Theor Appl Genet (1993) 86:985–990

THEORETICAL AND APPLIED GENETICS

Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs)

M. R. Thomas, N. S. Scott

CSIRO Division of Horticulture, GPO Box 350, Adelaide, SA 5001, Australia

Received: 1 October 1992 / Accepted: 15 January 1993

Abstract. Microsatellite repeat sequences were investigated as sequenced-tagged site (STS) DNA markers to determine the potential for genetic analysis of the grapevine genome. The PCR-generated markers detect codominant alleles at a single locus or site in the genome. The marker type is very informative detecting high heterozygosity (69%–88%) within individual grapevine cultivars and high genetic variation between cultivars, making it a useful marker type for plant genome mapping and genome typing. For five loci a screening of 26 *V. vinifera* cultivars found 13, 12, 8, 5, and 4 different length alleles respectively with some alleles more common than others. The genomic DNA sequences surrounding microsatellite sequences were conserved within the genus permitting STS primers to amplify STSs from other *Vitis* species. These *Vitis* species were found to have some unique alleles not present in *V. vinifera*.

Key words: Grapevine – Microsatellite DNA – DNA polymorphisms – STS – SSLP – STMS – *Vitis*

Introduction

DNA molecular markers represent a significant resource for creating genetic and physical genome maps, distinguishing individuals, investigating genetic relatedness, and studying genome organisation. Restriction fragment length polymorphism (RFLP) has been the dominant method or DNA marker type since the early 1980s (Soller and Beckmann 1983). The type of DNA sequence used for probes in RFLP analysis range from single-copy cDNAs to multi-locus repetitive non-coding sequences, with choice dependant on the objective of a study and the degree of polymorphism detected by the different DNA sequences. Recently there has been interest in the polymerase chain reaction (PCR) method for providing new DNA marker types that are easier to produce and analyse. The PCR marker types fall into two groups based on primer design; those known as sequence-tagged sites (STSs) (Olson et al. 1989), with PCR primers designed from a known sequence, and those based on arbitrary primers (Welsh and McClelland 1990; Williams et al. 1990). The most informative or polymorphic STS marker appears to be one that amplifies a DNA region containing a microsatellite repeat sequence (Litt and Luty 1989; Tautz 1989; Weber and May 1989). Such an STS-based marker has been referred to as a simple-sequence length polymorphism (SSLP, Tautz 1989) or sequence-tagged microsatellite site (STMS; Beckmann and Soller 1990).

Microsatellite DNA consists of a small repeat unit, generally less than four nucleotides, that generates repeating regions less than 100 bp. These regions are highly interspersed throughout eukaryotic genomes (Hamada et al. 1982; Tautz and Renz 1984; Stallings et al. 1991) with some eukaryotic genomes suggested to have a microsatellite sequence distributed once every 10 kb (Tautz 1989). Microsatellite STSs appear to be the marker of choice for creating high density maps for human (e.g., Kwiatkowski et al. 1992), mouse (e.g., Love et al. 1990) and other mammalian genomes (e.g., Moore et al. 1991). An advantage of STS-type markers is the codominant mode of inheritance permitting easy transfer of markers between genetic maps of different crosses in contrast to the dominant PCR marker type

Communicated by J. S. Beckmann
Correspondence to: M. R. Thomas

based on arbitary primers which requires the generation of a new map for each cross. The STS marker type has also proved useful for genome typing individuals and studying genetic variation (e.g., Schlotterer et al. 1991; Edwards et al. 1992) and should have a similar role in agriculture, particularly with vegetatively-propagated crops like grapevine.

Knowledge about microsatellite sequences in plants is scarce (Beyermann et al. 1992; Thomas et al. 1992) with no information available on the application of STMSs for plant genome analysis. The present study characterises STMSs from the genome of grapevine (*Vitis*) and shows that they represent a highly polymorphic marker type well suited for the genetic analysis of the grapevine genome and presumably other plant genomes.

