Original Research

Journal Home page : www.jeb.co.in ★ E-mail : editor@jeb.co.in

Journal of Environmental Biology

JEB™

ISSN: 0254-8704 (Print)
ISSN: 2394-0379 (Online)
CODEN: JEBIDP

DOI : http://doi.org/10.22438/jeb/38/5(SI)/GM-26

UPGMA and artificial neural networks applications on wild type olives

Authors Info

M. Sesli[1]*, E.D. Yeğen oğlu[2], V. Altıntaş[3] and Y. Gevrekçi[4]

[1]Department of Tobacco Breeding, School of Tobacco Expertise, Manisa Celal Bayar University, Akhisar, Manisa- 45210, Turkey

[2]Department of Plant and Animal Production, Alasehir Vocational School, Manisa Celal Bayar University, Alasehir, Manisa-45400, Turkey

[3]Department of Computer Programming, Akhisar Vocational School, Manisa Celal Bayar University, Akhisar, Manisa-45210, Turkey

[4]Department of Animal Science, Unit of Biometry-Genetics, Bornova, Agriculture Faculty, Ege University, Izmir-35100, Turkey

*Corresponding Author Email : meltem.sesli@bayar.edu.tr

Key words
Artificial neural networks, ISSR, UPGMA, Wild olives

Publication Info
Paper received : 18.09.2016
Revised received : 20.05.2017
Accepted : 27.06.2017

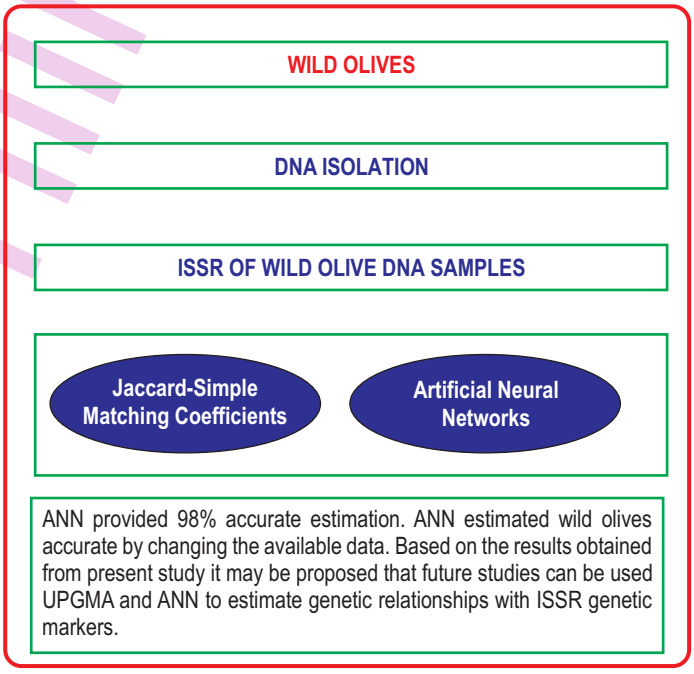
Abstract

Aim: Plant genetic sources are important to study genetic variability and richness of hereditary knowledge of plant species in gene pool. Local varieties, rural populations, wild types and old varieties are the primary ones. In this respect, wild type olives (*Olea europaea oleaster)* are valuable in terms of olive breeding, cultivation and ecosystem. The aim of the study was to determine genetic distances between olive varieties.

Methodology: Artificial Neural Networks intuitive algorithm application was performed on seven wild type olives grown in different regions of Turkey by using data obtained from twenty-two ISSR primers.

Results: UPGMA dendrograms were developed through Jaccard, simple matching coefficients, and similarity matrices; and genetic similarities and dissimilarities were exhibited.

Interpretation: It was concluded that Artificial Neural Networks would be beneficial for estimating olive types accurately based on the results obtained from earlier studies performed with genetic markers.



