MOLECULAR CLONING AND CHARACTERIZATION OF CONIFERYL ALCOHOL ACYLTRANSFERASE (CFAT) FROM *OCIMUM SP.*

Thesis submitted towards the partial fulfillment of

BS-MS Dual Degree Program

Indian Institute of Science Education and Research

By

MONTU PATAR

Under the guidance of

Dr. ASHOK P. GIRI

Sr. Scientist, Biochemical Sciences Division

CSIR - National Chemical Laboratory, Pune

CERTIFICATE

This is to certify that this dissertation entitles "**Molecular Cloning and Characterization of Coniferyl Alcohol Acyltransferase (CFAT) from *Ocimum Sp.***" towards the partial fulfillment of the BS-MS Dual Degree program at the Indian Institute of Science Education and Research, Pune represents original research carried out by "**Montu Patar** at **CSIR-National Chemical Laboratory, Pune**" under the supervision of "**Dr. Ashok P. Giri**, Sr. Scientist, Biochemical Sciences Division, NCL" during the academic year 2013-2014.

Dr. Ashok P. Giri

Senior Scientist, Biochemical Sciences Division

CSIR-National Chemical Laboratory

DECLARATION

I hereby declare that the matter embodied in the report entitled "**Molecular Cloning and Characterization of Coniferyl Alcohol Acyltransferase (CFAT) from *Ocimum Sp.***" are the results of the investigations carried out by me at the Division of Biochemical Sciences, CSIR- National Chemical Laboratory, Pune under the supervision of **Dr. Ashok P. Giri** and the same has not been submitted else for any other degree.

Montu Patar

BS-MS Dual Degree Program

IISER Pune

Acknowledgement

I would like to thank Dr. Ashok P. Giri for giving me this wonderful opportunity to work on this project. I am ever grateful to him for his valuable guidance through every step of this project.

I would also extend my deepest gratitude to Dr. H. V. Thulasiram for providing me a chance to join his lab group. I am sincerely thankful to him for his continuous assistance and encouragement throughout the course of my project.

I would like to thank CSIR-National Chemical Laboratory for providing me state of art amenities to carry out my project successfully.

I wish to express my sincere thanks to Atul (a^2) for being an excellent guide throughout my project. I am especially thankful to him for his patience with my ignorance and constant encouragement which served as the driving force behind all the progress of this project. I also thank Ramesha H.J. and Supriya for sharing their ideas and experience and providing me with beneficial guidance at each step.

My gratitude to my labmates, Prabhakar, Krithika, Devdutta, Shiva, Avinash, Dipesh, Rani, and Uttara for providing me with their valuable advice and help, every single day at lab. I would also like to thank Yashpal and Vijay for their help (in lab) as well as for accompanying me to the cafeteria.

I can never thank enough (the indispensable part) my beloved friends, Upendra, Kush, Abhishek, Dhanashree and Ritika who made my life at NCL as well as IISER so memorable. I am also grateful to all my 09' batchmates who provided me constant gossips and fun that helped me relax during my stay at IISER.

Finally, I would take this opportunity to thank my parents and uncle for their love and blessings at every phase of my life. I am grateful from the deepest of my heart to my elder brother for his immense support and selfless love.

Abstract

The genus *Ocimum* serves as a large pool of metabolites including isoprenoids and phenylpropanoids. Phenylpropanoids, such as eugenol, isoeugenol, eugenol methylether, chavicol, anol etc., are products of the diverted pathway from the lignin biosynthesis. The phenylpropanoid biosynthesis use acetylated monolignols such as coniferyl and coumaryl acetate as their substrate. These acetylated monolignols in turn are synthesized by acyltransferases which are solely responsible in diverting the metabolic flux away from lignin biosynthesis towards phenylpropanoid biosynthesis. This study aims at studying a similar enzyme, coniferyl alcohol acyltransferase (CFAT) which catalyzes the synthesis of coniferyl acetate by using coniferyl alcohol as a substrate. Coniferyl acetate in turn serves as the unique substrate used by eugenol synthase (EGS) for eugenol biosynthesis. This study aims at developing an intensive understanding of the role of CFAT in regulating the levels of phenylpropanoids in *Ocimum*. Transcriptome data of *Ocimum* was analyzed to search for putative CFAT contigs. These contigs were analyzed and out of 3 selected sequences, contig_13604 was filtered out as the most promising candidate gene. Further, the contig was cloned and the full length gene was obtained using RACE PCR. Real time and semi-quantitative PCR results added to the understanding of expression levels of this gene across leaf (young and mature) and stem tissues of *Ocimum*. The work proceeded towards recombinant protein expression with enzyme activity assay and chemical characterization of enzyme underway

Table of Contents

List of figures

List of tables

Chapter 1

Introduction

1.1 *Ocimum* and its importance

1.1.1 *Ocimum* morphology

Ocimum genus belongs to the Lamiaceae/Labiatae family. The Lamiaceae family contains more than 200 genera and around 3200 species (Sciences et al. 2013). They are widespread across the temperate and tropical regions. Plants belonging to the Lamiaceae family possess a square stem with opposite and decussate leaves and many gland dots (Nahak et al. 2011). *Ocimum* essentially belongs to the category of angiosperms (flowering plants) with flowers that are morphologically zygomorphic and have two lips. These plants possess upright stems and grow up to a height of 20-80 cm.

1.1.2 Medicinal significance of *Ocimum*

Ocimum is considered as an important economic and medicinal herb. The genus *Ocimum* is well known for its strong aromatic characteristic due to essential oils which comprises of monoterpenes, sesquiterpenes, and phenylpropanoids. They produce these chemicals in various parts like leaves, seeds, flowers, stem and roots. The parts are known to possess therapeutic potentials and have been used as expectorant, analgesic, anticancer, antiasthmatic, antiemetic, diaphoretic, antidiabetic, antifertility, hepatoprotective, hypotensive, hypolipidmic and antistress agents (Prakash and Gupta 2005). *Ocimum* has been studied for its ability to treat fever, bronchitis, arthritis, convulsions etc. (Prakash and Gupta 2005). Extracted oils are shown to possess antioxidant, insecticidal and fungistatic properties (Deshpande and Tipnis, 1977; Chogo and Crank, 1981; Reuveni et al., 1984). In addition, essential oils extracted from *Ocimum* plants are popularly used as food preservatives and to flavor various food

products. Few of the species of *Ocimum* that are widely exploited for their medicinal properties are: *Ocimum sanctum. L, Ocimum gratissium, Ocimum canum, Ocimum basilicum* (Figure 1, 2)*, Ocimum kilimandscharicum, Ocimum ammericanum, Ocimum camphora* and *Ocimum micranthum*.

Figure 1 *Ocimum* basilicum (Sweet Basil)

Figure 2 Inflorescence - *Ocimum* basilicum (Sweet Basil)

1.2 Natural Products (Secondary Metabolites)

1.2.1 Secondary metabolites in plants

Plants produce a vast and diverse collection of organic compounds known as metabolites. A fraction of these metabolites are directly involved in the growth and development of the plant such as photosynthesis, protein or nucleic acid synthesis,

respiration and are broadly known as primary metabolites. In contrast, the fraction of metabolites that does not seem to participate directly in these processes, are classified as secondary metabolites. Secondary metabolites were initially considered to be dispensable to the immediate survival of the organism and hence, categorized as secondary metabolites. However, the interest of organic chemists and their extensive investigation of these compounds have resulted in understanding their importance. Few examples of primary metabolites like Phytosterols, acyl lipids, nucleotides, amino acids and organic acids play critical metabolic roles and are found in all plants. In contrast, secondary metabolites vary widely in their distribution across plant species (Rhodes, 1994). Thousands of secondary metabolites have been identified, characterized and their functions have being elucidated.

Extensive studies on structure and function of secondary metabolites have revealed the immense diversity in plant metabolites. Major groups of secondary metabolites such as alkaloids, terpenoids and phenylpropanoids have been acknowledged to influence ecological interaction between the plant and its environment. Biosynthetically, all terpenoids are derived from a five-carbon precursor isopentenyl diphospphate (IPP). The alkaloids contain one or more nitrogen atoms and are biosynthesized from amino acids. The remaining phenolic compounds are formed either by way of shikimic acid pathway or the malonate/acetate pathway.

1.2.2. Phenylpropanoids

Phenylpropanoids are a diverse group of compounds known to play a vital role in plant structural support, defense and survival (Vogt, 2010). A bulk of these plant phenolic compounds was assumed to have cell wall structural roles. In addition, they also have nonstructural constituents playing various roles as defending plants, determining certain features of woods and barks and contributing to flower color and flavors (Buchanan et al, 2004). Most plant phenolics are products of the carbon skeleton of phenylalanine and these phenols account for about 40% of organic carbon circulating in biosphere and are vital to the interaction of plants with their biotic and abiotic environment (Buchanan et al, 2004; Fraser and Chapple, 2011). Phenylalanine is an end product of the

shikimate pathway, which also gives rise to the aromatic amino acids tyrosine and tryptophan (Herrmann and Weaver, 1999; Tzin and Galili, 2010).

