



DIPARTIMENTO DI SCIENZE AGRARIE ALIMENTARI E AMBIENTALI
CORSO DI LAUREA IN: SCIENZE E TECNOLOGIE AGRARIE

IL MIGLIORAMENTO GENETICO DELLE PIANTE NELL'ERA PAN-GENOMICA PLANT BREEDING IN PAN-GENOMIC ERA

TIPO TESI: compilativa

Studente:
MATTEO GATTARI

Relatore:
PROF. ROBERTO PAPA

Correlatore:
GAIA CORTINOVIS

ANNO ACCADEMICO 2020-2021

Alle quattro persone più importanti per me
Benedetta, Paola, Stefano e Valentina
e a Damiano che ha saputo stravolgere la mia vita
rendendola meravigliosa.

SOMMARIO

ELENCO DELLE TABELLE.....	5
ELENCO DELLE FIGURE	6
ACRONIMI E ABBREVIAZIONI	7
INTRODUZIONE E SCOPO DELLA TESI	9
CAPITOLO 1 L'ORTICOLTURA ITALIANA: PROBLEMATICHE E TENDENZE DI UN SETTORE	10
1.1 Cambiamenti Climatici: problematiche e ripercussioni in agricoltura ed orticoltura .	10
1.2 Evoluzione delle nuove fisiopatie e avversità.....	12
1.3 Le varietà orticole e il rischio di uniformazione	13
1.4 Le politiche agricole comunitarie	13
1.4.1 La Strategia Farm To Fork.....	13
1.4.2 Nuova Politica Agricola Comunitaria.....	14
1.5 Tendenze dei consumatori e del mercato	14
CAPITOLO 2 LA SELEZIONE NEL MIGLIORAMENTO GENETICO	15
2.1 La selezione da un punto di vista genetico	15
2.1.1 Caratteri monogenici.....	15
2.1.2 Caratteri poligenici o Quantitativi	16
2.1.3 Selezione per caratteri quantitativi.....	17
2.2 Limiti genetici della selezione	18
CAPITOLO 3 OPERARE LA SELEZIONE: STRUMENTI MOLECOLARI	20
3.1 Marcatori molecolari.....	20
3.2 Tipologie di Marcatori Molecolari.....	21
3.2.1 RFLP.....	21
3.2.2 AFLP.....	21
3.2.3 SSR o Microsatelliti.....	23
3.2.4 SNP	23
CAPITOLO 4 SELEZIONE ASSISTITA DA MARCATORI E SELEZIONE GENOMICA	25

4.1	Costruzione di mappe genetiche e mappaggio di geni.....	25
4.2	QTL e mappaggio	26
4.2.1	Mappaggio di QTL per Linkage	26
4.2.2	Mappaggio di QTL per associazione	27
4.2.3	Analisi del mappaggio per associazione	29
4.2.4	GWAS, un approccio innovativo nel mappaggio dei QTL.....	29
4.2.5	Pregi e difetti del mappaggio di QTL per associazione	30
4.3	La MAS – Selezione Assistita da Marcatori	31
4.3.1	MAS per geni ad effetto maggiore.....	32
4.3.2	La Selezione Genomica (GS).....	35
CAPITOLO 5 PROSPETTIVE FUTURE DEL MIGLIORAMENTO GENETICO		38
5.1	L’era Pan-genomica.....	38
5.1.1	Struttura del Pan-Genoma.....	40
5.1.2	Possibili impieghi del Pan-Genoma nel miglioramento genetico delle piante ...	41
CONCLUSIONI		44
BIBLIOGRAFIA		45
RINGRAZIAMENTI.....		50

ELENCO DELLE TABELLE

Tabella 1-1: Anomalie di temperatura media in Italia dal 1961 al 2019 in confronto al resto del mondo.....	11
Tabella 1-2: Anomalie medie annuali del numero di giorni con onde di calore in Italia rispetto al valore normale 1961 – 2019.....	11
Tabella 4-1: Esempio di Manhattan Plot	30

ELENCO DELLE FIGURE

Figura 2-1: Curva normale di una popolazione teorica con media μ e deviazione stand .	16
Figura 2-2: Differenziale di selezione (S).....	17
Figura 3-1: Schema del processo di AFLP	22
Figura 4-1: Schema dei due sistemi di mappaggio QT	28
Figura 4-2: Schema di reincrocio.....	33
Figura 4-3: Confronto fra reincrocio tradizionale e con MABC.....	34
Figura 5-1Diversi tipi di variazioni strutturali (SV) rilevabili nel pan genoma.....	39
Figura 5-2:Confronto tra l'utilizzo di un singolo genoma di riferimento e il modello pangenomico	40
Figura 5-3:Rappresentazione schematica della struttura del pan-genoma.....	41

ACRONIMI E ABBREVIAZIONI

TOBRFV	Tomato Brown Rugose Fruit Virus: Virus del frutto marrone rugose del pomodoro
ToLCNDV	Tomato Leaf Curl New Delhi Virus: Virus dell'accartocciamento fogliare "New Delhi" del pomodoro
PAC	Politica Agricola Comune
PCR	Polymerase Chain Reaction: Reazione a catena della polimerasi
MAS	Marker Assisted Selection: Selezione assistita da marcatore
RFLP	Restriction Fragment Length Polymorphism: Polimorfismo da lunghezza dei frammenti di restrizione
AFLP	Amplified Fragment Length Polymorphism: Polimorfismo da lunghezza di frammenti amplificati
SSR	Single Sequence Repeats: Ripetizioni di sequenze singole
STR	Short Tandem Repeats: Ripetizioni di sequenza corte in tandem
SNP	Single Nucleotide Polymorphism: Polimorfismo da singolo nucleotide
QTL	Quantitative Trait Locus: Locus per carattere quantitativo
GS	Genomic Selection: Selezione Genomica
RIL	Recombinant Inbred Lines: Linee ricombinanti inbreed
LD	Linkage Disequilibrium: Disequilibrio da Linkage
GWAS	Genome-Wide Association Study: Studio di associazione dell'intero genoma
MABC	Marker Assisted BackCross: Reincrocio assistito da marcatore
MARS	Marker Assisted Recurrent Selection: Selezione ricorrente assistita da marcatore
BV	Breeding Value: Valore dell'incrocio
GEBV	Genomic Estimated Breeding Value: Stima genomica del valore dell'incrocio

NGS	New Generation Sequencing: Sequenziamento di nuova generazione
TGS	Third Generation Sequencing: Sequenziamento di terza generazione
SV	Structural Variation. Variazione di struttura
CNV	Copy Number Variant: Variazione nel numero di copie
PAV	Presence/Absence Variant: Variazione di presenza/assenza
PHG	Practical Haplotype Graph: Grafico pratico dell'aplotipo

INTRODUZIONE E SCOPO DELLA TESI

Il mondo di oggi è pervaso da enormi cambiamenti: clima, società, economia, crescita demografica sono temi che l'uomo si trova ad affrontare in una sfida epocale che non conosce uguali nella storia della nostra specie.

L'agricoltura deve provvedere a fornire cibo salutare, sostenibile e in quantità sufficiente per sfamare tutte le donne e gli uomini che abitano il pianeta Terra, e lo deve fare in maniera tale da non intaccare le già limitate risorse naturali di cui disponiamo, ma anzi provvedendo a conservarle e, possibilmente, rinnovarle.

La dieta mediterranea, la migliore al mondo secondo diversi studi, si basa sull'alto consumo di frutta e verdura; in particolar modo le specie orticole hanno visto, negli ultimi decenni, un potente impiego di mezzi e tecnologie nei programmi di miglioramento genetico, tanto da creare numerose varietà in grado di adattarsi alle più diverse situazioni.

Ma oggi, le problematiche accennate, innalzano ancora di più il livello della sfida tanto da non rendere più sufficienti i traguardi raggiunti.

Le nuove scoperte, i metodi di analisi, l'impiego delle nuove tecnologie in ambito di miglioramento genetico delle piante, possono, anzi devono aiutare gli agricoltori di tutto il mondo a coltivare riducendo gli input e garantendosi un giusto compenso.

Questa tesi ha lo scopo di illustrare i progressi e le tecniche ottenute oggi in ambito di plant breeding, specialmente in orticoltura, e cercare di dare uno sguardo al prossimo futuro, per capire quali innovazioni si stanno affacciando per cercare di dare una risposta alle crescenti problematiche di questa era moderna.

Capitolo 1

L'ORTICOLTURA ITALIANA: PROBLEMATICHE E TENDENZE DI UN SETTORE

Con una superficie di oltre 410.000 ha - di cui 37.800 coperti da serre - (Istat, 2021) l'Italia, insieme alla Spagna, rappresenta il vertice europeo della produzione orticola.

Sul territorio italiano operano 63 società sementiere orticole, molte delle quali sono impegnate anche nel settore della ricerca (Assosementi, 2021).

Negli ultimi anni la continua e veloce evoluzione del mercato, una sempre crescente attenzione dei consumatori a ciò che mettono nel piatto, la necessità di resilienza nei confronti del cambiamento climatico, stanno determinando delle sfide che il settore agricolo può vincere solo grazie ad una serie di interventi, nei quali il miglioramento genetico riveste un ruolo fondamentale.

1.1 Cambiamenti Climatici: problematiche e ripercussioni in agricoltura ed orticoltura

L'agricoltura di oggi e anche quella del futuro si trova di fronte ad una sfida epocale: da un lato si pone l'obiettivo di sfamare i quasi 9 miliardi di persone nel mondo, dall'altro lo deve fare senza andare a depauperare ulteriormente le risorse naturali, già fortemente minacciate. Inoltre, la produzione agricola deve rispondere alle esigenze di resilienza imposte dalla minaccia dei cambiamenti climatici, in parte causati dalla stessa agricoltura.

La comunità scientifica è concorde nell'affermare che il cambiamento del clima sia causato in gran parte dalle attività umane, in cui sono comprese le attività agricole.

La FAO stima che dal 2000 al 2018, almeno il 90% della deforestazione nel mondo, sia da attribuire all'espansione delle attività agricole.

A livello nazionale, l'agricoltura rappresenta circa il 7% delle emissioni di gas ad effetto serra (Cristofaro, 2020).

La comunità scientifica è concorde nell'affermare che il cambiamento del clima sia causato in gran parte dalle attività umane, in cui sono comprese le attività agricole.

Ma, al tempo stesso, l'agricoltura subisce pesantemente il riscaldamento globale.

A titolo di esempio vengono riportati alcuni dati che meglio aiutano a delineare il quadro generale:

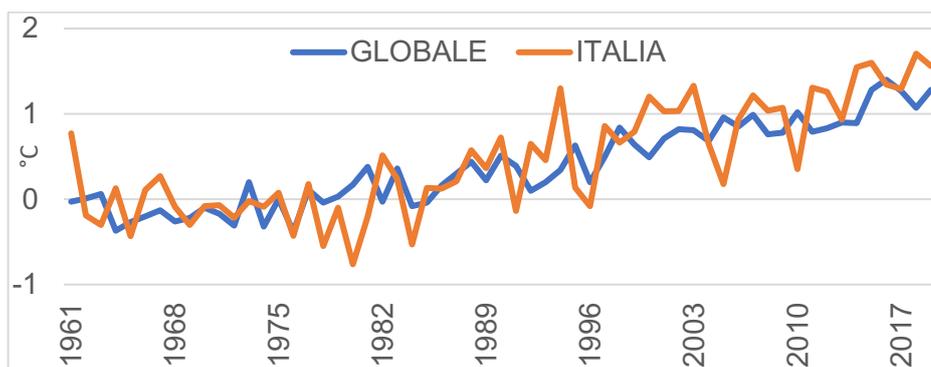


Tabella 1-1: Anomalie di temperatura media in Italia dal 1961 al 2019 in confronto al resto del mondo (ISPRA - Istituto Superiore per la Protezione e la ricerca Ambientale, 2021)

La Tabella 1 mostra le anomalie delle temperature medie in Italia e nel mondo, nel periodo che va dal 1981 al 2019; si evince chiaramente come il nostro Paese si stia scaldando in media di più del resto del globo.

