

## **Application of Cluster Analysis and Principal Component Analysis in the Development of Table Seeded and Seedless Vine Cultivars (*Vitis Vinifera* L.)**

V. ROYTCHEV

*Agricultural University, BG - 4000 Plovdiv, Bulgaria*

### **Abstract**

ROYTCHEV, V., 2007. Application of cluster analysis and principal component analysis in the development of table seeded and seedless vine cultivars (*Vitis vinifera* L.). *Bulg.J. Agric. Sci.*, 13: 409-421

The possibilities for application of cluster analysis and principal component analysis in the investigation of  $F_1$  progeny of a cross between a seeded and seedless vine cultivar have been studied. It has been established that their combined application allows obtaining more profound ampelographic information on the character and significance of each trait in the grouping of genotypes. The selection of elite hybrid forms should be carried out according to those traits, which explain to the greatest extent variation in the first and second principal component. The division of seedlings into two groups - seeded and seedless, increases the efficiency of both analyses in the development of new table vine cultivars possessing valuable economic characteristics. Promising elite hybrid forms which should be used in future hybrid combinations, have been selected from the population of seeded and seedless seedlings in accordance with the criteria of the applied methods. Other hybrid forms deserving special attention have been pointed out as well.

*Key words:* development of seedless and seeded vine cultivars, elite hybrid forms,  $F_1$  progeny, cluster analysis, principal component analysis

### **Introduction**

Cluster analysis is a method used in viticulture to allow differentiate and classify large groups of cultivars, characterized by significant diversity of morphological, biochemical, agrobiological and technological traits. Svanepel, Villiers

(1987), Cameiro, Lina (1989), Ortis et al. (1990). Rodriguez et al. (1994). Troshin et al. (1998) apply this method successfully in studies of morphological correspondence between hybrid vine forms and their parents, as well as in intrageneric and intraspecific classification of cultivars (*Vitis* L.) based on ampelographic leaf traits. At

present this method finds wider application in pursuing hemotaxonomic objectives by determining the amount of anthocyanins, proteins, isoenzyme systems and other biochemical components in the grapes and wine of a considerable number of cultivars, forming specific groups (Altube et al., 1991; Ferreira et al., 1996; Carreno et al., 1997; Gao, Cahoon, 1998; Moreno-Arribas et al., 1999). Cluster analysis has been used for selection purposes in vine (*V. vinifera* L.) by Krastanova (1986), Molle, Krastanova (1987), Judez et al. (1995). According to Spagnoletti Zeuli, Qualset (1987) this method should be employed in different plant species. It is assumed that the traits used in cluster analysis participate to a different extent in the differentiation of cultivar remoteness. The relative significance of the traits having the strongest influence on division of genotypes into groups is determined by means of principal component analysis (PCA). The higher the value of the variation coefficient of a certain trait from a studied group is, the greater its influence is. The preliminary evaluation of the genetic remoteness of parent genotypes helps to make the right choice as regards the concrete purpose of selection. The result of each hybrid combination could be predicted as well (Goodman, 1973; Dubles, Jain 1980; Cox et al. 1985; Murphy et al. 1986). Bassi et al. (1995), Yao et al. (1995), Mattheou et al. (1995), Tomazic, Korosec-Koruza (1998), Goto-Yamamoto (2000), Boselli et al. (2000), Cervera et al. (2002), Labra et al. (2002), Hvarleva et al. (2004), Hvarleva T., A. Atanassov (2006) and Varga et al. (2006) determine aspects of cultivar identification related to their utilization in conventional ampelography and pomology combined with numerical taxonomy.

The purpose of this investigation is to determine the genetic similarity and remoteness of the seeded and seedless plants from F<sub>1</sub> progeny of the Super Early Bolgar x Ruby Seedless cross, to explain the reasons for their clustering and the possibilities for more efficient selection of valuable elite hybrid vine forms.

### Materials and Methods

The experiment included 30 seedlings from F<sub>1</sub> progeny of a cross between the seeded table cultivar Super Early Bolgar and the seedless cultivar Ruby Seedless. The plants were grown to a training system based on a bi-lateral cordon of equal loading during pruning with 6 nodes carrying 2 winter buds and 2 fruiting shoots carrying 18 winter buds. For a period of 4 consecutive years 22 agrobiological traits which characterize the economic value of each plant, were reported: yield (kg); coefficient of shoot fertility; coefficient of main shoot fertility; coefficient of fruiting shoot fertility; millerandage berries (%); average bunch weight (g); bunch length (cm); bunch width (cm); average weight of 100 berries (g); berry length (mm); berry width (mm); index of berry shape; budding-flowering (days); flowering-softening (days); softening-technological maturity (days); budding-technological maturity (days); sugars (%); acids (g/dm<sup>3</sup>); total number of buds; total shoot number; total fruiting shoot number; total bunch number (Table 1). The values of the quantitative indices were calculated in accordance with the established viticulture methods. Cluster analysis of the average values for the four-year study was used to determine the homogeneous groups on the basis of genetic similarity and remoteness (Ward, 1963; Duran and Odell, 1974; Everitt,

