Theor Appl Genet (2000) 100:723–726 

**K.D. Scott · P. Eggler · G. Seaton · M. Rossetto**
**E.M. Ablett · L.S. Lee · R.J. Henry**

# Analysis of SSRs derived from grape ESTs

Received: 4 June 1999 / Accepted: 21 September 1999

**Abstract** One hundred and twenty four microsatellites were isolated from analysis of 5000 *Vitis* expressed sequence tags (ESTs). A diversity of dinucleotide and trinucleotide simple sequence repeat (SSR) motifs were present. Primers were designed for 16 of these SSRs and they were tested on seven accessions. Ten of the sixteen primer pairs resulted in PCR products of the expected size. All ten functional primers were polymorphic across the accessions studied. Polymorphisms were evident at the level of cultivars, *Vitis* species, and between related genera. SSRs that were from the 3′ untranslated region (3′UTR) were most polymorphic at the cultivar level, the 5′ untranslated region (5′ UTR) SSRs were most polymorphic between cultivars and species, and those SSRs within coding sequence were most polymorphic between species and genera. These results show that EST-derived SSRs in *Vitis* are useful as they are polymorphic and highly transferable. With EST SSRs being applicable to studies at several taxonomic levels, the large number of SSRs (approximately 1000) that will be available from an expanded EST database of 45 000 will have many potential applications in mapping and identity research.

**Key words** ESTs · Microsatellites · *Vitis vinifera*

## Introduction

Microsatellites [simple sequence repeats (SSRs) or hypervariable sequences] are arrays of short motifs of 1–4 base pairs in length. These single-locus markers are characterised by their hypervariability, abundance, reproducibility, Mendelian inheritance and codominant nature. Microsatellite markers require the design of primers for the conserved flanking regions of the microsatellite, and the PCR-amplification of the repeat region.

Applications of microsatellites to table and wine grapes include the construction of molecular maps, accurate cultivar identification and cultivar parentage studies (Sefc et al. 1999). Some grape cultivars are difficult to differentiate using ampelographic methods based on their phenotype (partly due to the long history of cultivation and the wide distribution of the species) (Thomas et al. 1993). In these instances microsatellites offer an objective means of identifying such cultivars. Microsatellite markers have also been utilised to verify or clarify the genotype and parentage of particular grape lines (Bowers and Meredith 1997; Lamboy and Alpha 1998). Microsatellites may also have applications in the characterisation of new and unique cultivars for the purpose of patents and plant variety rights.

In the past, SSRs have been expensive to develop and thus often limited to applications in the larger commercial crops. Enrichment protocols are reducing this cost; however, recently, a new alternative source of microsatellites has been utilised. This is the identification of microsatellites from expressed sequence tag (EST) databases. These microsatellites are obviously limited to species for which this type of database exists. To-date this is available for some important species, including *Arabidopsis* (Delseny et al. 1997) rice (Sasaki et al. 1994; Yamamoto and Sasaki 1997), and maize (Wang and Bowen 1998). There are also a small number of similar studies where SSRs have been derived from public databases for use in plants, for example in potato (Milbourne et al. 1998).

This paper describes ten SSRs derived from the grape EST database and tests them for levels of polymorphism and transferability. As the EST-derived microsatellites are within transcribed regions of the DNA, it would be expected that these markers are more transferable than anonymous SSRs. The corresponding concern is that the

Communicated by P. Langridge

K.D. Scott (✉) · P. Eggler · E.M. Ablett · R.J. Henry
Centre for Plant Conservation Genetics, PO Box 157,
Lismore NSW 2480, Southern Cross University, Australia
e-mail: kscott@scu.edu.au
Tel.: +61 2 66203141, Fax: +61 2 66222080

G.Seaton, M.Rossetto, L.S. Lee,
Australian Agriculture Research Institute, PO Box 157,
Lismore NSW 2480, Southern Cross University, Australia

