Quantitative image analysis of berry size and berry shape of different grapevine (*Vitis vinifera* L.) accessions

Péter Bodor-Pesti[1], Eszter Somogyi[1], Tamás Deák[1], Diána Ágnes Nyitrainé Sárdy[1] and Márta Ladányi[2]

[1]Institute of Viticulture and Enology, Hungarian University of Agriculture and Life Sciences
HU-1118 Budapest, Villányi Str. 29-43
[2]Department of Applied Statistics, Institute of Mathematics and Basic Science, Hungarian University of Agriculture and Life Sciences
HU-1118 Budapest, Villányi Str. 29-43
E-Mail: nyitraine.sardy.diana.agnes@uni-mate.hu

Morphological diversity of the grapevine (*Vitis vinifera* L.) berry size and berry shape were investigated by means of digital image analysis. Thirteen genotypes, representing a large morphological variability, were evaluated with the Tomato Analyzer software. For each accession, 50 berries were analyzed with respect to 21 traits, in particular berry weight, seed number and 19 descriptors related to the size and shape of the berries. Statistical analysis revealed significant uvometric differences between the genotypes. According to discriminant analysis (DA), the first five canonical functions (DF1 to DF5) were significant and among these, the first three explained 91 % of the variance. DA classification was carried out at genotype and group-of-genotypes level. The average correct classification of the samples based on the berry traits was 55.3 % on the individual genotype level and 88.4 % if genotypes with similar berry morphology were grouped. We demonstrate that uvometric traits represent similar groups of genotypes more effectively than the individual accessions. Our study revealed that image analysis is a powerful and timesaving tool for grapevine uvometric characterization.
Keywords: morphology, uvometry, diversity, discriminant analysis, digital image analysis

Quantitative Bildanalyse der Beerengröße und Beerenform verschiedener Rebgenotypen (*Vitis vinifera* L.). Die morphologische Vielfalt der Beerengröße und Beerenform der Weinrebe (*Vitis vinifera* L.) wurde in dieser Studie mittels digitaler Bildanalyse untersucht. Dreizehn Genotypen, die eine große morphologische Variabilität darstellen, wurden mit der Tomato-Analyzer-Software ausgewertet. Für jede Akzession wurden 50 Beeren nach 21 Merkmalen analysiert, insbesondere Beerengewicht, Samenzahl und 19 Deskriptoren bezüglich der Größe und Form der Beeren. Die statistische Analyse ergab signifikante Unterschiede von Beereneigenschaften zwischen den Genotypen. Nach der Diskriminanzanalyse (DA) waren die ersten fünf kanonischen Funktionen (DF1 bis DF5) signifikant, und von diesen erklärten die ersten drei 91 % der Varianz. Die durchschnittliche korrekte, auf Beerenmerkmalen basierende Klassifizierung der Genotypen betrug 55,3 % auf Genotyp-Ebene und 88,4 %, wenn Genotypen mit ähnlicher Beerenmorphologie zusammengefasst wurden. Das Ergebnis zeigt, dass uvometrische Merkmale Genotyp-Gruppen mit ähnlichen Merkmalen effektiver repräsentieren als die einzelnen Akzessionen. Unsere Studie zeigte, dass die Bildanalyse ein leistungsstarkes und zeitsparendes Werkzeug für die Charakterisierung von Beereneigenschaften ist.
Keywords: Morphologie, Uvometrie, Diversität, Diskriminanzanalyse, digitale Bildanalyse

Ampelography involves the morphological and phenological descriptions of grape (*Vitis* sp.) genotypes, including traits that are important in identification, vineyard maintenance and winemaking. According to the International Organization of Vine and Wine (OIV), there are more than 100 trait descriptors defined and standardized for shoot, leaf, bunch, berry and seed, which most frequently refer to size, shape and number of the grape organs (OIV, 2009). Measurement-based evaluation of grapevine is usually referred to as ampelometry. Most studies associate the expression with leaf morphometry (Ravaz, 1902), although it can be considered as an umbrella term of methods aiming to describe the leaf (foliometry), the flower (florimetry), and the seed (carpometry) according to measurable size and shape attributes. Among the metric methods, uvometry deals with the analysis of bunch and berry, based on weight, seed number, shape, size and further physical parameters of the fruit.