1.2.3. Phenylpropanoid biosynthetic pathway

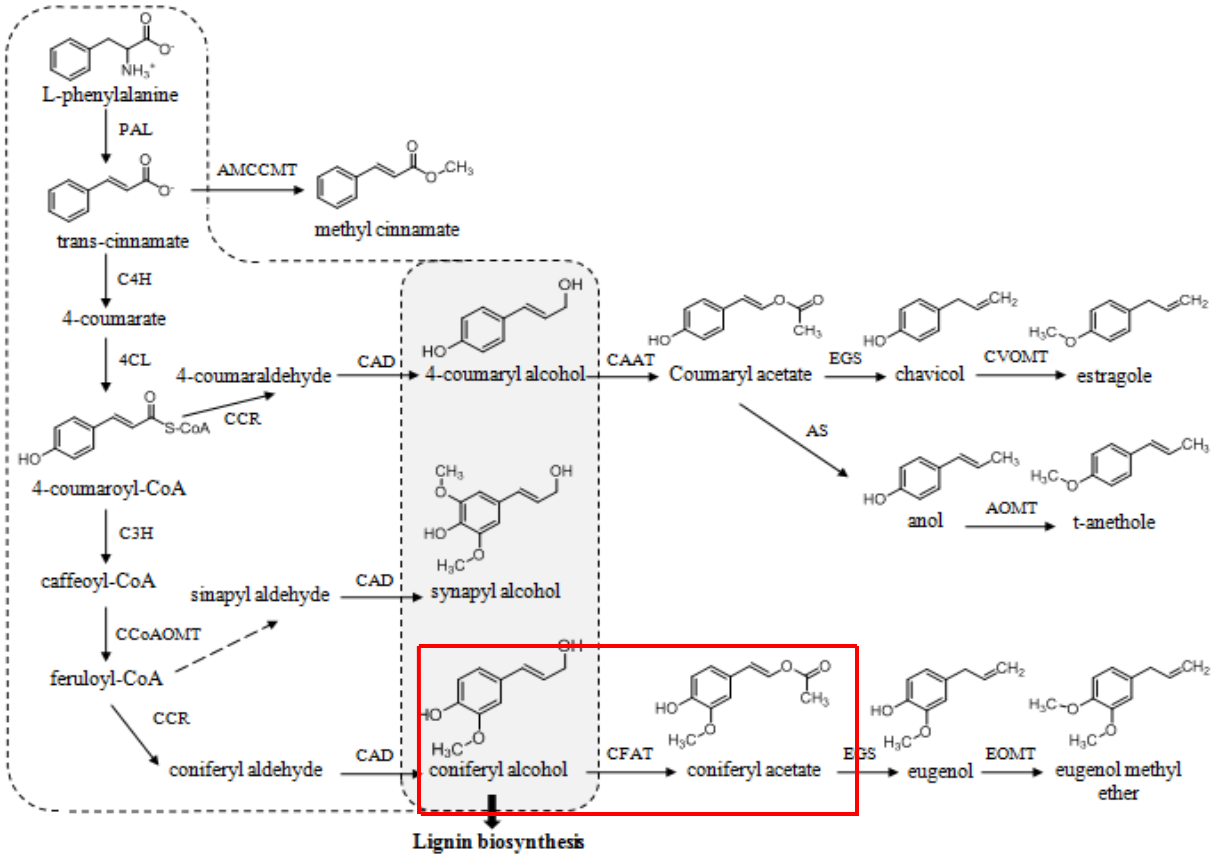
The phenylpropanoid biosynthesis pathway is imperative to plants because of its function in the production of hydroxycinnamyl alcohols (monolignols) (Boerjan, Ralph, and Baucher 2003). This pathway gives rise to flavonoids, coumarins, hydroxycinnamic acid conjugates and lignans (Vogt, 2010). It yields several commercially important metabolites like Eugenol, Eugenol methyl ether, Anethol, Enol etc. The shikimate pathway produces phenylalanine which acts as the initial step to the biosynthesis of phenylpropanoids. Phenylalanine ammonia lyase (PAL) catalyzes the oxidative deamination of phenylalanine to *trans*-cinnamate and ammonia (Figure 3). This acts as the first step in directing the carbon flow from shikimate pathway towards the various branches of phenylpropanoid metabolism(Vogt 2010). In the next step, cinnamate 4-hydroxylase (C4H) catalyzes the regiospecific hydroxylation at the *para*-position of *trans*-cinnamate converting it to 4-coumarate (also known as *p*-coumarate) (Figure 3). C4H is an oxygen requiring, NADPH-dependent, cytochrome P450 enzyme. Following this the 4-coumarate:CoA ligase (4CL) then catalyzes the ATP based conversion of the CoA thioester 4-coumaryl CoA (also known as *p*-coumaryl CoA) (Figure 3) (Fraser and Chapple 2011). The significance of 4CL could be appreciated in the light of the roles served by *p*-coumaryl CoA. The activated thioester *p*-coumaryl CoA serves at a substrate for various other secondary metabolites in addition to its role in phenylpropanoid biosynthesis(Vogt 2010). For example, chalcone synthase (CHS) diverts the pathway towards the production of flavones from this branching point. Similarly, biosynthesis of proanthocyanidins, tannins and flavonoids and also other pathways generating phenylpropenes such as coumarins, lignins and lignans initiate using *p*-coumaryl CoA as a substrate.

The phenylpropanoid pathway diverges immediately after the synthesis of *p*-coumaryl CoA. One leg of the pathway is catalyzed by two sequential NADPH-dependent reductases to yield a monolignol in just two steps. In the first step, cinnamoyls-CoA reductase (CCR) catalyzes the conversion of *p*-coumaryl CoA to *p*-coumaraldehyde

(Figure 3). Following this, the second enzyme cinnamoyl alcohol dehydrogenase (CAD) leads to the formation of monolignol, p-coumaryl alcohol (Figure 3). In the other leg of the pathway multiple enzymetic steps are involved as *p*-coumaryl CoA is converted to two other monolignols, coniferyl and sinapyl alcohols. In this route, *p*-coumaryl CoA is initially catalyzed by Hydroxycinnamoyl-coenzyme A shikimate:quinate hydroxycinnamoyl-transferase (HCT) forming *p*-coumaryl shikimate (Hoffmann et al. 2003). Next, p-coumaroyl shikimate 3' hydroxylase (C3'H) catalyzes the formation of caffeoyl shikimate which is then converted to caffeoyl CoA by HCT (Fraser and Chapple 2011). The caffeoyl CoA thus synthesized undergoes a methyl-transfer reaction catalyzed by Caffeoyl CoA 3-O-methyltransferase (CCoAOMT) to yield feruloyl CoA (Figure 3).

Similar to the fate of *p*-coumaryl CoA, feruloyl CoA too undergoes two sequential NADPH-dependent reductions to yield two monolignols. Firstly, Cinnamoyl-CoA reductase (CCR) catalyzes hydroxycinnamoylCoA thioesters to form hydroxycinnamaldehydes, i.e. coniferaldehyde and sinapaldehyde (Fraser and Chapple 2011). Secondly, cinnamoyl alcohol dehydrogenase (CAD) catalyzes the NADPH dependent reduction of hydroxycinnamaldehydes to the respective alcohols (coniferyl and sinapyl alcohol) as a final step in the synthesis of monolignols (Figure 3).

The monolignols, coniferyl and sinapyl alcohol are oxidized and further undergo polymerization process to form lignin (Fraser and Chapple 2011). However, the biosynthesis of phenylpropenes diverges from the lignin biosynthetic pathway at this branching point utilizing the monolignols, coniferyl and *p*-coumaryl alcohol.



Boerjan, Ralph, and Baucher 2003

Figure 3 Phenylpropanoid biosynthesis pathway in *Ocimum*.

PAL, phenylalanine lyase; C4H, trans-cinnamate 4-monoxygenase; 4CL , 4-coumarate CoA ligase; C3H, 4-coumarate-3-hydroxylase; CCoAOMT, caffeoyl-CoA O-methyl transferase; AMCCMT, S-adenosyl L-methionine cinnamic acid carboxy methyl transferase; CCR, cinnamoyl-CoA/coniferyl-CoA reductase; CAD, coumaryl/coniferyl alcohol dehydrogenase; CAAT, 4-coumaryl/coniferyl alcohol acetyl transferase; EGS, eugenol synthase; AS, anol synthase; CVOMT, chavicol O-methyl transferase; AOMT, anol O-methyl transferase, EOMT, eugenol O-methyl transferase

Many phenylpropenes are crucial for the plant because they provide characteristic aroma and flavor to the plants as well as to its fruits and flowers as floral attractants of pollinators. In addition, many phenylpropenes synthesized by plants are also used for defense against herbivores and pathogens due to their toxic effects (Koeduka et al. 2008). Phenylpropenes could be categorized into Allyl phenylpropenes (Chavicol, Methylchavicol, Eugenol, methyleugenol) and Iso phenylpropenes (Anol, Anethole, Isoeugenol, Methylisoeugenol) based on the difference in the position of double bond in the propene side chain (Koeduka et al. 2008). Volatile phenylpropenes like eugenol has been very well recognized for its varied array of aromatic as well as antibacterial and medicinal properties. Phenylpropenes also serve as the starting materials for many lignans which also possess antimicrobial and antioxidant properties. Given the historical and economical importance of phenylpropenes, it has acquired special attention of researchers since a long time. However, the enzymes and biochemical processes involved in the synthesis of eugenol, isoeugenol and related compounds are yet to be clarified.

1.3 Coniferyl alcohol acyltransferase in phenylpropanoid pathway

1.3.1 Role of coniferyl alcohol acyltransferase

A very significant enzyme involved in the complex phenylpropanoid pathway is coniferyl alcohol acetyltransferase (CFAT). CFAT is solely responsible in diverting the metabolic flux away from lignin biosynthesis by catalyzing the synthesis of coniferyl acetate from coniferyl alcohol (red outlined box in Figure 3). The ester of coniferyl alcohol (coniferyl acetate) is used as a substrate by eugenol synthase (EGS) to catalyze the formation of eugenol by reduction of the ester. Following the synthesis of eugenol, the enzyme eugenol *O*-methyltransferase (EOMT) further catalyzes the conversion of eugenol to eugenol methyl ether (Renu et al. 2014).

1.3.2 Identification and characterization of coniferyl alcohol acyltransferase from *Petunia x hybrida*

There has been only a single incident where a very similar enzyme named PhCFAT (coniferyl alcohol acyltransferase) has been isolated and characterized from *Petunia x hybrid* (1836 bp, 454 aa) (Dexter et al. 2007). The specificity of this enzyme towards

coniferyl alcohol as a substrate to produce isoeugenol was confirmed. A suppression of PhCFAT expression resulted in lower synthesis of isoeugenol and several other volatiles in Petunia. Also, it was shown that PhCFAT acetylates coniferyl alcohol and several other substrates like sinapyl, cinnamyl alcohol etc. in a pH dependent manner.

Based on significant sequence identity and conserved motifs CFAT is categorized as a member of the BAHD group of acyltransferases (Dexter et al. 2007). Acylation of oxygen and nitrogen containing substrates is the most routine type of chemical modification of secondary metabolites (D'Auria 2006). Plant BAHD acyltranferases utilize CoA thioesters to transfer acyl moiety to acceptor molecules. These transferases are significantly involved in the biosynthesis of various plant natural products such as lignin, alkaloids, phenolics, anthocyanins and other volatile esters (Grienenberger et al. 2009). The BAHD family of acyltransferases has gained its name based on the first letter of the first four characterized enzymes of this family (BEAT, AHCT, HCBT and DAT) (St-Pierre and De Luca, 2000). All identified BAHD acyltransferases are monomeric enzymes with molecular mass varying between 48 to 55kD and an average of 445 amino acids (D'Auria 2006). Enzymes of the BAHD family share two conserved motifs, HxxxD and DFGWG (St-Pierre and De Luca, 2000). The HxxxD domain is located near the center portion and is directly involved in the catalysis at the active site of the enzyme. The DFGWG motif is located near to carboxyl terminus and is shown to play a structural role in stabilizing the enzyme (Ma et al. 2005).

1.4. *Ocimum* as a model organism to study phenylpropanoid pathway

Some of the oils (secondary metabolites) that provide characteristic properties to *Ocimum* are eugenol, chavicol and their derivatives which belong to the group of phenylpropenes. Peltate glandular trichomes of *Ocimum Basilicum* was demonstrated to accumulate volatile oil constituents eugenol and methylchavicol (Gang et al. 2014). Thus, *Ocimum* could serve as an ideal model plant to develop an understanding of the biochemical process leading to differential expression levels of eugenol across various tissue types in the plant. Hence, elucidating the properties of the enzyme CFAT is vital as it is responsible for catalyzing the synthesis of the substrate for EGS to produce eugenol. This project aims at isolating and characterization of the enzyme CFAT from

Ocimum. Comparison of expression levels of eugenol and eugenol methyl ether with that of CFAT would help us demonstrate the role of CFAT in phenylpropanoid pathway.

List of *Ocimum* plants under study:

1. *Ocimum gratissium (Og)*

2. *Ocimum kilimandscharicum(Ok)*

3. *Ocimum tenuiflorum* L*. (Ot)*

4. *Ocimum basilicum* (*Oba*)

5. *Ocimum americamum* L*. (Oa)*

Further, two sub-types of *Ocimum tenuiflorum L.* named as OtI and Ot II, four sub-types of *Ocimum basilicum* named as Oba I, Oba II, Oba III, Oba IV and one unidentified *Ocimum r*. (*Or*) were included in study. Hence, ten taxa of plants in total were used in the extensive study to understand the expression levels of secondary metabolites across *Ocimum* plants.

