Theor Appl Genet (1999) 98: 171–177 

C. Tessier · J. David · P. This
J. M. Boursiquot · A. Charrier

# Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L.

Received: 22 April 1998 / Accepted: 9 June 1998

**Abstract** The aim of this study was to develop a cultivar identification tool based on molecular analysis and a statistical approach. From the PIC parameter we defined the D parameter, which evaluates the efficiency of a primer for the purpose of identification of varieties; i.e. the probability that two randomly chosen individuals have different patterns. D can be used to compare different types of markers even if only the allelic frequencies are known. We used this parameter to develop an algorithm for selecting the optimal combination of primers necessary to identify a set of varieties. The optimal combination of primers determined for a small elite group of varieties applied on a larger set induces a risk of confusion involving 1 of the elite varieties. We estimated the risk of confusion using the D value of each primer of the combination. We applied this methodology on a set of 224 varieties of *Vitis vinifera* screened with 21 RAPD primers and two microsatellite loci. The discriminating power of the primers did not only depend on the number of patterns it generates but also on the frequencies of the different patterns. A combination of 8 primers (6 RAPD and two microsatellite) was found to be optimum for the discrimination of these 224 varieties. A subset of 38 elite varieties was also investigated. The determined optimal combination of 4 primers (3 RAPD and one microsatellite) applied on the 224 varieties gave 9 risks of confusion involving 1 of the elite varieties. Confusion can happen between varieties with the same origin as well as between varieties of very diverse geographical origins.

Communicated by G. Wenzel

C. Tessier · P. This · J. M. Boursiquot · J. David
A. Charrier (✉)
U. R. GAP, INRA-ENSAM, 2 Place Viala,
34060 Montpellier, France

**Key words** Varietal identification · RAPD · Microsatellite · *Vitis vinifera* L.

## Introduction

The grape vine (*Vitis vinifera* L.) is a vegetatively propagated plant. More than 6000 varieties have been identified on the basis of their ampelographical characters, i.e. morphological criteria (Alleweldt and Possingham 1988). Since the grape vine is of important economical value, the viticulture industry has been interested in the identification and conformity analysis of the different vegetatively propagated lines. In particular, problems revolve around the identification of young plants during the process of multiplication, international exchanges and disputes between wine growers and nurseries as well as the concerns of breeders with the protection of varietal names, especially with respect to the table grape market. Classical phenotypic methods of identification are not always sufficient to solve these problems because of the instability of the morphological characters (clonal and environmental variability, Levadoux 1954), as well as an inability to use such information for identification at juvenile stages or of isolated plant parts. Isozyme (Benin et al. 1988; Wolf 1976) and restriction fragment length polymorphisms (RFLPs) (Bourquin et al. 1993; Gorgocena et al. 1993), as well as random amplified polymorphic DNAs (RAPDs) and microsatellites have been widely used for identifying grapevine varieties (Thomas et al. 1994; Botta et al. 1995; Moreno et al. 1995), but such studies do not focus on the important problem, that is the way to optimally apply these new techniques for elite variety identification purposes. In particular, the greatest challenges are to reduce the cost of analysis, (i.e. the number of amplifications, and thus the number of primers) as well as the risk of confusing one of these elite genotypes with a randomly chosen genotype taken from a larger sample.

The aim of the study presented here was first to develop a parameter by which the efficiency of a primer (used alone or in combination with others) can be evaluated for the purpose of identification of varieties. We then proposed an algorithm to select the optimal combination of primers necessary to identify a set of varieties and looked for a link between the combination of primers and the efficiency of these primers. Finally, we evaluated the risk of confusion when applying on a set of 224 varieties the primers necessary for the discrimination of a reduced set of elite varieties constituted by the 38 most cultivated varieties in France.

