

Genetic diversity of weedy rice (*Oryza sativa* f. *spontanea*) populations in Sri Lanka: An application of Self Organizing Map (SOM)

SHYAMA R. WEERAKOON*, SENEVIRATNE SOMARATNE

Department of Botany, The Open University of Sri Lanka, Nawala, Nugegoda 10250, Sri Lanka. *email: shyamaweerakoon@gmail.com

Manuscript received: 29 November 2020. Revision accepted: 21 March 2021.

Abstract. Weerakoon SR, Somaratne S. 2021. Genetic diversity of weedy rice (*Oryza sativa* f. *spontanea*) populations in Sri Lanka: An application of Self Organizing Map (SOM). *Asian J Agric* 5: 35-43. Weedy rice (WR) (*Oryza sativa* f. *spontanea*) has become a major threat in rice cultivation. Discrimination of WR from cultivated rice is difficult since agro-morphology of WR and cultivated rice are overlapping. Molecular markers are useful and can be an informative tool for estimating genetic diversity and relationships in closely related WR eco-types. Self-Organizing Maps (SOM) is an interesting and promising classification tool employing an innovative and data-driven classification method based on unsupervised artificial neural networks. The present study focused on exploring the potential use of SOM to classify WR populations of different eco-climatic zones in Sri Lanka using agro-morphological and molecular data. Separate SOMs for each set of variables, agro-morphological and molecular data were developed. The best SOM was chosen based on the error performance. Findings of SOM analyses showed that certain morphological characters (seedling height, leaf blade width, leaf blade length, culm strength, panicle shattering, seed coat color and leaf angle) and certain molecular characters detected from SSR primers (RM 11, RM 21, RM 14, and RM 280) are important in separation of different WR eco-types satisfactorily. SOM clustering of cultivated, wild, and WR eco-types indicated specific patterns of grouping with respect to climatic conditions of the country. WR eco-types in dry zone and wet zone of the country are closely related to *Oryza nivara* and *O. rufipogon* respectively.

Keywords: Genetic diversity, *Oryza sativa* f. *spontanea*, Self-Organizing Map, Sri Lanka, weedy rice

INTRODUCTION

Weedy rice (WR) (*Oryza sativa* f. *spontanea*) belongs to the Family Poaceae and is widely distributed in rice-planting areas all over the world, particularly in South and South-East Asia, South and North America, and Southern Europe (Mortimer et al. 2000; Chauhan and Johnson 2010; Chauhan 2012). Emergence of WR leads to high production costs, reduction of yield (Azmi et al. 2008) and lowered commodity value by staining the grains with undesirable pericarp color (Mortimer et al. 2000). Presently, WR has reached a considerable competitive level of infestation threatening rice cultivation sustainability, especially in Asian countries (Londo et al. 2007). WR was first reported in 1992 from Ampara District, Sri Lanka, and has spread into many areas of the country. WR occurs in all agro-ecological zones of the country with varying population densities (Abeysekara et al. 2010). Factors that contribute to the distribution of WR are: gene flow between wild and cultivated rice varieties, old rice varieties becoming feral, and crosses between cultivated rice and wild rice (Marambe et al. 2000; Karunarathna et al. 2014).

Microsatellite markers were coupled with morphological characteristics to explore genetic diversity and possible origins of weedy rice in Taizhou city, Jiangsu Province, China. With wild rice, hybrid rice and cultivated rice revealing low diversity amongst weedy rice populations and suggesting that weedy rice has originated from segregating populations of hybrid rice that hybridized naturally with cultivated rice (Zhang et al. 2008). Results

of the molecular studies on the Sri Lankan weedy rice populations carried out by He et al. (2014) have shown that a high level of genetic diversity as estimated by 23 SSR loci, which is comparable with that estimated based on morphological characterization. The genetic features of weedy rice (*Oryza sativa* f. *spontanea*) have been studied based on morpho-physiology, random amplified polymorphic DNA (RAPD) (Suh et al. 1997), restriction fragment length polymorphism (RFLP) (Cho et al. 1995), amplified fragment length polymorphism (AFLP) (Federici et al. 2001) and simple sequence repeat (SSR) (Cao et al. 2006; Ma et al. 2008).

SSR markers have been identified as a powerful tool due to their abundance in eukaryotic genomes, co-dominance, and high polymorphisms (Gwag et al. 2010; Zhang et al. 2011) and thus, SSR markers have been used as allele-specific and co-dominant markers in population genetic and evolutionary studies of many plants (McKhann et al. 2004; Upadhyaya et al. 2006).

More recently, with the advancement of computer technology and algorithms, computerized keys with interactive systems have been used for plant identification which allows rapid identification and information retrieval. There have been several major approaches made to use computer algorithms in identification of biological specimens. These include computer-stored dichotomous keys; computer constructed keys, simultaneous characteristic-set methods, and automated pattern recognition (Dallwitz 1980). The Artificial Neural Network (ANN) has been developed to simulate the activity of the

human brain (Simon 1999). Among the ANN learning algorithms, the Self-Organizing Map (SOM) is a learning algorithm that has been proposed by Kohonen (1982). A SOM algorithm also known as Kohonen Map or Self-Organizing Feature Map is an unsupervised neural network based on competitive learning (Kohonen 1998; Kohonen 2001). SOM is considered as an effective and innovative, data-driven classification method built on unsupervised Artificial Neural Networks (ANN) (Wang et al. 2014). SOM has been applied in varied disciplines such as agriculture and environment management due to prediction, estimation, classification, remote sensing image classification, etc. (Agarwal and Skupin 2008; Brosse et al. 2001). Further, SOM reduces the high-dimensional input data onto a low-dimensional space.

In general, SOM consists of two types of units *viz.*, an input layer and an output layer (input variables). The array of input units operates simply as a flow-through layer for the input vectors and has no further significance. SOM often consists of a two-dimensional network of neurons arranged in a square grid or lattice in the output layer. Each neuron is connected to its nearest neighbors on the grid. The neurons store a set of weights, each of which corresponds to one of the data inputs (Vesanto and Alhoniemi 2000). During these processes, weights are adjusted through iteration. The outcome weight vectors of the SOM nodes are allocated to have characteristic data patterns. An important feature of SOMs that distinguishes them from Vector Quantization techniques is that the neurons are organized on a regular grid. During training, both Best-Matching Neuron (BMU) and its topological neighbors are updated.

