CLUSTER ANALYSIS IN PUBESCENT OAK TAXA FROM SERIES LANUGINOSAE: A CASE STUDY

C.M. ENESCU1  N. ŞOFLETEA1  A.L. CURTU1

Abstract: The main purpose of this study was to compare the different clustering techniques in order to identify the one with the best discriminating power among oak species. By using two groups of trees, one corresponding to pedunculate oak and the other one to pubescent oak, Ward’s Method with Manhattan distances provided the best separation between the two oak species. By contrast, different results were obtained for the two taxa from series Lanuginosae (pubescent oak and Italian oak), due mainly to their similarities in leaf and fruit traits. In conclusion, by using a wide variety of clustering combinations within STATISTICA software, no separation between pubescent oak and Italian oak was achieved.

Key words: Cluster analysis, Lanuginosae, pubescent oak, Italian oak.

1. Introduction

In general, the term cluster analysis refers to a broad series of techniques used for classification of individuals [8]. In particular, cluster analysis is a common statistic tool used in studies aimed to discriminate between closely related oak species (genus Quercus, Fagaceae) [1], [4], [12], [14], [15], or to group different populations of the same species according to their geographic origin [2], [7], [13]. Its goal is to sort different objects (i.e. trees) into homogenous groups. Therefore, on the basis of similarity, the trees within a group (cluster) will be related to one another and unrelated to the trees from other groups.

Nowadays, there are several clustering techniques provided by different statistical software packages and sometimes selecting the best one is complicated. In the case of the oaks (Quercus spp.), many amalgamation (linkage) rules and distances were used for cluster analysis. The most frequent linkage rules were: i) Complete Linkage [3], [10]; ii) Single Linkage or the Nearest Neighbour [11]; iii) Unweighted Pair Group Method with Arithmetic Mean [6], [24] and iv) Ward’s Method [19], [23], while the usual distances used were: i) the Euclidean [6], [7], [10], [15]; ii) the Square Euclidean [17]; iii) the Manhattan [1] and iv) the Mahalanobis [13] ones.

Given that the oaks constitute one of the most species-rich genera and the fact that the identification of species is often complicated by the overlapping morphological traits, the taxonomical status of some oak species is sometimes a subject of debate. If the taxonomical status of pubescent or downy oak (Quercus

1 Dept. of Silviculture, Transilvania University of Braşov.
According to different taxonomical classifications, Italian oak (Q. virgiliana Ten.) is sometimes considered as a distinct species [5], [20], sometimes it is not [18].

2. Objectives

Firstly, the objective of this study was to compare different clustering techniques in the case of two easily identified by morphological criteria oak species, namely pedunculate oak (*Quercus robur* L.) and pubescent oak (*Q. pubescens*). Secondly, the best clustering technique was used in the case of the two taxa from series *Lanuginosae*, in which morphological differentiation is difficult or questionable.

3. Materials and Methods

The methodology for evaluating the leaf morphology was also used in other similar studies [9], [16], [22]. So, five leaves per tree were chosen and 14 leaf descriptors were assessed: five dimensional characters - lamina length (LL), petiole length (PL), lobe width (LW), sinus width (SW), length of lamina at largest width (WP), two counted variables - number of lobes (NL) and number of intercalary veins (NV), two observed variables - abaxial laminar pubescence (PU), evaluated according to the grading system from 1 (no pubescence) to 6 (dense hairness) with an microscope (x30) and basal shape of the lamina (BS), scored as an index varying from 1 to 9 and five transformed variables - obversity (OB), petiole ratio (PR), lobe depth ratio (LDR), percentage venation (PV) and lobe width ratio (LWR). In addition, for the two taxa from series *Lanuginosae* the length of fruit peduncle (Lp) was also assessed.

In order to achieve the main objective of this study, the first data set for analysis was composed of fifty pedunculate oaks and fifty pubescent oaks. The reason for choosing these two oak species is the obvious difference regarding the leaf morphology. Particularly, compared to pubescent oak, pedunculate oak has a longer lamina, but a shorter petiole, an ear-like basal shape of the lamina and no hairs on the abaxial part of the leaf [21]. So, the proper cluster technique should be able to provide a good separation of the two oaks.

Secondly, the best clustering techniques resulted from the first data set were used to separate the two taxa from series *Lanuginosae* (75 pubescent oak trees and 25 Italian oaks). The mean values of all descriptors, except the calculated ones are given in Table 1.

