Morphological Classification of Wild Banana (*Musa acuminata* Colla) by Classical Taxonomy and Cluster Analysis Technique in Lower Northern Thailand

Det Wattanachaiyingcharoen [1/]

Kumthorn Therakhupt [1/] Surin Kangkun [2/]

ABSTRACT

Three hundred and nine samples of wild banana (*Musa acuminata* Colla) were collected in the five national parks namely a wildlife sanctuary and a forest park during March 2002 to March 2003. All samples were examined through classical taxonomy, then major qualitative characteristics were classified into 7 groups. Meanwhile, Hierarchical and K-mean cluster analysis were numerical taxono being used to report examination with these characteristics. There are 7 characteristics of species identification criterion, while other 13 characteristics were subspecies. All samples were classified into 7 groups at 5 units of re-scaled distance with 80 % of similarity by Hierarchical cluster analysis. In addition, the similar classification results which were re-approved by K-means cluster analysis and non-significantly different at P=0.05 from Hierarchical cluster analysis. Discriminant analysis were shown the probability of precise separation and prediction into 7 groups of both cluster techniques at 83-100 % with the original and cross-validated methods. Both results from numerical morphological classification were synchronised and non-significantly different from the result of classical morphological taxonomy with Pearson Chi-square test (P = 0.05). Sample in the group 6 was highly different from *M. acuminata* Colla with its erect rachis position that should be a new wild banana group. It is very interesting to focus on re-identification of this sample through banana taxonomic system. The prominent strong convolute with significant 1/3 bright light green tip of male bud of the group 2 which was also interesting and might be a new promising subspecies of *M. acuminata* Colla.

Key words : banana, classification, *Musa acuminata* Colla, morphology, cluster analysis, discriminant analysis.

[1/] Department of Biology, Faculty of Science, Chulalongkorn University, Pathum Wan, Bangkok 10330

[2/] Department of Agricultural Science, Faculty of Agriculture, Natural Resources and Environmental Science, Naresuan University, Phitsanulok province 65000

Moreover, the samples in group 1, 3, 4, 5, 7 indicated strong characteristics of the subspecies *malaccensis* (Ridl.) Simmonds, *malaccensis* (Ridl.) Simmonds, *malaccensis* (Ridl.) Simmonds, *siamea* Simmonds, and *burmannica* Simmonds respectively. Most of samples from 7 groups were also presented the best characters in genetic and breeding material for future utilization.

INTRODUCTION

Wild banana (*M. acuminata* Colla) is originated in South-east Asia especially in Thailand (Silayoi and Chomchalow, 1983). Thai common names are Kluai Kae (Northern Thai), Kluai Tuan (Southern Thai), Kluai Ling (Uttaradit), Kluai Hmon (Chaing Mai) (Suvatti, 1978). Moreover, there are 4 subspecies namely (1) *M. acuminata* Colla ssp. *siamea* Simmonds (2) *M. acuminata* Colla ssp. *burmannica* Simmonds (3) *M. acuminata* Colla ssp. *malaccensis* (Ridl.) Simmonds (4) *M. acuminata* Colla ssp. *microcarpa* (Becc.) Simmonds. Their distribution is in natural forest geographically through Thailand (Suvatti, 1978; Silayoi and Chomchalow, 1983). Fundamental classification of banana genome and cultivar are employed morphological scoring techniques with 15 characteristics (Simmonds and Shepherd, 1955). Cytology and chromosome counting (Silayoi and Chomchalow, 1983), molecular chemistry and DNA techniques (Rekha and *et al.*, 2001) were also successfully applied to identify the genome group of banana, especially the hybridization of *M. acuminata* Colla or/and M. *balbisiana* Colla (Simmonds and Weatherup, 1988; Swennen and Vuylsteke, 1986). Those techniques are required high expertise and delicate laboratory works, which are not practical for a person to identify banana promptly (Argent, 2000). However, the classical taxonomy may not be modernized but is suitable to support fieldwork and complement with a rapidly identified result (Simmonds and Shepherd, 1955; Valmayor and Pascua, 1985; Valmayor, 2001).

Moreover, the classical taxonomy and classification of subspecies have to be confirmed to harmonise with the modern taxonomy trend (Nasution, 1993). Numerical taxonomy techniques could enhance the efficiency utilization of classical taxonomy as well as other techniques. The objectives of this research were to explore new genetic materials, and to compare of the classical numerical taxonomy techniques.

METERIALS AND METHODS

(1) Line Transect Survey method was employed cut through the research sites (Sutherland, 1996). Samples and morphological

information were recorded (Anon, 1997).

(2) Three handred and nine samples were collected from 7 research sites (Figure 1). Arcview GIS software version 3.2 was used as Desktop Mapping and Geographic Information System to analyze the collected and spatial data (Haines-Young *et al.*, 1993).

(3) Classical taxonomy was used to identify the samples into groups (Simmonds, 1956). There were 7 qualitative characteristics, namely pseudostem height, pseudostem appearance, blotch at petiole base, blotch colour, petiole cannel leaf, wax on leaf, peduncle hairiness, bunch position, rachis position, male bud shape, bract imbrication, colour of the bract external face, colour of the bract internal face, wax on the bract, compound tepal basic colour, lobe colour of compound tepal, colour on the bract apex, colour stripe on the bract, fruit apex, and fusion of pedicel.

(4) Meanwhile, these 20 characteristics of 309 samples were standardized and analyzed by Hierarchical cluster analysis with inter-group linkage or average linkage between group or Unweighted Pair-Group Method using Arithmetic average, UPGMA (Vanitbuncha, 2000). The distance and similarity value were measured by Euclidean Distance (Anon, 1993). In

