

Università degli Studi di Firenze



SCUOLA DI DOTTORATO UBALDO MONTELATICI
DOTTORATO DI RICERCA IN SCIENZE GENETICHE
X X I I I ° CICLO
SETTORE SCIENTIFICO DISCIPLINARE BIO18

Caratterizzazione molecolare della famiglia genica della Calcone Sintasi in *Olea europaea* ed analisi della variabilità genetica in cultivar italiane

Coordinatore:

Prof.ssa Milvia Luisa Racchi

Tutor Prof. Marcello Buiatti

Candidato Dott. Francesco Nieddu

Indice

1 Introduzione	Pag.1
1.1 La storia dell'Olivo è parte integrante della storia dell'uomo.	Pag.1
1.2 L'olivo: la botanica, la sistematica, i tratti generali.	Pag.5
1.3 L'olio di oliva	Pag.7
1.4 L'olivo selvatico	Pag.18
1.5 I Marcatori Molecolari	Pag.20
1.6 Il pathway dei fenilpropanoidi	Pag.24
1.7 La famiglia genica della Calcione Sintasi.	Pag.33
2 Scopo del lavoro	Pag.40
3 Materiali e metodi	Pag.42
4 Risultati	Pag.65
4.1 Isolamento e caratterizzazione molecolare della famiglia genica della calcione sintasi in <i>Olea europaea</i> L.	Pag.65
4.1.1 Analisi in silico dei frammenti chsA e chsB.	Pag.66
4.1.2 Isolamento e caratterizzazione molecolare della sequenza completa di chsA e chsB mediante l'uso di primer degenerati.	Pag.71
4.1.3 Analisi del numero di copie dei geni della famiglia chs in olivo mediante Southern blot	Pag.83
4.1.4 Identificazione di un terzo paralogo del gene chs	Pag.86
4.2 Isolamento dei geni chsA, chsB e chsC in cultivar di olivo della Toscana.	Pag.99
4.2.1 Premessa	Pag.99
4.2.2 La scelta delle cultivar	Pag.100
4.2.3 Isolamento delle sequenze chsA, chsB e chsC mediante amplificazione PCR.	Pag.108

4.2.3.1. Isolamento di chsA in cultivar toscane.	Pag.109
4.2.3.2. Isolamento di chsB in cultivar toscane.	Pag.115
4.2.3.3. Isolamento di chsC in cultivar toscane.	Pag.121
4.3 L'introne 1 di chsA come marcatore	Pag.129
4.4 I marcatori molecolari ISSR: analisi della variabilità genetica in cultivar calabresi e siciliane di olivo	Pag.138
5 Discussione e conclusioni	Pag.142
6 Bibliografia	Pag.151

Capitolo 1

Introduzione

Capitolo 1.1. La storia dell'Olivo è parte integrante della storia dell'uomo

Atena, dea greca della sapienza e della saggezza, delle arti e della tessitura, era in competizione con Poseidone, dio del mare, dei cavalli e dei terremoti, per diventare la divinità protettrice di Atene, città ai tempi ancora senza un nome. Le due divinità si accordarono in questo modo: entrambi avrebbero fatto un regalo ai cittadini, e sarebbe poi spettato a loro decidere. Poseidone allora piantò il suo tridente nel terreno e da esso sgorgò una sorgente d'acqua. Questa garantiva ai cittadini una fonte di abbeveraggio e di commercio. L'acqua però era salmastra e non molto gradevole da bere. Atena offrì loro il primo albero di olivo adatto alla coltivazione.

L'olivo regalò al popolo legname, cibo e un eccellente olio. In virtù di questo dono i cittadini scelsero la dea Atena e la città venne chiamata Atene, in suo onore.

L'olivo dunque nei miti greci e in generale nella cultura classica greca era un dono diretto degli déi: questo ci fa ben comprendere, assieme alle numerose testimonianze paleo-biologiche l'importanza di questa coltivazione nella cultura classica che a partire da mille anni prima di Cristo abbracciò e governò l'intera area mediterranea.

Tuttavia la coltivazione dell'olivo iniziò molto tempo prima, ed è davvero significativo come la storia agronomica di questa coltivazione sia intimamente correlata ed abbia costantemente accompagnato le vicende di tutte le grandi civiltà del passato che si sono susseguite nel bacino del mediterraneo.

Per indagare le origini antiche di questa coltivazione occorre arrivare letteralmente all'età della pietra, più precisamente all'ultimo dei tre periodi che la contraddistinguono: il neolitico. Reperti preistorici risalenti a questo periodo indicano la presenza di olive nella dieta umana. Esistono altre testimonianze assai interessanti, ad esempio il ritrovamento di alcuni frantoi primitivi localizzati

nell'area mediorientale, che confermano l'utilizzo dell'olivo e delle sue drupe a partire almeno dal 5000 AC (Zohary and Hopf, 1994).

Ed è proprio il Medio Oriente, universalmente riconosciuto come la culla dell'agricoltura, il luogo di origine della coltivazione dell'olivo, che nei secoli a venire colonizzerà progressivamente l'intero bacino del mediterraneo e in seguito il resto del mondo.

Secondo diversi autori la prima grande diffusione dell'olivo attraverso il Mar Mediterraneo sarebbe avvenuta passando per almeno tre centri di diversificazione.

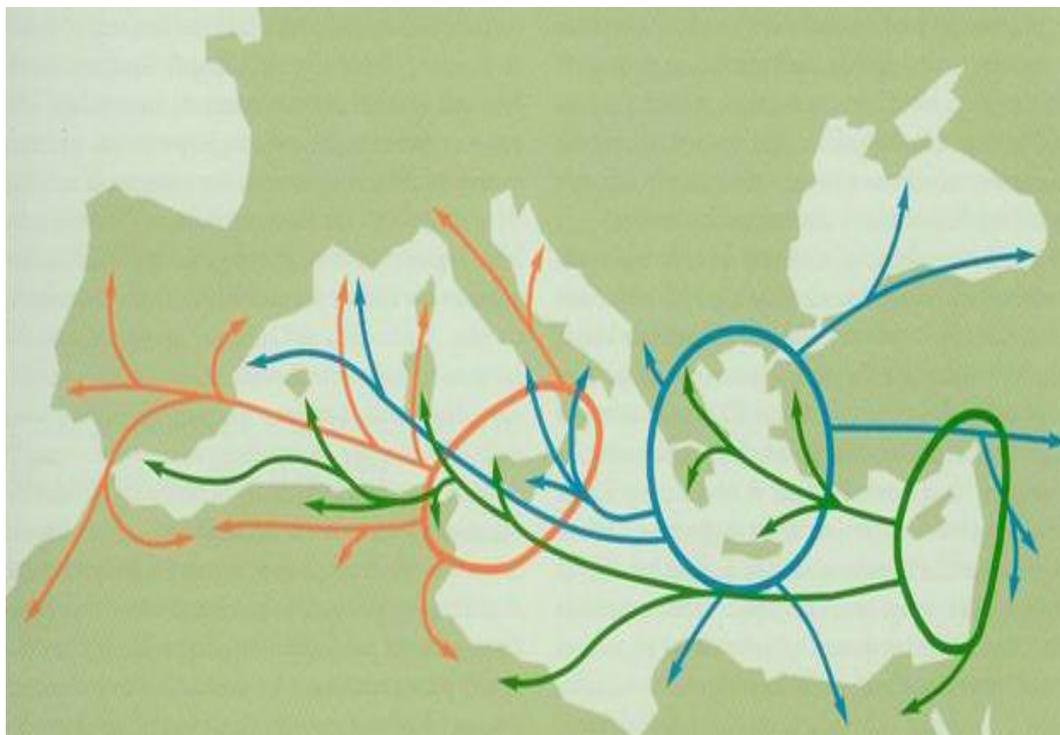


Figura 1. I centri di diversificazione: Primario (verde) secondario (azzurro) e terziario (arancio).

Il primo di questi è rappresentato dalla zona di origine, ovvero l'area mediorientale (tra il V e III millennio a.C.), il centro di diversificazione secondario, tra il III e il II millennio A.C. nella attuale Grecia e più in generale nella zona bagnata dal mare Egeo. Il centro di diversificazione terziario è invece

localizzato tra l'attuale Italia meridionale e l'Africa settentrionale durante l'epoca romana (Acerbo, 1937; Zohary, 1973; Zohary and Spiegel-Roy, 1975).

Appare chiaro come la diffusione dell'olivo, come del resto quella di altre piante coltivate, sia stata segnata dalle grandi civiltà del passato. Quello che occorre evidenziare è che in questo caso davvero tutte le civiltà che si sono succedute nei secoli nel bacino del Mediterraneo, a partire dalla civiltà egizia, passando per la civiltà minoica, greca, fenicia, etrusca, romana, sono fortemente legate all'olivo.

Numerosissime testimonianze di ciascuna civiltà sono a disposizione di paleontologi e storici. Tra i reperti più importanti citiamo, uno per tutti, il famoso codice di Hammurabi (XVIII sec. a.C.) che riporta un'ampia e dettagliata legislazione sul commercio dell'olivo.

Tra le grandi civiltà citate fino ad adesso occorre rimarcare quanto l'antica civiltà greca abbia contribuito, e in modo assolutamente determinante, alla diffusione verso l'occidente dell'olivo. Come abbiamo visto, nella penisola greca è localizzato il grande centro di diversificazione secondario. Ed è qui che in tutta probabilità è stato effettuato il primo grande lavoro di selezione. Ed è sempre grazie alla Magna Grecia che la coltivazione dell'olivo si è diffusa nelle colonie mediterranee, situate nell'Italia meridionale, particolarmente l'attuale Sicilia e Calabria, che insieme al regno Cartaginese (situato approssimativamente entro i confini dell'attuale Tunisia) costituiscono il centro di diversificazione terziario. La diffusione successiva dell'olivo in tutto il bacino del Mediterraneo si è poi consolidata e potenziata durante le civiltà che si susseguiranno in Europa. Meritevoli di citazione sia gli Etruschi che in seguito i Romani, per il loro apporto sostanziale al consolidamento dell'olivo nella penisola italiana inizialmente nella parte centro meridionale, poi anche in pianura padana, in Liguria e in generale tutto il nord Italia.

Perlopiù in epoca romana l'olivo, in particolare grazie al suo prodotto d'elezione, l'olio di oliva, raggiunge una grandissima diffusione e diventa un genere di larghissimo consumo, sia per scopi gastronomici che come ingrediente principale di creme e unguenti, medicali e di bellezza.

Naturalmente la grande richiesta fa progredire in modo esponenziale la coltivazione sia in termini quantitativi che qualitativi, mirati questi ultimi essenzialmente ad una selezione delle piante più produttive.

Il consumo di olio di oliva è assolutamente trasversale nella società romana: è consumato da tutte le classi sociali, dagli imperatori come dagli schiavi. Tuttavia si assiste ad una diversificazione del prodotto su base economica che risponde ad una differente qualità. L'olio di oliva viene classificato e suddiviso in cinque fasce di qualità, che possiamo definire come i precursori delle attuali denominazioni di origine protetta e dei disciplinari di produzione. Il prodotto commerciato risponde a parametri ben definiti, in particolare basati sul grado di maturazione delle olive al momento della frangitura. La classificazione comprende le tipologie "oleum ex albis ulivis" proveniente dalla spremitura delle olive verdi, "oleum viride" proveniente da olive a uno stadio leggermente più avanzato di maturazione, "oleum maturum" proveniente da olive mature, "oleum caducum" proveniente da olive cadute a terra e "oleum cibarium" proveniente da olive passite. Quest'ultima fascia qualitativa è essenzialmente destinata all'alimentazione degli schiavi.

L'olio di oliva è presente come alimento centrale in particolare nel rancio dell'esercito romano, numericamente assai rilevante, che notoriamente rappresenta il cuore pulsante della società ed è distribuito in un territorio molto esteso, che abbraccia più continenti. Per cui per i romani è fondamentale che l'olio di oliva sia ben distribuito e diffuso in tutto il territorio dell'impero e venga trasportato costantemente e incessantemente tutte le grandi vie di comunicazione. Successivamente alla caduta dell'Impero Romano di Occidente la produzione dell'olio di oliva conosce la prima grande "crisi di settore". Tuttavia la sensibile diminuzione della domanda non porta affatto all'abbandono delle coltivazioni in senso generale, che anzi resistono, seguendo fasi altalenanti, a tutti i passaggi storici per arrivare come sappiamo al nostro tempo.

Infatti la coltivazione dell'olivo rimane diffusa pressoché in tutto il bacino del Mediterraneo, aiutata anche, in particolare durante l'epoca Medioevale, dal crescente utilizzo non alimentare, ad esempio come unguento religioso e oppure come combustibile nelle lanterne.

In epoca rinascimentale, in particolare per quanto riguarda l'olivicoltura toscana, l'olivo conosce un nuovo periodo d'oro. La Famiglia de' Medici rilancia notevolmente la produzione con sostanziose agevolazioni, non ultime quelle fiscali, contribuendo al radicamento che progressivamente dagli etruschi ad oggi, ha plasmato le campagne e rende inscindibile la terra toscana dai suoi oliveti.

Capitolo 1.2 L'olivo: la botanica, la sistematica, i tratti generali.

L'olivo (*Olea europaea* L.) è una pianta diploide ($2n=46$) prevalentemente allogama, che mostra differenti livelli di autocompatibilità a seconda della cultivar e con un alto grado di eterozigosi. All'interno della sub specie *Olea europaea* subsp. *europaea* l'olivo si diversifica in un grande numero di varietà, stimate in circa 1200, delle quali almeno 600 presenti in Italia, con caratteri altamente eterogenei. Le cultivar possono essere considerate come varietà di origine sconosciuta, propagate da millenni per via vegetativa tramite talee ed innesti (Zohary and Hopf 2000) anche se è possibile che in passato l'olivo sia stato propagato per seme (Breton et al. 2006).

Molte cultivar hanno areali di distribuzione estremamente ridotti (Besnard et al., 2001) a fronte di alcune che invece sono estremamente diffuse in zone di produzione di diversi continenti (Bronzini et al., 2002; Rotondi et al., 2003).

Anche all'interno delle cultivar esiste una certa documentata eterogeneità che consente di definire le varietà come cultivar policlonali con fenotipi eterogenei (Mozzalupo et al., 2010). Dal punto di vista della sistematica l'olivo appartiene all'Ordine *Lamiales* (della classe Dicotiledoni e Divisione Angiosperme) e alla Famiglia *Oleaceae*, che comprende ben 27 generi e dalle 400 alle 900 specie (a seconda degli autori) tutte legnose e adattate ai climi temperati o tropicali. La specie *Olea europaea* (L.) comprende a sua volta le subspecie *europaea*, della quale fanno parte l'olivo coltivato e selvatico, e le subspecie *cerasiformis*, *cuspidata*, *guanchica*, *laperrinei* e *maroccana*.

L'olivo è una pianta arborea sempreverde, la cui fase vegetativa è pressoché continua, eccetto che per una breve attenuazione nel periodo invernale. È una pianta assai longeva con una aspettativa di vita che si attesta sui 500 anni (Rhizopoulou, 2007): questa importante caratteristica è sicuramente dovuta a numerosi fattori intrinseci della pianta, ma un ruolo determinante è sicuramente attribuibile alla grande capacità rigenerativa delle parti della pianta che subiscono un danneggiamento, siano esse ipogee o epigee.

È una specie tipicamente basitone, che cioè assumerebbe la forma tipicamente conica, senza l'intervento dell'uomo. Il suo tronco è contorto, la corteccia è grigia e liscia e con l'età assume una tendenza alle spaccature.

Il legno dell'olivo è assai pregiato e profumato, viene utilizzato prevalentemente per la fabbricazione di mobili di pregio in legno massello.

Caratteristica del tronco, sin dalla forma giovanile, è la formazione di iperplasie nella zona del colletto appena sotto la superficie del terreno; simili strutture si possono ritrovare sulla branche. Queste formazioni non sono attribuibili alla presenza di fattori di tipo parassitario ma a squilibri ormonali oppure ad eventi di tipo microclimatico.

Le radici sono prevalentemente di tipo fittonante nei primi tre anni di età, ma dal quarto anno in poi si trasformano quasi completamente in radici di tipo avventizio, superficiali. Queste garantiscono alla pianta un'ottimo vigore anche su terreni difficili e inospitali, ad esempio quelli rocciosi, dove lo strato di terreno che contiene sostanze nutrienti è limitato a non più di alcune decine di centimetri. L'olivo presenta foglie opposte, coriacee, semplici, intere, ellittico-lanceolate, con picciolo corto e margine intero, spesso revoluta e con la pagina superiore glabra, di colore verde glauco. La pagina inferiore è invece dotata di peli stellati ed è di colore argentato - bianco.

Le gemme sono per lo più di tipo ascellare e il fiore è ermafrodito, piccolo, con calice di 4 sepali e corolla di petali bianchi.

I fiori sono raggruppati in numero dai dieci ai quindici, in infiorescenze a grappolo, chiamate mignole, emesse all'ascella delle foglie dei rametti dell'anno precedente. La mignolatura precede la fioritura *sensu stricto* ed ha inizio verso marzo-aprile. Chiaramente il periodo di fioritura è dipende sia dalla zona che dalla cultivar, ma in ogni caso avviene tra maggio e giugno.

Il frutto, come nelle altre *Oleaceae*, è una drupa globosa, elissoideale o ovoidale, talvolta asimmetrica. Il suo peso è variabile e va da 1 grammo a 6 grammi, a seconda della varietà in esame, della tecnica di coltura utilizzata e dell'andamento climatico stagionale. Generalmente il peso varia da uno a tre grammi per le olive da olio e dai quattro ai sei grammi per le olive da tavola.

L'oliva è uno dei pochi frutti dal quale è possibile estrarre un olio per spremitura diretta. Generalmente gli altri oli vengono estratti a partire dal seme e utilizzando generalmente procedimenti estrattivi di tipo chimico.

Il colore dell'esocarpo è piuttosto variabile e dipende dalla cultivar di appartenenza: ad esempio può cambiare dal verde al rosso-viola.

Il mesocarpo, è carnoso e contiene il 25-30 % di olio. Il seme è contenuto in un endocarpo legnoso e ovoidale, ruvido, di colore marrone.

Capitolo 1.3 L'olio di oliva

L'olio di oliva è un prodotto alimentare caratterizzato da un contenuto elevato di grassi mono e polinsaturi. Nella denominazione “vergine” si ricava esclusivamente dalla frangitura e spremitura delle olive con metodi meccanici. In altre tipologie merceologiche, l'olio di oliva può essere ottenuto per rettificazione degli oli vergini o per estrazione chimica mediante l'utilizzo di solventi a partire dalla sansa di olive. Naturalmente negli ultimi due casi la qualità, le proprietà alimentari e organolettiche oltre al costo finale del prodotto variano decisamente.

La raccolta delle olive avviene da ottobre a dicembre a seconda della latitudine, delle condizioni climatiche e delle cultivar. La raccolta con metodi tradizionali è stata perlopiù sostituita dall'utilizzo di abbacchiatori meccanici anche se in alcune zone (specie in oliveti di piccole dimensioni) resiste la raccolta tradizionale mediante l'utilizzo di pertiche, “pettini”.

E' noto come resista, particolarmente nelle regioni meridionali, la raccolta a mano, estremamente faticosa e lenta, che prevede però una accurata selezione delle olive.

La produzione dell'olio vergine avviene per spremitura in oleifici che possono utilizzare sia macine in pietra che impianti di movimentazione meccanica o idraulica. La chiave qualitativa che determina le proprietà del prodotto finale è rappresentata dalla temperatura alla quale avviene la spremitura. Temperature alte riducono proporzionalmente la qualità dell'olio di oliva prodotto. Per questo attualmente vengono privilegiati oleifici che mantengono le macine in pietra o macchinari con raffreddamento ad acqua che permettono di controllare la temperatura mediante termosonde situate nei vari passaggi del processo.

In seguito alla molitura viene estratto il mosto d'olio e separato dalla fase acquosa che viene successivamente chiarificato e poi imbottigliato.

Dalla restante sansa di olive è possibile l'estrazione chimica di olio di sansa di oliva.

Sulla base della qualità, la Unione Europea suddivide l'olio di oliva in diverse denominazioni commerciali e all'interno di queste esistono i marchi e le denominazioni di qualità delle quali accenneremo in seguito. La direttiva comunitaria 136/66/CEE, il regolamento 2568/91 e il regolamento 1989/03 disciplinano la denominazione del prodotto e lo suddividono in diverse categorie, dipendenti da numerosi fattori come ad esempio il metodo di spremitura e la acidità espressa in peso di acidità di acido oleico.

L'olio di oliva viene classificato come:

1. Olio extravergine di oliva, ottenuto esclusivamente tramite spremitura meccanica a bassa temperatura ed acidità minore di 0.8%;
2. Olio vergine di oliva: ottenuto per spremitura meccanica, acidità minore di 2%.

Gli oli lampanti sono classificati come:

1. Olio di oliva vergine lampante, che è ottenuto per spremitura meccanica, e presenta una alta acidità e non è commerciabile in questa forma;
2. Olio di oliva rettificato, che viene è trattato chimicamente per eliminare la alta acidità del precedente, non è commerciabile direttamente (è completamente privo delle qualità organolettiche tipiche dell'olio di oliva);
3. Olio di oliva: che è composto sia di oli oliva vergini che lampanti e non deve presentare una acidità maggiore di 1%.

Gli oli di sansa sono invece suddivisi in:

1. Olio di sansa di oliva greggio, ottenuto con metodi chimici e non commerciabile direttamente;
2. Olio di sansa di oliva rettificato, è ottenuto dall'olio di sansa di oliva greggio per rettificazione
3. Olio di sansa di oliva, ottenuto dall'olio di sansa di oliva rettificato che viene miscelato con olio di oliva vergine per essere commercializzato con tale denominazione.

Come discusso nel paragrafo precedente, oltre che dal punto di vista alimentare l'olio di oliva rappresenta una importante risorsa economica e commerciale nell'intero bacino del Mediterraneo fin dai tempi antichi. Nonostante questo florido e millenario commercio, il consumo di olio di oliva fino all'inizio del

nuovo millennio era essenzialmente ristretto ai Paesi produttori, nei quali costituisce la base della dieta Mediterranea (Nestle, 1995; Ness, 2002) e in generale della cosiddetta “Cucina Mediterranea” (Figura 2).

Per capita consumption (food) of olive oil (liter/capita/year) in selected EU15 countries, 1972–2003.

Region	Country	1972/73	1982/83	1992/93	2002/03
Producing countries (EU15)	Spain	9.8	10.4	12.7	12.9
	Italy	13.4	11.5	12.6	14.4
	Greece	23.3	22.4	20.0	16.7
	Portugal	7.6	5.0	4.3	5.5
	France	0.4	0.4	0.7	1.8
Non-producing countries (EU15)	United Kingdom	0.06	0.05	0.2	0.7
	Germany	0.06	0.09	0.2	0.5
	Austria	0.05	0.07	0.2	0.6
	NPC average	0.05	0.07	0.2	0.6

Note: Data on consumption are averaged over the indicated years; NPC, non-producing countries.

Source: Based on FAO (2005, 2008b).

Figura 2: Consumo di olio di oliva alimentare nella dieta espresso in litri pro-capite per anno nei paesi selezionati EU15 ovvero Austria, Belgio, Danimarca, Finlandia, Francia, Germania, Grecia, Irlanda, Italia, Lussemburgo, Olanda, Portogallo, Spagna, Svezia e Regno Unito.

Recenti studi come ad esempio il lavoro di Scheidel e Krausmann (2010) e i dati messi a disposizione dai centri statistici della FAO, testimoniano però come questa tendenza stia pian piano modificandosi.

Questo nuovo trend può essere collegato a numerosi fattori; ad esempio nei Paesi non produttori, negli ultimi dieci anni si sono moltiplicate campagne di marketing, portate avanti sia dalle imprese produttrici locali o straniere che dalla commissione europea, affiancate da iniziative educative dei Governo e Ministeri nazionali dedicate alla salute e alle buone abitudini alimentari.

Oltre a questo ha inciso sicuramente anche una sempre più capillare diffusione di associazioni di cittadini, in particolare nei Paesi di cultura anglosassone, che

promuovono le virtuose abitudini alimentari della dieta mediterranea, anche nei negozi di alimentari dei Paesi non produttori si stanno diffondendo gli oli di altissima qualità, non destinati quindi all'utilizzo esclusivo nella cottura. L'arrivo dell'olio di oliva di qualità nelle grandi distribuzioni dei Paesi non produttori è una evidenza di uno spostamento della tendenza di mercato e di consumo verso questo prodotto, fino a dieci anni fa considerato di nicchia ed estremamente costoso, adesso promosso consumato e in taluni casi "scoperto" come alimento di consumo quotidiano. Numerosi produttori e giornalisti di settore sottolineano che il vero passo in avanti e la grande vittoria per il marketing dell'olio di oliva avverrà quando, anche nelle dispense dei paesi anglosassoni sarà presente una bottiglia di olio "buono" da utilizzare "per occasioni speciali" proprio come avviene per il vino o lo Champagne.

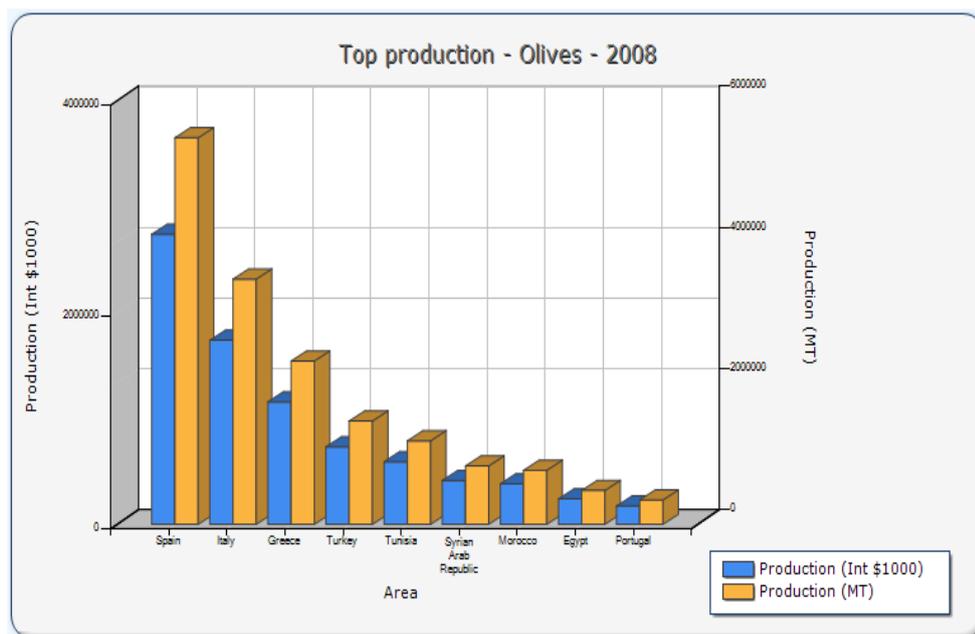


Figura 3: Campagna pubblicitaria australiana che promuove l'utilizzo di olio extravergine

Chiaramente siamo solo all'inizio di una inversione di tendenza, in quanto ad oggi dobbiamo registrare un utilizzo sempre prevalente in mercati come quello nord americano o nord europeo di grassi animali o di diversi oli di semi.

Per quanto riguarda la produzione dell'olio di oliva, gli ultimi dati ufficiali della FAO che abbiamo a disposizione, che si riferiscono all'anno 2008, evidenziano come la produzione rimanga per la stragrande maggioranza dislocata nel bacino mediterraneo. A conferma di questo e per ribadire l'importanza relativa del comparto olivicolo nel settore agroalimentare è davvero significativa la stima effettuata dalla Commissione Europea per la quale almeno un agricoltore su tre nella Unione Europea lavora nella filiera dell'olivicoltura (Commissione Europea - Direzione Generale dell'Agricoltura, giugno 2002).

Nonostante l'esame approfondito delle statistiche ci mostri un dato importante ovvero che la coltivazione olivicola si sta progressivamente diffondendo nel mondo, ovviamente in Paesi con climi assimilabili al bacino Mediterraneo, come ad esempio il Sudafrica (Costa, 1998), a ricordarci quanto questa tendenza sia ancora all'inizio resta il fatto che, per trovare un produttore extra-Mediterraneo, per l'anno 2008 si deve scendere al dodicesimo posto, dove si colloca l'Argentina (Figura 4).



Rank	Area	Production (Int \$1000)	Flag	Production (MT)	Flag
1	Spain	2739128	*	5475300	
2	Italy	1737737	*	3473600	
3	Greece	1157152	*	2313055	F
4	Turkey	732519	*	1464248	
5	Tunisia	591819	*	1183000	
6	Syrian Arab Republic	413739	*	827033	
7	Morocco	382896	*	765380	
8	Egypt	240165	*	480071	
9	Portugal	172993	*	345800	
10	Algeria	127102	*	254067	
11	Libyan Arab Jamahiriya	90048	*	180000	*
12	Argentina	75040	*	150000	F
13	Jordan	47059	*	94068	
14	Occupied Palestinian Territory	42908	*	85770	
15	Lebanon	38120	*	76200	F
16	United States of America	30316	*	60600	
17	Peru	28865	*	57700	
18	Australia	28576	*	57123	F
19	Albania	28115	*	56200	
20	Israel	24388	*	48750	

* : Unofficial figure

[]: Official data

F : FAO estimate

Figura 4. Elenco dei Paesi produttori di olio di oliva nell'anno 2008

Come documentato in figura 4, il Paese leader della produzione olivicola si conferma di gran lunga la Spagna, seguita da Italia e Grecia.

Per quanto riguarda la situazione italiana, i dati ISTAT aggiornati al 2007 assegnano alla Calabria il primato della produzione, seguita dalla Puglia. Queste due regioni rappresentano una grande fetta della produzione olivicola italiana, ed insieme valgono qualcosa come quattro milioni di quintali all'anno. Al settimo posto si conferma la Toscana, con circa centosettantamila quintali annui.

La produzione risulta marcatamente proiettata verso il mercato di alta qualità, in quanto dei circa sei milioni di quintali prodotti in media in un anno nel nostro Paese, circa quattro milioni sono riconducibili ad oli di fascia alta, ovvero extra-vergini.

E' evidente che la maggior parte degli imprenditori agricoli italiani ha scelto di giocare questa partita sul campo della qualità e sfidare su questo terreno le Nazioni concorrenti, sia quelle "storiche" che quelle "emergenti".

La fascia alta del prodotto si rispecchia essenzialmente nelle certificazioni di qualità.

L'importanza cruciale delle denominazioni, in particolare le DOP, protette da disciplinari che tutelano le origini di produzione e le coltivazioni autoctone è riflessa nelle 37 certificazioni DOP che vanta il nostro paese, a fronte delle 32 spagnole, nonostante il gap di produzione.

Riconosciuto dunque il valore del marketing, della promozione della qualità e della tipicità e della promozione pubblicitaria per il nostro paese, chi scrive stenta a comprendere come i rappresentanti politici e istituzionali oltre ai lobbisti di settore che rappresentano l'Italia nella Unione Europea abbiano acconsentito il cambio dei colori del marchio DOP, motivato con la volontà (quella sì, a nostro modo di vedere, corretta) di differenziare meglio i prodotti DOP dagli IGP (Indicazione Geografica Protetta). Ci convince poco innanzitutto la decisione assunta dalla Unione Europea di modificare i colori del marchio DOP e non quelli del Marchio IGP, questi ultimi con disciplinari meno rigidi e in linea di massima di qualità minore rispetto ai primi. Ci convince ancora meno la soluzione cromatica adottata, che si distacca dai colori europei per adottarne altri, che francamente ricordano molto (forse troppo) i colori nazionali spagnoli (Figura 5). A nostro modesto parere sarebbe stata più opportuna la adozione di una colorazione che non richiamasse esplicitamente la bandiera di uno stato.



Figura 5: Il marchio IGP e il marchio DOP con la nuova colorazione, in basso la bandiera dell'Unione Europea e del Regno di Spagna.

Il marchio DOP può essere applicato esclusivamente a prodotti alimentari dove l'intero ciclo produttivo viene svolto entro una area geografica ben definita

La peculiarità del prodotto deve riferirsi sia a fattori intrinseci naturali che a fattori antropici, vale a dire sia nelle materie prime utilizzate che nei metodi tradizionali di produzione e lavorazione, stabiliti e codificati nei Disciplinari.

I prodotti a marchio registrato devono caratterizzarsi come peculiari e non riproducibili con gli stessi standard qualitativi del prodotto finale come indicato nell'articolo 2 del regolamento CEE 2081/92.

La provenienza del prodotto e le fasi di lavorazione devono essere garantite e deve esserne dimostrata la tracciabilità.

Per controllare che tutto questo venga rispettato, le aziende che intendono commercializzare prodotti col marchio DOP sottostanno al controllo di conformità di un Ente di Certificazione indipendente nominato dal Ministero delle Politiche Agricole e Forestali.

L'Indicazione Geografica Protetta (IGP) è anch'esso un marchio di qualità della Unione Europea. Differisce dal precedente da un disciplinare meno "rigido" che attesta esclusivamente che una fase, specifica e connotante il prodotto finale, venga eseguita nel luogo di origine.

Per quanto riguarda la regione Toscana, nell'anno 2007 risulta la settima regione in Italia per quantità di olio prodotto, e vanta una tradizione millenaria e una sempre rinnovata attenzione alla qualità.

La produzione di olio, diffusa in tutte le province toscane, è estremamente frazionata e capillare: sono ben quattrocento i gli oleifici che costellano il territorio. Il valore economico dell'olivicoltura toscana è ingente ed ammonta a circa 75 milioni di Euro (ISTAT 2007).

La vocazione qualitativa dell'olio di oliva prodotto in Toscana è rintracciabile nel numero di denominazioni protette riconosciute dalla Unione Europea che sono ben tre DOP e una IGP, in particolare l' Olio extravergine di oliva Chianti Classico DOP, l'Olio extravergine di oliva Lucca DOP, l'Olio extravergine di oliva Terre di Siena DOP e l' Olio extravergine di oliva Toscano IGP (Figura 6).



Figura 6: Gli oli di oliva di qualità prodotti in Toscana: in alto i tre marchi DOP e in basso il marchio IA titolo esemplificativo abbiamo scelto di illustrare brevemente i disciplinari appartenenti alle due importanti denominazioni protette ovvero l'olio Chianti Classico DOP e dell'olio Toscano IGP.

L'olio di oliva "Chianti Classico" DOP deve essere prodotto con le olive di oliveti costituiti per almeno l'80% da piante delle varietà "Frantoio", "Correggiolo", "Moraiolo", "Leccino", da sole o congiuntamente, ed un massimo del 20% da piante di altre varietà di seguito elencate: Allora, Americano, Arancino, Ciliegino, Colombino, Correggiolo di Palesse, Cuoricino, Da Cuccare, Filare, Frantoiano di Montemurlo, Ginestrino, Giogolino, Grappolo, Gremigna Tonda, Gremigno di Fauglia, Gremigno di Montecatini, Gremignolo, Gemignolo di Bolgheri, Grossaio, Grossolana, Larcianese, Lastrino, Lazzero, Lazzero della Guadalupe, Lazzero di Prata, Leccio del Corno, Madonna dell'Impruneta, Madremignola, Mansino, Maremmano, Marzio, Maurino, Melaiolo, Mignolo, Mignolo Cerretano, Morcaio, Morchiaio, Morcone, Morello a Punta, Martellino, Olivastra di Populonia, Olivastra di Suvereto, Olivastra Seggianese, Olivo Bufalo, Olivo del Mulino, Olivo del Palone, Olivo di Casavecchia, Olivo di San Lorenzo, Ornellaia, Pendagliolo, Pendolino, Pesciatino, Piangente, Pignolo, Piturzello, Punteruolo,

Quercetano, Rama Pendula, Razzaio, Razzo, Rosino, Rossellino, Rossellino Cerretano, Rossello, Salcino, S. Francesco, S. Lazzero, Santa Caterina, Scarlinese, Selvatica Tardiva, Tondello, Trillo.

L'olio del "Chianti Classico", la cui zona di produzione coincide con quella del vino Chianti Classico DOC, comprende, i territori amministrativi dei comuni di Castelina in Chianti, Gaiole in Chianti, Greve in Chianti, Radda in Chianti per tutto il loro territorio, ed, in parte, Barberino Val d'Elsa, Castelnuovo Berardenga, Poggibonsi, San Casciano in Val di Pesa e Tavernelle Val di Pesa.

La produzione di olio ha precise limitazioni di tipo quantitativo: non può superare 650 chilogrammi per ettaro per oliveti con densità di almeno 200 piante. Per gli impianti con densità inferiore, la produzione non può superare 3,25 chilogrammi a pianta.

Per quanto concerne la raccolta, deve avvenire entro e non oltre il 31 Dicembre. Per ottenere il marchio DOP, l'olio deve essere conforme a numerose analisi sia chimiche che organolettiche.

Per l'olio Chianti Classico DOP Le analisi chimiche debbono necessariamente risultare conformi ai seguenti parametri:

1. acidità (espressa in acido oleico) massimo. 0,5%;
2. numero di perossidi massimo. 12 (mq di ossigeno);
3. estinzione all'ultravioletto K232 massimo 2,1 e K270 massimo. 0,2;
4. alto tenore di acido oleico, del 74%;
5. CMP totali (antiossidanti fenolici, metodo della Stazione sperimentale per le industrie degli oli e dei grassi) maggiori di 150 ppm;
6. Tocoferoli totali maggiori di 150 ppm.

Per quanto riguarda la valutazione organolettica, l'olio deve risultare di colore da verde intenso a verde con sfumature dorate; con aroma netto di olio di oliva e di fruttato. In particolare la scheda di assaggio con Panel-test effettuato da una giuria composta da assaggiatori professionisti deve risultare:

1. fruttato di oliva 2-4;
2. erba e/o olio 0-3;
3. amaro 1-4;

4. piccante 1-3.
5. Non è ammesso alcun tipo e livello di difetto organolettico.

Per quanto riguarda l'olio Toscano IGP, l'indicazione geografica protetta "Toscano", senza alcuna menzione geografica aggiuntiva, deve essere ottenuta dalle seguenti varietà di olive presenti, da sole o congiuntamente, negli oliveti: Americano, Arancino, Ciliegino, Frantoio, Grappolo, Gremignolo, Grossolana, Larcianese, Lazzero, Leccino, Leccio del Corno, Leccione, Madonna dell'Impruneta, Marzio, Maurino, Melaiolo, Mignolo, Moraiolo, Morchiaio, Olivastra Seggianese, Pendolino, Pesciatino, Piangente, Punteruolo, Razzaiolo, Rossellino, Rossello, San Francesco, Santa Caterina, Scarlinese, Tondello e loro sinonimi. Possono, inoltre, concorrere altre varietà presenti negli oliveti fino ad un massimo del 5%. Per quanto riguarda le rese, al fine di garantire la reale consistenza della produzione ogni anno sono definite le rese massime in olive in olio per menzione geografica aggiuntiva, area omogenea o provincia a seguito delle rilevazioni effettuate dalle Associazioni di produttori riconosciute dal Consorzio di tutela. All'atto dell'immissione al consumo l'olio extravergine di oliva a indicazione geografica protetta "Toscano" deve rispondere alle seguenti caratteristiche: colore: dal verde al giallo oro con variazione cromatica nel tempo; odore: di fruttato accompagnato da sentore di mandorla, carciofo, altra frutta matura, verde di foglia; sapore: di fruttato marcato; punteggio al panel test: livelli consentiti dalla normativa vigente.

Per quanto riguarda i parametri chimici:

1. acidità massima totale espressa in acido oleico, in peso, non eccedente grammi 0,6 per 100 grammi d'olio;
2. acido palmitico: 8 – 14 %,
3. acido palmitoleico: 0,5 - 1,5 %,
4. acido stearico: 1,1 –3 %,
5. acido oleico: 73- 83 %,
6. acido linoleico: minore di 9 %,
7. acido linolenico: minore di 0,9 %,
8. acido arachico: minore di 0,6 %,
9. acido eicosenoico: minore di 0,4 %,
10. polifenoli totali: maggiore o uguale a 60mg/Kg,

11. tocoferoli: maggiore o uguale a 40mg/Kg.

Capitolo 1.4 L'olivo selvatico

Come descritto in precedenza, le cultivar di olivo hanno una origine genetica incerta. E' dimostrato tuttavia che alcune varietà coltivate, autoctone in diversi siti nel bacino del Mediterraneo, sono state selezionate da popolazioni spontanee di olivi selvatici (Besnard et al., 2001, 2002; Breton et al., 2006). L'olivo selvatico o olivastro è generalmente riconosciuto come l'antenato selvatico dell'olivo coltivato. Ritrovamenti fossili dimostrano la presenza di olivi selvatici già in era paleolitica (Lipshitz et al, 1991). L'utilizzo da parte dell'uomo dell'olivo selvatico perdura poi in diverse parti del mediterraneo durante l'era neolitica (Lipshitz et al, 1991; Zohary and Hopf, 1993).

Le cultivar locali possono derivare da una selezione di olivi selvatici seguita da propagazione vegetativa tramite talee sia piantate direttamente sia innestate su portainnesti selvatici (Lumaret et al. 2004). In ogni caso è assodato come l'olivo selvatico abbia contribuito in modo determinante alla variabilità genetica intraspecifica presente nelle cultivar.



Figura 7: Pianta di olivo selvatico

L'olivastro (Figura 7) da millenni è parte integrante della vegetazione mediterranea (Zohary and Hopf, 2000), anche se la presenza di alcune popolazioni isolate è stata documentata in regioni nella fascia nord della penisola iberica (Vargas and Kadereit, 2001).

Loumaret e colleghi (2004) hanno proposto l'ipotesi secondo la quale le popolazioni di olivi selvatici "autentiche", non comprendenti dunque le forme domestiche che sono sfuggite alla domesticazione e hanno revertito il proprio fenotipo, sono oramai ristrette a poche ed isolate aree delle foreste Mediterranee mentre le altre popolazioni con fenotipo tipico dell'olivastro sarebbero appunto da ricondurre a forme "revertite".

In realtà la distinzione tra cultivar e olivastri non è affatto scontata, anzi, basandoci esclusivamente su caratteristiche morfologiche, potrebbe dimostrarsi davvero difficoltosa se non in taluni casi impossibile. Ad esempio la forma dell'albero o della foglia possono essere molto simili, e nel caso specifico della drupa, le cultivar che presentano frutti piccoli (ad esempio Frantoio) possono essere confuse (Besnard & Berville', 2000; Lumaret & Ouazzani, 2001).

L'olivastro si riproduce mediante impollinazione anemofila e i semi prodotti vengono dispersi soprattutto tramite gli uccelli (Herrera, 1995). Generalmente presenta germogli spinescenti e come detto drupe molto piccole (da 0.5 a 1.2 cm), con endocarpo meno carnoso e conseguentemente una resa in olio estremamente più bassa rispetto ai coltivati.

La sua fase giovanile è molto lunga e può perdurare anche diverse decine di anni. A prescindere dalla origine parentale, sia essa selvatica o coltivata, la discendenza nata a partire dal seme tenderà a presentare un fenotipo selvatico (Zohary and Spiegel-Roy, 1975).

Per quanto riguarda la sistematica dell'olivastro, il dibattito sulla collocazione tassonomica e sulla classificazione rispetto all'olivo coltivato è tuttora aperto e controverso. Vengono infatti descritti come due specie differenti: *Olea oleaster* e *Olea europaea* da Link; oppure come due differenti sottospecie della stessa specie, ovvero *O. europaea* L. subsp. *sylvestris* e subsp. *Europaea* prima da Turrill (1951) poi da Morettini (1972).

La versione che sembra ottenere maggiori consensi nella comunità scientifica ad oggi è invece quella di Amaral e Rocha (1972) confermata prima da Pignatti (1982) poi da Green (2002) che vede olivastri e cultivar appartenere alla stessa

specie e stessa sottospecie *Olea europaea* subsp. *europaea* ma a due diverse varietà: *sylvestris* ed *europaea*. Anche se il dibattito non si è chiuso, dobbiamo evidenziare che ci sono studi di genetica anche recenti che cercano di risolvere la controversia sistematica (Angiolillo et al. 1999; Hamman-Khalifa et al., 2007; Erre et al., 2010).

La maggioranza di questi sembra indicare come corretta la teoria delle due varietà (Besnard e Berville 2000, De Caraffa et al. 2002, Rallo et al. 2003, Lumaret et al. 2004, Reale et al. 2006).

Riassumendo, appare chiaro che una caratterizzazione varietale in olivo, sia per quanto riguarda le cultivar, come vedremo nel prossimo capitolo che per quanto riguarda la distinzione tra olivi selvatici e coltivati, non possa prescindere dall'utilizzo di strumenti certi, come i marcatori molecolari.

Capitolo 1.5 I Marcatori Molecolari

I marcatori molecolari attualmente sono basati essenzialmente sulla tecnica PCR, e sono utilizzati nella ricerca genetica con finalità molteplici, quali ad esempio il breeding assistito dalla mappatura dei QTL, gli studi di matrice filogenetica, i lavori di genomica comparativa o per la caratterizzazione molecolare di specie o varietà (Prasad et al., 2000; Stein et al., 2004; Varshney et al., 2005; Crepieux et al., 2005; Diaz et al., 2006).

Per quest'ultima applicazione, prendendo come esempio l'olivo, esiste una grande necessità di identificare le numerosissime cultivar e di chiarire le relazioni che intercorrono tra di esse. E' stata infatti ben documentata la significativa presenza di numerose omonimie e sinonimie all'interno delle varie cultivar (Barranco e Rallo, 2000). Considerato ciò risulta chiaro come una caratterizzazione univoca sia ad oggi oltremodo urgente e necessaria.

E' chiaro come la classificazione e la assegnazione di una accessione alla relativa cultivar, in passato eseguita prevalentemente (o esclusivamente) utilizzando i caratteri morfologici, si presti a letture ed assegnazioni quantomeno ambigue. Questo è dovuto a diversi fattori, tra i quali la soggettività dell'analisi, la generale

incertezza sulle origini delle cultivar, e in modo ancor più stringente a causa della numerosissima presenza di caratteri intermedi. Questo è imputabile sia ad una differenza di genotipo causata dall'eterogeneità genetica delle cultivar ma anche all'ambiente esterno ossia per esempio dalle condizioni climatiche o del suolo.

Per di più una classificazione non ambigua delle cultivar è la base su cui si poggia la certificazione controllata dei marchi di qualità, in particolare per prevenire la tagliatura con oli di provenienza o di qualità differente. Inoltre una ampia e completa analisi del genoma di olivo con marcatori molecolari permetterebbe la salvaguardia del germoplasma e incrementerebbe e potenzierebbe le strategie per la conservazione dello stesso.

Tale analisi altresì consentirebbe la selezione di genotipi con una migliore adattabilità in diversi contesti ambientali (quindi una maggiore e migliore diffusione della coltura) e contribuirebbe a chiarire le origini della domesticazione dell'olivo coltivato e il rapporto genetico (oltre che sistematico) con la varietà selvatica.

La spiccata variabilità del germoplasma delle cultivar è stata evidenziata in alcuni studi utilizzando vari marcatori molecolari come i chloroplast-DNA RFLPs (Amane et al., 1999; Lumaret et al., 2000), o AFLPs (Angiolillo et al., 1999; Baldoni et al., 2006), RAPDs (Bogani et al., 1994; Fabbri et al., 1995; Gemas et al., 2000; Besnard et al., 2001) e microsatelliti (Sefc et al., 2000; Rallo et al., 2000; Cipriani et al., 2002).

Anche le relazioni filogenetiche e le distanze genetiche che intercorrono tra varie cultivar sono oggetto importante di interesse scientifico, testimoniato da diverse pubblicazioni (Angiolillo et al., 1999; Vargas e Kadereit 2001; Belaj et al., 2004; Diaz et al., 2006; Erre et al., 2010).

Attualmente sono a disposizione una moltitudine di marcatori molecolari, alla quale proliferazione abbiamo assistito nell'ultimo decennio.

Ognuno di questi dimostra le proprie caratteristiche, e la propria maggiore o minore rapidità di esecuzione, economicità, capacità discriminativa, riproducibilità.

All'interno della vastissima gamma di marcatori molecolari è opportuno soffermarsi su alcuni di essi.

Gli ISSR ad esempio sono marcatori molecolari estremamente utilizzati. Sono basati su una tecnica che prevede l'amplificazione PCR di un frammento di DNA

presente tra due microsatelliti identici a distanza amplificabile, orientati in direzione opposta. I polimorfismi avvengono e possono essere individuati con questa tecnica nel caso un microsatellite venga alterato o deletato oppure una inserzione modifichi la distanza tra microsatelliti.

