

BACKGROUND

The genetic diversity within the species *Cannabis sativa*, combined with the natural heterogeneity of its genetic makeup, presents a significant hurdle in developing marker-based identification systems that can reliably distinguish between different varieties. In the frame of the project **CaRiFIT2022 (Canapa e Ricerca Fileri Italiana 2022)**, <https://carifit.crea.gov.it/>, a 5-year national research and development project (2023-2028), funded by the Italian Ministry of Agriculture, Food Sovereignty and Forests (MASAF), the genetic characterization of eight Italian hemp varieties registered in the European catalogue of plant varieties and maintained at CREA - Research Centre for Cereal and Industrial Crops - is underway. A whole-genome resequencing (RWGS) approach was chosen, with the aim of developing genotype-specific SNP markers for varietal characterization. To achieve this goal, a total of 20 Italian and European hemp certified varieties have been sequenced using Illumina WGS technology with paired-end reads of 150 bp and a throughput of 50 Gbp/sample.

MATERIALS & METHODS

PLANT MATERIAL:

GENOTYPE	TYPE	ORIGIN	END USE
ELETTA CAMPANA	DIOECIOUS	ITALY	FIBER/BIOMASS
CODIMONO	MONOECIOUS	ITALY	SEED/BIOMASS
ASSO	DIOECIOUS	ITALY	FIBER/BIOMASS
HENOLA	MONOECIOUS	POLAND	SEED/BIOMASS
CARMALEONTE	MONOECIOUS	ITALY	SEED/BIOMASS
EARLINA	MONOECIOUS	FRANCE	SEED/BIOMASS
FELSINEA	DIOECIOUS	ITALY	FIBER/BIOMASS
OLIVIA	MONOECIOUS	ROMANIA	FIBER/SEED
USO31	MONOECIOUS	UKRAINE	SEED/BIOMASS
CARMAGNOLA	DIOECIOUS	ITALY	FIBER/BIOMASS
FUTURA75	MONOECIOUS	FRANCE	SEED/BIOMASS
FIBROR79	MONOECIOUS	FRANCE	SEED/BIOMASS
FIBRANTE	DIOECIOUS	ITALY	FIBER/BIOMASS
FERIMON	MONOECIOUS	FRANCE	SEED/BIOMASS
NASHINOIDE15	MONOECIOUS	FRANCE	FIBER/SEED
FIBRANOVA	DIOECIOUS	ITALY	FIBER/BIOMASS
KOMPOLTI	DIOECIOUS	HUNGARY	FIBER/BIOMASS
FINOLA	DIOECIOUS	FINLAND	SEED/BIOMASS
SANTHICA27	MONOECIOUS	FRANCE	SEED/BIOMASS
FEDORA17	MONOECIOUS	FRANCE	SEED/BIOMASS

Tab 1. List of the 20 hemp varieties registered subjected to RWGS. The Italian varieties maintained by CREA are indicated in red.

