

GENOME-WIDE DIVERSITY STUDY ON A COLLECTION OF THE ITALIAN OLIVE GERMLASM USING GENOTYPE-BY-SEQUENCING

Francesca Taranto (1), Nunzio D'Agostino (2), Salvatore Camposeo (3), Giacomo Mangini (4), Valentina Fanelli (1), Cecilia Miacola (1), Susanna Gadaleta (1), Monica Marilena Miazzi (4), Stefano Pavan (4), Valentina di Rienzo (1), Wilma Sabetta (1), Samanta Zelasco (5), Enzo Perri (5), Cinzia Montemurro (1,4).
1) SINAGRI S.r.l. - Spin Off of the University of Bari "Aldo Moro"; 2) CREA Research Centre for Vegetable and Ornamental Crops, Pontecagnano Faiano, Italy Bari, Italy; 3) Department of Agricultural and Environmental sciences, University of Bari "Aldo Moro", Bari, Italy; 4) Department of Soil, Plant and Food Sciences, University of Bari "Aldo Moro", Bari, Italy; 5) CREA Research Centre for Olive, Citrus and Tree Fruit, Rende, Italy

AIM OF RESEARCH

Herein we present our study aimed to use the GBS technology for SNP discovery and genotyping studies on a collection of 94 Italian olive varieties. The SNPs catalogue we developed was used to (i) measure genetic variation and establish the relationships among all individuals across the population as independently assessed by a parametric and a non-parametric population structure analysis software; (ii) formulate hypotheses about the origin and spread of olive cultivars on Italian territory; (iii) resolve cases of synonymy in olive germplasm;

Background

Olea europaea L. is one of the most widely grown species in the Mediterranean basin and phenotypic and genotypic diversity of modern varieties were derived either from the inter-crossing of ancient cultivars, or from the inter-breeding with wild plants. The world olive germplasm includes more than 2,600 different varieties, but olive genetic resources are poorly exploited and studied. Increasing our knowledge on olive tree genetics is a crucial step for breeding purpose and represents the future of olive growing. As far as we know, no genetic diversity studies have been published yet this species using Genotyping by Sequencing (GBS).

1 Genotyping by sequencing analysis

A set of 94 cultivars belonging to *Olea europaea* L. subsp. *europaea* var. *europaea* representing the Italian olive biodiversity, was filtered out from the collection of more than 500 olive genotypes available at CREA-OLI based on results from a pre-existing screening on technological and morphological traits. Genomic DNA was extracted from young leaves and GBS analysis was performed as described by Elshire et al (2001). Genome complexity was reduced by digesting total genomic DNA from individual samples with the *ApeKI* methylation sensitive restriction enzyme.

The olive draft genome recently published by Cruz et al. (2016) was used as reference for SNP calling. High quality SNP markers were filtered out using the *vcftools* (vcftools.github.io) with the following parameters: minimum allele frequency (MAF) ≥ 0.05 , max-missing = 0.90, Hardy Weinberg Equilibrium (hwe) = $p \leq 0.001$ and min-mean depth = 5. Furthermore, we removed single nucleotide InDels. The filtering step produced 37,792 high quality SNPs (Fig 1), that were categorized according to nucleotide substitution as either transitions (C \leftrightarrow T or A \leftrightarrow G) or transversions (A \leftrightarrow C, C \leftrightarrow G, A \leftrightarrow T, G \leftrightarrow T). A higher frequency of transitions (65%) over transversions (35%) was observed (Fig. 2), and 10,087 out of 37,792 SNPs (28.8%) were distributed in genic regions. Among these, 2,690 SNPs (26.75%) fall within annotated exons and affect 1,302 genes (Fig. 3).

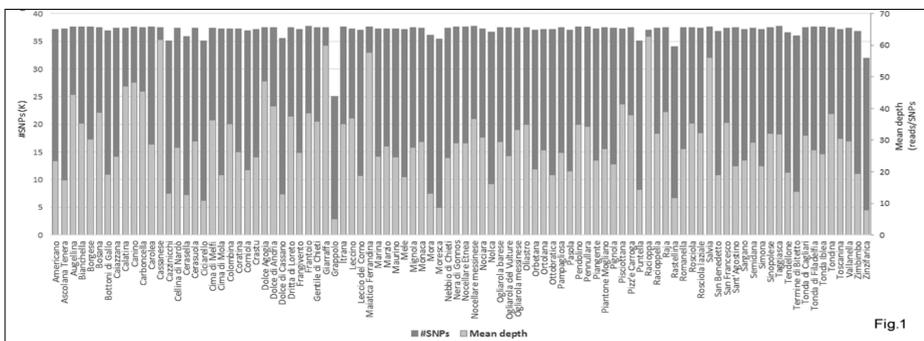


Figure 1. High quality SNP count and mean depth per olive genotype.

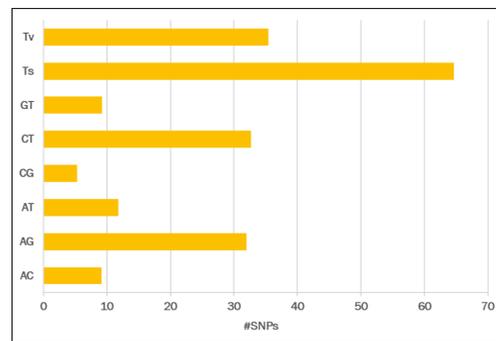


Figure 2 SNP types identified in the panel of 78 cultivars.