Berry size is a priority parameter in viticulture as it influences yield and bunch compactness (Tello and Ibanez, 2018), besides that it is considered a key factor in winemaking affecting color of the must through the pulp-to-skin ratio (Barbagallo et al., 2011). Moreover, in fresh grape production and breeding, size is one of the main targets, as consumers prefer large berries (Piva et al., 2006). However, although digital image analysis is widely applied in horticultural crop phenotyping, individual grapevine berry size is in most studies evaluated by means of manual measurements, simply based on the maximum length and width of the fruit. However, size and shape of the berry are much more complex, so the evaluation of further traits is also recommended. For this reason, we aimed to describe berry morphological diversity of 13 grape accessions according to traditional morphometry using the Tomato Analyzer software (Brewer et al., 2006). The purpose of our study was to show the effectiveness of this digital image analysis software as an assisting tool for grapevine breeding, clonal selection and traditional ampelographic observations. We also aimed to analyze the berry size attributes influenced by the seed numbers and find those characteristics that could discriminate grapevine genotypes or groups of genotypes.

Materials and methods

Plant material

Thirteen grapevine genotypes representing a large berry size and shape variability were investigated: '13/10', 'Cornichon-like' (COR), 'Ferenc József' (FER), 'Gyűszű szőlő' (GYU), 'Halhólyag fehér' (HAL), 'Italia' (ITA), 'KM.193', 'KM.238', 'Mecsta' (MEC), 'Perlona' (PER), 'Szusenszkij belüj' (SZU), 'Usztojcsivüj gyikij' (USZ) and '*Vitis* typ. Weiss' (VIT). Fifty berries from at least ten bunches of each genotype were collected in the gene bank of the Research Institute for Viticulture and Enology, Hungarian University of Agriculture and Life Sciences (Kecskemét, Hungary, VIVC: HUN047) in 2020 at full ripeness. Berries were stored in plastic boxes at 4 °C until further investigations.

Morphological characterization

Individually labeled berries were weighed on an Ohaus Explorer Pro EP114C analytical scale (Ohause Corporation, Pine Brook, NJ, USA). Berries were then cut into half and seed number was counted and recorded. Digitalization of one section per each berry was carried out with an Epson V370 scanner (Seiko Epson Corporation, Suwa, Japan) at 200 dpi. TIFF type images were analyzed with the software Tomato Analyzer 3.0 (TA) (The Ohio State University, Columbus, OH, USA, 2021) according to the protocol reported by Brewer et al. (2006). Nineteen attributes were included in this study: basic measurements (perimeter, area, width mid-height, maximum width, height mid-width, maximum height, curved height, curved width) (Fig. 1), fruit shape indices (fruit shape index external I (FSIE I), fruit shape index external II (FSIE II), curved fruit shape index (CFSI)), blockiness (proximal fruit blockiness (PFB), distal fruit blockiness (DFB), fruit shape triangle (FST)) and asymmetry (obovoid, ovoid, v.asymmetry, h.asymmetry.ob, h.asymmetry.ov, width widest position). To avoid any ambiguities in the denomination of the attributes, original variable names from the TA user manual are used consequently in the paper. Definitions of the traits are detailed in Hurtado et al. (2013).
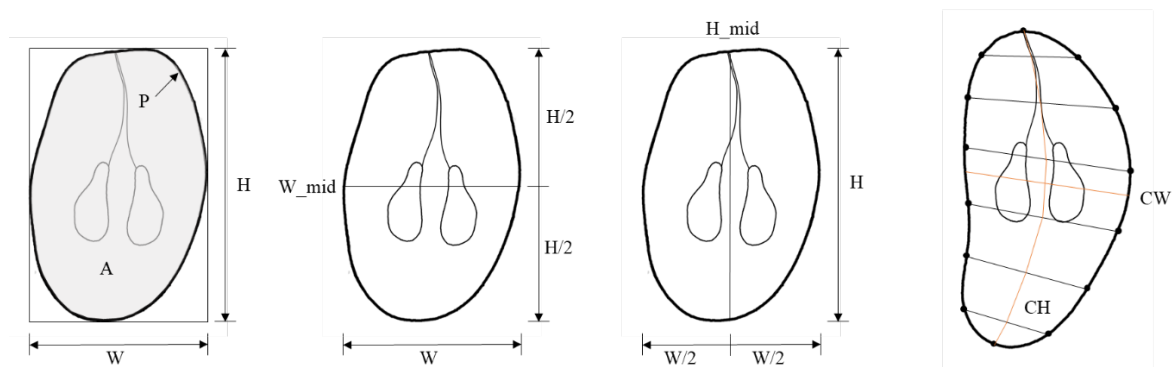
Fig. 1: Basic measurements performed in this study: P = perimeter, A = area, H = maximum height, W = maximum width, H_mid = width mid-height, W_mid = height mid-width, CH = curved height, CW = curved width (according to the manual of the Tomato Analyzer)

