Microsatellite Profiles as a Basis for Intellectual Property Protection in Grape

J. Ibañez[1] and F.A. van Eeuwijk[2]
[1] Instituto Madrileño de Investigación Agraria y Alimentaria. Apdo 127. 28800 Alcala de Henares. Spain.
[2] Lab. of Plant Breeding, Dpt of Plant Sciences, Wageningen University, P.O. Box 386, 6700 AJ Wageningen, The Netherlands

Keywords: essential derivation, forensic statistics, matching probability, Plant Breeders' Rights, SSRs, Vitis, varietal identification.

Abstract

The use of microsatellite analysis in a forensic procedure for establishing infringement on plant breeders' rights in vegetatively propagated crops was evaluated. A reference collection of 45 seedless grape varieties was chosen as reference collection. Matching probabilities of grape microsatellite genotypes were calculated under the assumption of independent breeding programs. After testing for independence of alleles within and between loci a set of 5 microsatellite loci was chosen as basis for the calculation of matching probabilities. Even the highest matching probabilities were so low that a forensic procedure using microsatellites seems to provide a sound basis for assessing infringements on plant breeders' rights for the case of grape.

INTRODUCTION

Development of new varieties is becoming an increasingly costly history. Therefore, a good legal protection system has become a high priority for breeders. The current plant variety protection system is based on UPOV regulations (UPOV 1991). Within the UPOV system for the granting of breeders' rights a dominant role is given to the testing system for Distinctness, Uniformity and Stability, the DUS system. Focussing on distinctness, only those candidate varieties that are found sufficiently different on one or more of a set of morphological/ anatomical characteristics from any other variety of common knowledge over a 2-3 year testing period will receive breeders' rights. Furthermore, the candidate varieties should comply with the subsidiary requirement of the intra varietal heterogeneity not exceeding that of comparable varieties, where comparable refers to the mean expression of the characteristics used for assessing distinctness. Finally, the plant material of the variety must remain unchanged after repeated propagation.

The plant variety protection conferred by the UPOV provides a legal framework that offers sufficient protection on average. However, infringements and piracy on plant breeders' rights do occur. Typical examples are the following: one protected variety can be exploited under a different name without the authorization (legally needed) of the holder of the right. Another one, a bit more complex, occurs when an already protected variety, the initial variety, is exploited or even resubmitted, usually after some 'cosmetic' breeding to create an illusion of phenotypic difference, under another name for DUS testing in the same or another country as the country that granted plant breeders' rights before. When the DUS tests are successful, a paradoxical situation originates in which the 'same' variety can be sold under two names, where it is clear that the company that resubmitted the variety under a new name, the derived variety, infringes on the rights of the company that had the original rights. Such a situation is nowadays alluded to as a simple case of essential derivation (see ASSINSEL position paper on intellectual property rights on http://www.worldseed.org). Even when the difference between the initial variety and the derived variety is not 'cosmetic', but affects any important agronomical trait like, for instance, when a somatic mutation for berry colour occurs in a protected variety, the exploitation of the derived variety requires the authorization of the owner of the initial

variety. Otherwise, the exploitation of the derived variety is illegal.

Infringement on plant breeders' rights of the above types do occur in table grape varieties and, in fact, some cases have been taken to court. Grapevines are vegetatively propagated and as such all plants of a variety will be genetically identical. Therefore, infringement by exploitation of an already protected or an essentially derived variety should in principle be easily detectable by comparison of DNA profiles, since these cases will usually lead to identical DNA profiles. Ibañez (2001) gives an example of the use of DNA profiles for assessing fraud of this type using RAPDs. The statistical elements of the procedure are like those used in human forensics (Evett and Weir, 1998). As the grape example is relatively uncomplicated, the issue is to calculate the probability of the occurrence of a match between the profile of initial and suspect variety under the assumption of independent breeding programs.

In this paper, the procedure presented in Ibañez (2001) will be elaborated in two ways. First, sequence-tagged microsatellite sites (STMSs) will be used as markers for grape characterization (Bowers et al. 1996; Sefc et al. 1999; Thomas and Scott 1993). Second, calculation of match probabilities will be based on allele frequencies in the reference population, the population within which possible identity or essential derivation cases might occur. In addition, an error margin for the probability of identity in profile will be presented.