Introduction

It is genetic variability that confers for plant species to accommodate the environmental conditions. Meanwhile, the plant genetic sources serve as raw materials required for improving plant species in the field of biotechnology that has been progressing rapidly in recent years. Turkey has immense area for genetic studies on genetic variation since it is a gene center for many plants cultivated globally for agriculture purposes. Turkey is being integrated and shaped with the diversity of its geographical structure, high endemism, genetic variability, ecosystems, various habitats and other geographical features of various regions; and this causes the emergence of different plant species in different regions. This shows the biological variability of Turkey. The homeland of olive is Anatolia has (Altındal and Akgün, 2015) rich wild olive sources. Wild types are assessed as all the wild forms of plants are consumed by humans and domestic animals. Wild types are characterized as living in natural habitats, and thus existing only in specific geography. Most important feature of these sources is that they contain many agriculturally important characteristics such as resistance to diseases or environment conditions; even though their agricultural performance is considerably weak and their direct use in agriculture is rather limited. Wild varieties are quite rich especially in terms of genes providing resistance against diseases and pests as they have high economical value (Şakiroğlu, 2010). UPGMA family trees were developed through different coefficients as Jaccard, Simple Matching (Jaccard, 1901; Sokal and Michener, 1958) and similarity matrices to determine the genetic similarities and dissimilarities of wild type olives by using ISSR markers (Zietkiewicz *et al.*, 1994) since they have such important characteristics and ISSR does not need prior information for plant DNA (Lenka et al., 2015). The Unweighted Pair Group Method with Arithmetic Averages (UPGMA) is frequently used for cluster analysis in olive varieties (Asadiar *et al.,* 2013; Noormohammadi *et al.,* 2012). It is a simple method for constructing taxonomic phenograms. Constructing trees reflect similarities between Operational Taxonomic Units (OTUs) (Opperdoes, 1997).

Artificial Neural Networks (Visen *et al*., 2002; Dubey *et al.*, 2006) application which has a significant potential for the classification and definition of agricultural products and used in biological practices frequently for classifications and product definitions in recent years was performed. Artificial Neural Networks (ANN) are parallel and distributed structures consisting of process elements developed through inspiration from human brain and linked together through weighted connections. They are most importantly characterized by learning from experiences (Uğur and Kınacı, 2006). One of the important areas that ANN is used is estimation. ANN may reveal the relations unknown and difficult to distinguish among the data (Zhang *et al.,* 1998); and it may also allow nonlinear modeling without the need of any preliminary information and making any assumptions among the input and output variables (Kaastra and Boyd,1996).

Materials and Methods

Seven wild olives were obtained from Muğla, Manisa, İzmir, Aydın provinces. DNAs of all samples were isolated from fresh leaves according to Doyle and Doyle (1987, 1990) method.

Following the isolation of DNAs of wild olives (*oleasters*), spectrophotometric method was used in the determination of DNA quantities. Spectrophotometric analyses were performed at 260 nm in Gamma Helios spectrophotometer. Samples were put in an Eppendorfs and kept in deep-freezer before the PCR process. Then, PCRs of samples were performed with ISSR primers. ISSR amplification reactions were carried out in 25 µl volume containing 25 ng DNA, primer 4 µl, PCR buffer 2.5 µl, dNTP stock 2 µl, Taq NA polymerase 0.5 µl. PCR cycles were taken from Martins-Lopes *et al.* (2007). Annealing temperatures were determined based on the melting temperatures of ISSR primers. The amplification reactions of ISSR were carried out following these steps: initial denaturation 94 ºC 5 min; followed by 45 cycles of 94 ºC 30 sec, 52 ºC 45 sec., 72 ºC 2 min, and final extension of 72 ºC 5 min. Table 1 shows the provinces, districts and villages from which *oleaster* olives are obtained and Table 2 gives the properties of ISSR primers. The pre-selected UBC (University of British Columbia) ISSR primers were used.

Agarose-gel-electrophoresis was performed for the seperation of DNA fragments that were reproduced following the polymerase chain reaction. The gels were stained with ethidium bromide before casting. During electrophoresis, 1.2% agarose-gel was used for ISSR products and Fermentas Gene Ruler Ladder was used as molecular weight standart (100-5000 bp),

Table 1 : Provinces, districts and villages from where *oleaster* olives were procured

Province	District	Village	Samples
Manisa	Akhisar	Araplar	Wild 1
Manisa	Saruhanli	Seyitoba	Wild 2
İzmir	Selcuk	Selcuk	Wild 3
İzmir	Urla	Ovacik	Wild 4
Mugla	Yatagan	Yatagan	Wild 5
Aydin	Bozdogan	Bozdogan	Wild 6
Mugla	Ortaca	Dalyan	Wild 7

later molecular weights from ISSR primers were used for ANN.

