available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/compag

Photometric clustering of regenerated plants of gladiolus by neural networks and its biological validation

*V.S.S. Prasad, S. Dutta Gupta**

Department of Agricultural and Food Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, India

ARTICLE INFO

Article history:
Received 29 July 2005
Received in revised form
6 February 2007
Accepted 21 May 2007

Keywords:
Adaptive resonance theory
Fuzzy ART
Cluster analysis
Machine vision
Organogenic potential

ABSTRACT

Photometric clustering of regenerated plants of gladiolus was described using fuzzy adaptive resonance theory (ART) and the resultant grouping pattern was compared with ART 2, and self-organizing map (SOM) neural network modules. Classical clustering techniques such as hierarchical (HC) and *k*-means clustering (KM) were also applied to analyze the same data set to evaluate the performance of the artificial neural network (ANN)-based clustering. Regenerated plants were clustered into two groups in varying numbers by ART 2, SOM, HC and KM. With ART 2, 19 of 55 plants were sorted into group '0' and the remaining 36 plants were placed in group '1', whereas; SOM distributed the regenerated plants in the ratio of 28:27. The clustering ratios of HC and KM were 34:21 and 26:29, respectively. However, a refined clustering of regenerated plants into seven groups was observed with Fuzzy ART. There was a similarity in the number of generated clusters between the training and validation data sets indicating the network efficiency. Biological validation of photometric clustering of regenerated plants was also assessed by indexing the corm induction potential of the sorted groups. A significant difference in corm induction potential between the groups was noted only with ART 2. Fuzzy ART-assisted grouping patterns are not conducive to segregate the potential corm producing shoots. ART 2-aided clustering of the regenerated plants appeared to be more promising for selecting group of plants capable of corm development than did other clustering approaches.

1. Introduction

Micropropagation through tissue culture techniques is widely used for multiplication of elite plant species to produce genetically identical plants. However, one of the major bottlenecks in commercialization of micropropagation is the poor survival of regenerated plants upon *ex vitro* transfer. *In vitro* environmental factors such as gradients in humidity and/or CO_2 concentration, differential distribution of light intensity and air temperature inside the culture vessel has significant impact on growth and quality of the regenerated plants during micropropagation (Ibaraki, 2006). Such inconsistencies in the cultural environment, even under controlled conditions, might cause variations in the regenerated plants which affect uniformity in plantlet quality. Regenerated plants may differ in their *in vitro* behavioral aspects, viz., root and storage organ (such as corm) development ability, hyperhydric status, and adaptability to *ex vitro* conditions. These kinds of variation are not commensurate with that of well-documented aspects of somaclonal variation (Larkin and Scowcroft, 1981) and deserve attention for studies to characterize the *in vitro* and/or *ex vitro* behavioral aspects of the plantlets.

Development of an automatic decision-making entity reflecting the variations of *in vitro* regenerated plants is

* *Corresponding author*. Tel.: +91 3222 283114; fax: +91 3222 255303.
E-mail address: sdg@agfe.iitkgp.ernet.in (S. Dutta Gupta).
0168-1699/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2007.05.006

necessary to ensure high rate of success in micropropagation. The decision-making may be made in the form of grouping or clustering of the regenerated plants based on their photometric properties, and correlating the ability of the sorted groups with respect to organogenic potential and *ex vitro* survival. Since the physiological and behavioral variations among the regenerated plants are difficult to be resolved by human visual evaluation, machine vision-coupled neural network-based clustering might be an efficient alternative to select plants or groups of plants with organogenic potential and high percent of *ex vitro* survival.

The main objective of any clustering model would be to find a valid organization of the data with respect to the inherent structure and relations among the inputs. Classical clustering methods, e.g. hierarchical clustering (HC) (Eisen et al., 1998), *k*-means algorithm (KM) (Hartigan and Wong, 1979), and self-organizing maps (SOM) (Tamayo et al., 1999), have been restricted in classifying human cancerous cell lines and gene expression analysis of baker's yeast with considerable pitfalls (Eisen et al., 1998; Tamayo et al., 1999; Ross et al., 2000). The ANN-based modeling approach has been found to be more flexible, effective and versatile in dealing with non-linear relationships prevalent in plant cell culture practices (Albiol et al., 1995). In plant tissue culture systems, artificial neural networks (ANN) have been used for pattern recognition of somatic embryos, growth evaluation, estimation of shoot length, online estimation of biomass and photometric assessment of regenerated plants (Prasad and Dutta Gupta, 2006). In the field of neural networks, the adaptive resonance theory (ART) was introduced by Grossberg (1976) as a theory of human cognitive information processing. ART is a kind of self-organized clustering, which clusters a given set of input patterns into some groups in an unsupervised manner. The most distinctive feature of an ART model lies in its ability to adapt to new input patterns while maintaining the temporal stability of stored patterns. As a member of ART networks, ART 2 (Carpenter and Grossberg, 1987a) was designed to accept and process analog or binary vectors in contrast to ART 1 (Carpenter and Grossberg, 1987b) which exclusively deal with binary forms of inputs. We have successfully demonstrated the ART 2 neural network-aided image processing method for photometric clustering of regenerated plants of gladiolus based on the trichromatic (RGB) features of leaves (Mahendra et al., 2004). However, as the photometric variations are highly inconsistent with an exhaustive range of overlapping, a more discrete approach is required to channel such variable patterns into stable recognition categories. The fuzzy set theory and complement coding incorporated in Fuzzy ART neural networks may allow refinement in grouping pattern. Fuzzy ART has been introduced by Carpenter et al. (1991) for rapid stable learning of recognition categories in response to analog or binary input patterns. Properties of ART networks depend on two main parameters, vigilance (ρ) and learning rate (β). The vigilance parameter (VP) defines the minimum similarity between patterns in one cluster. The resulting number of clustered groups depends on the similarity patterns of all input patterns. The incorporation of the VP value that controls the grouping pattern within the inputs provides leverage to the user to decide as to how many reasonable and logical groups are expected from autonomous and adaptive clustering of the ART network. A learning procedure adjusted the weight vector W_j of cluster j, the vector represents the pattern of cluster j.

In the present work, we describe the photometric clustering of regenerated plants of gladiolus by Fuzzy ART. To compare the efficiency of the clustering results, we also applied ART 2, SOM, HC, and KM to the same data. Biological validation of the clustering of regenerated plants into groups has been assessed by indexing the *in vitro* corm development potential of the respective groups.