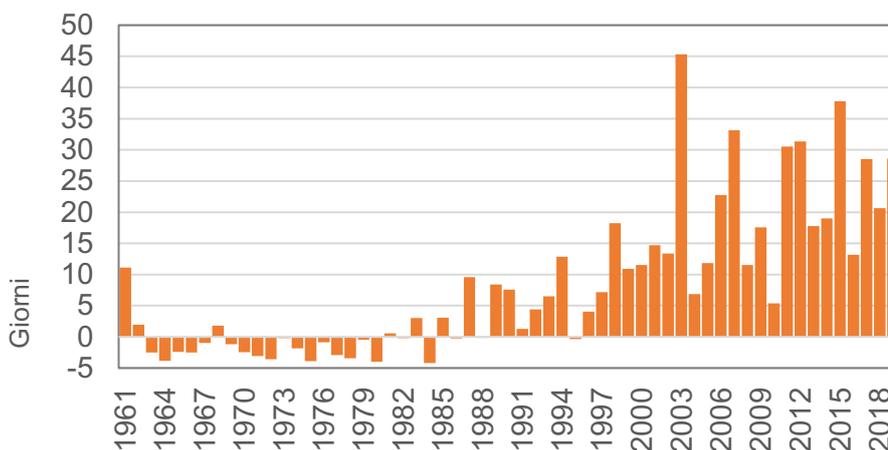


Tabella 1-2: Anomalie medie annuali del numero di giorni con onde di calore in Italia rispetto al valore normale 1961 – 2019 (ISPRA - Istituto Superiore per la Protezione e la ricerca Ambientale, 2021).

La Tabella 2 mostra la media dei giorni con ondate di calore nello stesso periodo di riferimento: anche questo dato indica come le temperature stiano salendo notevolmente in Italia.

In tale scenario l'agricoltura e specificamente l'orticoltura italiana sono sottoposte a nuove problematiche. Ad esempio, la coltivazione delle brassiche vede spesso enormemente cambiati i propri cicli colturali; infatti, capita che, a seguito di forti ondate di calore o di periodi di siccità, (come l'estate 2021), i cicli di produzione vengano completamente disattesi, con varietà da 85 gg che arrivano a maturazione in 120 gg o più. Ciò determina, da un lato carenza di prodotto nei periodi preventivati per la raccolta con eventuale copertura delle mancanze con prodotti esteri, dall'altro si ha un surplus di produzione in altri momenti della stagione, con conseguente abbassamento dei prezzi e deterioramento della profittabilità delle az. Agricole.

Altro esempio, nella coltivazione dello zucchini, i frequenti ritorni di freddo delle primavere degli ultimi anni, stanno determinando una forte problematica nell'emergenza delle semine più precoci, costringendo, di fatto gli agricoltori a riseminare o in qualche caso a posticipare alcuni periodi di semina causando carenza di produzione.

1.2 Evoluzione delle nuove fisiopatie e avversità

L'orticoltura è da sempre impegnata nel contrastare le malattie e i fitofagi che determinano importanti cali produttivi. Fin dall'avvento della "rivoluzione verde", la lotta alle avversità è stata condotta grazie all'ausilio delle molecole di sintesi contenute negli agrofarmaci; insieme alle resistenze genetiche, ne permettevano un efficace controllo.

Dopo anni di uso indiscriminato di tali sostanze, unito alla poca rotazione dei terreni e alla massiccia perdita di biodiversità dell'agroecosistema stanno inducendo reazioni di resistenza sia negli insetti, che nelle malattie fungine e batteriche. Ad esempio, la *Bremia lactucae* evolve spesso in nuove razze via via più aggressive; proprio a giugno 2021 è stata ufficializzata la nuova razza 37 (International Bremia Evaluation Board Europe (IBEB-EU), 2021).

Anche le virosi, da sempre considerate fisiopatie in grado di nuocere gravemente alle colture se non contrastate adeguatamente, stanno intensificando la loro aggressività.

Il movimento delle merci, lo spostamento delle persone, l'arrivo sul mercato nazionale di produzioni estere, facilitano l'ingresso in Europa e in Italia di virus alieni; ne è un esempio recente l'allarme riguardante il "Tomato brown rugose fruit virus – ToBRFV" (CREA Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, 2020) (European and

Mediterranean Plant Protection Organization (EPPO), 2020) identificato per la prima volta in Giordania nel 2015, è in grado di attaccare le solanacee con perdite importanti di produzione, o il “Tomato Leaf Curl New Delhi Virus – ToLCNDV” (EPPO , 2015) che attacca principalmente lo zucchini, e che è divenuto endemico nel sud Italia e nella zona di Latina.

1.3 Le varietà orticole e il rischio di uniformazione

Come già accennato, in Italia operano oltre 60 società sementiere, molte delle quali hanno un proprio programma di ricerca e breeding. Queste società private hanno la capacità economico/finanziaria di sviluppare programmi di miglioramento varietale complessi e variegati, in modo da offrire al mercato soluzioni innovative.

Quando i programmi di breeding vengono condotti direttamente nei Paesi nei quali andranno commercializzate le varietà, si ottengono risultati in grado di adattarsi agli ambienti pedoclimatici di destinazione.

Capita spesso, però, che alcune società, per meglio gestire i costi, focalizzino la ricerca in pochi Paesi, per poi “esportare” le varietà in giro per il mondo.

Dagli studi di genetica, sappiamo che i caratteri quantitativi (che verranno spiegati nel prossimo capitolo, e che sono i più interessati dai programmi di breeding), sono influenzati anche dall’ambiente esterno in cui le piante sono poste. In un Paese come l’Italia, caratterizzato da orografia e morfologia uniche, posto all’interno di un mare chiuso e quindi soggetto a condizioni climatiche anche molto diverse, questa scelta non è sempre vincente, in quanto le condizioni pedoclimatiche italiane potrebbe non garantire un efficace espressione genetica della varietà. Ad esempio, varietà di lattuga iceberg selezionate in Paese mediorientali, caratterizzati da climi caldo asciutti, che non hanno resistenze genetiche, mal si adatteranno alle condizioni climatiche italiane e alle pressioni delle malattie tipiche delle nostre zone.

Nonostante le performance produttive possano essere notevoli, si sarà costretti a intervenire con un numero più alto di trattamenti fitosanitari, con aggravio di costi economici e rischi per la salute per l’agricoltore, ambientali per l’ecosistema e di salubrità per il consumatore.

1.4 Le politiche agricole comunitarie

1.4.1 La Strategia Farm To Fork

La commissione Europea, all’interno dello European Green Deal varato nel 2020, ha previsto una strategia per la produzione di cibo equo, salutare e sostenibile chiamata Farm To Fork. Fra gli obiettivi che il piano si è prefissato, c’è quello della riduzione dell’utilizzo del

50% degli agrofarmaci in agricoltura e del 50% di quelli classificati come più pericolosi entro il 2030; inoltre si auspica la conversione ad agricoltura biologica del 25% del totale europeo delle superfici destinate alla produzione agricola. (Commissione Europea, 2020) Questa direzione, pur prefiggendosi obiettivi sfidanti ma giusti, potrebbe creare degli svantaggi alle nostre produzioni.

Già oggi chi fa orticoltura vede affermarsi avversità biotiche sempre più aggressive (vedi paragrafo precedente), e si ritrova molto spesso con prodotti fitofarmaci non in grado di garantire un'efficacia elevata (ne è un esempio la lotta alle virosi che prevede solo trattamenti preventivi, o ai funghi terricoli come *Fusarium spp.*). Anche la conversione al bio preclude l'impiego di soluzioni di difesa a base di prodotti chimici di sintesi.

Diventa, pertanto, fondamentale un approccio integrato alla difesa, dove il miglioramento genetico, grazie alle resistenze genetiche delle varietà, riveste un ruolo tutt'altro che secondario.

1.4.2 Nuova Politica Agricola Comunitaria

Anche la nuova PAC, prevista per il periodo 2023/2027, ha nei suoi obiettivi la sostenibilità ambientale e la riduzione delle emissioni, con pagamenti più alti per coloro che sceglieranno di intraprendere soluzioni ecocompatibili. Anche se i vari stati membri stanno ancora discutendo sui tecnicismi, la strada sembra segnata, pertanto il breeding può rappresentare un valido aiuto nel perseguire gli obiettivi prefissati (Commissione Europea, 2020).

1.5 Tendenze dei consumatori e del mercato

Da diversi anni, ormai, si assiste al cambio di tendenza nelle scelte dei consumatori italiani, nel settore orticolo.

Accanto alla sempre crescente richiesta di prodotti bio oggi le grandi catene di distribuzione sono sempre più interessate al made in italy, con particolare attenzione alla produzione locale e alle tipicità dei vari luoghi di origine. (Ismea, 2021)

Si pone, anche, particolare attenzione a tutta quella produzione in convenzionale che ha un occhio di riguardo nei confronti della sostenibilità (ne sono un esempio i prodotti ortofrutticoli a marchio COOP o i prodotti "Verso Natura Conad").

Tutto ciò si ripercuote nell'aumento di prodotti italiani che siano il più possibile freschi, salubri, con pochi o senza residui e che pongano particolare attenzione alla sostenibilità ambientale, sociale ed economica di chi li produce.

Capitolo 2

LA SELEZIONE NEL MIGLIORAMENTO GENETICO

Le considerazioni del capitolo 1 sottolineano come un buon programma di miglioramento genetico sia imprescindibile per la buona riuscita di una produzione orticola moderna e funzionale.

La selezione è il cuore stesso del concetto di Plant Breeding.

2.1 La selezione da un punto di vista genetico

Ogni volta che il breeder sceglie nelle popolazioni di piante gruppi di individui che sono migliorativi per il carattere voluto, di fatto opera una *selezione*, che lo porterà ad operare degli incroci per ottenere generazioni successive con i tratti desiderati.

La selezione, però, non può essere effettuata solamente sulla base del fenotipo, ma deve tenere conto di come i caratteri voluti siano controllati a livello genetico.

Esistono fondamentalmente due tipologie di controllo genetico: i caratteri monogenici e i caratteri poligenici.

2.1.1 Caratteri monogenici

Sono anche detti qualitativi, poiché la loro espressione si presenta con stati discreti e non continui. Essi sono generalmente ascrivibili a uno o pochi geni quindi con numero di loci limitato.

Le differenze fenotipiche di questi caratteri sono facilmente individuabili, e normalmente prevedibili in quanto si trasmettono nelle generazioni secondo le leggi di Mendel e l'ambiente esterno non influenza la loro espressività.

Di solito non sono oggetto di miglioramento poiché non rivestono importanza economica.

Un esempio è il colore del tegumento nelle specie di pisello (carattere indagato dallo stesso Mendel). (Lorenzetti, et al., 2018) (Ceccarelli, et al., 1980)

2.1.2 Caratteri poligenici o Quantitativi

I caratteri quantitativi sono portati da un numero alto di geni e sono a variabilità continua (i fenotipi non possono essere classificati entro classi discrete, ma possono essere descritti e misurati mediante parametri numerici).

L'alto numero dei loci coinvolti e il piccolo controllo che ciascuno di essi, preso singolarmente, esercita nel carattere, fa sì che l'ambiente possa influenzare notevolmente la sua espressione.

Il quadro è ulteriormente complicato a causa del tipo di azione genica, che nel caso di caratteri quantitativi, può essere additiva, di dominanza o di epistasi tra loci diversi; per questo motivo tali caratteri non hanno andamento Mendeliano.

Preso un carattere quantitativo, una popolazione avrà una distribuzione di tipo Gaussiano, dove sull'asse delle ascisse verranno indicati i valori fenotipici e sull'asse delle ordinate le frequenze; sarà inoltre possibile visualizzare graficamente il valore medio della popolazione che corrisponderà al punto più alto della curva.

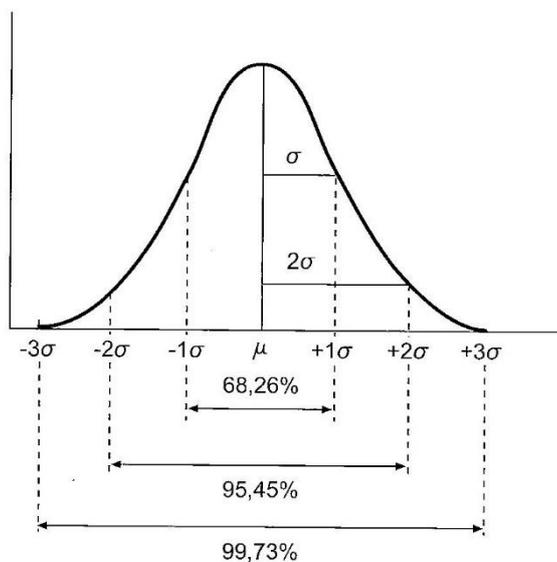


Figura 2-1: Curva normale di una popolazione teorica con media μ e deviazione standard σ (Ceccarelli, et al., 1980)

Dalla Figura 1 si può dedurre che oltre il 68% delle popolazioni si trovano fra $\mu \pm \sigma$ e il 95% circa si trova compreso fra $\mu \pm 2\sigma$, mentre solo il 4% della popolazione si troverà ai margini della curva.

I caratteri quantitativi sono inseriti nei programmi di breeding in quanto spesso di importanza economica (es. la resa, l'habitus vegetativo o le resistenze a malattie). (Ceccarelli, et al., 1980)

2.1.3 Selezione per caratteri quantitativi

Lo scopo della genetica dei caratteri quantitativi non è quello di indagare il numero dei geni coinvolti nell'espressione del carattere, quanto quello di restituire gli effetti in termini statistici di tali geni (media, varianza, correlazione e regressione).

