MARDI Res. J. 19(1) (1991): 31–42

Numerical analysis of variation among *Mangifera indica* L. accessions

(Analisis berangka bagi perbezaan antara asesi mangga *Mangifera indica* L.)

I. Salma*, T. Zahrah** and H. Masrom*

Key words: classification, numerical taxonomy, cluster analysis, *Mangifera indica* L.

Abstrak

Kebolehubahan antara 26 klon mangga (*Mangifera indica* L.) telah diselidiki dengan menggunakan kaedah rantaian lengkap, purata kelompok dan kaedah Wards serta analisis komponen utama dengan putaran varimaks berdasarkan 53 ciri morfologi. Hasil daripada analisis purata kelompok menunjukkan persamaan dengan ordinasi analisis komponen utama pada paksi 1, 2 dan 3. Empat kumpulan asesi telah diperoleh. Analisis komponen utama menunjukkan bahawa ciri-ciri utama yang membezakan klon-klon mangga ialah bentuk, saiz, bau dan rasa buah, panjang dan warna tangkai bunga, saiz dan jumlah urat daun serta bentuk pangkal daun.

Abstract

Variabilities among 26 clones of mango (*Mangifera indica* L.) were studied using complete linkage, unweighted pair group method using averages (UPGMA) and Wards method, and principal component analysis (PCA) with varimax rotation based on 53 morphological characters. Results of UPGMA agreed with that of PCA ordination on axes 1, 2 and 3. Four groups of accessions were revealed. The PCA showed that the main distinguishing characters among mango clones were fruit shape, fruit size, aroma, flavour, panicle length, panicle colour, leaf size, number of secondary veins and shape of leaf base.

Introduction

Mango types in cultivation are generally divided into three main groups i.e. monoembryonic, polyembryonic and horticultural varieties (Chaudhri 1976). The monoembryonic mangoes are generally referred to the Indo-Pakistan type while the polyembryonic ones are found in the tropics such as Malaysia, Indo-China, the Philippines, Hawaii, Mexico, Brazil and West Indies. The clonal progenies of the monoembryonic seedlings arise to the horticultural varieties.

A few different classification schemes have been elaborated for mango varieties. Hartless (1913) emphasised the importance of floral characteristics in classifying mango varieties. Similarly, Popenoe (1932) classified mango varieties based on flower and fruit characteristics. On the other hand, Naik and Gangolly (1950), and Singh and Singh (1956) used leaf and fruit characteristics for their mango classification. Most classification schemes elaborated were based on conventional taxonomy but the numerical approach has been neglected.

Numerical taxonomic techniques may help towards achieving precise and flexible classifications (Baum 1987). Numerical taxonomy is a method which involves the

*Basic Research Division, MARDI, P.O. Box 12301, 50774 Kuala Lumpur, Malaysia
**Techno-Economic and Social Studies Division, MARDI, P.O. Box 12301, 50774 Kuala Lumpur, Malaysia
Authors' full names: Salma Idris, Zahrah Talib and Masrom Hasran

Table 1. The clone number, local name and geographical origin of 26 mango clones

Code no.	Clone no.	Local name	Geographical origin
1	MA 2	-	Malaysia
2	MA 3	-	Malaysia
3	MA 112	Mempelam Siam	Thailand
4	MA 121	Gundu	India
5	MA 127	Lebai Mohamed	Malaysia
6	MA 128	Harumanis	Indonesia
7	MA 129	Kuala Selangor 1	Malaysia
8	MA 134	Orange	Florida
9	MA 154	Paris	Florida
10	MA 156	Fairchild	Florida
11	MA 159	Sungai Siput	Malaysia
12	MA 165	MAHA 65	Malaysia
13	MA 173	Alfonso	India
14	MA 179	Rim Green/ Telok Anson	Malaysia
15	MA 190	Ampalavi	India
16	MA 194	Apple	India
17	MA 195	Bahagia	Malaysia
18	MA 200	Malgoa	India
19	MA 206	Tetenene	Indonesia
20	MA 207	Simpang Empat	Malaysia
21	–	Cheteau	India
22	–	Colombo Black	India
23	–	Lenggong	Malaysia
24	–	Luzon	Philippines
25	–	Malgoa Tan	Malaysia
26	–	Sri Lanka	India

grouping of taxonomic units into taxa on the basis of their character states (Sokal and Sneath 1963). The numerically coded characteristics are analysed to produce a phenogram which shows the relationship of the taxonomic units. This technique has been widely used in the classifications of several crops to establish cultivar groups such as in barley (Molino-Cano 1978; Baum 1987), yams (Martin and Rhodes 1977; Okoroda 1983), sweet corn (Rhodes and Carmer 1966), avocado (Rhodes et al. 1971) and eggplant (Martin and Rhodes 1979). Based on the numerical analysis, the genetic relationship of 40 cultivars of *Mangifera indica* and one specimen each from *M. odorata* and *M. zeylanica* was investigated by Rhodes et al. (1970).

