

STR genotyping and genetic clustering analysis in the grapevine (*Vitis vinifera*)

Sara Raquel de Almeida Ribeiro dos Santos

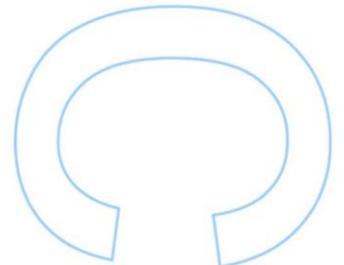
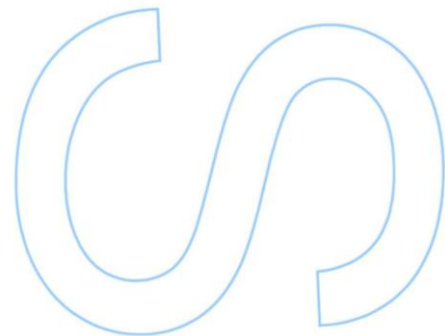
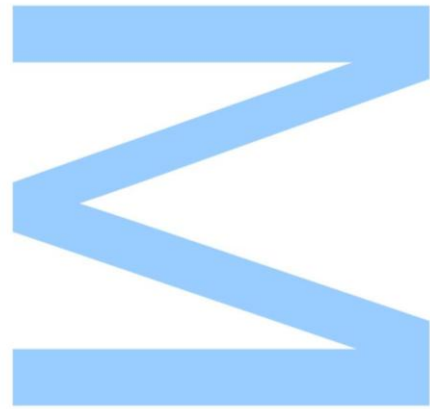
Mestrado em Genética Forense

Departamento de Biologia

2013

Orientador

Barbara van Asch, PhD, IPATIMUP

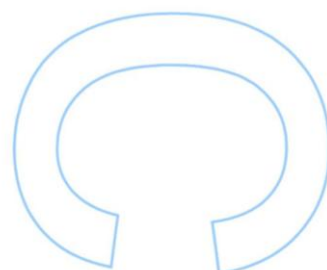
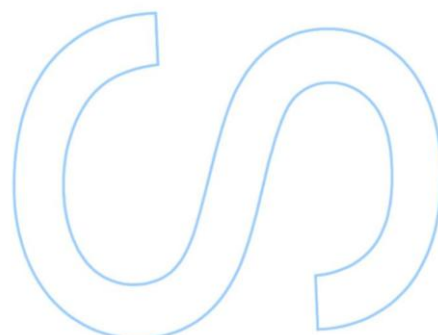
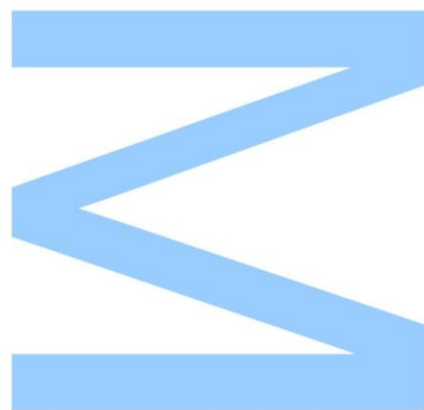




Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____/____/____



Dissertação de candidatura ao grau de mestre em
Genética Forense submetida à Faculdade de
Ciências da Universidade do Porto.

O presente trabalho foi desenvolvido no Instituto de
Patologia e Imunologia Molecular da Universidade do
Porto sob orientação da Doutora Barbara van Asch.

Dissertation for applying to a Master's degree in
Forensic Genetics, submitted to the Faculty of
Sciences of the University of Porto.

The present work was developed at the Institute of
Molecular Pathology and Immunology of the
University of Porto under the scientific supervision of
Barbara van Asch, PhD.

Agradecimentos

Em primeiro lugar quero agradecer à minha orientadora Barbara van Asch por toda a ajuda, todo o apoio, toda a preocupação, toda a insistência, pelas conversas e pelo trabalho de “Ama-seca” por que passou, foi sem dúvida graças a si que pude concluir a tese e entrega-la. Um muito obrigado do fundo do coração.

Quero agradecer ao Professor Amorim a oportunidade que me deu de tirar este mestrado, continuando assim os meus estudos.

Ao grupo de Genética Populacional um obrigado, em especial à Joana por estar sempre disponível a esclarecer qualquer dúvida e ajudar no laboratório.

Agradeço à minha mãe todo o esforço que fez para eu tirar este mestrado, sei que não foi fácil. Mas esta etapa já acabou. Obrigada mãe. Amo-te.

Um beijo muito especial para a minha avó Júlia que me brindava, todos os fins-de-semana, com a sua maravilhosa frase: “Agarra-te aos estudos minha filha, para ver se acabas rápido”.

Um obrigado enorme para os meus primos Rita e Rúben por me terem lá em casa este ano que passou, não há palavras para agradecer tudo o que fizeram por mim, espero que este singelo agradecimento possa de alguma forma reconhecer a minha gratidão.

À tia Xana, ao tio Roberto, e a toda a minha família um beijo e um obrigado pelo apoio e por toda a ajuda.

A ti, Moina, por seres um chato de vez em quando, mas por me fazeres feliz e suportares os meus momentos mais caóticos, obrigada (já passou esta fase, mais etapas esperam por nós).

Um obrigada a todos que de forma direta ou indireta fizeram parte deste meu percurso, os coleguinhas de mestrado, os amigos verdadeiros que ficaram de

Bragança, e, claro, a todos os meus amigos de Vinhais que fazem parte da minha vida e sem os quais nada seria o mesmo.

Abstract

The grapevine (*Vitis vinifera* L.) is a diploid plant that has a relatively small genome size (475-500 Mb) organized in 19 chromosomes. This plant is one of the most valuable and oldest horticultural crops used to produce table fruit, raisins, juice and wine. The nomenclature of grape varieties has resulted in a complex pattern of synonyms and homonyms.

Variety identification is necessary to grape growers, winemakers, regulatory authorities, and consumers. The increasing interest in cultivar identification results from the need of planting and producing correctly identified grapes and derived products, due to the economic value of conformity with protected designations of origin. Accurate identification of grapevine cultivars is crucial and so is the importance of a good identification system throughout the winemaking process, including variety provenance. Nowadays, identification methods include ampelography, chemical and biochemical methods and molecular methods. Amongst the molecular methods, microsatellite markers (STRs) have become the most widely used type of DNA marker for identification of grapevine cultivars since their properties enable a broad range of applications. Due to their high discriminative power, the finding of identical genotype in two different plants is strong evidence that these plants belong to the same cultivar. Therefore, STR analysis can be used to determine cultivar identity and to identify plant material of unknown varietal origin by comparing the genotype obtained from the sample with reference genotypes of cultivars stored in a database. However, utilizing di-nucleotide STRs present significant disadvantages such as stuttering and requires very accurate and reliable protocols for allele separation and identification to avoid allele miscalling. A more reliable method is the use of STRs with long core motif, namely tetra-nucleotides.

Aiming for the use of tetra-nucleotide STRs and the standardization of identification methods in grapevine, we introduce a critical review of STR markers for grapevine cultivar identification applied to both international and Portuguese varieties in the context of forensic genetics. We have analyzed the STR markers utilized in grapevine identification by comparing genotypes of di-nucleotide markers (sequencing alleles in 12 Portuguese varieties, using reviewed and fully characterized 8 STR markers

distributed through 5 chromosomes recommended by the International Organization for Vine and Wine and The European *Vitis* Database with genotypes of tetra-nucleotide markers (newly described STR markers retrieved from the *Vitis vinifera* whole genome sequence). We also performed genetic clustering analysis using a comprehensive collection of previously published grapevine genotypes, using Structure software.

Keywords: genetic identification, Grapevine, *Vitis vinifera*, Short Tandem Repeats (STR).

Resumo

A vinha (*Vitis vinifera* L.) é uma planta diploide com um genoma relativamente pequeno (475-500 Mb) organizado em 19 cromossomas. Esta planta é uma das mais antigas e valiosas culturas usadas para a produção de fruta, passas de uva, sumo e vinho. A nomenclatura das castas da videira resultou num complexo padrão de sinonímia e homonímia.

A identificação de castas é importante para os viticultores, produtores de vinho, autoridades reguladoras e consumidores. O aumento do interesse na identificação de castas resultou de uma necessidade de plantar e produzir uvas e produtos derivados corretamente identificados, devido ao valor económico em conformidade com designações de origem protegida. A correta identificação de castas da videira é crucial, deste modo é necessário desenvolver métodos eficientes de identificação da proveniência da casta e também durante o processo de produção de vinho.

Hoje em dia os métodos de identificação incluem ampelografia, métodos químicos e bioquímicos e métodos moleculares. Dentro dos métodos moleculares baseados em DNA, os microssatélites (STR) tornaram-se o marcador de DNA mais usado na identificação de castas da videira. Devido ao seu elevado poder discriminatório, ao encontrar um genótipo idêntico em duas diferentes plantas é uma forte evidência que essas plantas pertencem de facto à mesma casta. Por isso, a análise de STRs pode ser usada para confirmar ou determinar a identidade de castas comparando o genótipo obtido da amostra com genótipos de referência de castas alojadas em bases de dados. Contudo, a utilização de STRs di-nucleotídicos apresentam grandes desvantagens como *stuttering* e requerem protocolos corretos e fiáveis para a separação e identificação de alelos, de modo a evitar erros na leitura. Um método mais fiável passa pela utilização de STRs com repetições mais extensas (nomeadamente tetra-nucleotídicos).

Apontando para a utilização dos STRs tetra-nucleotídicos e uma padronização dos métodos de identificação na vinha, apresentamos uma revisão crítica dos STRs utilizados (em castas internacionais e portuguesas) na identificação de castas da videira, inseridas num contexto forense. Analisamos os marcadores STRs, comparando os genótipos de marcadores di-nucleotídicos (sequenciando os alelos em 12 castas Portuguesas, usando 8 marcadores de STRs revistos e caracterizados

distribuídos por 5 cromossomas, recomendados pela Organização Internacional da Vinha e do Vinho e pela Base de dados Europeia da *Vitis* com os genótipos de novos marcadores tetra-nucleotídicos obtidos através sequência genómica completa de *Vitis vinifera*. Foi feita ainda análise de *clustering* genético utilizando uma compilação extensa de genótipos da videira obtidos de estudos prévios usando o software STRUCTURE

Palavras-chave: *Vitis vinifera*, Vinha, identificação genética, Short Tandem Repeats

Contents

Agradecimentos	VII
Abstract	IX
Resumo.....	XI
Contents	XIII
Abbreviations	XV
Figures index	XVII
Tables index.....	XIX
1. INTRODUCTION	1
1.1. The domestic grapevine (<i>Vitis vinifera</i> L.)	1
1.2. Identification of grapevine cultivars	2
1.3. Grapevine Identification Methods	3
1.3.1. Ampelography	3
1.3.2. Chemical and biochemical methods	4
1.3.3. Molecular methods	5
1.4. Short Tandem Repeats	6
1.4.1. Development of STR markers in <i>Vitis</i>	7
1.4.2. The use of STR in <i>Vitis</i>	8
1.4.3. <i>Vitis</i> database	13
1.4.4. Recommended STR markers	15
1.4.5. Repeat structure of STRs	16
1.5. Other types of DNA markers and their utilization	19
1.5.1. Single Nucleotide Polymorphisms.....	19
1.5.2. Chloroplast markers	20
1.6. Portuguese grapevines.....	21
1.6.1. Studies on Portuguese grapevines	23
1.7. The forensic genetics approach	24
1.8. Aims.....	25
2. METHODOLOGY	27
2.1. Sampling and storage.....	27

2.2.	Selection of STR markers.....	28
2.2.1.	STR markers used in the European <i>Vitis</i> Database	28
2.2.2.	New long core-repeat STR markers.....	29
2.3.	Primer Design	30
2.4.	DNA Extraction	31
2.5.	Polymerase Chain Reaction.....	32
2.6.	Polyacrylamide gel electrophoresis with silver nitrate staining	33
2.7.	PCR product purification and sequencing reaction.....	34
2.8.	Genetic clustering analysis.....	35
3.	RESULTS AND DISCUSSION.....	37
3.1.	Selection of new tetra-nucleotide STR markers.....	37
3.2.	Primers	38
3.3.	Amplification and sequencing results	40
3.3.1.	Di-nucleotide and tetra-nucleotide STRs	40
3.4.	Structure analysis	43
3.5.	Recommendations for forensic genotyping of grapevine.....	49
4.	CONCLUSION AND FUTURE PERSPECTIVES.....	51
5.	REFERENCES	53

Abbreviations

A-Adenine

bp- base pairs

C- Cytosine

Cp- Chloroplast

cpDNA- Chloroplast genome

DNA- Desoxyribonucleic Acid

G- Guanine

INDEL- Insertion/Deletion

Mb- Mega bases

PCR- Polymerase Chain Reaction

RAPD- Random Amplified Polymorphic DNA

RFLP- Restriction Fragment Length Polymorphism

SNP- Single Nucleotide Polymorphism

SSR- Simple Sequence Repeats

STR- Short Tandem Repeats

T- Timine

Ta- Annealing Temperature

Figures index

Fig. 1 - Ampelographic characteristics of grapevine leaf. A) Diagram of mature leaf parts showing lobing delineated by the presence of sinuses, and the teeth at the leaf edges; B) Shoot tips/young leaves; C) Teeth shape on the mature leaf [11].	3
Fig. 2 - Electrophoretic pattern of true alleles (highlighted by blue dots) and stuttering peaks in a tetra-nucleotide STR (VChr5c) and a di-nucleotide STR (VVS2) currently used for grapevine fingerprinting [27].	16
Fig. 3 - Electropherogram of the cultivar Tinto Cão for locus VVMD32.	40
Fig. 4 - Electropherogram of the cultivar Souzao on locus VV16.	41
Fig. 5 - Electropherogram of the cultivar Tinto Cão on locus VV16.	41
Fig. 6 - Bar plotting of the results obtained with Structure software showing the genetic clustering of 535 grapevine genotypes when 6 populations are assumed in the dataset ($K = 6$). Each vertical line represents one individual, with color segmentation proportional to the Q coefficient of membership (Y axis) in the K clusters.	44
Fig. 7- Graphic output of L(k) using the Structure Harvest software. The horizontal axis represents K and the vertical axis represent L(k)	43
Fig. 8 - Graphic output of L(k) using the Structure Harvest software. The horizontal axis represents K and the vertical axis represents L(k).	45
Fig. 9- Bar plotting of the results obtained with Structure software showing the genetic clustering of 386 grapevine genotypes when 6 populations are assumed in the dataset ($K = 6$). Each vertical line represents one individual, with color segmentation proportional to the Q coefficient of membership (Y axis) in the K clusters.	48

