TESTING BAYESIAN ALGORITHMS TO DETECT GENETIC STRUCTURE IN TWO CLOSELY RELATED OAK TAXA

CRISTIAN MIHAI ENESCU[1,2*], NECULAE ŞOFLETEA[2], ALEXANDRU LUCIAN CURTU[2]

Keywords: structure, *Quercus*, *Lanuginosae*, oak

Abstract: The aim of this study was to test the Bayesian algorithm implemented in the software STRUCTURE in order to detect the number of clusters, by using microsatellite data from four oak species. Several assignment models, with or without a priori grouping of individuals to species, were proposed. Better results were obtained by using the sampling location information and when only two taxa were analyzed. Particularly, pedunculate oak and sessile oak formed distinct clusters whatever the assignment model we use. By contrast, no separation between the two oaks from series *Lanuginosae* was observed. This can be explained, on one hand, by the small sampling size for Italian oak, or by the genetic similarities of the two pubescent oaks, namely *Quercus pubescens* and *Q. virgiliana*, on the other hand. Our findings support the hypothesis according which Italian oak is an intraspecific taxonomic unit of pubescent oak.

INTRODUCTION

Nowadays, there are a set of software packages developed for delineating clusters of individuals on the basis of their genotypes by using a Bayesian algorithm. The software STRUCTURE (Pritchard et al, 2000) is one of the most used programs which can be applied to multilocus microsatellite data to define the number of clusters corresponding to number of sampled taxa. It can be also used to infer the population structure or to detect the hybrids, especially in species-rich genera, such genus *Quercus* L. (oaks).

It was used in the case of closely related European white oak species, such as *Quercus pubescens* and *Q. frainetto* (Curtu et al, 2011) or *Q. robur*, *Q. petraea*, *Q. pubescens* and *Q. pyrenaica* (Lepais et al, 2009) or the two endemic Californian oaks, namely *Q. lobata* and *Q. douglasii* (Craft et al, 2002).

The aim of this study was to test the Bayesian algorithm implemented in the software STRUCTURE version 2.3.3 (Pritchard et al, 2000) in order to detect the number of clusters (k), by using microsatellite data for four Romanian oak species, namely pubescent oak (*Q. pubescens*), Italian oak (*Q. virgiliana*), pedunculate oak (*Q. robur*) and sessile oak (*Q. petraea*).

MATERIAL AND METHODS

A total of 162 oak individuals were genotyped at seven microsatellite loci (Table 1). Among them, according to the Dendrological Romanian literature (Şofletea and Curtu, 2007), 61 trees were pubescent oak trees, 18 were Italian oak trees, 37 were pedunculate oaks and 46 were sessile oaks. Special attention was given to certain leaf and fruit descriptors, such as: lamina length, petiole length, basal shape of the lamina and length of the cupula peduncle. According to some recent studies (Enescu et al, 2011; Şofletea et al, 2011), these morphological characters proved to had the highest discriminating power between the Romanian oak species. While the oaks from series *Lanuginosae* were sampled from pure or mixed stands across Romania, the pedunculate oaks and the sessile oaks were sampled from two pure stands, namely Podul Iloaiei (Iaşi County, NE Romania) for *Q. robur* and Cristian (Braşov County, Central Romania) for *Q. petraea*, respectively.

DNA was extracted from winter buds using the Qiagen DNeasy 96 Plant Kit following the manufacturer protocol, but without liquid nitrogen (Toader et al, 2009). Then, the DNA was kept by -60º C until use. The seven genomic SSRs (gSSRs) were amplified using Polymerase Chain Reaction (PCR). The primers were combined into two PCR multiplexes on the basis of annealing temperature and fluorescent label. The first multiplexing reaction included four gSSRs (ssrQpZAG112, ssrQpZAG96, ssrQpZAG11 and ssrQpZAG110), while the second one only three (ssrQpZAG87, ssrQpZAG20 and ssrQpZAG7). More information about the seven microsatellite loci is given in Table 1. The reactions were performed in a 10 µl volume containing 1 µl template DNA (1:40…1:80; 2 µl DNA : 38 µl H_20… 2 µl DNA : 158 µl H_2O), 2 µl PCR Buffer 5x, 0.90 µl $MgCl_2$ (25 mM), 1 µl dNTPs (2mM) and 0.10 µl Promega *Taq* DNA polymerase (5 U/ µl). For primers concentrations see Table 1. Amplification was carried out in an Eppendorf Master Cycler. The PCR profile was as follows: 3 minutes of denaturation at 94º C followed by 30 cycles of 45 s denaturation at 94º C, a 35 s annealing step at 51º C, a 1 min 50 s elongation step at 69º C and a final extension step at 69º C for 15 min. The correct amplification of loci was checked by using 2 µl of PCR products mixed with 3 µL of Dye and migrated on 1.5 % agarose gels for 25 minutes at 100V. Amplification products were run on a Beckman Coulter Genetic Analyser