2. Materials and methods

2.1. Plant regeneration and in vitro corm induction

The primary leaves of the sprouted corms of *Gladiolus hybridus* Hort. were surface disinfected with 0.1% $HgCl_2$ followed by three to four rinses in sterile double distilled water. The basal meristematic portions of the innermost leaves were dissected, blotted dry and inoculated on Murashige and Skoog medium (MS; Murashige and Skoog, 1962) containing 2.0 mg/l α-naphthaleneacetic acid (NAA), 3% (w/v) sucrose and 0.8% agar. The leaf-derived calluses were then transferred onto MS medium supplemented with 0.2 mg/l NAA and 2.0 mg/l 6-benzyladenine (BAP) for the induction of meristematic bud clusters as described previously (Dutta Gupta and Datta, 2004). The differentiated multiple shoot clusters were transferred onto basal MS medium devoid of growth regulators in five GA-7 vessels (Osmotek, Israel) and incubated for a period of 2 weeks. There were 5 clusters per vessel, each with 2–5 shoot buds. The training set comprised of 25 leaf images, each having its origin from a regenerated plant per cluster as detailed in Mahendra et al. (2004).

For testing as well as biological validation (hereafter referred to as validation set) a total of 55 regenerated plants were sampled randomly. The outermost expanded leaves of the regenerated plants were excised and used for photometric feature extraction. Digital images of 1–55 regenerated leaves belonging to 1–55 shoots were acquired and numbered from R1 to R55. The corm induction potential of R1–R55 regenerated plants was monitored. For corm induction, the corresponding regenerated plants without the outermost leaves were individually inoculated in culture tubes (25 mm × 150 mm) with 20 ml of MS medium supplemented with 0.5 mg/l NAA, 6% sucrose and 0.8% agar. The cultures were incubated for 90 d. Perimeters, fresh weights and dry weights of the developed corms were measured. The pH of all the media was adjusted to 5.6 before autoclaving for 15 min at 121 °C and 106 kPa. All cultures were kept at 16 h photoperiod (irradiance of 50 $\mu mol\ m^{-2}\ s^{-1}$), temperature of 25 °C and relative humidity of 55%.

2.2. Extraction of photometric parameters of leaves

Photometric data that included the mean brightness, grey-scale level for the maximum pixel count and maximum pixel count in the luminosity and trichromatic (red, green and blue) components were extracted from digitized leaf images as described by Mahendra et al. (2004). The outermost expanded leaves of the regenerated plants were scanned under constant luminosity and the digitized images were saved in Adobe Pho-

toshop; Adobe Systems Incorporated, USA (*.psd) format with 8 bits per pixel having 256 grey-scale levels. The pixel properties of the images were evaluated using Adobe Photoshop 7.0 software. Each leaf image represented a single, randomly selected regenerated plant per cluster.

2.3. *Cluster analysis of the photometric data*

For grouping the regenerated plants into cluster, the photometric data of the regenerated leaves procured from the digitized leaf images were subjected to adaptive resonance theory (ART 2), self-organizing map (SOM), hierarchical clustering (HC), *k*-means algorithm (KM) and fuzzy adaptive resonance theory (Fuzzy ART) cluster analyses as described below. Significant difference among the groups with regard to corm parameters was ascertained by *P* value obtained from t-tests of the two samples assuming heteroscedasticity. Component steps of photometric clustering of regenerated plants of gladiolus and its biological validation are presented in Fig. 1.

2.3.1. *ART 2- and SOM-based cluster analyses*

The ART 2 neural network trained with a set of 25 leaf input patterns as reported earlier (Mahendra et al., 2004), was utilized to cluster the data set of 55 leaf input patterns at a VP set at 0.999 to test and biologically validate the groups in terms of *in vitro* corm induction. Summary of the algorithm for ART 2 along with its exemplified numerical interpretation has been described in Mahendra et al. (2004). A bi-nodal self-organizing map (Kohonen neural network; Kohonen, 1997) was utilized to sort the validation set leaf input patterns whose corresponding shoots were meant to be indexed for corm inducibility. The mapping was performed on a MATLAB Ver. 7.0 platform. The various network parameters were set as follows: (a) input data normalization function: Prestd; (b) topology function: Hextop;

