Theor Appl Genet (2003) 106:912–922
DOI 10.1007/s00122-002-1158-z

**V. Decroocq · M.G. Favé · L. Hagen · L. Bordenave
S. Decroocq**

Development and transferability of apricot and grape EST microsatellite markers across taxa

Received: 2 May 2002 / Accepted: 23 August 2002 / Published online: 3 December 2002

Abstract EST microsatellite markers were developed in apricot (*Prunus armeniaca* L.) and grape (*Vitis vinifera* L.). cDNA libraries from either apricot leaves or grape roots were used in an enrichment procedure for GA and CA repeats. The transferability of EST simple sequence repeat (SSR) markers from apricot and grapevine to other related and unrelated species was examined. Overall, grape primers amplified products in most of the *Vitaceae* accessions while the apricot primers amplified polymorphic alleles only in closely related species of the *Rosaceae*. In this taxonomic family, ten EST SSR loci were tested, and one single primer pair, PacB22, was amplified across species and sections in the *Prunoideae* and *Maloideae*. Sequencing of EST SSR loci in other species and genera confirmed a higher level of conservation in the microsatellite motif and flanking regions in the *Vitaceae* compared to the *Rosaceae*. Two distinct fragments of the PacB22 locus amplified across the *Malus* and *Pyrus* genera; however, while the coding region was highly conserved, the microsatellite repeat motif was no longer present. The banding pattern was explained by base substitution and insertion/deletion events in the intronic region of PacB22. This study includes the determination of the degree of polymorphism detected among species and genera in two unrelated taxonomic families and the evaluation of the information provided by the microsatellite repeats and the flanking regions.

Communicated by C. Möllers

V. Decroocq (✉) · M. G. Favé · L. Bordenave · S. Decroocq
INRA Centre de Bordeaux, U.R.E.F.V, 71 Avenue E. Bourleaux, 33883 Villenave d'Ornon France,
e-mail: decroocq@bordeaux.inra.fr
Tel.: 33/557122383, Fax: 33/557122384

L. Hagen
INRA, U.G.A.F.L, Domaine St Maurice, BP 94, 84143 Montfavet Cédex, France,

INRA Centre de Bordeaux, IBVM – UMR GDPP – Virology, BP 81, 33883 Villenave d'Ornon, France,

Keywords SSR · *Prunus* · *Vitis* · *Rosaceae* · *Vitaceae*

Introduction

Recently, interest has been directed towards the comparative studies of genome organisation in natural and cultivated populations. For this purpose, hypervariable markers such as SSRs (simple sequence repeats, also called microsatellites) have been characterised in a number of crop and tree species (see Powell et al. 1996 for a review). In many cases, microsatellite markers were shown to be transferable across taxa (Rossetto 2001). The conservation of flanking regions and primer binding sites in different genera has been reported, for example, between wheat and barley (Erpelding et al. 1996), soybean and other legume genera (Peakall et al. 1998), *Malus* and *Pyrus* (Yamamoto et al. 2001) and among the *Vitaceae* family (Thomas and Scott 1993; Rossetto et al. 2002). In the *Rosaceae*, cross-species amplification was tested on *Prunus* species and between *Malus, Fragaria* and *Prunus* genera (Cipriani et al. 1999; Downey and Iezzoni 2000; Sosinski et al. 2000; Dirlewanger et al. 2002). In cases where codominant and hypervariable markers can be transfered over a set of related and unrelated species, microsatellites can be of great interest for the analysis of genetic diversity and for studies on the evolution of species. However, they are usually distinguished by size, which is determined on high-resolution sequencing gels. Compensatory effects within the microsatellite locus such as base substitutions and insertion/deletion events are therefore not detected and the fragments amplified may not be those anticipated. This might lead to errors due to the high chance of independently arising, equally sized alleles (homoplasies) (Grimaldi and Crouau-Roy 1997). Moreover, transferability means not only that amplification but also that the microsatellite primers developed for one species can be used to detect polymorphism at homologous loci in related species. However, successful amplification does not guarantee the presence of the repeat motif within the sequence, leading to a loss in polymorphism across species.

We successfully used an enrichment procedure to isolate microsatellite repeats from genes expressed in grape roots and apricot leaves. We produced a set of microsatellite markers linked to ESTs (expressed sequence tags) and known genes that identify single variable loci. In the current report, the possibility of transferring EST SSR information across species and genera of the *Vitaceae* and *Rosaceae* taxa is investigated. The purpose here is not to define phylogenetic relationships across related and unrelated species in the *Vitaceae* and *Rosaceae* but rather to determine how informative a set of microsatellites is across two unrelated taxonomic families. Two types of information are compared in this study: the degree of polymorphism detected on high-resolution sequencing gels and the sequences of the microsatellite and its flanking regions.

Materials and methods

Construction of cDNA libraries enriched for microsatellites and primer design (Table 1)

Total RNA was extracted from mature leaves of apricot (*Prunus armeniaca* L., cultivar Stark Early Orange, SEO) and grape roots (*Vitis vinifera* L. cv Cabernet Sauvignon). RNA was extracted following the procedure described in Chang et al. (1993) and treated with DNAase RQ1 (Promega). PolyA+ RNA was purified using the Oligotex kit (QIAGEN). Double-stranded cDNA was synthesised with the Promega Riboclone Kit and cDNA larger than 200 bp was selected using spin columns (Pharmacia). The fragment ends were polished and, after purification, 3 ng of cDNA was ligated with the *Mlu*1 adaptor. Prior to enrichment, 1/20th of the ligation was amplified for 25 cycles with the 21-mer *Mlu*1 primer (Edwards et al. 1996). The enrichment procedure of Edwards et al. (1996) was used with the modifications described in Butcher et al. (2000). Enrichment for microsatellites was carried out using the total volume of the former amplification which was hybridised on CA- and GA-bound Nylon membranes (Appligene). Following two rounds of enrichment, 8 µl of eluted DNA was amplified by PCR with 0.8 µM of the 21-mer *Mlu*1 primer in 40 µl of 20 mM Tris, 50 mM KCl, 2 mM $MgCl_2$, 0.2 mM of each dNTP and 1 U of *Taq* polymerase (Life Technologies) using a Perkin Elmer 9700 thermocycler; 1/20th of the PCR product was cloned in 50 ng of pGEM-T vector (Promega) prior to transformation into DH5α *Escherichia coli* cells.

Randomly picked clones were amplified using M13 universal primers. Insert DNA was purified on Qiaquick PCR purification columns (QIAGEN) and directly sequenced using an ABI 377 automated DNA sequencer (Applied Biosystems). Oligonucleotide primers complementary to the regions flanking the identified repeat motifs were designed using the program Primer version 0.5 (National Biosciences, Plymouth, Minnesota), setting an annealing temperature of 57.5 °C. Database searches were carried out using the Advanced Blast program at the National Center for Biotechnology Information (Bethesda, Md.).

