«SOS for endangered traditional vine varieties-VineSOS»

DNA analysis of Bulgarian grapevine varieties - Final Report

Work Package 3: Diagnosis and project Strategy planning (D3.5.2: Scientific work for the project implementation)

Part A: D3.5.2. «External expertise for DNA analysis»

Dimitrios Loukovitis

Agriculturist, MSc, PhD

Introduction

Molecular Biology

Molecular Biology studies the molecular basis of biological interactions between biomolecules, various systems, anabolic and catabolic reactions in cells, including interactions between DNA, RNA, proteins and biosynthesis of them, as well as the regulation mechanisms of these interactions. Molecular Biology is directly linked to Genetics and has offered valuable tools for Molecular Genetics science, such as DNA cloning using plasmid vectors, Polymerase Chain Reaction (PCR), electrophoresis methods and DNA/RNA sequencing.

Molecular Genetics and techniques

Molecular Genetics studies the structure and function of genes in a molecular level (nucleotide pairs), using the aforementioned Molecular Biology methodologies, in combination with classical statistical methods of Genetics science.

The study of chromosomes and gene expression in organisms can offer valuable information for heritability, genetic drift and mutations in animal/plant genomes as well as their frequency in population level.

DNA extraction

In order for the above methods to be successful, a DNA molecule is necessary to serve as a template for the amplification procedure. This molecule can be isolated from cell nucleus, mitochondria, chloroplasts or be produced from homologous mRNA using reverse transcriptase enzyme. This enzyme is used for cDNA (complementary DNA) production coming from single stranded mRNA. In nature, reverse transcriptase is found mostly in retroviruses.

For DNA or mRNA extraction, cell lysis (cell membrane, cell wall, nuclear membrane, histones etc.) is necessary, which is achieved by using the appropriate buffers and enzymes (proteinase K, lysozyme, lysis buffer, wash buffer etc.) as well as the recommended experimental protocol.

DNA amplification

For the production of a large number of DNA copies in 70's, cloning method was applied using plasmid vectors. However, this proved to be costly and time consuming, giving its position to Polymerase Chain Reaction method in the middle 80's, which brought revolutionary changes in the field of genome analysis and, specifically, in molecular genetics. PCR mimics *in vitro* the mechanism of DNA replication, which happens naturally in cells, using thermostable reagents.

PCR starts with the denaturation of a double stranded DNA molecule that contains the target sequence. Amplification of this sequence is achieved through many cycles of copy, using a pair of oligonucleotide primers (usually 20-30 bases) that bind to the single stranded DNA molecules (at the 3' end of each strand), in the sequence of interest, due to complementarity.

Steps of PCR are as follows:

- Heating of the reaction mixture to 94-95 °C for denaturing double stranded DNA molecules
- The reaction temperature is lowered to 37-65 °C, allowing the annealing of primers to each of the single-stranded DNA templates (the annealing depends on the sequence and length of primers)
- Heating of the reaction mixture to 68-72 °C and extension/elongation of primers, where a thermostable DNA polymerase (e.g. *Taq* polymerase from the thermophilic bacterium *Thermus aquaticus*) synthesizes a new DNA strand complementary to the DNA template strand by adding free dNTPs from the reaction mixture that are complementary to the template in the 5'-to-3' direction
- The processes of denaturation, annealing and elongation constitute a single cycle. Multiple cycles (25-50) are required to amplify the target sequence to millions of copies

The duration of each PCR cycle is 1-3 minutes and under optimal conditions (i.e., if there are no limitations due to limiting substrates or reagents), at each extension/elongation step, the number of DNA target sequence is doubled. With each successive cycle, the original template strands plus all newly generated strands become template strands for the next round of elongation, leading to exponential (geometric) amplification of the specific DNA target region.

Scope of the study

The aim of this study was to analyze the genetic diversity of various Bulgarian grapevine (*Vitis vinifera* L.) varieties, in order to assess whether they can be genetically differentiated from each other, based on their DNA profile in specific molecular markers.

Materials and Methods

Sampling

Young leaves were collected from Bulgarian colleagues, including nine different grapevine varieties from several locations in Bulgaria. After collection, the leaves were immediately stored in a freezer (-20 ^{o}C) and sent to the Laboratory of Agrobiotechnology and Inspection of Agricultural Products, Department of Agriculture, International Hellenic University, Sindos, Thessaloniki, Greece. Detailed information for leaf samples are given in Table 1.