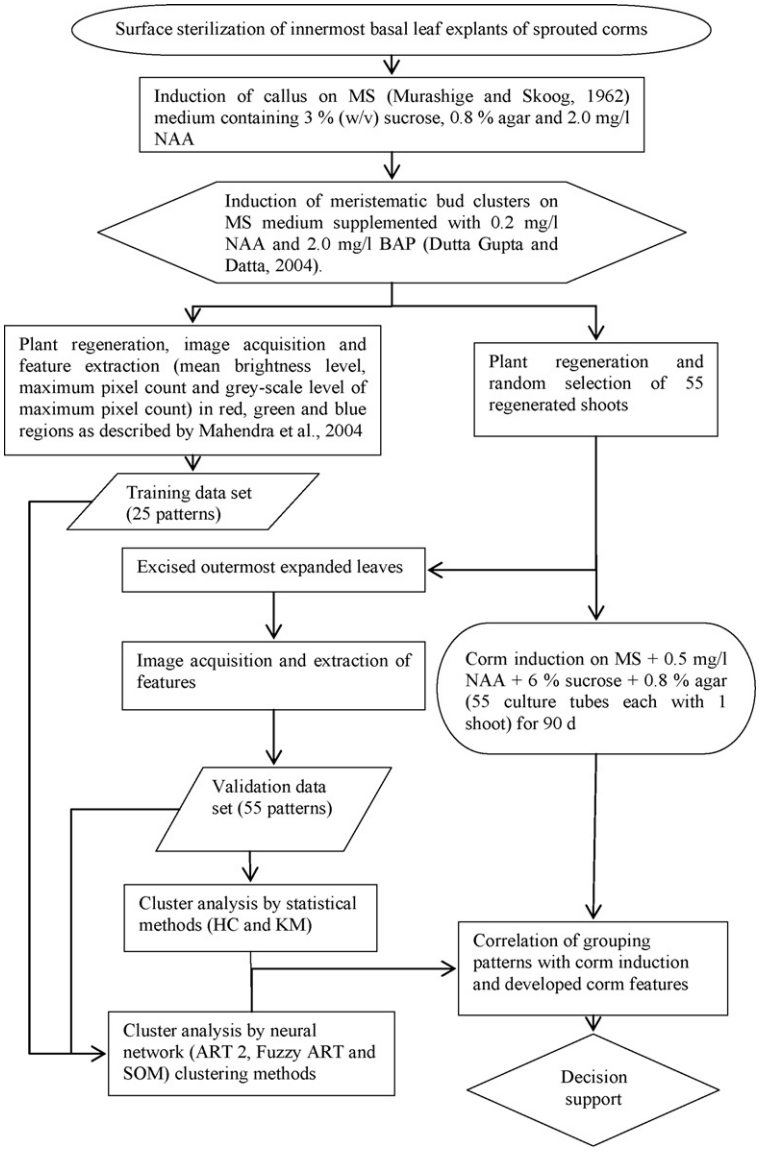

Fig. 1 – Step-wise illustration of photometric clustering of *in vitro* regenerated plants of gladiolus and its biological validation.

Table 1 – Number of leaf image parameters for each leaf input pattern

Leaf image attribute	Domain (luminosity + red + green + blue)	Total
Mean brightness level	1+1+1+1	4
Grey-scale level for the maximum pixel count	1+1+1+1	4
Maximum pixel count	1+1+1+1	4
Total no. of attributes/input pattern		12

(c) distance function: Linkdist; (d) ordering phase learning rate: 0.9; (e) epochs: 100; (f) tuning phase learning rate: 0.02; (g) neighbourhood distance: 1.0. The SOM network was initialized with the minimum and maximum values of the training data (Mahendra et al., 2004) of 25 leaf input patterns.

2.3.2. *HC- and KM-based cluster analyses*

Hierarchy of the progressively agglomerative clusters of validation-set leaf input patterns was constructed by calculating the Euclidean distance between the variables on STATISTICA Ver. 6.0 platform; StatSoft India Pvt. Ltd., New Delhi. The complete linkage method was used to assemble the patterns.

For *k*-means clustering, two cluster centroids were randomly generated. Then, the input patterns were assigned to the nearest cluster center. The new cluster centers were recalculated and the process was repeated until convergence (maximum inter-cluster variance and minimum intra-cluster variance) was reached. *k*-Means algorithm was run on STATISTICA Ver. 6.0 platform.

2.3.3. *Fuzzy ART neural network-based clustering analysis*

The training data set (Mahendra et al., 2004) and validation data sets represented by 25 and 55 leaf input patterns, respectively, having a string of values for 12 distinct properties (Table 1) were subjected to fuzzy adaptive resonance theory (Fuzzy ART) neural network-aided clustering analysis.

2.3.3.1. Data normalization. The values of mean brightness level, and grey-scale level for the maximum pixel count in the luminosity and trichromatic components of the leaf image pixels ranged from 0 to 255. The maximum pixel count ranged from 1 to 16. Each input datum representing a particular parameter for all the leaf images was scaled to a value lying between 0 and 1 using Eq. (1):

$$\text{Normalized value} = 0.5 \times \frac{(\text{Original value} - \text{mean value})}{(\text{Maximum value} - \text{minimum value})} + 0.5 \quad (1)$$

Such data pre-processing is done to ensure the uniform statistical distribution of each input and output value, and also to match the range of fuzzy neurons for efficient and fast functioning. The fuzzy set values of training and validation patterns are presented in Fig. 2(a) and (b), respectively.

As large numbers of analog input patterns disturbs the norm of weight vectors, complement coding was implemented to maintain the amplitude of inherent information:

$$\text{Complement of input pattern (Ic)} = 1 - \text{Original input pattern} \quad (2)$$

Subsequent to complement coding, input data were fed to the recognition system as 2×12 dimensional input vectors. An MS DOS-based executable 'C' program compilation of Fuzzy ART algorithm developed by Tomida et al. (2002) was utilized to perform the cluster analysis. The structural architecture of Fuzzy ART is depicted in Fig. 3. The dynamic computations in Fuzzy ART were determined by a choice of parameters α (value is assigned >0; presently 0.01), a learning rate parameter β (value is assigned >0 and <1; presently 0.1) and a vigilance parameter (value is assigned >0 and $\leq$1; presently 0.5–1.0).

2.3.3.2. Working principle of Fuzzy ART algorithm. The network is initialized by setting the connection weights (W_{ij}) to unity. When an input pattern is presented, choice function is calculated for every single node. A cluster category is established holding a maximum value for choice function (T_j):

$$T_j(I) = \frac{|I \wedge W_j|}{\alpha + |W_j|} \quad (3)$$

where I is the leaf input pattern/vector of 12 analogue values representing 12 photometric features, T_j the choice function, α the choice parameter, W_j the weight vector, and $\wedge$ is the fuzzy AND operator defined as: $(x \wedge y)_i = \min(x_i, y_i)$, where

$$|x| \equiv \sum x_i \quad (4)$$

The following example illustrates the clustering during the simulation of Fuzzy ART algorithm with validation set leaf input patterns:

$$\begin{aligned} I_{\text{leaf input pattern}} &= |0.89, 0.60, 0.73, 0.59, 0.58, 0.72, 0.66, 0.64, 0.73, 0.51, 0.49, 0.66|, \\ |W_{\text{example}}| &= |0.93, 0.65, 0.58, 0.63, 0.62, 0.72, 0.66, 0.65, 0.73, 0.39, 0.64, 0.52| = 7.76, \\ |I_{\text{leaf input pattern}} \wedge W_{\text{example}}| &= |0.89 \wedge 0.93 + 0.60 \wedge 0.65 + 0.73 \wedge 0.58 + 0.59 \wedge 0.63 \\ &\quad + 0.58 \wedge 0.62 + 0.72 \wedge 0.72 \\ &\quad + 0.66 \wedge 0.66 + 0.64 \wedge 0.65 + 0.73 \wedge 0.73 + 0.51 \wedge 0.39 + 0.49 \wedge 0.64 + 0.66 \wedge 0.52| \\ &= |0.89 + 0.60 + 0.58 + 0.59 + 0.58 + 0.72 + 0.66 + 0.64 + 0.73 + 0.39 + 0.49 + 0.52| = 7.45 \end{aligned} \quad (5)$$

