E

BMT/10/9
ORIGINAL: English
DATE: October 17, 2006

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES AND DNA PROFILING IN PARTICULAR

Tenth Session
Seoul, November 21 to 23, 2006

THE USE OF A BIONUMERICS DATABASE FOR THE RAPID IDENTIFICATION OF POTATO CULTIVARS

Document prepared by experts from the United Kingdom

THE USE OF A BIONUMERICS DATABASE FOR THE RAPID IDENTIFICATION OF POTATO CULTIVARS

Alex Reid

Scottish Agricultural Science Agency, 1 Roddinglaw Road,
Edinburgh, EH12 9FJ, United Kingdom
email: alex.reid@sasa.gsi.gov.uk

Introduction

The need for an efficient system to be able to store, retrieve and interrogate data (both molecular and morphological) pertaining to agriculturally important crops was discussed at the ninth session of the Working Group on Biochemical and Molecular Techniques, and DNA-Profiling in Particular (BMT), held in Washington, D.C., June 21 to 23, 2005 (see document TWC/23/7-BMT/9/3 "Creation of databases for molecular markers: one approach taking into account possible cooperation between authorities/laboratories"). This paper will give an overview of the system currently in use at the Scottish Agricultural Science Agency (SASA) (United Kingdom) and Naktuinbouw (Netherlands) for the Community Plant Variety Office (CPVO) funded project to construct a database containing SSR and key morphological characteristics for the potato varieties on the European Union common catalogue.

Allele scoring and data storage

The potato varieties on the EU common catalogue were analyzed with nine SSR markers which yielded a total of 88 possible alleles. Alleles were scored as present or absent and stored in an Access database using a separate table for each marker.

The BioNumerics platform

BioNumerics is a software package developed and distributed by Applied Maths (www.applied-maths.com). The software is designed to allow the integration of different types of biological data as well as the analysis of these data. Links are easy to establish with external databases (such as Access) using ODBC (open database connectivity) protocols. The different types of data ('Experiment types') that can be stored in BioNumerics include:

(1) 1-D gel electrophoresis such as RFLP and AFLP (imported as TIFF files). Size standards run on these gels allow the normalization of different electrophoresis runs with high accuracy;
(2) Characteristic types including binary data (e.g. SSRs) and multi-state characteristics (e.g. phenotypic characteristics or fatty acid profiles);
(3) Sequence data for both DNA and RNA (the program accepts EMBL, Fasta and GenBank formats). In addition raw chromatograms from automated sequencers can be imported and assembled within BioNumerics;
(4) 2-D gel electrophoresis (also imported as TIFF files); and
(5) Similarity matrix data generated by other applications.

Once these data have been imported into BioNumerics they can be analyzed by several different methods:

(1) Clustering methods such as UPGMA, Ward, Neighbor Joining, maximum parsimony with bootstrapping and maximum likelihood. Both rooted and unrooted trees can be produced.
(2) Dimensioning techniques such as principle component analysis and multi-dimensional scaling.

It is possible to analyze data from a single experiment type (in the case of SSRs this could be a single marker) or to pool data from several experiments in composite data sets. This can be either different data sets of the same type (e.g. different SSR markers) or completely different types of data (e.g. sequence data, SSRs and morphological characteristics). The individual sets of data can be weighted depending on their importance if desired.

The identification of unknown isolates can be achieved by comparison to a library created within BioNumerics containing certified taxa. Again, different experiment types and composite data sets can be chosen to achieve the identification. It is possible to link numerous entries for each taxon to an individual library entry and to train neural networks so that any variation within the data can be taken into account when identifications are performed. Detailed reports of the results are produced and can be exported.

The program has various database sharing tools to enable BioNumerics databases to be interrogated via the internet with varying levels of security access which can be controlled by the database host.

[End of document]