Cross-species amplification of EST SSR loci

Eight grape microsatellites isolated from an enriched root cDNA library of *V. vinifera* were tested among 62 individuals out of 46 different species in the *Vitaceae* family (see Table 2). Classification was as reported by Galet (1967). All accessions were obtained from the national collection of INRA (Institut National de la Recherche Agronomique) based at Montpellier and Bordeaux.

In the *Rosaceae* family, 21 *Prunus* accessions including interspecific crosses, one pear (*Pyrus communis* L.) and six apple (*Malus* × *domestica* Bortch.) cultivars and an interspecific cross between *Cydonia oblonga* (Miller) and *P. communis* were amplified with each of ten apricot EST SSR primer pairs (see Table 3). These accessions were obtained from the INRA germplasm collection at Bordeaux, Avignon and Angers. Classification was as reported by Rehder (1949).

Genomic DNA from *Rosaceae* individuals was prepared as described in Lefort and Douglas (1999). In the *Vitaceae*, DNA was prepared with the QIAGEN Plant DNEASY mini kit. Starting with 10 ng of genomic DNA as a template, PCR conditions were as follows: 2 mM of $MgCl_2$, 0.1 mM of each dNTP, 0.2 µM of each primer, 1× Buffer and 0.5 units of *Taq* polymerase (Sigma) in 25 µl final volume. Thermocycling conditions were 30 s at 94 °C, followed by 30 cycles of (94 °C for 30 s, 57.5 °C for 30 s and 72 °C for 30 s) and an extension step at 72 °C for 7 min, in a GeneAmp 9700 thermal cycler (Applied Biosystems).

Separation of alleles was performed on a 6% acrylamide sequencing gel in 0.5× TBE buffer at 80 W for 100 to 120 min. One to two microliters of PCR product was mixed with 3 µl of denaturing loading buffer containing 95% formamide, 0.25% bromophenol blue and 0.25% xylen cyanol, and 10 mM of EDTA. This mixture was heated for 5 min at 94 °C to denature the DNA before loading. Gels were stained with silver nitrate following the protocol detailed by Chalhoub et al. (1997) and bands were scored after gel scanning.

PCR cloning of SSR loci and sequence analysis

A dinucleotide $(CA/GT)_{18}$ microsatellite locus was selected in apricot, PacB22, and screened for sequence variation in the *Rosaceae* family. Another dinucleotide repeat, vvc34, was screened from *Vitaceae* genomic DNA. The PacB22 primer pair amplifies a fragment of about 260 bp in *P. armeniaca* L. and the vvc34 primers amplify a fragment of 200 bp in *V. vinifera*. PCR conditions were as indicated above. PCR products were purified on QiaQuick PCR purification columns (Qiagen) and cloned in the pGEM-T vector (Promega). Inserts were sequenced in duplicate from various individuals using an automated sequencing system (Genaxis, Nîmes, France). DNA sequences were aligned using the ClustalW software from InfoBiogen (http://www.infobiogen.fr).

Results

Identification of EST SSR loci

Within both enriched cDNA libraries, the insert size ranged from 300 to 1,500 bp with an average of 800 bp. Several clone sequences were discarded prior to primer design for several reasons which included insufficient length of the flanking sequences to design primers, too short sequences between the repeat motif and the poly-A tail or multiple cloned sequences. Details on primers which amplified a single locus in apricot and grape are summarised in Table 2. About 500-bp long sequences from the 5′ and/or 3′ ends of each clone were obtained. TBLAST-X analysis (http://www.ncbi.nlm.nih.gov/BLAST/) identified significant homology to known plant-coding regions for 81% of the apricot and 70% of the grape SSR containing clones (see Table 1). The remaining apricot and grape cDNAs represent novel transcripts. Enrichment in CA and GA motifs was up to 90% but this does not take into account clone redundancy. In apricot, 51.5% of the positive clones were unique based on sequencing data. One clone, PacC3, accounted for up to 19% of the overall library.

Table 1 cDNA microsatellite markers developed in apricot and grape: repeat motif, primer sequences and sequence homology. Sequences of the cDNA clones have been loaded in the dbEST Genbank Database under the following accession numbers: from BQ106729 to BQ106736 for the grape ESTs and BQ134639 to BQ134649 for the apricot ESTs

SSR name	Name and GenBank accession number of the most similar match	BLAST Expect value	Location of SSR on the sequence	Core motif	Primer sequences (5′ → 3′) forward and Reverse
P. armeniaca					
PacA10	*A. thaliana* chlorophyll a/b-binding protein T04049	1e-32	5′UTR	(GA)14	tgagcataattggggcaggccagagaagccatttcagt
PacA18	*A. thaliana* hypothetical protein AB007649	1e-04	5′UTR	(GA)20	tccaaacctaccgtttctcatcaacagcacaaacagaaccac
PacA33	*Nicotiana tabacum* chloroplast inner envelope membrane precursor T03230	4e-40	3′UTR	(GA)16	tcagtctcatcctgcatacgcatgtggctcaaggatcaaa
PacA49	Photosystem I subunit PSI-E (*N. sylvestris*) S72358	7e-36	5′ UTR	(GA)12	tcacagccagccagagagcagctgatgccatgccac
PacB22	*A. thaliana* unknown protein linked tail to tail to a hexokinase ATU19925	2e-32	ORF[a]	(CA)18	gaggtgcggcgatagttcgagctgaaagtcaattcagagtagtt
PacB26	*A. thaliana* MADS-box ANR1 protein AC007210	2e-03	ORF[a]	(CA)19	ccaatcatgaaatcataaagcaatgggatgtcctattgttttca
PacB35	No similarity	–	3′UTR[b]	(CA)14(GA)11	attgcgatttcggtctgttccatcccaaattgcttactt
PacC3	*N. tabacum* allyl alcohol deshydrogenase BAA89423	1e-34	3′UTR	(GA)16	tgacttgatcagactcgacattgcatttgcatttacaataga
PacC13	No similarity	–	3′ UTR[b]	(GA)10	gcttgctgctcatcatttacaataacaaccatattggagtatttac
PacC25	*A. thaliana* hypothetical protein AC007843	2e-20	3′UTR	(CA)15	gtgttttgacaagaaatgaattgtccattcgcagtaaaattaaac
V. vinifera					
vvc5	*A. thaliana* unknown protein T00623	6e-08	3′UTR	(GT)14(GA)12	ttagcttccatgcctgcttgtggcatctgaaccttaa
vvc6	No similarity	–	nd	(GA)14	ggttgaggactgaccattgacacaatccaagaagcatcctat
vvc7	*N. tabacum* CEN-like protein 4 AF145261	2e-63	5′UTR	(GA)9	gtccagtcccacaggctcccccaatcactctaccaacaa
vvc19	No similarity	–	3′UTR[b]	(GGA)2(GA)3ggaca(GA)7	tcagaatcagtctctttaatcctttgcggctgtttaaggcttt
vvc34	*A. thaliana* SKOR potassium channel AJ223357	4e-25	3′UTR	(CA)18	aggatgaaatgacatggatgacccatgttgatgagttcac
vvc62	No similarity	–	Nd	(GGA)7	tgtgcatgtggctaagctagaaaagggaactcaccacaaa
vvc71	*A. thaliana* ubiquitin-specific protease 2 AC004809	6e-17	3′UTR	(AG)3(CA)3(AG)3ggt(GTA)2(GCA)4	gagatgttgtgcttctgtgagttggcatgcaaagcagacat
vvc82	*A. thaliana* GDP-mannose pyrophosphorylase AC013258	6e-20	3′UTR	(GA)16	tgctgtatggcagaattgaacccaacaagactctcaatgtt