The comparison of results obtained from the SOMs with the results obtained from conventional multivariate statistical methods, such as PCA and correspondence analysis, revealed that SOM facilitates both visualizations of the sample units and finds underline structures in ecological communities (Lek and Gue gan 2000). Application of SOM in an environmental data set for predicting the species richness of aquatic insects in streams has been successfully achieved (Brosse et al. 2001). The SOM, the MLP, and a network based on the adoptive resonance theory (ART) used in animal science applications (Fernandez et al. 2006). Various algorithms (including k-means and fuzzy c-mean clustering techniques, SOM, and fuzzy adaptive resonance theory (ART) have been applied to the dive profiles of penguins and seals (Schreer et al. 1998). Unsupervised classification methods such as SOM (Li 2002) have been attempted to identify and classify crop weeds (Moshou et al. 2001; Hemming and Rath 2001). The long-term uses of SOM-assisted discrimination of weedy rice from cultivated rice can be potentially employed in monitoring edible rice cultivations, combating unwanted eco-types of rice, supporting weedy rice combustion with artificial intelligence tools used by drones, etc.

In agriculture, visualization of complex agro-morphological and even molecular data of different crop

varieties in two-dimensional space is very important. In such situation, SOMs can be used as supplementary to conventional statistical analyses. However, there is scarcity in the studies that focus on complex patterns of agro-morphological and molecular characters across different rice varieties in different origins of ecology. The use of artificial intelligence systems in controlling weeds, including weedy rice in rice ecosystems, is important for widely distributed weeds. SOM provides a way to recognize the patterns reflected from the output data and the relationships between the input data. Thus, there are benefits of application of SOM to solve the problems of species identification *viz.* minimize the need for user interaction and improve the capability of the system to learn previously vaguely defined complex visual patterns (Walley et al. 2000).

The objective of this study was to explore the potential use of the SOM in classification of weedy rice populations that widely occur in different climatic zones of Sri Lanka using agro-morphological and molecular data.

MATERIALS AND METHODS

Agro-morphological characterization

The seeds of weedy rice eco-types, cultivated rice varieties, and wild rice varieties were collected from rice fields located in twelve (12) Districts (Ampara [12 WR eco-types], Anuradhapura [8 WR eco-types], Batticaloa [11 WR eco-types], Hambantota [11 WR eco-types], Kandy [10 WR eco-types], Kurunegala [15 WR eco-types], Matale [10 WR eco-types], Matara [24 WR eco-types], Mannar [10 WR eco-types], Polonnaruwa [10 WR eco-types], Puttalam [11 WR eco-types], and Vauniya [10 WR eco-types] in different eco-climatic zones (Dry, Wet and Intermediate zones) in Sri Lanka (Figure 1).

Collected seeds were subjected to dormancy, breaking treatments, and sown in plastic trays in a plant house at the Open University of Sri Lanka, Nawala, Sri Lanka. Five replicates each with a single plant were planted in plastic pots with paddy soils. Replicates were arranged in a Complete Randomized Design (CRD). Morphological characterization using thirty-six (36) characters of different rice varieties (Table 1) was made using the Standard Rice Characterization Catalogue (PGRC 1999).

Molecular characterization

Total genomic DNA was extracted from 7-day old seedlings of WR eco-types, cultivated rice varieties, and wild rice varieties using Ceygen Plant total DNA purification kit. Ten SSR primer pairs were used. The primer sequences and amplification conditions for primers were obtained from <http://www.gramene.org/>. A four-primer system (Schuelke 2000) was used, which included a universal M13 oligonucleotide (TGTAACGACGGCC AGT), labeled with one of four fluorescent dyes (6-FAM, NED, PET, or VIC) (Table 2).

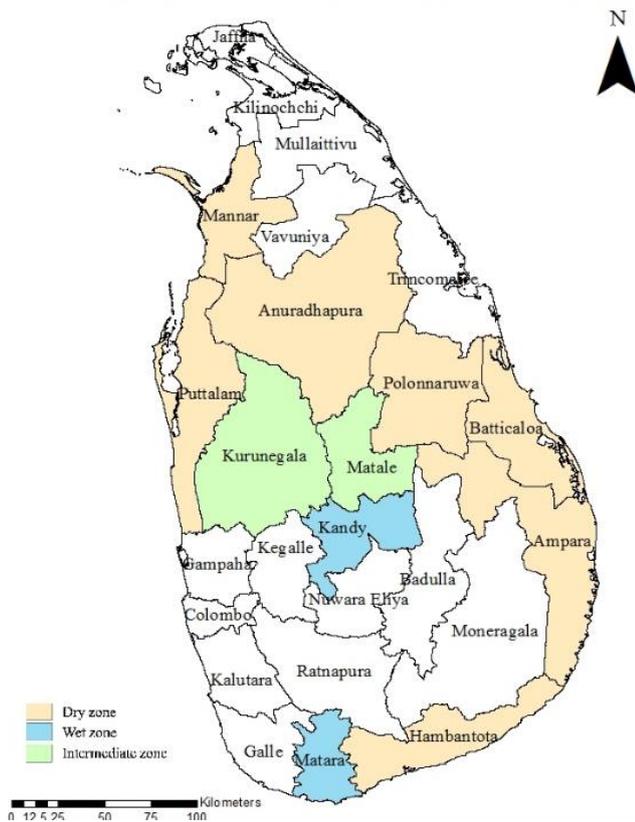


Figure 1. Geographical locations of the weedy rice populations sampled in Sri Lanka.

The fluorescent dyes allowed the products to be perplexed during electrophoresis; a special forward primer composed of a concentration of the M13 oligonucleotide; and the pigtail reverse primer for SSR PCR amplification. All amplification reactions were carried out in 30 μ l volumes of which contained 1 x PCR buffer, 1mM dNTPs, 2 μ M SSR primers, 2mM MgCl₂, 50ng of genomic DNA, and 0.5 Units of *Taq DNA polymerase*. The reaction conditions were: 95 °C for 1min, followed by 30 cycles of 95°C (30 sec), 55 °C (1 min), and 72°C (1 min), with 10 subsequent cycles of 95 °C (30 sec), 53°C (45 sec), and 72 °C (1 min), and a final extension at 72°C for 10 min. The SSR alleles were resolved on an ABI Prism 3100 DNA sequencer using GeneScan 4.1 software and sized precisely using GeneScan600LIZladder. Fragment analysis using capillary electrophoresis was performed using GENE MAPPER software and identified different peaks among weedy rice, cultivated rice varieties and wild rice varieties. Collected data were normalized (z-score) to have a zero mean and standard deviation of one (1).

