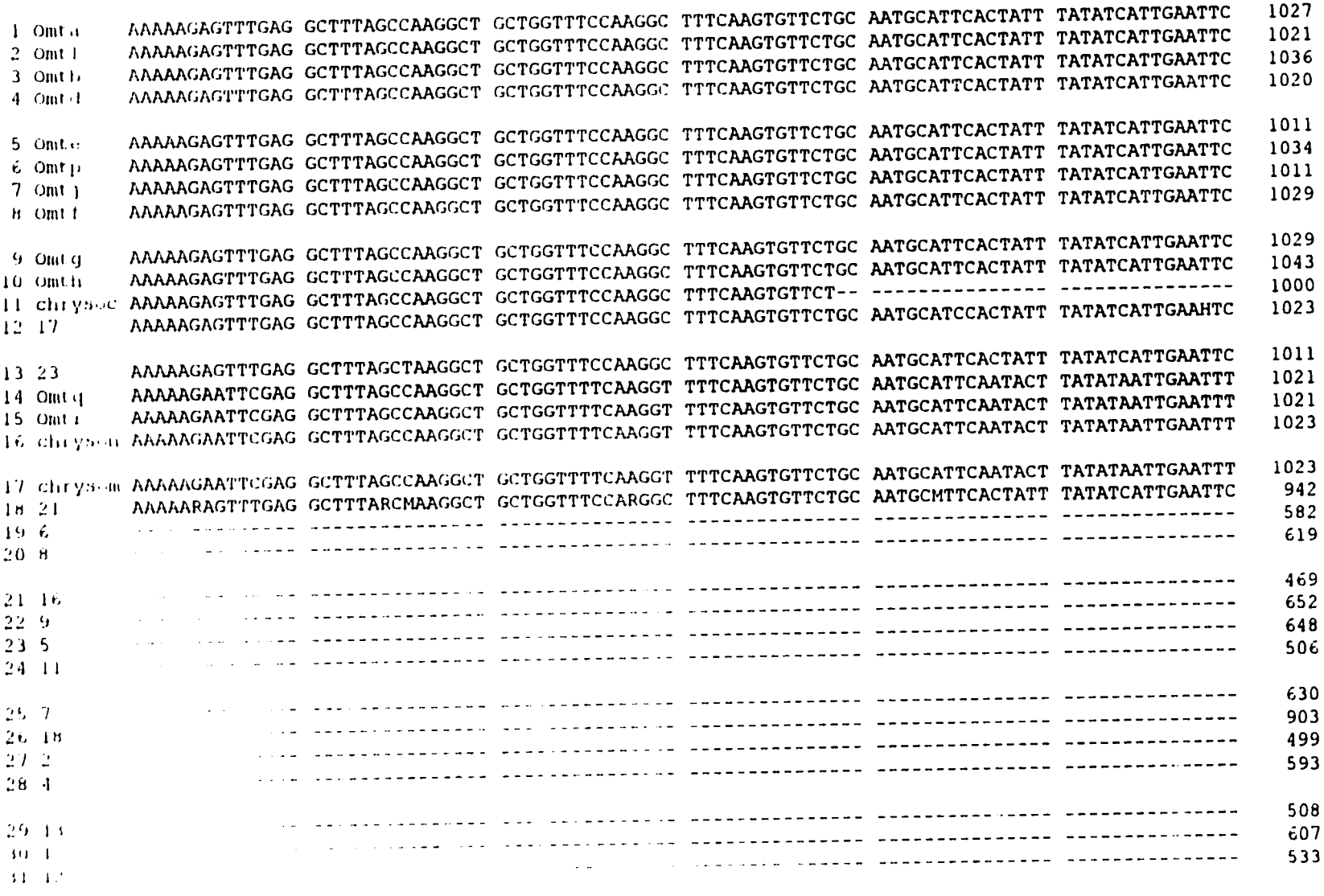

1 Omta TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 580
2 Omtl TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 574
3 Omtb TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 589
4 Omtd TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 573
5 Omte TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 564
6 Omtp TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 587
7 Omtj TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 564
8 Omtf TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 582
9 Omtg TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 582
10 Omth TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 596
11 chrysoc TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 585
12 17 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 574
13 23 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 562
14 Omtq ATTCTACGATTATCA TGAAGAAAGTTTTCC AAACTTATCAAGGGT TCCAAGGACTAACAT CTTTGGTTGATGTC GGTGGTGGTACCGGA 574
15 Omtr ATTCTACGATTATCA TGAAGAAAGTTTTCC AAGCTTATCAAGGGT TCCAAGGACTAACAT CTTTGGTTGATGTC GGTGGTGGTACCGGA 574
16 chryson ATTCTACGATTATCA TGAAGAAAGTTTTCC AAACTTATCAAGGGT TCCAAGGACTAACAT CTTTGGTTGATGTC GGTGGTGGTACCGGA 576
17 chrysom ATTCTACGATTATCA TGAAGAAAGTTTTCC AAGCTTATCAAGGGT TCCAAGGACTAACAT CTTTGGTCGATGTC GGTGGTGGTACCGGA 576
18 21 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 595
19 6 ATTCCTCGATTATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAAGAT CGGTGGTTGATGTT GGTGGTGGTACCGGA 474
20 8 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 502
21 16 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTA---- ---------- ---------- 469
22 9 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 559
23 5 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT NCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 563
24 11 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT CCAA---------- ---------- ---------- 506
25 7 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 544
26 18 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 591
27 2 TTT ------ -- ---------- -- ---------- ---------- ---------- ---------- 499
28 4 TTTCCTCGATGATCA TCAAGANAATTCTGG AAACTTACAAAGGAT NCCAANNACTAACAT CTTTGGTCGATGTT GGTNGGTGGTACCC- 593
29 13 TTT ------ -- ---------- ---------- ---------- ---------- ---------- ---------- 508
30 1 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACTTACAAAGGAT TCCAAGGACTAACAT CTTTGGTCGATGTT GGTGGTGGTACCGGA 594
