FOREST GENETICS 8(2):151-158, 2001

NATURAL INTERSPECIFIC HYBRIDIZATION BETWEEN *PINUS BRUTIA* (TEN.) AND *PINUS HALEPENSIS* (MILL.), VERIFIED BY USING THE LOGISTIC REGRESSION MODELING ON MORPHOLOGICAL CHARACTERS

Katerina D. Dounavi[1], Nikos Koutsias[2] & Konstantinos P. Panetsos[3]

[1)] Niedersächsische Forstliche Versuchsanstalt (Abt. C), Forstamtstr. 6, D-34355 Staufenberg-Escherode, Deutschland, phone: 0049-5543-940854, fax: 0049-5543-940861, e-mail: Katerina Dounavi <nfv-abtc@t-online.de>
[2)] Lab. of Forest Management and Remote Sensing, Department of Forestry and Natural Environment, P. O. Box 248, Aristotelian University, GR-54006 Thessaloniki, Greece
[3)] Lab. of Forest Genetics and Breeding, Department of Forestry and Natural Environment, Aristotelian University, GR-54006 Thessaloniki, Greece

Received October 24, 2000; accepted April 3, 2001

ABSTRACT

The natural hybridization and introgression between *Pinus brutia* (Ten.), *Pinus halepensis* (Mill.) is of a great importance for the evolution of the Mediterranean forests, since they are characteristic tree species of the Mediterranean ecosystems. The two species in Greece were well isolated and they might have come in contact through artificial plantations or destruction of previously existing natural isolation zones. Logistic regression modeling was applied as an alternative classification method for the identification of populations of the two species and their putative natural hybrids on Rhodes island (Greece). The discrimination problem was defined as the relation of a qualitative dependent variable, which takes only two values to one or more independent explanatory variables. The dependent variable was species identity (*P. brutia* or *P. halepensis*), while the independent variables were the morphological characters measured. The best performed models were applied for all the sampled populations to identify the putative natural hybrids. These models consisted of the explanatory variables (a) number of resin canals on the ventral face of the needle, and (b) combination of the characters peduncle length + cone width, with an overall accuracy of 98.08 % and 98.11 %, respectively.

Key words: conifers, hybridization, introgression, logistic regression, morphological markers, *Pinus*.

INTRODUCTION

Pinus brutia Ten. and *Pinus halepensis* Mill. are two very common pine species of the Mediterranean forests. They are both very well adapted to the Mediterranean abiotic factors (climate, fire events), which exert a strong influence on the composition and structure of the Mediterranean-type ecosystems. Data obtained from artificial crosses, as well as from morphological and ecological studies lead to the conclusion that they are two independent distinct species (PANETSOS 1981; 1986).

The climatic changes at the end of Tertiary and the beginning of Quaternary lead to a differentiation of the common distribution of the two species (MIROV 1967; PANETSOS 1981). *P. brutia,* being more tolerant to cold than *P. halepensis,* occupied the eastern Mediterranean region and parts of the Middle East. Today it grows in north Iraq, Lebanon, Syria, Cyprus, Turkey, and on the Greek islands of the Aegean Sea, such as Thassos, Chios, Samos, Rhodes, Lesvos and Crete, as well as in Thrace (MIROV 1967; WEINSTEIN 1989). In addition, *P. halepensis* extends, in the Mediterranean region, from Spain eastwards to central Greek Macedonia (PANETSOS 1981).

Natural hybridization between *P. brutia* and *P. halepensis* has been reported in areas where the two species come in contact (PANETSOS 1975). The evidence of hybridization is of great importance for the evolution of the two species. PANETSOS (1975 1981) reported the occurrence of hybridization on the island Rhodes, where *P. halepensis* was artificially introduced some hundreds years ago in the autochthonous *P. brutia* population. In this case, hybridization and introgression can spread out fast if no other barriers preventing hybridization, besides spatial isolation, are operating. Univariate and multivariate statistical analysis based on discriminant estimators have already been applied for the identification of the two species and their hybrids (PANETSOS 1975; PANETSOS *et al.* 1997). The identification of *P. brutia, P. halepensis* and their putative hybrids with logistic regression modeling validate its use in biology and specifically in the taxonomy of biological species.

