CHARACTERIZATION OF TAXONOMICALLY RELATED SOME TURKISH OAK (*QUERCUS* L.) SPECIES IN AN ISOLATED STAND: A MORPHOMETRIC ANALYSIS APPROACH

THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

CANER AKTAŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY

APRIL 2010

Approval of the thesis

CHARACTERIZATION OF TAXONOMICALLY RELATED SOME TURKISH OAK (*QUERCUS* L.) SPECIES IN AN ISOLATED STAND: A MORPHOMETRIC ANALYSIS APPROACH

submitted by **CANER AKTAŞ** in partial fulfillment of the requirements for the degree of **Master of Science in Biology Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _____________________

Prof. Dr. Musa Doğan
Head of Department, **Biology** _____________________

Prof. Dr. Zeki Kaya
Supervisor, **Biology Dept., METU** _____________________

Assoc. Prof. Dr. Sertaç Önde
Co-Supervisor, **Biology Dept., METU** _____________________

Examining Committee Members:

Prof. Dr. Musa Doğan
Biology Dept., METU _____________________

Prof. Dr. Zeki Kaya
Biology Dept., METU _____________________

Prof. Dr. Hayri Duman
Biology Dept., Gazi University _____________________

Assoc. Prof. Dr. Ayşegül Çetin Gözen
Biology Dept., METU _____________________

Assoc. Prof. Dr. İrfan Kandemir
Biology Dept., Ankara University _____________________

Date: 29. 04. 2010

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Caner Aktaş

Signature:

ABSTRACT

CHARACTERIZATION OF TAXONOMICALLY RELATED SOME TURKISH OAK (*QUERCUS* L.) SPECIES IN AN ISOLATED STAND: A MORPHOMETRIC ANALYSIS APPROACH

Aktas, Caner
M.S., Department of Biology
Supervisor: Prof. Dr. Zeki Kaya
Co-Supervisor: Assoc. Prof. Dr. Sertaç Önde

April 2010, 65 pages

The genus *Quercus* L. is represented with more than 400 species in the world and 18 of these species are found naturally in Turkey. Although its taxonomical, phytogeographical and dendrological importance, the genus *Quercu*s is still taxonomically one of the most problematical woody genus in Turkish flora. In this study, multivariate morphometric approach was used to analyze oak specimens collected from an isolated forest (Beynam Forest, Ankara) where *Quercus pubescens* Willd., *Q. infectoria* Olivier subsp. *boissieri* (Reuter) O. Schwarz and *Q. macranthera* Fisch. & C. A. Mey. ex Hohen. subsp. *syspirensis* (C.Koch) Menitsky taxa are belonging to section *Quercus* sensu stricto (s.s.) are found. Additional oak specimens were included in the analysis for comparison. Morphometric study was based on 52 leaf characters such as, distance, angle, and area as well as counted, descriptive and calculated variables. Morphometric variables were calculated automatically by use of landmark and outline data. Random forest classification method was used to select discriminating variables and predict unidentified specimens by use of pre-identified training group. The results of the random forest variable selection procedure and the principal component analysis (PCA) showed that the morphometric variables could distinguish the specimens of *Q. pubescens* and *Q. macranthera* subsp. *syspirensis* mostly based on the overall leaf size and number of intercalary veins while the specimens of *Q. infectoria* subsp.

boissieri were separated from others based on lobe and lamina base shape. Finally, micromorphological observations of abaxial lamina surface have been performed by scanning electron microscope (SEM) on selected specimens which were found useful to differentiate, particularly the specimens of *Q. macranthera* subsp. *syspirensis* and its putative hybrids from other taxa.

Keywords: *Quercus pubescens, Q. infectoria* subsp. *boissieri, Q. macranthera* subsp. *syspirensis,* hybrids, traditional morphometrics, principal component analysis, random forest classification, micromorphology, trichomes, abaxial lamina waxes.

ÖZ

İZOLE BİR MEŞCEREDE BULUNAN VE TAKSONOMİK OLARAK YAKIN OLAN BAZI TÜRK MEŞE TÜRLERİNİN (*QUERCUS* L.) MORFOMETRİK ANALİZ YÖNTEMLERİYLE KARAKTERİZASYONU

Aktas, Caner
Yüksek Lisans, Biyoloji Bölümü
Tez Yöneticisi: Prof. Dr. Zeki Kaya
Ortak Tez Yöneticisi: Doç.Dr. Sertaç Önde

Nisan 2010, 65 sayfa

Quercus L. cinsi dünya çapında 400'den fazla türle temsil edilmektedir ve bu türlerden 18'i doğal olarak Türkiye'de bulunmaktadır. Sahip olduğu taksonomik, fitocoğrafik ve dendrolojik öneme rağmen, *Quercus* cinsi, Türkiye Florası'nda taksonomik olarak hala en sorunlu odunlu bitki cinslerinden birini oluşturmaktadır. Bu çalışmada, *Quercus* sensu stricto (s.s.) seksiyonuna ait *Quercus pubescens* Willd., *Q. infectoria* Olivier subsp. *boissieri* (Reuter) O. Schwarz ve *Q. macranthera* Fisch. & C.A.Mey. ex Hohen. subsp. *syspirensis* (C.Koch) Menitsky taksonlarının bulunduğu izole bir ormandan (Beynam Ormanları, Ankara) toplanan meşe örnekleri, çokdeğişkenli morfometrik yöntemler kullanılarak incelenmiştir. Karşılaştırma yapmak için ek meşe örnekleri incelemeye dahil edilmiştir. Morfometrik çalışmada uzaklık, açı, alan ve ayrıca sayılabilen, betimleyici ve hesaplanan 52 yaprak değişkeni kullanılmıştır. Morfometrik değişkenler, "landmark" ve "outline" verileri kullanılarak otomatik olarak hesaplanmıştır. Ayırt edici değişkenlerin bulunması ve tanısı yapılamamış örneklerin ait oldukları taksonların belirlenmesi için, öntanısı yapılmış örneklerden oluşan öğrenme kümesi üzerinde "random forest" sınıflandırma yöntemi uygulanmıştır. "Random forest" değişken seçim yöntemi ve temel bileşenler analizi (PCA) sonuçları göstermektedir ki, morfometrik değişkenler, *Q. pubescens* ve *Q. macranthera* subsp. *syspirensis* bireylerini genel olarak yaprak boyutu ve interkalar damarların sayısı ile, *Q. infectoria* subsp. *boissieri* bireylerini diğerlerinden lob ve yaprak ayası taban şekli ile ayırmaktadır. Son olarak, seçilmiş bireylerin yaprak ayası alt yüzeylerinde taramalı elektron mikroskobuyla (SEM) yapılan incelemeler sonucunda, mikromorfolojik özelliklerin, özellikle *Q. macranthera* subsp. *syspirensis* bireylerini ve melezlerini diğer taksonlardan ayırmada kullanışlı olduğu bulunmuştur.

Anahtar Kelimeler: *Quercus pubescens, Q. macranthera* subsp. *syspirensis, Q. infectoria* subsp. *boissieri,* geleneksel morfometri, temel bileşenler analizi, random forest sınıflandırma yöntemi, mikromorfoloji, tüyler, yaprak ayası alt yüzü mumsu tabakası

ACKNOWLEDGMENTS

I am deeply thankful to my supervisor Prof. Dr. Zeki Kaya for his guidance, suggestions and great patience till the end of this study.

I would like to express my thanks to my co-supervisor Assoc. Prof. Dr. Sertaç Önde and all the jury members for their criticism and comments which gave me a lot of motivations also for future researches.

I want to specially thank to Prof. Dr. Hayri Duman for supporting the scanning electron microscope investigations in this study.

I would like to thank to all my lab mates, especially Kubilay Yıldırım and Ali Murat Gülsoy for their help in field trips.

I want to mention about my special thanks to my friends Yasin İlemin, Okan Ürker and Selçuk Köylü for their help in field trips and sharing their experience.

Finally, I would like to thank to my family, mom and dad; special thanks go to Tuba and my close friends for their love, support and great tolerance over the study period.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

µm	Micrometer
cpDNA	Chloroplast DNA
csv	Microsoft Excel Comma Separated Values File
LDA	Linear discriminant analysis
oob	Out-of-bag
PCA	Principal component analysis
PC	Principal component
PC1	First principal component
PC2	Second principal component
SEM	Scanning electron microscope
a	Alacaatli
b	Beynam Forest
c	Cubuk-Karagol
e	Elmadag
k	Kemer
p	*Q. pubescens*
i	*Q. infectoria* subsp. *boissieri*
m	*Q. macranthera* subsp. *syspirensis*
pi	*Q. pubescens* x *Q. infectoria* subsp. *boissieri*
mi	*Q. macranthera* subsp. *syspirensis* x *Q. infectoria* subsp. *boissieri*
mp	*Q. macranthera* subsp. *syspirensis* x *Q. pubescens*
nid	Not identified
lmi	initial landmark

CHAPTER 1

INTRODUCTION

1.1 Literature review over genus *Quercus* L. and the field of morphometrics

The genus *Quercus* L. comprises more than 400 species distributed through the temperate and subtropical zone of the Northern Hemisphere. The distribution area ranges from the montane tropics to latitude 60^{o}N in the Baltic region (Axelrod, 1983). Members of oak genus have been adapted various kinds of ecological and biogeographical conditions in different parts of its distribution zone (Nixon, 1993; 2006). Oak species have many ecological and economical values both for wildlife and humans. Oaks are providing living space for other organisms such as mycorrhizal fungus, other beneficial microorganisms (Landis *et al*., 2004) and insects (Tovar-Sánchez and Oyama, 2006). Acorns and leaves are important food source for insects, birds and mammals (Goodrum *et al*., 1971). Members of the genus are important for protecting soils against erosion, for their valuable wood for furniture, manufacture and fuelwood, and the oaks are used in other purposes such as cork and tannin production (Yaltırık, 1984).

The genus consists of 18 species of deciduous and evergreen trees in Turkey. Today, including the sub-species and varieties, totally 24 taxa have been identified in the latest revision of the genus *Quercus* and the recent publications (Hedge and Yaltirik, 1982; Zielioski *et al*., 2006). The distribution of the genus is varying from the temperate rain forest in the East Black Sea coasts to the Mediterranean Region, and the upper forest zones on the mountains to the dry woodlands in the steppes (Yaltirik, 1984) where the Beynam Forest also lies. *Q. macranthera* subsp. *syspirensis* is an endemic sub-species in Turkey. It is distributed mainly in the mountainous areas of inner-northern Anatolia. In the western part of its distribution zone, particularly in the dry woodlands, it is found with *Q. pubescens* and rarely with *Q. infectoria* subsp. *boissieri*. *Q. pubescens* and *Q. infectoria* subsp. *boissieri* are typical sub-Mediterranean species. They are often found in the coastal mountains of the western half of Turkey and inner Anatolia. *Q. infectoria* subsp. *boissieri* is also distributed in the south east Anatolia and the Mediterranean coastal region where the species exhibit a semi-evergreen habit. On the other hand, *Q. pubescens* is found rarely in maquis.

Turkish oaks have been classified at sectional level according to their leaf habit, wood type, maturation period of acorns, cupule scales, leaves and bark features. Three sections have been recognized: White oaks (section *Quercus*), Red oaks (section *Cerris*), and Evergreen oaks (section *Ilex*) (Hedge and Yaltirik, 1982). This classification is consistent with the most current phylogenetic reconstruction of the genus at sectional level based on morphological and molecular data (Manos *et al*., 1999; Bellarosa *et al*., 2005). Oaks are, taxonomically most problematical woody genera in Turkish flora. In the earlier studies, definition of the species was taken narrowly so that unnecessarily many local variants were recognized as species. Hedge and Yaltırık (1982) simplified the taxonomic scheme by eliminating those “ill-defined” species. Although their efforts has been great and meritorious; the taxonomical, biogeographical and phylogenetic position of oaks in Turkey are still far from being completely understood. There have been a limited number of studies conducted on oaks in comparing with Europe and North America, and there is little known about the species dispersal, hybridization, morphological and genetic variation within and between populations in Turkey.

Oaks are easy-to-study as a genus because they are in the form of trees and widely distributed. Characteristically, oak species show apparent intraspecific-morphological variations which may be due to introgressive-hybridization between species, genetically determined phenotypic plasticity or environmental variations (Palmer, 1948; Burger, 1975; Rushton, 1993; Bacilieri *et al*., 1996; Howard *et al*., 1997). The morphological variations observed in several taxonomic entities within oak genus can lead to identification problems (Burger, 1975). Moreover, high amount of allele sharing among closely related oak species are well reported in the literature (Whittemore and Schaal, 1991; Howard *et al*., 1997). Allele sharing was explained by the recurrent gene flow or shared ancestral polymorphism between species (Muir and Schlötterer, 2005; Lexer *et al*., 2006). In addition to the classical taxonomical studies on genus *Quecus*, these patterns make the oaks an important model organism for the studies of interspecific variations, hybridization and speciation (Burger, 1975).

Species differentiation for decades has been of general interest of the systematists. Morphological data has been and still form the basis of the plant taxonomy since morphological features have the advantage of being easily screened by various techniques. Thus, it used to constitute a rapid method needed to discriminate between closely related species such as oaks where they coexist in a sympatric population (Kremer *et al*., 2002).

Although virtually all morphological features used in taxonomy have genetic backgrounds, low genetic differentiation among closely related oak species is limiting the use of solely molecular data for the species identification/differentiation. Even though molecular markers would have enough resolution and could be used for delimitation of species in the future; morphological descriptions are still needed in systematics as a visual interface for displaying the variations of biological forms. The groups would be defined morphologically at first, and then these groups are used in further studies, such as molecular systematics (Jensen, 2003).

The classification science, taxonomy is traditionally based on the qualitative descriptions of morphological data. The analysis of morphology has turned into a quantitative science during the last century (Adams *et al*., 2004). The number of quantitative characters included in traditional studies of plant taxonomy is still limited, and in many cases such as in identification keys and descriptions, the quantitative data have been summarized by use of descriptive statistics. The development of multivariate methods allows investigating multiple morphological traits which is providing more objective and repeatable methods for comparing taxonomic groups (Rohlf and Bookstein, 1990). Combination of the multivariate techniques and the aids to make taxonomy a quantitative science, the concept of numerical taxonomy has been arisen (Sneath and Sokal, 1973). Morphometrics, in a broader concept, is a field of study dealing with biological forms (shape and size) in a quantitative manner. Traditional school of morphometrics uses sequentially collected size and shape variables such as distances and angles. Multivariate statistics is the most commonly used methods so the term multivariate morphometrics has been used interchangeably with the traditional morphometrics (Adams *et al.*, 2004; Henderson, 2006). Commonly employed multivariate techniques are exploratory analyses with no *a priori* knowledge of taxonomic groupings such as ordination techniques or cluster analysis; and confirmatory analyses which need *a priori* knowledge of taxonomic groupings such as linear discriminant analysis (Henderson, 2006).

Ordination methods which are common in the taxonomic studies are used to summarize the multidimensional variable space in a few dimensions in a various ways (Henderson, 2006, Cron *et al.*, 2007). Principle component analysis (PCA) is one of the well-known ordination techniques. The main purpose in the PCA is to find a set of orthogonal eigenvectors of the multivariate data swarm. Each eigenvector (= principle component) are the linear combinations of the original variables that accounts successively for the maximum variance available in the multivariate data set. The resulting eigenvalues for each axis are the respective eigenvector lengths that denote the amount of variance represented by each eigenvector. The calculation of the PCA can be done by singular value decomposition (SVD) on the pairwise covariance or correlation matrix of the multivariate data (R Development Core Team, 2008). Although the numbers of principle components are equal to the number of original variables, most of the variance is usually accounted by the first few components. Thus PCA is also a dimension reducing method (Henderson, 2006).

Clustering is an example of unsupervised classification method which seeks to determine how the data are organized. The main purpose is grouping a collection of individuals into subsets or “clusters” based on the degree of similarity or dissimilarity between them such that the members of each cluster are more closely related to one another than the members of other clusters (Hastie *et al.*, 2009).

As a confirmatory method, linear discriminant analysis (LDA) is used to find linear combinations of variables which discriminate between previously determined groups. LDA is also used for classification by using the previously identified individuals as a training group for predicting unidentified individuals. Although it is a well known statistical classifier and its theoretical background is well understood, the method makes the assumptions of multivariate normality, homogeneity of variance/covariance matrices across groups, linearity of the data and the absence of high multicollinearity. LDA can give misleading results depending on the degree of the violation of one or more of these assumptions (Worth and Cronin, 2003). Furthermore, the results may be unstable under resampling or cross-validation (Feldesman, 2002).

A machine learning method, classification tree learning is a non-parametric alternative to LDA when the assumptions are not met (An, 2005; Kotsiantis, 2007). A classification tree is a decision tree where the dependent variable is categorical. It is a tree structured model where each internal node involves only one splitting predictor variable and each leaf (terminal) node is associated with a unique state for dependent variable (Figure 1.1) (Gehrke, 2005).

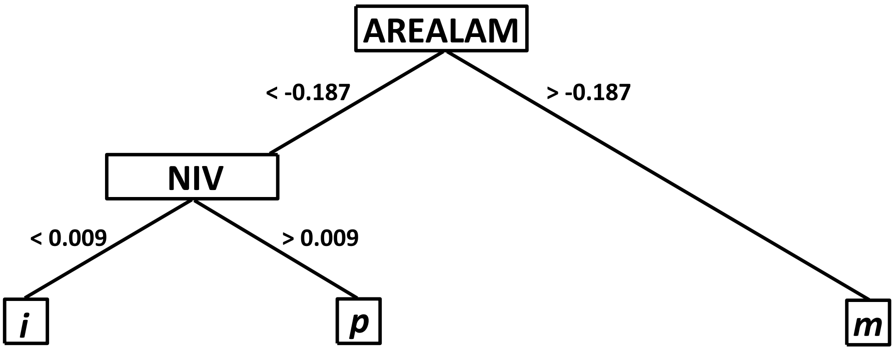


Figure 1.1 A single example classification tree. AREALAM (leaf area) and NIV (number of intercalary veins) variables were used as splitting predictor variables, and each terminal node is associated with i (*Q. infectoria* subsp. *boissieri*), p (*Q. pubescens*) or m (*Q. macranthera* subsp. *syspirensis*) species respectively (See Results for further explanation).

Classification tree learning is a supervised learning technique where the training data set is used to produce a model classifier to predict the labels of unidentified individuals (An, 2005; Gehrke, 2005; Kotsiantis, 2007). Predicting accuracy of a single classification tree may be highly unstable against small perturbations in training data set (Breiman, 1998). On the other hand, the perturbation itself can be used to improve accuracy. Multiple trees may be constructed by repeatedly perturbing the training data set or the construction method. Aggregating these trees to obtain a final predictor improves the accuracy of prediction (Breiman, 1998). Bagging is a perturbing and combining method which generates bootstrap samples repeatedly within a training set to produce an ensemble of different structured classification trees; and the majority of the votes (e.g. labels of species) which comes from each tree are used to predict final classes (Breiman, 1998; Wu, 2003). Random Forest (Breiman, 2001) classifier combines the idea of bagging and random variable selection at each internodes of the tree to construct a forest of decision trees.

Random Forest is constructed basically as follows (Breiman and Cutler, 2004; Svetnik *et al.*, 2003; Kim and Kim, 2007):

Before the calculations, assume that the number of individuals in the training set is *S*, and during the bagging procedure the number of individuals taken from training set as a bootstrap sample is *B*. The total number of predictor variables used in the analysis is *M*, and the number of different predictor variables tried at each split of the tree is *m* (the value of *m* is determined before analysis and $m \leq M$). The number of trees in the forest is *n* which is also determined before running the algorithm.

1. Use the bootstrap sample to construct a single classification tree *T* to predict the classes of individuals in the training set that are not in the bootstrap sample. These predicted individuals are out-of-bag (oob) samples (with the sample size of *S*-*B*) and the step is called oob estimation. During the constructions of each node of *T*, *m* variables out of *M* are randomly chosen for each node; than only a variable m_k**,** which gives a best split at that node is used.
2. Measure the importance of predictor variables. This can be done by using two types of measurements:

a. Mean decrease in accuracy: After a tree grown, the oob individuals are put down to the tree and the number of correct classifications is counted. Then the values of variable m_k in the oob data are randomly permuted and the class memberships are calculated again. The number of the correct classes in the original oob data is subtracted from the number of the correct classes obtained by permuted m_k oob data. Mean decrease in accuracy is the average value of this over all the trees in the forest which gives a raw importance for variable m_k. The larger the value of the mean decrease accuracy, the more important the variable is.

b. Gini importance: Gini index is the default splitting criterion in the tree construction of the random forest and uses as a node impurity measure. Node impurity measurement provides information about the homogeneity of the terminal nodes which reflects the goodness-of-split. The Gini impurity criterion for the two child nodes is always less than the parent node (if that variable has a discriminating property between the groups) so that at each split on variable m_k, gini index decreases. The mean decrease in gini index for each variable over all trees is used as a variable importance measure in random forest. Again, the larger value indicates the importance of variable. Often it is very consistent with the accuracy measurement.

3. Repeat each steps *n* times for growing a random forest classifier.
4. Use the constructed random forest classifier to predict the unknown specimens (the test set).

Distance measurements (including perimeter and area) are usually highly correlated with the size of measured object. Size does not reflect the natural relationships between and within taxonomic groups always, where they could be highly dependent on the ecological and ontogenic factors. Although size variables inherently take place in taxonomical studies, it has to be removed from the analysis in order to displaying true morphological structure and the shape. Several methods of size removal have been proposed in traditional morphometrics:

1) In order to describe the shape, ratios of the distance measurements could be used instead of raw data. Ratios are widely used in plant taxonomy for addressing some simple shape features, e.g. length/wide ratio for describing the leaf shape (Metcalfe and Chalk, 1979). The use of ratios can be misleading. Also, they may be size related (Atchley *et al.* 1976).
2) A second and more usual method for removing the size is applying multivariate techniques like PCA to construct new shape variables from linear combinations of the morphometric variables. Usually the first principle component (PC1) accounts for the variation in size (Gage and Wilkin, 2008). The rest of the components are expected to be size-free and reflect the shape variation.

