

## Identificazione *in silico* di polimorfismi SNP in sequenze espresse di arancio dolce e clementine

Concetta Licciardello<sup>1</sup>, Jamila Bernardi<sup>2</sup>, Raffaele Mazza<sup>2</sup>, Giuseppe Reforgiato Recupero<sup>1</sup> e Adriano Marocco<sup>2\*</sup>

<sup>1</sup> Consiglio per la Ricerca e Sperimentazione in Agricoltura - ACM Centro di Ricerca per l'Agrumicoltura e le Colture Mediterranee, Acireale (CT)

<sup>2</sup> Istituto di Agronomia, Genetica e Coltivazioni erbacee, Università Cattolica del Sacro Cuore, Piacenza

### ***In silico* SNP detection in sweet orange and clementine EST sequences**

**Abstract.** Sweet orange [*Citrus sinensis* (L.) Osbeck] and clementine (*C. clementina* Hort. ex Tan.), limited to Mediterranean area, represent the most important grown citrus species. Cultivars, belonging to these two species, are difficult to recognize as they have moderate level of diversity, due to their origin from spontaneous mutation or nucellar selection. Despite the great number of sequences available and the draft assembly of sweet orange and clementine, there are not currently SNPs databases to be exploited for genotyping. For this purpose the QualitySNP pipeline was used to discover potential and reliable SNPs using a stringent cleaning procedure. A total of 3,634 EST sequences were associated to enzymes belonging to 121 metabolic pathways. The highest number of EST sequences belongs to secondary metabolite pathway, followed by the biosynthesis of plant hormones and biosynthesis of phenylpropanoids. *C. clementina* and *C. sinensis* SNP analysis produced 9,017 and 6,862 reliable SNP respectively. A selection of 163 SNPs were validated. From the genotyping analysis 78 polymorphic sites were found, 36 out them were heterozygous. The base variations were found between species (33 sweet orange against 17 clementine), while no variability was discovered between navel and blood orange groups. Furthermore, a selection of SNPs were further validated by sequencing.

**Key words:** *Citrus sinensis*, *Citrus clementina*, single nucleotide polymorphism, EST.

### Introduzione

Gli agrumi costituiscono una delle specie da frutto più ampiamente coltivata nel mondo, la cui produzione globale nel 2009-2010 è risultata essere 7,4 milioni di tonnellate. I maggiori Paesi produttori sono Cina,

Brasile, Stati Uniti, Spagna e Italia. Tra le varie specie, l'arancio dolce [*Citrus sinensis* (L.) Osbeck] ed il clementine (*C. clementina* Hort. ex Tan.), limitatamente all'area Mediterranea, determinano le maggiori superfici investite. Il maggiore interesse per queste due specie ha determinato l'isolamento di mutazioni migriorative, producendo una notevole diversificazione ed ampliamento del germoplasma, soprattutto per caratteristiche di importanza agronomica riguardanti principalmente il frutto.

In *Citrus* l'utilizzo di marcatori molecolari è stato oggetto in passato di vari tentativi per caratterizzare e distinguere le diverse specie e cultivar. Lo studio della tassonomia è risultato complesso, perché molte specie di agrume sono state coltivate per migliaia di anni, presentano in molti casi compatibilità sessuale, hanno un alto livello di embrionia nucellare ed un'elevata frequenza di mutazioni somatiche, che determinano una base genetica ristretta nelle diverse cultivar di una specie.

I microsatelliti o SSR (*Simple Sequence Repeats*), caratterizzati da un'elevata variabilità e riproducibilità, sono stati utilizzati per verificare la diversità genetica e la struttura della popolazione in una collezione di germoplasma agrumicolo (Barkley *et al.*, 2006; Jannati *et al.*, 2009), anche se erano state escluse dallo studio le accessioni originate da mutazione per la conseguente bassa diversità genetica (Barkley *et al.*, 2006). In uno studio per verificare l'efficienza di tre differenti marcatori molecolari al fine di caratterizzare la diversità genetica del germoplasma agrumicolo, un più alto livello di polimorfismo si era rilevato per marcatori SSR e SRAP (*Sequence Related Amplified Polymorphism*), rispetto ai CAPS-SNP (*Cut Amplified Polymorphic Sequence-Single Nucleotide Polymorphism*) (Hamdy Amar *et al.*, 2011). Recentemente l'utilizzo di STS (*Sequence Tagged Sites*) è stato rivolto all'individuazione di SNP (*Single Nucleotide Polymorphism*), al fine di verificare la potenzialità di questi marcatori per studi di filogenesi (Novelli *et al.*, 2004).

\* adriano.marocco@unicatt.it

Licciardello *et al.*

Gli SNP rappresentano la forma di polimorfismo più frequente e diffusa nel genoma di piante e animali. Nonostante la loro frequenza aumenti nelle sequenze non codificanti, possono trovarsi anche all'interno di geni. Diversi studi hanno utilizzato i marcatori SNP per discriminare cultivar con tratti agronomici di interesse, come quelli condotti su orzo (Close *et al.*, 2009), riso (McNally *et al.*, 2009), mais (Van Inghelandt *et al.*, 2011), frumento (Trebbi *et al.*, 2011), agrumi (Jiang *et al.*, 2010), patata (Hamilton *et al.*, 2011), melo (Micheletti *et al.*, 2011) e soia (Chaisan *et al.*, 2010). I genomi di arancio e di clementine sono stati sequenziati alla fine del 2010 e sono consultabili dalla comunità scientifica (<http://www.phytozome.net/Citrus.php>). Nonostante ciò solo un basso numero di SNP è disponibile e ad oggi non è stato ancora chiarito il potere discriminatorio di questa classe di marcatori sulle diverse cultivar.

Nel presente lavoro viene implementato un approccio *in silico* con lo scopo di identificare SNP da sequenze espresse (EST). Un numero selezionato di SNP è stato validato in 50 accessioni di *C. sinensis* (33) e *C. clementina* (17), allo scopo di verificare la loro capacità come marcatori funzionali, in grado di distinguere la variabilità inter- e intra-specifica.

## Materiali e metodi

Tutte le EST disponibili in banca dati NCBI appartenenti a 20 specie di *Citrus* (tra cui *C. sinensis* e *C. clementina*) sono state utilizzate per l'analisi bioinformatica. Sono stati eliminati i vettori associati alle sequenze con cross-match e le EST di ciascuna specie sono state assemblate in contig. Per ciascun contig sono state identificate le ontologie associate alla proteina grazie al software Blast2GO e gli enzimi sono stati raggruppati nelle vie metaboliche. Le sequenze di ciascun contig sono state analizzate con algoritmo implementato QualitySNP che individua i putativi SNP e ricostruisce gli aplotipi.

Un totale di 163 SNP, 105 di arancio e 58 di clementine, sono stati validati con il metodo Kaspar (Kbioscience). I dati sono stati visualizzati grazie al software SNPviewer, che permette di distinguere gli alleli in base alla loro distribuzione in un piano ortogonale. Alcuni SNP sono stati ulteriormente validati mediante sequenziamento.

## Risultati e discussione

### Procedura bioinformatica per l'identificazione di SNP

La ricerca degli SNP è stata condotta utilizzando una procedura *in silico* mediante il software

QualitySNP, partendo dalle sequenze EST di tutte le specie di *Citrus* presenti in banca dati. Circa 550.000 EST sono state allineate alle corrispondenti 25.000 sequenze di Unigene utilizzando criteri di alta stringenza. L'analisi degli aplotipi è stata usata per identificare e scartare i geni paraloghi; ciascun SNP individuato è stato selezionato per un *confidence score* (basato sulla regione ad esso fiancheggiante). Successivamente ogni contig, contenente almeno un putativo SNP, è stato opportunamente annotato. Sono stati scelti geni le cui annotazioni ricadevano in 21/22 vie metaboliche (tab. 1), comprendenti 269 contig per clementine e 343 per arancio.

### Validazione degli SNP

L'utilizzo dell'approccio bioinformatico è servito per restringere il numero degli ipotetici SNP a quelli più attendibili. Un ulteriore sistema di validazione è stato comunque necessario per accertare la validità dei marcatori su diverse accessioni di arancio dolce e

Tab. 1 - Vie metaboliche selezionate per l'analisi dei geni che presentano SNP e le specie in cui sono state individuate.

Tab. 1 - Metabolic pathway related to genes including SNPs and relative sweet orange and/or clementine species.

Vie metaboliche	Specie
Glicolisi / Gluconeogenesi	Arancio/Clementine
Metabolismo di amido e saccarosio	Arancio/Clementine
Biosintesi di fenilalanina, tirosina e triptofano	Arancio/Clementine
Biosintesi di ormoni vegetali	Arancio/Clementine
Metabolismo della fenilalanina	Arancio/Clementine
Biosintesi dei fenilpropanoidi	Arancio/Clementine
Biosintesi delle antocianine	Arancio/Clementine
Biosintesi dei flavonoidi	Arancio/Clementine
Metabolismo di ascorbato e aldarato	Arancio/Clementine
Metabolismo del gliossilato e dicarbossilato	Arancio/Clementine
Interconversione pentoso e glucuronato	Arancio/Clementine
Via del pentoso fosfato	Arancio/Clementine
Metabolismo dei zuccheri aminici e nucleotidici	Arancio/Clementine
Metabolismo del glutatione	Arancio/Clementine
Fissazione del carbonio in organismi foto sintetici	Arancio/Clementine
Ciclo dell'acido citrico	Arancio/Clementine
Degradazione del limonene e pinene	Arancio/Clementine
Biosintesi dei monoterpenoidi	Arancio/Clementine
Biosintesi dei carotenoidi	Arancio/Clementine
Biosintesi dei diterpenoidi	Arancio/Clementine
Biosintesi dei terpenoidi	Arancio/Clementine
Biosintesi dei flavoni e flavonoli	Arancio

clementine. In particolare sono stati scelti 50 campioni, comprendenti 33 accessioni di arancio dolce (di cui 15 del gruppo Navel e 18 di quello pigmentato) e 17 di clementine. I campioni sono stati scelti in funzione di alcune caratteristiche qualitative, quali la pigmentazione antocianica, il contenuto in zuccheri e acidità e l'epoca di maturazione. Un totale di 163 SNP (105 per arancio e 58 per clementine), dai 52 contig selezionati, sono stati validati sulle 50 accessioni utilizzando il metodo Kaspar (KBioscience). Il metodo di genotipizzazione ha prodotto 135 SNP utilizzabili per l'analisi dei polimorfismi, di cui 44 sono risultati polimorfici nelle due specie (32,6%) e 36 eterozigoti (26,7%). I polimorfismi evidenziati in tutti gli SNP (ad eccezione di due, caratterizzati da un elevato potere discriminante) sono risultati imputabili esclusivamente a variazioni tra le due specie *C. sinensis* e *C. clementina*, e non in grado di discriminare il gruppo di arancio navel da quello pigmentato.