Various applications of the logistic regression modeling as a classification method on different fields can be found in the literature, like the application on medicine (ELY *et al.* 1996; GALLO 1996; HOSMER *et al.* 1997), on social sciences (DATTALO 1994) and other fields. To our knowledge this method has not yet been applied for the taxonomic identification of biological species. The main targets of the study were: a) identification of the morphological characters, that could be used to identify the species, and b) the examination of the suitability of such a statistical analysis for the identification of groups of individuals which, based on their high morphological similarity, are considered to belong to the same species.

MATERIAL AND METHODS

Sampling

Mature cones and foliage samples were collected in the spring time from 120 mother trees of three populations on Rhodes island for the investigation of hybridization between *P. brutia* and *P. halepensis*. The native species on this island is *P. brutia*, while *P. halepensis* was planted about 100 years ago. The populations, from which the parental species were sampled, were considered to be typical *P. brutia* and *P. halepensis* populations. The intermediate zone includes "putative hybrids" of the F_1 generation as well as introgressive forms and backcrosses and even *P. brutia* and *P. halepensis* trees. The location of the sampled populations are shown in Figure 1. Within the population, mother trees were sampled randomly.

Material

Five vigorous branches were collected from different parts of each mother tree, bearing at least 10 closed mature cones of the current year, as well as 5 branch shoots. The morphological characters of 10 closed, healthy cones per mother tree measured, were (a) cone length, (b) cone width, (c) peduncle length, (d) angle between the axis of the cone and the branch bearing it. The morphological characters of 10 seeds taken from 5 cones of each mother tree and measured using a micrometer were (a) seed length, (b) seed width, (c) seed thickness, (d) (seed + wing) length.

Concerning the branch shoots, the sheath length of 10 fascicles per mother tree was measured. From each fascicle, one needle was taken and the morphological and anatomical characters measured were: (a) needle length, (b) needle width, (c) needle thickness, (d) number of teeth per cm, (e) number of resin canals, (f) number of rows of stomata on the dorsal surface, (g) number of rows of stomata on the ventral surface, (h) number of stomata per cm on the dorsal surface, (i) number of stomata per cm on the ventral surface. All morphological characters were measured at the widest point of

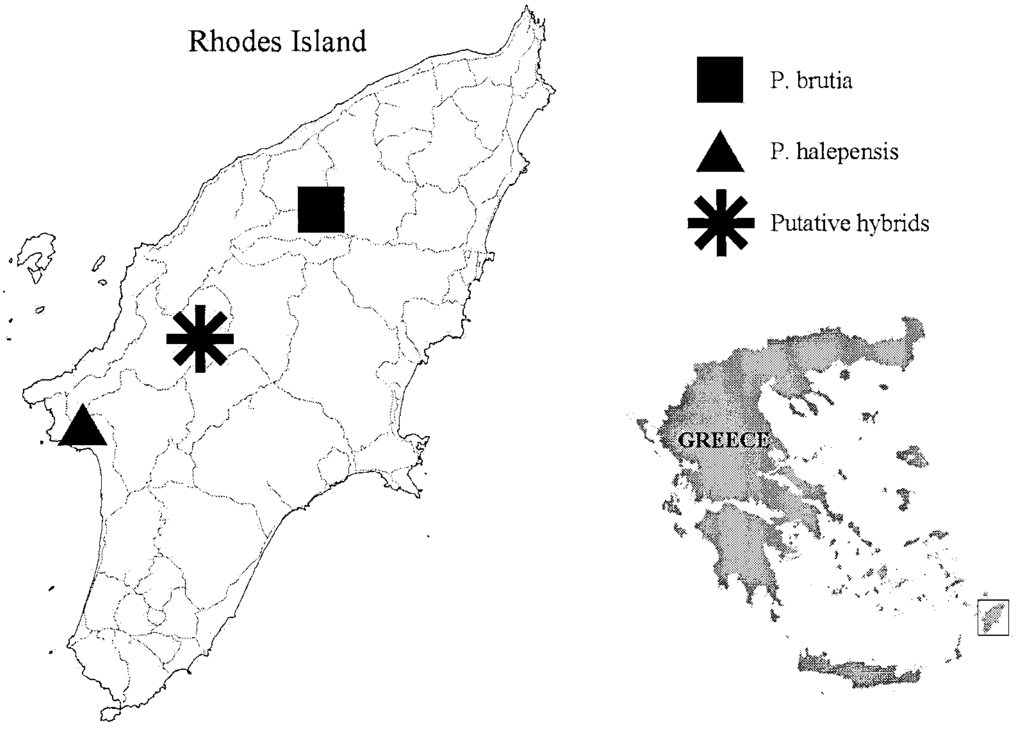

Figure 1. Sampled populations on Rhodes island, Greece.

