

**UNIVERSITÀ DEGLI STUDI DELLA TUSCIA DI VITERBO**

**DIPARTIMENTO DI SCIENZE E TECNOLOGIE PER L'AGRICOLTURA, LE FORESTE, LA  
NATURA E L'ENERGIA**

**Corso di Dottorato di Ricerca in**

**ORTOFLOROFRUTTICOLTURA - XXVI Ciclo**

**STUDI PRELIMINARI DI MAPPAGGIO PER ASSOCIAZIONE IN PESCO**

**(s.s.d. AGR/07)**

**Tesi di dottorato di:**

Dott.ssa Valeria Aramini

**Coordinatore del corso**

Prof. Giuseppe Colla

Firma .....

**Tutore**

Dott. Ignazio Verde

Firma .....

**Co-tutore**

Dott.ssa Sabrina Micali

Firma.....

**07 Luglio 2015**

# INDICE

<b>1. INTRODUZIONE.....</b>	<b>1</b>
1.1 IL PESCO.....	1
1.2 I MARCATORI MOLECOLARI.....	3
1.2.1 I microsatelliti (SSR).....	8
1.2.2 I marcatori SNP.....	11
1.3 LE MAPPE GENETICHE.....	19
1.3.1 Il mappaggio dei caratteri quantitativi.....	26
1.4 IL MAPPAGGIO PER ASSOCIAZIONE.....	29
1.4.1 Il Linkage Disequilibrium.....	32
1.4.2 Metodi statistici per il mappaggio per associazione.....	37
1.4.3 Studi di mappaggio per associazione nelle specie vegetali.....	38
1.5 IL SEQUENZIAMENTO DEL GENOMA DEL PESCO.....	41
1.6 IL CHIP IPSC 9K SNP.....	45
1.7 SCOPO DEL LAVORO.....	46
<b>2. MATERIALI E METODI.....</b>	<b>48</b>
2.1 MATERIALE VEGETALE.....	48
2.2 QUANTIFICAZIONE DEL DNA GENOMICO.....	48
2.3 GENOTIPIZZAZIONE CON MARCATORI SSR.....	50
2.4 GENOTIPIZZAZIONE CON MARCATORI SNP E DEFINIZIONE DELLA CORE COLLECTION.....	52
2.5 FENOTIPIZZAZIONE DELLA CORE COLLECTION.....	52
2.6 ANALISI DELLA STRUTTURA DI POPOLAZIONE, DELLE RELAZIONI DI PARENTELA TRA GENOTIPI E ANALISI DI ASSOCIAZIONE.....	53
<b>3. RISULTATI.....</b>	<b>56</b>
3.1 GENOTIPIZZAZIONE CON MARCATORI SSR.....	56
3.2 GENOTIPIZZAZIONE CON MARCATORI SNP E DEFINIZIONE DELLA CORE COLLECTION.....	58
3.3 FENOTIPIZZAZIONE DELLA CORE COLLECTION.....	65
3.4 ANALISI DELLA STRUTTURA DI POPOLAZIONE.....	67

3.5 ANALISI DI ASSOCIAZIONE.....	70
3.5.1 Caratteri qualitativi.....	70
3.5.1.1 Tomentosità del frutto.....	70
3.5.1.2 Colore della polpa.....	71
3.5.1.3 Tipo di Fiore.....	72
3.5.1.4 Forma delle ghiandole fogliari.....	73
3.5.2 Caratteri quantitativi.....	74
3.5.2.1 Epoca di fioritura.....	74
3.5.2.2 Epoca di maturazione.....	76
3.5.2.3 Acidità titolabile.....	78
3.5.2.4 Contenuto in solidi solubili (Brix).....	79
<b>4. DISCUSSIONE.....</b>	<b>83</b>
4.1 GENOTIPIZZAZIONE CON MARCATORI SSR.....	83
4.2 FENOTIPIZZAZIONE DELLA CORE COLLECTION.....	85
4.3 ANALISI DELLA STRUTTURA DI POPOLAZIONE.....	85
4.4 ANALISI DI ASSOCIAZIONE.....	86
4.4.1 Caratteri qualitativi.....	87
4.4.1.1 Tomentosità del frutto.....	87
4.4.1.2 Colore della polpa.....	88
4.4.1.3 Tipo di Fiore.....	88
4.4.1.4 Forma delle ghiandole fogliari.....	89
4.4.2 Caratteri quantitativi.....	89
4.4.2.1 Epoca di fioritura.....	89
4.4.2.2 Epoca di maturazione.....	90
4.4.2.3 Acidità titolabile.....	91
4.4.2.4 Contenuto in solidi solubili (Brix).....	92
<b>5. CONCLUSIONI.....</b>	<b>94</b>
<b>6. BIBLIOGRAFIA.....</b>	<b>99</b>

# 1. INTRODUZIONE

## 1.1 IL PESCO

La famiglia delle Rosaceae appartiene al gruppo delle Angiosperme (piante a fiore) e comprende circa 90 generi e 3000 specie, morfologicamente molto diverse tra loro. Le specie di questa famiglia vengono coltivate per i loro frutti (ad esempio pesche, albicocche, mele, fragole), per il legname pregiato (ad esempio il legno di ciliegio) o per il loro valore ornamentale (come le rose).

In base alle statistiche della FAO del 2012 (<http://faostat3.fao.org>), le Rosaceae includono le specie da frutto economicamente più importanti, con una produzione mondiale di circa 143 milioni di tonnellate l'anno (circa il 20% della produzione totale di frutta). La famiglia è suddivisa in tre sottofamiglie: le Dryadoideae, le Rosoideae e le Spiraeoideae (Potter et al., 2007); all'interno di quest'ultima troviamo il genere *Prunus* al quale appartengono diverse specie i cui frutti forniscono vitamine, minerali, fibre e composti antiossidanti alla nostra dieta. Tra queste specie, il pesco [*Prunus persica* (L.) Batsch] si colloca ai vertici per importanza economica, con una produzione mondiale annua di 21 milioni di tonnellate (FAOSTAT 2012, <http://faostat3.fao.org>).

Il centro di origine primario del pesco è l'Ovest della Cina, dove la specie è stata addomesticata circa 4000-5000 anni fa (Faust & Timon, 1995). La reale provenienza geografica è stata accertata in epoca relativamente recente, poiché per molto tempo le informazioni risultavano frammentarie anche a causa di notizie incomplete provenienti dai paesi orientali. Il nome della specie deriva dalla Persia (l'attuale Iran), che per molti secoli fu erroneamente ritenuta il luogo di origine. Il pesco, invece, giunse in Persia solo all'inizio del II secolo a.C., poco prima dell'avanzata dell'esercito romano in questa regione. In Italia, come documentato da scrittori greci e romani, la pianta di pesco fu introdotta nel primo secolo d.C. e raggiunse, indipendentemente e in modo quasi contemporaneo, la Francia attraverso la via dei Balcani e del Mar Nero. Infine, secondo quanto riportato da Hedrick (1917), il pesco arrivò nel continente americano, subito dopo la scoperta di Colombo, ad opera dei colonizzatori spagnoli e trovò condizioni così favorevoli che si diffuse molto rapidamente anche tra gli Aztechi in Messico. Un'altra introduzione sempre ad opera degli spagnoli sarebbe avvenuta, verosimilmente, in Florida intorno al 1500.

La coltivazione commerciale del pesco in California ebbe inizio intorno alla metà del XIX secolo e la produzione aumentò così rapidamente che questa coltura divenne la più importante della California.

Intorno al 1850, negli USA ci furono importazioni dirette di peschi dalla Cina: una cultivar a polpa bianca denominata 'Chinese Cling' venne inviata a New York e si diffuse notevolmente, tanto che 25 anni dopo questa varietà era coltivata in frutteti di ogni parte degli Stati Uniti. Numerose varietà moderne derivano da questo gruppo cinese, tra cui ad esempio le più conosciute cultivar americane a polpa gialla 'Elberta', 'Redhaven' e 'J.H. Hale', che vennero introdotte in Europa nella prima metà del XX secolo.

Data la sua grande importanza economica tra le specie da frutto, nel corso degli ultimi 120 anni sono stati condotti numerosi programmi di miglioramento genetico del pesco. Da allora, la gamma di tipologie di frutto disponibili per il consumatore è stata ampliata e diversificata: pesche e nettarine di grosse dimensioni, polpa gialla e bianca, polpa tenera e succosa o soda e croccante, forma rotonda o piatta, gusto della polpa dolce o acido, calendario di maturazione sempre più ampio. Il rinnovo varietale è proceduto a ritmi serrati, basti pensare che dal 1980 al 2008 sono state rilasciate circa 3200 nuove cultivar (Della Strada e Fideghelli, 2011).

Attualmente, i programmi di miglioramento genetico non solo del pesco ma, più in generale anche quelli delle altre colture, sono focalizzati soprattutto sul concetto di qualità. L'obiettivo odierno è quello di trovare un compromesso tra la domanda dell'industria di grandi volumi e le sempre maggiori richieste da parte del consumatore di qualità, salubrità e disponibilità sul mercato del prodotto.

Il concetto di "qualità" ha un significato che sottintende *un grado di eccellenza di un prodotto o la sua adeguatezza a un particolare uso*. Questa definizione assume valori diversi all'interno della filiera ortofrutticola: per il produttore rappresenta il controllo di costi e produttività, per la struttura di lavorazione la facilità di manipolazione e l'attitudine alla conservazione, per la distribuzione l'omogeneità del prodotto. Tuttavia la somma di questi valori non garantisce un prodotto di buona qualità al consumo.

Nel caso del frutto, la qualità può essere determinata a quattro diversi livelli: commerciale (pezzatura, colorazione, resistenza alle manipolazioni e attitudine alla conservazione), nutrizionale (composizione chimica del prodotto e contenuto in vitamine, antiossidanti ed elementi minerali), sanitario (assenza di residui chimici) e organolettico (grado di soddisfazione del consumatore relativamente a sapore, odore e sensazioni visive).

La finalità della filiera è quella di soddisfare le esigenze del consumatore finale, fornendo beni che rispondano al suo desiderio di qualità.

È quindi evidente che, essendo gli obiettivi in gioco molto più complessi rispetto al passato, la sfida dei breeder è oggi certamente più impegnativa. Tuttavia, attualmente abbiamo a disposizione conoscenze genetiche e biotecnologiche che possono essere opportunamente utilizzate per identificare geni utili in un miglioramento mirato a trasferire tutti i caratteri desiderati nelle cultivar commerciali.

## **1.2 I MARCATORI MOLECOLARI**

Uno dei più importanti progressi della biologia molecolare è stato lo sviluppo e l'utilizzo dei marcatori molecolari per analizzare la variabilità genetica nelle piante e negli animali.

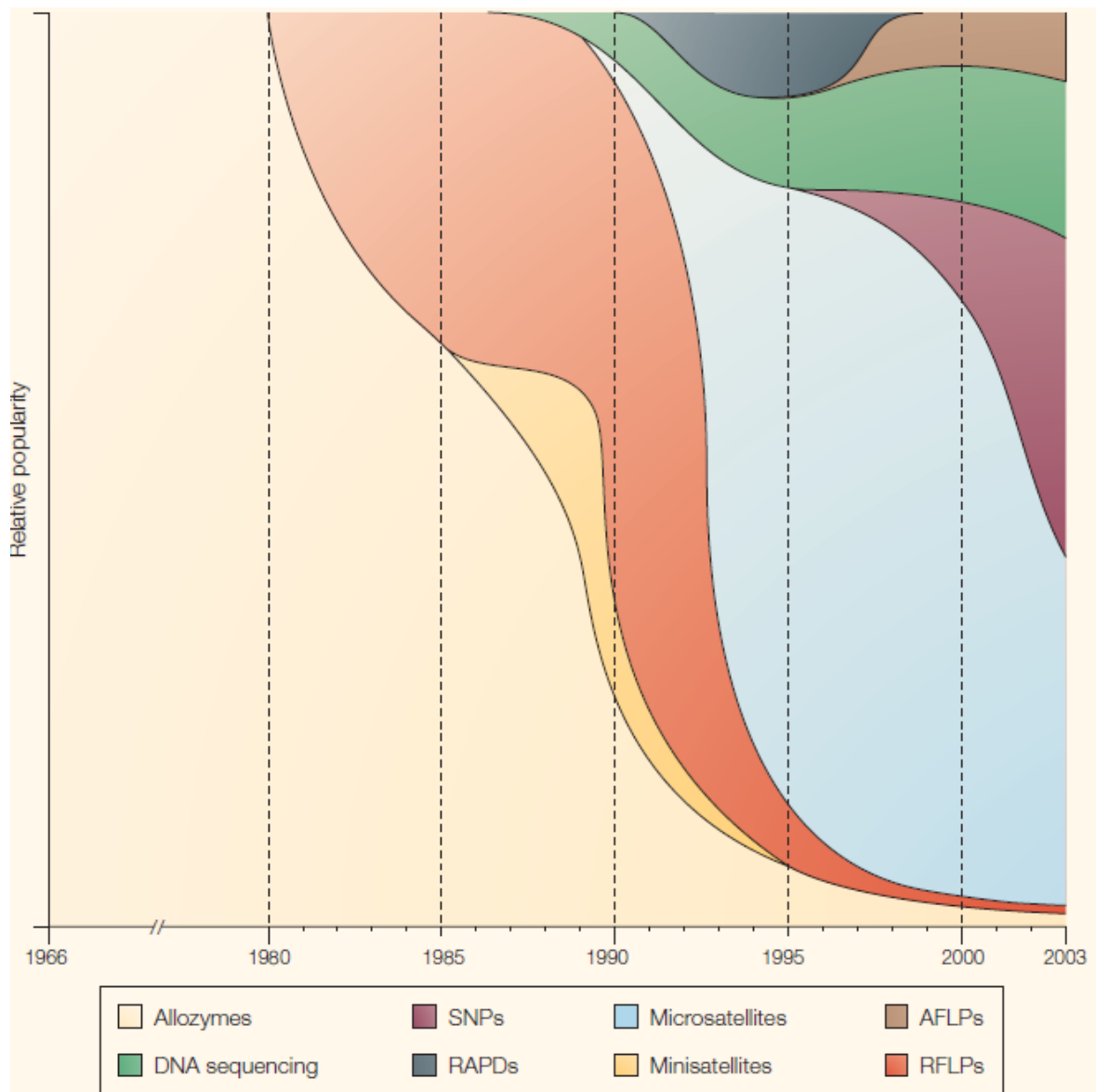
Il concetto più generico di marcatore genetico era già noto ai tempi di Mendel, il quale utilizzò dei marcatori morfologici nei suoi esperimenti nel XIX secolo. Successivamente, impiegando lo stesso tipo di marcatori, Sturtevant (1913) produsse la prima mappa genetica di *Drosophila melanogaster*.

Tuttavia tali marcatori, pur essendo semplici da individuare, in quanto la loro rilevazione si basa sull'analisi del fenotipo, presentano molte limitazioni: non sono molto numerosi, dipendono dalla specie di riferimento, sono influenzati dall'ambiente e non permettono di analizzare caratteri di tipo quantitativo.

Nel corso degli anni, i progressi della biologia molecolare hanno portato all'introduzione di numerosi nuovi tipi di marcatori (Fig. 1), dai marcatori biochimici (come gli isoenzimi e gli alloenzimi) fino ai marcatori molecolari, in grado di analizzare direttamente la sequenza del DNA, di cui attualmente esistono numerose tipologie.

Un marcatore molecolare può essere definito come “quel locus genomico, rilevabile con sonde o primer specifici che, in virtù della sua presenza, contraddistingue in maniera inequivocabile il tratto cromosomico con il quale si identifica e le regioni che lo circondano alle estremità 3' e 5'” (Barcaccia e Falcinelli, 2006).

L'analisi della variabilità genetica mediante i marcatori molecolari consiste nell'identificazione delle variazioni della sequenza nucleotidica dei marcatori stessi, cioè nell'analisi dei loro polimorfismi. Un marcatore molecolare, pertanto, consente di individuare polimorfismi specifici all'interno della sequenza di DNA di un individuo.



**Fig.1** Rappresentazione grafica della storia dello sviluppo dei marcatori molecolari. L'asse orizzontale rappresenta gli anni. In corrispondenza di ciascun anno, sull'asse verticale è rappresentato il livello di utilizzo dei marcatori molecolari. Com'è possibile osservare dall'immagine, l'aumento e la perdita di importanza delle diverse classi di marcatori riflette i progressi che sono stati fatti relativamente alla capacità di analizzare la variabilità genetica nel corso degli anni. Da Schlötterer, 2004.

I marcatori molecolari ideali devono presentare i seguenti requisiti:

- polimorfici: devono permettere di individuare differenze tra individui in regioni specifiche del genoma;
- devono essere uniformemente distribuiti nel genoma;
- multiallelici: devono presentare più di due alleli alternativi in uno stesso locus genico;
- stabili: non devono essere influenzati dall'ambiente;
- codominanti: devono consentire di distinguere la condizione eterozigote da quella dei due omozigoti;

- neutri: non devono avere effetto a livello fenotipico;
- non epistatici: non devono esserci interazioni inter o intralocus;
- semplici e veloci da rilevare;
- a basso costo;
- riproducibili.

I marcatori molecolari vengono suddivisi in due classi principali:

1. Marcatori basati su tecniche di ibridazione e restrizione. Sono i primi marcatori introdotti e utilizzano la tecnica Southern Blotting, mediante la quale i frammenti di DNA sono trasferiti su una membrana di nylon o di nitrocellulosa, ibridati con sonde marcate e successivamente esposti su lastra radiografica e visualizzati come bande discrete. Appartengono a questa categoria i marcatori RFLP e VNTR (o minisatelliti).
2. Marcatori basati sulla reazione a catena della polimerasi (PCR). Sfruttano la tecnica della PCR, che consente di amplificare per via enzimatica uno specifico frammento di DNA localizzato tra due regioni di sequenza nucleotidica nota, producendone un numero elevato di copie, attraverso una serie di cicli di denaturazione del DNA, appaiamento degli inneschi oligonucleotidici (primer) e polimerizzazione dei nuovi filamenti. Appartengono a questa categoria i marcatori RAPD, SSR, I-SSR, AFLP, CAPS, SCAR, SNP.

È inoltre possibile un'ulteriore classificazione dei marcatori molecolari in funzione del numero di loci che consentono di analizzare in un singolo esperimento. I marcatori multi-locus consentono di esaminare diversi loci contemporaneamente, in quanto prevedono l'amplificazione di sequenze nucleotidiche casuali utilizzando primer a sequenza nota arbitraria (ad esempio RAPD, AFLP, VNTR). Sono solitamente marcatori dominanti, cioè ad ogni locus si può evidenziare la presenza o l'assenza della banda ma non è possibile distinguere la condizione omozigote da quella eterozigote. I marcatori a singolo-locus, invece, permettono l'analisi di un unico locus specifico, poiché si basano sull'ibridazione o l'amplificazione di regioni genomiche note mediante l'utilizzo di sonde o primer specifici (ad esempio RFLP, SSR, SCAR, STS, SNP). Sono marcatori di tipo codominante, cioè permettono di distinguere la condizione omozigote da quella eterozigote.

Le tipologie di marcatori attualmente a disposizione sono molto numerose. Nella tabella 1 sono schematizzate le caratteristiche dei principali marcatori utilizzati e il loro impiego nel genere *Prunus*.

Nei paragrafi successivi verranno descritti in dettaglio i marcatori utilizzati in questo lavoro sperimentale: i microsatelliti (o SSR) e i marcatori che rilevano polimorfismi di singoli nucleotidi (SNP).

Marcatore	Abbondanza nei genomi	Riproducibilità	Grado di polimorfismo	Locus-specifico	Codominanza	Utilizzo nel genere <i>Prunus</i>
RFLP	Elevata	Elevata	Medio	Sì	Codominante	Rajapakse et al., 1995; Joobeur et al., 1998a; Dettori et al., 2001; Hauck et al., 2001; Vilanova et al., 2008
RAPD	Elevata	Bassa	Basso	No	Dominante	Lu et al., 1996; Warburton & Bliss, 1996; Bartolozzi et al., 1998; Rajapakse et al., 1995; Dirlewanger et al., 1998; Joobeur et al., 2000; Baránek et al., 2006
AFLP	Elevata	Elevata	Medio	No	Dominante	Hurtado et al., 2002; Geuna et al., 2003; Verde et al., 2005; Dirlewanger et al., 2006
CAPS	Bassa	Elevata	Basso	Sì	Codominante	Martins et al., 2005; Olmstead et al., 2008
SCAR	Bassa	Elevata	Medio	No	Dominante	Lecoals et al., 1999; Jun et al., 2002
IRAP/REMAP	Elevata	Elevata	Medio	Sì	Dominante	Yuying et al., 2011; Senková et al., 2013
SSR	Media	Elevata	Elevato	Sì	Codominante	Testolin et al., 2000; Dirlewanger et al., 2004; Aranzana et al., 2010; Cao et al., 2012; Li et al., 2013
SNP	Molto elevata	Elevata	Elevato	Sì	Codominante	Aranzana et al., 2012; Martínez-García et al., 2013a; Martínez-García et al., 2013b; Yang et al., 2013

**Tab. 1** Confronto delle caratteristiche dei principali marcatori molecolari utilizzati. RFLP: Restriction Fragment Length Polymorphism, RAPD: Random Amplified Polymorphic DNA, AFLP: Amplified Fragment Length Polymorphism, CAPS Cleaved Amplified Polymorphic Sequence, SCAR: Sequence Characterized Amplified Region, IRAP/REMAP: Inter-Retrotransposon Amplified Polymorphism/Retrotransposon-Microsatellite Amplified Polymorphism, SSR: Simple Sequence Repeats, SNP: Single Nucleotide Polymorphism.

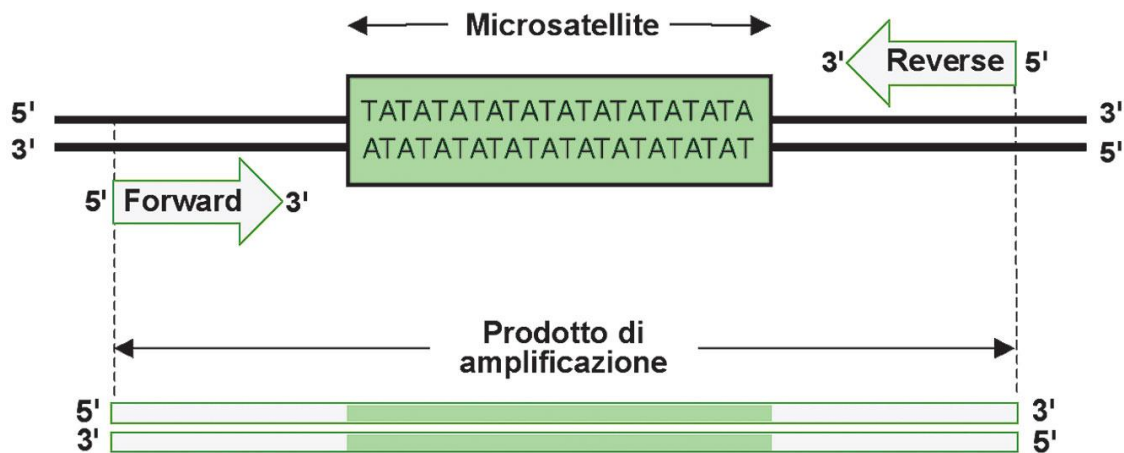
### 1.2.1 I microsatelliti (SSR)

I marcatori SSR derivano la loro denominazione da Simple Sequence Repeat (semplici sequenze ripetute) e sono anche noti come microsatelliti, termine utilizzato per la prima volta da Litt e Luty nel 1989. Sono delle corte sequenze di basi (1-6 bp) ripetute in tandem e disperse nei genomi eucariotici e procariotici, sia nelle regioni codificanti che nelle regioni non codificanti (Tautz e Renz, 1984).

In base alla lunghezza del motivo ripetuto i microsatelliti possono essere classificati come: i) dinucleotidici (ad esempio  $(CA)_n$ ,  $(TA)_n$ ); ii) trinucleotidici (ad esempio  $(GCC)_n$ ); iii) tetranucleotidici (ad esempio  $(AAAT)_n$ ); iv) pentanucleotidici (ad esempio  $(TCCCC)_n$ ). La maggior parte degli SSR individuati nei genomi sono dinucleotidici, ma la frequenza dei diversi motivi ripetuti è differente nelle piante e nell'uomo. Nei genomi dell'uomo e di molti mammiferi, ad esempio, il motivo  $(AC)_n$  è il più comune, seguito dal motivo  $(AT)_n$ . Nelle piante, invece, la ripetizione  $(AT)_n$  è la più frequente (Toth et al., 2000). Tuttavia, i microsatelliti con un numero di ripetizioni maggiore sono generalmente più polimorfici rispetto a quelli con minori ripetizioni. (Hellegren, 2004).

I polimorfismi vengono messi in evidenza utilizzando la PCR con primer specifici disegnati in modo da essere complementari alle regioni fiancheggianti le sequenze ripetute, poiché queste sono altamente conservate (Fig. 2). I primer così disegnati consentono di amplificare singoli microsatelliti che possono differire tra individui non per il motivo di base, ma per il numero di volte che questo motivo è ripetuto. Inizialmente i prodotti di amplificazione venivano solitamente sottoposti ad elettroforesi su gel di agarosio ad alta definizione (ad esempio Metaphor) o su gel di poliacrilammide poiché queste matrici consentono di evidenziare polimorfismi di pochi (2-5) nucleotidi.

Oggi la rilevazione di marcatori SSR è basata principalmente sull'uso di primer marcati con molecole fluorescenti e sull'impiego di sistemi automatizzati per l'analisi elettroforetica capillare con sequenziatori. Questi strumenti, oltre a consentire un'analisi più rapida, possiedono una risoluzione molto elevata e quindi permettono di discriminare frammenti che differiscono anche per una singola coppia di basi.



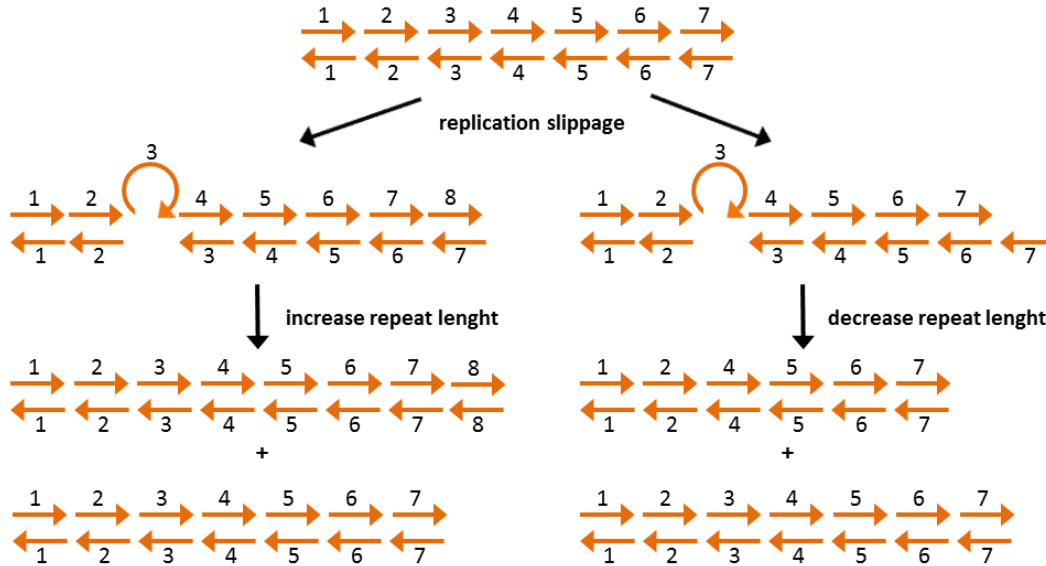
**Fig.2** Rappresentazione schematica dell'analisi di marcatori microsatelliti. Da Barcaccia e Falcinelli, 2006.

Sono stati proposti due modelli per spiegare l'origine dei polimorfismi osservati nei microsatelliti. Il primo è il cosiddetto 'slippage' (scivolamento) della DNA polimerasi. Tale meccanismo consiste in una sorta di slittamento della DNA polimerasi lungo il filamento copia del DNA durante il processo di replicazione. A volte, infatti, può accadere che l'enzima DNA polimerasi commetta degli errori durante la replicazione del DNA con conseguente aggiunta o delezione di basi ripetute. Solitamente questi errori vengono corretti dal sistema di riparazione delle basi del DNA, ma talvolta questo meccanismo non funziona correttamente e si generano i polimorfismi a livello delle sequenze microsatelliti (Schlötterer e Tautz, 1992; Richards e Sutherland, 1994; Fig. 3A). Il secondo modello, invece, prevede che le differenze a livello delle sequenze ripetute possano essere causate da eventi di crossing-over ineguale tra regioni ripetute durante la meiosi (Harding et al., 1992; Hancock, 1999; Richard & Paques, 2000; (Fig.3B).

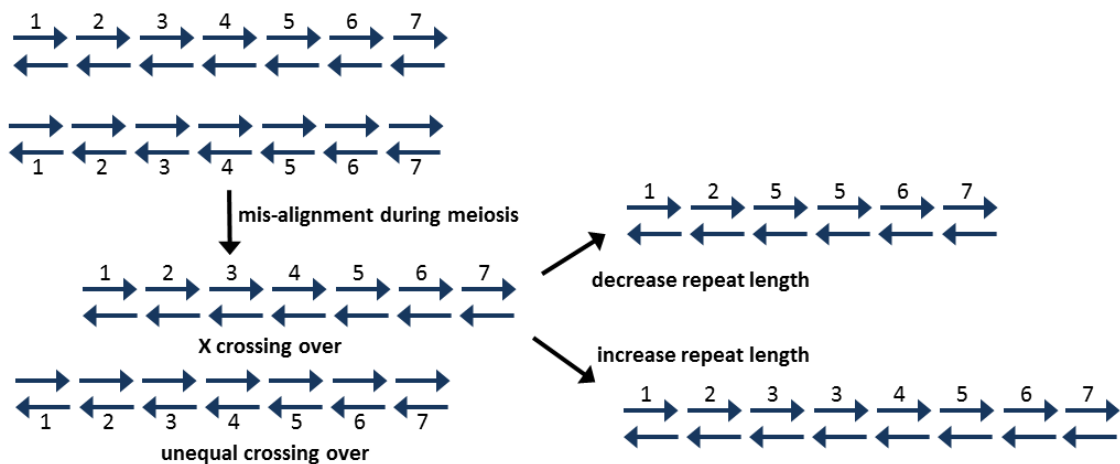
I marcatori microsatelliti possiedono numerosi vantaggi: sono codominanti, generalmente a singolo-locus, molto polimorfici (poichè presentano un elevato numero di alleli per locus), altamente riproducibili, sono ubiquitari nel genoma e sono semplici da analizzare. Quando furono introdotti, la principale limitazione al loro utilizzo era legata, per molte specie, alla possibilità di avere a disposizione sequenze dei microsatelliti e delle regioni fiancheggianti. Inizialmente, uno dei metodi più usati per isolare sequenze SSR a partire da DNA genomico era lo screening di librerie genomiche arricchite per sequenze microsatelliti (Squirrell et al., 2003). Oggi, grazie alla diminuzione dei costi di sequenziamento del DNA, per molte specie sono disponibili ampie collezioni di EST (Expressed Sequence Tag, cioè parti di sequenze espresse, di solito di ridotta lunghezza, ottenute dal sequenziamento di cloni di cDNA) e la sequenza completa del DNA genomico. La

disponibilità di quest'ultima, in particolare, consente, tramite l'utilizzo di strumenti bioinformatici, di isolare il catalogo completo di microsatelliti di una specie.

**A. Replication slippage during DNA replication**



**B. Unequal crossing-over during meiosis**



**Fig. 3.** Rappresentazione grafica dei due meccanismi che generano i polimorfismi SSR. Da Park et al., 2009.

L'impiego dei marcatori SSR nelle piante è stato documentato per la prima volta da Morgante e Olivieri nel 1993. Da allora, sono stati ampiamente utilizzati per l'analisi della variabilità genetica, per l'identificazione varietale, per studi evolutivisti e per la costruzione di mappe genetiche.

Per quanto riguarda la famiglia delle Rosaceae, sono state isolate e pubblicate moltissime coppie di primer microsatelliti. All'interno di questa famiglia, i primi SSR sono stati

descritti in pesco (Cipriani et al., 1999). Successivamente, sono stati sviluppati in mandorlo (Mnejja et al., 2005), albicocco (Hagen et al., 2004), susino giapponese (Mnejja et al., 2004), ciliegio (Vaughan & Russell, 2004), melo (Liebhard et al., 2002), pero (Yamamoto et al., 2002), fragola (Monfort et al., 2006), lampone (Graham et al., 2002), e rosa (Kimura et al., 2006).

Numerosi studi hanno dimostrato la trasferibilità dei marcatori SSR da una specie all'altra all'interno del genere *Prunus* (Wünsch & Hormaza, 2002; Dirlwanger et al., 2002; Mnejja et al., 2005; Wünsch, 2009; Vendramin et al. 2007). Questa caratteristica è molto importante, in quanto consente di utilizzare i medesimi loci SSR per studiare e confrontare la variabilità genetica di specie diverse o per la costruzione di mappe genetiche utilizzando lo stesso set di marcatori. Recentemente è stato individuato un set di 12 marcatori SSR, polimorfici in sei specie del genere *Prunus* (pesco, mandorlo, susino europeo, susino giapponese, ciliegio e albicocco) e ben distribuiti nel genoma, che potranno essere impiegati per caratterizzare la variabilità genetica in queste specie (Mnejja et al., 2010).

Un'altra importante applicazione dei marcatori microsatelliti riguarda lo studio della struttura di popolazione e del linkage disequilibrium (LD) in accessioni di pesco occidentali e orientali (Aranzana et al., 2010; Cao et al., 2012; Li et al., 2013). La comprensione dell'andamento del LD nel genoma di una specie è fondamentale per mettere a punto studi di mappaggio per associazione e i marcatori SSR si sono rivelati utili anche a questo scopo.

### **1.2.2 I marcatori SNP**

I marcatori SNP (Single Nucleotide Polymorphism), come suggerisce il loro nome, consentono di evidenziare differenze nella sequenza nucleotidica dovute a mutazioni puntiformi (sostituzioni) oppure a inserzioni o delezioni (indel).

Anche se, teoricamente, in ciascuna posizione di una sequenza nucleotidica ognuno dei quattro nucleotidi (A, C, T, G) potrebbe essere ugualmente presente, in realtà solo due di queste quattro possibilità sono in genere osservate in ogni posizione. Gli SNP, pertanto, sono considerati marcatori biallelici. Un primo motivo di questo fenomeno è la bassa frequenza con cui avvengono le sostituzioni di singoli nucleotidi all'interno di una sequenza genomica: è stato stimato che nei mammiferi le mutazioni puntiformi avvengono con una frequenza che varia tra  $1 \times 10^{-9}$  e  $5 \times 10^{-9}$  per nucleotide per anno, pertanto la probabilità che nella stessa posizione si verificano due eventi indipendenti di sostituzione

con due basi differenti è molto bassa (Vignal et al., 2002). Un altro motivo è la prevalenza di alcune mutazioni rispetto ad altre. Le mutazioni puntiformi possono essere di due tipi: i) transizioni, ossia sostituzione di una purina con un'altra purina ( $A \leftrightarrow G$ ) o di una pirimidina con un'altra pirimidina ( $C \leftrightarrow T$ ); ii) transversioni, cioè sostituzione di una purina con una pirimidina o viceversa ( $A \leftrightarrow C$ ,  $A \leftrightarrow T$ ,  $G \leftrightarrow C$ ,  $G \leftrightarrow T$ ). Vari studi hanno dimostrato che le transizioni avvengono più frequentemente rispetto alle transversioni (Wakeley, 1994, 1996; Batley et al., 2003, Aranzana et al., 2012).

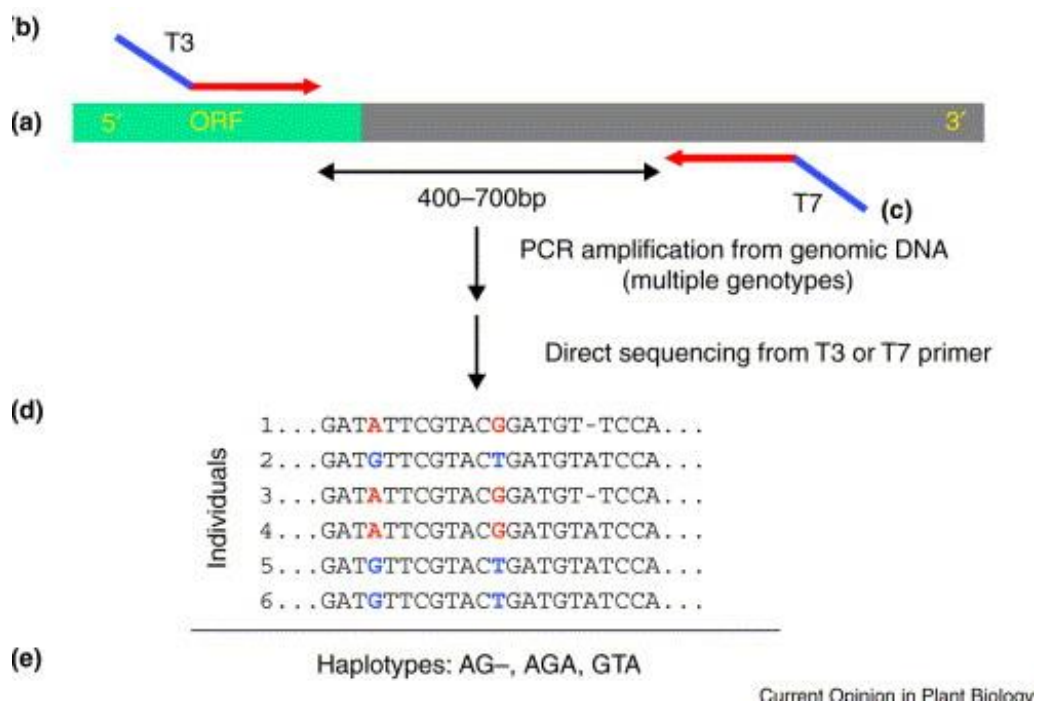
Nonostante gli SNP siano marcatori biallelici e quindi meno polimorfici di altri marcatori multiallelici come gli SSR, hanno ricevuto un forte interesse da parte della comunità scientifica, in quanto sono molto abbondanti nei genomi di piante e animali e possono essere presenti sia nelle regioni codificanti ("coding SNPs"), sia in quelle non codificanti ("non-coding SNPs"). Nell'uomo ad esempio, ne sono stati individuati più di 120 milioni ([http://www.ncbi.nlm.nih.gov/SNP/snp\\_summary.cgi](http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi)). Mediamente, nelle piante sono presenti con una frequenza di uno SNP ogni 40-500 pb, variabile a seconda delle specie: in cassava è stata stimata una frequenza di 1 SNP ogni 121 pb (Kawuki et al., 2009), in *Arabidopsis* una frequenza di 1 SNP/500 pb (Atwell et al., 2010), in mais di 1 SNP/43 pb (Jones et al., 2009), in vite di 1 SNP/117 bp (Lijavetzky et al., 2007), in melo di 1 SNP/52 pb (Micheletti et al., 2011), mentre in pesco è stato osservato uno SNP ogni 390 pb nelle regioni non-codificanti e uno ogni 1850 pb nelle regioni codificanti (Aranzana et al., 2012).

