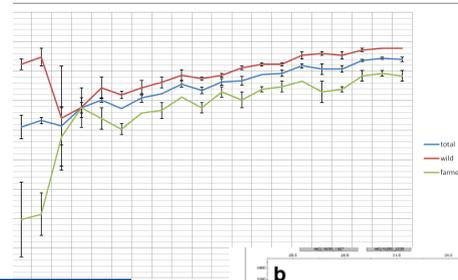
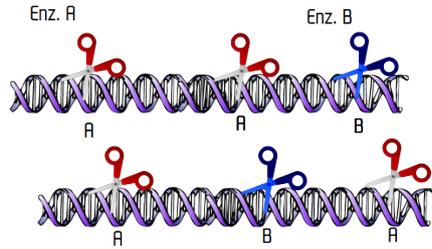
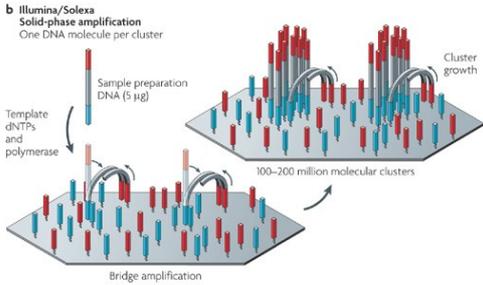
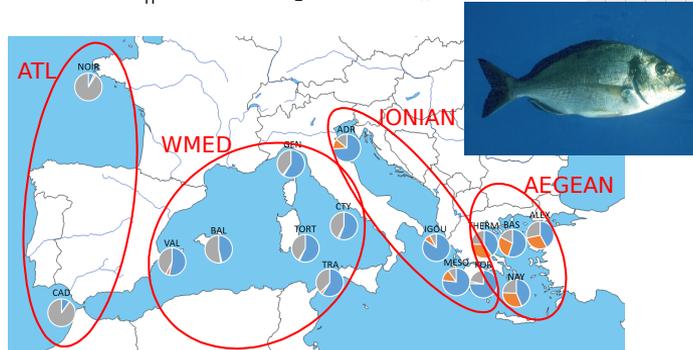
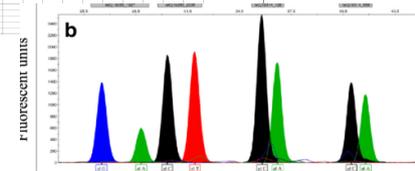


Tecniche genomiche per la tracciabilità geografica e di specie



Indiv. 1	A/A	Indiv. 1	T/T
Indiv. 2	A/A	Indiv. 2	T/T
	A/A	Indiv. 3	T/T



Francesco Maroso
mrsfnc1@unife.it

Sommario

- Dalla genetica alla genomica
 - Next Generation Sequencing
 - Riduzione della complessità genomica e RAD
 - Analisi preliminari *in silico*
- Tracciabilità geografica
- Tracciabilità di specie e ibridi

Dalla genetica alla genomica

- Genetica

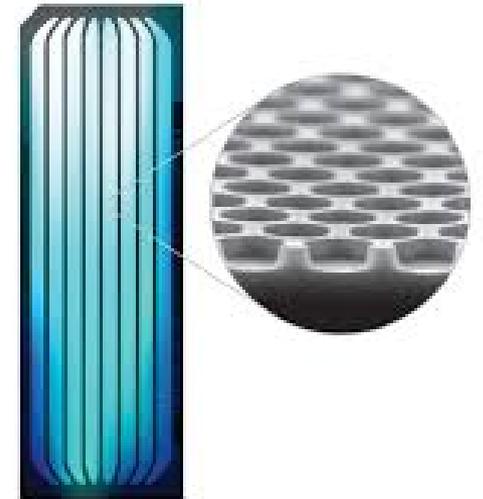
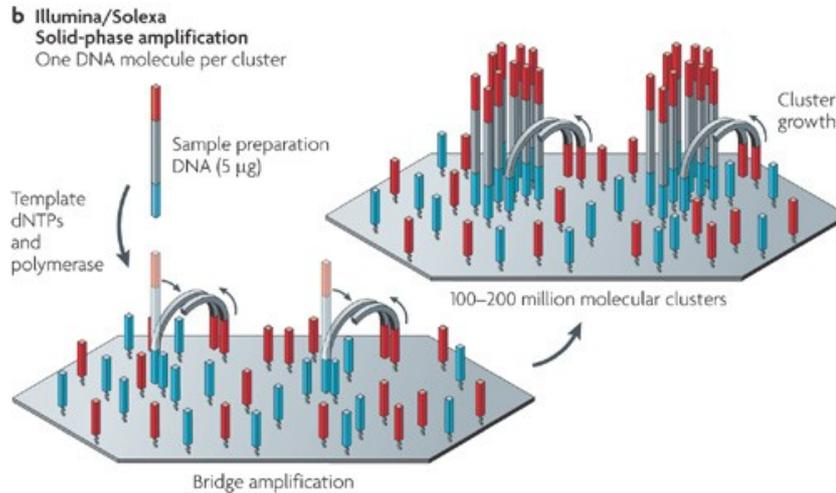
- Pochi geni/sequenze
- Sequenze target
- “Poca” informazione, ma utile!
- Poca flessibilità
- Costi meno elevati
- Più semplici da analizzare

- Genomica

- Moltissime sequenze
- Genome-wide
- Non target
- Molta informazione “inutile”
- Flessibile/modulabile
- Costi più elevati
- Bioinformatica!

