Luka Marinov[1], Goran Zdunić[1]*, Gabriele Di Gaspero[2], Gabriele Magris[2,3], Michele Morgante[2,3], Katarina Lukšić[1], Katarina Hančević[1], Edi Maletić[4], Marijan Bubola[5], Ivan Pejić[6]

Analysis of Croatian wild and cultivated grapevine diversity by genotyping by sequencing

Affiliations
[1]Institute for Adriatic Crops and Karst Reclamation, Split, Croatia
[2]Istituto di Genomica Applicata, Udine, Italy
[3]University of Udine, Department of Agricultural, Food, Environmental and Animal Sciences, Udine, Italy
[4]University of Zagreb, Faculty of Agriculture, Department of Viticulture and Enology, Zagreb, Croatia
[5]Institute of Agriculture and Tourism, Poreč, Croatia
[6]University of Zagreb, Faculty of Agriculture, Department of Plant Breeding, Genetics and Biometrics, Zagreb, Croatia

Correspondence
Luka Marinov: luka.marinov@krs.hr, Goran Zdunić*: goran.zdunic@krs.hr, Gabriele Di Gaspero: digaspero@appliedgenomics.org, Gabriele Magris: gmagris@appliedgenomics.org, Michele Morgante: michele.morgante@uniud.it; Katarina Lukšić: katarina.luksic@krs.hr, Katarina Hančević: katarina.hancevic@krs.hr, Edi Maletić: emaletic@agr.hr, Marijan Bubola: marijan@iptpo.hr, Ivan Pejić: ipejic@agr.hr

Summary

Minor varieties represent a significant part of the grapevine germplasm (*Vitis vinifera* L. subsp. *vinifera*) in Croatia. During the long history of grape cultivation in insular, coastal and continental Croatia, numerous local varieties were cultivated, many of which are still used today. There are also several populations of wild grapevine (*Vitis vinifera* subsp. *sylvestris* Hegi Gmel) that have survived until the present in natural sites. Here, we developed a single primer enrichment technology (SPET) panel consisting of 61,308 probes for targeted sequencing of the *V. vinifera* gene space. In doing so, we examined a total of 28,092 gene models, 88.2% of all predicted genes, interrogating an average of 19 Mb nucleotides per individual genome, which corresponds to 4% of the haploid genome length. This is the first time that SPET-based sequencing has been applied to Croatian grapevine germplasm to generate multilocus genotype data of 126 cultivated accessions and 50 wild specimens. We identified 531,900 variants sites, 208,802 of which reside in coding sequences, 140,836 in introns, 118,416 in UTRs, and 63,424 in the nearby intergenic space that revealed the presence of 33 clonally propagated specimens representing both synonymies and clones within local varieties (22) as well as synonymies between local and international varieties (11), mostly from neighbouring countries. To investigate possible contribution of wild grapevine (*Vitis vinifera* subsp. *sylvestris* Hegi Gmel) to the development of present cultivars, four wild grapevine populations were included. None of the cultivated varieties showed close kinship with local wild specimens. This study opens new possibilities for studying the genetic diversity of Croatian grapevine germplasm and provides additional information with respect to SSR genotyping and phenotyping.

Keywords

single nucleotide polymorphism, native grapevines, Croatian germplasm, first-degree relationship, genetic diversity

Introduction

Research on Croatian grapevine germplasm began in the early 19th century. One of the most important ampelographic studies was presented by Bulić (1949) with the description of 200 varieties grown in Dalmatia. The occurrence of wild grapevine (*Vitis vinifera* subsp. *sylvestris*) on Croatian territory was documented by Šulek in the late 19th century and the first detailed morphological descriptions were published by Turković (1953). Much of the genetic diversity was lost in the last century due to phylloxera (*Daktulosphaira vitifoliae*) and the adoption of international or high yield varieties. Nevertheless, a high phenotypic diversity can be observed in Croatian grapevine germplasm. Set of 140 grapevine varieties were described with OIV descriptors and genetic profiles (9 SSR markers), 103 of them considered traditional Croatian varieties (Maletić *et al.*, 2015). Later, 145 unique accessions of varieties were analysed with a set of 36 microsatellite (SSR) markers, reporting 24 full parentages and identifying 'Plavac Mali' as the variety with the highest number of parent-offspring relationships within the Croatian set (Žulj Mihaljević *et al.*, 2020). Analysis of wild accessions from Croatia, the Mediterranean region, and Central Asia with SSR markers revealed a decrease in diversity among *sylvestris* from east to west, high haplotype diversity, and allele richness of accessions from the Middle East and Caucasus, suggesting a primary centre of domestication (Butorac *et al.*, 2018; Riaz *et al.*, 2018; Zdunić *et al.*, 2013, 2019, 2020).