**Table 1** SSR primers developed from a *V. vinifera* Chardonnay EST database. For each SSR the repeat unit, length and location are shown. The annealing temperature ($T_a$) as calculated using MacVector 6.0 (Oxford Molecular Group 1996) and the expected length in base pairs (L) of PCR products are given

| Primer | Repeat | Position | Forward primer (5′→3′) | Reverse Primer (5′→3′) | $T_a$ | L |
|---|---|---|---|---|---|---|
| scu01vv | $(CT)_9X_{60}(CT)_{10}AT(CT)_5$ | 5'UTR | TTTGGAATCCGTCACGAG | GGTTGACCTTTTTCCGAC | 52.3 | 164 |
| scu02vv | $(GA)_9$ | 5′UTR | TTGCAAGAATDCGGCACGAGG | GAATTGTGCGTGGTAGGCTCA | 55.0 | 144 |
| scu03vv | $(GA)_8$ | 5′UTR | TTCGGCACGAGGTTTTAG | ATTAGGCAGAGAAGAGCGG | 54.1 | 165 |
| scu04vv | $(CT)_8$ | 5′UTR | TGTCCTCTTTCCCTCTCCCAAC | CAGTCTGTCATCTGACCATGTAGCC | 55.2 | 175 |
| scu05vv | $(AT)_{13}$ | 3′UTR | CAAGCAGTTATTGAAGCTGCAAGG | TCATCCATCACACAGGAAACAGTG | 51.2 | 174 |
| scu06vv | $(AT)_8$ | 3′UTR | CCTAATGCCAGGAAGGTTGC | CCCTAGTCTCTCTACCTATCCATG | 49.7 | 171 |
| scu07vv | $(ACC)_5$ | CDS | CCGAAGAGGAATATGGGTTTGAG | CCTAACTTGAAACGAAAGGACTGC | 54.9 | 203 |
| scu08vv | $(GGT)_5$ | 5′UTR | CGAGACCCAGCATCGTTTCAAG | GCAAAATCCTCCCCGTACAAGTC | 57.7 | 180 |
| scu09vv | $(GGT)_5$ | 3′UTR | AAGCAGCAGTTATTGGCG | CAGATACTGAGGGTTTAAGCTC | 50.7 | 122 |
| scu10vv | $(CAA)_6$ | 5′UTR | TACCCCCACAACCCTTTTTCCC | TTCTCCGCCACCTCCTTTTCAC | 55.9 | 205 |
| scu11vv | $(CTT)_8$ | 5′UTR | AATTGATAGTGCCACGTTCTCGCC | AACGCCGACAAGAATCCCAAGG | 57.3 | 248 |
| scu12vv | $(TCT)_{10}$ | 5′UTR | GAATTCGGCACGAGGAACTA | ACAGTGGAGAGGTGAATGCA | 50.0 | 130 |
| scu13vv | $(CTT)_6$ | CDS | AATTCGGCACGAGTCAACACC | GCTTTGGGAAATGGAGTACAGAAG | 53.9 | 188 |
| scu14vv | $(GAA)_6$ | 3′UTR | CTGCACTTGAATACGAGCAGGTC | TGTTATATGATCCTCCCCCTCCTC | 53.6 | 182 |
| scu15vv | $(GAA)_6$ | CDS | GCCTATGTGCCAGACCAAAAAC | TTGGAAGTAGCCAGCCCAACCTTC | 53.5 | 195 |
| scu16vv | $(GAA)_5$ | CDS | CAAAGACAAAGAAGCCACCGAC | ACCCTCTAAAGCACACACAGGAAC | 54.4 | 170 |

coding character of these SSRs will limit their polymorphism. This pretext was addressed in this study by including two *Vitis vinifera* cultivars, three non-*vinifera* commercial species and two Australian native relatives from related genera, in the testing. The EST-derived SSRs in this study are also compared to genomic enrichment-derived SSRs developed using enrichment techniques from the *Vitis* Microsatellite Consortium. With the large number of SSRs available from existing EST databases at little cost, our EST database provides a unique source of many SSRs for *Vitis* that can be applied to identity and mapping applications.