Tables index

Table. 1- List of recommended STR markers by “The European <i>Vitis</i> Database” for grapevine cultivar analyses, repeat type, allele frequencies and bibliographic references.....	15
Table. 2- List of 19 STR markers with tetra-nucleotide repeats, distributed one per chromosome, and their discrimination power (PD) [27].	18
Table. 3- The most economically relevant grapevine cultivars utilized for wine production in Portugal, designated by the official name and synonym	22
Table. 4- Culivars, type, numbers and origin of grapevine cultivars sampled for this study.....	277
Table. 5- New tetra- nucleotide STR markers isolated in silico from the <i>Vitis</i> genome and chosen for characterization in this study.	377
Table. 6- Newly designed primers for previously described di-nucleotide STRs, their chromosomal location, annealing temperature, GC % and length in bp.	388
Table. 7- Design primers for the tetra-nucleotide markers obtain in silico, their annealing temperature, GC content and length.....	399
Table. 8- Repeat motif of four Grapevine cultivars on <i>locus</i> VV16.....	422

1. INTRODUCTION

1.1. The domestic grapevine (*Vitis vinifera* L.)

Vitis vinifera Linnaeus, commonly known as grapevine, belongs to the kingdom Plantae, division Magnoliophyta, class Magnoliopsida, order Vitales, family Vitaceae and genus *Vitis*. This species is a diploid plant that has a small genome size (475 - 500 Mb) consisting of 19 chromosomes.

The grapevine is one of the most valuable and oldest horticultural crops used to produce table fruit, raisins, juice and wine [2-4]. It is also the only remaining Mediterranean/ Western Asiatic representative of the *Vitus* genus and its domestication originated cultivars that are adapted to an extensive range of climates and tastes [5]. The grapevine is cultivated worldwide, and many cultivars important nowadays have been selected throughout the centuries. The great genetic diversity originated from the old practice of growing seedlings chosen by chance and breeding activity [2]. It is estimated that 5,000 to 6,000 cultivars of the *Vitis vinifera* exist worldwide, although only a relatively small number of these are used commercially [1].

In many regions cultivars were renamed leading to synonyms (different names for the same cultivar) as well as homonyms (different cultivars identified under the same name) [3].

1.2. Identification of grapevine cultivars

Usually only full-grown leaves and fruit are taken into consideration for morphological identification. Nevertheless, vine plant material is exchanged in the form of woody canes, and due to this fact cultivar identification becomes extremely difficult. When problems occur, the mistake is commonly noticed only after a vineyard is planted. Another problem arises from the fact that grafted rootstocks are not allowed to develop leaves in vineyards, and their genotypes have a major influence over the growth of the grafted scion and the quality of the grapes. For that reason, the selection of proper rootstock cultivars for diverse environmental conditions is a major economic factor in viticulture, and their misclassification could present economic disadvantages to the viticulturists [6].

The quality and market value of wine depends on grapevine variety, together with other factors such as ‘terroir’ (the combination of soil, climate, vineyard location and aspect), vineyard husbandry and winemaking technique [7, 8]. Therefore, wine is commonly marketed with labeling information regarding the grape variety used, the designation of the cultivation region, and the specific vintage year [9]. Depending on the wine production region, especially for wines with Controlled Designation of Origin (geographical), only a limited number of varieties are allowed and the inclusion of other varieties is only permitted under legally defined percentages [8]. According to wine labeling laws and international trade regulations, cultivars must be correctly identified on the bottle [8, 10]. The fact that different varieties may be used in wine production allows for fraudulent practices [10].

Taking into account these factors, accurate identification of grapevine cultivars is crucial and so is the importance of a good identification system throughout the winemaking process, enforcing wine regulations and protecting the grape growers, vineyard owners and winemakers from mistakes or fraudulent manufacturing practices that can have of severe commercial consequences [6, 8, 10].

1.3. Grapevine Identification Methods

1.3.1. Ampelography

From the Greek *ampelos*- grapevine and *graphos*- description, ampelography is a traditional identification method that relies on the visual inspection of several morphological vine organs [6, 11]. The main descriptors of the vine parts include the morphology of the vine (leaves, shoots, inflorescence, grape bunches and individual grapes), their phenology and other characteristics (budburst, flowering, the beginning of ripening, maturity, leaf fall, yield, grape and wine quality, resistance to disease and pests) [11, 12].

Ampelography can be easily applied to some grapevine cultivars with distinctive morphology and has been the most used tool in the characterization of grapevine germplasm as well as it occurs in most of the plant collections because it is a fast and cost effective method [13]. However, these descriptions are time consuming and prone to ambiguity in the cases of morphologically similar grapevines [14]. Ampelographic descriptions for a particular cultivar can slightly vary from one area to another depending on factors such as health status (including pest or disease load), vigor, interpretation of the observer (ampelography experts do not have complete knowledge on of thousands of different cultivars used worldwide), environmental, cultural, and genetic variations [6, 14]. Furthermore, juvenile plants are nearly impossible to identify because they do not exhibit the typical morphological traits of adult plants until 4 or 5 years of growth [11].

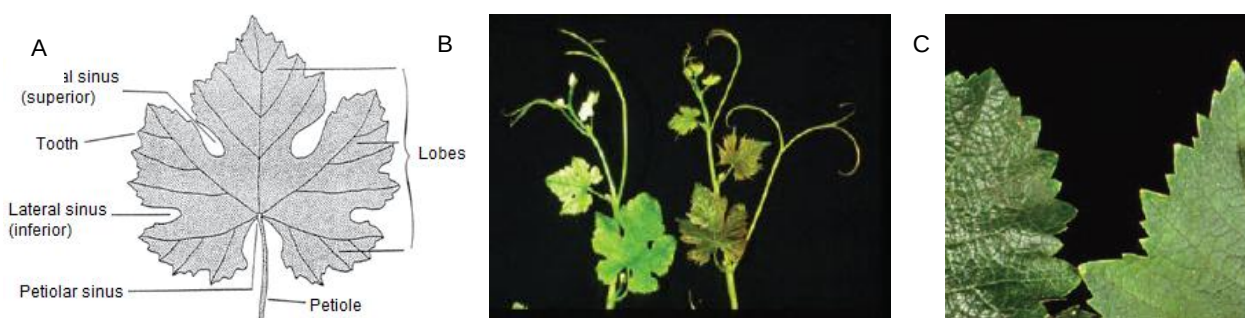


Fig. 1 - Ampelographic characteristics of grapevine leaf. A) Diagram of mature leaf parts showing lobing delineated by the presence of sinuses, and the teeth at the leaf edges; B) Shoot tips/young leaves; C) Teeth shape on the mature leaf [5].

1.3.2. Chemical and biochemical methods

Chemical and biochemical methods allow for the characterization of grapevine cultivars with regards to geographical differentiation. The principal methods used for authentication of geographical origin (production area) are the profiling of volatile compounds, amino-acids, minerals and analysis of stable isotopes and organic compounds. The principal methods used for the authentication of grape cultivars in wine are sensory analysis, mineral profile analysis, amino-acid and protein profile analysis, analysis of polyphenolic profile and analysis of volatile compounds profile. Must and wine characterization and differentiation have been mostly based on the analysis of these parameters, which can be used as proof of authenticity by comparison to a control sample [15]. However, these methods are time consuming and do not always give definitive results since they are influenced by a set of parameters such as soil composition, weather conditions, vinification methods and wine aging [1].

1.3.3. Molecular methods

The past decade has witnessed the introduction of molecular characterization, namely DNA technology, to the art of ampelography, helping to identify varieties, their origin and their parentage [6, 10, 12].

The first method described was based on isoenzyme analysis of grapevine cultivars through an enzyme system in order to detect possible synonyms or homonyms [6, 13]. However the expression of enzymes depends on the phenological stage of the plant and/or on environmental conditions [6].

After that, scientists began to develop molecular markers based on the analysis at the DNA level. DNA-based analyses do not have the same limitations as the previous method since plant DNA is identical in all cells of any tissue at any stage of development. Therefore it can be obtained from every plant tissue available and its characteristics are not influenced by environmental or sanitary conditions [6].

RFLP (Restriction Fragment Length Polymorphism) analysis was successfully employed to detect cultivar-specific DNA fingerprints for grapevine and rootstock cultivars. Compared with isoenzyme analysis, RFLP offered the advantages of robustness in various environments and higher levels of detectable polymorphisms. Nevertheless, it was time-consuming, required high quality DNA and the results had low reproducibility [6, 16].

RAPD (Random Amplified Polymorphic DNA) analysis was a cost-effective and faster method than RFLP. However, the greatest disadvantage is that different experimental conditions (e.g. thermal cycling equipment, Taq polymerases, DNA and primer concentrations) often led to different results between laboratories. Reproducibility of the results could be achieved by carefully performing standardized reaction conditions, but standardization of the RAPD procedure between laboratories is a difficult goal [6].

Because none of those methodologies met the requirements of an ideal identification system, the search for more promising genetic markers led to the development of STRs (Short Tandem Repeats), also known as Simple Sequence Repeats (SSRs) or microsatellites, for cultivar identification [6, 17, 18].

1.4. Short Tandem Repeats

The existence of repeated simple sequence motifs in plant nuclear DNA was demonstrated by Delseney et al in 1983 [22]. It was subsequently shown that STRs were abundant in most organisms, including plant and organelle genomes, and that these sequences represented a major source of genetic variation suitable for plant genetics [6]. Although the existence of STRs in plant nuclear DNA had been known for many decades, they did not become useful tools for grapevine genetics until the advent of Polymerase Chain Reaction (PCR) technology, which allowed them to be targeted as single genetic loci. STRs consist of tandem repeated DNA sequences with a core unit of 1-6 bp (base pairs) [6]. According to their repeat sequence structure they can be classified as: perfect STRs, if there is no interruption of the sequence by a base that does not belong to the repeat motif; complex STRs, if there is an interruption on the repeat motif by variable size sequences, and composed STRs, if there are two or more adjacent repeat motifs [21].

In eukaryotes, an estimated 10^4 to 10^5 STR loci are scattered randomly throughout the genome constituting an almost unlimited source of polymorphic sites that can be exploited as genetic markers [6]. They are selectively neutral, given that they are not sited inside or close to a coding sequence, where they may cause gene function disruption or affected by selection pressures on a gene in their vicinity [6].

STR markers have many advantages: a) high level of polymorphism (high level of variability in the number of repeats of the core motif occasionally presenting dozens of alleles at each locus); b) the co-dominant mode of inheritance (allowing for the discrimination of homozygotes and heterozygotes); [3, 6, 10, 11].

Currently, genotype information is given through profiles that are represented by allele sizes detected at the analyzed loci and expressed in base pairs.

1.4.1. Development of STR markers in *Vitis*

In order to identify grape cultivars using repetitive DNA sequences, including STRs, pioneering analysis were performed in the laboratories of CSIRO in Australia by Thomas in 1993. They demonstrated an abundance of STR sequences in grapevines that were suitable for identifying *Vitis vinifera* cultivars while primer sequences for these loci were conserved across other *Vitis* species [16]. Most importantly, STR alleles were inherited in a co-dominant Mendelian way, confirming their suitability for genetic mapping and their potential use for identification of genetic relatedness [23]. This study was followed by the development of additional grapevine STR markers around the world [17, 18, 24, 25]. The method was demanding on resources and time: genomic libraries were constructed from grapevine or rootstock cultivar, the library was screened with STR probes, the STR-containing clones were sequenced, PCR primers were designed from the sequences flanking the STR, and finally the PCR conditions were optimized to characterize the particular STR polymorphism.

STR markers have multiple uses in grapevine breeding and genetics, including cultivar identification and germplasm management, mapping of interest traits, and estimation of genetic diversity [26].

1.4.2. The use of STR in Vitis

1.4.2.1. Cultivar Identification

Grapevine genotyping is presently based on STR markers, which have a number of characteristics that makes them more suitable than any other molecular marker developed for DNA fingerprinting (e.g. RAPD and RFLP) [2, 3, 27]. One of the major applications of STR markers in grapevines is the identification and discrimination of cultivars in order to facilitate the management of collections and control the trade of plant material [28-34].

There have been reports of mistakes on cultivar identification throughout the years mostly due to synonymy and homonymy. Several studies, using STR markers together with ampelography, were made to solve some of these identification problems [18, 30, 35, 36].

1.4.2.2. Parentage and pedigree

Some of the modern grapevine cultivars are the result of the domestication of wild vines while others are the result of spontaneous crosses between wild vines and cultivars or crosses between two cultivars. The appearance of DNA profiling has served to clarify a long interest on the historical origins of current grapevine cultivars. Identifying the parentage of all modern grapevines is possible if the parents still exist in cultivation or collections [17, 37].

However, previous information of the parents is a rare event in grapevine genetics, and generally there is no information on the chronological order of appearance of the different cultivars.