Materials and methods

Plant material used for nucleic acid extraction was obtained from *Vitis* collections in Australia located at the CSIRO Division of Horticulture, Merbein, Victoria and the South Australia Department of Agriculture Nuriootpa Research and Advisory Centre, Nuriootpa, South Australia. Both collections have individual vine identification numbers as well as cultivar accession/import numbers. This information is available on request. DNA was extracted as previously described (Thomas et al. 1993).

Cloning, selection and sequencing of microsatellite sequences

DNA isolated from *V. vinifera* (L.) cv Sultana was digested with the restriction enzyme *Taq*I, ligated into the *Acc*I-cut dephosphorylated (HK Phosphatase; Epicentre Technologies) vector pBluescript SK(+) (Stratagene) and used to transform competent *E. coli* Sure (Stratagene) cells. Colony lifts were performed according to the procedure supplied with the nylon membrane (Bio-Rad). Oligonucleotide probes $(GA)_{10}$ and $(GT)_{10}$ were hybridized to the membranes as described previously (Thomas et al. 1993) to select colonies containing similar sequences. The hybridization solution for oligonucleotide probes was 0.5 M Na_2HPO_4 (pH 7.2), 1 mM EDTA, 7% sodium dodecyl sulphate (SDS) and 10 µg/ml yeast tRNA. Hybridization temperature for the individual oligonucleotides was 5 °C below the T_d values of 56 °C for $(GT)_{10}$ and 50 °C for $(GA)_{10}$. Double-stranded DNA sequencing of plasmid clones was performed with a Sequenase Version 2 kit (USB) using alpha [^{35}S] dATP (43.7 TBq/mmol).

Polymerase chain reactions

Selection of oligonuleotide primers flanking microsatellite sequences was done with the use of a computer program (Rychlik and Rhoads 1989) prior to synthesis. Oligonucleotides were synthesised with an Applied Biosystems 391 DNA synthesiser. PCR products were labelled during synthesis with [^{35}S] dATP in a 10-µl PCR mix containing 0.5 ng of plasmid or 25 ng *Vitis* genomic DNA, 0.625 U of *Taq* DNA polymerase (Promega), buffer (Promega), 1.5 mM $MgCl_2$, 25 pmol of each primer, 111 KBq alpha [^{35}S] dATP (43.7 TBq/mmol), 200 µM of dGTP, dCTP, dTTP and 2.5 µM of dATP. After 5 min at 95 °C, 26 cycles of PCR were performed (denaturation: 1 min at 94 °C; annealing: 50 s at 50 °C; ramp: 1 min from 50–72 °C; elongation: 1 min at 72 °C) followed by a 7-min elongation step using a Perkin-Elmer Cetus automated thermal cycler. An equal volume of formamide, containing 0.5% xylene cyanol, was added to individual PCR reactions. Samples were denatured at 95 °C for 3 min prior to loading onto 40-cm long sequencing gels (5% polyacrylamide, 7 M urea). Fixed and dried gels were exposed of Fuji RXG film for 24–76 h.

Results

Characterisation of microsatellite sequences

Our previous study (Thomas et al. 1993) demonstrated that the grapevine genome has many different microsatellite classes, $(GT)_n$, $(GA)_n$, $(CAC)_n$, $(GACA)_n$, and $(GATA)_n$, with the microsatellite GA and GT dinucleotide repeat classes highly represented and dispersed through the genome. Individual microsatellite repeats suitable for use as sequence-tagged site DNA markers were isolated from a Sultana genomic library. The library was made from *Taq*I-digested DNA to generate clones with small insert sizes to facilitate subsequent sequencing. Clones containing microsatellite repeats had insert sizes between 0.2 and 2.5 kb and 1.2% and 0.5% of colonies were positive for $(GA)_n$ and $(GT)_n$ microsatellites respectively.

Several different categories of repeats (see Weber 1990) were found in the microsatellite-positive clones sequenced so far. There were seven perfect repeats having no interruptions, the smallest of which was $(CT)_8$, seven imperfect repeats interrupted by no more than three consecutive non-repeat bases and five compound repeat sequences where two or more repeat runs were adjacent. The smallest repeat sequenced was a $(GT)_6$ (e.g., VVS3, Table 1) indicating that conditions used for screening the library allowed selection of repeats smaller than the $(GT)_{10}$ oligonucleotide probe.

