Genetic diversity and relationships in the grapevine germplasm collection from Central Asia

A. Marrano[1], L. Grzeskowiak[1], P. Moreno Sanz[1], S. Lorenzi[1], M. L. Prazzoli[1], A. Arzumanov[2], M. Amanova[2], O. Failla[3], D. Maghradze[4] and M. S. Grando[1]

[1] Fondazione Edmund Mach, Research and Innovation Centre, San Michele all'Adige (Trento) Italy
[2] Uzbek Research Institute of Plant Industry, Tashkent, Uzbekistan
[3] Dipartimento di Scienze Agrarie e Ambientali, Università degli Studi di Milano, Italy
[4] Institute of Horticulture, Viticulture and Oenology, Agricultural University of Georgia, Georgia

Summary

The mountainous region between the Caucasus and China is considered the center of diversity for many temperate fruit crops. Also the transitional types of grapes, including wild forms of the subsp. *Vitis sylvestris*, cultivated landraces and ancient local varieties, were once common in this region. Despite Central Asia is considered a focal region of the world regarding grapevine development, limited information about the extent and distribution of grapevine genetic variation is available.

Here we report the first assessment of genetic diversity, relationships and structure of 80 grapevine cultivars and 21 *V. sylvestris* accessions originated from the regions of Uzbekistan, Tajikistan and Kyrgyzstan. We expanded the coverage of this survey to include a set of 53 traditional Georgian varieties and homologous SSR genotypes of 107 cultivars representing four *V. vinifera* ancestral subpopulations. This allowed us to evaluate the contribution of the Central Asian grapevine germplasm to diversification of the cultivated grapevine gene pool.

K e y   w o r d s : SSR marker profiles; Georgia; Uzbekistan; Tajikistan; Kyrgyzstan.

Introduction

Based on archaeological evidence dating from 8000 BC and the large genetic diversity, South Caucasus and Anatolia have long been regarded as homelands of viticulture (Vavilov 1931, Vouillamoz *et al.* 2006). Historical records suggest that cultivation of *V. vinifera* was spread to North Africa by the end of the fifth millennium BC, and it was established in Europe during the first millennium BC. Grape culture is supposed to have reached Afghanistan and the oases of Central Asia by the fourth century BC, and China in the second century BC (Lutz 1922, Vavilov 1931, Negrul 1946, Levadoux 1956, Mc Govern 2003).

According to Negrul (1946) who traveled widely throughout Europe and Central Asia, the grapevines found in the wide area extending from eastern Georgia, Armenia, Azerbaijan to the former Soviet republics in Central Asia and the region of the Near East have clear distinguishing features and were placed in the Proles *orientalis*. Negrul recognised two sub-proles within this main group: *caspica*, composed of ancient vines used for vinification before the advent of Islam (from AD 500-1100), and the *antasiatica* including cultivars for table grape and raisins of more recent origin. Varietal ecotypes found from Georgia to the Balkans were instead designated *P. pontica* sub-proles *georgica* and sub-proles *balkanica*, respectively.

Further extensive field investigations into natural populations of *V. vinifera* led Negrul to conclude that cultivars from the region of the Caspian Sea (sub-proles *caspica*) were so different from the Proles *pontica* that they must have arisen from a different wild form. He called it *V. sylvestris* var. *aberrans* the vine form with hairless leaf surface as opposed to the most widespread *V. sylvestris* var. *typical* having hairy leaves.

Molecular analysis has provided, for almost two decades, new insights on genetic diversity of *V. vinifera* in relation to wild relatives, origin of cultivars and specific alleles linked to selected traits (Arroyo-Garcia *et al.* 2006, This *et al.* 2006, Emanuelli *et al.* 2010). However, despite Central Asia is considered a focal region of grapevine development, information about the amount and distribution of grapevine genetic variation have only recently started to emerge and it is based on accessions from Central Asian countries maintained in European and USA germplasm repositories. These materials were included in genetic studies aimed at interpreting the population structure of cultivated varieties as well as to further investigate the intriguing resistance to *Erysiphe necator* found in some eastern *V. vinifera* forms (Hoffmann *et al.* 2008, Bacilieri *et al.* 2013, Riaz *et al.* 2013).

