

DOCUMENTO  
TECNICODa sinistra:  
G. Faes,  
M. Salmaso

\* **Marzia Salmaso**  
\*\* **Giorgia Faes**

\* Dipartimento di Agronomia  
Ambientale e Produzioni Vegetali,  
Università di Padova, Agripolis  
Legnaro (PD)

\*\* Molecular Stamping - Povo (TN)

## ANALISI DELLA DIVERSITÀ GENOMICA DELLA VITE MEDIANTE POLIMORFISMI A SINGOLO NUCLEOTIDE (SNP)

Obiettivo di questo lavoro è stata la valutazione del livello di variabilità presente nel genoma di vite mediante polimorfismi a singolo nucleotide. La presenza di SNP è stata determinata analizzando 25 EST in nove diversi genotipi. Questo ha permesso di ottenere una stima della diversità nucleotidica intra e inter-specifica della vite.

### Introduzione

Nella vite lo sviluppo di nuove varietà è ostacolato dal lungo ciclo generazionale di questa pianta e dagli effetti negativi causati dalla depressione da *inbreeding* [1]. Negli ultimi anni, lo sviluppo di tecniche molecolari ha permesso di superare queste difficoltà e ha aperto una nuova era per l'attività di miglioramento genetico. Grazie alla reazione a catena della polimerasi (PCR) e ai progressi nelle procedure di

sequenziamento [26] è possibile oggi determinare l'eredità di polimorfismi a singolo nucleotide (SNPs) e di inserzioni/delezioni (indels), che rappresentano le differenze genetiche più frequenti tra gli individui di una popolazione (circa il 90% dei polimorfismi nel genoma umano sono dovuti a SNPs [9]). Altre metodiche sono state impiegate con successo per la genotipizzazione e per gli studi di associazione. Tra queste citiamo la tecnica SSCP che valuta il polimor-

fismo conformazionale di un singolo filamento [24], l'analisi "heteroduplex" (HA) che si basa sulla denaturazione e lenta rinaturazione di un mix di frammenti omologhi [38] e gli *electronic* SNPs (eSNPs) dedotti dal confronto di sequenze appartenenti a librerie di differenti individui [29].

Gli SNPs sono marcatori genetici co-dominanti e risultano molto adatti all'integrazione come marcatori funzionali in mappe genetiche ad elevata risoluzione

[6][19]. Inoltre, gli SNP presentano un elevato livello di potenziale automazione che aumenta l'efficienza nella mappatura [17]. In alcuni casi gli SNP risultano correlati con cambiamenti amminoacidici non conservativi che portano a variazioni fenotipiche, fornendo così una prima evidenza del possibile coinvolgimento di un gene nell'espressione di un certo carattere [21].

Il loro impiego nella rilevazione di associazioni tra alleli e fenotipi è già stato dimostrato nell'uomo, studiando ad esempio gli alleli APOE  $\epsilon$ 2,  $\epsilon$ 3,  $\epsilon$ 4 nel morbo di Alzheimer [14] e nel mais nel controllo del tempo della fioritura [36].

Nelle piante superiori, lo sviluppo di SNPs non è stato ancora sufficientemente implementato, sebbene la mappatura del genoma di *Arabidopsis* abbia evidenziato l'efficacia della tecnica. Analisi SNPs in orzo [16], mais [34], soia [39] e barbabietola da zucchero [31] hanno dimostrato la frequente presenza di polimorfismi e l'esistenza di un numero limitato di aplotipi entro i geni, spesso molto differenziati fra di loro.

La vite (*Vitis vinifera* L.) è una delle specie più importanti dell'area mediterranea, è ampiamente presente nei climi temperati e la sua coltivazione si sta diffondendo rapidamente in diverse parti del mondo. Nonostante ciò le conoscenze molecolari sono ancora ridotte, come per la maggior parte delle piante arboree perenni. Recentemente sono state sviluppate alcune mappe molecolari di vite [10] [11] [15] [13] [27], che favoriscono l'inizio di una nuova era per il *breeding* assistito da marcatori.

In questo lavoro al fine di valutare l'impiego dei polimorfismi SNPs per lo sviluppo di marcatori molecolari è stato utilizzato un approccio "EST-candidate gene". Sono state selezionate alcune centinaia di "Expressed sequence tag" (TAG) appartenenti a vie metaboli-

che di interesse in termini di qualità e resistenza alle malattie (metabolismo degli zuccheri, meccanismi di trasduzione del segnale, vie metaboliche degli antociani e meccanismi di difesa). Molti dei geni utilizzati sono stati recentemente mappati nella vite [10]. La presenza di polimorfismi è stata determinata sequenziando 25 di questi frammenti genici in nove diversi genotipi (sei cultivars di *Vitis vinifera*, *Vitis riparia* e un ibrido interspecifico). Sono state determinate la frequenza degli SNPs e l'esistenza di aplotipi allelici. In questo modo è stato possibile ottenere una stima della diversità nucleotidica intra e interspecifica.

## Materiali e metodi

**Materiale vegetale.** Per questo studio sono stati scelti sette cloni di *Vitis vinifera* L. (cvs Regent, Lemberger, Moscato bianco, Teroldego rotaliano, Riesling italico, Pinot noir, Syrah), un ibrido interspecifico Freiburg 993-60 (un genotipo complesso derivato da incroci multipli che comprende nel proprio pedigree anche specie americane quali *V. rupestris* e *V. lincecumii*) e un clone di *Vitis riparia* Mchx. Questi cloni sono stati utilizzati per valutare il grado di polimorfismo molecolare in *Vitis*.