Artificial neural networks were modeled as 78 inputs and 1 output. Input data were taken as molecular weights of bands and output data were taken as wild olives. MATLAB 2013a software (Saudagare and Chaudhari, 2012) was used in the modeling of artificial neural networks and Feed Forward Back Propagation was used as learning algorithm. The structure and parameters of ANN are shown in Table 5. The performance obtained and the training graph of network from ANN is shown in Fig. 3. To develop the ability of estimation for ANN, different parameters were investigated in teaching time. The structure of ANN giving the optimal result were accepted as teaching parameter.

Table 2 : ISSR primers and properties

ISSR primers	Base sequence (5'-3')	Base number
UBC 810	GAGAGAGAGAGAGAT	17
UBC 811	GAGAGAGAGAGAGAC	17
UBC 812	GAGAGAGAGAGAGAA	17
UBC 817	CACACACACACACAA	17
UBC 818	CACACACACACACAG	17
UBC 823	TCTCTCTCTCTCTCG	17
UBC 825	ACACACACACACACT	17
UBC 826	ACACACACACACACC	17
UBC 834	AGAGAGAGAGAGAGYT	18
UBC 850	GTGTGTGTGTGTGTYC	18
UBC 855	ACACACACACACACYT	18
UBC 889	DBDACACACACACACAC	15
IMA 9	GAGAGAGAGAGAGAGACG	18
UBC 801	ATATATATATATATTT	17
UBC 802	ATATATATATATATTG	17
UBC 803	ATATATATATATATTC	17
UBC 804	TATATATATATATTAA	17
UBC 805	TATATATATATATTAC	17
UBC 806	TATATATATATATTAG	17
UBC 807	AGAGAGAGAGAGAGAGT	17
UBC 808	AGAGAGAGAGAGAGAGC	17
UBC 809	AGAGAGAGAGAGAGAGG	17

Table 3 : Results obtained by Jaccard Similarity Coefficient

	Wild 1	Wild 2	Wild 3	Wild 4	Wild 5	Wild 6	Wild 7
Wild 1	1.00						
Wild 2	0.08	1.00					
Wild 3	0.04	0.07	1.00				
Wild 4	0.04	0.07	0.04	1.00			
Wild 5	0.04	0.00	0.07	0.04	1.00		
Wild 6	0.05	0.13	0.04	0.00	0.00	1.00	
Wild 7	0.00	0.07	0.03	0.04	0.10	0.00	1.00

Table 4 : Results obtained by Jaccard Similarity Coefficient

	Wild 1	Wild 2	Wild 3	Wild 4	Wild 5	Wild 6	Wild 7
Wild 1	1.00						
Wild 2	0.68	1.00					
Wild 3	0.66	0.65	1.0				
Wild 4	0.69	0.68	0.66	1.00			
Wild 5	0.66	0.59	0.66	0.66	1.00		
Wild 6	0.73	0.75	0.69	0.71	0.67	1.00	
Wild 7	0.63	0.65	0.62	0.66	0.67	0.67	1.00

Results and Discussion

Following the agarose-gel electrophoresis, the dyed gels were viewed and the bands provided by each ISSR primer in wild olives were evaluated as an independent individual locus; and a data matrix was developed to have 1 for the existence of bands and 0 for their absence. In data analysis, the NTSYS pc ver. 2.1 program was used and genetic distance matrices were developed in wild olives through Jaccard (1901) and Simple matching (Sokal and Michener, 1958) Coefficients; UPGMA (Sokal and Michener, 1958) trees were formed thereover (Rohlf, 1998).