Here we report the first assessment of genetic diversity, relationships and structure of grapevine cultivars conserved in the local collection of the Uzbek Research Institute of Plant Industry ('UzRIPI'; Tashkent region, Republic of Uzbekistan) including several *V. sylvestris* accessions. We expanded the coverage of this preliminary survey to include a set of traditional Georgian varieties and homologous SSR genotypes of cultivars representing four *V. vinifera* ancestral subpopulations (Emanuelli *et al.* 2013). This allowed us to evaluate the contribution of Central Asian grapevine germplasm to diversification of the cultivated gene pool.

Material and Methods

P l a n t   m a t e r i a l   a n d   S S R   a n a l y s i s : A grapevine (*V. vinifera* L.) collection of 80 cultivated and 21 supposed wild accessions from the region of Central Asia (Uzbekistan, Tajikistan and Kyrgyzstan) and 53 cultivars from Georgia was analyzed (Tab. 1). Leaf samples were placed in 96-well microtube plates and freeze-dried. DNA extraction was performed using DNAeasy 96 plan mini kit (Qiagen, Germany).

T a b l e   1

Origin of grapevine accessions analysed in the present study.

Area of origin	Number of accessions	Institutional source
Tajikistan	52	UzRIPi
Uzbekistan	23	UzRIPi
Kyrgyzstan	4	UzRIPi
Central Asia*	22	UzRIPi
Georgia	53	GEO015

*the exact geographic location is unknown

Twenty two SSR markers, including at least one locus per chromosome, were chosen to profile the whole collection of 154 accessions. This set includes the nine SSR markers proposed by the European Project GrapeGen06 for the characterization of regional cultivars (Maul *et al.* 2012) and the loci VVIQ52, VVIN16, VVIV37, VVIH54, VVIN73, VVIP31, VVIB01, VVIV67 (Merdinoglu *et al.* 2005), VVMD21, VVMD24 (Bowers and Meredith 1999), VMC4F3.1 (BV722689), VMC4F8 (BV102437) and VMC1B11 (BV681754).

Nine multiplex panels of fluorescently labeled markers were used as reported in Emanuelli *et al.* (2013). The PCR products were denatured and size fractioned using capillary electrophoresis on an ABI 3130 Genetic Analyzer (Applied Biosystems). GeneMapper v3.5 (Applied Biosystems) was used for the alleles size estimation.

G e n e t i c   d i v e r s i t y   a s s e s s m e n t : The final dataset of non-redundant genotypes was used to estimate the main diversity statistics, such as total number of different alleles per locus (N_A), number of effective alleles (N_E, the number of equally frequent alleles required to give the observed level of heterozygosity), observed (H_O) and expected (H_E) heterozygosity and fixation index (F, inbreeding coefficient) through GenAlex v6.501 (Peakall and Smouse 2006, 2012).

A n a l y s i s   o f   p o p u l a t i o n   s t r u c t u r e : The genetic structure was first assessed by principal coordinate analysis (PCoA), implemented in GenAlex v6.501. Genotypic data were then subjected to the Bayesian clustering analysis, implemented in STRUCTURE 3.2 (Pritchard *et al.* 2000) using the admixed and correlated allele frequency models. Ten independent runs for K values ranging from 1 to 10 were performed with a burn-in length of 10,000 followed by 100,000 iterations. The most likely subdivision (K) was established by plotting the log probability L(K) and ΔK of the data over ten runs, as implemented in Structure Harvester v0.6.94 (Evanno *et al.* 2005, Earl and von Holdt 2012).

The unique genetic profiles at 22 SSR loci were further subjected to cluster analysis using the Darwin software package v6.0 (Perrier and Jacquemoud-Collet 2006). A weighted neighbour-joining tree was constructed based on the simple matching dissimilarity matrix with 100 bootstrap replicates. Further cluster analysis was performed including the genetic profiles of 107 cultivars which belong to the FEM germplasm collection (ITA362) and represent four subpopulations of *V. vinifera*, in accordance with the eco-geographic origin of the cultivars (Emanuelli *et al.* 2013). In addition, the SSR profile of 11 grape rootstock (*Vitis* spp.) varieties were used for outgroup comparisons.

