Theor Appl Genet (1992) 85:394–402

THEORETICAL AND APPLIED GENETICS

Hierarchical clustering of 54 races and strains of the mulberry silkworm, *Bombyx mori* L: Significance of biochemical parameters

S. N. Chatterjee and R. K. Datta

Genetics Laboratory, Central Sericultural Research and Training Institute, Mysore-570008, Karnataka, India

Received May 22, 1992; Accepted June 19, 1992
Communicated by E. J. Eisen

Summary. A detailed analysis was undertaken to test the efficacy of hierarchical agglomerative clustering (UPGMA method) in grouping the races and strains of the mulberry silkworm, *Bombyx moti* L., and to ascertain the importance of biochemical parameters in the clustering process. The analysis was based on data from two rearing seasons with 54 selected races/strains of different geographic origin and varying yield potentials. The results indicate that seven clusters can be realised with yield parameters alone, whereas the inclusion of biochemical parameters in clustering resulted into two broad groups: one having all the breeds with high cocoon weight and shell weight, the other having all the low-yielding silkworm strains both from India and from other countries. Further sub-grouping under these two groups highlights genetical differences associated with the differentiation of various groups of races in temperate and tropical areas as well as their significance for silkworm breeding. Estimates of all ten variables were further subjected to 'quick clustering' and the results showed that cluster 5, constituted by 38 low-yielding strains of India, China and Europe, had the highest values of the final cluster centre for amylase and the effective rate of rearing (ERR), while clusters 1 and 4 had the highest values for invertase and alkaline phosphatase. The evolutionary aspect of the genetic channelisation of silkworm races from various countries is discussed against the background of differences in the biochemical parameters and yield variables.

Key words: Hierarchical clustering – *Bombyx mori* – Biochemical and yield parameters

Correspondence to: S. N. Chatterjee

Introduction

During the last 50 years, genetic material of *Bombyx mori* from Japan, China and Russia has spread to many other sericultural countries and is currently being used as a base material for breeding work in India and elsewhere. During the breeding process, a large number of lines have been isolated and stabilised whose maintenance and utilisation in India warrants proper classification and listing. Such classification is also important for breeding work aimed at improving adaptability and other yield components (Spagnoletti Zeuli and Qualset 1987). Attempts made so far in India for such a genetic classification have been sporadic and made solely on the basis of Mahalanobis D^2 distances (Rao et al. 1989; Jolley et al. 1989). In addition to Mahalanobis D^2 analysis, computer algorithms developed in the field of multivariate statistics have been used by plant breeders for classifying germplasm collections (Murphy et al. 1986) and analyzing host-parasite interactions (Lebeda and Jendrulek 1987). Currently, more and more studies are being undertaken to test the suitability of these methods of analysis for handling large collections of materials and for utilising both qualitative and quantitative information (Peters and Martinelli 1989).

Against this background, an attempt was made to use the method of hierarchical clustering for deciphering genetic similarity among silkworm breeds. The objectives of the present study were: (1) to test the effectiveness of 'hierarchical clustering' in classifying silkworm strains and (2) to test whether biochemical parameters provide more useful information for clustering the genetic stocks of silkworm. The biochemical parameters are considered because of the recent information generated on their correlation or interaction

with yield attributes (Gamo 1983; Moon and Seol 1983; Chatterjee et al. 1988, 1989).

Materials and methods

Silkworm races/strains

A Total of 54 stocks maintained in the Genetics Laboratory, CSRTI, Mysore, were used in this study, and they can be described under the following groups:

Group A. Indian races known for more than 100 years: Nistari, Sarupat, Moria, Boropalu, Pure Mysore, C'nichi. Of these, Boropalu is characterised by hibernation at the egg stage and the formation of both yellow and white cocoons (Mukherjee 1912). Other breeds are multivoltine. All of these breeds are low yielding (poor cocoon characters) but are highly adaptable to tropical climates.

Group B. Races received from China during 1984–85 and characterised by poor cocoon characters but laying hibernating eggs. The races are Rong Dazao, Dong 34, Nong 51, Hu 204, and they show higher adaptability to tropical conditions.

Group C. Multivoltine strains evolved in India, using traditional Indian races (e.g., Nistari and Pure Mysore) and high-yielding bivoltine breeds from Japan or China. They are Kolar Gold, Kollegal Jawan, A14Dy, AW8a, PA1, etc.

Group D. Genetic stocks (or their segregrants) characterised by high-yielding cocoon characters. These were received at various times from Japan (J122, G92, G73, NB1, C124, Eq plain), Europe (AWC, French plain, French batik etc.), Brazil (JZHpo), and Korea (NBHpo, NBJHpc).

Group E. Bivoltine breeds evolved in India (NB18, NB4D2, KA, NN6D etc.) or their sublines derived from Japanese/Chinese high-yielding bivoltine and mutlivoltine stocks or others.

Rearing of silkworms

Rearing was done during two seasons (January–February and April–May, 1990) following the standard procedure described by Krishnaswami (1978) and Narasimhanna (1968). The record of temperature and humidity in the rearing house during these months shows clearly that the latter period was distinctly warmer and less humid than the former. For each race/breed, a minimum of ten layings (total number of eggs laid by ten mother moths) was brushed *en-masse* and, after the IIIrd ecdysis, divided into three replications with 450 larvae for each replication.