Pertanto lo scopo della selezione è quello di spostare nella direzione voluta la media del carattere della popolazione.

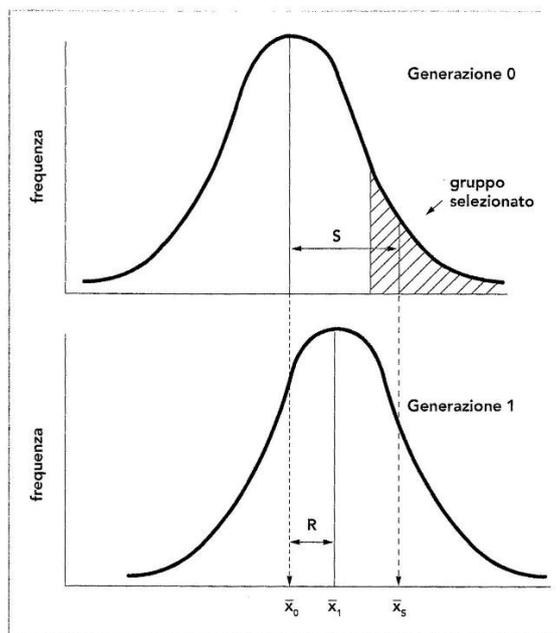


Figura 2-2: Differenziale di selezione (S) (Lorenzetti, et al., 2018)

Affinché la selezione dia buoni effetti, occorre scegliere una quota di individui, all'interno della popolazione di riferimento che abbia una media (\bar{x}_s) differente dalla media della popolazione originaria (\bar{x}_0). La differenza fra le due medie sarà chiamata Differenziale di Selezione (S) ($S = (\bar{x}_s - \bar{x}_0)$).

Andando ad incrociare gli individui selezionati si otterrà, per la maggior parte, una progenie con una media (\bar{x}_1) che sarà compresa fra le media del gruppo selezionato e la media della popolazione originaria, dando come risultato la risposta della selezione (R) o guadagno conseguito con la selezione ($R = (\bar{x}_1 - \bar{x}_0)$) (Figura 2). (Lorenzetti, et al., 2018)

La risposta della selezione (R) sarà sempre inferiore al differenziale di Selezione (S) poiché le differenze fra gli individui delle varie popolazioni sono basate sull'osservazione fenotipica. I caratteri quantitativi, però, sono influenzati dall'ambiente e non sono pienamente ereditabili: la loro risposta alla selezione dipenderà quindi sia da S che dalla ereditabilità (h^2) (Ceccarelli, et al., 1980).

Per meglio interpretare gli effetti della selezione si ricorre a modelli matematici e statistici: uno dei più utilizzati per l'analisi della varianza è il modello ANOVA.

L'ANOVA ha come obiettivo quello di determinare se la variazione tra le medie campionarie dei gruppi sia maggiore di quella attesa per effetto del caso (Schluter & C., 2010), cioè è in grado di stimare di quanto il genetista sia riuscito a spostare la media della nuova generazione, con la selezione, senza includere spostamenti dovuti alla casualità (o all'errore); è pertanto uno strumento fondamentale per il breeder.

2.2 Limiti genetici della selezione

La selezione produce un miglioramento nel/nei carattere/i desiderati finché tutti gli alleli favorevoli risultino fissati.

La fissazione degli alleli varia, come visto, a seconda se i caratteri sono qualitativi o quantitativi.

Nel caso dei monogenici, la selezione sarà tanto più efficace, quanto più si riesca a riconoscere l'omozigosi che interessa dagli altri genotipi; ad esempio la dominanza rallenta il processo di fissazione degli omozigoti e ancora di più se si presenta l'epistasi.

Ad ogni modo la fissazione di caratteri monogenici si raggiunge in tempi relativamente brevi.

Nella fissazione di caratteri poligenici, invece, la risposta alla selezione (R) nella popolazione ottenuta da interincrocio del gruppo selezionato, dipende in larga misura dalla variabilità genetica del gruppo di partenza e dalla sua utilizzazione.

Occorre ricordare come la selezione determini cambiamenti nella frequenza dei singoli geni e della loro combinazione: inoltre la selezione consente l'ottenimento di combinazioni nuove.

Tali fatti determinano delle modificazioni che possono essere riassunte in

- Cambiamenti delle frequenze geniche, le quali conducono anche a cambiamenti delle frequenze genotipiche e fenotipiche del carattere in selezione, con la conseguente comparsa di nuova media nella popolazione;
- Comparsa di nuovi genotipi;
- Cambiamento nella varianza (variabilità) della popolazione. (Lorenzetti, et al., 2018)

In linea di massima le risposte alla selezione più comuni che si attendono nella selezione di caratteri quantitativi sono raggruppabili in 6 gruppi:

1. Risposta iniziale rapida seguita da un lungo periodo di risposta lenta;
2. Risposta lenta e costante per un lungo periodo;
3. Risposta lenta fino a raggiungere limite oltre il quale non si avrà più nessun miglioramento;
4. Risposta limitata o nulla;
5. Risposta iniziale rapida fino ad un valore che rimane stabile per alcune generazioni, per poi riprendere a salire fino al raggiungimento di un nuovo limite;
6. Risposta asimmetrica nella selezione divergente.

La figura 3 riporta in maniera grafica le 6 risposte sopra menzionate.

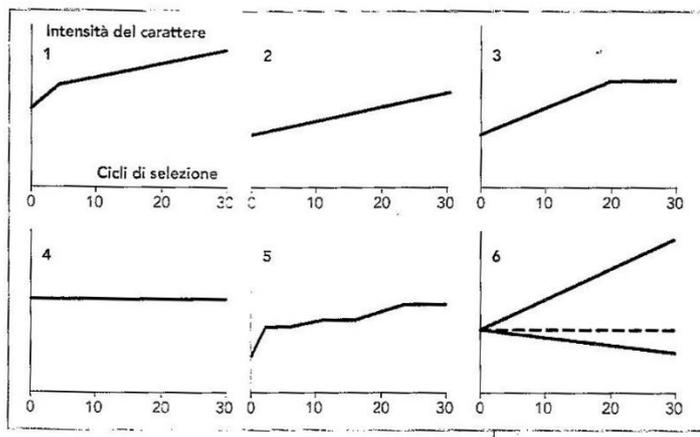


Figura 2-3: Risposta alla selezione (Lorenzetti, et al., 2018)

Capitolo 3

OPERARE LA SELEZIONE: STRUMENTI MOLECOLARI

Come si è visto la selezione è fondamentale nella gestione del miglioramento genetico. Quando si ha, però, a che fare con caratteri poligenici la sola selezione su base fenotipica non conduce a risultati soddisfacenti, in quanto la pressione dell'ambiente può incidere notevolmente nell'espressione genica.

La strada da percorrere è quella che prevede lo studio del genoma, che è in grado di trasferire i caratteri da una generazione all'altra, in modo più o meno marcato (ereditabilità).

In questa ottica, gli strumenti ideali per lo studio della parte genetica di un fenotipo sono i marcatori molecolari.

3.1 Marcatori molecolari

I marcatori molecolari sono frammenti di DNA genomico chiaramente definiti. Essi rappresentano dei caratteri ereditabili e polimorfici a livello di specie, di popolazione e di individui.

I marcatori non sono direttamente ascrivibili a geni ma si basano sulla variabilità allelica della sequenza del DNA genomico fra gli individui. Lo studio di tali differenze conduce ad un'analisi estremamente precisa, in grado di dare un gran numero di informazioni.

La costruzione di un marcatore si basa sulla possibilità di creare frammenti di DNA grazie a tecniche che utilizzano enzimi di restrizione e PCR. Considerando che il DNA genomico è piuttosto ampio, teoricamente si possono creare un gran numero di marcatori.

Chiaramente ogni marcatore dovrà essere identificato e occorrerà conoscerne l'inizio, la fine e la lunghezza, oltre ad avere la caratteristica delle ripetibilità.

Le diverse forme di un marcatore si dicono alleli, prendendo in prestito questo termine dalla genetica classica.

Quando due forme di marcatori (cioè due alleli) sono uguali, si dicono omozigoti, mentre se sono diversi sono eterozigoti.

I marcatori che permettono di identificare sia individui omozigoti che eterozigoti si dicono *codominanti*, viceversa quelli che danno risposte solo di presenza/assenza sono *dominanti*.

Un buon marcatore molecolare dovrebbe essere polimorfico, codominante, ripetibile, facilmente individuabile e con un meccanismo di trasmissione di tipo mendeliano.

I marcatori consentono lo studio del genoma da cui derivano numerose applicazioni, fra le quali si annoverano la costruzione di mappe genetiche, la genotipizzazione (genetic fingerprinting) e la selezione assistita da marcatori molecolari (MAS).

La loro classificazione si basa essenzialmente sul tipo di tecnica utilizzata. Al momento esistono due tecniche di base:

- La *Southern blot hybridation*: che si basa sull'impiego di enzimi di restrizione (endonucleasi), i quali riconoscono specifiche sequenze del DNA, per poi tagliarla, producendo una quantità di frammenti con lunghezze nucleotidiche diverse. Le differenziazioni dei segmenti tagliati vengono eseguite mediante specifiche sequenze di DNA marcate con isotopi radioattivi (di solito ^{32}P) legati ad un nucleotide
- *PCR, Polymerase Chain Reaction*: tecnica che permette di amplificare in vitro delle specifiche sequenze di DNA in poche ore, utilizzando delle DNA polimerasi termostabili (Lorenzetti, et al., 2018).

3.2 Tipologie di Marcatori Molecolari

Esistono diversi tipi di Marcatori Molecolari derivanti dalle tecniche sopra illustrate:

3.2.1 RFLP

I marcatori RFLP (Restriction Fragment Length Polymorphism) utilizzano la tecnica del Southern blot hybridation. La capacità di riconoscere entrambe le sequenze alleliche negli individui eterozigoti, grazie alla presenza degli isotopi, dona a questa tipologia la caratteristica di essere codominante, Sono ormai in disuso a causa dell'alto costo di produzione e delle poche informazioni che riescono a fornire. (Young, 1992)

3.2.2 AFLP

La tecnica di produzione degli AFLP (Amplified Fragment Length Polymorphism) prevede due fasi: nella prima si utilizzano due enzimi di restrizione che tagliano il DNA; uno riconosce una porzione di 6 basi (esempio EcoR1) l'altro di 4(ad esempio Mse1). Entrambe gli enzimi

lasciano alle estremità dei frammenti delle sporgenze dette coesive (sticky ends). A tali sporgenze vengono legati degli oligonucleotidi (adattatori) in modo da non riformare i siti di restrizione. Si ottengono così un gran numero di frammenti i quali saranno stati generati da enzimi da 6 basi, 4 basi o entrambe (una per ogni estremità del frammento). La frequenza dei frammenti non è costante perché i siti di restrizione da 6 basi sono mediamente meno frequenti nei genomi

Nella seconda fase vengono effettuati due cicli di PCR dove i frammenti vengono amplificati grazie a primer specifici.

Questa tecnica è di tipo dominante e rileva polimorfismi dovuti a mutazioni puntiformi nei siti di restrizione. Il vantaggio è quello di poter analizzare contemporaneamente un gran numero di loci in un unico esperimento, ed inoltre non necessita di informazioni preliminari sulla sequenza del genoma. (Lorenzetti, et al., 2018)

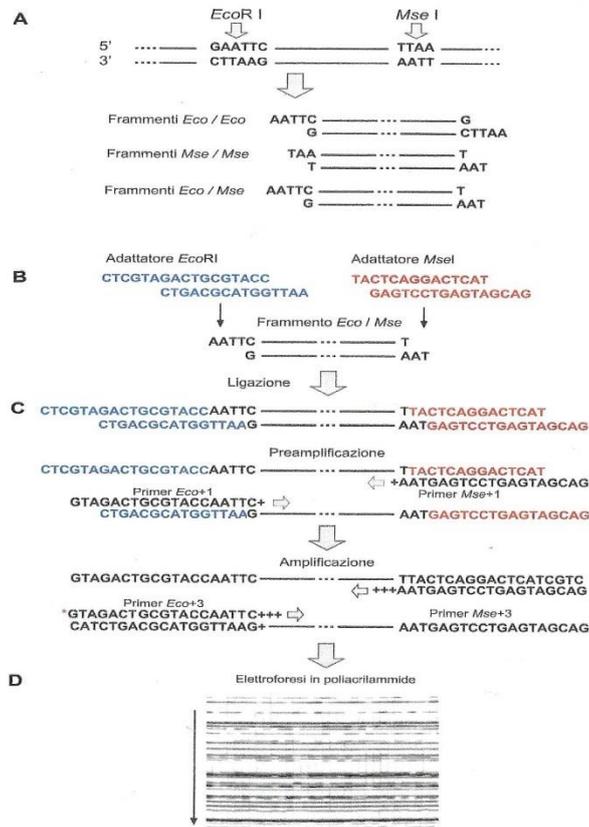


Figura 3-1: Schema del processo di AFLP (Lorenzetti, et al., 2018)

3.2.3 *SSR o Microsatelliti*

Gli SSR (Single Sequence Repeats o STR Short Tandem Repeats) o Microsatelliti sono sequenze molto corte (1-6 nucleotidi) ripetute in tandem, presenti in tutti i genomi. Sono formati da ripetizione di una coppia di nucleotidi (ad esempio AT) ripetute n volte. È proprio la variabilità del numero di ripetizione delle coppie (n) a determinare il polimorfismo.