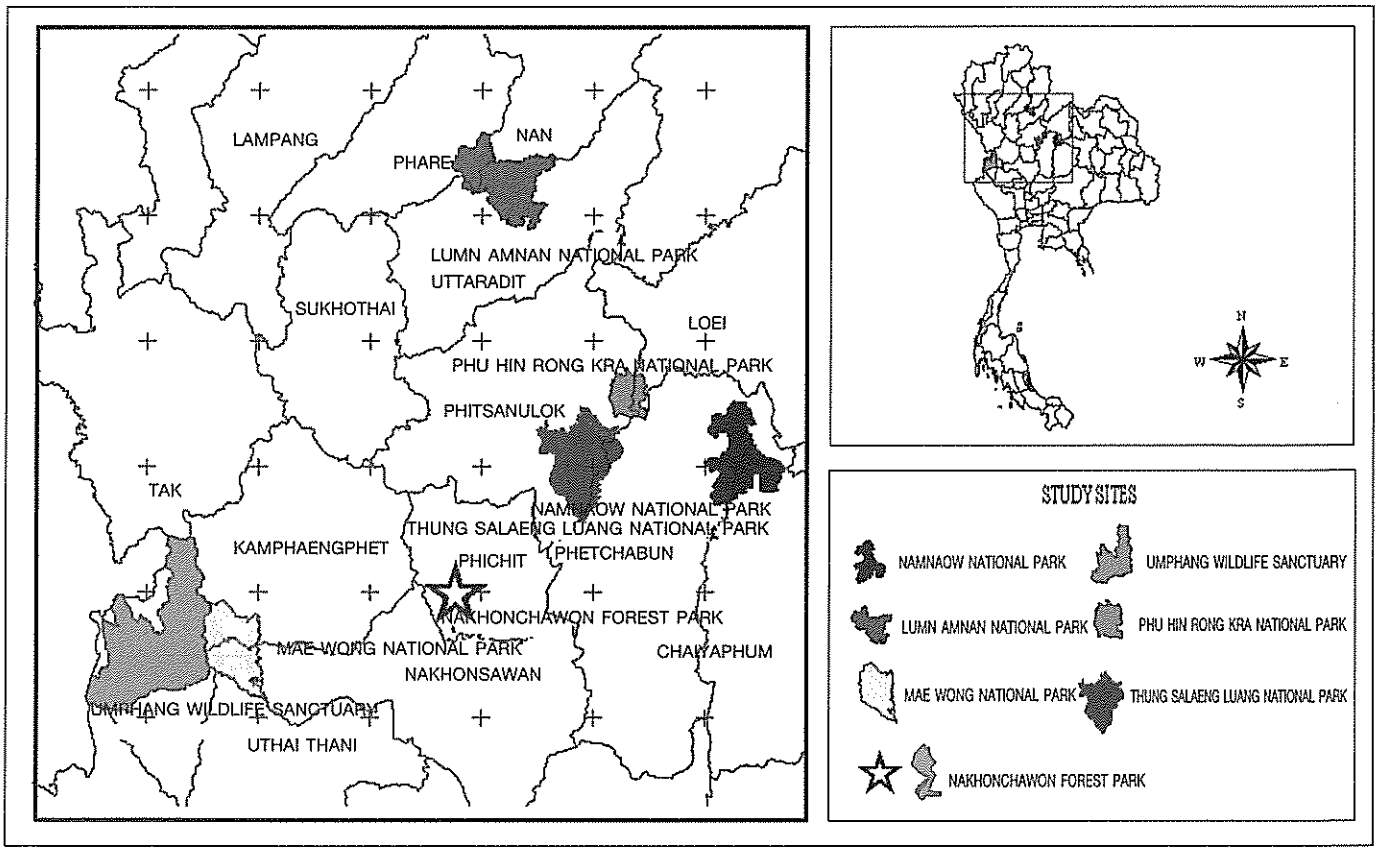


Figure 1. Map of lower northern Thailand shown the site locations of 5 national parks, a wildlife sanctuary, and a forest park

addition, Agglomerative Hierarchical cluster analysis or Agglomerative Schedule has been a method for cluster combination. The results were presented in Hierarchical dendrogram (Anon, 1993).

(5) K - means cluster analysis was also employed when the number of samples is higher than 200. This analysis was run and repeated as iteration which provided the table and member list in each cluster (Anon, 1993)

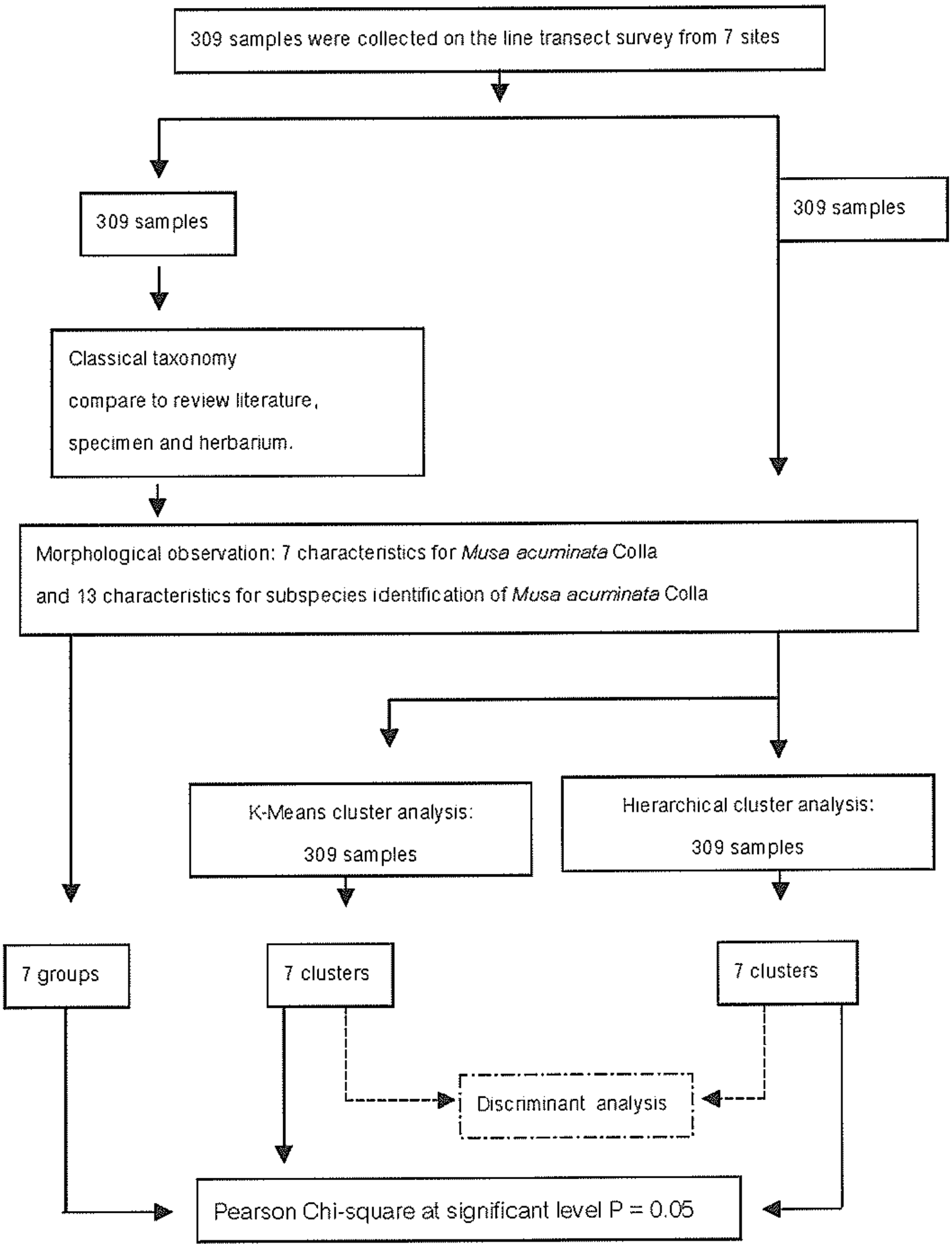


Figure 2. Flow - chart diagram of the research procedure

(6) Relationship or correlation of classical taxonomy with both cluster analysis techniques were statistically analysed by Pearson Chi-square (Anon, 1993).

(7) Discriminant analysis were tested the result form both cluster analysis with the original and cross-validated methods to accept the prediction of the cluster and grouping (Anon, 1993).

Table 1. Sampling locations and sample size of *Musa acuminata* at 7 sites

Site	Sample	Total number
Umphang Wildlife Sanctuary	up 1 - up 41	41
Thung Salaeng Luang National Park	tl 1 - tl 55	55
Mae Wong National Park	mw 1 - mw 49	49
Phu Hin Rong Kra National Park	ph 1 - ph 61	61
Lumnamnan National Park	ln 1 - ln 55	55
Namnao National Park	nw 1 - nw 48	48
Nakhonchaibowon Forest Park	cb$^{1/}$	0
Total -	309	

Remarks : 1/ sample was not found in habitat due to tiny and fragile forest (1,728 km^2) surrounded with agricultural farmland.

Table 2. Seven groups of wild banana (*M. acuminata*) in lower Northern Thailand and its members after being used the classical taxonomy in the different groups with different fronts.

