



# UNIVERSITA' DEGLI STUDI DI TRIESTE

---

Dottorato di Ricerca in  
**Scienza, Tecnologia ed Economia nell'industria del caffè**

[ Settore scientifico-disciplinare: BIO 18]

- **XX ciclo** -

***Caratterizzazione di geni di *C. arabica* L.  
correlati alle caratteristiche di qualità in tazza  
della bevanda di caffè***

*Dottorando:*  
**Lorenzo Del Terra**

*Coordinatore del Collegio dei Docenti:*  
Chiar.mo Prof. **Tullio Gregori**  
(Università di Trieste)

*Tutore:*  
**Furio Suggi Liverani**  
(illycaffè SpA)

*Relatore:*  
Chiar.mo Prof. **Alberto Pallavicini**  
(Università di Trieste)



## RIASSUNTO

La bevanda di caffè è bevuta ed apprezzata in tutto il mondo per le sue qualità organolettiche. Queste sono numerose e sfaccettate in molti diversi aspetti, che vanno dal contenuto in caffeina alle proprietà antiossidanti. Inoltre, il caffè presenta un ricchissimo profilo aromatico, composto da circa un migliaio di composti volatili che contribuiscono a dare alla bevanda il suo aroma caratteristico.

Questo complesso aroma è fortemente influenzato da numerosi fattori quali l'origine geografica della pianta, la sua varietà botanica e le modalità di lavorazione del frutto e del seme dopo la raccolta. Tutti questi fattori contribuiscono a dare a ciascun tipo di caffè un aroma unico, discriminabile sia dal punto di vista sensoriale che chimico.

Le sostanze che compongono il profilo aromatico della bevanda di caffè sono state studiate sotto molti aspetti, da quello analitico a quello sensoriale. Non è invece stato svolto alcuno studio sull'origine biochimica e biomolecolare di tali composti nella pianta di *Coffea arabica*, nonostante ricerche in tal senso siano già state svolte in altre piante importanti dal punto di vista alimentare ed economico.

Il presente lavoro si è posto l'obiettivo di identificare e caratterizzare alcuni geni della pianta di *Coffea arabica*, che potessero essere correlati alla qualità percepita dal consumatore finale nella tazzina di caffè. I geni che meglio soddisfacevano questi requisiti sono stati individuati nelle monoterpene sintasi. I monoterpene sono infatti tra i più importanti composti odorosi vegetali, responsabili delle note aromatiche positive di tè, caffè, vino, presenti negli olii essenziali di numerose piante aromatiche quali salvia, basilico, menta, e nelle resine di pini ed abeti.

Sono stati quindi disegnati dei primer degenerati sulla base delle sequenze di monoterpene sintasi presenti nei database pubblici. Questi primer hanno consentito di amplificare 3 trascritti di *Coffea arabica*, la cui sequenza completa è poi stata ottenuta mediante la tecnica RACE. Queste putative monoterpene sintasi sono state isolate da vari cDNA sintetizzati da fiore, drupa a vari stadi di maturazione, e seme. L'analisi di queste sequenze mediante BLAST e analisi filogenetica ha permesso di stabilire la loro elevata omologia con geni noti di monoterpene sintasi, confermando quindi il loro status di putative monoterpene sintasi.

Questi trascritti, denominati CaMTS (*Coffea arabica* MonoTerpene Sintasi) sono i primi geni isolati dalla pianta di caffè a poter essere correlati alla qualità percepita in tazza in termini di aroma, percezione olfattiva e gustativa della bevanda, e quindi -in ultima analisi- di apprezzamento edonistico del prodotto. La loro importanza è quindi notevole, anche se la complessità della composizione chimica del caffè suggerisce l'esistenza di una famiglia genica molto numerosa, di cui i 3 geni individuati sono sicuramente solo una piccola parte. Come evidenziato dalle analisi filogenetiche, però, i geni di monoterpene sintasi tendono ad essere molto simili all'interno di una stessa specie, per cui la conoscenza dei primi enzimi di questa famiglia in *Coffea arabica* sarà presupposto fondamentale per la caratterizzazione di ulteriori geni di questo tipo, fatto che porterà gradualmente ad approfondire la conoscenza dei meccanismi che legano la genetica della pianta di caffè alla qualità del prodotto finale.



# INDICE

<b>Riassunto.....</b>	<b>I</b>
<b>Indice.....</b>	<b>III</b>
<b>1. Introduzione .....</b>	<b>1</b>
1.1- La pianta di caffè .....	1
1.2- Dalla pianta al chicco tostato: la filiera di lavorazione .....	3
1.3- Il concetto di “qualità”, in generale e nel mondo del caffè .....	5
1.4- Aspetti sensoriali, fisiologici e genetici della qualità del caffè .....	6
1.4.1 – Sviluppo e maturazione del frutto e del seme di caffè .....	7
1.4.2 – Carboidrati, zuccheri e polisaccaridi .....	9
1.4.3 – Lipidi .....	9
1.4.4 – Acidi clorogenici .....	10
1.4.5 – Alcaloidi: caffeina e trigonellina .....	10
1.4.6 – Proteine e aminoacidi liberi .....	11
1.4.7 – Composti volatili che contribuiscono alla formazione dell’aroma del caffè .....	11
1.4.8 – Espressione genica durante il trattamento post-raccolta .....	13
1.4.9 – Ricerca genomica nel caffè .....	14
1.5- Terpeni e monoterpeni .....	15
1.5.1 – I terpeni .....	15
1.5.2 – Monoterpeni e monoterpene sintasi .....	20
<b>2. Scopo della tesi .....</b>	<b>21</b>
<b>3. Materiali e metodi .....</b>	<b>23</b>
3.1- Estrazioni di RNA totale da <i>C. arabica</i> .....	23
3.1.1 – Estrazione di RNA da fiori e drupe freschi .....	23
3.1.2 – Estrazione di RNA da semi di caffè essiccati: protocollo CTAB .....	24
3.1.3 – Estrazione di RNA da semi di caffè essiccati: protocollo “Hot borate” .....	25
3.1.4 – Quantizzazione dell’RNA estratto .....	27
3.1.5 – RT-PCR .....	28
3.2- Ricerca di geni di <i>C. arabica</i> omologhi a monoterpene sintasi .....	30
3.2.1 – Analisi di similarità e disegno dei primer degenerati .....	30
3.2.2 – PCR con primer degenerati da DNA genomico .....	30
3.2.3 – Clonaggio degli amplificati con kit TOPO-TA Cloning® .....	31
3.2.4 – Amplificazione degli inserti clonati .....	33
3.2.5 – Sequenziamento, assemblaggio ed analisi delle sequenze .....	34
3.2.6 – Disegno dei primer specifici .....	34
3.3- Ottenimento delle sequenze complete .....	35
3.3.1 – Sintesi del cDNA con tecnica SMART .....	35
3.3.2 – PCR da cDNA con primer specifici .....	37
3.3.3 – RACE 3’ e 5’ .....	37
3.3.4 – Nested PCR .....	40
3.3.5 – Sequenziamento ed assemblaggio dei trascritti completi .....	40
3.4- Analisi filogenetica delle monoterpene sintasi di <i>C. arabica</i> .....	41

<b>4. Risultati e discussione .....</b>	<b>43</b>
4.1- Sviluppo di protocolli per l'estrazione di RNA totale da semi di caffè essiccati .....	43
4.1.1 – Protocolli sviluppati: CTAB e “Hot Borate” .....	43
4.1.2 – Qualità dell'RNA estratto .....	44
4.2- Isolamento di 3 putative monoterpene sintasi di <i>C. arabica</i> (CaMTS) .....	46
4.2.1 – Analisi di similarità e disegno dei primer degenerati .....	46
4.2.2 – PCR con primer degenerati da DNA genomico .....	48
4.2.3 – Clonaggio delle PCR ed amplificazione degli inserti .....	49
4.2.4 – Sequenziamento, assemblaggio ed analisi delle sequenze .....	50
4.2.5 – Disegno dei primer specifici .....	58
4.2.6 – PCR da cDNA con primer specifici .....	59
4.2.7 – RACE 3' e 5', nested PCR e clonaggio degli amplificati .....	60
4.2.8 – Sequenziamento, assemblaggio ed analisi dei trascritti “full length” .....	62
4.3- Analisi di espressione dei geni CaMTS nei diversi tessuti .....	68
4.4- Analisi filogenetica dei geni CaMTS .....	69
<b>5. Conclusioni .....</b>	<b>75</b>
<b>6. Bibliografia .....</b>	<b>79</b>

# 1. INTRODUZIONE

## 1.1 - La pianta di caffè

Il genere *Coffea* appartiene dal punto di vista tassonomico alla sottoclasse *Asterids*, ordine *Gentianales*, e alla famiglia delle *Rubiaceae*. In natura comprende approssimativamente un centinaio di specie, tutte originarie dell'Africa tropicale e subtropicale. La produzione di caffè a fini commerciali si basa su due sole di queste specie: *C. arabica* L. e *C. canephora* Pierre ex Froehn, comunemente note come caffè Arabica e Robusta. Una terza specie, *C. liberica* Bull ex Hiern (commercializzata come caffè liberica o excelsa), viene coltivata in modo estremamente limitato e contribuisce a meno dell'1% della produzione mondiale.

Le popolazioni selvatiche di *C. arabica* si rinvencono sugli altipiani dell'Etiopia sud-occidentale, ad altitudini comprese tra i 1000 e i 2000m slm, con precipitazioni medie annue di 1500-2500mm, concentrate in una stagione piovosa ben definita. *C. canephora*, invece, cresce spontaneamente nelle foreste equatoriali dalla Guinea all'Uganda, ad altitudini molto inferiori (fino al livello del mare) con temperature ed umidità decisamente più elevate.

Nonostante le due specie di caffè siano state ampiamente diffuse nel mondo a scopi commerciali, in regioni estremamente varie e distanti dall'Africa centrale, la loro distribuzione rispecchia fedelmente questo andamento climatico. *C. arabica* risulta quindi diffusa nella fascia subtropicale, ad altitudini più elevate, temperature inferiori e precipitazioni più concentrate in una stagione specifica. Le regioni che meglio rispondono a queste caratteristiche, e che sono difatti famose per produrre l'Arabica di miglior qualità, sono le aree montuose di Brasile (il maggior produttore mondiale), Colombia, CentroAmerica ed India, oltre naturalmente alla sua zona d'origine, gli altipiani dell'Africa centro-orientale (Etiopia, Kenya).

Il caffè Robusta è invece coltivato soprattutto nel suo areale nativo, l'Africa centro-occidentale (Uganda, Costa d'Avorio), nelle regioni a bassa quota di Brasile e India (che sono tra i principali produttori anche di questa specie), e nel Sud-Est asiatico, principalmente in Indonesia e Vietnam. In questi ultimi due paesi la coltura di *C. canephora* è stata introdotta in tempi relativamente recenti (dagli anni '90 in poi) e attuata in maniera sistematica e intensiva, portando in breve questi paesi ai vertici in termini di volumi prodotti; il Vietnam in particolare è ormai il secondo produttore mondiale, dopo aver sopravanzato la Colombia che storicamente occupava questa posizione alle spalle del Brasile.

La Tab. 1.1 riporta i dati ufficiali dell'ICO (International Coffee Organization) relativi alla produzione di caffè verde nel 2007, nella quale si ritrovano i principali paesi produttori di Arabica e Robusta, descritti pocanzi. La tabella evidenzia come Arabica costituisca attualmente i due terzi della produzione mondiale, con il terzo restante rappresentato da Robusta.

Paese	Arabica	Robusta	Totale	%
Brasile	30	4	34	30
Vietnam	0	16	16	14
Colombia	12,4	0	12,4	11
Indonesia	1,4	5,6	7	6
Etiopia	5,7	0	5,7	5
India	1,6	3,2	4,8	4
Messico	4,3	0	4,3	4
Guatemala	4	0	4	3
Honduras	3,5	0	3,5	3
Perù	3	0	3	3
Uganda	0	2,8	2,8	2
Cote d'Ivoire	0	2,4	2,4	2
Costa Rica	1,9	0	1,9	2
Nicaragua	1,8	0	1,8	2
El Salvador	1,5	0	1,5	1
Ecuador	1	0	1	1
Papua Nuova Guinea	1	0	1	1
Kenya	0,9	0	0,9	1
Tailandia	0	0,9	0,9	1
Venezuela	0,9	0	0,9	1
Camerun	0	0,8	0,8	1
Tanzania	0,8	0	0,8	1
Filippine	0	0,7	0,7	1
Altri	1,7	1,2	2,9	3
<b>TOTALE</b>	<b>77,4</b>	<b>37,6</b>	<b>115</b>	<b>100</b>

% Arabica	67%
% Robusta	33%

Tab. 1.1 – Produzione mondiale di caffè verde nel 2007, espressa in milioni di sacchi da 60kg  
(Elaborazione dell'autore da dati ICO 2007)

Uno dei motivi principali della diffusione della coltura di *C. canephora* è la sua adattabilità a diverse condizioni crescita, la sua elevata produttività, e soprattutto –da qui il nome commerciale Robusta– la sua resistenza a molte delle malattie che affliggono invece le piantagioni di *C. arabica*, in primo luogo la ruggine delle foglie, patologia dovuta all'infezione del fungo *Hemileia vastatrix* (Clarke RJ e MacRae R, 1988).

La pianta di caffè (Illy e Viani, 2005) è un arbusto sempreverde che può raggiungere i 4-6 m di altezza in Arabica, e gli 8-12 m in Robusta, anche se nelle piantagioni le piante sono potate ad altezze di 2-3 metri per agevolare la raccolta dei frutti, quasi sempre effettuata a mano. I fiori si schiudono quando alla stagione secca seguono le prime piogge; 8-10 giorni dopo le precipitazioni, i fiori si aprono e avviene l'impollinazione, che in *C. arabica* è autogama (organismo autofertile, i fiori sono per il 90% autoimpollinati), mentre in *C. canephora* è eterogama, in quanto questa specie possiede un sistema gametofitico di auto-incompatibilità che impedisce al polline di un fiore di fecondare l'ovario di fiori dello stesso individuo (Lashermes *et al*, 1996).

Una volta avvenuta la fecondazione, lo sviluppo del frutto richiede un tempo estremamente lungo, 7-9 mesi per Arabica, 9-11 mesi per Robusta. Al termine di questo periodo il frutto maturo è costituito da una drupa di colore rosso acceso (giallo in alcune varietà) comunemente chiamata "ciliegia", contenente un sottile strato di polpa, uno strato più lasso e idratato detto "mucillagine", e due semi appaiati ciascuno racchiuso da un involucro denominato "pergamino".

Anche la composizione chimica dei semi delle due specie è diversa: le principali differenze sono:

- il contenuto in caffeina: circa l'1.2% del peso di un seme di Arabica è costituito da caffeina, contro il 2.5% di Robusta, che ne contiene quindi circa il doppio.



- il contenuto in lipidi: Arabica è più ricco in lipidi totali di Robusta (16% contro 10%, rispettivamente).
- il contenuto dei diterpeni cafestolo e kahweolo, che sono tipici del genere *Coffea* e non si ritrovano in altre specie vegetali; Arabica li contiene entrambi, mentre in Robusta il kahweolo è assente. Il rapporto tra questi composti viene quindi utilizzato come indicatore nella composizione delle miscele.

La pianta di *C. canephora* è diploide ( $2n=22$  cromosomi), così come tutte le altre specie conosciute del genere *Coffea*, con l'eccezione di *C. arabica* che è allotetraploide ( $2n=44$ ) e si è probabilmente originata da un'incrocio interspecifico spontaneo seguito da duplicazione. Gli studi più recenti hanno indicato che molto probabilmente i progenitori di *C. arabica* sono stati *C. eugenoides* e *C. canephora* (Raina *et al.*, 1998, Lashermes *et al.*, 1999).

*C. arabica* e *C. canephora* sono diverse non solo nelle caratteristiche genetiche, agronomiche e chimiche, ma anche e soprattutto nella qualità della bevanda ricavata dai rispettivi semi. La tazzina di caffè Arabica è più acida e dolce, con note aromatiche più piacevoli (tipicamente pantostato, caramello, cioccolato e sentori fruttati e floreali). Il caffè Robusta invece produce una bevanda più amara ed astringente, con sentori terrosi, legnosi e medicinali. Questo fa sì che dal punto di vista commerciale Arabica sia il caffè più pregiato e costoso, mentre Robusta viene normalmente utilizzato nella composizione di miscele e per la produzione di caffè solubile.

## 1.2 - Dalla pianta al chicco tostato: la filiera di lavorazione

Per comprendere alcuni aspetti inerenti alla qualità del caffè verde, è indispensabile fare prima alcuni cenni alla lavorazione che viene eseguita per preparare il chicco di caffè alla tostatura. L'obiettivo del trattamento post-raccolta è innanzitutto eliminare la buccia, la polpa e la mucillagine, tessuti ricchi di acqua e zuccheri, che dopo il distacco del frutto dalla pianta, iniziano in brevissimo tempo a fermentare. Questo è un fenomeno da evitare assolutamente, in quanto la fermentazione è la causa dei peggiori difetti nell'aroma del caffè, con la produzione di composti volatili estremamente spiacevoli all'olfatto responsabili dei difetti classificati come "fermentato" e "stinker" (Illy e Viani, 2005).

Il secondo obiettivo della lavorazione del caffè verde è l'eliminazione della maggior parte dell'acqua dal chicco. Infatti, una volta essiccato fino a un contenuto in acqua residuo del 10-12%, il seme può essere conservato in magazzino per mesi o anni senza perdere le sue caratteristiche qualitative. L'essiccazione inoltre rallenta o impedisce lo sviluppo di muffe, responsabili della contaminazione dei lotti di caffè con tossine estremamente pericolose per la salute, in particolare l'ocratossina A, sintetizzata dal fungo *Aspergillus ochraceus*.

Questi obiettivi vengono raggiunti mediante due metodi principali: la lavorazione per via secca (caffè "naturale", anche noto come dry-processed coffee, sundried coffee) e quella per via umida (caffè "lavato", altrimenti detto washed coffee, wet-processed coffee). Questi metodi sono schematizzati in Fig. 1.2.



Fig. 1.2 – rappresentazione schematica dei passaggi di lavorazione del caffè verde.

Il metodo “naturale” è il più diretto e semplice, e consiste nello stendere ad essiccare al sole le ciliegie di caffè subito dopo la raccolta, in grandi patii di terra battuta, mattoni o cemento. L’asciugatura può durare dalle 2 settimane fino ad un mese, e richiede una costante manodopera per rimescolare lo strato di caffè durante la giornata, e coprirlo rapidamente in caso di pioggia. Al termine dell’essiccazione, i frutti vengono privati meccanicamente della scorza e del pergamino, liberando così i semi.

Con il metodo “lavato” invece scorza, polpa e mucillagine vengono rimosse prima dell’asciugatura dei semi. I frutti freschi vengono prima spolpati meccanicamente, poi le bucce e la polpa vengono separate dai semi, che vengono lasciati fermentare per 12-24 ore in vasche piene d’acqua. Questa fermentazione serve a rimuovere lo strato di mucillagine, che è particolarmente viscoso e non viene rimosso dai passaggi precedenti. Una volta terminata la fermentazione i semi vengono lavati per rimuovere la mucillagine in sospensione, e possono infine venir asciugati in patio per 6-10 giorni, o in asciugatori meccanici ad aria calda in 24-48 ore. Il prodotto finale di questo procedimento è il chicco di caffè essiccato e ancora racchiuso nel pergamino che in genere non viene rimosso, in quanto sembra che il seme in pergamino mantenga più a lungo la sua qualità durante lo stoccaggio in magazzino.

Entrambi i metodi presentano vantaggi e svantaggi. Il metodo naturale richiede meno infrastrutture e macchinari industriali, ed è quindi più economico e praticabile anche da piccoli produttori o da fattorie a conduzione familiare. Per contro, l’asciugatura del caffè è estremamente lenta ed avviene col seme ancora racchiuso da polpa e mucillagine, che sono le parti del frutto più idratate e ricche di zuccheri. Il prodotto è quindi esposto a tutti i rischi da ciò derivanti: innanzitutto la fermentazione che può facilmente rovinare la qualità del caffè verde; in secondo luogo la pioggia, verso cui questo metodo ha ben poche

difese, che innesca nei frutti cicli di essiccazione/reimbibizione che comportano ancora fermentazione, sviluppo di muffe ed una generale perdita di qualità.

Il metodo umido invece elimina subito polpa e mucillagine, riducendo così il rischio di fermentazione dei semi ed accorciando di molto i tempi di asciugatura, in quanto per prima cosa viene rimossa la parte più idratata del frutto. Questo metodo è quindi superiore dal punto di vista qualitativo, ed infatti è impiegato dalle nazioni che tradizionalmente producono caffè di alta qualità, principalmente Colombia, Centroamerica ed Etiopia. Gli svantaggi di questa tecnica sono da individuarsi nei costi e nella complessità tecnica. Per applicare questa lavorazione infatti è necessario un vero e proprio impianto industriale: ogni passaggio richiede macchinari dedicati, ed il processo nel suo insieme consuma enormi quantità di energia e di acqua per la movimentazione, fermentazione e lavaggio della massa di caffè; acqua che poi viene in genere scaricata direttamente nei corsi d'acqua, creando così anche gravi problemi di inquinamento e ecosostenibilità. Il procedimento umido viene quindi generalmente applicato da grandi impianti di lavorazione posseduti dai maggiori coltivatori, oppure gestiti da cooperative che raccolgono e lavorano il caffè proveniente da molte piantagioni circostanti.

Quale che sia il metodo utilizzato, il caffè verde viene poi selezionato per eliminare i principali difetti (chicchi rotti, tarlati, scoloriti o neri, frammenti di scorze o pergamino, materiali estranei come foglie, rami, pietre, terra). La selezione può essere fatta a mano o mediante macchine selezionatrici ottiche che leggono forma, dimensioni e colore dei chicchi, eliminando quelli che non rientrano nei parametri di accettabilità.

Al termine della selezione, il caffè verde è pronto per essere insaccato nei tradizionali sacchi di juta (tuttora la soluzione più usata), o in grandi "bulk bag" di plastica, utilizzati dai grandi importatori, o anche in sacchi di plastica confezionati sottovuoto, soluzione moderna che viene utilizzata soprattutto dai tostatori alla ricerca della massima qualità. Il caffè così insaccato viene quindi caricato nei container ed inviato via nave ai magazzini di tutto il mondo, prima di venir tostato, macinato e confezionato in vista dell'utilizzo finale.

### 1.3 - Il concetto di "qualità", in generale e nel mondo del caffè

E' difficile dare una definizione univoca di "qualità", in quanto questo termine è usato —e spesso abusato— nei contesti più disparati. La definizione ufficiale adottata dall'ISO (International Organization for Standardization) è:

*"Il livello a cui un insieme di caratteristiche intrinseche (fisiche, sensoriali, comportamentali, temporali, ergonomiche, funzionali, etc.) soddisfa le richieste, ove richieste significa necessità o aspettative che possono essere esplicite, ma generalmente implicite."* (ISO, 2000)

Altre definizioni possono essere "conformità alle richieste", "essere adatto all'utilizzo", "la soddisfazione delle aspettative del cliente, ad un costo che rappresenta per lui un valore", e molte altre.

Queste definizioni generiche devono ovviamente essere adattate al campo specifico di applicazione. Nel caso del caffè, la multidisciplinarietà dell'oggetto di studio impone

una corrispondente molteplicità delle definizioni di “qualità” ad esso applicabili. Infatti le discipline che ruotano intorno al mondo del caffè spaziano dall’agronomia, all’agricoltura, alla botanica, alla genetica, all’ingegneria, all’economia, alla sociopolitica, all’analisi sensoriale, per elencare solo le principali.

Lo stesso termine “caffè” cambia in realtà significato a seconda dell’utente che ne fa uso: con questa parola, un produttore si riferirà ad una pianta, un importatore a dei sacchi di juta pieni di chicchi essiccati, un torrefattore agli stessi chicchi tostati e macinati, un consumatore invece penserà come prima cosa a una bevanda scura che borbotta nella caffettiera, o alla pausa al bar con i colleghi.

Naturalmente, impieghi ed esigenze così diverse richiedono diverse definizioni di “qualità”. Per un agricoltore, le principali qualità di una pianta di caffè saranno la facilità di coltivazione, il vigore della pianta, la necessità di acqua e fertilizzanti, la produttività e la resistenza alle malattie, oltre naturalmente al ricavato ottenibile dalla vendita di quella varietà di caffè.

Un commerciante di caffè verde sarà interessato invece alla qualità organolettica, al contenuto in acqua, e alla possibilità di tracciare le origini del prodotto.

Un’azienda di decaffeinizzazione andrà a valutare naturalmente il contenuto in caffeina (in larga misura dipendente dalla specie, arabica o robusta).

Un’industria di caffè solubile sarà interessata alla frazione di solidi solubili ed insolubili, in particolare alla composizione in carboidrati, che varia da specie a specie ed influenza la resa e la qualità del processo industriale.

Un impianto di torrefazione deve tener conto dei parametri che possono influenzare la qualità della tostatura: la massa e umidità iniziale dei chicchi, il calore specifico e la conduttività termica del caffè, le dimensioni dei chicchi. Importante è inoltre l’omogeneità granulometrica del lotto e l’eventuale presenza di difetti quali chicchi rotti o frammenti di pergamino: durante la tostatura questi contaminanti di piccole dimensioni bruciano, dando all’intero lotto uno sgradevole aroma affumicato.

Infine, per il consumatore finale, la qualità sarà data fondamentalmente dalla percezione sensoriale, gustativa ed olfattiva, della bevanda che gli viene servita al bar o che si prepara lui stesso a casa. Difatti, a differenza di altri prodotti alimentari la cui qualità è data dall’unione della componente sensoriale con quella nutrizionale, il caffè non possiede qualità nutritive degne di nota, essendo il suo contributo calorico pressochè nullo e il suo apporto di nutrienti trascurabile. Ciò che il consumatore ricerca nell’assunzione di caffè è quindi solamente l’aspetto edonistico, dato dall’effetto stimolante della caffeina e dalla gratificazione sensoriale dovuta al ricco profilo aromatico della bevanda.

## **1.4 - Aspetti sensoriali, fisiologici e genetici della qualità del caffè**

Alla luce di quanto illustrato sinora, diventa chiaro che la qualità della bevanda di caffè è solo l’ultimo anello di una catena di lavorazione molto lunga e, dal punto di vista del consumatore finale, è data fondamentalmente dalla composizione chimica della bevanda. Tale composizione è assai varia e complessa (Clarke e Vitzthum, 2001): basti pensare che nel caffè tostato sono state isolate alcune migliaia di specie chimiche, comprendenti carboidrati ad alto e basso peso molecolare, acidi organici, lipidi, aminoacidi e proteine, oltre ad una grande varietà di composti volatili (terpeni, furani, pirazine, fenoli, tioli) che sono i principali responsabili dell’aroma. Questa enorme varietà viene ulteriormente ampliata dal processo di tostatura, le cui alte temperature (circa 200-220°C)

innescano svariate reazioni, in particolare la degradazione termica di molti composti, e la reazione di Maillard in cui mono- e disaccaridi si combinano con molecole che espongono un gruppo aminico libero, come peptidi e aminoacidi liberi. Il prodotto della reazione di Maillard sono le cosiddette melanoidine, una classe di sostanze molto ampia, che include molti composti volatili che vanno a contribuire all'aroma del caffè, e sono inoltre le sostanze che conferiscono al chicco tostato il caratteristico colore marrone.

E' quindi estremamente difficile cercare di correlare questa complessità finale con il primo anello della catena, ovvero la pianta di caffè e gli eventi biochimici e biomolecolari che avvengono durante la maturazione del frutto e del seme.

E' probabilmente per questa ragione che, mentre altri aspetti qualitativi sono ampiamente studiati da lungo tempo, è invece appena agli inizi la comprensione dei fenomeni metabolici e biosintetici che avvengono nel frutto e nella lavorazione del caffè verde.

Il caffè verde è un oggetto di studio di particolare interesse per svariati motivi. Innanzitutto è la *commodity* su cui si basa l'intero commercio mondiale del caffè. Da ciò consegue che, mentre lo studio degli aspetti biologici, agronomici e biotecnologici della pianta di *Coffea* riveste un'enorme importanza per i paesi produttori, i paesi importatori sono più interessati al caffè verde, che dal loro punto di vista è la materia prima per i successivi processi industriali.

In secondo luogo, se consideriamo che con la tostatura e macinazione il seme viene ucciso e distrutto meccanicamente, il caffè verde è il materiale *biologico* più prossimo alla lavorazione industriale, e quindi quello più strettamente legato alla qualità del prodotto finale, inteso come la tazzina di caffè bevuta dal consumatore.

Di seguito vengono brevemente passate in rassegna le principali caratteristiche biologiche, chimiche e molecolari del caffè verde, focalizzandosi in particolare sulle loro correlazioni con la qualità della bevanda.

#### ***1.4.1 – Sviluppo e maturazione del frutto e del seme di caffè***

La drupa di caffè (Fig. 1.3) è costituita da un esocarpo (la scorza esterna), un mesocarpo che ne costituisce la polpa, e un endocarpo lignificato che racchiude i semi ed è denominato "pergamino". Il seme è a sua volta costituito dall'embrione che andrà a formare la nuova pianta, che ha dimensioni estremamente ridotte in proporzione al resto del seme; la maggior parte del volume del chicco, infatti, è costituita da endosperma, un tessuto di riserva che si sviluppa gradualmente nel corso della maturazione fino ad occupare la quasi totalità dello spazio disponibile nel loculo. Il sottile strato che separa l'endosperma dal pergamino è ciò che rimane del perisperma, il tessuto di origine materna che originariamente riempiva l'intera cavità, e di cui a maturazione completata non rimane che una sottile membrana, denominata "silver skin" o pellicola argentea.

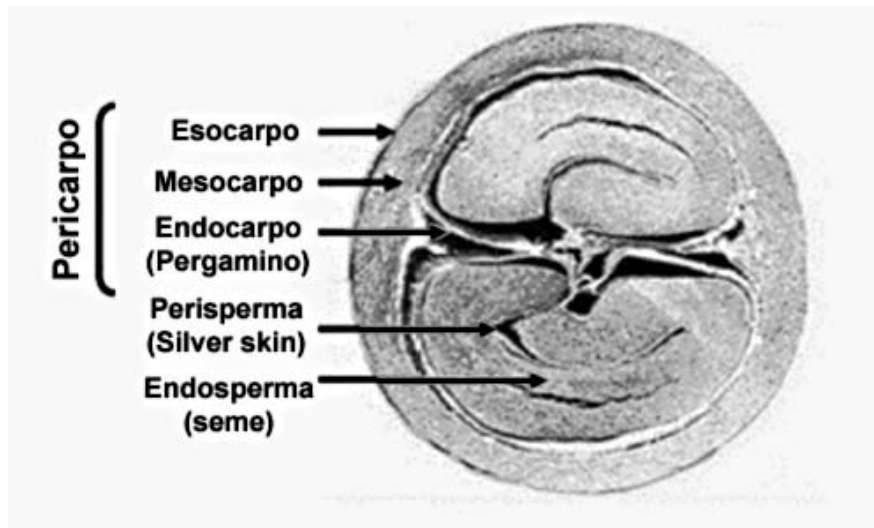


Fig. 1.3 – rappresentazione schematica dei tessuti in una drupa di caffè matura (adattata da De Castro e Marraccini, 2006)

Queste diverse parti del frutto e del seme si sviluppano in momenti diversi durante la maturazione (Fig. 1.4). In una prima fase, fino alla 10ma settimana circa dopo la fioritura, c'è un rapido aumento di volume del seme, che fino a questo momento è costituito in gran parte da perisperma. La crescita del perisperma si arresta verso la 11-12ma settimana, quando l'endosperma inizia a svilupparsi riassorbendo il perisperma fino alla 24ma settimana circa, momento in cui la sua crescita è completata. A questo punto le dimensioni del frutto e del seme sono pressochè quelle definitive, e le settimane successive sono dedicate alla sintesi ed all'accumulo delle sostanze di riserva del seme.

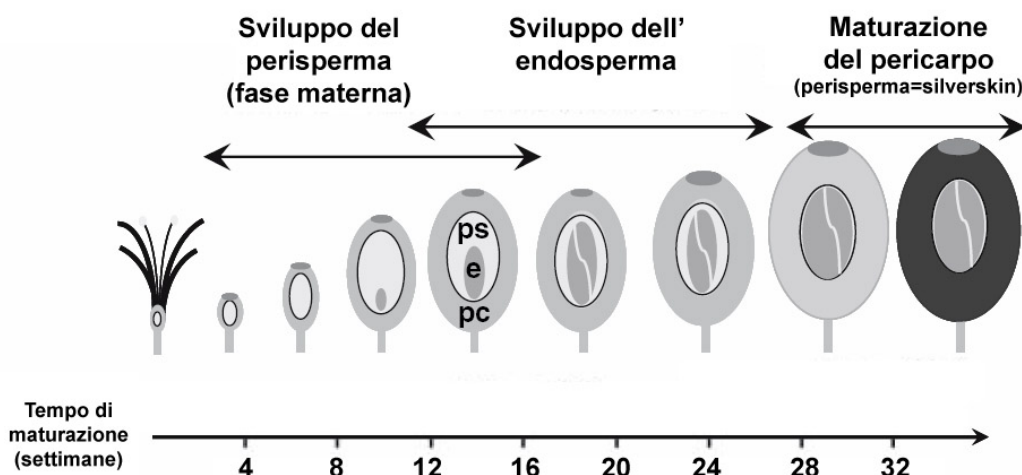


Fig. 1.4 – cambiamenti tissutali nel corso della maturazione della drupa di caffè (adattata da Salmona *et al.*, 2008)

### 1.4.2 – Carboidrati, zuccheri e polisaccaridi

Circa il 50% del peso secco del seme di caffè è formato da carboidrati, in gran parte (43-45%) galattomannani, arabinogalattani e cellulosa, i polisaccaridi costituenti la parete cellulare (Fischer *et al*, 2001). Gli arabinogalattani nel caffè sono una popolazione estremamente eterogenea, a causa della varietà nel grado di ramificazione, e nella composizione monosaccaridica delle catene laterali; inoltre sono covalentemente legati a proteine, apparentemente con funzioni strutturali nella parete cellulare di cui costituiscono il 25% circa (Redgwell *et al*, 2002a). Questi composti presentano una bassa solubilità nel caffè verde, che aumenta significativamente con la tostatura. Nel caffè torrefatto inoltre si riscontra una perdita di carboidrati tra il 10 e il 40%, a seconda del grado di tostatura. Questo è probabilmente dovuto a degradazione termica delle catene polisaccaridiche, che rilasciano monosaccaridi che a loro volta vengono coinvolti in reazioni di Maillard. La rottura delle catene polisaccaridiche spiegherebbe anche l'aumento di solubilità, parametro estremamente importante nella produzione del caffè solubile (Redgwell *et al* 2002b; Oosterveld *et al* 2003).

In aggiunta ai polisaccaridi, in Arabica circa l'8% del peso secco del seme è costituito da saccarosio (in Robusta la percentuale è minore, circa il 4%), di cui solo tracce rimangono dopo la tostatura, a causa del suo coinvolgimento nelle reazioni di Maillard e della sua degradazione per pirolisi (caramellizzazione).

La saccarosio sintetasi di *C. canephora* è stata studiata da Leroy *et al* (2005) facendo uso di una libreria BAC di Robusta, la prima costruita al mondo.

