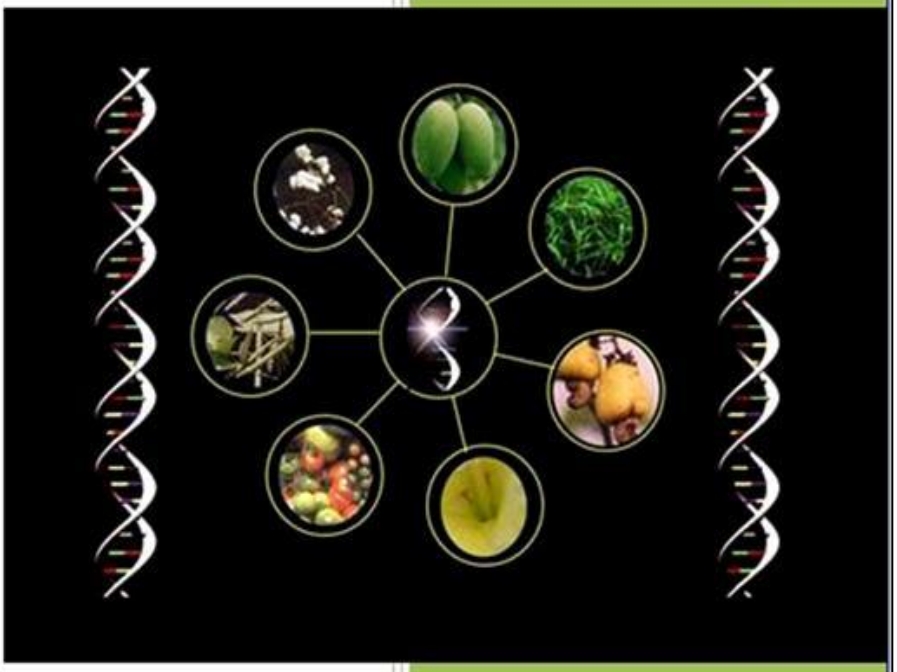
User Manual

Molecular Binary Data Analyses Software



NBPGR

Madhu Bala Priyadarshi
Scientist (Sr. Scale)

National Bureau of Plant Genetic Resources
New Delhi-110012

Table of Contents

Molecular Binary Data Analyses Software

Introduction

"Molecular Binary Data Analyses Software" is developed using Visual basic 2008 integrated development environment at front end and SQL server at back end. It is interactive software that stores and retrieves information according to the choice of user and performs data analysis. Visual Basic 2008 is one of the latest versions of Visual Basic launched by Microsoft in 2008. It is a full fledged Object-Oriented Programming (OOP) Language. VB2008 enables Rapid Application Development (RAD) of graphical user interface (GUI) applications, access to databases using Data Access Object (DAO), Remote Data Objects (RDO), or Active X Data Objects (ADO), 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 as well. 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. "Molecular Binary Data Analyses Software" is designed to store and analyze profile tables of crops fingerprinted. Software is dedicated to store all necessary information regarding varieties and primers in profile tables. In addition to that, it performs some of the important statistical analyses. Module for Jaccard's, dice and simple matching coefficient analysis of the software helps to know similarity between two varieties. Facility is made to do comparison between varieties by one to one or one to many, within table or across two tables. In order to find best informative primer, modules of polymorphic information content and average number of bands per cultivar analyses is used. Genetic relationships among different primers are found by using gene diversity and resolving power analyses. Module of barcode generation develops band map for all primers in a particular profile table. The Help module is developed to provide working assistance to users. Facility has been developed to upload data directly from MS Excel worksheet to database. In addition to that profile table of already stored data in database can be saved at desired location to do any modification if required. Different types of reports were developed for different types of analyses.

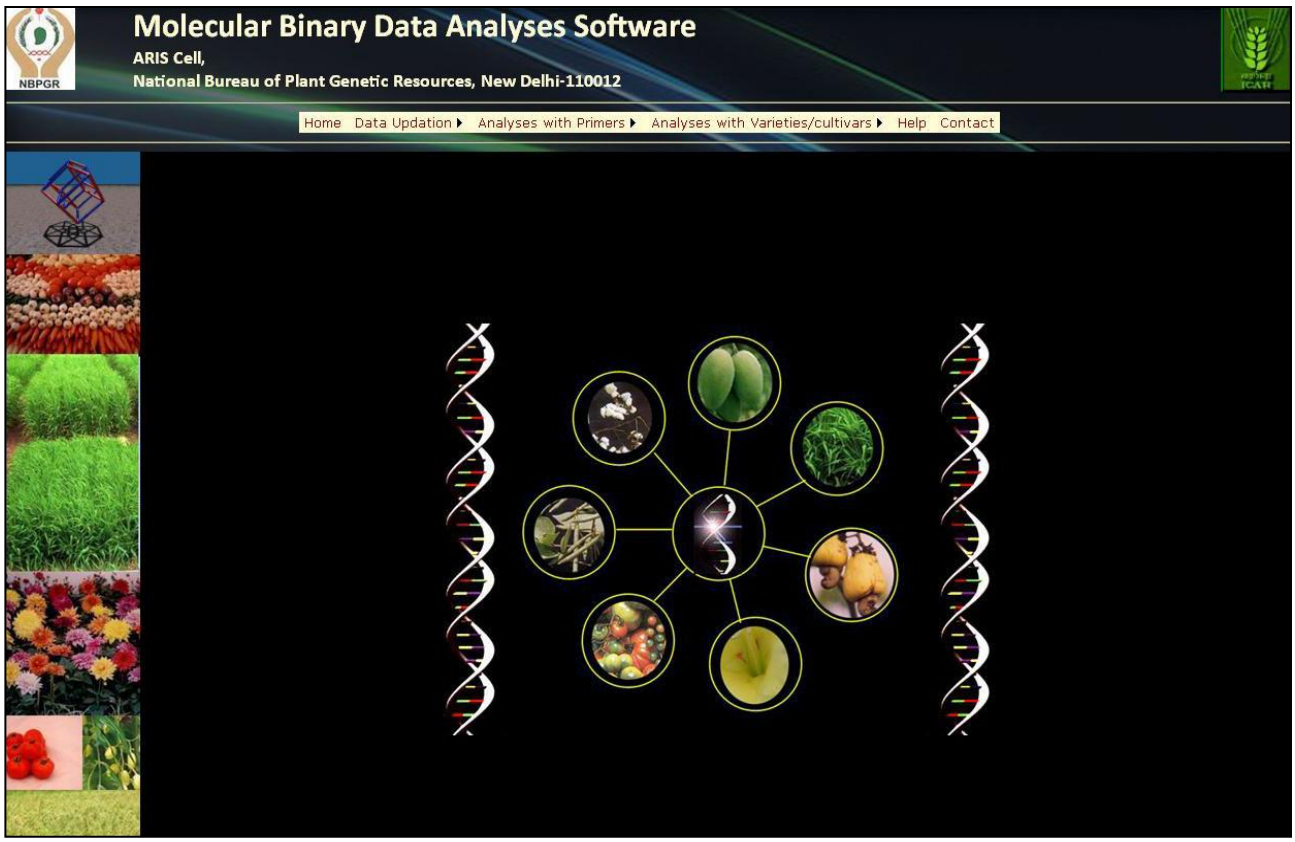


Figure1. Startup screen of software

Major components of software

1. Uploading binary profile data
2. Performing statistical analyses

Hardware and software requirement

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

Software: Window XP or later version

Menu bar of Software

Table1. Menu Bar

Home	Data Updation	Analyses on Primers	Analysis on Varieties/Cultivars	Help	Contact
	Insert New Profile Data View and Edit Data Delete Data View Stored Data Profile Data in Barcode Format	Average number of bands per cultivar Polymorphic Information Content (PIC) Gene Diversity Resolving Power Shannon Diversity Index Marker Informativeness Polymorphic Loci Observed Heterozygosity Expected Heterozygosity Frequency of Occurrence of marker Probability of chance identity	Coefficients Calculation		

2. Data Management

Data Updation

Data Updation on the menu bar deals with adding, deleting, editing and viewing data. Click on menu option of Data Updation leads to following submenus.