1979). The processing of the data was conducted using the statistical program SPSS. The genetic remoteness was calculated by finding the Euclidian distance between each pair of objects in the multi-dimensional space, whose coordinates depend on the values of the studied traits. The forming of clusters (genotype groups) is presented graphically in dendrograms showing the order in which objects unite and individual groups form. Principle component analysis (PCA) was performed to establish the relative significance of traits for the division of genotypes into groups (Philippea, 1990). Owing to the seedlings being grown on their own roots the values of certain traits naturally decreased. The entire  $F_1$  progeny was conditionally divided into seeded and seedless plants and the mentioned analyses were carried out in these two plant groups as well.

## Results and Discussion

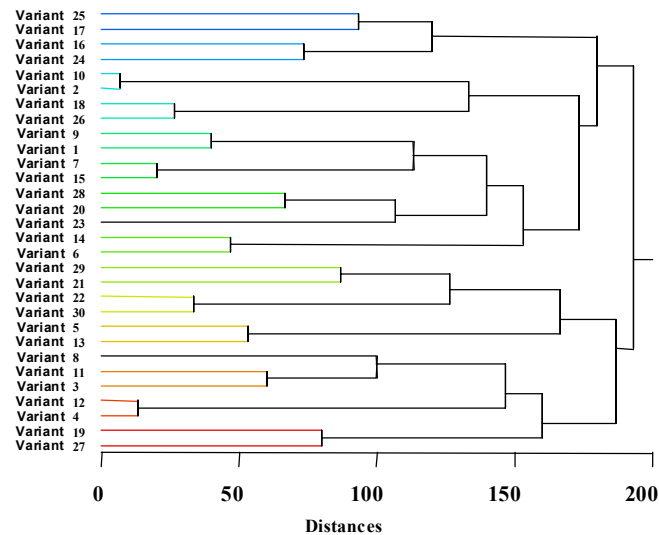
Depending on the relative distance between them, the studied 30 genotypes have been divided into two large groups and five subgroups (Figure 1). The first cluster (genotype subgroup) includes the following plants – 25, 17, 16, 24; the second cluster – 10, 2, 18, 26; the third one – 9, 1, 7, 15, 28, 20, 23, 14, 6; the fourth one – 29, 21, 22, 30, 5, 13 and the fifth one – 8, 11, 3, 12, 4, 19, 27. In each of the groups the variants are characterized by a high degree of phenotypic similarity and they are considered to be approximately identical. The formation of a large number of subgroups reveals the great diversity of traits in  $F_1$  progeny of this hybrid combination and the possibilities for more efficient selection. The relative distances between individual variants range from

0.000 to 197.557.

The results from the analysis of the possible 22 principle components show that only 7 of them are enough to explain 87.775 % of total variation (Table 2). The traits: yield, total bunch number, total fruiting shoot number, coefficient of shoot fertility and average weight of 100 berries prove to exert the greatest influence on the division of genotypes into groups, since their relative variation rate correlates to the greatest extent with the first principle component, which explains 26.500 % of the total variation.

The second principle component explains 19.430 % of the total variation, and the traits: total bud number, bunch width and acids possess the highest correlation coefficients. The third principle component explains 13.820 % of the total variation of traits mainly through: bunch length, index of berry shape, budding-flowering. The remaining four components - the fourth, fifth, sixth and seventh one, explain respectively: 9.586 %; 7.414 %; 5.685 % and 5.340 % of the total variation, the traits with higher correlation coefficients being: coefficient of fruiting shoot fertility, softening-technological maturity, budding- technological maturity and acids.

These results show that the selection in the entire seedling population will be the most efficient when aiming at improving the following traits: yield, total bunch number, total fruiting shoot number, coefficient of shoot fertility and average weight of 100 berries. The seedlings (genotypes) numbered 8, 21, 22, 29 and 30 possess the comparatively highest absolute values of those indices (Table 1). In future selection research seedlings from the first subgroup of the first main cluster and from the fifth subgroup of the second



**Fig. 1. Dendrogram of cluster analysis of the genotypes of  $F_1$  progeny of the hybrid combination Super Early Bolgar x Ruby Seedless**

main cluster (group) should be included since they are the most genetically remote.

The seeded seedlings in this hybrid combination are divided into two large groups and four subgroups (Figure 2). The first subgroup includes plants 6 and 22; the second one – 18, 2, 14, 30, 29, 13; the third one – 10, 26, 15, 24, 8; and the fourth one – 25 and 9. The genotypes from the first and the fourth group are separated from all the rest, being the most genetically remote. The relative distances between individual variants range from 0.000 to 131.371.