Results and Discussion

Pairwise comparisons based on SSR profiles at 9 loci led to the identification of 11 and 10 synonymous groups in the Central Asia and Georgian subsets respectively, comprising 35 accessions overall. The final dataset of distinct SSR profiles was composed of 13 wild and 66 cultivated genotypes from Central Asia and 40 Georgian cultivars.

A comparison of the SSR genotypes with those reported in the European *Vitis* Database (www.eu-vitis.de) revealed that the three different Georgian accessions 'Saperavi Budeshuriseburi' (a 'Saperavi' mutant), 'Kisi' and 'Ikaltos Tsiteli' matched at all the nine tested loci with the following entries: 'Saperavi' (DEU098-1993-253, ITA035-118), 'Goruli mtsvane' (ITA035-69) and 'Rkatsiteli' (DEU098-1980-083, AUT024-319), respectively. It is worth noting that the cultivated Uzbek varieties, 'Bishti' and 'Ruzbari', had identical SSR profiles as 'Lambrusque Carranques' 3, a *V. sylvestris* accession (FRA139-8500Mtp164), and 'Rund Weiss' (FRA139-0Mtp1002) a cultivar thought to have originated in Azerbaijan, respectively. These findings deserve additional investigation on the accessions' morphological descriptors which were not integrated in the dataset.

The panel of 119 unique genotypes was characterized at 13 more SSR loci in order to estimate the main indexes of genetic diversity separately in the three subsets: cultivated and wild accessions from Central Asia and cultivars from Georgia, and to assess the relationship among the accessions. The average number of different alleles per locus in the whole sample was 11.2 and 64 % of alleles were shared among these three groups. Genetic diversity parameters, summarized in Tab. 2, revealed higher levels of expected and observed heterozygosity in the cultivated compartment, compared to the small group of wild individuals. The amount of variation was similar to that reported for larger samples of *V. vinifera* germplasm (Bacilieri *et al.* 2013; Emanuelli *et al.* 2013).

Genetic relationships were investigated using the principle coordinate analysis (PCoA) and Structure approaches. The PCoA, based on a genetic distance matrix with data standardization, explored how the Georgian group may be differentiated from the Central Asia populations. Plotting of the first two principle coordinates showed a clear separation between cultivated accessions from

T a b l e 2

Summary statistics for 119 genotypes from three populations assessed using 22 SSR markers (N = sample size; N_a = N° of different alleles per locus; N_e = N° of effective alleles; N_p = N° of private alleles; Ho = observed heterozygosity; He = expected heterozygosity; F = Fixation Index)

Population	N	N_a	N_e	Np	H_o	H_e	F
Central Asia	66	9.55	4.86	54	0.76	0.77	0.01
Wild Central Asia	13	4.91	3.39	0	0.64	0.66	0.02
Georgia	40	8.05	4.24	34	0.75	0.73	-0.03

Central Asia and Georgian varieties along the first axis, whereas the wild and cultivated genotypes of Central Asia were distinguished, though to a lesser degree, along the second coordinate (Fig. 1). The subdivision of populations originating from Central Asia, with respect to those from the Caucasus region, was not very evident in the previous structure analysis of the large grape collection of Vassal (INRA, France) performed by BACILIERI *et al.* (2013). However, similarly to our findings genotypes from the eastern regions subdivided into two sub-groups according to the main local use of grapevines: wine, for the Caucasian cultivars and table, for the Central Asian cultivars.