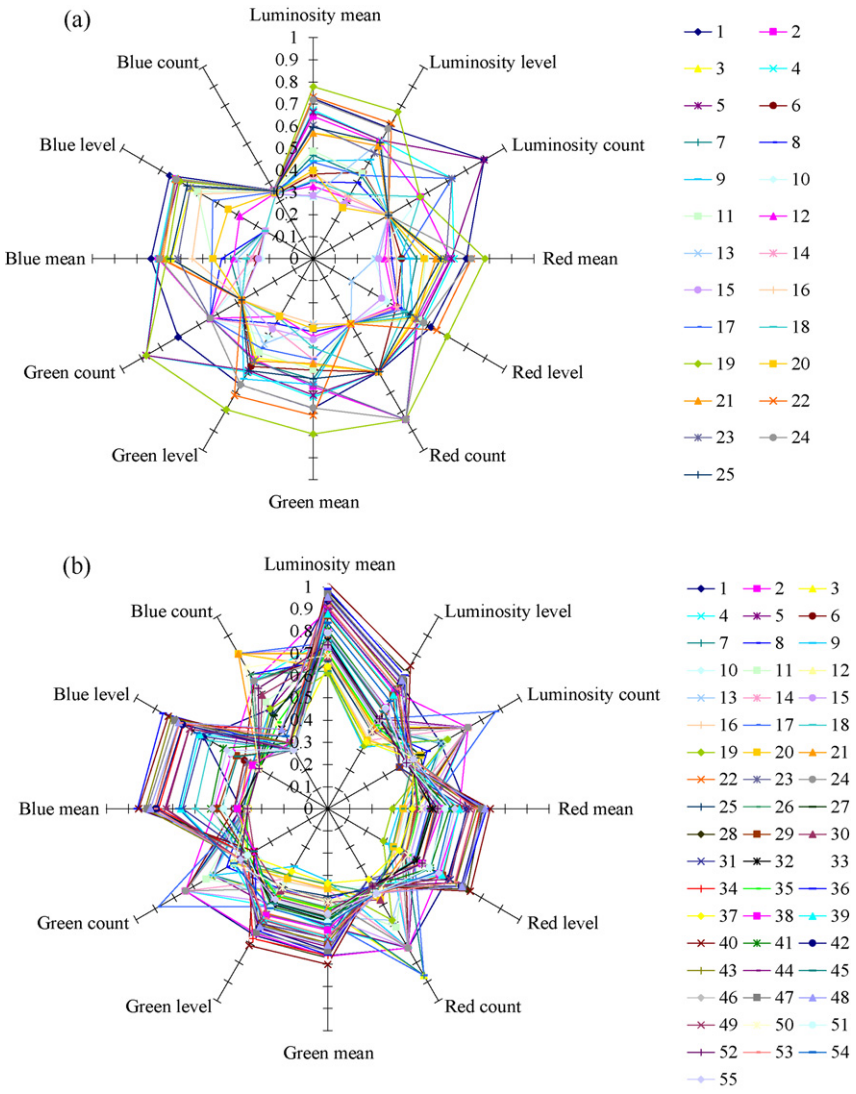

Fig. 2 – Fuzzified values of photometric data of (a) training and (b) validation data sets. (For interpretation of the references to color in the text, the reader is referred to the web version of the article.)

$$T_{\text{example}} = \frac{7.45}{0.01 + 7.76} = 0.95 \tag{6}$$

The maximal T_j is defined as the 'winner' cluster for input pattern. When more than one T_j is maximal, the output nodes become committed to cluster categories in the order of $j = 1, 2, 3, \ldots$

The match function is represented by Eq. (7):

$$M_j = \frac{|I \wedge W_j|}{|I|} \tag{7}$$

where M_j is the match function, I the input vector, and W_j is the weight vector.

If $M_j \geq \rho(\text{VP})$, then the resonance occurs and learning (weight vector update) takes place for the established category following Eq. (8):

$$W_j(\text{new}) = \beta[I \wedge W_j(\text{old})] + (1 - \beta)W_j(\text{old}) \tag{8}$$

where W_j is the weight vector, I the input vector, and β is the learning rate parameter.

On the contrary, if the match-function is less than the vigilance parameter value, then the mismatch reset occurs. Due to mismatch reset, a new cluster is generated that has the next maximal T_j value.

The weight vector of a new cluster is calculated as follows:

$$W_j(\text{new cluster}) = \beta\,[I \wedge W_j(\text{initial})] = I \tag{9}$$

where W_j is the weight vector, and β is the learning rate parameter.

When a new input pattern is subjected, calculations are repeated from Eq. (3).

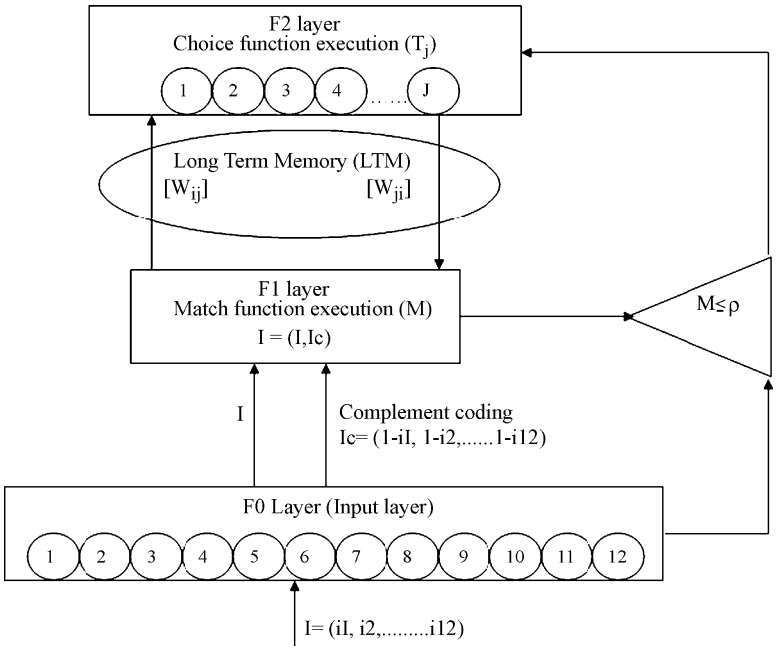

Fig. 3 – Fuzzy ART module.