I marcatori SNP, quindi, sono più vantaggiosi rispetto alle generazioni precedenti di marcatori poiché, essendo molto frequenti nei genomi, possono essere trovati più facilmente all'interno dei geni di interesse (Syvanen, 2005). Infatti, in una data popolazione non tutti gli SNP saranno polimorfici, ma la loro elevata frequenza aumenta la probabilità di trovare almeno uno di questi marcatori all'interno dei geni di interesse o nelle immediate vicinanze. Questo rappresenta un grosso vantaggio per la selezione assistita da marcatori molecolari (MAS), in quanto i marcatori usati prima degli SNP erano, nella migliore delle ipotesi, strettamente associati ai loci di interesse (e non al loro interno) in una determinata progenie e questa associazione poteva facilmente essere persa quando tali marcatori erano applicati ad altre popolazioni con differente origine.

Il metodo più diretto per la rilevazione dei marcatori SNP è il sequenziamento di prodotti di PCR (Polymerase Chain Reaction) di campioni di DNA genomico isolati da un certo numero di individui geneticamente differenziati. I primer per la PCR possono essere disegnati in due modi: sfruttando le sequenze nucleotidiche di geni disponibili nelle banche

dati pubbliche, oppure selezionando cloni omologhi da genoteche. I prodotti di PCR vengono poi sequenziati in entrambe le direzioni e le sequenze ottenute vengono allineate per individuare le posizioni che presentano polimorfismi dovuti alla sostituzione di singoli nucleotidi (Fig. 4).

Tuttavia, questo approccio è piuttosto costoso, soprattutto se effettuato su larga scala, in quanto è necessario costruire primer locus-specifici, è limitato alle regioni per cui sono disponibili dati di sequenza e a volte possono essere generati falsi SNP, dovuti ad errori di sequenziamento.



**Fig. 4** Rappresentazione della rilevazione di marcatori SNP in una sequenza EST. In questo esempio, una sequenza in prossimità della regione 3'-UTR di una EST **(a)** è stata utilizzata per disegnare una coppia di primer per la PCR **(b, c)**, ai quali sono state agganciate le sequenze dei promotori T3 e T7 al fine di facilitare la successiva fase di sequenziamento. Dopo aver effettuato la PCR con tali primer su una serie di individui geneticamente differenti, i prodotti della PCR vengono purificati e sequenziati e le sequenze ottenute vengono allineate **(d)** per poter individuare polimorfismi di singoli nucleotidi **(e)**. Da Rafalski, 2002.

Un altro approccio per individuare marcatori SNP può essere utilizzato in quelle specie in cui sono stati avviati progetti di sequenziamento. In questo caso vengono allineate e confrontate sequenze di cloni BAC (Bacterial Artificial Chromosome) che si sovrappongono. Il problema di questo metodo è che gli SNP identificati non saranno uniformemente distribuiti nel genoma, in quanto la loro rilevazione dipende dal numero di cloni BAC di differente origine genetica che si sovrappongono, disponibili lungo il genoma (Vignal et al., 2002).

Oggi l'identificazione dei marcatori SNP è molto più rapida, in quanto, grazie allo sviluppo delle piattaforme di sequenziamento di nuova generazione (Next Generation Sequencing, NGS), è possibile ottenere il sequenziamento di interi genomi con costi e tempi ridotti rispetto alla tecnologia Sanger utilizzata in precedenza (Metzker, 2010). Se si ha a disposizione la sequenza dell'intero genoma di un genotipo di una specie (reference), il ri-sequenziamento di altre varietà della stessa specie (tramite le tecnologie NGS) consente di identificare rapidamente e in modo affidabile milioni di SNP, a seguito di allineamento delle sequenze delle varietà ri-sequenziate sulla sequenza originale, che funge da riferimento.

Le tecniche utilizzate per genotipizzare differenti individui con marcatori SNP sono numerose:

- PCR-RFLP. È stato un metodo ampiamente utilizzato in passato, quando gli SNP da analizzare erano localizzati all'interno dei siti di restrizione. Sostituzioni di singole basi all'interno di un sito di restrizione possono portare alla perdita o alla creazione del sito di restrizione stesso. Con questa tecnica, frammenti di PCR di differenti individui vengono digeriti con enzimi di restrizione e successivamente separati mediante elettroforesi. I vari individui presenteranno diversi pattern secondo la presenza o meno di SNP nei siti di restrizione.
- SSCP (Single Strand Conformation Polymorphism). Questo approccio si basa sul principio che la mobilità elettroforetica di frammenti di DNA a singolo filamento non dipende solo dalla loro lunghezza, ma anche dalla loro conformazione e quindi dalla sequenza nucleotidica. I singoli filamenti di DNA, infatti, formano delle strutture secondarie differenti a seconda della loro sequenza nucleotidica.  
Per effettuare questa tecnica, primer specifici vengono disegnati sulle regioni fiancheggianti la regione bersaglio e vengono poi utilizzati per amplificare il DNA genomico di differenti individui. I prodotti di PCR così ottenuti vengono denaturati in modo da ottenere frammenti di DNA a singolo filamento e successivamente vengono separati mediante elettroforesi in condizioni non denaturanti. Un frammento con una mutazione o uno SNP ha una conformazione differente rispetto al frammento non mutato e di conseguenza i due frammenti correranno in modo diverso su gel. Questa tecnica è stata utilizzata per analizzare marcatori SNP nel genere *Picea* (Germano & Klein, 1999).
- DGGE (Denaturing Gradient Gel Electrophoresis). Questa tecnica sfrutta il principio per cui la temperatura di denaturazione del DNA (Temperatura di Melting,  $T_m$ )

dipende dalla sua sequenza nucleotidica ed è peculiare di ogni frammento. Variazioni di un singolo nucleotide modificano la temperatura di denaturazione del frammento e di conseguenza la sua mobilità elettroforetica.

Innanzitutto viene effettuata una reazione di PCR con primer modificati mediante l'aggiunta di una coda stabilizzante ricca in G e C (GC clamp), al fine di ottenere amplificati fiancheggiati da regioni altamente stabili con temperature di melting elevate (maggiore è il contenuto in G e C in una sequenza, più elevata è la temperatura di denaturazione del frammento). Gli amplificati vengono poi fatti migrare su un gel di poliacrilammide a gradiente denaturante crescente, utilizzando denaturanti chimici e alte temperature. Il raggiungimento della temperatura di denaturazione comporta una parziale apertura della doppia elica e quindi un rallentamento nella progressione del frammento sul gel. Frammenti che differiscono anche per una singola sostituzione nucleotidica hanno temperature di melting diverse e si dissociano a differenti concentrazioni del gel, formando così bande visualizzabili a diverse altezze.

- HA (Heteroduplex Analysis). Questa tecnica si basa sul presupposto che la mobilità elettroforetica degli eteroduplex è inferiore rispetto a quella degli omoduplex. Questi duplex si formano quando un frammento amplificato di DNA mutato e uno non mutato vengono denaturati termicamente e lasciati appaiare a temperatura ambiente. L'appaiamento di due filamenti identici nella fase di re-annealing porta alla formazione di omoduplex, mentre se vi è una qualsiasi variazione di sequenza della molecola originale (wild type) si ha la formazione di eteroduplex, combinazione di due catene di DNA non perfettamente complementari, caratterizzate dalla presenza di un disappaiamento (mismatch), al sito della mutazione e quindi la non perfetta complementarità.

Per analizzare marcatori SNP mediante questo approccio, innanzitutto viene fatta una reazione di PCR con il campione wild type e i campioni mutati. Gli amplificati del controllo e dei campioni mutati vengono poi miscelati in parti uguali, denaturati a 95°C e lasciati riassociare a temperatura ambiente, in modo da permettere la formazione degli eteroduplex. I frammenti vengono poi fatti correre su gel e quelli mutati vengono visualizzati come bande che hanno una differente posizione sul gel, in quanto, come detto prima, le bande dell'eteroduplex migrano nel gel più lentamente rispetto alle corrispondenti bande dell'omoduplex.

- DHPLC (Denaturing High Performance Liquid Chromatography). Questa è una tecnica che deriva dalla HA, in quanto si basa sulla differente velocità di migrazione,

in un mezzo simile ad un gel quale è la colonna HPLC, per gli eteroduplex e gli omoduplex. È stata ampiamente utilizzata in quanto è completamente automatizzabile, sensibile e rapida.

L'eteroduplex si comporta cromatograficamente in modo diverso rispetto all'omoduplex non mutato, i tempi di ritenzione sono differenti, in particolare l'eteroduplex è solitamente più veloce (meno trattenuto) dell'omoduplex e da ciò si può caratterizzare la presenza di una mutazione in un campione.

La DHPLC rileva la differenza tra la molecola omoduplex, che ad una determinata temperatura (chiamata temperatura di "quasi-denaturazione") è ancora sotto forma di doppia elica e quelle di eteroduplex, che alla stessa temperatura mostrano una parziale denaturazione in corrispondenza del sito dove si è verificato un mis-appaiamento. La presenza di una mutazione si evidenzia sotto forma di picchi non presenti nel cromatogramma del wild type.

La tecnica DHPLC impiega il meccanismo di ripartizione in fase inversa ad accoppiamento ionico (RP-IP), che separa le molecole e le eluisce sulla base della dimensione crescente delle molecole.

Anche in questo caso prima di tutto viene fatta una reazione di PCR con il campione wild type e i campioni mutati. Gli amplificati del controllo e dei campioni mutati vengono poi miscelati in parti uguali, denaturati a 95°C e lasciati riassociare a temperatura ambiente, in modo da permettere la formazione degli eteroduplex. Le coppie di omoduplex e di eteroduplex vengono poi iniettate nel sistema DHPLC e un software collegato allo strumento controlla la temperatura del campione. Alla temperatura di "quasi-denaturazione" (che è specifica per ogni frammento), la molecola omoduplex è intatta, mentre quelle degli eteroduplex mostrano una parziale denaturazione nel sito di non perfetto appaiamento dovuto alla presenza della mutazione. Gli eteroduplex eluiscono dalla colonna prima rispetto agli omoduplex. Ciò si verifica perchè nel punto della mutazione si formano delle porzioni di DNA a singola elica (dette bolle), dovute al fatto che le singole eliche non sono in grado di formare legami a idrogeno in corrispondenza di basi non correttamente appaiate. La proprietà fondamentale di tali bolle, nel DNA a singolo filamento, è quella di avere una minore densità di gruppi fosfato e quindi una minore densità di carica rispetto al doppio filamento, garantendo una minore ritenzione della molecola di DNA sulla fase stazionaria e un tempo di eluizione inferiore a quello dell'omoduplex.

La presenza di una mutazione si riconosce quando il tracciato cromatografico mostra più di un singolo picco.

- HRM (High Resolution Melting). Questa tecnica viene usata per caratterizzare campioni di DNA in base al loro comportamento durante la dissociazione del doppio filamento che si verifica con l'aumento della temperatura (melting curve) e rileva la presenza di mutazioni puntiformi in frammenti di circa 100-300 bp.

Prima di tutto viene fatta una reazione di PCR utilizzando primer disegnati in modo da amplificare la regione contenente lo SNP. Alla mix di PCR viene aggiunto anche un fluoroforo specifico in grado di legare solo il DNA a doppio filamento e non quello a singolo filamento. Gli ampliconi contenenti il fluoroforo vengono poi portati ad una temperatura crescente e contestualmente lo strumento misura il livello di fluorescenza. Man mano che la temperatura aumenta, viene raggiunta la temperatura di melting specifica di ogni frammento, il quale si denatura e rilascia il fluoroforo ad esso legato e di conseguenza la fluorescenza diminuisce. Il risultato viene visualizzato graficamente come una serie di curve di melting, ognuna specifica di un frammento, rappresentate come variazione del livello di fluorescenza rispetto alla temperatura. Come detto precedentemente, la temperatura di denaturazione di un frammento di DNA dipende dalla sua sequenza nucleotidica, pertanto gli individui che hanno mutazioni SNP vengono riconosciuti in quanto mostrano curve di melting differenti rispetto gli individui che non hanno la mutazione.

Negli ultimi anni sono state sviluppate delle piattaforme che permettono di genotipizzare in un'unica analisi milioni di marcatori SNP per singolo campione di DNA. Queste tecnologie sono basate sull'utilizzo di un supporto solido (array o chip) sul quale vengono legate le sonde contenenti gli SNP e sul quale viene ibridato il campione di DNA da analizzare (Gunderson et al., 2005); ne è un esempio la tecnologia BeadArray della ditta Illumina. Attualmente questi array sono stati realizzati in molte specie vegetali: in mais (Ganal et al., 2011), in riso (Zhao et al., 2011), in melo (Chagnè et al., 2012; Bianco et al., 2014), in ciliegio (Peace et al., 2012), in vite (Myles et al., 2011), in pesco (Verde et al., 2012) e in fragola (Bassil et al., 2015).

Grazie alle tecniche di sequenziamento di nuova generazione (NGS) e allo sviluppo di questi chip, oggi i marcatori SNP possono essere facilmente utilizzati per diversi tipi di studi. Innanzitutto con un numero molto elevato di marcatori è possibile costruire mappe genetiche ad alta risoluzione e questo, come detto prima, consente di individuare marcatori strettamente associati ai caratteri di interesse. Inoltre, con mappe sature è più facile

confrontare mappe genetiche di differenti popolazioni e individuare con maggior precisione differenze nel genoma o riarrangiamenti cromosomici. Mappe genetiche con migliaia di marcatori consentono anche di mappare in modo più accurato caratteri quantitativi (Quantitative Trait Loci, QTL; Ganal et al., 2012).

Gli SNP stanno diventando i marcatori d'elezione per l'analisi della variabilità e per mappare caratteri di interesse nel genere *Prunus*. Sono stati utilizzati in *Prunus mume*, in mandorlo, pesco, cilieglio dolce, ciliegio acido e albicocco per l'identificazione varietale e per l'analisi della variabilità genetica (Li et al., 2010; Fang et al., 2006; Wu et al., 2010; Aranzana et al., 2012; Fernández i Marti et al., 2012; Sun et al., 2013; Rosyara et al., 2014; Salazar et al., 2015). Più recentemente, Bielsa et al. (2014) hanno impiegato i marcatori SNP per caratterizzare differenti portinnesti del genere *Prunus*. In mandorlo, Wu et al. (2009) e Tavassolian et al. (2010) hanno creato le prime mappe genetiche che combinano gli SNP con altri marcatori, quali RAPD e SSR, mentre in pesco sono state costruite mappe genetiche con un'elevata densità di marcatori SNP al fine di individuare QTL relativi alla qualità del frutto, ai danni da freddo, ai composti aromatici volatili, alla resistenza alla maculatura batterica e al marciume bruno, all'epoca di maturazione (Martínez-García et al., 2013a; Martínez-García et al., 2013b; Frett et al., 2014; da Silva et al., 2015; Eduardo et al., 2012; Yang et al., 2013; Pacheco et al., 2014; Pirona et al., 2013). Anche in ciliegio dolce sono state analizzate due differenti progenie intraspecifiche mediante marcatori SNP al fine di individuare marcatori associati a caratteri di interesse agronomico (Klagges et al., 2013).

In pesco, inoltre, l'analisi degli SNP mediante la piattaforma Sequenom è stata utilizzata per mappare in modo fine le regioni genomiche che controllano la data di maturazione nel gruppo di associazione 4 (Pirona et al., 2013) e il locus *G* (carattere pesco/nettarina) sul gruppo di associazione 5 (Vendramin et al., 2014). In entrambi i casi, tale approccio ha permesso di individuare geni candidati responsabili di questi caratteri.

Infine, l'analisi dell'intero genoma di pesco con migliaia di marcatori SNP è stata utilizzata per effettuare un'analisi della variabilità genetica, dell'estensione del linkage disequilibrium e uno studio di mappaggio per associazione (Micheletti et al., sottomesso). Questo tipo di analisi, come verrà discusso più avanti, consente di identificare caratteri qualitativi e quantitativi associati a determinati fenotipi con grande precisione.

### 1.3 LE MAPPE GENETICHE

Lo sviluppo e la disponibilità di un numero elevato di marcatori molecolari hanno dato un grande impulso alla costruzione di mappe genetiche sature, che coprono l'intero genoma di molte specie di interesse agrario.

Una mappa genetica è la rappresentazione grafica dei singoli cromosomi di una specie, sui quali vengono indicati la posizione e l'ordine dei loci marcatori l'uno rispetto all'altro.

La costruzione delle mappe genetiche si basa sull'analisi delle frequenze di ricombinazione ricavate analizzando la segregazione di marcatori molecolari in progenie ottenute con appropriati programmi di incrocio. La frequenza di ricombinazione, rappresenta la frequenza con cui si verifica il crossing-over tra due loci marcatori, ed è funzione della loro distanza: maggiore è la distanza tra i loci, maggiore è la probabilità che un crossing-over si verifichi e interrompa la loro associazione. Sulla base di questo principio, per realizzare una mappa genetica vengono confrontati i dati di segregazione dei marcatori molecolari a coppie e in questo modo vengono definite le associazioni tra di essi, vengono calcolate le distanze genetiche e ad ogni marcatore viene assegnata una posizione in relazione agli altri marcatori dello stesso gruppo di associazione. La distanza tra i marcatori è espressa in unità di mappa o centiMorgan (cM): 1 cM è pari ad una distanza che fa ottenere un gamete ricombinante ogni 100 prodotti della meiosi, ossia ogni 25 meiosi. Un cM, quindi, equivale alla distanza tra due marcatori che dà una frequenza di ricombinazione dell'1%.

Le popolazioni segreganti maggiormente utilizzate nei vegetali per lo sviluppo di mappe genetiche sono le  $F_2$ , ottenute mediante autofecondazione dell'ibrido  $F_1$ , e le  $BC_1$ , prodotte attraverso il reincrocio dell'ibrido  $F_1$  con una delle due linee parentali. Tuttavia, nelle specie annuali che non si propagano vegetativamente, queste popolazioni non possono essere mantenute nel tempo e quindi non consentono l'analisi di caratteri di tipo quantitativo, fortemente influenzati dall'ambiente, che richiedono prove replicate dello stesso genotipo per avere una buona stima del valore fenotipico. Per ovviare a questo inconveniente sono state proposte strategie alternative, come le RIL (Recombinant Inbred Lines) e le linee di-aploidi o linee DH (Double Haploids). Per ottenere le RIL è necessario effettuare 5-6 cicli di autofecondazione dopo la produzione di una popolazione  $F_2$  a partire da genitori omozigoti (linee pure). In  $F_6$  le piante che compongono la popolazione sono altamente omozigoti e risultano fissate per corti blocchi di associazione. Queste linee possono essere mantenute mediante autofecondazione ed offrono la possibilità di effettuare prove replicate nello spazio (località diverse) e nel tempo (annate successive). Una

popolazione di linee di-aploidi, invece, può essere prodotta mediante androgenesi attraverso coltura di antere o polline e successivo raddoppiamento cromosomico. In questo modo è possibile ottenere velocemente linee omozigoti a tutti i loci che possono essere mantenute nel tempo mediante autofecondazione e, ugualmente alle RIL, consentono di aggiungere in tempi successivi nuovi marcatori ad una mappa precedentemente prodotta.

Per il mappaggio con marcatori molecolari è comunque possibile utilizzare anche la popolazione  $F_1$ , se ottenuta attraverso l'incrocio di due genotipi altamente eterozigoti oppure di un genotipo eterozigote e una linea omozigote.

Le mappe genetiche possono essere costruite per due scopi: 1) per realizzare la mappa di riferimento di una data specie, sulla quale poter posizionare progressivamente nuovi marcatori; 2) per determinare il numero, la localizzazione e gli effetti di loci che controllano specifici tratti di interesse (Peace & Norelli, 2009).

Numerose mappe genetiche, basate su incroci intraspecifici o interspecifici, sono state ottenute all'interno del genere *Prunus*: in albicocco (Hurtado et al., 2002; Lambert et al., 2004; Dondini et al., 2007; Lalli et al., 2008; Olukolu et al., 2009), in ciliegio (Olmstead et al., 2008; Clarke et al., 2009), in mandorlo (Viruel et al., 1995; Joobeur et al., 2000, Sánchez-Pérez et al., 2007; Tavassolian et al., 2010). In pesco la prima mappa è stata realizzata da Chaparro et al (1994), utilizzando 83 marcatori RAPD, un isoenzima e 4 caratteri morfologici. Poco dopo sono state pubblicate altre due mappe, costruite utilizzando marcatori RFLP: la prima è stata prodotta con una popolazione  $F_2$  di pesco (Rajapakse et al., 1995), la seconda con una popolazione  $F_2$  ottenuta da un incrocio intraspecifico pesco x mandorlo (Foolad et al., 1995). Le mappe genetiche costruite successivamente in pesco hanno integrato marcatori dominanti, come RAPD e AFLP, e marcatori codominanti e morfologici (Dirlewanger et al., 1998) oppure sono state costruite interamente con marcatori AFLP (Lu et al., 1998). Queste mappe, tuttavia, non coprivano l'intero genoma, generalmente individuavano un numero di gruppi di associazione inferiore agli 8 attesi per il genere *Prunus* (numero cromosomico di base  $x=n=8$ ) e avevano una ridotta densità media di marcatori (un marcatore ogni 4.5-8.5 cM).

La prima mappa satura di pesco è stata ottenuta nell'ambito di un progetto europeo, utilizzando un ibrido mandorlo ('Texas') x pesco ('Earligold') e marcatori molecolari trasferibili da una specie all'altra (11 isoenzimi e 226 RFLP; Joobeur et al., 1998a). Tutti questi marcatori erano distribuiti negli otto gruppi di associazione con una distanza di 491 cM e con una densità media di un marcatore ogni 2 cM. Questa mappa, nota come T x E, è stata successivamente sviluppata con l'aggiunta di 176 marcatori SSR e 123 RFLP

(Aranzana et al., 2003; Dirlewanger et al., 2004), molti dei quali derivanti da *Arabidopsis thaliana*.

Molti marcatori sono stati localizzati sulla mappa T x E mediante il metodo della mappatura selettiva, descritta da Howad et al. (2005). Grazie a questo approccio, analizzando i marcatori su un piccolo sottoinsieme di individui della popolazione di partenza, scelti in modo da essere rappresentativi, nel loro insieme, di tutti gli eventi di ricombinazione (definiti BIN) distribuiti su tutti i cromosomi, è possibile identificare in modo rapido la posizione dei marcatori stessi sulla mappa.

Per la popolazione T x E, il sottoinsieme per la mappatura selettiva consiste di solo 8 individui (il parentale 'Earligold', l'ibrido F<sub>1</sub> e 6 piante della progenie F<sub>2</sub>) e ogni marcatore codominante polimorfico ricade in uno dei 67 BIN definiti sulla mappa.

Grazie alla strategia della mappatura selettiva, sono stati inseriti molti altri marcatori all'interno della mappa T x E: marcatori SSR, STS, geni candidati (CG), marcatori RosCOS (Rosaceae Conserved Ortholog Set; Howad et al., 2005; Chen et al., 2008; Ogundiwin et al., 2008,2009; Cabrera et al., 2009; Sargent et al., 2009; Illa et al., 2011). I marcatori SSR sono stati ampiamente utilizzati per la costruzione di mappe genetiche in pesco, in quanto sono altamente polimorfici in questa specie, mostrano un elevato livello di eterozigosità osservata e consentono di confrontare facilmente mappe differenti (Peace & Norelli, 2009; Dirlewanger et al., 2005b).

La comunità scientifica ha deciso di adottare la T x E come mappa di riferimento del genere *Prunus*; sulla base di questa mappa è stata definita una terminologia comune per numerare i gruppi di associazione, ad esempio il gruppo di associazione 1 (G1) è il più lungo e corrisponde al cromosoma 1, anche esso identificato come il più lungo (Salesses & Mouras, 1977). Questa mappa è stata un'importante risorsa per gli studi relativi al genere *Prunus*, in quanto, essendo basata su un incrocio interspecifico, è molto polimorfica e consente di mappare marcatori che non segregano in altri incroci intraspecifici. Inoltre, sulla mappa T x E sono localizzati numerosi marcatori trasferibili ad altre specie (isoenzimi, RFLP e SSR) che sono stati utilizzati per la costruzione di mappe in altre progenie del genere *Prunus*, definiti loci 'ancora'. In uno studio condotto dieci anni fa, confrontando la posizione dei marcatori 'ancora' sulla mappa T x E con quella su altre 13 mappe costruite utilizzando diverse specie del genere *Prunus*, è stato osservato che i genomi delle specie diploidi (2n=16) pesco, mandorlo, albicocco, ciliegio, *P. davidiana*, *P. cerasifera*, *P. ferganensis* sono sostanzialmente colineari (Dirlewanger et al., 2004).

La disponibilità di una mappa di riferimento saturo è stata molto utile anche per altri motivi: i marcatori RFLP posizionati su di essa sono stati utilizzati come punto di partenza per la costruzione della mappa fisica del genere *Prunus* (Zhebentyayeva et al., 2008) e successivamente la mappa T x E è stata utilizzata per ancorare ed orientare su di essa gli scaffold ottenuti dal sequenziamento del genoma del pesco, al fine di individuare la loro collocazione cromosomica (Verde et al., 2013).

Grazie all'esistenza di una singola mappa di riferimento è stato inoltre possibile definire la posizione di caratteri qualitativi (major genes) e QTL di interesse agronomico che segregano in popolazioni differenti (Tab. 2). Tuttavia, nonostante siano disponibili tutte queste informazioni, l'utilizzo dei marcatori molecolari per l'ottenimento di nuove cultivar commerciali è ancora limitato. La selezione assistita da marcatori (MAS) in pesco è stata utilizzata solo per introdurre i due geni indipendenti *Mi* e *Ma* che conferiscono resistenza a nematodi (*Meloidogyne*) in nuovi portinnesti (Dirlewanger et al., 2005a). Non è ancora stato riportato l'impiego della selezione assistita con marcatori associati a geni che regolano caratteri legati alla qualità del frutto. Ciò è dovuto al fatto che non vi è una sufficiente associazione tra i marcatori disponibili e i geni e al fatto che la maggior parte dei caratteri legati al frutto (epoca di maturazione, contenuto in solidi solubili, capacità antiossidante, massa del frutto, etc) sono di tipo quantitativo (QTL) e quindi la loro dissezione genetica è piuttosto complessa. Diversi sono gli studi relativi ai QTL responsabili della qualità del frutto in pesco (Dirlewanger et al., 1999; Verde et al., 2002; Etienne et al., 2002a/2002b; Ogundiwin et al., 2009; Eduardo et al., 2012; Martínez-García et al., 2013a), ma è necessaria una maggiore conoscenza del numero, degli effetti e della posizione di questi QTL affinché possano essere usati regolarmente nei programmi di selezione assistita da marcatori.

Carattere	Gruppo di associazione	Simbolo	Popolazioni	Riferimento bibliografico
<b>Caratteri morfologici</b>				
‘Evergrowing’	G1	<i>Evg</i>	‘Empress op dwarf’ x PI442380	Wang et al., 2002
Lunghezza degli internodi	G1	QTL	( <i>P. ferganensis</i> x IF7310828) BC <sub>1</sub>	Verde et al., 2002
Colore del fiore	G1	<i>B</i>	‘Garfi’ x ‘Nemared’	Jaregui, 1998
Habitus ‘pillar’ o ‘broomy’	G2	<i>Br</i>	Varie progenie	Scorza et al., 2002
Fiore doppio	G2	<i>Dl</i>	NC174RL x PI	Chaparro et al., 1994
Colore delle antere (giallo/rosso)	G3	<i>Ag</i>	‘Texas’ x ‘Earligold’	Joobeur, 1998b
Altezza della pianta	G4	QTL	‘Venus’ x ‘BigTop’	Cantín et al., 2010
Ovaio pluricarpellare	G3	<i>Pcp</i>	‘Padre’ x 54P455	Bliss et al., 2002
Colore del fiore	G3	<i>Fc</i>	‘Akame’ x ‘Jusetou’	Yamamoto et al., 2001
Altezza della pianta (normale/nano)	G6	<i>Dw</i>	‘Akame’ x ‘Jusetou’	Yamamoto et al., 2001
Forma delle foglie (stretta/larga)	G6	<i>Nl</i>	‘Akame’ x ‘Jusetou’	Yamamoto et al., 2001
Maschiosterilità	G6	<i>Ps</i>	‘Ferjalou Jalousia®’ x ‘Fantasia’	Dirlewanger et al., 1998; 2006
Colore delle foglie (rosso/giallo)	G6-G8	<i>Gr</i>	‘Garfi’ x ‘Nemared’; ‘Akame’ x ‘Jusetou’	Verde et al., 2002, Yamamoto et al., 2001
Forma delle ghiandole fogliari (reniformi/globose/assenti)	G7	<i>E</i>	( <i>P. ferganensis</i> x IF310828) BC1	Dettori et al., 2001
Morfologia del fiore	G8	<i>Sh</i>	‘Contender’ x Fla.92-2C; Pop-DG	Fan et al., 2010; Ogundiwin et al., 2009

<b>Caratteri fenologici</b>				
Fabbisogno in freddo e in caldo, epoca di fioritura	G1	QTL	‘Contender’ x Fla.92-2C	Fan et al., 2010
Epoca di maturazione	G2	QTL	( <i>P. ferganensis</i> x IF310828) BC <sub>1</sub>	Verde et al., 2002
Epoca di fioritura	G2	QTL	‘Contender’ x Fla.92-2C, ‘Summergrand’ x P1908	Fan et al., 2010; Quilot et al., 2004
Epoca di fioritura, epoca di maturazione, periodo di sviluppo del frutto	G4	QTL	‘Ferjalou Jalousia®’ x ‘Fantasia’; ( <i>P. ferganensis</i> x IF7310828) BC <sub>1</sub> ; Venus’ x ‘BigTop’; ‘Summergrand’ x P1908	Etienne et al., 2002b; Verde et al., 2002; Cantín et al., 2010, Quilot et al., 2004
Fabbisogno in freddo, epoca di fioritura	G4-G5-G6-G7	QTL	‘Contender’ x Fla.92-2C	Fan et al., 2010
Epoca di maturazione	G6	QTL	( <i>P. ferganensis</i> x IF310828) BC <sub>1</sub>	Verde et al., 2002
Fabbisogno in freddo e in caldo	G8	QTL	‘Contender’ x Fla.92-2C	Fan et al., 2010
<b>Caratteri legati alla qualità del frutto</b>				
Colore della polpa (bianco/giallo)	G1	<i>Y</i>	‘Padre’ x 54P455	Bliss et al., 2002; Warburton et al., 1996
Colore della buccia, contenuto in solidi solubili	G2	QTL	( <i>P. ferganensis</i> x IF7310828) BC <sub>1</sub>	Verde et al., 2002
Peso del frutto, diametro del frutto, contenuto in glucosio	G3	QTLs	‘Suncrest’ x ‘Bailey’	Abbott et al., 1998
Colore della polpa intorno al nocciolo	G3	<i>C<sub>s</sub></i>	‘Akame’ x ‘Jusetou’	Yamamoto et al., 2001
Peso del frutto, diametro del frutto, contenuto in glucosio	G3	QTL	‘Suncrest’ x ‘Bailey’	Abbott et al., 1998
Aderenza della polpa al nocciolo (clingstone/freestone)	G4	<i>F</i>	( <i>P. ferganensis</i> x IF7310828) BC <sub>1</sub> ; ‘Akame’ x ‘Jusetou’	Dettori et al., 2001; Yamamoto et al., 2001

Contenuto in solidi solubili, fruttosio e glucosio	G4	QTL	‘Ferjalou Jalousia®’ x ‘Fantasia’; Venus’ x ‘BigTop’; ‘Summergrand’ x P1908	Cantín et al., 2010; Quilot et al., 2004; Dirlewanger et al., 2006.
Caratteri legati ai danni da freddo	G4	QTL	‘Venus’ x ‘BigTop’	Cantín et al., 2010
Dimensioni del frutto	G4	QTL	‘Venus’ x ‘BigTop’; ‘Summergrand’ x P1908; ‘Contender’ x ‘Ambra’	Cantín et al., 2010; Quilot et al., 2004; Eduardo et al., 2011
pH, acidità titolabile	G4	QTL	‘Venus’ x ‘BigTop’	Cantín et al., 2010
Frutto non acido	G5	<i>D</i>	‘Ferjalou Jalousia®’ x ‘Fantasia’	Dirlewanger et al., 1998; 1999; Boudehri et al., 2009.
Saccarosio, malato, acidità titolabile, pH	G5	QTL	‘Ferjalou Jalousia®’ x ‘Fantasia’; ‘Summergrand’ x P1908	Etienne et al., 2002b; Quilot et al., 2004.
Pesco/nettarina	G5	<i>G</i>	‘Ferjalou Jalousia®’ x ‘Fantasia’; ‘Padre’ x 54P455; ‘Contender’ x ‘Ambra’	Dirlewanger et al., 1998; 1999; Bliss et al., 2002; Vendramin et al., 2014
Dimensioni del frutto	G5	QTL	‘Summergrand’ x P1908	Quilot et al., 2004
Sapore del nocciolo (dolce/amaro)	G5	<i>Sk</i>	‘Padre’ x 54P455	Bliss et al., 2002
Colore della buccia, contenuto in solidi solubili	G6	QTL	( <i>P. ferganensis</i> x IF310828) BC1	Verde et al., 2002
Forma del frutto (piatta/rotonda)	G6	<i>S*</i>	‘Ferjalou Jalousia®’ x ‘Fantasia’	Dirlewanger et al., 1998; 1999; 2006
Frutto abortivo	G6	<i>Af</i>	‘Ferjalou Jalousia®’ x ‘Fantasia’	Dirlewanger et al., 2006
Colore della buccia	G6-G8	<i>Sc</i>	‘Akame’ x ‘Jusetou’	Yamamoto et al., 2001
Chinasi	G8	QTL	‘Ferjalou Jalousia®’ x ‘Fantasia’	Etienne et al., 2002b

Resistenza a stress biotici				
Resistenza ad Oidio	G1-G6	QTL	'Summergrand' x P1908	Foulongne et al., 2003
Resistenza a Plum Pox Virus	G1-G2-G4-G5-G6-G7	QTL	'Summergrand' x P1908; 'Summergrand' x P1908 F2; 'Rubira' x P1908	Decroocq et al., 2005; Marandel et al., 2009; Rubio et al., 2010.
Resistenza a nematodi galligeni	G2	<i>Mi(c)</i>	P.2175 x GN22; 'Akame' x 'Jusetou', 'Lovell' x 'Nemared', 'Garfi' x 'Nemared', 'Padre' x 54P455	Claverie et al., 2004; Yamamoto et al., 2001; Lu et al., 1998; Bliss et al., 2002; Jaregui, 1998
Resistenza alla Bolla del Pesco	G3-G6	QTL	'Summergrand' x P1908	Viruel et al., 1998
Resistenza ad Oidio	G6	Vr2	'Rubira' x 'Pamirskij 5'	Pascal et al., 2010
Resistenza ad Oidio	G7	QTL	( <i>P. ferganensis</i> x IF7310828) BC <sub>1</sub>	Verde et al., 2002
Resistenza ad Oidio	G8	QTL	'Summergrand' x P1908 ( <i>P. ferganensis</i> x IF7310828) BC <sub>1</sub>	Foulongne et al., 2003; Verde et al., 2002.

**Tab. 2** Major genes e QTL che controllano caratteri morfologici o agronomici in pesco. Da Verde et al., 2013.

### 1.3.1 Il mappaggio dei caratteri quantitativi

La maggior parte dei caratteri di interesse agronomico delle specie vegetali non possono essere descritti come alternativi o antagonisti perché le differenze risultano essere graduali lungo una scala continua di valori. Tali caratteri vengono detti quantitativi e a differenza di quelli qualitativi variano in modo continuo nella popolazione, non possono essere classificati secondo classi discrete ma possono essere misurati e quindi descritti mediante parametri numerici. I caratteri quantitativi sono controllati da numerosi geni (chiamati poligeni), ognuno dei quali ha un diverso effetto sul carattere e sono notevolmente influenzati dall'ambiente. Le posizioni occupate da questi geni sui cromosomi sono chiamate loci per caratteri quantitativi e vengono indicate con l'acronimo QTL (Quantitative Trait Loci).

La variabilità fenotipica osservabile in una popolazione per un determinato carattere quantitativo può quindi avere due componenti: una genetica, determinata dai geni che controllano il carattere e una ambientale, dovuta alle condizioni pedo-climatiche e alle pratiche colturali adottate.

Il mappaggio dei QTL è una pratica di base utilizzata nella genetica quantitativa che ha avuto inizio con l'avvento dei marcatori molecolari e la costruzione delle mappe genetiche che permettono di analizzare numerosi caratteri di interesse agronomico. L'obiettivo del mappaggio di uno o più QTL è quello di associare un carattere quantitativo ad uno o più marcatori molecolari aventi possibilmente una posizione di mappa nota.

Il primo passaggio in qualunque esperimento di mappaggio di QTL è la costituzione di una popolazione segregante mediante l'incrocio tra individui parentali con caratteristiche antagoniste per il carattere quantitativo di interesse. Successivamente tale popolazione deve essere genotipizzata mediante marcatori molecolari al fine di ottenere una mappa genetica in cui questi marcatori siano ordinati nei vari gruppi di associazione corrispondenti ai cromosomi del corredo di base della specie. È importante utilizzare un numero elevato di marcatori, in quanto più densamente satura è la mappa genetica ottenuta, maggiore è la probabilità di trovare una associazione stretta tra loci coinvolti nel controllo del carattere e i marcatori e quindi di localizzare i QTL con precisione.

Per completare l'analisi è necessario raccogliere i dati fenotipici, ossia misurare il carattere quantitativo negli individui della popolazione, ed effettuare l'analisi statistica per individuare le associazioni tra i marcatori e il carattere (Doerge, 2002).

Per il mappaggio dei QTL sono disponibili diversi software (come ad esempio MAPQTL, JoinMap, QTL Cartographer), ognuno dei quali è in grado di effettuare un'analisi statistica dei dati attraverso l'utilizzo di diversi approcci statistici che variano in base al numero di marcatori utilizzati contemporaneamente per l'analisi. Esistono due differenti metodi per il mappaggio dei QTL: l'analisi al singolo marcatore e l'analisi a intervalli (semplice, SIM oppure composto, CIM).

L'analisi al singolo marcatore indica quali marcatori sono associati al carattere quantitativo di interesse e quindi rileva l'esistenza di potenziali QTL. In questo approccio, per ogni locus marcatore la popolazione viene suddivisa in gruppi distinti in base alla loro composizione allelica a quel marcatore e per ogni gruppo viene calcolato il valore fenotipico medio del carattere quantitativo. Applicando poi un test statistico (come ANOVA, t di Student o l'analisi di regressione) viene valutata la significatività della differenza tra le medie. L'ipotesi nulla testata è che la media del valore fenotipico è

indipendente dal genotipo per un dato marcatore. L'ipotesi nulla viene rigettata quando il valore del test statistico è più grande di un valore soglia e ciò indica la presenza di un QTL in prossimità del locus marcatore in esame. Gli svantaggi di questa tecnica sono che non consente di definire con accuratezza l'effetto del QTL sui valori fenotipici e non permette di localizzare con sufficiente precisione la posizione più probabile di un QTL nell'intervallo tra due marcatori.