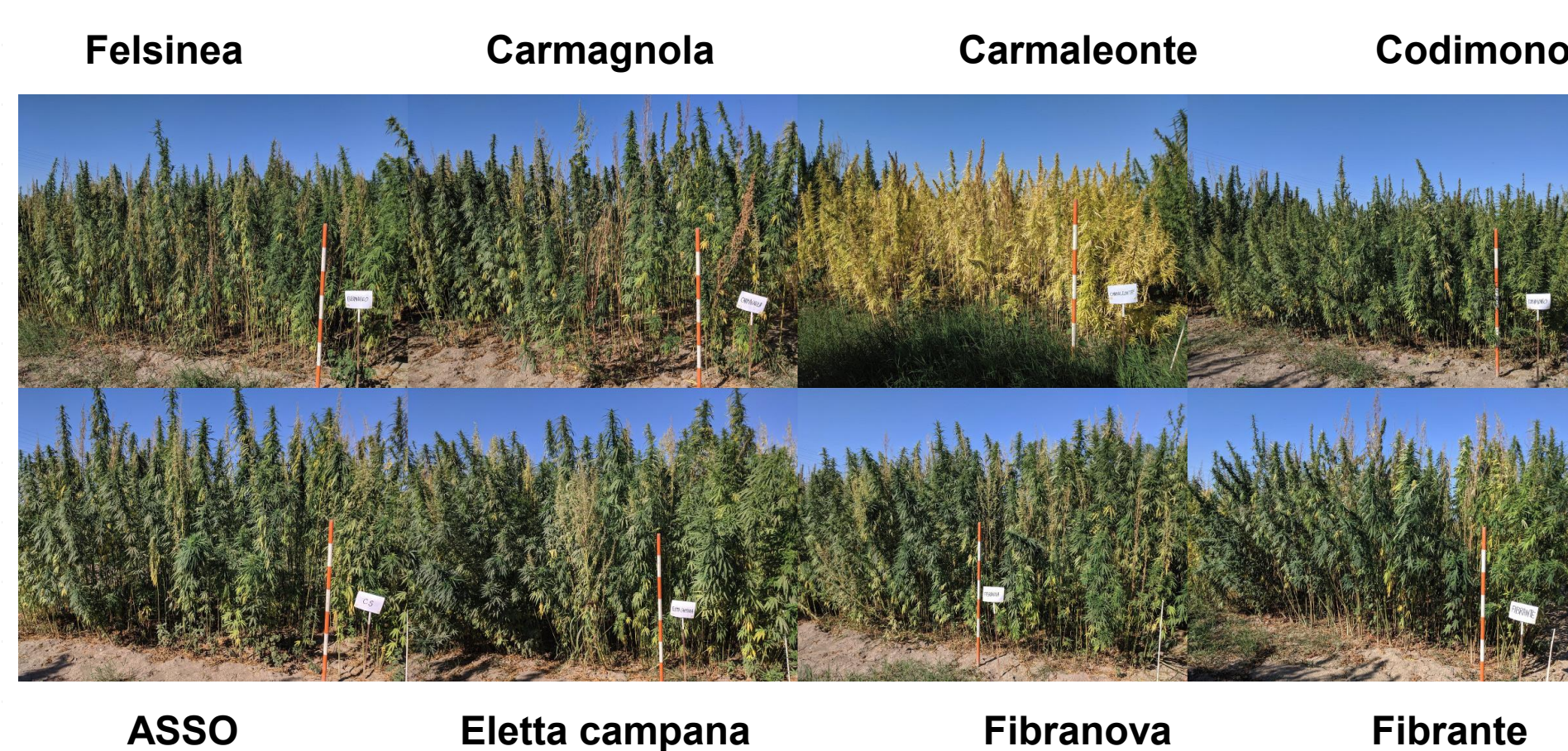


Fig 1. Cultivation in field (CREA CI, Rovigo, Italy) of the Italian hemp varieties registered in the European catalogue of plant varieties maintained at CREA and subjected to RWGS within the Task. 1.9 of the WP1 of the CARIFIT project.

METHOD:

Genetic characterization of Italian hemp varieties through **RGWS**

Step 1 DNA extraction: For 20 hemp varieties, a pool of 20 leaves collected from different shootlets was used for genomic DNA extraction using the DNeasy Plant PRO kit, Qiagen.

Step 2 RWGS: PCR free library prep and Illumina sequencing PE 150 with a throughput of 50Gb for sample were performed at BMKGNE, Germany.

Step 3 Data analysis. Bioinformatics analysis was performed by Sequentia Biotech SL, Barcelona. Workflow consist in QC of the data, Trimming and Clipping, Mapping on the *Cannabis sativa* reference genome https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_029168945.1/, Variant calling (SNP/INDELS) for identification of a panel of SNPs able to discriminate the different genotypes.

RESULTS

In total 1030.51 Gb of sequencing data were obtained with an average yield of 51.53 Gb sequencing data per sample and the percentage of Q30 bases in each sample was above 96.68%.

A subset of high-quality SNPs (4,575,635 in total) was obtained after bioinformatics analysis of RWGS data.

Performing another filtering step based on allele frequency, we selected a set of 454 SNPs based on unique genotype patterns within the genotypes of the dataset.

A minimum set of 9 SNPs able to distinguish uniquely among varieties under study was obtained, also .

The discriminating power of 454 SNPs identified among the hemp genotypes under examination is graphically shown in the PCA plot and in the Heat Map on the 20 hemp genotypes (**Fig. 2 and 3**).