## Materials and methods

Microsatellites were isolated from a Chardonnay leaf and berry EST database of 5000 sequences (Ablett et al. submitted) by converting the *Vitis* EST database from FastA format to BLAST-2 format using the program Formatdb. Potential microsatellite sites were identified using BLASTN for all possible mononucleotide, dinucleotide and trinucleotide repeat patterns against the BLAST 2 formatted database. Only mononucleotides of ten or more repeats, dinucleotides of seven or more repeats, or trinucleotides of five or more repeats were extracted. This method found 46 dinucleotide repeats and 78 trinucleotide repeats.

Sixteen primer pairs were designed and tested. The 16 primers were chosen to represent dinucleotide and trinucleotide repeats located in 5′ untranslated (5′UTR), 3′ untranslated (3′UTR) and coding sequence (CDS) regions (Table 1). The assignment of an SSR location to a gene region, was based on high homology (>60% identity) of the EST to genes in other species. Of the 16 primers eight were from the 5′UTR, four from the 3′UTR and four from CDS regions. The four SSRs from the CDS region were all trinucleotide repeats.

The samples used to screen these primers were: two *V. vinifera* cultivars (Cabernet Sauvignon and Riesling), three non-*vinifera* cultivars (Ramsey, Riparia and 1616C) and two non-*Vitis* genera from within the Vitaceae (*Cissus cardiophylla* and *Cayratia japonica*). DNA was extracted from these species using the protocol of Scott and Playford (1996). PCR conditions for screening the SSRs were as follows: 25 ng of DNA, 2.5 mM of $MgCl_2$, 0.2 µM of each primer, 10 mM of Tris-HCl, 50 mM of KCl, 1 unit of Amplitaq Gold (Perkin Elmer) and 0.2 mM of dNTPs. Reaction volumes were 25 µl, and they were cycled at 94°C for 1 min, 50°C for 1 min, and 73°C for 1 min repeated for 45 cycles on a Perkin Elmer 9700. Resulting products were resolved and scored manually on ethidium bromide-stained 8% denaturing polyacrylamide gels.

## Results

Non-redundant microsatellites, which were dinucleotides of seven or more repeats and trinucleotides of five or more repeats, were 2.5% of the total population of cDNA clones in the library. From our pilot EST database a total of 46 non-redundant dinucleotide and 78 non-redundant trinucleotide repeats were isolated. The proportions of different dinucleotide and trinucleotide repeat motifs are listed in Table 2. Many dinucleotide and trinucleotide repeat motifs are represented; however, no mononucleotide repeats of a useful length were found.

Of the 16 primer pairs tested, ten pairs produced products of the expected size under the given conditions (e.g. Fig. 1). The number and size of alleles from the 16 primer pairs are shown in Table 3. Primers scu01vv, scu02vv, scu03vv, scu09vv, scu12vv and scu13vv either failed to amplify or did not produce products of the expected size. All ten functional primers revealed polymorphisms. Variation was evident between cultivars for five of the ten primers, between *Vitis* species for eight of ten, and across genera in four of the ten primers. Primer pairs scu04vv, scu08vv, scu10vv and scu11vv amplified an SSR from the 5′UTR region and, of these, one was polymorphic between cultivars, three were polymorphic between species and one was variable across genera. The 3′UTR primers (scu05vv, scu06vv and scu14vv) were all variable across cultivars and species, and one of the three was polymorphic across genera. Of the CDS SSRs (scu07vv, scu15vv and scu16vv) one was variable between cultivars and two were variable across species and genera.