Table 1. Description of the morphological characters used for the logistic regression analysis.

Description of the variables	Variables
Angle between the cone and the branch	*angle*
Small twig connecting the cone with the branch	*pdcle*
Cone length	*conel*
Cone width	*conew*
Needle length	*ndlel*
Needle width	*ndlew*
Needle thickness	*ndleth*
Length of the sheath that keep the needles connected in fascicles	*sthl*
Number of resin canals on the dorsal (upper) face of the needle	*rcu*
Number of resin canals on the ventral (down) face of the needle	*rcd*
Number of rows of stomata on the dorsal face of the needle	*stupr*
Number of rows of stomata on the ventral face of the needle	*stdwn*
Number of teeth per cm	*tcm*
Seed length	*sdl*
Seed width	*sdw*
Seed thickness	*sdth*
Length of the (seed + wing)	*sw*

the cone or the needle using a micrometer while the anatomical ones with the help of a microscope. All the characters tested with the logistic regression models are shown in Table 1.

Methodology

The species identification can be analyzed by logistic regression modeling, since the dependent variable can be expressed in a dichotomous way. This means that an individual tree may belong either to the *P. brutia* group or to the *P. halepensis* group. In order to answer this taxonomical question in a dichotomous way, a dummy or indicator variable E was created, which takes the value 1 if the respondent belongs to *P. brutia* group or the value 0 if it belongs to *P. halepensis*. The mean value of each character was used as a multivariate data set for the construction of a logistic regression model, in order to find the variables or the combination of them that best identify the two species.

A careful univariate analysis of each variable was the first step in the process of the variable selection. Univariate logistic regression models were constructed using eighteen morphological characters measured. The estimated potential differentiation of the two parental species offered by each morphological character was used to build up logistic regression models for the species identification. Moreover, models were developed by adding one or more variables to each univariate model.

The contribution of each independent variable to predict the dependent variable is difficult to be estimated in regression analyses, because the behavior of each variable depends on the others presented in the model (NORUSIS 1990). For that reason, eighteen logistic models were constructed, where in each of them only one morphological character was used as independent variable. This way provided a good estimation about the characters which best identify the two species and could be used as well to identify their putative hybrids. The most important variables were estimated by the Wald statistic, while the best fitted models were analyzed by the percent of correct classification, the -2log likelihood and the goodness of fit estimators.

Logistic regression models were constructed in order to: (1) estimate the discriminator ability of each morphological character and (2) classify all the individual trees of the study into one of the considered species. The samples used to construct the logistic regression models consisted of 25 individuals of the *P. brutia* group and 27 individuals of the *P. halepensis* group. Afterwards, the best performed characters were used as independent variables to develop a logistic model which was applied to the whole sampled data set, including also the putative hybrids, in order to classify all the individuals. The criterion used to classify an individual tree into one of the two parental species or into putative hybrids was the probability $E(y)$ estimated from each model.

The final evaluation of the specified logistic models and selection of the best performed ones were based on four criteria: (a) the final overall accuracy of the model (higher the accuracy better the performance of the model), (b) the distribution of the estimated probabilities of the response variable (more discrete groups means better the performance of the model), (c) the statistics estimated for each model, and (d) the number of morphological characters included in the model (less characters used means easier and quicker identification of the two species).

Logistic Regression Modeling

A variety of statistical multivariate techniques exists for the prediction of a dependent response variable from a set of independent ones or for the classification of individuals. These are for example, the discriminant analysis, multiple regression, etc. (NORUSIS 1990). However, their valid application presupposes a set of assumptions such as multivariate normal distributions, continuous variables etc. (AFIFI & CLARK 1991). In cases where either the multivariate normal model is not assumed or the set of the independent measurements consists of continuous and scalar variables, the logistic regression modeling can be applied (PRESS & WILSON 1979; AFIFI & CLARK 1990).