3. Results and discussion

3.1. *Cluster analysis of the regenerated plants of gladiolus*

A validation data set comprising 55 leaf input patterns (Fig. 2(b)) was subjected to both neural network-based (Fuzzy ART, ART 2 and SOM) and statistical (hierarchical and *k*-means) autonomous clustering analyses to sort the regenerated plants into groups based on leaf photometric features in trichromatic domains. The sorted groups were compared in terms of *in vitro* corm induction potential of the representative shoots to ascertain the biological validation of clustering.

3.1.1. *ART 2- and SOM-aided clustering*

The validation data set were grouped by the ART 2 algorithm into two clusters (group '0' and '1') with the vigilance parameter value of 0.999. Of 55 plants, 19 plants were sorted into group '0' and the remaining 36 plants were placed in group '1'. Grouping pattern and clustering resolution of the validation data set are presented in Table 2. The VP value of 0.999 has already been found to be critical in deciphering the grouping

Table 2 – Grouping pattern and clustering resolution of regenerated shoots of gladiolus generated following Hierarchical, *k*-means algorithm, self-organized mapping and ART 2 approaches (value shown in the parentheses indicates the number of regenerated plants)

Analytical approach	Module	No. of groups	Regenerated leaf pattern no.
Statistical	Hierarchical clustering	A	R1, R2, R3, R4, R5, R7, R11, R13, R15, R17, R18, R22, R23, R28, R29, R31, R33, R34, R36, R38, R39, R40, R41, R42, R43, R44, R45, R47, R48, R49, R51, R53, R54, R55 (34)
		B	R6, R8, R9, R10, R12, R14, R16, R19, R20, R21, R24, R25, R26, R27, R30, R32, R35, R37, R46, R50, R52 (21)
	k-means algorithm	1	R1, R2, R18, R22, R23, R28, R29, R31, R33, R34, R36, R38, R39, R40, R41, R42, R43, R44, R45, R47, R48, R49, R51, R53, R54, R55 (26)
		2	R3, R4, R5, R6, R7, R8, R9, R10, R11, R12, R13, R14, R15, R16, R17, R19, R20, R21, R24, R25, R26, R27, R30, R32, R35, R37, R46, R50, R52 (29)
Artificial neural network	Self-organizing map	1	R3, R4, R5, R6, R7, R8, R9, R10, R11, R12, R13, R14, R15, R16, R17, R19, R20, R21, R24, R25, R26, R27, R30, R35, R37, R46, R50, R52 (28)
		2	R1, R2, R18, R22, R23, R28, R29, R31, R32, R33, R34, R36, R38, R39, R40, R41, R42, R43, R44, R45, R47, R48, R49, R51, R53, R54, R55 (27)
	Adaptive resonance theory 2 (at VP of 0.999)	0	R1, R5, R8, R9, R13, R15, R22, R25, R26, R30, R31, R34, R38, R39, R44, R46, R47, R51, R52 (19)
		1	R2, R3, R4, R6, R7, R10, R11, R12, R14, R16, R17, R18, R19, R20, R21, R23, R24, R27, R28, R29, R32, R33, R35, R36, R37, R40, R41, R42, R43, R45, R48, R49, R50, R53, R54, R55 (36)

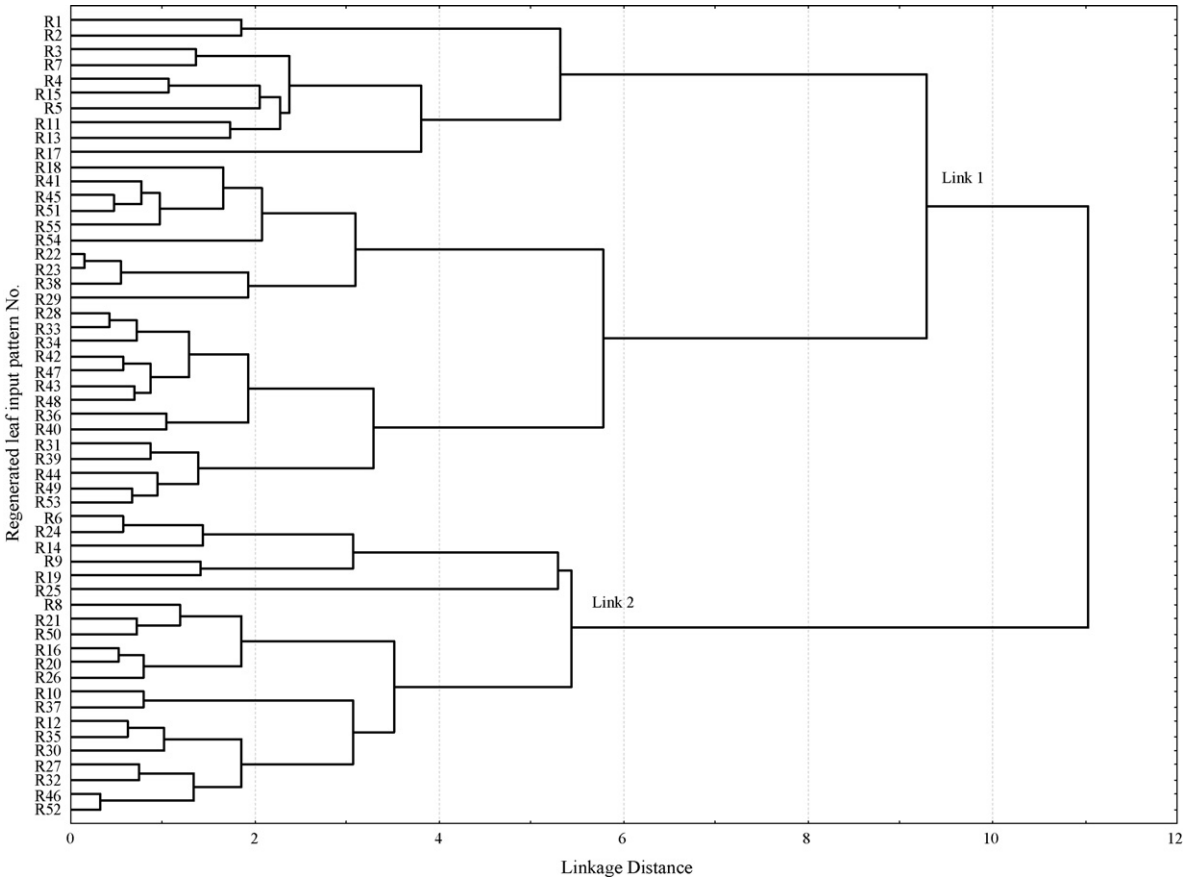

Fig. 4 – Hierarchical clusters of regenerated leaf input patterns.

pattern of the training data sets (Mahendra et al., 2004). Clustering of the validation set leaf input patterns did not result in any variation in terms of the number of groups that are generated from the training set. In our earlier work, the training set was sorted into two groups with ART 2 (Mahendra et al., 2004). The self-organized mapping analysis of the validation data set also resulted in two distinct clusters. Regenerated plants were distributed in the ratio of 28:27 in two groups (Table 2).