MATERIAL AND METHODS

Plant Material

Table 1 shows the 45 seedless grape varieties studied. The list includes accessions producing grapes of distinct sizes and colours (yellow-green/rose-red/black-blue-violet), of different seasons of ripening (early, mid, late) and of the two types of seedlessness (parthenocarpy and stenospermy), as well as different genetic relationships (full-sibs, half-sibs, parents and progeny). Two different plants were studied per variety. Plant material was mainly obtained from the Vitis Germplasm Bank (BGV) at the Finca El Encin, Alcala de Henares (Spain). Some cultivars were obtained from CIDA (Murcia).

DNA Extraction

DNA extractions were carried out from young leaves following the CTAB method described by Lodhi et al. (1994). After RNase treatment, DNAs were further purified by PCIA (Phenol:Chlorophorm:Isoamylic Alcohol 25:24:1) extraction and then by CIA (24:1) extraction. DNA concentrations were established in agarose gels by comparing fluorescent intensity against Lambda DNA standards. DNAs from two different plants were extracted per accession.

STMS Analysis

A total of nine STMS previously described were studied: VVS1, VVS2, VVS5, VVS29 (Thomas and Scott 1993), VVMD5, VVMD7 (Bowers et al. 1996), ssrVrZAG 47, ssrVrZAG 62 and ssrVrZAG 79 (Sefc et al. 1999). Six of the nine loci (all but VVS1, VVS5 and VVS29) were chosen as a core set for the screening of grapevine collections in Europe covered by the GENRES#081 research project (www.genres.de/vitis/summary3.htm). Initially, each locus was individually amplified in a reaction volume of 20 µl containing: 10 mM Tris-HCl pH 8.3; 50 mM KCl; 200 µM of each dNTP; 1.5 mM MgCl2; 0.5 µM upper primer, with an attached fluorescent dye; 0.5 µM lower primer; 0.025 U/µl of AmpliTaq Gold DNA Polymerase and 1 ng/µl of genomic DNA. The amplification program was: 1 cycle [94 °C, 5 min], 30 cycles [94 °C, 45 s; 50 °C, 1 min; 72 °C, 2 min] and 1 cycle [72 °C, 10 min]. Later, the STMSs were grouped in three multiplex PCRs: set A included microsatellites VVS5 (labelled with the fluorochrome 6-FAM), VVS29 (HEX) and VVMD7 (TET), set B included the loci VVS1 (TET), VVS2 (HEX) and VVMD5 (6-FAM) and set C included ssrVrZAG 47 (6-FAM), ssrVrZAG 62 (TET) and ssrVrZAG 79 (HEX). The reaction mixes were like in the

individual PCRs except for the primers and polymerase concentration: Set A contained VVS5 primers 0.5 µM each; VVS29 primers 0.3 µM each; VVMD7 primers 0.15 µM each and 0.038 U/µl of Amplitaq Gold DNA Polymerase (PE Biosystems). Set B contained VVS1 primers 0.3µM each; VVS2 primers 0.05 µM each; VVMD5 primers 0.5 µM each and 0.05 U/µl of Amplitaq Gold DNA Polymerase (PE Biosystems). Set C contained ZAG47 primers 0.5µM each; ZAG62 primers 0.1 µM each; ZAG79 primers 0.5 µM each and 0.05 U/µl of Amplitaq Gold DNA Polymerase (PE Biosystems). The amplification program was the same for the three sets: 1 cycle [95 °C, 5 min], 40 cycles [94 °C, 45 s; 50 °C, 1 min; 72 °C, 1.5 min] and 1 cycle [72 °C, 7 min]. The separation of the amplified fragments was carried out in an ABI PRISM 377 Sequencer (sets A and B) or in an ABI PRISM 310 (set C) (PE Biosystems). Electrophoresed fragments were sized with the GENESCAN software, using TAMRA 350 or 500 respectively as an internal marker.

Microsatellite VVS1 was further analysed individually in the following PCR conditions: 20 mM Tris-HCl pH 8.8; 10 mM KCl; 10 mM (NH4)2SO4; 0.1% Triton X-100; 3 mM MgSO4; 400 µM of each dNTP; 0.4 µM upper primer, with an attached fluorescent dye; 0.4 µM lower primer; 0.02 U/µl of Vent DNA Polymerase (New England Biolabs) and 5 ng/µl of genomic DNA. PCR program: 1 cycle [95 °C, 5 min], 35 cycles [95 °C, 30 s; 50 °C, 1 min; 72 °C, 30 s] and 1 cycle [72 °C, 10 min]. The separation of the amplified fragments was carried out in an ABI PRISM 310. Electrophoresed fragments were sized with the GENESCAN software, using TAMRA 500 as an internal marker. Allele binning of GENESCAN values was carried out following the algorithm described by Ghosh et al. (1997).