Construction of Self-Organizing Map

A Self-organizing map (SOM) is a form of an ANN that can provide an objective way of classifying data through self-organizing networks of artificial neurons. It is a feed-forward ANN that uses an unsupervised training algorithm and can be trained to learn or find relationships between inputs or organize data to discover unknown patterns or structures. The SOM algorithm has been applied to a

variety of real-world problems (Oja 2002). The main advantage of applying the algorithm comes from the easy visualization and interpretation of the map clusters. One of the main reasons (Kohonen 2002) for using a SOM for exploratory data analysis and data mining is that it is a numerical method and can treat numerical statistical data naturally and represent graded relationships. Other reasons for using the SOM algorithm include that it is a non-parametric method; no assumptions about the distribution of data need be made in advance, and it is a method that can detect unexpected pattern structures by learning without supervision. The SOM can be used deductively and is able to produce results even if the data set is incomplete, which is extremely important when dealing with biological material requiring identification. Another advantage of the SOM is that it can provide a probability for each species of an unidentified biological specimen belonging to a particular species. At the same time, the SOM can also be used to investigate the differences between clusters of species, and in the process, it is possible that some new features that are discern between the species might be revealed. The SOM can also be used to determine which features or characteristics are the most important or diagnostic ones to consider when discerning between given species.

The SOM training utilizes a competitive learning strategy during which a weighted vector associated with each neuron in a neural network is modified and is gradually developed to become sensitive to a set of patterns from a specific domain of the input space. The result of the training process is that different neurons specialize to represent different types of input patterns. This specialization is enforced by competition among the neurons. Neighboring neurons are allowed to learn, those neurons will also gradually specialize to represent similar patterns, and consequently, the representations on the output layer will become ordered (Kaski 1996). Usually, the neurons are arranged on a regular 1- or 2-dimensional lattice type of array with a hexagonal and oblong arrangement. This arrangement can represent the data clusters better than a rectangular arrangement and fits the data input distribution more easily (Deboeck and Kohonen 1998). The number and positions of neurons on the grid are defined and fixed, when the map is created and depends on the purpose for which the SOM will be used and the amount of input data. Sometimes the number of neurons used is determined by a heuristic formula such as $5\sqrt{N}$ or \sqrt{N} , where N is the number of training patterns (Vesanto 2000).

The SOM is defined in the training phase. During the training, assumptions must be made about several parameters of the map, such as learning parameters, map topology and map size. These features influence the final map; thus it is very important to choose these parameters carefully in order to reach the appropriate map (Kohonen 2001). Once tested with different choices, it is possible to use some measures to evaluate the quality of the map and select the optimal one to represent the experimental data. Several measures have been used to evaluate the quality of a SOM. A widely used measurement is the quantization

error. This error measures the average distance between each data vector and its best matching unit (BMU). The quantization error (Eq) is calculated as shown in equation

(1), where N is the number of data-vectors, X_i is the current vector patterns, m_c is the BMN of the corresponding X_i input vector.

Table 1. Agro-morphological characters used in characterization of WR eco-types cultivated rice varieties, and wild rice species in Sri Lanka (PGRC Rice Catalog 1999).

Character no.	Character	Abbr.	Description
1	Seedling height (cm)	Sdlh	Recorded at the five-leaf stage
2	Leaf-blade length (cm)	LBL	Measured from topmost leaf below the flag leaf on the main culm at late vegetative stage.
3	Leaf-blade width (mm)	LBW	Measured at the widest portion of the leaf blade
4	Leaf-blade pubescent	LBP	1. Glabrous, 2. Intermediate, 3. Pubescent
5	Leaf-blade color	LBC	1. Pale green, 2. Green, 3. Dark green, 4. Purple tips, 5. Purple margins, 6. Purple blotch, 7. Purple
6	Basal leaf sheath color	BLSC	1. Green, 2. Purple lines, 3. Light purple, 4. Purple
7	Leaf angle	LeafA	1. Erect, 2. Intermediate, 3. Horizontal, 4. Descending
8	Flag leaf angle	FleafA	1. Erect, 2. Intermediate, 3. Horizontal, 4. Descending
9	Ligule length (mm)	LiguleL	Measured at late vegetative stage
10	Ligule color	LiguleC	0. Absent 1. White, 2. Purple lines, 3. Purple
11	Collar color	CollorC	1. Pale green, 2. Green, 3. Purple
12	Auricle color	AuricalC	0. Absent 1. Pale green, 2. Purple
13	Days of heading	DaysofH	No. Of days from effective seeding to, 50% heading
14	Culm length (cm)	CulmL	From ground level to the base of the panicle
15	Culm number	CulmNo	Total no. of grain bearing and non-bearing tillers
16	Culm angle	CulmA	1. Erect, 3. Intermediate, 5. Open, 7. Spreading, 9. Procumbent
17	Internode color After Full Heading	INCAF	1. Green, 2. Light gold, 3. Purple lines, 4. Purple
18	Culm strength	CulmS	1. Strong, 3. Moderately strong, 5. Intermediate, 7. Weak, 9. Very week
19	Panicle length (cm)	PanicleL	From the base to the tip of the panicle
20	Panicle type	PanicleT	1. Compact, 5. Intermediate, 9. Open
21	Secondary branching	SecB	0. Abscent 1. Light, 2. Heavy, 3. Clustering
22	Panicle exertion	PanicleE	0. Well exertion, 3. Moderately, 5. Justexterted, 7. Partlyexserted, 9. Enclosed
23	Awning after full heading	AwnAFH	0. Absent 1. Short and partly awned, 5. Short and fully awned, 7. Long and partly awned, 9. Long and fully awned
24	Apicus color	ApiculC	1. White, 2. Straw, 3. Brown, 4. Red, 5. Red apex, 6. Purple, 7. Purple apex
25	Lemma and palea color	LemmaPC	0. Straw, 2. Gold, 3. Brown spot on straw, 4. Brown, 5. Reddish to light purple, 6. Purple spots on straw, 7. Purple, 8. Black, 9. White
26	Lemma and palea pubescence	LPP	1. Glabrous, 2. Hairs on lemma keel, 3. Hairs on upper portion, 4. Short hairs, 5. Long hairs
27	Sterile lemmacolor	SLC	1. Straw, 2. Gold, 3. Red, 4. Purple
28	Sterile lemma length	SLL	1. Short, 3. Medium, 5. Long, 7. Extra-long, 9. Asymmetrical
29	100 grain weight	Gw100	A random sample of 100 well-developed grains dried 13% moisture content
30	Seed coat color	SeedCC	1. White, 2. Light brown, 3. Speckled brown, 4. Brown, 5. Red, 6. Variable purple, 7. Purple
31	Ligule shape at late vegetative stage	LiguleS	0. absent, 1. acute to acuminate, 2. 2-cleft, 3. truncate
32	Leaf senescence	Leaf S	1. Late and low senescence-, 2 or more leaves retain their green color at maturity, 5. Intermediate, 9. Early and fast senescence- leaves are dead when the grains have become fully ripened
33	Awn color at maturity	AwnC	1. Straw, 2. Gold, 3. Brown, 4. Red, 5. Purple, 6. Black
34	Panicle threshability	PanicleT	1. difficult, 5. intermediate, 9. easy
35	Panicle axis at maturity	PanicleA	1. Straight, 2. Droopy
36	Panicle shattering	PanicleS	1. Very low, 3. Low, 5. Moderate, 7. Moderately high, 9. High