1.4.2.3. Genetic diversity

Sexual reproduction, vegetative propagation and somatic mutations contribute significantly to the development of cultivated grapevines. New genotypes are produced by sexual reproduction, either by crossing or self-fertilization. Because individual grapevine plants have highly heterozygous genotypes, any progeny produced from seed is a novel combination of parental alleles, resulting in phenotypic variation and the segregation of traits in the progeny. Once a desirable trait is identified, vegetative propagation (asexual) by cuttings is a method of maintaining and multiplying a genotype so that a whole vineyard can be planted with a single clone from a cultivar. Cuttings are also a convenient method of transporting cultivars from one region to another and cultivars grown today are maintained by vegetative propagation. Although clonal propagation should ensure that all plants grown from cuttings have the same genotype, the occurrence of somatic mutations in individual cuttings may eventually lead to plants of the same cultivar having a slightly different genotype and sometimes a different phenotype, referred to as clonal variation [38]. Several studies about genetic diversity in grapevine cultivars have been developed so far, and their results provided insights about the levels of distribution of genetic diversity in the existing resources of *Vitis*, and into genetic subdivision within the European germplasm. Also they demonstrated the ineffectiveness of some STRs when genetic diversity within clone collections is addressed [5, 39-42].

1.4.2.4. Geographical Origin

In 2000, Sefc et al. [43] suggested that according to cultivars genotypes they could be connected to their regions of origin. Due to the predominance of certain alleles, or the occurrence of null alleles in some populations, the information content of a given marker may vary between the cultivars from different regions. Moreover, intra-cultivar variability can result from epigenetic modifications in response to the environment or from the presence of phytopathological agents such as virus and viroids [44].

In 2012, Meneghetti et al. [45] developed a strategy that permitted to identify different biotypes, accessions, and clones among the intra-varietal genetic variability, and to correlate the genetic differences to their geographical origins or morphological traits.

1.4.2.5. Linkage mapping

The development of a genetic linkage map based on a skeleton of transferable markers is the first step required for finding the chromosomal position of genes for traits of interest, and exploring the genetics underlying the observed phenotypic variation in natural germplasm and breeding lines. A large amount of effort in developing this type of map is intended to identifying valuable polymorphic markers, that can be used in other pedigrees and related *taxa*. Several STR markers were identified in grapevine over the past decade and were incorporated into several genetic maps [46-49].

1.4.3. *Vitis* database

1.4.3.1. The history of *Vitis* databases

In 1983 the inventory of the world-wide existing *Vitis* species varieties and genotypes grown in grapevine collections was started by the Geilweilerhof Institute for Grapevine Breeding. Initial support was obtained from the International Plant Genetic Resources Institute (IPGRI) and the International Organization of Vine and Wine (OIV) resulting in the *Vitis* International Variety Catalogue (VIVC), which provided an inventory of the grapevine genetic resources and became accessible via the internet in 1996 (<http://www.genres.de/idb/vitis/vitis.htm>).

In 1994, a database of grapevine genetic identity profiles at 6 nuclear STR loci was publicly presented for more than 200 cultivars, with the initial aim to offer this resource to the research community [23]. The database included complete information about the plant used as the source of the cultivar DNA profile with all plants sourced from national or state germplasm collections. Lately after the publication, commercial interest in using the database as the basis of a commercial grapevine DNA identification service resulted in restricted access. A similar approach inspired a second database service, under the appellation of “The Greek *Vitis* database” (<http://gvd.biology.uoc.gr/gvd/index.htm>). This database is aimed not only to make the genetic information about Greek cultivars widely available but also and especially to facilitate germplasm management of the Greek grapevine resources disseminated in several collections.

In 2002, one of the objectives of the EU-project GENRES081 entitled “European network for grapevine genetic resources conservation and characterization”, was to establish a European *Vitis* Database (ECVD) that became available at <http://www.genres.de/eccdb/vitis/>.

In the meanwhile, a searchable catalogue of *ex situ* collections in Europe, called European Internet Search Catalogue (EURISCO) (<http://eurisco.ecpgr.org>) was created (<http://www.ecpgr.cgiar.org/EPGRIS/Index.htm>). EURISCO is based on reviewed Food and Agriculture Organization of the United Nations/International Plant Genetic Resources Institute - Multicrop Passport Descriptors (FAO/IPGRI-MCPD) which are acknowledged by international genetic resources databases and are promoted to be used worldwide.

In 2003, the FAO/IPGRI-MCPD was adopted for the European *Vitis* database (EVDB). The EVDB is an accession linked database; each accession is identified by its accession number which is crucial due to the high number of misnamed, synonymous or homonymous grapevine varieties. Thus, every record was assigned to the corresponding accession from which the information was taken. It recommended including at least six STR loci (designated VVS2, VVMD5, VVMD7, VVMD27, ssrVrZAG62 and ssrVrZAG79) that could enable immediate comparison with the variety identification data obtained by the GENRES 081 project [50, 51].

1.4.4. Recommended STR markers

Although a considerable amount of STR data has been published for grape, inter-laboratory variations have made the comparisons of results difficult. The characterization of grapevine cultivars by STRs should work universal and independently from analysis systems and laboratory equipment, i.e. each analysis of identical DNA samples should produce identical allelic profiles.

In 2004, This et al. [11] proposed to compare different methods of STR-based profiling for reproducibility among the study participants, and to standardize allele scoring by defining reference alleles. Although some participants obtained different absolute allele sizes and different relative allele sizes, a set of coded alleles based on well-known reference cultivars and proposed six STR markers (VVS2, VVMD5, VVMD7, VVMD27, ssrVrZAG62 and ssrVrZAG79) was developed to be adopted as a minimal set of standard marker for grapevine cultivar analyses. This set of markers has been widely used in subsequent studies (Table 1).

Table. 1- List of recommended STR markers by “The European *Vitis* Database” for grapevine cultivar analyses, repeat type, allele frequencies and bibliographic references.

Marker	Type of repeat	Allele frequencies	Reference
VVS2	Di-nucleotide	Yes	[5, 11, 14, 23-25, 32-37, 39, 42, 45, 48, 52-72]
VVMD5	Di-nucleotide	Yes	[5, 11, 14, 17, 18, 25, 32-37, 39, 42, 44, 45, 48, 52-72]
VVMD7	Di-nucleotide	Yes	[5, 11, 14, 17, 18, 25, 32-37, 39, 42, 44, 45, 48, 52-72]
VVMD27	Di-nucleotide	Yes	[5, 11, 14, 17, 18, 33, 34, 36, 37, 39, 45, 48, 54-66, 72]
ssrVrZAG62	Di-nucleotide	Yes	[5, 11, 14, 25, 32-37, 39, 42, 45, 48, 54-66, 69-72]
SsrVrZAG79	Di-nucleotide	Yes	[5, 11, 14, 25, 32-37, 39, 42, 45, 48, 54-66, 69-72]

1.4.5. Repeat structure of STRs

STRs have a high polymorphism and exhibit a very high mutation rate [73]. Their mutation rate depends of their intrinsic characteristics such as the number of repeat units, the size of consensus motifs and the repetitive motif in itself. Several studies indicate that with increased size of the repeated structure (tri-, tetra-, penta-nucleotides) increases the likelihood occurrence of punctual mutations, dividing the initial motif in other structurally smaller and with lower mutation rates [74-76]. The highly polymorphic nature of STR can be explained by an inherent instability since these regions have a higher mutation rate (gain or loss of one or more repetition units) than other regions of the DNA.

Di-nucleotide markers need extremely precise and reliable protocols for allele separation and identification to avoid allele miscalling. This problem also derives from the frequent addition of one adenine nucleotide by DNA polymerase during PCR amplification, which produces 'artifact' alleles very close in size and complex to differentiate. Another problem in using di-nucleotides is the high amount of stuttering that they originate during PCR, which makes the interpretation of electropherograms and the call of true alleles extremely difficult (Fig. 2) [27]. These problems can be partially solved with the use of STR markers with longer repeat length [3].

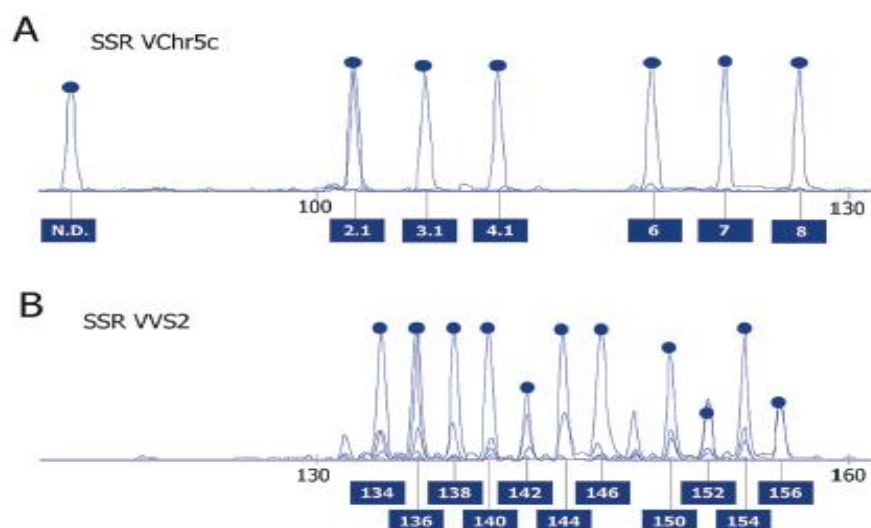


Fig. 2 - Electrophoretic pattern of true alleles (highlighted by blue dots) and stuttering peaks in a tetra-nucleotide STR (VChr5c) and a di-nucleotide STR (VVS2) currently used for grapevine fingerprinting [27].

Most of discrepancies between laboratories in scoring di-nucleotide alleles are due to arbitrary decisions in allele binning, which is the process that converts raw allele lengths into allele classes usually expressed by integer numbers [31]. In humans, long nucleotide repeats (tetra and penta-nucleotide repeats) were adopted early on genetic identification. Still, it is imperative to take into account the fact that different analytical systems might produce different allele sizes and as a result different bins, increasing the difficulty of comparing genotype tables produced by different laboratories [27].

Numerous electrophoresis procedures can be used to separate STR alleles. The currently accepted method for human DNA in forensic disputes is based on PCR with dye-labeled primers; PCR products are analyzed by capillary electrophoresis in automatic sequencers and alleles sized with reference allelic ladders constructed for each *locus*. Resembling protocols are being developed for animals, such as domestic dogs, but in plants such a robust and reliable procedure is rare. More simple and less expensive protocols are frequently used, including the use of manually cast gels, the detection of DNA fragments by silver staining, and the estimate of allele size by comparison with anonymous ladders or plasmid sequences loaded on the gel in the adjacent lanes. Taking into account these factors, the grapevine community has adopted allelic ladders for a limited number of markers [13, 27, 30].

In 2008, Cipriani et al. [27] presented a list of 38 new STR markers (tri-,tetra- and penta-nucleotides). From each chromosome was selected one marker with the highest discrimination power, thereby a list of 19 STR markers with long core repeats was chosen for grapevine genotyping (Table 2).

Table. 2- List of 19 STR markers with tetra-nucleotide repeats, distributed one per chromosome, and their power of discrimination (PD) [27].

Marker designation	Forward primer (3'-5')	Reverse primer (3'-5')	Length of the sequenced allele (bp)	Motif	PD
VChr1b	AGATGGGTGGCATTAGCAA G	TTATTTCCCTCCCTCGCTGT	112	ATCC	0.821
VChr2b	CCTCCTGCGAACAAGTCTG T	GTTGCTGGATTTGTGGAAGG	123	AGCT	0.714
VChr3a	CAATCATATGAGCAAGGCA TGT	GCTTCCTGAAATTTGTGTCCA	199	AAT	0.933
VChr4a	CAACTGGGATCCAAGACCT C	CAGCTTCACAGGTAACCACA	197	AAAG	0.787
VChr5b	CTTCTCGGTCATGGTCATT G	CTCCTTCCACCTCTGGTTCA	198	AAAG	0.899
VChr6a	AATGTTGAGCTTTGGGCTT G	CCAATTCTTCCATACCTCAAAA	184	AATC	0.733
VChr7b	AAAGGGCCTAAACTCTTAAT AACTTG	TGCTTTATAGACACTAACCCAC AAA	188	ACAT	0.866
VChr8b	TGTGTGATGTTTTGTCGATG G	TGAACCAAGTTCTAATTTACATT TCC	142	AAG	0.956
VChr9a	GCGACAGCATCACTTCAAT C	GAATTGCCAAGGACAAGGAG	114	AAG	0.925
VChr10b	CCATGTCCAACCGAAACAA C	CAGAAATCTCGTGTGCTCA	140	AAC	0.798
VChr11b	TGAGTTGAGCTATTGGCTTT GA	AGCAACTCTGTCCATCCATGT	163	AGAT	0.874
VChr12a	GCTTTAAATGTTAGATTAGG GCACTC	TCCATGTTGTTTGCTCTTTCC	136	AATT	0.814
VChr13a	TGGCAGAGCAAATGAATCA A	TTGGATGGATTGGAATGACC	155	AAAAG	0.855
VChr14b	CAATTGAACACTTACACTCA CAATCA	TGTGACTAAAGGTTATTAGCAG GA	195	ATC	0.834
VChr15b	GGGTCCAATTCCTTTTGTT	CGAAAGACTCAATTGCCACA	124	AAT	0.898
VChr16a	TTCATGTGTGACACCCCTTT	AATGTCCATGCTTCAAAATACC	162	AAAT	0.796
VChr17a	AGGAAGAGGATTGATCACC A	GTGCCAACCCTTGCACTATT	187	AACC	0.585
VChr18a	TTCCCACCCGGTAAATATG A	CATCCAAACATCACGCTGAG	167	AAGG	0.852
VChr19a	TGGATTCACCATGTCTCA	CGAGGATACCAACAAGAATGAA	143	AAG	0.902

1.5. Other types of DNA markers and their utilization

1.5.1. Single Nucleotide Polymorphisms

Sequencing projects have created a large quantity of sequence information and nucleotide polymorphisms. These belong to two basic types: Single Nucleotide Polymorphisms (SNPs) and Insertions-Deletions of different lengths (INDELs). They are the most common class for detection of the smallest unit of genetic variation among individuals within a species and are usually bi-allelic variations between individuals that occur in genes (promoters, exons or introns) or between genes (intragenic) [77]. Although the polymorphism information content for SNP is lower than for STR markers, many SNP markers can be easily used when required. SNPs are highly reproducible among different laboratories and detection techniques since the alleles are not discriminated according to their size but according to the nucleotide present at a given position. Due to these characteristics and their unlimited availability, SNP are becoming the chosen markers for the development of identification panels in many animal and plant species [3]. In recent years there have been an increasing number of studies for grapevine identification using SNP markers [3, 77, 78].