Chapter 2

Materials and Methods

2.1 Plant Material from *Ocimum sp.*

Samples of young leaves and inflorescence of six species were collected from the field grown plants at the Botanical Garden of National Chemical Laboratory, Pune. The plant tissue were washed with sterile water, wiped and were packed into 50 mL falcon tubes. The tubes were flash frozen in liquid nitrogen and subsequently stored at -80°C until processed.

2.2 RNA Isolation and Gene Expression analysis

2.2.1 RNA isolation

Frozen tissue samples were crushed to powder in liquid nitrogen using mortar and pestle. For each sample, 100 mg of the grinded tissue was weighed in 2 mL pre chilled microcentrifuge tubes and stored at -80°C. RNA was isolated from the tissue using the Spectrum Plant Total RNA Kit (Sigma-Aldrich) following the manufacturer's protocol. All RNA samples were analyzed on 1% agarose gel to check for degradation before proceeding to DNAse treatment. The RNA samples were treated with DNaseI (Sigma-Aldrich) following the manufacturer's protocol. DNaseI treatment is imperative for removing any traces of genomic DNA present in RNA samples. Obtained RNA samples were essentially stored in -80°C until further processed.

Qualitative and quantitative analysis of RNA was done using Thermo-scientific NanoDrop2000c spectrophotometer. All samples were analyzed on 1% agarose gel.

Eight µL of 1:10 dilution of each RNA sample was mixed with 2 µL of RNA loading dye and run in 1% TAE buffer (prepared using DEPC treated water) at 100 V.

2.2.2 cDNA synthesis

An equivalent amount of RNA, i.e. 4 µg of samples across all species was used for reverse transcription reaction. Reverse transcription for cDNA synthesis was done using Superscript III-First-Strand-Synthesis system for RT-PCR (Invitrogen) following manufacturer's protocol.

Total RNA extracted from tissues was used to synthesize cDNA as mentioned above. The synthesized cDNA were checked for 18S gene amplification by PCR (Polymerase Chain Reaction) using 18S gene specific primers. Reaction mixtures of 110 uL was prepared using 50 µL of Jumpstart mastermix, 10 µL of forward and reverse primer each of 18S and final volume was made up with nuclease free water. Aliquots of 10 µL each in 200 µL pcr tubes were made for different PCR amplifications reactions using 10 different synthesized cDNA as template. One µL of cDNA was added to each aliquot respectively. A 10 µL aliquot with no added cDNA was used as a negative control. PCR reaction was carried out as follows– (1) initial denaturation is carried out at 95 °C for 5 minutes (2) subsequent denaturation was carried out at 95 °C for 30 seconds (3) annealing was done at 58 °C for 30 seconds (4) extension at 72 °C for 30 seconds (5) final extension at 72 °C for 7 minutes was carried out and stored at 4 °C. Steps (2) to (4) was repeated for 35 cycles.

All PCR samples were analyzed on a 1% agarose gels. 10 µL of each sample was mixed to 2 µL of loading dye and run in 1% TAE buffer at 100 V. 1kb DNA ladder from Invitrogen was used to run along-with the samples. Gel pictures were taken using Gel documentation system (Syngene) equipped with Gene Snap software for analysis.

2.2.3 Semi-quantitative PCR

PCR reaction mixture of 100 µL was prepared using 55 µL of Jumpstart mastermix, 10 µL of forward and reverse primer each (specific to 18S) and volume was made up by adding water. Aliquots of 10 µL each in 200 µL pcr tubes were made for different PCR amplifications reactions using 10 different synthesized cDNA as template. 1 µL of

vacuum concentrated cDNA (optimized amount) was added to each aliquot. A 10 µL aliquot with no added cDNA was used as a negative control. PCR reaction was carried out as follows– (1) initial denaturation at 95 °C for 5 minutes (2) subsequent denaturation at 95 °C for 30 seconds (3) annealing at 58 °C for 30 seconds (4) extension at 72 °C for 30 seconds (5) final extension at 72 °C for 7 minutes and stored at 4 °C. Steps (2) to (4) was repeated for 30 cycles for 18S. The 18S rRNA fragment used as an internal control was amplified using 18 specific primers. A similar reaction was performed using the primers specific to EGS, EOMT and Contig_13604 (specific to full length) to check expression levels of the gene across all species. The amplified PCR products were analyzed on 1% agarose gel.

2.2.4 Real Time PCR

Taq-man primers and probes for EGS, EOMT and Actin, and Taqman Fast Advanced master mix were obtained from Invitrogen. The reaction set up was put as per the manufacturer's instruction. In short, 10 µL reactions was setup using 5 µL of mastermix, 1 µL of cDNA and 0.5 µL of primer-probe mix. PCR was done in ABI 7500 Fast Real Time PCR machine on standard program mode. Standard curve method was followed for calculating Ct values of EGS and EOMT and was normalized with the Ct value of Actin for calculating relative expression levels.

2.3 Primer designing

2.3.1 Set I

In the first attempt, 3 pair of primers (Table 1) were designed based on the reported nucleotide sequence of coniferyl alcohol acyltransferase from *Petunia x hybrida* (*PhCFAT*) (GenBank: DQ767669.1).

Table 1 List of primers - set I

No.	Oligo Name	Forward primers sequence
1.	CFAT F1	5'-ATTGGCTGGAGAAGTGG-3'
2.	CFAT F2	5'-TATCGACAGCACGTATGTCT-3'
3.	CFAT RTF	5'-TTGGACAATGTTGCTACGA-3'

No.	Oligo Name	Reverse primer sequence
1.	CFAT R1	5’-TGGTCGTTGAGGAATCGT-3’
2.	CFAT R2	5’-ACACAATCCAATCACCATTCTT-3’
3.	CFAT RTR	5’-TAGACCAGAGGAGACAACG-3’

2.3.2 Set II

In the second attempt primers were designed (Figure 2, 3) based on the reported nucleotide sequence of *PhCFAT* but taking into consideration (i) codon bias and (ii) conserved motifs of BAHD group of acyltransferases.

Table 2 List of Forward primers - set II

No.	Oligo Name	Sequence
1.	CFAT F1A	5’-ATGGGAAACACAGATTTCCAT-3’
2.	CFAT F1B	5’-ATGGGAAACACTGACTTCCAT-3’
3.	CFAT F1C	5’-ATGGGAAACACAGATTTCCAC-3’
4.	CFAT F1D	5’-ATGGGAAACACAGACTTTCAC-3’
5.	CFAT F2A	5’-ATGCATCATGAGCACTGGTTGCCCATG-3’
6.	CFAT F2B	5’-ATGCATCATGAGCACTGGCTCCCAATG-3’
7.	CFAT PF1	5’-GCAAGTGTTGGACAGCTCAAGCAA-3’
8.	CFAT PF2	5’-TGATCATAGAGTAGCTGATGC-3’

Table 3 List of Reverse primers - set II

No.	Oligo Name	Sequence
1.	CFAT R1	5’-CATATACACAATCCAATCACCATT -3’
2.	CFAT R2	5’-TTCATCATTGCACTCATCTTTGCA-3’
3.	CFAT R3A	5’-GTTCCCGAAATAGTTTTTCAT-3’
4.	CFAT R3B	5’-GTTTCCGAAATAGTTTTTCAT-3’
5.	CFAT PR1	5’-TCTTGTTAGGGCTAGGCATTGGCA-3’
6.	CFAT PR2	5’-ACACCCCCAGCCAAAGTTTAC-3’

* Nucleotides marked in red denote the different combination of codons coding for same amino acid.

2.3.3 Set III

In the third attempt primers were designed (Table 5, 6) specific to the selected results obtained through analysis of transcriptome data of *Ocimum.* The sequence of *PhCFAT* was used as a query for tblastn against the database of *Ocimum* transcripts using Scoring Matrix Blosum62. The details of the sequences selected for synthesizing primers are mentioned in Table 4.

Table 4 List of selected sequences from transcriptome

No.	Name	Nucleotide length	Region of query covered in tblastn	Score (bits)
1.	Contig_13604	1631	31-433	147 (371)
2.	Locus_50463	561	185-355	135 (341)
3.	Locus_137834	201	384-448	100 (247)

Table 5 List of Forward primers - set III

No.	Oligo Name	Sequence
1.	L 50463_F1	5'-CCATCATTCCGCCGTTCCGTCTTC-3'
2	L 50463_F2	5'-TTGCTACGACATTTCTGTCGAC-3'
3	C13604_F1	5'-ATGAAGATCGTCGTGAAAGATTC-3'
4.	C13604_F2	5'-ATGCAGCACCACGCGGCCGAC-3'
5.	L137834_F1	5'-ATGGGGTGGGGTTTACCGGTGTTTG-3'

Table 6 List of Reverse primers - set III

No.	Oligo Name	Sequence
1.	L 50463_R1	5'- ATGGGATTGATAAAACGTTCCCG-3'
2	L 50463_R2	5'- TCCTCCACCCAATCTATCAACCC -3'
3	C13604_R1	5'- TCAGATCTCATAGAGCAGCTTCTCA -3'
4.	C13604_R2	5'- CATATGCTCCGCCTGCAGAG -3'
5.	L137834_R1	5'- AAAACATATAGTCAGGAGTAAGAGGAG -3'

2.3.4 Set IV

For 5' RACE PCR reactions reverse primers with higher number of nucleotides and high Tm were designed for specific binding. The primers (Table 7) were based on the sequence of contig_13604 from the transcriptome of *Ocimum.*

Table 7 List of reverse primers for 5'RACE PCR - set IV

No.	Oligo Name	Sequence
1.	13MR	5'-GTCGGCCGCGTGGTGCTGCAT-3'
2	13FLR1	5'- TTAAATCTCATAGAGCAGCTTCTCA-3'
3	13FLR2	5'- TCAGATCTCATACAGCAGCTTCTCG-3'

2.3.5 Set V

Based on the sequence of full length contig_13604 received after RACE PCR reaction, these primers (Table 8) were designed to clone full length of contig 13604.

Table 8 List of primers for full length sequencing - set V

No.	Oligo Name	Sequence
1.	CFATFF (forward)	5'-ATGAAGATCGTCGTGAAAGATTCGACACTGG-3'
2	CFATFR (reverse)	5'-TCAGATCTCATACAGCAGCTTCTCGAAAAC-3'

2.4 Molecular Cloning

2.4.1 PCR reaction (Polymerase chain reaction)

The components for a PCR reaction were brought on ice. Components were added to prepare the reaction mixture, such that each 10 µL of mixture would contain - Jumpstart mastermix: 5 µL, forward and reverse primers (1:10 dilution): 1 µL each and final volume was made up with nuclease free water. 1 µL (10 ng) of cDNA was added to each aliquot of reaction mixture respectively. However, for cloning of full length gene AccuPrime *Pfx* Supermix (9 µL) was used instead of Jumpstart mastermix. A 10 µL aliquot of reaction mixture with no added cDNA was used as a negative control.

General description of PCR reaction:

PCR reaction was carried out as follows – (1) initial denaturation at 95 °C for 5 minutes (2) subsequent denaturation at 95 °C for 30 seconds (3) annealing at X °C for 30 seconds (4) extension at 72 °C for Y seconds (5) final extension at 72 °C for 10 minutes was carried out and stored at 4 °C. Steps (2) to (4) was repeated for 35 cycles.

Here, X and Y imply annealing temperature and extension time respectively. However, the values for X and Y vary for individual PCR reactions (Table 9).