1. Insert New Profile Data
2. View & Edit Data
3. Delete Data
4. View Stored Data
5. Profile Data in Barcode Format

2.1 Insert New Profile Data

Format of profile data

Sno	Bandsize	Primer	Variety1	Variety2	Variety3	Variety4	Variety5	Variety6	Variety7	Variety8	Variety9	Variety10	Variety11	Variety12	Variety13
1	0	P1	1	1	1	1	1	1	1	1	1	0	1	0	1
2	0	P1	0	1	0	0	0	1	1	1	0	1	1	0	0
3	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	1
4	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	1
5	0	P1	1	0	0	0	0	0	0	0	0	0	0	0	0
6	0	P1	1	1	1	1	0	1	1	1	1	1	1	0	1
7	0	P1	1	0	1	1	0	0	0	0	1	1	1	0	1
8	0	P1	1	1	1	1	1	1	1	1	1	1	1	0	1
9	0	P1	1	1	1	0	1	1	1	0	1	1	1	0	0
10	0	P1	1	1	1	1	0	1	1	1	1	1	1	0	1
11	0	P1	0	1	0	0	0	1	0	0	0	1	0	0	0
12	0	P1	1	1	1	1	1	0	1	1	1	1	1	0	1
13	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	1
14	0	P1	1	1	1	1	1	1	1	1	1	1	1	0	1
15	0	P1	1	1	1	1	0	1	1	1	1	1	1	0	1
16	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	0
17	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	1
18	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	0
19	0	P2	1	1	1	1	0	1	1	1	1	1	1	0	1
20	0	P2	1	1	0	0	0	1	1	1	1	1	1	0	0
21	0	P2	1	1	1	0	0	1	1	1	1	1	1	0	1
22	0	P2	1	1	0	0	0	1	1	1	1	1	1	0	1
23	0	P2	0	1	0	0	0	1	1	1	0	1	1	0	0
24	0	P2	1	1	1	0	0	1	1	1	1	1	1	0	1
25	0	P2	1	1	0	0	0	1	1	1	1	1	1	0	0
26	0	P2	0	1	0	0	0	0	0	1	0	1	1	0	0
27	0	P2	0	1	0	0	0	1	1	0	0	1	1	0	0
28	0	P2	0	1	0	0	0	1	1	1	0	1	1	0	0
29	0	P2	0	1	0	0	0	1	1	1	1	1	1	0	0
30	0	P3	1	1	1	1	1	1	1	1	1	1	1	0	1
31	0	P3	1	1	1	1	0	1	1	1	1	1	1	0	1
32	0	P3	1	1	1	0	0	1	1	1	1	1	1	0	0
33	0	P3	1	1	1	1	0	1	1	1	1	1	1	0	1
34	0	P3	1	1	1	0	0	1	1	0	1	0	1	0	0
35	0	P3	1	1	1	0	0	1	1	1	1	1	1	0	0
36	0	P3	1	1	1	0	0	1	1	1	1	1	1	0	0
37	0	P3	1	1	0	0	0	1	1	1	1	1	1	0	1
38	0	P3	1	0	0	0	0	0	0	0	0	0	0	0	0

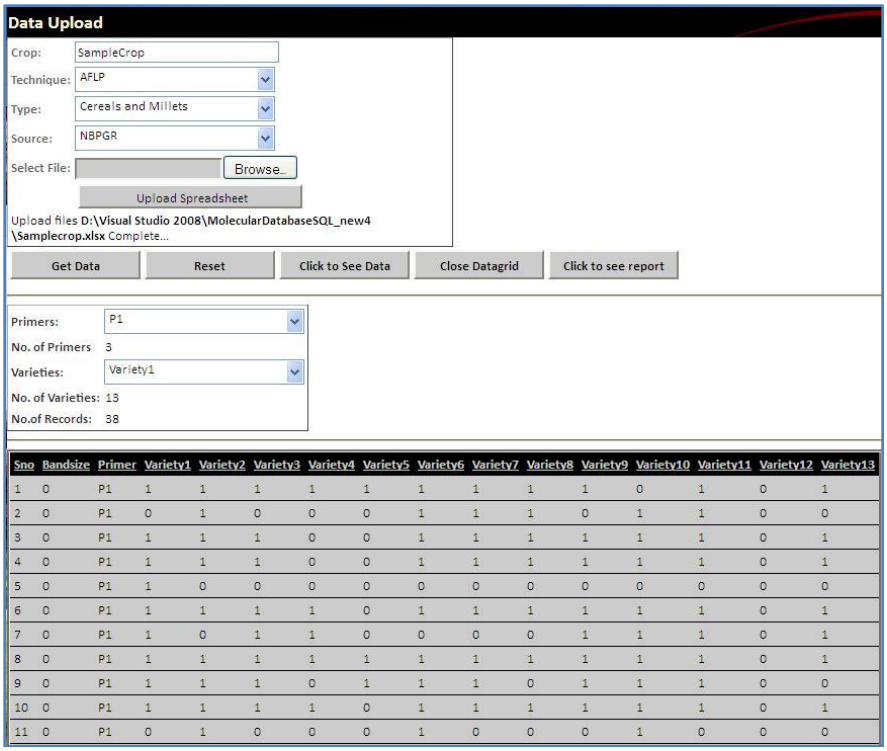
Figure2. Sample profile data

- The profile data is diploid data set and is saved in MS Excel worksheet.

- The first two columns are fixed for serial number and bandsize. Third column is fixed for name of primers, and from fourth column name of varieties/ cultivars are stored as shown in figure 2.

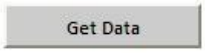
- As shown in figure 2, Variety1, Variety2 … Variety13 are names of varieties/cultivars of profile data. Primers are stored in third column. Figure 2 shows data for two primers, P1 and P2.

- Presence of band is represented by value 1. Absence of band is shown by value 0 and missing bands are represented by value 3.



Sno	Bandsize	Primer	Variety1	Variety2	Variety3	Variety4	Variety5	Variety6	Variety7	Variety8	Variety9	Variety10	Variety11	Variety12	Variety13
1	0	P1	1	1	1	1	1	1	1	1	1	0	1	0	1
2	0	P1	0	1	0	0	0	1	1	1	0	1	1	0	0
3	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	1
4	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	1
5	0	P1	1	0	0	0	0	0	0	0	0	0	0	0	0
6	0	P1	1	1	1	1	0	1	1	1	1	1	1	0	1
7	0	P1	1	0	1	1	0	0	0	0	1	1	1	0	1
8	0	P1	1	1	1	1	1	1	1	1	1	1	1	0	1
9	0	P1	1	1	1	0	1	1	1	0	1	1	1	0	0
10	0	P1	1	1	1	1	0	1	1	1	1	1	1	0	1
11	0	P1	0	1	0	0	0	1	0	0	0	1	0	0	0

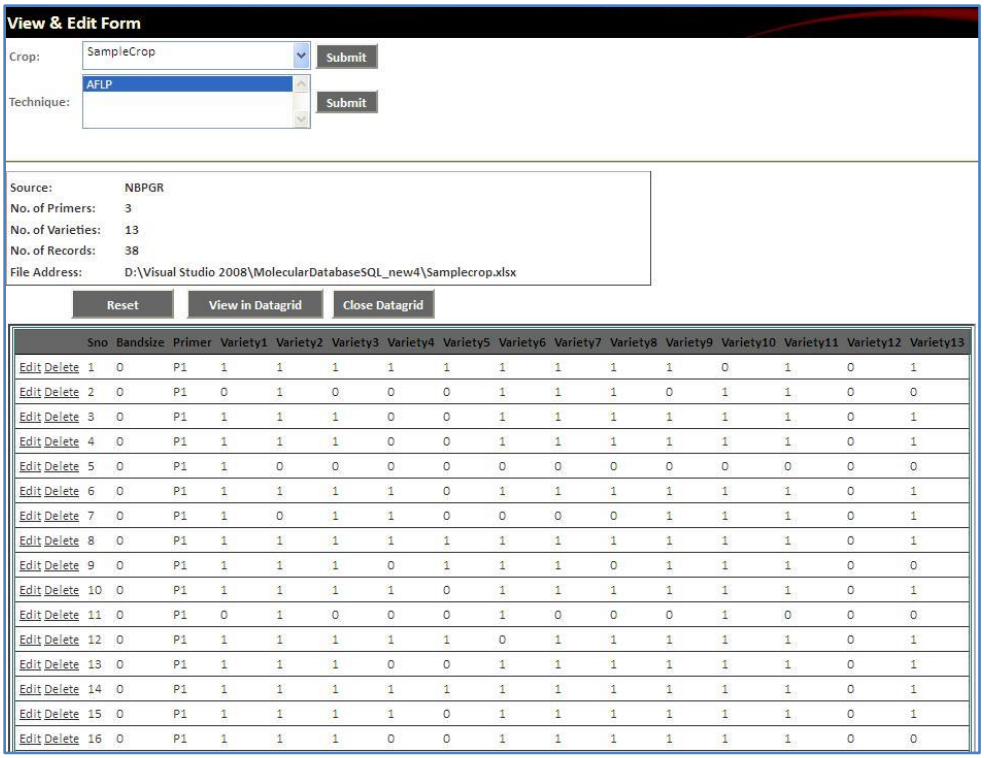
Figure3. Web form for data upload

Web form for data upload has provision to enter name, select technique, type and source for profile data. Browse button is there to select data to be uploaded. User can fill the necessary information and click on 

button to upload data. As soon as data is uploaded, web form shows list of variety/cultivars names in variety combo box, names of primers in primer combo box, total number of primers, varieties and records. Click on "Click to see Data" button, user can see uploaded data in data grid. Figure 3 shows web form for data upload.