**Estrazione del DNA genomico, sviluppo dei primer e amplificazione tramite PCR.** Il DNA è stato estratto utilizzando il protocollo Doyle & Doyle (1990) [12], con alcune modifiche. La rottura della parete cellulare delle foglie (1gr) è stata fatta in azoto liquido con mortaio e pestello. Immediatamente dopo sono stati aggiunti 5ml di tampone CTAB (acetiltrimetilammonio bromuro) preriscaldato a 60°C, 0,2%v/v di  $\beta$ -mercaptoetanolo e 1%w/v di PVP (polivinilpolipirrolidone) per isolare ed eliminare poi i polifenoli. I campioni sono

stati incubati a 60°C per 40 min e successivamente lasciati raffreddare a temperatura ambiente. Per favorire la separazione della fase organica sono stati aggiunti 4ml di IAC. I campioni sono stati centrifugati a 6000g per 15 min per isolare il surnatante. A questo è stato aggiunto un volume di isopropanolo pari a due terzi del volume di surnatante. L'aggiunta di isopropanolo ad una temperatura di -20°C favorisce la precipitazione del DNA. Dopo centrifugato a 5000g per 15 min, il precipitato è stato lavato con una soluzione di etanolo 76%v/v e ammonio acetato 7.5M per almeno 2 ore. Al termine del lavaggio i campioni sono stati centrifugati a 5000g per 10 min ed il precipitato così raccolto è stato lasciato asciugare sotto cappa. Infine il DNA è stato risospeso in 200  $\mu$ l di tampone TE e sottoposto a trattamento con Rnasi 10 $\mu$ g/ml.

Per l'analisi dei polimorfismi è stato usato un approccio ESTs-SNPs. Sono state utilizzate 1920 sequenze appartenenti a due librerie cDNA (di germoglio e infiorescenza di Regent), disponibili presso l'Istituto Agrario di S. Michele all'Adige. Le sequenze sono state classificate sulla base della loro omologia mediante PSI-BLAST dell'NCBI [2] e sono state selezionate quelle appartenenti a quattro classi di interesse: metabolismo degli zuccheri (S), meccanismi di trasduzione del segnale (C), metabolismo degli antociani (A) e meccanismi di difesa (D) (Tab. 1). Mediante il programma *Primer3 Output* (<http://www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi>) [28] sono stati sviluppati i primer per amplificare queste sequenze in modo da ottenere frammenti di lunghezza pari a 300-500 bp (<http://genomics.research.iasma.it>).

Invece, per amplificare i geni CHI (Chalcone isomerase), DFR (Dihydroflavonol reductase) e UFGT (UDP-glucose:flavonoid 3-O-glucosyltransferase), sono stati

**Tab. 1 - Primer utilizzati per le analisi. Per ciascun gene viene riportata l'omologia in database e la putativa categoria funzionale**

Gene	Categoria funzionale putativa e omologia PSI-BLAST
CHI	A Chalcone isomerase
I-a06	C Myb related protein 5
I-b02	D Cf2/Cf5 disease resistance protein
I-d04	D putative bZIP DNA-binding protein
I-e04	C Protein kinase family
I-f01	D Caffeoil-CoA o-methyltransferase
I-f08	C Remorin family protein
I-g04	A Cinnamoyl-CoA reductase
I-h09	C Gibberellin response modulator
II-a02	C myo-inositol-1-phosphate synthase
II-a05	S Probable glycerate dehydrogenase
II-b05	A Dicyanin
II-c08	C Zinc finger protein
II-e02	C DNA-binding protein 3
II-g05	C MYB transcription factor R2R3 type
III-b08	C Putative receptor like protein kinase
III-b09	A Putative dihydroflavonol reductase
III-c03	C Putative ethylene response factor ERF3b
III-c08	C Response regulator 9
III-c12	D Pathogenesis-related protein 10
IV-g10	S N-acetylglucosaminyltransferase - related
IV-h04	C Squamosa promoter binding protein
IV-h06	S Aldose-1-epimerase-like protein
IV-h09	D Phosphoesterase family
UFGT	A Flavonol 3-O-glucosyltransferase

sviluppati i primer sulle sequenze disponibili nella banca dati dell'NCBI (GenBank accession number X75965, AF280768 e AF000372).

I primer sono stati utilizzati per amplificare il DNA genomico dei 9 diversi genotipi.

Le condizioni di amplificazione sono state ottimizzate per ottenere un unico amplificato, variando in particolare la temperatura di *annealing* e la concentrazione di  $Mg^{2+}$ .

Tutti i geni selezionati presentano un rapporto di

segregazione atteso sulla base del genotipo dei parentali. Per la reazione di PCR sono stati usati 0.2 mM dNTPs, 1.5 mM  $MgCl_2$ , 0.2 mM di ciascun primer e 0.2 U Taq DNA polymerase. Il protocollo di amplificazione prevede 35 cicli di 45 s a 94°C, 30 s a 55-60°C e 90 s a 72°C, preceduti da 4 min di denaturazione a 94°C e seguiti da 10 min di estensione a 72°C. Dopo la reazione di amplificazione i frammenti di PCR sono stati sequenziati in entrambe le direzioni per assicurare un'accurata determinazione della sequenza.

## Rivelazione dei polimorfismi

**Rivelazione dei polimorfismi a singolo nucleotide.** I metodi per l'identificazione di mutazioni attraverso il sequenziamento del DNA si basano sull'allineamento e la comparazione degli elettroferogrammi prodotti. L'analisi delle sequenze *forward* e *reverse* è stata fatta impiegando il *software* Pregap4/Gap4 [4].

Questo *software* assegna un punteggio di qualità agli elettroferogrammi, elimina le porzioni di sequenza di bassa qualità, crea allineamenti multipli di sequenze, effettua una scansione per identificare le mutazioni, consente un'efficace visualizzazione dei tracciati, produce dei *report* e permette l'archiviazione degli allineamenti ottenuti.

Tramite Pregap4, gli elettroferogrammi vengono allineati e comparati con la traccia di riferimento e viene evidenziata ogni possibile mutazione.

