

COMPUTATIONAL GO ANNOTATION OF LARGE ORGAN-SPECIFIC EST REPERTORIES REVEALS COMMON FEATURES OF AFLP TECHNOLOGY FOR NON-MODEL PLANT TRANSCRIPTOMICS

GALLA G., BOTTON A., RAMINA A., BARCACCIA G.

Department of Environmental Agronomy and Crop Science, Faculty of Agriculture of the University of Padova, Viale Università 16, Campus of Agripolis, 35020 Legnaro (Italy)

gene ontology, functional annotation, metabolic pathways, regulatory networks, candidate genes

After 10-year-use of AFLP (Amplified Fragment Length Polymorphism) technology for DNA fingerprinting and mRNA profiling, large repertoires of genome- and transcriptome-derived sequences are available in public databases for model, crop and tree species. AFLP-based marker systems have been and are being extensively exploited for genome scanning and gene mapping, as well as cDNA-AFLP for transcriptome profiling and differentially expressed gene cloning. The evaluation, annotation and classification of genomic markers and expressed transcripts would be of great utility for both functional genomics and systems biology research in plants. This goal may be achieved by means of the Gene Ontology (GO), consisting in three structured vocabularies describing genes, transcripts and proteins of any organism in terms of their associated cellular component, biological process and molecular function in a species-independent manner. Our work deals with the computational GO annotation of about 8,000 AFLP-derived EST clones retrieved from both publically available databases and unpublished collections.

Descriptive statistics on the type, size and nature of gene sequences obtained by means of AFLP technology were preliminary calculated. The gene products associated with mRNA transcripts were then classified according to basic terminology of the three main GO vocabularies for either single species or botanic families. The adopted Blast2GO software was shown to represent a comprehensive bioinformatics solution for an annotation-based functional analysis. A comparison of the functional content of cDNA-AFLP records sorted by plant organs was also performed by splitting the sequence database into monocots and dicots and by comparing them to all annotated EST datasets of rice and Arabidopsis, respectively. Our findings demonstrated that reliable GO annotations of AFLP-derived sequences can only be gathered through the optimization of the experimental steps and the statistical parameters adopted. Finally, organ-specific cDNA-AFLP sequences from dicots and monocots were used to query the Arabidopsis and rice proteomes, respectively, and the two output sequence collections were then linked to the original microarray probe datasets provided by Affimetrix. The vast majority of our transcript-derived fragments were successfully incorporated in biosynthetic pathways and regulatory networks by using the MapMan software. Qualitative metabolism maps along with regulatory and biosynthetic overviews were developed for specific plant organs across botanic families leading to the definition of key functional categories and groups. Such an exhaustive annotation seems to offer a suitable platform for performing functional genomics and selecting candidate genes, particularly useful in non-model species. According to the whole set of computational-inferred GO annotations, the AFLP technology generates thorough information for angiosperm gene products and shares common

features across angiosperm species and families. Actual possibilities of using AFLP-derived tags on cDNA fragments produced directly by sequencing-by-synthesis technologies opens up the possibility of not only identifying very large numbers of expressed genes, but also retrieving large-scale SNP collections. The utility of this technology for structural and functional genomics in plants can be implemented by serial annotation analyses of genome-anchored fragments and organ/tissue-specific repertoires of transcriptome-derived fragments.