Microsatellite sequence-tagged site polymorphisms

Conversion of sequenced clones containing microsatellite DNA into a sequence-tagged site marker involved the design of primers flanking the microsatellite repeats such that the microsatellite repeat and flanking DNA region between the primers could be amplified in a PCR reaction. The primers and the repeat type for the different microsatellite loci are presented in Table 1. In addition to $(GA)_n$ and $(GT)_n$ repeats, other microsatellite repeat types were found fortuitously in some clones during sequencing and could also be turned into STS markers (e.g., VVS1, Table 1). The size of the PCR products was kept between 100–220 bp by judicious choice of primer site to allow accurate sizing and differentiation, within 1 bp, of PCR products when run on sequencing gels.

The microsatellite STS analysis detects loci that are polymorphic, allowing *V. vinifera* cultivars to be distin-

Table 1. *V. vinifera* microsatellite loci

Locus	Primers	Repeat type[a]	Length classes[b]	Heterozygosity[b]
VVS1	ACAATTGGAAACCGCGTGGAG CTTCTCAATGATATCTAAAACCATG	Imperfect A_{17}, Imperfect $(TAA)_7$	8	69%
VVS2	CAGCCCGTAAATGTATCCATC AAATTCAAAATTCTAATTCAACTGG	Perfect $(GA)_{22}$	13	88%
VVS3	TGCCCTATCAATTAGTTCACCTA TCGACTTTGATATATTGATGATT	Compound $(GTAT)_4\,(GT)_6$	4	73%
VVS4	CCATCAGTGATAAAACCTAATGCC CCCACCTTGCCCTTAGATGTTA	Perfect $(AG)_9$	5	71%
VVS5	ATTGATTTATCAAACACCTTCTACAT TAGAAAGATGGAAGGAATGGTGAT	Perfect $(GT)_{21}$	12	88%

[a] Repeat type present in sequenced clone
[b] Observed values based on a screening of 26 *V. vinifera* cultivars