I dati così prodotti sono resi disponibili a Gap4 sotto forma di database. Utilizzando Gap4 è possibile archiviare gli allineamenti, visualizzare gli elettroferogrammi, controllare visivamente le differenze tra la traccia di riferimento e il campione, ed infine tabulare tutti i polimorfismi riscontrati.

## Diversità nucleotidica

**Diversità nucleotidica, determinazione dell'aplotipo e PIC.** Il livello di polimorfismo del DNA è stato valutato attraverso uno specifico parametro: la diversità nucleotidica  $\pi$  [23]. La diversità nucleotidica per un singolo gene rappresenta la proporzione di nucleotidi che differiscono tra due sequenze, mediata su tutte le possibili coppie di comparazioni genotipiche. Per ciascuna coppia di genotipi  $\pi$  è pari a  $K/L$ , dove  $K$  è il numero di differenze per sito nucleotidico ed  $L$  la lunghezza della sequenza in bp. Quando sono presenti più di due sequenze, tale valore è dato dalla media aritmetica di tutte le possibili combinazioni [23]. Il valore medio della diversità nucleotidica tra diversi frammenti genici è stato calcolato dividendo il numero totale di siti varianti per il numero totale di nucleotidi sequenziati [18].

Un aplotipo è una specifica combinazione di nucleotidi, uno per ciascun sito polimorfico presente in un singolo cromosoma [32]. Gli aplotipi allelici sono stati definiti allineando le sequenze dei frammenti amplificati. Per la derivazione computazionale degli aplotipi è stato impiegato l'algoritmo di Clark [7].

È stata analizzata la combinazione degli alleli presenti in ciascun sito polimorfico, utilizzando il metodo della massima parsimonia per ricavare i diversi aplotipi in un insieme di campioni. L'algoritmo assegna direttamente un aplotipo agli individui omozigoti in tutti i siti e a quelli eterozigoti per un unico sito polimorfico. Questi aplotipi vengono utilizzati per identificare i genotipi dei rimanenti individui (eterozigoti multipli). Questo processo termina quando tutti gli aplotipi vengono ricostruiti, oppure quando non è più possibile trovare ulteriori aplotipi. L'algoritmo fallisce nel riconoscimento di un aplotipo

quando, comparando i diversi genotipi, risulta presente un individuo eterozigote per due aplotipi unici.

L'obiettivo dell'analisi svolta è stato la valutazione del numero corretto di aplotipi, piuttosto che la loro determinazione, e questo algoritmo si è rivelato molto utile allo scopo. Infatti è stato possibile dedurre il numero totale di aplotipi presenti nel nostro insieme di campioni, sebbene per alcuni genotipi non siano stati definiti i due alleli. Questi aplotipi sono stati identificati impiegando un piccolo sottoinsieme delle popolazioni F1. Seguendo questo approccio gli aplotipi allelici sono stati verificati sequenziando i prodotti di amplificazione di diversi individui scelti casualmente nelle popolazioni segreganti.

Il valore del PIC (*Poly-morphism Information Content*) indica la capacità di un marcatore di rilevare loci polimorfici in esperimenti di mappatura e nel *breeding* assistito [5].

L'eterozigosità è stata calcolata come  $H = 1 - \sum_i p_i^2$  dove  $\sum p_i^2$  è la somma di ciascuna frequenza allelica al quadrato [22]. Il PIC è stato calcolato per gli aplotipi come  $1 - \sum p_i^2 - \sum_{ij} p_i^2 p_j^2$  in cui  $\sum p_i^2$  è la somma di ogni *i*-esima frequenza di allele al quadrato.

## Risultati della ricerca

**Natura e frequenza degli SNPs identificati tramite sequenziamento.** I risultati ottenuti dal sequenziamento sono stati utilizzati anche per valutare la frequenza e il tipo di SNPs identificati, la diversità nucleotidica, gli aplotipi e il PIC. I prodotti di amplificazione sono stati allineati con i rispettivi cDNA per identificare esoni, introni e polimorfismi nella sequenza. In totale sono stati esaminati 11629 nucleotidi per ciascuno dei 9 genotipi, di cui 7653 derivanti da esoni. Sono stati identificati 247 SNPs, di cui 177 in regioni codificanti

(cSNPs). Questi risultati sono riportati in Tab. 2. Circa il 50% delle variazioni esoniche sono state trovate nella terza posizione del codone, ma è stato anche identificato un numero significativo di variazioni non sinonime (100 su 177 siti polimorfici osservati nelle regioni codificanti). Gli SNPs sono stati classificati sulla base del loro orientamento 5'-3' in transizioni (sostituzioni tra purine o tra pirimidine) e transversioni (sostituzioni di una purina con una pirimidina e viceversa). Il 61% degli SNPs sono rappresentati da transizioni (SNP tipo 1 e 2 in Tab. 2), sebbene essi costituiscano solo un terzo di tutti i possibili tipi di mutazioni. Nei 9 genotipi studiati, 3 frammenti genici presentano solamente uno o due alleli: un aldose-1-epimerase (IVH06), un GHMyb9 (proteina Myb-related) (IIG05) e un acetilglucosaminyltransferase (IVG10). È stato riscontrato un solo caso di delezione.

Il valore medio di SNPs rilevati è stato ricavato mediando il numero di SNP di ogni possibile coppia di genotipi in esame, come mostrato in Tab. 3. Tale valore calcolato per la totalità dei genotipi è pari a 113.79 (77.70 negli esoni). La diversità nucleotidica media tra i frammenti genici è stata calcolata dividendo il numero totale medio di siti varianti (K) per il numero totale di nucleotidi sequenziati (L). Il valore di  $\pi$  medio, considerando tutti i genotipi è risultato pari a  $9.79 \times 10^{-3}$ ; si riduce invece a  $8.41 \times 10^{-3}$  per le cultivar di *V. vinifera*. Come atteso, tale valore è maggiore nel confronto tra *V. vinifera* e *V. riparia* ( $12.85 \times 10^{-3}$ ) così come tra l'ibrido e i cloni *V. vinifera* ( $10.28 \times 10^{-3}$ ).

