Rivista Italiana di Paleontologia e Stratigrafia	volume 102	numero 3	pagine 367-374	Dicembre 1996

COMPARISON BETWEEN CLADISTIC AND PHENETIC METHODS IN LARGER FORAMINIFERA ANALYSIS

FABRIZIO MAIA

Key-words: Miogypsinidae, Phenetic analysis, Cladistic analysis.

Riassunto. Oggetto di questo lavoro è l'analisi numerica dei Miogypsinidi, macroforaminiferi presenti nelle facies terrigene e bioclastiche alternate a facies pelitiche della Collina di Torino e del Monferrato, nell'intervallo Oligocene sup. - Miocene inf. In accordo con la letteratura, lo studio dei Miogypsinidi si basa sulla valutazione dei parametri biometrici relativi all'apparato embrionale e nepionico. I dati ottenuti dalle misurazioni vengono trattati statisticamente e, successivamente, viene effettuato un confronto con i dati relativi alle unità tassonomiche riconosciute in letteratura. Alcune attribuzioni specifiche ottenute in questo modo appaiono poco affidabili: ciò significa che, in diversi casi, l'elaborazione statistica tradizionale risulta inefficace. Al fine di superare tale problema e trovare nuovi metodi non soggettivi per assegnare nuove popolazioni alle rispettive specie, vengono esaminate alcune tecniche fenetiche e cladistiche. Nel presente lavoro vengono quindi confrontati vantaggi e difficoltà nell'applicare l'analisi delle componenti principali, l'analisi dei clusters, l'analisi discriminante e l'analisi filogenetica col sistema della parsimonia. L'analisi discriminante sembra fornire i risultati più interessanti.

Abstract. The analysis is focused on Miogypsinidae, larger Foraminifera characterizing the terrigenous and bioclastic facies that alternate with pelitic facies in the Turin Hill and in the Monferrato area (NW Italy) from Upper Oligocene to Early Miocene. According to the literature, the study of Miogypsinidae is based on biometry of embryonic and nepionic apparatus characters. Measurements are processed statistically and comparison is also made with the values for taxonomic units recognized in literature. Certain specific determinations so obtained result ambiguous: this means that, in many cases, the traditional statistical elaboration appears to be inefficient. To overcome this problem and to find a new method to assign new populations to their species in a non-subjective way, an attempt is made to use either phenetic or cladistic systems. In this work there is a comparison between the advantages and the difficulties in using techniques like principal components analysis, cluster analysis, discriminant analysis and phylogenetic analysis using parsimony. Discriminant analysis seems to provide the best results.

Introduction.

Miogypsinidae are benthic polythalamic non-sessiles larger Foraminifera, belonging to the super-family of Orbitoidacea. They originate in Late Oligocene and disappear in Burdigalian. Each specimen is sectioned and studied in equatorial plane, then the data of the specimens of a sample are collected to obtain an average of the whole population, according to Drooger (1952) and subsequent papers.

Attribution of a population to a certain species is based on biometry of embryonic apparatus parameters pertinent to the *juvenarium* (protoconch and deuteroconch) and to the nepionic protoconchal spirals (Fig. 1). These parameters are the main spiral lenght (X), the γ angle between the medio-embryonic line and the frontal-apical line, the symmetry of the nepiont (V), the protoconchal diameter (D1), the deuteroconchal diameter (D2), the ratio D2/D1, the distance between the center of the protoconch and the apical margin (ε), the ratio ε/D1.

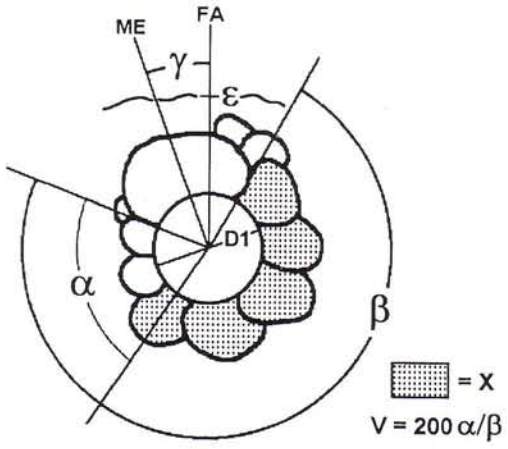

Fig. 1 - Schematic drawing showing the measures of the internal features in embryonic-nepionic stages of *Miogypsina*. Meaning of the symbols: FA = frontal-apical line; ME = medio-embryonic line; α = angle made by the shortest spiral around the protoconch; β = angle made by both spirals around the protoconch; γ = angle between ME and FA lines; D1 = diameter of the protoconch; ε = distance between the center of the protoconch and the apical margin; X = number of chambers of the principal spiral around the protoconch; V = degree of simmetry of the nepiont.

- Dipartimento di Scienze della Terra - Via Accademia delle Scienze, 5 - 10123 Torino.

