



Numerical taxonomy of phytoseiidae mite family : a multivariate approach

Pavan Kumar S.T. , ¹Debasis Mazumdar and ¹Krishna Karmakar

College of Community Science, Central Agricultural University, West Garo Hills, Meghalaya-794005

¹Bidhan Chandra Krishi Viswavidyalaya, Mohanpur -741252, Nadia, West Bengal.

Received : 19-03-2019 ; Revised : 08-06-2019 ; Accepted : 10-06-2019

ABSTRACT

Available data of 171 species of Phytoseiidae family were classified into three distinct subfamily clusters i.e., Amblyseiinae (Cluster-I), Phytoseiinae (Cluster-II) and Typhlodrominae (Cluster-III) based on key setae characters. PCA was deployed to know the strength of characters in enhancing the separations of the clusters and Discriminant analysis was applied to validate the cluster analysis results, and to find out the discriminating characters in the classification of three important subfamily of Phytoseiidae. Validation results confirmed that 171 species were classified with 99.4% accuracy in the 3 clusters.

Keywords : Cluster analysis, Wilk's Lamda, discriminant analysis, Phytoseiidae, subfamily.

1. INTRODUCTION

Taxonomy is the science of the description and classification of organisms, essential in theoretical and applied biology. The development of statistics in phylogeny and the new taxonomic funding initiatives and global projects are giving some light. It is very important to know the living organisms around us, careful and accurate identification and classification are of vital importance (Kapoor, 1998). Without taxonomy, nobody would be sure of the identity of organisms they were interested in or whether they belonged to the same or different species as the organisms studied by others and without taxonomy, we could not begin to understand biodiversity and the related issue of conservation. As Kapoor (1998) pointed out, taxonomy is essential in theoretical and applied biology (agriculture and forestry, biological control, public health, wild life management, mineral prospecting through the dating of rocks by their enclosed fauna and flora, national defense, environmental problems, soil fertility, commerce, etc).

In the early 1960s, a group of statisticians and biologists introduced a new approach, known as numerical taxonomy or phenetics, the phonetic method was all the rage in the early and mid-sixties, popularized primarily by Robert R. Sokal and Peter H. A. Sneath (1963). It considers that organisms should be grouped on the basis of overall similarity. Many characters are analyzed and taxa are arranged using clustering methods based on overall similarity. The grouping by numerical methods of taxonomic units in to taxa on the basis of their character states.

Among the family of mites, the present study mainly focused on the predatory mite group i.e., Phytoseiidae family. The family Phytoseiidae has received worldwide attention because of their importance in biological control of parasite mites and some of the soft-bodied insect pests of various field and plantation crops. This is probably the most explored and exploited among all the predatory mites. In view of their importance, these mites have also been explored in India, from where many new species have been described since 1960. Apart from these, some work has also been carried out in India on their bio-ecology, predator-prey interactions, and the effects of pesticides on these mites, which have been reviewed in Gupta (2003a). From India, 207 species under 21 genera of eight tribes and three subfamilies arranged as per the classificatory scheme of Chant and McMurtry (2007).

In addition, another 4 species two under *Euseius* and one each under *Neoseiulus* and *Amblyseius* are being described separately by Karmakar and Gupta, (2014 in press) and hence the total now comes to 211 species described. In the present study 171 phytoseiidae mites discovered in India with their important morphological characters were considered. Keeping importance of Phytoseiidae mite family, classification based on morphological characters using statistical methodology was carried out and tried to include important discriminating characters in the model to predict and classify the unknown species of Phytoseiidae in the respective group in the future.