The existence of many mango clones from different origins in MARDI's germplasm has created problems in distinguishing the accessions. Therefore, there is a need to know the characteristics which are more important in accounting for the major variations and differences among accessions. This study aims to identify the principal characteristics for mango variation among *Mangifera indica* accessions and to establish clonal groups of these accessions.

Materials and methods

The trees used in this study are located at MARDI station, Kuala Kangsar, Perak. The trees are 5–6 years old maintained under standard horticultural practice (Norlia n.d.).

Data for leaf characteristics were recorded during 1982–1983, while those of flower and fruit were recorded during the seasons in 1983–1985. Data were recorded from three trees for each clone. The code number, clone name and reputed geographical origin are presented in *Table 1* while the characteristics and character states are in *Table 2*.

Two multivariate methods, cluster analysis and principal component analysis (PCA), were used to analyse the data. These two techiques have been shown by Rhodes

Table 2. The characters and character states

No.	Character	Character states
1	Petiole length (cm)	(1) <3, (2) 3–4.5, (3) >4.5
2	Leaf length (cm)	(1) <18, (2) 18–22, (3) >22
3	Leaf width (cm)	(1) <4, (2) 4–5.5, (3) >5.5
4	Leaf area (sq. cm)	(1) <60, (2) 60–90, (3) >90
5	No. of secondary veins	(1) <12 pairs, (2) >12 pairs
6	Leaf tip	(1) acute, (2) subacuminate, (3) acuminate
7	Leaf base	(1) acute, (2) obtuse, (3) round, (4) acute to obtuse, (5) obtuse to round
8	Shoot colour	(1) light green, (2) brownish green, (3) brownish red
9	Leaf colour	(1) green, (2) dark green
10	Leaf orientation	(1) upheld to outheld, (2) outheld, (3) outheld to downheld
11	Leaf habit	(1) flat to slightly impressed, (2) slightly impressed, (3) slightly impressed to impressed, (4) deeply impressed
12	Tip twisted	(1) no, (2) yes
13	Leaf margin	(1) smooth to slightly wavy, (2) slightly wavy, (3) slightly wavy to wavy, (4) wavy
14	Leaf shape	(1) elliptic lanceolate, (2) oval lanceolate, (3) ovate lanceolate, (4) elliptic to oval lanceolate, (5) oval to ovate lanceolate
15	Panicle width (cm)	(1) <18, (2) 18–35, (3) >35
16	Panicle length (cm)	(1) <30, (2) 30–50, (3) >50
17	Branchlet number	(1) <20, (2) 20–30, (3) >30
18	Flower diameter (mm)	(1) ≤ 0.7, (2) >0.7
19	Style length (mm)	(1) ≤0.25, (2) >0.25
20	Stamen length (mm)	(1) ≤0.21, (2) >0.21
21	Petal shape	(1) oval, (2) round
22	Panicle shape	(1) conical, (2) pyramid, (3) broadly pyramidal
23	Fragrance	(1) none to very slight, (2) slight, (3) strong
24	Density of hair on panicle	(1) few, (2) intermediate, (3) dense
25	Height of staminode (mm)	(1) <0.25, (2) 0.25–0.50, (3) >0.50
26	Panicle colour	(1) green, (2) greenish pink, (3) reddish green, (4) red
27	Fruit length (cm)	(1) <10, (2) 10–15, (3) >15
28	Fruit width (cm)	(1) <7, (2) 7–10, (3) >10
29	Fruit shape	(1) round, (2) ovate oblong, (3) long
30	Fruit weight (g)	(1) <350, (2) 350–500, (3) >500
31	Sugar content (% Brix)	(1) ≤14, (2) >14
32	Pulp thickness (cm)	(1) <1, (2) 1–2, (3) >2
33	Fruit surface at stalk attachment	(1) elevated, (2) level, (3) depressed
34	Dorsal shoulder	(1) abruptly falling, (2) moderately falling, (3) level
35	Ventral shoulder	(1) rising, (2) level, (3) falling
36	Beak	(1) not prominent, (2) slightly prominent, (3) prominent
37	Fruit apex (breadth view)	(1) curved, (2) symmetrical
38	Fruit apex (thickness view)	(1) acute, (2) obtuse, (3) acute to obtuse
39	Basal cavity	(1) absent, (2) present
40	Skin colour	(1) yellowish green, (2) light green, (3) green, (4) dark green
41	Blush present on fruit	(1) absent, (2) present
42	Blush colour	(1) yellowish orange, (2) reddish
43	Sinus	(1) absent, (2) slight, (3) deep
44	Pulp colour	(1) light yellow, (2) yellow, (3) yellowish orange, (4) orange to deep orange
45	Pulp texture	(1) soft, (2) intermediate, (3) firm
46	Amount of fibre	(1) none to scanty, (2) moderate, (3) abundant
47	Presence of aroma	(1) no, (2) yes
48	Flavour	(1) mild, (2) rich
49	Presence of acid	(1) no, (2) slight, (3) strong
50	Edible portion (%)	(1) ≤70, (2) >70
51	Skin thickness (mm)	(1) <0.1, (2) 0.1–0.15, (3) >0.15
52	Stone shape	(1) squarish to ovate, (2) oblong, (3) oblong elongate
53	Embryo type	(1) monoembryonic, (2) polyembryonic

and Martin (1977) as complementary to each other.