Although the general expectation is that the PCA is sufficient for separating size and shape variations, first component can also contain significant amounts of shape variation and subsequent components can also be size-related (Marcus, 1990; Parsons *et al*., 2003; Zelditch *et al*., 2004; Gage and Wilkin, 2008). There is no common agreement on which method is more preferable for size removal than the others (Parsons *et al*., 2003; Jamniczky and Russell, 2004). Another problem with the traditional morphometrics is that the oversimplification of shapes through the limited number of measurement points taken on the object which leads to the possibility of two different shapes can be found identical (Dickinson *et al*., 1987).

Although traditional morphometrics has several drawbacks that have limited its use in shape analysis, both shape and size components of variation are important in the taxonomy, and the traditional morphometrics is useful in identifying the patterns of variations within a set of taxonomic units (Jensen, 2003).

Micromorphological features such as trichomes and wax characteristics have proven their usefulness in the oak taxonomy (Hardin, 1976, 1979; Uzunova *et al*., 1997; Bussotti and Grossini, 1997; Scareli-Santos *et al*., 2007). For example, increasing complexity of the trichomes was considered to reflect the evolutionary trends (Hardin, 1976). On the other hand, abaxial lamina wax properties for the genus *Quercus* section *Quercus* and section *Cerris* were well separated where the former has waxes arranged in vertical scales, and the latter has smooth waxes (Bussotti and Grossini, 1997).

1.2. Significance of the study

Quercus is an extremely important genus in Turkey both ecologically and economically. Species richness for oaks is notably high in Turkey, but more importantly they are the main stand forming forest trees existing in the semi-arid regions of Turkey. These areas constitute a large part of the country and are mainly found in the Central Anatolia Plateau and southeastern Turkey. The mean annual rainfall is sufficient for tree growth in most parts of these semi-arid areas. Natural lower limit of the oak woodlands on steppes varied due to aridity which is related to altitude, slope exposure and the general topography of the area. Today, it is very difficult to distinguish such a border due to the anthropogenic activities that have continued since the ancient times. In the past, steppes expanded gradually with the destruction of forests, and now those steppe lands and remaining forests are both endangered by agriculture and urbanization. These practices made oak woodlands be scattered in small patches across the landscape. It is still possible to find well developed oak individuals or stands around the semi-arid natural lower forest borders, but they remained usually in more humid regions outside the Central Anatolia basin where the natural regeneration is easier. Oaks have been chosen as a national tree for many countries as they are the symbol of endurance and strength. Likewise, in The Victory Monument at the Ulus Square in Ankara, young shoots emerge from an old oak block has been chosen to portray the young Turkish Republic which emerged from the Ottoman Empire (Yaltırık, 1984).

Although, its importance, taxonomists mostly hesitated to study oaks because the existing difficulties. Since they have no striking flowers and the variations seem rampant in the genus for a traditional taxonomist, the genus has not been well understood and studied in Turkey.

1.3. Objectives of the study

There are three main objectives of this study.

1) Firstly, to differentiate closely related and possible hybridizing oak species by the use of multivariate morphometrics and micromorphological features where these species exist in an isolated oak forest.

2) Secondly, to examine the discriminating power of the morphometric variables and to introduce a relatively new method, random forest classification.

3) Finally, to make a contribution to the field of semi-automated taxon identification studies, by proposing an automated calculation method for most of the variables

CHAPTER 2

MATERIALS AND METHODS

2.1 Plant Material and Sampling Methods

During the fall of 2008, mature branches were collected from trees in the Beynam Forest. The Beynam Forest is located south of Ankara (N39^{o} 40' - E32^{o} 55', at 1200-1500m). This stand was selected because it is a natural and isolated stand of three possibly hybridizing oak species. The Beynam Forest is a protected remnant *Pinus nigra* subsp. *pallasiana* forest where the black pine is in one of its most extreme distributions for inner Anatolia. Although most of the oaks are old coppices, they are protected and forestry activities not conducted within the area since 1960's. *Quercus pubescens*, *Q. macranthera* subsp. *syspirensis* and *Q. infectoria* subsp. *boissieri* species which are belonging to section *Quercus* sensu stricto (s.s.) are naturally found within the area. Reference specimens were included additionally in the analysis for comparing the collected taxa from Beynam Forest (b). The number and sample locations for additional specimens are as follows: four from Cayyolu/Alacaatli (a), Ankara; one from Cubuk Karagol (c), Ankara; 10 from Mount Elmadag (e), Ankara and two from Kemer (k), Antalya (Figure 2.1).

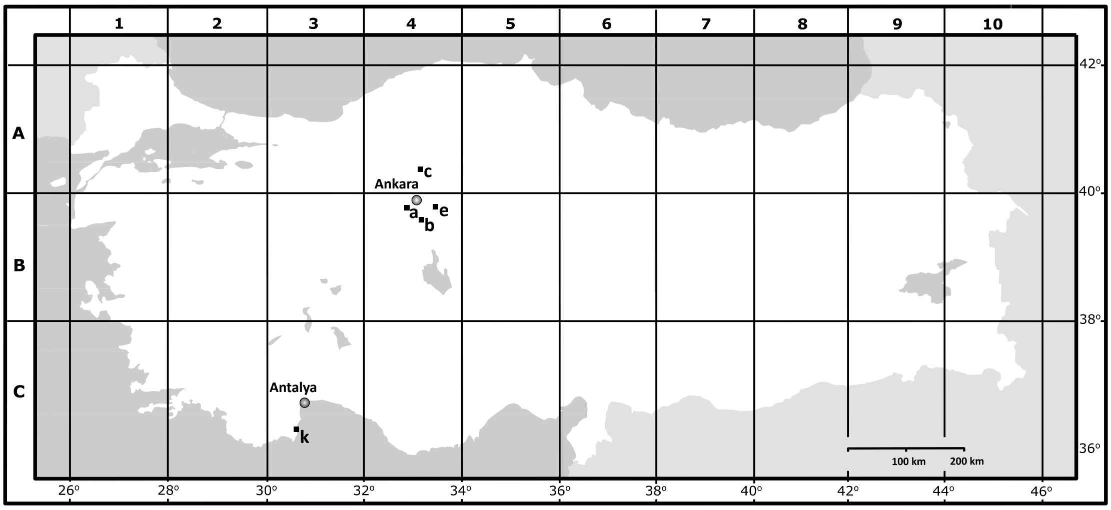


Figure 2.1 Distribution map showing the locations of studied oak populations. The grid system which was used in the Flora of Turkey (Hedge and Yaltirik, 1982). (b: Beynam; a:Alacaatli; c: Cubuk Karagol; e: Elmadag; k: Kemer)

A total of 139 trees were sampled. To avoid between-years variation in the leaf shape most of the samples were collected within the same year. Only the specimens from Kemer were collected in the following year. The time which was chosen for collecting leaves was after the leaf growth stopped in the autumn to prevent seasonal bias. Sampled trees located at least 40m apart along a transect and randomly sampled. Exceptions were only the specimens coded with an “x” affix in their operational taxonomic unit (OTU) name (e.g. bx6). Those individuals were collected less than 40m and they were selected *a priori* for any remarkable trait which was observed during the field trips. Five different branches within each tree were chosen randomly. Sun-exposed and wherever possible south-facing branches were harvested for uniform sampling in all specimens. As many as possible undamaged leaves were detached from the branches and pressed by standard herbarium techniques. Ten substantially intact leaves were selected randomly for the study. Initial identification of specimens was done according to Hedge and Yaltirik (1982). Forty-eight of those specimens were identified as a species without a doubt and were marked with the acronyms of its respective species name p (*Quercus pubescens*), i (*Q. infectoria* subsp. *boissieri)* or m (*Q. macranthera* subsp. *syspirensis*). Thirty-one of specimens were identified as a putative hybrid pi (*Q. pubescens* x *Q. infectoria* subsp. *boissieri*), mi (*Q. macranthera* subsp. *syspirensis* x *Q. infectoria* subsp. *boissieri*) or mp (*Q. macranthera* subsp. *syspirensis* x *Q. pubescens*). The rest, 60 of them do not represent ‘typical’ attributes of any taxa nor the clear hybrid origin so they were marked as nid (not identified) (Table 2.1). Identification results for each specimen were arranged in a csv (Microsoft Excel Comma Separated Values File, Microsoft Inc.) table.

Table 2.1 The result of the initial identifications.

Species / Group name	**Code**	**Number**
Q. pubescens	p	17
Q. infectoria subsp. *boissieri*	i	11
Q. macranthera subsp. *syspirensis*	m	20
Q. pubescens x *Q. infectoria* subsp. *boissieri*	pi	8
Q. macranthera subsp. *syspirensis* x *Q. infectoria* subsp. *boissieri*	mi	14
Q. macranthera subsp. *syspirensis* x *Q. pubescens*	mp	9
Not Identified	nid	60

2.2 Image Capturing and Obtaining Macromorphometric Data

2.2.1 Collecting Morphological Data

Dried leaves from each specimen were digitized in 600 dpi resolution, using a flatbed scanner (HP Scanjet 4370; Hewlett Packard, Palo Alto, CA, USA). Leaves were placed on a black background, abaxial side up and vertically on the scanner. Ten leaves were scanned in a single image at one time, if possible. But if the leaf sizes were too large for scanning all leaves at once, two plates with five leaves each were prepared for a specimen. Two sets of jpeg images were created from the original scans. One was used for locating landmarks on leaves (Figure 2.2), and the other was used for outline extraction (Figure 2.3). For the outline detection, binary images were produced by thresholding the images of the second set (Figure. 2.3). The tpsDig2 program (Rohlf, 2005) was used for locating landmarks manually and capturing outline by placing equally spaced semilandmarks automatically along leaf margins. Totally, 32 landmarks were chosen on each leaf for multivariate morphometrics analysis (Figure 2.4). Only four truly biologically homologous landmarks could have been detected on lamina (Landmarks 1, 2, 3 and 28, Figure 2.4) (also see Jensen, 2003). The rest are mathematical or geometrical landmarks (Landmark 4 to 27) which show no exact homological consistency between leaves, but used to locate lamina base lobes and six widest lobes. Landmark 29 and 30 have no exact position on a leaf at all where they were used only for representing the position of the midrib for automatic calculations. Similarly, landmark 31 and 32 were located on the curvature of the lamina base wheresoever's suitable for measuring the auricle angles, AURIL and AURIR. Two sets of text files were created which contain the XY-coordinate list of the landmarks and outline data respectively. For manually entering the counted and observed variables data, a csv table was used.

These files were then transferred to R program (R Development Core Team, 2008) for automated calculations of morphometric variables from coordinate data, achieving multivariate analysis and graphical presentations.

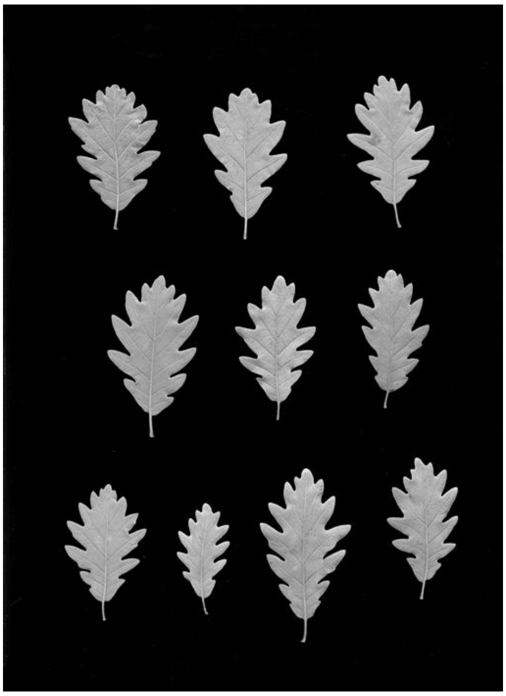


Figure 2.2 Scanned oak leaf-images, used for landmark registrations.

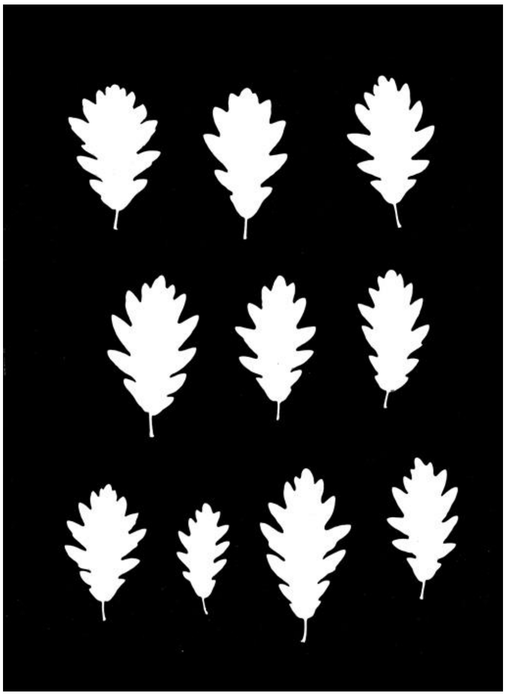


Figure 2.3 Threshold oak leaf-images for outline extraction.

Morphological analyses were carried out on previously selected ten leaves of each subsample. The characters used in analysis were chosen considering previous literature (Kremer *et al*., 2002; Ponton *et al*., 2004) and also new variables were identified for the study or new calculation methods were introduced in this study for some of the variables chosen from previous literature (Table 2.2). Characters in multivariate study are classified in six classes (Table 2.2): a) Distance variables, b) Angle variables, c) Area variables, d) Calculated (Ratio) variables, e) Counted variables and f) Descriptive variables. Counted (e) and descriptive (f) variables were gathered manually by the use of leaf images (e.g. Figure 2.2). R codes have been written for morphometric measurements (first four character classes) which were done automatically by the use of landmarks and outline coordinate data.

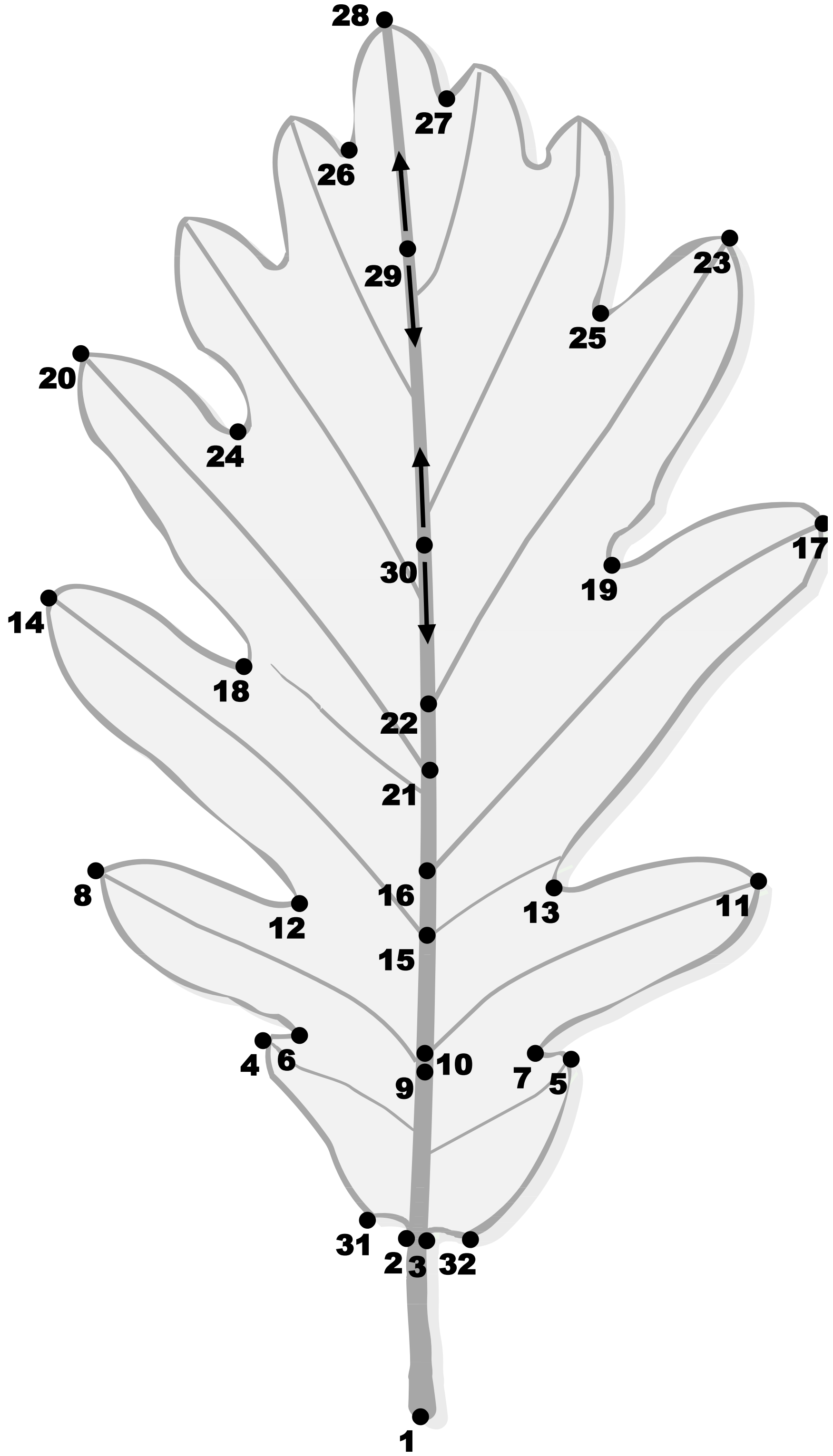


Figure 2.4 The locations of landmarks on oak leaves.

Table 2.2 Characters which were used in multivariate morphometric analysis.

#	Code	Character description
	DISTANCE VARIABLES	
1	PL	Petiole length$^{\mathbf{2}}$.
2	PW	Petiole width$^{\mathbf{2}}$.
3	LL	Lamina length$^{\mathbf{2}}$.
4	LW	Maximum lamina width$^{\mathbf{2}}$.
5	LLW	Length of the lamina to the largest width$^{\mathbf{1}}$.
6	LOBL	Mean of the length of the six largest lobes$^{\mathbf{1}}$.
7	LLBL	Length between lamina base and first lobes$^{\mathbf{2}}$.
8	LOBWD	Mean of the width of the six largest lobes between primary vein and lower sinus$^{\mathbf{2}}$.
9	LOBWU	Mean of the width of the six largest lobes between primary vein and upper sinus$^{\mathbf{2}}$.
10	SINVB	Mean length between sinus bases to six primary veins base$^{\mathbf{2}}$.
11	SW	Mean of the width of the sinus$^{\mathbf{1}}$.
12	APW	Wide of lamina apex$^{\mathbf{2}}$.
13	VLV	Variation of the distances between lobe veins$^{\mathbf{2}}$.
14	LBLV	Total distance on the midrib between six largest lobes$^{\mathbf{2}}$.
	ANGLE VARIABLES	
15	AURIL	Angle of the auricle at the left lamina base$^{\mathbf{2}}$.
16	AURIR	Angle of the auricle at the right lamina base$^{\mathbf{2}}$.

$^{\mathbf{1}}$: Variables which were chosen from previous literature without a change in their calculation methods (Kremer *et al*., 2002; Ponton *et al*., 2004). $^{\mathbf{2}}$: New variables which were identified in this study or variables which were chosen from previous literature with new calculation methods introduced in this study.

Table 2.2 *Continued*

#	Code	Character description
17	AVVU	The vein angles of the six largest lobes with the upper vein base or landmark 30 on the midrib[2].
18	ASIN	Mean of the four sinus angles in between the six largest lobes[1].
19	ALOB	Mean of the angles of the six largest lobes[1].
	AREA VARIABLES	
20	PERILAM	Lamina outlines perimeter[1].
21	AREALAM	Lamina surface area[2].
22	LOBTSL	Mean of triangular lobe area between two sinuses and lobe tip on six largest lobes[2].
23	LOBTSV	Mean of triangular lobe area between two sinuses and midrib on six largest lobes[2].
24	TAA	Triangular area of lamina apex[2].
	CALCULATED VARIABLES	
25	PRL	Relative length of the petiole[1].
26	PRW	Relative width of the petiole[2].
27	LRW	Relative width of the lamina[1].
28	LWRL	Relative length of lamina at largest width[1].
29	ARPE	Surface area to perimeter ratio[1].
30	ISOP	Isoperimetric deficit[1].
31	ELD	Elliptic deficit[1].
32	IVLOB	Number of intercalary veins per lobe[1].
33	PERINL	Perimeter to number of lobes ratio[1].
34	PERINLL	Perimeter to number of lobes ratio invariant of LL[2].

Table 2.2 *Continued*

#	Code	Character description
35	LUBLOB	Number of lobules per lobe[1].
36	LOBTAR	Lobe triangular area ratio[2].
37	LOBAAR	LOBTSL to lamina apex triangular area ratio[2].
38	PLWR	Petiole length width ratio[2].
39	LWS	Lobe width symmetry[2].
43	AURISIN	Sinus angle mean AURI ratio[2].
44	AURIVVU	Mean AURI to AVVU ratio[2].
45	LLLBLV	Lamina length LBLV ratio[2].
46	LLBLLR	LLBL to lamina length ratio[2].
	COUNTED VARIABLES	
47	NLOB	Number of lobes[1].
48	NLUB	Number of lobules[1].
49	NIV	Number of intercalary veins[1].
	DESCRIPTIVE VARIABLES	
50	BSL	Basal shape of lamina[1].
51	LOBTPS	Shape of lobe tips[2].
52	TEETH	Presence or absence of teeth on the tips of the lobes[1].