The principal component analysis (PCA) shows that five of these components are enough to explain 94.207 % of the total variation (Table 3). The traits: index of berry shape, budding-flowering, yield, total number of bunches–fruiting shoots–shoots and bunch length explain 36.246 % of the total variation of seedlings in the studied population. They

influence the grouping of genotypes to the greatest extent since their relative variation rate correlates to the greatest extent with the first principal component. The traits: berry length, coefficient of fruiting shoot fertility, bunch width, total fruiting shoot number and total bud number are characterized by the highest correlation coefficients with the second principal component. The second principal component explains 26.616 % of the total variation. The third principal component explains 13.804 % of the total variation, mainly by means the traits: acids, total bud number and softening-technological maturity. The fourth and the fifth principal components explain respectively 9.469 % and 8.072 % of the total variation, the traits: budding-technological maturity, flowering-softening and bunch length having the highest correlation coefficients. Seedlings 8, 22, 29 and 30 possess the highest values of the studied traits, correlating to the

**Table 1**  
**Values of the studied traits of F<sub>1</sub> progeny of the hybrid combination Super Early Bolgar x Ruby Seedless**

Traits	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	2.223	1.12	1.12	1.46	27.78	117	13.50	8.40	230	16.80	13.80	1.21	59	48	29	136	29.64	5.00	18	17	13	19
2	1.105	0.85	0.85	1.31	23.10	65	14.80	7.50	200	15.60	13.90	1.12	57	55	19	131	22.11	5.25	30	20	13	17
3	1.287	0.55	0.63	1.00	26.02	117	17.40	9.10	340	18.60	15.50	1.20	58	44	33	135	21.90	5.25	39	20	11	11
4	0.590	0.70	0.72	1.00	14.83	118	16.80	8.30	355	19.40	16.30	1.19	59	41	33	133	20.93	3.23	18	10	7	7
5	1.592	0.91	0.30	1.13	15.94	199	17.50	7.40	300	19.05	14.60	1.30	61	48	26	135	22.07	3.39	38	20	16	18
6	0.952	0.28	0.28	1.14	25.18	119	14.60	9.10	220	16.10	14.30	1.12	52	69	25	146	27.84	4.05	49	29	7	8
7	2.240	0.83	0.83	1.25	20.48	112	18.80	6.30	240	16.90	14.80	1.14	62	43	33	138	17.86	4.95	38	24	16	20
8	4.181	1.16	1.26	1.68	20.11	113	18.70	6.40	340	21.40	15.10	1.42	64	38	33	135	20.87	4.06	41	32	22	37
9	2.241	1.00	1.12	1.33	28.62	130	16.45	9.51	247	17.91	15.8	1.13	57	46	30	133	20.44	4.10	28	20	15	20
10	1.124	0.77	0.84	1.13	25.00	85	16.40	8.50	210	16.50	14.90	1.10	58	56	20	134	20.21	5.15	36	22	15	17
11	1.293	0.50	0.58	1.00	25.12	127	17.20	10.10	339	18.72	16.40	1.14	60	45	33	138	20.80	5.52	41	28	14	14
12	0.610	0.55	0.58	1.25	15.81	120	16.46	8.39	365	18.80	16.40	1.15	56	42	34	132	19.95	4.25	28	18	8	10
13	1.586	0.78	0.85	1.13	16.84	196	18.50	8.40	310	19.20	15.60	1.23	60	47	28	135	20.07	4.49	40	23	16	18
14	0.974	0.44	0.50	1.09	24.12	121	15.50	10.00	231	17.10	13.30	1.28	54	68	27	149	22.74	4.34	48	27	11	12
15	2.256	0.77	0.83	1.11	24.33	119	17.60	9.30	251	17.90	13.80	1.29	63	44	32	139	16.63	4.85	40	26	18	20
16	1.050	0.40	0.48	1.27	12.82	75	10.20	6.80	150	13.40	12.40	1.08	60	38	36	138	22.20	4.47	45	35	11	14
17	1.088	0.43	0.46	1.45	12.83	68	19.70	7.50	120	12.50	10.80	1.16	64	50	24	138	21.16	3.88	47	37	11	16
18	0.704	0.57	0.57	1.33	28.70	88	10.80	7.00	200	15.30	13.30	1.15	57	43	33	133	22.92	3.00	14	14	6	8
19	0.912	0.92	0.98	1.00	18.40	76	10.30	6.90	310	19.20	15.20	1.26	61	41	33	135	23.71	3.54	15	13	12	12
20	2.024	0.71	0.73	1.10	25.01	92	17.80	8.40	220	16.90	14.10	1.20	57	45	29	131	20.84	4.20	38	31	20	22
21	4.000	0.84	0.89	1.60	19.47	200	19.20	10.40	220	16.20	14.80	1.09	58	59	26	143	18.75	5.45	25	19	10	16
22	3.960	0.85	1.00	1.37	24.19	180	18.50	9.80	240	17.40	14.50	1.20	53	45	30	128	14.50	4.16	22	26	16	22
23	1.888	0.43	0.43	1.14	13.10	118	15.60	9.50	210	16.20	13.70	1.18	56	49	28	133	14.50	3.93	46	37	14	16
24	0.600	0.44	0.51	1.00	14.32	85	13.20	7.80	170	14.30	13.50	1.06	61	37	34	132	22.20	4.32	50	39	17	17
25	0.680	0.46	0.62	1.30	18.61	88	20.70	13.50	140	14.50	11.80	1.22	62	51	25	138	22.16	3.74	38	28	10	13
26	0.724	0.68	0.77	1.10	26.60	98	15.80	10.00	220	16.40	14.30	1.15	58	44	34	136	21.92	3.10	24	16	10	11
27	0.552	0.79	0.85	1.07	21.40	86	12.30	11.80	320	19.38	14.30	1.35	58	53	24	135	21.11	3.44	26	19	14	15
28	2.124	0.72	0.79	1.09	25.10	98	18.12	12.38	230	17.70	15.00	1.18	56	46	28	130	19.40	4.20	40	33	22	24
29	3.920	0.80	0.80	1.26	19.24	205	19.05	10.22	222	16.15	14.20	1.14	56	57	25	138	18.65	5.41	38	30	19	24
30	3.864	0.92	1.10	1.35	23.33	182	18.32	9.66	238	17.47	14.70	1.19	51	43	29	123	14.31	4.12	23	25	17	23