Genetical and Statistical Analysis

Matching probabilities were calculated from allele frequencies assuming Hardy Weinberg equilibrium within loci and Linkage Equilibrium between loci. Thus, when p_j and p_j are the frequencies of the alleles *i* and *j*, at a particular locus, the frequencies for homozygotes at this locus will be p_i^2 and p_j^2, where heterozygotes will have frequency $2p_i\,p_j$. The probability of a matching profile at a locus is just the probability of that genotype,. p_i^2 , p_j^2, or $2p_i\,p_j$ The extension to more alleles at the same locus is obvious. For the matching probability at multiple loci just multiply the matching probabilities at the individual loci.

An important assumption in the calculation of the matching probabilities along these lines is the independence of alleles within (Hardy-Weinberg Equilibrium) and between loci (Linkage Equilibrium). We explicitly tested for these assumptions using the test for Hardy Weinberg by Guo and Thompson (1992), and the test for Linkage Equilibrium by Slatkin and Excoffier (1996), in the form as they are implemented in Arlequin (Schneider et al., 2000). We used a Bonferroni correction for multiple testing (Weir, 1996), leading to a comparison-wise test level of 0.05/9 = 0.0056 for the Hardy Weinberg tests, and a comparison-wise test level of 0.05/28 = 0.0018 for Linkage Equilibrium.

To calculate an upper bound to the matching probabilities we used the formulae for the variances of one locus genotype probabilities $Var(\hat{p}_i^2) \approx \frac{1}{2n} 4p_i^3(1-p_i)$ for homozygotes and $Var(2\hat{p}_i\hat{p}_j) \approx \frac{1}{2n} 4p_i p_j (p_i + p_j - 4p_i p_j)$ for heterozygotes (Evett and Weir 1998, Box 5.2). Under independence, the variance of multi-locus genotypes can be approximated by $Var(\prod_l \hat{P}_l) = (\prod_l P_l)^2 \left(\prod_l \left[1 + \frac{Var(\hat{P}_l)}{P_l^2} \right] - 1 \right)$ (Evett and Weir 1998, Box 5.6), where P_l refers to the one locus genotype for the locus *l*, Of course, for evaluation of these formulae the population parameters, without hat, should be replaced

by their estimates.

RESULTS AND DISCUSSION

STMS Analysis

The results of the DNA analyses performed on the 45 seedless varieties allowed to identify all cultivars. No substantial differences were found among GENESCAN values obtained for the two plants analyzed within each variety. Allele binning was carried out essentially following the algorithm described by Ghosh et al. (1997). For most of the loci, a few bins were manually created or removed. As a result of the automatic and manual binning, a genotype table was constructed. The resolving power of the system was very high: only 3 loci (VVS2, VVS5 and VVMD7) were needed to distinguish all the cultivars.

Genetical and Statistical Analysis

Before calculating the matching probabilities we tested for Hardy Weinberg and Linkage Equilibrium. At the test levels given above, only VVS5 was found not to be in Hardy-Weinberg Equilibrium. This is very probably because of the presence of null alleles in a high frequency at this locus (Ibañez, submitted), which causes an apparent 'deficiency' in heterozygotes, and, consequently, a deviation from equilibrium. With respect to linkage disequilibrium tests, we found significant disequilibrium within the pairs VVMD7 and ZAG62, ZAG47 and ZAG79, and VVS2 and VVMD5. Combining the results of the tests for Hardy Weinberg and Linkage Disequilibrium we found a few sets of five loci exhibiting independence of alleles within and between loci.