using Frag-3 method and Size Standard 400. The products were then analyzed using Fragment Analysis Software using default parameters and PA ver 1 dye correction.

Table 1. Characteristics of the seven microsatellite loci

Locus	Nucleotide motif	Linkage group (LG)	Beckman Dye	Primer concentration (uM)	Allele size (bp)
ssrQpZAG112	di	12	D4	0.20	82-112
ssrQpZAG96	di	10	D3	0.80	140-180
ssrQpZAG11	di	10	D3	0.60	242-289
ssrQpZAG110	di	8	D4	0.90	205-243
ssrQpZAG87	di	1	D3	0.55	103-183
ssrQpZAG7	di	2	D4	0.65	116-157
ssrQpZAG20	di	1	D3	0.80	159-213

Species assignments were evaluated using the program STRUCTURE version 2.3.3 and all possible combinations (with two, three or four taxa) were tested (see Table 2). In addition, an extra combination with all four taxa (162 oak trees), by changing the order of four individuals (the pubescent oak no. 5 changed its place with the pedunculate oak no. 95 and the pubescent oak no. 6 changed its place with the sessile oak no. 120, respectively) in the input file was also done.

The admixture model assuming correlated allele frequencies was used. In all cases, two model approaches have been used, namely with or without *a priori* grouping of individuals to species. Three runs were done for each case. In every run the length of *burnin* period was set to 50 000, while the number of MCMC iterations after *burnin* was 100 000. The number of clusters was estimated according to ΔK values (Evanno et al, 2005) by the aid of STRUCTURE HARVESTER software (Earl, 2011).

RESULTS AND DISCUSSIONS

Regarding the eleven possible combinations (Table 2), better results were obtained by using the sampling location information, on one hand, and when only two taxa were analyzed, on another hand. Particularly, pedunculate oak and sessile oak formed distinct clusters, whatever the assignment model was.

By contrast, no separation between the two oaks from series *Lanuginosae*, namely pubescent oak and Italian oak was observed in combinations comprising three or four taxa (cases 7, 8 and 11). Structure clustering results obtained for the latter three cases with *LocPrior* model (with *a priori* grouping of individuals to species) are illustrated in Figures 1, 2 and 3, respectively. Each individual is represented by a vertical bar partitioned into two or three color segments proportional to its membership in each genetic cluster.

Table 2. K values for the 11 combinations (with or without *a priori* grouping of oaks to species)

Case	Combination	K		Case	Combination	K	
		without	with			without	with
1	STP, ST	2	2	7	STP, STV, ST	**2**	**2**
2	STP, STV	**3**	2	8	STP, STV, PET	**2**	**2**

3	STP, PET	**3**	2	9	STV, ST, PET	**5**	3
4	STV, ST	2	2	10	STP, ST, PET	3	3
5	STV, PET	**3**	2	11	STP, STV, ST, PET	**3**	**3**
6	ST, PET	2	2				
Abbreviations: STP-*Q. pubescens*, STV-*Q. virgiliana*, ST-*Q. robur*, PET-*Q. petraea*							

If we take into consideration the hypothesis according which the pubescent oak and the Italian oak represent a solely morphological and genetic entity, by analyzing the 2-D 100% Stacked Column Graphs from Figures 1, 2 and 3 we can say that only a few individuals were wrongly assigned. Among them, were the individuals 38 and 54 (identified in the field as being pubescent oaks) and the tree number 106 (from Figures 1 and 3), which was considered a *Q. robur*-like individual according to its twig and leaf morphology.