Table 2. Four labeled primers used for the Capillary electrophoresis.

Oligo name	Oligo sequence (5'-3')	Color of the Label Primer
5'- FAM- M13 (-21)	5'(FAM) TGT AAA ACG ACG GCC AGT 3'	Blue
5'- NED- M13 (-21)	5'(NED) TGT AAA ACG ACG GCC AGT 3'	Yellow
5'- PET- M13 (-21)	5'(PET) TGT AAA ACG ACG GCC AGT 3'	Red
5'- VIC- M13 (-21)	5'(VIC) TGT AAA ACG ACG GCC AGT 3'	Green

$$Eq = 1/N \sum_{i=1}^N (X_i - mc) \quad (1)$$

This error evaluates the fitting of the neural map to the data. Thus, the optimal map is expected to yield the smallest average quantization error. The smaller the quantization error, the smaller the average of the distance from the vector data to the prototypes, and that means, that the data vectors are closer to their prototypes.

An alternative approach for measuring topology preservation is to use input samples to determine how continuous the mapping from the input space to the map grid is. One of the most extended indices for this purpose is the topographic error (Kiviluoto 1996). It is also one of the errors proposed by Kohonen (Kohonen 2001). This error measures the proportion of all data vectors for which first and second best-matching units (BMU) are not adjacent vectors. Therefore, the lower the topographic error (Et) is, the better the SOM preserves the topology.

$$Et = 1/N \sum_{i=1}^N U(X_i) \quad (2)$$

The topographic error is calculated as shown above, where the equation 2. $U(x_i) = 1$, if the 1st BMN and 2nd BMN are not adjacent, otherwise $U(x_i) = 0$. Separate SOMs for each set of variable *i. e.*, morphological and molecular data were developed (SOMmor and SOMmol) (Uriarte and Martin 2005). The best SOM was chosen based on the error performance of quantization error (Eq) and topographic error (Et).

RESULTS AND DISCUSSION

There was a variation of agro-morphological characters (parametric only) across different WR eco-types in different eco-climatic zones in Sri Lanka. Analysis of variance of parametric variables used in construction of SOM for the morphological characters interactions with samples; climatic zones are shown in Table 3. According to Table 3, two (panicle length and 100-grain weight) out of eight parametric variables were not statistically significant ($p < 0.05$).

Similarly, χ^2 test of non-parametric variables shown in Table 4 indicated that two out of 28 variables such as leaf blade pubescence and culm number were not different across the rice populations. The summary of analysis of variance of molecular data of rice populations is shown in Table 5. According to the table, the markers RM11A1, RM21A1, RM21A2, RM44A1, RM44A2 and RM280 A2 were significantly varied across the rice populations (Table 5).

Different rice populations resulted from SOM analysis of morphological variables were grouped into labeled as Weedy rice populations in the dry zone - D1W, D2W, D3W, D4W and D5W; weedy rice populations in the intermediate zone - I1W and I2W; and weedy rice populations in the wet zone - WW1 and WW2. The group

of cultivated rice resulting in the SOMmor, were also indicated by cultivated rice population in the dry zone – DC1, DC2, DC3; and cultivated rice population in the intermediate zone I1C and I2C; cultivated rice population in the wet zone – WC (*unpublished data*).

The analysis of variance performed on quantitative morphological data of SOM groups is shown in Table 6. Except, 100-grain weight, the rest of the variables were significantly between the weedy rice populations ($p < 0.05$). Similarly, all the qualitative characters of rice populations differed significantly between SOM groups of WR eco-type across the different climatic zones, and association between the climate zone and non-parametric variables included in the study (data not shown).

The resulting SOM grouping patterns of WR eco-type populations indicate that SOM can differentiate WR eco-type populations in different climatic zones of the country. The qualitative and quantitative variables were used to develop two SOM models, SOMmor and SOMmol for morphological and molecular data respectively. The error performance, Et and Eq of best SOMmor developed in the study were 0.005 and 0.580 respectively. The labels in SOM matrices (Figure 2) show the “position” of the data samples in the SOM plane. The component planes of agro-morphological characters (qualitative and quantitative) (Figures not shown) characters such as seedling height, leaf blade width, leaf blade length, culm strength, panicle shattering, seed coat color and leaf angle emerge as salient characters which distinguish weedy rice populations in the different climatic zones of the country.