Cluster analyses were carried out with STATISTICA software v.8.0. Several clustering amalgamation rules (Complete Linkage, Single Linkage, Unweighted pair-group average, Weighted pair-group average and Ward’s Method) and all available distances within the software were applied to the data. In total, 35 combinations (five linkage rules multiplied with seven distances) were made.

<table>
<thead>
<tr>
<th>Variable</th>
<th>1st set</th>
<th>2nd set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q. robur</td>
<td>Q. pubescens</td>
<td>Q. virgiliana</td>
</tr>
<tr>
<td>PU</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>BS</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>NL</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>NV</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>LL [mm]</td>
<td>114</td>
<td>81</td>
</tr>
<tr>
<td>PL [mm]</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>LW [mm]</td>
<td>34</td>
<td>29</td>
</tr>
<tr>
<td>SW [mm]</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>WP [mm]</td>
<td>71</td>
<td>44</td>
</tr>
<tr>
<td>Lp [mm]</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Mean values for descriptors

<table>
<thead>
<tr>
<th>Variable</th>
<th>1st set</th>
<th>2nd set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q. robur</td>
<td>Q. pubescens</td>
<td>Q. virgiliana</td>
</tr>
<tr>
<td>PU</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>BS</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>NL</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>NV</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>LL [mm]</td>
<td>114</td>
<td>81</td>
</tr>
<tr>
<td>PL [mm]</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>LW [mm]</td>
<td>34</td>
<td>29</td>
</tr>
<tr>
<td>SW [mm]</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>WP [mm]</td>
<td>71</td>
<td>44</td>
</tr>
<tr>
<td>Lp [mm]</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
4. Results and Discussions

As revealed by the tree diagram (Figure 1), Ward’s Method with Manhattan (City-Block) distances provided the best separation between Quercus pubescens and Q. robur. It can be seen that the cluster analysis delimited two distinct groups which correspond to the two oak species.

Similarly, two homogenous groups were also observed in the case of other related oak species, namely Q. robur and Q. petraea [15], or Q. virginiana and Q. geminata [4]. Moreover, good results were also obtained for the following combinations: Ward’s Method with Euclidean distances, Weighted pair-group Average with Manhattan distances or Unweighted pair-group average with percent disagreement.

On the contrary, bad results were obtained when Single Linkage was used as amalgamation rule. In this case, the cluster analysis produced more than two groups and no species assignment could be observed (Figure 2).

---

**Fig. 1. Dendrogram of the first data set (Ward’s Method).**  
Abbreviations: P - Q. pubescens, R - Q. robur

**Fig. 2. Dendrogram of the first data set (Single Linkage).**  
Abbreviations: P - Q. pubescens, R - Q. robur
Based on these results, the best combination of cluster analysis was applied to the second set of data. Interestingly, for this data set, two mixed groups were observed (Figure 3).

These findings support the hypothesis that the two taxa from series *Lanuginosae* cannot be recognized as different species [18], due mainly to their similarity in leaf and fruit traits [21]. This is in agreement with the results from Principal Component Analysis conducted for 931 pubescent oaks from Romania, where only one morphological group was observed [10].

Furthermore, regarding the same one hundred analysed trees from series *Lanuginosae* no significant differences were observed in allele frequencies at seven microsatellite loci (unpublished data).

**5. Conclusions**

By using all the clustering techniques provided by STATISTICA software no separation between the two taxa from series *Lanuginosae* was possible. This was mainly due to their similarities in leaf and fruit traits, the only descriptor which somehow separates them being the length of the peduncle (Lp) [10].

On the other hand, Ward’s Method with Manhattan distances proved to be the best choice for discriminating between pedunculate oak and pubescent oak.

Our findings from this statistical-morphological survey correlated with those from genetic investigations done for the same trees by using seven nuclear microsatellite loci (unpublished data) suggest that *Q. virgiliana* is rather an intra-specific unit of *Q. pubescens* than a separate species.

**Acknowledgements**

This work was supported by CNCS-UEFISCDI, project number PN II-RU-TE-73/2010. Cristian Mihai Enescu was supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU/88/1.5/S/59321.

**References**

1. Aas, G.: *Taxonomical Impact of Morphological Variation in Quercus*
17. Minotta, G., Degioanni, D.: Naturally Regenerated English oak (Quercus robur L.) Stands on Abandoned