Statistical analysis

Mean, standard deviation and the coefficient of variation $(\mathrm{CV} = 100 * \mu/\sigma\,[\,\%\,])$ were calculated for each accession and for the whole sample set. The multivariate outliers were detected by Mahalanobis distance and those with significance rate $p < 0.001$ were eliminated. The total number of eliminated outliers was 8 and no more than 2 were eliminated from a set of one accession. Pearson's correlation was calculated to investigate the relation between seed number and berry weight that were previously considered as normally distributed as the absolute values of their skewness and kurtosis were both below 1. Berry weight, maximum width and maximum height of the berries were compared along the seed numbers by one-way ANOVA. The normality of the residuals was accepted again by skewness and kurtosis and the homogeneity of variances was tested by ratio test (maximum variance/minimum variance values were below 3). Discriminant analysis was performed to explain the difference of accessions by the observed traits. The number of variables was reduced to 12 in order to avoid serious collinearity. The variables that were highly significantly correlating with another variable were excluded. To analyze and visualize the relation of seed number and berry weight, quantile regression was performed. The significances of the linear trends of the quantiles 5 %, 10 %, 25 %, 50 %, 75 %, 90 % and 95 % were tested by Student's t-test.

Results and discussion

Uvometric traits (weight, size and shape) influenced by genotype and seed number

In this study the average berry weight ranged from 3.02 g ('COR') to 6.46 g ('KM.193') and the same pattern was observed when the ten largest berries were evaluated: that of 'COR' was 3.9 g and of 'KM.193' it was 8.75 g. We assessed berry size variability according to Bioletti (1938), evaluating the *W/w* quotient, where *W* = is the mean weight of the ten largest berries, while *w* = is the mean weight of all the berries. The lowest variability was observed in the case of 'COR' (1.22) while the highest was found with 'GYU' (1.41). This is in accordance with the berry weight coefficient of variation for which 'COR' had the lowest CV value (15.77 %) and the most variable accession was 'GYU' (28.7 %). Berry size was evaluated according to traits linked to perimeter, area, length and width. Perimeter and area were the largest in the case of 'KM.193' (81.86 mm and 450.98 mm², respectively) while the smallest average values were recorded for 'SZU' (63.1 mm and 281.73 mm², respectively). Length of berries was measured as the maximum height and as the height intercepting the middle-point of the maximum width. The longest berries belonged to 'USZ' (27.48 mm) while the shortest were the berries of accession 'SZU' (19.72 mm). The widest berries were obtained from 'KM.193' (20.41 mm) while the narrowest were the berries of accession 'HAL' (15.73 mm). The curved height of the berries was defined as the length of the berries along the midpoints at different positions of the berries. The longest curved height belonged to 'USZ' (31.1 mm), the shortest to 'SZU'