Table 9 Primer set for PCR reaction and corresponding values for X and Y

Set/ Type	Primer pair	X in °C	Y in sec	Y' in sec
Set I	CFAT F1 – CFAT RTR	52	40	60
Set II	CFAT PF2 – CFAT PR1	48	50	70
	CFAT F1B – CFAT R1	50	72	82
	CFAT F1D – CFAT R1	50	72	82
Set III	C13604_F1 – C13604_R1	54	78	88
	C13604_F2 – C13604_R2	54	54	74
3'RACE	C13604_F2 – GeneRacer 3' primer	64	80	100
	L 50463_F2 – GeneRacer 3' primer	68	70	90
5'RACE	C13604_R2 – GeneRacer 5' primer	58	108	128
	13 MR - GeneRacer 5' primer	72	60	80
	L 50463_R2 – GeneRacer 5' primer	64	60	80
Full length	CFATFF - CFATFR	56	78	98

2.4.2 Gel Extraction and PCR Purification

The excised band from the 1% agarose gel was processed using PureLink® Quick Gel Extraction and PCR Purification COMBO Kit (Invitrogen), following the manufactures protocol. The PCR products too were purified using the above mentioned kit. The final elution was done in warm water (heated at 65 °C). Qualitative and quantitative analysis of the purified product was done using Thermo-scientific NanoDrop2000c spectrophotometer. 40-50 ng of purified product was mixed with 2 µL of loading dye and analyzed on 1% agarose gel.

2.4.3 Ligation

The purified amplicon obtained through PCR reaction was ligated into pCR 2.1 linearized vector using Original TA Cloning Kit (Invitrogen). The amount of PCR product was determined as per the following formula:

$$x \text{ ng of PCR product} = \frac{(y \text{ bp of pcr product})(50 \text{ ng of pCR 2.1 vector})}{(\text{size in bp of pCR 2.1 vector} \sim 3900)}$$

Where x ng is the amount of PCR product of y base pairs to be ligated for a 1:1 (vector:insert) molar ratio.

Used (vector:insert) molar ratio =1:3

The ligation mixture was prepared by adding 5X T4 DNA Ligase Buffer (2 µL), pCR 2.1 vector (2 µL), DNA Ligase (1 µL), PCR product and volume was made up to 10 µL using water. The ligation mixture was incubated at 14 °C overnight.

2.4.4 Preparation of Chemically competent *E.coli* cells (TOP10 and Rosetta)

Materials

Media – Luria Broth (LB), 0.1 M $MgCl_2$, 0.1 M $CaCl_2$, 40% glycerol, Dimethyl sulfoxide (DMSO) Autoclaved centrifuge tubes and 1.5 mL collection tubes.

Method

5 mL of glycerol stock (TOP10 *E.coli* cells) was streaked on a LB agar plate and incubated at 37°C overnight. Next day a single colony was inoculated in 5 mL LB media and incubated overnight at 37 °C with shaking at 200 rpm. 500 µL of the grown culture was inoculated in 50 mL of LB in 250 mL conical flask. The culture was incubated at 37 °C with shaking at 200 rpm until the OD_{600} reaches ~0.4. The culture was then cooled at ice for ~30 minutes. Following this, the culture was transferred into centrifuge tubes and centrifuged at 2500 rcf for 5 minutes at 4 °C. The cell pellet obtained was resuspended in 25 mL of 0.1 M of filter sterilized ice-cold $MgCl_2$ and stored on ice for ~20 minutes. Next, the solution was centrifuged at 2500 rcf for 5 minutes at 4 °C. The cell pellet

obtained is resuspended in 25 mL of 0.1 M of filter sterilized ice-cold $CaCl_2$ and stored on ice for ~20 minutes. The solution was gain centrifuged at 2500 rcf for 5 minutes at 4 °C. The obtained cell pellet was now dissolved in 1.3 mL $CaCl_2$, 500 µL of 40% glycerol and 200 µL of DMSO. 100 µL aliquots of the solution were transferred in 1.5 mL chilled sterile collection tubes and stored at -80 °C.

Rosetta 2(DE3) competent cells were also prepared using the same method.

2.4.5 Transformation

The ligated product was transformed into TOP10 competent cells. 10 µL of the ligated product was added to a 100 µL aliquot of TOP10 cells and stored on ice for 20-30 minutes. The mixture was given a heat shock for 45 seconds at 42 °C and then immediately transferred to ice. After an interval of 5-10 minutes 250 µL of S.O.C. medium was added. The vials were incubated in a shaking incubator at 37 °C and 200 rpm for 1 hour. The whole content of transformation vial was spread on LB agar plates containing 100 µg/mL ampicillin. The plates were incubated overnight at 37 °C.

2.4.5 Screening of positive colonies using PCR

Colonies were picked up from the incubated plates and inoculated in 200 µL tubes containing 10 µL of prepared PCR mixture. Each 10 µL aliquot of PCR mixture contained Jumpstart mastermix: 5 µL, M13 Forward and M13 Reverse primer: 500 µL each and water: 4 µL. PCR reaction was carried out as follows – (1) initial denaturation at 95 °C for 5 minutes (2) subsequent denaturation at 95°C for 30 seconds (3) annealing at 54 °C for 30 seconds (4) extension at 72 °C for Y' seconds (5) final extension at 72 °C for 10 minutes was carried out and stored at 4 °C. Steps (2) to (4) was repeated for 35 cycles.

Here Y' is the extension time (Table 7) which is ~12 seconds more than that used for the respective PCR reaction to obtain the desired amplicon.

2.4.6 Plasmid purification

The positive colonies were inoculated in 5 mL LB with 100 µg/mL ampicillin. The culture was incubated overnight at 37 °C and 200 rpm. The culture was then centrifuged at 13400 rpm for 2 minutes. The harvested cell pellet was processed as per the

manufacturer's protocol using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich) to obtain purified plasmids. Qualitative and quantitative analysis of the purified product was done using Thermo-scientific NanoDrop2000c spectrophotometer. 40-50 ng of plasmid was mixed with 2 µL of loading dye and analyzed on 1% agarose gel.

2.5 RACE (Rapid Amplification of cDNA Ends)

The GeneRacer™ Kit (Invitrogen) was used to obtain full length sequence of desired gene. The manual says that "kit provides a method to obtain full length 5' and 3' ends of a cDNA using known cDNA sequence from expressed sequence tags (ESTs). It ensures the amplification of only full-length transcripts via elimination of truncated messages from the amplification process. The obtained RACE PCR products could then be easily cloned using TA cloning kit". RACE ready cDNA was prepared using the manufacturer's protocol. It involved the steps: dephosphorylating RNA, removing the mRNA cap structure, ligating the RNA Oligo to decapped mRNA and reverse transcribing mRNA. 3' and 5' RACE PCR was carried out using the cDNA of *Ocimum tenuiflorum*. The conditions and primer sets were as mentioned in **2.4.1** (Table *2.*7). The 3' RACE PCR was done using gene specific forward primer and GeneRacer 3' primer - 5′-GCTGTCAACGATACGCTACGTAACG-3′. Similarly, the 5' RACE PCR was carried out using gene specific reverse primer and GeneRacer 5' primer - 5′-CGACTGGAGCACGAGGACACTGA-3′.

2.7 Recombinant protein expression

2.7.1 Cloning in expression vector pET 32c

PCR product obtained with M13 forward and reverse primer using plasmid containing the full length gene was purified and digested using *EcoRI*. The reaction mixture for Restriction Enzyme (RE) digestion contained 1 µg of PCR product, 0.5 µL (10 units) of *EcoRI*, 2 µL of 10X reaction buffer, 2µL of 10X BSA and water added to make up volume to 20 µL. Similarly, 1 µg of pET 32c vector was digested by *EcoRI*. Full length gene was cloned into the pET32c vector by *EcoRI* single digestion. For this both mixtures were incubated at 37 °C for 3 hours and then heat inactivated at 65°C for 20 minutes. The digested mixture was purified using GenElute™ PCR Clean-up Kit (Sigma).

Ligation reaction was done with T4 DNA ligase, using 1:10 molar ratio. The ligated mixture was incubated at 23 °C for 3 hours. The mixture was then transformed in TOP10 cells and spread on a Luria Agar (LA) agar plate containing 100 µg/mL ampicillin. The plate was incubated overnight at 37 °C. The colonies on plate were screened by doing a colony PCR using T7 promoter and T7 terminator primers. The positive colonies were inoculated in 5 mL LB and grown overnight at 37 °C with shaking at 200 rpm. Cells were harvested from the culture and processed for purification of plasmid. Qualitative and quantitative analysis of the purified product was done and analyzed on 1% agarose gel.

2.7.2 Expression of gene in Rosetta 2(DE3) cells

The plasmid containing the insert (full length Contig_13604) was transformed into Rosetta 2(DE3) competent cells and spread to grow on a LB agar plate containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. Positive colonies were screened and inoculated in 5 mL LB (primary culture) and grown overnight at 37 °C with shaking at 200 rpm.

2.7.3 Analysis of expressed protein

50 µL of primary culture was inoculated in four tubes with aliquots of 5 mL Terrific Broth (TB) each containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. All four

tubes were incubated at 37 °C with shaking at 200 rpm until the OD_{600} reached 0.8. At this point two tubes were induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). The two un-induced tubes were used as a negative control. One set of tubes were incubated overnight at 16 °C and other set at 28 °C for 6 hours. The cells were harvested and resuspended in 1 mL of lysis buffer containing 50 mM Tris-HCl, 300 mM NaCl, 10% v/v glycerol, 1mM PMSF, 0.5% CHAPS and 1mg/mL lysozyme. The solution was subjected to sonication and then centrifuged at 13000 rpm for 5 minutes. The pellet and supernatant were separated. Pellet obtained was dissolved in 1 mL buffer containing 50 mM Tris-HCl, 300 mM NaCl, 10% v/v glycerol. 20 µL of each (dissolved pellet as well as supernatant) was mixed with 10 µL SDS gel loading dye and heated at 70 °C for 10 minutes. The samples were analyzed on 12% SDS-PAGE.

Chapter 3

Results and Discussion

3.1 Phenylpropanoids

The table below (Table 10) provides a brief overview of the major phenylpropanoids that are present in the leaf tissues of 5 representative species of *Ocimum.* Eugenol and eugenol methyl ether are the most significant phenylpropanoids found in *Ocimum* that are derived from the monolignol, coniferyl alcohol. Coniferyl alcohol acyltransferase plays a crucial role in diverting the metabolic flux away from lignin biosynthesis by catalyzing the synthesis of coniferyl acetate from coniferyl alcohol (Figure 4).

Table 10 List of phenylpropanoids present in leaf tissues across *Ocimum* species.

Metabolite / Species	Eugenol	Eugenol Methyl Ether	Estragole	c-Methyl cinnamate	t-Methyl cinnamate	Anethol
O. gratissimum	78.25	0.06	n.d.	n.d.	n.d.	n.d.
O. tenuiflorum I	0.05	60.41	n.d.	n.d.	n.d.	n.d.
O. tenuiflorum II	n.d.	72.69	n.d.	n.d.	n.d.	n.d.
O. kilimandscharicum	0.05	n.d.	n.d.	n.d.	n.d.	n.d.
O. basilicum I	n.d.	n.d.	0.38	2.24	63.77	n.d.
O. basilicum II	n.d.	0.47	3.48	11.48	46.35	n.d.
O. basilicum III	n.d.	n.d.	0.86	5.23	54.73	n.d.
O. basilicum IV	0.04	0.02	82.39	n.d.	0.03	0.1
O. americanum	0.90	n.d.	n.d.	n.d.	n.d.	n.d.