La diversità nucleotidica varia molto anche tra i diversi frammenti analizzati, oscillando tra 0 e  $30.26 \times 10^{-3}$ . È stato osservato un maggior numero di polimorfismi nelle regioni codificanti rispetto a quelle non codificanti ( $\pi = 10.15 \times 10^{-3}$  nelle regioni codificanti, rispetto a un totale di  $9.79 \times 10^{-3}$ ). Nelle regio-

ni codificanti la diversità nucleotidica è pari a  $8.63 \times 10^{-3}$  tra le cultivar, a  $13.66 \times 10^{-3}$  tra *V. vinifera* e *V. riparia*, e a  $10.90 \times 10^{-3}$  tra *V. vinifera* e l'ibrido.

## Determinazione degli aplotipi

Per capire come la variabilità nucleotidica nei genotipi considerati si organizza in aplotipi entro i frammenti analizzati si è cercato di inferire per via computazionale gli aplotipi presenti in ciascuna varietà nei 25 frammenti genici.

Non è facile determinare i diversi aplotipi. Infatti, partendo da una singola sequenza, l'identificazione del corretto aplotipo non è immediata, poiché in un genotipo eterozigote la fase di due SNPs adiacenti non è immediatamente risolvibile. Ad esempio, se un locus presenta un polimorfismo C/T e il locus adiacente un T/G, non è possibile sapere se i due aplotipi sono rappresentati da C (locus1) T (locus2) (aplotipo 1) e T (locus1) G (locus2) (aplotipo 2) o dalle altre due possibili combinazioni. Tra i vari metodi sviluppati per la derivazione del corretto aplotipo, per quest'analisi è stato utilizzato l'algoritmo di Clark [7]. In questo modo è stato possibile dedurre il numero totale di aplotipi presenti, sebbene per alcuni non sia stato possibile determinare l'esatto allele. Nei casi più semplici, l'aplotipo corretto può essere determinato dalla comparazione del genotipo eterozigote con quello omozigote. Gli altri aplotipi possono così venire risolti per sottrazione di questi aplotipi "certi". Questa procedura continua finché si trovano tutti gli aplotipi o finché non si riesce più a dedurre nessun nuovo aplotipo. Nella maggior parte dei casi è stato identificato un aplotipo principale (allele A in Tab. 4) e un discreto numero di aplotipi diversi da questo. Nei 25 frammenti genici considerati, il numero degli aplotipi identificati per

**Tab. 2 - SNPs riscontrati nelle sequenze di 25 frammenti genici studiati in 9 genotipi**

Gene	Frammento (bp)	Regione codificante (bp)	SNP tot	cSNP	Tipo di SNP						Posizione nel codone			Sostituzioni non sinonime
					1	2	3	4	5	6	1	2	3	
CHI-for	598	409	5	2	1	2	1	0	1	0	0	2	0	2
CHI-rev	550	387	8	2	4	3	0	0	1	0	0	0	2	0
IA06	357	357	7	7	0	3	1	1	0	2	2	3	2	6
IB02	481	481	19	19	3	8	0	3	4	1	6	4	9	11
ID04	419	0	12	0	2	4	2	0	3	1	0	0	0	0
IE04	425	425	15	15	6	1	1	2	3	2	4	6	5	12
IF01-for	165	0	3	0	1	0	0	1	0	1	0	0	0	0
IF01-rev	607	303	11	4	4	3	2	0	1	1	1	0	3	1
IF08-for	554	335	4	3	1	1	1	0	0	1	1	0	2	0
IF08-rev	565	170	6	2	1	3	1	1	0	0	0	0	2	0
IG04	357	138	9	6	0	3	2	2	1	1	0	2	4	3
IH09	565	565	5	5	3	2	0		0	0	1	0	4	1
IIA02	405	108	8	5	3	1	1	2	1	0	2	1	2	3
IIA05	190	190	9	9	3	0	1	0	3	2	2	1	6	4
IIB05	533	203	20	12	6	6	1	2	3	2	5	4	3	11
IIC08	418	418	12	12	6	2	0	1	1	2	2	0	10	2
IIE02	329	329	11	11	6	2	2	1	0	0	4	6	1	10
IIG05	258	258	1	1	0	1	0	0	0	0	0	0	1	0
IIIB08	292	292	4	4	0	2	1	0	1	0	2	2	0	4
IIIB09	934	174	10	1	1	5	1	2	0	1	0	1	0	1
IIIC03	228	228	25	25	11	8	1	2	0	3	6	9	10	15
IIIC08	232	161	3	2	0	1	1	1	0	0	1	0	1	1
IIIC12	412	241	12	7	3	5	2	1	1	0	5	0	2	5
IVG10	398	398	1	1	0	0	1	0	0	0	0	0	1	0
IVH04	276	202	2	1	1	0	0	0	1	0	0	0	1	0
IVH06	254	107	0	0	0	0	0	0	0	0	0	0	0	0
IVH09	369	316	9	5	2	6	0	0	1	0	0	0	5	0
UFGT	458	458	16	16	5	5	1	3	2	0	4	5	7	8
<b>Totali (1)</b>	<b>11629</b>	<b>7653</b>	<b>247</b>	<b>177</b>	<b>73</b>	<b>77</b>	<b>24</b>	<b>25</b>	<b>28</b>	<b>20</b>	<b>48</b>	<b>46</b>	<b>83</b>	<b>100</b>
<b>Totali (2)</b>	<b>104661</b>	<b>68877</b>												