[a] Microsatellite is within an intron disrupting the predicted ORF and leading to alternative splicing (PacB22), or the stop codon could not be determined precisely (PacB26)
[b] 3′-prime end position determined by the presence of a polyA tail downstream of the microsatellite motif

Table 2 Amplification and Polymorphism of 8 eight EST SSR loci in a variety of *Vitaceae* species and genera showing allele sizes (base pairs) detected. The left column indicates species and accession numbers chosen for testing transferability of eight grapevine EST SSRs; mb indicates multiple banding patterns. Classification is as reported by Galet (1967)

Individuals/species tested	Geographic origin	VVC5		VVC6		VVC7		VVC19		VVC34		VVC62		VVC71		VVC82	
Ampellocissus chantinii	Africa (Cameroun)	178		115	119	152		83		182	192	161		93		–	–
Cissus quadrangularis	indochinese peninsula	162		111		118		97		mb	–	161		–	–	256	
Cissus voinieriana[b]	indochinese peninsula	–	–	111		–	–	87		–	–	–	–	–	–	–	–
Parthenocissus henryana	Asia Minor/Europe	144	172	109		–	–	85		182		184	195	85		197	201
Parthenocissus semicordata	Asia Minor/Europe	154		105		96		–	–	–	–	–	–	–	–	–	–
Ampelopsis orientalis	Asia Minor	176	211	116	119	98	112	97		–	–	–	–	–	–	–	–
Ampelopsis japonica	Eastern Asia	154	162	105		94		97		mb	–	–	–	–	–	–	–
Ampelopsis aconitifolia	China	161	167	105		–	–	97		–	–	203	212	–	–	–	–
Ampelopsis aegirophylla	China164		105		96		97		192		181		90		250		
Ampelopsis delavayana	China	162		105		94		97		mb		204		88		200	
Ampelopsis heterophylla #2245	China	–	–	–	–	94		–	–	–	–	202	206	–	–	208	222
Ampelopsis indivisa #10085	USA	169		107		132	136	97		180	208	196		85	98	–	–
Muscadinia rotundifolia est	USA/Florida	–	–	111	119	98	102	99	102	188		208		99	126	231	245
Muscadinia rotundifolia ouest	USA/Florida	180	182	111		91	92	100	106	188		195	202	91		216	
Vitis riparia #10128	Northern America	145	153	119	121	–	–	102	104	178	190	189		103		–	–
Vitis riparia #10709	Northern America	153	170	119		–	–	102		178	190	187		107		235	
Vitis riparia cv. RGM	French rootstock	145	153	119	121	–	–	102	104	178	190	189		103		202	218
Vitis candicans #10189	USA	179	200	111	113	99	112	100	106	186	200	186	189	99	103	213	219
Vitis candicans #10196	USA	180		111		96		93		186		–	–	–	–	–	–
Vitis coriacea	USA	163	179	111	113	99	114	94	100	188		185	203	85		211	213
Vitis Simpsonii #10968	USA	168	198	113		98	120	103		180	188	189		101		191	200
Vitis champinii #10164	USA	145	179	111	117	96	98	92	104	186	191	189	200	89	119	–	–
Vitis longii #587-02	USA	169	186	117	119	98	102	100	102	178	191	186	190	97	101	208	237
Vitis solonis= longii	USA	146	162	117	121	–	–	100	104	178	190	196	200	–	–	196	220
Vitis doaniana #10165	USA	153	179	111	129	96	100	92	104	178	188	193	200	93		–	–
Vitis doaniana #10179	USA	196		119		96	104	100	102	178	191	189	196	99	107	214	222
Vitis mexicana	USA	171		119	121	98	102	100	102	190	196	198	202	101		198	199
Vitis arizonica	USA	175		117	119	–	–	98	106	190		193	196	–	–	194	204
Vitis berlandieri #10099	USA	162	164	111	115	96		100		192		199		88	99	–	–
Vitis berlandieri #10178	USA	157	169	111		96	110	100		190	198	198	201	82	103	204	206
Vitis berlandieri #10594	USA	169		111		96	100	100	102	186	190	198		84		199	
Vitis labrusca #10308	USA	156	165	113		100	102	106		179	184	186		99		207	209
Vitis labrusca #11056	USA	156		113		100	102	106		185		186	188	93	95	217	222
Vitis cinerea #10139	USA	152		111		100		–	–	186		–	–	–	–	197	201
Vitis cinerea #10943	USA	172	178	111	113	96		–	–	191		191		83	101	216	238
Vitis rupestris #10334	USA	153	170	119		98		108		178	207	188	189	99	107	201	235
Vitis rupestris #10400	USA	159	176	119		98		102	112	200	207	186	188	99	107	215	218
Vitis rupestris #10738	USA	153	157	121		98	106	102	108	178	190			99	107	207	218
Vitis aestivalis #10055	USA	154	184	113		96		99		180	208	188		85	89	196	200
Vitis aestivalis #11051	USA	176	177	113		96		98		180	190	189	192	85	91	197	210
Vitis bicolor #11218	USA	180	202	113		96		96	98	178	180	189		85	87	181	197
Vitis bicolor #11219	USA	–	–	–	–	96		–	–	–	–	189	197	–	–	192	204
Vitis lincecumii #10988	USA	174	210	113		96	98	98		178		189		85		183	211
Vitis cordifolia #11018	USA	–	–	–	–	94	96	–	–	–	–	–	–	109	109	203	231
Vitis cordifolia #11551	USA	195	196	–	–	108		–	–	188	194	193	197	–	–	216	218
Vitis monticola #10111	USA	163	171	113		94	100	101		–	–	185	192	83	107	–	–
Vitis rubra #10168	USA	154		113		108		98		184	198	189		93		189	
Vitis rubra #10919	USA	–	–	–	–	108	112	–	–	179	191	189		91		203	231
Vitis amurensis I	Eastern Asia	–	–	118	120	98	102	103		179	208	189	195	103	112	202	213
Vitis amurensis II	Eastern Asia	145		118	120	102	104	103		201	208	190	196	103	113	231	
Vitis thunbergii	Eastern Asia	–	–	–	–	106		–	–	179	191	196		112		203	
Vitis flexuosa #2484	China	151	188	118		96	98	97		179	191	193	199	89	97	207	
Vitis piazeskii	China	189		116		96		97		188		189	198	79	88	201	212
Vitis betulifolia	China	166	185	116	118	96	98	97		179	191	193	196	81	103	210	216
Vitis pagnucii	China	172	176	119		98	106	101	103	196	200	190		83	101	187	211
Vitis pentagona	China	160	170	113		96		97		192		199	202	73	81	193	195
Vitis reticulata	China	171		118		–	–	103		196	200	187	189	83	101	196	219
Vitis microsperma	China	152		120	122	98	102	103		–	–	189	192	107		–	–
Vitis davidii	China	165		118	120	96	98	97		186	–	194		103		203	219
Vitis romaneti	China	–	–	116	128	–	–	97		184	–	195		97	107	207	
Vitis sylvestris	Europe	–	–	113		96		97		185	–	212		91	105	203	220
Vitis vinifera cv Cabernet Sauvignon	Europe	161	166	113	125	96		96	98	200	207	182	208	89	97	214	232
Hybride riparia cinerea (börner)	German accession	145	149	113	119	96	106	100	102	178	190	197	201	87	103	203	
Total number of allelles amplified in the Vitacees			43		16		19		17		20		28		27		42

