Poster Abstract – D.51

ON THE EVALUATION OF COMPUTATIONAL TOOLS FOR MAPPING EXPRESSED SEQUENCES ONTO GENOMIC DATA

A. TRAINI, M. L. CHIUSANO

Department of Structural and Functional Biology, University 'Federico II', 80134 Naples, Italy

spliced-alignment, genome annotation, gene model

Curated genome annotation is a challenging computational problem in genomics. Many stand-alone programs are available for spliced aligning expressed sequence tags onto a genome sequence. With the goal to contribute to the bioinformatics efforts of the Tomato Genome Sequencing Project we tested the software proposed by the International Committee for cDNA/EST to genome mapping. We investigated which algorithm is more reliable and in which context, comparing and evaluating some of the most frequently used specialized software, to provide a reference data-set of tomato *gene models*. We tested sim4 (Florea *et al.*, Genome Research 1998) based on the BLASTZ algorithm (Schwartz *et al.*, Genome Research 2003), GALAHAD, included in the Grail-EXP package (Hyatt *et al.*, Genome Sequencing and Biology Meeting May 2000), blEST (Florea *et al.*, Genome Research 1998) and SIBsim4 (Florea *et al.*, Genome Research 1998), all based on sim4 algorithm, and GeneSeqer (Brendel *et al.*, Bioinformatics 2004) as software commonly proposed for solving the task of mapping cDNAs/ESTs to genomic sequences. To provide a reliable informative workbench for tomato genome annotation based on experimental data, we set up a Gbrowse based platform reporting the results from the different methods. The datasets used in the present analysis are the tomato expressed sequences available from dbEST and the BAC sequences from the genome sequencing effort at the SOL Genomics Network (http://www.sgn.cornell.edu). We present here the results of our evaluation of the software considered and we propose multiple algorithm usage under different constraints to provide exhaustive and reliable information when experimentally annotating genome sequence data.