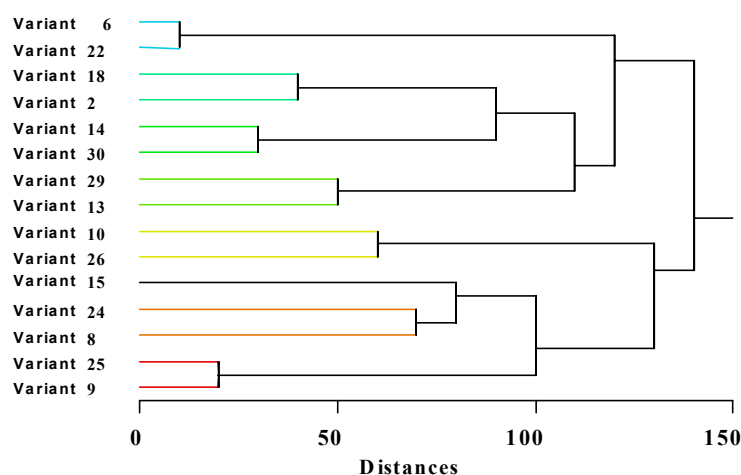
*Traits:* 1. Yield, kg; 2. Coefficient of shoot fertility; 3. Coefficient of main shoot fertility; 4. Coefficient of fruiting shoot fertility; 5. Millerandage berries, %; 6. Average bunch weight, g; 7. Bunch length, cm; 8. Bunch width, cm; 9. Average weight of 100 berries, g; 10. Berry length, mm; 11. Berry width, mm; 12. Index of berry shape; 13. Budding-flowering (days); 14. Flowering-softening (days); 15. Softening-technological maturity (days); 16. Budding - technological maturity (days); 17. Sugars, %; 18. Acids, g/dm<sup>3</sup>; 19. Total bud number; 20. Total shoot number; 21. Total fruiting shoot number; 22. Total bunch number.

**Table 2**  
**Results from Principal Component Analysis of F<sub>1</sub> progeny**  
**of the hybrid combination Super Early Bolgar x Ruby Seedless**

Traits	Principal components						
	1	2	3	4	5	6	7
1	0.852	-0.124	-0.190	-0.313	0.265	-0.096	-0.008
2	0.772	0.242	-0.036	-0.450	-0.152	-0.079	0.148
3	0.434	0.457	-0.240	-0.454	0.189	-0.210	0.192
4	0.257	0.022	0.382	-0.685	-0.054	0.314	-0.149
5	0.563	0.156	0.442	0.400	0.047	-0.140	0.065
6	0.466	0.393	0.102	0.454	0.003	-0.308	0.164
7	0.208	0.364	0.757	0.143	-0.231	-0.036	-0.030
8	0.564	-0.675	0.247	0.223	0.119	0.213	0.126
9	0.716	-0.573	0.177	0.177	0.196	0.203	-0.016
10	0.594	-0.491	0.462	0.134	-0.006	0.261	0.155
11	0.595	-0.439	-0.277	0.160	0.395	0.022	-0.258
12	0.036	-0.142	-0.744	0.187	0.260	-0.233	0.443
13	-0.220	0.450	0.677	-0.042	0.495	-0.058	-0.107
14	0.103	-0.368	-0.317	0.085	-0.515	0.428	0.389
15	-0.308	0.346	0.296	0.147	0.528	0.183	0.539
16	-0.472	-0.114	-0.108	-0.400	0.528	0.325	0.049
17	0.154	0.568	0.134	-0.334	-0.196	0.220	0.392
18	-0.159	0.666	-0.181	0.415	0.155	0.463	-0.042
19	-0.118	0.771	-0.339	0.310	-0.084	0.318	-0.187
20	0.695	0.373	-0.346	0.123	0.057	0.244	-0.285
21	0.774	0.450	-0.368	-0.041	0.104	0.107	-0.149
22	0.784	0.521	0.064	0.031	-0.133	-0.083	0.096
Explained % of the total variation	26.500	19.430	13.820	9.586	7.414	5.685	5.340

greatest extent with the first and second principal components. According to group formation only plant 22 from the first subgroup should participate in hybrid combinations with plants from the fourth subgroup. The remaining plants should be taken into consideration during selection.