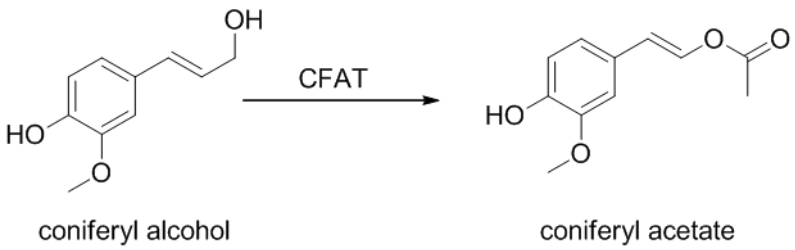


Figure 4 Reaction catalyzed by CFAT (coniferyl alcohol acyltransferase)

3.1 Cloning

3.1.1 PCR using primer set I

An initial set of primers based on the available nucleotide sequence of coniferyl alcohol acyltransferase from *Petunia x hybrida* (GenBank: DQ767669.1) was used to clone the gene from *Ocimum*. Out of the three pairs designed, a combination of CFATF1 and CFAT RTR seemed to generate an amplicon with leaf tissue cDNA of *Ot* I, *Ot* II, *Oba* I and *Oba* II (Figure 5). The PCR product of *Ot* I was cloned in pCR 2.1 vector and transformed into TOP10 cells. Colony PCR with M13 forward and reverse primers was done to select positive colonies (Figure 6).

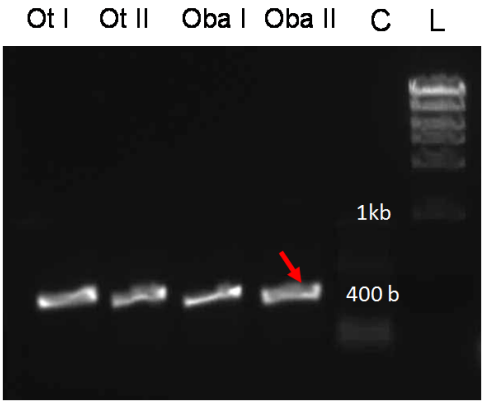


Figure 5 PCR amplification results of CFAT F1-CFAT RTR

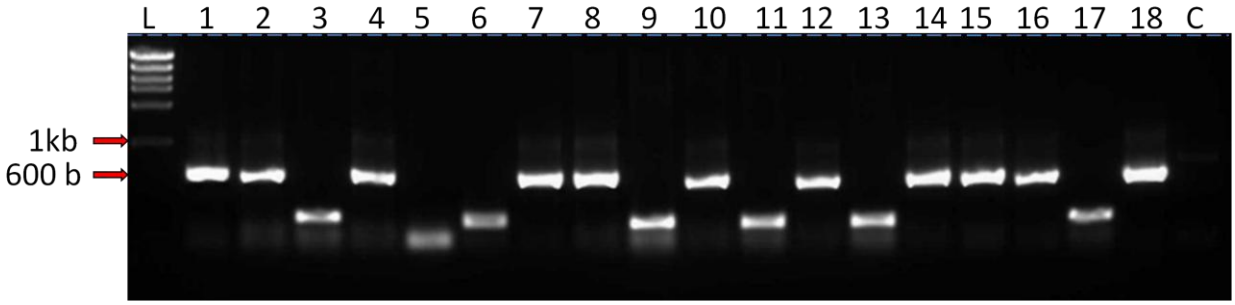


Figure 6 Colony PCR with M13 F and M13 R primers.
Numbers represent colony number. Positive colonies – 1,2,4,7,8,10,12,14,15,16,18

The combination of CFAT F1 and CFAT RTR as a primer set was expected to generate ~898 bp amplicon. However, the sequencing results were indicative of a ~400 bp amplicon. The sequence did not correspond to acyltransferase as per BLAST results against the database and showed a similarity of <16 % with respect to the reported sequence, *Ph*CFAT.

3.1.2 PCR using primer set II

Combination that showed positive results with *Og* (Young Leaf) cDNA are summarized in Table 10. The PCR products obtained are as presented in figure 8 and figure 9.

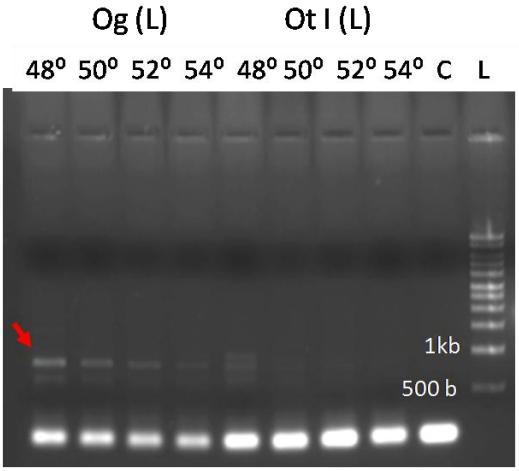


Figure 8 PCR products using CFAT PF2-CFAT PR1 using cDNA from young leaves of *Og* and *Ot*I at 48°C, 50°C, 52°C and 54°C

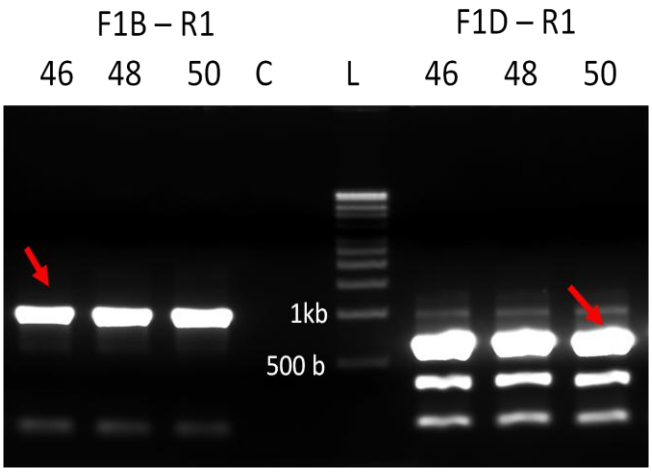


Figure 7 PCR products using CFAT F1B/D – CFAT R1 with cDNA from young leaves of *Og* at 46°C, 48°C and 50°C

Table 10 Summary of sequenced PCR product using primer set II

Name of Combination	Expected Amplicon size	Obtained Amplicon size	Similarity %
CFAT PF2 – CFAT PR1	~785 bp	544 bp	~19%
CFAT F1B – CFAT R1	~1200 bp	820 bp	~6%
CFAT F1D – CFAT R1	~1200 bp	618 bp	~14%

Ligation of these PCR products in pCR 2.1 vector was followed by transformation in TOP10 cells. Positive colonies were screened to obtain plasmid, which were purified and sent for nucleotide sequencing. The sequencing results were analyzed by aligning and BLAST search against available database and finally summarized as in Table 10.

The BLAST results were not indicative of coniferyl alcohol acyltransferases as could be inferred from figure 9.

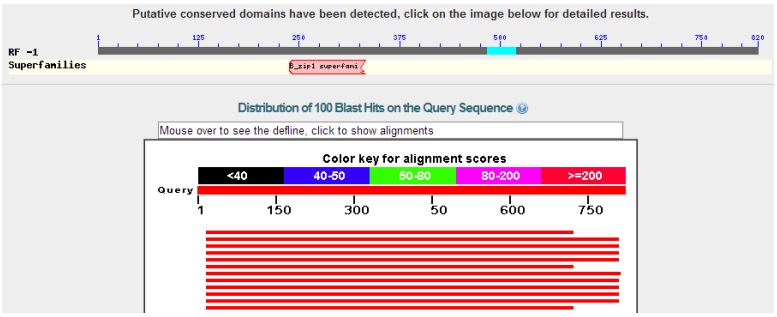


Figure 9 Blast results for CFAT F1B - CFAT R1

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics

Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: transcription factor MYC2-like [Solanum lycopersicum]	278	278	85%	3e-85	58%	XP_004244391.1
MYC transcription factor 2 [Solanum lycopersicum]	279	279	96%	4e-85	53%	AGZ94899.1
PREDICTED: LOW QUALITY PROTEIN: transcription factor MYC2 [Solanum lycopersicum]	279	279	96%	5e-85	53%	XP_004245895.1
MYC1b transcription factor [Nicotiana tabacum] >gb|ADU60100.1| MYC2a transcription factor [Nicotiana tabacum]	278	278	96%	1e-84	54%	ADH04268.1
MYC1a transcription factor [Nicotiana tabacum] >gb|ADU60101.1| MYC2b transcription factor [Nicotiana tabacum] >gb|ADU60102.1| MYC2c transcription fac	278	278	96%	1e-84	54%	ADH04267.1
MYC transcription factor [Solanum tuberosum]	276	276	85%	2e-84	56%	CAF74711.1

The primers CFAT PF2 was designed such that it would cover the region of conserved motif HxxxD. As mentioned earlier HxxxD is a characteristic conserved motif of BAHD group of acyltransferases.

The primers CFAT F1B, CFAT F1D and CFAT R1 were designed using the concept of codon bias. We know that more than one codon codes for a particular amino acid. The frequency of a codon coding a specific amino acid varies across organisms. Thus, a region of amino acids was selected from the reported sequence of PhCFAT. For coding the selected stretch of amino acids the codons most frequent to *Ocimum* were arranged to get the desired sequence of nucleotides as a primer.

But the amplicons generated indicated that these set of primers were not able to bind specifically to the desired regions of the gene, coniferyl alcohol acyltransferase in *Ocimum* plants.

3.2 Transcriptomic analysis

Table 11 List of selected transcripts

Sl.	Name	Length	Region of query covered	Score(bits)
1	Contig_13604	1631	31-433	147 (371)
2	Locus_50463	561	185-355	135 (341)
3	Locus_137834	201	384-448	100 (247)
4	Locus_4888	620	5-171	94 (232)
5	Locus_31505	714	60-283	84 (207)
6	Locus_51064	812	8-199	82 (200)
7	Locus_59548	874	8-287	79 (192)
8	Locus_678882	1181	64-390	70 (168)
9	Locus_91576	380	64-175	66 (160)
10	Locus_25941	628	64-220	58 (140)
11	Locus_6389	461	94-235	56 (134)
12	Locus_16093	588	94-261	55 (130)
13	Locus_116133	477	114-242	49 (115)

Length is in bp, the region of query sequence that they cover in bp and respective alignment score.

Query: PhCFAT - 1836bp, 454aa

1-3 indicate the transcripts selected based on the tblastn search from the Terpene Biosynthesis pathway.

4-13 indicate the transcripts selected based on tblastn search results from the Lignin Biosynthesis pathway.

After the unsuccessful attempt at cloning the gene using primers designed based on the sequence of *PhCFAT*, the transcriptomic data of *Ocimum* was considered for analysis. The database under study consisted of transcripts annotated on (i) Terpene biosynthesis pathway and (ii) Lignin biosynthesis pathway and created using leaf, stem and inflorescence tissues of 3 species of *Ocimum*. The sequence of reported *PhCFAT* was used as a query for local BLAST against the database of *Ocimum* transcripts using Scoring Matrix Blosum62. The blastn with *PhCFAT* as a query resulted in hits with very low scores and hence, were not selected for primer designing. The blastn is a nucleotide–nucleotide BLAST, it accepts a DNA query and returns the most similar DNA sequence.