Gli SNP sono stati ulteriormente validati mediante amplificazione e sequenziamento, accertando i dati di genotipizzazione Kaspar.

## Conclusioni

Il presente lavoro conferma la bassa variabilità genetica presente negli agrumi, in contrasto con la variabilità agronomica, giustificabile dalla selezione clonale indirizzata all'isolamento di caratteristiche specifiche. Il basso potere discriminante, che gli SNP hanno mostrato nel nostro lavoro, è risultato idoneo a caratterizzare con un *fingerprinting* unico esclusivamente un'accessione (Tarocco Gallo). Il polimorfismo evidenziato nei due SNP merita di essere successivamente confermato su un numero maggiore di accessioni, per rilevare il loro potere discriminante nella caratterizzazione del germoplasma o nel miglioramento genetico.

## Riassunto

Sfruttando il database di 540.000 sequenze espresse (ESTs) è stato utilizzato un approccio *in silico* per identificare polimorfismi di sequenza (SNP) nelle specie di arancio dolce [*Citrus sinensis* (L.) Osbeck] e clementine (*C. clementina* Hort. ex Tan.). Con la pipeline QualitySNP sono stati identificati 6.862 potenziali SNP in arancio dolce e 9.017 in clementine. Inoltre 163 SNP di geni appartenenti a 22 vie metaboliche sono stati usati per differenziare 50 accessioni, 33 di arancio (18 arance pigmentate e 15

bionde) e 17 di clementine. Il potere discriminante degli SNP tra specie è alto mentre non sono stati identificati polimorfismi cultivar-specifici.

**Parole chiave:** *Citrus sinensis*, *Citrus clementina*, marcatori molecolari, EST.

## Bibliografia

- BARKLEY N.A., ROOSE M.L., KRUEGER R.R., FEDERICI C.T., 2006. *Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs)*. Theor. Appl. Gen. 112: 1519-1531.
- CHAISSAN T., VAN K., KIM M.Y., KIM K.D., CHOI B-S., LEE S-H., 2010. *In silico single nucleotide polymorphism discovery and application to marker-assisted selection in soybean*. Molecular Breeding, 29 (1): 221-233.
- CLOSE T.J., BHAT P.R., LONARDI S., WU Y., ROSTOKS N., RAMSAY L., DRUKA A., STEIN N., SVENSSON J.T., WANAMAKER S., BOZDAG S., ROOSE M.L., MOSCOU M.J., CHAO S., VARSHNEY R.K., SZÜCS P., SATO K., HAYES P.M., MATTHEWS D.E., KLEINHOF A., MUEHLBAUER G.J., DEYOUNG J., MARSHALL D.F., MADISHETTY K., FENTON R.D., CONDAMINE P., GRANER A., WAUGH R., 2009. *Development and implementation of high-throughput SNP genotyping in barley*. BMC Genomics, 10: 582.
- HAMDY AMAR M., BISWAS M.K., ZHANG Z., GUO W-W., 2011. *Exploitation of SSR, SRAP and CAPS-SNP markers for genetic diversity of Citrus germplasm collection*. Sci Hort. 128 (3): 220-227.
- HAMILTON J.P., HANSEY C.N., WHITTY B.R., STOFFEL K., MASSA A.N., VAN DEYNZE A., DE JONG W.S., DOUCHES D.S., BUELL C.R., 2011. *Single nucleotide polymorphism discovery in elite north american potato germplasm*. BMC Genomics, 12: 302.
- JANNATI M., FOTOUHI R., ABAD A.P., SALEHI Z., 2009. *Genetic diversity analysis of Iranian citrus varieties using micro satellite (SSR) based markers*. J. Hort. Forestry 1 (7): 120-125.
- JIANG D., YE Q-L., WANG F-S., CAO L., 2010. *The mining of Citrus EST-SNP and its application in cultivar discrimination*. Agricultural Sciences in China, 9 (2): 179-190.
- MCNALLY K.L., CHILDS K.L., BOHNERT R., DAVIDSON R.M., ZHAO K., ULAT V.J., ZELLER G., CLARK R.M., HOEN D.R., BUREAU T.E., STOKOWSKI R., BALLINGER D.G., FRAZER K.A., COX D.R., PADHUKASAHASRAM B., BUSTAMANTE C.D., WEIGEL D., MACKILL D.J., BRUSKIEWICH R.M., RÄTSCH G., BUELL C.R., LEUNG H., LEACH J.E., 2009. *Genomewide SNP variation reveals relationships among landraces and modern varieties of rice*. Proc Natl Acad Sci U S A 106(30): 12273-8.
- MICHELETTI D., TROGGIO M., SALAMINI F., VIOLA R., VELASCO R., SALVI S., 2011. *On the evolutionary history of the domesticated apple*. Nature Genetics 43: 1044-1045.
- NOVELLI V.M., TAKITA M.A., MACHADO M.A., 2004. *Identification and analysis of single nucleotide polymorphisms (SNPs) in citrus*. Euphytica 138: 227-237.
- VAN INGHELANDT D., REIF J.C., DHILLON B.S., FLAMENT P., MELCHINGER A.E., 2011. *Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm*. Theor. Appl. Gen. 123 (1): 11-20.
- TREBBI D., MACCAFERRI M., DE HEER P., SØRENSEN A., GIULIANI S., SALVI S., SANGUINETI M.C., MASSI A., ANDRIES E., VAN DER VOSSEN G., TUBEROSA R., 2011. *High-throughput SNP discovery and genotyping in durum wheat (Triticum durum Desf.)*. Theor. Appl. Gen. 123: 555-569.

## Studio della biologia riproduttiva e della maturazione del clementine attraverso l'analisi di mutanti naturali

Gaetano Distefano<sup>1</sup>, Marco Caruso<sup>2</sup>, Sergio Currò<sup>1</sup>, Alessandra Gentile<sup>1\*</sup>, Stefano La Malfa<sup>1</sup>, Giuseppina Las Casas<sup>1</sup> e Paolo Rapisarda<sup>2</sup>

<sup>1</sup> Dipartimento di Scienze delle Produzioni Agrarie e Alimentari, Università di Catania

<sup>2</sup> Consiglio per la Ricerca e Sperimentazione in Agricoltura - ACM Centro di Ricerca per l'Agrumicoltura e le Colture Mediterranee, Acireale (CT)

### Studies of clementine reproductive biology and ripening by means of natural mutants

**Abstract.** Clementine [*Citrus clementina* (Hort. ex Tan.)] represents an important citrus species especially for Mediterranean countries. In this study, two important aspects of the reproductive biology of this species have been investigated, namely the gametophytic self-incompatibility mechanism and the ripening process, with a special emphasis on flavedo pigmentation. Self-incompatibility is an important trait in citrus, since, when coupled with parthenocarpy, it determines the production of seedless fruits in solid blocks. The pigmentation process is also very important especially for out of season products. The availability of natural mutants such as Comune and Monreal clementines differing for self-incompatibility mechanism, and Comune and Tardivo clementines differing for ripening period, represented an important starting point for these analyses. The study of these mutants at biochemical, histological, physiological and molecular level allowed us to shed light on some aspects of the two characters and to identify some of the genes involved in self-incompatibility mechanism and ripening. The comparison between Comune and its late ripening mutant Tardivo revealed a difference and phytoene synthase gene expression and in the accumulation of phytoene, which represents a key precursor of the colored carotenoids. Also, an altered sensitivity to ethylene was observed in Tardivo. Transcriptome analyses of self-pollinated and non-pollinated whole styles with stigmas and stylar canal cells of Comune (self-incompatible) and Monreal (self-compatible) identified several candidate genes involved in pollen-pistil interaction, among which a set of tightly linked genes likely associated with the breakdown of self-incompatibility in Monreal.

**Key words:** *Citrus clementina* (Hort. ex Tan.), self-incompatibility, ethylene, carotenoids, transcriptome.

### Introduzione

I mutanti naturali o indotti rappresentano uno strumento efficace per lo studio di specifici aspetti della fisiologia degli agrumi (Talon *et al.*, 2011). Nell'ultimo decennio la comparazione tra i mutanti e le varietà originarie si è dimostrata particolarmente utile per l'identificazione di geni chiave coinvolti nella regolazione dell'acidità, nella pigmentazione dei frutti e nell'epoca di maturazione. Anche diversi studi sulla biologia riproduttiva sono stati condotti attraverso la caratterizzazione di mutanti, in particolare di pummelo e di mandarino-simili (Chai *et al.*, 2011; Miao *et al.*, 2011).

In questo lavoro sono stati studiati alcuni aspetti della maturazione e della biologia riproduttiva del clementine (*Citrus clementina* Hort. ex Tan.), specie di rilevante interesse per l'agrumicoltura mediterranea e particolarmente ricca di mutanti naturali. In particolare, sono state condotte analisi comparative su frutti di clementine Comune e clementine Tardivo per comprendere i meccanismi fisiologici e molecolari che regolano l'epoca di maturazione. La seconda cultivar infatti, si distingue dalla prima principalmente per il ritardo nella colorazione del flavedo.

Inoltre, il clementine Comune è stato comparato con il clementine Monreal per lo studio delle basi molecolari dell'incompatibilità gametofitica. I due genotipi sono caratterizzati da comportamenti opposti nella manifestazione del carattere "autoincompatibilità". Il clementine Comune è infatti autoincompatibile e produce frutti apireni per partenocarpi, mentre il Monreal, originatosi dal primo per mutazione naturale, è autocompatibile, e quindi ricco di semi.

### Materiali e metodi

#### *Studio del processo di maturazione del clementine*

Sono state monitorate la sintesi e l'accumulo dei carotenoidi, della clorofilla e dell'etilene e sono state ricercate sequenze differenzialmente espresse durante

\* gentilea@unict.it

il processo di invecchiamento in frutti delle due varietà di clementine: Comune e Tardivo.