Group 1	*Group 2*	Group 3	Group 4	Group 5	Group 6	Group 7
up5, up10, up16, mw2, mw12, mw33, mw35, mw38, mw41, ph2, ph5, ph30, ph31, ph33, ph34, ph43, ph47,ph48, ph50, ln21, ln23, ln27, ln30, ln46, nw7, nw22, nw24	mw10, nw41	up25, up26, up27, up28, up29, up30, up31, up32 up33, up34, up35,	up37, up40, up41	up1-up4, up6-up9, up12, up14, up19, up20, up22-up24, up38, up39, mw5,mw9, mw11, mw13-mw15, mw17- mw30, mw34,mw36, mw37, mw39, mw43, mw44, mw46, mw47, mw49, mw50, ph1, ph4, ph6-ph9, ph11-ph17, ph28, ph41, ln1, ln2, ln4, ln5, ln6, ln9- ln20, ln24, ln25-ln26, ln28, ln41, ln42, ln45, ln47-ln50, nw1, nw4, nw5, nw8, nw9, nw12-nw14, nw18-nw20, nw23, nw25-nw28, nw33-nw35, nw37, nw39, nw40, nw42, nw46, nw48, tl1-tl3, tl12-26, tl28-tl38, tl47, tl53, tl55, tl56	up36	up11,up13,up15, up17-18, up21, mw31, mw32, mw40, mw42, mw48, ph3, ph14, ph18, ph24-ph25, ph27, ph32, ph35-ph36, ph39-ph40, ph42, ph44-ph46, ph55-ph60, ln3, ln7, ln8, ln22, ln43-44, nw2-nw3, nw6, nw10, nw11, nw15-nw17, nw21, nw29, nw30-nw32, nw36, nw38, tl4-tl11, tl13, tl19, tl27, tl39-46, tl48-tl52, tl54
29	2	11	3	187	1	76

RESULTS AND DISCUSSION

The list of 309 samples were collected in 7 research sites (Table 1) and were morphologically identified by classical taxonomy (Simmonds, 1956; Simmonds and Shepherd, 1955; Simmonds and Weatherup, 1988) into 7 groups (Table 2).Twenty morphological characteristic data of each sample were analysed as factors of each case with Hierarchical cluster analysis by SPSS for Windows version 10.0. The pair distance and similarity values were calculated by Square Euclidean Distance of each pair-case (Table 3).

After being combined cluster within group linkage Agglomeration Schedule was conducted and was found that it the coefficiend value was smalles, next stages or next samples could be combined into the cluster (Table 4). In this Hierarchical cluster analysis process clearly provided a hierarchical dendrogram that could classified 309 samples into 7 clusters at 5 re-scaled distance units or at 80 % of similarity (Figure 3). Initial cluster centre values of 20 characteristics were employed to separate all samples into 7 clusters. Their values were the average of initial cluster centre values and standardised by K - means cluster analysis (Table 5). Then, the average mean of each cluster centers were recalculated in each iteration (cycle) until 36 iterations. The changes in all cluster centers were reached the final value at 0.000.

Table 3. The example of 18 banana samples presents the distance between pair-case and clusters after Square Euclidean Distance Technique.

Case	1:up1	2:up2	3:up3	4:up4	5:up5	6:up6	7:up7	8:up8	9:up9	10:up10	11:up11	12:up12	13:up13	14:up14	15:up15	16:up16	17:up17	18:up18
1:up1		23.697	12.953	25.910	20.455	34.964	22.586	27.885	23.946	20.902	10.542	26.125	51.375	33.313	23.476	19.564	14.937	11.467
2:up2	23.697		18.897	46.193	40.990	50.885	47.218	29.834	50.514	36.817	37.109	47.395	78.329	51.215	50.044	38.154	17.600	14.080
3:up3	12.953	18.897		29.961	24.757	45.309	33.619	32.187	34.291	25.892	20.887	36.470	70.078	37.615	33.821	24.555	16.289	12.769
4:up4	25.910	46.193	29.961		4.767	11.601	8.621	15.913	28.085	10.306	14.680	4.394	35.276	24.015	1.746	11.643	23.594	23.791
5:up5	20.455	40.990	24.757	4.767		12.877	9.209	11.146	29.360	4.851	15.956	5.670	30.997	19.249	3.021	6.198	23.007	23.203
6:up6	34.964	50.885	45.309	11.601	12.877		12.378	15.009	36.194	8.714	22.790	7.207	27.165	23.460	9.655	15.399	35.893	32.363
7:up7	22.596	47.218	33.619	8.621	9.209	12.378		11.342	32.527	4.358	19.123	3.539	28.789	16.770	6.188	3.021	33.592	30.072
8:up8	27.885	29.834	32.187	15.913	11.146	15.009	11.342		40.507	6.983	27.102	11.519	36.846	21.381	14.168	8.320	27.514	23.993
9:up9	23.946	50.514	34.291	28.085	29.360	36.194	32.527	40.507		34.211	13.405	28.988	54.315	21.800	26.339	35.548	26.498	26.694
10:up10	20.902	36.817	25.892	10.306	4.851	8.714	4.358	6.983	34.211		20.807	5.223	27.799	15.086	7.872	1.337	29.233	25.713
11:up11	10.542	37.109	20.887	14.680	15.956	22.790	19.123	27.102	13.405	20.807		15.583	40.910	35.204	12.934	22.144	13.093	13.290
12:up12	26.125	47.395	36.470	4.394	5.670	7.207	3.539	11.519	28.988	5.223	15.583		25.327	19.621	2.649	6.561	28.677	28.873
13:up13	51.375	78.329	70.078	35.276	30.997	27.165	28.789	36.846	54.315	27.799	40.910	25.327		47.622	27.976	31.810	48.042	54.836
14:up14	33.313	51.215	37.615	24.015	19.249	23.460	16.770	21.381	21.800	15.086	35.204	19.621	47.622		22.270	13.748	42.255	38.735
15:up15	23.476	50.044	33.921	1.746	3.021	9.655	6.188	14.168	26.339	7.872	12.934	2.649	27.976	22.270		9.209	26.028	26.224
16:up16	19.564	38.154	24.555	11.643	6.198	15.399	3.021	8.320	35.548	1.337	22.144	6.561	31.810	13.748	9.209		30.570	27.050
17:up17	14.937	17.600	16.289	23.594	23.007	35.893	33.592	27.514	26.498	29.233	13.093	28.677	48.042	42.255	26.028	30.570		3.520
18:up18	11.467	14.080	12.769	23.791	23.203	32.363	30.072	23.993	26.694	25.713	13.290	28.873	54.836	38.735	26.224	27.050	3.520	
19:up19	36.768	17.548	32.463	38.656	41.071	52.366	50.075	37.357	44.562	45.717	31.158	41.444	64.625	58.738	44.092	47.054	13.130	17.033