As shown on the dendrogram, which presents cluster formation of seedless seedlings, the seedlings are divided into two large groups and three subgroups including a total of 15 plants (Figure 3). The phenotypic differentiation at the highest level demonstrates the genotypic similarity



**Fig. 2. Dendrogram of cluster analysis of the genotypes of F<sub>1</sub> progeny of the hybrid combination Super Early Bolgar x Ruby Seedless – seeded plants**

of the seedless plants from this cross. On the first level of similarity plants 19, 3, 1, 17 relate to the first subgroup; plants 16, 7, 23, 20, 4, 5, 21 – to the second subgroup and plants 28, 12, 11, 27 – to the third subgroup. The seedlings in these subgroups are characterized by considerably close values of agrobiological indices and they could be treated as identical in practical selection. The relative distances between individual variants are small – from 0.000 to 97.841.

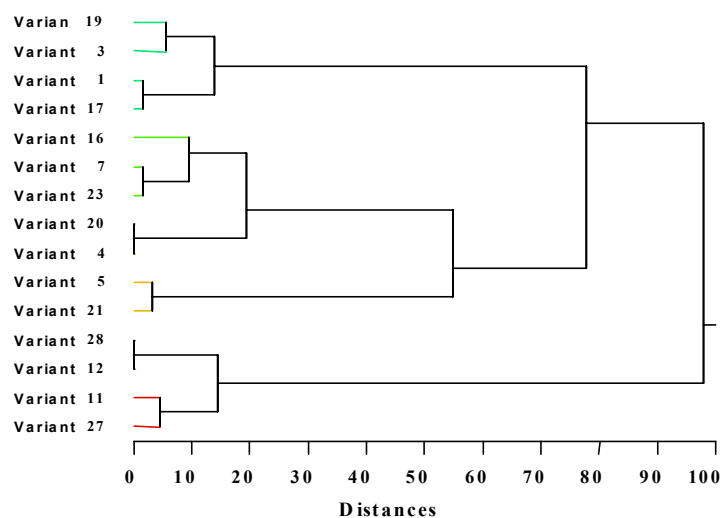
The results from the principal component analysis show that five of them are enough to explain 91.661 % of the total variation (Table 4). The traits: total bunch number, millerandage berries, bunch length, coefficient of shoot fertility, yield, total fruiting shoot number and berry length explain 37.788 % of the total variation in the seedless plants' population. They exert the greatest influence on the division of genotypes into groups, since their relative variation rate correlates to the greatest extent with the first principal component.

The traits: acids, total bud number,

bunch width and berry width possess the highest correlation coefficients with the second principal component. These traits explain 25.312 % of the total variation. The third principal component explains 12.158 % of the total variation, mainly by means of the traits: budding-flowering, softening-technological maturity and flowering-softening. The other two components – the fourth and the fifth one – explain respectively 9.496 % and 6.907 % of the total variation, the traits: coefficient of main shoot fertility and average bunch weight having the highest correlation coefficients.

Owing to peculiarities in the cytoembryological processes which determine seedlessness in vine, the hybridization between the most genetically remote seedlings from the first and the third group could not be carried out under field conditions. The results from both analyses show that the most efficient selection in seedless seedlings could be performed according to the traits possessing the highest correlation with the first and the second principal component.





**Fig. 3. Dendrogram of cluster analysis of the genotypes of  $F_1$  progeny of the hybrid combination Super Early Bolgar x Ruby Seedless – seedless plants**

In the studied seedless plants' population only seedling 21 comes close to the mentioned criteria in phenotype.

The application of cluster analysis and principal component analysis in studies of  $F_1$  seedlings of the hybrid combination Super Early Bolgar x Ruby Seedless allows the obtaining of more detailed information on the significance of individual traits for the division of genotypes into groups. Judging by the obtained results it could be assumed that the selection will be most efficient when aiming at increasing the values of the traits: total number of bunches and fruiting shoots, bunch length and berry length, yield, coefficient of shoot and main shoot fertility, bunch width, acids and total bud number, index of berry shape; and at the same time aiming at decreasing the values of the traits millerandage berries and budding-flowering.