Marcatori molecolari ISSR sono stati utilizzati in diverse specie vegetali come riso (Joshi et al., 2000), grano (Nagaoka and Ogihara, 1997), nel genere *Vigna* (Ajibade et al., 2000) con la finalità di stimare la diversità genetica.

Sono stati impiegati anche in olivo: ad esempio Hess e collaboratori hanno utilizzato gli ISSR per uno studio filogenetico geografico su 24 cultivar provenienti dalla Macaronesia. Pasqualone e collaboratori invece hanno applicato i marcatori ISSR a dieci cultivar italiane.

In generale possiamo dire che in diversi lavori gli ISSR si sono dimostrati uno strumento molecolare utilizzabile per l'identificazione varietale in olivo (Terzopoulos et al., 2005; Essadki et al. 2006; Martins Lopes et al., 2007).

I maggiori vantaggi che questa tecnica consente rispetto agli altri marcatori risiede nel non richiedere conoscenze precedenti allo studio per la costruzione dei primer PCR; inoltre non necessita della costruzione di librerie genomiche costose sia in materiali che in tempo (Rakoczy et al. 2004).

Marcatori molecolari in olivo sono stati utilizzati anche per valutare le relazioni e le distanze filogenetiche presenti tra olivo coltivato e selvatico. Come descritto in precedenza questo ha implicazioni genetiche ma anche agronomiche, storiche e tassonomiche e la caratterizzazione morfologica appare sensibilmente ambigua. Fino ad adesso anche una caratterizzazione molecolare semplice e diretta che distingue la varietà coltivata da quella selvatica è indisponibile. L'analisi molecolare potrebbe anche permettere di risolvere la questione ad oggi sempre dibattuta e incerta dell'origine delle cultivar e dei rapporti genetici che a tutt'oggi intercorrono tra le varietà coltivata e selvatica.

In aggiunta una caratterizzazione semplice e non ambigua consentirebbe una immediata riconoscibilità anche delle forme coltivate che sono evase dalla coltivazione e hanno revertito il proprio fenotipo, mostrando caratteri morfologici identici ai selvatici "autentici".

Un lavoro di questo tipo (Hamman-khalifa et al., 2007) è stato pubblicato e sono stati a questo fine utilizzati marcatori ILP. I marcatori ILP sono considerati rapidi

economici e ripetibili ma hanno per contro la necessità di una previa conoscenza della sequenza e dei primer da utilizzare.

Gli ILP (Intron length Polymorphism) sono causati da una o più inserzione o delezione nelle sequenze introniche. Evidenziarli è piuttosto semplice e rapido in quanto è necessaria esclusivamente una reazione di amplificazione PCR chiamata EPIC PCR (Exon-Primed Intron-Crossing PCR) dove i primer vengono costruiti sulle sequenze di “confine” tra introni ed esoni.

I marcatori ILP sono stati sviluppati in poche specie vegetali, come ad esempio ciliegio (Marchese et al., 2010), *Clianthus* (Song et al. 2008) e genziana (Shimada et al. 2009). Sono stati utilizzati anche per la classificazione intraspecifica, con l'acronimo SSILP (Sub-species Specific Intron Length Polymorphism) ad esempio in riso (Zhao et al., 2009).

I marcatori molecolari SNPs (Single Nucleotide Polymorphism) sono stati ampiamente applicati in diversi campi di studio e ad oggi sono estremamente utilizzati anche grazie alla diffusione di metodi di sequenziamento ad alta processività. Ad oggi praticamente tutti i metodi pubblicati per l'individuazione degli SNPs seguono due passaggi obbligati: l'amplificazione (generalmente con tecnica PCR) di una sequenza target seguita da sequenziamento.

Sono numerosi i vantaggi che i marcatori SNPs possono offrire come l'abbondanza nel genoma, la presenza sia in regioni codificanti che non, il fatto che spesso si tratti di marcatori ereditati stabilmente, biallelici e codominanti

Considerando l'uso massiccio che ne viene fatto ad esempio nella genetica umana (Marnellos 2003; Salisbury et al., 2003) risultano essere ancora piuttosto limitati gli utilizzi sia in generale nella genetica vegetale che in particolare in olivo, naturalmente fatte salve alcune eccezioni (Muleo et al., 2009).

I marcatori molecolari SNPs, se vengono sviluppati in zone codificanti hanno un ulteriore e importante significato funzionale.

Sappiamo che gli SNPs assieme agli InDels possono originare cambiamenti rilevanti nella sequenza codificante e successivamente nel caso di mutazioni di “senso errato” o a maggior ragione nel caso di mutazioni “non senso” possono originare in modo diretto cambiamenti rilevanti della sequenza amminoacidica.

Queste modificazioni possono determinare in primis una modificazione della struttura secondaria e terziaria della proteina, e questo potrebbe ripercuotersi in una variazione della sua funzione o addirittura in una sua parziale o totale

inattivazione. Naturalmente tutto questo può originare una modificazione fenotipica della pianta che può essere sia favorevole che non per la fitness dell'individuo e in seconda istanza può rappresentare o meno una modificazione interessante e rilevante dal punto di vista agronomico.

E' chiaro come quest'ultimo aspetto sia direttamente proporzionale alla importanza del gene o dei geni che vengono scelti come candidati allo sviluppo di marcatori SNPs, ad esempio alla loro rilevanza nel metabolismo della pianta. Oppure nella loro influenza nei caratteri di interesse agrario, quindi correlabile a caratteristiche quali l'incremento della resa agronomica, alla difesa dalle avversità climatiche o parassitarie, oppure alla qualità organolettica del prodotto di interesse. Questi marcatori offrono quindi la possibilità di caratterizzare e classificare i vari genotipi e allo stesso momento operare su geni di grande importanza, le quali variazioni potrebbero avere delle implicazioni funzionali.

Per queste considerazioni nel presente studio sono stati selezionati come marcatori molecolari i marcatori SNPs, e come geni candidati, quelli appartenenti ad una famiglia genica, la Calcone Sintasi, che svolge un ruolo chiave all'interno del pathway dei fenilpropanoidi, descritto nel paragrafo successivo.

Capitolo 1.6 Il pathway dei fenilpropanoidi

“In theory we could be creating a world of plants richer in flavonoids.”

(Monique S. J. Simmonds 2003)

Circa quattrocentocinquanta milioni di anni fa le specie di piante pioniere si diffusero dagli ambienti acquatici e andarono gradualmente ad occupare una nuova ed estesa nicchia ecologica: le terre emerse.

Per compiere questa sensazionale evoluzione i geni dovettero adattarsi, modularsi e modificarsi per consentire quegli adattamenti fisiologici, determinati dalla formazione e dall'adattamento di nuove vie metaboliche, che sono indispensabili per la le nuove condizioni della vita emersa (Waters., 2006).

Tra le vie metaboliche che consentono la vita in ambienti non acquatici è di cruciale importanza il pathway dei fenilpropanoidi, praticamente ubiquitario per

le specie vegetali. Inizia da una deaminazione dell'aminoacido fenilalanina per arrivare a decine e decine di composti del metabolismo secondario delle piante. Questi hanno una moltitudine di effetti fisiologici e metabolici tutti fondamentali in ambienti emersi, tra i quali la protezione strutturale della lignina, la protezione dai raggi Ultravioletti dei flavoni e ancora la regolazione e la segnalazione di flavonoidi e isoflavoni.

Dobbiamo però aggiungere che, per quanto riguarda la protezione dai raggi UV determinata dai flavonoidi, è stato ipotizzato da Stafford e colleghi (1991), che probabilmente questa non sia stata la funzione originaria della via biosintetica. Secondo questa teoria infatti per ottenere una effettiva azione protettiva occorre in primo luogo una presenza massiccia di tali composti, oltre ad una opportuna localizzazione, generalmente superficiale.

Altre funzioni proprie del pathway necessitano invece di una minor concentrazione di flavonoidi e in più non necessitano di una regolazione dell'espressione genica tessuto-specifica.

Per Stafford dunque queste si candidano con una maggiore probabilità ad essere la funzione "primitiva" ed originale della via biosintetica.

Questo approccio ci convince solo parzialmente, in quanto sembra non tenere di conto che il percorso che ha portato all'emersione delle prime specie vegetali va letto, a nostro parere, con una lente maggiormente evolucionistica.

L'emersione delle piante è stato un processo graduale che probabilmente è passato da vegetali che inizialmente vivevano in prossimità delle sponde o sulla superficie del mare, (Figura 8) per cui è plausibile che la fitness di alcuni individui sia stata premiata anche con caratteri intermedi e una lenta selezione naturale abbia gradualmente selezionato la maggiore presenza di flavonoidi.

Anche se il dibattito scientifico resta aperto, possiamo affermare con certezza che i fenilpropanoidi sono stati essenziali all'evoluzione delle piante e alla loro emersione (Ferrer et al., 2007).



Figura 8: Ricostruzione dell'ambiente terrestre nel periodo siluriano (435-395 m.a.), nel quale iniziò la colonizzazione delle prime specie vegetali. Fonte: Sito web <http://www.msn.unifi.i/>.

E' stato ipotizzato che la via biosintetica dei fenilpropanoidi si sia originata a partire dagli enzimi del metabolismo primario appartenenti alla via biosintetica degli acidi grassi (Stafford, 1991): in seguito alla pubblicazione della struttura cristallina della CHS di *Medicago Sativa* (Ferrer et al., 1999) si sono evidenziati numerosi elementi che accomunano le strutture degli enzimi appartenenti alle vie biosintetiche degli acidi grassi e dei fenilpropanoidi, quali ad esempio le disposizioni del sito attivo o il ripiegamento caratteristico $\alpha\beta\alpha\beta$ (conservato in tutte le β chetoacilsintasi).

Ci sono dunque forti prove sia strutturali che biochimiche che dimostrano che la CHS, come altre polichetide sintasi (PKSs), derivino dalla β chetoacil ACP sintasi (KASIII) della biosintesi degli acidi grassi. Allo stesso modo si ipotizza

che la CoA ligasi dei fenilpropanoidi derivi dall'acetil CoA ligasi, sempre del metabolismo primario degli acidi grassi (Austin e Noel 2003; Ferrer et al., 2008). Il pathway dei fenilpropanoidi recluta il flusso di carbonio proveniente dal metabolismo primario mediante la fenilalanina.

E' interessante notare come i primi passaggi del pathway (Figura 9) siano altamente conservati anche tra specie vegetali filogeneticamente molto distanti.

Questi step consistono in una deaminazione catalizzata dalla PAL che forma l'acido cinnamico che a sua volta viene idrossilato dalla C4H per formare l'acido para cumarico, il quale gruppo carbossilico viene attivato grazie al legame tioestere con CoA catalizzato dalla p-cumaroil Coa ligasi (4CL) per formare p-cumaroil CoA.

Alcuni funghi e alcuni batteri iniziano la via biosintetica da questo composto, grazie all'enzima Tirosina ammonio ligasi che utilizza come substrato la tirosina e consente il "risparmio" di due enzimi (Hwang et al., 2003).

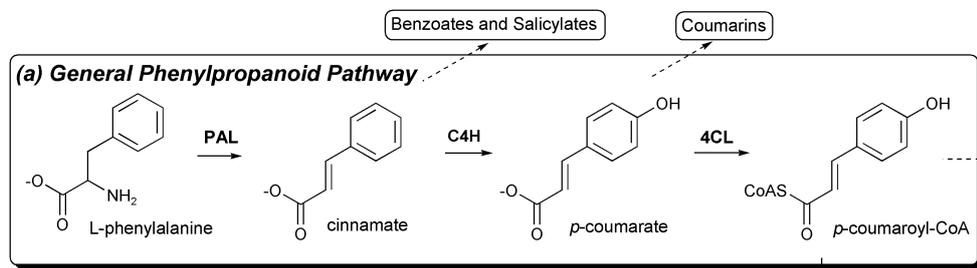


Figura 9: I primi step del pathway dei fenilpropanoidi.

La via biosintetica dei flavonoidi, che è parte integrante del pathway dei fenilpropanoidi, si avvia a partire dal p-cumaroil-CoA con una estensione di polichetidi seguiti da ciclizzazione. Questa azione è esplicata dall'enzima ubiquitario Calcione Sintasi (CHS, classificazione Enzyme Commission EC 2.3.1.74) che rappresenta l'enzima chiave che avvia la biosintesi dei flavonoidi. Questi composti derivano tutti dalla struttura del calcione. La chs catalizza la condensazione e successiva ciclizzazione intra-molecolare di tre unità acetato

derivate dalla condensazione decarbossilativa del malonil CoA sul p-cumaroil CoA per formare la Naringenina- Calcone (Figura 10). La CHS agisce in forma di un omodimero formato da due polipeptidi di 42 Kda (Figura 11).

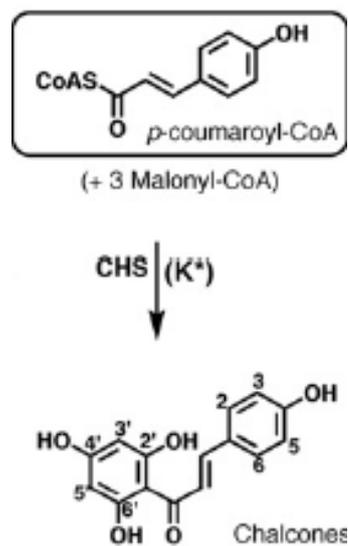
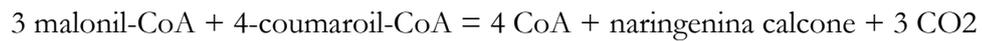


Figura10: L'azione enzimatica della CHS

In seguito a questo passaggio lo scheletro del calcone viene sottoposto ad una serie di modificazioni chimiche catalizzate da enzimi come isomerasi, reductasi, glicosiltransferasi, idrossilasi, aciltransferasi, che diversificano la molecola e la trasformano nei circa 9000 prodotti finali della via biosintetica dei flavonoidi che possiamo suddividere in numerose sottoclassi afferenti a sei grandi gruppi indicati in figura 11: flavoni, isoflavoni, flavonoli, flavandioli, protoantocianidine e antocianine

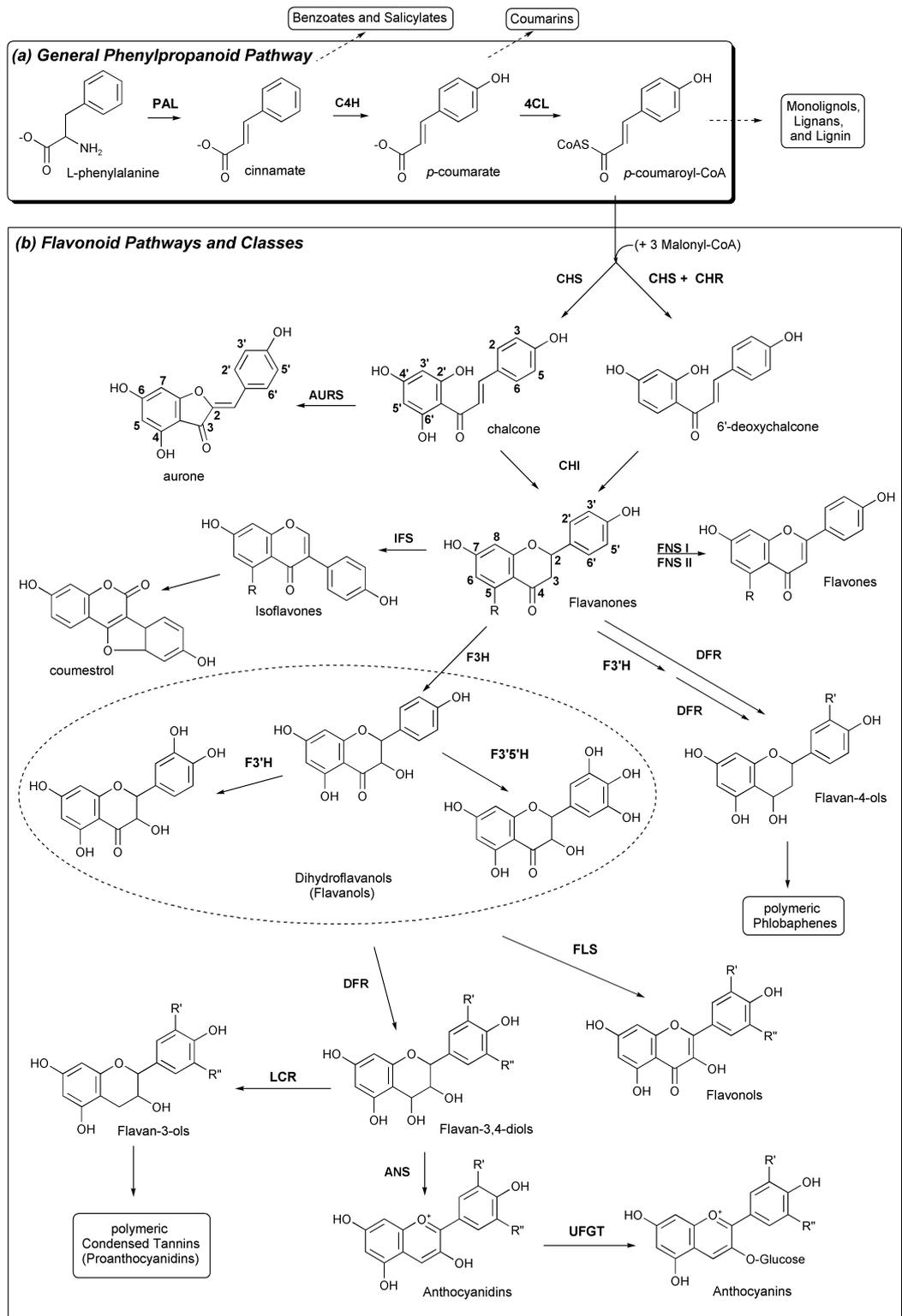


Figura 11: La via biosintetica dei flavonoidi

Ci preme rimarcare il ruolo che questa grande classe di composti occupa all'interno della storia della genetica e più in generale, vista la più che rilevante portata delle scoperte, della intera scienza.

Per citarne solo alcune a termine di esempio, iniziamo dalla scoperta degli elementi trasponibili che valsero il nobel a Barbara Mc Clintock (1983) alla più recente scoperta di vari gruppi di scienziati dell'RNA interference in Petunia.

Le varie classi di composti sintetizzati a partire dalla naringenina calcione svolgono numerose funzioni cruciali per la vita della pianta ma oltre a questo svolgono compiti estremamente importanti anche nell'uomo, chiaramente se vengono assimilati con la dieta.

Numerosi studi (Sankawa 2008, whiting 2001, Jang et al., 1997; Gehm et al., 1997; Zwaagstra et al., 1997; Li et al., 1995; Edwards et al., 1990) dimostrano come alcuni composti fenolici hanno spiccate proprietà anticancerogene, estrogeniche, antimalariche, antiasmatiche e antiossidanti, dunque benefiche per la salute umana.

Tali conoscenze hanno portato sia i breeder che i ricercatori a far si che questi composti fossero presenti in maggiori quantità nelle piante coltivate, sfruttando sia tecniche classiche di miglioramento genetico sia le biotecnologie e la transgenosi (Johnson and Felton, 2001; Galili et al., 2002). Oltre a questo per le medesime ragioni si è assistito ad un incremento della coltivazione di piante officinali e medicinali che generalmente presentano alti livelli di polifenoli e flavonoidi. Nelle ultime decadi l'incremento incontrollato della produzione dei gas serra, l'inquinamento atmosferico, ed in generale la fase di riscaldamento (l'Intergovernmental Panel on Climate Change delle Nazioni Unite indica un potenziale aumento della temperatura, durante il XXI secolo, compreso tra 1,4 °C e 5,8 °C) che il pianeta sta vivendo, correlata ad un avanzamento delle aree siccitose e al progressivo deperimento della fascia di ozono, portano a stress tutti riconducibili alla necessità per la pianta di una iper-espressione dei geni relativi a questa via biosintetica, che come descriveremo in seguito ha un ruolo di fondamentale importanza nella risposta allo stress, nella difesa dai raggi UV e in generale in tutto il comparto difensivo della pianta.

Questa serie di causa effetti spiegano la frase con la quale abbiamo iniziato questo paragrafo: "In theory we could be creating a world of plants richer in flavonoids."

Una frase a nostro parere efficace nel riassumere alcune funzionalità conferite da questa grande famiglia di prodotti del metabolismo secondario delle piante.

Abbiamo brevemente accennato al contributo positivo dei flavonoidi nella salute umana, come principi attivi di piante officinali e farmaci (De Bruyne et al., 1999; Kong et al., 2003; Marles et al., 2003; Yilmaz e Toledo 2004).

Per le piante i loro effetti sono pleiotropici e coinvolgono numerose funzioni essenziali per la sopravvivenza della pianta.

Sappiamo che le piante, non essendo vagili, sono costrette a rispondere alle avversità con la plasticità del loro genoma e la adattabilità del loro metabolismo. Possiamo dire che sotto certi aspetti, concedendoci una metafora un poco grossolana, il metabolismo rappresenta per la pianta ciò che gli arti rappresentano per gli animali. Vale a dire che gli animali ad esempio, essendo in larga maggioranza vagili, possono rispondere alle avversità con il movimento. In caso di avversità climatiche o metereologiche, possono ripararsi o crearsi un rifugio, in casi limite possono arrivare a spostamenti estremi, addirittura da un continente all'altro, per sfuggire ad esse. Per la riproduzione possono risalire correnti e attraversare deserti. Per sfuggire ad un predatore possono correre più velocemente di esso, arrampicarsi più abilmente, o sfidarlo e prevalere con la forza, sempre dovuta al movimento.

Le piante invece essendo organismi sessili, hanno dovuto trasferire e tradurre questa grande molteplicità di reazioni conferita agli animali dal movimento attivo in una sensazionale plasticità e duttilità genomica, che infatti presentano in modo decisamente maggiore rispetto agli animali in generale. Basta pensare per esempio alle mutazioni genetiche, alle delezioni cromosomiche, agli eventi di poliploidizzazione che leggiamo nei genomi vegetali e che spesso per un genoma animale risulterebbero letali.

I flavonoidi sono tasselli fondamentali di queste risposte e in generale delle interazioni che la pianta ha con l'ambiente esterno. Ad esempio per quanto riguarda la foto - protezione dai raggi UV che rappresenta uno dei ruoli chiave dei flavonoidi (Schmelzer et al., 1988; Stafford, 1991; Winkel-Shirley, 2001), la resistenza al freddo (Chalker-Scott e Kramer 1989) ed alle alte temperature (Coberly e Raucher, 2003), alla siccità (Pizzi e Cameron 1986; Moore et al., 2004; Tannini et al., 2004). I flavonoidi consentono alla pianta di sopportare alte

concentrazioni di metalli tossici, come per esempio l'alluminio, nel terreno (Barcelò e Poschenrieder, 2002).

Oltre al rapporto con l'ambiente esterno i flavonoidi assolvono la funzione di molecole segnale verso batteri e simbionti che vivono a stretto contatto con la pianta ospite. Per la segnalazione tra legumi e batteri azoto - fissatori sono impegnati diversi flavonoidi che vengono emessi dalle radici ed inducono la trascrizione di geni batterici le cui proteine sono necessarie per il processo di infezione (Hungria e Stacey 1997; Broughton et al., 2003; Mathesius 2003; Kobayashi et al., 2004; Cooper, 2004).

Sembra che i flavonoidi agiscano in modo positivo nella simbiosi con i funghi micorrizali (Ponce et al., 2004): ad esempio in *Medicago trunculata* le radici colonizzate da *Glomus versiforme* mostrano elevati livelli di espressione sia di PAL che di CHS (Harrison e Dixon 1994).

Ad essere attratti non sono esclusivamente batteri che hanno impatti positivi sulla crescita e la salute: ad esempio due microorganismi patogeni come *Agrobacterium tumefaciens* e *Pseudomonas syringae*, sono attirati, e i loro e i loro geni di virulenza indotti, dai flavonoidi.

Questi ultimi sono implicati anche in processi interattivi pianta-pianta come composti allelochimici (Bais et al., 2003).

Un altro meccanismo complesso che li vede protagonisti è la difesa della pianta: possiamo suddividere i flavonoidi utilizzati a questo scopo tra preformati e indotti. I primi sono espressi costitutivamente e stoccati in compartimenti strategici per poi essere usati sia per la fotoprotezione che in caso di attacco da parte di insetti erbivori (Feeny 1976). Sono infatti numerosi gli insetti sensibili per i quali è stata dimostrata una repulsione (Brignolas et al., 1998; Berhow e Vaughn 1999; Hoffmann-Campo et al., 2001; Widstrom e Snook 2001; Haribal e Feeny 2003; Thoison et al., 2004; Chen et al., 2004).

I geni codificanti i flavonoidi “indotti” si attivano invece in modo selettivo in seguito all'attacco di patogeni o insetti nocivi

I flavonoidi rappresentano i pigmenti da fiore più numerosi (Gleiss et al., 2001; Koes et al., 2005) e agiscono come agenti attrattori per gli insetti e in generale per gli agenti impollinatori.

Capitolo 1.7 La famiglia genica della Calcone Sintasi.

L'enzima CHS in molte specie di piante è codificato da una piccola famiglia multigenica (Durbin et al., 1995; Koes et al., 1999).

Per quanto riguarda la provenienza evolutiva di questa famiglia genica è stato ipotizzato che si sia originata a partire dagli enzimi del metabolismo primario appartenenti alla via biosintetica degli acidi grassi. A loro volta altre polichetide sintasi delle piante, come ad esempio la stilbene sintasi (STS), l'acridone sintasi (ACS), la floroisovalerofenone sintasi (PVPS) sembra che si siano evolute a partire proprio dalla calcone sintasi (Tropf et al., 1994; Lukacin et al., 1999; Paniego et al., 1999).

A rimarcare una plausibile origine condivisa, tutti questi enzimi mostrano similitudini sia per quanto riguarda la funzione che per quanto riguarda la regolazione. In particolare tutti catalizzano reazioni di condensazione tra molecole "starter CoA" e malonil-CoA e mostrano una identità di sequenza con la Calcone Sintasi che si aggira approssimativamente sul 60%.

Gli enzimi in questione sono quindi stati classificati con la denominazione comune: enzimi CHS-like, che assieme alla CHS fanno parte di una superfamiglia denominata superfamiglia CHS (CHS superfamily) (Schroder, 1997).

Abbiamo accennato brevemente ai legami che intercorrono tra i diversi enzimi appartenenti a vie biosintetiche anche molto differenti, e abbiamo visto come ci siano dei tratti vistosamente comuni all'interno di superfamiglie e famiglie geniche. Ci chiediamo quindi quali siano le dinamiche genetiche che hanno portato alla differenziazione e all'evoluzione dei geni che codificano per questi importanti enzimi.

Sicuramente eventi molecolari come la duplicazione genica e di riflesso la ridondanza genetica hanno rappresentato (e rappresentano) un autentico motore dell'evoluzione. Hanno consentito l'aumento della complessità delle funzioni geniche e la loro modulazione, spingendo gli organismi alla conquista di nuovi habitat e permettendo loro la possibilità di adattamento (Hooper e Berg 2003; Moore e Purugganan, 2003; Blanc e Wolfe, 2004).

Questo in generale per tutte le specie, ma in particolar modo nelle piante: la duplicazione genica infatti costituisce un aspetto cruciale e di fondamentale

importanza quando parliamo di genomi vegetali, in quanto essi hanno una maggiore tolleranza di questi eventi molecolari rispetto ai genomi animali.

La duplicazione può derivare sia da un evento singolo di duplicazione che da un evento di poliploidizzazione. La pressione selettiva che si sviluppa su entrambi i geni può portare al silenziamento di un paralogo dovuto a mutazioni degenerative che portano alla perdita di funzione, oppure può esserci un effetto protettivo (assunto che il sovradosaggio non sia critico) rispetto alla selezione naturale, di un paralogo rispetto all'altro, che può quindi “permettersi” di accumulare mutazioni. Può avvenire anche una riduzione nell'espressione di entrambi i paraloghi che porta ad una complementarità degli stessi o ad una differenziazione nell'espressione che può essere tessuto specifica o specifica per alcuni stadi di sviluppo della pianta. Vediamo in seguito alcuni esempi che ci dimostrano quanto illustrato.

Prendiamo *Petunia hybrida*, nella quale sono state isolate 8 copie complete di CHS (oltre a 4 parziali). Tra queste in particolare CHSB mostra una maggiore differenza di sequenza rispetto alle altre CHS che può essere attribuita ad una duplicazione antica seguita da una rapida evoluzione.

CHSC, CHSG e CHSL invece mostrano una elevata omologia di sequenza sia nelle zone codificanti che nella 5'UTR e nelle sequenze introniche.

Possiamo ipotizzare che si siano evolute in seguito ad un recente evento di duplicazione.

CHSA e CHSJ sono espresse in particolar modo nella corolla e nel seme ma solo in seguito ad irradiazione con raggi UV. Anche CHSB e CHSG sono espressi in seme come conseguenza di irradiazione con UV, ma non si esprimono nel tessuto florale.

Per questi paraloghi è stato dimostrato che la similarità di sequenza non è correlata all'espressione. CHSD invece ha subito una mutazione che ne ha determinato la perdita di attività. Infatti il codone di stop ha subito una mutazione che ha determinato l'estensione della regione codificante del gene, che probabilmente è responsabile dell'inattivazione (Koes et al., 1989; Yang et al., 2006).

Un altro caso studio interessante da discutere è relativo a *Ipomoea violacea*. I geni CHS in questo caso sono 6, CHSA CHSB CHSC CHSD CHSE oltre ad uno pseudogene denominato CS. Le sequenze di questi geni sarebbero da assegnare a

due differenti sub-famiglie, una comprendente CHSA CHSB CHSC e PS, l'altra con CHSD e CHSE. Utilizzando la tecnica dell'orologio molecolare è possibile datare la divergenza dei due sottogruppi a circa 100 milioni di anni fa. E' stato osservato come solo CHSD e CHSE catalizzano la reazione caratteristica della CHS e sintetizzano la naringenina-calcone, mentre le altre sintetizzano la bis-noriangonina.

E' stato stimato da Durban e collaboratori che CHSA CHSB e CHSC sono evolute ad una velocità 2,7 volte maggiore (ai siti non sinonimi). E' stata suggerita una selezione positiva di tipo darwiniano (Yang et al., 2004).

Per quanto riguarda l'espressione, CHSD risulta essere il più espresso nell'apparato fiorale e il responsabile dell'accumulo di pigmento mentre CHSA, CHSB, CHSC e CHSE sono maggiormente espressi nelle parti non pigmentate, con espressioni differenziate in funzione dello stadio di sviluppo della pianta. Questi elementi fanno ipotizzare una sostanziale differenza nella regolazione dei paraloghi.

E' opportuno sottolineare a livello del promotore e nell'introne di CHSD la presenza di sequenze riconducibili a elementi mobili MELs e ripetizioni dirette Drs (Hoscino et al., 2002) mentre in CHSE è stata descritta esclusivamente la presenza di un elemento mobile MELs6-4 nella regione fiancheggiatrice a valle della sequenza codificante (Durbin et al., 2001). Gli altri paraloghi identificati non presentano elementi mobili nella sequenza isolata.

Questo caso ci consente di citare brevemente una altra componente fondamentale nelle dinamiche evolutive ovvero gli elementi mobili, sequenze che hanno avuto un ruolo determinante nella diversificazione dei paraloghi sia dal punto di vista della sequenza genica che nell'espressione del prodotto.

Andando avanti con gli esempi di differenziazione dei paraloghi, nel lavoro pubblicato da Ito e collaboratori emerge con chiarezza la differenziazione funzionale dei paraloghi che nel caso specifico di *Pisum sativum* La specie in questione conta almeno 8 paraloghi CHS, numerati da CHS1 a CHS8, che possono essere suddivisi in tre gruppi in base all'omologia della sequenza isolata, andando a confrontare sequenze che comprendono il gene e il suo promotore.

CHS6 e CHS7 (gruppo 3) mostrano una minore (< 80%) similarità rispetto al gruppo 1 (CHS1-CHS2-CHS8) e al gruppo 2 che tra loro invece hanno una

similarità maggiore dell'80%. Questo riflette in modo significativo le differenze nell'espressione dei tre gruppi di paraloghi.

Scendendo nel dettaglio è possibile evidenziare come il gruppo 1 e il gruppo 2 siano fortemente indotti dall'irradiamento con raggi UV: il gruppo 1 è però espresso in modo costitutivo nelle radici e nel fiore mentre il gruppo 2 solo nelle radici. Il gruppo 3 è invece caratterizzato da una espressione bassa dal punto di vista quantitativo e non sembra essere influenzato dai raggi UV.

L'isolamento dei rispettivi promotori ha consentito di individuare alcune sequenze cis-agenti, assenti in CHS6 e CHS7 (Ito et al 1997). In *Phaseolus vulgaris* si osserva una situazione analoga che vede come le sequenze altamente omologhe dei paraloghi che vanno dal 96% al 98% di omologia sono espresse in modo differente.

Ad esempio CHS1 viene indotto dalla luce a differenza di CHS4, CHS14 e CHS17 che si attivano in seguito a stress. Tra questi tre geni esiste una ulteriore differenziazione in quanto CHS4 e CHS14 si attivano fortemente mentre CHS17 mostra livelli piuttosto bassi di trascrizione.

E' stato proposto che la modulazione differenziale di CHS in fagiolo sia in parte da attribuire alla combinazione di due elementi presenti in cis nei promotori di tali geni, H-Box e G-Box.

Ci sono altri esempi meritevoli di citazione che fanno emergere una diversificazione dei prodotti dei diversi paraloghi, una differenziazione tessuto-specifica come in *Viola cornuta* dove sono stati individuati 3 diversi geni CHS che si esprimono in modo differenziale a partire dagli stadi iniziali di colorazione del fiore (Farzad et al., 2003) oppure una differenziazione sia tessuto specifica che legata ai differenti stadi di sviluppo della pianta come dimostrato da Yi e collaboratori (2010) per CHS7 e CHS8 di soia.

Dallo studio di questi lavori sembra emergere che la differenziazione funzionale dei paraloghi sia spesso correlata a modificazioni che sono presenti nelle sequenze regolative e non nelle codificanti.

Riassumendo, abbiamo esaminato alcuni casi in cui si dimostra che la duplicazione dei geni può portare ad una divergenza e a un differenziamento dei due paraloghi. A differenza dei geni ortologi la cui divergenza è dovuta alla differente storia evolutiva delle specie ed alla variabilità genetica, che nelle piante

è strettamente connessa alla modalità di riproduzione, i paraloghi hanno gioco forza la medesima storia evolutiva. Tuttavia la selezione naturale generalmente spinge verso una differenziazione dei paraloghi, dimostrata dalla variabilità riscontrata all'interno delle famiglie geniche (si veda ad esempio hm1 e hm2 in mais: Zhang et al., 2002). Tale diversificazione può portare ad una diversificazione tra i paraloghi dal punto di vista della funzione come della regolazione (Ohno, 1970; Linch, 2002).

Come abbiamo visto in *Petunia hybrida*, *Ipomoea violacea*, *Pisum sativum*, *Phaseolus vulgaris*, *Viola cornuta* e *Glycine max*, la famiglia genica della calcione sintasi è composta da un numero altamente variabile di paraloghi.

Altri esempi per supportare questo dato ce li possono offrire i 9 geni chs che compongono la famiglia genica in trifoglio (Howles et al., 1995), i 2 geni CHS che sono stati isolati in *Populus* (Tuskan et al., 2006), i 3 paraloghi che sono stati isolati in *Vitis vinifera* (Sparvoli et al., 1994; Jaillon et al., 2007).

In *Antirrhinum majus* invece è stata isolata una singola copia del gene CHS (Sommer e Sadle, 1986).

Ad oggi quindi, molti geni della famiglia CHS sono stati isolati e caratterizzati in numerose specie vegetali, e abbiamo visto che esiste una notevole differenziazione per quanto riguarda il numero di paraloghi della famiglia genica CHS presenti nelle specie. Vediamo adesso le analogie e divergenze presenti nella struttura dei geni CHS.

La struttura del gene risulta essere altamente conservata all'interno dei vari taxa e consiste in due esoni separati da un introne. L'unica eccezione strutturale ad oggi nota in letteratura riguarda *Antirrhinum majus*, che presenta una struttura a tre esoni e due introni.

La lunghezza dell'introne varia in modo significativo nelle differenti specie, e può spaziare dalle decine di paia di basi a numerose kilobasi. La sua posizione però è altamente conservata ed è situata tra la prima e la seconda base codificanti il residuo Cisteina.

Per quanto concerne gli esoni, generalmente il primo risulta essere più corto e codifica per 37-64 residui. Il secondo esone, di lunghezza meno variabile, codifica per circa 340 residui, tra i quali vi sono tutti i residui aminoacidici del sito attivo (Fliegmann et al., 1992; Jinxia et al., 2002).

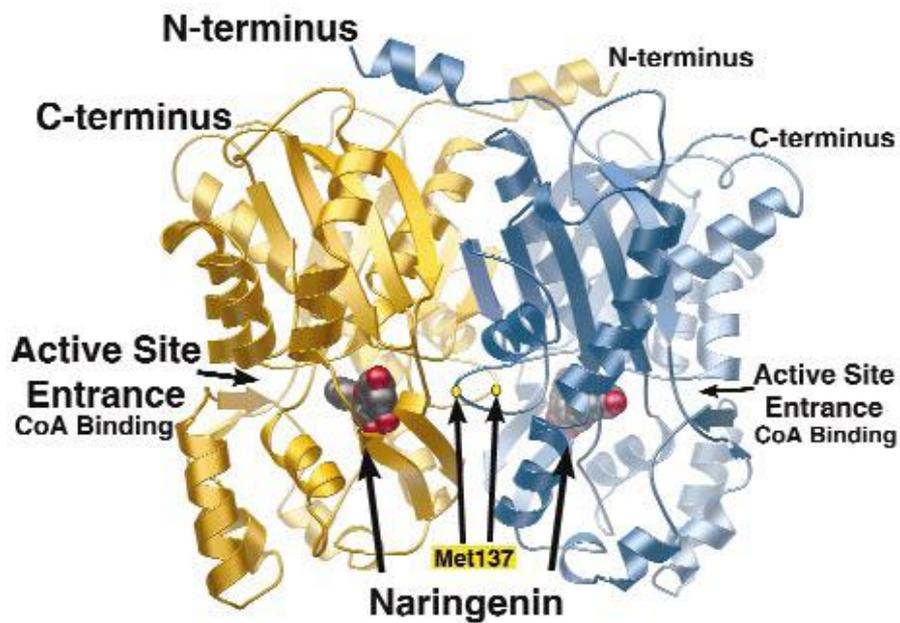


Figura 12. La proteina CHS

Tra le numerose CHS isolate è possibile mettere in evidenza residui aminoacidici altamente conservati, che sono stati raggruppati in quattro gruppi in base alla funzione svolta (Yasuyo et al., 2005):

- 1 residui del sito attivo;
- 2 residui catalitici;
- 3 residui del sito di aggancio del Coenzima A;
- 4 altri residui conservati.

In seguito alla determinazione della struttura cristallina della CHS 2 di *Medicago sativa* da parte di Ferrer e collaboratori è possibile collegare i residui conservati con la loro funzione.

I quattro residui Cys164, Phe215, His303 e Asn 336, conservati in tutte le CHS isolate fino ad oggi definiscono il sito attivo.

I cinque residui Ser133, Glu192, Thr194, Thr197 e Ser 338 formano il sito di aggancio del Coumaril-CoA. I sette residui che formano la tasca di ciclizzazione

sono Thr132, Met137, Phe215, Ile254, Gly256, Phe265 e Pro375. Oltre a questi, altri residui risultano essere localizzati in posizione tale da influenzare la geometria del sito attivo ovvero Pro138, Gly167, Leu214, Asp217, Gly262, Pro304, Gly 305, Gly306, Gly 374, Pro375 e Gly376 e risultano anch'essi altamente conservati (Ferrer et al., 1999) (Figura 12).

Capitolo 2

Scopo del lavoro

Gli obiettivi principali del lavoro si articolano seguendo due tematiche principali: la caratterizzazione varietale delle cultivar di olivo e lo studio della variabilità presente nella specie. La caratterizzazione delle cultivar, nello studio e nella coltivazione dell'olivo riveste una importanza centrale da vari punti di vista. Sappiamo come ad oggi questa venga effettuata principalmente seguendo la caratterizzazione morfologica dei caratteri fenotipici della pianta e delle drupe.

In aggiunta a questa metodica non priva di problematiche e criticità, esposte diffusamente nell'introduzione, si stanno predisponendo una serie di tecniche alternative, in primis i marcatori genetici molecolari, che riescano a rendere la caratterizzazione varietale scempia da ambiguità.

Il nostro lavoro si orienta in questa direzione, assegnando una priorità alla caratterizzazione molecolare di cultivar rilevanti per la produzione di oli di qualità, elencate e disciplinate nelle denominazioni di origine protetta, consci che la valorizzazione dei prodotti del nostro territorio e la loro collocazione strategica nel mercato internazionale debba necessariamente focalizzarsi sulla qualità che le nostre tecniche agronomiche tradizionali e i nostri germoplasmi autoctoni possono esprimere.

Per questa motivazione tra le numerosissime cultivar che il germoplasma olivicolo italiano ci offre abbiamo preferito indirizzare il nostro studio su varietà tipiche del territorio nel quale operiamo ovvero la Toscana.

Per la caratterizzazione varietale abbiamo scelto di utilizzare i marcatori molecolari SNPs in una famiglia genica che riveste una importanza centrale nel metabolismo secondario della pianta.

Questo ci consente di perseguire l'obiettivo della tipizzazione varietale con strumenti di biologia molecolare e allo stesso tempo di caratterizzare dal punto di vista genetico questa importante famiglia genica, ampiamente studiata in altre specie vegetali ma sulla quale a nostra conoscenza non sono presenti pubblicazioni in letteratura per quanto concerne *Olea europaea*.

Il nostro lavoro è mirato alla caratterizzazione genetica della famiglia genica della Calcione Sintasi in olivo e comprende l'isolamento delle sequenze complete dei geni che la compongono: inizialmente in una cultivar di riferimento e successivamente in tutte le cultivar selezionate.

Questa finalità viene perseguita utilizzando varie applicazioni della tecnica PCR, il clonaggio e il sequenziamento delle sequenze amplificate, la loro analisi e caratterizzazione strutturale *in silico* con metodiche di bioinformatica e lo studio del numero dei componenti mediante tecnica di ibridazione Southern blot.

Il nostro obiettivo di fondo però non si esaurisce allo studio di cultivar toscane ma il nostro lavoro vuole estendersi ad altre cultivar importanti, caratteristiche ed autoctone di altre realtà produttive di primissima linea nell'olivicultura italiana come la Calabria, la Puglia, la Campania, la Sicilia.

Intendiamo dunque perseguire il secondo obiettivo centrale del nostro studio ovvero la analisi della variabilità genetica presente all'interno della specie *Olea europaea*, sia tra cultivar appartenenti allo stesso areale di produzione (la Toscana) che estendendo la analisi a cultivar coltivate nelle regioni citate in precedenza.

Inoltre intendiamo comprendere il grado di variabilità genetica presente all'interno di una cultivar, nel nostro caso la Cultivar calabrese Carolea, per dare un contributo ad una altra tematica ampiamente dibattuta in letteratura.

In aggiunta intendiamo studiare i rapporti genetici che intercorrono tra l'olivo coltivato e selvatico e possibilmente individuare un marcatore molecolare semplice ed economico, attualmente non presente in letteratura, che sia in grado di diversificare le due varietà.

Infine intendiamo utilizzare anche un marcatore molecolare con uno spettro di azione più ampio, quali i marcatori molecolari ISSR, che ci permettano di analizzare la variabilità genetica presente in due areali prossimi tra loro ovvero la Sicilia e la Calabria.

Capitolo 3

Materiali e Metodi

Capitolo 3.1 Estrazione di DNA genomico da tessuto vegetale

Per l'estrazione del DNA genomico dalle foglie di olivo è stato applicato il metodo Doyle & Doyle.

Il materiale vegetale di partenza (100-150 mg), viene congelato in azoto liquido, quindi ridotto in polvere finissima, ed aliquotato in campioni di 2 g circa ciascuno. Ad ogni aliquota vengono poi aggiunti 8 ml di soluzione d'estrazione, contenente CTAB, un detergente in grado di complessarsi con il DNA e di proteggerlo durante l'estrazione, mentre polisaccaridi ed altri componenti cellulari vengono degradati.

I campioni vengono quindi lasciati in incubazione per 30 min a 60°C, e successivi 10 min a temperatura ambiente.

Vengono poi aggiunti 5 ml di una soluzione cloroformio:isoamilalcol (24:1) e la miscela viene centrifugata per 5 min a 3000 rpm; l'aggiunta della soluzione permette la formazione di due fasi: nella fase acquosa si distribuiscono DNA e RNA, in quella organica, formata da cloroformio ed isoamilalcol, le molecole idrofobiche, ed all'interfaccia quelle anfipatiche, come le proteine di membrana. Si recupera quindi la fase acquosa, alla quale vengono poi aggiunti 7 ml di isopropanolo, che provoca la precipitazione del DNA sotto forma di flocculo lattescente.

Questo viene prelevato tramite una pipetta Pasteur uncinata e trasferito in 2 ml di soluzione contenente etanolo al 76% e NaAc 0.2 M, dove è lasciato per almeno 20 min.

Un ulteriore lavaggio del DNA isolato è compiuto trasferendo nuovamente il precipitato in 2 ml di soluzione contenente etanolo al 76% e NH₄Ac 10 mM. La soluzione viene poi eliminata ed il precipitato è lasciato ad asciugare, dopo di che viene risospeso in 200 µl di soluzione Tris-HCl 100 mM (pH=8).

Soluzione d'estrazione

Tris-HCl pH=7.5	1 M
NaCl	5 M
EDTA	0.5 M
CTAB	2% w/v
β -mercaptoetanolo	0.2%

Qualunque metodo di estrazione sia stato applicato, l'integrità del DNA genomico ottenuto è verificata tramite elettroforesi su gel d'agarosio.

Le soluzioni di DNA vengono conservate a -20°C.

Capitolo 3.2 Purificazione del DNA

Per eliminare l'RNA residuo nel campione estratto, questo viene trattato con 0.1 mg/ml di RNAasi (Boehringer-Mannheim) per un'ora a 37°C. L'enzima e l'eventuale RNA degradato vengono poi eliminati precipitando il DNA. Si aggiunge 1 volume di una miscela di fenolo:cloroformio:isoamilalcol (50:49:1), e la soluzione ottenuta viene centrifugata per 5 min a 8000 rpm. Si ottengono così due fasi separate, la superiore delle quali (di natura acquosa), contenente il DNA, viene prelevata. Alla soluzione ottenuta viene quindi aggiunto 1/10 di volume di una soluzione di NaAc 3M (pH=5.2) e 2.5 volumi di etanolo assoluto. La precipitazione viene lasciata procedere per tutta la notte. Infine il campione viene centrifugato per 30 min a 10000 rpm ad una temperatura di 4°C, eliminato il sovrantante, aggiunto etanolo al 70%, e di nuovo centrifugato a 10000 rpm, alla temperatura di 4°C, per 20 minuti.

Il DNA precipitato ottenuto è quindi lasciato ad asciugare, ed infine risospeso in 200 μ l di soluzione Tris-HCl 100 mM (pH=8).

Capitolo 3.3 Quantizzazione tramite biofotometro

La stima della concentrazione delle soluzioni di acidi nucleici viene effettuata con il metodo spettrofotometrico. In particolare, i campioni oggetto del nostro studio sono stati quantizzati nella loro concentrazione in acidi nucleici tramite il *Biophotometer Eppendorf*.

Tale macchina, come tutti i metodi spettrofotometrici, sfrutta la caratteristica dello spettro d'assorbimento degli acidi nucleici di avere un picco a 260 nm di lunghezza d'onda. La legge di Lambert-Beer ($A = \epsilon \cdot c \cdot l$; dove A è l'assorbanza, ϵ il coefficiente d'estinzione molare, c la concentrazione molare, e l lo spessore della cuvetta) permette la conversione dei livelli di assorbanza a 260 nm nella concentrazione del campione analizzato.

Il biofotometro permette la misurazione dell'assorbanza a tre diverse lunghezze d'onda: 230 nm, 260 nm, 280 nm. Si tratta delle lunghezze d'onda dei picchi d'assorbimento rispettivamente di carboidrati, acidi nucleici, proteine (in particolare degli aminoacidi aromatici). Anche la lunghezza d'onda a 320 nm può essere settata, come indice di contaminazioni da particolati in soluzione.