2.2 View & Edit Data

View & Edit Data submenu is for viewing and editing stored data. Figure 4 shows web form for view & edit data. In this web form user has option to select crop and its technique and web form displays the respective information of the data stored. Whole of the profile data is visible in data grid with edit/delete function at each row, to do any modification if needed.

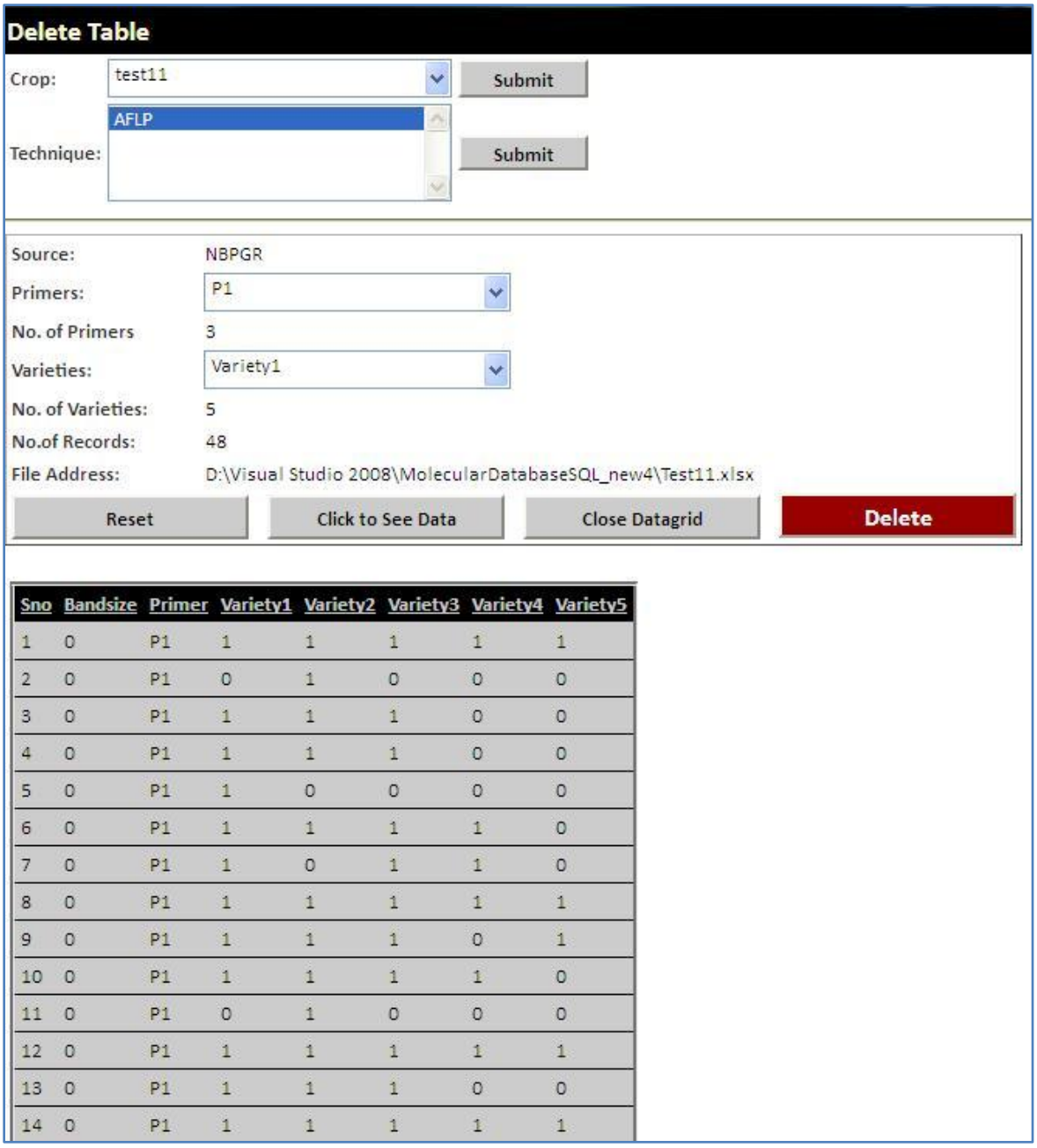


	Sno	Bandsize	Primer	Variety1	Variety2	Variety3	Variety4	Variety5	Variety6	Variety7	Variety8	Variety9	Variety10	Variety11	Variety12	Variety13
Edit Delete	1	0	P1	1	1	1	1	1	1	1	1	1	0	1	0	1
Edit Delete	2	0	P1	0	1	0	0	0	1	1	1	0	1	1	0	0
Edit Delete	3	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	1
Edit Delete	4	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	1
Edit Delete	5	0	P1	1	0	0	0	0	0	0	0	0	0	0	0	0
Edit Delete	6	0	P1	1	1	1	1	0	1	1	1	1	1	1	0	1
Edit Delete	7	0	P1	1	0	1	1	0	0	0	0	1	1	1	0	1
Edit Delete	8	0	P1	1	1	1	1	1	1	1	1	1	1	1	0	1
Edit Delete	9	0	P1	1	1	1	0	1	1	1	0	1	1	1	0	0
Edit Delete	10	0	P1	1	1	1	1	0	1	1	1	1	1	1	0	1
Edit Delete	11	0	P1	0	1	0	0	0	1	0	0	0	1	0	0	0
Edit Delete	12	0	P1	1	1	1	1	1	0	1	1	1	1	1	0	1
Edit Delete	13	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	1
Edit Delete	14	0	P1	1	1	1	1	1	1	1	1	1	1	1	0	1
Edit Delete	15	0	P1	1	1	1	1	0	1	1	1	1	1	1	0	1
Edit Delete	16	0	P1	1	1	1	0	0	1	1	1	1	1	1	0	0

Figure4. Web form for view & edit data

2.3 Delete Data

Web form for delete data is made to delete stored profile data from database. Figure 5 shows web form for deletion of data.



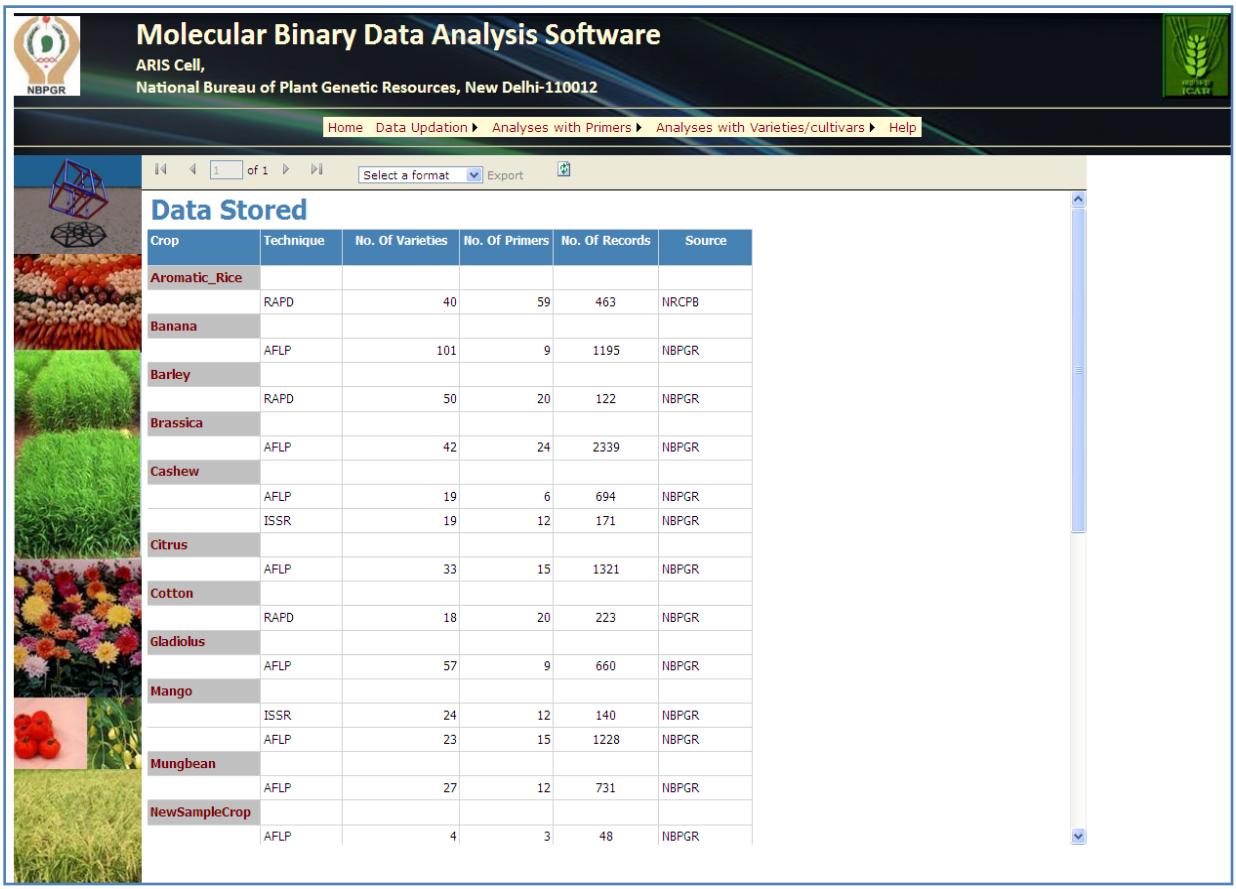
Sno	Bandsize	Primer	Variety1	Variety2	Variety3	Variety4	Variety5
1	0	P1	1	1	1	1	1
2	0	P1	0	1	0	0	0
3	0	P1	1	1	1	0	0
4	0	P1	1	1	1	0	0
5	0	P1	1	0	0	0	0
6	0	P1	1	1	1	1	0
7	0	P1	1	0	1	1	0
8	0	P1	1	1	1	1	1
9	0	P1	1	1	1	0	1
10	0	P1	1	1	1	1	0
11	0	P1	0	1	0	0	0
12	0	P1	1	1	1	1	1
13	0	P1	1	1	1	0	0
14	0	P1	1	1	1	1	1

