

Chapter 6. Multivariate morphometrics and its application to monography at specific and infraspecific levels

Karol Marhold

Introduction

Multivariate morphometrics is a powerful tool for assessment of patterns of variation at the specific and infraspecific levels. Unlike the cladistic approach (and other phylogenetic methods) that aims to reconstruct evolutionary relationships among established taxa, morphometrics is particularly useful for drawing lines between taxa, to ascertain differences between different cytotypes or geographical races, or to discover the most important characters that differentiate taxa. As these are issues that often concern the monographer, multivariate techniques can be particularly useful. Multivariate morphometrics comprises a wide spectrum of methodological approaches. The choice of a particular method or approach depends on the data being used, hypotheses to be tested, and questions being asked. Although current use of multivariate methods in taxonomy stems from the phenetic approach in classification (also called numerical taxonomy, statistical systematics, or numerical phenetics), established at the end of the 1950s, it is now acknowledged that this approach has limitations and should be used in situations where similarity reasonably reflects close relationship (i.e., mostly on the level of species and below).

There are numerous examples of successful applications of multivariate morphometrics, especially if it is applied in concert with other methodological approaches such as molecular systematics and/or evaluation of ploidy level either by chromosome counting or flow cytometry (e.g., Brysting & Elven, 2000; Perný & al., 2004, 2005; Smith & Waterway, 2008; Kučera & al., 2010; Rivero-Guerra, 2011). Both these latter methodological approaches are crucial for understanding plant variation. It is common now to first delimit groups of populations (potential taxa) using genetic data (both molecular and ploidy level), subsequently to search for morphological differences among such genetically defined groups,

and finally, to consider appropriate rank for the recognized taxa (e.g., Andrés-Sánchez & al., 2009; Barrett & Freudenstein, 2009; Španiel & al., 2011).

The current role of the “traditional morphometrics” was appropriately summarized by Henderson (2006) in relation to the systematics of palms: “... if we can move toward a more scientific systematics of palms ..., incorporating morphometric methods, then we will have a more stable systematics. We will also have a sound basis on which to study other problems, such as subspecific variation, hybrids and hybrid zones, species complexes and biogeographical patterns. Furthermore, we should complete descriptive systematic studies before we attempt phylogenetic analysis, although these things are often carried out the wrong way round.”

Overall, multivariate methods can be divided into (1) those used for generating taxonomic hypotheses, and (2) those employed for their testing. The first category includes a wide spectrum of clustering and ordination methods such as principal components analysis, principal coordinates analysis, or non-metric multidimensional scaling. The hypotheses-testing methods comprise several kinds of discriminant analyses. Here we provide only a brief account of the multivariate methods used in taxonomy. For more detail the books by Dunn & Everitt (1982, reprinted 2004), Krzanowski (1990), Podani (1994, 2000), or Legendre & Legendre (1998) are recommended. Clustering techniques are nicely explained by Everitt (1986), and discriminant analyses by Klecka (1980). For ordination methods, an excellent source of information is the web page “Ordination methods for ecologists” by Mike Palmer (<http://ordination.okstate.edu/>).

History (phenetic approach)

The phenetic approach in taxonomy dates to the papers by Michener & Sokal (1957) and Sneath (1957a). Charles Michener and Robert Sokal demonstrated a more objective approach to classification of the bee family Megachilidae, whereas Peter Sneath applied a similar approach to classification of bacteria. Subsequently, Sneath and Sokal published their ideas in the book *Principles of numerical taxonomy* (1963), which they developed further in *Numerical Taxonomy: The Principles and Practice of Numerical Classification* (Sneath & Sokal, 1973). In their so-called “neo-Adansonian” principles, they postulated that the greater the content of information supporting taxa in a classification and the more characters on which it is based, the better the classification will be. They considered every character to have equal value in creating “natural” taxa, and they suggested basing classifications on overall similarities. Although it is now broadly acknowledged that the phenetic approach is not often efficacious at higher taxonomic levels, the most important message from numerical taxonomy, i.e., stressing the importance of detailed study of as many characters as possible, is still fully valid.

Moreover, comparisons of individuals, populations, ploidy levels, and/or infra-specific taxonomic entities, still based on overall similarity, can help us improve our classifications. This approach is currently used mostly in morphometric studies, but some methods are applied also in evaluation of molecular data such as in analysis of DNA fragment length polymorphisms. The phenetic approach is based on a wide spectrum of methods of multivariate analysis. Rapid spread of usage of these methods at the end of the 1950s and beginning of the 1960s was closely linked with development of computer technologies. Although most of the multivariate techniques were known for decades, their practical applications were hampered by computational difficulties.

Characters and character states

Morphological, molecular and other characters that characterize plant taxa can be measured or scored in the following scales (Anderberg, 1973; Podani, 2000):

1. Nominal scale—in this case the only valid mathematical operators are equality (=) or inequality (\neq) and character states are non-ordered. Character states (e.g., ovate, obovate, or lanceolate shapes of leaves) are usually coded by numbers, but their choice is purely arbitrary.
2. Ordinal scale—here apart from equality or inequality also operators $<$ and $>$ can be applied, which gives the possibility to order objects according to the degree of presence of a certain property (e.g., density of hairs on leaves). Nevertheless, the exact differences among character states are not captured in this scale.
3. Interval scale—in this scale subtraction and addition also apply, which enable expression of degrees of difference among objects with respect to measured characters. The position of zero on this scale is chosen arbitrarily.
4. Ratio scale—gives the possibility of expressing ratios among objects. Operator of division can also be applied here. The value of zero here means absence of a particular characteristic.

There is also another way to classify characters:

1. Qualitative characters: (a) binary—characters with two character states, mostly presence or absence, usually coded by 0 and 1; (b) multistate—characters with three or more character states, usually coded by integer values 0 to x .
2. Semiquantitative characters (rank-ordered, ordinal)—character states represent in this case usually a small number of ordered classes and the differences between neighboring values are not constant (or precisely defined). Such characters can be treated either as quantitative (with certain reser-

- vations) or qualitative, depending on the particular type of analysis or nature of character.
3. Quantitative characters—here the differences between neighboring values are constant and these characters can be divided into: (a) discontinuous (discrete, meristic) characters with the values expressed in natural numbers without intermediate values, (b) continuous characters with indefinite amount of character states expressed in real numbers.

Sampling design and selection of characters

The sampling design is probably the most important step in the morphometric analysis. It mostly depends on the particular questions or aims of analyses; there are also properties of particular groups of plants that must be taken into consideration. Sampling itself should involve a random element and, at the same time, samples must be taken so as to maximize variation of the collected material. The wider the morphological variation represented in the sample, the better we assess the variability of the species. In estimating sample size, we particularly need to take into account the reproductive system. For example, in autogamous (selfing) plants the amount of sampled populations should be higher, possibly at cost of the number of sampled plants per population, whereas in allogamous (out-crossing) ones, both individual and populational variation should be well represented.¹ The population size should be considered as well, especially in the case of endangered taxa. Sampling populations in the field is obviously much more representative than use of specimens from herbaria. In the latter case one has to be aware that the intra-population variation might be hidden; there are some characters that are not preserved on herbarium specimens. Botanists also tend to collect “unusual” specimens, which are likely to be over-represented in herbaria. In spite of these limitations, however, herbarium collections are irreplaceable in cases of remote areas or highly endangered species. In sampling design, one also has to consider the distribution of studied taxa and type localities, karyological variation (all known cytotypes should preferably be included in the analyses), as well as different habitats occupied. In the selection of morphological characters, those traditionally used in Floras and identifications keys should always be included (to test their relevancy for the particular taxa). However, one should avoid characters strongly influenced by ecological conditions or those from not yet fully developed organs. For some analyses highly correlated pairs of characters or those invariable within particular groups should be avoided (see below).

¹ Gilmartin (1974) and Gilmartin & Hart (1986) proposed a method using the coefficient of phenetic variation for estimating the number of specimens to be sampled in a population (or taxon).

Resemblance coefficients

Once the data matrix (with the objects in rows and characters in columns, or vice versa) is available, the next step in multivariate analyses is the computation of a secondary data matrix, expressing pair-wise interrelationships among objects (similarity, dissimilarity) or characters (correlation). There is a wide spectrum of available resemblance coefficients and only few representative ones are given here. More details can be found, e.g., in Podani (1994, 2000) or Legendre & Legendre (1998).