Vengono calcolati poi i rapporti di assorbanze a 260 nm/280 nm e 260 nm/230 nm, indicativi del grado di purezza del campione.

Si considerano campioni con un buon grado di purezza quelli che hanno il rapporto A_{260}/A_{280} compreso tra 1.7 e 2, e il rapporto A_{260}/A_{230} superiore a 2.

Le cuvette usate sono cuvette monouso. Il biofotometro viene prima azzerato con una cuvetta contenente 50 μ l di acqua distillata sterilizzata in autoclave (lettura del "bianco").

L'acqua distillata sterile è il solvente utilizzato nelle misurazioni dei campioni, che vengono diluiti 50 volte nella cuvetta (il volume nella cuvetta è sempre 50 μ l).

La differenza nel calcolo della concentrazione tra campioni di DNA e di RNA sta nel valore del coefficiente d'estinzione molare: 50 per il DNA, 40 per l'RNA.

Capitolo 3.4 Quantizzazione tramite fluorimetro

In alcuni casi la quantificazione con il biofotometro è stata confermata dalla quantizzazione utilizzando il fluorimetro. Il suo funzionamento è il seguente: una lampada emette ad una determinata lunghezza d'onda e la luce va ad eccitare una sostanza fluorescente che si lega in modo proporzionale al DNA. Gli elettroni del composto fluorescente si trovano ad uno stato eccitato e per tornare al livello fondamentale, emettono luce ad una certa lunghezza d'onda. L'intensità del raggio emesso indica la concentrazione del campione: maggiore è l'intensità più materiale fluorescente si è legato, più DNA è presente nel campione.

Anche in questo caso per risalire alla concentrazione viene usata una formula, che deriva dalla legge di Lambert- Beer:

$$I = I_0 e^{-\epsilon \lambda c d}$$

Per fare le letture al fluorimetro abbiamo usato un kit, Quant –It dsDNA RB Assay Kit, composto da un buffer, due standard e un colorante che si intercala al DNA.

Capitolo 3.5 Quantizzazione di DNA su gel di agarosio

La quantificazione di DNA su gel viene fatta con l'aiuto di un marcatore di peso molecolare, le cui bande siano di quantità nota, in questo modo il marcatore oltre ad essere un indicatore di peso molecolare, può essere utilizzato per valutare la concentrazione del campione. Il marcatore durante la corsa elettroforetica si separa su gel di agarosio nei singoli frammenti da cui è composto. Ciascun frammento identifica una determinata concentrazione, espressa in ng/ μ l. Nei

nostri esperimenti abbiamo utilizzato il ladder 1KB o il 100bpPLUS (Fermentas). Abbiamo utilizzato 1 μ l delle soluzioni stock diluendole in 8 μ l di acqua di acqua distillata e 1 μ l della soluzione stock del colorante blu di bromofenolo (BFB). Così facendo abbiamo potuto confrontare le nostre bande con quelle del ladder di riferimento..

Capitolo 3.6 Elettroforesi su gel di agarosio

L'elettroforesi su gel di agarosio permette di separare frammenti di DNA a diverso peso molecolare a scopi sia analitici che preparativi. Questa metodica necessita di un supporto solido, ovvero di un gel di agarosio utilizzato a concentrazioni diverse a seconda della necessità, che funzioni da "setaccio" per le molecole che lo attraversano, e la separazione sarà tanto migliore quanto più strette saranno le maglie che lo formano. La separazione elettroforetica si basa sul principio secondo cui le molecole di acido nucleico poste in una soluzione a pH 8.0, come il tampone per l'elettroforesi, possiedono una carica negativa e pertanto, se sottoposte ad un campo elettrico, migrano verso l'anodo. Le molecole di acido nucleico migrano attraverso la matrice del gel in modo inversamente proporzionale al logaritmo in base dieci del loro peso molecolare: molecole più grandi migrano più lentamente a causa delle maggiori forze di attrito e attraversano le maglie del gel con maggiore difficoltà.

A bassi voltaggi esiste una relazione lineare tra il logaritmo della mobilità elettroforetica e la concentrazione del gel, mentre ad alti voltaggi la capacità di separazione di un gel diminuisce.

L'agarosio (Boehringer-Mannheim), solido a temperatura ambiente, è un polimero lineare estratto dalle alghe. Il gel viene preparato solubilizzando, in tampone TAE, dell'agarosio (Tris-Acetato-EDTA) in quantità diverse per avere concentrazioni finali che possono variare da 0.8% a 3% . La solubilizzazione avviene a caldo, dopodiché la soluzione viene fatta raffreddare in agitazione.

Prima che la soluzione venga versata sul supporto per la corsa elettroforetica viene aggiunto bromuro di etidio in concentrazione 1 µg/ml. Tale sostanza ha la capacità di intercalarsi nella doppia elica del DNA e di emettere fluorescenza se colpita da luce ultravioletta: quando presente all'interno del gel elettroforetico, il bromuro di etidio si lega ai frammenti di DNA migranti permettendone l'osservazione sotto lampada UV.

La soluzione viene quindi versata su di un apposito supporto munito di un pettine che permette di formare dei pozzetti nei quali verrà posto il DNA all'inizio della corsa.

Una volta raffreddato, il gel viene trasferito all'interno di una vasca elettroforetica, nella quale è immerso in tampone TAE; la vasca elettroforetica è munita, alle due estremità, di un catodo e di un anodo. Le soluzioni di DNA da analizzare vengono immerse all'interno dei pozzetti dopo l'aggiunta di una soluzione colorante, il blu di bromo fenolo (BFB), che permette di seguire la corsa elettroforetica e rende i campioni più densi, riducendo il rischio di fuoriuscita dai pozzetti.

In un pozzetto viene caricato il marcatore di peso molecolare, una miscela di frammenti di DNA di dimensioni note che permettono la stima per comparazione della lunghezza dei frammenti dei campioni migranti.

Soluzioni utilizzate:

Tampone TAE 50x:

Tris-HCl 2M

EDTA 100mM

La soluzione viene portata a pH 8.0 con acido acetico glaciale e, prima dell'uso, viene filtrata con apposita carta da filtro sottovuoto. Si conserva a temperatura ambiente.

Gel loading buffer:

Questa soluzione viene aggiunta ai campioni da separare in modo da visualizzarne la corsa sul gel. Il BBF in essa contenuto ha mobilità anodica come il DNA.

Ogni 10 μ l di campione da analizzare, vengono aggiunti 2 μ l di una soluzione stock 6X composta da 0.25% di BBF e 40% w/v di glucosio in acqua distillata.

Soluzione di bromuro di etidio:

Il bromuro di etidio (Sigma) viene solubilizzato in acqua distillata sterile alla concentrazione di 10 mg/ml, filtrato e conservato a 4°C.

Capitolo 3.7 PCR

La PCR, o reazione a catena della polimerasi (Mullis et al., 1986), è una metodica che permette l'amplificazione esponenziale del numero di specifici segmenti di DNA. La reazione si basa su tre fasi ripetute ciclicamente nelle quali il DNA viene prima denaturato ad alta temperatura (fase di denaturazione, normalmente condotta a 95° C), poi due oligonucleotidi, specifici per le regioni laterali del segmento da amplificare e usati come inneschi per la reazione di polimerizzazione, vengono fatti appaiare alle rispettive regioni complementari tramite abbassamento della temperatura (fase di *annealing*). Infine la temperatura viene nuovamente alzata fino a 72° C, temperatura che permette la catalisi della reazione di polimerizzazione da parte di una DNA polimerasi termostabile proveniente dal microrganismo *Thermus aquaticus* (Taq-DNA polimerasi) (Innis et al.1988).

Le reazioni di amplificazione PCR sono state condotte in un volume di 25 μ l o di 100 μ l, nel caso fosse necessaria una maggiore quantità di prodotto, ad esempio per sequenziamento diretto del prodotto PCR.

Per le reazioni condotte in 25 µl utilizzando 100 ng di DNA genomico come stampo, 5 pmol di ogni primer, 5 µmol di deossinucleotidi trifosfati, 1U di Taq polimerasi (Pharmacia) e il corrispondente buffer 10X.

Per le reazioni condotte in 100 µl utilizzando 400 ng di DNA genomico come stampo, 20 pmol di ogni primer, 20 µmol di deossinucleotidi trifosfati, 4U di Taq polimerasi (Pharmacia) e il corrispondente buffer 10X.

L'amplificazione è stata condotta in un termocicizzatore MJ PTC-200 seguendo un programma consistente in:

- una fase di denaturazione iniziale di 10 minuti a 95° C,
- 30 cicli di amplificazione consistenti in

denaturazione a 95° C per 45 secondi,

annealing per 45 secondi,

extension a 72° C,

- una fase finale di estensione a 72° C per 8 minuti.

La temperatura di *annealing* dei *primers* è stabilita in base alla loro composizione in nucleotidi; La formula

$$T_m = 4x(G+C) + 2x(A+T)$$

permette il calcolo della temperatura di *melting* (temperatura alla quale il primer appaiato alla regione complementare si distacca da esso). La temperatura di *annealing* invece può essere stimata solo sperimentalmente, essendo normalmente 5-10° C più bassa di quella di *melting*.

Per valutare i risultati delle reazioni di amplificazione PCR, in termini di presenza dei frammenti di dimensioni attese all'interno del campione, i prodotti vengono fatti correre elettroforeticamente su gel di agarosio. Le dimensioni dei frammenti amplificati, osservabili sul gel, vengono valutate tramite confronto con marcatori di peso molecolare.

In figura 13 sono indicate le condizioni di reazione delle principali PCR eseguite. In figura 14 sono indicati i primer utilizzati.

Coppia di primer	Utilizzo	amplificato	T annealing	Tempo extension
5'CHSAFW/CHSCODA1rev	Amplificazione sequenza completa <i>chsA</i>	2200 bp	56°C	2 min
5'CHSBFW/CHSCODA3rev	Amplificazione sequenza completa <i>chsB</i>	2140	57°C	2 min
CHSEX1F/CodaRev1	Prima PCR Nested	2100	56°C	2 min
CHSINTRONFW/oeCHSex3rev	Seconda PCR Nested Touchd.	1800-2050 bp	60°-0.2° per ciclo	1min 30 sec
chsATG1fw/chsCintr2rev	Parte 5' <i>chsC</i>	1586 bp	54°C	1 min 30 sec
ChsCintr1Fw/CHSCODARE V2:	Parte 3' <i>chsC</i>	2070 bp	55°C	2 min
CHSX1F/CHSX2RN	Amplificazione Introne 1 <i>chsA</i> <i>chsB</i>	699bp/614bp/570bp	52°C	30 sec
CHSX1F/chsacvrev1	Amplificazione Introne 1 <i>chsA</i> varA	428	54°C	30 sec
CHSX1F/ intrsylvrev1	Amplificazione Introne 1 <i>chsA</i> varS	432	54°C	30sec

Figura 13. Condizioni di reazione delle principali PCR eseguite nel presente lavoro.

nome primer	Sequenza
OECHSPIIDL2C11FW(A)	CAAAGCATTCTAAGACATTGG
OECHSUTR5'FW(A)	TAAACCGGACCGGGCCGG
OECHSISOC11EX1FW (A)	GTAGCCAACACCCTCAGAACC
OECHSISOC11EX2FW (A)	ATGTAGTATTGCTTTTGCTCTG
CHSEX2FW	TTGCAGGTGATAAATCGATG
OECHSISOC34EX2FW (B)	TGTTGCATTGGAGATGATCTT
OE CHSDEGEX2REV	GTGAG(AGC)CC5GG(GCT)CC(AG)AA(GCTCC
ANTIRREV2	ACT ATG CAG CAC CAC CGT
ANTIRREV2DEG	ACT NTG CA5 CAC NAC 5GT
CHSINT2REV	GCCGACTGATTCATGTCAC
CHSINT2EX3REV	GAGTTCGAATCAGATATACC
OECHSESONE2REV	GTGAGGCCGGGCCCAAAG
CHSX1F	AGGAAAAATTCAAGCGCATG
CHSX2RN	TTCAGTCAAGTGATGTAACG
CHSVIOLAFW	TAYCARCARGGNTGYTTYGC
CHSVIOLAREV	GGRTGDGCDATCCARAAVA
CHSGENBFW	GACATGCCTGGYGCYDAYTAYCA
CHSGENBREV	ACRCATGCRCCTGACATGTTCCRTA
CHSEX2FW1	GACTTGGCGGAGAACATAA
CHSEX2FW2	CGGTTACTTTCCGGGGAC
OECHSEX2TOP1	ATGATGTACCAACAGGG
OECHSEX2TOPDEG1	ATGATGTACCAACAGGGNTGCT
OECHSEX2TOPDEG2	ATGATGTANCANCANGGNTGN
ISOADEGREV	GCA CNA CNG TCT CMA CNG TG
INTRSYLVFW1	GCAATTGAAAGCCCAAATATG
INTRSYLVREV1	CATATTTGGGCTTTCAATTGC
CHSACVFW1	CT CTG TCT ATA TTG TTA TTA GGC
CHSACVREV1	GCCTAATAACAATATAGACAGAG
OECHSISOC34EX1FW	CAAGCATAGAGGCTGTGCCAC
OECHSISOC11EX1FW	GTAGCCAACACCCTCAGAACC
OECHSUTR5'FW	TAAACCGGACCGGGCCGG
OECHSPIIDL2C11FW	CAAAGCATTCTAAGACATTGG
OECHSISOC34EX2FW	TGTTGCATTGGAGATGATCTT
OECHSISOC11EX2FW	ATGTAGTATTGCTTTTGCTCTG
OECHSEXON_ISOAGWFW1	GCAATCAAGGAATGGGGCCAGTCTAA
OECHSEXON_ISOAGWFW2	AAATCACCCATTTAGTCTTCTGCACCAC
OECHSINTRONCHSAFW1	ATCACCATGCAATTGAAAGCTTTTCTC
OECHSINTRONCHSAFW2	GGTTTTCAATTCGATTGGACAGTGTTA
OECHSINTRONCHSBFW1	GAAAGCACGATGCGACTGTGGATGACA
OECHSINTRONCHSBFW2	TCTTCCATTCCCATTTTGCTCGTCGT
ISOA-LIKEDEGREV1	CACNACNGTCTCNACNGTGAG
ISOA-LIKEDEGREV2	CTNTGNAGCACNACNGTCTC
ISOA-LIKEDEGREV3	CCCANTCNAGNCCNGTNGT
INTRSYLVFW2	CTCATCTCTTGAACGAGCTTAC
OECHSEX3DEGREV2	AGA AMT WSG AGT TCC ART C
OECHSEX3DEGREV3	CCR GGR TGW GCR ATC CAG AA
CHSCODAREV1	TTAATTAAGCGGCACACTATG
CHSCODAREV2:	TTAATTAAGTGGCTCACTGTG
CHSCODAREV3	TTAATTAAGNGGCNCACTATG

CHSCODAREV4	TCAAACCCTGGCTTGCTTAA
CHSCODAREV5	CAAATCAAACCCTGGCTTGC
CHSINTRON2ISOAFW	CCAAATTGGACAGCTCTGGT
CHSINTRON2ISOAFW2	TCTCCCAAATTGGACAGCTC
CHSINTRON2ISOBFW	AATCAGTCGGCGCACATT
OEGENBREV1	CACCTTCACTATCTGGTAGT
OEGENBREV2	GTTCTTTGATATCAATCCCG
M13 FOR	TTGTAACACGACGGCCAGT
M13 REV	CAGGAAACAGCTATGACC
CHSCINTR2FW1	GCCCCATTTCGAGTAGCTTA
CHSCINTR2FW2	TCTATCTCAAAGGGCGAACA
CHSCINTR2REV	TGTTCCGCCCTTTGAGATAGA
CHSATG1FW	AAT GGT NAC NGT NGA GGA GNT
CHSATG2FW	GAGGAGNTTCGCCNNGCTCAA
CHSATG3FW	ACNGTNGAGGAGNTTCGCC
CHSCINTR1FW	AAATGCTGCTATGTCATCATCG
CHSCINTR1REV	TTGTCCTAAGCAGCTTTTAATCA
CHSAINTRON2REV	ACC AGA GCT GTC CAA TTT GG
CHSAINT1F	ATT CAA CGA AAT GTC ACC ATG
CHSCINT1FW2	GGACATGGATTTTGCATGAG
CHSCINT1REV2	CACCGGGATCTCTGTTTATTC
CHSEX1FW	GATTACTATTTCCGCATCACA
OECHSEX3REV	GACATGTTGCCATACTCGCT
CHSINTRONFW	AGAAAATTCAAGCGCATGTG

Figura 14. Primer utilizzati

Capitolo 3.8 Nested PCR

La Nested PCR (Figura 15) è una tecnica utilizzata generalmente per ridurre la possibilità di ottenere amplificazioni aspecifiche e prevede la messa a punto di due PCR distinte, in successione. La seconda PCR utilizza una aliquota del prodotto di reazione della prima PCR o una sua diluizione come DNA stampo. Le due coppie di primer dovranno essere disegnate in modo che amplifichino una zona contenuta nel primo prodotto di reazione. In questo modo si riducono le amplificazioni aspecifiche in quanto è estremamente improbabile che gli

ampliconi non target possiedano sequenze complementari ad entrambe le coppie di primer.

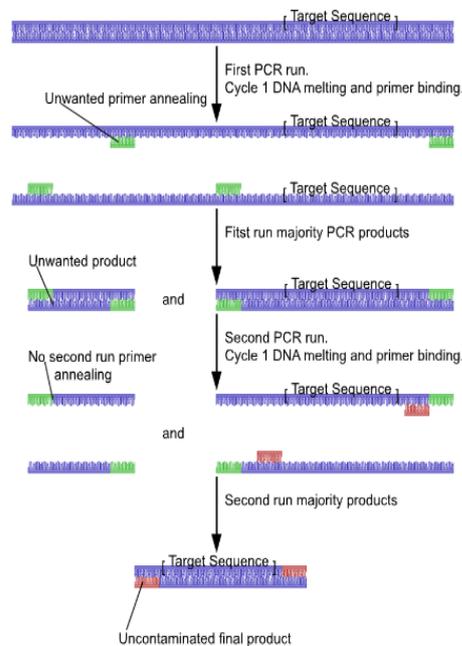


Figura 15. Schematizzazione delle fasi della Nested PCR.

Capitolo 3.9 PCR Touchdown

La PCR Touchdown è una particolare messa a punto della PCR che anche in questo caso è finalizzata ad eliminare eventuali amplificazioni aspecifiche. Sappiamo che la temperatura di annealing è un parametro fondamentale per la selettività della amplificazione. Nella PCR touchdown i primi cicli hanno una temperatura di annealing molto alta e prossima alla temperatura di melting. All'aumentare del numero di cicli la temperatura di annealing si abbassa proporzionalmente. Il primer si appaierà alla temperatura più alta che gli consenta l'appaiamento, le prime sequenze amplificate saranno quelle con la più alta affinità, quindi presumibilmente quelle di interesse. La sequenza target sarà amplificata nei cicli successivi e determinerà una alta specificità di reazione,

anche ai cicli nei quali la temperatura si abbassa ulteriormente. Nel nostro studio abbiamo utilizzato le seguenti condizioni PCR:

- una fase di denaturazione iniziale di 10 minuti a 95° C,
- 30 cicli di amplificazione consistenti in

denaturazione a 95° C per 1 minuto,
annealing a 60° C 1 minuto, -0,2° C per ciclo,
extension a 72° C,

- una fase finale di estensione a 72° C per 8 minuti.

Capitolo 3.10 PCR con primer degenerati

La reazione a catena della polimerasi generalmente usa primer specifici, costruiti su una sequenza nota del DNA. Esiste una variante della PCR che utilizza primer degenerati. Questa tecnica viene utilizzata in diverse applicazioni. Nel nostro caso abbiamo utilizzato questa tecnica per amplificare la sequenza completa dei geni *chsA* e *chsB*. Di questi geni conosciamo una parte di sequenza, che abbiamo utilizzato per disegnare primer specifici. Questi sono stati utilizzati con primer degenerati costruiti a partire dall'allineamento di sequenze di calcione sintasi di altre specie presenti in rete all'interno della banca di sequenze Genbank. In particolare abbiamo selezionato le zone a maggiore identità, quindi più conservate e probabilmente presenti anche nei geni *chs* di olivo. Abbiamo disegnato quindi diversi primer con un numero variabile di degenerazioni.

Capitolo 3.11 Purificazione di prodotti PCR

A seguito della reazione di amplificazione, i frammenti ottenuti, per poter essere utilizzati in successivi protocolli di clonaggio e sequenziamento, devono essere purificati dai sali e dalle altre componenti della miscela di reazione PCR.

Per tale purificazione è stato utilizzato il kit QIAquick® (Qiagen). Seguendo il protocollo descritto nel kit, si aggiungono 5 volumi di Buffer PBI per volume di reazione PCR. Una volta controllato, tramite analisi del colore della soluzione, che il pH non sia virato, si trasferisce la soluzione in un apposita colonnina (*QIAquick spin column*) contenente una membrana chimicamente affine al DNA, e si centrifuga per 1 minuto. Il DNA, ora legato alla membrana, viene lavato aggiungendo 750 µl di Buffer PE e centrifugando per 1 minuto; dopo aver eliminato il filtrato, si centrifuga per un minuto in modo da rimuovere completamente i residui di etanolo del Buffer PE.

La colonnina viene quindi inserita in una eppendorf da 1.5 ml e aggiunti 10 µl di Buffer EB (10mM Tris.Cl, pH 8.5) alla membrana. Dopo circa 15 minuti si centrifuga per 1 minuto in modo da recuperare il DNA nella eppendorf.

Capitolo 3.12 Clonaggio di prodotti PCR

Il clonaggio di frammenti amplificati in PCR è stato effettuato usando il kit TOPO-TA® Cloning (Invitrogen). Tale kit prevede l'utilizzo di un vettore linearizzato (pCRII-TOPO), attivato alle estremità libere (consistenti in un 3' aggittante composto da poli-T) dalla presenza della topoisomerasi I, ad esse legata in modo covalente. La Taq-polimerasi utilizzata nella reazioni PCR, ha un'attività indipendente dallo stampo per la quale aggiunge alla fine della reazione di polimerizzazione una sequenza di adenine al 3'. Questa caratteristica dei frammenti amplificati viene sfruttata per il loro clonaggio nei vettori TOPO-TA. I plasmidi ricombinanti che si ottengono possono essere usati sia per trasformare

cellule chimicamente competenti sia cellule elettrocompetenti. Nel nostro caso sono state trasformate cellule chimicamente competenti.

L'attivazione del vettore con il legame alle estremità della topoisomerasi I garantisce una ligazione efficiente in un tempo di circa 5 minuti.

La reazione di clonaggio viene allestita secondo il seguente schema:

Prodotto PCR	da 0.5 μ l a 4 μ l
Salt Solution	1 μ l
vettore TOPO-TA	1 μ l
dH ₂ O sterile	fino a un volume di 6 μ l

La *salt solution* è una soluzione salina con NaCl 1.2 M e MgCl₂ 0.06 M, necessaria per la trasformazione di cellule chimicamente competenti.

La reazione di clonaggio viene fatta procedere per 15 minuti a temperatura ambiente. La miscela di ligazione ottenuta può essere conservata a -20°C.

Capitolo 3.13 Trasformazione di *Escherichia coli*

Le cellule usate per la trasformazione sono cellule di *E. coli* fornite dalla Invitrogen e denominate “One Shot TOP10 Chemically Competent”. Sono cellule con alta efficienza di trasformazione e alta velocità di proliferazione. Vengono conservate a -80°C e per l'uso vengono scongelate in ghiaccio. A queste vengono aggiunti 2 μ l della miscela di ligazione precedentemente allestita. La sospensione cellulare viene mantenuta in ghiaccio per 25 minuti, quindi le cellule vengono sottoposte a uno shock termico a 42°C per 30 sec e immediatamente riposte in ghiaccio.

A questo punto si aggiungono 250 μ l di terreno SOC, fornito con il kit, e le cellule vengono incubate a 37° C per 1 ora in agitazione continua (200 rpm).

Le cellule vengono poi piastrate su un terreno di coltura massimo per batteri denominato LB, supplementato degli opportuni marcatori di selezione, che nel

nostro caso sono stati l'antibiotico ampicillina e il β -galattoside X-gal, essenziale per lo *screening* dei trasformanti.

Le piastre con i batteri vengono fatte crescere per una notte a 37°C. L'ampicillina seleziona le cellule trasformanti, in quanto sul vettore TOPO-TA è presente un gene marcatore di resistenza a questo antibiotico. Grazie alla presenza di X-gal nel terreno si può invece eseguire lo *screening* bianco-blu: l'inserzione del frammento nel vettore avviene in un sito *polylinker* localizzato tra il gene lacZ (*lacZ α*) e il suo promotore (Plac). I batteri trasformati che hanno catturato il plasmide nativo sono in grado di idrolizzare l'X-gal attraverso l'attività β -galattosidasica fornita loro dalla proteina codificata da *lacZ α* . Questa idrolisi produce un composto colorato di blu intenso, e quindi le colonie derivate da batteri con plasmide nativo sono blu. Per contro, le colonie che risultano bianche sono derivate da batteri che portano il plasmide ricombinante: in questi, l'inserzione del frammento nel sito *polylinker* impedisce l'espressione di *lacZ α* . Successivamente 50-100 colonie bianche vengono ristriciate su una nuova piastra selettiva, che viene incubata per una notte a 37°C per far crescere i batteri.

Capitolo 3.14 Screening delle colonie trasformanti via PCR

Per avere la certezza che una colonia porti il plasmide ricombinante con il giusto inserto clonato viene fatta una PCR sul lisato cellulare derivato dalla colonia.

I primers usati sono forniti dal kit e disegnati ai lati del sito *polylinker*, fiancheggiando così il frammento inserto.

Le cellule vengono prelevate dalle piastre con anse sterili e vengono stemperate in un volume di 10 μ L di acqua distillata sterile.

La sospensione ottenuta viene lisata a 100°C per 5 min.

A questo punto vengono aggiunti i reagenti di una reazione PCR e si fa procedere l'amplificazione nel termocicizzatore. Il controllo dell'amplificazione del frammento desiderato viene effettuato mediante gel elettroforesi.

Soluzioni di ampicillina e X-gal

L'ampicillina (Sigma) viene solubilizzata in acqua distillata sterile per ottenere delle soluzioni stock a concentrazione 50 mg/ml. Queste soluzioni vengono filtrate con filtri da 0.22 µm e conservate a -20°C. La concentrazione dell'ampicillina nei terreni di coltura quando viene usata come agente selettivo è 50-100 µg/ml.

L'X-gal (Inalco) viene sciolto in N,N-dimetilformammide e viene conservato a una concentrazione di 40 mg/ml a -20°C, al buio. Un totale di 1.6 mg di X-gal vengono usati per ogni piastra selettiva.

Terreno LB (Sambrook *et al.*,1989):

L'LB (Luria Broth) è un terreno di coltura massimo ideale per la crescita di *E.Coli*. Può essere preparato sia solido, per le colture in piastra, che liquido per le colture in sospensione. La composizione è la seguente:

Bacto-triptone (DIFCO)	10 g/l
NaCl (Baker)	10 g/l
Estratto di lievito (DIFCO)	5 g/l
Agar (DIFCO)	15 g/l (solo per terreno solido)
pH 7.5	

Il terreno deve essere sterilizzato in autoclave per 20 minuti a 120°C e 2 atm.

Le condizioni di crescita ottimali per *E.Coli* prevedono una temperatura di 37°C e, nel caso in cui le cellule siano in sospensione, un'agitazione costante a 225 rpm.

Capitolo 3.15 ESTRAZIONE DI DNA PLASMIDICO DA *E. coli*

Per l'estrazione dei plasmidi dalle colonie di *E. coli* trasformate è stato utilizzato il kit "Nucleospin® Plasmid" della Macherey-Nagel. Le cellule vengono fatte crescere in 4 ml di terreno LB liquido selettivo (contiene ampicillina in concentrazione di 100 µg/ml) per una notte a 37°C. La sospensione cellulare viene poi centrifugata in eppendorf da 2 ml per 5 min a 13000 rpm. Il pellet è poi risospeso in 250 µl di buffer A1.

Vengono quindi aggiunti 250 µl di buffer A2 prima, dopo di che la miscela è lasciata a temperatura ambiente per circa 5 minuti. Vengono quindi aggiunti 300 µl di buffer A3, ed il campione è successivamente centrifugato per 10 min a 13200 rpm.

Il sovranatante ottenuto è poi trasferito in una colonnina di filtraggio Nucleospin® Plasmid e centrifugato per 1 minuto a 13200 rpm. Eliminata la soluzione filtrata, la membrana presente nella colonnina è lavata prima con 500 µl di buffer AW e poi con 600 µl di buffer A4: dopo ogni lavaggio la colonna è centrifugata per 1 minuto a 13200 rpm per eliminare il buffer. L'eluizione dei plasmidi legati alla membrana è infine compiuta aggiungendo alla colonnina 50 µl di buffer AE, e centrifugando per 1 minuto a 13200 rpm.

L'integrità del DNA plasmidico estratto è controllata mediante elettroforesi su gel d'agarosio.

I frammenti PCR, una volta clonati nel vettore TOPO-TA sono stati sequenziati presso il Centro Interdipartimentale per le Biotecnologie di Interesse Agrario Chimico e Industriale (C.I.B.I.A.C.I.) dell'Università di Firenze.

Per sequenziare i prodotti PCR clonati in plasmide è necessario preparare una miscela contenente i seguenti componenti: una quantità di DNA variabile a seconda delle dimensioni del frammento, 3.2 pmol di primer e acqua distillata sterile fino ad un volume finale di 11 µl.

Il metodo di sequenziamento utilizzato sfrutta dideossinucleotidi trifosfati marcati con quattro fluorocromi diversi. Il kit usato è il BigDye Terminator v1.1 Cycle

Sequencing Kit (Applied Biosistem). con un sequenziatore automatico ABI Prism 310 (Perkin Elmer, Applied Biosystem).

Capitolo 3.16. Estrazione dei frammenti di DNA da gel di agarosio

Con questa procedura si può recuperare il DNA da gel di agarosio dopo la corsa elettroforetica. Per questa procedura è stato utilizzato il kit “PCR clean-up Gel extraction”(Macherey-Nagel). Il protocollo del Kit è stato seguito fedelmente in tutte le sue indicazioni.

- Si espone il gel alla luce ultravioletta.
- Si taglia il frammento di gel contenente la banda di interesse e si mette in una provetta da 1.5 mL .
- Si pesa una provetta da 1.5 mL vuota e poi si pesa quella contenente il pezzettino di gel in modo da risalire al peso del frammento recuperato.
- Si aggiungono 200µl di buffer NT per ogni 100 ng di agarosio.
- Si incuba nel bagnetto a 50°C per 5-10 minuti e poi si agita il campione per 2-3 minuti fino a che il gel non dissolve del tutto.
- Si pone una colonnina di estrazione “Nucleo Spin Extract II” in una provetta da 2mL e si aggiunge il campione (ciò serve per legare il DNA alla membrana).
- Si centrifuga per 1 minuto a 11.000x g per legare il DNA alla membrana.
- Si aggiungono 600µl di buffer NT3 che permette il lavaggio della membrana.
- Si centrifuga per 1 minuto a 11.000x g e si butta il sovrnatante.
- Si centrifuga per 2 minuti a 11.000 x g per rimuovere il buffer NT3 e poi si rimuove il sovrnatante.
- Si mette la colonnina di estrazione in una provetta da 1,5mL nuova.
- Si aggiunge da 15µl a 50µl di buffer di eluizione NE e si lascia a temperatura ambiente per qualche minuto e poi si centrifuga per 1 minuto a 11.000x g (si ripetono questi tre passaggi due volte).

Capitolo 3.17. Digestione del DNA con enzimi di restrizione

Un enzima di restrizione (o endonucleasi di restrizione) riconosce una specifica sequenza di coppie di basi, detta sito di restrizione, e taglia il DNA a livello di quella sequenza, idrolizzando l'ossatura zucchero fosfato. Tutti gli enzimi di restrizione tagliano il DNA tra il carbonio 3' ed il gruppo fosfato del legame fosfodiesterico, in modo che i frammenti prodotti per digestione possiedono estremità 5'fosfato e 3'-OH.

Gli enzimi di restrizione sono usati sia per produrre un pool di frammenti da clonare che per analizzare la posizione dei siti di restrizione in un frammento clonato o in frammento di DNA genomico.

Nel nostro caso la digestione del DNA genomico serve per frammentare il DNA in modo tale che, dopo corsa su gel d'agarosio e trasferimento su membrana, attraverso l'uso di una sonda opportuna, potessimo risalire al numero di copie di *chs* nel genoma di ulivo. Gli enzimi sono scelti per l'analisi tra quelli che non tagliano nella sequenza della sonda. Questa scelta si motiva perchè altrimenti l'ottenimento di due bande di ibridazione poteva essere dovuto sia al frammento originale tagliato in due, sia alla effettiva presenza di due copie. Per eliminare questa ambiguità si sceglie di non tagliare all'interno della sonda. Inoltre sono stati selezionati enzimi a sei basi perchè altrimenti enzimi a taglio più frequente avrebbero frammentato eccessivamente il DNA.

Noi abbiamo scelto EcoRI, BamHI, HindIII e abbiamo fatto anche le digestioni doppie EcoRI-BamHI, BamHI-HindIII.

- Per digerire 20 µg di DNA si calcola la quantità di DNA da prendere. Le digestioni sono state fatte in un volume totale di 250µl per le singole e 300µl per le doppie con tanti µl corrispondenti ai 20 µg di DNA, 20µl di enzima (nel caso di digestioni doppie 15µl di un enzima e 15µl dell'altro) 25µl di buffer (30µl nel caso delle doppie) e si porta a volume (250µl per le singole e 300µl per le doppie) con H₂O.

- Le digestioni si sono tenute in bagnetto a 37°C per 6 h.

Il buffer viene fornito insieme all'enzima ed è costituito da una miscela di sali che garantiscono il massimo di attività dell'enzima, alla concentrazione 10X, ovvero da diluire 1:10.

- Trascorso questo tempo si aggiunge una miscela costituita da 10µl di enzima (nel caso delle doppie 5µl di un enzima e 5µl dell'altro), 5µl di buffer 10X (10µl per le doppie) e si porta al volume totale con H₂O. Il volume totale della miscela deve essere 300µl per le digestioni singole e 400µl per le doppie.

- Si rimette a 37°C over night.

Capitolo 3.18. Marcatura della sonda

Per preparare la sonda da utilizzare in ibridazione abbiamo effettuato una PCR utilizzando primer opportuni per amplificare il DNA corrispondente alla sequenza codificante del gene di interesse. I reagenti utilizzati sono i comuni reagenti di una PCR standard (paragrafo 3.3), ad eccezione della miscela di deossiribonucleotidi trifosfati, comprendente oltre a 1mM di dATP, 1 mM di dGTP, 1 mM di dCTP, anche 0.65 mM di dTTP e 0.35 mM di dUTP-digossigenina, che è un tracciante non radioattivo. Si tratta di una molecola stereoidea estratta di *Digitalis purpurea* che, coniugata con dUTP, vien incorporata nella sequenza da marcare. Il frammento amplificato corrispondente alla sonda viene fatto correre su gel d'agarosio.

La miscela di deossiribonucleotidi trifosfati da usare per la marcatura del DNA (dNTP labeling mixture) è fornita con il kit "DIG High Prime DNA Labeling and Detection Starter Kit II" (Roche).

In Risultati questi i dNTPs marcati con la digoxigenina sono stati indicati con la sigla "dig lab".

Abbiamo fatto la seguente mix: Buffer 10X 2.5µl, dNTPs marcati 2µl, primer CHS ex2 FW 1µl, primer *chs*INT 2 rev , Taq 0.2 µl, dH₂O.

Diluizioni della sonda

Per verificare l'avvenuta marcatura della sonda abbiamo allestito una membrana di nylon dove abbiamo posto delle diluizioni della sonda e di uno standard.

Si sono effettuate le diluizioni 1:5, 1:10, 1:100, 1:1000 della sonda.

Abbiamo inoltre usato uno stock di DNA marcato con digoxigenina, fornito dal kit della Roche, e abbiamo fatto le diluizioni. Lo stock è alla concentrazione iniziale di 5,2 ng/ µl e lo abbiamo portato alle concentrazioni finali di 1 ng/ µl, 0,5 ng/ µl e 0,25 ng/ µl e da queste abbiamo fatto le diluizioni 1:10,1:100,1:1000. Il volume totale finale è stato di 10 µl.

Per la preparazione della membrana si deve porre uno spottino di 1 µl del tal quale della sonda, delle diluizioni 1:5, 1:10, 1:100, 1:1000 della sonda e dello standard sulla membrana, si lascia asciugare la membrana e si fissa il DNA alla membrana con UV crosslinking SPECTROLINKER CORPORATION XL-100.

Detection della sonda

Per rilevare il DNA dalla membrana si lava la membrana brevemente, per 2 minuti, in Maleic Acid buffer

- Si incuba per 30 minuti in 10 mL di Blocking Solution fonita dal produttore alla concentrazione 10X e si diluisce in Maleic Acid Buffer.
- Si incuba per 30 minuti in 10 mL di Antibody solution (diluire l'anticorpo, anti-digoxigenina, 1:10000 nella Blocking solution).
- Si fanno 2 lavaggi di 15 minuti ciascuno in 10 mL di Washing buffer per equilibrare la membrana.
- Si aggiungono 40µl di NBT/BCIP a 2 mL di Detection Buffer sulla membrana con il DNA rivolto verso l'alto.

-Dopo che si è sviluppato il colore si blocca la reazione con il TE.

Il rilevamento si basa su una reazione immunoenzimatica, dove la digossigenina funziona come aptene sul quale si lega una anticorpo specifico legato a sua volta all'enzima fosfatasi alcalina. Il colorante (NBT/BCIP) è costituito da un sale di

tetrazolio e da un fosfato, substrati della fosfatasi che, dalla reazione, origina un precipitato colorato sulla membrana.

Tutto ciò ci permette di valutare, per confronto con il DNA standard marcato, la concentrazione della sonda espressa in ng/ μ l.

La sonda da noi utilizzata ha una concentrazione di 5 ng/ μ l. In un esperimento di Southern Blot sono necessari circa 25 ng di sonda per ogni mL di Dig Easy Hib. La Dig Easy Hib è la soluzione di ibridazione che viene fornita da DIG High Prime DNA Labeling and Detection Starter Kit II sotto forma di granuli che si sciolgono attraverso l'aggiunta di 64 mL di acqua distillata e attraverso riscaldamento per 5 minuti a 37°C.

Riepilogo soluzioni utilizzate.

Maleic Acid buffer

Maleic acid 0.1 M

NaCl 0.15 M

pH 7.5

Washing buffer

Maleic acid 0.1 M

NaCl 0.15 M

Tween 20 0.3% (v/v)

pH 7.5

T.E.

Tris-HCl 10 mM

EDTA 1 mM

Capitolo 3.19. Trasferimento del DNA su membrana (Southern, 1975)

- Dopo la corsa elettroforetica si tratta il gel con HCl 0,25M per 15 minuti che serve a depurinare il DNA e ciò è necessario per il trasferimento di DNA ad alto peso molecolare.
- Dopo questo passaggio si denatura il DNA mediante trattamento per 30 minuti con una soluzione composta da NaCl 1,5M e NaOH 0,5 M.
- Si adagia il gel su una membrana di nylon, precedentemente montata su un dispositivo capace di creare il vuoto detto VACUUM-BLOT (Pharmacia). La creazione del vuoto facilita il trasferimento del DNA dal gel alla membrana, che essendo carica positivamente ha affinità per il DNA.
- Si mantiene il gel costantemente ricoperto da una soluzione di SSC 20X per tutta la durata del trasferimento (circa 3,5 ore).
- Si separa la membrana dal gel, si lava brevemente in SSC 2X e si mette ad asciugare su un foglio di carta assorbente.
- Si fissa il DNA alla membrana mediante raggi UV crosslinking SPECTROLINKER CORPORATION XL-100.

Riepilogo soluzioni utilizzate

Soluzione di depurazione

HCl 0.25 N

Soluzione di denaturazione

NaCl 1.5 M

NaOH 0.5 M

SSC 20X

NaCl 3M

Na-Citrato 0.3M

SSC 2X

È una diluizione 1:10 si SSC 20X.

Preibridazione

Per la reazione di ibridazione si usa un dispositivo chiamato ibridatore HYBRIDASER HB-200 (Techne). L'ibridatore è uno strumento dotato di un termostato, e capace di fare ruotare lentamente lungo il proprio asse dei cilindri di vetro chiusi alle estremità, in cui vengono inserite le soluzioni utilizzate per l'ibridazione. La membrana aderisce alle pareti interne del recipiente di ibridazione, e la rotazione fa sì che le soluzioni bagnino uniformemente tutta la superficie del filtro.

- La membrana sulla quale è stato trasferito il DNA viene prima messa in pre ibridazione, nel tubo apposito, nell'ibridatore, per 4-5 h con 10mL/ 100cm² di DIG Easy Hyb che è la soluzione di pre ibridazione fornita da "DIG High Prime DNA Labeling and Detection Starter Kit II"(Roche). La soluzione di pre ibridazione ha la stessa composizione della soluzione di ibridazione, ma non contiene la sonda marcata.

L'ampiezza della membrana per il Souther Blot in cui il DNA genomico è stato digerito mediante gli enzimi EcoRI, BamHI, HindIII, BamHI-EcoRI, BamHI-HindIII è di 131,2 cm². Sono stati utilizzati 20mL di DIG Easy Hyb.

L'ampiezza della membrana ibridazione per il Souther Blot in cui il DNA genomico è stato tagliato con EcoRI, BamHI, BamHI-EcoRI è di 150 cm². Sono stati utilizzati 20mL di DIG Easy Hyb.

Nel nostri esperimenti la preibridazione è sempre stata effettuata a 42°C per 4,5h.

Ibridazione

- Si denatura la sonda marcata (circa 25 ng/mL DIG Easy Hyb) con digoxigenina mettendola per 5 minuti in dH₂O.
- Si mette immediatamente in ghiaccio.
- Si toglie la soluzione di pre ibridazione dal tubo e si mette la soluzione di ibridazione contenente la sonda .
- Si rimette nell'ibridatore over night.

Poiché la nostra sonda ha una concentrazione di circa 5 ng sono stati utilizzati 100 μ l di sonda.

La temperatura di ibridazione per il Souther Blot in cui il DNA genomico è stato digerito mediante gli enzimi EcoRI, BamHI, HindIII, BamHI-EcoRI, BamHI-HindIII è stata, nel primo caso di 65°C, nel secondo caso di 60°C. A queste temperature sono stati effettuati anche i successivi lavaggi per la stringenza.

La temperatura di ibridazione per il Souther Blot in cui il DNA genomico è stato tagliato con EcoRI, BamHI, BamHI-EcoRI è stata di 60°C. A questa temperatura sono stati effettuati anche i successivi lavaggi per la stringenza.

Lavaggi stringenza

- Si effettuano lavaggi per eliminare la sonda che non si è legata in modo specifico: 2 lavaggi di 5 minuti ciascuno in 2 X SSC, 0.1% SDS a 15-25°C in costante agitazione e 2 lavaggi di 15 minuti ciascuno in 0.5 X SSC, alla temperatura alla quale è stata effettuata l'ibridazione over night.

Detection della membrana

- Si lava la membrana brevemente, da 1 a 5 minuti, in Washing buffer
- Si incuba per 30 minuti in 100 mL di Blocking Solution (diluire la Blocking solution 1:10 in Maleic Acid Buffer).
- Si incuba per 30 minuti in 20 mL di Antibody solution (diluire l'anticorpo, anti-digoxigenina, 1:10000 nella Blocking solution). L' anticorpo è coniugato con la fosfatasi alcalina
- Si fanno 2 lavaggi di 15 minuti ciascuno in 100 mL di Washing buffer.
- Si equilibria la membrana mettendola per 2-5 minuti in Detection Buffer.
- Si applica 1mL di CSPD, fornito dal Kit, sulla membrana con il DNA rivolto verso l'alto. IL CSPD è alla concentrazione di 0.25mM ed è il substrato della

fosfatasi alcalina. La defosforilazione del CSPD da parte della fosfatasi alcalina porta alla formazione di una molecola instabile che si decompone spontaneamente emettendo luce visibile.

- Si mette la membrana coperta con la stagnola per 10 minuti a 37° nella stufa.
- Si lascia la membrana, per 2/3 ore, a contatto con una lastra fotografica in camera oscura, che viene così impressionata dai fotoni emessi nella degradazione del substrato. Lo sviluppo della lastra mostra quindi segnali scuri in corrispondenza delle bande ibridate. Per verificare l'altezza delle bande nella corsa del DNA da trasferire viene inserito un marcatore di peso molecolare, marcato con digossigenina e quindi evidenziato sulla lastra fotografica.

Riepilogo soluzioni utilizzate

Maleic Acid buffer

Maleic acid 0.1 M
NaCl 0.15 M
pH 7.5

Washing buffer

Maleic acid 0.1 M
NaCl 0.15 M
Tween 20 0.3% (v/v) pH= 7.5

Detection Buffer

Tris-HCl 0.1 M
NaCl 0.1 M
pH 9.5

Capitolo 3.20. Analisi ISSR

Dodici primer ancorati in 5' o 3' (MWG-Biotech) sono stati utilizzati nell'analisi. Sette primer tutti costituiti da ripetizioni di-nucleotidiche hanno mostrato pattern chiari e riproducibili e sono stati utilizzati per la analisi ISSR. L'amplificazione è stata effettuata in un volume di 25 µl. Le reazioni sono state allestite con i seguenti reagenti:

- ❖ 20 ng di DNA genomico,
- ❖ 1 unità di Taq-polymerase (GE- Healthcare),
- ❖ 0.2 mM dNTPs (Fermentas),
- ❖ 1,32 pmol/μl di primer,
- ❖ 1X PCR buffer.

Le reazioni PCR sono state effettuate nel Termociclatore PTC-100™ P Thermal Cycler (MJ Research, INC., USA).

Le condizioni di reazione sono state le seguenti:

- ❖ 4 min at 94°C per la denaturazione iniziale,
- 35 cicli di:
- ❖ 60 s a 94°C (denaturazione),
 - ❖ 60 s a 45°C (annealing),
 - ❖ 90 s a 72°C (extension),
- 1 ciclo finale di extension di 5 min a 72°C

I prodotti PCR sono stati separati mediante elettroforesi su gel di agarosio-NuSieve 3:1 (Lonza) al 3% in buffer TBE 0.5X (0.0445M Tris, 0.0445M di acido borico e 0.001M EDTA), addizionato con 10 μg ml⁻¹ di bromuro di etidio e visualizzato con luce UV.

Per stimare il numero di bp dei frammenti amplificati è stato utilizzato il ladder 100bp plus (Fermentas). Tutte le reazioni ISSR sono state ripetute 2 volte e le bande che non erano riproducibili in tutte le reazioni sono state scartate. I pattern selezionati presentavano ampliconi che variavano da 150 a 2000 bp.

A partire dai profili di amplificazione ISSR abbiamo costruito una matrice di similarità basata sulla presenza (1) e sull'assenza (0) della banda visualizzata sul gel. Per la analisi della matrice binaria abbiamo utilizzato il programma *NTSYS-pc2.02i* (*Exeter Software*, New York, USA). La matrice di similarità ottenuta è stata utilizzata per costruire un albero filogenetico con il metodo UPGMA (unweighted pair-group method with arithmetic means). L'albero filogenetico ottenuto è stato validato utilizzando il programma MXCOMP che confronta la matrice di similarità e la matrice cofenetica (Rohlf, 1993).

Capitolo 3.21. Analisi bioinformatica

Disegno primers.

Alcuni primers utilizzati nel nostro lavoro sono stati disegnati tramite software Primer3 (<http://frodo.wi.mit.edu/>) disponibile in rete.

Allineamento sequenze

L'allineamento delle sequenze, sia nucleotidiche che aminoacidiche, è stata condotta tramite i programmi

Multalin (<http://bioinfo.genopole-toulouse.prd.fr/multalin/>),

ClustalW (<http://www.ebi.ac.uk/Tools/clustalw/index.html>),

Lalign (http://www.ch.embnet.org/software/LALIGN_form.html),

utilizzando matrici Blosum per i calcoli di similarità tra sequenze nucleotidiche e matrici PAM per le sequenze aminoacidiche.

