

Sequence analysis of a microsatellite and its flanking regions in intraspecific hybrids of grapevine (*Vitis vinifera* L.)

E. MASI¹, R. VIGNANI², E. SENSI², M. SCALI², M. CRESTI² and M. BOSELLI¹

¹) Dipartimento di Ortoflorofruitticoltura, Università degli Studi di Firenze, Italia

²) Dipartimento di Scienze Ambientali “G. Sarfatti”, Università degli Studi di Siena, Italia

Summary

Microsatellite (MS) VVMD21 (BOWERS *et al.* 1999) was taken as a model to explore the molecular basis of polymorphism in a panel of 6 grapevine accessions (*Vitis vinifera* L.), consisting of Sangiovese and Cabernet Sauvignon and 4 F₁ plants derived from crossing both varieties. The 12 alleles of both parents and the progeny were cloned and sequenced. The microsatellite repeat (AG)_{n>6} was found in each sequence, together with a poly-T rich region that showed irregularity. Furthermore, single nucleotide deletion or exchange (point mutations) were found in the microsatellite flanking regions.

Key words: MS polymorphism, DNA, sequence, *Vitis vinifera*.

Introduction

Microsatellite (MS) DNA sequences consist of relatively short repeats of 1-5 base pair units found both in prokaryotes and eukaryotes, whose length polymorphism have been largely used, *e.g.* for genotype identification, parentage tests and DNA mapping. Although in prokaryotes distinct biological functions have been demonstrated, *e.g.* contingency genes (MEYER 1987), the role of MS in eukaryotes is less clear, even if some cases of trinucleotide repeats seem to be associated with diseases, such as Huntington's disease (BROOK *et al.* 1992). Furthermore, the instability of microsatellite (MSI), especially distinct in single nucleotidic repeats, seems to have a dynamic role in the origin and progression of several tumoral forms where it could have a negative effect in the right mismatch (MMR) repair (FRANK 2003; ANDREW and PETERS 2001), the DNA repair system that detects and replaces wrongly paired, mismatched, bases in newly replicated DNA. Besides, repetitive DNA variations seem to be involved in evolution (JOBLING and TYLER-SMITH 2003)

Investigated in detail is also the issue of microsatellite origin; two phenomena are probably involved: (1) DNA slippage and (2) unequal crossing-over (LEVINSON and GUTMAN 1987; SCHLÖTTERER and TAUTZ 1992). When these processes occur in areas of tandem repeats, causing modification in the number of bases, microsatellite polymorphisms originate. According to these theories, microsatellite polymorphism should be linked only with the length of the repeat

region, and the variation in length should be multiple of the repetitive motif ($\pm 3n$ if we have a trinucleotide repeat).

There are, however, many possibilities to observe deviation from this rule; in many cases the errors are probably inherent in the methodology (binning errors, PCR generation of false positive/negative), but what are the probabilities to observe the “correct” phenomenon of microsatellite generation? Besides, very few studies have been performed to confirm the nature of the so-called conserved flanking regions, and recently, mutations in the MMR system have been supposed to play a significant role in expansion and contraction of microsatellite sequences and in MSI (VAISH and MITTAL 2002).

While microsatellite polymorphism is well established in the animal and human genome, and the information on microsatellites in plants, especially in crop plants, has increased over the last few years, only few reports describe microsatellite nature in woody plants (EHT *et al.* 1996, 1999; DI GASPERO *et al.* 2000; ELSIK *et al.* 2000; DEVEY *et al.* 2002). Elucidation of the processes that lead to MS formation might shed light on their possible significance and the understanding of the evolution of the genome eukaryotes.

The aim of the present work was to study microsatellites in cross-originated individuals and their parents; in particular, the sequencing of the PCR products can be a valid tool for the investigation of the nature of the polymorphism detected on the MS locus of *Vitis vinifera*.