2.2.2. Automatic Calculation Procedure for Morphometric Variables.

Raw coordinate data and csv file (created for counted and descriptive variables) were imported to R. Some variables were measured directly with the use of the coordinate data. These measurements were simple interlandmark distances, angles or areas between coordinate points. R functions have been written to perform these principle measurements and they were used as basic internal functions for other algorithms as well. Interlandmark distance (*d*) was calculated by Euclidean distance formula for two dimensions which were shown in Figure 2.5 (Bookstein, 1991). Variables LL, LOBL, SINVB and APW are simple distances or mean distances between related landmarks (Figure 2.5, 2.6 and 2.7). Angles were calculated between three respective landmarks by applying the cosine formula on interlandmark distances and the formula was given in the Figure 2.6 (Bookstein, 1991). AVVU, ASIN and ALOB are mean angles between their respective landmarks (Figure 2.8). Due to the arching of petioles, 10th semilandmarks on petiole outline for both sides were used instead of landmark-1 for minimizing the error in the calculation of the lamina base angles. AURIL is the angle between 10th semilandmark (left) – landmark 2 – landmark 31, and AURIR is the angle between 10th semilandmark (right) – landmark 3 – landmark 32 (Figure 2.8). Cumulative chordal distance around outline and coordinates of landmarks were also used directly for calculating the area and perimeter variables. Determinant method was used for measuring the polygonal area demarcated by landmarks or semilandmarks which were represented in Figure 2.9 (Bookstein, 1991). LOBTSV and LOBTSL are the mean areas between respective landmarks, and TAA is the area between those three lamina apex landmarks (Figure 2.9). AREALAM is the enclosed lamina area where the semilandmarks form a high-dimensional polygon around lamina outline (Figure 2.9). Euclidean distance formula was applied cumulatively to the semi-landmarks for calculating the chordal distance for variables PERILAM and PL (Figure 2.9 and 2.10). Since petioles are usually arching, for the PL, using the chordal distance instead of the simple interlanmark distance measurement between landmark-1 and the lamina base is more accurate (Figure 2.10).

Other type of variables which were derived from landmark data have been obtained by use of series of mathematical procedures applied in R. VLV was used to measure the degrees of regularity in venation of lobes (Figure 2.7). It is invariant to size and a lower value indicates a more regular distribution of lobe veins on leaves. Upper and lower lobe widths (LOBWU and LOBWD) were calculated by the use of one landmark (e.g. landmark 19 in the Figure 2.6) and a line (e.g. lines between landmarks 22-23 for LOBWU and 16-17 for LOBWD, in the Figure 2.6, which are also equal to lobe lengths for the fourth and fifth lobes).

Finally, some variables such as PW, LW and LLW, which could not be achieved by using only previously defined landmarks, were measured automatically in R by the use of landmarks and outline data together.

A method for calculating the petiole width (PW) was summarized in the Figure 2.11. An initial landmark point and first 20 semi-landmarks (set A) on the petiole were used in the calculation. The criterion to choose the initial landmark (lmi) was the lmi's y-coordinate value. In the most oak leaves, lamina base is not symmetric so the distance between landmark-2 and landmark-3 ($d_{2\text{-}3}$) can not be considered as the petiole width. Here the main objective was to find the orthogonal distance between lower landmark (2 or 3) and the outline segment in the opposite side (set A). The landmark with a lower y-value was chosen as initial landmark. To draw a line which best represents the cloud of points in the set A and oriented in the same direction of the beginning of petiole, first eigenvector of the covariance matrix of A was used. PW is the orthogonal distance between lmi and the first eigenvector of A (Figure 2.11).

The measurement for the widest part of the lamina (LW) must be orthogonal to the midrib so the position and the direction of the midrib should be determined. Since many oak leaves do not have perfect symmetry and midrib deviates from a straight line, using a single line (e.g. the line between lamina base and apex points) would be inaccurate to represent the midrib. Geometric position of the midrib was defined in seven points by the use of 11 midrib landmarks in this study. In order to reduce calculations and for getting reliable results especially when the landmarks 9-10,15-16, 21-22 were overlapping with each other, arithmetic averages were taken for landmark pairs 2-3,9-10,15-16 and 21-22, respectively (Figure 2.5). Coordinate data of landmark 28, 29 and 30 were used directly. These points divide the midrib in six lines and so the lamina in six parts. At each part, local maximum lamina width was calculated. This was done by rotating all the landmarks and semi-landmarks coordinate data until the midrib line would be parallel to the y-axis for each part (e.g. for the first local lamina width, all coordinate data for that leaf were rotated in a way that the line between A and B would be parallel to the y-axis in the Figure 2.12.) After the rotation, semi-landmarks which were falling in-between the midrib line were chosen (e.g. the semi-landmark points which get higher y-values than the point A and lower than the point B). Then, XY-coordinate matrix of selected semilandmark data was sorted by their y-values. Adjacent rows of x-values in this matrix were subtracted from each other. The maximum value obtained by these subtractions gives the local maximum lamina width for that part (e.g. in the Figure 2.12, X1 is a line between the point pair, which has the greatest possible x-value difference where the y-value difference between these points ensured to be lowest). Calculation was repeated six times and largest local maximum lamina width was taken as lamina width (LW) for that leaf. LLW is the distance between LW line and the base of lamina (point A in the Figure 2.12).

In the following figures (Figure 2.5 – Figure 2.12), calculation procedures, formulation of variables and important basic equations were provided. The presented landmarks on each figure are the only ones which were used for calculating those variables given in that figure, but only one representative example was given for each variable, instead of repeating the illustration. For example, variable LOBTSV is the mean of the areas between landmarks: 6-9-12, 12-15-18, 18-21-24, 19-22-25, 13-16-19, 7-10-13; but only the calculation for the sixth lobe was represented in the Figure 2.9. Similarly variable ASIN, which is the mean value of four sinus angles between six largest lobes, was represented only for the first sinus (Figure 2.8).

$\boldsymbol{d_{i-k}}$: **Euclidean distance between landmark*s j and k* , where;**

$$d_{j-k} = \sqrt{(xj - xk)^2 + (yj - yk)^2}$$

$\mathbf{A = middle\ point\ on\ the\ d_{2-3}}$

$\mathbf{B = middle\ point\ on\ the\ d_{9-10}}$

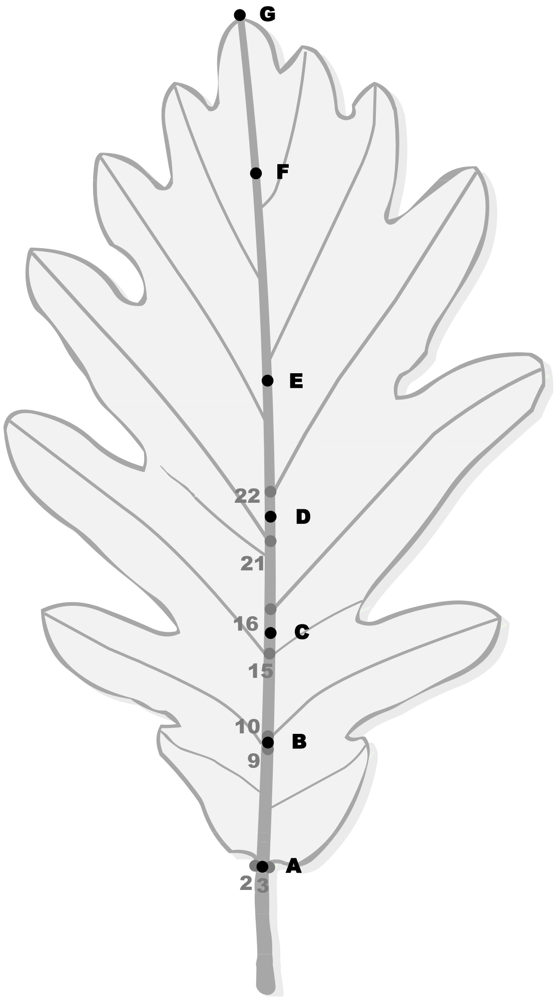
$\mathbf{C = middle\ point\ on\ the\ d_{15-16}}$

$\mathbf{D = middle\ point\ on\ the\ d_{21-22}}$

$\mathbf{E = landmark\ 30}$

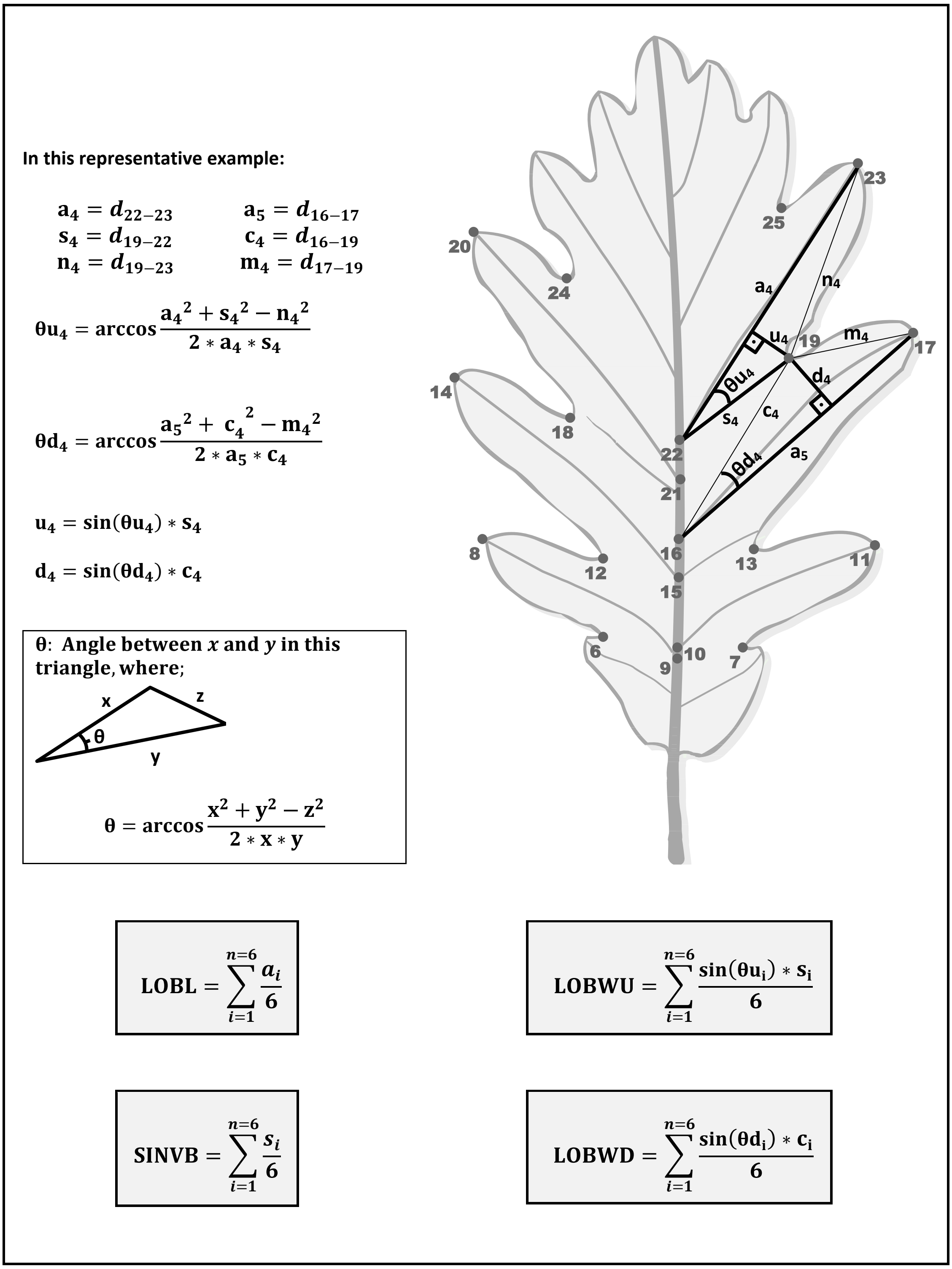
$\mathbf{F = landmark\ 29}$

$\mathbf{G = landmark\ 28}$



$$\mathbf{LL} = \boldsymbol{d_{A-B} + d_{B-C} + d_{C-D} + d_{D-E} + d_{E-F} + d_{F-G}}$$

Figure 2.5 Formulation of interlandmark distance and variable LL.



In this representative example:

$$\mathbf{a_4} = d_{22-23} \qquad \mathbf{a_5} = d_{16-17}$$

$$\mathbf{s_4} = d_{19-22} \qquad \mathbf{c_4} = d_{16-19}$$

$$\mathbf{n_4} = d_{19-23} \qquad \mathbf{m_4} = d_{17-19}$$

$$\theta u_4 = \arccos \frac{a_4{}^2 + s_4{}^2 - n_4{}^2}{2 * a_4 * s_4}$$

$$\theta d_4 = \arccos \frac{a_5{}^2 + \ c_4{}^2 - m_4{}^2}{2 * a_5 * c_4}$$

$$u_4 = \sin(\theta u_4) * s_4$$

$$d_4 = \sin(\theta d_4) * c_4$$

θ: Angle between *x* and *y* in this triangle, where;

$$\theta = \arccos \frac{x^2 + y^2 - z^2}{2 * x * y}$$

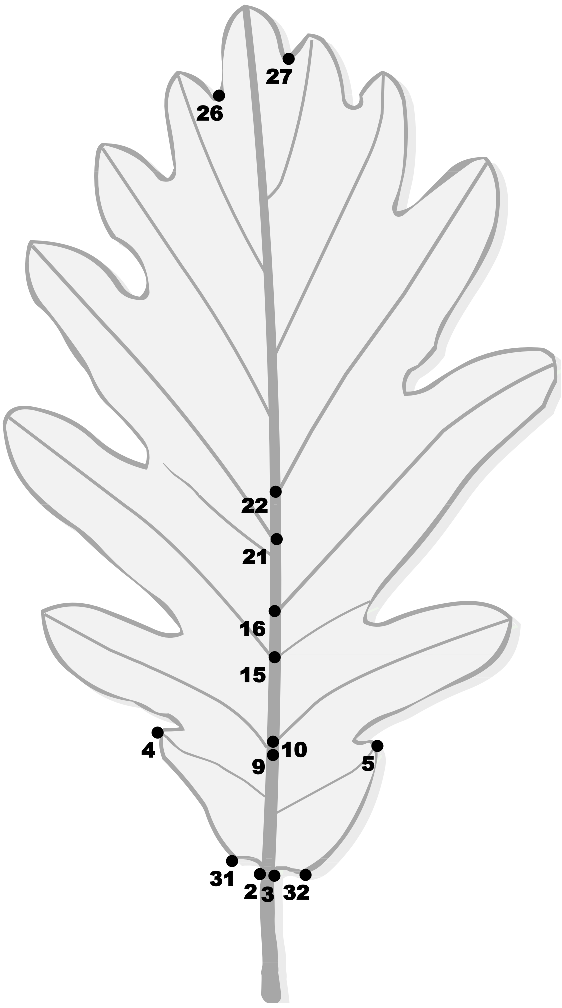
$$\mathrm{LOBL} = \sum_{i=1}^{n=6} \frac{a_i}{6}$$

$$\mathrm{LOBWU} = \sum_{i=1}^{n=6} \frac{\sin(\theta u_i) * s_i}{6}$$

$$\mathrm{SINVB} = \sum_{i=1}^{n=6} \frac{s_i}{6}$$

$$\mathrm{LOBWD} = \sum_{i=1}^{n=6} \frac{\sin(\theta d_i) * c_i}{6}$$

Figure 2.6 The cosine formula and formulations of the variables LOBL, SINVB, LOBWU, LOBWD and representative example for the measuring procedure.



$$\mu = \frac{\mathbf{LBLV}}{\mathbf{4}}$$

$$\boldsymbol{X} = [\,\boldsymbol{d}_{9-15}\ \boldsymbol{d}_{15-21}\ \boldsymbol{d}_{10-16}\ \boldsymbol{d}_{16-22}]$$

$$\sigma^2 = \frac{\sum(\boldsymbol{X} - \mu)^2}{\mathbf{4}}$$

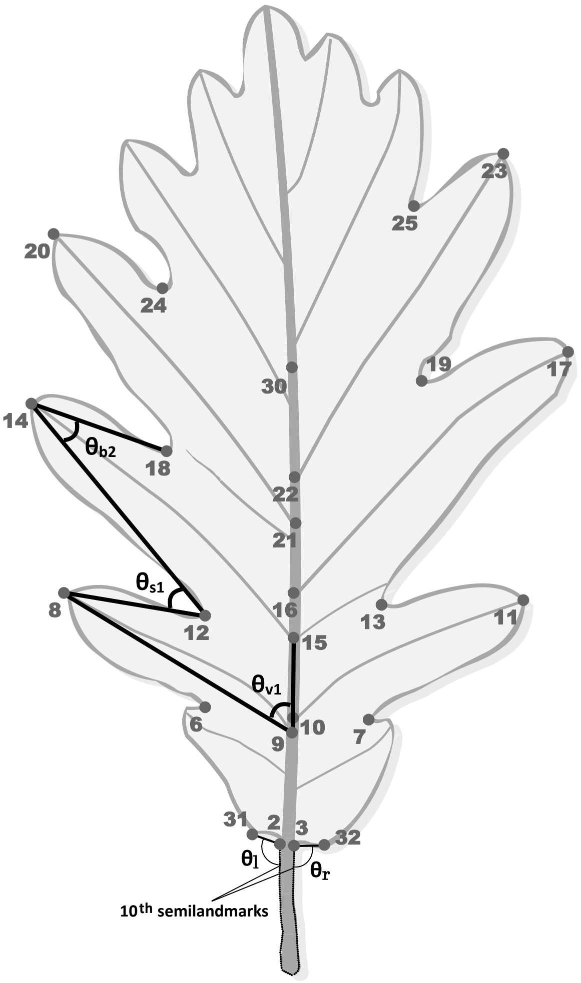
$$\mathbf{APW} = \boldsymbol{d}_{26-27}$$

$$\mathbf{LBLV} = \boldsymbol{d}_{9-15} + \boldsymbol{d}_{15-21} + \boldsymbol{d}_{10-16} + \boldsymbol{d}_{16-22}$$

$$\mathbf{LLBL} = \frac{\boldsymbol{d}_{2-4} + \boldsymbol{d}_{3-5}}{\mathbf{2}}$$

$$\mathbf{VLV} = \frac{\mathbf{100} * \sigma^2}{\mu^2}$$

Figure 2.7 Formulations of the variables APW, LBLV, LLBL, and VLV.



In this representative example:

$$\boldsymbol{\theta_{v1} = \angle \;\; landmarks\,(8-9-15)}$$

$$\boldsymbol{\theta_{s1} = \angle \;\; landmarks\,(8-12-14)}$$

$$\boldsymbol{\theta_{b2} = \angle \;\; landmarks\,(12-14-18)}$$

$$\mathbf{AVVU} = \sum_{i=1}^{n=6} \frac{\theta_v i}{6} * \frac{180}{\pi}$$

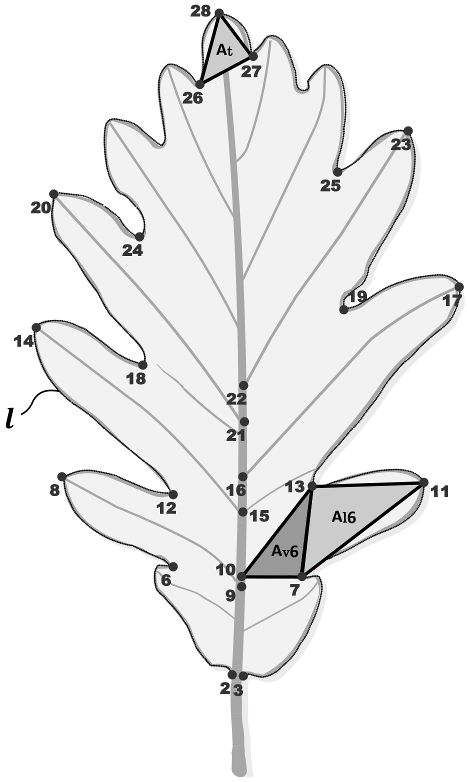
$$\mathbf{ASIN} = \sum_{i=1}^{n=4} \frac{\theta_s i}{4} * \frac{180}{\pi}$$

$$\mathbf{ALOB} = \sum_{i=1}^{n=6} \frac{\theta_b i}{6} * \frac{180}{\pi}$$

$$\mathbf{AURIL} = \theta_l * \frac{180}{\pi}$$

$$\mathbf{AURIR} = \theta_r * \frac{180}{\pi}$$

Figure 2.8 Formulations of the variables AVVU, ASIN, ALOB, AURIL and AURIR.



In this representative example:

$$\mathbf{a_v6} = \Delta\ \boldsymbol{landmarks}\ (\mathbf{7-10-13})$$

where;

$$\mathbf{a_v6} = \frac{\mathbf{1}}{\mathbf{2}}\left(\begin{vmatrix} x7 & x10 \\ y7 & y10 \end{vmatrix} + \begin{vmatrix} x10 & x13 \\ y10 & y13 \end{vmatrix} + \begin{vmatrix} x13 & x7 \\ y13 & y7 \end{vmatrix}\right)$$

$$\mathbf{a_l6} = \Delta\ \boldsymbol{landmarks}\ (\mathbf{7-11-13})$$

where;

$$\mathbf{a_l6} = \frac{1}{2}\left(\begin{vmatrix} x7 & x11 \\ y7 & y11 \end{vmatrix} + \begin{vmatrix} x11 & x13 \\ y11 & y13 \end{vmatrix} + \begin{vmatrix} x13 & x7 \\ y13 & y7 \end{vmatrix}\right)$$

A: The area of the polygon formed by the landmarks (or semilandmarks) 1 to *n*, where;

$$\mathbf{A} = \frac{\mathbf{1}}{\mathbf{2}}\left(\begin{vmatrix} x1 & x2 \\ y1 & y2 \end{vmatrix} + \begin{vmatrix} x2 & x3 \\ y2 & y3 \end{vmatrix} + \cdots \begin{vmatrix} xn & x1 \\ yn & y1 \end{vmatrix}\right)$$

$$\mathbf{TAA} = \frac{\mathbf{1}}{\mathbf{2}}\left(\begin{vmatrix} x26 & x27 \\ y26 & y27 \end{vmatrix} + \begin{vmatrix} x27 & x28 \\ y27 & y28 \end{vmatrix} + \begin{vmatrix} x28 & x26 \\ y28 & y26 \end{vmatrix}\right)$$

$$\mathbf{PERILAM} = \boldsymbol{l} = \sum_{i=1}^{n=1999} d_{s(i)-s(i+1)}$$

$$\mathbf{LOBTSL} = \sum_{i=1}^{n=6} \frac{a_l i}{\mathbf{6}}$$

$$\mathbf{LOBTSV} = \sum_{i=1}^{n=6} \frac{a_v i}{\mathbf{6}}$$

$$\mathbf{AREALAM} = \frac{\mathbf{1}}{\mathbf{2}}\left(\begin{vmatrix} sx1 & sx2 \\ sy1 & sy2 \end{vmatrix} + \begin{vmatrix} sx2 & sx3 \\ sy2 & sy3 \end{vmatrix} + \cdots \begin{vmatrix} sx2000 & sx1 \\ sy2000 & sy1 \end{vmatrix}\right)$$

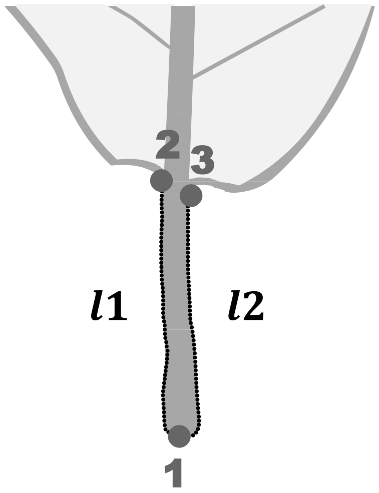
Note:

1. sx, sy **: semilandmarks coordinates.**

2. ℓ **= cumulative chordal distance for lamina outline, it is the total sum of Euclidean distances between semilandmarks which starts from landmark-2 and ends in landmark-3.**

3. Number of outline points in representations may not be equal with original semilandmark number 2000.

Figure 2.9 The determinant method for calculating area and formulations of the variables LOBTSV, LOBTSL, TAA, AREALAM and PERILAM.