Very good selection results will be achieved if the selection in seeded and seedless seedlings is carried out according

to the traits, which explain the highest percentages of the total variation in the first and second principal component. Conditionally they could be considered as selection indices of the first and second rank, and the others – of the third rank. According to the estimated genetic similarity, the seedlings from the first and the third or fourth subgroups in this hybrid combination are the most relatively remote. Depending on the character of traits on different levels and on the concrete selection objective, higher or lower values of these traits could be desired and pursued. In almost each case separate plants, which should also be taken into consideration during selection, remain outside this conditional scheme. Their presence shows that despite the accuracy of the methods, improvisation as a part of the selection of elite forms should not be underestimated or ignored, since numerous factors influence the phenotype of vine seedlings.

In most cases the traits in the first



**Table 3**  
**Results from Principal Component Analysis of F<sub>1</sub> progeny of the hybrid combination**  
**Super Early Bolgar x Ruby Seedless – seeded plants**

Traits	Principal components				
	1	2	3	4	5
1	0.782	0.468	-0.224	0.326	-0.107
2	0.562	0.537	-0.55	0.001	0.258
3	0.477	0.797	-0.079	0.212	0.08
4	-0.445	0.629	-0.173	-0.044	0.533
5	0.255	-0.374	0.655	0.453	-0.094
6	0.691	-0.37	0.333	-0.521	-0.034
7	-0.712	-0.295	-0.064	0.297	0.502
8	0.505	-0.747	0.044	0.08	0.351
9	0.732	-0.503	0.189	0.213	0.272
10	0.302	-0.86	-0.07	-0.231	0.263
11	0.769	-0.025	0.317	0.481	0.175
12	0.958	-0.039	-0.094	-0.054	-0.102
13	-0.845	0.324	0.356	0.082	-0.111
14	0.507	-0.432	-0.041	-0.217	0.609
15	-0.491	0.176	0.636	-0.1	0.357
16	-0.537	0.343	0.048	0.698	0.304
17	-0.186	0.55	-0.514	-0.491	0.321
18	-0.131	0.366	0.794	-0.38	0.095
19	0.009	0.643	0.657	-0.293	0.183
20	0.733	0.573	0.231	-0.055	-0.151
21	0.734	0.663	0.122	0.032	-0.032
22	0.744	0.584	0.151	0.025	0.257
Explained % of the total variation	36.246	26.616	13.804	9.469	8.072

principal component are almost identical in the population and after its division into two plant groups. In the following components the differences are significant and these components include other and fewer traits. When the analyses are carried out with the participation of all F<sub>1</sub> plants, the number of groups and subgroups is

relatively larger. In this case more seeded plants are found to be of a greater economic value and interest. In seedless seedlings there are relatively fewer traits with high correlation coefficients and in the general plant population there are more principal components. The phenotypic equalization in seedless seedlings means

**Table 4**  
**Results from Principal Component Analysis of F<sub>1</sub> progeny of the hybrid**  
**combination Super Early Bolgar x Ruby Seedless – seedless plants**

Traits	Principal components				
	1	2	3	4	5
1	0.803	-0.457	0.168	0.101	0.192
2	0.848	0.172	-0.127	-0.354	-0.212
3	0.295	0.398	0.263	-0.654	0.036
4	0.464	-0.522	-0.099	-0.48	-0.166
5	0.876	0.297	0.132	-0.24	0.126
6	0.511	0.614	0.135	0.137	-0.502
7	0.865	0.437	0.051	-0.036	-0.074
8	0.572	-0.721	0.119	0.271	0.229
9	0.654	-0.676	0.032	0.287	0.146
10	0.76	-0.495	0.045	0.131	0.379
11	0.195	-0.72	-0.002	0.459	-0.436
12	-0.745	0.293	0.385	0.305	0.039
13	0.477	0.369	0.716	0.225	-0.072
14	-0.274	-0.284	-0.61	-0.454	0.476
15	-0.171	0.513	0.682	0.093	0.454
16	-0.724	-0.121	0.04	0.109	0.25
17	0.455	0.708	0.055	0.16	0.454
18	-0.203	0.793	-0.472	0.24	0.081
19	-0.104	0.741	-0.594	0.254	-0.02
20	0.647	0.132	-0.506	0.482	0.101
21	0.785	0.287	-0.4	0.245	0.05
22	0.891	0.375	-0.032	-0.153	0.03
Explained % of the total variation	37.788	25.312	12.158	9.496	6.907

that their selection will last longer, seedlessness being a qualitative trait which strongly influences the agrobiological indices of hybrid forms. The distance from the borderline of complete similarity between separate plants considerably decreases from the general population to the seedless forms' group. The division of

plants into groups of seeded and seedless provides greater clarity of the results from both methods and enhances selection efficiency.