Material and Methods

Plant material and DNA extraction: Two varieties of *Vitis vinifera* L. (Sangiovese and Cabernet Sauvignon) were used to generate intraspecific hybrids. Four F₁ plants were selected and marked with ID numbers 1 to 4. Genomic DNA was extracted from young leaves using a modified method reported by MULCAHY *et al.* (1993). Purified DNA was stored at 4 °C.

Amplification: The 6 DNAs were PCR amplified for the locus VVMD21 generated by BOWERS *et al.* (1999). PCR amplification and amplicons detection were performed according to Masi *et al.* (2001), using Cy5 primers (CyTM5 Amidite, Pharmacia Biotech), except for the use of a proof-reading *Taq*-DNA polymerase (Invitrogen). Further amplification was performed with unlabelled primers since labelling had revealed negative interference with the cloning procedure.

Microsatellite cloning and sequencing: Cloning reactions were carried out with the TOPO TA Cloning® Kit for Sequencing version F (Invitrogen). The plasmids were extracted using the QIAGEN® Plasmid Purification Kit (Qiagen), sequenced using the ALFexpress™ AutoRead™ Sequencing Kit (Amersham Pharmacia Biotech Inc., Piscataway, NJ, USA) and analyzed on a polyacrylamide PAA gel run on a semi-automated DNA sequencer, the ALFexpressII. Post run data analysis was performed with the ALFwin Sequence Analyzer 2.00 software while the generation of the consensus sequence and the pairwise sequence comparison was carried out with DNAsys 2.10 software. The analysis was performed in order to observe at least three times each of the 12 VVMD21 amplicons generated by the accessions.

Results and Discussion

Tab. 1 presents VVMD21 allele sizes for the analyzed accessions, as they were detected on a PAA gel electrophoresis conducted on a semi-automated DNA sequencer. Three different alleles were differentiated, 244 bp, 250 bp and 258 bp (later indicated with a, b, c respectively). Mendelian segregation was confirmed, supporting the accuracy of the analysis. Conversely, when sequenced, suspicious differences were found in the dimension of the alleles detected (Tab. 2). This is probably due to some limitation of the sizing method, *e.g.* the anomalous migration of the fragments: on a polyacrylamide gel, DNA fragments having AT-rich regions migrate slower than other DNA fragments of similar size (STELLWAGEN 1983).

The sequencing assay highlighted the presence of the repeat motif, which is an (AG)_n repeat (Figure), one of the most common in woody plant genomes (MORGANTE and

Table 1

Allele size (bp) of parents (Sangiovese and Cabernet Sauvignon) and progeny (1-4) utilized for the study; size was detected with a semi-automated DNA sequencer

Sangiovese	Cabernet Sauvignon	No. 1	No. 2	No. 3	No. 4
244				244	244
250	250	250-250	250		250
	258		258	258	

OLIVIERI 1993; DOW *et al.* 1995) and in grapevine as well (THOMAS and SCOTT 1993; DI GASPERO *et al.* 2000).

The microsatellite showed further modifications in the adjacent poly-T rich region, which occurred as an imperfect motif (T)_nC(T)_nC(T)_n.

Besides the expected variations in repeat length and in the poly-T region, the comparison of the sequences showed a deletion (position 9) and two single-nucleotide substitutions, G→A (positions 105 and 153). In particular, the deletion in position 9 was present only in the shorter allele (*a*)

Table 2

Comparison of allele size between values obtained by a semi-automated DNA sequencer (electropherogram size) and by sequencing