(1) Numero di paia di basi sequenziate per genotipo. (2) Numero di paia di basi sequenziate nei 9 genotipi SNPs riscontrati nelle sequenze di 25 frammenti genici studiati in 9 genotipi. Sono tabulate transizioni/transversioni, posizioni dei cSNPs nella corrispondente codone e il conseguente cambiamento aminoacidico. cSNP = SNP nella regione codificante. Tipo di SNP: transizioni 1 = C↔T, 2 = G↔A; transversioni 3 = C↔G, 4 = T↔A, 5 = C↔A, 6 = T↔G. I frammenti di DNA sono stati sequenziati in entrambe le direzioni, quando non è stata raggiunta la sovrapposizione le due sequenze sono state scritte in corsivo, separatamente, indicando con -for e -rev il 5' e il 3' del gene.

gene varia da 1 a 10. Se gli SNPs fossero associati casualmente tra loro all'interno di un gene e la popolazione fosse infinitamente grande, ci potrebbero essere circa  $2^n$  possibili aplotipi per frammento genico dove  $n$  è il numero di SNP in un frammento.

In questo studio, a prescindere dal numero di SNP osservati, potrebbero essere presenti al massimo 18 diversi aplotipi (uno diverso per ognuno dei 18 cromosomi presenti nei 9 cloni di *Vitis*). Come conseguenza della mancanza di *linkage*

*equilibrium* fra SNP che deriva dalla ridotta ricombinazione e dall'azione di selezione e deriva genetica in popolazioni di dimensione finita, il numero di aplotipi risulta minore del massimo numero teoricamente possibile [8]. Il numero di aplotipi trovati è stato in media di 6.04 diversi aplotipi per gene, spesso con un aplotipo in frequenza maggiore accompagnato da una serie di altri aplotipi (Tab. 4).

Inoltre, è stata calcolata l'eterozigosità attesa per ciascun frammento genico considerando gli aplotipi. L'etero-

zigosità attesa si può definire come la probabilità che due cromosomi scelti a caso nella popolazione portino alleli diversi, in questo caso aplotipi diversi. Per quanto riguarda i 25 frammenti genici considerati, l'eterozigosità aplotipica varia da 0 a 0.87 nei 9 genotipi con un valore medio pari a 0.64. Nell'uomo, l'eterozigosità media è di 0.534 [32] e nella soia di 0.52 [39]. Viene così confermata l'elevata eterozigosità della vite. La massima eterozigosità attesa possibile per un singolo SNP bialeleico è pari a 0.5. Dei 25 geni, 22 presentano

una diversità genica superiore a 0.5, pertanto l'elevata eterozigosità e la natura multiallelica degli aplotipi li rendono più informativi dei bialeleici SNPs.

Per indicare il valore dei marcatori, per esperimenti di mappatura o di selezione assistita, viene calcolato il PIC (*Polymorphism Information Content*). I loci più desiderabili sono quelli con molti alleli e un PIC prossimo ad 1. Il valore del PIC degli aplotipi del nostro campione di 25 frammenti genici varia da 0 a 0.86 e ha un valore medio di 0.61.

**Tab. 3 - SNPs e cSNPs presenti in 7 genotipi di *V. vinifera*, *V. riparia* e nel genotipo complesso Freiburg 993-60 (tutti i genotipi)**

Comparazioni	Frammenti amplificati (bp)	Regioni codificanti (bp)	SNPs totali	$\pi$ totale ( $10^{-3}$ )	cSNPs totali	$\pi$ regione codificante ( $10^{-3}$ )
Tutti i genotipi	da 165 a 934	da 0 a 565	113.79	9.79	77.70	10.15
Cult vs Cult	"	"	97.76	8.41	66.06	8.63
R vs Cult	"	"	149.49	12.85	104.53	13.66
F vs Cult	"	"	119.57	10.28	83.41	10.90

SNPs e cSNPs presenti in 7 genotipi di *V. vinifera*, *V. riparia* e nel genotipo complesso Freiburg 993-60 (tutti i genotipi). Sono riportate le comparazioni tra le diverse cultivar di *V. vinifera* (Cult vs Cult), tra le cultivar e il clone di *V. riparia* (R vs Cult) e tra le cultivar e l'ibrido complesso (F vs Cult). È stata calcolata la diversità nucleotidica media, sia per i frammenti interi che per le porzioni codificanti.

## Discussione dei risultati

Principale obiettivo di questo lavoro è l'analisi del genoma della vite per stimolare nuove strategie di *breeding* assistito da marcatori molecolari. Attualmente, i microsatelliti sono i marcatori più utilizzati per la caratterizzazione e la mappatura del genoma [35]. Tuttavia, se comparati con gli SNPs, risultano meno adatti agli studi di associazione a causa di fenomeni di omoplasia [37] e poiché sono molto meno frequenti degli SNPs nelle regioni codificanti. Inoltre, la rilevazione di SNPs può essere facilmente automatizzata per caratterizzare e mappare geni e per definire i diversi possibili aplotipi [26]. Il crescente numero di sequenze geniche disponibili nei database facilita lo sviluppo di un elevato numero di marcatori SNP nelle specie modello così come in quelle coltivate.