Species	Samples		X	ϒ	V	D1	ε	ε/D1
M. gunteri	Superga	1	10.44	-101.06	4.69	182	312	1.74
M. gunteri	Civera	7 bis	10.43	-109.00	1.89	151	259	1.75
M. gunteri	Monferrato	MC18	9.40	-49.00	-	154	255	1.67
M. gunteri	Monferrato	MC12	9.90	-78.00	-	155	254	1.70
M. gunteri	Monferrato	MC2	9.23	-43.00	-	160	257	1.60
M. gunteri-tani	Superga	33 bis	9.16	-46.79	4.76	181	295	1.64
M. gunteri-tani	Baldissero	28	9.63	-43.34	7.22	189	303	1.61
M. gunteri-tani	Baldissero	34	9.32	-49.16	5.04	172	308	1.81
M. gunteri-tani	Monferrato	MU1	8.52	-24.00	-	166	241	1.48
M. gunteri-tani	Monferrato	PC2E	9.18	-27.00	6.90	155	241	1.57
M. gunteri-tani	Monferrato	PC2C	8.97	-34.00	3.70	161	251	1.58
M. tani	Monferrato	RO	8.50	-19.00	8.60	177	267	1.50
M. tani	Monferrato	VM	8.69	-18.00	8.40	178	253	1.42
M. tani	Monferrato	SM3	8.07	-13.00	11.00	166	257	1.57
M. tani	Monferrato	SM2	8.50	-20.00	-	163	-	-
M. tani	Monferrato	SM1	7.00	-17.00	12.10	190	269	1.52
M. tani	Monferrato	SB1	7.97	-30.00	6.50	171	268	1.62
M. tani	Monferrato	FB2	7.91	-15.00	-	176	261	1.50
M. tani	Monferrato	FA2	8.30	-44.00	5.60	168	261	1.56
M. tani	Monferrato	FA1	8.40	-38.00	10.60	163	270	1.68
M. globulina	Superga	93	6.95	16.48	35.47	172	351	2.10
M. globulina	Civera	10	6.17	22.00	27.22	137	213	1.58
M. globulina	Civera	13	6.53	27.00	24.51	178	296	1.70
M. globulina	Civera	16	6.42	19.17	23.79	150	232	1.55
M. globulina	Monferrato	CA1b	6.73	15.30	28.20	162	241	1.49
M. globulina-intermedia	Superga	84	6.60	18.45	44.13	182	311	1.71
M. globulina-intermedia	Civera	29	5.88	29.69	48.19	162	279	1.75
M. globulina-intermedia	Civera	36	5.40	34.50	40.28	141	262	1.89
M. globulina-intermedia	Civera	961	5.45	30.50	52.14	164	273	1.68
M. globulina-intermedia	Baldissero	68	5.57	27.14	44.43	145	250	1.70
M. globulina-intermedia	Monferrato	M10	5.83	25.50	42.50	185	292	1.58
M. globulina-intermedia	Monferrato	M8	5.78	27.80	40.60	189	311	1.68
M. socini	Superga	43	8.47	-35.71	19.87	184	312	1.72
M. socini	Bric Palouch	3a	8.30	-10.10	22.74	172	309	1.81
M. socini	Rivodora	3F	8.92	-42.90	14.43	168	312	1.88
M. socini	Rivodora	4	8.63	-17.63	18.23	172	353	2.09
M. socini	Monferrato	VDMU	8.43	-38.00	17.30	141	241	1.71
M. socini-burdigalensis	Superga	52	10.00	-63.00	28.18	162	309	1.91
M. socini-burdigalensis	Baldissero	48	9.21	-52.20	13.64	157	320	2.06
M. burdigalensis	Baldissero	41b	7.97	-	30.99	166	-	-
M. burdigalensis	Vergnana	1	7.45	-4.83	29.83	178	301	1.70
M. burdigalensis	Bric Palouch	3b	7.09	-	38.61	-	-	-
M. negrii	Superga	72	5.33	-	68.10	-	-	-
M. negrii	Monferrato	CA1a	5.16	-	75.90	188	-	-
M. negrii	Monferrato	RS1	5.30	-	55.30	183	-	-

Tab. 1 - Mean values of counts and measurements on internal characteristics of *Miogypsina* populations studied in Piedmont.

During the evolution of this group, parameter changes permit to reconstruct an hypothetical phylogenesis. Up to now, the determination of one population of Miogypsinidae was based on the comparison of its parameters with the values for taxonomic units recognised in literature. Student's "t" test was used to compare the specific determinations thus obtained with all the populations assigned to these forms by other workers and hence confirm our species attribution.

The aim of this work is to find identification methods more reliable to assign a new population to a predetermined taxon, using more objectivity as possible and taking advantage of all the measurement collected. Such methods have been identified in the realm of phenetic analysis (Barrai, 1984; Camussi et al., 1986; Dunn & Everitt, 1982; Elliott, 1977) and cladistic analysis (Forey et al., 1992).

The input data for the analysis correspond to the mean values of the populations studied in Piedmont (Tab. 1) and those of the populations reported in literature (Tab. 2). On the base of available measurements, for each analysis we used the one or the other data set. The species considered are those from *M. gunteri* to *M. Globulina-intermedia* along the main branch of the phyletic tree, and those from *M. socini* to *M. negrii* along the lateral branch. We have chosen to use the mean values rather than individual values, because the objects of our study are the populations rather than the individuals that compose the populations. The mean values have been standardized before exploiting phenetic analysis: standardization consists in subtracting the mean of each variable from each value and dividing the difference by the standard deviation.