However, tblastn with the peptide sequence of PhCFAT as a query resulted in significant hits. The tblastn is a protein-nucleotide BLAST, it accepts a protein query and compares it against the all six reading frames of a nucleotide sequence database. Transcripts obtained were further filtered based on their alignment score, profile, and the presence of conserved motifs (HxxxD and DFGWG). Thirteen transcripts were

selected in total (Table 11), which were further aligned with the query sequence to figure out the overlapping regions.

Initially, only the top three results were taken into account for designing primers due to their scores. Hence, two pairs of primers were designed based on the sequences of Contig_13604 and Locus_50463 each and one pair for Locus_137834 (primer set III).

3.2.1 PCR using primer set III

Results of PCR reaction with primers designed for Contig_13604 are shown in figure 10 and figure 11.

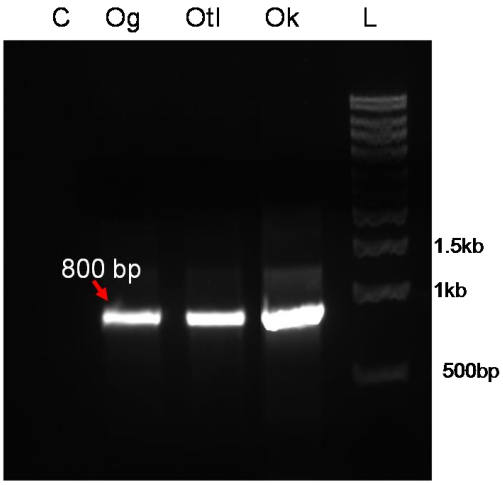


Figure 11 PCR using C13604 F2-C13604 R2 with young leaf cDNA of *Og, Otl, Ok*

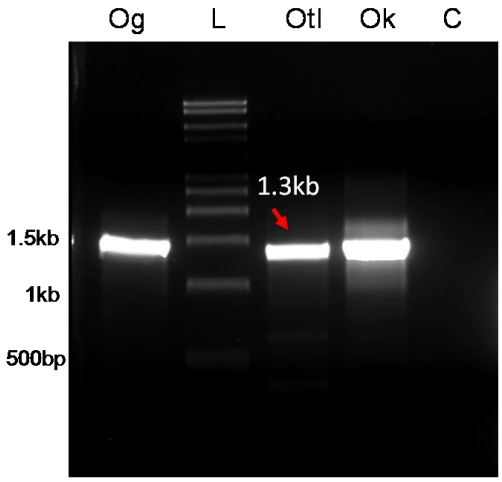


Figure 10 PCR using C13604 F1-C13604 R1 with young leaf cDNA of *Og, Otl, Ok*

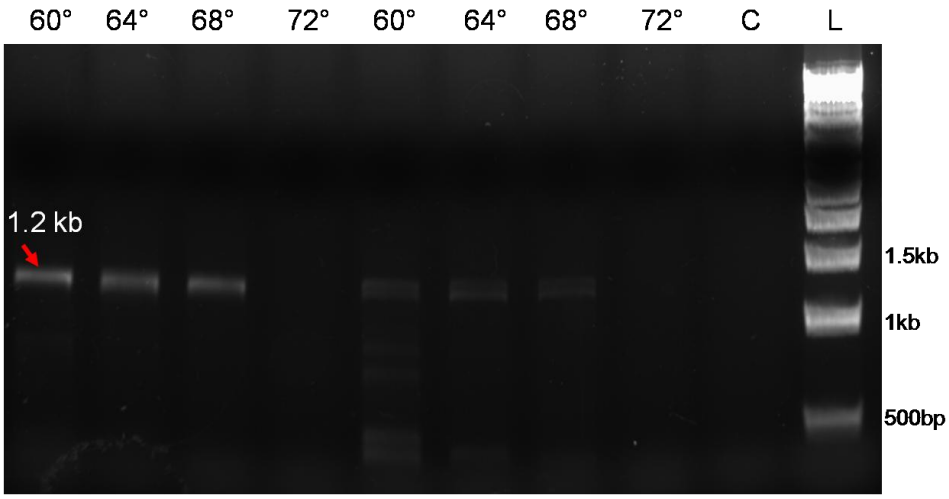


Figure 12 3'RACE PCR with C13604 F2 and using young leaf cDNA of *Otl*

PCR product obtained using C13604 F2 and C13604 R2 was confirmed by sequencing. Next step was to obtain 5' and 3' ends of the Contig_13604 by RACE PCR reaction. 3' RACE PCR using C13604 F2 as forward primer generated a unique band with cDNA of *Ot* I (Figure 12). This band also corresponded to the expected size of ~1300 bp. However, 5'RACE PCR using C13604 R2 generated multiple bands of varying sizes. After scanning a range of temperature conditions, we switched to using a different set of reverse primers (set IV) for 5'RACE PCR reaction. The reverse complement of C13604 F2 named as 13MR was used as the reverse primer for producing the 5' end of the Contig_13604. This strategy successfully generated bands of expected size (Figure 13). The band of interest was purified using gel extraction and further processed to obtain the full length 5' end of the gene.

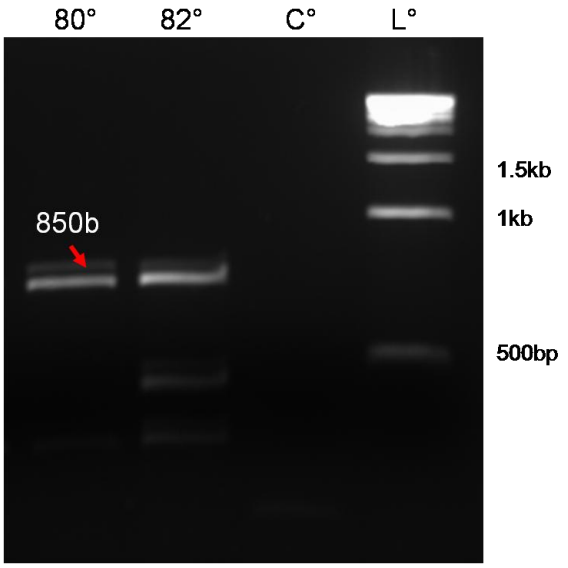


Figure 13 5'RACE PCR with 13MR using young leaf cDNA of *Ot*I

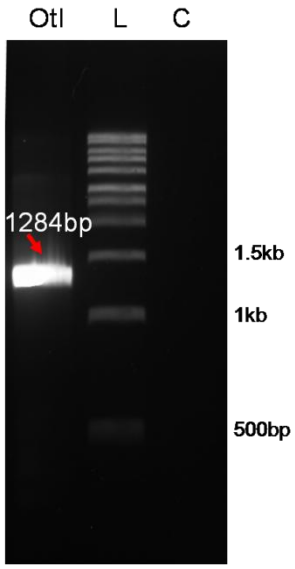


Figure 14 PCR for full length gene with CFATFF and CFATFR using *Ot*I cDNA

3.3. Full length cloning

The open reading frame (ORF) for full-length contig_13604 encodes a protein of 427 amino acids, 1284 nucleotides long, with a calculated molecular mass of 46.9 kDa.

The final step in cloning the contig_13604 was to obtain its full length DNA sequence. A pair of primers (set V) was designed specific to nucleotide stretch at the 5' and 3' ends, based on the sequencing reports of 3' and 5' RACE PCR products. The PCR product generated by CFAT FF and CFAT FR (Figure 14) was purified and transformed into cloning vector pCR 2.1. The product was verified by nucleotide sequencing.

The sequence was translated to peptide sequence and aligned to PhCFAT which resulted in a score of 21. The sequence was analyzed by using it as a query for BLAST against the database. The results of BLAST were indicative of similar genes but from different species as generated when the reported PhCFAT is given as a query.

Since, the genus *Petunia* and *Ocimum* belong to different order in the kingdom Plantae, so it might be expected that the proteins are significantly different. However, similar BLAST results for both proteins can be a positive indication that this gene could be a coniferyl alcohol acyltransferase in *Ocimum*.

Hence, the BLAST results for PhCFAT as well as contig_13604 were further analyzed (**Sec 3.4**).

However, confirmation of this gene being a coniferyl alcohol acyltransferase could be achieved by enzyme activity assay. Hence, the next step was to successfully express the recombinant protein in *E.Coli* cells.

3.4 Sequence and Expression analysis

3.4.1 Multiple Sequence Alignment

The peptide sequences of PhCFAT and its top two BLAST results were aligned with peptide sequence of full length Contig_13604 and its corresponding BLAST results to observe sequence similarity. The alignment clearly presents the motifs HxxxD and BAHD to be highly conserved across the class of BAHD acyltransferases (Figure 16).

3.4.2 Dendrogram

The peptide BLAST results of Contig_13604 and PhCFAT were subjected to phylogenetic analysis (Figure 15). This analysis was carried out to get information about the difference in evolutionary origin of acyltransferase in relatively distant family of

plants. As *Petunia* and *Ocimum* are distantly related genus, hence they lie separate in the Dendrogram along with proteins closely related to them grouped separately. This also added to the assumption that contig_13604 could be a coniferyl alcohol transferase in *Ocimum*.

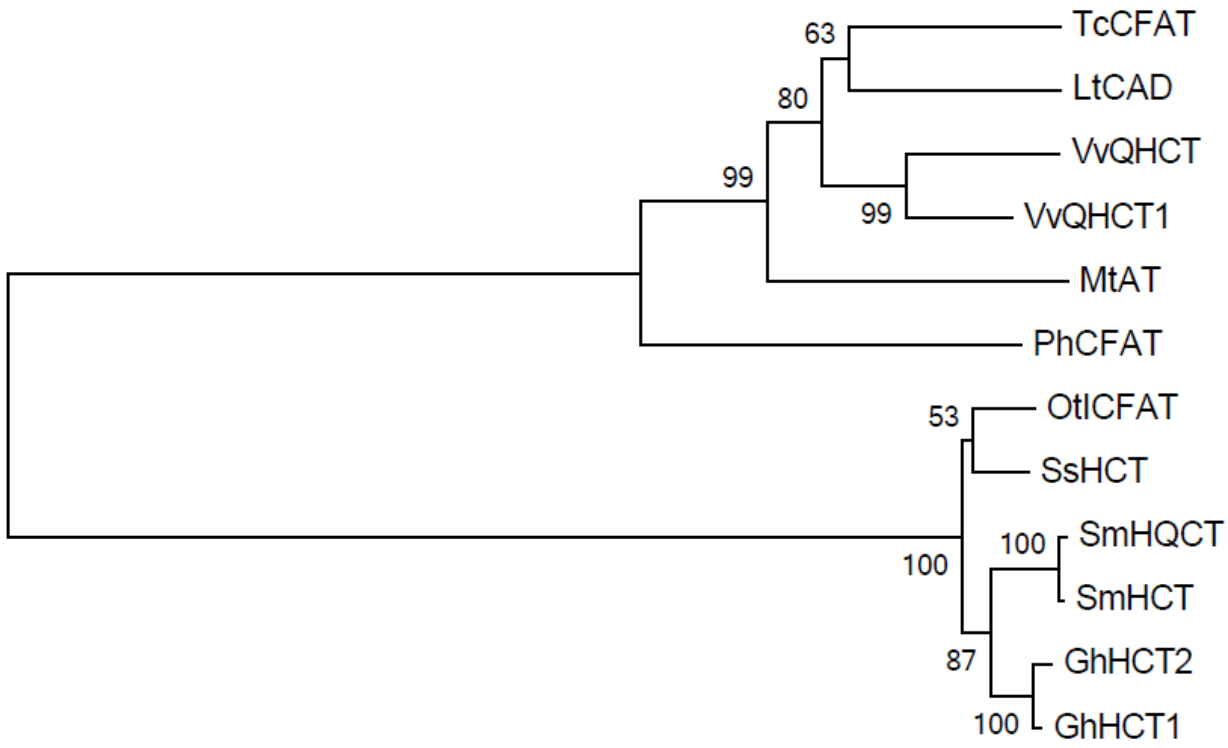


Figure 15 Dendrogram contig_13604 and related proteins with PhCFAT and related proteins.