Resemblance coefficients can be generally classified into the following four categories:

(1) Coefficients of distance for quantitative and binary characters (metric distances):

The most widely used coefficient of distance is the Euclidean one:

$$EU_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

where x_{ik} is the value of character k for the object i , x_{jk} is the value of character k for the object j , p is the total number of characters included in the analysis.

An alternative possibility is represented by the Manhattan (Hemmings, city-block) metric:

$$CB_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

This metric gives greater weight to differences among higher numbers of characters (its value does not depend so much on the considerable difference in just one character).

The Minkovski metric represents a more generalized case:

$$MNK_{ij} = \sqrt[r]{\sum_{k=1}^p (x_{ik} - x_{jk})^r}$$

This includes Manhattan metrics (where $r = 1$) and Euclidean distance (where $r = 2$) as special cases.

It should be taken into consideration that all the above-mentioned coefficients are dependent on the scale at which characters are measured. Therefore, if

characters are measured at different scales (which is often the case), they should be standardized prior to the analyses to avoid unequal influences on the results. One commonly used standardization method is by standard deviation:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i}$$

where s_i is a standard deviation of character i , x_{ij} is the value of the character i on the object j and \bar{x}_i is the mean value of the character i .

Mahalanobis generalized distance eliminates the overweighting, which might be due to correlation of characters, and which appears, e.g., when Euclidean distance is used (Podani, 2000):

$$D_{ij} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{W}^{-1} (\mathbf{x}_i - \mathbf{x}_j)} = \sqrt{\sum_{k=1}^p \sum_{r=1}^p (x_{ik} - x_{jk}) w_{kr} (x_{ir} - x_{jr})}$$

where \mathbf{x}_i and \mathbf{x}_j are column vectors corresponding to objects i and j , respectively, \mathbf{W}^{-1} is the inverse of the variance-covariance matrix and w_{kr} is one of its elements (where k and r are characters). This distance coefficient is used in the discriminant analyses.

(b) Similarity coefficients for binary data:

Coefficients designed for binary data should be used when the primary data matrix consists exclusively of binary characters (presence/absence data). There is a large number of coefficients for binary data that are used in taxonomic and especially ecological applications. A detailed account of such coefficients can be found in Podani (2000); only the most important ones are presented here. Suppose that we have objects i and j , characterized in the contingency table as follows:

		Object i	
		+	-
Object j	+	a	b
	-	c	d

a , the number of characters for which both compared objects have a positive value (+ or 1) (positive match)

b , the number of characters for object i with a negative value (- or 0) and object j with a positive value (+ or 1) (mismatch)

c , the number of characters for object i with a positive value (+ or 1) and object j a negative value (– or 0) (mismatch)

d , the number of characters for which both compared objects have a negative value (– or 0) (negative match).

There are two possible basic alternatives for computing the similarity coefficient, depending on whether negative matches should be taken into consideration. In the simple case where we code an ovate leaf shape as 1 and lanceolate as 0, both positive and negative matches make sense and both values “ a ” and “ d ” should be included in computing the similarity coefficient. However, in the case that negative or zero value means an absence of a given character, and this absence can be caused by multiple reasons, then a negative match does not necessarily mean that the objects are similar with respect to this character. In such a case value “ d ” should not be used in computing the similarity coefficient.

Depending on whether negative matches are taken into consideration or not, two coefficients are most widely used: simple matching

$$SM = \frac{a + d}{a + b + c + d},$$

and Jaccard

$$JAC = \frac{a}{a + b + c}.$$

There are also coefficients that accord “ a ” and “ d ” values in an asymmetrical way, giving different weights to “ b ” and “ c ” compared to “ a ” and “ b ”.

(c) Coefficients for mixed data:

The above-mentioned coefficients do not apply to cases where the primary matrix consists of a mixture of binary, multistate qualitative, ordinal, and quantitative characters. For such cases coefficients for mixed data are available (Gower, 1971; Podani 1980, 1999).

(d) Correlation coefficients:

Correlation coefficients, unlike the above-mentioned ones that measure similarities or dissimilarities among objects, give evidence about the relations among measured characters.

The Pearson product-moment correlation coefficient measures linear dependence between two variables in the scale from –1 to +1. Its use requires a normal distribution of characters:

$$r_{12} = \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 \sum_{i=1}^n (x_{i2} - \bar{x}_2)^2}},$$

where n is the number of objects, x_{i1} is the value of character 1 for the object i , x_{i2} is the value of character 2 for the object i , \bar{x}_1 and \bar{x}_2 are the mean values of characters 1 and 2, respectively.

In the case that monotonic, non-linear dependence of characters is expected, rank (non-parametric) coefficients should be used. Spearman's rank correlation coefficient is then one of the possible choices:

$$sp_{12} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n},$$

where d_i is the difference in the order of object i on characters x_1 and x_2 (for all characters, the order of objects is determined, and values of characters x_{i1} to x_{in} are replaced by the first n integers, i.e., by the order of objects on each character).

Covariance is in a sense similar to the product-moment coefficient, except that it is conditioned upon commensurability (it should be used only in cases where characters are measured in the same scales) and has neither lower nor upper bounds:

$$Cov(x_1, x_2) = \frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2).$$

Cluster analysis

Clustering is a generic term for a wide range of procedures that can be used for the building of hierarchical or non-hierarchical classifications from multivariate data. Everitt (1986) described them as follows: "Given a number of objects or individuals, each of which is described by a set of numerical measures, devise a classification scheme for grouping of objects into a number of classes such that objects within classes are similar in some respect and unlike those from other classes." The results of the hierarchical clustering are generally presented as dendrograms, while the non-hierarchical methods sort objects into a predefined number of non-hierarchical groups. Nevertheless, in taxonomic applications only a small

portion of available clustering methods is used. They mostly comprise sequential, agglomerative, hierarchical methods creating non-overlapping clusters (often abbreviated as SAHN methods). Agglomerative methods as opposite to divisive ones proceed in successive fusions of objects into groups and sequential methods proceed in subsequent steps as opposed to those that proceed in a single step. The SAHN methods comprise, among others, single linkage, complete linkage, average linkage, median, centroid and Ward's methods. They are based on the optimization of distance among objects and/or clusters, except the last one, which is based on optimization of homogeneity of clusters. More details than given here can be found in Podani (2000) or in Everitt (1986).

Each cluster analysis begins with computing the similarity or distance matrix. The choice of a similarity or distance coefficient obviously depends on the properties of data. If the characters are not measured in the same scale, they should be standardized prior to computation of the matrix.

Single linkage clustering (the nearest neighbor method; Florek & al., 1951; Sneath, 1957b) defines the distance between clusters as the distance between the most similar (or least dissimilar) pair of objects, considering only one object from each cluster (Fig. 6.1). The main disadvantage of this method is the so-called chain effect, which is that the initial clusters tend to attract other objects one by one, resulting in the pattern depicted in Fig. 6.6A.

Complete linkage clustering (the furthest neighbor method; Sørensen, 1948; Lance & Williams, 1967), is the opposite of the single linkage method. It defines the distance between clusters as the distance between the least similar (or most dissimilar) pair of objects, considering again only one object from each cluster (Fig. 6.2).

The average linkage clustering method (group-average clustering, UPGMA—unweighted pair-group method using arithmetic averages; Sokal & Michener, 1958;

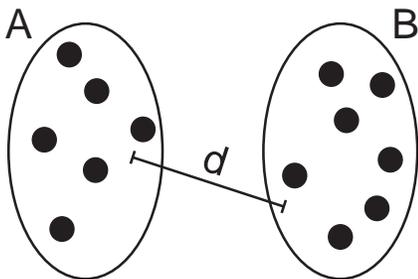


Fig. 6.1. Geometric interpretation of the single linkage clustering method. The distance between clusters A and B is defined as the distance between the most similar (or least dissimilar) pair of objects.

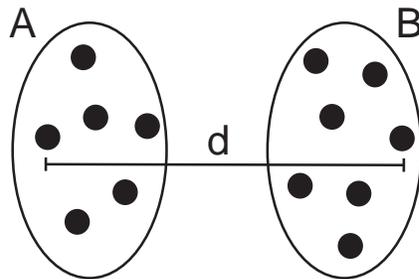


Fig. 6.2. Geometric interpretation of the complete linkage clustering method. The distance between clusters A and B is defined as the distance between the most dissimilar (or least similar) pair of objects.