31 12 TTTCCTCGATGATCA TCAAGAAAATTCTGG AAACCC--------- ---------- ---------- ---------- 533

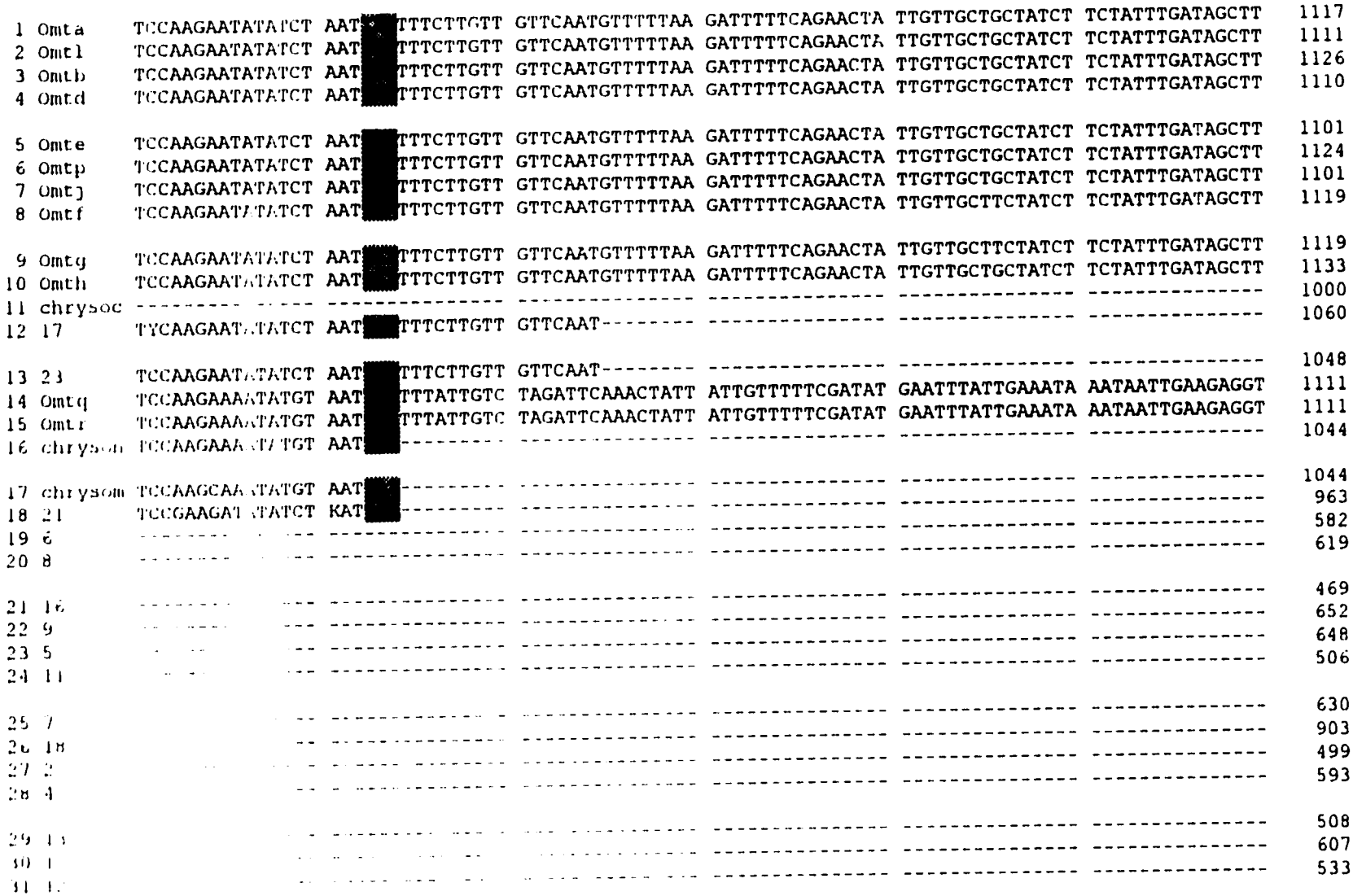

1 Omta GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 670
2 Omtl GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 664
3 Omtb GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 679
4 Omtd GCAACACTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 663
5 Omte GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 654
6 Omtp GCAACACTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 677
7 Omtj GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 654
8 Omtf GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 672
9 Omtg GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 672
10 Omth GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCGACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 686
11 chryson GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 675
12 17 GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 664
13 23 GCAACACTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 652
14 Omtq GCAACGCTTACCATG ATCCTCTCCAAGTAC CCTACAATTCGATGC ATCAACTTTGATTTG CCTCATGTCATTGAG GATGCCCCTGAATAT 664
15 Omtr GCAACGCTTACCATG ATCCTCTCCAAGTAC CCTACAATTCGATGC ATCAACTTTGATTTG CCTCATGTCATTGAG GATGCCCCTGAATAT 664
16 chryson GCAACGCTTACCATG ATCCTCTCCAAGTAC CCTACAATTCGATGC ATCAACTTTGATTTG CCTCATGTCATTGAG GATGCCCCTGAATAT 666
17 chryson GCAACGCTTACCATG ATCCTCTCCAAGTAC CCTACAATTCGATGC ATCAACTTTGATTTG CCTCATGTCATTGAG GATGCCCCTGAATAT 666
18 21 GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCGACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 585
19 6 GCAACGCTTACCATG ATCCTCTCCAAGTAC CCGACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCCCCTGAATAT 564
20 8 GCAACACTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGANTAT 592
21 16 ---- ---------- ---------- ---------- ---------- ---------- ---------- 469
22 9 GCAACACTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTNCTGAATAT 649
23 5 GCAACACTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGANGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTNCTT----- 648
24 11 --- ---------- ---------- ---------- ---------- ---------- ---------- 506
25 7 GCAACACTTACCAAG ATCCTCTCCAAGTAC CCTACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTNCTGG---- 630
26 18 GCAACGCTTACCAAG ATCCTCTCCAAGTAC CCGACAATTCGAGGC ATCAACTTTGATTTG CCTCATGTCATACAG GATGCTCCTGAATAT 681
27 2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- 499