Per valutare i possibili sviluppi di marcatori SNPs per il *breeding* assistito, in questo studio sono stati considerati i geni coinvolti nel metabolismo degli zuccheri, nei meccanismi di trasduzione del segnale, nelle vie metaboliche degli antociani e nei meccanismi di difesa. I marcatori molecolari risultano essere particolarmente informativi quando presentano un elevato valore di PIC. Per questo motivo, i microsatelliti sono molto usati nella costruzione di mappe molecolari; infatti, presentano in

media un elevato numero di alleli, che corrisponde ad un PIC di 0.6 in soia [25] e di 0.77 in mais [33]. Invece, RFLP e AFLP mostrano valori inferiori, pari rispettivamente a 0.32 e 0.41 in soia [25], 0.33 e 0.32 in frumento [3] e 0.56 per RFLP in mais [33]. La nostra ricerca ha preso in considerazione 25 geni in 9 genotipi di vite, ottenendo un valore medio del PIC basato sugli aplotipi intragenici pari a 0.61. L'elevato valore del PIC e la natura multiallelica degli aplotipi rendono quindi gli SNPs maggiormente informativi quando sono scelti per monitorare aplotipi intragenici. Va ricordato che per ottenere valori così elevati di PIC è necessario analizzare più di uno SNP per frammento genico. Risultati simili sono stati trovati in *Beta vulgaris*, ottenendo tramite analisi SSCP e HA un valore del PIC di 0.47 [31].

Gli studi sulla variabilità genomica in vite sono appena iniziati. Relativamente poco è noto sulla variabilità tra cultivar, ibridi e tra diverse specie di *Vitis*. In questa ricerca, comparando i 9 diversi genotipi (sette cloni di *V. vinifera*, l'ibrido interspecifico e il clone di *V. riparia*), sono stati scoperti in totale 247 siti polimorfici in 11629 bp di sequenze genomiche, con una media di un SNP ogni 47 basi (11629 bp totali/247 SNP totali). Nel genoma di *Vitis vinifera*, considerando una dimensione di 475 Mb [20], si può dunque stimare che siano presenti  $4 \times 10^6$  SNPs, assumendo che il

campione di varietà da noi analizzato sia sufficientemente rappresentativo del germoplasma di vite esistente. Questi valori sono comparabili a quelli trovati in altre specie, come barbabietola da zucchero [31] e soia [39]. Un maggior numero di polimorfismi è stato osservato nelle regioni codificanti rispetto a quelle non codificanti ( $7563/177 = 23.1$  SNPs per kb vs  $3976/70 = 17.6$ ; nelle regioni codificanti  $\pi = 10.15$  contro un valore totale di 9.79). Si deve comunque notare che le regioni non codificanti utilizzate in questo studio si trovano nelle immediate vicinanze dei geni. Questa diversità nucleotidica potrebbe dunque essere indicativa di funzioni regolative o di splicing delle sequenze non codificanti, anche se appare difficile pensare a pressioni selettive maggiori su queste regioni di quelle che agiscono sulle regioni codificanti. Quando si considerano solo le regioni codificanti, dei 177 siti polimorfici osservati, 100 codificano per sostituzioni amminoacidiche. Queste scoperte meritano ulteriori approfondimenti; ciononostante si può ipotizzare che questi significativi cambiamenti amminoacidici possano alterare le funzioni delle proteine prodotte, contribuendo così all'estesa variabilità fenotipica osservata tra le cultivar di *V. vinifera*. I 25 frammenti genici sono stati selezionati per l'inferenza degli aplotipi. Un aplotipo è la specifica combinazione di nucleotidi osservata a ciascun sito poli-

morfico in un singolo cromosoma [32] ed è anche definito come genotipo multilocus aploide. È stata osservata una media di circa 6.04 diversi aplotipi per gene, che corrisponde a circa 0.7 volte il numero medio di SNPs identificati per gene. Quando i geni della vite saranno completamente caratterizzati, sarà interessante correlare le combinazioni aplotipiche con le differenze fenotipiche. Per questi motivi, dovrebbe essere considerato l'impiego degli aplotipi per mappare, tramite associazione, i geni responsabili della variazione di caratteri quantitativi [36].

## Riassunto

Con questo studio si è voluto approfondire la conoscenza del genoma di una delle più importanti colture fruttifere del mondo, la vite (*Vitis vinifera* L.).

Obiettivo principale è stato la valutazione del livello di polimorfismo presente all'interno del genoma. Grazie alle tecniche di amplificazione molecolare e ai progressi nelle procedure di sequenziamento è oggi possibile determinare la presenza di polimorfismi a singolo nucleotide (SNP) e di inserzioni/delezioni (indel), che rappresentano le differenze genetiche più frequenti tra gli individui di una popolazione. Gli SNP sono marcatori molecolari co-dominanti molto utili in quanto, grazie alla loro abbondanza ed ubiquità, consentono di studiare geni codificanti proteine a funzione nota e non solo trat-