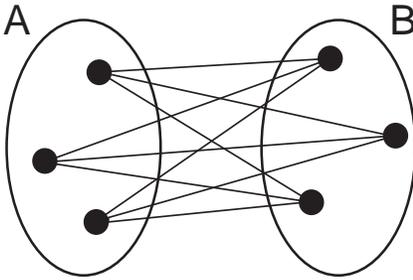


Fig. 6.3. Geometric interpretation of the average linkage clustering method. The distance between clusters A and B is defined as the average of the distances (or similarities) between all pairs of objects.

Lance & Williams, 1966) defines the distance between two clusters as the average of the distances (or similarities) between all pairs of objects that comprise one object from each cluster (Fig. 6.3).

In the centroid clustering method (Gower's method, UPGMC—unweighted pair-group method using centroids; Sokal & Michener, 1958), once a cluster is formed it is replaced by the mean values of all characters (by centroids) and the distance between clusters is defined as that between their centroids (Fig. 6.4).

The median method (WPGMC—weighted pair-group method using centroids, weighted centroid clustering; Gower, 1967) differs from the centroid one in that the new centroid of two merged clusters is halfway between their centroids, which is closer to the smaller cluster, compared to the situation when the size of clusters is taken into consideration (Fig. 6.5).

The Ward's method (minimization of the increase of error sum of squares; Ward, 1963) is based on the idea that the loss of information resulting from the merger of individuals into clusters can be measured by the increase of the sum of squared deviations of every point from the centroid of the cluster to which it belongs. In each step, all possible pairs of clusters are considered and those two that result in the minimum increase of the error sum of squares are merged. Compared with the other clustering methods, Ward's method produces more compact clusters (see Fig. 6.6D).

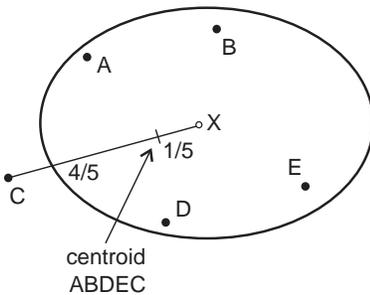


Fig. 6.4. Geometric interpretation of the centroid clustering method.

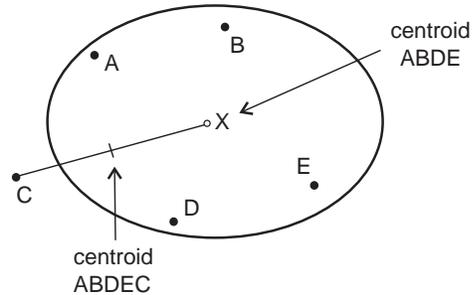


Fig. 6.5. Geometric interpretation of the median clustering method.

The general rule in clustering methods is that there is no single method that is best in any situation. Therefore it is advisable to use more than one clustering method and to search for similar clusters in resulting dendrograms. It is likely that the clusters that appear on dendrograms resulting from several clustering methods reflect structure in the dataset, whereas those that appear only in one dendrogram reflect properties of a particular clustering method, rather than properties of data. The example on Fig. 6.6 shows the results of four different cluster analyses based on the same dataset of morphological characters (binary and quantitative) of populations of *Cardamine amara* from the Carpathian and Sudeten mountains, using Euclidean distance (*Cardamine amara* dataset; Marhold, 1992). There are only two clusters of populations merging on the highest

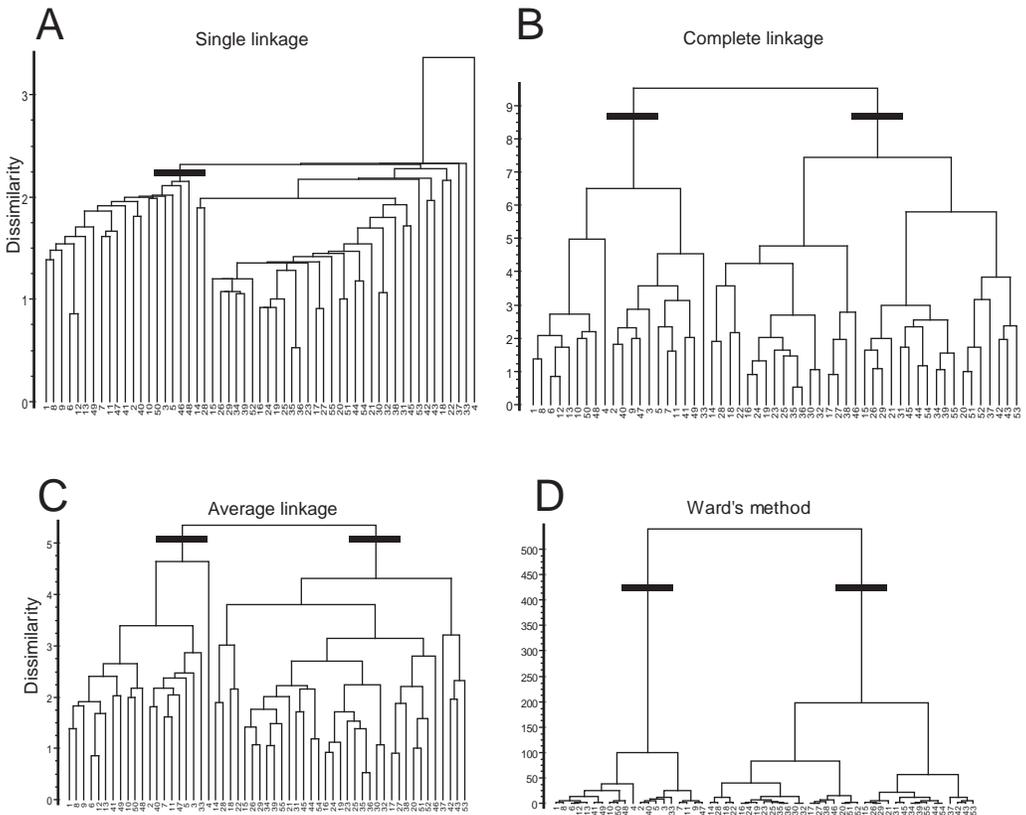


Fig. 6.6. Dendrograms resulting from the four different clustering methods (A, single linkage; B, complete linkage; C, average linkage; D, Ward's methods) based on the same dataset (*Cardamine amara* dataset, with populations characterized by mean values of characters as objects; Marhold, 1992). Only division into two clusters on the highest level is common to all these dendrograms (except the single linkage method, where only one of these clusters appears). Cluster analyses were computed with the SYN-TAX 2000 software (Podani, 2001).

level that are common to these dendrograms (with the sole exception of the single linkage method). Most of the remaining subclusters are artifacts of the particular methods.

Dunn & Everitt (1982) concluded that the single linkage method is in most cases the least successful (due to phenomenon of chaining), whereas average linkage and Ward's methods do fairly well, overall.

The results of cluster analyses can be seriously distorted by the presence of outliers, which may represent atypical individuals or populations or even misidentified material belonging to a different taxon. Outliers, therefore, are better excluded from the analyses. Clustering methods are generally appropriate for datasets that describe hierarchical variation, but they are least convenient for datasets dealing with clinal variability, i.e., where variation of the characters reflects some environmental gradient.

