FONDAZIONE
EDMUND
MACH

VIII CONVEGNO NAZIONALE
DEL CASTAGNO
14, 15 e 16 settembre 2022 - PORTICI - NAPOLI

Assemblaggio del genoma di castagno (*Castanea sativa*) e sviluppo di un array di genotipizzazione ad alta densità

Diego Micheletti

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Sara Alessandri

Luca Dondini

Consiglio Nazionale delle Ricerche

Claudia Mattioni

Federico Sebastiani

Paola Pollegioni

FONDAZIONE
EDMUND MACH

Luca Bianco

Stefano Piazza

Cristiano Vernesi

Mickael Malnoy

Michela Troggio

Mirko Moser

Luisa Palmieri

UNIVERSITA DEGLI STUDI DI TORINO

Daniela Torello Marinoni

Roberto Botta

Vera Pavese

UNIVERSITÀ POLITECNICA DELLE MARCHE

Sergio Murolo

Outline

- Assemblaggio del genoma di *Castanea sativa*
- Risequenziamento e studio della diversità genetica
- Sviluppo di un array di genotipizzazione ad alta densità

Outline

- Assemblaggio del genoma di *Castanea sativa*

- Risequenziamento e studio della diversità genetica

- Sviluppo di un array di genotipizzazione ad alta densità

Input

- *C. sativa* varietà Marrone di chiusa pesio
  - > 100x Illumina PE150
  - > 100x Oxford Nanopore (N50 >15Kb)
- Mappa genetica da dati GBS (Torello Marinoni, Nishio et al 2020)
  - 3515 SNPs (200+200 bps flanking regions)

Assemblaggio

Method:

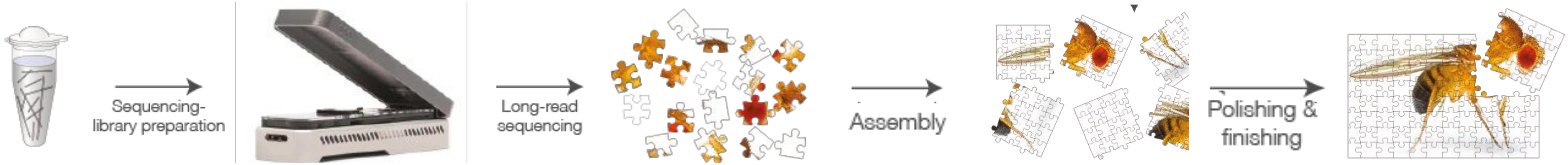
NextDenovo + NextPolish

Results

238 scaffolds [51kb - 91.9 Mb]
N50: 21.8 Mb
N80: 7.1 Mb
Total assembly size: 750 Mb

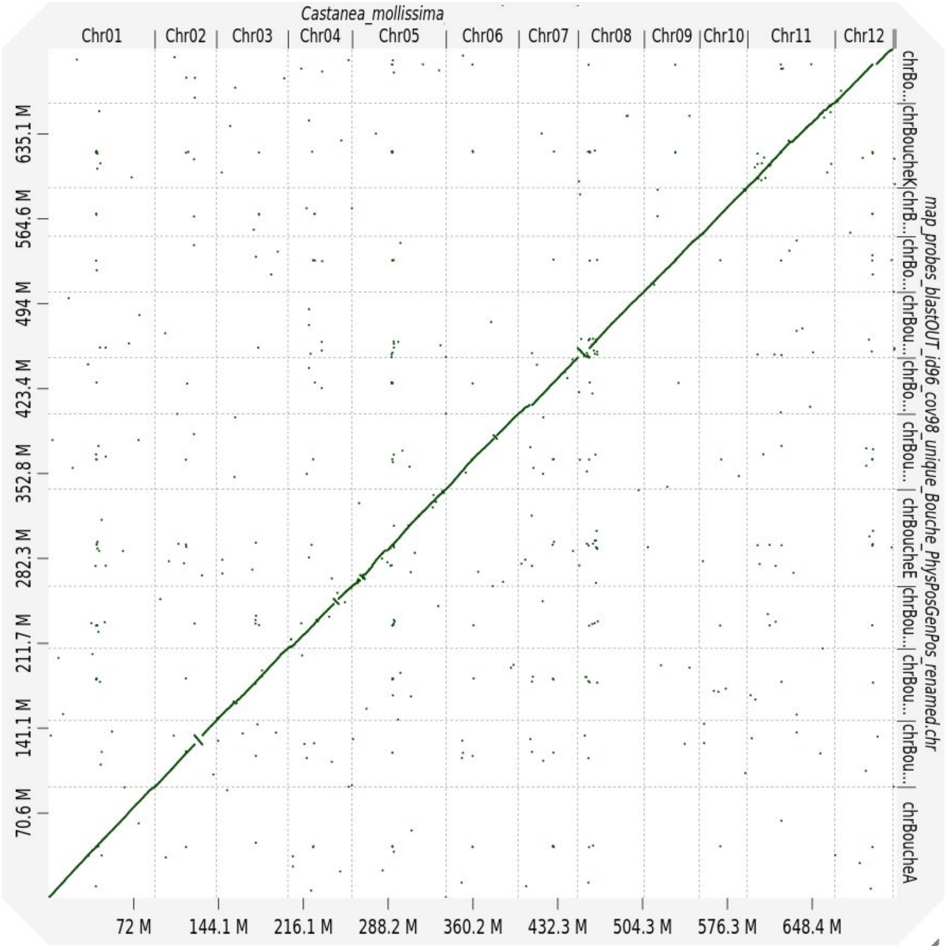
Busco genes:

98.6%	Complete BUSCOs
95.4%	Complete and single-copy BUSCOs
3.2%	Complete and duplicated BUSCOs
0.9%	Fragmented BUSCOs
0.5%	Missing BUSCOs
1614	Total BUSCO groups searched



Ancoraggio sulla mappa genetica di Bouche de Betizac

- 705 Mb anchored in 12 chromosomes
- chromosomes ranging from 40 Mb to 91 Mb
- 45 Mb unanchored (177 sequences)



Outline

- Assemblaggio del genoma di *Castanea sativa*

- Risequenziamento e studio della diversità genetica

- Sviluppo di un array di genotipizzazione ad alta densità

Input

50 accessioni di castagna e marrone principalmente italiane

- 8 campioni da alberi monumentali
- Campioni sequenziati autonomamente dai partner del progetto finanziato in Auto-Crowdfunding
- 20-30x coverage di Illumina o DNBSEQ PE150

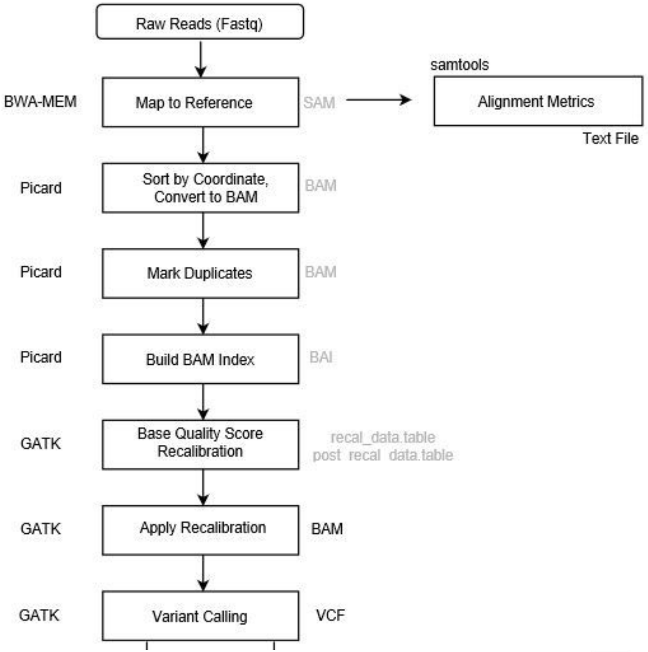
Identificazione delle varianti

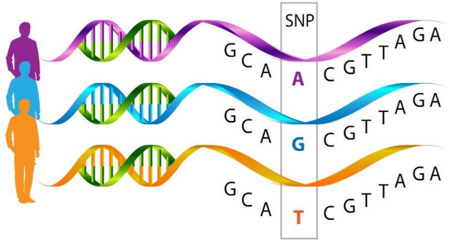
Metodo

- BWA-MEM + GATK
- C. mollissima come reference
- Filter out:
  - Non biallelic SNPs
  - Quality <50
  - MAF < 0.05
  - DP < 8