Leggere il DNA

- Next Generation Sequencing (NGS)
 - Sequenziamento parallelo di miliardi di sequenze corte* di DNA (...bioinformatica!)



Leggere il DNA

- Next Generation Sequencing (NGS)

- Sequenziamento parallelo di miliardi di sequenze corte* di DNA
- 4 miliardi di pagine di un libro
- file di testo da 6 Terabyte
- 3000 genomi umani

* tecnologia Nanopore: frammenti contigui di centinaia di migliaia di basi



	NextSeq 2000	NovaSeq 6000
Run Time	24-48 hours	~13 - 38 hours (dual SP flow cells) ~13-25 hours (dual S1 flow cells) ~16-36 hours (dual S2 flow cells) ~44 hours (dual S4 flow cells)
Maximum Output	300 Gb*	6000 Gb
Maximum Reads Per Run	1 billion*	20 billion
Maximum Read Length	2 x 150 bp	2 x 250**



Riduzione della “complessità genomica”

Per rispondere a molte delle “domande della ricerca”, è sufficiente molta meno informazione che tutta quella contenuta nel genoma



- Reduced Representation Library (RRL)

Solo una minima parte del genoma viene “letto” e analizzato, di solito molto meno dell’ 1%

Bisogna prima *frammentare* il genoma per poi recuperarne solo alcuni pezzi

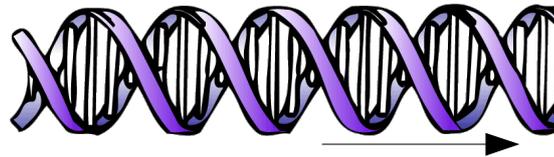
Reduced Representation Library (RRL)

- Restriction site-Associated DNA (RAD) analysis



Vengono usati degli enzimi di restrizione per tagliare il DNA in frammenti di lunghezza variabile

DNA



Enzima di restrizione



Sito di riconoscimento



...GGTCAG**AATTC**GCTGA...
...CCAGT**CTTAA**GCGACT...



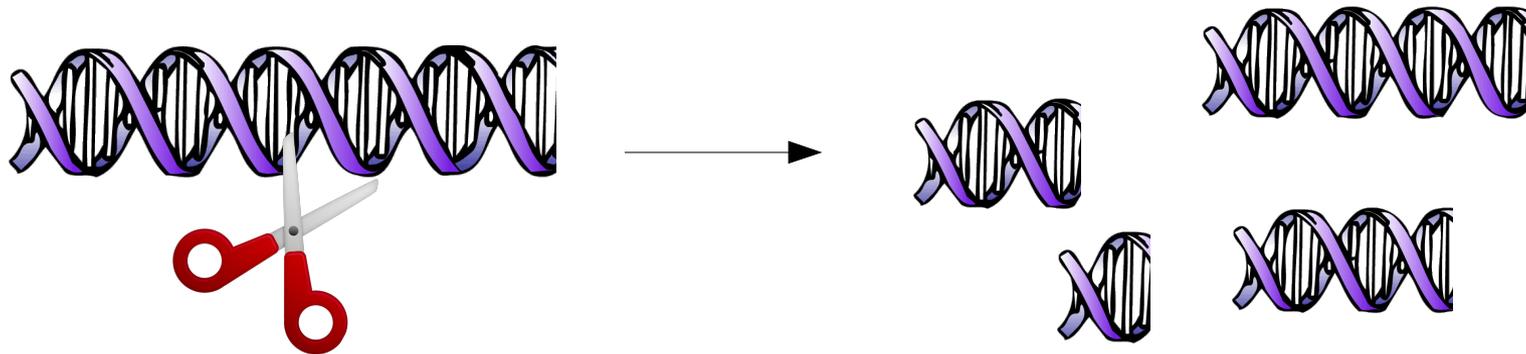
...GGTCAG **AATTC**GCTGA...
...CCAGT**CTTAA** **GCGACT**...

Reduced Representation Library (RRL)

- Restriction site-Associated DNA (RAD) analysis



Vengono usati degli enzimi di restrizione per tagliare il DNA in frammenti di lunghezza variabile

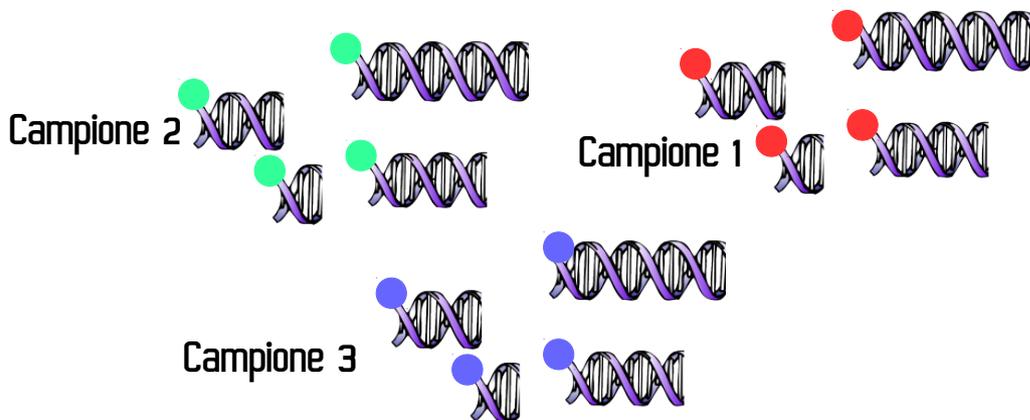


Riduzione della "complessità genomica"

