



The Potential of HTS Approaches for Accurate Genotyping in Grapevine (*Vitis vinifera* L.)

Urban Kunej, Aida Dervishi, Valérie Laucou, Jernej Jakše, Nataša Štajner

► To cite this version:

Urban Kunej, Aida Dervishi, Valérie Laucou, Jernej Jakše, Nataša Štajner. The Potential of HTS Approaches for Accurate Genotyping in Grapevine (*Vitis vinifera* L.). *Genes*, 2020, 11 (8), 16 p. 10.3390/genes11080917 . hal-02990444

HAL Id: hal-02990444

<https://hal.inrae.fr/hal-02990444>




Submitted on 5 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

The Potential of HTS Approaches for Accurate Genotyping in Grapevine (*Vitis vinifera* L.)

Urban Kunej ¹, Aida Dervishi ² , Valérie Laucou ³, Jernej Jakše ¹  and Nataša Štajner ^{1,*} 

¹ Department of Agronomy, Biotechnical Faculty, University of Ljubljana, 1000 Ljubljana, Slovenia; urban.kunej@bf.uni-lj.si (U.K.); jernej.jakse@bf.uni-lj.si (J.J.)

² Department of Biotechnology, Faculty of Natural Sciences, University of Tirana, Blv Zog I, 25/1, 1001 Tirana, Albania; aida.dervishi@fshn.edu.al

³ AGAP, Univ Montpellier, CIRAD, INRAE, Montpellier SupAgro, F-34070 Montpellier, France; valerie.laucou@inrae.fr

* Correspondence: natasa.stajner@bf.uni-lj.si; Tel.: +386-1-3203-255

Received: 8 July 2020; Accepted: 6 August 2020; Published: 10 August 2020



Abstract: The main challenge associated with genotyping based on conventional length polymorphisms is the cross-laboratory standardization of allele sizes. This step requires the inclusion of standards and manual sizing to avoid false results. Capillary electrophoresis (CE) approaches limit the information to the length polymorphism and do not allow the determination of a complete marker sequence. As an alternative, high-throughput sequencing (HTS) offers complete information regarding marker sequences and their flanking regions. In this work, we investigated the suitability of a semi-quantitative sequencing approach for microsatellite genotyping using Illumina paired-end technology. Twelve microsatellite loci that are well established for grapevine CE typing were analysed on 96 grapevine samples from six different countries. We redesigned primers to the length of the amplicon for short sequencing (~100 bp). The primer pair was flanked with a 10 bp overhang for the introduction of barcodes on both sides of the amplicon to enable high multiplexing. The highest data peaks were determined as simple sequence repeat (SSR) alleles and compared with the CE dataset based on 12 reference samples. The comparison showed that HTS SSR genotyping can successfully replace the CE system in further experiments. We believe that, with next-generation sequencing, genotyping can be improved in terms of its speed, accuracy, and price.

Keywords: *Vitis vinifera* L.; microsatellites; high-throughput sequencing; SSR markers; genotyping

1. Introduction

Molecular marker technologies have changed plant genetics research enormously since their introduction in the 1980s and have provided researchers with a tool that is able to analyse an unlimited number of markers independent of environmental influences. Since then, there has been an impressive improvement in technology, which has moved from single to highly multiplexed analysis that includes southern hybridization probing [1]; random and specific PCR amplification methods [2,3]; quantitative PCR approaches [4]; microarrays [5]; and, more recently, next-generation (NGS) sequencing for plant genotype determination [6]. This progress has led to numerous publications describing dense genetic maps [7], finding quantitative trait loci (QTL) of great agronomic interest [8] and completely genotyped germplasm resources [9], to name a few.

The characterization of plant varieties or germplasm resources, such as grapevine, *Vitis vinifera* L., is a requirement driven by related economic interests, seed certification, plant variety rights, and scientific knowledge. Molecular marker methods for variety identification have undoubted advantages, including microsatellites, which have proven to be powerful tools for the identity,

parentage, and kinship analysis of a wide range of plant species. Since their introduction in 1993 as a tool for plant genetic research [10], they have become one of the most widely used molecular markers in various fields of research, including plant genotyping. They are described as the best marker system for determining inter-variety polymorphisms [11]. Microsatellites, simple sequence repeats (SSRs) [12], or simple tandem repeat (STRs) [13] are the most commonly used DNA sequence features in plant genotyping due to their ubiquity in plants, their wide genomic distribution, their codominant inheritance, and their high degree of polymorphism [14,15]. Microsatellites are DNA regions consisting of tandem repeating units of 1–6 nucleotides. The number of repeats is highly variable between individuals due to the high rates of DNA polymerase slippage events [16] or unequal crossing-over [17], which makes them the ultimate multi-allelic marker system.

Microsatellite analysis is routinely based on multiplex fluorescence PCR, accompanied by capillary electrophoresis (CE) and the sizing of the resolved products. This is a fast and well-established technique with certain limitations: it is semi-quantitative, and the standardization of the identified alleles is required. When the CE methodology is applied between laboratories and the data subset is compared, the relative size values must be standardised against each other. In this step, manual sizing and processing are required, mainly due to the rounding of allele sizes, which must be very accurate to avoid false differences between samples from two data sets. The information provided by such an approach refers only to the length of the polymorphism and does not include the determination of the complete sequence of certain microsatellite loci.

Alternatively, new high-throughput sequencing platforms (HTS) enable the simultaneous sequencing of millions of sequences in a single run at enormous cost reductions [18]. The HTS analysis of microsatellite loci provides more information regarding SSR sequences, including the identification of sequence variants of STR loci that would be interesting for discriminating alleles, resolving mixed samples, and parentage analysis. Initial experiments successfully employed HTS platforms, such as Illumina and 454 sequencers for SSR genotyping, in the field of human forensic genetics, and showed the high applicability of powerful STR genotyping platforms [19,20]. Darby et al. [21] showed that such microsatellite genotyping is an ideal tool for population genetic structure studies, as it can detect a higher number of unique alleles compared to CE systems.

Recently, the term simple sequence repeats sequencing (SSRseq) was introduced to describe the application of HTS microsatellite genotyping. The authors developed a workflow for an efficient SSRseq setup for a wide range of situations [22]. In addition, the electrophoresis conditions associated with the polymer type [23], buffer conditions, or the use of alternative fluorescent dyes bound to primers [24] may also have an effect on DNA migration and the further sizing of microsatellite alleles. As denaturing electrophoresis resolves DNA fragments based on the length of the amplified alleles, fragments of equal length with different nucleotide compositions cannot be distinguished. This phenomenon is called size homoplasy [25], and can only be detected by sequencing the alleles.

In this work, the power of HTS for microsatellite genotyping was evaluated and a comparative genotyping study between HTS and a microsatellite CE analysis of grapevine cultivars was carried out. A standard set of 12 microsatellite loci was used to HTS-genotype 96 unique grapevine cultivars. In addition, a bioinformatic method is proposed using publicly available tools for sequence analysis. The microsatellite HTS analysis approach facilitates the high multiplexing capability of the loci and also allows the identification of variations that remain hidden in conventional SSR genotyping based on length polymorphisms.

2. Materials and Methods

2.1. SSRs and Cultivars

The genotyping of 96 grapevine cultivars (Table 1) obtained from six different countries (France, 12; Slovenia, 18; Bosnia and Herzegovina, 15; Serbia, 22; Montenegro, 5; Albania, 16; and North Macedonia, 8) was performed on 12 standard SSR loci using newly designed primers to shorten the

product length below 150 bp (Table 2). The primers were designed using the Primer3 software [26]. A subset of HTS data was compared with the CE data of cultivars from the French collection (Table 1), obtained in a previous study of grapevine SSR genotyping, performed at National Research Institute for Agriculture, Food and Environment (INRAE), France [27].

