Proceedings of the 52nd Italian Society of Agricultural Genetics Annual Congress
Padova, Italy – 14/17 September, 2008
ISBN **978-88-900622-8-5**

Oral Communication Abstract – 1.04

HALFWAY CHECKPOINT! WHAT CAN WE LEARN FROM THE DRAFT AND INCOMPLETE TOMATO GENOME?

CHIUSANO M.L., TRAINI A., D'AGOSTINO N., FRUSCIANTE L.

Dept. of Soil, Plant, Environmental and Animal Production Sciences, University of Naples
'Federico II', Via Università 100, 80055 Portici, Naples

tomato genome, tomato transcriptome, genome annotation, genome properties

The 950Mb tomato genome is structured into distal, gene-rich euchromatin and gene-poor peri-centromeric heterochromatin. As a consequence, the sequencing strategy within the international Tomato Genome Project is to attempt to initially sequence the euchromatic, gene rich portions of the genome, which make up one quarter (220Mb) of the tomato genomic sequence (Peterson *et al*., 1996), while including more than 90% of the genes (Wang *et al.*, 2006).

To render the emerging tomato sequence immediately useful to the community, while the international Tomato Annotation Group (ITAG) is still setting up an official annotation, we designed and implemented ISOL@ (http://biosrv.cab.unina.it/isola/), an **I**talian **SOLA**naceae genomics resource. ISOL@ was designed to provide full-web-access on details of the genome annotation based on experimental evidence as derived from EST/full-length cDNA sequences (Chiusano et al., 2008). The platform, which is conceived as a multi-level computational environment, can be currently accessed through two convenient gateways. The '*transcriptome*' gateway provides an access point to explore sixteen EST collections from different *Solanaceae* (14) and *Rubiaceae* (2) species and the corresponding virtual transcripts generated by assembling ESTs into Tentative Consensus sequences (TCs).

The '*genome*' gateway allows to browse the annotation of the *S. lycopersicum* BAC sequences which consists mainly on the spliced-alignments of the EST/TC collections included in the local EST databases (TomatEST, PotatEST, SOLEST). The EST-based identification of the '*expressed'* loci is exploited by the GeneModelEST tool (D'Agostino et al. 2007) for the definition of reliable gene models and for the detection of putative alternative transcripts.

Annotations also include repeat analysis performed by RepeatMasker which scans for known repeats according to the SGN Tomato UniRepeats database.

A preliminary view of the draft tomato genome highlights that a number of BACs show low gene and a high repeat content. In addition, a strong negative correlation between gene and repeat content supports the general assumption that genes are predominantly present in the relatively repeat-poor euchromatin. The tomato heterochromatin consists of the bulk of the repetitive DNA fraction, which nevertheless contains some genes as has been described by Yasuhara and Wakimoto (2006). Finally, a preliminary functional classification of gene products, the distribution of gene family along chromosomes and comparative genomic studies have been undertaken exploiting ISOL@ as well as *Soolgle*. This last is an in-house-developed web search engine representing a quick route for identifying ortholog sequences between tomato and the model plant *Arabidopsis thaliana*.