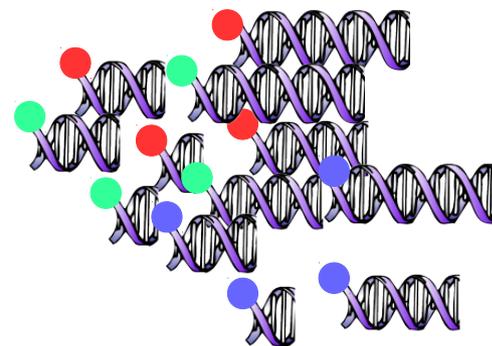
Sfruttare la enorme capacità di lettura per analizzare più campioni contemporaneamente

- Multiplexing

Barcoding



Sequenziamento unico



Flessibilità, dati/risorse

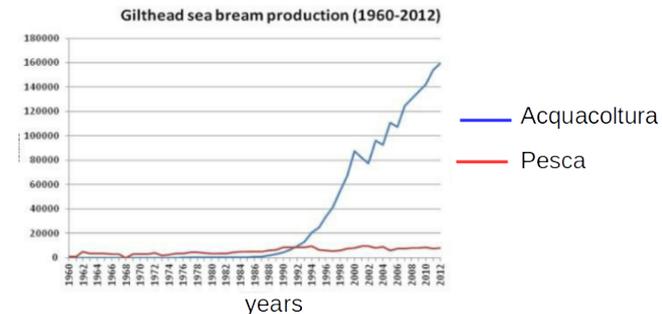
- Tipo di enzima /combinazione di enzimi (frequenza di taglio)
- Selezione dei frammenti
 - Per taglia
 - Per sequenza
- Strumenti semplificati di analisi (chip...)
 - Una volta identificate i frammenti più informativi, si possono sviluppare test più economici e rapidi, mirati solo a analizzare queste sequenze

Analisi preliminari

- *Test in silico*
 - Identificare la quantità di informazione sufficiente (numero di frammenti o numero atteso di SNPs)
 - Simulare diverse combinazioni RAD *in silico* e identificare la migliore
 - Genoma della specie in esame o simile per lunghezza/composizione
- Multiplexing, calcolo dei campioni per *run* di sequenziamento
 - e.g. Illumina NextSeq 2000, 1×10^9 frammenti: 100 campioni a 10^7 frammenti
- Occhio alla precisione e affidabilità dei test!

RAD e tracciabilità geografica

- Tracciabilità geografica
 - Risalire all'origine (geografica o allevato/selvatico)
 - Fenotipica
 - Chimica
 - Genetica