Table 1. The 96 cultivars analysed in this study, sorted by barcodes and assigned to their country of origin.

Barcode	Cultivar	Country of Origin	Barcode	Cultivar	Country of Origin
F1-R1	Kratošija	North Macedonia	F7-R1	GnetKras	Slovenia
F1-R2	Neznana Bela	Slovenia	F7-R2	Kratošija I	Montenegro
F1-R3	Rebula	Slovenia	F7-R3	Muštoš Feher	Serbia
F1-R4	Sremska Zelenika	Serbia	F7-R4	Rebula Portalis	Slovenia
F1-R5	Zimsko Belo	Serbia	F7-R5	Smederevka	Bosnia and Herzegovina
F1-R6	Manastirsko Belo	North Macedonia	F7-R6	Vranac	Bosnia and Herzegovina
F1-R7	Dobrogostina	Bosnia and Herzegovina	F7-R7	Belovina	North Macedonia
F1-R8	Godominka	Serbia	F7-R8	Gnjat	Slovenia
F2-R1	Bagrina	Serbia	F8-R1	Kreaca	Serbia
F2-R2	DrenakCrni	Serbia	F8-R2	Refosco	Slovenia
F2-R3	Kadarka Bela	Serbia	F8-R3	Stanušina	North Macedonia
F2-R4	Krkošija Šupljica	Serbia	F8-R4	Žametovka	Bosnia and Herzegovina
F2-R5	Prokupac	Bosnia and Herzegovina	F8-R5	Ohridsko Belo	North Macedonia
F2-R6	Ružica	Serbia	F8-R6	Refošk	Slovenia
F2-R7	Bela Zgodnja	Slovenia	F8-R7	DolgiGrozdi	Slovenia
F2-R8	Chardonnay	Slovenia	F8-R8	Gročanka	Serbia
F3-R1	Drenak	Bosnia and Herzegovina	F9-R1	PlovinaCrna	Serbia
F3-R2	Kadarka	Serbia	F9-R2	Rezaklija	Bosnia and Herzegovina
F3-R3	Kujundžuša	Bosnia and Herzegovina	F9-R3	Stari Rizling VI	Montenegro
F3-R4	Prokupac	Serbia	F9-R4	Žlozder	Bosnia and Herzegovina
F3-R5	Ružica V	Montenegro	F9-R5	Debinë e Zezë	Albania
F3-R6	TamjanikaCrna	Serbia	F9-R6	Kallmet	Albania
F3-R7	Merlot	Slovenia	F9-R7	Potek e Zezë	Albania
F3-R8	Beli Medenac	Serbia	F9-R8	Shesh i Zi	Albania
F4-R1	Ružica VI	Montenegro	F10-R1	Stambolleshë	Albania
F4-R2	TrbljanBeli	Serbia	F10-R2	Sheshi Bardhë	Albania
F4-R3	Pinot noir	Slovenia	F10-R3	Kosinjot	Albania
F4-R4	Bena	Bosnia and Herzegovina	F10-R4	Vlosh	Albania
F4-R5	Elezovka	Bosnia and Herzegovina	F10-R5	Tajgë e Zezë	Albania
F4-R6	Kavčina	Serbia	F10-R6	Meresnik	Albania
F4-R7	Marburger	Slovenia	F10-R7	Korith i Bardhë	Albania
F4-R8	Prošip	Bosnia and Herzegovina	F10-R8	Tajgë e Bardhë	Albania
F5-R1	Sipa	Slovenia	F11-R1	Pulëz	Albania
F5-R2	Trnjak	Bosnia and Herzegovina	F11-R2	Razaki e Kuqe	Albania
F5-R3	Cabernet Sauvignon	Slovenia	F11-R3	Serinë e Bardhë	Albania
F5-R4	Blatinal	Bosnia and Herzegovina	F11-R4	Debinë e Bardhë	Albania
F5-R5	Furmint	Serbia	F11-R5	Furmint	France
F5-R6	Menigovka	Bosnia and Herzegovina	F11-R6	Chardonnay	France
F5-R7	Radovača VII	Montenegro	F11-R7	Pinot Noir	France
F5-R8	Sipon	Slovenia	F11-R8	Mourvedre	France
F6-R1	Sultanine	Slovenia	F12-R1	Ugni B/Trebbianotoscano	France
F6-R2	CrnValandovskiDrenok	North Macedonia	F12-R2	Muscat a petit grains	France
F6-R3	Gavran	Serbia	F12-R3	Muscat d'Alexandrie	France
F6-R4	Končanka	North Macedonia	F12-R4	Merlot	France
F6-R5	Muskat Ruža	Serbia	F12-R5	Cabernet Sauvignon	France
F6-R6	Slankamenka Crvena	Serbia	F12-R6	Cabernet franc	France
F6-R7	Touriga Nacional	Slovenia	F12-R7	Sultanine	France
F6-R8	Čauš Bel	North Macedonia	F12-R8	Richter110	France

* Bolded cultivars are, in general, considered as references.

2.2. DNA Extraction

The grapevine samples were obtained from different countries (Table 1), and DNA was extracted from fresh young leaves at the Biotechnical Faculty, University of Ljubljana, Slovenia. For this purpose, the modified cetyl trimethylammonium bromide (CTAB) method [28] was used. After measuring the concentrations (Amersham Biosciences DyNAQuant 200), the DNA samples were stored in a TE Buffer (Invitrogen™, Carlsbad, CA, USA) at -20°C .

Table 2. The reference sequence, microsatellite core repeat, and reference length of each locus.

[illegible]

Table 2. Cont.

Locus	Reference Sequence	Microsatellite Core Repeat	Reference Length
VVMD32-NGS	TGAAACGTCTCGCCAT TACCCCTCCCTCTCTCTC TCTCTCTCTCTCTCTCTC TCTCTCTCTCTCTCTCTC TCAAGCCAGGCGTCAAAAC ATGAACTGTTTGTC CCTCAAGCAGCGTATCCATA GCGAGTGGAGGAGAGAGAG AGAGAGAGAGAGAGAGAGA GAGTGAAGCGCCAAAGAGAG AGGGAGGAGGG	(CT) _n	109
VVMD7-NGS		(CT) _n	88

2.3. PCR Amplification

The grapevine cultivars were genotyped using SSR amplicon sequencing. The primers were redesigned to the length of the amplicon for short sequencing (~100 bp) (Table 2) and amplified according to the established protocol. For this purpose, two rounds of PCR amplification were performed (Figure 1) according to the protocols of Gohl et al. [29] and Vartia et al. [30]. The modified protocol consisted of amplification with locus-specific primers (forward and reverse) adapted to contain a universal primer sequence (Figure 1; Table 3), and the incorporation of two barcodes by two barcoded universal primers into both ends of the resulting amplicons. A total of 12 forward and 8 reverse DNA barcodes enabled the recovery of 96 unique individuals (Supplementary Material, Figure S1).