Grazie alla PCR sono diventati uno degli strumenti più utilizzati come marcatori, specialmente per la creazione di mappe genetiche e per la caratterizzazione varietale.

L'elevato polimorfismo dovuto alle differenze di ripetizioni della sequenza di base comporta la presenza di numerosi alleli per locus; questa grande varietà di differenze sono dovute a errori di replicazione o ad crossing over integrale nelle regioni SSR.

L'analisi più utilizzata per determinare i microsatelliti è la PCR, seguita da elettroforesi.

Sono marcatori codominanti molto usati per il fingerprinting di individui singoli. (Landi, 2008) (Gargani, s.d.)

3.2.4 *SNP*

Gli SNP (Single Nucleotide Polymorphism) sono marcatori di polimorfismo dovuti ad un singolo nucleotide. Una posizione del genoma in cui è possibile trovare una differenza di un solo nucleotide forma un SNP. Quando tale differenza ha frequenza maggiore dell'1% nella popolazione o fra popolazioni, non è più considerata un effetto del caso o dell'errore e può essere utilizzata come SNP. (Morin, et al., 2004)

Non sono considerati SNP le differenze nel genoma dovute a inserzione o delezione ma solamente quelle dovute a sostituzioni, le quali sono abbondanti nel genoma ed hanno un tasso di mutazione relativamente basso.

Gli SNP sono molto abbondanti nel genoma, ad esempio il mais ne presenta uno ogni 100bp, soprattutto nelle zone non codificanti, ma spesso si trovano associate a regioni codificanti. (Lorenzetti, et al., 2018)

Per la loro natura sono molto utilizzati per associare il polimorfismo di sequenza con caratteri specifici o per indagini filogenetiche, ma soprattutto sono usati in gruppi adiacenti per definire aplotipi, cioè combinazioni di più alleli marcatori strettamente associati che caratterizzano uno specifico tratto cromosomico.

Gli aplotipi sono interessanti poiché risulta più semplice associare geni con tratti cromosomici piuttosto che con singoli SNP. (Lorenzetti, et al., 2018)

Esistono diverse tecniche di genotipizzazione di SNP; la scelta di utilizzarne una piuttosto che un'altra dipende sia dal numero di campioni da analizzare sia dal numero di SNP da rilevare.

Nel caso di analisi di pochi SNP su pochi individui il costo è relativamente basso, mentre se l'analisi prevede un altro numero di SNP su di un numero elevato di individui aumentano sia la precisione della risposta, sia il costo.

Capitolo 4

SELEZIONE ASSISTITA DA MARCATORI E SELEZIONE GENOMICA

I marcatori molecolari sono gli strumenti utili alla costruzione delle mappe genetiche e per il mappaggio dei geni qualitativi, quantitativi e loro impiego (QTL).

Essi sono fondamentali anche per operare la selezione assistita (MAS) e la Selezione Genomica (GS).

4.1 Costruzione di mappe genetiche e mappaggio di geni

Una mappa genetica è una rappresentazione grafica della distanza che separa i geni.

La costruzione delle mappe utilizza i marcatori molecolari, andando a determinare la loro posizione relativa lungo i cromosomi. Le posizioni relative vengono stimate sulla base delle frequenze di ricombinazione meiotica che si basano sulle distanze espresse in centimorgan. Pertanto per la determinazione delle mappe genetiche bisogna incrociare due individui di genotipo diverso, con marcatori polimorfici, allevare una popolazione segregante (es. F_2) ed estrarre il DNA genomico di singoli individui per analizzare la segregazione dei marcatori polimorfici nei due parentali (ad esempio nel cromosoma 1 del pomodoro due caratteri Elasticity *ela1_1* e TG17 sono fra loro distanti 2,10 cM (Network Sol Genomics, s.d.)).

Esistono due categorie di mappaggio:

- Mappaggio tramite linkage: che sfrutta popolazioni sperimentali ottenute per incrocio fra due o più individui che differiscono per il carattere o i caratteri interessati. Tali popolazioni sono derivate da genitori geneticamente diversi e lontani e si trovano in disequilibrio di linkage (LD). Il disequilibrio rappresenta la condizione per cui due o più alleli in loci diversi non si trovano associati in maniera casuale, e quindi vengono ereditati insieme con frequenze di ricombinazione conosciute e stimate. Ciò consente di valutare il livello di risoluzione a cui è possibile mappare i loci che interessano; (Slatkin, 2008)
- Mappaggio tramite associazione: che prevede l'utilizzo di popolazioni naturali o popolazioni in cui il tasso di parentela è ignoto. In questa situazione la frequenza

di ricombinazione fra il/i gene/i non è conosciuta e il livello di LD tra geni e marcatori va stimato empiricamente. Questa tipologia viene utilizzata quando non è possibile ottenere popolazioni sperimentali (es. specie appartenenti a collezioni di germoplasma). Il fatto di adoperare popolazioni naturali non derivanti da incroci di individui geneticamente conosciuti, consente l'esplorazione di una maggiore diversità genetica.

Quando si procede a mappare un gene o più geni, si determinano i marcatori (che sono effettivamente da considerare alleli) il più possibile vicino al gene di interesse, in quanto il marcatore avrà maggiore probabilità di essere ereditato insieme al gene/i che interessa.

Ciò comporta che nelle generazioni successive la presenza del/dei marcatore/i associato/i al gene/i, darà una probabilità maggiore di avere ereditato anche il gene/i di interesse.

4.2 QTL e mappaggio

Un QTL (Quantitative Traits Locus) è una specifica regione del cromosoma che contiene uno o più geni in grado di determinare una variazione fenotipica di un carattere quantitativo.

Gli individui all'interno della popolazione che sono scelti per studiare il carattere di interesse, devono essere valutati sia per il fenotipo che per il genotipo, con un numero sufficientemente alto di marcatori molecolari.

Le tecniche di mappaggio dei QTL si dividono, anch'esse, in mappaggio per linkage e mappaggio per associazione, e sono del tutto simili a quelle indicate precedentemente.

4.2.1 Mappaggio di QTL per Linkage

Come già parzialmente descritto, il mappaggio per linkage sfrutta una popolazione selezionata segregante per ottenere un'associazione statistica tra il QTL e i genotipi ai loci marcatori.

Questo processo prevede quattro fasi distinte che possono essere così riassunte:

1. Creazione di una popolazione sperimentale segregante;
2. Uso di marcatori per la genotipizzazione;
3. Fenotipizzazione del carattere/i quantitativo preso in esame;
4. Utilizzo di test statistici adeguati a stabilire la co-segregazione fra QTL e marcatori (fig. 4-1).

Le popolazioni segreganti e i test statistici per dimostrare le associazioni, possono essere diversi. (Collard & Mackill, 2007)

Nel caso di piante annuali con riproduzione per seme, come sono le orticole, normalmente si usano popolazioni biparentali, con successivi incroci, autofecondazione e reincroci (F_2 o BC_1 = Back Cross, reincrocio). Da tali generazioni si possono derivare linee o famiglie di piante sulle quali mappare i QTL.

Una tipologia di popolazione particolarmente adatta al mappaggio di QTL per questo genere di piante è la RIL.

La popolazione RIL (Recombinant Inbreed Lines, linee inbreed ricombinati), deriva da una serie di autofecondazioni di piante F_2 , fino ad ottenere un'alta omozigosi. (Huang, et al., 2011)

Da ogni individuo F_2 autofecondato, si ottengono linee omozigotiche differenti, tanto da avere come risultato una collezione di ricombinanti.

Dal punto di vista delle analisi statistiche, per quanto riguarda il mappaggio di QTL per linkage il test maggiormente adoperato è l'ANOVA.

Questo tipo di mappaggio può rendersi utile nel caso di creazione di ibridi commerciali, laddove siano disponibili i parentali, in modo da poter sfruttare al massimo l'eterosi (Lorenzetti, et al., 2018).

4.2.2 Mappaggio di QTL per associazione

Il mappaggio per associazione prevede l'uso di individui non imparentati derivanti da popolazioni naturali o collezioni di germoplasma, nei quali sono sconosciute il posizionamento dei QTL e dei marcatori nei parentali.

Dato questo presupposto il metodo migliore per il mappaggio consiste nell'utilizzare quanti più marcatori possibili distribuiti in maniera omogenea sul genoma; questo perché il numero dei marcatori e la probabilità che uno o più di essi siano vicini ai QTL è direttamente proporzionale; di conseguenza anche il livello LD aumenta con l'aumentare del numero dei marcatori.

Pertanto il livello di LD determina il numero di marcatori da impiegare negli studi di associazione.

Il livello di LD si forma principalmente per selezione e mutazione, viceversa la ricombinazione lo limita.

Quindi il mappaggio per associazione non fa altro che utilizzare il residuo di LD derivante dagli incroci, tra i marcatori e i QTL per mappare quest'ultimi. (fig. 4-1)

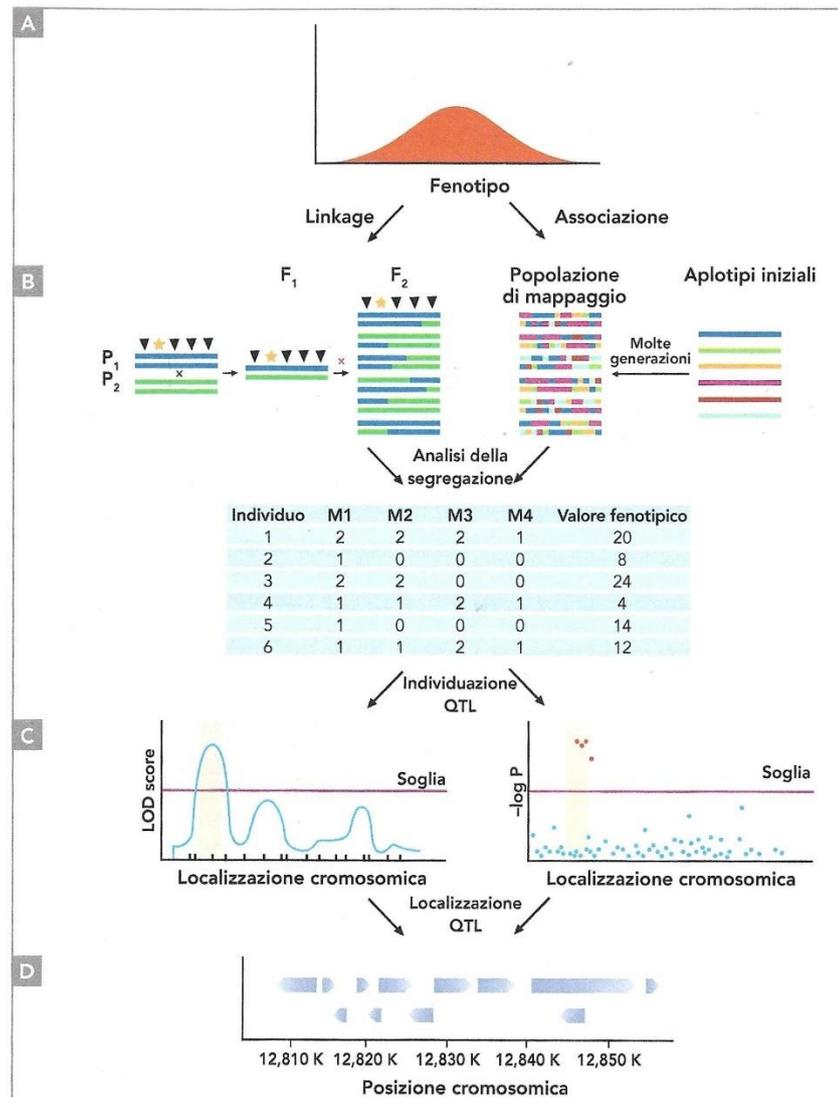


Figura 4-1: Schema dei due sistemi di mappaggio QTL: A) entrambe le popolazioni mostrano una distribuzione gaussiana del QTL; B) il QTL (stella) segrega insieme ai marcatori (M1, M2...); a sinistra il mappaggio linkage, a destra il mappaggio per associazione; C) le analisi statistiche consentono di individuare la probabilità (linea o puntini blu) il QTL nel cromosoma; la regione sopra la linea di soglia rappresenta la zona del cromosoma con più alta probabilità di presenza del QTL (intervallo di supporto del QTL); D) rappresentazione schematica del contenuto dei geni della regione cromosomica che contiene il QTL.