During the migration events, there were two main transit routes of grapevine varieties towards Europe, the northern

Submitted/accepted for publication: January 31, 2023/June 5, 2023

route from the Middle East through the Balkans and Central Europe, and the Southern Mediterranean route through Egypt to Spain (Bacilieri *et al.*, 2013). It has been shown that the identification of clear geographic patterns of genetic diversity is difficult at a small geographic scale or using small germplasm collections (Aradhya *et al.*, 2003). Nevertheless, with analysis of 20 SSR markers Bacilieri *et al.* (2013) have confirmed the eco-geographic grouping proposed by Negrul using a large set of genotypes. The varieties from the East (Eastern Mediterranean, Caucasus and Middle East) appeared to be the most diverse in terms of allele number, number of private alleles and unbiased heterozygosity, which is consistent with the hypothesis of an Eastern event of grapevine domestication (Aradhya *et al.*, 2003; Arroyo-García *et al.*, 2006; Bacilieri *et al.*, 2013). Analysis of the genetic structure of Croatian germplasm, based on 36 SSR markers, confirmed that the Croatian germplasm is predominantly part of the Balkan grape gene pool (Žulj Mihaljević *et al.*, 2020). Although most of varieties grown in Croatia are used for winemaking, Magris *et al.* (2021) pointed out that most Mediterranean wine grapes from the Balkans, Sicily, and southern Italy largely correspond to *pontica balcanica* and appear to be genetically more similar to *orientalis* ancestors (table grapes) than to *pontica georgica* ancestors (Caucasian wine grapes), which was concluded on basis of 6357 SNPs.

Next-generation DNA sequencing (NGS) technologies have led to rapid and genome-wide detection of Single Nucleotide Polymorphisms (SNP) in various plant species (Taranto *et al.*, 2016; Alipour *et al.*, 2017; D'Agostino *et al.*, 2018). For identification and pedigree reconstruction, SSR markers are still the most commonly used genotyping markers in long-lived and vegetatively propagated perennials, although SNPs are biallelic and less hypervariable than SSR markers, which may represent an advantage rather than a limitation for pedigree reconstruction in those species. SNPs are abundant in the genome and genotypic calling is relatively easy to perform in a reproducible and cost-effective manner, as well as associations with phenotypic differences (Rafalski, 2002; Deschamps *et al.*, 2012; Scheben *et al.*, 2017). In order to define the events that have led to the origin of varieties it is necessary to investigate as much grapevine genetic diversity as possible. The development of 'Vitis9kSNP' (Myles *et al.*, 2010) and 'Vitis18KSNP' arrays (https://urgi.versailles.inra.fr/Projects/Achieved-projects/GrapeReSeq) led to a more comprehensive SNP analysis of large grapevine collections (Laucou *et al.*, 2018; D'Onofrio *et al.*, 2021). SNP analysis allowed reliable analysis of parentage and genetic structure, confirmation of the initial domestication in the Middle East, and detection of introgression from wild *sylvestris* (Myles *et al.*, 2011; Laucou *et al.*, 2018; D'Onofrio *et al.*, 2021; Dong *et al.*, 2023). In addition to the widely used genotyping arrays, whole-genome sequencing (WGS) data analysis was performed on 472 *Vitis* accessions (Liang *et al.*, 2019) and 204 *V. vinifera* accessions (Magris *et al.*, 2021) to discover a large number of SNPs that revealed associations between SNPs and desirable agronomic traits and provided better insight into evolutionary biology.

To date, only one genotyping study with SNP markers has been performed on Croatian grapevine germplasm using a small panel of 48 pre-ascertained SNPs (Žulj Mihaljević *et al.*, 2020). As part of NGS, genotyping by sequencing (GBS) is a widely used method for simultaneous SNP discovery and SNP genotyping. A major limitation of GBS is the reliance on restriction enzymes to generate fragment sizes to be captured and sequenced and thus the inability to generate markers that are systematically localized within or nearby genes and could have functional significance (Elshire *et al.*, 2011; Scheben *et al.*, 2017; Barchi *et al.*, 2019). However, recent studies have introduced a new GBS concept that combines a highly reproducible and cost-effective approach with precise targeting of reduce-representation genome sequencing to the desired genomic fractions (Scaglione *et al.*, 2019). SNP discovery by single primer enrichment technology (SPET) has already been successfully applied to annual (Scaglione *et al.*, 2019) and horticultural (Gramazio *et al.*, 2020) species, as well as to perennial crops such as apricot (Baccichet *et al.*, 2022).

In this study, we present a GBS analysis protocol for Croatian germplasm using a targeted sequencing system based on single primer enrichment technology. Using this method with state-of-the-art sequencing platforms allowed us both to discover novel polymorphisms that increase the depth of the analysis and to score known polymorphisms that allow for cross-comparisons with existing datasets. We verified the presence of homonyms and synonyms and determined the identity of Croatian cultivated and wild grapevine accessions with more than 500 000 SNP markers.