28 4 ---------- ---------- ---------- ---------- ---------- ---------- ---------- 593
29 13 ---------- ---------- ---------- ---------- ---------- ---------- ---------- 508
30 1 GCAACGCTTACCC-- ---------- ---------- ---------- ---------- ---------- ---------- 607
31 12 ---------- ---------- ---------- ---------- ---------- ---------- ---------- 533

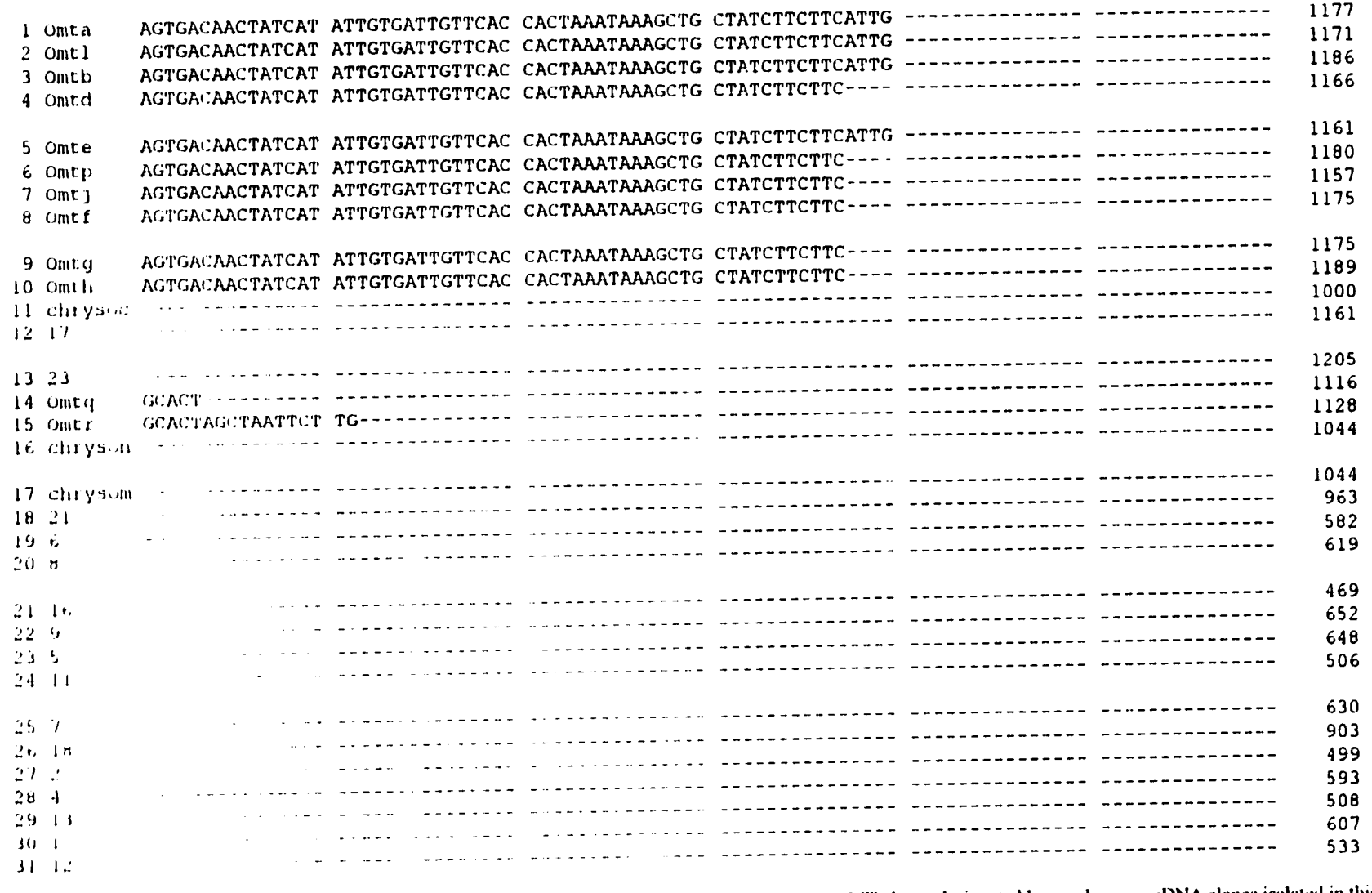

1 Omta CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 760
2 Omt1 CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 754
3 Omtb CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 769
4 Omtd CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 753
5 Omte CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 744
6 Omtp CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 767
7 Omtj CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 744
8 Omtf CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 762
9 Omtq CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 762
10 Omth CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 776
11 chrysoc CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 765
12 17 CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 754
13 23 CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 742
14 Omtq CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAGC 754
15 Omtr CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAGC 754
16 chryson CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAGC 756
17 chrysom CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAGC 756
18 21 CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 675
19 6 CCTGGTATGCATATC GAG---------- ---------- ---------- ---------- ---------- 582
20 8 CCCGGCATTGAGCAT GTTGGTGGAGAC--- ---------- ---------- ---------- ---------- 619
21 16 ---------- ---------- ---------- ---------- ---------- ---------- 469
22 9 CCC ---------- ---------- ---------- ---------- ---------- ---------- 652
23 5 --- ---------- ---------- ---------- ---------- ---------- ---------- 648
24 11 ---------- ---------- ---------- ---------- ---------- ---------- 506
25 7 ---------- ---------- ---------- ---------- ---------- ---------- 630
26 18 CCCGGCATTGAGCAT GTTGGTGGAGACATG TTTGTTAGTGTCCCC AAAGGAGATGCCATT TTCATGAAGTGGATA TGTCATGATTGGAAC 771
27 2 ---------- ---------- ---------- ---------- ---------- ---------- 499
28 4 ---------- ---------- ---------- ---------- ---------- ---------- 593
29 13 ---------- ---------- ---------- ---------- ---------- ---------- 508
