Theor Appl Genet (2002) 104:61–66

M. Rossetto · J. McNally · R.J. Henry

Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae

Received: 15 December 2000 / Accepted: 12 April 2001

Abstract Three EST-derived microsatellite loci from *Vitis vinifera* were amplified and sequenced across eight species of Vitaceae from four different genera. Phylogenetic analysis of the microsatellite's flanking regions produced informative results in congruence with previous studies. Generic relationships were respected and the data produced sufficient inter-specific variation to distinguish between *Cayratia acris* and *Cayratia saponaria*, two very closely related species. Overall, the sequence alignments showed that priming sites were conserved, whereas microsatellite repeats were present in most cases but structurally variable. The sequence data provided information on the evolutionary patterns of various microsatellite repeats and their correlation to evolutionary relationships among taxa.

Keywords Microsatellites · *Vitis* · Mutation · Flanking regions · ESTs

Introduction

Microsatellites (or simple sequence repeats, SSRs) are considered as being amongst the most-efficient molecular markers. They are abundantly and uniformly distributed throughout the genome, inherited codominantly, are highly reproducible and have simple screening requirements. These properties make SSRs particularly advantageous for applications such as population studies and mapping. The development of increasingly efficient enrichment strategies (Edwards et al. 1996; Cordeiro et al. 1999), the mining of expressed sequence tags (EST) for SSRs (Scott et al. 2000), and the success of cross-species transferability (Rossetto 2001) have expanded the popularity of these markers. Within the Vitaceae, SSRs have been previously developed for *Vitis vinifera* from genomic DNA libraries (Thomas and Scott 1993) as well as from EST databases (Scott et al. 2000). As previously shown in a number of other taxa, *V. vinifera* SSRs are transferable across cultivars and species (Scott et al. 2000, Rossetto 2001). The success of heterologous PCR amplification usually depends upon the evolutionary distance between the source and target species, with higher genomic homology being likely to translate in greater conservation of SSR flanking regions.

Despite their versatility, SSRs are generally not considered as optimal tools for phylogenetic studies. Empirical studies have shown that, at higher taxonomic levels, SSR alleles identical by size are not necessarily identical by descent (Orti et al. 1997; Doyle et al. 1998). The evolutionary mode of microsatellite repeats is such that new alleles are created following slipped-strand mispairing during DNA replication (Levinson and Gutman 1989; Schlötterer and Tautz 1992; Wierdl et al. 1997). It has been suggested that such events follow stepwise mutation models (single-step, random, or a combination of both: Shriver et al. 1993; Di Rienzo et al. 1994; Estoup et al. 1995; Goldstein and Pollock 1997) rather than the infinite-allele model. Consequently there may be constraints on allelic size-range limiting the genetic distances between distinct taxa and potentially counteracting the effects of genetic drift (Garza et al. 1995). As a result, one of the main problems associated with the use of SSRs for phylogenetic purposes is the absence of a fully validated evolutionary model that can be applied to any type of microsatellite repeat (Goldstein et al. 1995). Since interspecific allelic differences at the one locus are often more complex than simple changes in repeat number, allele frequency and distribution alone are generally inadequate to assess long-term evolutionary divergence.

Phylogenetic reconstructions based on sequence data have the advantage of relying on a large number of characters that follow defined evolutionary patterns (Li 1997). As a result, comparing gene genealogies is be-

Communicated by P. Langridge

M. Rossetto (✉) · J. McNally · R.J. Henry
Centre for Plant Conservation Genetics,
Southern Cross University, PO Box 157 Lismore NSW,
Australia 2480
e-mail: mrossett@scu.edu.au
Tel.: +61-2-66203458, Fax: +61-2-66222080

coming an increasingly popular phylogenetic approach, even if finding regions containing sufficient information can sometimes be problematic. Phylogenetic resolution between recently diverged taxa can be low because of the lack of informative characters, even when assessing rapidly evolving non-coding regions. For instance, in a recent study of the Vitaceae the combination of sequence information obtained from plastid DNA (*trn*L intron) and nuclear ribosomal DNA (ITS1) was not sufficient to clearly differentiate between closely related taxa such as *Cayratia acris* and *Cayratia saponaria* (Rossetto et al., submitted).