1.5.2. Chloroplast markers

The arrangement of the circular chloroplast genome (cpDNA) is extremely conserved with genes generally occurring in the same order. With few exceptions, cpDNA is divided in two single regions: the large single copy (LSC) and the small single copy (SSC), which are separated by an inverted repeat.

The chloroplast genome has a lower evolutionary rate than the nuclear genome, and given the maternal inheritance it can only be disseminated by seeds or cuttings. For this reasons cpSTRs make an excellent tool to investigate geographical origin [33, 79].

1.6. Portuguese grapevines

Due to its mild climate, Portugal is by excellence a wine-producing country. Portugal is very rich in grapevine biodiversity, and when compared with other western countries with a long history in grapevine cultivation it is considered the Mediterranean country with the highest density of autochthonous grape cultivars [5, 80]. The grapevine has a significant part as a border culture and extensive crop in the different Portuguese agro-ecosystems.

The Portuguese National Ampelographic Collection records approximately 450 varieties, most of which are considered autochthonous [71]. Portugal has a high number of cultivars and due to their dissemination all over the country some cultivars assume different names according to the region where they are grown, resulting in synonyms and homonyms, a problem for viticulture and for germplasm management [14]. Despite the high number of cultivars currently recognized, many of them are hardly used or face the risk of extinction. In fact, although 341 varieties are officially authorized for wine production, only 51 are economical relevant [5, 14]. Approximately 15 native cultivars represent the majority of those utilized for viticulture. Non-autochthonous cultivars are also of great importance in viticulture in Portugal (Table 3) [14].

Table. 3- The most economical relevant grapevine cultivars utilized for wine production in Portugal, designated by the official name and synonym, when applicable [4, 6, 14, 79].

Cultivar designation	Berry color	Origin	Synonyms
Alvarinho	Green yellow	Minho (Portugal)	-
Antão Vaz	Green yellow	Alentejo (Portugal)	-
Aragonez	Black blue	Galiza (Spain)	Tinta Roriz (Portugal)
Arinto	Green yellow	Minho (Portugal)	-
Baga	Black blue	Bairrada (Portugal)	-
Castelão	Black blue	Alentejo (Portugal)	-
Fernão Pires	Green yellow	Bairrada (Portugal)	-
Loureiro	Green yellow	Minho (Portugal)	-
Moscatel Graúdo	Green yellow	Spain	Moscatel de Setúbal (Portugal)
Síria	Green yellow	Alentejo (Portugal)	-
Tália	Green yellow	Tuscany (Italy)	-
Tinta Barroca	Black blue	Douro and Dão (Portugal)	-
Tinto Cão	Black blue	Portugal	-
Tinta Miúda	Black blue	Spain	-
Touriga Franca	Black blue	Douro and Dão (Portugal)	-
Touriga Nacional	Black blue	Dão (Portugal)	-
Trincadeira	Black blue	Portugal (found in the oldest vineyards)	-
Vinhão	Black blue	North of Portugal	Souson (Spain)
Viosinho	Green yellow	Portugal	-

1.6.1. Studies on Portuguese grapevines

STR markers have been used to improve the management of the Portuguese Grapevine National Collection (verifying synonyms and homonyms), and understanding the genetic relatedness of the cultivars used for wine making in Portugal (identifying and discriminating) [5, 14, 35, 71].

Other studies using cpSTRs have been developed to assess cpDNA diversity in Portuguese grapevines, and to examine genetic relations between Portuguese grapevine cultivars and wild vines [12, 80, 81].

Due to the fact that all Portuguese grapevine identification studies only use dinucleotide STR markers, comparison of results with other datasets is not straightforward.

1.7. The forensic genetics approach

The need to correctly identify different grapevine cultivars has stimulated the development of DNA technology for more accurate discrimination methods [10].

In the early 1990s, initial research using repetitive DNA fingerprinting technology by CSIRO laboratories in Australia showed its potential as an objective technique to identify grapevine cultivars [10, 24].

Individual fingerprinting with molecular markers based on DNA has grown to be the most utilized for population genetics studies and analysis of genetic diversity in germplasm collections, including the solution of synonymy/homonymy, paternity and kinship analysis [2, 7]. DNA-based identification of grapevine varieties is similar to the DNA profiling in humans long used in forensic criminology and to investigate kinship among individuals [27]. Due to plant DNA characteristics (identical in each cells of all tissues at any stage of development), DNA-based identification of grapevines is theoretically more objective than ampelographic and chemical methods, and can ascertain the grape variety parentage as it can prove paternity in humans [7]. This identification method is particularly suitable for grapevines because they are not propagated by seeds but vegetatively by cuttings or buds, resulting in genetic identity for all clones within a particular variety [7].

The benefit of DNA technology is that it may overcome the problems of morphological variability and, given a sufficient and verified database, may aid in the identification of unknown cultivars. Ultimately, DNA-based identification may offer an objective method for characterizing cultivars that can be adopted and shared by laboratories around the world [7, 18].

1.8. Aims

We propose to perform a critical review of grapevine cultivar identification based in STRs in the context of forensic genetics. We focused on STR markers utilized in international and Portuguese cultivars and discuss current trends and recommendations for achieving robust standardized genotyping in the *Vitis vinifera*.

We also assessed the potential of genetic clustering to investigate the origin and identification using Structure analysis on a comprehensive dataset of grapevine genotypes.

2.METHODOLOGY

2.1. Sampling and storage

Eight Portuguese grapevine cultivars (Mourisco, Tinta Barroca, Moscatel de Hamburgo, Tinta Roriz, Touriga Nacional, Touriga Franca, Sousão, Tinto Cão, Alvarinho, Fernão Pires, Loureiro and Viosinho) were obtained *in situ* from several grape producers in June 2012 (Table 4).

The leaves were stored at -20 °C in order to maintain the integrity of the tissues until DNA extraction.

Table. 4- Cultivars, type, numbers and origin of grapevine varieties sampled for this study.

Cultivars	Number of Samples	Origin
	Leaf	
Mourisco	1	Covas do Douro
Tinta Barroca	3	Covas do Douro
Moscatel de Hamburgo	2	Covas do Douro
Tinta Roriz	3	Covas do Douro
Touriga Nacional	3	Covas do Douro
Touriga Franca	3	Covas do Douro
Sousão	3	Covas do Douro
Tinto Cão	3	Covas do Douro
Alvarinho	3	Amares
Fernão Pires	3	Amares
Loureiro	3	Amares
Viosinho	3	Provesende
Total	33	

2.2. Selection of STR markers

2.2.1. STR markers used in the European *Vitis* Database

A set of 8 STR markers distributed through 5 chromosomes and recommended by the International Organization for Vine and Wine [82] and utilized in The European *Vitis* Database [83] was selected for the survey in Portuguese grapevine cultivars, aiming at the characterization of the loci at the sequence level. New primers were designed in order to amplify and sequence a longer fragment than the original recommended loci. This had the objective to allow the sequencing of the primer annealing region of the recommended loci to assess the presence of polymorphisms.

2.2.2. New long core-repeat STR markers

The *Vitis vinifera* Genome Database [84] allowed the retrieval of the complete sequence for each chromosome. *In silico* isolation of new potential tetra-nucleotide STRs was performed using Tandem Repeats Finder software in order to isolate repeated sequences in each chromosome at a time [85]. The entire sequence of each chromosome was inserted on Tandem Repeats Finder, and the program's parameters were set to a Minimum Alignment Score, report repeat of 80, Maximum Period Size of 4, and choosing the option of showing the Flanking sequence.

2.3. Primer Design

Regarding the previously described markers, the genetic coordinates were retrieved using the NCBI database [86]. The sequences for each marker were compared to the *Vitis vinifera* Genome Database [84] for confirmation of the sequence and chromosomal location. The sequence search was extended 2000 bp to allow primer re-design. After selecting potential STR markers, primer design was based on the following criteria: avoiding polymorphisms and poly-A or poly-C type repeated segments in the flanking regions, annealing temperature between 58°C and 60°C, primer size between 18 and 22 bp and a GC content exceeding 45%.

All primers were tested for primer dimer and hairpin formation using Auto Dimer v.1.0 software [24, 87].

2.4. DNA Extraction

DNA was extracted according to an established CTAB-based protocol [88] with in-house adaptations. Plant sample was submerged on 500 μ L CTAB buffer and incubated at 60°C for one hour.

Samples were then transferred to phase lock gel tubes with 575 μ L phenol making up the final volume with deionized water (DNase-, RNase- and Proteinase-free). Samples were centrifuged at 14,000 g for 3 minutes to separate the aqueous phase (DNA) and the lower phenol phase (degraded proteins, lipids and secondary compounds). The content of the tubes was transferred to new phase lock gel tubes with chloroform-isoamyl alcohol, and centrifuged at 14,000 g for 3 minutes, after which was transferred to new tubes with 1 mL ethanol (96%) and refrigerated at -20°C for 30 minutes. Samples were centrifuged at 14,000 g for 15 minutes at 4°C, transferred to new tubes with 1 mL ethanol (70%) and centrifuged again at 14,000 g for 5 minutes. The content of the tubes was discarded and the samples were air dried at room temperature. DNA was resuspended in 100 μ L deionized water and stored at -20°C until used.

2.5. Polymerase Chain Reaction

In order to test the efficiency of the DNA extraction, a control PCR amplification was performed on the chloroplast *trnL* (UAA) gene, using the forward primer 5'CGAAATCGGTAGACGCTACG3' and the reverse primer 5'GTTCAAGTCCCTCTATCCCC3'. The genomic location of the amplified fragment in *Nicotiana tabacum* cpDNA reference sequence (NC_001879.2) is 49312-49888 [89].

For both the chloroplast *trnL* (UAA) gene and STR loci, PCR reactions were carried out with a final volume of 10 µL, combining 5 µL PCR kit [MyTaq HS Mix 2x (BIOLINE) or Fermentas PCR Master Mix 2x], 2 µM of each primer, 1 µL of DNA and 2 µL of DNase-, RNase- and Protease- free water (5 Prime).

Amplifications were carried out in GeneAmp® PCR System 2700 (Applied Biosystems) thermal cycler or a 2720 thermal cycler (Applied Biosystems), under the following conditions for MyTaq HS Mix 2x: initial denaturation at 95°C for 10 minutes followed by 35 cycles at 95°C for 30 seconds, primer annealing temperature for 60 seconds and extension for 72°C for 90 seconds, and final extension at 72°C for 5 minutes. For Fermentas PCR Master Mix, thermal cycling was as follows: initial denaturation at 95°C for 3 minutes followed by 35 cycles at 95°C for 30 seconds, primer annealing temperature for 30 seconds and extension for 72°C for 60 seconds, and final extension at 72°C for 15 minutes.

2.6. Polyacrylamide gel electrophoresis with silver nitrate staining

PCR amplification of DNA samples was confirmed by polyacrylamide gel electrophoresis. Electrophoresis gel was composed of 5 ml acrylamide solution (9%), 510 μ L APS (2.5%) and 21 μ L TEMED. Hydrophilic film (GelBond Film®) was placed on glass supports in order to obtain a polymerized gel with 1 mm thickness. Electrophoresis buffer consisted of 0.125 M Tris/Glicine (pH = 8.8) solution used to soak strips of Whatman® paper, placed over the gel at the anode and cathode. Bromophenol blue dye was added to the buffer placed on the cathode buffer to control the progress of the run. Electrophoresis was performed horizontally in a Multiphor plate (Pharmacia) at a constant temperature of 4°C, voltage rate of 300 V and an amperage rate of 80 A for one hour.

In order to visualize the amplified fragments the electrophoresis gel was stained according to the silver nitrate method as follows: DNA fixing by the gel immersion in ethanol (10%) for 10 minutes, followed by immersion in nitric acid (1%) for 5 minutes, rinsing with deionized water twice for 10 seconds, immersion in silver nitrate (0.2%) with stirring for 20 minutes protected from light, rinsing with deionized water twice for 10 seconds, DNA revelation was performed with a sodium nitrate solution (0.28 M) and formaldehyde (0.02%), fixed by immersion in acetic acid (10%) for 30 seconds, rinsing with deionized water and drying at room temperature.

2.7. PCR product purification and sequencing reaction

PCR products were purified by digestion of excess primers and unincorporated dNTPs using 1 µL of ExoSAP-IT® (USB® Corporation, USA) for 1.5 µL of each PCR product. The reaction was held at 37°C for 15 minutes, followed by enzymatic denaturation at 85°C for 15 minutes.

Sequencing reactions were performed in a final volume of 5 µL, combining 2.5 µL of purified PCR product, 0.5 µM primer and 2 µL Kit BigDye® Terminator v1.1 (Applied Biosystems) diluted to 50% with Sequencing buffer 5X (Applied Biosystems) as follows: initial denaturing cycle at 96°C for 2 minutes followed by 35 cycles at 96°C for 15 seconds, 50°C for 9 seconds, and 60°C for 2 minutes. Sequencing reactions were performed in a thermal cycler GeneAmp® PCR System 2700 (Applied Biosystems).