For cluster analysis, standard numerical taxonomic procedures (Sokal and Sneath 1963) were followed. The steps used, involved standardisation of the characters to have a mean of 0 and a variance of 1. Euclidean distance coefficient was then computed and a similarity matrix was established. The resulting similarity matrix was clustered using the single linkage, complete linkage, average linkage and Wards method. The cluster results were summarised in the form of phenograms *(Figure 1* to *Figure 3)*. The phenograms obtained were compared using the cophenetic correlation (Sokal and Rohlf 1962).

The process of PCA resulted in a transformation of the original variables into a new set of axes called components (Abbot et al. 1985). The first component accounts for as much as possible the variability in the original data, and the second for the next largest portion of the variability. Varimax rotation was used to redistribute the principal component results, so that the solution approached an orthogonally simple structure (Rhodes and Martin 1972). The relative positions among clones and the trend in variation among characteristics are graphically shown when plotted as two dimensional figures with reference to the components.

All computations were done on the IBM 4381–11 computer at MARDI, Serdang, Selangor, Malaysia.

Results and discussion

Cluster analysis

The three methods compared here are the complete linkage, unweighted pair group method using averages (UPGMA) and the Wards method. Based on the three phenograms *(Figure 1* to *Figure 3)*, the 26 clones can be divided generally into four major groups at phenon line 2.65, 2.05 and 3.23 respectively.

A general comparison within clusters among the three phenograms indicates that cluster 2 is identical i.e. having the same set of clones. Cluster 4 is identical only in two phenograms (*Figure 1* and *Figure 3*), but in the other phenogram (*Figure 2*) clone MA 134 has been excluded. Cluster 3 is similar between two phenograms (*Figure 2* and *Figure 3*) but agrees fairly well with the other phenograms produced by complete linkage (*Figure 1*). Cluster 1 shows the greatest degree of discordance in the three phenograms. The degree of agreement between the phenograms produced by different clustering methods and between the different phenograms, and the distance in the similarity matrix on which they are based is referred to as cophenetic correlation (Sokal and Rohlf 1962). The higher the correlation, the better the phenogram summarises the results of the similarity matrix (Sokal and Rohlf 1962). The cophenetic correlations are summarised as shown in *Table 3*. The correlation between the similarity matrix and the phenogram distances from the corresponding clustering methods are generally low. The correlation with the phenogram distance resulting from the average linkage method is fairly high, 0.66. These similarities between the phenograms produced by the average linkage and complete linkage, and between average linkage and the Wards method are quite high, 0.79 and 0.68 respectively. The similarity between the phenograms produced by complete linkage and Wards method is 0.61. These values 0.79, 0.68 and 0.61 indicate a fair representation of the phenograms. The cophenetic correlation for the single linkage method is the lowest and therefore, the results are not discussed.

Since the average method has been found to produce a good summary of the results, therefore the analysis elaborated then is based on the dendrogram produced by this method (*Figure 2*). The dendrogram separates the 26 clones into four major groups. There is no clear separation of the clones according to their countries of origin. However, the clones generally have been