## Materials and methods

### Plant material

All of the plant material used in this study belongs to the species *Vitis vinifera*. It was sampled from the INRA grape collection "Le Domaine de Vassal" (France). This collection consists of more than 2200 varieties from 35 countries. Two hundred and twenty-nine genotypes were sampled to represent the morphological diversity and geographical origins available in the collection (Table 1). These 229 genotypes represent 224 varieties since for several of them, two to four, clones were collected. The 38 most cultivated varieties in France also form part of this sample.

### Molecular methods

We used young leaves harvested in the spring and stored at $-80°C$. DNA extractions was carried out with leaf extracts according to Bowers et al. (1993). Twenty-one RAPD primers (Bioprobe, Montreuil-sous-Bois, France) and two microsatellite loci (VVS1, Thomas and Scott 1993 and VVS29, Thomas et al. 1994) were used (Table 2). Both RAPD and microsatellite amplifications were achieved on a Biomed thermocycler (Theres, Germany) according to the protocols of This et al. (1997) and Loureiro et al. (1998), respectively.

### Scoring of marker genotypes

Intense and reproducible RAPD bands were scored by a 1/0 system. Because of the codominance of the markers, microsatellites were scored as homozygotic and heterozygotic genotypes.

## Methodology

To compare the efficiency of the markers in varietal identification, we estimated the discriminating power ($D$) of each primer. If $C$ is the confusion probability, i.e. the probability that two randomly chosen individuals from the sample of 224 varieties have identical banding patterns, then $D = 1 - C$ represents the probability that two randomly chosen individuals have different patterns, and thus are distinguishable from one another.

In a set of $N$ individuals, it is possible to draw $N(N-1)/2$ different pairs. For the $i$th pattern of the given $j$th primer, present at frequency $p_i$ in this set of varieties, the confusion probability $c_i$ is:

$$c_i = p_i \frac{(Np_i - 1)}{N - 1}.$$

For the $j$th primer, the confusion probability $C_j$ is equal to the sum of the different $c_i$ for all $I$ patterns generated by the primer:

$$C_j = \sum_{i=1}^{I} c_i = \sum_{i=1}^{I} p_i \frac{(Np_i - 1)}{N - 1}$$

Thus, the discriminating power of the $j$th primer is equal to:

$$D_j = 1 - C_j = 1 - \sum_{i=1}^{I} p_i \frac{(Np_i - 1)}{N - 1}.$$

As $N$ tends towards infinity, the limit of $D_j$, equal to $D_L$, provides an estimate of the discriminating power of each of the $j$ primers.

$$\lim(D_j) = \lim\left(1 - \sum_{i=1}^{I} p_i \frac{(Np_i - 1)}{N - 1}\right) = D_L = 1 - \sum_{i=1}^{I} p_i^2$$

This is an extension of the Polymorphism Information Content or PIC (Anderson et al. 1993) available from the frequencies of the different banding patterns (or genotypes) generated by a primer. For any type of primers (RAPD, microsatellite, AFLP …) the frequencies of the different markers (or alleles) can be used to obtain the expected frequencies of the different patterns (or genotypes), and thus an estimation of the discriminating power $D_L$.

Theoretically, the total number of non-differentiated pairs of varieties for the $j$th primer is given by $x_j = (N(N-1)/2)\, C_j$. For a given combination of $k$ primers, this number $X_k$ is equal to the product of the $x_k$'s, under the hypothesis of independence of the considered primers patterns:

$$X_k = \frac{N(N-1)}{2} \prod_{j=1}^{k} C_j.$$

To find the optimal combination of primers among $n$ available primers, we then chose the primers one after the other, and minimized at each step $X_k$, the number of non-differentiated pairs of varieties for the given primer combination. In the first step, we retained the primer which differentiated the largest number of pairs among the $N(N-1)/2$ pairs, i.e. the primer which maximizes $D$. In the second step, numerical tests of each of the $n-1$ remaining primers in association with the previously chosen primer were carried out to determine the most efficient combination of two primers, that is to say the combination that reduced the most the number of undifferentiated pairs. In the following steps we applied the same principle to determine for each subsequent primer which should be retained and which should be discarded. This procedure gave, at each step, the total number of non-differentiated pairs of varieties for each primer alone (in the first step) and in combination with the previously chosen primers (in the following steps).

