Notulae Botanicae Horti Agrobotanici
Cluj-Napoca

Identifying the Phenotypic Resemblances of the Vine Breeds by Means of Cluster Analysis

Liliana ROTARU [1]

[1] *University of Agricultural Sciences and Veterinary Medicine Iaşi, 3 M. SadoveanuAl., Iaşi 700490, Romania; lirotaru@univagro-iasi.ro*

Abstract

The use of the multi-variate statistical methods for the description and recognition of vine breeds is a new preoccupation in the field of ampelography. Besides other well known methods (genetic and/or enzyme pattern analysis) cluster analysis offer the possibility to study genotypes through their phenotypic expressions manifested in vine leaves. 15 authochtonous vine breeds werepart of this study. They belonged to the proles pontica and orientalis. With the help of cluster analysis it was possible to Fig. out three polythetic groups highlighting the degree of phenotypic resemblance among the analyzed breeds. The results and especially the grouping can be only confirmed with the help of an additional genetic analysis.

Keywords: vine breed, cluster analysis, ampelometry, dendrogram, phenotype

Introduction

To cluster or clustering denotes "to form a group". The cases building a group are linked amongst themselves by objective affinity (kinship) and the agglomeration (settlement) within the group is formed hierarchically. The method was first proposed for statistics by Sokal and Sneath (1963) its application was realized by Lance and Williams (1967) cited by Biji *et al*. (1991). Its use in ampelography was proposed by Campastrini F. *et al*. (1993); Boursiquot J. M. *et al*. (1989, 1997); Cid Alvarez Nuria *et al*. (1994); Rotaru L. (1999); Santiago J. L. *et al*. (2005, 2007).

Theoretical aspects

The cluster method admits the existence of the polythetic groups whose elements are equivalent or similar for several criteria, but not for all characteristics. The constitution of groups is made on the basis of some criteria and decision functions. They measure at the same time the similitude of the elements from the group and the difference among groups.

The principle of this method consists of making a division in each stage of formation by the aggregation of two elements. "*Elements*" are individuals or the objects subjected to the classification itself, or the individual regroupings generated by the classification algorithm.

According to the algorithm a division in q-classes of an assembly of n-individuals, in accordance with a "division hierarchy". This hierarchy has the shape of a "tree" or "dendrogram" and contains n-1 divisions, and it may give an idea about the number classes existing within one population (Silvestroni *et al*. 1996).

The fundamental algorithm of ascendant hierarchical classification comprises the following stages (fig. 1):

Stage 1 – there is a statistic population comprising n-elements (n-individuals);

Stage 2 –the distance matrix among the n-elements is determined and the two elements closest together form a new element. This aggregate forms a new element resulting in the first stage of division in n – 1 classes;

Stage 3 –a new distance matrix is formed after the aggregation and the distances between the new element and the remaining are calculated. This is similar stage 1, but with only n – 1 close elements that aggregate. A new division in n – 2 classes is obtained where the first is included too;

Stage m –new distances are calculated and the aggregation process is repeated until there is a single element resulted from the regrouping of all those n-objects subjected to grouping.

This successive grouping is displayed as a dendrogram in which we insert orderly the values of the index or the distances corresponding to different levels of aggregation.

The classification technique according to the *variance principle* or *inertia calculation* tries to optimize the best division resulted by the aggregation of the elements.

According to Huygens's relation, the decomposition of the total inertia quantity is made in intraclass inertia and interclass inertia (Stefano *et al.,* 1996).