RAD e tracciabilità geografica

- Tracciabilità su base genetica

- Poco intrusiva
- Prodotto lavorato

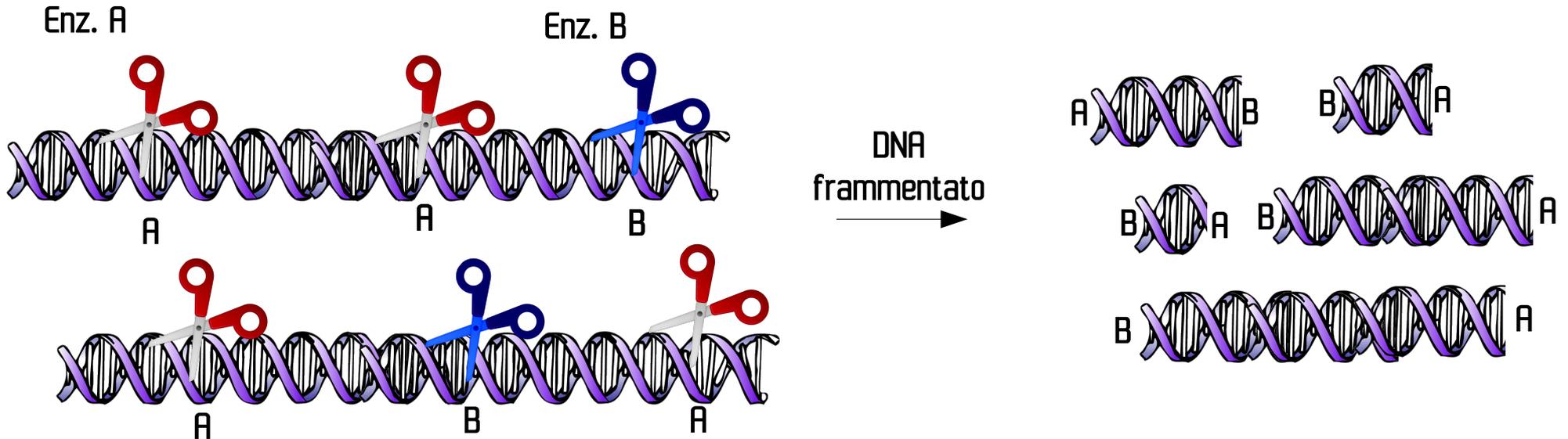


- Informazione genetica ricavata con metodologia ddRAD

- Coppia di enzimi di restrizione
- Size selection

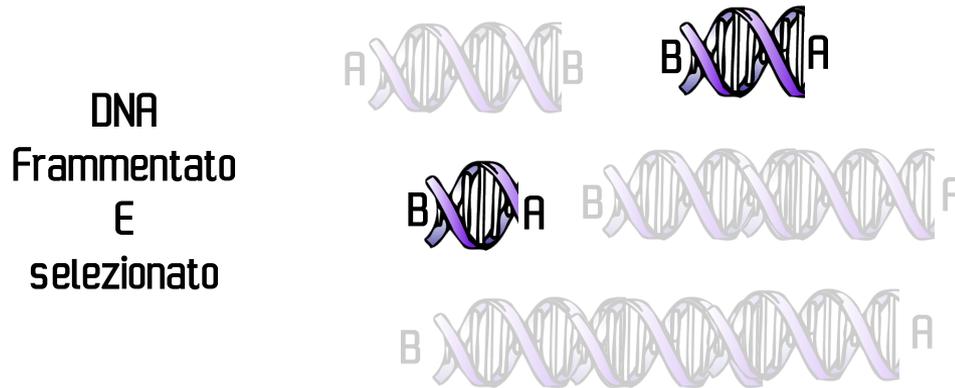
RAD e tracciabilità geografica

- Informazione genetica ricavata con metodologia ddRAD
 - Coppia di enzimi di restrizione



RAD e tracciabilità geografica

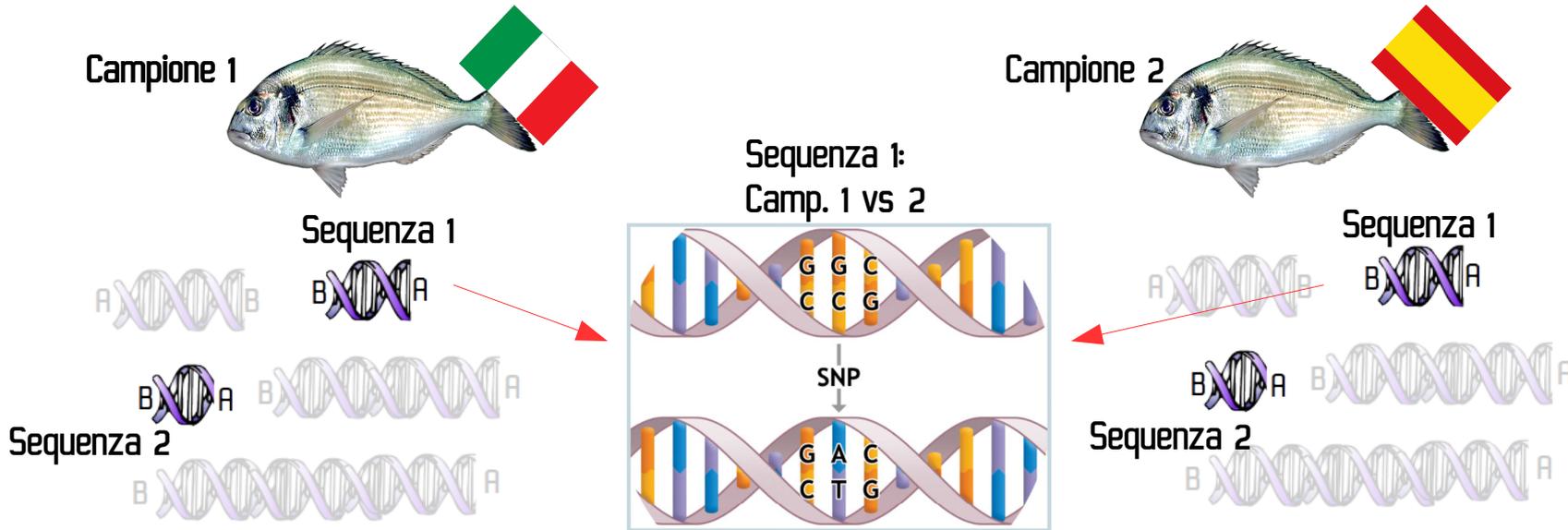
- Informazione genetica ricavata con metodologia ddRAD
 - Coppia di enzimi di restrizione + selezione per dimensione



- La frazione effettivamente analizzata è circa 0,04% dell'intero genoma
- Stessi frammenti in tutti i campioni