Letting $x$ be the size of a subsample of the set of size $N$, the number of possible confusions between a variety of this subsample and a variety of the complementary sample of size $N - x$, under the hypothesis that the frequency of the different patterns (or genotypes) is identical in both subsamples, is:

$$E = x(N - x) \prod_{j=1}^{k} C_j$$

As when working on pattern frequencies, the discriminating power $D$ can be used to compare the efficiency of different types of markers. As we had only two microsatellite loci in our study, we used data published for four other microsatellite loci (Bowers et al. 1996) on a set of 70 varieties of *Vitis vinifera* to compare the efficiency of the RAPD and microsatellite markers.

Bowers et al. (1996) did not take the homozygotic genotypes into account in calculating the allelic frequencies, i.e homozygotic count only for one copy of the corresponding allele. At a locus, the sum of the frequencies of the alleles are thus not equal to 1. We thus recalculated allele frequencies by adding the missing frequency

**Table 1** Names and geographical origins of the 224 varieties of *Vitis vinifera* used in this study

*Unknown origin*
Burgrave de Hongrie

*Afghanistan*
Kandari
Naosé

*Algeria*
Ahmeur bou ahmeur
Amokrane
Farana
Toutrissin

*Argentina*
Criolla chica no. 2 (obt)[a]
Torrontès riojano (obt)

*Austria*
Sylvaner[b]

*Bulgaria*
Corinthe blanc
Dolosata
Gros Maroc
Kichmich chichraou
Mirni

*China*
Pin el pou tao

*Cyprus*
Maratheftico
Sultanine monococco[c, 3]

*Czechoslovakia*
Kichmich rond

*Egypt*
Roumi noir

*France*
Alicante Henri Bouschet[b] (obt)
Altesse
Aramon noir[b]
Argant
Aspiran gris
Aubun[b]
Béclan
Blancorna
Bonne Vituaigne
Brun argenté
Cabernet franc[b]
Cabernet-Sauvignon[b]
Canari
Castets
Chardonnay[b]
Chasselas apyrène
Chasselas blanc
Chatus
Chenin[b]
Cinsaut[b]
Clairette blanche[b]
Colombard[b]
Colombaud
Cot[b]
Counoise
Enfariné
Farbfränkisch (obt)
Folle blanche[b]
Gamay Mourot[1]
Gamay noir[b, 1]
Gibi
Gouget noir
Grec rouge
Grolleau noir[b]
Gros vert
Gueuche blanc
Jurançon noir[b]
Lacryma Christi
Madeleine Céline (obt)
Mancin
Mauzac[b]
Melon[b]
Merlot noir[b]
Meunier[b]
Mondeuse
Mouyssaguès
Muscadelle[b]
Pascal blanc
Petit Bouschet (obt)
Petit Verdot
Pignol
Précoce de Malingre (obt)
Roi des précoces
Roussanne
Saint Macaire
Sauvignon blanc[b]
Sémillon[b]
Serénèze de Moirans
Syrah[b]
Tannat[b]
Teinturier
Terret gris[b]
Tibouren
Trousseau
Viognier[b]

*Germany*
Faber (obt)
Frankenthal
Riesling[b]

*Greece*
Corinthe noir
Heptakilo
Koritsanos rouge
Kouroupitsa
Muscat blanc à petits grains[b]
Opsimos edessis
Phileri kokkino
Romaico

*Hungary*
Dinka vörös
Furmint
Harslevelu
Kadarka török
Kövidinka
Perle de Csaba (obt)
Pozsonyi feher