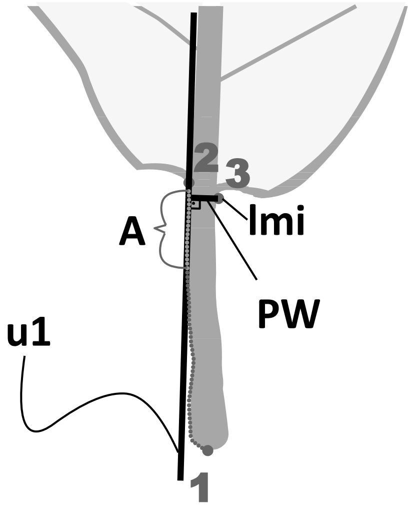


$l1$ = cumulative chordal distance for the left part of the petiole outline.

$l2$ = cumulative chordal distance for the right part of the petiole outline.

$$\mathbf{PL} = \frac{l1 + l2}{2}$$

Figure 2.10 Calculation of variable PL.



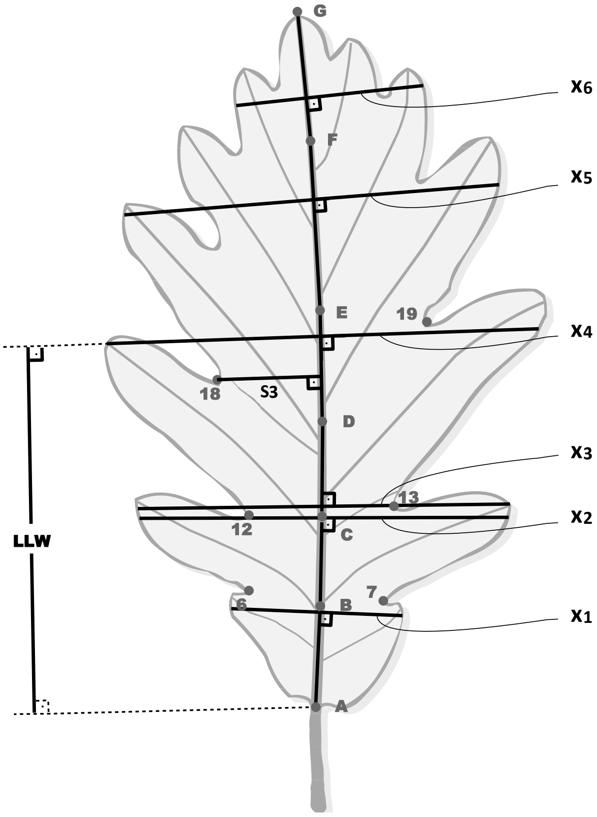
lmi: initial landmark for calculating PW, it is either lm2 or lm3, depending on its y-coordinate value. The one which has lower y-value is lmi.

A : A set of 20 semilandmark point which are closest to initial landmark (*lmi*) on the other side of petiole outline.

u1: Overlapping line with the first eigenvector of the covariance matrix of *A*'s XY-coordinate data.

PW: Petiole width which is the orthogonal distance between the *lmi* point and the line of *u1*.

Figure 2.11 Calculation procedures for variable PW.



In this representative example:

X4 = LW

S3 = Sinus width under third lobe.

$\mathrm{X}i =$ A local maxima of the lamina width between each respective midrip landmark pairs, where the mesurements were orthogonal to the connecting lines between midrip landmarks*.
Si = The sinus widths under six largest lobes.
$\mathrm{Vx}i =$ A vector of *Xi* (X*1* to X6) values.

$\mathrm{LW} =$ Largest value of the $\mathrm{X}i$ in the vector $\mathrm{Vx}i$.

$$\mathbf{SW} = \sum_{i=1}^{n=6} \frac{Si}{6}$$

$\mathrm{LLW} =$ The orthogonal distance between the largest part and the lamina base.

*** See the text and Figure 2.5 for information about midrib landmarks.**

***Notes*:**
1. The rotation process in the calculation of LW is not represented in the figure.
2. Due to the arching of the midrip, if more than one orthogonal distace do exist for Si, the shortest one is taken.

Figure 2.12 Formulations and calculation of variables LW, LLW and SW.

2.2.3. Calculated Variables

Most of the calculated variables were simple ratios between two variables. The formulations of calculated variables were presented in the Table 2.3.

Table 2.3 Formulations of calculated variables (see Table 2.2 for codes).

$\mathrm{PRL} = 100 * \frac{\mathrm{PL}}{\mathrm{PL} + \mathrm{LL}}$	$\mathrm{LOBTAR} = \frac{\mathrm{LOBTSL}}{\mathrm{LOBTSV}}$
$\mathrm{PRW} = 100 * \frac{\mathrm{PW}}{\mathrm{PW} + \mathrm{LL}}$	$\mathrm{LOBAAR} = \frac{\mathrm{LOBTSL}}{\mathrm{TAA}}$
$\mathrm{LWR} = 100 * \frac{\mathrm{LW}}{\mathrm{LL}}$	$\mathrm{PLWR} = \frac{\mathrm{PW}}{\mathrm{PL}}$
$\mathrm{LWRL} = 100 * \frac{\mathrm{LLW}}{\mathrm{LL}}$	$\mathrm{LWS} = \frac{\mathrm{LOBWU}}{\mathrm{LOBWD}}$
$\mathrm{ARPE} = \frac{\sqrt{\mathrm{AREALAM}}}{\mathrm{PERILAM}}$	$\mathrm{LLLR} = \frac{\mathrm{LOBL}}{\mathrm{LL}}$
$\mathrm{ISOP} = 1 - 4 * \pi * \frac{\mathrm{AREALAM}}{\mathrm{PERILAM}^2}$	$\mathrm{LLWR} = \frac{\mathrm{LOBL}}{\mathrm{LW}}$
$\mathrm{ELD} = \frac{\pi * \mathrm{LL} * \mathrm{LW}}{4 * \mathrm{AREALAM}}$	$\mathrm{LOBLWR} = \frac{\mathrm{LOBWU} + \mathrm{LOBWD}}{\mathrm{LOBL}}$
$\mathrm{IVLOB} = 100 * \frac{\mathrm{NIV}}{\mathrm{NLOB}}$	$\mathrm{AURISIN} = \frac{\mathrm{ASIN} * 2}{\mathrm{AURIL} + \mathrm{AURIR}}$
$\mathrm{PERINL} = \frac{\mathrm{PERILAM}}{\mathrm{NLOB}}$	$\mathrm{AURIVVU} = \frac{\mathrm{AVVU} * 2}{\mathrm{AURIL} + \mathrm{AURIR}}$
$\mathrm{PERINLL} = \frac{\mathrm{PERILAM}}{\mathrm{NLOB} * \mathrm{LL}}$	$\mathrm{LLLBLV} = \frac{\mathrm{LBLV}}{\mathrm{LL}}$
$\mathrm{LUBLOB} = \frac{\mathrm{NLUB}}{\mathrm{NLOB}}$	$\mathrm{LLBLLR} = \frac{\mathrm{LLBL}}{\mathrm{LL}}$

2.2.4. Counted Variables

Number of lobes (NLOB) is the total number of lobes (both left and right part) except the terminal lobe (the apex). In order to be considered as a lobe, every lobe must have at least a clearly identified second order lobe vein. Lobules are the lobes which have been irrigated by the third order veins. Number of lobules (NLUB) is the total number of the lobules on the lamina. An intercalary vein is also a second order vein but irrigating a sinus instead of a lobe. A second order vein was considered as an intercalary vein and counted in the number of intercalary veins (NIV) if it is extending at least half way from midrib directly to the sinus base (Kremer *et al*., 2002)

2.2.5. Descriptive Variables

Lobe shape (LOBTPS) was scored using the nine-step index given in the Figure 2.13 which was prepared from the samples used in this study. TEETH is a binary variable for the presence or absence of teethes on the leaf margins. Lamina base shape (BSL) was scored by the use of nine-step grading system which was presented in the Kremer *et al*. (2002).

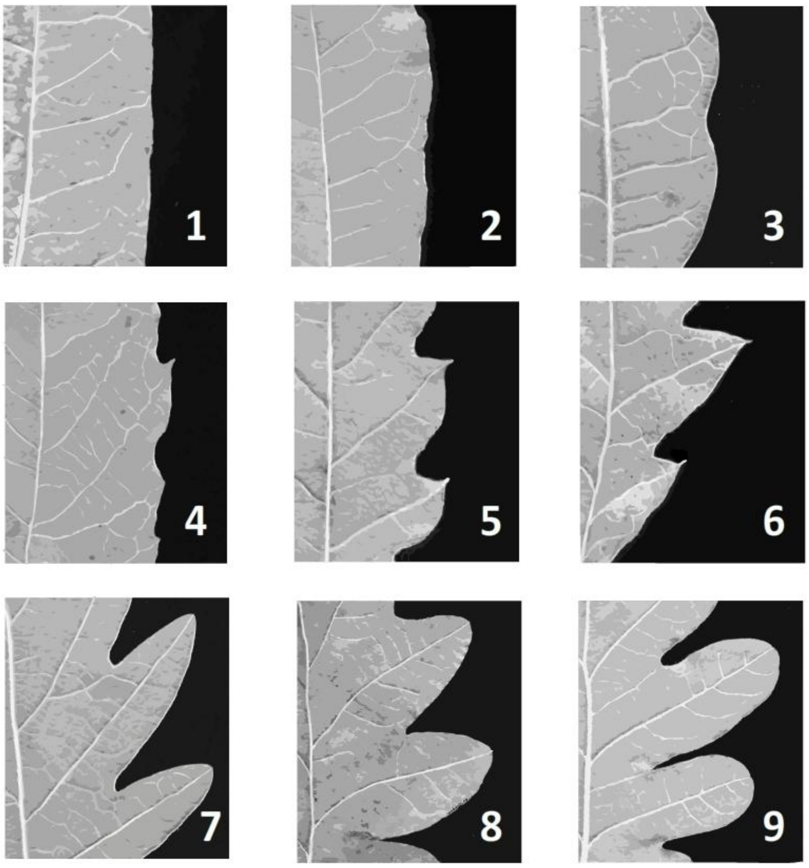


Figure 2.13 Nine-step lobe shape index which was prepared from the samples used in this study.

2.3 Data Analysis

Final data matrix for specimen means (not shown) was prepared in R. Each row represents a single specimen while each column stands for a variable. The data matrix was scaled in order to have unit variance and zero mean in each column (converted to correlation matrix). The csv table, which was comprised of the identification information for the specimens, was imported to R. Previously identified specimens, which were clearly assigned to a species (p, i or m) were selected as a training data. Random forest classifier in the R package randomForest (Liaw and Wiener, 2002) was used for measuring the discrimination importance of the predictor variables for the training data set. The variables, which have positive values for mean decrease in accuracy and mean decrease in Gini impurity, were chosen for constructing new data matrix. First 10 important variables were represented as boxplots with the use of package graphics (R Development Core Team, 2008). A new training set was reconstructed from new matrix. Random forest was run with new training set to predict the classes of full data set, so every individual in the study was assigned one species or other. Posterior probabilities of class memberships for all specimens were also predicted. Initial identifications, biplots and final probabilities of class membership were shown on PCA plots respectively. PCA was performed by the package stats (R Development Core Team, 2008). Since the first two principle components gave the best separation between groups, first two PC were represented in these plots. Small pie-charts on PCA plot were used to show the probabilities of class membership by use of package plotrix in R (Lemon *et al*., 2009).

2.4. Micromorphological Methods

Dry materials have been used for micromorphological observations on abaxial lamina trichomes and epicuticular waxes around the stomas. Initial observations were done by a stereomicroscope. Small pieces were cut from the middle part of two leaves from each of 24 specimens which were chosen for micromorphological study. Pieces were fixed on aluminum stubs and gold coated. Investigations were done with the JEOL JSM–6060 scanning electron microscope (SEM) in the Gazi University Faculty of Arts and Science Department of Biology. The terminology used for the classification of trichomes is taken from Hardin (1976) while the terminology for the epicuticular waxes taken from Bussotti and Grossini (1997).

CHAPTER 3

RESULTS

3.1 Random Forest for Variable Selection

In the random forest constructed for variable selection, forty eight of all specimens in the training set were classified into correct classes (majority of votes determine the classes so the posterior probabilities for class membership were greater than 0.5 for these individuals) with a zero oob error rate (Table 3.1). A single classification tree which was picked up from 20.000 trees in the random forest was given as an example in the Figure 1.1. In this tree, all specimens in the training set were correctly classified, where the specimens were identified as (m) if the variable AREALAM gets higher value than -0.187. Specimens which gets lower than -0.187 were assigned to (p) or (i), and then, if NIV was higher than 0.009 specimen was identified as (p), else (i).

Table 3.1 Confusion matrix for training data set.

OOB estimate of error rate: 0%

	i	m	p	class error
i	11	0	0	0
m	0	20	0	0
p	0	0	17	0

Variable-importance-measure produced by random forest turns out that the 38 variables out of 52 were taken a part in discrimination for the training set. These variables get positive values for both mean decrease in accuracy and node impurity (Figure 3.1 and 3.2). This means that when these variables are excluded from the study, mean predicting accuracy and node impurity are decreasing so that the ability of the classification also decreases. However, the remaining 14 variables have either no effect on discrimination and get zero or play as a confounder by getting negative values.

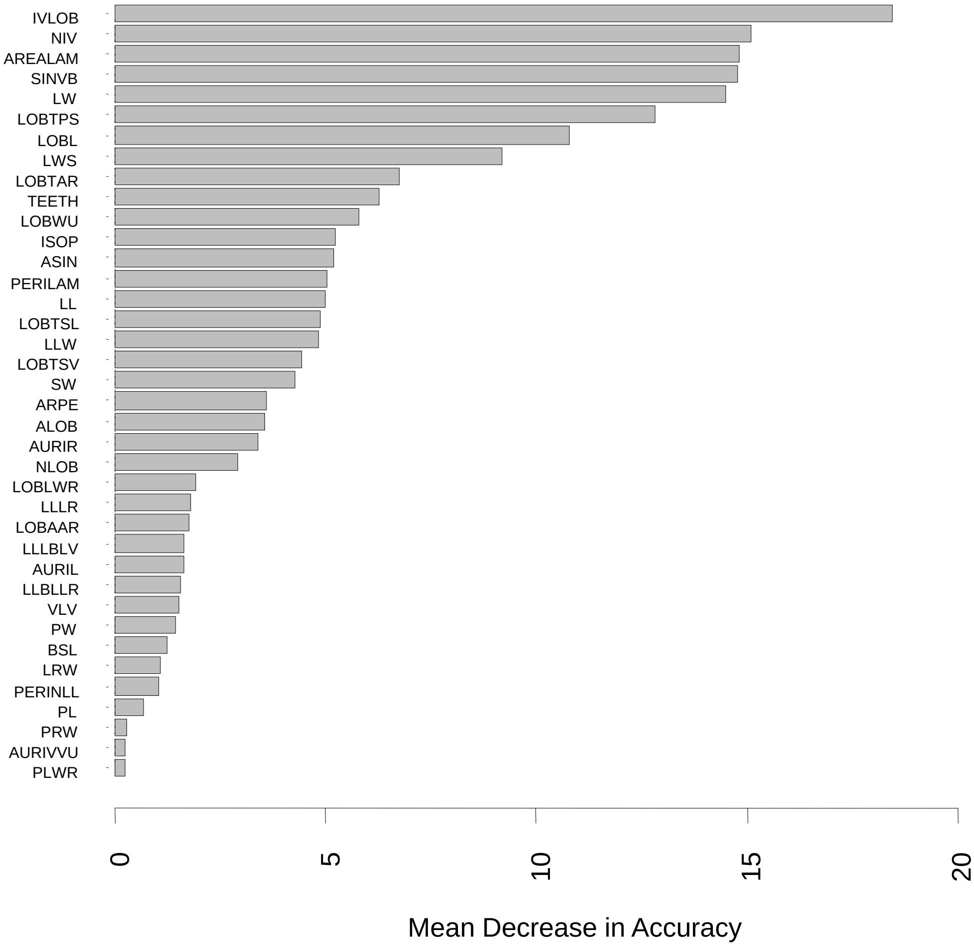


Figure 3.1 Variables which have mean decrease in accuracy measure greater than zero. They were chosen for further analysis (presented in decreasing order).

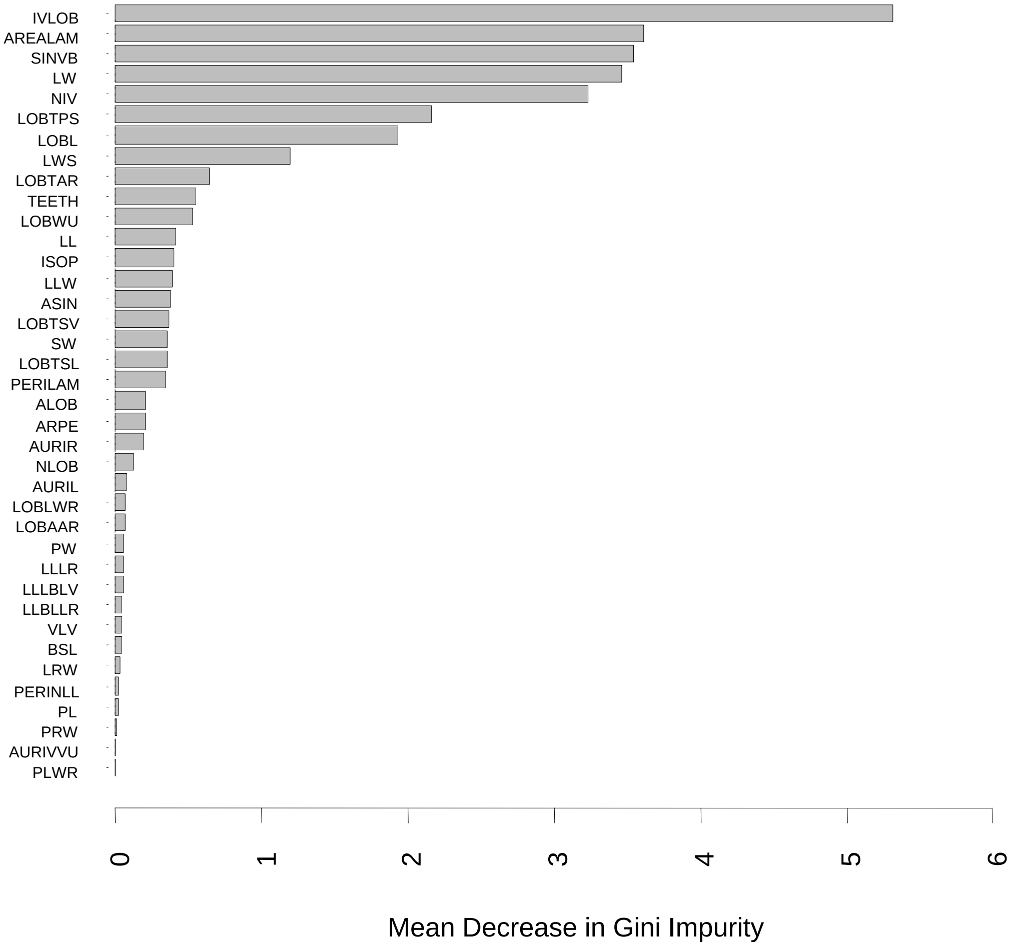


Figure 3.2 Variables which have mean decrease in Gini impurity measure greater than zero. They were chosen for further analysis (presented in decreasing order)

Negative values arise when the number of correct classes obtained by randomly permuted variables is greater than the number of the correct classes obtained by the original data. The predictive accuracy increases by permutation so that the real states of these variables are not useful in discrimination. Negative values for mean decrease in Gini impurity reflect the increasing heterogeneity between nodes of the tree which decreases the goodness-of-split. First ten important variables, were plotted as box-plots to show the difference between groups (including putative hybrids and nid specimens) (Figure 3.3)

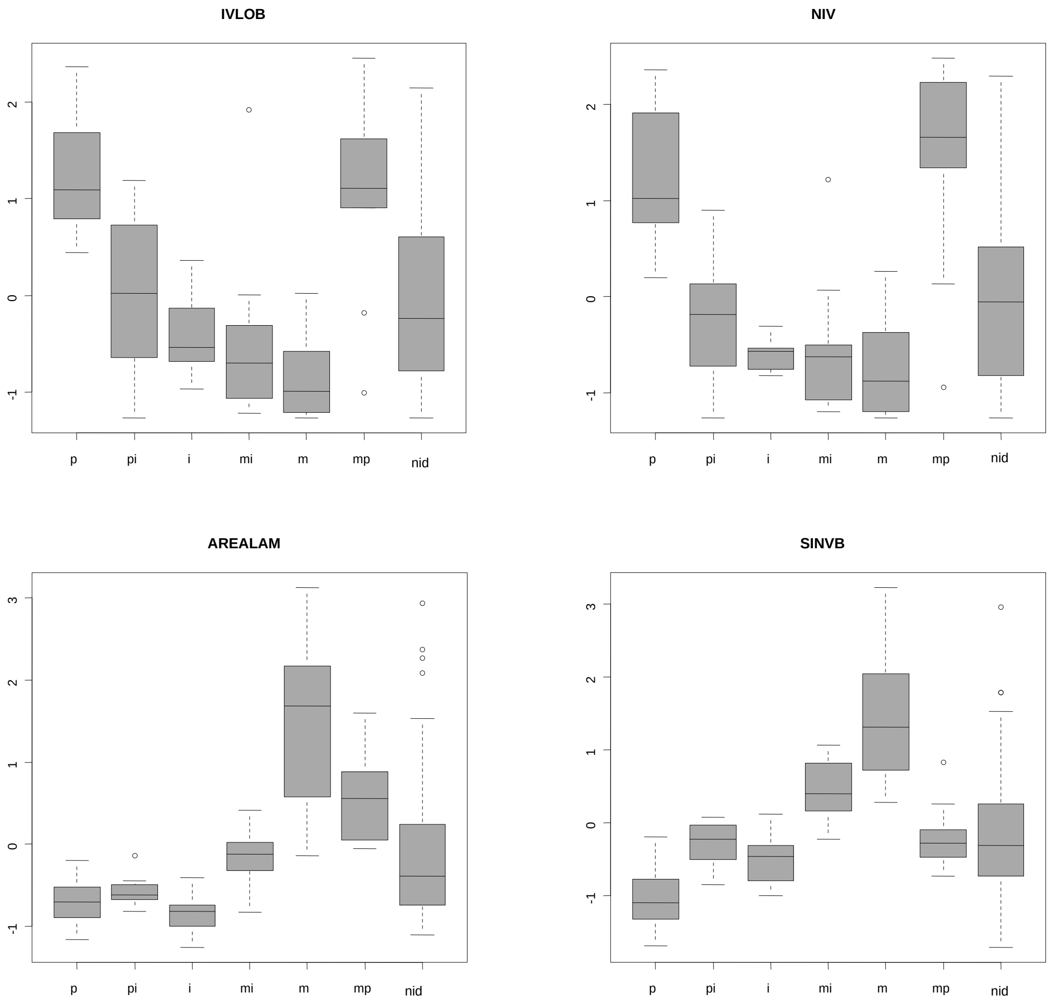


Figure 3.3 Box-plots representation for the first ten variables selected by random forest importance measurements (See Table 2.1 for group codes).