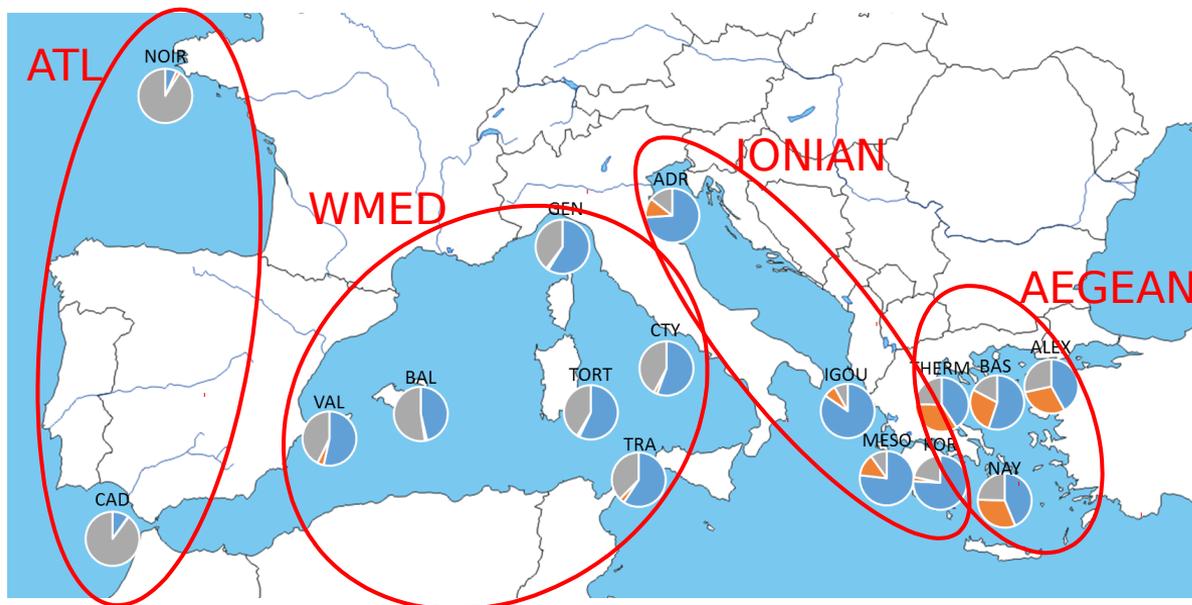
RAD e tracciabilità geografica

- Identificazione delle differenze tra popolazioni (SNPs)
 - Circa 1000 SNPs di alta qualità individuati



RAD e tracciabilità geografica

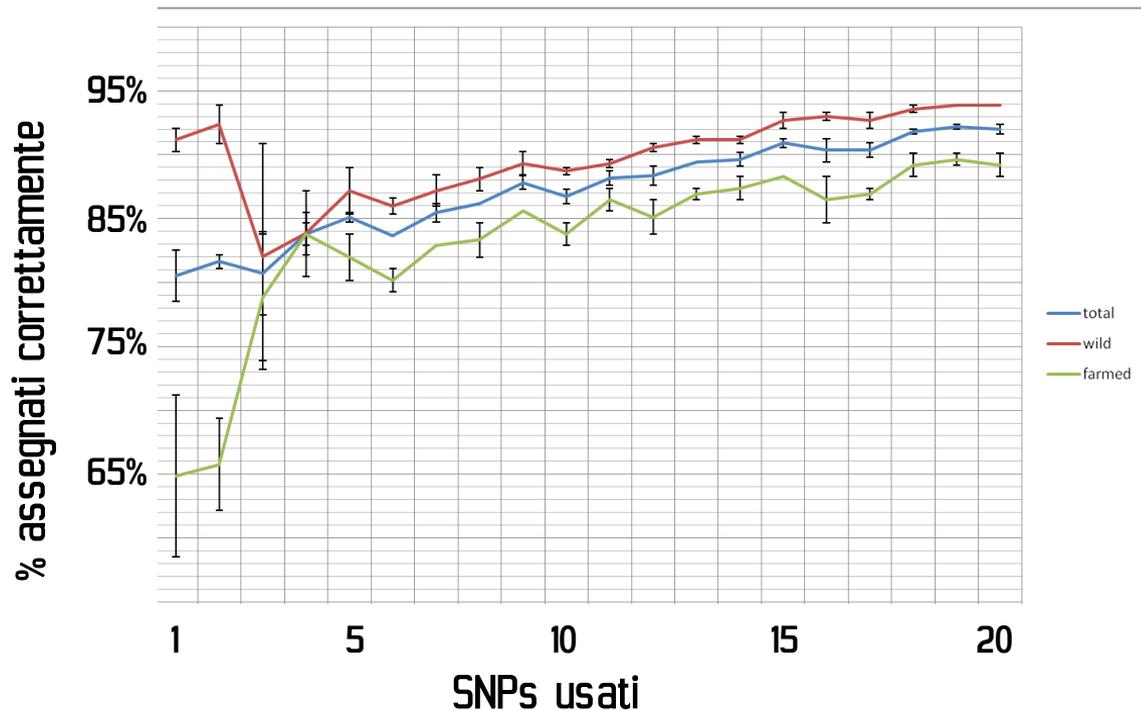
- Identificazione della *struttura genetica*
 - 4 gruppi di riferimento individuati



RAD e tracciabilità geografica

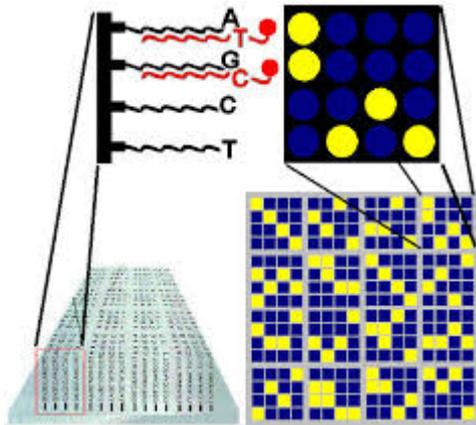
- Selezione dei marcatori più informativi

Rank	SNP	Occurrences in top 10 F_{ST} ranked SNPs of 70 pairs of samples
1	2879_85	65
2	901_49	48
3	7610_82	36
4	8727_39	31
5	6793_54	25
6	11607_56	23
7	239_15	18
8	10951_7	16
9	13514_87	16
10	11632_5	14
11	7954_59	14
12	12501_69	13
13	12192_68	13
14	8327_14	12
15	11192_6	12
16	5068_16	12
17	13129_86	11
18	13674_61	10
19	686_88	10
20	60_69	9