Figure5. Web form for deletion of data

In this web form user has facility to select crop and its technique from combo boxes. Upon selection of desired crop, form shows the primary details of the selected table as source, names and number of primers, names and number of varieties/cultivars, number of records and file address. Click on "Click to see Data" button shows profile table in data grid. Click on "Delete" button deletes profile table from database.

2.4 View Stored Data

This option in the menu bar is to get report on data stored in database. Figure 6 shows report of data stored in database.



Crop	Technique	No. Of Varieties	No. Of Primers	No. Of Records	Source
Aromatic_Rice					
	RAPD	40	59	463	NRCPB
Banana					
	AFLP	101	9	1195	NBPGR
Barley					
	RAPD	50	20	122	NBPGR
Brassica					
	AFLP	42	24	2339	NBPGR
Cashew					
	AFLP	19	6	694	NBPGR
	ISSR	19	12	171	NBPGR
Citrus					
	AFLP	33	15	1321	NBPGR
Cotton					
	RAPD	18	20	223	NBPGR
Gladiolus					
	AFLP	57	9	660	NBPGR
Mango					
	ISSR	24	12	140	NBPGR
	AFLP	23	15	1228	NBPGR
Mungbean					
	AFLP	27	12	731	NBPGR
NewSampleCrop					
	AFLP	4	3	48	NBPGR

Figure6. Report of data stored in database

2.5 Profile Table in Barcode Format

This facility in this software is developed to view profile table in barcode format. In this format presence of band is represented by a bar and absence or missing band is represented by blank. Figure7 shows profile table in barcode format in datagrid of “samplecrop”.

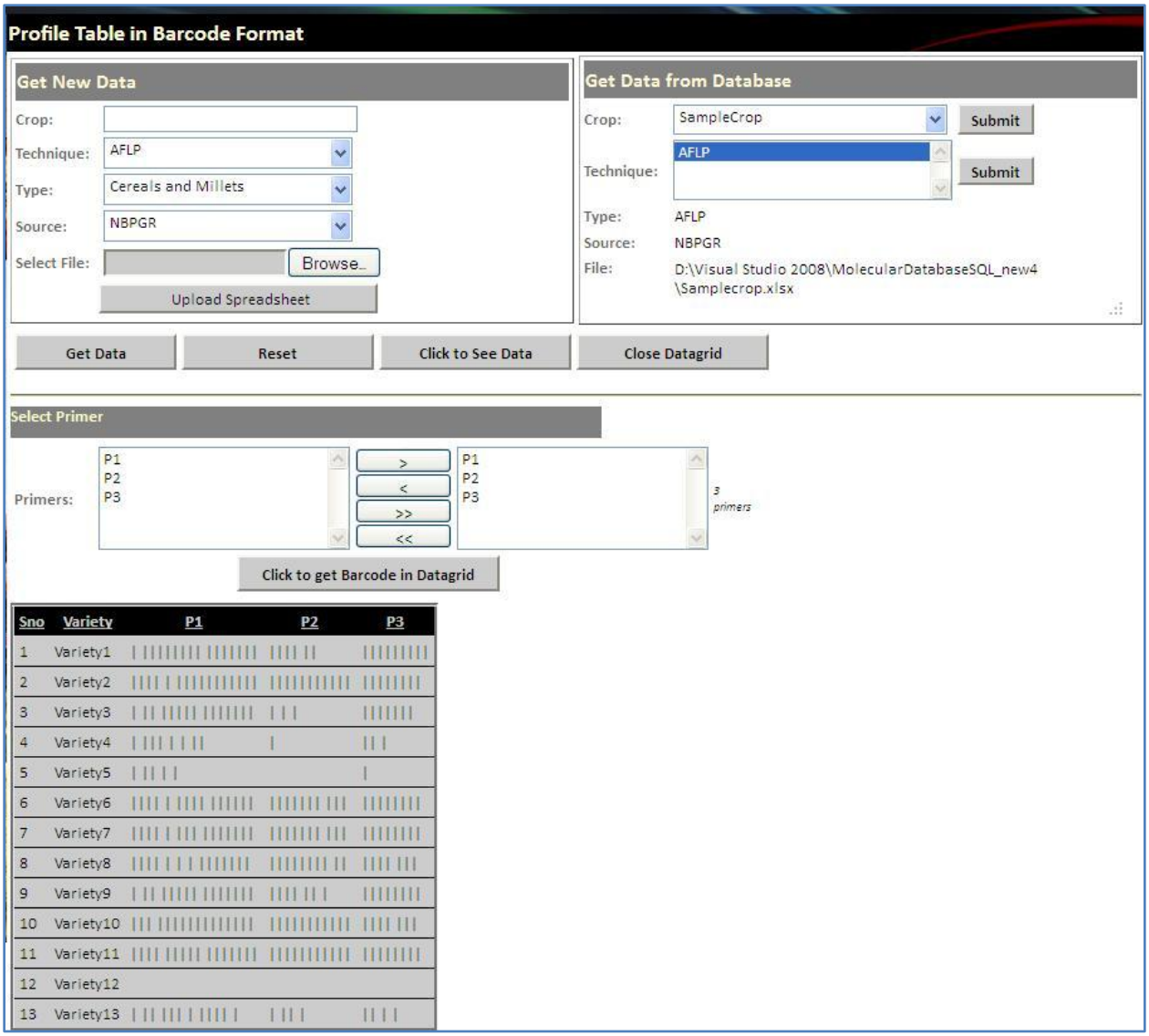


Figure7. Barcode format of profile table of “SampleCrop” data

3. Statistical Analyses

Statistical Analyses

Statistical analysis is one of the important components of software. In this software there are two types of analyses: (1) Analyses with Primers (2) Analyses with Varieties/Cultivars.

In the first case user makes choice for primers and in second case user makes choice for primer and varieties/cultivars.

3.1 Analyses with Primers

Analyses with Primer option on menu bar deals with eleven types of statistical analyses with primers. Figure 8 shows web form for analyses with primers. Following are eleven types of statistical modules:

1. Average number of bands per cultivar
2. Polymorphic information content
3. Gene diversity
4. Resolving power
5. Shannon diversity index