### 1.4.3 – Lipidi

I lipidi sono una componente importante nel caffè verde, costituendo il 15-16% del peso in un seme di Arabica, e circa il 10% in uno di Robusta. Tra i lipidi presentano particolare interesse i diterpeni cafestolo e kahweolo, che rappresentano circa il 10% dell'olio di caffè, e sono esclusivi del genere *Coffea* (Speer e Kolling-Speer, 2001). La funzione di queste molecole nella fisiologia del seme è tuttora ignota, ma nell'uomo è stata riportata una loro influenza sulla sintesi del colesterolo (Urgert *et al*, 1997) e una loro azione protettiva nei confronti di alcuni tipi di tumore (Cavin *et al*, 1998).

Il cafestolo si ritrova sia in Arabica che in Robusta, mentre il kahweolo è presente soltanto nei tessuti di *C. arabica*, in particolare nell'endosperma. Nei semi di *C. canephora* è stato invece isolato un altro diterpene, il 16-O-metilcafestolo, che non viene prodotto da Arabica. Questo composto non viene degradato in tostatura, e può essere quindi utilizzato come marcatore per rilevare la presenza di Robusta in miscele di Arabica (Speer *et al*, 1991).

Dal punto di vista genetico, è possibile ipotizzare l'esistenza di una "kahweolo sintasi" espressa solo nei semi di *C. arabica*, e di una ipotetica "cafestolo metilasi" presente in *C. canephora* (De Castro e Marraccini, 2006).

#### 1.4.4 – Acidi clorogenici

Gli acidi clorogenici sono una famiglia di acidi fenolici esterificati nelle posizioni 3, 4 e 5 dell'acido chinico, per formare i vari acidi caffeoil-, p-coumaroil- e feruloil-quinico (Clifford, 1999). Questi composti ammontano a circa il 10% del peso del caffè verde Arabica, mentre in Robusta la loro quantità è di circa il 4%.

E' stata osservata una regolazione temporale della sintesi di acidi clorogenici nel corso dello sviluppo del seme, in particolare durante l'accrescimento dell'endosperma si rileva un notevole accumulo di 3-, 4- e 5-CQA (acido caffeoilquinico). Questa sintesi sembra poi diminuire drasticamente alla fine della maturazione.

Sembra che gli acidi clorogenici, e i loro derivati prodotti durante la tostatura, siano tra i principali antiossidanti contenuti nel caffè, e nei consumatori abituali svolgono quindi un'importante azione protettiva contro i radicali liberi (Natella *et al* 2002).

#### 1.4.5 – Alcaloidi: caffeina e trigonellina

La caffeina è probabilmente l'alcaloide vegetale più studiato al mondo. La quasi totalità degli studi, però, riguardano i suoi effetti fisiologici sull'uomo, che sono effettivamente ben documentati e vanno ad influenzare una molteplicità di apparati: nervoso, respiratorio, cardiocircolatorio, urinario etc.

Invece la letteratura sul ruolo della caffeina nella fisiologia della pianta di caffè non è particolarmente ampia, anche se si ipotizza un suo ruolo tossico nella difesa da parassiti e patogeni; l'elevata quantità di acidi clorogenici presenti nel chicco sembra abbia la funzione di complessare le molecole di caffeina, impedendone la diffusione attraverso le membrane e limitandone così l'autotossicità (Mosli-Waldhauser e Baumann, 1996).

Come già accennato, Robusta ha un contenuto in caffeina del 2.2-2.4%, all'incirca doppio rispetto ad Arabica che si attesta sull'1.2-1.3%. Esistono in natura alcune specie del genere *Coffea* che producono livelli estremamente bassi, o addirittura nulli, di caffeina, ad esempio *Coffea pseudozanguebariae*.

La pianta sintetizza caffeina in molti tessuti, inclusa la foglia e il fiore. Nel frutto in maturazione, la caffeina viene sintetizzata nel perisperma e nel pericarpo (Sondahl e Baumann, 2001), anche se vi sono indizi di una sua sintesi anche nell'endosperma in formazione. L'incremento maggiore del contenuto in caffeina nei semi si verifica tra l'8° e la 16ma settimana di sviluppo, in corrispondenza della fase di espansione dell'endosperma. Dopo questo periodo, il contenuto di caffeina nel pericarpo rimane costante, indicando un arresto della sua sintesi, o un suo attivo trasferimento nel seme mano a mano che viene sintetizzata.

La caffeina (1,3,7-trimetilxantina) viene sintetizzata per metilazioni successive della xantosina, una base purinica. Le dimetilxantine teobromina, teofillina e paraxantina costituiscono intermedi di reazione e prodotti di degradazione della via metabolica della caffeina, e infatti si ritrovano anch'esse nel seme di *Coffea*. La prima caffeina sintasi è stata isolata in *Camellia sinensis*. Sono poi stati trovati i corrispondenti geni che operano le tre metilazioni successive nel caffè (Crozier *et al* 1997; Misako e Kouichi 2004).

Un altro alcaloide presente in quantità non trascurabili nel caffè (0.6-1.3% in Arabica, 0.3-0.9% in Robusta) è la trigonellina, derivata dalla metilazione dell'acido



nicotinico. La sua funzione fisiologica nella pianta non è ancora stata chiarita, ma è un composto importante dal punto di vista qualitativo, in quanto durante la tostatura partecipa a reazioni di Maillard per dare origine a vari composti aromatici quali alchilpirazine e pirroli.

#### **1.4.6 – Proteine e aminoacidi liberi**

Il caffè verde ha un contenuto di circa il 10% in proteine, di cui all'incirca l'1% costituito da aminoacidi liberi, che scompaiono con la tostatura e come già detto sono importanti come reagenti nelle reazioni di Maillard.

Il 45% delle proteine totali è rappresentato da proteine di riserva 11S, caratterizzate da Rogers *et al* (1999), che ne ha osservato l'espressione specifica nel periodo di espansione dell'endosperma.

#### **1.4.7 – Composti volatili che contribuiscono alla formazione dell'aroma del caffè**

Gli organismi vegetali sono dei potentissimi laboratori di sintesi chimica, che producono una quantità e varietà incommensurabile di specie chimiche. Tra queste, le più affascinanti e misteriose sono probabilmente i composti volatili. Ne sono state isolate diverse migliaia, sintetizzate da ogni parte della pianta, dal fiore, alle foglie, al fusto, addirittura alla radice. In generale queste sostanze sono lipofile con massa molecolare inferiore a 300 (Dudareva *et al* 2004). In ordine di massa decrescente, le classi principali sono rappresentate da: terpenoidi, derivati degli acidi grassi (inclusi i prodotti della via metabolica delle lipoossigenasi), benzenoidi, fenilpropanoidi, composti ramificati a 5 atomi di carbonio, e vari composti azotati e solforati.

Una tale varietà di composti deve sicuramente assolvere una quantità di usi, di cui molti sono ancora sconosciuti. Certamente una delle funzioni più importanti, specialmente per i composti responsabili per la profumazione dei fiori, è l'attrazione degli impollinatori (Dudareva *et al* 2000; Shalit *et al* 2003).

Un'altra funzione dimostrata è la comunicazione e segnalazione intra- e intercellulare (ormoni vegetali). Inoltre, molte di queste sostanze hanno proprietà antimicrobiche ed anti-erbivore, suggerendo una loro funzione tossica nel difendere le parti della pianta più importanti ed appetibili.

Questa funzione è suffragata da vari elementi: innanzitutto le parti meristematiche (gemme apicali, foglie giovani) e riproduttive (fiori, frutto) della pianta sono in genere quelle che presentano le massime concentrazioni di queste molecole. In secondo luogo, è stato dimostrato che il profilo dei volatili emessi da una pianta è dinamico, e muta sia in composizione che in intensità a seconda delle condizioni esterne; in particolare, dopo l'attacco di erbivori tutta la pianta (non solo la parte predata) risponde con un massiccio aumento nella produzione di composti volatili, sintetizzando sia composti deterrenti per l'erbivoro, sia composti che attraggono i suoi predatori naturali (Pichersky *et al* 2002; Paré e Tumlinson 1997).

In generale, l'evoluzione della sintesi e della regolazione di questi metaboliti secondari sembra essere regolata da eventi di duplicazione genica e cambiamenti nel livello di regolazione dell'espressione genica (Dudareva *et al* 1996; Pichersky *et al* 2000;

van der Hoeven *et al* 2000). Questi eventi evolutivi sono piuttosto comuni nel regno vegetale, il che contribuisce a spiegare l'enorme varietà di queste sostanze e dei geni preposti alla loro sintesi (cfr. anche § 4.4 e § 5).

Il caffè verde contiene circa 300 composti volatili (Bonnelaender 2005; Flament 2001). Alcuni vengono degradati dalla tostatura, altri rimangono stabili. Nella maggior parte dei casi, però, la tostatura fa aumentare decisamente la concentrazione di queste sostanze. Non solo, le reazioni di idrolisi, pirolisi e Maillard che avvengono durante la torrefazione portano alla creazione di oltre 650 nuove specie chimiche volatili, portando il totale a un migliaio circa.

Lo studio di queste sostanze si basa su procedure di estrazione differenziate per poter ottenere un estratto il più possibile rappresentativo della grande complessità chimica della miscela studiata: esempi di queste tecniche sono SDE (Simultaneous Distillation Extraction), HVD (High Vacuum Distillation), SAFE (Solvent-Assisted Flavour Extraction), e SPME (Solid-Phase Micro Extraction), forse la più rapida e versatile attualmente a disposizione.

Le molecole estratte vengono poi studiate mediante GC-olfattometria, una tecnica che combina l'analisi gas-massa della miscela con lo "sniffing" delle sostanze man mano che vengono separate e analizzate dal gascromatografo: in questo modo è possibile determinare quali composti sono olfattivamente attivi, e la loro contribuzione relativa alla formazione dell'aroma complessivo.

Questa si calcola mediante la tecnica della flavour dilution (FD): diluendo progressivamente l'estratto aromatico, si annota fino a che diluizione l'odore è ancora percepibile. I composti che contribuiscono maggiormente all'aroma complessivo sono quelli percepibili alle maggiori diluizioni. Composti come il  $\beta$ -damascenone, il guaiacolo, molte pirazine e vari composti solforati, hanno valori FD di 1000 o più.

Fare un elenco completo di tutte le sostanze che compongono l'aroma di caffè sarebbe impossibile in questa sede, si riportano solo le categorie principali di questi composti e la loro modalità di formazione.

- le aldeidi e gli acidi volatili si formano principalmente per pirolisi e caramellizzazione dei carboidrati (Yeretian *et al* 2002);
- Le proteine e gli aminoacidi liberi reagiscono con i carboidrati per dare i prodotti delle reazioni di Maillard;
- Gli acidi clorogenici si degradano a fenoli;
- Le trigonelline si degradano tramite reazioni di Maillard in alchilpirazine e pirroli.
- I composti con i più alti valori FD sono: MMBF (3-mercapto-3-metilbutilformiato), 2-furfuriltiolo, metionale,  $\beta$ -damascenone, alcune pirazine e furanoni.
- Alcuni composti terpenici, in particolare svariati monoterpeni (linalolo, limonene, geraniolo,  $\alpha$ -terpineolo) sono i composti responsabili di alcune delle principali note positive dell'aroma di caffè, in particolare dei sentori floreali, agrumati e fruttati. Questi composti sono presenti nel seme sia in forma libera, sia immagazzinati in forma glicosilata. Questi ultimi vengono liberati dalla componente glicosidica durante la tostatura, mentre nella fisiologia del seme è lecito attendersi la presenza di specifiche glucosidasi che idrolizzano questo legame rilasciando gli aromi, come è stato dimostrato nel caso del tè (Mizutani *et al*, 2002).

#### 1.4.8 – *Espressione genica durante il trattamento post-raccolta*

L'espressione genica nel corso dello sviluppo del frutto è stata studiata in alcuni singoli geni di interesse, come accennato nei precedenti paragrafi. Di recente, il nascente interesse per la genomica del caffè ha portato alla costruzione di alcune decine di migliaia di EST sia da *C. arabica* che da *C. canephora*. Questa crescente massa di informazioni ha consentito i primi studi su larga scala del trascrittoma di caffè, in particolare nel corso dello sviluppo e maturazione della drupa (cfr. § 1.4.9).

Un altro aspetto che nel corso degli ultimi anni ha attirato l'attenzione dei ricercatori, è lo studio degli eventi metabolici che avvengono nel seme nel breve periodo tra la raccolta del frutto e la completa essiccazione. Dal punto di vista biologico, il seme disidratato è ancora vitale, ma in stato di quiescenza indotta dal ridotto contenuto in acqua; non dovrebbe quindi presentare attività di sintesi proteica, ma è lecito attendersi che nella sua popolazione di mRNA sia conservata una sorta di "memoria" degli eventi che il seme ha vissuto fino al momento in cui è entrato in quiescenza.

In tale periodo infatti il seme è ancora metabolicamente attivo, ed è sottoposto a un grande stress soprattutto idrico e osmotico (ma anche termico nel caso dell'asciugatura meccanica ad aria calda), causato dalla disidratazione forzata. E' quindi naturale che al suo interno si attivino dei meccanismi di reazione allo stress.

Lo studio di questi eventi è stato iniziato dall'equipe del Prof. Selmar (Selmar *et al* 2001; Selmar *et al* 2004a), che ha messo in evidenza come il chicco di caffè non sia un inerte "serbatoio" di caffeina ed aromi, bensì un organismo vivo che reagisce in modo differenziato ai diversi stress e alle diverse situazioni in cui viene a trovarsi. Ancor più importante, è stato notato che queste reazioni possono avere un'influenza sulla qualità del prodotto. Le prime indicazioni di attività metabolica nel seme si sono avute con l'osservazione che a seconda della modalità di trattamento, i campioni di caffè verde presentavano diverse quantità di acido  $\gamma$ -aminobutirrico (GABA), un noto metabolita da stress idrico. In particolare, il livello di GABA è risultato molto più alto nei semi trattati per via secca. Dato che la lavorazione per via secca dura circa il doppio di quella per via umida, è facilmente ipotizzabile che il seme rimanga attivo più a lungo, e quindi abbia più tempo per reagire allo stress idrico sintetizzando maggiori quantità di GABA (Bytof *et al* 2004; Bytof *et al* 2005).

La chiave di queste osservazioni è stata la scoperta che nel corso della lavorazione post-raccolta, nel seme di caffè si attivano processi di germinazione, indicati dall'attività dei geni marcatori  $\beta$ -tubulina ed isocitrato liasi (Selmar *et al* 2004b; Selmar *et al* 2006; Bytof *et al* 2007). Inoltre, questa attività germinativa ha dinamiche temporali diverse nei diversi metodi di trattamento: nel corso del processo lavato la massima attività si ha 2 giorni dopo la raccolta, mentre nel caffè sundried ciò si verifica circa 1 settimana dopo l'inizio del trattamento.

La cessazione della germinazione si verifica, in entrambi i casi, quando il contenuto in acqua del seme scende a livelli troppo bassi per consentire attività metaboliche (inferiore al 25%). Data la durata molto diversa dei due trattamenti, questo punto si raggiunge molto prima nel caffè lavato. I fenomeni germinativi hanno quindi più tempo per svilupparsi nel corso del trattamento sundried, in cui l'estrazione di acqua avviene in maniera più lenta e graduale, mentre nel caffè lavato i processi di germinazione si attivano subito, ma altrettanto rapidamente si estinguono.

Questi fenomeni sembrano in parte spiegare le diverse qualità organolettiche tradizionalmente riconosciute ai due metodi di lavorazione post-raccolta, in particolare la differenza nel contenuto degli zuccheri fruttosio e glucosio, che è decisamente inferiore nel caffè lavato rispetto a quello naturale. Questa osservazione è compatibile con i diversi livelli di attivazione e mobilizzazione delle riserve metaboliche nel seme, in funzione della germinazione (Knopp *et al*, 2006).

#### **1.4.9 – Ricerca genomica nel caffè**

Negli ultimi anni si è verificato un crescente interesse per la genomica vegetale, stimolato dai sempre più numerosi progetti di sequenziamento di interi genomi e dalla costruzione di grandi librerie di EST in svariati organismi.

Sull'onda di queste ricerche, anche nello studio del caffè si è assistito alla nascita di alcuni progetti di genomica, su scala via via maggiore.

Innanzitutto, non va dimenticato che i primi passi in questo campo sono stati compiuti dal Laboratorio di Genetica del Dipartimento di Biologia dell'Università di Trieste, che ha sviluppato la prima collezione EST di *C. arabica* (Pallavicini *et al*, 2001) ed il primo cDNA microarray sempre di *Coffea arabica* (Pallavicini *et al*, 2001; De Nardi *et al*, 2006).

Come già citato in precedenza, la prima libreria BAC di *C. canephora* è stata costruita da Leroy *et al* (2005).

Vieira *et al* (2006) hanno sviluppato il Brazilian Coffee Genome Project, che ha sequenziato circa 150'000 EST di *C. arabica* e *C. canephora*.

La Cornell University, con la collaborazione della Nestlé, ha sequenziato circa 50'000 ulteriori EST (Lin *et al*, 2005) da *C. canephora*.

Infine, Salmona *et al* (2008) hanno sviluppato un macroarray di cDNA di drupa a vari stadi di maturazione, contenente sia sequenze di *C. canephora* che di *C. arabica*.

Questo nuovo patrimonio di conoscenza, e soprattutto il fatto che la maggior parte di questi dati sono stati resi pubblicamente disponibili (con l'eccezione delle EST brasiliane), consente di ipotizzare che nel prossimo futuro si assisterà ad una proliferazione di studi e ricerche di questo tipo. E' auspicabile che questo permetta di chiarire aspetti della genomica del caffè rimasti finora oscuri per la mancanza di sufficienti conoscenze di base su questa pianta.

## 1.5 - Terpeni e monoterpeni

### 1.5.1 – I terpeni

Come accennato nel § 1.4.7, nel profilo aromatico del caffè sia verde che tostato, si ritrovano composti terpenici ed in particolare monoterpenici (linalolo, limonene, geraniolo,  $\alpha$ -terpineolo), responsabili di alcune importanti note olfattive percepite positivamente dai degustatori quali frutto, fiore, fruttato, agrume.

I terpeni (anche detti isoprenoidi, dato che sono sintetizzati dalla condensazione di unità di isoprene a 5 atomi di carbonio) sono la classe strutturalmente più ampia e varia di composti di origine vegetale.

Sono suddivisi in 5 classi: mono-, di-, tri-, tetra- e sesqui-terpeni, sulla base delle unità monomeriche che li compongono. Si conoscono migliaia di diversi terpeni, in letteratura sono riportati più di 1000 monoterpeni, più di 3000 diterpeni e più di 7000 sesquiterpeni (Bohlmann *et al*, 1998).

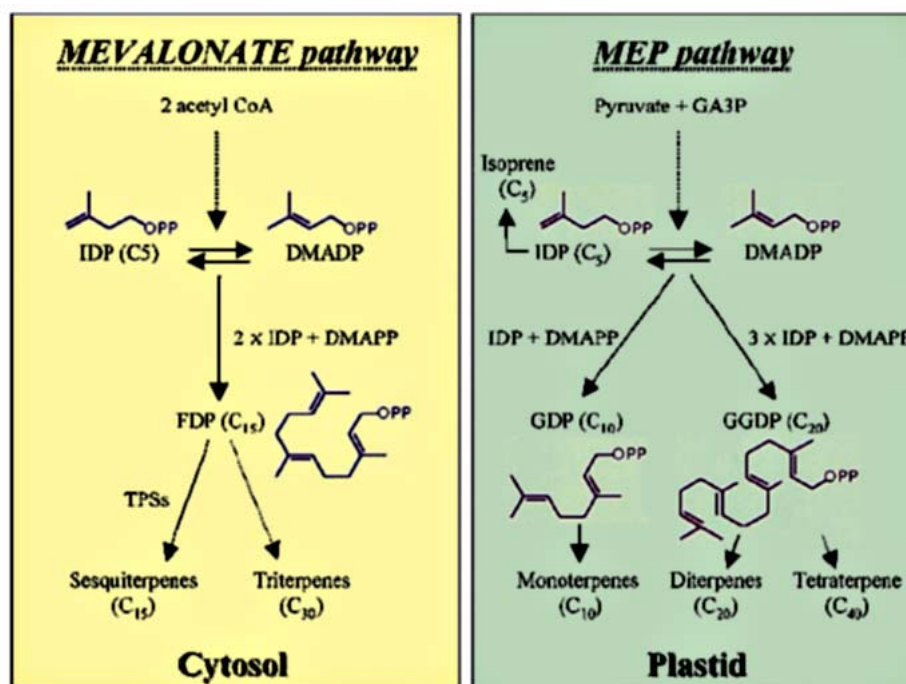


Fig. 1.5 – rappresentazione schematica delle vie metaboliche di sintesi delle diverse classi di terpeni.

GA3P=gliceraleide-3-fosfato; TPSs=terpene sintasi; IDP=isopentenil difosfato;  
DMADP=dimetilallil difosfato; FDP=farnesil difosfato; MEP=metileritritol-4-fosfato;  
GDP=geranil difosfato; GGDP=geranigeranil difosfato (adattata da Aharoni *et al*, 2006)

In Fig. 1.5 sono illustrate le vie metaboliche di biosintesi delle diverse classi terpeniche. Si nota come tutti questi composti nascono da un numero variabile di unità di isopentenil difosfato (IDP) e del suo isomero dimetilallil difosfato (DMAPP). Queste si combinano a formare i tre precursori di tutte le molecole terpeniche: il geranil difosfato (GDP, anche detto geranil pirofosfato - GPP) dà origine ai monoterpeni, il geranilgeranil

pirofosfato (GGPP) a di- e tetraterpeni, e il farnesil difosfato (FDP) a sesqui- e triterpeni. Dalla figura si vede anche che la sintesi di questi composti è compartimentata spazialmente, infatti sesqui- e triterpeni sono sintetizzati nel citosol, mentre la sintesi delle altre classi avviene nei plastidi.

La Fig. 1.6 illustra le stesse vie metaboliche, evidenziando però un altro aspetto importante, ovvero che i terpeni partecipano sia al metabolismo primario, sia a quello secondario. In generale si può dire (tenendo presente le dovute eccezioni) che i tri- e tetraterpeni sono impegnati in processi primari, mentre mono-, di- e sesquiterpeni sono di solito metaboliti secondari. Il più importante terpene coinvolto nel metabolismo primario, nonché una delle molecole più abbondanti in natura, è il diterpene aciclico fitolo, che costituisce la catena laterale della clorofilla. Altri terpeni importanti nel metabolismo primario sono quelli che svolgono ruoli di ormoni vegetali e che costituiscono i pigmenti carotenoidi. L'enorme varietà di metaboliti secondari terpenici svolge invece ruoli di pigmentazione, profumo e aroma di frutti e fiori, attrazione degli impollinatori o effetto deterrente verso predatori e parassiti, tossicità e difesa, e molte altre ancora.

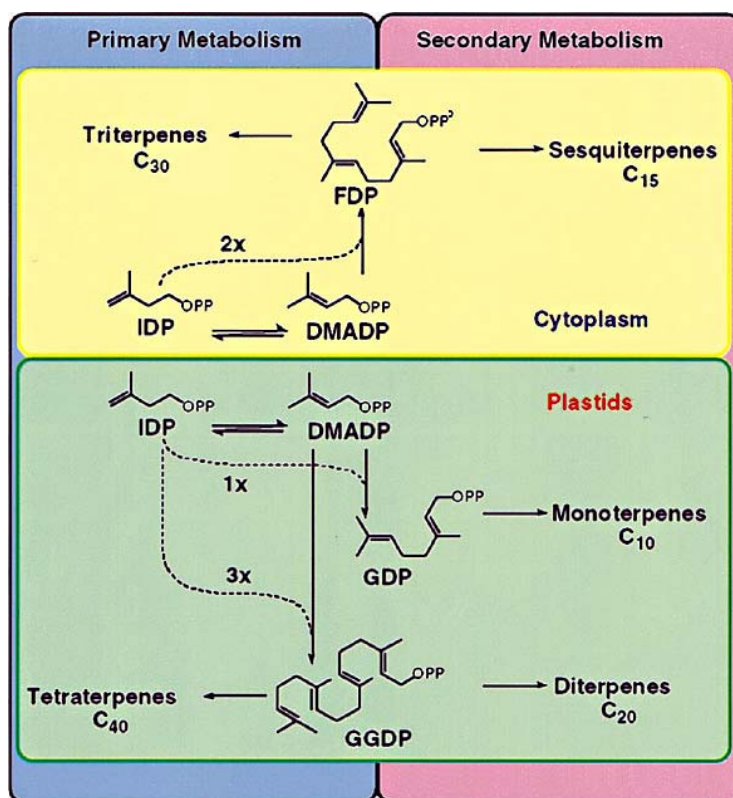


Fig. 1.6 – destinazione delle varie classi di composti terpenici al metabolismo primario o secondario. (Adattata da Bohlmann *et al*, 1998)

Per meglio comprendere come dallo schema di base riportato qui sopra, si origini la grandissima varietà strutturale dei terpeni, le Fig. 1.7, 1.8 e 1.9 riportano alcuni esempi delle strutture rispettivamente di mono-, di- e sesquiterpeni.

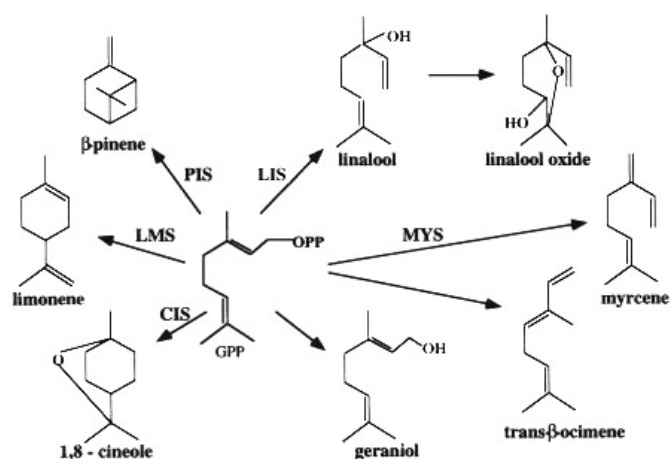


Fig. 1.7 – Biosintesi dal GPP e struttura di alcuni monoterpeni rappresentativi (Adattata da Dudareva *et al*, 2000)

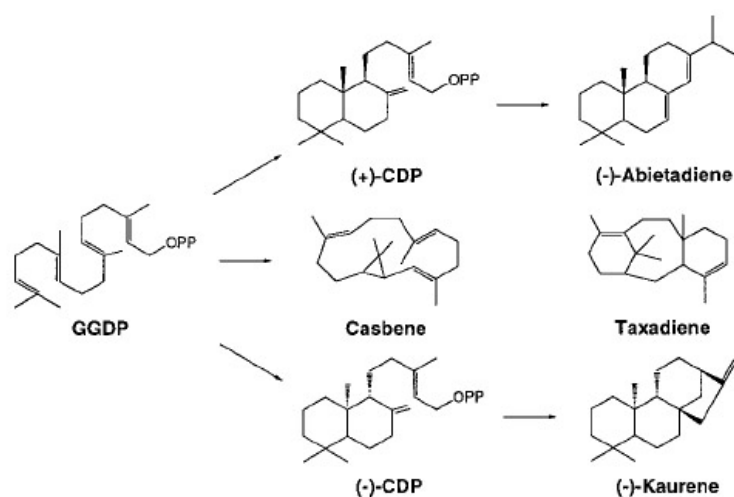


Fig. 1.8 – Biosintesi dal GGDP e struttura di alcuni diterpeni rappresentativi (Adattata da Bohlmann *et al*, 1998)

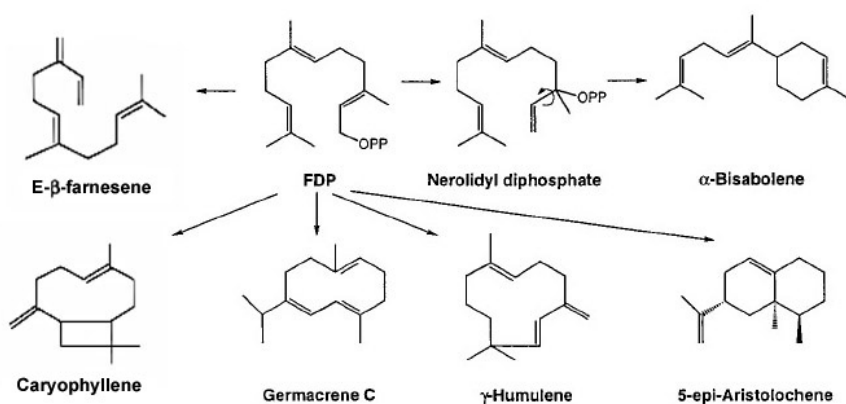


Fig. 1.9 – Biosintesi dall’FDP e struttura di alcuni sesquiterpeni rappresentativi (Adattata da Bohlmann *et al*, 1998 e Dudareva *et al*, 2000)

Oltre al loro ruolo fondamentale in natura, i terpeni hanno un'enorme importanza economica, in quanto sono responsabili di buona parte del profilo aromatico di numerosi frutti, ortaggi e piante aromatiche, nonché di prodotti quali il caffè, il tè, il vino. Per questo sono diventati ormai ingredienti industriali fondamentali nella formulazione di profumi, cosmetici, aromi alimentari, farmaci, antimicrobici ed insetticidi.

Da ciò consegue che la loro chimica, biochimica e biologia molecolare sia stata oggetto di studi particolarmente accurati in un gran numero di specie.

In dettaglio, Bohlmann *et al* (2000) e Chen *et al* (2003) hanno studiato l'emissione di questi volatili in *Arabidopsis*. Il risultato di questi studi è stato in parte sorprendente, in quanto *Arabidopsis* è una pianta autofertile che si riproduce preferenzialmente per autoimpollinazione, ed i cui fiori sono piccoli e non emettono odori percepibili. Eppure, delle 32 terpene sintasi identificate nel genoma di *Arabidopsis*, ben 20 sono risultate essere espresse nei fiori, evidenziando che anche una pianta non profumata e che non produce resine o olii essenziali come *Arabidopsis*, possiede un metabolismo attivo dei terpenoidi secondari.

Una delle prime monoterpene sintasi ad essere isolate è stata la linalolo sintasi di *Clarkia breweri*, attraverso un lungo lavoro che è cominciato dalla localizzazione dell'attività enzimatica nelle diverse parti del fiore (Pichersky *et al*, 1994), proseguendo poi con la purificazione e caratterizzazione enzimatica della proteina (Pichersky *et al*, 1995), per finire con l'ottenimento della sequenza nucleotidica ed aminoacidica (Dudareva *et al*, 1996). I risultati sono stati poi riepilogati in Cseke *et al* (1998), in cui viene ricostruita la filogenesi delle linalolo sintasi del genere *Clarkia*. In realtà, alla luce degli studi successivi, questi enzimi risultano piuttosto diversi dalle monoterpene sintasi isolate in altri organismi vegetali, sia come lunghezza della proteina (circa 900aa contro i 600 della maggioranza delle altre monoterpene sintasi), sia nell'analisi filogenetica, al punto da necessitare la creazione di un gruppo di classificazione distinto (vedi alberi filogenetici al § 4.4). Altri studi di questo gruppo si sono concentrati sulla sintesi, sempre in *Clarkia*, di altri composti odorosi, gli esteri benzenici e i fenilpropanoidi (Wang *et al*, 1997; Dudareva *et al*, 1998).

Uno dei maggiori filoni di ricerca sulle terpene sintasi ha riguardato gli enzimi di questa classe responsabili della sintesi dei numerosi composti terpenici che compongono la resina delle gimnosperme. Vale la pena citare i lavori su *Picea abies* di Faldt *et al* (2003), Martin *et al* (2003), Martin *et al* (2004), e quelli sulle di- e sesqui-terpene sintasi di *Abies grandis* (Stofer-Vogel *et al* 1996; Steele *et al* 1998).

Un'altra famiglia particolarmente studiata è quella delle *Lamiaceae*, che include numerosi generi di piante aromatiche ricche di olii essenziali quali *Salvia*, *Rosmarinus*, *Ocimum* (basilico), *Lavandula* e *Mentha*. Da svariate specie di questi generi sono stati isolate numerose monoterpene sintasi (Wise *et al* 1998; Iijima *et al* 2004; Crowell *et al* 2002). Al genere *Mentha* è stata dedicata particolare attenzione, con studi sulla regolazione funzionale della sintesi dei monoterpene nelle sue foglie e nei tricomi ghiandolari (Gershenzon *et al* 2000; McConkey *et al* 2000), e la recente caratterizzazione con cristallografia a raggi X della struttura di una limonene sintasi di *Mentha spicata* (Hyatt *et al* 2007).

Altre piante che sono state studiate includono i generi *Antirrhinum*, le cui monoterpene sintasi sembrano appartenere a una sottofamiglia filogeneticamente distinta dalle altre (Dudareva *et al*, 2003; cfr. anche § 4.4); *Perilla* (Hosoi *et al*, 2004); *Artemisia* (Jia *et al* 1999) e le orchidee del genere *Phalaenopsis* (Hsiao *et al*, 2006).



Tutti questi lavori hanno utilizzato un approccio classico, basandosi sull'analisi di similarità per amplificare anche nell'organismo studiato, tramite primer degenerati, delle sequenze di terpene sintasi che poi sono state ricostruite e completate.

Più di recente, la ricerca sulle terpene sintasi di alcune piante di grande interesse scientifico ed economico ha beneficiato delle nuove tecnologie di indagine su larga scala, e della grande quantità di informazione genetica derivante dai progetti di genomica completati ed in corso.

Sono così stati pubblicati studi estremamente interessanti, come ad esempio i lavori di Aharoni sulla fragola, in cui un cDNA microarray è stato utilizzato per individuare geni di monoterpene sintasi, ed esaminare le differenze nell'espressione di questi geni tra varietà selvatiche e coltivate (Aharoni *et al* 2000, Aharoni *et al* 2004).

Un approccio simile è stato utilizzato in diverse varietà di rosa da Guterman *et al* (2002) e Shalit *et al* (2003). Nei loro lavori, un database annotato di EST ricavate da petali di rosa, è stato utilizzato per costruire un cDNA microarray ed individuare nuovi geni correlabili allo specifico profilo aromatico di ciascuna varietà.

Nel caso del pomodoro, van Schie *et al* (2007) hanno isolato due geni di linalolo sintasi e ne hanno studiato l'induzione attraverso acido jasmonico.

Nel caso di questa specie, inoltre, la disponibilità di un'ampia collezione di EST unita a un'accurata mappa fisica del genoma, ha permesso di risalire alle sequenze regolatrici di alcune sesquiterpene sintasi, ricostruendone il sistema di controllo genico e la storia evolutiva (van der Hoeven *et al*, 2000).

Il genoma di *Vitis vinifera* è stato recentemente completato (The French-Italian Public Consortium for Grapevine Genome Characterization, 2007), e questo ha permesso di affiancare l'analisi di un intero genoma al lavoro di Martin *et al* (2004), che della vite aveva isolato e caratterizzato una  $\alpha$ -terpineolo sintasi.

Da questa analisi *in silico* sono state ricavate informazioni preziose sul numero e la tipologia delle terpene sintasi di una pianta che produce una grande quantità e varietà di composti aromatici (cfr. § 4.4).