I campionamenti sono stati realizzati, per un triennio, con cadenza quindicinale, a partire dal 30 settembre di ogni anno, fino alla completa maturazione dei frutti (30 dicembre per il Comune, 15 marzo per il Tardivo) da piante adulte allevate presso l'Azienda Agraria Sperimentale dell'Università di Catania. I rilievi morfologici, i trattamenti con etilene ed il suo inibitore metilciclopropene (MCP1), le analisi biochimiche e molecolari per lo studio della via biosintetica dei carotenoidi e per l'identificazione di geni espressi in maniera differenziale nel flavedo dei due genotipi sono stati eseguiti come riportato da Distefano *et al.* (2009a, 2011).

#### *Studio delle interazioni polline-pistillo e del sistema di autoincompatibilità gametofitica*

I campionamenti dei fiori, le analisi istologiche sugli stili, le osservazioni al microscopio, la microdissezione laser delle cellule dei canali stilari e le analisi del trascrittoma dei due genotipi (clementine Comune e Monreal) sono state effettuate come descritto in Distefano *et al.* (2009b) e Caruso *et al.* (2012).

## Risultati e discussione

#### *Studio del processo di maturazione del clementine*

I carotenoidi sono componenti dell'apparato fotosintetico, contribuiscono alla cattura dell'energia luminosa e alla protezione dei tessuti dalla fotossidazione. Essi caratterizzano la variazione di colore che si manifesta negli agrumi durante la maturazione, periodo in cui all'aumentata sintesi di tali pigmenti si accompagna la degradazione delle clorofille a e b e la graduale trasformazione del cloroplasto in cromoplasto (Zacarias *et al.*, 2011).

Dai rilievi morfologici sul peso e sulle dimensioni dei frutti non si è riscontrata differenza tra i due genotipi per l'intero periodo di sviluppo del frutto, così come le analisi dell'acidità totale e dei solidi solubili che hanno mostrato uniformità con differenze che non si rivelano significative (dati non mostrati). I due genotipi hanno mostrato invece marcate differenze sia a livello biochimico che molecolare per quanto riguarda la sintesi e l'accumulo di carotenoidi nel flavedo dei frutti. In particolare, il genotipo mutato Tardivo mostra un'alterata regolazione della trascrizione del gene fitoene sintasi (PSY), gene chiave della sintesi dei carotenoidi, ed un differente accumulo del corrispondente pigmento (fig. 1). In generale, il clementine Tardivo ha mostrato un ritardo, rispetto al Comune, nell'espressione di geni coinvolti nella

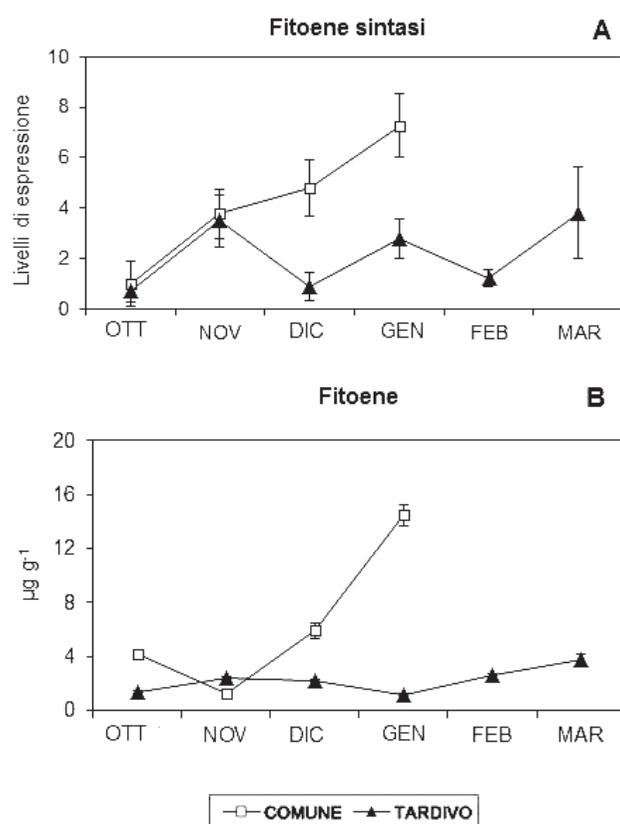


Fig. 1 - Livelli di espressione del gene Fitoene sintasi (A) e contenuto di fitoene (B) nel flavedo del clementine Comune e Tardivo, campionati il primo giorno del mese durante lo sviluppo e la maturazione del frutto (media di tre diverse misurazioni).

Fig. 1 - Analysis of relative transcript levels of *Phytoene sintase* (A) and phytoene content (B) in the flavedo of Comune and Tardivo clementines sampled every first day of the month. Data are means of three different measurements.

sintesi dei carotenoidi e nell'accumulo dei pigmenti corrispondenti come caroteni, luteine e xantofille. Altri geni di cui è stata monitorata l'espressione sono la clorofillasi (CHLASE1), due membri della famiglia dei geni dell'ACC sintasi (ACS1 e ACS2), quello dell'ACC ossidasi (ACO1) e ancora due geni della famiglia dei recettori dell'etilene di tipo ETR1 (ETR1 ed ESR1). Una significativa differenza di espressione del gene CHLASE1 è stata osservata durante l'invecchiamento, raggiungendo i più alti valori nei frutti del Comune. Anche i geni coinvolti nella sintesi e nella percezione dell'etilene hanno mostrato differenti livelli di espressione nei due genotipi. In particolare per ACS1, ACS2 e ACO1, i livelli dei trascritti sono stati uguali nei due genotipi finché i frutti erano ancora verdi, ma hanno dato livelli di espressione differenti in novembre - quando il Comune va incontro all'invecchiamento, mentre i frutti di Tardivo si presentano ancora verdi - e in dicembre, tornando poi su valori omogenei in gennaio quando entrambi i genotipi presentano frutti del tutto colorati. I recettori dell'etilene

Distefano *et al.*

hanno mostrato livelli di espressione diversi in novembre quando i livelli di trascritto in Tardivo sono stati di 2 e 3 volte superiori rispetto al Comune (Distefano *et al.*, 2009a).

Al fine di identificare geni coinvolti nella regolazione del processo di invecchiamento del flavedo, è stata inoltre effettuata un'analisi del trascrittoma attraverso la tecnica cDNA-AFLP. Utilizzando 64 combinazioni di primer *EcoRI/MseI* sono stati comparati i profili trascrizionali del flavedo di Comune e Tardivo campionato in 3 date differenti antecedenti ed in corrispondenza dell'invecchiamento (tra metà novembre e metà dicembre). Tale analisi ha permesso di identificare 85 trascritti potenzialmente implicati nell'evoluzione del colore del flavedo. Le principali classi funzionali dei geni identificati sono state quelle inerenti l'attività fotosintetica, la via di sintesi di pigmenti, il metabolismo lipidico, il trasporto, la risposta allo stress ed il metabolismo ormonale, e sono stati inoltre identificati diversi fattori di trascrizione. La differente espressione di alcuni geni selezionati è stata validata mediante PCR real time (dati non mostrati).

Attraverso analisi biochimiche sono state monitorate la sintesi e l'accumulo di carotenoidi e clorofilla a seguito di trattamento con etilene e con metilciclopropene (un inibitore della percezione dell'etilene) nei due genotipi studiati che hanno mostrato una differente risposta ai trattamenti. La differenza più marcata si è riscontrata sul flavedo ancora verde (*mature green*), in questo caso il trattamento con etilene ha determinato la totale deverdizzazione del frutto nel Comune, mentre tale effetto risulta meno marcato nel Tardivo. In più, in questo stadio di maturazione, il flavedo dei frutti di Tardivo, sia quelli trattati con etilene sia quelli non trattati, ha mostrato un contenuto di clorofilla maggiore rispetto al Comune (Distefano *et al.*, 2009b). L'insieme dei risultati ottenuti dalle diverse prove sperimentali mette in evidenza come la mutazione avvenuta nel Tardivo abbia causato un'alterazione della risposta all'etilene. Per quanto riguarda i trattamenti con MCP1 i frutti dei due genotipi comparati allo stesso stadio di maturazione (*mature green* e *color break*) non hanno mostrato differenze significative (fig. 2).

#### Studio dell'incompatibilità gametofitica del clementine

Analisi istologiche condotte su pistilli delle due varietà autoimpollinate ed incrociate hanno consentito di stabilire che la mutazione da cui ha avuto origine il genotipo autocompatibile Monreal ha interessato il determinante stilare e non quello pollinico. Infatti, l'incrocio Comune x Monreal risulta incompatibile perché il pistillo del Comune riconosce il polline di Monreal come se fosse il proprio e dunque non ne

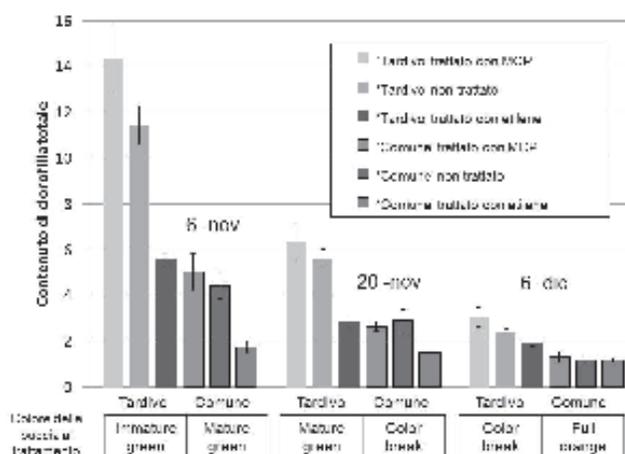


Fig. 2 - Contenuto di clorofilla totale nel corso della maturazione di frutti di clementine Comune e Tardivo trattati con etilene, trattati con metilciclopropene e non trattati.

Fig. 2 - Total chlorophyll (CHL) content during flavedo ripening of ethylene and MCP-treated and untreated fruits of Comune and Tardivo clementine.

consente lo sviluppo (Distefano *et al.*, 2009b).

L'attività di ricerca è stata quindi condotta sui pistilli dei due genotipi. In una prima fase l'analisi del trascrittoma è stata condotta analizzando mRNA isolato da stili e stigmi vergini ed autoimpollinati. L'analisi, condotta con la tecnica cDNA-AFLP, ha permesso di identificare un set di geni potenzialmente coinvolti nell'interazione polline-pistillo (Distefano *et al.*, 2009b). In particolare, oltre ad aver individuato geni coinvolti nel metabolismo ormonale, nel metabolismo del calcio e nella risposta a stress, sono stati isolati diversi frammenti di elementi trasponibili, gran parte dei quali espressi maggiormente nel Monreal. Questo risultato suggerisce che l'attivazione di retrotrasposoni possa avere avuto un ruolo nell'alterazione dell'autoincompatibilità in questo genotipo.