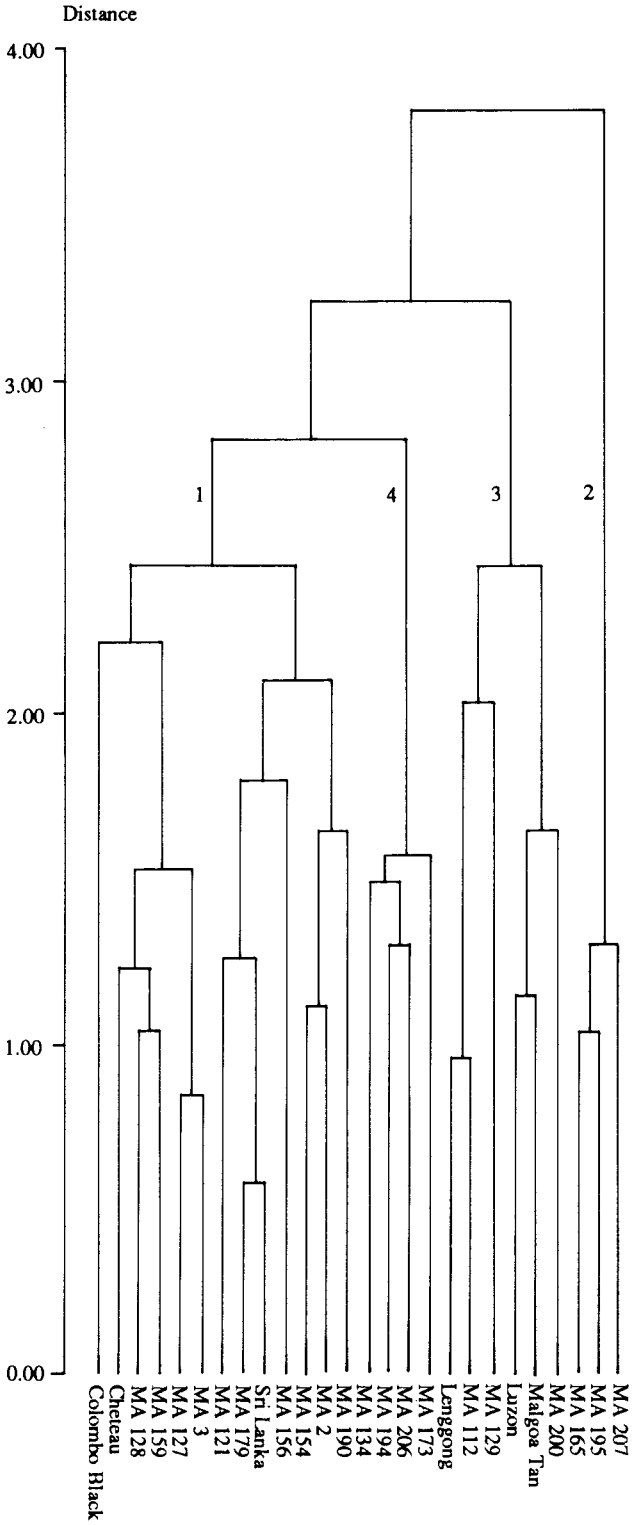

Figure 1. Distance phenogram produced by complete linkage

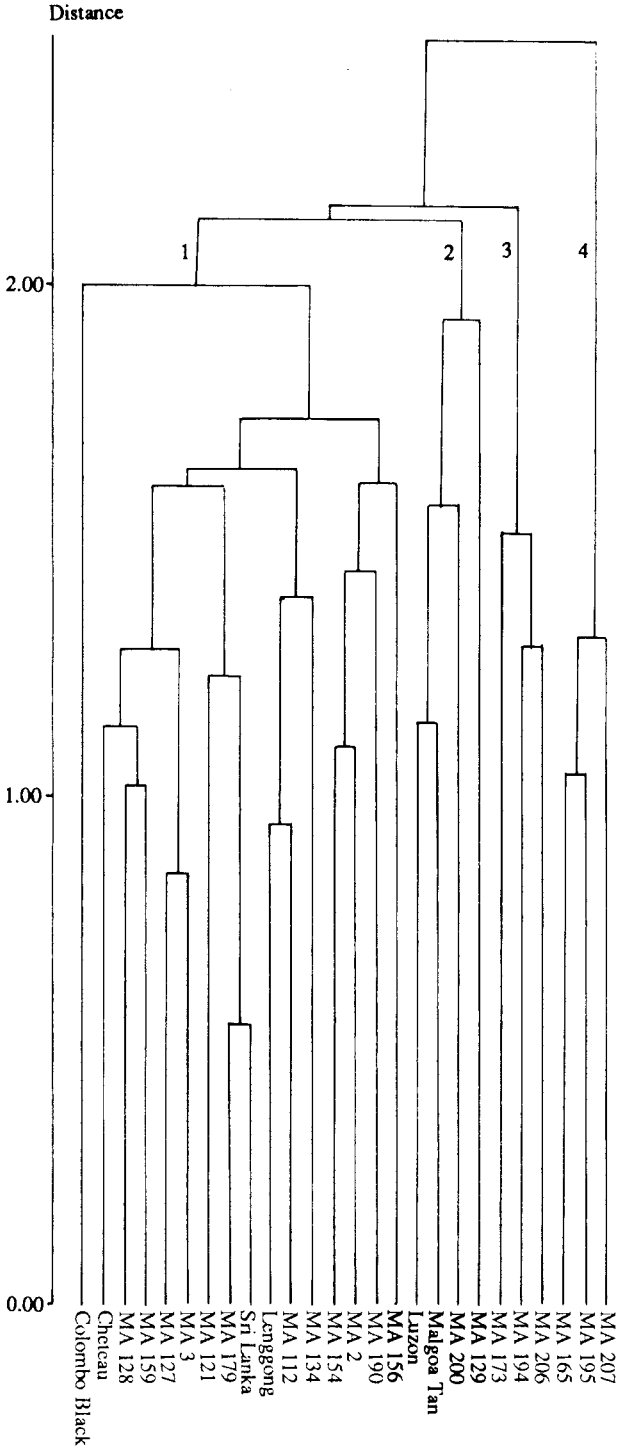

Figure 2. Distance phenogram produced by average linkage

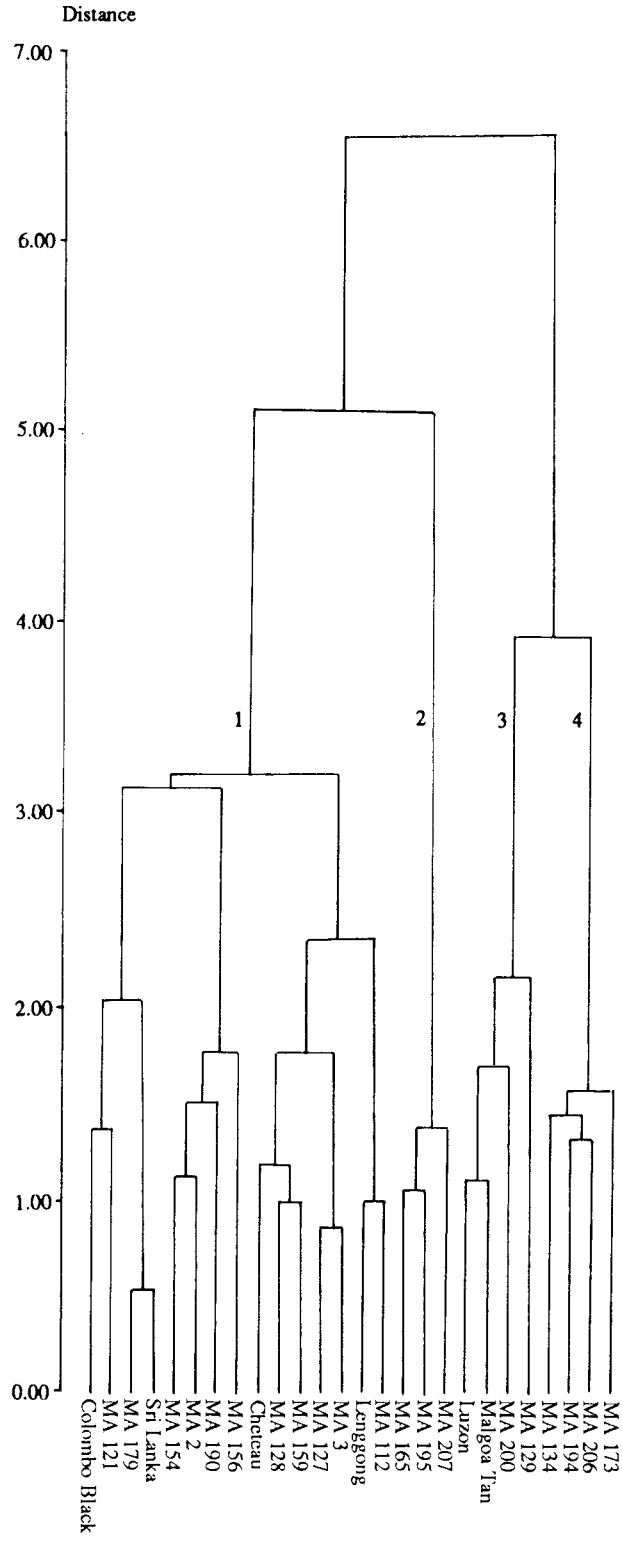

Figure 3. Distance phenogram produced by Wards method

Table 3. Cophenetic correlation matrix

	r	SL	AL	CL	Wards
r	-				
SL	0.47	-			
AL	0.66	0.48	-		
CL	0.55	0.23	0.79	-	
Wards	0.57	0.40	0.68	0.61	-

r denotes the original matrix of correlations between forms from which the dendrograms were derived
SL = single linkage
AL = average linkage
CL = complete linkage

found to group well according to the fruit shape and size. Group 1, which is the largest in number, consists of 16 clones, of which 5 are from India, 6 from Malaysia, 3 from Florida, 1 from Indonesia and 1 from Thailand. With the exception of clone MA 134 and MA 156, most clones in this group are differentiated from the other groups by having oval fruit. In the complete linkage (*Figure 1*) and Wards method dendrograms (*Figure 3*), clone MA 134 joined Group 4. Group 2 consists of three clones of Malaysian origin. These clones are very closely related. They are distinguished by having long fruit. Flower characteristics such as panicle length, style length and petal shape, and also shape of leaf base as well as qualitative characters of fruit are very similar in these three clones. Group 3 consists of four clones i.e. 2 from Malaysia, 1 from the Philippines and 1 from India. This group possesses slightly round and big fruit. Group 4 which are represented by three clones (2 from India and 1 from Indonesia) is distinguished by having round but smaller fruit compared with those of Group 3. The grouping of these clones based on fruit shape seems to be agreeable to the work done by Mukherjee (1949) who classified the mango varieties in India into three groups also based on fruit shape. Similarly, Rhodes et al. (1970) also found that mango cultivars grouped according to their fruit shape.