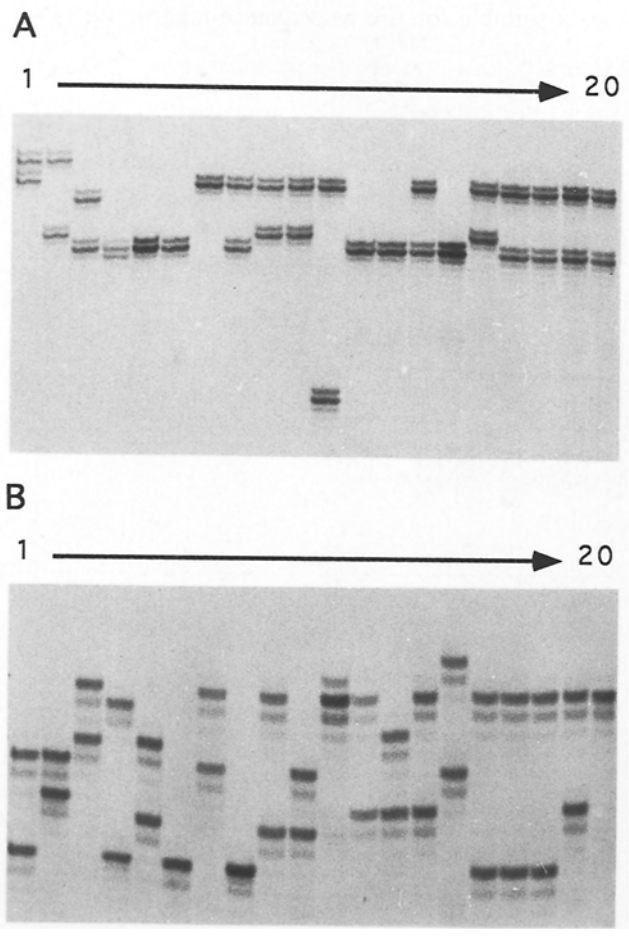

Fig. 1A, B. STMS genotypes of grapevine cultivars at the VVS1 locus (**A**) and VVS2 locus (**B**). *Lane 1*, Villard Blanc; *lane 2*, Villard Noir; *lane 3*, Sultana; *lane 4*, Muscat Gordo Blanco; *lane 5*, Grenache; *lane 6*, Shiraz; *lane 7*, Riesling; *lane 8*, Semillon; *lane 9*, Pinot Noir; *lane 10*, Chardonnay; *lane 11*, Traminer; *lane 12*, Cabernet Sauvignon; *lane 13*, Cabernet Franc; *lane* 14, Fer-Servadou; *lane 15*, Petit Verdot; *lane 16*, Malbec; *lane 17*, Sauvignon Blanc; *lane 18*, Sauvignon Rose; *lane 19*, Merlot; *lane 20*, Merlot Blanc

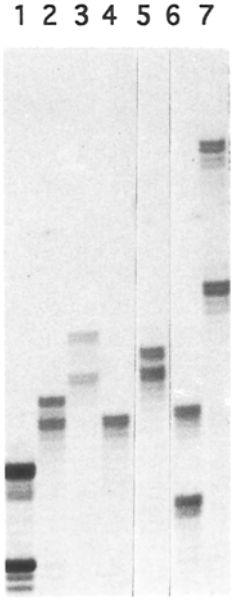

Fig. 2. STMS genotypes of non-*V. vinifera* species at the VVS2 locus. *Lane 1*, *V. labrusca*; *lane 2*, *V. champini*; *lane 3*, *V. riparia*; *lane 4*, *V. rupestris*; *lane 5*, *V. longii*; *lane 6*, *V. amurensis*; *lane 7*, *V. rotundifolia*

guished based on their genotype at different STS loci (Fig. 1). The two STMS loci in Fig. 1 show that different degrees of polymorphism are detected at different loci within the same population of cultivars. Most of the grapevine cultivars are heterozygous with two different-sized alleles, with only a few homozygous for a single allelic length. Only the cultivar Shiraz (lane 6) is homozygous at both loci. Combining the genetic variation observed for these two loci produces a unique STMS-combined genotype for all the cultivars shown in Fig. 1 except for Fer-Servadou (lane 14) and Merlot (land 19). DNA typing by combining the five loci identified in Table 1 has allowed all grapevine cultivars

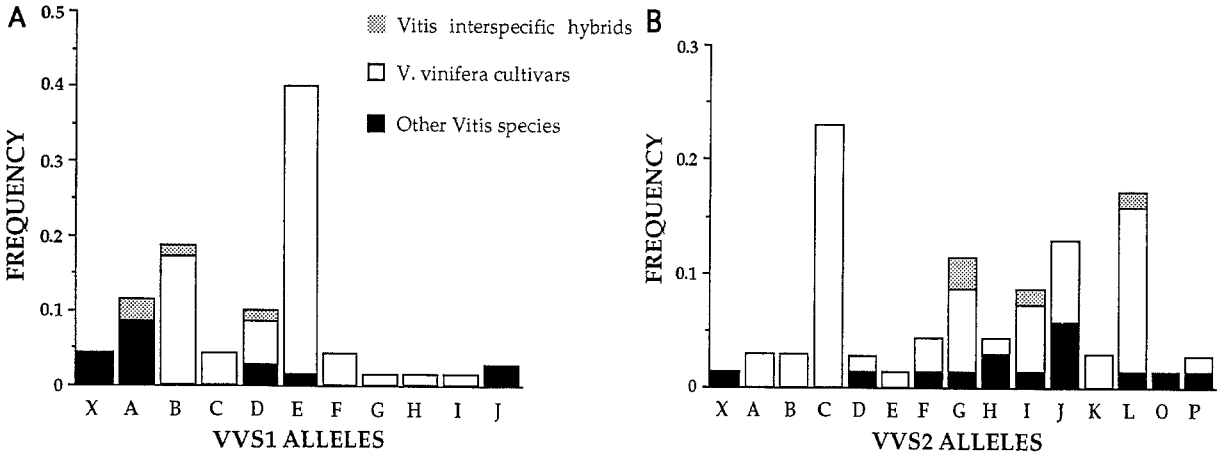

Fig. 3A, B. Distributions of microsatellite alleles at the VVS1 (**A**), and VVS2 (**B**) locus in *Vitis*. Each allele size in Figs. 3 and 4 is represented by a *letter*