In seguito, il progetto di ricerca ha previsto l'utilizzo di microarray Affymetrix (Citrus GeneChip<sup>®</sup>, contenente 30.171 *probe sets* che rappresentano 33.879 trascritti) per lo studio del trascrittoma delle cellule del canale stilare isolate mediante microdissezione laser (fig. 3). La tecnica della microdissezione laser ha consentito di isolare la porzione dello stilo coinvolta nell'interazione polline-pistillo, e di eliminare porzioni di stilo e stigma non interessate dal processo di incompatibilità, quali quelle costituite da cellule parenchimatiche e da fasci vascolari. Le comparazioni del trascrittoma hanno permesso di identificare un altro set di geni coinvolti nel processo di incompatibilità (Caruso *et al.*, 2012). Un'analisi *time course* è stata effettuata per studiare i livelli di espressione di alcuni geni candidati in stili e stigmi autoimpollinati durante le diverse fasi di interazione tra polline e pistillo (germinazione del polline, crescita del tubo

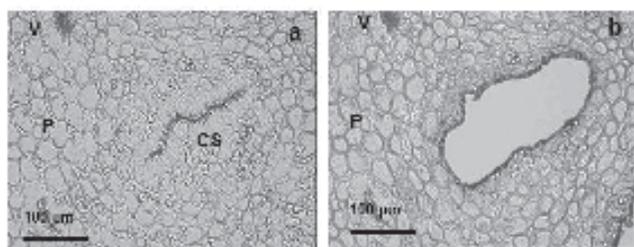


Fig. 3 - Sezioni trasversali di stili di clementine Comune prima (a) e dopo (b) la microdissezione delle cellule dei canali stilari (CS).

V: vasi; P: parenchima.

Fig. 3 - *Transversal sections of Comune clementine styles before (a) and after (b) microdissection of styler canals (CS).*

V: vascular bundles; P: parenchyma.

pollinico, inibizione della fecondazione). I geni candidati presi in esame sono stati: Cit.7568.1.S1\_at, corrispondente ad un proteina putativa F-box con dominio Skp2-like; Cit.11563.1.S1\_at, Cit.5456.1.S1\_at e Cit.5776.1.S1\_s\_at, corrispondenti a tre proteine dalla funzionalità sconosciuta, aventi in comune domini ricchi in acido aspartico e denominate quindi “*aspartic-acid rich (asp-rich) proteins*”. L’analisi di espressione genica, ha permesso di evidenziare una notevole differenza di espressione tra i due genotipi in concomitanza con l’arresto della crescita del tubo pollinico, avvenuta intorno al 5°-6° giorno (fig. 4). Da ricerche effettuate nei database NCBI, Uniprot, Harvest e Phytozome, è emerso che i geni Asp-rich si collocano in un *cluster* di circa 10 kb nello *scaffold* 9 del genoma di clementine (versione 0.9). Inoltre, i geni asp-rich sono strettamente associati ad un altro gene (DELLA) differenzialmente espresso negli stili di Comune e Monreal (Distefano *et al.*, 2009). È emerso quindi che la differente regolazione dei geni presenti in questa regione ha probabilmente portato alla manifestazione del carattere dell’autocompatibilità del clementine Monreal.

## Conclusioni

L’utilizzo di mutanti naturali, assai frequenti in molte specie di agrumi, e segnatamente nel clementine, rappresenta una utile strategia per la comprensione di complessi meccanismi fisiologici quali quelli dell’incompatibilità gametofitica e della maturazione, presi in considerazione in questo studio. La disponibilità di tale materiale biologico, se abbinato a potenti tecniche di analisi quali la microdissezione laser (LCM) che consente di isolare cellule da tessuti specifici, o le tecniche molecolari per l’analisi del trascrittoma, consentono di ottenere risultati di estremo interesse che potranno inoltre essere approfonditi grazie ai dati provenienti dai progetti di sequenziamento del genoma delle specie e varietà di agrumi.

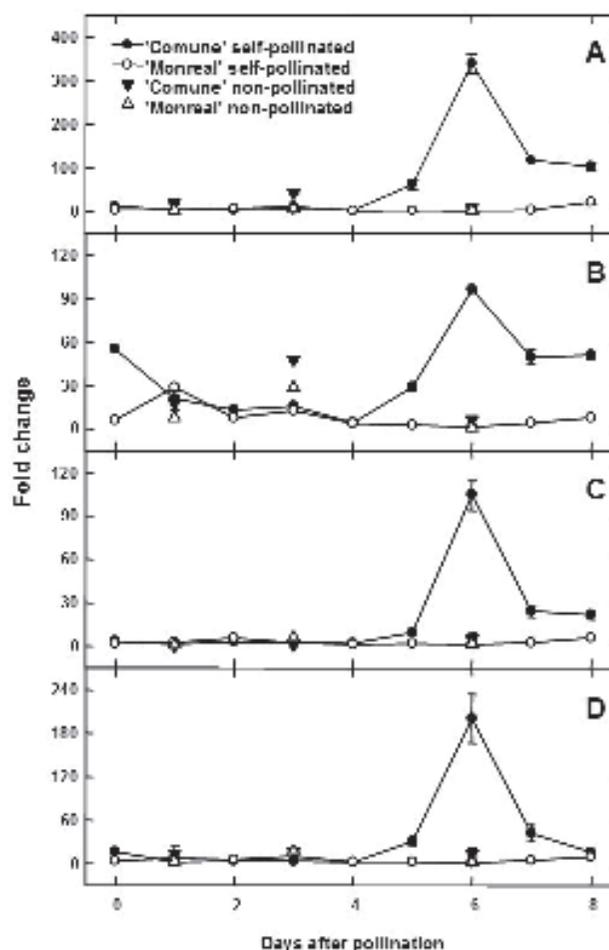


Fig. 4 - Analisi di espressione di 4 geni coinvolti nell’interazione polline-pistillo, eseguita su stili e stigmi di clementine Comune e Monreal. A: cit.7568 (F-box); B: cit.11563 (Asp-rich protein gene 1); C: cit.5456 (Asp-rich protein gene 2); D: cit.5776 (Asp-rich protein gene 3). Stili e stigmi sono stati campionati da 0 (fiori vergini) a 8 giorni dopo l’autoimpollinazione. Campioni di stili e stigmi vergini sono stati campionati il primo, il terzo ed il sesto giorno dopo l’antesi per verificare che i geni fossero indotti dall’autoimpollinazione.

Fig. 4 - *Expression analysis of four genes involved in pollen-pistil interaction, performed on styles and stigmas of Comune and Monreal clementine. A: cit.7568 (F-box); B: cit.11563 (Asp-rich protein gene 1); C: cit.5456 (Asp-rich protein gene 2); D: cit.5776 (Asp-rich protein gene 3). Styles and stigmas were sampled from 0 (non pollinated flowers) to 8 days after self-pollination (DAP). Samples of virgin styles and stigmas were sampled one, three and six DAP in order to determine if the genes were induced by self-pollination.*

## Riassunto

Attraverso una serie di analisi fisiologiche, istologiche e molecolari sono stati indagati due aspetti particolarmente importanti del ciclo di fruttificazione del clementine [*Citrus clementina* (Hort. ex Tan.)], specie di rilevante interesse per l’agrumicoltura mediterranea. In particolare sono stati presi in considerazione il sistema di auto incompatibilità gametofitica ed i processi di maturazione e colorazione del flavedo. Entrambi gli aspetti indagati rivestono importanza per

Distefano *et al.*

i riflessi sugli aspetti qualitativi del frutto (assenza di semi ed epoca di maturazione). La disponibilità di mutanti naturali, clementine Comune e Monreal nel caso dello studio dell'incompatibilità gametofitica, clementine Comune e Tardivo nel caso degli studi di maturazione, ha rappresentato un indubbio vantaggio per l'indagine svolta. Le analisi effettuate, nel loro complesso, hanno contribuito a una migliore conoscenza delle basi fisiologiche e genetiche dei due importanti caratteri presi in considerazione e di individuare alcuni geni coinvolti nei meccanismi di incompatibilità gametofitica e di pigmentazione del flavedo.

**Parole chiave:** *Citrus clementina* (Hort. ex Tan.), autoincompatibilità, etilene, carotenoidi, trascrittoma.

### Bibliografia

- CARUSO M., MERELO P., DISTEFANO G., LA MALFA S., LO PIERO A.R., TADEO F. R., M. TALON, GENTILE A., 2012. *Comparative transcriptome analysis of stylar canal cells identifies novel candidate genes implicated in the self-incompatibility response of Citrus clementina*. BMC Plant Biol., 12:20.
- CHAI LJ, GE XX, XU Q, DENG XX., 2011. *CgSL2, an S-like RNase gene in 'Zigui shatian' pummelo (Citrus grandis Osbeck), is involved in ovary senescence*. Mol Biol Rep, 38:1-8.
- DISTEFANO G., CARUSO M., LA MALFA S., GENTILE A., TRIBULATO E., 2009b. *Histological and molecular analysis of pollen-pistil interaction in clementine*. Plant Cell Rep., 28:1439-1451.
- DISTEFANO G., CARUSO M., LAS CASAS G., LA MALFA S., GENTILE A., TRIBULATO E., SIMSEK O., RAPISARDA P., 2011. *Maturation in Fruits of Clementine: Biochemical and Molecular Analysis*. Acta Hort., 892: 329-334.
- DISTEFANO G., LAS CASAS G., CARUSO M., TODARO A., RAPISARDA P., LA MALFA S., GENTILE A., TRIBULATO E., 2009a. *Physiological and molecular analysis of the maturation process in fruits of clementine mandarin and one of its late-ripening mutants*. J. Agr. Food Chem., 57:7974-7982.
- MIAO HX., QIN YH., TEIXEIRA DA SILVA JA., YE ZX., HU GB., 2011. *Cloning and expression analysis of S-RNase homologous gene in Citrus reticulata Blanco cv. Wuzishatangju*. Plant Sci, 180:358-367.
- TALON M., CERCOS M., IGLESIAS D.J., COLMENERO-FLORES J.M., IBÁÑEZ V., BRUMOS J., HERRERO-ORTEGA M.A., RIOS G., TEROL J., TADEO F.R., 2011. *Citrus Genomics and Breeding: Identification of Candidate Genes by the Use of Mutants and Microarrays*. Acta Hort., 892:19-25.
- ZACARIAS L., ALQUEZAR B., RODRIGO M.J., 2011. *New Insights into the Regulation of Carotenoid Diversity in Citrus Fruits*, Acta Hort., 892:321-327.