Sequencing reaction products were purified using Sephadex™ G-50 Fine DNA Grade columns (GE Healthcare, United Kingdom) and resuspended in 12 µL formamide [Hi-Di Formamide (Appllichem, Germany)].

Electrophoretic separation of the sequencing reaction products was performed on a Genetic Analyzer 3130xl (Applied Biosystems) with 36 cm capillaries using POP-7 polymer (Applied Biosystems). The injection of 12 µL sample was at 60°C for 15 seconds.

2.8. Genetic clustering analysis

A comprehensive dataset of grapevine genotypes was compiled from the literature in order to perform genetic clustering analyses. The dataset was based on 10 different studies [11, 14, 34, 35, 45, 59, 69-72] comprising 535 grapevine cultivars for 6 STR loci (VVS2, VVMD5, VVMD7, VVMD27, VrZAG62 and VrZAG79). The different datasets were converted to the allele sizes estimated by This et al. 2004, using the grapevine cultivars common to this study. These conversions assumed that all individual samples of the same cultivars were clones, therefore, having the same genotype, moreover we verified if the differences between allele pairs at a given *locus* in one study matched the same pair for the same sample in other study.

The software Structure available at <http://pritchardlab.stanford.edu/structure.html> [90] implements a model-based cluster algorithm to STR data for inferring population structure (i.e. to identify genetically distinct subpopulations on the basis of patterns of allele frequencies). The algorithm was used using the following settings: 'no prior population information, admixture model and K = 1 to K = 20 to test for population structure among the dataset of genotypes.

Structure Harvester software available at <http://taylor0.biology.ucla.edu/structureHarvester/> was used to plot the 'Ln Probe of Data' values in order to allow for choosing the value of K (number of inferred cluster in the dataset) that that captured most of the structure in the dataset.

3.RESULTS AND DISCUSSION

3.1. Selection of new tetra-nucleotide STR markers

The markers obtained by the Tandem Repeats Finder that presented a Period Size and a Consensus Size of four, a Copy Number between 16 and 20, an identity percentage superior to 90%, and did not present poly-A or poly-C repeats on the flanking sequence were taken into account. Thus, four STRs were selected (Table 5).

Table. 5- New tetra- nucleotide STR markers isolated *in silico* from the *Vitis* genome and chosen for characterization in this study.

Chromosome	Genetic coordinates	Repeat motif in the reference sequence
2	6121223-6122282	(CATA) ₁₅
8	521914-522995	AA(TAAA) ₁₇ TAAG(TAAA) ₂
10	11419033-11420131	ATA(CATA) ₂ GGTA(CATA) ₃ GGTA(CATA) ₂ GGTA(CATA) ₁₄
16	3187435-3188506	(ATTT) ₁₈

3.2. Primers

New primers for the 8 markers recommended by the International Organization for Vine and Wine [91] selected for review in this study were designed according to the criteria before mentioned (section 2.3 Primer Design). The objective was to survey by sequencing not only the repeated region but also the primer annealing region of the recommended fragment.

Alterations were made to the previously published primers so that annealing temperatures were similar for each pair, the G/C content were closer to 45% and size in between 20-24 bp (Table 6).

Table. 6- Newly designed primers for previously described di-nucleotide STRs, their chromosomal location, annealing temperature, GC % and length in bp.

Primer	Chromosome	Forward primer	Ta (°C)	% GC	Length (bp)	Reverse primer	Ta (°C)	% GC	Length (bp)
VVS2	11	GACGTAGAAGTA GAAGAATAGTC	60	39	23	GCAAGCATCCAT ATGCTTTGAG	60	45	22
VVMD5	16	GGGCTTGGAGA AAGGTAAGAG	60	52	21	CTCCACTTAATG GGACTACAT G	60	45	22
VVMD7	7	GATCAACAGGTT CTGATCGGA	58	48	21	CGAACTTACCAT AGTTTTTATAGC	58	33	24
VVMD25	11	GGGAGGGTCAG AAGTATACTC	60	52	21	CTACAGCTCCAT CTCCAAC TG	60	52	21
VVMD27	5	GGAATGAGAAAA TGCCGGACG	60	52	21	GTACCAGATCTG AATACATCCG	60	45	22
VVMD32	4	GGAAAGATGGG ATGACTCGC	58	55	20	TTTCACGTACGG CCGGTGG	58	63	19
VrZAG 62	7	CCATGTCTCTCC TCAGCTTCT	60	52	21	TCTTTTCTTGGG CTCCACTGC	60	52	21
VrZAG 79	5	CTTTGCACACCT CATCTGTGG	60	52	21	TGCCCCCATTTT CAAAC TCCC	60	52	21

For the new *in silico* tetra-nucleotide markers, it was not possible for all primers to have the characteristics mentioned before, due to the presence of poly-A and poly-T stretches on the loci's flanking sequences (Table 7).

Table. 7- Design primers for the tetra-nucleotide markers obtain *in silico*, their annealing temperature, GC content and length.

Primer	Forward primer	Ta (°C)	% GC	Length (bp)	Reverse primer	Ta (°C)	% GC	Length (bp)
VV 2	GATGGATGTCCAA GACTTCTTT	58	40	22	AATTTCCTAACTTTAAT GCACATAAA	58	27	23
VV 8	GTGAATAAGTAAAT GAATAAACT	54	21	24	TCTTTTCTAAGCCACC TTGG	54	45	20
VV 10	CATATGTTGAGGT GCATCTTCTA	60	39	23	AACACACCTCAAATTT AGGGGG	60	45	22
VV 16	GAAAATAAGGTTTT AATTGATTAAC	56	20	25	AATTAAGCCTTAGTCA ACAAGC	56	36	22

3.3. Amplification and sequencing results

3.3.1. Di-nucleotide and tetra-nucleotide STRs

The primers of the chloroplast *trnL* (UAA) gene used as control allowed positive amplification proving adequate DNA extraction for all samples.

The di-nucleotide STRs provided good amplification results, with exceptions of marker VVMD25 and VrZAG62 for which approximately 60% of PCR amplifications failed. The amplification results for the *loci* VV2, VV8 and VV10 were inconsistent as approximately 50% of PCR failed. This could be due to polymorphisms on the annealing region of the primers. This is a common cause for apparent deviations from the Hardy-Weinberg equilibrium at STR *loci* and may interfere with correct identification. Consistent PCR amplification was obtained for locus VV16.

The aim was the characterization of recommended di-nucleotide *loci* at the sequence level. However, the attempt for this characterization based on the number of repeats failed due to the amount of stuttering (artifact sequences) produced in the PCR reaction. The electropherogram of the cultivar Tinto Cão on *locus* VVMD32 clearly shows ambiguity in the sequence with regards to the number of repeats, a common event when di-nucleotide markers are surveyed (Fig.3). The sample had shown to be homozygous in the acrilamye gel, therefore was chosen for sequencing. This ambiguity in reading the correct number of repeats in the electropherogram is caused by superimposition of the real sequence with artifacts.

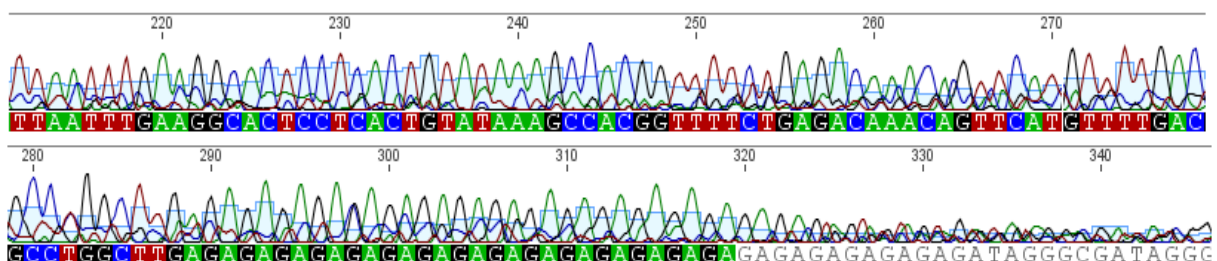


Fig. 3 - Electropherogram of the cultivar Tinto Cão for *locus* VVMD32

The new tetra-nucleotide *locus* VV16 was sequenced for four homozygous grapevine cultivars and the electropherograms for the cultivars Souzão and Tinto Cão are presented below (Fig.4 and Fig.5). They do not show ambiguity in the electropherogram caused by PCR artifacts and the number of repeats is easily assessed, making its interpretation straightforward, in contrast with the electropherograms obtained for the di-nucleotide *locus* VVMD32 (Fig. 3).

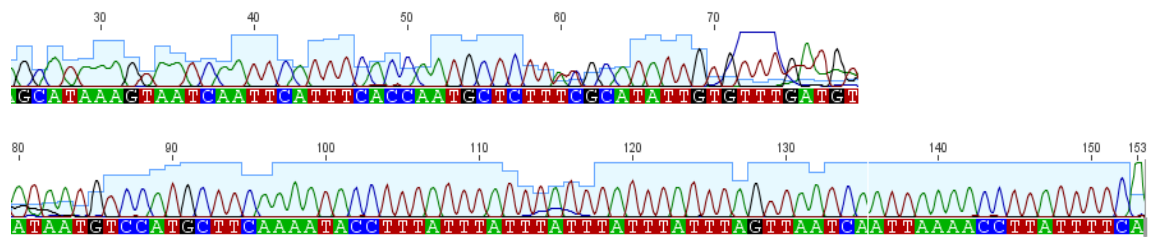


Fig. 4 - Electropherogram of the cultivar Souzão on *locus* VV16.

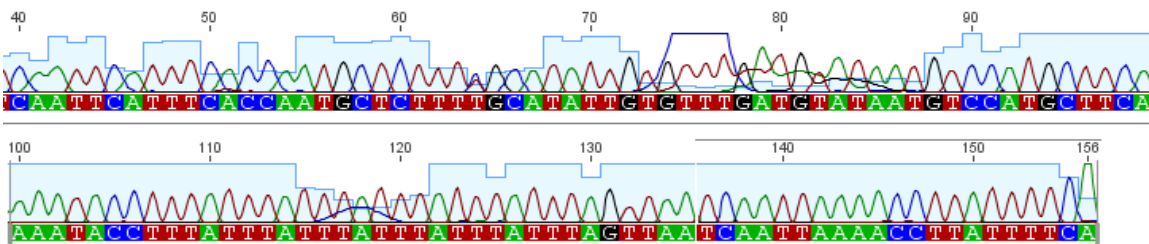


Fig. 5 - Electropherogram of the cultivar Tinto Cão on *locus* VV16.

Di-nucleotide STRs are generally highly polymorphic, more frequently presenting an allelic incremental step of two bp, which outcomes in the peaks of true alleles overlapping the stuttering peaks of the closest alleles that interfere with the scoring of the number of repeats at a given *locus* (Fig. 4 and Fig. 5). On the contrary, tetra-nucleotide STRs have a lower number of alleles, peak distances are larger, and stuttering peaks are attenuated, which makes the scoring sequence repeat numbers of these STRs more reliable allowing for the characterization of these *loci* at the sequence level.

Recommended primers for di-nucleotide were redesigned to survey the annealing region of the original primers in terms of polymorphisms. Despite the high amount of stuttering

on the electropherograms which made the sequence interpretation difficult, no apparent polymorphisms were found in these regions.

In order to genotype the grapevine cultivars according to the numbers of repeats using tetra-nucleotide markers, four grapevine cultivars were characterized for *locus* VV16 according to their number of repeats (Table 8).

Table. 8- Repeat motif of four Grapevine cultivars on *locus* VV16.

Primer	Grapevine cultivars	Repeat motif
VV16	Fernão Pires	(TAAA) ₅
	Souzão	(TAAA) ₅
	Tinto Cão	(TAAA) ₅
	Touriga Franca	(TAAA) ₅

For all four cultivars, the number of repeats was the same which means that *locus* VV16 cannot be used for discrimination purposes, at least among these cultivars. This also suggests that this marker may not be polymorphic in the *Vitis vinifera* and, therefore, not useful for the grapevine identification. This example shows the difficulty of finding adequate polymorphic STR markers in the *Vitis* genome. The methodology for mining for new tetra-nucleotide STRs used here was successfully employed before in the dog genome, where an abundance of polymorphic STRs with perfect repeat structure was found [92, 93].

These results highlight the difficulty of characterizing di-nucleotide markers at the sequence level and the apparent scarcity of adequate tetra-nucleotide markers in the *Vitis* genome. They also illustrate the need to employ long-core repeat markers (e.g. tetra-nucleotide) that allow for full characterization at the sequence level for grapevine identification.

In 2008, Cipriani et al. [27] presented a set of STR with long core repeats suitable for grapevine identification, this panel consisted of tri-, tetra- and penta-nucleotide markers. Due to the lack of use of tetra-nucleotide STRs on grapevine identification and given their great advantages, the purpose was to find and characterize this type of marker, thus contributing to more robust grapevine genotyping (promoting the use of fully characterized markers with a good discrimination power and the production of results that could be easily exchanged and compared between laboratories).

3.4. Structure analysis

In order to obtain a picture of the genetic diversity of the Portuguese cultivars in the context of other world regions, a population clustering analysis was performed using the Structure software [90]. Genetic structure analysis in large collections of grapevine genotypes can enable the better understanding of the extent and distribution of grape diversity. It may also aid in the clarification of the origin of complex traits in genetic resources using association genetics.

Given the possibility for sexual reproduction in the grapevine, the most appropriate model to analyze the data was the admixture model and a 'naïve' clustering algorithm, i.e. using no prior population information. This last allows for the software to blindly assign individuals to genetic clusters. Since there was no previous indication of the putative number of genetic clusters present in the dataset, the total compiled dataset (535 samples) was started to run for K up to 20. This preliminary test showed that a high percentage of the samples had a high admixture component (Fig. 6). K=6 was chosen as the best number of clusters capturing most of the genetic structure in the dataset. This was performed using the Structure Harvester software [90] which finds the most pertinent level of population subdivision (choosing the less negative K before plateau phase is achieved) (Fig.7).