Vitaceae are woody climbers comprising 13 genera and about 700 species worldwide (Mabberley 1995). The family is characterised by leaf-opposed tendrils, which may be modified to form an inflorescence. The majority of the genera have predominantly tropical and subtropical distributions mainly from Africa through Asia and into the Pacific islands. In these latter regions, the Vitaceae often play an important role in rainforest habitats and vine thickets. Originally, only *Vitis* was recognised within Australia by Bentham (1863), but after later revisions this genus is now considered to be absent. Five other genera are represented in the Australian flora, with a total of 30 named species. Despite recent revisions, some uncertainty still exists about the phylogeny of the Australian Vitaceae, particularly *Cissus* for which remarkable morphological polymorphism has been recorded. Recent phylogenetic studies suggest that at least four taxa have been incorrectly included within that genus (Rossetto et al. 2001).

As previous studies have shown that expressed-sequences containing SSR repeats from *V. vinifera* are conserved across a number of taxa (Scott et al. 2000), the present research proposes to investigate the potential of these loci in providing useful phylogenetic information. Such an investigation was made possible by the availability of comparative evolutionary data on intra- and inter-generic relationships in the Vitaceae (Rossetto et al. 2001). The aim of this study was to investigate the potential resolution of SSR flanking-regions in phylogenetic studies, as well as the evolutionary patterns of SSR repeats. A total of eight species and three EST-derived SSR loci were selected and analysed for this purpose.

Materials and methods

Species selection and SSR amplification

Details on the study species, their origin and voucher details are presented in Table 1. Species selection in this study was aimed at representing inter-generic diversity within the Australian Vitaceae. Previous phylogenetic investigations based on plastid and nrDNA found that the genus *Cissus* is not monophyletic but structured into three distinct groups (Rossetto et al. 2001). These groups are represented in this study by *Cissus opaca* (belonging to the genus *Clematicissus*), *Cissus penninervis* (a true *Cissus*) and *Cissus sterculiifolia* (belonging to a new, yet to be named, genus). Thus these three taxa, despite being all *Cissus*, truly represent three different genera. The three *Cayratia* species were selected to represent the genus but also to include two very closely related taxa, *C. acris* and *C. saponaria*. *Tetrastigma nitens* was included as a genus-representative and finally *V. vinifera* was the species from which the SSRs were originally characterised.

Locus selection (scu06vv, scu08vv, scu15vv) was based on preliminary transfer studies across a number of different grape accessions (Scott et al. 2000) and other Vitaceae species (unpublished data). These three loci were chosen because they produced clear amplification products under standard PCR conditions for the majority of the eight test species. They also represent an unbiased sample in repeat length and position within the expressed sequence (scu08vv is located within the 5′UTR, scu06vv within the 3′UTR and scu15vv within the coding sequence).

As this study was aimed at assessing the potential of SSR flanking regions in examining taxonomic relationships, and not at defining phylogenetic relationships within the Vitaceae, sufficient information could be obtained from the species/loci combination selected.

PCR, cloning and sequencing

PCR conditions were as described in Scott et al. (2000), with the only difference being that the annealing temperature was lowered to 48°C for cross-species amplifications. PCRs that produced a homozygous product could be sequenced directly (forward and reverse), but if heterozygous patterns were obtained then cloning was required. At least two separate PCR products or clones were sequenced for each locus/species combination. Cloning was performed using the pGEM-T Easy vector system (specifically designed for cloning PCR products) following standard supplier's protocols (Promega). PCR amplification of ITS1 and the *trn*L (UAA) intron are described in Rossetto et al. (submitted) and the GenBank accession numbers for each sequence is indicated in Table 1. Cycle-sequencing was carried out using the Big-Dye terminator (Perkin Elmer) following standard supplier's protocols on an ABI 3700 thermocycler. Sequences were visualised on an ABI 377 automatic sequencer at the Australian Genome Research