[b] Formerly classified as Cyphostemma voinieriana

SSR length polymorphism

Identical PCR conditions and thermal reactions were used for all EST SSR primers. Optimisation of the PCR conditions especially for relatively distant taxa was not attempted in order to avoid amplification of false positives. Separation of the fragments by standard polyacrylamide sequencing gels and visualisation by silver staining gave, in general, a common "stuttered" banding pattern, but the fragments could easily be distinguished at a resolution of 1 bp. The allele sizes were recorded in base pairs and, for this, the most intense upper band was used. Alleles were scored twice independently by the authors. The degree of polymorphism was counted as the absolute number of different alleles per microsatellite locus.

Of the eight primer pairs that amplified across *Vitaceae* DNA, all EST SSR loci were variable, two being highly polymorphic, vvc5 and vvc83, with 43 and 42 alleles respectively among 62 accessions tested (Table 2). In the family *Vitaceae*, most of the EST SSR loci in *Cissus* and *Ampelopsis* were homozygous, while most primers failed to amplify DNA from *Cissus voinieriana* and *Parthenocissus semicordata* with the exception of two or three loci. However, since only one accession was tested per species, we can not state whether the EST SSR loci are truly homozygous or not in those species and genera. Surprisingly, the vvc34 primer pair amplified more than two fragments in *Cissus quadrangularis*. This could be explained by gene duplication but cannot be confirmed without segregation analysis in progeny arrays.

During cross-taxa amplification, most of the blanks due to mis-amplification were detected in genera distant from *Vitis*. These results confirm a negative correlation between the phylogenetic distance and successful amplification. The same tendency was noticeable when considering the level of polymorphism. Interestingly, for the less polymorphic SSR loci (vvc6, vvc7 and vvc19), the low level of allelic variation was not due to a failure in amplification but instead reflects the amplification of predominant alleles.

The most successful cross-species amplification of polymorphic EST SSR loci was within the genus *Vitis*. There were however a number of exceptions including *Vitis rubra* accession #10168 which was homozygous at all, but one (vvc34), loci.

All ten apricot EST SSR loci transferred to closely related species such as *Prunus domestica*, which is a member of the same subgenus as *P. armeniaca* L., *Prunophora* (Table 3). In the *Prunophora* subgenus, the level of ploidy can differ greatly from 2n = 2*x* to 6*x* but in most cases, EST SSR loci were polymorphic with two to five alleles per genome. Loci presenting two alleles in the apricot genome were the most polymorphic in related species, except for PacC3 where null alleles were determined by analysis in plum full-sib segregating pedigrees (data not shown). Null alleles can be due to changes in flanking regions, for example at the PacC3 primer binding sites, that prevent primer annealing and result in no amplification.

Across different subgenera, amplification of heterozygote alleles was attained within the subgenus *Cerasus*, mostly in *Prunus cerasus* species (2n = 4*x*), while the *Prunus persica* accessions were in general homozygous at the respective loci. Table 3 does not show any predominant allele except in intraspecific microsatellite amplification of *P. persica* DNA, confirming a high level of homozygosity in this species.

Among ten EST SSR loci tested, only the PacB22 primer pairs yielded an unambiguous banding pattern across the *Rosaceae* genera. Interestingly, in the *Maloideae*, this locus showed a very similar allelic pattern between the two species, *Malus* × *domestica* and *Pyrus communis*, as well as successful amplification in the *Cydonia* × *Pyrus* interspecific accession, *Pyronia veitchii*. Two clear bands of equal size with no stutters were detected in four of the *Malus* cultivars and in the pear and *P. veitchii* accessions.

With the exception of PacB22, cross-genera amplification of EST SSR loci failed or produced a number of bands that could not be interpreted (see PacA10 and PacB35). PacB22 is one of the most-polymorphic EST SSR loci, together with loci PacA18 and PacC13, resulting in 22 to 23 alleles for 29 accessions tested (Table 3). However, this cannot be compared to the most polymorphic vvc loci in the *Vitaceae* since, in the *Rosaceae*, we have to account for polyploidy. This can increase considerably the number of alleles observed per locus.

Comparison of nucleotide sequences of the microsatellite locus vvc34 among the *Vitaceae* species

The PCR products amplified across taxa by the vvc34 primer pair were cloned and sequenced from ten species of the genus *Vitis*. The sequences were analysed to verify the nature of the amplification products compared with one of the two alleles sequenced in *V. vinifera*, cv Cabernet Sauvignon. The vvc34 microsatellite repeat was systematically present but shorter than in the source cultivar. The alignment by ClustalW analysis revealed a high degree of conservation of vvc34 SSR flanking regions especially across the *Vitis* genus (Fig. 1). In this genus, allelic variation and length variability within a species and across the *Euvitis* and *Muscadinia* subgenus was mainly due to a length variation of the microsatellite alone, combined with point mutation and short insertion/deletion events (2–4-bp long, Fig. 1) within the flanking regions. An interrupted repeat was detected in the related *Vitis* subgenus, *Muscadinia* but not in the *Euvitis* species. Interestingly, a 2-bp indel placed just downstream of the vvc34 forward primer is observed in two geographically distinct species, *Vitis mexicana* (American) and *Vitis davidii* (Asian).