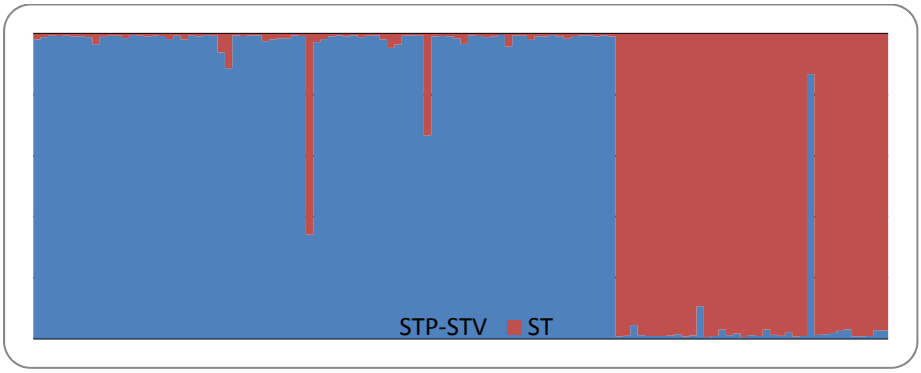

Figure 1. Structure clustering results (K=2) for case 7 (61 STP, 18 STV and 37 ST)

Interesting, regarding the same two pubescent oaks (individuals 38 and 54) different results were obtained in the case number 8, by assigning only the trees from series *Lanuginosae* and the sessile oaks. In this case, the memberships of the two pubescent oaks in the sessile oak cluster were less (Figure 2: 15% and 4%, respectively), compared with those from the pedunculate oak cluster from case 7 (Figure 1: 66% and 33%, respectively). Moreover, for the same two individuals lesser membership values were obtained in the case 11, when four taxa were analyzed (Figure 3).

By contrast, no significant differences were recorded for tree number 106. Its membership values to STP-STV cluster were 87% (case 7) and 84 % (case 11), respectively. Similar results were recorded also for pedunculate oak cluster (ST), namely 13% (case 7) and 14% (case 11), respectively.

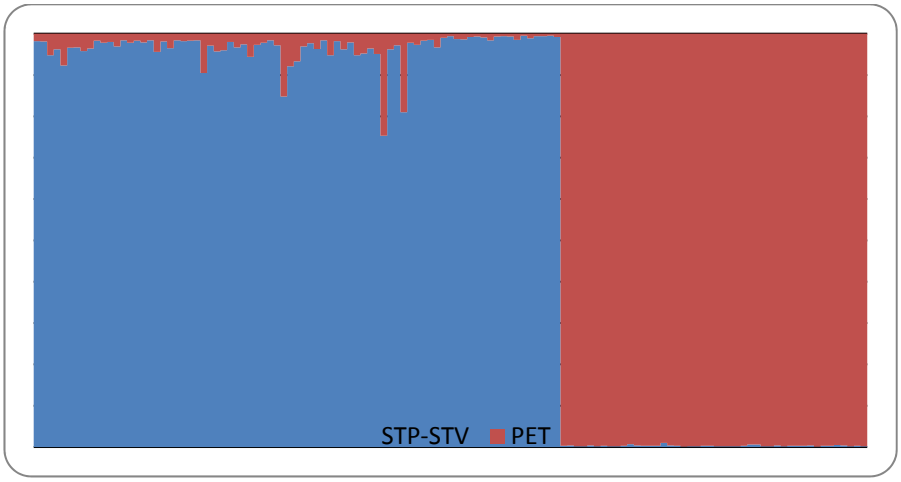

Figure 2. Structure clustering results (K=2) for case 8 (61 STP, 18 STV and 46 PET)

It can be seen from Figure 3 that most of the pubescent oak individuals (no. 1-61) present around 5 to 10 % membership to sessile oak cluster (PET) and only 1 to 3 % to pedunculate oak cluster (ST). Instead, the *Q. virgiliana*-like individuals (oaks no. 62-79) had around 99 % membership to STP-STV cluster.