Il mappaggio ad intervalli semplice (SIM, Simple Interval Mapping) stima la probabilità della presenza di un QTL per ogni intervallo di mappa formato da coppie di loci marcatori adiacenti. Il metodo statistico su cui si basa è quello del LOD (Logarithm of the Odds) che equivale al  $\log_{10} (L_1/L_0)$ , dove  $L_1$  e  $L_0$  rappresentano rispettivamente la probabilità di associazione e di indipendenza tra QTL e marcatore molecolare. Qualora il LOD calcolato risulti superiore ad un valore soglia prestabilito, si può accettare la presenza di un QTL nel tratto cromosomico delimitato dai due marcatori analizzati. L'inconveniente di questo metodo di mappaggio è che non consente di valutare gli effetti di altri QTL associati sullo stesso cromosoma o presenti in cromosomi diversi. Questo problema è stato superato grazie al metodo del mappaggio ad intervalli composto (CIM, Composite Interval Mapping), il quale combina il mappaggio di un QTL in un dato intervallo sulla base dei valori di LOD con una analisi di regressione multipla di marcatori associati con altri QTL. Con questo approccio è possibile localizzare i QTL con maggior precisione.

Numerosi esperimenti di analisi QTL sono stati effettuati in pesco (Tab. 2) permettendo di studiare le basi genetiche di diversi caratteri quantitativi. Tuttavia, come detto precedentemente, la maggior parte di questi QTL non vengono utilizzati per la selezione assistita da marcatori molecolari nei programmi di breeding. Ciò è in massima parte dovuto al fatto che il mappaggio dei QTL mediante l'utilizzo di popolazioni bi-parentali non consente di avere un'elevata risoluzione di mappa e di conseguenza per molti caratteri le regioni dove sono stati localizzati i QTL sono molto ampie, possono variare da pochi cM a decine di cM e possono comprendere milioni di paia di basi e centinaia di geni.

## 1.4 IL MAPPAGGIO PER ASSOCIAZIONE

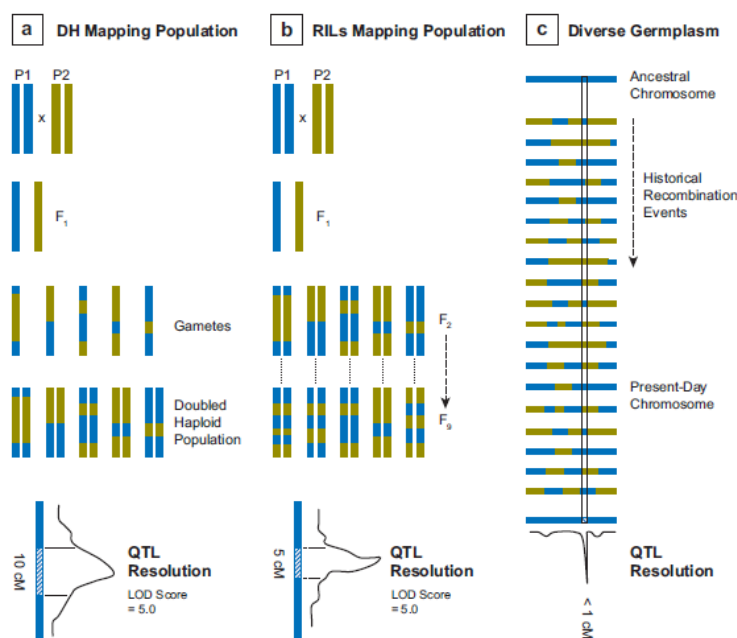
L'obiettivo principale dei *breeders* è quello di analizzare la variabilità fenotipica e spiegare come questa sia associata a cambiamenti nella sequenza del DNA. La costruzione di mappe genetiche mediante l'utilizzo di popolazioni segreganti è stato il metodo più comunemente usato per raggiungere questo obiettivo (Myles et al., 2009). Un approccio alternativo, che consente il mappaggio di caratteri qualitativi e quantitativi con maggior accuratezza, è quello basato sulla stima del linkage disequilibrium, noto come mappaggio per associazione. Questa tecnica consiste nel cercare associazioni tra genotipo e fenotipo in un campione di individui di diversa origine genetica, come ad esempio popolazioni naturali o collezioni di germoplasma.

Sia il mappaggio con popolazioni bi-parentali che il mappaggio per associazione si basano sulla capacità della ricombinazione di suddividere il genoma in frammenti che possono essere correlati alla variabilità fenotipica; la differenza sostanziale tra i due metodi è il controllo che si ha sulla ricombinazione (Fig. 5). Da un lato, infatti, il mappaggio con popolazioni bi-parentali è un esperimento altamente controllato: due individui vengono incrociati al fine di generare una popolazione segregante in cui le relazioni di parentela sono note. Si crea così un sistema chiuso e si usano dei marcatori molecolari per inferire la posizione dei siti di ricombinazione e costruire in questo modo una mappa genetica. Avendo a disposizione questi strumenti, il ricercatore può determinare se un frammento cromosomico compreso tra due siti di ricombinazione è associato ad un carattere fenotipico. Utilizzando questo sistema però si hanno a disposizione solo gli eventi di ricombinazione che si sono verificati durante la costituzione della popolazione segregante. In questo caso la ricombinazione non ha avuto abbastanza tempo per suddividere il genoma in piccoli frammenti e i QTL vengono in genere localizzati in regioni genomiche molto ampie; per poter avere una risoluzione di mappa più fine è necessario costituire popolazioni molto numerose (Fig 5a,b). Inoltre, i QTL individuati sono popolazione-specifici, cioè dipendono dai parentali utilizzati per ottenere la popolazione (e solo dagli alleli portati da questi) e non è detto che segreghino in altre popolazioni.

D'altro lato il mappaggio per associazione non è un esperimento controllato, ma piuttosto un esperimento naturale. In questo tipo di approccio, vengono raccolti dati genotipici e fenotipici da una popolazione in cui le relazioni di parentela non sono controllate dallo sperimentatore e vengono cercate le associazioni tra questi dati genetici e fenotipici. Il principale vantaggio del mappaggio per associazione è che sfrutta tutti gli eventi di

ricombinazione che si sono verificati durante la storia evolutiva della specie e di conseguenza consente di avere una più elevata risoluzione di mappa rispetto a quella ottenuta con popolazioni bi-parentali (Fig. 5c), ma è difficile stabilire quando e dove si è verificata la ricombinazione. Un altro importante vantaggio è che tramite il mappaggio per associazione è possibile mappare molti caratteri di interesse agronomico in una singola collezione di germoplasma, anziché dover costruire una popolazione specifica per ogni carattere da analizzare, come è necessario fare per il mappaggio con popolazioni bi-parentali. Inoltre, il numero di QTL che possono essere mappati per un dato fenotipo non è limitato a quelli che segregano tra i parentali di un incrocio, ma è possibile mappare un numero molto più ampio di QTL che controllano il carattere, che varia a seconda del grado con cui la popolazione naturale utilizzata racchiude la totale diversità genetica disponibile in natura (Zhu et al., 2008). Teoricamente, avendo a disposizione una popolazione che rappresenta tutta la variabilità genetica di una data specie, tramite il mappaggio per associazione sarebbe possibile mappare tutti i QTL che controllano un dato carattere.

Lo svantaggio del mappaggio per associazione è che si tratta di un esperimento meno controllato e ci potrebbero essere problemi di struttura della popolazione che possono portare ad associazioni spurie tra genotipo e fenotipo, come verrà discusso più in dettaglio successivamente.



**Fig. 5** Confronto della risoluzione di mappa ottenuta usando popolazioni bi-parentali e una collezione di germoplasma. In a) è stata utilizzata una popolazione Di-Aploide (DH), in b) una popolazione Recombinant Inbred Line (RIL) e in c) una collezione di germoplasma.

Sebbene il mappaggio per associazione presenti numerosi vantaggi rispetto al mappaggio con popolazioni bi-parentali, questi approcci sono spesso utilizzati insieme, soprattutto per validare le associazioni individuate con il mappaggio per associazione, al fine di ridurre le associazioni spurie (Soto-Cerda & Cloutier, 2012).

Esistono due diverse metodologie di mappaggio per associazione: 1) l'approccio per geni candidati (candidate gene association) e 2) l'analisi dell'intero genoma (Whole Genome Scan o Genome-Wide Association mapping). Con l'approccio per geni candidati viene testata l'ipotesi che ci sia una correlazione tra variazioni della sequenza nucleotidica in un determinato gene e il carattere di interesse. Ad esempio, si può fare un esperimento in cui si analizza se in una collezione di germoplasma di mais ci sia una correlazione tra variazioni della sequenza di DNA degli alleli della fitoene sintasi (o di qualunque altro gene coinvolto nella biosintesi dei carotenoidi) e contenuto di carotenoidi nei semi (Palaisa et al., 2003; Harjes et al., 2008). Un approccio di questo tipo presuppone una buona conoscenza delle caratteristiche biochimiche e genetiche del carattere di interesse. Nel caso in cui non si conoscano nel dettaglio tutti i geni che costituiscono una determinata via biosintetica si può utilizzare il mappaggio per associazione mediante analisi dell'intero genoma. Questo metodo consiste nell'analizzare interamente il genoma di un dato organismo, mediante la genotipizzazione con marcatori molecolari uniformemente distribuiti lungo tutto il genoma. L'ipotesi che viene testata è che uno o più marcatori molecolari siano associati al carattere di interesse.

La messa a punto di un esperimento di mappaggio per associazione mediante analisi dell'intero genoma richiede i seguenti passaggi: 1) selezione di un gruppo di individui da una popolazione naturale o da una collezione di germoplasma con una buona copertura della diversità genetica della specie; 2) misurazione dei caratteri fenotipici di interesse all'interno della popolazione, preferibilmente effettuata in diversi ambienti e in diversi anni; 3) genotipizzazione della popolazione con marcatori molecolari; 4) quantificazione dell'estensione del linkage disequilibrium (LD), utilizzando i dati ottenuti dalla genotipizzazione con marcatori molecolari; 5) valutazione della struttura della popolazione (ossia della suddivisione della popolazione in gruppi distinti in cui le frequenze alleliche sono significativamente differenti) e delle relazioni di parentela tra gli individui della popolazione (cioè calcolo del coefficiente di parentela tra coppie di individui della popolazione); 6) sulla base delle informazioni ottenute dalla quantificazione dell'LD e dalla struttura di popolazione, correlazione dei dati fenotipici e genotipici mediante

l'utilizzo di appropriati programmi statistici ed eventuale individuazione di marcatori associati ai caratteri di interesse.

Il punto di partenza fondamentale per qualunque esperimento di mappaggio per associazione è l'analisi dell'estensione dell'LD, in quanto il successo di tale approccio dipende soprattutto dal livello dell'LD tra i marcatori genotipizzati e i QTL coinvolti nel controllo genico dei caratteri di interesse.

#### **1.4.1 Il Linkage Disequilibrium**

L'LD è una condizione per la quale, in una popolazione, gli alleli di un locus non sono associati in modo casuale con gli alleli di un altro locus. L'LD indica la correlazione tra polimorfismi (ad es. in marcatori SNP) dovuta alla loro storia condivisa di mutazioni ed eventi di ricombinazione (Flint-Garcia et al., 2003). In una popolazione ampia, dove gli incroci avvengono in modo casuale, con loci che segregano in maniera indipendente e in assenza di selezione, mutazione o migrazione, i loci polimorfici saranno in linkage equilibrium. Al contrario, fenomeni come la selezione, le unioni non casuali, la struttura di popolazione, la deriva genetica, aumenteranno i livelli dell'LD all'interno della popolazione stessa.

Talvolta l'LD viene confuso con il linkage, ma in realtà sono due cose ben distinte: il linkage è una condizione per cui due o più loci situati vicini sullo stesso cromosoma vengono trasmessi sempre insieme alla discendenza, mentre l'LD si riferisce alla correlazione tra alleli all'interno di una popolazione ed è il risultato di fenomeni come la selezione, la mutazione, il sistema di incrocio, la struttura di popolazione, ecc. (Soto-Cerda & Cloutier, 2012). Sebbene una stretta associazione tra alleli sullo stesso cromosoma spesso si traduce in elevato LD, ci possono essere significativi livelli di LD anche tra loci distanti sullo stesso cromosoma o tra loci localizzati su cromosomi diversi.

Sia il mappaggio con popolazioni bi-parentali che il mappaggio per associazione sono basati sull'LD tra marcatori molecolari e loci funzionali. Nel mappaggio con popolazioni bi-parentali l'LD è generato dal sistema di incrocio ed è influenzato solo dalla ricombinazione, mentre nel mappaggio per associazione l'LD è specifico per la collezione di germoplasma in esame e dipende dalla ricombinazione ma anche da fattori evolutivi, quali quelli definiti sopra (Stich & Melchinger, 2010).

Il concetto di LD è stato descritto per la prima volta da Jennings nel 1917 e nel tempo sono stati proposti diversi approcci statistici per la sua misurazione. Tuttavia, i coefficienti

maggiormente utilizzati per calcolare il valore di LD per coppie di loci sono due:  $D'$  (Lewontin, 1964) e  $r^2$  (Hill & Robertson, 1968).

Si considerino, ad esempio, due loci (locus 1 e locus 2), ciascuno avente due alleli: gli alleli A e a al locus 1, con frequenze pari a  $P_A$  e  $P_a$  rispettivamente, e gli alleli B e b al locus 2 con frequenze pari a  $P_B$  e  $P_b$  rispettivamente. Le frequenze aplotipiche saranno le seguenti:  $P_{AB}$ ,  $P_{Ab}$ ,  $P_{aB}$ ,  $P_{ab}$ . La componente di base di entrambi i coefficienti  $D'$  e  $r^2$  è la differenza (D) tra la frequenza osservata dell'aplotipo AB e il prodotto delle frequenze dei singoli alleli A e B:

$$D = P_{AB} - P_A P_B$$

Il coefficiente  $D'$  ha la seguente formula:

$$D' = D / D_{\max}$$

dove  $D_{\max}$  è il più alto valore possibile per una data serie di frequenze alleliche ai due loci considerati. In particolare,  $D_{\max}$  equivale al valore più basso tra  $P_A P_b$  e  $P_a P_B$  se D è positivo e al valore più basso tra  $P_A P_B$  e  $P_a P_b$  se D è negativo.

Il coefficiente  $r^2$ , invece, introdotto da Hill e Robertson nel 1968, viene calcolato nel seguente modo:

$$r^2 = D^2 / P_A P_a P_B P_b$$

Entrambi i parametri possono assumere valori compresi tra 0 (completo linkage equilibrium) e 1 (completo linkage disequilibrium). La scelta del parametro più appropriato per la quantificazione dell'LD dipende dal tipo di studio che si intende effettuare. In alcuni studi sull'LD in popolazioni animali è stato utilizzato  $D'$  per misurare l'LD utilizzando marcatori microsatelliti (Du et al., 2007). Tuttavia, il valore di  $D'$  aumenta in caso di piccole dimensioni della popolazione e di basse frequenze alleliche, pertanto valori intermedi di  $D'$  non sono affidabili per effettuare analisi comparative di differenti studi e andrebbero confrontati con il parametro  $r^2$  prima di essere utilizzati per la quantificazione dell'LD (Oraguzie et al., 2007). Il coefficiente  $D'$ , inoltre, è influenzato soltanto dalla ricombinazione, mentre  $r^2$  dipende sia dalla ricombinazione, sia dalle differenze nelle frequenze alleliche ai due loci e quindi viene considerato migliore rispetto a  $D'$  per valutare l'associazione tra marcatori molecolari e caratteri agronomici. Attualmente, la maggior parte degli studi sull'LD nelle piante utilizzano il coefficiente  $r^2$  per la quantificazione del LD (Flint-Garcia et al., 2003; Gupta et al., 2005). Solitamente, valori di  $r^2$  pari a 0.1 o 0.2 vengono considerati come soglia minima per stabilire che due loci sono in LD e per definire il valore minimo di LD per individuare associazioni tra marcatori e caratteri di interesse (Zhu et al., 2008).

Per verificare se i valori di LD trovati in un determinato studio sono statisticamente significativi solitamente viene utilizzato il test esatto di Fisher, inserendo i dati in una tabella di contingenza 2 x 2. Il valore soglia (P-value) di 0.05 è generalmente usato per indicare mancanza di indipendenza degli alleli ai due loci, cioè i due loci sono associati (Gupta et al., 2005).

È molto importante analizzare l'andamento dell'LD all'interno di una popolazione prima di effettuare uno studio di mappaggio per associazione, in quanto la risoluzione di mappa e il numero di marcatori molecolari richiesti per analizzare l'intero genoma dipendono dal tasso di decadimento dell'LD lungo la distanza fisica del genoma. Se l'LD è basso, quindi decade rapidamente, è possibile ottenere un mappaggio con alta risoluzione, benché siano richiesti molti marcatori per individuare l'associazione. Viceversa, quando l'LD è alto la risoluzione di mappa risulta più bassa, ma il numero di marcatori necessari per analizzare l'intero genoma è inferiore.

L'entità dell'LD è estremamente variabile tra diverse specie, ma anche tra individui della stessa specie e all'interno del genoma di un singolo individuo. I fattori che influenzano l'LD sono:

- La ricombinazione. La frequenza di ricombinazione non è omogenea lungo il genoma, quindi, poiché la ricombinazione rompe l'LD, in alcuni punti del genoma saranno attesi grandi blocchi di LD, mentre in altri ci saranno blocchi più piccoli. Nelle regioni centromeriche ad esempio, si osservano livelli più elevati di LD rispetto ad altre regioni cromosomiche, in quanto in queste aree gli eventi di ricombinazione sono meno frequenti (Soto-Cerda & Cloutier, 2012).
- Modalità di riproduzione. La modalità di riproduzione ha profondi effetti sull'LD (Myles et al., 2009). Nelle specie autogame la frequenza degli eventi di ricombinazione è minore, in quanto gli individui hanno una percentuale maggiore di loci omozigoti rispetto agli individui delle specie allogame (Flint-Garcia et al., 2003). Nelle specie allogame, come mais, vite e segale (Tenailon et al., 2001; Myles et al., 2011; Li Y. et al., 2011), l'LD decade più rapidamente rispetto alle specie che utilizzano l'autofecondazione come modalità di riproduzione, quali il riso, Arabidopsis, il frumento e il pesco (Garris et al., 2005; Nordborg, 2000; Zhang et al., 2010; Aranzana et al., 2010; Cao et al., 2012; Li et al., 2013). Di conseguenza, nelle specie autogame i polimorfismi genetici tendono a rimanere collegati e sono attesi blocchi di LD che si estendono lungo grandi distanze nel genoma (Gaut & Long, 2003). Nelle specie

allogame, poiché l'LD tende ad essere basso è attesa una maggiore risoluzione di mappa e quindi un mappaggio più fine dei caratteri di interesse.

- Scelta del germoplasma. La popolazione naturale utilizzata gioca un ruolo fondamentale nella variazione dell'LD, in quanto l'estensione dell'LD dipende dal livello di variabilità genetica presente all'interno della popolazione stessa. In generale, maggiore è la variabilità genetica all'interno della popolazione, più rapidamente decade l'LD, in conseguenza dell'elevato numero di eventi di ricombinazione che si sono verificati durante la storia evolutiva della popolazione (Soto-Cerda & Cloutier, 2012). Gli effetti della scelta del germoplasma sull'LD sono evidenti in mais, in quanto è stato osservato che l'LD decade entro 1 Kb nelle *landraces*, nelle linee inbred decade nel raggio di circa 2 kb (il doppio), mentre nelle accessioni commerciali, frutto del miglioramento genetico, si estende anche per parecchie centinaia di Kb (Jung et al., 2004). In pesco, Li et al., (2013) hanno effettuato uno studio della variabilità genetica e dell'LD utilizzando una popolazione costituita da accessioni cinesi, americane ed europee. La popolazione totale è risultata suddivisa in tre sottopopolazioni (accessioni orientali, accessioni occidentali e *landraces*) ed è stato osservato che l'LD decade più velocemente (di circa il 43%) nelle accessioni orientali rispetto alle occidentali.
- La selezione. Sia la selezione naturale che la selezione artificiale hanno effetto sull'LD. Per quanto riguarda la selezione naturale, se gli alleli a due loci sono in LD ed entrambi i loci influenzano positivamente la *fitness* riproduttiva, la risposta alla selezione di un locus potrebbe essere accelerata dall'influenza positiva che ha l'altro locus sulla *fitness* riproduttiva. Pertanto, la selezione positiva aumenta l'LD tra un locus favorevole e un locus situato nelle vicinanze, anche se quest'ultimo è selettivamente neutrale; questo fenomeno è noto come "genetic hitchhiking" (autostop genetico; Slatkin, 2008).

Anche la selezione artificiale influenza notevolmente l'LD: grossi blocchi di LD sono osservati soprattutto nelle regioni genomiche che comprendono geni legati a caratteri agronomici. I processi di domesticazione delle specie coltivate, hanno modificato la struttura genomica delle colture, determinando una riduzione della variabilità genetica e creando una struttura di popolazione, che sono i maggiori fattori che influenzano il mappaggio per associazione.

- La deriva genetica. In popolazioni di piccole dimensioni la deriva genetica porta ad una consistente perdita di alleli rari, con conseguente aumento dell'LD.

Un aumento dell'LD si può osservare anche in popolazioni che hanno subito una riduzione delle loro dimensioni (ad esempio a causa della domesticazione), con conseguente deriva genetica. Questo fenomeno viene chiamato “collo di bottiglia” e porta alla perdita di alcuni aplotipi, a cui segue un aumento dell'LD (Flint-Garcia et al., 2003).

- La struttura di popolazione. Per struttura di popolazione si intende la suddivisione della popolazione iniziale in sottogruppi distinti in cui le frequenze alleliche sono significativamente differenti tra i sottogruppi stessi. La struttura di popolazione può essere determinata da deriva genetica, domesticazione o unioni non casuali tra individui a causa della loro separazione fisica. Il successo di un esperimento di mappaggio per associazione può essere fortemente ridotto a causa della struttura di popolazione, in quanto quest'ultima è uno dei principali fattori che può determinare LD tra loci non associati (Balding, 2006). Un elevato LD tra loci non realmente associati porta ad individuare associazioni falso-positive tra marcatori e caratteri di interesse. Si supponga, ad esempio, che in una sottopopolazione siano fissati gli alleli A e B a due diversi loci non associati, mentre in un'altra sottopopolazione siano fissati gli alleli a e b per gli stessi loci. Gli individui della prima sottopopolazione presenteranno solo l'aplotipo AB e quelli della seconda solo l'aplotipo ab. Se si prendono un po' di individui dal primo sottogruppo e un po' di individui dal secondo per creare una nuova popolazione, si osserverà che i due loci sono in perfetto LD, quando in realtà non lo sono (Slatkin, 2008).

Thornsberry et al., nel 2001 hanno condotto uno studio in cui hanno individuato associazioni significative tra marcatori molecolari situati all'interno del gene *Dwarf8* di mais e l'epoca di fioritura, ma loro stessi hanno affermato che circa l'80% di queste associazioni erano spurie, in quanto causate dalla struttura di popolazione. Il verificarsi di associazioni spurie è più frequente nei caratteri legati all'adattamento ambientale, in quanto questi mostrano una correlazione positiva con le variabili ambientali sotto cui si sono evoluti e di conseguenza, le regioni genomiche che controllano questi caratteri sono fortemente correlate alla struttura di popolazione.

Sono stati sviluppati diversi modelli statistici per analizzare la struttura di popolazione: gli algoritmi di Pritchard e Rosenberg (1999) implementati nel software STRUCTURE (Pritchard et al., 2000a), l'analisi delle componenti principali (PCA) e l'analisi delle coordinate principali (PCoA).

#### 1.4.2 Metodi statistici per il mappaggio per associazione

I metodi statistici di base per le analisi di associazione, in una situazione ideale, sarebbero la regressione lineare, l'analisi della varianza (ANOVA), il t test o il test del chi-quadro (Zhu et al., 2008). Tuttavia, poiché la struttura di popolazione può portare ad associazioni spurie tra genotipo e fenotipo, sono stati sviluppati quattro diversi approcci statistici per controllare gli effetti della suddivisione della popolazione in sottogruppi sui test di associazione: 1) *genomic control* (GC) (Devlin & Roeder, 1999), 2) *structured association* (SA) (Falush et al., 2003; Pritchard & Rosenberg, 1999; Pritchard et al., 2000b), 3) *mixed linear model* (Yu et al., 2006), 4) approccio delle componenti principali (Price et al., 2006). Il metodo GC assume che la struttura di popolazione abbia lo stesso effetto su tutti i loci genotipizzati. In questo metodo, un piccolo set di marcatori (ad esempio quelli meno probabilmente associati ai caratteri di interesse) viene utilizzato per stimare l'influenza della struttura di popolazione sul test statistico di associazione. Il livello di significatività del test di associazione (P-value), in questo modo, viene corretto sulla base della struttura di popolazione. Nell'analisi SA, al contrario, prima viene utilizzato un insieme di marcatori per valutare la struttura di popolazione e definire una matrice Q, in cui per ogni individuo viene indicata la probabilità di appartenenza a ciascuna sottopopolazione, e successivamente tale matrice viene incorporata nell'analisi statistica di associazione. Il metodo SA è implementato nel General Linear Model (GLM) del software TASSEL (Bradbury et al., 2007).

Nel 2006, Yu et al., hanno sviluppato un nuovo modello statistico, il Mixed Linear Model (MLM), in cui un insieme di marcatori molecolari viene usato per valutare sia la struttura di popolazione che le relazioni di parentela tra gli individui della popolazione (Kinship). Per eseguire il MLM è necessario effettuare i seguenti passaggi: 1) analizzare la struttura di popolazione e definire una matrice Q, utilizzando ad esempio il programma STRUCTURE. 2) Analizzare le relazioni di parentela tra gli individui e definire una matrice K. La matrice K è una matrice  $n \times n$  ( $n$  = numero di individui della popolazione) che riporta i coefficienti di parentela tra coppie di individui. Tali coefficienti vengono calcolati in base al livello di similarità tra i genotipi, confrontando lo stato allelico dei marcatori utilizzati (Yu et al., 2006). Un programma che fa questo tipo di analisi è ad esempio SPAGeDI (Hardy & Vekemans, 2002). 3) Utilizzare le matrici Q e K nel test statistico di associazione, allo scopo di ridurre le associazioni spurie tra genotipo e fenotipo. Al fine di chiarire l'effetto

combinato che la struttura di popolazione e le relazioni di parentela hanno sul mappaggio per associazione, questi fattori vengono usati come variabili indipendenti nel modello di regressione che definisce la correlazione tra genotipo e fenotipo (Ersoz et al., 2007). Anche il MLM, come il GLM, è implementato all'interno del software TASSEL. Altri programmi sono disponibili per effettuare analisi di associazione, oltre TASSEL, come ad esempio GAPIT (Lipka et al., 2012).

Un altro tipo di MLM usato nel mappaggio per associazione, prevede l'impiego dell'analisi delle componenti principali (PCA), al posto della matrice Q calcolata con STRUCTURE (Price et al., 2006). La PCA permette di individuare gli assi principali (le componenti principali) che spiegano la variabilità osservata con i marcatori molecolari all'interno della popolazione. Nella matrice che si ottiene, le colonne rappresentano gli assi principali, mentre i valori riportati per ciascun individuo indicano la posizione di quell'individuo lungo gli assi. Questo metodo è stato introdotto per ovviare ad alcuni inconvenienti che presenta il programma STRUCTURE: innanzitutto il calcolo della matrice Q attraverso questo software è piuttosto laborioso dal punto di vista computazionale (Balding, 2006; Price et al., 2006), in secondo luogo STRUCTURE è stato realizzato per popolazioni che contengono individui non imparentati e che sono in equilibrio di Hardy-Weinberg (Pritchard et al., 2000b), ma la maggior parte delle collezioni di germoplasma delle specie vegetali non si trovano in questa condizione e di conseguenza i risultati di STRUCTURE devono essere interpretati scrupolosamente (Camus-Kulandaivelu et al., 2007). Il metodo MLM con la PCA è stato utilizzato in diverse specie (Zhao et al., 2007; Weber et al., 2008; Stich & Melchinger, 2009) e ha dato buoni risultati.

### **1.4.3 Studi di mappaggio per associazione nelle specie vegetali**

I primi studi di mappaggio per associazione nelle piante furono effettuati in mais (*Zea mays*; Bar-Hen et al., 1995), in riso (*Oryza sativa*; Virk et al., 1996) e in avena (*Avena sativa*; Beer et al., 1997). Bar-Hen et al. (1995) e Virk et al. (1996) hanno studiato l'associazione di caratteri quantitativi con marcatori RAPD e isoenzimi, rispettivamente, mentre Beer et al. (1997) hanno individuato 13 QTL associati a marcatori RFLP usando 64 varietà di avena. In questi studi furono testati un ridotto numero di marcatori distribuiti in tutto il genoma e non venne presa in considerazione la struttura di popolazione.

Il primo esperimento di mappaggio per associazione, mediante l'approccio per geni candidati, che ha tenuto conto della struttura di popolazione è stato effettuato in mais, in cui si è cercata associazione tra marcatori localizzati nel gene *D8* e l'epoca di fioritura (Pritchard, 2001).

In *Arabidopsis* (*Arabidopsis thaliana*) i primi studi di mappaggio per associazione sono stati effettuati mediante l'approccio per geni candidati. Il primo esperimento di mappaggio per associazione mediante l'impiego di marcatori localizzati in tutto il genoma è stato effettuato da Aranzana et al., (2005), i quali hanno individuato geni legati all'epoca di fioritura e tre geni di resistenza, già identificati in altri studi.

Negli ultimi dieci anni, gli studi di mappaggio per associazione sono aumentati notevolmente, grazie soprattutto all'introduzione delle nuove tecnologie di sequenziamento che hanno consentito di individuare in modo rapido, efficiente e non troppo costoso un elevato numero di marcatori molecolari, soprattutto SNP (Soto-Cerda & Cloutier, 2012).

In *Arabidopsis*, sono stati condotti nuovi studi per cercare di individuare marcatori associati all'epoca di fioritura utilizzando il mappaggio per associazione mediante l'analisi dell'intero genoma con 231.497 marcatori SNP (Li et al., 2010). In mais, recentemente è stato effettuato uno studio di mappaggio per associazione mediante analisi dell'intero genoma, in cui sono stati analizzati 41 differenti caratteri fenotipici in circa 5000 linee inbred al fine di individuare associazioni con marcatori SNP e *copy-number variants* (CNVs). Le accessioni sono state genotipizzate con 28.9 milioni di SNP e circa 800 mila CNVs. (Wallace et al., 2014). In riso, Huang et al., (2010) hanno analizzato la dissezione genetica di 14 caratteri agronomici in 517 *landraces*, utilizzando circa 3.6 milioni di marcatori SNP, mentre recentemente Begum et al., (2015) hanno utilizzato una popolazione di accessioni tropicali di riso (*Oryza sativa*) coltivato per studiare l'associazione di 19 caratteri agronomici con marcatori SNP. La popolazione è stata genotipizzata con 71.710 SNP e sono stati identificati 52 QTL che controllano 11 caratteri. In orzo (*Hordeum vulgare*), Cockram et al., (2010) hanno utilizzato una popolazione di 500 individui genotipizzata con 1536 marcatori SNP con l'obiettivo di individuare polimorfismi all'interno del gene *ANT-2*, il quale regola la produzione di antocianine in orzo. In erba medica (*Medicago sativa*), che è una specie tetraploide, 15 marcatori SSR hanno mostrato una forte associazione con la produttività in esperimenti condotti in diversi ambienti (Li X. et al., 2011). Nella barbabietola da zucchero (*Beta vulgaris*), è stata studiata la dissezione genetica di sei caratteri agronomici mediante il mappaggio per

associazione e sono stati individuati QTL ad effetto principale sui caratteri analizzati e altri con effetti epistatici (Würschum et al., 2011).

In pesco, sono stati effettuati diversi studi per valutare la variabilità, la struttura di popolazione e l'estensione dell'LD (Aranzana et al., 2010; Cao et al., 2012; Li et al., 2013; Micheletti et al., sottomesso), parametri fondamentali da analizzare prima di intraprendere uno studio di mappaggio per associazione. Aranzana et al. (2010) hanno analizzato 224 cultivar commerciali di pesco (ottenute attraverso i moderni programmi di miglioramento genetico) di origine spagnola e americana, con 50 marcatori SSR distribuiti in tutto il genoma e hanno osservato che, in seguito all'analisi di struttura, la loro popolazione si divideva in tre sottogruppi: pesche a polpa fondente, nettarine a polpa fondente e pesche e nettarine a polpa soda. In tutte e tre le sottopopolazioni i livelli di LD erano alti e decadevano con la distanza, entro 13-15 cM. Cao et al. (2012) hanno analizzato 104 *landraces* di pesco con 53 marcatori SSR, provenienti da sei regioni della Cina, e hanno osservato che in questa popolazione l'LD decadeva entro 6 cM. Hanno anche effettuato uno studio di mappaggio per associazione, utilizzando il GLM implementato in TASSEL, e hanno individuato alcuni di questi marcatori SSR associati a caratteri relativi alla qualità del frutto. Li et al. (2013) hanno esaminato 653 cultivar di pesco di origine cinese, nord-americana ed europea utilizzando 48 marcatori SSR e hanno osservato che le accessioni orientali presentavano una maggiore variabilità rispetto alle occidentali. La popolazione totale è risultata suddivisa in tre sottopopolazioni (accessioni orientali, accessioni occidentali e *landraces*) ed è stato osservato che l'LD decadeva più velocemente (di circa il 43%) nelle accessioni orientali rispetto alle occidentali, mentre nelle *landraces* non vi era variazione dell'LD rispetto alla distanza fisica sul genoma. Font I Forcada et al. (2013) hanno effettuato uno studio di mappaggio per associazione in una popolazione di 94 accessioni di pesco di origine occidentale, utilizzando il GLM al fine di individuare marcatori SSR associati a caratteri agronomici rilevanti. Infine, Micheletti et al. (sottomesso) hanno valutato 1580 accessioni di pesco provenienti da 5 diverse collezioni di germoplasma (quattro localizzate in Europa e una in Cina) utilizzando un chip contenente 8144 SNP, l'IPSC 9k SNP array (Verde et al., 2012). In questo caso, la popolazione è risultata suddivisa in tre gruppi (accessioni occidentali moderne, ottenute attraverso il miglioramento genetico, accessioni occidentali antiche e accessioni orientali), più un elevato numero di accessioni che non sono state assegnate ad alcuna popolazione e sono state quindi considerate admixed. In questo studio è stato osservato che l'LD decadeva con la distanza in tutte le sottopopolazioni e tale decadimento era più elevato

negli individui *admixed*. In questo caso non è stato osservato un più elevato decadimento dell'LD nelle accessioni orientali rispetto alle occidentali, come avevano evidenziato Li et al. (2013), probabilmente perché le accessioni orientali in questo studio erano per la maggior parte accessioni ottenute con il miglioramento genetico, non erano incluse specie antiche o *landraces* (Micheletti et al., sottomesso). Micheletti et al., hanno inoltre effettuato un'analisi di mappaggio per associazione (mediante il MLM) e hanno individuato marcatori SNP significativamente associati a sette caratteri qualitativi.

In conclusione, da tutti questi lavori è emerso che le accessioni di pesco coltivato presentano una ridotta variabilità genetica e una elevata conservazione dell'LD, pertanto per effettuare uno studio di mappaggio per associazione che includa queste varietà non è necessario utilizzare un numero molto elevato di marcatori distribuiti lungo tutto il genoma.

## 1.5 IL SEQUENZIAMENTO DEL GENOMA DEL PESCO

Il pesco [*Prunus persica* (L.) Batsch] è sempre stata la specie più studiata dal punto di vista genetico all'interno della famiglia delle Rosaceae. Il pesco è una specie economicamente molto importante (è il terzo frutto maggiormente coltivato nelle zone temperate dopo melo e pero) ed è considerata una specie modello per la famiglia delle Rosaceae, poiché presenta diversi vantaggi dal punto di vista genetico e riproduttivo (Arús et al., 2012). Contrariamente ad altre specie del genere *Prunus*, quali susino coltivato e ciliegio acido che sono poliploidi, è una specie diploide ( $2n=16$ ) e ha un genoma di piccole dimensioni (227 Mbp), circa il doppio di quello di *Arabidopsis thaliana*. Il pesco, inoltre, è una specie autogama e ha un periodo improduttivo limitato (2-3 anni) se paragonato a quello di altre specie da frutto che presentano una fase giovanile di 5-10 anni (Sosinski et al., 2009).

Negli ultimi venti anni, anche grazie allo sviluppo dei marcatori molecolari, in pesco sono stati effettuati numerosi lavori per studiare la dissezione genetica di importanti caratteri agronomici, sono state sviluppate mappe genetiche, mappe fisiche e collezioni di EST (Arús et al., 2012). Per tutte queste ragioni il pesco è stato scelto come specie su cui investire per un progetto di sequenziamento. Nel 2009 è stato istituito un consorzio internazionale, IPGI (International Peach Genome Initiative), che comprendeva laboratori e ricercatori di Stati Uniti, Italia, Francia, Spagna, Cile e che ha guidato il sequenziamento del genoma del pesco.

Il genotipo individuato per il sequenziamento è il doppio aploide della cultivar Lovell. La scelta di un diaploide è motivata dal fatto che un individuo omozigote a tutti i loci rende più semplice la fase di assemblaggio della sequenza. È stato utilizzato un approccio Whole Genome Shotgun e le reads (i frammenti di sequenza) sono state ottenute mediante il metodo Sanger. In totale sono state ottenute 3 729 679 reads, con una copertura finale della sequenza di 8.5X. Utilizzando il programma Arachne, queste reads sono state assemblate in 391 scaffolds principali (>1 kb), per un totale di 226.6 Mb. Questi scaffold sono poi stati ancorati e orientati alla mappa di riferimento del genere *Prunus* T x E utilizzando 827 marcatori molecolari. Il 1 aprile 2010 è stata rilasciata la sequenza completa del genoma del pesco (Peach v1.0), organizzata in 8 pseudomolecole (che corrispondono agli 8 cromosomi del pesco) che comprendono 224.6 Mb.

L'assemblaggio del genoma del pesco è stato confrontato con quello di altre specie sequenziate ed è stato osservato che, sulla base dei parametri definiti da Chain et al., 2009 (escludendo i genomi di *Arabidopsis* e riso che sono le uniche specie in cui la sequenza del genoma è perfettamente completa), la sequenza del genoma del pesco è una delle migliori tra quelle di altre specie vegetali sequenziate e certamente la migliore tra le specie arboree (Verde et al., 2013).

In seguito all'annotazione genica sono stati individuati 27 852 geni che codificano per proteine, 28 689 trascritti che codificano per proteine, più altri RNA non codificanti. È stato osservato che il contenuto genico in pesco è notevolmente più basso rispetto a quello del melo (57 386 geni) e del pioppo (45 654 geni) ma è simile a quello di vite (30 434 geni) e di *Arabidopsis* (27 416 geni) (Verde et al., 2013). Inoltre, la densità genica in pesco (1.22 geni ogni 10 kb in media) è risultata più elevata rispetto al melo (0.78 geni) ma più bassa rispetto ad *Arabidopsis* (2.29 geni; Verde et al., 2013).