*Iran*
Bidaneh ghelmez
Gora chirine
Khalili piskakes
Ozaan daii
Phakri
Yaghasti

*Israel*
Hebron blanc
Marawi
Nehelescol

*Italy*
Annamaria (obt)
Avarengo
Brachetto
Catarratto bianco lucido
Ciliegiolo
Colombana nera
Corinto bianco
Corniola bianca di Milazzo
Corvina veronese
Foglia tonda
Gewürztraminer[b]
Grignolino
Malvasia di Sardegna
Marsigliana nera
Molinara
Montonico bianco
Muscat d'Alexandrie[b]
Neretta cuneese
Piccolit
Pizutello nero
Prunesta nera
Sangiovese[b]
Santa Paula
Susumaniello
Timorasso
Ugni blanc[b]
Uva di Troia
Verdeca
Vermentino

*Lebanon*
Ahmar
Asmi assouad
Assouad karech
Inab el mir

*Morocco*
Bezoul el aouda
Bouchouka
Lambrusque A
Lambrusque C
Maticha

*Portugal*
Alva
Azal tinto
Boal de Alicante moscatel (obt)
Borraçal
Dedo de dama
Fernao Pires
Galego de Montemor
Jampal
Loureiro
Monvedro
Mourisco tinto
Muscat rouge de Madère
Tavrida
Tinta pinheira
Tinto cao

*Rumania*
Armas
Braghina
Cabasma alba
Coarna neagra
Crimposie
Gordin gurguiat
Negru mare
Pamid
Plavaï
Samoveanca

*Russia*
Askeri
Basicata
Chaani biely
Chaouch blanc
Gros Colman
Katta kourgan
Kefessia
Khoussaïné blanc
Kizil
Kouldjinsky
Krasnostop zolotowskii
Pletchistik
Rhoditis
Rkatsitelli
Sabalkanskoï
Saperavi
Sourkhak kitabsky
Tarnaou (obt)
Tchiliaki blanc
Tzitzka

*South Africa*
Cape currant

*Spain*
Ariño
Bobal
Carignan noir[b]
Crujillon
Doradilla
Gorgollosa
Grenache blanc[b, 2]
Grenache noir[2]
Jaen
Listan
Macabeu[b]
Molinera gorda
Morrastel
Mourvèdre[b]
Parellada
Santa Morena
Tempranillo[b]
Turruntes

*Switzerland*
Arvine
Landroter

*Tunisia*
Bezoul el khadem
Bou rouguia

*Turkey*
Dattier de Beyrouth
Sultanine blanche[3]
Sultanine rose[3]
Sultanine rouge[3]

*United States of America*
Black rose (obt)
Mission (obt)
Red globe (obt)

*Yemen*
Bayad

*Yugoslavia*
Crvena slaubanic
Posip
Vugava

[a] Obt: Obtention
[b] Subset of 38 French elite varieties
[c] Clonal selections of a common origin followed by the same number

**Table 2** Primer discrimination power calculated (D) and estimated from the $p_i^2$ ($D_L$) on the subsample of 224 varieties

| Primer | Number of markers | Number of banding patterns | D | $D_L$ | Orders of D and $D_L$ |
|---|---|---|---|---|---|
| A9 | 4 | 13 | 0.826 | 0.822 | 2 |
| P17 | 3 | 8 | 0.765 | 0.762 | 3 |
| D16 | 3 | 8 | 0.734 | 0.731 | 4 |
| D11 | 2 | 4 | 0.726 | 0.723 | 5 |
| P2 | 4 | 11 | 0.711 | 0.708 | 6 |
| C4 | 2 | 4 | 0.684 | 0.681 | 8 |
| C13 | 4 | 11 | 0.674 | 0.671 | 9 |
| A7 | 3 | 6 | 0.642 | 0.639 | 10 |
| B17 | 2 | 4 | 0.636 | 0.633 | 11 |
| C6 | 2 | 4 | 0.605 | 0.603 | 12 |
| P10 | 1 | 2 | 0.502 | 0.500 | 13 |
| P8 | 1 | 2 | 0.500 | 0.498 | 15 |
| A20 | 2 | 4 | 0.500 | 0.497 | 14 |
| A18 | 1 | 2 | 0.471 | 0.469 | 16 |
| D4 | 2 | 4 | 0.449 | 0.447 | 17 |
| A11 | 1 | 2 | 0.410 | 0.408 | 18 |
| B1 | 2 | 4 | 0.355 | 0.354 | 19 |
| C9 | 2 | 3 | 0.224 | 0.223 | 20 |
| C15 | 1 | 2 | 0.213 | 0.212 | 21 |
| D18 | 1 | 2 | 0.156 | 0.155 | 22 |
| P1 | 1 | 2 | 0.077 | 0.077 | 23 |
| VVS1 | 10 | 29 | 0.895 | 0.891 | 1 |
| VVS29 | 5 | 9 | 0.697 | 0.694 | 7 |

