Proceedings of the Joint Meeting AGI-SIBV-SIGA
Assisi, Italy – 19/22 September, 2011
ISBN **978-88-904570-2-9**

Oral Communication Abstract – 2A.06

INTRIGUING ISSUES FROM A HIGHLY DUPLICATED GENOME: AN EXAMPLE FROM TRANSCRIPTION FACTOR GENE PARALOGS

VIGILANTE A[#], SANGIOVANNI M.[#], FRUSCIANTE L., CHIUSANO M.L.

Department of Soil, Plant, Environmental and Animal Production Sciences, University of Naples "Federico II", Via Università 100, 80055 Portici (Italy)
#) These authors contribute equally to the work.

Transcription factors, whole genome duplication (WGD), paralogs

Gene duplication followed by functional diversification of the duplicated genes (paralogs) is a major driver of evolution. There are evolutionary scenarios where paralogs are significantly over-retained following whole genome duplication events (WGDs) but at the same time they may exhibit lower retention rates after smaller scale duplications.

The model diploid plant *Arabidopsis thaliana* underwent several rounds of WGDs, followed by reduction and reshuffling of the gene content. Therefore, an all-against-all protein sequence similarity search allowed the identification of all the possible pair-wise similarities between genes, classifying structurally related ones into networks of paralogs. The data have been organized in a user-friendly web accessible database.

To further exploit these data, we focused on transcription factor genes. In fact, since the presence of widespread intra-genome duplications, together with the loss of gene copies, the interpretation and the study of the evolution of transcription factor gene families is very complicated and this threats the role of this genome as a reference in plant comparative genomics. Moreover, due to their key roles in gene regulation, transcription factor are among the best examples of dosage-sensitive genes.

Our effort required to overcome one of the major limitation to studying TFs in *A. thaliana*, i.e. the lack of a reliable and unique annotation, a challenge that is compounded by the presence of many dedicated databases and several methods for the identification of genes encoding DNA-binding domains. In a first step, we focused on well-known TFs collections to validate and integrate their data. Then, we performed a deep investigation on transcription factors organization within the *A. thaliana* genome, via the analysis of paralogs.

Our approach provides support to the classification of TFs in *A. thaliana* and represents a step forward to understand TF family organization and evolution.

The analysis here presented confirms the usefulness in exploiting the collection of network of paralog genes in *A. thaliana*, since it permits an appropriate investigation of gene families and reveals interesting issues concerning this reference plant genome.