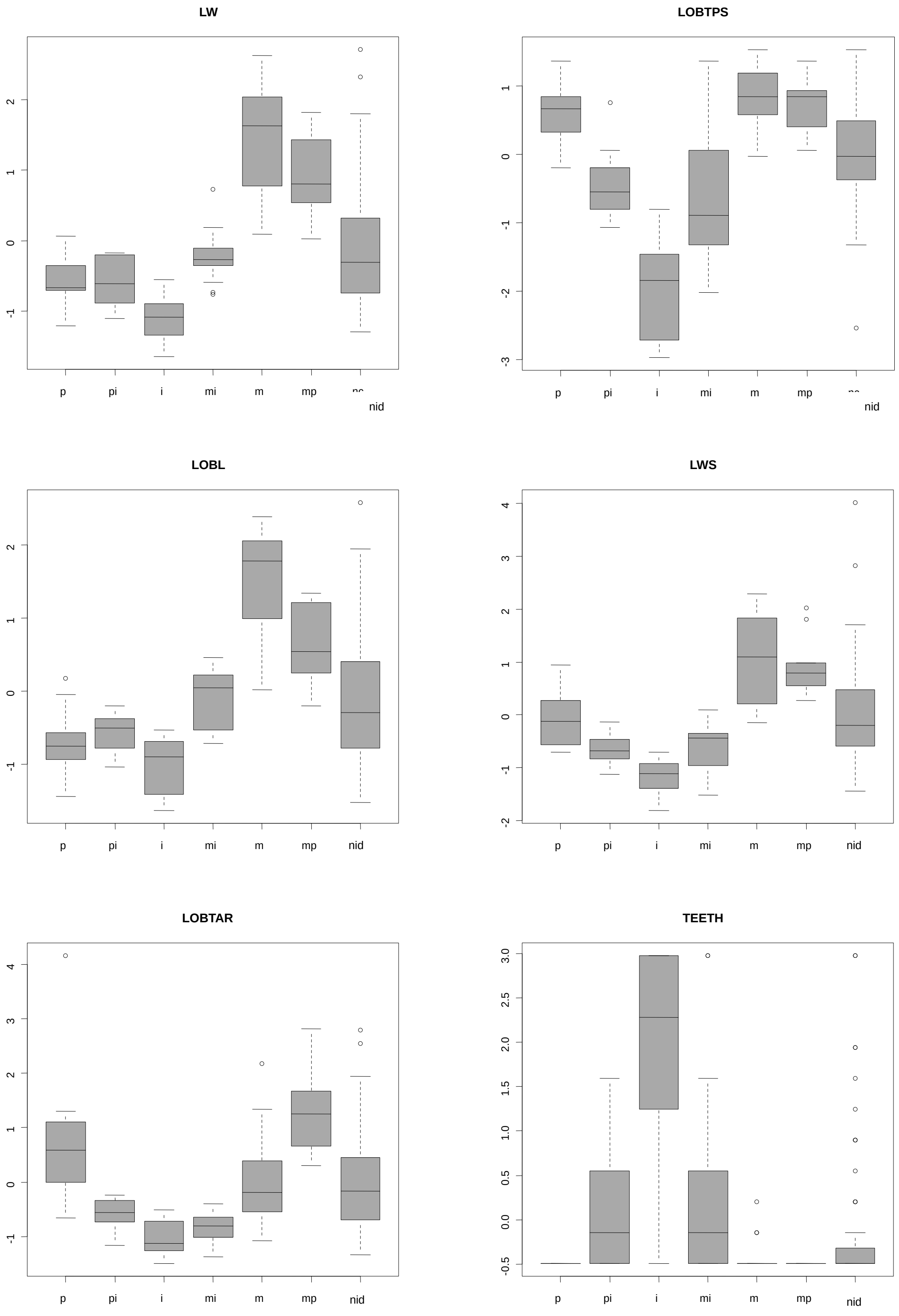


Figure 3.3 C*ontinued.*

Some variables are diagnostic for discriminating among at least the typical members of these species. For instance, the variables IVLOB and NIV discriminate *Quercus pubescens* and putative hybrid *Q. macranthera* subsp. *syspirensis* x *Q. pubescens* from other taxa while size related AREALAM, LW and LOBL separate *Q. macranthera* subsp. *syspirensis* from others. Lobe shape variables such as LOBTPS, LWS, LOBTAR, and the presence of teeth on the lobe tips (TEETH) successfully differentiated *Q. infectoria* subsp. *boissieri* from *Quercus pubescens* and *Q. macranthera* subsp. *syspirensis.* Putative hybrids between *Q. infectoria* subsp. *boissieri* and other species show intermediate characteristics for these variables. *Q. macranthera* subsp. *syspirensis* x *Q. pubescens* has relatively deeper lobes so the variable LOBTAR gets slightly greater value than its putative parents. Some variables are highly correlated such as IVLOB and NIV (Pearson correlation coefficients = 0.949). However both of these variables were included in the study for characterization of the hybrids.

3.2 Principle Component Analysis on Selected Variables

The loadings of the first and second principle components were provided in Figures 3.4 and 3.5, respectively. The first principle component explains 40.6% of the total variation and the second PC explains 18.605 % of the total variation (Figure 3.6, 3.7 and 3.8). There was not complete separation between groups particularly when the unidentified specimens were added in the analysis. First principle component (PC1) accounts for mostly the variation in size (Figure 3.4) and discriminates *Q. macranthera* subsp. *syspirensis* from the *Quercus pubescens* and *Q. infectoria* subsp. *boissieri* based on the size of the leaf (Figure 3.7) which has loaded on first component positively (Figure 3.4 and 3.7).

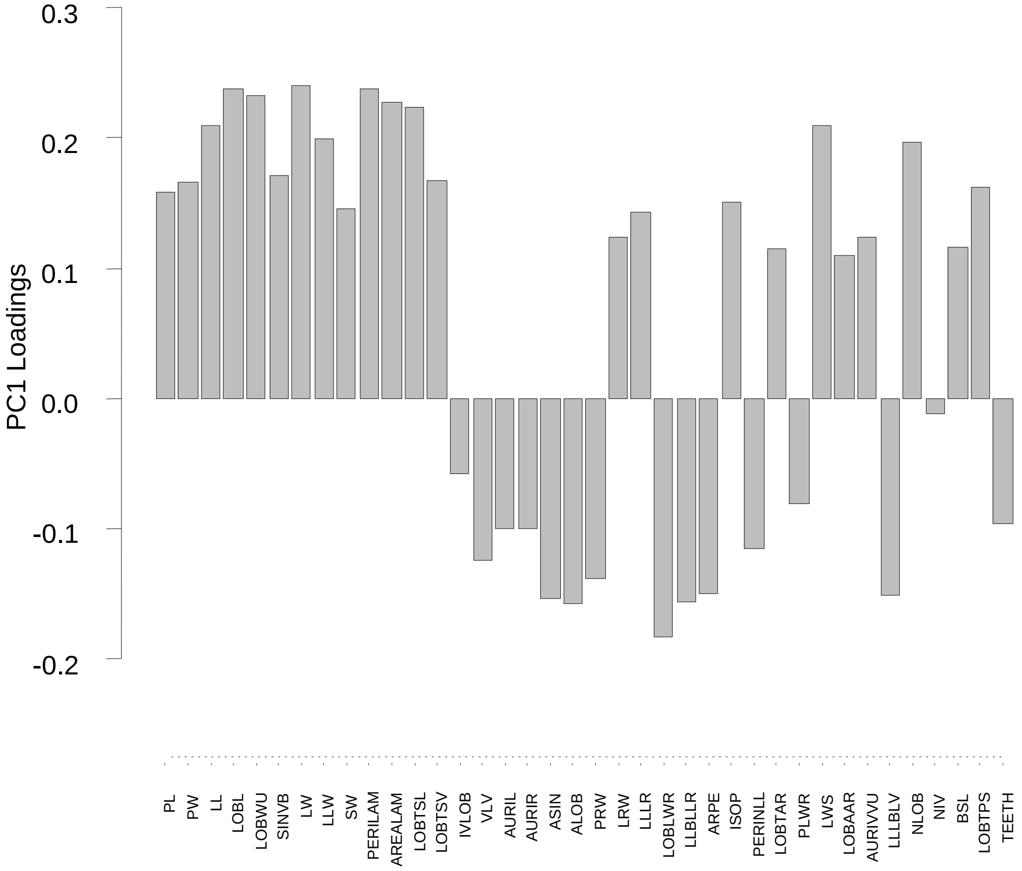


Figure 3.4 Loadings of the variables on the first principle component (PC1).

Second principle component is more or less related to leaf and lobe shape, and the number of intercalary veins (Figure 3.5). NIV has negative loading on the second PC and *Quercus pubescens* are differentiated with increasing number of intercalary veins (NIV) and the relative petiole thicknesses, PRW and PLWR (Figure 3.7). *Q. infectoria* subsp. *boissieri* specimens are separated based on lobe and leaf shape variables which have positive loadings on PC2 (Figure 3.5 and 3.7).

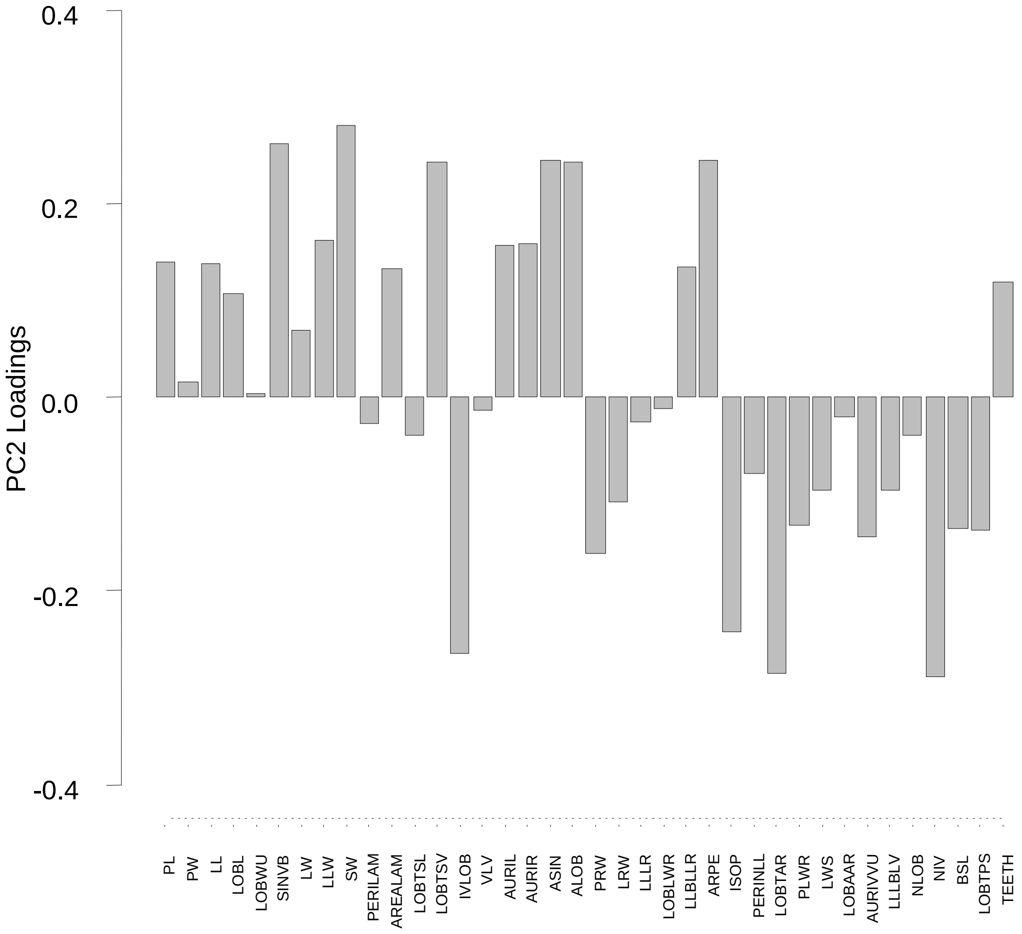
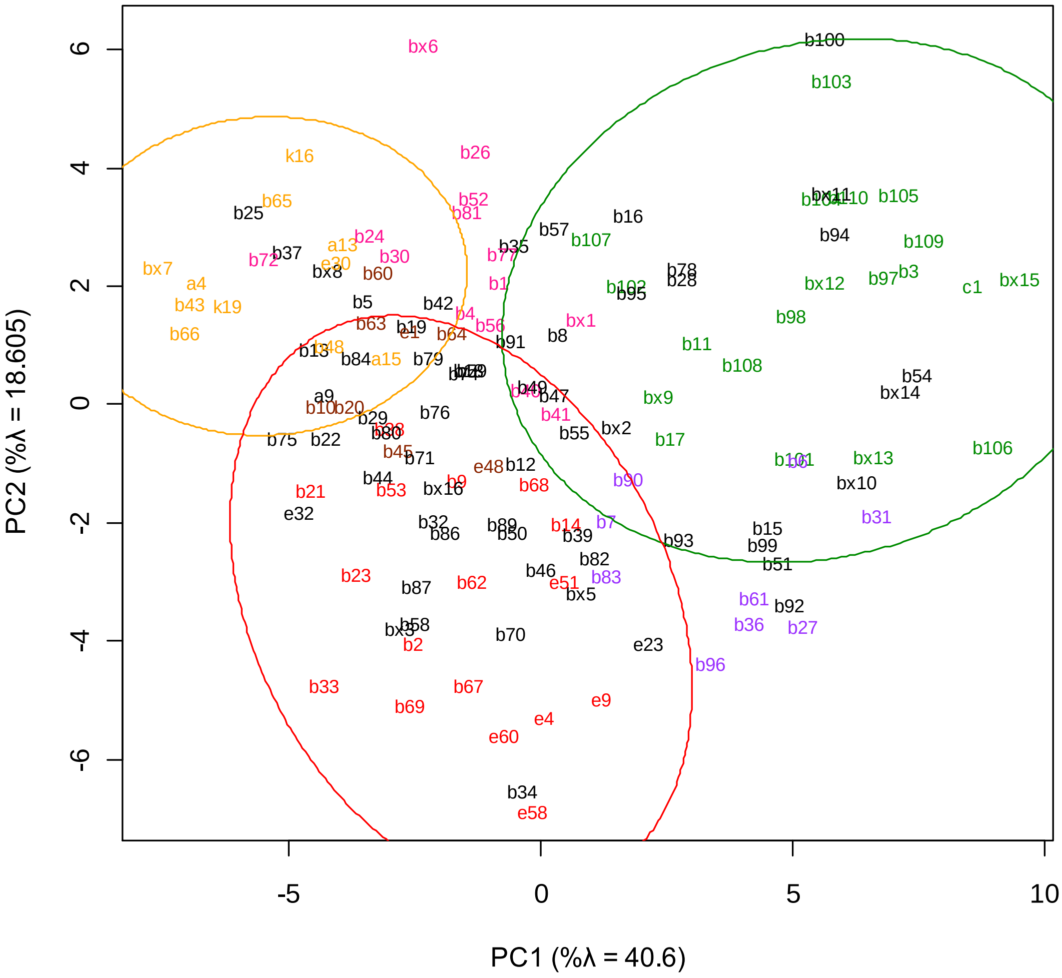


Figure 3.5 Loadings of the variables on the second principle component (PC2).

Putative hybrids were somewhere intermediate between parent clusters. However, some showed different morphometric pattern, rather than being the intermediates between putative parents, particularly for the specimens of bx6, b26, b27, b36, b61 and b92 (Figure 3.6).



Colors

🟥 *Q. pubescens* (p)

🟧 *Q. infectoria* subsp. *boissieri* (i)

🟩 *Q. macranthera* subsp. *syspirensis* (m)

🟫 *Q. pubescens* x *Q. infectoria* subsp. *boissieri* (pi)

🟥 *Q. macranthera* subsp. *syspirensis* x *Q. infectoria* subsp. *boissieri* (mi)

🟪 *Q. macranthera* subsp. *syspirensis* x *Q. pubescens* (mp)

⬛ Not identified (nid)

Codes

b : Beynam
e : Elmadag
a : Alacaatli
k : Kemer
c : Cubuk-Karagol
bx : Extras from Beynam

Figure 3.6 First two axes represented for the principle component analysis with specimen codes (95% confidence ellipses were drawn for the groups in the training set).

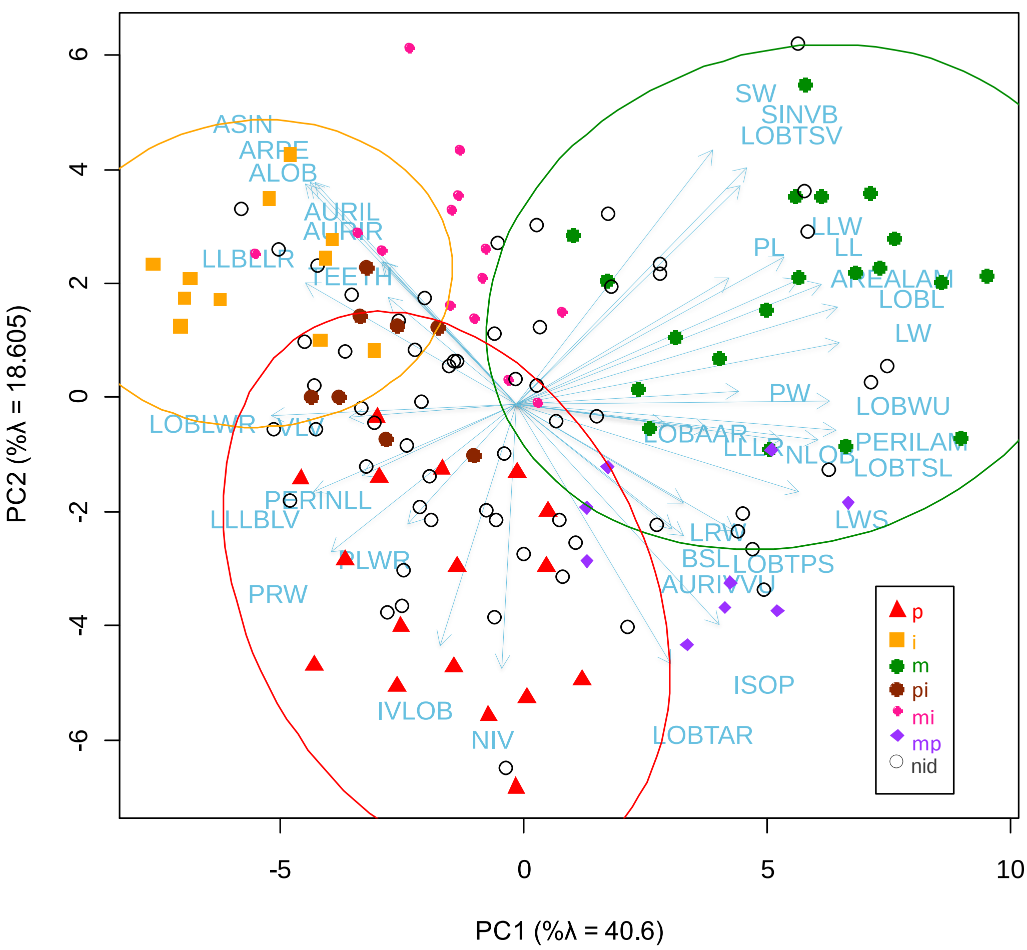


Figure 3.7 Biplot graphical representation for the first two PC axes and the variable loadings. Color legends used to show the individuals and groupings (95% confidence ellipses were drawn for the groups in the training set).

3.2 Random Forest for Predicting Posterior Probabilities of Class Membership

Random forest was constructed again from new training set including only 38 variables. All the forty-eight specimens in the training set were classified in correct class with a zero oob error rate. This new forest was used to predict the probabilities for class membership of the full data set (including training data). Since the training set was compromised of only the identified species, individuals have three probabilities for class membership and the results was presented by pie-charts on the PCA plot (Figure 3.8).