Al fine di analizzare la diversità nucleotidica e il percorso di domesticazione del pesco, sono state risequenziate 11 accessioni di *Prunus persica* con diversa origine geografica e genetica (includendo anche l'accessione diaploide Lovell P<sub>Lov2-2N</sub> come controllo) e una accessione di *Prunus ferganensis*, una di *Prunus kansuensis*, una di *Prunus davidiana* e una di *Prunus mira*. Dopo il risequenziamento, tutte le accessioni sono state confrontate con il genoma di riferimento ed è stato osservato, come atteso, che tre delle specie selvatiche (*P. kansuensis*, *P. davidiana* e *P. mira*) erano molto diverse dal genoma di riferimento e da tutte le altre accessioni, con quasi un milione di SNP individuati in ciascuna di esse. Al contrario *P. ferganensis* era indistinguibile dalle altre accessioni di *P. persica* in termini di frequenza e distribuzione degli SNP. Una possibile spiegazione è che

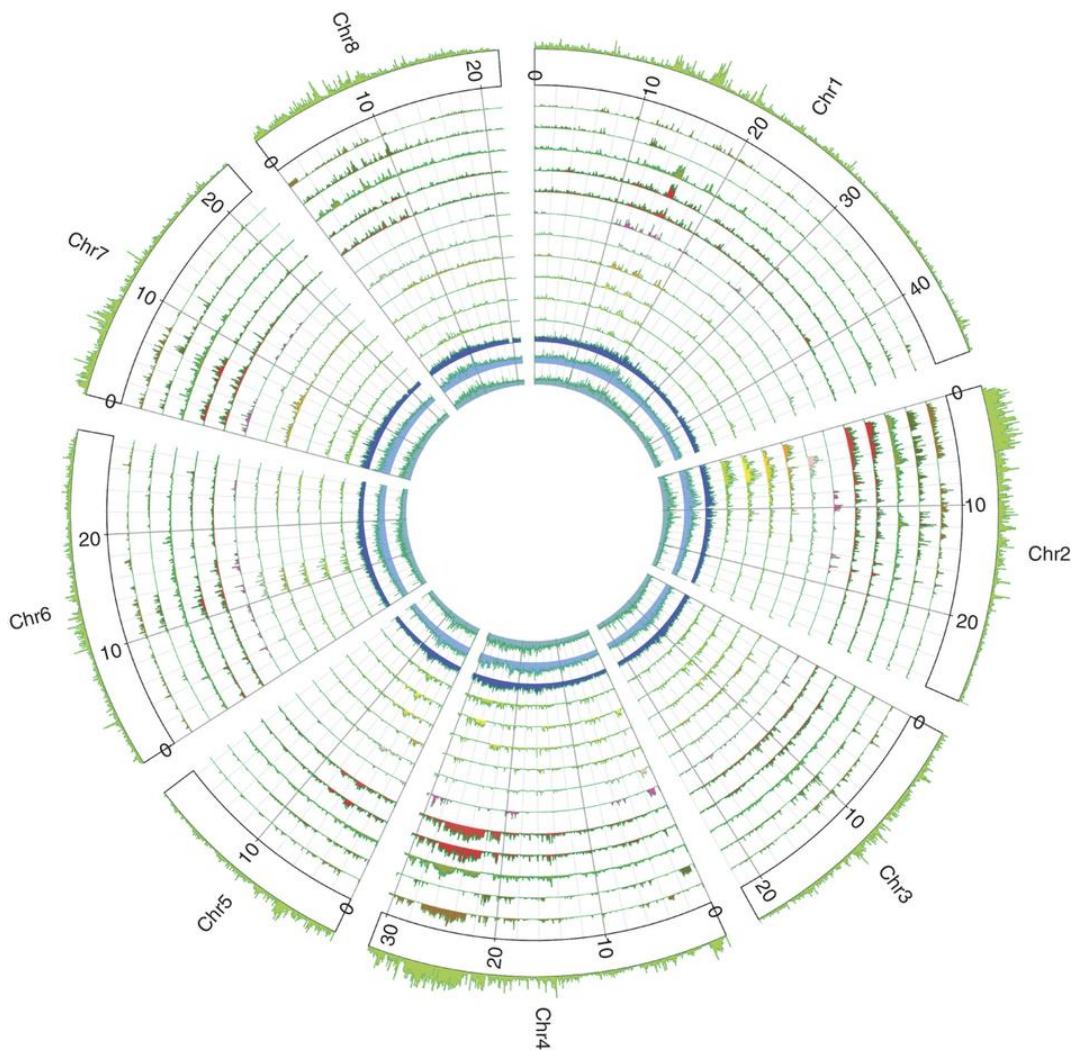
*P. ferganensis* e *P. persica* siano in realtà la stessa specie e che *P. ferganensis* sia una accessione selvatica di pesco (Verde et al., 2013).

In seguito a questo lavoro di risequenziamento è stato selezionato un set di 953 357 SNP distribuiti negli 8 cromosomi e informativi in almeno 4 accessioni, con cui è stata effettuata un'analisi di diversità nucleotidica lungo gli 8 pseudocromosomi. La diversità nucleotidica media a livello dell'intero genoma è di  $1.5 \times 10^{-3}$ , con notevoli differenze tra i cromosomi, variando da un minimo di  $1.1 \times 10^{-3}$  per la pseudomolecola 1 ad un massimo di  $2.2 \times 10^{-3}$  per la pseudomolecola 2 (Fig. 6). Inoltre, una maggiore variabilità degli SNP rispetto alla media è stata osservata all'inizio della pseudomolecola 2 e alla fine della pseudomolecola 4. L'inizio del cromosoma 2 comprende geni di resistenza a nematodi, infatti è stato osservato che questa regione presenta una densità di 5 volte maggiore di geni NB-LRR rispetto al resto del genoma. Le regioni che comprendono geni di resistenza evolvono molto rapidamente e questo potrebbe spiegare l'insolito elevato livello di variabilità della sequenza all'inizio del cromosoma 2 (Verde et al., 2013). La fine del cromosoma 4, invece, include geni coinvolti nell'epoca di maturazione del frutto. Poiché le pesche, una volta raccolte, possono essere conservate solo per poche settimane, è necessario avere a disposizione sul mercato diverse varietà di pesche con diversa epoca di maturazione durante tutta la stagione di raccolta di questo frutto (in Italia ad esempio va da maggio ad ottobre). I programmi di miglioramento genetico per questo carattere seguono diversi schemi e ciò è compatibile con il mantenimento di un elevato livello di variabilità nella regione che comprende geni relativi all'epoca di maturazione. Ad ulteriore conferma di questa ipotesi, non è stata osservata questa elevata variabilità di sequenza nella stessa regione nelle specie selvatiche *P.kansuensis*, *P. mira* e *P. davidiana*, che non sono state oggetto del miglioramento genetico (Fig. 6; Verde et al., 2013).

Un altro obiettivo del lavoro di risequenziamento delle 15 accessioni di *Prunus* era quello di valutare gli effetti della domesticazione del pesco. Si suppone che durante la storia di domesticazione del pesco si siano verificati tre colli di bottiglia che hanno portato a una riduzione della variabilità genetica: 1) il primo processo di domesticazione, che si ritiene abbia avuto luogo in Cina circa 4000-5000 anni fa, seguito dall'inizio della moltiplicazione vegetativa; 2) la diffusione del pesco dalla Cina in Europa, attraverso la Persia; 3) la più recente introduzione del pesco negli Stati Uniti (intorno al XVI-XIX secolo), in cui sono state selezionate alcune accessioni (fondatori) da cui derivano la maggior parte delle varietà coltivate occidentali. Per verificare il primo collo di bottiglia, è stata confrontata la diversità nucleotidica dell'accessione di *P. davidiana* (una specie selvatica allogama) con

quella delle accessioni asiatiche di pesco coltivato. La diversità nucleotidica è stata stimata di  $4.8 \times 10^{-3}$  per *P. davidiana* e di  $1.6 \times 10^{-3}$  per le accessioni asiatiche e questo evidenzia la forte riduzione di variabilità che si è avuta in seguito ai primi processi di domesticazione in Cina.

Analizzando le differenze nei valori di diversità nucleotidica tra le accessioni orientali e occidentali ( $1.6 \times 10^{-3}$  e  $1.1 \times 10^{-3}$  rispettivamente), è stato confermato un secondo collo di bottiglia, legato alla diffusione del pesco dalla Cina in Europa e all'attività di miglioramento genetico, che ha portato ad una notevole perdita di variabilità nelle accessioni orientali rispetto a quelle occidentali. Infine, l'ultimo collo di bottiglia è dimostrato dalla perdita di alcuni alleli rari nelle accessioni moderne e ad un decadimento più lento del LD rispetto a quello di altre specie (Verde et al., 2013), quali ad esempio soia, erba medica e *Arabidopsis* (Lam et al., 2010; Branca et al., 2011; Cao et al., 2011).



**Fig. 6** Distribuzione della diversità nucleotidica negli 8 cromosomi di pesco. La linea più esterna rappresenta la diversità nucleotidica stimata da un campione di 23 genotipi. Le linee più interne blu rappresentano la diversità nucleotidica analizzata nelle specie selvatiche (dall'interno verso l'esterno sono: *P. mira*, *P. davidiana* e *P. kansuensis*), mentre le altre linee mostrano l'analisi della variabilità nelle altre accessioni di *P. persica* risequenziate. Da Verde et al., 2013.

Il completamento del sequenziamento del genoma del pesco è stato un punto fondamentale per la comprensione dell'evoluzione di questa specie. Grazie alla disponibilità della sequenza completa di pesco, inoltre, si sono moltiplicati gli studi per cercare di individuare geni che controllano specifici caratteri di interesse, quali la resistenza alle malattie, il colore della polpa, il fabbisogno in freddo (Verde et al., 2013).

Il lavoro di sequenziamento del genoma del pesco ha prodotto una sequenza completa ed accurata che può essere utilizzata come riferimento per assemblare il genoma di molte altre specie all'interno del genere *Prunus*, come albicocco, susino e ciliegio dolce (Verde et al., 2013). Infine, la sequenza di pesco è stata un punto di partenza per costruire strumenti di ricerca molto utili, ossia i chip Illumina contenenti migliaia di marcatori SNP che sono stati sviluppati in pesco (Verde et al., 2012) e in ciliegio (Peace et al., 2012).

## **1.6 IL CHIP IPSC 9K SNP**

Fino a pochi anni fa, lo studio della dissezione genetica di caratteri agronomici rilevanti nelle piante è stato effettuato prevalentemente utilizzando popolazioni bi-parentali e un ridotto numero di marcatori molecolari. Grazie all'introduzione delle tecnologie di sequenziamento di nuova generazione (Next Generation Sequencing, NGS), oggi è possibile sequenziare in tempi rapidi diverse accessioni di una specie e confrontarle con il genoma di riferimento. Ciò consente di individuare un elevato numero di marcatori SNP e di inserzioni/delezioni.

Un'altra importante innovazione nel campo della ricerca genetica è stata lo sviluppo di piattaforme che permettono di genotipizzare in un'unica analisi milioni di marcatori SNP per singolo campione di DNA; ad esempio la piattaforma Infinium BeadArray della ditta Illumina è un sistema di genotipizzazione che consente di analizzare fino a 2.5 milioni di SNP per singolo campione di DNA (Gunderson, 2009).

In pesco, essendo disponibile un genoma di riferimento completamente sequenziato, è stato istituito un consorzio internazionale (The International Peach SNP Consortium, IPSC) che ha messo a punto un array Illumina contenente 9K marcatori SNP (Verde et al., 2012).

Inizialmente sono state risequenziate 56 accessioni di pesco coltivato utilizzando le piattaforme Genome Analyzer (Illumina) e Roche 454. Queste accessioni erano fondatori, progenitori intermedi o individui utilizzati come parentali nei programmi internazionali di miglioramento genetico e sono stati scelti in modo da avere una buona copertura della

variabilità genetica del pesco coltivato. In totale, tramite algoritmi specifici di rilevazione di polimorfismi, sono stati individuati 1 022 354 marcatori SNP. La verifica di questi polimorfismi è stata effettuata su un sottoinsieme di marcatori tramite la piattaforma Illumina GodenGate ed è stato osservato che gli SNP esonici ed intronici erano i migliori, con circa il 75% dei polimorfismi verificati, mentre solo un terzo degli SNP intergenici sono stati verificati.

Attraverso una serie di operazioni di filtraggio, sono stati selezionati 8144 SNP che sono stati posizionati sul chip Illumina (IPSC peach 9K SNP array). Di questi marcatori, 8125 (99.8%) sono localizzati negli 8 cromosomi e 19 (0.2%) nella frazione non mappata del genoma. Gli SNP localizzati sul chip sono tutti SNP esonici e sono ben spaziati nel genoma, con una distanza media tra di essi di 26.7 kb.

Per valutare l'efficienza del chip, quest'ultimo è stato impiegato per genotipizzare 709 accessioni, divise in due differenti gruppi: alcune cultivar provenivano dall' Europa, mentre altre dagli Stati Uniti. In totale, 6869 marcatori SNP sono risultati polimorfici in tutte le 709 accessioni analizzate, cioè l'84.3% degli SNP presenti sul chip. I 6869 marcatori polimorfici sono distribuiti nel genoma con una distanza media tra di essi di 31.5 kb. Questi quasi 7000 SNP localizzati sul chip e polimorfici in un elevato numero di accessioni rappresentano una importante fonte di marcatori che possono essere utilizzati nelle analisi di mappaggio genetico con popolazioni bi-parentali, negli studi di variabilità genetica e negli esperimenti di mappaggio per associazione (Eduardo et al., 2012; Yang et al., 2013; Pirona et al., 2013; Romeu et al., 2014; Frett et al., 2014; Pacheco et al., 2014; da Silva et al., 2015; Micheletti et al., sottomesso).

## **1.7 SCOPO DEL LAVORO**

Il pesco è tra le specie frutticole più importanti per l'economia italiana ed è considerata una specie modello per la famiglia delle Rosaceae, in quanto presenta diversi vantaggi dal punto di vista genetico e riproduttivo. È infatti una specie diploide, autogama, con un genoma molto contenuto. Al contrario di altre specie frutticole ha un periodo improduttivo limitato (2-3 anni).

Presso il Centro di Ricerca per la Frutticoltura del Consiglio per la Ricerca e l'analisi dell'economia agraria (CRA) di Roma è disponibile una collezione di 877 accessioni di pesco che hanno diversa origine e potenzialmente contengono un'ampia diversità allelica che si traduce in una notevole variabilità fenotipica.

D'altro lato, nel 2010 è stata rilasciata la sequenza completa del genoma del pesco (Verde et al., 2013) e successivamente sono state risequenziate 56 accessioni di *Prunus persica*. Ciò ha permesso di individuare un milione di marcatori SNP (Single Nucleotide Polymorphism) sparsi nel genoma; di questi ne sono stati selezionati 8144 localizzati nelle regioni esoniche ed è stato realizzato un chip Illumina (IPSC 9K SNP array) (Verde et al., 2012).

In questo contesto, l'obiettivo del presente dottorato di ricerca è quello di analizzare la variabilità genetica presente nel germoplasma di pesco e di individuare marcatori SNP strettamente associati a caratteri agronomici rilevanti effettuando uno studio di mappaggio per associazione.

Il lavoro è suddiviso nelle seguenti fasi:

- Analisi genetica dell'intera collezione di germoplasma con 16 marcatori SSR;
- Selezione di un sottoinsieme di 339 accessioni;
- Genotipizzazione delle 339 accessioni con il chip IPSC 9K SNP array;
- Definizione di una core collection di 165 accessioni;
- Fenotipizzazione della core collection;
- Analisi dei dati.

## 2. MATERIALI E METODI

### 2.1 MATERIALE VEGETALE

Tutte le 877 accessioni di *Prunus* presenti presso il Centro di Ricerca per la Frutticoltura di Roma sono state innanzitutto analizzate con marcatori SSR, al fine di effettuare uno studio della variabilità genetica. Di queste accessioni, 876 erano di *Prunus persica* e una di *Prunus davidiana*. I genotipi di *P. persica* comprendevano accessioni orientali e occidentali, cultivar moderne e *landraces*.

Per ciascuna delle 877 accessioni è stato prelevato, in primavera, 1 g di tessuto fogliare (apici fogliari). Gli apici fogliari appena raccolti sono stati congelati in azoto liquido e sono stati conservati a -80 °C fino al momento dell'utilizzo.

Inizialmente, i campioni di tessuto sono stati polverizzati con mortaio e pestello in azoto liquido ed è stato estratto il DNA utilizzando il DNeasy Plant Mini Kit (Qiagen). I campioni sono poi stati quantificati impiegando lo spettrofotometro NanoDrop ND-1000 UV-Vis (Thermo Scientific) ed è stata ottenuta una concentrazione di DNA di circa 70 ng/μl per campione. Poiché tale resa era troppo bassa per consentire la successiva analisi degli acidi nucleici con il chip IPSC 9K SNP e anche al fine di ridurre la degradazione del campione e il tempo richiesto per la preparazione del lisato, si è deciso di liofilizzare gli apici fogliari prima di procedere all'estrazione del DNA. È stato utilizzato un liofilizzatore Edwards Freeze Dryer Modulyo, che opera ad una temperatura di -50 °C ed è dotato di una pompa che genera un vuoto pari a -1,5 mbar. Gli apici fogliari, conservati in tubi da 2 ml, sono stati prelevati dal congelatore -80 °C e sono stati rapidamente posti nel liofilizzatore, dove sono stati lasciati per 3 giorni. All'interno di ciascun campione liofilizzato sono state inserite due biglie di tungsteno e, tramite un vortex, i tessuti sono stati polverizzati. A questo punto si è proceduto all'estrazione del DNA impiegando ugualmente il DNeasy Plant Mini Kit (Qiagen). Tramite questo procedimento è stato possibile ottenere una resa media di 200 ng/μl.

### 2.2 QUANTIFICAZIONE DEL DNA GENOMICO

Tutti i campioni sono stati quantificati utilizzando lo spettrofotometro NanoDrop ND-1000 UV-Vis (Thermo Scientific).

Tra le 877 accessioni di *Prunus*, ne sono state selezionate 339 di *Prunus persica* che sono state genotipizzate con il chip IPSC 9K SNP. Poiché la reazione di ibridazione del chip è molto sensibile, in quanto si basa su polimorfismi di un singolo nucleotide, è stato necessario verificare in modo preciso la qualità e la concentrazione del DNA estratto da queste 339 accessioni. A tal fine, i campioni di DNA sono stati quantificati usando il fluorocromo Picogreen, che si lega selettivamente al DNA a doppia elica. È stato utilizzato il Quanti-iT PicoGreen® dsDNA Assay Kit (Invitrogen, USA) ed un lettore di fluorescenza Fluoroskan Ascent (Thermo Scientific, Finlandia).

Il kit include:

- soluzione stock di Picogreen 200X;
- soluzione TE 20X;
- soluzione stock di  $\lambda$ DNA, fornita ad una concentrazione di 100 ng/ $\mu$ l in TE.

#### PREPARAZIONE DEI REAGENTI:

##### *BUFFER TE1X:*

Prima di iniziare con la quantificazione è necessario preparare una soluzione di TE 1X, diluendo 20 volte lo stock 20X in acqua sterile e priva di DNasi. Il volume di TE da preparare dipende dal numero di campioni che devono essere quantificati.

##### *WORKING SOLUTION DI PICOGREEN:*

Diluire 200 volte la soluzione stock di Picogreen 200X in TE 1X, al fine di ottenere una soluzione di Picogreen 1X. Proteggere la soluzione così preparata dalla luce e utilizzarla entro poche ore.

##### *CURVA STANDARD DI CALIBRAZIONE*

La soluzione stock di  $\lambda$ DNA, fornita ad una concentrazione di 100 ng/ $\mu$ l, deve essere diluita 50 volte in TE 1X, al fine di ottenere una concentrazione finale di  $\lambda$ DNA di 2 ng/ $\mu$ l. Successivamente, diluire il  $\lambda$ DNA [2 ng/ $\mu$ l] in differenti tubi, al fine di ottenere diverse concentrazioni di  $\lambda$ DNA per la curva di calibrazione, come riportato nella tabella 3:

Volume (µl) di TE 1X	Volume (µl) di λDNA [2 ng/µl]	Volume (µl) di PicoGreen 1X	Concentrazione finale di λDNA
0	1000	1000	1 ng/µl
900	100	1000	0.1 ng/µl
990	10	1000	0.01 ng/µl
999	1	1000	0.001 ng/µl
1000	0	1000	0 ng/µl - Bianco

**Tab. 3** Protocollo per preparare i campioni per la curva di taratura.

Dopo aver aggiunto 1 ml di Picogreen 1X a ciascuna diluizione del λDNA, come riportato in tabella 3, miscelare bene i tubi e incubare per 2-5 minuti a temperatura ambiente.

#### PROCEDURA PER LA QUANTIFICAZIONE:

1. Prendere una piastra da 96 pozzetti e dispensare in ciascun pozzetto 400 µl di ogni diluizione del λDNA, compreso il bianco, facendo tre ripetizioni per ogni diluizione.
2. Diluire i campioni di DNA genomico da quantificare 1:1000 (inserire 1 µl di DNA in 999 µl di TE 1X).
3. Prelevare 200 µl di ciascun campione di DNA genomico diluito e dispensarlo nella piastra di lettura.
4. Aggiungere 200 µl di Picogreen 1X a ogni campione di DNA genomico nella piastra di lettura.
5. Misurare la fluorescenza dei campioni.

La fluorescenza è stata rilevata mediante il fluorimetro Fluoroskan Ascent (Thermo Scientific, Finlandia) utilizzando filtri di eccitazione a 485 nm e filtri di emissione 538 nm.

### **2.3 GENOTIPIZZAZIONE CON MARCATORI SSR**

Tutte le 877 accessioni di *Prunus* presenti presso il Centro di Ricerca per la Frutticoltura di Roma sono state analizzate con 16 marcatori SSR. Le reazioni di amplificazione sono state effettuate dal Parco Tecnologico Padano di Lodi.

Gli SSR utilizzati sono marcatori scelti da un pool internazionale di ricercatori esperti di genetica del pesco, per le seguenti caratteristiche: 1) elevato polimorfismo; 2) singolo locus; 3) ottimale distribuzione nel genoma (2 marcatori per gruppo di associazione) con

conseguente segregazione indipendente degli stessi (Jouy et al., 2012). I marcatori microsatelliti utilizzati sono riportati in tabella 4.

Nome Primer	Crom	Sequenza (5'→3')	Motivo ripetuto
UDP98-022	1	Fw: CTAGTTGTGCACACTCACGC Rv: GTCGCAGGAACAGTAAGCCT	(GT) <sub>12</sub> AG <sub>(24)</sub>
UDP96-005	1	Fw: GTAACGCTCGCTACCACAAA Rv: CCTGCATATCACCACCCAG	(AC) <sub>16</sub> TG(CT) <sub>2</sub> CA(CT) <sub>11</sub>
BPPCT001	2	Fw: AATTCCCAAAGGATGTGTATGAG Rv: CAGGTGAATGAGCCAAAGC	(GA) <sub>27</sub>
CPPCT044	2	Fw: ACATGCCGGGTAATTAGCAA Rv: AAAATGCACGTTTCGTCTCC	-
BPPCT007	3	Fw: TCATTGCTCGTCATCAGC Rv: CAGATTTCTGAAGTTAGCGGTA	(AG) <sub>22</sub> (CG) <sub>2</sub> (AG) <sub>4</sub>
UDP96-008	3	Fw: TTGTACACACCCTCAGCCTG Rv: TGCTGAGGTTTCAGGTGAGTG	(CA) <sub>23</sub>
BPPCT015	4	Fw: ATGGAAGGGAAGAGAAATCG Rv: GTCATCTCAGTCAACTTTCCG	(AG) <sub>13</sub>
CPDCT045	4	Fw: TGTGGATCAAGAAAGAGAACCA Rv: AGGTGTGCTTGCACATGTTT	(GA) <sub>16</sub>
BPPCT017	5	Fw: TTAAGAGTTTGTGATGGGAACC Rv: AAGCATAATTTAGCATAACCAAGC	(GA) <sub>28</sub>
BPPCT038	5	Fw: TATATTGTTGGCTTCTTGCATG Rv: TGAAAGTGAAACAATGGAAGC	(GA) <sub>25</sub>
BPPCT025	6	Fw: TCCTGCGTAGAAGAAGGTAGC Rv: CGACATAAAGTCCAAATGGC	(GA) <sub>29</sub>
UDP98-412	6	Fw: AGGGAAAGTTTCTGCTGCAC Rv: GCTGAAGACGACGATGATGA	(AG) <sub>28</sub>
EPPCU5176	7	Fw: ATGACCACACAGAATCACCC Rv: GATCCTCAGCCCGAGTCAAT	-
CPPCT022	7	Fw: GGAGCTGCAATATTGCTG Rv: GTTAGGGAAGCATCTCAC	(CT) <sub>28</sub> CAA(CT) <sub>20</sub>
CPPCT006	8	Fw: GTCCTCCTCCCAGCTTCTCT Rv: GGTTTAGCGCAAAGCTTCA	(CT) <sub>16</sub>
UDP98-409	8	Fw: GCTGATGGGTTTTATGGTTTTTC Rv: CGGACTCTTATCCTCTATCAACA	(AG) <sub>19</sub>

**Tab. 4** Marcatori SSR utilizzati per la genotipizzazione delle accessioni di *Prunus*.

Gli elettroferogrammi sono stati analizzati tramite il software GeneMarker (<http://www.softgenetics.com/GeneMarker.html>) e sono stati riuniti in un'unica matrice contenente le lunghezze di tutti i frammenti ottenuti con le varie coppie di primer. Questo file è stato poi elaborato mediante il software GENETIX (<http://kimura.univ-montp2.fr/genetix>), al fine di determinare alcuni parametri di popolazione, quali il numero di alleli, le frequenze alleliche a ciascun locus e l'eterozigosi attesa e osservata a ciascun locus.

Le relazioni tra gli individui sono state analizzate mediante il programma NTSYSpc2.1. La matrice di similarità è stata calcolata con l'indice BAND, mentre l'analisi dei cluster è stata effettuata tramite il metodo UPGMA. Con i cluster così ottenuti è stato disegnato un dendrogramma (modulo TREE) che esprime le relazioni fra le accessioni analizzate.

## 2.4 GENOTIPIZZAZIONE CON MARCATORI SNP E DEFINIZIONE DELLA CORE COLLECTION

Tra le 877 accessioni di *Prunus*, ne sono state selezionate 339 di *Prunus persica* che sono state genotipizzate con l'IPSC 9K SNP array (Verde et al., 2012). I dati di ibridazione sono stati analizzati mediante il software Genome Studio, filtrando per GenTrain>0.4 ed eliminando i marcatori con un numero di dati mancanti (no call) superiore al 5%. Successivamente, mediante il software PowerMarker, è stata individuata una core collection, la quale è stata utilizzata per le successive analisi di struttura, delle relazioni di parentela tra le accessioni e per l'analisi di associazione. La genotipizzazione delle accessioni con l'IPSC 9K SNP array e la definizione della core collection sono state effettuate dalla Fondazione Edmund Mach di San Michele all'Adige (TN) nell'ambito del progetto Europeo FP7 FruitBreedomics.

## 2.5 FENOTIPIZZAZIONE DELLA CORE COLLECTION

I caratteri fenotipici analizzati sono riportati nella tabella 5.

Caratteri qualitativi	QTL
Tomentosità del frutto	Epoca di fioritura
Colore della polpa	Epoca di maturazione
Tipo di fiore	Sovracolore
Forma delle ghiandole fogliari	Peso del frutto
	Acidità titolabile
	Contenuto di solidi solubili (Brix)

**Tab. 5** Caratteri fenotipici analizzati, suddivisi in caratteri qualitativi (monogenici) e caratteri quantitativi (QTL).

I caratteri qualitativi sono stati valutati solo nell'anno 2012, in quanto rimangono invariati nel tempo. I rilievi dei caratteri quantitativi, invece, sono stati effettuati durante gli anni 2012, 2013 e 2014.

Per quanto riguarda i caratteri quantitativi, si è deciso di effettuare le osservazioni in campo tre volte la settimana. Relativamente ai caratteri legati alla maturazione dei frutti, si è stabilito di effettuare un forte diradamento, lasciando circa tre frutti per branca.

I criteri seguiti per i rilievi dei caratteri quantitativi sono i seguenti:

- Epoca di fioritura: è stata considerata come data di inizio fioritura il momento in cui il 2-3% delle gemme a fiore totali erano in piena fioritura (stadio F di Baggiolini).
- Epoca di maturazione: si è considerata come data di inizio della maturazione il momento in cui il 2-3% dei frutti rispetto al totale si trovavano nello stadio di maturazione fisiologica.
- Peso del frutto: media del peso di 10 frutti, dopo aver pesato ciascuno singolarmente.
- Contenuto di solidi solubili (SSC): Gli stessi 10 frutti usati per calcolare la massa del frutto e il sovracolore sono stati centrifugati e ne è stato ottenuto il succo. Successivamente sono state effettuate tre misurazioni con un rifrattometro digitale (Refracto 30PX, METTLER TOLEDO) e i valori sono stati espressi in gradi Brix.
- Acidità titolabile: È stato utilizzato il succo degli stessi 10 frutti considerati per gli altri caratteri. Sono state effettuate tre misurazioni dell'acidità titolabile, mediante un titolatore automatico G20 con rondolino da 9 postazioni della METTLER TOLEDO, e i valori sono stati espressi in meq/100 ml.
- Sovracolore: è stata considerata la percentuale della superficie del frutto colorata rispetto al colore di fondo. Per la valutazione del sovracolore sono stati utilizzati sempre gli stessi 10 frutti e la valutazione è stata effettuata in maniera empirica.

## **2.6 ANALISI DELLA STRUTTURA DI POPOLAZIONE, DELLE RELAZIONI DI PARENTELA TRA I GENOTIPI E ANALISI DI ASSOCIAZIONE**

La struttura della popolazione e il più probabile numero di cluster (sottopopolazioni) K sono stati calcolati utilizzando il software STRUCTURE v.2.3.4 (Pritchard et al., 2000). Il modello Bayesiano implementato in STRUCTURE valuta la presenza dell'equilibrio di Hardy Weinberg o di linkage disequilibrium nel campione in esame e tenta di raggruppare la popolazione in cluster in modo da minimizzare il disequilibrium e le deviazioni dalle frequenze alleliche di equilibrio all'interno di ogni gruppo. Il programma stima quindi la

probabilità che ha un individuo di appartenere ad ogni cluster così definito, consentendo di assegnarlo al gruppo con maggiore probabilità. Il modello scelto per queste analisi considera che gli individui possano aver avuto origine da una popolazione mescolata (admixture model). Sono state effettuate 10 repliche per ogni analisi, con K che variava da 1 a 5 (K= numero di cluster in cui suddividere le popolazioni), con una fase di burn-in di 10000 step e 10000 iterazioni MCMC (Markov chain Monte Carlo).

Il valore più probabile di K è stato stimato applicando il metodo di Evanno et al. (2005), implementato nel software Harvester (Earl & vonHoldt, 2012). È stato poi utilizzato il programma CLUMPP v 1.1.2 per ottenere un'unica matrice di struttura. Questo programma restituisce una matrice in cui per ogni accessione è indicato un coefficiente di appartenenza ad ognuna delle sottopopolazioni. Tali coefficienti vengono definiti calcolando la media tra tutti i coefficienti ottenuti dalle repliche che sono state effettuate con il programma STRUCTURE. Il valore del coefficiente di appartenenza può variare da 0 a 1 e per ogni accessione, la somma di tutti i coefficienti di appartenenza ad ogni sottopopolazione è uguale a 1.

Infine, la matrice di struttura è stata inserita nel programma DISTRUCT v 1.1 (Rosenberg, 2004) al fine di ottenere una visualizzazione grafica della struttura di popolazione.

Un'ulteriore analisi di struttura è stata effettuata mediante l'analisi delle coordinate principali (PCoA), utilizzando il programma PAST (Hammer et al., 2001). Le prime tre coordinate della PCoA, sono state successivamente utilizzate come matrice Q per effettuare l'analisi di associazione.

L'analisi delle relazioni di parentela è stata effettuata utilizzando il software SPAGeDI (Hardy & Vekemans, 2002) ed è stata ottenuta la matrice K, la quale è una matrice n x n (n=numero di individui della popolazione) che riporta i coefficienti di parentela tra coppie di individui. Tali coefficienti vengono calcolati in base al livello di similarità tra i genotipi, confrontando lo stato allelico dei marcatori utilizzati e possono variare da 0 (massima diversità tra gli individui) a 2 (individui identici). Per poter utilizzare tale matrice nella successiva analisi di associazione tra marcatori e caratteri fenotipici, tutti i valori negativi sono stati sostituiti con 0.

L'analisi di associazione è stata condotta utilizzando il programma TASSEL v 5.2.3 (Bradbury et al., 2007). L'analisi dei caratteri qualitativi è stata effettuata mediante il General Linear Model (GLM), descritto dalla seguente formula: Fenotipo = effetto del marcatore + struttura della popolazione (Q) + residuo. I marcatori significativamente associati ai caratteri qualitativi analizzati sono stati individuati applicando la correzione di

Bonferroni, ad un livello di significatività di  $\alpha=0.0001$ . Per l'analisi di associazione dei caratteri quantitativi, invece, è stato impiegato il Mixed Linear Model (MLM), in cui la struttura della popolazione e le relazioni di parentela tra gli individui vengono utilizzati come covariate. Il MLM è descritto dalla seguente formula: Fenotipo = effetto del marcatore + struttura della popolazione (Q) + relazioni di parentela tra gli individui (K) + residuo. Per individuare i marcatori significativamente associati ai caratteri quantitativi analizzati, i p-value sono stati corretti sulla base del False Discovery Rate (FDR) (Benjamini & Hochberg, 1995), separatamente per ciascun carattere, utilizzando il pacchetto p.adjust, implementato nel programma R (The R development core Team, 2008). Il livello di significatività per individuare associazioni è stato stabilito a 0.01. I Manhattan plot mostrano una visualizzazione grafica dei risultati ottenuti mediante l'analisi di associazione e sono stati realizzati, per tutti i caratteri (qualitativi e quantitativi), utilizzando il pacchetto qqman (Turner, 2014), implementato nel programma R. Tutte le posizioni dei marcatori o dei geni elencati nel presente lavoro, si riferiscono alla versione Peach v1.0 della sequenza del genoma del pesco (Verde et al., 2013).

## 3. RISULTATI

### 3.1 GENOTIPIZZAZIONE CON MARCATORI SSR

Le 877 accessioni di *Prunus* mantenute presso il Centro di Ricerca per la Frutticoltura di Roma sono state genotipizzate con 16 marcatori SSR, al fine di analizzare la variabilità genetica presente nel germoplasma di pesco.

Dei 16 marcatori SSR utilizzati, due (UDP96-008 e CPPCT022) sono stati eliminati dall'analisi finale in quanto gli elettroferogrammi non erano di buona qualità per la maggior parte dei campioni.

Le 14 coppie di primer utilizzate hanno prodotto un totale di 185 alleli, con una media di 13.21 alleli per locus ed un totale di 34 alleli unici, cioè presenti solo in una accessione fra tutte quelle analizzate (Tab. 6). L'eterozigosi osservata varia da un valore minimo di 0.31 (UDP98-409) ad un massimo di 0.57 (CPPCT044) con un valore medio di 0.46, mentre l'eterozigosi attesa varia da un minimo di 0.42 (UDP98-409) ad un massimo di 0.77 (BPPCT025), con un valore medio di 0.64. Per ogni locus, più della metà degli alleli individuati sono rari, cioè presenti con una frequenza inferiore al 5%, per un totale di 135 alleli rari.

Il dendrogramma ottenuto (Fig. S1 posta alla fine della tesi) mostra un coefficiente di similarità che varia da un valore minimo di 0.00 relativo a *Prunus davidiana*, che risulta pertanto il genotipo geneticamente più distante, ad un valore massimo di 1.00 relativo alle cultivar di *Prunus persica* che non si sono discriminate.

Data l'elevata complessità del dendrogramma, sono stati numerati solo i nodi principali che individuano cluster significativi (Fig. S1 posta alla fine della tesi). Il nodo numero 1 individua un grande cluster che include prevalentemente nettarine. Dal nodo numero 2 si origina un cluster molto ampio che include accessioni sia orientali che occidentali, di origine antica e moderna e di tipologia pesca e nettarina. Il nodo numero 3, invece, origina un cluster in cui sono localizzate prevalentemente accessioni occidentali antiche. Il nodo numero 4 individua un cluster di 32 accessioni, per la maggior parte orientali. Questo cluster comprende anche le accessioni 'Sicilia 1', 'Sicilia 2', 'Tabacchiera', 'Platicarpa a', 'Platicarpa b', 'Platicarpa c' e 'Platicarpa pirovano'.

SSR	Cromo- soma	N <sub>a</sub>	N° alleli unici	N° alleli rari (MAF<0.05)	H <sub>e</sub>	H <sub>o</sub>	F <sub>IS</sub>
UDP98-022	1	10	1 ( <i>P. davidiana</i> )	6	0.68	0.41	0.39
UDP96-005	1	14	2 ('Chui huang tao', 'Yu bay')	9	0.64	0.40	0.38
BPPCT001	2	16	2 ('Citation', <i>P. davidiana</i> )	10	0.76	0.49	0.34
CPPCT044	2	20	7 ('Delice', 'Chui huang tao', AB/4, PHP 91-10, 'Nectadiana', 'Lampetella', <i>P. davidiana</i> )	16	0.74	0.57	0.22
BPPCT007	3	14	4 ('Citation', 'Xia Guang', 'Yu Bay', <i>P. davidiana</i> )	11	0.63	0.50	0.21
BPPCT015	4	12	2 ('Pesco fiore rosa', <i>P. davidiana</i> )	10	0.56	0.37	0.34
CPDCT045	4	9	2 (IF817023, AB/6)	6	0.63	0.49	0.23
BPPCT017	5	18	2 ('Fiorenza', 'Higama')	16	0.59	0.46	0.22
BPPCT038	5	13	3 ('Chui huang tao', 'Ermelinda', 'Shan dong')	10	0.58	0.44	0.24
BPPCT025	6	16	1 ( <i>P. davidiana</i> )	11	0.77	0.48	0.38
UDP98-412	6	12	2 (IF817023)	8	0.70	0.52	0.26
EPPCU517 6	7	11	3 (IF817023, 'Citation', 'Queen Ruby')	8	0.64	0.49	0.24
CPPCT006	8	10	1 (IF817023), 2 <i>P. davidiana</i> )	7	0.59	0.44	0.26
UDP98-409	8	10	0	7	0.42	0.31	0.26
<i>Media</i>		<i>13.21</i>		<i>9.6</i>	<i>0.64</i>	<i>0.46</i>	<i>0.28</i>

**Tab. 6** Marcatori SSR usati per genotipizzare le accessioni di pesco. Cromosoma= cromosoma sul quale sono localizzati gli SSR. N<sub>a</sub>=numero di alleli amplificati ad ogni locus. N alleli rari= numero di alleli con frequenza minore al 5%. H<sub>e</sub>=eterozigosi attesa. H<sub>o</sub>=eterozigosi osservata. F<sub>IS</sub>=indice di fissazione, secondo Weir & Cockerham.

In seguito a questa analisi con marcatori SSR, è stato scelto un sottoinsieme di 339 accessioni di *Prunus persica* che sono state successivamente genotipizzate con marcatori SNP. Tali cultivar sono state selezionate sia in modo da essere il più possibile rappresentative della variabilità genetica presente in pesco, sia in base alla loro origine geografica, dando, ove possibile, preferenza alle accessioni locali italiane.

### 3.2 GENOTIPIZZAZIONE CON MARCATORI SNP E DEFINIZIONE DELLA CORE COLLECTION

Le 339 accessioni di *Prunus persica* selezionate sono state genotipizzate mediante l'IPSC 9K SNP array (Verde et al., 2012), sul quale sono posizionati 8144 marcatori SNP. Le reazioni di ibridazione dei campioni di DNA con il chip e il successivo filtraggio sono stati effettuati presso la Fondazione Edmund Mach di San Michele all'Adige (TN).