Concludiamo questa rapida panoramica con i primi approcci all'ingegnerizzazione di piante con geni di terpene sintasi. Questi lavori sono stati resi possibili dalla solida base di informazioni che si è consolidata in tempi recenti: da un lato riguardo alla conoscenza di interi genomi vegetali, dall'altro in merito alle sequenze, strutture e meccanismi di regolazione delle sintasi dei composti terpenici.

Sono stati compiuti esperimenti per aumentare i livelli di linalolo prodotti nel pomodoro, andando ad agire sui meccanismi della via biosintetica dei terpenoidi (Lewinsohn *et al*, 2001).

Lucker *et al* (2004) hanno trasformato con successo una pianta di tabacco con tre monoterpene sintasi di limone, incrementandone la quantità e varietà di monoterpene emessi.

Aharoni *et al* (2003) hanno confrontato il profilo dei composti volatili emessi da linee selvatiche di *Arabidopsis*, con quello di linee che sovraesprimevano due diverse monoterpene sintasi, riscontrando livelli di emissione di linalolo ed altri monoterpene fino a 60 volte superiori agli individui *wild-type*.

Questi risultati hanno permesso di affermare che l'ingegnerizzazione delle vie metaboliche dei terpeni nelle piante è possibile, e apre la strada ad ampie possibilità di miglioramento varietale, sia dal punto di vista organolettico, sia da quello delle relazioni tra la pianta e i suoi parassiti (Aharoni *et al*, 2006).

Una ulteriore conferma dell'interesse non solo accademico che gravita attorno a queste molecole e ai geni che ne sovrintendono la sintesi, è che sia le sequenze della linalolo sintasi di *Clarkia*, sia quelle di fragola (*Fragaria x ananassa*), sono state brevettate per tutelare il loro possibile utilizzo a fini industriali (European Patent EP1006190, European Patent EP1231273, United States Patent 5849526).

Anche nel caffè, questi geni sono stati oggetto di ricerche, seppur non dal punto di vista biomolecolare. Weckerle *et al* (2003) hanno caratterizzato tramite NMR e spettrometria di massa la struttura di due linalol-disaccaridi presenti in semi di *Coffea arabica*. Bonnlaender *et al* (2004) hanno analizzato il rapporto tra gli enantiomeri di linalolo e la sua variazione tra caffè verde e caffè tostato.

### **1.5.2 – Monoterpeni e monoterpene sintasi**

Nell'enorme varietà e complessità della famiglia genica delle terpene sintasi, i geni più omogenei come struttura sono le monoterpene sintasi. A parte le eccezioni costituite dai geni di *Clarkia* e *Antirrhinum* sopradescritti, la maggioranza delle monoterpene sintasi è costituita da un cDNA di approssimativamente 2Kb, che codifica per una proteina circa 600 aminoacidi. Questa proteina catalizza la trasformazione del GPP in monoterpene in una reazione single-step (cfr. §1.5.1; Fig. 1.7).

La semplicità di questa via metabolica ha suggerito di concentrarsi su questi geni per lo svolgimento del presente lavoro, che mirava a ricercare i primi geni di *C. arabica* correlabili alle caratteristiche di qualità in tazza della bevanda.

## 2. SCOPO DELLA TESI

Il presente lavoro si è posto l'obiettivo di isolare i primi geni di *Coffea arabica* che possano esser messi in relazione con la composizione aromatica della bevanda. A tale scopo è stato deciso di focalizzare la ricerca sulle monoterpene sintasi, ovvero gli enzimi che sintetizzano i composti appartenenti alla classe dei monoterpeni. Questi composti sono estremamente importanti nell'industria alimentare in quanto costituiscono la componente principale degli aromi gradevoli di frutta, verdura e bevande di origine vegetale quali tè, vino e caffè. La loro importanza si estende però anche in campo farmaceutico e cosmetico, essendo componenti fondamentali nella formulazione di profumi, farmaci e prodotti di cosmesi.

Tutti i monoterpeni vengono sintetizzati a partire dal precursore comune geranyl pirofosfato (GPP) tramite reazioni single-step controllate ciascuna da un singolo enzima. Una via metabolica così semplice e lineare è apparsa essere l'ideale per l'apertura di un filone di ricerca inedito in un organismo mai studiato prima sotto questo punto di vista.

Il presente lavoro si è quindi posto gli obiettivi di:

- isolare uno o più geni di *C. arabica* putativamente classificabili come monoterpene sintasi
- ricavarne la sequenza codificante completa
- studiarne l'espressione genica nei tessuti più strettamente correlati alla qualità finale del prodotto: il fiore, il frutto ed il seme
- caratterizzarli tramite analisi di similarità e analisi filogenetica



### 3. MATERIALI E METODI

#### 3.1 - Estrazioni di RNA totale da *C. arabica*

##### 3.1.1 – Estrazione di RNA da fiori e drupe freschi

La sperimentazione sui frutti di caffè è stata svolta utilizzando l'RNA già raccolto dal Laboratorio di Genetica nel corso delle passate fioriture delle piante presenti nella serra del Dipartimento di Biologia dell'Università di Trieste. Tale RNA è stato ricavato dalle drupe di piante di Catuai Red a 10, 15, 25 settimane di maturazione, e infine dai frutti a maturazione completa. In aggiunta, è stato isolato l'RNA totale da fiori di *C. arabica* provenienti dal raccolto indiano della stagione 2006, gentilmente forniti dal Dott. Manoj Mishra. Il protocollo utilizzato è un adattamento del protocollo con CTAB di Chang, Puryear e Cairney (1993).

##### SOLUZIONI UTILIZZATE:

- Tampone di estrazione:

TrisHCl (pH 8.0)	100 mM
NaCl	2.0 M
Spermidina	0.5 g/L
EDTA	25 mM
CTAB	2% (p/v)
PVP 30 K *	2% (p/v)
$\beta$ -mercaptoetanolo *	2% (v/v)

\* = PVP e  $\beta$ -mercaptoetanolo vanno aggiunti freschi, immediatamente prima di procedere con l'estrazione.

Tutte le soluzioni vanno preparate con H<sub>2</sub>O DEPC. Il buffer finale va sterilizzato per filtrazione e conservato a temperatura ambiente.

- LiCl 8M: aggiungere 33.9g di LiCl a 50ml di H<sub>2</sub>O DEPC. Portare la soluzione al volume finale di 100ml. Sterilizzare per filtrazione e conservare a temperatura ambiente.

##### GIORNO 1

- 1- Aggiungere alla soluzione di estrazione PVP e  $\beta$ -mercaptoetanolo, e preriscaldarla in stufa a 65°C.
- 2- Tagliuzzare con il bisturi 1 g di tessuto fresco in frammenti di piccole dimensioni. Per evitare l'ossidazione del tessuto, può essere conveniente aggiungervi qualche ml di tampone di estrazione.
- 3- trasferire il tessuto sminuzzato in un tubo Falcon da 50 ml contenente 10 ml di tampone di estrazione caldo, e omogenizzare il più possibile il tessuto.
- 4- estrarre con 1 volume di CHL:IAA (24:1), mescolando per inversione. Centrifugare 25 min a 4000 rpm, 4°C, e trasferire il surnatante in Falcon da 50 ml pulita.
- 5- estrarre una seconda volta come dal punto 4. Trasferire la fase acquosa in un tubo Falcon da 15 ml.
- 6- aggiungere 0.3 volumi di LiCl 8M, mescolare per inversione e precipitare overnight a 4°C.

## GIORNO 2

- 7- centrifugare 25 min a 7000 rpm, 4°C.
- 8- rimuovere completamente la fase acquosa.
- 9- lavare bene il pellet con 1 ml di EtOH 75%.
- 10- centrifugare 15 min a 7000 rpm, 4°C. Eventualmente centrifugare prima brevemente in centrifuga swinging-bucket. Rimuovere l'etanolo e asciugare bene il pellet, avendo però cura di non seccarlo troppo.
- 11- risospendere il pellet in 100 µl di H<sub>2</sub>O DEPC e trasferire in Eppendorf da 1.5 ml.
- 12- leggere una diluizione della soluzione allo spettrofotometro (lunghezze d'onda 230, 260, 280 e 310 nm).
- 13- conservare a -80°C.

### 3.1.2 – Estrazione di RNA da semi di caffè essiccati: protocollo CTAB

Questo protocollo si basa sul metodo proposto da Chang *et al* (1993), modificato aggiungendo al buffer di estrazione 0.5 mg/ml di Proteinasi K e un'incubazione a 55°C per 45 minuti, ottimizzando le condizioni proposte da Azevedo *et al* (2003). La proteinasi degrada la frazione proteica dell'omogenato, proteggendo l'RNA dalla degradazione per opera delle nucleasi endogene. Inoltre, l'incubazione a caldo favorisce la solubilizzazione dei composti e la fuoriuscita del contenuto cellulare, aumentando la resa dell'estrazione.

Dopo questo passaggio l'omogenato viene brevemente centrifugato a 4°C per precipitare il PVP, che porta con sé il particolato residuo e la componente fenolica, aumentando di molto la pulizia degli step successivi. Da questo punto il protocollo ricalca quello di Chang *et al* (1993), con estrazioni in cloroformio/alcol isoamilico (24:1) e precipitazione selettiva dell'RNA in LiCl 2M.

#### SOLUZIONI UTILIZZATE:

- Tampone di estrazione:

Tris-HCl (pH 8)	100 mM
NaCl	2.0 M
Spermidina	0.5 g/L
EDTA	25 mM
CTAB	2% (p/v)
PVP 90 K *	2% (p/v)
β-mercaptoetanolo *	2% (v/v)
proteinase K *	0.5 mg/ml

\* = PVP, β-mercaptoetanolo e proteinasi K vanno aggiunti freschi, immediatamente prima di procedere con l'estrazione.

Tutte le soluzioni vanno preparate con H<sub>2</sub>O DEPC. Il buffer finale va sterilizzato per filtrazione e conservato a temperatura ambiente.

- LiCl 8M: aggiungere 33.9g di LiCl a 50ml di H<sub>2</sub>O DEPC. Portare la soluzione al volume finale di 100ml. Sterilizzare per filtrazione e conservare a temperatura ambiente.

## GIORNO 1

- 1- Aggiungere alla soluzione di estrazione PVP,  $\beta$ -mercaptoetanolo e proteinasi K, e preriscaldarla in stufa a 65°C.
- 2- pesare 2 semi (300-350 mg) e sminuzzarli finemente col bisturi, avendo cura che i frammenti siano di dimensioni inferiori a 1 mm. Questo agevolerà molto l'omogenizzazione, migliorando la resa dell'estrazione.
- 3- trasferire i semi sminuzzati in un tubo Falcon da 15 ml contenente 10 ml di tampone di estrazione caldo, e omogenizzare il più possibile il tessuto.
- 4- incubare l'omogenato a 55°C per 45 min.
- 5- centrifugare 15min a 7000 rpm, 4°C, per precipitare il PVP.
- 6- trasferire il surnatante in Falcon da 50 ml ed estrarre con 1 volume di CHL:IAA (24:1), mescolando per inversione. Centrifugare 25 min a 4000 rpm, 4°C, e trasferire il surnatante in Falcon da 50 ml pulita.
- 7- estrarre una seconda volta come dal punto 6. Trasferire la fase acquosa in un tubo Falcon da 15 ml.
- 8- aggiungere 0.3 volumi di LiCl 8M, mescolare per inversione e precipitare overnight a 4°C.

## GIORNO 2

- 9- centrifugare 25 min a 7000 rpm, 4°C.
- 10- rimuovere completamente la fase acquosa.
- 11- lavare bene il pellet con 1 ml di EtOH 75%.
- 12- centrifugare 15 min a 7000 rpm, 4°C. Eventualmente centrifugare prima brevemente in centrifuga swinging-bucket. Rimuovere l'etanolo e asciugare bene il pellet, avendo però cura di non seccarlo troppo.
- 13- risospendere il pellet in 100  $\mu$ l di H<sub>2</sub>O DEPC e trasferire in Eppendorf da 1.5 ml.
- 14- leggere una diluizione della soluzione allo spettrofotometro (lunghezze d'onda 230, 260, 280 e 310 nm).
- 15- conservare a -80°C.

*Opzionale:*

Nel caso la purezza dell'RNA sia insufficiente, può essere opportuno effettuare una seconda precipitazione in LiCl, ripetendo i passaggi da 8 a 15.

**3.1.3 – Estrazione di RNA da semi di caffè essiccati: protocollo “Hot borate”**

Questo protocollo si basa sul metodo di Wan e Wilkins (1994). Ne esistono numerose varianti ed adattamenti, in base alla specie ed al tessuto da cui estrarre l'RNA (Wu *et al* 2002, Smart *et al* 1999). Il protocollo utilizzato in questo lavoro è stato adattato da quello di Birtic e Kranner (2006), che è specificamente ottimizzato per l'estrazione di RNA da semi ricchi in polisaccaridi, polifenoli e lipidi. Questo metodo è analogo a quello con CTAB nel suo fare uso di PVP come inibitore dei polifenoli, e di elevate concentrazioni di Proteinasi K per degradare le proteine, in particolare le Rnasi.

Sotto tutti gli altri aspetti, però, se ne differenzia in modo sostanziale: innanzitutto nella composizione del buffer di estrazione, che non contiene CTAB ma sodio tetraborato decaidrato, sostituisce l'EDTA con l'analogo EGTA, e utilizza come riducente il ditiotreitolo in luogo del  $\beta$ -mercaptoetanolo. L'SDS e il sodio deossicolato agiscono come detergenti per solubilizzare le membrane e facilitare la rottura delle pareti cellulari. Altra

importante differenza risiede nel fatto che questo protocollo non prevede estrazioni con solventi organici (CHL:IAA), sostituendo ad esse l'azione dei detergenti e precipitazioni saline successive. I lipidi vengono separati dalla fase acquosa durante la prima centrifugazione, mentre nel contempo il PVP sequestra i composti fenolici e il KCl favorisce la precipitazione dei polisaccaridi carichi positivamente. Dopo il passaggio overnight in LiCl, l'aggiunta di potassio acetato serve a precipitare le proteine residue prima della risospensione finale.

#### SOLUZIONI UTILIZZATE:

- Tampone di estrazione (pH 9):

Sodio tetraborato decaidrato	200 mM
SDS	1% (p/v)
Sodio deossicolato	1% (p/v)
EGTA	30 mM
(correggere il pH a 9 con NaOH 1M)	
PVP 90 K *	2% (p/v)
Ditiotreitolo (DTT) *	10 mM
proteinase K *	1.5 mg/ml

\* = PVP, DTT e proteinasi K vanno aggiunti freschi, immediatamente prima di procedere con l'estrazione.

Tutte le soluzioni vanno preparate con H<sub>2</sub>O DEPC. Il buffer finale va sterilizzato per filtrazione e conservato a temperatura ambiente.

- LiCl 8M: aggiungere 33.9g di LiCl a 50ml di H<sub>2</sub>O DEPC. Portare la soluzione al volume finale di 100ml. Sterilizzare per filtrazione e conservare a temperatura ambiente.

- KCl 2M: aggiungere 7.455g di KCl a 20ml di H<sub>2</sub>O DEPC. Portare la soluzione al volume finale di 50ml. Sterilizzare per filtrazione e conservare a temperatura ambiente.

- KOAc 2M: aggiungere 9.815g di KOAc a 20ml di H<sub>2</sub>O DEPC. Portare la soluzione al volume finale di 50ml. Sterilizzare per filtrazione e conservare a temperatura ambiente.

#### GIORNO 1

- 1- Aggiungere alla soluzione di estrazione PVP, DTT e proteinasi K, e preriscaldarla in stufa a 65°C.
- 2- pesare 1 seme (circa 150 mg) e sminuzzarlo finemente col bisturi, avendo cura che i frammenti siano di dimensioni inferiori a 1 mm. Questo agevolerà molto l'omogenizzazione, migliorando la resa dell'estrazione.
- 3- trasferire il seme sminuzzato in un tubo da 2 ml contenente 1 ml di tampone di estrazione caldo, e omogeneizzare il più possibile il tessuto.
- 4- incubare l'omogenato a 42°C per 1.5 h.
- 5- aggiungere 87 µl di KCl 2M (conc. finale 160 mM).
- 6- incubare in ghiaccio 45 min.
- 7- centrifugare 20 min a 13'000 rpm, 4°C.
- 8- scartare il pellet e trasferire il surnatante in un tubo Eppendorf pulito, facendo attenzione a non trasferire i lipidi rimasti nella fase superiore.
- 9- aggiungere 1/3 vol LiCl 8M (conc. finale 2M).
- 10- incubare overnight a 4°C.



## GIORNO 2

- 11- preparare una diluizione 1:4 dello stock LiCl 8M (finale 2M)
- 12- centrifugare 20 min a 13'000 rpm, 4°C.
- 13- scartare il surnatante e lavare il pellet con 750 µl LiCl 2M
- 14- centrifugare 15 min a 13'000 rpm, 4°C.
- 15- ripetere il lavaggio secondo i punti 13-14.
- 16- scartare il surnatante e risospendere il pellet in 100µl H<sub>2</sub>O DEPC.
- 17- aggiungere 1/10 vol KOAc 2M.
- 18- incubare in ghiaccio 15 min.
- 19- centrifugare 15 min a 13'000 rpm, 4°C.
- 20- trasferire il surnatante in un tubo Eppendorf pulito.
- 21- aggiungere 1/10 vol NaAc 3M, 2.5 vol EtOH 100% e precipitare a -80°C per 2h (*eventualmente overnight*).
- 22- centrifugare 30 min a 13'000 rpm, 4°C.
- 23- scartare il surnatante e lavare il pellet con 500 µl EtOH 70%
- 24- centrifugare 15 min a 13'000 rpm, 4°C.
- 25- rimuovere l'etanolo e asciugare bene il pellet, avendo però cura di non seccarlo troppo.
- 26- risospendere l'RNA in 50µl H<sub>2</sub>O DEPC.
- 27- leggere una diluizione della soluzione allo spettrofotometro (lunghezze d'onda 230, 260, 280 e 310 nm).
- 28- conservare a -80°C.

### 3.1.4 – Quantizzazione dell'RNA estratto

Indipendentemente dal metodo di estrazione utilizzato, l'RNA è stato quantizzato tramite misura spettrofotometrica dell'assorbanza (A) del campione alle lunghezze d'onda di 230nm, 260nm, 280 e 310nm. La concentrazione del campione è stata calcolata mediante la formula:

$$\text{Concentrazione (ng/}\mu\text{l)} = (A_{260} - A_{310}) \times 40$$

La valutazione di un'eventuale contaminazione del campione da parte di proteine (P), è stata calcolata nel seguente modo:

$$P = (A_{260} - A_{310}) / (A_{280} - A_{310})$$

Un'eventuale contaminazione da carboidrati (C), invece, è stata verificata con la formula:

$$C = (A_{260} - A_{310}) / (A_{230} - A_{310})$$

In un RNA di buona qualità, i valori di P e C dovrebbero essere pari o superiori a 2.

L'RNA è stato infine corso in gel di agarosio denaturante per verificarne visivamente l'integrità, giudicabile dalla presenza e intensità delle due tipiche bande degli RNA ribosomiali 18 S e 28 S.

### 3.1.5 – RT-PCR

Con la misura spettrofotometrica si controlla la quantità e la purezza dell'RNA totale. Per verificarne la qualità in termini di integrità e possibilità di utilizzo per reazioni successive, è stata fatta una RT-PCR di controllo con geni *housekeeping*, che si possono presupporre espressi in qualunque tessuto.

In primo luogo l'RNA è stato trattato con Dnase I per rimuovere eventuali contaminazioni da DNA genomico.

Reagente	Volume	Concentrazione finale
RNA totale	q.b per 1 µg	100 ng/µl
10x DNase Buffer	1 µl	1x
DNase I	1 µl	-
H <sub>2</sub> O	q.b a 10 µl	-
TOTALE	10 µl	

- incubare 15 minuti a temperatura ambiente
- inattivare la DNase con 0.8 µl di EDTA 25mM

Ogni campione di RNA viene ora diviso in due aliquote da 5 µl, su cui verranno svolte in parallelo due reazioni di retrotrascrizione: in quella positiva (+) si aggiunge effettivamente la trascrittasi inversa per sintetizzare il cDNA; nel controllo negativo (-) non viene aggiunto l'enzima, in modo che, se la successiva PCR dà un prodotto di amplificazione, questo è indice di una residua contaminazione da DNA genomico. La retrotrascrizione viene effettuata usando come innesco un oligo-d(T) che si appaia al poli-A situato al 3' degli mRNA.

Miscela di reazione:

Reagente	Volume	Concentrazione finale
5X RT-PCR Buffer	6 µl	1X
MgCl <sub>2</sub> 2mM	3 µl	200 µM
DNTP mix (dATP, dCTP, dGTP, dTTP, 10 mM ciascuno)	3 µl	1 mM ciascuno
DTT 100mM	3 µl	10 mM
Oligo-d(T) <sub>16</sub> 10µM	0.75 µl	250 nM
RNase Inhibitor	0.75 µl	-
RNA totale	5 µl ≈ 500ng	≈ 15 ng/µl
H <sub>2</sub> O	8.5 µl	-
TOTALE	30 µl	

Condizioni di retrotrascrizione:

Step	Temperatura	Durata
Denaturazione iniziale	65° C	10'
Annealing	55° C	10'
Retrotrascrizione	42° C	50'

Non appena parte la fase di retrotrascrizione a 42°C, ai soli campioni (+) si aggiungono 0.45 µl di MMLT-Reverse Transcriptase.

Una volta sintetizzato il cDNA, si effettua una normale PCR per controllarne la qualità. Come primer di controllo, sono state utilizzate delle sequenze specifiche per il gene della poliubiquitina di *C. arabica*.

Primer utilizzati:

Oligonucleotide	Sequenza (5'-3')
Ubiquitina_For	GCA GAT TTT CGT CAA GAC ATT GAC TG
Ubiquitina_Rev	ACC AAG GCA AGT TCA GTT CAA CAC

Miscela di reazione:

Reagente	Volume	Concentrazione finale
10X PCR Buffer (100mM Tris-HCl pH 9, 500mM KCl, 1% Triton X100)	2.5 µl	1X
MgCl <sub>2</sub> 25mM	2.5 µl	2.5 mM
DNTP mix (dATP, dCTP, dGTP, dTTP, 10 mM ciascuno)	0.5 µl	200 µM ciascuno
Forward primer 10 µM	0.5 µl	200 nM
Reverse primer 10 µM	0.5 µl	200 nM
Taq DNA Polimerasi	0.5 µl ≈ 0.5U	0.025 U/µl
cDNA template	1 µl	-
H <sub>2</sub> O	17 µl	-
TOTALE	25 µl	

Condizioni di amplificazione:

Step	Temperatura	Durata	n° cicli
Denaturazione iniziale	94° C	3'	1
Denaturazione	94° C	30"	} 32
Annealing	55° C	30"	
Estensione	72° C	1'	
Estensione finale	72° C	10'	1

Il risultato dell'amplificazione è stato controllato in gel di agarosio all'1% in tampone TBE (Tris-HCl 80mM, H<sub>3</sub>BO<sub>3</sub> 80mM, EDTA 2.55mM, pH 8.8), con etidio bromuro 0.5 µg/ml.

### 3.2 - Ricerca di geni di *C. arabica* omologhi a monoterpene sintasi

#### 3.2.1 – Analisi di similarità e disegno dei primer degenerati

La ricerca preliminare di dati sulle monoterpene sintasi (articoli, sequenze nucleotidiche e aminoacidiche, analisi di similarità) è stata svolta principalmente sul database Entrez del National Center for Biotechnology Information (NCBI), sfruttando il motore di ricerca PubMed per reperire pubblicazioni pertinenti, e gli algoritmi blastn, blastp, blastx, tblastn e tblastx (Altschul *et al.*, 1997) per le analisi di similarità, tutti accessibili tramite il suddetto portale Entrez (<http://www.ncbi.nlm.nih.gov/sites/gquery>).

Ulteriori dati sono stati recuperati dal portale dell'European Bioinformatics Institute <http://www.ebi.ac.uk>.

Gli allineamenti di sequenze multiple sono stati eseguiti con i programmi ClustalW (Higgins *et al* 1994; Larkin *et al* 2007; <http://www.ebi.ac.uk/tools/clustalw2/index.html>) e Multalin (Corpet 1988, <http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html>).

I primer degenerati sono stati disegnati manualmente, ricercando le regioni geniche a minor ridondanza, e utilizzando la nomenclatura standard per sequenze degenerate.

#### 3.2.2 – PCR con primer degenerati da DNA genomico

Per verificare la funzionalità dei primer degenerati, è stata effettuata una PCR da DNA genomico di *C. arabica* cv. Caturra, utilizzando tutte le 4 combinazioni possibili di primer. La miscela di reazione è descritta nella tabella sottostante.

Reagente	Volume	Concentrazione finale
10X PCR Buffer (100mM Tris-HCl pH 9, 500mM KCl, 1% Triton X100)	2.5 µl	1X
MgCl <sub>2</sub> 25mM	2.5 µl	2.5 mM
DNTP mix (dATP, dCTP, dGTP, dTTP, 10 mM ciascuno)	0.5 µl	200 µM ciascuno
Forward primer 10 µM	0.5 µl	200 nM
Reverse primer 10 µM	0.5 µl	200 nM
Taq DNA Polimerasi	0.5 µl ≈ 0.5U	0.025 U/µl
DNA template	1 µl ≈ 25ng	1 ng/µl
H <sub>2</sub> O	17 µl	-
TOTALE	25 µl	

Le reazioni sono state eseguite in termociclatore Eppendorf Mastercycler EP Gradient S secondo il seguente programma:

Step	Temperatura	Durata	n° cicli
Denaturazione iniziale	94° C	3'	1
Denaturazione	94° C	30"	} 35
Annealing	45° C	1'	
Estensione	72° C	30"	
Estensione finale	72° C	10'	1

Il risultato dell'amplificazione è stato controllato in gel di agarosio all'1% in tampone TBE (Tris-HCl 80mM, H<sub>3</sub>BO<sub>3</sub> 80mM, EDTA 2.55mM, pH 8.8), con etidio bromuro 0.5 µg/ml.

### 3.2.3 - Clonaggio degli amplificati con kit TOPO-TA Cloning®

Le amplificazioni risultate positive sono state innanzitutto purificate per eliminare i residui della PCR (proteine, sali, nucleotidi e primer) che possono interferire con la successiva reazione di clonaggio. Per la purificazione è stato usato il kit MinElute (Qiagen), seguendo le indicazioni del produttore. L'eluizione finale è stata fatta in un volume di 10µl di H<sub>2</sub>O.

Le PCR purificate sono state poi clonate tramite la tecnologia TOPO-TA Cloning® (Invitrogen). Questo metodo richiede che gli inserti da clonare possiedano una A sporgente al 3' (*3' A-overhang*); in primo luogo è quindi necessario aggiungere questa base con una Taq polimerasi *non* proof-reading; in questo caso è stata usata la EuroTaq (EuroClone), componendo la mix nel modo seguente.

Reagente	Volume	Concentrazione finale
10X Euro Buffer	1.2 µl	1X
MgCl <sub>2</sub> 50mM	0.6 µl	2.5 mM
dATP 10 mM	0.25 µl	200 µM
EuroTaq DNA Polimerasi (5U/µl)	0.15 µl ≈ 0.75 U	0.0625 U/µl
PCR	10 µl	-
TOTALE	12 µl	

- incubare 15' a 72°C

Al termine dell'incubazione, la reazione è stata immediatamente posta in ghiaccio. A questo punto si è proceduto con la reazione di clonaggio sotto descritta.

Reagente	Volume
PCR purificata	1 µl
Salt solution (1.2M NaCl, 0.06M MgCl <sub>2</sub> )	0.25 µl
PCR II-TOPO vector	0.25 µl
TOTALE	1.5 µl

- incubare 20' a temperatura ambiente

Notare come non sia necessario l'utilizzo di una DNA ligasi, perchè al vettore è legata covalentemente una topoisomerasi che taglia il DNA a doppio filamento e riforma il legame fosfodiesterico, integrando così l'inserto nel vettore pCR II-TOPO®, le cui caratteristiche sono illustrate in Fig. 3.1.

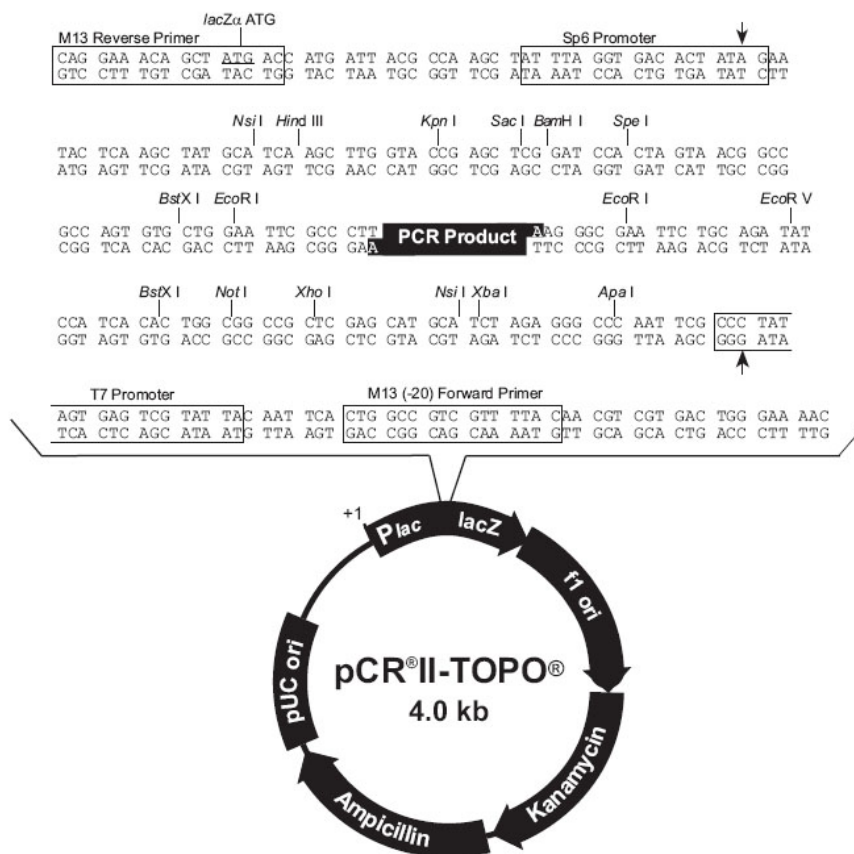


Fig. 3.1 – mappa del vettore pCR II-TOPO utilizzato per il clonaggio

Una volta terminato il clonaggio, le reazioni sono state purificate tramite precipitazione con sodio acetato ed etanolo, per eliminare le proteine e soprattutto i sali che andrebbero altrimenti a interferire con la successiva elettroporazione.

Reagente	Volume
Vettore con inserto	1.5 µl
Glicogeno (10mg/ml)	3 µl
H <sub>2</sub> O	15.5 µl
NaAc 3M, pH 5.2	2 µl
EtOH 100%	44 µl
TOTALE	66 µl

- precipitare 1h a -20°C
- centrifugare 20', 13'000rpm, 4°C
- eliminare il surnatante e lavare il pellet con EtOH 70%
- risospendere in 2µl H<sub>2</sub>O

Il vettore purificato contenente l'inserto è stato a questo punto elettroporato in cellule elettrocompetenti, che sono poi state piastrate su terreno LB-agar contenente kanamicina (50 µg/ml) come agente di selezione. Le piastre sono quindi state incubate overnight a 37°C per verificare la crescita delle colonie.

### 3.2.4 – Amplificazione degli inserti clonati

Allo scopo di conservare i cloni batterici per utilizzi futuri, le colonie sono state innanzitutto prelevate dalla piastra Petri e trasferite in una piastra a 96 pozzetti per colture batteriche liquide, contenente terreno LB-agar con Kanamicina (50 µg/ml) e glicerolo 10%. Questa piastra è stata lasciata overnight a 37°C per favorire la crescita dei batteri.

Per verificare la presenza degli inserti, i singoli cloni batterici sono stati amplificati tramite le sequenze M13 che fiancheggiano il sito di clonaggio. Da ogni pozzetto è stata prelevata un'aliquota che è stata trasferita direttamente nella posizione corrispondente sulla piastra da PCR, come template per la reazione (la fase di denaturazione iniziale a 94°C è sufficiente a lisare le cellule batteriche, rilasciando in soluzione il DNA).

Primer utilizzati:

Oligonucleotide	Sequenza (5'-3')
M13 Forward	GTAAAACGACGGCCAG
M13 Reverse	CAGGAAACAGCTATGAC

Miscela di reazione:

Reagente	Volume	Concentrazione finale
10X PCR Buffer (100mM Tris-HCl pH 9, 500mM KCl, 1% Triton X100)	2.5 µl	1X
MgCl <sub>2</sub> 25mM	2.5 µl	2.5 mM
DNTP mix (dATP, dCTP, dGTP, dTTP, 10 mM ciascuno)	0.5 µl	200 µM ciascuno
M13 Forward 10 µM	0.5 µl	200 nM
M13 Reverse 10 µM	0.5 µl	200 nM
Taq DNA Polimerasi	0.5 µl ≈ 0.5U	0.025 U/µl
DNA template	colonia batterica	-
H <sub>2</sub> O	18 µl	-
TOTALE	25 µl	

Condizioni di amplificazione:

Step	Temperatura	Durata	n° cicli
Denaturazione iniziale	94° C	5'	1
Denaturazione	94° C	30"	} 35
Annealing	55° C	30"	
Estensione	72° C	1' 30"	
Estensione finale	72° C	10'	1

Il risultato dell'amplificazione è stato controllato in gel di agarosio all'1% in tampone TBE (Tris-HCl 80mM, H<sub>3</sub>BO<sub>3</sub> 80mM, EDTA 2.55mM, pH 8.8), con etidio bromuro 0.5 µg/ml.

### 3.2.5 – Sequenziamento, assemblaggio ed analisi delle sequenze

I cloni risultati positivi (presenza di una singola banda all'altezza prevista) sono stati sequenziati presso il Servizio di Sequenziamento del centro BMR <http://www.bmr-genomics.it/>, spin-off ufficiale del CRIBI dell'Università di Padova (<http://www.cribi.unipd.it/>). Le sequenze sono state poi assemblate in contig con il programma *Seqman* del pacchetto Lasergene (DNA Star Inc.), per raggruppare i cloni ridondanti e distinguere i diversi geni presenti nella popolazione amplificata dai primer degenerati.

I contig così ottenuti sono stati analizzati mediante i programmi *blastn*, *blastx* e *tblastx* (<http://ncbi.nlm.nih.gov/blast>), per verificare il loro livello di similarità a sequenze di monoterpene sintasi note.

### 3.2.6 – Disegno dei primer specifici

Avendo individuato 3 putative sequenze di monoterpene sintasi, si è proceduto al disegno di primer specifici per amplificare selettivamente ciascuno di questi geni.

Per il disegno di tutti i primer è stato utilizzato il programma *Primer3* disponibile in Internet al sito <http://frodo.wi.mit.edu/>.

Sono stati disegnati innanzitutto dei primer interni per effettuare PCR di controllo sulle sequenze ricavate dai primi assemblaggi fin qui descritti.

Sono stati poi disegnati svariati primer per l'amplificazione delle estremità 3' e 5' (RACE). Per ogni amplificazione è stata disegnata una coppia di primer: uno più interno al gene e uno esterno, più vicino all'estremità. Questo ha reso possibile effettuare una prima amplificazione, che è poi stata a sua volta ri-amplificata con il primer interno ("nested PCR"): questo processo di amplificazioni sequenziali aumenta notevolmente la specificità della reazione.