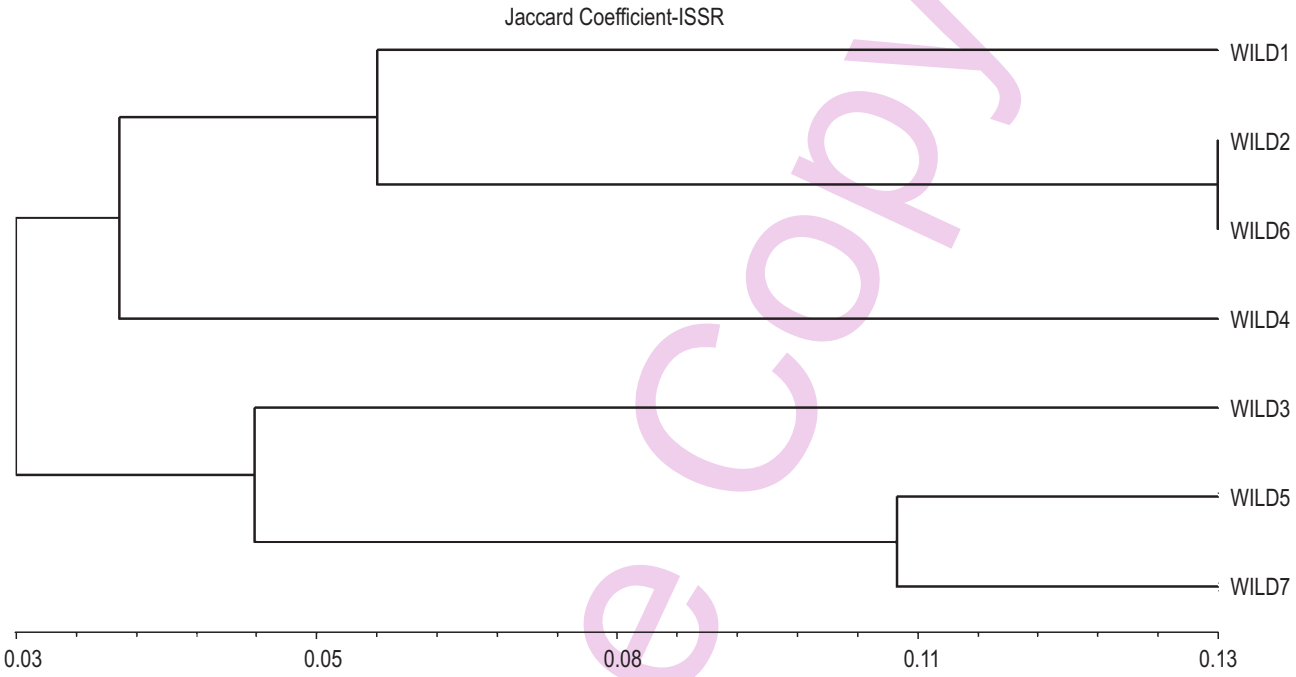


Fig. 1 : Jaccard Similarity Coefficient UPGMA dendrogram

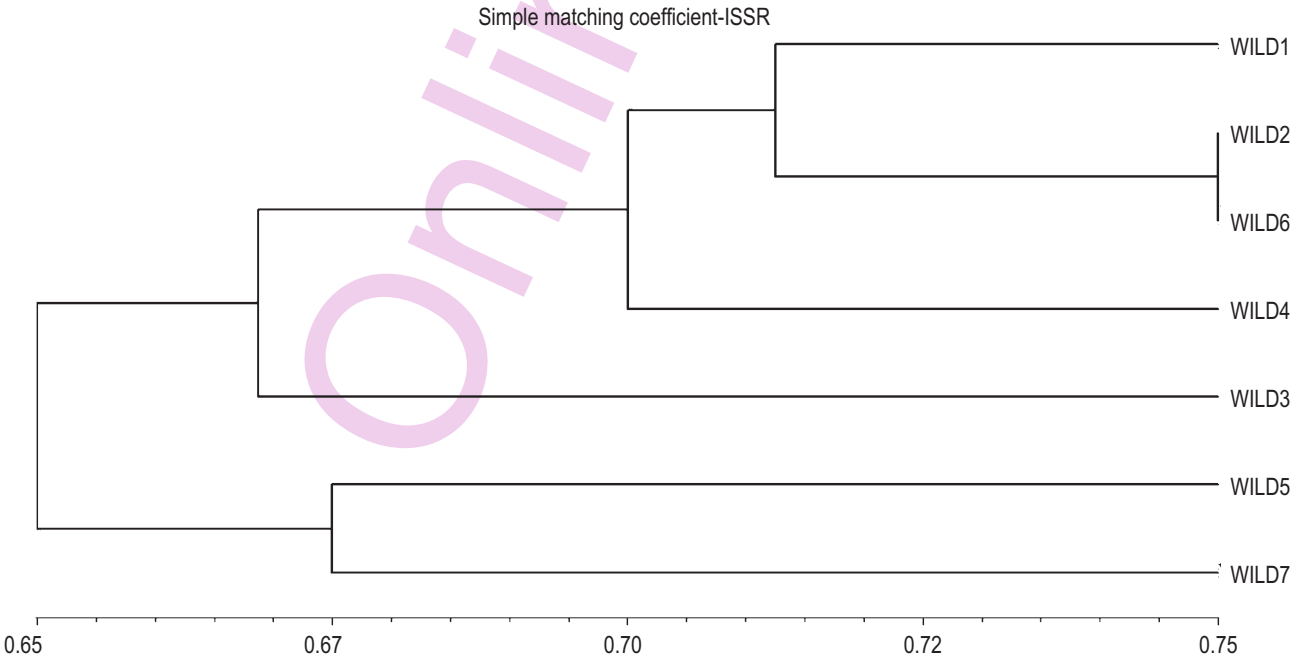


Fig. 2 : Simple Matching Coefficient UPGMA dendrogram

Table 3 and Fig. 1 show the results of Jaccard Similarity Coefficient and UPGMA dendrogram. When the matrix and dendrogram are examined, the genetic distance values were between 0.03 (Wild 7 and Wild 3) and 0.13 (Wild 2 and Wild 6). Thus, the samples closest to each other based on the genetic distance values were Wild 7 and Wild 3; and the samples farthest to each other based on genetic distance values were Wild 2 and Wild 6.