In seguito all'analisi di filtraggio, 4268 marcatori sono risultati polimorfici e sono stati utilizzati per le analisi successive. Questi marcatori sono così distribuiti negli 8 cromosomi: 572 sul cromosoma 1, 606 sul cromosoma 2, 461 sul cromosoma 3, 891 sul cromosoma 4, 311 sul cromosoma 5, 541 sul cromosoma 6, 438 sul cromosoma 7, 435 sul cromosoma 8 e 13 sugli scaffold non mappati.

Utilizzando i dati di genotipizzazione con i marcatori SNP è stato possibile costruire una core collection di 165 accessioni (Tab. 7). Le accessioni della core collection sono state quindi fenotipizzate per i caratteri di interesse.

Accessione	Parentale femminile	Parentale maschile	Stratificazione
Akizora	Nishino Hakuato	Akatsuki	Orientale
Albatros	Princess of Wales op	-	Occidentale antica
Alexandra	Fayette	RoyalGold	Struttura mista
Amarillo de Agosto 1	-	-	Struttura mista
Angelo Marzocchella	Vesuvio mut	-	Occidentale antica
Anita	-	-	Struttura mista
Ap 31 (6339 - 2n)	-	-	Occidentale antica
Appia	(Southland x Pesco Noce1) op	-	Occidentale moderna

<b>Accessione</b>	<b>Parentale femminile</b>	<b>Parentale maschile</b>	<b>Stratificazione</b>
Armking	(Palomar x Springtime) self	-	Occidentale moderna
Aso 43	-	-	Occidentale antica
Atalanta	(Bonetti 2 x Impero) x Redhaven	-	Struttura mista
Babygold 5	PI35201	(J.H. Hale x Goldfinch) F <sub>2</sub>	Struttura mista
Baekmi Josaeng	Mishima	Nunome Wase	Orientale
Baladi 1	-	-	Occidentale antica
Bambino	-	-	Struttura mista
Bayue Cui	Lu Hua 5	Okubo	Orientale
Bella di Gambettola	-	-	Struttura mista
Bella Lucia	-	-	Struttura mista
Bianca di Carini	Sicilian sel	-	Occidentale antica
Bianca di s. Tomè	-	-	Struttura mista
Bianca prof. Gambi	-	-	Struttura mista
Bianca Tardiva	-	-	Occidentale antica
Big Top	-	-	Occidentale moderna
Blanca de Jolon	-	-	Occidentale antica
BR 2	Aldrighi	Cerrito	Struttura mista
BR 3	Pala op	-	Struttura mista
Branca	Goldmine	Pala	Struttura mista
Buco incavato	Morellona op	-	Occidentale antica
Caldesi 2000	Redgold	Snow Queen	Occidentale moderna
Calred	(Fireglobe x 3-4M) self	-	Occidentale moderna
Camden	FV89-14	Springtime	Occidentale moderna
Capdeboscq	(Lake City x Intermediario) op	-	Struttura mista
Centenaria	Docura 2 op	-	Struttura mista

<b>Accessione</b>	<b>Parentale femminile</b>	<b>Parentale maschile</b>	<b>Stratificazione</b>
Cerrito	(Lake City x Interludio) op	-	Struttura mista
Champion	-	-	Occidentale antica
Changbang Mutant	-	-	Struttura mista
China	-	-	Struttura mista
Chiyomaru	Nakatsu Hakuto x Nunome Wase	Nunome Wase op#2	Orientale
Chui Huang Tao	-	-	Struttura mista
Chun Lei	Sunago wase	Bai Xiang Lu	Orientale
Ciccio Petrino	-	-	Struttura mista
Circe	USDA dwarf sdkg op	-	Occidentale moderna
City 32-82	-	-	Struttura mista
Cotogna Cicalini	-	-	Struttura mista
CP 88/2	-	-	Struttura mista
Crasiommolo Rosso	-	-	Occidentale antica
Crown Princess	Red Diamond	Peach sel	Struttura mista
De Wet	NJC18/21	SN45/3	Orientale
Diamante	Convenio	(Cardeal x Aldrighi)	Struttura mista
Diamond Ray	Red Diamond	Peach sel	Occidentale moderna
Docura	Alo-Docura	Tutu	Struttura mista
Early Risinger	-	-	Struttura mista
Earlynette	-	-	Occidentale moderna
Elegant Lady	EarlyOHenry	JulyLady	Occidentale moderna
Entella	-	-	Struttura mista
Fairtime	(Rodeo x Kirkman Gem) op	-	Occidentale moderna
Fay Elberta	Elberta op	-	Occidentale moderna
Fei Cheng Bai Li 17	-	-	Occidentale moderna
Fei Cheng Tao	-	-	Struttura mista

Accessione	Parentale femminile	Parentale maschile	Stratificazione
Ficarazzo	-	-	Struttura mista
Fidelia	(OHenry x GiantBabcock)	(MayGrand x SamHouston)	Occidentale moderna
Fior di Maggio	-	-	Occidentale antica
Fior di Novembre	-	-	Struttura mista
Flordahome	(Chico 11 x <i>Prunus davidiana</i> (C-26712))F2	-	Struttura mista
Flordastar	Flordagold	EarliGrande	Struttura mista
FruherRote Ingelheimer	-	-	Occidentale antica
Fuoco di Romagna	-	-	Struttura mista
Gemini	-	-	Occidentale antica
GF 305	-	-	Occidentale antica
Gianni Laura Dolce	-	-	Struttura mista
Gran Sun	May Grand mut	-	Struttura mista
Greta	Stark SunGlo	IF7131649	Struttura mista
Higama	-	-	Struttura mista
Huangshaneardew	-	-	Orientale
IF 8810268	IF 8431354 op	-	Struttura mista
Ingo Peach	-	-	Struttura mista
Iris Rosso	Porpurea	(J.H. Hale x Redbird)	Occidentale moderna
Jacquotte	-	-	Struttura mista
Jing Ju	Okubo	Okitsu	Struttura mista
Jingshaneardew	-	-	Orientale
July Lady	JHHale	MerrillGem	Occidentale moderna
La Pecher	La Feliciana op	-	Struttura mista
Lampetella	-	-	Occidentale antica
Leonforte 1	-	-	Occidentale antica
Luisa Berselli	-	-	Struttura mista

Accessione	Parentale femminile	Parentale maschile	Stratificazione
Madonna di Agosto	-	-	Occidentale antica
Majestic	Meadow Lark	op	Struttura mista
Maribelle	-	-	Struttura mista
May Flower	-	-	Occidentale antica
Maycrest	Springcrest mut	-	Struttura mista
Mb Petrini	-	-	Struttura mista
Michelini	-	-	Struttura mista
Miorita	-	-	Struttura mista
Miraflores	-	-	Occidentale antica
Montclair	-	-	Occidentale antica
Moscatello	-	-	Occidentale moderna
Natalino di Leonforte	-	-	Occidentale antica
Nett. Russa 598-81	-	-	Struttura mista
NJ weeping	-	-	Struttura mista
O.P.G.	<i>P. persica</i>	<i>P. dulcis</i>	Struttura mista
October Red	-	-	Occidentale antica
Ouromel 3	Tutu	Columbina	Struttura mista
Palazzina	-	-	Occidentale antica
Pelo Tardivo	-	-	Struttura mista
Percoca di Romagna 7	-	-	Occidentale antica
Percoco Bianco	-	-	Occidentale antica
Percoco di Turi	-	-	Occidentale antica
Percoco Novembre	-	-	Occidentale antica
Persi de Vin	-	-	Occidentale antica
Pesca a Cuore Capozzi	-	-	Struttura mista
Pesca Rossa della Vigna	-	-	Occidentale antica
Pesca Settembrina a polpa rossa	-	-	Occidentale antica

<b>Accessione</b>	<b>Parentale femminile</b>	<b>Parentale maschile</b>	<b>Stratificazione</b>
Pesco dei Santi	-	-	Struttura mista
Pesco fiore rosa	-	-	Orientale
PHN 91-12	-	-	Struttura mista
PHN 91-14	-	-	Occidentale moderna
PHP 91-05	-	-	Occidentale moderna
Pillar	-	-	Struttura mista
Precocinho	Diamante op	-	Struttura mista
Primerose	-	-	Struttura mista
<i>Prunus ferganensis</i>	-	-	Struttura mista
Rancho Resistant	S-37 op	-	Struttura mista
Red Delight	(May Grand x Red Grand) op	-	Occidentale moderna
Regina di Londa	-	-	Struttura mista
Rich Lady	Amparo op	-	Occidentale moderna
Rojo di Albesa	-	-	Struttura mista
Rosalina	(Real x Sunlite) op	-	Struttura mista
Rossa de Cecchi	-	-	Struttura mista
Rossa di Lugo	-	-	Struttura mista
Rossa di S. Carlo	-	-	Struttura mista
Royal Lee	-	-	Occidentale moderna
Rubirich	-	-	Occidentale moderna
Russotto	-	-	Struttura mista
S. Anna Balducci	-	-	Occidentale antica
Sahua Hong Pantao	-	-	Orientale
San Giorgio	-	-	Struttura mista
Sanguigna di Savoia	-	-	Struttura mista
Sanguinella	-	-	Struttura mista
Sensation	-	-	Occidentale moderna
Settembrina Turca	-	-	Occidentale antica

Accessione	Parentale femminile	Parentale maschile	Stratificazione
Shen Zhou Mi Tao	-	-	Struttura mista
Shizukured	(Okitsu x NjN17) op	-	Struttura mista
Sicilia 1	-	-	Orientale
Silver Rome	Redgold	SnowQueen	Occidentale moderna
Silvette 2	-	-	Struttura mista
Snow Queen	-	-	Occidentale moderna
Souvenir Nikitski	-	-	Struttura mista
Spring Bright	May Diamond	sel	Occidentale moderna
Stark Saturn	Pallas	NJ 602903	Struttura mista
Starlite	FV8914	Springtime	Occidentale moderna
Summer Pearl	NJ554039	NJ585399	Struttura mista
Suncrest	Alamar	GoldDust	Occidentale antica
Tabacchiera	-	-	Orientale
Tardibelle	-	-	Struttura mista
Tardiva di S. Vittorino	-	-	Occidentale antica
Tatura Dawn	Levis	Levis	Occidentale antica
Terzarola Gialla	-	-	Occidentale antica
Vaccaro Roccalmunto	-	-	Struttura mista
Vigna Verde	-	-	Occidentale antica
Vittorio Emanuele III	-	-	Struttura mista
Xia Guang	-	-	Struttura mista
Yoshihime	(Nakatsu Hakuto x NunomeWase)	Akatsuki	Occidentale moderna
Yu Bai	-	-	Struttura mista
Yumyeong	Yamato-Wase	Nunome-Wase	Orientale
Zansetsushidare	-	-	Struttura mista

**Tab. 7** Lista delle 165 accessioni della Core Collection. Per ogni accessione sono riportati i parentali e la sottopopolazione a cui appartiene, in base al risultato della stratificazione ottenuta mediante il programma STRUCTURE (vedere §3.4).

### 3.3 FENOTIPIZZAZIONE DELLA CORE COLLECTION

I caratteri presi in considerazione sono stati sia di tipo qualitativo che quantitativo. In tabella 8 sono riassunti i risultati relativi alla fenotipizzazione per i caratteri monogenici.

Carattere qualitativo		N° di accessioni
Tomentosità del frutto	Pesco	139
	Nettarina	25
	Dati mancanti	1
Colore della polpa	Bianco	82
	Giallo	63
	Dati mancanti	20
Tipo di fiore	Rosaceo	105
	Campanulaceo	54
	Dati mancanti	6
Forma delle ghiandole fogliari	Reniformi	126
	Globose	34
	Dati mancanti	5

**Tab. 8** Rilievi dei caratteri qualitativi

Per quanto riguarda il carattere pesco/nettarina, nella nostra popolazione erano presenti 139 pesche e 25 nettarine, più una accessione (O.P.G) che è stata considerata come dato mancante, in quanto non produce frutti.

La popolazione, per il colore della polpa era composta da 63 accessioni a polpa gialla e 82 accessioni a polpa bianca. Erano presenti anche 6 accessioni a polpa rossa che sono state trattate come dati mancanti in quanto, essendo il carattere 'polpa rossa' epistatico sul giallo o bianco, non consente di riconoscere il fenotipo per questo carattere. Altre 16 accessioni non hanno prodotto frutti o i frutti sono stati danneggiati da *Ceratitis capitata* e sono state ugualmente considerate come dato mancante.

Per quanto riguarda il carattere tipo di fiore, nella popolazione erano presenti 105 accessioni con fiore rosaceo e 54 accessioni con fiore campanulaceo, mentre per il

carattere forma delle ghiandole fogliari, 126 accessioni presentavano ghiandole fogliari reniformi e 34 accessioni ghiandole fogliari globose.

In tabella 9 sono riportati il valore minimo, il valore massimo, il valore medio e la deviazione standard per i caratteri quantitativi analizzati negli anni 2012, 2013, 2014. Per i caratteri relativi all'epoca di maturazione e fioritura, le date sono indicate come nel calendario giuliano.

<b>Carattere quantitativo</b>	<b>Min</b>	<b>Max</b>	<b>Media</b>	<b>Dev Standard</b>
Data fioritura 2012	51	87	77.3	7.2
Data fioritura 2013	57	95	81.3	8.2
Data fioritura 2014	48	82	67.6	7.2
Data maturazione 2012	152	284	218.9	34.6
Data maturazione 2013	151	304	206.2	32.9
Data maturazione 2014	148	258	201.9	28.5
Sovracolore 2012 (%)	0	100	40.8	30.8
Sovracolore 2013 (%)	0	95	42.1	29.9
Sovracolore 2014 (%)	0	100	58.4	28.8
Peso del frutto 2012	26.9	342.1	164.0	58.8
Peso del frutto 2013	39.8	332.3	186.2	62.9
Peso del frutto 2014	35.6	368.8	193.8	71.1
Acidità titolabile 2012	3.3	26.5	13.0	4.9
Acidità titolabile 2013	4.0	27.0	13.0	5.4
Acidità titolabile 2014	3.0	23.5	12.4	4.2
SSC (Brix) 2012	7.3	19.5	12.9	2.7
SSC (Brix) 2013	6.3	18.2	11.7	2.3
SSC (Brix) 2014	7.2	16.9	11.5	2.1

**Tab. 9** Valore minimo, valore massimo, media e deviazione standard dalla media per i caratteri quantitativi fenotipizzati negli anni 2012, 2013, 2014.

Per verificare la relazione esistente tra i dati rilevati per ciascun carattere nei tre anni, è stato calcolato il coefficiente di Pearson confrontando, per coppie di anni, i rilievi effettuati per ogni carattere (Tab. 10).

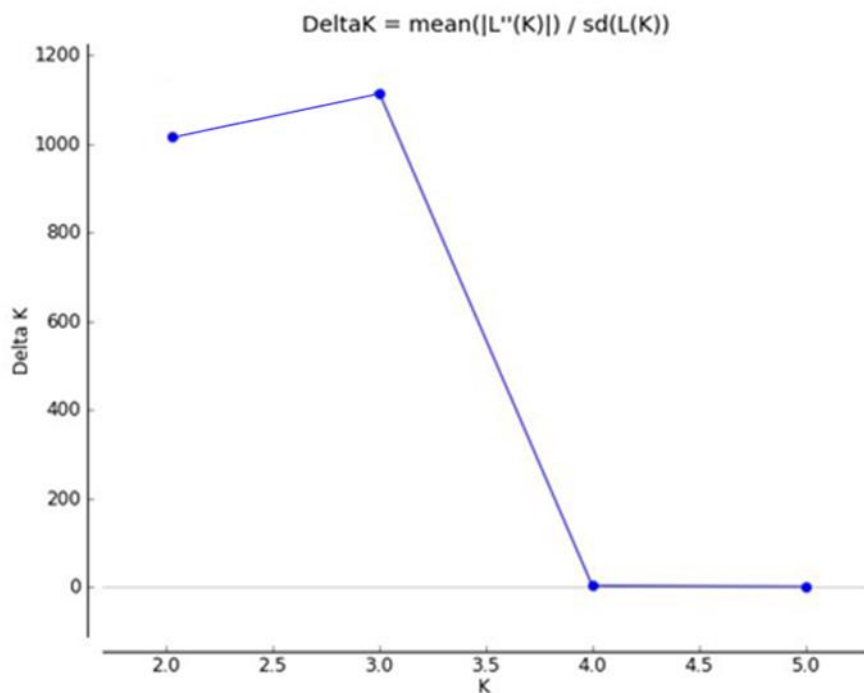
<b>Carattere</b>	<b>r 2012/2013</b>	<b>r 2012/2014</b>	<b>r 2013/2014</b>
Data fioritura	0.96	0.86	0.91
Data Maturazione	0.99	0.99	0.99
Sovracolore	0.85	0.81	0.81
Peso del frutto	0.82	0.73	0.72
Acidità titolabile	0.76	0.77	0.82
SSC (Brix)	0.67	0.58	0.68

**Tab. 10** Valori del coefficiente di correlazione di Pearson (r) per i diversi anni di analisi dei caratteri quantitativi.

I dati fenotipici acquisiti nei 3 anni di osservazione effettuata sono risultati in generale altamente correlati con valori di  $r > 0.7$  per tutti i caratteri fatta eccezione per il contenuto in solidi solubili (SSC) che ha restituito nel confronto 2012/2014 il coefficiente di Pearson più basso ( $r=0.58$ ) e negli anni 2012/2013 e 2013/2014 un valore di 0.67 e 0.68, rispettivamente. La data di maturazione è il carattere che ha ottenuto il più alto indice di correlazione in tutti e tre i confronti effettuati ( $r=0.99$ ).

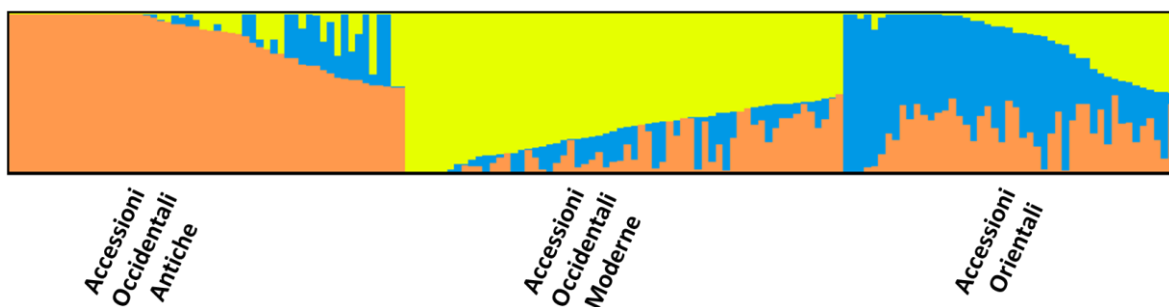
### **3.4 ANALISI DELLA STRUTTURA DI POPOLAZIONE**

Una prima analisi della struttura di popolazione è stata effettuata mediante il programma STRUCTURE ed è stato osservato che il più probabile numero di sottopopolazioni (K) in cui è divisa la popolazione iniziale è 3 (Fig. 7). Le accessioni sono state considerate appartenenti ad una sottopopolazione se presentavano un coefficiente di appartenenza maggiore o uguale a 0.8. In ciascuna sottopopolazione è possibile riconoscere delle caratteristiche uniche che accomunano le accessioni appartenenti a quel dato gruppo, in particolare le tre sottopopolazioni sono così suddivise: accessioni orientali (13 accessioni) che includono anche le accessioni ‘Sicilia 1’ e ‘Tabacchiera’, accessioni occidentali antiche (38 accessioni) e accessioni occidentali moderne, ottenute mediante i programmi di miglioramento genetico (29 accessioni). Le rimanenti 85 accessioni (52%) avevano coefficienti di appartenenza inferiori a 0.8 e sono state considerate a struttura mista (admixed; Tab. 7).



**Fig. 7** Grafico che mostra il numero più probabile di sottopopolazioni (K) in cui è divisa la popolazione iniziale, individuato mediante il metodo di Evanno et al., 2005.

Una rappresentazione grafica della suddivisione in base alla struttura della core collection analizzata è mostrata in figura 8.

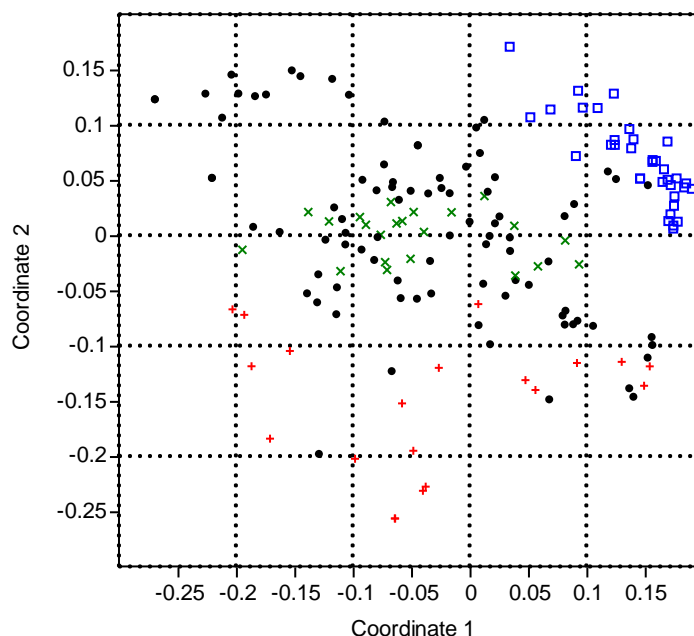


**Fig. 8** Analisi della struttura di popolazione effettuata con STRUCTURE per K=3. L'immagine è stata ottenuta mediante il programma DISTRUCT v1.1.

Un'ulteriore studio della struttura della popolazione è stato effettuato mediante l'analisi delle coordinate principali (PCoA). Le prime tre coordinate spiegano il 16%, il 9% e il 7% della variabilità totale, per una variabilità cumulativa del 32%. Tutte le altre coordinate spiegano ognuna una varianza inferiore al 4%.

I risultati della PCoA sono mostrati in figura 9 sullo scatterplot. Le accessioni sono state colorate sulla base dei risultati di stratificazione ottenuti con il programma STRUCTURE: le accessioni orientali sono indicate con croci rosse, le accessioni occidentali antiche con quadratini blu, le accessioni occidentali moderne con croci verdi e le accessioni admixed con puntini neri.

Come si può osservare, le accessioni occidentali antiche si raggruppano nel quadrante in alto a destra, le accessioni orientali nei quadranti in basso e le accessioni occidentali moderne nella parte centrale del grafico, insieme alle accessioni admixed.



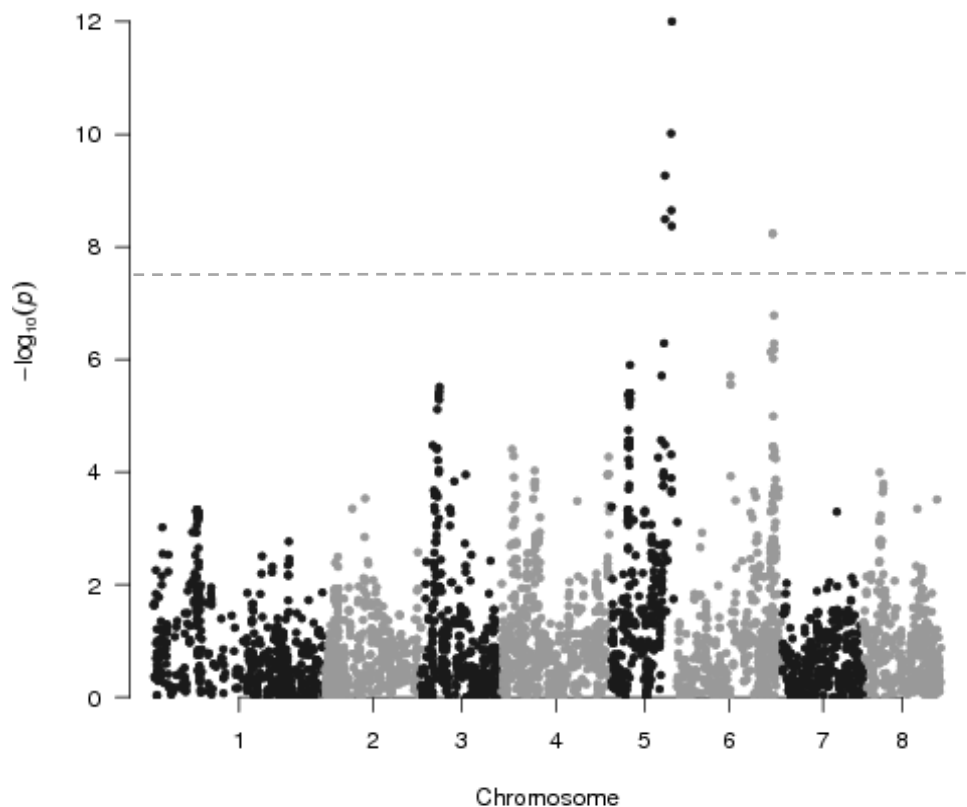
**Fig. 9** Grafico dell'analisi delle coordinate principali (PCoA) effettuata con 165 accessioni e 4268 marcatori SNP. I colori con cui sono indicate le accessioni sono stati attribuiti sulla base dei risultati di stratificazione ottenuti con STRUCTURE. Le accessioni orientali sono indicate con croci rosse, le accessioni occidentali antiche con quadratini blu, le accessioni occidentali moderne con croci verdi e le accessioni admixed con puntini neri

## 3.5 ANALISI DI ASSOCIAZIONE

### 3.5.1 Caratteri qualitativi

#### 3.5.1.1 Tomentosità del frutto

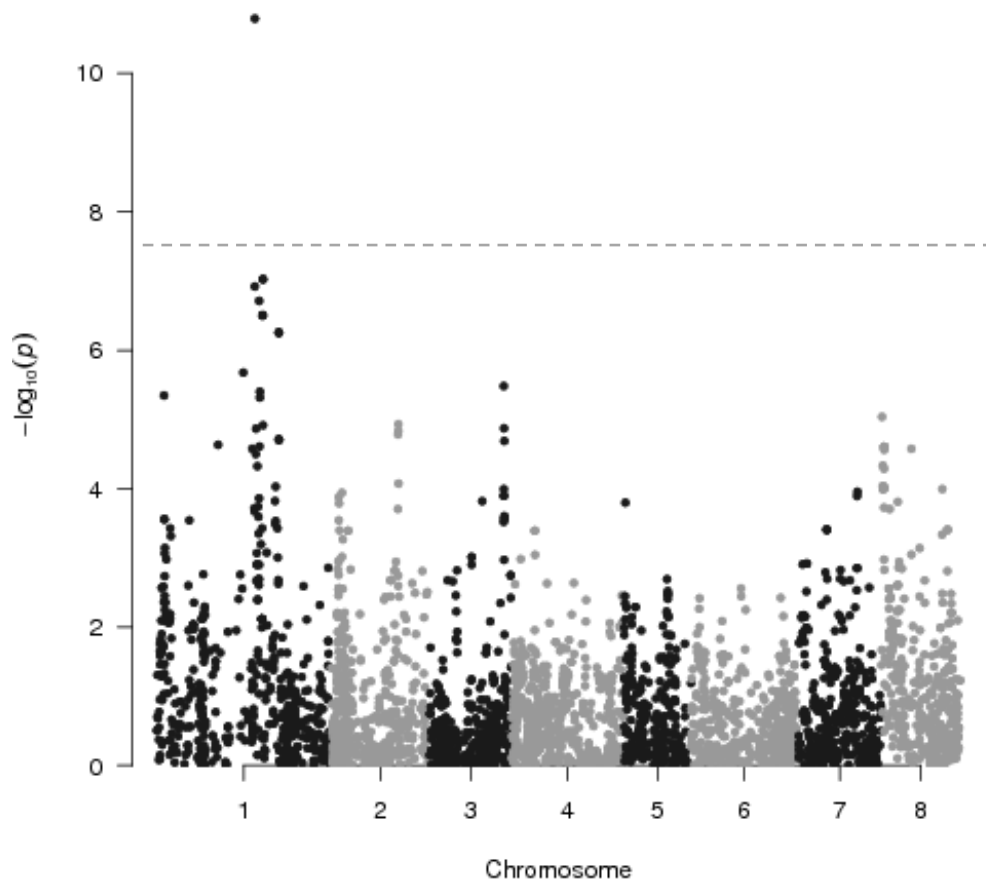
Sono stati individuati 6 marcatori associati al carattere tomentosità del frutto localizzati sul cromosoma 5, in una regione compresa tra 14 893 692 e 16 774 236. Ulteriori 2 marcatori associati sono localizzati sul cromosoma 6 (scaffold\_6: 26 245 947.. 26 247 380) (Tab. 11). Lo SNP maggiormente associato al carattere tomentosità del frutto è lo SNP\_IGA\_603047 localizzato sul cromosoma 5 in posizione 16 774 236 (Fig. 10) con un p-value di 9.89E-13.



**Fig. 10** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere tomentosità del frutto. L'analisi è stata effettuata utilizzando il GLM e i p-value sono stati corretti in base alla correzione di Bonferroni, ad un livello di significatività di  $\alpha = 0.0001$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate sono riportati i valori di significatività (p-value) espressi come  $-\log_{10}$  p-value.

### 3.5.1.2 Colore della polpa

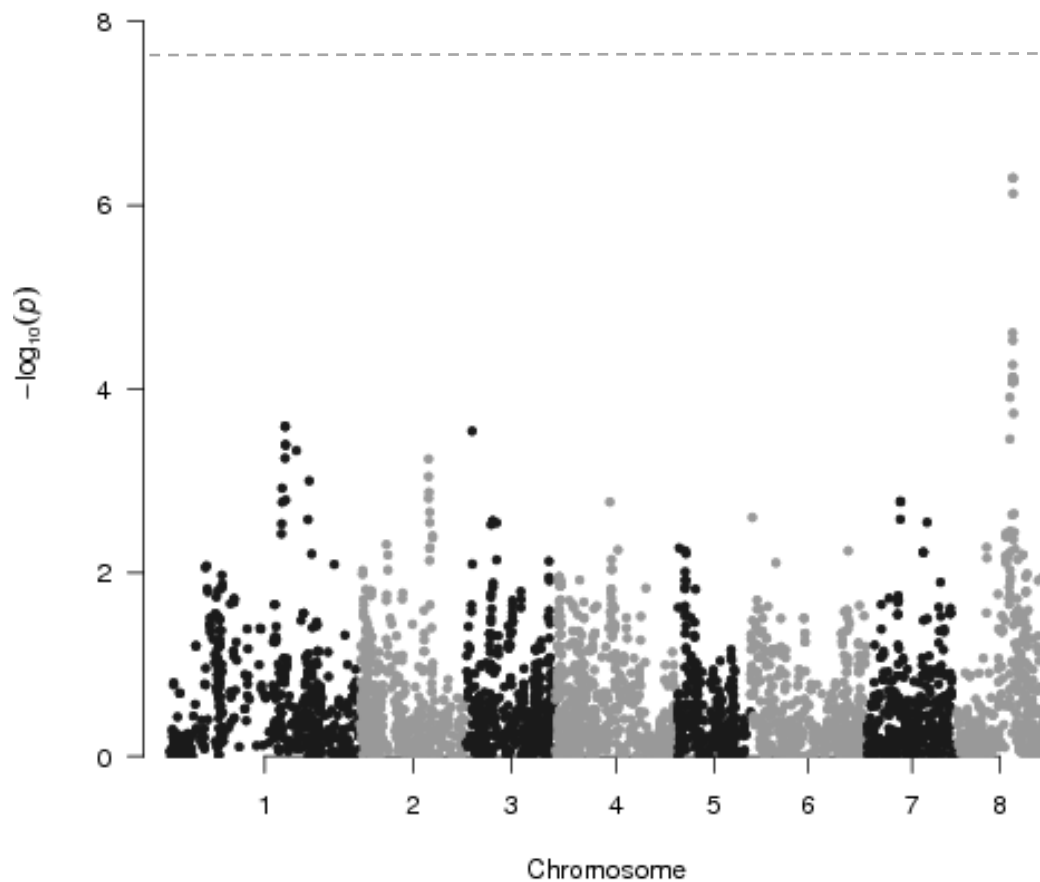
L'analisi di associazione ha individuato un marcatore associato al carattere colore della polpa localizzato sul cromosoma 1, lo SNP\_IGA\_90213 (scaffold\_1:26 479 525) con un p-value di 1.62E-11 (Tab. 11; Fig. 11).



**Fig. 11** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere colore della polpa. L'analisi è stata effettuata utilizzando il GLM e i p-value sono stati corretti in base alla correzione di Bonferroni, ad un livello di significatività di  $\alpha = 0.0001$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate sono riportati i valori di significatività (p-value) espressi come  $-\log_{10}$  p-value.

### 3.5.1.3 Tipo di fiore

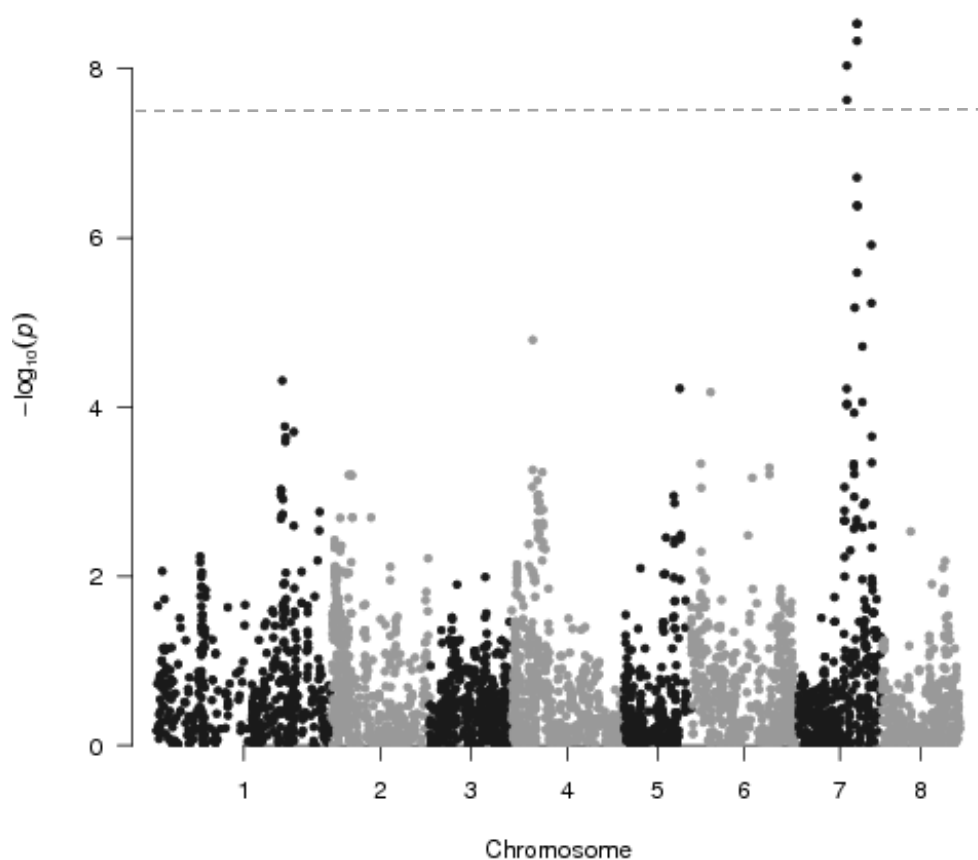
È stato individuato un marcatore associato al carattere tipo di fiore localizzato sul cromosoma 8, lo SNP SNP\_IGA\_864149 a 13 756 987, con un p-value di 3.56E-09 (Tab.11; Fig. 12).



**Fig. 12** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere tipo di fiore. L'analisi è stata effettuata utilizzando il GLM e i p-value sono stati corretti in base alla correzione di Bonferroni, ad un livello di significatività di  $\alpha = 0.0001$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate sono riportati i valori di significatività (p-value) espressi come  $-\log_{10}$  p-value.

### 3.5.1.4 Forma delle ghiandole fogliari

Sono stati individuati 5 marcatori SNP associati alla forma delle ghiandole fogliari, localizzati sul cromosoma 7 in una regione compresa tra 13 461 563 e 16 219 910 (Tab. 11). Gli SNP maggiormente associati sono: lo SNP\_IGA\_779439 (a 16 194 545) con un p-value di 2.94E-09, lo SNP\_IGA\_779520 (a 16 209 516) con un p-value di 4.69E-09 e lo SNP SNP\_IGA\_779540 (a 16 219 910) con un p-value di 2.94E-09 (Fig. 13).



**Fig. 13** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere forma delle ghiandole fogliari. L'analisi è stata effettuata utilizzando il GLM e i p-value sono stati corretti in base alla correzione di Bonferroni, ad un livello di significatività di  $\alpha = 0.0001$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate sono riportati i valori di significatività (p-value) espressi come  $-\log_{10}$  p-value.

