

A multivariate morphometric analysis of diagnostic traits in southern Italy and Sicily pubescent oaks

Romeo Di Pietro¹, Antonio Luca Conte², Piera Di Marzio², Lorenzo Gianguzzi³, Giovanni Spampinato⁴, Orazio Caldarella³, Paola Fortini²

¹Department PDTA, University of Rome Sapienza, 00196 Rome, Italy.

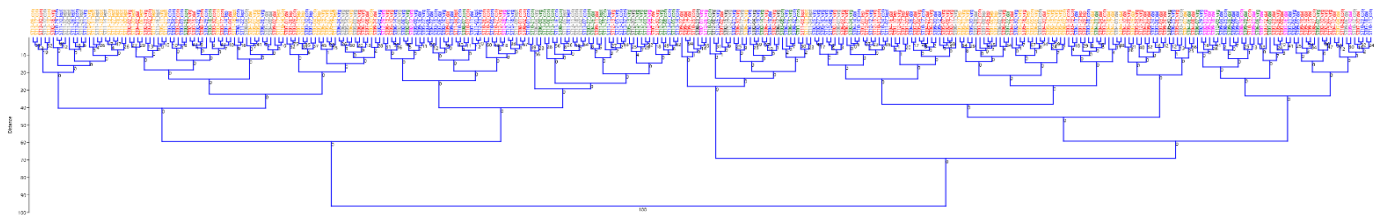
²Department of Bioscience and Territory, University of Molise, 86090 Pesche (IS) Italy.

³Department of Scienze Agrarie e Forestali, University of Palermo, Italy.

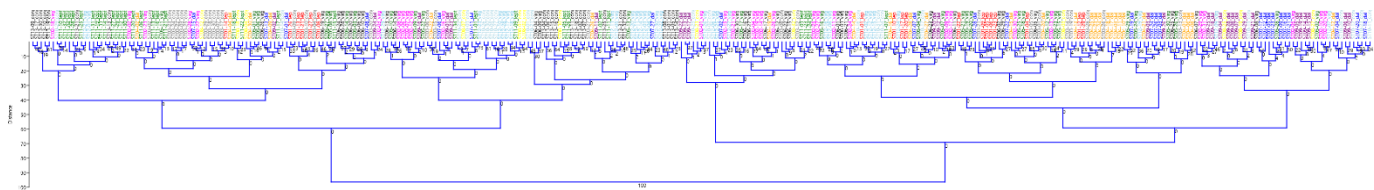
⁴Department of AGRARIA, Mediterranean University of Reggio Calabria, Italy.

✉: Romeo Di Pietro, romeo.dipietro@uniroma1.it

Online Resource 4. Multivariate analyses results.



Oak specimens are coloured following the expert classification: *Quercus dalechampii*, *Q. congesta*, *Q. ichnusae*, *Q. leptobalanos*, *Q. pubescens*, *Q. virgiliana*, doubtful specimens.



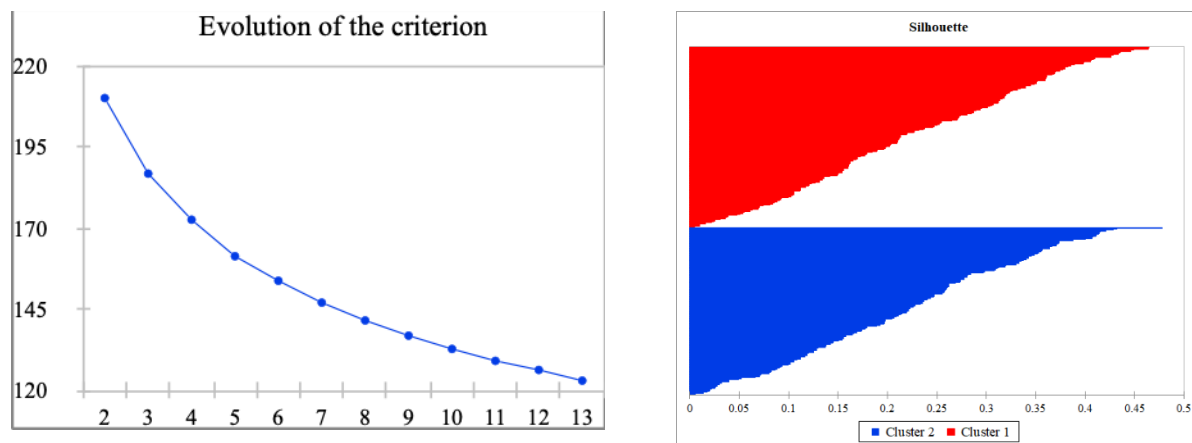
Oak specimens are coloured following the thirteen population sites: ST01 (red), ST02 (fuchsia), ST03(gray), ST04 (orange), ST05 (darkgreen), ST06 (black), ST07(blue), ST08 (yellow), ST09 (purple), ST10 (lightblue), ST11 (green), ST12 (skyblue), ST13 (darkgolden)