Material and Methods

Plant material and DNA extraction

A total of 176 grapevine accessions, including 126 cultivated (*V. vinifera* subsp. *vinifera*) and 50 wild accessions (*Vitis vinifera* subsp. *sylvestris*) were collected and genotyped. Plant material of varieties was obtained from the germplasm repositories at the Institute for Adriatic Crops and Karst Reclamation (Split) (83), the Institute of Agriculture and Tourism (Poreč) (10), and the University of Zagreb Faculty of Agriculture (Zagreb) (33). Wild specimens were collected in their natural habitat at three different and distanced sites in Croatia (Psunj, Paklenica, Modro jezero) and one in Bosnia and Herzegovina (Cerovica). The geographical location of germplasm repositories and wild populations is shown in Fig. 1, while the complete list of analysed accessions can be found in Supplementary table 1.

Three to five young leaves were collected from individual plants and dried in silica gel. DNA was extracted using the NucleoSpin Plant II kit (Macherey-Nagel, Düren, Germany). Genomic DNA was quantified using the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA).

Probe design

Probes were designed to target CDS, introns, and UTRs in the *V. vinifera* gene space using the 12Xv0 assembly of the strain PN40024 (GCA_000003745.2) and the gene models in the V2.1 version of gene prediction. Targets for probe design were defined as follows. For monoexonic genes, one target was placed 25 bp 3'–apart from the 5' end of the exon and the other target was placed 25 bp 5'–apart from the 3' end of the exon. For mo-

Figure 1: Sampling locations. Blue circle represents repositories (1- Institute for Adriatic Crops and Karst Reclamation, Split, 2- Institute for Agriculture and Tourism, Poreč, 3- Faculty of Agriculture, Zagreb), red rectangle represents sampling locations of spontaneous populations (1- Cerovica, Bosnia and Herzegovina, 2- Modro jezero, Croatia, 3- Paklenica, Croatia, 4- Psunj, Croatia). Source: Google maps

noexonic genes > 500 bp in length, a third target was placed in the middle position of the predicted transcript. For a random half of the monoexonic genes, probes were designed 3' to the targets in antisense orientation to sequence intervals that include each target and the 5' region. For the other half probes were designed 5' to the targets in sense orientation to sequence intervals that include each target and the 3' region. For multiexonic genes regardless of their size, one target was placed 25 bp 3'–apart from the start of the first exon, the second target was placed 25 bp 5'–apart from the end of the last exon and the third target was placed in the middle position of the internal exon that was situated the closest to the middle point of the transcript. For a random half of the multiexonic genes, probes were designed 5' to each target in sense orientation and for the other half 3' to each target in antisense orientation. One of the targets and the corresponding probe within each gene was removed if two targets were < 200 bp apart. The targets and the corresponding probe were also removed if the target site and a ± 75-bp interval overlapped with a predicted transposable element or any other type of repetitive DNA. As a result, 68,782 targets were used for primer design and oligonucleotides could be synthesized by Tecan Trading AG (Männedorf, Switzerland) for 61,308 targets. As many as 28,092 genes out of the 31,845 gene models of the V2.1 gene prediction is targeted by at least one primer included into the SPET primer panel.

Library preparation and SPET sequencing

Libraries were prepared using the Allegro Targeted Genotyping Kit (Tecan Trading AG) with 100 ng of DNA as input and following the manufacturer's instructions. Libraries were quantified using a Qubit 2.0 Fluorometer and their size was checked using the High Sensitivity DNA assay from Caliper LabChip GX (Caliper Life Sciences, Alameda CA, USA). Libraries were sequenced using an Illumina NovaSeq 6000 sequencer (Illumina, San Carlos, CA, USA).

Bioinformatics analysis

Raw sequences were trimmed for quality and adapter sequences using ERNE (del Fabbro *et al.*, 2013) and Cutadapt (Martin, 2011) with default parameters. Reads were aligned to the reference genome (GCA_000003745.2) using BWA-MEM (Li, 2013; Li and Durbin, 2009) with default parameters. Raw variant sites were called using uniquely mapping reads and GATK Haplotype Caller (Poplin *et al.*, 2018). Individual genotypic calls were retained if read coverage was comprised between 10 and 1000. We used the thresholds of reference/alternative allele coverage ratio between 0.25 and 0.75 for calling heterozygous genotypes. Homozygous genotypes were called with no read carrying mismatches. Variant sites were retained if genotype calls passed the filters in > 50% of the accessions. Variant sites overlapping with primers sequences or repetitive DNA intervals were removed from the dataset. The first report of a relationship analysis based on identity-by-state (IBS), was used to identify possible synonyms between our dataset using PLINK version 1.07. (Purcell *et al.*, 2007).

