



## Evaluation of simple sequence repeats (SSR) and single nucleotide polymorphism (SNP)-based methods in olive varieties from the Northwest of Spain and potential for miniaturization

Joana Carvalho<sup>a,b</sup>, Shambhavi Yadav<sup>c</sup>, Alejandro Garrido-Maestu<sup>a</sup>, Sarah Azinheiro<sup>a,b</sup>, Isabel Trujillo<sup>d</sup>, Jorge Barros-Velázquez<sup>b</sup>, Marta Prado<sup>a,\*</sup>

<sup>a</sup> International Iberian Nanotechnology Laboratory (INL), Food Quality and Safety Research Group, Braga, Portugal

<sup>b</sup> Department of Analytical Chemistry, Nutrition and Food Science, College of Pharmacy/School of Veterinary Sciences, University of Santiago de Compostela, Campus Vida, E-15782 Santiago de Compostela, Spain

<sup>c</sup> Forest Research Institute, Dehradun, India

<sup>d</sup> Agronomy Department of University of Córdoba, International Campus of Excellence on Agrofood (ceiA3), Rabanales Campus, Córdoba, Spain

### ARTICLE INFO

#### Keywords:

Cultivated olive  
Simple sequence repeats  
Single nucleotide polymorphisms  
HRM  
Allele-specific qPCR  
Miniaturization

### ABSTRACT

Miniaturization of DNA-based techniques can bring interesting advantages for food analysis, such as portability of complex analytical procedures. In the olive oil industry, miniaturization can be particularly interesting for authenticity and traceability applications, through *in situ* control of raw materials before production and/or the final products. However, variety identification is challenging, and implementation on miniaturized settings must be carefully evaluated, starting from the selected analytical approach. In this work, SSR- and SNP-based genotyping strategies were investigated for the identification and differentiation of two olive varieties from the Northwest of Spain. For the selected SNPs two genotyping methods were tested: real-time allele-specific PCR and high resolution melting analysis. These methods were compared and evaluated regarding their potential for integration in a microfluidic device. Both SNP-based methods proved to be successful for identification of the selected varieties, however real-time allele-specific PCR was the one that achieved the best results when analyzing mixtures, allowing the identification of both monovarietal samples and mixtures of the varieties tested with up to 25%.

### 1. Introduction

The worldwide recognition of olive oil for its health benefits and nutritional properties, and the high appreciation for its aroma and taste, make it a price premium product in the agro-food industry. Consequently, it is also very prone to adulteration either by blending with cheaper olive oil or other vegetable oils, by using unapproved production processes or by misrepresentation regarding origin, cultivar or category (Yan, Erasmus, Aguilera Toro, Huang, & van Ruth, 2020). Regulations have been implemented to protect consumers and producers against these fraudulent practices and to ensure product traceability, define product origin and proper labelling (European Parliament and European Council, 2012). However, in addition to the richness of its genetic patrimony, the high occurrence of synonyms (different names

given to the same cultivar) and homonyms (same name given to different cultivars) among many olive cultivars around the world makes its classification particularly difficult (Ben Amar, Souabni, Saddoud-Debbabi, & Ali Triki, 2021; Trujillo et al., 2014). DNA-based methods are an attractive alternative for varietal identification and are highly specific and with high sensitivity, and stability.

In the last decades, simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) have become the most popular markers for describing genetic variation of a wide range of organisms. The most common SSR genotyping methods include amplification by polymerase chain reaction (PCR) and amplicon size determination by agarose gel electrophoresis, polyacrylamide gel electrophoresis, or capillary electrophoresis with fluorescent markers (Vietina, Agrimonti, Marmiroli, Bonas, & Marmiroli, 2011). On the other hand, for SNP

\* Corresponding author at: Food Quality and Safety Research Group, International Iberian Nanotechnology Laboratory (INL), Av. Mestre José Veiga s/n, 4715-330 Braga, Portugal.

E-mail address: [marta.prado@inl.int](mailto:marta.prado@inl.int) (M. Prado).

<https://doi.org/10.1016/j.fochms.2021.100038>

Received 22 December 2020; Received in revised form 10 July 2021; Accepted 31 July 2021

Available online 4 August 2021

2666-5662/© 2021 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

genotyping there are several techniques available depending on the number of SNPs and population size to be studied. More recently, high resolution melting (HRM) (Pereira, Gomes, Barrias, Fernandes, & Martins-Lopes, 2018), and sequencing techniques have also been implemented for both SSR (Xanthopoulou et al., 2014) and SNP genotyping (Scheben, Batley, & Edwards, 2017).

Although these marker-assisted techniques are routinely used, they require sophisticated tools and highly trained personnel, hindering their applicability in resource-limited settings. With this in mind, interest in microfluidics for DNA-based analysis has grown exponentially, due to their applicability for miniaturization of analysis. The main advantages of these platforms include their lower sample and reagent consumption, faster analysis, lower cost, portability, and potential for automation. Consequently, a diversity of PCR-based microfluidic devices has been developed (Y. Zhang & Ozdemir, 2009), allowing rapid heating/cooling rates which results in faster amplification with lower power consumption. Integration of such systems with genotyping methods has been explored to achieve shorter analysis times, improved sensitivity, and high-throughput analysis. Some examples include commercial microfluidic arrays (e.g. Fluidigm®) (Von Thaden et al., 2017), microfluidic integration of sample preparation steps and/or amplification and detection steps, either based on allele-specific hybridization strategies for SNP genotyping (Zec et al., 2018), or combined with microscale capillary electrophoresis ( $\mu$ CE) for SSR markers (Y. T. Kim et al., 2016). Other strategies include integration of microfluidic DNA amplification with microarrays (Choi et al., 2012), primer extension on microbeads (Chang et al., 2017) or implementation of HRM analysis (Velez et al., 2017) at microscale. Such platforms can have a high impact in the food and agricultural sector for varietal discrimination, plant breeding and food authenticity applications by allowing *in situ* control of the raw materials before production and/or the final products.

In this work, genotyping strategies implementing these popular markers (SSRs and SNPs) have been explored for the identification and differentiation of two varieties from the Northwest of Spain: 'Brava' and 'Mansa de Figueiredo'. These varieties have been explored for the production of high quality extra virgin olive oil (EVOO) (Reboredo-Rodríguez, González-Barreiro, Cancho-Grande, Simal-Gándara, & Trujillo, 2018), with different studies being performed to evaluate their business potential (Zamuz et al., 2020) and their potential nutraceutical properties (Figueiredo-González et al., 2018). This varieties are mainly cultivated in Galicia, in the NW of Spain, where the area under olive cultivation increased from 10 ha in 2007 to 275 ha in 2019 (MAPA, 2019). Those particular varieties, although just a part of this total area, are highly appreciated and efforts towards the identification of other varieties are currently going on (Gago, Santiago, Boso, & Martínez, 2019).

In the current work, SSR genotyping was performed by end-point PCR combined with gel electrophoresis and a commercial chip-based capillary electrophoresis system for amplicon size discrimination and results compared with previous studies (Reboredo-Rodríguez et al., 2018). Regarding SNP-based approaches, to our knowledge no previous studies are available in literature for these Galician varieties. Therefore, SNP identification was performed for the first time for 'Brava' and 'Mansa de Figueiredo' in this work. Selected SNPs were then studied using two genotyping methods: real-time allele-specific PCR and HRM analysis. The effect of different DNA extraction and purification methods was also evaluated in this work. The main focus was the comparison and evaluation of the selected methods regarding their potential for integration in a microfluidic setting, using these two varieties as proof of concept to evaluate the possibilities for miniaturization.

## 2. Materials and methods

### 2.1. Samples

A total of 9 samples of *Olea europaea* were included in this study,

which comprised 7 leaf samples and 2 samples from fruit. Single leaves and fruits were used as the starting plant material for this work. Leaves were collected from different trees belonging to three varieties: 'Brava' (3 trees), 'Mansa de Figueiredo' (3 trees), and 'Arbequina' (1 tree), including 3 samples from 'Brava', 3 from 'Mansa de Figueiredo' and 1 from 'Arbequina'. Regarding fruits, 1 sample of 'Brava' and 1 sample of 'Mansa de Figueiredo' were included. The plant material was previously located, characterized and authenticated (Reboredo-Rodríguez et al., 2018). Detailed information about the samples analyzed in this work is provided in Table S1 (Supplementary material).

### 2.2. Reagents

All reagents used for buffer preparation, including cetyltrimethyl ammonium bromide (CTAB), sodium chloride (NaCl), (Ethylenedinitrilo)tetraacetic acid (EDTA), Tris-Base, Tris-HCl, polyvinylpyrrolidone (PVP- 40),  $\beta$ - mercaptoethanol, ethanol, and chloroform-isoamyl alcohol (24:1) were purchased from Sigma-Aldrich® (Darmstadt, Germany). Solutions were prepared with DNase, RNase-free deionized (DI) water obtained with Milli-Q® system from Merk KGaA (Darmstadt, Germany).

### 2.3. DNA extraction and purification

Fresh leaves and fruit samples were homogenized by pestle and mortar in the presence of liquid nitrogen. In the case of leaf samples, three different DNA extraction and purification protocols were used for comparison purposes, including two commercial kits (NucleoSpin® Food kit and NucleoSpin® Plant kit) and a precipitation method using CTAB buffer. These methods were performed in triplicate. DNA extraction and purification methods were selected for being previously reported as providing high DNA yield from olive oil and other vegetal samples (Costa, Mafra, & Oliveira, 2012). For fruit samples, DNA extraction was performed using both commercial kits only.