Table 1. Information for collected grapevine leaf samples

PRODUCER	LOCATION	VARIETY	N
BRATANOV WINERY	SHISHMANOVO	TAMYANKA	20
BRATANOV WINERY	SHISHMANOVO	RUBIN	20
BRATANOV WINERY	SHISHMANOVO	MAVRUD	10
SVILEN GEORGIEV	KOLAROVO	PAMID	10
DIMITAR DJEMPERLIEV	DIMITROVCHE	MAVRUD	20
DIMITAR DJEMPERLIEV	DIMITROVCHE	TAMYANKA	10
DIMITAR DJEMPERLIEV	DIMITROVCHE	DIMYAT	20
VALERI LAZAROV GALAJOV	VZANYA	RUEN	20
VALERI LAZAROV GALAJOV	VZANYA	SHIZOKA MELNISHKALOZA	30
VALERI LAZAROV GALAJOV	VZANYA	KERATSUDA	10
KOSTADIN CHOZBADJISHI	MITINO	SHIZOKA MELNISHKALOZA	20
KOSTADIN CHOZBADJISHI	STAZCHEVO	MISHEJ SANDANSKI	30
VASIL KIZOV TZAYANOV	SLIVNITSA	KERATSUDA	20
IVAN PLAMENOV BOYCHEV	SLIVNITSA	KERATSUDA	20
DIMITAZ ANDEEZ AKAZEEV	DOLNA GZADESHNIDZA	KERATSUDA	10
VANGEL ZONEV	DOLNA GZADESHNIDZA	KERATSUDA	10
VLADIMIZ ANDZEEV	DOLNA GZADESHNIDZA	KERATSUDA	10
RUMEN MALCHEV	KZESHA	KERATSUDA	10

EVGENY STOYANOV	KZESHA	KERATSUDA	10
GEORGI SHTEZEV	KZESHA	KERATSUDA	10
VALENTIN ANGELOV	KZESHA	KERATSUDA	1
GEORGI IVANOV	KZESHA	KERATSUDA	1
ILIA ANGELOV	KZESHA	KERATSUDA	1
GEORGI STEZEV	KZESHA	KERATSUDA	1
IORDAN PAUVLOV	KZESHA	KERATSUDA	1
BOZIS MONEV	KZESHA	KERATSUDA	1
NIKOLAI STEFANOV	KZESHA	KERATSUDA	1
GEORGI MARKOV	KZESHA	KERATSUDA	1
BLAGOI STOYMENOV	KZESHA	KERATSUDA	1
PETAN PETROV	KZESHA	KERATSUDA	1
OLEG ILIEV	KZESHA	KERATSUDA	1
VASIL MITOV	KZESHA	KERATSUDA	1
PETAN GALTSHEV	KZESHA	KERATSUDA	1
ANDREI TTAZIZANOV	KZESHA	KERATSUDA	1
IVAN TSINKAZOV	KZESHA	KERATSUDA	1
VALENTIN LAZAROV	GZADESHNITZA	KERATSUDA	1
DRAGAM ISGOBAMOV	GZADESHNITZA	KERATSUDA	1
MLADEN GEORGIEV	GZADESHNITZA	KERATSUDA	1
ANGEL VAKADINOV	GZADESHNITZA	KERATSUDA	1
STOYAN AKAZEV	GZADESHNITZA	KERATSUDA	1
ASEN BOEV	GZADESHNITZA	KERATSUDA	1
VANGEL YAMEV	GZADESHNITZA	KERATSUDA	1
PETAN TASEV	SLIVNIKA	KERATSUDA	1
YORDAN GEORGIEV	SLIVNIKA	KERATSUDA	1
STZAHIL GYANCHEV	SLIVNIKA	KERATSUDA	1
VASIL TZAYANOV	SLIVNIKA	KERATSUDA	1
ATOMAS MILER	SLIVNIKA	KERATSUDA	1
SLAVCHO PANAYOTOV	SLIVNIKA	KERATSUDA	1
GEORGI TZAYAMOV	SLIVNIKA	KERATSUDA	1

GEORGI HZISTOV	SLIVNIKA	KERATSUDA	1
		Total number of samples	**350**

DNA extraction

Genomic DNA was isolated from 20-30 mg of leaf tissue (350 samples in total) using the 'NucleoSpin Plant II' DNA extraction kit (Macherey-Nagel, Germany), according to the following protocol:

1. Leaf samples are stored in -20 ^{o}C immediately after sampling, for DNA integrity purposes
2. Leaves are gradually (24 samples each time) taken out form freezer and left to thaw completely until they reach 5-10 ^{o}C
3. A piece of tissue from each sample, weighing 20-30 mg, is placed in an 'Eppendorf' 1.5 ml tube, which has a unique code according to sample code
4. 400 ul of lysis buffer are added in all tubes with a pipette
5. All tissues are homogenized using sterile pestles
6. All tubes are shaked in vortex mixer for 12-15 seconds and then they are transferred in a water bath for incubation in 65 °C for 2 hours
7. After incubation tubes are centrifuged for 5 minutes in 16.500 rpm
8. Using a pipette the supernatant is carefully (without disturbing the debris) taken from each tube and transferred to a new sterile tube
9. 450 ul of PC buffer are added in all new tubes with a pipette
10. Tubes are shaked in vortex mixer for 12-15 seconds
11. The whole content from each tube is transferred with a pipette to a coded filter column (silica membrane) that is attached to a collection tube
12. Filter columns are centrifuged for 2 minutes in 12.000 rpm for the DNA to be collected in each filter
13. Collection tubes are discarded and the columns are placed in new ones
14. 400 ul of wash buffer PW1 are added in each column with a pipette and then all columns are centrifuged for 1 minute in 11.000 rpm
15. Collection tubes are discarded and the columns are placed in new ones
16. 680 ul of wash buffer PW2 are added in each column with a pipette and then all columns are centrifuged for 1 minute in 11.000 rpm

17. Collection tubes are discarded and the columns are placed in new ones
18. All filter columns are centrifuged for 2 minutes in 11.000 rpm for the filters to dry completely
19. Collection tubes are discarded and the columns are placed in new sterile 'Eppendorf' tubes (1.5 ml)
20. Using a pipette 50 ul of elution buffer PE (preheated to 65 °C) are added carefully (in the center of each filter) in each column, followed by incubation in 65 °C for 5 minutes
21. All columns are centrifuged for 1 minute in 11.000 rpm for the DNA to be eluted and collected in each 'Eppendorf' tube
22. Using a pipette the eluted DNA is taken from each 'Eppendorf' tube and placed again in each column (carefully in the center of each filter), followed by incubation in 65 °C for 5 minutes
23. All columns are centrifuged again for 1 minute in 11.000 rpm for the DNA to be re-eluted and collected in each 'Eppendorf' tube
24. After centrifugation filter columns are discarded and the 'Eppendorf' tubes, containing the DNA solution, are sealed and stored in -20 °C

After DNA extraction all samples were checked for quantity and quality of DNA, using electrophoresis method in agarose gel. Preparation of agarose gel and electrophoresis of DNA samples were carried out using the following protocol:

1. 100 ml of TBE 1X solution are added in a conical flask
2. 0.7-0.8 gr of agarose (molecular biology grade) are also added in the flask
3. The flask is incubated in 100 °C (using a microwave) for the agarose to be diluted and the solution to become clear
4. After incubation the temperature in the agarose solution is gradually decreased using tap water (in the exterior of the flask) and gentle manual shaking
5. When the temperature in the solution reaches 45-60 °C, 1.5 ul of fluorescent (under UV radiation) nucleic acid dye 'Midori Green' are added with a pipette
6. For complete homogenization and staining the flask is shaked manually and gently at the same time, in order to avoid the formation of bubbles
7. After homogenization the flask solution is poured in a gel casting tray with comp, which is necessary for the formation of the wells

8. The solution is maintained in the casting tray for 35-45 minutes in order to be solidified. The procedure takes place in room temperature, in a dark room
9. After verification of complete solidification, the comp is removed with caution from the gel and the casting tray is placed in the electrophoresis tank
10. TBE 1X solution is added to the electrophoresis tank so as to cover completely the wells
11. 4 ul from each DNA sample (already thawed) are mixed with 1 ul of gel loading dye (blue/orange), using a pipette
12. 5 ul of each mixed solution (DNA and loading dye) are loaded with a pipette in separate wells of the agarose gel. 5 ul of DNA ladder (1 kb) are also added in a separate well of the gel
13. The electrophoresis tank is sealed and connected with cables (through electrodes) in a power supply
14. Electric current is then applied to the tank for 30 minutes at 120 Volts
15. As the DNA samples are negatively charged they migrate to the positive electrode
16. As they migrate through the agarose gel, the DNA samples form complexes with 'Midori Green' dye. This dye is a DNA intercalator, inserting itself into the spaces between the base pairs of the double helix.
17. After electrophoresis the power is turned off and the casting tray is taken out of the electrophoresis tank and placed in an ultraviolet (UV) transilluminator
18. DNA fragments are visualized as bands due to the fluorescent nature of 'Midori Green' dye under UV radiation
19. The band(s) from each DNA sample are compared with the ones of DNA ladder

