186
ANALYSIS OF GENETIC DIVERSITY IN THE ORYZA OFFICINALIS COMPLEX

FEDERICI, M.T. [1]; SHCHERBAN, A. [2].; CAPDEVIELLE, F. [1]; FRANCIS, M [1]; VAUGHAN, D. [3] National Institute of Agricultural Research (INIA), Uruguay [1]; Institute of Cytology and Genetics, Russia [2] ; National Institute of Agrobiological Resources (NIAR), Japan [3]

The *O. officinalis* complex is of particular interest because, in contrast to the single AA genome of all species in the *O.sativa* complex to which the rice cultigens belong, the *O.officinalis* complex has 4 of the 10 genomes so far designated for species in the genus *Oryza*. In addition, the *O. officinalis* complex includes both diploid and allotetraploid species. In the genus *Oryza*, all the other species complexes include either diploid species, such as species in the *O. sativa* complex, or allotetraploid species such as species in the *O. ridleyi* complex. Thus the *O. officinalis* complex not only furnishes broad *Oryza* genomic diversity for study but also may give insights into polyploidy in the genus *Oryza*. Genetic relationships among 34 accessions of wild rice from Asia, Africa, America and Australia were analysed using RFLP technique. After southern blotting, DNA digestion pattern was detected following hybridization with a highly repetitive DNA sequence from a "gypsy" family of mobile elements. This report includes some different accessions than previously reported by Shcherban (2000), representing all *O. Officinalis* species and genomes. Also, another set of enzymes was tested. A dendrogram was constructed from RFLP data, clustering species according to their genome designation (CC, BB, BBCC and CCDD genomes). Some species did not appear in the same group, for example, *O. eichingeri* from Africa and Sri Lanka clustered separately from each other. The same situation was observed for the accessions from China of *O. officinalis*, which clustered together showing a close relationship with *O. rhizomatis*, and *O. eichingeri* (both of CC genome). Also, the tetraploid BBCC from India of *O. officinalis* appears in the same cluster of *O. eichingeri* and *O. punctata* (both from Africa) suggesting close phylogenetic relationship with the African genomes BB, CC and BBCC. The results obtained confirmed previous studies but also gave new insights into phylogeny in the complex and the origin of some taxa.

Keywords: Oryza officinalis, phylogeny, RFLP, gypsy mobile element

188
DATA MINING APPROACHES USING MOLECULAR MARKER INFORMATION FROM GERMPLASM COLLECTIONS OF RICE

CAPDEVIELLE, F. [1]; PINSON, S. [2].; FJELLSTROM, R. [2]; OARD, J. [3] ; National Institute of Agricultural Research (INIA), Uruguay [1]; USDA Beaumont, USA [2] ; Louisiana State University, USA [3]

Data Mining (DM) approaches are applicable towards classification for large amounts of data, sorting objects into two or more labeled classes and deriving functions to optimally assign new object to the labeled classes. This study is part of an ongoing project focused on development and application of DM based on molecular marker information. Classification techniques were used to predict allocation of lines into predefined groups using two different datasets. In one case groups based on germplasm classes (*indica, japonica*-temperate, and *japonica*-tropical) and agronomic traits were used as training samples to evaluate a KNN classification approach using RAPD marker data from a collection of diverse rice accessions. In another case, classification procedures were applied to SSR marker data from a collection of breeding lines, and used to construct and evaluate predictive models for classification into different phenotypic groups. A stepwise discriminant analysis procedure was used in both cases to identify sets of markers that best reveal differences between predefined groups. A KNN algorithm using five to ten selected markers produced 90 % to 98 % correct classification of rice lines into *indica, japonica*-temperate or *japonica*-tropical germplasm classes. Percentages of correct classification into groups with contrasting phenotypes across and within germplasm classes were estimated for different traits. Using five markers, values ranged from 80 % (days to heading, cold tolerance, and seedling vigor) to 90 % (plant height) correct classification into groups defined without consideration of germplasm classes. Potential applications of DM classificatory approaches in germplasm improvement include marker-assisted allocation of accessions and development of classification models using marker information from reference genotypes. Association mapping within structured populations, using a set of unlinked markers to infer details of population structure and estimate ancestry of sampled individuals, is suggested based on the existence of differentiation at the molecular level among germplasm classes (*japonica* temperate, *Japonica* tropical and *indica*).