A logistic regression model estimates the probability of an event occurring and can be used as an alternative classification method. Actually, the logistic model was developed to be used in survival analysis where the dependent variable is dichotomous and can be expressed as 1 or 0 depending on whether the experimental question is true or false (MENDENHALL & SINCICH 1996). In the case of the species identification, logistic regression models can be constructed to estimate the probability, based on which an individual belongs to one or to another species and consequently, can be classified with the maximum probability as a member of the species. The independent variables are the morphological characters measured, for each sample unit. The development of the logistic regression models is achieved by:

- selecting the variables for the model,
- assessing the performance of the model, based on the individual variables and the overall fit of the model (HOSMER & LEMESHOW 1989).

Let the vector $\boldsymbol{x}' = (x_1, x_2, ... x_k)$ is k independent variables and $y \in \{0,1\}$ is a dichotomous dependent variable. The probability for $y = 1$, given the value of x, is assumed to be:

$$\mathrm{Prob}(y = 1) = \pi(x) \qquad [1]$$

and is defined from the model as:

$$\pi(x) = e^{g(x)} / (1 + e^{g(x)}) \qquad [2]$$

where $g(x) = a_0 + a_1 x_1 + a_2 x_2 + ... + a_k x_k$ is the logit transformation of the logistic regression model. Then according to HOSMER & LEMESHOW (1989) the probability for $y = 0$ is similarly assumed to be:

$$\mathrm{prob}(y = 0) = 1 - \pi(x) \qquad [3]$$

AFIFI & CLARK (1990) expressed the quantitative relationship between the dependent response variable and the independent explanatory variables according to logistic regression in another form:

$$\mathrm{prob}(y) = p = 1 / [1 + \exp(\hat{a} - \hat{\beta} x) \qquad [4]$$

where, $y = 1$ if category *A occurrs or* $y = 0$ if category N occurrs, $\hat{a}, \hat{\beta}$: the estimated coefficients, x: the independent variable.

Evaluation statistics

In order to fit the logistic regression model to a space, as it is defined in equation [4], the values of the unknown coefficients a and β should be estimated. Usually, the maximum likelihood method is used to estimate these unknown parameters. This method is based on a likelihood function, which expresses the probability of the observed data as a function of the unknown parameters and leads to values, the maximum likelihood estimators, for which the probability of obtaining the observed data set is maximized. The contribution of each pair (x_i, y_i) of the equations [1] and [2] to the likelihood function can be defined as:

$$\xi(x_i) = \pi(x)^{y_i} [1 - \pi(x)]^{1-y_i} \qquad [5]$$

The likelihood function, under the assumption that the variables are independent, can be formulated as the product of the $\xi(x_i)$:

$$l(A) = \Pi\, \xi(x_i) \qquad [6]$$

where, $A = (a, \beta)$ are estimates for the unknown parameters of the model. The log transformation of the likelihood estimator is often preferred because it is mathematically easier to work with (HOSMER & LEMESHOW 1989). The log-likelihood estimator is expressed as:

$$L(A) = \ln[l(A)] = \Sigma\{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \qquad [7]$$

The principle of the maximum likelihood method in this case is to find the value of A, A_{max}, that maximize the expression in equation [7]. The statistic $-2L(A)$ is estimated from software packages and is called *log likelihood statistic*. Actually, it tests the null hypothesis that the regression coefficients of the model are zero. A significant p value provides evidence that at least one of the coefficients for the explanatory variable is nonzero. The log likelihood statistic is an estimator of how good the model fits the data. The higher the value of $L(A)$ the better the model fits the data. In the case of the log likelihood estimator, the relation is opposite due to the multiplication by -2. Consequently, the best fitted models are the ones with the smallest values of $-2L(A)$.