Estimation of variables

The variables considered were four yield and six biochemical parameters. The yield attributes were the weight of ten mature larvae (g), the weight of a single cocoon (mg), cocoon shell weight (mg), and the effective rate of rearing, a measure of survival. The effective rate of rearing was estimated from the number of cocoons formed by 450 larvae for each replication. The mean weight of five male and five female cocoons and cocoon shells for each replication was considered as the weight of a single cocoon and cocoon shell.

The biochemical estimates used are the activity of amylase, invertase, protease (at neutral and alkaline pH), alkaline phosphatase in digestive juice, and the trehalose content of the blood. The techniques used for estimating the activity of the enzymes and the trehalose content are listed in Table 1. The biochemical estimation was carried out separately for each replication and each replication mean was based on three estimates from 15 larvae. Means of all the estimates for the two seasons were used for the present analysis.

Clustering methodology

Agglomerative hierarchical clustering was done using SPSS/PC+ software (Micro Software Dept., SPSS Inc., Chicago Ill.) employing the method of average linkage between groups (Romesburg 1984) under UPGMA (Unweighted Pair-Group Method using Arithmatic Averages). The clustering was based on the squared Euclidean distance, with Distance $(X,Y) = \Sigma_i(Xi - Yi)^2$. The average linkage between two groups is taken as the average of the distance between all pairs of cases with one member from each group. Hierarchical clustering analysis was carried out by considering: (1) yield attributes alone, (2) four yield attributes and two biochemical parameters, and (3) all ten parameters together.

Amylase and invertase were included separately with yield because both are important carbohydrases in the mulberry silkworm (Horie 1961) and an earlier study revealed that amylase was positively correlated with the effective rate of rearing (ERR), while invertase activity was negatively correlated with the same yield parameter (ERR).

The co-phenetic correlation (Sneath and Sokal 1973) between the original pairwise distances and those obtained from the dendrograms was worked out for three separate hierarchical clusterings to measure the fit of the dendrograms to the original data.

Finally, quick clustering, based on the K-mean method of MacQueen (1967), was used to group all 54 races into five clusters. The estimates of ten variables were subjected to Z-transformations prior to the analysis for quick clustering. The quick clustering method adopted initial clustering as the first step followed

Table 1. References for biochemical estimation methods used in the present study

Parameter	Reference	Substrate	Expressed
Amylase	Bernfeld (1955)	Starch	μg of maltodextrin released per 30 min per 20 μl of sample
Invertase	Ishaaya and Swirski (1970)	Sucrose	μg of glucose released per hr per 10 μl sample
Protease pH7 and pH10	Eguchi et al. (1972)	Casein	μg of L-tyrosine released per 15 min per 50 μl of sample
Alkaline phosphatase	Mihara et al. (1988)	P-Nitro-phenyl	μg of alkaline phosphatase per 50 μl phosphate of sample per 30 min
Trehalose	Roe (1955)		μg of trehalose per 10 μl of blood

by the grouping of all cases into five clusters, after which the classification cluster centres were estimated. Based on the 'classification cluster centres', the classification of members was further improved and 'final cluster centres' were obtained. Later, Euclidean distances between the 'final cluster centres' were worked out. The programme also has scope for variance analysis to compare variability between clusters against that realised within clusters.

Results

The ranges of variability of the ten variables are presented in Table 2. Among biochemical parameters, variability was maximum for amylase activity (CV = 0.53) followed by alkaline phosphatase (CV = 0.39) and invertase (CV = 0.30) activity in the digestive juice, whereas protease activity showed the minimum (CV = 0.14) variability.

Clustering analysis with yield attributes

Classification on the basis of 'yield attributes' by the UPGMA method led to the formation of seven clusters (Fig. 1). The three 'nearest neighbour' pairs were found to be G-race↔A4 (distance = 73.07), Pure Mysore↔

Table 2. Mean and range estimates of the ten variables (six biochemical and four yield attributes) in 54 strains (average of two crops) used for clustering analysis

Estimate	Variable									
	AML	TRE	PRO7	PRO10	INV	ALK	LWT	CWT	SWT	ERR%
Mean	470.42	111.74	8.92	10.17	550.29	17.88	24.85	1090.79	180.23	80.66
SD± *	250.97	28.37	1.23	2.48	166.41	6.92	5.95	221.84	58.19	10.25
Maximum	939.70	210.55	11.91	21.91	948.75	44.50	39.23	1544.00	311.00	93.84
Minimum	34.18	42.15	6.08	7.14	169.30	5.27	14.78	662.50	84.00	37.20
CV	0.53	0.25	0.14	0.24	0.30	0.39	0.24	0.20	0.32	0.13

AML, Amylase; TRE, Trehalose; PRO7, Protease (pH7); PRO10, Protease (pH10); INV, Invertase; ALK, Alkaline Phosphatase; LWT, Larval wt (g); CWT, Cocoon wt (mg); SWT, Shell wt (mg); ERR %, Effective rate Rearing.
* SD was calculated on the basis of 54 entries for 54 strains