**Tab. 4 - Numero di aplotipi e loro abbondanza rilevata per 25 geni**

Gene	bp	Nr. cromosomi	Numerosità di ciascun aplotipo nei 9 genotipi analizzati										Nr. aplotipi	Nr. cromosomi con aplotipi ≠ A	H	PIC
			A	B	C	D	E	F	G	H	I	J				
CHI-for	5382	16	7	3	1	1	0	0	0	0	0	0	4	5	0.58	0.53
CHI-rev	4950	16	7	2	2	1	1	1	0	0	0	0	6	7	0.69	0.66
IA06	3213	18	4	3	3	3	2	2	1	0	0	0	7	14	0.84	0.82
IB02	4329	18	5	5	1	1	1	1	1	1	1	1	10	13	0.82	0.80
ID04	3771	18	7	3	3	1	1	1	1	1	0	0	8	11	0.78	0.76
IE04	3825	18	7	2	2	2	2	1	1	1	0	0	8	11	0.79	0.77
IF01-for	1485	18	9	6	2	1	0	0	0	0	0	0	4	9	0.62	0.56
IF01-rev	5463	18	5	3	2	2	1	1	1	1	1	1	10	13	0.85	0.83
IF08-for	4986	16	10	2	2	1	1	0	0	0	0	0	5	6	0.57	0.54
IF08-rev	5085	18	10	2	2	1	1	1	1	0	0	0	7	8	0.65	0.63
IG04	3213	18	7	6	1	1	1	1	1	0	0	0	7	11	0.72	0.68
IH09	5085	14	6	4	3	2	1	0	0	0	0	0	5	10	0.74	0.70
IIA02	3645	16	7	4	1	1	1	0	0	0	0	0	5	7	0.65	0.60
IIA05	1710	18	8	3	3	1	1	1	1	0	0	0	7	10	0.73	0.70
IIB05	4797	18	5	3	2	2	2	1	1	1	1	0	9	13	0.85	0.83
IIC08	3762	18	8	7	2	1	0	0	0	0	0	0	4	10	0.64	0.57
IIE02	2961	18	7	3	3	2	1	1	1	0	0	0	7	11	0.77	0.74
IIG05	2322	18	17	1	0	0	0	0	0	0	0	0	2	1	0.10	0.10
IIIB08	2628	18	13	3	2	0	0	0	0	0	0	0	3	5	0.44	0.42
IIIB09	8406	14	5	3	2	1	1	1	1	0	0	0	7	9	0.79	0.76
IIIC03	2052	18	9	3	2	2	1	1	0	0	0	0	6	9	0.69	0.66
IIIC08	2088	18	14	2	1	1	0	0	0	0	0	0	4	4	0.38	0.35
IIIC12	3708	18	7	2	2	1	1	1	1	1	1	1	10	11	0.80	0.78
IVG10	3582	18	17	1	0	0	0	0	0	0	0	0	2	1	0.10	0.10
IVH04	2484	18	12	4	2	0	0	0	0	0	0	0	3	6	0.49	0.49
IVH06	2286	18	18	0	0	0	0	0	0	0	0	0	1	0	0.00	0.00
IVH09	3321	16	3	2	2	2	2	1	1	1	0	0	8	11	0.86	0.84
UFGT	4122	18	4	3	2	2	2	1	1	1	1	1	10	14	0.87	0.86
Valore medio			8.5	3.04	1.79	1.18	0.86	0.61	0.5	0.29	0.18	0.14	6.04	8.57	0.64	0.61

Numero di aplotipi e loro abbondanza rilevata per 25 geni. Per ogni gene sono stati indicati il valore del PIC e dell'eterozigotità. Con A è stato designato l'aplotipo più frequente a ciascun locus.

ti anonimi di DNA come altri tipi di marcatori. La presenza di polimorfismi SNPs all'interno del genoma della vite è stata determinata analizzando 25 ESTs in nove diversi genotipi (sei cultivars di *Vitis vinifera*, *Vitis riparia* e un ibrido interspecifico). I geni considerati nello studio appartengono a vie metaboliche di interesse in termini di qualità e resistenza alle malattie (metabolismo degli zuccheri, meccanismi di trasduzione del segnale, vie metaboliche degli antociani e meccanismi di difesa). Gli SNPs identificati hanno inoltre permesso di ottenere una stima della diversità nucleotidica intra e inter-specifica della vite.

## Bibliografia

[1] Alleweldt G. 1997. Genetics of grapevine breeding. *Progress in Botany* 58: 441-454.  
 [2] Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W. and Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.  
 [3] Bohn M., Utz F. and Melchinger A.E. 1999. Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs and SSRs and their use for predicting progeny variance. *Crop Sci.* 39: 228-

237.

[4] Bonfield J.K., Rada C. and Staden R. 1998. Automated detection of point mutations using fluorescent sequence trace subtraction. *Nucleic Acids Res.* 26: 3404-3409.

[5] Botstein D., White R.L., Skolnick M. and Davis R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.

[6] Cho R.J., Mindrinos M., Richards D.R., Sapolsky R.J., Anderson M., Drenkard E., Dewdney J., Reuber T.L., Stammers M., Federspiel N., Theologies A., Yang W.H., Hubbell E., Au M., Chung E.Y., Lashkary D., Lemieux B., Dean C., Lipshutz R.J.,

Ausubel F.M., Davis R.W. and Oefner P.J. 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genet.* 23: 203-207.

[7] Clark A.G. 1990. Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol. Biol. Evol.* 7: 111-122.

[8] Clark A.G., Weiss K.M., Nickerson D.A., Taylor S.L., Buchanan A., Stengård J., Salomaa V., Vartiainen E., Perola M., Boerwinkle E. and Sing C.F. 1998. Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am. J. Hum. Genet.* 63: 595-612.

[9] Collins F.S., Brooks L.D. and Chakravarti A. 1998.