Varimax solution

For the principal component analysis, 16 eigenvalues (component variances) were obtained. The eigenvalues were greater than 1, which represented 92% of the total variance in the character correlation matrix. The first three components explain only 31.9% of the total variance and therefore the scores that are plotted on *Figure 4* and *Figure 5* for each clone represent a fairly satisfactory representation of the data.

The clones, as shown in *Figure 4* and *Figure 5*, can be separated into four groups and their positions agree quite well with their respective groupings in *Figure 1*, *Figure 2* or *Figure 3*.

The relative positions of the clones in *Figure 4* and *Figure 5* depend on the character loadings (*Table 4*) as compared with *Figure 1* to *Figure 3* which depend on the overall similarities. Characteristics which are effective in bringing about the separation are those with high loadings. *Table 4* lists 17 characteristics that express their highest variance in one of the first three components. The arrows in *Figure 4* and *Figure 5* indicate the trend in variation for certain characteristics showing their maximum influence on the respective components. The direction of an arrow indicates an increase in value for a particular characteristic. The arrow labelled shape along component 1 (*Figure 4*) indicates that clones to the far right have long fruit and those on the far left have round fruit. Clone MA 200 and MA 129 possess round and big fruit, compared with those in group 4. Thus, these clones do not group together with the clones in group 4. But clone MA 156, which is in cluster 1 (*Figure 1* to *Figure 3*), tends to be able to group with other clones which have small and round fruit in group 4 (*Figure 4*). Clones in the upper part of the diagram near component 2 tend to have rich flavour, strong aroma and sweet taste. Component 1 emphasises on fruit shape, fruit length, dorsal shoulder, basal cavity, stone shape, panicle length and leaf length. Component 2 reflects on fruit quality such as flavour, aroma and sugar content, and also on fruit size. The choice of clones with strong aroma, rich flavour and high sugar