Genealogical relationships

Genealogical relationships were analysed with pairwise identical-by-descent (IBD) estimated in 100 Kb windows of non-repetitive DNA as described in Magris *et al.*, 2021. The degree of kinship was assigned based on the length and distribution of IBD = 0 (no shared haplotypes), IBD = 1 (one shared haplotype) and IBD = 2 (two shared haplotypes) windows using the following parameters. Parent-offspring: IBD = 0 cumulative length < 50 Mb, no IBD = 0 segment > 3 Mb; IBD = 1 cumulative length > 50% of the genome size, standard deviation across the genome (st.dev.genome) of IBD = 0 segment length <0.1 Mb; clonal variants or duplicated samples: IBD = 0 < 3%, IBD = 1 > 3%, IBD = 2 > 90% of the genome length (Magris *et al.*, 2021).

Results and Discussion

For identity determination and diversity analysis of Croatian germplasm we genotyped 126 *V. vinifera* accessions and 50 *V. sylvestris* individuals. Genotyping was performed using GBS based on SPET, which allowed us to discover a large number of SNPs that proved to be a reliable input for the analysis of genealogical relationships with pairwise IBD, as used in similar studies on grapevine with SNP markers (D'Onofrio *et al.*, 2021).

The designed single primer enrichment technology (SPET) panel consisted of 61,308 probes for targeted sequencing of the *V. vinifera* gene space, which collectively investigated 28,092 gene models, 88.2% of all predicted genes, at an average of 19 Mb nucleotides per individual genome, which corresponds to 4% of the haploid genome length. Using SPET-based sequencing, we generated multilocus genotype data of 178 grapevine (*V. vinifera*) accessions at 531,900 variant sites, 208,802 of which reside in coding sequences, 140,836 in introns, 118,416 in UTRs, and 63,424 in the nearby intergenic space. Data were filtered for repetitive regions and homozygous calls with either 0 or 1 allele read coverage ratio. After filtering for MAF < 0.05, a total of 212,120 SNPs were retained for further analysis.

An initial relationship analysis based on IBS was performed in PLINK to identify synonyms and clonal replicates in our dataset. Total of 17 clonal replicates among varieties were found (Table 1). Nine clones of 'Tribidrag', four clones of 'Pošip bijeli', two clones of 'Malvazija istarska' and two clones of 'Welschriesling' were detected. Within the dataset, five cases of synonymy and somatic mutations were observed. Synonymy between 'Chasselas blanc' and 'Chasselas rouge', 'Plavac mali crni' and 'Plavac mali sivi', 'Babić' and 'Rogoznička', 'Vlaška' and 'Tanetova loza', 'Maraština' and 'Rukatac' was already known (Maletić *et al.*, 2015; Žulj Mihaljević *et al.*, 2020). In this dataset, 11 varieties matched SNP profiles of accessions present in a WGS dataset

Table 1: List of accessions with the same genotype (synonyms) based on SNP markers and calculated Identity by State (IBS) coefficients within the analysed data set (internal matches) and compared to the Whole Genome Sequencing (WGS) dataset of Magris *et al.* 2021.

No.	Accession 1	Internal matches	Matches with WGS data set (Magris *et al.*, 2021)
1	Pribidrag (HR07)	Tribidrag VV079 (HR13), Tribidrag IP130 (HR14), Zinfadel 01 (HR31), Prim FPS03 (HR10), Prim FPS 06 (HR09), Zin-Her 1-10 (HR08), Zin-Her 1-24 (HR11), Zin-Her 1-31 (HR12)	Tribidrag
2	Pošip bijeli obični (HR05)	Pošip B1 (HR03), Pošip B2 (HR04), Pošip sitni (HR06)	
3	Rukatac (HR01)	Maraština (HR02)	Malvasia Bianca Lunga
4	Graševina (HR17)	Graševina OB402 (TLV2)	Welschriesling
5	Babić (HR29)	Rogoznička (HR35)	
6	Plemenka bijela (HR46)	Plemenka crvena (HR84)	Chasselas Blanc
7	Plavac mali crni (HR33)	Plavac Mali Sivi (HR65)	
8	Malvazija istarska (IPTPO01)	Malvasija VCR 26 (TLV1)	Malvasia Istriana
9	Vlaška (HR15)	Tanetova loza (HR16)	
10	Belina starohrvatska (AF13)		Heunisch Weiss
11	Malvasia dubrovačka (HR21)		Malvasia di Lipari
12	Verdić (AF19)		Glera
13	Muškat momjanski (IPTPO06)		Muscat à Petits Grains Blancs
14	Bombino (HR44)		Bombino Bianco
15	Cabernet sauvingon (IPTPO12)		Cabernet Sauvignon
16	PS 177 (Sy21)	PS 177/a (Sy22), PS 178 (Sy23), PS 180 (Sy25), PS 180/a (Sy26), PS 181 (Sy27)	
17	PS NN 1 (Sy31)	PS NN 2 (Sy32), PS NN 4 (Sy34)	
18	PAK 88 (Sy37)	IJK88 (TLV4)	
19	PAK 110 (Sy39)	IJK120 (TLV5)	

(Magris *et al.*, 2021). All synonyms found are in agreement with previous results based on SSR markers (Maletić *et al.*, 2015; Žulj Mihaljević *et al.*, 2020). Among wild accessions of *V. sylvestris*, five identical to 'PS 177', two identical to 'PS NN 1', one identical to 'PAK88' and one identical to 'PAK110' accessions were found, although leaf samples from Psunj location were taken from different plants in close proximity (3-5 m). Thus, it can be assumed that some wild individuals vegetatively propagated in their habitat.