(a) Dendrograms obtained from the hierarchical clustering with PAST vers. 4.03 (Euclidean distance, Ward’s method). Oak specimens labelled following the putative species identified by expert classification (dendrogram above) and following the populations sites of collection (dendrogram below)

(b) **Table 1** Hierarchical clustering with JASP vers. 0.13 (estimation method “moment”, model optimized with respect to the silhouette value)

Clusters	N	R ²	Silhouette
2	333	0.102	0.180

Cluster information	1	2
Size	205	128
Explained proportion within-cluster heterogeneity	0.682	0.318
Within sum of squares	5689.908	2653.405

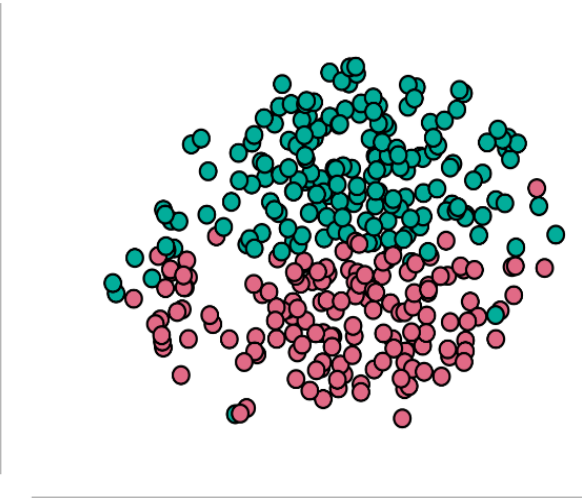


(c) Fuzzy k-means clustering with XLSTAT vers. 2020.3.1 (coefficient of fuzziness=1.1, method: Cosine, 20 repetitions, not weighted, initialization method: K||): evolution of the criterion and silhouette plot

(d) **Table 2** Results of the k-means clustering with JASP vers. 0.13 (estimation method “moment”, model optimized with respect to the silhouette value

Clusters	N	R ²	Silhouette
2	333	0.135	0.120

Cluster information	1	2
Size	151	182
Explained proportion within-cluster heterogeneity	0.493	0.507
Within sum of squares	3967.586	4075.427
Silhouette score	0.084	0.157



(e) t-SNE Cluster Plot of the k-means clustering with JASP vers. 0.13 (estimation method “moment”, model optimized with respect to the silhouette value). Cluster 1 is shown in red, cluster 2 is shown in green

(f) Table 3 Number of specimens per group

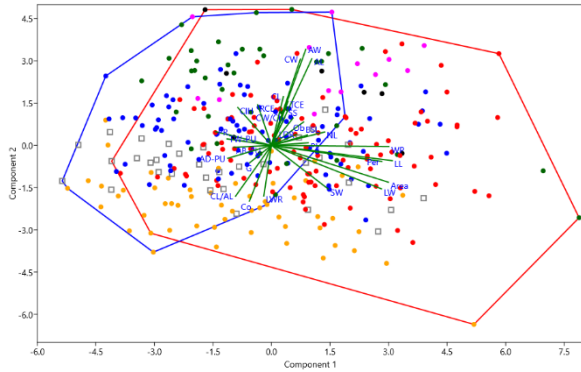
	HCP	HCJ	FKM	KMJ
Group 1	161	205	172	151
Group 2	172	128	161	182

HCP – Hierarchical clustering with PAST, HCJ – Hierarchical clustering with JASP, FKM – Fuzzy k-means clustering with XLSTAT, KMJ – k-means clustering with JASP. The highest value is shown in bold type

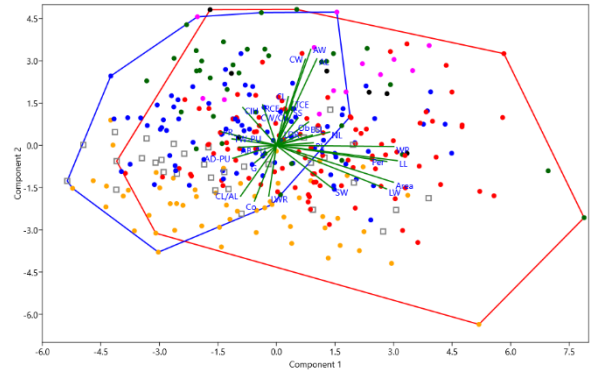
(g) Concordances of specimens in the two groups, regardless of the order of the groups