The combined application of cluster analysis and principal component analysis in F<sub>1</sub> progeny of crosses between seeded and seedless table cultivars allows the

selection of plants, proved to possess valuable economic characteristics. The obtained results show that selection efficiency will be considerably higher when the studied population is divided into two groups – of seeded and seedless plants. Also, the choice of appropriate future parents, when one is required for selection purposes, is more precise and accurate. Selection is aided by the division of seedlings with similar trait values into separate groups, but the revealing of an elite form is mostly possible through analyzing the results from both methods used after the division of the population according to the traits seededness-seedlessness.

### Conclusions

In the entire plant population in  $F_1$  progeny of the hybrid combination Super Early Bolgar x Ruby Seedless, the selection will be most efficient when aiming at improving the values of the traits: yield, total bunch and fruiting shoot number, coefficient of shoot fertility and average weight of 100 berries, since these traits influence to the greatest extent group formation of genotypes. Seven of the principal components explain 87.775 % of their total variation. In order to obtain new elite forms in future selection research, seedlings from the first and the fifth subgroup should be used, since they are the most genetically remote.

In seeded seedlings 94.207 % of the total variation of traits is explained by five principal components. The most important of these are: index of berry shape, budding-flowering, yield, total number of bunches-fruiting shoots-shoots, and bunch length. Seedlings 8, 22, 29 and 30 possess the highest values of these traits, correlating

most with the first and second principal component. According to the division of plants into groups, only plant 22 - from the first subgroup - should participate in hybrid combinations with plants from the fourth subgroup. The remaining plants should be taken into consideration during selection.

Five principal components explain 91.661 % of the total variation of traits in seedless seedlings. 37.788 % of them are determined by: total bunch number, millerandage berries, bunch length, and coefficient of shoot fertility, yield, total fruiting shoot number and berry length. According to the criteria of the methods, only seedling 21 out of the seedless plants' population could be selected as a promising hybrid form.

The application of cluster analyses and principal component analyses in the investigations of  $F_1$  progeny of crosses between seeded and seedless vine cultivars allows the obtaining of more profound ampelographic information, concerning the character and significance of separate traits for the grouping of genotypes. The selection of elite hybrid forms should be carried out according to the traits, which explain to the greatest extent the total variation in the first and second principal component. The division of seedlings into two groups – seeded and seedless – increases the efficiency of both methods in the development of new table vine cultivars with valuable economic characteristics.

### References

- Altube, H., F. Gabello and J. M. Ortiz, 1991. Caracterización de variedades y portainjertos de vid mediante isoenzimas de los samientos. *Vitis*, **30**: 203-212.
- Bassi, D., O. Silvestroni and S. Sansavini,

1995. Cultivar identification and numerical taxonomy in grape and fruit crops. *Agro Bio Frut*, Cesena, Italia. **6 maggio**: 47-57.
- Boselli, M., C. Corso, A. Monaco, A. Bouquet and J. M. Boursiquot**, 2000. Ampelographic characterization of white grape varieties in Campania (Southern Italy) by multivariate analysis. Proceedings of the Seventh International Symposium on Grapevine Genetics and Breeding, Montpellier, France, 6-10 July 1998, 1, *Acta Horticulturae*, **528**: 75-81.
- Cameiro, L. C. and M. B. Lina**, 1989. Comparison of fresh and dried leaves of grapevine for ampelographic purposes using numerical taxonomy. *Rivista di Viticoltura e di Enologia*, **1**: 53-58.
- Carreno, J., L. Almera, A. Martinez and J. A. Fernandez-Lopes**, 1997. Chemotaxonomical Classification of Red Table Grapes based on Anthocyanin Profile and External Colour. *Lebensmittel-Wissenschaft und Technologie/FST*, **30** (3): 259-285.
- Cervera, M. T., J. A. Cabezas, I. Rodriguez-Torres, J. Chaves, F. Cabello and J. M. Martinez-Zapater**, 2002. Varietal diversity within grapevine accessions of cv. Tempranillo. *Vitis*, **41**(1): 33-36.
- Cox, T. S., Y. T. Kiang, M. B. Gorman and D. M. Rodgers**, 1985. Comparison of coefficient of parentage and genetic similarity estimates in the soybean. *Crop. Sci.*, **25**: 529-532.
- Duran, B. S. and P. L. Odell**, 1974. Cluster analysis: a survey. *Springler-Verlag*. Berlin and New York.
- Dubles, R. and A. K. Jain**, 1980. Cluster methodologies in Exploratory Data analysis. Advances in computers, *Academic Press*, New York, **19**: 113-228.
- Everitt, B. S.**, 1979. Unresolved Problems in Cluster Analysis. *Biometrics*, **35**: 169-181.
- Ferreira, V., P. Fernandez and J. F. Cacho**, 1996. A Study of Factors Affecting Wine Volative Composition and Application in Discriminant Analysis. *Lebensmittel-Wissenschaft und Technologie/FST*, **29** (3): 251-259.
- Gao, Yu and G. A. Cahoon**, 1998. Cluster Thinning Effects on Fruit Weight, Juice Quality and Fruit Skin Characteristics in Reliance' Grapes. *Eruit Crops: A Summary of Research 1998*, Research Circular 299-99, pp. 1-8.
- Goodman, M. M.**, 1973. Genetic distances: Measuring dissimilarity among populations. *Jearb. Phys. Anthropol.*, **17**: 1-38.
- Goto-Yamamoto, N.**, 2000. Phenolic clustering of grapes (*Vitis* spp.) by AFLP analysis. *Breeding-Science*, **50** (1): 53-57.
- Hvarleva, T., K. Rusanov, F. Lefort, I. Tsvetkov, A. Atanassov and I. Atanassov**, 2004. Genotyping of Bulgarian *Vitis vinifera* L. Cultivars by microsatellite analysis. *Vitis*, **43** (1): 27-34.
- Hvarleva, T. and A. Atanassov**, 2006. Establishment of genetic authenticity of vine cultivars by means of DNA-analysis. *Viticulture and Wine Research*, **5**: 23-30.
- Judez, L., J. Iitago, J. Yuste, A. Soldevilla and F. Martinez**, 1995. Statistical procedure to guide the first stages of clonal selection of the variety Tinta del Pais. *Journal International des Sciences de la Vigne et du Vin.*, **29** (4): 183-191.
- Krastanova, S.**, 1986. Investigation of intraspecific diversity of some economic traits in the vine variety Pamid (*Vitis vinifera* L.) aiming at genetically improving it by means of clonal selection. Dissertation, Plovdiv, 170 pp. (Bg).
- Labra, M., G. Moriondo, A. Schneider, F. Grassi, O. Failla, A. Scienza and F. Sala**, 2002. Biodiversity of grapevines (*V. vinifera* L.) grown in the Aosta Valley. *Vitis*, **41**(2): 89-92.
- Mattheou, A., N. Stavropoulos and S. Samaras**, 1995. Studies on table grape