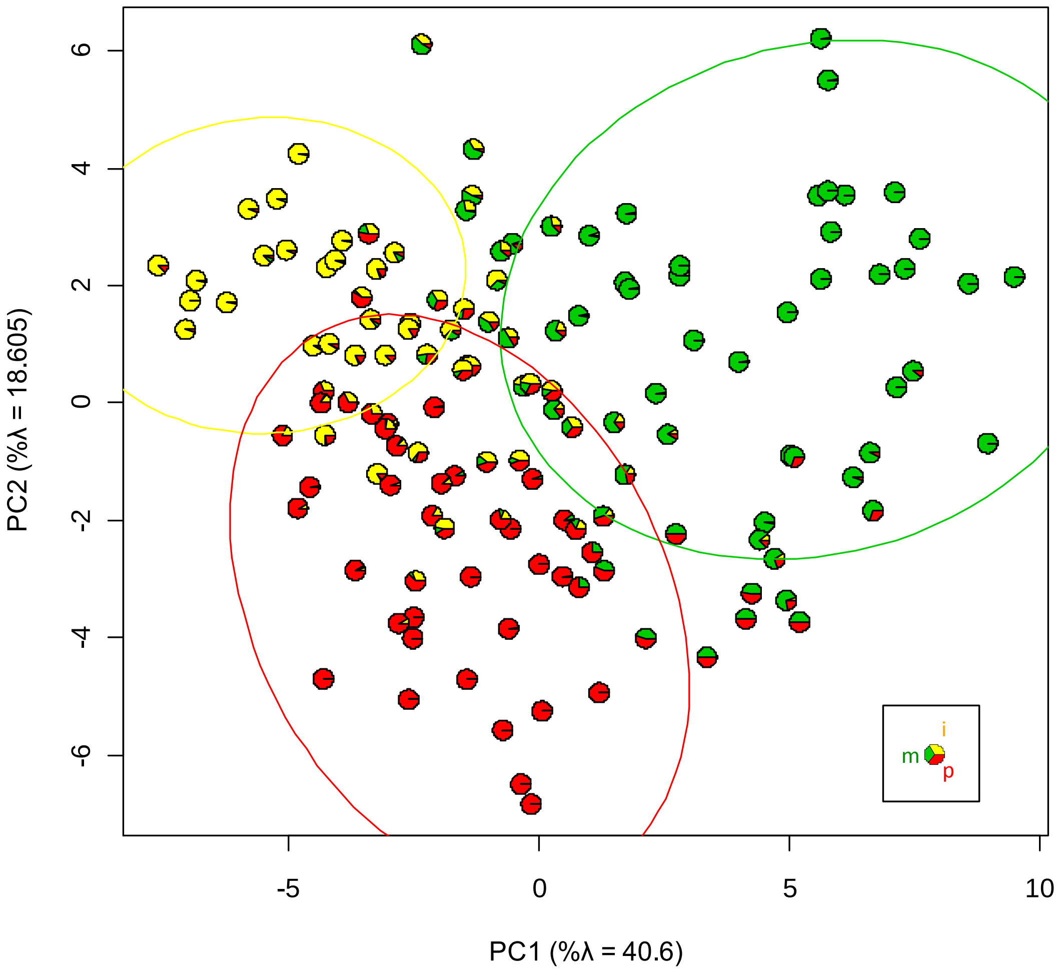


Figure 3.8 Posterior probabilities of class membership represented with pie-charts on the first two PC axes. Reds: *Q. pubescens*, Yellows: *Q. infectoria* subsp. *boissieri* and Greens: *Q. macranthera* subsp. *syspirensis*. (95% confidence ellipses were drawn for the groups in the training set).

Although the putative hybrids have intermediate probabilities for both parents classes, class memberships were not fully coherent with the PCA results. Some specimens in the centers of clusters have different class probabilities.

3.3 Micromorphological Investigations

Some of the used terms and explanations for micromorphological features were illustrated in Figure 3.9. The SEM micrographs of the selected specimens were provided from Figures 3.10 to 3.15.

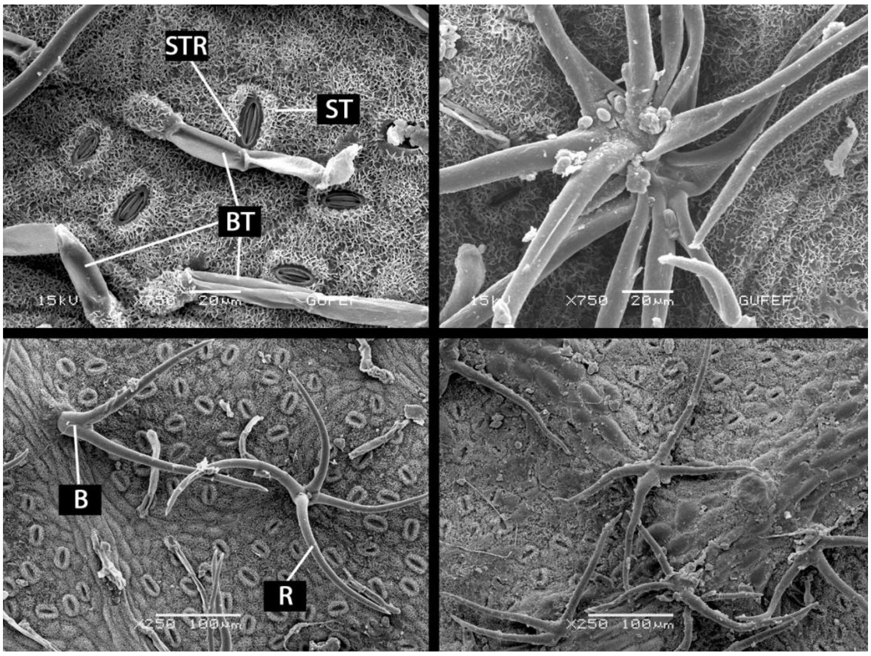


Figure 3.9 Some basic micromorphological features and terms which were mentioned in this study. **Upper-Left:** Abaxial lamina surface of the specimen b78 (group code nid), **ST:** Stoma, **STR:** Stomal rim, **BT:** Bulbous trichomes (X750, scale bar = 20µm). **Upper-Right:** Multiradiate trichome which radiates from more than one level, seen on the abaxial lamina surface of the specimen bx5 (p) (X750, scale bar = 20µm). **Below-Left:** Stipitate-fasciculate trichomes on *Q. macranthera* subsp. *syspirensis*, c1 (m), **B:** Stipitated trichome base, **R:** Trichome ray (X250, scale bar = 100µm). **Below-Right:** Flattened fasciculate trichomes of specimen e51 (p). (X250, scale bar = 100µm).

In *Q. pubescens,* the abaxial surface has fasciculate or stipitate-fasciculate trichomes with 5-8 rays, which are about 100-200 µm long. The cell walls of some trichomes have been collapsed. They are usually flattened onto the lamina so they may be easily confused with stellate trichomes (Figure 3.10.A). Bulbous trichomes are not present in typical *Q. pubescens* specimens, b2 and e60 (Figure 3.10.A and 3.10.E). But they are present in b68 and bx5 which were also identified as *Q. pubescens* at the beginning of the study. Abaxial lamina trichomes of *Q. infectoria* subsp. *boissieri* are very similar to *Q. pubescens* with slightly shorter ray lengths (100 µm) and 4 to 8 rays (Figure 3.11.I and 3.11.J). Bulbous trichomes are usually present in *Q. infectoria* subsp. *boissieri* but not regularly distributed over the lamina. *Q. macranthera* subsp. *syspirensis* has stipitate-fasciculate trichomes with 2 and/or 4 rays (about 150-300 µm long) on the abaxial lamina surface (Figure 3.11.N, 3.11.O, 3.11.P and 3.12.Q). Typically, the rays are regularly erect and the cell walls are never collapsed. Bulbous trichomes are generally numerous and uniformly distributed over the lamina. Unidentified specimens and putative hybrids which are related to *Q. macranthera* subsp. *syspirensis* and *Q. pubescens* have trichomes generally with longer rays (100-450 µm) with varying number of arms (4-12). Multiradiate trichomes, which radiates from more than one level, are also found on some of these specimens (3.10.G, 3.11.M, 3.12.W).

Abaxial lamina surfaces are covered by waxes as expected, since all the specimens belong to section *Quercus* which is characterized by vertically arranged epicuticular wax (Bussotti and Grossini, 1997). Stomal rim of *Q. pubescens* and *Q. infectoria* subsp. *boissieri* specimens are at least partially covered by wax except the specimen e51 (Figure 3.13.D). Wax scales were rather large and covered the abaxial surface densely in the *Q. infectoria* subsp. *boissieri* especially in a4 (Figure 3.14.I). *Q. macranthera* subsp. *syspirensis* has less densely distributed wax scales, stomal rim is more or less free from waxes and wax scales are forming a crown around the rim (Figure 3.14.N, 3.14.O, 3.14.P and 3.15.Q), but the distinction was not much with the *Q. pubescens* as stated in the Bussotti and Grossini (1997). Althought it was identified initially as *Q. pubescens,* scale properties of b68 is much similar to *Q. macranthera* subsp. *syspirensis* (Figure 3.13.B). Putative hybrids showed characteristics of one parent or another, such as b1 is very similar to *Q. macranthera* subsp. *syspirensis* based on wax density and the freedom of rim (Figure 3.14.K) while specimen b24 is more similar to the *Q. infectoria* subsp. *boissieri* but having smaller wax scales (Figure 3.14.L).

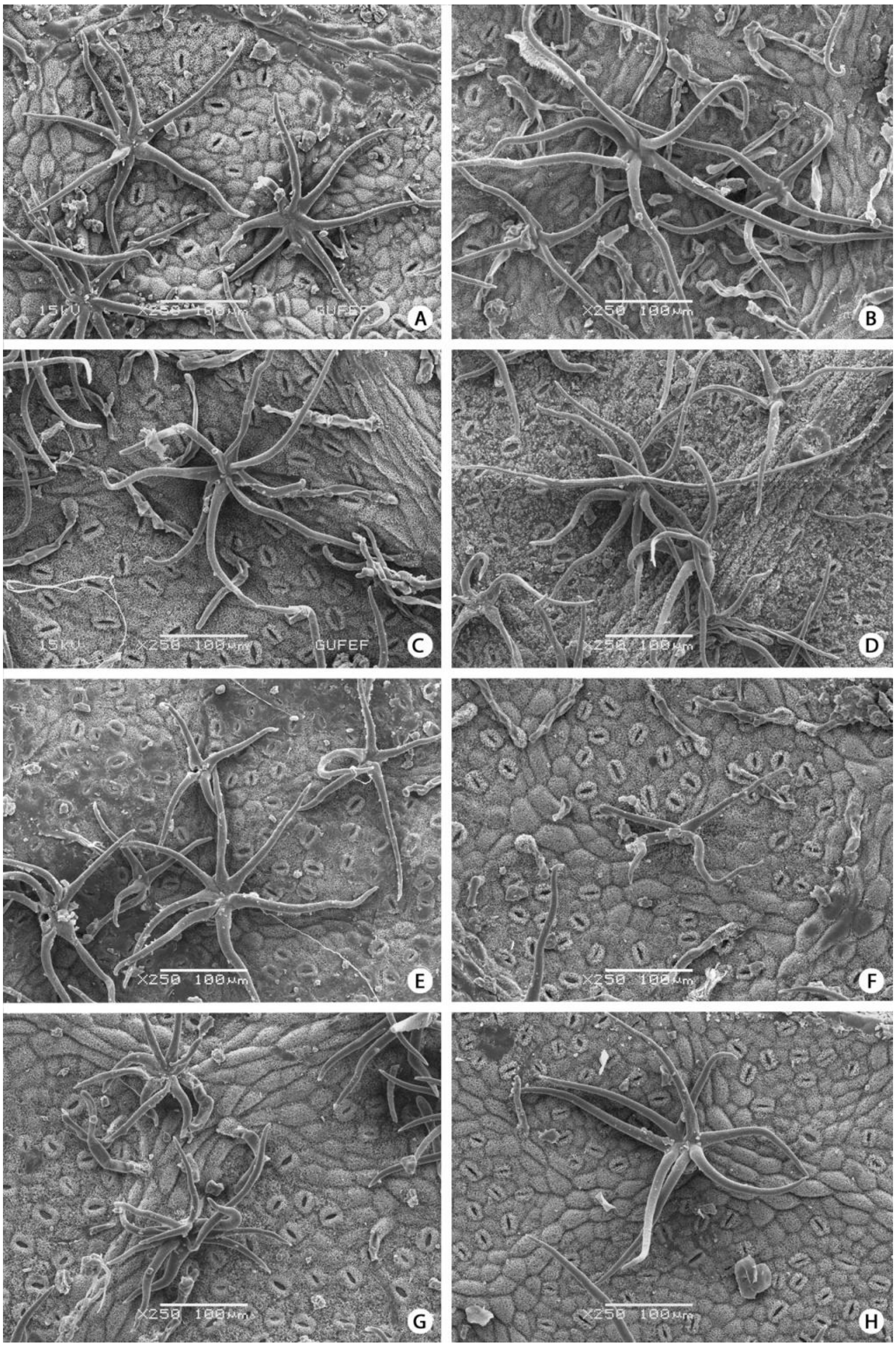


Figure 3.10 Micrographs of trichomes. **A:** Specimen b2 (Species p), flattened fasciculate trichomes. **B:** b68 (p), fasciculate trichomes. **C:** bx5 (p), fasciculate trichomes. **D:** e51 (p), fasciculate trichomes. **E:** e60 (p), flattened fasciculate trichomes. **F:** e1 (pi), fasciculate trichomes. **G:** b5 (nid), fasciculate and multiradiate trichomes. **H:** e32 (nid), fasciculate trichomes. SEM (X250), Scale bar = 100µm.

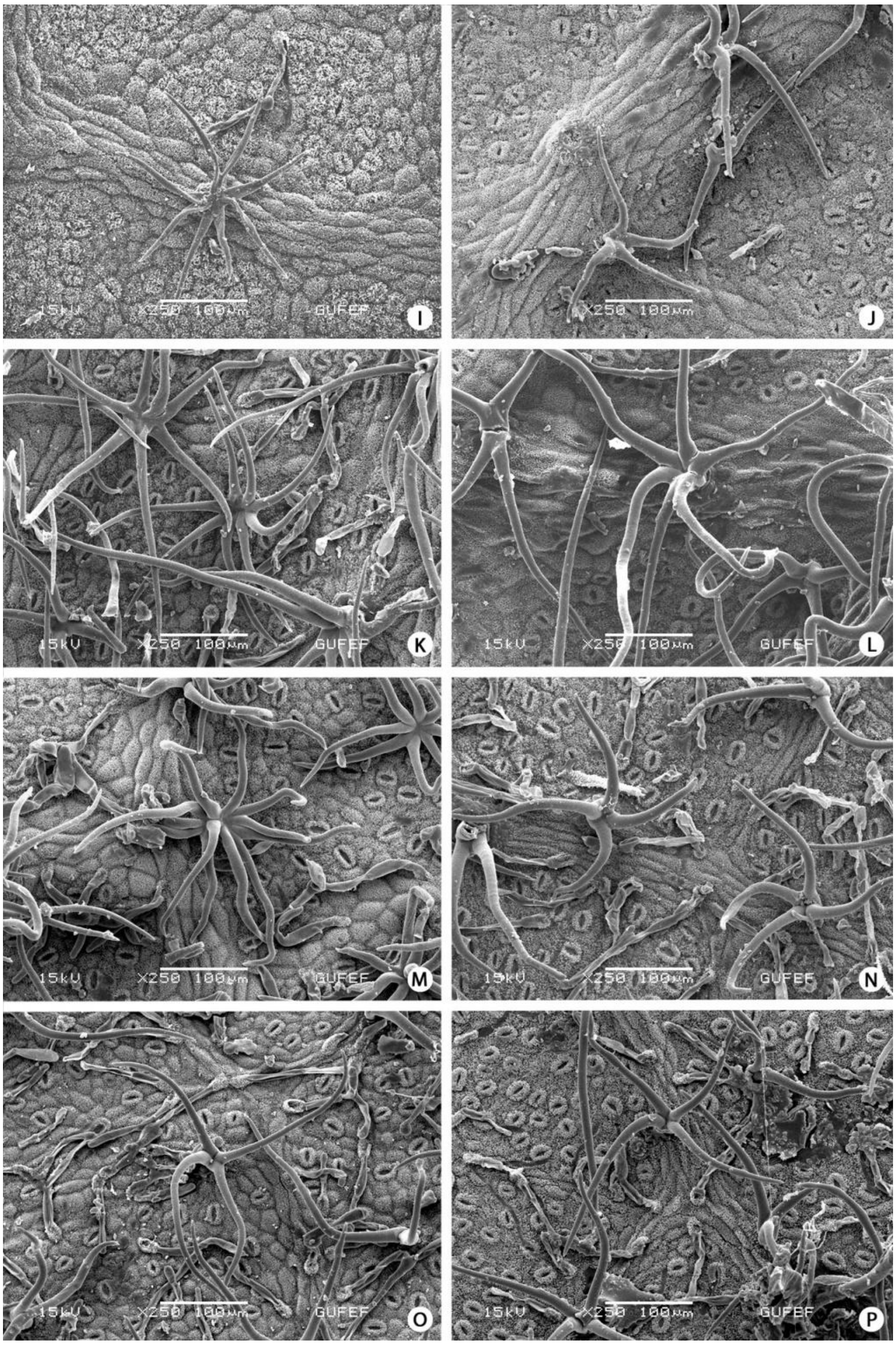


Figure 3.11 Micrographs of trichomes. **I:** a4 (i), fasciculate trichomes. **J:** a13 (i), fasciculate trichomes. **K:** b1 (mi), fasciculate trichomes. **L:** b24 (mi), fasciculate and stipitate-fasciculate trichomes. **M:** b8 (nid), fasciculate and multiradiate trichomes. **N:** b3 (m), stipitate-fasciculate trichomes. **O:** b17 (m), stipitate-fasciculate trichomes. **P:** b110 (m), stipitate-fasciculate trichomes. SEM (X250), Scale bar = 100µm.

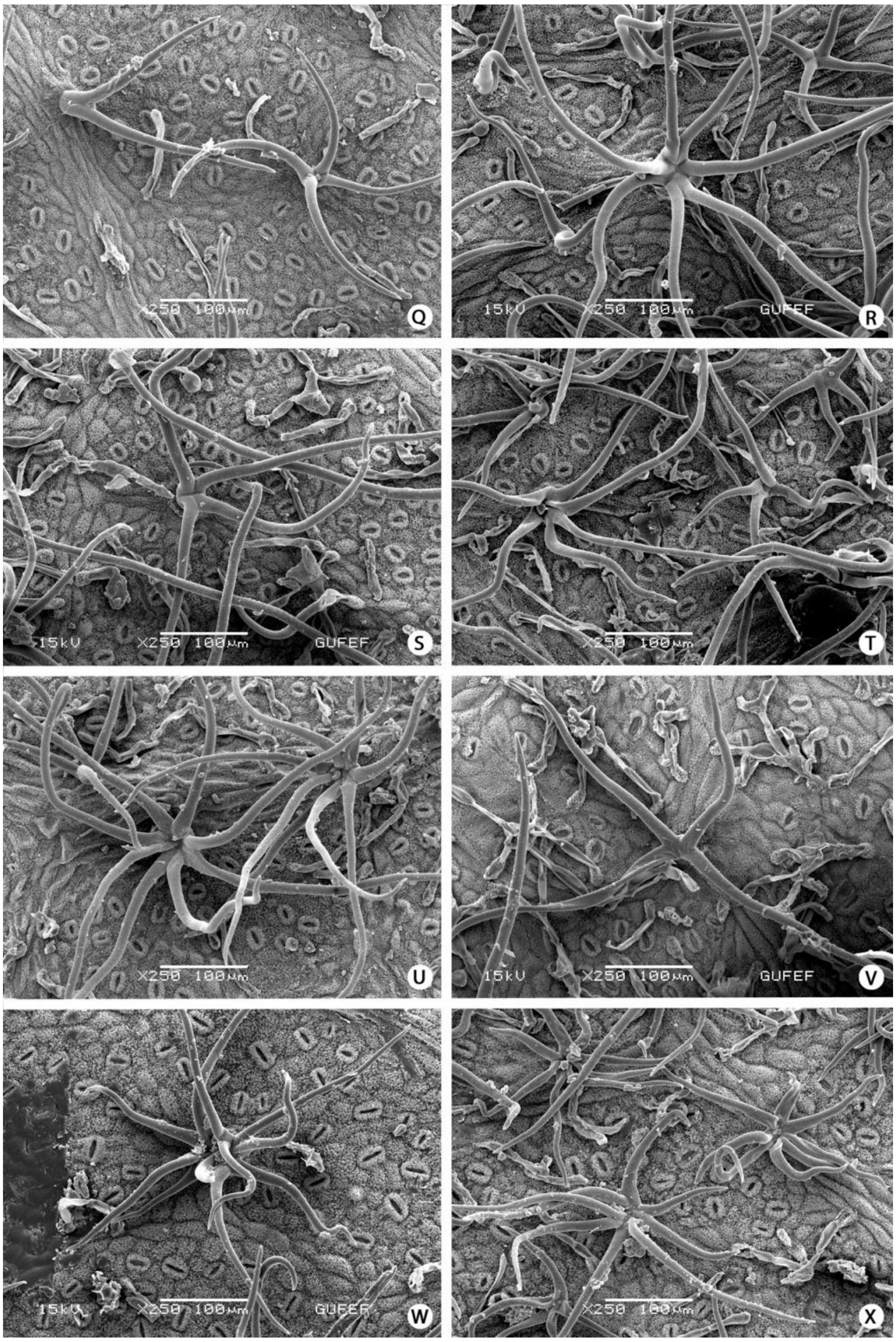


Figure 3.12 Micrographs of trichomes. **Q:** c1 (m), stipitate-fasciculate trichomes. **R:** b31 (mp), fasciculate and stipitate-fasciculate trichomes. **S:** b54 (nid), stipitate-fasciculate trichomes. **T:** b99 (nid), fasciculate and stipitate-fasciculate trichomes. **U:** b51 (nid), fasciculate trichomes. **V:** b78 (nid), fasciculate trichomes. **W:** bx10 (nid), multiradiate trichomes. **X:** b82 (nid), fasciculate and multiradiate trichomes. SEM (X250), Scale bar = 100µm.