RAD e tracciabilità geografica

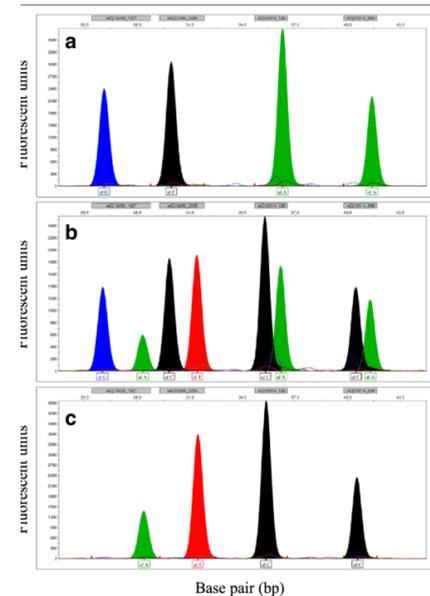
- Selezione dei marcatori più informativi
- Sviluppo di strumenti rapidi/economici
 - SNP chip/array → si analizzano pochi marcatori target (100s-1000s)
 - SnaPshot (10-20 SNPs)



Campione 1

Campione 2

Campione 3



RAD e identificazione di ibridi

- Esempio: identificazione di specie e ibridi:

Cerastoderma edulis e *Cerastoderma glaucum*



RAD e identificazione di ibridi

- Una tecnica alternativa al ddRAD: il 2bRAD
 - gli enzimi 2b (e.g. BsaXI)

tactgttcgacacacACccttaaCTTCggctagctgacgtagctag
atgacaagctgtgtgTGgaattGAAGccgatcgactgcatcgatc

RAD e identificazione di ibridi

- Una tecnica alternativa al ddRAD: il 2bRAD
 - gli enzimi 2b (e.g. BsaXI)
 - Riconoscono un sito

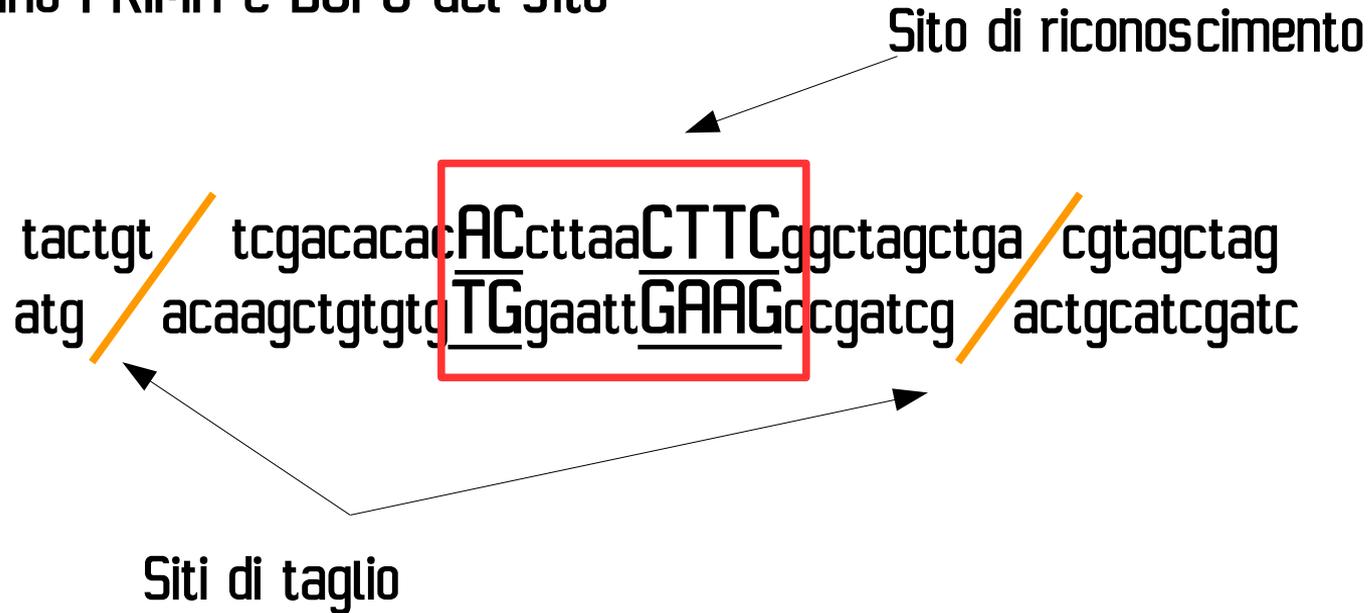
Sito di riconoscimento



tactgttcgacacacACcttaaCTTCggctagctgacgtagctag
atgacaagctgtgtgTGgaattGAAGccgatcgactgcatcgatc

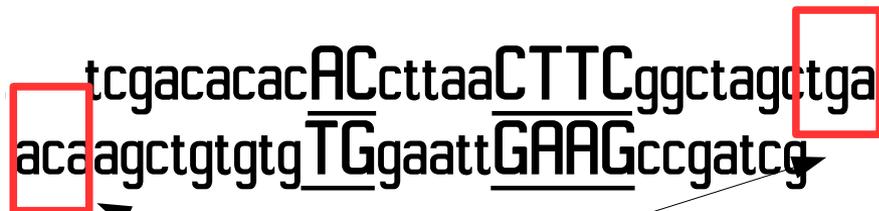
RAD e identificazione di ibridi

- Una tecnica alternativa al ddRAD: il 2bRAD
 - gli enzimi 2b (e.g. BsaXI)
 - Tagliano PRIMA e DOPO del sito