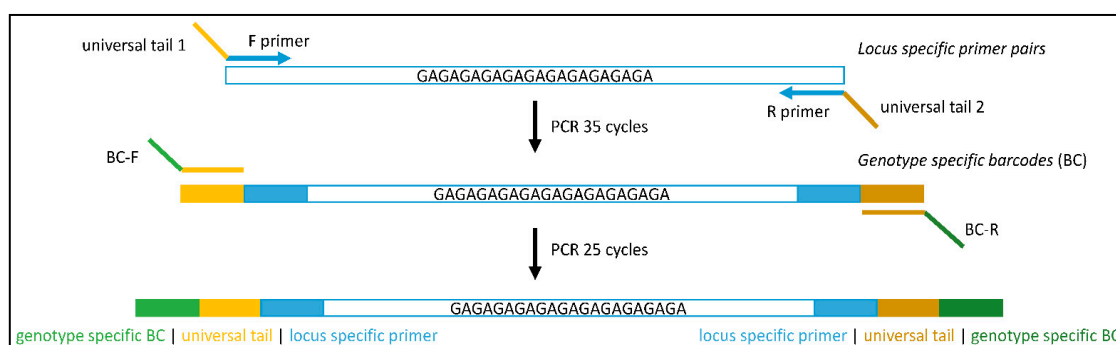


Figure 1. Workflow to amplify short sequence repeats in high-throughput sequencing (HTS) analysis. Amplifying begins with locus-specific amplification (step 1) using locus-specific forward (F) and reverse (R) primers extended with universal tails (Table 2); tail 1 (for F primer) is AATTAACCCCT, tail 2 (for R primer) is CAGTCGGGCG. In step 2, the loci are pooled by sample and re-amplified to integrate the barcoding primers (BC-F, BC-R) listed in Supplementary Material, Figure S1.

2.3.1. PCR for Locus-Specific Amplification

Primary PCR amplification was performed in a final volume of 10 µL containing 5 µL of 5X Q5 Hot Start HiFi buffer, 0.3 µL of 10 mM dNTPs, 5 µL of Q5 Enhancer, 0.1 µL of Q5 Hot Start HiFi Polymerase, 0.25 µL (10 µM) of each locus-specific primer (forward and reverse), and 20 ng of DNA. The cycling conditions were as follows: initial denaturation at 95 °C for 5 min, followed by 35 cycles of 98 °C for 10 s, 65 °C for 20 s, and 72 °C for 10 s. A final extension was performed at 72 °C for 2 min, and then the reaction was cooled down to 4 °C.

Table 3. Table of the simple sequence repeat (SSR) locus-specific primers with universal tail (letters in bold), linkage group, and reference.

SSR Name	Linkage	Locus Specific Forward Primer with Universal Tail	Locus Specific Reverse Primer with Universal Tail	Reference
VMC1b11-NGS	8	AATTAACCTT GACCTAAGTT TCTGAGGCTTTGA	CAGTCGGGCG TGCCTCTCAT TTTGTGTTGC	BV681754
VrZAG79-NGS	5	AATTAACCTT TTAGCCGAAG CCATCTCTGT	CAGTCGGGCG TCAAGAAACA CAAGCACCTCA	[31]
VVIb01-NGS	2	AATTAACCTT CCTGTGAAAC CACCCTATCC	CAGTCGGGCG TGGTTGCACA AGGTAAACGA	[32]
VVIn73-NGS	17	AATTAACCTT AGGCTTCAAA GCCCTCTCAT	CAGTCGGGCG ACACCCTTGT GGGTGTCAGT	[32]
VVIp31-NGS	19	AATTAACCTT TTGGGAAACC ACAGAAGTGA	CAGTCGGGCG AATGGGTTGT TCCTGGAATG	[32]
VVIq52-NGS	9	AATTAACCTT CAGGAAAGTG TTCAATGGTTAC	CAGTCGGGCG AAGGATGGTA GATGACAGAACCA	[32]
VVIv37-NGS	10	AATTAACCTT ACCAGTATTA AGAACGCAGTCAC	CAGTCGGGCG CCCTACTCTT AACTTCCCACTCA	[32]
VVMD24-NGS	14	AATTAACCTT AGAAGACTTG TCTCTCTCAA TCAA	CAGTCGGGCG CGCAAAGAAA TTGAGACAAGG	[33]
VVMD25-NGS	11	AATTAACCTT TGAAAAGTGT AGTGACCCTTTGA	CAGTCGGGCG TGGAGGCTTG TCTTGTGGTT	[33]
VVMD27-NGS	5	AATTAACCTT CCTCTCTCTC CGGCGGTA	CAGTCGGGCG ACCAGATCTG AATACATCCGTAA	[33]
VVMD32-NGS	4	AATTAACCTT TGAAACGTCT CGCCATTACC	CAGTCGGGCGG ACAAACAGTT CATGTTTTGACG	[33]
VVMD7-NGS	7	AATTAACCTT CCTCAAGCAG CGTATCCATAG	CAGTCGGGCG CCCTCCTCCC TCTCTCTTG	[33]

2.3.2. PCR for Barcode Integration

We performed the second dual barcoding PCR in a volume of 10 μ L containing 5 μ L of primary PCR at a 1:100 dilution, 3 μ L of 5 μ M oligo for each index/barcode, 1.5 μ L of 10x KAPA HiFi buffer, 0.3 μ L of 10 mM dNTPs, and 0.08 μ L of KAPA HiFi Polymerase. The following cycling conditions allowed the efficient incorporation of barcodes to PCR amplicons: initial denaturation at 95 °C for 5 min, followed by 25 cycles of 98 °C for 30 s, 45 °C for 30 s, and 72 °C for 1 min. A final extension was performed at 72 °C for 8 min, and the reaction was cooled down to 4 °C.

2.4. Pooling and Sequencing

After the second dual indexing PCR, the amplification products were checked using agarose gel electrophoresis across all loci and diluted appropriately to minimise the amplification rate differences between samples. Two microliters of each PCR product (across all loci and all specimens) were pooled together and cleaned using the Illustra GFX PCR and a gel band purification kit (GE Healthcare, Chicago, IL, USA), following the recommended procedures to remove shorter oligonucleotides. The cleaned sample was eluted in 25 μ L, analysed with a highly accurate DNA electrophoresis Bioanalyzer

2100 system using a DNA 1000 kit (Agilent, Santa Clara, CA, USA), diluted to the final concentration of 20 ng/μL, and submitted for the Illumina 150 bp paired-end sequencing at GATC Biotech (Ebersberg, Germany). The project was designed to obtain approximately 5 M paired-end reads per DNA library. The reads were delivered as two FASTQ non-interleaved files.

2.5. Bioinformatics Analysis

Reference loci sequences were acquired through the Grape genome browser (12X coverage) (<http://www.cns.fr/externe/GenomeBrowser/Vitis/>) and adapted to shorter lengths (Table 2). The raw sequencing reads were mapped to the reference sequences using the “Map Reads to reference” tool implemented in CLC Genomics Workbench 20 (Version 20.0.3) (Qiagen, Hilden, Germany) to obtain the sequencing statistics per locus.

We used two different approaches to assign amplicon sequences to each cultivar and locus. The first approach consisted of mapping the raw sequencing data against the Pinot Noir genomic reference sequences. In the second approach, we demultiplexed the sequencing data by the cultivar- and locus-specific sequences present in the amplicon sequences. Briefly, the pair-end sequencing data were demultiplexed in two steps using the fastq-multx tool [34]. In the first step, the sequencing reads were demultiplexed based on the cultivar-specific barcodes introduced into amplicons in the second PCR step and, thus, sorted into the corresponding cultivar samples. After this, Cutadapt ver. 1.18 [35] was used to trim the cultivar-specific barcode sequences from the 3′ and 5′ ends of the reads.

In the second step, demultiplexing based on primer sequences, which are considered as locus-specific barcode sequences, was performed for each cultivar, and reads with locus-specific sequences on both ends of the reads were kept, thus retaining only full-length sequences. With this procedure, we filtered out incomplete amplicons and kept the reads that fully defined the microsatellite region. The filtered FASTQ files were converted to FASTA files and analysed using (1) the MISA Perl script [36] for the presence of perfect as well as compound microsatellites and (2) the Infoseq tool [37] to obtain the number of sequences with the same length.