Each band of the DNA ladder is of known molecular weight (nanograms) and length (in base pairs). Comparing the bands' position and intensity of our DNA samples with those of DNA ladder, we can draw safe conclusions regarding the quantity and quality of the isolated DNA (Figure 1). After we verified that DNA extraction was successful for all leaf samples, we moved on to the next step of the analysis, which is to amplify DNA samples using the PCR method.

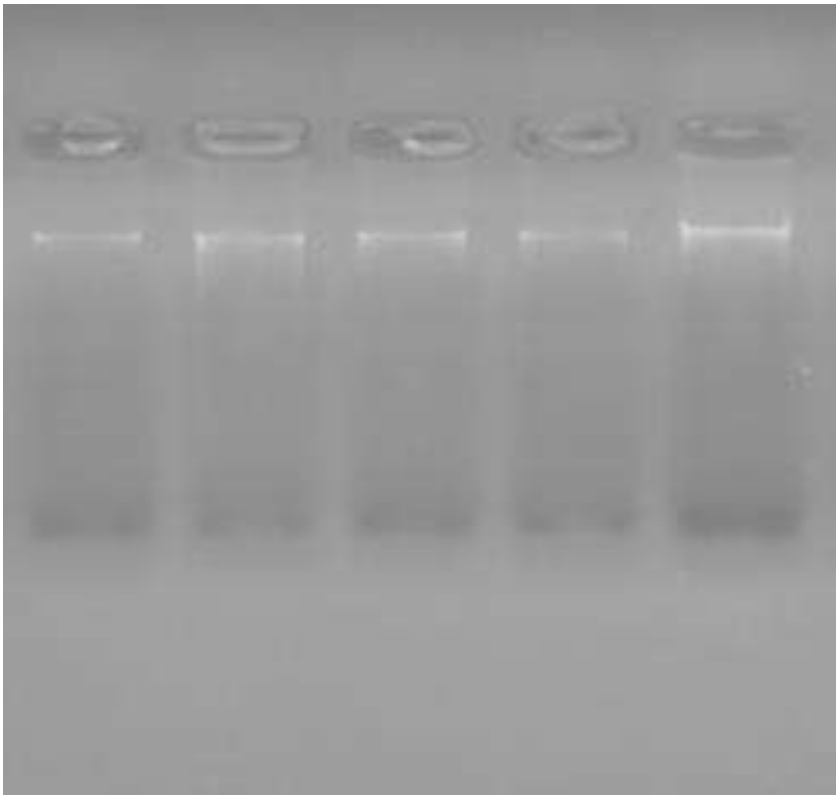

Figure 1. Isolated DNA from leaf samples

DNA amplification

In the present study, the method of Multiplex-PCR was applied, where in one reaction many different DNA fragments are amplified simultaneously, using a mix of primers that consists of 10 different primer pairs (Table 2). These pairs of primers target DNA fragments that harbor microsatellite sequences (microsatellite markers). The selected microsatellites are recommended by various studies in scientific literature and are proven to be very successful in unraveling the genetic variation and differentiation of grapevine varieties. Multiplex-PCR reactions were set up using the PCR mix 'KAPA2G Fast Multiplex PCR Kit' (KAPABIOSYSTEMS) and according to the following protocol:

1. PCR mix is taken out from freezer and left to thaw completely, followed by shaking in vortex mixer for 10-15 seconds in maximum speed
2. The volume of each reagent is calculated according to the final (total) volume of the PCR reaction. Herein, the final volume for each DNA sample, that is going to be amplified, is 5 ul and consists of:

- 1 ul of extracted DNA from each leaf sample
- 2 ul of PCR mix 'KAPA2G Fast Multiplex PCR Kit'
- 2 ul of primer mix, containing 0.1 ul from each primer on average. Therefore, there are 4 ul of master mix (2 ul PCR mix+ 2 ul primer mix) and 1 ul of DNA, in total