- germplasm grown in northern Greece- II. Seedlessness, berry and must characteristics, *Vitis*, **34** (4): 217-220.
- Molle, E. and S. Krastanova**, 1987. Application of a clustering method in vine selection. *Horticultural Sciences*, **11**, XXIV: 81-85.
- Moreno-Arribas, M. V., F. Cabello, M. C. Polo, P. J. Martín-Alvarez and E. Puevo**, 1999. Assessment of the native electrophoretic analysis of total grape must proteins for the characterization of *V. vinifera* L. cultivars. *3 Agric Food Chem*, **47** (1): 114-120.
- Murphy, J. P., T. S. Cox and D. M. Rodgers**, 1986. Cluster Analysis of Red Winter Wheat Cultivars. *Crop. Sci.*, **26**: 672-676.
- Ortis, J. M., F. Cabello, H. Altude and V. Sotes**, 1990. Multifactorial analysis of leaf parameters for identification of *Vitis* cultivars. *Riv. Vitic. Enol.*, **1**: 3-8.
- Philippeau, G.**, 1990. In "Principal component analysis: How to use the results". ITCF, Paris, p. 9.
- Rodríguez, C. M., J. L. G. Mantilla and D. L. Rodríguez**, 1994. Description ampelographica sobre Hoja Adulta de cepas de *Vitis vinifera* L. denominadas "Albarino". *Viticultura/Enología Profesional*, **34**: 30-40.
- Spagnoletti, Zeuli P. L. and C. O. Qualset**, 1987. Geographical Diversity for Quantitative Spike Characters in a World Collection of Durum Wheat. *Crop. Sci.*, **27**:235-241.
- Svanepel, J. J. and C. E. Villiers**, 1987. A numerical-taxonomic classification of *Vitis* spp. and cultivars based on leaf characteristics. *South African Journal for Enology and Viticulture*. Stellenbosch., **8**: 31-35.
- Tomazic, I. and Z. Korosec-Koruza**, 1998. Grapevine variety identification by means of ampelographic and phyllometric descriptors. *Research Reports Biotechnical Faculty-University of Ljubljana*, Agricultural Issue., **71**: 173-177.
- Troshin, L. P., A. A. Poluljah and V. I. Risovannaia**, 1998. An assessment of the taxonomic relations of the vine cultivars *V. v. s. p. balkanica* Negr. and *V. v. s. p. meridionali balcanica* Trosh. based on a morphological leaf trait. *Grape and Wine of Russia*, **3**: 41-42.
- Varga, Z. S., F. Bünyei and A. Ferenczy**, 2006. Examination of the relations of origin of old grape vine cultivars by cluster analysis in Tokaj. 9<sup>th</sup> International Conference on Grape Genetics and Breeding, Udine, Italy 2-6 July, 2006.
- Ward, J. H.**, 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**: 236-244.
- Yao-YunCong, Wang-YouNian, Shen-RenQian, Yao-YC, Wang-YN, Shen-RQ and Zhu-DW**, 1995. Fuzzy clustering

Received May, 12, 2007; accepted July, 23, 2007.