Risultati

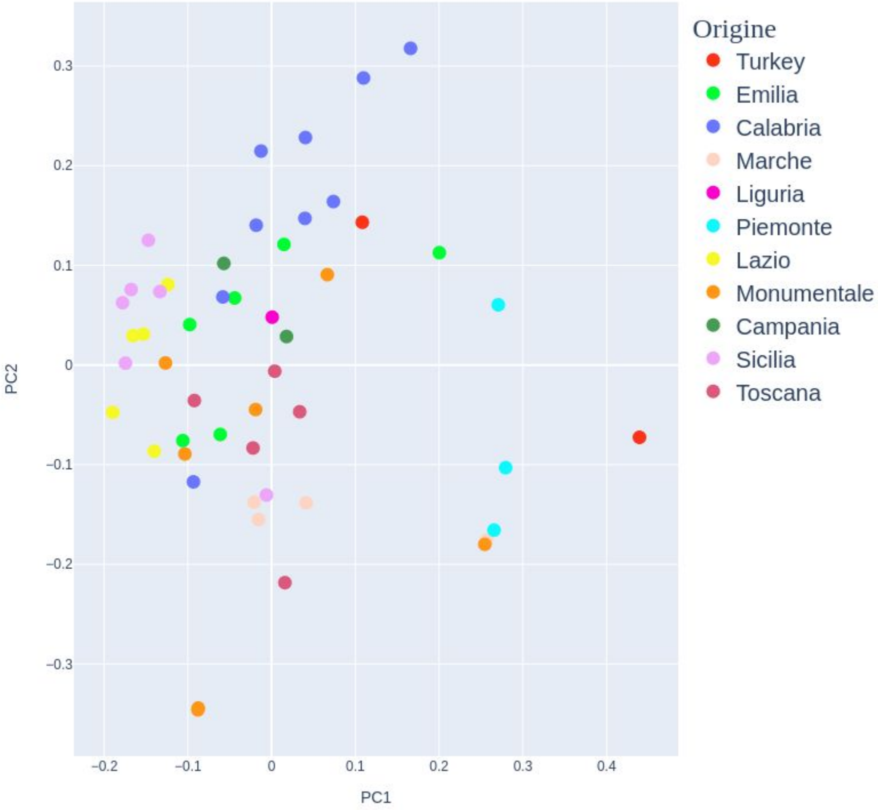
- 14 971 862 varianti con missing data < 20%
- 10 280 585 varianti con missing data < 10%





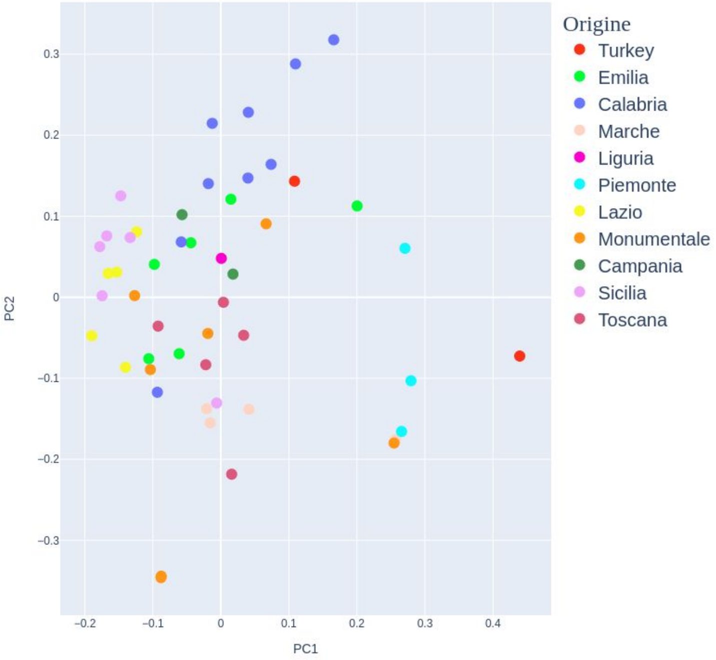
Diversità genetica

PCA basata su 10.2 milioni di SNP

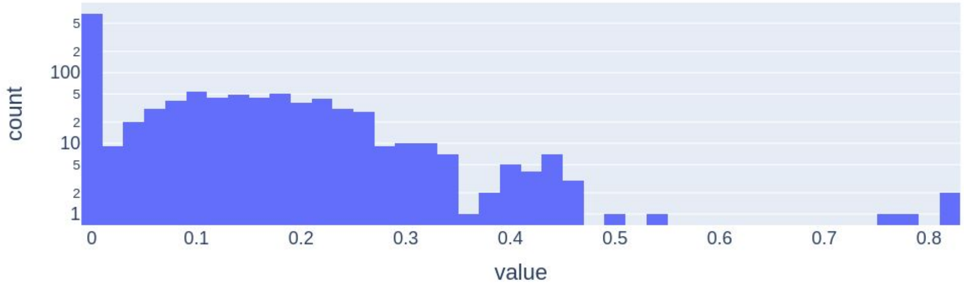


Diversità genetica

PCA basata su 10.2 milioni di SNP



Pairwise identity by descent



IID1	IID2	%IBS0	%IBS1	%IBS2
M._Val_Susa	M29	0.329100	13.706386	85.964514
M._Val_Susa	Marrone_classico	0.150837	11.242212	88.606951
M29	Marrone_classico	0.198301	11.833729	87.967970