3. The volume of master mix is calculated according to the number of DNA samples that are chosen for amplification

4. The master mix is set up in an 'Eppendorf' tube (1.5 ml) and placed temporarily in a fridge (4 °C)

5. 1 ul of DNA from each leaf sample is added to a PCR plate with numbered wells

6. The PCR plate is centrifuged for 5 seconds in 3.700 rpm

7. The 'Eppendorf' tube containing the master mix is taken out from the fridge and homogenized completely with manual pipetting (10-15 times)

8. 4 ul of master mix is added to each well with a pipette

9. The PCR plate is centrifuged for 5 seconds in 3.700 rpm

10. The plate is sealed with caps and placed in the PCR machine. The Multiplex-PCR procedure for the amplification of microsatellite loci is ready to start

The PCR machine is equipped with heated lid adjusted to 105 °C. This function prevents water condensation on the lid and evaporation of the PCR reagents in the wells of the plate. The Multiplex-PCR conditions for the microsatellite marker amplification were set up according to the following protocol:

1. The lid of the PCR machine is heated to 105 °C

2. PCR plate is incubated for 3 minutes in 95 °C

3. PCR plate is incubated for 30 seconds in 95 °C in order to denaturate the double stranded DNA molecules

4. PCR plate is incubated for 1 minute in 57 °C so as the primers to anneal (hybridization) in the microsatellite loci of interest, according to complementarity

5. PCR plate is incubated for 1 minute in 72 °C for elongation of primers, where the DNA polymerase synthesizes a new DNA strand complementary to the target sequence, by adding free dNTPs from the reaction mixture that are complementary to the template in the 5'-to-3' direction

6. Steps from 3-5 constitute a PCR cycle and they are repeated 34 times (35 cycles in total). With each successive cycle, the original template strands plus all newly

generated strands become template strands for the next round of elongation, leading to exponential (geometric) amplification of the specific DNA target regions (microsatellite sequences)

7. PCR plate is incubated for 10 minutes in 72 °C (final elongation) to ensure that any remaining single-stranded DNA is fully elongated, as well as, for the DNA polymerase to add an extra adenine in the 3' end of each amplified fragment

When the Multiplex-PCR reaction is completed, the plate is taken out from the PCR machine and placed in a freezer (-20 °C) until fragment analysis, applying capillary electrophoresis of the PCR products in an automated sequencer.

Table 2. Microsatellite markers used in the present study

MICROSATELLITE LOCI	PRIMERS	FLUORESCENT DYE
VVS2	F:CAGCCCGTAAATGTATCCATC R:AAATTCAAAATTCTAATCACTGG	FAM
VVMD5	F:CTAGAGCTACGCCAATCCAA R:TATACCAAAAATCATATTCCTAAA	ROX
VVMD7	F:AGAGTTGCGGAGAACAGGAT R:CGAACCTTCACACGCTTGAT	FAM
VVMD25	F:TTCCGTTAAAGCAAAAGAAAAAGG R:TTGGATTTGAAATTTATTGAGGGG	HEX
VVMD27	F:GTACCAGATCTGAATACATCCGTAAGT R:ACGGGTATAGAGCAAACGGTGT	HEX
ssrVrZAG47	F:GTTCTTGGTCTGAATACATCCGTAAGT R:ACGGTGTGCTCTCATTGTCATTG	TAMRA
ssrVrZAG62	F:GGTGAAATGGGCACCGAACACACGC R:CCATGTCTCTCCTCAGCTTCTCAGC	FAM
ssrVrZAG67	F:ACCTGGCCCGACTCCTCTTGTATGC R:TCCTGCCGGCGATAACCAAGCTATG	HEX
ssrVrZAG79	F:AGATTGTGGAGGAGGGAACAAACCG R:TGCCCCCATTTTCAAACTCCCTTCC	TAMRA
scu05vv	F:CAAGCAGTTATTGAAGCTGCAAGG R:ATCATCCATCACACAGGAAACAGTG	ROX

Fragment analysis and statistical analysis

Fluorescently labeled PCR products were separated on an ABI 3500 Genetic Analyzer (Applied Biosystems) through capillary electrophoresis. Alleles were sized using the GeneScan LIZ500 size standard (Applied Biosystems) and individuals were genotyped using the STRand 2.4.59 software (Figure 2).