The results were analysed with bash tools using the following procedure. The sizes of the microsatellites (no. of repeats or length of alleles) were reported for each read or amplicon sequence, and the number of unique values (sizes) were reported in a table-wise manner. The number of sequencing reads with obtained SSR sizes (MISA output) and the number of sequencing reads with obtained lengths (Infoseq output) were further used as an input for SONiCS [38], a tool that enables stutter noise correction and the determination of true alleles. The tool was run in Monte Carlo mode, with 5000 simulation repetitions. Analyses with SONiCS were applied for only a subset of data (12 French cultivars), for which we were able to make a comparison on the previously reported CE data [27].

3. Results and Discussion

3.1. Sequencing Analysis

The Illumina paired-end sequencing yielded 24,360,664 reads with an average size of 151 nt, yielding a total of 3,678,460,264 (3.68 Gb) bp of data. Theoretically, the even distribution over 12 loci should be approximately 306.5 Mb. The mapping of the reads to the reference alleles (Table 4) showed that the majority of the reads were of high quality, as 22 M of reads (90.7%) were assigned to 12 loci. However, the distribution of the reads across the loci was not uniform, with an acceptable range between 0.79 M for locus VVIq52 and 3.6 M for locus VMC1b11. This is most likely the consequence of competition among loci in the PCR during the library preparation.

Table 4. Sequencing statistics for 96 grapevine cultivars over 12 loci.

Locus	Reference Allele Length	Mapped Reads ¹	Amount of Data [bp] after Mapping ²	Average Coverage after Mapping	No. of Sequences after Filtering ³	Amount of Data (bp) after Demultiplexing	Average Coverage after Filtering
VMC1b11	106	3,649,804	551,120,404	2,888,215	2,927,098	305,691,052	2,883,878
VrZAG79	102	1,192,832	180,117,632	1,148,842	1,694,202	156,444,090	1,533,766
VVIb01	87	2,713,933	409,803,883	2,712,447	2,444,143	223,916,517	2,573,753
VVIn73	86	1,085,003	163,835,453	1,084,374	957,723	81,434,445	946,912
VVIp31	101	1,963,722	296,522,022	1,819,017	1,447,397	149,040,267	1,475,646
VVIq52	79	792,078	119,603,778	791,626	704,921	56,007,493	708,956
VVIv37	101	1,119,484	169,042,084	1,022,452	1,264,019	115,785,796	1,146,394
VVMD24	105	1,813,906	273,899,806	1,810,404	1,588,685	162,085,558	1,543,672
VVMD25	107	2,348,776	354,665,176	2,192,870	2,000,291	221,586,577	2,070,903
VVMD27	112	2,911,241	439,597,391	2,895,057	2,468,380	277,195,183	2,474,957
VVMD32	109	1,556,828	235,081,028	1,163,432	1,037,861	92,983,922	853,064
VVMD7	88	949,509	143,375,859	937,902	899,408	75,335,819	856,089
Total		22,097,116	3,336,664,516		19,434,128	1,917,506,719	

¹ Raw reads mapped to reference alleles using CLC genomics Workbench/Server. ² Total number reported with unaligned part included. ³ Full length sequences starting and ending with amplification primer.

The approach of using reference microsatellite sequences and further demultiplexing sequences based on mapping results did not prove to be the method of choice in our example. Microsatellite repeats can be similar between loci, which leads to incorrect mapping, especially for long alleles. Therefore, we chose a demultiplexing approach based on filtering out those sequences that contained correct locus-specific primer-to-primer information and were considered for the final genotyping. The final number of obtained reads was slightly lower than the number of mapped reads (19.4 M, 79.8%); however, they represented high-quality data that were confirmed twice by sequencing (the paired-end approach). Similarly, the demultiplexing approach yielded from 0.7 M (VVIq52) to 2.9 M (VMC1b11) full-length amplicons per locus (Table 4). Using the mapping approach, we obtained a slightly higher number of sequences for most loci; this was likely mainly due to the inclusion of sequences that did not cover the entire microsatellite sequences.

The minimum length of the amplicons demultiplexed by the locus ranged from 73 nt (VVIq 52 and VVIv37) to 99 nt (VVMD25), and the maximum length ranged from 85 nt (VVIq52) to 131 nt (VVMD25), corresponding to the allele lengths shown in the Supplementary Material, Figure S2.

3.2. Comparison of CE and HTS Approaches

The results of the comparisons between the HTS and CE methods for microsatellite analyses are presented in the Supplementary Material, Table S1. In examining the HTS approach, the sequences were analysed according to the number of microsatellite repeats (MISA script) and the full lengths of the sequenced amplicons (Infoseq script). The SSR lengths obtained by the MISA script and the amplicon lengths obtained by Infoseq were first analysed with SONiCS. During the visual inspection of the results, we found some allele calling errors when using automated SONiCS analyses, and thus we concluded that the approach using solely SONiCS was not appropriate for the determination of true alleles.

In the past, some other bioinformatics tools have been developed for retrieving SSRs from HTS data, such as LobSTR [39], RepeatSeq [40], STRViper [41], STR-FM [42], PSR [43], rAmpSeq [44], and STRScan [45]. We decided to use the software SONiCS, as it performs simulations of PCR reactions to correct allele calling due to the stutter bands, which are amplified at most grapevine SSR loci used in this study. SONiCS uses the length and depth of the sequenced alleles as input data, and, after each simulation, the log likelihood is calculated to estimate the probability of generating the observed data (input data) from the assumed simulated results. SONiCS selects the alleles for which the model has the highest likelihood. In 144 comparisons (12 loci × 12 cultivars) between MISA or Infoseq and the CE approach, SONiCS showed a 58% success rate in genotyping using MISA data, as 75 alleles were correctly called and 8 alleles differed only by 1 bp. When calling genotypes based on sequence length

(Infoseq), SONiCS performed better compared to the approach using the MISA data, as it showed a 77% success rate in genotyping, as 102 alleles were correctly called and 9 alleles differed only by 1 bp.

However, due to missing some longer alleles with lower read counts, we continued to call alleles from the Infoseq output data by visual determination. The CE approach served as a standard. The comparison of the differences for the two alleles (per locus per sample) revealed some discrepancies between the HTS and CE methods, as shown in the Supplementary Material, Table S1. When comparing the MISA data with the CE data for 144 data points (12 loci \times 12 cultivars), we obtained 75 alleles that showed the same difference between the alleles within the locus and 8 that differed only by 1 bp. Comparing the Infoseq data with the CE data for 144 data points, we obtained 102 alleles that showed the same difference between the alleles within the locus and 9 that differed only by 1 bp. The reported differences could be due to the development of new primers for HTS analyses that could lead to new null alleles, so that, in some cases, the homozygosity was higher than the expected heterozygosity for the HTS approach (Richter110, locus VVMD25), and, conversely, in some cases the homozygosity was higher than the expected heterozygosity for the CE approach (e.g., Merlot, locus VrZAG79).

The clustering of cultivars based on simple-matching dissimilarity coefficients was performed for the CE and HTS allelic data and resulted in two trees (Figure 2), with bipartition complexities of 0.94 and 0.91. The value for the consensus tree was 0.52, and the obtained distance between the trees was 0.82. Certain clusters supported with high bootstrapping values (e.g., a cluster of Muscat cultivars and cluster of Pinot Noir–Chardonnay) appeared equally in both approaches, and the Richter 110 rootstock was the most different from other *V. vinifera* cultivars in both approaches (Figure 2).

