Gene mining in barley: generation and database analysis of ESTs

W. Michalek, A. Graner and K.-P. Pleissner

Institute for Plant Genetics and Crop Plant Research, Gatersleben, Germany

email: pleiss@ipk-gatersleben.de

The partial sequencing of cDNA clones allows the identification of genes in a relatively cost effective manner. In conjunction with the fast evolving information content of public databases it is possible to assign putative functions to a growing number of ESTs.

Currently we are sequencing clones within three different cDNA libraries made from mRNA of the barley cultivar `Barke´, which is a high quality spring barley variety used for malting. The libraries were made from etiolated seedlings, roots, and developing caryopsis respectively. Until now, 4279 EST sequences were generated, mainly from the 5´-end.

For the assignment of putative functions the EST sequences were used for similarity search against the SwissPirPlus database. We employed the BlastX2–program available at the Heidelberg Unix Sequence Analysis Resources (HUSAR). The searches were performed only with ESTs exceeding a sequence length of 200 bp. Using BlastX2 output a BARLEY EST-DATABASE managing these search results was created. The database consists of a search table and the BlastX2 documents in HTML-format. The search table contains the names of the documents as well as the entire text information of each protein found by BlastX2 search. Using this table a search for a protein (text as any word, several words or partial strings) can be carried out to link all documents containing these search pattern. The text string search is realized by a CGI-script written in C. Finally, a document can be selected to present the entire result of the BlastX2 search and the original EST name for which the search was done. The BARLEY EST-DATABASE is currently available as an Intranet solution.

Apart from continued sequencing the further goals are the development of new libraries, the advancement of the BARLEY EST-DATABASE and the implementation of bioinformatic tools (e.g. analysis of redundancy, EST clustering)