In this sequencing analysis, DNA sequences from *Parthenocissus henryana, Ampellocissus chantinii* and *C. quadrangularis* were added as representatives of the more distant genera. Figure 1 shows that, despite point mutations and rather important length variation, the microsatellite repeat is still present. However, it can

Table 3 Amplification and polymorphism of ten EST SST loci in a variety of *Rosaceae* species and accessions. The asterisks mark the individuals from which the apricot PacB22 microsatellite locus was sequenced. In italics, amplified products of unexpected size. *n* indicates the level of ploidy; *nd*, non-determined Classification as reported by Rehder (1949)

Species	individuals tested	n	PacA10	PacA18	PacA33	PacA49	PacB22	PacB26	PacB35	PacC3	PacC13	PacC25
Prunoideae	*Subgenus Prunophora*											
P. armeniaca	SEO	2*x*	107	174 178	188 196	107	238 260*	200 209	207 209	164 166	110	193 205
	Polonais	2*x*	104	168 176	186 188	107	237	200 202	209 224	159 170	110	188 193
	Scréara	2*x*	107	163 176	187 188	107	238 250	197 200	209	133 166	114	188
	Goldrich	2*x*	104	166 174	186 196	107	238 250	197 200	209 224	166 170	110	195
	Moniqui	2*x*	107	168	188	107	238 250	202 206	209 224	166 170	110 114	188
P. domestica	Cacanska najbolja	6*x*	86 88	151 171	174 184	100 102	232 249	170 184	214 218	140	114 116	182 185
			90	172 178	192 200	104	279	191 193	220 222		126 128	194
				186					230		144	
	Jojo	6*x*	86 88	167 168	187 189	102 116	232 240	170 186	218 220	140	114 116	180 185
			90				249 258	191	224 230		128 142	187 194
							279		232			
	Cacanska rodna	6*x*	86 90	149 171	181 187	102 110	232 245	170 184	208 218	140	116 126	180 182
				172 179	201	116	258	186 191	224 230		142 146	185 194
				180					232			
Interspecific	*P. marianna* × *cerasifera*	3x	84 86	166 188	171 197	102 103	228 240	188 192	209 231	157 159	98 118	192 195
	or domestica, cv GF8.1		120									200
	P. cerasifera ×	2*x*	86 129	164 170	170	93 102	238	190	208 214	132	101 123	192 195
	(*P. amygdalus* × *persica*)											
	P. japonica × *spinosa*	4*x*	86	150 159	171 187	100 102	235* 244*	175 180	220 230	140 144	105 125	182 188
	cv. Jaspi			163	189 197			190			131 150	190 192
Prunoideae	*Subgenus Amygdalus*											
P. persica	Summergrand	2*x*	102	0	182	93	252	190	229	132	102	184
	GF 305	2*x*	102	189	182	93	250*	190	228	132	102	184
	S2660	2*x*	102	174	170 182	93	251	190	228	132	102	184
Prunoideae	*Subgenus Cerasus*											
P. avium	Burlat	2*x*	106	145	0	120 126	239*	178	217	0	124 131	0
	Summit	2*x*	106	145	0	122 126	237 239	178 180	215 227	0	114	0
P. cerasus	V 2868	4*x*	92 106	0	0	0	0	0	0	0	114	0
	V 3711	4*x*	92 106	145	0	118 124	239	178	213 226	130 134	107 109	190 194
											114 130	
	V 3715	4*x*	92 106	145 168	0	118 124	237 239	178 180	213 226	130 134	107 109	190
				180 183							114 130	
	V 2327	4*x*	91 94	145 168	0	97 118	239*	178	213 226	130 134	nd	194
			105			126						
Maloideae	Discovery	2*x*	204	0	0	0	220* 227*	0	0	0	0	0
id Malus	TN 10.8	2*x*	204	0	0	0	220 227	0	0	0	0	0
× domestica	MM 105	2*x*	0	0	0	0	220* 227*	0	0	0	0	0
	MM 106	2*x*	0	0	0	0	0	0	0	0	0	0
	×3460	2*x*	0	0	0	0	220 227	0	0	0	0	0
	×3485	2*x*	0	0	0	0	0	0	199	0	0	0
P. communis	perry cv. Pine	2*x*	0	0	0	0	214* 221*	0	207	0	0	0
Interspecific *Cydonia* × *Pyrus*	*P. veitchii* cv Guillaumin	2*x*	86	0	0	0	214 221*	0	0	0	0	0
Total number of alleles amplified in the *Rosaceae*			15	23	15	15	22	16	22	13	22	13

```
1-   AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCGCCCACACACACACACACACACACACACACAATGATTATTAAATT
2-   AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCGCC------CACACACACACACACACACACACAATGATTATTAAATT
3-   AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTTGCGCGCGCG----------------CACACACACACAAAATGCTTATTAAATT
4-   AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTC-----------------CACACACACACACACACACACAATGCTTATTAAATT
5-   AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCGCGCG----------CACACACACACACACACACAATGCTTATTAAATT
6-   AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTC-----------CACACACACACACACACACACACACAATGCTTATTAAATT
7-   AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCGCGCG----------CACACACACACACACACACAATGCTTATTAAATT
8-   AGGATGAAATGACATGGATGACAAA--GAC--GAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCGCGCG--CACATACACACACACACACACACACAAAATGCTTATTAAATT
9-   AGGATGAAATGACATGGATGACAAA--GAC--GAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCGCGCG------------CACACACACACACAAAATGCTTATTAAATT
10-  AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCGCGCG----------CACACACACACACACACACAATGCTTATTAAATT
11-  AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTTGCGCGCGCG--------------CGCACACACACACAAAATGCTTATTAAATT
12-  AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCGCGCG--------------CACACACACACACACAATGCTTATTAAATT
13-  AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCG------------CACACACACACGCATACACACAATGCTTATTAAATT
14-  AGGATGAAATGACATGGATGACATAAAGATTGGAGTGTATTTATACAGCAAAAGAACAAATGATATTTTACTGAGTT-------------------CACACATACACAAACACAAGGCTTATTAAATT
15-  AGGATGAAATGACATGGATGACATAAAGACT-GAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTT--------------------CACACACACACACACACACAAAGCTTATTAAATT
16-  AGGATGAAATGACATGGATGACATAAAGACT-GAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTT----------------------CACACACACACACACACAAAGCTTATTAAATT
17-  AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCGCGCAAA--CACACACACACACACACACACACACAATGCTTATTAAATT
18-  AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCGCGCAAACACACACACACACACACACACACACACAATGCTTATTAAATT
19-  AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCACGC------------------ACACACACACACAATGCTTATTAAATT
20-  AGGATGAAATGACATGGATGACAAA--GACTGGAGTGTGTTTATACAACAAAAGAACAAATGATATTTTACTTTCGCGCGCGCG----------CACACACACACACACACACAATGCTTATTAAATT
     ************************   ***  ******************************************                  ******* ** ***** *   ** * **********

1-   AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
2-   AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
3-   AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
4-   AAAA-----ATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
5-   AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
6-   AAAA-----ATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
7-   AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
8-   AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
9-   AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
10-  AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
11-  AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
12-  AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
13-  AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
14-  AAAATTTGAATGCCTTGTCATAACATGGCATGAAAAA-CTGTAAGGGTCAAAGAGTGAACTCATCAACATGGGT
15-  AAAATCTGAATGCCCTGTCATAAGATGGCATGAAAAAACTGCAAGACTCAAGGGTGAACTCATCAACATGGGT
16-  AAAATCTGAATGCCCTGTCATAAGATGGCATGAAAAAACTGCAAGACTCAAAGGGTGAACTCATCAACATGGGT
17-  AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
18-  AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
19-  AAAA-----ATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
20-  AAAATCTGAATGCCTTGTCATAAGATGGCATGAAAAAACTGTAAGGCTCAAAGAGTGAACTCATCAACATGGGT
     ****     *****************************************************************
```