## La regolazione della biosintesi delle antocianine durante la frigo-conservazione post raccolta delle arance rosse

Angela Roberta Lo Piero\*

Dipartimento di Scienze delle Produzioni Agrarie e Alimentari, Università di Catania

### The anthocyanin biosynthesis as affected by the post harvest cold storage of blood oranges

**Abstract.** In our work, the transcriptome analysis based on subtractive hybridization was performed in order to emphasize the overall induction in gene expression after the exposure of blood oranges [(*Citrus sinensis*) L. Osbeck Tarocco Sciara] to low temperature. The cold induction of several gene expression was then validated by RT real time PCR. The results show that cold stress induces transcriptomic modifications directed towards the increase of flavonoid biosynthesis, including those reactions involved in anthocyanin biosynthesis, as well as of the metabolic pathways supplying it. The overall profile of genes induced by cold storage, mainly involved in defense mechanisms, in osmoregulation processes and lipid desaturation, seems to be similar to that of the chilling acclimated species. Moreover, the data highlights that the EST encoding the transcription factor NAC domain protein is selectively induced by cold in blood oranges but not in common oranges thus proposing it as candidate gene specifically involved in the blood orange response to cold exposure.

**Key words:** *Citrus sinensis*, suppressive subtractive hybridization, pigmented orange, cold stress, anthocyanin.

#### Introduzione

L'arancio dolce è una coltura economicamente importante coltivata nelle zone tropicali e sub tropicali del mondo. In particolare, l'Italia produce circa il 4% del raccolto mondiale ed il 37% degli Stati europei. Nella campagna di commercializzazione 2009-2010, la produzione italiana di agrumi ha raggiunto i 3,8 milioni di tonnellate, gran parte costituita dalle varietà pigmentate, coltivate prevalentemente in

Sicilia e utilizzate quasi esclusivamente per il consumo fresco (Baldi, 2011). Le arance rosse [(*Citrus sinensis*) L. Osbeck] sono caratterizzate dalla presenza, nella scorza, ma soprattutto nella porzione commestibile del frutto, di antociani (Cotroneo *et al.*, 2006) che rappresentano i composti flavonoidi più abbondanti di frutta e verdura. Tra i flavonoidi, gli antociani sono stati intensamente studiati con particolare riguardo alle loro proprietà biofunzionali che sono ritenute utili nella prevenzione di alcune malattie degenerative (De Pascual-Teresa e Sanchez-Ballesta, 2008). Il contenuto di pigmenti antociani può variare da frutto a frutto a causa di differenti condizioni ambientali come l'esposizione alla luce, lo stato nutrizionale, trattamenti ormonali o xenobiotici e la temperatura (de Pascual-Sanchez-Teresa e Ballesta, 2008). In particolare, Lo Piero *et al.* (2005) hanno dimostrato che il contenuto di pigmento di frutti appena raccolti può aumentare in maniera significativa durante la conservazione post raccolta in celle frigorifere grazie alla stimolazione della trascrizione dei geni coinvolti nella biosintesi delle antocianine. Per quanto riguarda la durata del periodo di conservazione a freddo, è stato trovato un punto di convergenza, per la temperatura e la durata di stoccaggio, in modo da ottenere il massimo accumulo di antociani e un accettabile gradimento dei frutti da parte dei consumatori (4 °C per 77 giorni) (Lo Piero *et al.*, 2005). In questo lavoro è stata effettuata un'analisi globale dei cambiamenti indotti nel trascrittoma della polpa di arancio rosso [(*Citrus sinensis*) L. Osbeck Tarocco Sciara] dopo l'esposizione a 4 °C per 77 giorni, utilizzando un approccio sperimentale basato sulla costruzione e l'analisi di una libreria sottrattiva di cDNA. I dati di espressione differenziale di geni selezionati sono stati validati tramite real time RT-PCR. Tra i geni analizzati, è stato misurato il livello di espressione di un fattore di trascrizione appartenente alle proteine NAC domain, sia nelle arance rosse che nelle arance bionde sottoposte a frigoconservazione, al fine di valutare il suo possibile coinvolgimento nella risposta all'esposizione al freddo.

\* rlopiero@unict.it

Lo Piero

## Materiali e metodi

Frutti di arancio pigmentato [(*Citrus sinensis*) L. Osbeck Tarocco Sciara] sono stati raccolti nel mese di gennaio 2009 presso l'azienda sperimentale del Centro di Ricerca per l'Agrumicoltura e le Colture Mediterranee nel territorio di Palazzelli (Italia). Le arance appena raccolte sono state trattate come descritto in Lo Piero *et al.* (2005). L'RNA totale è stato estratto e quantificato come descritto in Lo Piero *et al.* (2005). I cDNA purificati, sono stati generati utilizzando il Super SMART kit con 1 µg di RNA secondo le istruzioni del produttore (Clontech, Palo Alto, CA). L'ibridazione sottrattiva (SSH) è stata effettuata tra il campione di polpa di arancia sottoposto a conservazione a freddo ("tester", 4 °C per 77 giorni) e il campione di polpa di arancia conservato a 25 °C per 77 giorni ("driver"), utilizzando il kit di PCR select (Clontech, Palo Alto, CA). I cDNA ottenuti dalla sottrazione sono stati clonati nel vettore pGEM-T Easy (Promega, Madison, WI). Gli inserti dei plasmidi ricombinanti sono stati sequenziati e ulteriormente analizzati per identificare similarità con sequenze presenti nella banca dati pubblica Genbank (programma BLASTX). La real time RT-PCR è stata eseguita come descritto in Lo Piero *et al.* (2005) e la quantificazione relativa dell'espressione genica è stata calcolata utilizzando il metodo della comparazione del ciclo soglia ( $C_T$ ). Gli esperimenti di PCR quantitativa sono stati eseguiti su tre gruppi di campioni in modo da generare un valore medio del ciclo soglia  $C_T$  e calcolare la deviazione standard. Per l'esecuzione dell'esperimento di breve frigo conservazione (15 giorni) le arance (Tarocco Sciara e Navel) sono state conservate come descritto in Lo Piero *et al.* (2005), ed i prelievi sono stati eseguiti al tempo 0 e ogni tre giorni per un periodo sperimentale complessivo di 15 giorni.

## Risultati e discussione

La porzione più abbondante di cloni specificatamente espressi nei campioni conservati a freddo è rappresentata da geni codificanti per proteine che proteggono la cellula dai danni ossidativi, come le deidrine, le metallotionine e le tioredoxine. Si è osservato che tra i cloni differenziali sono presenti geni codificanti per enzimi coinvolti nel metabolismo degli acidi grassi e dei lipidi, come le desaturasi che stabilizzano le membrane contro i danni da freddo, geni codificanti per proteine coinvolte nella regolazione della traduzione del segnale e fattori di trascrizione che regolano geni indotti da stress. Per quanto riguarda il metabolismo primario e secondario, sono stati isolati geni codificanti per enzimi che intervengono nelle vie fermentative (esochinasi, fruttosio-bifosfato aldolasi, la piruvato chinasi e la gliceraldeide-3-fosfato deidrogenasi); geni appartenenti alla via metabolica che conduce alla fenilalanina ed enzimi coinvolti nella biosintesi delle antocianine (dato non mostrato). In particolare, per alcuni geni codificanti gli enzimi coinvolti nel pathway delle antocianine, è stato studiato il differente livello di espressione a 4 °C e 25 °C tramite PCR quantitativa in real time. L'aumento dei livelli della 3-deossi-D-arabino-eptulosonato-7-fosfato deidrogenasi (DAHP) e della corismato mutasi (CM1) a 4 °C suggeriscono che la via dello shikimato è fortemente attivata dalle basse temperature (tab. 1). I dati relativi all'espressione differenziale della fenilalanina ammonio liasi (PAL), della 4-cumaroiCoA sintetasi (4CL), della calcone sintasi (CHS), dell'antocianidina sintetasi (ANS) hanno confermato l'aumento dei livelli di trascrizione di questi geni (tab. 1). Tra le proteine regolatrici, fattori di trascrizione appartenenti alla famiglia delle proteine NAC domain sono stati isolati nella libreria sottrattiva (clone 378,

Tab. 1 - Analisi dell'espressione di differenti EST coinvolte nella biosintesi dei composti flavonoidi tramite RT-real time PCR. Ciascun valore rappresenta la media di tre repliche  $\pm$  SE (or SD) (nm: non misurabile).

Tab. 1 - Analysis of the expression of different blood orange ESTs involved in the flavonoid biosynthesis by RT-real time PCR. Each value represents the mean value of three replications  $\pm$  SE (or SD) (nm: not measurable).

Expressed sequence tag	Quantità normalizzata campioni conservati a 25°C $2^{-\Delta\Delta CT}$	Quantità normalizzata campioni del tempo 0 $2^{-\Delta\Delta CT}$
3-deossi-D-arabino-eptulosio 7-fosfato sintetasi (DAHP)	21 $\pm$ 0,7	18 $\pm$ 1,1
Corismato mutasi (CM1)	>1.000 (nm*)	>1.000 (nm**)
Fenilalanina ammonio liasi (PAL)	85,5 $\pm$ 2,1	79,8 $\pm$ 2,6
4-Coumarato-coenzima A ligasi (4CL)	52 $\pm$ 1,1	48,3 $\pm$ 1,1
Calcione sintetasi (CHS)	45 $\pm$ 3	43 $\pm$ 2,5
Antocianidina sintetasi (ANS)	85 $\pm$ 1,5	82 $\pm$ 2,0

FK826627). I fattori di tipo NAC hanno un ruolo fondamentale nella risposta delle piante di stress (Olsen *et al.*, 2005). I livelli di espressione della EST codificante il fattore NAC domain clone 378 (FK826627) sono stati determinati sia nelle arance rosse che nelle arance bionde sottoposte ad una breve frigo-conservazione (4 °C x 15 giorni) al fine di valutare il suo possibile coinvolgimento nella risposta all'esposizione al freddo (fig. 1). L'analisi rivela che, nelle arance pigmentate, l'espressione del gene NAC domain (clone 378) aumenta durante la conservazione a freddo raggiungendo livelli di 20 volte più elevati rispetto al tempo 0 (fig. 1A). Al contrario, l'esperimento effettuato su arancia bionda evidenzia che la EST codificante il clone 378 non è indotta né nei campioni controllo (25 °C), né in campioni conservati a freddo (4 °C) (fig. 1B). Pertanto, la EST FK826627 si propone come gene specificamente coinvolto nella risposta all'esposizione al freddo nelle arance rosse.