(complementary to 1) proportionally to the respective frequency of the different alleles.

## Results and discussion

With the 21 RAPD primers, 44 reliable markers could be selected on the set of 224 varieties. The number of polymorphic markers per primer varied from 1 to 4 and generated 2–13 different banding patterns per primer (Table 2). The two microsatellite loci VVS1 and VVS29 generated 10 and 5 alleles, respectively, which corresponded to 29 and 9 genotypes. Among the sample of 224 varieties, 224 RAPD and microsatellite patterns were obtained. As already published (Jean-Jacques et al. 1993; Loureiro et al. 1998) no variability within cultivars was detected since two clonal variants of 'Grenache' (noir and blanc), two clonal variants of 'Gamay' (Noir and Mourot), and four clonal variants of 'Sultanine' (blanche, rose, rouge and Monoccoco) gave the same patterns. The discriminating power *D* and the optimal combination of primers were estimated from these 224 patterns, retaining only 1 individual of each clonal variant.

### Evaluation of the discrimination power of primers

The analysis of power discrimination revealed that the efficiency of a given primer does not depend only on the number of patterns it generates (Fig. 1 and Table 2). For example, even if 2 primers produce the same number of patterns, they can have very different discriminating powers, the scale of variation ranging from 1 to 10 (e.g. primers P1 and P8). On the contrary, 2 primers with quite a different number of patterns can have similar discrimination powers (e.g. primers C4 and C13; 4 and 11 patterns, respectively). This result can be explained by the frequency differences between the patterns generated with these primers. A primer has a maximal discriminating power ($D_{max}$) when it generates patterns at the same frequency (the isofrequency situation). The farther it is from this situation, the more its discrimination power diminishes. For instance, primers P8 and D11 are nearly isofrequent, and their discriminating power is thus close to $D_{max}$. In contrast, primers P1 and C13 have a skewed abundance or rare pattern (e.g. in 96% and 5% of the varieties, respectively), and this explains their relatively low discriminating power (Fig. 1). Microsatellite loci efficiency is also subject to this rule, even though a large number of genotypes seem to yield an advantage for locus VVS1 compared to RAPD primers.

### Determination of the optimal primers combination

Three optimal combinations of 8 primers were obtained to discriminate among the 224 patterns. They were composed of 6 stable primers (VVS1, A9, P17, D16, P2, VVS29), respectively primer B1 and primers (C4, C13 or D4) were needed to discriminate between the two last pairs of individuals. In this study, we chose the first 4 primers on the basis of their discrimination power (Table 2). Two hypotheses can explain the deviation observed from the fifth primer. The first is the non-statistical independence of the patterns generated by the primers. The efficiency of a primer in combination with others does not depend on discrimination

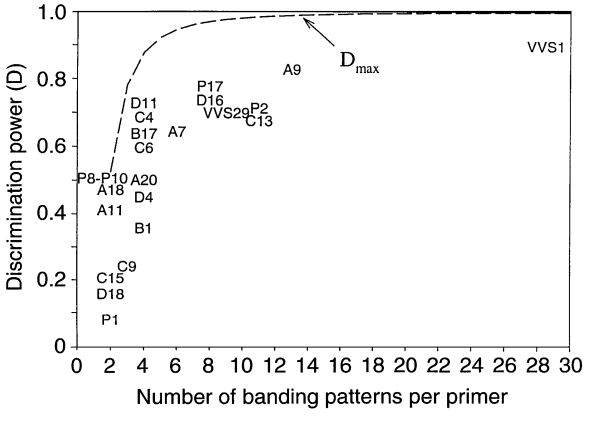


**Fig. 1** Value of the discriminating power D of the primers as a function of their number of banding patterns

**Table 3** Comparison of the real and theoretical efficiency of a primer combination under the hypothesis of independence of their patterns