After removing redundant genotypes, the IBD data were used for genealogical relationship analysis. IBD analysis allowed us to reveal the complete pedigree of 14 varieties. Thirteen parentages were already proposed based on 20 SSR markers (Žulj Mihaljević *et al.*, 2020), but one novel parentage of cv. 'Surina' was discovered. Within the Croatian germplasm, the parent-offspring relationship was determined for 54 varieties (42,8%) with the second parent missing in the analysed set. The analysis identified 'Plavac mali' as the main progenitor in our data set with the highest number of parent-offspring (PO) relationships (13), followed by 'Bombino bianco' with 11 PO, 'Belina starohrvatska' with 10 PO, 'Bljuzgavac' with 8 PO and 'Tribidrag' with 6 PO relationships. The genealogical results and the constructed map of the studied grapevine accession will be elaborated elsewhere after the validation of the pedigree analysis for all possible trios. Here, as an example, we present the full pedigree of cv. 'Pošip bijeli' with shared chromosomal segments (Fig. 2), which supports the previously reported pedigree with SSR markers (Piljac *et al.*, 2002; Žulj Mihaljević *et al.*, 2020). Haplotype sharing and the cumulative length of IBD = 0 between P1 and P2 indicates a low degree of relatedness between the parents. Mendelian inconsistencies that seemingly conflict with the proposed direction of the trio are only 0.66% of the analysed SNPs and they are randomly distributed across the genome, suggesting that they may be attributed to genotyping errors and that the genotyping error rate is extremely low.

Croatian grapevine germplasm has been genotyped several times with microsatellite markers (SSR) (Maletić *et al.*, 2015; Žulj Mihaljević *et al.*, 2020; Zdunić *et al.*, 2013, 2020). This study represents the first preliminary application of genotyping by sequencing (GBS) based on a targeted sequencing system of single primer enrichment technology (SPET). The application of this method on state-of-the-art sequencing platforms allowed