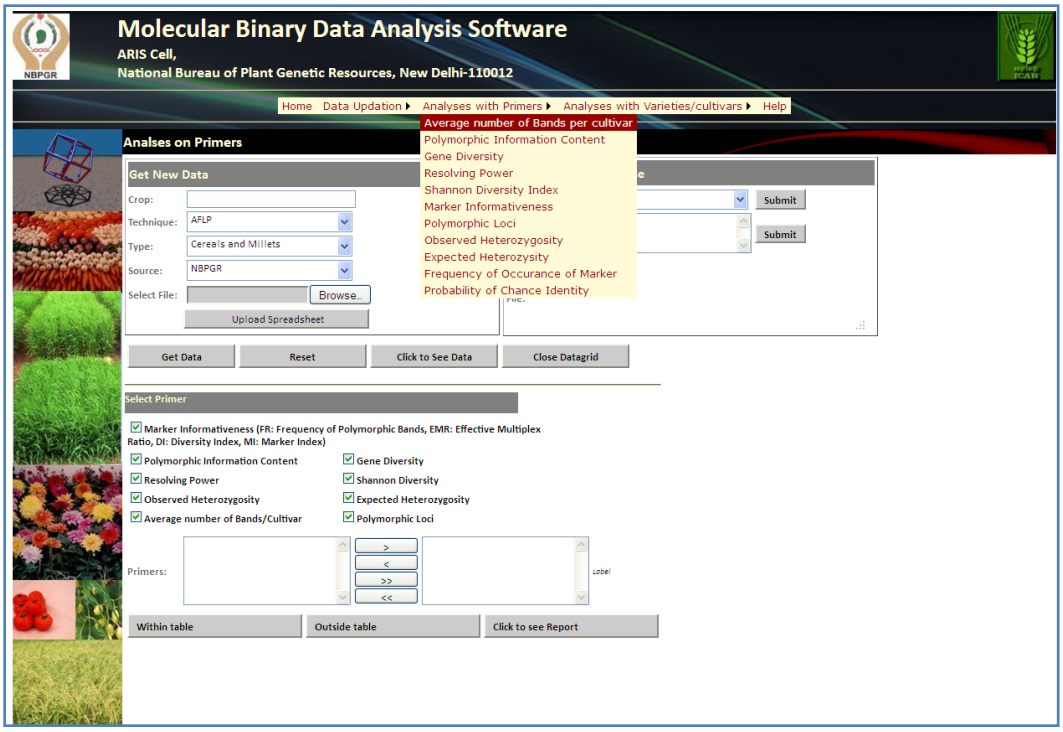


Figure8. Web form for statistical analyses with primers

6. Marker informativeness
7. Polymorphic loci
8. Observed heterozygosity
9. Expected heterozygosity
10. Frequency of occurrence of marker
11. Probability of chance identity

There are two options to get data for analyses.

- Get New Data
- Get Data from Database

Get New Data option is used, when user wants to do analyses on some new data, which is not stored in database. Whereas, second option is used to do analyses on data stored in database.

When user chooses to select first option, i.e. Get New Data, then user provides basic information such as name of crop, technique, type and source. Also, selects desired file for analyses by clicking browse button, hence click Upload spreadsheet button. As soon as file is uploaded message appears as shown in figure 9.

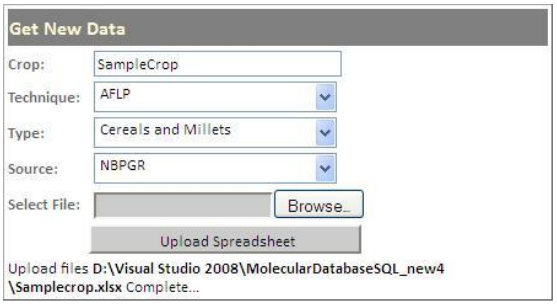


Figure9. Get New Data

When user chooses to select second option i.e. Get Data from Database, then user selects file from drop down box in the second option, as shown in figure 10.

Figure10. Get Data from Database

Upon selecting file by either of ways, user click 

button to get data for analyses. This invokes program to select data and show name of primers into list box of primers and inform about total number of primers in data.

Also, here user can see list of nine statistical modules with choice option as true. User can mark selected statistical analyses module to analyze as shown in figure 11.

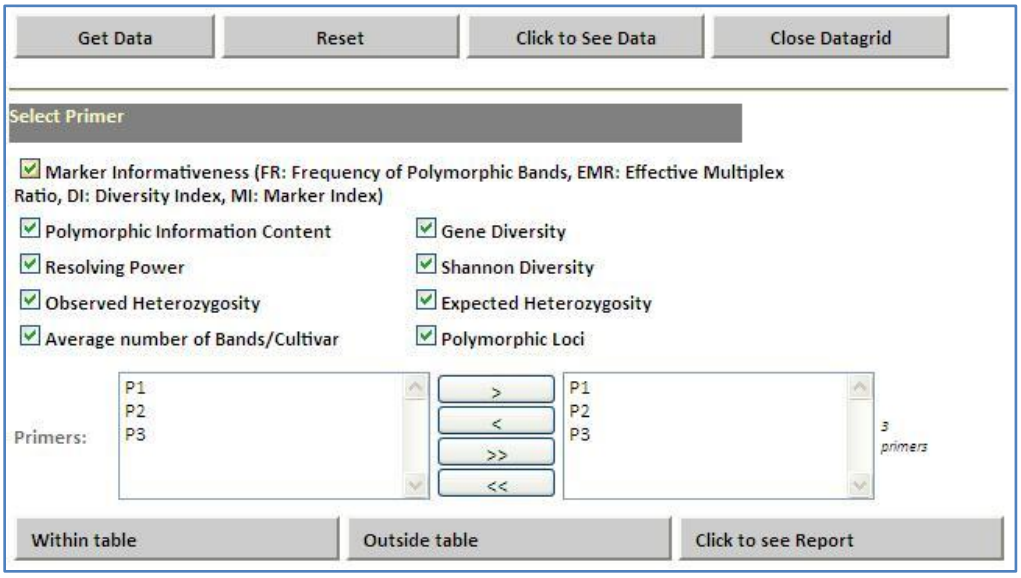


Figure11. Choice to do analyses

Depending upon the choice on statistical module further analysis is being done. After making choice on statistical modules, User has to decide whether the analyses is to be done within table or with other profile table. If option of within table is clicked, then analyses starts and values are calculated from the same table. Figure 12 shows report of analyses being done with selected primers and modules, within table. If user decides to do calculation in addition with some new profile table then, Click on “Outside table” button is done. This will invoke program to select desired file as shown in figure 12.

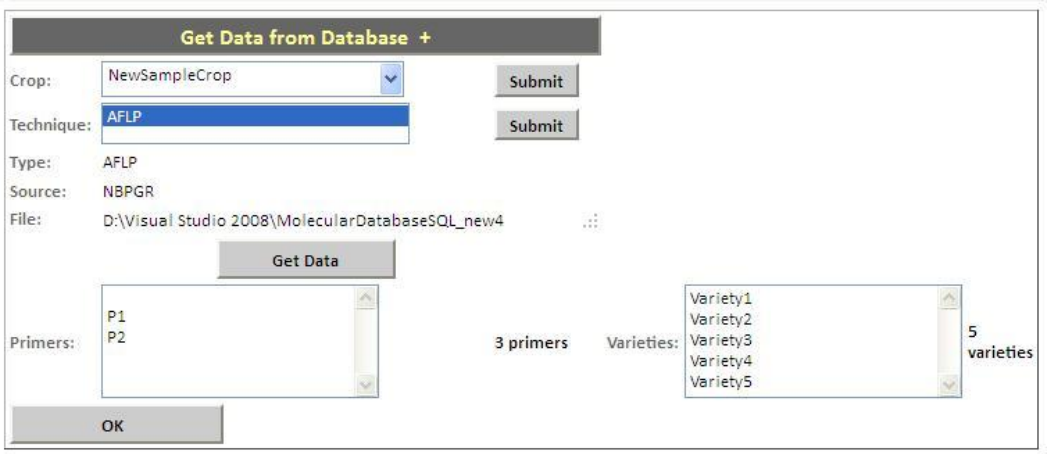
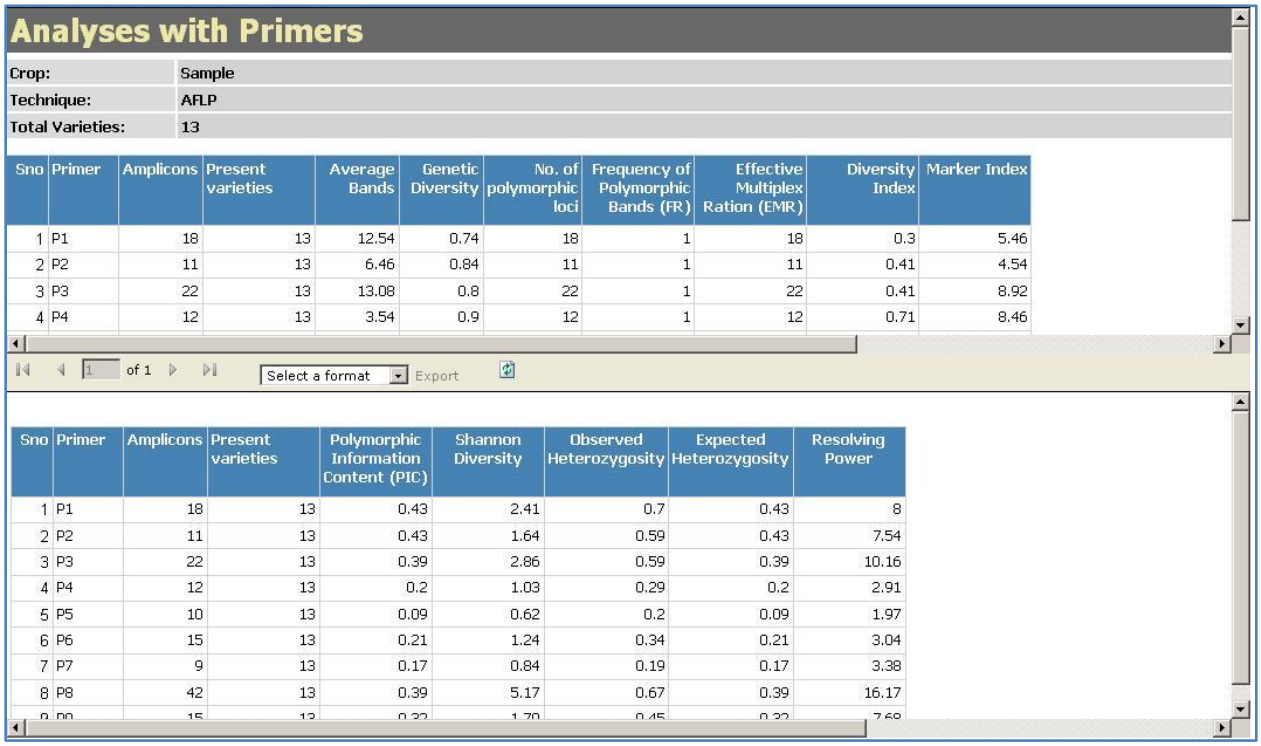