tested so far to be separated. Lanes 17 and 18, with identical STMS genotypes, represent two Sauvignon clones or selections, with the Sauvignon Rose sport having rose coloured berries instead of the normal white berries. The different berry colour is thought to be due to the occurrence of a somatic mutation.

The locus VVS1 is an example of an STMS containing imperfect repeats whilst the locus VVS2 contains a perfect repeat (Table 1). The primers for STMS analysis (Table 1), designed from sequenced *V. vinifera* cv Sultana DNA, also work in other *Vitis* species (Fig. 2) indicating the versatility of STMS markers within a genus. Even *V. rotundifolia*, which has been placed in a separate section (*Muscadinia*) to the other *Vitis* species (*Euvitis* section) analysed, has sufficient DNA homology at the primer sites to allow efficient amplification of these microsatellite sites (Fig. 2).

Genotype and allele frequency distributions

The graphs in Fig. 3 show the relative frequency profiles of STMS alleles at two loci in the *Vitis* population analysed. Both loci are multi-allelic with 11 different length alleles detected at the VVS1 locus and 15 at the VVS2 locus. Some alleles are more prevalent than others in the population, and a feature of both loci is the presence of some alleles unique to only non-*V. vinifera* species and others unique to *V. vinifera*. The majority of the non-*V. vinifera* species originate from Noth America with one, *V. amurensis*, originating from Japan, while *V. vinifera* cultivars are of Asia minor origin (Alleweldt et al. 1990). As one would expect, complex interspecific hybrids between *V. vinifera* cultivars and non-*V. vinifera* species share alleles common to both groups (Fig. 3).

An example of genotype distribution at the VVS2 locus for *V. vinifera* cultivars only is shown in Table 2.

Table 2. *V. vinifera* genotype distributions at the VVS2 STMS locus

Alleles	X	A	B	C	D	E	F	G	H	I	J	K	L	O	P
X	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
A		–	–	–	–	–	–	1	–	–	–	–	–	–	1
B			–	1	–	–	–	–	–	–	–	–	1	–	–
C				1	–	–	1	3	–	3	2	1	3	–	–
D					–	–	–	–	–	–	–	–	1	–	–
E						–	–	–	–	1	–	–	–	–	–
F							–	–	–	–	1	–	–	–	–
G								–	1	–	1	–	–	–	–
H									–	–	1	–	–	–	–
I										–	–	–	–	–	–
J											–	–	–	–	–
K												–	1	–	–
L													2	–	–
O														–	–
P															–

Possible genotypes = 91. Observed genotypes = 19
Heterozygotes = 23. Homozygotes = 3

An analysis of 26 cultivars produced 19 different observed genotypes out of a possible 91. It is likely that more of the possible genotypes will be found as more cultivars are DNA typed but it may also be that there has been a selection, natural or man-made, for cultivars possessing certain chromosome combinations. Table 1 shows the number of length alleles detected at different STMS loci. The two longer dinucleotide loci are genetically more variable and detect more alleles than those with short repeat lengths (Table 1), an observation in agreement with those of Weber (1990). In the grapevine STMS loci tested so far heterozygosity values are high and those in Table 1 range from 69% to 88% despite the different number of allele-length classes detected in each one.