Minimum spanning trees

The method of building minimum spanning trees (Gower & Ross, 1969) is closely related to clustering methods, particularly to single linkage. A minimum spanning tree is a graph that connects all objects in a way that the resulting sum of the edge lengths is the minimum and at the same time leaving no loops in the graph (Fig. 6.7). The main difference from the single linkage clustering method is

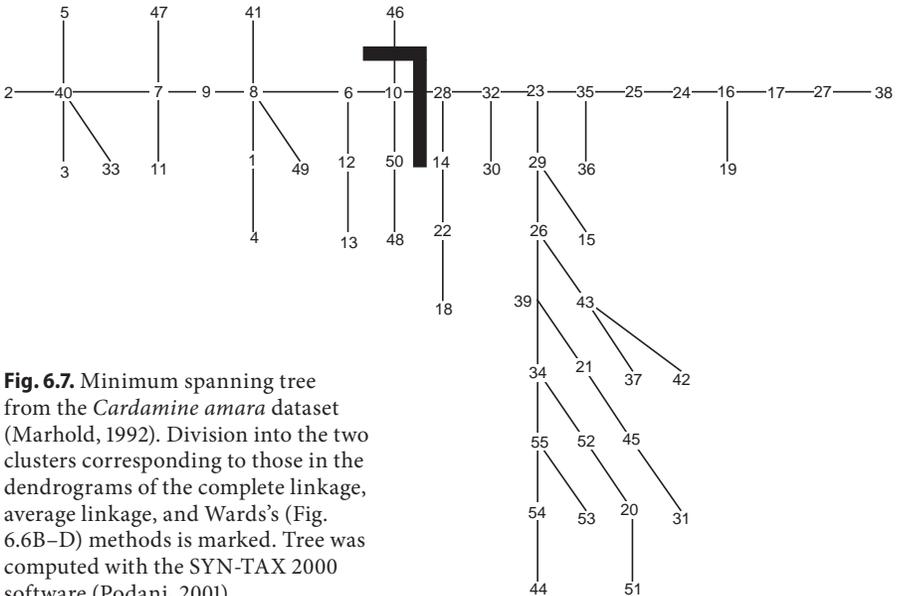


Fig. 6.7. Minimum spanning tree from the *Cardamine amara* dataset (Marhold, 1992). Division into the two clusters corresponding to those in the dendrograms of the complete linkage, average linkage, and Ward's (Fig. 6.6B–D) methods is marked. Tree was computed with the SYN-TAX 2000 software (Podani, 2001).

that each vertex on the graph corresponds to a concrete object. Minimum spanning trees are essential in checking the validity of two-dimensional ordination displays (Dunn & Everitt, 1982; Podani, 2000). The tree projected onto the ordination diagram (Fig. 6.8) shows whether the close position of the objects on the diagram represents an artifact or not. It can reveal, for instance, the “horseshoe effect” on the ordination diagram of the principal components analysis.

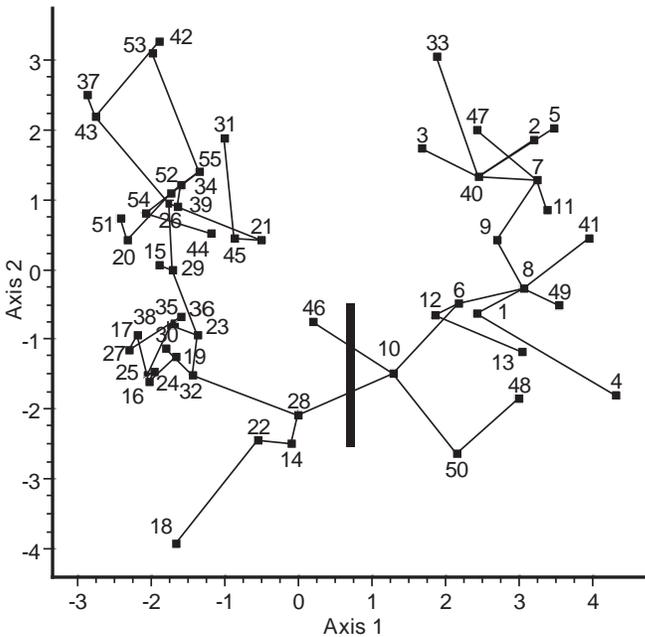


Fig. 6.8. Minimum spanning tree from the *Cardamine amara* dataset (with populations characterized by mean values of characters as objects; Marhold, 1992) projected onto the ordination diagram of the principal components analysis. The horseshoe effect is visible here. Division into two clusters corresponding to those in the dendrograms of the complete linkage, average linkage, and Ward’s (Fig. 6.6B–D) methods is marked. Diagram was computed with the SYN-TAX 2000 software (Podani, 2001).

Ordination methods

Objects being studied can be interpreted as points in a p -dimensional space, where p equals the number of characters measured and/or scored on these objects. In the case that the objects are characterized by three characters only,

one can easily check the distances among them and subsequently also their relationships in 3-D space. Nevertheless, with an increasing number of characters such possibility is lost. For such cases, methods faithfully depicting the relationships among objects in a space of reduced number of dimensions are available. Goodall (1954) coined the term “ordination” for such methods, which is currently interpreted as “any technique that extracts artificial variables in order to reduce the dimensionality of data” (Podani, 2000). Probably the most widely used ordination method is principal components analysis, followed by principal coordinates analysis and non-metric multidimensional scaling. Canonical discriminant analysis, treated here separately, is also sometimes placed into the category of ordination methods.

Principal components analysis

Principal components analysis (PCA; Pearson, 1901; Hotelling, 1933) is a technique that, as with other ordination techniques, reduces dimensionality of the original character space. It transforms an original set of p characters x_1, x_2, \dots, x_p into a new set of uncorrelated characters y_1, y_2, \dots, y_p , where each character y is a linear combination of the original set of characters x (i.e., $y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p$; $y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p$; \dots $y_p = a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pp}x_p$).

The first component axis is derived to encompass the highest percentage of variation among objects (Fig. 6.9). Similarly, the second, the third, and remaining component axes are derived to explain the highest percentage of variation left after derivation of previous axes. At the same time, relative positions of objects in

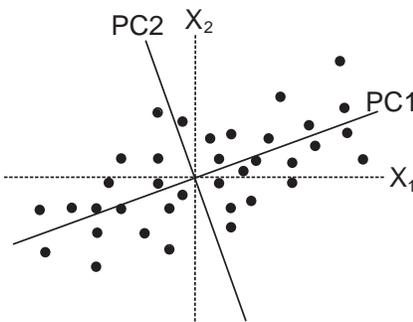


Fig. 6.9. Schematic presentation of the original character plane (defined by the axes x_1 and x_2) and corresponding principal component axes (PC1 and PC2), positioned in the direction of the highest (PC1) and the second highest (PC2) variation among objects.

the original character space and that in the space determined by the principal components are the same (just the coordinate systems are changed). As the first few new dimensions capture most of the variation of the original dataset, it is possible to check visually the relationships of objects on a two-dimensional plane (Fig. 6.10) or in three-dimensional space (Fig. 6.11) instead of having as many dimensions as original characters (which is impossible to control visually).

Nevertheless, it must be taken into consideration that in the case that only the first few principal components are depicted, the spatial posi-

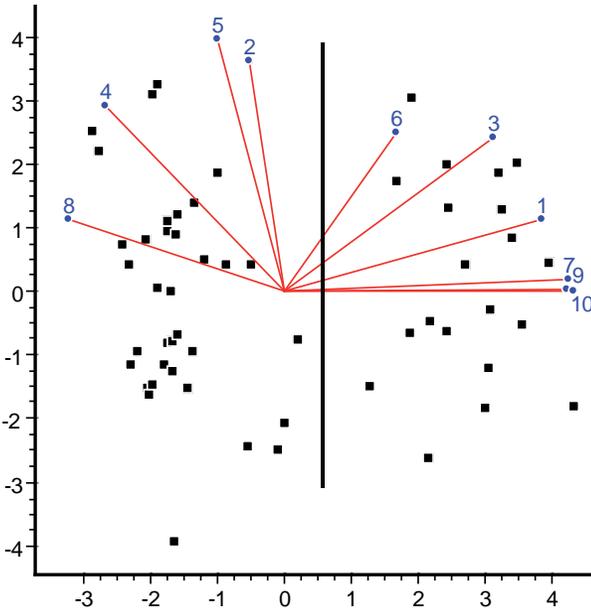


Fig. 6.10. Biplot diagram of the principal component analysis of the *Cardamine amara* dataset (the same as depicted in Fig. 6.11 on a three-dimensional ordination diagram). Scale factor for variables is 4.49. The lines of eigenvectors (printed in red) of the characters 7, 9 and 10 are almost parallel with the first axis, indicating the highest influence of these characters on the first principal component axis and thus on the division of two groups of objects as defined by the complete linkage, average linkage and Ward's (Fig. 6.6B–D) methods (marked by the vertical line). Diagram was computed with the SYN-TAX 2000 software (Podani, 2001).