#### 2.3.1. Commercial NucleoSpin® Food kit

NucleoSpin® Food kit (Macherey-Nagel, Düren, Germany) was used according to the supplier specifications with some modifications. The protocol started with the lysis step by mixing 200 mg of homogenized sample with 550  $\mu$ L of buffer CF (preheated at 65 °C) and 10  $\mu$ L of Proteinase K (10 mg mL<sup>-1</sup>), and incubating for 30 min at 65 °C with constant agitation at 1200 rpm. Then, a second incubation with 10  $\mu$ L of RNase A (10 mg mL<sup>-1</sup>) was performed for 30 min at room temperature, followed by centrifugation for 10 min at 11000  $\times$  g. The supernatant was collected and mixed with the same volume of buffer C4 and ethanol. After that, the mixture was loaded to the column and centrifuged for 1 min at 11000  $\times$  g. The column was then washed in three steps: the first using 400  $\mu$ L of buffer CQW and centrifuging for 1 min at 11000  $\times$  g, the second using 700  $\mu$ L of buffer C5 and centrifuging in the previous conditions, and the third using 200  $\mu$ L of buffer C5 followed by centrifugation for 2 min at 11000  $\times$  g. In the elution step, DNA elution was done in two steps, adding 50  $\mu$ L of Buffer CE (preheated at 70 °C) each time.

#### 2.3.2. Commercial NucleoSpin® Plant kit

NucleoSpin® Plant kit (Macherey-Nagel, Düren, Germany) was used following the supplier specifications with few modifications. The protocol started with the lysis step by mixing 100 mg of homogenized sample with 400  $\mu$ L of buffer PL1 and 10  $\mu$ L of RNase A (10 mg mL<sup>-1</sup>), and incubating for 30 min at 65 °C with continuous agitation at 1200 rpm. A second incubation with 10  $\mu$ L of Proteinase K (10 mg mL<sup>-1</sup>) was performed in the same conditions. After lysis, the crude lysate was first centrifuged for 10 min at 11000  $\times$  g, and the precleared supernatant was then filtered using NucleoSpin® Filters by centrifuging for 2 min at 11000  $\times$  g. The filtrate was mixed with 450  $\mu$ L of buffer PC before loading in the column and centrifuging for 1 min at 11000  $\times$  g. The column was then washed in three steps: the first using 400  $\mu$ L of buffer

PW1 and centrifuging for 1 min at  $11000 \times g$ , the second using 700  $\mu\text{L}$  of buffer PW2 and centrifuging in the previous conditions, and the third using 200  $\mu\text{L}$  of buffer PW2 followed by centrifugation for 2 min at  $11000 \times g$ . The elution step was done in two stages, adding 50  $\mu\text{L}$  of Buffer PE (preheated at  $65^\circ\text{C}$ ) and incubating for 5 min at room temperature before centrifuging for 1 min at  $11000 \times g$ .

### 2.3.3. Precipitation method using CTAB buffer

The protocol used for DNA extraction by precipitation with CTAB was adapted from *De la Rosa et al.* (De La Rosa, James, & Tobutt, 2002). A total of 0.8 mL of CTAB buffer pH 8 (2% (w/v) CTAB, 1.4 mM NaCl, 20 mM EDTA, 100 mM Tris, 2% (w/v) PVP 40 and 1% (v/v)  $\beta$ -mercaptoethanol), pre-heated at  $65^\circ\text{C}$ , and 10  $\mu\text{L}$  of RNase A (10  $\text{mg mL}^{-1}$ ) were added to 100 mg of homogenized sample and incubated for 30 min at  $65^\circ\text{C}$  with continuous agitation (1200 rpm). The lysate was centrifuged for 10 min at  $13000 \times g$ , and the supernatant was then mixed by gentle inversion with an equal volume of chloroform-isoamyl alcohol (24:1), followed by a second centrifugation for 10 min at  $13000 \times g$ . The volume of aqueous phase was collected and mixed with 0.75 volumes of cold isopropanol (stored overnight at  $-20^\circ\text{C}$ ). This mixture was left to incubate overnight at  $-20^\circ\text{C}$ . After incubation, a centrifugation for 10 min at  $13000 \times g$  and  $4^\circ\text{C}$  was performed to obtain a pellet, which was washed with 200  $\mu\text{L}$  of ethanol 70% (v/v) for 10 min at  $4^\circ\text{C}$  with continuous agitation (500 rpm). The excess of ethanol was removed, and the pellet was left to dry for 30 min at  $37^\circ\text{C}$ . Finally, the dried pellet was resuspended in 50  $\mu\text{L}$  of TE buffer (10 mM Tris-HCl, 1 mM EDTA).

## 2.4. DNA quantification

After DNA extraction and purification, total DNA quantification and DNA purity based on absorbance ratios A260/A280 and A260/A230 were determined for each sample using the NanoDrop™ 2000c spectrophotometer (Thermo Scientific™, Waltham, MA, United States). PCR products targeting *OEX* and *OEW* genes from the 'Brava', 'Mansa de Figueiredo' and 'Arbequina' samples analyzed in this work were quantified using Invitrogen™ Qubit™ 4 Fluorometer and the Qubit™ dsDNA High Sensitivity (HS) Assay Kit (Thermo Scientific™, Waltham, MA, United States) before sequencing.

## 2.5. SSR analysis

SSR analysis was performed by end-point PCR, followed by amplicon size determination by gel electrophoresis and chip-based capillary electrophoresis.

### 2.5.1. SSR marker selection

A total of 14 SSR markers were analyzed, including UDO99-011, UDO99-019, UDO99-024, UDO99-043, *ssrOeUA-DCA3*, *ssrOeUA-DCA9*, *ssrOeUA-DCA11*, *ssrOeUA-DCA15*, *ssrOeUA-DCA16*, *ssrOeUA-DCA18*, GAPU59, GAPU71B, GAPU101, and GAPU103A. These markers were adapted from previously published works, and they have been described as very efficient for olive cultivar identification studies (Baldoni et al., 2009; Reboredo-Rodríguez et al., 2018; Trujillo et al., 2014). The list of primers used for SSR analysis is described in Table S2 (Supplementary material).

### 2.5.2. End-point PCR

Amplification was carried out in Veriti 96-well Thermal Cycler (Applied Biosystems™, Foster City, CA, USA) and the thermal profile included an initial denaturation step at  $95^\circ\text{C}$  for 5 min, followed by 35 cycles of dissociation at  $95^\circ\text{C}$  for 20 s, annealing for 30 s and extension at  $72^\circ\text{C}$  for 30 s. Final extension step was performed at  $72^\circ\text{C}$  for 8 min. Annealing temperature and primer concentration were optimized for each set of primers, being the optimal conditions described in Table S2 (Supplementary material). The final reaction volume was 20  $\mu\text{L}$ , with 2  $\mu\text{L}$  of template DNA, 1x PCR buffer (Biotools B&M Labs S.A., Madrid,

Spain), 1.5 mM  $\text{MgCl}_2$  (Biotools B&M Labs S.A., Madrid, Spain), 200  $\mu\text{M}$  dNTPs (Thermo Fisher Scientific Inc., Waltham, MA, USA), and 0.025 U  $\mu\text{L}^{-1}$  Taq polymerase (Biotools B&M Labs S.A., Madrid, Spain). The remaining volume was completed with DNase, RNase free water.

### 2.5.3. Evaluation of SSR markers

Two methods were used for fragment separation and amplicon size measurement. The first one was agarose gel electrophoresis, which was used to easily confirm the success of the amplification reaction and to determine the fragment sizes, in particular for the SSR markers with larger base pair differences. Gel electrophoresis was performed using 2% (w/v) agarose gel, prepared with 100 mL of sodium borate (SB) buffer and 5  $\mu\text{L}$  of GreenSafe Premium dye for staining (NZYtech Lda, Lisbon, Portugal). Gel loading was done using 6x NZYDNA loading dye (NZYtech Lda, Lisbon, Portugal) mixed with the samples in a total volume of 6  $\mu\text{L}$ . The ladder used for amplicon size reference was NZYDNA Ladder VI (NZYtech Lda, Lisbon, Portugal). The gel was run in SB buffer at 125 V for 1 h30 using Midi 13 horizontal gel electrophoresis system (VWR International LLC, Lutterworth, UK), and gel imaging was obtained using Gel Doc EZ System (Bio-Rad Laboratories Inc., Hercules, CA, USA). The second method for amplicon size determination was a chip-based capillary electrophoresis, performed with the 2100 Bioanalyzer system (Agilent technologies Inc., Santa Clara, CA, USA) and the Agilent DNA 1000 Kit (Agilent technologies Inc., Santa Clara, CA, USA) according to manufacturer's instructions. Fragment sizes were determined using the 2100 Expert Software (Agilent technologies Inc., Santa Clara, CA, USA). Results of SSR analysis with both methods were then compared with previous studies (Reboredo-Rodríguez et al., 2018) for the genotypic characterization of 'Brava' and 'Mansa de Figueiredo', which used an automatic capillary sequencer and fluorescently-labelled primers. This method was performed for the pure varieties tested using DNA extracts obtained from leaves with the commercial NucleoSpin® Plant kit. For comparison purposes, some of the markers were also tested using DNA extracts obtained from leaves with the commercial NucleoSpin® Food kit and the precipitation method with CTAB.