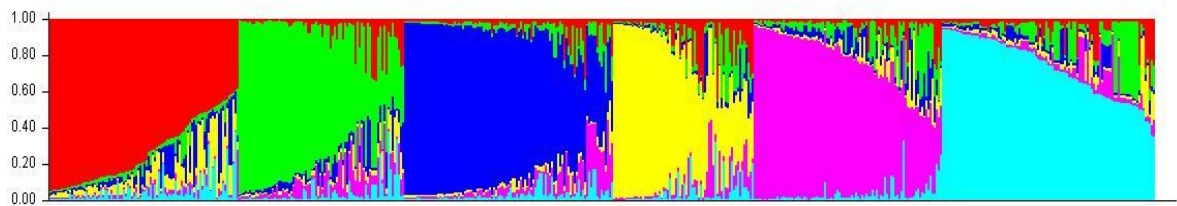


Fig. 6- Bar plotting of the results obtained with Structure software showing the genetic clustering of 535 grapevine genotypes when 6 populations are assumed in the dataset (K = 6). Each vertical line represents one individual, with color segmentation proportional to the Q coefficient of membership (Y axis) in the K clusters.

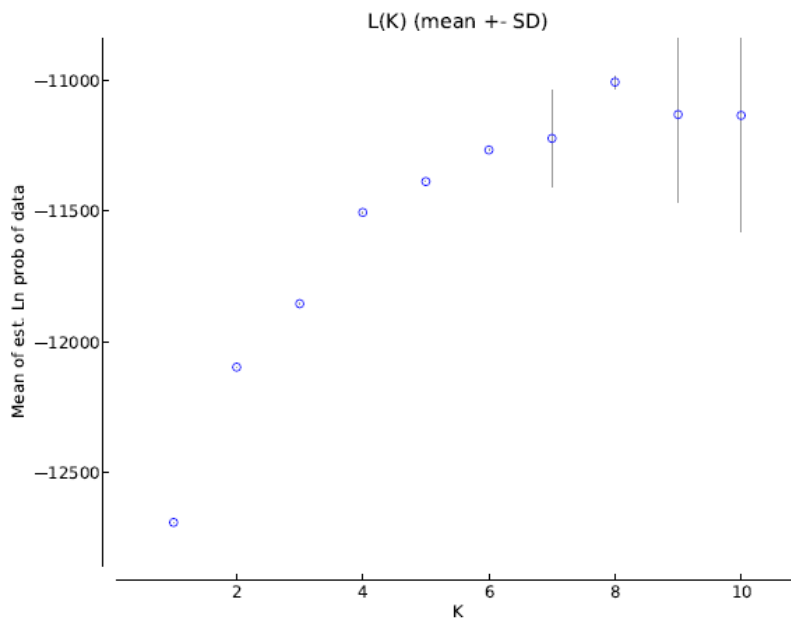


Fig. 7 - Graphic output of L(k) using the Structure Harvest software. The horizontal axis represents K and the vertical axis represent L(k).

Given the high level of sub-structuring in the dataset, we decided to perform a second genetic clustering analysis in a sub-dataset that excluded the samples that had a Q (membership coefficient) lower than 75%. This sub-dataset comprised 386 genotypes. We run the genetic clustering algorithm using the same setting described above and the K values was estimated with the same method and also resulted in six putative clusters (Fig. 8).

The bar plotting of the Structure analysis with 386 samples (Fig. 9) showed expected sharing of clusters by some varieties. For example, the Portuguese Touriga Nacional, Tourigo Francês, Touriga Franca and Tourigo do Douro shared cluster 3. Also strikingly, Cabernet Sauvignon, Chardonnay and Carbernet Franc also share the same cluster (cluster 5), all of them with a very Q (>90%). Although these results are preliminary, this method of genetic clustering analysis shows that it may be useful to aid in the identification of the relationships among grapevine cultivars.

Provided that a comprehensive database of genotypes obtained with adequate STRs, genetic clustering may also be developed to assign an unknown sample to a given cultivar or cluster of cultivars, thus contributing to the identification of cultivars for a

broad range of forensic applications. This type of analysis may also contribute to clarify the origins and history of grapevine diversity.

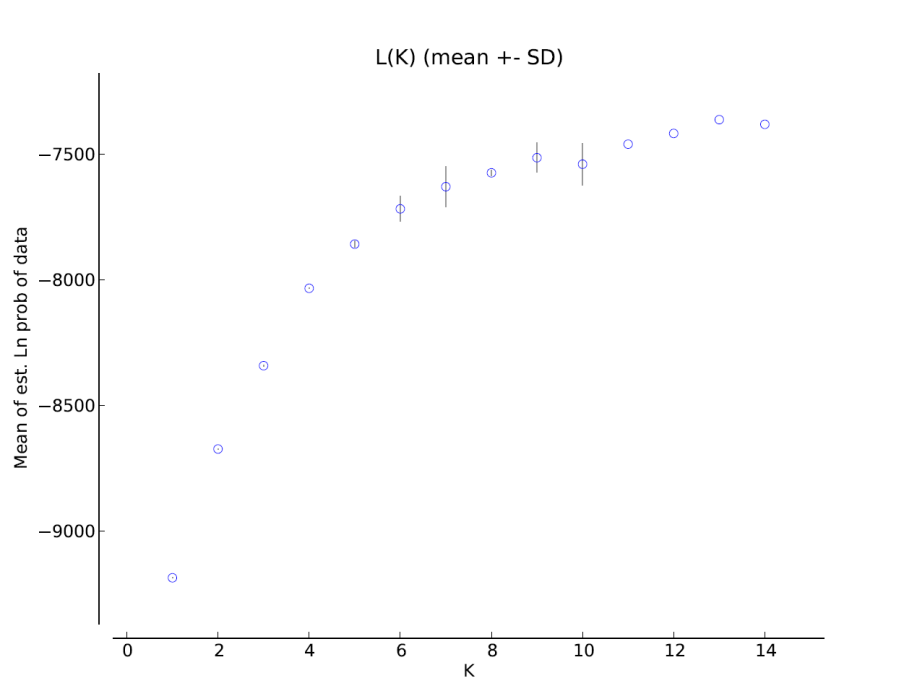
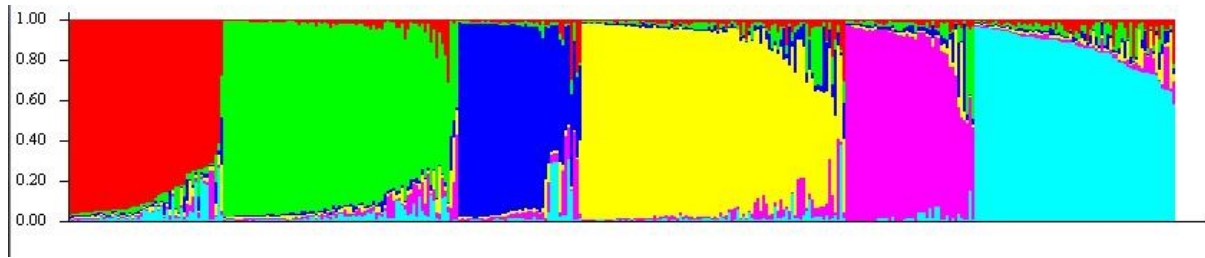
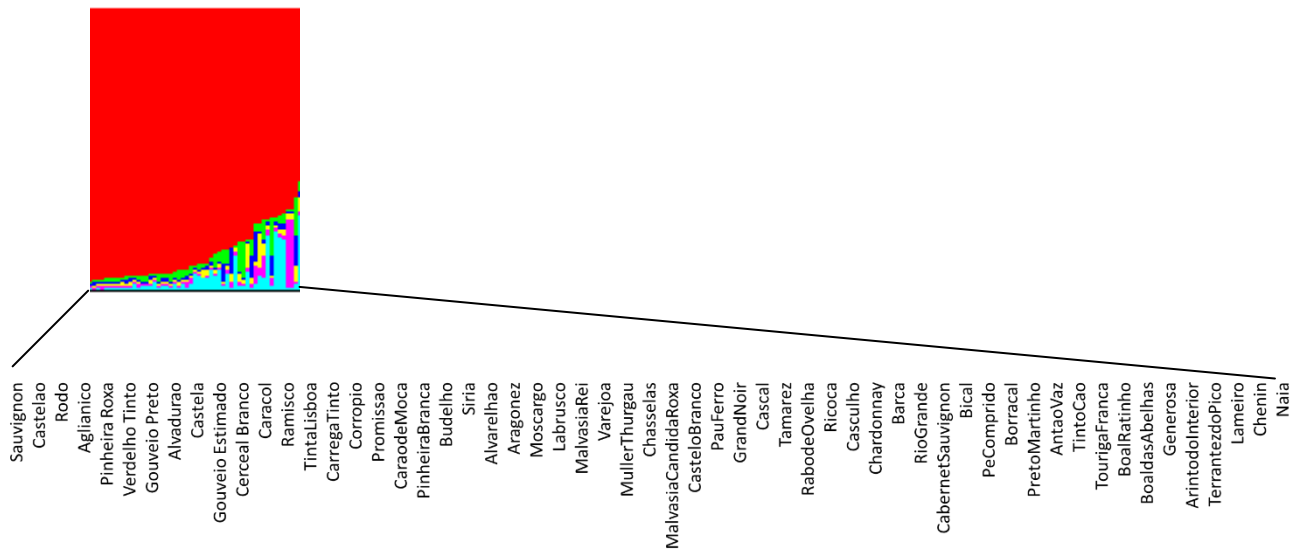


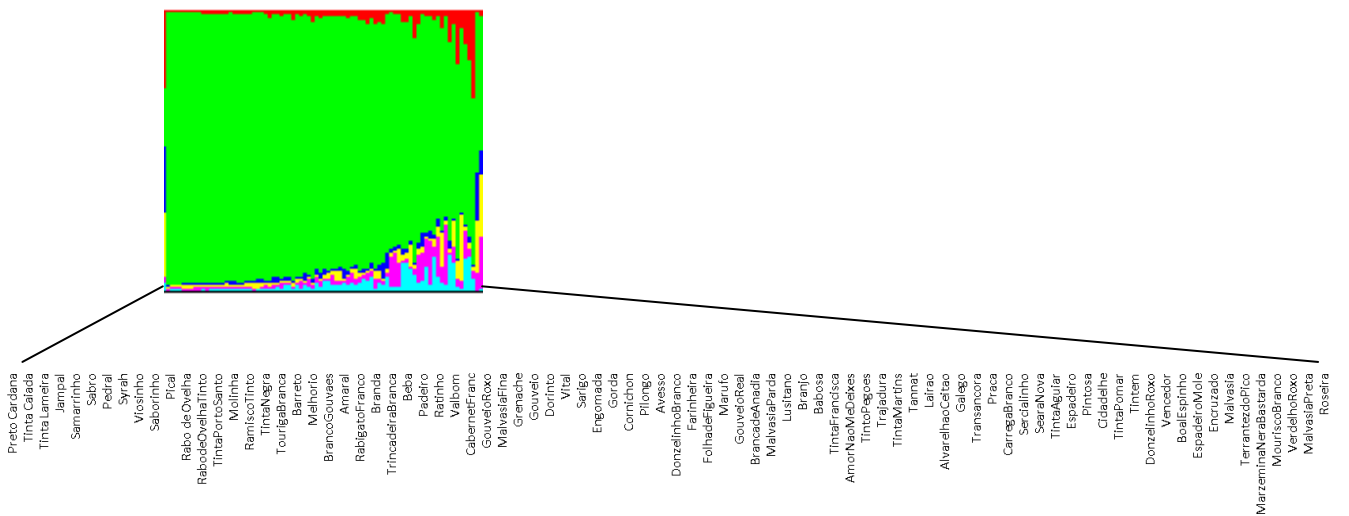
Fig. 6 - Graphic output of L(k) using the Structure Harvest software. The horizontal axis represents K and vertical axis represents L(k).