| | Number of indistinguishable pairs | |
|---|---|---|
| | Experimentally observed | Expected under the independence hypothesis |
| VVS1 | 2668 | 2671.2 |
| VVS1 + A9 | 475 | 464.8 |
| VVS1 + A9 + P17 | 125 | 107.8 |
| VVS1 + A9 + P17 + D16 | 32 | 28.6 |
| VVS1 + A9 + P17 + D16 + P2 | 9 | 8.3 |
| VVS1 + A9 + P17 + D16 + P2 + VVS29 | 2 | 2.5 |
| VVS1 + A9 + P17 + D16 + P2 + VVS29 + B1 | 1 | 1.6 |
| VVS1 + A9 + P17 + D16 + P2 + VVS29 + B1 + C4 | 0 | 0.5 |

power alone but also on its independence from the set of primers already selected. The experimental values of the total number of non-differentiated pairs obtained with the optimal combination of primers are not very different from the values estimated with $D$ under the independence hypothesis (Table 3). This can explain the preferential choice of primer P2 instead of primer D11 at the fifth step. Primer D11 leaves 12 non-distinguishable pairs instead of 7.8, as expected under the independence hypothesis, and 9 and 8.3 respectively for primer P2. The second hypothesis concerns essentially the choice of the last primers. It implies that the primers are kept for their ability to discriminate the few non-distinguishable pairs and not for their global efficiency. Primers C4 ($D = 0.684$) C13 ($D = 0.674$) and D4 ($D = 0.449$) are the sole ones able to differentiate the last pair of individuals, and this despite their different discriminating powers.

## Application to published data

$D_L$ values estimated from the pattern frequencies of the primers are very close to the real discriminating power D. This result shows that $D_L$ may be achieved for a sample of relatively small size (Table 2).

The $D_L$ values estimated for 3 of the four microsatellite loci published by Bowers et al. 1996 (VVMD5 $D_L = 0.959$, VVMD6 $D_L = 0.914$ and VVMD7 $D_L = 0.923$) are higher than those for all the RAPD primers of our study ($D_L = 0.891$, Table 4). Applied to the sample of 224 varieties, locus VVMD5 would statistically leave only 1024 non-distinguishable pairs of varieties compared to 2615 for locus VVS1. Under the independence hypothesis, the 4 loci VVMD5, VVMD6, VVMD7 and VVS1 would be sufficient to discriminate the 224 patterns of our sample. But this comparison is subject to two assumptions in addition to the hypothesis concerning the discrimination of the last few non-distinguishable pairs of varieties. The first assumption is that allele frequencies are stable from one sample to another. This does not seem to be a very severe constraint for *Vitis vinifera* because of the low degree of differentiation among the cultivated varieties of the species (data not shown). The second assumption concerns the error present when recalculating the allelic frequencies. The procedure we applied can effectively overestimate the allelic frequency of the most frequent alleles and underestimate the frequency of the rarest alleles and thus genotypes. But this situation leads to a departure from the isofrequency situation (for genotypes) and, thus, to an underestimation of the discrimination power $D_L$ for these loci.