30 1 ---------- ---------- ---------- ---------- ---------- ---------- 607
31 12 ---------- ---------- ---------- ---------- ---------- ---------- 533


1 Omta GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 850
2 Omtl GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 844
3 Omtb GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 859
4 Omtd GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 843
5 Omte GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 834
6 Omtp GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 857
7 Omtj GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 834
8 Omtf GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 852
9 Omtg GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 852
10 Omth GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 866
11 chrysoc GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 855
12 17 GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 844
13 23 GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 832
14 Omtq GACGAGCACTGCTTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTCTTGCTGAA TGCATACTCCCGGAG 844
15 Omtr GACGAGCACTGCTTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTCTTGCTGAA TGCATACTCCCGGAG 844
16 chrysom GACGAGCACTGCTTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTCTTGCTGAA TGCATACTCCCGGAG 846
17 chrysom GACGAGCACTGCTTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTCTTGCTGAA TGCATACTCCCGGAG 846
18 21 GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 765
19 6 582
20 8 619
21 16 469
22 9 652
23 5 648
24 11 506
25 7 630
26 18 GAAGAACAGTGCCTG AAATTATTGAAAAAT TGTTATGATGCCCTT CCGAATAATGGGAAG GTGATTGTTGCTGAA TACATACTCCCCGTG 861
27 2 499
28 4 593
29 13 508
30 1 607
31 12 533


1 Omta GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATCGTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 938
2 Omtl GTACCGGACAGTAGC CTGGCGCGCAAATTA AGCGTTACTGCCGAT GTTATGATCGTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 932
3 Omtb GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATCGTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 947
4 Omtd GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATCGTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 931
5 Omte GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATCGTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 922
6 Omtp GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATCGTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 945