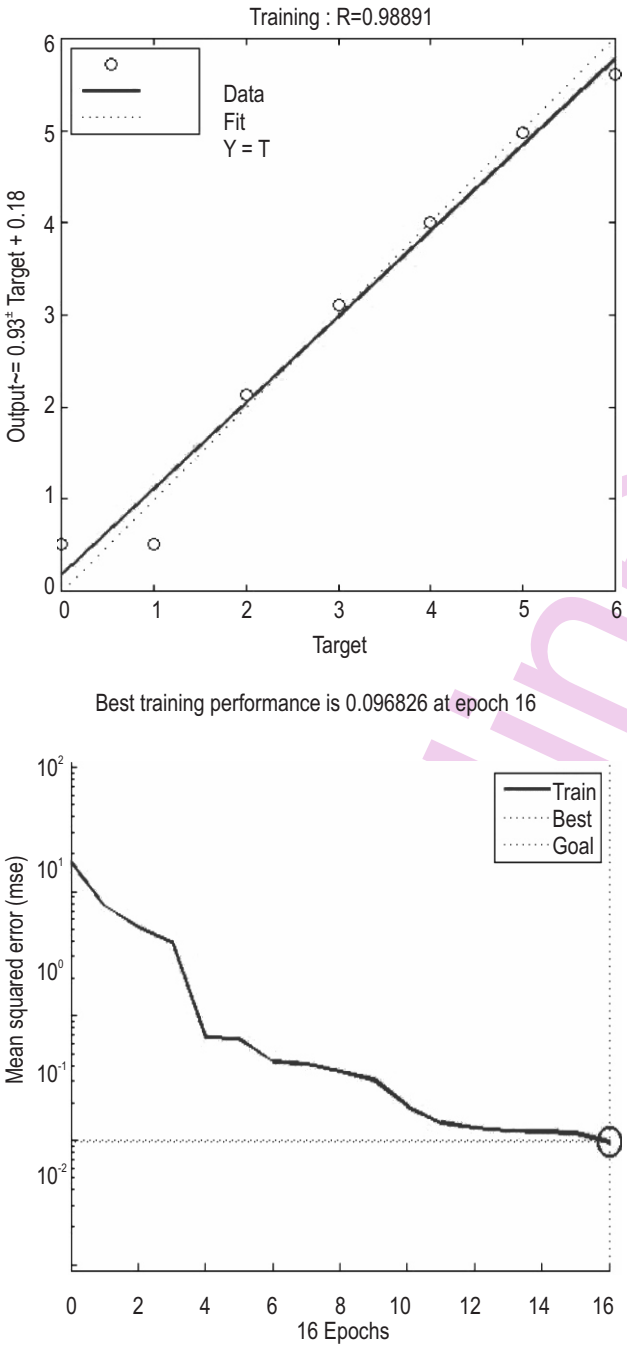


Fig. 3 : Performance and training graph of network

Table 4 and Fig. 2 presents the results from Simple Matching Coefficient and UPGMA dendrograms of wild olives. When the matrix and dendrogram were examined, the genetic distance values were between 0.59 (Wild 2 and Wild 5) and 0.75 (Wild 2 and Wild 6). Thus, the samples closest to each other based on the genetic distance values were Wild 2 and Wild 5; and the samples farthest to each other based on genetic distance values were Wild 2 and Wild 6.

ISSR markers are dominant markers; each band obtained from these represents a single bi-allelic locus in phenotype. The existence of bands in studies with markers was interpreted as heterozygote or dominant homozygote, and the absence of bands was interpreted as recessive homozygote. The basic data structure in ISSR markers consists of binary (0/1) matrix. Based on this, Jaccard and Simple Matching Coefficients

Table 5 : Training parameters and artificial neural networks (ANN) structure

	Artificial Neural Networks (ANN)
Number of coefficients	3
Weight values	Random
Activation function	Tansig
Learning function	Back propagation
Learning rate	0.8
Mean-squared error	1e-10

Table 6 : ANN modelling

Marker	**Test values**
UBC889-2	579
UBC889-5	15
UBC817-2	600
UBC817-6	250
UBC818-4	792
UBC818-19	544
UBC818-23	25
UBC818-25	411
UBC850-9	631
UBC850-18	447
UBC850-22	15
UBC850-28	332
UBC850-33	117
UBC834-2	1133
UBC834-13	707
UBC834-17	622
UBC834-21	554
UBC855-4	740
IMA9-9	695
IMA9-10	20
Network Predicted	0,9841
Expected Result	1
Wild 2	

were used to determine genetic similarities and diversities; Jaccard Coefficient, different from Simple Matching, considers the existence of DNA bands, not the absence of bands, when calculating the genetic similarity (Debnath, 2006). As a result, when the dendrograms obtained from two different coefficients through UPGMA method were examined, it