The grouping pattern of WR eco-types in the label matrix (Figure 2) revealed a grouping trend associated with eco-climatic conditions of the country. Further, there was an association of WR eco-types such as AKW2, APW2, KIW2, ADW etc., which were closely related to *O. nivara*, a wild rice species commonly found in the dry zone of the country suggesting that *O. nivara* is potentially hybridized with weedy rice eco-types.

The results obtained in SOMmol (Eq = 0.005 and Et = 0.331) derived from the molecular data (Tamayo et al. 1999) of the rice populations were indicated in Figure 3. The comparison of component planes showed (Figures not given), indicated four primer pairs that reflect considerable variation among the populations of WR eco-types. Among the primer pairs used in the study, RM 11 A1, RM 21 A1, RM 21 A2, RM 14 A1 and RM 280A2 showed capability of identifying populations of WR eco-type along the climatic gradient of the country. The grouping pattern of rice populations produced by SOMmol is shown in Figure 3. (Label matrix). A total of three groups (dry, wet, and intermediate rice populations) were represented in the label matrix. WR eco-types in the dry zone were closely associated with *O. nivara* and wet zone WR eco-types are grouped with *O. rufipogon* reflecting their close relationships. Application of SOM with molecular data produced clear SOMmol groups than that of the SOMmor developed from the agro-morphological characters.

Table 3. The effect of parametric variables used in constructing SOMs.

Variable	Source	df	Mean square	F ratio	Sig. (p)
SDLH	Climate	2	1145.604	24.066	s
	Samples	2	2682.572	56.354	s
	Climate*Samples	3	158.100	3.321	s
LBL	Climate	2	248.482	4.100	s
	Samples	2	6552.096	108.112	s
	Climate*Samples	3	939.210	15.497	s
LBW	Climate	2	284.242	28.939	s
	Samples	2	13.879	1.413	ns
	Climate*Samples	3	40.505	4.124	s
PanilcL	Climate	2	1195.631	52.628	s
	Samples	2	961.600	42.327	s
	Climate*Samples	3	29.202	1.285	ns
GW100	Climate	2	.324	.352	ns
	Samples	2	3.789	4.119	s
	Climate*Samples	3	.652	.709	ns
LiguleL	Climate	2	125.505	5.232	s
	Samples	2	155.722	6.491	s
	Climate*Samples	3	301.821	12.582	s
DaysofH	Climate	2	2501.815	53.313	s
	Samples	2	1317.238	28.070	s
	Climate*Samples	3	1387.490	29.567	s
CulmL	Climate	2	2342.816	47.842	s
	Samples	2	1285.035	26.241	s
	Climate*Samples	3	539.453	11.016	s

Note: s: significant at $p \leq 0.05$, ns: not significant $p \geq 0.05$

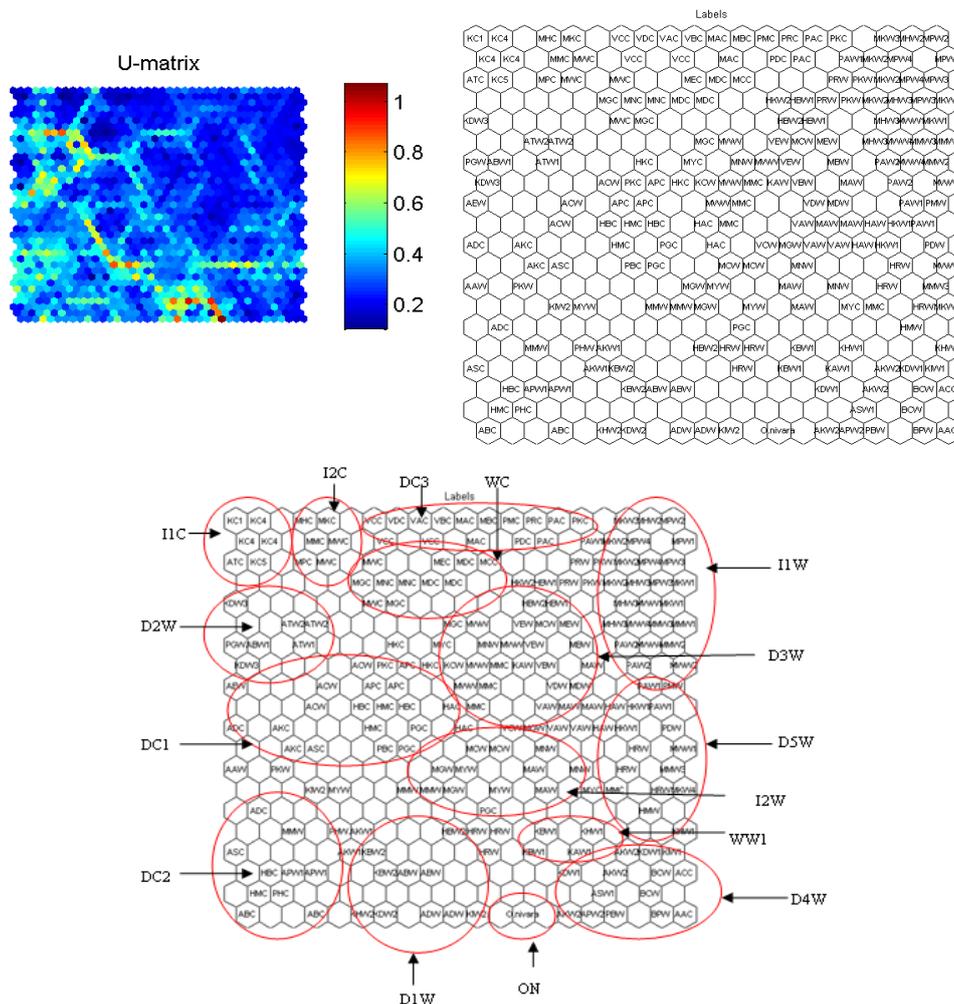


Figure 2. U matrix and the label matrix of the SOM developed from the agro-morphological characters.

