Proceedings of the LXVI SIGA Annual Congress
Bari, 5/8 September, 2023
ISBN: **978-88-944843-4-2**

Poster Communication Abstract – 6.21

GENOME-WIDE ANALYSIS OF GENE HETEROZYGOSIS AND ALLELE DIFFERENTIAL EXPRESSION IN THE FIG TREE

USAI G.*, VANGELISTI A.*, CASTELLACCI M.*, SIMONI S.*, MASCAGNI F.*, NATALI L.*, CAVALLINI A.*, GIORDANI T.*

*) Department of Agriculture, Food and Environment, University of Pisa, Pisa, Italy

heterozygosity, phased genome, differential allelic expression, fig tree

Heterozygosity is a genetic condition known since Mendel's foundation of Genetics. It is exploited in agriculture, particularly in developing hybrid cultivars and vegetatively propagated heterozygous crops, including most fruit trees. Heterozygosity can occur at both the DNA sequence and epigenetic levels. However, understanding the genome-wide effects of heterozygosity remains challenging due to the complexities involved in assembling the two haplotypes of diploid heterozygous individuals.

In this study, we investigated whole-genome heterozygosity and its impact on gene and allele expression in the fig tree (*Ficus carica* L.), a nutritionally important fruit tree resilient to environmental changes. Using long DNA read sequencing and chromosome conformation capture technologies, we accurately sequenced and assembled the approximately 356 Mbp fig genome into two homologous chromosomal sets. This enabled us to investigate the extent and effects of heterozygosity on genome function, particularly gene and allele expression.

Overall, we identified 20,441 allelic gene pairs in the haplotype-phased fig genome. Considering the CDS, 13,331 gene pairs were homozygous, while 7,110 exhibited heterozygosity. The heterozygous gene pairs were further categorized into three functional categories: 3,311 structural proteins, 2,664 enzymes, 477 transcription factors, and 658 remained uncharacterized. These categories exhibited an average Ka/Ks ratio of 0.66.

Gene expression was then studied in the leaves of plants subjected or not to saline stress for 48 days. Out of the 7,110 heterozygous gene pairs,

5,067 were found to be expressed. Among the expressed genes, 14.41% in the control group and 18.77% in the treated group exhibited differential allelic expression (DAE), occurring in either the control, the salt treatment, or in both conditions. The percentage was higher for genes encoding structural proteins and lower for those encoding transcription factors. Generally, the more expressed allele was found to be more expressed in both the control and treated groups. Only 0.14% of the differentially expressed genes (DEGs) showed DAE only in the control or only in the treated group, indicating that only one of the two alleles was regulated by saline stress.

In addition, a reduced impact of sequence variations in the promoter regions related to allelic expression was observed. At genome-wide level, minor variations in upstream gene promoters were apparently related to higher DAE levels than large variations in the same regions.

In conclusion, our findings evidence the rate of DAE at genome-wide level. The difference in proximal promoter regions appears important in determining DAE. However, further investigations are required to evaluate the contributions of distant enhancers and epigenetic changes to DAE. This genome-wide analysis represents an initial step towards understanding the broader implications of genome heterozygosity.