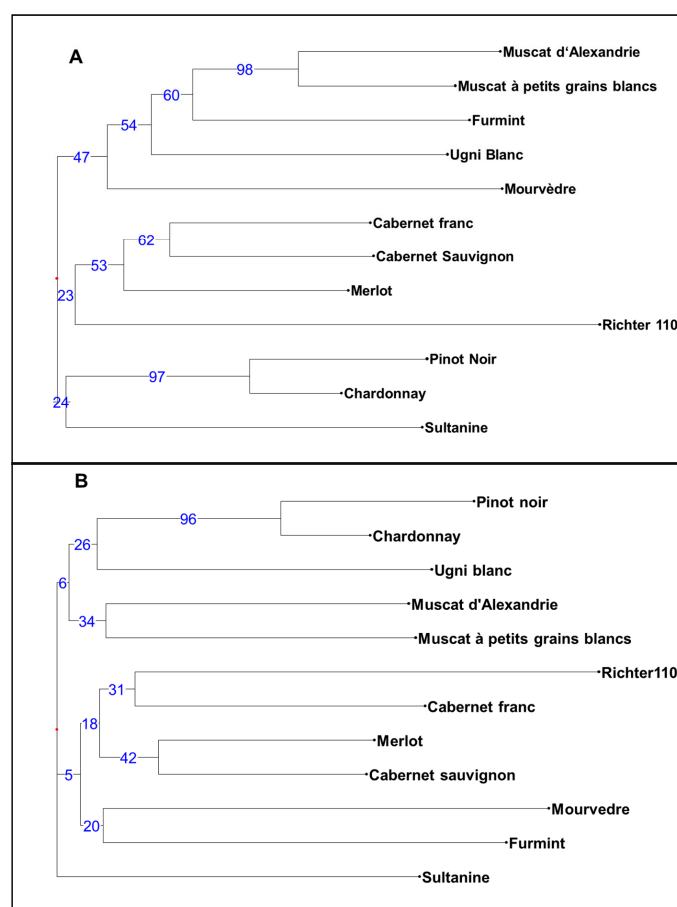


Figure 2. Tree construction based on simple-matching dissimilarity coefficient and the weighted neighbour-joining clustering method using alleles (A) obtained by capillary electrophoresis (CE) analysis (B) and by HTS (Infoseq) analysis. The numbers on the branches indicate the percentage of bootstrap analysis (1000).

3.3. The HTS Approach Creates a Bias in Calling True Alleles for Some Loci

The number of read counts of full-length sequences (alleles) for 12 cultivars over 12 loci are presented as histograms (Supplementary material, Figure S2), with the corresponding alleles determined (Supplementary Material, Table S1; columns K and L). We observed that some loci are more problematic for the HTS approach than others; e.g., for the loci VVIq52, VVIb01, and VVMD24, we did not observe any discrepancies in the intra-allelic length comparison between different approaches (Supplementary Material, Table S1), whereas for locus VVMD27, for example, 6 out of 12 comparisons resulted in inconsistencies (Supplementary Material, Table S1). In locus VVIb01, the alleles were short (from 87 to 97 bp), and were similarly so in locus VVIq52 (from 75 to 83 bp) and VVMD 24 (from 97 to 108 bp), while in locus VVMD27 the allele lengths were from 110 to 125 bp and certain long-sized alleles could be overlooked due to their poor sequence coverage (Figure 3, Furmint, allele 125 bp). A similar problem was observed for the locus VVMD25 (Figure 3, Mourvedre, allele 131 bp).

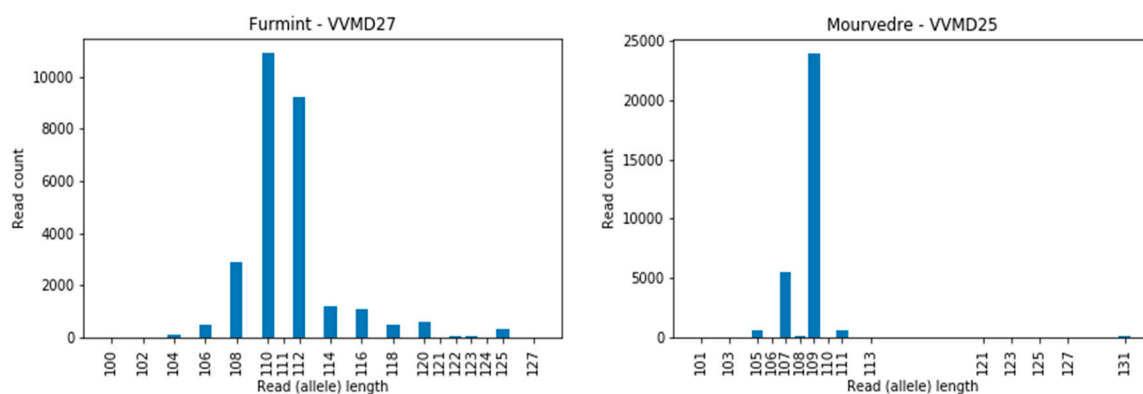


Figure 3. Example of the low sequence coverage for long alleles in the cultivar Furmint at locus VVMD27, allele 125 bp, and in cultivar Mourvedre at locus VVMD25, allele 131 bp.

In locus VrZag79, in many cases (for cultivars Muscat Blanc a Petits Grains, Muscat d’Alexandrie, Mourvedre, Furmint, Cabernet franc, etc.) a three-allelic profile or high debris (reads of 83 and 89 bp) appeared. Figure 4 shows the Mourvedre cultivar for locus VrZag79 with a tri-allelic profile (83, 89, and 97 bp). The three-allelic profiles discovered for this locus were previously observed in studies when extracting DNA from leaves. The presence of a third allele in leaf tissue indicates a periclinal chimera [46].

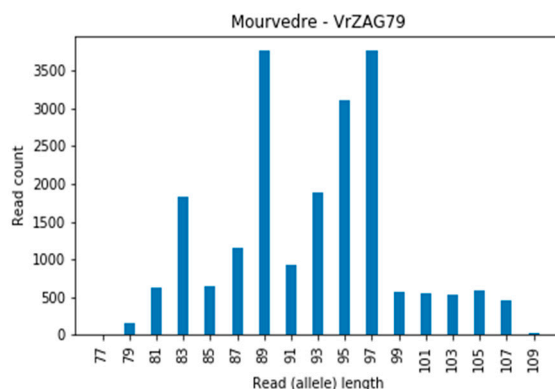


Figure 4. Example of the triallelic profile of cultivar Mourvedre at locus VrZag79.

The locus VVMD7 showed, in some cases, a very intensive amplification of stuttering bands (Figure 5), which can hinder the calling of true alleles. Small and unexpected mutations associated

with locus VVMD7 were also reported earlier [46–50] and may, in some cases, be a consequence of the impaired allele calling.

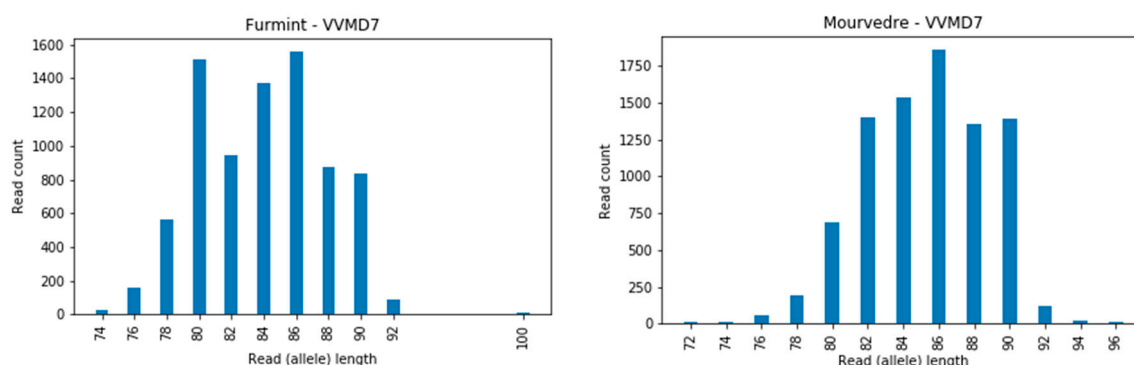


Figure 5. Example of the intense amplification of stutter bands at locus VVMD7 for two cultivars, Furmint and Mourvedre.