3.1.2. *Hierarchical clustering*

In Hierarchical clustering, the extent of similarity or dissimilarity between the regenerated leaf input patterns was estimated by calculating the Euclidean distance function. Thereafter, based upon the proximity level, the input patterns are linked together by a method of complete linkage function until all the clusters are grouped together in a hierarchical format. The hierarchical clustering results showed that the greatest linkage distance was between Link 1 and Link 2 (Fig. 4). This inconsistency of linkage distance indicated prominent differences existed between the input patterns at this level of hierarchy. Therefore, the cluster of leaf input patterns grouped under the Link 1 was considered to be distinctly different from the cluster grouped under Link 2. The grouping patterns of 55 regenerated plants are shown in Table 2.

3.1.3. *k-Means clustering*

The dual-grouping property of both the ART 2 and HC methods prompted us to heuristically set two cluster centroids for *k*-means clustering of the validation-set leaf input patterns. Table 3 presents the mean values of variables in each cluster. The resultant two groups of input patterns generated by *k*-means clustering are distinct from each other due to significant differences between the mean values of all the 12 features.

Table 3 – Mean values of each variable in cluster 1 and cluster 2

Variable	Cluster 1	Cluster 2
Luminosity-mean	0.872528	−0.782267
Luminosity-level	0.879119	−0.788175
Luminosity-count	−0.517605	0.464059
Red-mean	0.879108	−0.788166
Red-level	0.880380	−0.789307
Red-count	−0.451130	0.404461
Green-mean	0.802724	−0.719683
Green-level	0.809589	−0.725839
Green-count	−0.402454	0.360821
Blue-mean	0.815030	−0.730717
Blue-level	0.854906	−0.766467
Blue-count	−0.796882	0.714446

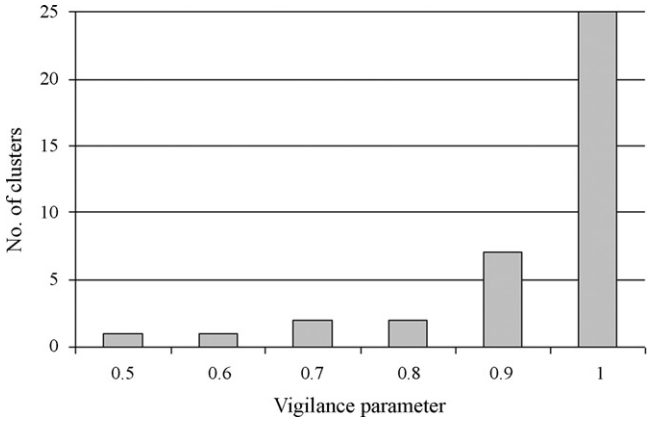

Fig. 5 – Effect of vigilance parameter on the number of clusters generated by Fuzzy ART.

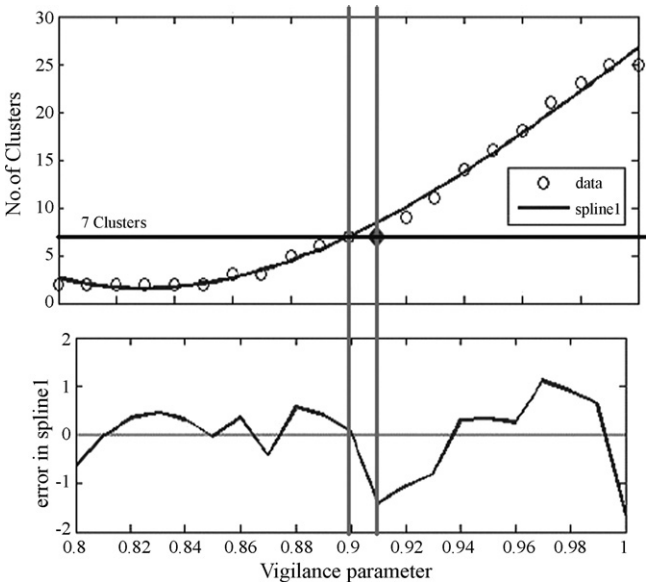

Fig. 6 – B-spline interpolation of number of clusters generated.

The partitioning of validation-set leaf input patterns into two distinct clusters is shown in Table 2. It appears that the regenerated plants are not uniform in terms of leaf photometric features. Plants having maximum similarity in trichromatic domain fall in a particular group and they were clustered into two distinct groups, both by classical discriminant analysis techniques and the ART 2-based artificial neural network. However, the distribution of regenerated plants in two clusters appeared to be dependent on the method of analysis.

3.1.4. *Fuzzy ART-based clustering*

Initially, training set data of leaf images were subjected to Fuzzy ART clustering analysis with vigilance parameter values ranging from 0.1 to 1.0 at an interval of 0.01 increments. The number of generated clusters increased with higher VP values. For VP values less than 0.8, there was little change in the number of generated clusters (Fig. 5).

For VP values above 0.91, the number of clusters increased sharply. Similar kinds of variation in clustering pattern with a change in VP value were also noted in the analysis of gene expression profile using Fuzzy ART (Tomida et al., 2002). To identify the optimal number of clusters, we considered the error value obtained from B-spline interpolation (Fig. 6). From Fig. 6, it is evident that the network displayed a stable clustering up to a VP of 0.91. Inconsistency in network performance was observed with VP > 0.91. The number of groups corresponded to VP = 0.91 was 7 (Table 4) and seemed to be reasonable for Fuzzy ART clustering.