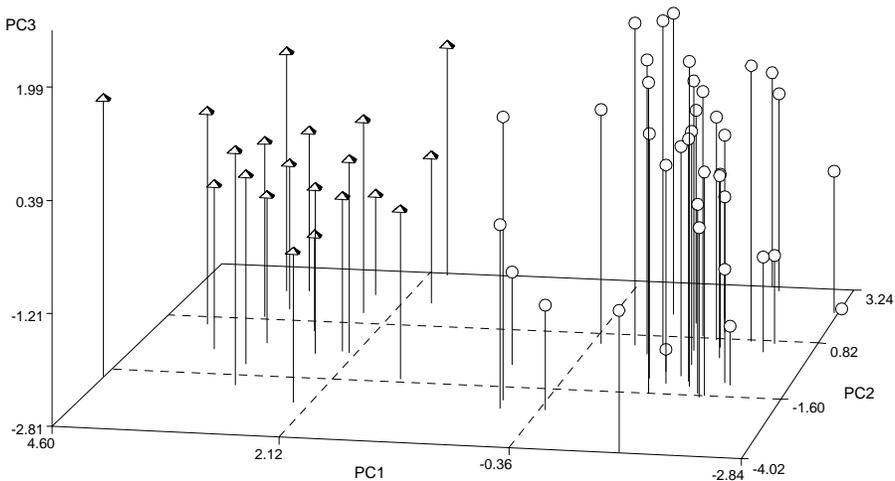


Fig. 6.11. Three-dimensional ordination diagram of principal components analysis from the *Cardamine amara* dataset (with populations characterized by mean values of characters as objects; Marhold, 1992). Based on eigenvalues, the first three components account for 50.3%, 25.9%, and 11.3% of variation of the original dataset, respectively. Objects (populations) belonging to the two different clusters in the dendrograms of the complete linkage, average linkage, and Ward's (Fig. 6.6B–D) methods are marked by different symbols. PCA was computed using SAS 9.1.3 software (SAS Institute, 2007).

tion of the objects is only an approximation of their relative positions in the original character space, and that their Euclidean distances are also only an approximation.

PCA is based on reduction of correlations among characters (the new axes, principal components are uncorrelated), so its success depends on the amount of correlations among the original characters. The higher the correlation is among them, the more successful is the result of PCA (i.e., fewer axes are needed to display the principal relationships among objects).

Computation in principal component analysis is based on eigenanalysis of the symmetrical matrix (i.e., matrix with the same number of columns and rows). This can be either a correlation or covariance matrix among characters. In the case that characters have not been measured in the same scale and a covariance matrix is used, data should be standardized prior to analysis. The eigenanalysis results in a set of eigenvalues and eigenvectors. There are as many eigenvalues and eigenvectors as rows in the symmetrical matrix and as principal components. The eigenvalues indicate the proportion of variation of the original dataset expressed by the particular component axis, and they are usually presented as a percentage of their total sum. In the case that a correlation matrix is used, the sum of the eigenvalues equals the number of characters; in the case of the covariance matrix, it equals the sum of the variances of all characters. The eigenvectors (direction cosines) express the direction of vectors characterizing the influence of the original characters on the principal component axes. Eigenvectors can be depicted in an ordination diagram of variables or in a joint display (called the biplot) with the result of ordination of objects (Fig. 6.10). Such joint display of ordination of objects and characters requires multiplication of the character scores by an appropriate factor, as the coordinates of objects and characters are expressed on different scales. As an aid for interpretation, lines are drawn from the origin towards the points representing characters (Fig. 6.10; for details see Podani, 2000).

Although the principal components analysis was originally developed for quantitative characters with normal distribution, it can be used, with certain reservations, also for binary and semiquantitative characters and is considerably robust also with respect to departures of characters from normal distribution. Nevertheless, presence of binary data often causes the horseshoe effect (Fig. 6.8). The only strict limitation of PCA is with regard to the number of analyzed characters. It should be lower than the number of analyzed objects.

Principal components analysis is often confused with factor analysis. It should be stressed that these are two different multivariate methods. Unlike PCA, in factor analysis new axes (factors) are extracted to account maximally for the covariances (the shared variances) of variables. The main aim is to reveal the correlation structure of variables. The results of factor analysis are usually presented by an ordination of variables.

Principal coordinates analysis (metric multidimensional scaling, classical scaling)

Principal coordinates analysis (PCoA; Torgerson, 1952; Gower, 1966) is a method related to PCA. The procedure places analyzed objects into ordination space defined by the principal coordinates (new axes, equivalent to principal components). Euclidean distances among objects in the PCoA ordination diagram approximate the original distances among objects based on any coefficient of similarity or distance. This method is appropriate especially in cases when one has to deal with characters that cannot be used in principal component analyses, such as binary characters, where negative matches are not meaningful, or with mixed characters that include multistate qualitative characters, where Gower's coefficient (Gower, 1971) is appropriate. This method is useful also in cases when the number of analyzed characters exceeds that of analyzed objects.

Somewhat simplified, the principal coordinates analysis involves two steps. In the first step, a secondary similarity (or dissimilarity) matrix is computed from the primary data matrix, using any similarity or dissimilarity coefficient. In the second step, a symmetric matrix, equivalent to the correlation or covariance matrix in PCA, is computed, which is, in turn, subject of the eigenanalysis, resulting in an ordination diagram and set of eigenvalues. If the secondary matrix is based on Euclidean distance, the results of PCA and PCoA are fully identical. The main difference between PCA and PCoA is that the relationship between the original characters and principal coordinates is not linear, and therefore we do not have information on the influence of particular characters on coordinate axes. For quantitative characters we can get at least some information in this respect by computing the correlation coefficients between given character values and score on the coordinate axes.

Non-metric multidimensional scaling (NMDS)

Non-metric multidimensional scaling (Kruskal, 1964) is another method, the aim of which is to reduce the number of dimensions in the original character space. The main difference among the NMDS on one hand and PCA and PCoA on the other, is that NMDS does not preserve exact distances among objects in the original character space and approximates only the order of distances among objects. Let us suppose that there are four objects and six dissimilarity values for pairs of objects that can be ordered in the following way: $\delta_{23} < \delta_{12} < \delta_{34} < \delta_{13} < \delta_{24} < \delta_{14}$ (the second and third objects are most similar and the first and fourth are most dissimilar). Further suppose that these objects are placed as points in Euclidean space and their distances are as follows: $d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, d_{34}$. In the

non-metric multidimensional scaling these distances are considered to be in full agreement with the original dissimilarity values if $d_{23} \leq d_{12} \leq d_{34} \leq d_{13} \leq d_{24} \leq d_{14}$ (Kruskal, 1964; Dunn & Everitt, 1982). In other words, the order of distances among objects in new Euclidean space is in agreement with the order of the original dissimilarities (distances are monotonic with the dissimilarities). Like PCoA, NMDS is not limited to Euclidean distances and can be used in combination with any other (dis)similarity coefficient. The number of dimensions (two or three) is a matter of choice of the user, depending on the value of stress (Kruskal, 1964) for a given number of dimensions.

Discriminant analyses

Discriminant analysis comprises several statistical techniques convenient mostly for hypotheses testing. Generally, they can be divided into techniques used for the interpretation of differences among predefined groups of objects (canonical discriminant analysis) and into those whose aim is assignment of objects into groups (classificatory discriminant analysis). A separate category is represented by stepwise discriminant analysis, the aim of which is to select the most useful discriminating characters.

Canonical discriminant analysis gives the possibility of answering two questions: (1) If and to what extent can predefined groups of objects be distinguished based on available characters? (2) Which characters contribute to this differentiation? Classificatory discriminant analyses, on the other hand, derive one or more functions with the aim to identify objects. While canonical discriminant function maximizes separation of predefined groups, classificatory functions minimize the number of misidentified objects (minimize the error rate). For these purposes sometimes different algorithms are needed.

It is important to realize that discriminant analyses do not provide opportunities to test whether there is any other, more successful, grouping of objects in the analyzed dataset than that which is being tested. Groups of objects tested can be defined by certain morphological characters, ploidy levels, distributional areas, or ecological requirements. In any case, the characters that define the groups should not be involved in the analyses to avoid circular argumentation.

There are certain requirements for data in discriminant analyses (Klecka, 1980; Legendre & Legendre, 1998): (1) objects should be characterized by quantitative and binary characters, only; (2) no character can be a linear combination of any other character; (3) any pair of characters cannot be highly correlated; (4) covariance matrices should be approximately equal (homogeneity of covariance matrices is tested by the Bartlett test); (5) the distribution of characters within each group should be multivariate normal; (6) no character can be invariant in any predefined group; and (7) for the number of groups (g), characters (p) and

total number of objects (n) should hold: $0 < p < (n - g)$, and there must be at least two groups, and in each group there must be at least two objects.