## 2.6. SNP-based analysis

For SNP-based analysis, no previous studies were found in literature for 'Brava' and 'Mansa de Figueiredo' varieties. Therefore, DNA sequencing was performed targeting two genes for SNP identification in these varieties. Selected SNPs were then analyzed by two approaches: one based on real-time allele-specific PCR and the other based on HRM analysis.

### 2.6.1. Target genes and PCR amplification

Genes *OEX* (cycloartenol synthase) and *OEW* (lupeol synthase), which have been previously used for olive cultivar discrimination (Bazakos et al., 2012; Hakim, Kammoun, Makhouloufi, & Rebaï, 2010), were targeted for SNP identification in 'Mansa de Figueiredo', 'Brava' and 'Arbequina' varieties. Two different sets of primers were selected from literature for amplification of partial genomic fragments of each target gene, as detailed in Table S3 (Supplementary material). In this step, DNA extracts obtained from leaves using NucleoSpin® Food kit were amplified by PCR. Amplification was carried out in Veriti 96-well Thermal Cycler (Applied Biosystems™, Foster City, CA, USA) with the following thermal profile: initial denaturation step at  $94^\circ\text{C}$  for 2 min, followed by 40 cycles of dissociation at  $94^\circ\text{C}$  for 15 s, annealing at  $60^\circ\text{C}$  for 15 s and extension at  $68^\circ\text{C}$  for 15 s. Each reaction was performed in a final volume of 20  $\mu\text{L}$ , with 2  $\mu\text{L}$  of template and 10  $\mu\text{L}$  of 2x Invitrogen Platinum II Taq Hot-Start PCR Master Mix (Thermo Fisher Scientific Inc., Waltham, MA, USA) and DNase, RNase free water. The reaction was optimized attending to primer concentration, being the optimal conditions for each set of primers described in Table 1. DNA quantification and PCR product purification using NucleoSpin® Gel and PCR Clean-Up kit (Macherey-Nagel, Düren, Germany) were performed before

**Table 1**

List of olive varieties included in the alignments for SNP discovery, including country of origin and respective GenBank accession numbers.

Target Gene	Olive Varieties for sequence alignment	Country	GenBank Accession number
OEW (Lupeol Synthase)	Brava*	Spain	MW033194
	Mansa de Figueiredo*	Spain	MW033195
	Arbequina*	Spain	MW033193
	Mastoidis	Greece	JN656241.1
	Mignola	Italy	JN656235.1
	Ottobratica	Italy	JN656237.1
	Picual	Spain	AY847066.1
	Caiazzana	Italy	JN656238.1
	Cassanese	Italy	JN656236.1
	Dolce	Italy	JN656239.1
OEX (Cycloartenol Synthase)	Kerkyras	Greece	JN656240.1
	Brava*	Spain	MW033197
	Mansa de Figueiredo*	Spain	MW033198
	Arbequina*	Spain	MW033196
	Cordovil de Serpa	Portugal	AY847065.1

\*Sequences obtained in this work (submitted to GenBank database).

sequencing.

### 2.6.2. DNA sequencing

Purified PCR products were sent to Macrogen (Macrogen Inc., Seoul, South Korea) for Sanger sequencing. Bidirectional sequencing was performed using the respective forward and reverse PCR primers to sequence both strands for each sample, and reduce sequencing errors. In addition, each sample was sequenced at least twice, making it a minimum of 4 sequences per sample for each set of primers. Two different sets of primers were used for each target gene: primers Cyc1 and Cyc2 for cycloartenol synthase, and primers Lup1 and Lup2 for lupeol synthase. Consequently, at least 8 partial sequences were obtained per sample for each target gene, making a total of at least 56 partial sequences per target gene considering all the samples sequenced. In addition to Sanger sequencing, PCR products targeting cycloartenol synthase gene obtained with primers Cyc1 (longest fragment) were also sequenced with MinION (Oxford Nanopore Technologies, Oxford Science Park, UK) using the Rapid Barcoding sequencing kit, obtaining >7000 reads per sample.

### 2.6.3. SNP identification

After Sanger sequencing, sequences were filtered regarding base call quality, and aligned using Geneious Prime software Version 2019.2.3 (Biomatters Ltd., Auckland, New Zealand) to generate the consensus partial sequence of each target gene for each sample. Regarding MinION sequencing, the vast number of reads obtained for each sample was first filtered by size, according to the expected size of the PCR product obtained. The filtered reads were then aligned to obtain the consensus sequences of each sample by performing the Multiple Alignment using Fast Fourier Transform (MAFFT) in Geneious Prime. Consensus sequences obtained for 'Brava', 'Mansa de Figueiredo' and 'Arbequina' samples with both sequencing methods were compared. The consensus sequences obtained for each sample were then aligned with sequences of other olive varieties available in GenBank database (National of library Medicine, 2006) for the same target genes. For OEW gene (lupeol synthase), there were 8 sequences of olive varieties available in GenBank, including 5 varieties from Italy ('Caiazzana', 'Cassanese', 'Dolce', 'Mignola', and 'Ottobratica'), 2 from Greece ('Kerkyras', 'Mastoidis'), and 1 from Spain ('Picual'). For OEX gene (cycloartenol synthase), only 1 sequence was available, namely an olive variety from Portugal ('Cordovil de Serpa'). After the respective alignments, SNP identification was performed for each target gene using Geneious Prime tools. A list of the olive varieties and samples included in the alignments, with the respective GenBank accession numbers, is described in Table 1. A total of 5 SNP positions were then selected for the SNP genotyping

experiments performed in this study.

### 2.6.4. Real-time allele-specific PCR method

Primers for real-time allele-specific amplification were designed targeting the 5 SNPs selected for cycloartenol synthase gene (SNP1-SNP5). A total of 3 primers were designed for each SNP position, including 2 forward primers, each one specifically targeting one of the alleles, and 1 reverse primer, common to both alleles. Forward primers were designed to include the target SNP site at the 3' end, and an extra mismatch at the antepenultimate nucleotide (3rd to the terminal) to improve allele discrimination, by inhibiting or significantly delaying amplification, when mismatch at the SNP site is observed. This position has been reported as the best to place the additional mismatch base in, allowing higher polymorphic percentages (J. Liu et al., 2012). The list of primers designed for the selected SNPs is described in Table 2. For each sample, 2 PCR reactions were performed for genotyping of each SNP position, using the primer sets that target each allele separately. Primers designed only including the target SNP site at the 3' end (without an extra mismatch) did not significantly affect the amplification efficiency when the mismatch at the SNP site was observed (data not shown). Amplification was carried out in QuantStudio™ 5 Real-time PCR System (Applied Biosystems™, Foster City, CA, USA) with QuantStudio™ Design & Analysis Software v1.5.1. The final reaction volume was 20 µL, including 2 µL of template DNA and 10 µL of PowerUp™ SYBR™ Green Master Mix (Thermo Fisher Scientific Inc., Waltham, MA, USA). The concentration of primers used was 200 nM. The remaining volume was completed with DNase, RNase free water. The thermal profile included an initial Uracil-DNA Glycosylase (UDG) treatment at 50 °C for 2 min and a hot start polymerase activation at 95 °C for 2 min, followed by 40 cycles of dissociation at 95 °C for 15 s and annealing-extension at 60 °C for 1 min. A melt curve step was included after the qPCR amplification, consisting of heating up to 95 °C for 15 s, cooling down to 60 °C for 1 min, and heating up again to 95 °C for 15 s, acquiring fluorescence every 0.3 °C. All samples were analyzed in duplicates and a non-template control (NTC) was always included.

### 2.6.5. HRM method

Primers for high resolution melting analysis were also designed targeting the same 5 SNPs selected for cycloartenol synthase gene (SNP1-SNP5). In this case, primers were designed targeting conserved regions flanking the target SNP position. Considering that CYC-SNP3, CYC-SNP4, and CYC-SNP5 were relatively close to each other, only one set of primers was designed for this group of SNPs. A total of 3 different sets of primers were designed for HRM analysis, targeting CYC-SNP1, CYC-SNP2, and CYC-SNP3-5, respectively. The detailed list of primers designed for this SNP genotyping approach is described in Table 2. PCR amplification was carried out in StepOne Plus™ Real-Time PCR system (Applied Biosystems™, Foster City, CA, USA) with StepOne™ Software v2.2.2. Duplicates of each sample were amplified in a final reaction volume of 20 µL, including 2 µL of template DNA and 10 µL of MeltDoctor™ HRM Master Mix (Thermo Fisher Scientific Inc., Waltham, MA, USA). The primer concentration was 300 nM, and the remaining volume was completed with DNase, RNase free water. The thermal profile included an initial step at 95 °C for 10 min, followed by 40 cycles of dissociation at 95 °C for 15 s, and annealing-extension at 60 °C for 1 min. The melting curve was obtained in continuous by heating up to 95 °C for 15 s, cooling down to 60 °C for 1 min, heating up again to 95 °C for 15 s, acquiring fluorescence every 0.3 °C, and cooling down once more to 60 °C for 15 s. High Resolution Melt Software v3.0.1 (Applied Biosystems™, Foster City, CA, USA) was used to analyze the data.

### 2.6.6. Evaluation of SNP-based methods

For SNP genotyping by real-time allele-specific PCR, result evaluation was performed by comparison of the amplification results obtained for each sample using the two allele-specific primer sets designed for



**Table 2**

List of primers designed for SNP genotyping by real-time allele-specific PCR and high resolution melting methods.