In alcuni casi queste prime RACE non sono state sufficienti a raggiungere l'estremità del trascritto, a causa di una sua lunghezza eccessiva che non è stata coperta interamente dalla corsa di sequenziamento. Si è allora proceduto con un processo iterativo: le nuove sequenze sono state assemblate con i contig preesistenti in modo da estendere la regione nota. Sulla base di queste nuove informazioni sono stati disegnati primer sempre più vicini alle estremità, e così fino al completamento delle sequenze *full-length* dei 3 geni in esame.



### 3.3 - Ottenimento delle sequenze complete

#### 3.3.1 – Sintesi del cDNA con tecnica SMART

Per la retrotrascrizione in cDNA dell'RNA totale è stato usato il protocollo “BD SMART<sup>TM</sup> RACE cDNA Amplification Kit” (Clontech, Takara Bio Company) modificato. Questo sistema consente di ottenere cDNA di elevata qualità, con un'elevata percentuale di trascritti a sequenza completa. La trascrittasi inversa (RT) non è sempre in grado di trascrivere l'intera sequenza di mRNA, e le estremità 5' dei geni tendono ad essere sottorappresentate nella popolazione di cDNA. Questa tecnica sopperisce a questo inconveniente, sfruttando una caratteristica intrinseca della RT, combinata con un oligonucleotide specifico.

Quando la RT raggiunge l'estremità 5' dell'mRNA, aggiunge alcuni nucleotidi addizionali, generalmente Citosine, al 3' del cDNA. Questo fenomeno è il cosiddetto “dC tailing”, ed è favorito dalla presenza di MnCl<sub>2</sub>.

L'oligonucleotide SMART T7 possiede delle ribo-Guanosine al 3' che si possono appaiare alle Citosine della prima elica del cDNA, creando un'estensione del template. Quando la trascrittasi inversa incontra questo oligo, cambia filamento stampo e continua la replicazione fino alla fine dell'oligonucleotide, introducendo così una coda a sequenza nota al 3' del cDNA, che potrà essere utilizzata come sito di priming per l'amplificazione dell'estremità 5' del gene.

Nel caso invece in cui l'enzima interrompesse la trascrizione prima della fine del template, l'aggiunta di Citosine è molto meno efficiente che per gli ibridi *full-length* cDNA-RNA, per cui nella popolazione di cDNA i trascritti incompleti saranno sottorappresentati.

Il primo filamento di cDNA viene diluito e usato come stampo per la RACE; ciò riduce notevolmente la sintesi di prodotti aspecifici e incompleti. L'amplificazione dei cDNA, specie dei *full-length*, risulta più efficiente e fedele rispetto alle PCR convenzionali. La tecnologia BD SMART<sup>TM</sup> RACE è riassunta in Fig. 3.2, e di seguito viene illustrato il protocollo usato.

Come materiale di partenza è stato usato l'RNA totale estratto da 6 tessuti: fiore, drupa a 10, 15 e 25 settimane di maturazione, drupa matura, e seme essiccato.

L'RNA di fiori e semi è stato ricavato da campioni di *C. arabica* cv. Bourbon Red; tutti gli RNA di drupa sono stati estratti da piante di *C. arabica* cv. Catuai Red.

Primer utilizzati:

Oligonucleotide	Sequenza (5'-3')
attB2_oligodT	biotin-GGG GAC CAC TTT GTA CAA GAA AGC TGG GCG GCC GC(T) <sub>20</sub> VN
SMART_T7	ACT CTA ATA CGA CTC ACT ATA GGG A rGrGrG

- Preparare la seguente mix:

Reagente	Volume	Conc. finale
RNA totale	q.b.	3 µg
attB2oligodT 10µM	1 µl	2 µM
H <sub>2</sub> O	a volume	-
TOTALE		5 µl

- Denaturare 5' a 65°C, mettere immediatamente in ghiaccio.
- Preparare la reazione di sintesi del cDNA:

Reagente	Volume	Conc. finale
RNA	5 µl	-
Super Script II Reverse Trascriptase	1 µl (20 U)	1U/µl
First Strand Buffer 5x	4 µl	1x
Oligo SMART-T7-rGGG	0.5 µl	0.3 µM
DTT 0.1M	1 µl	5 mM
dNTPs (10 mM ciascuno)	0.5 µl	250 µM
H <sub>2</sub> O	15 µl	-
TOTALE	20 µl	

- Incubare a 42°C per 1h.
- Aggiungere 1µl di MnCl<sub>2</sub> 20mM
- Incubare a 42°C per altri 15'
- Diluire con H<sub>2</sub>O fino ad un volume finale di 100µl

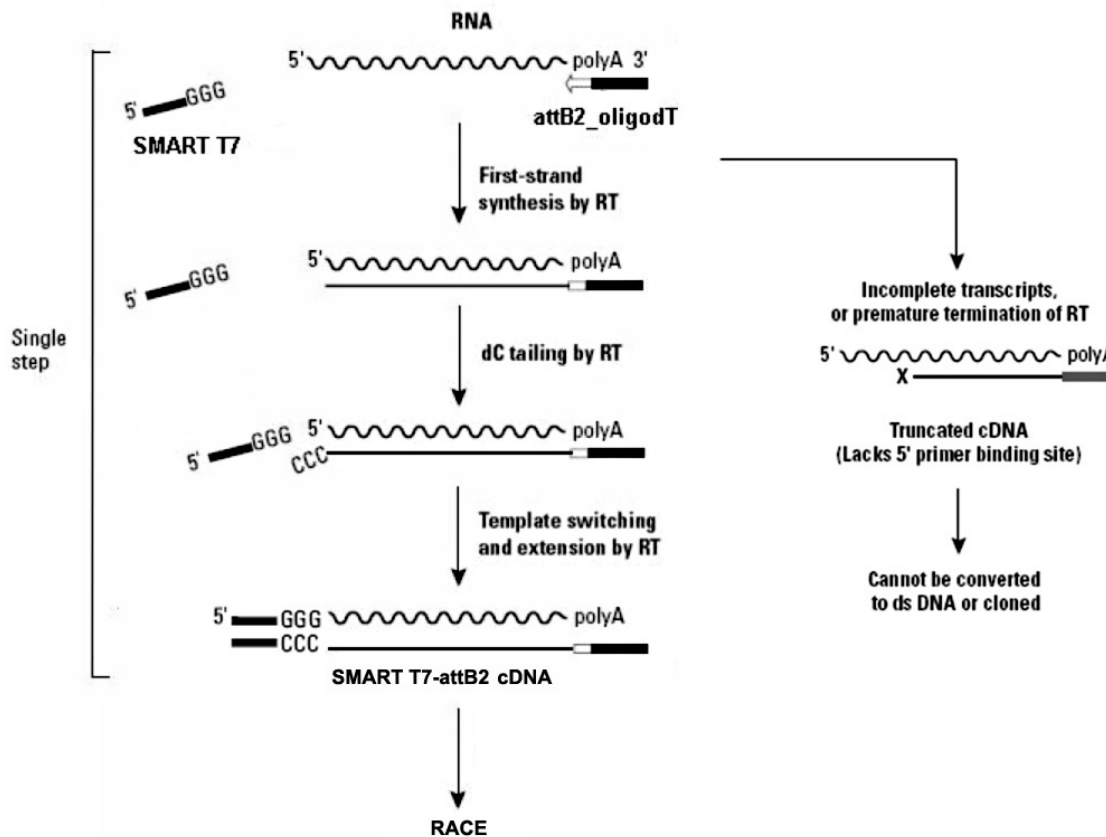


Fig. 3.2 - Schema della tecnologia BD-SMART™ RACE (adattato da Fig.1. in Clontech SMART™ RACE cDNA Amplification Kit User Manual, Clontech Laboratories Inc., Takara Bio Company).

### 3.3.2 – PCR da cDNA con primer specifici

Per verificare se ed in quali tessuti siano espressi i geni per le putative terpene sintasi, è stata fatta una PCR dai cDNA con i primer specifici disegnati al punto 3.2.6. Come controllo positivo per accertarsi che la sintesi del cDNA sia avvenuta in modo corretto, è stata fatta un'ulteriore amplificazione con primer per la poliubiquitina esomerica, gene *housekeeping* espresso in tutti i tessuti.

Primer utilizzati:

Oligonucleotide	Sequenza (5'-3')
Ubiquitina_For	GCA GAT TTT CGT CAA GAC ATT GAC TG
Ubiquitina_Rev	ACC AAG GCA AGT TCA GTT CAA CAC

Miscela di reazione:

Reagente	Volume	Concentrazione finale
10X PCR Buffer (100mM Tris-HCl pH 9, 500mM KCl, 1% Triton X100)	2.5 µl	1X
MgCl <sub>2</sub> 25mM	2.5 µl	2.5 mM
DNTP mix (dATP, dCTP, dGTP, dTTP, 10 mM ciascuno)	0.5 µl	200 µM ciascuno
Primer Forward 10 µM	0.5 µl	200 nM
Primer Reverse 10 µM	0.5 µl	200 nM
Taq DNA Polimerasi	0.5 µl ≈ 0.5U	0.025 U/µl
cDNA	1 µl ≈ 25ng	1 ng/µl
H <sub>2</sub> O	17 µl	-
TOTALE	25 µl	

Condizioni di amplificazione:

Step	Temperatura	Durata	n° cicli
Denaturazione iniziale	94° C	3'	1
Denaturazione	94° C	30"	} 35
Annealing	58° C	30"	
Estensione	72° C	1'	
Estensione finale	72° C	10'	1

### 3.3.3 – RACE 3' e 5'

Per l'amplificazione della regione 3' dei geni (Fig. 3.3), è stato usato come primer Rev l'attB2 adapter, che si appaia alla sequenza attB2 introdotta accanto al poli-A dall'oligo attB2-dT usato nella sintesi del cDNA. Come primer For sono stati usati i primer specifici per ogni gene, disegnati al punto 3.2.6.

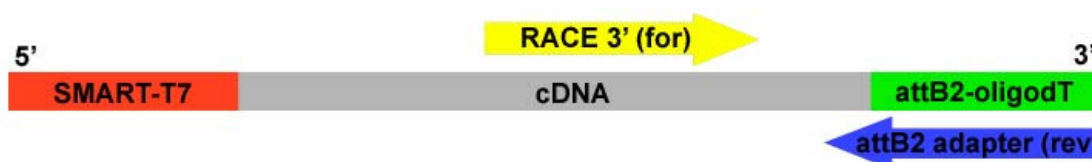


Fig.3.3 – rappresentazione schematica della tecnica RACE per l'amplificazione delle estremità 3'

Simmetricamente, per l'amplificazione della regione 5' dei geni (Fig. 3.4), è stato usato come primer For l'adattatore 12attB1-T7, che si appaia alla sequenza T7 introdotta al 5' dal primer SMART usato nella sintesi del cDNA. Come primer Rev sono stati usati i primer specifici per ogni gene, disegnati al punto 3.2.6.

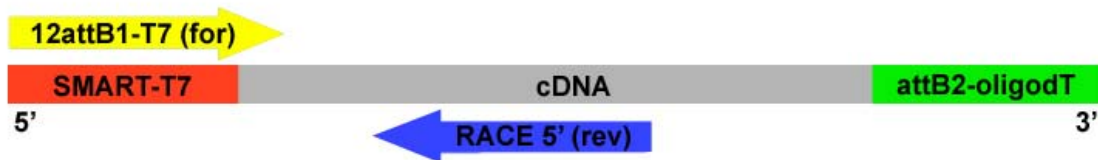


Fig.3.4 – rappresentazione schematica della tecnica RACE per l'amplificazione delle estremità 5'

Si riportano di seguito le sequenze di tutti i primer usati:

Oligonucleotide	Sequenza (5'-3')
12attB1-T7	AAA AAG CAG GCT AAT ACG ACT CAC TAT AGG GA
attB2 adapter	GGG GAC CAC TTT GTA CAA GAA AGC TGG G
<i>Primer RACE per il trascritto CaMTS 1</i>	
LiS_C1aAr_FOR	GGAGGCAAGGTGGTTCATAG
LiS_C1aAr_REV	CATTCCATGTTGAGGGTCA
LiS_C1aAr_3END	GAGCAGTACAGGCCTAGCAGA
LiS_C1aAr_3ENDnxt	GACCTCAACATGGAAATGC
LiS_C1aAr_5END	GTGGAAGCTCCAAAGCATGT
LiS_C1aAr_5ENDnxt	TGCTTCGAGTCTTGATCAATTC
C1a_5'RACE_nx2	CACAAGCTTGAAATCCCAACTG
C1a_3'_seq	CGGCTAGCAAATGATTTAGGAAC
<i>Primer RACE per il trascritto CaMTS 4</i>	
LiS_C4Ar_FOR	ACGAGAGAAGTCCCAACGTG
LiS_C4Ar_REV	CCGTACTGAGGCTCGAAGAC
LiS_C4Ar_3END	ACATACGAATCGCGGAAAAG
LiS_C4Ar_3ENDnxt	GGCAATAGGAGCTGTCTTCG
LiS_C4Ar_5END	CACCTCAGGGGAAGCTCTAGT
LiS_C4Ar_5ENDnxt	AATTATTGAGGTGCTTAGTTGTGAA
C4Ar_walk	CACCAGGGGAGATGAAAAGA
<i>Primer RACE per il trascritto CaMTS Cc</i>	
LiS_C1Ca_FOR	CCAAGCAGCACATCAGAATG
LiS_C1Ca_REV	CCACGTTGAGGGTCTGAAAT
LiS_C1Ca_REV2	GGTACATGCATAGTCCCATCC
CAN_FOR_2	AATCTGGAAATGGCAAGACG
CAN 3' RACE	TGTACCAAAATCGATACGATGC
CAN 5' RACE	ATCAAAGCAACGACTTTTGTCA
CAN_5'RACE_nx	CGTCAAAACCATGCTGCCTA
CaMTS_Cc_5'	CAAGAATTTCCGAAAACCTTAGCA

Per l'amplificazione è stata usata la Taq polimerasi *Advantage* (Clontech, Takara Bio Company)

Miscela di reazione:

Reagente	Volume	Concentrazione finale
10X <i>Advantage</i> Buffer	2.5 µl	1X
DNTP mix (dATP, dCTP, dGTP, dTTP, 10 mM ciascuno)	0.5 µl	200 µM ciascuno
Primer Forward 10 µM	0.5 µl	200 nM
AttB2 adapter (reverse) 10 µM	0.5 µl	200 nM
<i>Advantage</i> DNA Polimerasi	0.2 µl ≈ 1U	0.04 U/µl
cDNA	1 µl	-
H <sub>2</sub> O	17 µl	-
TOTALE	25 µl	

Per l'amplificazione è stato usato un programma di tipo “*Touch-down*” ovvero con temperatura di annealing variabile, inizialmente molto elevata (63°C) e poi gradualmente calante, fino a stabilizzarsi a 55°C. In questo modo i primi cicli vengono compiuti ad alta stringenza sintetizzando dei prodotti il più possibile specifici; questi vengono poi amplificati ulteriormente a stringenza medio-bassa per massimizzare l'efficienza del resto della reazione. La temperatura di estensione è stata fissata a 68°C (invece degli usuali 72°C) come indicato dal manuale della Taq *Advantage*.

Condizioni di amplificazione:

Step	Temperatura	Durata	n° cicli
Denaturazione iniziale	94° C	5"	1
Denaturazione	94° C	30"	} 4
Annealing	63° C	30"	
Estensione	68° C	1' 30"	
Denaturazione	94° C	30"	} 4
Annealing	61° C	30"	
Estensione	68° C	1' 30"	
Denaturazione	94° C	30"	} 4
Annealing	59° C	30"	
Estensione	68° C	1' 30"	
Denaturazione	94° C	30"	} 4
Annealing	57° C	30"	
Estensione	68° C	1' 30"	
Denaturazione	94° C	30"	} 20
Annealing	55° C	30"	
Estensione	68° C	1' 30"	
Estensione finale	68° C	10'	1

### 3.3.4 – Nested PCR

Nei casi in cui la prima amplificazione tramite RACE non ha prodotto una banda sufficientemente buona per essere clonata e/o sequenziata direttamente, è stata svolta un'amplificazione sequenziale (nested PCR): il prodotto RACE è stato innanzitutto diluito (da 20 a 100 volte, in proporzione variabile a seconda dell'intensità della banda) e poi amplificato una seconda volta con un primer "nested", ovvero qualche decina di basi più interno rispetto al primer originale. Dato che la reazione va ad agire su un DNA già amplificato, deve essere estremamente stringente e con un basso numero di cicli.

Miscela di reazione:

Reagente	Volume	Concentrazione finale
10X PCR Buffer (100mM Tris-HCl pH 9, 500mM KCl, 1% Triton X100)	2.5 µl	1X
MgCl <sub>2</sub> 25mM	2.5 µl	2.5 mM
DNTP mix (dATP, dCTP, dGTP, dTTP, 10 mM ciascuno)	0.5 µl	200 µM ciascuno
Primer Forward 10 µM	0.5 µl	200 nM
Primer Reverse 10 µM	0.5 µl	200 nM
Taq DNA Polimerasi	0.5 µl ≈ 0.5U	0.025 U/µl
Prodotto RACE diluito	1 µl	-
H <sub>2</sub> O	17 µl	-
TOTALE	25 µl	

Condizioni di amplificazione:

Step	Temperatura	Durata	n° cicli
Denaturazione iniziale	94° C	3'	1
Denaturazione	94° C	30"	} 20
Annealing	65° C	30"	
Estensione	72° C	1' 30"	
Estensione finale	72° C	10'	1

### 3.3.5 – Sequenziamento ed assemblaggio dei trascritti completi

Tutte le amplificazioni sopradescritte sono state controllate in gel di agarosio all'1% in tampone TBE (Tris-HCl 80mM, H<sub>3</sub>BO<sub>3</sub> 80mM, EDTA 2.55mM, pH 8.8), con etidio bromuro 0.5 µg/ml.

Le amplificazioni di buona qualità (unica banda nitida) sono state sequenziate direttamente; quelle che presentavano smear o bande multiple sono state clonate (se necessario dopo essere state purificate da gel con kit "GFX PCR DNA and gel band purification kit" (Amersham Biosciences), selezionate come descritto ai paragrafi 3.2.3 e 3.2.4, ed infine sequenziate. Ove necessario, per lo stesso frammento sono stati utilizzati più primer, fino ad ottenerne la sequenza completa.

Tutte le sequenze sono state eseguite dal Servizio di Sequenziamento del centro BMR (<http://www.bmr-genomics.it/>). Le sequenze sono state poi assemblate con il programma *Seqman* del pacchetto Lasergene (DNA Star Inc.), per unirle ai frammenti genici precedentemente isolati, e ricostruire così la sequenza completa.

I consensus così ottenuti sono stati analizzati mediante i programmi blastn, blastx e tblastx, per verificare il loro livello di similarità a sequenze di monoterpene sintasi note.

### 3.4 - Analisi filogenetica delle monoterpene sintasi di *C. arabica*

Oltre alla semplice analisi di similarità, è stata svolta un'analisi ad ampio spettro con il fine di stabilire la distanza filogenetica tra i geni scoperti in *C. arabica* ed le sequenze di terpene sintasi vegetali già note e presenti in rete.

A questo scopo è stata innanzitutto costruita una collezione di sequenze specifica, ottenuta raccogliendo le sequenze proteiche di terpene sintasi del regno *Viridiplantae* dai database pubblici (<http://www.ncbi.nlm.nih.gov/sites/gquery> e <http://www.ebi.ac.uk>). Per avere una popolazione campione il più possibile rappresentativa ma non ridondante, i geni sono stati selezionati secondo i seguenti criteri:

- includere geni appartenenti a tutte le classi di terpene sintasi (mono-, di- e sesqui-terpene sintasi. Non sono state trovate nei database sequenze di tri- e tetra-terpene sintasi)
- includere geni sia da angiosperme che da gimnosperme
- includere almeno un gene per ogni organismo di cui siano note sequenze di terpene sintasi
- scartare sequenze ridondanti e isoforme dello stesso gene
- scartare sequenze putative, ipotetiche o incomplete

Alle sequenze proteiche così individuate sono state aggiunte le 3 sequenze di *C. arabica* isolate nel corso del progetto. L'insieme di questi dati è stato sottoposto ad analisi tramite il programma MEGA Molecular Evolutionary Genetics Analysis, ver.4.1 (Tamura *et al*, 2007; Kumar *et al*, 2004).

Gli alberi filogenetici sono stati assemblati secondo il criterio neighbor-joining con i seguenti parametri (sia per l'allineamento pairwise che per quello multiplo): gap opening penalty 10; gap extension penalty 2; numero di repliche bootstrap 1000.





## 4. RISULTATI E DISCUSSIONE

### 4.1 - Sviluppo di protocolli per l'estrazione di RNA totale da semi di caffè essiccati

#### 4.1.1 – Protocolli sviluppati: CTAB e “Hot Borate”

Per lo studio dei geni espressi nel corso della maturazione del frutto e dello sviluppo del seme, il prerequisito fondamentale è la disponibilità di un RNA totale di buona qualità. Presso il nostro laboratorio è in uso un protocollo sviluppato internamente, basato sul metodo di Chang, Puryear e Cairney (1993) opportunamente modificato, che si è dimostrato adatto a ottenere RNA di buona qualità da tessuti freschi molli (foglie, fiori, frutti). Purtroppo, il medesimo protocollo si è rivelato inefficace quando applicato ai semi di caffè essiccati per uso commerciale. I principali problemi riscontrati sono stati una resa piuttosto bassa, e soprattutto un alto livello di impurezze, rappresentate soprattutto da carboidrati. E' stato quindi deciso di sperimentare alcune metodiche specifiche per l'estrazione di RNA totale da questo tessuto.

Il cosiddetto “caffè verde” è un materiale a basso contenuto in acqua ricco in carboidrati, composti fenolici (in gran parte acidi clorogenici) e lipidi, che sono le principali classi di sostanze che interferiscono con le estrazioni degli acidi nucleici. I composti fenolici tendono ad ossidarsi a chinoni, che si legano covalentemente agli acidi nucleici degradandoli ed abbassando la resa dell'estrazione (Loomis, 1974). I polisaccaridi invece, nelle condizioni di estrazione tipiche dei protocolli standard, hanno caratteristiche fisico-chimiche simili all'RNA, e di conseguenza tendono a essere co-estratti e co-precipitati con l'acido ribonucleico, contaminando le estrazioni. I lipidi si associano a proteine e carboidrati rendendone più difficile la separazione dall'RNA. Infine, la consistenza legnosa del seme ostacola notevolmente la sua manipolazione meccanica, rendendo necessario un laborioso sminuzzamento manuale con un bisturi, prima di poterlo omogenizzare.

Per ovviare a queste difficoltà, il protocollo di Chang *et al* (1993) basato sull'uso di CTAB, è stato modificato aggiungendo al buffer di estrazione 0.5 mg/ml di Proteinasi K e un'incubazione a 55°C per 45 minuti, ottimizzando le condizioni proposte da Azevedo *et al* (2003). Questo protocollo si è rivelato piuttosto affidabile nel produrre un RNA di buona qualità, eventualmente effettuando una doppia precipitazione in LiCl per rimuovere eccessive contaminazioni da polisaccaridi. Gli svantaggi di questo metodo sono una resa di RNA totale non altissima, e suscettibile di ampie oscillazioni dovute a fattori contingenti come la manualità dell'operatore e la qualità dell'omogenizzazione. Altri inconvenienti minori sono i grandi volumi di reazione utilizzati (10ml di tampone per 2 semi/300mg di tessuto) e l'utilizzo di solventi organici ad elevata tossicità.

Per queste ragioni sono stati sperimentati alcuni protocolli alternativi, basati su sali di guanidinio (Sharma *et al* 2003, Singh *et al* 2003) o sull'uso di urea (Tai *et al*, 2004) che non hanno dato risultati soddisfacenti e sono stati abbandonati. Un metodo che invece ha avuto successo è basato sul metodo “Hot borate” di Wan e Wilkins (1994), come modificato da Birtic e Kranner (2006). Questa procedura è leggermente più laboriosa

dell'estrazione con CTAB, in quanto include un maggior numero di passaggi di centrifugazione e precipitazione. Ha però il grande vantaggio pratico ed economico di utilizzare volumi di reazione 10 volte minori (1ml di tampone per 1 seme/150mg), con un notevole risparmio di reagenti e una facilità di manipolazione molto maggiore, senza inoltre far uso di solventi organici pericolosi.

#### 4.1.2 – Qualità dell'RNA estratto

La tabella 4.1 illustra i risultati ottenuti con i vari protocolli di estrazione. Vengono confrontati il protocollo originario secondo Chang *et al* (senza Proteinasi K), il protocollo CTAB con Proteinasi K, e il protocollo Hot Borate. Nel caso del protocollo CTAB, quando la purezza dell'RNA è scarsa, può essere conveniente ripetere la precipitazione in LiCl. In tal modo, a fronte di una perdita di circa il 50% della resa, la purezza migliora sensibilmente, in particolare per quanto riguarda la contaminazione da carboidrati.

Metodo di estrazione	Resa ( $\mu\text{g RNA/g semi}$ )	Purezza dell'RNA A260 / A280	Purezza dell'RNA A260 / A230
<b>CTAB</b> (Chang <i>et al</i> 1993) $n=8$	15.98 $\pm$ 5.84	1.86 $\pm$ 0.16	0.71 $\pm$ 0.43
<b>CTAB + Proteinase K</b> (Azevedo <i>et al</i> 2003) $n=39$	83.02 $\pm$ 17.87	2.18 $\pm$ 0.05	1.99 $\pm$ 0.27
<b>CTAB + Proteinase K</b> (doppia precipitazione LiCl) $n=5$	32.12 $\pm$ 7.35	2.20 $\pm$ 0.08	2.29 $\pm$ 0.58
<b>Hot borate</b> (Birtic Kranner 2006) $n=6$	158.45 $\pm$ 8.58	2.22 $\pm$ 0.14	2.33 $\pm$ 0.06

Tab. 4.1 – Resa e purezza dell'RNA totale ottenuto da semi essiccati di caffè, tramite i metodi descritti. I risultati sono espressi come media  $\pm$  deviazione standard di  $n$  repliche biologiche.

Per verificare, oltre alla quantità e purezza, anche l'integrità e la qualità dell'RNA totale ai fini degli utilizzi successivi, i campioni sono stati controllati in gel denaturante ed è stata effettuata una RT-PCR di controllo utilizzando dei primer per la poliubiquitina esomerica, gene *housekeeping* espresso in tutti i tessuti.

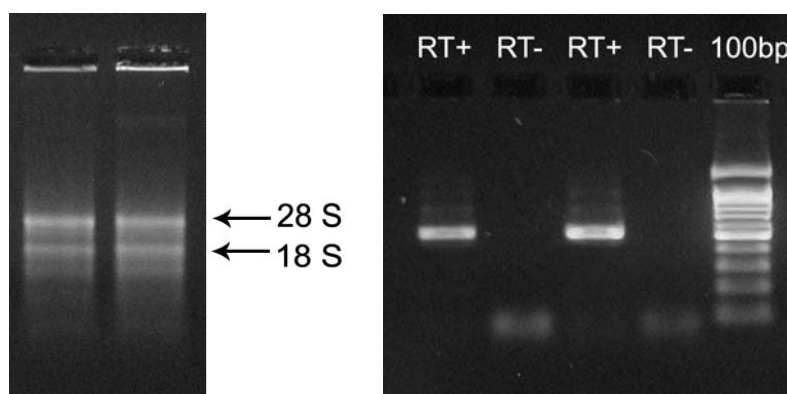


Fig. 4.2 – Gel denaturante e RT-PCR di campioni di RNA estratti da semi secchi, con protocollo CTAB+Proteinase K.

In Fig. 4.2 si mostra il controllo di due estrazioni effettuate con il protocollo CTAB. Si può apprezzare come l'RNA totale sia integro, essendo ben evidenti le bande dell'RNA ribosomiale 28 S e del 18 S. La RT-PCR ha avuto successo, ed il fatto che i controlli negativi (RT-) non presentino alcuna banda indica che l'RNA non è contaminato da DNA genomico.

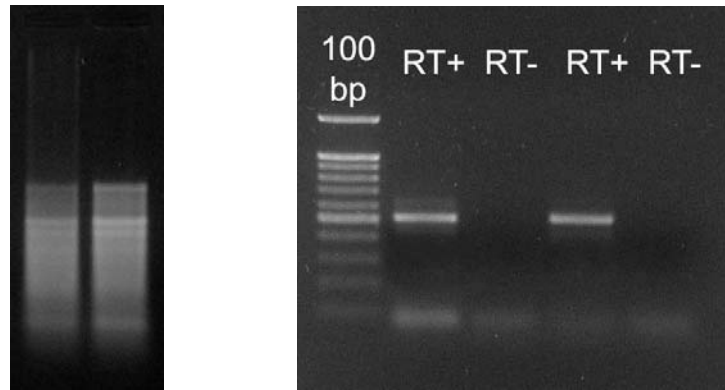


Fig. 4.3 – Gel denaturante e RT-PCR di campioni di RNA estratti da semi secchi, con protocollo Hot Borate.

La Fig. 4.3 riporta invece il controllo dei campioni estratti con protocollo Hot Borate. Nella corsa in gel denaturante non sono chiaramente distinguibili le bande dei ribosomiali, segno che l'RNA è parzialmente degradato. Tale degradazione non è comunque tale da compromettere la retrotrascrizione e l'amplificazione, che difatti sono riuscite e non mostrano segni di contaminazione da DNA genomico.

In definitiva, si può concludere che entrambi i protocolli CTAB e Hot Borate, ottimizzati per i semi essiccati di caffè verde, risultano validi e danno un RNA totale di buona qualità. Il protocollo CTAB è stato quello meglio testato, e per questo usato routinariamente nel corso della sperimentazione. Il protocollo Hot Borate ha fornito un RNA leggermente degradato, ma comunque utilizzabile per le successive applicazioni, e decisamente superiore in termini di resa e pulizia da contaminazioni.

## 4.2 - Isolamento di 3 putative monoterpene sintasi di *C. arabica* (CaMTS)

### 4.2.1 – Analisi di similarità e disegno dei primer degenerati

A causa della mancanza di qualunque informazione sulle monoterpene sintasi di *C. arabica*, l'unico modo di isolare dei geni di questa famiglia dalla pianta di caffè è l'utilizzo di primer degenerati, basati su sequenze conservate identificate in altre piante meglio conosciute.

In primo luogo è stata quindi svolta una ricerca preliminare in banca dati sui geni di questa classe già noti in altri organismi vegetali. Data la vastità della famiglia delle monoterpene sintasi e la loro varietà per sequenza nucleotidica e struttura genica, è stato deciso di concentrarsi su una sola proteina che sintetizzasse un prodotto il più possibile specifico e rappresentativo; a questo scopo è stata selezionata la *linalolo sintasi*.

Tra le linalolo sintasi presenti nei database (una ventina) è stata arbitrariamente scelta quella di *Perilla frutescens*, che è poi stata usata come sonda per cercare sequenze simili, restringendo la ricerca alla SottoClasse *Asterids*, cui il genere *Coffea* appartiene, con l'aggiunta delle sequenze di *Vitis vinifera* (organismo di cui è recentemente stato completato il genoma, che si è rivelato molto ricco di geni della famiglia delle terpene sintasi).

Le sequenze proteiche così individuate sono state poi allineate allo scopo di identificare le regioni più adatte alla costruzione di primer, come illustrato in Fig. 4.4. Sono state individuate 4 regioni conservate in posizioni adatte a disegnare 2 primer forward e 2 reverse, in modo da ottenere prodotti PCR di lunghezza stimata compresa tra le 200 e le 1000 paia di basi.

Le sequenze di queste 4 zone conservate sono state poi confrontate con i database nucleotidici EST e nr (not redundant). In questo modo, per ogni regione è stato ottenuto il maggior numero possibile di sequenze nucleotidiche, su cui sono stati manualmente disegnati i primer degenerati come schematizzato in Fig. 4.5.



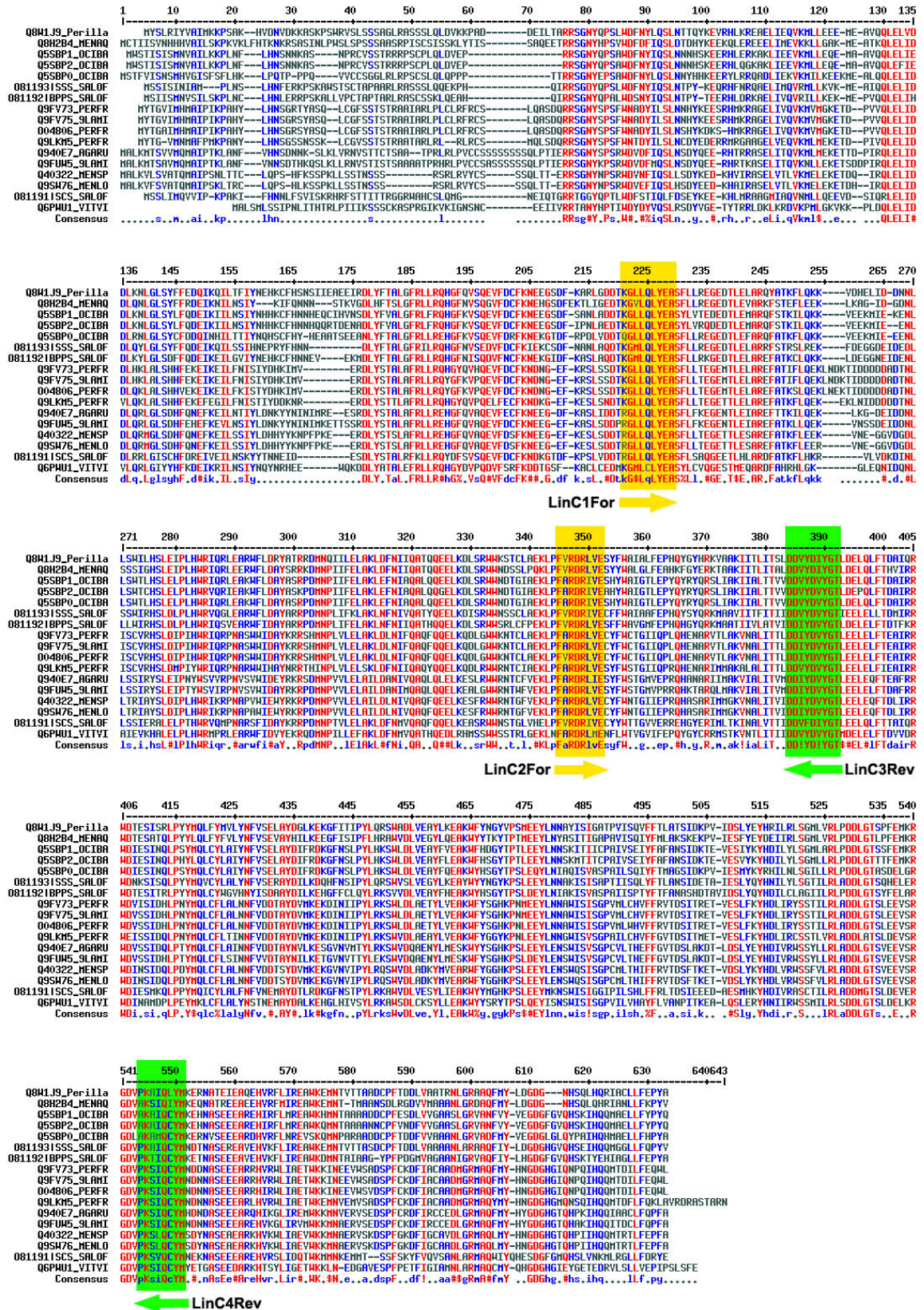


Fig. 4.4 – Allineamento delle sequenze aminoacidiche di linalolo sintasi ed altre monoterpene sintasi: in rosso le regioni ad alta similarità (>90%), in blu quelle a similarità media (>50%). I box colorati evidenziano le regioni in corrispondenza delle quali sono stati disegnati i primer.