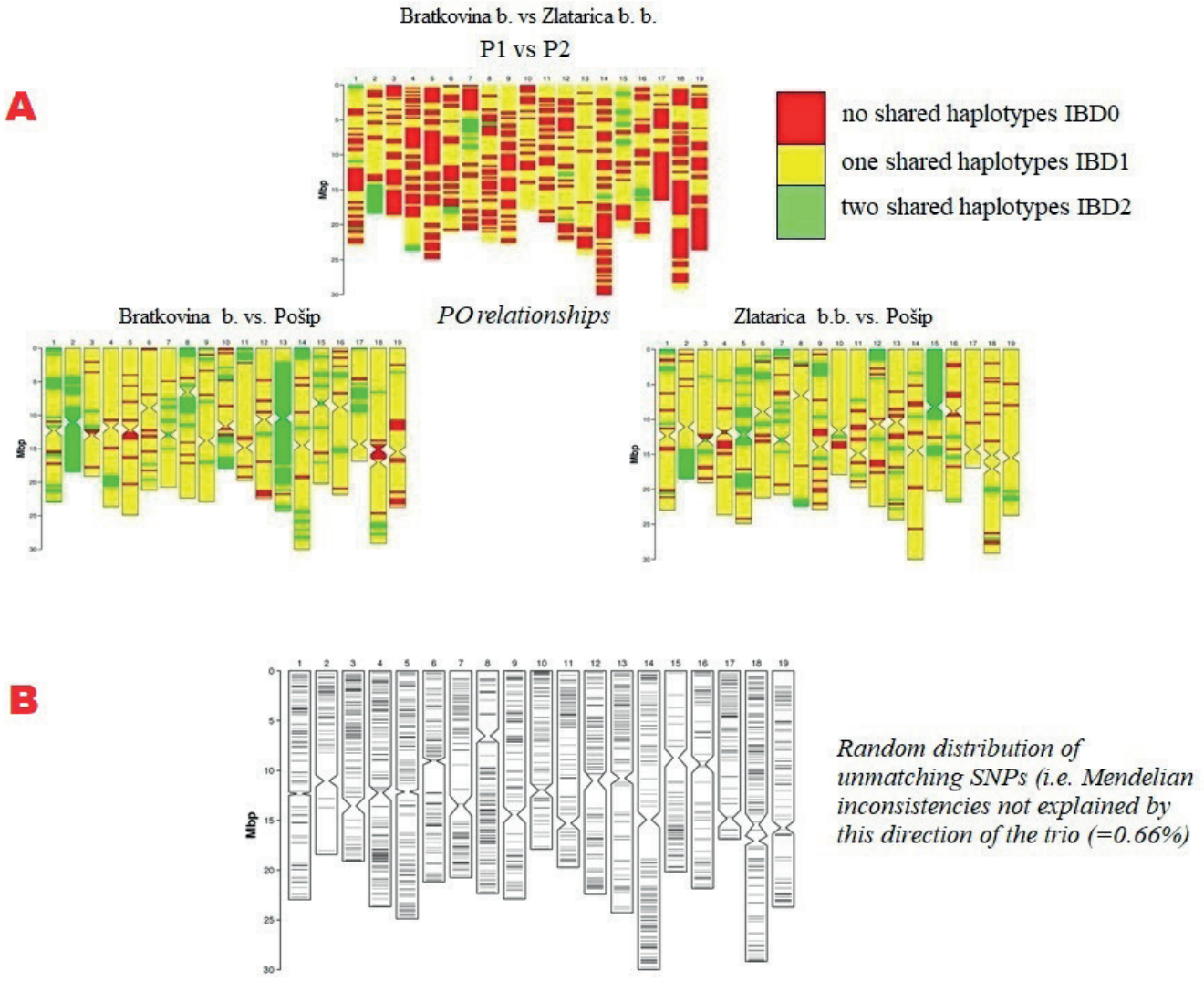

Figure 2: An example of the graphical representation of pairwise IBD relationship for cv. 'Pošip Bijeli' (*Vitis vinifera* L.) complete pedigree and random distribution of unmatching SNP's (A), Mendelian inconsistencies distributed across genome (B)

us to discover more than 500 000 sites of polymorphisms that reside in coding sequences, in introns, in UTRs, and in the nearby intergenic space. The large number of SNP markers obtained allowed us to gain a more detailed insight into the genealogical relationships between the observed accessions. SNP data will be used in future research to perform principal component analysis, to analyse population structure, and to study genetic diversity and intravarietal variability. The SNPs derived from the Croatian germplasm will allow us to combine genetic data and phenotypic data in genome-wide association studies (GWAS) and help us identify gene loci that control some phenotypic traits beneficial to grapevine breeding.

Acknowledgement

This research was supported by the project "Assessment of drought tolerance of Croatian grapevine germplasm-TOLVIN" (grant number: KK.05.1.1.02.0010) funded by the European Regional Development Fund (ERDF) and by the Croatian Science Foundation under the project "Young researchers' career development—training new doctoral students".

Conflicts of interest statement

The authors declare that they do not have any conflicts of interest.

References

Alipour, H., Bihamta, M. R., Mohammadi, V., Peyghambari, S. A., Bai, G., Zhang, G., 2017: Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. Frontiers in Plant Science, 8. DOI: 10.3389/fpls.2017.01293.

Aradhya, M. K., Dangl, G. S., Prins, B. H., Boursiquot, J. M., Walker, M. A., Meredith, C. P., Simon, C. J., 2003: Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. Genetical Research, 81 (3), 179–192. DOI: 10.1017/S0016672303006177.

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., Gorislavets, S., Grando, M. S., Klein, B. Y., McGovern, P. E., Merdinoglu, D., Pejić, 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. Molecular Ecology, 15 (12), 3707–3714. DOI: 10.1111/j.1365-294X.2006.03049.x.

Baccichet, I., Chiozzotto, R., Scaglione, D., Bassi, D., Rossini, L., Cirilli, M., 2022: Genetic dissection of fruit maturity date in apricot (*P. armeniaca* L.) through a Single Primer Enrichment Technology (SPET) approach. *BMC* Genomics, 23 (1), 712. DOI: 10.1186/s12864-022-08901-1.

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 Biology, 13 (1). DOI: 10.1186/1471-2229-13-25.

Barchi, L., Acquadro, A., Alonso, D., Aprea, G., Bassolino, L., Demurtas, O., Ferrante, P., Gramazio, P., Mini, P., Portis, E., Scaglione, D., Toppino, L., Vilanova, S., Díez, M. J., Rotino, G. L., Lanteri, S., Prohens, J., Giuliano, G., 2019: Single Primer Enrichment Technology (SPET) for high-throughput genotyping in Tomato and Eggplant germplasm. Frontiers in Plant Science, 10. DOI: 10.3389/fpls.2019.01005.

Butorac, L., Hančević, K., Lukšić, K., Škvorc, Ž., Leko, M., Maul, E., & Zdunić, G., 2018: Assessment of wild grapevine (*Vitis vinifera* ssp. *sylvestris*) chlorotypes and accompanying woody species in the Eastern Adriatic region. PLoS ONE, 13(6). DOI: 10.1371/journal.pone.0199495.