4.2.3 *Analisi del mappaggio per associazione*

Il mappaggio per associazione prevede due approcci generali.

Nel primo, chiamato “studio di associazione al gene candidato”, vengono determinati il/i gene/i che potrebbero controllare l’espressione del carattere di interesse.

In seguito vengono studiate le associazioni statistiche fra le varianti alleliche ed il fenotipo.

Il secondo metodo è chiamato “studio di associazione sull’intero genoma” o “Genome-wide association study, GWAS”.

4.2.4 *GWAS, un approccio innovativo nel mappaggio dei QTL*

La GWAS (Genome-Wide Association Study) è un approccio nell’analisi del mappaggio di QTL per associazione, che non prevede lo studio di uno o pochi geni di interesse, ma, l’impegno di tutto il genoma. (Brachi, et al., 2011)

I marcatori che rendono possibile questo modello sono i SNP, i quali sono in numero molto alto e sono distribuiti nell’intero genoma.

Nelle piante il mappaggio dei QTL era originariamente effettuato tramite incroci biparentali (come nel caso del linkage), ma, come visto, questo approccio ha valenza solo nel caso di diversità allelica e quindi ha una risoluzione genomica limitata.

La GWAS permette di superare queste limitazioni procurando un’alta risoluzione (spesso a livello di singoli geni) usando individui che provengono da studi di popolazioni precedenti.

L’utilizzo di alte densità di SNP permette la scansione dell’intero genoma alla ricerca degli aplotipi che sono correlati con i caratteri quantitativi. (Brachi, et al., 2011) (Uffelma, et al., 2021)

I test statistici utilizzati per i caratteri a variabilità continua nella GWAS, sono il già citato ANOVA o la regressione lineare.

Il mappaggio per associazione presenta, però, una problematica: ovvero la presenza di “Struttura” o “stratificazione della popolazione”.

La Struttura è una condizione per cui all’interno della popolazione in esame esistono due o più sottogruppi ciascuno dei quali presenta, al suo interno, individui imparentati.

Questo comporta che mediamente gli individui dello stesso sottogruppo, che hanno rapporti di parentela, presenteranno una probabilità più alta di avere gli stessi alleli condivisi a qualunque locus, determinando la co-segregazione di alleli a geni diversi anche in assenza di linkage. (Khan, et al., 2021)

Ad esempio se un fenotipo di interesse, viene espresso a seguito di una mutazione in un individuo appartenente a un certo sottogruppo, lo stesso tipo di fenotipo comparirà molto raramente nei sottogruppi che non presentano la mutazione o che abbiano una nuova mutazione per lo stesso allele. Pertanto, in uno studio GWAS gli individui che saranno analizzati, saranno maggiormente appartenenti al sottogruppo dove è presente la mutazione originaria, i quali sono imparentati tra loro.

Conseguentemente oltre a condividere l'allele mutato andranno a condividere anche alleli marcatori che non hanno nulla a che fare con il carattere di interesse. Ciò determina frequenze di associazione marcatori/QTL non di interesse elevate con alto numero di falsi positivi.

Come risultato finale dell'analisi GWAS, si ottiene un grafico, chiamato Manhattan Plot, (tabella 4-1) dove la posizione del marcatore sul cromosoma è data nell'asse delle ascisse, mentre la significatività statistica per ogni singolo marcatore è indicata nell'asse delle ordinate.

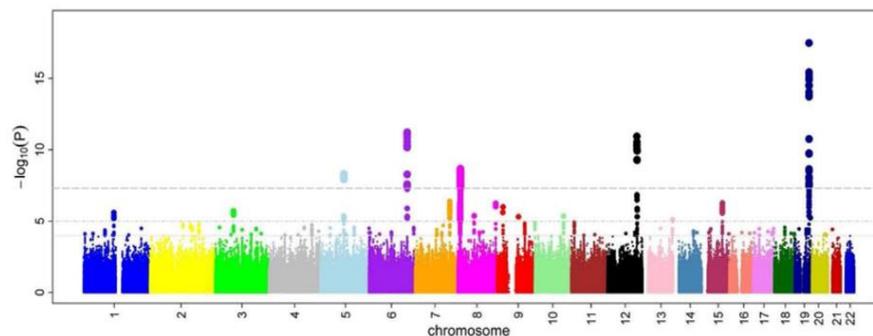


Tabella 4-1: Esempio di Manhattan Plot By M. Kamran Ikram et al - Ikram MK et al (2010) Four Novel Loci (19q13, 6q24, 12q24, and 5q14) Influence the Microcirculation In Vivo. PLoS Genet. 2010 Oct 28;6(10): e1001184. doi:10.1371/journal.pgen.1001184.g001, CC BY 2.5, <https://commons.wikimedia.org/w/index.php?curid=18056138>

4.2.5 Pregi e difetti del mappaggio di QTL per associazione

Il metodo per associazione è stato utilizzato in passato ampiamente nello studio del genoma umano, a causa dell'evidente indisponibilità di popolazioni sperimentali. Oggi tale metodologia è stata sdoganata anche nei programmi di miglioramento genetico delle piante.

I principali vantaggi del metodo per associazione sono:

- Applicabilità quando, per diverse ragioni, non si può disporre di popolazioni sperimentali;
- Studio di maggiore diversità genetica, soprattutto con l'analisi GWAS;
- Più alto potere di risoluzione grazie all'ausilio dei crossing over "storici" avvenuti nelle popolazioni naturali;

- Velocità di operazione, in quanto si utilizzano genotipi già disponibili senza dover attendere la creazione di popolazioni sperimentali;
- Talvolta disponibilità immediata di valori fenotipici perché raccolti in altri ambiti per altri motivi.

Gli svantaggi sono:

- Impossibilità di controllare le frequenze alleliche nella popolazione;
- Come conseguenza del punto precedente la scarsa possibilità di indentificare alleli rari potenzialmente di interesse;
- Struttura della popolazione con aumento di falsi positivi;

4.3 La MAS – Selezione Assistita da Marcatori

Il miglioramento genetico tramite selezione fenotipica ha permesso, in passato, di contribuire in maniera determinante all'incremento dei raccolti, mentre il mappaggio tramite marcatori dei caratteri maggiormente d'interesse, ha aperto un nuovo modo di costruire i programmi di miglioramento.

La necessità di incrementare la quantità del raccolto per sfamare una popolazione in costante crescita rappresenta una sfida senza precedenti per i genetisti e gli scienziati che si occupano di agricoltura.

Oggi e nel futuro prossimo la genetica delle piante avrà un ruolo determinante negli sforzi per aumentare la produzione agricola.

A causa delle problematiche viste nel capitolo 1, l'aumento di produzione dovrà necessariamente essere stabile e sostenibile, e le colture dovranno avere caratteri genetici in grado di conferire una duratura resistenza alle malattie, tolleranza agli stress abiotici, un incremento dell'efficienza nell'uso di acqua e nutrienti.

In questo contesto la selezione assistita da marcatori può massimizzare i tempi e i costi in modo tale che l'efficienza e la precisione della selezione possano essere fortemente aumentati.

La MAS (Marker Assisted Selection) è uno strumento dove la selezione degli individui è operata sulla base del genotipo attraverso i marcatori molecolari. (J.M. & D., 1998)

Tale tecnica può integrare la classica selezione basata sul fenotipo, permettendo un'efficienza maggiore del programma di miglioramento genetico riducendo i tempi e i costi.

Rispetto ad una selezione unicamente fenotipica, la MAS presenta diversi vantaggi:

- È più semplice ed ha meno costi con tempi più corti (un esempio è la ricerca del carattere di resistenza ai nematodi, che nella selezione fenotipica richiede tempi molto lunghi per essere verificata);

- La selezione può essere eseguita a livello di plantula (si pensi alla determinazione di un carattere nel corimbo di un cavolfiore a lungo ciclo, o alla resa in colore di un pomodoro determinato);
- Può essere selezionata anche una sola pianta: nella selezione classica le piante vengono allevate in plot replicati per minimizzare gli effetti dell'ambiente, con la MAS si vanno a ricercare i caratteri desiderati a livello genomico e quindi senza influenza ambientale);
- La MAS necessita di meno linee da testare e conseguentemente si ha meno bisogno di spazio, sia in serra, che in pieno campo.

Sulla base del livello di conoscenza dei loci che interessano la MAS può essere distinta in due macrocategorie:

1. *MAS per geni ad effetto maggiore;*
2. *Selezione genomica (GS);*

4.3.1 *MAS per geni ad effetto maggiore*

Questa tipologia di MAS si basa sulle informazioni ricevute dai marcatori associati a geni e/o QTL che hanno un effetto genetico rilevante.

Viene usata per la selezione di caratteri importanti geneticamente semplici (esempio resistenze a malattie o habitus vegetativo), ma per i quali la valutazione fenotipica risulterebbe lunga e costosa.

Affinché si possa attuare questa tipo di selezione occorre preliminarmente conoscere l'esatta posizione dei marcatori, i quali dovranno essere associati con le regioni QTL di interesse.

Occorre anche verificare l'effettivo effetto linkage fra i marcatori e i QTL, pena la mancanza di co-segregazione.

Le applicazioni maggiormente utilizzate sono:

- La selezione assistita per i materiali di miglioramento – Marker assisted evaluation of breeding material.

Questa applicazione può essere utilizzata prima degli incroci (ibridizzazione) o nello sviluppo delle linee. Infatti, con tale tecnica si possono verificare l'identità delle cultivar di partenza, la diversità genetica e la purezza delle linee.

Prima dell'avvento della MAS queste verifiche erano effettuate esclusivamente su base fenotipica e visiva con le tutte le limitazioni del caso.

- Il reincrocio assistito da marcatori – Marker-assisted backcross o MABC;

è un metodo utilizzato per incorporare uno o più geni portatori di caratteri migliorativi, in varietà preesistenti.

La tecnica consiste in una serie di reincroci a partire dalla generazione F_1 , originatasi da due linee parentali (P_1 e P_2). I genitori sono indicati come genitore Ricorrente, quando non possiede la caratteristica da inserire, ma proviene dalla varietà preesistente, e genitore Donatore, quando ha nel suo genoma, il tratto desiderato da aggiungere. Dalla generazione F_1 , il reincrocio (backcrossing) sarà effettuato per n volte con il genitore ricorrente, finché nel genoma finale non rimanga solo il carattere che interessa aggiungere proveniente dal genitore Donatore e siano mantenute tutte le caratteristiche volute del genitore ricorrente (fig. 4-2).

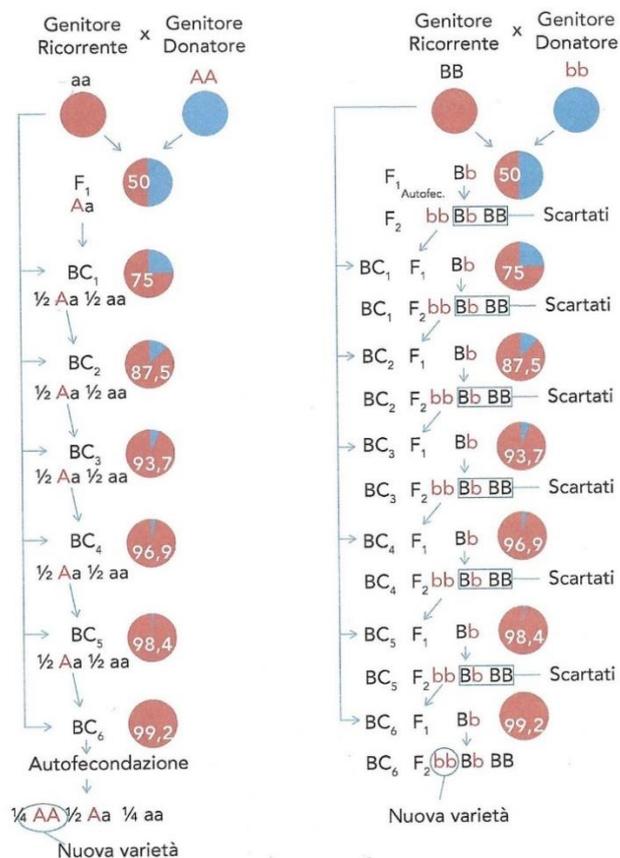


Figura 4-2: Schema di reincrocio (Lorenzetti, et al., 2018) a sinistra per allele dominante a destra per il recessivo.