A true incentive to undertake this work has come from the development of computer packages suitable for many types of numerical analysis. Among them, we have selected NTSYS-pc (a software for phenetic analysis), STATGRAPHICS (statistics) and PAUP (cladistic analysis).

About the phenetic analysis, in this work we will describe the principal component analysis, the cluster analysis and the discriminant analysis. About cladistic analysis, we will discuss the method of maximum parsimony, used to infer the phylogenetic tree of Miogypsinidae from their characters. First the principle of similarity will be introduced.

Species	References	Samples
M. gunteri	Drooger, 1954(a)	11
M. gunteri	Ferrero, 1965	1
M. gunteri-tani	Drooger, 1952	11a
M. gunteri-tani	Drooger et al., 1955	22a"; 34; Mor222
M. tani	Drooger, 1952	12
M. tani	Drooger et al., 1955	2b; 13-15; 15a; 16-18; M1-6; 26"
M. tani	Delicati & Schiavinotto, 1985	PMT16; PMT7
M. tani	Vilizzi, 1991	OT2; OT4
M. tani-globulina	Drooger, 1952	18
M. tani-globulina	Raju, 1974	G1437
M. tani-globulina	De Mulder, 1975	A179; DM363
M. tani-globulina	Fermont & Troelstra, 1983	80PC02
M. globulina	Drooger, 1952	19-22
M. globulina	Drooger, 1954(a)	5; 6; 10; 18
M. globulina	Drooger, 1954(b)	240b
M. globulina	Drooger et al., 1955	1; 2; 5; 20a; 24-25; 35-36; 43-44; M7
M. globulina	Ujiié & Oshima, 1969	Shuk.A
M. globulina	Matsumaru, 1971	CH; HO
M. globulina	Raju, 1974	KR36-B; G1401-B; G1401AB; G1406-B; G1406AB; G1421-B; Jag.W.B.
M. globulina	De Mulder, 1975	DM114; DM116; DM117
M. globulina	Schüttenhelm, 1976	SM-28/87/81/268/157A/110/131/135/166/251/248/469/140/199/347
M. globulina	Schiavinotto, 1979	TLS76
M. globulina	Delicati & Schiavinotto, 1985	PMT-6/80
M. globulina-intermedia	Drooger, 1952	24; 25
M. globulina-intermedia	Drooger, 1954(a)	7; 8; 9
M. globulina-intermedia	Drooger et al., 1955	22a'; 29; 3
M. globulina-intermedia	De Mulder, 1975	A194; DM106; DM608
M. globulina-intermedia	Schüttenhelm, 1976	SM-22/481/258
M. globulina-intermedia	Wildemborg, 1991	JT5124; JT5063
M. intermedia	Drooger et al., 1955	21
M. intermedia	De Mulder, 1975	DM684; DM107; DM140
M. intermedia	Schüttenhelm, 1976	SM-205/237/440
M. intermedia	Schiavinotto, 1985(a)	AC5
M. intermedia	Schiavinotto, 1985(b)	Ca82
M. intermedia	Wildemborg, 1991	JT-7980/5117/5112/5107/7978/7977/7976/7975/5102/5098/5094/7974/7177/5077
M. socini	Drooger, 1954(a)	12 13
M. socini	Vervloet, 1966	326-3
M. socini	Schüttenhelm, 1976	SM-482/483/234/267/447/266/116/124/287/444/283/188/187/184/226/227/298
M. socini	De Bock, 1977	M13
M. socini	Schiavinotto, 1979	TLS107
M. socini	Delicati & Schiavinotto, 1985	PMT3; PMT2
M. socini-burdigalensis	Schüttenhelm, 1976	SM-233/373
M. burdigalensis	Schüttenhelm, 1976	SM-478/479
M. burdigalensis	Schiavinotto, 1979	TLS-42/39
M. burdigalensis-negrii	Schiavinotto, 1979	TLS-109/33
M. negrii	Schiavinotto, 1979	TLS110

Tab. 2 - Populations of *Miogypsina* reported in literature and considered in this work. See Ferrero et al. (1992, 1994) for references that are not cited at the end of this work.

Similarity.

The principle of similarity is essential to deal with some problems related to the principal component analysis and the cluster analysis.

Similarity is the resemblance or affinity among the taxonomic units, based on their characters; in other words, it's their phenetic relationship. The complement of the similarity of taxonomic units is their dissimilarity or phenetic distance, measured by means of processes that satisfy mathematical properties which make them particularly suitable for phenetic analysis.

Among the available measures of dissimilarity, in this work we used the average taxonomic distance. Its expression is similar to that of euclidean distance, resulting from the Pythagora's theorem.

Principal Component Analysis.