3.4. Analyses of 96 *V. vinifera* Samples

The sequencing analyses (i.e., the number of reads for the sequenced amplicons) for 96 different *V. vinifera* cultivars over 12 loci are presented in the Supplementary Material, Table S2. In the analysed data set, we included five counterparts from French and Slovenian collections (Chardonnay, Merlot, Pinot Noir, Cabernet Sauvignon, and Sultanine), and the comparison over 12 loci yielded 55 exact matches and 5 discrepancies (Supplementary data, Table S2); three out of five were different for only two bp for the compared alleles and two were within the locus VVMD27, which was previously confirmed as one of the loci with triallelic profiles (chimerism) that showed a high intra-clonal variability [51,52]. Discrepancies were found in the Merlot and Pinot Noir cultivars, with previously reported intra-clonal genetic variation [46,51,52]. Studies have previously reported polymorphisms identified by microsatellite markers, which indicate the presence of triallelic loci, referred to in grapevines as chimeras [46,49], caused by mutations in the cells of the meristem layers L1 and L2 [53].

3.5. HTS Genotyping Economy

HTS systems offer extremely cost-effective sequencing generation for large amounts of data. Therefore, HTS systems are already used in genotyping projects that employ different strategies to find polymorphisms, such as genotyping by sequencing [54], capturing strategies [55–57], or the shotgun sequencing of entire genomes [58]. Microsatellites are multiallelic markers, which makes them ideal for the management of plant germplasm. In our project, we investigated the possibility of using a sequence counting approach for genotyping microsatellite alleles.

There are also economic reasons behind switching from capillary-based systems to HTS platforms. The first important reason is the price of a capillary-based instrument, which is higher than for medium-throughput NGS systems. The price of the instrument is worth considering, especially for those laboratories that are considering either replacing their capillary systems or buying new ones. The second reason is the operating costs. The sequencing cost of our project was 531 € (VAT excluded), and we have produced more than 12 million sequences. Our data contained 1152 data points (96 cultivars by 12 loci), which means 0.46 € per data point. However, the sequencing coverage was extremely high (10,000× on average). We believe that we were able to reduce the coverage by at least five times, which is 0.09 € per data point. The running costs for capillary instruments are higher than 1 € per sample (data point), and genotyping providers usually charge 2.5–3 € per sample. Therefore, the economic situation speaks in favour of HTS typing.

4. Conclusions

The remarkable advances in high-throughput sequencing technologies have significantly increased their application in genetic diversity studies, population structure analyses, and conservation genetics. The HTS approach has the advantage of the large-scale genotyping of individuals at multiple loci simultaneously using an amplicon barcoding system that allows large-scale analysis, generating a large amount of data in less time and at a surprisingly lower cost [59,60]. The HTS approach showed significant advantages over the fragment length variation-based approach using conventional capillary and gel electrophoresis [21,30,59,61]. Studies [21,59] reported that HTS technology increased the number of detected alleles compared to the electrophoresis-based method, overcoming the effect of microsatellite length homoplasy, resolving the hidden variations, and maximizing the genetic information obtained. While homoplasy was reported in certain previous studies, it was not detected in any of the loci we investigated. Homoplasy is more likely to be detected in less closely related genotypes.

According to our observations, the limitation of HTS-SSR genotyping is in the automation of allele retrieval, which is crucial for HTS approaches with high multiplexing and large amounts of data. Due to the high degree of mismatching observed for some microsatellite loci when using SONiCS bioinformatics tools for retrieving SSRs from HTS data, we recommend that other tools should be investigated and/or improvements made to the existing tool (e.g., the normalization of the read counts according to the amplicon length and sequencing depth of the libraries) to reduce the distortion obtained from the amplification and sequencing process.

The HTS-SSR approach has huge potential in terms of its speed and cost effectiveness. As our study is one of the first studies of this kind presented for plants, an additional optimization and validation process should be performed before the routine use of HTS genotyping instead of the CE approach, especially as we have shown that not all loci are equally suitable for the sequencing approach.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/11/8/917/s1>: Figure S1: (a) Forward and reverse barcodes with universal tail (in bold); (b) Barcoding system in a 96-well plate; 12 forward barcodes and 8 reverse barcodes enabling the barcoding of 96 samples. Figure S2: Histograms generated from the number of read counts of full-length sequences (alleles) obtained with the Infoseq approach for twelve *V. vinifera* cultivars at twelve different loci. Table S1: Comparison of three different approaches to determine the genotypes of 12 different *V. vinifera* cultivars at 12 different loci, i.e., by capillary electrophoresis (CE), by the calling length of SSR (MISA), and by the calling allele lengths (Infoseq). The genotype data obtained by capillary electrophoresis are publicly available [27]; SONiCS was used to call genotypes from the data obtained by MISA, and visual determination of genotypes was done to call the alleles obtained by Infoseq. Table S2: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VMC1b11 obtained using the HTS approach. Table S2.1: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VrZAG79 obtained using the HTS approach. Table S2.2: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VVIb01 obtained using the HTS approach. Table S2.3: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VVIn73 obtained using the HTS approach. Table S2.4: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VVIp31 obtained using the HTS approach. Table S2.5: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VVIq52 obtained using the HTS approach. Table S2.6: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VVIv37 obtained using the HTS approach. Table S2.7: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VVMD24 obtained using the HTS approach. Table S2.8: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VVMD25 obtained using the HTS approach. Table S2.9: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VVMD27 obtained using the HTS approach. Table S2.10: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VVMD32 obtained using the HTS approach. Table S2.11: The No. of reads for the sequenced amplicons of 96 different *V. vinifera* cultivars of locus VVMD7 obtained using the HTS approach.

Author Contributions: Conceptualization, J.J. and N.Š.; Data curation, U.K.; Formal analysis, U.K., A.D., V.L., J.J., and N.Š.; Investigation, N.Š.; Methodology, U.K. and N.Š.; Resources, V.L.; Software, U.K.; Supervision, N.Š.; Validation, V.L. and J.J.; Visualization, U.K.; Writing—original draft, U.K., A.D., J.J., and N.Š.; Writing—review and editing, U.K., A.D., J.J., and N.Š. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Slovenian Research Agency (ARRS), grant number P4-0077, research programme: Genetics and Modern Technologies of Crops. The APC was funded by sources of the same grant.

