

# Revealing the chromosome organization of the emerging tomato genome

Di Filippo M<sup>1</sup>, Traini A<sup>1</sup>, D'Agostino N<sup>1</sup>, Frusciante L<sup>1</sup>, Chiusano ML<sup>1</sup>

## Motivation

In the attempt to provide further insights into the nature, composition and role of heterochromatin and euchromatin in plant genomes, we analyzed the first tomato (*Solanum lycopersicum*) genome draft (Mueller et al., 2009) which was obtained by a BAC (Bacterial Artificial Chromosome) based sequencing effort focused on the euchromatin region (~250 Mb) of the tomato genome which is also expected to be gene richer (Peterson et al., 1996).

## Methods

ISOL@ is conceived as a multi-level computational environment accessible through different gateways. The 'transcriptome' gateway provides an access point to explore publicly available EST collections from Solanaceae species. The 'genome' gateway allows to browse the EST-based annotation of the tomato BAC sequences. The resulting identification of the 'expressed' loci is exploited for the definition of gene models and for the identification of putative alternative transcripts. Annotations of repetitive elements, based on the Plant Repeats database at Michigan State University, are included as well. Both gene and repeat content for each BAC were then calculated as percentage of nucleotides covered by *S. lycopersicum* ESTs and by annotated repeats.

## Results

Solanaceae transcriptome data which are part of the Italian Solanaceae Platform (ISOL@, Chiusano et al., 2008, accessible at <http://biosrv.cab.unina.it/isola/>), integrated with the available tomato genome sequences, allowed us to perform preliminary investigations on structural features of the tomato genome. We considered the repeat content and gene distribution for each BAC, providing a quick and an high resolution preview of the genome composition. In fact, despite the generally accepted idea that the euchromatic regions were gene rich, we found a number of BACs showing low gene and high repeat content outside the pericentromeric regions. In addition, the screening of the gene and the repeat content per BAC revealed: (i) large variability in both repeat and gene coverage along chromosomes; (ii) in spite of this variability, repeat- or gene-richer BACs generally were organized as blocks sharing similar composition; (iii) repeat-richer or gene-

---

<sup>1</sup> Dept. of Soil, Plant, Environmental and Animal Production Sciences, University of Naples Federico II

richer blocks can be associated to heterochromatic or to euchromatic regions, respectively; iv) repeat-rich blocks, which are evident in both pericentromeric and extra-pericentromeric regions, show similar compositional properties; v) regions of inverted repeats are found mainly associated to heterochromatin, suggesting their role in chromatin compaction. In conclusion, although the tomato genome here analysed is only made of 120 Mb, an integrated bioinformatics platform and novel computational strategies permitted: i) to reveal a typical design of the emerging tomato chromosomes and pave the way for further investigations on the relationship between DNA primary structure and chromatin organization in Solanaceae genomes

**Availability**

<http://biosrv.cab.unina.it/isola/>

**Contact e-mail**

[miriam.difilippo@gmail.com](mailto:miriam.difilippo@gmail.com)