The unique profiles at 22 SSR loci were used for Bayesian clustering analysis implemented in STRUCTURE. The most likely number of clusters (K), obtained using the ΔK method proposed by Evanno et al. (2005), was equal to K = 2. Using a threshold of cluster membership coefficient equal to 0.80, 37 out of 40 Georgian genotypes were assigned to the cluster K_1 and all 79 individuals from Central Asia were included into the group K_2. Thus, a very low level of admixture characterizes the Caucasian sample analyzed in this study. This is in agreement with the findings of IMAZIO *et al.* (2013) on a different portion of the Georgian germplasm. The absence of admixture observed in the Central Asian populations is intriguing, and it may raise questions regarding the spatial and temporal patterns of grapevine domestication.

The topology of the weighted neighbor joining dendrogram including the genotypes from Georgia and Central Asia reflected as well two major groups, in accordance with the geographic origin of the samples (data not shown). Moreover, within the Central Asian grapevines, most of the supposed wild accessions clustered together, and no evidence of genetic differentiation was observed among the subsets from the UzRIPI germplasm collection considered to have originated from the regions of Uzbekistan, Tajikistan and Kyrgyzstan.

To gain a broader understanding of the genetic relationship of these Georgian and Central Asian grapevines we performed a further cluster analysis including the homologous genetic profiles of 107 additional cultivars at the same set of 22 SSR loci. These accessions belong to four ancestral subpopulations of *V. vinifera* ssp. *sativa* (VV) which were detected within a large sample of grapevine accessions following a hierarchical clustering approach (EMANUELLI *et al.* 2013). In particular, we included the cluster of Italian and Greek wine grapes (VV1), representing the proles *pontica*, the French and German wine varieties (VV4), representing the proles *occidentalis*, and the Muscat table and wine grapes (VV3) reflecting the proles *orientalis* subpr. *caspica*. The cluster VV2 was more heterogeneous, and it was composed of both table grape varieties related to 'Sultanina' (proles *orientalis* subpr. *antasiatica*) and Spanish wine grapes.
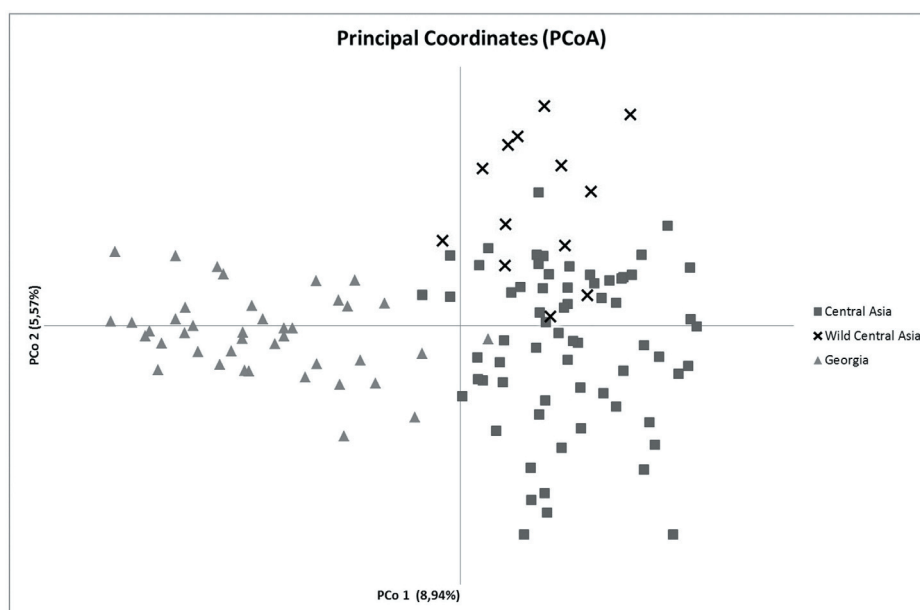