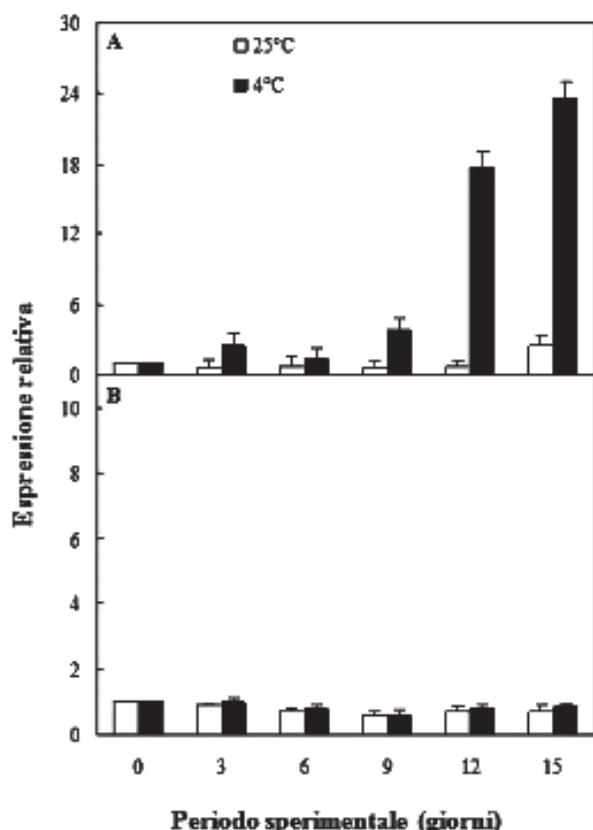


Fig. 1 - Analisi in RT- real time PCR dell'espressione del gene codificante la proteina NAC domain (FK826827) in A) arancia rossa (Tarocco Sciara) and B) arancia comune (Navel) entrambe sottoposte a conservazione a freddo. Ciascun punto rappresenta il valore medio di tre repliche  $\pm$  SE (or SD). Ogni replica è composta da tre campioni di mRNA indipendentemente isolati.

Fig. 1- RT- real time analysis of NAC domain protein (FK826827) in A) blood (Tarocco Sciara) and B) common (Navel) oranges subjected to cold storage. Each point represents the mean value of three replications  $\pm$  SE (or SD). Each replication was composed of three independently isolated mRNA.

## Riassunto

Lo scopo di quest'attività è stato l' identificazione di geni specificatamente espressi durante la frigo-conservazione nella polpa di arancia rossa, tramite la tecnica della ibridizzazione sottrattiva. Il profilo dei geni indotti dalla frigo conservazione (coinvolti nei meccanismi di difesa, nell'osmoregolazione, nella desaturazione dei lipidi, e nel metabolismo primario e secondario) suggerisce che le arance rosse possiedono una tipologia di risposta allo stress riconducibile alle specie acclimatate al freddo. Inoltre, l'analisi comparativa tra arance pigmentate e non pigmentate ha consentito l'isolamento di un fattore di trascrizione di tipo NAC specificatamente coinvolto nella risposta allo stress nelle arance rosse.

**Parole chiave:** *Citrus sinensis*, ibridizzazione sottrattiva, arance pigmentate, stress da basse temperature, antocianine.

## Bibliografia

- BALDI S.; USDA Foreign Agricultural Service, Global Agricultural Information Network. 2011, <http://gain.fas.usda.gov/>.
- COTRONEO P.S., RUSSO M.P., CIUNI M., REFORGIATO RECUPERO G., LO PIERO A. R., 2006. *Quantitative real-time RT-PCR profiling of anthocyanin biosynthetic genes during orange fruit ripening*. J. Amer. Soc. Hort. Sci., 131: 537-543.
- DE PASCUAL-TERESA S., SANCHEZ-BALLESTA M. T., 2008. *Anthocyanins: from plant to health*. Phytochem. Rev., 7: 281-299.
- LO PIERO A.R., PUGLISI I., RAPISARDA P., PETRONE G., 2005. *Anthocyanin accumulation and related gene expression in blood orange fruit induced by low temperature storage*. J. Agric. Food Chem., 53: 9083-9088.
- OLSEN A.N., ERNST H.A., LO LEGGIO L., SKRIVER K., 2005. *NAC transcription factors: structurally distinct, functionally diverse*. Trends Plant Sci., 10: 79-87.

## Identificazione di cloni bac di *Citrus sinensis* contenenti geni rilevanti per la qualità dei frutti

Sergio Pietro Ferrante<sup>1\*</sup> e Mikeal L. Roose<sup>2</sup>

<sup>1</sup> CNR - Istituto di Genetica Vegetale, Palermo

<sup>2</sup> Department of Botany and Plant Sciences, University of California Riverside

### Identification of *Citrus sinensis* bac clones containing genes relevant to fruit quality

**Abstract.** We used the C38 EST assembly (<http://harvest.ucr.edu/>) and OligoSpawn software (<http://138.23.178.42>) to design the overgo probes employed to screen a *Citrus sinensis* Vaniglia BAC library (19X). For BAC library screening we selected 81 overgo probes associated with unigenes that putatively code for enzymes relevant to fruit quality (flavonol, anthocyanin, carotenoid, chlorophyll, cellulose, starch, ascorbic acid, aromatic amino acid and lignin biosynthesis; sucrose catabolism; glycolysis; oxidative/nonoxidative pentose phosphate pathway; fatty acid biosynthesis and oxidation; Krebs cycle). A two-dimensional 9 X 9 overgo pooling strategy was used. Hybridization probes were pooled and hybridized in groups of intersecting rows and columns to high-density BAC filters, followed by a deconvolution process that established BAC-probe addresses. BAC addresses were obtained for 75 of the 81 overgo probes initially selected, for a total of 1018 BAC clones, a number consistent with the depth of coverage of the BAC library. BAC end sequencing was carried out and sequence similarity searches (BLAST) were conducted against draft assemblies of the *Citrus clementina* genome (<http://phytozome.net>). The BAC clones corresponding to each probe were located within the same scaffold as the target gene, demonstrating that the approach we used was successful in isolating the target genes.

**Key words:** Unigene, repeat element, overgo probe, pooling strategy, deconvolution.

### Introduzione

Sono molti gli aspetti che contribuiscono a determinare la qualità dei frutti. Oltre alle caratteristiche

fisiche (dimensioni, forma, colore, struttura, numero di semi, sbucciabilità) molto importanti sono le caratteristiche chimiche, come il contenuto in zuccheri e l'acidità, le proprietà organolettiche, le componenti nutritive e biofunzionali. La percezione di questi aspetti influenza le decisioni dei consumatori, che si confrontano con una gamma sempre più diversificata di opzioni di acquisto, e il loro miglioramento si traduce direttamente in un aumento della domanda, che in ultima analisi, determina il successo dell'industria degli agrumi. Il *breeding* tradizionale, come strategia di miglioramento genetico per lo sviluppo di cultivar di agrumi, presenta numerosi limiti: sterilità sessuale, incompatibilità, lunga fase giovanile, bassa variabilità genetica all'interno dei gruppi economicamente più importanti di agrumi (arance dolci, pompelmi, limoni, satsuma, clementine). Questi ostacoli non soltanto riducono la probabilità di ottenere ed identificare individui superiori nelle progenie, ma rendono estremamente difficile, se non impossibile, ottenere cultivar migliorate con caratteristiche del frutto che i consumatori riconoscono come agrumi tipici. Per mantenere le caratteristiche essenziali e ottenere miglioramenti significativi nei tratti critici, identificare, clonare e manipolare i geni che influenzano i tratti desiderabili e indesiderabili è di fondamentale importanza. Le genoteche BAC rappresentano uno dei principali strumenti utilizzati per la mappatura fisica e il sequenziamento del genoma negli eucarioti. Il loro screening con sonde gene-specifiche e tecniche multiplex ha consentito di accelerare l'isolamento dei geni (Zhang e Wu, 2001). Le sonde possono infatti essere ordinate in matrici di righe e colonne che si intersecano (*pooling strategy*), individualmente marcate e ibridizzate in gruppi, ciascun gruppo corrispondente ad una riga o colonna (Evans e Lewis, 1989). Tuttavia, un grosso limite di questo approccio sta sia nella difficoltà di conseguire una marcatura uniforme di tutte le sonde in ciascun gruppo, sia nella presenza di elementi ripetuti o di elementi conservati nelle tipologie di sonde comunemente utilizzate (frammenti di DNA o

\* sergiopietro.ferrante@igv.cnr.it

cDNA clonati, PCR amplimer, oligonucleotidi), che spesso confondono i risultati dell'ibridazione e non consentono di individuare la corrispondenza tra BACs e sonde (Gardiner *et al.*, 2004). Rispetto alle tipologie di sonde gene-specifiche comunemente utilizzate, il metodo di ibridazione overgo (Ross *et al.*, 1999) presenta diversi vantaggi (migliore cinetica di ibridazione, maggiore specificità e minore background di ibridazione, minore tasso di falsi positivi, elevato throughput, facilità di gestione). Il fondamento di questi vantaggi è rappresentato dal ridotto numero di nucleotidi di cui è costituita una sonda overgo, che aumenta la probabilità di individuare regioni *single copy* in sequenze EST e unigenes. Le sonde overgo, lunghe 36/40-bp, sono infatti disegnate sullo stampo di filamenti di cDNA che sono stati precedentemente filtrati allo scopo di eliminare tutti gli elementi ripetuti conosciuti. Questo tipo di approccio presuppone ovviamente la disponibilità di una banca dati completa che permetta di individuare le sequenze ripetute presenti nel gene target. HarvEST: Citrus (<http://harvest.ucr.edu/>), un database di sequenze espresse (EST) sviluppato presso l'Università della California a Riverside, presenta 141 *libraries* e 469.618 ESTs di *Citrus* e *Poncirus*. Utilizzando la EST *assembly* C38, che comprende 36.980 citrus unigenes, e il software "OligoSpawn" (<http://138.23.178.42>) abbiamo disegnato 31.230 sonde overgo. Nell'articolo descriviamo i metodi e i risultati sperimentali ottenuti per le sonde overgo selezionate.