The Wald statistic is an estimator, which can be used as an indicator for the significance of a variable in the model and is computed by dividing the estimated coefficient of interest (expressed with the symbol ˆ) $\hat{a}_1, \hat{a}_2, ... \hat{a}_k$ by its standard error:

$$W = \hat{a}_k / \text{s.e.}\ (\hat{a}_k) \qquad [8]$$

A critical value with a specific level of significance can be used as a measure for the evaluation of the variables as significant or not. This value is often the value 2, which would give a level of significance of 0.05 approximately (HOSMER & LEMESHOW 1989). The higher the value of the Wald statistic, the more important the variable for the model.

It is shown that the likelihood ratio statistic gives better results in small to moderate samples, while for studies of large samples the two estimates give similar results (KLEINBAUM 1994).

RESULTS

Assessment of the information content of each morphological character

The results of fitting the univariate logistic regression models to the data are summarized in Table 2. Based on the overall percent of correct classified observations, the best characters for the identification of both species are the angle and the rcd which resulted in an overall accuracy of 98.11 % consisted of 100 % of correct classified observations for *P. brutia* and 96.43 % for *P. halepensis*. Similarly, the second best characters are the tcm and sthl which provide an overall accuracy of 96.23 %. The tcm gave a 100 % correct classification for *P. brutia* and 92.86 % for *P. halepensis*, while the sthl classified 96 % correctly the *P. brutia* individuals and 96.43 % the *P. halepensis* individuals. The seed characters sdl and sdw seem to be important, since the percent of correct classification is high (96.15 %) and the other statistics indicate also a good fit of the models. The character conl provided the poorest classification ability for the individuals belonging to the *P. brutia* group (24 %) but with a medium identification ability for the individuals of the *P. halepensis* group (75 %). The character sw shows also a very low percent of correct classification for the identification of the trees belonging to both groups. All the other characters show a good discriminator ability for the identification of both species (83.02–96.15 %). Similar results, about the performance of each character used, are taken from the -2log likelihood, Wald statistic and the goodness of fit statistic provided as well in Table 2. The character rcd shows the smallest -2log likelihood value (7.433) and the character sdw has the smallest goodness of fit value, meaning that it has the smallest difference between the expected and the observed values for each species. The model constructed with the character Angle as independent variable does not fit good to the data, since it shows high values of -2log likelihood and goodness of fit statistics.

Development of Logistic Regression Models

The best fitted models were then chosen to be used for the identification of the sample individuals of two species and their putative hybrids. Best performed univariate model was considered to be the one with the

Table 2. Percent of correct classified individuals, –2log likelihood, Wald statistic and goodness of fit parameters estimating from the logistic regression models. Estimated coefficients for the best fitted models used for the identification of the sampled individuals and standard errors for these coefficients.

Variable	Percent of correct classified individuals (%)			Goodness of fit	Wald statistic	–2Log likelihood	B_0	S.E.
	P. brutia	*P. halepensis*	Overall					
rcd	100.00	96.43	98.11	9.870	6.536	7.433	–19.386	2.031
constant							6.701	1.111
pdcl	100.00	96.43	98.11	4.769	5.549	4.939	–29.977	4.223
conw							8.230	1.134
constant							– 5.794	0.345

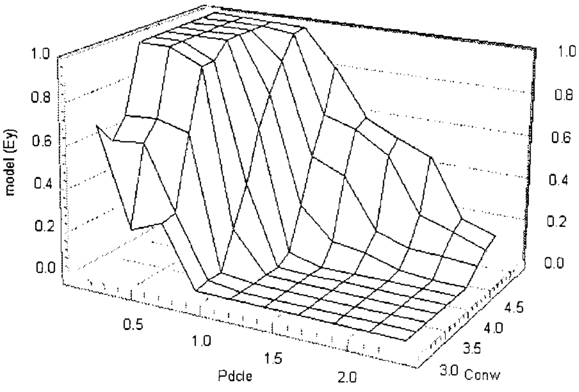

Figure 2. The hybrid index E(y) estimated by the model used the characters Peduncle length + Cone width.