- A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* 8: 1229-1231.
- [10] Dalbó M.A., Ye G.N., Weeden N.F., Steinkellner H., Sefe K.M. and Reisch B.I. 2000. Gene controlling sex in grapevine placed on a molecular marker-based genetic map. *Genome* 43(2): 333-340.
- [11] Doligez A., Bouquet A., Danglot Y., Lahogue F., Riaz S., Meredith C.P., Edwards K.J. and This P. 2002. Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTLs for seedlessness and berry weight. *Theor. Appl. Genet.* 105: 780-795.
- [12] Doyle J.J. and Doyle J.L. 1990. Isolation of plant DNA from fresh tissue. *Focus Biotech.* 12: 13-15.
- [13] Fischer B.M., Salakhutdinov I., Akkurt M., Eibach R., Edwards K.J., Topfer R. and Zyprian E.M. 2004. Quantitative trait locus analysis of fungal disease resistance factors on a molecular map of grapevine. *Theor. Appl. Genet.* 108(3): 501-515.
- [14] Fullerton S.M., Clark A.G., Weiss K.M., Nickerson D.A., Taylor S.L., Stengard J.H., Salomaa V., Vartiainen E., Perola M., Boerwinkle E. and Sing C.F. 2000. Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. *Am. J. Hum. Genet.* 67: 881-900.
- [15] Grando M.S., Bellin D., Edwards K.J., Pozzi C., Stefanini M. and Velasco R. 2003. Molecular linkage maps of *Vitis vinifera* and *Vitis riparia*. *Theor. Appl. Gen.* 106: 1213-1224.
- [16] Kanazin V., Talbert H., See D., DeCamp P., Nevo E. and Blake T. 2002. Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). *Plant Mol. Biol.* 48: 529-537.
- [17] Landgreen U., Nilsson M. and Kwok P.Y. 1998. Reading bits of genetic information. Methods for single nucleotide polymorphism analysis. *Genome Res.* 8: 769-776.
- [18] Li W.H. and Sadler L.A. 1991. Low nucleotide diversity in man. *Genetics* 129: 513-523.
- [19] Lindblad-Toh K., Wibchester E., Daly M.J., Wang D.G., Hirschhorn J.N., Lavolette J.P., Ardlie K., Reich D.E., Robinson E., Sklar P., Shah N., Thomas D., Fan J.B., Gigeras T., Warrington J., Patil N., Hudson T.J. and Lander E.S. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genet.* 24: 381-386.
- [20] Lodhi M.A. and Reisch B.I. 1995. Nuclear DNA content of *Vitis* species, cultivars, and other genera of the Vitaceae. *Theor. Appl. Genet.* 90: 11-16.
- [21] McCallum C.M., Comai L., Greene E.A. and Henikoff S. 2000. Targeted screening for induced mutations. *Nature Biotech.* 18: 455-457.
- [22] Nei M. and Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* 76: 5269-5273.
- [23] Nei M. and Tajima F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* 97: 145-163.
- [24] Orita M., Iwahana H., Kanazawa H., Hayashi K. and Sekiya T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single strand conformational polymorphisms. *Proc. Natl. Acad. Sci. USA* 86: 2766-2770.
- [25] Powell W., Morgante M., Andre C., Hanafey M., Vogel J., Tingey S.D. and Rafalsky A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2: 225-238.
- [26] Rafalski A. 2002. Application of single nucleotide polymorphisms in crop genetics. *Curr. Op. Plant Biol.* 5: 94-100.
- [27] Riaz S., Dangi G.S., Edwards K.J. and Meredith C.P. 2004. A microsatellite marker based framework linkage map of *Vitis vinifera* L. *Theor. Appl. Genet.* 108(5): 864-872.
- [28] Rozen S. and Skaletsky H.J. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S., Misener S. (eds), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, New Jersey, pp. 365-386.
- [29] Sachidanandam R., Weissman D., Schmidt S.C., Kakol J.M., Stein L.D., Marth G., Sherry S., Mulikin J.C., Mortimore B.J. and Willey D.L. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409: 928-933.
- [30] Salmaso M., Malacarne G., Troglio M., Faes G., Stefanini M., Grando S., Velasco R. A grapevine (*Vitis vinifera* L.) genetic map integrating the position of 141 expressed genes (submitted)
- [31] Schneider K., Weishaar B., Borchardt D.C. and Salamini F. 2001. SNPs frequency and allelic haplotype structure of Beta vulgaris expressed genes. *Mol. Breed.* 8: 63-74.
- [32] Stephens J.C., Schneider J.A., Tanguay D.A., Choi J., Acharya T., Stanley S.E., Jiang R., Messer C.J., Chew A., Han J.H., Duan J., Carr J.L., Lee M.S., Koshy B., Kumar A.M., Zhang G., Newell W.R., Windemuth A., Xu C., Kalbfleisch T.S., Shaner S.L., Arnold K., Schulz V., Drysdale C.M., Nandabalan K., Judson R.S., Ruano G. and Vovis G.F. 2001. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293: 489-493.
- [33] Taramino T. and Tingey S.D. 1996. Simple sequence repeats for germplasm analysis and mapping in maize. *Genome* 39: 277-287.
- [34] Tennailon M.I., Sawkins M.C., Long A.D., Gaut R.L., Doebley J.F. and Gaut B.S. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *Mays* L.). *Proc. Natl. Acad. Sci. USA* 98: 9161-9166.
- [35] Thomas M.R., Matsumoto S., Cain P. and Scott N.S. 1993. Repetitive DNA of grapevine: classes present and sequences suitable for cultivar identification. *Theor. Appl. Genet.* 86: 286-289.
- [36] Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D. and Buckler E.S. 4<sup>th</sup>. 2001. Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.* 28(3): 286-289.
- [37] Viard F., Frank P., Dubois M.P., Estoup A. and Jarne P. 1998. Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *J. Mol. Evol.* 47: 42-51.
- [38] White M.B., Carvalho M., Derse D., O'Brien S.J. and Dean M. 1992. Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* 12: 301-305.
- [39] Zhu Y.L., Song Q.J., Hyten D.L., Van Tassel C.P., Matukumalli L.K., Grimm D.R., Hyatt S.M., Fickus E.W., Young N.D. and Cregan P.B. 2003. Single-Nucleotide Polymorphisms in Soybean. *Genetics* 168: 1123-1134.

**Ringraziamenti.** Parte di questo lavoro fa riferimento alla pubblicazione Salmaso M., Faes G., Segala C., Stefanini M., Salakhutdinov I., Zyprian E., Topfer R., Grando M.S., Velasco R. (2005) *Genome diversity and gene haplotypes in grapevine (Vitis vinifera L.), as revealed by single nucleotide polymorphisms Mol. Breed.* 14: 385-395. Lo svolgimento di questo lavoro è stato possibile grazie alla collaborazione esistente con il Dipartimento di Genetica e Biologia Molecolare dell'Istituto Agrario di S. Michele all'Adige (TN), responsabile dott. Riccardo Velasco.

La ricerca rientra nelle finalità dei progetti "Biologia avanzata applicata a vite, melo e salmoneidi" e "Analisi della struttura del genoma della vite al fine di isolare geni rilevanti per il miglioramento della qualità delle uve", finanziati dalla Fondazione delle Casse di Risparmio di Trento e Rovereto (TN) e dalla Provincia Autonoma di Trento.