Table 2 shows the most extreme frequencies for the individual loci, together with 95% upper limit bounds. These values give an idea about how decisive individual loci could be in disputes on breeders' rights. Best suited for settling disputes on a one locus basis would be VVS2 and ZAG79. In contrast, VVS1 and VVS29 seemed less useful, at least for the current set of varieties. It is clear that a protection procedure based on just 1 microsatellite locus will not work. We then studied the different sets of five loci fulfilling HWE and LE and found that VVS1, VVS2, VVS29, VVMD7, and ZAG79, was the one with lowest average matching probability. Using this set, the most extreme matching probability was only 0.00031(for variety Danuta, Table 3). In other words, combining these 5 loci, the highest probability with which any two seedless grape profiles out of the reference set of 45 will match by chance, amounts to only once in about 3200. Such a low probability provides a solid ground for decisions in potential fraud cases. A conservative approach can be followed, using the 95% upper limits for the matching probabilities. These values are between 1.7 and 4.1 times higher than the estimated ones. In the case of Danuta, the value found was 0.00053, or once in 2000, which still would be high enough. The interpretation of these values is a matter of debate. Evett and Weir (1998) proposed for use in human forensics verbal equivalents for a set of probabilities. Above once in 1000 was the most extreme class of probabilities, considered to stand for 'very strong support' of the hypothesis of the profiles not matching by chance. In our case, it would mean that the suspect sample with high probability belongs to the variety (or is an EDV of the variety). There are, by far, less seedless varieties than human beings, so the strength would be even higher for the grape case. As Danuta is the least favourable case, this means that the system as a whole will be appropriate for decisions on suspect variety samples matching with any of the varieties included in the reference collection.

Each particular case requires separate study. The system might work well for one variety and not so well for another. This will depend on the rarity of the involved alleles in the reference collection. Matching probabilities and corresponding upper 95% confidence limits are given in Table 3 for a limited set of interesting varieties using the set of 5 loci. The extreme cases among the 45 seedless varieties were Danuta (see above) and Chasselas apyrene. This variety is a mutant of Chasselas blanc, a wine and table purpose variety that has not been a regular choice for breeding seedless grapevines. In contrast, Sultanina, also called Thompson seedless, *has* been included in many breeding

programs, and most of the varieties of the collection are descendants of it. Still, for this variety the probability of identity by chance is still low, once in about 15000 (or once in about 8000 using the 95% upper limit).

There were several parthenocarpic varieties within the reference collection (Table 1), with less common alleles. Their matching probabilities were quite low (see Black currant in table 3).

Some of the varieties in Table 3 are, or have been, legally protected and so their analysis can illustrate potential cases for legal disputes. The least favourable cases are Dawn and Superior seedless, which had a relatively high proportion of frequent alleles, but still their matching probabilities were low, similar to that of Sultanina. With respect to Blush and Centennial seedless, the probabilities are even much lower. Therefore, the forensic method described above, allows breeders to demonstrate the illegal exploitation of their variety. Simultaneously, a breeder can make plausible that a new variety might be an essentially derived variety.

We have followed a very conservative approach, which is convenient in a legal framework. We have calculated not only the probabilities of identity by chance using the estimated frequencies, but also using their corresponding 95% upper limits, and we have excluded four loci from the calculations because they apparently were not in HWE or LE. Nevertheless, genotypic identity for one or more of these other four loci will constitute additional evidence in favour of the hypothesis of varietal identity or essential derivation, although this additional evidence will be more difficult to quantify.

A key point of the forensic method concerns the 'population' for which allele frequencies were estimated. As we take just a sample of the reference population, it is essential to try to represent in the sample the variation present in the population. In the sample we took care to include both types of seedlessness (parthenocarpy and stenospermy), different berry colours, and different genetic relationships (full-sibs, half-sibs, parents and progeny). We checked with various statistical methods like hierarchical and model based clustering whether any group structure was present in our sample, and thus in the population. No indications for group structure were found, and thus no corrections were necessary in the calculations for the matching probabilities to take account of possible group structure (Evett and Weir, 1998).

As we were working with a sample out of the reference collection, the upper limit of the matching probability depended on the sample size. In the present sample only seedless varieties were included. However, from a genetical and breeding point of view there is no reason why seeded varieties could not be included in our sample as well, with as a consequence also the extension of the reference population. This would lower the upper limit for the matching probabilities.

A last point to consider is the selection of loci. The protocol is currently being improved by the inclusion of extra unlinked microsatellite loci, where the absence of linkage is safeguarded by the availability of genetic maps. This will allow us to use a set of loci that are in HWE and on different chromosomes for calculating matching probabilities. The result will be even lower matching probabilities and higher sensitivity of the procedure as a whole for detecting infringements on breeders´ rights.