The Principal Component Analysis (PCA) is useful to find new variables that are linear combinations of the original measures and describe the sample without the abundance of information rising from correlation among the original measures. The new variables have to be uncorrelated, so that it is possible to choose those showing the greatest variance. The first few measures that account for most of the variation of the sample correspond to the principal components. The results of this transformation can be better explained by calculating the correlation among the original measures and the new variables, so that a loading matrix is obtained in which the absolute value and the sign of the correlations allow to understand the connections among measures and principal components. A geometrical interpre-

Principal components

Component	Eigenvalue	Percentage of Variance	Cumulative Percent
1	2.71035	67.76	67.76
2	1.08278	27.07	94.83
3	0.15513	3.88	98.71
4	0.05173	1.29	100.00

Loading matrix

Parameter	Comp. 1	Comp. 2	Comp. 3
X	0.974	-0.115	-0.080
ɣ	-0.962	0.110	0.210
V	-0.914	-0.267	-0.303
ε/D1	0.027	-0.993	0.114

Tab. 3 - Results of principal components analysis. The eigenvalues (latent roots) are proportional to the variance accounted for by each of the first four components. The component loadings (latent vectors) for the first three principal components are shown in the loading matrix, in which it appears that the first component loads heavily on X, γ and V, while the second loads heavily on ε/D1.

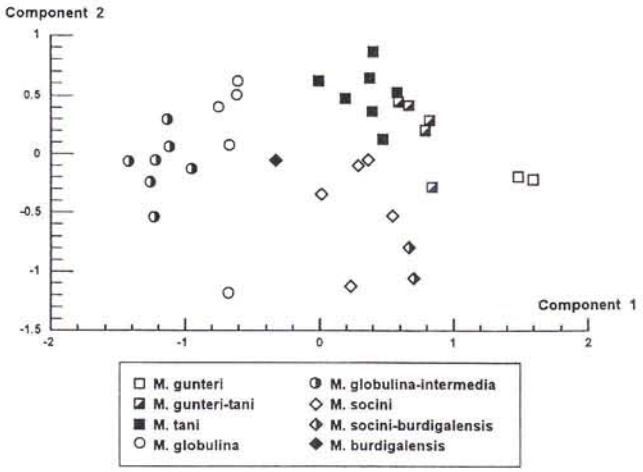

Fig. 2 - Plot of the first two principal component scores for populations studied in Piedmont.

tation of the PCA is possible if we imagine to find the axis (or dimension) that express the variation of the characters, that is the axis which maximizes the variance of the projections of the values onto itself. This axis is given by the line that minimizes the sum of squares of the distances between the values and itself. If there are "p" characters, the first principal component is the best-fitting straight line in the p-dimensional space. A very useful visualization of the results of PCA is given by a scatterplot of the first component score for each taxonomic units against the second. How well this scatterplot describes the configuration in the original p-dimensional space may be measured by the proportion of the variance in the data acccounted for by the first two principal components.

The analysis has been carried out either on the data of the populations reported by literature or on the data of the populations studied in Piedmont. In both cases the first two principal components account for more than 90% (cumulative) of the total variance.

Only the results concerning the populations studied in Piedmont are shown (Tab. 3). The loading matrix shows the respective significance of each of the original measures (X, γ, V and ε/D1) in making the components. In the scatterplot based on the first two principal components (Fig. 2), the groups of the populations belonging to each species appear well distinct.

To verify how well this two-dimensional mapping preserves the original distances among populations (measured with the average taxonomic distance, previously discussed), we can use the Minimum Spanning Tree (MST) of the distance matrix. The spanning tree is a set of straight-line segments joining pairs of points such that no closed loops occur, each point is touched by at least one line and the tree has continuous link between any pair of points. If a weight is assigned to each segment, than the lenght of the tree is defined to be the sum of these weights. The MST is defined as the spanning tree of minimum lenght. This tool helps to detect local distorsions in the diagram resulting from the ordination technique, like the case of pairs of populations which look close together in the two-dimensional representation, but actually are far apart if other dimensions are taken into account. For example, in the scatterplot of the PCA applied to the data of Piedmont (Fig. 2), populations SUP52 and BAL48, belonging to *M. socini-burdigalensis*, appear to be well separated from

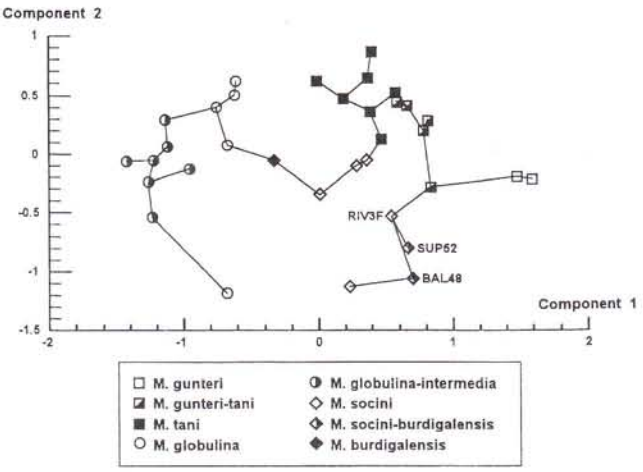

Fig. 3 - Minimum spanning tree superimposed on scatterplot of the first two principal component scores for populations studied in Piedmont.