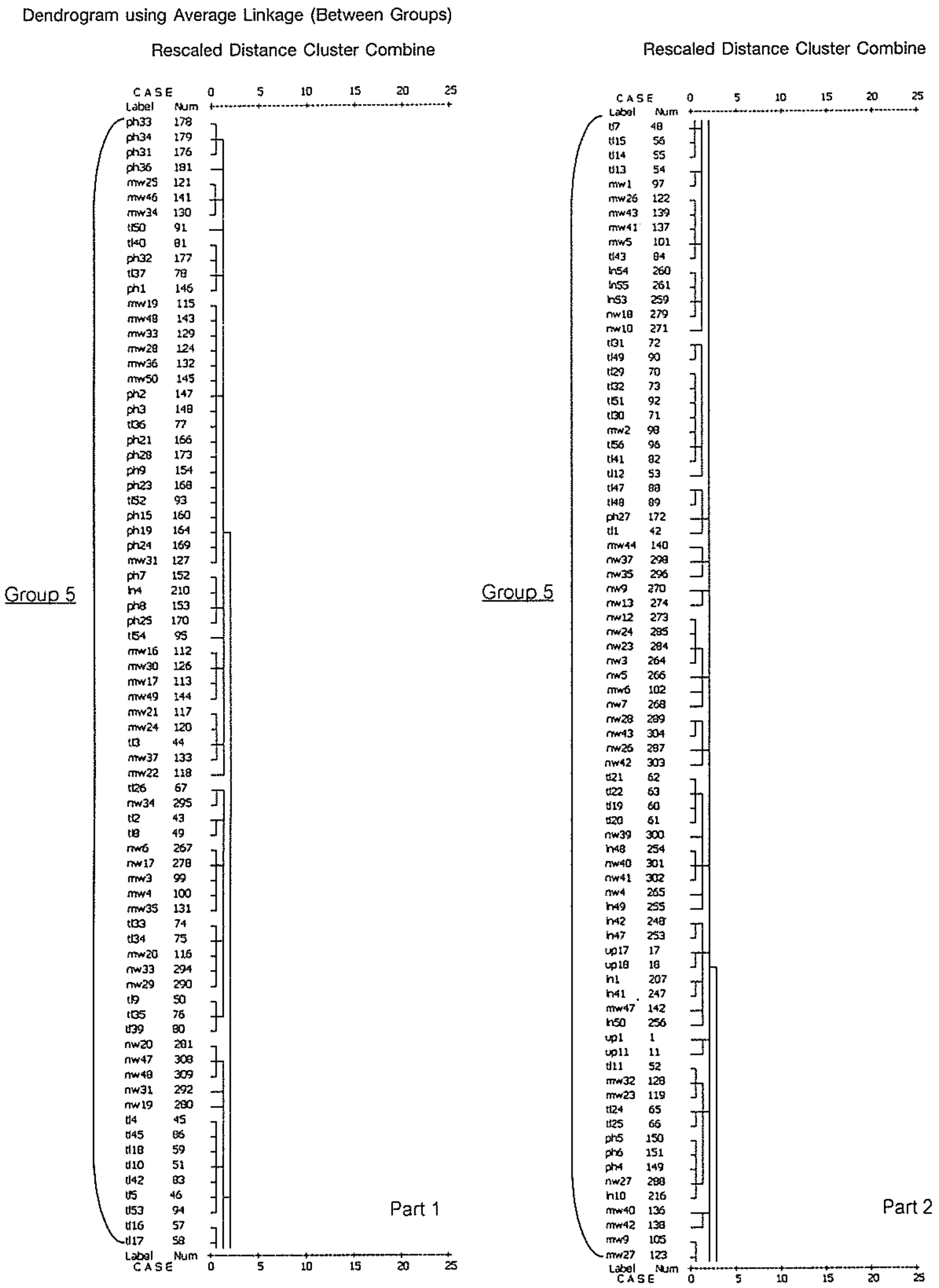
Remarks: higher distance coefficient values between each pair-case mean that higher separation levels of the case into the different cluster, if it is lower mean that cases should be combined together.

Table 4. Cluster combining of 309 wild bananas with between-group linkage of Agglomeration schedule : when the coefficient value was smaller, the combining cluster stage of each pair-case would comprise the next stages or the next samples would be combined into the same cluster.

Agglomeration Schedule

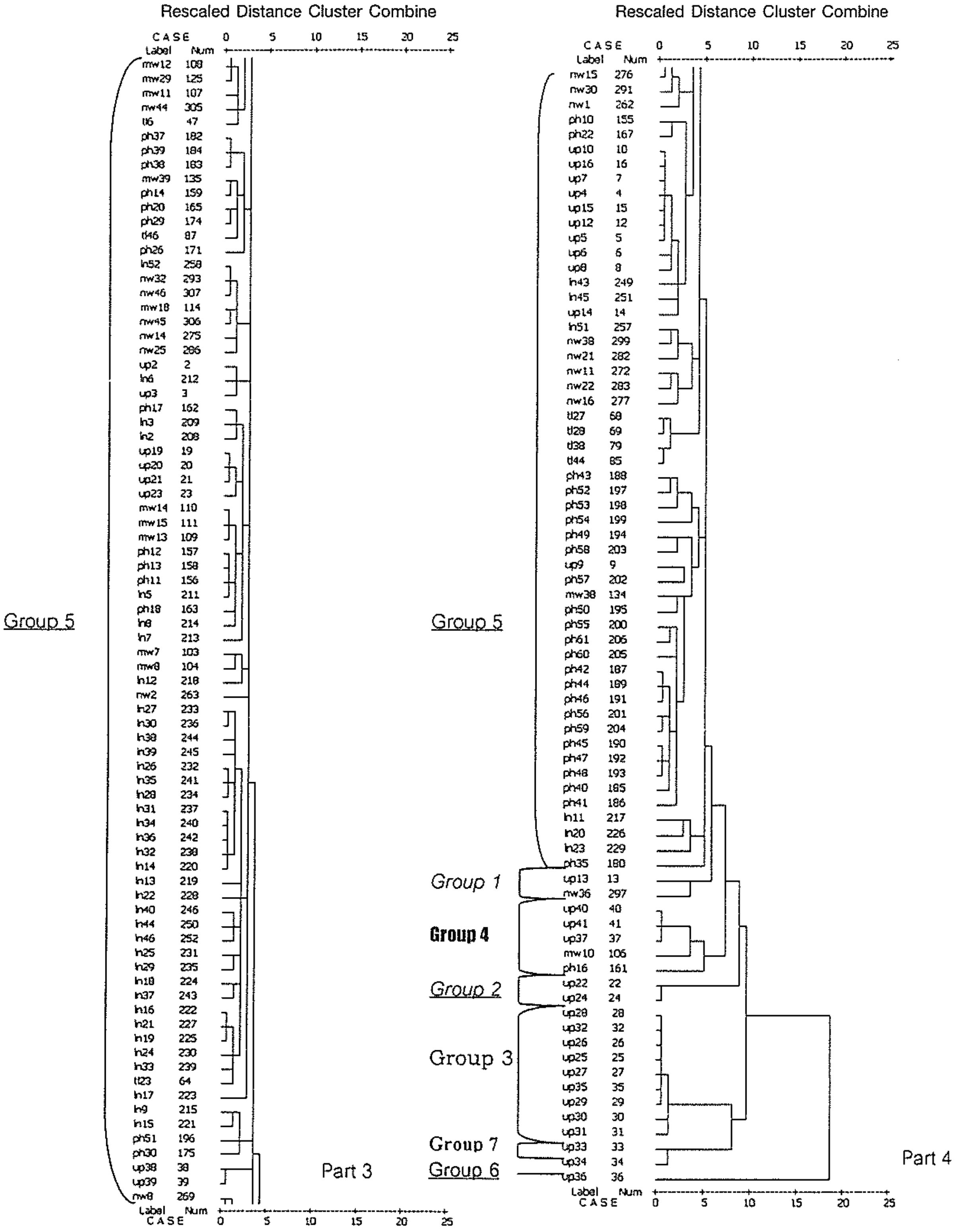
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	178	179	.000	0	0	81
2	166	173	.000	0	0	37
3	154	168	.000	0	0	18
4	93	160	.000	0	0	18
5	157	158	.000	0	0	6
6	156	157	.000	0	5	100
7	150	151	.000	0	0	48
8	147	148	.000	0	0	70
9	99	100	.000	0	0	99
10	68	69	.000	0	0	201
11	40	41	.000	0	0	12
12	37	40	.000	0	11	294
13	28	32	.000	0	0	14
14	26	28	.000	0	13	20
15	25	27	.000	0	0	20
16	19	20	.000	0	0	124
17	116	294	1.073	0	0	58
18	93	154	1.073	4	3	34
19	115	143	1.337	0	0	50
20	25	26	1.337	15	14	55
21	10	16	1.337	0	0	59
22	281	308	1.660	0	0	110
23	222	227	1.660	0	0	121
24	152	210	1.660	0	0	68
25	124	132	1.660	0	0	52
26	112	126	1.660	0	0	76
27	117	120	1.660	0	0	65
28	70	73	1.660	0	0	38
29	74	75	1.662	0	0	73
30	57	58	1.662	0	0	67
31	4	15	1.746	0	0	54
32	105	123	1.858	0	0	57
33	187	189	1.905	0	0	117
34	93	164	1.938	18	0	36
35	260	261	2.208	0	0	62
36	93	169	2.369	34	0	37
37	93	166	2.530	36	2	82
38	70	92	2.688	28	0	49
39	62	63	2.732	0	0	113
40	48	56	2.735	0	0	67
41	122	139	2.900	0	0	75
42	51	83	2.900	0	0	61
43	45	86	3.001	0	0	106
44	110	111	3.021	0	0	104
45	65	66	3.061	0	0	173
46	72	90	3.106	0	0	144
47	273	285	3.204	0	0	66
48	149	150	3.322	0	7	102
49	70	71	3.381	38	0	86
50	115	129	3.401	19	0	83
51	121	141	3.405	0	0	84
52	124	145	3.487	25	0	83
53	17	18	3.520	0	0	154
54	4	12	3.522	31	0	79
55	25	35	3.556	20	0	77