Fig. 1: Scatter plot of the first two principal coordinate analysis axes for the SSR data of 119 genotypes.
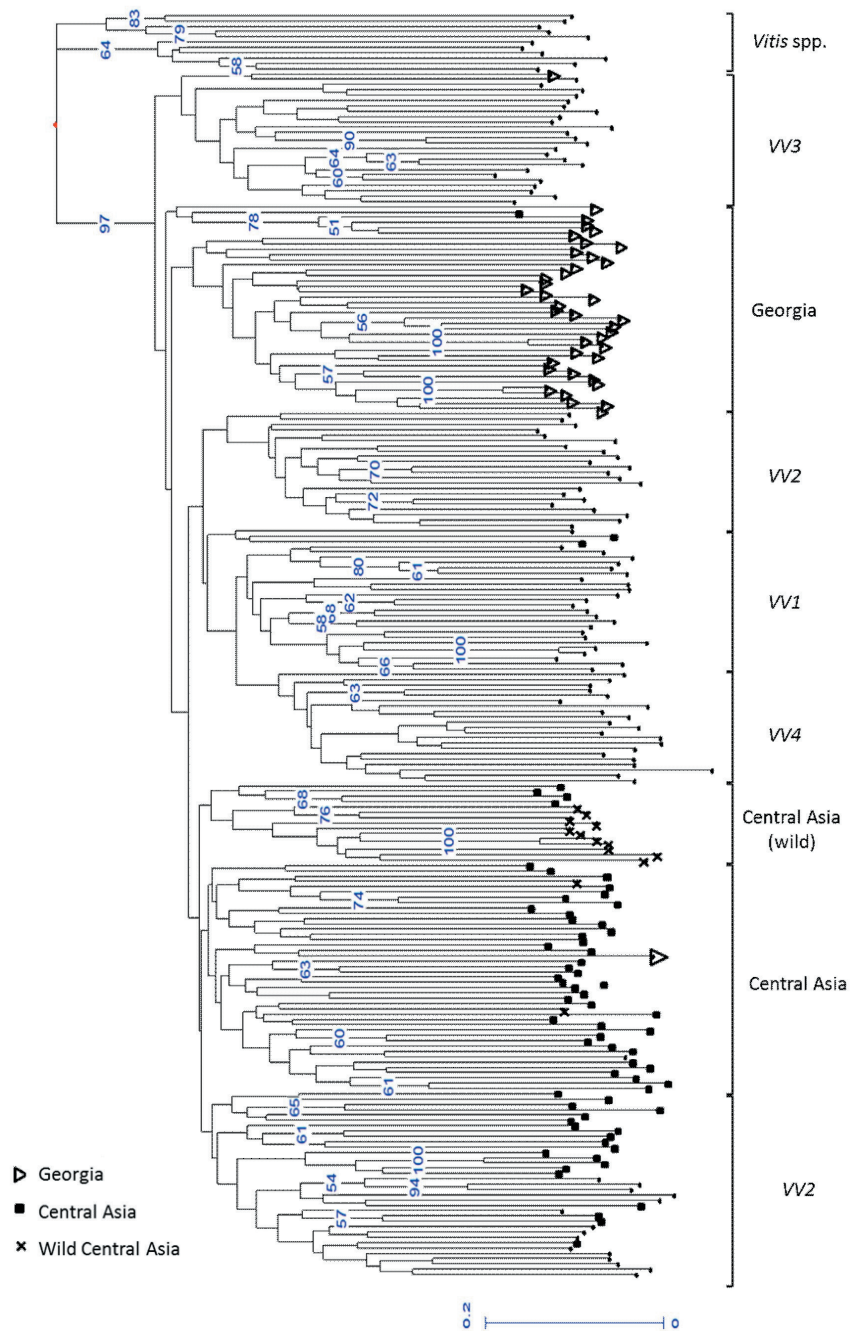
Fig. 2: Neighbor joining weighted tree based on a dissimilarities matrix calculated from SSR alleles at 22 loci for 226 *V. vinifera* genotypes and 11 rootstocks (*Vitis* spp.) as an outgroup. Only bootstraps superior to 50 are presented.

In this context, almost all the Georgian genotypes formed an additional well distinct clade, which likely corresponds to the proles *pontica* subpr. *georgica* Negr (Fig. 2). Most of Central Asian genotypes also grouped together in a large cluster composed of three subclusters, including the table grapes portion of the VV2 population. In fact, since 20 Central Asian genotypes fell within the VV2 population, the previous subgroup of the Spanish wine cultivars was now separated. Most of the cultivated and *V. sylvestris* accessions from the UzRIPI collection composed two distinct subgroups, although it is worth noting that some wild accessions were included within the subgroup of cultivars and *vice versa*.