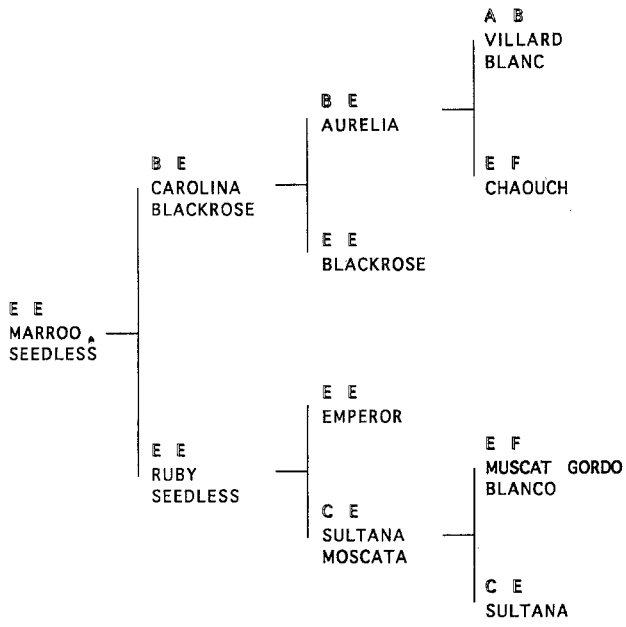

Fig. 4. Pedigree of Marroo Seedless and segregation of VVS1 alleles

Segregation of STMS alleles

Grapevine STMS alleles are codominant DNA markers that follow expected Mendelian autosomal inheritance. Figure 4 shows the flow of VVS1 alleles through the pedigree of Marroo Seedless a table-grape cultivar recently bred at CSIRO (Clingeleffer and Possingham 1988). Marroo Seedless is a single progeny from the cross between Carolina Blackrose and Ruby Seedless. As grapevine cultivars are propagated vegetatively the STMS genotype for Marroo Seedless, as well as those of all other grapevine cultivars, can be expected to remain fixed. Analysis of progeny from the Carolina Blackrose (BE) and Ruby Seedless (EE) cross segregated for the expected genotype of either EE or BE at the VVS1 locus.

Discussion

The highly-abundant microsatellite sequences in the grapevine genome can be used as informative sequence-tagged site molecular markers. The examples presented show that both perfect, imperfect and compound microsatellite repeat types are useful both for detecting heterozygosity within individuals and genetic variation between individuals. The degree of genetic polymorphism detected is high at some individual STMSs, especially those sites with long repeat lengths. The simple interpretation and genetic analysis of single-locus markers like STMSs make them superior to multi-locus DNA marker types, especially for map construction and DNA typing. Combining STMS genotype data from a number of sites has the potential to provide a unique DNA identification or DNA profile for individual cultivars, with all of the grapevine cultivars tested so far found to be genetically distinct based on the combined genotype data from the five loci detailed in Table 1.

The *V. vinifera* cultivar population tested is seen to be genetically very diverse with eight and 13 different alleles detected for the VVS1 and VVS2 loci. Inclusion of other grapevine species into the analysis increases the number of alleles to 11 and 15 respectively. This genetic variety not only demonstrates the hypervariability of microsatellite repeat sequences but most likely reflects the diverse historical origins of different *V. vinifera* cultivars. Grapevine cultivation had its origins with the domestication of *V. vinifera* around 4000 BC (Olmo 1976). The spread of civilisation in Eurasia and associated grape cultivation and wine production is thought to have contributed to the diverstiy of *V. vinifera* genotypes due to the introduction of cultivated *V. vinifera* forms into different geographical regions, the direct cultivation of regional wild *V. vinifera* forms, and hybridisations between them (Negrul 1938).

Cultivars of *V. vinifera* make up the majority of cultivars grown worldwide for fruit and wine production. Those grown for wine production are generally very old having been established before the 20th century. Active breeding programs are mainly restricted to cultivars grown for the production of table grapes and dried fruit. Due to the reliance on morphological characteristics for cultivar descriptions, the vegetative propagation of cultivars, and the long history of viticulture, accurate cultivar identification is often difficult. It has been shown previously that the use of RFLP molecular markers for grapevine identification offers an objective and viable alternative to phenotypic identification by ampelographic methods (Striem et al. 1990; Thomas et al. 1993) and recently there has been interest in producing a genetic map of grapevine using RFLPs (Mauro et al. 1992). This present study demonstrates that STMSs represent a PCR DNA marker type with an extremely large number of potential markers available in the grapevine genome, the individual markers being highly polymorphic or informative. In addition, the PCR technology is simpler and faster than Southern-blot technology. Due to the universal application of STMS primers within the genus Vitis this marker type would appear to be extremely useful for the genetic mapping of the grapevine genome and for accurate cultivar identification.