## Materiale e metodi

### *BAC library*

È stata utilizzata una BAC library di *Citrus sinensis* cultivar Vaniglia (Ferrante *et al.*, 2007). La library contiene 58.368 cloni (155 piastre da 384-well) con una copertura genomica pari a 19 equivalenti del genoma aploide e una dimensione media degli inserti di 125 kb. La library è contenuta su 4 filtri ad alta densità (3 filtri 22cm x 22cm contenenti ciascuno 18.432 cloni unici presenti in duplice copia, e un filtro 11,9 cm x 7,8 cm contenente 3.072 cloni unici presenti in duplice copia).

### *Selezione delle sonde overgo*

È stata utilizzata la EST *assembly* C38, che comprende 36.980 citrus unigenes (<http://harvest.ucr.edu/>), e il software "OligoSpawn" (<http://138.23.178.42>). Sono state disegnate 31.230 sonde overgo (sonde 36 bp specifiche per un determinato gene). Al software OligoSpawn è stato chiesto di selezionare sonde con un contenuto in guanina e citosina compreso tra il

45% e il 55% e una temperatura di melting compresa tra 55 °C e 70 °C. Le sonde selezionate dal software OligoSpawn sono state controllate per verificare che non ricadessero in corrispondenza delle zone di giunzione tra introni ed esoni, utilizzando il software "HarvEST BLAST Search" e il database "JGI Citrus" (<http://138.23.178.42/blast/blast.html>), e il software "Intron Finder" ([http://solgenomics.net/tools/intron\\_detection/find\\_introns.pl](http://solgenomics.net/tools/intron_detection/find_introns.pl)). Delle 31.230 sonde "overgo" individuate, ne sono state selezionate 81, putativamente associate con unigene che codificano per gli enzimi strutturali delle vie biosintetiche aventi un ruolo rilevante nel determinare la qualità dei frutti (tab. 1). Ciascuna sonda overgo è rappresentata da una coppia di primers 22-mer le cui estremità 3' presentano 8 basi complementari. A seguito dell'annealing ciascuna coppia di nucleotidi crea, in posizione 5', una estremità sporgente (*overhang*) di 14 basi. In presenza del frammento di Klenow (*E. coli* DNA polymerase I) vengono sintetizzate le basi mancanti nel filamento incompleto (*filling in*), incorporando i nucleotidi marcati. I primers sono stati acquistati presso Integrated DNA Technologies ([www.idtdna.com/Home/Home.aspx](http://www.idtdna.com/Home/Home.aspx)), con una scala di sintesi pari a 25 nmol. La sequenza dei primers è disponibile su richiesta agli autori.

### *Marcatore delle sonde overgo, ibridizzazione, autoradiografia*

È stata effettuata come descritto in Madishetty *et al.*, 2007, con modifiche minori. Per la sonda del background è stata utilizzata una concentrazione pari a 1/10 di quella descritta nel suddetto articolo.

### *Multiplex pooling strategy*

Le ibridazioni sono state condotte con gruppi (pool) di sonde, disposte in una matrice di righe e colonne che si intersecano, ciascuno contenente 9 sonde. Questa metodologia consente teoricamente di completare lo screening delle 81 sonde ricorrendo soltanto a 18 ibridazioni (efficienza della matrice = 4,5).

### *Analisi delle autoradiografie*

Ciascuna autoradiografia è stata analizzata (score) 2 volte e ulteriormente controllata in caso di discrepanza tra le 2 letture (fig. 1).

### *Deconvoluzione*

Per attribuire i cloni BAC isolati a ciascuna delle sonde utilizzate (*matrix deconvolution*) è stato utilizzato un software derivato da HybSweeper (Lazo *et al.* 2005) e un Perl deconvolution script sviluppato presso il Clemson University Genomics Institute (USA).

## Ferrante e Roose

Tab. 1 - Mappatura degli unigene e dei corrispondenti BACs all'interno del genoma di *Citrus clementina*.

A.A.A.: amino acidi aromatici; ANT.: antocianine; ASC.: acido ascorbico; CAR.: carotenoidi; CEL.: cellulosa; CHL.: clorofilla; F.A.B.: biosintesi acidi grassi; F.A.O.: ossidazione acidi grassi; FLA.: flavonoli; GLY.: glicolisi; LIG.: lignina; OXI.: via dei pentoso fosfati (ossidativa/non ossidativa); STA.: amido; TRI.: ciclo degli acidi tricarbossilici (ciclo di Krebs). UNI. N.: numero unigene all'interno della EST assembly C38. BAC N.: numero BAC individuati per ciascuna sonda. SCAF. N.: numero dello scaffold. UNI. POS.: mappatura unigene all'interno dello scaffold. BACs POS.: mappatura BAC ends all'interno dello scaffold.

*Tab. 1 - Genomic mapping of unigenes and BACs against Citrus clementina genome.*

*A.A.A.: aromatic amino acid; ANT.: anthocyanin; ASC.: ascorbic acid; CAR.: carotenoid; CEL.: cellulose; CHL.: chlorophyll; F.A.B.: fatty acid biosynthesis; F.A.O.: fatty acid oxidation; FLA.: flavonol; GLY.: glycolysis; LIG.: lignin; OXI.: oxidative/nonoxidative pentose phosphate pathway; STA.: starch; TRI.: tricarboxylic acid cycle (Krebs cycle). UNI. N.: unigene number (EST assembly C38). BAC. N.: number of BAC identified per probe. SCAF. N.: scaffold number. UNI. POS.: unigene mapping against the scaffold. BACs POS.: BAC ends mapping against the scaffold.*

BIOSINTESI	UNI. N.	BAC N.	SCAF. N.	UNI. POS.	BACs POS.
A.A.A.	222	9	8	745668-747550	643038-816856
A.A.A.	331	4	31	1782246-1783769	1652897-1838141
A.A.A.	1644	8	1	11721561-11725384	11625304-11853813
A.A.A.	1754	7	123	17810-20287	5234-109179
A.A.A.	5580	15	9	4435011-4438849	4327210-4581129
A.A.A.	5805	1	3	656907-658322	652001-773320
A.A.A.	6214	12	16	3273917-3275997	3219960-3323807
A.A.A.	6686	3	80	625013-626346	531114-673576
A.A.A.	7359	1	80	623225-625077	531114-696367
A.A.A.	14407	21	7	4284009-4289634	4269725-4377269
A.A.A.	15815	11	24	2130469-2133204	2114293-2212414
A.A.A.	16190	14	49	1229453-1229860	1177290-1303853
A.A.A.	18320	15	53	307350-311266	273433-399652
A.A.A.	18662	14	41	965836-970231	871522-973510
A.A.A.	18961	24	25	2891770-2895237	2850028-2969276
A.A.A.	25469	17	16	225454-228270	191555-311593
A.A.A.	28012	2	5	2816700-2817389	2686395-2921500
A.A.A.	31595	1	5	2819719-2820939	2788714
ANT.	1432	36	43	1294036-1294586	1145596-1512892
ANT.	9790	14	2	6450148-6450758	6316637-6454056
ANT.	17156	25	31	477079-497154	392827-521033
ANT.	19185	27	84	700089-701852	680415-818205
ANT.	19782	4	22	2126783-2128516	2095019-2218552
ASC.	1273	22	3	1270880-1273888	1211039-1313733
ASC.	1506	8	1	4984531-4987010	4910106-5054600
CAR.	4	13	30	203554-215057	116373-216876
CAR.	9	10	47	183132-187902	123787-226240
CEL.	9485	14	39	641785-643924	603885-729960
CHL.	2313	2	69	970654-972829	966116-1142876
F.A.B.	9358	2	2	7075336-7081201	7057900-7180247
F.A.B.	11189	13	31	1326320-1330100	1282540-1439306
F.A.O.	11290	8	89	721038-724977	629518-936718
F.A.O.	18201	24	7	5800467-5802492	5693335-5810083
FLA.	2234	1	95	82349-83957	49151-112527
FLA.	4888	21	25	2400955-2404458	2325085-2425087
FLA.	19290	7	16	980731-983321	925336-1029283
FLA.	29909	19	4	2327276-2327629	2174727-2341821
GLY.	1280	23	67	589086-593675	493032-643204
GLY.	3648	18	8	1613881-1617624	1504882-1665467

segue

segue

BIOSINTESI	UNI. N.	BAC N.	SCAF. N.	UNI. POS.	BACs POS.
GLY.	6588	5	2	6065529-6067930	6059891-6149156
GLY.	7333	20	6	6386509-6390784	6339847-6471082
GLY.	9792	12	1	12149674-12153427	12087972-12202660
GLY.	10672	27	58	493930-500265	470025-622697
GLY.	12486	36	6	6291449-6294607	6286582-6294820
GLY.	12567	21	22	245044-250510	243144-352104
GLY.	19335	19	29	892574-893860	786721-899856
LIG.	1886	11	26	824075-826146	741874-865228
LIG.	2003	19	30	479226-481519	414441-595955
LIG.	3589	1	9	3459033-3464431	3393458-3484363
LIG.	5323	21	26	550513-552366	521485-595978
LIG.	5581	8	47	607323-610469	505972-633771
LIG.	11308	17	12	4251373-4253077	4152459-4262830
LIG.	16768	12	25	2158801-2159557	2127402-2226991
OXI.	524	15	65	337029-341044	300320-346934
OXI.	2641	19	56	1126358-1127586	1122842-1146862
OXI.	6342	61	11	510569-512169	486337-648749
OXI.	7213	5	3	648672-652301	488740-652741
OXI.	9822	10	5	1788483-1795270	1740830-1868739
OXI.	10059	7	18	15880-18015	3660-93357
STA.	5779	1	56	555654-562521	557371-655870
STA.	12980	17	1	11281456-11286979	11190194-11308207
SUC.	673	16	7	6041050-6043439	5957441-6050580
SUC.	4759	4	2	7041567-7045778	6928334-7048580
SUC.	18988	15	3	4457889-4463255	4412829-4531494
TRI.	1389	10	7	6115259-6118250	6099322-6210022
TRI.	2149	22	78	104836-109619	90141-147277
TRI.	4187	6	11	3608058-3610762	3513921-3651616
TRI.	4970	7	3	4399279-4405305	4363576-4459597
TRI.	5004	12	25	3073657-3074802	3064345-3208718
TRI.	6277	5	25	3288693-3293350	3342543-3343297
TRI.	6363	2	5	2431964-2434204	2425381-2600449
TRI.	10489	15	3	7349353-7353312	7300128-7474365
TRI.	16528	16	25	1536987-1540457	1475298-1575382
TRI.	18474	8	29	2497506-2500092	2434707-2546869
TRI.	18670	26	4	7201405-7206793	7178320-7322019

### Re-arrey dei cloni BAC, BAC e sequencing

Il re-arrey dei cloni BAC individuati con le sonde overgo in piastre da 96-well e il sequenziamento delle terminazioni dei cloni BAC (metodo Sanger) è stato effettuato presso l'Amplicon Express (<http://www.genomex.com>).