In the initial stage, the intraclass inertia is zero and the interclass inertia is equal to total inertia since at this level of aggregation each element constitutes a class. In the final
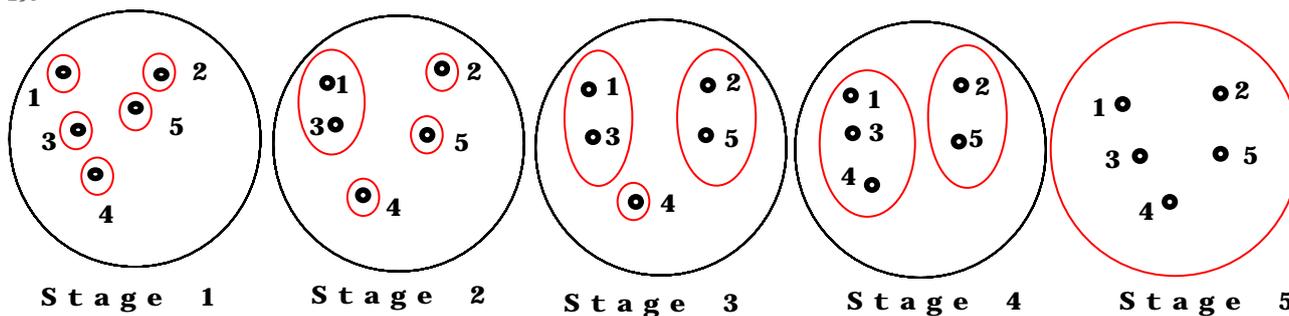
250

Fig. 1. Algorithm of classification – progressive agglomeration of the 5 points - individuals

stage, the interclass inertia is zero and the intraclass inertia is equal to total inertia since now there is a class containing all elements. Consequently, as aggregations are being executed by regrouping of elements, we notice that the intraclass inertia increases whereas the interclass inertia decreases.

Consequently, the principle of the aggregation algorithm according to the variance (inertia) resides in looking for the best division for each class where internal variance of every class may be minimal and thus the interclass variance may be maximal.

The application of this method in ampelography allows the establishment of the phenotypic resemblances of breeds on the basis of the "*closeness*" among the elements subjected to grouping. So, the smaller the chaining index of two breeds, the bigger their phenotypic similitude (Diaz G. *et al.,* 1991; Martinez de Toda F. *et al.,* 1997).

Materials and Methods

We studied 15 autochthonous vine breeds belonging to different ecological-geographical groups (proles).

Within the proles pontica, subproles balcanica the following vine breeds are found: Bătută neagră, Zghihară de Huşi, Galbenă de Odobeşti, Mustoasă de Măderat, Berbecel, Frâncuşă. To the subproles georgica belong the breeds Grasă de Cotnari, Coarnă albă, Fetească regală, Tămâioasă românească, Busuioacă de Bohotin. To the proles orientalis, subproles antasiatica belongs only Coarna neagră, and to the subproles caspica belong the breeds Băbească neagră, Fetească neagră and Fetească albă. From these autochthonous breeds (considered monothetic groups) we harvested during 3 years, 10 adult leaves situated in the middle of summershoots, because in this area the variability of the ampelographic characters is lowest.