7 Omtj GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATCGTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 922
8 Omtf GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATCGTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 940
9 Omtg GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATCGTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 940
10 Omth GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATCGTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 954
11 chrysoc GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATCGTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 943
12 17 GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATGCTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 934
13 23 GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATGCTGACG CAGAATTTCGGGTGG GGAAGGAGAGGACTG 922
14 Omtq GTACCGGACAGCAGC CTAGCAACTAAAGGA GTTGTACATATCGAT GTTATAACTGTGGCG CATAATCCGGG--TG GGAAGGAGAGGACTG 932
15 Omtr GTACCGGACAGCAGC CTAGCAACTAAAGGA GTTGTACATATCGAT GTTATAACTGTGGCG CATAATCCGGG--TG GGAAGGAGAGGACTG 932
16 chryson GTACCGGACAGCAGC CTAGCAACTAAAGGA GTTGTACATATCGAT GTTATAACTGTGGCG CATAATCCGGG--TG GGAAGGAGAGGACTG 934
17 chrysom GTACCGGACAGCAGC CTAGCAACTAAAGGA GTTGTACATATCGAT GTTATAACTGTGGCG CATAATCCGGG--TG GGAAGGAGAGGACTG 934
18 21 GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCCGAT GTTATGATGCTGACG CAGAATTCGGG--TG GGAAGGAGAGGACTG 853
19 6 582
20 8 619
21 16 469
22 9 652
23 5 648
24 11 506
25 7 630
26 18 GTACCGGACAGTAGC CTGGCGAGTAAATTA AGCGTTACTGCC--- 903
27 2 499
28 4 593
29 13 508
30 1 607
31 12 533


1 Omt a AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATTCACTATT TATATCATTGAATTC 1027
2 Omt l AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATTCACTATT TATATCATTGAATTC 1021
3 Omt b AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATTCACTATT TATATCATTGAATTC 1036
4 Omt d AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATTCACTATT TATATCATTGAATTC 1020
5 Omt e AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATTCACTATT TATATCATTGAATTC 1011
6 Omt p AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATTCACTATT TATATCATTGAATTC 1034
7 Omt j AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATTCACTATT TATATCATTGAATTC 1011
8 Omt f AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATTCACTATT TATATCATTGAATTC 1029
9 Omt g AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATTCACTATT TATATCATTGAATTC 1029