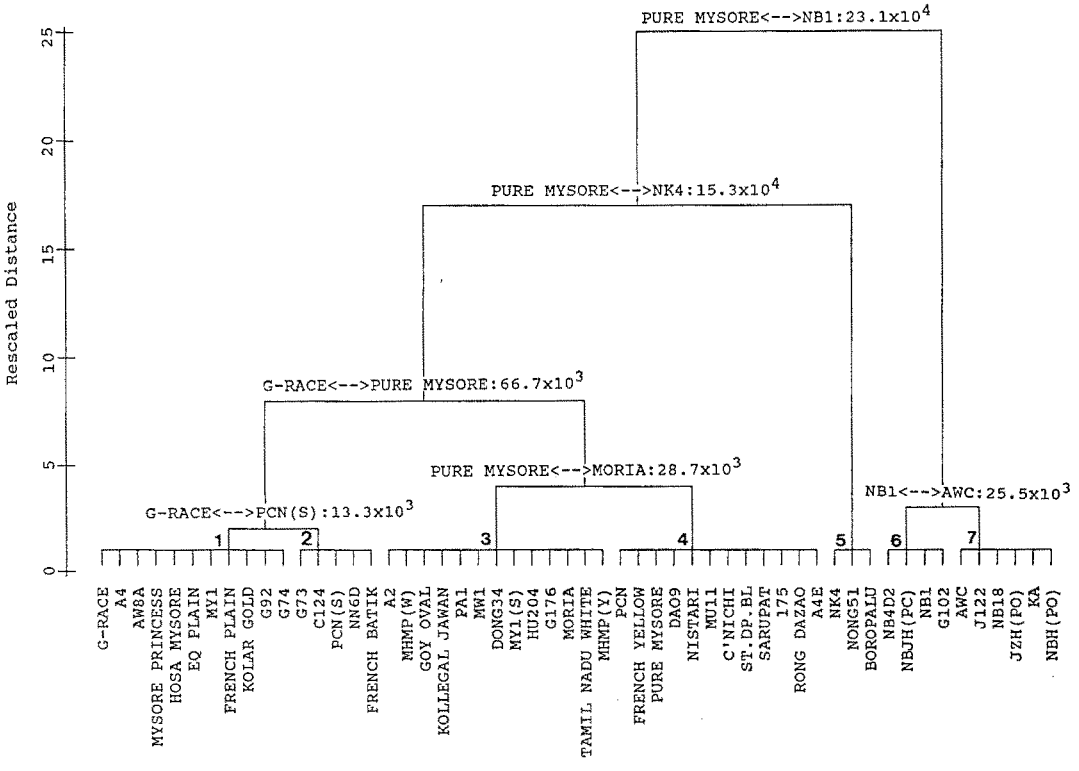

Fig. 1. Dendrogram obtained from the hierarchical clustering analysis of average estimates for four yield attributes from 54 silkworm strains (see Materials and methods) over two seasons. The 'rescaled distance' indicates rescaling of maximum distance realised to the scale of 0 to 25

Dao9 (81.92) and My1↔Hu 204 (91.75). The maximum distance realised was between Pure Mysore↔NB1 (23.1×10^4), preceded by Pure Mysore↔NK4 (15.3×10^4). Other important observations apparent from the dendrogram (Fig. 1) are: (1) the clustering together of NK4, Nong51 and Boropalu, with distances ranging from 3.1×10^3 (NK4↔Boropalu) to 1.9×10^3 (NK4↔Nong51), and (2) the clustering of the high-yielding breeds NB4D2, NBJH(pc), NB1, G102 and of AWC, J122, NB18, JZH(po), KA, NBH(po) into two clusters with the distance between NB1 and AWC being 25.5×10^3. The co-phenetic correlation index was high ($^rcs = 0.901$).

Clustering on the basis of yield attributes, amylase and invertase

It is evident from Fig. 2 that the inclusion of amylase and invertase as variables resulted in a more complex dendrogram with five subgroups (a–e) classified into two major groups: A and B (see Fig. 2 and Table 3). Of these small clusters, the minimum distance was found between NB18↔NBH(po) (2.57×10^3) followed closely by Moria↔PCN (2.9×10^3) and C'nichi↔Staple deep black (3.0×10^3). The addition of the two biochemical parameters (amylase and invertase) has broken the association of NK4 with Nong 51 and Boropalu,

Table 3. Clustering on the basis of yield attributes, amylase and invertase

Group A:	
Subgroup a	
Cluster-1	NB18, NBH(po), KA, AWC, NBJH(pc), G73
Cluster-2	NB1 and J122
Subgroup b	
Cluster-1	PCN(s), G102, French batik, C124, NN6D, Eq plain, AW8a
Cluster-2	G176, Hu 204, French Plain, Dong 34, G74
Cluster-3	French Yellow
Subgroup c	
Cluster-1	NB4D2, JZH(po)
Cluster-2	G92
Group B:	
Subgroup d	
Cluster-1	Pure Mysore, MU11, Nistari, MHMP(y), Goy oval, Kollegal Jawan, A4, Mysore Princess, MW1, PA1, A2, MHMP(w), Tamil Nadu White, MY1
Cluster-2	G-race, Hosa Mysore, Kolar Gold
Cluster-3	A4E
Subgroup e	
Cluster-1	NK4, Rong Dazao
Subgroup f	
Cluster-1	Nong51, Boropalu
Cluster-2	C'nichi, Staple deep black (ST. DP. BL.), Sarupat, MY1(S)
Cluster-3	Moria, PCN, Dao9, 175