Genotyping Method	Target SNP	Primer	Primer Sequence (5'→3')*	Mismatch(J. Liu et al., 2012)	Amplicon size (bp)
Real-time Allele Specific PCR	CYC-SNP1	SNP1_a1_F	ACTGGATGCCAGGCGATTGGAT	3': TG Extra: AG	231
		SNP1_a2_F	ACTGGATGCCAGGCGATTGGAC		
		SNP1_a_R	GCTACAGCAACAAAATCCAGGA		
	CYC-SNP2	SNP2_a1_F	TCCTGGATTTTGTGCTGTAGCTTA	3': GA Extra: TT	212
		SNP2_a2_F	TCCTGGATTTTGTGCTGTAGCTTC		
		SNP2_a_R	TGTCCTGTAAGATGCTTATTTTGA		
	CYC-SNP3	SNP3_a1_F	AATTTAGCAGAAAATGATAGGTTAA	3': CA Extra: CT	226
		SNP3_a2_F	AATTTAGCAGAAAATGATAGGTTAG		
		SNP3_a_R	GTAAGCCCGACTACTATCGCAAGA		
	CYC-SNP4	SNP4_a1_F	CTTGAGACGTGTAGATTGAATGG	3': GT Extra: TT	186
		SNP4_a2_F	CTTGAGACGTGTAGATTGAATGA		
		SNP4_a_R	CAGTACTATCGCAAGATTTTCTC		
	CYC-SNP5	SNP5_a1_F	AAGGATAACTGGATAATTGTA	3': GA Extra: CT	167
		SNP5_a2_F	AAGGATAACTGGATAATTGTAC		
		SNP5_a_R	CCAGTACTATCGCAAGATT		
High Resolution Melting	CYC-SNP1	SNP1_F	ACTGGATGCCAGGCGATTG	-	105
		SNP1_R	CATCACAAATGTAATGTATTACAACCA		
	CYC-SNP2	SNP2_F	GAAAGGGATAATTTTACAGAAATTGG	-	139
		SNP2_R	TCACCTATCATTTTCTGCTAAATTTCA		
	CYC-SNP3-5	SNP3-5_F	TTTAGCAGAAAATGATAGGTGA	-	124
		SNP3-5_R	AATTCCTCCAAGCATTATTATCGA		

\* Nucleotides marked in **bold** represent the extra mismatch added to the allele-specific primers; Underlined nucleotides represent the SNP position in allele-specific primers.

each SNP position. For instance, for a particular SNP position, varieties in which both allele variants are present (heterozygous) are expected to have similar amplification efficiency for both PCR reactions. On the other hand, varieties in which only one of the allele variants is present (homozygous) are expected to have amplification for that allele variant only, and no amplification (or a significant delay) for the other target allele. With this in mind, it was possible to verify which alleles were present at a particular SNP position for each variety. Combining the information obtained in each SNP position, it was possible to define the profile expected for the pure varieties tested. For HRM, analysis was performed using the same set of primers to amplify both alleles, and any variation was identified through the analysis of the melting curves for each SNP position. Combining the information obtained from each set of primers also allowed to define the profile expected for the pure varieties using this method. Both methods were tested for pure varieties using DNA extracts obtained from leaves and fruits with NucleoSpin® Food and NucleoSpin® Plant kits to evaluate the effect of the plant tissue and DNA extraction protocol on the SNP genotyping results. In addition, mixtures of DNA extracts from the different varieties were prepared in different proportions and tested using both genotyping methods to evaluate their potential to distinguish pure from mixed samples.

### 3. Results and discussion

#### 3.1. DNA extraction and purification

Three DNA extraction and purification protocols were used to obtain DNA extracts from leaf samples. After quantification, it was observed that, although the DNA concentration of extracts obtained using NucleoSpin® Food kit was up to 10 times higher than NucleoSpin® Plant kit, DNA extracts from both commercial kits showed similar purity, presenting an average absorbance ratio of 1.7 for A260/A280, and 1.2 for A260/A230. The differences observed regarding concentration of DNA extracts may not only be related to the higher amount of initial sample material used in NucleoSpin® Food kit but also due to the extra filtration step that was performed in NucleoSpin® Plant kit, increasing the number of steps of the protocol and the amount of DNA lost in the process. DNA extraction using CTAB was the method that resulted in DNA extracts with the highest concentration, being up to 8 times higher than NucleoSpin® Food kit. In addition, the purity ratios obtained with this method were also higher than the commercial kits, presenting an

average A260/A280 ratio of 2.0, and A260/A230 of 1.6. In the case of fruit samples, DNA extraction and purification was performed with the commercial kits. Similar to the leaf samples, DNA concentration of extracts obtained using NucleoSpin® Food kit was also higher than the one obtained with NucleoSpin® Plant kit for fruit samples, presenting an average absorbance ratio of 1.4 for A260/A280, and 1.0 for A260/A230. The precipitation method using CTAB allowed higher DNA extraction yields and higher purity ratios, however it is much more laborious and time consuming, and requires the use of hazardous reagents. Methods based on solid-phase extraction have shown its potential to be easily transferred into miniaturized devices, versus precipitation methods, making the integration with further DNA analysis steps easier (Carvalho et al., 2018).

#### 3.2. SSR analysis

##### 3.2.1. Evaluation of SSR markers

After amplification, amplicon sizes of the 14 SSR markers tested were determined by agarose gel electrophoresis and by microchip-based capillary electrophoresis (Bioanalyzer). The results obtained were evaluated and compared with the reference study for 'Brava' and 'Mansa de Figueiredo' regarding the number of alleles identified, quality of the bands obtained, and amplicon size. From the results obtained by agarose gel electrophoresis, 6 out of the 14 SSR markers tested were consistent with the reference study regarding these parameters, namely UDO99-019, UDO99-024, GAPU71B, GAPU101, GAPU103A, and ssrOeUA-DCA11. Although the number of monomorphic markers revealed through agarose gel electrophoresis was consistent with literature, except for ssrOeUA-DCA18 and ssrOeUA-DCA16, the amplicon sizes and/or number of alleles were sometimes different, as described in Table S4. On the other hand, from the results obtained with microchip-based capillary electrophoresis, a total of 9 out of the 14 SSR markers were comparable to the reference study when considering the same parameters, including UDO99-019, UDO99-024, GAPU59, GAPU71B, GAPU101, GAPU103A, ssrOeUA-DCA3, ssrOeUA-DCA11, and ssrOeUA-DCA18. In the case of ssrOeUA-DCA3 marker, it was observed that the size of the two alleles reported for 'Mansa de Figueiredo' were quite close (4 bp difference), which explains why only one allele was identified for this variety using agarose gel and microchip-based capillary electrophoresis methods. The lower number of alleles identified using these two methods is a consequence of their lower resolution capacity

compared to the automatic capillary sequencer used on the reference study.

Overall, from the selected SSR markers a total of 5 markers were polymorphic for 'Brava' and 'Mansa de Figueiredo', namely UDO99-024, GAPU103A, *ssrOeUA-DCA3*, *ssrOeUA-DCA11* and *ssrOeUA-DCA18*, presenting a different allelic pattern for each variety, and therefore allowing to distinguish them. The remaining 4 selected SSR markers, were found to be monomorphic, with similar number of alleles as found in the reference study. The remaining SSR markers, although polymorphic, were inconsistent with the reference study. The discrepancies observed were the absence or very low intensity of expected bands, as it was observed for markers UDO99-011, UDO99-043, *ssrOeUA-DCA15*, and *ssrOeUA-DCA16*, and the presence of extra bands, as it was the case of marker *ssrOeUA-DCA9*, for which an extra 200 bp band was observed in all 'Mansa de Figueiredo' samples tested but not described in the reference study. Additionally, it was observed that if the number of PCR cycles or primer concentration were increased for markers with a missing or very low intensity band, several artifacts such as stutter bands would be detected, making it difficult to correctly identify the alleles. Moreover, *ssrOeUA-DCA16*, which was described as monomorphic marker in reference study, showed polymorphism between 'Brava' and 'Mansa de Figueiredo' with amplicon sizes also quite different from literature. When comparing results obtained from different DNA extraction methods, both the commercial kits and the precipitation method using CTAB provided similar results regarding amplicon size. SSR markers have several advantages over SNPs, such as being highly polymorphic and easily transferable between closely related species (Guichoux et al., 2011). Also, with the advances on next-generation sequencing technologies, identification and selection of SSR markers have become cheaper and faster (Palumbo & Barcaccia, 2018). However, it was observed that different SSR markers have different detection efficiencies, which affect the limit of the detection between markers, requiring further optimization steps. In addition, discrepancies in allele sizing determination affected the reproducibility of the assay and made comparison of data between laboratories difficult, as evident in this study. Hence, selection and use of SSR markers should be done cautiously while using them as cultivar identification tools.

