Proceedings of the Joint Congress SIBV-SIGA
Pisa, Italy – 19/22 September, 2017
ISBN **978-88-904570-7-4**

Poster Communication Abstract – 2.19

RAD-TAG SEQUENCING AND ASSOCIATION MAPPING IN A GLOBE ARTICHOKE GERMPLASM COLLECTION

ACQUADRO A.*, CERRUTI E.*, COMINO C.*, RODRIGUEZ M.**, PISANU A.B.***, BAGHINO L.***, SANNA D.***, ATTENE G.**, PORTIS E.*, RAU D.**

*) DISAFA Plant Genetics and Breeding, University of Torino, Grugliasco (Italy)
**) Department of Agriculture, SACEG, University of Sassari, Sassari (Italy)
***) Agris Sardegna - Agenzia Regionale per la ricerca in agricoltura - Servizio Ricerca sui Sistemi Colturali Erbacei, Cagliari (Italy)

association mapping, Cynara cardunculus, genome, population structure, RADseq

Knowledge of genetic diversity, population structure, and linkage disequilibrium (LD) extension in a target association mapping population, composed of representative and genetically diverse panel of genotypes, is a prerequisite for LD-based mapping. Until recently, these topics have not been explored in globe artichoke (*Cynara cardunculus* var. *scolymus*) due to the lack of a high-throughput, high-density marker system. Advances in "next generation sequencing" (NGS), through multiplexed sequencing of barcoded samples in a single sequencing run, have decreased experimental costs and allowed the re-sequencing of entire germplasm collections with the aim of providing a comprehensive resource on species variation.

A genome wide association (GWA) analysis in globe artichoke is here reported, based on an association panel composed of 111 varietal types came from the living collection maintained at AGRIS (Sardinia, Italy). Accessions are subdivided into seven groups based on their different morpho-phenotypic traits, primarily related to head characteristics. An extensive phenotyping of the collection for key breeding capitulum (*e.g.*, number per plant, length, diameter, fresh weight, receptacle diameter and height), plant (*e.g.*, height, number of branches, growth habit) and leaf (e.g., dimension, shape) traits have been performed for the season 2015/2016 at the AGRIS experimental station. The collection, previously investigated by means of microsatellite (SSRs) genotyping (Rau *et al.* 2015, Proceedings of the Joint Congress SIBV-SIGA), has been here characterized with a genotyping by sequencing approach, using a two-enzymes RADseq (restriction-site associated DNA sequencing) technology.

RAD-tag libraries for each genotypes were produced, pooled and sequenced on a Next-Seq500 Illumina platform (SE 1×75). Raw reads were de-multiplexed and assigned to each sample by genotype-specific barcodes (P1). Reads were adapter trimmed and quality filtered using Sickle and Scythe scripts (https://github.com/ucdavis-bioinformatics). Alignment to the globe artichoke reference genome (www.artichokegenome.unito.it) was carried out using BWA mem with default parameters and SNPs calling was performed using a Samtools-based pipeline. The resulting vcf file was filtered to keep loci having, on average, a minimum of 15 reads for sample (DP>15) and used to conduct the GWAS analysis. About 250 M raw reads (SE × 75 bp) were produced, trimmed and quality cleaned to 240 M of useful reads (∼2M reads/genotype). Mapped sequences showed an extensive coverage alongside the 17 pseudomolecules. In all 9,694 robust SNPs (DP>15) have been identified, of which about 1,400 belonged to the gene space.

Using the whole SNP dataset, modelling of population structure was obtained by applying UPGMA (unpaired group method of average), principal coordinates analysis (PCoA) and the Bayesian clustering provided by *FastStructure v1.0* software. The total genotype panel could be divided into three main groups according with their level of membership (threshold level of 705), estimated by using the admixture model for the ancestry of individuals and correlated allele frequencies. Association analyses were performed by applying the software *Tassel*. To avoid the population structure effect, a Mixed Linear Model was applied which considers both the subpopulation membership of each line inferred with *FastStructure*, and the relationship between each line provided by a kinship matrix. As a result, the high throughput sequencing of RAD-tags allowed the establishment of correlations between SNP alleles and different key breeding traits, which will prove useful for assisting in marker-aided selection.