**Table 4** Discrimination power estimated on the microsatellite loci published by Bowers et al. (1996)

| | Number of alleles on 77 varieties | Number of heterozygotic genotypes | $D_L$ estimate |
|---|---|---|---|
| VVMD5 | 8 | 26 | 0.959 |
| VVMD6 | 5 | 13 | 0.914 |
| VVMD7 | 11 | 27 | 0.923 |
| VVMD8 | 6 | 13 | 0.825 |

## Evaluation of the risks of confusion

To discriminate among the 38 varieties corresponding to the French elite varieties, we obtained two combinations. In both VVS1 was associated with 3 RAPD primers (Table 5). These two combinations tested on the whole set of 224 varieties generated 47 and 66 confusions, among which 9 and 12 involved 1 of the 38 varieties. To solve these problems, we need to add 3 or 4 primers (Table 5). To identify the 38 varieties and certify that no confusion is possible with the 186 remaining varieties, we require at least 7 primers. Table 6 indicates for the most efficient primer combination the varieties most subject to confusion.

These confusions are not negligible as at least 8 of the 38 varieties were subjected to confusion. Our results show that taking only the minimum number of primers

**Table 5** Number of confusions involving 1 of the 38 elite varieties when using the primer combination necessary for their discrimination on the whole set of 224 varieties

| Primers combination | Number of confusions involving 1 of the 38 elite varieties | Primers required to avoid these confusions | Total number of primers |
|---|---|---|---|
| A9 + VVS1 + D11 + C4 | 9 | D16 + C13 + A18 | 7 |
| A9 + VVS1 + D11 + C13 | 12 | C6 + D4 + P17 + B1 | 8 |

**Table 6** Couples of varieties involving 1 of the 38 elite varieties when applying the primer combination (A9 + VVS1 + D11 + C4) on the whole set of 224 varieties

| Varieties of the elite subset | Varieties among the 186 remaining varieties |
|---|---|
| Aramon noir | Samoveanca |
| Viognier | Colombana nera |
| Grenache noir | Listan |
| Riesling blanc | Madeleine Céline |
| Terret gris | Roi des précoces |
| Gewürztraminer | Grignolino |
| Melon | Pozsonyi feher |
| Mourvèdre | Corniola bianca di Milazzanó |
| Mourvèdre | Maratheftico |

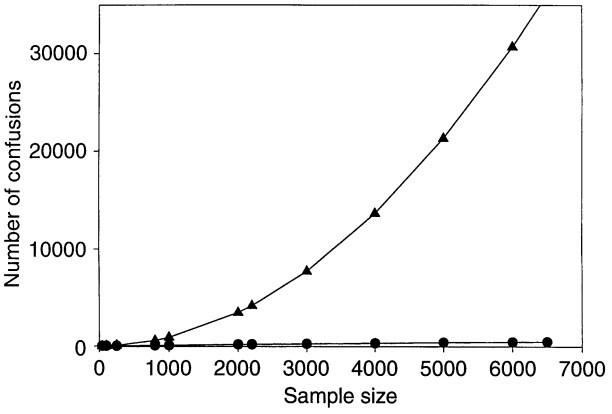


**Fig. 2** Estimation of the number of confusions, either total (●) or involving 1 of the 38 elite varieties (▲), when applying the combination of primers A9 + VVS1 + D11 + C4 on an increasing number of varieties

necessary to identify a set of varieties may not be sufficient to avoid confusion with a larger sample. This has to be taken into account when placing confidence in expertise results at both the international and local level of investigation. Confusions can happen between varieties of the same origin, 'Grenache' and 'Listan' for example, or between varieties of very diverse geographical origins, 'Mourvèdre' and 'Maratheftico' or 'Aramon noir' and 'Samoveanca' for example (Table 6).

If we use the 4 primers necessary to identify the 38 varieties on a larger sample, the total number of confusions grows exponentially with the size of this sample (Fig. 2). For example, 3491 confusions are expected in a sample of 2200 varieties, the size of the collection at "Le Domaine de Vassal", and 25 980 confusions in a sample of 6000 varieties, which is the approximate overall number of *Vitis vinifera* varieties. The 4 primers published by Bowers et al. 1996 associated with the locus VVS1 should permit the avoidance of any risk of confusion between 1 of the 38 elite varieties and any varieties of a sample of 2200 varieties. Only 1 more locus would theoretically be necessary for a sample of 6000 varieties as the 5 pre-cited loci leave only 1 possibility of confusion. Moreover, these numbers of confusions can be underestimated, as in our study where 47 confusions were observed for 31 expected on the basis of the $D$ values. The number of confusions involving 1 of the 38 varieties and 1 of the other varieties evolves linearly with the size of the second sample. The estimation obtained with our 224 varieties sample is precise enough, 10.2 confusions estimated for 9 in reality. We would find 118.6 and 327.1 confusions when using the 4-primer combination determined (Table 5) on a 2200 and 6000 variety sample respectively.