### 3.3.2. Potential for miniaturization of SSR-based methods

SSR markers have been extensively employed in plant genetic studies and food authentication (Pasqualone, Montemurro, Caponio, & Blanco, 2004; Reale et al., 2006), being PCR-based amplification the most commonly used method. After amplification, the amplicon sizes are usually measured in order to identify polymorphisms. The polymorphisms identified for a specific locus are related to variations in the amplicon size, which depend on the number of repeats of a particular DNA motif. Therefore, the accuracy of amplicon size determination is extremely important for SSR-based analysis. Automated capillary analyzers combined with fluorescently-labelled primers are usually the preferred technique for amplicon size determination due to their high accuracy. However, these instruments are quite expensive, large, and require specialized and experienced users. On the other hand, the use of microchips for SSR separations can bring several advantages, such as the reduced reagent and sample volumes, decreasing the cost per analysis and sample consumption. Other advantages of such miniaturized systems are the faster separations achieved, and the possibility for an easier integration with other miniaturized analytical steps, such as sample preparation and PCR-based amplification. When transferring capillary-based techniques to microchip-based devices, the main challenge is in the development of channel geometries allowing high-resolution separations in a small microchip footprint, without compromising resolution. The development of microchip-based SSR separation systems has been explored with the aim of moving towards the development of fully integrated and portable miniaturized devices for SSR analysis (G. Kim, Moon, Moh, Lim, & Guk, 2015). In this study, we were interested in evaluating if the SSR analysis of the 14 markers previously described

could be easily transferred from a highly specialized and accurate capillary analyzer to a more affordable microchip-based system, such as Bioanalyzer. Bioanalyzer is a commercial lab-on-a-chip technology based on traditional gel electrophoresis principles, providing automated sizing and quantification. According to the manufacturer's specifications, the Agilent DNA 1000 Assay has a coefficient of variation of 5% for sizing reproducibility, a sizing resolution of 5 bp, and a sizing accuracy of  $\pm 10\%$ . One of the main problems found in this study regarding SSR analysis using microchip-based capillary electrophoresis was defining the size of each amplicon. This was a result of variation on the sizing reproducibility observed between different chips, and even between different wells on the same chip. This variation is an important limitation for SSR-based analysis, particularly when analyzing samples with similar amplicon sizes. However, for samples in which the amplicon size differences were large enough to overcome this limitation, it was possible to confirm their polymorphisms. Nevertheless, this method allowed to identify more amplification products than the conventional gel electrophoresis, therefore providing more complete information regarding the samples. Overall, based on the results obtained using microchip-based capillary electrophoresis, the main limitations of microchip-based devices for SSR genotyping were the amplicon sizing accuracy and resolution, limiting the applicability of this methodology for more complex samples.

### 3.3. SNP-based analysis

#### 3.3.1. Identification of SNPs

The sequences obtained for 'Brava', 'Mansa de Figueiredo' and 'Arbequina' varieties were aligned with sequences of other olive varieties available in GenBank. The SNPs identified for lupeol synthase and cycloartenol synthase genes, as well as the allele frequency determined for the olive varieties included in the alignments, are listed in Table S5 (Supplementary material). For lupeol synthase gene, a total of 3 SNPs were identified after the alignments, while for cycloartenol synthase gene a total of 7 SNPs were identified. In the case of the SNPs identified for cycloartenol synthase, two sequencing methods were performed: Sanger sequencing and sequencing with MinION. The sequencing results obtained with both methods were similar, with the exception that the results from MinION identified an additional SNP for cycloartenol synthase gene. In the current study, only the SNPs obtained by both methods were explored. Most of the SNPs identified corresponded to transitions, which occur between purines ( $A \leftrightarrow G$ ) or between pyrimidines ( $C \leftrightarrow T$ ). Transitions accounted for approximately two-thirds of all the SNPs identified, being in accordance with previous observations (Jiang, Wu, Zhang, Michal, & Wright, 2008) in which this is the most abundant type of SNP in plants and animals. The genotypes obtained for each SNP position, after the sequence alignments for both target genes, are described in Table S6 (Supplementary material). After careful analysis of the alignments, a total of 5 SNP positions were selected for the SNP genotyping experiments performed in this study to compare two different SNP-based genotyping methods for the differentiation of 'Brava' and 'Mansa de Figueiredo' varieties.

#### 3.3.2. Evaluation of real-time allele-specific PCR method

The 5 SNP positions selected for genotyping by real-time allele-specific PCR were CYC-SNP1, CYC-SNP2, CYC-SNP3, CYC-SNP4, and CYC-SNP5. Although, the genotyping experiments were mainly focused on distinguishing 'Brava' and 'Mansa de Figueiredo' varieties, samples of 'Arbequina' were also included in this study. The SNPs identified in lupeol synthase gene were not selected for the genotyping experiments. Although it was possible to distinguish 'Brava' from 'Mansa de Figueiredo' and 'Arbequina' with any of these SNPs, it was not possible to differentiate 'Mansa de Figueiredo' from 'Arbequina' with any of them. In addition, CYC-SNP6 and CYC-SNP7 were also not included in the genotyping experiments since they were located close to the end of the DNA sequences obtained for these samples, limiting the quality of

primer designing. Amplification reactions were optimized for the 5 selected SNPs, however for CYC-SNP3 and CYC-SNP5 only late or no amplification was observed for all the samples tested. As a result, these two target SNPs were excluded from this genotyping method.

**3.3.2.1. Pure olive varieties.** Amplification results obtained for pure samples of 'Brava', 'Mansa de Figueiredo' and 'Arbequina' varieties using the allele-specific primers designed for the remaining 3 SNP positions (CYC-SNP1, CYC-SNP2, and CYC-SNP4) are described in Figure S3. As previously described, this genotyping method requires two amplification reactions for each SNP position, targeting each allele variant separately. For CYC-SNP1, it was observed that 'Mansa de Figueiredo' and 'Arbequina' varieties showed similar amplification for both allele-specific PCR reactions, indicating that both allele variants (T and C) were present in these samples (heterozygous TC). However, for 'Brava', a significant delay on the PCR reaction targeting allele variant 2 (C) was observed in comparison to allele variant 1 (T), indicating that in this variety both alleles were the same (homozygous TT). For CYC-SNP2, 'Brava' and 'Arbequina' showed similar amplification for both allele variants (A and C), being both present in these varieties (heterozygous AC). On the other hand, for 'Mansa de Figueiredo', amplification was only observed for the allele variant 1 (A), indicating that both alleles were the same (homozygous AA). Finally, for CYC-SNP4, 'Brava' only showed amplification for the allele variant 1 (G), being both alleles the same (homozygous GG). In the case of 'Mansa de Figueiredo' and 'Arbequina', amplification was achieved for both allele variants targeted (G and A), being both present in these varieties (heterozygous GA). For this SNP position, it was noticed that, for 'Mansa de Figueiredo' and 'Arbequina', there was a short delay of approximately 3 cycles on the PCR reaction targeting allele variant 2 (A) compared to allele variant 1 (G). Although this short delay was observed, it was not significant when compared to the amplification profile obtained for 'Brava', which only presented one of the allele variants. This short delay can be related to the lower stability of the annealing for the primers targeting allele variant 2

(A) when compared with the primers targeting allele variant 1 (G), combined with the strong destabilization provided by the extra mismatch included in these primers, as described in Table 2 Overall, the SNP genotyping results obtained for CYC-SNP1, CYC-SNP2 and CYC-SNP4 using this method for pure 'Brava', 'Mansa de Figueiredo' and 'Arbequina' varieties were in accordance with the sequencing results previously described, validating the applicability of this methodology. In addition, this method was performed using DNA extracts from different plant tissues (leaves and fruits) and obtained with different commercial kits (NucleoSpin® Food kit and NucleoSpin® Plant kit), showing consistent results for all the conditions tested and, therefore, establishing good reproducibility of the method.

**3.3.2.2. Mixtures of olive varieties.** In addition to pure samples of 'Brava', 'Mansa de Figueiredo' and 'Arbequina', mixtures of DNA extracts from these varieties, prepared in different proportions, were also tested by real-time allele-specific PCR to evaluate the discriminatory potential of the developed methods in mixtures. The amplification results obtained for these mixtures targeting CYC-SNP1, CYC-SNP2 and CYC-SNP4 from cycloartenol synthase gene are described in Table 3. The results obtained for each mixture were compared with the results obtained for the controls (pure varieties) in order to evaluate the ability of this method to identify the authenticity of each sample, evaluating if the sample being tested could be a pure 'Brava' or a pure 'Mansa de Figueiredo'. According to the results described in Table 3, amplification profiles obtained for CYC-SNP2 were enough to confirm that none of the samples were pure 'Mansa de Figueiredo', since none of them showed no amplification (or a significant delay) as expected for this variety. On the other hand, amplification profiles obtained for CYC-SNP4 were enough to confirm that none of the samples were pure 'Brava', since none of them showed no amplification as expected for this variety at this SNP position. Therefore, real-time allele-specific PCR targeting CYC-SNP2 and CYC-SNP4 would be enough to confirm that these samples were neither pure 'Mansa de Figueiredo' nor pure 'Brava', however it would

**Table 3**

Real-time allele-specific PCR results obtained for mixtures of 'Brava', 'Mansa de Figueiredo' and 'Arbequina' varieties in different proportions using primer sets targeting SNP1, SNP2 and SNP4 identified in cycloartenol synthase gene.