Figure12. Getting data from another profile table for calculation

Depending upon the choice of user analyses are being done. Figure 13 shows result of analyses done on profile table "SampleCrop" and figure 14 shows result of analyses being done on "SampleCrop" and "NewSampleCrop".



Analyses with Primers

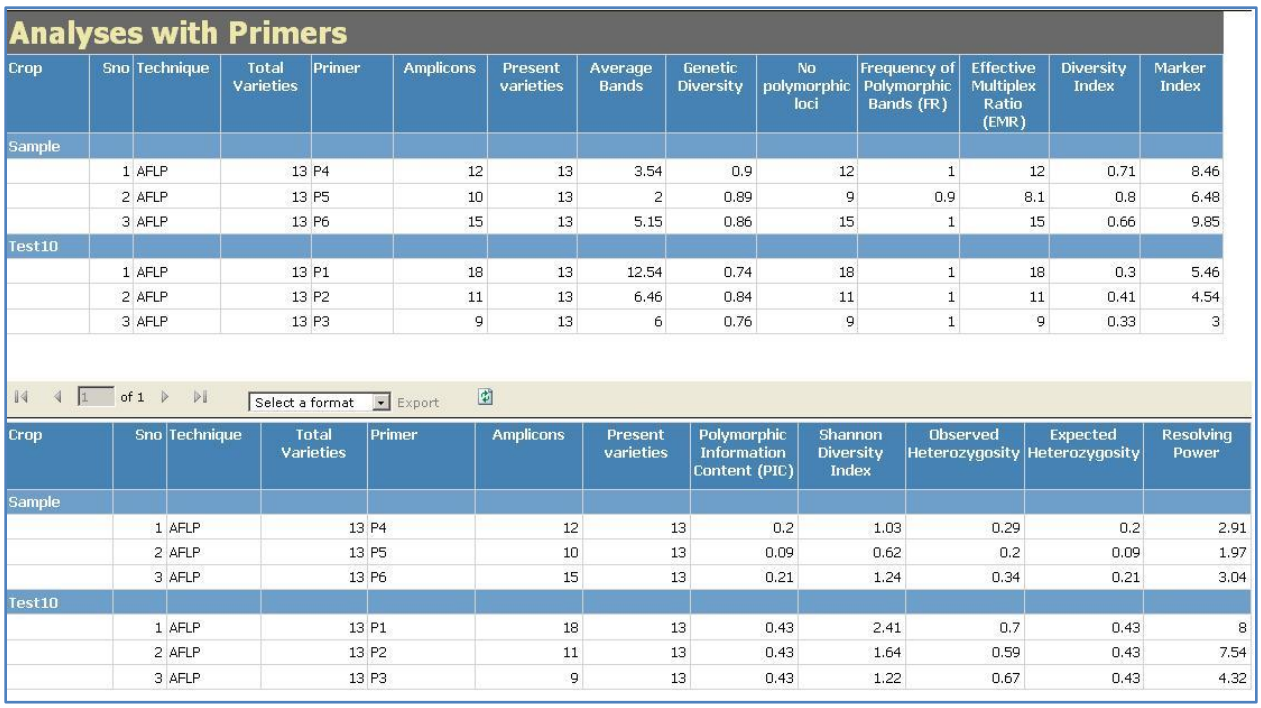
Crop:	Sample
Technique:	AFLP
Total Varieties:	13

Sno	Primer	Amplicons	Present varieties	Average Bands	Genetic Diversity	No. of polymorphic loci	Frequency of Polymorphic Bands (FR)	Effective Multiplex Ration (EMR)	Diversity Index	Marker Index
1	P1	18	13	12.54	0.74	18	1	18	0.3	5.46
2	P2	11	13	6.46	0.84	11	1	11	0.41	4.54
3	P3	22	13	13.08	0.8	22	1	22	0.41	8.92
4	P4	12	13	3.54	0.9	12	1	12	0.71	8.46

1 of 1 Select a format Export

Sno	Primer	Amplicons	Present varieties	Polymorphic Information Content (PIC)	Shannon Diversity	Observed Heterozygosity	Expected Heterozygosity	Resolving Power
1	P1	18	13	0.43	2.41	0.7	0.43	8
2	P2	11	13	0.43	1.64	0.59	0.43	7.54
3	P3	22	13	0.39	2.86	0.59	0.39	10.16
4	P4	12	13	0.2	1.03	0.29	0.2	2.91
5	P5	10	13	0.09	0.62	0.2	0.09	1.97
6	P6	15	13	0.21	1.24	0.34	0.21	3.04
7	P7	9	13	0.17	0.84	0.19	0.17	3.38
8	P8	42	13	0.39	5.17	0.67	0.39	16.17

Figure13. Report of Analyses with primers for "sample" crop profile data



Analyses with Primers

Crop	Sno	Technique	Total Varieties	Primer	Amplicons	Present varieties	Average Bands	Genetic Diversity	No polymorphic loci	Frequency of Polymorphic Bands (FR)	Effective Multiplex Ratio (EMR)	Diversity Index	Marker Index
Sample													
	1	AFLP	13	P4	12	13	3.54	0.9	12	1	12	0.71	8.46
	2	AFLP	13	P5	10	13	2	0.89	9	0.9	8.1	0.8	6.48
	3	AFLP	13	P6	15	13	5.15	0.86	15	1	15	0.66	9.85
Test10													
	1	AFLP	13	P1	18	13	12.54	0.74	18	1	18	0.3	5.46
	2	AFLP	13	P2	11	13	6.46	0.84	11	1	11	0.41	4.54
	3	AFLP	13	P3	9	13	6	0.76	9	1	9	0.33	3

1 of 1 Select a format Export

Crop	Sno	Technique	Total Varieties	Primer	Amplicons	Present varieties	Polymorphic Information Content (PIC)	Shannon Diversity Index	Observed Heterozygosity	Expected Heterozygosity	Resolving Power
Sample											
	1	AFLP	13	P4	12	13	0.2	1.03	0.29	0.2	2.91
	2	AFLP	13	P5	10	13	0.09	0.62	0.2	0.09	1.97
	3	AFLP	13	P6	15	13	0.21	1.24	0.34	0.21	3.04
Test10											
	1	AFLP	13	P1	18	13	0.43	2.41	0.7	0.43	8
	2	AFLP	13	P2	11	13	0.43	1.64	0.59	0.43	7.54
	3	AFLP	13	P3	9	13	0.43	1.22	0.67	0.43	4.32

Figure14. Report of Analyses with primers for “sample” and “test10” crop profile data

Average Number of Bands per Cultivar

In order to find best informative primer module of average number of bands per cultivar analysis is used. The formula used for calculating average number of bands per cultivar is as follows:

$$F = \frac{\text{(Sum of present bands/ cultivars)}}{\text{Total number of cultivars}}$$

This analysis helps to select best primer in a profile table. Higher the value of F better is primer.

Polymorphic Information Content

The polymorphic information content (PIC) value is commonly used in genetics as a measure of polymorphism for a marker locus used in linkage analysis (Botstein *et al.)*. According to Dr. K.V. Bhat polymorphism information content (PIC) for each marker is determined separately using the following equation:

$PIC= 1-[pi^2 + (1-pi^2)]$

Where pi= frequency of i^{th} marker in the data set.

In markers having null alleles the formula is changed as PIC= 2pi (1-pi)

Where pi= frequency of the i^{th} null allele

Higher the value of PIC better is the primer.

Gene Diversity Analysis

Genetic relationship among different primers is found by using gene diversity analysis. The formula used for gene diversity is

$$H_{es} = (1 - \Sigma pi_2)/N$$

Where pi : mean frequency of the allele

N : Total number of present cultivars in a primer

Higher the value of gene diversity better is the primer

Resolving Power Analysis

To find overall suitability of a primer for the purpose of identification, formula given by Prevost and Wilkinson on Resolving Power(Rp) calculation is used.