Nel sistema di selezione tradizionale, questa tecnica prevedeva lunghi cicli di reincrocio e spesso lunghi tempi di attesa per a valutazione visiva per ogni generazione, con aumento dei costi.

L'uso dei marcatori consente un notevole risparmio di tempi e costi.

Si pensi alla necessità di introdurre una resistenza ad un patogeno; se in una coltivazione ad esempio di lattuga, compare una nuova razza di un fungo antagonista (es. *Bremia*), la velocità è un fattore determinante nello sviluppo di varietà resistenti, pena l'arrivo tardivo sul mercato varietale, con il rischio di trovarsi di fronte in poco tempo una nuova razza che vanifica in parte gli sforzi profusi.

I marcatori maggiormente utilizzati in questo processo sono i SSR e gli SNP (J.M. & D., 1998)

Rispetto ad un reincrocio classico l'uso della MABC porta notevoli vantaggi tra cui una selezione maggiormente efficace nel caso in cui l'allele da trasferire al genitore ricorrente abbia una ereditabilità bassa, o meno passaggi per recuperare totalmente il genoma del genitore ricorrente (background selection) (Lorenzetti, et al., 2018)).

Ma fra i vantaggi uno spicca in maniera decisa, cioè la riduzione dei tempi nel caso di trasferimento di alleli recessivi; infatti, nel reincrocio classico, quando il carattere del genitore donatore è recessivo, a seguito di ogni passaggio di reincrocio occorre operare una autofecondazione. Con la MABC questo non è più necessario in quanto i marcatori utilizzati sono codominanti e quindi in grado di riconoscere e selezionare gli eterozigoti.

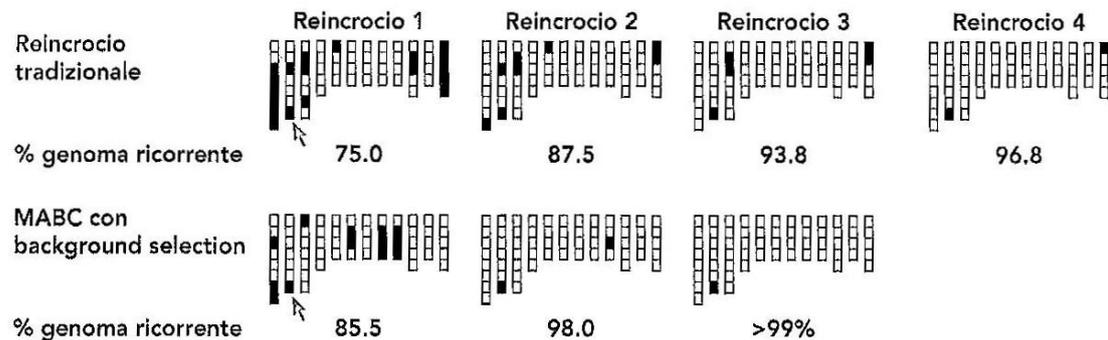


Figura 4-3: Confronto fra reincrocio tradizionale e con MABC (Lorenzetti, et al., 2018)

i quadratini neri indicano il genoma del genitore donatore, i bianchi del ricorrente. La freccia indica il gene da trasferire.

- La selezione ricorrente assistita da marcatori – Marker Assisted Recurrent Selection o MARS:

spesso i QTL sono a bassa ereditabilità e la stima precisa della loro posizione è difficile; inoltre, la loro variabilità in base alle condizioni ambientali non è sempre facilmente rilevabile.

Per cercare di ovviare a questi problemi si ricorre alla MARS soprattutto per quei caratteri complessi.

La MARS si esegue in due fasi distinte in due anni: nella prima, le popolazioni F_2 sono valutate fenotipicamente e genotipicamente dopo essere state allevate in campo, in diverse località. La loro valutazione prevede lo studio dei QTL. Vengono quindi selezionate le famiglie in F_3 sulla base dei dati fenotipici e genotipici ai QTL.

Nella seconda fase, tali famiglie vengono allevate in ambiente protetto (serra), dove l'ereditabilità si riduce a causa della diversità dell'effetto dell'ambiente e le famiglie F_3 vengono interincrociate per poi subire la MAS sulla base dei dati di interconnessione QTL/marcatori derivati dalla fase 1.

Questa tecnica è stata utilizzata spesso nei programmi di miglioramento del mais, per conferire maggiore tolleranza agli stress abiotici, resistenza alle malattie e incremento della resa (J.M. & D., 1998).

4.3.2 *La Selezione Genomica (GS)*

Quando i caratteri di interesse sono portati da diversi loci, ognuno dei quali apporta piccoli effetti presi singolarmente, la MAS può non fornire risultati apprezzabili.

Questo perché l'introggressione di alleli QTL richiede molti cicli di selezione, e spesso i loci a piccolo effetto non sono ben rilevabili dalle analisi dei QTL.

La MAS prevede cicli continui di incroci e reincroci per l'individuazione di QTL ad effetto maggiore (piramidazione), concentrando in un unico genotipo gli alleli favorevoli. (Collard & Mackill, 2007)

Nel caso in cui il carattere sia portato da numerosi loci disposti lungo tutto il genoma, i cui singoli effetti siano piccoli, questo processo incontra delle difficoltà già nella fase di identificazione dei QTL, perché le analisi statistiche volte a predire le differenze fra gli alleli, non sono in grado di stimare sufficientemente le deviazioni piccole fra le medie.

A seguito della disponibilità sempre maggiore delle mappe genetiche di marcatori ad alta densità, si è potuto in parte superare questi limiti grazie alla selezione genomica.

La Selezione Genomica (Genomic Selection – GS) è un programma che prevede l'analisi di un gran numero di marcatori (normalmente SNP) in tutto il genoma. (Bhat, et al., 2016)

Si presuppone che fra tutti i marcatori ci siano anche quelli associati ai QTL ad effetto minore di interesse; non si effettueranno test di significatività per singolo marcatore, ma verranno presi in esame tutti.

Si supponga di avere 2000 individui in F_2 , e che il carattere in oggetto sia controllato da 300 QTL. È ragionevole pensare che nessuno dei 2000 individui abbia fissati in omozigosi tutti e 300 i marcatori. Ad ogni modo, all'interno della popolazione ci saranno sicuramente degli individui che avranno una frequenza maggiore degli altri, di alleli favorevoli.

Quest'ultimi saranno oggetto di identificazione e selezione e vi sarà possibile usare i marcatori molecolari per scovare, con un modello predittivo, le piante migliori di tutta la popolazione.

Il modello predittivo, fondamentale per l'individuazione degli individui migliori, consente la misurazione del più alto breeding value (BV), cioè la somma degli effetti medi degli alleli in un individuo, sulla base dell'informazione genomica (Genomic Estimated Breeding Value – GEBV).

Nella costruzione del modello predittivo, che si basa su equazioni di previsione e modelli statistici, occorre avvalersi di una popolazione di training e di una popolazione di validazione.

La popolazione di training ha lo scopo di addestrare il modello di previsione; per fare ciò deve essere fenotipizzata in un numero elevato di ambienti e genotipizzata con un alto numero di marcatori. (Lorenzetti, et al., 2018)

La popolazione di validazione è anch'essa una popolazione fenotipizzata e genotipizzata accuratamente ed ha lo scopo di verificare gli effetti di associazione del genotipo ai vari marcatori nella popolazione studiata.

La popolazione di validazione è un sottoinsieme della popolazione intera per la quale si conoscono i dati fenotipici e i dati dei marcatori.

Si ponga ad esempio una popolazione di 300 piantine, tutte caratterizzate con gli stessi marcatori per il carattere della resa.

La popolazione training sarà di 250 individui, mentre quella di validazione di 50 individui.

La suddivisione nei due sottogruppi sarà ripetuta parecchie volte, attribuendo ogni volta gli individui in maniera casuale nelle due popolazioni. Questo processo aumenterà l'accuratezza della previsione. Alla fine tutti e 300 gli individui saranno analizzati ottenendo l'equazione di previsione.

Una tipica equazione di previsione è:

$$G = \sum g_i x_i$$

Dove: G= valore genotipico (in questo caso anche GEBV)

g_i = effetto associato al genotipo del marcatore i -esimo

x_i = vettore con i genotipi ai diversi marcatori

Questa equazione verrà utilizzata per stimare il valore genotipico (G) di tutti gli individui, all'interno della popolazione totale, consentendo di individuare la popolazione test, cioè gli individui che saranno soggetti a selezione. (Lorenzetti, et al., 2018)

La popolazione test sarà solamente genotipizzata (con gli stessi marcatori delle pop. training e validazione) e non fenotipizzata, cioè le piante della popolazione test saranno

valutati esclusivamente per il loro valore genotipico per mezzo dei marcatori; solamente gli individui selezionati saranno successivamente valutati per il carattere che interessa.

L'accuratezza della previsione è variabile in funzione dell'ereditabilità e dell'ampiezza della popolazione training. Finora si è supposto che sia la popolazione training sia quella test provengano dalla stessa popolazione iniziale, e che quindi condividano gli stessi alleli. È possibile però, utilizzare modelli costruiti su popolazioni, impiegati, poi, per effettuare selezioni su popolazioni diverse. In questo caso l'elemento discriminante dell'accuratezza della previsione è il grado di parentela delle due popolazioni.

La selezione genomica si avvantaggia di un alto numero di marcatori a basso costo, cosicché possa essere più conveniente rispetto alla selezione fenotipica.

La massima applicabilità della selezione genomica sulla fenotipica, si ha quando i caratteri hanno un'alta ereditabilità nella popolazione training e bassa nella popolazione test; questo sta a significare che nella popolazione test la selezione fenotipica sarà inefficace.

Capitolo 5

PROSPETTIVE FUTURE DEL MIGLIORAMENTO GENETICO

Gli strumenti visti nel precedente capitolo mostrano come, adottando progetti di miglioramento genetico moderni e funzionali, si possano notevolmente ridurre i costi e tempi di sviluppo di nuovi ibridi e varietà.

Le tecniche di selezione tramite marcatori o di selezione genomica sono basate sullo studio di un singolo genoma di riferimento per specie; oggi la ricerca sta facendo notevoli progressi verso uno studio multiplo di genomi in grado di catturare la completa diversità genetica presente in una specie: tale approccio si basa sullo sviluppo e sull'analisi del Pangenoma (Agnieszka A. Golicz, 2016)

5.1 L'era Pan-genomica

Nel 1977 il chimico britannico Frederick Sanger si fece conoscere al mondo come il primo scienziato capace di completare il sequenziamento del DNA di un batteriofago, dando vita al metodo di sequenziamento che porta il suo nome. (Lorenzetti, et al., 2018)

Da allora, le tecnologie e le conoscenze in campo biochimico ed informatico si sono evolute con grande velocità determinando la comparsa di nuovi sistemi di sequenziamento del genoma, tra cui i più recenti chiamati Next Generations Sequencing (NGS) (Slatko, et al., 2018). Tali approcci sono in grado di sequenziare una gran mole di genomi riuscendo ad elaborare enormi quantità di dati con tempi e costi estremamente ridotti ((Ansorge, 2009) (Rhee & Burns, 2006)): i genomi di riferimento sono così diventati disponibili ad un ritmo esponenziale per diverse specie. Inoltre, gli assemblaggi del genoma ad alta qualità e il sequenziamento ad alto rendimento hanno anche incoraggiato le metodologie di ri-sequenziamento (Stratton, 2008), in cui le reads di ciascun campione vengono allineate ad un singolo genoma di riferimento per catturare la diversità genetica presente in una specie, generalmente in termini di Single Nucleotide Polymorphisms (SNPs) (Yuan, et al., 2017).

Gli SNPs sono marcatori biallelici e co-dominanti, caratterizzati da una vasta distribuzione lungo tutto il genoma e da semplici modelli mutazionali (Morin, et al., 2004). Per tutte queste

ragioni, negli ultimi anni sono diventati i marcatori di scelta nell'analisi della diversità genetica di dati umani, animali, microbici e vegetali.

Tuttavia, gli SNPs da soli non sono sufficienti per rappresentare il repertorio genetico completo di una specie (Springer NM, 2009), (Saxena, 2014).

