SHORT COMMUNICATION

Bandsize-Binary: A Program to Convert Size Data to Binary Data

Madhu Bala Priyadarshi
NRC on DNA Fingerprinting, National Bureau of Plant Genetic Resources, New Delhi-110012

Bandsize-Binary is windows based computer program, developed in Visual Basic 6 environment for converting electrophoresis gel size data to binary data. It is a user friendly program provided with several features to perform accurate analysis and display of inputed molecular biology data. On the click of *submit* button programs runs to convert microsatellite data to binary data. In the end, it flashes the message of successful data conversion on the right hand side of the form. In addition, the program also provides facility to calculate heterozygosity of markers in the data sheet. Heterozygosity (*H*) is a widely used measure of the allelic diversity or informativeness of a genetic marker. The informativeness of a genetic marker increases as *H* increases. This program provides graphical user interface that makes it more accessible for casual computer users and more convenient for the experienced computer user. Bandsize-Binary program is a very useful tool for researchers and persons involved in DNA fingerprinting.

Key Words: Binary data, DNA fingerprinting, Electrophoresis gel size data, Molecular marker

Molecular markers have become essential tools for population genetics, conservation biology and evolutionary studies as well as for molecular mapping and gene tagging (Jarne and Lagoda, 1996). They have proven to be powerful tools in the assessment of genetic variation and in the elucidation of genetic relationships within and among species. Several molecular markers, *viz.*, RFLP (Becker *et al.,* 1995), RAPD (Tingey and Deltufo, 1993), SSRs (Levinson and Gutman, 1987), ISSRs (Albani and Wilkinson, 1998), AFLP (Mackill *et al.*, 1996) and SNPs (Vieux *et al*., 2002) are presently available to assess the variability and diversity at molecular level (Joshi *et al.,* 2000). Information regarding genetic variability at molecular level could be used to help, identify and develop genetically unique germplasm that compliments existing cultivars.

There are several bioinformatics applications for analysis of molecular biology data in public domain on World Wide Web (www). The sequence analysis software tools such as FASTA, BLAST, CLUSTALW and other sequences analysis tools became very popular. However, there are not many other easy to use tools available for conversion of molecular biology lab data (such as DNA Polymorphic band data) to binary format. These data analyses softwares are common in the science, as the scientists in need of new analyses develop algorithms, and then crystallize the algorithms as software. Most of the basic biosequence analyses are developed by scientist such as FASTA, BLAST etc. In this paper

an effort has been made to describe a program entitled 'Bandsize-Binary'. This program has been developed to provide a facility of data conversion from band size to binary format and also to know informativeness of markers by heterozygosity analysis.

Bandsize-Binary is a windows based computer program. It is developed in Visual Basic 6 environment. Visual Basic is an event driven programming language. It enables Rapid Application Development (RAD) of Graphical User Interface (GUI) applications. It helps to access databases using Data Access Objects (DAO), Remote Data Objects (RDO), or Activex Data Objects (ADO). The language not only allows programmers to easily create simple GUI applications, but also provides flexibility to develop fairly complex applications. On hardware site it requires an IBM PC or compatible with an Intel Pentium IV processor or heigher CPU and 64 or more MB of Random Access Memory (RAM). A mathematical coprocessor is required to achieve a reasonable computing speed. Window XP or later version of Windows are suitable platforms for running the program.

In case of co-dominant markers (such as SSRs-microsatellites/ minisatellites or RFLPs), alleles are coded simply as their integer size in base pair. In addition to that, Genscan a gene prediction tool (Reference no. 12) inputs automatically genome sequence data and outputs allelic band size data. Researchers try to put the data into data analyses softwares and perform desired

analysis to interpret data. In general, most of the softwares for molecular data analysis require data into binary format such as NTSYS PC. Band size-Binary program has been developed to convert huge bandsize data to binary format. The genotype data is recorded in terms of a two-way matrix. The program reads all the band sizes with their locus and markers from an excel worksheet as input file. The input file of Bandsize–Binary program is identical to the Fig. 1, showing sample of genotype data in Excel worksheet. Varieties or cultivars are entered in the first row of Excel worksheet and primers are entered into the rows. The input form, as shown in Fig. 2, prompts to get name of crop and Excel file. On click of *submit* button programs converts size data to binary data. In the end, it flashes the message of successful data conversion on the right hand side of the form. In addition to this, it also informs about number of primers, varieties and records of data sheet. The output file in Excel is saved into the C:\ directory of the computer system. In binary data, presence of band is represented by 1 and absence of band is represented by 0. The format of binary data includes columns as *Sno, Bandsize, Primer* and *names of varieties*. The values of varieties corresponding to their primer in microsatellite genotype data is represented into *Bandsize* column. Fig. 3a shows binary format of sample data in Excel worksheet.