(24.5 mm). In this study, seed numbers were significantly different along the genotypes (F(4;644)=32.00; p<0.001). The average seed number in the sample set was 1.63, the lowest was 1.36 ('VIT' and '13/10') while the highest was 2.04 ('FER'). Seed number had significant effect on both weight and size of the berries. The weight of the seedless berries was as low as 2.87 g in average, while the largest berries were those with four and five seeds (5.98 g and 5.95 g, respectively). The coefficient of variation of berry weight was the highest in the case of the seedless berries with 37.71 % while more uniform berry weight was recorded in the case of the one-seeded berries (28.56 %). In accordance with Mullins et al. (1992) and Barbagallo et al. (2011), we found a significant positive correlation (r=0.406; p<0.001) between the seed number and the weight of the individual berries. Regarding the correlation of berry weight and seed numbers, our quantile regression analysis (Fig. 2) showed a significant positive slope at all percentiles (p<0.05). Similarly to berry weight, size traits correlated also positively with seed numbers (maximum width r=0.38, p<0.01;

maximum height r=0.26, p<0.01) and the traits showed significant differences between the berries with different seed number (maximum width: F(4;644)=27.99; p<0.001; maximum height: F(4;644)=14.29; p<0.001).

Among the uvometric traits, berry shape is also an important aspect of the consumers' fruit purchase; therefore, genotypes with attractive forms are spreading (Ferrara et al., 2017). The diversity in the fruit shape showed that 'VIT' is the least variable in shape (CV =4.8 %), while 'USZ' is the most variable (CV =10.19 %). The mean berry shape index of 'SZU' was the lowest (FSIE I=1.08; FSIE II=1.08) while the highest values belong to 'HAL' (FSIE I=1.52; FSIE II=1.53). The same results were obtained in the case of the CFSI. The PFB and the DFB showed low variability in the sample set ranging from 0.55 to 0.66 ('COR' and 'KM.238', respectively) and from 0.57 to 0.65 ('13/10' and 'COR', respectively). The FST showed higher variability ranging from 0.85 ('COR') to 1.12 ('13/10'). The asymmetry of the samples showed that 'COR', 'FER', 'ITA' and 'SZU' have obovoid berries (wider on the bottom than on the top) while the other accessions are ovoid.