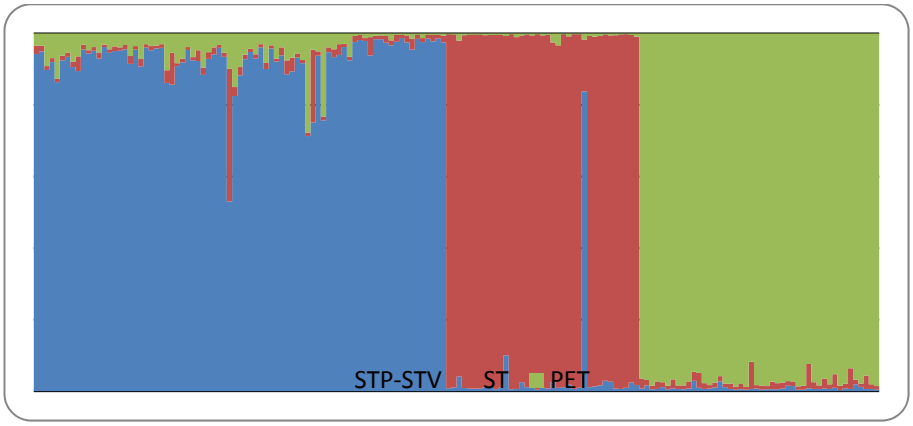

Figure 3. Structure clustering results (K=3) for case 11 (61 STP, 18 STV, 37 ST and 46 PET)

Regarding the extra case (Figure 4), the program identified all the four changes from the input file, being more evident for the first three cases, namely the oaks no. 5, 6 and 95, respectively.

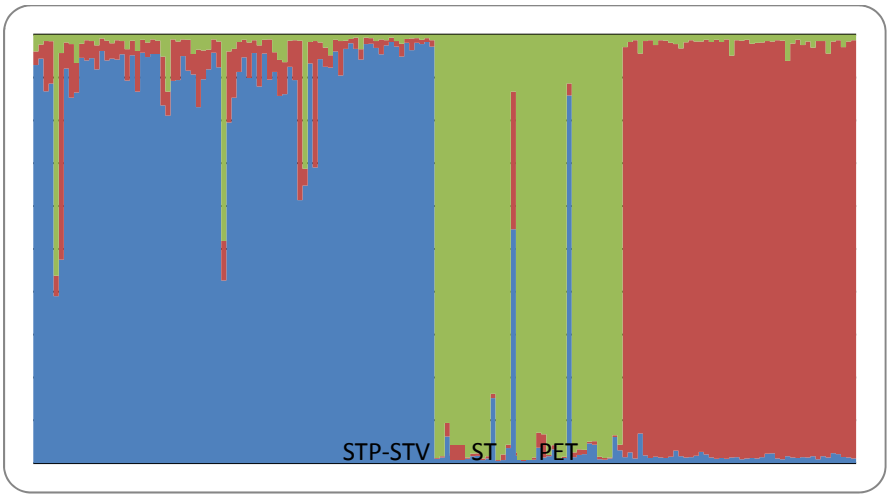

Figure 4. Structure clustering results (K=3) for the extra case

The different results obtained for the oaks from series *Lanuginosae* (cases 1-5, 7, 8 and 11) can be explained, on one hand, by the small sampling size for Italian oak, or by the genetic similarities of the two pubescent oaks, namely *Quercus pubescens* and *Q. virgiliana*, on the other hand. Nevertheless, if we take into consideration the Figure 3 we can say that the pubescent oak individuals shared a bigger part from their genome with the sessile oak, around 5-10%, compared with the pedunculate oak, with only 1-3 %.

By contrast, the distinct clusters formed by the pedunculate oak and sessile oak individuals indicated the existence of two different genetic entities. Same results were communicated by Moldovan (2011) for the two oak species in eastern Romania by using both microsatellite and cpDNA markers or by Neophytou and his colleagues (2010) by sampling individuals from three different European stands (Greece, Bulgaria and Germany).