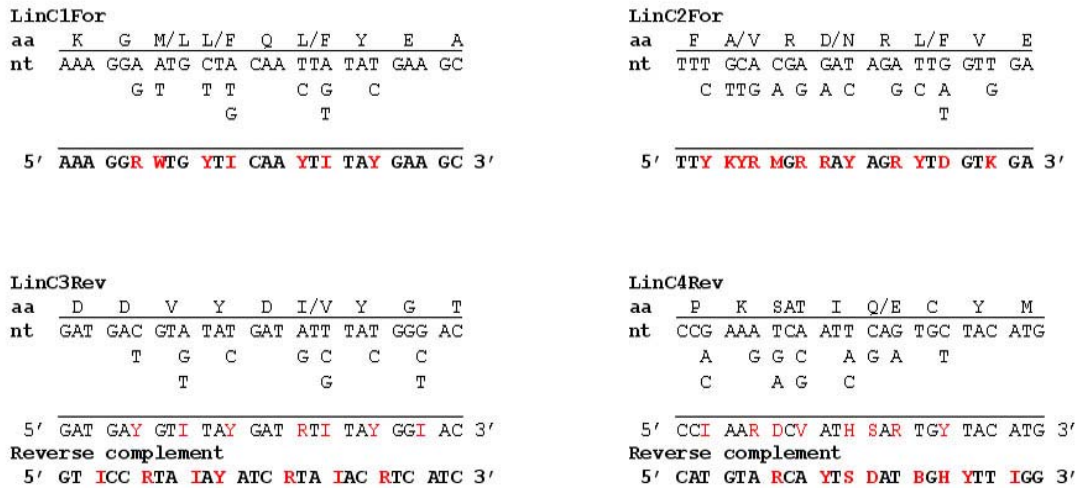


Fig. 4.5 – disegno dei primer degenerati (in rosso le basi degenerate).

#### 4.2.2 – PCR con primer degenerati da DNA genomico

Per verificare la funzionalità dei primer degenerati, è stata effettuata una PCR da DNA genomico, utilizzando tutte le 4 combinazioni possibili di primer, come schematizzato dalla tabella sottostante. Per cercare la massima variabilità, aumentando così la probabilità di individuare geni diversi, è stato amplificato DNA di più varietà di *C. arabica*, in particolare il cultivar Caturra e semi etiopici provenienti dalle regioni Yirga ed Agaro.

Combinazione	Primer Forward	Primer Reverse	Dimensioni stimate dell'amplificato
A	LinC1For	LinC3Rev	540 bp
B	LinC1For	LinC4Rev	975 bp
C	LinC2For	LinC3Rev	171 bp
D	LinC2For	LinC4Rev	630 bp

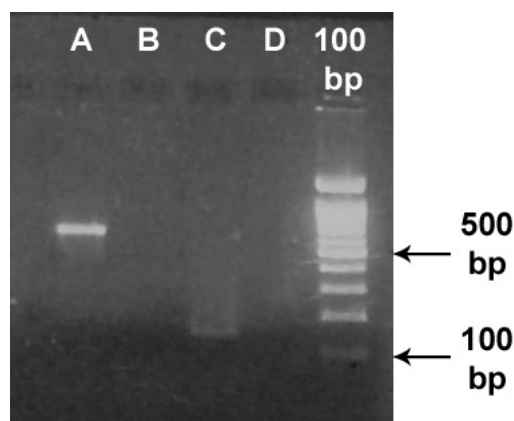


Fig. 4.6 – PCR con le 4 combinazioni di primer degenerati.

In Fig. 4.6 è rappresentata la corsa elettroforetica di una delle PCR. Come si può notare, il primer LinC4Rev non ha dato alcun prodotto di amplificazione. Questo è stato confermato in tutte le successive PCR tentate con questo primer, il cui uso è stato di conseguenza abbandonato. La combinazione LinC2For–LinC3Rev ha dato un prodotto delle dimensioni attese (circa 150bp), mentre la combinazione LinC1For–LinC3Rev mostra una banda di approssimativamente 650bp, circa un centinaio di basi più lunga del previsto. Trattandosi di un'amplificazione da DNA genomico, è lecito ipotizzare che possa trattarsi di un piccolo introne.

#### 4.2.3 – Clonaggio delle PCR ed amplificazione degli inserti

Le PCR ottenute mediante i primer degenerati non possono essere sequenziate direttamente, in quanto è molto probabile che ogni reazione abbia amplificato più sequenze tra loro simili. Per isolare le sequenze l'una dall'altra è stato quindi necessario clonare gli amplificati, mediante il kit TOPO-TA Cloning® (Invitrogen).

Dopo l'amplificazione, sono stati giudicati positivi i cloni di lunghezza corrispondente alla PCR originaria, con l'aggiunta di circa 160 bp dovute al vettore. Le lunghezze attese erano dunque:

LinC2For–LinC3Rev  $\approx$  300 bp

LinC1For–LinC3Rev  $\approx$  800 bp

In Fig. 4.7 e 4.8 si riportano alcuni esempi dei gel di analisi degli inserti amplificati, con evidenziate le bande corrispondenti alla lunghezza prevista.

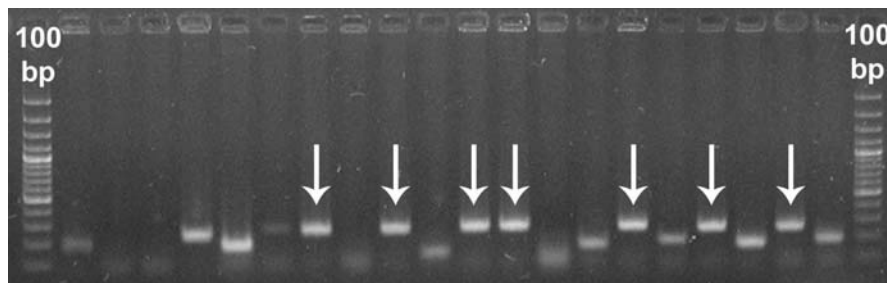


Fig. 4.7 – analisi degli inserti clonati dalla PCR ottenuta con la combinazione di primer degenerati LinC2For–LinC3Rev



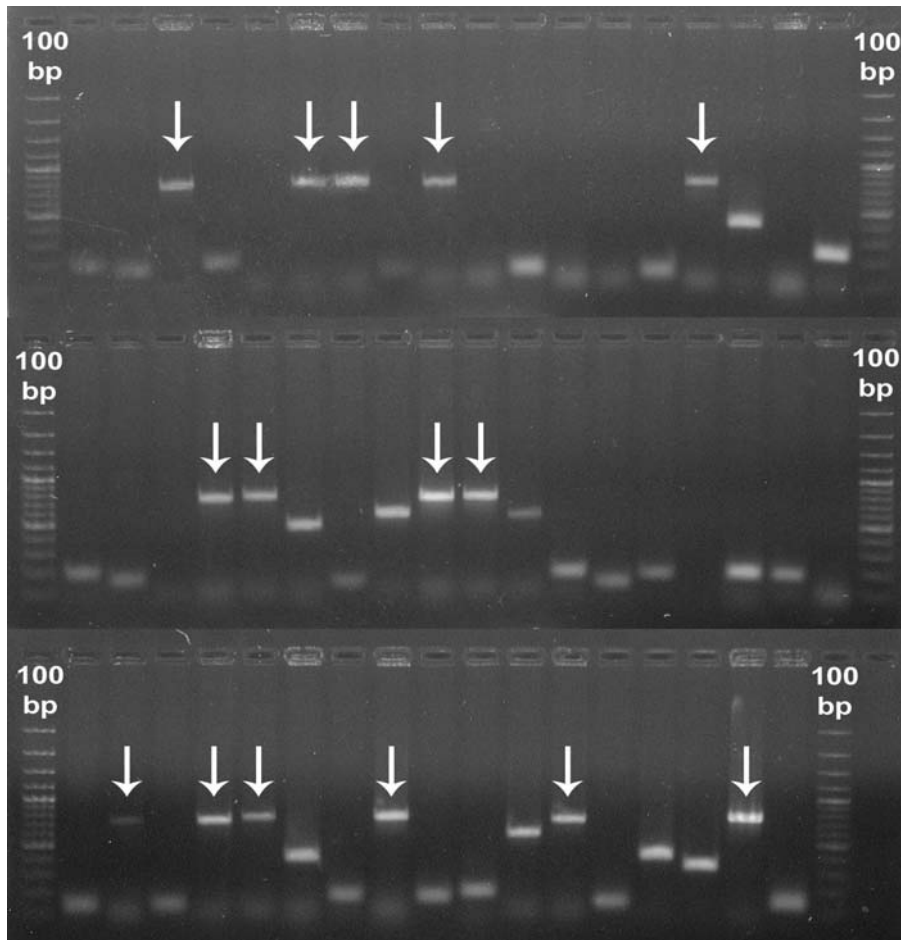


Fig. 4.8 – analisi degli inserti clonati dalla PCR ottenuta con la combinazione di primer degenerati LinC1For–LinC3Rev

#### 4.2.4 – Sequenziamento, assemblaggio ed analisi delle sequenze

Come schematizzato dalla Fig. 4.9, sono stati sequenziati in totale 45 cloni, di cui 36 ottenuti con la coppia LinC1For–LinC3Rev, e 9 con la combinazione LinC2For–LinC3Rev (contenuta all’interno della sequenza precedente).

Le sequenze sono state controllate per eliminare le estremità del vettore e le sequenze non riuscite o di bassa qualità. Sono state poi assemblate in 10 contig, di lunghezza variabile tra le 111 e le 642 bp, e la sequenza consensus di ogni contig è stata confrontata con le sequenze nei database pubblici tramite blastn, blastx e tblastx.



Name	Length	Seqs	Pos	Conflict Split
▼Unlocated Contigs				
CaMts1	630	26	0	?
CaMts2	638	6	0	?
CaMts3	143	3	0	?
CaMts4	642	3	0	?
Contig 4	222	1	0	?
Contig 5	209	1	0	?
Contig 6	157	1	0	?
Contig 10	111	1	0	?
Contig 11	163	1	0	?
genomico	458	2	0	?

Fig. 4.9 – schermata del programma SeqMan, con il riepilogo dell’assemblaggio delle sequenze dei prodotti PCR ottenuti con i primer degenerati

Tramite l’analisi di similarità è stato possibile stabilire che i 4 contig più rappresentati (che insieme raggruppavano 38 sequenze delle 45 totali) presentavano somiglianze significative con sequenze di terpene sintasi. Dei rimanenti 6, 5 erano costituiti da brevi sequenze singole senza similarità significative, mentre l’ultimo presentava deboli omologie con sequenze genomiche non codificanti, ed era quindi il risultato di un’amplificazione aspecifica di DNA genomico.

Le sequenze putativamente classificabili come terpene sintasi sono state ribattezzate con la sigla CaMts (*Coffea arabica* MonoTerpene Sintasi) ed i numeri da 1 a 4 (dal numero del contig di origine).

Il contig **CaMts 1** è risultato quello a maggiore ridondanza, essendo composto da ben 26 sequenze su 45 totali. Interessante notare la presenza di due diversi alleli ciascuno con frequenza esattamente del 50% (13 sequenze su 26 totali, Fig. 4.10).

Alignment of Contig 'CaMts1'

Position: 1

400 410 420 430 440 450 460 470 480 490 630 bp

Translate Consensus ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

04\_yirga\_F1\_clean.abl(1>626) ← ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

11\_raia\_D4\_clean.abl(1>622) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

17\_abdul\_4\_clean.abl(1>625) ← ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

12\_raia\_F5\_clean.abl(1>566) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

02\_yirga\_A3\_clean.abl(1>619) ← ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

03\_yirga\_C1\_clean.abl(1>620) ← ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

LiS\_1\_riseq.abl(1>829) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

LiS\_4.abl(8>805) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

LiS\_2.abl(13>813) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

LiS\_3.abl(15>812) ← ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

LiS\_6.abl(9>805) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

LiS\_8.abl(8>810) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

07\_yirga\_G2\_clean.abl(1>615) ← ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

eli\_7.abl(1>624) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

eli\_2.abl(1>546) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

eli\_6.abl(1>625) ← ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

01a.scf(1>1219) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

06\_yirga\_A2\_clean.abl(1>592) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

08\_yirga\_F3\_clean.abl(1>598) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

13\_raia\_H5\_clean.abl(1>592) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

09\_yirga\_G3\_clean.abl(1>618) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

05\_yirga\_G1\_clean.abl(1>618) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

15\_raia\_D6\_clean.abl(1>617) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

07\_yirga\_C2\_clean.abl(1>608) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

eli\_1.abl(1>580) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

eli\_8.abl(50>576) → ATTATGCAAAATTTTTTAAATCTCTCAAAAACATTTTTCATGTTTTCTGTCCAGATGGTGGAGCAGTACAGGCCCTAGCGAAAAACTACCAITTTGCTA

Fig. 4.10 – allineamento (eseguito col programma SeqMan) delle sequenze del contig CaMts 1, con evidenziate in giallo e verde le due varianti alleliche

La Fig 4.11 riporta i risultati del blastx della sequenza CaMTS 1, con il dettaglio delle accession in banca dati risultate più simili.

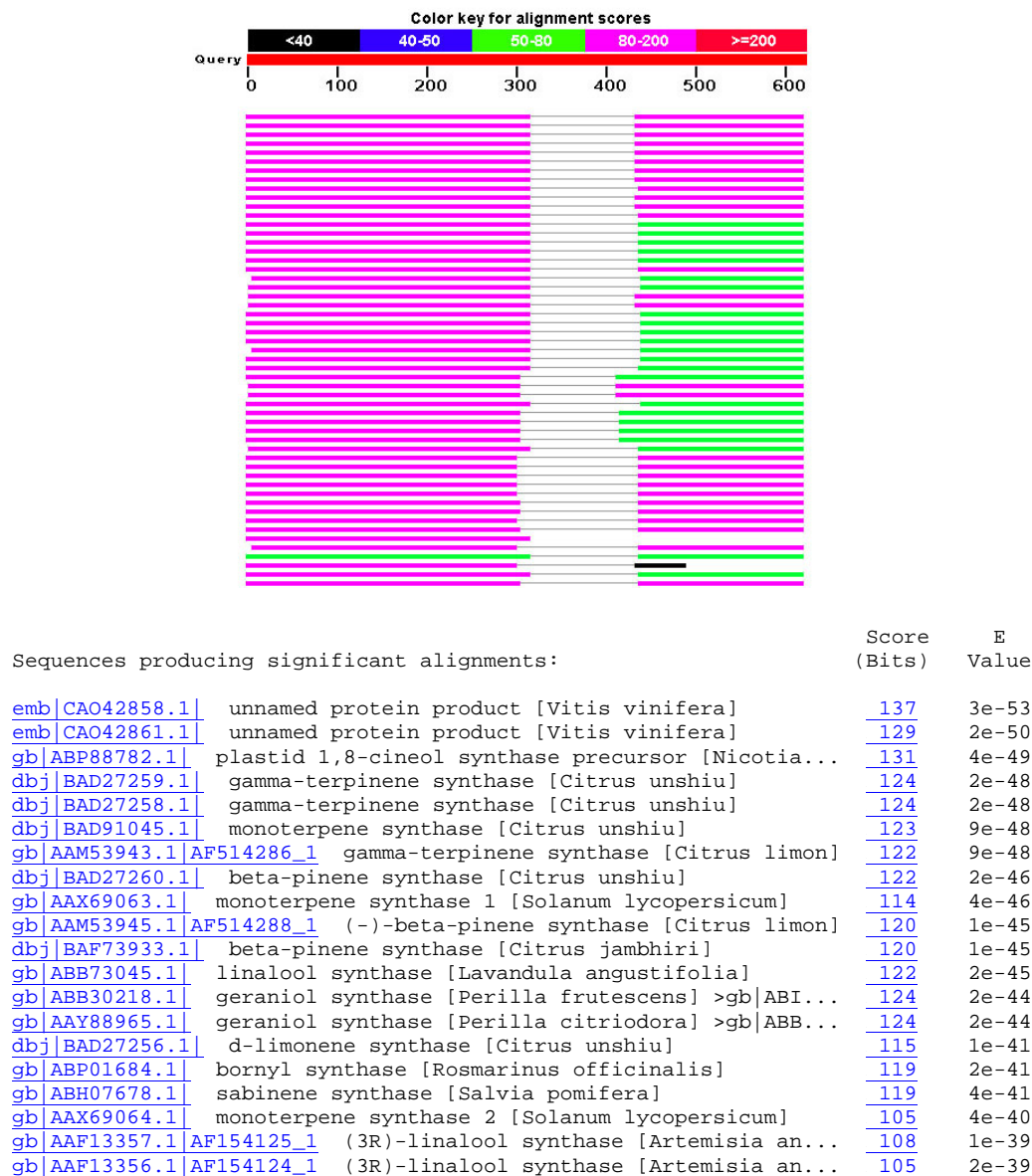


Fig. 4.11 – blastx della sequenza consensus CaMTS 1

Si può notare come le maggiori similarità si abbiano con geni codificanti per monoterpene sintasi (linalool, limonene, pinene, geraniol synthase, etc). Questo ci consente di classificare il gene CaMTS 1 come una putativa monoterpene sintasi. Notare la zona senza similarità al centro della sequenza, che a questo punto viene senza più dubbi identificata come un introne, la cui posizione precisa verrà rintracciata tramite un'analisi accurata dei siti di splicing e si rivelerà lungo 128bp.

Il contig **CaMTS 2**, costituito da 6 sequenze, ha mostrato invece similarità con geni di sesquiterpene sintasi ((+)-delta-cadinene synthase, (-)-germacrene D synthase), come illustrato dalla Fig. 4.12. Anche in questo caso è evidente il “vuoto” corrispondente all'introne.

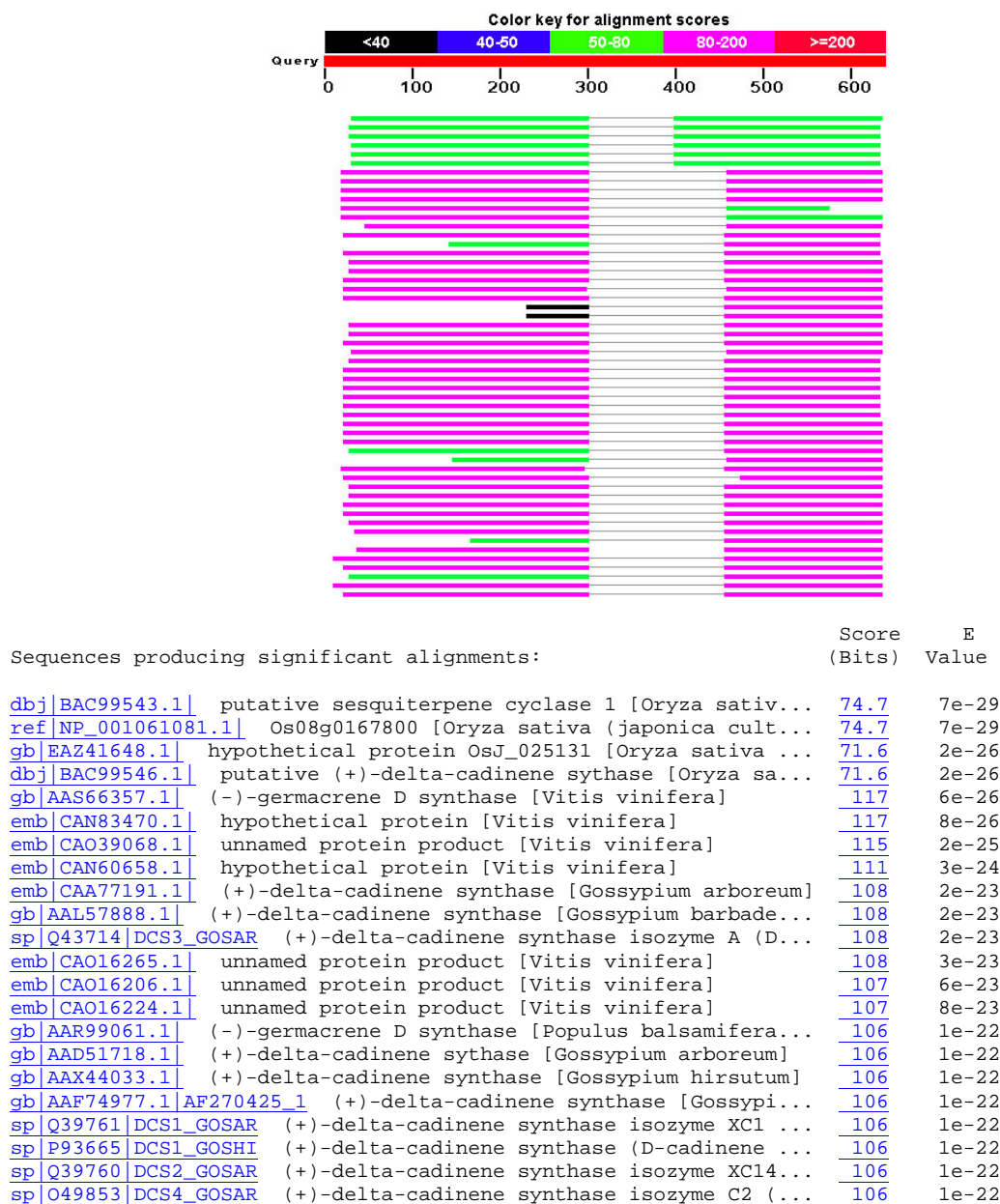
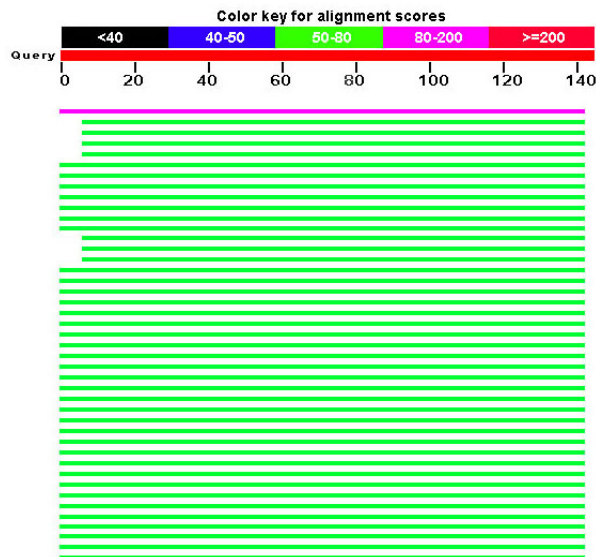


Fig. 4.12 – blastx della sequenza consensus CaMTS 2

Anche il contig **CaMTS 3**, costituito da 3 sequenze, ha mostrato similarità con geni di sesquiterpene sintasi (Fig. 4.13). E' vero che nella lista compaiono anche alcune monoterpene sintasi, ma in questo caso le somiglianze sono piuttosto deboli e confuse, sia per i bassi punteggi E-value, sia per la brevità della sequenza (130 bp contro le 600 circa degli altri contig).



Sequences producing significant alignments:		Score (Bits)	E Value
<a href="#">gb AY860851.1</a>	Artemisia absinthium sesquiterpene synthase ge...	<a href="#">84.9</a>	4e-16
<a href="#">gb EF197888.1</a>	Artemisia annua amorpha-4,11-diene synthase mR...	<a href="#">79.9</a>	1e-14
<a href="#">gb AY006482.1</a>	Artemisia annua amorpha-4,11-diene synthase mR...	<a href="#">79.9</a>	1e-14
<a href="#">gb AF138959.1</a>	<a href="#">AF138959</a> Artemisia annua amorpha-4,11-diene syn...	<a href="#">79.9</a>	1e-14
<a href="#">gb DQ241826.1</a>	Artemisia annua amorpha-4,11-diene synthase mR...	<a href="#">79.9</a>	1e-14
<a href="#">gb AY789791.1</a>	Actinidia deliciosa germacrene-D synthase (Adg...	<a href="#">79.4</a>	2e-14
<a href="#">gb AF061285.1</a>	<a href="#">AF061285</a> Capsicum annuum sesquiterpene cyclase ...	<a href="#">79.0</a>	3e-14
<a href="#">gb AF317695.1</a>	<a href="#">AF317695</a> Perilla frutescens var. frutescens lim...	<a href="#">78.5</a>	4e-14
<a href="#">gb AF241792.1</a>	<a href="#">AF241792</a> Perilla frutescens strain 5598 limonen...	<a href="#">78.5</a>	4e-14
<a href="#">gb AF241790.1</a>	<a href="#">AF241790</a> Perilla citriodora limonene synthase g...	<a href="#">78.5</a>	4e-14
<a href="#">gb AF233894.1</a>	<a href="#">AF233894</a> Perilla citriodora limonene synthase m...	<a href="#">78.5</a>	4e-14
<a href="#">gb AC212768.1</a>	Solanum lycopersicum DNA sequence from clone L...	<a href="#">78.0</a>	5e-14
<a href="#">emb AJ251751.1</a>	<a href="#">AAN251751</a> Artemisia annua mRNA for amorpha-4,1...	<a href="#">78.0</a>	5e-14
<a href="#">gb AF327527.1</a>	<a href="#">AF327527</a> Artemisia annua amorpha-4,11-diene syn...	<a href="#">78.0</a>	5e-14
<a href="#">gb AF327526.1</a>	<a href="#">AF327526</a> Artemisia annua amorpha-4,11-diene syn...	<a href="#">78.0</a>	5e-14
<a href="#">emb AJ271793.1</a>	<a href="#">AAN271793</a> Artemisia annua mRNA for putative se...	<a href="#">77.6</a>	7e-14
<a href="#">gb AF472361.1</a>	Artemisia annua beta-caryophyllene synthase QH...	<a href="#">77.6</a>	7e-14
<a href="#">gb EF599333.1</a>	Zea mays cultivar B73 monoterpene synthase (tp...	<a href="#">77.1</a>	9e-14
<a href="#">gb EF599332.1</a>	Zea mays cultivar CML139 monoterpene synthase ...	<a href="#">77.1</a>	9e-14
<a href="#">gb EF599331.1</a>	Zea mays cultivar CML139 monoterpene synthase ...	<a href="#">77.1</a>	9e-14

Fig. 4.13 – blastx della sequenza consensus CaMTS 3

Il contig **CaMTS 4**, infine, composto da 3 sequenze per una lunghezza di 642bp, ha mostrato similarità con geni di monoterpene sintasi (Fig. 4.14). Inoltre i punteggi E-value sono relativamente alti, dell'ordine di e-50, comparabili quindi con quelli ottenuti da CaMTS 1.

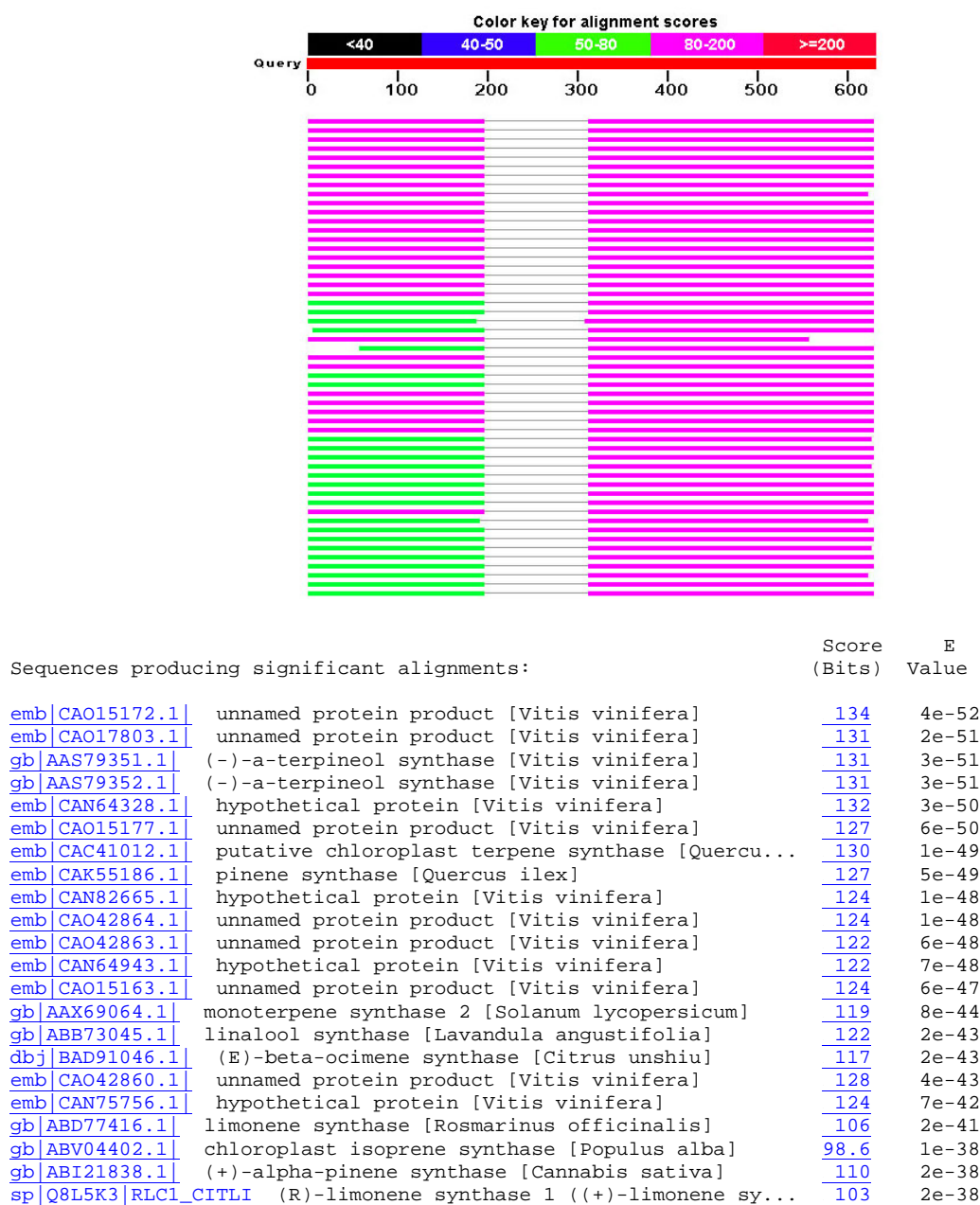
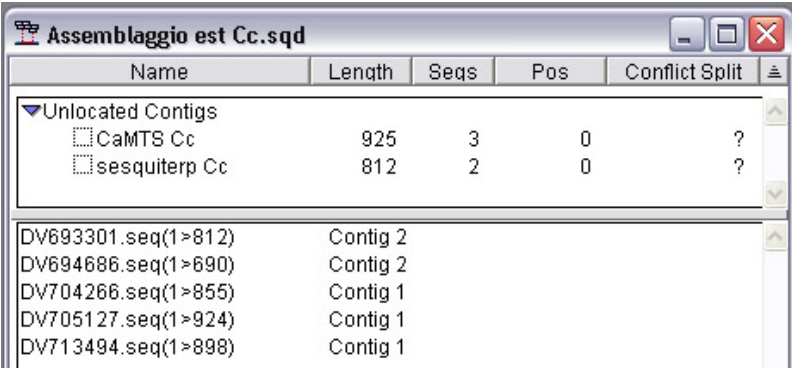


Fig. 4.14 – blastx della sequenza consensus CaMTS 4

Oltre alle sequenze finora descritte, la sequenza di un altro putativo gene CaMTS è stata ottenuta tramite un'analisi *in silico* della libreria EST di *Coffea canephora* (caffè Robusta) costruita dalla Cornell University (Lin *et al*, 2005), e condivisa con la comunità scientifica tramite i database pubblici e il sito CoffeeDNA, sviluppato dal Laboratorio presso il quale è stato svolto questo lavoro di tesi.

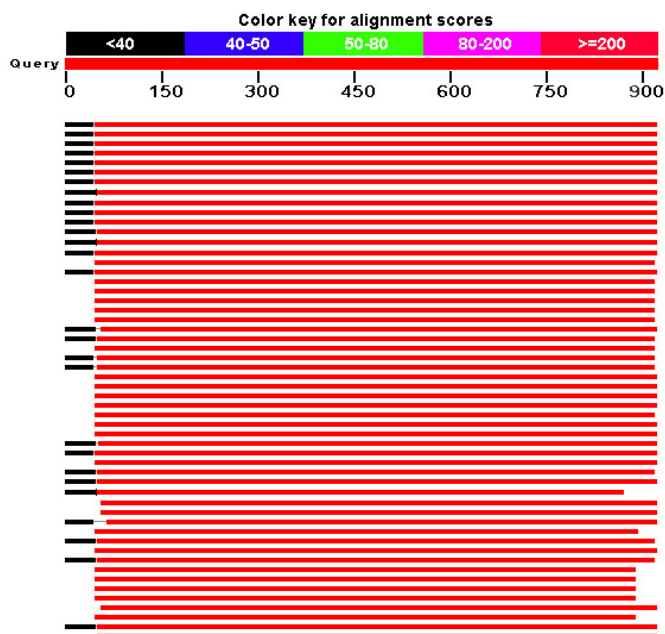
La suddetta libreria EST contiene circa 50'000 sequenze, delle quali 5 sono risultate simili a terpene sintasi. Sono state quindi assemblate mediante il programma SeqMan, che ha individuato due contig formati rispettivamente da 3 e 2 EST (Fig. 4.15).



Name	Length	Seqs	Pos	Conflict Split
Unlocated Contigs				
CaMTS Cc	925	3	0	?
sesquiterp Cc	812	2	0	?
Sequence Mappings				
DV693301.seq(1>812)	Contig 2			
DV694686.seq(1>690)	Contig 2			
DV704266.seq(1>855)	Contig 1			
DV705127.seq(1>924)	Contig 1			
DV713494.seq(1>898)	Contig 1			

Fig. 4.15 – schermata del programma SeqMan, con il riepilogo dell'assemblaggio delle sequenze delle EST di *C. canephora* simili a geni noti di terpene sintasi.

Il contig **CaMTS Cc** (ad indicare la sua origine da *C. canephora*), composto da 3 sequenze per una lunghezza di 925bp, ha mostrato un'elevata similarità con geni di monoterpene sintasi, come mostrato in Fig. 4.16.