2. MATERIAL AND METHODS

Data for taxonomy study was collected from literatures and books (S.K. Gupta, 2003a, Chant and McMurtry, 2007). The database was treated with multivariate technique viz., Gower's similarity index- the distance which accepts all scales of measurements since morphological characters of different scales were used in the analysis, Cluster

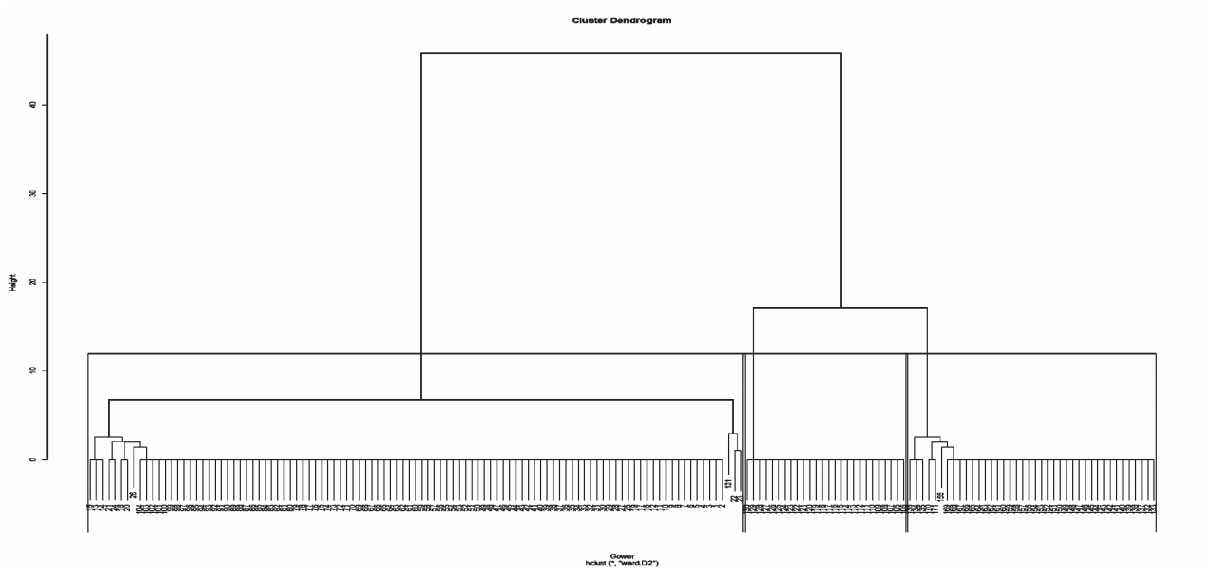


Fig. 1: Dendrogram for subfamily classification of phytoseiidae family using Ward's Method (Gower Distance)

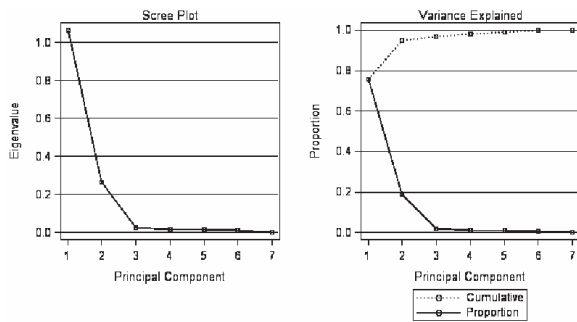


Fig. 2: Scree plot showing total and proportional variance explained by the each extracted principal components.

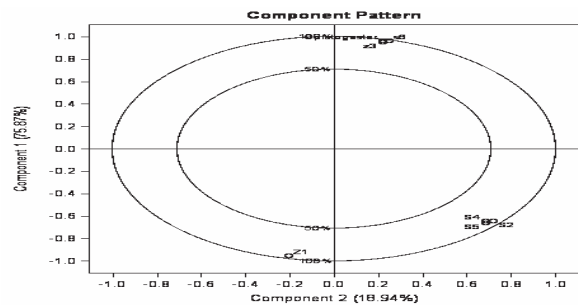


Fig. 4: Component pattern for the morphological characters plotted against the component 1 by the component 2.