Carattere qualitativo	Marcatore	Cromo- soma	Posizione (Peach v1.0)	p-value
Tomentosità del frutto	SNP_IGA_600509	5	14 893 692	3.23E-09
	SNP_IGA_600517	5	14 894 563	5.41E-10
	SNP_IGA_602331	5	16 556 187	9.66E-11
	SNP_IGA_602512	5	16 616 848	2.25E-09
	SNP_IGA_602901	5	16 707 379	4.28E-09
	<b>SNP_IGA_603047</b>	<b>5</b>	<b>16 774 236</b>	<b>9.89E-13</b>
	SNP_IGA_694521	6	26 245 947	5.80E-09
	SNP_IGA_694526	6	26 247 380	5.80E-09
Colore della polpa	<b>SNP_IGA_90213</b>	<b>1</b>	<b>26 479 525</b>	<b>1.62E-11</b>
Tipo di fiore	<b>SNP_IGA_864149</b>	<b>8</b>	<b>13 756 987</b>	<b>3.56E-09</b>
Forma ghiandole fogliari	SNP_IGA_772183	7	13 461 563	2.34E-08
	SNP_IGA_772231	7	13 466 067	9.21E-09
	<b>SNP_IGA_779439</b>	<b>7</b>	<b>16 194 545</b>	<b>2.94E-09</b>
	<b>SNP_IGA_779520</b>	<b>7</b>	<b>16 209 516</b>	<b>4.69E-09</b>
	<b>SNP_IGA_779540</b>	<b>7</b>	<b>16 219 910</b>	<b>2.94E-09</b>

**Tab. 11** Schema riassuntivo dei marcatori individuati in associazione ai caratteri qualitativi.

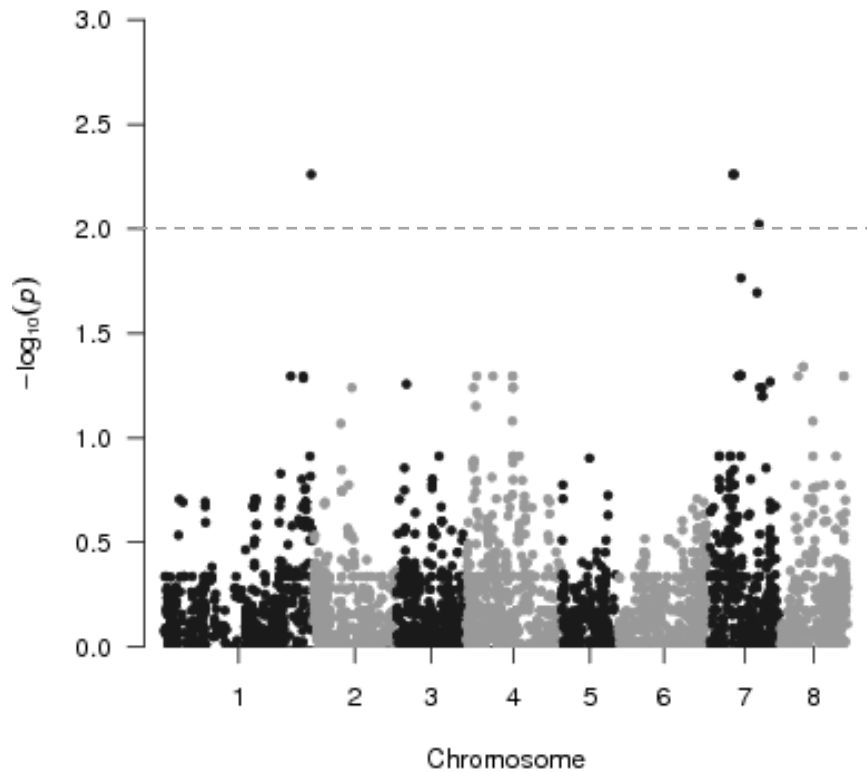
### 3.5.2 Caratteri quantitativi

#### 3.5.2.1 Epoca di fioritura

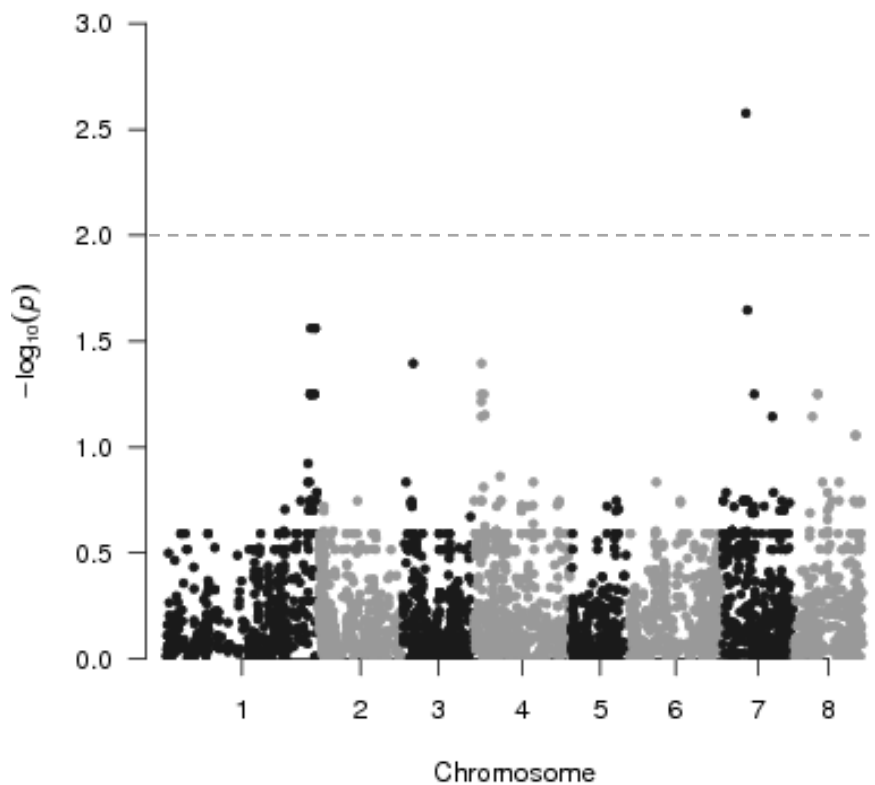
Per quanto riguarda l'epoca di fioritura sono stati individuati marcatori associati a questo carattere per gli anni 2012 e 2013 (Tab. 12).

Per l'anno 2012 è stato individuato uno SNP sul cromosoma 1, lo SNP\_IGA\_135188 in posizione 46 542 316, con un p-value (corretto con FDR) di 5.49E-03 e un  $R^2$  di 0.19; altri 3 SNP sono stati identificati sul cromosoma 7: lo SNP\_IGA\_748301 e lo SNP\_IGA\_751897 in una regione compresa tra 7 681 738 e 8 151 156, entrambi con un p-value di 5.49E-03 e lo (SNP\_IGA\_778535), in posizione 15 904 306, con un p-value di 9.49E-03 (Fig. 14).

Per l'anno 2013, un solo marcatore risulta associato all'epoca di fioritura: lo SNP\_IGA\_748301 sul cromosoma 7 a 7 681 738 con un p-value di 2.65E-03 (Fig. 15). Questo marcatore è uno di quelli individuati anche nell'anno 2012. Nessuna associazione è stata identificata per il 2014.



**Fig. 14** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere epoca di fioritura per l'anno 2012. L'analisi è stata effettuata utilizzando il MLM e i p-value sono stati corretti in base al FDR, ad un livello di significatività di  $\alpha = 0.01$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate  $-\log_{10}$  p-value.



**Fig. 15** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere epoca di fioritura per l'anno 2013. L'analisi è stata effettuata utilizzando il MLM e i p-value sono stati corretti in base al FDR, ad un livello di significatività di  $\alpha = 0.01$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate  $-\log_{10}$  p-value.

### 3.5.2.2 Epoca di maturazione

Relativamente all'epoca di maturazione, sono stati osservati marcatori associati a tale carattere sui cromosomi 2, 4 e 6 (Tab. 12, Fig. 16, 17, 18).

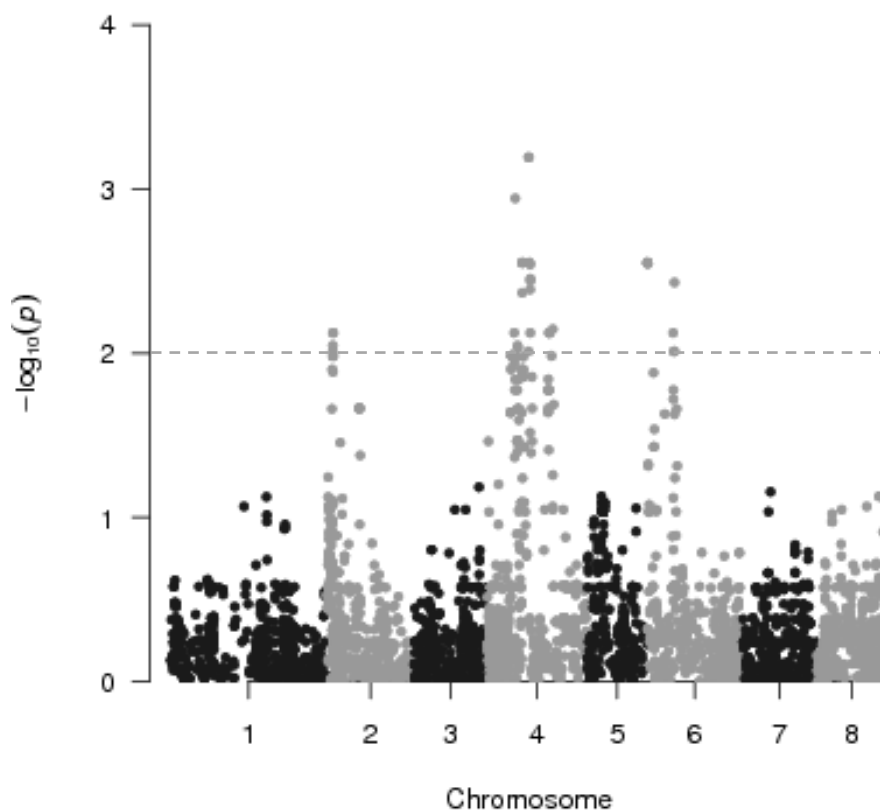
I marcatori sul cromosoma 2 sono stati individuati solo per l'anno 2012 e sono localizzati all'inizio del cromosoma 2, in una regione genomica compresa tra 2 094 415 e 2 186 582.

Sul cromosoma 4 è stata individuata una regione associata all'epoca di maturazione per tutti e tre gli anni di osservazione. Nel 2012, questa regione è molto ampia e comprende 18 SNP che coprono una distanza di 4.9 Mb, da posizione 8 222 657 a 13 112 264. Nel 2013 e 2014 questa regione è compresa tra 10 615 885 e 10 685 500.

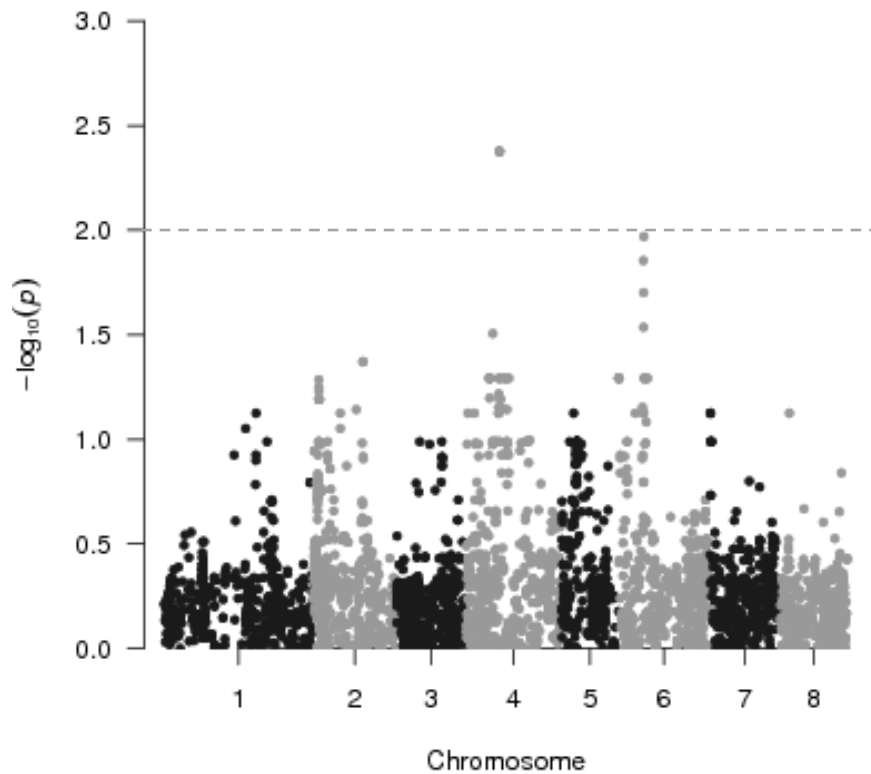
Solo per l'anno 2012 è stata individuata un'altra regione significativamente associata al carattere sul cromosoma 4, compresa tra 18 742 047 e 19 905 501.

Negli anni 2012 e 2014 sono stati individuati marcatori associati al carattere epoca di maturazione all'inizio del cromosoma 6, in una regione compresa tra 45 035 e 203 384.

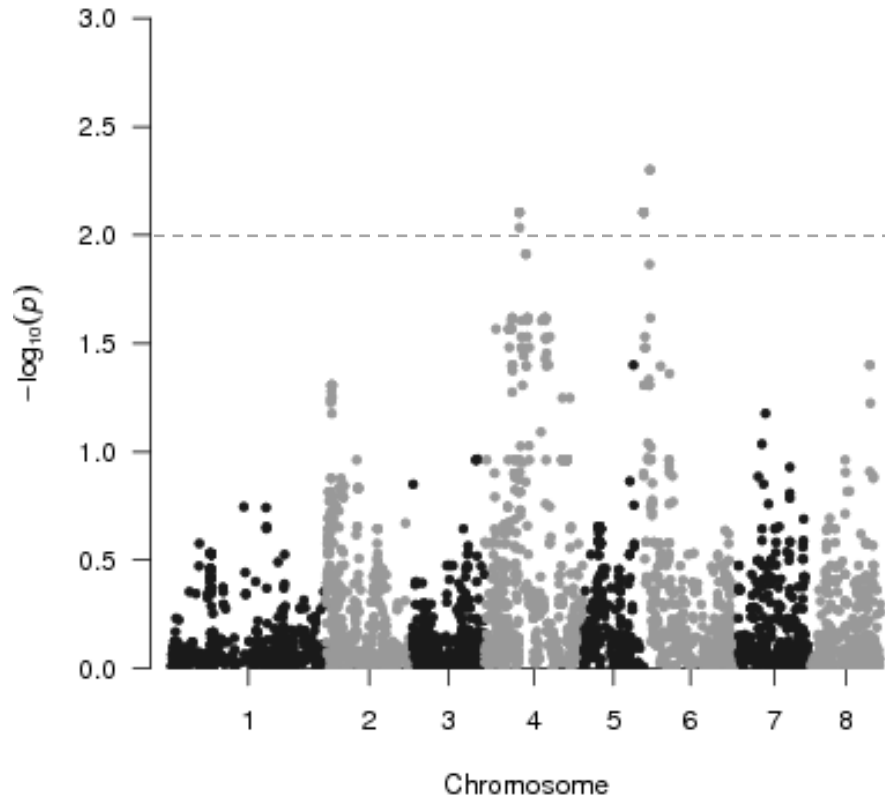
Nel 2014, questa regione si estende fino a 2 044 947. Solo per il 2012, inoltre, è stata individuata un'altra regione sul cromosoma 6, compresa tra 7 833 639 e 8 238 299.



**Fig. 16** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere epoca di maturazione per l'anno 2012. L'analisi è stata effettuata utilizzando il MLM e i p-value sono stati corretti in base al FDR, ad un livello di significatività di  $\alpha = 0.01$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate  $-\log_{10}$  p-value.



**Fig. 17** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere epoca di maturazione per l'anno 2013. L'analisi è stata effettuata utilizzando il MLM e i p-value sono stati corretti in base al FDR, ad un livello di significatività di  $\alpha = 0.01$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate  $-\log_{10}$  p-value.



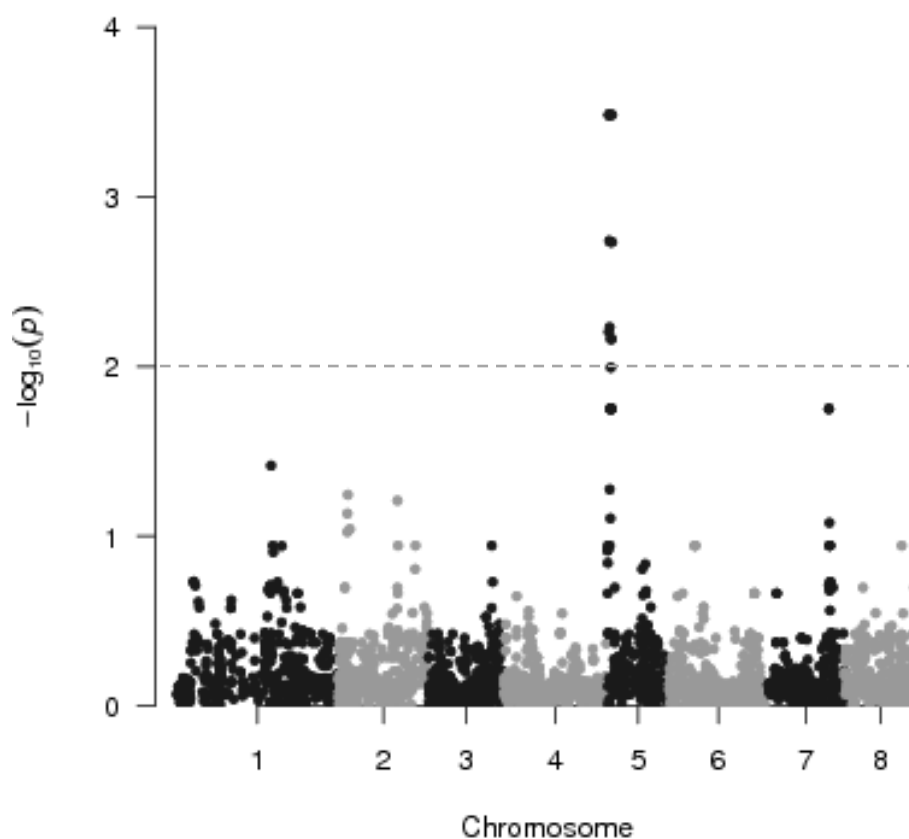
**Fig. 18** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere epoca di maturazione per l'anno 2014. L'analisi è stata effettuata utilizzando il MLM e i p-value sono stati corretti in base al FDR, ad un livello di significatività di  $\alpha = 0.01$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate  $-\log_{10}$  p-value.

### 3.5.2.3 Acidità titolabile

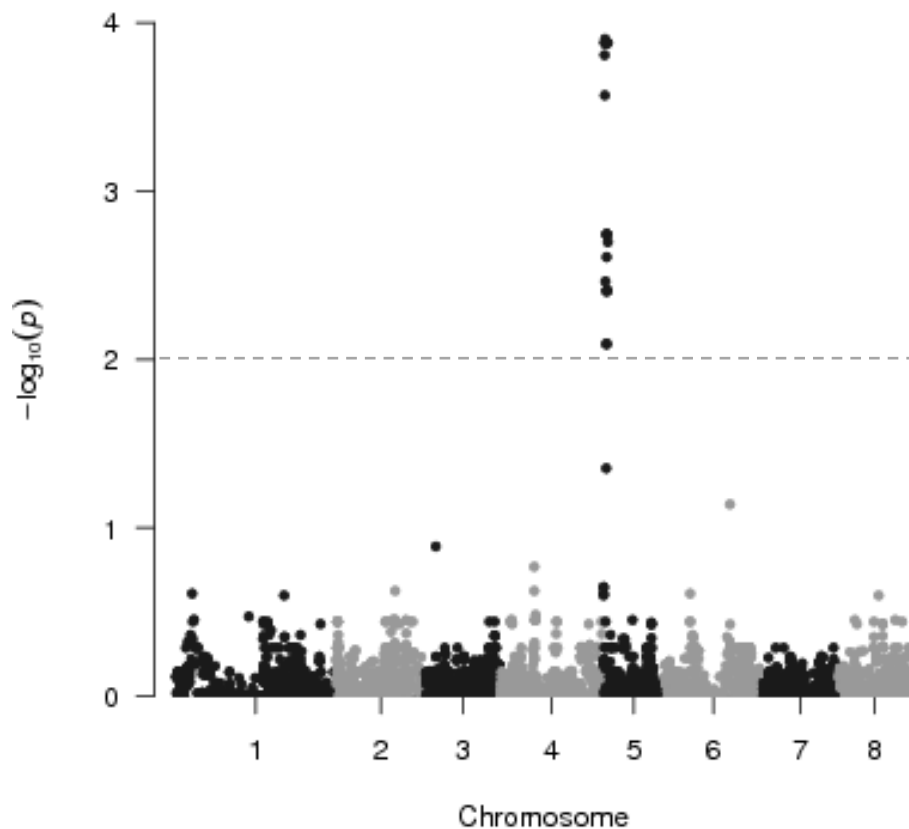
Per quanto riguarda l'acidità titolabile, sono stati individuati marcatori associati a tale carattere per gli anni 2012 e 2014, tutti localizzati sul cromosoma 5 (Tab. 12).

Per l'anno 2012, sono stati individuati 12 marcatori posizionati sul cromosoma 5 in una regione di circa 1 Mb, che va da 557 503 a 1 518 380, in cui lo SNP che presenta il valore massimo di varianza spiegata ( $R^2=0.26$ ) è lo SNP\_IGA\_548512, in posizione 1 503 401 con un p-value di  $3.28E-04$  (Fig. 19).

Per l'anno 2014, sono stati osservati 19 marcatori associati al carattere acidità sul cromosoma 5, che ricadono nella stessa regione dei marcatori individuati per l'anno 2012, ossia da 557 503 a 1 518 380 (Fig. 20). Per questo anno lo SNP che presenta il valore massimo di varianza spiegata ( $R^2=0.45$ ) è lo SNP\_IGA\_544961, in posizione 698 214 con un p-value di  $1.25E-04$ .



**Fig. 19** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere acidità titolabile per l'anno 2012. L'analisi è stata effettuata utilizzando il MLM e i p-value sono stati corretti in base al FDR, ad un livello di significatività di  $\alpha=0.01$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate  $-\log_{10}$  p-value.

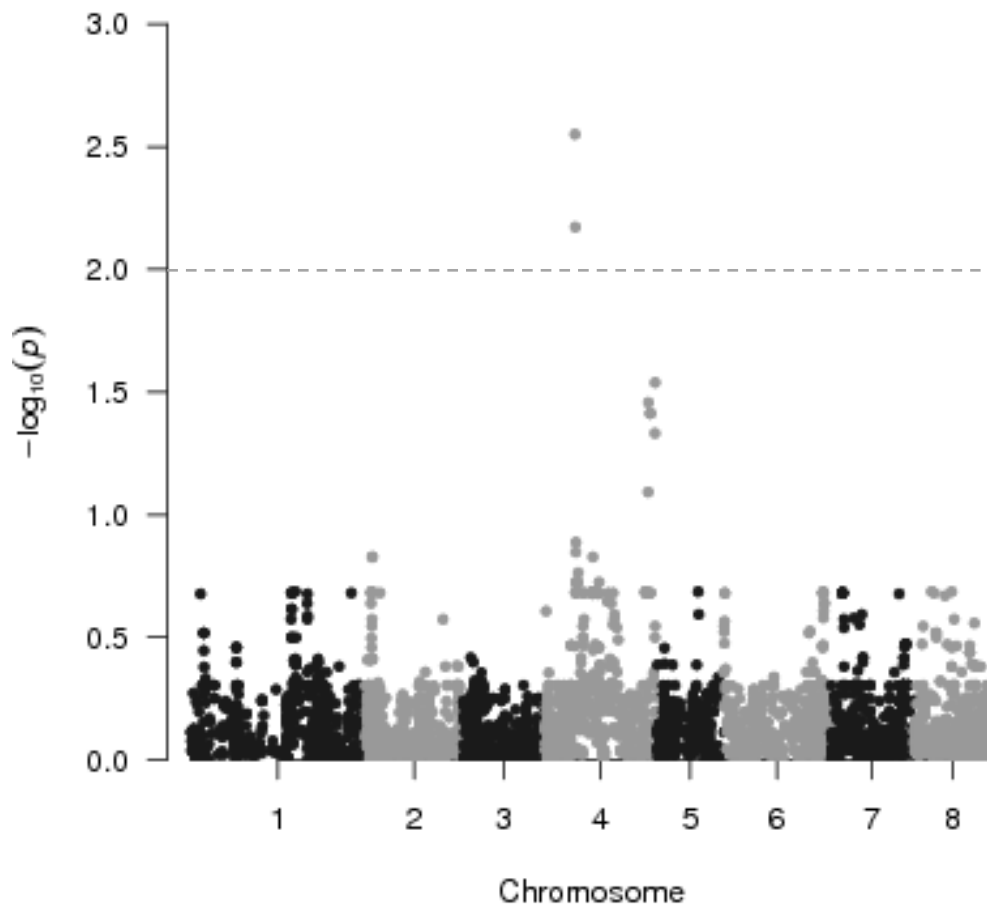


**Fig. 20** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere acidità titolabile per l'anno 2014. L'analisi è stata effettuata utilizzando il MLM e i p-value sono stati corretti in base al FDR, ad un livello di significatività di  $\alpha = 0.01$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate  $-\log_{10} p$ -value.

#### 3.5.2.4 Contenuto in solidi solubili (Brix)

Relativamente al contenuto in solidi solubili, sono stati individuati due marcatori associati a questo carattere per l'anno 2012 (SNP\_IGA\_401886 e lo SNP\_IGA\_402081) che coprono una regione ristretta compresa tra 8 222 657 e 8 308 422 (Tab. 12) e con valori di significatività pari a 2.82E-03 e 6.72E-03, rispettivamente (Fig. 21).

Non sono stati individuati marcatori associati a tale carattere per gli anni 2013 e 2014.



**Fig. 21** Manhattan plot che mostra i risultati ottenuti per l'analisi di associazione sul carattere contenuto in solidi solubili per l'anno 2012. L'analisi è stata effettuata utilizzando il MLM e i p-value sono stati corretti in base al FDR, ad un livello di significatività di  $\alpha = 0.01$ . Sull'asse delle ascisse sono indicati i cromosomi, sull'asse delle ordinate  $-\log_{10}$  p-value.

Carattere quantitativo	Anno	Marcatore	Crom	Posizione (Peach v1.0)	p-value	p-value corretto con FDR	R <sup>2</sup>
Epoca di fioritura	2012	SNP_IGA_135188	1	46 542 316	3.63E-06	5.49E-03	0.19
		SNP_IGA_748301	7	7 681 738	3.87E-06	5.49E-03	0.17
		SNP_IGA_751897	7	8 151 156	2.35E-06	5.49E-03	0.17
		SNP_IGA_778535	7	15 904 306	8.92E-06	9.49E-03	0.15
	2013	SNP_IGA_748301	7	7 681 738	6.23E-07	2.65E-03	0.20
Epoca di maturazione	2012	SNP_IGA_157433	2	2 094 415	7.70E-05	9.74E-03	0.13
		SNP_IGA_157556	2	2 119 637	6.32E-05	8.96E-03	0.13
		SNP_IGA_157983	2	2 182 219	4.10E-05	7.51E-03	0.14
		SNP_IGA_158011	2	2 186 582	4.20E-05	7.51E-03	0.14
		SNP_IGA_401886	4	8 222 657	3.99E-05	7.51E-03	0.14
		SNP_IGA_402416	4	8 538 214	8.03E-07	1.14E-03	0.19
		SNP_IGA_404165	4	9 239 957	6.21E-05	8.96E-03	0.13
		SNP_IGA_404570	4	9 396 554	6.77E-05	9.29E-03	0.13
		SNP_IGA_410134	4	10 615 885	2.72E-06	2.81E-03	0.18
		SNP_IGA_410265	4	10 638 625	7.77E-06	2.81E-03	0.16
		SNP_IGA_410336	4	10 665 019	1.91E-05	4.28E-03	0.15
		SNP_IGA_410398	4	10 685 500	4.51E-06	2.81E-03	0.17
		SNP_IGA_415301	4	12 512 099	2.04E-07	6.40E-04	0.21
		SNP_IGA_415382	4	12 535 151	7.77E-05	9.74E-03	0.13
		SNP_IGA_415799	4	12 715 296	3.01E-07	6.40E-04	0.21
		SNP_IGA_415813	4	12 719 686	3.93E-06	2.81E-03	0.17
		SNP_IGA_416194	4	12 770 981	7.92E-06	2.81E-03	0.16
		SNP_IGA_417310	4	13 023 877	1.73E-05	4.09E-03	0.15
		SNP_IGA_417637	4	13 067 436	9.46E-06	2.88E-03	0.16
		SNP_IGA_417666	4	13 081 053	4.61E-05	7.51E-03	0.13
		SNP_IGA_417715	4	13 097 715	1.36E-05	3.62E-03	0.15
		SNP_IGA_417840	4	13 112 264	1.24E-05	3.52E-03	0.15
		SNP_IGA_443304	4	18 742 047	4.94E-05	7.51E-03	0.13
		SNP_IGA_444204	4	18 861 832	4.72E-05	7.51E-03	0.13
		SNP_IGA_444291	4	18 881 545	4.52E-05	7.51E-03	0.13
		SNP_IGA_449112	4	19 905 501	3.35E-05	7.13E-03	0.14
		SNP_IGA_604703	6	45 035	7.58E-06	2.81E-03	0.16
		SNP_6_56169	6	56 169	7.88E-06	2.81E-03	0.16
		SNP_IGA_604834	6	108 105	5.75E-06	2.81E-03	0.16
		SNP_IGA_604859	6	114 296	6.27E-06	2.81E-03	0.16
		SNP_IGA_605045	6	203 384	8.69E-06	2.84E-03	0.16
		SNP_IGA_628424	6	7 833 639	4.86E-05	7.51E-03	0.13
	SNP_IGA_629558	6	8 000 787	8.01E-05	9.74E-03	0.13	
	SNP_IGA_630266	6	8 232 664	7.85E-05	9.74E-03	0.13	
	SNP_IGA_630302	6	8 238 299	1.48E-05	3.70E-03	0.15	
	2013	SNP_IGA_410134	4	10 615 885	3.95E-06	4.20E-03	0.28
		SNP_IGA_410265	4	10 638 625	2.16E-06	4.20E-03	0.30
		SNP_IGA_410336	4	10 665 019	2.81E-06	4.20E-03	0.29
		SNP_IGA_410398	4	10 685 500	3.91E-06	4.20E-03	0.28
	2014	SNP_IGA_410134	4	10 615 885	1.75E-05	7.87E-03	0.24
		SNP_IGA_410265	4	10 638 625	1.49E-05	7.87E-03	0.24
		SNP_IGA_410336	4	10 665 019	1.85E-05	7.87E-03	0.24
		SNP_IGA_410398	4	10 685 500	2.39E-05	9.24E-03	0.23
SNP_IGA_604703		6	45 035	1.22E-05	7.87E-03	0.25	
snp_6_56169		6	56 169	1.07E-05	7.87E-03	0.25	
SNP_IGA_604834		6	108 105	1.33E-05	7.87E-03	0.24	
SNP_IGA_604859		6	114 296	1.21E-05	7.87E-03	0.25	
SNP_IGA_605045	6	203 384	8.70E-06	7.87E-03	0.25		

		SNP_IGA_610487	6	2 028 395	2.35E-06	5.00E-03	0.29
		SNP_IGA_610531	6	2 044 947	2.35E-06	5.00E-03	0.29
Acidità	2012	SNP_IGA_544428	5	557 503	1.47E-05	6.25E-03	0.17
		SNP_IGA_544495	5	610 568	2.65E-07	3.28E-04	0.24
		SNP_IGA_544640	5	629 641	3.31E-07	3.28E-04	0.24
		SNP_IGA_544657	5	635 221	3.86E-07	3.28E-04	0.23
		SNP_IGA_544961	5	698 214	2.56E-06	1.82E-03	0.20
		SNP_IGA_545261	5	821 372	1.24E-05	5.86E-03	0.17
		SNP_IGA_546094	5	987 702	3.68E-07	3.28E-04	0.23
		SNP_IGA_546467	5	1 075 612	3.47E-06	1.85E-03	0.20
		SNP_IGA_546987	5	1 166 305	1.87E-05	6.88E-03	0.17
		SNP_IGA_548037	5	1 376 489	1.94E-05	6.88E-03	0.17
		SNP_IGA_548512	5	1 503 401	7.72E-08	3.28E-04	0.26
		SNP_IGA_548597	5	1 518 380	3.44E-06	1.85E-03	0.20
	2014	SNP_IGA_544428	5	557 503	2.56E-07	1.56E-04	0.39
		SNP_IGA_544495	5	610 568	1.29E-07	1.32E-04	0.41
		SNP_IGA_544640	5	629 641	1.09E-07	1.32E-04	0.41
		SNP_IGA_544657	5	635 221	5.07E-07	2.70E-04	0.37
		SNP_IGA_544961	5	698 214	2.94E-08	1.25E-04	0.45
		SNP_IGA_545261	5	821 372	1.13E-05	3.43E-03	0.28
		SNP_IGA_546094	5	987 702	1.75E-07	1.32E-04	0.40
		SNP_IGA_546467	5	1 075 612	1.86E-07	1.32E-04	0.40
		SNP_IGA_546765	5	1 141 551	3.61E-05	8.08E-03	0.25
		SNP_IGA_546791	5	1 147 084	3.61E-05	8.08E-03	0.25
		SNP_IGA_546835	5	1 155 466	1.37E-05	3.88E-03	0.27
		SNP_IGA_546987	5	1 166 305	4.67E-06	1.81E-03	0.30
		SNP_IGA_547017	5	1 168 341	7.53E-06	2.46E-03	0.29
		SNP_IGA_547211	5	1 190 216	1.58E-05	3.95E-03	0.27
		SNP_IGA_547473	5	1 216 777	4.58E-06	1.81E-03	0.30
		SNP_IGA_547510	5	1 219 848	4.58E-06	1.81E-03	0.30
		SNP_IGA_548037	5	1 376 489	1.46E-05	3.88E-03	0.27
		SNP_IGA_548512	5	1 503 401	1.75E-07	1.32E-04	0.40
		SNP_IGA_548597	5	1 518 380	5.66E-06	2.01E-03	0.30
		SSC (Brix)	2012	SNP_IGA_401886	4	8 222 657	6.62E-07
SNP_IGA_402081	4			8 308 422	3.16E-06	6.72E-03	0.20

**Tab. 12** Schema riassuntivo dei marcatori individuati in associazione ai caratteri quantitativi. L'alternanza grigio-bianco differenzia le regioni genomiche individuate in associazione a un dato carattere.

Per quanto riguarda i caratteri sovracalore e peso del frutto, non stati individuati marcatori significativamente associati a questi caratteri per nessuno dei tre anni di osservazione.

## 4. DISCUSSIONE

### 4.1 GENOTIPIZZAZIONE CON MARCATORI SSR

È stata effettuata un'analisi della variabilità all'interno di una collezione di 877 accessioni di *Prunus persica* e una accessione di *Prunus davidiana*, utilizzando 14 marcatori SSR.

I genotipi di *P. persica* comprendevano una porzione rilevante della diversità genetica di pesco e includevano accessioni orientali e occidentali, cultivar moderne e *landraces*.

In totale, i 14 marcatori SSR utilizzati hanno prodotto 185 alleli, di cui 34 sono alleli unici, cioè presenti in una sola accessione. Otto alleli unici appartengono a *P. davidiana* e questo è atteso, in quanto si tratta di una specie diversa da *P. persica*, mentre per quanto riguarda gli alleli unici appartenenti a *P. persica*, si tratta di accessioni orientali, selezioni locali e, frequentemente, l'accessione 'Citation'. L'analisi del pedigree ha rivelato che questa accessione è frutto di un incrocio complesso che comprende anche il susino e questo spiega la maggiore variabilità genetica.

Per tutti i loci analizzati l'eterozigosi osservata ( $H_o$ ) è inferiore a quella attesa ( $H_e$ ) e l'indice di fissazione ( $F_{is}$ ) è positivo. Ciò indica una deviazione dall'equilibrio di Hardy-Weinberg, dovuta a un eccesso di omozigoti, come atteso in una specie prevalentemente autogama come il pesco.

Il numero medio di alleli per locus è di 13.21, l'eterozigosi media osservata è di 0.46, quella media attesa è di 0.64. Il 73% degli alleli (pari a 135 sul totale degli alleli osservati) presenta una frequenza inferiore al 5%. Questi risultati sono in accordo con quanto osservato da Li et al., (2013), i quali hanno analizzato 653 accessioni di pesco di origine cinese, nord-americana ed europea con 48 marcatori SSR. In questo studio, i 48 primer utilizzati hanno prodotto 588 alleli, con una media di 12.25 alleli per locus e con il 73.9% di alleli rari, cioè presenti con una frequenza minore del 5%; l'eterozigosi osservata era di 0.47, mentre quella attesa era di 0.60.

In uno studio effettuato su 224 accessioni occidentali di pesco (spagnole e americane) con 50 marcatori SSR (Aranzana et al., 2010), è stata ottenuta un'eterozigosi media osservata di 0.34, mentre quella attesa era di 0.46. Questi valori sono più bassi rispetto a quelli ottenuti nella nostra analisi e ciò potrebbe essere dovuto al fatto che nello studio di Aranzana et al. (2010) non erano compresi genotipi orientali.

L'analisi della variabilità ha raggruppato le accessioni in 4 cluster principali, divisi in base alla loro origine e alla loro storia di domesticazione. Il primo cluster comprende nettarine

di origine occidentale. La suddivisione tra pesche e nettarine è stata osservata anche da Aranzana et al. (2010) e da Li et al., (2013). La spiegazione di questa separazione risiede presumibilmente nella storia della costituzione delle cultivar dei due tipi. Fino a qualche decennio fa, infatti, nuove cultivar di pesco sono state ottenute a partire da genitori pesco, così come nuove cultivar di nettarine sono state ottenute utilizzando parentali di tipologia nettarina.

Il secondo cluster è molto ampio e comprende accessioni antiche, accessioni moderne (frutto del breeding) e numerose accessioni non identificabili in questi due gruppi. In uno studio condotto su 1240 accessioni di pesco di origine orientale e occidentale, genotipizzate con 4271 marcatori SNP (Micheletti et al., sottomesso), l'analisi di struttura ha rivelato che la popolazione iniziale era suddivisa in tre sottogruppi: accessioni orientali, accessioni occidentali antiche e accessioni occidentali moderne sviluppate dal breeding. Circa la metà delle accessioni (il 54%) formano un gruppo con una struttura mista e questo suggerisce, come indicano gli autori, che spesso nei programmi di miglioramento genetico, al fine di sfruttare la variabilità, sono stati incrociati individui appartenenti a diverse sottopopolazioni, con il conseguente ottenimento di accessioni a struttura mista.

Il terzo cluster include soprattutto accessioni antiche. Anche nel già citato studio di Micheletti et al., (sottomesso), le accessioni tradizionali si separano chiaramente dalle accessioni ottenute mediante il miglioramento genetico moderno.

Il quarto cluster, infine, comprende le accessioni orientali, più le accessioni 'Sicilia 1', 'Sicilia 2', 'Tabacchiera', 'Platicarpa a', 'Platicarpa b', 'Platicarpa c' e 'Platicarpa pirovano'. Queste sono tutte pesche piatte, diffuse in Sicilia da diversi secoli e presumibilmente derivano da accessioni orientali a forma piatta introdotte in Sicilia nei secoli scorsi e questo spiega perché sono raggruppate insieme. La stessa cosa accade per una varietà tradizionale piatta spagnola ('Paraguay Delfin') che nell'analisi di Li et al. (2013) si raggruppa nello stesso cluster delle piatte cinesi, indicando un'origine comune tra questi gruppi. Le altre cultivar a forma piatta, quali alcune della serie 'Ufo', invece, derivano da 'Stark Saturn' che è una accessione piatta derivante dal miglioramento genetico moderno (Okie, 1998). Queste si raggruppano nel cluster 2 e sono ben separate dalle accessioni della serie 'Sicilia', 'Platicarpa' e 'Tabacchiera'.

## 4.2 FENOTIPIZZAZIONE DELLA CORE COLLECTION

Tutti i caratteri quantitativi analizzati per tre anni nelle 165 accessioni di pesco, hanno mostrato un'ampia variabilità fenotipica, come si può osservare dalla tabella 9 nella sezione risultati.

In tutti e tre gli anni, ed in particolare nel 2013 e 2014, durante la stagione di raccolta del pesco, a partire dall'ultima decade di luglio, ci sono stati forti attacchi di *Ceratitis capitata*, un insetto polifago e di difficile controllo. Le piante colpite da questo insetto presentavano frutti danneggiati, e per alcune accessioni non è stato possibile effettuare i rilievi relativi alla qualità del frutto, come il sovracoloro, il peso del frutto, l'acidità titolabile e il contenuto in solidi solubili.

La correlazione tra i vari anni di osservazione è comunque stata medio-alta per tutti i caratteri.

## 4.3 ANALISI DELLA STRUTTURA DI POPOLAZIONE

Lo studio della struttura effettuato sulle 165 accessioni di pesco mediante l'analisi delle componenti principali (PCoA) e il software STRUCTURE, ha consentito di individuare tre sottogruppi all'interno della popolazione di partenza: le accessioni occidentali tradizionali (38 accessioni), le accessioni occidentali moderne ottenute mediante il miglioramento genetico (29 accessioni) e le accessioni orientali (13 accessioni), che includono anche le accessioni 'Tabacchiera' e 'Sicilia 1'. Analogamente a quanto osservato in seguito alla genotipizzazione con marcatori SSR, queste due accessioni locali siciliane a forma piatta si raggruppano insieme alle accessioni orientali, confermando ulteriormente l'ipotesi di un'origine comune con queste ultime. Le rimanenti 85 accessioni (il 52%) non sono state assegnate a nessuna sottopopolazione e sono state considerate accessioni a struttura mista (admixed).