Fig. 1 Nucleotide comparison of vvc34 EST SSR locus in ten related species and four genera of the *Vitaceae* family. From *1 to 12*, the subgenus *Euvitis* comprising European (*1–2*) the *V. vinifera* accession number in Genbank BQ106729; *3*, *Vitis sylvestris*), american (*V. riparia 4–5*, *labrusca 6* and *mexicana 7–8* respectively) and asian (from *9 to 12*, *V. davidii*, *Vitis amurensis*, *Vitis piazeskii*, *V. romaneti* respectively) species; *13* subgenus *Muscadinia*, species *Vitis rotundifolia*; *14 P. henryana*; *15 and 16 A. chantinii*; *17 to 20 C. quadrangularis* The microsatellite repeat is *underlined*. The *dot* above the sequence indicates interruption in the perfect (CA) repeat in *Muscadinia rotundifolia*, and *arrows* the forward and reverse primers

be either a compound microsatellite repeat as it is in *V. vinifera* (succession of GC and CA repeats), or a perfect (CA) repeat. It is also interrupted in two distinct positions in *P. henryana*. Four distinct fragments were amplified and cloned from *C. quadrangularis*. Surprisingly, one of those presents a deletion downstream to the motif repeat similar to *Vitis riparia* and *Vitis labrusca*, while those species are definitely not related. This might be due to the fact that this indel belongs to an extensive hypervariable region, prone to mutation and deletion. It draws attention to potential pitfalls that arise from using flanking sequences too close to the hypervariable microsatellite region for phylogenetic analysis.

Length variation and sequence analysis of the PacB22 flanking regions in the *Rosaceae*

The PCR products amplified across taxa by the PacB22 primer pairs were cloned and sequenced from three subgenera of the *Prunoideae* and two *Maloideae* genera. Along with the microsatellite repeat, 66 nucleotides of the upstream coding region were cloned. They displayed almost perfect conservation at the nucleotide level across species and genera (Fig. 2). One base substitution was observed at position 32 which distinguished members of the *Prunoideae* from the *Maloideae*. Single or double consecutive base substitutions were detected in the partial coding region but they were either species-specific (position 19–20) or allele-specific.

Across the *Prunoideae* subfamily, the microsatellite repeat was present but much shorter, down to eight (CA/GT) motifs in *P. cerasus* and *P. avium* (Fig. 2). This would explain the low level of polymorphism observed across sections in the *Prunoideae*. Interestingly, the PacB22 flanking region was perfectly conserved apart from the few base substitutions within the coding region.