Sample	Composition	Allele	Presence of target alleles			
			SNP1	SNP2	SNP4	
<b>CONTROLS</b>	Brava	100% Brava	a1	Yes	Yes	Yes
			a2	No Significant delay <sup>a</sup>	Yes Ct similar to a1	No No amplification
	Mansa de Figueiredo	100% Mansa de Figueiredo	a1	Yes	Yes	Yes
			a2	Yes Ct similar to a1	No No amplification or Significant delay <sup>a</sup>	Yes <sup>b</sup> Short delay
	Arbequina	100% Arbequina	a1	Yes	Yes	Yes
			a2	Yes Ct similar to a1	Yes Ct similar to a1	Yes <sup>b</sup> Short delay
<b>MIXTURES</b>	m1	Equal proportions of Brava, Mansa de Figueiredo and Arbequina	a1	Yes	Yes	Yes
			a2	Yes Short delay	Yes Ct similar to a1	Yes Short delay
	m2	50% Brava 25% Mansa de Figueiredo 25% Arbequina	a1	Yes	Yes	Yes
			a2	Yes Short delay	Yes Ct similar to a1	Yes Delay 6 Cycles
	m3	25% Brava 50% Mansa de Figueiredo 25% Arbequina	a1	Yes	Yes	Yes
			a2	Yes Short delay	Yes Ct similar to a1	Yes Short delay
	m4	25% Brava 25% Mansa de Figueiredo 50% Arbequina	a1	Yes	Yes	Yes
			a2	Yes Short delay	Yes Ct similar to a1	Yes Short delay
	m5	50% Brava 50% Mansa de Figueiredo	a1	Yes	Yes	Yes
			a2	Yes Short delay	Yes Ct similar to a1	Yes Delay 7 Cycles
	m6	75% Brava 25% Mansa de Figueiredo	a1	Yes	Yes	Yes
			a2	Yes Delay 6 Cycles	Yes Short delay	Yes Delay 9 Cycles
	m7	25% Brava 75% Mansa de Figueiredo	a1	Yes	Yes	Yes
			a2	Yes Ct similar to a1	Yes Short delay	Yes Short delay
	m8	50% Brava 50% Arbequina	a1	Yes	Yes	Yes
			a2	Yes Short delay	Yes Ct similar to a1	Yes Short delay
	m9	50% Mansa de Figueiredo 50% Arbequina	a1	Yes	Yes	Yes
			a2	Yes Ct similar to a1	Yes Ct similar to a1	Yes Short delay

Ct – Cycle threshold,

<sup>a</sup> Significant delay (> 8 cycles in SNP1; > 15 cycles in SNP2) for allele 2 compared to allele 1.

<sup>b</sup> Short delay (< 5 cycles) observed for allele 2 compared to allele 1.

not be possible to identify the correct composition of the samples. In the case of the amplification profiles obtained for CYC-SNP1, it was observed that none of the samples were pure 'Brava', except for sample m6 (75 % 'Brava', 25 % 'Mansa de Figueiredo') for which the amplification delay observed could raise some doubts. In addition, it was also observed that none of the samples were pure 'Mansa de Figueiredo' or pure 'Arbequina', except for samples m7 (75 % 'Mansa de Figueiredo', 25 % 'Arbequina') and m9 (50 % 'Mansa de Figueiredo', 50 % 'Arbequina') for which it was not possible to distinguish, since these samples were both composed by varieties with similar profiles for this SNP position. Although the genotyping experiments were mainly focused on distinguishing 'Brava' and 'Mansa de Figueiredo' varieties, it was also possible to confirm that none of the samples tested were pure 'Arbequina' using this method, except for sample m9 (50 % 'Mansa de Figueiredo', 50 % 'Arbequina'), for which this confirmation was not possible since it showed the same profile as a pure 'Arbequina'. This variety presents a genotype similar to 'Brava' for CYC-SNP2 and similar to 'Mansa de Figueiredo' for CYC-SNP1 and CYC-SNP4.

### 3.3.3. Evaluation of HRM method

Similar to the real-time allele-specific PCR methods, the SNP

positions selected for genotyping by HRM were CYC-SNP1, CYC-SNP2, CYC-SNP3, CYC-SNP4, and CYC-SNP5. In this case, primers were designed to hybridize in regions flanking the target SNP positions. Therefore, since the positions of CYC-SNP3, CYC-SNP4, and CYC-SNP5 were relatively close, the same set of primers was used to target these SNPs simultaneously (CYC-SNP3-5), as described in Table 2.

**3.3.3.1. Pure olive varieties.** The derivative melting curves and respective aligned melting curves obtained for pure samples of 'Brava', 'Mansa de Figueiredo' and 'Arbequina' varieties using the HRM method are illustrated in Figure S4. It was observed that for varieties presenting both allele variants (heterozygous), two close peaks were identified in the derivative melting curves, as it was the case of 'Mansa de Figueiredo' and 'Arbequina' for CYC-SNP1, and 'Brava' and 'Arbequina' for CYC-SNP2. The difference between these two peaks was more significant in the case of 'Mansa de Figueiredo' and 'Arbequina' for CYC-SNP3-5, since these results were targeting three SNP positions simultaneously, and all of them presented both allele variants. On the other hand, for varieties presenting only one of the allele variants (homozygous), only a single peak was identified in the respective derivative melting curve, as it was the case of 'Brava' for CYC-SNP1 and CYC-SNP3-5, and the case of

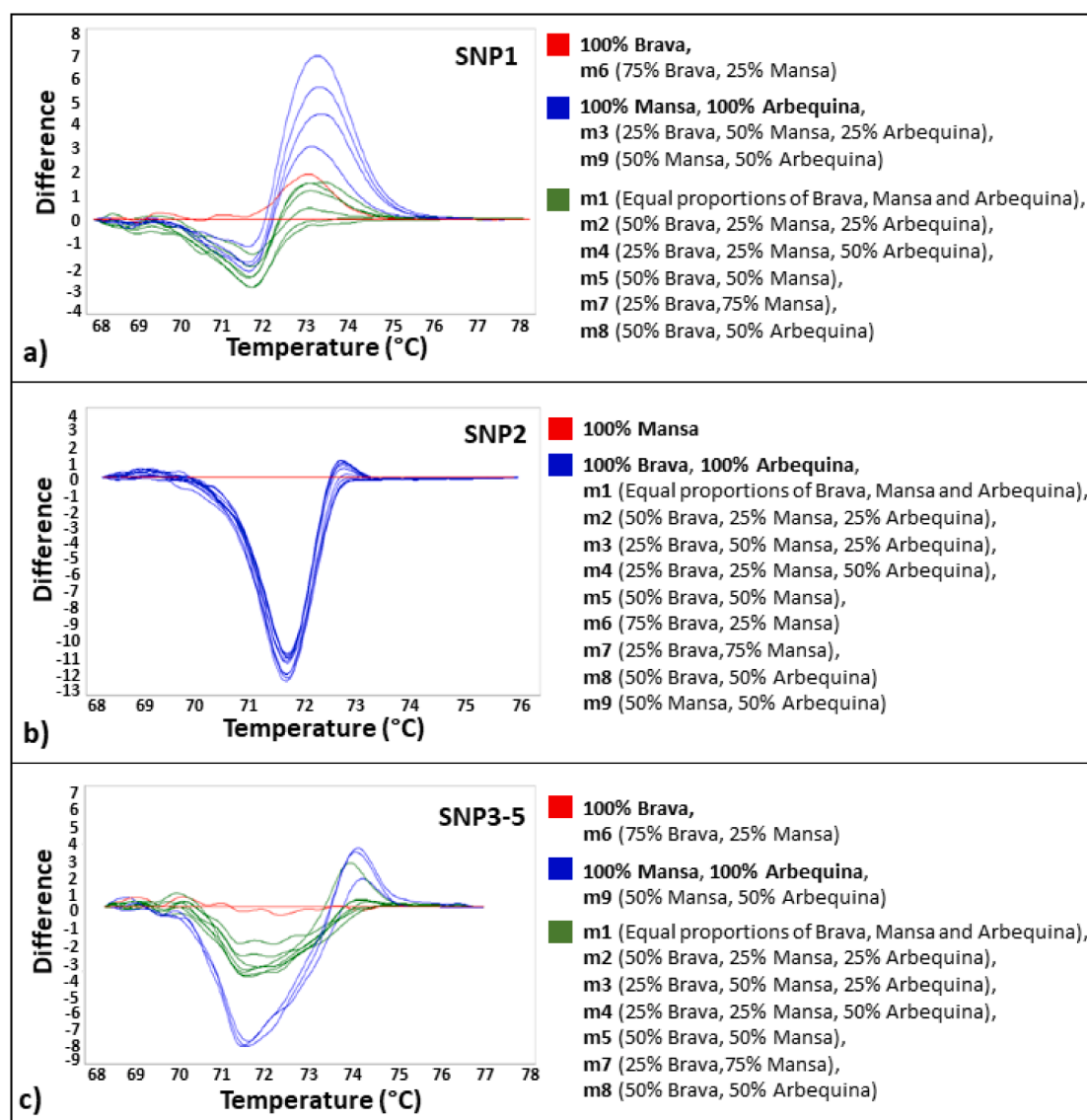


Fig. 1. High resolution melting analysis of mixtures of 'Brava', 'Mansa de Figueiredo' and 'Arbequina' varieties targeting the SNPs identified in cycloartenol synthase gene.



'Mansa de Figueiredo' for CYC-SNP2. Any of the SNP positions targeted using these sets of primers can be used to distinguish pure 'Brava' from pure 'Mansa de Figueiredo'. In order to also differentiate pure 'Arbequina', a combination of CYC-SNP1 with CYC-SNP2 or CYC-SNP2 with CYC-SNP3-5 is required because this variety presents a profile similar to 'Mansa de Figueiredo' for CYC-SNP1 and CYC-SNP3-5, and similar to 'Brava' for CYC-SNP2. From the aligned melting curves obtained it was possible to distinguish the profile of varieties with only one of the allele variants or both the allele variants in a particular SNP position. With this method, it was not only possible to distinguish the varieties due to their different melting curve profiles, but it was also possible to confirm if they presented only one or both allele variants for the target SNP positions. Similar to the real-time allele-specific PCR method, this approach was also evaluated using different plant tissues and different commercial kits, showing reproducible results for all the conditions tested, and being in accordance with the sequencing results previously obtained.