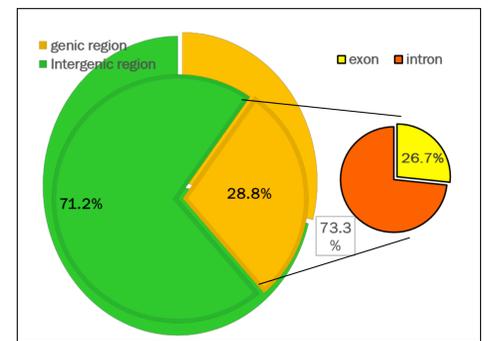


Figure 3. Pie charts describing of the genomic distribution of SNPs.

2 Population structure analysis and genetic differentiation

In this study we applied two complementary clustering methods to assess genetic diversity and establish relationships among individuals in the population under investigation: a parametric Bayesian clustering (STRUCTURE) and a non-parametric distance-based hierarchical clustering (AWclust). Structure Harvester and AWclust indicated K=4 and K=5 as the optimal number of sub-populations in which divided the olive population. Both clustering methods usually assign cultivars to similar clusters. Notably, the clustering was based on morphological trait with high heritability such as fruit size and fruit weight. It's clear that cluster C1 grouped together all olive cultivars with smaller fruit size and weight (< 2g) used as olive for oil. Cluster 2 and 3 included olive with medium fruit size and weight (2-4 g), while cluster 4 comprised cultivars with higher fruit size and fruit weight (>6 g) and used as table olive.

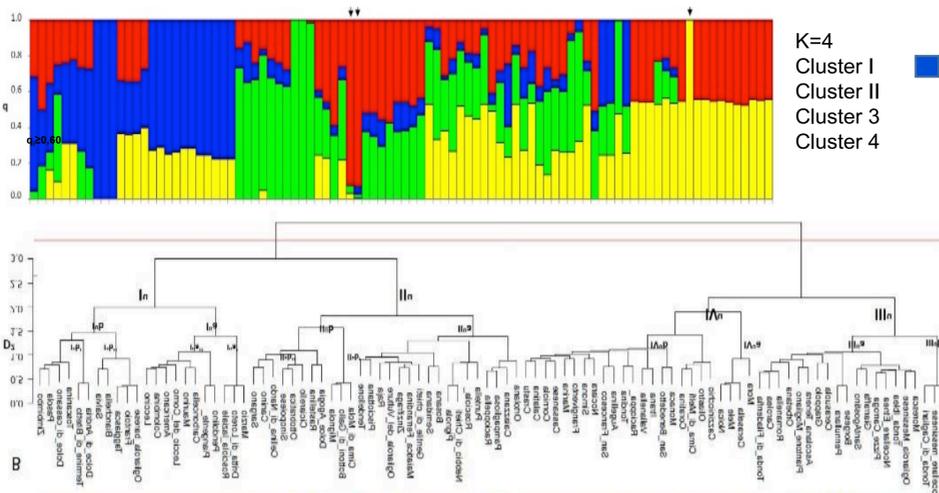


Figure 4 A) Bar-plot describe the population structure estimated by STRUCTURE. Each bar is separated into K coloured segments each representing the ancestry coefficient membership (qi) proportion in each individual. B) Dendrogram generated by Awclust.

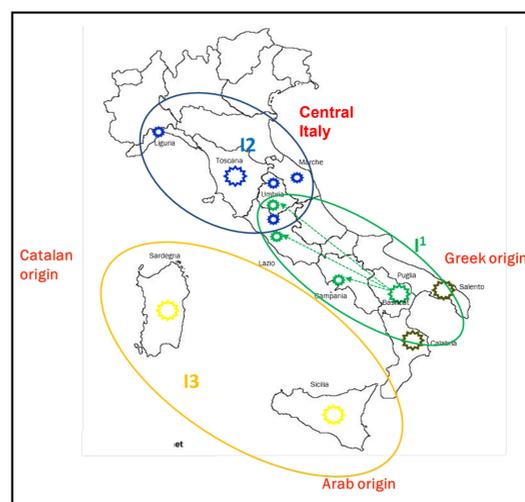


Figure 5. Origin and spread of Italian olive cultivars

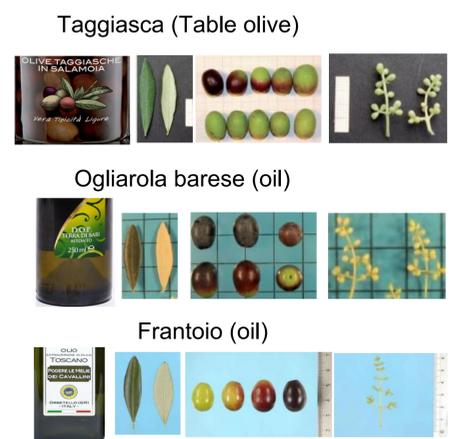


Figure 6. Cases of synonymies

On the basis of the genetic structure, we are able to identify three main gene pools within the population under investigation: I¹ includes most of the cultivars of Magno-Greek origin; I² represents most of the Italiote cultivars with admixed ancestry and I³ consists of cultivars of Arab-Catalan origin (Figure 5). In addition, the genetic population structure, discloses known or novel/unexpected cases of synonymies. (Figure 6).

ACKNOWLEDGEMENTS

PROGRAMMA SVILUPPO RURALE FEASR 2014-2020
Asse II "Miglioramento dell'Ambiente e dello Spazio Rurale" Misura 10.2.1 "Progetti per la conservazione e valorizzazione delle risorse genetiche in agricoltura" trascinamento della Misura 214 Az. 4 sub azione a) del PSR 2007-2013
Progetti integrati per la biodiversità Progetto Re.Ge.O.P."Recupero del Germoplasma Olivicolo Pugliese" Progetto di continuità e prosecuzione delle attività

