

Software for DNA fingerprinting analyses

In the current scenario, wild and cultivated plant varieties and their genes are being increasingly recognized as resources of high economic value. Consequently there is worldwide policy shift from free exchange and unhindered exploitation to controlled access to these resources. The World Trade Organization (WTO) agreement of which India is signatory, envisages adoption of either a patent system or some form of effective *sui generis* system or by combination thereof, for protecting the intellectual property rights of plant breeders. Therefore, there is an urgent need for a perfect identification system to enforce the protection regimes for plant varieties and germplasm.

Although there are a number of statistical softwares available in the Internet on public domain, they are either difficult to use or expensive. With reference to this, often statistical packages such as SAS (SAS Institute, 1985) or R (R Development Core Team, 2003) require some programming to perform the desired statistical analysis. There is a need to develop cheap, easily accessible and user-friendly specialized software aimed to store and analyse data at common interface.

When the thought of database comes to our mind, we think of two basic important functions. First, storing of large and scattered data and secondly, a retrieval system to show the desired result of a query. Data stored in the software can help in the cases of unauthorized use of varieties and provide insight to the breeders to view differences between heterozygous and homozygous varieties of plant. In addition, uniqueness of the DNA fingerprinting profile table helps to legally protect new varieties of plants, whether they are developed by genetic engineering, tissue culture or traditional methods.

At the National Research Centre on DNA Fingerprinting (NRCDF), New Delhi huge molecular profile data were available on crop varieties and genetic stocks. It was observed that scientists required a simple, powerful and intuitive tool to manipulate data in infinite and unpredictable combination. In addition, there was a need to build a library of crop DNA profiles to store any new DNA sample, including its matches with the

DNA samples already recorded. To fulfil this need a software system entitled 'Crop DNA Fingerprint Database' has been developed and implemented at NRCDF.

The Crop DNA Fingerprint Database has been developed in Visual Basic-6 (VB-6) environment in front-end and MS Access as back-end. VB is an event-driven programming language and associated development environment from Microsoft. VB-6 enables Rapid Application Development of Graphical User Interface (GUI) applications, access to databases using Data Access Objects, Remote Data Objects, or Activex Data Objects, and creation of ActiveX controls and objects. The language not only allows programmers to easily create simple GUI applications, but also has the flexibility to develop fairly complex applications. Programming in VB is a combination of visually arranging components or controls on a form, specifying attributes and actions of those components, and writing additional lines of code for more functionality.

Hardware: An IBM PC or compatible with an Intel Pentium IV processor or higher CPU and 64 or more MB of Random Access Memory. A mathematical coprocessor is required to achieve reasonable computing speed.

Software: Window XP or later version.

The Crop DNA Fingerprint Database is a menu-driven software. Most features are accessed by making selections from the menus. Binary dataset is stored in this database, where 0 represents absence of a band, 1 represents presence of a band and 3 represents missing data.

For easy retrieval of data from the database a powerful search facility has been created with different user-friendly search options. There are four types of search options:

1. Search by crop
2. Search by technique
3. Search by primer
4. Search by variety.

Statistical analysis is one of the important components of the software. Drop-down menu of statistical analysis consists of eleven types of statistical methods. The user can select any of the options and invoke the desired analysis. Statistical modules for Jaccard, Dice and Simple

Matching coefficient analyses indicate whether two varieties are different or similar. They also help to know the extent of similarity between varieties. Comparison may be done on one-to-one or one-to-many varieties. Module for probability of chance identity (I)¹, is used to find the probability of two random primers displaying the same genotype using the following formula. The polymorphism information content (PIC) value is commonly used in genetics as a measure of polymorphism for a marker locus used in linkage analysis². In order to find best informative primer, modules of PIC and average number of bands per cultivar analyses were used. Genetic relationship among different primers was found using gene diversity analysis. The Shannon diversity index (H) is commonly used to characterize species diversity in a community. It accounts for abundance of the present species. To find the overall suitability of a primer for the purpose of identification, the formula by Prevost and Wilkinson for the calculation of resolving power was used. Genotype differentiation between markers was found using the module for marker informativeness. Marker informativeness analysis deals with fraction of polymorphic loci, effective multiplex ratio, diversity index and marker index.