51 reference points in the leaf architecture were defined achieving 68 direct ampelometric measurements points. With these data 53 ampelometric values were calcultaed (sums, ratios, product) (Fig. 2). As for the case of symmetric characters we measured and calculated both values. These allowed the build up from a statistic population made of simple series (variation series) of 30 values for all 121 characters established and for all breeds analyzed (15). The elaborated dendrogram of breeds constructed
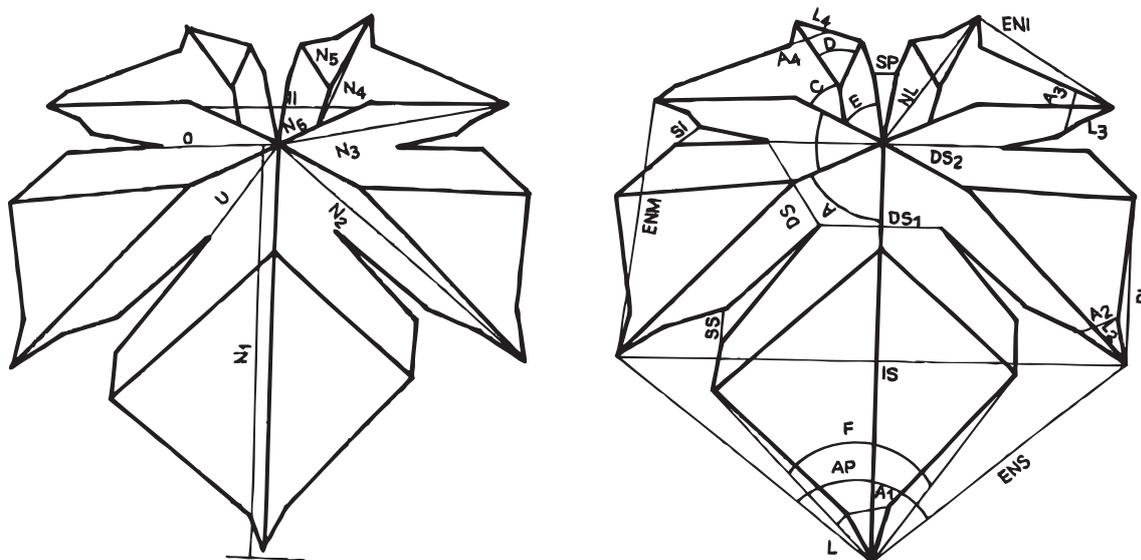

Fig. 2. Ampelometric measurements effectuated on the leaf of vine breeds

Rotaru, L./ Not. Bot. Hort. Agrobot. Cluj 37 (1) 2009, 249-252

251

with the help of the cluster analysis is based on the established groups (branches) which depend on the size of Euclidian distance of the elements subjected to grouping.

The data assembly was represented by the average value of each character analyzed.

Results and discussions

The elaboration of the dendrogram of hierarchized classification of breeds (Fig. 3) was achieved to the principle of minimal inertia loss (Ward criterion generalized). From the analysis of the dendrogram the existence of 3 optimal groups (branches) could be identified and the existence of a level "rupture" between knots 12-13.

Branch A is made of the breeds: Coarnă albă, Berbecel, Frâncuşă, Tămâioasă românească, Busuioacă de Bohotin, Băbească neagră.

Branch B results from the aggregation of: Coarnă neagră, Fetească regală, Grasă de Cotnari, Fetească albă, Fetească neagră.

The third main branch C is formed by the chaining among the breeds Galbenă de Odobeşti, Mustoasă de Măderat, Zghihară de Huşi and Bătută neagră.

The three main groups (branches) may be characterized as follows (Tab. 1):

Group A, is the least homogenous because it aggregates at the highest dissimilarity index 12.1520 and is made from breeds having leaves of different shapes: orbicular slightly cuneiform (Tămâioasă românească, Busuioacă de Bohotin, Frâncuşă, Băbească neagră); orbicular (Coarnă albă); taper (Berbecel), but generally most of the ampelometric data had the lowest values;

Group B is a little more sTab. because the value of the dissimilarity index is less than 8.9445 In this group with
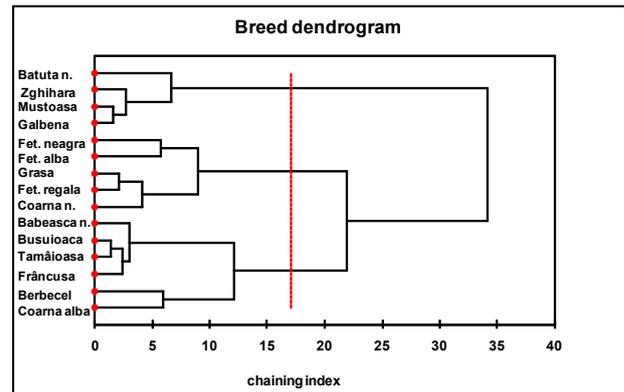