Rp= Sum of Band Informativeness(IB)

IB= 1-(2*|0.5-p|)

Where p is the proportion of accessions containing the band.

Higher the value of Resolving Power (Rp) better is the primer.

Shannon Diversity Index Analysis

The Shannon diversity index (*H*) is another index that is commonly used to characterize species diversity in a community. Shannon's index accounts for abundance of the species present. The proportion of species *i* relative to the total number of species (p_i) is calculated, and then multiplied by the natural logarithm of this proportion ($\ln p_i$). The resulting product is summed across species, and multiplied by -1:

$$H = -\sum_{i=1}^{n} pi \ln (pi)$$

Higher the value of Shannon Diversity Index (*H*) better is the primer.

Marker Informativeness Analysis

Genotype differentiation between markers is found by using module for Marker Informativeness. Marker Informativeness analysis deals with Fraction of polymorphic loci (FR), Effective multiplex ratio (EMR), Diversity index (DI) and Marker index (MI).

According to Powell *et al*. (1996) Diversity index (DI) for genetic markers is calculated from the sum of the squares of allele frequencies using following formula.

$DI = 1-\sum pi^2$

where 'pi' is the allele frequency of the ith allele.

Effective multiplex ratio EMR (E) is the product of the fraction of polymorphic loci (FR) and the number of polymorphic loci for a particular primer.

$EMR\ (E) = n_p = (n_p / n)$

where 'n_p' is the number of polymorphic loci and n is the total number of loci.

Fraction of polymorphic loci (FR) is ratio total number of polymorphic loci with total number of loci for a particular primer.

$FR = (n_p / n)$

Marker index (MI) is defined as the product of the average diversity index for polymorphic bands in any assay and the EMR for that assay, MI=DI * E.

Polymorphic Loci Analysis

Percent polymorphism of primers in a profile table is calculated by using following formula.

$$\text{Ps of a Primer} = \frac{\text{Polymorphic amplicons} \times 100}{\text{Total Amplicons}}$$

Expected (H_E) and Observed (H_o) Heterozygosity

The variation in alleles is critical to the survival of a species and allows organisms to adapt to changing environments. Allele frequency, or the frequency at which alleles are found at any locus of interest, is used to estimate the frequency of a given genetic profile. Every diploid cell has two alleles, one inherited from each parent. If an individual has two different alleles at a specific locus, the individual is heterozygous at that locus; if the two alleles are the same, the individual is homozygous. Allele frequency is used to characterize the genetic diversity, or richness of the gene pool, in a population. Populations need variation. The measure of the amount of heterozygosity across loci can be used as a general indicator of the amount of genetic variability.

Frequency of Occurrence of Marker

The formula used for calculating frequency of occurrence of marker is:

$$X = (2 \times A) / (Va + Vb)$$

A: Total number of matched present bands between cultivar a and cultivar b.

Va: Total number of present bands in cultivar a.

Vb: Total number of present bands in cultivar b.

Report of Frequency of Occurance of Marker

Crop:	Sample
Technique:	AFLP
Total Varieties:	13

Sno	Primer	Cultivars I	Cultivars II	A Matched cultivars	Va Present cultivars of I Cultivar	Vb Present cultivars of II Cultivar	X Frequency of Occurance of Marker
	P1						
		Variety1					
1			Variety2	14	16	16	0.88
2			Variety3	15	16	15	0.97
3			Variety4	8	16	8	0.67
4			Variety5	5	16	5	0.48
5			Variety6	13	16	15	0.84
6			Variety7	14	16	15	0.9
7			Variety8	13	16	14	0.87
8			Variety9	15	16	15	0.97
9			Variety10	14	16	16	0.88
10			Variety11	15	16	16	0.94
11			Variety12	0	16	0	0
12			Variety13	12	16	12	0.86
		Variety10					
73			Variety11	15	16	16	0.94
74			Variety12	0	16	0	0
75			Variety13	11	16	12	0.79
		Variety11					
76			Variety12	0	16	0	0

Figure15a. Report for frequency of occurance of marker calculation for “sample” data

Report of Frequency of Occurance of Marker

Crop:	Sample
Technique:	AFLP

Sno	Primer	Total Varieties	Average of Frequency Of Occurance of Marker X
1	P1	13	52.33
2	P2	13	36.57
3	P3	13	46.14
4	P4	13	46.54
5	P5	13	48.86
6	P6	13	54.8
7	P7	13	6.76
8	P8	13	55.33
9	P9	13	42.32

Figure15b. Report for frequency of occurance of marker calculation for "sample" data

Probability of Chance Identity

Module for probability of chance identity (*I*) (Paetkau *et al.*, 1995), is used to find the probability of two random primers displaying the same genotype using following formula.

Probability of chance/genetic identity (*I*) = X^F

X = Frequency of occurrence of marker and

F = Average number of bands per cultivar.

This analysis informs about chance of probability of two unrelated primers having the same pattern. Lower the probability better is the primer.

Analyses with Primer

Crop: SampleCrop
Technique: AFLP Total Varieties: 13

Sno	Primer	Present varieties	F: Average no. of Bands per cultivars	X: Frequency of Occurance of Marker	Probability of Chance Identity
1	P3	13	6	0.63	0.0625
2	P2	13	6.46	0.47	0.0076
3	P1	13	12.54	0.67	0.0066

Figure16. Report for probability of chance identity calculation for "samplecrop" profile data

3.2 Analyses with Varieties/Cultivars

Assessment of similarity from DNA fingerprints is based on comparison of band profiles. DNA fragments, appearing as bands, that have moved the same length on electrophoresis gel and thus bear same molecular weights are considered to be identical. The proportion of these shared bands is indicative of the relationship among the samples under comparison.

To compare similarity between two varieties, three modules of similarity coefficient analysis were developed.

Simple Matching Coefficients (*SM*): It is done for Random Amplification of Polymorphic DNA (RAPD) and Amplified Fragment Length Polymorphism (AFLP) Markers i.e. for dominant markers. This coefficient considers that absence of variable corresponds to homozygous loci.

		Sample B		
Sample A		+	-	
	+	AB	Ab	∑A
	-	aB	ab	∑a
		∑B	∑b	N

$$SM = (\sum AB + \sum ab) / (\sum AB + \sum ab + \sum Ab + \sum aB)$$

a. Jaccard's Coefficients (*J*): It is done for AFLP and RAPD markers i.e. for dominant markers. In this case absence of variety is considered as missing data.

$$J = \sum AB / (\sum AB + \sum Ab + \sum aB)$$

b. Dice Coefficients (*D*): It is also called as Nei and Li's Coefficients. This coefficient is for co-dominant markers such as Restriction Fragment Length Polymorphism (RFLP) and Simple Sequence Repeat (SSR). This coefficient counts percentage of shared bands among two individual varieties and gives more weight to those bands that are present in both. It considers that absence of variety has less biological significance. Therefore, this coefficient has complete meaning in DNA similarity.

$$D = 2\sum AB / (2\sum AB + \sum Ab + \sum aB)$$

Coefficients analysis helps to know the extent of similarity between varieties.

Comparison may be done on one to one and one to many varieties.

<u>Calculation of similarity coefficient between a variety of a profile table with all other varieties</u>

There are two types of comparison for similarity coefficient calculation.

1. Comparison within the same profile table
2. Comparison with new profile table having the same primer.

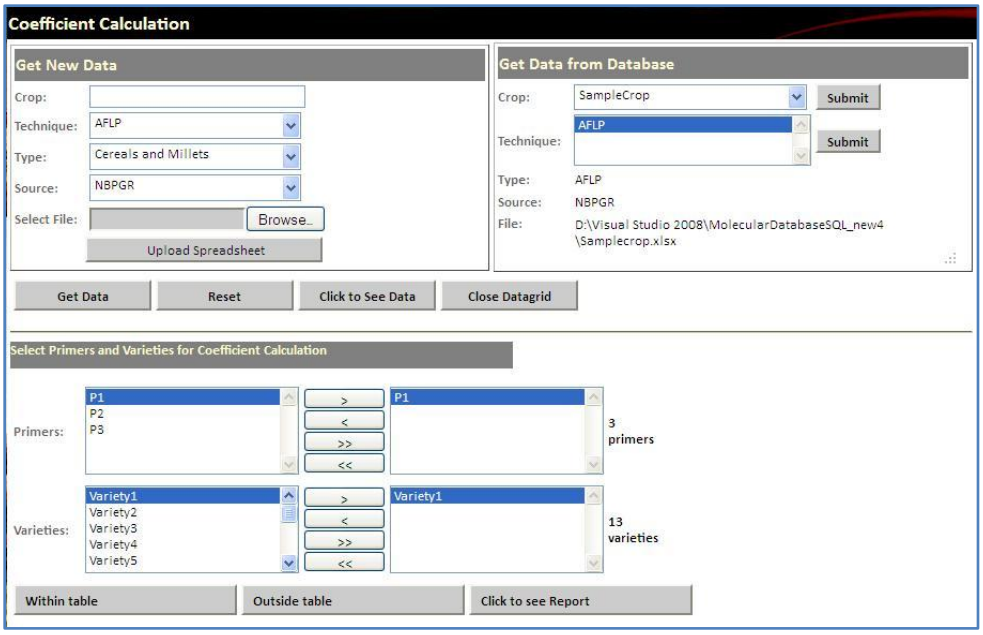


Figure17. Web form for coefficient calculation

If user chooses to compare within the same profile table then click the button 

. As soon as choice is made calculation starts. In the above case, If user has made choice of "Within table " then comparison of variety1 is done with 13 other varieties of samplecrop data, and result is displayed in datagrid as shown in figure16 .