Most of the transition events were detected between alleles from the *Prunoideae* subfamily and species of the *Maloideae* subfamily. Eleven nucleotide substitutions in the region upstream to the primary tandem repeat clearly distinguish the *Prunoideae* and *Maloideae* accessions. Moreover, Fig. 2 shows that electrophoretic size variation detected in the *Maloideae* was not due to variation in the microsatellite repeat numbers but instead to the occurrence of two deletion events present in the PacB22 flanking region. The first indel and several nucleotide substitutions demonstrate similarities between alleles (named A and B arbitrarily) across two *Malus* cultivars and the *Pyrus* and *Pyronia veitchii* accessions. To determine whether these were from different loci, allele segregation was analysed in *Malus* × *domestica* F1 progeny cv Dis-

```
Primer PacB22b Forward
         10        20        30        40        50        60        70        80        90       100       110       120       130
   G  A  A  I  V  V  V  W  L  S  S  I  V  V  G  A  V  N  S  V  P  L [
1-  GAGGTGCGGCGATAGTCGTGGTGTGGTTGTCTTCAATAGTTGTTGGTGCCGTCAATTCAGTGCCACTTGTAAGAGATTTCTTTTGAAACCAGTTGCCTCTAGATTTTGCAATTTAGAATCTATTTTGATTAT
2-  GAGGTGCGGCGATAGTCGTGGTGTGGTTGTCTTCAATAGTTGTTAGTGCCGTCAATTCAGTGCCACTTGTAAGAGATTTCTTTTGAAACCAGTTGCCTCTAGATTTTGCAATTTAGAATCTATTTTGATTAT
3-  GAGGTGCGGCGATAGTCGCTGTGTGGTTGTCTTCAATAATTGTTGGTGCCGTCAATTCAGTGCCACTTGTAAGAGATTTCTTTTGAAACCAGTTGCCTCTAGATTTTGCAATTTAAATTCTATTTTGATTAT
4-  GAGGTGCGGCGATAGTCGCTGTGTGGTTGTCTTCAATAGTTGTTAGTGCCGTCAATTCAGTGCCACTTGTAAGAGATTTCTTTTGAAACCAGTTGCCTCTAGATTTTGCAATTTAGAATCTATTTTGATTAT
5-  GAGGTGCGGCGATAGTCGCTGTGTGGTTGTCTTCAATAGTTGTTAGTGCCGTCAATTCAGTGCCACTTGTAAGAGATTTCTTTTGAAACCAGTTGCCTCTAGATTTTGCAATTTAGAATCTATTTTGATTAT
6-  GAGGTGCGGCGATAGTCGCTGTGTGGTTGTCTTCAATAGTTGTTGGTGCCGTCAATTCAGTGCCACTTGTAAGAGATTTCTTTTGAAACCAGTTGCCTCTAGATTTTGCAATTTAGAATCTATTTTGATTAT
7-  GAGGTGCGGCGATAGTCGCTGTGTGGTTGTCTTCAATAGTTGTTGGTGCCGTCAATTCAGTGCCACTTGTAAGAGATTTCTTTTGAAACCAGTTGCCTCTAGATTTTGCAATTTAGAATCTATTTTGATTAT
8-  GAGGTGCGGCGATAGTCGCAGTGTGGTTGTCCTCAATAGTTGTTGGTGCCGTCAATTCAGTGCCACTTGTAAGTGATTTCTTCGGAAACCACATACCTCTAGGTTTTTCAGTTTTGAA------------AT
9-  GAGGTGCGGCGATAGTCGCTGTGTGGTTGTCCTCAATAGTTGTTGGCGCCGTCAATTCAGTGCCACTTGTAAGTGATTTCTCTGGAAACCACATACCTCTAGGTTTTTCAATTTTGAGTCCATTT-G---AT
10- GAGGTGCGGCGATAGTCGCAGTGTGGTTGTCCTCAATAGTTGTTGGTGCCGTCAATTCAGTGCCACTTGTAAGTGATTTCTTCGGAAACCCCATACCTCTAGGTTTTTCAGTTTTGAA------------AT
11- GAGGTGCGGCGATAGTCGCTGTGTGGTTGTCCTCAATAGTTGTTGGCGCCGTCAATTCAGTGCCACTTGTAAGTGATTTCTCTGGAAACCACATACCTCTAGGTTTTTCAATTTTGAGTCCATTT-G---AT
12- GAGGTGCGGCGATAGTCGCTGTGTGGTTGTCCTCAACAGTTGTTGGTGCCGTCAATTCAGTGCCACTTGTAAGTGATTTCTTTGGAAACCACATACCTCTAGGTTTTTCAGTTTTGAA------------AT
13- GAGGTGCGGCGATAGTCGCTGTGTGGTTGTCCTCAACAGTTGTTGGCGCCGTCAATTCAGTGCCACTTGTAAGTGATTTCTCTGGAAACCACATACCCCTAGCTTTTTCAAGTTTGAGTCCATTT-GATTAT
14- GAGGTGCGGCGATAGTCGCTGTGTGGTTGTCCTCAACAGTTGTTGGTGCCGTCAATTCAGTGCCACTTGTAAGTGATTTCTTTGGAAACCACATACCTCTAGGTTTTTCAGTTTTGAA------------AT
C-  ******************  *********** **** ******* * ************************** *******   *******  * ******* **** **  ** **            **

        140       150       160       170       180       190       200       210       220       230       240       250       260
                                                              o     o                                                            ]
1-  GGAAA-TGATGAATCTATATTTCAATTTACAACTTTGTGTGTGTGTGTGTGTGTGTGTGAGTGTGAGTGTGTGTTTGCATTATCTCCTGTTTAGAAATTAATT-AACTACTCTGAATTGACTTTCAGCT
2-  GGAAA-TGATGAATCTATATTTCAATTTACAACTTTGTGCGTGTGTGTGTGTGTGTGTGT----GAGTGTGTGTTTGCATTATCTCCTGTTTAGAAATTAATT-AACTACTCTGAATTGACTTTCAGCT
3-  GGAAA-TGATGAATCTATATTTCAATTTACAACTTTGTGTGTGCGTGCGTGTATCTCTGTGT----------GTGTTTGCATTATCTCCTGTTTAGAAATTAATT-AACTACTCTGAATTGACTTTCAGCT
4-  GGAAA-TGATGAATCTATATTTCAATTTACAACTTTGTGCGTGTGTGTGTGT-----------------------TTGCATTATCTCCTGTTTAGAAATTAATTGAACTACTCTGAATTGACTTTCAGCT
5-  GGAAA-TGATGAATCTATATTTCAATTTACAACTTTGTGCGTGTGTGTGTGT-----------------------TTGCATTATCTCCTGTTTAGAAATTAATTGAACTACTCTGAATTGACTTTCAGCT
6-  GGAAA-TGATGAATCTATATTTCAATTTACAACTTTGTATGTGTGTGTGTGTGTGTGT----------------TCGCATTATCTCCTGTTTAGAAATTAATT-AACTACTCTGAATTGACTTTCAGCT
7-  GGAAA-TGATGAATCTATATTTCAATTTACAACTTTGTGTGTGTGTGTGT-------------------------TCGCATTATCTCCTGTTTAGAAATTAATT-AACTACTCTGAATTGACTTTCAGCT
8-  GGAAAGTGATGAATCTATTTTTCAAGTTACAACTTTGTCCGTGT-CGT----------------------------TGAATTATGTCCTCTTTAGAAATTAATTGAACTACTCTGAATTGACTTTCAGCT
9-  GGAGAATGATGAATCTATTTTTCAAC-TACAACTTTGTGCGTGT-CAT----------------------------TGAATTACCTCCTCTTTGGAAATTAATTGAACTACTCTGAATTGACTTTCAGCT
10- GGAAAGTGATGAATCTATTTTTCAAGTTACAACTTTGTCCGTGT-CGT----------------------------TGAATTATGTCCTCTTTAGAAATTAATTGAACTAATCTGAATTGACTTTCAGCT
11- GGAAAATGATGAATCTATTTTTCAAC-TACAACTTTGTGCGTGT-CAT----------------------------TGAATTACCTCCTCTTTGGAAATTAATTGAACTACTCTGAATTGACTTTCAGCT
12- GGAAAGTGATGAATCTATTTTTCAAGTTACAACTTTGTCCGTGT-CGT----------------------------TGATTTATCTCCTCTTTAGAAATTAATTGAACTACTCTGAATTGACTTTCAGCT
13- GGGAAATGATGAATCTATTTTTCAACTTACAACTTTGTGCGTGT-CAT----------------------------TGA------------------ATTAATTGAACTACTCTGAATTGACTTTCAGCT
14- GGAAAGTGATGAATCTATTTTTCAAGTTACAACTTTGTCCGTGT-CGT----------------------------TGATTTATCTCCTCTTTAGAAATTAATTGAACTACTCTGAATTGACTTTCAGCT
C-  **  * ************ ******   ***********  ***                                *                  *******  ***** *******************
                                                                                                        Primer PacB22 revers
```

Fig. 2 Nucleotide comparison of PacB22 EST SSR locus in species of the *Rosaceae* family. Different alleles in the same accessions are indicated by the name of the accession followed by A or B. *1 P. armeniaca* L. SEO (260 bp); *2 Prunus davidiana* #1908 (256 bp); *3 P. persica* GF305 (250 bp); *4 P. avium* Burlat (239 bp); *5 P. cerasus* V2327 (239 bp); *6 Prunus japonica* × *Prunus spinosa* cv Jaspi A (244 bp); *7 P. japonica* × *P. spinosa* cv Jaspi B (234 bp); *8 Malus* × *domestica* cv discovery fragment A (220 bp); *9 Malus* × domestica cv discovery fragment B (227 bp); *10 Malus* × *domestica* cv MM105 A (220 bp); *11 Malus* × *domestica* cv MM105 B (227 bp); *12 P. communis* cv Pine seedlings A (220 bp); *13 P. communis* cv Pine seedlings B (213 bp); *14 P. veitchii* (220 bp); *C* consensus sequence. The primary tandem repeat is *underlined* while gaps are represented by *broken lines*. *Bold characters* indicate base substitutions in between *Prunoideae* and *Maloideae*, as well as nucleotide similarities across *Maloideae* alleles. *Dots* above the sequences show interruption in the apricot microsatellite repeat and *uppercase letters* above the first 70 bp correspond to the partial 5′ coding region of B22 (accession number BQ134643). Forward and reverse PacB22 primers are displayed by *arrows*

covery × TN10.8 (data not shown). Twelve individuals were tested along with the parents Discovery and TN10.8 following the above PCR conditions. All 12 progeny and two parents displayed two distinct and equally sized fragments. Therefore, the PacB22 locus in the *Maloideae* does not follow a single-locus segregation.

Another EST SSR locus, PacA10, was sequenced across the *Rosaceae* species and genera (data not shown) but the sequence in *Malus* did not correspond to the initial apricot EST clone. While the PacA10 locus in apricot displayed homology with a chlorophyll a/b binding protein (Table 1), the PCR product amplified with the PacA10 primer pair in *Malus* × *domestica* showed similarity with a gene coding for a nonsense-mediated mRNA decay transacting factor in *Arabidopsis thaliana* (accession number AB017068, id.= 4e-08).