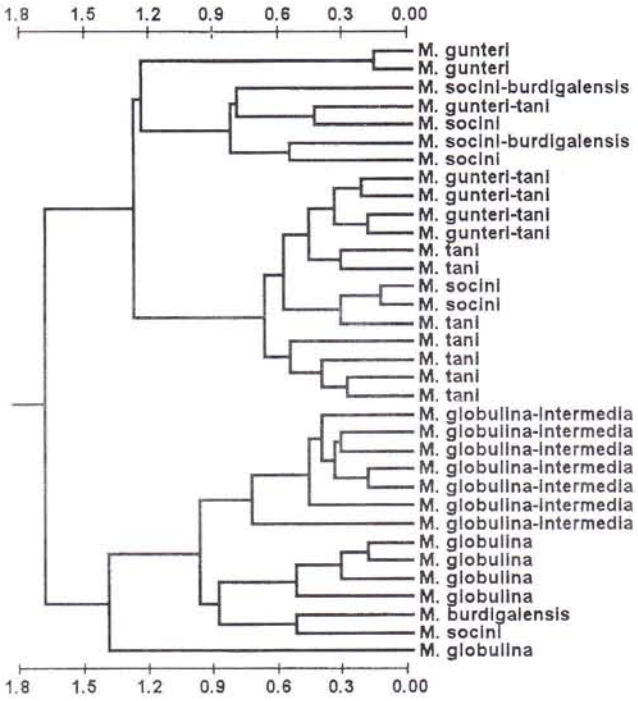

Fig. 4 - Dendrogram showing the results of group-average clustering applied to the populations studied in Piedmont. Cophenetic correlation coefficient = 0.75.

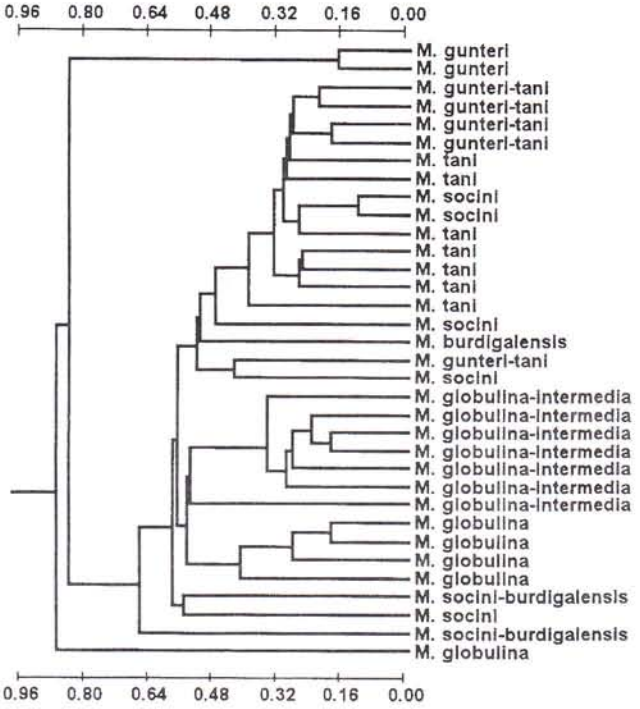

Fig. 5 - Dendrogram showing the results of single-linkage clustering applied to the populations studied in Piedmont. Cophenetic correlation coefficient = 0.64.

the other groups. But if the MST is superimposed on the plot (Fig. 3), we see that the two populations are "closer" to RIV3F (belonging to *M. socini*) than to each other. This means that the projection of the populations onto the first two principal components axes has not preserved, in the case of *M. socini-burdigalensis*, the original structure of the phenetic distances.

Cluster analysis.

A cluster can be defined like a maximally connected set. For the analysis of Miogypsinidae we have selected agglomerative hierarchical clustering techniques, that proceed by a series of successive fusions of the taxonomic units into groups. The two populations showing the smaller distance between them (measured with the average taxonomic distance, discussed previously) are grouped together, then distances between this two-member cluster and each of the remaining populations are calculated. The process continues with the number of groups being reduced by one at each stage, until all the populations are grouped into a single cluster. A useful means to display the results is a diagram called "dendrogram".

The methods available differ in the algorithm they use to calculate the distance between two clusters. In group-average clustering (Fig. 4), the distance between two clusters is the average of the distances between all

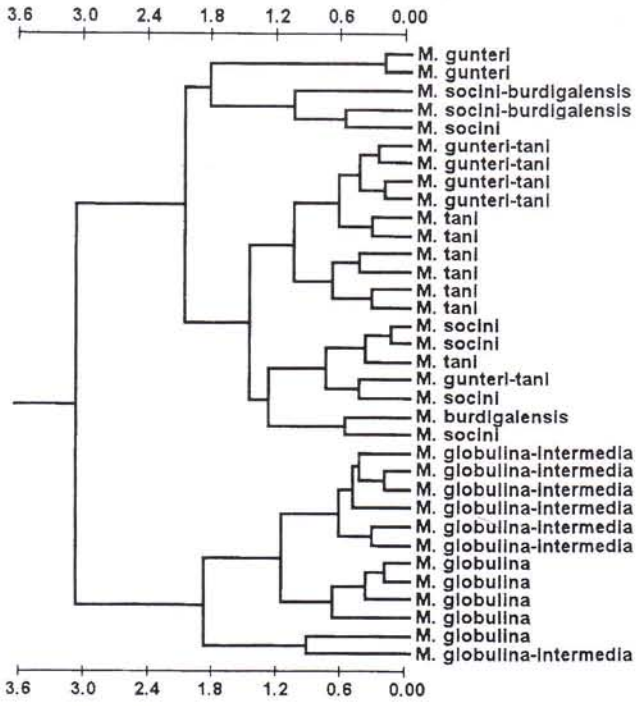

Fig. 6 - Dendrogram showing the results of complete-linkage clustering applied to the populations studied in Piedmont. Cophenetic correlation coefficient = 0.77.