Analisi similarità

Le analisi di similarità delle sequenze isolate durante il nostro studio sono state condotte tramite programma BLAST contro diverse banche dati, a seconda dei casi: banca dati EBI delle sequenze nucleotidiche (<http://www.ebi.ac.uk/blast2/nucleotide.html>), banca dati SWISSPROT di sequenze proteiche (<http://expasy.org/tools/blast/>), e le banche dati di domini proteici SMART (<http://smart.embl-heidelberg.de/>), PROSITE (<http://expasy.org/tools/scanprosite/>) e InterPro (<http://www-archbac.u-psud.fr/genomics/lbmgeiprscan.html>).

Traduzione in silico

La traduzione *in silico* delle sequenze nucleotidiche, in entrambi i filamenti e in tutte e tre le fasi di lettura, è stata operata tramite software Expasy Translate Tool (<http://www.expasy.ch/tools/dna.html>).

Predizione della struttura secondaria della proteina

Per la predizione della struttura secondaria delle proteine è stato utilizzato il software:

SOPMA(http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)

Predizione della conformazione tridimensionale della proteina

Per la predizione della struttura terziaria e tridimensionale della proteina è stato utilizzato il software freeware Accelrys Software Discovery Studio Visualizer 2.5,

Predizione segnali intronici

L'analisi delle sequenze nucleotidiche ottenute da campioni di DNA genomico e cDNA sono state analizzate per la ricerca di potenziali segnali intronici tramite programma NetPlantGene Server (<http://www.cbs.dtu.dk/services/NetPGene/>) in grado di compiere tale analisi sulla base di una banca dati di segnali di splicing di *Arabidopsis thaliana*.

Capitolo 4

Risultati

Capitolo 4.1 Isolamento e caratterizzazione molecolare della famiglia genica della calcone sintasi in *Olea europaea* L.

La calcone sintasi (CHS) è un enzima chiave della biosintesi dei flavonoidi. Il suo ruolo cruciale nel metabolismo secondario ha fatto sì che questo enzima e i geni che lo codificano siano ampiamente studiati. In letteratura sono presenti numerosi lavori di analisi genetica e biochimica che hanno evidenziato come CHS sia generalmente codificata da una piccola famiglia multigenica (Durbin et al., 1995; Koes et al., 1999). Tuttavia sono mancanti in olivo studi di caratterizzazione genetica, sia strutturale sia funzionale della CHS, con l'eccezione di un lavoro pubblicato nel 2004 da Ferrante e collaboratori relativo ad una sequenza di mRNA di 571 nucleotidi ottenuta dal pericarpo di drupa della cultivar Manzanillo in stadio di maturazione. La sequenza è stata depositata nel database *GenBank* del National Center for Biotechnology Information (NCBI) con il N° di accessione AF384049 (Figura 16)

GenBank: AF384049.1

[GenBank: AF384049.1](#)

Olea europaea chalcone synthase mRNA, partial cds

Change Region Shown
Customize View

Features Sequence

LOCUS AF384049 571 bp mRNA linear PLN 26-JUN-2001

DEFINITION Olea europaea chalcone synthase mRNA, partial cds.

ACCESSION AF384049

VERSION AF384049.1 GI:14550119

KEYWORDS .

SOURCE Olea europaea (common olive)

ORGANISM [Olea europaea](#)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Pick Primers
Design and test primers for this sequence

Olea europaea chalcone synthase mRNA, partial cds

(chalcone synthase olea)... (1)

Figura 16. Annotazione in GenBank della sequenza CHS di *Olea europaea*.

Successivamente al lavoro di Ferrante et al. (2004), nel nostro laboratorio è stato allestito un esperimento di “*genome walking*” a partire dalla sequenza depositata in GenBank, con lo scopo di isolare ed identificare la sequenza di CHS a monte di quella nota. A questo scopo è stato estratto il DNA da tessuto fogliare di *Olea europaea* cv Leccino, una delle cultivar con maggiore diffusione in Italia e mediante amplificazione PCR con primer specifici per la sequenza nota sono state identificate due sequenze, una di circa 2000 paia di basi e l’altra di circa 1500 bp. denominate *chsA* e *chsB* (Scialpi et al., 2005).

Capitolo 4.1.1 Analisi *in silico* dei frammenti *chsA* e *chsB*.

Le sequenze nucleotidiche dei frammenti denominati *chsA* e *chsB* sono state utilizzate come query per interrogare la banca dati NCBI, in particolare il database “nucleotide collection”, utilizzando il programma BLAST (Basic Local Alignment Search Tool). Questa analisi ha evidenziato per la sequenza di 2002 bp una alta identità (maggiore dell’80%) con sequenze di calcione sintasi di altre specie come *Digitalis Lanata*, *Malus x domestica*, *Nelumbo nucifera*. Per la sequenza di 1466 bp è stata dimostrata una alta identità (maggiore dell’80%) con numerose sequenze di calcione sintasi, ad esempio *Vitis vinifera*, *Malus x domestica* e *Persea americana*. Per entrambi i frammenti identificati tuttavia, la zona ad alta identità con le sequenze di CHS corrisponde alla parte finale della sequenza ottenuta (Figura 17) e cioè ad una parte codificante.

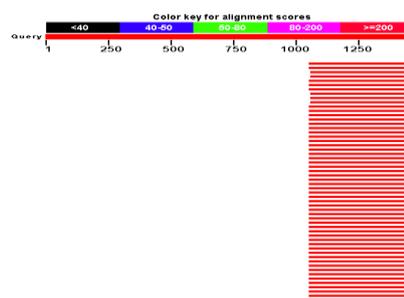


Figura 17. Risultato della analisi con il programma BLAST utilizzando come query la sequenza di 1466 bp.

La zona corrispondente alla regione 5'upstream, invece, essendo una regione non codificante a monte del gene, è molto più variabile e quindi diversa nelle diverse specie analizzate.

L'analisi *in silico* delle sequenze *chsA* (2002 bp) e *chsB* (1466 bp), ha evidenziato in entrambe la presenza di due Open Reading Frame (ORF) che iniziano con le basi ATG.

In *chsA* è stato possibile individuare una prima ORF a partire dalla base 871 fino alla base 1049. Per *chsB* la ORF inizia in corrispondenza della base 239 e termina alla base 417.

In entrambi i frammenti quindi è stata individuata una prima sequenza ORF di 178 paia di basi.

Inoltre è stato possibile individuare in entrambe un'altra ORF di 409 bp, che nel caso di *chsA* si estende dalla base 1592 alla 2001 e nel caso di *chsB* dalla base 1057 alla 1466. In figura 18 vengono paragonate le due sequenze *chsA* e *chsB*, ambedue costituite da due esoni intervallati da un introne.

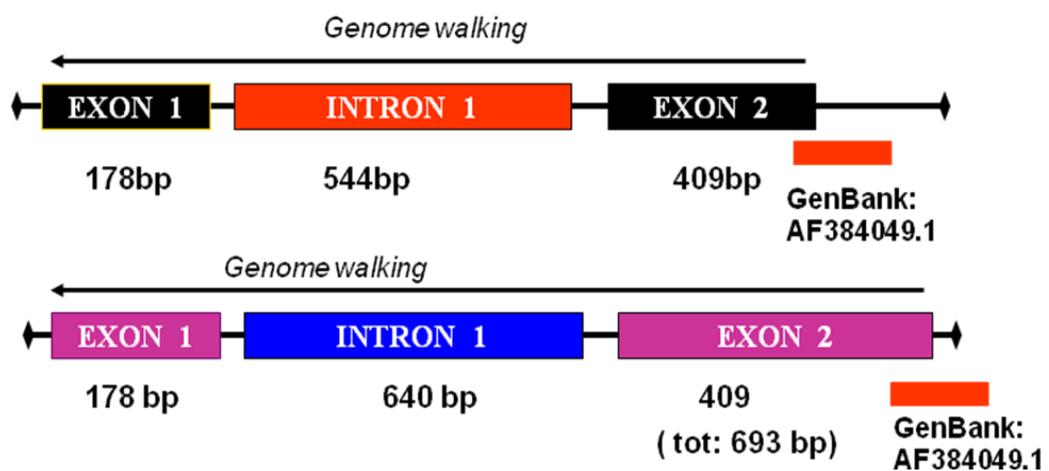


Figura 18. La struttura delle sequenze parziali di *chsA* e *chsB* ottenute con la tecnica Genome walking. Sulla parte destra della figura si noti la posizione della sequenza AF384049.1 depositata in Genbank.

Entrambe le sequenze tradotte *in silico* codificano per un polipeptide di 195 amminoacidi. Le sequenze proteiche dedotte da quella nucleotidica sono state utilizzate come query nel programma BLASTP di ricerca di similarità utilizzando database di sequenze di proteine. La ricerca ha evidenziato identità molto alte, fino al 100%, con proteine CHS presenti nel database.

A titolo di esempio sono elencate le prime 4 specie a identità più alta: *Nicotiana alata*, *Nicotiana benthamiana*, *Nicotiana attenuata*, *Misopates orientium*. Inoltre entrambe le sequenze codificanti isolate contengono alcuni residui amminoacidici che in letteratura sono considerati fondamentali per la funzione catalitica dell'enzima CHS ovvero Thy132, Met137, Cys164 (Ferrer et al., 1999).

Lo splicing *in silico* ha inoltre dimostrato che la sequenza non codificante dell'introne 1 che segue la prima ORF, si inserisce all'interno del codone che codifica il residuo di cisteina in posizione 60: questo punto di inserzione dell'introne 1 secondo la letteratura è altamente conservato (Yang et al., 2006).

Questi risultati hanno indicato che le due sequenze identificate corrispondano a due geni *chs* incompleti di *Olea europaea* e in particolare ne costituiscono in serie:

- ❖ la regione 5'-upstream,
- ❖ l'esone 1,
- ❖ l'introne 1
- ❖ parte dell'esone 2

L'allineamento delle sequenze geniche di *chsA* e *chsB* è mostrato in figura 19. L'allineamento delle due regioni codificanti di *chsA* e *chsB* mostra una identità reciproca del 92%, mentre l'allineamento delle due sequenze amminoacidiche mostra una identità del 93%

Come atteso dalla loro alta variabilità, le sequenze introniche mostrano invece una identità estremamente bassa (10%).

L'allineamento tra le sequenze *chsA* e *chsB*, e la sequenza numero AF384049 presente in GenBank mostra diversi livelli di identità per i due geni isolati, rispettivamente del 92% e del 99%.



Figura 19. Allineamento delle sequenze geniche parziali di *chsA* e *chsB*

Questo ci ha indicato che la sequenza GenBank AF384049 corrisponde a parte dell'esone 2 di *chsB* (Figura 20).

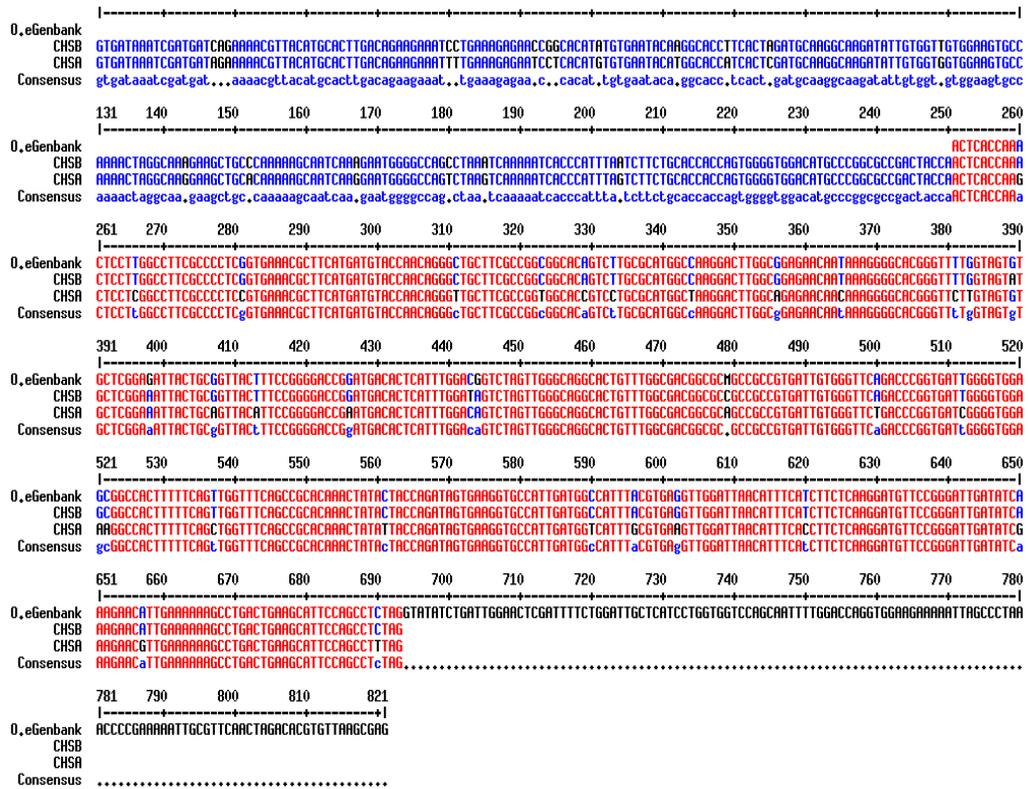


Figura 20. Allineamento tra l'esone 2 di *chsA*, *chsB* e la sequenza GenBank AF384049

La differenza di identità tra la sequenza *chsB* della cultivar Leccino e la sequenza *chsB* presente in Genbank può essere attribuita a variabilità tra due diverse varietà.

Capitolo 4.1.2 Isolamento e caratterizzazione molecolare della sequenza completa di *chsA* e *chsB* mediante l'uso di primer degenerati.

Le analisi descritte nel precedente paragrafo hanno indicato che le sequenze *chsA* e *chsB* non sono complete. Per cercare di identificare l'intera sequenza dei due geni *chs* precedentemente isolati mediante la tecnica di *genome walking* è stato utilizzato DNA genomico estratto da foglie giovani di *Olea europaea* cv. Leccino. L'estrazione e la successiva purificazione sono descritte nella sezione Materiali e Metodi. Per raggiungere questo obiettivo abbiamo scelto la tecnica dell'amplificazione PCR con primer degenerati anche se alternativamente potevano essere scelti approcci diversi, come ad esempio la 3'RACE PCR (Frohmann et al., 1988) oppure nuovamente il *genome walking*.

. Questa scelta si è basata sulle seguenti diverse motivazioni:

- ❖ la conoscenza parziale della sequenza ci ha permesso di utilizzare il primer *forward* ancorato;
- ❖ la presenza nelle sequenze isolate di una porzione non codificante corrispondente all'introne 1, ci ha consentito di utilizzare un primer *forward* specifico e in grado di selezionare *chsA* o *chsB* ed evitare in questo modo amplificazioni multiple, che avrebbero complicato notevolmente l'identificazione delle 2 diverse sequenze *chs*;
- ❖ la tecnica utilizzata ha garantito l'ottenimento diretto della sequenza completa del gene.
- ❖ Questa procedura risponde a criteri di economicità

Per la costruzione dei primers degenerati sono state allineate circa 30 sequenze *chs*, sia nucleotidiche, sia aminoacidiche presenti in Genbank avendo cura di scegliere sequenze complete e relative a specie filogeneticamente lontane da *Olea*. Per l'allineamento abbiamo utilizzato vari software disponibili in rete come ad esempio Multalin (Corpet, 1988; /), Lalign (Huang and Miller, 1991), ClustalW (Chenna et al., 2003).

L'allineamento delle zone 3'-terminali dei geni, ha evidenziato diversi motivi conservati subito a monte del codone di stop e una breve sequenza conservata immediatamente successiva ad esso. Utilizzando queste informazioni abbiamo disegnato 5 primer PCR nei quali abbiamo inserito un numero variabile di "basi degenerate", in particolare dalle 3 alle 8 basi, e con diversi gradi di degenerazione (dai due, ad esempio A/T, ai tre, ad esempio A/T/C).

Per la degenerazione non si sono usate basi modificate come ad esempio l'Inosina, ma miscele diverse di nucleotidi. In questo modo una degenerazione completa ATGC vede ogni base presente in quella data posizione nel primer, in quantità relativa di ¼. Come primer *forward* abbiamo utilizzato primer selettivi per *chsA* o *chsB*, costruiti nella regione a monte del sito di inizio della ORF di entrambi i geni, oppure costruiti all'interno dell'introne 1 di entrambe le sequenze. Tra le numerose combinazioni di primer e le varie condizioni di reazione provate e messe a punto, la coppia di primer PCR 5'CHSAFW/CHSCODA1rev e 5'CHSBFW/CHSCODA3rev ha permesso di amplificare con successo due frammenti, rispettivamente di 2200 e 2140 bp.

I frammenti ottenuti sono stati clonati e sequenziati come descritto in Materiali e Metodi.

L'analisi mediante allineamento multiplo delle sequenze ottenute e di quelle precedentemente isolate ha evidenziato tra queste una identità del 100%.

Si è quindi così dimostrato che gli ampliconi ottenuti mediante PCR con primer degenerati corrispondono alle sequenze complete dei geni *chsA* e *chsB*.

L'analisi di similarità tra le sequenze complete ottenute e la banca dati dell'EMBL-EBI effettuata con il software FASTA, ha dato i seguenti risultati:

per la sequenza *chsA*

- ❖ una identità dell'80% con la sequenza *chs3* di *Vitis vinifera*, (GenBank, N° di accessione AB066274);
- ❖ una identità del 79.1% con la sequenza *chs* di *Sorbus aucuparia* (GenBank, N° di accessione DQ286037);

- ❖ una identità del 79.1% con la sequenza CHS di *Gossypium hirsutum* (GenBank, N° di accessione EF643507).

per la sequenza *chsB*

- ❖ una identità del 79.1% con la sequenza *chs3* di *Vitis vinifera* (GenBank, N° di accessione AB066274);
- ❖ una identità del 78.9% con *Digitalis lanata* (GenBank, N° di accessione AJ002526);
- ❖ una identità dell'78.3 con *Misopates orontium* (GenBank, N° di accessione AM162205).

Negli esperimenti finora descritti entrambe le sequenze mostravano una rilevante eccezione strutturale, ovvero una inattesa interruzione di lettura della ORF in corrispondenza del nucleotide 693 a valle dell'inizio dell'esone 2 (Figura 21).

La sequenza in ORF riprendeva poi in entrambi i geni ed arrivava correttamente al codone di stop.

Si è visto poi che nel caso di *chsA* la ORF rientra in fase 492 bp e nel caso di *chsB* 271 bp a valle dell'interruzione per cui è ipotizzabile la presenza di una possibile seconda sequenza intronica posizionata dopo la metà dell'esone 2.

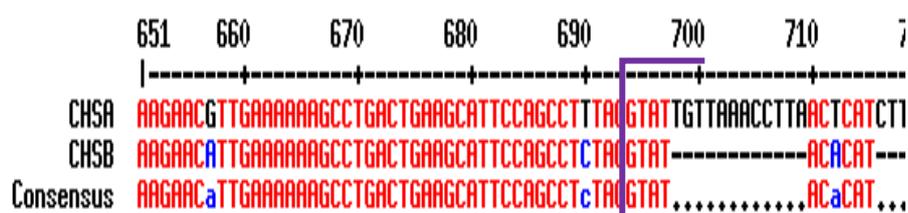


Figura 21: Interruzione della ORF situata a 693 bp a valle dell'inizio dell'introne 2 in CHSA e CHSB, segnalata dalla linea viola.

Per verificare questa ipotesi abbiamo effettuato una analisi di previsione delle sequenze consenso introniche con diversi software tra i quali il programma NetPlant gene service (<http://www.cbs.dtu.dk/services/NetPGen/>), Netgene2 ed altri (vedi Materiali e Metodi) che hanno confermato tutti, con una confidenza maggiore o uguale al 95%, la presenza di siti consenso di splicing in corrispondenza dell'interruzione e del ripristino della lettura *in frame* dopo di essa.

La presenza di un secondo introne nella stessa posizione in entrambi i paraloghi ci ha notevolmente sorpreso dato che in letteratura la struttura genica delle calcione sintasi veniva universalmente ritenuta estremamente conservata e corrispondente a due esoni separati da un introne (Yang et al., 2006) con la eccezione di *Antirrhinum majus* che presenta una struttura dei geni *chs* a tre esoni e due introni (Sommer e Sadle, 1986). In questo caso specifico il secondo introne si inserisce però in posizione differente rispetto a quanto da noi ritrovato sia per il gene *chsA*, sia per il gene *chsB*.

Questi ultimi risultati indicano quindi che la sequenza isolata corrispondente alla sequenza completa della calcione sintasi chiamata *chsA* contiene:

- ❖ l'esone 1 (178bp),
- ❖ l'introne 1 (543 bp),
- ❖ l'esone 2 (693 bp),
- ❖ l'introne 2 (492 bp),
- ❖ l'esone 3 (299 bp)
- ❖ 4 bp della regione 3'- downstream.

La sequenza completa di *chsB* comprende

- ❖ l'esone 1 (178 bp),
- ❖ l'introne 1 (634 bp),
- ❖ l'esone 2 (693 bp),
- ❖ l'introne 2 (271 bp),

- ❖ l'esone 3 (299 bp)
- ❖ 19 bp della regione 3' downstream

La struttura dei geni è rappresentata schematicamente nella figura 21.

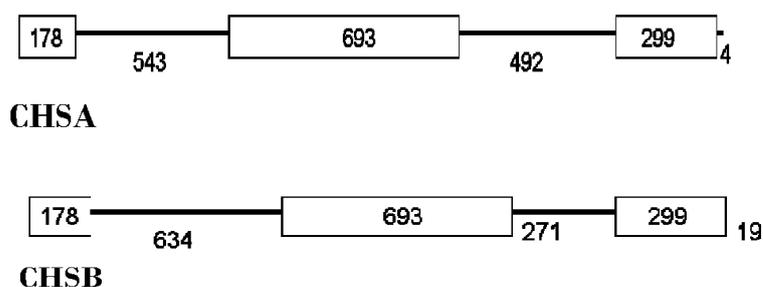


Figura 22. La struttura schematizzata di *chsA* e *chsB*. Le linee corrispondono agli introni mentre i rettangoli corrispondono agli esoni. In figura sono indicate le lunghezze in bp.

La struttura degli esoni appare costante nei due geni: in entrambi l'esone 1 risulta essere il più piccolo in dimensione visto che codifica per 59 residui, mentre l'esone 2 risulta essere il più grande dei tre, codificando per 231 residui. L'esone 3 codifica per 100 residui amminoacidici. Per quanto concerne gli introni è stata riscontrata una diversità di lunghezza nei geni *chsA* e *chsB*, in accordo con quanto riportato per i geni *chs* isolati in altre specie (Xu et al., 2008). Gli introni 1 e 2 di entrambi i geni seguono la regola "GT-AG" delle sequenze di consenso. L'introne 1 è posizionato tra la prima e la seconda base della tripletta che codifica per l'aminoacido cisteina in accordo a quanto osservato nelle altre specie. L'introne 2, è posizionato tra la seconda e la terza base della tripletta che codifica per il residuo amminoacidico glicina. E' da notare che la posizione del secondo introne in *Antirrhinum majus* è diversa: in questa specie infatti, il secondo introne si

inserisce tra le basi codificanti i residui di Glu162 e Gly163 (Sommer and Saedler 1986). Inoltre, l'allineamento tra le due sequenze (figura 23) ha indicato che gli esoni di *chsA* e *chsB* mostrano una identità del 91.3%. Le differenze di sequenza sono dovute a variazioni di singola base (SNPs). L'analisi della diversità di sequenza condotta con il software DNAsp (Rozas et al., 2003), riporta negli esoni la presenza di 1068 siti monomorfici e di 102 siti polimorfici.

La traduzione *in silico* della sequenza, effettuata con il software ExPASy disponibile in rete, ci ha indicato che i siti polimorfici comportano 79 mutazioni sinonime e 23 mutazioni missenso.

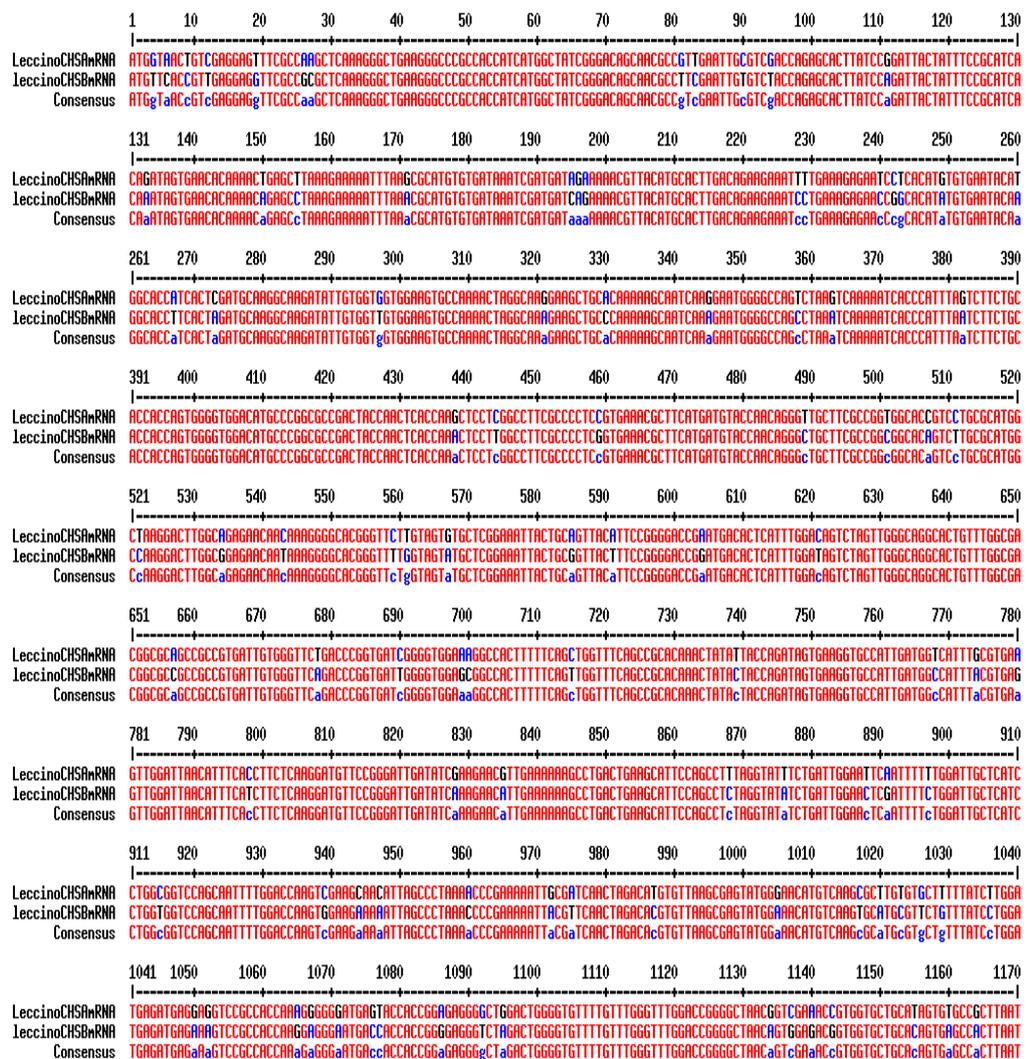


Figura 23. Allineamento tra le sequenze codificanti complete di *chsA* e *chsB*

Come atteso le sequenze introniche hanno mostrato, invece, una identità molto bassa, pari al 50.9%.

Entrambe le sequenze *chs* in olivo esprimono una proteina di 390 amminoacidi che differisce per il punto isoelettrico teorico, calcolato con il software Expasy, corrispondente a 5.85 in CHSA e a 6.24 in CHSB. La proteina CHSA presenta, inoltre un peso molecolare di 42.73 kDa, mentre CHSB ha un peso molecolare di 43.03 kDa. Entrambi i paraloghi conservano i quattro residui reattivi del sito attivo ossia Cys164, Phe215, His303, Asn336; i residui del coumaroyl binding pocket Ser133, Glu192, Thr194, Thr197, Ser338 e i residui del cyclization pocket Thr132, Met137, Phe215, Ile254, Gly256, Phe265, Pro375 (Ferrer et al. 1999) che sono conservati nelle proteine CHS.

Nella Figura 24 e 26 sono riportate le sequenze amminoacidiche rispettivamente di CHSA e CHSB e gli amminoacidi conservati, indicati con diversi colori. Inoltre in figura 25 e 27 è riportata la struttura secondaria.

MVTVEEFRQAQRAEGPATIMAIGTATPLNCVDQSTYPDYYFRITDSEHKTELKEKFKR
 MCDKSMIEKRYMHLTEEILKENPHMCEYMAPSLDARQDIVVVEVPKLGKEAAQKAI
 KEWGQSKSKITHLVFCTTS **G**VDM **P**GADYQLTKLLGLRPSVKRFMMY **Q** **O** **G** **C** **F** **A** **G** **G** **T**
 VLRLMAKDLAENNKGARVLVVCSE **E** **I** **A** **V** **T** **F** **R** **G** **P** **N** **D** **H** **L** **D** **S** **L** **V** **G** **Q** **A** **I** **G** **D** **G** **A** **A** **V** **I** **V** **G** **S**
DPVIGVERPLFQLVSAAQTLPDSEGA **I** **D** **G** **H** **L** **R** **E** **V** **G** **L** **T** **F** **H** **L** **L** **K** **D** **V** **P** **G** **L** **I** **S** **K** **N** **V** **E** **K** **S** **L** **T**
 AFQPLGISDWNSIFWIA **H** **P** **G** **G** **P** **A** **I** **L** **D** **Q** **V** **E** **A** **T** **L** **A** **L** **K** **P** **E** **K** **L** **R** **S** **T** **R** **H** **V** **L** **S** **E** **Y** **G** **N** **M** **S** **S** **A** **C** **V** **L**
 FILDEMRRSATKGGMSTTGEGLDWGVLF **G** **F** **G** **P** **L** **T** **V** **E** **T** **V** **V** **L** **H** **S** **V** **P** **L** **N**

Figura 24. La sequenza amminoacidica dedotta dalla sequenza nucleotidica di *chsA*. In grigio i residui codificati dai nucleotidi dell'esone 1. Evidenziati in giallo i residui costituenti il sito attivo; evidenziati in verde i residui costituenti il coumaroyl- binding pocket; evidenziati in azzurro i residui costituenti la cyclization pocket; evidenziati in viola i residui importanti per geometria del sito attivo. I residui sottolineati fanno parte di domini conservati.

La Tabella in figura 28 mostra il confronto tra le sequenze amminoacidiche di CHSA e CHSB di olivo, la CHS1 di vite, una PKS di tipo III di platano e una CHS like di pioppo eseguito con il software clustalW. La Tabella indica i punteggi assegnati in base al livello di identità nell'allineamento migliore, diviso il numero di residui confrontati; sono esclusi da questo confronto i "gap" ovvero le inserzioni e le delezioni (Wilbur e Lipman, 1983).

E' evidente la similitudine tra le due CHS di olivo (94%) e tra CHS di olivo e vite, che sia per CHSA che per CHSB è pari all'85%. La PKS di tipo III di platano risulta essere marcatamente differente dalle altre, con un punteggio che non supera mai il 40%. Più vicina alle CHS di vite e olivo invece risulta essere la CHS-like di pioppo, con punteggi del 65%. Utilizzando il software Accelrys Software Discovery Studio Visualizer 2.5, è stato infine possibile riprodurre, a partire dalla sequenza amminoacidica la struttura terziaria tridimensionale di entrambe le proteine CHS di olivo (Figure 29 e 30).



Figura 29. Riproduzione della struttura tridimensionale della proteina CHSA.

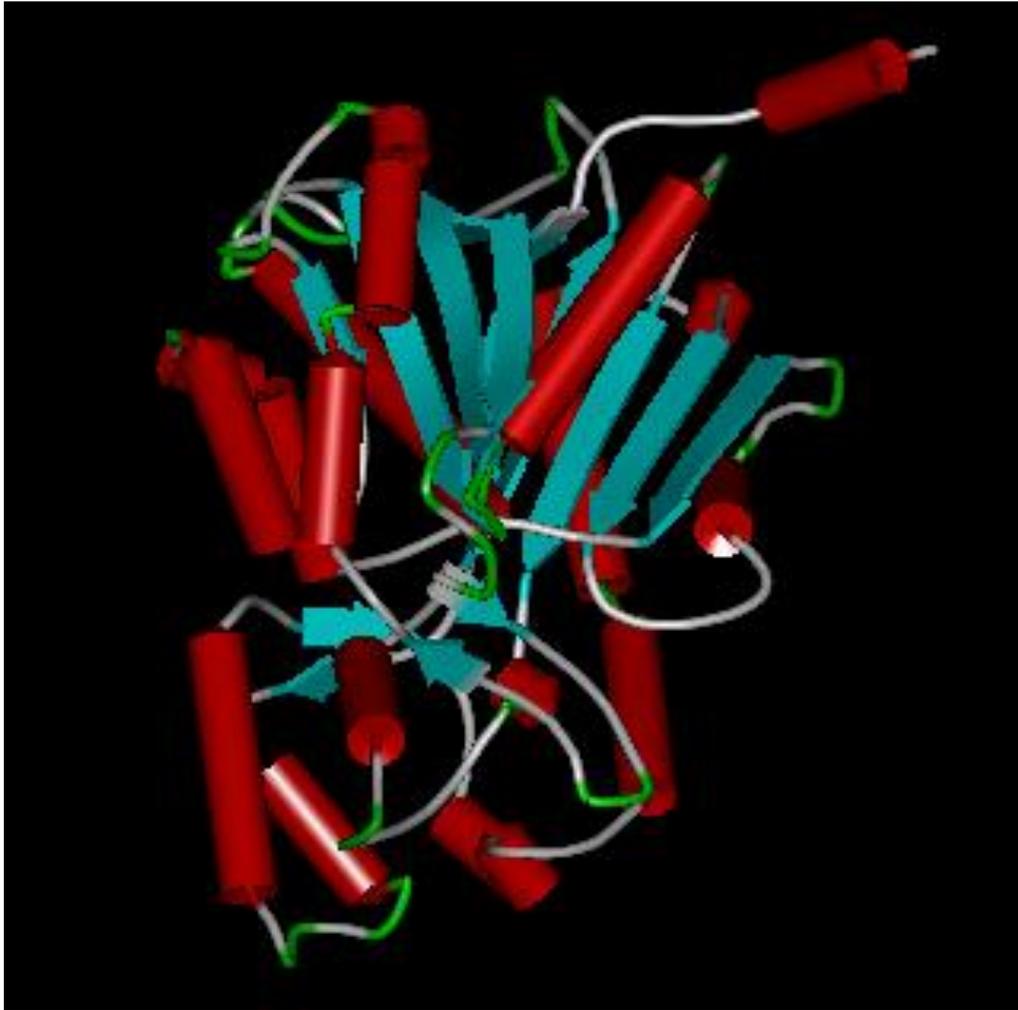


Figura 30. Riproduzione della struttura tridimensionale della proteina CHSB

Confrontando ed allineando le sequenze tradotte *in silico* delle proteine CHS di *Olea europaea* con altri geni CHS appartenenti a diversi ordini e famiglie (Figura 31), più o meno correlate dal punto di vista filogenetico, è possibile osservare come ci siano numerose zone e motivi conservati (evidenziati in rosso) in tutte le specie considerate, distribuiti in tutta la sequenza, e in particolare nella parte centrale della proteina

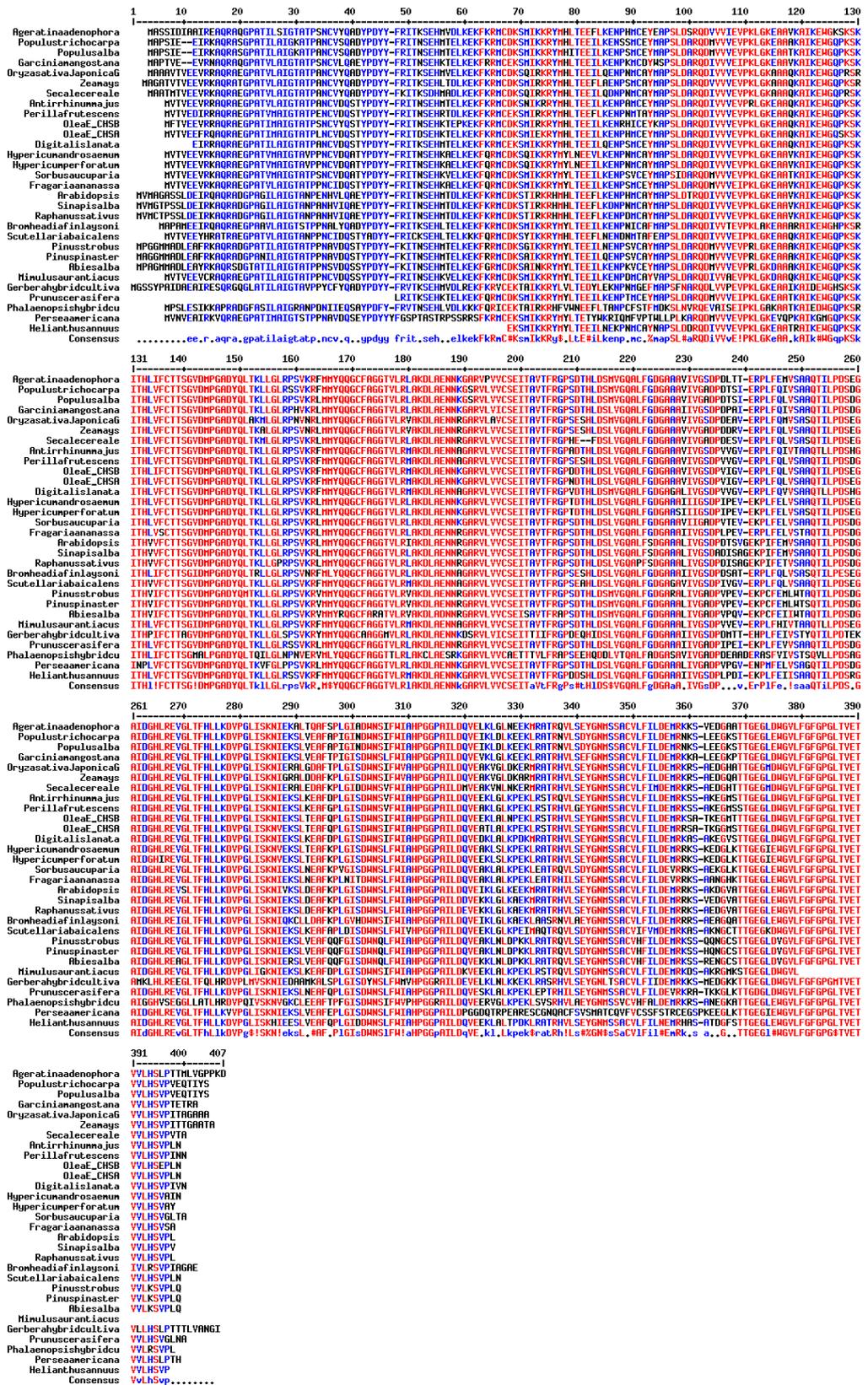


Figura 31. Allineamento delle sequenze tradotte *in silico* delle proteine CHS di *Olea europaea* con altri geni CHS appartenenti a diversi ordini e famiglie

Capitolo 4.1.3 Analisi del numero di copie dei geni della famiglia *chs* in olivo mediante Southern blot

Il numero di geni che in olivo compongono la famiglia genica della Calcione Sintasi è stato stimato utilizzando la tecnica di ibridazione Southern blot.

Gli esperimenti di Southern Blot sono stati eseguiti utilizzando DNA genomico di *Olea europaea* cultivar Leccino, digerito con enzimi di restrizione che non tagliano dentro la sonda utilizzata per l'ibridazione (vedi Materiali e Metodi).

Come DNA sonda è stata scelta la sequenza dell'esone 2 di *chsB* (Figura 32)



Figura 32. La sequenza scelta come sonda utilizzata per l'ibridazione.

che tuttavia condivide una elevata omologia di sequenza con *chsA*. Questa sonda ci garantisce da un lato la capacità di individuare sia *chsA* che *chsB* ma, trattandosi di una zona altamente conservata, potenzialmente anche altri eventuali geni facenti parte della famiglia CHS.

L'individuazione degli enzimi di restrizione che non tagliano dentro la sonda è stata effettuata con il programma Webcutter 2.0 disponibile in rete (<http://bio.lundberg.gu.se/cutter2/>). Tra tutti quelli individuati dal programma abbiamo selezionato HindIII, EcoRI, BamHI, tutti caratterizzati da una sequenza consenso di 6 basi. Sono state quindi condotte digestioni con un solo enzima e digestioni doppie. A questo scopo 20 µg di DNA genomico sono stati digeriti per 12 h con gli enzimi di restrizione BamHI; EcoRI; HindIII; BamHI-EcoRI; BamHI-HindIII. Come descritto in Materiali e Metodi la digestione è stata condotta in un volume finale di reazione di 250 µl nel caso di digestioni singole e di 300 µl nel caso di digestioni doppie. Al termine della reazione, i prodotti di digestione sono stati precipitati con alcol e sali, risospesi e caricati su un gel di agarosio 1% e separati tramite elettroforesi (Figura 33).

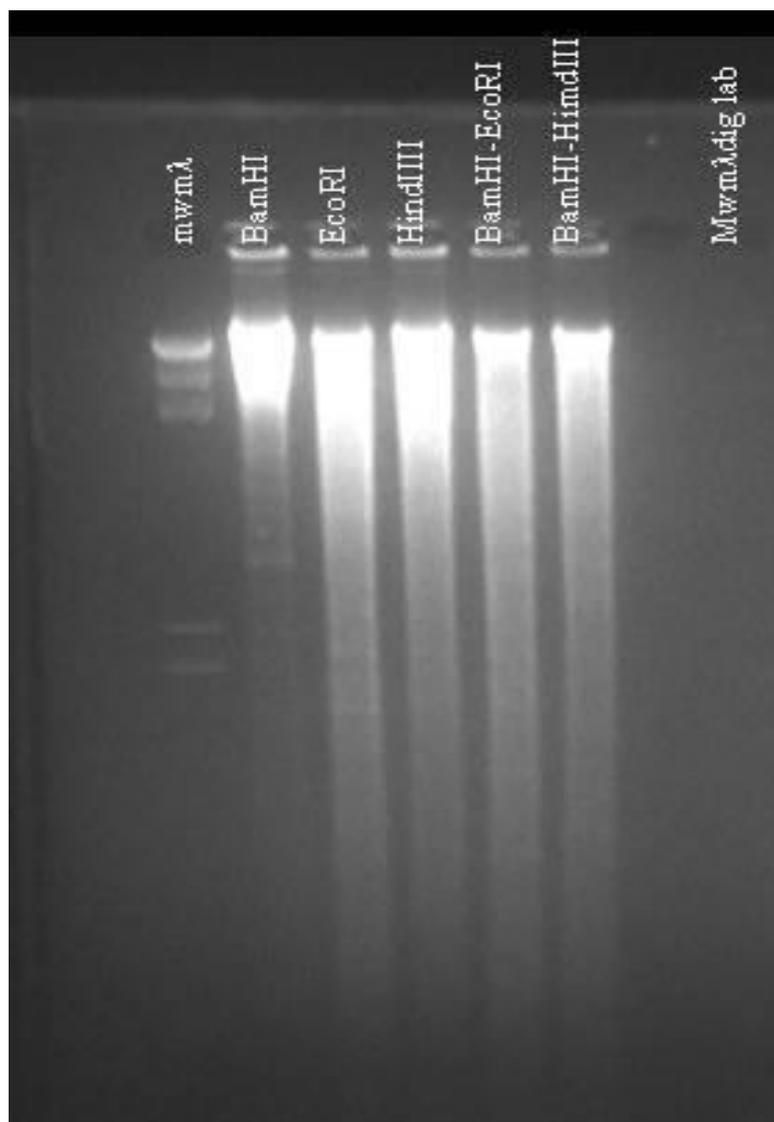


Figura 33. Elettroforesi su gel di agarosio delle reazioni di digestione con gli enzimi di restrizione BamHI; EcoRI; HindIII; BamHI-EcoRI; BamHI-HindIII.

Al termine della corsa il gel è stato trasferito su una membrana di Nylon (Boehringer Monnheim) mediante blotting e successivamente ibridato con la sonda costruita sull'esone 2 di *chsB* marcata con digossigenina, alla temperatura di 60°C. Per la reazione di ibridazione e la marcatura della sonda è stato utilizzato il *DIG-UTP, Non Radioactive Labeling and Detection Kit* (Roche®). Per il

rilevamento del segnale sono state utilizzate lastre ultrasensibili, sviluppate dopo 3 ore di esposizione (vedi Materiali e Metodi).

I risultati ottenuti hanno messo in evidenza la presenza di due bande in corrispondenza delle digestioni **BamHI-HindIII**; **HindIII**; **EcoRI** (Figura 34). Questi risultati sono pertanto compatibili con la presenza di due copie del gene nella famiglia genica della Calcone sintasi in *Olea europaea* e confermano i dati ottenuti mediante amplificazione PCR (“*genome walking*”, amplificazione con primer degenerati).

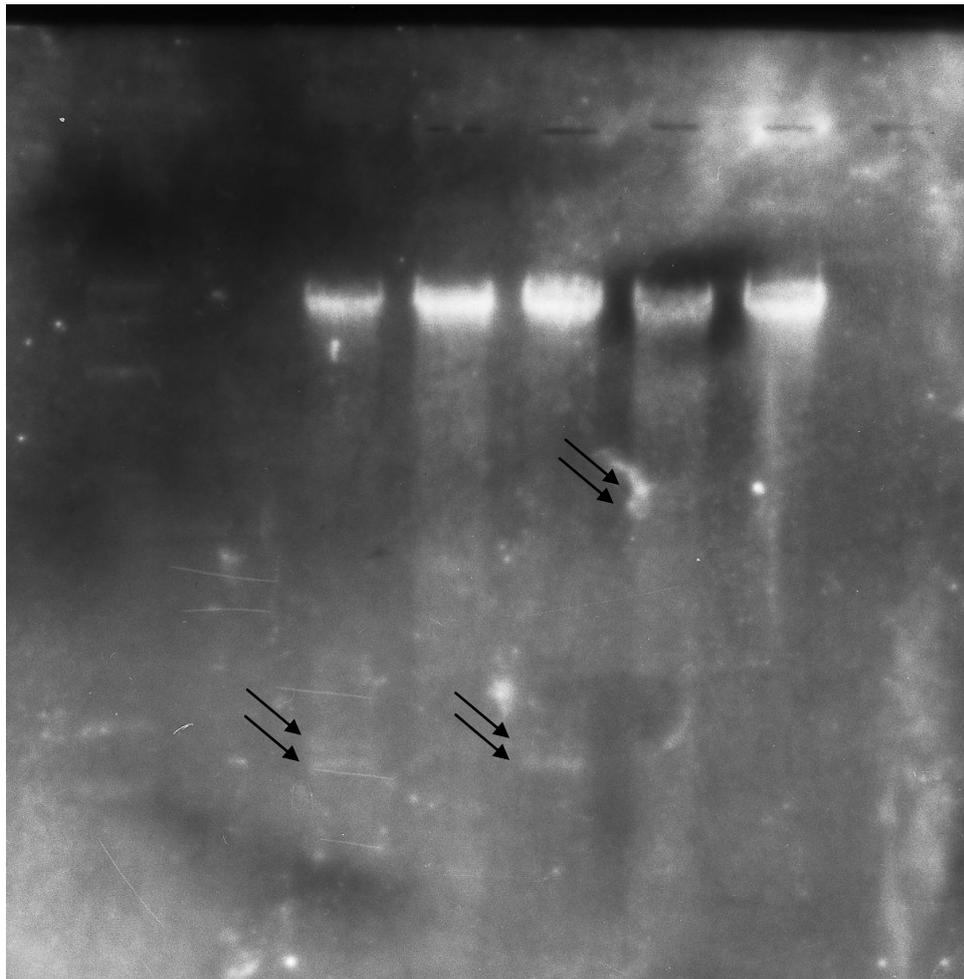


Figura 34. Southern blot del DNA genomico digerito con enzimi di restrizione di *Olea europaea* cv Leccino ibridato con una sonda marcata con digossigenina corrispondente all'esone 2 di CHS. I segnali sono evidenziati dalle frecce. In figura (speculari rispetto alla foto della digestione): **BamHI-HindIII**; **BamHI-EcoRI**; **HindIII**; **EcoRI**; **BamHI**.