## Conclusion

This study shows that confusion risks must not be neglected in varietal identification. It also emphasizes the necessity of a good basic knowledge of the varietal diversity of a species before choosing the best primer combination for a small set of varieties. The discriminating power $D$ can be considered to be a good estimator of the efficiency of a primer or a locus. It allows one to compare different types of molecular markers. It also can be used to predict the efficiency of primers taken in combination, and the risks of confusion due to the use of this combination. This parameter can be of great interest for varietal identification by molecular techniques especially to evaluate the cost in terms of amplifications.

**Acknowledgements** The suggestions of Dan Schoen, Joëlle Ronfort and Eric Jenczewski to improve an earlier draft of the manuscript are very much appreciated. We particularly thank the Direction Générale de l'Enseignement et de la Recherche for supporting this work.

## References

Alleweldt G, Possingham JV (1988) Progress in grapevine breeding. Theor Appl Genet 75:669–673

Anderson JA, Churchill GA, Autroque JE, Tanksley SD, Swells ME (1993) Optimising selection for plant linkage map. Genome 36:181–186

Benin M, Gasquez J, Mahfoudi A, Bessis R (1988) Caractérisation biochimique des cépages de *Vitis vinifera* L. par électrophorèse d'isoenzymes foliaires: essai de classification des variétés. Vitis 27 : 157–172

Botta R, Scott NS, Eynard I, Thomas MR (1995) Evaluation of microsatellite sequence-tagged site markers for characterizing *Vitis vinifera* cultivars. Vitis 34 : 99–102

Bourquin JC, Sonko A, Otten L, Walter B (1993) Restriction fragment length polymorphism and molecular taxonomy in *Vitis vinifera* L. Theor Appl Genet 87 : 431–438

Bowers JE, Bandmann EB, Meredith CP (1993) DNA fingerprinting characterization of some wine grape cultivars. Am J Viticult Enol 44 : 266–27

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

Gorgocena Y, Arulsekar S, Dandekar AM, Parfitt DE (1993) Molecular markers for grape characterization. Vitis 32 : 183–185

Jean-Jaques I, Defontaine A, Hallet JN (1993) Characterization of *Vitis vinifera* cultivars by Random Amplified Polymorphic DNA markers. Vitis 32 : 189–190

Levadoux L (1954) La connaissance des cépages. Cahiers vitivinicoles de la revue Chambres d'Agriculture [Suppl] 49 : 1–9

Loureiro D, Martinez C, Boursiquot JM, This P (1998) Molecular markers analysis of *Vitis vinifera* 'Albariño' and some similar grapevine cultivars. J Am Soc Hortic Sci 123 : 842–848

Moreno S, Gorgocena Y, Oritz JM (1995) The use of RAPD markers for identification of cultivated grapevine (*Vitis vinifera* L.). Sci Hortic 62 : 237–243

This P, Cuisset C, Boursiquot JM (1997) Development of stable molecular markers based on RAPD for the identification of grapevine rootstocks. Am J Enol Viticult 48 : 492–501

Thomas M, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). Theor Appl Genet 86 : 985–990

Thomas M, Cain P, Scott NS (1994) DNA typing of grapevines: a universal methodology and database for describing cultivars and evaluating genetic relatedness. Plant Mol Biol 25 : 939–949

Wolf WH (1976) Identification of grape varieties by isozyme banding pattern. Am J Enol Viticult 27 : 68–73