Discrim. Function	Eigenvalue	Percentage of Variance	Cumulative Percent
1	12.812	86.55	86.55
2	1.654	11.17	97.72
3	0.337	2.28	100.00

Functions Derived	Wilks Lambda	Chi-Square	Degree of Freedom	Signific. Level
0	0.020	686.94	33	0
1	0.282	223.53	20	0
2	0.748	51.29	9	0

Tab. 4 - Canonical discriminant functions. The eigenvalues are proportional to the variance accounted for by each function. The significance level of the three functions is tested with the Wilks - Λ statistics and the χ - square statistics, that show a high degree of significance (probability = 0).

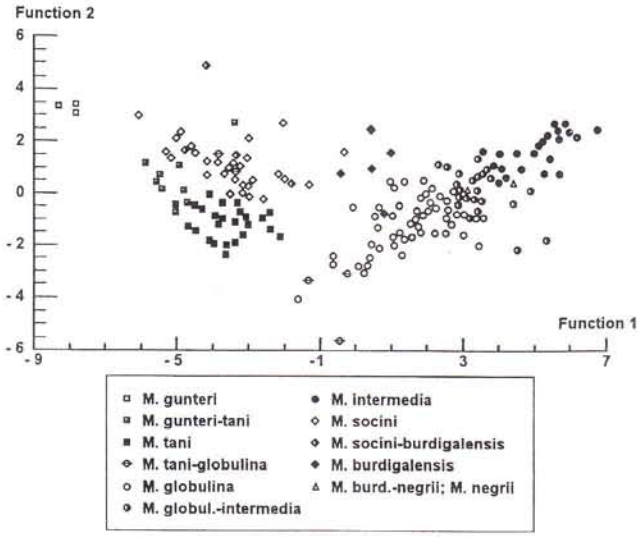

Fig. 7 - Plot of population means relative to the first two canonical discriminant functions. Data from literature and from Piedmont.

pairs of populations that are made up of one population from each group. In single-linkage clustering (or nearest neighbour method) (Fig. 5), the distance between two clusters is that of their most similar pair of populations. In complete-linkage clustering (or furthest neighbour method) (Fig. 6), the distance between two clusters is that of their least similar (or most dissimilar) pair of populations.

To evaluate if the original relationships among the populations (described by their measured distances) fit well with the hierarchical structure imposed on data by clustering, we can use the cophenetic correlation coefficient. It corresponds to the correlation of the original distance matrix with the cophenetic matrix, which consists of the set of similarities produced by clustering. A value above 0.8 is sufficient to give evidence of a good fit between dendrogram and distance matrix.

The cluster analysis of the populations of Miogypsinidae studied in Piedmont, applying the complete linkage method, produced the most meaningful grouping and the highest cophenetic correlation coefficient (Fig. 6). The resulting classification of populations is not the same as that produced applying only taxonomic criteria, but it provides important informations. For example, the dendrogram shows that *M. globulina* and *M. globulina-intermedia* are well distinct as regards to all the other species: this suggests an evolutionary trend that isolates these two taxonomic units from all the others. Populations belonging to the phyletic branch *M. socini* - *M. burdigalensis* show characters so different that it appears very difficult to group them in a single cluster. Finally, *M. gunteri* results well distinct from the other species.

Discriminant analysis.

The discriminant analysis technique assumes that significative differences would exist among the mean vectors of the species to which populations belong. In certain respects, it is similar to PCA, but PCA seeks transformed axes that account for most of the global variation of the data, while in discriminant analysis the transformed axes permit to separate the mean vectors of groups. In the case of more than two groups to be discriminated, the method is called canonical variate analysis and consists in seeking one or more new variables that would be linear functions of the original varia-

	Predicted Group (percentage)											
Actual Group	gunt.	g.-tani	tani	t.-glob.	glob.	gl.-int.	interm.	socini	s.-burd	burd.	b.-neg.	negrii
M. gunteri	100	0	0	0	0	0	0	0	0	0	0	0
M. gunteri-tani	0	78	0	0	0	0	0	11	11	0	0	0
M. tani	0	4	89	7	0	0	0	0	0	0	0	0
M. tani-globulina	0	20	0	60	0	0	0	20	0	0	0	0
M. globulina	0	0	0	9	65	16	0	0	0	4	5	0
M. globulina-intermedia	0	0	0	0	0	74	0	0	0	0	4	22
M. intermedia	0	0	0	0	0	17	70	0	0	0	0	13
M. socini	0	17	3	0	0	0	0	55	17	7	0	0
M. socini-burdigalensis	0	25	0	0	0	0	0	50	25	0	0	0
M. burdigalensis	0	0	0	0	20	0	0	0	0	80	0	0
M. burdigalensis-negrii	0	0	0	0	0	0	0	0	0	0	100	0
M. negrii	0	0	0	0	0	0	0	0	0	0	0	100

Tab. 5 - Classification results for species in discriminant analysis. The actual groups are classified correctly when they correspond to predicted groups with a high percentage.