There is no general agreement to what extent departures can be tolerated from the above-mentioned requirements, particularly departures from the multivariate normal distribution of characters and equality of the group covariance matrices. Nevertheless, the prevailing opinion is that discriminant analysis is a rather robust technique, which can tolerate some deviations from these assumptions (Lachenbruch, 1975; Klecka, 1980). Tests of significance of canonical discriminant functions (e.g., Wilk's lambda or its associated chi-square test) and classifications based on probability of group membership are among those that are likely to be influenced by departures from the multivariate normal distribution of data (Klecka, 1980).

Canonical discriminant analysis (canonical variates analysis)

Canonical discriminant analysis provides the opportunity to view relationships among objects in space defined by canonical axes. In some respects it resembles the above-mentioned ordination methods (PCA, PCoA, NMDS); nevertheless, it differs from them as the main canonical axes maximize differences among predefined groups rather than among individual objects (those characters that considerably differ among groups and have low within-group variation have the highest influence on canonical axes). The method was originally proposed by Fischer (1936), who incidentally in his original paper used a dataset from plant taxonomy, namely length and width of sepals and petals of 150 plants of three *Iris* species (*Iris setosa*, *I. variegata*, *I. virginica*), measured by Edgar Anderson from the Missouri Botanical Garden. Generally, the linear discriminant function (Klecka, 1980) is a linear combination of discriminant characters: $f_{km} = a_0 + a_1x_{1km} + a_2x_{2km} + \dots + a_px_{pkm}$, where f_{km} is a score of canonical discriminant function for the case m in the group k , x_{ikm} is a value of the discriminative character x_i for the case m in the group k (total amount of characters is p), and a_i are coefficients of discriminant functions ($i = 0, 1, \dots, p$). Coefficients (a) for the first function are derived so that the group centroids (in terms of Mahalanobis distance) are as different as possible. Those for the second function are also derived to achieve maximum distance of centroids, but at the condition that values of the first function are not correlated with the values of the second. The maximum number of discriminant functions that we can derive is equal to the number of variables or groups minus one, whichever value is lower. The extent of between-group variation extracted by the discriminant functions is expressed by corresponding eigenvalues. In interpretation of canonical discriminant analyses it is usually enough to interpret the first few axes as they are ordered according to their importance. To interpret differences among predefined groups we consider relative positions of the group centroids

and individual objects on the histograms (in the case of a single discriminant function, Fig. 6.12) or ordination diagrams (in the case of two or three functions, Figs. 6.13, 6.14).

The actual values of coefficients of discriminant function depend on the scale from which the given character is measured, and therefore they are called unstandardized coefficients. When standardized, they express unique contributions of a given character for the discriminant function. Nevertheless, if two variables share nearly the same discriminating information (i.e., if they are highly correlated), they share also their contribution to the score and consequently, their standardized coefficients may be smaller than when only one of these variables is used. Therefore, the correlation of the original characters and canonical axes is better reflected by the product-moment correlation coefficients between them (called total canonical structure or total structure coefficients). They can be considered to be the cosines for the angles formed by the characters and the discriminant function (Klecka, 1980).

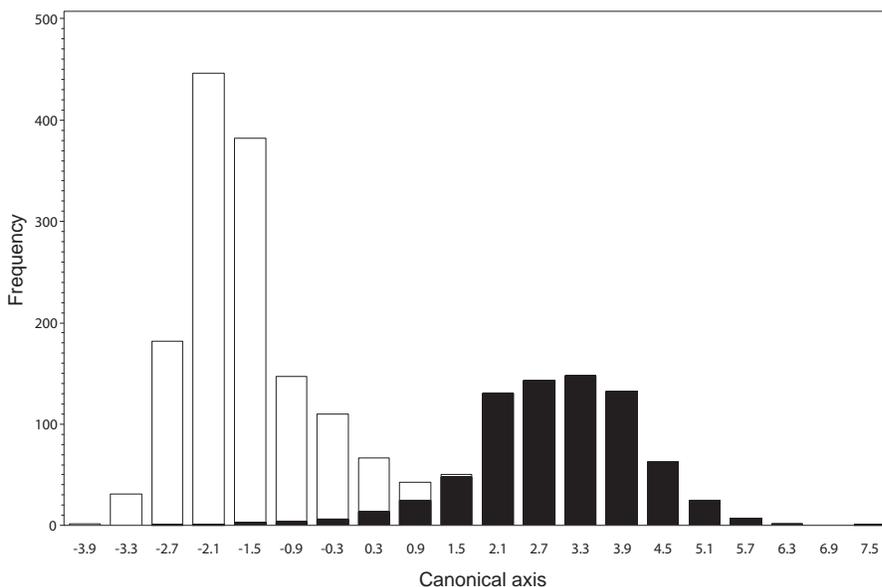


Fig. 6.12. Histogram of the canonical discriminant analysis of the *Cardamine amara* dataset (with individual plants as objects; Marhold, 1992). Two groups in the discriminant analysis correspond to those defined by complete linkage, average linkage, and Wards's (Fig. 6.6B–D) methods (based on the populations, characterized by the mean values of characters). The aim of the analysis was to test whether two groups revealed in the analyses based on population means are recognizable also on the level of individual plants (which is not always the case, see, e.g., analysis of the *Cardamine pratensis* group by Marhold, 1996). Canonical discriminant analysis was computed using SAS 9.1.3 software (SAS Institute, 2007).

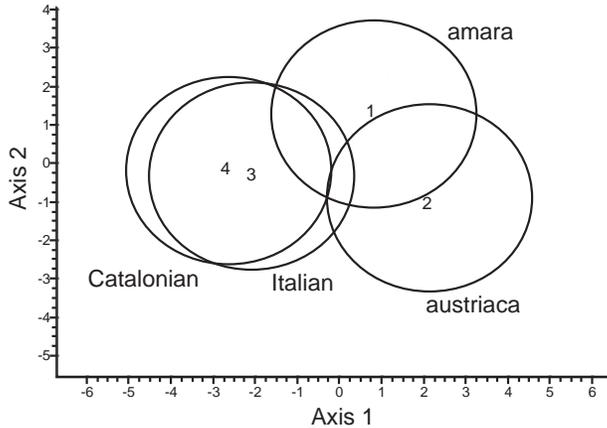


Fig. 6.13. Ordination diagram of the canonical discriminant analysis of the individuals of *Cardamine amara* subsp. *amara* (1), subsp. *austriaca* (2), and Italian (3) and Catalanian (4) populations of *C. amporitana* based on morphological characters (Lihová & al., 2004). 95% isodensity circles, expected to contain 95% of the members of the group (Podani, 2000, 2001) calculated for the two canonical axes are drawn around the centroids. Individual objects are not shown. The centroids are marked by the group numbers. The radius of the circle is independent of the number of objects analyzed. Canonical discriminant analysis was computed using SYN-TAX 2000 software (Podani, 2001).

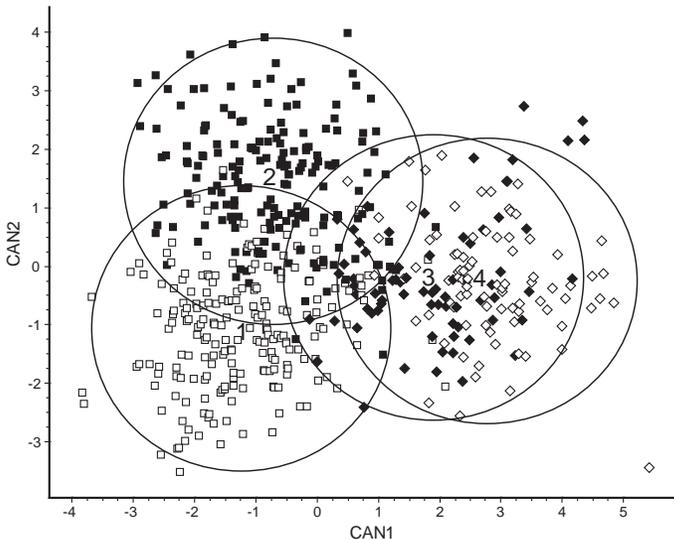


Fig. 6.14. Ordination diagram of the canonical discriminant analysis of the individuals of *Cardamine raphanifolia*, with groups defined based on geographical origin of populations (Perný & al., 2005). 95% isodensity circles, expected to contain 95% of the members of the group, are drawn around the centroids. Individual objects are shown and marked by different symbols. Canonical discriminant analysis was computed using SYN-TAX 2000 software (Podani, 2001).