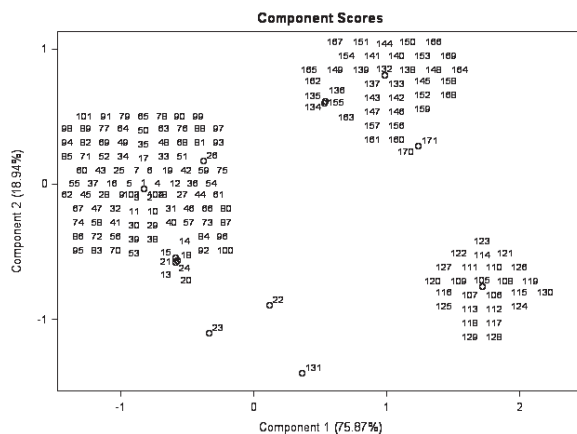


Fig. 3: Scatter plot of 171 studied taxa plotted against the first component by the second component.

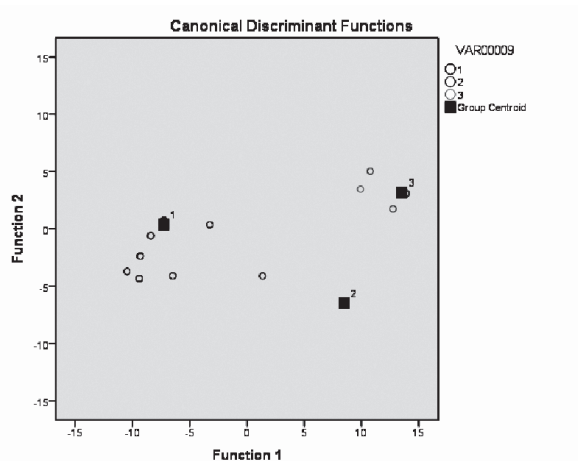


Fig. 5: Group centroid plot for Canonical Discriminant function 1 by function.

analysis for grouping of species in to subfamily, Principal Component analysis (PCA) was deployed to know the strength of characters in the grouping of species into subfamily and Discriminant analysis (DA) to validate the cluster analysis results, and to predict the group membership to avoid misclassification due to clustering.

3. RESULTS AND DISCUSSION

For the classification of subfamily level of Phytoseiidae mite family morphological characters (Table.1) were considered to cluster individual species in different clusters based on similarity. The binary character states of z3, s6, Z1, S2, S4, S5 and JV3 setae were considered for 171 species of phytoseiidae family to cluster them based on the similarity. Descriptive statistics was studied to understand the preliminary idea regarding the character states (Table 2).

Table 1: Morphological characters and character states used in the analysis of subfamily classification of phytoseiidae mite family

Characters	Characters states	Code
z3	Present	1
	Absent	0
s6	Present	1
	Absent	0
Z1	Present	1
	Absent	0
S2	Present	1
	Absent	0
S4	Present	1
	Absent	0
S5	Present	1
	Absent	0
Opistogaster Setae (JV3)	Present	1
	Absent	0

Table 2 : Descriptive statistics for the characters of subfamily classification of phytoseiidae mite family

Characters	Mean	Median	SD	SE	Range	Min	Max	Variance	Skewness	Kurtosis	n
z3	0.39	0	0.49	0.04	1	0	1	0.24	0.45	-1.82	171
s6	0.39	0	0.49	0.04	1	0	1	0.24	0.47	-1.8	171
Z1	0.63	1	0.48	0.04	1	0	1	0.23	-0.55	-1.72	171
S2	0.82	1	0.39	0.03	1	0	1	0.15	-1.67	0.8	171
S4	0.82	1	0.38	0.03	1	0	1	0.15	-1.72	0.98	171
S5	0.81	1	0.4	0.03	1	0	1	0.16	-1.57	0.47	171
JV3	0.39	0	0.49	0.04	1	0	1	0.24	0.45	-1.82	171