Acknowledgments: We acknowledge the help of Tjaša Cesar regarding DNA isolation and barcoding PCR management.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Southern, E.M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *JMB* **1975**, *98*, 503–517. [\[CrossRef\]](#)
2. Saiki, R.K.; Scharf, S.; Faloona, F.; Mullis, K.B.; Horn, G.T.; Erlich, H.A.; Arnheim, N. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **1985**, *230*, 1350–1354. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Williams, J.G.; Kubelik, A.R.; Livak, K.J.; Rafalski, J.A.; Tingey, S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **1990**, *18*, 6531–6535. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Semagn, K.; Babu, R.; Hearne, S.; Olsen, M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Mol. Breed.* **2014**, *33*, 1–14. [\[CrossRef\]](#)
5. Ganai, M.W.; Polley, A.; Graner, E.M.; Plieske, J.; Wieseke, R.; Luerssen, H.; Durstewitz, G. Large SNP arrays for genotyping in crop plants. *J. Biosci.* **2012**, *37*, 821–828. [\[CrossRef\]](#) [\[PubMed\]](#)
6. He, J.; Zhao, X.; Laroche, A.; Lu, Z.X.; Liu, H.; Li, Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* **2014**, *5*, 484. [\[CrossRef\]](#)
7. Deokar, A.A.; Ramsay, L.; Sharpe, A.G.; Diapari, M.; Sindhu, A.; Bett, K.; Warkentin, T.D.; Tar'an, B. Genome wide SNP identification in chickpea for use in development of a high density genetic map and improvement of chickpea reference genome assembly. *BMC Genom.* **2014**, *15*, 708. [\[CrossRef\]](#)
8. Su, Q.; Zhang, X.; Zhang, W.; Zhang, N.; Song, L.; Liu, L.; Xue, X.; Liu, G.; Liu, J.; Meng, D.; et al. QTL detection for kernel size and weight in bread wheat (*Triticum aestivum* L.) using a high-density SNP and SSR-based linkage map. *Front. Plant Sci.* **2018**, *9*, 1484. [\[CrossRef\]](#)
9. Cipriani, G.; Spadotto, A.; Jurman, I.; Di Gaspero, G.; Crespan, M.; Meneghetti, S.; Frare, E.; Vignani, R.; Cresti, M.; Morgante, M.; et al. The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. *Theor. App. Genet.* **2010**, *121*, 1569–1585. [\[CrossRef\]](#)
10. Morgante, M.; Olivieri, A.M. PCR-amplified microsatellites as markers in plant genetics. *Plant J.* **1993**, *3*, 175–182. [\[CrossRef\]](#)
11. Stępień, Ł.; Mohler, V.; Bocianowski, J.; Koczyk, G. Assessing genetic diversity of Polish wheat (*Triticum aestivum*) varieties using microsatellite markers. *Genet. Resour. Crop Evol.* **2007**, *54*, 1499–1506. [\[CrossRef\]](#)
12. Jacob, H.J.; Lindpaintner, K.; Lincoln, S.E.; Kusumi, K.; Bunker, R.K.; Mao, Y.P.; Ganten, D.; Dzau, V.J.; Lander, E.S. Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell* **1991**, *67*, 213–224. [\[CrossRef\]](#)
13. Edwards, A.; Civitello, A.; Hammond, H.A.; Caskey, C.T. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* **1991**, *49*, 746. [\[PubMed\]](#)
14. Parida, S.K.; Dalal, V.; Singh, A.K.; Singh, N.K.; Mohapatra, T. Genic non-coding microsatellites in the rice genome: Characterization, marker design and use in assessing genetic and evolutionary relationships among domesticated groups. *BMC Genom.* **2009**, *10*, 140. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Powell, W.; Machray, G.C.; Provan, J. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* **1996**, *1*, 215–222. [\[CrossRef\]](#)
16. Richard, G.F.; Hennequin, C.; Thierry, A.; Dujon, B. Trinucleotide repeats and other microsatellites in yeasts. *Res. Microbiol.* **1999**, *150*, 589–602. [\[CrossRef\]](#)
17. Richard, G.F.; Pâques, F. Mini- and microsatellite expansions: The recombination connection. *EMBO Rep.* **2000**, *1*, 122–126. [\[CrossRef\]](#)
18. Van Dijk, E.L.; Auger, H.; Jaszczyzsyn, Y.; Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet.* **2014**, *30*, 418–426. [\[CrossRef\]](#)

19. Bornman, D.M.; Hester, M.E.; Schuetter, J.M.; Kasoji, M.D.; Minard-Smith, A.; Barden, C.A.; Nelson, S.C.; Godbold, G.D.; Baker, C.H.; Yang, B.; et al. Short-read, high-throughput sequencing technology for STR genotyping. *BioTech. Rapid Dispatches* **2012**, *2012*, 1–6. [\[CrossRef\]](#)
20. Fordyce, S.L.; Ávila-Arcos, M.C.; Rockenbauer, E.; Børsting, C.; Frank-Hansen, R.; Petersen, F.T.; Willerslev, E.; Hansen, A.J.; Morling, N.; Gilbert, M.T. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. *Biotechniques* **2011**, *51*, 127–133. [\[CrossRef\]](#)
21. Darby, B.J.; Erickson, S.F.; Hervey, S.D.; Ellis-Felege, S.N. Digital fragment analysis of short tandem repeats by high-throughput amplicon sequencing. *Ecol. Evol.* **2016**, *6*, 4502–4512. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Lepais, O.; Chancerel, E.; Boury, C.; Salin, F.; Manicki, A.; Taillebois, L.; Dutech, C.; Aissi, A.; Bacles, C.F.E.; Daverat, F.; et al. Fast sequence-based microsatellite genotyping development workflow. *PeerJ* **2020**, *8*, e9085. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Albarghouthi, M.N.; Buchholz, B.A.; Doherty, E.A.; Bogdan, F.M.; Zhou, H.; Barron, A.E. Impact of polymer hydrophobicity on the properties and performance of DNA sequencing matrices for capillary electrophoresis. *Electrophoresis* **2001**, *22*, 737–747. [\[CrossRef\]](#)
24. Tu, O.; Knott, T.; Marsh, M.; Bechtol, K.; Harris, D.; Barker, D.; Bashkin, J. The influence of fluorescent dye structure on the electrophoretic mobility of end-labeled DNA. *Nucleic Acids Res.* **1998**, *26*, 2797–2802. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Estoup, A.; Jarne, P.; Cornuet, J.M. Homoplasmy and mutation model at microsatellite loci and their consequences for population genetic analysis. *Mol. Ecol.* **2002**, *11*, 1591–1604. [\[CrossRef\]](#)
26. Koressaar, T.; Lepamets, M.; Kaplinski, L.; Raime, K.; Andreson, R.; Remm, M. Primer3_masker: Integrating masking of template sequence with primer design software. *Bioinformatics* **2018**, *34*, 1937–1938. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Laucou, V.; Lacombe, T.; Dechesne, F.; Siret, R.; Bruno, J.P.; Dessup, M.; Dessup, T.; Ortigosa, P.; Parra, P.; Roux, C.; et al. High throughput analysis of grape genetic diversity as a tool for germplasm collection management. *Theor. Appl. Genet.* **2011**, *122*, 1233–1245. [\[CrossRef\]](#)
28. Kump, B.; Javornik, B. Evaluation of genetic variability among common buckwheat (*Fagopyrum esculentum* Moench) populations by RAPD markers. *Plant Sci.* **1996**, *114*, 149–158. [\[CrossRef\]](#)
29. Gohl, D.M.; MacLean, A.; Hauge, A.; Becker, A.; Walek, D.; Beckman, K.B. An optimized protocol for high-throughput amplicon-based microbiome profiling. *Protoc. Exch.* **2016**. [\[CrossRef\]](#)
30. Vartia, S.; Villanueva-Cañas, J.L.; Finarelli, J.; Farrell, E.D.; Collins, P.C.; Graham, H.M.; Carlsson, J.E.L.; Gauthier, D.T.; McGinnity, P.; Cross, T.F.; et al. A novel method of microsatellite genotyping by sequencing using individual combinatorial barcoding. *R. Soc. Open Sci.* **2016**, *3*, 150565. [\[CrossRef\]](#)
31. Sefc, K.M.; Regner, F.; Turetschek, E.; Glossl, J.; Steinkellner, H. Identification of microsatellite sequences in *Vitis riparia* and their applicability for genotyping of different *Vitis* species. *Genome* **1999**, *42*, 367–373. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Merdinoglu, D.; Butterlin, G.; Bevilacqua, L.; Chiquet, V.; Adam-Blondon, A.F.; Decroocq, S. Development of a large set of microsatellite markers in grapevine (*Vitis vinifera* L.) suitable for multiplex PCR. *Mol. Breed.* **2005**, *15*, 349–366. [\[CrossRef\]](#)
33. Browsers, J.E.; Dangl, G.S.; Vignani, R.; Meredith, C.P. Isolation and characterization of new polymorphic simple sequence repeat loci in grape (*Vitis vinifera* L.). *Genome* **1996**, *39*, 628–633. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Aronesty, E. Comparison of sequencing utility programs. *Open Bioinform. J.* **2013**, *7*, 1–8. [\[CrossRef\]](#)
35. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*. [\[CrossRef\]](#)
36. Beier, S.; Thiel, T.; Munch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Rice, P.; Longden, I.; Bleasby, A. EMBL: The European molecular biology open software suite. *Trends Genet.* **2000**, *508*, 276–277. [\[CrossRef\]](#)
38. Kedzierska, K.Z.; Gerber, L.; Cagnazzi, D.; Krutzen, M.; Ratan, A.; Kistler, L. SONiCS: PCR stutter noise correction in genome-scale microsatellites. *Bioinformatics* **2018**, *34*, 4115–4117. [\[CrossRef\]](#)
39. Gymrek, M.; Golan, D.; Rosset, S.; Erlich, Y. lobSTR: A short tandem repeat profiler for personal genomes. *Genome Res.* **2012**, *22*, 1154–1162. [\[CrossRef\]](#)