In conclusion, the grapevine gene pool of Georgia and Central Asia surveyed in this study has a significant amount

of genetic variation and exhibits high levels of population differentiation which may reflect a limited historical gene flow between the two regions. In addition to the first molecular description of the genetic diversity of the Central Asian grape germplasm collection currently maintained in local repository, this study contributed to the integration of genotype information on these extremely valuable grapevine genetic resources into The European *Vitis* Database.

References

Arroyo-García, R.; Ruiz-García, L.; Bolling, L.; Ocete, R.; López, M. A.; Arnold, C.; Ergul, A.; Söylemezoğlu, G.; Uzun, H. I.; Cabello, F.; Ibáñez, J.; Aradhya, M. K.; Atanassov, A.; Atanassov, I.; Balint, S.; Cenis, J. L.; Costantini, L.; Goris-Lavets, S.; Grando, M.

S.; KLEIN, B. Y.; MCGOVERN, P. E.; MERDINOGLU, D.; PEJIC, I.; PELSY, F.; PRIMIKIRIOS, N.; RISOVANNAYA, V.; ROUBELAKIS-ANGELAKIS, K. A.; SNOUSSI, H.; SOTIRI, P.; TAMHANKAR, S.; THIS, P.; TROSHIN, L.; MALPICA, J. M.; LEFORT, F.; MARTINEZ-ZAPATER, J. M.; 2006: Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. Mol. Ecol. **15,** 3707-3714.

BACILIERI, R.; LACOMBE, T.; LE CUNFF, L.; DI VECCHI-STARAZ, M.; LAUCOU, V.; GENNA, B.; PÉROS, J. P.; THIS, P.; BOURSIQUOT, J. M.; 2013: Genetic structure in cultivated grapevines is linked to geography and human selection. BMC Plant Biol. **13,** 25.

BELKHIR, K.; BORSA, P.; CHIKHI, L.; RAUFASTE, N.; BONHOMME, F.; Genetix 4.05. CNRS UMR 5171. Montpellier: Université de Montpellier II; http://www.genetix.univmontp2.fr/genetix/intro.htm.

BOWERS, J. E.; DANGL, G. S.; MEREDITH, C. P.; 1999: Development and characterization of additional microsatellite DNA markers for grape. Am. J. Enol. Vitic. **50,** 243-246.

EARL, D. A.; VON HOLDT, B. M.; 2012: STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. **4,** 359-361.

EMANUELLI, F.; BATTILANA, J.; COSTANTINI, L.; LE CUNFF, L.; BOURSIQUOT, J. M.; THIS, P.; GRANDO, M. S.; 2010: A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). BMC Plant Biol. **10,** 241.

EMANUELLI, F.; LORENZI, S.; GRZESKOWIAK, L.; CATALANO, V.; STEFANINI, M.; TROGGIO, M.; MYLES, S.; MARTINEZ-ZAPATER, J. M.; ZYPRIAN, E.; MOREIRA, F. M.; GRANDO, M. S.; 2013: Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biol. **13,** 39.

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.

HOFFMANN, S.; DI GASPERO, G.; KOVÁCS, L.; HOWARD, S.; KISS, E.; GALBÁCS, Z.; TESTOLIN, R.; KOZMA, P.; 2008: Resistance to *Erysiphe necator* in the grapevine 'Kishmish vatkana' is controlled by a single locus through restriction of hyphal growth. Theor. Appl. Genet. **116,** 427-438.

IMAZIO, S.; MAGHRADZE, D.; DE LORENZIS, G.; BACILIERI, R.; LAUCOU, V.; THIS, P.; SCIENZA, A.; FAILLA, O.; 2013: From the cradle of grapevine domestication: molecular overview and description of Georgian grapevine (*Vitis vinifera* L.) germplasm. Tree Genet. Genomes **9,** 641-658.