Sequences represented as OtICFAT (Contig_13604), SsHCT (GeneBank: CBI83579.1), SmHQCT (GeneBank: ACA64049.1), SmHCT (ADG46003.1) GhHCT2 (CDG56249.1), GhHCT1 (CDG56248.1), TcCFAT (GeneBank: EOY26190.1), LtCAD (GeneBank: AHA90802.1), VvQHCT (XP_002270252.1), VvQHCT1 (XP_002270251.1), MtAT (XP_003617368.1) and PhCFAT (GeneBank: DQ767669.1). The complete amino acid sequences were analyzed by MEGA6 and tree was built using the Neighbour joining Method.

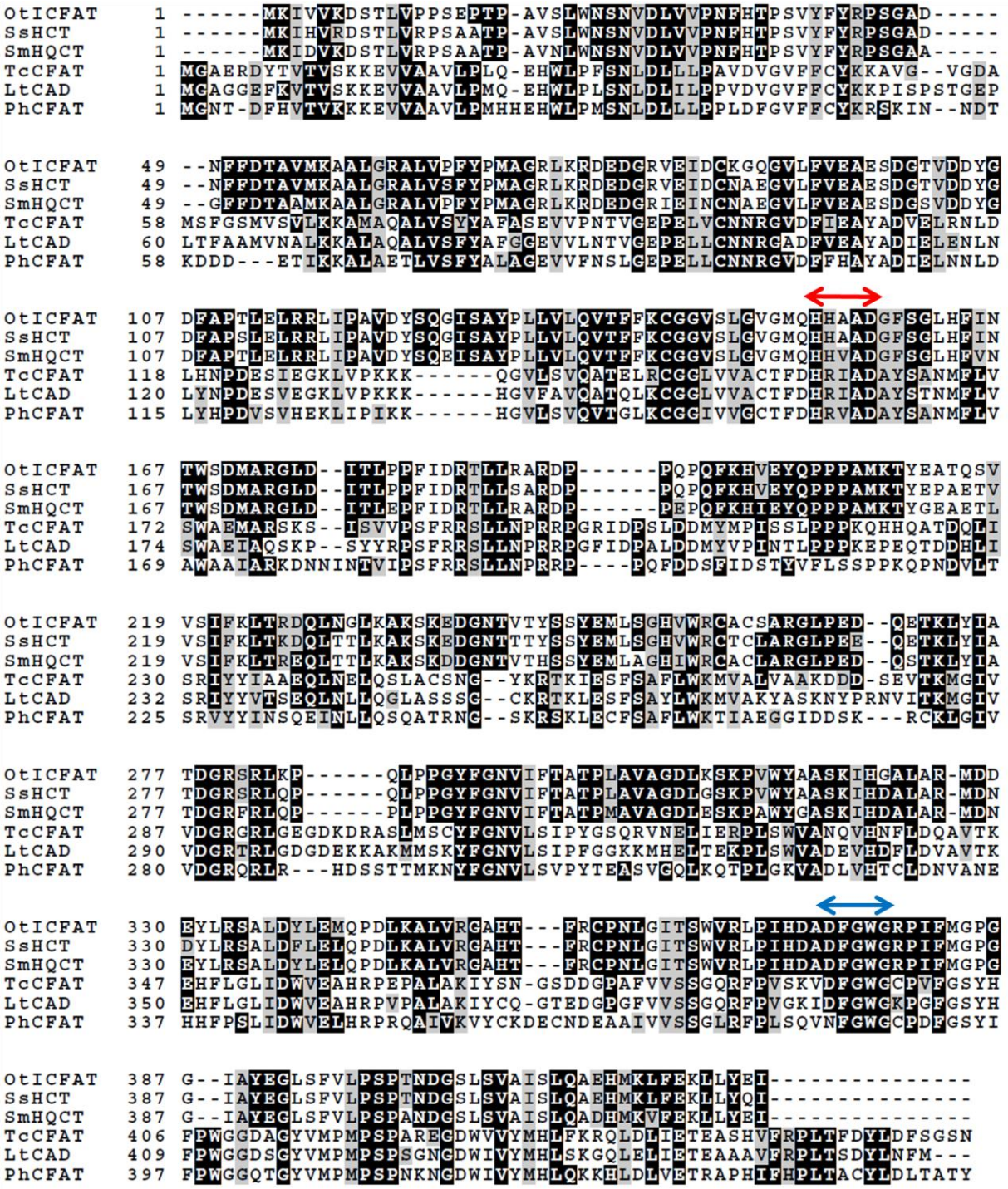


Figure 16 Peptide sequence alignment of contig_13604 and related proteins with PhCFAT and related proteins.

Sequences represented as OtICFAT (Contig_13604), SsHCT (GenBank: CBI83579.1), SmHQCT (GenBank: ACA64049.1), TcCFAT (GenBank: EOY26190.1), LtCAD (GenBank: AHA90802.1) and PhCFAT (GenBank: DQ767669.1). Sequences were aligned using CLUSTALX and shaded using BOXSHADE 3.21. Residues highlighted in black represent identical matches in at least three of the sequences, while residues highlighted in gray represent similar matches. Dashes indicate gaps in the sequence inserted to attain optimal sequence alignment. The HxxxD and DFGWG conserved motifs are identified by horizontal red and blue colored arrows respectively.

3.4.3 Expression analysis

3.4.3.1 Semi-quantitave PCR

The expression of contig_13604, EGS and EOMT was investigated by RT-PCR across all *Ocimum sp.* under study. The presented data (Figure 18) consists only of 5 representative species (for simplicity). The contig_13604 (labeled as CFAT in figures) is strongly expressed in the leaf tissues of *Ot*I but expression is decreased in *Og* and *Oa*. However, expression of eugenol is relatively more in *Og* than in *Ot* I and *Oa*. In addition, the expression of EOMT is strongest in *Ot*I than and is barely expressed in others. These results correlated well with the observed eugenol and eugenol methyl ether (EME) levels in leaves of different species. As eugenol is substrate for EME synthesis, in case of *Ot* I, even though Eugenol is detected in traces in oil, EGS expression was observed. Apart from expression levels of contig_13604 in young leaf tissues, expression in other tissues like stem and mature leaves was also studied (Figure 19). It was observed that expression levels of contig_13604 in stem and mature leaf tissues of *Og*, *Oa* and *Ot* I are similar, indicative that tissue specific expression is variable.

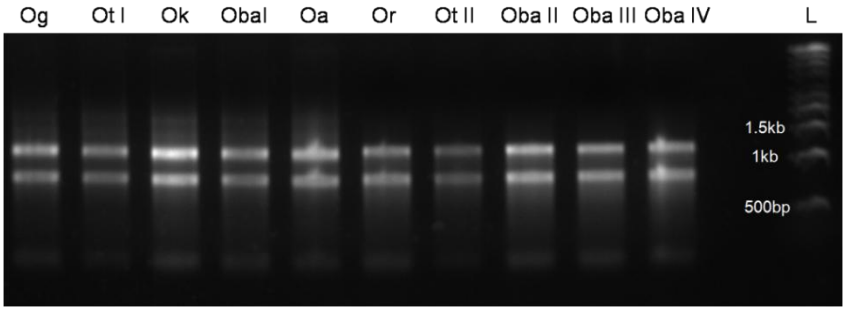


Figure 17 Isolated RNA after DNAse treatment from *Ocimum* plants

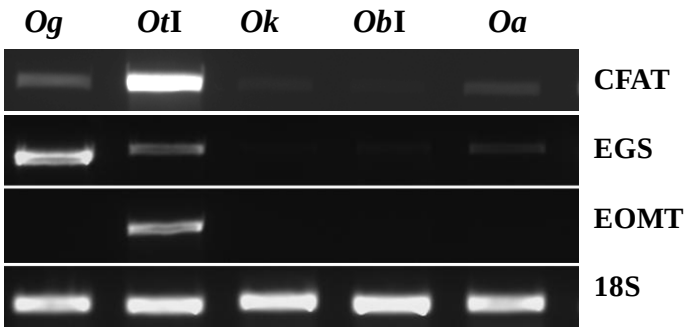


Figure 18 Expression of contig_13604 (here represented as CFAT), EGS and EOMT in young leaf tissue.

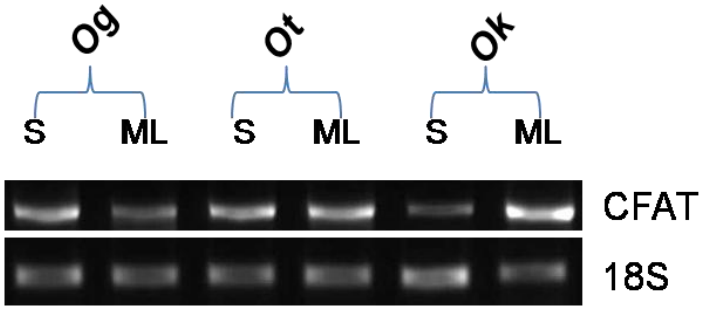


Figure 19 Expression of contig_13604 (here represented as CFAT in stem and mature leaf tissue.

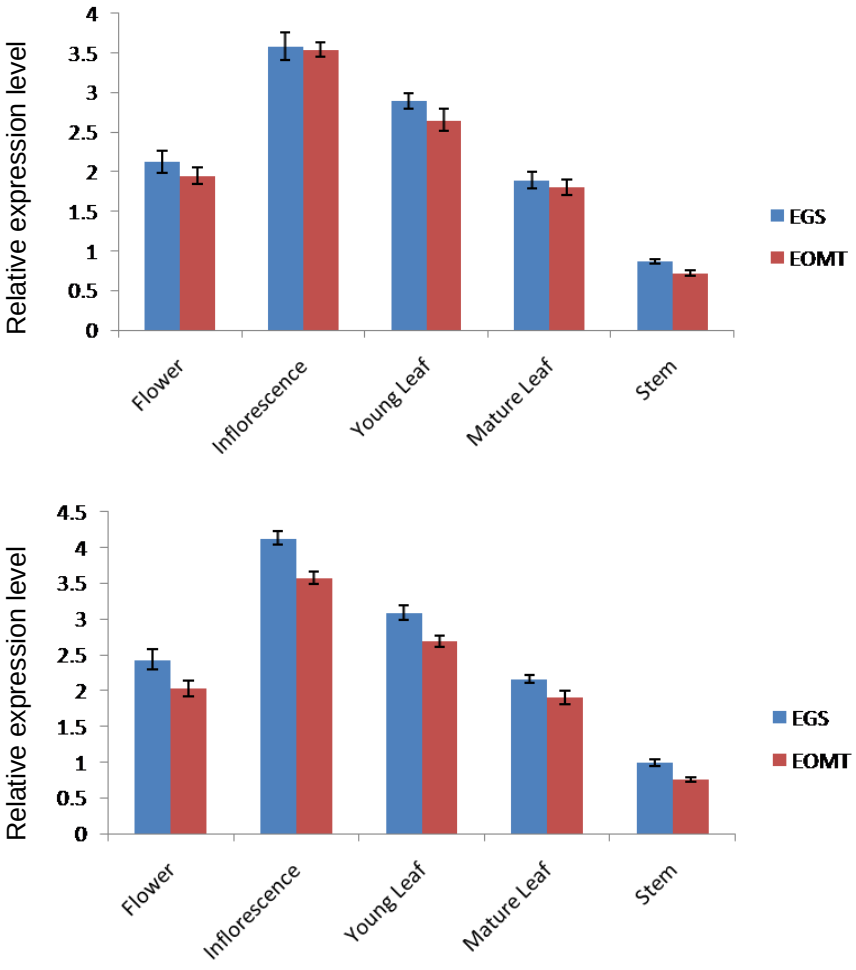


Figure 20 Relative expression levels of EGS and EOMT in tissues of (A) *Ot* I (B) *Ot* II. Bar indicates SE

2.4.3.2 Real Time PCR

Expression levels of EGS and EOMT was checked in 5 tissues of *Ot* I and *Ot* II, and it was observed that Inflorescence tissue which is rich in phenylpropanoid content, had the highest expression levels of EGS and EOMT (Figure 20). It was also observed that young plant parts such as young leaves and inflorescence had higher expression levels of these genes in comparison to mature parts e.g. mature leaves and flowers. This observation is consistent with the role of phenylpropanoids in plant protection, acting as anti-feedent, deterrent etc. Real time PCR of CFAT will yield insights into the levels of production of these compounds, which will be carried out after the contig_13604 confirmation.