**Table 2** Comparison of the characteristics of EST-derived SSRs to those obtained through genomic enrichment (VMC repeat data based on the clones from which primers were designed)

| Item | EST SSRs | Enriched Genomic SSRs |
|---|---|---|
| Frequency in library (%) | 2.6 | 47 |
| Repeat types within each length class (%) | 4.3 CA/GT | 0.9 CA/GT |
| | 19.6 AT/TA | 0 AT/TA |
| | 67.4 CT/GA | 97.3 CT/GA |
| | 8.7 compound dinucleo. | 1.8 compound dinucleo. |
| | 2.5 TTA/AAT | 0 TTA/AAT |
| | 2.6 CAA/GTT | 0 CAA/GTT |
| | 3.8 CGT/GCA | 0 CGT/GCA |
| | 5.1 CAT/GTA | 0 CAT/GTA |
| | 9 CAG/GTC | 40 CAG/GTC |
| | 10.3 CAC/GTG | 6.7 CAC/GTG |
| | 11.5 CCT/GGA | 0 CCT/GGA |
| | 12.8 CGG/GCC | 26.7 CGG/GCC |
| | 41 CTT/GAA | 13.3 CTT/GAA |
| | 1.4 compound trinucleo. | 13.3 compound trinucleo. |
| Average repeat length | 7.9 | 20.2 |

**Table 3** Allele lengths (base pairs) of the EST SSRs. Polymorphisms (Poly) between cultivars (C), species (Spp) and genera (G) are shown, in addition to the transferability (Trans) of the markers. No amplification products shown by (–)

| Primer pair | Riesling *V. vinifera* | Cab.Sauv. *V. vinifera* | Ramsey *V. champini* | 1616c *V. longii*× *V. riparia* | Riparia *V. riparia* | *Cissus* | *Cayratia* | Poly | Trans |
|---|---|---|---|---|---|---|---|---|---|
| scu04vv | 171 | 175 | 171 | 171 | 171 | – | – | C | Spp |
| scu05vv | 164 | 168 | 160 | 160 | 160 | – | – | C, Spp | Spp |
| scu06vv | 155 | 175 | 175 | 191 | 171 | – | 155 | C, Spp, G | Spp, G |
| scu07vv | 203 | 203 | 206 | 206 | 209/235 | – | – | Spp | Spp |
| scu08vv | 180 | 180 | 180 | 159/180 | 159/180 | – | – | Spp | Spp |
| scu10vv | 205/274 | 205 | 205/274 | 205/244 | 205/244 | – | 307 | C, Spp, G | Spp, G |
| scu11vv | 248 | 248 | 242 | 242 | 242 | – | – | Spp | Spp |
| scu14vv | 167/185 | 167 | 188 | 188 | 188 | – | – | C, Spp | Spp |
| scu15vv | 207/252 | 201/252 | 201/252 | 201 | 201 | – | 201/222 | C, Spp, G | Spp, G |
| scu16vv | 170 | 170 | 170 | 170 | 170 | 170 | 179 | G | Spp, G |

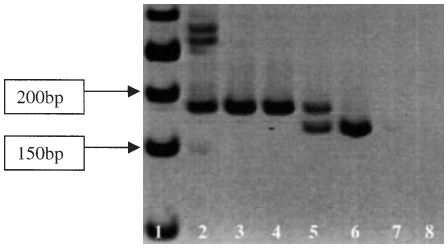


**Fig. 1** Primer scu14vv products separated on an 8% polyacrylamide gel. Lane 1–50 base pair ladder, 2 - Ramsey, 3 – 1616 C, 4 – Riparia, 5 - Riesling, 6 - Cabernet Sauvignon, 7 – *Cissus cardiophylla*, 8 – *Cayratia japonica*

## Discussion

SSRs in the ESTs studied were relatively abundant (2.5% of total clones), and all functional primer pairs revealed polymorphisms and were highly transferable across cultivars, species and genera. Half of the SSRs examined were useful for differentiating the two *V. vinifera* cultivars, while eight out of the ten SSRs were able to discriminate rootstock species. If more cultivars were screened, the number of primer pairs that were polymorphic at this level would most likely also increase. The level of transferability of the EST SSRs was good, with four of the ten primer pairs amplifying alleles from a related genus.