Remarks: this table shows present only 1-55 stages from 309 stages (samples) after Agglomeration Schedule of the Hierarchical cluster analysis.



Remarks: part 1, 2, 3, 4 are meaning that the consequence picture of the same dendrogram

Figure 3. Dendrogram of 309 wild banana (*M. acuminata*) in lower Northern Thailand using the Hierarchical cluster analysis



.Remarks: part 1, 2, 3, 4 are meaning that the consequence picture of the same dendrogram

Figure 3. Dendrogram of 309 wild banana (*M. acuminata* Colla) of wild banana in lower Northern Thailand by Hierarchical cluster analysis. (continue)

Table 5. Average mean value of initial cluster centers of 20 characteristic variables from 309 samples within the 7 clusters of wild banana (*M. acuminata*) in lower Northern Thailand.

Initial Cluster Centres

	Cluster						
	1	2	3	4	5	6	7
Zscore: Pseudostem Height	-.17066	1.11971	1.11971	1.11971	-1.46104	-1.46104	1.11971
Zscore: Pseudostem Appearance	.24049	24049	-4.14484	.24049	.24049	-4.14484	.24049
Zscore: Blotch at the Petiole Base	.53515	.53515	.53515	-1.53857	.53515	.53515	.53515
Zscore: Blotch Color	-2.42270	-1.42398	.57346	1.57218	.57346	.57346	.57346
Zscore: Petiole Cannal Leaf	1.68660	-2.40880	3.05173	.32147	-1.04366	-2.40880	-1.04366
Zscore: Wax on Leaves	1.05511	-.47332	4.11198	-.47332	-.47332	-.47332	-.47332
Zscore: Peduncle Hairiness	-.87796	-.87796	.49663	.49663	.49663	.49663	1.87122
Zscore: Bunch Position	-1.69383	.67677	-1.69383	.67677	.67677	1.86207	-2.87913
Zscore: Rachis Position	-.47814	-.47814	1.28643	-.47814	4.81557	6.58014	-.47814
Zscore: Male Bud Shape	-.56915	1.11120	1.11120	2.79155	-.56915	-.56915	1.11120
Zscore: Bract Imbrication	-1.46725	-1.46725	-1.46725	.15720	.15720	-1.46725	.15720
Zscore: Color of the Bract External Face	4.40829	-2.54037	-2.54037	.93396	-.22415	2.09207	-1.38226
Zscore: Color of the Bract Internal Face	-.05444	-.05444	-.05444	.52754	.52754	-1.80041	-2.96438
Zscore: Color on the Bract Apex	-.30208	-.30208	3.29967	3.29967	-.30208	-.30208	-.30208
Zscore: Color Stripes on Bract	-.30208	-.30208	-.30208	-.30208	-.30208	-.30208	3.29967
Zscore: Wax on the Bract	-.85894	.62861	-.85894	-.85894	.62861	-.85894	-.85894
Zscore: Compound Tepal Basic Color	1.10004	-.63982	-.63982	-.63982	-.63982	6.31962	1.10004
Zscore: Lobe Color of Compound Tepal	1.33559	-.46455	-.46455	1.33559	-.46455	1.33559	-.46455
Zscore: Fruit Apex	.15459	-6.23500	-6.23500	.15459	.15459	-6.23500	.15459
Zscore: Fusion of Pedicel	-.27071	1.02038	1.02038	-1.56181	-.27071	-1.56181	1.02038

This means that all samples are finally grouping into 7 clusters while the maximum distance between all initial centres is 8.636. Each cluster center of 7 clusters are far form the centre of the point with the same distance (Table 6).

The result of sample grouping under 36 iterations by K - means cluster analysis were also combined 309 samples into 7 clusters. Each cluster membership was combined the samples when have minimum distance from the cluster centre values (Table 7).

The K - means cluster analysis techniques also provided the average distance between each cluster centres (Table 8). For example, the result showed that the distance between cluster No 6 and cluster No 3 was 13.696. This mean that higher value were separated from other clusters.

The results of classical taxonomy, Hierarchical cluster analysis and K - Means cluster analysis of 309 samples as well as groups classification from 20 qualitative morphological data and their relationship were compared (Table 9) and tested with

Table 6. The average mean of cluster centres change until the end of 36 Iterations of wild banana (*M. acuminata*) in the lower Northern Thailand

Iteration history

	Change in Cluster Centres						
Iteration	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	6.127	.922	6.369	3.637	5.339	.000	1.959
2	2.504	.307	.531	5.713	.713	.000	.653
3	.318	.102	4.423E-02	.952	7.653E-02	.000	3.579
4	.361	3.416E-02	3.686E-03	.159	.165	.000	2.231
5	.971	1.139E-02	3.071E-04	2.645E-02	.225	.000	1.055
6	.441	3.796E-03	2.559E-05	4.408E-03	.123	.000	.519
7	1.840E-02	1.265E-03	2.133E-06	7.347E-04	2.599E-02	.000	.144
8	7.665E-04	4.218E-04	1.777E-07	1.224E-04	1.140E-04	.000	3.511E-03
9	3.194E-05	1.406E-04	1.481E-08	2.041E-05	4.999E-07	.000	8.563E-05
10	1.331E-06	4.686E-05	1.234E-09	3.401E-06	2.193E-09	.000	2.089E-06
11	5.545E-08	1.562E-05	1.029E-10	5.669E-07	9.617E-12	.000	5.094E-08
12	2.310E-09	5.207E-06	8.571E-12	9.448E-08	4.206E-14	.000	1.243E-09
13	9.626E-11	1.736E-06	7.143E-13	1.575E-08	3.986E-16	.000	3.031E-11
14	4.011E-12	5.786E-07	5.951E-14	2.624E-09	.000	.000	7.385E-13
15	1.662E-13	1.929E-07	4.852E-15	4.374E-10	.000	.000	1.794E-14
16	7.421E-15	6.429E-08	5.088E-16	7.290E-11	.000	.000	8.026E-16
17	6.794E-16	2.143E-08	.000	1.215E-11	.000	.000	.000
18	.000	7.143E-09	.000	2.025E-12	.000	.000	.000
19	.000	2.381E-09	.000	3.378E-13	.000	.000	.000
20	.000	7.936E-10	.000	5.642E-14	.000	.000	.000
21	.000	2.645E-10	.000	9.248E-15	.000	.000	.000
22	.000	8.818E-11	.000	1.384E-15	.000	.000	.000
23	.000	2.939E-11	.000	5.226E-16	.000	.000	.000
24	.000	9.798E-12	.000	1.110E-16	.000	.000	.000
25	.000	3.266E-12	.000	.000	.000	.000	.000
26	.000	1.089E-12	.000	.000	.000	.000	.000
27	.000	3.629E-13	.000	.000	.000	.000	.000
28	.000	1.209E-13	.000	.000	.000	.000	.000
29	.000	4.030E-14	.000	.000	.000	.000	.000
30	.000	1.356E-14	.000	.000	.000	.000	.000
31	.000	4.418E-15	.000	.000	.000	.000	.000
32	.000	1.520E-15	.000	.000	.000	.000	.000
33	.000	5.551E-16	.000	.000	.000	.000	.000
34	.000	1.388E-16	.000	.000	.000	.000	.000
35	.000	8.327E-17	.000	.000	.000	.000	.000
36	.000	.000	.000	.000	.000	.000	.000

a. Convergence achieved due to no or small distance change. The maximum distance by which any center has changed is .000. The current iteration is 36. The minimum distance between initial centers is 8.636.