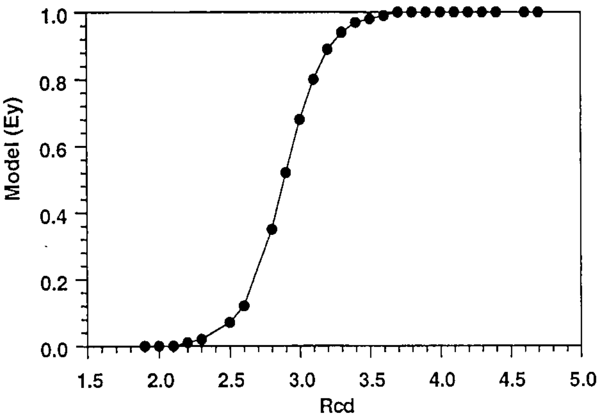

Figure 3. The hybrid index E(y) estimated by the model used the characters number of resin canals on the ventral face of needles.

rcd as independent variable, according to the percentage of correct classification, the other statistics estimated and the distribution of the estimated probabilities for each parental group of individuals sampled (Table 2). The second best performed one was considered the one with the sdw as independent variable. This model shows a lower percent of correct classification and a more scattered distribution of the probabilities but has the lowest value of goodness of fit, a relative low value of -2log likelihood estimator and a value >2 for the Wald statistic, meaning that this model seems to fit the data very well. The model with the character angle as independent variable was not used for the identification of the putative hybrids, because it seems to be not reliable, since it has high values of goodness of fit and -2log likelihood statistics.

All the models developed with two independent variables showed a good performance with the best one presented in Table 2. As the best two performed models were chosen the ones with explanatory variables the combinations *pdcle* + *conw* and *ndlel* + *sdw*.

The identification of the putative hybrids was performed with the models using rcd and a combination pdcle+conw as independent variables because of their most successful and reliable classification results.

Table 2 summarizes the estimated coefficients used to build each model to each population sampled in order to identify the putative hybrids.

Application of the models for the identification of hybrids

The two best performed models, which were developed using individuals of *P. brutia* and *P. halepensis*, were applied in all individuals sampled in the study area. The estimated probability E(y) from each of the two models was the criterion to classify an individual in one of the three considered groups: *P. brutia*, *P. halepensis* and putative hybrids. In the structure of the model the parental species *P. brutia* was coded with the value 1 and while *P. halepensis* was coded with the value 0. Thus it is clear that, if E(y) tends to the value 1, then the candidate individual was classified as *P. brutia*, while if E(y) tends to 0, then it was classified as *P. halepensis*. In the case of putative hybrids, since the values of their characters constitute an intermediate range in between the two parental species, it is obvious that if E(y) tends to value 0.5, then the candidate individual was classified as hybrid. In more details, when E(y) tends to the value 0.5, then it was considered as typical hybrid, while if E(y) is close to one of two extremes (1 or 0) of the model, then the individual was classified as hybrid with characters close either to *P. brutia* or to *P. halepensis*, respectively. The probability range of E(y) used to classify an individual tree into one of the two parental species or into the putative hybrid group was defined as following:

	>0.7	*P. brutia*
Prob(y)	{0.3–0.7}	Putative hybrids
	<0.3	*P. halepensis*

This classification takes into consideration that the two parental species show a considerable variation in their morphology and that the putative hybrid trees could show intermediate values, as well as a morphological variation according to a higher similarity to one of the two parental species.

All the individuals sampled as *P. brutia* are classified by all the models as *P. brutia*.

The identification of the trees sampled as *P. halepensis* gives different results. All the trees sampled are identified as *P. halepensis*, except one of them, which is classified from one model as *P. brutia* and from the other model as a putative hybrid.

The diagram showing the tendency of the *E*(y) values of the models for the characters used in the models could supply a hybrid index based on these characters (Figures 2, 3). The hybrid index would be the value $E(y)$ and would provide an interval for each character, where each species and the hybrids would be defined. These intervals, obtained using the number of resin canals on the ventral face of the needle, could be 1–2.5 for *P. halepensis*, >3 for *P. brutia* and 2.5–3 for the hybrids. The model using two characters provided intervals cone width 3–5 cm and peduncle length >0.8 cm for *P. halepensis*, cone width >4 cm and peduncle length 0–0.8 cm for *P. brutia* and cone width >3.5 cm and peduncle length >0.5 cm for the hybrids.