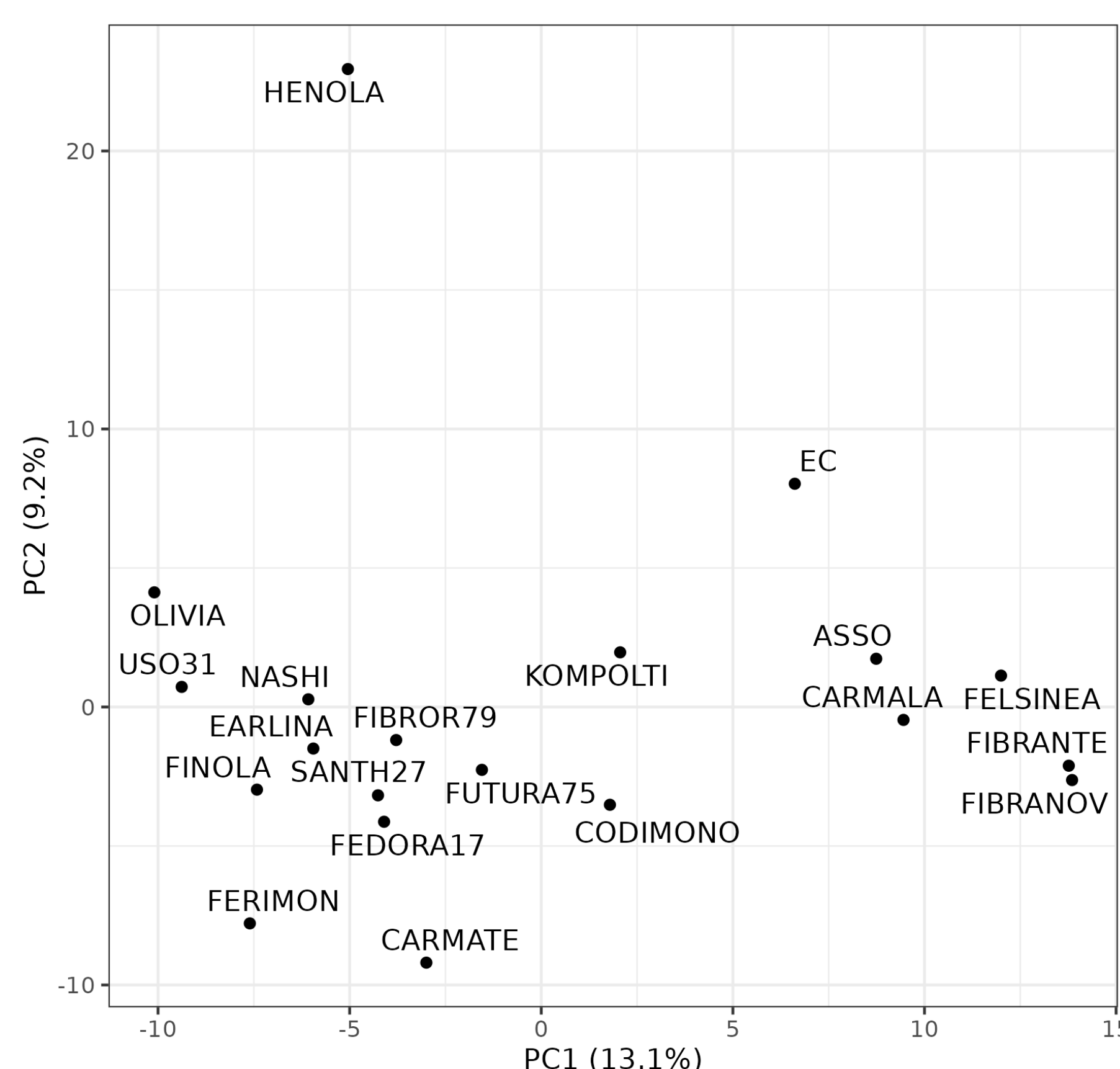


Fig 2. PCA using the 454 SNPs on the 20 hemp genotypes under study.