Capitolo 4.1.4 Identificazione di un terzo paralogo del gene *chs*

Al fine di identificare ulteriori possibili isoforme del gene codificante la calcione sintasi in *Olea europaea* non rilevabili per ibridazione, abbiamo utilizzato ancora una volta una strategia basata su amplificazione PCR, questa volta una “touch-down” nested-PCR (vedi Materiali e Metodi).

Come schematizzato in Figura 35, sono stati disegnati diversi primer, utilizzati in due successive amplificazioni PCR. Per la prima reazione PCR abbiamo utilizzato i primer indicati dalle frecce rosse. Il primer *forward* è stato disegnato in prossimità della giunzione tra il primo esone e il primo introne, in una zona altamente conservata, non solo all’interno di una specie o di un genere ma anche tra famiglie molto distanti dal punto di vista evolutivo.

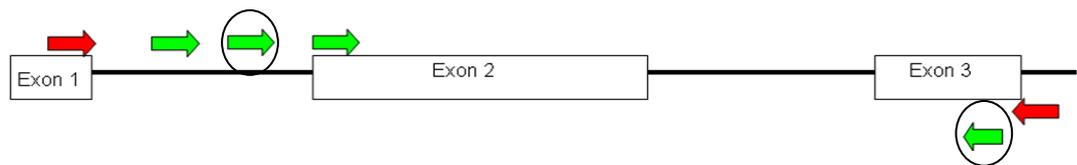


Figura 35. Schema della Nested-PCR. Le frecce rosse indicano i primer utilizzati nella prima PCR, le frecce verdi indicano i primer utilizzati nella seconda PCR.

Il primer reverse è stato disegnato in corrispondenza dell'estremità 3' dell'esone 3, anche questo dunque in una zona altamente conservata. Le condizioni di bassa stringenza utilizzate nel protocollo sperimentale della prima reazione di amplificazione PCR e illustrate in Materiali e Metodi, poco restrittive sia per quanto riguarda la temperatura di *annealing*, sia per quanto riguarda il tempo di estensione sono state concepite per ottenere la massima probabilità di sintesi di altri geni *chs*. La PCR successiva è stata eseguita usando come stampo una diluizione 1:10 del prodotto di reazione della prima PCR. Come primer forward sono state utilizzate diverse sequenze, sia in zone conservate come l'inizio

dell'esone 2 che in zone non conservate in corrispondenza dell'introne 1 (frece verdi della Figura 35). In particolare è stato utilizzato un primer denominato INTRONFW, disegnato sulla sequenza di *chsA*, evidenziato con un cerchio in figura 34.

In questa amplificazione PCR le condizioni di reazione utilizzate sono state stringenti e per quanto riguarda la temperatura di annealing è stata utilizzata la tecnica touchdown con un abbassamento di 0,2°C per ciclo di reazione.

I risultati di questa amplificazione sono riportati nella Figura 35. La seconda reazione PCR effettuata con la coppia di primer INTRONFW/CHSex3REV ha prodotto un frammento di dimensione attesa di 1800 bp ed un ulteriore frammento di maggiori dimensioni (2050 bp) (Figura 36).

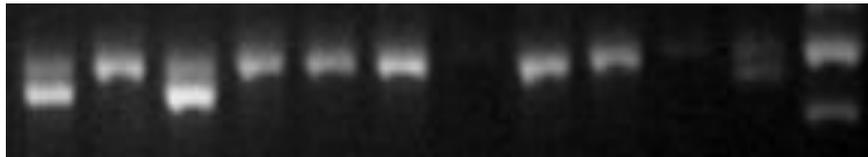


Figura 36. Elettroforesi dei prodotti di reazione della seconda PCR. Nella ultima corsia del gel il marcatore di 2000 bp e di 1500 bp del Ladder 100bp plus (Fermentas).

Il frammento di 2050 bp è stato clonato e sequenziato. E' stato quindi effettuato un allineamento tra le sequenze di *chsA*, *chsB* e del nuovo prodotto PCR. I risultati di questo allineamento hanno messo in evidenza l'80% di identità di sequenza di questo frammento con i geni *chsA* e *chsB*.

Come possiamo vedere in figura 37, in una serie successiva di esperimenti abbiamo disegnato ulteriori coppie di primer in corrispondenza delle zone non conservate del prodotto di amplificazione di 2050 bp .



Figura 37. Schematizzazione delle PCR effettuate per l'isolamento completo della sequenza del gene CHSC. Il riquadro azzurro indica la porzione di sequenza isolata con l'esperimento Nested PCR illustrato precedentemente; in basso la struttura del gene isolato con questo esperimento

In particolare in figura è illustrata la strategia che ha portato all'isolamento della sequenza completa: la freccia viola indica il primer forward CHSATG1Fw costruito in una zona conservata di *chsA* e *chsB* in prossimità dell'inizio della sequenza codificante. La freccia rossa indica il primer reverse utilizzato in coppia col precedente e disegnato su una sequenza del frammento isolato non conservata con *chsA* e *chsB* in modo da ottenere l'amplificazione esclusiva della parte in 5' della nuova sequenza.

Per ottenere l'isolamento della parte in 3' della nuova abbiamo disegnato un primer forward in una altra sequenza non conservata del frammento ottenuto con la Nested PCR (in figura rappresentato con una freccia verde) e lo abbiamo utilizzato con un primer reverse costruito in prossimità del termine della sequenza codificante di *chsA* e *chsB* in una zona conservata.

Queste amplificazioni hanno condotto all'isolamento delle sequenze a monte e a valle della sequenza ottenuta con la Nested PCR.

L'intera sequenza è risultata essere un terzo paralogo *chs* che abbiamo denominato *chsC*.

L'analisi di similarità con le sequenze presenti nei database dell'EMBL-EBI effettuata con il software FASTA ha individuato infatti una alta identità e similarità con i geni *chs* di varie specie tra cui *Forsythia x intermedia*

appartenente alla famiglia delle *Oleaceae*, con una percentuale di identità pari all'88.6% (Figura 38).

Alignment	DB:ID	Source	Length	Identity%	Similar%	Overlap	E()
1 <input type="checkbox"/>	EM_PL:X93225	Forsythia x intermedia DNA encodi	553	88.6	88.6	542	4e-105
2 <input type="checkbox"/>	EM_PL:AB512766	Camellia japonica CHS mRNA fo	1446	76.9	76.9	649	9.6e-92
3 <input type="checkbox"/>	EM_PL:D26593	Camellia sinensis CHS1 mRNA for c	1390	76.7	76.7	649	2.9e-91
4 <input type="checkbox"/>	EM_PL:AY169403	Camellia sinensis chalcone sy	1385	76.7	76.7	649	2.9e-91
5 <input type="checkbox"/>	EM_PL:AB011467	Hydrangea macrophylla chs mRN	1510	76.8	76.8	647	3.3e-91
6 <input type="checkbox"/>	EM_PL:AF456448	Hydrangea macrophylla chalcon	1483	76.5	76.5	647	3e-90
7 <input checked="" type="checkbox"/>	EM_PL:D26594	Camellia sinensis CHS2 mRNA for c	1405	76.1	76.1	649	2.5e-89
8 <input type="checkbox"/>	EM_PL:AM472295	Vitis vinifera contig VV78X16	4340	76.0	76.0	649	6.8e-89
9 <input type="checkbox"/>	EM_EST:EC924887	WIN027.TB24.1_A06 Cab Sauv f	946	76.0	76.0	649	7.7e-89
10 <input type="checkbox"/>	EM_EST:EC933813	WIN0419.C21_G08 Cab Sauv per	924	76.0	76.0	649	7.7e-89

Figura 38. Analisi di similarità effettuata con il software FASTA, utilizzando come query la sequenza *chsC*

In particolare, l'allineamento tra la sequenza parziale del gene *chs* di *Forsythia x intermedia* (X93225) e *chsC* ha messo in evidenza in quest'ultimo una inserzione di circa 150 bp (Figura 39).

Nell'allineamento tra i geni *chs* di olivo questa inserzione che interrompe la sequenza dell'esone 2 è risultata essere il secondo introne di *chsC* (Figura 40). Questo risultato spiega quindi come la sonda costruita sull'esone 2 di *chsB* non sia riuscita ad evidenziare la presenza di una terza isoforma *chs* nell'esperimento di ibridazione precedentemente descritto.

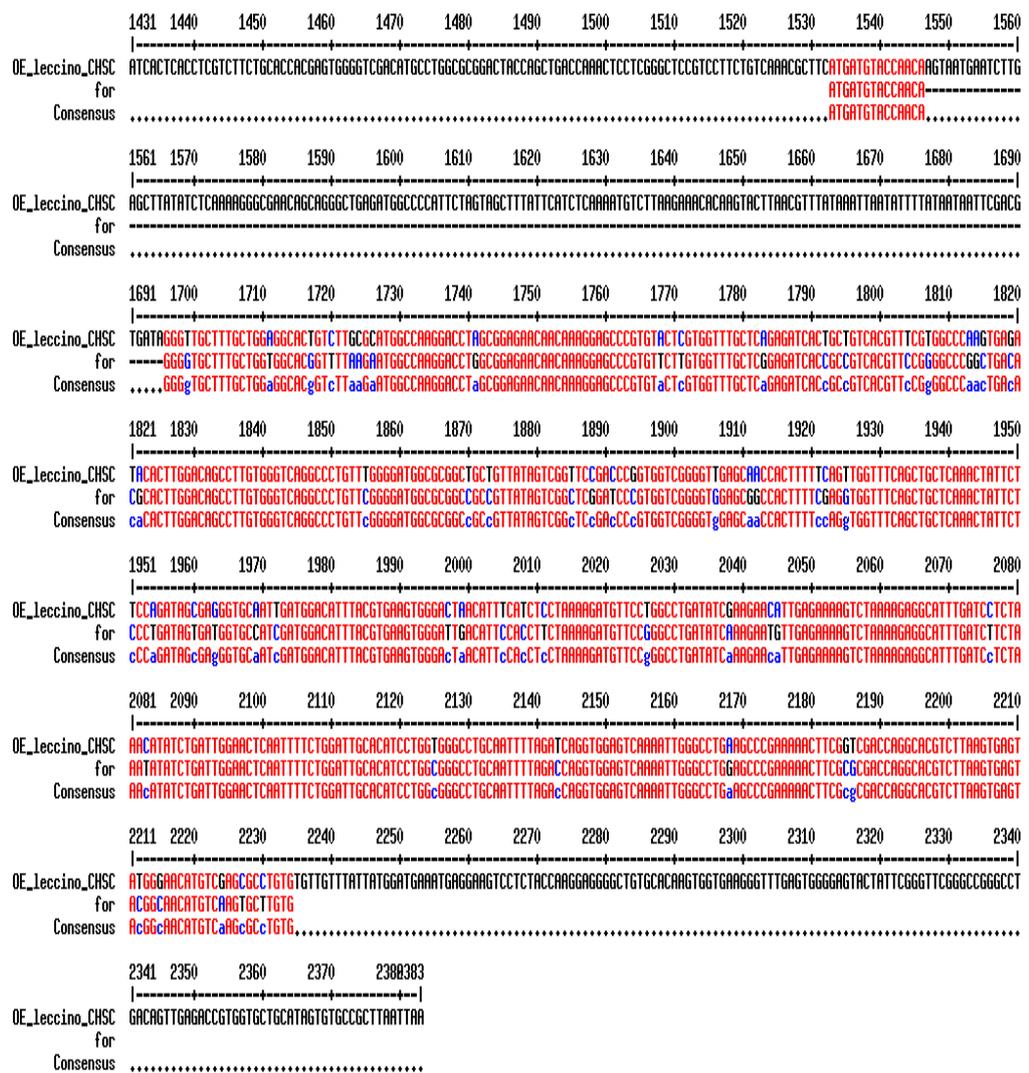


Figura 39. Allineamento della sequenza *chsC* con la sequenza di *Forsythia x intermedia* (X93225) presente in Genbank.

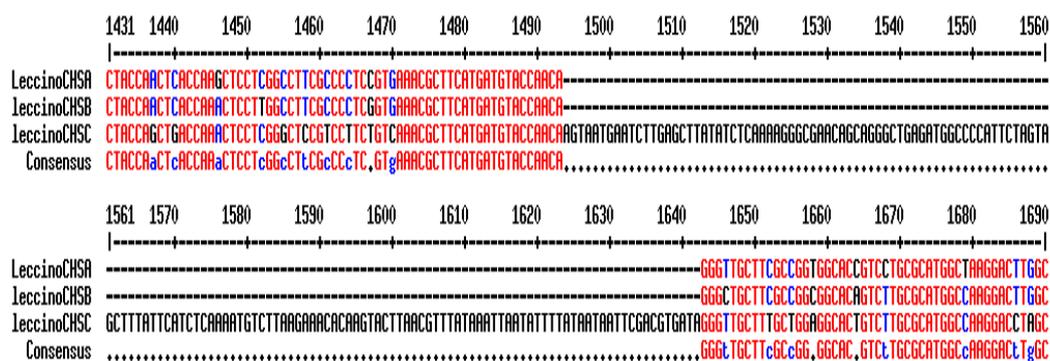


Figura 40. Allineamento delle sequenze genomiche di CHSA CHSB e CHSC, dove si mostra la presenza dell'introne 2 di CHSC

L'allineamento tra gli esoni di *chsC* con i geni *chsA* e *chsB* ha evidenziato in entrambi i casi una alta identità, pari all'80.7% tra *chsC* e *chsA* e pari all'80.4% tra *chsC* e *chsB*.

Questi risultati suggeriscono che *chsC* sia effettivamente un altro gene appartenente alla famiglia della Calcione sintasi di *Olea europaea*.

Nella Figura 41 è riportato lo schema che raffigura la struttura del gene *chsC* di olivo.

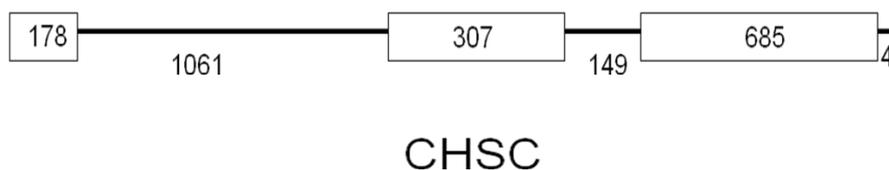


Figura 41. Struttura del gene *chsC*

Il gene *chsC* mostra una ORF di 1170 bp interrotta da due introni rispettivamente di 1061 bp e di 149 bp.

Anche *chsC* è costituito da tre esoni e due introni, in analogia con quanto osservato per *chsA* e *chsB*. Nella Figura 42 è riportato il confronto dell'organizzazione strutturale dei geni *chsA*, *chsB* e *chsC*.

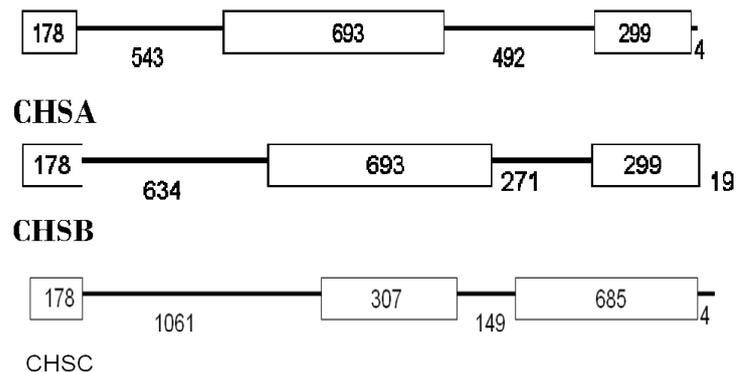


Figura 42. Confronto delle strutture dei geni *chsA*, *chsB* e *chsC*.

Come si può osservare sono presenti alcune importanti differenze strutturali tra i tre geni. Le analogie e le differenze di struttura tra i tre geni *chs* possono essere così riassunte:

- ❖ la grandezza dell'esone 1 resta costante (in analogia con altri geni *chs* in diverse specie);
- ❖ il sito di inserzione dell'introne 1 è lo stesso in tutti i paraloghi (in analogia con altri geni *chs* in diverse specie);
- ❖ l'introne 1 e l'introne 2 presentano una lunghezza variabile nei tre geni ;
- ❖ nel gene *chsC* l'introne 2 va ad inserirsi in una diversa posizione rispetto a *chsA* e *chsB*: il residuo dove si inserisce infatti è posizionato tra il residuo glutammina 162 (Gln162) e glicina 163 (Gly163), nella stessa posizione di

inserzione del secondo introne di *Anthirrhinum majus* (quest'ultimo al posto di Gln presenta acido Glutammico (Glu));

- ❖ l'esone 2 e l'esone 3 del gene *chsC* hanno lunghezze diverse rispetto a quelle dei geni *chsA* e *chsB*;
- ❖ l'esone 2 di *chsA* e *chsB* codifica per la maggioranza dei residui amminoacidici conservati della proteina;
- ❖ nel gene *chsC* i residui altamente conservati vengono ripartiti tra l'esone 2 e l'esone 3.

La proteina CHSC tradotta *in silico* (Figura 43) mostra il 90.8% di identità di sequenza con CHSA e il 90% con CHSB (Figura 44) e presenta tutti i residui conservati e specifici per la disposizione del sito attivo, l'attività catalitica, il sito del cumaroil CoA. In figura 45 è rappresentata la struttura secondaria predetta con il software SOPMA e in figura 46 possiamo visualizzare la struttura terziaria tridimensionale della proteina *chsC* ottenuta a partire dalla sequenza del gene, utilizzando il software Accelrys Software Discovery Studio Visualizer 2.5.

MVTVEELRHAQRAEGPATIMAIGTATPTNCVDQSSYPDFYFRITNSE
HMTDLKEKFKRMCDKSMIKKRYMYLNEEILKENPNMCAYMAPSLDAR
QDIVVVEVPKLGKEAAQRAIKEWGQPKSKITHLVFCTTSGVDMPGAD
YQLTKLLGLRPSVKRFMMYQQCFAGGTVLRMAKDLAENNKGARVLV
VCSEITAVTFRGPSEIHLDSL VGQALFGDGAAAVIVGSDPVVGVEQP
LFQLVSAAQ TILPDSEGADDEHLREVGLTPEHLLKDVPGLISKNIEKS
LKEAFDPLNISDWNSIFWIAEPGGPAILDQVESKLGKPEKLRSTRH
VLSEYGNMSSACVLFIMDEMRSSTKEGLCTSGEGFEWGVLF GFGPG
LTVETVVLHSVPLN

Figura 43. La sequenza amminoacidica dedotta dalla sequenza nucleotidica di *chsC*. In grigio scuro i residui codificati dai nucleotidi dell'esone 1, in grigio chiaro i residui codificati dall'esone 3. Evidenziati in giallo i residui costituenti il sito attivo; evidenziati in verde i residui costituenti il cumaroyl- binding pocket; evidenziati in azzurro i residui costituenti la cyclization pocket; evidenziati in viola i residui importanti per geometria del sito attivo.

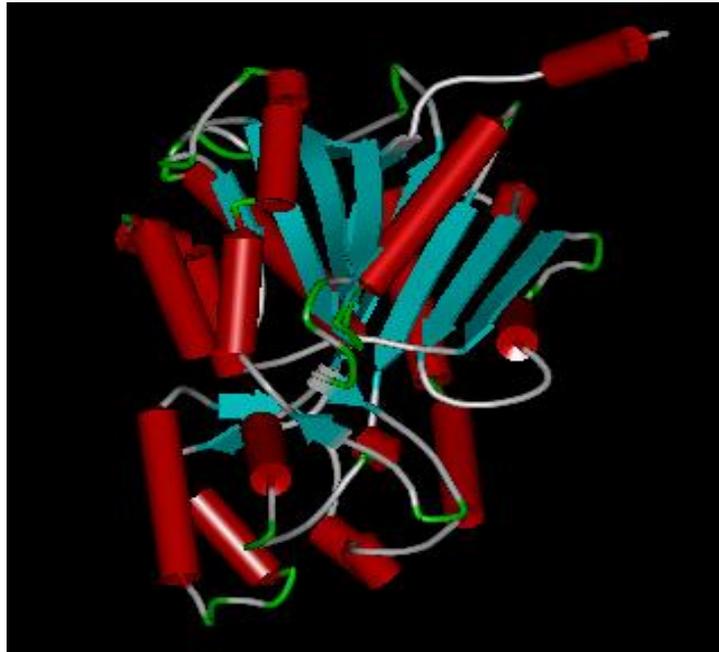


Figura 46. La struttura terziaria tridimensionale predetta della proteina chsC.

Dal confronto degli mRNA dedotti dalle sequenze geniche di CHSA CHSB e CHSC della cultivar Leccino (Figura 47) si evidenzia l'assenza di delezioni o inserzioni in un totale di 1170 basi di mRNA. I siti monomorfici sono 902, pari al 77% e i siti polimorfici sono 268 per un totale di 291 mutazioni.

L'allineamento delle sequenze proteiche CHS di 30 specie appartenenti a generi diversi e delle tre proteine CHS di olivo, effettuato utilizzando il software Multalin (Figura 48), ha permesso di costruire un albero filogenetico utilizzando il software Mega4. L'albero filogenetico, illustrato in Figura 49 e basato sul metodo UPGMA, mostra come le specie siano raggruppate in cluster omogenei che riflettono le posizioni sistematiche. Ad esempio l'Ordine *Lamiales* si raggruppa in un unico cluster con l'unica eccezione di *Scutellaria baicalensis* che appare isolata. Si può osservare come CHSA, CHSB e CHSC di olivo formino un raggruppamento omogeneo con *Perilla frutescens* anch'essa appartenente alla famiglia delle *Lamiaceae*. Questo raggruppamento è vicino ad altri due sub-cluster dell'ordine *Lamiales*, rispettivamente quello costituito da *Antirrhinum*

majus e *Digitalis lanata* (famiglia *Plantaginaceae*) e da *Mimulus aurantiacus* della famiglia delle *Phrymaceae*.

La tendenza delle specie a raggrupparsi insieme ad altre appartenenti allo stesso Ordine è confermata in altri cluster. Ad esempio le famiglie *Salicaceae* e *Clusiaceae* che clusterizzano insieme e formano un raggruppamento omogeneo dell'Ordine *Malpighiales*.

In questo tipo di analisi sono state inserite come out-group due sequenze CHS appartenenti a due specie di due diversi regni, rispettivamente il cianobatterio unicellulare *Synechococcus elongatus* batteri ed il fungo ascomiceto *Microsporium canis*. Come è possibile osservare essi vengono raggruppati a margine dell'albero filogenetico, ben separati dalle specie vegetali.

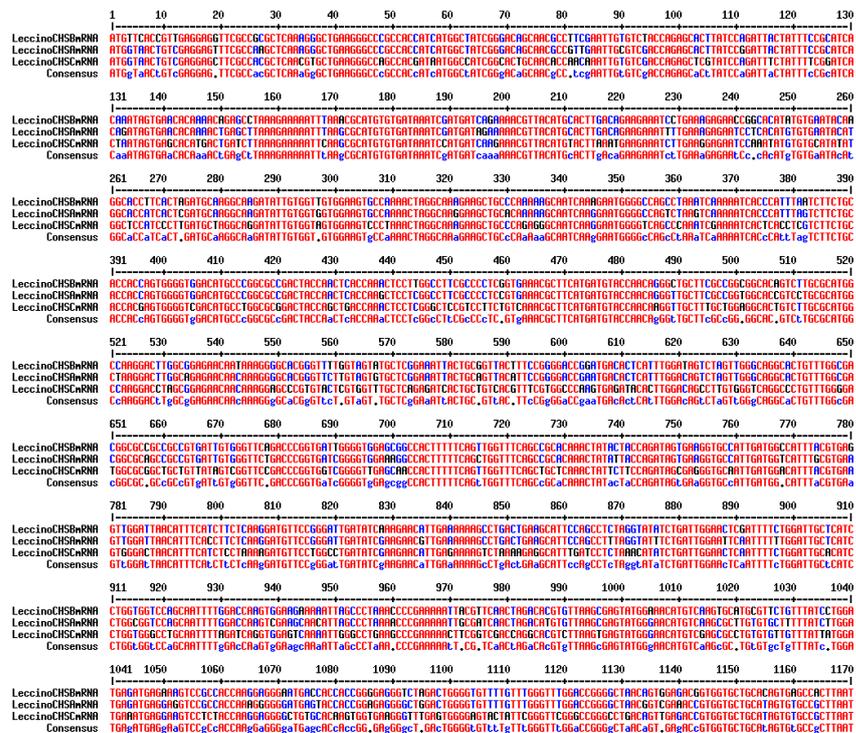


Figura 47. Allineamento delle sequenze di mRNA di *chsA*, *chsB* e *chsC* predette in silico.

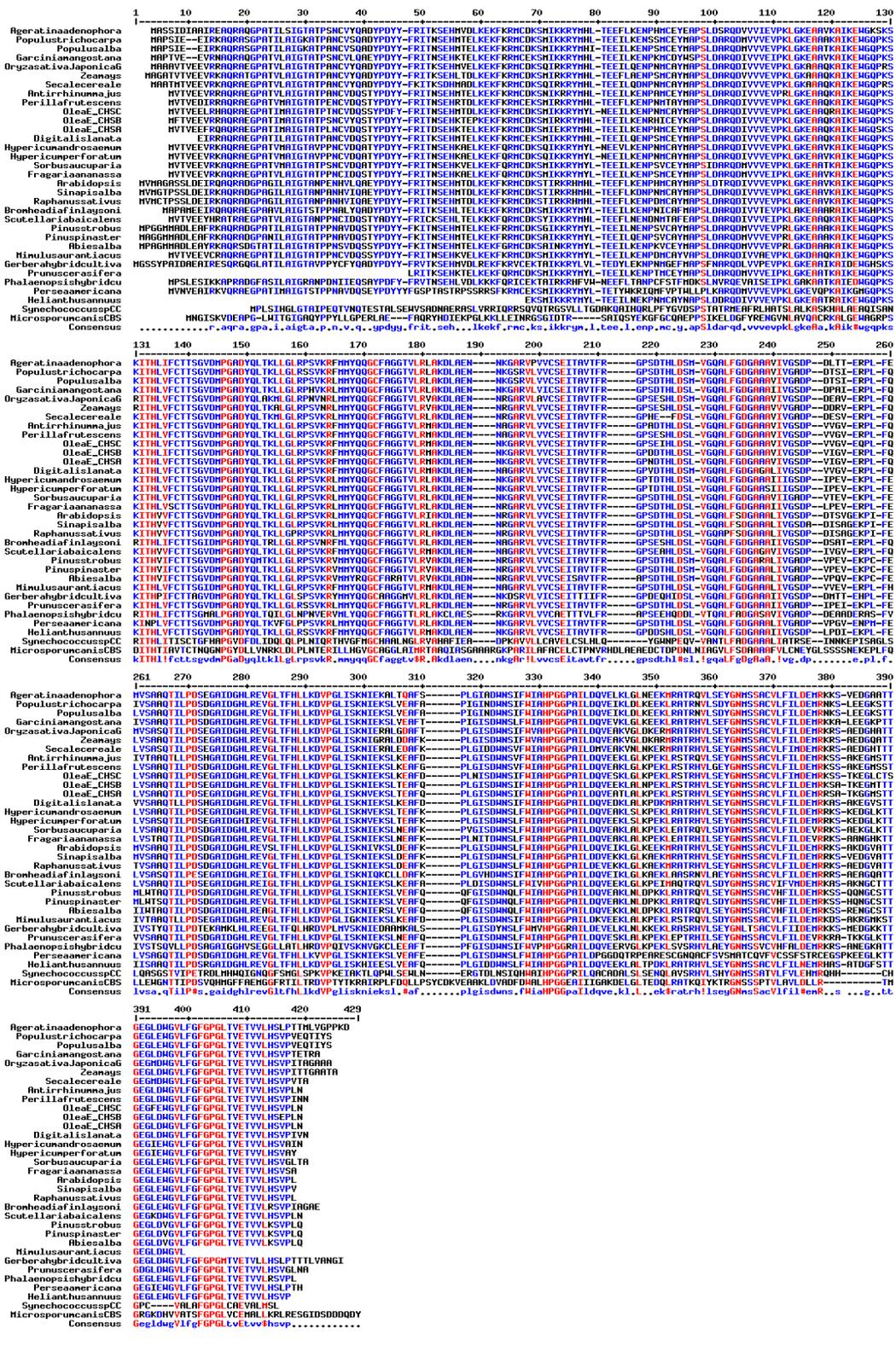


Figura 48. Allineamento delle sequenze proteiche CHS di 33 specie appartenenti a generi diversi.

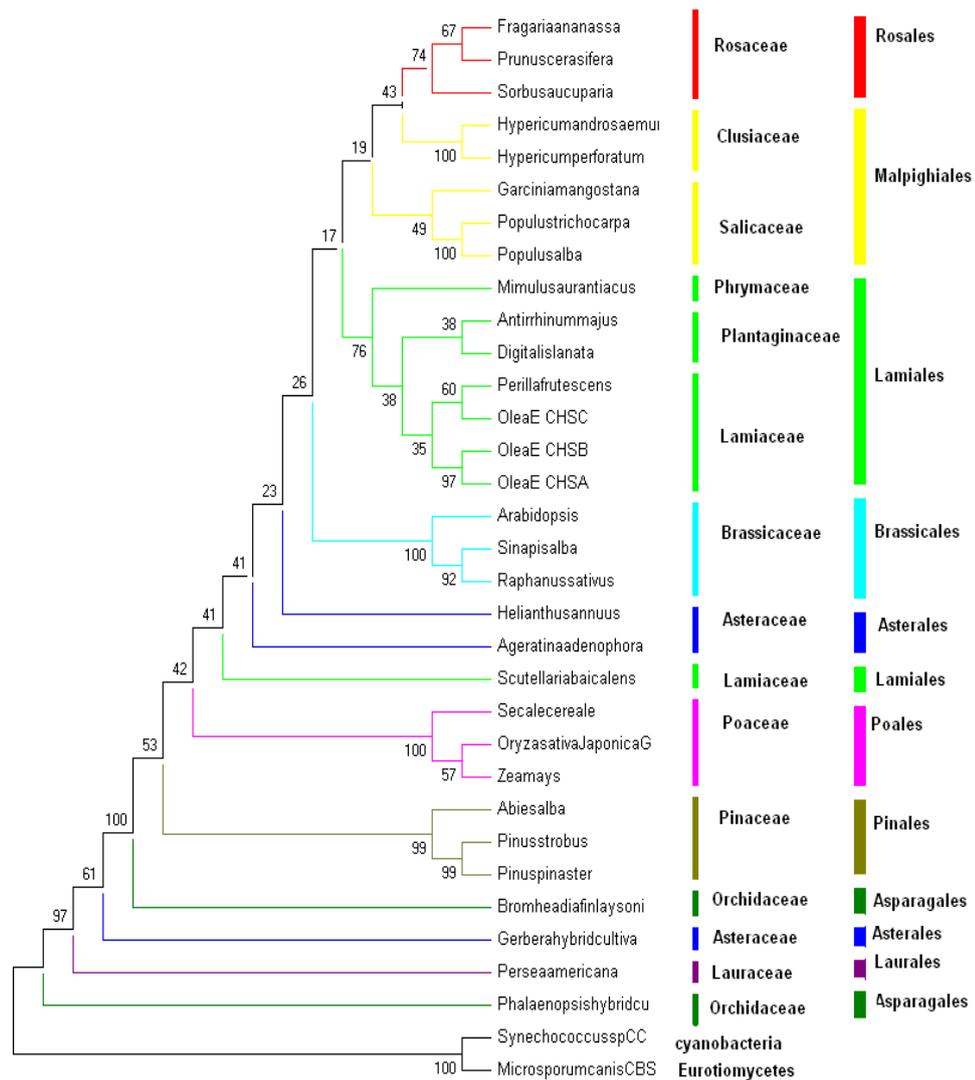


Figura 49. Albero filogenetico UPGMA costruito a partire dall'allineamento delle tre proteine CHS di olivo con le sequenze proteiche della calcon sintasi di 30 specie appartenenti a generi diversi.

Capitolo 4.2 Isolamento dei geni *chsA*, *chsB* e *chsC* in cultivar di olivo della Toscana.

Capitolo 4.2.1 Premessa

Come detto nella Introduzione, in letteratura tra le cultivar di olivo sono documentate numerose sinonimie (la stessa cultivar chiamata in modi differenti) ed omonimie (cultivar differenti che hanno lo stesso nome). Per fare un esempio, la cultivar Santa Caterina è chiamata, secondo il database OleaDB (<http://www.oleadb.it/>) curato dal CNR-IVaSA di Firenze, in almeno sette modi diversi ovvero Katerina, Oliva da Indolcire, Oliva di San Biagio, Oliva di San Giacomo, Oliva Lucchese, Santa Catalina e Santa Caterina.

E' chiaro che in olivo, dove l'origine sconosciuta delle cultivar e la longevità della pianta complicano notevolmente le cose, c'è un assoluto bisogno di una classificazione certa delle cultivar, che abbia come obiettivo principale l'assegnazione non ambigua ed univoca delle accessioni alle cultivar corrispondenti. Fino ad ora risulta ancora estremamente utilizzata la caratterizzazione basata sui caratteri morfologici (Figura 50) anche in conseguenza dell'assenza di disponibilità di marcatori molecolari accreditati.

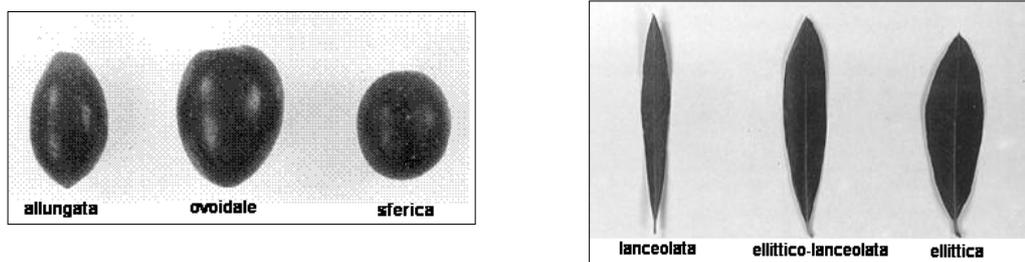


Figura 50. Esempi di caratterizzazione basata su caratteri morfologici di drupa e foglia di olivo.

La caratterizzazione morfologica presenta forti problematiche sia di ambiguità, dovuta all'ampia gamma di caratteri intermedi difficilmente classificabili, sia di condizionamenti dovuti all'influenza dell'ambiente o anche alla soggettività del classificatore. Anche la caratterizzazione chimica, molto usata per la classificazione degli oli in particolare quelli di origine certificata non può da sola risolvere questa problematica.

La certificazione di qualità garantita, sia DOP che IGP, sulla quale si basa una buona parte degli olivicoltori italiani, necessita di una chiara e non ambigua assegnazione delle olive e delle piante dalle quali derivano a cultivar precise presenti nei disciplinari.

D'altra parte la conoscenza delle sequenze dei geni codificanti la Calcone sintasi di olivo, vista l'importanza della proteina nel metabolismo secondario in particolare nella difesa da patogeni o insetti dannosi, potrebbe fornire dati utili sulla differente sensibilità ad esempio alla mosca dell'olivo, una delle principali cause di abbattimento di resa nei raccolti, o con le proprietà organolettiche delle drupe (presenza di antociani, contenuto in polifenoli, ecc.).

La conoscenza dei geni *chs* delle diverse cultivar di olivo potrebbe inoltre permettere di dimostrare correlazioni significative tra eventuali polimorfismi nella sequenza codificante e i dati fenotipici.

Capitolo 4.2.2 La scelta delle cultivar

Per una selezione delle cultivar da analizzare funzionale agli scopi di questa ricerca, abbiamo inizialmente raccolto e schematizzato le caratteristiche principali riguardanti le cultivar toscane, presenti nei database scientifici e nella letteratura, coadiuvati dal supporto tecnico scientifico del Dott. Cimato (IVaISA-CNR).

Questo lavoro preliminare, particolarmente complesso a causa della scarsità di informazioni disponibili e della presenza di documentazioni spesso frammentarie

o poco documentate e talvolta anche contraddittorie, ci ha tuttavia permesso di costruire un database informativo e di immediato e semplice utilizzo, che potrà essere utilizzato anche per il futuro (Figura 51).

Abbiamo selezionato 71 cultivar toscane e abbiamo scelto alcuni parametri potenzialmente correlabili con le funzioni dei flavonoidi come la resistenza al freddo e la resistenza a malattie. Oltre a questo abbiamo inserito parametri correlati con l'”*escape strategy*” alla mosca dell'olivo, quali ad esempio la maturazione precoce.

Abbiamo considerato anche la produttività come un fattore principale di interesse agronomico.

Anche la localizzazione nel territorio può essere considerata un fattore determinante. A questo riguardo sono state scelte cultivar che fossero legate al territorio dove operiamo e presenti nei disciplinari. Nel database abbiamo incluso anche altre caratteristiche legate al principale utilizzo delle drupe, dando priorità alle drupe da olio. Nell'ultima colonna abbiamo inserito l'eventuale rischio di estinzione (Fonte: www.oleadb.com) in quanto si considera di importanza fondamentale la conservazione del germoplasma e della biodiversità della specie. L'incrocio tra tutti questi fattori, inclusa la quantità di componenti minori polari (CMP) presenti nella drupa (A. Cimato, comunicazione personale) ci ha portato a selezionare le 10 cultivar illustrate in Figura 52.

nome	sinonimi	localizzazione	produttività	resistenza al freddo	resistenza a malattie	maturazione	altro	note	rischio	
									estimazione	
ALLORA	Allorina, Allorino.	Lucchesia di Pistoia e Firenze		media	sensibile alla rogna e alla mosca buona tolleranza al cicloconio	media precocita'		ornamentale		▲
AMERICANO		aretino e diffusa solo in alcune zone interne della Toscana	produttività media	resistente ai ricorrenti freddi primaverili	particolare tolleranza al cicloconio	mediamente precoce e contemporanea		esclusivamente alla produzione di olio.		▲
ARANCINO	Pignolo.	sporadico solo nel territorio della provincia pistoiese	Produttiva ma soggetta ad alternanza produttività e ridotta ed alternante		limitata resistenza al cicloconio	media precocita' e piuttosto uniforme		alla raccolta presentano una colorazione rosso vinoso molto accesa.		▲
CILIEGINO				Scarse	Scarse	uniforme, piuttosto lenta, non omogenea		colorazione rosso vinoso dei frutti durante la maturazione		▲
COLOMBINO	Felcisaio, Olivastro colombino.	nord di Lucca	produttiva ma alternante	Resiste bene al freddo ed alla mosca	sensibile a fumaggine e rogna	medio-tardiva e contemporanea				▲
CORREGGIOLO	Pignatello, Fiorentino, Pendagliolo, Frantoiano, Gentile, Grossaio	diffusa in tutta la Toscana Presente nelle regioni del centro Italia	resa in olio è elevata	scarsa tolleranza al freddo ed ai venti	Elevata e' la sensibilità alla rogna, media al cicloconio			prodotto è ritenuto di ottima qualità		
CORREGGIOLO DI PALLESE		colline dei monti pisani	colline dei monti pisani	scarsa tolleranza	scarsa tolleranza			ottimo qualità		
CUORICINO		costa livornese	oliveti della costa livornese		scarsa resistenza a tignola, fumaggine, cocciniglia e rogna			elevata attitudine naturale delle talee alla rizogenesi		▲
DA CUCCARE	Oliu di Segromigno, Cuccola, Cucca.	collina lucchese	produzione risulta piuttosto elevata	buona tolleranza al freddo						▲
FRANTOIANO DI MONTEMURLO	Barilsia, Frantoio di Montemurlo	sporadicamente in provincia di Pisa e soprattutto nel comune di Terricciola	La produttività è buona ma piuttosto alternante	limitata tolleranza al freddo	elevata sensibilità alla rogna e cicloconio	in epoca tardiva ed in modo scalare		eccellente qualità		
FRANTOIO	Oliu lunga, Correggiolo, Frantoiano, Comune, Solcario, Nostrato, Raggio, Rajo, Razzo, Raggiolo, Crognolo, Razzo, Gentile, Bressa Fina, Infrantoio, Laurino, Pendagliolo, Pignatello, Stringona	in tutta Italia ed in quasi tutte le zone olivicole del mondo	elevata e costante	scarsa tolleranza alle avversità climatiche	elevata sensibilità al cicloconio alla rogna ed alla mosca	scalare e tardiva		L'olio è molto pregiato, fine, sapido, aromatico		
GRAPPOLO	Oliu a grappolo.	comprendorio pistoiese collina fiorentina	produttività e' elevata e costante	notevole resistenza al freddo	resistente al cicloconio					▲
GREMIGNA TONDA		zona al confine tra le province di Pisa e Livorno	produttività e' buona e piuttosto costante		sensibile alla mosca, alla rogna ed alla margaronia	medio-tardiva		colorazione rossa molto intensa		▲

GREMIGNO DI FAUGLIA		sud della provincia di Pisa	produzione e' alquanto elevata e costante	ha resistito alle basse temperature del gennaio 1985	elevata tolleranza alle malattie			elevata qualita'	⚠
GREMIGNOLO	Mignolo	comprensorio livornese, fiorentino e pistoiese	produttivita' e' alternante	rusticita' agli agenti climatici	scarsa resistenza alle principali avversita' parassitarie			Scarsa in genere e' la resa in olio.	
GREMIGNOLO DI BOLGHERI		Bolgheri comprensorio fiorentino	produzione e' media ed alternante	Tollera poco quasi tutte le principali avversita' dell'olivo	Tollera poco quasi tutte le principali avversita' dell'olivo		resistenza ai venti salsi marini	utilizzata anche come pianta ornamentale	⚠
GREMIGNO DI MONTECATINI	Gremigno	sporadicamente negli oliveti a sud di Pisa	produttivita' e' elevata e costante	tollera freddo	Tollera bene cicloconio suscettibilita' a fumaggine, cocciniglia e rogna	maturano tardivamente			
GROSSAIO		Monti Pisani collina fiorentina			elevata sensibilita' a cicloconio (1) e margaronia.	scalare e tardiva		fruttato e delicato	⚠
GROSSOLANA		provincia di Livorno	produttivita' risulta alternante			tardivamente		olive da tavola L'olio e' di buona qualita'	⚠
LARCIANESE		Lamporecchio		scarsamente tollerante		tardiva ed avviene in modo scalare			
LASTRINO		Casale Marittimo	produttivita' e' alternante	limitata tolleranza	limitata tolleranza	lento e scalare			⚠
LAZZERO	Lazzero grosso, Lazzero tipo, Lazzera.	Valle del fiume Cecina	elevata anche se piuttosto alternante	discreta resistenza	discreta resistenza				⚠
LAZZERO DELLE GUADALUPE		Volterra		discreta resistenza al freddo	discreta resistenza all'occhio di pavone	media precocita'		essiccati ed utilizzati per il consumo diretto	⚠
LAZZERO DI PRATA		Prata Savereto	alternante e di media entita'	buona tolleranza al freddo ed agli sbalzi termici	elevata la sensibilita' a cicloconio e alla mosca.			"fine" e di buona qualita'	⚠
LECCINO	Leccio, Premice, Silvestrone	tutte le zone olivicole italiane e nei principali areali del mondo		tolleranza alle avversita' climatiche	tolleranza ad alcune patologie (rogna, cicloconio e carie) sensibilita' alla fumaggine	precoce ed uniforme		utilizzata sia per l'estrazione dell'olio che come olive "nere" da menza	
LECCIO DEL CORNO		San Casciano val di Pesa		particolare resistenza alle basse temperature dell'inverno 1929	notevole tolleranza al cicloconio				
LECCIONE		provincia di Pistoia, sporadico	Buona e' risultata in genere la produttivita' anche se l'alternanza e' piuttosto elevata	buona resistenza alle basse temperature	resistente cicloconio				⚠
MADONNA DELL'IMPRUNETA		solo in provincia di Firenze	produttivita' e' scarsa ed alternante	ottima tolleranza	tolleranza agli agenti parassitari	maturazione piuttosto uniforme			
MADREMIGNOLA	Madremignolo	collina pisana	produttivita' e' buona e costante	buona tolleranza ai ristagni di aria umida molto sensibile a freddo				pianta ornamentale ma anche olio apprezzato	⚠
MANSINO	Manzino.	collina fiorentina	produttivita' e' buona ma alternante	buona tolleranza alle basse temperature	discreta tolleranza alle piu' frequenti fitopatie	maturano piuttosto tardivamente			⚠

MAREMMANO	Americano, Leccio maremmano, Olivastro maremmano	sporadica diffusione nei comprensori di Firenze, Siena, Arezzo e Grosseto	resa al frantoio e' discreta (1), ma la qualita' dell'olio non e' molto apprezzata		sensibilita' alla mosca, tolleranza cicloconio					▲
MARZIO		provincia di Pistoia	elevata produttivita'	discreta resistenza al freddo		media precocita' ma scalare				▲
MAURINO	Razzola, Maurino lucchese.	tutta la Toscana ed anche in altre zone dell'Italia centrale	buona e' la resa in olio	discreta resistenza al freddo	resistenza al cicloconio	medio-precocita' e contemporanea				
MELAILO		non ha trovato larga diffusione	produttivita' piuttosto scarsa ed alternante	abbastanza resistente	abbastanza resistente				L'olio non e' molto apprezzato dagli agricoltori	▲
MIGNOLO	Mignola, Minuto, Prugnolo.	comprensorio fiorentino si e' diffuso in modo sporadico in quasi tutte le province toscane	produttivita' non e' elevata.	non molto resistente	discreta tolleranza al cicloconio					▲
MIGNOLO CERRETANO	Mignola domestica, Cerretano	confine tra le province di Firenze e Pistoia.	Produzione non elevata	molto sensibile alle basse temperature	buona tolleranza all'asfissia radicale, margaronia, mosca e cicloconio				ritenuto di pregio perche' verde, fruttato e fine contenuti elevati	▲
MORAILO	Ruzzolino, Morinello, Morellino, Morello, Oriolo, Corniolo, Cimignolo, Fosco, Nostrale, Assisano, Aserina, Bucino, Carboncella, Migno, Morella, Morichello, Morina, Muragliolo, Neriolo, Nerella, Nerina, Neriolo, Ogliolo, Oliva nera, Oliva tonda, Petrosello, Tondello, Tondolina, Tondorina.	larga diffusione in Italia e negli altri Paesi del Mediterraneo	all'elevata resa in olio delle drupe ed alle particolari capacita' produttive	molto sensibile	molto sensibile al cicloconio, fumaggine	media precocita'	resistenza ai venti anche salsi			
MORCAIO		province di Pisa e Livorno	buona e costante	elevata resistenza al freddo	discreta tolleranza agli agenti fitopatogeni				ricco di "morchia"	▲
MORCHIAIO	Morcaio, Morchiolo, Fecciaio, Felcisio, Propolo, Bozzolo, Colombino, Morcanel, Maschio, Morchiaccio.	tutta la Toscana	limitata, e' anche alternante	tolleranza al freddo	sensibilita' al cicloconio	scalare e tardiva				▲
MORCHIONE		colline fiorentine che negli oliveti della provincia di Arezzo	elevata solo in alcune annate	tollerante ai frequenti freddi che colpiscono la zona aretina ed anche alla gelata del gennaio 1985	molto sensibile agli attacchi della mosca					▲
MORCONE		Bolgheri		Tollera bene le basse temperature	sensibilita' alle diverse malattie dell'olio ed in particolare alla mosca.					▲
MORELLO A PUNTA		zona pisana	produzione risulta elevata solo nelle annate particolarmente favorevoli	discreta tolleranza ai freddi primaverili						
MORTELLINO	Mortellino, Mortatele, Maschion, Olivastro mortellino, Olivastro maschio.	province di Lucca, Massa e Pisa	frutti, molto piccoli, non consentono di ottenere elevate quantita' di prodotto	Tollera discretamente il freddo,	sensibilita' a rogna, cicloconio, mosca e fumaggine.	medio-tardiva ed avviene piuttosto lentamente			funzione di impollinatori.	
OLIVO DEL PALONE		Bolgheri	produzione e' elevata e costante	buona tolleranza alle piu' frequenti avversita' climatiche	buona tolleranza				Non ha trovato il favore degli agricoltori	
OLIVASTRA DI POPULONIA		comprensorio sud-occidentale di Livorno	piuttosto scarsa	scarsa tolleranza	scarsa tolleranza					▲

OLIVA STRA DI SUVERETO		Suvereto	buona.	buona tolleranza	buona tolleranza	medio-tardiva e piuttosto lenta		resa delle drupe non e' elevata ma l'olio e' molto apprezzato	⚠
OLIVA STRA SEGGIANESE	Seggianese, Seggianina, Olivastra di Montalcino, Olivastra di Seggiano, Olivo dell'Amiata.	Monte Amiata province di Grosseto e Siena	produttivita' e' buona anche se alternante	elevata resistenza al freddo	discreta al cicloconio e rognia			caratteristiche organolettiche peculiari	
OLIVO BUFALO	Bufalo, Bufale	sporadicamente nel comprensorio di origine tra le province di Siena e Pisa.	buona resa in olio e sono molto apprezzati come olive da mensa	buona tolleranza	buona tolleranza			sia per l'estrazione dell'olio che per il consumo diretto	⚠
OLIVO DI SAN LORENZO		Suvereto	elevata ma piuttosto alternante		discreta tolleranza alle piu' frequenti avversita' parassitarie che colpiscono l'olivo nella zona			particolare colorazione rosso-vinosa scura	⚠
PENDAGLIOLO		zona sud della provincia di Pisa	produzione e' elevata e costante		discreta (tardiva		L'olio e' ritenuto pregiato	⚠
PENDOLINO	Piagnente, Msurino fiorentino.	tutte le zone olivicole della Toscana	mediamente elevata		sensibile a fletotribo, cicloconio e fumaggine	media precocita' e contemporanea		pregevoli caratteristiche ornamentali apprezzato per le caratteristiche organolettiche	
PESCIATINO		?				tardiva			
PIANGENTE		zone olivicole toscane	l'elevata produttivita', non sempre ha confermato queste caratteristiche	discreta resistenza	discreta resistenza	tardiva		R. in particolare ai venti salsi di mare	⚠
PIGNOLO	Pignola.	fascia settentrionale della provincia di Pistoia	elevata anche se non molto costante			tardivamente		L'olio che si ottiene e' considerato di buona qualita'	⚠
PITURZELLO		zona meridionale della provincia di Siena	buona e molto costante	piu' tollerante del "Moraiolo" al freddo					
PUNTERUOLO	Trillo, Occhino.	zone settentrionali della Toscana		in presenza di condizioni ambientali fresche.	buona resistenza del "Punteruolo" al cicloconio				⚠
QUERCETANO	Quercetano, Olivastra quercetano, Trillo, Minutaio.	Versilia,	produzione abbastanza elevata		buona tolleranza alla mosca ed all'occhio di pavone mentre e' molto sensibile alla rognia				⚠
RAMA PENDULA		zone collinari della provincia pistoiese	anni alterni e la resa in olio non e' elevata	Tollerante al freddo	Tollerante al cicloconio (1), alla mosca ed alla cocciniglia elevata sensibilita' al mal del piombo (cercozopora).	elevata sensibilita' alla siccita'			⚠
RAZZAIO		comprensori fiorentino e pistoiese	produttivita' e' in genere piuttosto scarsa	scarsamente tollerante	scarsamente tollerante				
RAZZO		Livorno, Pisa, Arezzo e Pistoia	molto elevata e costante	discreta tolleranza	discreta tolleranza mentre e' molto sensibile a cicloconio, rognia e	tardiva e scalare	venti marini	L'olio ha caratteristiche qualitative di pregio perche' fruttato ed elegante	
ROSINO		Bolgheri	scarsa	scarsa tolleranza	scarsa tolleranza			mpollinatrice	⚠
ROSSELLINO	Rossellina.	comprensorio fiorentino	produttivita' risulta buona		scarsa tolleranza al cicloconio	media precocita'		colorazione rosso-vinosa dei frutti durante la maturazione	⚠
ROSSELLINO CERRETANO	Rossellino di Cerreto Guidi	Cerreto Guidi		Buona	Buona discreta sensibilita' al cicloconio	medio-precoco			⚠

ROSSELLO		provincia di Pistoia	scarsa	sensibilita'	sensibilita'			le olive rimangono sulla pianta per molto tempo con una caratteristica colorazione rosso vinoso	⚠
SAN FRANCESCO		province di Firenze e Pistoia	mediamente alternante	buona resistenza al freddo	buona resistenza ed al cicloconio			produzione di olive da tavola	⚠
SANTA CATERINA	Oliva di S.Biagio, Oliva di San Giacomo, Oliva lucchese.	Lucca, Firenze e nel grossetano zone dell'Italia centrale	piuttosto elevata e costante	Buona	res limitata al cicloconio			la cultivar da tavola toscana piu' importante	⚠
SCARLINESE		Scarolino	abbastanza elevata.	buona resistenza	buona resistenza	mediamente precoce			
SELVATICA		province di Livorno e Pisa.	resa non e' elevata ma l'olio e' ritenuto di eccellente qualita'	discreta tolleranza	discreta tolleranza in particolare alla mosca	tardiva e scalare (1)		rustica con aspetto caratteristico	⚠
TONDELLO		Bolgheri	analoghe al "Morsuolo"	scarsamente resistente	scarsamente resistente	analoghe al "Morsuolo"			⚠
TRILLO		Monti Pisani	resa in olio risulta generalmente limitata	tollera bene freddo	tollera bene cicloconio, fumaggine				⚠

Figura 51. Database contenente le principali cultivar toscane. Per ogni varietà sono indicati: nome, sinonimie, localizzazione, produttività, resistenza al freddo, resistenza a malattie, maturazione, altre notizie rilevanti e particolarità, eventuale rischio di estinzione.