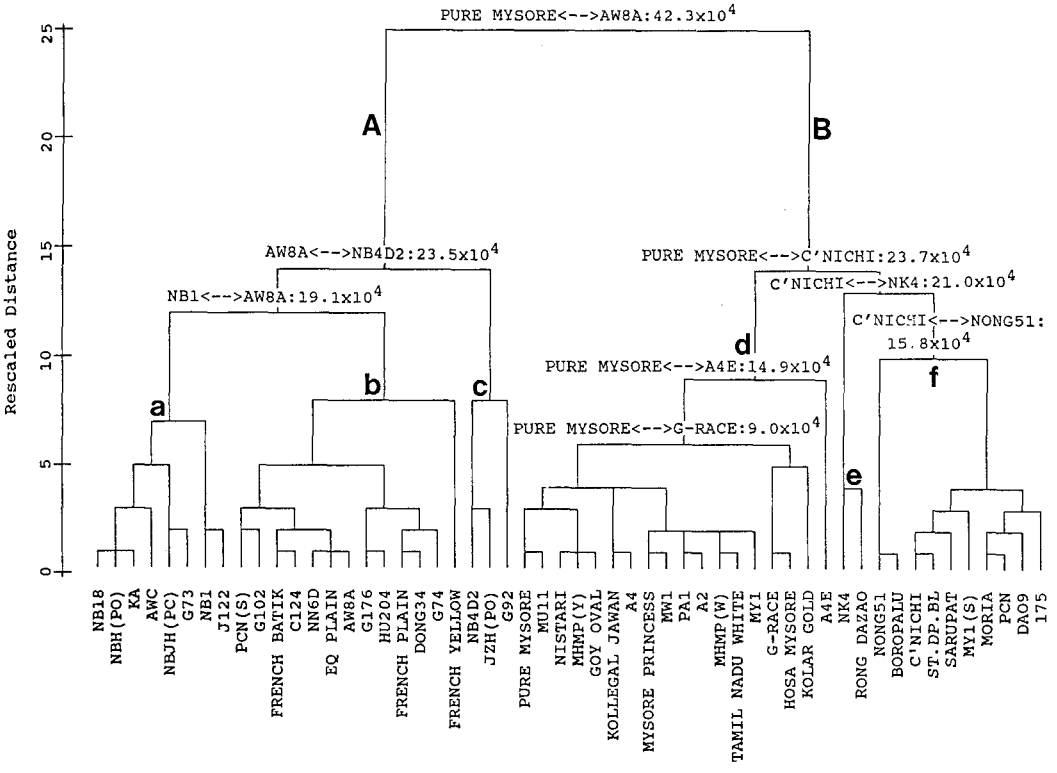

Fig. 2. Dendrogram realised for clustering on the basis of estimates for the four yield attributes and two biochemical parameters (digestive amylase and invertase)

bringing NK4 closer to Rong Dazao. Also of interest is the disassociation of C'nichi, Staple deep black (ST. DP. BL.), Sarupat, 175, Rong Dazao and A4E (Fig. 2) from the earlier cluster 4 (Fig. 1) containing 12 members. The addition of amylase and invertase variables reduced the co-phenetic correlation (Sneath and Sokal 1973) from 0.904 to 0.851.

Clustering on the basis of ten variables

The result of combining six biochemical parameters and four yield attributes revealed two major groups comprising five subgroups as realised earlier, and very little change occurred in the co-phenetic correlation index ($r_{cs} = 0.862$).

Examination of Fig. 3 reveals that even with the addition of the other variables, the relationship has not changed for certain pairs, though the position of others was changed. It is evident from Figs. 2 and 3 that NB4D2, JZH(po) and G92 have remained in the same relative position within the main group. Similarly, the paired association between C'nichi and Staple deep black (ST. DP. BL.), Nong 51 and Boropalu, NK4 and Rong Dazao, PA1, A2, MHMP(w), Tamilnadu White and MY1, or that between Pure Mysore and Mull, has remained unaltered. On the other hand, the relative position of French Yellow (French YL.) has changed and it now appears as a single entity. Likewise, the spatial relationships between NB18 and AWC, or that between AW8a, Eq plain and NN6D, also reflect changes. However, no change was found in the relationship between Pure Mysore↔C'nichi, Boropalu↔Nong 51 or Pure Mysore↔AW8a.

Application of quick clustering

The perusal of earlier dendrograms (Figs. 2 and 3) indicated the presence of six and five subgroups forming the two major groups A and B. Consequently the 'quick cluster' provision of SPSS/PC+ was used with a 'five cluster' command. The analysis of variance for the ten variables and the final cluster centres for these variables in the five clusters are presented in Tables 4 and 5, respectively. It is evident from Table 4 that except for 'trehalose content' all other cluster variances are substantially greater than the error variance. The results presented in Table 5 show, additionally, that cluster 4 has the highest indices for larval weight (LWT), cocoon weight (CWT) and alkaline phosphatase (ALK) activity. On the other hand, cluster 5 has the highest indices for effective rate of rearing (ERR) and amylase activity (AML). The perusal of cluster membership (Table 6) further reveals that cluster 5 includes all low-yielding breeds/races of India, China, Vietnam