3.5 Expression of recombinant protein

In order to express *O. tenuiflorum* contig_13604, the 1284 bp ORF was cloned into the pET 32c expression vector by *EcoRI* single digestion. The vector was transformed in Rosetta 2 (DE3) cells and expressed in small scale (5 mL) to optimize conditions for recombinant protein expression.

The recombinant protein of the gene with His Tag coding sequence had a calculated mass of 65kDa. The expression profile (Figure 21) suggested that the expressed protein of interest was observed in the pellet fraction of Rosetta 2 (DE3) cells.

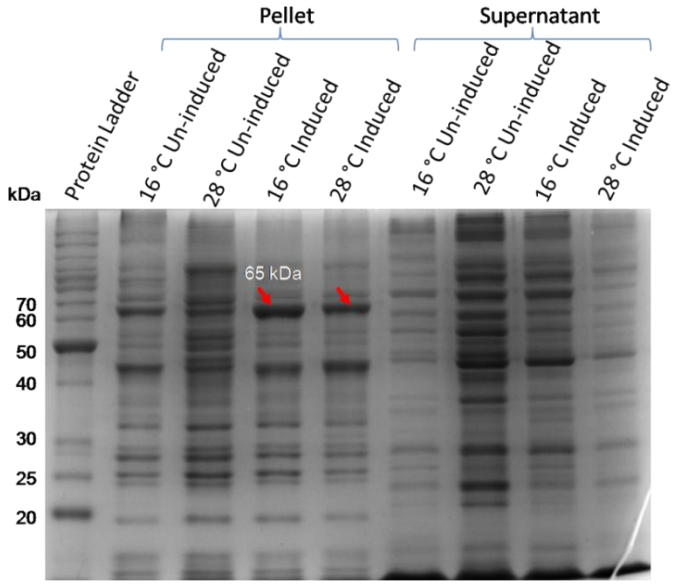


Figure 21 SDS Gel of protein expression profile at 16°C and 28°C

Further work comprises of:

1. In gel trypsin digestion of expected protein to peptides for mass spectroscopy analysis.
2. If the mass spectroscopy confirms the protein then large scale expression will be performed. In addition efforts to optimize solubility condition to obtain the protein in soluble fraction would also be required.
3. Enzyme activity assay using coniferyl alcohol as substrate in the presence of acetyl-CoA would provide a final confirmation of the gene as coniferyl alcohol acyltransferase.

References

B.B.Buchanan, W.Gruissem, R.L.Jones (2004). Natural products. In Biochemistry and Molecular Biology of Plants (1250-1318). I.K.International Pvt.Ltd.

Boerjan, Wout, John Ralph, and Marie Baucher. 2003. “Lignin Biosynthesis.” *Annual review of plant biology* 54: 519–46. http://www.ncbi.nlm.nih.gov/pubmed/14503002 (January 29, 2014).

Chogo JB, Crank G. 1981. Chemical Composition and biological activity of Tanzanian Plant *Ocimum* suave. *Journal of Natural Products* 44(3), 308-311

D’Auria, John C. 2006. “Acyltransferases in Plants: A Good Time to Be BAHD.” *Current opinion in plant biology* 9(3): 331–40. http://www.ncbi.nlm.nih.gov/pubmed/16616872 (March 20, 2014).

Deshpande, R.S., and Tipnis H.P. (1977). Insecticidal activity of *Ocimum* basilicum L. Pesticides, 11, 1e12.

Dexter, Richard, Anthony Qualley, Christine M Kish, Choong Je Ma, Takao Koeduka, Dinesh a Nagegowda, Natalia Dudareva, Eran Pichersky, and David Clark. 2007. “Characterization of a Petunia Acetyltransferase Involved in the Biosynthesis of the Floral Volatile Isoeugenol.” *The Plant journal : for cell and molecular biology* 49(2): 265–75. http://www.ncbi.nlm.nih.gov/pubmed/17241449 (February 11, 2014).

Fraser, Christopher M, and Clint Chapple. 2011. “The Phenylpropanoid Pathway in Arabidopsis.” 1–19.

Gang, David R, Jihong Wang, Natalia Dudareva, Kyoung Hee Nam, James E Simon, Efraim Lewinsohn, and Eran Pichersky. 2014. “An Investigation of the Storage and Biosynthesis of Phenylpropenes in Sweet Basil 1.” 1048.

Grienenberger, Etienne, Sébastien Besseau, Pierrette Geoffroy, Delphine Debayle, Dimitri Heintz, Catherine Lapierre, Brigitte Pollet, Thierry Heitz, and Michel Legrand. 2009. “A BAHD Acyltransferase Is Expressed in the Tapetum of Arabidopsis Anthers and Is Involved in the Synthesis of Hydroxycinnamoyl Spermidines.” *The Plant journal : for cell and molecular biology* 58(2): 246–59. http://www.ncbi.nlm.nih.gov/pubmed/19077165 (March 20, 2014).

Herrmann, K.M., and Weaver, L.M.(1999). The shikimate pathway. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. 50:473-503.

Hoffmann, Laurent, Stephane Maury, Francoise Martz, Pierrette Geoffroy, and Michel Legrand. 2003. “Purification, Cloning, and Properties of an Acyltransferase Controlling Shikimate and Quinate Ester Intermediates in Phenylpropanoid Metabolism.” *The Journal of biological chemistry* 278(1): 95–103. http://www.ncbi.nlm.nih.gov/pubmed/12381722 (March 20, 2014).

Koeduka, Takao, Gordon V Louie, Irina Orlova, Christine M Kish, Mwafaq Ibdah, Curtis G Wilkerson, Marianne E Bowman, Thomas J Baiga, Joseph P Noel, Natalia Dudareva, and Eran Pichersky. 2008. “The Multiple Phenylpropene Synthases in Both Clarkia Breweri and Petunia Hybrida Represent Two Distinct Protein Lineages.” *The Plant journal : for cell and molecular biology* 54(3): 362–74. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2741023&tool=pmcentrez&rendertype=abstract (February 11, 2014).

Ma, Xueyan, Juergen Koepke, Santosh Panjikar, Günter Fritzsch, and Joachim Stöckigt. 2005. “Crystal Structure of Vinorine Synthase, the First Representative of the BAHD Superfamily.” *The Journal of biological chemistry* 280(14): 13576–83. http://www.ncbi.nlm.nih.gov/pubmed/15665331 (March 22, 2014).

Nahak, Gayatri, R C Mishra, R K Sahu, and Central Rice. 2011. “Taxonomic Distribution , Medicinal Properties and Drug Development Potentiality of *Ocimum* (Tulsi).” 3(6): 95–113.

Rhodes, M.J.C., 1994. Physiological Roles for Secondary Metabolites in Plants : Some Progress , Many Outstanding Problems. Plant Mol. Biol. 24:1-2020.

Prakash, P, and Neelu Gupta. 2005. “REVIEW ARTICLE THERAPEUTIC USES OF *OCIMUM* SANCTUM LINN (TULSI) WITH A NOTE ON EUGENOL AND ITS PHARMACOLOGICAL ACTIONS : A SHORT REVIEW.” 49(2): 125–31.

Renu, Indu Kumari, Inamul Haque, Manish Kumar, Raju Poddar, Rajib Bandopadhyay, Amit Rai, and Kunal Mukhopadhyay. 2014. “Characterization and Functional Analysis of Eugenol O-Methyltransferase Gene Reveal Metabolite Shifts, Chemotype Specific Differential Expression and Developmental Regulation in *Ocimum* Tenuiflorum L.” *Molecular biology reports*. http://www.ncbi.nlm.nih.gov/pubmed/24420851 (February 11, 2014).

Reuveni, R., Fleischer, A. and Putievsky, E. (1984) Fungistatic activity of essential oils from *Ocimum* basilicumchemotypes. Phytopathologische Zeitschrift 110, 20–22

Sciences, Biological, Mohamed A Nassar, Mohamed U El-segai, and Samah N A Mohamed. 2013. “RTICLES Botanical Studies on *Ocimum* Basilicum L.(Lamiaceae).” 9(5): 150–63.

St Pierre B, De Luca V: Evolution of acyltransferase genes: origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism.In *Recent Advances in Phytochemistry* Vol 34. Evolution of Metabolic Pathways. Edited by Romeo JT, Ibrahim R, Varin L, De Luca V. Elsevier Science Ltd; 2000:285-315

Tzin, V., and Galili, G.(2010). The biosynthetic pathways for shikimate and aromatic amino acids in Arabidopsis thaliana. The Arabidopsis Book 8:e0132. doi:10.1199/tab.0132.

Vogt, Thomas. 2010. “Phenylpropanoid Biosynthesis.” *Molecular plant* 3(1): 2–20. http://www.ncbi.nlm.nih.gov/pubmed/20035037 (March 19, 2014).

Supplementary information

Jumpstart™ *Taq* ReadyMix™ (Sigma-Aldrich)

Composition: 20 mM Tris-HCl, pH 8.3, 100 mM KCl, 3 mM, MgCl2 , 0.002% gelatin, 0.4 mM of each dNTP (dATP, dCTP, dGTP, TTP), stabilizers, 0.1 unit/mL *Taq* DNA Polymerase, JumpStart *Taq* antibody.

S.O.C medium (Invitrogen)

Composition: 2% Tryptone, 0.5% yeast Extract,10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose

Primer sequences

18 Forward: 5'-TCGAAACCTGCAAAGCAGACC-3'

18S Reverse: 5'-GATTCTGCAATTCACACCAAGTATCG-3'

M13 Forward: 5'-GTAAAACGACGGCCAG-3'

M13 Reverse: 5'-CAGGAAACAGCTATGAC-3'

Size of bands in agarose gel:

The size of the bands corresponding to PCR products were confirmed by nucleotide sequence verification. Hence, the sizes of amplicons mentioned on all figures are sequence verified and not just based on comparison with DNA ladder.