Allele	Electropherogram size	Sequencing size	Δbp
<i>a</i>	244	237	7
<i>b</i>	250	245	5
<i>c</i>	258	256	2

and was confirmed for all three accessions in which was scored (Sangiovese, No. 3 and No. 4). The deletion disagrees with the supposed mechanism of polymorphism generation only concerning the repeat motif. Similar consideration can be made for the single base exchange in position 105: Sangiovese showed a G in both alleles (*a* and *b*), Cabernet Sauvignon showed an A in its alleles (*b* and *c*), and the progeny exhibited the two nucleotides according with a mendelian segregation (accession No. 1 inherited one of the allele *b* from Sangiovese and the other from Cabernet Sauvignon; accession No. 2 inherited allele *b* from Sangiovese and allele *c* from Cabernet Sauvignon; accession No. 3 inherited allele *a* from Sangiovese and allele *c* from Cabernet Sauvignon; accession No. 4 inherited allele *a* from Sangiovese and allele *b* from Cabernet Sauvignon). The base exchange in position 153, instead, does not confirm Mendel's laws: only allele *b* of Cabernet Sauvignon showed a G instead of an A.

Conclusion

The study represents a model for a deep exploration on the nature of MS polymorphism. Even though it should be applied to a wider level, some interesting conclusions can be drawn. Some of the alterations observed in the MS sequences do not belong to the repeats region *sensu stricto*; this fact confirms that the length variability does not follow a single and simple mutation mechanism. Besides, we have observed some alteration in flanking regions that, indirectly, lead back to the topic of microsatellite instability. These alterations might not be always conserved through evolution or, as in this case, through segregation from parents to progeny. Sequencing MS alleles, it has been highlighted the presence of a poly-T rich region which is considered a sequence strictly related with instability (BACON 2001). This suggests that the MS containing poly-T region may not be first choice as molecular marker, and in future works of MS isolation they might be discarded during selection.

	10	20	30	40	50	60
S allele <i>a</i>	GGTTGTCT-T	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
No. 3 allele <i>a</i>	GGTTGTCT-T	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
No. 4 allele <i>a</i>	GGTTGTCT-T	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
S allele <i>b</i>	GGTTGTCTAT	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
C allele <i>b</i>	GGTTGTCTAT	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
No. 1 allele <i>b</i>	GGTTGTCTAT	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
No. 1 allele <i>b</i>	GGTTGTCTAT	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
No. 2 allele <i>b</i>	GGTTGTCTAT	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
No. 4 allele <i>b</i>	GGTTGTCTAT	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
C allele <i>c</i>	GGTTGTCTAT	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
No. 2 allele <i>c</i>	GGTTGTCTAT	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
No. 3 allele <i>c</i>	GGTTGTCTAT	GGAGTTGATG	TTGCTTTCAT	AGTTTTC TAG	GAAACAGCCTT	TGTATCAACA
	70	80	90	100	110	120
S allele <i>a</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
No. 3 allele <i>a</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
No. 4 allele <i>a</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
S allele <i>b</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
C allele <i>b</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
No. 1 allele <i>b</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
No. 1 allele <i>b</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
No. 2 allele <i>b</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
No. 4 allele <i>b</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
C allele <i>c</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
No. 2 allele <i>c</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
No. 3 allele <i>c</i>	TGGTGACTGA	AAAGTTTCAG	AACA AAAAGCA	TTATAT TAGA	GATGTGATC	AGGGCTCTGT
	130	140	150	160	170	180
S allele <i>a</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT-	-----TTTTT
No. 3 allele <i>a</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT-	-----TTTTT
No. 4 allele <i>a</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT-	-----TTTTT
S allele <i>b</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT	-----TTTTT
C allele <i>b</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT	-----TTTTT
No. 1 allele <i>b</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT	-----TTTTT
No. 1 allele <i>b</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT	-----TTTTT
No. 2 allele <i>b</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT	-----TTTTT
No. 4 allele <i>b</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT	-----TTTTT
C allele <i>c</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT	TTCTTTT
No. 2 allele <i>c</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT	TTCTTTT
No. 3 allele <i>c</i>	CTCCAGGTGG	TGGA AACCTA	GGTAGTTAAC	CCAACCCCA	TTTCTTTT	TTCTTTT

Figure: DNAsys multiple sequence alignment: VVMD21 microsatellite and poly-T motif, single-base exchange and deletion are shaded. Gaps (-) indicate length difference. **S** denotes Sangiovese, **C** Cabernet Sauvignon, *a*, *b* and *c* are the short, medium and long alleles, respectively.

