Poster Communication Abstract – 6A.28

A COMPARISON OF METHODS FOR ASSOCIATION MAPPING IN RICE

PEDRETTI A.*, GRECO R.**, CASELLA L.**, BRUSCHI G.**, STELLA A.*,
PIFFANELLI P.**, BIFFANI S.*

*) Bioinformatics Unit, Parco Tecnologico Padano, Lodi (Italy)
**) Rice Genomics Unit, Parco Tecnologico Padano, Lodi (Italy)

Oryza sativa, Mixed Linear Model, Random Forest, association mapping, SNP

The main objective of this study was to compare alternative methods for association mapping with those implemented in the software package TASSEL, which is commonly used in plant association studies. An empirical rice data set was used which comprised phenotypic (14 morpho-physiological traits) and genotypic (383 SNPs selected in 150 target genes) data of 152 accessions of *O. sativa*.

Three methods available in R packages and a JAVA-based software (RanFog) were tested and an ad-hoc pipeline was designed to run the association analyses using R/Java scripts.

All those methods adjust for complex population structure among genotypes, which induces inflated false positive rates. The first two methods tested (EGSCORE and GRAMMAR) were taken from the R package GenABEL and are based upon a mixed model approach. The former adjusts for possible stratification by principal components, while the latter copes with the problem of population structure by using a marker-based kinship matrix. In both cases phylogenetic information obtained by STRUCTURE software was added to the mixed model as a fixed effect. The third method, EMMA, was taken from the R package EMMA (Efficient Mixed-Model Association). This method takes into account the population structure (as fixed effect) and is similar to Grammar, correcting for population structure by means of a marker-based kinship matrix. The fourth method used was a Java implementation of the Random Forest algorithm. One of the main advantages of this method is the speed of analysis, which would become more relevant in case of association studies using high-density SNP panels (GWAS).

Using the results obtained from TASSEL as the reference set, we found that EMMA attained the best performances giving the higher number of SNPs in common with TASSEL significantly associated with the analyzed traits.