	HCP / HCJ	HCP / FKM	HCP / KMJ	HCJ / FKM	HCJ / KMJ	FKM / KMJ
number	284	292	290	275	269	323
%	85.29	87.69	87.09	82.58	80.78	97.00

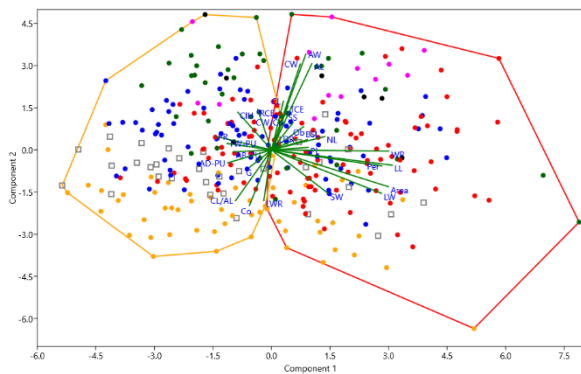
HCP – Hierarchical clustering with PAST, HCJ – Hierarchical clustering with JASP, FKM – Fuzzy k-means clustering with XLSTAT, KMJ – k-means clustering with JASP. The highest value is shown in bold type



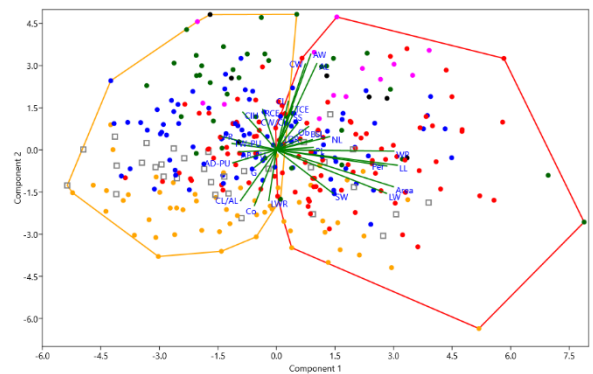
HCP



HCJ



FKM



KMJ

(h) Principal Component Analysis (PCA) biplot for 28 leaf and fruit variables in 333 oak specimens, with superimposed classifications. The first two axes explain 32.14% of the variance. HCP – Hierarchical clustering with PAST, HCJ – Hierarchical clustering with JASP, FKM – Fuzzy k-means clustering with XLSTAT, KMJ – k-means clustering with JASP. Individuals were labelled following the expert classification: Dots – *Quercus dalechampii*, *Q. congesta*, *Q. ichnusae*, *Q. leptobalanos*, *Q. pubescens*, *Q. virgiliana*; Squares – doubtful specimens.

(i) Non-parametric Kruskal-Wallis's test p-values

Variable	HCP	HCJ	FKM	KMJ
Area	<0.0001	<0.0001	<0.0001	<0.0001
Per	<0.0001	<0.0001	<0.0001	<0.0001
PL	<0.0001	0.718	<0.0001	0.001
LL	<0.0001	<0.0001	<0.0001	<0.0001
LW	<0.0001	<0.0001	<0.0001	<0.0001
SW	<0.0001	<0.0001	<0.0001	<0.0001
WP	<0.0001	<0.0001	<0.0001	<0.0001
NL	<0.0001	<0.0001	<0.0001	<0.0001
BSL	<0.0001	<0.0001	<0.0001	<0.0001
Co	0.282	0.021	0.015	0.194
Ob	0.015	0.504	0.202	0.023
PR	<0.0001	<0.0001	<0.0001	<0.0001
LDR	0.001	0.117	<0.0001	<0.0001
LWR	0.426	0.233	0.586	0.461
AB-PU	<0.0001	<0.0001	0.002	<0.0001
AD-PU	<0.0001	<0.0001	<0.0001	<0.0001
TW-PU	<0.0001	<0.0001	<0.0001	<0.0001
CL	0.948	0.282	0.062	0.775
CW	0.027	0.997	0.047	0.014
AL	<0.0001	0.157	0.005	0.002
AW	<0.0001	0.591	0.002	0.001
CW/CL	0.034	0.695	<0.0001	0.015
CL/AL	<0.0001	0.087	<0.0001	0.001
RCE	0.081	0.001	0.094	0.064
SS	0.023	<0.0001	0.001	0.017
G	0.104	<0.0001	<0.0001	0.049
CIH	0.005	<0.0001	0.006	0.004
TCE	0.507	0.631	0.380	0.527