40. Highnam, G.; Franck, C.; Martin, A.; Stephens, C.; Puthige, A.; Mittelman, D. Accurate human microsatellite genotypes from high-throughput resequencing data using informed error profiles. *Nucleic Acids Res.* **2012**, *41*, e32. [[CrossRef](#)]
41. Cao, M.D.; Tasker, E.; Willadsen, K.; Imelfort, M.; Vishwanathan, S.; Sureshkumar, S.; Balasubramanian, S.; Bodén, M. Inferring short tandem repeat variation from paired-end short reads. *Nucleic Acids Res.* **2013**, *42*, e16. [[CrossRef](#)] [[PubMed](#)]
42. Fungtammasan, A.; Ananda, G.; Hile, S.E.; Su, M.S.-W.; Sun, C.; Harris, R.; Medvedev, P.; Eckert, K.; Makova, K.D. Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res.* **2015**, *25*, 736–749. [[CrossRef](#)] [[PubMed](#)]
43. Cantarella, C.; D'Agostino, N. PSR: Polymorphic SSR retrieval. *BMC Res. Notes* **2015**, *8*, 525. [[CrossRef](#)] [[PubMed](#)]
44. Buckler, E.S.; Ilut, D.C.; Wang, X.; Kretzschmar, T.; Gore, M.A.; Mitchell, S.E. rAmpSeq: Using repetitive sequences for robust genotyping. *BioRxiv* **2016**. [[CrossRef](#)]
45. Tang, H.; Nzabarushimana, E. STRScan: Targeted profiling of short tandem repeats in whole-genome sequencing data. *BMC Bioinforma.* **2017**, *18*, 398. [[CrossRef](#)]
46. Riaz, S.; Garrison, K.E.; Dangl, G.S.; Boursiqot, J.-M.; Meredith, C.P. Genetic divergence and chimerism within ancient asexually propagated winegrape cultivars. *J. Am. Soc. Hortic. Sci.* **2002**, *127*, 508–514. [[CrossRef](#)]
47. Crespan, M. Evidence on the evolution of polymorphism of microsatellite markers in varieties of *Vitis vinifera* L. *Theor. Appl. Genet.* **2004**, *108*, 231–237. [[CrossRef](#)]
48. Ibanez, J.; De Andres, M.T.; Borrego, J. Allelic variation observed at one microsatellite locus between the two synonym grape cultivars Black Currant and Mavri Corinthiaki. *Vitis* **2000**, *39*, 173–174.
49. Hocquigny, S.; Pelsy, F.; Dumas, V.; Kindt, S.; Heloir, M.C.; Merdinoglu, D. Diversification within grapevine cultivars goes through chimeric states. *Genome* **2004**, *47*, 579–589. [[CrossRef](#)]
50. Štajner, N.; Rusjan, D.; Korosec-Koruza, Z.; Javornik, B. Genetic Characterization of Old Slovenian Grapevine Varieties of *Vitis vinifera* L. by Microsatellite Genotyping. *Am. J. Enol. Viticult.* **2011**, *62*, 250–255. [[CrossRef](#)]
51. Koncilja, K. Intravarietal Variability analysis of Grapevine Variety 'Merlot' (*Vitis vinifera* L.) with Microsatellites Markers. Master's Thesis, University of Ljubljana, Ljubljana, Slovenia, 2010.
52. Vélez, M.D.; Ibáñez, J. Assessment of the uniformity and stability of grapevine cultivars using a set of microsatellite markers. *Euphytica* **2012**, *186*, 419–432. [[CrossRef](#)]
53. Thompson, M.M.; Olmo, H.P. Cytohistological studies of cytochimeric and tetraploid grapes. *Am. J. Bot.* **1963**, *50*, 901–906. Available online: <https://www.jstor.org/stable/2439777> (accessed on 9 August 2020). [[CrossRef](#)]
54. Deschamps, S.; Llaca, V.; May, G.D. Genotyping-by-sequencing in plants. *Biology* **2012**, *1*, 460–483. [[CrossRef](#)] [[PubMed](#)]
55. Mertes, F.; Elsharawy, A.; Sauer, S.; van Helvoort, J.M.; van der Zaag, P.J.; Franke, A.; Nilsson, M.; Lehrach, H.; Brookes, A.J. Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief. Funct. Genom.* **2011**, *10*, 374–386. [[CrossRef](#)] [[PubMed](#)]
56. Zhou, L.; Holliday, J.A. Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. *BMC Genom.* **2012**, *13*, 703. [[CrossRef](#)] [[PubMed](#)]
57. Hill, C.B.; Wong, D.; Tibbits, J.; Forrest, K.; Hayden, M.; Zhang, X.Q.; Westcott, S.; Angessa, T.T.; Li, C. Targeted enrichment by solution-based hybrid capture to identify genetic sequence variants in barley. *Sci. Data* **2019**, *6*, 12. [[CrossRef](#)] [[PubMed](#)]
58. Lachagari, V.; Gupta, R.; Lekkala, S.P.; Mahadevan, L.; Kuriakose, B.; Chakravartty, N.; Katta, A.M.; Santhosh, S.; Reddy, A.R.; Thomas, G. Whole genome sequencing and comparative genomic analysis reveal allelic variations unique to a purple colored rice landrace (*Oryza sativa* ssp. *indica* cv. Purpleputtu). *Front. Plant Sci.* **2019**, *10*, 513. [[CrossRef](#)]
59. Šarhanová, P.; Pfanzelt, S.; Brandt, R.; Himmelbach, A.; Blattner, F.R. SSE-R-seq: Genotyping of microsatellites using next-generation sequencing reveals higher level of polymorphism as compared to traditional fragment sizes coring. *Ecol. Evol.* **2018**, *8*, 10817–10833. [[CrossRef](#)]

60. Curto, M.; Winter, S.; Seiter, A.; Schmid, L.; Scheicher, K.; Barthel, L.M.F.; Plass, J.; Meimberg, H. Application of a SSR-GBS marker system on investigation of European Hedgehog species and their hybrid zone dynamics. *Ecol. Evol.* **2019**, *9*, 2814–2832. [[CrossRef](#)]
61. Farrell, E.D.; Carlsson, J.E.L.; Carlsson, J. Next Gen Pop Gen: Implementing a high-throughput approach to population genetics in boarfish (*Capros aper*). *R. Soc. Open Sci.* **2016**, *3*, 160651. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).