RAD e identificazione di ibridi

- Una tecnica alternativa al ddRAD: il 2bRAD
 - gli enzimi 2b (e.g. BsaXI)
 - Lasciano delle basi libere (VARIABILI DA FRAMMENTO A FRAMMENTO)



tcgacacacACccttaaCTTCggctagctga
acaagctgtgtgTGgaattGAAGccgatcg

The diagram shows a DNA sequence with two recognition sites for the BsaXI enzyme. The first site is 'ACccttaa' and the second is 'CTTC'. The sequence is shown as two strands: the top strand is 'tcgacacacACccttaaCTTCggctagctga' and the bottom strand is 'acaagctgtgtgTGgaattGAAGccgatcg'. The 'ACccttaa' and 'CTTC' sites are underlined. Red boxes highlight the 'A' in 'ACccttaa' and the 'G' in 'CTTC' on the top strand, and the 'A' in 'acaagctgtgtg' and the 'G' in 'GAAGccgatcg' on the bottom strand. Arrows point from the text below to these highlighted bases.

Basi libere a cui si attacca
l'adattatore per l'amplificazione PCR

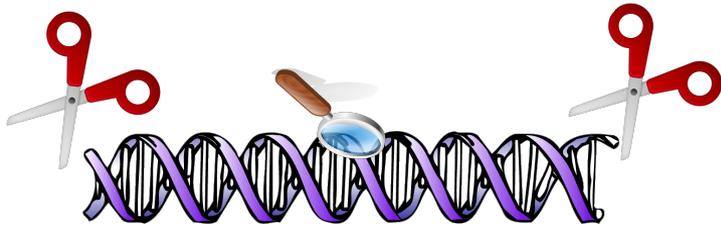
RAD e identificazione di ibridi

- 2bRAD: Selezione per adattatore
- Come per il ddRAD, si può fare un'ulteriore selezione dei frammenti tagliati dall'enzima, selezionando solo quei frammenti che hanno delle basi specifiche agli estremi
- Anche questa tecnica offre molta flessibilità nel numero di sequenze analizzate

2bRAD vs ddRAD

- 2bRAD

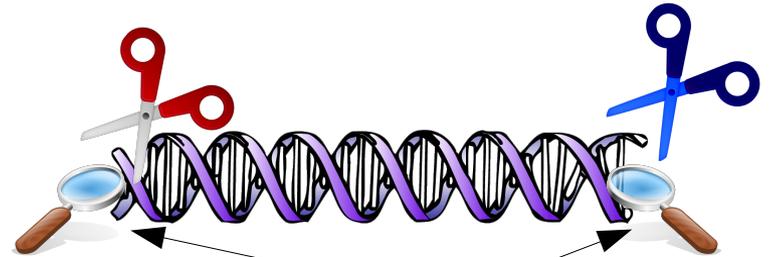
- Sequenze corte (30-40 bp)
- Lunghezza omogenea
- Selezione con adattatori
- Pochi enzimi



Sito di riconoscimento E taglio
non coincidono

- ddRAD

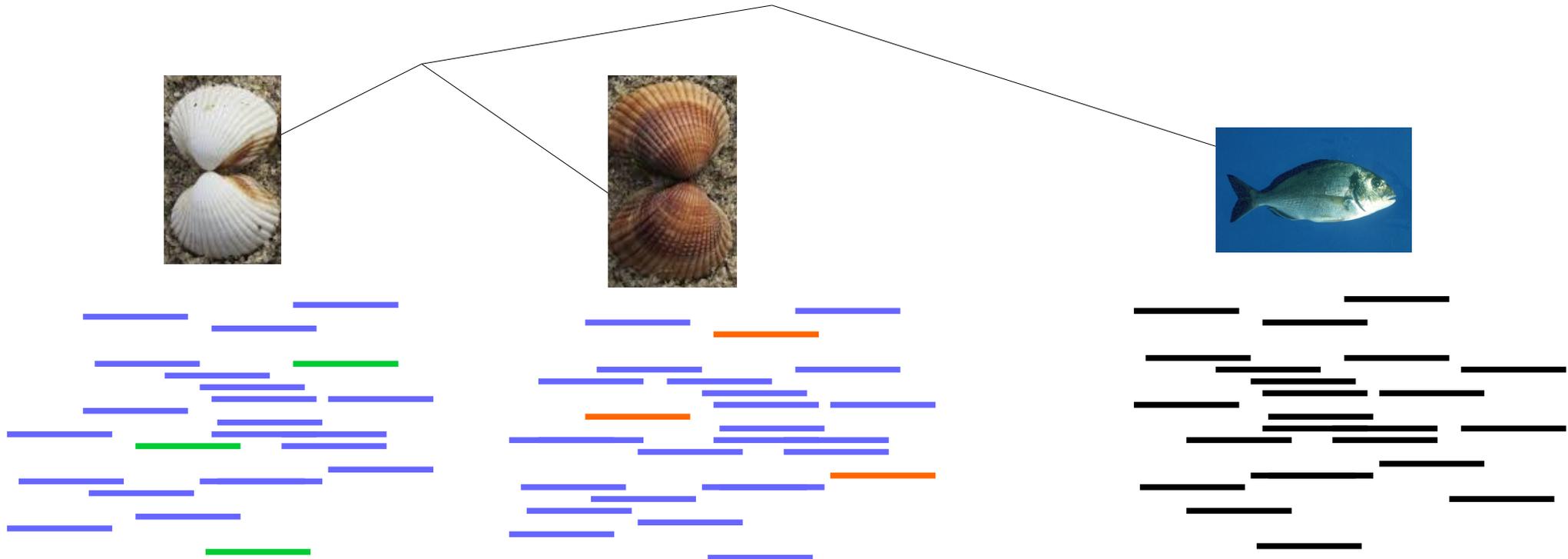
- Sequenze lunghe (200-500 bp)
- Lunghezza disomogenea
- Selezione per taglia
- Molte combinazioni di enzimi