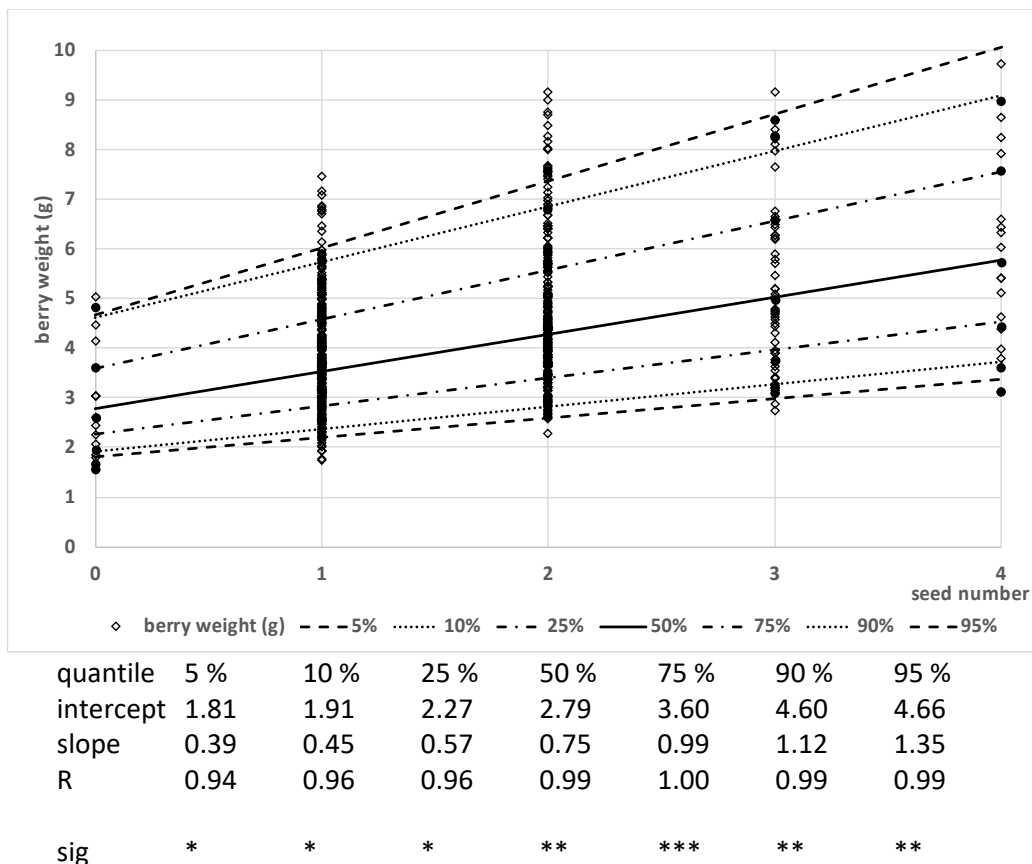| quantile	5 %	10 %	25 %	50 %	75 %	90 %	95 %
intercept	1.81	1.91	2.27	2.79	3.60	4.60	4.66
slope	0.39	0.45	0.57	0.75	0.99	1.12	1.35
R	0.94	0.96	0.96	0.99	1.00	0.99	0.99
sig	*	*	*	**	***	**	**

Fig. 2: Seed number and berry weight (g) of the observed accessions with the results of their quantile regression (intercepts together with slopes of the linear trends and Pearson's correlations); the slopes and correlations are significant at *p<0.05, **p<0.01, ***p<0.001.

Multivariate classification of varieties based on berry morphometrics

To explore the discriminative potential of berry morphological traits, multivariate Z-score based linear discriminant analysis was carried out with

.

those 12 traits that showed significant differences between the samples (Table 1).

Table 1: Tests of group means equality (Wilk's lambda values with F values and their significance levels) and the discriminant function loadings with absolute values above 0.3 (main loadings) for the discriminant analysis model with 13 accessions and for the one with four similarity groups; correct classification rates and their cross-validated rates are indicated.

Traits	Accession groups (13) Correct classification rate: 55.3% Leave-one-out cross-validation: 51.1%					Similarity groups (4) Correct classification rate: 88.43% Leave-one-out cross-validation: 87.2%				
	Tests of Equality of Group Means		Main Loadings			Tests of Equality of Group Means		Main Loadings		
	Wilk's lambda	F(12;629)	DF1	DF2	DF3	Wilk's lambda	F(3;489)	DF1	DF2	DF3
FSIE II	0.25	154.14***	0.88			0.31	367.35***	0.83		
CFSI	0.39	82.91***	0.66			0.44	204.91***	0.64		
curved height	0.46	61.23***	0.39	-0.53		0.48	176.18***	0.49		
maximum width	0.47	59.32***		-0.68		0.53	146.27***		-0.68	
berry weight	0.50	52.02***		-0.68		0.54	137.09***		-0.63	
area	0.48	55.81***		-0.67		0.51	154.82***		-0.60	
perimeter	0.53	46.89***		-0.59		0.54	137.93***		-0.53	
PFB	0.70	22.62***			-0.35	0.69	72.85***			0.66
width widest	0.70	22.50***			0.59	0.74	57.46***			-0.55
seed number	0.90	5.88***			0.30	0.95	8.27***			-0.36
DFB	0.69	23.24***			0.74	0.83	33.62***			-0.40
v, asymmetry	0.85	9.17****	0.16†			0.89	57.46***	0.19†		

† even the highest loading is below 0.20 *** significant at p < 0.001

Wilks' lambda values of the first five discriminant functions (DF1 to DF5) were significant (p<0.001). Based on the Wilks' lambda values below 0.4 (p<0.001) and the explained variance ratios, three discriminant functions were involved in the analysis with canonical correlations 1 to 3 from r=0.88 to r=0.60. Discriminant function 1 (DF1) explained 53.7 % of the total variance, while DF2 and DF3 explained 29.2 % and 8.3 %, respectively. The first three DF explained more than 91 % of the total variance.