Recenti studi hanno identificato altre fonti di diversità genetica, chiamate variazioni strutturali (SVs), che sono variazioni genomiche in segmenti di DNA > 1 kbp e includono variazioni di presenza/assenza (PAV), variazioni del numero di copie (CNV) e altre varie variazioni nella forma di inversioni, trasversioni e traslocazioni inter/intra-cromosomiche

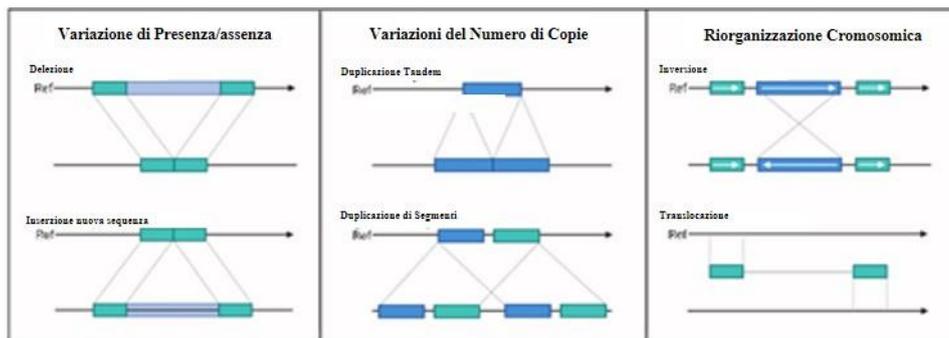


Figura 5-1 Diversi tipi di variazioni strutturali (SV) rilevabili nel pangenoma (Della Colletta Rafael, 2021)

Nella genetica umana, le SVs sono state ampiamente studiate per la loro associazione con malattie croniche (Jan O. Korbel, 2007), (McCarroll, 2007).

Recentemente, le SVs sono state identificate anche in diverse specie vegetali, tra cui *Arabidopsis thaliana*, orzo, mais, riso, sorgo, soia e frumento, evidenziando la loro associazione con importanti variazioni fenotipiche (Muñoz-Amatriaín, 2013).

Sebbene sia stato scoperto che le SVs hanno un ruolo fondamentale nelle piante, la loro caratterizzazione usando un singolo genoma di riferimento è limitata dal fatto che si tende a perdere l'informazione relativa a tutte quelle regioni altamente polimorfiche o che non sono presenti nel genoma utilizzato come reference. Il singolo genoma di riferimento risulta quindi essere insufficiente per rappresentare l'intera diversità genetica di una data specie, mettendo in luce l'urgente bisogno di passare ad un sistema di riferimento pan-genomico (Jordan M. Eizenga, 2020).

Per definizione, il pan-genoma è l'architettura genomica totale di una specie sviluppata dal sequenziamento e dall'analisi di accessioni multiple (Tettelin & Medini, 2020). Utilizzare il pan-genoma anziché il singolo genoma di riferimento permette di migliorare l'identificazione delle varianti e dei geni associati a tratti agronomici chiave, in quanto tutte le relazioni di

sequenza vengono catturate tramite il confronto diretto tra il genoma di riferimento scelto e le restanti accessioni rappresentative di una data specie.

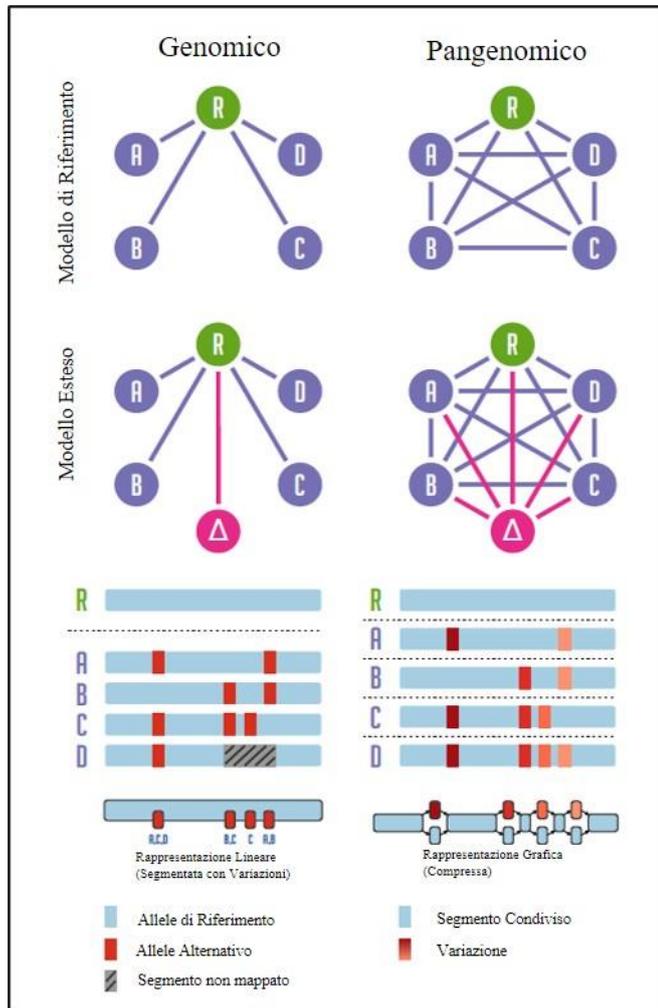


Figura 5-2: Confronto tra l'utilizzo di un singolo genoma di riferimento e il modello pangenomico (Jordan M. Eizenga, 2020)

5.1.1 Struttura del Pan-Genoma

Il pan-genoma è strutturato principalmente in due parti:

- Il Genoma Core: rappresenta il set di sequenze condivise completamente da tutti gli individui della specie o delle famiglie prese in esame. Sono descritte come il genoma minimo per garantire la sopravvivenza cellulare;
- Il Genoma Accessorio: è composto da sequenze di DNA parzialmente condivise (Shell) e/o non condivise (Cloud) tra gli individui. Diversamente dal Genoma Core, il Genoma Accessorio è correlato a geni responsabili

dell'adattamento e della sopravvivenza in diversi agroambienti, come la resistenza alle malattie e le risposte allo stress

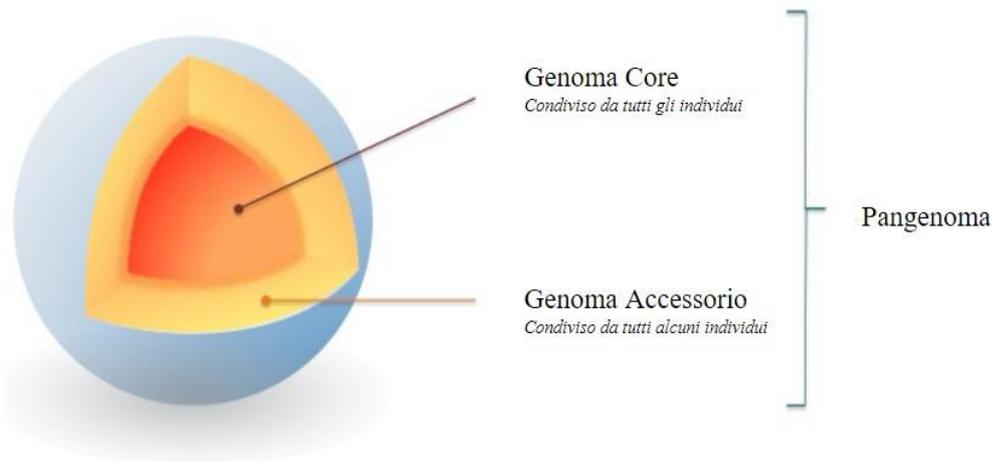


Figura 5-3: Rappresentazione schematica della struttura del pan-genoma (Christine Tranchant-Dubreuil, 2019)

Il primo lavoro sul pan-genoma in mais ha mostrato che solo la metà della struttura genomica è conservata tra due individui e che la diversità privata è responsabile di importanti variazioni fenotipiche. Un altro studio condotto sul genoma della varietà di vite Tannant, coltivata prevalentemente in Uruguay, mostra l'importanza dell'approccio pangenomico.

La caratteristica di tale varietà è quella di avere un alto grado di polifenoli sia nei semi sia nella buccia (fino a 6 volte in più del contenuto del Pinot Nero), tanto da conferire al vino che ne deriva, interessanti proprietà salutistiche. Lo studio ha dimostrato come i geni deputati alla sintesi polifenolica siano contenuti nel Genoma Accessorio. Queste SVs non sono riscontrabili nel genoma di riferimento, ma sono in grado di apportare una gran differenza fenotipica. Solamente il sequenziamento di diversi genomi di diverse cultivar, ha permesso di rilevare tale polimorfismo. (Da Silva, et al., 2013).

5.1.2 Possibili impieghi del Pan-Genoma nel miglioramento genetico delle piante

Come già visto, molti dei QTL di interesse sono a piccolo effetto, e possono essere studiati e selezionati grazie alla selezione Genomica (GS).

I modelli predittivi per lo sviluppo della GS si basano, ad oggi, sullo studio di numerosi SNP.

Gli SNP sono normalmente identificati rispetto ad un singolo genoma di riferimento; ciò comporta una possibile serie di limitazioni e distorsioni. Infatti l'utilizzo di un solo genoma di riferimento non permette l'esplorazione della diversità genetica all'interno della specie.

Il passaggio da uno studio GWAS con genoma di riferimento ad uno studio pangenomico con rilevazione di SV_s può portare innumerevoli vantaggi.

Oggi giorno esistono tecniche che consentono di fare questo con tempi e costi relativamente ridotti. Una di queste tecniche consiste nel mappare le reads risequenziate su un genoma di riferimento, assemblare de novo le reads non mappate, per poi aggiungere i contigs all'assemblato di riferimento (questa tecnica è chiamata approccio map to pan). (Della Colletta Rafael, 2021)

Questa strategia comporta la riduzione degli errori sfruttando appieno le conoscenze già disponibili di un genoma di riferimento di alta qualità, ma le posizioni genomiche dei contigs assemblati de novo rimangono sconosciute senza ulteriori indagini.

Un'altra tecnica consiste nel creare un grafico nel quale qualsiasi variante riscontrata nell'assemblaggio (SV_s o SNP) è aggiunta al genoma di riferimento come un "nodo" nella posizione genomica in cui viene scoperto.

Per i piccoli genomi è possibile studiare le variazioni assemblando tutti i genomi, confrontandoli, poi con una comparazione generale.

Per i genomi più ampi questo non è possibile a causa degli assemblaggi de novo, pertanto la costruzione di un grafico in grado di evidenziare i polimorfismi può facilitare il lavoro (Della Colletta Rafael, 2021) (Garrison, et al., 2018).

Recentemente è stata sviluppata una strategia ibrida tra il confronto lineare e il grafico basato sul genoma di riferimento. La tecnica prevede che le reads siano mappate su di un genoma basato sul grafico e gli aplotipi siano associati ad uno dei genomi di riferimento utilizzato per costruire il grafico. Le reads sono, quindi, riassociate a questo genoma consentendo una mappatura più accurata rispetto all'approccio basato solo sulla costruzione del grafico. (Della Colletta Rafael, 2021) (Grytten, et al., 2020)

Un altro problema che affligge la selezione con marcatori per QTL, è la difficile previsione dell'interazione ambientale nell'espressione dei geni.

Come si è visto esistono dei metodi che consentono di ridurre in parte, le influenze dell'ambiente, aumentando l'accuratezza della previsione.

Introdurre delle regioni SV nei modelli predittivi potrebbe aumentare notevolmente la precisione delle analisi, questo perché le SV hanno dimostrato di svolgere un ruolo determinante nell'adattamento delle specie ai vari ambienti.

Ad esempio la predizione dell'altezza della pianta di mais in condizioni di scarso apporto di azoto è aumentata notevolmente quando sono stati aggiunti al modello predittivo, poche centinaia di CNV_s all'analisi di circa 20000 SNP_s (Lyra, et al., 2019)

Naturalmente non tutte le SV possono essere etichettate con SNP_s, e le differenze fenotipiche di queste variazioni sfuggono ai modelli predittivi. (Della Colletta Rafael, 2021)

Ad oggi esistono poche specie coltivabili di cui siano stati studiati i pan-genomi; nelle orticole per il momento si hanno studi solo nel segmento delle brassiche, inoltre i costi sono al momento piuttosto alti.

Le informazioni delle variazioni strutturali provenienti dal pan-genoma possono essere efficacemente utilizzate dai genetisti se esistono tecnologie di genotipizzazione degli SNP. Viceversa la caratterizzazione degli SV, che non hanno associazione con gli SNP, rimane conveniente solo se il guadagno genetico si è tale da giustificare l'alto costo.

L'utilizzo di studi pan-genomici potrebbe permettere, in futuro, non solo il miglioramento di specie già molto studiate ed inserite nei programmi di breeding, ma anche affrontare lo studio di quelle specie che hanno genomi difficili e poco conosciuti ed anche l'approccio alla domesticazione di specie selvatiche.

CONCLUSIONI

Dagli studi esaminati e trattati nei capitoli precedenti, emerge quanto la selezione tramite marcatori e le sue forme di applicazione possano attivamente contribuire al raggiungimento degli obiettivi che oggi il nostro pianeta ci chiede.