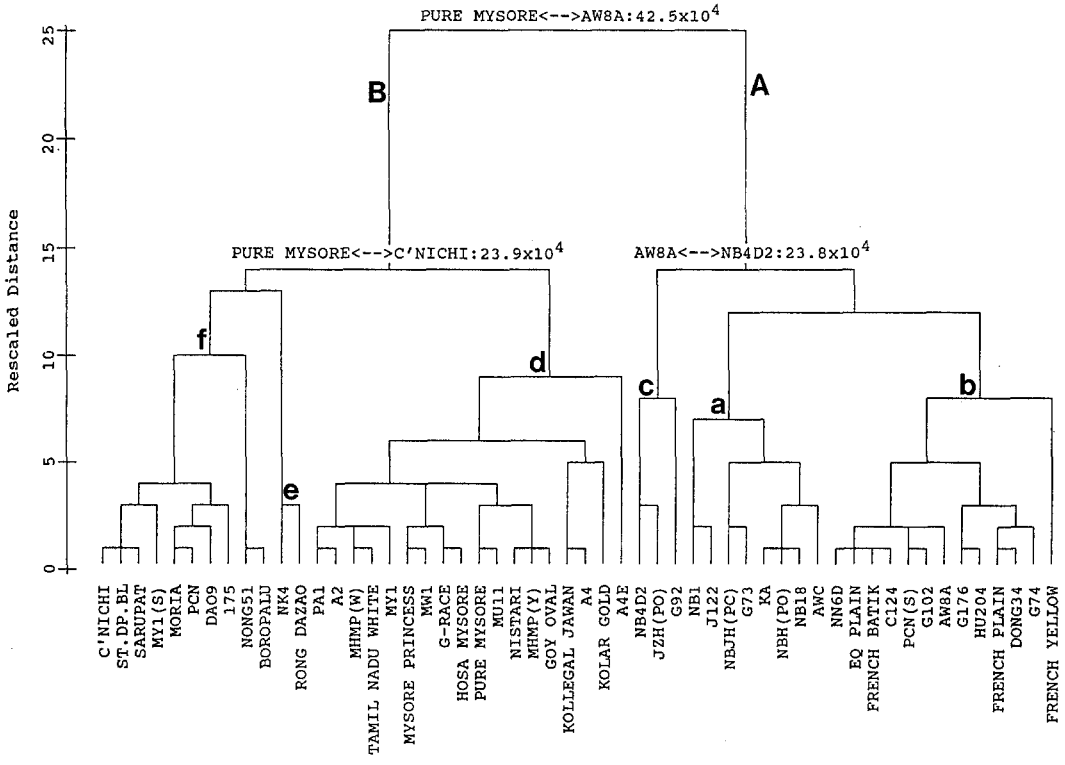

Fig. 3. Dendrogram realised for clustering on the basis of estimates of all the ten variables (see Materials and methods). Other explanations as in Fig. 1. The numbering of clusters is as in Fig. 2

Table 4. Analysis of variance for ten variables as realised from the operation of the SPSS/PC+ 'quick-cluster' programme. The programme was executed after Z-transformation of the estimates of variables

Variable	Cluster MS (df = 4)	Error MS (df = 49)	F
ZLWT	8.3998	0.3959	21.2148
ZCWT	7.6882	0.4540	16.9334
ZSWT	8.9481	0.3512	25.4804
ZERR	9.2579	0.3259	28.4089
ZAML	6.0667	0.5864	10.3457
ZTRE	1.8763	0.9285	2.0208
ZPRO7	2.7447	0.8576	3.2005
ZPRO10	10.5811	0.2179	48.5659
ZINV	3.0827	0.8300	3.7141
ZALK	4.0042	0.7548	5.3053

Explanation of abbreviations given in Table 2

Table 7. Matrix of cluster distances as realised from 'quick clustering with a five cluster command'

Cluster	1	2	3	4
2	4.0154			
3	5.7868	6.5022		
4	3.0204	4.4616	7.4690	
5	4.3522	4.1545	5.5013	4.4003

and Europe. Clusters 1 and 4 include high-yielding breeds of India and other countries. Cluster 3 has only one member (Boropalu) while four members (G102, JZHpo, G74 and French yellow) together constitute cluster 2.

The distance matrix presented in Table 7 shows that the distance between clusters 3 and 4 is maximum

Table 5. Result of 'quick cluster' with five-cluster command showing 'Final Cluster Centers' for 10 variables (with Z-transformation) as realised for five different clusters

Cluster	ZLWT	ZCWT	ZSWT	ZERR	ZAML	ZTRE	ZPRO7	ZPRO10	ZINV	ZALK
1	1.1276	1.0468	1.5455	−1.5105	−1.0883	0.8957	−0.5205	−0.4489	1.4199	0.0614
2	0.7089	0.6250	0.7771	−0.7163	−0.5929	0.3064	−0.1279	3.0456	0.2183	0.4446
3	−1.3368	−1.8970	−1.5193	−4.2014	−1.1912	1.0408	0.0741	−0.3362	0.5413	−0.9101
4	1.6289	1.5357	1.3596	−0.0587	−1.1984	−0.5328	−1.0906	−0.7556	−0.0629	1.4104
5	−0.4450	−0.3961	−0.4598	0.3940	0.4262	−0.0934	0.2522	−0.1334	−0.2141	−0.2536

Table 6. Listing of entries under different clusters as realised from the operation of "quick cluster with five cluster command

Race	Origin	Cluster	Distance	Race	Origin	Cluster	Distance
G92	Japan	1	1.205	Sarupat	India	5	3.457
NBJH(pc)	Korea	1	4.346	MY1(S)	India	5	2.69
NB4D2	India	1	3.136	G-Race	India	5	2.802
G73	Japan	1	3.291	A4E	India	5	2.558
NB18	India	1	2.37	Kollegal Jawan	India	5	2.974
JZH(po)	Brazil	2	3.254	A2	India	5	3.399
G74	Japan	2	0.754	MW1	India	5	3.103
French Yellow	France	2	4.369	Staple Deep Black	Japan	5	3.497
G102	Japan	2	3.183	MU11	India	5	2.736
Boropalu	India	3	0.0	NK4	Thailand	5	3.752
NB1	Japan	4	2.569	A4	India	5	1.801
J122	Japan	4	3.097	Tamilnadu White	India	5	3.841
C124	China	4	0.738	MHMP(w)	India	5	2.727
NBH(po)	Korea	4	4.166	AW8A	India	5	4.934
KA	India	4	3.302	G176	Vietnam	5	2.885
AWC	Europe	4	2.955	HU204	China	5	4.229
Nong 51	China	5	3.913	Dong 34	China	5	4.088
Pure Mysore	India	5	3.11	MY1	India	5	2.916
DA09	China	5	3.098	NN6D	India	5	3.686
C'nichi	India	5	2.718	Goy Oval	India	5	3.204
175	India	5	3.383	Eq Plain	Japan	5	3.633
Moria	India	5	2.732	Rong Dazao	China	5	3.58
PCN(S)	India	5	2.878	French Plain	France	5	3.806
PCN	India	5	3.772	MHMP(y)	India	5	4.477
Mysore Princess	India	5	2.699	Hosa Mysore	India	5	3.269
Nistari	India	5	3.6	Kolar Gold	India	5	2.897
PA1	India	5	2.238	French Batik	France	5	3.568