CULTIVAR TOSCANE DI OLIVO	GRADO DI PIGMENTAZIONE DELLE DRUPE	Tot.CMP
ALBATRO		241
ALLORA		113
COLOMBINO		900
LECCINO		
MADREMIGNOLA		420
MELAIOLO		130
MIGNOLO CERRETANO		642
MORAIOLO		
ROSSELLO		350
SAN FRANCESCO		234

Figura 52: Cultivar toscane di olivo selezionate per il nostro lavoro. Lo schema comprende la indicazione dei componenti minori polari (mg/Kg). Per le cultivar Leccino e Moraiolo le analisi dei CMP non sono disponibili.

Capitolo 4.2.3 Isolamento delle sequenze *chsA*, *chsB* e *chsC* mediante amplificazione PCR.

Da tessuto fogliare delle cultivar selezionate è stato innanzitutto estratto il DNA utilizzando il metodo Doyle and Doyle (1990) descritto in Materiali e Metodi. Per amplificare le sequenze complete dei geni *chsA*, *chsB* e *chsC* abbiamo allestito una serie di reazioni PCR utilizzando coppie di primer specifiche per ogni gene. I prodotti di amplificazione sono stati visualizzati mediante corsa elettroforetica su gel di agarosio, e gli ampliconi di interesse sono stati purificati utilizzando il kit Nucleospin extract II (Macherey-Nagel) e clonati in vettori Topo TA (Invitrogen) in cellule competenti di *E.coli*. Per ogni reazione di clonaggio è stato necessario effettuare uno screening molto ampio delle colonie (>100) in quanto l'inserimento in un plasmide di 3900 bp di un amplicone di dimensioni rilevanti (circa 2400 bp) risulta particolarmente difficoltoso. I frammenti clonati sono stati sequenziati con il sequenziatore ABI PRISM 310 (Applied Biosystem), utilizzando il kit di sequenza BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems), presso il centro servizi CIBIACI (Centro Intredipartimentale Biotecnologie Interesse Agrario Chimico e Industriale) dell'Università di Firenze.

Per sequenziare l'intero frammento (Figura 53) sono stati utilizzati due primer intermedi, localizzati approssimativamente in corrispondenza della metà dell'esone 2 a circa 1200 bp dall'inizio della sequenza per *chsA* e *chsB* ed all'inizio dell'esone 2 nel caso di *chsC*.

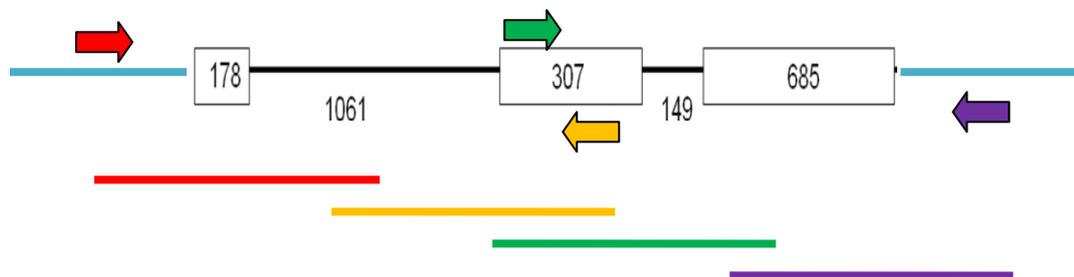


Figura 53. Primer utilizzati per il sequenziamento dei geni CHS nelle cultivar toscane.

Come schematizzato in figura 53, per ottenere l'intera sequenza di ciascun gene lunga 2400 bp sono stati necessari almeno 4 primer, due dei quali, M13FW e M13REV, universali e forniti con il kit, situati sul plasmide, e due, EX2FW e EX2REV, disegnati direttamente sulle sequenze d'interesse. Per ottenere il costrutto finale comprendente l'intera sequenza del gene, le sequenze ottenute devono avere le estremità sovrapposte per consentire il corretto allineamento. Tutti i frammenti clonati sono stati sequenziati in triplice copia partendo da colonie indipendenti.

Capitolo 4.2.3.1. Isolamento di *chsA* in cultivar toscane.

In Figura 54 è illustrato l'allineamento effettuato con il software Multalin delle sequenza *chsA* nelle 10 cultivar toscane selezionate. Nell'allineamento di Figura 54 è importante indicare la presenza di due inserzioni cultivar-specifiche, una delle quali è in corrispondenza del secondo introne ed è una ripetizione microsatellite (ATT)₄ che che nella cultivar Mignolo cerretano si ripete 14 volte. L'altra nella cultivar Leccino, in corrispondenza del sito 1250 che è costituita da una ripetizione della sequenza CAAAGGCTTATA.

E' possibile inoltre rilevare la presenza dell'inserzione di una Citosina in corrispondenza del sito 710, esclusiva della cultivar Moraiolo e di una delezione di una Timina nel sito 253, nelle cultivar Rossello, San Francesco e Leccino.

L'analisi di diversità nucleotidica effettuata con il software DNAsp è riportata in Figura 55. Questa analisi ha rilevato la presenza di 2227 siti che, escludendo quelli che presentano gaps, si sono ridotti a 2201 totali. Di questi 2150 sono risultati monomorfici (98%) e 51 (23%) polimorfici. In figura 56 è illustrato il grafico che riporta la diversità nucleotidica Pi in funzione della posizione della

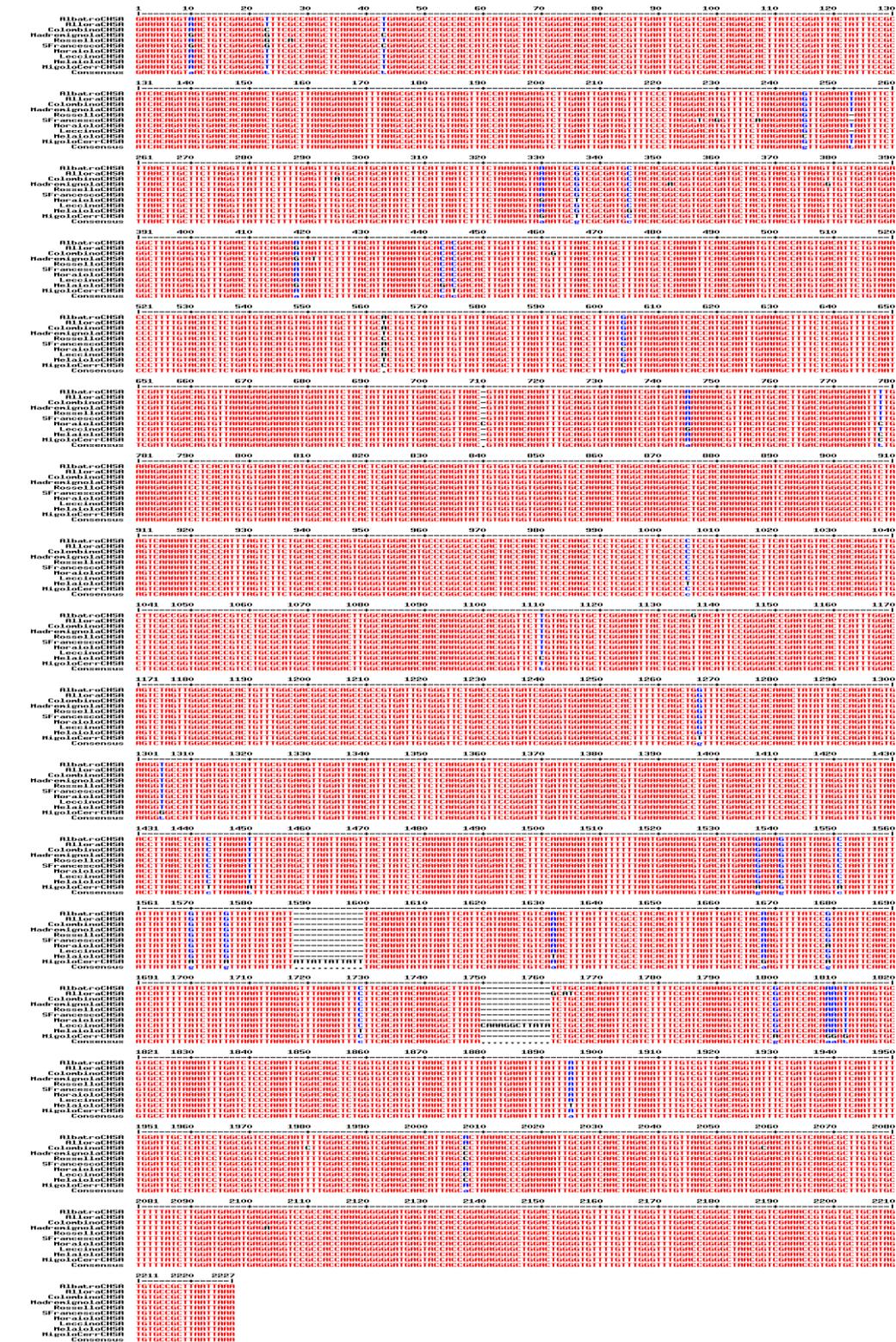


Figura 54. Allineamento delle sequenze del gene *chsA* nelle 10 cultivar.

```

Number of sequences: 10   Number of sequences used: 10
Selected region: 1-2227   Number of sites: 2227
Total number of sites (excluding sites with gaps / missing data): 2201

Number of polymorphic (segregating) sites, S: 51
Total number of mutations, Eta: 53

Number of Haplotypes, h: 10
Haplotype (gene) diversity, Hd: 1,000
Variance of Haplotype diversity: 0,00200
Standard Deviation of Haplotype diversity: 0,045

Nucleotide diversity, Pi: 0,00589
Theta (per site) from Eta: 0,00851

Theta (per site) from S, Theta-W: 0,00819
Variance of theta (no recombination): 0,0000119
Standard deviation of theta (no recombination): 0,00345
Variance of theta (free recombination): 0,0000013
Standard deviation of theta (free recombination): 0,00115

Finite Sites Model
Theta (per site) from Pi: 0,00593
Theta (per site) from S: 0,00832
Theta (per site) from Eta: 0,00859

Average number of nucleotide differences, k: 12,956
Stochastic variance of k (no recombination), Vst(k): 32,564
Sampling variance of k (no recombination), Vs(k): 8,178
Total variance of k (no recombination), V(k): 40,742
Stochastic variance of k (free recombination), Vst(k): 4,319
Sampling variance of k (free recombination), Vs(k): 0,960
Total variance of k (free recombination), V(k): 5,278

Theta (per sequence) from S, Theta-W: 18,028
Variance of theta (no recombination): 57,784
Variance of theta (free recombination): 6,373

```

Figura 55. Analisi della diversità nucleotidica delle 10 sequenze di *chsA*

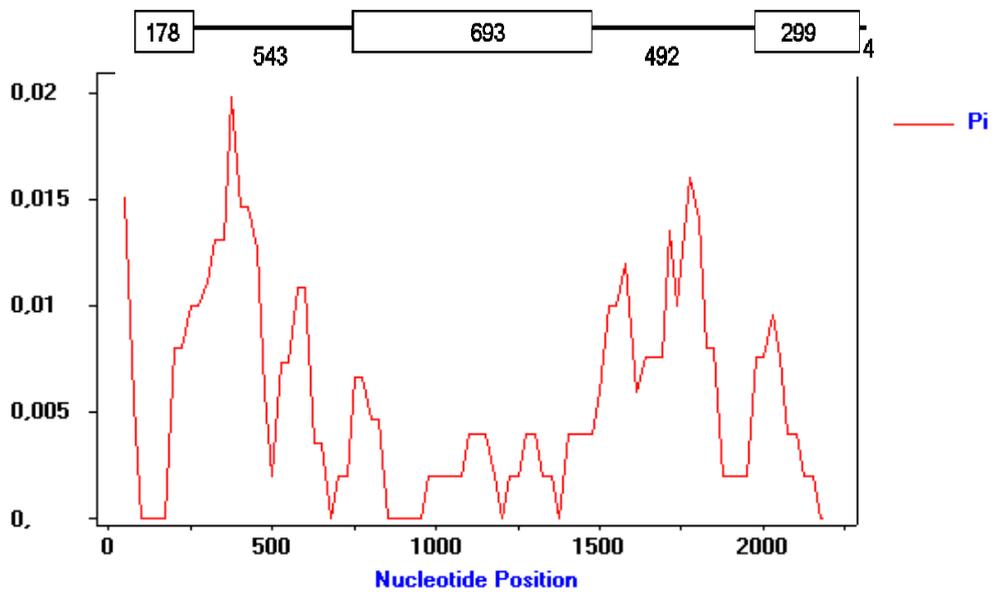


Figura 56. Diversità nucleotidica Pi in funzione della posizione del nucleotide. In alto la struttura del gene

posizione all'interno del gene. Si vede chiaramente come in corrispondenza degli introni ci sia una maggiore variabilità tra le sequenze analizzate, come atteso. In corrispondenza dell'esone 2 invece, che ospita la maggioranza dei nucleotidi che codificano per gli amminoacidi implicati in funzioni cruciali nella proteina CHS, troviamo una variabilità molto bassa. Anche le giunzioni tra introni ed esoni rappresentano zone con bassa variabilità, che diventa prossima allo zero all'inizio della sequenza codificante (ATG) e alla fine di questa.

L'allineamento delle sequenze di mRNA dedotte *in silico* è mostrato in Figura 57. Su un totale di 1170 siti, 1155 sono monomorfici (99%) e 15 polimorfici (1%). In totale l'allineamento mette in evidenza 16 mutazioni, delle quali 10 sinonime e 6 non sinonime. Per quanto riguarda le sequenze proteiche anche esse dedotte, (Figura 58) sono presenti 5 cambiamenti, che riguardano:

- ❖ il residuo 7 che risulta il più variabile e presenta gli amminoacidi Fenilalanina (6 cultivar), Leucina (solo la cultivar Colombino) e Valina (3 cultivar);
- ❖ il residuo 66 con la cultivar Leccino che presenta Glu al posto di Lys;
- ❖ il residuo 196 che varia in Albatro (Glicina al posto di Valina);
- ❖ il residuo in posizione 240 nel quale Fenilalanina sostituisce Valina nella cultivar Mignolo Cerretano.

Nessuno di questi cambiamenti coinvolge residui documentati come altamente conservati e implicati in funzioni cruciali per l'enzima.

In figura 59 è illustrato il profilo dei marcatori SNPs identificati nelle sequenze di *chsA* che caratterizza le cultivar.

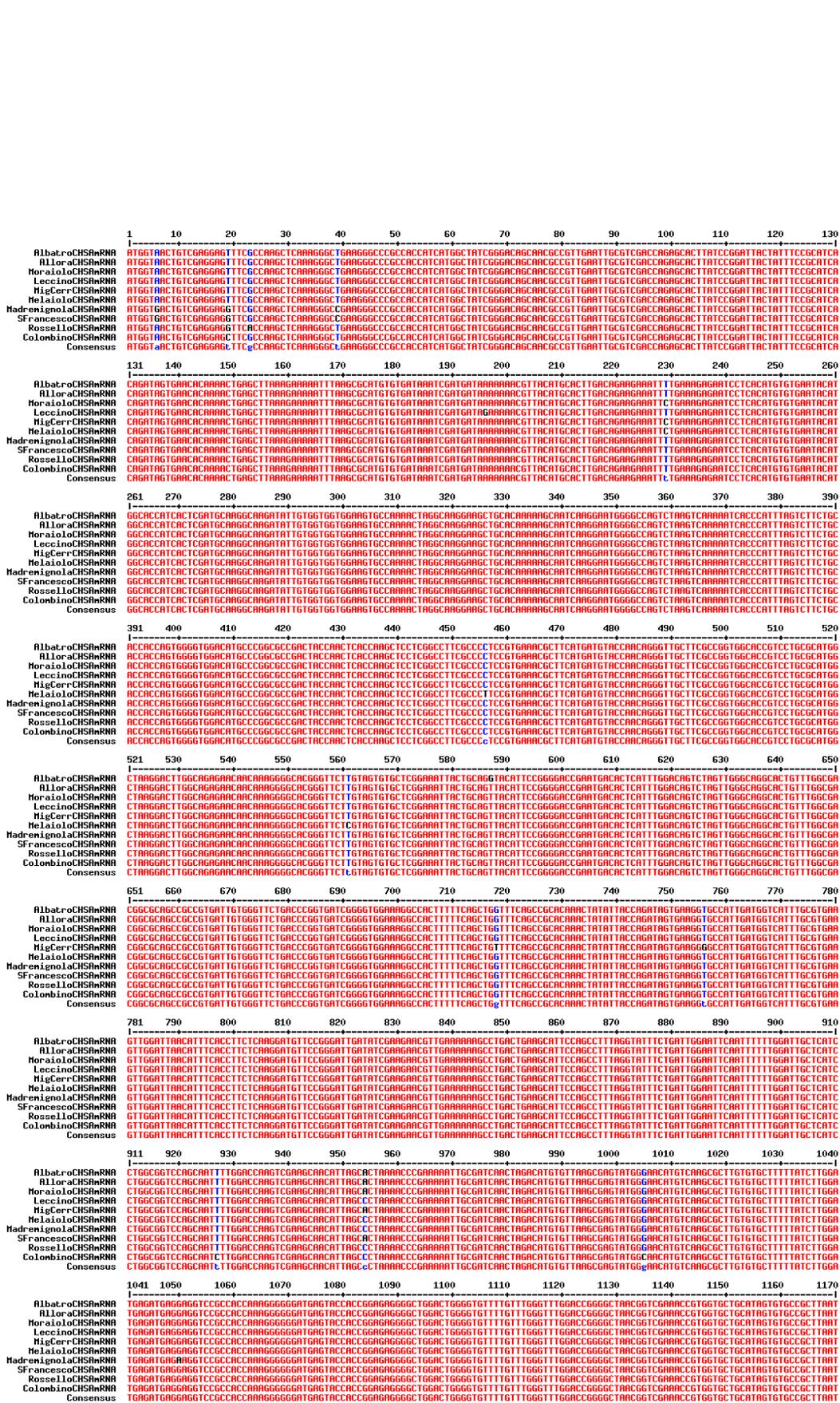


Figura 57. Allineamento delle sequenze di mRNA dedotte dalle sequenze geniche di *chsA* nelle cultivar toscane selezionate

Capitolo 4.2.3.2. Isolamento di *chsB* in cultivar toscane.

La sequenza del gene *chsB* è stata amplificata clonata e sequenziata nelle 10 cultivar toscane selezionate.

In totale sono presenti 2157 siti; se escludiamo le delezioni (25) sono computabili 2132 siti. Tra questi risultano monomorfici 2046 siti pari al 96% del totale e polimorfici 86 siti, dei quali 40 sono variazioni di singola base.

Come si può osservare dal grafico in Figura 60 che riporta la variabilità nucleotidica P_i in funzione della posizione nella sequenza, in corrispondenza dell'inizio della sequenza codificante, la variabilità nucleotidica è minore, come dimostrato dai valori di P_i che indica il numero medio delle differenze per sito tra le sequenze analizzate. A livello della base 750, in corrispondenza della giunzione tra l'introne 1e l'esone 2, in una zona che in letteratura è riconosciuta come altamente conservata, sorprendentemente si può notare un notevole aumento di diversità. Molto variabile appare anche la prima regione dell'esone 2.

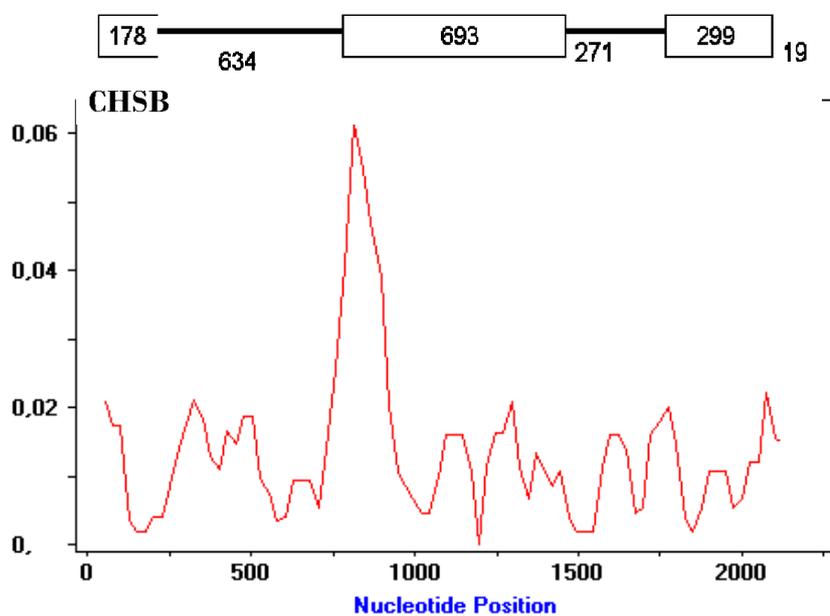


Figura 60. Variabilità nucleotidica P_i delle sequenze di *chsB* nelle cultivar toscane selezionate.

Si noti in figura 60 come in questo caso la variabilità, particolarmente quella riscontrata in corrispondenza del picco localizzato all'altezza della giunzione introne1-esone2, sia sensibilmente più alta rispetto ai valori ottenuti in *chsA*

In Figura 61 è riportato l'allineamento tra le sequenze di *chsB* amplificate nelle 10 cultivar di olivo toscane.

In Figura 62 possiamo osservare una Tabella riassuntiva che riporta il livello di polimorfismo presente nella sequenza *chsB* delle cultivar toscane. Dal confronto con lo stesso tipo di analisi per la sequenza *chsA*, appare evidente, come nella analisi precedente, una maggiore variabilità del gene *chsB*.

<p>Number of sequences: 10 Number of sequences used: 10 Selected region: 1-2157 Number of sites: 2157 Total number of sites (excluding sites with gaps / missing data): 2132</p> <p>Number of polymorphic (segregating) sites, S: 86 Total number of mutations, Eta: 87</p> <p>Number of Haplotypes, h: 10 Haplotype (gene) diversity, Hd: 1,000 Variance of Haplotype diversity: 0,00200 Standard Deviation of Haplotype diversity: 0,045</p> <p>Nucleotide diversity, Pi: 0,01352 Theta (per site) from Eta: 0,01442</p> <p>Theta (per site) from S, Theta-W: 0,01426 Variance of theta (no recombination): 0,0000348 Standard deviation of theta (no recombination): 0,00590 Variance of theta (free recombination): 0,0000024 Standard deviation of theta (free recombination): 0,00154</p> <p>Finite Sites Model Theta (per site) from Pi: 0,01377 Theta (per site) from S: 0,01465 Theta (per site) from Eta: 0,01466</p> <p>Average number of nucleotide differences, k: 28,822 Stochastic variance of k (no recombination), Vst(k): 151,901 Sampling variance of k (no recombination), Vs(k): 38,499 Total variance of k (no recombination), V(k): 190,400 Stochastic variance of k (free recombination), Vst(k): 9,607 Sampling variance of k (free recombination), Vs(k): 2,135 Total variance of k (free recombination), V(k): 11,742</p> <p>Theta (per sequence) from S, Theta-W: 30,400 Variance of theta (no recombination): 158,126 Variance of theta (free recombination): 10,746</p>

Figura 62. Analisi della diversità nucleotidica delle 10 sequenze di *chsB*

A conferma di quanto detto il grafico in Figura 63 che rappresenta il numero di siti segreganti in funzione della posizione dei siti stessi, computato a partire dall'allineamento dalle sequenze di mRNA di *chsB* predette *in silico* (Figura 64) ci evidenzia la presenza di un picco di diversità nucleotidica situato attorno al sito 250, in prossimità dell'inizio dell'introne 2.

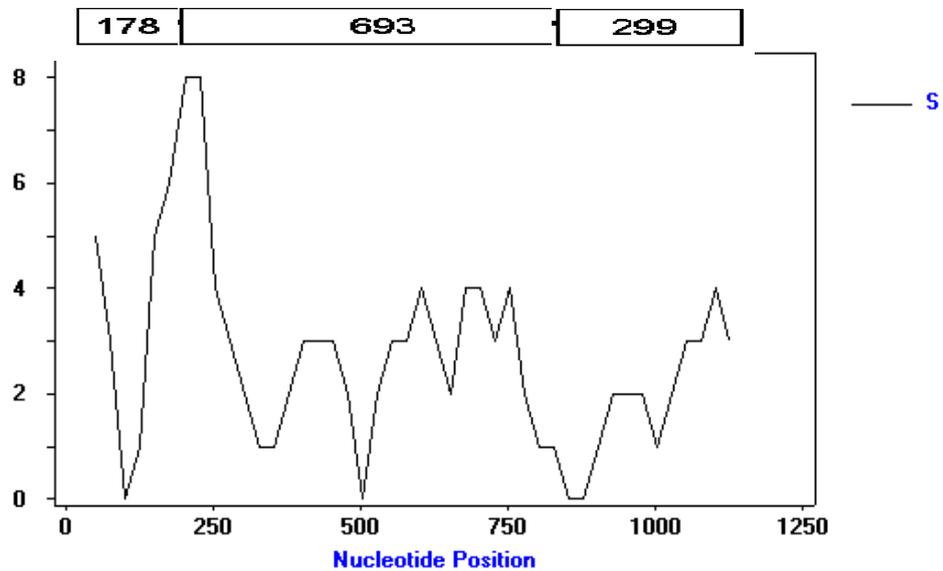


Figura 63. Numero di siti segreganti delle sequenze di mRNA di *chsB* in funzione della posizione. In alto la struttura dell'mRNA.

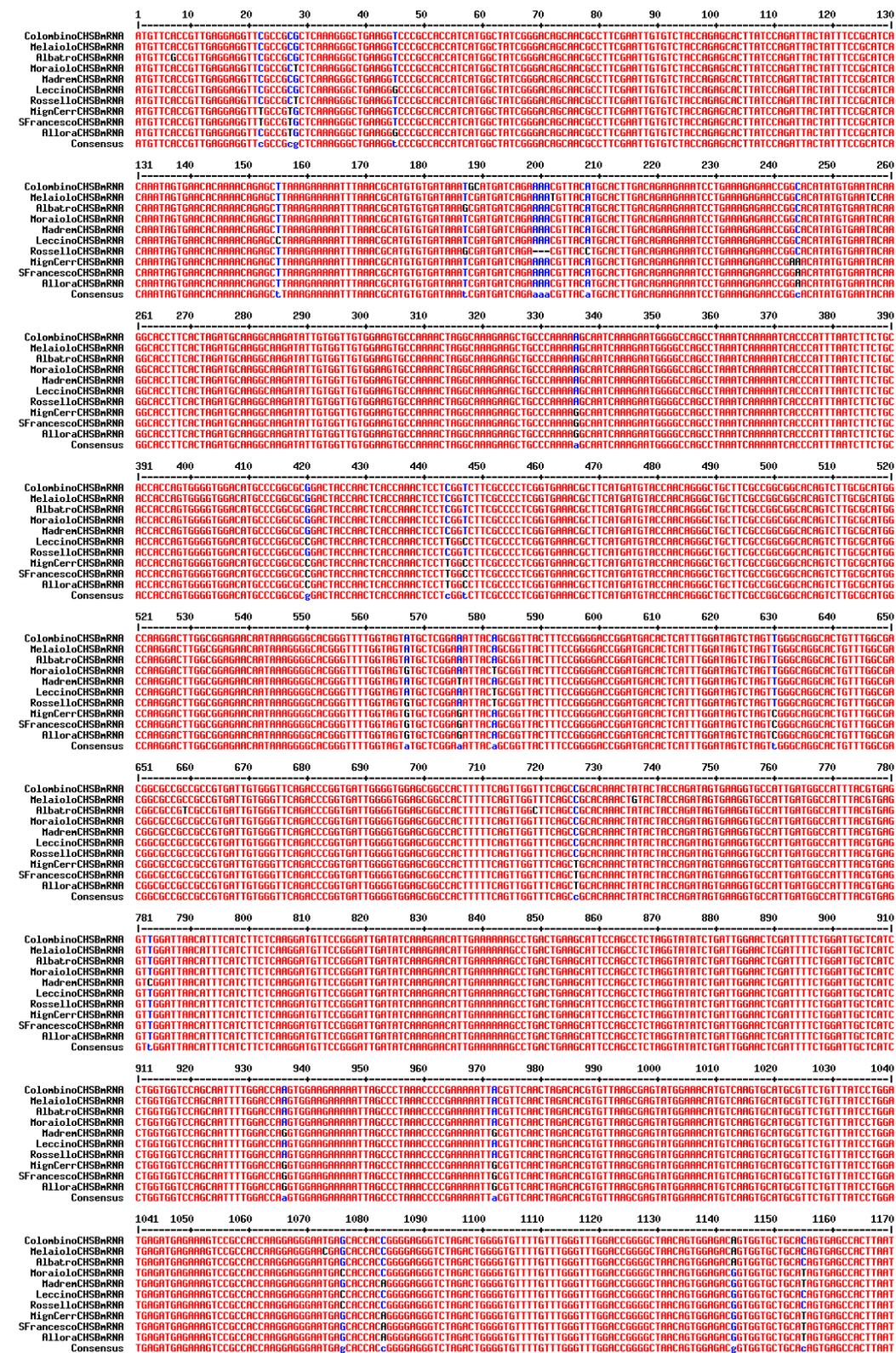


Figura 64. Allineamento delle sequenze di mRNA dedotte dalle sequenze geniche di *chsB* nelle cultivar toscane selezionate

In figura 65 è illustrato l'allineamento delle sequenze proteiche tradotte *in silico*. In questo caso è presente una modificazione strutturale rilevante, in quanto la delezione di una tripletta nella sequenza codificante nella cultivar Rossello porta alla delezione nella sequenza proteica di un residuo di Lisina. Altre variazioni di lunghezza nella sequenza della proteina non sono state individuate negli altri geni analizzati. Il maggiore livello di polimorfismo a livello delle sequenze porta ad un aumento della variabilità nelle sequenze proteiche dedotte. Le mutazioni individuate sono 16.

Le posizioni 8, 63, 82 e 359 coinvolgono 2 o più cultivar mentre le altre variazioni risultano essere cultivar specifiche.

Anche per il gene *chsB* non è stata trovata nessuna modificazione che coinvolgesse i residui biochimicamente più reattivi e importanti.

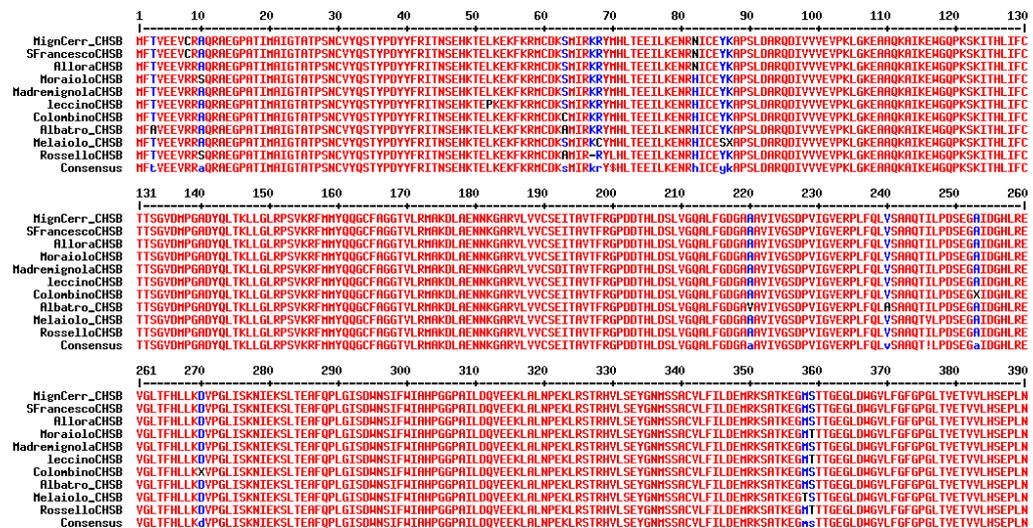


Figura 65. Allineamento delle sequenze amminoacidiche dedotte dalle sequenze geniche di *chsB* nelle cultivar toscane selezionate

Capitolo 4.2.3.3. Isolamento di *chsC* in cultivar toscane.

L'allineamento delle sequenze di *chsC* isolate nelle cultivar toscane è mostrato in figura 67. In questo caso sono stati rilevati 2383 siti, che escludendo i siti che presentano delezioni si riducono a 2380.

E' presente infatti una delezione 3 bp in posizione 860 nella cultivar Allora, che si trova in una parte non codificante della sequenza.

Del totale dei siti analizzati, 2311 sono monomorfici (97%), mentre 69 sono polimorfici. Anche in questo caso circa la metà dei siti polimorfici sono variazioni di singola base. In totale per *chsC* sono state riscontrate 72 mutazioni.

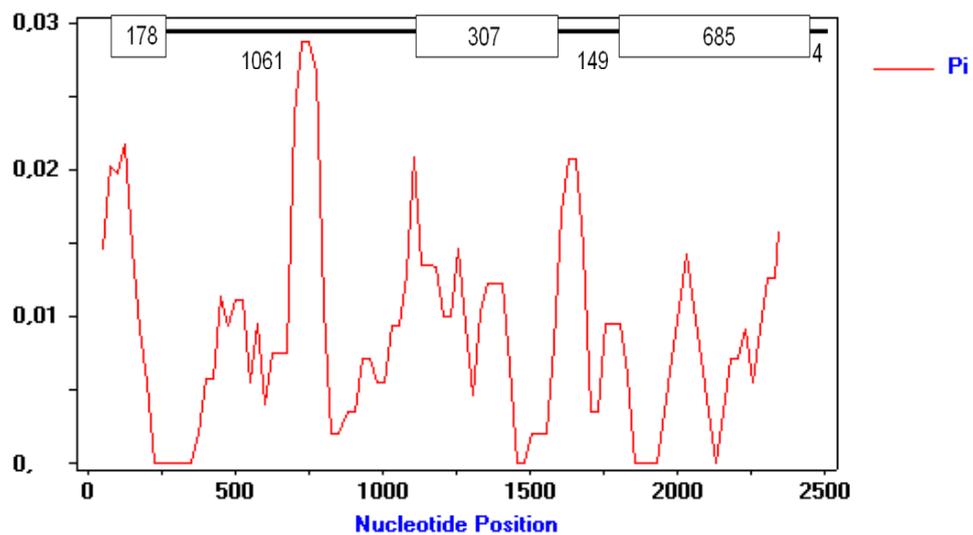


Figura 66. Analisi della diversità nucleotidica Pi in funzione della struttura del gene *chsC*

L'analisi della diversità nucleotidica lungo l'intera sequenza del gene (Figura 66) ha indicato il maggiore picco di variabilità a livello del sito 750, a metà della sequenza dell'introne 1.

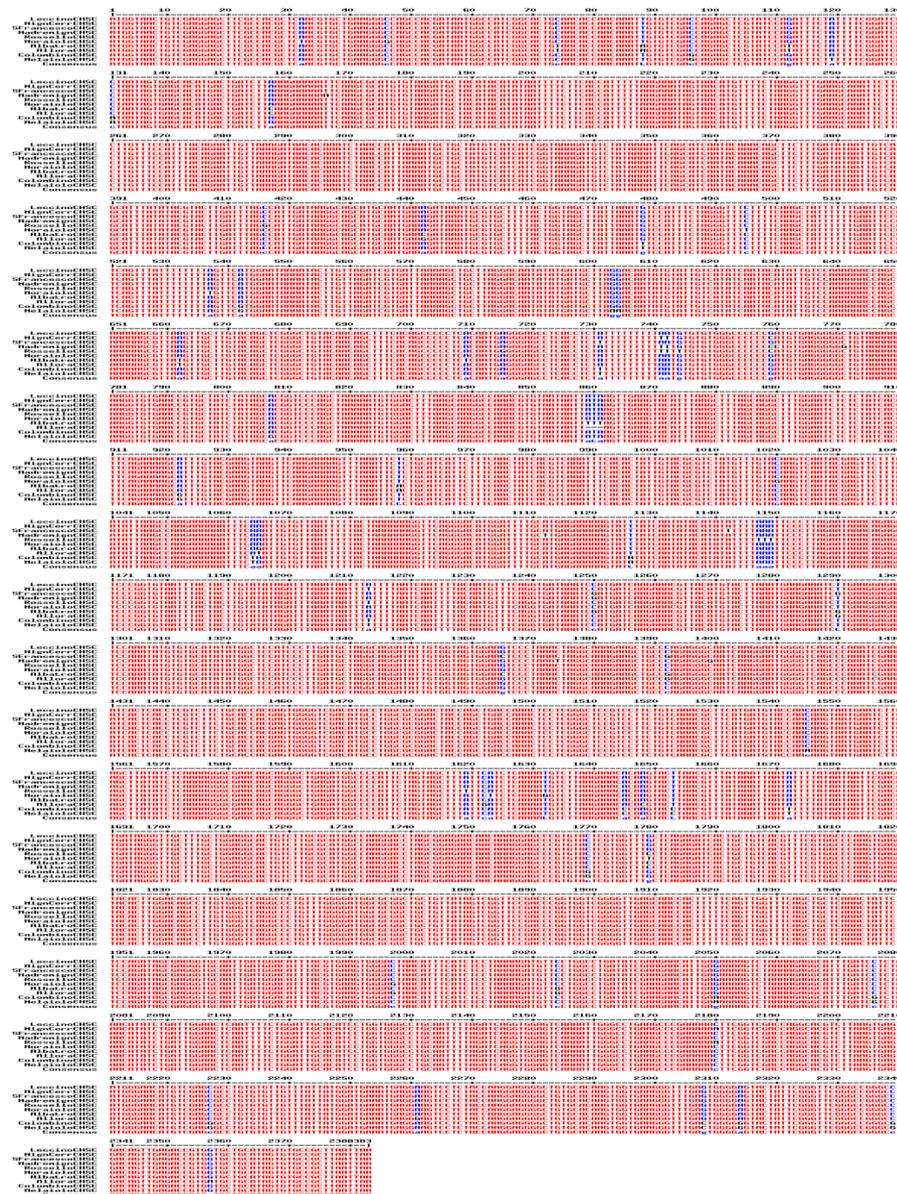


Figura 67. Allineamento delle sequenze geniche di *chsC* nelle cultivar toscane selezionate

Anche in questo caso, come riscontrato in *chsA* e come atteso, le zone che presentano una maggiore variabilità sono quelle non codificanti dell'introne 1 e 2. Le zone a minore variabilità che coincidono con l'esone 2 e l'esone 3 sono riconducibili a parti della sequenza altamente conservate che codificano per aminoacidi chiave per la funzione di una calcone sintasi.

Nelle figure 68 e 69 sono riportati rispettivamente gli allineamenti delle sequenze degli mRNA e le aminoacidiche CHSC dedotte *in silico*. Anche in questo caso la sequenza aminoacidica non presenta variazioni nei residui conservati.

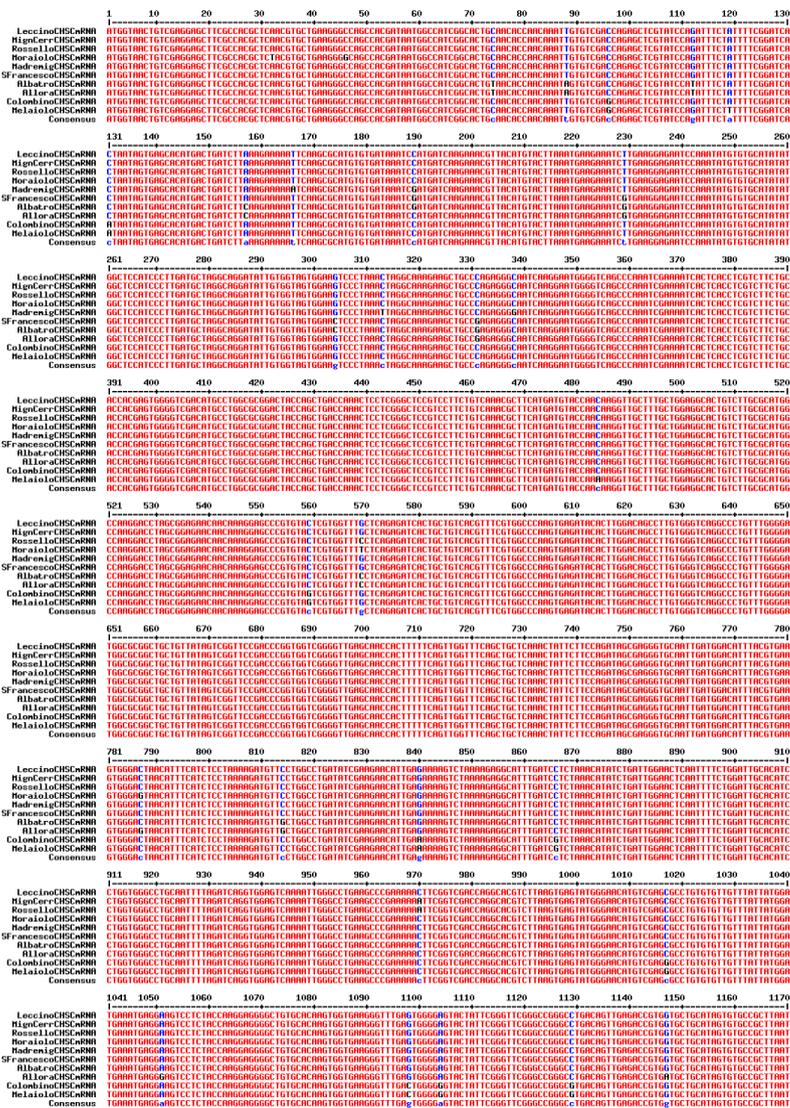


Figura 68. Allineamento delle sequenze di mRNA dedotte dalle sequenze geniche di *chsC* nelle cultivar toscane selezionate.

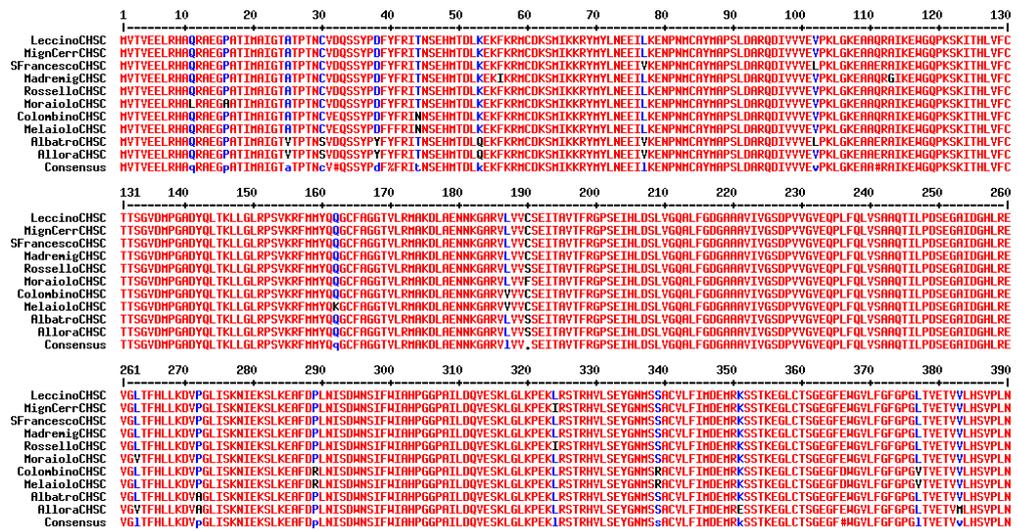


Figura 69. Allineamento delle sequenze amminoacidiche dedotte dalle sequenze geniche di *chsC* nelle cultivar toscane selezionate.