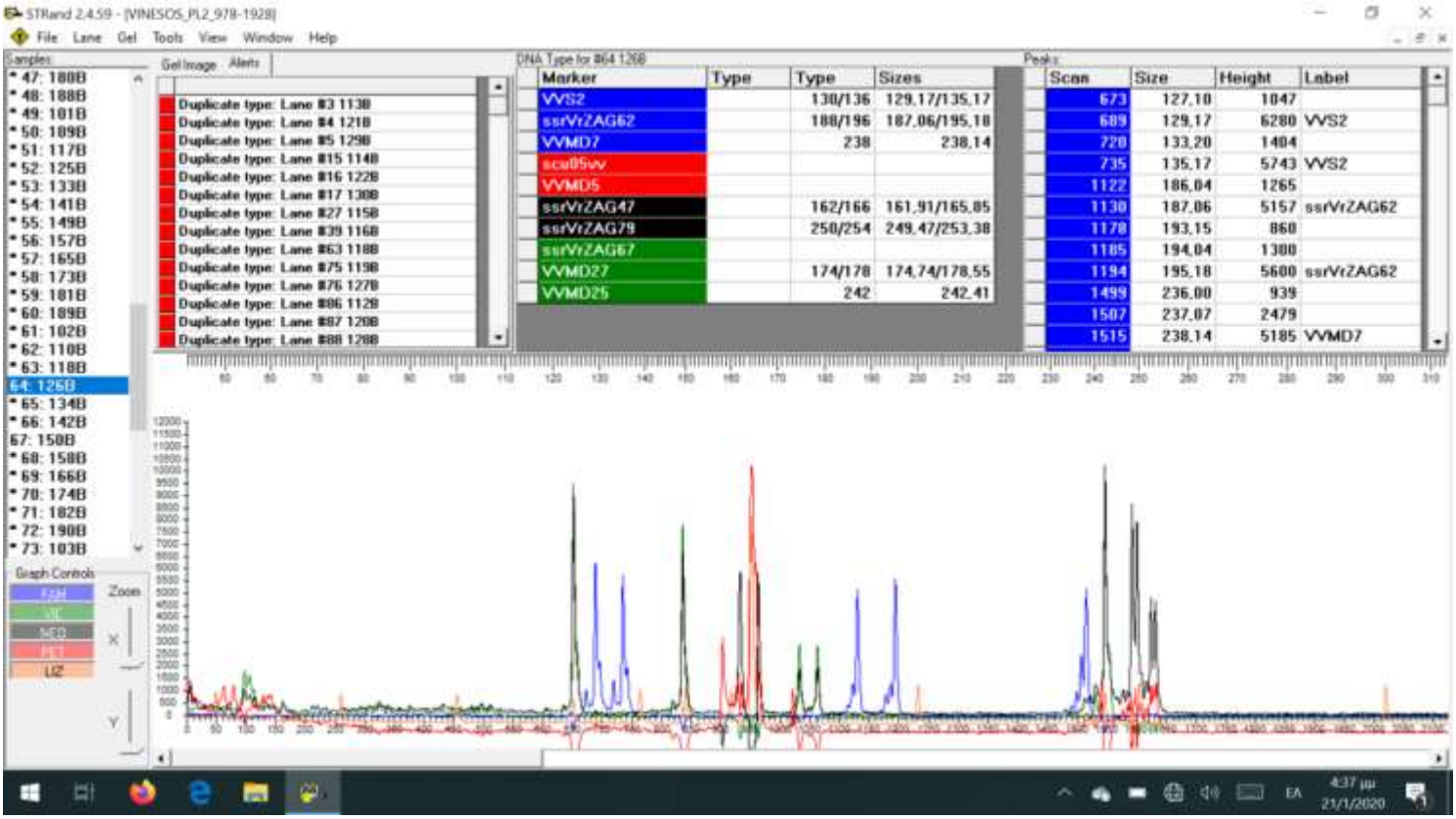

Figure 2. Genotyping (allele calling) of samples using the STRand software

As a first step for the statistical analysis, we defined populations based on producer, location and variety information. As a consequence, we had (in some cases) populations belonging to the same variety but coming from different producers and sampling areas. For example, we had Tamyanka variety from BRATANOV WINERY (producer) and SHISHMANOVO (location) but we also had Tamyanka variety from DIMITAR DJEMPERLIEV (producer) and DIMITROVCHE (location). These were treated as two different populations. According to preliminary results (data not shown), we noticed that all populations of the same variety were of very high genetic similarity to each other. For example, Keratsuda variety was represented from many different populations (producers and locations). However, all these populations were genetically similar (almost identical) to each other, regardless of the producer and location of each population. This scenario was evident (with no exception) for all the varieties studied herein. Consequently, we redefined the populations for the statistical analysis purposes, based exclusively on variety information (Table 3).