The efficiency of the learning process was checked with the validation data set. The test as well as validation leaf input patterns were allocated into seven groups (Table 5). Similarity in the number of groups between the training and validation data sets indicates the efficiency of network classification and its ability to recognize the photometric variation in leaf image properties.

Compared to the performance of classical discriminant analysis techniques and ART 2, the Fuzzy ART generated more clusters (7) in both the training and validation data sets. The combination of the fuzzy set theory and adaptive resonance theory resulted in refined grouping and thereby efficiently projected the variation among the regenerated plants in terms of leaf trichromatic features.

One of our aims was to find whether the autonomously sorted groups exhibit any difference in their organogenic (corm induction) potential. The organogenic potential of the *in vitro* shoots obviously would take only two probable routes, i.e. either they would or would not produce *in vitro* corms. Therefore, from a plant tissue culturist point of view, clustering into two groups as obtained by HC, *k*-means, SOM and ART 2 seems to be more logical than the refined grouping of Fuzzy ART. Compared to ART 2, Fuzzy ART uses the degree of an input pattern being fuzzy subset of a stored prototype to measure the similarity between the patterns. This property makes Fuzzy ART highly sensitive to additional noise on trained input patterns and resulted in refined grouping with seven clusters at optimum VP of 0.91.

Table 4 – Fuzzy ART mediated grouping and distribution of training set leaf input patterns at VP of 0.91

	Groups						
	A	B	C	D	E	F	G
No. of regenerated shoots	3	5	4	7	4	1	1
Leaf input pattern no.	1, 4, 5	2, 21, 23, 24, 25	6, 7, 9, 18	8, 10, 12, 13, 14, 15, 16	3, 11, 17, 20	19	22

Table 5 – Fuzzy ART mediated distribution and grouping of validation set leaf input patterns with potential for corm induction

	Groups						
	A	B	C	D	E	F	G
No. of regenerated shoots	2	12	2	11	8	12	8
Leaf input pattern no.	1, 2	3, 4, 5, 6, 7, 11, 13, 14, 15, 17, 21, 24	9, 19	8, 10, 12, 16, 20, 25, 26, 30, 35, 37, 50	18, 27, 32, 41, 46, 52, 54, 55	28, 33, 34, 36, 40, 42, 43, 44, 47, 48, 49, 53	22, 23, 29, 31, 38, 39, 45, 51
No. of plants with corms	0	7	1	6	5	8	5
% Corm induction	0	58.3	50	54.5	62.5	66.6	62.5

3.2. *Biological validation of the grouping of regenerated shoots in terms of in vitro corm induction potential*

Biological validation of grouping of regenerated plants, i.e. which group of plants is more suitable to induce corm was also investigated. The culture-derived corms are analogous to synthetic seeds and may be conveniently stored, distributed and sown at any time of the year, bypassing the hardening stage and simultaneously eliminating the problem of dormancy exhibited by *in vivo* corms. Non-invasive prediction of the behavior of the *in vitro* regenerated shoots that are capable of producing corms might enhance the overall efficacy of micropropagation.

Corm induction features, such as percent induction, perimeter, fresh and dry weights, were compared among the groups generated by various clustering approaches and are presented in Table 6. The groups generated by HC, KM and SOM clustering methods did not show any significant difference with respect to corm induction features. A significant difference in percent corm induction among the groups was noted with ART 2. Corm induction in group '1' was 69.4%, while 36.8% regenerated plants of group '0' developed corms. Quantitative features of the developed corms, such as perimeter, fresh and dry weights, have also been found to be significantly different between the two groups (Table 6).

A significantly high percentage of corm development in group '1' with improved corm features indicates the biological potential of the regenerated shoots categorized in that group. However, the grouping pattern of Fuzzy ART rendered itself inconclusive while trying to distinguish the groups with significant difference in corm induction potential (Table 5). Neuro-fuzzy clustering algorithms were used successfully in characterizing the genetic and metabolic functions of prokaryotic organisms. Fuzzy ART has been used to cluster sporulation-specific gene expression profiles of *Saccharomyces* (Tomida et al., 2002) and to analyze the gene expression of heat shock protein (Kato et al., 2002). Fuzzy ART was also

Table 6 – Correlation of grouping pattern with *in vitro* corm induction and quantitative features of developed corms

Group no.	No. of plants	No. of plants producing corms	% Corm induction	Perimeter (mm)	Diameter (mm)	FW (g)	DW (g)	Relative dry weight (DW/FW)
Hierarchical clustering								
A	34	20	58.8	21.60 ± 5.88	6.88 ± 1.87	0.24 ± 0.10	0.14 ± 0.07	0.55 ± 0.08
B	21	12	54.5	23.58 ± 7.91	7.51 ± 2.52	0.29 ± 0.16	0.17 ± 0.12	0.54 ± 0.11
				P = 0.461 ns	P = 0.461 ns	P = 0.303 ns	P = 0.363 ns	P = 0.906 ns
k-Means algorithm								
1	26	16	61.5	22.25 ± 5.96	7.09 ± 1.90	0.24 ± 0.11	0.14 ± 0.08	0.54 ± 0.07
2	29	16	55.1	22.44 ± 7.49	7.15 ± 2.39	0.28 ± 0.14	0.16 ± 0.10	0.55 ± 0.11
				P = 0.938 ns	P = 0.938 ns	P = 0.470 ns	P = 0.464 ns	P = 0.816 ns
Self-organizing map								
1	28	15	53.5	22.60 ± 7.73	7.20 ± 2.46	0.29 ± 0.14	0.17 ± 0.10	0.56 ± 0.10
2	27	17	62.9	22.12 ± 5.80	7.04 ± 1.85	0.24 ± 0.11	0.13 ± 0.08	0.53 ± 0.08
				P = 0.844 ns	P = 0.844 ns	P = 0.325 ns	P = 0.293 ns	P = 0.396 ns
Adaptive resonance theory 2								
0	19	7	36.8	16.14 ± 2.97	5.14 ± 0.95	0.19 ± 0.05	0.11 ± 0.04	0.54 ± 0.06
1	36	25	69.4	24.08 ± 6.38	7.67 ± 2.03	0.28 ± 0.14	0.16 ± 0.10	0.54 ± 0.10
				P = 0.0001*	P = 0.00011*	P = 0.0146*	P = 0.0315*	P = 0.976 ns

Perimeter = $2 \times \pi \times$ radius; diameter = $2 \times$ radius.