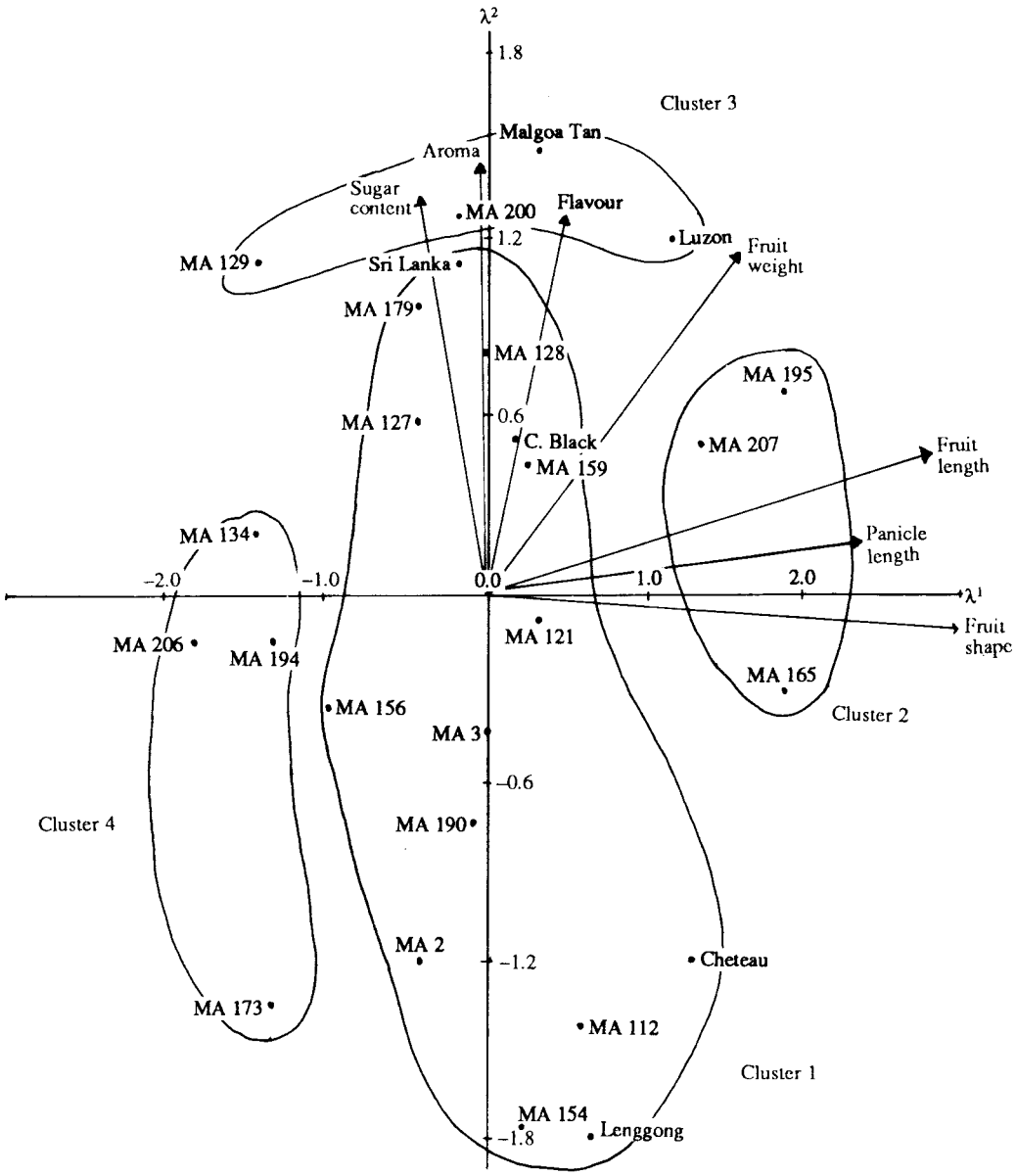

Figure 4. Varimax solution for component 1 and 2. Solid enclosed lines used to show clusters of clones. Arrows indicate character trends

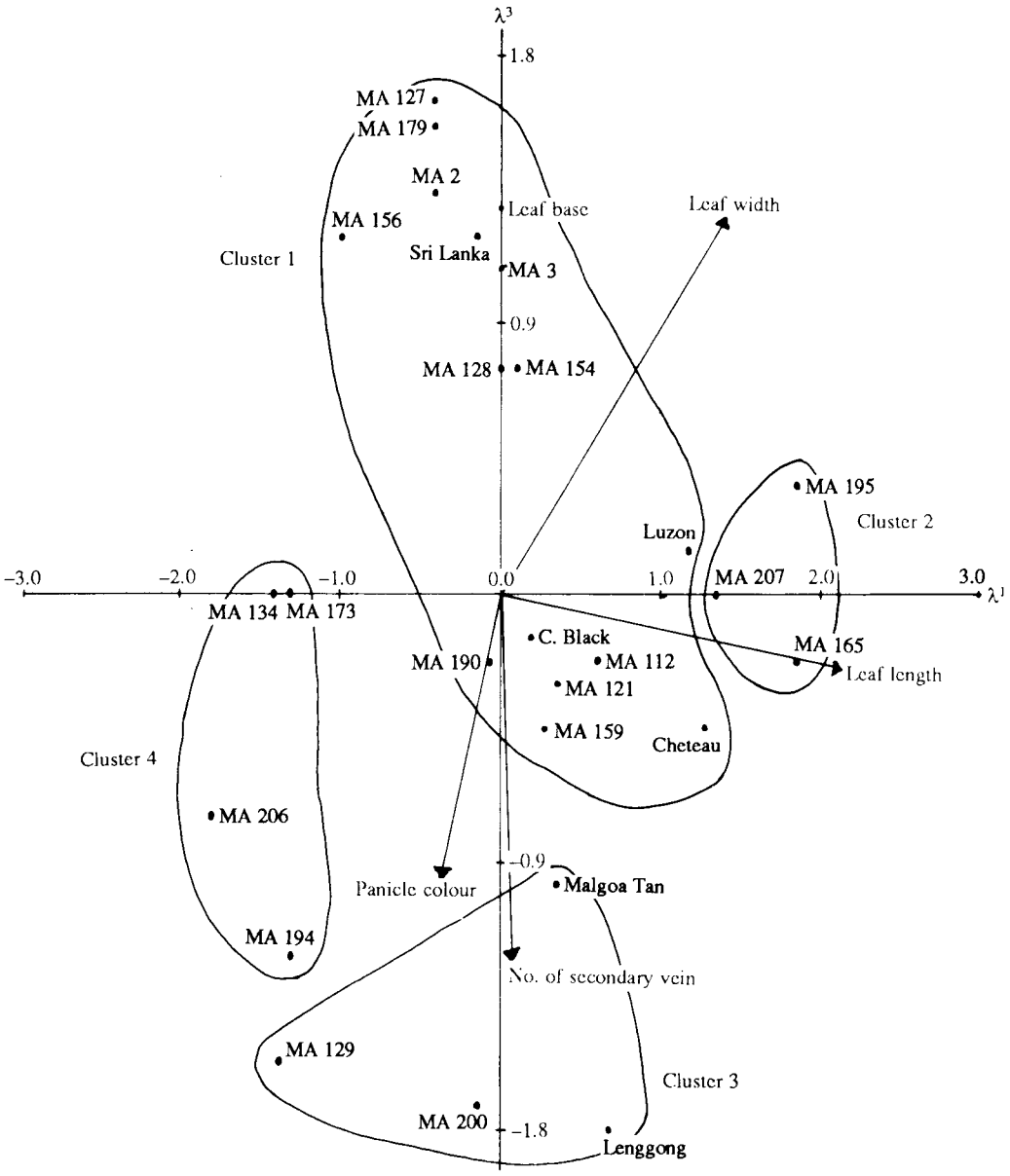

Figure 5. Varimax solutions of component 1 and 3. Solid enclosed lines used to show clusters of clones. Arrows indicate character trends