In Table 1, loadings with absolute values above 0.3 show that DF1 is highly correlated with FSIE II, CFS and curved height, while DF2 with curved height, maximum width, berry weight, area and perimeter, and DF3 is highly correlated with PFB, width widest pos., seed number and DFB. trait v. asymmetry is not highly correlated with any of the DFs. The classification showed that 55.3 % of

the samples (i.e. berries) were grouped correctly. The leave-one-out cross-validated classification was close to this successful classification rate with its 51.1 %. Accession 'COR' was classified correctly with the highest rate (88.0 %) followed by '10/13' (62.0 %) and 'FER' (59.2 %). The highest misclassification was observed in the case of 'PER' and 'USZ' where 68.0 % of the samples were misclassified.

We found that the correct classification rate is not high in this model, however, we noticed that the misclassifications show a well-detectable pattern: namely, accessions 'GYU', 'HAL' and 'VIT' behave as a group: misclassifications of an accession from this group mainly classified into another member of this group. The same is valid for accession group '13/10', 'KM193', 'KM238', 'MEC' and 'USZ' as well as for group 'FER', 'ITA', 'PER' and 'SZU' while 'COR' constitutes a separate

group alone. Thus, as a next step, we performed a multivariate linear discriminant analysis with the same variables but with the above four accession groups.

Wilks' lambda values of the three discriminant functions (DF) were all significant (p<0.001) with canonical correlations 1 to 3 from r=0.86 to r=0.51. Discriminant function 1 (DF1) explained 57.0 % of the total variance, while DF2 and DF3 explained 36.0 % and 6.9 %, respectively. The first three DF explained more than 99 % of the total variance. In Table 1, we can see that the loadings of the DFs show a similar pattern as in the 13-accession model. The classification resulted in 88.4 % correct classifications. The leave-one-out cross-validated classification was again close to this successful classification rate with its 87.2 %. According to this model, 92 % of

'COR' (group "A") samples were classified correctly. This variety has small berries and in terms of the OIV (2009), it has ovoid shape, which is wider on the bottom than on the top (Fig. 3). The accessions 'GYU', 'HAL', 'VIT' belong to group "B" and 82.1 % of them were classified correctly. The members of this group have long, cylindrical and horn-shaped berries. Accessions 'FER', 'ITA', 'PER' and 'SZU' belonged to the group "C" that was classified correctly with the highest rate (95.3 %). The members of this group have globose and broad ellipsoid berries. This is followed by group "D" of accessions '10/13', 'KM193', 'KM238', 'MEC' and 'USZ' with long and wide berries performing a correct classification rate of 82.1 %. This latter group consists of those accessions that have large elongated ellipsoid berries.
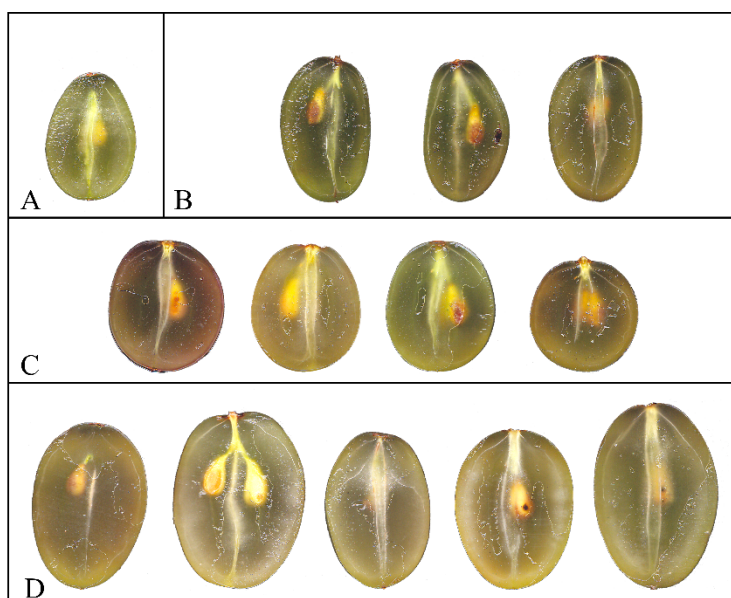