Classificatory discriminant analysis

There are two main goals of the classificatory discriminant analyses: (1) to derive an identification (classification) criterion to classify objects of unknown status, and (2) to test accuracy of the classification criterion to assess exactness of group separation.

In the first case, classification function or functions are derived based on the training dataset of two or more groups of objects, and then this function is applied to the objects of unknown group membership to determine to which group they most likely belong (or which group they most resemble). There are several possibilities here: (1) The simplest version of the classificatory discriminant analysis is the classification of objects according to the score on the canonical discriminant function. An example of such is the Atkinson discriminant function, which is used for the determination of two species of the genus *Betula* in the *New Flora of the British Isles* (Stace, 2010: 295) (Fig. 6.15), and which was developed by Atkinson & Codling (1986). (2) It is also possible to classify objects according to projection in the canonical ordination space. (3) The other option is to derive a separate function for each group and to compute scores for each object and for each function; the object is subsequently classified into the group with

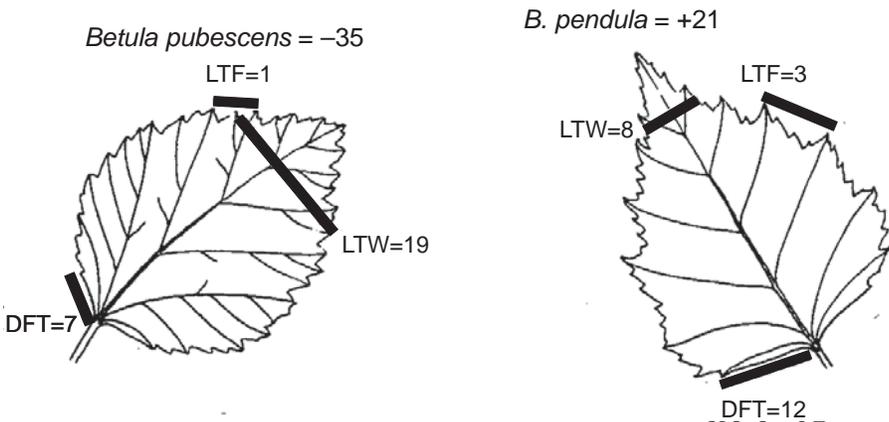


Fig. 6.15. Atkinson discriminant function for identification of *Betula pendula* and *B. pubescens* (Stace, 2010: 295): $12LTF + 2DFT + 2LTW - 23$ (based on means from five short-shoot leaves). LTF, Leaf tooth factor (number of teeth projecting beyond line connecting tips of main teeth at ends of third and fourth lateral veins, subtracted from total number of teeth between these two main teeth); LTW, leaf tip width (width, in mm, of leaf 1/4 distance from apex to base; DFT, distance to first tooth (distance, in mm, from apex of petiole to first tooth). If the solution is greater than zero, the tree is likely to be *B. pendula*; if it is less than zero, the tree is likely to be *B. pubescens*. This method gives a correct rate of classification of 93% (tested against chromosome number; Atkinson & Codling, 1986).

the highest score. (4) The next option is to calculate Mahalanobis distances from the individual objects of each of the group centroids; objects are then classified into the closest group. (5) The option of the classificatory discriminant analysis, usually incorporated in statistical program packages (e.g., SAS Institute, 2007), is based on probability models. Under the assumptions of multivariate normal distribution of characters, most objects cluster near the group centroids, whereas the density of objects diminishes further away from the centroid. Based on this assumption, the probability of group membership is computed for each object and each group. If the dataset conforms to the requirement of having equality of covariance matrices, a simple linear discriminant function (based on the pooled covariance matrix) is computed; otherwise a quadratic discriminant function (based on the within-group covariance matrices) is used. In the case that the data do not show a multivariate normal distribution, non-parametric methods, such as k -nearest neighbors, can be employed.

The accuracy of the classification criterion can be tested either by resubstitution, where computation of the classification criterion is based on the entire dataset, and this criterion is subsequently tested using the same dataset, or by cross-validation, wherein the classification criterion is derived from the dataset of $n-1$ objects and then applied to classify the one individual left out. In the first instance there is only one discriminant function computed, but in the latter there are as many functions as objects. Both approaches can be applied to linear and quadratic functions using parametric methods or to functions derived using non-parametric methods. Results of the analysis are presented by a classification table with the proportion (%) of correctly classified individuals in each group. It is also possible to determine the probability of group membership for each object. An example of results from classificatory discriminant analysis based on a dataset from individual plants in *Cardamine amara* (Marhold, 1992) is given in Table 1.

Table 6.1. Results of the classificatory discriminant analysis of the dataset of *Cardamine amara* (Marhold, 1992). Comparison of the results of the parametric discriminant analysis and non-parametric k -nearest neighbors method. The aim of the analysis was to test whether two groups revealed in the analyses based on population means are recognizable also on the level of individual plants.

Actual group		Predicted group membership (Number of observations and percent classified into groups)	
		Group 1	Group 2
Group 1	Parametric method	1323 / 97.35%	36 / 2.65%
	Nearest neighbor	1353 / 99.56%	6 / 0.44%
Group 2	Parametric method	13 / 1.73%	740 / 98.27%
	Nearest neighbor	7 / 0.93%	746 / 99.07%

Stepwise discriminant analysis

There are some situations when a reduction of the number of characters is needed in order to get a subset of most useful characters that discriminate predefined groups of objects. Some characters might have only low discriminative power, whereas others might be good discriminators on their own. Nevertheless, as they share discriminating information with other characters, their contribution becomes somewhat redundant. The way to eliminate such unnecessary characters is the use of a stepwise discriminant analysis (Klecka, 1980). Characters are added to the subset in subsequent steps. First, the best discriminating character is selected. Then, in a series of steps further characters are added one by one. At each step all remaining characters are tested using the F-to-enter statistic, which tests the additional discriminative power introduced by the character, taking into account the discriminative power achieved by the characters already selected. At the same time, at each step all characters already included are tested using F-to-remove statistic, testing the significance of the decrease of total discriminative power, should the particular character be removed from the list of those already selected (Klecka, 1980). After the final step, the characters are ordered according to the F-to-remove statistic, the character with the highest F-to-remove value being the most important for separation of the predefined groups. As the number of characters is not limited in stepwise discriminant analysis (unlike in canonical and classificatory ones), Álvarez Fernández & Nieto Feliner (2001) used this method to reduce the number of characters from the initial character set, to satisfy the criterion of the maximum possible number of characters (in relation to the number of analyzed individuals and predefined groups) for the subsequent canonical discriminant analysis.

Conclusion

For the monographer, some of the most complex taxonomic situations often occur at the specific and especially infraspecific levels. While intuitive pattern recognition efforts may be successful in disentangling sets of populations and yielding a structure of information suitable for direct classification, sometimes help is required to see more clearly what patterns actually exist (if any). Morphometric methods offer a large tool box of ways to analyze and dissect variation at the populational level in complex situations. Whereas phylogenetic (cladistic) methods are extremely valuable for constructing classifications above the species level within a genus, they are not able to deal with complex mosaic patterns often encountered at the infraspecific level. This is the role and real strength of morphometrics for the monographer.

Acknowledgements

This paper was prepared with financial support from the Ministry of Education, Youth and Sports of the Czech Republic (grant no. 0021620828) and of the Visiting Professorship at the Center of Ecological Research, Kyoto University, Japan. Zuzana Fialová kindly helped with the preparation of several figures. The permission of Cambridge University Press and Clive Stace to reproduce part of the figure from the *New Flora of the British Isles* is very much appreciated.

