

The ESTree db as an engine for peach EST related information retrieval

Caprera A.(1), Lazzari B.(2), Vecchietti A.(2), Stella A.(2), Milanesi L.(3), Pozzi C.(2)

(1) CISI, Milano (2) Parco Tecnologico Padano CERSA, Lodi (3) ITB-CNR, Segrate (MI)

Motivation

The ESTree db represents a collection of *Prunus persica* expressed sequenced tags (ESTs) and is intended as a resource for peach functional genomics. With this aim, the db has been structured to be a repository of information and links related to the sequences and to provide a user-friendly interface to allow easy querying of all the db fields. Within the month of March 2005 the third release of ESTree will be online and will include 18630 sequences, encompassing 8 libraries at different peach fruit developmental stages. A second version of the db, including only the 6155 sequences produced by the FFTP-CERSA group will also be online. The major data resources that are included in the ESTree db are: annotation both with BLASTx versus the NCBI nr database and with BLASTx versus the GO viridiplantae subset of sp-trembl; contig assembly and display; SNP analysis with AutoSNP and links to the KEGG metabolic pathways and the enzyme entries of NiceZyme (Expasy). Gene Ontology statistics are also presented both for the whole set of sequences that are included in the db and for library-specific subsets.

Methods

The ESTree db core structure is a MySQL relational database at the moment composed of 15 tables where all the results of the sequence analysis pipeline are stored. This analysis pipeline is a fully automatic procedure that starting from raw EST sequences fills up all the tables in the database. The SNP analysis has been performed with AutoSNP and Gene Ontology annotation has been performed online with the GOblet program. Statistics on Gene Ontology are shown in a graphical representation of the Gene Ontology Tree dynamically created using ontology definitions and their child-parent relations. The web interface allows querying the db via a text search facility that performs independent or integrated queries on all the db fields and also allows restricting the queries to a specific library or to the unigene or putative-SNP-containing sequences subsets. BLAST analysis is allowed on the db sequences and on the already described sequence subsets. Download of sequences and program outputs is also allowed.

Results

The resulting db is a comprehensive resource of data and links related to the peach EST sequences. The Sequence Report and Contig Report pages work as the web interface core structures, giving quick access to data related to each sequence/contig. Surfing and querying the db is intended to be user-friendly and data and links are accessible from multiple pages, to allow easy information retrieval.

Contact email: andrea.caprera@itb.cnr.it

URL: <http://www.itb.cnr.it/estree/>