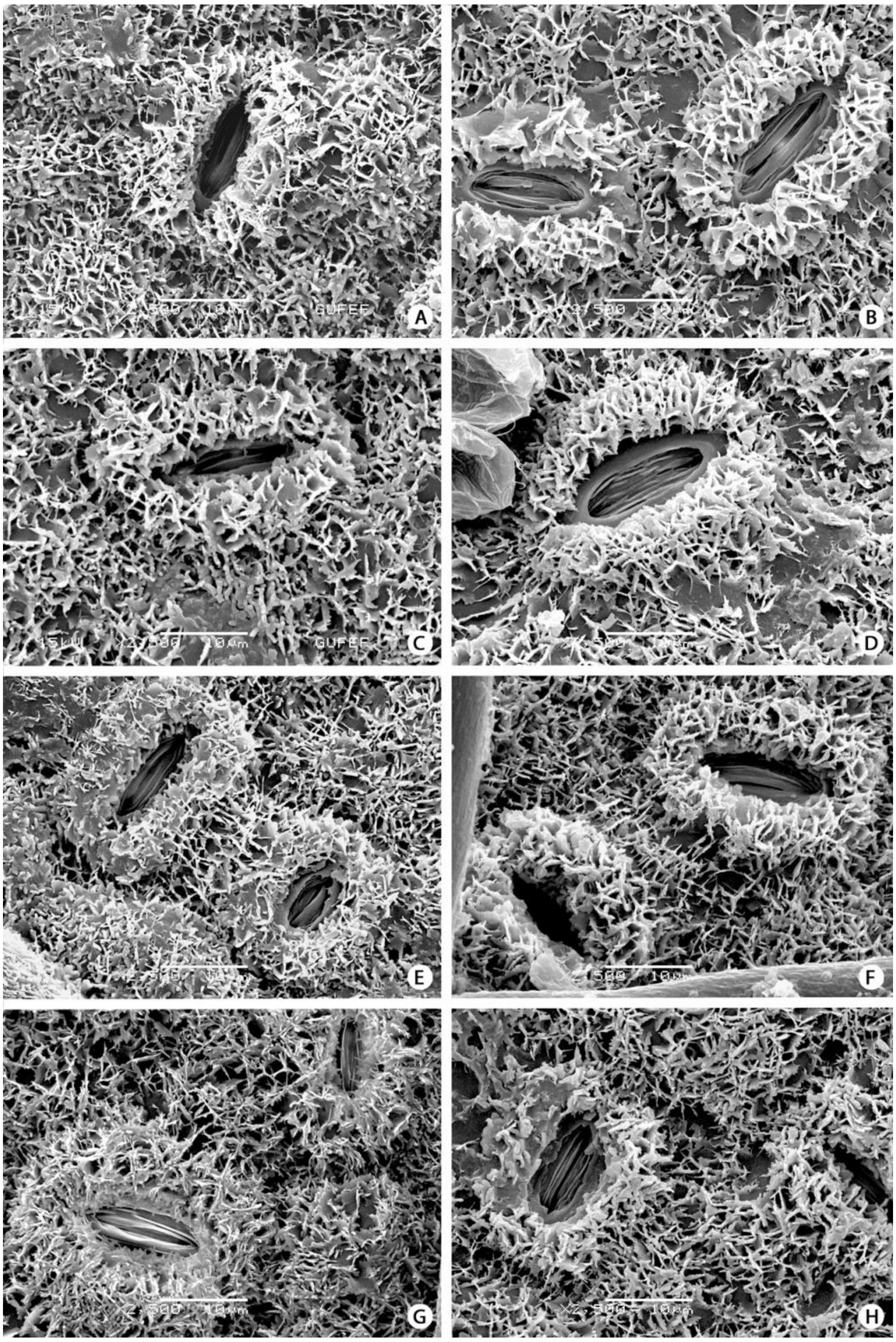


Figure 3.13 Micrographs of epicuticular wax and stoma. **A:** b2 (p). **B:** b68 (p). **C:** bx5 (p). **D:** e51 (p). **E:** e60 (p) **F:** e1(pi) **G:** b5 (nid). **H:** e32 (nid). SEM (X2500), Scale bar = 10µm.

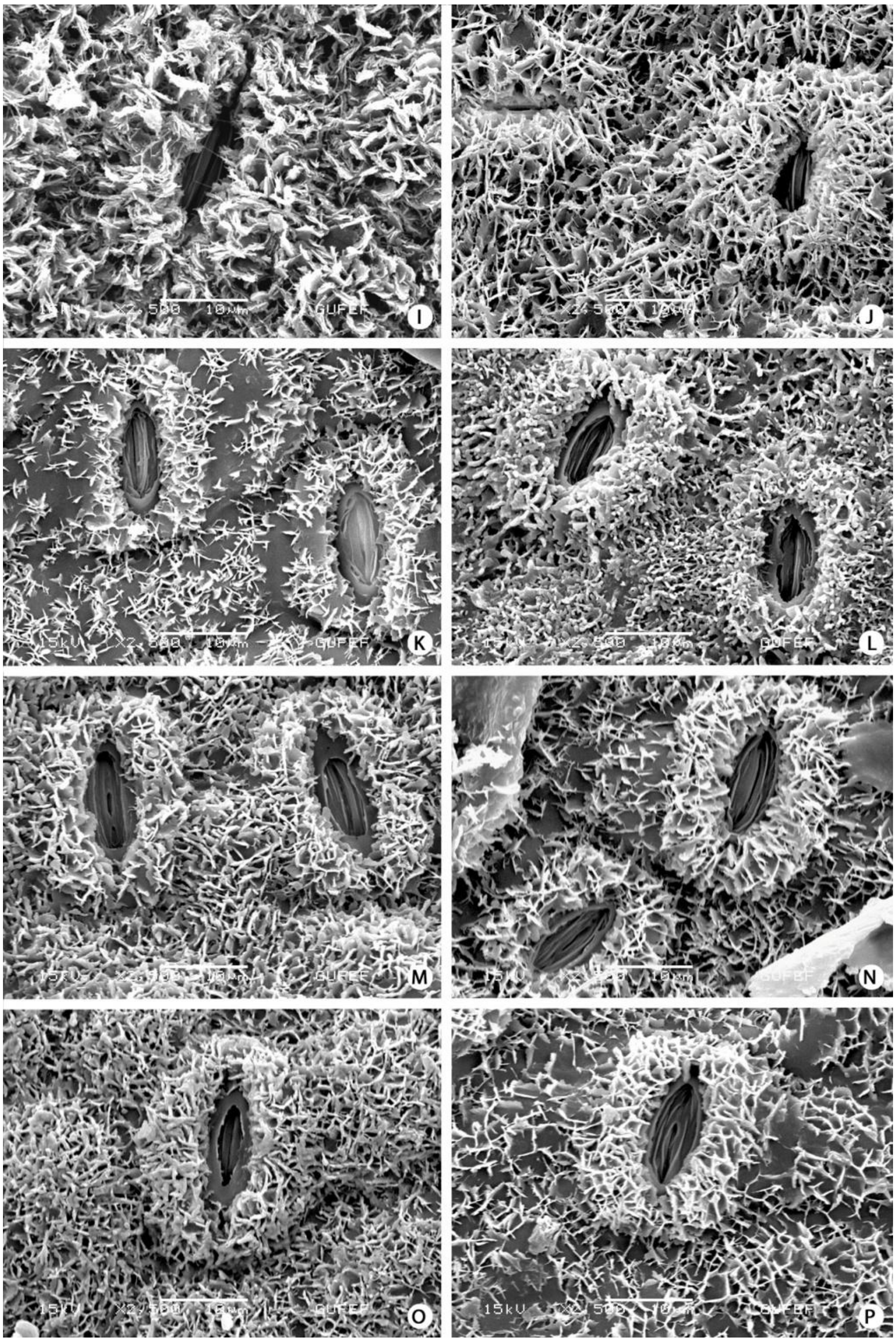


Figure 3.14 Micrographs of epicuticular wax and stoma**. I:** a4 (i). **J:** a13 (i). **K:** b1 (mi). **L:** b24 (mi). **M:** b8 (nid). **N:** b3 (m). **O:** b17 (m). **P:** b110 (m). SEM (X2500), Scale bar = 10µm.

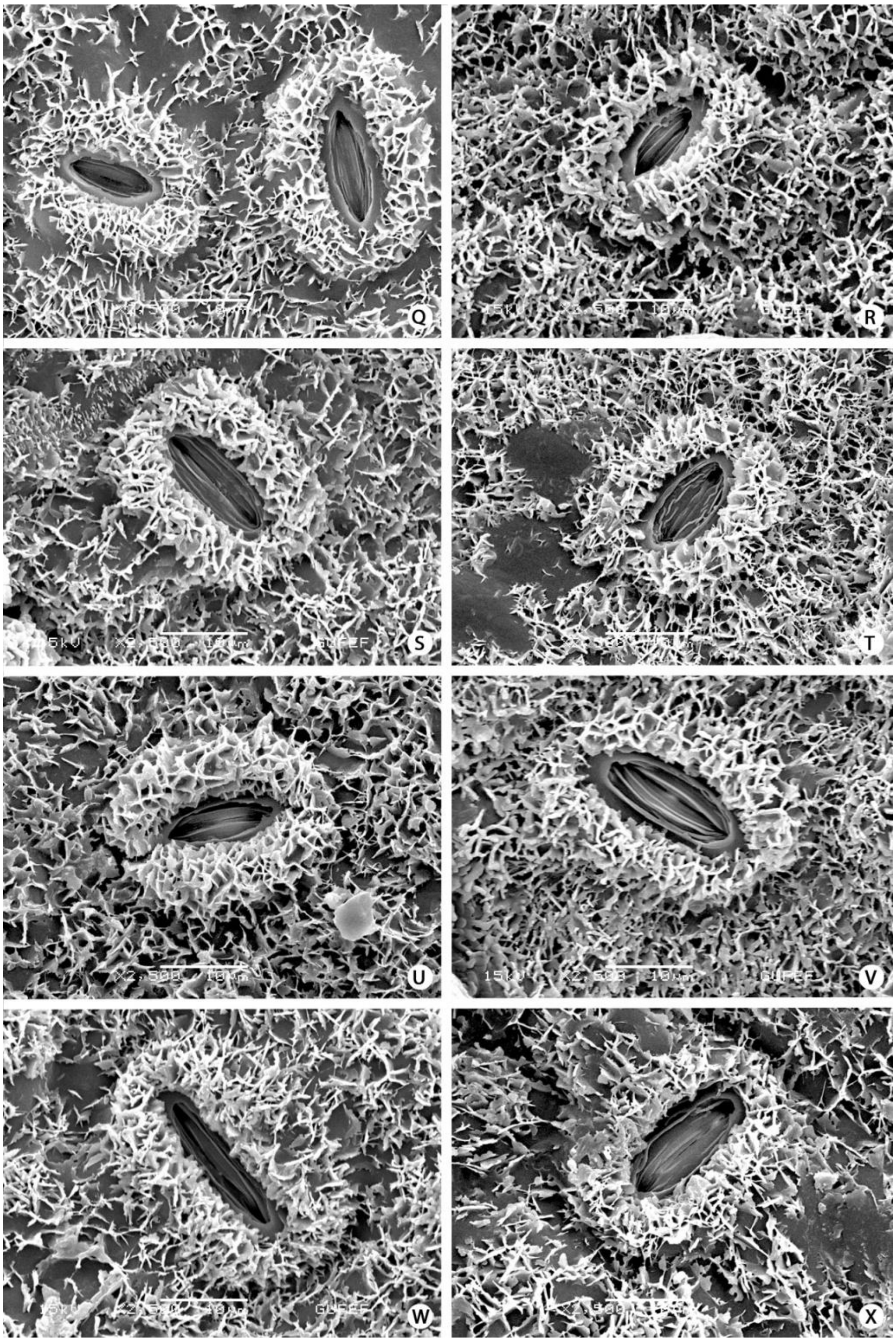


Figure 3.15 Micrographs of epicuticular wax and stoma. **Q:** c1 (m). **R:** b31 (mp). **S:** b54 (nid). **T:** b99 (nid). **U:** b51 (nid). **V:** b78 (nid). **W:** bx10 (nid). **X:** b82 (nid). SEM (X2500), Scale bar = 10µm.

CHAPTER 4

DISCUSSION

Leaf is one the most important morphological data source in plant taxonomy. Although floral parts are extensively used in classification of flowering plants, in particular groups such as *Quercus*, leaf features are more suitable source for identification and classification at species level (Hedge and Yaltirik, 1982; Jensen *et al*., 2002). In genus *Quercus*, floral features are illustrating little variation between species and do not have much uses in taxonomy. Acorn characteristics are useful particularly in the sectional delimitation or some subspecific categorization for Turkish oaks (Hedge and Yaltirik, 1982), but acorn bearing specimens is not available throughout the time or stands. The age of trees, crown types of the stand, the occurrence of the vegetative reproduction and year to year variations in acorn production (Goodrum *et al*., 1971), make it impossible to constitute a large-scaled morphometrical study solely on the acorn characteristics. Even though leaves may have similar weaknesses, they are more accessible and accurate samplings generally overcome such problems.

The posterior class probabilities obtained from random forest are generally consistent with PCA results with some exceptions. Exceptions may arise because of the difference between the PCA and the random forest classification. PCA plots may represent only some amount of variations and mostly the main trends in the multivariate data set. On the other hand, random forest classification is a supervised learning technique where the result depends on the variable discriminating ability between groups in the training data set rather than the relationships between variables.

In this study 38 variables found to be informative in discriminating between these oak species. Roughly the first ten of those variables seem to be diagnostic for species differentiation. These variables were also highly loaded on the first two principle components which indicate that they also represent the main trends in the data set.

Phenetic classifications are generally consistent with the phylogenetic schemes derived from molecular data (see also Sneath and Sokal, 1973), but this is not the case where the phenotypes of an organisms effected by various types of developmental or ecological factors and local environmental adaptations. The posterior class probabilities obtained by random forest could reflect the genetic assignment of individuals, but they may not necessarily be consistent. It is better to be considered as only an effective way of summarizing and reflecting the underlying morphometric variations. For example, *Q. macranthera* subsp. *syspirensis* was not known from Elmadag (e) and also not detected during the field trips; however, individual e23 got the probability of 0.461 for *Q. macranthera* subsp. *syspirensis* and 0.53840 for *Q. pubescens*. In fact, based on other traits, such as trichomes, venation and twig properties, it is likely to be identified as *Q. pubescens*. Specimen e23 was collected from a valley bottom on the shady side of a creek so the ecological factors may cause this appearance. Morphological similarity based on leaf traits may be analogous between unrelated individuals without any genetic base. Since the role of the leaf is photosynthesis rather than reproduction; phenotypic plasticity and developmental instability might also increase the morphological variations within populations or even within individual trees (Ponton *et al*., 2004; Gonzalez-Rodriguez and Oyama, 2005).

There is considerably low morphological differentiation between the species in the present study, particularly some degree of overlap occurred between clusters. Traditionally, it was claimed that this was due to the introgressive hybridization between species which is also challenged the biological species concept. It was well reported that the different species found in sympatry have more similar cpDNA than have the members of same species inhabiting different locations (Whittermore and Schaal, 1991), and morphological traits show continuous geographical gradient in part independent of species (Gonzalez-Rodriguez and Oyama, 2005). However, it is possible to recognize at least some degree of polymodal distribution of morphometric variables where two or more oak species are present (Knops and Jensen, 1980; Kremer *et al.*, 2002; Gonzalez-Rodriguez and Oyama, 2005, Penaloza-Ramirez *et al*., 2010). The level of hybridization was controversial issue for oaks that it was found below %10 in some recent studies (Dupouey and Badeau, 1993; Valbuena-Carabana *et al*., 2007; Curtu *et al*., 2007). However, it was found relatively higher in others (Mir *et al*., 2009; Penaloza-Ramirez *et al.*, 2010). Although hybridizing which may brings out morphological or genetic intermediates seems evident in many cases between closely related species particularly in the sympatry or contact zones (Kelleher *et al*., 2005; Rubio de Casas *et al*., 2007; Penaloza-Ramirez *et al*., 2010), intermediate morphotypes are not necessarily hybrids (Curtu *et al*., 2007) or hybrids are not necessarily show intermediate morphology

(Mir *et al*., 2009). Indeed, in the present study, some previously identified putative hybrids were observed to be displaying slightly different morphometric patterns on PCA plot rather than being intermediates between putative parent groups.

This suggests that the lack of characterization of morphological attributes in addition to some degree of introgressive hybridization might be the explanation of the continuous variation between these species. Also, the size and the composition of the training data become more of an issue if a classifier is used to find discriminating variables. The smaller size of the training set does not represent enough variation and intermediate individuals are likely to appear when the classifier was used. These reasons can also explain the tendency of the agglomeration of unidentified individuals in the center of the PCA plot.

Recognized hybrids between *Q. macranthera* subsp. *syspirensis* and *Q. pubescens* b27, b36, b61, b83 and b96 had intermediate or slightly greater leaf sizes closer to *Q. macranthera* subsp. *syspirensis* than being intermediate between putative parents, while the number of intercalary veins were as higher as or much higher than the *Q. pubescens*. The number of intercalary veins is somewhat proportional to the leaf size, since increasing in leaf size may results in increasing in number of intercalary veins. The truth is these putative hybrids had relatively same number of intercalary veins with *Q. pubescens* individuals when leaf size was standardized by calculating the number of intercalary veins per lobe (IVLOB) variable. High positive loadings of size related variables on first principle component, and the high negative loadings of the variables NIV and IVLOB on second principle component explains why these putative hybrids fall outside both parent clusters on PCA plots. Individuals b7 and b90, which were identified as hybrids between these species were intermediate values for both leaf size and intercalary veins so they were fall in-between two putative parent clusters. The degrees of lobe deepness (variable LOBTAR) for hybrids between these species seems generally higher than the parents but the *Q. pubescens* individual e58 has much deeper lobes among all individuals in this study. Isoperimetric deficit is the deviation from isoperimetric equality. Isoperimetric equality is equals to one for a circle. Variable ISOP gets higher values for these putative hybrids because they have deeper lobes similar to *Q. pubescens* and increased leaf sizes as in the *Q. macranthera* subsp. *syspirensis* which together increase the deviation from a circle. Leaf size of the putative hybrids between *Q. macranthera* subsp. *syspirensis* and *Q. infectoria* subsp. *boissieri* were intermediate but much closer to the former species. Lobe numbers in these putative hybrids were less than *Q. macranthera* subsp. *syspirensis,* and both the number and order of lobes were more similar to *Q. infectoria* subsp. *boissieri*. The lamina base shape and angles of auricles (AURIR and AURIL) also have an increased state than the parents (Fig 4.1). This might be due to the hybridization, and the lamina base shape might have deformed in order to have longer leaf similar to *Q. macranthera* subsp. *syspirensis* with a less number of lobes as like in the typical *Q. infectoria* subsp. *boissieri* (Fig 4.2).

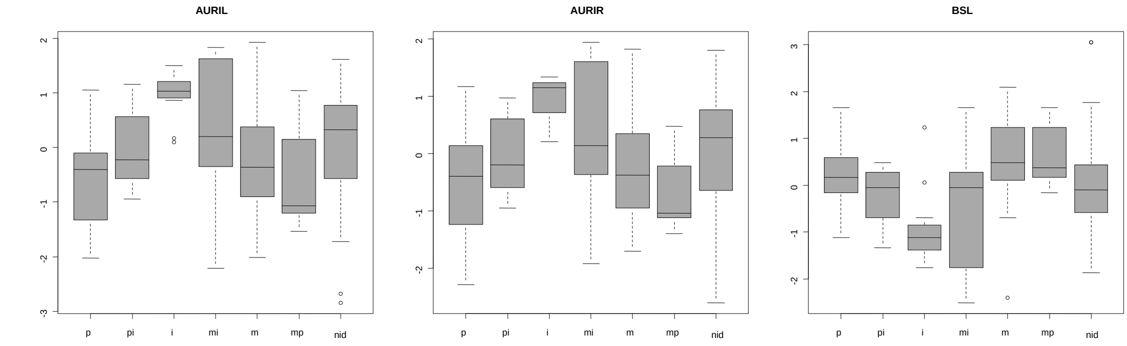


Figure 4.1 Additional boxplot representations for variable AURIL, AURIR and BSL (See Table 2.1 for group codes).

As expected, measurements on leaves are highly correlated with the overall leaf size. Much of the ratio variables are also correlated with those size variables since the numerator or denominator of these ratios are unequally affected from leaf size variation (see also Atchley *et al.*, 1976). In order to remove the size effects, the ratio could be taken between highly correlated variables. This minimizes the correlation between first principle component, so the size, and the ratios (e.g. PRL and LLWR). Although size is unwanted variation in shape analysis, size variables is traditionally used to distinguish between *Q. macranthera* subsp. *syspirensis* and *Q. pubescens* species (Hedge and Yaltirik, 1982) which explains the separation between species on PC1. The second PC has also some amount of size variation, but it was correlated with other diagnostic features used in differentiation of these species, such as the number of intercalary veins. It was described in Turkish Flora that, intercalary veins were absent or at most one or two for a typical *Q. macranthera* subsp. *syspirensis* individual (Hedge and Yaltirik, 1982). It was reported that, hybrids between *Q. macranthera* subsp. *syspirensis* and *Q. pubescens* species were common (Hedge and Yaltirik, 1982). Results support that the traditional taxonomic practices are sufficient to discriminate between these species, and separation can be done more precisely by including some calculated variables which were also used for detecting those putative hybrids between these species.