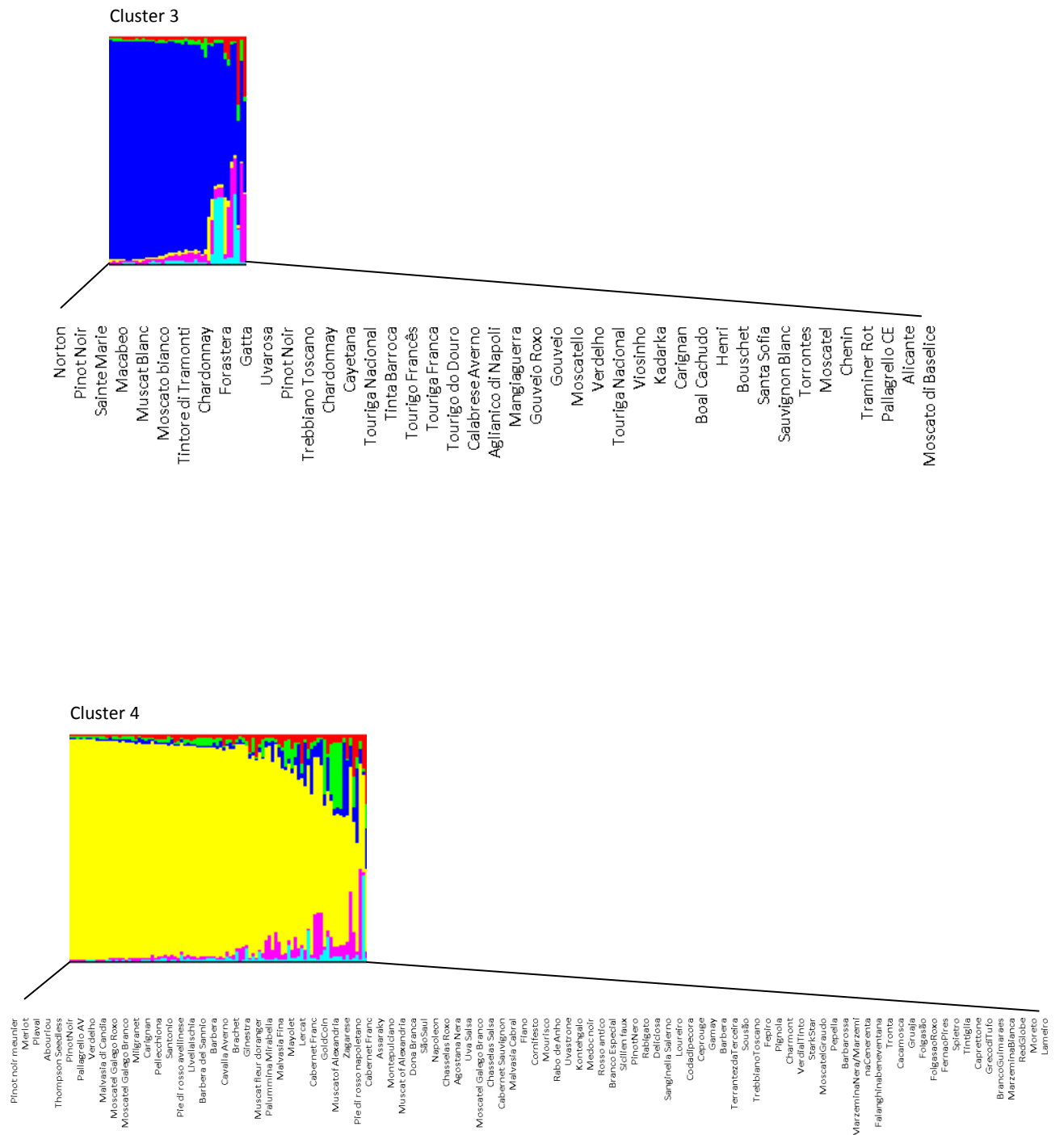


Cluster 1



Cluster 2





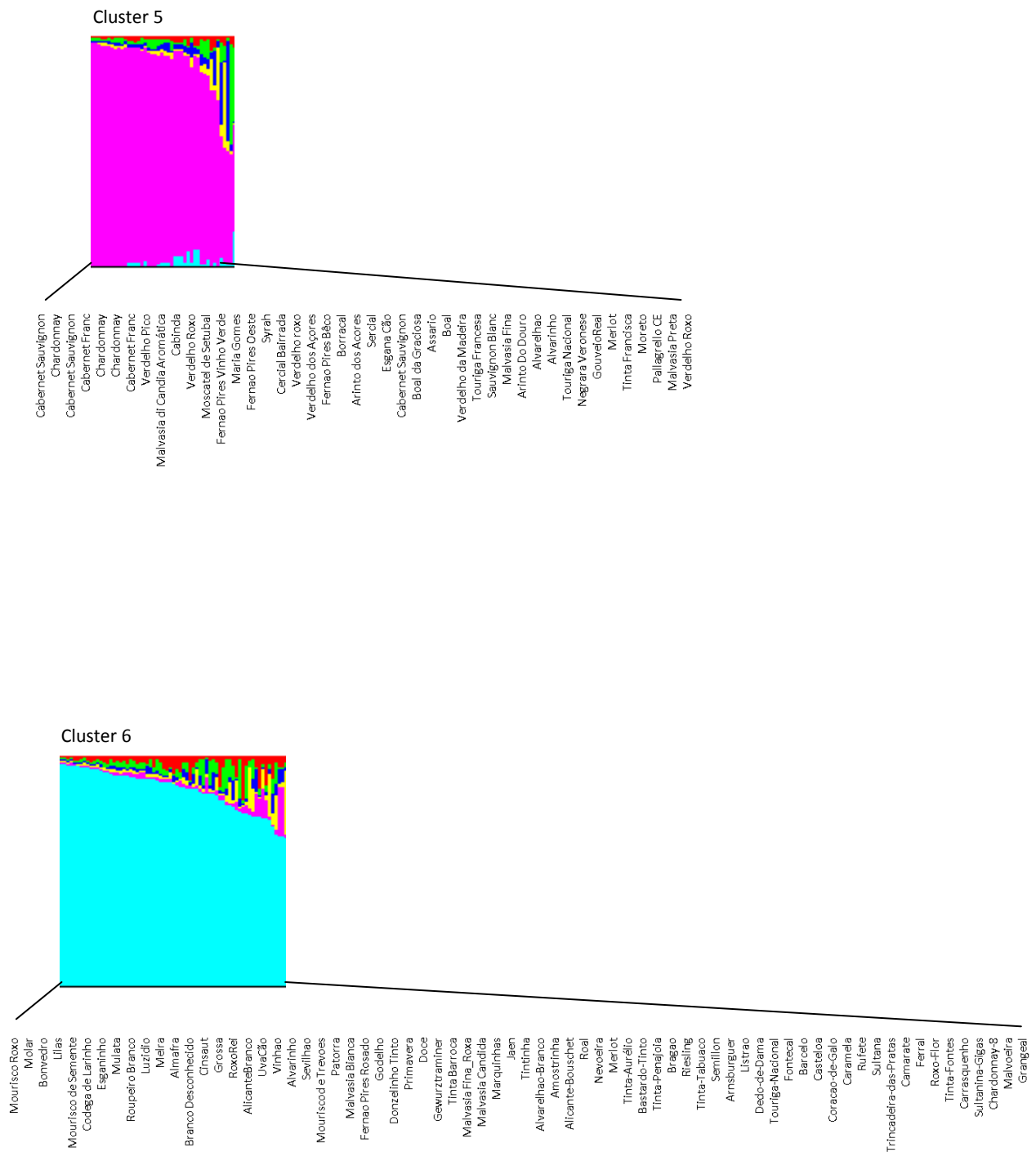


Fig. 7- Bar plotting of the results obtained with Structure software showing the genetic clustering of 386 grapevine genotypes when 6 populations are assumed in the dataset ($K = 6$). Each vertical line represents one individual, with color segmentation proportional to the Q coefficient of membership (Y axis) in the K clusters. Clusters are also shown separately with names of the samples below.

3.5. Recommendations for forensic genotyping of grapevine

A guideline for grapevine DNA genotyping which includes the correct nomenclature is yet to be created. Hence, in agreement with the recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations by Linacre et al. 2011 [94], which is a recommendation document on the use of non-human DNA testing in forensic investigations, and adapting these to grapevine identification, we compiled a set of recommendations that are essential for the development of high quality genotyping in the grapevine:

- a) It is important that the samples used in identification studies and in forensic investigations are “true-to-type”, there has to be certainty in the identity of the sample;
- b) In identification studies, the designed primers should produce consistent results and its specificity must be tested by sequencing primers annealing regions. In order to facilitate information exchange between laboratories, the results must be publicly available;
- c) All alleles obtained using STR markers in identification or other type of study should be sequenced, it is recommended the use of tetra-nucleotide markers that provide more reliable results, moreover, when di-nucleotides are used parameters like the expected heterozygosity should be presented;
- d) To ensure a trusty nomenclature of the alleles, all allelic ladders should be sequenced and the results should be referred as the number of the repeats;
- e) In parentage and pedigree analysis with STRs there should be consciousness of mutation rates when results are being evaluated;
- f) A high amount of individuals must be typed in order to obtain a credible allele frequency databases;
- g) It is important to use accurate values of the kinship factor (F_{ST}), on which it affects the chance that two members of the same population share an allele as they have a common genetic ancestor.

4. CONCLUSION AND FUTURE PERSPECTIVES

The advantages of international collaborations for grapevine molecular research it is clear, the uncovering of synonymy, the exchange of developed STR markers and plant material in order to develop more efficient and consistent methodologies on genotyping. It is expected that these collaborations will continue to play a major role in the future. For this to be possible, standardized methods and protocols in grapevine identification should be achieved, most important, and as said before, adequate genetic markers must allow for a high power of discrimination, be stable and produce consistent and reproducible results among different laboratories and detection platforms.

In the forensic context, di-nucleotide STRs remain problematic due to PCR artifacts (i.e. stuttering peaks) and narrow distances between adjacent alleles that complicate binning. For this reason, di-nucleotide STRs have been discarded in human fingerprinting in favor of STRs with longer core repeats. STR markers with long repeats, namely penta-, tetra-, and tri-nucleotides, are the recommended markers for individual genotyping, allowing correct binning and sizing of alleles, therefore allows the construction of robust databases of individual profiles. They offer a good level of polymorphism, are relatively easy to score, and alleviate the problems with binning and allele calling produced by the di-nucleotide repeat markers of the past. Furthermore, the standardization of allele nomenclature based on the number of repeats in the sequence aids in the reduction of allele miscalling and definitively would allow grape fingerprinting methods to converge with the current standards of forensic genetics established for human analyses.

The great amount of cultivars genetic profiles based on di-nucleotide markers accumulate in databases slows down the utilization of a more robust genotyping marker as tetra-nucleotide. Since there are few genotypes based on tetra-nucleotide markers in databases scientists prefer to keep using di-nucleotide markers, a great effort has to be made in order to change this.

A robust genotyping protocol for the grapevine should include: (a) the use of tetra-nucleotide loci; (b) the use of capillary electrophoresis so the PCR amplicons can be separated; (c) data analysis with a software that uses binning algorithms to call and size alleles; (d) the use of specific allelic ladders constructed for each *locus*.

Intra-cultivar genetic variation raises some concern about how to identify such variants. Some STR markers have presented allelic variation amongst clones of the same variety, so this has to be taken into account in forensics and parentage reconstruction. A strategy to overcome this problem could be the use of SNPs.

Tetra-nucleotide STRs can be the future of grapevine identification. For that to be accomplished, a more reliable and standardize genotyping method, based on fully characterized markers on the sequence level and the availability of allelic ladders, which allows information exchanges between laboratories, should be fomented by the forensic community.

5. REFERENCES

1. Tassie, L., *Vine identification- knowing what you have*, G.a.W.R.a.D. Corporation, Editor 2010: <http://www.gwrdc.com.au/wp-content/uploads/2012/09/2010-08-FS-Vine-Identification.pdf>.
2. Gaspero, G.d., et al., *Conservation of microsatellite loci within the genus Vitis*. Theoretical and Applied Genetics, 2000. **101**(1/2): p. 301-308.
3. Cabezas, J., et al., *A 48 SNP set for grapevine cultivar identification*. BMC plant biology, 2011. **11**(1): p. 153.
4. HP, O., *Grapes*. Evolution of crop plants, ed. S. NW (1976) Longman, London. 294-298.
5. Almadanim, M., et al., *Genetic diversity of the grapevine (*Vitis vinifera* L.) cultivars most utilized for wine production in Portugal*. VITIS-GEILWEILERHOF-, 2007. **46**(3): p. 116.
6. Sefc, K., et al., *Microsatellite markers for grapevine: a state of the art*, in *Molecular biology and biotechnology of grapevine*, K.A. Roubelakis-Angelakis, Editor 2001, Kluwer Academic Publishers. p. 433-463.
7. GEORGE, R., et al., *Wine a Scientific Eplanation*. Archaeology, ed. M. Sandler and R. Pinder. Vol. 21. 2003: Taylor & Francis Inc.
8. Pereira, L., H. Guedes-Pinto, and P. Martins-Lopes, *An enhanced method for *Vitis vinifera* L. DNA extraction from wines*. American Journal of Enology and Viticulture, 2011. **62**(4): p. 547-552.
9. Takeoka, G. and S. Ebeler. *Progress in authentication of food and wine*. in *American Chemical Society Symposium Series*. 2011. American Chemical Society Symposium Series.
10. Faria, M., et al., *Vitis vinifera must varietal authentication using microsatellite DNA analysis (SSR)*. Journal of agricultural and food chemistry, 2000. **48**(4): p. 1096-1100.
11. This, P., et al., *Development of a standard set of microsatellite reference alleles for identification of grape cultivars*. Theoretical and Applied Genetics, 2004. **109**(7): p. 1448-1458.
12. Cunha, J., et al., *Portuguese traditional grapevine cultivars and wild vines (*Vitis vinifera* L.) share morphological and genetic traits*. Genetic Resources and Crop Evolution, 2009. **56**(7): p. 975-989.
13. Ortiz, J.M., et al., *Molecular and morphological characterization of a Vitis gene bank for the establishment of a base collection*. Genetic Resources and Crop Evolution, 2004. **51**(4): p. 403-409.
14. VELOSO, M., et al., *Microsatellite database of grapevine (*Vitis vinifera* L.) cultivars used for wine production in Portugal*. Ciência Téc. Vitiv, 2010. **25**(2): p. 53-61.
15. Rapeanu, G., C. Vicol, and C. Bichescu, *Possibilities to assess the wines authenticity*. Inn Rom Food Biotechnol, 2009. **5**: p. 1-9.
16. Thomas, M.R., et al., *Repetitive DNA of grapevine: classes present and sequences suitable for cultivar identification*. Theor. Appt. Genet., (1993). **86** p. 173-180.
17. Bowers, J.E., G.S. Dangl, and C.P. Meredith, *Development and characterization of additional microsatellite DNA markers for grape*. American Journal of Enology and Viticulture, 1999. **50**(3): p. 243-246.