**3.3.3.2. Mixtures of olive varieties.** Mixtures of 'Brava', 'Mansa de Figueiredo' and 'Arbequina' varieties, prepared in different proportions, were also tested using the HRM method targeting CYC-SNP1, CYC-SNP2 and CYC-SNP3-5. The difference plots obtained with the High Resolution Melt Software v3.0.1 for the different samples tested are described in Fig. 1. Pure varieties presenting only one of the allele variants were always the ones selected as reference to obtain the difference plots. With this in mind, for CYC-SNP1 and CYC-SNP3-5 the variety selected as reference was 'Brava', while for CYC-SNP2 the reference was 'Mansa de Figueiredo'. From the results obtained for CYC-SNP2, it was possible to confirm that none of the mixtures tested was pure 'Mansa de Figueiredo'. In addition, when targeting CYC-SNP1 and CYC-SNP3-5, it was also possible to confirm that none of the mixtures tested was pure 'Brava', except for sample m6 (75 % 'Brava', 25 % 'Mansa de Figueiredo') which presented a profile similar to 'Brava'. This sample was mainly composed by 'Brava', making it more difficult to differentiate it from pure 'Brava' based on the melting curve profile. Although the main focus was to distinguish 'Brava' and 'Mansa de Figueiredo' varieties, it was also possible to confirm that none of the mixtures tested were pure 'Arbequina', with the exception of sample m9 (50% 'Mansa de Figueiredo', 50 % 'Arbequina') which showed the same profile as a pure 'Arbequina', through the analysis of the difference plots targeting CYC-SNP3-5 and CYC-SNP2.

### 3.3.4. Potential for miniaturization of SNP-based methods

The development of microfluidic systems for SNP detection has been rapidly growing, including a much larger diversity of techniques available compared to microfluidic devices for SSR genotyping. Several approaches have been reported, showing the potential of SNP-based methods in a miniaturized setting. Particularly for allele-specific PCR, a microfluidic system combining this technique with a microarray-chip was developed and successfully applied for SNP genotyping of indigenous and imported beef cattle (Choi et al., 2012). Miniaturization of allele-specific PCR combined with fluorescently-labeled oligonucleotides has also been achieved for different applications, such as ancestry inference using KASP™ markers in forensic science (Ren et al., 2019), or SNP detection directly from whole blood samples using TaqMan® markers (L. Zhang et al., 2016). A microfluidic chip using microbeads has also been developed (Chang et al., 2017), showing promising results for the implementation of a bead-based allele-specific qPCR methodology onto a microchip for SNP genotyping. Similarly, melting curve analysis has also been implemented for SNP genotyping in microfluidic devices. Several approaches have been developed for this purpose, including continuous-flow systems (Crews, Wittwer, Montgomery, Pryor, & Gale, 2009), bead-based devices (Kao et al., 2014), and droplet-based platforms (F. W. Liu, Ding, Lin, Lu, & Jang, 2017). Overall, it has been demonstrated that miniaturization of SNP-based methods can further reduce cost of analysis and risk of contamination, as well as

expedite assay time, making detection at the point-of-need a possibility. In general, SNPs are widely distributed and highly abundant in the genome, present high genetic stability and great repeatability, and allow high-throughput automated analysis (Yang, Kang, Yang, Lin, & Fang, 2013), being these some of their advantages over SSR markers. In addition, as observed in this study, data analysis can also be simpler in the case of allele-specific PCR, since the results can be condensed to the presence or absence of a target allele. In the case of SSR markers, size determination can be affected by the PCR reaction, and by the accuracy and reproducibility of the equipment used for size determination, which makes it difficult to compare data from different laboratories. In our opinion, when considering miniaturization, SNP-based methods seem to be the simplest to integrate in a microfluidic system. However, in the case of allele-specific PCR, it is required that the primers designing and the amplification reaction have been properly optimized, while for HRM it is necessary that the miniaturized system is able to provide defined melting curves to properly distinguish the different profiles, which can be a limitation particularly when analyzing more complex mixtures.

## 4. Conclusions

SSR-based and SNP-based methods, were compared and evaluated regarding their potential for integration in a miniaturized device. The methods were evaluated for the identification of two olive varieties from the Northwest of Spain: 'Brava' and 'Mansa de Figueiredo'. For the SSR-based methods it was observed that the commercial microfluidic-based capillary electrophoresis did not provide the required accuracy and reproducibility for amplicon size determination, which can make the analysis particularly challenging when analyzing complex mixtures of different varieties. Regarding SNP-based methods, two approaches were explored in this study: real-time allele-specific PCR and HRM analysis. The HRM method required more time for result analysis, particularly for mixtures of varieties, being highly dependent on the quality of the melting curves obtained, which can be a limitation when considering miniaturization. Considering all the aspects described in this work, allele-specific PCR might be the best option for miniaturization, since the results can be condensed to a presence/absence type of answer for the target allele, and it showed the best results for both pure and mixed varieties.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the project NANO EATERS: Valorization and transfer of NANOTECHNOLOGIES to EARLY ADOPTERS of the Euroregion Galicia-Norte Portugal (200005902 - NANO EATERS - Aceite de oliva), supported by INTERREG V-A España-Portugal (POCTEP) 2014-2020 and by the project Nanotechnology Based Functional Solutions (NORTE-01-0145-FEDER-000019), supported by Norte Portugal Regional Operational Programme (NORTE2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund. S. A. acknowledges the Portuguese funding institution FCT – Fundação para Ciência e Tecnologia for Ph.D. scholarship SFRH/BD/140396/2018. The authors would also like to thank to the Department of Analytical Chemistry and Food Science from the University of Vigo (Spain) for providing the plant material used in this work.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fochms.2021.100038>.