D'Agostino, N., Taranto, F., Camposeo, S., Mangini, G., Fanelli, V., Gadaleta, S., Miazzi, M. M., Pavan, S., di Rienzo, V., Sabetta, W., Lombardo, L., Zelasco, S., Perri, E., Lotti, C., Ciani, E., Montemurro, C., 2018: GBS-derived SNP catalogue unveiled wide genetic variability and geographical relationships of Italian olive cultivars. Scientific Reports, 8 (1). DOI: 10.1038/s41598-018-34207-y.

del Fabbro, C., Scalabrin, S., Morgante, M., Giorgi, F. M., 2013: An extensive evaluation of read trimming effects on illumina NGS data analysis. PLoS ONE, 8 (12). DOI: 10.1371/journal.pone.0085024.

Deschamps, S., Llaca, V., May, G. D., 2012: Genotyping-by-sequencing in plants. Biology, 1 (3), 460–483. DOI: 10.3390/biology1030460.

Dong, Y., Duan, S., Xia, Q., Liang, Z., Dong, X., Margaryan, K., Musayev, M., Goryslavets, S., Zdunić, G., Bert, P.-F., Lacombe, T., Maul, E., Nick, P., Bitskinashvili, K., Dénes Bisztray, G., Drori, E., De Lorenzis, G., Cunha, J., Florentina Popescu, C., Arroyo-Garcia, R., Arnold, C., Ergül, A., Zhu, Y., Ma, C., Wang, S., Liu, S., Tang, L., Wang, C., Li, D., Pan, Y., Li, J., Yang, L., Li, X., Xiang, G., Yang, Z., Chen, B., Dai, Z., Wang, Y., Carrasco, D., Rodriguez-Izquierdo, A., Muñoz, G., Uysal, T., Özer, C., Kazan, K., Xu, M., Wang, Y., Zhu, S., Lu, J., Zhao, M., Wang, L., Jiu, S., Zhang, Y., Sun, L., Yang, H., Weiss, E., Wang, S., Zhu, Y., Li, S., Sheng, J., Chen, W., 2023: Dual domestications and origin of traits in grapevine evolution. *Science* 379 (6635), 892-901. DOI: 10.1126/science.add8655.

D'Onofrio, C., Tumino, G., Gardiman, M., Crespan, M., Bignami, C., de Palma, L., Barbagallo, M. G., Muganu, M., Morcia, C., Novello, V., Schneider, A., Terzi, V., 2021: Parentage atlas of Italian grapevine varieties as inferred from SNP genotyping. Frontiers in Plant Science, 11. DOI: 10.3389/fpls.2020.605934.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., Mitchell, S. E., 2011: A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE, 6 (5). DOI: 10.1371/journal.pone.0019379.

Gramazio, P., Jaén-Molina, R., Vilanova, S., Prohens, J., Marrero, Á., Caujapé-Castells, J., Anderson, G. J., 2020: Fostering conservation via an integrated use of conventional approaches and high-throughput SPET genotyping: A case study using the endangered Canarian endemics *Solanum lidii* and *S.*

vespertilio (*Solanaceae*). Frontiers in Plant Science, 11. DOI: 10.3389/fpls.2020.00757.

Laucou, V., Launay, A., Bacilieri, R., Lacombe, T., Adam-Blondon, A. F., Bérard, A., Chauveau, A., de Andrés, M. T., Hausmann, L., Ibáñez, J., le Paslier, M. C., Maghradze, D., Martinez-Zapater, J., Maul, E., Ponnaiah, M., Töpfer, R., Péros, J. P., Boursiquot, J. M., 2018: Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. PLoS ONE, 13 (2). DOI: 10.1371/journal.pone.0192540.

Li, H., 2013: Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. http://arxiv.org/abs/1303.3997.

Li, H., Durbin, R., 2009: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25 (14), 1754–1760. DOI: 10.1093/bioinformatics/btp324.

Liang, Z., Duan, S., Sheng, J., Zhu, S., Ni, X., Shao, J., Liu, C., Nick, P., Du, F., Fan, P., Mao, R., Zhu, Y., Deng, W., Yang, M., Huang, H., Liu, Y., Ding, Y., Liu, X., Jiang, J., Zhu, Y., Li, S., He, X., Chen, W., Dong, Y., 2019: Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. Nature Communications, 10 (1). DOI: 10.1038/s41467-019-09135-8.

Magris, G., Jurman, I., Fornasiero, A., Paparelli, E., Schwope, R., Marroni, F., di Gaspero, G., Morgante, M., 2021: The genomes of 204 *Vitis vinifera* accessions reveal the origin of European wine grapes. Nature Communications, 12 (1). DOI: 10.1038/s41467-021-27487-y.

Maletić, E., Pejić, I., Karoglan Kontić, J., Zdunić, G., Preiner, D., Šimon, S., Andabaka, Ž., Žulj Mihaljević, M., Bubola, M., Marković, Z., Stupić, D., Mucalo, A., 2015: Ampelographic and genetic characterization of Croatian grapevine varieties. Vitis 54 (Special Issue), 93-98) DOI: 10.5073/vitis.2015.54.special-issue.93-98.