18. Bowers, J., et al., *Isolation and characterization of new polymorphic simple sequence repeat loci in grape (Vitis vinifera L.)*. Genome, 1996. **39**(4): p. 628-633.
19. Edwards A, C.A., Hammond HA, Caskey CT, *DNA typing and genetic mapping with trimetric tandem repeats*. Am J Hum Genet, 1991. **49**: p. 746-756.
20. Polymeropoulos, M.H., H. Xiao, and C.R. Merrill, *Tetranucleotide repeat polymorphism at the human myelin basic protein gene (MBP)*. Human molecular genetics, 1992. **1**(8): p. 658-658.
21. Urquhart, A., et al., *Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers*. International journal of legal medicine, 1994. **107**(1): p. 13-20.
22. Delseny, M., M. Laroche, and P. Penon, *Detection of sequences with Z-DNA forming potential in higher plants*. Biochemical and Biophysical Research Communications, 1983. **116**(1): p. 113-120.
23. Thomas, M.R., P. Cain, and N.S. Scott, *DNA typing of grapevines: A universal methodology and database for describing cultivars and evaluating genetic relatedness*. Plant Molecular Biology, 1994. **25**(6): p. 939-949.
24. Thomas, M. and N. Scott, *Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs)*. Theoretical and Applied Genetics, 1993. **86**(8): p. 985-990.
25. Sefc, K., et al., *Reconstruction of a grapevine pedigree by microsatellite analysis*. Theoretical and Applied Genetics, 1998. **97**(1-2): p. 227-231.
26. Reisch, B.I., C.L. Owens, and P.S. Cousins, *Grape*, in *Fruit breeding* 2012, Springer. p. 225-262.
27. Cipriani, G., et al., *A set of microsatellite markers with long core repeat optimized for grape (Vitis spp.) genotyping*. BMC plant biology, 2008. **8**(1): p. 127.
28. López, M., et al., *Microsatellite and AFLP analysis of autochthonous grapevine cultivars from Galicia (Spain)*. American Journal of Enology and Viticulture, 2009. **60**(2): p. 215-222.
29. Zdunić, G., et al., *Genetic Characterization of Grapevine Cultivars Collected throughout the Dalmatian Region*. American Journal of Enology and Viticulture, 2013. **64**(2): p. 285-290.
30. Cipriani, G., 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*. Theoretical and Applied Genetics, 2010. **121**(8): p. 1569-1585.
31. Cipriani, G., M.T. Marrazzo, and E. Peterlunger, *Molecular characterization of the autochthonous grape cultivars of the region Friuli Venezia Giulia-North-Eastern Italy*. Vitis, 2010. **49**(1): p. 29-38.
32. HARTA, M., et al., *Identification of Romanian Vitis vinifera L. Cultivars in Must Using Nuclear Microsatellite Markers*. Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Horticulture, 2010. **67**(1): p. 198-203.
33. Dzhambova, T., et al., *Genetic diversity in native Bulgarian grapevine germplasm (Vitis vinifera L.) based on nuclear and chloroplast microsatellite polymorphisms*. Vitis, 2009. **48**: p. 115-121.
34. Costantini, L., et al., *Genetic relationships among local Vitis vinifera cultivars from Campania (Italy)*. Vitis, 2005. **44**(1): p. 25-34.
35. Lopes, M.S., et al., *The use of microsatellites for germplasm management in a Portuguese grapevine collection*. Theoretical and Applied Genetics, 1999. **99**(3-4): p. 733-739.
36. Schneider, A., et al., *Verifying synonymies between grape cultivars from France and Northwestern Italy using molecular markers*. VITIS-GEILWEILERHOF-, 2001. **40**(4): p. 197-204.

37. De Lorenzis, G., et al., *Pedigree Reconstruction of the Italian Grapevine Aglianico (Vitis vinifera L.) from Campania*. Molecular biotechnology, 2012: p. 1-9.
38. This, P., T. Lacombe, and M.R. Thomas, *Historical origins and genetic diversity of wine grapes*. TRENDS in Genetics, 2006. **22**(9): p. 511-519.
39. Pelsy, F., et al., *An extensive study of the genetic diversity within seven French wine grape variety collections*. Theoretical and Applied Genetics, 2010. **120**(6): p. 1219-1231.
40. Emanuelli, F., et al., *Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape*. BMC plant biology, 2013. **13**(1): p. 39.
41. Aradhya, M.K., et al., *Genetic structure and differentiation in cultivated grape, Vitis vinifera L.* Genetical research, 2003. **81**(3): p. 179-192.
42. Martinez, L., et al., *SSR-based assessment of genetic diversity in South American Vitis vinifera varieties*. Plant Science, 2006. **170**(6): p. 1036-1044.
43. Sefc, K., et al., *Microsatellite variability in grapevine cultivars from different European regions and evaluation of assignment testing to assess the geographic origin of cultivars*. Theoretical and Applied Genetics, 2000. **100**(3-4): p. 498-505.
44. Donini, P., et al., *Genetic diversity and geographical dispersal in grapevine clones revealed by microsatellite markers*. Genome, 2006. **49**(11): p. 1459-1472.
45. Meneghetti, S., et al., *Study of intra-varietal genetic variability in grapevine cultivars by PCR-derived molecular markers and correlations with the geographic origins*. Molecular biotechnology, 2012. **50**(1): p. 72-85.
46. Riaz, S., et al., *A microsatellite marker based framework linkage map of Vitis vinifera L.* Theoretical and Applied Genetics, 2004. **108**(5): p. 864-872.
47. Di Gaspero, G., et al., *Linkage maps of grapevine displaying the chromosomal locations of 420 microsatellite markers and 82 markers for R-gene candidates*. Theoretical and Applied Genetics, 2007. **114**(7): p. 1249-1263.
48. Battilana, J., et al., *Linkage mapping and molecular diversity at the flower sex locus in wild and cultivated grapevine reveal a prominent SSR haplotype in hermaphrodite plants*. Molecular biotechnology, 2013: p. 1-7.
49. Doligez, A., et al., *An integrated SSR map of grapevine based on five mapping populations*. Theoretical and Applied Genetics, 2006. **113**(3): p. 369-382.
50. Maul, E., *The European Vitis Database*.
51. Maul, E., et al. *GENRES 081-a basis for the preservation and utilization of Vitis genetic resources*. in Report of a working group on Vitis, first meeting, Palic', Serbia and Montenegro, 12-14 June, 2003. 2008. Bioversity International.
52. Grando, M.S. and C. Frisinghelli, *Grape microsatellite markers: Sizing of DNA alleles and genotype analysis of some grapevine cultivars*. VITIS-GEILWEILERHOF-, 1998. **37**: p. 79-82.
53. Loureiro, M.D., et al., *Molecular Marker Analysis of Vitis vinifera Albariño and some Similar Grapevine Cultivars*. Journal of the American Society for Horticultural Science, 1998. **123**(5): p. 842-848.
54. Crespan, M., *Evidence on the evolution of polymorphism of microsatellite markers in varieties of Vitis vinifera L.* Theoretical and Applied Genetics, 2004. **108**(2): p. 231-237.
55. Hvarleva, T., et al., *Genotyping of Bulgarian Vitis vinifera L. cultivars by microsatellite analysis*. VITIS-GEILWEILERHOF-, 2004. **43**(1): p. 27-34.
56. Hvarlena, T., et al., *Genotyping Vitis vinifera L. cultivars of Cyprus by microsatellite analysis*. Vitis-Journal of Grapevine Research, 2005. **44**(2): p. 93-97.

57. Santiago, J., et al., *Characterisation and identification of grapevine cultivars (Vitis vinifera L.) from northwestern Spain using microsatellite markers and ampelometric methods*. Vitis, 2005. **44**(2): p. 67-72.
58. HESZKY, L. and E. KISS, *Identification of grapevine cultivars using microsatellite-based DNA barcodes*. Vitis, 2009. **48**(1): p. 17-24.
59. Stover, E., et al. *Investigations into the origin of Norton's grape using SSR markers*. in *Proc. Fla. State Hortic. Soc.* 2010.
60. Malenica, N., et al., *Whole genome amplification and microsatellite genotyping of herbarium DNA revealed the identity of an ancient grapevine cultivar*. Naturwissenschaften, 2011. **98**(9): p. 763-772.
61. Bergamini, C., et al., *Evidences for an alternative genealogy of 'Sangiovese'*. Molecular biotechnology, 2012: p. 1-11.
62. De Andrés, M., et al., *Genetic diversity of wild grapevine populations in Spain and their genetic relationships with cultivated grapevines*. Molecular Ecology, 2012. **21**(4): p. 800-816.
63. Meneghetti, S., A. Calò, and L. Bavaresco, *A strategy to investigate the intravarietal genetic variability in Vitis vinifera L. for clones and biotypes identification and to correlate molecular profiles with morphological traits or geographic origins*. Molecular biotechnology, 2012. **52**(1): p. 68-81.
64. Sant'Ana, G., et al., *Comparison of a retrotransposon-based marker with microsatellite markers for discriminating accessions of Vitis vinifera*. Genetics and Molecular Research, 2012. **11**(2): p. 1507-1525.
65. Vezzulli, S., et al., *Pinot blanc and Pinot gris arose as independent somatic mutations of Pinot noir*. Journal of experimental botany, 2012. **63**(18): p. 6359-6369.
66. Harangozó, T., et al., *Assessment of morphological and molecular similarity of Hungarian white grape varieties*. Acta Biologica Hungarica, 2013. **64**(2): p. 231-248.
67. Pellerone, F., K. Edwards, and M. Thomas, *Grapevine microsatellite repeats: isolation, characterisation and use for genotyping of grape germplasm from Southern Italy*. VITIS-GEILWEILERHOF-, 2001. **40**(4): p. 179-186.
68. Siret, R., et al., *Analysis of grape Vitis vinifera L. DNA in must mixtures and experimental mixed wines using microsatellite markers*. Journal of agricultural and food chemistry, 2002. **50**(13): p. 3822-3827.
69. Pinto-Carnide, O., et al., *Characterization of grapevine (Vitis vinifera L.) cultivars from northern Portugal using RAPD and microsatellite markers*. VITIS-GEILWEILERHOF-, 2003. **42**(1): p. 23-26.
70. García-Beneytez, E., et al., *Application of a DNA analysis method for the cultivar identification of grape musts and experimental and commercial wines of Vitis vinifera L. using microsatellite markers*. Journal of agricultural and food chemistry, 2002. **50**(21): p. 6090-6096.
71. Lopes, M.S., et al., *Discrimination of Portuguese grapevines based on microsatellite markers*. Journal of Biotechnology, 2006. **127**(1): p. 34-44.
72. Bonen, L., et al., *Gene pool variation and phylogenetic relationships of an indigenous northeast Italian grapevine collection revealed by nuclear and chloroplast SSRs*. Genome, 2008. **51**(10): p. 838-855.
73. Goldstein, D., et al., *Genetic absolute dating based on microsatellites and the origin of modern humans*. Proceedings of the National Academy of Sciences, 1995. **92**(15): p. 6723-6727.
74. Bell, G.I. and J. Jurka, *The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process*. Journal of molecular evolution, 1997. **44**(4): p. 414-421.
75. Kruglyak, S., et al., *Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations*.

- Proceedings of the National Academy of Sciences, 1998. **95**(18): p. 10774-10778.
76. Calabrese, P.P., R.T. Durrett, and C.F. Aquadro, *Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models*. Genetics, 2001. **159**(2): p. 839-852.
 77. Dong, Q.-H., et al., *Discovery and characterization of SNPs in Vitis vinifera and genetic assessment of some grapevine cultivars*. Scientia Horticulturae, 2010. **125**(3): p. 233-238.
 78. Myles, S., et al., *Genetic structure and domestication history of the grape*. Proceedings of the National Academy of Sciences, 2011. **108**(9): p. 3530-3535.
 79. ARROYO.GARCÍA, R., et al., *Multiple origins of cultivated grapevine (Vitis vinifera L. ssp. sativa) based on chloroplast DNA polymorphisms*. Molecular Ecology, 2006. **15**(12): p. 3707-3714.
 80. Castro, I., et al., *Chloroplast Genome Diversity in Portuguese Grapevine (Vitis vinifera L.) Cultivars*. Mol Biotechnol, 2013: p. 1-13.
 81. Cunha, J., et al., *The Portuguese Vitis vinifera L. germplasm: genetic relations between wild and cultivated vines*. Ciência e Técnica Vitivinícola, 2010. **25**(1): p. 25-37.
 82. (<http://www.oiv.int/oiv/cms/index?lang=en>).
 83. (<http://www.eu-vitis.de/index.php>).
 84. (<http://www.plantqdb.org/VvGDB/>).
 85. Benson, G., *Tandem repeats finder: a program to analyze DNA sequences*. Nucleic Acids Research
1999. **Vol. 27**(No. 2): p. 573-580.
86. (<http://www.ncbi.nlm.nih.gov/>).
 87. Vallone, P. and J. Butler, *AutoDimer: a screening tool for primer-dimer and hairpin structures* Biotechniques, 2004. **37**: p. 226-231.
 88. Doyle, J.J. and J.L. Doyle, *A rapid DNA isolation procedure for small quantities of fresh leaf tissue*. Focus, 1987. **12**: p. 13-15.
 89. Taberlet, P., et al., *Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding*. Nucleic Acids Res, 2007. **35**(3): p. e14-e14.
 90. Pritchard, J.K., W. Wen, and D. Falush, *Documentation for STRUCTURE software: version 2*, 2003.
 91. Lijavetzky, D., et al., *High throughput SNP discovery and genotyping in grapevine (Vitis vinifera L.) by combining a re-sequencing approach and SNPlex technology*. BMC genomics, 2007. **8**(1): p. 424.
 92. Van Asch, B. and F. Pereira, *State-of-the-art and future prospects of canine STR-based genotyping*. Open Forensic Science Journal, 2010. **3**: p. 45-52.
 93. van Asch, B., et al., *A framework for the development of STR genotyping in domestic animal species: Characterization and population study of 12 canine X-chromosome loci*. Electrophoresis, 2010. **31**(2): p. 303-308.
 94. Linacre, A., et al., *ISFG: recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations*. Forensic Science International: Genetics, 2011. **5**(5): p. 501-505.