

## A GENOME-WIDE SNP PANEL AND HIGH RESOLUTION PHENOTYPING HELP TO DECIPHER *LOCI* UNDER SELECTION ASSOCIATED WITH FRUIT MORPHOLOGY IN *CAPSICUM*

COLONNA V.\*, D'AGOSTINO N.\*\*, GARRISON E.\*\*\*, ALBRECHTSEN A.\*\*\*\*, MEISNER J.\*\*\*\*, FACCHIANO A.\*\*\*\*\*, CARDI T.\*\*, TRIPODI P.\*\*

\*) Institute of Genetics and Biophysics, National Research Council (CNR), Naples (Italy)

\*\*) CREA Research Centre for Vegetable and Ornamental Crops, Pontecagnano Faiano (Italy)

\*\*\*) Wellcome Trust Sanger Institute, Hinxton (UK)

\*\*\*\*) Bioinformatics Center, Department of Biology, University of Copenhagen (Denmark)

\*\*\*\*\*) Institute of Food Science, National Research Council (CNR), Avellino (Italy)

*pepper, population structure, GWAS, fruit traits, selection process*

The *Capsicum* genus includes species commonly known as peppers, which among vegetables are widely appreciated for their economic importance and nutritional value. This crop, whose centre of origin is located in tropical and temperate areas of South America, includes about 30 species with different ploidy and number of chromosomes. Five diploid species ( $2n=2x=24$ ) have been domesticated and spread throughout the world after the XVI century. During these processes, fruit size and shape have been particularly subjected to selection pressure. As a result, a large number of cultivars have been developed and diverse phenotypes have been described. Herein we present the broadest and deepest study on *Capsicum* spp. using a sequencing-based genotyping approach to obtain 1.5 million SNPs across 373 accessions belonging to 11 species from 51 countries. The genome-wide SNP panel developed was used to study population structure, natural selection and genome-wide association with 45 phenotypic traits with a precision that gives new insight into the species complexes within the genus. Phenotypes were precisely evaluated through an automated procedure aimed to dissect single components of fruit size and shape. Population structure analysis recovers known species subdivisions and indicates very little genetic exchange among species. We find that the group with largest fruit has lower effective population size than one with smaller fruit. The cultivated species *C. annuum* cluster mainly in two genetic group. Two species, *C. pubescens* and *C. chacoense* evidence the lowest genetic variation, lowest effective population size, the highest kinship and the longest regions in linkage disequilibrium. These results, match observations from breeding experiments demonstrating how these species contributed relatively little to the genetic background of the common cultivars of the genus. Following this findings, we investigated positive selection and statistical association between genotypes and phenotypes related to fruit size within *C. annuum*. Twenty-one SNPs, of which five falling within genes, showed genome wide significant association with 11 traits. Several independent analyses converge on a non-synonymous variant associated to variation in most of the traits under investigation and located downstream *fs3.1*, a major QTL previously identified through bi-parental mapping. The structure and the function of the underlying protein as well as the expression of the gene are under investigation.