Fig. 3: Similarity groups based on the discriminant analysis according to the investigated berry morphological traits (group "A": 'COR'; group "B": 'GYU', 'HAL', 'VIT'; group "C": 'FER', 'ITA', 'PER' and 'SZU'; group "D": '10/13', 'KM193', 'KM238', 'MEC', 'USZ')

Conclusions

In this study, we found that there are obvious differences between the grapevine berry samples belonging to the same genotype and this variability is very much influenced by seed number. Discriminant analysis highlighted those traits, which provide high-level correct classification of the samples, however our results showed that discriminant potential of the berry variables is more linked to some similarity groups of varieties than to individual accessions. We found that digital image analysis is a powerful tool to evaluate berry size and shape diversity and Tomato Analyzer could successfully be applied in ampelography to examine large numbers of samples. This methodology provides timesaving and accurate description of the samples with valuable information about the uvometric diversity. We suggest to adopt attributes belonging to basic measurements and fruit shape indices of the TA in traditional ampelographic investigations to provide a more detailed characterization of the grapevine genotypes.

References

Barbagallo, M.G., Guidoni, S., Hunter, J.J. 2011: Berry size and qualitative characteristics of Vitis vinifera L. cv. Syrah. South African Journal for Enology and Viticulture 32(1): 129–136.

Bioletti, F.T. 1938: Outline of ampelography for the vinifera grapes in California. Hilgardia 116: 227-293.

Brewer, M.T., Lang, L., Fujimura, K., Dujmovic, N., Gray, S., van der Knaap, E. 2006: Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species. Plant Physiology 141: 15–25.

Ferrara, G., Gallotta, A., Pacucci, C., Matarrese, A. M. S., Mazzeo, A., Giancaspro, A., Gadaleta, A., Piazzolla, F., Colelli, G. 2017: The table grape 'Victoria' with a long shaped berry: a potential mutation with attractive characteristics for consumers. Journal of the Science of Food and Agriculture 97(15): 5398–5405.

Hurtado, M., Vilanova, S., Plazas, M., Gramazio, P., Herraiz, F. J., Andújar, I., Prohens, J. 2013: Phenomics of fruit shape in eggplant (Solanum melongena L.) using Tomato Analyzer software. Scientia Horticulturae 164: 625–632.

Mullins, M.G., Bouquet, A., Williams, L.E., 1992: Biology of the grapevine. Cambridge University Press, Cambridge.

OIV 2009: OIV descriptor list for grape varieties and Vitis species 2nd edition. Office International de la Vigne et du Vin, Paris, France.
Piva, C.R., Garcia, J.L.L., Morgan, W. 2006: The ideal table grapes for the Spanish market. Rev. Bras. Frutic. 28(2): 258-261. 10.1590/S0100-29452006000200023

Ravaz, L. 1902: Les Vignes Americaines: Porte-Greffes et Producteurs Directs (Caracteres Aptitudes). Coulet et Fils, Montpellier. 376.
Tello, J., Ibañez, J. 2018: What do we know about grapevine bunch compactness? A state-of-the-art review. Australian Journal of Grape and Wine Research 24(1): 6-23.