Not significant p-values are shown in red. the most significant p-values are shown in bold type. HCP – Hierarchical clustering with PAST, HCJ – Hierarchical clustering with JASP, FKM – Fuzzy k-means clustering with XLSTAT, KMJ – k-means clustering with JASP

(l) Results of the Linear Discriminant Analysis (LDA) based on the two groups resulting from the four cluster analyses

Linear discriminant		Method	N(Train)	N(Test)	Test Accuracy
HCP	1	Moment	267	66	0.894
HCJ	1	Moment	267	66	0.939
FKM	1	Moment	267	66	0.955
KMJ	1	Moment	267	66	0.939

HCP – Hierarchical clustering with PAST, HCJ – Hierarchical clustering with JASP, FKM – Fuzzy k-means clustering with XLSTAT, KMJ – k-means clustering with JASP. The highest value of accuracy is shown in bold type

(m) Confusion matrices resulting from the test of the Linear Discriminant Classification (LDC)

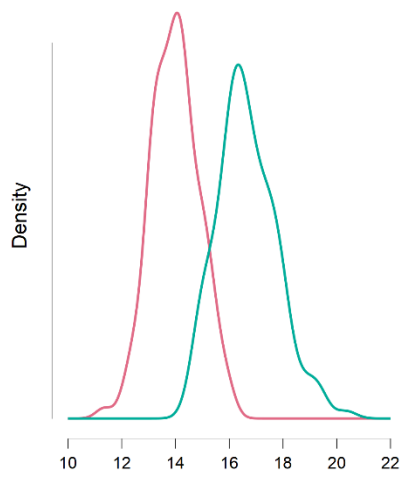
HCP		<i>Predicted</i>	
<i>Observed</i>		1	2
	1	32	4
	2	3	27

HCJ		<i>Predicted</i>	
<i>Observed</i>		1	2
	1	39	2
	2	2	23

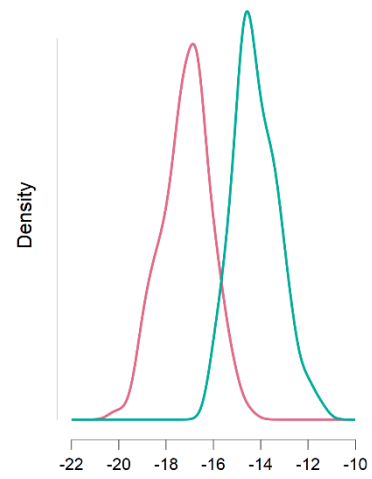
FKM		<i>Predicted</i>	
<i>Observed</i>		1	2
	1	36	2
	2	1	27

KMJ		<i>Predicted</i>	
<i>Observed</i>		1	2
	1	26	1
	2	3	36

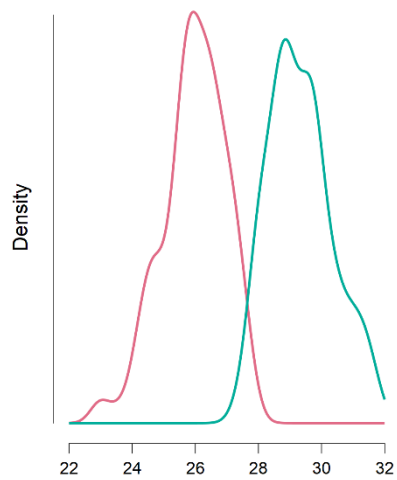
HCP – Hierarchical clustering with PAST, HCJ – Hierarchical clustering with JASP, FKM – Fuzzy k-means clustering with XLSTAT, KMJ – k-means clustering with JASP



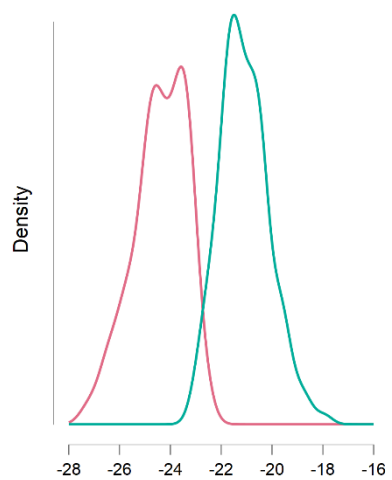
HCP



HCJ



FKM



KMJ

(n) Densities on the linear discriminant 1 (LD1). HCP – Hierarchical clustering with PAST, HCJ – Hierarchical clustering with JASP, FKM – Fuzzy k-means clustering with XLSTAT, KMJ – k-means clustering with JASP. Group 1 is shown in red, group 2 is shown in green