CONCLUSIONS

Even if we used only seven microsatellite loci only a few individuals were wrongly classified (Figures 1-3). This could be explained, on one hand, by the fact that these individuals could be putative hybrids or, on another hand, by the DNA contamination. It resulted also that the only seven microsatellite loci were able to separate *Q. robur* from *Q. petraea* and the group pedunculate oak – sessile oak from the oaks belonging to series *Lanuginosae*.

Our findings support the hypothesis according which Italian oak is an intraspecific taxonomic unit of pubescent oak. This is in accordance with the results from a morphological survey according which the length of the cupula peduncle was the only descriptor which somehow discriminate the two taxa (Enescu et al, 2012).

It was proven that the software STRUCTURE was able to highlight the changed made in the input file. This could be helpful if someone wants, for example, to determine to which taxa a certain sample belongs. In other words, a bigger number of samples and microsatellite markers will increase the precision of the assignment.

REFERENCES

Craft, K.J., Ashley, M.V., Koenig, W.D., (2002). *Limited hybridization between Quercus lobata and Quercus douglasii (Fagaceae) in a mixed stand in Central Costal California*. Am J Bot, 89(11), 1792-1798.
Curtu, A.L., Moldovan, I.C., Enescu, C.M., Crăciunesc, I., Şofletea, N., (2011). *Genetic Differentiation between Quercus frainetto Ten. and Q. pubescens Willd. in Romania*. Not Bot Horti Agrobo, 39(1), 275-282.
Earl, D.A., von Holdt, B.M., (2011). *STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method.* Cons Gen Res, DOI: 10.1007/s12686-011-9548-7.
Enescu, C.M., Şofletea, N., Curtu, A.L., (2011). *Leaf multivariate analysis in four Romanian oak species*, 57-66. Proceedings of the 4th International Conference of Young Scientists, Prague, Czech Republic.
Enescu, C.M., Şofletea, N., Curtu, A.L., (2012). *Cluster Analysis in Pubescent Oak Taxa from Series Lanuginosae: A Case Study*, Bulletin of the *Transilvania* University of Braşov, Series II 54(1), 79-84.
Evanno, G., Regnaut, S., Goudet, J., (2005). *Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study*. Mol Ecol, 14, 2611-2620.
Lepais, O., Petit, R.J., Guichoux, E., Lavabre, et al., (2009). *Species relative abundance and direction of introgression in oaks*. Mol Ecol, 18, 2228-2242.
Moldovan, I.C. 2011. PhD Thesis, *Transilvania* University of Braşov. *Evaluarea variabilităţii genetice a stejarului pedunculat (Quercus robur L.) şi a gorunului [Quercus petraea (Matt.) Liebl.] în populaţii din estul României cu ajutorul markerilor ADN*.
Neophytou, C., Aravanopoulos, F.A., Fink, S., Dounavi, A., (2010). *Detecting interspecific and geographic differentiation patterns in two interfertile oak species (Quercus petraea (Matt.) Liebl. and Q. robur L.) using small sets of microsatellite markers*. For Ecol Manage, 259(10), 2026-2035.
Pritchard, J.K., Stephens, M., Donnelly, P., (2000). *Inference of population structure using multilocus genotype data*. Genetics, 155(2), 945-959.
Toader, V.A., Moldovan, I.C., Şofletea, N., Abrudan, I.V., Curtu, A.L., (2009). *DNA isolation and amplification in oak species (Quercus spp.).* Bulletin of *Transilvania* University of Braşov, Series II 51, 45-50.
Şofletea, N., Curtu, A.L., (2007). *Dendrologie*, Braşov, Editura "Pentru Viaţă".
Şofletea, N., Enescu, C.M., Curtu, A.L., (2011). *Small-scale morphological descriptors analysis in four Romanian oak stands reported to series Lanuginosae Simk*. Bulletin of *Transilvania* University of Braşov, Series II 53(2), 77-84.

[1] Department of Soil Sciences, University of Agronomic Sciences and Veterinary Medicine of Bucharest, Romania
[2] Department of Forest Sciences, *Transilvania* University of Brasov, Romania
[*] mihaienescu21@yahoo.com

Acknowledgements: This paper is supported by the Sectorial Operational Program Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU/88/1.5/S/59321 and by CNCS-TE-73-2010 project.