All the statistical modules of this software generate two types of reports: standardized and detailed. The standardized report shows values of the results obtained after calculation, whereas the detailed report shows the steps of calculation and this report is generated in MS Excel worksheet.

Input form for statistical analysis is shown in Figure 1. The name of the analysis flashes at the top of form on execution of the program. This form has provision to select crop, technique, primer and variety. Type of desired report can also be selected from the right side of the form.

Barcode representation is another form of representing the profile table in MS Excel. In this case the presence of the band is shown by bar and absence/missing of band is shown with blank space. To provide working assistantship to the user, the help menu is used to open the standard Microsoft Help Window contain-

Figure 1. Input form for statistical analysis.

ning information on how to use different features of the software package.

The Crop DNA Fingerprint Database is a user-friendly Windows-based computer package for storing and analysing profile data of crop varieties and genetic stocks. The package provides Windows GUI that makes software more accessible for the casual computer user and more convenient for the experienced computer user. Simple menus and dialogue box selections enable users to perform stati-

stical analysis and produce a scientifically sound report, thereby assisting in analysing the profile data using computational tools. This will be an important tool for scientists, researchers, plant breeders and those involved in DNA fingerprinting of crops. It would also provide an interface where the DNA profile can be stored and analysed.

1. Paetkau, D., Calvert, W., Stirling, I. and Strobeck, C., *Mol. Ecol.*, 1995, 4, 347–354.

2. Botstein, D., White, R. L., Skolnick, M. and Davis, R. W., *Am. J. Hum. Genet.*, 1980, 32, 314–331.
3. Bhat, K. V., Molecular Data Analysis. In Anonymous (eds.), NRC on DNA Fingerprinting Training Manual on Techniques for Plant DNA Fingerprinting 19–28 November 2001, NRC on DNA Fingerprinting, National Bureau of Plant Genetic Resources, New Delhi, 2001, pp. 46–58.
4. Bhattacharya, E., Dandin, S. B. and Ranade, S. A., *J. Biosci.*, 2005, 30, 669–677.
5. Kirst, M., Cordeiro, C. M., Rezende, G. D. S. P. and Grattapaglia, D., *J. Hered.*, 96, 161–166.
6. Nagaraju, J., Reddy, K. D., Nagaraja, G. M. and Sethuraman, B. N., *Heredity*, 2001, 86, 588–597.

ACKNOWLEDGEMENTS. I thank my colleagues of NRCDF for help during the development of the Crop DNA Fingerprint Database software.

Received 23 September 2007; accepted 5 October 2007

MADHU BALA

National Research Centre on DNA Fingerprinting,
National Bureau of Plant Genetic Resources,
New Delhi 110 012, India
e-mail: madhu74_nbpgpr@yahoo.com

Population status of *Gymnocladus assamicus*, a critically endangered tree species in Arunachal Pradesh

Arunachal Pradesh, northeastern state of India and part of the Himalayan hot-spot, harbours a large number of rare, endangered and endemic plant species due to its vivid climatic conditions. *Gymnocladus assamicus* is one such tree species of ecological and economic importance. It is categorized as critically endangered for its extremely small population and given conservation priority for effective management¹. The species is also listed for national recovery programme in India². It was first described and published in 1934 from Khasi Hills, Meghalaya. However, further information regarding

its demography and niche requirements are unknown. The present correspondence describes the existing population of *G. assamicus* in Arunachal Pradesh along with other autecological details.

Gymnocladus is a small primitive genus (Fabaceae: Caesalpinoideae) with only five species of which three species, namely *G. chinensis* Baill., *G. assamicus* Kanjilal ex. P.C. Kanjilal and *G. burmanicus* Parkinson are confined to the trijunction of India, China and Myanmar (Burma). While *G. diocus* (L.) K. Koch is native to the midwest of North America³, *G. assamicus* and *G. chinensis* are

known only from the northeastern states of India⁴. A phylogenetic study of eleven *Gleditsia* species along with two species of its sister genus *Gymnocladus*, showed that the two genera appear to have originated in eastern Asia during the Eocene⁵.

Extensive field visits were conducted during 2004–07 throughout Arunachal Pradesh to document the populations of *G. assamicus*. The populations were geographically marked with global positioning system (GPS) and the number of mature trees, saplings and seedlings was counted. Flower from each of the mature trees was collected and dissected care-