Table 3 : Distribution of 171 species into 3 subfamily clusters based on Ward's method

Cluster	Count	Per cent	Species List
I (Amblyseiniinae)	105	61.40	<i>aceriae, baraki, cucumeris, cynodoneae, fallacis, ficusi, imbricatus, indicus, lablabi, longispinosus, neoghaniai, paspalivorus, laeensis, himalayana, sikkimensis, yazuliensis, kumaonensis, heveae, hyauliangensis, mucifera, apocynaeavagrans, arjunae, multidentatus, scleroticus, andamanicus, asiaticus, arecae, assamensis, eucalypticus, hapoli, meghalayensis, officinaria, polyantheae, sapienticola, sorghumae, suknaensis, syzygii, tetranychivorus, adhatodae, aeralis, apocynae, channabasavannai, charui, coffeae, crotolariae, cucurbitae, eujeniae, excelsus, fraterculus, guajavae, hapoliensis, herbicolus, indirae, jarrooa, koothaliensis, kulini, targoensis, manipurensis, mcmurtryi, mohanasundaramai, mulberricus, muraledharani, neocrotalariae, neorykei, orientalis, paraaerialis, pseudorientalis, rangatensis, rubiae, shoreae, sijiensis, arunachalensis, keralaensis, peltatus, synachattiensis, bakeri, salebrosus, fragariae, mumai, eharai, ghaiae, ricini, swaga, ahaioensis, alstoniae, bambusae, chikmagalurensis, chitradargae, coccineae, coccococius, concordis, delhiensis, eucalypti, finlandicus, kalimpongensis, macrospatulatus, mangiferae, neococcineae, ovalis, papayensis, prasadi, rhododendronis, scutalis, vignus, mumai</i>
II (Phytoseiinae)	26	15.20	<i>bandipurensis, brevicrimis, coheni, corniger, crinitus, domesticus, indicus, intermedius, jujuba, kapuri, macropilis, macrosetosus, maldahaensis, meyeriae, minutus, mixtus, namdaphaensis, neocorniger, neoferox, nipponicus, punjabensis, rachelae, roseus, rugosus, swirskii, wainsteini</i>
III (Typhlodrominae)	40	23.39	<i>kuzini, neosoleiger, celtis, nesbitti, transitans, arunachalensis, bambusicolus, channabasavannai, chrysanthemii, communis, dalii, darjeeligenis, denmarki, divergentis, fleschneri, garhowalicus, gopali, hadii, himalayensis, homalii, kodaikanelensis, majumderi, manipurensis, meerutensis, mori, neorhenanus, neotransvaalensis, nilgiriensis, orissaensis, persicus, pruni, rhenanus, rhododendroni, rickeri, sijiensis, sonprayagensis, umbratus, zafari, confusus, roshanlali</i>
Total	171	100.00	

Table 4: Variation explained by the extracted principal component for subfamily classification

Components	Eigen value	Difference	Proportion	Cumulative
1	1.065	0.799	0.759	0.759
2	0.266	0.238	0.189	0.948
3	0.028	0.010	0.020	0.968
4	0.018	0.002	0.013	0.980
5	0.015	0.005	0.011	0.991
6	0.010	0.008	0.007	0.998
7	0.002		0.002	1.000

Table 5: Factor loadings showed the most intrinsic characters enhanced separations of the subfamilies

Characters	Factor loadings						
	Extraction : Principal components						
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
z3	0.4514	0.2073	0.2666	0.4975	0.3852	-0.5159	-0.1423
s6	0.4558	0.2366	-0.0546	0.0017	0.0435	0.2766	0.8093
Z1	-0.4469	-0.1960	0.1338	0.4878	0.5436	0.4374	0.1383
S2	-0.2404	0.5383	-0.1533	-0.5119	0.5938	-0.1172	-0.0231
S4	-0.2383	0.5047	0.7496	0.0082	-0.3244	0.1460	0.0047
S5	-0.2542	0.5252	-0.5623	0.5024	-0.2987	-0.0346	-0.0215
JV3	0.4578	0.2056	-0.0805	-0.0016	0.0847	0.6556	-0.5520
Proportion variance	0.759	0.189	0.020	0.013	0.011	0.007	0.002
Cumulative variance (%)	75.87	94.81	96.78	98.03	99.11	99.84	100

