Proceedings of the LX SIGA Annual Congress
Catania, Italy – 13/16 September, 2016
ISBN **978-88-904570-6-7**

Poster Communication Abstract – 8.16

MAIZE PANGENOME ESTIMATION FROM NEXT-GENERATION SEQUENCING DATA ANALYSIS

ZAPPAROLI E.*,**, MAGRIS G.*,**, VIDOTTO M.*,**, MARRONI F.*,**, MORGANTE M.*,**

*) Dipartimento di Scienze agroalimentari, ambientali e animali (Di4a), University of Udine, Via delle Scienze 206, 33100 Udine (Italy)
**) IGA, Istituto di Genomica Applicata, Via Jacopo Linussio 51, 33100 Udine (Italy)

genomics, plants pan-genomes, Zea mays, structural variants, transposable elements

The pan-genome describes all the genomic sequences present in a population of different varieties of the same species and it is commonly applied in bacterial genomics.

The sequences present in all varieties represent the "core" genome, while sequences present in some varieties and absent in others are named "dispensable" genome and represent the variable portion of the pan-genome.

The bacteria pan-genome concept has been successfully applied to plants in recent years.

Some plants pan-genomes (i.e. *A. thaliana, G. soja, Z. mays, V. vinifera*), are currently under study in order to quantify and characterize the shared and not shared portion of the genome between varieties, respectively the core genome and the dispensable genome.

Structural variants (SVs) are an important source of genetic variation in plants, mostly due to large (>1000bp) insertions and deletions of transposable elements (TEs).

The identification of structural variants (SVs) is a strategy to characterize the dispensable genome of plants.

Here, we apply this strategy to characterize the maize pan-genome of 6 varieties selected from the parental lines of the MAGIC maize population: A632, H99, HP301, F7, Mo17, W153R.

SVs were identified using next-generation sequencing (NGS) data of each variety and analysed in order to characterize the dispensable portion of the pan-genome in respect to the varieties analysed.

Integrating paired-end mapping (PEM) and split-read mapping (SR) approaches, we obtained a wide collection of high-confident deletions (20~30,000 hits per variety) and insertions (14~29,000 hits per variety) within varieties.

While the maize B73 reference genome size is around 2,5 Gb (2 Gb excluding scaffolds), we identified 561 Mb present in the reference and absent in at least one of our varieties (namely deletions, at least 221 Mb for one sample), and 967 Mb absent in the reference and present in other varieties (namely insertions, at least 173 Mb for one sample).

This suggests that a large part of the reference genome is dispensable, confirming the previous estimations.

Further efforts are underway to improve characterization of the pan-genome leveraging information obtained with de-novo assembly and with the use of longer reads.