DISCUSSION

Information content provided from each morphological character

The two independent variables, that best identify *P. brutia* and *P. halepensis* are the number of resin canals on the ventral surface of the needle (*rcd*) and the seed width (*sdw*). The number of resin canals on the dorsal face of the needle (*rcu*) does not show a high discrimination ability, meaning that not only the number but also the position of the resin canals plays an important role. Similar results were reported from SCHÜTT and HATTEMER (1959), who mentioned that the number of resin canals is not a useful diagnostic character for species discrimination, while the location of the resin canals on the needle provides very useful information. The combination of the characters peduncle length and cone width developed the best performed model, having more than one independent explanatory variable. The high discriminator ability of these morphological characters is in accordance with the results published by PANETSOS (1975) who used a simpler statistical analysis for the discrimination of the same species based on morphological characters. The results of the Principal Component Analysis showed the importance of the seed characters and of all the morphological and anatomical characters of the needle for the identification of the two species and their putative hybrids (PANETSOS *et al.* 1997). These results are in agreement with those of the logistic regression model, concerning the information content of these characters. This is not the case for the characters cone length and length of seed + wing, which seem to be quite informative according to the results of the PCA but seem to provide the lowest information value of all the characters according to the results of the logistic regression model. Except of the results concerning the seed characters, those regarding the needle and the cone characters are similar with those reported from KOROL *et al.* (1995), who found that only the seed characters show no differences between the two species and their hybrids. The needle characters seem to be useful morphological traits in pine taxonomy. CALAMASSI *et al.* (1988) mentioned significant differences in each needle character measured between *P. brutia* and *P. eldarica* populations.

Application of the models for the identification of hybrids

It has to be mentioned, that these two logistic regression models classified some of the trees sampled as putative hybrids differently. These trees could be hybrids but they are more similar to one of the parents or these individuals are not hybrids. They were sampled in this group although they do not really belong to it. It is obvious, that the F_1 generation is well discriminated by these models. Furthermore, the models are also sensitive for the identification of individuals of the F_2 generation and backcrosses. The individuals that are classified from both the models in the same way as the parental species do not seem to be hybrids. These trees were probably missampling, indicating that parental species are also present in the intermediate region. The fact that hybridization and active introgression in two directions occur in the area where the two species come in contact could explain such results. On the other hand, misclassifications caused due to model errors could also be the reason of different results. In order to evaluate the performance of the method and to acquire an indication about the accuracy of the classification, the results were compared with those acquired from the multivariate analysis of the morphological characters and the isoenzymatic analysis of the same data set (PANETSOS *et al.* 1997). The study of PANETSOS *et al.* (1997) gave similar results. Specifically, the models which in the present study designated most of these individuals as one of the two species had as explanatory variables those, which according to PANETSOS *et al.* (1997), showed also mean values closer to one or the other parent species. In the same way, the models that classified the same trees as hybrids had as independent variables those, which according to the same work show intermediate mean values (*pdcle, ndlel*). Interestingly, the hierarchical tree produced by PANETSOS *et al.* (1997), by using the average linkage clustering method based on the unweighted pair group algorithm, shows a very high similarity in the clustering of the trees with the classification of the same trees from the logistic regression models.

Based on the effectiveness of the isoenzyme analysis in the study of the natural hybridization and the ability of the isoenzyme gene markers to elucidate the phenomenon of introgression (PANETSOS *et al.* 1997), a comparison between the results of the logistic regression and the isoenzyme analysis was made. The findings of the isoenzyme analysis corroborated for most of the individuals with the interpretation of the logistic regression results, while in some cases the logistic models are more sensitive in identifying hybrids. Specifically, some individuals that are identified from the isoenzyme gene markers to be *P. brutia* individuals, could be hybrids but they do not possess the marker alleles of *P. halepensis*. Such cases can be better identified by the use of logistic regression models.