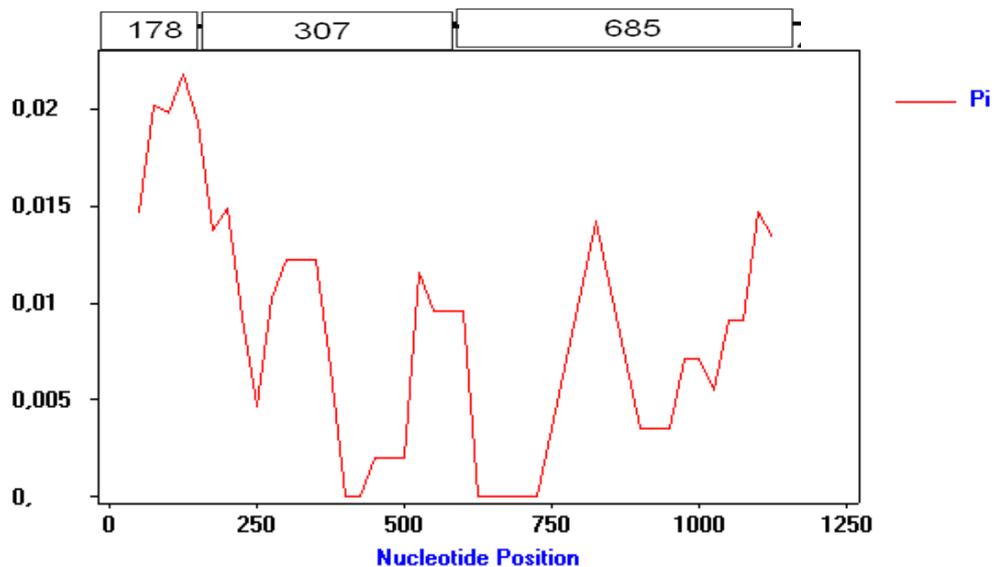


Figura 70. Diversità nucleotidica delle sequenze di mRNA di *chsB* in funzione della posizione. In alto la struttura dell'mRNA

La analisi di diversità nucleotidica (Figura 70) effettuata in funzione della posizione dei siti variabili nell'mRNA ci evidenzia come le posizioni più conservate nel messaggero siano i siti 400 e 600, corrispondenti rispettivamente alla parte centrale e finale dell'esone 2 e iniziale dell'esone 3. In particolare in queste regioni si trovano i residui più conservati **G163**, **C164**, F165, A166, **G167** e G252, A253, **I254**, D255 e **G256**. Come riportato in letteratura C164 rappresenta uno dei quattro residui che definisce il sito attivo, G163 e G167 sono importanti per definire la geometria di quest'ultimo mentre I254 e G256 sono residui fondamentali per il sito di ciclizzazione dell'enzima.

I valori di polimorfismo registrati nei geni *chsA*, *vhsB* e *chsC* sono riassunti nella tabella in figura 71.

	Siti analizzati	h	siti segreganti S	Numero Mutazioni Eta	Diversità Nucleotidica Theta per site S	Nucleotide diversity Pi	Numero medio di differenze K
CHSA	2201	10	51	53	0.009	0.0059	12.956
CHSB	2132	10	86	87	0.014	0.0135	28.822
CHSC	2380	10	69	72	0.010	0.0088	20.966

Figura 71. Analisi della variabilità genetica delle sequenze di *chsA*, *chsB* e *chsC* nelle 10 cultivar selezionate.

A partire dall'allineamento delle sequenze geniche di *chsA*, *chsB* e *chsC* abbiamo costruito un albero di similarità utilizzando il metodo UPGMA (Figura 74).

I valori riportati sopra i nodi rappresentano i livelli di bootstrap computati con 1000 repliche. Le distanze sono state misurate utilizzando il "Composite Likelihood method" valutando il numero di sostituzioni di base per sito (Tamura et al. 2004). L'analisi è stata effettuata con il software Mega4. Il dendrogramma

mostra due cluster principali. Uno dei due cluster è costituito dalle cultivar Allora, San Francesco e Mignolo Cerretano, che raggruppa tre cultivar coltivate nella provincia di Pistoia. Un sub-cluster raggruppa Albatro e Colombino con Moraiolo e Rossello. Tre di queste cultivar, Moraiolo Rossello e Albatro sono particolarmente sensibili alle basse temperature e alle altre avversità climatiche e molto suscettibili alla Mosca dell'Olivo.

L'allineamento delle sequenze di mRNA e delle sequenze aminoacidiche di CHSA CHSB e CHSC ottenute *in silico* a partire dalle sequenze genomiche è mostrato rispettivamente nelle figure 73 e 72 .

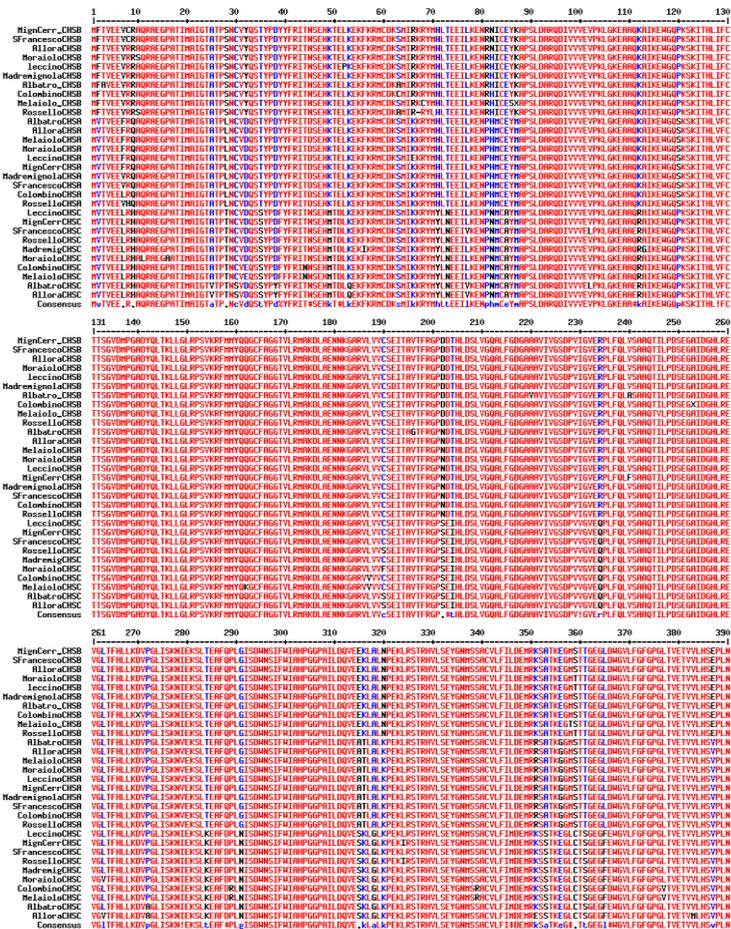


Figura 72. Allineamento delle sequenze aminoacidiche di CHSA, CHSB e CHSC delle cultivar toscane selezionate

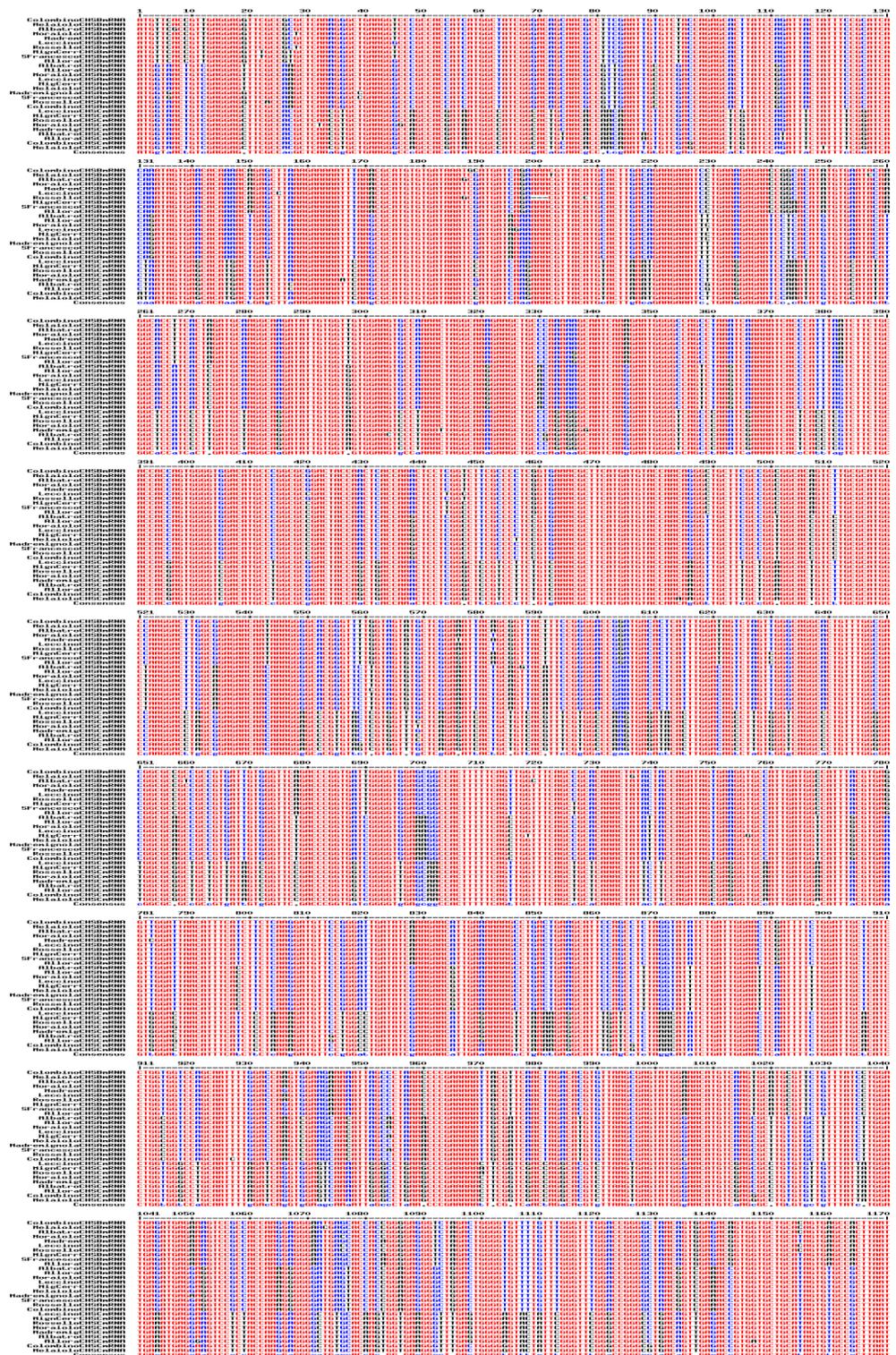


Figura 73. Allineamento delle sequenze di mRNA di *chsA*, *chsB* e *chsC* delle cultivar toscane selezionate.

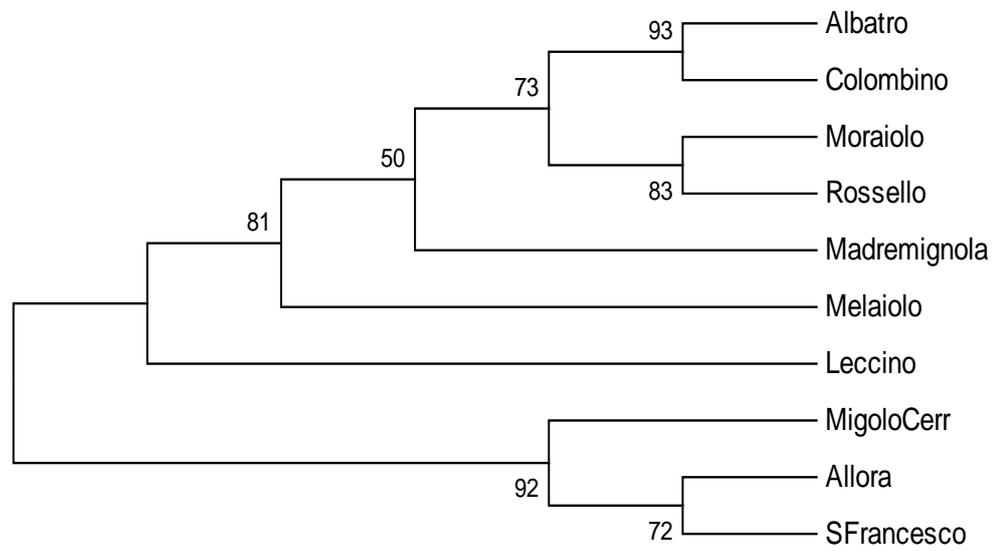


Figura 74. Albero di similarità costruito a partire dall'allineamento delle sequenze geniche di *chsA*, *chsB* e *chsC* utilizzando il metodo UPGMA. In prossimità dei nodi sono riportati i valori di bootstrap.

Capitolo 4.3 L'introne 1 di *chsA* come marcatore molecolare in cultivar italiane di olivo

Per indagare ulteriormente la variabilità genetica presente nella specie *Olea europaea* necessitavamo di ampliare il nostro studio ad un numero più elevato di cultivar. Oltre alle cultivar toscane abbiamo esteso il nostro studio a numerose altre cultivar originarie di diverse regioni dell'Italia del Sud e ivi coltivate. Abbiamo selezionato 58 accessioni di olivo che ci sono state fornite dagli enti di ricerca CRA, Istituto Sperimentale per l'Olivicoltura (Rende, Italia), dal Giardino Botanico dell'Università della Calabria (Arcavacata di Rende, Cosenza, Italia), dal CNR di Firenze.

Le cultivar esaminate in base alla provenienza regionale sono riportate di seguito:

- ❖ 12 dalla Toscana,
- ❖ 6 dalla Campania,
- ❖ 1 dalla Basilicata,
- ❖ 6 dalla Puglia,
- ❖ 5 dalla Calabria,
- ❖ 14 dalla Sicilia.

Inoltre per valutare la variabilità presente tra olivo coltivato e olivo selvatico abbiamo analizzato 6 accessioni di *Olea europaea* subsp. *europaea* var *sylvestris*, anche queste provenienti da diverse regioni dell'Italia meridionale ovvero

- ❖ 2 dalla Campania,
- ❖ 2 dalla Basilicata,
- ❖ 1 dalla Calabria,
- ❖ 1 dalla Sardegna.

Per studiare la variabilità all'interno di una singola cultivar abbiamo deciso di analizzare anche 8 accessioni appartenenti alla stessa cultivar, la cultivar calabrese

Carolea, campionate in diverse collezioni coltivate da Enti di ricerca e aziende agricole custodi in tre diverse località della Calabria: Cetraro, Cosenza e Rende.

Come marcatore molecolare abbiamo scelto di utilizzare l'introne 1 di *chs* che si è dimostrato sufficientemente variabile pur essendo conservato nelle zone di giunzione con gli esoni 1 e 2.

Per l'amplificazione dell'esone 1 di *chs* abbiamo scelto di seguire una strategia PCR utilizzando la coppia di primer universali CHSX1F/CHSX2RN disegnati da Strand e collaboratori (1997).

CHSX1F è disegnato sull'esone 1 e CHSX2RN sull'esone 2, entrambi in prossimità della giunzione esone-introne1 in una zona conservata. La coppia di primer ha amplificato con successo le specie del genere *Aquilegia* (Strand et al., 1997).

Le reazioni di amplificazione hanno prodotto in tutte le accessioni esaminate 2 o 3 bande. Gli ampliconi rispettivamente di 699bp, 614bp e 570bp sono stati visualizzati mediante una corsa elettroforetica su gel di agarosio 3 %. I risultati delle amplificazioni hanno indicato complessivamente che:

- ❖ Tutte le accessioni analizzate presentavano il frammento di 699bp;
- ❖ 20 cultivar e le 9 accessioni della cultivar Carolea presentavano sia il frammento di 699 bp, sia il frammento di 614 bp;
- ❖ le restanti 24 cultivar (le calabresi Ciciarello, Ottobratica e Tondina, le siciliane Cerasuola Minuta e Verdello, le campane Marinese, Ogliara, Ogliastro Grande, Ortice e Ravece, la lucana Rotondella, le pugliesi Cellina di Nardò, Cima di mola, Coratina, Ogliarola Barese e Peranzana, le toscane Albatro, Frantoio, Madremignola, Melaiolo, Rossello, San Francesco, Olivo Della Strega) presentavano tutte e tre le bande;
- ❖ le 6 accessioni di olivo selvatico a differenza delle cultivar analizzate presentavano oltre al frammento di 699bp anche un frammento di 580 bp

Successivamente i prodotti PCR di differente lunghezza, amplificati da 10 accessioni, compresa la varietà di olivo selvatico, sono stati clonati in plasmide e sequenziati. Abbiamo analizzato numerosi plasmidi, sequenziandoli in entrambe

le direzioni. Abbiamo quindi allineato le sequenze ottenute con i geni *chsA chsB* e *chsC* di *Olea*. I risultati dell'allineamento hanno indicato che il frammento di 614bp corrispondeva all'introne 1 del gene *chsA* e il frammento di 699 bp all'introne 1 di *chsB*, entrambi con una identità di circa il 100%. Il frammento di 570 bp mostrava invece l'84.1% d'identità con l'introne 1 di *chsA* (Figura 75) ed una bassa identità con l'introne 1 di *chsB* (Figura 76). L'allineamento del frammento di 614bp con il frammento di 699 bp ha mostrato una identità del 54.5% (Figura 77).

I risultati ottenuti suggeriscono che il frammento di 570bp, che abbiamo denominato *chsA-varS*, ed il frammento di 614bp, denominato *chsA-varA* sono due varianti dell'introne 1 del gene *chsA*. L'analisi di omologia BLAST non ha dimostrato identità significative con le sequenze presenti nel database GenBankTM/EMBL.

L'allineamento tra le sequenze di *chsA-varA* e *chsA-varS* evidenzia la presenza di tre diverse inserzioni: una di 6bp, una di 10bp e una inserzione di un microsatellite poli-T. All'interno delle sequenze varS questo microsatellite mostra un grado molto elevato di variabilità sia negli olivastri che nelle cultivar. VarA invece, mostra due inserzioni di 18bp e 39 bp.

Per identificare in tutte le cultivar analizzate e nelle accessioni di olivastro la presenza e/o l'assenza di varS e varA abbiamo disegnato due primer reverse PCR, specifici per una particolare variante o per l'altra, che abbiamo chiamato "chsAcvREV1" e "intryslv_rev1", il primo specifico per varA e il secondo per varS. Questi due primer sono stati utilizzati in combinazione con il primer CHSX1F. In questo modo con la combinazione CHSX1F - chsAcvREV1 è stato possibile amplificare una porzione di 428bp dell'introne 1 di *chsA-varA* in tutte le 53 accessioni di *Olea europaea* var *europaea*. Le 6 accessioni di olivo selvatico non hanno mostrato alcun segnale. Inoltre abbiamo amplificato una porzione di 432bp dell'introne 1 di *chsA-varS* in 24 cultivar e in tutte le 6 accessioni di olivo selvatico. Complessivamente 24 cultivar presentano sia *chsA-varA* che *chsA-varS*; le altre 29 accessioni di olivo coltivato presentavano solo *chsA-varA* e le accessioni di *Olea europaea* var *sylvestris* presentano solo *chsA-var S*.

```

        10      20      30      40      50      60
614  AAGAAAAATTTAAGCGCATGGCATGTGTAAGTTACCATAAGAAGTCTTGAATTGATAGTT
      .....
570  AAGAAAAATTTAAGCGCATGGCATGTGTAAGTTACCATAAGAAGTCTTGAATTGATAGTT
        10      20      30      40      50      60

        70      80      90      100     110     120
614  TTCCCTAGGGACATGTTTTCTAAGAAAAGTTGAAAATAATTTCTTTAACTTGCTTCTTAG
      .....
570  TTCCCTAGGGACATGTTTTCTAAGAAAAGTTGAAAATAATTTCTTTAACTTGCTTCTTAG
        70      80      90      100     110     120

        130     140     150     160     170     180
614  GTTATTTCTTTTGAGTTTGTGCATGCATATCTTCATTAATCTTCTAAAAGTAGAATGCT
      .....
570  GTTATTTCTTTTGAGTTTGTGCATGCATCTCTTCATTAATCTTCTAAAAGTAGAATGCT
        130     140     150     160     170     180

        190     200     210     220     230     240
614  TCGCGATGCTACACGGCGGTGGCGATGCTACGTAACGTTAAGTTGTTGCATGGAGGCTTA
      .....
570  TCGCGATGCTACACTGTGGTGGCGATGCTACGTAGCGTTAGGTTGTTGCATGGAGGCTTA
        190     200     210     220     230     240

        250     260     270     280     290
614  TGAGTGTTTGAACTGTCAGAAATAATTCTTTTACATTAATAAATGCAC--ACGACACTTGA
      .....
570  TGAGTGTTTGCACTGTCAGAAATAATACTTGTATATTAATAAATGCACACTGACAACACTTGA
        250     260     270     280     290     300

300      310      320      330      340
614  TTTACTGTTTTAACTATGCTTTATGCT-----CAAAATTCACGAAATGTCACC
      .....
570  TTTATTGTTTTAACTATGCTTTTTTTTTTTTTTTTTTCCAAATTCACGAAATGTCATC
        310      320      330      340      350      360

        350     360     370     380     390     400
614  ATGTGACATTCTGTAAACCCTTTTGTACATCTCTGATGTACATGTAGTATTGCTTTTGC
      .....
570  ATGTGACATTCTGTAAACCCTTTTGTGCATATCCGATGTACATGTAGTATTGC-----
        370     380     390     400     410

        410     420     430     440     450     460
614  CTGTCTATATGGTTATTAGGCTTTAATTTGCTACCTTTATGATTAAGAAATCACCATGCA
      .....
570  -----AATT-----GAAAGCCCAA----
                          420

        470     480     490     500     510     520
614  ATTGAAAGCTTTTCTCAGGTTTTCAATTCGATTGGACAGTGTAAAGAAGAAAAATGAAT
      .....
570  ---ATATGTTTTCTTAGGTTTTCAATTCGATTGGACAATGTAAAGAAGAAAAATGAAT
        430     440     450     460     470     480

        530     540     550     560     570     580
614  ATCTACTATTATATGAACGGTTAACGTATAACAAATTTGCAGGTGATAAATCGATGATA
      .....
570  ATCTACTATTATATGAACGGTTAATGTATGACAAATTTGCAGGTGATAAATCGATGATA
        490     500     510     520     530     540

```

Figura 75. Allineamento tra il frammento di 614 bp e il frammento di 570 bp isolati nella cultivar Ottobratica (identità dell'84.1%)

```

          10      20      30      40      50
570  AAGAAAAATTTAAGCGCATGGCATGTGTAAGTTACCATAAGAAGTCTTGAA--TTGATA
      ::::::::::: ::::: ::::: : : : : : : : : : : : : : : : : :
699  AAGAAAAATTTAAACGCATG----TGTAAGCT-CCTCTCTCTTTCTTTCTCCTTCGAAC
          10      20      30      40      50

          60      70      80      90      100     110
570  GTT--TTCCTAGGGACATGTTTCTAAGAAAAGTTGAAAATAATTTCTTTAACTTGCTT
      :: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699  GTGCCTATCATAATTATTTGTTTATAC-ATGACACGGACTAAAATTTCTTGAAAACTCGA
          60      70      80      90      100     110

          120     130     140     150     160     170
570  CTTAGGTTATTTCTTTGAGTT--TGTGCATGCATCTCTTCATTAATCTTTCTAAAAGTA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699  AAAATCTTGATTATTTGACATGTTTTCCAAGAAAATTTGAA--AATCAAACGAAAATT
          120     130     140     150     160     170

          180     190     200     210     220     230
570  AAATGCTTCGCGATGCTACACTGTGGTGGCGATGCTACGTAGCGTTAGGTTGTCATGG
      :: :::: : : : : : : : : : : : : : : : : : : : : : : : :
699  TAA-GCTT-GAAATCAGACTAAATAGTTAGACCAGTATGTAACATTTTGTATTGCATGG
          180     190     200     210     220

          240     250                               260     270
570  AGGCTTATGAGTGTTCACACT-----GTCAGAAAATAACTTGTAT-AT
      : ::::: : ::::: : : : : : : : : : : : : : : : : : : :
699  AAGCTTATTTGTGTTTGCACACTACTTATTTACTACTAGCCTTTAATGCTACCTATATCAT
230      240     250     260     270     280

          280                               290     300     310     320
570  TAAAAATGCAC-----TGACAACACTTGATTATTGTTTAACTATGCTT-
      :: : : : : : : : : : : : : : : : : : : : : : : : : : :
699  TAAGAAAGCACGATGCGACTGTGGATGACATCAAAAAGTTTGTACGTCAAA-ATGCTTA
290      300     310     320     330     340

                               330     340     350
570  57 -----TTTTTTTTTTTTTTTCCAAATTTCAACGA-A
      : : ::::: : : : : : : : : : : : : : : : : : : : : :
699  CATGACAAGAAAATGGTGGACGACCCACTTAATTTTTTATTAATTTATCTTTAATCAAGA
350      360     370     380     390     400

          360     370                               380     390     400
570  ATGTATCATGTGACATTCTGTAAA-----CCCTTT-TGTGCATATCCGAT---GTAC
      ::::: : : : : : : : : : : : : : : : : : : : : : : : : :
699  ATGTCTCTATATAATTTATTGTACTTTATTTCGCTCTTTCTCTGCATTTTCTCTTTGTAC
410      420     430     440     450     460

          410     420                               430
570  ATGTAGT-ATTGCAATTGAAAGCCCA-----AATA-----TGTTT
      :: : : : : : : : : : : : : : : : : : : : : : : : : : :
699  ATCTATTGAATGTGCTCTTCTTCCCATTTTGCTCGTCGTATACAATAATTCACACTGTTG
470      480     490     500     510     520

          440     450     460     470     480
570  -----TCTTAGGTTTTCAATTCGATTGGACAATGTTAAAGAA-GAAAAATGAA
      ::::::::::: : : : : : : : : : : : : : : : : : : : : : : :
699  CATTGGAGATGATCTTAGGTTTTTCAGTTGGACAAGACGAAATTAATAAAGAAAAA
530      540     550     560     570     580

```

Figura 76. Allineamento tra il frammento di 570 bp e il frammento di 699 bp isolati nella cultivar Ottobratica (identità del 52.2%)

```

        10      20      30      40      50
614  AAGAAAAATTTAAGCGCATGGCATGTGTAAGTTACCATAAGAAGTCTTGAA---TTGATA
      ::::::::::: :::::  ::::: : : : : : : : : : : : : : : : : :
699  AAGAAAAATTTAAACGCATG-----TGTAAGCTCCTCTCTCTT-TCTTTCTCCTTCGAAC
        10      20      30      40      50

        60      70      80      90      100     110
614  GTT--TTCCCTAGGGACATGTTTTCTAAGAAAAGTTGAAAATAATTTCTTTAACTTGCTT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699  GTGCCTATCATAATTATTTGTTTTATAC-ATGACACGGACTAAAATTTCTTGAAAACCTCGA
        60      70      80      90      100     110

        120     130     140     150     160     170
614  CTTAGGTTATTTCTTTTGA---GTTTGTGCATGCATATCTTCATTAATCTTTCTAAAAGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699  AAAATCTTGATTATTTGACATGTTT-TCCAAGAAAATTTGAA--AATCAAACCTGAAAAT
        120     130     140     150     160     170

        180     190     200     210     220     230
614  AGAATGCTTCGCGATGCTACACGGCGGTGGCGATGCTACGTAACGTTAAGTTGTTGCATG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699  TTAA-GCTT-GAAATCAGACTAAATAGTTAGACCAGTATGTAACATTTTGTATTGCATG
        180     190     200     210     220

        240     250                               260     270
614  GAGGCTTATGAGTGTTTGAACT-----GTCAGAAAATAATTTCTTTTA-CA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699  GAAGCTTATTTGTGTTGCACTACTTATTTTACTACTAGCCTTAAATGTACCTATATCA
230      240     250     260     270     280

        280     290     300     310     320
614  TTAAAAATGCAC-ACGACACT-TGATTTACTGTTTTAACTATG-----CTTTATGCTCA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699  TTAAAGAAAGCAGATGCGACTGTGGATGACATCAAAAAGTTTGTACAGTCAAAAATGCTTA
290      300     310     320     330     340

        330     340                               350     360     370
614  AAATTCAACGAAATGTCA-----CCATGTGACATTCTGTAAACCTTTTGTGA--CA---
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699  CATGACAAGAAAATGGTGGACGACCCACTTAATTTTTTATTAATTTATCTTTAATCAAGA
350      360     370     380     390     400

        380     390     400     410     420
614  ---TCTCTGATGTACATGTAGTATTGCTTTTGTCTGTCTCT---ATATGGTTATTAGGC-
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699  ATGTCTCTATATAATTTATTTGACTTTATTCGCTCTTTCTCTGCATTTTCTCTTTTGTAC
410      420     430     440     450     460

        430     440                               450     460
614  -----TTTAATTTGCTA-----CCTTTATGATT-----AAGAAATCACCATGCA-
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699  ATCTATTGAATGTGCTCTTCTTCCCATTTTGTCTCGTGTATACAATAATTCACACTGTTG
470      480     490     500     510     520

        470     480     490     500     510     520
614  -ATTGAAAGCTTTTCTCAGGTTTTCAATTCGATTGGACAGTGTAAAGAA-GAAAAATGA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
699  CATTGGA-GATGATCTTAGGTTTTCAGTTGGACAAGACGAAATTAATAAAGAAAAAAA
530      540     550     560     570     580

```

Figura 77. Allineamento tra il frammento di 614 bp e il frammento di 699 bp isolati nella cultivar Ottobratica (identità del 54.5%)

In particolare, come riassunto in figura 78 :

- ❖ le cultivar calabresi Cassanese, Ciciarello, Tondina, presentano solo la var A mentre Ottobratica presenta sia var A che varS;
- ❖ le siciliane Biancolilla, Brandorfinò, Buscionetto, Giarraffa, Moresca, Nocellara del Belice, Nocellara Etnea Ovale, Ogliarola Messinese, Tonda Iblea presentano solo VarA, mentre Cerasuola e Minuta presentano entrambe le varianti;
- ❖ le campane Marinese, Ortice e Ravece presentano varA e varS, mentre Carpellese solo varA; la lucana Rotondella presenta var A e varS;
- ❖ le pugliesi Cima di Mola, Ogliarola Barese e Peranzana presentano varA e varS) mentre Rotondella Di Foggia solo varA;
- ❖ le 6 accessioni di *Olea europaea var sylvestris* presentano solo varS.

Per investigare il polimorfismo intra-cultivar abbiamo selezionato anche 8 delle 9 accessioni della cultivar calabrese Carolea. I risultati degli esperimenti hanno indicato innanzitutto che le 8 accessioni di Carolea presentano tutte la stessa sequenza. L'analisi dei polimorfismi mostrata in Figura 79 indica anche che esiste una maggiore variabilità all'interno di *chsA*-varS rispetto a *chsA*-var A. Sulla base di questi polimorfismi è stato costruito un dendrogramma utilizzando l'algoritmo NJ (vedi Materiali e Metodi). Come si vede in figura 80 le due varianti A e S vengono separate chiaramente in due clusters differenti.

	n°	Accessions name	Provenance
	1	Carolea	Calabria
	2	Carolea	Calabria
	3	Carolea	Calabria
	4	Carolea	Calabria
	5	Carolea	Calabria
	6	Carolea	Calabria
	7	Carolea	Calabria
	8	Carolea	Calabria
	9	Carolea	Calabria
	10	Cassanese	Calabria
	11	Ciciarello	Calabria
	12	Ottobratica	Calabria
	13	Tondina	Calabria
	14	Biancolilla	Sicily
	15	Brandofino	Sicily
	16	Buscionetto	Sicily
	17	Cerastola	Sicily
	18	Giarraffa	Sicily
	19	Minuta	Sicily
	20	Moresca	Sicily
	21	Nocellara del Belice	Sicily
	22	Nocellara Etnea ovale	Sicily
	23	Nocellara etnea	Sicily
	24	Nocellara messinese	Sicily
	25	Ogliarola messinese	Sicily
	26	Tonda iblea	Sicily
	27	Verdello	Sicily
614	Var A		
580	Var S		
698	CHSB		

	n°bp	Accessions name	Provenance
	28	Carpellese	Campania
	29	Marinese	Campania
	30	Ogliara	Campania
	31	Ogliastro grande	Campania
	32	Ortice	Campania
	33	Ravece	Campania
	34	Rotondella	Basilicata
	35	Cellina di Nardò	Apulia
	36	Cima di Mola	Apulia
	37	Coratina	Apulia
	38	Ogliarola barese	Apulia
	39	Peranzana	Apulia
	40	Rotondella di Foggia	Apulia
	41	Albatro	Tuscany
	42	Alloria	Tuscany
	43	Colombino	Tuscany
	44	Frantoio	Tuscany
	45	Leccino	Tuscany
	46	Madremignola	Tuscany
	47	Melaiolo	Tuscany
	48	Mignolo Cerretano	Tuscany
	49	Moraiole	Tuscany
	50	Rossello	Tuscany
	51	S.Francesco	Tuscany
	52	Olivo Della Strega	Tuscany
	53	O.e. var <i>Syhexstris</i>	Calabria
	54	O.e. var <i>Syhexstris</i>	Campania
	55	O.e. var <i>Syhexstris</i>	Campania
	56	O.e. var <i>Syhexstris</i>	Basilicata
	57	O.e. var <i>Syhexstris</i>	Basilicata
	58	O.e. var <i>Syhexstris</i>	Sardinia
614	Var A		
580	Var S		
698	CHSB		

Figura 78. Schema riassuntivo della analisi effettuata nelle cultivar italiane di olivo.

Variant	N°sequences	N°Haplotypes	N°mutations	Haplotype (gene) diversity	Nucleotide diversity (per site) Pi	Average number of nucleotide differences, k:	Theta (per sequence) from S, Theta-W:	Theta (per site) from S, Theta-W:	G+C content
var A	52	10	7	0.803	0,00484	1,96983	1,54908	0,00381	0,346
var S	19	12	12	0.924	0,00535	2,51462	3,43337	0,00731	0,316

Figura 79. Analisi dei polimorfismi di varA e varS nelle cultivar analizzate.

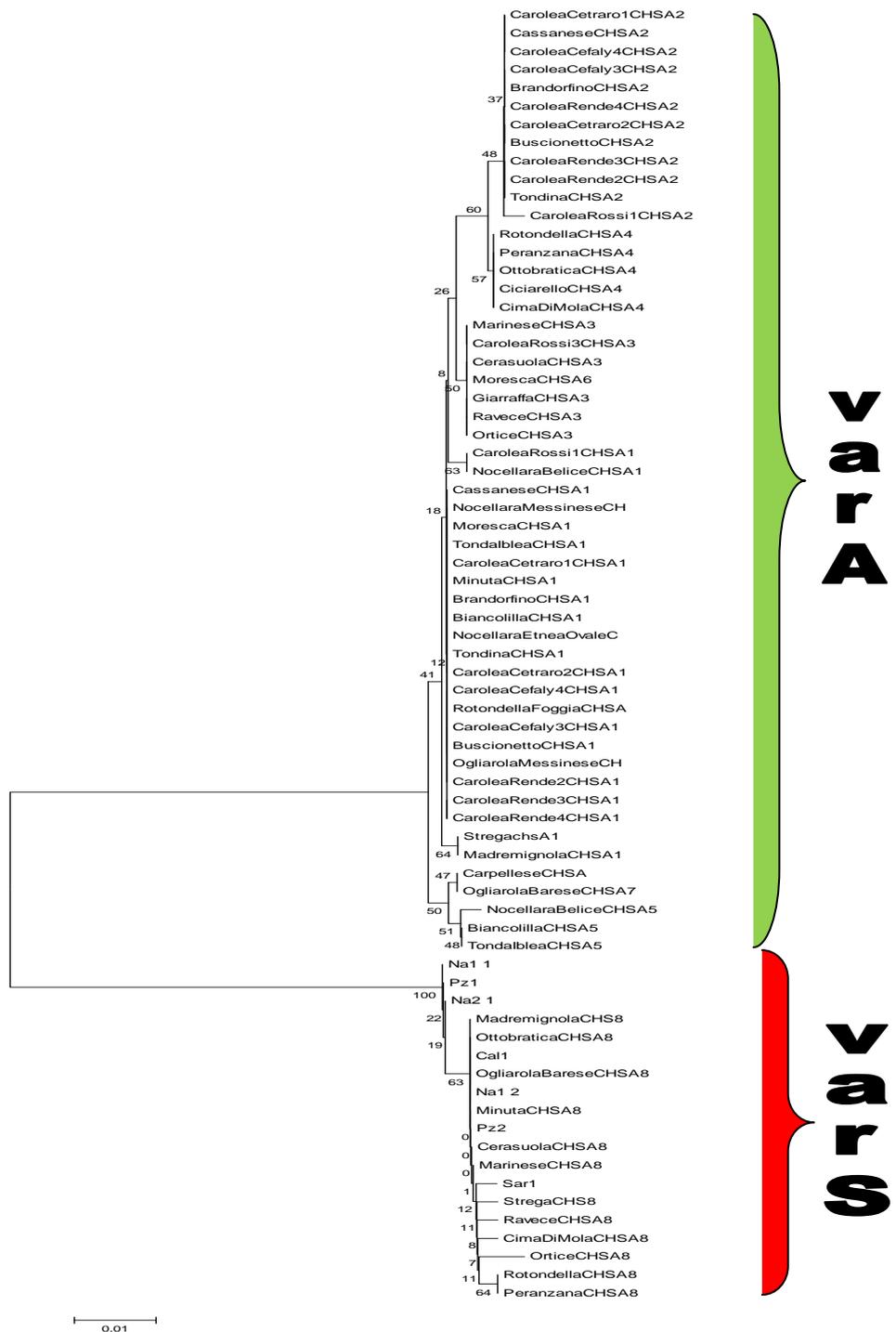


Figura 80. Dendrogramma costruito utilizzando l’algoritmo NJ con le sequenze varA e varS isolate nelle cultivar italiane di olivo analizzate.

Capitolo 4.4 I marcatori molecolari ISSR: analisi della variabilità genetica in cultivar calabresi e siciliane di olivo

In questa parte del lavoro abbiamo voluto indagare ulteriormente il polimorfismo presente tra diverse cultivar di olivo provenienti da due areali vicini ma separati da una barriera naturale, quello calabrese e quello siciliano e tra varie accessioni della stessa cultivar, la Carolea. A questo scopo abbiamo scelto di utilizzare un'altra tipologia di marcatore *random whole-genome*. Abbiamo infatti utilizzato i marcatori molecolari ISSR per studiarne i patterns di amplificazione in 19 cultivar siciliane e calabresi, in 9 accessioni di Carolea e in una accessione di olivo selvatico proveniente dalla Calabria (Figura 81).

SPECIES	CULTIVARS	CRA-OLI* CODE	ORIGIN	TREE-VIGOUR	LEAF	FRUIT SHAPE	PURPOSE
<i>O. europaea</i>	Carolea.Citraro1	A38	Calabria	Medium	Elliptic-lanceolate	Elliptic	Double-purpose
"	Carolea Cefaly 3	A7	Calabria	Medium	Elliptic-lanceolate	Elliptic	Double-purpose
"	Carolea Cefaly 4	A7	Calabria	Medium	Elliptic-lanceolate	Elliptic	Double-purpose
"	Carolea Rende 2	F1P6	Calabria	Medium	Elliptic-lanceolate	Elliptic	Double-purpose
"	Carolea Rende 3	F1P8	Calabria	Medium	Elliptic-lanceolate	Elliptic	Double-purpose
"	Carolea Rende 4	F1P11	Calabria	Medium	Elliptic-lanceolate	Elliptic	Double-purpose
"	Carolea Rende 5	F1P14	Calabria	Medium	Elliptic-lanceolate	Elliptic	Double-purpose
"	Carolea Rossi 1	A8	Calabria	Medium	Elliptic-lanceolate	Elliptic	Double-purpose
"	Carolea Rossi. 3	A8	Calabria	Medium	Elliptic-lanceolate	Elliptic	Double-purpose
"	Cassanese	20	Calabria (Cosenza)	Strong	Elliptic	Elliptic	Double-purpose
"	Ciciarello	115	Calabria	Strong	Elliptic	Rounded	Oil
"	Ottobratica	25	Calabria (Reggio Calabria)	Strong	Elliptic	Elliptic	Oil
"	Tondina (Cosenza) var. <i>syvestris</i>	30	Calabria	Strong	Elliptic-lanceolate	Rounded	Double-purpose
"	Biancolilla	347	Sicily	Weak	Elliptic-lanceolate	Elliptic	Oil
"	Brandorfino	78	Sicily (Messina)	Weak	Elliptic	Elliptic	Double-purpose
"	Buscionetto	79	Sicily	Medium-Strong	Elliptic	Elliptic	Oil
"	Cerasuola	80	Sicily (Agrigento)	Medium-Strong	Elliptic-lanceolate	Elliptic	Oil
"	Giarraffa	81	Sicily (Palermo)	Medium	Elliptic-lanceolate	Elliptic	Table
"	Minuta	83	Sicily (Messina,Palermo)	Medium	Elliptic-lanceolate	Rounded	Oil
"	Moresca	84	Sicily (Ragusa,Siracusa)	Medium-Strong	Elliptic	Oval	Oil
"	Nocellara Belice	85	Sicily (Trapani)	Weak	Elliptic-lanceolate	Elliptic	Oil
"	Nocellara Etnea	183	Sicily (Catania)	Medium	Lanceolate	Elliptic	Double-purpose
"	Nocellara Etnea Ovale	86	Sicily (Catania)	Medium	Lanceolate	Oval	Table
"	Nocellara Messinese	87	Sicily (Messina)	Medium	Elliptic	Elliptic	Double-purpose
"	Ogliarola Messinese	88	Sicily (Messina)	Medium	Elliptic	Elliptic	Oil
"	Tonda Iblea	90	Sicily (Ragusa)	Medium	Elliptic-lanceolate	Rounded	Double-purpose
"	Verdello	91	Sicily (Messina)	Medium	Elliptic	Elliptic	Oil

*CRA-OLI= Istituto Sperimentale per l'Olivicoltura (Rende, Italy).

Figura 81. Cultivar utilizzate: sono indicati alcuni caratteri morfologici e fisiologici, origine geografica e codici della collezione di provenienza.

Dapprima 12 primer dinucleotidici, riportati nella tabella in figura 82, sono stati esaminati in un subset di campioni per valutarne la capacità di originare pattern leggibili e riproducibili.

ISSR1	(GT) ₆ Y*R*N*	-
ISSR2	(CT) ₈ TRA	+
ISSR3	GGGC(GA) ₈	-
ISSR4	CV*W*(AG) ₆	-
ISSR5	(GT) ₆ YR	-
ISSR6	(GT) ₆ AY	+
ISSR7	(CA) ₆ RY	+
ISSR8	(CA) ₆ RG	+
ISSR9	(CT) ₈ RG	+
ISSR10	(GT) ₆ RG	+
ISSR11	(GT) ₈ RYRY	-
ISSR12	YGYACAY(TG) ₇	+

*Type of degenerate nucleotide: Y= (C+T); R= (A+G); N = (A+C+G+T); V = (G+A+C); W = (A+T). + primers selected for the analysis, - primers discarded.

Figura 82. Nome e sequenza dei primer utilizzati.

Dai risultati di queste amplificazioni sono stati eliminati dalla successiva analisi i primer ISSR5 e ISSR 11 a causa di un pattern non chiaro oppure ISSR1 e ISSR4 a causa di un pattern con poche bande.

I primer ISSR2, ISSR6, ISSR7, ISSR8, ISSR9, ISSR10, ISSR12 hanno prodotto invece un numero ottimale di bande che variava da 6 a 11.

Ad esempio in Figura 83 è mostrato il pattern ottenuto con ISSR2.

Questa analisi ha permesso l'amplificazione di 43 frammenti riproducibili di dimensioni variabili da 150 a 2000 bp.

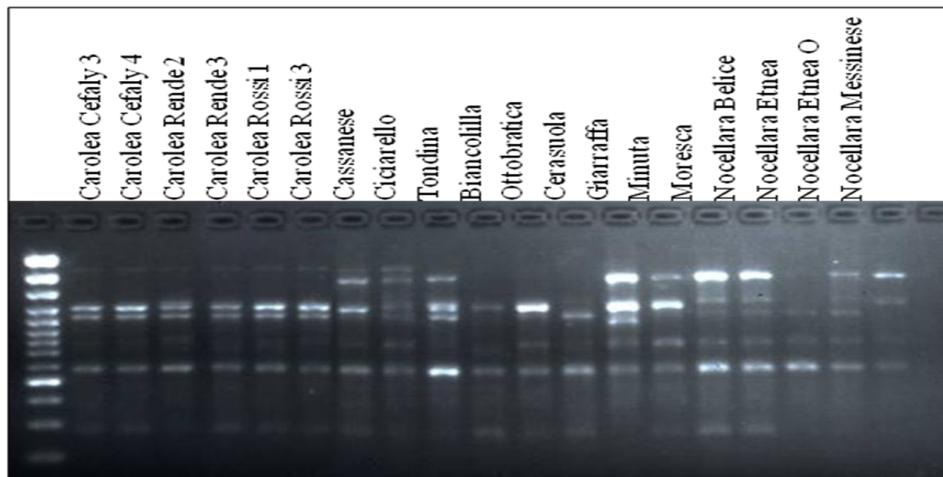


Figura 83. Elettroforesi su gel di agarosio del pattern ottenuto con ISSR2.

Frammenti della stessa dimensione sono stati considerati appartenenti allo stesso locus. La presenza e l'assenza sono state indicate in una matrice binaria 0/1- La matrice è stata utilizzata per calcolare una distanza di similarità con il coefficiente "Simple Matching" incluso nel software NTSYS (NTSYS software package, version 2.1 (Exeter Software, Setauket, New York).

I valori ottenuti variavano da 1 a 0.43. La matrice di distanza è stata utilizzata per costruire un fenogramma UPGMA (Figura 84), validato con una analisi di correlazione cofenetica calcolata con il programma MXCOMP. La correlazione espressa ($r=0.84$) indica che le differenze osservate sono significative. Sulla base di questa analisi le cultivar vengono chiaramente discriminate in due differenti cluster che raggruppano ciascuno le cultivar della stessa regione, con l'eccezione di una accessione di Carolea, in particolare Carolea Rende2 che viene posizionata assieme alle cultivar siciliane Brandorfino ed Ottobratica. A questo stesso cluster appartiene anche *O. europaea* var *sylvestris*. Inoltre tutte le cultivar "Nocellara" sono raggruppate, con l'eccezione di Nocellara Messinese. Allo stesso modo Giarruffa e Nocellara del Belice, considerate da alcuni autori sinonime,

clusterizzano nello stesso sottogruppo. Biancolilla e Moresca, considerate anch'esse sinonime da alcuni autori, sono invece separate in due diverse ramificazioni.

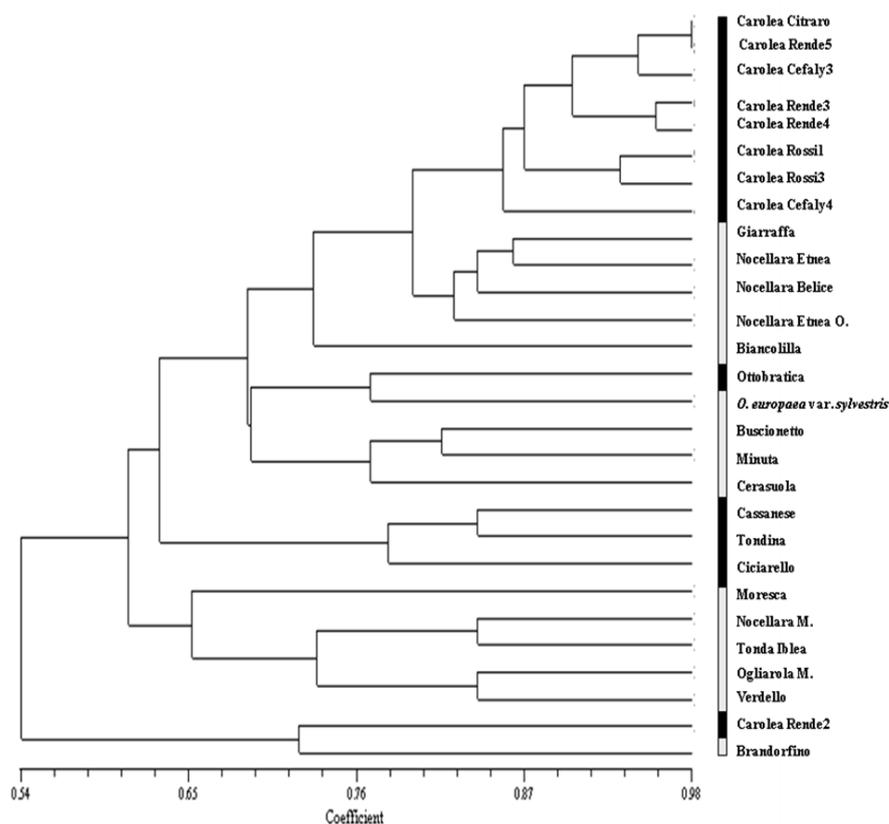


Figura 84. Fenogramma UPGMA delle cultivar analizzate. Le linee di fianco all'albero indicano la provenienza geografica (in nero le cultivar calabresi, in bianco quelle siciliane).

