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

Liliana ROTARU 1)

1) University of Agricultural Sciences and Veterinary Medicine Iași, 3 M. Sadoveanu Al., Iași 700490, Romania; lrotaru@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 offers the possibility to study genotypes through their phenotypic expressions manifested in vine leaves. 15 autochthonous vine breeds were part of this study. They belonged to the proles pontica and orientalis. With the help of cluster analysis it was possible to uncover 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 similarity 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
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, Tămăioasă românească, Busuioacă de Bohotin. To the proles orientalis, subproles antasiatica belongs only Coarnă 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 calculated (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
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, Tâmâioasă românească, Busuiocă 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 (Tâmâioasă românească, Busuiocă 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 because the value of the dissimilarity index is less than 8.9445 In this group with Tab. 1. Levels of breed chaining in the elaboration of dendrogram

<table>
<thead>
<tr>
<th>Knot making</th>
<th>No. of breeds in the knot</th>
<th>Value of chaining index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tâmâioasa românească - Busuiocă de Bohotin</td>
<td>2</td>
<td>1.3897</td>
</tr>
<tr>
<td>Galbena de Odobesti - Mustoasa de Măderat</td>
<td>2</td>
<td>1.5740</td>
</tr>
<tr>
<td>Feteasca regala - Grasa de Cotnari</td>
<td>2</td>
<td>2.1244</td>
</tr>
<tr>
<td>Frâncusa - Tâmâioasa românească - Busuiocă de Bohotin</td>
<td>3</td>
<td>2.4340</td>
</tr>
<tr>
<td>Galbena de Odobesti - Mustoasa de Măderat - Zghihară de Huşi</td>
<td>3</td>
<td>2.7144</td>
</tr>
<tr>
<td>Frâncusa - Tâmâioasa românească - Busuiocă de Bohotin - Babeasca neagra</td>
<td>4</td>
<td>2.9856</td>
</tr>
<tr>
<td>Coarna neagra - Feteasca regala - Grasa de Cotnari</td>
<td>3</td>
<td>4.0913</td>
</tr>
<tr>
<td>Feteasca alba - Feteasca neagra</td>
<td>2</td>
<td>5.7765</td>
</tr>
<tr>
<td>Coarna alba - Berbecel</td>
<td>2</td>
<td>5.9572</td>
</tr>
<tr>
<td>Galbena de Odobesti - Mustoasa de Măderat - Zghihară de Huşi - Barută neagra</td>
<td>4</td>
<td>6.6457</td>
</tr>
<tr>
<td>Coarna neagra - Feteasca regala - Grasa de Cotnari - Feteasca alba - Feteasca neagra</td>
<td>5</td>
<td>8.9445</td>
</tr>
<tr>
<td>Coarna alba - Berbecel - Frâncusa - Tâmâioasa românească - Busuiocă de Bohotin - Babeasca neagra</td>
<td>6</td>
<td>12.1520</td>
</tr>
<tr>
<td>Coarna alba - Berbecel - Frâncusa - Tâmâioasa românească - Busuiocă de Bohotin - Babeasca neagra</td>
<td>11</td>
<td>21.9544</td>
</tr>
<tr>
<td>Total inertia</td>
<td>15</td>
<td>34.1896</td>
</tr>
</tbody>
</table>

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-

![Breed dendrogram](image-url)
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 Fetecăscă neagră, Coarnă neagră, Fetecăscă albă, Fetecăscă regală and Grasă de Cotnari, with a medium stability, and a dissimilarity index of 8.9445;

- Group C with the breeds Galbenă de Odobești, Mustoasă de Măderat, Zghihară de Huși 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 Odobești, Zghihară de Huși and Bătută neagră have a common genetic origin. This is certified because these three breeds 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

This study is financed by the National Council for Higher Education Research (CNCSIS), the Executive Unit for the Funding of Higher Education and Academic Research (UEFISCU), Romania, project PCE-IDEI, code CNCSIS 1141.

References