## References

- Baldoni, L., Cultrera, N. G., Mariotti, R., Ricciolini, C., Arcioni, S., Vendramin, G. G., ... Testolin, R. (2009). A consensus list of microsatellite markers for olive genotyping. *Molecular Breeding*, 24(3), 213–231. <https://doi.org/10.1007/s11032-009-9285-8>
- Bazakos, C., Dulger, A. O., Uncu, A. T., Spaniolas, S., Spano, T., & Kalaitzis, P. (2012). A SNP-based PCR-RFLP capillary electrophoresis analysis for the identification of the varietal origin of olive oils. *Food Chemistry*, 134(4), 2411–2418. <https://doi.org/10.1016/j.foodchem.2012.04.031>
- Ben Amar, F., Souabni, H., Saddoud-Debbabi, O., & Ali Triki, M. (2021). Reliability of Morphological Characters in Identification of Olive (*Olea europaea* L.) Varieties in Ex-Situ Conditions. *Current Research in Agricultural Sciences*, 8(2), 56–64. <https://doi.org/10.18488/journal.68.2021.82.56.64>
- Carvalho, J., Puertas, G., Gaspar, J., Azinheiro, S., Diéguez, L., Garrido-Maestu, A., ... Prado, M. (2018). Highly efficient DNA extraction and purification from olive oil on a washable and reusable miniaturized device. *Analytica Chimica Acta*, 1020, 30–40. <https://doi.org/10.1016/j.aca.2018.02.079>
- Chang, Y. M., Ding, S. T., Lin, E. C., Wang (Alex), L., & Lu, Y. W. (2017). A microfluidic chip for rapid single nucleotide polymorphism (SNP) genotyping using primer extension on microbeads. *Sensors and Actuators, B: Chemical*, 246, 215–224. <https://doi.org/10.1016/j.snb.2017.01.160>
- Choi, J. Y., Kim, Y. T., Byun, J. Y., Ahn, J., Chung, S., Gweon, D. G., Kim, M. G., & Seo, T. S. (2012). An integrated allele-specific polymerase chain reaction-microarray chip for multiplex single nucleotide polymorphism typing. *Lab on a Chip*, 12(24), 5146. <https://doi.org/10.1039/c2lc40878c>
- Costa, J., Mafra, I., & Oliveira, M. B. P. P. (2012). Advances in vegetable oil authentication by DNA-based markers. *Trends in Food Science and Technology*, 26(1), 43–55. <https://doi.org/10.1016/j.tifs.2012.01.009>
- Crews, N., Wittwer, C. T., Montgomery, J., Pryor, R., & Gale, B. (2009). Spatial DNA melting analysis for genotyping and variant scanning. *Analytical Chemistry*, 81(6), 2053–2058. <https://doi.org/10.1021/ac801495w>
- De La Rosa, R., James, C. M., & Tobutt, K. R. (2002). Isolation and characterization of polymorphic microsatellites in olive (*Olea europaea* L.) and their transferability to other genera in the Oleaceae. *Molecular Ecology Notes*, 2(3), 265–267. <https://doi.org/10.1046/j.1471-8286.2002.00217.x>
- European Parliament and European Council. (2012). Regulation (EU) No 1151/2012 of the European Parliament and of the Council of 21 November 2012 on quality schemes for agricultural products and foodstuffs. *Official Journal of the European Union*, L343, 1–29.
- Figueiredo-González, M., Reboredo-Rodríguez, P., González-Barreiro, C., Carrasco-Pancorbo, A., Simal-Gándara, J., & Cancho-Grande, B. (2018). Nutraceutical potential of phenolics from brava and mansa extra-virgin olive oils on the inhibition of enzymes associated to neurodegenerative disorders in comparison with those of picual and cornicabra. *Molecules*, 23(4). <https://doi.org/10.3390/molecules23040722>
- Gago, P., Santiago, J. L., Boso, S., & Martínez, M. C. (2019). The forgotten, ancient olive trees of the Spanish northwest: A first molecular and botanical analysis. *Spanish Journal of Agricultural Research*, 17(2). <https://doi.org/10.5424/sjar/2019172-13572>
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., ... Petit, R. J. (2011). Current trends in microsatellite genotyping. *Molecular Ecology Resources*, 11(4), 591–611. <https://doi.org/10.1111/j.1755-0998.2011.03014.x>
- Hakim, I. R., Kammoun, N. G., Makhloufi, E., & Rebaï, A. (2010). Discovery and potential of SNP markers in characterization of Tunisian olive germplasm. *Diversity*, 2(1), 17–27. <https://doi.org/10.3390/d2010017>
- Jiang, Z., Wu, X. L., Zhang, M., Michal, J. J., & Wright, R. W. (2008). The complementary neighborhood patterns and methylation-to-mutation likelihood structures of 15,110 single-nucleotide polymorphisms in the bovine genome. *Genetics*, 180(1), 639–647. <https://doi.org/10.1534/genetics.108.090860>
- Kao, P. C., Ding, S. T., Lin, E. C., Li, K. C., Wang, L., & Lu, Y. W. (2014). A bead-based single nucleotide polymorphism (SNP) detection using melting temperature on a microchip. *Microfluidics and Nanofluidics*, 17(3), 477–488. <https://doi.org/10.1007/s10404-014-1331-0>
- Kim, G., Moon, J. H., Moh, C. Y., Lim, & Guk, J. (2015). A microfluidic nano-biosensor for the detection of pathogenic Salmonella. *Biosensors and Bioelectronics*, 67, 243–247. <https://doi.org/10.1016/j.bios.2014.08.023>
- Kim, Y. T., Lee, D., Heo, H. Y., Sim, J. E., Woo, K. M., Kim, D. H., ... Seo, T. S. (2016). Total integrated slidable and valveless solid phase extraction-polymerase chain reaction-capillary electrophoresis microdevice for mini Y chromosome short tandem repeat genotyping. *Biosensors and Bioelectronics*, 78, 489–496. <https://doi.org/10.1016/j.bios.2015.11.079>
- Liu, F. W., Ding, S. T., Lin, E. C., Lu, Y. W., & Jang, J. S. R. (2017). Automated melting curve analysis in droplet microfluidics for single nucleotide polymorphisms (SNP) genotyping. *RSC Advances*, 7(8), 4646–4655. <https://doi.org/10.1039/C6RA26484K>
- Liu, J., Huang, S., Sun, M., Liu, S., Liu, Y., Wang, W., ... Hua, W. (2012). An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods*, 8(1). <https://doi.org/10.1186/1746-4811-8-34>
- MAPA. (2019). Superficies y producciones anuales de cultivos. Ministerio de Agricultura, Pesca y Alimentación. <https://www.mapa.gob.es/es/estadistica/temas/estadisticas-agrarias/agricultura/superficies-producciones-anuales-cultivos/>
- National of library Medicine. (2006). GenBank Overview.
- Palumbo, F., & Barcaccia, G. (2018). Critical Aspects on the Use of Microsatellite Markers for Assessing Genetic Identity of Crop Plant Varieties and Authenticity of their Food Derivatives. In *Rediscovery of Landraces as a Resource for the Future*. InTech. <https://doi.org/10.5772/intechopen.70756>
- Pasqualone, A., Montemurro, C., Caponio, F., & Blanco, A. (2004). Identification of Virgin Olive Oil from Different Cultivars by Analysis of DNA Microsatellites. *Journal of Agricultural and Food Chemistry*, 52(5), 1068–1071. <https://doi.org/10.1021/jf0348424>
- Pereira, L., Gomes, S., Barrias, S., Fernandes, J. R., & Martins-Lopes, P. (2018). Applying high-resolution melting (HRM) technology to olive oil and wine authenticity. In *Food Research International* (pp. 170–181). Elsevier Ltd.. <https://doi.org/10.1016/j.foodres.2017.10.026>
- Reale, S., Doveri, S., Díaz, A., Angiolillo, A., Lucentini, L., Pilla, F., ... Lee, D. (2006). SNP-based markers for discriminating olive (*Olea europaea* L.) cultivars. *Genome*, 49(9), 1193–1205. <https://doi.org/10.1139/G06-068>
- Reboredo-Rodríguez, P., González-Barreiro, C., Cancho-Grande, B., Simal-Gándara, J., & Trujillo, I. (2018). Genotypic and phenotypic identification of olive cultivars from north-western Spain and characterization of their extra virgin olive oils in terms of fatty acid composition and minor compounds. *Scientia Horticulturae*, 232, 269–279. <https://doi.org/10.1016/j.scienta.2018.01.015>
- Ren, P., Liu, J., Zhao, H., Fan, X. P., Xu, Y. C., & Li, C. X. (2019). Construction of a rapid microfluidic-based SNP genotyping (MSG) chip for ancestry inference. *Forensic Science International: Genetics*, 41, 145–151. <https://doi.org/10.1016/j.fsigen.2019.04.006>
- Scheben, A., Batley, J., & Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. In *Plant Biotechnology Journal* (pp. 149–161). Blackwell Publishing Ltd.. <https://doi.org/10.1111/pbi.12645>
- Trujillo, I., Ojeda, M. A., Urdirro, N. M., Potter, D., Barranco, D., Rallo, L., & Diez, C. M. (2014). Identification of the Worldwide Olive Germplasm Bank of Córdoba (Spain) using SSR and morphological markers. *Tree Genetics and Genomes*, 10(1), 141–155. <https://doi.org/10.1007/s11295-013-0671-3>
- Velez, D. O., Mack, H., Jupe, J., Hawker, S., Kulkarni, N., Hedayatnia, B., ... Fraley, S. I. (2017). Massively parallel digital high resolution melt for rapid and absolutely quantitative sequence profiling. *Scientific Reports*, 7(1), 1–14. <https://doi.org/10.1038/srep42326>
- Vietina, M., Agrimonti, C., Marmiroli, M., Bonas, U., & Marmiroli, N. (2011). Applicability of SSR markers to the traceability of monovarietal olive oils. *Journal of the Science of Food and Agriculture*, 91(8), 1381–1391. <https://doi.org/10.1002/jsfa.4317>
- Von Thaden, A., Cocchiara, B., Jarausch, A., Jüngling, H., Karamanlidis, A. A., Tiesmeyer, A., ... Muñoz-Fuentes, V. (2017). Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. *Scientific Reports*, 7(1), 1–13. <https://doi.org/10.1038/s41598-017-10647-w>
- Xanthopoulou, A., Ganopoulos, L., Koubouris, G., Tsafaris, A., Sergendani, C., Kalivas, A., & Madesis, P. (2014). Microsatellite high-resolution melting (SSR-HRM) analysis for genotyping and molecular characterization of an *Olea europaea* germplasm collection. *Plant Genetic Resources: Characterisation and Utilisation*, 12(3), 273–277. <https://doi.org/10.1017/S147926211400001X>
- Yan, J., Erasmus, S. W., Aguilera Toro, M., Huang, H., & van Ruth, S. M. (2020). Food fraud: Assessing fraud vulnerability in the extra virgin olive oil supply chain. *Food Control*, 111, Article 107081. <https://doi.org/10.1016/j.foodcont.2019.107081>
- Yang, W., Kang, X., Yang, Q., Lin, Y., & Fang, M. (2013). Review on the development of genotyping methods for assessing farm animal diversity. In *Journal of Animal Science and Biotechnology* (Vol. 4, Issue 1, p. 2). BioMed Central. <https://doi.org/10.1186/2049-1891-4-2>
- Zamuz, S., Purriños, L., Tomasevic, I., Domínguez, R., Brncić, M., Barba, J. F., & Lorenzo, J. M. (2020). Consumer Acceptance and Quality Parameters of the Commercial Olive Oils Manufactured with Cultivars Grown in Galicia (NW Spain). *Foods*, 9(4), 427. <https://doi.org/10.3390/foods9040427>
- Zec, H. C., Zheng, T., Liu, L., Hsieh, K., Rane, T. D., Pederson, T., & Wang, T.-H. (2018). Programmable microfluidic genotyping of plant DNA samples for marker-assisted selection. *Microsystems & Nanoengineering*, 4(1), 1–10. <https://doi.org/10.1038/micronano.2017.97>
- Zhang, L., Cai, Q., Wiederkehr, R. S., Fauvart, M., Fiorini, P., Majeed, B., ... Stakenborg, T. (2016). Multiplex SNP genotyping in whole blood using an integrated microfluidic lab-on-a-chip. *Lab on a Chip*, 16(20), 4012–4019. <https://doi.org/10.1039/c6lc01046f>
- Zhang, Y., & Ozdemir, P. (2009). Microfluidic DNA amplification-A review. *Analytica Chimica Acta*, 638(2), 115–125. <https://doi.org/10.1016/j.aca.2009.02.038>