Sno	Crop	Primer	Variety	Compared_variety	Matched	Unmatched	Jaccards_Coefficient	Dice_Coefficient	Simple_Matching_Coefficient
1	SampleCrop	P1	Variety1	Variety12	0	16	0	0	0
2	SampleCrop	P1	Variety1	Variety5	5	11	0.31	0.48	0.39
3	SampleCrop	P1	Variety1	Variety4	8	8	0.5	0.67	0.56
4	SampleCrop	P1	Variety1	Variety6	13	5	0.72	0.84	0.72
5	SampleCrop	P1	Variety1	Variety2	14	4	0.75	0.86	0.78
6	SampleCrop	P1	Variety1	Variety8	13	4	0.76	0.87	0.78
7	SampleCrop	P1	Variety1	Variety10	14	4	0.78	0.88	0.78
8	SampleCrop	P1	Variety1	Variety13	12	4	0.78	0.88	0.78
9	SampleCrop	P1	Variety1	Variety7	14	3	0.82	0.9	0.83
10	SampleCrop	P1	Variety1	Variety11	15	2	0.88	0.94	0.89
11	SampleCrop	P1	Variety1	Variety3	15	1	0.94	0.97	0.94
12	SampleCrop	P1	Variety1	Variety9	15	1	0.94	0.97	0.94
13	SampleCrop	P1	Variety1	Variety1	16	0	1	1	1

Figure18. Datagrid showing similarity coefficient analysis with "samplecrop" for Primer "P1"

Click on "Click to See Report" button leads to show the formatted report of comparison as shown in figure 17.

Report of Similarity Coefficient Analysis

Crop: SampleCrop

Primer	Sno	Variety	Compared variety	Matched	Unmatched	Jaccards Coefficient	Dice Coefficient	Simple Matching Coefficient
P1								
	1	Variety1	Variety12	0	16	0	0	0
	2	Variety1	Variety5	5	11	0.31	0.48	0.39
	3	Variety1	Variety4	8	8	0.5	0.67	0.56
	4	Variety1	Variety6	13	5	0.72	0.84	0.72
	5	Variety1	Variety2	14	4	0.75	0.86	0.78
	6	Variety1	Variety8	13	4	0.76	0.87	0.78
	7	Variety1	Variety10	14	4	0.78	0.88	0.78
	8	Variety1	Variety13	12	4	0.78	0.88	0.78
	9	Variety1	Variety7	14	3	0.82	0.9	0.83
	10	Variety1	Variety11	15	2	0.88	0.94	0.89
	11	Variety1	Variety3	15	1	0.94	0.97	0.94
	12	Variety1	Variety9	15	1	0.94	0.97	0.94
	13	Variety1	Variety1	16	0	1	1	1

Figure19. Report showing similarity coefficient analysis for “samplecrop” profile data

If comparison is to be made with new profile table then click the button 

. When user clicks button of “Outside table” then program invokes provision to select new data from database, as shown in figure 18.

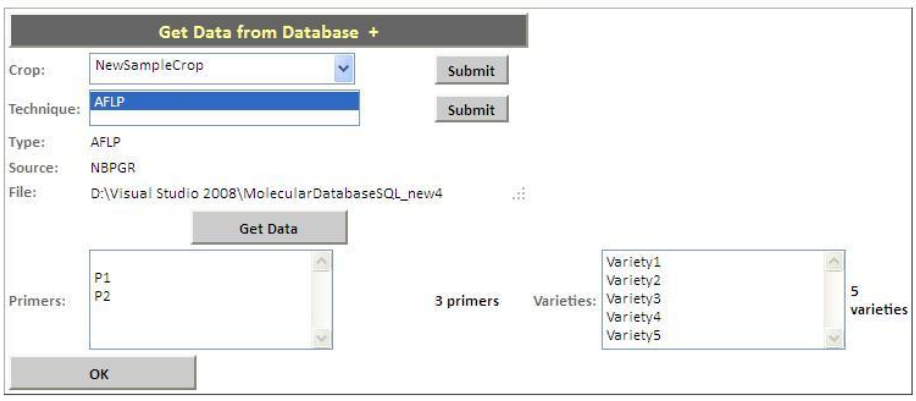


Figure20. Provision to select data from database, for doing comparison with new profile table

As shown in figure 18, if user has choosen “NewSampleCrop” for comparing with “SampleCrop” profile data, then it should be having the same primer. In this case

the profile data of NewSample Crop is shown in figure 19. In this Profile table there are two primers P1 and P2 and five varieties.

Sno	Bandsize	Primer	Variety1	Variety2	Variety3	Variety4	Variety5
1	0	P1	1	1	1	1	1
2	0	P1	0	1	0	0	0
3	0	P1	1	1	1	0	0
4	0	P1	1	1	1	0	0
5	0	P1	1	0	0	0	0
6	0	P1	1	1	1	1	0
7	0	P1	1	0	1	1	0
8	0	P1	1	1	1	1	1
9	0	P1	1	1	1	0	1
10	0	P1	1	1	1	1	0
11	0	P1	0	1	0	0	0
12	0	P1	1	1	1	1	1
13	0	P1	1	1	1	0	0
14	0	P1	1	1	1	1	1
15	0	P1	1	1	1	1	0
16	0	P1	1	1	1	0	0
17	0	P1	1	1	1	0	0
18	0	P1	1	1	1	0	0
19	0	P2	1	1	1	1	0
20	0	P2	1	1	0	0	0
21	0	P2	1	1	1	0	0
22	0	P2	1	1	0	0	0
23	0	P2	0	1	0	0	0
24	0	P2	1	1	1	0	0
25	0	P2	1	1	0	0	0
26	0	P2	0	1	0	0	0
27	0	P2	0	1	0	0	0
28	0	P2	0	1	0	0	0
29	0	P2	0	1	0	0	0

Figure21. Profile data of “NewSampleCrop” data

For conducting analysis if Primer “P1” of “SampleCrop” is considered then comparison will be done with primer “P1” of “NewSampleCrop”. After calculation of similarity coefficients, datagrid shows results of analysis.

Sno	Crop	Compared Crop	Primer	Variety	Compared variety	Matched	Unmatched	Jaccards Coefficient	Dice Coefficient	Simple Matching Coefficient
1	SampleCrop	NewSampleCrop	P1	Variety1	Variety5	5	11	0.31	0.48	0.39
2	SampleCrop	NewSampleCrop	P1	Variety1	Variety4	8	8	0.5	0.67	0.56
3	SampleCrop	NewSampleCrop	P1	Variety1	Variety2	14	4	0.78	0.88	0.78
4	SampleCrop	NewSampleCrop	P1	Variety1	Variety3	15	1	0.94	0.97	0.94
5	SampleCrop	NewSampleCrop	P1	Variety1	Variety1	16	0	1	1	1

Figure22. Datagrid showing similarity coefficient analysis with “SampleCrop” and new crop profile table of “NewSampleCrop” for Primer “P1”

Figure 20 and figure 21, which is report of analysis depicts that comparison of “variety1” from “SampleCrop” is being done with all five varieties of “NewSampleCrop” of Primer “P1”.

Report of Similarity Coefficient Analysis

Crop: SampleCrop

Compared Crop: NewSampleCrop

Primer	Sno	Variety	Compared variety	Matched	Unmatched	Jaccards Coefficient	Dice Coefficient	Simple Matching Coefficient
P1								
	1	Variety1	Variety5	5	11	0.31	0.48	0.39
	2	Variety1	Variety4	8	8	0.5	0.67	0.56
	3	Variety1	Variety2	14	4	0.78	0.88	0.78
	4	Variety1	Variety3	15	1	0.94	0.97	0.94
	5	Variety1	Variety1	16	0	1	1	1

Figure23. Report showing similarity coefficient analysis with “SampleCrop” and new profile table of “NewSampleCrop” for Primer “P1”

****