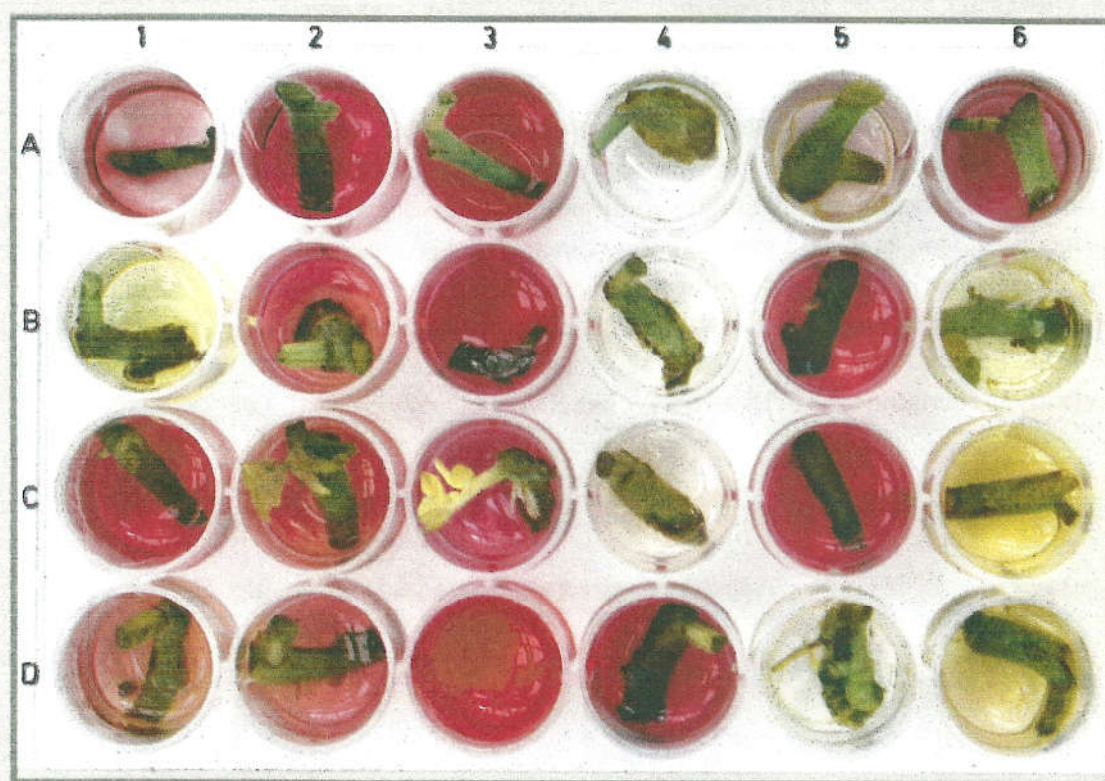# 6th International Symposium in the Series
RECENT ADVANCES IN PLANT BIOTECHNOLOGY

FROM LABORATORY TO BUSINESS

Book of Abstracts

September 12–16, 2005
České Budějovice, Czech Republic

DEVELOPMENT OF UTILITIES FOR TRA AND E-TRA *IN SILICO* MINING PROGRAMS: ORGMINER AND KEYWORD FINDER

M Karaca. M. Bilgen, A.G. İnce, A.N. Onus

Akdeniz University, Faculty of Agriculture, 07059, Antalya, Turkey

Recently we developed two tandem repeat analyzer programs: combining sequence motif searches with keywords such as organs, tissues, cell lines or development stages for finding exact, inexact and compound tandem repeats (TRs). Tandem Repeats Analyzer 1.5, TRA, and Exact Tandem Repeats Analyzer, E-TRA. These two programs differ from other repeat finding software such as Tandem Repeats Finder, TRF, REPutcr, Simple Sequence Repeat Identification Tool, SSRIT, Simple Sequence Repeat Finder, SSRF, Search for Tandem Repeats IN Genomes, STRING, and Microsatellite Search, MISA. With the exception of TRA and E-TRA, none of these programs provide information about the distribution of repeats among organisms, organs, tissues, cell types or development stages when multi-sequences or organs are used. The difference between TRA and E-TRA is that compound, imperfect, and extended compound repeats are only detectable in E-TRA. With the advent of new biotechnological tools number of DNA sequences being collected in GenBank are rapidly proliferating. Such as the date of May 13, 2005, there were 26,858,818 ESTs sequences deposited in GenBank (http://www.ncbi.nlm.nih.gov). The mining and automatic identification of these sequences using computational methods have become a necessary step. Here we report on the development of two computational tools that make easier to sort, collect DNA sequence data to be used in TRA and E-TRA programs. Briefly, the OrgMiner utility sorts GenBank and EST formatted files according to user request. For instance researcher may wish to collect ESTs or GenBank sequences of a particular organism that would be mined in TRA or E-TRA. The number of organism in the organism list file is unlimited. The other utility reported here is the KeyWord Finder program. This program collects keyword (such as organs, tissues, cell lines or development stages) for TRA and E-TRA. Each utility has its detailed help file. Its outputs show results such as number of keywords found for particular organisms, organism lists and library names. Two utility programs were written in C++ using Microsoft Visual C++ software and can run on Windows 98, Windows NT, Windows ME and Windows XR These programs are freely available for non-commercial users by request from the authors

Keywords:
Tandem Repeats, SSR-Markers, Microsatellites, Expansion, Database