In conclusion, it appears that the logistic regression analysis of morphological traits has been successfully applied in this study, as a method of taxonomic classification. It seems to be reliable for species identification and the designation of hybrids resulting from F_1 or advanced generations and backcrosses. Each logistic regression model allowed a clear separation between two pure species and a good identification of the hybrids, while the combination of the results of more than one model is required for studying introgression. Logistic regression modeling is a useful method in examining the best fitting and biologically meaningful model to describe relationships between a dependent variable which takes only two dichotomous values and a set of independent variables consisted of continuous quantitative or scalar qualitative variables. In our study, according to the results recorded, the contribution of this method to the identification of the two species and their natural hybrids was significant. Logistic regression modeling, applied on morphological characters, seems to be a valuable method for studying also the phenomenon and dynamics of introgession. The results of this study confirm that the application of logistic regression on morphological datasets can be very helpful in taxonomical studies, as well as in studies of evolutionary processes, such as interspecific hybridization and introgression. It is also evident that hybridization takes place on Rhodes island and more ecological studies have to be performed to identify the intensity of the phenomenon and the potential of hybrid genotypes to occupy novel habitats.

AKNOWLEDGEMENT

This work has been supported by EU funding under the AIR–CT93–0803 project. The authors thank the Mediterranean Agronomic Institute of Chania for the financial support and the offer of all the material, equipment software and hardware needed for this work.

REFERENCES

AFIFI, A. A. & CLARK, V. 1990: Computer-Aided Multivariate Analysis. Van Nostrand Reinhold Company, New York.

CALAMASSI, R., PUGLISI, S. R. & VENDRAMIN, G. G. 1988: Genetic variation in morphological and anatomical needle characteristics in *P. brutia* Ten. *Silvae Genetica* **37**: 199–206.

DATTALO, P. 1994: A comparison of discriminant analysis and logistic regression. *Journal of Social Service Research* **19**: 121–144.

ELY, J. W., DAWSON, J. D., MERH, D. R. & BURNS, T. L. 1996: Understanding logistic regression analysis through example. *Family Medicine* **28**: 134–140.

GALLO, J. J. 1996: Making sense of logistic regression. *Family Medicine* **28**: 141–143.

HOSMER, D. W. JR., HOSMER, T., LE CESSIE, S. & LEMESHOW, S. 1997: A comparison of goodness of fit tests for the logistic regression model. *Statistics in Medicine* **16**: 965–980.

HOSMER, D.W. JR. & LEMESHOW, S. 1989: Applied Logistic Regression. Wiley-Interscience Publication, New York.

KLEINBAUM, D. G. 1994: Logistic Regression-A Self-Learning Text: Springer-Verlag, New York, Inc.

KOROL, L., MADMONY, A., RIOV, V. & SCHILLER, G. 1995: *P. halepensis* x *P. brutia* subs. *P. brutia* hybrids? Identification using morphological and biochemical traits. *Silvae Genetica* **44**: 186–190.

MENDENHALL, W. & SINCICH, T. 1996: A Second Course in Statistics: Regression Analysis. Prentice-Hall, New Jersey.

MIROV, N. T. 1967: The Genus *Pinus*. Ronald Press, New York.

NORUSIS, M. J. 1990: SPSS/PC+Advanced Statistics^{TM} 4.0 for the IBM PC/XT/AT and PS/2., SPSS Inc.

PANETSOS, K. P. 1975: Natural hybridization between *P. halepensis* and *P. brutia* in Greece. *Silvae Genetica* **24**: 163–168.

PANETSOS, K. P. 1981: Monograph of *Pinus halepensis* (Mill.) and *Pinus brutia* (Ten.). *Ann. Forest.* **9**: 39–77.

PANETSOS, K. P. 1986: Genetics and breeding in the group halepensis. *Forêt mediterranéenne* **13**: 5–12.

PANETSOS, K. P., SCALTSOYIANNES, A., ARAVANOPOULOS, F. A., DOUNAVI, K. & DEMETRAKOPOULOS, A. 1997: Identification of *P. brutia* Ten., *P. halepensis* Mill. and their putative hybrids. *Silvae Genetica* **46**: 253–257.

PRESS, J. S. & WILSON, S. 1979: Choosing between logistic regression and discriminant analysis. The Rand Corporation, California.

SCHÜTT, P. & HATTEMER, H. H. 1959: Die Eignung von Merkmalen des Nadelquerschnits für Kiefern-Bastarddiagnose. *Silvae Genetica* **8**: 93–99.

WEINSTEIN A. 1989: Geographic variation and phenology of *P. halepensis, P. brutia* and *P. eldarica* in Israel. *Forest Ecology and Management* **27**: 99–108.