	X	ɣ	V	Coeff.
M. gunteri	39.283	0.134	0.682	-200.51
M. gunteri-tani	38.447	0.490	0.571	-165.37
M. tani	35.419	0.574	0.456	-135.55
M. tani-globulina	33.942	0.683	0.509	-123.27
M. globulina	32.491	0.765	0.970	-124.73
M. globulina-intermedia	31.097	0.730	1.336	-129.54
M. intermedia	30.982	0.702	1.738	-148.71
M. socini	39.398	0.593	0.793	-172.82
M. socini-burdigalensis	40.101	0.545	0.934	-183.46
M. burdigalensis	32.102	0.511	1.182	-127.20
M. burdigalensis-negrii	28.125	0.564	1.326	-108.66
M. negrii	29.489	0.691	1.485	-126.52

Tab. 6 - Coefficients obtained by discriminant analysis, useful for classifying new populations. The last column contains a constant in each function.

bles. The coefficients of these functions (discriminant functions) are calculated in such a way to maximize the between-groups variance and covariance matrix on the base of within-groups variance and covariance matrix. The first canonical variate axis is required to be in the direction of greatest variability between the means of the different species, the second axis is chosen to be orthogonal to the first and inclined in the direction of the next greatest variability, and so on.

Discriminant analysis was performed on the data (parameters X, γ and V) of the populations of Miogypsinidae studied in Piedmont and those reported by literature, and three significant discriminant functions were obtained (Tab. 4). The scatterplot built on the base of the first two functions (which summarize the most part of discrimination) is useful for displaying the distinction between groups of populations (Fig. 7). In the table of results of classification for species (Tab. 5), the actual and predicted groups are shown: it appears that in many cases the populations are not classified correctly. But the most important result is the set of coefficients for use in classifying new populations (Tab. 6). A new population is classified by evaluating one function for each character and each species and assigning the population to the species corresponding to the highest function value. We have put these coefficients into a spreadsheet to automatically assign a new population to one species with the highest probability.

Cladistic analysis.

The aim of the cladistic analysis is to construct phylogenies by studying the phylogenetic relationship between species, that is to construct evolutionary trees by considering the transformations of morphological characters during evolution. To apply this approach to the study of Miogypsinidae, first of all we have had to deal with the problem of coding the measured parameters. In fact, mathematical algorithms for the analysis of phylogenetic data require alphanumeric codes that represent character states, but the populations of larger Foraminifera are studied by the measurement of quantitative, continuous parameters. To not derive discrete codes from quantitative data in an arbitrary fashion, we have assigned character states to the parameters of different taxa using statistically significative differences between species (or homogeneous groups of species) resulting from the Analysis Of Variance (ANOVA).

Because of the results of the ANOVA applied to the populations of Miogypsinidae suggest that the variations of parameters between species are highly significant, we used a multiple range test based on the confidence intervals to separate these species in homogeneous groups for each parameter (Tab. 7).

Then we applied a maximum parsimony method, that is a technique to search a phyletic tree that minimizes the amount of evolution needed to explain the avalaible data. As this method requires a prespecified set of constraints upon permissible character changes, we have established to consider parameters X, γ and V like ordered characters (in a progressive series of character states, the transformation of one state to another that requires to skip an intermediate state is not allowed), and parameters D1, D2/D1, ε and ε/D1 like unordered characters (any state of the character is capable of transforming directly to any other state).

The cladogram resulting by applying the analysis to the populations of Miogypsinidae studied in Piedmont (Fig. 8) is better comparable to a pattern of the similarities between species rather than to a hierarchical statement regarding genealogical relationships. This branching diagram, however, shows interesting affinities with the hypothetic phylogenesis of Miogypsinidae proposed by the literature, but shows also unexpected deviations along the *M. socini - M. burdigalensis* branch.

	X		ɣ		V		D1		D2/D1		ε		ε/D1	
Species	Average	Group	Average	Group	Average	Group	Average	Group	Average	Group	Average	Group	Average	Group
M. gunteri	9.90	0	-77.8	0	2.94	0	160.3	1	1.15	0	267.8	1	1.69	1
M. gunt.-tani	9.17	1	-39.0	1	5.43	0	172.8	3	1.13	1	279.2	1	1.63	1
M. tani	8.16	3	-22.9	2	9.16	0	170.5	3	1.11	2	264.0	0	1.57	0
M. globulina	6.50	5	20.2	3	27.83	2	157.3	0	1.14	0	266.4	1	1.71	3
M. glob.-int.	5.84	6	27.5	3	45.04	4	172.1	3	1.22	3	289.8	2	1.71	3
M. socini	8.71	2	-33.3	1	16.90	1	169.4	2	1.16	0	312.8	4	1.87	4
M. burdigal.	7.39	4	-5.0	2	32.83	3	177.3	4	1.11	1	301.1	3	1.70	2

Tab. 7 - Multiple range test applied on the populations studied in Piedmont. The averages of the measurements and the character states (groups) are reported for each parameter. *M. socini-burdigalensis* and *M. negrii* are excluded because of missing data.