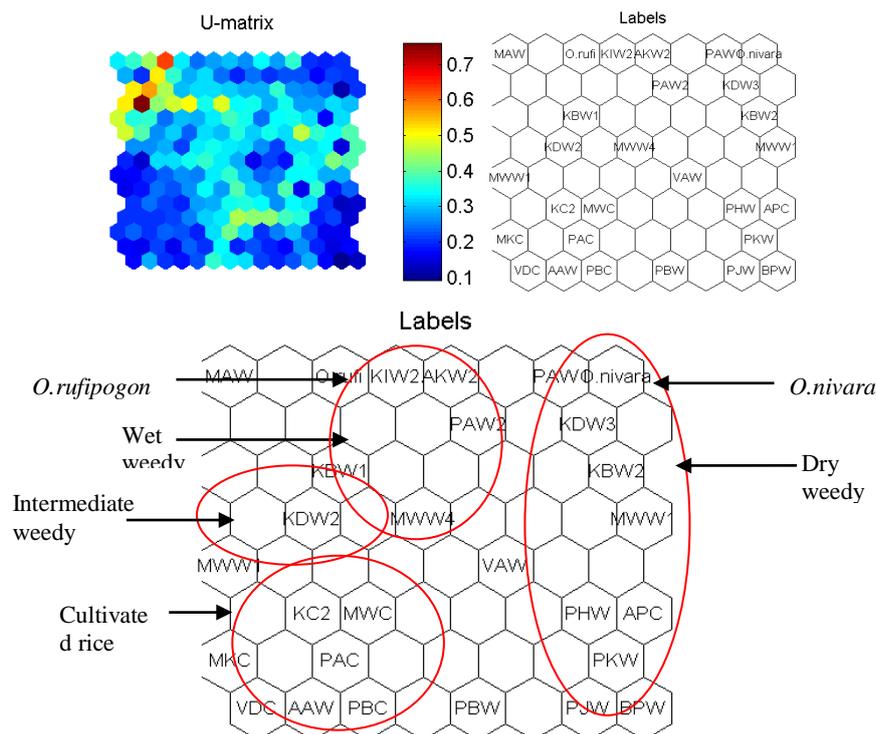


Figure 3. U matrix and label matrix of SOM developed for molecular data.

Table 4. The effect of climatic zone on qualitative morphological features of rice populations (CLIM).

Variable	χ^2 value	Df	Probability
BLSC*CLIM	43.849	6	s
INCAF * CLIM	114.251	6	s
LBC * CLIM	180.398	10	s
LBP * CLIM	8.175	4	ns
PanicleA * CLIM	22.790	4	s
PanicleE * CLIM	65.701	8	s
PanicleT * CLIM	50.088	4	s
SecB * CLIM	26.097	6	s
PanicleS * CLIM	119.253	12	s
PanicleT * CLIM	143.788	14	s
AwnAFH * CLIM	132.786	8	s
AwnC * CLIM	124.642	10	s
ApiculC * CLIM	52.431	4	s
SeedCC * CLIM	17.102	6	s
Leaf S * CLIM	191.292	8	s
LemmaPC * CLIM	82.196	14	s
LPP * CLIM	47.384	4	s
SLC * CLIM	197.225	4	s
SLL * CLIM	126.854	4	s
LeafA * CLIM	32.154	6	s
FLeafA * CLIM	232.013	10	s
LiguleC * CLIM	36.778	2	s
LiguleS * CLIM	32.467	4	s
CollorC * CLIM	88.948	4	s
AuricalC * CLIM	45.764	2	s
CulmNo * CLIM	19.738	16	ns
CulmA * CLIM	141.045	10	s

Note: s: significant at $p \leq 0.05$, ns: not significant $p \geq 0.05$

The application of SOM analysis in grouping WR eco-types, with respect to eco-climatic zones of Sri Lanka, was carried out using agro-morphological and molecular data. The relationships among the agro-morphological characters (36 characters) and molecular data (derived from 10 primer pairs) were examined and generated hypotheses and tested. The SOM is supportive of other statistical analyses such as Principle Component Analyses, Cluster by complementing ability to examine the relationships between different types of variables in a visual presentation of the data (Mele and Crowley 2008). Visualization of the component planes derived from the SOM (SOMmor and SOMmol) indicates that there are patterns within the sets of variables. The SOMmor model developed indicated that seedling height, leaf blade width, leaf blade length, culm strength, panicle shattering, seed coat color, and leaf angle play an important role in grouping WR eco-type populations in Sri Lanka (Figure 3).

Meanwhile, the molecular banding patterns of primers such as RM 11 A1, RM 21 A1, RM 21 A2, RM 14 A1, and RM 280A2 revealed that they are important in grouping different populations of WR eco-types in Sri Lanka (Figures not given).

Table 5. The effect of variance of molecular data .

Marker	Sum of squares	df	Mean square	F ratio	Sig.
RM11 A1	16828. 725	2	8414. 363	3. 715	s
RM11 A2	593. 789	2	296. 895	0. 505	ns
RM21 A1	38527. 956	2	19263. 978	7. 727	s
RM21 A2	36618. 273	2	18309. 136	8. 884	s
RM14 A1	4731. 476	2	2365. 738	1. 054	ns
RM14 A2	7185. 571	2	3592. 786	2. 06	ns
RM44 A1	9823. 883	2	4911. 942	9. 322	s
RM44 A2	15850. 684	2	7925. 342	29. 516	s
RM84 A1	0. 265	2	0. 132	0. 002	ns
RM84 A2	30. 333	2	15. 167	0. 389	ns
RM167 A1	3576. 074	2	1788. 037	0. 926	ns
RM167 A2	3620. 583	2	1810. 292	0. 918	ns
RM205 A1	67. 681	2	33. 841	2. 417	ns
RM205 A2	46. 718	2	23. 359	1. 967	ns
RM211 A1	4943. 52	2	2471. 76	0. 457	ns
RM211 A2	5749. 015	2	2874. 507	0. 525	ns
RM280 A1	4666. 254	2	2333. 127	0. 535	ns
RM280 A2	4090. 331	2	2045. 165	0. 555	s
RM332 A1	148. 989	2	74. 495	0. 18	ns
RM332 A2	103. 283	2	51. 641	0. 009	ns

Note: s= significant at $p \leq 0. 05$, and ns = not significant $p \geq 0. 05$

Table 6. The effect of variance among the group of cultivated, wild rice species and weedy rice eco-types in Sri Lanka.