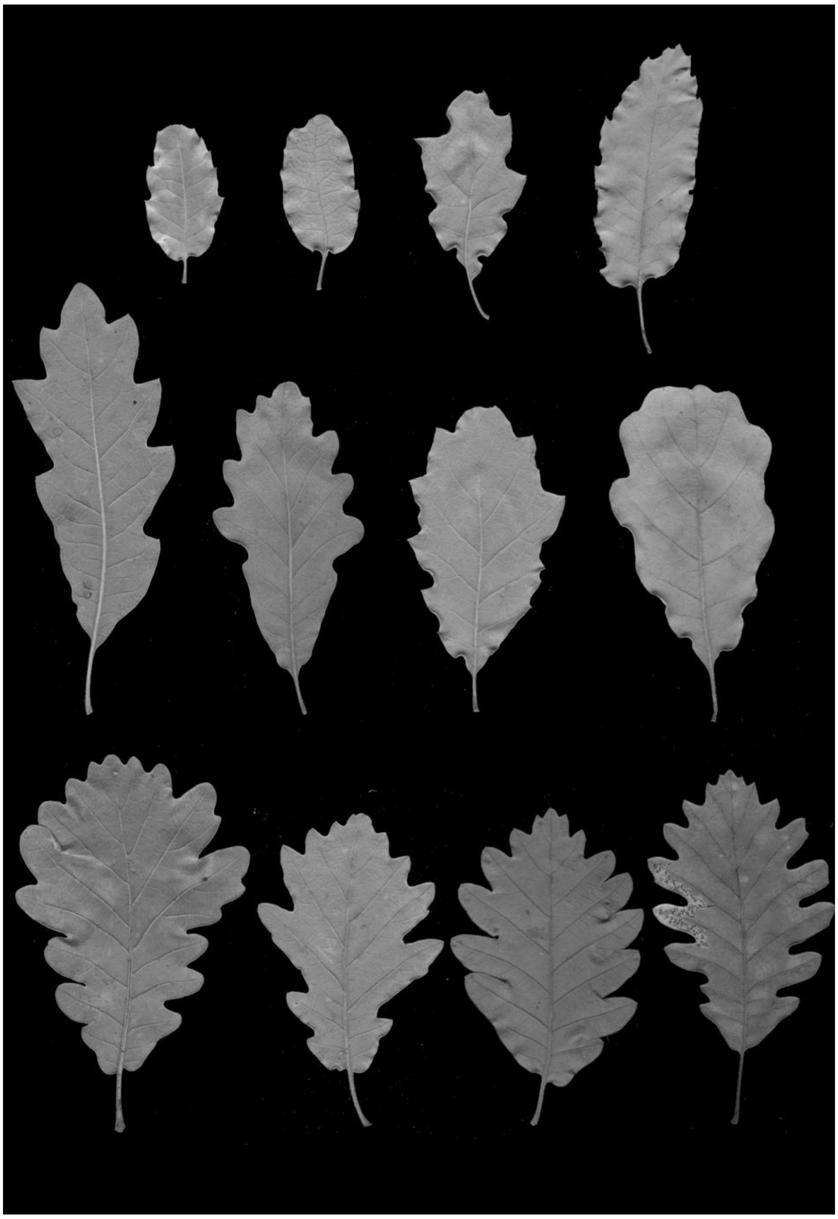


Figure 4.2 Leaf samples for visualizing leaf shapes. *Q. infectoria* subsp. *boissieri* leaves (upper ones, a4, b66, e30 and k16 from left to right), the putative hybrids *Q. macranthera* subsp. *syspirensis* X *Q. infectoria* subsp. *boissieri* leaves (middle ones, b24, b26, b52 and

bx6 from left to right) and *Q. macranthera* subsp. *syspirensis* leaves (below ones, b3, b11, b98 and bx12).

Two closely related species *Q. infectoria subsp. boissieri* and *Q. pubescens* represents a very problematic case in Turkish Flora. Although typical members of these species can be distinguished easily, there is no common diagnostic trait found to differentiate these species. The problem becomes more apparent when these species found together in a sympatric population. Typical *Q. infectoria subsp. boissieri* have variable leaves, lobed to entire, where the lobes generally have teeth at their tips and the lamina base shape is oblique to cuneate. Individuals found in Kemer (specimen k16 and k19) were the typical members of *Q. infectoria subsp. boissieri*, where the only white oak species in that area is *Q. infectoria subsp. boissieri*. Clearly identified *Q. infectoria subsp. boissieri* specimens were also collected in Alacaatli (a4) and Beynam (b65 and b66). Linear measurements and other quantitative practices related to multivariate morphometrics have reduced ability to differentiate between different shapes but leaf shape variables such as the LOBTPS, LWS and LOBTAR, lamina base angle variables AURIL and AURIR, and again, number of intercalary veins (NIV) separated at least the typical (here the training set) *Q. infectoria subsp. boissieri* and *Q. pubescens* specimens. However clusters of these species were not far from each other and their putative hybrids were scattered mostly within those clusters, supports the affinity between these species at least in leaf morphology (Schwarz, 1993).

Micromorphological features are useful in oak taxonomy particularly at specific and supra-specific levels (Uzunova *et al.*, 1997; Bussotti and Grossini, 1997). This was also supported in this study especially for characterization of the *Q. macranthera* subsp. *syspirensis*. The trichomes for its putative hybrids with other species generally have high number and longer rays than both parents. Bulbous trichomes and glands give the yellowish-brown color of abaxial leaf surface in the *Q. macranthera* subsp. *syspirensis*. Bulbous trichomes are present in b68 and bx5 unlike other typical *Q. pubescens* specimens which was supported by streomicroscobical investigations. It is interesting that these specimens were also closer to the *Q. macranthera* subsp. *syspirensis* on PCA plots, where they may have a hybrid ancestry. Number of glandular trichomes (bulbous trichomes are glandular trichomes) were accepted as species-specific marker to differentiate *Q. pubescens* from *Q. petraea* in Bruschi *et al.*, (2000) where the mean value of the number of glandular trichomes were found 9.75 for *Q.*

pubescens and 94.10 for *Q. petraea* on the 10mm^2 abaxial lamina surface. This result was supported in this study for *Q. pubescens* species.

Introgressive hybridization within this isolated population may be the reason for lack of differentiation based on micromorphological features particularly the wax properties. But it must be considered that earlier studies (Uzunova *et al.*, 1997; Bussotti and Grossini, 1997) were done with less number of individuals which were collected from distinct locations. This may be eliminating the effects of ecological conditions (and also possible hybridizations) which may be influential on the formation of micromorphological features, especially the wax characteristics.

Certainly, the next step in taxonomy after the development of phenetics will be the automated taxon identification. It has been a dream among systematists for a long time (Macleaod, 2007). The first step of automated identification, if it is based on the morphological features of the objects, is image acquisition. Feature extraction is the next and the most important step where the computers need much of human assistance today. Here, in this study, the features were selected manually by designating landmarks and semilandmarks on leaf images. Automated measuring and analyzing procedures follow the feature extraction in this study.

Amazingly, after preparing the landmarks and semilandmarks coordinate data matrix, and entering counted and descriptive variables to a table, final data matrix for 52 variables were calculated approximately only in two minutes. Preparing leaf samples and designating landmarks was the only time consuming part of this study. Certainly, preparation is expected to be time-consuming in any kind of morphometric study and measuring of at least 19 distance and angles variables in 1390 leaves, takes long time with rulers and calipers or even with the measuring tools of some computer programs. Moreover, some measurements such as PW, LOBWU and LOBWD would be difficult to measure out manually. In addition, human factors such as carelessness and tiredness were also minimized in the measuring procedure presented here. Previously written R codes could be use in further studies on new data sets, and for future works, automated calculations opens the opportunity of saving time for modifying the study and/or landmark designation instead of measuring variables from the start.

CHAPTER 5

CONCLUSION

Here in this study, several leaf morphological traits were analyzed to understand the variations within isolated oak stand. Some degrees of separation were seen between the species where the morphometric variables show three-modal distribution on PCA plots. First two PC's accounts roughly the 60% of the total variation. Thirty-eight of 52 variables had discriminating power between the studied oak species and about 10 of them were more informative and might be assumed as diagnostic variables for species differentiation.

Findings of this study are generally consistent with the work of Hedge and Yaltirik (1982) where the leaf size and the abaxial lamina hairs differentiate *Q. macranthera* subsp. *syspirensis* from others. *Q. infectoria subsp. boissieri* and *Q. pubescens* are highly related species but leaf shape variables provide good characterization of the former specimens.

Typical *Q. pubescens* in this study has no glandular-bulbous trichomes on the abaxial lamina and *Q. macranthera* subsp. *syspirensis* has typical trichomes with 2 and/or 4 rays. Although these micromorphological variables have taxonomical value, one should be careful about the generalization of these results.

Taxonomic status and discrimination problems between oak species, particularly in the section *Quercus* is still a big challenge for taxonomists. In this study only 48 specimens out of 139 were readily identified. Supervised learning methods, such as random forest classifier is a useful tool for taxonomists to identify unknown specimens by use of available information without thinking the statistical assumptions as in the parametric methods. Automated measurements and calculations, which were practiced in the current study, have an advantage of time saving, repeatability and modifiability for future works. These methods give the opportunity to the taxonomist to give less stress on measuring procedure and identification before analysis.

REFERENCES

ADAMS, D. C, ROHLF, F. J, and SLICE, D. E., 2004. Geometric morphometrics: ten years of progress following the 'revolution.' *Ital J. Zool* 71:5–16.

AN, A., 2005. Classification Methods. In J. Wang (Ed.), Encyclopedia of Data Warehousing and Mining, Idea Group Inc., 144-149.

ATCHLEY, W. R., GASKINS, C. T., and ANDERSON, D., 1976. Statistical properties of ratios. I. Empirical results. *Systematic Zoology* 25:137–148.

AXELROD, D., 1983. Biogeography of Oaks in the Arcto-Tertiary Province, *Annals of the Missouri Botanical Garden,* 70: 629-657.

BACILIERI R., DUCOUSSO A., PETIT R.J, and KREMER A., 1996. Mating system and asymmetric hybridization in a mixed stand of European oaks, *Evolution* 50: 900–908.

BELLAROSA, R., SIMEONE, M.C., PAPINI, A., and SCHIRONE, B., 2005. Utility of ITS sequence data for phylogenetic reconctruction of Italian *Quercus* spp. *Mol Phyl Evol* 34: 355–370.

BOOKSTEIN, F. L., 1991. Morphometric tools for landmark data. Geometry and biology. Cambridge University Press: New York.

BREIMAN, L., 1998. Arcing Classifiers. *The Annals of Statistics* vol. 26, no. 3, pp. 801-849

BREIMAN, L., 2001. Random Forests. *Machine Learning* 45 (1): 5–32.

BREIMAN, L., and CUTLER, A., 2004. Random forests (online manual). Version 5.0. http://www.stat.berkeley.edu/users/breiman/RandomForests/.

BRUSCHI, P., VENDRAMIN, G. G., BUSSOTTI, F. and GROSSONI, P., 2000. Morphological and Molecular differentiation between *Quercus petrea* (Matt.) Liebl. and *Quercus pubescens* Willd (Fagaceae) in Northern and Central Italy. *Annals of Botany* 85: 325–333.

BURGER, W.C., 1975. The species concept in *Quercus*. *Taxon* 24:45-50.

BUSSOTTI, F., and GROSSONI, P., 1997. European and Mediterranean Oaks (genus *Quercus* L.). SEM characterization of the micromorphology of the abaxial leaf surface waxes, stomata and trichomes. *Botanical Journal of the Linnaean Society* 124: 183-199.

CRON, G. V., K. BALKWILL, and KNOX, E. B., 2007. Multivariate analysis of morphological variation in *Cineraria deltoidea* (*Asteraceae, Senecioneae*). *Bot. J. Linn. Soc.* 154: 497–521.

CURTU, A.L., GAILING, and O., FINKELDEY, R. 2007. Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. BMC *Evolutionary Biology* 7: 218

DICKINSON, T. A., PARKER, W. H., and STRAUSS, R. E., 1987. Another approach to leaf shape comparisons. *Taxon* 36(1): 1-20.

DUPOUEY, J.L., and BADEAU, V. 1993. Morphological variability of oaks (*Quercus robur* L., *Quercus petraea* (Matt.) Liebl., *Quercus pubescens* Willd.) in northeastern France: preliminary results. *Annales des Sciences Forestières*, 50:35-40.

FELDESMAN, M.R., 2002. Classification Trees as an Alternative to Linear Discriminant Analysis. *American Journal of Physical Anthropology* 119:257–75.

GAGE, E., and WILKIN, P., 2008. A morphometric study of species delimitation in *Sternbergia lutea* (*Alliaceae*, *Amaryllidoideae*) and its allies *S. sicula* and *S. greuteriana*. *Botanical Journal of the Linnean Society*, 158, 460–469.

GEHRKE, J., 2005. Classification and Regression Trees. In J. Wang (Ed.), Encyclopedia of Data Warehousing and Mining, Idea Group Inc., 141-143.

GONZALEZ-RODRIGUEZ, A., and OYAMA, K. 2005. Leaf Morphometric Variation In *Quercus affinis* and *Q. laurina* (*Fagaceae*), Two Hybridizing Mexican Red Oaks. *Botanical Journal of the Linnean Society,* Vol 147(4): 427 - 435

GOODRUM, P.D., REID, V.H., and BOYD, C.E., 1971. Acorns yields, characteristics, and management criteria of oaks for wildlife. *Journal of Wildlife Management*. 35: 520-532.

HARDIN, J.W., 1976. Terminology and classification of *Quercus* trichomes. *J Elisha Mitchell Sci Soc* 92: 151–161.

HARDIN, J,W., 1979. Patterns of variation in foliar trichomes of Eastern North American *Quercus*. *American Journal of Botany* 66: 576–585.

HASTIE, T., TIBSHIRANI, R., and FRIEDMAN, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd edition). Springer-Verlag, New York.

HEDGE, I.C., and YALTIRIK, F., 1982. *Quercus* L. In: Davis PH, ed. Flora of Turkey and the East Aegean Islands, 7. Edinburgh: Edinburgh University Press, 659–683.

HENDERSON, A., 2006. Traditional morphometrics in plant systematics and its role in palm systematics. *Bot. J. Linn. Soc.* 151:103–111.

HOWARD, D. J., PRESZLER, R. W., WILLIAMS, J., FENCHEL, S. and BOECKLEN, W. J., 1997. How discrete are oak species? Insights from a hybrid zone between *Quercus grisen* and *Quercus gambelii*. *Evolution* 51:747-755.

JAMNICZKY, H.A., and RUSSELL, A.P., 2004. Cranial arterial foramen diameter in turtles: quantitative assessment of size-independent phylogenetic signal. *Animal Biology* 54(4): 417-436.

JENSEN, R. J., CIOFANI, K. M., and MIRAMONTES, L. C., 2002. Lines, outlines, and landmarks: morphometric analyses of leaves of *Acer rubrum*, *Acer saccharinum* (*Aceraceae*) and their hybrid.*Taxon* 51: 475-492.

JENSEN, R. J., 2003. The conundrum of morphometrics. *Taxon* 52:663–671.

LANDIS, F.C., GARGAS, A., and GIVNISH T.J., 2004. Relationships among arbuscular mycorrhizal fungi, vascular plants and environmental conditions in oak savannas. *New Phytologist* 164: 493–504.

LEMON, J., BOLKER, B., OOM, S., KLEIN, E., ROWLINGSON, B., WICKHAM, H.,*et al*. 2009. plotrix: Various plotting functions. R package version 2.6-1.

LEXER, C., KREMER, A, and PETIT, R.J., 2006. Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology,* 15: 2007–2012.

LIAW, A., and WIENER, M., 2002. Classification and Regression by randomForest. *R News* 2(3), 18-22.

KELLEHER , C.T., HODKINSON, T.R., DOUGLAS,G.C., and KELLY, D.L. 2005.
Species Distinction in Irish Populations of *Quercus petraea* and *Q. robur*: Morphological versus Molecular Analyses. *Ann Bot* 96: 1237-1246.

KIM, Y., and KIM, H., 2007. Application of Random Forests to Association Studies Using Mitochondrial Single Nucleotide Polymorphisms. *Genomics & Informatics* Vol. 5(4) 168-173

KNOPS, J.F., and JENSEN R.J. 1980. Morphological and Phenolic Variation in a Three Species Community of Red Oaks. *Bulletin of the Torrey Botanical Club*, Vol. 107, No. 3., pp. 418-428.

KOTSIANTIS, S. B., 2007. Supervised Machine Learning: A Review of Classification Techniques, *Informatica Journal* 31 249-268.

KREMER, A., DUPOUEY, J.L., DEANS, J.D., COTTRELL, J., CSAIKL, U., FINKELDEY, R., *et al*., 2002. Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annals of Forest Science* 59:777–787.

MACLEOD, N. (ed), 2007. Automated Taxon Identification in Systematics: Theory, Approaches and Applications. Systematics Association Special Volume, 74. Boca Raton: Taylor & Francis.

MANOS, P.S., DOYLE, J.J. and NIXON, K.C., 1999. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (*Fagaceae*). *Molecular Phylogenetics and Evolution*, 12: 333-349.

MARCUS, L. F., 1990. Traditional morphometrics. In F. J. Rohlf and Bookstein, F. L. eds. Proceedings of the Michigan morphometrics workshop. Special publication no. 2. Univ. of Michigan Museum of Zoology, Ann Arbor, ML.

METCALFE, C.R., and CHALK, L., 1979. Anatomy of the Dicotyledons. Volume 1 – 2nd Edition. *Clarendon Press*: Oxford.

MIR, C., JARNE, P., SARDA, V., BONIN, A**.**, and LUMARET, R., 2009. Contrasted nuclear and cytoplasmic reciprocal exchanges between two phylogenetically distant oak species (*Quercus suber* L. and *Q. ilex* L.) in Southern France. *Plant Biology* 11. 213–226.

MUIR, G., and SCHLÖTTERER, C., 2005. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology,* 14: 549–561.

NIXON, K.C., 1993. Infrageneric classification of *Quercus* (*Fagaceae*) and typification of sectional names. *Ann. Sci. Forest*. 50 (Suppl. 1): 25s–34.

NIXON, K.C., 2006. Global and Neotropical Distribution and Diversity of Oak (genus *Quercus*) and Oak Forests. *Ecological Studies,* 185 : 3-16.

PALMER, E. J., 1948. Hybrid oaks of North America. J.Arnold Arbor Harv. Univ. 29: 1–48.

PARSONS, K. J., ROBINSON, B.W., and HRBEK T., 2003. Getting into shape: An empirical comparison of traditional truss-based morphometric methods with a newer geometric method applied to New World cichlids. *Environmental Biology of Fishes* 67: 417–431.

PENALOZA-RAMIREZ, J. M., GONZALEZ-RODRIGUEZ, A. , MENDOZA-CUENCA, L., CARON, H., KREMER, A., and OYAMA K. 2010. Interspecific gene flow in a multispecies oak hybrid zone in the Sierra Tarahumara of Mexico. *Ann. Bot.*, March 1, ; 105(3): 389 - 399.

PONTON, S., DUPOUEY, J.L., and DREYER, E., 2004. Leaf morphology as species indicator in seedlings of *Quercus robur* L. and *Q. petraea* **(**Matt.**)** Liebl: modulation by irradiance and growth flush. *Annals of Forest Science* 61:73–80.

R DEVELOPMENT CORE TEAM, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

ROHLF, E J., and BOOKSTEIN, E. L., 1990. Proceedings of the Michigan morphometries workshop. University of Michigan Museum of Zoology Special Publication No.2. Ann Arbor.

ROHLF, F. J., 2005. tpsDig, digitize landmarks and outlines, version 2.05. Department of Ecology and Evolution, State University of New York at Stony Brook.

RUBIO DE CASAS, R., CANO, E., BALAGUER, L., PÉREZ-CORONA, E., MANRIQUE, E., GARCÍA-VERDUGO, C. and VARGAS, P. 2007. Taxonomic identity of *Quercus* *coccifera* L. in the Iberian Peninsula is maintained in spite of widespread hybridisation, as revealed by morphological, ISSR and ITS sequence data. *Flora* 202: 488-499

RUSHTON, B. S., 1993. Natural hybridization within the genus *Quercus* L. *Ann. Sci. For.* 50:73-90.

SCARELI-SANTOS, C., HERRERA-ARROYO, M.L., MONDRAGÓN, M.L.S., GONZÁLEZ-RODRÍGUEZ, A., BACON, J. and OYAMA, K., 2007. Comparative analysis of micromorphological characters in two distantly related Mexican oaks, *Quercus conzattii* and *Q. eduardii* (*Fagaceae*), and their hybrids. *Brittonia* 59(1), 37-78**.**

SCHWARZ, O. 1993. *Quercus* L. In: Tutin TG, Burges NA, Chater AO, Edmondson JR, Heywood VH, Moore DM, Valentine DH, Walters SM, Webb DA, eds. Flora europaea, Vol. I. 2nd edn. Cambridge: Cambridge University Press.

SNEATH, P. H. A, and SOKAL, R.R., 1973. Numerical taxonomy. San Francisco: W.H. Freeman.

SVETNIK, V., LIAW, A.,TONG, C., CULBERSON, J. C., SHERIDAN, R. P., and FEUSTON, B. P., 2003. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. J. Chem. Inf. Comput. Sci., 43, 1947-1958.

TOVAR-SÁNCHEZ, E., and OYAMA, K., 2006. Effect of hybridization of the *Quercus crassifolia - Quercus crassipes* complex on the community structure of endophagous insects. Oecologia 147: 702–713

UZUNOVA, K., PALAMAREV, E., and EHRENDORFER, F., 1997: Anatomical changes and evolutionary trends in the foliar epidermis of extant and fossil Euro-Mediterranean oaks (*Fagaceae*). *Pl. Syst. Evol.* 204: 141-159.

VALBUENA-CARABAÑA, M., GONZÁLEZ-MARTÍNEZ, S.C., HARDY, O.J., and GIL, L. 2007. Fine-scale spatial genetic structure in mixed oak stands with different levels of hybridization. *Molecular Ecology* 16:1207–1219.

WHITTEMORE, A. T., and SCHAAL, B., 1991. Interspecific gene flow in sympatric oaks. *Proc. Natl. Acad. Sci*. USA 88:2540-2544.

WORTH, A.P., and CRONIN, M.T.D., 2003. The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. *Theochem*, 622, 97-111.

WU, B., 2003. Comparison of statistical methods forclassification of ovarian cancer using mass spectrometry data. Bioinformatics, 19, 1636-1643.

YALTIRIK, F., 1984. Turkiye meseleri. Yenilik Basimevi. – Istanbul.

ZELDITCH, M. L., D. L. SWIDERSKI, H. D. SHEETS, and W. L. FINK., 2004. Geometric Morphometrics for biologists: a primer. *Elsevier Academic Press*: London. 443 pp.

ZIELIOSKI, J., PETROVA, A. and TOMASZEWSKI, D., 2006. *Quercus trojana* subsp. *yaltirikii* (*Fagaceae*), a new subspecies from southern Turkey. *Willdenowia* 36: 845-849.