Figure, continued

	190	200	210	220	230	240
S allele <i>a</i>	CCTTCVCAGA	GAGAG-----	-----AGA	GATTTACCTG	CATTGCAAGA	AATTGCAAT
No. 3 allele <i>a</i>	CCTTCVCAGA	GAGAG-----	-----AGA	GATTTACCTG	CATTGCAAGA	AATTGCAAT
No. 4 allele <i>a</i>	CCTTCVCAGA	GAGAG-----	-----AGA	GATTTACCTG	CATTGCAAGA	AATTGCAAT
S allele <i>b</i>	CCTTCVCAGA	GAGAGAGAG-	-----AGAGA	GATTTACCTG	CATTGCAAGA	AATTGCAAT
C allele <i>b</i>	CCTTCVCAGA	GAGAGAGAG-	-----AGAGA	GATTTACCTG	CATTGCAAGA	AATTGCAAT
No. 1 allele <i>b</i>	CCTTCVCAGA	GAGAGAGAG-	-----AGAGA	GATTTACCTG	CATTGCAAGA	AATTGCAAT
No. 1 allele <i>b</i>	CCTTCVCAGA	GAGAGAGAG-	-----AGAGA	GATTTACCTG	CATTGCAAGA	AATTGCAAT
No. 2 allele <i>b</i>	CCTTCVCAGA	GAGAGAGAG-	-----AGAGA	GATTTACCTG	CATTGCAAGA	AATTGCAAT
No. 4 allele <i>b</i>	CCTTCVCAGA	GAGAGAGAG-	-----AGAGA	GATTTACCTG	CATTGCAAGA	AATTGCAAT
C allele <i>c</i>	CCTTCVCAGA	GAGAGAGAGAG	GAGAGAGAGAG	GATTTACCTG	CATTGCAAGA	AATTGCAAT
No. 2 allele <i>c</i>	CCTTCVCAGA	GAGAGAGAGAG	GAGAGAGAGAG	GATTTACCTG	CATTGCAAGA	AATTGCAAT
No. 3 allele <i>c</i>	CCTTCVCAGA	GAGAGAGAGAG	GAGAGAGAGAG	GATTTACCTG	CATTGCAAGA	AATTGCAAT
	250	260				
S allele <i>a</i>	CCCTTTTAC	TGAAGC-----				
No. 3 allele <i>a</i>	CCCTTTTAC	TGAAGC-----				
No. 4 allele <i>a</i>	CCCTTTTAC	TGAAGC-----				
S allele <i>b</i>	CCCTTTTAC	TGAAGC-----				
C allele <i>b</i>	CCCTTTTAC	TGAAGC-----				
No. 1 allele <i>b</i>	CCCTTTTAC	TGAAGC-----				
No. 1 allele <i>b</i>	CCCTTTTAC	TGAAGC-----				
No. 2 allele <i>b</i>	CCCTTTTAC	TGAAGC-----				
No. 4 allele <i>b</i>	CCCTTTTAC	TGAAGC-----				
C allele <i>c</i>	CCCTTTTAC	TGAAGC-----				
No. 2 allele <i>c</i>	CCCTTTTAC	TGAAGC-----				
No. 3 allele <i>c</i>	CCCTTTTAC	TGAAGC-----				

