Assessment of Genetic Diversity of mustard (*Brassica juncea* (L.) Czern & Coss) germplasm in Sri Lanka: An Application of Self-Organizing Map (SOM) with dimension reduction

P.K.D. Peiris[1], S. R. Weerakoon[2] and S. Somaratne[2]
[1]Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, 94, Cumarathunga Munidasa Mawatha, Colombo 03, Sri Lanka
[2]Department of Botany, The Open University of Sri Lanka, P. O. Box 21, Nawala, Sri Lanka.

INTRODUCTION

Mustard (*Brassica juncea* L.) is one of the economically most important species in the genus *Brassica* of the family Brasicaceae. Mustard has been grown in the Indian subcontinent for many hundreds of years as an oil seed crop (Paramjit *et al*., 1991; Labana and Gupta, 1993). However, in Sri Lanka, mustard is use as a condiment and oil is use for cooking and in therapeutic purposes and it is grown as a subsidiary crop in the dry and intermediate zones (Andrahennadi *et al*., 1991). In Sri Lanka, there are over 60 local mustard accifferent fields of studies in solving complex problems involving nonlinearity and uncertainty. The rapid increase in the computation power and new algorithms has made it simple to use the neural networks. There is an increasing tendency in the application of ANN in many fields of studies. The ANNs are capable of dealing with the datasets with high level of noise and complexity as well as with their power to unravel patterns from datasets in which there are complex non-linear interactions. A self-organizing map (SOM) or self-organizing feature map (SOFM) is a type of ANN that is trained using unsupervised learning to produce a low dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map. Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space (Kohonen, 1984). The main objective of the study was to train a Self-Organizing Maps (SOMs) and used to study the grouping patterns of mustard accession reflected from the agro-morphological characters of 44 different mustard accessions in order to differentiate the mustard germplasm in Sri Lanka.

MATERIALS AND METHODS

Forty four local (44) mustard (*Brassica juncea*) accesions were obtain from the plant Genetic Resource Center Gannoruwa, Sri Lanka. A total of 44 accessions used in this study and an additional two accessions were given in Table 1.

The seeds of each accession were sown in plastic seedbeds in a plant house at Open University of Sri Lanka. The seeds of each genotype were soaked in water and allowed to germinate under natural condition. A total of twenty seedlings of each accession were planted in black polythene bags with standard potting mixture. Subsequently, the seedlings (3-4-leaf stage) were transferred to plastic pots of 13 cm diameter. Each of these replicate was arranged in Randomized Complete Block Design (RCBD). The distance between pots was 30 cm and between row was 60 cm. The Row lengths were kept as 360 cm. A total of ten plants were chosen at random from each mustard genotype for measuring/recording.

The traits selection and method of measurement were based on International Board for Plant Genetic Resources (IBPGR, 1991), Descriptors for *Brassica* and *Raphanus* and Morphological Descriptors for mustard (Rabbani *et al*., 1993). The measure/recorded agro-morphological traits were measures/recorded from the seedling stage o the harvesting stage of the crop. A total of 35 agro-morphological traits were measured/recorded. Classification and Regression Trees (CART) analysis was performed on the on the dataset for the purpose of objectively selection of variables prior to training of SOM. Based on the results of CART analysis, a total of five variables (leaf length, leaf width, seeds per plant, silique length and stem coloration) were selected.

RESULTS AND DISCUSSION

Self-Organizing Maps (SOMs) were trained and used to study the grouping patterns reflected from the agro-morphological characters of 44 different mustard accessions, in order to differentiate the mustard germplasm in Sri Lanka. The agro-morphological dataset which consists of 35 variables were reduced to five variables (leaf length, leaf width, seeds per plant, silique length and stem coloration) using CART. The constructions of SOMs were made after the pre-processing of data by normalizing to range from 0 to 1 to avoid scaling effect of the measurement used in the variables. A number of SOMs were trained with different sized (from 6 X 6 to 16 X 16 neurons) and the best size was chosen based on the RMSE (Root Mean Sums of Square Error). The SOM that yield the best grouping of the mustard accessions was obtained for a 12 X 12 size (RMSE = 0.0124), trained for 300 epochs. An appreciable grouping of mustard accessions was observed (Fig. 1. F). A total of seven groups were presented. Group 1 included accession Nos. 346 8640, 7789, 1396 (Orange); Group 2 - 327 and 508 (Yellow); Group 3- 8658 (Red); Group 4 - 2181, 2184, 9725, 790, 1040, 2181, 2182, 2184 and 9723 (Green); Group 5 – 346, 8640, 1256, 501, 1813, 1814, 7700, 7786, 7815, 8722, 8831, 8852, 327, 346, 8640, 1396, 2186 and 5041 (Black); Group 6 - 2376, 1244, 9724 and 2180 (Sierra Brown); Group 7 – 1146, 2310, 7792, 7814, 1774, 1847, 9418, 1040, 1146, 790, 2056, 2182, 9723, 2186, 5041 and 5184 (White). The accession number such as 8658 well separated (Group 1) and certain accession numbers become almost mixed indicating that these accessions may be represent single accession which has been collected from different localities of the island. The comparison of weight mps and SOM classification of mustard accessions (Fig. 1), reveals that the grouping pattern of the accessions was mainly accounted by the variables such as leaf breadth, leaf length and number of seeds per plant and to lesser extent average pod length and stem color.