Table 1 List of the eight taxa used in this study. Provenance (Qld, Queensland), voucher details (JCT, James Cook University Herbarium) and accession numbers for the plastid and nrDNA sequences are included

Genus	Species	Provenance	Voucher	GenBank # *trn*L	GenBank # ITS1
Cayratia	*acris* (F. Muell.) Domin.	Millaa Millaa (Qld)	B.Jackes 9960 JCT	AF300314	AF365985
	cardiophylla Jackes	Chillagoe (Qld)	B.Jackes 9950 JCT	AF300316	AF365987
	saponaria (Seem. Ex Benth.) Domin.	Millaa Millaa (Qld)	B.Jackes 9959 JCT	AF300315	AF365986
Cissus	*opaca* F. Muell.	Mt. Fox (Qld)	B.Jackes 9854 JCT	AF300301	AF365981
	penninervis (F. Muell.) Planch.	Paluma (Qld)	B.Jackes 9850 JCT	AF300300	AF365982
	sterculiifolia (F. Muell. ex Benth.) Planch.	Paluma (Qld)	B.Jackes 9841 JCT	AF300302	AF365983
Tetrastigma	*nitens* (F. Muell.) Planch.	Townsville (Qld)	B.Jackes 9903 JCT	AF300311	AF365984
Vitis	*vinifera* L.	–	–	AF300295	AF365988