Table 3. Populations-varieties used in the statistical analysis

Population	Variety
population 1	TAMYANKA
population 2	MAVRUD
population 3	RUBIN
population 4	PAMID
population 5	DIMYAT
population 6	RUEN
population 7	SHIZOKA MELNISHKALOZA
population 8	MISHEJ SANDANSKI
population 9	KERATSUDA

The genetic variation parameters were calculated using GENETIX 4.05.2 and FSTAT 2.9.3.2 statistical packages. The FSTAT software was also used to test for statistically significant differentiation between all possible pairs of the nine studied populations. This was done by running multiple pairwise tests of differentiation with a 5% level of significance. In order to visualize the degree of differentiation among the nine varieties, we used factorial correspondence analysis (FCA) through GENETIX software. Population relationships were also explored by estimating F_{st} values, which are widely used to estimate the genetic distance between two populations.

Results

DNA extraction was successful for all 350 leaf samples, while amplification of DNA samples with the Multiplex PCR method and genotyping of PCR products in ABI 3500 Genetic Analyzer were also 100% successful. Regarding statistical analysis, the pairwise F_{st} values were quite high ranging from 0.1871 to 0.5184 (Table 4), meaning that there is a high degree of genetic differentiation (or low degree of genetic similarity) between the nine grapevine varieties.

Table 4. Estimated F_{st} values among all possible pairs of the nine studied populations

	POP1	POP2	POP3	POP4	POP5	POP6	POP7	POP8	POP9
POP1	**0.0000**	0.2701	0.2612	0.3977	0.3969	0.3859	0.3925	0.1871	0.4215
POP2	0.2701	**0.0000**	0.3536	0.3192	0.2987	0.3392	0.2762	0.2331	0.3735
POP3	0.2612	0.3536	**0.0000**	0.5107	0.5089	0.4338	0.4828	0.3977	0.4709
POP4	0.3977	0.3192	0.5107	**0.0000**	0.4031	0.4541	0.4671	0.4115	0.2524
POP5	0.3969	0.2987	0.5089	0.4031	**0.0000**	0.5184	0.5017	0.3947	0.4422
POP6	0.3859	0.3392	0.4338	0.4541	0.5184	**0.0000**	0.2919	0.3382	0.3587
POP7	0.3925	0.2762	0.4828	0.4671	0.5017	0.2919	**0.0000**	0.2933	0.3490
POP8	0.1871	0.2331	0.3977	0.4115	0.3947	0.3382	0.2933	**0.0000**	0.3852
POP9	0.4215	0.3735	0.4709	0.2524	0.4422	0.3587	0.3490	0.3852	**0.0000**

The pairwise tests of differentiation showed that all nine varieties are, with no exception, genetically different from each other with a statistical significance of 5%. This result verifies the high degree of genetic differentiation previously reported by F_{st} values. According to the factorial correspondence analysis (FCA), there is a clear separation of all nine varieties, confirming the low degree of genetic similarity between them. It is worth mentioning that TAMYANKA, RUBIN and RUEN varieties tend to isolate (genetically) more compared to all others (Figure 3).