10 Omt h AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATTCACTATT TATATCATTGAATTC 1043
11 chrysoc AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCT-- ---------- ---------- 1000
12 17 AAAAAGAGTTTGAG GCTTTAGCCAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATCCACTATT TATATCATTGAAHTC 1023
13 23 AAAAAGAGTTTGAG GCTTTAGCTAAGGCT GCTGGTTTCCAAGGC TTTCAAGTGTTCTGC AATGCATTCACTATT TATATCATTGAATTC 1011
14 Omt q AAAAAGAATTCGAG GCTTTAGCCAAGGCT GCTGGTTTTCAAGGT TTTCAAGTGTTCTGC AATGCATTCAATACT TATATAATTGAATTT 1021
15 Omt i AAAAAGAATTCGAG GCTTTAGCCAAGGCT GCTGGTTTTCAAGGT TTTCAAGTGTTCTGC AATGCATTCAATACT TATATAATTGAATTT 1021
16 chrysm AAAAAGAATTCGAG GCTTTAGCCAAGGCT GCTGGTTTTCAAGGT TTTCAAGTGTTCTGC AATGCATTCAATACT TATATAATTGAATTT 1023
17 chrysm AAAAAGAATTCGAG GCTTTAGCCAAGGCT GCTGGTTTTCAAGGT TTTCAAGTGTTCTGC AATGCATTCAATACT TATATAATTGAATTT 1023
18 21 AAAAARAGTTTGAG GCTTTARCMAAGGCT GCTGGTTTCCARGGC TTTCAAGTGTTCTGC AATGCMTTCACTATT TATATCATTGAATTC 942
19 6 582
20 8 619
21 16 469
22 9 652
23 5 648
24 11 506
25 7 630
26 18 903
27 2 499
28 4 593
29 15 508
30 1 607
31 17 533


1 Omta TCCAAGAATATATCT AAT TTTCTTGTT GTTCAATGTTTTTAA GATTTTTCAGAACTA TTGTTGCTGCTATCT TCTATTTGATAGCTT 1117
2 Omtl TCCAAGAATATATCT AAT TTTCTTGTT GTTCAATGTTTTTAA GATTTTTCAGAACTA TTGTTGCTGCTATCT TCTATTTGATAGCTT 1111
3 Omtb TCCAAGAATATATCT AAT TTTCTTGTT GTTCAATGTTTTTAA GATTTTTCAGAACTA TTGTTGCTGCTATCT TCTATTTGATAGCTT 1126
4 Omtd TCCAAGAATATATCT AAT TTTCTTGTT GTTCAATGTTTTTAA GATTTTTCAGAACTA TTGTTGCTGCTATCT TCTATTTGATAGCTT 1110
5 Omte TCCAAGAATATATCT AAT TTTCTTGTT GTTCAATGTTTTTAA GATTTTTCAGAACTA TTGTTGCTGCTATCT TCTATTTGATAGCTT 1101
6 Omtp TCCAAGAATATATCT AAT TTTCTTGTT GTTCAATGTTTTTAA GATTTTTCAGAACTA TTGTTGCTGCTATCT TCTATTTGATAGCTT 1124
7 Omtj TCCAAGAATATATCT AAT TTTCTTGTT GTTCAATGTTTTTAA GATTTTTCAGAACTA TTGTTGCTGCTATCT TCTATTTGATAGCTT 1101
8 Omtf TCCAAGAATATATCT AAT TTTCTTGTT GTTCAATGTTTTTAA GATTTTTCAGAACTA TTGTTGCTTCTATCT TCTATTTGATAGCTT 1119
9 Omtg TCCAAGAATATATCT AAT TTTCTTGTT GTTCAATGTTTTTAA GATTTTTCAGAACTA TTGTTGCTTCTATCT TCTATTTGATAGCTT 1119
10 Omth TCCAAGAATATATCT AAT TTTCTTGTT GTTCAATGTTTTTAA GATTTTTCAGAACTA TTGTTGCTGCTATCT TCTATTTGATAGCTT 1133
11 chrysoc --------- ----- --------------- --------------- --------------- --------------- --------------- 1000
12 17 TYCAAGAATATATCT AAT TTTCTTGTT GTTCAAT-------- --------------- --------------- --------------- 1060
13 23 TCCAAGAATATATCT AAT TTTCTTGTT GTTCAAT-------- --------------- --------------- --------------- 1048
14 Omtq TCCAAGAAAATATGT AAT TTTATTGTC TAGATTCAAACTATT ATTGTTTTTCGATAT GAATTTATTGAAATA AATAATTGAAGAGGT 1111