References

- ANDREW, S. E.; PETERS, A. C.; 2001: DNA instability and human disease. *Am. J. Pharmacogen.* **1**, 21-28.
- BACON, L. A.; DUNLOP, N. G.; FARRINGTON, S. M.; 2001: Hypermotability at a poly(A/T) tract in the human germline. *Nucl. Acids Res.* **29**, 4405-4413.
- BOWERS, J. E.; DANGL, G. S.; MEREDITH, C. P.; 1999: Development and characterization of additional microsatellite DNA markers for grape. *Am. J. Enol. Vitic.* **50**, 243-246.
- BROOK, J.; MCCURRACH, M. E.; HARLEY, H. G.; 1992: Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**, 799-808.
- DEVEY, M. E.; BELL, J. C.; UREN, T. L.; MORAN, G. F.; 2002: A set of microsatellite markers for fingerprinting and breeding applications in *Pinus radiata*. *Genome* **45**, 984-989.
- DI GASPERO, G.; PETERLUNGER, E.; TESTOLIN, R.; EDWARDS, K. J.; CIPRIANI, G.; 2000: Conservation of microsatellite loci within the genus *Vitis*. *Theor. Appl. Gen.* **101**, 301-308.
- DOW, B. D.; ASHLEY, M. V.; HOWE, H. F.; 1995: Isolation and characterization of highly variable (GA/CT)_n microsatellites in the bur oak, *Quercus macrocarpa*. *Theor. Appl. Gen.* **91**, 137-141.
- ECHT, C. S.; MAY-MARQUARDT, P.; HSEIH, M.; ZAHORCHAK, R.; 1996: Characterization of microsatellite markers in eastern white pine. *Genome* **39**, 1102-1108.
- ECHT, C. S.; VENDRAMIN, G. G.; NELSON, C. D.; MARQUARDT, P.; 1999: Microsatellite DNA as shared genetic markers among conifer species. *Can. J. for. Res.* **29**, 365-371.
- ELSIK, C. G.; MINIHAM, V. T.; HALL, S. E.; SCARPA, A. M.; WILLIAMS, C. G.; 2000: Low-copy microsatellite markers for *Pinus taeda*. *Genome* **43**, 550-555.
- FRANK, T. S.; 2003: Topics in Molecular Oncology: Microsatellite Instability. University of Michigan School of Medicine, Department of Pathology, MI, USA. Available from: <http://www.pds.med.umich.edu/users/frank/MIN.html>.
- JOBLING, M. A.; TYLER-SMITH, C.; 2003: The human Y chromosome: An evolutionary marker comes of age. *Nat. Rev. Genet.* **4**, 598-612.
- LEVINSON, G.; GUTMAN, G. A.; 1987: Slipped-strand misrepairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203-221.
- MASI, E.; VIGNANI, R.; DI GIOVANNANTONIO, A.; MANCUSO, S.; BOSELLI, M.; 2001: Ampelographic and cultural characterisation of the Casavecchia variety. *Adv. Hort. Sci.* **15**, 47-55.
- MEYER, T. F.; 1987: Molecular basis of surface-antigen variation in *Neisseria*. *Trends Genet.* **3**, 319-324.
- MORGANTE, M.; OLIVIERI, A. M.; 1993: PCR-amplified microsatellites as markers in plant genetics. *Plant J.* **3**, 175-182.
- MULCAHY, D. L.; CRESTI, M.; SANSAVINI, S.; DOUGLAS, G. C.; LINSKENS, H. F.; BERGAMINI-MULCAHY, G.; VIGNANI, R.; PANCALDI, M.; 1993: The use of random amplified polymorphic DNAs to fingerprint apple genotypes. *Sci. Hortic.* **54**, 89-96.
- SCHLÖTTERER, C.; TAUTZ, D.; 1992: Slippage synthesis of simple sequence DNA. *Nucl. Acid Res.* **20**, 211-215.
- STELLWAGEN, N. C.; 1983: Anomalous electrophoresis of deoxyribonucleic acid restriction fragments on polyacrylamide gels. *Biochemistry* **22**, 6186-6193.
- THOMAS, M. R.; SCOTT, N. S.; 1993: Microsatellite repeats in grapevine reveal DNA polymorphism when analyzed as sequence-tagged sites (STSs). *Theor. Appl. Genet.* **86**, 985-990.
- VAISH, M.; MITTAL, B.; 2002: DNA mismatch repair, microsatellite instability and cancer. *Indian J. Exp. Biol.* **40**, 989-994.

Received November 4, 2003