Risultati analoghi sono stati ottenuti da Micheletti et al. (sottomesso), analizzando una popolazione molto più ampia di 1240 accessioni orientali e occidentali di *Prunus* che includevano anche le 165 accessioni della nostra popolazione. Anche in questo lavoro, la collezione di partenza è risultata suddivisa in tre sottogruppi che si differenziavano sulla base della loro origine e della loro storia di domesticazione, mentre le accessioni admixed erano il 54%, un valore simile a quanto ottenuto in questo studio.

Come suggerito anche da Micheletti et al. (sottomesso), la struttura individuata nella nostra popolazione di 165 accessioni corrisponde ai colli di bottiglia (che hanno portato ad una riduzione della variabilità genetica) descritti da Verde et al. (2013) analizzando 12 accessioni occidentali e orientali di pesco con circa un milione di SNP. Il primo collo di bottiglia è rappresentato dal processo di domesticazione del pesco, che si ritiene abbia avuto luogo in Cina circa 4000-5000 anni fa e corrisponde al sottogruppo delle accessioni orientali. Il secondo collo di bottiglia è legato alla diffusione del pesco dalla Cina in Europa e ai programmi di miglioramento genetico iniziati in Europa e negli Stati Uniti circa un secolo fa, che hanno portato all'ottenimento delle moderne cultivar commerciali; questo secondo collo di bottiglia è rappresentato dal sottogruppo delle accessioni occidentali antiche e da quello delle accessioni occidentali moderne.

Precedenti studi avevano comunque dimostrato che il pesco presenta una forte struttura di popolazione. Aranzana et al. (2010) hanno analizzato 224 cultivar commerciali di pesco (di origine spagnola e americana) con 50 marcatori SSR e, in seguito all'analisi di struttura, hanno osservato che la loro popolazione si suddivideva in tre sottogruppi: pesche a polpa fondente, nettarine a polpa fondente e pesche e nettarine a polpa soda. Li et al. (2013), analizzando con marcatori SSR una popolazione più ampia di 653 accessioni di pesco di origine orientale e occidentale, hanno osservato una stratificazione della loro popolazione in tre sottogruppi: le accessioni occidentali, le accessioni orientali e le *landraces*. La sottopopolazione delle accessioni occidentali era ulteriormente suddivisa al suo interno in pesche e nettarine. Un altro studio che ha confermato l'esistenza di una forte struttura di popolazione in pesco è stato effettuato da Font i Forcada et al. (2013), che hanno analizzato 94 accessioni di pesco (comprendenti accessioni tradizionali spagnole e cultivar moderne occidentali) con marcatori SSR. Anche in questo caso, l'analisi di struttura ha rivelato la suddivisione della popolazione in due sottogruppi, quello delle accessioni tradizionali e quello delle cultivar moderne.

#### **4.4 ANALISI DI ASSOCIAZIONE**

Per effettuare le analisi di associazione solitamente vengono utilizzati due modelli: il Mixed Linear Model (MLM) o il General Linear Model (GLM). Il MLM è più stringente rispetto al GLM, in quanto tiene conto sia della struttura di popolazione che delle relazioni di parentela tra le accessioni per correggere gli errori di tipo I, derivanti cioè da associazioni spurie. Nel GLM, invece, viene utilizzata solo la struttura di popolazione e di

conseguenza alcuni marcatori significativamente associati individuati mediante questo modello potrebbero essere falsi positivi poiché non considera le relazioni di parentela tra le accessioni (Zhang et al., 2014). Proprio per questo motivo, è fondamentale stabilire in modo accurato il livello di significatività da adottare, in base al modello impiegato per effettuare l'analisi di associazione. Se si utilizza un valore soglia troppo basso si possono perdere alcune associazioni significative (errore di tipo II), mentre scegliendo come valore soglia un p-value più elevato è più facile ottenere associazioni spurie (errore di tipo I) (Zhang et al., 2014).

Nella nostra popolazione, l'analisi di associazione, inizialmente, è stata effettuata mediante il Mixed Linear Model (MLM) implementato nel software TASSEL, sia per i caratteri qualitativi che per i caratteri quantitativi; tuttavia per tutti i caratteri qualitativi, ad eccezione del colore della polpa, non erano stati individuati marcatori associati. Si è deciso quindi di provare ad effettuare l'analisi di associazione dei caratteri qualitativi mediante il GLM. Poiché questo modello, come detto prima, è meno stringente rispetto al MLM è stata applicata la correzione di Bonferroni ad un livello di significatività di  $\alpha=0.0001$ . Per l'analisi di associazione dei caratteri quantitativi, invece, il MLM ha dato buoni risultati e per individuare i marcatori significativamente associati ai caratteri analizzati, i p-value sono stati corretti mediante il False Discovery Rate (FDR), ad un livello di significatività  $\alpha=0.01$ .

#### **4.4.1 Caratteri qualitativi**

##### 4.4.1.1 Tomentosità del frutto

La tomentosità del frutto è un carattere monogenico (*G*), con il fenotipo nettarina recessivo rispetto al frutto tomentoso (pesco). Il locus *G* era stato mappato nella parte finale del cromosoma 5, in una regione di 1.189 Mb (Dirlewanger et al., 2006; Le Dantec et al., 2010). Recentemente è stato isolato il gene *PpeMYB25* (ppa023143; scaffold\_5:15897836..15899262) responsabile del carattere pesco/nettarina sul cromosoma 5 ed è stato osservato che il fenotipo recessivo nettarina è causato dall'inserzione di un retrotrasposone nell'esone 3 di questo gene (Vendramin et al., 2014).

Nel presente lavoro, in seguito all'analisi di associazione, il marcatore SNP maggiormente associato al carattere tomentosità del frutto è lo SNP\_IGA\_603047 localizzato sul cromosoma 5 in posizione 16 774 236, e quindi si trova nella regione in cui è stato isolato

il gene responsabile di questo carattere. Lo stesso SNP è stato individuato da Micheletti et al. (sottomesso).

Per quanto riguarda i due marcatori significativamente associati localizzati sul cromosoma 6, si tratta probabilmente di associazioni spurie, infatti nello studio di Micheletti et al (sottomesso) in cui l'analisi di associazione è stata effettuata con un numero molto più elevato di accessioni (1240) che includono anche quelle analizzate nel presente lavoro, questi marcatori non sono stati individuati.

#### 4.4.1.2 Colore della polpa

Il colore della polpa è un carattere controllato da un solo locus (*Y*), localizzato sul cromosoma 1 e il colore bianco è dominante sul giallo (Bliss et al., 2002). Recentemente il locus *Y* è stato mappato finemente in un intervallo di circa 500 kb (scaffold\_1: 25 584 537 – 26 004 830) (Martinez-Garcia et al., 2013a). Successivamente, Falchi et al., (2013) hanno isolato il gene *PpCCD4* (ppa006109; scaffold\_1: 25 639 445..25 641 500), responsabile del carattere colore della polpa e hanno dimostrato che tre mutazioni indipendenti a carico di questo gene determinano il colore giallo: una mutazione in un microsatellite sul primo esone del gene, una in uno SNP sul secondo esone e la terza è data dall'inserzione di un retrotrasposone.

Nella nostra popolazione di 165 accessioni di pesco, il mappaggio per associazione ha portato ad individuare un marcatore associato al carattere colore della polpa localizzato sul cromosoma 1, lo SNP\_IGA\_90213, in posizione 26 479 525. Anche per questo carattere, lo SNP associato si trova nella stessa regione in cui è stato clonato il gene causale. Questo SNP, inoltre, è stato individuato anche da Micheletti et al. (sottomesso) come lo SNP maggiormente associato al carattere colore della polpa.

#### 4.4.1.3 Tipo di fiore

In pesco, il fiore può essere rosaceo o campanulaceo; è un carattere monogenico (*Sh*) e il tipo di fiore campanulaceo è dominante sul rosaceo. Ogundiwin et al. (2009), utilizzando una popolazione derivata da un incrocio intraspecifico tra due cultivar di pesco, hanno mappato questo carattere sul cromosoma 8, a 5.4 cM dal marcatore SSR CPPCT006, localizzato a 13 659 021. Successivamente, Fan et al (2010) hanno ristretto l'intervallo a 1.4 cM dallo stesso marcatore. In questa stessa regione, in seguito all'analisi di

associazione effettuata nel presente lavoro, è stato individuato un marcatore associato al carattere tipo di fiore: lo SNP\_IGA\_864149 a 13 756 987. Anche Micheletti et al. (sottomesso) individuano lo stesso marcatore associato a questo carattere.

#### 4.4.1.4 Forma delle ghiandole fogliari

Il locus *E*, che controlla la forma delle ghiandole fogliari, è stato mappato sul cromosoma 7 (Dettori et al., 2001). Le ghiandole fogliari possono essere globose o reniformi e il carattere reniforme è dominante sul globoso. Verde et al. (2005) hanno individuato un marcatore SSR (pchgms44; scaffold\_7:15 145 564) strettamente associato, a 2 cM dal locus.

Dei 5 marcatori SNP individuati nel presente lavoro, i 3 maggiormente associati alla forma delle ghiandole fogliari, sono localizzati sul cromosoma 7, in una regione corrispondente a quanto già osservato mediante il mappaggio con una popolazione bi-parentale.

Nella stessa regione, Micheletti et al (sottomesso) hanno individuato uno SNP associato a questo carattere (SNP\_IGA\_776161, scaffold\_7:14 870 521).

### **4.4.2 Caratteri quantitativi**

#### 4.4.2.1 Epoca di fioritura

Nel 2012 è stato individuato uno SNP associato all'epoca di fioritura, localizzato sul cromosoma 1 in posizione 46 542 316. In questa stessa regione è stato mappato un *major* QTL che controlla l'epoca di fioritura in uno studio condotto utilizzando una popolazione bi-parentale (Fan et al., 2010). Questa regione, inoltre, coincide con quella in cui è stato mappato il locus *Evergrowing* (*EVG*) (Wang et al., 2002; Bielenberg et al., 2008).

Nel 2012, inoltre, sono stati individuati altri tre marcatori associati all'epoca di fioritura, posizionati sul cromosoma 7: 2 (SNP\_IGA\_748301 e SNP\_IGA\_751897) in una regione compresa tra 7 681 738 e 8 151 156 e uno (SNP\_IGA\_778535), in posizione 15 904 306. Lo SNP\_IGA\_748301, localizzato sul cromosoma 7, in posizione 7 681 738, è stato individuato significativamente associato a questo carattere anche nel 2013.

Fan et al., (2010) hanno mappato un *major* QTL relativo all'epoca di fioritura (che spiegava più del 30% della varianza fenotipica) sul cromosoma 7, a 5.2 cM dal marcatore CPPCT033 (scaffold\_7:16 702 195). Nella stessa regione, anche Eduardo et al. (2011),

mediante il mappaggio con una popolazione bi-parentale, hanno individuato un QTL che controlla l'epoca di fioritura a 2 cM dal marcatore CPPCT033.

#### 4.4.2.2 Epoca di maturazione

Sono stati osservati marcatori associati all'epoca di maturazione sui cromosomi 2, 4 e 6. Sul cromosoma 2 sono stati individuati 4 SNP, solo nell'anno 2012, e sono localizzati in una piccola regione, da 2 094 415 a 2 186 582. In altri lavori sono stati individuati marcatori associati all'epoca di maturazione sul cromosoma 2, ma localizzati in un'altra regione rispetto a quella individuata nel nostro studio. Cao et al. (2012) hanno individuato un marcatore SSR associato a questo carattere (UDP98-411;scaffold\_2:20 1727 01) mediante uno studio di mappaggio per associazione, effettuato su 104 *landraces* di pesco orientali con 53 marcatori SSR. In questa stessa regione del cromosoma 2, Verde et al., (2002), utilizzando una popolazione BC<sub>1</sub>, hanno mappato un QTL che controlla l'epoca di maturazione, stabile in 5 anni di osservazione.

I marcatori localizzati sul cromosoma 4 coprono una regione di 4.9 Mb, da 8 222 657 a 13 112 264, nell'anno 2012. Negli anni 2013 e 2014, sono stati individuati 4 SNP localizzati in una piccola regione, da 10 615 885 a 10 685 500, che ricade all'interno della regione più ampia di 4.9 Mb, osservata nel 2012. Quilot et al. (2004), utilizzando un incrocio interspecifico, hanno individuato un QTL che controlla l'epoca di maturazione sul cromosoma 4, in prossimità del marcatore UDP96-003 (scaffold\_4: 8 757 450). Nella stessa regione Etienne et al. (2002b), utilizzando un incrocio intraspecifico di pesco, hanno identificato un QTL, in prossimità del marcatore UDP97-402 (10 486 180). Utilizzando la stessa popolazione, Dirlwanger et al., (2012) hanno mappato un QTL ad effetto principale relativo al carattere, localizzato sul cromosoma 4 tra i marcatori SSR UDP97-402 (10 486 180) e AMPA103 (13 509 355). In questa stessa regione, Eduardo et al. (2011) hanno mappato il locus MD (Maturity Date) e successivamente Pirona et al (2013) hanno individuato un gene candidato (un fattore di trascrizione NAC; ppa008301m, scaffold\_4:11 105 983..11 107 728) che controlla l'epoca di maturazione e hanno osservato che un'inserzione di 9 pb nell'ultimo esone di questo gene, potrebbe essere la causa dei fenotipi a maturazione precoce. Sempre Edoardo et al. (2011) hanno individuato un altro QTL che controlla l'epoca di maturazione, associato al marcatore M12a (scaffold\_4: 9 208 608) sul cromosoma 4. Font I Forcada et al. (2013), invece, hanno effettuato uno studio di mappaggio per associazione in una popolazione di 94 accessioni di pesco, utilizzando 40

marcatori SSR e hanno individuato due marcatori strettamente associati a questo carattere, il BPPCT015 e il CPPCT028 (a 2 086 534), localizzati sul cromosoma 4 in un'altra regione rispetto a quelle individuate nel presente lavoro.

Solo per l'anno 2012 è stata individuata un'altra regione significativamente associata al carattere sul cromosoma 4, compresa tra 18 742 047 e 19 905 501. In base alle nostre conoscenze, ad oggi in questa regione non stati individuati QTL per il carattere.

Negli anni 2012 e 2014 sono stati individuati marcatori associati al carattere all'inizio del cromosoma 6, in una regione compresa tra 45 035 e 203 384. Nel 2014, questa regione si estende fino a 2 044 947. In questa regione, nella versione Peach v1.0, sono stati individuati degli errori di assemblaggio che sono stati recentemente corretti nella nuova versione Peach v2.0 (Verde et al., 2015). I 7 SNP associati al carattere, nella nuova versione, ricadono in una regione di 933 kb, anziché di 2 Mb, compresa tra 3 163 668 e 4 096 357. In tabella 13 sono riportate le posizioni di questi SNP nelle due versioni del genoma.

Sul cromosoma 6 Verde et al., (2012) hanno mappato un QTL che controlla l'epoca di maturazione localizzato vicino al marcatore UDP98-416 (in posizione 425 587 nella versione Peach v1.0).

Solo per il 2012, inoltre, è stata individuata un'altra regione sul cromosoma 6, compresa tra 7 833 639 e 8 238 299. In base alle nostre conoscenze, ad oggi in questa regione non stati individuati QTL per il carattere.

<b>Marcatore</b>	<b>Cromo- soma</b>	<b>Posizione (Peach v1.0)</b>	<b>Posizione (Peach v2.0)</b>
SNP_IGA_610531	6	2 044 947	3 163 668
SNP_IGA_610487	6	2 028 395	3 180 220
SNP_IGA_604703	6	45 035	3 937 990
snp_6_56169	6	56 169	3 949 124
SNP_IGA_604834	6	108 105	4 001 078
SNP_IGA_604859	6	114 296	4 007 269
SNP_IGA_605045	6	203 384	4 096 357

**Tab. 13** Posizioni dei 7 SNP associati alla data di maturazione, localizzati all'inizio del cromosoma 6, nelle due versioni del genoma.

#### 4.4.2.3 Acidità titolabile

In pesco, il carattere acido/sub-acido è stato mappato finemente come carattere monogenico (locus *D*, allele sub-acido dominante) all'inizio del cromosoma 5, utilizzando

popolazioni bi-parentali (Dirlewanger et al., 1998; Boudehri et al., 2009). Micheletti et al. (sottomesso) hanno mappato questo carattere, utilizzando un approccio di mappaggio per associazione, e hanno individuato 21 marcatori SNP associati ad esso, che coprono una regione di 1.8 Mb (scaffold\_5: 467 067 - 2 270 122) all'inizio del cromosoma 5. Nel loro studio, il carattere acido/sub-acido è stato considerato monogenico, suddividendo le accessioni in acide (con acidità titolabile maggiore di 8.2 meq/100ml) e sub-acide.

Nella nostra analisi, il carattere acidità è stato considerato come carattere quantitativo e per ogni accessione è stato considerato il valore dell'acidità titolabile (espresso in meq/100 ml). Sono stati individuati marcatori SNP associati al carattere che coprono una regione di circa 1 Mb, all'inizio del cromosoma 5 (scaffold\_5: 557 503 - 1 518 380). Mediante un'analisi effettuata con una popolazione bi-parentale, Dirlewanger et al., (1999) hanno individuato un QTL ad effetto principale, stabile per due anni, che controlla l'acidità titolabile, localizzato ad 8 cM dall'inizio del cromosoma 5. Questo QTL spiegava l'87% della varianza fenotipica nel primo anno e il 78 % nel secondo anno.

#### 4.4.2.4 Contenuto in solidi solubili (Brix)

Per quanto riguarda il contenuto in solidi solubili, sono stati individuati marcatori associati a questo carattere solo nel 2012: lo SNP\_IGA\_401886, localizzato sul cromosoma 4 a 8 222 657 e lo SNP\_IGA\_402081 anch'esso localizzato sul cromosoma 4, in posizione 8 308 422. QTL che controllano il contenuto in solidi solubili sono stati mappati all'inizio del cromosoma 4, in prossimità del marcatore UDP97-402 (10 486 180) da Etienne et al. (2002b) e da Quilot et al. (2004). Anche Eduardo et al. (2011) hanno individuato sul cromosoma 4 un QTL stabile per due anni relativo a questo carattere, ma è localizzato nella parte finale del cromosoma.

I due marcatori SNP trovati in associazione al contenuto in solidi solubili si trovano nella stessa regione in cui è stato mappato il locus MD (Eduardo et al., 2011) ed uno di questi (lo SNP\_IGA\_401886 a 8 222 657) è stato individuato associato anche all'epoca di maturazione. Come suggerito da Edoardo et al. (2011), questo è dovuto all'effetto pleiotropico che il carattere epoca di maturazione ha su altri caratteri. In pesco, spesso, QTL che controllano l'epoca di maturazione sono localizzati nella stessa posizione di QTL relativi al contenuto in solidi solubili o all'acidità del frutto ((Dirlewanger et al. 1999; Quarta et al. 2000; Etienne et al. 2002b; Verde et al. 2002; Quilot et al. 2004). Ad esempio, Dirlewanger et al. (1999) hanno individuato un cluster che raggruppa QTL relativi

all'epoca di maturazione, al contenuto in saccarosio, al contenuto in fruttosio e all'acidità titolabile sul gruppo 4, vicino al marcatore BPPCT15, in una mappa ottenuta da un incrocio intraspecifico tra due cultivar di pesco. Nella stessa regione, Quilot et al. (2004) hanno mappato un cluster di QTL, associati all'epoca di maturazione e al contenuto di sorbitolo, fruttosio, maltosio e acido citrico.

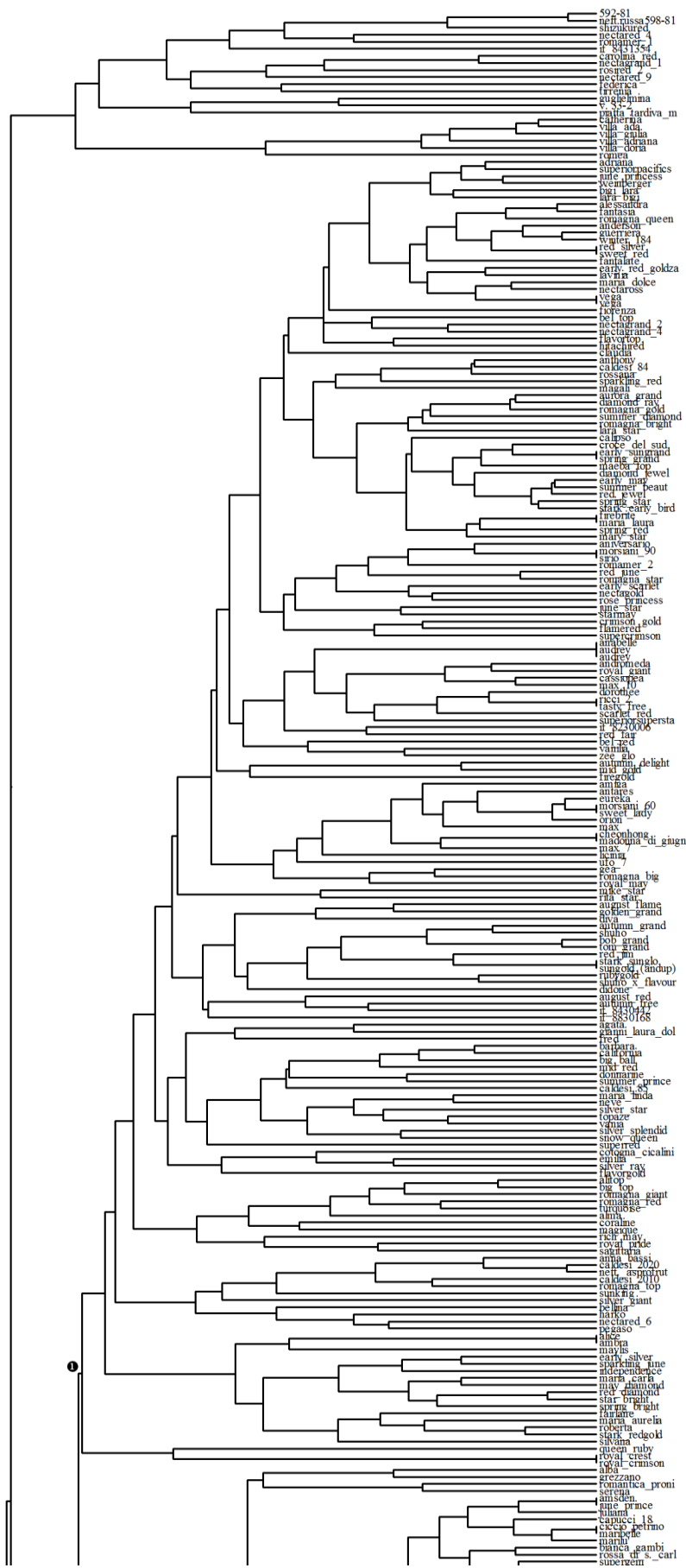
## 5. CONCLUSIONI

L'obiettivo del presente lavoro era quello di individuare marcatori strettamente associati a caratteri agronomici rilevanti in pesco. Associazioni significative sono state trovate, sia per i caratteri qualitativi che quantitativi, con marcatori e/o regioni già identificate in altri lavori utilizzando approcci e materiali vegetali diversi. Nel presente lavoro, inoltre, l'analisi di associazione ha consentito di identificare per alcuni caratteri QTL in regioni genomiche mai descritte prima. Questo risultato conferma la potenzialità del mappaggio per associazione, che a differenza del mappaggio con popolazioni bi-parentali, sfrutta tutti gli eventi di ricombinazione che si sono verificati durante la storia evolutiva della specie e di conseguenza il numero di QTL che possono essere mappati per un dato fenotipo non è limitato a quelli che segregano tra i parentali di un incrocio, ma è possibile individuare un numero molto più ampio di QTL che controllano il carattere.

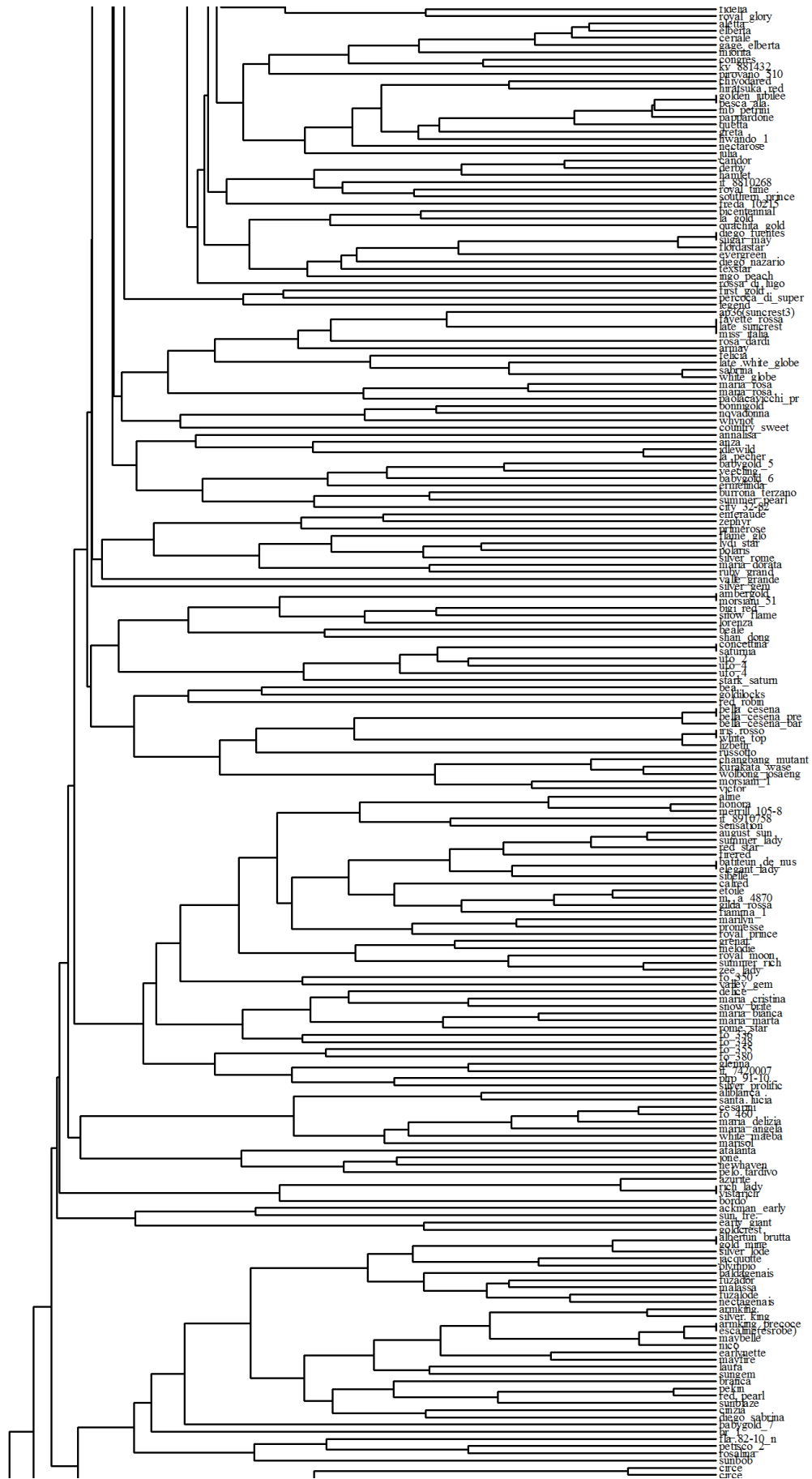
I marcatori individuati nel presente lavoro potranno essere utilizzati per la selezione genetica assistita da marcatori (MAS), che consente la selezione precoce dei semenzali con le caratteristiche desiderate con un importante risparmio di tempo, di spazio e di risorse economiche.

Dal presente studio inoltre è stata verificata l'importanza, ampiamente discussa in numerosi altri lavori, dei modelli statistici da utilizzare per la ricerca di associazioni significative e come non esista un modello statistico univoco in grado di gestire efficientemente gli errori di tipo I e II, ma come questo debba essere testato e verificato empiricamente secondo il carattere, il genotipo o la popolazione studiata.

Infine, i risultati ottenuti hanno dimostrato la fattibilità di studi di mappaggio per associazione nella core collection utilizzata. Quest'ultima rappresenta pertanto un'importante risorsa da poter utilizzare nel tempo per analizzare nuovi caratteri e individuare i marcatori ad essi associati.









## 6. BIBLIOGRAFIA

- Abbott AG**, Rajapakse S, Sosinski B, Lu ZX, Sossey-Alaoui K, Gannavarapu M, Reighard G, Ballard RE, Baird WV SR, Callahan A (1998) Construction of saturated linkage maps of peach crosses segregating for characters controlling fruit quality, tree architecture and pest resistance. *Acta Horti*, 465: 41–49.
- Aranzana MJ**, Pineda A, Cosson P, Dirlewanger E, Ascasibar J, Cipriani G, Ryder CD, Testolin R, Abbott A, King GJ, Iezzoni AF, Arús P (2003) A set of simple-sequence repeat (SSR) markers covering the *Prunus* genome. *Theor Appl Genet*, 106: 819–825.
- Aranzana MJ**, Kim S, Zhao K, Bakker E, Horton M, Jacob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics*, 1: e60.
- Aranzana MJ**, Abbassi EK, Howad W, Arús P (2010) Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. *BMC Genet*, 11: 69.
- Aranzana MJ**, Illa E, Howad W, Arús P (2012) A first insight into peach [*Prunus persica* (L.) Batsch] SNP variability. *Tree Genetics & Genomes*, 8: 1359-1369.
- Arús P**, Verde I, Sosinski B, Zhebentyayeva T, Abbott AG (2012) The peach genome. *Tree Genetics & Genomes*, 8: 531-547.
- Atwell S**, Huang YS, Vilhjalmsón BJ, Willems G, Horton M, et al. (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465: 627–631.
- Balding D** (2006) A tutorial on statistical methods for population association studies. *Nature Review Genetics*, 7: 781-791.
- Baránek M**, Raddová J, Pidra M (2006) Comparative analysis of genetic diversity in *Prunus* L. as revealed by RAPD and SSR markers. *Sci. Horti*. 108: 253–259.
- Bar-Hen A**, Charcosset A, Bourgoïn M, Guiard J (1995) Relationship between genetic markers and morphological traits in a maize inbred line collection. *Euphytica*, 84: 145-154.
- Barcaccia G**, **Falcinelli M** (2006) *Genetica e Genomica. Volume III Genomica e Biotecnologie genetiche*. Liguori Editore.

- Bassil NV**, Davis TM, Zhang H, Ficklin S, Mittmann M, Webster T, Mahoney L, Wood D, Alperin ES, Rosyara UR, Koehorst-vanc Putten H, Monfort A, Sargent DJ, Amaya I, Denoyes B, Bianco L, van Dijk T, Pirani A, Iezzoni A, Main D, Peace C, Yang Y, Whitaker V, Verma S, Bellon L, Brew F, Herrera R, van de Weg E (2015) Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry *Fragaria × ananassa*. *BMC Genomics*, DOI 10.1186/s12864-015-1310-1.
- Batley J**, Barker G, O’Sullivan H, Edwards KJ, Edwards D (2003) Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiol.*, 132: 84–91.
- Bartolozzi F**, Warburton ML, Arulsekar S, Gradziel TM (1998) Genetic characterization and relatedness among California almond cultivars and breeding lines detected by randomly amplified polymorphic DNA (RAPD) analysis. *J. Am. Soc. Hort. Sci.*, 123: 381-387.
- Beer S**, Siripoonwivat W, O’donoghue L, Souza E, Mathews D, Sorrels M (1997) Associations between molecular markers and quantitative traits in oat germplasm pool: can we infer linkages? *Journal of Agricultural Genomics*, Vol. 3.
- Begum H**, Spindel JE, Lalusin A, Borromeo T, Gregorio G, Hernandez J, Virk P, Collard B, McCouch SR (2015) Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). *PLoS ONE* 10(3): e0119873.
- Benjamini Y & Hochberg Y** (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol*, 57: 289–300.
- Bianco L**, Cestaro A, Sargent DJ, Banchi E, Derdak S, Di Guardo M, Salvi S, Jansen J, Viola R, Gut I, Laurens F, Chagné D, Velasco R, van de Weg E, Troglio M (2014) Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (*Malus × domestica* Borkh). *PLoS ONE*, 9 e110377.
- Bielenberg DG**, Wang Y, Li Z, Zhebentyayeva T, Fan S, Reighard GL, Scorza R, Abbott AG (2008) Sequencing and annotation of the evergrowing locus in peach [*Prunus persica* (L.) Batsch] reveals a cluster of six MADS-box transcription factors as candidate genes for regulation of terminal bud formation. *Tree Genetics and Genomes*, 4: 495–507.

- Bielsa B**, Jiwan D, Fernández i Martí A, Dhingra A, Rubio-Cabetas MJ (2014) Detection of SNP and validation of a SFP InDel (deletion) in inverted repeat region of the *Prunus* species chloroplast genome. *Sci Horti*, 168: 108–112.
- Bliss FA**, Arulsekhar S, Foolad MR, Becerra V, Gillen AM, Warburton ML, Dandekar AM, Kocsisne GM, Mydin KK (2002) An expanded genetic linkage map of *Prunus* based on an interspecific cross between almond and peach. *Genome*, 45: 520–529.
- Boudehri K**, Bendahmane A, Cardinet G, Troadec C, Moing A, Dirlwanger E (2009) Phenotypic and fine genetic characterization of the D locus controlling fruit acidity in peach. *BMC Plant Biol*, 9: 59.
- Bradbury P**, Zhang Z, Kroon D, Casstevens T, Ramdoss Y, Buckler E (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23: 2633-2635.
- Branca A**, Paape TD, Zhou P, Briskine R, Farmer AD, Mudge J, Bharti AK, Woodward J E, May GD, Gentzbittel L, Ben C, Denny R, Sadowsky MJ, Ronfort J, Bataillon T, Young ND, Tiffin P (2011) Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *P. Natl. Acad. Sci. USA*, 108 (42) E864–E870.
- Cabrera A**, Kozik A, Howad W, Arus P, Iezzoni AF, van der Knaap E (2009) Development and bin mapping of a Rosaceae Conserved Ortholog Set (COS) of markers. *BMC Genomics*, 10: 562.
- Camus-Kulandaivelu L**, Veyrieras JB, Gouesnard B, Charcosset A, Manicacci D (2007) Evaluating the reliability of structure outputs in case of relatedness between individuals. *Crop Sci*, 47: 887–892.
- Cantín CM**, Crisosto CH, Ogundiwin EA, Gradziel T, Torrents J, Moreno MA, Gogorcena Y (2010) Chilling injury susceptibility in an intra-specific peach [*Prunus persica* (L.) Batsch] progeny. *Postharvest Biol Technol*, 58: 79–87.
- Cao J**, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Muller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet*, 43: 956–963.
- Cao K**, Wang L, Zhu G, Fang W, Chen C, Luo J (2012) Genetic diversity, linkage disequilibrium, and association mapping analyses of peach (*Prunus persica*) landraces in China. *Tree Genet Genomes*, 8 (5): 975–990.

- Chagné D**, Crowhurst RN, Troglio M, Davey MW, Gilmore B, Lawley C, Vanderzande S, Hellens RP, Kumar S, Cestaro A, Velasco R, Main D, Rees JD, Iezzoni A, Mockler T, Wilhelm L, Van de Weg E, Gardiner SE, Bassil N, Peace C (2012) Genomewide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS ONE*, 7 e31745.
- Chain PS**, Grafham DV, Fulton RS, Fitzgerald MG, Hostetler J, Muzny D, Ali J, Birren, B Bruce, DC Buhay, C Cole, JR Ding, Y Dugan, S Field, D Garrity, GM Gibbs, R Graves, T Han, et al. (2009) Genome project standards in a new era of sequencing. *Science*, 326: 236-237.
- Chaparro JX**, Werner DJ, O'Malley D, Sederoff RR (1994) Targeted mapping linkage analysis of morphological, isozyme, and RAPD markers in peach. *Theor Appl Genet*, 87: 805-815.
- Chen L**, Zhang S, Illa E, Song L, Wu S, Howard W, Arús P, van de Weg E, Chen K, Gao Z (2008) Genomic characterization of putative allergen genes in peach/almond and their synteny with apple. *BMC Genom*, 9: 543.
- Cipriani G**, Lot G, Huang WG, Marrazzo MT, Peterlunger E, Testolin R (1999). AC/GT and AG/CT microsatellite repeats in peach (*Prunus persica* (L.) Batsch): isolation, characterisation and cross-species amplification in *Prunus*. *Theor Appl Genet*, 99: 65–72.
- Claverie M**, Bosselut N, Lecouls AC, Voisin R, Lafargue B, Poizat C, Kleinhentz M, Laigret F, Dirlewanger E, Esmenjaud D (2004) Location of independent root-knot nematode resistance genes in plum and peach. *Theor Appl Genet*, 108: 765–773.
- Clarke J**, Sargent D, Bošković R, Belaj A, Tobutt K (2009) A cherry map from the interspecific cross *Prunus avium* ‘Napoleon’ × *P. nipponica* based on microsatellite, gene-specific and isoenzyme markers. *Tree Genet Genomes*, 5:41–51.
- Cockram J**, White J, Zuluaga D, Smith D, Comadran J, Macaulay M, Luo Z, Kearsey M, Werner P, Harrap D, Tapsell C, Liu H, Hedley P, Stein N, Schulte D, Steuernagel B, Marshall D, Thomas W, Ramsay L, Mackay I, Balding D, AGOUEB Consortium, Waugh R, O'Sullivan D (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome *Proc Natl Acad Sci U S A*, 107:21611-21616.
- da Silva C**, Bassi D, Bianco L, Pacheco I, Pirona R, Rossini L (2015) Genetic dissection of fruit weight and size in an F<sub>2</sub> peach (*Prunus persica* (L.) Batsch) progeny. *Mol Breeding*, 35: 71.

- Decroocq V**, Foulongne M, Lambert P, Le Gall P, Mantin C, Pascal T, Schurdi-Levraud V, Kervella J (2005) Analogues of virus resistance genes map to QTLs for resistance to sharka disease in *Prunus davidiana*. *Mol Genet Genom*, 272: 680–9.
- Della Strada G, Fideghelli C** (2011) Le varietà di fruttiferi introdotte nel mondo dal 1980 al 2008. MiPAAF, pag 343.
- Dettori MT**, Quarta R, Verde I (2001) A peach linkage map integrating RFLPs, SSRs, RAPDs and morphological markers. *Genome*, 44: 783-790.
- Devlin B & Roeder K** (1999) Genomic control for association studies. *Biometrics*, 55: 997–1004.
- Dirlewanger E**, Pronier V, Parvery C, Rothan C, Guye A, Monet R (1998) Genetic linkage map of peach [*Prunus persica* (L.) Batsch] using morphological and molecular markers. *Theor Appl Genet*, 97: 888-895.
- Dirlewanger E**, Moing A, Rothan C, Svanella L, Pronier V, Guye A, Plomion C, Monet R (1999) Mapping QTL controlling fruit quality in peach (*Prunus persica* (L.) Batsch). *Theor Appl Genet*, 98:18–31.
- Dirlewanger E**, Cosson P, Tavaud M, Aranzana MJ, Poizat C, Zanetto A, Arús P, Laigret F (2002) Development of microsatellite markers in peach (*Prunus persica* (L.) Batsch) and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theor Appl Genet*, 105: 127–138.
- Dirlewanger E**, Graziano E, Joobeur T, Garriga-Calderé F, Cosson P, Howad W, Arús P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crop. *Proc Natl Acad Sci U S A*, 101: 9891-9896.
- Dirlewanger E**, Kleinhentz M, Laigret F, Gomez-Aparisi J, Rubio-Cabetas MJ, Claverie M, Bosselut N, Voisin R, Esmenjaud D, Xyloyannis C, Dichio B, Poessel JL, Di Vito M, Arús P., Howad W. (2005a) Breeding for a new generation of *Prunus* rootstocks based on marker-assisted selection: A European initiative. *Acta Hort*, 663: 829–833.
- Dirlewanger E**, Arús P (2005b): **Markers in Fruit Tree Breeding: Improvement of Peach.** In *Molecular Marker Systems in Plant Breeding and Crop Improvement. Volume 55*. Edited by Lörz H, Wenzel G. Springer: 279-302.
- Dirlewanger E**, Cosson P, Boudierhi K, Renaud C, Capdeville G, Tauzin Y, Laigret F, Moing A (2006) Development of a second-generation genetic linkage map for peach [*Prunus persica* (L.) Batsch] and characterization of morphological traits affecting flower and fruit. *Tree Genet Genomes*, 3: 1–13.