Pearson Chi-square (Table 10). There were some relatively strong confound of member's interception in the cluster 1, 5 and 7 from three techniques (Table 9) due to the closer characteristic relationship among major subspecies of *M. acuminata* Colla (Figure 4-5).

However, there was non-significant difference among these three techniques on cluster or group separation (Table 10). The result from classical taxonomy and Hierarchical cluster analysis relationship testing with Pearson Chi-square was equal 28.000 but lower than $\chi^2_{(0.05,20)} = 36.42$. While, asymtotic significance (2 - sided) is equally 0.260 which higher than 0.05. Moreover, the results in cluster grouping from both techniques were non-significantly different at P = 0.05 (Table 10).

Secondly, classical taxonomy and K - means cluster analysis relationship testing with Pearson Chi-square is 42.000 but lower than $\chi^2_{(0.05,30)} = 50.964$. While asymtotic significance (2 - sided) is 0.227 which was higher than 0.05. This results showed that

Table 7. Demonstration on the cluster number of samples and minimum distance from their cluster centre values with only 62 examples from 309 wild banana samples

Cluster Membership

Case Number	Short name	Cluster	Distance
1	up 1	5	4.326
2	up 2	5	4.341
3	up 3	5	4.234
4	up 4	1	2.596
5	up 5	1	2.538
6	up 6	1	3.599
7	up 7	1	2.451
8	up 8	1	2.832
9	up 9	1	4.608
10	up 10	1	2.387
11	up 11	1	4.020
12	up 12	1	2.203
13	up 13	1	5.173
14	up 14	1	3.221
15	up 15	1	2.564
16	up 16	1	2.424
17	up 17	5	2.712
18	up 18	5	3.046
19	up 19	5	4.130
20	up 20	5	4.130
21	up 21	5	3.716
22	up 22	2	1.384
23	up 23	5	4.027
24	up 24	2	1.384
25	up 25	3	1.838
26	up 26	3	1.588
27	up 27	3	1.838
28	up 28	3	1.588
29	up 29	3	2.275
30	up 30	3	2.643
31	up 31	3	3.248
32	up 32	3	1.588
33	up 33	3	6.948
34	up 34	3	6.136
35	up 35	3	2.109
36	up 36	6	.000
37	up 37	4	2.503
38	up 38	5	4.988
39	up 39	5	5.344
40	up 40	4	2.503
41	up 41	4	2.503
42	tl 1	5	3.328
43	tl 2	5	3.214
44	tl 3	5	2.832
45	tl 4	5	2.327
46	tl 5	5	2.531
47	tl 6	7	4.132
48	tl 7	5	3.162
49	tl 8	5	2.972
50	tl 9	5	2.556
51	tl 10	5	2.781
52	tl 11	5	2.876
53	tl 12	5	3.207
54	tl 13	5	2.864
55	tl 14	5	3.262
56	tl 15	5	2.689
57	tl 16	5	2.874
58	tl 17	5	2.779
59	tl 18	5	2.687
60	tl 19	5	2.517
61	tl 20	5	3.576
62	tl 21	5	3.549

the result from classical taxonomy and K-means cluster analysis techniques were non-significant different in clustering and grouping these wild banana samples (Table 10).

Finally, relationship of identification between Hierarchical cluster analysis and K-means cluster analysis techniques were tested with Pearson Chi-square is equal 28.000 but also lower than $\chi^2_{(0.05,30)}$ = 36.42. While, asymtotic significance (2 - sided) was equal 0.260 which was higher than 0.05. These confirmed that there was non-significantly different (P=0.05) between both techniques (Table 10).

Where results of both cluster analysis or numerical morphological classification are synchronized and non-

Table 8. Distances between the final Cluster Centres after K - means cluster analysis of 309 wild banana samples with 20 characteristics.

Cluster	1	2	3	4	5	6	7
1		8.961	8.853	7.166	3.903	13.430	4.252
2	8.961		10.824	10.574	7.792	12.812	8.241
3	8.853	10.824		9.894	8.435	13.696	7.632
4	7.166	10.574	9.894		6.934	12.017	6.074
5	3.903	7.792	8.435	6.934		13.165	2.992
6	13.430	12.812	13.696	12.017	13.165		13.224
7	4.252	8.241	7.632	6.074	2.992	13.224	

Table 9. Comparison on the number of sample frequency separated into 7 clusters under classical taxonomy namely Hierarchical cluster analysis and K - means cluster analysis techniques.

Techniques	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Classical taxonomy	29	2	11	3	187	1	76
Hierarchical cluster analysis	2	2	9	5	288	1	2
K-means cluster analysis	25	2	11	5	219	1	46

Table 10. Relationship testing with Pearson Chi-square between three techniques on morphological classification of wild banana *M. acuminata* of the lower Northern Thailand into 7 clusters.

Comparison detail	Pearson Chi - square	Degree of freedom	Asymtotic significant	$\chi^2_{(0.05,df)}$
(1) Classical taxonomy χ Hierarchical cluster analysis	28.000 (Ns)	24	0.260	36.42
(2) Classical taxonomy χ K - means cluster analysis	28.000 (Ns)	36	0.277	50.964
(3) Hierarchical cluster analysis χ K - means cluster analysis	42.000 (Ns)	24	0.260	36.42

NS = Non-significant different at P = 0.05

Table 11. Discriminant analysis on the grouping results from both cluster analysis techniques with the original and cross-validated methods, the percentage values indicated the acceptance on separation predicting of the 7 clusters.

Discriminant analysis	Cluster1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Original-validated (%)	87.5	100	100	100	95.9	100	93.5
Cross-validated (%)	83.3	100	100	100	94.0	100	93.5

significantly different from the result of Classical taxonomy with Pearson Chi-square test (P = 0.05) in the Table 10.

In addition, discriminant analysis shown the probability of precise separation and prediction into 7 groups of both cluster techniques at 83-100% with the original and cross-validated methods (Table 11).

In conclusion, classical taxonomy, Hierarchical and K-mean cluster analysis techniques were convenient and suitable for employing as standard classification method for *M. acuminata* Colla in the lower Northern Thailand. Furthermore, 20 major qualitative characteristics were effectivly and precisly to identify wild banana in these groups as cross check by Discriminant analysis with higher value of the original and cross-validated methods. After being critical analyed the data into 7 groups from classical taxonomy, there are also highly subgroup characteristic variations in the group 5. Further research should be focus on this confound. However, primary identification key and cluster classification was finalized. In conclusion, sample in group 6 is highly different from *M. acuminata* Colla with its erect rachis Position, and should be a new wild banana group. It is very interesting to focus on re-identification of this sample through banana taxonomic system. The prominent strong convolute with significant 1/3 bright light green tip of male bud of the group 2 which is also interesting and might be a new promising subspecies of of *M. acuminata* Colla. Moreover, the samples in group 1, 3, 4, 5, 7 indicated strong characteristics of the subspecies of *M. acuminata* Colla spp. *malaccensis* (Ridl.) Simmonds, *malaccensis* (Ridl.) Simmonds, *malaccensis* (Ridl.) Simmonds, *siamea* Simmonds, and *burmannica* Simmonds, respectively. Most of samples from 7 groups are also presented the potential characters in genetic and breeding material for future utilization.