Capitolo 5

Discussione e conclusioni

Come discusso precedentemente le *chs* costituiscono una famiglia genica isolata e caratterizzata in numerose specie vegetali ma per quanto riguarda *Olea europaea* a nostra conoscenza non esistono per ora in letteratura studi esaurienti e anche in Genbank si trova sequenziato solo una parte dell'introne 2. Nel presente lavoro è stata effettuata la caratterizzazione molecolare della famiglia genica della calcione sintasi in *Olea europaea*, si è analizzata la variabilità fra i diversi geni presenti in questa specie e si è poi passati alla caratterizzazione e organizzazione in un albero filogenetico della calcione sintasi di una serie di cultivar, operazione molto utile per la protezione delle varietà locali e tipiche in attesa della attuazione della direttiva europea sulle varietà di conservazione. Per rendere più rapida ed efficiente questa operazione sono stati sviluppati, sulla base delle conoscenze acquisite, marcatori molecolari potenzialmente utilizzabili sia per la protezione che, eventualmente per la MAS (Marker Assisted Selection) in olivo.

Lo studio comparativo delle sequenze dei geni *chs* isolati in diverse specie vegetali e riportate in letteratura e in GenBank può essere sintetizzato nel modo seguente. :

1. La famiglia genica della *chs* è composta da un numero di geni variabile a seconda delle specie.
2. La famiglia genica *chs* è relativamente piccola dal punto di vista numerico e varia generalmente da 1 a 10
3. La struttura genica a due esoni ed un introne appare conservata in tutte le specie nelle quali è stata isolata con la eccezione di *Antirrhinum majus*
4. All'interno della sequenza amminoacidica sono identificabili alcuni residui fondamentali dal punto di vista funzionale estremamente conservati
5. La posizione dell'introne è conservata (in conseguenza lo è anche il numero di basi che compongono l'esone 1)

Il presente lavoro ha dimostrato, in accordo con le caratteristiche comuni sopra elencate che la calcione sintasi in olivo è codificata da una piccola famiglia multigenica composta da tre geni, chiamati *chs A*, *chs B*, *chs C*. L'introne 1 si inserisce nella posizione conservata dopo le 178 bp del primo esone.

Tutti e tre i geni isolati presentano nella traduzione in silico i residui conservati nelle *chs* isolate nelle specie già caratterizzate che si trovano in posizione strategica nel sito attivo o comunque in posizioni significative dal punto di vista funzionale. Ne sono esempi i residui del sito attivo Cys164, Phe215, His303, Asn336; i residui del "coumaroyl binding pocket" Ser133, Glu192, Thr194, Thr197, Ser338 e i residui del "cyclization pocket" Thr132, Met137, Phe215, Ile254, Gly256, Phe265, Pro375, come dimostrato dalla struttura cristallina della *chs* di *Medicago sativa*.

Tuttavia i geni *chs* da noi isolati in olivo si distaccano in modo sostanziale dalla struttura conservata a due esoni e un introne. In olivo in tutti e tre i geni si ritrova una struttura a tre esoni e due introni isolata in precedenza esclusivamente nella specie *Antirrhinum majus*. Va notato tuttavia che nel caso della *Bocca di Leone comune*, è stato isolato un solo gene della famiglia *chs* e quindi manca del tutto il gene che presenta la struttura conservata mentre in *Olea europaea* i geni sono tre. Quindi in *Antirrhinum majus* come in *Olea europaea*, le eccezioni strutturali a ben vedere sono due: la struttura a tre esoni e due introni e l'assenza della struttura conservata, considerata in letteratura ancestrale rispetto alle altre.

Analizzando in dettaglio la struttura dei geni e allineando *chs A*, *chs B* e *chs C* si nota immediatamente come i primi due appaiano più simili tra loro rispetto al terzo in termini di livello di identità

Risultati analoghi sono stati ottenuti anche analizzando la famiglia genica *chs* in altre specie come ad esempio *Petunia hybrida* (Koes et al., 1989; Yang et al., 2006).

Come accennato in precedenza, esistono differenze sostanziali fra i tre geni *chs* isolati anche dal punto di vista delle sequenze: le tre sequenze infatti presentano variabilità nel numero di basi degli introni (come atteso) ma anche negli esoni.

Questo comporta che l'introne 2 si inserisca in posizione diversa in *chs C* rispetto a *chs A* e *chs B*. Va inoltre notato che il punto di inserzione dell'introne 2 in *chs C* è il medesimo di *chs* di *Antirrhinum majus* e che complessivamente

Antirrhinum majus e *Olea europaea*, sono molto vicine dal punto di vista filogenetico dato che appartengono ambedue all'Ordine delle *Lamiales*.

Un risultato analogo si ottiene dalla analisi delle sequenze proteiche di numerose *chs* che conferma la coerenza fra l'albero di questo gene e la organizzazione tassonomica accettata che vede insieme i geni di specie dello stesso ordine, come avviene per *Antirrhinum majus*, *Olea europaea*, *Digitalis lanata*, *Perilla frutescens*, tutte appartenenti alle *Lamiales*.

Non é stato invece possibile escludere che la struttura caratteristica a due introni e tre esoni sia presente in altre specie appartenenti a questo ordine, visto che per quanto riguarda quelle prima citate è disponibile esclusivamente la sequenza di cDNA e non di DNA genomico.

Più in generale, la presenza nelle banche dati di un numero molto elevato di *chs* ottenute per retrotrascrizione a partire dallo RNA messaggero corrispondente, invece di sequenze ottenute per amplificazione di DNA genomico, potrebbe avere in parte “mascherato” la presenza di un secondo introne nella struttura dei geni *chs*.

E' probabile d'altra parte che nella storia evolutiva di *chs* come i molti altri casi si siano susseguiti eventi di duplicazione e divergenza o estinzione.

Da questo punto di vista la più alta identità dimostrata tra *chs A* e *chs B* rispetto a *chs C* può farci ipotizzare che il gene ancestrale abbia subito una duplicazione che ha portato alla formazione di *chs C* e di un altro paralogo che ha poi subito una ulteriore duplicazione originando *chs A* e *chs B*.

Questa ipotesi è solo in parte confermata dall'albero NJ, che distingue *chs C* da *chs A* e *chs B*, ma con livelli inadeguati di bootstrap.

Nella parte introduttiva alla tesi è stato descritto come attualmente sia presente una disputa scientifica sull'attribuzione tassonomica dell'olivo selvatico, che viene identificato da alcuni autori come una specie a sé, da altri come una sub-specie, da altri ancora come una varietà.

In più sappiamo come sia spesso difficoltoso il riconoscimento reciproco di olivi coltivati e selvatici, in particolare le accessioni sfuggite alla domesticazione che hanno revertito i loro caratteri fenotipici, all'interno di un generale processo di “re-inselvaticamento” della pianta.

Le finalità di questa parte di lavoro sono state mirate alla messa a punto di un marcatore molecolare che costituisca una metodica di riconoscimento alternativa, rapida e allo stesso tempo economica, ma soprattutto non ambigua.

Utilizzando una particolare PCR chiamata EPIC - PCR (exon - primed intron - crossing PCR) siamo stati in grado di individuare un marcatore molecolare ILPs (intron length polymorphism) in grado di discriminare le varietà coltivate dalla selvatica.

Oltre alla caratterizzazione delle varietà selvatica e coltivata, questo approccio ci ha permesso di confermare possibili ibridazioni tra esse. Ne è prova la presenza dell'allele caratteristico delle forme selvatiche in numerose varietà coltivate. Tali varietà sono state individuate indiscriminatamente a prescindere dalla provenienza regionale. Da notare però una differente percentuale di presenza dell'allele varS a seconda della provenienza regionale, per cui potremmo ipotizzare una differente diffusione della variante varS nelle diverse regioni.

La dimostrazione di questa ipotesi fa però parte degli sviluppi futuri di questo Lavoro.

Compiuta quindi la caratterizzazione e localizzazione filogenetica del gene per la calcon sintasi in olivo si è proceduto alla seconda parte del lavoro e cioè alla analisi della variabilità varietale ed allo sviluppo di marcatori molecolari, ambedue utili per la protezione del germoplasma e, eventualmente per la MAS, in un gene che, come si è detto nella introduzione, potrebbe essere di grande interesse per il miglioramento di una serie di caratteri.

Come già discusso in parte precedentemente, infatti, la caratterizzazione genetica delle cultivar è importante per monitorare e certificare con precisione la qualità del prodotto di fascia alta costruendo una base scientifica di riferimento sia per questioni legate al mercato e alla promozione dei prodotti tipici, che per individuare eventualmente la presenza di tagli nel prodotto finito.

In questa seconda parte del lavoro è stata analizzata la variabilità varietale di 10 cultivar toscane, selezionate in base alla provenienza locale, alla diffusione e all'interesse agronomico, alla presenza nei tre disciplinari DOP e nel disciplinare IGP che descrivono e certificano i quattro oli di oliva di qualità tipici della regione Toscana, alla presenza di Componenti Minori Polari (CMP) nelle drupe, a

caratteristiche morfologiche e fenotipiche riconducibili alle funzioni correlate ai flavonoidi.

Analizzando i dati di sequenza è stato possibile dimostrare, per ciascuno dei tre geni analizzati, la presenza di patterns diversi di variabilità genetica nelle cultivars analizzate. Dalla analisi delle sequenze di *chs A*, *chs B* e *chs C* nelle 10 cultivar, si è notato che *chs B* è il gene più variabile, in termini di frequenza di mutazione, mentre *chs A* è il gene che mostra meno variabilità intervarietale. Questo dato potrebbe derivare da un diverso grado di pressione selettiva sui diversi membri della stessa famiglia genica.

E' interessante notare a questo riguardo che *chs B* presenta una zona altamente variabile che non è localizzata in corrispondenza delle sequenze introniche, ma comprende la parte in 5' dell'esone 2 mentre questa peculiarità non è riscontrabile in *chs A* e *chs C* in cui le zone di massima variabilità nucleotidica sono all'interno delle sequenze introniche, come ci si sarebbe atteso in base alla letteratura.

In alcuni casi si è osservato che la variabilità delle sequenze codificanti ha portato a cambiamenti della sequenza amminoacidica predetta *in silico*. Tali modificazioni non coinvolgono tuttavia in nessun caso i residui coinvolti nelle strutture critiche della proteina, in particolare nel sito attivo, nel coumaroyl binding pocket, nel cyclization pocket e negli altri residui che svolgono una funzione cruciale dal punto di vista strutturale. Questi residui, che si sono dimostrati estremamente conservati nelle proteine CHS fino ad oggi caratterizzate, sono invariati in tutte le cultivar analizzate.

Dall'albero filogenetico derivante dalla analisi di similarità complessiva delle sequenze dei geni *chs A*, *chs B* e *chs C* delle cultivar toscane, effettuata mediante la costruzione di un dendrogramma utilizzando l'algoritmo Neighbor-joining, è risultato che le tre cultivar caratteristiche dell'areale pistoiese clusterizzano assieme, un dato che non conferma le osservazioni pubblicate in alcuni lavori presenti in letteratura (Sanz-Courtés 2003) in cui non sono state trovate correlazioni fra la formazione di clusters e la distribuzione geografica delle cultivar studiate. Il nostro dato probabilmente deriva dal fatto che la olivicoltura pistoiese è particolarmente circoscritta, omogenea e isolata, se confrontata con gli areali fiorentini o pisani. E' possibile quindi che le tre varietà pistoiesi derivino tutte da un numero limitato di piante iniziali.

Il set di marcatori SNPs sviluppato nel presente lavoro potrà essere utilizzato sia in quanto tale, per reazioni di amplificazione e sequenziamento utilizzando i primer da noi disegnati e le condizioni di reazione da noi messe a punto, ma potrà anche servire come base scientifica per altre tecniche di genotipizzazione, ad esempio che prevedono una conoscenza preventiva dei polimorfismi SNPs, come la Tetra-primer ARMS-PCR.

Oltre a SNPs caratterizzanti le cultivar, sono state identificate in *chs A* tre inserzioni cultivar-specifiche di Mignolo Cerretano e di Leccino e di oltre a una delezione caratteristica delle cultivar Rossello, San Francesco e Leccino. Nelle sequenze di *chs B*, come detto maggiormente variabili, sono presenti numerose delezioni in parti non codificanti della sequenza.

L'unico caso di cambiamento del numero di amminoacidi presenti nella sequenza proteica predetta è riscontrabile nella sequenza di *chs B* di Rossello, nella quale sequenza genica codificante è presente una delezione di 3bp che porta alla delezione di un residuo nella sequenza amminoacidica. Nelle sequenze di *chs C* è stata scoperta una delezione di 3 bp nella cultivar. Infine le sequenze ottenute sono state utilizzate per continuare l'analisi della variabilità genetica all'interno della specie *Olea europaea var europaea* ed estendere la nostra valutazione alle 33 cultivar italiane esaminate.

Inoltre abbiamo esteso le nostre analisi all'interno di una particolare cultivar, Carolea, che presenta una diffusione molto ampia, per chiarire la eventuale presenza di polimorfismo intra-varietale, dimostrato in alcuni precedenti lavori (Muzzalupo et al, 2010).

Nella parte introduttiva alla tesi è stato descritto come attualmente sia presente una disputa scientifica sull'attribuzione tassonomica dell'olivo selvatico che viene identificato da alcuni autori come una specie a sé, da altri come una sub-specie, da altri ancora come una varietà.

E' noto inoltre che è spesso difficile distinguere olivi coltivati da quelli selvatici, e dalle accessioni inselvatichite che hanno revertito i loro caratteri fenotipici. D'altra parte, come vedremo in seguito, anche per la protezione del germoplasma italiano analisi come quelle descritte fino ad ora, che si basano sulla variabilità complessiva dei geni studiati, sono lente e costose mentre servirebbero metodi che accelerino le procedure pur mantenendo l'alto grado di risoluzione che è

necessario per la discriminazione fra diversi genotipi. Per queste ragioni abbiamo sviluppato una serie di marcatori molecolari a partire dalle sequenze precedentemente descritte che potranno anche essere potenzialmente utilizzati per la selezione assistita da marcatori (MAS)

Per quanto riguarda la discriminazione fra olivi coltivati e selvatici utilizzando una particolare PCR chiamata EPIC - PCR (exon - primed intron - crossing PCR) siamo stati in grado di individuare un marcatore molecolare ILPs (intron length polymorphism) in grado di discriminare le varietà coltivate dalla selvatica.

Oltre alla caratterizzazione delle varietà selvatiche e di quelle coltivate, questo approccio ci ha permesso di studiare eventuali casi di incrocio tra di esse dimostrando così la presenza dell'allele caratteristico delle forme selvatiche in numerose varietà coltivate indipendentemente dalla provenienza regionale. Va però notato che la percentuale di presenza dell'allele varS varia a seconda della provenienza regionale..

In ogni caso, questi risultati confermano i dati ottenuti da Erre e colleghi (2010), che, analizzando alcune cultivar caratteristiche della regione Sardegna e alcuni olivastri autoctoni hanno dimostrato che alcuni di essi condividono lo stesso pattern allelico.

Nel caso in esame si contraddice però l'ipotesi di Lumaret e colleghi (2004) che propongono che la variabilità presente nelle cultivar rappresenti un subset della variabilità genetica più ampia presente negli olivastri. L'analisi delle sequenze ha fatto emergere la presenza di marcatori SNPs in grado di caratterizzare alcune varietà testate.

Tuttavia i marcatori molecolari ora descritti non hanno sempre dato risultati positivi. Un esempio di questo è dato dalla cultivar Carolea che si è dimostrata completamente omogenea dal punto di vista genetico se saggiata con il primo set di marcatori.

Per affrontare meglio questo problema siamo dovuti ricorrere ad altri marcatori molecolari: gli ISSR, da utilizzare per la discriminazione inter ed intravarietale (Piagnani et al., 2002; Terzopoulos et al. 2005).

Inoltre l'utilizzo di questo tipo di marcatore molecolare ci ha permesso di valutare un altro importante aspetto dell'analisi della variabilità genetica della specie *Olea europea*: negli esperimenti descritti in precedenza abbiamo valutato insieme la

variabilità tra cultivar dello stesso areale e quella tra cultivar di diverse zone di origine e coltivazione.

Per chiarire il peso di ambienti geografici diversi abbiamo quindi scelto di quantificare la variabilità presente tra cultivar appartenenti a due areali molto particolari in quanto prossimi dal punto di vista geografico ma separati da una barriera fisica importante: Sicilia e Calabria. Inoltre abbiamo analizzato una accessione di *Olea europaea* var *Sylvestris* proveniente dalla regione Calabria.

Il risultato di questi esperimenti è stato innanzitutto la dimostrazione che i marcatori ISSR sono in grado di discriminare 7 delle 9 accessioni della cultivar Carolea, il che dimostra che il marcatore molecolare ILP messo a punto sull'introne 1 di *chs A* è in grado di discriminare tra cultivar e varietà selvatica ma non è sufficiente per una analisi di variabilità intravarietale. Questa è probabilmente dovuta alla diffusione ampia della cultivar come riportato in altri lavori (Cipriani et al 2002) ed è stato osservato in altre specie arboree come ad esempio la vite (Bellin et al. 2001).

Anche i metodi di propagazione delle cultivar potrebbero avere avuto un ruolo determinante nel generare tale variabilità genetica: ad esempio cloni con le stesse caratteristiche sia morfologiche che fisiologiche potrebbero essere stati prelevati e successivamente propagati a partire da branche differenti di grandi alberi secolari. Questi cloni, in tutto simili fenotipicamente possono tuttavia potenzialmente presentare un diverso genotipo.

Il dendrogramma costruito a partire dalla matrice di similarità ci mostra che le accessioni di carolea clusterizzano tutte assieme eccetto una, che si raggruppa in un diverso cluster- Per quanto riguarda l'accessione di olivo selvatico si noti che clusterizza insieme ad alcune cultivar siciliane e alla cultivar calabrese Ottobratica e non si separa dalle varietà coltivate. Questo in accordo con gli studi citati in precedenza che dimostrano che olivi coltivati e selvatici condividono gli stessi pattern genetici.

Infatti, analisi effettuate con diversi marcatori (ad esempio RAPD e AFLP) danno risultati analoghi a quelli da noi ottenuti che suggeriscono che cultivar originarie di areali vicini non si separino in base alla provenienza geografica. Angiolillo e collaboratori (1999) invece dimostrano una eccezione, non confermata dal nostro lavoro, per le cultivar siciliane, che tendono invece a separarsi dalle cultivar provenienti da areali vicini. Infine gli studi riportati in questa tesi hanno

contribuito a risolvere una falsa sinonimia: le cultivar Moresca and Biancolilla, considerate sinonime nel database oleadb, presentano patterns molecolari differenti e in conseguenza sono localizzate in due differenti raggruppamenti.

Capitolo 6

Bibliografia

- Acerbo G., 1937. La marcia storica dell'olivo nel Mediterraneo. Atti della Società per il Progresso delle Scienze, Riun. XXV, Vol. I, Fasc. 2. 1-22.
- Ajibade S.R., Weeden N.F., Chite S.M., 2000. Inter Simple Sequence Repeat analysis of genetic relationship in genus *Vigna*. *Euphytica* 111. 47-55.
- Amane M., Lumaret R., Hany V., Ouazzani N., Debain C., Vivier G., Deguilloux M. F., 1999. Chloroplast-DNA variation in cultivated and wild olive *Olea europaea* L.. *Theor Appl Genet*, 99. 133-139.
- Amaral J., Rocha M.L., 1972. *Olea* L. In: Tutin TG, Heywood VH, Burges NA, Valente DH, Walters M, Webb DA eds *Flora Europaea*, Vol 3. Cambridge University Press, Cambridge, 370.
- Angiolillo A., Mencuccini M., Baldoni L., 1999. Olive genetic diversity assessed using amplified fragment length polymorphisms. *Theor Appl Genet* 98. 411–421.
- Austin M.B., Noel J.P., 2003. The chalcone synthase superfamily of type III polyketide synthases, *Nat. Prod. Rep.* 20 79-110.
- Bais H.P., Vepachedu R., Gilroy S., Callaway R.M., Vivanco J.M., 2003. Allelopathy and exotic plant invasion. from molecules and genes to species interaction. *Science* 301.1377–1380.

Baldoni L., Tosti N., Ricciolini C., Belaj A., Arcioni S., Pannelli G., 2006. Genetic structure of wild and cultivated olives in the Central Mediterranean Basin. *Ann.Bot.*, 98. 935-942.

Barceló J., Poschenrieder C., 2002. Fast root growth responses, root exudates, and internal detoxification as clues to the mechanism of aluminium toxicity and resistance. a review. *Environ Exp Bot* 48.75–92.

Barranco, D., Rallo L., 2000. Olive cultivars in Spain. *HortTechnology* 10.107–110.

Bellin, D., Velasco, R. and Grando, M.S. 2001. Intravarietal DNA polymorphisms in grapevine *Vitis vinifera* L. *Acta Hort.*, 546. 343-349.

Berhow M.A., Vaughn S.F. 1999. Higher plant flavonoids. biosynthesis and chemical ecology. In. Inderjit, Dakshini KMM, Foy CL eds *Principles and practices in plant ecology*. CRC Press, Boca Raton, 423–438.

Besnard G., Baradat P., Berville A., 2001. Genetic relationships in the olive *Olea europaea* L. reflect multilocal selection of cultivars. *Theoretical and Applied Genetics* 102. 251–258.

Besnard G., Khadari B., Baradat P., Bervillè A., 2002. *Olea europea* Oleaceae phylogeography based on chloroplast DNA polymorphism. *Theor. Appl. Gen.*; 104. 1353-1361.

Besnard, G., Berville A., 2000. Multiple origins for mediterranean olive *Olea europaea* L. ssp. *europaea* based upon mitochondrial DNA polymorphisms. *C R Acad Sci Ser III.Vie* 323.173–181.

Blanc, G., Wolfe, K.H., 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16, 1679–1691.

Bogani P., Cavalieri D., Petruccelli R., Polsinelli L., Roselli G., 1994. Identification of olive tree cultivars by using random amplified polymorphic DNA. *Acta Hort*, 356. 98-101.

Breton C., Besnard G., Berville A.A., 2006. Using multiple types of molecular markers to understand olive phylogeography. in. Zeder MA, Bradley DG,

Emshwiller E, Smith BD eds Documenting domestication. new genetic and archeological paradigms. University of California Press, California, 143–152.

Brignolas F., Lieutier F., Sauvard D., Christiansen E., Berryman A.A., 1998. Phenolic predictors for Norway spruce resistance to the bark beetle *Ips typographus* Coleoptera. Scolytidae and an associated fungus, *Ceratocystis polonica*. *Can J For Res* 28.720–728.

Bronzini de Caraffa V., Maury J., Gambotti C., Breton C., Berville A., Giannettini J., 2002. Mitochondrial DNA variation and RAPDmark oleasters, olive and feral olive from Western and Eastern Mediterranean. *Theoretical and Applied Genetics* 104. 1209–1216.

Broughton W.J., Zhang F., Perret X., Staehelin A., 2003. Signals exchanged between legumes and *Rhizobium*. agricultural uses and perspectives. *Plant Soil* 252.129–137.

Chalker-Scott L., Krahmer R.L., 1989. Microscopic studies of tannin formation and distribution in plant tissues. In. Hemingway RW, Karchesy JJ eds *Chemistry and significance of condensed tannins*, Plenum Press, New York, 345–368.

Chen K., Ohmura W., Doi S., Aoyama M., 2004. Termite feeding deterrent from Japanese larch wood. *Bioresource Technol* 95.129–134.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G. and Thompson, J.D. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.*, 31, 3497–3500.

- Cipriani G., Marrazzo M.T., Marconi R., Cimato A., Testolin R. 2002. Microsatellite markers isolated in olive *Olea europaea* L. are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. *Theor Appl Genet*, 104. 223-228.
- Clegg M.T., Durbin M.L., 2003. Tracing floral adaptations from ecology to molecules. *Nat Rev Genet*. 4.206–215.
- Coberly, L.C., Rausher, M.D., 2003. Analysis of a chalcone synthase mutant in *Ipomoea purpurea* reveals a novel function for flavonoids. amelioration of heat stress. *Mol. Ecol*. 12, 1113–1124.
- Cooper J.E., 2004. Dominant role of flavonoids among signal molecules involved in the formation of legume-rhizobia symbiosis. *Polyphenol Commun* 2004.87.
- Corpet F., 1988. Multiple sequence alignment with hierarchical clustering, *Nucl. Acids Res*. 16.10881-10890.
- Costa C., 1998. Olive production in South Africa. A handbook for olive growers. Agricultural Research Council, Pretoria, 124.
- Crepieux S., Lebreton C., Flament P., Charmet G., 2005. Application of a new IBD-based QTL mapping method to common wheat breeding population. analysis of kernel hardness and dough strength. *Theor Appl Genet*, 111.1409-1419.
- de Bruyne, M., Clyne, P., and Carlson, J. 1999. Odor coding in a olfactory organ. the *Drosophila* maxillary palp. *J. Neurosci*. 19, 4520–4532.
- De Caraffa V.B., Maury J., Gambotti C., Breton C., Berville A., Giannettini J., 2002. Mitochondrial DNA variation and RAPD mark oleasters, olive and feral olive from Western and Eastern Mediterranean. *Theor Appl Genet* 1046–7. 1209–1216.

Diaz, A., De la Rosa R., Martin A., Rallo P., 2006. Development, characterization and inheritance of new microsatellites in olive *Olea europaea* L. and evaluation of their usefulness in cultivar identification and genetic relationship studies. *Tree gen. genom.* 3 165-175.

Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12.13-15.

Durbin M.L., Learn G.H., Huttley G.A., Clegg M.T., 1995. Evolution of the chalcone synthase gene family in the genus *Ipomoea*. *Proc Natl Acad Sci USA.* 92.3338–3342.

Durbin, M. L., Denton, A. L., Clegg, M. T., 2001. Dynamics of mobile elements activity in chalcone synthase loci in the common morning glory *Ipomoea purpurea*, *Proc. Natl. Acad. Sci. USA*, 98. 5084-5089.

Edwards, M.L., Stemerick, D.M. & Sunkara, P.S., 1990. Chalcones. a new class of antimitotic agents. *J. Med. Chem.* 33, 1948–1954.

Erre P., Chessa I., Munoz-Diez C., Belaj A., Rallo L., Trujillo I., 2010. Genetic diversity and relationships between wild and cultivated olives *Olea europaea* L. in Sardinia as assessed by SSR markers. *Genet Resour Crop Evol* 57.41–54.

Essadki M., Ouazzani N., Lumaret R., Moumni M., 2006. Issr variation in olive-tree cultivars from Morocco and other western countries of the Mediterranean Basin. *Genet. Resour. Crop Evol.* 53. 475-482.

Fabbri A., Hormazy J. J., Polito V.S., 1995. Random Amplified Polymorphic DNA Analysis of Olive *Olea europaea* L. cultivars. *J. Amer. Soc. Hort. Sci* 120.538-542.

Farzad M., Griesbach R., Hammond J., Weiss M.R., Elmendorf H.G., 2003. Differential expression of three key anthocyanin biosynthetic genes in a color-

changing flower, *Viola cornuta* cv. Yesterday, today and tomorrow. *Plant Sci.* 165.1333–1342.

Feeny P.P., 1976. Plant apparancy and chemical defense. *Rec Adv Phytochem* 10.1–40.

Ferrer J., Jez J.M., Bowman M.E., Dixon R.A., Noel J.P., 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis, *Nature* 6 775-784.

Ferrer, M., Golyshina, O., Beloqui, A., and Golyshin, P.N., 2007. Mining enzymes from extreme environments. *Curr Opin Microbiol* 10. 207–214.

Fliegmann J., Schroder G., Schanz S., Britsch L., Schroder J., 1992. Molecular analysis of chalcone and dihydropinosylvin synthase from Scots pine *Pinus sylvestris*, and differential regulation of these and related enzyme activities in stressed plants. *Plant MolBiol* 18.489–503.

Frohman M., Dush M., Martin G., 1988. Rapid production of full-length cDNAs from rare transcripts. amplification using a single genespecific oligonucleotide primer. *Proceeding of the National Academic of Science of the United States of America*, 23, 8998-9002.

Galili G., Galili S., Lewinsohn E., Tadmor Y., 2002. Genetic, molecular, and genomic approaches to improve the value of plant foods and feeds. *Crit Rev Plant Sci* 21. 167-204.

Gehm, B.D., McAndrews, J.M., Chien, P.-Y. Jameson, J.L. 1997. Resveratrol, a polyphenolic compound found in grapes and wine, is an agonist for the estrogen receptor. *Proc. Natl. Acad. Sci. USA* 94, 14138–14143.

Gemas V. J., Rijo-Johansen M. J., Tenreiro R., Fevereiro P., 2000. Inter- and intra-varietal analysis of three *Olea europaea* L. cultivars using the RAPD technique. *J. Hortic. Sci. Biotech.*, 75. 312-319.

Gleiss, P.M., Stadler, P.F., Wagner, A., Fell, D.A., 2001. Small cycles in small worlds. *Adv. Complex Syst.* 4, 207–226.

Green PS 2002 A revision of *Olea* L. *Oleaceae*. *Kew Bull* 57. 91–140

Gulick A.M., Starai V.J., Horswill A.R., Homick K.M., Escalante-Semerena J.C., 2003. The 1.75 Å crystal structure of acetyl-coA synthetase bound to adenosine-50-propylphosphate and coenzyme A, *Biochemistry* 42 10 2866-2873.

Hamman-Khalifa A., Navajas-Perez, de la Herra R., Ruiz Rejo Garrido-Ramos M., Ruiz Rejo C., Rosu J., 2007. Establishing the genetic relationships between the wild and cultivated olives using a nuclear intron from nitrate reductase *nia-i3*. *Pl Syst Evol* 269. 63–73.

Haribal M., Feeny P., 2003. Combined roles of contact stimulant and deterrents in assessment of host-plant quality by ovipositing zebra swallowtail butterflies. *J Chem Ecol* 29.653–670.

Harrison M.J., Dixon R.A., 1994. Spatial pattern of expression of flavonoid/isoflavonoid pathway genes during interactions between roots of *Medicago truncatula* and the mycorrhizal fungus *Glomus versiforme*. *Plant J* 6.9–20.

Herrera, C.M., 1995. Plant-vertebrate seed dispersal systems in the Mediterranean. ecological, evolutionary, and historical determinants. *Annual Review of Ecology and Systematics*, 26, 705 – 727.

Hess J., Kadereit W., Vargas P., 2000. The colonization history of *Olea europaea* L. in Macaronesia based on internal transcribed spacer 1 ITS-1 sequences, randomly amplified polymorphic DNAs RAPDs, and intersimple sequence repeats ISSRs. *Mol. Ecol*, 9. 857-868.

Hoffmann-Campo C.B., Harborne J.B., McCaffery A.R., 2001. Preingestive and post-ingestive effects of soya bean extracts and rutin on *Trichoplusia ni* growth. *Entomol Exp Appl* 98.181–194.

- Hooper, S.D., Berg, O.G., 2003. On the nature of gene innovation. duplication patterns in microbial genomes. *Mol. Biol. Evol.* 20, 945–954.
- Hoshino, A., Johzuka-Hisatomi, Y., Iida, S., 2001. Gene duplication and mobile genetic elements in the morning glories, *Gene*, 2001, 265. 1-10.
- Howles P.A., Arioli T., Weinman J.J., 1995. Nucleotide sequence of additional members of the gene family encoding chalcone synthase in *Trifolium subterraneum*. *Plant Physiol* 107.1035–1036.
- Huang X. and Miller W., 1991. A Time-Efficient, Linear-Space Local Similarity Algorithm, *Advances in Applied Mathematics*, 12, 337-357.
- Hungria M., Stacey G., 1997. Molecular signals exchanged between host plants and rhizobia. basic aspects and potential application in agriculture. *Soil Biol Biochem* 29.819–830.
- Hwang E.I., Kaneko M., Ohnishi Y., Horinouchi S., 2003. Production of plant specific flavanones by *Escherichia coli* containing an artificial gene cluster, *Appl. Environ. Microbiol.* 69 5 2699-2706.
- Ito, M., Ichinose, Y., Kato, H., 1997. Molecular evolution and functional relevance of the chalcone synthase genes of pea, *Mol. Gen. Genet.*, 1997, 255. 28-37.
- Jaillon O., Aury J.M., Noel B., et al. (56 co-autori). 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 449.463–468.
- Jang, M.. 1997. Cancer chemopreventive activity of resveratrol, a natural product derived from grapes. *Science* 275, 218–220.

Jinxia H., Hongya G., Yang Z., Ziheng Y., 2002. Duplication and adaptive evolution of the chalcone synthase genes of *Dendranthema* Asteraceae. *Mol Biol Evol* 19.1752–1759.

Johnson K. S., Felton G. W., 2001. Plant phenolics as dietary antioxidants for herbivorous insects. A test with genetically modified tobacco. *J. Chem. Ecol.* 27.2579Y2597.

Joshi S.P., Gupta V.S., Aggarwal R.K., Ranjekar P.K., Brar D.S., 2000. Genetic diversity and phylogenetic relationship as revealed by Inter Simple Sequence Repeat ISSR polymorphism in the genus *Oryza*. *Theor Appl Genet*, 100.1311-1320.

Kobayashi H., Naciri-Graven Y., Broughton W.J., Perret X., 2004. Flavonoids induce temporal shifts in gene-expression of nod-box controlled loci in *Rhizobium* sp. NGR234. *Mol Microbiol* 51.335– 347.

Koes, R., Verweij, W., Quattrocchio, F., 2005. Flavonoids. a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* 10, 236–242.

Koes, R.E., Spelt, C.E., van den Elzen, P., 1989. Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida*. *Gene* 81, 245–257.

Kong J.M., Chia L.S., Goh N.K., Chia T.F., Brouillard R., 2003. Analysis and biological activities of anthocyanins. *Phytochemistry* 64.923–933.

Li, R. 1995. In vitro antimalarial activity of chalcones and their derivatives. *J. Med.Chem.* 38, 5031–5037.

Liphschitz N., Gophna R., Hartman M., Biger G., 1991. The beginning of *Olea europaea* L. Cultivation in the old world. a reassessment. *J Archaeol Sci* 18. 441–453.

Lukacin, R., Springob, K., Urbanke, C., 1999. Native acridone synthases I and II from *Ruta graveolens* L. form homodimers, *FEBS Lett.*, 1999, 448. 135-140.

Lumaret R., Ouazzani N., Michaud H., Vivier G., Deguilloux M.F., Di Giusto F., 2004. Allozyme variation of oleaster populations wild olive tree *Olea europaea* L. in the Mediterranean Basin. *Heredity* 92.343–351.

Lumaret R., Amane M., Ouazzani N., Baldoni L., Debain C., 2000. Chloroplast DNA variation in the cultivated and wild olive taxa of the genus *Olea* L. *Theor. Appl. Genet.* 101.547-553.

Lumaret R., Ouazzani N., Michaud H. and Villemur P. 1997. Cultivated olive and oleaster. two very closely connected partners of the same species *Olea europaea* L. evidence from enzyme polymorphism. *Boccone* 7. 39–42.

Lumaret, R. & Ouazzani, N., 2001. Ancient wild olives in Mediterranean forests. *Nature* 413. 700.

Lynch, M., 2002. Gene duplication and evolution. *Science* 297, 945–947.

Marchese A., Bošković R. Caruso T., Tobutt K. 2010. Intra-allelic variation in introns of the S13-RNase allele distinguishes sweet, wild and sour cherrie. *Tree Genetics & Genomes*.

Marles M.A.S., Ray H., Gruber M.Y., 2003. Newperspectives on proanthocyanidin biochemistry and molecular regulation. *Phytochemistry* 64.367–383.

Marnellos G., 2003. High-throughput SNP analysis for genetic association studies. *Curr. Opin. Drug. Discov. Devel.* 6. 317–321.

Martins-Lopes P., Lima-Brito J., Gomes S., Meirinhos J., Santos L., Guedes-Pinto H., 2007. RAPD and ISSR molecular markers in *Olea europaea* L. genetic variability and cultivar identification. – *Genet. Resour. Crop. Evol.* 54.117–128.

Moore J.P., Westall K.L., Ravenscroft N., Farrant J.M., Lindsey G.G., Brandt W.F., 2004. The predominant polyphenol in the leaves of the resurrection plant *Myrothamnus flabellifolius*, 3,4,5 tri-O-galloylquinic acid, protects membranes against desiccation and free radical induced oxidation. *Polyphenol Commun* 20 04-79.

Moore, R. C., Purugganan M., 2003. The early stages of duplicate gene evolution. *Proc. Natl. Acad. Sci. USA* 100. 15682– 15687.

Moore, R.C., Purugganan, M.D., 2003. The early stages of duplicate gene evolution. *Proc. Natl. Acad. Sci. USA* 100, 15682–15687.

Morettini A., 1972. *Olivicoltura*. REDA, Roma.

Muleo R., Colao M. C., Miano D., Cirilli M., Intrieri M. C., Baldoni L., E. Rugini, 2009. Mutation scanning and genotyping by high-resolution DNA melting analysis in olive germplasm. *Genome* 52. 252–260.

Mullis K., Faloona F., Scharf S., Saiki R., Horn G., Erlich H., 1986. Specific enzymatic amplification of DNA in vitro. the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol.* ;51 Pt 1.263–273.

Muzzalupo I., Chiappetta, A., Benincasa C., Perri E. 2010. Intra-cultivar variability of three major olive cultivars grown in different areas of central-southern Italy and studied using microsatellite markers *Scientia Horticulturae* 126. 324-329.

Nagaoka T., Ogihara Y., 1997. Applicability of intersimple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theor. Appl. Genet*, 94. 597-602.

Ness, A., 2002. Commentary. Is olive oil a key ingredient in the Mediterranean recipe for health? *International Journal of Epidemiology* 31 2, 481– 482.

Nestle, M., 1995. Mediterranean diets. historical and research overview. *The American Journal of Clinical Nutrition* 61 6, 1313S–1320.

Ohno S., *Evolution by Gene Duplication*, Springer-Verlag, Berlin, New York, 1970.

Paniego, N. B., Zuurbier, K. W. M., Fung, S., 1999. Phlorisovalero-phenone synthase, a novel polyketide synthase from hop *Humulus lupulus* L. cones, *Eur. J. Biochem.*, 1999, 262. 612-616.

Pasqualone A., Caponio F., Blanco A., 2001. Inter-Simple Sequence Repeat DNA markers for identification of drupes from different *Olea europaea* L. cultivars. – *Eur. Food Res. Technol.* 213.240–243.

Pignatti S., 1982. *Flora D'Italia*, Vol 2. Edagricole, Bologna, 732.

Pizzi A., Cameron F.A., 1986. Flavonoid tannins—structural wood components for drought-resistant mechanisms of plants. *Wood SciTechnol* 20.119.

Ponce M.A., Scervino J.M., Erra-Balsells R., Ocampo J.A., Godeas A.M., 2004. Flavonoids from shoots and roots of *Trifolium repens* whiteclover grown in presence or absence of the arbuscularmycorrhizal fungus *Glomus intraradices*. *Phytochemistry* 65.1925–1930.

Prasad M., Varshney R.K., Roy J.K., Balyan H.S., 2000. The use of microsatellites for detecting DNA polymorphisms, genotype identification and genetic diversity in wheat. *Theor Appl Genet*, 100.584-592.

Rakoczy–Trojanowska M. and Bolibok H., 2004. Characteristics and a comparison of three classes of microsatellite-based markers and their application in plants. *Cell. Mol. Biol. Lett.* 9. 221–238.

Rallo P., Tenzer I., Gessler C., Baldoni L., Dorado G., Marti'n A., 2003. Transferability of olive microsatellite loci across the genus *Olea*. *Theor Appl Genet* 107. 940–946.

Reale S., Doveri S., Di'az A., Angiolillo A., Lucentini L., Pilla F., Marti'n A., Donini P., Lee D., 2006. SNP-based markers for discriminating olive *Olea europaea* L. cultivars. *Genome* 49. 1193–1205.

Rhizopoulou S., 2007. *Olea europaea* L. A Botanical Contribution to Culture American-Eurasian *J. Agric. & Environ. Sci.*, 2 4. 382-387.

Rotondi A., Magli M., Ricciolini C., Baldoni L., 2003. Morphological and molecular analyses for the characterization of a group of Italian olive cultivars. *Euphytica* 132. 129–137.

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X. and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19. 2496-2497.

Salisbury, B. A., Pungliya, M., Choi, J. Y., Jiang, R., Sun, X. J., and Stephens, J. C. 2003. SNP and haplotype variation in the human genome. *Mutat. Res.* 526.53–61.

Sankawa U., 1999. Polyketides and other secondary metabolites including fatty acids and their derivatives, in: D.H.R. Barton, K. Nakanishi, O. Meth-Cohn Eds., *Comprehensive Natural Products Chemistry*, vol. 1, Elsevier Science Ltd., Oxford, UK.

Scheidel A., Krausmann F., 2010. Diet, trade and land use. a socio-ecological analysis of the transformation of the olive oil system. *Land Use Policy* 28 47–56.

Schmelzer E., Jahnen W., Hahlbrock K., 1988. In situ localization of light-induced chalcone synthase mRNA, chalcone synthase, and flavonoid end products in epidermal cells of parsley leaves. *Proc Natl Acad Sci U S A.* May;85.2989–2993.

Schröder J., 1997. A family of plant-specific polyketide synthases. facts and predictions. *Trends Plant Sci* 2. 373-378.

Sefc K. M., Lopes M. S., Mendonça D., Rodrigues Dos Santos M., Laimer Da Câmara Machado M., Da Câmara Machado A., 2000. Identification of microsatellite loci in olive *Olea europaea* and their characterization in Italian and Iberian olive trees. *Mol. Ecol.* 9.1171-1173.

Shimada S., Ogawa T., Kitagawa S., Suzuki T., Ikari C., Shitsukawa N., Abe T., Kawahigashi H., Kikuchi R., Handa H., Murai K., 2009. A genetic network of Xowering time genes in wheat leaves, in which an APETALA1/FRUITFULL-like gene, VRN1, is upstream of FLOWERING LOCUS T. *Plant J* 58.668–681.

Sommer, H. Saedler H., 1986. Structure of the chalcone synthase gene of *Antirrhinum majus*, *Mol. Gen. Genet.* 202 429–434.

Sparvoli F., Martin C., Scienza A., Gavazzi G., Tonelli C., 1994. Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape *Vitis vinifera* L. *Plant Mol Biol.* 24.743–755.

Stafford H., 1991. Flavonoid evolution. An enzymic approach, *Plant Physiol.* 96 680e685.

Stein N., Graner A., 2004. Map-based gene isolation in cereal genomes. In *Cereal genomics* Edited by. Gupta PK, Varshney RK. Kluwer. Dordrecht;.331-360.

Strand, A. E., 1997. Inte ratin population genetics anddemography in *Aquiegia*. fh. D. dissertation, Department of fisiology, new Mexico State University, Las Cruces.

Tamura K., Nei M., Kumar S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences USA* 101.11030-11035.

Tattini M., Galardi C., Pinelli P., Massai R., Remorini D., Agati G., 2004. Differential acumulation of flavonoids and hydroxycinnamates in leaves of

Ligustrum vulgare under excess light and drought stress. *New Phytol* 163.547–561.

Terzopoulos P.J., Kolano B., Bebelia P.J., Kaltsikes P.J., Metzidakis I., 2005. Identification of *Olea europaea* L. cultivars using inter-simple sequence repeat markers. – *Sci. Hortic.* 105. 45-51.

Thoison O., S'évenet T., Niemeyer H.M., Russell G.B., 2004. Insect antifeedant compounds from *Nothofagus dombeyi* and *N. pumilio*. *Phytochemistry* 65.2173–2176.

Tropf, S., Lanz, T., Rensing, S. A. 1994. Evidence that stibene synthases have developed from chalcone synthases several times in the course of evolution, *J. Mol. Evol.*, 1994, 38. 610-618.

Turrill, W.B., 1951. *Modern Trends in The Classification of Plants.* Taxon 1 2.

Tuskan GA, DiFazio S, Jansson S, et al. (104 co-authors), 2006. The genome of black cottonwood, *Populus trichocarpa* Torr. & Gray. *Science.* 313.1596–1604.

Vargas, P., Kadereit J.W. 2001. Molecular fingerprinting evidence ISSR for a wild status of *Olea europaea* L. Oleaceae in the Eurosiberian North of the Iberian Peninsula. *Flora* 196. 142–152.

Varshney R.K., Sigmund R., Borner A., Korzun V., Stein N., Sorrells M.E., Langridge P., Graner A., 2005. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Sci.* 168.195-202.

Wade H., Bibikova T.N., Valentine, W.J., Henkins, C.I., 2001. Interactions within a network of phytochrome, cryptochrome and UV-B phototransduction pathways regulate chalcone synthase gene expression in *Arabidopsis* leaf tissue. *Plant J.* 25, 675–685.

Wang W., Schaal B., Chioua Y., Murakami M., Ge X., Huang C., Chiang T., 2007. Diverse selective modes among orthologs/paralogs of the chalcone synthase Chs gene family of *Arabidopsis thaliana* and its relative *A. halleri* ssp. *Gemmifera* *Molecular Phylogenetics and Evolution* 44 503–520.

Whiting D.A., 2001. Natural phenolics compounds 1900–2000. a bird's eye view of a century's chemistry, *Nat. Prod. Rep.* 18 583–606.

Widstrom N.W., Snook M.E., 2001. Recurrent selection for maysin, a compound in maize silks, antibiotic to earworm. *Plant Breed* 120.357–359.

Wilbur, W. J. and Lipman, D. J. 1983. Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA*, 80. 726-730.

Winkel-Shirley, B., 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126, 485–493.

Yang, J., Gu, H., Yang, Z., 2004. Likelihood analysis of the chalcone synthase genes suggests the role of positive selection in morning glories *Ipomoea*, *J. Mol. Evol.*, 2004, 58. 54-63.

Yasuyo S., Praveen R.J., Isao F., Katsuhiko K., 2005. Discovery of a novel superfamily of type III polyketide synthases in *Aspergillus oryzae*. *Biochem Biophys Res Commun* 331.253–260.

Yilmaz Y., Toledo R.T., 2004. Health aspects of functional grape seed constituents. *Trends Food Sci Technol* 15422–33.

Zhang, L., Peek, A.S., Dunams, D., Gaut, B.S., 2002. Population genetics of duplicated disease-defense genes, *hm1* and *hm2*, in maize *Zea mays* ssp. *Mays L.* and its wild ancestor *Zea mays* ssp. *parviglumis*. *Genetics* 162, 851–860.

Zohary D., Hopf M., 1993. Domestication of plants in the old world. The origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley. Oxford. Clarendon Press.

Zohary D., Hopf M., 2000. Domestication of plants in the old world. the origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley, 3rd edn. Oxford University Press, Oxford.

Zohary D., Spiegel-Roy P. 1975. Beginnings of fruit growing in the Old World. Science, 187 4174. 319-327.

Zohary D., Spiegel-Roy P., 1994. Domestication of Plants in the Old World second edition. Clarendon, Oxford. 279.

Zohary M., 1973. Geobotanical foundations of the Middle East. Vol.1, Gustav Fischer Verlag, Stuttgart, Germany.

Zwaagstra, M.E., 1997. Synthesis and structure–activity relationships of carboxylated chalcones. J. Med. Chem. 40, 1075–1089.