### Mappatura degli unigene e delle BAC ends

Per la mappatura degli unigene e delle sequenze delle terminazioni dei cloni BAC individuati lungo il genoma di *Citrus clementina*, è stato utilizzato Phytozome (<http://www.phytozome.net/>).

### Risultati e discussione

Mediante il software OligoSpawn sono state selezionate 31.230 sonde overgo. OligoSpawn ha individuato 22.444 *unique oligos* (oligonucleotidi lunghi 36 bp ciascuno presente in un solo unigene ma non presente in nessun altro unigene). Per i restanti 8.786 unigene OligoSpawn non è stato capace di individuare *unique oligos* e ha selezionato sonde overgo comuni a più unigene. Questo risultato è imputabile al fatto che l'assembly C38 occasionalmente colloca alleli alternativi dello stesso gene in unigene differenti, e, probabil-

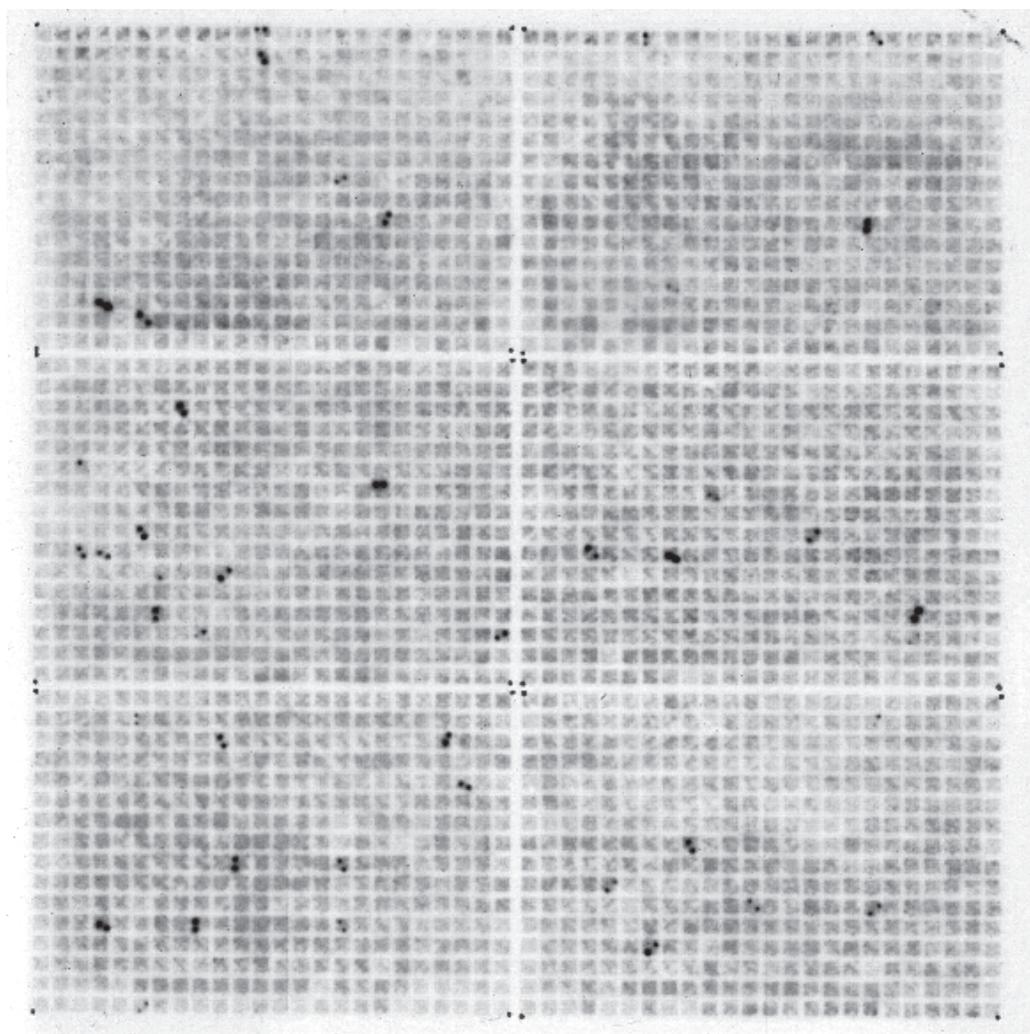


Fig. 1 - Autoradiografia di un filtro di nitrocellulosa 22 x 22cm.  
 Fig. 1 - Autoradiograph of a nitrocellulose 22 x 22cm filter.

mente, al fatto che nel genoma di *Citrus* sono presenti geni paraloghi molto simili. Degli 81 geni selezionati, 75 sono stati individuati (92,6%), corrispondenti ad un totale di 1018 cloni BAC (tab. 1). Delle 75 corrispondenti sonde overgo, 69 (92%) hanno ibridizzato con 2 o più cloni, mentre solo tre sonde (4%) hanno ibridizzato con più di 30 cloni. In media ciascuna sonda ha ibridizzato con 13,6 cloni BAC, un numero compatibile con la copertura genomica della library (19X). Sei sonde overgo non hanno ibridizzato in modo specifico con alcun clone. Questo risultato è attribuibile a diversi fattori: disegno della sonda non ottimale, fallimento della reazione di marcatura, assenza del corrispondente segmento di DNA nella library. Il software OligoSpawn si è rivelato particolarmente efficace nel filtrare le sequenze ripetute presenti in citrus e soltanto una sonda (OXI. 6342) ha ibridizzato con un numero elevato di cloni (tab. 1). Su 960 dei 1.018 cloni BAC individuati è stato effettuato il sequenziamento delle terminazioni degli inserti. I risultati sono stati positivi

(*pass rate*) nel 93% dei casi e sono state ottenute 1.786 sequenze. La lunghezza delle sequenze ottenute è risultata in media pari a 550 basi. La mappatura delle sequenze ottenute lungo il genoma di *Citrus clementina* ha evidenziato che i cloni BAC corrispondenti a ciascuna sonda sono posizionati all'interno dello stesso scaffold contenente il gene target (tab. 1), dimostrando che i cloni BACs individuati sono *true positives* e che la metodologia seguita è stata vincente nell'isolare i geni oggetto dello studio.

### Conclusioni

A integrazione del lavoro di sequenziamento del genoma di *Citrus sinensis*, per identificare regioni codificanti e non codificanti, elementi di regolazione inclusi, responsabili delle caratteristiche biologiche di questa specie. Nel presente lavoro, utilizzando una BAC library di *C. sinensis* cv "Vaniglia" e una tecnica di screening multiplex con sonde overgo, abbiamo

isolato 75 geni appartenenti alle vie biosintetiche aventi un ruolo rilevante nel determinare la qualità dei frutti. I geni isolati sono stati mappati all'interno del genoma di *C. clementina*. Questi risultati dimostrano l'idoneità della BAC library e della tecnica utilizzata per l'isolamento dei geni, la mappatura fisica e gli studi di genomica comparativa in *Citrus*. Nonostante le nuove tecniche di sequenziamento oggi disponibili sono sempre più veloci ed economiche, le genoteche BAC rimarranno probabilmente un mezzo fondamentale per colmare le lacune e risolvere le anomalie dovute alle nuove tecnologie.

### Riassunto

Utilizzando la EST assembly C38 (<http://harvest.ucr.edu/>) e il software "OligoSpawn" (<http://138.23.178.42>), sono state selezionate 81 sonde overgo associate con unigene che codificano per gli enzimi strutturali delle vie biosintetiche aventi un ruolo rilevante nel determinare la qualità dei frutti. Con le sonde selezionate è stato effettuato lo screening di una BAC library (19X) di *Citrus sinensis* cv Vaniglia per individuare i cloni contenenti i corrispondenti geni. Degli 81 geni selezionati, 75 sono stati individuati, corrispondenti ad un totale di 1018 cloni BAC. La mappatura dei BACs individuati contro il genoma di *Citrus clementina* (<http://phytozome.net>) ha dimostrato che i cloni corrispondenti a ciascuna sonda sono posizionati all'interno dello stesso

scaffold contenente il gene target, confermando che la metodologia seguita è stata vincente nell'isolare i geni oggetto del presente studio.

**Parole chiave:** Unigene, elemento ripetuto, sonda overgo, *pooling strategy*, deconvolution.

### Bibliografia

- EVANS GA., LEWIS KA., 1989. *Physical mapping of complex genomes by cosmid multiplex analysis*. Proc. Nat. Acad. Sci. USA., 87: 5030-5034.
- FERRANTE S., GERACI G., SCARANO MT., 2007. *Construction of a bacterial artificial chromosome library of a seedless clone of Citrus sinensis cv "Vaniglia"*. Italus Hortus 14 (6): 22-23.
- GARDINER J., SCHROEDER S., POLACCO M. L., SANCHEZ-VILLEDA H., FANG Z., MORGANTE M., LANDEWE T., FENGLER K., USECHE F., HANAFEY M., TINGEY S., CHOU H., WING R., SODERLUND C., COE E. H. JR., 2004. *Anchoring 9,371 Maize Expressed Sequence Tagged Unigenes to the Bacterial Artificial Chromosome Contig Map by Two-Dimensional Overgo Hybridization*. Plant Physiology, 134: 1317-1326.
- LAZO G. R., LUI N., GU Y. Q., KONG X., COLEMAN-DERR D., ANDERSON O. D., 2005. *Hybsweeper: a resource for detecting high-density plate gridding coordinates*. BioTechniques 39: 320-324.
- MADISHETTY K., CONDAMINE P., SVENSSON J. T., RODRIGUEZ E., CLOSE T. J., 2007. *An improved method to identify BAC clones using pooled overgos*. Nucleic Acids Res. 35(1): e5.
- ROSS M. T., LABRIE S., MCPHERSON J., STANTON VP. JR., 1999. *Screening large insert libraries by hybridization*. In A. Boyl ed., "Current Protocols in Human Genetics". Wiley, New York, 5.6.1-5.6.52.
- ZHANG H. B., WU C., 2001. *BACs as tools for genome sequencing*. Plant Physiol Biochem, 39: 195-209.