was determined that both Jaccard and Simple Matching Coefficients provided the farthest samples to each other based on the genetic distance values as Wild 2 and Wild 6; however, the samples closest to each other based on the genetic distance values were different in results obtained from both coefficients.

Table 6 shows the data of value found by the network by changing some input marker values. Other input values were entered as 0 for this trial. It was also set forth through determining the expected result as wild 2 in the modeling that the wild olives estimated with the artificial neural networks model and the wild olives known actually showed genetic distance to each other. The wild olive which was determined jointly by both the Jaccard and Simple Matching coefficients based on the genetic distance value was wild 2. Consequently, it was observed that determination of genetic relationships based on ANN estimation and markers supported each other and this modelling technique provided quite successful results.

As a result of modelling, it was determined that ANN provided 98% accurate estimation. It was also checked as to whether ANN estimated the wild olives accurate by changing the available data. It was concluded that ANN would be beneficial for estimating olive types accurately. Based on these results it is concluded that ANN would be beneficial for estimating the olive types accurately based on the data obtained from further studies with genetic markers**.**

Acknowledgment

This study has been supported by the Scientific Research Projects Coordination Unit of Manisa Celal Bayar University, Manisa, Turkey.

References

Altındal, D. and I. Akgün: Plant genetic resources and the status in cereals. *J. Adn. Mend. Univ. Agricult. Fac.,* **12**, 147 – 153 (2015).

Bitki Genetik KaynaklarıveTahıllardaki Durumu.*Adnan Menderes Üniversitesi Ziraat Fakültesi Dergisi*., **12**, 147 – 153 (2015).