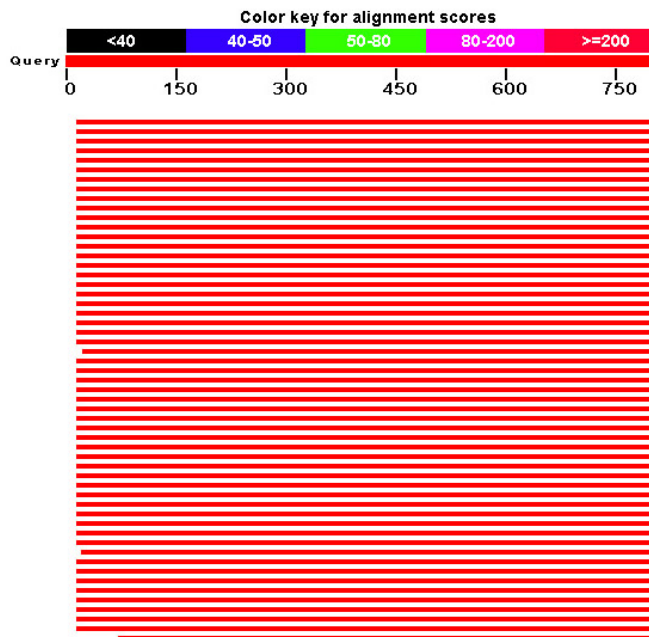




Sequences producing significant alignments:		Score (Bits)	E Value
<a href="#">emb CAO42858.1 </a>	unnamed protein product [Vitis vinifera]	<a href="#">393</a>	2e-110
<a href="#">gb AAS79351.1 </a>	(-)-a-terpineol synthase [Vitis vinifera]	<a href="#">384</a>	1e-106
<a href="#">emb CAK55186.1 </a>	pinene synthase [Quercus ilex]	<a href="#">377</a>	1e-104
<a href="#">gb ABP88782.1 </a>	plastid 1,8-cineol synthase precursor [Nicotia...	<a href="#">361</a>	1e-100
<a href="#">emb CAC41012.1 </a>	putative chloroplast terpene synthase [Quercu...	<a href="#">364</a>	1e-100
<a href="#">emb CAO15163.1 </a>	unnamed protein product [Vitis vinifera]	<a href="#">362</a>	2e-99
<a href="#">dbj BAD91045.1 </a>	monoterpene synthase [Citrus unshiu]	<a href="#">334</a>	9e-93
<a href="#">gb AAM53943.1 </a>	<a href="#">AF514286_1</a> gamma-terpinene synthase [Citrus limon]	<a href="#">330</a>	1e-91
<a href="#">dbj BAD27258.1 </a>	gamma-terpinene synthase [Citrus unshiu]	<a href="#">330</a>	2e-91
<a href="#">gb AAG31438.1 </a>	limonene synthase [Perilla frutescens]	<a href="#">330</a>	1e-89
<a href="#">gb AAF65545.1 </a>	limonene synthase [Perilla citriodora]	<a href="#">329</a>	1e-89
<a href="#">dbj BAA08367.1 </a>	limonene cyclase [Perilla frutescens var. acu...	<a href="#">329</a>	2e-89
<a href="#">gb AAK06663.1 </a>	limonene synthase [Perilla frutescens var. fru...	<a href="#">328</a>	4e-89
<a href="#">gb AAG31437.1 </a>	limonene synthase [Perilla frutescens]	<a href="#">327</a>	5e-89
<a href="#">gb AAx69064.1 </a>	monoterpene synthase 2 [Solanum lycopersicum]	<a href="#">323</a>	1e-88
<a href="#">gb AAF76186.1 </a>	mercene synthase [Perilla frutescens]	<a href="#">324</a>	6e-88
<a href="#">dbj BAF73933.1 </a>	beta-pinene synthase [Citrus jambhiri]	<a href="#">317</a>	2e-87
<a href="#">ref NP_189212.1 </a>	ATTPS-CIN (TERPENE SYNTHASE-LIKE SEQUENCE-1,...	<a href="#">312</a>	5e-87
<a href="#">gb AAx69063.1 </a>	monoterpene synthase 1 [Solanum lycopersicum]	<a href="#">314</a>	6e-86
<a href="#">dbj BAD91046.1 </a>	(E)-beta-ocimene synthase [Citrus unshiu]	<a href="#">312</a>	1e-85
<a href="#">gb AAM89254.1 </a>	<a href="#">AF527416_1</a> 3-carene synthase [Salvia stenophylla]	<a href="#">316</a>	2e-85
<a href="#">gb AAM53945.1 </a>	<a href="#">AF514288_1</a> (-)-beta-pinene synthase [Citrus limon]	<a href="#">310</a>	2e-85
<a href="#">gb AAL17636.1 </a>	d-limonene synthase [Agastache rugosa]	<a href="#">312</a>	2e-84
<a href="#">dbj BAD27260.1 </a>	beta-pinene synthase [Citrus unshiu]	<a href="#">306</a>	4e-84
<a href="#">gb ABW86881.1 </a>	limonene synthase [Mentha x piperita]	<a href="#">310</a>	7e-84
<a href="#">gb AAC37366.1 </a>	4S-limonene synthase	<a href="#">310</a>	9e-84
<a href="#">gb AAD50304.1 </a>	<a href="#">AF175323_1</a> limonene synthase [Mentha longifolia]	<a href="#">308</a>	3e-83
<a href="#">emb CAO24192.1 </a>	unnamed protein product [Vitis vinifera]	<a href="#">308</a>	4e-83
<a href="#">gb ABR15422.1 </a>	limonene synthase [Mentha haplocalyx var. pipe...	<a href="#">308</a>	4e-83
<a href="#">gb AAO85532.1 </a>	monoterpene synthase [Arabidopsis thaliana]	<a href="#">301</a>	6e-82
<a href="#">gb ABD77416.1 </a>	limonene synthase [Rosmarinus officinalis]	<a href="#">304</a>	6e-82

Fig. 4.16 – blastx della sequenza consensus CaMTS Cc

Il secondo contig ha evidenziato similarità con sequenze di sesquiterpene sintasi, ed è per questo stato rinominato **Sesquiterp\_Cc** (Fig. 4.17).



Sequences producing significant alignments:	Score (Bits)	E Value
<a href="#">dbj BAA82092.1 </a> vetispiradiene synthase [Solanum tuberosum]	<a href="#">308</a>	2e-83
<a href="#">gb AAG09950.1 AF171218_1</a> vetispiradiene synthase [Lycopersico...	<a href="#">306</a>	1e-82
<a href="#">gb AAA19216.1 </a> 5-epi-aristolochene synthase	<a href="#">303</a>	1e-81
<a href="#">gb AAP05760.1 AF484123_1</a> 5-epi-aristolochene synthase 12 [Nic...	<a href="#">302</a>	2e-81
<a href="#">sp Q40577 5EAS_TOBAC</a> Aristolochene synthase (5-epi-aristoloch...	<a href="#">301</a>	3e-81
<a href="#">dbj BAA82108.1 </a> vetispiradiene synthase [Solanum tuberosum]	<a href="#">301</a>	3e-81
<a href="#">pdb 5EAT A</a> Chain A, 5-Epi-Aristolochene Synthase From Nicotia...	<a href="#">301</a>	3e-81
<a href="#">pdb 5EAS A</a> Chain A, 5-Epi-Aristolochene Synthase From Nicotia...	<a href="#">301</a>	3e-81
<a href="#">gb AAP05762.1 AF484125_1</a> 5-epi-aristolochene synthase 37 [Nic...	<a href="#">300</a>	6e-81
<a href="#">gb AAO85555.1 </a> 5-epi-aristolochene synthase [Nicotiana attenu...	<a href="#">300</a>	6e-81
<a href="#">gb AAD02268.1 </a> putative vetispiradiene synthase 3 [Solanum tu...	<a href="#">300</a>	8e-81
<a href="#">gb AAG17667.1 AF272244_1</a> 5-epi-aristolochene synthase [Nicoti...	<a href="#">300</a>	1e-80
<a href="#">gb AAD02269.1 </a> putative vetispiradiene synthase 4 [Solanum tu...	<a href="#">300</a>	1e-80
<a href="#">gb AAA86337.1 </a> vetispiradiene synthase	<a href="#">298</a>	2e-80
<a href="#">gb AAD02223.1 </a> vetispiradiene synthase [Solanum tuberosum]	<a href="#">297</a>	6e-80
<a href="#">gb AAC61260.1 </a> sesquiterpene cyclase [Capsicum annum]	<a href="#">296</a>	1e-79
<a href="#">pdb 1HX9 A</a> Chain A, Crystal Structure Of Teas W273s Form 1 >p...	<a href="#">296</a>	1e-79
<a href="#">pdb 1HXG A</a> Chain A, Crystal Structure Of Teas W273sC440W	<a href="#">296</a>	1e-79
<a href="#">gb AA16077.1 </a> valencene synthase [Perilla frutescens var. fr...	<a href="#">295</a>	2e-79
<a href="#">emb CAA06614.1 </a> 5-epi-aristolochene synthase [Capsicum annum...	<a href="#">295</a>	3e-79
<a href="#">gb AAK15641.1 AF326117_1</a> sesquiterpene cyclase [Capsicum annum]	<a href="#">294</a>	5e-79

Fig. 4.16 – blastx della sequenza consensus Sesquiterp Cc

Alla luce di queste analisi, è stato deciso di basarsi sulle sequenze CaMTS 1 e CaMTS 4 per la prosecuzione degli esperimenti.

Sulla base della considerazione che il genoma di *C. canephora* presenta grandi omologie con quello di *C. arabica*, era probabile che anche il gene CaMTS Cc, o un suo omologo altamente simile, si potesse ritrovare in Arabica (come difatti è stato). E' stato quindi deciso di includere nella ricerca anche questo gene.

#### 4.2.5 – Disegno dei primer specifici

Sulla base delle sequenze consenso parziali ottenute, sono stati disegnati i primer specifici per l'amplificazione delle estremità del cDNA (RACE) dei geni CaMTS 1, CaMTS 4 e CaMTS Cc. In base alla lunghezza dei tratti di sequenza nota e al confronto con le monoterpene sintasi presenti in rete, i cui cDNA sono in genere lunghi circa 2 Kb, si è potuto stimare la lunghezza approssimativa delle estremità ignote ancora da sequenziare (Fig. 4.17).

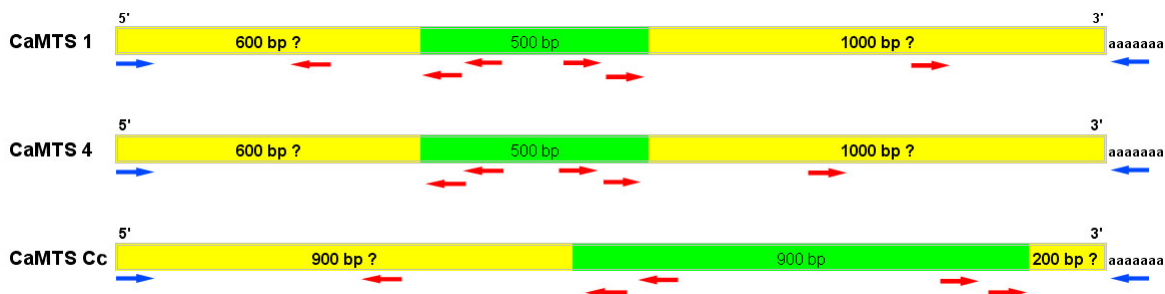


Fig. 4.17 – lunghezza stimata delle estremità da amplificare nei 3 geni studiati e strategia delle RACE; in blu i primer 12attB1-T7 e attB2 adapter, in rosso i primer specifici RACE, esterni e “nested”



#### 4.2.6 – PCR da cDNA con primer specifici

Per individuare i tessuti in cui le tre terpene sintasi putative sono espresse, per prima cosa è stata fatta una PCR di controllo su cDNA sintetizzato con tecnica SMART dai tessuti più direttamente coinvolti con la qualità del prodotto finale: il fiore, la drupa a 10, 15 e 25 settimane di maturazione, la drupa matura, e il seme essiccato. Per verificare la qualità del cDNA è stato fatto un controllo amplificando il gene per l'ubiquitina; per verificare l'eventuale contaminazione residua da DNA genomico è stato inserito anche un controllo con DNA di *C. arabica* cv. Bourbon.

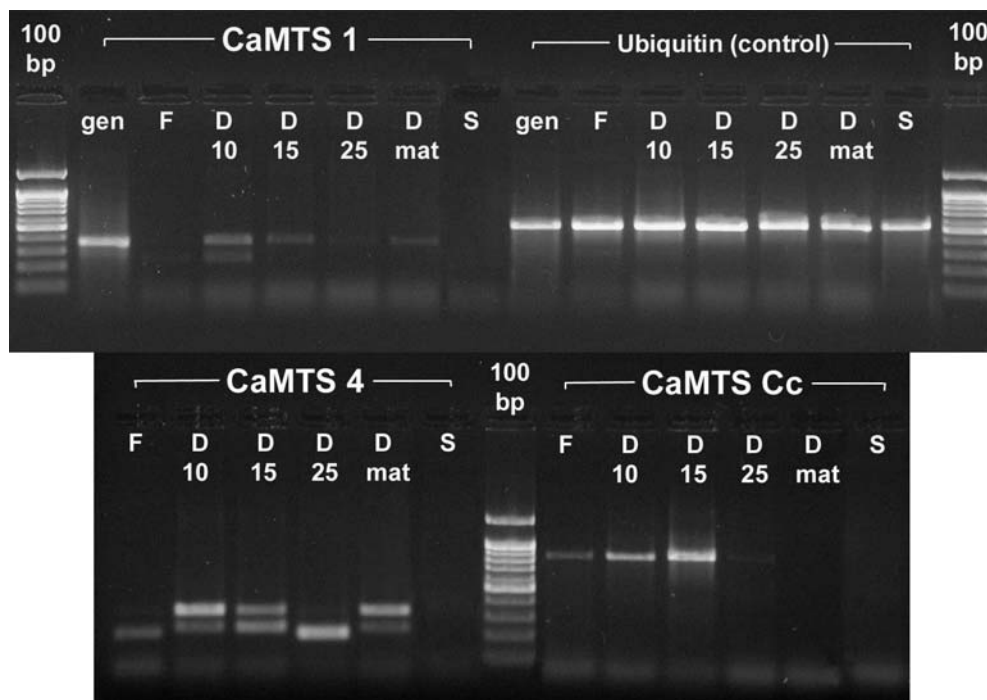


Fig. 4.17 – PCR di controllo della presenza dei trascritti CaMTS in cDNA di diversi tessuti: F=fiore; D10=drupa a 10 settimane di maturazione; D15=drupa a 15 settimane di maturazione; D25=drupa a 25 settimane di maturazione; Dmat= drupa matura; S=seme essiccato; gen= controllo da DNA genomico.

Come si vede in Fig. 4.17, la sintesi del cDNA è avvenuta con successo in tutti i tessuti (come si evince dal controllo ubiquitina, positivo per tutti i cDNA); è tuttavia presente una contaminazione da DNA genomico residuo, che il protocollo di sintesi non è riuscito ad eliminare. Ciò non costituisce un problema grave, in quanto l'amplificazione del cDNA avviene in maniera specifica grazie alle sequenze 12attB1-T7 e attB2-oligod(t), che non sono presenti nel DNA genomico.

E' stato quindi possibile procedere con le RACE, utilizzando il cDNA D10 per il gene CaMTS 1, D25 per CaMTS 4, ed F, D10 e D15 per CaMTS Cc.

#### 4.2.7 – RACE 3' e 5', nested PCR e clonaggio degli amplificati

Le reazioni RACE sono state condotte per ciascun gene come descritto nella sezione Materiali e Metodi. La prima amplificazione ha dato in genere bande multiple, non idonee al sequenziamento o al clonaggio. E' stata quindi eseguita una seconda amplificazione sequenziale con primer interni ("nested PCR"), che ha permesso di isolare bande nettamente migliori e clonare (o sequenziare direttamente) le estremità.

In Fig. 4.18 si possono osservare i risultati dell'amplificazione delle estremità del trascritto CaMTS 1. Interessante notare come tra i prodotti della RACE (Fig. 4.18 A) i più abbondanti non siano della lunghezza prevista (circa 600 basi per il 5' e 1 Kb al 3'). La nested PCR, invece (Fig. 4.18 B), favorisce in maniera molto selettiva i prodotti specifici, evidenziando le bande più intense proprio ai pesi molecolari previsti. La banda dell'estremità 5' è risultata di qualità sufficiente ad essere sequenziata direttamente; quella al 3' è stata invece prima purificata ed estratta dal gel con kit GFX, e poi clonata, con i risultati illustrati in Fig. 4.18 C.

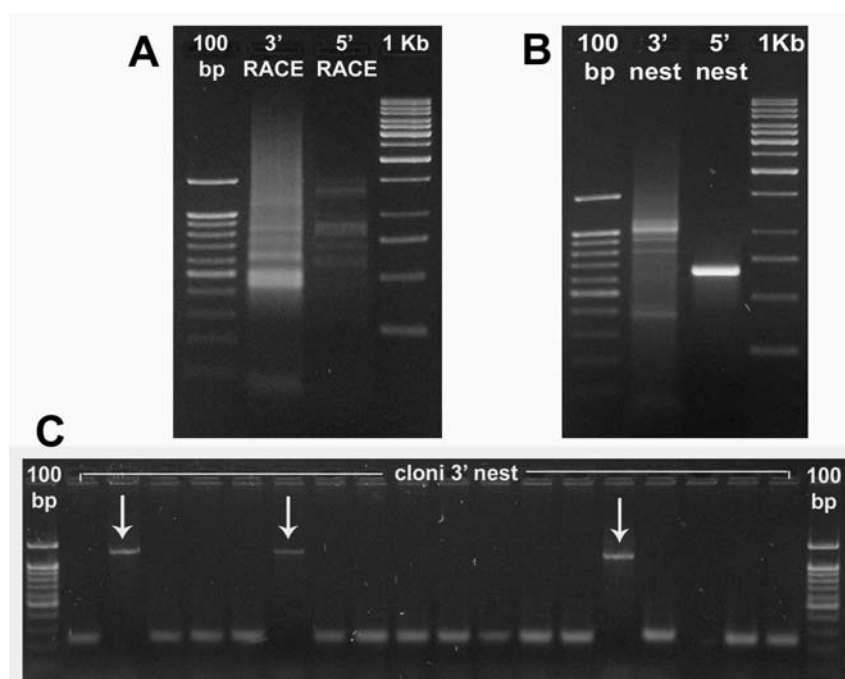


Fig. 4.18 – (A) RACE, (B) nested PCR e (C) clonaggio delle estremità 3' e 5' del gene CaMTS 1

La Fig. 4.19 mostra i risultati della stessa procedura applicata al gene CaMTS 4. Come si può vedere, nonostante le RACE corrispondessero alle lunghezze attese (Fig. 4.19 A), in questo caso la PCR nested non ha prodotto un miglioramento determinante nella specificità dei prodotti di amplificazione (Fig. 4.19 B). E' stato quindi necessario estrarre da gel entrambe le bande e clonarle, selezionando i cloni positivi come da Fig. 4.19 C.

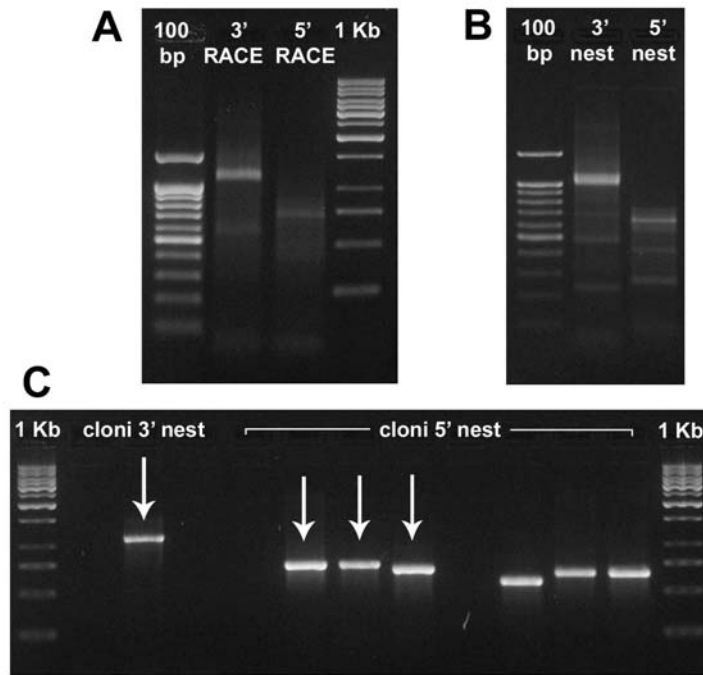


Fig. 4.19 – (A) RACE, (B) nested PCR e (C) clonaggio delle estremità 3' e 5' del gene CaMTS 4

Infine, la Fig. 4.20 illustra i prodotti di amplificazione di CaMTS Cc. Si nota come l'estremità 5' è presente nella RACE come un banda molto debole, ma di lunghezza corrispondente a quella stimata (Fig. 4.20 A). Nelle amplificazioni successive non è stato tuttavia possibile recuperare l'intera lunghezza del frammento, nonostante l'utilizzo di nested PCR, taglio ed estrazione della banda, e clonaggio. Questo ha portato a una sequenza probabilmente non del tutto completa, in quanto troncata al 5'.

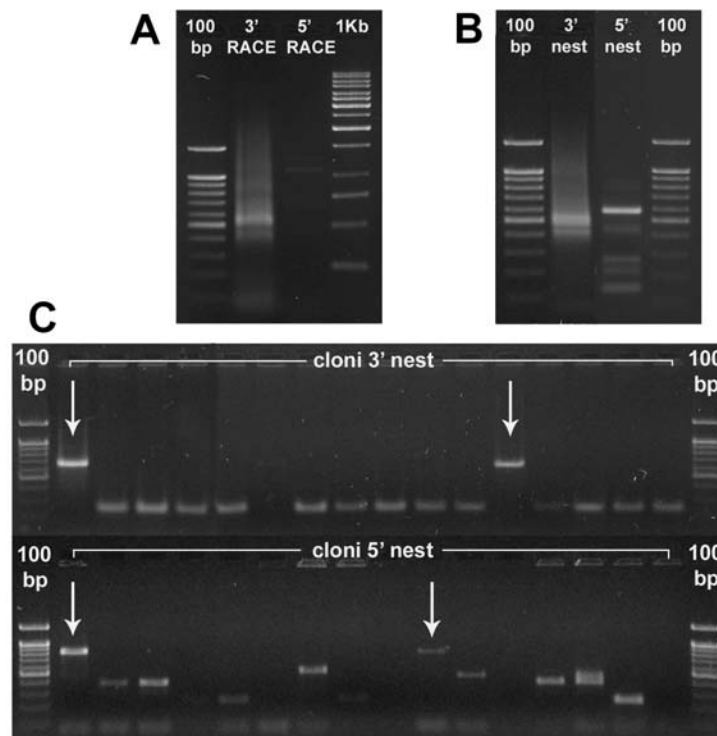


Fig. 4.20 (A) RACE, (B) nested PCR e (C) clonaggio delle estremità 3' e 5' del gene CaMTS Cc

#### 4.2.8 – Sequenziamento, assemblaggio ed analisi dei trascritti “full length”

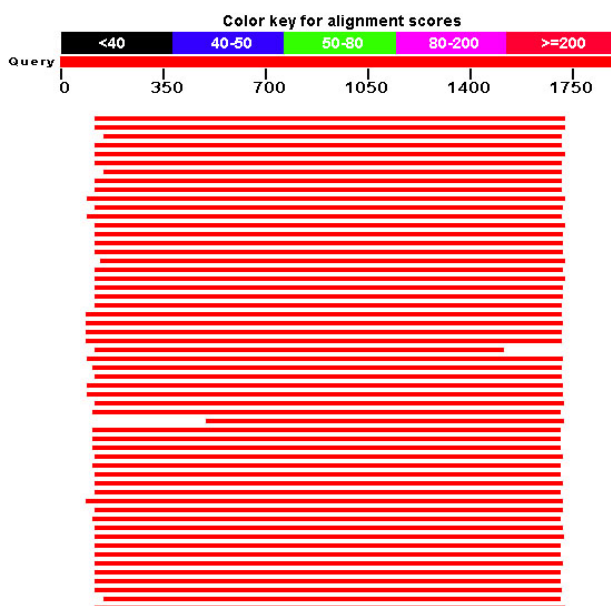
I cloni così selezionati sono stati sequenziati ed assemblati con i contig precedentemente costruiti, in modo da completare le sequenze dei trascritti CaMTS. Di seguito si riportano le sequenze nucleotidiche dei 3 geni, e le sequenze aminoacidiche ricavate traducendo l’open reading frame più esteso individuato all’interno di ciascun cDNA. Di seguito si riportano le sequenze nucleotidiche ed aminoacidiche dei tre geni, e le analisi BLAST che confermano la loro alta similarità con sequenze di monoterpene sintasi.

##### CaMTS 1 (1886 nt)

```
GAgAGGTGGAGGAacCCAAGACAAAAATCAGTGCTTCTACAGCTGAAGCCTCTAGTAGCCGAATTTTCATCTGCTAAAAATG
ACTGCTGATGGCACTATCAAGCTTGGTGATCAGTCACCTCTAAAACAATCAGAAAAGGACCACCCTGTCAGTTGGGATTT
CAAGCTTGTGAGTCACTAAGAAATGAATATGCAGATGAAAGGTACATAAGTCGGTCTGCAATGCTGGATCAAGAAATGA
ATGTGGTAAATCTGTTGGAGTTAATTGATAACTTGCAAAGGCTTGGCTTATCTTACCATTTTGAGGACAAAATCCGCTCA
ATTTTAAGTGGAAATTTACAACACGATTAATAATGCGGAACCCAGAGGGATTATATGCAACAGCTCTTGAGTTCCAGACTCCG
TAGACAACATGGTTTTATGTCCCTCAAGAGATCTTTGAGTCTTTCAAGGATGAAAACGGAGATTTCAATCATAGTCTTT
GTGAGGATTTGAAGGATTTGTGTACTTGTACGAAGCTTCATACCTCGAAAAAGAAAATGAAAGCAATTTGGAAATGGCT
AGAGAATTCACAGCAAAGCATCTTAAGAAGATCTTAAGGAGAAGAGAATTGATCAAGAACTCGAAGCATTAGTTCACA
TGCTTTGGAGGTTCCACTCCATTGGAGAATGATGCGATTGGAGGCAAGGTGGTTCATAGACATATATGAGGCAAGATCAG
ACAGAAATCCATTTTTGCTTGAGCTTGCAAAGTTAGATTTTAACTTGTCCAAGCAATACATCAGAATGATCTCGAATGT
ACACTAAGATGGTGGAGCAGTACAGGCTTAGCAGAAAACCTATCATTGCTAGGGACATAATGGTGGAAAATTTCTTTTG
GACCGTAGGAATTTCTGACCCTCAACATGGAATGCAAGAAGACTATTGACAAAGGTCGCTGCTTTGGTAACTGCCTA
TAGATGATGTATACGATCAATATGGTACTGAGGATGAACCTGGAACCTTCTACTAGTGTGTTGAAAGATGGGATGTCAAC
TCAATCGATCAACTCCAGACTACATGAAAATATGTTTTCTTGCACTCTTCAACTTTGTTAATGAAATGGCTTATGATGC
TCTAAAAGAAGAAGGAGTCAACATTTATCCCTACCTAAGAAAAGCATGGGCAGATTTGTGCAAGCATACTTGCAAGAGG
CAAAATGGTTTTTCAGTGGACATATACCAACGCTACAGCAATACCTTAATAATGCCTGGACTTCAATTTGAGTCTCTCTG
GTGGTAGTGYATGCTTATTTTTGTGTTGATTATCCGATTAACARGGACCACGTGGAATACTTAGAGAAATGTCACAAGAT
TATTCGTTGTTTCATCGATGATTATACGGCTAGCAAATGATTTAGGAACTTCTCCAGAATCGGAGGTGCTGAAAAGTCTG
ATGTACAAAAATCGATACAATGCTATGTGAAGGAGACAGGTGCATGTGAAGAAAAGGCACGTGAATACCTGAGGTTTCTG
ATTATTGAGGCAYGGAAGCAAATGAACGAAGCTCAACTGTAGATTCTCCATTTTCTTCGACATTTAAAGGATTTGCAGT
GAATGTTGCCAGGATGGACAATGCATGTAYCAGCATGGAGATGGCCATGCCATGAAAATTTCTGAGCTAGGGATCGTA
TTTTGTCACTGCTCTTTGAGCCCATATCTTCTTTGCTTAGACCTAATATTCGGTTTATAATCTAATAATCGGAATTGGG
TGGCAGTATCTGCAAMATCTTGACCATCCGATTRATCTCTCTCTCTTTAAGATCTTCYTAAGATGTTATAGAGTTAAT
TCTGCTTTGCACCCTTAAATTAActggAAAAAAAAAAAAAAAAAAAA
```

##### CaMTS 1 (548 aa)

```
MTADGTIKLGDQSPKQSEKDHVPVSWDFKLVQSLRNEYADERYISRSAMLDQEMNVNLLLELIDNLQRLGLSYHFEDKIR
SILSGIYNTIKMRNPEGLYATALEFRLRRQHGFYVPEIFESFKDENGDFNHSLCEDLKGLLYLYEASYLEKENESNLEM
AREFTAKHLKILKEKRIDQLEALVQHLELPLHWRMMRLEARWFIDIYEARS DRNPILLELAKLDFNIVQAIHQNDLE
CTLRWWSSTGLAEKLSFARDIMVENFFWTVGTISDPQHGNARLLTKVAALVTAIDDVYDQYGTEDLELFTSVVERWDV
NSIDLPLDYMKICFLALFNFNEMAYDALKEEGVNIIPYLKAWADLCKAYLQEAkWFFSGHIP TLQQYLNNAWTSISAP
LVVVXYAFCVDYPINXDHVEYLEKCHKIRCSMIIRLANDLGTSPSEVLKSDVPKSIQCYVKETGACEEKAREYLRFL
LIEAXKQMNBAQTVDSPFSSTFKGFVNVARMGQCMYQHGDGHAHQNSEPRDRILSLLFEPIS SFA
```



Sequences producing significant alignments:		Score (Bits)	E Value
<a href="#">gb AAS79351.1 </a>	(-)-a-terpineol synthase [Vitis vinifera]	<a href="#">637</a>	0.0
<a href="#">emb CAO42858.1 </a>	unnamed protein product [Vitis vinifera]	<a href="#">630</a>	1e-179
<a href="#">emb CAN62729.1 </a>	hypothetical protein [Vitis vinifera]	<a href="#">625</a>	4e-178
<a href="#">emb CAK55186.1 </a>	pinene synthase [Quercus ilex]	<a href="#">610</a>	1e-173
<a href="#">emb CAC41012.1 </a>	putative chloroplast terpene synthase [Quercu...]	<a href="#">597</a>	9e-170
<a href="#">emb CAN64328.1 </a>	hypothetical protein [Vitis vinifera]	<a href="#">592</a>	2e-168
<a href="#">gb ABP88782.1 </a>	plastid 1,8-cineol synthase precursor [Nicotia...]	<a href="#">572</a>	3e-162
<a href="#">emb CAO15161.1 </a>	unnamed protein product [Vitis vinifera]	<a href="#">569</a>	2e-161
<a href="#">emb CAN64943.1 </a>	hypothetical protein [Vitis vinifera]	<a href="#">567</a>	1e-160
<a href="#">emb CAN60797.1 </a>	hypothetical protein [Vitis vinifera]	<a href="#">562</a>	3e-159
<a href="#">emb CAO15163.1 </a>	unnamed protein product [Vitis vinifera]	<a href="#">556</a>	2e-157
<a href="#">dbj BAD27259.1 </a>	gamma-terpinene synthase [Citrus unshiu]	<a href="#">553</a>	2e-156
<a href="#">dbj BAD91045.1 </a>	monoterpene synthase [Citrus unshiu]	<a href="#">546</a>	2e-154
<a href="#">gb AAM53943.1 </a>	<a href="#">AF514286_1</a> gamma-terpinene synthase [Citrus limon]	<a href="#">545</a>	4e-154
<a href="#">dbj BAD27258.1 </a>	gamma-terpinene synthase [Citrus unshiu]	<a href="#">543</a>	2e-153
<a href="#">emb CAO15174.1 </a>	unnamed protein product [Vitis vinifera]	<a href="#">531</a>	5e-150
<a href="#">dbj BAF73933.1 </a>	beta-pinene synthase [Citrus jambhiri]	<a href="#">526</a>	2e-148
<a href="#">gb AAX69063.1 </a>	monoterpene synthase 1 [Solanum lycopersicum]	<a href="#">526</a>	3e-148
<a href="#">gb AAX69064.1 </a>	monoterpene synthase 2 [Solanum lycopersicum]	<a href="#">526</a>	3e-148
<a href="#">gb AAM53945.1 </a>	<a href="#">AF514288_1</a> (-)-beta-pinene synthase [Citrus limon]	<a href="#">525</a>	5e-148
<a href="#">dbj BAD27260.1 </a>	beta-pinene synthase [Citrus unshiu]	<a href="#">524</a>	1e-147
<a href="#">dbj BAD91046.1 </a>	(E)-beta-ocimene synthase [Citrus unshiu]	<a href="#">518</a>	7e-146
<a href="#">dbj BAA08367.1 </a>	limonene cyclase [Perilla frutescens var. acu...]	<a href="#">517</a>	1e-145

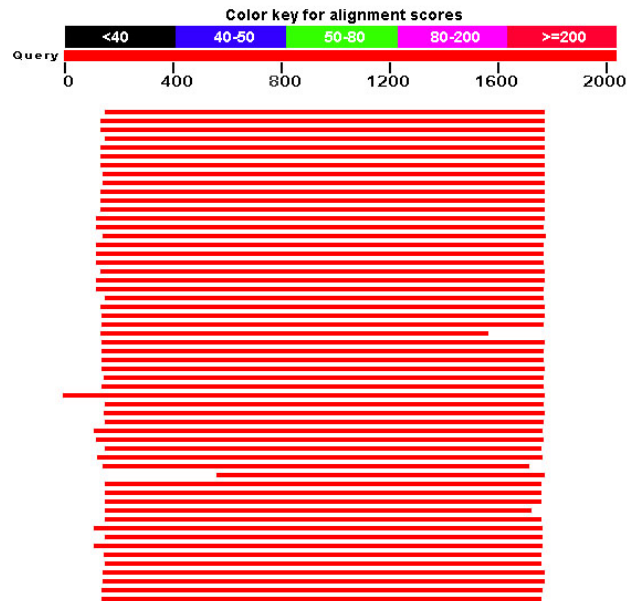
Fig. 4.21 – blastx della sequenza completa del cDNA CaMTS 1