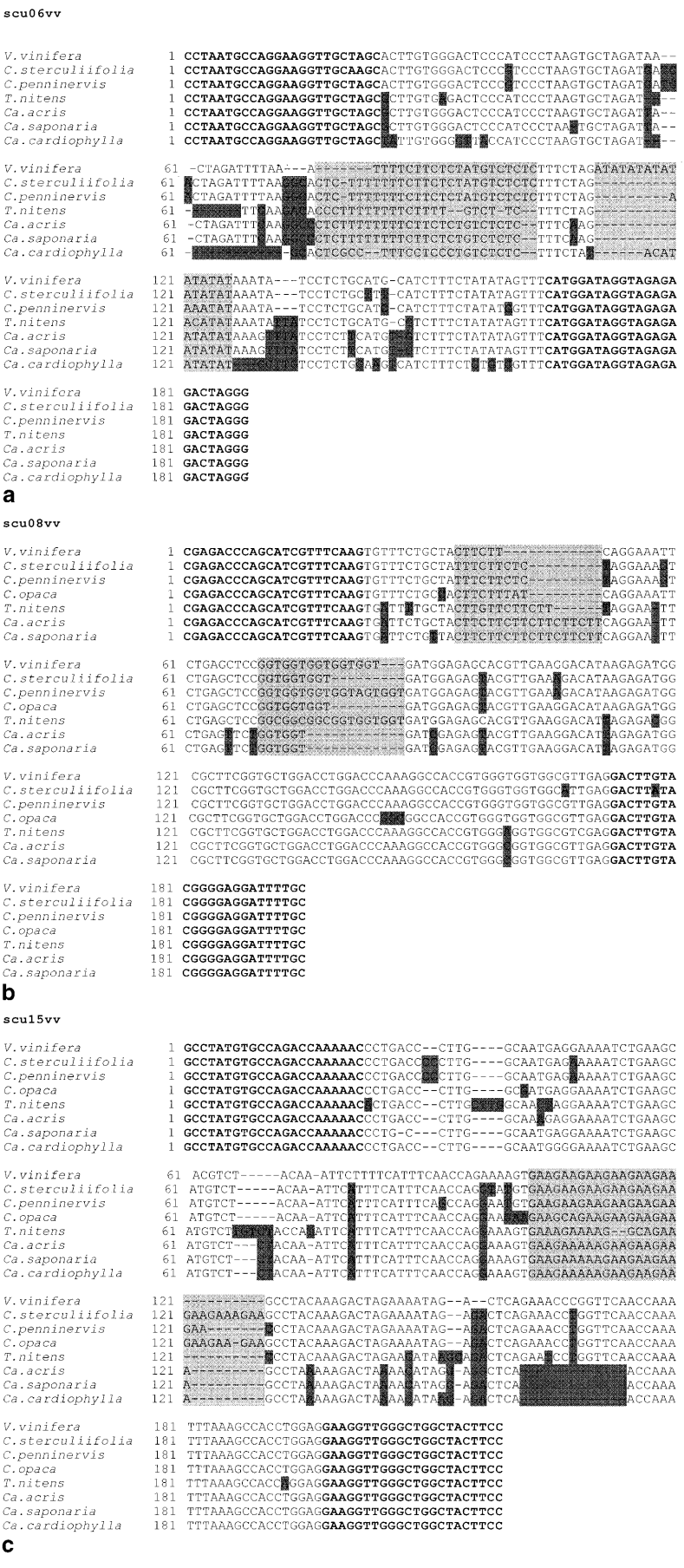

Fig. 1a–c Full sequence alignment of the three SSR sequences. The forward and reverse primer sites are in *bold*, SSR repeats and character substitutions are *highlighted* (*light and dark grey* respectively).
a scu06vv alignment,
b scu08vv alignment,
c scu15vv alignment

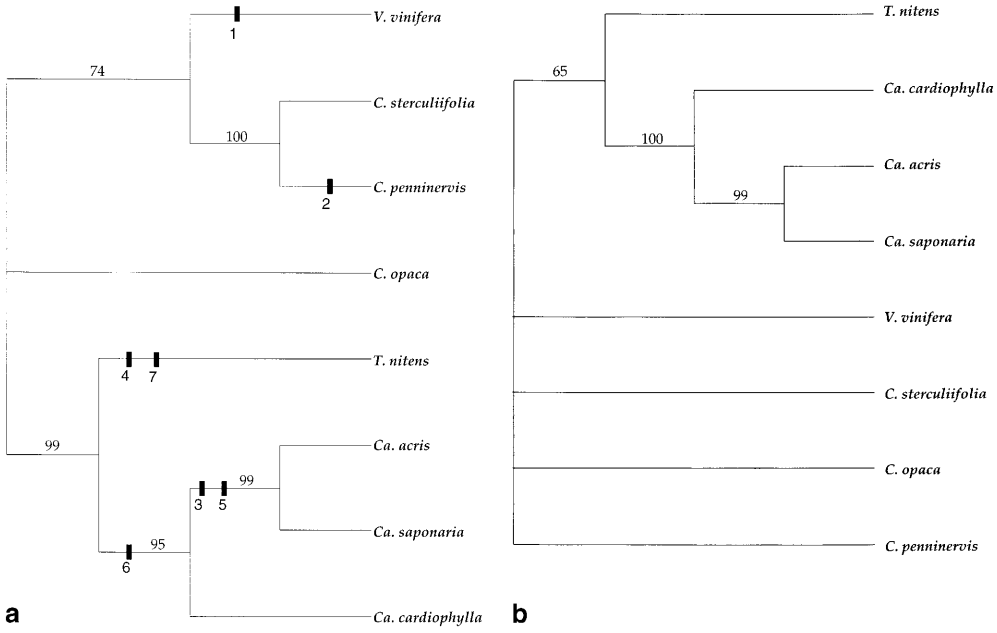

Fig. 2 **a** Strict consensus tree obtained from the combined analysis of the flanking regions from three EST-derived SSR loci. Bootstrap values are indicated on the branches. The position of the SSR repeat mutations are also *marked and numbered* (see text for details). **b** Strict consensus tree obtained from the combined analysis of the *trn*L intron and ITS1. Bootstrap values are indicated on the branches

Facility (Brisbane, Australia). The sequences obtained were aligned using Clustal W (Thompson et al. 1994).