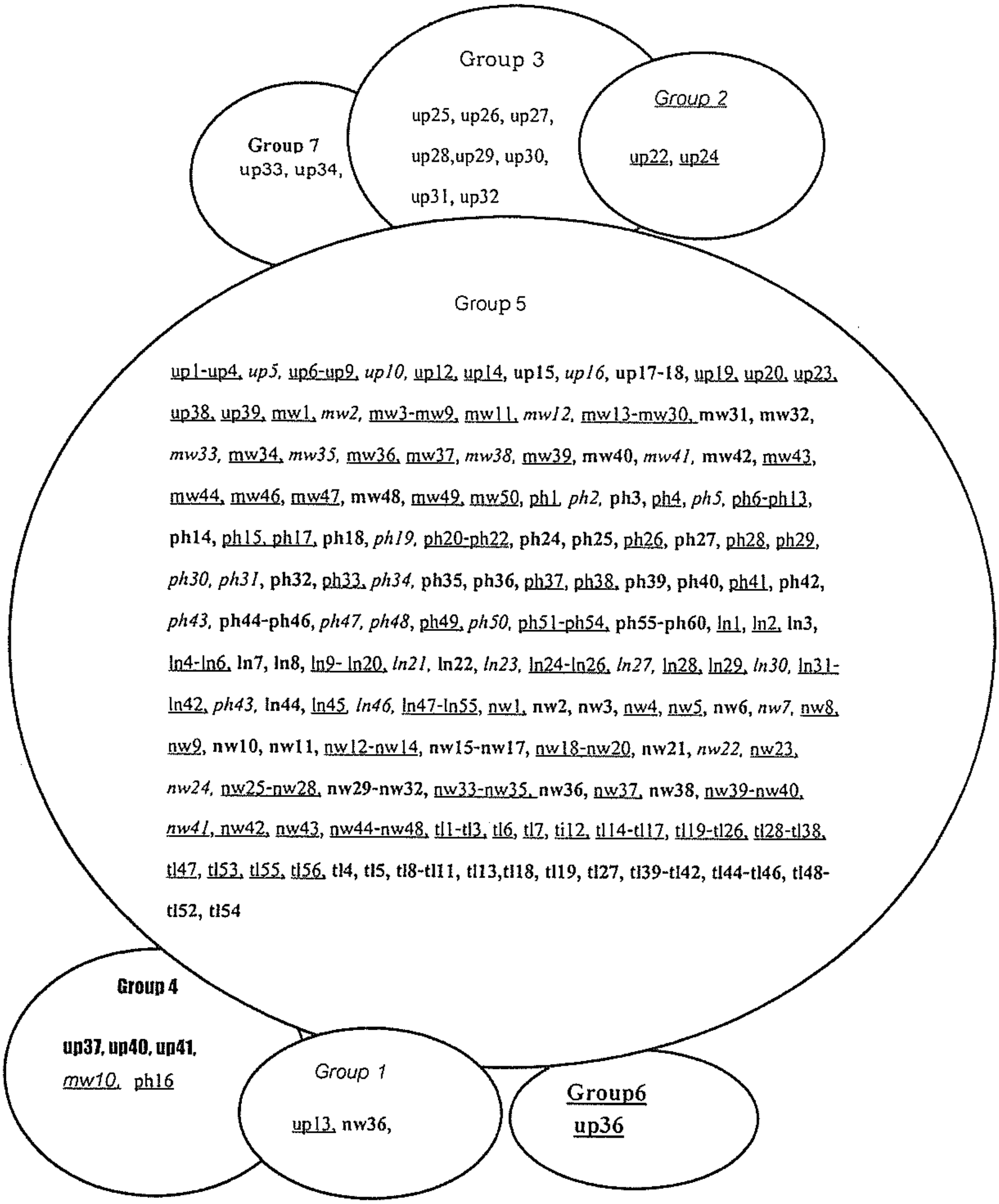


Figure 4. The member of 7 groups after Hierarchical cluster analysis of 309 wild bananas (*M. acuminata*) in lower Northern Thailand, the group members with different in front are originally of from the classical taxonomy groups.