Martin, M., 2011: Cutadapt removes adapter sequences from high-throughput sequencing reads. https://journal.embnet.org/index.php/embnetjournal/article/viewFile/200/458.

Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., Prins, B., Reynolds, A., Chia, J. M., Ware, D., Bustamante, C. D., Buckler, E. S., 2011: Genetic structure and domestication history of the grape. Proceedings of the National Academy of Sciences of the United States of America, 108 (9), 3530–3535. DOI: 10.1073/pnas.1009363108.

Myles, S., Chia, J. M., Hurwitz, B., Simon, C., Zhong, G. Y., Buckler, E., Ware, D., 2010: Rapid genomic characterization of the genus *Vitis*. PLoS ONE, 5 (1). DOI: 10.1371/journal.pone.0008219.

Piljac, J., Maletić, E., Karoglan Kontić, J., Dangl, G. S., Pejić, I., Mirošević, N., & Meredith, C. P., 2002: The parentage of Pošip bijeli, a major white wine cultivar of Croatia. https://www.researchgate.net/publication/291794502.

Poplin, R., Ruano-Rubio, V., Depristo, M. A., Fennell, T. J., Carneiro, M. O., van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., Macarthur, D. G., Banks, E., 2018: Scaling accurate genetic variant discovery to tens of thousands of samples. BioRxiv preprint, 1-22. DOI: 10.1101/201178.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., Sham, P. C., 2007: PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81 (3), 559–575. DOI: 10.1086/519795.

Rafalski, A., 2002: Applications of SNP Related papers Single Nucleotide Polymorphisms and Linkage Disequilibrium in Sunflower. Current Opinion in Plant Biology (Vol. 5). http://www.intl-pag.org/pag/9/

Riaz, S., de Lorenzis, G., Velasco, D., Koehmstedt, A., Maghradze, D., Bobokashvili, Z., Musayev, M., Zdunic, G., Laucou, V., Andrew Walker, M., Failla, O., Preece, J. E., Aradhya, M., Arroyo-Garcia, R., 2018:. Genetic diversity analysis of cultivated and wild grapevine (*Vitis vinifera* L.) accessions around the Mediterranean basin and Central Asia. *BMC Plant Biology*, *18*(1). DOI: 10.1186/s12870-018-1351-0

Scaglione, D., Pinosio, S., Marroni, F., Centa, E. di, Fornasiero, A., Magris, G., Scalabrin, S., Cattonaro, F., Taylor, G., Morgante, M., 2019: Single primer enrichment technology as a tool for massive genotyping: A benchmark on black poplar and maize. Annals of Botany, 124 (4), 543–551. DOI: 10.1093/aob/mcz054.

Scheben, A., Batley, J., Edwards, D., 2017: Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnology Journal 15 (2), 149–161). DOI: 10.1111/pbi.12645.

Taranto, F., D'Agostino, N., Greco, B., Cardi, T., Tripodi, P., 2016: Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. BMC Genomics, 17 (1). DOI: 10.1186/s12864-016-3297-7.

Zdunić, G., Lukšić, K., Nagy, Z. A., Mucalo, A., Hančević, K., Radić, T., Butorac, L., Jahnke, G. G., Kiss, E., Ledesma-Krist, G., Regvar, M., Likar, M., Piltaver, A., Žulj Mihaljević, M., Maletić, E., Pejić, I., Werling, M., Maul, E., 2020: Genetic structure and relationships among wild and cultivated grapevines from central Europe and part of the western Balkan peninsula. Genes 11 (9), 1–15. DOI: 10.3390/genes11090962.

Zdunić, G., Maul, E., Lukšić, K., Mucalo, A., Hančević, K., Radić, T., Butorac, L., Budić Leto, I., Leko, M., Regvar, M., Likar, M., Piltaver, A., Žulj Mihaljević, M., Pejić, I., Maletić, E., 2019: Inventory and descriptions of wild grapevine (*Vitis vinifera* subsp. *sylvestris*) from Slovenia, Croatia and Bosnia and Herzegovina. Acta Horticulturae 1248, 63–67. DOI: 10.17660/ActaHortic.2019.1248.9.

Zdunić, G., Preece, J. E., Aradhya, M., Velasco, D., Koehmstedt, A., Dangl, G. S., 2013: Genetic diversity and differentiation within and between cultivated (*Vitis vinifera* L. ssp. *sativa*) and wild (*Vitis vinifera* L. ssp. *sylvestris*) grapes. Vitis 52 (1), 29-32. DOI: 10.5073/vitis.2013.52.29-32.

Žulj Mihaljević, M. Ž., Maletić, E., Preiner, D., Zdunić, G., Bubola, M., Zyprian, E., Pejić, I., 2020: Genetic diversity, population structure, and parentage analysis of Croatian grapevine germplasm. Genes, 11(7), 1–35. DOI: 10.3390/genes11070737.