In addition, the program also provides facility to calculate heterozygosity of markers in the data sheet. Heterozygosity is one of the analyses to measure informativeness. The Informativeness of genetic marker is measured by the number of alleles and their frequencies. Heterozygosity (*H*) is a widely used measure of the allelic diversity or informativeness of a genetic marker. The informativeness of a genetic marker increases as

Formula for Heterozygosity

$$P_i = \frac{1}{N} \sum_{i=1}^{n} f_i \qquad h = 1 - \sum_{i=1}^{n} P_i^2$$

f_i = Presence of specific band
n = Number of plants in the population
N = Total number of locus i bands in a population
P_i = Allele frequencies
h = Heterozygosity (Nei 1987)

H increases.

Fig. 3b shows calculated heterozygosity report of all the primers in Excel worksheet. At the end of processing, the user gets two output files in Excel format stored in C: drive of computer. Of the two output files, the first file is binary format of size data and second file is the heterozygosity report file. This program can be downloaded from following address: http://www.nbpgr.ernet.in/NRCWEB08/BandsizeBinary.htm. Click on the hyperlink will allow the user to download the setup file for Bandsize-Binary program. This setup

cultivar	Var1	Var2	Var3	Var4	Var5
Primer1	152	155	160	162	168
Primer1	152	155	152	152	152
Primer2	238	238	243	240	237
Primer2	241	238	243	243	243
Primer3	101	106	106	111	103
Primer3	101	106	110	111	111

Fig. 1: Sample of genotype data in Excel worksheet
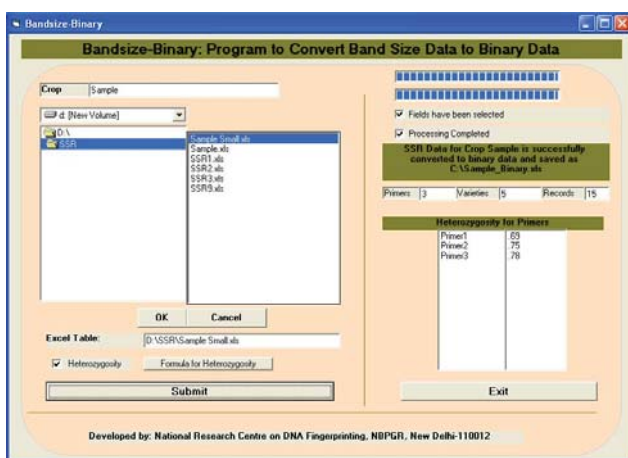
Fig. 2: Form to convert band size data to binary data

Bandsize	Primer	Var1	Var2	Var3	Var4	Var5
152	Primer1	1	0	1	1	1
155	Primer1	0	1	0	0	0
160	Primer1	0	0	1	0	0
162	Primer1	0	0	0	1	0
168	Primer1	0	0	0	0	1
237	Primer2	0	0	0	0	1
238	Primer2	1	1	0	0	0
240	Primer2	0	0	0	1	0
241	Primer2	1	0	0	0	0
243	Primer2	0	0	1	1	1
101	Primer3	1	0	0	0	0
103	Primer3	0	0	0	0	1
106	Primer3	0	1	1	0	0
110	Primer3	0	0	1	0	0
111	Primer3	0	0	0	1	1

Fig. 3a: Binary format of sample data in Excel worksheet

Fig. 3b: Heterozygosity results generated in Excel worksheet

file has to be installed on computer to run. This is a free of cost program and author can also be contacted for email version.

Band size-Binary Program has been registered at The Registrar of Copyrights, Copyright Office, New Delhi-110001.

This program provides graphical user interface that makes it more accessible for casual computer users and more convenient for the experienced computer user. A wider look of similar type of program in internet took to the conclusion that although there are number of statistical software available in internet on public domain, unfortunately, they are either difficult to use or very expensive. Henceforth, Band size-Binary program will be very useful tool for researchers, and persons involved in DNA fingerprinting.

The author declares that she has no competing interests.

References

Albani MC and MJ Wilkinson (1998) Inter simple sequence repeat polymerase chain reaction for the detection of somaclonal variation. *Plant Breed.* **117:** 573-575.

Becker J, P Vos, M Kuiper, F Salamini and M Heun (1995) Combined mapping of AFLP and RFLP markers in barley. *Mol. Gen. Genet.* **249:** 65-73.

Jarne P and PJL Lagoda (1996) Microsatellites from molecules to populations and back. *Trends Ecol. Evol.* **11:** 424-429.

Joshi SP, VS Gupta, RK Aggarwal, PK Ranjekar and DS Brar (2000) Genetic diversity and phylogenetic relationship as revealed by Inter simple sequence repeat polymorphism in the genus *Oryza*. *Theor. Appl. Genet.* **100:** 1311-1320.

Levinson G and GA Gutman (1987) Slipped-strand mis-pairing: a major mechanism of DNA sequence evolution. *Mol. Biol. Evol.* **4:** 203-221.

Mackill DJ, Z Zhang, ED Redona and PM Colowit (1996) Level of polymorphism and genetic mapping of AFLP markers in rice. *Genome* **39:** 969-977.

Tingey SV and JP Deltufo (1993) Genetic analysis with Random Amplified Polymorphic DNA. *Plant Physiol.* **101:** 349-352

Vieux EF, PY Kwork and RD Miller (2002) Primer design for PCR and sequencing in high-throughput analysis of SNPs. *Biotechniques* **32:** 28-30.