Asadiar, L.S., F. Rahmani and A. Siami: Assessment of genetic diversity in the Russian olive *(Elaeagnus angustifolia*) based on ISSR genetic markers.Rev. Ciênc. Agron. 44 (2). http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-66902013000200013 (2013).

Debnath, S.C.: Inter simple sequence repeat (ISSR) to assess genetic diversity within a collection of wild lingonberry (*Vaccinium vitis-idaea* L.) clones. *Cana J. Plant Sci*., **2**, 337-344 (2006).

Doyle, J.J. and J.L. Doyle: A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*., **19**, 11-15 (1987).

Doyle, J.J. and J.L. Doyle: Isolation of plant DNA from fresh tissue. *Focus*., **12**, 13-15 (1990).

Dubey, B.P., S.G. Bhagwat, S.P. Shouche and J.K. Sainis: Potential of artificial neural networks in varietal identification using morphometry of wheat grains. *Biosyst. Engg.*, **95**, 61-67 (2006).

Jaccard, P.: Étude comparative de la distribuitionfloraledansune portion des Alpeset des Jura. *Bull. Soc. Vandoise. Sci. Nat.,* **37**, 547-579 (1901).

Kaastra, I. and M. Boyd: Designing a neural network for forecasting financial and economic time series. *Neurocomputing,* **10**, 215- 236 (1996).

Lenka, D., S.K. Tripathy, R. Kumar, M. Behera and R. Ranja: Assessment of genetic diversity in quality protein maize (QPM) inbreds using ISSR markers. *J. Env. Biol.,* **36**, 985-992 (2015).

Martins-Lopes, P., J.S. Lima-Brito Gomes, J. Meirinhos, L. Santos and H. Guedes-Pinto: RAPD and ISSR molecular markers in *Olea europaea* L. Genetic variability and molecular cultivar identification. *Genet. Res. Crop. Evol*., **54**, 117-128 (2007).

Noormohammadi, Z., H. Samadi-Molayousefi and M. Sheidai: Intra-specific genetic diversity in wild olives (*Olea europaea ssp. cuspidata*) in Hormozgan Province. *Iran. Genet. Mol. Res.,* **11**, 707-716 (2012).

Opperdoes, F.: Construction of a distance tree using clustering with the Unweighted Pair Group Method with Arithmatic Mean (UPGMA). http://www.inf.u-szeged.hu/~ormandi/ai2/02-UPGMA-example.pdf. Access: April, 2017.

Rohlf, F.J.: NTSYSpc: Numerical taxonomy and multivariate analysis system version 2.02 Exeter Software, Setauket, NY (1998).

Saudagare, P.V. and D.S. Chaudhari: Facial expression recognition using neural network –An cverview. *Int. J. Soft Comp. Enginee. (IJSCE)*, **2**, (2012).

Sokal, R.R. and C.D. Michener: A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.,* **38**, 1409- 1438 (1958).

Şakiroğlu, M.: International sharing problem of plant genetics sources. *Seta Analysis,* **25**, 7 (2010).

Uğur, A. and A.C. Kınacı: Artificial Intelligence Techniques and Classification of Web Pages by using Artificial Neural Networks. Inet-tr'06 - XI. Reports of "Internet in Turkey" Conference (2006). Yapay Zeka Tekniklerive Yapay Sinir Ağları Kullanılarak Web Sayfalarının Sınıflandırılması. Inet-tr'06 - XI. "Türkiye'de Internet" Konferansı Bildirileri. Zhang, G., B.E. Patuwo and M.Y. Hu: Forecasting with Artificial Neural Networks: The State of the Art. *Int. J. Forecas*.,**14**, 35- 62 (1998).

Visen, N.S., J. Paliwal, D.S. Jayas and N.D.G. White: Specialist neural networks for cereal grain classification. *Biosyst. Engin.*, **82**, 151-159 (2002).

Zietkiewicz, E., A. Rafalski and D. Labuda: Genome fingerprinting by simple sequence repeats (SSR)-anchored PCR amplification. *Genomics*, **20**, 176-183 (1994).