Fig. 3. Dendrogram of hierarchized classification of breeds

cuneiform leaves (Fetească neagră, Coarnă neagră), reniform (Fetească albă), orbicular reniform (Fetească regală), orbicular-cuneiform (Grasă de Cotnari); in contrary the ampelometric data sizes of the breeds were medium;

Group C has the highest homogeneity because the value of the dissimilarity index is 6.6457. It includes breeds with big leaves which had the highest values of the ampelometric data. The shape of the leaves is close together: orbicular-cuneiform (Galbenă de Odobeşti, Mustoasă de Măderat), orbicular cordiform (Zghihară de Huşi) or cordiform (Bătută neagră).

The graphic representation of the chaining index values may be seen in fig. 4.

CONCLUSIONS

Cluster analysis admits the existence of polythetic groups (similar groups of breeds but not for all characters) and therefore allow the division of breeds under study in branches according to the existing dissimilarity or simili-

Tab. 1. Levels of breed chaining in the elaboration of dendrogram

| Knot making	No. of breeds in the knot	Value of chaining index
Tamâioasa româneasca - Busuioaca de Bohotin	2	1.3897
Galbena de Odobesti - Mustoasa de Maderat	2	1.5740
Feteasca regala - Grasa de Cotnari	2	2.1244
Frâncusa - Tamâioasa româneasca - Busuioaca de Bohotin	3	2.4340
Galbena de Odobesti - Mustoasa de Maderat - Zghihara de Husi	3	2.7144
Frâncusa - Tamâioasa româneasca - Busuioaca de Bohotin - Babeasca neagra	4	2.9856
Coarna neagra - Feteasca regala - Grasa de Cotnari	3	4.0913
Feteasca alba - Feteasca neagra	2	5.7765
Coarna alba - Berbecel	2	5.9572
Galbena de Odobesti - Mustoasa de Maderat - Zghihara de Husi - Batuta neagra	4	6.6457
Coarna neagra - Feteasca regala - Grasa de Cotnari - Feteasca alba - Feteasca neagra	5	8.9445
Coarna alba - Berbecel - Frâncusa - Tamâioasa româneasca - Busuioaca de Bohotin - Babeasca neagra	6	12.1520
Coarna alba - Berbecel - Frâncusa - Tamâioasa româneasca - Busuioaca de Bohotin - Babeasca neagra - Coarna neagra - Feteasca regala - Grasa de Cotnari - Feteasca alba - Feteasca neagra	11	21.9544
Total inertia	15	34.1896
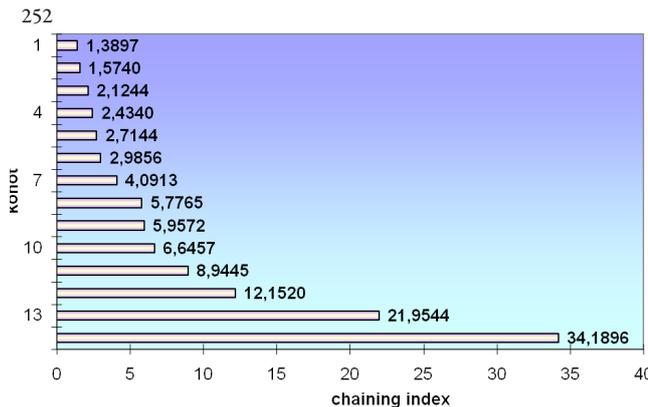
Fig. 4. Histogram of breed chaining

tude. There were three polythetic groups of breeds (A, B, C):

- Group A with the breeds Tămâioasă românească, Busuioacă de Bohotin, Frâncuşă, Băbească neagră, Coarnă albă and Berbecel, as the least homogenous, with the highest value of the dissimilarity of 12.1520;

- Group B with the breeds Fetească neagră, Coarnă neagră, Fetească albă, Fetească regală and Grasă de Cotnari, with a medium stability, and a dissimilarity index of 8.9445.