Data analysis

The SSR sequences obtained were evaluated by maximum parsimony. The regions were analysed separately and as a combined data set. Information on parsimony informative indels (insertion/deletions shared by two or more taxa) was added to the sequence data as a presence/absence matrix, which carried equal weight as single base-pair substitutions. Phylogenetic relationships among the eight species were inferred with 1000 random sequence addition and tree bisection-reconnection (TBR) branch swapping, permitting five trees to be held at each step then swapped to completion. Tree searches were conducted via Fitch parsimony (equal-weight unordered characters state, Fitch 1971) using PAUP 4.0b4a (Swofford 2000). Clade robustness was measured by the bootstrap method (Felsenstein 1985) with 500 replicates, and using 1000 random sequence addition and tree bisection-reconnection (TBR) branch swapping, but again permitting only five trees to be held at each step. Fragments that did not amplify (scu06vv for *C. opaca* and scu08vv for *Cayratia cardiophylla*) were treated as missing data. The same method was used to analyse plastid and nrDNA regions (the GenBank accession numbers for these sequences are shown in Table 1).

Results

Amplification and sequencing were possible for all eight species with scu15vv, for all but *C. opaca* with scu06vv, and for all but *C. cardiophylla* with scu08vv (Fig. 1). The aligned matrices were 188-bp long for scu06vv (147 bp without the SSR repeats), 195 bp for scu09vv (160 bp without the SSR repeats), and 219 bp for scu15vv (189 bp without the SSR repeats). The combined data from the three sets of SSR flanking regions distinguished between all the taxa investigated, with a total of three substitutions (one at scu08vv and two at scu15vv) and two indels (one each at scu06vv and scu15vv) differentiating between *C. acris* and *C. saponaria*. The fact that each of the SSR loci could differentiate between these two species is particularly interesting, since no sequence differences were recorded for the *trn*L intron (Rossetto et al. 2001a) and a single indel distinguished these two taxa for ITS1 (Rossetto et al. 2001b).

As expected from the success of PCR amplifications, all priming regions were highly conserved (except for a single transition at locus scu08vv in *C. sterculiifolia*). Microsatellite repeats were present within most taxa at all loci, but length and repeat type varied (Fig. 1). At locus scu06vv a number of SSR repeats were present (Fig. 1a). The original repeat in *Vitis* was $(AT)_n$ but the same repeat was very small across the other taxa (maximum $(AT)_4$ in *C. cardiophylla*). However, two new repeats were found in most of the other taxa: a perfect $(T)_n$ followed by an imperfect $(TC)_n$ (Fig. 1a). Similarly, at locus scu08vv the original $(GGT)_n$ repeat from *V. vinifera* was shorter in most other taxa except for *C. penninervis* and *T. nitens*, for which the repeat was interrupted and compound respectively (Fig. 1b). As for the previous locus a new SSR repeat emerged $(CTT)_n$ but this time within *Cayratia* alone (Fig. 1b). Finally at locus scu15vv the original $(GAA)_n$ repeat was conserved across *Cissus*, interrupted for *Cayratia* and absent in *Tetrastigma* (Fig. 1c). The major changes in SSR repeat structure are marked in Fig. 2a.

The Fitch analysis for the combined flanking regions of the SSR sequences resulted in one most-parsimonious tree of 99 steps CI=0.94 and RI=0.91 (Fig. 2a). Strong bootstrap values supported two major clades, one grouping the three *Cayratia* and the one *Tetrastigma* analysed, the other including *C. penninervis*, *C. sterculiifolia* and *V. vinifera* (Fig. 2a). *C. opaca* was outside these two major clades. Within the major clades there was strong bootstrap support for two subclades grouping the two

Cissus and the three *Cayratia* species respectively. A very similar tree could be obtained from the data from locus scu15vv alone (data not shown). The equivalent tree obtained from the combined sequence data from ITS1 and the *trn*L intron is shown in Fig. 2b. This tree was less resolved than the SSR-based one, with only the *Tetrastigma-Cayratia* clade having bootstrap support and the rest of the species forming a polytomy.