ACKNOWLEDGEMENTS

We thank M. Palancar for her technical assistance.

Literature Cited

Bowers JE, Dangl GS, Vignani R, Meredith CP (1996) Isolation and characterization of new polymorphic simple sequence repeat loci in grape (*Vitis vinifera* L.). Genome 39:628-633

Evett IA, Weir BS (1998) Interpreting DNA Evidence. Statistical Genetics for Forensic Scientists. Sinauer Associates, Inc. Sunderland, Massachusettes

Ghosh S, Karanjawala ZE, Hauser ER, Ally DS, Knapp JI, Rayman JB, Musick A, et al. (1997) Methods for precise sizing, automated binning of alleles, and reduction of error

rates in large-scale genotyping using fluorescently labelled dinucleotide markers. Genome Res 7(2):165-178
Guo S, Thompson E (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-372
Ibáñez J (2001) Mathematical analysis of RAPD data to establish reliability of varietal assignment in vegetatively propagated species. Acta Hort 546, pp 73-79
Lodhi MA, Ye GN, Weeden NF, Reisch BI (1994) A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. Plant Mol Biol Rep 12:6-13
Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2.000: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland
Sefc KM, Regner F, Turetschek E, Glössl J, Steinkellner H (1999) Identification of microsatellite sequences in Vitis riparia and their applicability for genotyping of different *Vitis* species. Genome 42:367-373
Slatkin M, Escoffier L (1996) Testing for linkage disequilibrium in genotypic data using the EM algorithm. Heredity 76: 377-383
Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). Theor Appl Genet 86:985-990
UPOV: International Union for the Protection of New Varieties of Plants (1991) International Convention for the Protection of New Varieties of Plants
Weir BS (1996). Genetic Data Analysis II. Sinauer Associates, Inc. Sunderland, Massachusettes

Tables

Table 1. List of seedless grape cultivars used.

A-406-49 SRLH*	Corinto blanco*	Marroo seedless
Alvina	Chasselas apyrene	Moscatuel
Autumn seedless	Danuta	Pasiga
Basile Logothetis	Dawn seedless	Perlette
Bayad	Delight	Perlón
Beauty seedless	Emerald seedless	Pirovano 166A
Black Currant*	Emperatriz	Rodi
Black monukka	Fantasy	Ruby seedless
Black seedless	Flame seedless	Rutilia
Blush seedless	Graziella I	Selección Bruni 1
Bruni 116	Italia x Sultanina V-6	Slavjanka
Bruni 45	Italia x Sultanina VI-4	Sultana Crimson
Canner	Kischmisch Ali blanc	Sultana moscata
Cape currant*	Madina	Sultanina
Centennial seedless	Maria Pirovano	Superior seedless

* Parthenocarpic varieties

Table 2. Highest estimated frequencies across the sample of 45 seedless grape varieties + 95% upper limits, for 9 microsatellite loci

Locus	VVS1	VVS2	VVS5*	VVS29	VVMD5	VVMD7	ZAG47	ZAG62	ZAG79
Estimated frequency	0.4594	0.0978	0.1878	0.4356	0.1711	0.1867	0.197	0.2178	0.1106
95% Upper limit	0.5692	0.1360	0.2622	0.4794	0.2226	0.2383	0.2527	0.2985	0.1528

* Not in Hardy-Weinberg Equilibrium

Table 3. Matching probabilities using estimated allele frequencies and 95% upper limits

	Estimated frequencies			95% upper limits	
	Probability	Chance: 1 in		Probability	Chance: 1 in
Black Currant	0.000000267	3.745.318		0.000000879	1,137,656
Blush seedless*	0.000019537	51.185		0.000041459	24,120
Centennial seedless*	0.000039299	25.446		0.000084411	11,847
Chasselas apyrene	0.000000004	250.000.000		0.000000014	71,428,571
Dawn seedless*	0.000065755	15.208		0.000126029	7,935
Danuta	0.000309990	3.226		0.000534128	1,872
Flame seedless	0.000164112	6.093		0.000293817	3,403
Marroo seedless	0.000035064	28.519		0.000076765	13,027
Perlette	0.000014838	67.395		0.000029993	33,341
Sultanina	0.000066960	14.934		0.000123764	8,080
Superior seedless*	0.000068439	14.612		0.000137319	7,282

* Varieties that are or have been legally protected