Table 6: Test of equality of cluster means for subfamily classification

Characters	Wilks' Lambda	F	df1	df2	Sig.
z3	0.07	1081.47	2	168	<0.001
Z1	0.07	1120.44	2	168	<0.001
S2	0.19	363.71	2	168	<0.001
S4	0.16	456.05	2	168	<0.001
S5	0.25	251.80	2	168	<0.001
JV3	0.02	3371.79	2	168	<0.001

Table 7: Stepwise discriminant function analysis for the subfamily classification of phytoseiidae family

Step	Entered	Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	JV3	0.024	1	2	168	3371.789	2	168	<0.01
2	S4	0.004	2	2	168	1285.388	4	334	<0.01
3	S2	0.002	3	2	168	1125.394	6	332	<0.01
4	z3	0.002	4	2	168	995.006	8	330	<0.01
5	Z1	0.001	5	2	168	897.772	10	328	<0.01
6	S5	0.001	6	2	168	777.598	12	326	<0.01

Table 8: Canonical correlation and testing significance of discriminant function

Function	Eigen value	% of Variance	Cumulative %	Canonical correlation
1	87.427	90.7	90.7	0.994
2	8.924	9.3	100	0.948
Significance Test				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	0.001	1121.613	12	<0.01
2	0.101	379.812	5	<0.01

Table 9: Unstandardized canonical discriminant functions

Morphological characters	Function	
	1	2
z3	3.984	-0.389
Z1	-3.14	1.955
S2	2.073	3.121
S4	2.169	5.074
S5	1.156	1.338
JV3	13.993	4.679
(Constant)	-9.479	-10.734

Table 10: Cluster centroides for discriminant function

Cluster	Function	
	1	2
1	-7.248	0.385
2	8.498	-6.444
3	13.503	3.178

Table 11: Classification results and predicted group membership of subfamily classification of phytoseiidae family

Classification Results		Cluster	Predicted group membership			Total
			1	2	3	
Original	Count	1	104	1	0	105
		2	0	26	0	26
		3	0	0	40	40
	%	1	99	1	0	100
		2	0	100	0	100
		3	0	0	100	100

Note: 99.4% of original grouped cases correctly classified.

3.1 Cluster analysis

In order to ensure uniformity of programme output, multivariate statistical analysis procedures were run in either two or three independent computer programs (not all the procedures were available in all the three computer programs; Podani 1997; Kovach 1999; Hammer *et al.*, 2001). The input data of 171 species of each seven character states (Qualitative data) was considered. Gower's similarity index for mixed data (Gower 1971) was used to generate a distance matrix for cluster analysis by an agglomerative, hierarchical clustering algorithm (Ward 1963) otherwise known as the "incremental sum of squares method" (Podani 1994). Gower's similarity index was used to generate symmetrical association matrix from the original row-by-column 7 variable descriptors (Setae) for the 171 species for input to a Ward's linkage method for cluster analysis. 171 species of family Phytoseiidae were classified distinctly to 3 major subfamily clusters (Fig. 1), namely Amblyseiinae, Phytoseiinae and Typhlodrominae, out of 171 species 105 were classified as Amblyseiinae (Cluster-I), 26 species as Phytoseiinae (Cluster-II) and 40 species were classified as Typhlodrominae (Cluster-III) subfamily (Table 3). Cluster statistics from table revealed that, for Cluster-I the characters *viz.*, z3, S2, S4, S5 and JV3 were found responsible for the variation in the cluster. In cluster-II none of the characters produced variation while in Cluster-III setae z3, Z1, S5 produced variation within the cluster. Considering all, setae z3, s6, Z1, JV3 and S5 were responsible for the major proportion of variation towards total variation present in the subfamily classification than the S2 and S4 setae.