The level of polymorphism detected by SSRs from different EST regions (CDS, 5′UTR and 3′UTR) varied across the different taxonomic levels. In the 5′UTR, SSR polymorphism was present at all taxonomic levels; however, the majority of 5′UTR SSRs varied between species. The 5′UTR SSRs transferred to other species and genera from the Vitaceae. All 3′UTR SSRs were variable across cultivars and species, with some variation also evident across genera. The 3′UTR SSRs were as transferable as the 5′UTR SSRs. The SSRs within the CDS were all trinucleotides, and variation was present at all taxonomic levels; however, the CDS SSRs were most variable at the higher taxonomic levels. This is expected, as presumably there is less tolerance for mutation in the functional CDS region than in untranslated regions. For this same reason the CDS SSRs are also highly transferable. In summary, from the small number of primers tested the trend appears to be that 3′UTR SSRs are more

variable at the lower taxonomic levels (more related), the 5′UTR SSRs have intermediate variability, and the CDS SSRs tend to differentiate at the higher taxonomic level. These results will need to be verified with more screening. The availability of a selection of SSR sources (UTRs versus CDSs) may be valuable for targeting microsatellite use to studies of appropriate accessions.

Standard SSR libraries have a higher frequency of SSRs, up to 47% in the *Vitis* microsatellite consortium (VMC) compared with 2.5% in the EST-derived SSRs. The EST SSRs are, however, a by-product of other research, thus making the cost of finding them negligible, as they are filtered out computationally from the EST database for the generation of primers. The SSRs generated from ESTs are also more representative of all repeat motifs than are the enriched SSRs (Table 2). This is a direct result of the bias protocols for SSR enrichment. The proportion of $(GA)_n$ motifs in the EST-derived SSRs is much higher than the proportion of $(CA)_n$ repeat motifs. This characteristic is found in many plant species (Gupta et al. 1996). Powell et al. (1996) reported the $(AT)_n$ repeat to be the predominate motif in plants but this is apparently not the case in *Vitis*. As in other studies (Gupta et al. 1996), $(AAG)_n$ was the most common trinucleotide repeat in the EST SSRs; however, $(AAT)_n$ which is also often abundant in other plant species (Gupta et al. 1996) is apparently not common in *Vitis*. The average repeat length is longer in the VMC-enriched SSRs, in comparison to the EST SSRs.

In conclusion the EST SSRs are a viable source of polymorphic, highly transferable SSRs for mapping and identity applications in *Vitis*. Although only ten have been tested here, their utility has been illustrated and there is the potential to find approximately 1000 more useable *Vitis* SSRs from our EST project.

**Acknowledgments** We thank David OAG (DPI) for providing us with root stock plant material, and to the Australian Agriculture Research Institute for funding.

## References

Bowers JE, Meredith CP (1997) The parentage of a classic wine grape, Cabernet Sauvignon. Nature Genet 16:84–87

Delseny M, Cooke R, Raynal M, Grellet F (1997) The *Arabidopsis thaliana* cDNA sequencing projects. FEBS Lett 405:129–132

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci 70: 45–54

Lamboy WF, Alpha CG (1998) Using simple sequence repeats (SSRs) for DNA fingerprinting germplasm accessions of grape (*Vitis*) species. Am Soc Hort Sci 123:182–188

Milbourne D, Meyer RC, Collins AJ, Ramsay LD, Gebhardt C, Waugh R (1998) Isolation, characterisation and mapping of simple sequences repeat loci in potato. Mol Gen Genet 259: 233–245

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 7:215–222

Sasaki T, Song JY, Kogaban Y, Matsui E, Fang F, Higo H, Nagasaki H, Hori M, Miya M, Murayamakayano E, Takiguchi T, Takasuga A, Niki T, Ishimaru K, Ikeda H, Yamamoto Y, Mukai Y, Ohta I, Miyadera N, Havukkala I, Minobe Y (1994) Toward cataloguing all rice genes – large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. Plant J 6:615–624

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

Scott K, Playford J (1996) A DNA extraction technique for PCR in rainforest plant species. Biotechniques 20: 974–978

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

Wang X, Bowen B (1998) A progress report on corn genome projects at pioneer hi-bred. In: Plant and Animal genome VI Conference, SanDiego, California

Yamamoto K, Sasaki T (1997) Large-scale EST sequencing in rice. Plant Mol Biol 35:135–144