### CaMTS 4 (2039 nt)

TCCAGCGAAGTGAATAAACATAAACCATCTCAGGTCCTGTCTCCCTCCGGGCGAGCTACGTTCACTACTCTTAGTGCTGC  
 AGCCATGAGAAGTGCAACTATGGCAGCGGCTAACGTCGAGAGCAAAGTGGTCAGAAGCAGCAGCTCATTAATAGACGCT  
 CAGGGAACTACGAAGCTCCACTCTGGGAATTCGATTACATTCAGTCATTGAAAAATGAATATGCGGGTGATATTTACGTC  
 AGTCGGGCTAATGAGTTGAAGGAGCAAGTGAAGATGATGCTCGACGAGGAAGATATGAAGCTGCTGGATTGCATGGAGCT  
 TGTGTACGGGTTGAAAAGGCTAGGACTGGCTTATCACTTTGAGGGTGAATCAACAGACTATTAAGCAGCGATTACAAG  
 CTATTCATGAAGGCAATCATCAAAGAAACAAAGAGGATTTGTATGCTGCTCGAATTTAGAATCTTCAGGCAAAAT  
 GGCTTTAACGTCCTCAAGATATATTCATGATTTTCATAACTGAGGATGGTGAATTTGATGAAAGCCTTTCTGAGGATAC  
 AATGGGACTGCCTAAGTTGTATGAAGCATCTTTCCTGTCGTTGGAAGGTGAAGCCACCTGGATTTGGCAAGGGAATTC  
 CAACTAAGCACTCAATAATATCTAGGCAAGGAAAATACTGATCAAAATCTCAGGATTTTAGTGTACCATGCCTAGAG  
 CTTCCTCCCTGAGGTGGAGCGCCGAGGATAGAAGCTAGGTGATCATCGATGCATACGAGAGAAGTCCCAACCTGAATCC  
 TACTCTACTTGAGCTTGCAAAAATAGACTTCAACATTTGTTCAAGCAATACATCAGCAGGACCTAAAACATGTGTCTGGT  
 GGTGGAAGAACATACGAATCGCGGAAAAGTTGACATTTATCAGGACAGGATAGTGGAGAATTTCTTTTGGGCAATAGGA  
 GCTGCTCTCGAGCCTCAGTACGGAAGTTGTCGAAGAATGCTCACAAGGCTTTGCTTTGATTACAATGATAGATGACAT  
 ATACGATGTTTATGGAATTTTGAAGAATTTGGAATTTTACTGATGCAGTTGAGGATGTCGAGGATGTCGAGGATGATC  
 AACTTCCAGACTACATGAGAGTTGGATATCTTGGATTTTCAATCCATCAACGAGATGGCCTATGACGCTCTCAAAGAG  
 CAAGGCGTACATATAGTGAATACCTAAAAAAGTGTGGGCAGATCTGTGTAAGCATACTTACAAGAGGCAAAATGGTA  
 CTACGCTGGATACACCAACAGTGGAGGAATACCTGAAAAATGCATGGGTTTCAATGTCCGGTTCGGTAATGTTAATGC  
 ATGCTTATGCAGGGGTTACCAATCCCATGAATAAGGAAGCCATGGATGTCCTAGACACCCACGATATCGTTTCGCTGCTCT  
 TCATATCTTCTACGATTTGCAGATGATTTAGGAACATCACAGGGGAGATGAAAAGAGGTGATGTCCCGAAATTTGGTGCA  
 ATGTTACATGAAGGAAGCAGGTTGTTTCAAGAAGAGTCCGAGGGAACATGATGGTTTTTGGTGGAGGACCTGGAAGA  
 AGATGAACAAGGACAGTGAATGGGCGGAATCGCCTTTTTCCAAGACTTTTGTACAGCTGCAAGAAGCTTTGGAAGAGTG  
 GCCCTGGTATGTACCAATACGGAGATGGGCATGGCCTTCAATCCAATCCTGAGGCTAAGGATCGCATCTTGGCCTCACT  
 CTTCTCCCGAGTCCCGCCTGCTAGAGAATGGGCTGAATCACGTACGTGCAGAAACAAGCAGGCTCTGTTGATGATCATG  
 AATTTGGAAGTAAATGTTAAACCTCTCTTCTCTCCAGATATCTAATCTGTATTGAAGAAGAAATATCAGTTCTATA  
 TTTGGACAGAGTTTGTTCGCATAATAATCGTAGTGTTTTTCTGTTGTACATGAAATCTGTGTTATTGAGTTATGGTAAAG  
 TATATGTGATGAATGACCACAAAAA

### CaMTS 4 (567 aa)

MRSATMAANVREQSGQKQQLINRRSGNYEAPLWFEFYIQLKNEYAGDIYVSRANELKEQVKMMLDEEDMKLLDCMELV  
 DGLERLGLAYHFEGRIINRLLSSDYKAIHEGNHQNKEDLYAAALEFRIFRQNGFNVPQDIFNDFITEDGEFDESLSSEDTM  
 GLLSLYEASFLSLEGEATLDLAREFTTKHLNNYLKENTDQNLRIILVYHALEPLRWRAPRIEARWYIDAYERSPNVNPT  
 LLELAKIDFNIVQAIHQDLKHVSWWKNIRIAEKLTFIRDRIVENFFWAIIGAVFEPQYGSRRMLTKVFALITMIDDY  
 DVYGTLEELELFTDAVDRWDVKAIDQLPDYMRVGYLGFNSINEMAYDALKEQGVHIVEYLKVKVWADLCKAYLQEAQWYY  
 AGYTPVVEEYLENAWVMSVPMVMHAYAGVTNPMNKEAMDVLDTHTIVRCSYLLRFADDLGTSPEGMKRGDVPKLVQC  
 YMEKAGCSEESREHVWFLLERETWKKMNKDESWAESPFKFTVTAKNFGRVALVMYQYGDGHGLHSNPEAKDRILASLF  
 SPVPPA



Sequences producing significant alignments:

	Score (Bits)	E Value
<a href="#">gb AAS79352.1 </a> (-)-a-terpineol synthase [Vitis vinifera]	<a href="#">647</a>	0.0
<a href="#">gb AAS79351.1 </a> (-)-a-terpineol synthase [Vitis vinifera]	<a href="#">645</a>	0.0
<a href="#">emb CAO64901.1 </a> unnamed protein product [Vitis vinifera]	<a href="#">644</a>	0.0
<a href="#">emb CAK55186.1 </a> pinene synthase [Quercus ilex]	<a href="#">640</a>	0.0
<a href="#">emb CAC41012.1 </a> putative chloroplast terpene synthase [Quercu...]	<a href="#">633</a>	1e-180
<a href="#">emb CAN64328.1 </a> hypothetical protein [Vitis vinifera]	<a href="#">622</a>	2e-177
<a href="#">emb CAO15161.1 </a> unnamed protein product [Vitis vinifera]	<a href="#">608</a>	6e-173
<a href="#">emb CAN74705.1 </a> hypothetical protein [Vitis vinifera]	<a href="#">605</a>	3e-172
<a href="#">emb CAO42869.1 </a> unnamed protein product [Vitis vinifera]	<a href="#">585</a>	5e-166
<a href="#">gb ABP88782.1 </a> plastid 1,8-cineol synthase precursor [Nicotia...]	<a href="#">581</a>	8e-165
<a href="#">emb CAO42867.1 </a> unnamed protein product [Vitis vinifera]	<a href="#">580</a>	1e-164
<a href="#">emb CAN60797.1 </a> hypothetical protein [Vitis vinifera]	<a href="#">574</a>	9e-163
<a href="#">emb CAN82665.1 </a> hypothetical protein [Vitis vinifera]	<a href="#">573</a>	1e-162
<a href="#">gb AAX69064.1 </a> monoterpene synthase 2 [Solanum lycopersicum]	<a href="#">573</a>	2e-162
<a href="#">emb CAN75756.1 </a> hypothetical protein [Vitis vinifera]	<a href="#">570</a>	1e-161
<a href="#">dbj BAD27259.1 </a> gamma-terpinene synthase [Citrus unshiu]	<a href="#">570</a>	1e-161
<a href="#">dbj BAD91045.1 </a> monoterpene synthase [Citrus unshiu]	<a href="#">567</a>	1e-160
<a href="#">gb AAM53943.1 </a> <a href="#">AF514286_1</a> gamma-terpinene synthase [Citrus limon]	<a href="#">565</a>	3e-160
<a href="#">gb AAM53945.1 </a> <a href="#">AF514288_1</a> (-)-beta-pinene synthase [Citrus limon]	<a href="#">563</a>	2e-159
<a href="#">dbj BAF73933.1 </a> beta-pinene synthase [Citrus jambhiri]	<a href="#">561</a>	5e-159
<a href="#">dbj BAD27258.1 </a> gamma-terpinene synthase [Citrus unshiu]	<a href="#">561</a>	5e-159
<a href="#">dbj BAD91046.1 </a> (E)-beta-ocimene synthase [Citrus unshiu]	<a href="#">558</a>	7e-158
<a href="#">dbj BAD27260.1 </a> beta-pinene synthase [Citrus unshiu]	<a href="#">557</a>	9e-158
<a href="#">gb ABI21838.1 </a> (+)-alpha-pinene synthase [Cannabis sativa]	<a href="#">554</a>	1e-156
<a href="#">gb ABD77416.1 </a> limonene synthase [Rosmarinus officinalis]	<a href="#">547</a>	1e-154
<a href="#">gb ABI21837.1 </a> (-)-limonene synthase [Cannabis sativa]	<a href="#">546</a>	2e-154
<a href="#">gb AAX69063.1 </a> monoterpene synthase 1 [Solanum lycopersicum]	<a href="#">541</a>	7e-153

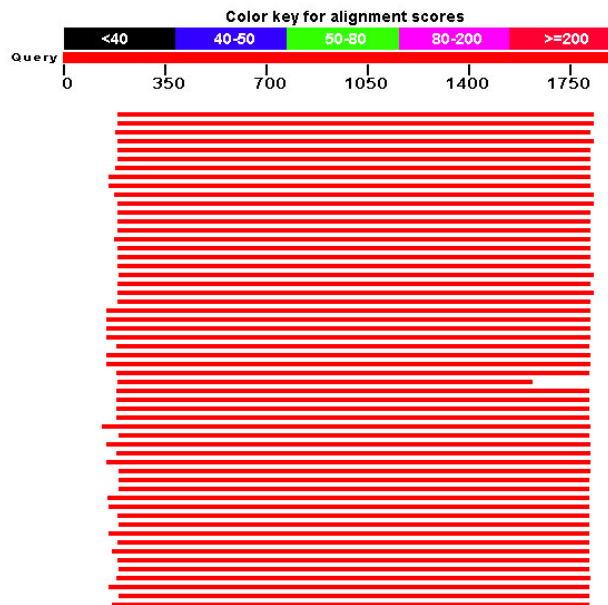
Fig. 4.22 – blastx della sequenza completa del cDNA CaMTS 4

**CaMTS Cc (1911 nt)**

AAACAAGACCAAGAATTTCCGAAAACCTTAGCAAATATAACGATGATCTCAACACTAAATCCCTTATTCACTACCCACAGA  
 AGTGGTGTATAGCCCAACAATTTTTGCATCTTCAGCTGCAGCCTCAATTAACAGTGTCTCTTCTCTTAAATCGCAGCA  
 TGCAGCAAAACAAAGCTTGTGATCAGTCACCATTAAAGACAATCAGGAAAcCACCAGTCTCTGAGTTCGGATTTCAATCA  
 TTTGCAGTCACTAAAGAATGACTATGCTGAGGAAAAGTACAAAAGTCGGTGTGAGGTGCTTAAGGAGCAAGTGAAGATGA  
 TGCTAGATCAAGAAATGGATGTGGTAAATCAGTTAGTGTAAATTGATGACTTGCAACGGCTTGGATTATCTTACCATTTT  
 GGAGACGAAATCACCTCAGTTTTAAGTGGAAATTTACAATCGGAAGTCGATGAATAAAATGCGGAACCAATGGGGATTATA  
 TGCAACATGTCTTGAGTTCAGACTCCTTAGGCAGCATGGTTTTGACGTTTCTCAAGAGATTTTGGATTGTTTCAAGGATG  
 AAAAAAGGGGATTTAGGCTAGTCTTTGCGAGGATTGGAAGGGATTGCTGTMCTTGTATGAAGCYTCAATWCCWCGAATCA  
 GAAAATGAAGAGAGCAATCTGGAAATGGCAAGACGGTTTGCAGCAAAACTCTTAAGAAGAATTTAGATGAAAAGAGAGT  
 TGATCAAGACCTTGTAGCATTAGTTCACATGCCCTGGAGCTTCCACTCCATTGGAGGATGATGCGTTTGGAGGCAAGGT  
 GGTTCATAGACATATACGAGGAAAGATCAAACAGGAATCCTATTCTGCTTGAGCTTGCAAACTAGATTTTAAATATTGTC  
 CAAGCAGCACATCAGAAATGATCTCACATATACATTAAGATGGTGGAGAAGTACATGTCTAGCAGAAAATTTGACATTTGC  
 CAGGGACATGATGGTAGAAAATTTCTTCTGGACTGTCCGAATAATTTAGACCCTCAACGTGGAAATGGTAGAAGAATAT  
 TGACAAAAGTCGTTGCTTTGATAACTGCTATAGATGATATATATGATTTGTTATGGTACTTTGGATGAACTGGAGTCTTC  
 ACTACTGCTGTTGAAAGATGGGATGTCAACTCAATTGATCAACTCCAGATTGCATGAAAATATGTTTTCTTGCACTCTA  
 CAACTTTGTCAATGAAATGGCCTATGATGCTCTAAAAGAACAGGAGTCAACATCATCCCATACCTAAGAAAATCATTGGG  
 CCGATTTATGCAAAAGCATACTTGAAGAGGCAAAATGGTTTTTTCAGTGGAGAAGTACCAACCCTACAGCAATACCTCAAT  
 AATGCCTGGATTTCAATATCAGCTCCAGCATTCTAGTCCAATGCTTATTTTTGTGTTGATTATCCCATAAACAGGATCA  
 CCTGCAATACCTGGATAACTATCACAAGATTATTCGTTGTTTCAGCAATGATTTTGGCGCTCACAAATGATTTAGGAACAT  
 CTCCAGAATCGAGGTTGTAATGTCGGTGTGATGATACCAAAATCGATACGATGCTATATGAAGGAGACAGGTGCATGTGAA  
 GAAAAGGCACGTGAACACCTAAGGTTTTCTGATTACTGAGGCATGGAAGCAATGGAAGAAGCTCAAACCTTTAGACTCTCC  
 CTTTTCTTACATTTAATGGAATTGCAGTTAATCTTGCAAGGATGGGACTATGCATGTACCAGCATGGAGATGGCCATG  
 GCCATCAAATTTCTGAGCCTAGGGATCGTATTTTGTCACTGCTCTTTGAGCCATATGTTGTCTTGCYTAGRCCTAATTT  
 TGGTTTCGAGTTAAATAACAATTTTCAGGAAAtgaaatagcttttaatacaaaaAAAAAAAAAAAAAAAAAAAA

**CaMTS Cc (65+505= 570 aa)**

TRPRI SENLANI TMI STL NPL F TTHRSGVIAQQFFASSAAAS INSVSLK SQHAAKQ SLL I SHH  
 MMLDQEMDVVNQLVLI DD LQRLGLSYHFGDEITSVLSGIYNRKSMNKMRNQWGLYATCLEFRLLRQHGFVDSQEIFDCCFK  
 DEKGFDRP SLCEDSKGLLXLYEASXXE SENEESNLEMARRFAAKTLKKNLDEKRVDQDLVALVQHALELPLHWRMMRLEA  
 RWFID IYEERSNRNPILLELAKLDFNIVQAAHQNDLTYTLRWRSTCLAEKLT FARDMMVENFFWTVGIISDPQRGNRR  
 ILLTKVVALITAI DDIYDCYGLDELEVFTTAV ERWDVNSIDQLPDCMKICFLALYNFVNEMAYDALKEQGVNIIPYL RKS  
 WADLCKAYLQEA KWFFS GEVPTLQYLNNAWISISAPAFVHVAYFCVDYPINKDHLQYLDNYHKIIRCSAMILRLTNDLG  
 TSP ESEVLNVGDV PKSIRCYMKETGACEEKAREHLRFLITEAWKQMEEAQTLDSPFSSTFNGLI AVNLARMGLCMYQHGDG  
 HGHQNSEPRDRILSLLFEPICCLA



Sequences producing significant alignments:		Score (Bits)	E Value
<a href="#">gb AAS79351.1</a>	(-)-a-terpineol synthase [Vitis vinifera]	<u>628</u>	5e-179
<a href="#">gb AAS79352.1</a>	(-)-a-terpineol synthase [Vitis vinifera]	<u>625</u>	3e-178
<a href="#">emb CAO42858.1</a>	unnamed protein product [Vitis vinifera]	<u>623</u>	2e-177
<a href="#">emb CAN62729.1</a>	hypothetical protein [Vitis vinifera]	<u>620</u>	1e-176
<a href="#">emb CAO42861.1</a>	unnamed protein product [Vitis vinifera]	<u>616</u>	2e-175
<a href="#">emb CAK55186.1</a>	pinene synthase [Quercus ilex]	<u>610</u>	1e-173
<a href="#">emb CAC41012.1</a>	putative chloroplast terpene synthase [Quercu...	<u>595</u>	3e-169
<a href="#">emb CAN64328.1</a>	hypothetical protein [Vitis vinifera]	<u>592</u>	3e-168
<a href="#">emb CAO15161.1</a>	unnamed protein product [Vitis vinifera]	<u>585</u>	5e-166
<a href="#">emb CAO42863.1</a>	unnamed protein product [Vitis vinifera]	<u>572</u>	4e-162
<a href="#">gb ABP88782.1</a>	plastid 1,8-cineol synthase precursor [Nicotia...	<u>570</u>	2e-161
<a href="#">emb CAN64943.1</a>	hypothetical protein [Vitis vinifera]	<u>570</u>	2e-161
<a href="#">emb CAO15163.1</a>	unnamed protein product [Vitis vinifera]	<u>555</u>	4e-157
<a href="#">dbj BAD91045.1</a>	monoterpene synthase [Citrus unshiu]	<u>547</u>	1e-154
<a href="#">dbj BAD27259.1</a>	gamma-terpinene synthase [Citrus unshiu]	<u>547</u>	1e-154
<a href="#">gb AAM53943.1</a>	<a href="#">AF514286_1</a> gamma-terpinene synthase [Citrus limon]	<u>544</u>	1e-153
<a href="#">dbj BAD27258.1</a>	gamma-terpinene synthase [Citrus unshiu]	<u>540</u>	1e-152
<a href="#">dbj BAA08367.1</a>	limonene cyclase [Perilla frutescens var. acu...	<u>531</u>	8e-150
<a href="#">dbj BAF73933.1</a>	beta-pinene synthase [Citrus jambhiri]	<u>529</u>	2e-149
<a href="#">gb AAM53945.1</a>	<a href="#">AF514288_1</a> (-)-beta-pinene synthase [Citrus limon]	<u>529</u>	2e-149
<a href="#">gb AAG31438.1</a>	limonene synthase [Perilla frutescens]	<u>529</u>	2e-149
<a href="#">emb CAO15174.1</a>	unnamed protein product [Vitis vinifera]	<u>528</u>	4e-149
<a href="#">gb AAG31435.1</a>	limonene synthase [Perilla citriodora]	<u>528</u>	5e-149

Fig. 4.23 – blastx della sequenza completa del cDNA CaMTS Cc

La sequenza del trascritto CaMTS Cc si è rivelata di più difficile analisi delle altre due, probabilmente a causa dei problemi incontrati nell'isolare la sua estremità 5' (cfr. § 4.2.7). La principale open reading frame del trascritto risulta sensibilmente più corta di quelle degli altri geni CaMTS (505 aminoacidi, rispetto ai 548 e 567 di CaMTS 1 e 4, rispettivamente), ma a monte di questa è presente un altro tratto di cornice di lettura aperta di 65 aminoacidi, mancante però di un codone ATG di inizio. Ciò lascia supporre che la sequenza non sia completa, oppure che esistano delle varianti di splicing dell'mRNA che hanno interferito con le reazioni di sequenziamento, confondendone i risultati.

I tre trascritti sono stati allineati mediante il programma Multalin ed il risultato è mostrato nella Fig. 4.24. E' immediatamente evidente come i geni siano molto simili, infatti ci sono ampie zone di identità separate da pochi aminoacidi differenti; in particolare il dominio DDXXD (posizione 319-323) e tutta la regione circostante sono altamente conservati. La regione più diversa è sicuramente la N-terminale, dove si può osservare che prima dell'aminoacido 66 non viene sostanzialmente trovata similarità. Da notare anche che delle tre sequenze, solo CaMTS 4 presenta il dominio RR<sub>x</sub>W (posizione 25-35). Questa caratteristica sarà discussa in relazione all'analisi filogenetica dei geni CaMTS, nel § 4.4.





### 4.3 - Analisi di espressione dei geni CaMTS nei diversi tessuti

Le PCR effettuate da cDNA di diversi tessuti (Fig. 4.25) permettono di fare alcune interessanti considerazioni sull'espressione differenziale dei trascritti CaMTS in fiori, frutti e semi di *Coffea arabica*.

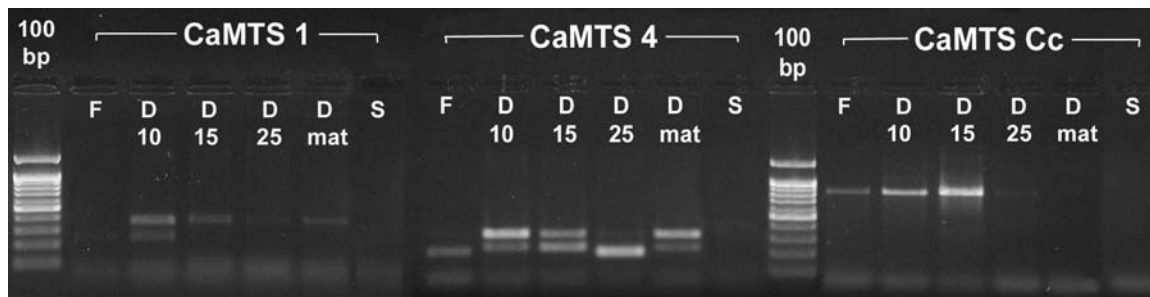


Fig. 4.25 – espressione differenziale dei geni CaMTS in diversi tessuti di *C. arabica*:  
F=fiore; D10=drupa a 10 settimane di maturazione; D15=drupa a 15 settimane di maturazione;  
D25=drupa a 25 settimane di maturazione; Dmat= drupa matura; S=seme essiccato

Innanzitutto si può notare che il trascritto CaMTS Cc, isolato da sequenze EST di *Coffea canephora*, è espresso anche in *C. arabica*, confermando l'elevata vicinanza evolutiva dei genomi di queste due piante. L'espressione di questo gene inizia già nel fiore, per aumentare fino alla 15ma settimana di maturazione del frutto. Dopo questa fase si trova ancora della debole attività residua a 25 settimane, che sparisce del tutto nel frutto maturo e nel seme.

Il trascritto CaMTS 1 presenta tracce di espressione nei fiori ed è indubabilmente espresso nella drupa alla decima settimana. Dopo questo momento l'espressione sembra disattivarsi (le ulteriori bande visibili nel gel sono ascrivibili a una contaminazione di DNA genomico – cfr. Fig. 4.17).

Il trascritto CaMTS 4, infine, è quello che presenta il quadro più complesso. Non solo sembra essere espresso in tutti i tessuti tranne il seme, ma la PCR ha amplificato una singola banda nei fiori e nella drupa alla 25ma settimana, e due bande (entrambe ad altezze diverse dalla prima) in drupe a 10, 15 settimane e a maturazione completa. La banda più alta potrebbe essere imputabile a DNA genomico, ma anche in questo caso la banda “corretta” presenta altezze diverse nei diversi tessuti: è quindi possibile che si tratti di un fenomeno di splicing alternativo, in cui uno stesso cDNA viene maturato in modi diversi a seconda del tessuto in cui viene espresso; un'altra possibilità è che i primer abbiano amplificato un altro gene molto simile, evento non improbabile in un genoma complesso come quello allotetraploide di *C. arabica*.

In generale, si può concludere che tutti i geni CaMTS sono trascritti già nel fiore e che la loro espressione continua, con varie modalità e durata, nel corso della maturazione del frutto. In ogni caso, comunque, l'espressione termina al più tardi con la completa maturazione della drupa (in alcuni casi ben prima), e tutti i geni sono inattivi nel seme maturo essiccato. Questo quadro è coerente con una visione generale dello sviluppo del frutto in cui i composti aromatici, di difesa e di riserva vengono sintetizzati ed accumulati nel frutto e nel seme durante la maturazione; questa sintesi si arresta quando frutto e seme hanno completato il proprio sviluppo, ed il seme entra in una fase di quiescenza (indotta anche dalla lavorazione post-raccolta e dalla disidratazione), in cui questi geni non sono attivati (De Castro e Marraccini, 2006).

#### 4.4 - Analisi filogenetica dei geni CaMTS

La ricerca in banca-dati ha prodotto una collezione di 124 sequenze di terpene sintasi, il più possibile rappresentativa e non ridondante. Le sequenze sono state identificate con un numero progressivo, il codice di accession GenBank, una sintetica descrizione del gene e l'organismo di appartenenza.

Queste informazioni sono sintetizzate nella tabella riassuntiva 4.26, che riporta anche quanti dei geni sono stati isolati da angiosperme o da gimnosperme. La tabella 4.27 riporta invece l'elenco completo dei geni utilizzati.

Famiglia genica	#	Organismo di provenienza	#
Monoterpene sintasi	62	Angiosperme	94
Diterpene sintasi	20	Gimnosperme	30
Sesquiterpene sintasi	42		
TOTALE	124	TOTALE	124

Tab 4.26 – Riepilogo delle caratteristiche dei geni utilizzati per la costruzione dell'albero filogenetico delle monoterpene sintasi vegetali

In base alla classificazione di Bohlmann (1998), le terpene sintasi si classificano in 7 sottofamiglie identificate con la sigla TPS (Terpene Synthase) e una lettera progressiva:

- TPS-a: sesquiterpene sintasi di angiosperme
- TPS-b: monoterpene sintasi di angiosperme
- TPS-c: alcune diterpene sintasi del metabolismo primario delle angiosperme, in particolare copalil-difosfato sintasi
- TPS-d: tutte le terpene sintasi (mono-, di- e sesqui-) di gimnosperme
- TPS-e: altre diterpene sintasi del metabolismo primario delle angiosperme, in particolare ent-kaurene sintasi
- TPS-f: un gruppo ristretto e filogeneticamente molto distante dagli altri, che include soprattutto (putative) sequenze di linalolo sintasi del genere *Clarkia*
- TPS-g: un altro gruppo isolato, che comprende sintasi di monoterpene aciclici prive del dominio RR<sub>x</sub>W, soprattutto del genere *Antirrhinum*

La Fig. 4.28 riporta l'albero filogenetico ottenuto con tutte le 124 sequenze, a cui sono state aggiunte le tre putative monoterpene sintasi di *C. arabica*.

Si nota innanzitutto che la classificazione di Bohlmann rimane valida, e i 7 gruppi sono ben distinguibili tramite le distanze filogenetiche.

Le sequenze dei geni CaMTS sono andate a posizionarsi nella sottofamiglia TPS-b delle monoterpene sintasi di angiosperme, in una ramificazione del sottoalbero che include i geni dell' $\alpha$ -terpineolo sintasi di *Vitis vinifera* e della pinene sintasi di *Quercus ilex*, che difatti anche nelle analisi blast risultavano tra le proteine a più elevata similarità.

Questo posizionamento era certamente prevedibile per CaMTS 4, mentre era incerto per CaMTS 1 e CaMTS Cc, che non presentano in prossimità dell'N-terminale il dominio RR<sub>x</sub>W che normalmente caratterizza le monoterpene sintasi del gruppo TPS-b. Questo fatto poteva far ipotizzare una loro appartenenza ad altre sottofamiglie geniche, ad esempio TPS-g, ma l'elevata somiglianza con la famiglia TPS-b fa propendere per un loro inserimento in tale gruppo. In tal caso, l'assenza del dominio conservato conferma l'ipotesi che la sequenza di CaMTS Cc sia incompleta al 5', mentre per CaMTS 1 solleva un nuovo interrogativo, a cui solo la caratterizzazione funzionale potrà dare risposta definitiva.





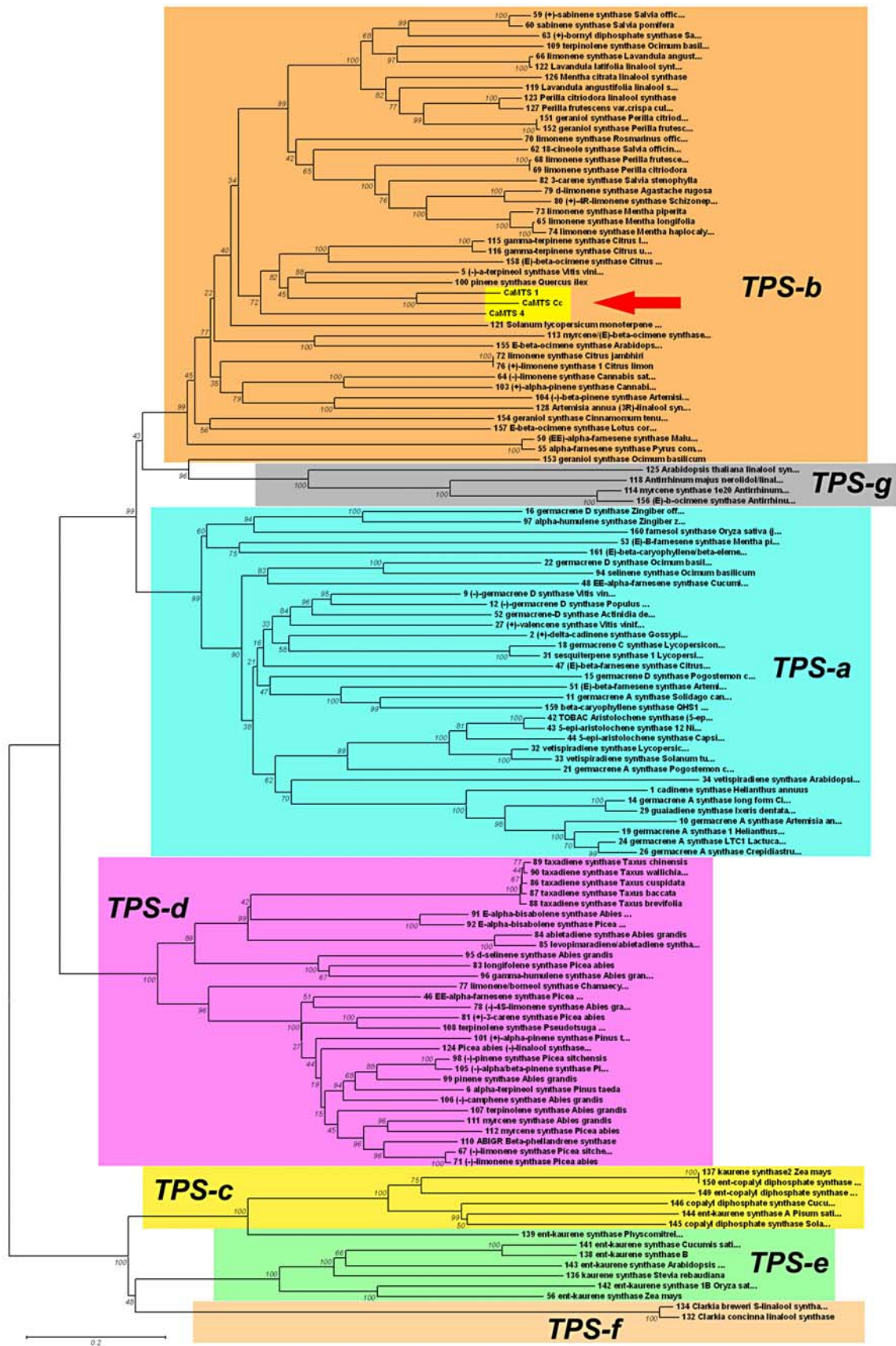


Fig.4.28 – Albero filogenetico di 124 terpene sintasi vegetali, cui sono state aggiunte le 3 monoterpene sintasi putative di *Coffea arabica* CaMTS1, CaMTS 4 e CaMTS Cc (evidenziate dalla freccia).

In Fig. 4.29 è rappresentato un albero radiale in cui sono state inserite soltanto le 62 sequenze di monoterpene sintasi, insieme alle 3 sequenze CaMTS.

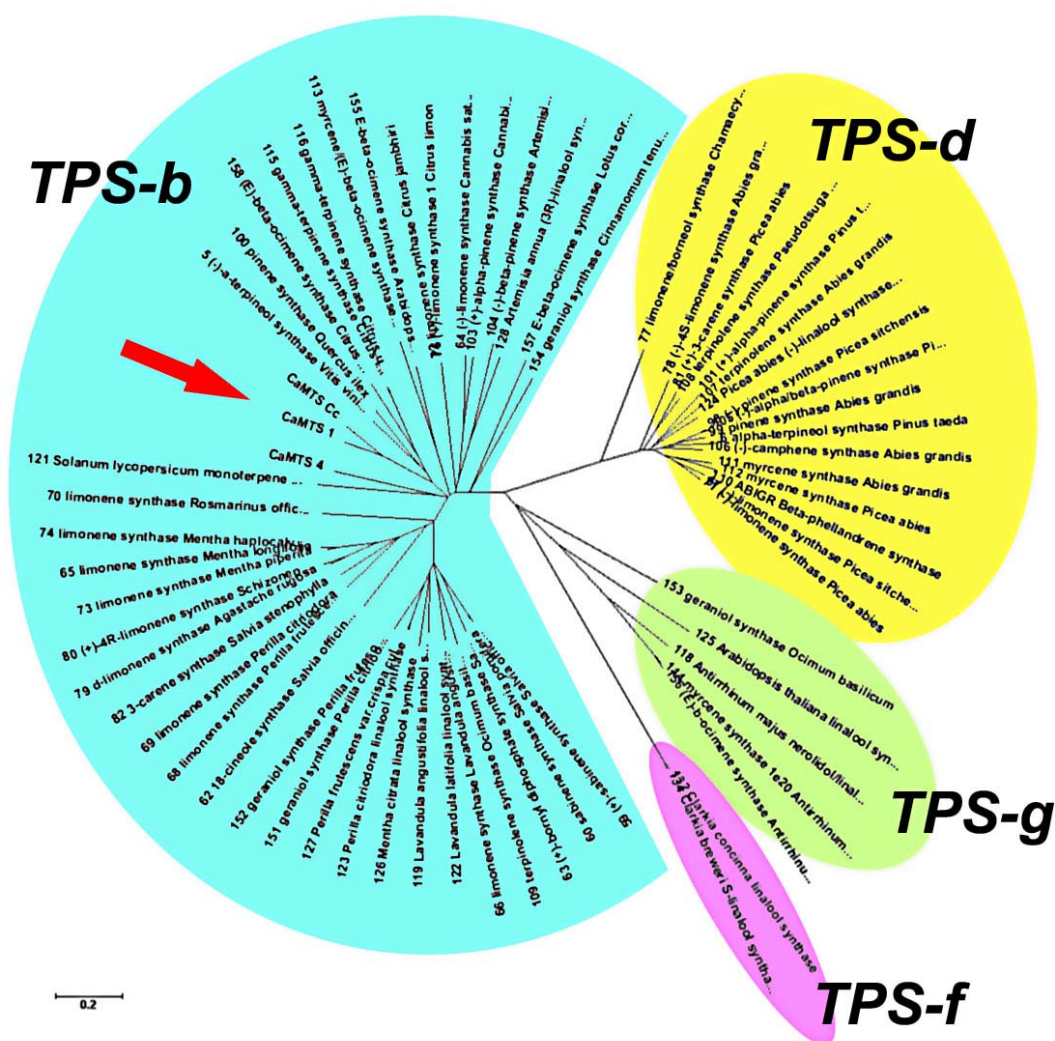


Fig.4.29 – Albero filogenetico radiale di 62 monoterpene sintasi vegetali, cui sono state aggiunte le 3 putative monoterpene sintasi di *Coffea arabica* CaMTS1, CaMTS 4 e CaMTS Cc (evidenziate dalla freccia).

In questo albero si può apprezzare meglio la distribuzione delle varie sottofamiglie. Innanzitutto si osserva come le monoterpene sintasi delle gimnosperme (TPS-d) siano notevolmente distanti da quelle delle angiosperme, anche quando sintetizzano la stessa molecola. Inoltre, si tende a verificare un raggruppamento di geni in base alla specie o famiglia, più che in base alla funzione. Ad esempio, è evidente l'isolamento delle sequenze di *Clarkia* e di *Antirrhinum*, come invece il fatto che sono raggruppate insieme la maggioranza delle sequenze di *Salvia*, *Rosmarinus*, *Ocimum*, *Lavandula* e *Mentha*, tutte appartenenti alla famiglia delle *Lamiaceae*.