(7.47) followed by that (6.50) between clusters 3 and 2. On the other hand, the distance between clusters 1 and 4 was found to be minimum. Also, the K-means method included Pure Mysore and C'nichi in the same group (cluster 5 in Table 6).

Discussion

The present work is the first attempt at using the agglomerative hierarchical clustering method for the classification of genetic stocks of the mulberry silkworm, *B. mori* L. As indicated earlier, the present analysis has been done on the basis of the 'average linkage between groups' or UPGMA. As others have shown (Sokal 1986; Rohlf and Wooten 1988; Peters and Martinelli 1989) UPGMA yields more accurate results for classification purposes than other hierarchical methods. The present paper also presents the result of non-hierarchical clustering (K-means method).

Analysis on the basis of Euclidean agglomerative clustering with four yield attributes resulted in the classification of the 54 entries in to seven clusters, which is meaningful insosfar as yield potential is concerned. This interpretation is evident from the clustering of all high-yielding strains (NB4D2, NBJHpc, NB1, G102, AWC, J122, NB18, JZHpo, KA, NBHpo) in one group with two clusters (nos. 6 and 7 in Fig. 1) with low- (clusters 4 and 5) or medium-yielding (clusters 1, 2 and 3) strains in other groups. The separation of NK4, Nong 51 and Boropalu (cluster 5) from other low-yielding strains (e.g., Pure Mysore, Nistari or C'nichi) is particularly interesting. This was due to the very low cocoon weight (66.6–77.5 cg*) and cocoon shell weight (9.1–11.1cg*) in comparison to that realised for Pure Mysore, Nistari and C'nichi (cocoon weight: 83.7–102.2 cg; shell weight: 8.7–13.7 cg*). However, clustering only on the basis of 'yield parameters' failed to project the background differences relating to the origin of the strains and other aspects of their genetical divergence. In this regard, the inclusion of biochemical parameters was of greater help.

Earlier studies on genetic divergence using Mahalanobis D^2 analysis of 49 bivoltine breeds resulted in only three (Jolly et al. 1989) clusters. As is evident from the canonical variance indicated therein, the small number of clusters resulted from the low range of variability for the phenotypic variables utilised for the analysis. On the other hand, a similar study by Rao et al. (1989) resulted in seven clusters. However, both studies used only bivoltine and univoltine high-yielding breeds, whereas the present study was based on multivoltines, univoltines and bivoltines, representing the total range of yield potential as realised in tropical regions having a high to medium temperature range and greater humidity variation. Thus, direct comparison of the present results with those of Jolly et al. (1989) and Rao et al. (1989) is not possible.

The use of biochemical parameters as variables for classification (clustering) of genetic stocks of mulberry silkworm has not previously been attempted. The results presented here clearly establish its usefulness in realising a better projection of the genetical difference between silkworm strains of different yield potentials. The inclusion of estimates of amylase and invertase activity has resulted in two groups, one having all the bivoltine stocks with a high potential for silk yield and the other including all the multivoltines and a few hibernating stocks having poor cocoon characters and high survival comparable to that of the multivoltine stocks (Fig. 2). This improvement over the clustering realised on the basis of yield parameters alone can be explained by several factors. For example, medium-yielding bivoltine strains, like G92 and G74 of Japanese origin, though genetically distant from the multivoltines evolved in India (e.g., Kolar Gold, Hosa Mysore or My1), were found clustered together in the first dendrogram (Fig. 1). However, in the dendrogram realised on the basis of yield and the two biochemical parameters (Fig. 2 and Table 3), G92 was found to be grouped together (subgroup e) with two other high-yielding bivoltine strains, NB4D2 and JZHpo, while G74 was clustered together with two other bivoltines, Dong 34 and French plain (cluster 2 of subgroup b of group A). Likewise, French plain, French yellow and French batik, although closely related, were placed in three different clusters when classified only on the basis of yield trait, but were placed in the same subgroup-e (Fig. 3) when yield parameters were considered along with biochemical parameters.

Even for the relationship between traditional Indian (multivoltine) races, such as Pure Mysore and/or C'nichi, biochemical parameters considered along with yield attributes helped to project the genetic distance between them (Figs. 2 and 3) which endorses the preference which Indian Farmers have shown for more than 200 years for hybrids between the two (Datta 1984). However, when clustering was done only with yield parameters, these two traditional Indian races were clustered together with Nistari and others (cluster 4 in Fig. 1). The present day C'nichi originated from bivoltine C'nichi brought from Japan (as indeed the word 'nichi' suggests), while Pure Mysore is considered to be an original Chinese race (Datta 1984). However, the long larval life of Pure Mysore indicates the possibility of it being a stabilised segregrant from a multi-bivoltine hybrid of Chinese and Japanese breeds.