- Group C with the breeds Galbenă de Odobeşti, Mustoasă de Măderat, Zghihară de Huşi and Bătută. It is the most homogenous, and a dissimilarity index of 6.6457.

The use of the cluster analysis allowed the examination of the hypothesis if the breeds Galbenă de Odobeşti, Zghihară de Huşi and Bătută neagră have a common genetic origin. This is certified because these three breds are located within the same knot. It is also possible to add Mustoasă de Măderat to this group. It belongs to the same ecologic-geographical group (proles pontica, subproles balcanica). Only the genetic analysis may confirm the kinship of these breeds.

The cluster analysis may also be extended to other plants provided that one should use the biometric methods of investigation for the phenotype expression.

Acknowledgements

References

Biji, E. T. Baron, L. Tovissi, P. Wagner, Al. Isac Maniu, M. Korka and D. Porojan (1991). Statistică teoretică si economică. Editura Didactică si Pedagogică, Bucuresti.

Boursiquot, J. M., L. Vignau and J. C. Boulet (1989). Ricerche sull'utilizzazione dell'ampelometria. Rev. Vitic. Enol. 42:37-52.

Boursiquot, J. M. and P. This (1997). Les nouvelles techniques utilisées en ampelographie: informatique et marquage. Journal International des Sciences de la Vigne et du Vin. 40(1):13-23.

Campastrini, F., R. Anzani, O. Failla, F. Iacono and A. Scienza (1993). Application de l'analyse phyllometrique a la clasification geographique de la population italienne de la vigne sauvage. Journal International de la Vigne et du Vin. 27(4):225-262.

Cid Alvarez, N., J. M. Boursiquot, S. Otero, M. Pilar and L. Romani Martinez (1994). Differenciation des cepages autochtones du Nord-Ouest de l'Espagne (Galice) et elaboration d'une cle de determination basse sur l'ampelometrie. Journal International de la Vigne et du Vin. 28(1):1-17.

Diaz, G., M. Setzu, A. Diana, I. Loi, B. De Martis, M. Pala and M. Boselli (1991). Analyse de Fourier de la forme de la feuille de vigne. Premiere application ampelometrique sur un echantillon de 34 cépages implantés en Sardaigne. Journal Internat. des Sciences de la Vigne et du Vin. 25(1):37-49.

Martinez de Toda, F. and J. C. Sacha (1997). Caractérisation ampélographique des cultivars rouge de Vitis vinifera L consérves en Rioja. Bull. OIV. 70(793-794), 221-234.

Rotaru Liliana, 1999-Analiza cluster în ampelometrie. Lucr. St. ale USAMV Iasi. seria Hortic. 1(42):53-60.

Santiago, J. L., S. Boso, M. Vilanova and M. C. Martínez (2005). Characterisation of cv. Albarín Blanco (Vitis vinifera L.). Synonyms, homonyms and errors of identification. Journal Internat. des Sciences de la Vigne et du Vin. 39:57-65.

Santiago, J. L., S. Boso, P. Gago, V. Alonso-Villaverde and M. C. Martínez (2007). Molecular and ampelographic characterisation of Vitis vinifera L. 'Albariño', 'Savagnin Blanc' and 'Caíño Blanco' shows that they are different cultivars. Spanish Journal of Agricultural Research. 5(3):333-340.

Silvestroni, O., C. Intrieri and D. Domizio (1996). Impiego di metodi fillometrici per la caracterizzazione di alcuni vitigni dell'Emilia Romagna. Rev. Vitic Enol. nr. 1:17-26.

Stefano, A., N. Vignozzi and M. A. Vignini (1996). AmpeloCADs (Ampelographic Computer-Aided Digitizing System): an integrated system to digitize, file and process biometrical data from Vitis ssp. leaves. American Journal Enol. and Vitic. 47(3):257-267.