* Significant at $P \leq 0.05$; ns, non-significant.

applied to time-series microarray data of oxidative stress in *Saccharomyces* to infer genetic interactions (Takahashi et al., 2002, 2003). Similarly, a time-course gene expression profile of *Escherichia coli* was subjected to Fuzzy ART to infer the role of genes in organic solvent tolerance (Shimizu et al., 2005). In plant tissue culture systems, the application of Fuzzy neural network was found to be useful in estimating the shoot length of regenerated rice (Honda et al., 1997).

Early non-invasive prediction of the potential of the regenerated shoots for corm induction appeared to be possible with ART 2 solution. Methods such as HC and KM processed the patterns in batch and lacked the property of recognizing and memorizing new input patterns, yet retaining the previously learned information. The primitive processing nature of SOM neural network perhaps could not recognize groups significantly different with respect to *in vitro* corm induction. Among the ART networks, Fuzzy ART with fuzzy sets made the network noise-sensitive, resulting in clustering of the patterns into more groups, which however, was inappropriate to index *in vitro* corm induction capacity. The merit of ART 2 in deciphering shoots with significant difference in *in vitro* corm development potential lies in its low noise-sensitivity and network plasticity.

4. Conclusions

The present work describes the photometric clustering of regenerated shoots and compares the performance of various clustering approaches in terms of corm induction potential of the sorted groups. ART 2-aided sorting of the regenerated shoots appeared to be more promising for selecting groups of plants capable of corm development than HC, KM and SOM autonomous clustering methods. Fuzzy ART resulted in larger numbers of clusters over ART 2, but the grouping pattern does not enable us to segregate the potential corm producing shoots. The significant difference in corm development percentages in two trichromatically grouped plants by ART 2 not only indicated a link between the photometric properties of the shoots to their organogenic potential but also the biological significance of photometric clustering of regenerated shoots.

REFERENCES

Albiol, J., Campmajo, C., Casas, C., Poch, M., 1995. Biomass estimation in plant cell cultures: a neural network approach. Biotechnol. Prog. 11, 88–92.

Carpenter, G.A., Grossberg, S., 1987a. ART 2: stable self-organization of pattern recognition codes for analog input patterns. Appl. Opt. 26, 4919–4930.

Carpenter, G.A., Grossberg, S., 1987b. A massively parallel architecture for a self-organizing neural pattern recognition machine. Comput. Vision Graph. Image Process. 37, 54–115.

Carpenter, G.A., Grossberg, S., Rosen, B., 1991. Fuzzy ART: fast stable learning and categorization of analog pattern by an adaptive resonance systems. Neural Networks 4, 759–771.

Dutta Gupta, S., Datta, S., 2004. Antioxidant enzyme activities during *in vitro* morphogenesis of gladiolus and the effect of application of antioxidants on plant regeneration. Biol. Plant. 47, 179–183.

Eisen, M.B., Spellman, P.T., Brown, P.O., Bostein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. (USA) 95, 14863–14868.

Grossberg, S., 1976. Adaptive pattern classification and universal recoding II: feedback, expectation, olfaction and illusions. Biol. Cybernatics 23, 187–202.

Hartigan, J.A., Wong, M.A., 1979. A *k*-means clustering algorithm. Appl. Stat. 28, 100–108.

Honda, H., Takikawa, N., Noguchi, H., Hanai, T., Kobayashi, T., 1997. Image analysis associated with a fuzzy neural network and estimation of shoot length of regenerated rice callus. J. Ferment. Bioeng. 84, 342–347.

Ibaraki, Y., 2006. Evaluation of photosynthetic capacity in micropropagated plants by image analysis. In: Dutta Gupta, S., Ibaraki, Y. (Eds.), Plant Tissue Culture Engineering. Focus on Biotechnology, 6. Springer, The Netherlands, pp. 15–29.

Kato, N., Kobayashi, T., Honda, H., 2002. Gene expression analysis heat shock response using Fuzzy ART. Genome Inf. 13, 272–273.

Kohonen, T., 1997. Self-Organizing Maps. Springer, Berlin.

Larkin, P.J., Scowcroft, W.R., 1981. Somaclonal variation—a novel source of variability from cell cultures for plant improvement. Theor. Appl. Genet. 60, 197–214.

Mahendra, Prasad, V.S.S., Dutta Gupta, S., 2004. Trichromatic sorting of *in vitro* regenerated plants of gladiolus using adaptive resonance theory. Curr. Sci. 87, 348–353.

Murashige, T., Skoog, F., 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant. 15, 473–497.

Prasad, V.S.S., Dutta Gupta, S., 2006. Applications and potentials of artificial neural networks in plant tissue culture. In: Dutta Gupta, S., Ibaraki, Y. (Eds.), Plant Tissue Culture Engineering. Focus on Biotechnology, 6. Springer, The Netherlands, pp. 49–69.

Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Ress, C., Spellman, P., Iyer, V., Jeffrey, S.S., Rijn, M.V.D., Waltham, M., Pergamenschikov, A., Lee, J.C.F., Lashkari, D., Shalon, D., Myers, T.G., Weinstein, J.N., Botstein, D., Brown, P.O., 2000. Systematic variation in gene expression patterns in human cancer cell lines. Nat. Genet. 24, 227–235.

Shimizu, K., Hayashi, S., Doukyu, N., Kobayashi, T., Honda, H., 2005. Time-course data analysis of gene expression profiles reveals purR regulon concerns in organic solvent tolerance in *Escherichia coli*. J. Biosci. Bioeng. 99, 72–74.

Takahashi, H., Tomida, S., Kobayashi, T., Honda, H., 2002. Genetic network analysis using F2D matrix. Genome Inf. 13, 377–379.

Takahashi, H., Tomida, S., Kobayashi, T., Honda, H., 2003. Inference of common genetic network using fuzzy adaptive resonance theory associated matrix method. J. Biosci. Bioeng. 96, 154–160.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., Golub, T.R., 1999. Interpreting patterns of gene expression with self organizing maps: methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. (USA) 96, 2907–2912.

Tomida, S., Hanai, T., Honda, H., Kobayashi, T., 2002. Analysis of expression profile using fuzzy adaptive resonance theory. Bioinformatics 18, 1073–1083, available at: http://www.nubio.nagoya-u.ac.jp/proc/ENGLISH/Fuzzy.html.