Detto in altre parole, l'omologia tra sintasi diverse della stessa specie o famiglia, è maggiore che tra sintasi di specie diverse che sintetizzano lo stesso prodotto (Hosoi *et al* 2004). Questo supporta l'ipotesi (Bohlmann *et al* 1998, Martin *et al* 2004), che questi enzimi si siano evoluti indipendentemente in angiosperme e gimnosperme, e che piccole mutazioni nella sequenza possano facilmente alterare la specificità dell'enzima.

Quest'ipotesi è plausibile, dato che i geni delle famiglie TPS-a, TPS-b, TPS-d, TPS-f e TPS-g sono coinvolti nel metabolismo secondario e sono presenti in un elevato numero di copie; ad esempio, nei genomi di piante come *Arabidopsis*, *Oryza* e *Populus*, (che pur non producono grandi quantità di sostanze odorose) sono state individuate 30-40 proteine putativamente appartenenti a questa famiglia. In *Vitis vinifera*, una pianta che invece sintetizza una grande quantità e varietà di composti odorosi, sono state ritrovate 89 terpene sintasi funzionali e 27 pseudogeni (The French-Italian Public Consortium for Grapevine Genome Characterization, 2007).

Essendo tali geni impiegati nel metabolismo secondario, una loro mutazione è difficilmente letale e può tramandarsi e venir selezionata positivamente. Questo è particolarmente vero in una famiglia genica come quella delle terpene sintasi, in cui un aumento del numero di geni, con conseguente arricchimento del profilo aromatico di fiori e frutti, è generalmente vantaggioso per la pianta che può così attirare un maggior numero e varietà di organismi pollinatori.

Come controprova, si possono osservare invece le diterpene sintasi che partecipano al metabolismo primario (TPS-c e TPS-e), che sono presenti in minor numero di copie e sono più simili ad altre proteine TPS-c e TPS-e di specie diverse, che non a terpene sintasi del metabolismo secondario della stessa specie. Tali caratteristiche ben si adattano a geni vitali per la pianta, la cui mutazione è sfavorita perché presenta un elevato tasso di letalità e sono quindi tendenzialmente piuttosto conservati.

I geni CaMTS si comportano nella maniera tipica per i geni della famiglia TPS-b, in quanto nell'albero filogenetico sono più vicini tra loro di quanto non siano a geni di altre specie. E' interessante notare che i geni risultati più simili a quelli di caffè siano alcune monoterpene sintasi dei generi *Vitis*, *Quercus* e *Citrus* nonostante queste piante, appartenenti alla sottoclasse delle *rosids*, siano filogeneticamente più lontane da *Coffea* delle *Lamiaceae*, che come il caffè appartenono alle *asterids* e sono ben rappresentate nell'albero da numerosi geni di salvia, basilico, lavanda e menta.

La singola specie che ha trovato la maggior omologia con i geni di *Coffea arabica*, sia come numero di geni omologhi che per il loro elevato grado di somiglianza, è sicuramente *Vitis vinifera*. In questo caso bisogna tener conto che il genoma della vite è stato di recente sequenziato completamente ed è inoltre risultato molto ricco di terpene sintasi, per cui questi geni di vite sono certamente sovrarappresentati nei database rispetto alle terpene sintasi di altre piante che non hanno beneficiato di progetti genoma dedicati.

E' comunque intrigante osservare come le prime monoterpene sintasi isolate in caffè trovino una così elevata somiglianza con geni di *Vitis*, un'altra pianta di grande importanza economica, e come il caffè utilizzata dall'uomo soprattutto per le caratteristiche aromatiche dei suoi frutti.





## 5. CONCLUSIONI

Dopo un'accurata analisi delle sequenze dei geni di monoterpene sintasi già noti in altre piante, sono stati disegnati alcuni primer degenerati che hanno permesso di isolare anche in *C. arabica* tre trascritti putativamente identificabili come monoterpene sintasi.

Questi trascritti (CaMTS 1, CaMTS 4 e CaMTS Cc) sono stati amplificati mediante la costruzione di primer specifici, e con la tecnica RACE è stato possibile ricostruirne le estremità 3' e 5' ed ottenerne la sequenza completa.

I tre geni sono risultati espressi nel fiore e nella drupa, ma la loro espressione cessa prima della completa maturazione del frutto –in momenti diversi per i diversi geni– ed è assente nel seme. Questo comportamento è caratteristico dei composti di riserva, sintetizzati durante lo sviluppo del frutto ed accumulati nel seme, che a maturazione completata entra in uno stato di quiescenza, cessando la quasi totalità delle sue attività biologiche fino al momento della germinazione.

Le tre sequenze CaMTS sono state poi caratterizzate tramite BLAST e analisi filogenetica. Tutti e 3 i geni sono risultati altamente simili a monoterpene sintasi di *Vitis vinifera* (sia putative che già caratterizzate), oltre che ad analoghi geni di *Quercus ilex* e *Citrus*. Il programma MEGA 4.1 è stato utilizzato per costruire un albero filogenetico delle più rappresentative terpene sintasi depositate nei database pubblici. All'interno dell'albero tutte le sequenze di *Coffea arabica*, che pur presentano alcune differenze soprattutto nella regione N-terminale, sono state raggruppate all'interno della famiglia TPS-b, che riunisce la gran maggioranza delle monoterpene sintasi delle angiosperme.

Come più volte indicato in letteratura (Bohlmann *et al* 1998, Hosoi *et al* 2004), la funzionalità specifica di una terpene sintasi non può essere dedotta esclusivamente dalla sequenza aminoacidica, in quanto questa tende ad essere più conservata tra sintasi diverse della stessa specie, che tra la stessa sintasi in specie diverse. In particolare, all'interno di una stessa specie sequenze molto simili possono codificare per enzimi che sintetizzano terpeni diversi. Ciò sta a significare che solo un'analisi funzionale può stabilire con certezza il prodotto di ogni singolo gene.

L'analisi funzionale delle terpene sintasi viene svolta di norma facendo sovraesprimere le proteine in un sistema batterico (Vogel *et al* 1996, Bohlmann *et al* 1998, Steele *et al* 1998, Dudareva *et al* 2003, Faldt *et al* 2003, Hosoi *et al* 2004, Martin *et al* 2004). La coltura batterica viene poi lisata, e la proteina purificata (o anche il semplice lisato grezzo) vengono utilizzate per il successivo test funzionale.

Questo consiste nel porre la proteina ricombinante in un appropriato ambiente di reazione, in cui trova le condizioni e i cofattori adatti per la sua attività enzimatica, ed aggiungervi il precursore appropriato: geranil pirofosfato (GPP) per le monoterpene sintasi, farnesil pirofosfato (FDP) per le sesquiterpene sintasi, geranilgeranil pirofosfato (GGDP) per le diterpene sintasi. La reazione viene lasciata procedere a temperatura ambiente in una provetta sigillata in cui è stata inserita una fibra SPME (Solid Phase Micro Extraction), che adsorbe i composti volatili consentendo poi una loro rapida analisi tramite GC-MS. Fornendo alla proteina tutti e tre i precursori in tre esperimenti separati, è possibile quindi stabilire se questa sia una mono-, di- o sesqui-terpene sintasi.

Il passo successivo di questa ricerca sarà quindi certamente la caratterizzazione funzionale dei geni CaMTS individuati, tramite le tecniche sopradescritte. Al momento della stesura di questa tesi (marzo 2008) il lavoro in questo senso è già stato avviato, grazie alla collaborazione con il laboratorio analitico AromaLab (illycaffè SpA), che ha una notevole esperienza di ricerca nel campo dei composti volatili che compongono l'aroma di caffè, e presso il quale saranno svolte le analisi gas-cromatografiche. E' già stato fatto un esperimento preliminare e si prevede di portare a compimento la caratterizzazione dei geni entro qualche mese.

L'isolamento e caratterizzazione dei geni CaMTS riveste una notevole importanza sia scientifica che applicativa.

Innanzitutto sono i primi geni di monoterpene sintasi a venire isolati nel caffè, pianta la cui grande importanza economica è legata anche al suo elevato contenuto di sostanze volatili, che ne compongono l'aroma e ne rendono piacevole il consumo. Lo studio di queste proteine potrà dare notevole impulso alla comprensione del complesso profilo aromatico del caffè, e alle relazioni tra le qualità organolettiche della bevanda e le caratteristiche genetiche della pianta d'origine.

In secondo luogo, va ricordato che i monoterpene sono ampiamente utilizzati come componenti di aromi e profumi in una varietà di industrie da quella alimentare, a quella cosmetica, a quella farmaceutica. Sono quindi una classe di composti di enorme importanza economica, e la conoscenza di geni per la loro sintesi può avere risvolti applicativi ed industriali rilevanti. Non a caso, svariate sequenze di monoterpene sintasi disponibili nei database risultano tutelate da brevetti (European Patent EP1006190, European Patent EP1231273, United States Patent 5849526) e, nel caso la funzionalità dei geni CaMTS venisse confermata, è sicuramente ipotizzabile la brevettazione di questa scoperta.

La ricerca in questo campo è certamente solo all'inizio. Va infatti ricordato che le terpene sintasi sono una famiglia genica piuttosto ampia, e il loro numero sembra essere direttamente correlato alla ricchezza del profilo aromatico della specie considerata. Infatti, tra gli organismi vegetali i cui genomi sono completati o in fase di completamento, specie che non producono grandi quantità di sostanze odorose come *Arabidopsis*, *Oryza* e *Populus*, risultano possedere 30-40 terpene sintasi. In *Vitis vinifera*, il cui profilo aromatico è invece estremamente ricco, sono state ritrovate 89 terpene sintasi funzionali e 27 pseudogeni (The French-Italian Public Consortium for Grapevine Genome Characterization, 2007).

E' ragionevole aspettarsi anche in *Coffea arabica* numeri simili, nell'ordine quindi del centinaio di geni o più. Questo innanzitutto perchè il seme è estremamente ricco di una grande varietà di molecole terpeniche; in secondo luogo, perché le grandi dimensioni e la complessità del genoma allotetraploide di *C. arabica* lasciano ipotizzare che possa esistere un gran numero di geni simili, derivati da eventi di duplicazione.

Le tecniche utilizzate nel presente lavoro potranno quindi venir impiegate nuovamente nella ricerca di questi ulteriori geni di terpene sintasi, utilizzando le informazioni dei database pubblici unite alle sequenze specifiche dei geni CaMTS caratterizzati in questo lavoro. Ciò permetterà di compiere uno studio più mirato per il genere *Coffea*, il che sarà importante dato che, come si è visto dall'analisi filogenetica, le sequenze di terpene sintasi tendono ad essere molto conservate all'interno di ogni singola

specie o famiglia; per questo è importante conoscere delle sequenze specifiche della specie studiata per andare ad isolare ulteriori sequenze omologhe.

Un'altra possibile applicazione è la trasformazione di piante per modificare l'espressione di questi geni, come ben illustrato in *Aharoni et al* (2006), che indica che l'ingegnerizzazione delle terpene sintasi nelle piante è fattibile e porta ad effetti misurabili ed osservabili.

In conclusione, si può affermare che le informazioni ricavate dallo studio dei geni CaMTS andranno ad arricchire la conoscenza di questa pianta, e permetteranno di far luce sui processi che sovrintendono alla nascita di quell'aroma che spinge ogni giorno milioni di persone a consumare una tazzina di caffè.



## 6. BIBLIOGRAFIA

**Aharoni A, Keizer L, Bouwmeester H, Sun Z, Alvarez-Huerta M, Verhoeven H, Blaas J, van Houwelingen A, De Vos R, van der Voet H, Jansen RC, Guis M, Mol J, Davis RW, Schena M, van Tunen AJ, O'Connell AP (2000):** Identification of the *SAAT* gene involved in strawberry flavor biogenesis by use of DNA microarrays. *The Plant Cell* 12: 647–661.

**Aharoni A, Giri AP, Deurlein S, Griepink F, de Kogel WJ, Verstappen F, Verhoeven HA, Jongsma MA, Schwab W, Bouwmeester HJ (2003):** Terpenoid metabolism in wild-type and transgenic Arabidopsis plants. *The Plant Cell* 15: 2866–2884.

**Aharoni A, Giri AP, Verstappen FW, Berteaux CM, Sevenier R, Sun Z, Jongsma MA, Schwab W, Bouwmeester HJ (2004):** Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *Plant Cell* 16(11): 3110-31.

**Aharoni A, Jongsma MA, Kim TY, Ri MB, Giri AP, Verstappen F, Schwab W, Bouwmeester HJ (2006):** Metabolic engineering of terpenoid biosynthesis in plants. *Phytochemistry Reviews* 5: 49–58.

**Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997):** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17): 3389- 402.

**Azevedo H, Lino-Neto T, Tavares RM (2003):** An improved method for high-quality RNA isolation from needles of adult maritime pine trees. *Plant Molecular Biology Reporter* 21: 333-338.

**Birtic S, Kranner I (2006):** Isolation of high-quality RNA from polyphenol-, polysaccharide- and lipid-rich seeds. *Phytochem Anal* 17: 144-148.

**BLAST: Basic Local Alignment Search Tool:**

<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>

**BMR Genomics: Servizio sequenziamento e analisi di DNA.**

<http://www.bmr-genomics.it/>

**Bohlmann J, Meyer-Gauen G, Croteau R (1998):** Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. USA* 95: 4126–4133.

**Bohlmann J, Martin D, Oldham NJ, Gershenzon J. (2000):** Terpenoid secondary metabolism in *Arabidopsis thaliana*: cDNA cloning, characterization, and functional expression of a myrcene/(E)-beta-ocimene synthase. *Arch Biochem Biophys* 375(2): 261-9.

**Bonnlaender B, Cappuccio R, Suggi Liverani S, Winterhalter P (2004):** Analysis of Enantiomeric Linalool Ratio in Green and Roasted Coffee. *20<sup>th</sup> International Conference on Coffee Science (ASIC)*, October 11<sup>th</sup>-16<sup>th</sup> 2004, Bangalore, India.

- Bonnlaender B (2005):** Volatile aroma compounds. In: *Illy A, Viani R (Editors): Espresso Coffee – The science of quality (2nd Edition)*, pp. 197-203. Elsevier Academic Press, London and New York.
- Bytof G, Knopp SE, Selmar D (2004):** Accumulation of the stress metabolite  $\gamma$ -aminobutyric acid (GABA) in coffee seeds (*Coffea arabica* L.) during post harvest processing. *20<sup>th</sup> International Conference on Coffee Science (ASIC)*, October 11<sup>th</sup>-16<sup>th</sup> 2004, Bangalore, India.
- Bytof G, Knopp SE, Schieberle P, Teutsch I, Selmar D (2005):** Influence of processing on the generation of  $\gamma$ -aminobutyric acid in green coffee beans. *Eur Food Res Technol* 220: 245-250.
- Bytof G, Knopp SE, Kramer D, Breitenstein B, Bergervoet J, Groot S, Selmar D (2007):** Transient occurrence of seed germination processes during coffee post-harvest treatment. *Annals of Botany* 100: 61–66.
- Cavin C, Holzhauser D, Scharf G et al (2002):** Cafestol and kahweol, two coffee specific diterpenes with anticarcinogenic activity. *Food Chem Toxicol* 40: 1155-1163.
- Chang S, Puryear J, Cairney J (1993):** Simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Rep* 11: 113-116.
- Chen F, Tholl D, D’Auria JC, Farooq A, Pichersky E, Gershenzon J (2003):** Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. *The Plant Cell* 15: 481-494.
- Clarke RJ, MacRae R (1988):** *Coffee. Volume 4 – Agronomy*. Elsevier Applied Science, London and New York.
- Clarke RJ, Vitzthum OG (2001):** *Coffee – Recent developments*. Blackwell Science, Oxford.
- Clifford MN (1999):** Chlorogenic acids and other cinnamates. Nature, occurrence and dietary burden. *Journ Sci Food Agric* 79: 362-372.
- ClustalW2 and ClustalX version 2:**  
<http://www.ebi.ac.uk/tools/clustalw2/index.html>
- Corpet F (1988):** Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res* 16(22): 10881-10890.
- Crozier A, Baumann TW, Ashihara H, Suzuki T, Waller GR (1997):** Pathways involved in the biosynthesis and catabolism of caffeine in *Coffea* and *Camellia*. *17<sup>th</sup> International Conference on Coffee Science (ASIC)*, Nairobi.
- Cseke L, Dudareva N, Pichersky E (1998):** Structure and evolution of linalool synthase. *Mol Biol Evol* 15(11): 1491-1498.

**CRIBI – Centro Interdipartimentale di Ricerca e Servizi per le Biotecnologie Innovative.** Padova, Italy. <http://www.cribi.unipd.it/>

**Crowell AL, Williams DC, Davis EM, Wildung MR, Croteau R (2002):** Molecular cloning and characterization of a new linalool synthase. *Arch Biochem Biophys* 405(1): 112-21.

**De Castro RD, Marraccini P (2006):** Cytology, biochemistry and molecular changes during coffee fruit development. *Braz J Plant Physiol* 18(1): 175-199.

**De Nardi B, Dreos R, Del Terra L, Martellosi C, Asquini E, Tornincasa P, Gasperini D, Pacchioni B, Rathinavelu R, Pallavicini A and Graziosi G (2006):** Differential responses of *Coffea arabica* L. leaves and roots to chemically induced systemic acquired resistance (SAR). *Genome* 49: 1594-1605

**Dudareva N, Cseke L, Blanc VM, Pichersky E (1996):** Evolution of Floral Scent in *Clarkia*: Novel Patterns of S-Linalool Synthase Gene Expression in the *C. breweri* Flower. *The Plant Cell* 8: 1137-1148.

**Dudareva N, Raguso RA, Wang J, Ross JR, Pichersky E (1998):** Floral Scent Production in *Clarkia breweri* - III. Enzymatic Synthesis and Emission of Benzenoid Esters. *Plant Physiol* 116: 599–604.

**Dudareva N, Pichersky E (2000):** Biochemical and Molecular Genetic Aspects of Floral Scents. *Plant Physiol* 122: 627–633.

**Dudareva N, Martin D, Kish CM, Kolosova N, Gorenstein N, Faldt J, Miller B, Bohlmann J (2003):** (E)- $\beta$ -ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon; function and expression of three terpene synthase genes of a new terpene synthase subfamily. *The Plant Cell* 15: 1227-1241.

**Dudareva N, Pichersky E, Gershenzon J (2004):** Biochemistry of Plant Volatiles. *Plant Physiol* 135: 1893–1902.

**Entrez, the life sciences search engine:**  
<http://www.ncbi.nlm.nih.gov/sites/gquery>

**European Bioinformatics Institute:**  
<http://www.ebi.ac.uk/>

#### **European Patent EP1006190**

*Fruit flavour related genes and use thereof.*

Verhoeven HA (NL), Van TAJ (NL), Aharoni A (IL), Luecker J (NL), O'connell AP (NL).

Application Number: EP19980204018

Filing Date: 12/02/1998; Publication Date: 06/07/2000

**European Patent EP1231273**

*Terpene synthase/cyclase and olefin synthase and uses thereof.*

Aharoni A (IL), Verhoeven HA (NL), Jongasma MA (NL), Bouwmeester HJ (NL).

Application Number: EP20010200488;

Filing Date: 02/12/2001; Publication Date: 08/14/2002.

**Faldt J, Martin D, Miller B, Rawat S, Bohlmann J (2003):** Traumatic resin defense in Norway spruce (*Picea abies*): methyl jasmonate-induced terpene synthase gene expression, and cDNA cloning and functional characterization of (+)-3-carene synthase. *Plant Mol Biol* 51: 119-133.

**Fischer M, Reimann S, Trovato V, Redgwell RJ (2001):** Polysaccharides of green Arabica and Robusta coffee beans. *Carbohydrate Research* 330: 93–101.

**Flament I (2001):** The volatile compounds identified in green coffee beans. In *Coffee flavor chemistry*. Chichester: J. Wiley & Sons, pp. 29-34.

**Gershenzon J, McConkey ME, Croteau RB (2000):** Regulation of Monoterpene Accumulation in Leaves of Peppermint. *Plant Physiol* 122: 205–213.

**Guterman I, Shalit M, Menda N, Piestun D, Dafny-Yelin M, Shalev G, Bar E, Davydov O, Ovadis M, Emanuel M, Wang J, Adam Z, Pichersky E, Lewinsohn E, Zamir D, Vainstein A, Weiss D (2002):** Rose scent: genomics approach to discovering novel floral fragrance-related genes. *The Plant Cell* 14: 2325-2338.

**Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ (1994):** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680.

**Hosoi M, Ito M, Yagura T, Adams RP, Honda G (2004):** cDNA isolation and functional expression of myrcene synthase from *Perilla frutescens*. *Biol Pharm Bull* 27(12): 1979-1985.

**Hyatt DC, Youn B, Zhao Y, Santhamma B, Coates RM, Croteau RB, Kang C (2007):** Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proc. Natl. Acad. Sci. USA* 104(13): 5360-5365.

**Hsiao YY, Tsai WC, Kuoh CS, Huang TH, Wang HC, Leu YL, Wu TS, Chen WH, Chen HH (2006):** Comparison of transcripts in *Phalaenopsis bellina* and *Phalaenopsis equestris* (Orchidaceae) flowers to deduce monoterpene biosynthesis pathway. *Plant Biology* 6(14).

**ICO - International Coffee Organization.** London, England.

<http://www.ico.org>



- Iijima Y, Davidovich-Rikanati R, Fridman E, Gang DR, Bar E, Lewinsohn E, Pichersky E (2004):** The Biochemical and Molecular Basis for the Divergent Patterns in the Biosynthesis of Terpenes and Phenylpropenes in the Peltate Glands of Three Cultivars of Basil. *Plant Physiol* 136: 3724-3736.
- Illy A, Viani R (2005):** *Espresso Coffee – The science of quality (2nd Edition)*. Elsevier Academic Press, London and New York.
- ISO (2000)** *Quality management systems, principles and terminology*. ISO 9000:2000. Geneva, International Organization for Standardization.
- Jia JW, Crock J, Lu S, Croteau R, Chen XY (1999):** (3R)-Linalool synthase from *Artemisia annua* L.: cDNA isolation, characterization, and wound induction. *Arch Biochem Biophys* 372(1): 143-9.
- Knopp S, Bytof G, Selmar D (2006):** Influence of processing on the content of sugars in green Arabica coffee beans. *Eur Food Res Technol* 223: 195-201.
- Kumar S, Tamura K & Nei M (2004):** MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Briefings in Bioinformatics* 5:150-163.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007):** ClustalW2 and ClustalX version 2. *Bioinformatics* 23(21): 2947-2948.
- Lashermes P, Couturon E, Moreau N, Pailard M, Louarn J (1996):** Inheritance and genetic mapping of self-incompatibility in *Coffea canephora*. *Theoret Appl Genet* 93: 458-462.
- Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A (1999):** Molecular characterization and origin of the *Coffea arabica* L. Genome. *Mol Gen Genet* 261: 259-266.
- Leroy T, Marraccini P, Dufour M, Montagnon C, Lashermes P, Sabau X, Ferreira LP, Jourdan I, Pot D, Andrade AC, Glaszmann JC, Vieira LG, Piffanelli P (2005):** Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. *Theor Appl Gene*. 111:1032-1041.
- Lewinsohn E, Schalechet F, Wilkinson J, Matsui K, Tadmor Y, Nam K-H, Amar O, Lastochkin E, Larkov O, Ravid U, Hiatt W, Gepstein S, Pichersky E (2001):** Enhanced Levels of the Aroma and Flavor Compound *S*-Linalool by Metabolic Engineering of the Terpenoid Pathway in Tomato Fruits. *Plant Physiol* 127: 1256-1265.
- Lin C, Mueller LA, McCarthy J, Crouzillat D, Petiard V, Tanksley SD (2005):** Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. *Theor Appl Genet* 112 (1): 114-130.

- Loomis WD (1974):** Overcoming problems of phenolics and quinones in the isolation of plant enzymes and organelles. *Meth Enzymol* 31: 528-545.
- Lucker J, Schwab W, van Hautum B, Blaas J, van der Plas LHW, Bouwmeester HJ, Verhoeven HA (2004):** Increased and Altered Fragrance of Tobacco Plants after Metabolic Engineering Using Three Monoterpene Synthases from Lemon. *Plant Physiol* 134: 510–519.
- Martin MD, Gershenzon J, Bohlmann J (2003):** Induction of Volatile Terpene Biosynthesis and Diurnal Emission by Methyl Jasmonate in Foliage of Norway Spruce. *Plant Physiol* 132: 1586–1599.
- Martin MD, Faldt J, Bohlmann J (2004):** Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiology* 135: 1908–1927.
- Martin MD, Bohlmann J (2004):** Identification of *Vitis vinifera* (-) $\alpha$ -terpineol synthase by in silico screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. *Phytochem* 65: 1223-1229.
- McConkey ME, Gershenzon J, Croteau RB (2000):** Developmental Regulation of Monoterpene Biosynthesis in the Glandular Trichomes of Peppermint. *Plant Physiol* 122: 215–223.
- Misako K, Kouichi M (2004):** Caffeine synthase and related methyltransferases in plants. *Front Biosci* 9:1833-42.
- Mizutani M, Nakanishi H, Ema J, Ma SJ, Noguchi E, Inohara-Ochiai M, Fukuchi-Mizutani M, Nakao M, Sakata K (2002):** Cloning of  $\beta$ -Primeverosidase from Tea Leaves, a Key Enzyme in Tea Aroma Formation. *Plant Physiol* 130: 2164–2176.
- Mosli-WaldhauserSS, Baumann TW (1996):** Compartmentation of caffeine and related purine alkaloids depends exclusively on the physical chemistry of their vacuolar complex formation with chlorogenic acids. *Phytochemistry* 42: 985-996.
- Multalin: Multiple sequence alignment with hierarchical clustering:**  
<http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html>
- Natella F, Nardini M, Giannetti I, Dattilo C, Scaccini C (2002):** Coffee drinking influences plasma antioxidant capacity in humans. *J Agric Food Chem* 50(21): 6211-6216.
- Oosterveld A, Harmsen JS, Voragen AGJ, Schols HA (2003):** Extraction and characterization of polysaccharides from green and roasted *Coffea arabica* beans. *Carbohydrate Polymers* 52: 285–296.

- Pallavicini A, Del Terra L, De Nardi B, Rovelli P, Graziosi G (2001):** A catalogue of genes expressed in *Coffea arabica* L. *19<sup>th</sup> International Conference on Coffee Science* (ASIC), May 11<sup>th</sup>-15<sup>th</sup>, Trieste, Italy.
- Pallavicini A, Del Terra L, Sondhal MR, Guerreiro-Filho O, Asquini E, Martellosi C, Rajkumar R, Dreos R, De Nardi B, Graziosi G (2004):** Transcriptomics of resistance response in *Coffea arabica* L. *20<sup>th</sup> International Conference on Coffee Science* (ASIC), October 11<sup>th</sup>-16<sup>th</sup>, Bangalore, India.
- Paré PW, Tumlinson JH (1997):** De Novo Biosynthesis of Volatiles Induced by Insect Herbivory in Cotton Plants. *Plant Physiol* 114: 1161-1167.
- Pichersky E, Raguso RA, Lewinsohn E, Croteau R (1994):** Floral Scent Production in *Clarkia* (Onagraceae) - I. Localization and Developmental Modulation of Monoterpene Emission and Linalool Synthase Activity. *Plant Physiol* 106: 1533-1540.
- Pichersky E, Lewinsohn E, Croteau R (1995):** Purification and characterization of S-linalool synthase, an enzyme involved in the production of floral scent in *Clarkia breweri*. *Archiv Biochem Biophys* 316: 803-807.
- Pichersky E, Gang DR (2000):** Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends in plant science* Vol. 5, No. 10: 439-445.
- Pichersky E, Gershenzon J (2002):** The formation and function of plant volatiles: perfumes for pollinator attraction and defense. *Curr Opin Plant Biol* 5: 237-243.
- Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A, Hamon P (2006):** SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Mol Genet Genomics* 276: 436-449.
- Primer3 PCR primer design program:**  
<http://frodo.wi.mit.edu/>
- Raina SN, Mukai Y, Yamamoto M (1998):** *In situ* hybridization identifies the diploid progenitor species of *Coffea arabica* (Rubiaceae). *Theoret Appl Genet* 97: 1204-1209.
- Redgwell RJ, Curti D, Fischer M, Nicolas P, Fay LB (2002a):** Coffee bean arabinogalactans: acidic polymers covalently linked to protein. *Carbohydrate Research* 337: 239-253.
- Redgwell RJ, Trovato V, Curti D, Fischer M (2002b):** Effect of roasting on degradation and structural features of polysaccharides in Arabica coffee beans. *Carbohydrate Research* 337: 421-431.

- Rogers WJ, Bèzard G, Deshayes A, Meyer I, Pètiard V, Marraccini P (1999):** Biochemical and molecular characterization and expression of the 11S-type storage protein from *Coffea arabica* endosperm.
- Salmona J, Dussert S, Descroix F, deKochko A, Bertrand B, Joet T (2008):** Deciphering transcriptional networks that govern *Coffea arabica* seed development using combined cDNA array and real-time RT-PCR approaches. *Plant Mol Biol* 66: 105-124.
- Selmar D, Bytof G, Knopp SE (2001):** New Aspects of Coffee Processing: The Relation Between Seed Germination and Coffee Quality. *19<sup>th</sup> International Conference on Coffee Science (ASIC)*, May 11<sup>th</sup>-15<sup>th</sup> 2001, Trieste, Italy.
- Selmar D, Bytof G, Knopp SE, Bradbury A, Wilkens J, Becker R (2004a):** Biochemical Insights into Coffee Processing: Quality and Nature of Green Coffees are Interconnected with an Active Seed Metabolism. *20<sup>th</sup> International Conference on Coffee Science (ASIC)*, October 11<sup>th</sup>-16<sup>th</sup> 2004, Bangalore, India.
- Selmar D, Hunecke D, Junghartchen I, Breitenstein B, Bytof G, Knopp SE (2004b):** Expression of isocitrate lyase and the abundance of  $\beta$ -tubulin: feasible markers for the estimation of germination processes in differently processed green coffees. *20<sup>th</sup> International Conference on Coffee Science (ASIC)*, October 11<sup>th</sup>-16<sup>th</sup> 2004, Bangalore, India.
- Selmar D, Bytof G, Knopp SE, Breitenstein B (2006):** Germination of Coffee Seeds and its Significance for Coffee Quality. *Plant Biol* 8: 260–264.
- Shalit M, Guterman I, Volpin H, Bar E, Tamari T, Menda N, Adam Z, Zamir D, Vainstein A, Weiss D, Pichersky E, Lewinsohn E (2003):** Volatile ester formation in roses. Identification of an acetyl-coenzyme A geraniol/citronellol acetyltransferase in developing rose petals. *Plant Physiol* 131: 1868-1876.
- Sharma AD, Gill PK, Singh P (2003):** RNA isolation from plant tissues rich in polysaccharides. *Anal Biochem* 314: 319-321.
- Singh G, Kumar S, Singh P (2003):** A quick method to isolate RNA from wheat and other carbohydrate-rich seeds. *Plant Mol Biol Rep* 21.
- Smart LB, Nall NM, Bennett AB (1999):** Isolation of RNA and Protein from Guard Cells of *Nicotiana glauca*. *Plant Mol Biol Rep* 17: 371–383.
- SMART<sup>TM</sup> cDNA Library Construction Kit User Manual** - Clontech Laboratories, Inc., Takara Bio Company, 2007.
- Sondahl MR, Baumann TW (2001):** Agronomy II: Developmental and cell biology. *In: Clarke RJ and Vitzthum OG (editors) Coffee: Recent Developments*, pp.202-223. Blackwell, Oxford.

- Speer K, Tewis R, Montag A (1991):** 16-O-methylcafestol – a quality indicator for coffee. *Proc. 14<sup>th</sup> ASIC Colloquium*, pp. 237-244.
- Speer K, Kolling-Speer I (2001):** Lipids. In: *Clarke RJ and Vitzthum OG (editors) Coffee: Recent Developments*, pp.224-234. Blackwell, Oxford.
- Steele CL, Crock J, Bohlmann J, Croteau R (1998):** Sesquiterpene synthases from grand fir (*Abies grandis*). *Journal of Biological Chemistry* Vol. 273, No. 4: 2078–2089.
- Stofer Vogel B, Wildung MR, Vogel G, Croteau R (1996):** Abietadiene synthase from grand fir (*Abies grandis*). *Journal of Biological Chemistry* Vol.271, No. 38: 23262-23268.
- Tai HH, Pelletier C, Beardmore T (2004):** Total RNA Isolation From *Picea mariana* Dry Seed. *Plant Molecular Biology Reporter* 22.
- Tamura K, Dudley J, Nei M & Kumar S (2007):** MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596-1599.
- The French-Italian Public Consortium for Grapevine Genome Characterization (2007):** The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463-468.
- TOPO TA Cloning® User Manual**, Invitrogen Corporation,2006.
- United States Patent 5849526**  
*Use of linalool synthase in genetic engineering of scent production.*  
Pichersky E.  
US Patent Issued on December 15, 1998
- Urgert R, Essed N, van der Weg G, Kosmeijer-Scuil TG, Katan MB (1997):** Separate effects of the coffee diterpenes cafestol and kahweol on serum lipids and liver aminotransferases. *Am J Clin Nutr* 65: 519-524.
- van der Hoeven RS, Monforte AJ, Breeden D, Tanksley SD, Steffens JC (2000):** Genetic Control and Evolution of Sesquiterpene Biosynthesis in *Lycopersicon esculentum* and *L. Hirsutum*. *The Plant Cell* 12: 2283–2294.
- van Schie CN, Haring MA, Schuurink RC (2007):** Tomato linalool synthase is induced in trichomes by jasmonic acid. *Plant Mol Biol* 64: 251-263.
- Vieira LGE, Andrade AC et al (2006):** Brazilian coffee genome project: an EST-based genomic resource. *Braz Journ Plant Physiol* 18: 95-108.
- Yeretzian C, Jordan A, Badoud R, Lindinger W (2002):** From the green bean to the cup of coffee: investigating coffee roasting by on-line monitoring of volatiles. *Eur Food Res Technol* 214: 92-104.

- Wan CY, Wilkins TA (1994):** A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L). *Anal Biochem* 223: 7-12.
- Wang J, Dudareva N, Bhakta S, Raguso RA, Pichersky E (1997):** Floral Scent Production in *Clarkia breweri* (Onagraceae) - II. Localization and Developmental Modulation of the Enzyme S-Adenosyl-i-Methionine: (Iso)Eugenol O-Methyltransferase and Phenylpropanoid Emission. *Plant Physiol* 114: 213-221.
- Weckerle B, Toth G, Schreier P (2003):** Linalool disaccharides as flavour precursors from green coffee beans (*Coffea arabica*). *Eur Food Res Technol* 216(1): 6-10.
- Wise ML, Savage TJ, Katahira E, Croteau R (1998):** Monoterpene Synthases from Common Sage (*Salvia officinalis*). *Journal of Biological Chemistry* Vol. 273, No. 24: 14891–14899.
- Wu Y, Llewellyn DJ, Dennis ES (2002):** A Quick and Easy Method for Isolating Good-Quality RNA From Cotton (*Gossypium hirsutum* L.) Tissues. *Plant Mol Biol Rep* 20: 213–218.