Acknowledgments. We thank Richard Cirami at the South Australian Department of Agriculture Nuriootpa Research Advisory Centre for advice and plant material. Trish Cain and Rita

Luters provided excellent technical assistance. This research was supported in part by grants from the Grape and Wine Research and Development Corporation. The Dried Fruits Research and Development Corporation, and the Commonwealth Department of Industry Technology and Regional Development.

References

Alleweldt G, Spiegel-Roy P, Reisch B (1990) Grapes (*Vitis*). In: Moore JN, Ballington Jr JR (eds) Genetic resources of temperate fruit and nut crops 1. Int Soc Hort Sci, Wageningen, pp 289–327

Beckmann JS, Soller M (1990) Toward a unified approach to the genetic mapping of eukaryotes based on sequence-tagged microsatellite sites Bio/Technology 8:930–932

Beyermann B, Nurnberg P, Weihe A, Meixner M, Epplen JT, Borner T (1992) Fingerprinting plant genomes with oligonucleotide probes specific for simple repetitive DNA sequences. Theor Appl Genet 83:691–694

Clingeleffer PR, Possingham JV (1988) Marroo Seedless: a new table grape variety. Agri Sci (Aust) 1:18–19

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253

Hamada H, Petrino MG, Kakunaga T (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily-diverse eukaryotic genomes. Proc Natl Acad Sci USA 79:6465–6469

Kwiatkowski DJ, Henske EP, Weimer K, Ozelius L, Gusella JF, Haines J (1992) Construction of a GT polymorphism map of human 9q. Genomics 12:229–240

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in-vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Love JM, Knight AM, McAleer MA, Todd JA (1990) Toward construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. Nucleic Acids Res 18:4123–4130

Mauro M-C, Strefeler M, Weeden NF, Reisch BI (1992) Genetic analysis of restriction fragment length polymorphisms in *Vitis*. J Hered 83:18–21

Moore SS, Sargeant LL, King TG, Mattick JS, Georges M, Hertzel DJS (1991) Conservation of dinucleotide microsatellites among mammalian genomes allows use of heterologous PCR primer pairs in closely related species. Genomics 10:645–660

Negrul AM (1938) Evolution of cultivated forms of grapes. CR Acad Sci USSR 18:585–588

Olmo HP (1976) Grapes. In: Simmonds NW (ed) Evolution of crop plants. Longman, London, pp 294–298

Olson M, Hood L, Cantor D, Botstein D (1989) A common language for physical mapping of the human genome. Science 245:1434–1435

Rychlik W, Rhoads RE (1989) A computor program for choosing optimal oligonucleotides for filter hybridization, sequencing and in-vitro amplification of DNA. Nucleic Acids Res 17:8543–8551

Schlotterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. Nature 354:63–65

Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:25–33

Stallings RL, Ford AF, Nelson D, Torney DC, Hildebrand CE, Moyzis RK (1991) Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. Genomics 10: 807–815

Striem MJ, Spiegel-Roy P, Ben-Hayyim G, Beckmann J, Gidoni D (1990) Genomic DNA fingerprinting of *Vitis vinifera* by the use of multi-loci probes. Vitis 29:223–227

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic markers. Nucleic Acids Res 17: 6463–6471

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12:4127–4138

Thomas MR, Matsumoto S, Cain P, Scott NS (1993) Repetitive DNA of grapevine: classes present and sequences suitable for cultivar identification. Theor Appl Genet 86:173–180

Weber JL (1990) Informativeness of human $(dC\text{-}dA)_n\cdot(dG\text{-}dT)_n$ polymorphisms. Genomics 7:524–530

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18: 7213–7218

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18: 6531–6535