- Dirlewanger E**, Quero-García J, Le Dantec L, Lambert P, Ruiz D, Dondini L, Illa E, Quilot-Turion B, Audergon JM, Tartarini S, Letourmy P, Arús P (2012) Comparison of the genetic determinism of two key phenological traits, flowering and maturity dates, in three *Prunus* species: peach, apricot and sweet cherry. *Heredity*, 109: 280–292.
- Doerge RW** (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature reviews. Genetics*, 3: 43-52.
- Dondini L**, Lain O, Geuna F, Banfi R, Gaiotti F, Tartarini S, Bassi D, Testolin R (2007) Development of a new SSR-based linkage map in apricot and analysis of synteny with existing *Prunus* maps. *Tree Genet Genomes*, 3: 239-249.
- Du F**, Clutter A, Lohuis M (2007). Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Sciences*, 3: 166-178.
- Earl DA & vonHoldt BM** (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*, 4: 359–361.
- Eduardo I**, Pacheco I, Chietera G, Bassi D, Pozzi C, Vecchietti A, Rossini L (2011) QTL analysis of fruit quality traits in two peach intraspecific populations and importance of maturity date pleiotropic effect. *Tree Genet Genomes*, 7: 323-335.
- Eduardo I**, Chietera G, Pirona R, Pacheco I, Troggio M, Banchi E, Bassi D, Rossini L, Vecchietti A, Pozzi C (2012) Genetic dissection of aroma volatile compounds from the essential oil of peach fruit: QTL analysis and identification of candidate genes using dense SNP maps. *Tree Genet Genomes*., doi:10.1007/s11295-012-0546-z.
- Ersoz ES**, Yu J, Buckler ES (2007) Applications of linkage disequilibrium and association mapping in crop plants. In: *Genomics-Assisted Crop Improvement*, R. Varshney, & R. Tuberosa, (Ed.), 97-119, Springer, ISBN 987-1-4020-6294-0, Dordrecht, the Netherlands.
- Etienne C**, Dirlewanger E, Cosson P, Svanella-Dumas L, Monet R, Moing A, Rothan C, Bodénès C, Plomion C, Kervella J (2002a) QTLs and genes controlling peach fruit quality. *Acta Hort*, 592: 253–258.
- Etienne C**, Rothan C, Moing A, Plomion C, Bodenes C, Dumas LS, Cosson P, Pronier V, Monet R, Dirlewanger E (2002b) Candidate genes and QTL for sugar and organic acid content in peach (*Prunus persica* (L.) Batsch). *Theor Appl Genet*, 105: 145–159.
- Evanno G**, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Eco*, 14: 2611–2620.

- Falchi R**, Vendramin E, Zanon L, Scalabrin S, Cipriani G, Verde I, Vizzotto G, Morgante M (2013) Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. *The Plant Journal*, 76: 175-187.
- Falush D**, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164: 1567–1587.
- Fan S**, Bielenberg DG, Zhebentyayeva TN, Reighard GL, Okie WR, Holland D, Abbott AG (2010) Mapping quantitative trait loci associated with chilling requirement, heat requirement and bloom date in peach (*Prunus persica*). *New Phytol*, 185: 917–930.
- Fang JG**, Twito T, Zhang Z, Chao CCT (2006) Genetic relationships among fruiting-mei (*Prunus mume* Sieb et Zucc.) cultivars evaluated with AFLP and SNP markers. *Genome*, 10: 1256–1264.
- Faust M e Timon B** (1995) Origin and dissemination of peach. *Hort. Reviews*, 17: 331–379.
- Fernández i Martí A**, Athanson B, Koepke T, Font i Forcada C, Dhingra A, Oraguzie N (2012) Genetic diversity and relatedness of sweet cherry (*Prunus avium* L.) cultivars based on single nucleotide polymorphic markers. *Front. Plant Sci*, 3: 116.
- Flint-Garcia S**, Thornsberry J, Bukler E (2003) Structure of linkage disequilibrium in plants. *Annual Review Plant Biology*, 54: 357-374.
- Font I Forcada C**, Oraguzie N, Igartua E, Moreno MÁ, Gogorcena Y (2013) Population structure and marker–trait associations for pomological traits in peach and nectarine cultivars. *Tree Genet Genomes*, 9:331–349.
- Foolad MR**, Arulsekhar S, Becerra V, Bliss FA (1995) A genetic map of *Prunus* based on an interspecific cross between peach and almond. *Theor Appl Genet*, 91(2): 262–269.
- Foulongne M**, Pascal T, Pfeiffer F, Kervella J (2003) QTL for powdery mildew resistance in peach × *Prunus davidiana* crosses: consistency across generations and environments. *Mol Breed*, 12: 33–50.
- Frett TJ**, Reighard G, Okie WR, Gasic K (2014) Mapping quantitative trait loci associated with blush in peach [*Prunus persica* (L.) Batsch]. *Tree Genet Genomics*, 10: 367-381.
- Ganal MW**, Polley A, Graner EM, Plieske J, Wieseke R, Luerssen H, Durstewitz G (2012) Large SNP arrays for genotyping in crop plants. *J Biosci*, 37: 821–828.
- Ganal MW**, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J, Le Paslier MC, McMullen M, Montalent P, Rose M, Schon CC, Sun Q, Walter H, Martin OC, Falque M (2011) A large maize (*Zea mays* L.)

- SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS ONE*, 6 e28334.
- Garris A**, Tai T, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics*, 169: 1631-1638.
- Germano J, Klein AS** (1999) Species specific nuclear and chloroplast single nucleotide polymorphisms to distinguish *Picea glauca*, *P. mariana* and *P. rubens*. *Theor Appl Genet*, 99: 37–49.
- Geuna F**, Toschi M, Bassi D (2003) The use of AFLP markers for cultivar identification in apricot. *Plant Breed*, 122: 526–531.
- Graham J**, Smith K, Woodhead M, Russell J (2002) Development and use of simple sequence repeat SSR markers in *Rubus* species. *Mol Ecol Notes*, 2: 250–252.
- Gunderson KL**, Steemers FJ, Lee G, Mendoza LG, Chee MS (2005) A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Genetics*, 37: 549–554.
- Gunderson KL** (2009) Whole-genome genotyping on bead arrays. *Meth MolBiol*, 529: 197–213.
- Gupta P**, Rustgi S, Kulwal P (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology*, 57: 461-485.
- Hagen LS**, Chaib J, Fady B, Decroocq V, Bouchet JP, Lambert P, Audergon JM (2004) Genomic and cDNA microsatellites from apricot (*Prunus armeniaca* L.). *Mol Ecol Notes*, 4: 742–745.
- Hammer Ø, Ryan P, Harper D** (2001) PAST: Paleontological Statistics software package for education and data analysis. *Palaeontol Electron*, 4: 9.
- Hancock JM** (1999) Microsatellites and other simple sequences: genomic context and mutational mechanisms. In *Microsatellites: Evolution and Applications*. Edited by D.B. Goldstein and C. Schlötterer. *Oxford University Press, Oxford* pp.1-9;
- Harding RM**, Boyce AJ, Clegg JB (1992) The Evolution of Tandemly Repetitive DNA: Recombination Rules. *Genetics*, 132: 847-859.
- Hardy O & Vekemans X** (2002) SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. *Mol Ecol Notes*, 2: 618-620.
- Harjes CE**, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES (2008): Natural

- genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science*, 319: 330-333.
- Hauck NR**, Iezzoni AF, Yamane H, Tao R (2001) Revisiting the *S*-allele nomenclature in sweet cherry (*Prunus avium*) using RFLP profiles. *J Am Soc Hort Sci*, 126: 654–660.
- Hedrick UP**, Howe GH, Taylor OM, Tubergen CB (1917) The peaches of New York. State of New York, Department of Agriculture. JB Lyon Company Printers, Albany, New York.
- Hellegren H** (2004) Microsatellites: Simple sequences with complex evolution. *Nat. Rev. Genet.*, 5: 435–445.
- Hill W, Robertson A** (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, 38: 226-231.
- Howad W**, Yamamoto T, Dirlewanger E, Testolin R, Cosson P, Cipriani G, Monforte AJ, Georgi L, Abbott AG, Arús P (2005) Mapping with a few plants, using selective mapping for microsatellite saturation of the *Prunus* reference map. *Genetics*, 171: 1305–1309.
- Huang X**, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler E, Qian Q, Zhang Q, Li J, Han B (2010) Genome-wide studies of 14 agronomic traits in rice landraces. *Nature Genetics*, 42: 961-967.
- Hurtado MA**, Romero C, Vilanova S, Abbott AG, Llácer G, Badenes ML (2002). Genetic diversity in apricot cultivars based on AFLP markers. *Euphytica*, 127: 297-301.
- Illa E**, Eduardo I, Audergon JM, Barale F, Dirlewanger E, Li X, Moing A, Lambert P, Le Dantec L, Gao Z, Poëssel JL, Pozzi C, Rossini L, Vecchiatti A, Arús P, Howad W (2011) Saturating the *Prunus* (stone fruits) genome with candidate genes for fruit quality. *Mol Breeding*, 28(4): 667–682.
- Jáuregui B** (1998) Identification of molecular markers linked to agronomic characters in an interspecific almond x peach progeny. Dissertation. University of Barcelona, Spain.
- Jennings H** (1917) The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. *Genetics*, 2: 97-154.
- Jones E**, Chu W-C, Ayele M, Ho J, Bruggeman E, Yourstone K, Rafalski A, Smith OS, McMullen MD, Bezawada C, Warren J, Babayev J, Basu S, Smith S (2009) Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (*Zea mays* L.) germplasm. *Mol Breeding*, 24: 165–176.

- Joobeur T**, Viruel MA, de Vicente MC, Jáuregui B, Ballester J, Dettori MT, Verde I, Truco MJ, Messeguer R, Batlle I, Quarta R, Dirlewanger E, Arús P. (1998a) Construction of a saturated linkage map for *Prunus* using an almond x peach F<sub>2</sub> progeny. *Theor Appl Genet*, 97: 1034-1041.
- Joobeur T** (1998b) Construction of a molecular marker map and genetic analysis of agronomic characters in *Prunus*. Dissertation. University of Lleida, Spain.
- Joobeur T**, Periam N, de Vicente MC, King GJ, Arús P (2000) Development of a second generation linkage map for almond using RAPD and SSR markers. *Genome*, 43: 649-655.
- Jouy C**, Gandelin MH, Collonnier C, Guitouni C, Zhang D, Panheleux AL, Fouilloux S, Pascal T, Tuero C, Füstös ZS, Szani ZS, Zubor L, Chomé P, Badenes MT, Soler G, Martinez J, Alonso JM, Verde I, Dettori MT, Della Strada G, Semon S (2012) Management of peach tree reference collections: ongoing research & development program relevant to the community plant variety rights protection system. *Acta Hort (ISHS)*, 962: 51-56.
- Jun JH**, Chung KH, Jeong SB, Lee HJ (2002) Development of RAPD and SCAR markers linked to flesh adhesion gene in peach. XXVI International Horticultural Congress. Toronto, Canada. p. 335.
- Jung M**, Ching A, Bhatramakki D, Dolan M, Tingey S, Morgante M, Rafalski A (2004) Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh1* locus in elite maize germplasm. *Theoretical and Applied Genetics*, 109: 681-689.
- Kawuki R**, Ferguson M, Labuschagne M, Herselman L, Kim D-J (2009) Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). *Molecular Breeding*, 23: 669–684.
- Kimura T**, Nishitani C, Iketani H, Ban Y, Yamamoto T (2006) Development of microsatellite markers in rose. *Mol Ecol Notes*, 2: 250–252.
- Klagges C**, Campoy JA, Quero-García J, Guzmán A, Mansur L, Gratacós E, Silva H, Rosyara UR, Iezzoni A, Meisel L, Dirlewanger E (2013) Construction and comparative analyses of highly dense linkage maps of two sweet cherry intra-specific progenies of commercial cultivars. *PLoS ONE*, 7: e54743.
- Lalli D**, Abbott A, Zhebentyayeva T, Badenes ML, Darmsteegt V, Polák J, Krška B, Salava J (2008) A genetic linkage map for an apricot (*Prunus armeniaca* L.) BC<sub>1</sub> population mapping Plum pox virus resistance. *Tree Genetics & Genomes* 4: 481–493.

- Lam HM**, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B, Li J, Jan M, Wang J, Shao G, Wang J, Sun SSM, Zhang G (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.*, 42: 1053–1059.
- Lambert P**, Hagen L S, Arús P, Audergon JM (2004) Genetic linkage maps of two apricot cultivars (*Prunus armeniaca* L.) compared with the almond Texas x peach Earlygold reference map for *Prunus*. *Theor Appl Genet*, 108: 1120-1130.
- Le Dantec L**, Cardinet G, Bonet J, Fouche´ M, Boudehri K, Monfort A, Poessel JL, Moing A, Dirlewanger E (2010) Development and mapping of peach candidate genes involved in fruit quality and their transferability and potential use in other Rosaceae species. *Tree Genet Genomes*, 6: 995–1012.
- Lecoals AC**, Rubio MJ, Cabetas JC, Minot R, Voisin A, Bonnet G, Salesses E, Dirlewanger E, Esmenjaud D (1999). RAPD and SCAR markers linked to the Ma1 root-knot nematode resistance gene in Myrobalan plum (*Prunus cerasifera* Ehr.). *Theor. Appl. Genet.*, 99: 328-335.
- Lewontin C** (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, 49 (1): 49-67.
- Li X**, Wang Y, Wang B, Wang C, Shangguan L, Huang Z, Fang J (2010) Genetic relationships between fruiting and flowering mei (*Prunus mume*) cultivars using SNP markers. *J Hortic Sci Biotechnol*, 85: 329–334.
- Li Y**, Huang Y, Bergelson J, Nordborg M, Borevitz J (2010) Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 107: 21119-21204.
- Li Y**, Haseneyer G, Schön C, Ankerst D, Korzun V, Wilde P, Bauer E (2011) High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biology*, 11: 1-14.
- Li X**, Wei Y, Moore K, Michaud R, Viands D, Hansen J, Acharya A, Brummer E (2011) Association mapping of biomass yield and stem composition in a tetraploid alfalfa breeding population *The Plant Genome*, 4: 24-35.
- Li XW**, Meng XQ, Jia HJ, Yu ML, Ma RJ, Wang LR, Cao K, Shen ZJ, Niu L, Tian JB, Chen MJ, Xie M, Arús P, Gao ZS, Aranzana MJ (2013) Peach genetic resources: diversity, population structure and linkage disequilibrium. *BMC Genet*, 14: 84.

- Lijavetzky D**, Antonio Cabezas J, Ibanez A, Rodriguez V, Martinez-Zapater JM (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics*, 8: 424.
- Liebhard R**, Gianfranceschi L, Koller B, Ruder CD, Tarchini R, Weg EVD, Gessler C (2002) Development and characterisation of 140 new microsatellites in apple (*Malus x domestica* Borkh.). *Mol Breed*, 10: 217–241.
- Lipka AE**, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics*, 28: 2397-2399.
- Litt M, Luty JA** (1989) A hypervariable microsatellite revealed by in vitro amplification of dinucleotide repeat within the cardiac muscle action gene. *Am.J. Hum. Genet.*, 44: 397-401.
- Lu ZX**, Reighard GL, Baird WV, Abbott AG, Rajapakse S (1996) Identification of peach rootstock cultivars by RAPD markers. *HortScience*, 31: 127-129.
- Lu ZX**, Sosinski B, Reighard GL, Baird WV, Abbott AG (1998) Construction of a genetic linkage map and identification of AFLP markers for resistance to root-knot nematodes in peach rootstocks. *Genome*, 41: 199–207.
- Marandel G**, Pascal T, Candresse T, Decroocq V (2009) Quantitative resistance to Plum pox virus in *Prunus davidiana* P1908 linked to components of the eukaryotic translation initiation complex. *Plant Pathol*, 58: 425–435.
- Martinez-Garcia P**, Parfitt D, Ogundiwin E, Fass J, Chan H, Ahmad R, Lurie S, Dandekar A, Gradziel T, Crisosto C (2013a) High density SNP mapping and QTL analysis for fruit quality characteristics in peach (*Prunus persica* L.). *Tree Genet. Genomes*, 9: 19–36.
- Martinez-Garcia P**, Fresnedo-Ramírez J, Parfitt DE, Gradziel TM, Crisosto CH (2013b) Effect prediction of identified SNPs linked to fruit quality and chilling injury in peach [*Prunus persica* (L.) Batsch]. *Plant Mol Biol.*, 81: 161-174.
- Martins M**, Sarmiento D, Batlle I, Vargas F, and Oliveira MM (2005) Development of SCAR/CAPS markers linked to tolerance/sensitivity to *Fusicoccum* in almond. *Options Méditerranéennes CIHEAM/IAMZ*, 63: 187–191.
- Metzker ML** (2010) Sequencing technologies – the next generation. *Nat. Rev. Genet.*, 11: 31–46.

- Micheletti D**, Troglio M, Zharkikh A, Costa F, Malnoy M, Velasco R, Salvi S (2011) Genetic diversity of the genus *Malus* and implications for linkage mapping with SNPs. *Tree Genetics & Genomes*, 7: 857–868.
- Micheletti D**, Dettori MT, Micali S, Aramini V, Pacheco I, Foschi S, Banchi E, Quilot B, Barreneche T, Lambert P, Pascal T, Iglesias I, Carbó J, Wang L, Ma RJ, Gao ZS, Caprera A, Troglio M, Bassi D, Rossini L, Verde I, Laurens F, Arús P, Aranzana MJ. Whole-genome analysis of diversity and SNP-major gene association in Peach Germplasm. *PLoS ONE*, submitted
- Mnejja M**, Garcia-Mas J, Howad W, Badenes ML, Arús P (2004) Simple-sequence repeat (SSR) markers of Japanese plum (*Prunus salicina* Lindl.) are highly polymorphic and transferable to peach and almond. *Mol Ecol Notes*, 4: 163–166.
- Mnejja M**, Garcia-Mas J, Howad W, Arús P (2005) Development and transportability across *Prunus* species of forty-two polymorphic almond microsatellites. *Mol Ecol Notes*, 5: 531–535.
- Mnejja M**, Garcia-Mas J, Audergon JM, Arús P (2010) *Prunus* microsatellite marker transferability across rosaceous crops. *Tree Genetics & Genomes*, 6: 689–700.
- Monfort A**, Vilanova S, Davis TM, Arús P (2006) A new set of polymorphic simple sequence repeat (SSR) markers from a wild strawberry (*Fragaria vesca*) are transferable to other diploid *Fragaria* species and to *Fragaria* × *ananassa*. *Mol Ecol Notes*, 6: 197–200.
- Morgante M**, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant J*, 3(1):175–182.
- Myles S**, Peiffer J, Brown P, Ersoz E, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell*, 21:2194–2202.
- Myles S**, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D, Bustamante CD, Buckler ES (2011) Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. USA*, 108: 3530–3535.
- Nordborg M** (2000) Linkage disequilibrium, gene trees and selfing: An ancestral recombination graph with partial self-fertilization. *Genetics*, 154: 923–929.
- Ogundiwin EA**, Marti C, Forment J, Pons C, Granell A, Gradziel TM, Peace CP, Crisosto CH (2008) Development of ChillPeach genomic tools and identification of cold-responsive genes in peach fruit. *Plant Mol Biol*, 68: 379–397.

- Ogundiwin EA**, Peace CP, Gradziel TM, Parfitt DE, Bliss FA, Crisosto CH (2009) A fruit quality gene map of *Prunus*. *BMC Genomics*, 10: 587.
- Okie WR** (1998) Handbook of peach and nectarine varieties: performance in the southeastern United States and index of names. (U.S. Dept. of Agriculture, Agricultural Research Service ; National Technical Information Service, distributor, 1998).
- Olmstead JW**, Sebolt AM, Cabrera A, Sooriyapathirana SH, Iriarte G, Wang D, Chen CY, Knaap E van der, Iezzoni AF (2008) Construction of an intra-specific sweet cherry (*Prunus avium* L.) genetic linkage map and synteny analysis with the *Prunus* reference map. *Tree Genet Genomes*, 4: 897-910.
- Olukolu BA**, Trainin T, Fan S, Kole C, Bielenberg DG, Reighard GL, Abbott AG, Holland D (2009) Genetic linkage mapping for molecular dissection of chilling requirement and budbreak in apricot (*Prunus armeniaca* L.). *Genome* 52: 819–828.
- Oraguzie N**, Wilcox P, Rikkerink H, de Silva H (2007) Linkage disequilibrium, In: *Association mapping in plants*, N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner & H.N. de Silve, (Ed.), 11-39, Springer, ISBN-13 9780387358444, New York, USA.
- Pacheco I**, Bassi D, Eduardo I, Ciacciuli A, Pirona R, Rossini L, Vecchiotti A (2014) QTL mapping for brown rot (*Monilinia fructigena*) resistance in an intraspecific peach (*Prunus persica* L. Batsch) F<sub>1</sub> progeny. *Tree Genet Genomes*., 10.1007/s11295-014-0756-7
- Palaisa KA**, Morgante M, Williams M, Rafalski A (2003): Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell*, 15: 1795-1806.
- Park YJ**, Lee JK, Kim NS (2009) Simple sequence repeat polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops. *Molecules*, 14: 4546-4569.
- Pascal T**, Pfeiffer F, Kervella J (2010) Powdery mildew resistance in the peach cultivar Pamirskij 5 is genetically linked with the Gr gene for leaf color. *Hortscience*, 45: 150–152.
- Peace CP**, Norelli JL (2009) Genomics approaches to crop improvement in the Rosaceae. In: Folta K, Gardiner S, eds. *Genetics and Genomics of Rosaceae*. New York: *Springer*. pp 19–53.
- Peace CP**, Bassil N, Main D, Ficklin S, Rosyara UR, Stegmeir T, Sebolt A, Gilmore B, Lawley C, Mockler TC, Bryant DW, Wilhelm L, Iezzoni A (2012) Development and

- evaluation of a genome-wide 6 K SNP array for diploid sweet cherry and tetraploid sour cherry. *PLoS ONE* 7:e48305.
- Pirona R**, Eduardo I, Pacheco I, Da Silva Linge C, Miculan M, Verde I, Tartarini S, Dondini L, Pea G, Bassi D, Rossini L (2013) Fine mapping and identification of a candidate gene for a major locus controlling maturity date in peach. *BMC Plant Biol*, 13: 166.
- Potter D**, Eriksson T, Evans RC, Oh S, Smedmark JEE, Morgan DR, Kerr M, Robertson KR, Arsenault M, Dickinson TA, Campbell CS (2007) Phylogeny and classification of Rosaceae. *Plant Syst Evol*, 266: 5-43.
- Price A**, Patterson N, Plenge R, Weinblatt M, Shadick N, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38: 904-909.
- Pritchard J & Rosenberg N** (1999) Use of unlinked genetic markers to detect population stratification in association studies. *The American Journal of Human Genetics*, 65: 220-228.
- Pritchard J**, Stephens M, Rosenberg N, Donnelly P (2000a) Association mapping in structured populations. *The American Journal of Human Genetics*, 67: 170-181.
- Pritchard J**, Stephens M, Donnelly P (2000b) Inference of population structure using multilocus genotype data. *Genetics*, 155: 945–959.
- Pritchard J** (2001) Deconstructing maize population structure. *Nature Genetics*, 28: 203-204.
- Quilot B**, Wu BH, Kervella J, Génard M, Foulongne M, Moreau K (2004) QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species *P. davidiana*. *Theor Appl Genet*, 109: 884–897.
- Rafalski A** (2002) Application of single nucleotide polymorphism in crop genetics. *Curr. Opin. Plant Biol.*, 5: 94–100.
- Rajapakse S**, Belthoff LE, He G, Estager AE, Scorza R, Verde I, Ballard RE, Baird WV, Callahan A, Monet R, Abbott AG (1995) Genetic mapping in peach using morphological, RFLP and RAPD markers. *Theor Appl Genet*, 90: 503-510.
- Richard GF, Paques F** (2000) Mini and Microsatellite expansion: the recombination connection. *Embo Reports*, 1(2):122-126.
- Richards RI, Sutherland GR** (1994) Simple repeat DNA is not replicated simply. *Nat. Genet.*, 6: 114-116.

- Romeu JF**, Monforte AJ, Sánchez G, Granell A, García-Brunton J, Badenes ML, Rios G (2014). Quantitative trait loci affecting reproductive phenology in peach. *BMC Plant Biol.*, 14: 52 10.1186/1471-2229-14-52.
- Rosenberg NA** (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes*, 4: 137–138.
- Rosyara UR**, Sebolt AM, Peace C, Iezzoni AF (2014) Identification of the parental parent of ‘Bing’ sweet cherry and confirmation of descendants using single nucleotide polymorphism markers. *J Am Soc Hortic Sci*, 139: 148–156.
- Rubio M**, Pascal T, Bachellez A, Lambert P (2010) Quantitative trait loci analysis of plum pox virus resistance in *Prunus davidiana* P1908: new insights on the organization of genomic resistance regions. *Tree Genet Genome*, 6: 291–304.
- Salazar JA**, Rubio M, Ruiz D, Tartarini S, Martínez-Gómez P, Dondini L (2015) SNP development for genetic diversity analysis in apricot. *Tree Genet Genome*, 11: 15.
- Salesses G & Mouras A** (1977) Tentative d’utilisation des protoplastes pour l’étude des chromosomes chez les *Prunus*. *Ann Amel Plant*, 27: 363-368.
- Sánchez-Pérez R**, Howad W, Dicenta F, Arús P, Martínez-Gómez P (2007) Mapping major genes and quantitative trait loci controlling agronomic traits in almond. *Plant Breed*, 126:310-319.
- Sargent DJ**, Marchese A, Simpson DW, Howad W, Fernandez-Fernandez F, Monfort A, Arus P, Evans KM, Tobutt KR (2009) Development of "universal" gene-specific markers from *Malus* spp. cDNA sequences, their mapping and use in synteny studies within Rosaceae. *Tree Genet Genomes*, 5: 133.
- Schlötterer C, Tautz D** (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acid Research*, 20: 211-215
- Schlötterer C** (2004) The evolution of molecular markers-just a matter of fashion? *Nature Reviews Genetics*, 5 (1): 63-69.
- Scorza R**, Melnicenco L, Dang P, Abbott AG (2002) Testing a microsatellite marker for selection of columnar growth habit in peach [*Prunus persica* (L.) Batsch]. *Acta Hort.*, 592: 285–289.
- Senková S**, Žiarovská J, Bežo M, Štefúnová V, Ražná K (2013): Utilization of IRAP technique for plums genotypes differentiation. *Biosc Res*, 10: 1-7.
- Slatkin M** (2008) Linkage disequilibrium: Understanding the evolutionary past and mapping the medical future. *Nature Review Genetics*, 9: 477-485.

- Sosinski B**, Shulaev V, Dhingra A, Kalyanaraman A, Bumgarner R, Rokhsar D, Verde I, Velasco R, Abbott AG (2009) Rosaceous Genome Sequencing: Perspectives and Progress. In Genetics and Genomics of Rosaceae, Plant Genetics and Genomics: Crops and Models 6. Edited by Folta KM, Gardiner SE. Springer Science+Business Media, LLC; 2009:601-615.
- Soto-Cerda BJ, Cloutier S** (2012) Association mapping in plant genomes. In: Caliskan, M eds. Genetic diversity in plants. InTech, Rijeka, pp. 29-54.
- Squirrell J**, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? *Mol Ecol*, 12: 1339–1348.
- Stich B & Melchinger AE** (2009) Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. *BMC Genomics*, 10
- Stich B, Melchinger A** (2010) An introduction to association mapping in plants. *CAB Reviews: Perspectives in agriculture, Veterinary Science, Nutrition and Natural Resources*, Vol. 5, No. 39, 1-9.
- Sturtevant A H** (1913) The linear arrangement of six sexlinked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.*, 14: 43-59.
- Sun LD**, Zhang QX, Xu ZD, Yang WR, Guo Y, Lu J, Pan H, Cheng T, Cai M (2013) Genome-wide DNA polymorphisms in two mei (*Prunus mume* Sieb. Et Zucc.). *BMC Genet*, 14: 98.
- Syvanen AC**, 2005. Toward genome-wide SNP genotyping. *Nat. Genet.*, 37: S5–S10.
- Tautz D, Renz M** (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acid Research*, 12: 4127-4238.
- Tavassolian I**, Rabiei G, Gregory D, Mnejja M, Wirthensohn MG, Hunt PW, Gibson JP, Ford CM, Sedgley M, Wu S-B (2010) Construction of an almond linkage map in an Australian population Nonpareil × Lauranne. *BMC Genomics*, 11: 551–561.
- Tenailon M**, Sawkins M, Long A, Gaut R, Doebley J, Gaut B (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA*, 98: 9169–9166.
- Testolin R**, Marrazzo T, Cipriani G, Quarta R, Verde I, Dettori MT, Pancaldi M, Sansavini S (2000) Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. *Genome*, 43: 512-520.

- The R development Core Team** (2008) Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Thornsberry J**, Goodman M, Doebley J, Kresovich S, Nielsen D, Buckler E (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics*, 28: 286-289.
- Toth G**, Gaspari Z, Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.*, 10: 967–981.
- Turner SD** (2014) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots
- Vaughan SP, Russell K** (2004) Characterization of novel microsatellites and development of multiplex PCR for large-scale population studies in wild cherry, *Prunus avium*. *Mol Ecol Notes*, 4: 429–431.
- Vendramin E**, Deitori MT, Giovinnazzi J, Micali S, Quarta R, Verde I (2007) A set of EST-SSRs isolated from peach fruit transcriptome and their transportability across *Prunus* species. *Mol Ecol Notes*, 7: 307-310.
- Vendramin E**, Pea G, Dondini L, Pacheco I, Dettori MT, Gazza L, Scalabrin S, Strozzi F, Tartarini S, Bassi D, Verde I, Rossini L (2014) A unique mutation in a MYB gene cosegregates with the nectarine phenotype in peach. *PLoS ONE*, 9: e90574.
- Verde I**, Quarta R, Cerdrola C, Dettori MT (2002) QTL analysis of agronomic traits in a BC1 peach population. *Acta Hort*, 592: 291–297.
- Verde I**, Lauria M, Dettori MT, Vendramin E, Balconi C, Micali S, Wang Y, Marrazzo MT, Cipriani G, Hartings H, Testolin R, Abbott AG, Motto M, Quarta R (2005) Microsatellite and AFLP markers in the *Prunus persica* [L. (Batsch)] x *P. ferganensis* BC1 linkage map: saturation and coverage improvement. *Theoretical and Applied Genetics*, 111: 1013-1021.
- Verde I**, Bassil N, Scalabrin S, Gilmore B, Lawley CT, Gasic K, Micheletti D, Rosyara UR, Cattonaro F, Vendramin E, Main D, Aramini V, Blas AL, Mockler TC, Bryant DW, Wilhelm L, Troglio M, Sosinski B, Aranzana MJ, Arus P, Iezzoni A, Morgante M, Peace CP (2012) Development and evaluation of a 9 K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PLoS ONE*, 7(4):e35668.
- Verde I**, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, Zuccolo A, Rossini L, Jenkins J, Vendramin E, Meisel

- L, Decroocq V, Sosinski B, Prochnik S, Mitros T, Policriti A, Cipriani G, Dondini L, Ficklin S, Goodstein DM, Xuan P, Del Fabbro C, Aramini V, Copetti D, Gonzalez S, Horner DS, Falchi R, Lucas S, Mica E, Maldonado J, Lazzari B, Bielenber D, Pirona R, Miculan M, Barakat A, Testolin R, Stella A, Tartarini S, Tonutti P, Arús P, Orellana A, Wells C, Main D, Vizzotto G, Silva H, Salamini F, Schmutz J, Morgante M, Rokhsar DS (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet*, 45: 487-494.
- Verde I**, Shu S, Jenkins J, Zuccolo A, Dettori MT, Dardick C, Rossini L, Grimwood J, Pirona R, Goodstein DM, Dondini, L, Vendramin E, Martínez-Gómez P, Silva H, Micali S, Falchi R, Scalabrin S, Bassi D, Main D, Orellana A, Vizzotto G, Tartarini S, Meisel L, Abbott AG, Morgante M, Rokhsar DS, Schmutz J (2015) The Peach v2.0 Release: An Improved Genome Sequence for Bridging the Gap Between Genomics and Breeding in *Prunus*. Atti International Plant & Animal Genome XXIII Conference / January 10-14, 2015 - San Diego, CA, USA pag 165 <https://pag.confex.com/pag/xxiii/webprogram/Paper14519.html>
- Vignal A**, Milan D, SanCistobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetic Selection and Evolution*, 34: 275-305.
- Vilanova S**, Sargent DJ, Arus P, Monfort A (2008) Synteny conservation between two distantly-related Rosaceae genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry). *BMC Plant Biol*, 8 :67.
- Virk P**, Ford-Lloyd B, Jackson M, Pooni H, Clemeno T, Newbury H (1996) Predicting quantitative variation within rice germplasm using molecular markers. *Heredity*, 76: 296-304.
- Viruel MA**, Messeguer R, de Vicente MC, Garcia-Mas J, Puigdomènech P, Vargas F, Arús P (1995) A linkage map with RFLP and isozyme markers for almond. *Theor Appl Genet*, 91: 964-971.
- Viruel MA**, Madur D, Dirlewanger E, Pascal T, Kervella J (1998) Mapping quantitative trait loci controlling peach leaf curl resistance. *Acta Hort*, 465: 79–87.
- Wallace JG**, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS ONE*, 10(12): e1004845.

- Wang Y**, Georgi LL, Reighard G L, Scorza R, Abbott AG (2002) Genetic mapping of the *evergrowing* gene in peach [*Prunus persica* (L.) Batsch]. *Journal Hered*, 93: 352–358.
- Wakely J**, (1994) Substitution-rate variation among sites and the estimation of transition bias. *Mol. Biol. Evol.*, 11: 436–442.
- Wakeley J** (1996) The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends Ecol. Evol.*, 11: 158–162.
- Warburton ML**, Bliss FA (1996) Genetic diversity in peach (*Prunus persica* L. Batch) revealed by randomly amplified polymorphic DNA (RAPD) markers and compared to inbreeding coefficients. *J. Am. Soc. Hort. Sci.*, 121: 1012-1019.
- Warburton ML**, Becerra-Velasquez VL, Goffreda JC, Bliss FA (1996) Utility of RAPD markers in identifying genetic linkages to genes of economic interest in peach. *Theor Appl Genet*, 93: 920–925.
- Weber A**, Clark RM, Vaughn L, Sanchez-Gonzalez JDJ, Yu J, Yandell BS, Bradbury P, Doebley JF (2008) Major regulatory genes in maize contribute to standing variation in Teosinte (*Zea mays* ssp. *parviglumis*). *Genetics*, 177: 2349–2359.
- Wu SB**, Tavassolian I, Rabiei G, Hunt PW, Wirthensohn M, Gibson JP, Ford CM, Sedgley M (2009) Mapping SNP-anchored genes using high-resolution melting analysis in almond. *Mol Genet Genomics*, 282: 273–281.
- Wu SB**, Tavassolian I, Francks TK, Hunt P, Wirthensohn MG, Gibson JP, Sedgley M (2010) Discrimination of SNP genotypes associated with complex haplotypes by high-resolution melting analysis in almond: implications for improved marker efficiencies. *Mol Breed*, 25: 351–357.
- Wünsch A** (2009) Cross-transferable polymorphic SSR loci in *Prunus* species. *Sci Hortic*, 120: 348–352.
- Wünsch A**, Hormaza JI (2002) Molecular characterisation of sweet cherry (*Prunus avium* L.) genotypes using peach (*Prunus persica* (L.) Batsch) SSR sequences. *Heredity*, 89: 56–63.
- Würschum T**, Maurer H, Kraft T, Janssen G, Nilsson C, Reif J (2011) Genome-wide association mapping of agronomic traits in sugar beet. *Theoretical and Applied Genetics*, 123: 1121-1131.
- Yamamoto T**, Shimada T, Imai T, Yaegaki H, Haji T, Matsuta N, Yamaguchi M, Hayashi T (2001) Characterization of morphological traits based on a genetic linkage map in peach. *Breed Sci*, 51: 271–278.

- Yamamoto T**, Kimura T, Shoda M, Ban Y, Hayashi T, Matsuta N (2002) Development of microsatellite markers in Japanese pear (*Pyrus pyrifolia* Nakai). *Mol Ecol Notes*, 2: 14–16.
- Yang N**, Reighard G, Ritchie D, Okie W, Gasic K (2013) Mapping quantitative trait loci associated with resistance to bacterial spot (*Xanthomonas arboricola* pv. Pruni) in peach. *Tree Genet Genomes* 9: 573–586.
- Yu J**, Pressoir G, Briggs W, Vroh Bi I, Yamasaki M, Doebley J, McMullen M, Gaut B, Nielsen D, Holland J, Kresovich S, Buckler E (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38: 203–208.
- Yuying S**, Xiajunb D, Fei W, Binhuaa C, Zhihonga G, Zhen Z (2011). Analysis of genetic diversity in Japanese apricot (*Prunus mume* Sieb. et Zucc.) based on REMAP and IRAP molecular marker. *Scientia Horticulturae*, 132: 50–58.
- Zhang D**, Bai G, Zhu C, Yu J, Carver B (2010) Genetic diversity, population structure, and linkage disequilibrium in U.S. elite winter wheat. *The Plant Genome*, 3: 117–127.
- Zhang P**, Liu X, Tong H, Lu Y, Li J (2014) Association Mapping for Important Agronomic Traits in Core Collection of Rice (*Oryza sativa* L.) with SSR Markers. *PLoS ONE*, 9(10): e111508.
- Zhao K**, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An arabidopsis example of association mapping in structured samples. *PLoS Genet*, 3:e4.
- Zhao K**, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2- 467.
- Zhebentyayeva TN**, Swire-Clark G, Georgi LL, Garay L, Jung S, Forrest S, Blenda AV, Blackmon B, Mook J, Horn R, Howad W, Arús P, Main D, Tomkins JP, Sosinski B, Baird WV, Reighard GL, Abbott AG (2008) A framework physical map for the peach, a model Rosaceae species. *Tree Genet Genomes*, 4: 745–756.
- Zhu C**, Gore M, Buckler ES, Yu J (2008). Status and prospects of association mapping in plants. *Plant Genome*, 1: 5–20.