Literature cited

- Álvarez Fernández, I. & Nieto Feliner, G. 2001. A multivariate approach to assess the taxonomic utility of morphometric characters in *Doronicum* (Asteraceae, Senecioneae). *Folia Geobot.* 36: 423–444.
- Anderberg, M.R. 1973. *Cluster analysis for applications*. New York: Academic Press.
- Andrés-Sánchez, S., Rico, E., Herrero, A., Santos-Vicente, M. & Martínez-Ortega, M.M. 2009. Combining traditional morphometrics and molecular markers in cryptic taxa: Towards an updated integrative taxonomic treatment for *Veronica* subgenus *Pentasepalae* (Plantaginaceae sensu APG II) in the western Mediterranean. *Bot. J. Linn. Soc.* 159: 68–87.
- Atkinson, M.D. & Codling, A.N. 1986. A reliable method for distinguishing between *Betula pendula* and *B. pubescens*. *Watsonia* 16: 75–87.
- Barrett, C.F. & Freudenstein, J.V. 2009. Patterns of morphological and plastid DNA variation in the *Corallorhiza striata* species complex (Orchidaceae). *Syst. Bot.* 34: 496–504.
- Bryusting, A.K. & Elven, R. 2000. The *Cerastium alpinum*–*C. arcticum* complex (Caryophyllaceae): Numerical analyses of morphological variation and taxonomic revision of *C. arcticum* Lange s.l. *Taxon* 49: 189–216.
- Dunn, G. & Everitt, B.S. 1982. *An introduction to mathematical taxonomy*. Cambridge: Cambridge University Press.
- Dunn, G. & Everitt, B.S. 2004. *An introduction to mathematical taxonomy*, reprint. Mineola, New York: Dower Publications.
- Everitt, B.S. 1986. *Cluster analysis*, ed. 2. New York: Gower, Halsted Press.
- Florek, K., Lukaszewicz, J., Perkal, J., Steinhaus, H. & Zubrzycki, S. 1951. Sur la liaison et la division des points d'un ensemble fini. *Colloq. Math.* 2: 282–285.
- Gilmartin, A.J. 1974. Variation within populations and classification. *Taxon* 23: 523–536.
- Gilmartin, A.J. & Hart, J.E. 1986. Computation of variation within and among population samples. *Taxon* 35: 682–684.
- Goodall, D.W. 1954. Objective methods for the classification of vegetation III. An essay in the use of factor analysis. *Austral. J. Bot.* 2: 304–324.
- Gower, J.C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53: 325–338.
- Gower, J.C. 1967. A comparison of some methods of cluster analysis. *Biometrics* 23: 623–637.
- Gower, J.C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27: 857–871.
- Gower, J.C. & Ross, G.J.S. 1969. Minimum spanning trees and single linkage cluster analysis. *Appl. Statist.* 18: 54–64.
- Henderson, A. 2006. Traditional morphometrics in plant systematics and its role in palm systematics. *Bot. J. Linn. Soc.* 151: 103–111.

- Hotelling, H.** 1933. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24: 417–441, 498–520.
- Klecka, W.R.** 1980. *Discriminant analysis*. Sage University Papers series on Quantitative Applications in the Social Sciences, no. 19. London & Beverly Hills: Sage Publications.
- Kruskal, J.B.** 1964. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29: 115–129.
- Krzanowski, W.J.** 1990. *Principles of multivariate analysis*. Oxford: Clarendon Press.
- Kučera, J., Marhold, K. & Lihová, J.** 2010. *Cardamine maritima* group (Brassicaceae) in the amph-Adriatic area: A hotspot of species diversity revealed by DNA sequences and morphological variation. *Taxon* 59: 148–164.
- Lachenbruch, P.A.** 1975. *Discriminant analysis*. New York: Hafner.
- Lance, G.N. & Williams, W.T.** 1966. A generalized sorting strategy for computer classifications. *Nature* 212: 218.
- Lance, G.N. & Williams, W.T.** 1967. A general theory of classificatory sorting strategies. I. Hierarchical systems. *Computer J.* 9: 373–380.
- Legendre, P. & Legendre, L.** 1998. *Numerical ecology*. Amsterdam: Elsevier Science.
- Lihová, J., Marhold, K., Tribsch, A. & Stuessy, T. F.,** 2004. Morphometric and AFLP re-evaluation of tetraploid *Cardamine amara* (Brassicaceae) in the Mediterranean. *Syst. Bot.* 29: 134–146.
- Marhold, K.,** 1992. A multivariate morphometric study of the *Cardamine amara* group (Cruciferae) in the Carpathian and Sudeten mountains. *Bot. J. Linn. Soc.* 110: 121–135.
- Marhold, K.** 1996. Multivariate morphometric study of the *Cardamine pratensis* group (Cruciferae) in the Carpathian and Pannonian area. *Pl. Syst. Evol.* 200: 141–159.
- Michener, C.D. & Sokal, R.R.** 1957. A quantitative approach to a problem in classification. *Evolution* 11: 130–162.
- Pearson, K.** 1901. On lines and planes of closest fit to systems of points in space. *Philos. Mag.*, ser. 6, 2: 559–572.
- Perný, M., Tribsch, A. & Anchev, M.E.** 2004. Intraspecific differentiation in the Balkan diploid *Cardamine acris* (Brassicaceae): Molecular and morphological evidence. *Folia Geobot.* 39: 405–429.
- Perný, M., Tribsch, A., Stuessy, T. & Marhold, K.** 2005. Taxonomy and cytogeography of *Cardamine raphanifolia* and *C. gallaecica* (Brassicaceae) on the Iberian Peninsula. *Pl. Syst. Evol.* 254: 69–91.
- Podani, J.** 1980. SYN-TAX: Számítógépes programcsomag ökológiai, cönológiai és taxonómiai osztályozások végrehajtására (Computer programs for classification of ecological, phytosociological and taxonomic data). *Abstr. Bot.* 6: 1–158.
- Podani, J.** 1994. *Multivariate data analysis in ecology and systematics—a methodological guide to the SYN-TAX 5.0 package*. Ecological Computations Series (ECS) 6. The Hague: SPB Academic Publishing.
- Podani, J.** 1999. Extending Gower's general coefficient of similarity to ordinal characters. *Taxon* 48: 331–340.
- Podani, J.** 2000. *Introduction to the exploration of multivariate biological data*. Leiden: Backhuys Publishers.
- Podani, J.** 2001. *SYN-TAX 2000: Computer programs for data analysis in ecology and systematics; User's manual*. Budapest: Scientia Publishing.
- Rivero-Guerra, A.** 2011. Morphological variation within and between taxa of the *Santolina rosmarinifolia* L. (Asteraceae: Anthemideae) aggregate. *Syst. Bot.* 36: 171–190.
- SAS Institute Inc.** 2007. *SAS OnlineDoc version 9.1.3.*. Cary: SAS Institute. <http://support.sas.com/onlinedoc/913/docMainpage.jsp>.
- Smith, T.W. & Waterway, M.J.** 2008. Evaluating the taxonomic status of the globally rare *Carex roanensis* and allied species using morphology and amplified fragment length polymorphism. *Syst. Bot.* 33: 525–535.

- Sneath, P.H.A.** 1957a. Some thoughts on bacterial classification. *J. Gen. Microbiol.* 17: 184–200.
- Sneath, P.H.A.** 1957b. The applications of computers to taxonomy. *J. Gen. Microbiol.* 17: 201–226.
- Sneath, P.H.A. & Sokal, R.R.** 1973. *Numerical taxonomy: The principles and practice of numerical classification*. San Francisco: Freeman.
- Sokal, R.R. & Mitchener, C.D.** 1958. A statistical method for evaluating systematic relationships. *Kansas Univ. Sci. Bull.* 38: 1409–1438.
- Sokal, R.R. & Sneath, P.H.A.** 1963. *Principles of numerical taxonomy*. San Francisco & London: Freeman.
- Sørensen, T.** 1948. A method for establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol. Skr.* 5(4): 1–34.
- Španiel, S., Marhold, K., Filová, B. & Zozomová-Lihová, J.** 2011. Genetic and morphological variation in the diploid-polyploid *Alyssum montanum* in Central Europe: Taxonomic and evolutionary considerations. *Pl. Syst. Evol.* DOI: 10.1007/s00606-011-0438-y.
- Stace, C.** 2010. *New flora of the British Isles*, ed. 3. Cambridge: Cambridge University Press.
- Torgerson, W.S.** 1952. Multidimensional scaling: I. Theory and method. *Psychometrika* 17: 401–419.
- Ward, J.H.** 1963. Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.* 58: 236–244.