Further genetical relationships between yield attributes and other genetical markers were shown by

* From raw data, but not projected in the results

Yoshitake (1968), Hirata (1974) and Gamo and Ohtsuka (1980). The genetic markers included both biochemical (blood albumen, acid phosphatase, lipoprotein complex, esterase, amylase etc.) and physiological attributes. Moreover, based on such observations, attempts were made (Chiang 1980; Gamo 1983) to assess genetical distances between different groups of silkworm races in tropical and temperate regions. The tropical races of southeast Asia were shown to have a higher number of gene substitutions than the Chinese, European and Japanese races (Hirobe 1968; Gamo 1983).

Recent studies in this laboratory, using multiple regression for these biochemical estimates, indicate that for the manifestation of the four yield attributes considered here, amylase, invertase and alkaline phosphatase are the most important factors (Chatterjee, unpublished). The present analysis also shows a higher contribution of these three biochemical estimates in the clustering of silkworm breeds and strains.

It is worth noting that sometimes a cluster includes members having different countries of origin; as, for example, NK4 of Japan clustered with Rong Dazao and Nong 51 of China, clustered with Boropalu of India. However, NK4, although collected from Japan, originated in Thailand (Kobayashi 1990). Thus its clustering with Chinese stocks is understandable. Boropalu is an Indian breed, known for more than 100 years (Mukherjee 1912), which lays hibernating eggs. Similarly, Moria, a white multivoltine also known for more than 100 years (Datta 1984), has clustered with PCN, the low-yielding hibernating breed developed during 1984–85 from some old Chinese stock maintained in Japan. All these cases indicate that clustering on the basis of estimates of phenotype does not always reflect geographical distance, as has also been pointed out by researchers working on the clustering of plant materials (Zohary 1970; Harlan 1971). Likewise Spagnoletti et al. (1987) pointed out that geographical position does not correspond with the phenotypic grouping for the origin of spike characteristics in Duram wheat. For the silkworm, domestication has played a major role in genetic diversification (Chang 1977; Gamo 1983). As sericultural regions of the world have different climatic conditions, physiological diversification has also been influenced by agro-climatic factors. Thus, given geographic isolation and limited cultural exchange, Nong 51 of China and Boropalu of India may have acquired similar genotypes due to similar selection pressures. On the other hand, it remains possible that there were exchanges of genetic material between the two countries at a still earlier time. Comparable examples are also available in the present analysis.

The non-hierarchical clustering on the basis of the K-mean method of MacQueen (1967) showed one positive aspect in providing a meaningful differentiation among high-yielding bivoltine/univoltine breeds, highlighting the differential association of alkaline phosphatase activity with high larval weight and high cocoon weight. On the other hand, high invertase activity showed an association with highest shell weight but comparatively lower cocoon weight, suggesting a higher shell ratio (%). This aspect is of particular importance for silkworm breeding because improvement of shell ratio is a priority area for the industry. Also of interest is the identification of Boropalu as a distinct race. Even so, the 'quick clustering' method failed to discriminate different low-yielding breeds and all 38 breeds/races fall within a single cluster.

The cluster analysis provides scope for adopting a recombinational breeding programme using distant cluster members. Thus, the subgrouping of high-yielding bivoltine strains offers an opportunity to exploit the genetical differences between high-yielding strains, as exemplified by clusters a and c of Fig. 2. The clustering also indicates the possibility for recombining low- and high-yielding members from genetically distant clusters. As an example, mention may be made of the hybrid vigour of multi-bivoltine (Pure Mysore, NB4D2 etc.) hybrids. Likewise, studies in this laboratory, and elsewhere, have shown a very high heterotic effect, especially in terms of cocoon yield and pupation rate, for the hybrid of Hosa Mysore and NB4D2 (Tayade 1987; Datta and Pershad 1988). In the present analysis these two breeds fall into two different distant clusters, under subgroups A and B. Thus, evaluation of combining-ability could be of help in testing the correlation between Euclidean distance coefficients and the degree of heterosis in total silk yield. The low-yielding breeds manifest a higher adaptability to tropical conditions than that realised for bivoltine breeds having better cocoon characters. The need for recombining both is a priority task for India (Chatterjee and Datta 1989).

The present report represents a novel approach for analysing the relationships of different silkworm races/breeds on the basis of Euclidean distance, suggests the prospect of easy handling of a larger number of silkworm races and strains, and supports the observations on the 'Euclidean system' proposed by others (Peters and Martinelli 1989). The results also substantiate the usefulness of biochemical parameters in obtaining more discrete information on genetic relationships. Hybrid performance between the near and distant relatives, and assessment of any positive correlation with the degree of heterosis realised, will be useful in testing the efficacy of this hierarchical clustering.

Acknowledgements. The authors express their appreciation to Shri. K. Sridharan for running the SPSS programme. Help from Dr. A. K. Patnaik and Dr. S. K. Ashwath in the preparation of the manuscript is also acknowledged.