Keywords: data mining, bioinformatics, germplasm, association, population structure

187
MARKER-ASSISTED CLASSIFICATION OF LEMONT X TEQING RILS INTO DISEASE RESPONSE GROUPS: COMPARISON OF DISCRIMINANT ANALYSIS AND NEURAL NETWORK ALGORITHMS

CAPDEVIELLE, F. [1]; PINSON, S. [2].; OARD, J. [3] National Institute of Agricultural Research (INIA), Uruguay [1]; USDA Beaumont, USA [2] ; Louisiana State University, USA [3]

Modern information technology, based upon the availability of powerful computer-based systems, is providing new tools to collect, transfer, store and combine agronomic and molecular data from breeding lines and germplasm collections. As a consequence, data mining approaches, based on techniques such as clustering, classification and association analysis, could be applied to help researchers discover useful patterns in their data. A discriminant analysis procedure coupling agronomic information and molecular marker profiles of breeding materials has been applied in our study to a population of recombinant inbred lines (RILs) of rice derived from a cross between Lemont and Teqing rice varieties. The main objective is to allocate breeding lines into "targeted" groups for selection using a classificatory approach based on molecular markers as predictors of resistant or susceptible response to blast disease, caused by *Pyricularia grisea* (Cooke) Sacc. (syn. *Magnaporthe grisea*). Two data mining approaches, namely K-Nearest Neighbor Discriminant Analysis (KNN) and Artificial Neural Networks (ANN), were selected for this study to treat the task of classification of rice lines into groups reflecting disease resistance as a learning problem. This comparison was based on training samples from extremes of the underlying distribution of a quantitative trait (Area Under Disease Progress Curve) reflecting the response of breeding lines to the blast pathogen and a set of measurement vectors (based on marker information) and their corresponding group assignment (resistant or susceptible), based on phenotypic information. ANN trained using a back-propagation algorithm to predict the allocation of lines using marker information was better (more than 94 % correct classification) compared with the KNN algorithm (84 % correct classification) in the case of groups with reduced differentiation. This is probably consequence of the learning process (error back-propagation) included in the adjustment of the ANN to recognize specific patterns in the data. In contrast, the KNN procedure relies only in the information provided by similar cases in the training sample used to define the groups.

Keywords: data mining, marker-assisted, classification, blast disease

189
APPLICATION OF IN VIVO EXPRESSION TECHNOLOGY (IVET) FOR THE STUDY OF RICE INFECTION BY THE NITROGEN-FIXING ENDOPHYTIC BACTERIUM PSEUDOMONAS STUTZERI A15

REDIERS, H. [1] ; BONNECARRERE, V. [2] ; VANDERLEYDEN, J. [1] ; DE MOT, R. [1] Centre of Microbial and Plant Genetics, K.U. Leuven, Belgium [1]; National Institute of Agricultural Research (INIA), Uruguay [2]

A bacterium repeatedly isolated from rice in China is *Pseudomonas stutzeri* A15 (Vermeiren *et al.*, 1999). This nitrogen-fixing endophyte may provide the rice plant with fixed nitrogen and hence promote plant growth. At present, the mechanisms which enable strain A15 to colonize and infect rice roots and survive within rice plants, are not known. To identify bacterial promoters specifically induced during the interaction with the host plant, the '*in vivo* expression technology' (IVET) is used (Rainey, 1999; Rainey and Preston, 2000). IVET is a promoter trapping technique based on the complementation of a mutation in an essential biosynthetic gene (Mahan *et al.*, 1993). The prerequisite for applying this technique under these circumstances is that the mutation cannot be complemented by production of plant metabolites. Therefore a *P. stutzeri* A15 *dapB* mutant was constructed. The *dapB* gene encodes a dihydrodipicolinate reductase and is involved in diaminopimelic acid biosynthesis. Diaminopimelic acid is an essential component of peptidoglycan and the precursor of lysine. A genome library was constructed by inserting DNA fragments in front of a promoterless *dapB* gene with the *P. stutzeri* A15 *dapB* mutant as the host strain. The screening of this library for specifically *in vivo* expressed promoters is in progress. Already, a number of genes with specific *in vivo* expression could be identified. These genes seem to be involved in stress response, nutrient acquisition, adaptation to different environments or regulation. In addition to these genes, some genes without significant homology to genes in the database or genes with unknown function were isolated as well.

Keywords: Pseudomonas stutzeri, endophytic, IVET, promoters, nitrogen-fixation