Variable	df	Mean square	F ratio	Sig.
Sdlh	15	933. 967	22. 146	s
LBL	15	1259. 932	27. 36	s
LBW	15	128. 407	15. 866	s
Gw100	15	1. 167	1. 119	ns
LiguleL	15	245. 899	11. 869	s
DaysofH	15	800. 506	20. 592	s
CulmL	15	1177. 008	40. 666	s

Note: s: significant at $p \leq 0. 05$, ns: not significant $p \geq 0. 05$)

The result of this preliminary study revealed that unsupervised artificial neural networks such as SOM are important tools in solving complicated biological problems associated with classification WR ecotypes with multivariate set of data. Further, results of the study led to hypothesis that the patterns reflected in the distribution of WR population are related to the varying climatic condition of the country. It appears that each WR eco-type represented each eco-climatic zone, especially closely associated with respective wild rice species. Further, WR eco-types of the dry and intermediate zones are closely related to *O. nivara*, commonly found in the country's dry zone. This observation suggests that the *O. nivara* possesses a higher potential of hybridization, with cultivated rice varieties and weedy rice eco-types in the dry and intermediate zones of Sri Lanka which is supported by the findings of Chen (Chen et al. 2004).

In conclusion, the finding of the study led to conclude that SOM is an important tool in tracing the patterns of variation of agro-morphological characters and molecular characters of WR eco-types with respect to the eco-climatic zones of the country. SOM can visualize the patterns of WR eco-type population distribution across different climatic conditions in a low-dimensional space. The characters such as seedling height, leaf blade width, leaf blade length, culm strength, panicle shattering, seed coat color, and leaf angle are the salient agro-morphological characters in identification of WR eco-type distributed in different eco-climatic zones of Sri Lanka. The primers such as RM 11, RM 21, RM 14, and RM 280 are of importance in delimitation of different WR populations in varying climatic conditions. The dimension reduction of input variables before presenting variables to SOM conserves the cost, labor, and computer processing time. The finding supports the idea capability of SOM in delimitation of WR populations through unsupervised artificial neural networks with limited multiple input variables. The findings of SOM analysis of the study's morphological data in the long run help develop intelligence systems to control weedy rice using areal targeting such as drones.

ACKNOWLEDGEMENTS

National Science Foundation in Sri Lanka is greatly appreciated for providing a research grant. A.S.K. Abesekara, RRDI, Bathalagoda, Sri Lanka and Ms. K.D.K. Karunarathna are also greatly acknowledged for their valuable support during the research.