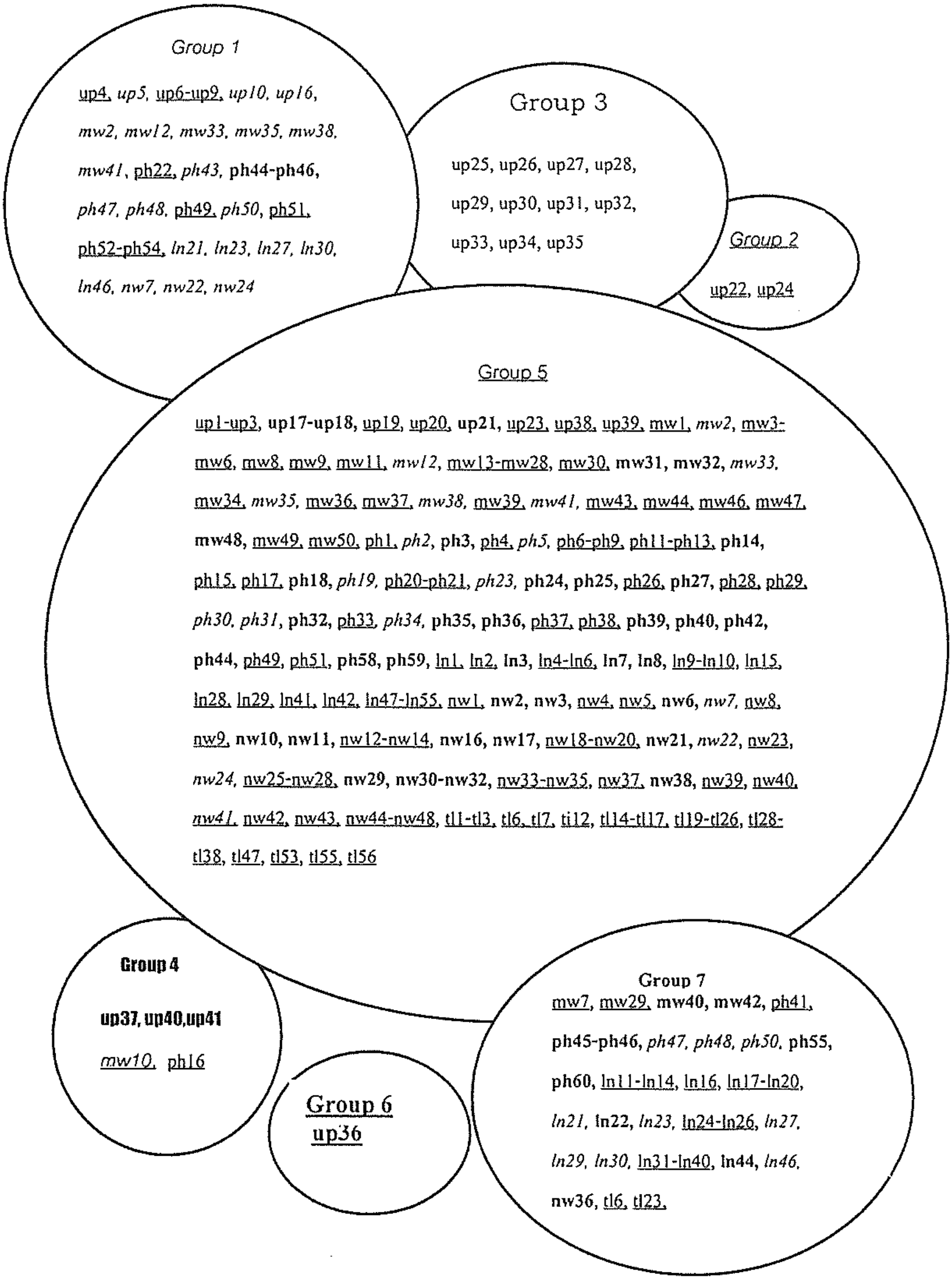
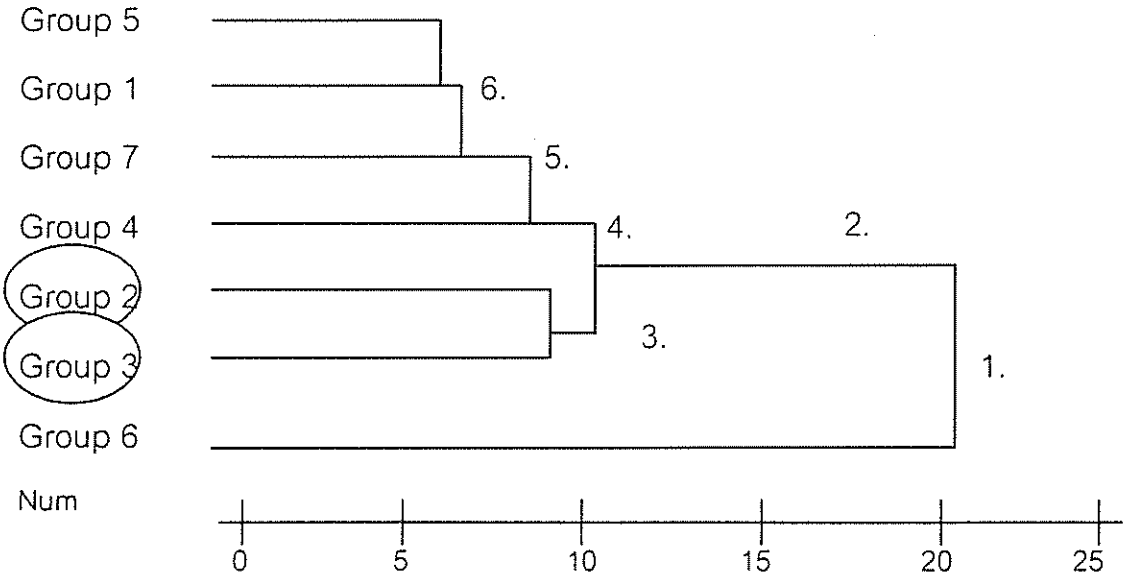


Figure 5. The members of 7 groups after K-means cluster analysis of 309 wild banana (*M. acuminata*) in lower Northern Thailand, the group members with the different in front are originally from of the different classical taxonomy groups.



Remarks: Identical sharing character within the Group 2 and Group 3 is waxy or very waxy on the leaf.

Figure 6. Dendrogram of 7 clusters (groups) with the identification and classification criteria

The identification key of *Musa acuminata* Colla in the lower Northern Thailand into 7 groups:

1. Rachis position: erect........................ Group 6
1. Rachis position: at an angle, horizontal or falling........................2.
    2. Wax on leaf: vary wax........................3.
    3. Male bud shape: lanceolate, colour of the bract external face: red-purple, colour of the bract internal face: pink-purple........................ Group 1
    3. Male bud shape: like a top, colour or the bract external face: red or red-purple, colour of the bract internal face: orange-red........................ Group 2
2. Wax on leaf: absent or visible, few wax, moderate........................4.
    4. Male bud shape: lanceolate, colour of the bract external face: red-purple, colour of the bract internal face: pink-purple........................ Group 4
    4. Male bud shape: lanceolate, ovoid, intermediat, like a top, colour of the bract external face: red, red-purple, purple-brown, purple, blue, colour of the bract internal face: yellow or green, orange-red, red, purple, red-purple, purple-brown, pink-purple........................5

5. Lobe colour of compound tepal: bright yellow.. Group 7

5. Lobe colour of compound tepal: yellow or orange..6

6. Color stripe on the bract: with discolour male bud shape: lanceolate and colour of the bract internal face : yellow or green, male bud shape: ovoid and colour of the bract Internal face: orange-red.. Group 3

6. Colour stripe on the bract: with discolour or without discolour, male bud shape: lanceolate, ovoid, like a top, colour of the bract internal face: yellow or green, orange-red, red, purple, red-purple, purple-brown, pink-purple, colour of the bract external face: red, red-purple, purple-brown, purple, blue... Group 5

References

Anon. 1997. *Musa* Descriptor. INIBAP/ IPGRI/ CIRAD, France. 75 p.

Anon. 1993. SPSS for Windows base system User's guide Release 6.0. Marija J. Norusis/ SPSS Inc. 828 p.

Argent, G. 2000. Two interesting wild *Musa* species (Musaceae) from Sabah, Malaysia. *Gardens' Bulletin Singapore* 52:203-210.

Haines-Young, R., Green, D. R. and S. H.,Cousins. 1993. Landscape ecology and geographic information systems. Taylor & Francis, London. 288 p.

Nasution, R. E. 1993. Rediscovery of two wild seeded bananas of Indonesia. *INFOMUSA*. 2(2) : 16-18.

Rekha, A.; Ravishankar, K. V., Anand, L. and S. C. Hiremath. 2001. Genetic and genomic diversity in banana (*Musa species and cultivars*) based on D^2 analysis and RAPD markers. *INFOMUSA*. 10(2) : 29-34.

Silayoi, B., and N. Chomchalow, 1983. Banana genetic resources exploration in Thailand. Report Submitted to IBPGR. Kasetsart University, Bangkok.

Simmonds, N. W. 1956. Botanical Results of the Banana Collecting Expedition, 1954-5. *Kew Bulletin*. 11: 463-489.

Simmonds, N. W. and K., Shepherd. 1955. The taxonomy and original cultivated banana. *J. Linn. Soc. (Botany)* 55: 302-312.

Simmonds, N. W. and S. T. C., Weatherup. 1988. Numerical taxonomy of the cultivated bananas. *Trop. Agric. (Trinidad)* 67(1): 90-92.

Sutherland, W. J. 1996. Ecological census techniques: A handbook. Cambridge University Press. 336 p.

Suvatti, C. 1978. Flora of Thailand. Royal Institute: Thailand. 394 p.

Swennen, R. and D., Vuylsteke. 1986. Morphological taxonomy of plantain (*Musa cultivars AAB*) in West Africa. Pages 165-171. *In* : Banana and Plantain Breeding Strategies : *Proceedings of an International Workshop*, Australia, 13-17 October, 1986.

Valmayor, R. V. 2001. Classification and Characterization of *Musa exotica*, M. *alimsanaya* and *M. acuminata* ssp. *errans*. *INFOMUSA* 10(2): 35-39.

Valmayor, R. V. and O. C., Pascua. 1985. Southeast Asian banana and plantain cultivar names and synonyms: A preliminary list. *IBPGR/ASEAN* 9(1): 9-10.

Vanitbunch, K. 2000. Multiple Variable Analysis with SPSS for Windows. Chulalongkorn University. 280 p.