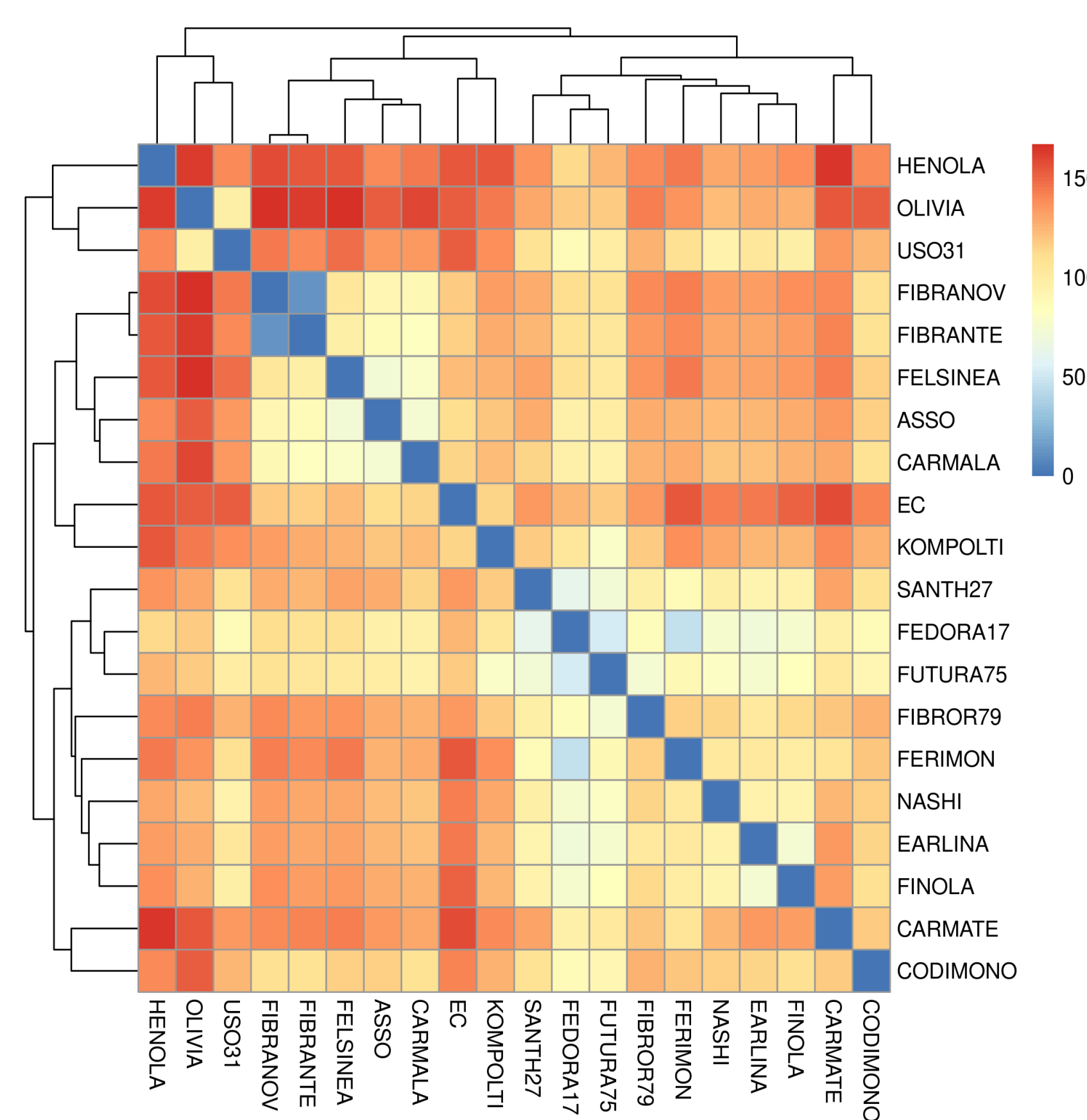


Fig 3. Heatmap showing the number of SNPs cultivar-differentiating SNPs

SUMMARY AND PROSPECTS

Based on the RWGS data, a subset of 4,575,635 high-quality SNPs was obtained. Within this SNP catalogue, a set of 454 SNPs based on unique patterns within the genotypes of the dataset, showing high reliability and effectiveness, was identified and it will be used for both breeding and intellectual property protection of the Italian hemp varieties. SNP-based molecular markers sets will be developed to define specific genetic profiles capable of distinguishing the CREA hemp varieties and obtain molecular passports for these varieties.