REFERENCES

- Abeysekara AS, Nugaliyadda L, Herath HM, Wickrame UB, Iqbal YB. 2010. Weedy Rice: a threat to direct-seeded rice cultivation in Sri Lanka. In: Rice Congress 2010, PGRC, Gannoruwa.
- Agarwal P, Skupin A. Self-Organising Maps: Applications in Geographic Information Science. John Wiley & Sons, New York.
- Azmi M, Karim SM. 2008. Weedy Rice: Biology, Ecology and Management. Malaysian Agricultural Research and Development Institute (MARDI), Kuala Lumpur, Malaysia.
- Brosse S, Giraudel JL, Lek S. 2001. Utilization of non-supervised neural networks and principal component analysis to study fish assemblages. *Ecol Model* 146 (1-3): 159-166. DOI: 10.1016/S0304-3800(01)00303-9.
- Cao Q, Lu BR, Xia HU, Rong J, Sala F, Spada A, Grassi F. 2006. Genetic diversity and origin of weedy rice (*Oryza sativa* f. *spontanea*) populations found in north-eastern China revealed by simple sequence repeat (SSR) markers. *Ann Bot* 98: 1241-1252. DOI: 10.1093/aob/mcl210.
- Chauhan BS, Johnson DE. 2010. Weedy rice (*Oryza sativa* L.). I. Grain characteristics and growth response to competition of weedy rice variants of five Asian countries. *Weed Sci* 58: 374-380. DOI: 10.1614/WS-D-09-00071.1.
- Chauhan BS. 2012. Weedy rice (*Oryza sativa*). II. Response of weedy rice to seed burial and flooding depth. *Weed Sci* 60: 385-388. DOI: 10.1614/WS-D-11-00213.1.
- Chen LJ, Lee DS, Song ZP, Suh HS, LU BR. 2004. Gene Flow from Cultivated Rice (*Oryza sativa*) to its Weedy and Wild Relatives. *Annals of Botany* 93: 67-73. DOI: 10.1093/aob/mch006.
- Cho YC, Chung TY, Suh HS. 1995. Genetic characteristics of Korean weedy rice (*Oryza sativa* L.) by RFLP analysis. *Euphytica* 86: 103-110. DOI: 10.1007/BF00022015.
- Dallwitz MJ. 1980. A general system for coding taxonomic descriptions. *Taxon* 29: 41-46. DOI: 10.2307/1219595.
- Deboeck G, Kohonen T. 1998. Visual explorations in finance with self-organizing maps. Springer-Verlag, London.
- Federici MT, Vaughan D, Norihiko T, Kaga A, Xin WW, Koji D, Francis M, Zorrilla G, Saldain N. 2001. Analysis of Uruguayan weedy rice genetic diversity using AFLP molecular markers. *Electron J Biotechnol* 4 (3): 5-6.
- Fernández C, Soria E, Martin JD, Serrano AJ. 2006. Neural Networks for animal science applications: Two case studies. *Expert Syst Appl* 31: 444-450. DOI: 10.1016/j.eswa.2005.09.086.
- Gwag JG, Dixit A, Park YJ, Ma KH, Kwon SJ, Cho GT, Lee GA, Lee SY, Kang HK, Lee SH. 2010. Assessment of genetic diversity and population structure in mungbean. *Genes Genomics* 32 (4): 299-308. DOI: 10.1007/s13258-010-0014-9.
- He Z, Jiang X, Ratnasekera D, Grassi F, Perera U, Lu BR. 2014. Seed-mediated gene flow promotes genetic diversity of weedy rice within populations: implications for weed management. *PLoS ONE* 9: e112778. DOI: 10.1371/journal.pone.0112778.
- Hemming J, Rath T. 2001. Computer-vision-based weed identification under field conditions using controlled lighting. *J Agric Eng Res* 78 (3): 233-243. DOI: 10.1006/jaer.2000.0639.
- Karunaratna KDK, Weerakoon SR, Somaratne S, Weerasena OVDSJ. 2014. Molecular and Agro-morphological affinities among weedy, wild and cultivated rice varieties in different climatic zones of Sri Lanka. *OUSL Ann Acad Sessions* 2014: 319-323.
- Kaski S. 1996. Exploratory data analysis by the self-organizing map: Structure of welfare and poverty in the world. In: *Neural Networks in Financial Engineering*. Proceedings of the Third International Conference on Neural Networks in the Capital Markets, Singapore, 498-507.
- Kohonen T. 1982. Self-organized formation of topologically correct feature maps. *Biol Cybern* 43: 59-69. DOI: 10.1007/BF00337288.
- Kohonen T. 1998. The self-organizing map. *Proc IEEE* 78 (9): 1464-1480.
- Kohonen T. 2001. Software Tools for SOM. In *Self-organizing maps*. Springer, Berlin, Heidelberg. DOI: 10.1109/5.58325.
- Kohonen T. 2002. *Self-organizing Maps*. Springer Series in Information Sciences, V. 30, Springer, Berlin.
- Kiviluoto K. 1996. Topology preservation in self-organizing maps. *Proc Int Conf Neural Netw* 1: 294-299. DOI: 10.1109/ICNN.1996.548907
- Lek S, Gue gan JF. 2000. Artificial Neuronal Networks: Application to Ecology and Evolution. Springer, Berlin.
- Li ST. 2002. A web-aware interoperable data mining system. *Expert Syst Appl* 22 (2): 135-146. DOI: 10.1016/S0957-4174(01)00050-1.
- Londo JP, Schaal BA. 2007. Origins and population genetics of weedy red rice in the USA. *Mol Ecol* 16: 4523-4535. DOI: 10.1111/j.1365-294X.2007.03489.x.
- Ma DR, Li MB, Wang N, Chen WF. 2008. Genetic diversity and population differentiation of weedy rice in Liaoning province of China. *Acta Agron Sin* 34: 403-411. [Chinese]
- Marambe B, Amarasinghe L. 2000. Weedy rice in Sri Lanka. In: Baki BB, Chin DV, Motimer AM (eds.) *Wild and Weedy Rice in Rice Ecosystems in Asia - A Review*. Intl Rice Res Inst Philippines 79-82.
- McKhann HI, Camilleri C, Bérard A, Bataillon T, David JL, Reboud X, Le Corre V, Caloustian C, Gut IG, Brunel D. 2004. Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J* 38 (1): 193-202.
- Mele PM, Crowley DE. 2008. Application of self-organizing maps for assessing soil biological quality. *Agric Ecosyst Environ* 126: 139-52. DOI: 10.1016/j.agee.2007.12.008
- Mortimer M, Pandey S, Piggin CI2000. Weedy rice approaches to ecological appraisal and implications for research priorities. In: Baki BB, Chin DV, Mortimer M (eds) *Proceedings of Wild and Weedy Rice in Rice Ecosystems in Asia*. A review, Los Banos, Philippines: Intl Rice Res Inst 97-105.
- Moshou D, Vrindts E, Ketelaere B, Baerdemaeker J, Ramon H. 2001. A neural network based plant classifier. *Comput Electron Agric* 31 (1): 15-16. DOI: 10.1016/S0168-1699(00)00170-8.
- Oja M, Kaski S, Kohonen T. 2002. Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum. *Neural Comput Surv* 3: 1-156.
- PGRC. 1999. Characterization Catalogue of Rice (*Oryza sativa*) Department of Agriculture. Ministry of Agriculture and Lands, Sri Lanka.
- Schreer JF, O'Hara Hine, RJ, Kovacs KM. 1998. Classification of dive profiles: A comparison of statistical clustering techniques and unsupervised artificial neural networks. *J Agric Biol Environ Stat* 3: 383-404. DOI: 10.2307/1400572.
- Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol* 18 (2): 233-234. DOI: 10.1038/72708.
- Simon H. 1999. *Neural networks: a comprehensive foundation*. Prentice-Hall, New York.
- Suh HS, Sato YI, Morishima H. 1997. Genetic characterization of weedy rice (*Oryza sativa* L.) based on morpho-physiology, isozymes and RAPD markers. *Theor Appl Genet* 94: 316-321. DOI: 10.1007/s001220050417.
- Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR. 1999. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 96: 2907-2912. DOI: 10.1073/pnas.96.6.2907.
- Upadhyaya HD, Gowda CL, Pundir RP, Reddy VG, Singh S. 2006. Development of core subset of finger millet germplasm using geographical origin and data on 14 quantitative traits. *Genet Resour Crop Evol* 53 (4): 679-685. DOI: 10.1007/s10722-004-3228-3.
- Uriarte EA, Martín FD. 2005. Topology Preservation in SOM. *Int J Appl Math Comput Sci* 1 (1): 19-22. DOI: 10.1.1.193.5889.
- Vesanto J, Alhoniemi E. 2000. Clustering of the Self-Organizing Map. *IEEE on Neural Netw* 11: 586-600. DOI: 10.1109/72.846731.
- Vesanto J. 2000. *Using SOM in Data Mining*. [Thesis]. Licentiate of Science in Technology. Espoo, Finland.
- Walley WJ, Martin RW, O'Connor MA. 2000. Self-organizing maps for the classification of river quality from biological and environmental data. In: Denzer R, Swayne DA, Purvis M, Schimak G (eds) *Environmental Software Systems: Environmental Information and Decision Support*. Kluwer Academic, Dordrecht.
- Wang JS, Li SX, Gao J. 2014. SOM neural network fault diagnosis method of polymerization kettle equipment optimized by improved PSO algorithm. *Sci World J* 2014: 937680. DOI: 10.1155/2014/937680.
- Zhang J, Burgos NR, Kun MA, Zhou YJ, Geng RM, Yu LQ. 2008. Genetic diversity and relationship of weedy rice in Taizhou city, Jiangsu Province, China. *Rice Sci* 15: 295-302. DOI: 10.1016/S1672-6308(09)60007-X.
- Zhang P, Li J, Li X, Liu X, Zhao X, Lu Y. 2011. Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. *PLoS ONE* 6 (12): 27565. DOI: 10.1371/journal.pone.0027565.