3.2 Principal Component Analysis (PCA)

To get a first insight into the morphometric variation of 171 species based on 7 morphometric characters (Setae characters) principal component analysis (PCA) was carried out. Three mutually isolated groups of 171 species were separated along the first and second ordination axis of PCA diagram and first two components (PCA1 and PCA2 components) were explained 94.8 % of variation among 171 species. Eigen value for the first component was high *i.e.*, 1.065 which exhibited 75.9 % variation and the second component (Eigen value was 0.266) exhibited 19.9 % proportional variation (Table 4). The factor loadings for characters of the factor1 showed that, the characters z3, s6 and JV3 were found contrasting to Z1, S2, S4 and S5. The factor loading for these characters in the component1 was less than the z3 (0.4514), s6 (0.4558) and JV3 (0.4578) but the setae S2 (0.5383), S4 (0.5047), S5 (0.5252) exhibited better factor loadings than the rest of the setae in the component2 (Table 5). Scatter plot of 171 species using these setae characters produced 3 distinct clusters when plotted against the component1 by the component2 (Fig. 3). The larger member group in the scatter plot was the Amblyseiinae, lower right corner group was Phytoseiinae and the upper right group was Typhlodrominae.

By looking at the scree plot one can also decide the number of clusters (at point 3) (Fig. 2). Contrasting characters which shown in the component pattern plots when component1 plotted against component2, where JV3 (Opistogaster setae), z3 and s6 were in the positive quadrant and Z1, S2, S4 and S5 were found in the negative quadrant (Fig. 4).

3.3 Canonical Discriminant Analysis

Canonical discriminant function is a supervised classification method where the prior cluster information was already known and it was applied here to check how best the clusters are correctly classified. Here stepwise discriminant function analysis was carried out using the setae z3, s6, Z1, S2, S4, S5 and JV3 and the testing of equality of cluster means was tested by using wilk's lamda (Λ). The morphological characters *viz.*, z3, Z1, S2, S4, S5 and JV3 were found highly significant ($p < 0.001$) because wilk's lamda (Λ) was less for these characters but wilk's lamda (Λ) for setae s6 was zero because within cluster sum of square was zero therefore it was excluded from the analysis (Table 6). The stepwise discriminant analysis found that, the setae JV3, S4, S2, z3, Z1 and S5 were considered for the group discrimination of Amblyseiinae, Phytoseiinae and Typhlodrominae. In the stepwise procedure setae JV3 entered first and then S4, S2, z3, Z1 and S5 were entered in later steps this was due to the wilk's lamda (Λ), the characters which are able to minimize the wilk's lamda (Λ) those characters can be able to enter in to the model. From table 7, it was observed that from JV3 the value of wilk's lamda (Λ) decreased gradually till the final character get enter into the model and significance of the entered characters were tested using F test and found that all the characters included in the model were significant at probability level < 0.001 .

The first discriminant function explained 90.7 % variation (Eigen value 87.427) and second discriminant function explained 9.3 % variation (Eigen value 8.924). Canonical correlation from the first discriminant function was 0.994 which specify that, prediction of group membership using first function provides better accuracy and through which an unknown species can be relocate to the predetermined groups without bias. Both function1 and function2 were found significant due to low wilk's lamda (Λ) value and tested using Chi-square test, the functions found highly

significant at 1% level (Table 8.). The unstandardized canonical discriminant functions (Table 9) could be used to predict the group membership of new/ unknown species to relocate them into the desired clusters and predicted score is compare with the respective cluster centroides if predicted score from both functions satisfied (Approximately near) centroides (Table 10).