Outline

- Assemblaggio del genoma di *Castanea sativa*

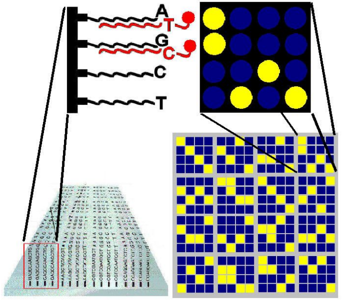
- Risequenziamento e studio della diversità genetica

- Sviluppo di un array di genotipizzazione ad alta densità

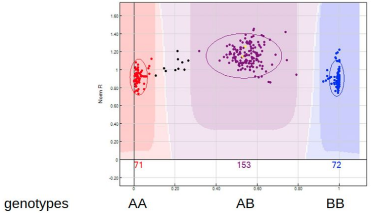
Microarray genotyping

- high density, high-throughput SNP genotyping methods
- hybridization based (i.e. allele specific oligonucleotide probes are present in the chip and hybridization signal is detected)
- closed platforms (i.e. once the probes are selected they are “fixed”)
    - prone to ascertainment bias
- highly-reproducible
- “easy” post-processing
    - software available by vendors

Affymetrix Gene Titan



[S.M. Car, Comp Biochem Physiol, 2008]



Input

Ri-sequenziamento ad alta profondità di :

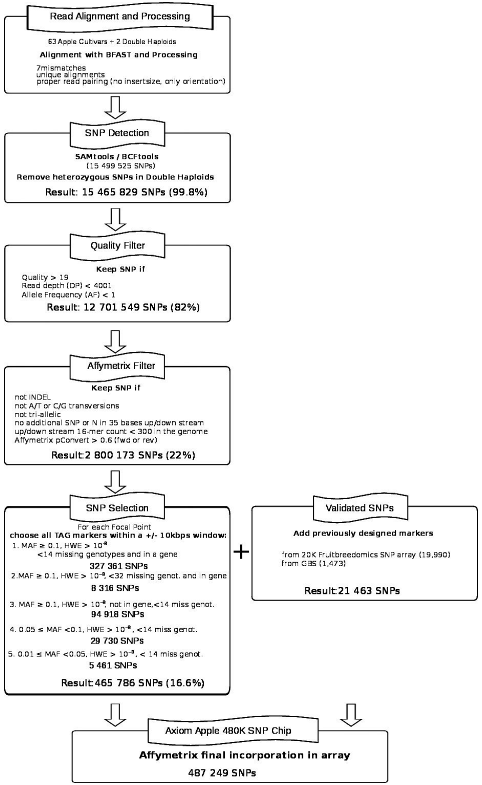
- 66 *C. sativa*
  - 50 consorzio genoma italiano
  - 16 J. Hollyday (Virginia Tech) e R. Revord (Missouri)
- 19 *C. mollissima*
  - J. Hollyday (Virginia Tech) e R. Revord (Missouri)
- 14 *C. crenata*
  - J. Hollyday (Virginia Tech) e R. Revord (Missouri)

Bioinformatics is crucial in the design

Regardless the technology…

1. Filtering of the reads, alignment and processing of the alignments
2. SNP prediction and quality filtering
3. Technology specific filtering to identify the most reliable SNPs with admissible probes
4. Selection of the target SNPs to tile in the array (ex. 487K for Apple480K)
    - project specific selection

… expertise and computational resources are needed.



[Bianco et. al, the Plant Journal, 2016]

Multispecies arrays

Two multispecies arrays to be developed by FEM for exclusive use of FEM and collaborators

- International collaboration coordinated by FEM:
  Oregon State University (hazelnut, chestnut), USDA (raspberry, hazelnut), UCDavis (grape), University of Missouri (pecan, black walnut), Cukurova University -Turkey- (pistachio), University of Catania (pistachio), Plant and Food Research -NZ- (red raspberry, blue raspberry)

FEM Axiom array 1 (provisional):

- grape (~20K), apple (~10K), red raspberry (8K), blue raspberry (~ 5K), pear (~5K), trout (~15K), walnut (~1K), chamois (~10K), pathogens (~1K)

FEM Axiom Array 2 (provisional):

- **chestnut (~25K)**, hazelnut (~25K), walnut (~1K), pistachio (~5K), black walnut (~3K), grape (~1K), apple (~1K), pathogens (~1K), …

SNP *Castanea*

~10K *C. sativa*

~7.5K *C. mollissima*

~7.5K *C. crenata*

Grazie per l’attenzione