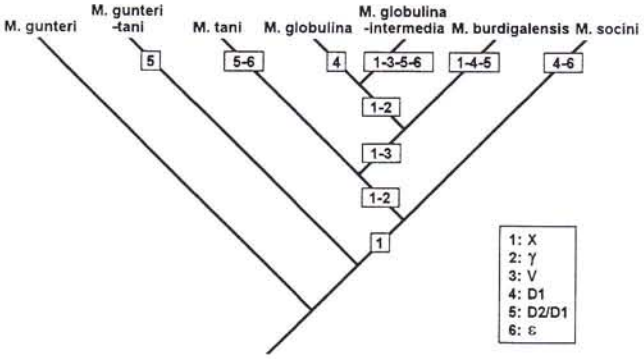

Fig. 8 - Cladogram of species studied in Piedmont. Numbers in square: characters that change unambiguously on branch. Treelenght (number of character changes) = 31.

Discussion.

From the working use point of view, if we compare the results of the different phenetic methods applied to the study of Miogypsinidae, we can see that the most effective tool consists in discriminant analysis. In fact, this technique provides the statistics to assign a new population to a certain species in a non-arbitrary manner. PCA and cluster analysis cannot be seen as tools for the production of a formal classification, but only for data exploration.

Concerning cladistc analysis, the preliminary results suggest to probe the research, especially in improving the characters coding and the selection of constraints upon permissible character changes.

A characteristic shared by cluster analysis and cladistic analysis consists in the big number of algorithms available. Sometimes it appears difficult to choose the best method of studying larger Foraminifera. On the other hand, the possibility of preparing a variety of classification using different techniques stimulates a global exploration of the various algorithms available, once the data are collected and coded for numerical use. This appears more and more valid as computer packages for numerical manipulation of the data are increasingly available.

If we consider that the data base used for this research has been intentionally restricted to a predeterminated number of species of Miogypsinidae, it appears evident that it will be possible to reach more significant results by including other species in the analysis. Moreover, it would be interesting to extend the research to other larger Foraminifera than Miogypsinidae. A study on the Lepidocyclinidae of the Piedmont Basin is in progress: we hope this research will help to verify if the problems of determination are linked to the type of organism or to the palaeontological classification criteria.

Acknowledgements.

The research was supported by the M.U.R.S.T. grants 60% assigned to P. Clari and E. Ferrero. The author wishes to thank E. Ferrero for his assistance throughout the study. Thanks are due also to M. Delpero for his advice and helpful discussion about cladistic analysis.

REFERENCES

Barrai I. (1984) - Metodi di regressione e classificazione in biometria. V. of 170 pp., Edagricole, Bologna.

Camussi A., Möller F., Ottaviano E. & Sari Gorla M. (1986) - Metodi statistici per la sperimentazione biologica. V. of 500 pp., Zanichelli, Bologna.

Drooger C.W. (1952) - Study of American *Miogypsinidae*. V. of 80 pp., Acad. Thesis, Univ. of Utrecht.

Dunn G. & Everitt B.S. (1982) - An Introduction to Mathematical Taxonomy. V. of 152 pp., Cambridge University Press, Cambridge.

Elliott J.M. (1977) - Statistical Analysis of Samples of Benthic Invertebrates. V. of 157 pp., Freshwater Biological Ass., Scientific Publication, 25, Ambleside.

Fermont W.j.j. & Troelstra S.r. (1983) - Early Miocene larger Foraminifera from Cruiser-Hyeres Sea mount Complex (Eastern North Atlantic). *Proc. Kon. Ned. Akad. Wetensch.*, ser.B, v. 86 (3), pp. 243-253, Amsterdam.

Ferrero E., Maia F. & Tonon M. (1992) - The evolutionary pattern of the *Miogypsinidae* in eastern Monferrato (NW Italy). *Paleontologia i Evolució*, n.24-25, pp. 209-217, Barcelona.

Ferrero E., Maia F. & Tonon M. (1994) - Le Miogypsine del Monferrato: aspetti morfologici e tassonomici. *Boll. Soc. Paleont. Ital.*, v. 33 (3), pp. 345-368, Modena.

Forey P.L., Humphries C.J., Kitching I.J., Scotland R.W., Siebert D.J. & Williams D.M. (1992) - Cladistics. V. of 191 pp., Oxford University Press, Oxford.

Schiavinotto F. (1985) - Different evolutionary stages in the Miogypsinidae from Sardinia. *Boll. Soc. Paleont. Ital.*, v. 23 (2), pp. 381-394, Modena.

Vilizzi L. (1991) - Studio biometrico e biostratigrafico sui Miogypsinidi e sui Lepidocyclinidi delle facies carbonatiche mioceniche del Monferrato orientale. Tesi inedita, Univ. Torino.

Received April 12, 1996; accepted October 3, 1996