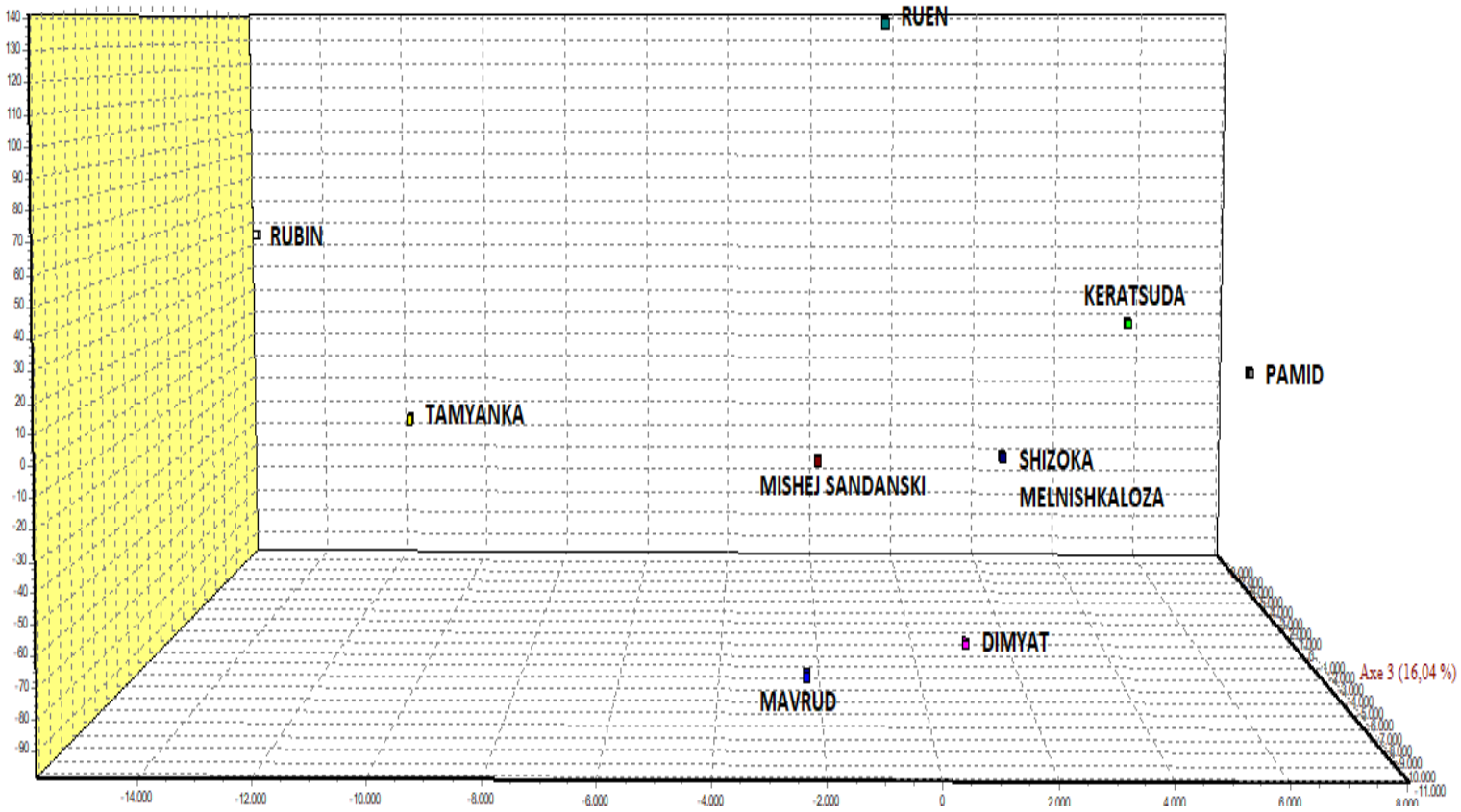

Figure 3. Factorial Correspondence Analysis (FCA) of the nine grapevine varieties

Conclusions

- All populations that are declared to belong to the same variety are genetically similar to each other, regardless of the producer and/or location they come from. This scenario applies (with no exception) in all varieties of the present study.

- Our findings show that genetic differentiation between the nine studied grapevine varieties is quite high.

- All nine varieties are, with no exception, genetically different from each other, which is supported statistically with a 5% level of significance.

- TAMYANKA, RUBIN and RUEN varieties tend to isolate (genetically) more compared to all other varieties.

- The genetic data produced herein give us the opportunity to establish a genetic database consisting of the genetic profiles of nine Bulgarian grapevine varieties, based on genotyping in 10 microsatellite markers. This can be used as a multi-purpose reference database. For example, if we have a leaf sample of unknown variety we can examine if that belongs to one of the nine varieties, by genotyping it (with low cost) in the above 10 molecular markers.

Project Title: "VineSOS", Project Code 1829
Co-funded by the European Regional Development Fund and by national funds of the countries in Interreg VA "Greece-Bulgaria 2014-2020" Cooperation Programme
http://www.greece-bulgaria.eu/
The content of this material does not necessarily represent the official position of the European Union