Discussion

The combined analysis of the flanking sequences from three EST-derived SSR loci was more phylogenetically informative than the combined analysis of ITS1 and the *trn*L intron. The SSR flanking regions contained a high number of informative substitutions and indels, and such variation proved to be particularly useful for distinguishing between very closely related taxa such as *C. acris* and *C. saponaria*. Overall, the main intra- and intergeneric relationships were successfully resolved (Fig. 2). A previous study on the Vitaceae (Rossetto et al. 2001) found that the genus *Cissus* was not monophyletic. This is supported by the SSR data which clearly distinguishes between *C. opaca* (likely to be a *Clematicissus*; Rossetto et al. 2001) and the other members of the genus tested (Fig. 2a). The close relationship between *Tetrastigma* and *Cayratia* is also clearly corroborated by the SSR data and, as expected, the three *Cayratia* species tested form a well-supported subclade (Fig. 2a). The resolving power of this study is clearly limited by the number of taxa used; nevertheless, the main aim of this investigation was not to define phylogenetic relationships in the Vitaceae but rather to show the potential application of SSR flanking regions to such studies. Clearly, the phylogenetic information obtained from the three loci was accurate enough to respect relationships among taxa and sufficiently informative to distinguish between two very closely related species.

The sequence variation found within the SSR loci was significantly higher than that obtained from the analysis of plastid DNA alone (Rossetto et al. 2001). Plastid DNA is generally less informative than nuclear DNA but has the advantage of being single-copy, non-recombinant and usually uniparentally inherited. Multiple gene families in nuclear DNA, such as the internal transcribed spacer regions (ITSs), are more informative because they are highly variable and undergo concerted evolution (which ensures sequence orthology). However, excess variability, especially among more distantly related taxa, can complicate the alignment of ITS regions thus reducing their overall informativeness. In the present study, the ease of alignment of the SSR loci is likely to have contributed to their greater resolution power when compared to ITS1. Small et al. (1998) found that the variation within a genic region (alcohol dehydrogenase, *Adh*) was phylogenetically more informative than that found within plastid and nrDNA. As for the SSR flanking regions, *Adh* was also easier to align and provided sufficient information to resolve relationships between tetraploid *Gossypium* species (Small et al. 1998). However, special care is required when dealing with nuclear genes, as orthology can be more complex to demonstrate.

Empirical studies on pedigrees have estimated SSR mutational events to be between 10^{-2} to 10^{-3} per locus, per gamete, per generation (Weber and Wong 1993). Because of their different evolutionary constraints, generally it is preferable not to include SSR repeats in phylogenetic analyses. However, comparing SSR repeat structure among a number of related species provides interesting information on their evolutionary patterns. Taylor et al. (1999) summarised two major events in SSR evolution, 'birth' and 'death', both originating from the combination of two mutational events. During the birth of a SSR, a first mutational event produces sufficient length for the second mutational event, slipped-strand mispairing, to occur and produce repeat length variations. Death of a SSR involves a first mutation which interrupts and stabilises a previously perfect repeat, followed by a second mutation which deletes a large portion of that repeat. The alignment matrices in Fig. 1 suggest that both events have occurred within the three loci sequenced. The presence of $(AT)_n$ at scu06vv in *Vitis* alone (mutation 1, Fig. 2a) appears to be an example of the birth of a SSR, as no interruptions are present among related taxa. The remaining SSR repeats in this study represent mainly SSR death events. The $(GGT)_n$ repeat of scu08vv is present in *Cissus* and *Vitis*, with an initial interruption appearing in *C. penninervis* (mutation 2, Fig. 2a). Within *Cayratia* and *Tetrastigma* this repeat sequence has followed different paths, being close to complete deletion in the former (mutation 3, Fig. 2a) and having evolved into a compound $(GGC)_n$ $(GGT)_n$ repeat in the latter (mutation 4, Fig. 2a). Interestingly, the death of the $(GGT)_n$ repeat in *Cayratia* appears to have been balanced by the birth of a new $(CTT)_n$ repeat at a different site (mutation 5 in Fig. 2a). Finally, while the $(GAA)_n$ repeat in locus scu15vv is conserved across all *Cissus* and *Vitis* investigated, an interruption is present in all *Cayratia* (mutation 6, Fig. 2a) and the repeat has disintegrated in *T. nitens* (mutation 7, Fig. 2a). Some evolutionary information can be extracted from these data and this study is in agreement with the concept of SSR interruptions being frequently found within species more recently evolved than the one from which the loci were characterised (Taylor et al. 1999). Such findings suggest the potential existence of a correlation between mutational events within SSR repeats and evolutionary relationships across taxa.