Sito di riconoscimento E taglio
coincidono

RAD e identificazione di ibridi

- Specie simili producono frammenti genomici molto simili tra di loro



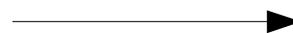
RAD e identificazione di ibridi

- Ricerca di caratteri specie-specifici per identificarle
 - Sequenze specie specifiche

Specie 1



Specie 2



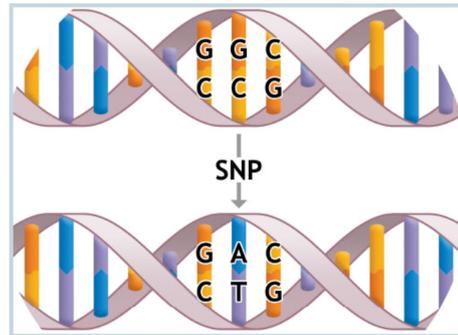
Sviluppo di protocolli
che amplificano solo
queste sequenze e
se, per esempio in
un campione ignoto
amplifica 
ma non 
Sarà della specie 2

RAD e identificazione di ibridi

- Ricerca di caratteri specie-specifici per identificarle
 - Varianti alleliche di tipo SNP nei frammenti da regioni genomiche omologhe



Nella stessa regione genomica, le due specie possono avere degli SNPs



RAD e identificazione di ibridi

- Gli SNP più informativi sono quelli in cui una delle due varianti è fissata in una specie, e l'altra variante è fissata nell'altra specie
 - e.g. al locus 1234 tutti gli individui della specie 1 sono omozigoti A/A e tutti gli individui della specie due sono omozigoti T/T



Indiv. 1 A/A
Indiv. 2 A/A
Indiv. 3 A/A

Indiv. 1 T/T
Indiv. 2 T/T
Indiv. 3 T/T

RAD e identificazione di ibridi

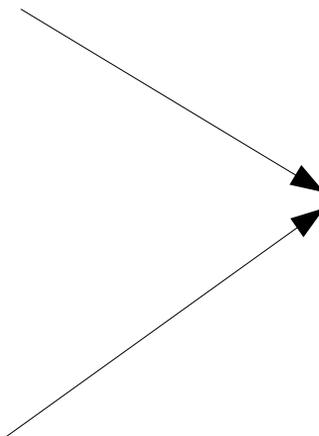
- E gli ibridi? Saranno eterozigoti per i due alleli!!



A/A



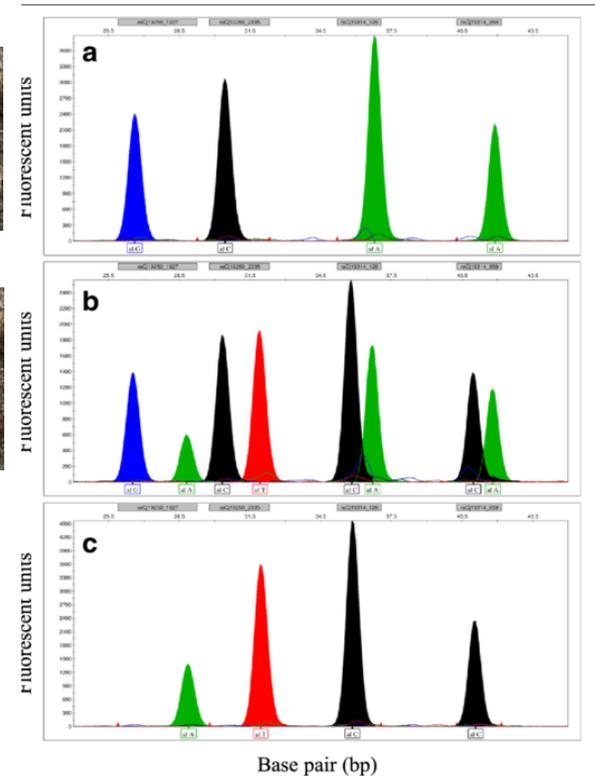
T/T



A/T

RAD e identificazione di ibridi

- Protocolli specifici per i frammenti target
 - Gli omozigoti saranno specie “pure”
 - Gli eterozigoti saranno ibridi



Temi affrontati

- Le tecnologie di analisi *genomica* ci permettono di analizzare più in profondità le differenze genetiche tra organismi
- Questi protocolli sono vari e modulabili per ottenere il miglior rapporto tra precisione e costi di analisi
- Esistono metodi di analisi preliminare (bioinformatica) che simulano i protocolli di laboratorio e permettono di risparmiare tempo e risorse
- Una volta identificati i marcatori di interesse, si possono sviluppare strumenti ancora più mirati e economici/rapidi