The results of the present study show that there was a significant genetic diversity among mustard germplasm in Sri Lanka and support the previous findings (Weerakoon, *et al*., 2007). However, the results of the present study suggest that there are cases in which the same accession were assigned different accession numbers and the different accessions are assigned same accession number. The ecological origin of the accession further reveals that numbering methods of the mustard accession seems to be less strict and which in turn led to more problems in classifying the musftard accessions in Sri Lanka. The future attempts at *Brassica* breeding should take these results into consideration. This study further indicated that application of SOM helps gene bank to characterize mustard germplasm avoiding confusion results from the duplication of accession numbers. Therefore, SOM is of great importance in classification and characterization of germplasm through which the proper management of gene banks is possible.

Table1. Total number of Sri Lankan mustard accessions and abbreviations used in the study (Source: PGRC, 1999).

Abbreviation	Accession no	Abbreviation	Accession no	Abbreviation	Accession no
AC1	327	AC18	1841	AC35	7789
AC2	346	AC19	1847	AC36	7792
AC3	347	AC20	2056	AC37	7814
AC4	501	AC21	2180	AC38	7815
AC5	508	AC22	2181	AC39	8640
AC6	790	AC23	2182	AC40	8658
AC7	1040	AC24	2184	AC41	8722
AC8	1098	AC25	2186	AC42	8831
AC9	1099	AC26	2223	AC43	8852
AC10	1146	AC27	2310	AC44	9418
AC11	1774	AC28	2376	AC45	9723
AC12	1244	AC29	5088	AC46	9724
AC13	1256	AC30	5044	AC47	9725
AC14	1396	AC31	5181		
AC15	1813	AC32	5184		
AC16	1814	AC33	7700		
AC17	1831	AC34	7786		

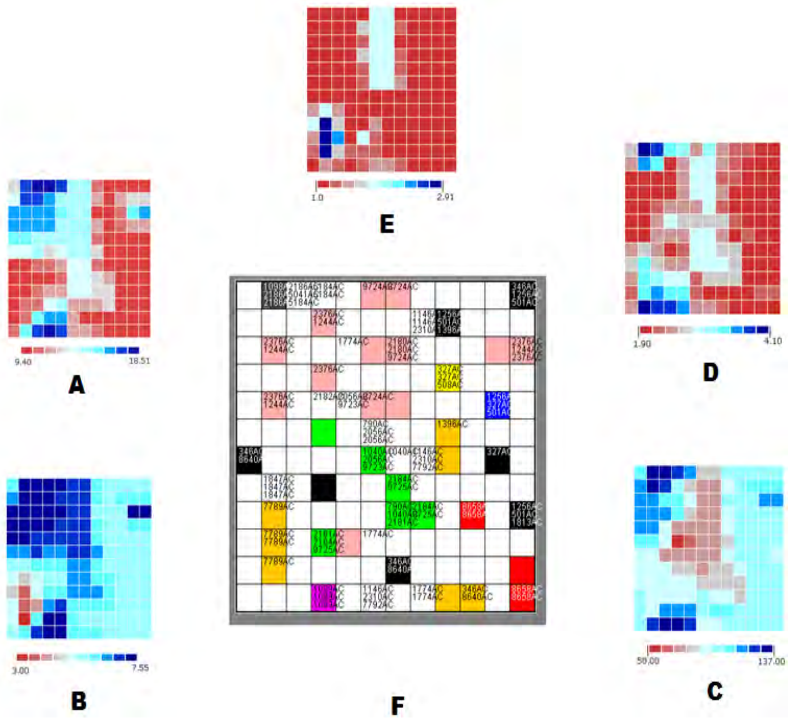

Fig. 1. Chosen SOM for the classification of Sri Lankan Mustard accessions (F) and the map of the layers of the weight for leaf length (A), leaf width (B), Number of seeds per plant (C), average pod length (D) and stem color (E). Color codes: Orange – Group 1, Yellow – Group 2, Red – Group 3, Green – Group 4, Black – Group 5, Sierra brown - Group 6 and White – Group 7.