Le sfide che l'orticoltura deve affrontare sono molteplici e non facili da superare; non esiste una unica soluzione capace di risolvere tutte le problematiche, ma bensì un approccio olistico che sia in grado di coniugare le conoscenze delle scienze agrarie, biologiche, fisiche, chimiche, biochimiche, genetiche, informatiche ecc.

In questo approccio gli studi di genetica e di genomica, penso possano dare un fattivo ed importante contributo.

La selezione di nuove varietà subisce, spesso, processi lunghi e farraginosi che portano alla presentazione sul mercato con tempi talmente lunghi da far risultare qualsiasi innovazione come superata. Un esempio è la resistenza a specifici ceppi di malattie che spesso sono surclassati da nuovi ceppi che rendono la varietà dotata di resistenza troppo vecchia per combattere la nuova razza.

La selezione attraverso marcatori, permette, come visto, di accelerare sui tempi di creazione di nuove varietà, consentendo anche un vantaggio in termini di costo.

Riuscire ad essere tempestivi rende possibile, in alcuni casi, la coltivazione in zone che presentano grosse problematiche (condizioni climatiche difficili, scarso approvvigionamento di acqua, presenza endemica di malattie o insetti che rendono difficile la coltivazione, salinità dei terreni ecc.), consentendo di fatto, di continuare a produrre cibo.

Penso che grazie alle nuove tecnologie ed ai nuovi studi, come il pan-genoma, le nuove metodologie di sequenziamento, ma anche la fenomica e le new breeding techniques (che non sono state trattate in questo testo) si possano incrementare le soluzioni a disposizione per progredire verso un'agricoltura sempre più sostenibile e capace di adattarsi alle molteplici condizioni con cui avremo a che fare.

BIBLIOGRAFIA

- Agnieszka A. Golicz, J. B. E., 2016. online library. wiley. *Plant Biotechnology Journal*, volume 14, issue 4, 04, pp. 1099-1105.
- Ansorge, W. J., 2009. Next-generation DNA sequencing techniques. *New Biotechnology*, volume 25, issue 4, pp. 195-203.
- Assosementi, 2021. *Assosementi/associate/orto*. [Online]
Available at: <http://www.sementi.it/associate/orto>
[Consultato il giorno 9 11 2021].
- Bhat, J. A., Sajad, A., K., S. R. & Mir Zahoor A., D. S. J. V. T. A. M. M. J. N. S. P. K. S. G. P. P. K. V., 2016. Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding. *Frontiers in Genetics*, Volume 7, p. 221.
- Brachi, J., Morris, B. & Borevits, G., 2011. Genome-Wide association studies in plants: the missing heritability is in the field. *Genome Biology*, 10 12, pp. 1-8.
- Ceccarelli, S., Lorenzetti, F., Rosellini, D. & Veronesi, F., 1980. *Genetica Agraria*. 2011 a cura di Bologna: Pàtron.
- Christine Tranchant-Dubreuil, M. R. F. S., 2019. Plant Pangenome: Impacts on Phenotypes And Evolution. *Annual Plant Review*, 1 03, pp. 1-8.
- Collard, B. C. Y. & Mackill, D. J., 2007. Marker-assisted selection: an approach for precision plant breeding in twenty-first century. *Philosophical Transaction of the Royal Society*, 22 10, 48(3), pp. 557-572.
- Commissione Europea, 2020. *European Commission - Farm To Fork Strategy*. [Online]
Available at: https://ec.europa.eu/food/horizontal-topics/farm-fork-strategy_it
- Commissione Europea, s.d. *Politica Agricola Comune*. [Online]
Available at: https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy_it

- CREA Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, 2020. *Fitogest*. [Online]
Available at: <https://www.crea.gov.it/-/virus-del-pomodoro-facciamo-il-punto-su-diffusione-e-difesa?inheritRedirect=true&redirect=%2Fricerca%3Fq%3Dtobrfv%26category%3D37390>
[Consultato il giorno 11 11 2021].
- Cristofaro, E. D., 2020. *Focus sulle emissioni da agricoltura e allevamento*, Roma: ISPRA - Istituto Superiore per la Protezione e la Ricerca Ambientale; Sistema Nazionale per la Protezione dell'Ambiente.
- Da Silva, C. et al., 2013. The High Polyphenol Content of Grapevine Cultivar Tannat Berries Is Conferred Primarily by Genes That Are Not Shared with the Reference Genome. *The Plant Cell*, 25(4777-4788), p. 12.
- Della Colletta Rafael, Q. Y. O. S. H. M. B. H. C. N., 2021. How the pan-genome is changing crop genomics and improvement. *Genome Biology*, 22 03, pp. 1-19.
- EPPO , 2015. *EPPO*. [Online]
Available at: https://www.eppo.int/ACTIVITIES/plant_quarantine/alert_list_viruses/tomato_leafcurl_newdelhi
[Consultato il giorno 11 11 2021].
- European and Mediterranean Palnt Protection Organization (EPPO), 2020. *Eppo*. [Online]
Available at: <https://gd.eppo.int/taxon/TOBRFV/documents>
[Consultato il giorno 11 11 2021].
- Gargani, M., s.d. *I Marcatori Microsatelliti per l'Indagine di Popolazioni*, s.l.: s.n.
- Garrison, E. et al., 2018. Variation graph toolkit improves read mapping by representing. *Nature Biotechnology*, Volume 26, pp. 875-879.
- Grytten, I., Ranf, K., Nederbragt, A. & Sandve , G., 2020. Assessing graph-based read mappers against a baseline approach highlights strengths and weaknesses of current methods. *BMC Genomics*, 06 4.
- Huang, X. et al., 2011. Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. *PNAS*, 108(11), pp. 4488-4493.

- International Breミア Evaluation Board Europe (IBEB-EU), 2021. *Plantum*. [Online]
Available at: www.plantum.nl
[Consultato il giorno 11 11 2021].
- Ismea, 2021. *AgriMercati: la congiuntura agroalimentare del terzo trimestre 2021 e le prospettive*, s.l.: ismea.
- ISPRA - Istituto Superiore per la Protezione e la ricerca Ambientale, 2021. *Annuario Dati Ambientali*. [Online]
Available at: https://annuario.isprambiente.it/sys_ind/macro/1
[Consultato il giorno 9 11 2021].
- Istat, 2021. *Istat.it*. [Online]
Available at: <http://dati.istat.it/Index.aspx?QueryId=33703>
[Consultato il giorno 09 11 2021].
- J.M., R. & D., H., 1998. *Marker-assisted selection: new tools and strategies*, s.l.: Elsevier Science.
- Jan O. Korb, A. E. U. J. P. A. B. G. F. G. J. F. S. P. M. K. D. P. N. J. C. L. D. B. E. T. Z. C. A. T. A., 2007. Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. *Science*, 19 10, pp. 420-426.
- Jordan M. Eizenga, A. M. N. J. A. S. S. H. A. G. G. H. X. C. J. D. S. R. R. J. E. M. R. S. G. B. P. T. M., 2020. Pangenome Graphs. *Annual Review of Genomics and Human Genetics*, 26 05, pp. 139-162.
- Khan, S., Saeed, S. & Khan, 2021. Advances and Challenges for QTL Analysis and GWAS in the Plant Breeding of High-Yielding: A Focus on Rapeseed. *Biomolecules*, 11(1516).
- Landi, V., 2008. *"DIFFERENZIAMENTO GENETICO E DINAMICHE DI POPOLAZIONE NEL MERINO EUROPEO*, s.l.: s.n.
- Lorenzetti, F. et al., 2018. *Miglioramento genetico delle piante*. 2021 a cura di Milano: Edagricole.
- Lyra, D. et al., 2019. Modeling copy number variation in the genomic prediction of maize hybrids. *Theoretical and Applied Genetics*, 1 1, pp. 273-288.
- McCarroll, S. A. D., 2007. Copy-number variation and association studies of human disease. *Nature Genetics* 39, pp. 37-42.

- McCarthy, C., 2018. *Microbiology Society*. [Online] Available at: <https://microbiologysociety.org/blog/uncovering-the-fungal-pangenome.html> [Consultato il giorno 12 2021].
- Morin, P. A., Luikart, G. & Wayne, R. K., 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, pp. 208-216.
- Muñoz-Amatriaín, M. E. S. W. T. e. a., 2013. Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome biology* 14, 14 6.
- Network Sol Genomics, s.d. *Sol Genomics Network*. [Online] Available at: https://solgenomics.net/cview/view_chromosome.pl?map_version_id=137&chr_nr=1&show_physical=0&show_IL=0&show_ruler=0&show_offsets=0&color_model=&comp_map_version_id=&comp_chr=0&zoom=1&size=0&hilite=&confidence=-2&show_zoomed=1&marker_type=&cM_start=0.00&cM
- Rhee, M. & Burns, M. A., 2006. Nanopore sequencing technology: research trends and applications. *Trends in Biotechnology*, volume 24, no. 12, 19 10, pp. 580-586.
- Saxena, R. & E. D. & V. R., 2014. Structural variations in plant genomes. *Briefings in functional genomics*, pp. 296-307.
- Schluter, D. & C., W. M., 2010. *Analisi Statistica dei Dati Biologici*. 2015 a cura di Bologna: Zanichelli.
- Slatkin, M., 2008. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future.. *Nature Reviews Genetics*, 9, pp. 477-485.
- Slatko, B. E., Gardener, A. F. & Ausubel, F. M., 2018. Overview of next-generation sequencing technologies. *Current Protocols in Molecular Biology*, 16 4, p. 122.
- Springer NM, Y. K. F. Y. J. T. Y. C. J. Y. W. W. R. T. K. J. R. H. I. A. B. W. J. J. N. D. S. P., 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet*.
- Stratton, M., 2008. Genome resequencing and genetic variation. *Nature Biotechnology*, 1, pp. 65-66.
- Tettelin, H. & Medini, D., 2020. *The Pangenome - Diversity and evolution of Genomes*. Baltimora USA - Siena IT: Springer OPEN.

- Uffelma, E. et al., 2021. Genome-wide association studies. *Nat Rev Methods Primers*, Volume 1.
- Young, N., 1992. Restriction Fragment Length Polymorphisms (RFLPS) and Crop Improvement. *Experimental Agriculture*, pp. 385-398.
- Yuan, Y., Bayer, P. E., Batley, J. & Edwards, D., 2017. Improvements in Genomic Technologies: Application to Crop Genomics. *Trends in Biotechnology*, volume 35, issue 6, 06, pp. 547-558.

RINGRAZIAMENTI

Giungo finalmente alla fine di questo emozionante percorso.

Percorso che ho voluto fortemente iniziare, ma che non avrebbe avuto il suo epilogo senza l'aiuto di tante persone che mi hanno saputo aiutare, spronare nei momenti di tentennamento o stanchezza, supportare e sopportare nelle lunghe sessioni di studio e che hanno gioito come me per ogni piccolo successo.

Perciò via con i ringraziamenti:

per prima cosa vorrei ringraziare il mio relatore, Prof. Roberto Papa e la mia correlatrice Dott.ssa Gaia Cortinovis, per avermi dato la possibilità di scrivere una tesi con loro e per avermi indirizzato e seguito nel lungo cammino della stesura del documento; i loro consigli sono stati realmente preziosi, grazie.

Grazie anche alle aziende per cui ho lavorato e con cui lavoro oggi, perché senza la possibilità che mi hanno dato di assentarmi dal lavoro per poter dare gli esami, non starei scrivendo queste righe.

Grazie ai colleghi, che spesso mi hanno supportato o per il solo fatto di avermi spesso chiesto “come va con l'università?”.

Grazie ai parenti tutti, acquisiti e non, ai cugini, zii e nonni, per il loro immenso supporto morale.

Grazie agli amici, in particolare a Andrea, Emanuele, Chiara, Marco, Agnese, Carlo, Francesco e Federica, per avermi aiutato a “distrarmi”.

Un ringraziamento speciale va a Stefano e Paola, genitori fantastici che mi hanno trasmesso l'amore per la cultura e la conoscenza e per la loro capacità di esserci sempre quando ne avevo bisogno: ancora grazie.

Un enorme grazie va a mia sorella Benedetta, perché con il suo esempio mi ha saputo indicare la strada da percorrere; anche se lontana l'ho sempre sentita vicina.

Non ci sono parole adatte per ringraziare mia moglie e compagna di vita Valentina: lei mi è stata vicina sempre, nei momenti di gioia quando gli esami andavano bene, ma anche nei momenti più faticosi, quando sarebbe stato più facile lasciare tutto; ha condiviso con me tutto il percorso aiutandomi di volta in volta a non perdere la via.

E grazie a Damiano, lui ancora non lo sa, ma il suo arrivo mi ha dato un motivo in più per credere di poter arrivare in fondo.