In conclusion, in the Vitaceae the benefits of investigating cross-species amplification of SSR loci are multiple. Firstly, the conservation of loci is conducive to their use in genetic studies across a number of related species. For instance, six loci characterised for *V. vinifera* are currently being used in population studies of *Cissus hypoglauca* and *C. sterculiifolia*. Secondly, the variability recorded within the flanking sequences of a small number of loci is sufficient to distinguish between closely related species and to

examine evolutionary relationships. For such applications, this approach is likely to be more reliable than the analysis of allelic frequencies alone. This strategy can be particularly beneficial to the study of closely related agricultural varieties, especially those with previously characterised SSR libraries and/or EST databases. Finally, the findings suggest that further large-scale studies on mutational events within microsatellite repeats are likely to advance our understanding of SSR evolution.

Acknowledgments This research was partly funded by the Australian Agriculture Research Institute. The authors wish to acknowledge the input from C. Arnold and K.D. Scott.

References

Bentham G (1863) Flora Australiensis, vol. I. Reeve and Co, London

Cordeiro GM, Maguire TL, Edwards KJ, Henry RJ (1999) Optimisation of microsatellite enrichment technique in *Saccharum* spp. Plant Mol Biol Rep 17:225–229

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Nat Acad Sci USA 91:3166–3170

Doyle JJ, Morgante M, Tingey SV, Powell W (1998) Size homoplasy in chloroplast microsatelites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). Mol Biol Evol 15:215–218

Edwards KJ, Barker JHA, DalyA, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques 20:759–760

Estoup A, Garnery L, Solignac M, Cornuet JM (1995) Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. Genetics 140:679–695

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Fitch WM (1971) Toward defining the course of evolution: minimum change for a specific tree topology. Systematic Zool 20:406–416

Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. Mol Biol Evol 12:594–603

Goldstein DB, Pollock DD (1997) Launching microsatellites: a review of mutation process and methods of phylogenetic inference. J Hered 88:335–342

Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471

Levinson G, Gutman GA (1989) High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. Nucleic Acids Res 15:5323–5338

Li WH (1997) Molecular evolution. Sinauer Associates, Sunderland, Massachusetts

Mabberley DJ (1995) Vitaceae. In: Dassanayake MD (ed) A revised handbook to the flora of Ceylon. Amerind Publishing Co, New Delhi, pp 446–482

Orti G, Pearse DE, Avise J (1997) Phylogenetic assessment of length variation at a microsatellite locus. Proc Natl Acad Sci USA 94:10745–10749

Rossetto M (2001) Sourcing SSR markers from related plant species. Chapter 14. In: Henry RJ (ed) Plant genotyping – the DNA fingerprinting of plants. CABI, UK, pp 211–224

Rossetto M, Jackes BR, Scott KD, Henry RJ (2001) Intergeneric relationships in the Vitaceae: new evidence from cpDNA analysis. Genet Res Crop Evol

Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:211–215

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics 134:983–993

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. Am J Bot 85:1301–1315

Swofford DL (2000) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

Taylor JS, Durkin JMH, Breden F (1999) The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. Mol Biol Evol 16:567–572

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

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:388–396

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146:769–779