Original cases (171 species) in the respective clusters were validated using discriminant functions and it was clear that 99.4 % of the original grouped cases (Species) were correctly classified in to three clusters (Table 11). One case from the first cluster was misclassified in the second cluster i.e Species *Phytoseius mumai* was classified in the first cluster from cluster analysis (Amblyseiinae) but this species actually belonged to cluster-II (Phytoseiinae) so that dicriminant function is validated the cluster analysis result and located this species in the second cluster, the predicted score for this species using discriminating character was 9.9 from the DF1 (First discriminant function) and -4.1 from the DF2 (Second discriminant function) which is very close to the centroides of second cluster (DF1- 8.49 and DF2- 6.44). Therefore *Phytoseius mumai* should be belong to Cluster-II (Phytoseiinae). It was confirmed from the plot which was plotted against DF1 by the DF2 (Fig.5), even the group centroides are very distinct therefore it was clear that any new species or unknown species belongs to Phytoseiidae family can be locate or relocate into 3 subfamilies (Amblyseiinae, Phytoseiinae and Typhlodriminae) without bias.

Subfamily classification was in accordance with the classificatory scheme of Chant and McMurtry, (2007).

4. CONCLUSION

In this study key characters considered were measured in different scales (Nominal, Ordinal, Ratio and Interval scale), therefore the Gower's distance similarity index was deployed. The advantage of using Gower's distance was, it accepts all scale of measurement. Total 171 species of Phytoseiidae family were classified into three distinct subfamily clusters i.e., Amblyseiinae (Cluster-I), Phytoseiinae (Cluster-II) and Typhlodrominae (Cluster-III) based on key setae characters like z3, s6, Z1, S2, S4, S5 and JV3. To know the strength of these characters in enhancing the separations of the clusters was treated with principal component analysis and it was found that characters z3, JV3, s6 were contrasting to Z1, S2, S4 and S5 whereas all the characters explained better variation. The morphological characters were treated with the discriminant analysis, setae character s6 was excluded from the model due to wilks lamda. Setae character z3, Z1, S2, S4, S5 and JV3 found discriminating between subfamily groups and hence these were considered to be key characters in separating the subfamily clusters.

REFERENCES

- Chant, D. A. and McMurtry, J. A. 2007. Illustrated Keys and Diagnoses for the Genera and Subgenera of the Phytoseiidae of the World (Acari: Mesostigmata). Indira Publishing House, West Bloomfield, MI. 220 pp.
- Gower, J. H. 1971. A general Coefficient of similarity and some of its properties. *Biometrics*, **27**:857-74
- Gupta, S. K. 2003a. A monograph on plant inhabiting predatory mites in India. Part II: Order: Mesostigmata. Mem. Zool. Surv. India. 20(1): 185pp.
- Hammer, O., Harper, D.A.T. and Ryan, P.D. 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4(1)**: 9.
- Kapoor, V.C. 1998. *Principles and practices of animal taxonomy*. Science Publishers.
- Karmakar, K. and Gupta, S. K. 2014. Description of four new species of phytoseiid mites (Acari: Mesostigmata) from West Bengal. Records zool. Surv. India (in press).
- Kovach, W.L. 1999. MVSP- A Multivariate Statistical Package for Windows, ver.3.1. Kovach Computing services, Pentraeth, Wales, UK
- Podani, J. 1997. SYN-TAX s.10-PC. Supplement to the Users Manual. Scientia Publishing, Budapest, Hungary.
- Podani, J. 1994. Multivariate Data Analysis in Ecology and Systematics- A Methodology Guide to the SYN-TAX 5.0 package. SPB academic Publishing, The Hague, The Netherlands, 316
- Sokal, R.R and Sneath, P. H. A. 1963. Principles of Numerical Taxonomy. The central work on phenetic approaches to classification. *Freeman V.H. and Co. San Francisco*.
- Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. *J. Amers. Stat. Assoc.* **58**: 236-44