15 Omtr TCCAAGAAAATATGT AAT TTTATTGTC TAGATTCAAACTATT ATTGTTTTTCGATAT GAATTTATTGAAATA AATAATTGAAGAGGT 1111
16 chrysom TCCAAGAAA AT TGT AAT --------- --------------- --------------- --------------- --------------- 1044
17 chrysom TCCAAGCAA ATATGT AAT --------- --------------- --------------- --------------- --------------- 1044
18 21 TCCGAAGAT ATATCT KAT --------- --------------- --------------- --------------- --------------- 963
19 6 --------- - --- --------------- --------------- --------------- --------------- --------------- 582
20 8 --------- - --- --------------- --------------- --------------- --------------- --------------- 619
21 16 --------- --- --------------- --------------- --------------- --------------- --------------- 469
22 9 --------- --- --------------- --------------- --------------- --------------- --------------- 652
23 5 --------- --- --------------- --------------- --------------- --------------- --------------- 648
24 11 --------- --- --------------- --------------- --------------- --------------- --------------- 506
25 7 --------- -- --------------- --------------- --------------- --------------- --------------- 630
26 18 --------- -- --------------- --------------- --------------- --------------- --------------- 903
27 2 --------- -- --------------- --------------- --------------- --------------- --------------- 499
28 4 --------- -- --------------- --------------- --------------- --------------- --------------- 593
29 15 --------- - --------------- --------------- --------------- --------------- --------------- 508
30 1 --------- -- --------------- --------------- --------------- --------------- --------------- 607
31 13 --------- -- --------------- --------------- --------------- --------------- --------------- 533



Appendix - Figure 1 Nucleotide base alignment of different OMT clones isolated from C. americanum. OMT clones designated by numbers are cDNA clones isolated in this work. cDNA clones from previous work (Gauthier, unpublished) are designated by the letters. OMT's a, l, b, d, e, p, j, f, g, h and chryso c are 3'5' flavonoid OMT cDNA clones. OMT's q,r, chryso n and chryso m are caffeic acid OMT cDNA clones. All other cDNA clones 17, 23, 21, 6, 8, 16, 9, 5, 11, 7, 18, 2, 4, 13, 1, 12 are 3'5' flavonoid OMT cDNA clones isolated in this work. Start codon (ATG) is highlighted for each cDNA clone. Stop codon (TAA) is highlighted for each cDNA clone.