KIMURA, M.; CROW, J. F.; 1964: The number of alleles that can be maintained in a finite population. Genetics **49,** 725-738.

LAUCOU, V.; LACOMBE, T.; DECHESNE, F.; SIRET, R.; BRUNO, J. P.; DESSUP, M.; DESSUP, T.; ORTIGOSA, P.; PARRA, P.; ROUX, C.; SANTONI, S.; VARÈS, D.; PÉROS, J. P.; BOURSIQUOT, J. M.; THIS, P.; 2011: High throughput analysis of grape genetic diversity as a tool for germplasm collection management. Theor. Appl. Genet. **122,** 1233-1245.

LEVADOUX, L. D.; 1956: Wild and cultivated populations of *Vitis vinifera* L. Ann. Amelior. Plant **6,** 59-118.

LUTZ, H. F.; 1922: Viticulture and Brewing in the Ancient Orient. J. C. Hinrichs'sche Buchhandlung, Leipzig.

MAUL, E.; TÖPFER, R.; EIBACH, R.; 2012: Vitis International Variety Catalogue. Julius Kühn Institut. www.vivc.de.

MERDINOGLU, D.; BUTTERLIN, G.; BEVILACQUA, L.; CHIQUET, V.; ADAMBLONDON, A. F.; DECROOCQ, S.; 2005: Development and characterization of a large set of microsatellite markers in grapevine (*Vitis vinifera* L.) suitable for multiplex PCR. Mol. Breed. **15,** 349-366.

MCGOVERN, P. E.; 2003: Ancient Wine: The Search for the Origins of Viniculture. Princeton: Princeton University Press;.

NEGRUL, A. M.; 1946: Origin and classification of cultivated grape. In: A. K. Y. BARANOV, M. A. LAZAREVSKI, T. V. PALIBIN, N. N. PROSMOSERDOV (Eds): The Ampelography of the USSR, vol 1, 159-216. Pischepromizdat, Moscow.

NEI, M.; 1973: Analysis of gene diversity in subdivided populations. PNAS **70,** 3321–3323.

PEAKALL, R.; SMOUSE, P. E.; 2006: GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecol. Notes **6,** 288-295.

PEAKALL, R.; SMOUSE, P. E.; 2012: GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics **19,** 2537-2539.

PERRIER, X.; JACQUEMOUD-COLLET, J. P.; 2006: DARwin software. http://darwin.cirad.fr

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P.; 2000: Inference of population structure using multilocus genotype data. Genetics **155,** 945-959.

RIAZ, S.; BOURSIQUOT, J. M.; DANGL, G. S.; LACOMBE, T.; LAUCOU, V.; TENSCHER, A. C.; WALKER, M. A.; 2013: Identification of mildew resistance in wild and cultivated Central Asian grape germplasm. BMC Plant Biol. **13,**149.

SEFC, K. M.; REGNER, F.; GLÖSSL, J.; STEINKELLNER, H.; 1999: Identification of microsatellite sequences in *Vitis riparia* and their applicability for genotyping of different *Vitis* species. Genome **42,** 367-373.

THIS, P.; LACOMBE, T.; THOMAS, M. R.; 2006: Historical origins and genetic diversity of wine grapes. Trends Genet **22,** 511-519.

VAVILOV, N. I.; 1931: Dikie rodichi plodovykh dereviev Aziatskoi chasti SSSR i Kavkaza i problema proiskhozhdenia plodovykh dereviev (Wild progenitors of the fruit trees of Turkestan and the Caucasus and the problem of the origin of fruit trees). Bull. Appl. Bot. Genet. Plant Breed. **6,** 85-134 (in Russian).

VOUILLAMOZ, J.; MCGOVERN, P. E.; ERGUL A SÖYLEMEZOĞLU, G.; TEVZADZE, G.; MEREDITH, C. P.; GRANDO M. S.; 2006: Genetic characterization and relationships of traditional grape cultivars from Transcaucasia and Anatolia. Plant Genet. Resour. **4,** 144-158.