Table 4. Varimax solution of the first three components for 17 characters (decimal points omitted)

Character	Component		
	1	2	3
Fruit length	84	26	-06
Fruit width	14	62	-37
Fruit shape	90	-06	23
Fruit weight	48	64	-08
Sugar content	-13	75	17
Dorsal shoulder	-86	-07	-12
Basal cavity	-60	44	-22
Aroma	-01	82	06
Flavour	16	72	-08
Embryo	07	-08	59
Stone shape	86	07	11
Panicle length	70	10	09
Panicle colour	-11	-25	-54
Leaf length	64	-07	-14
Leaf width	43	02	72
No. of secondary veins	02	00	-69
Leaf base	-00	-06	83
Eigenvalues	7.2	4.8	4.6

content are found in the upper part of *Figure 4*. Component 3 is highly loaded chiefly by leaf characteristics and appears to be quite useful for the group separation. Therefore, the component solution helps in showing the major trends of variation among clones as well as overlapping in variation between groups (*Figure 5*). Clone Luzon and Lenggong in *Figure 5* do not agree with their respective locations in *Figure 2* to *Figure 4*. This is due to their positions being determined by leaf characteristics.

Conclusions

Euclidean distance, UPGMA clustering and varimax solution have proven to be reliable and useful numerical techniques on the classification of mango clones. The mango clones are clustered into four major groups. Results of PCA agree well to the cluster analysis. The PCA shows that 17 characteristics are important in distinguishing the mango accessions. The mango clones are grouped mainly according to their fruit characteristics. However, leaf characteristics also proved to be useful in discriminating the genotypes.

Acknowledgements

The authors are grateful to Mr Shohime Ramli for his technical assistance and to the research assistants from MARDI Kuala Kangsar for their help.

References

Abbot, L. A., Bisby, F. A. and Rogers, D. J. (1985). *Taxonomic analysis in biology* p. 167–71. New York: Columbia University Press

Baum, B. R. (1987). Classification of cultivated barley (*Hordeum vulgare*). 2. Elaboration of cultivar groups. *Can. J. Bot.* **65**: 2152–60

Chaudhri, S. A. (1976). *Mangifera indica* - Mango. In *The propagation of tropical fruit trees* (Garner, R. J. and Chaudhri, S. A., ed.) p. 403. Kent: P.D.A. and Commonwealth Agric. Bur.

Hartless, A. C. (1913). The flowering of the mango. *Agric. J. India* **8**: 90

Martin, F. W. and Rhodes, A. M. (1977). Intraspecific classification of *Dioscorea alata*. *Trop. Agric, Trinidad* **54**: 1–13

—— (1979). Subspecific grouping of eggplant cultivars. *Euphytica* **28**: 367–83

Molino-Cano, J. L. (1978). A further contribution to the classification of barley cultivars: Use of numerical taxonomy and biochemical methods. *Seed Sci. and Technol.* **6**: 593–615

Mukherjee, S. K. (1949). Monograph of *Mangifera* L. *Lloydia* **12**: 73–136

Norlia, Y. (n.d.). Mango cultivation. *MARDI Siri Maklumat Penyelidikan* Serdang: MARDI

Naik, K. C. and Gangolly, S. R. (1950). *A monograph on classification and nomenclature of South Indian mangoes* Madras: Supt. Govt. Press

Okoroda, M. D. (1983). Principal component analysis and metroglyph of variation among Nigerian yellow yams. *Euphytica* **32**: 565–73

Popenoe, W. (1932). *Manual of tropical and subtropical fruits* New York: Macmillan

Rhodes, A. M., Campbell, C., Malo, S. E. and Carmer, S. G. (1970). A numerical taxonomic study of the mango *Mangifera indica* L. *J. Amer. Soc. Hort. Sci.* **95**: 252–6

Rhodes, A. M. and Carmer, S. G. (1966). Classification of sweet corn inbreds by methods of numerical taxonomy. *Proc. Amer. Soc. Hort. Sci.* **88**: 507–15

Rhodes, A. M., Malo, S. E., Campbell, C. and Carmer, S. G. (1971). A numerical taxonomic study of the avacado (*Persea americana* Mill.). *J. Amer. Soc. Hort. Sci.* **96**: 391–5

Rhodes, A. M. and Martin, F. W. (1972). Multivariate studies of variations in yams

(*Dioscorea alata* L.). *J. Amer. Soc. Hort. Sci. 97(5):* 685–8

Singh, I. B. and Singh, R. N. (1956). *A monograph on the mangoes of Uttar Pradesh* Lucknow: Supt. of Printing

Sokal, R. R. and Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon 11(2):* 33–40

Sokal, R. R. and Sneath P. H. A. (1963). *Principles of numerical taxonomy.* San Francisco and London: W. H. Freeman and Co.

Accepted for publication on 12 September 1990