References

Bernfeld P (1955) Amylases, α and β. Methods Enzymol I:149–158

Chang Y (1977) Variation in *Bombyx mori* L. under domestication of thousand years. Acta Entomol Sinica 20:345–351

Chiang Y (1980) Origin and differentiation of the silkworm, *Bombyx mori* L. Monogram Publication by The Sericultural Society of Zhejiang

Chatterjee SN, Datta RK (1989) The growth and importance of biotechnology: an overview. Indian Silk 28:60–67

Chatterjee SN, Chatterjee GK, Naidu WD (1988) Genetic variation of digestive juice amylase in bivoltine and multivoltine races. CSRTI Newslett 3:5–6

Chatterjee SN, Chatterjee GK, Rao GP, Kumar SN, Sundar B (1989) Variability of digestive amylase in mulberry silkworm. CSRTI Newslett 3:11–12

Datta RK (1984) Improvement of silkworm races, *Bombyx mori* L., in India. Sericologia 24:393–415

Datta RK, Pershad GD (1988) Combining ability among multivoltine × bivoltine silkworm (*Bombyx mori* L.) hybrids. Sericologia 28:21–29

Eguchi M, Furukawa S, Iwamoto A (1972) Proteolytic enzymes in the midgut of the pharate adult of the silkworm *Bombyx mori*. J Insect Physiol 18:2457–2467

Gamo T (1983) Biochemical genetics and its application to the breeding of silkworm. JARO 16:264–273

Gamo T, Ohtsuka Y (1980) Phylogenetic studies on the racial differentiation of the silkworm, *Bombyx mori*, on the basis of polymorphic genes in haemolymph proteins. Bull Seri Expt Sta Jpn 28:15–50

Harlan JR (1971) Agricultural origins: centres and non-centers. Science 174:468–474

Hirata Y (1974) Relations between the amylase activity of the larval digestive juice and several quantitative characters in *ae* strains of the silkworm, *Bombyx mori*. J Seri Sci Jpn 43:384–390

Hirobe T (1968) Evolution, differentiation and breeding of the silkworm – The silk road, past and present. In Proc Int Congr Genet, Tokyo, pp 25–36

Horie Y (1961) Physiological studies on the alimentary canal of the silkworm *Bombyx mori*. III. Absorption and utilisation of carbohydrates. Bull Seri Exp stn 16:289–309

Ishaaya I, Swirski E (1970) Invertase and amylase activity in the armoured scales *Chrysomphalus aonidum and Aonidiella aurantii*. J Insect Physiol 16:1599–1606

Jolly MS, Datta RK, Noamani MKR, Iyengar MNS, Nagaraj CS, Basavaraj HK, Kshama G, Rao PRM (1989) Studies on the genetic divergence in mulberry silkworm, *Bombyx mori* L. Sericologia 29:545–559

Kobayashi J (1990) Effect of photoperiod on the induction of egg diapause of tropical races of the domestic silkworm, *Bombyx mori*, and the wild silkworm, *B. mandarina*. JARQ 23:202–205

Krishnaswami S (1978) New technology of silkworm rearing. CSRTI Bulletin (2), Central Silk Board, India

Lebeda A, Jendrulek T (1978) Cluster analysis as a method for evaluation of genetic similarity in specific host-parasite interaction (*Lactuca sativa-Bremia lactucae*) Theor Appl Genet 75:194–199

MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. Proc Vth Berkley Symp Mathematical Statistics and Probability, vol I. University California Press, Berkley

Mihara Y, Saito A, Koga K, Sakaguchi B (1988) Changes in the alkaline phosphatase activity during embryogenesis of *Bombyx mori*. J Seric Sci Jpn 57:62–67

Moon TY, Seol GY (1983) Varietal difference in amylase activity of larval digestive fluid in the silkworm *Bombyx mori*. Seric J Korea 24:73–80

Mukherjee NG (1912) Handbook of sericulture. Bengal Secretariat Book Depot, Calcutta, India

Murphy JP, Cox TS, Rodgers DM (1986) Cluster analysis of red winter wheat cultivars based upon coefficients of parentage. Crop Sci 26:672–676

Narasimhanna MN (1986) Guidelines – Chawki rearing centres. Bulletin, Central Silk Board, India

Peters JP, Martinelli JA (1989) Hierarchical cluster analysis as a tool to manage variation in germplasm collection. Theor Appl Genet 78:42–48

Rao GS, Das SK, Das MK, Nandi S (1989) Genetic divergence among bivoltine races of silkworm (*Bombyx mori*). Ind J Agri Sci 59:767–765

Roe JH (1955) Determination of sugar in blood and spinal fluid with anthrone reagent. J Biol Chem 212:335–343

Rohlf FJ, Wooten MC (1988) Evaluation of the restricted maximum likelihood method for estimating phylogenetic trees using simulated allele frequency data. Evolution 2:581–595

Romesburg HC (1984) Cluster analysis for researchers. Lifetime Learning Publications, Belmont, California

Sneath PHA, Sokal RR (1973) Numerical taxonomy. WH Freeman and Co., San Francisco

Sokal RR (1986) Phenetic taxonomy: theory and methods. Annu Rev Ecol Sys 17:423–442

Spagnoletti Zeuli PL, Qualset CO (1987) Geographic diversity for quantitative spike characters in a world collection of durum wheat. Crop Sci 27:235–241

Tayade DS (1987) Heterosis effect on economic traits of new hybrids of silkworm *Bombyx mori* L. under Marathawada conditions. Sericologia 27:301–307

Yoshitake N (1968) Phylogenetic aspects on the origin of Japanese race of the silkworm, *Bombyx mori* L. J Seri Sci Jpn 37:83–87

Zohary (1970) Centres of diversity and centres of origin. In: Frankel OH, Bennett E (eds) Genetic resources in plants: Their exploration and conservation. Blackwell, Oxford, England, pp 33–42