Discussion

In this report, we show that SSRs are retrieved from expressed sequences of grape and apricot species, after little input and without extensive sequencing data. cDNA-SSR markers combine the advantages of microsatellite variability with the information content potentially carried by expressed sequences. This is an efficient and economical technique for accumulating and mapping cDNA sequences from different tissues and developmental stages and, by this means, increasing the density of gene markers on linkage maps of minor crops.

A putative function was assigned by sequence similarities for half of the grape and apricot cDNA-SSR loci; this is in concordance with systematic sequencing and functional annotation of genes in other species where 52% of the sequences were coding for unidentified proteins (Aubourg and Rouzé 2001). Some of them are good candidates for QTL mapping and for the assessment of phenotypic and adaptive variation, e.g. potassium import with the SKOR gene and its relation with fruit quality in grape.

Grape and apricot EST SSR primer pairs were tested for cross-species transferability in the family *Vitaceae* and *Rosaceae* respectively. Utility of microsatellite markers between related and less-related species was evaluated not only by the presence or absence of amplification products but also on heterozygous banding patterns and sequence variation in the flanking regions.

In the *Vitaceae*, high levels of length polymorphism across species and genera, and allele sequence data, confirmed the microsatellite nature of observed variations and the optimal utility of the EST SSR markers. In the *Rosaceae*, optimal utility of the apricot EST SSR markers was for closely related species belonging to the same subgenus *Prunophora*. Outside this subgenus, heterozygous banding patterns were rarely attained, except in the tetraploid *P. cerasus*. Specificity of SSRs for a given taxon can limit the cross-species utility of these markers in other cultivated species or in interspecific crosses. Many authors reported a decline in amplification success with increasing divergence and evolutionary distance between taxa (for example see Whitton et al. 1997). The more-closely related the species are, the better is the amplification and the higher is the allelic diversity. This was confirmed in both the *Vitaceae* and *Rosaceae* families. The threshold distance after which no amplification can be expected is shorter in the *Rosaceae* than in the *Vitaceae*. This may reflect large genetic distances among *Rosaceae* taxa and that speciation in the *Vitaceae* took place rather recently. According to Rossetto (2001), cross transferability within the same genus reaches 76.4%, with 86% displaying polymorphism. Across the *Vitaceae* genera, Arnold et al. (2002) suggested that cDNA microsatellites were transferred more readily than anonymous SSR markers. In the *Rosaceae*, such a level of transferability is apparently not the case; polymorphism is particularly low in the *Amygdalus* subgenus and in other non-source subgenera. More extensive data will be required in the *Rosaceae* to fully compare anonymous versus transcribed SSR markers.

Our results are consistent with studies performed in the *Vitaceae* using anonymous SSR markers (Di Gaspero et al. 2000), but in contrast with previous studies in the *Rosaceae* (Sosinski et al. 2000; Dirlewanger et al. 2002). However, in the two last reports, only amplification and detection of appropriate-sized fragments on Metaphor agarose gels were recorded; polymorphism and verification of true positives by sequencing were not considered. We believe that it is not possible to predict marker transferability into a given species on the basis of its taxonomic classification; it will mainly depend on the evolutionary histories of the respective taxonomic families.

In animals, the potential of interspecific amplification is high (for example see FitzSimmons et al. 1995). In plants, cross-species transferability has been particularly successful in several crop species and forest trees (Westman and Kresovich 1998; Karhu et al. 2000). Conservation of a microsatellite repeat at the waxy locus has been demonstrated in non-related species such as rice, barley and potato (Becker and Heun 1995; Ayres et al. 1997; Milbourne et al. 1998). While in some cases this locus is monomorphic in length (Washington et al. 2000; Domon et al. 2002), such conservation across plant taxa raises the question of the biological role for this microsatellite repeat.

It is important to note that we used unchanged PCR conditions to test the transferability of EST SSRs across taxa. Modifiying the PCR protocol by lowering of annealing temperature may increase transferability. However, false positives can appear and the fragments amplified may not be those anticipated, for example PacA10 in *Malus* × *domestica*. Such alleles will be incorrectly scored at an expected EST locus and are a major drawback in genetic comparison and marker-assisted localisation of expressed genes. In many surveys of SSR amplification across plant species and genera, products are selected as useful markers if they amplify consistently and are similar to SSRs in the source taxon in terms of size and polymorphism. This highlights the fact that results are easily misinterpreted when transferring SSR markers across taxa.

Overall, the sequence alignments showed that the microsatellite repeat in the *Vitaceae* was variable but present in most cases. This study is clearly limited by the number of taxa used and by the number of alleles sequenced. However, other authors recently reported the characterisation of microsatellite flanking regions in other species of the *V itaceae* family (Di Gaspero et al. 2000; Rossetto et al. 2002). Interestingly, Rossetto et al. (2002) evaluated the potential of EST SSR flanking regions for determining taxonomic relationships within two *Vitaceae* genera, *Cissus* and *Cayratia*. In our case, the purpose was not to perform a phylogenetic study but to compare the potential of information provided by the tandem repeat, as well as the flanking region, in comparative genetics. Clearly, in the *Vitaceae*, microsatellite and flanking regions are highly conserved. Similar results were reported by Di Gaspero et al. (2000) who showed that there was no difference at the sequence level between European, American and Asian species. However, interruptions in the microsatellite repeat occur in another subgenus, *Muscadinia*, and enables separation of the two *Vitis* subgenera, but this needs to be confirmed by more extensive sequencing. More information was obtained by analysis of the flanking regions and the occurrence of insertion/deletion events. We showed sequence similarities, especially a 2-bp deletion event, between two geographically distinct species, *V. mexicana* (American) and *V. davidii* (Asian). Along with the former study of Di Gaspero et al. (2000), this clearly demonstrates that classification in the genus *Vitis* based on the geographical distribution of species may not be reflecting the true phylogenetic relationship.

In view of the low success of cross-genera amplification and size polymorphism in the *Rosaceae* family, information based on flanking-region sequences makes those EST markers a potentially more useful tool for the study of genome variation and evolution between genera. It is well documented that within a microsatellite locus, length differences between alleles are not only due to a variation in the number of tandem repeat units (Grimaldi and Crouau-Roy 1997; Orti et al. 1997). In the case of the PacB22 EST SSR locus, no variability was found among different species of the *Prunoideae* subfamily except for microsatellite length shortening and a few base substitutions in the coding region. However, when amplifying across genera in the *Rosaceae*, base substitutions and deletion events in the flanking region are more informative than the core motif repeat itself. In fact, SSR length variation as normally scored in microsatellite assays is inadequate to assess long-term evolutionary divergence between the *Maloideae* and the *Prunoideae*, and in general between relatively distinct taxa. Effectively, we can observe that in distinct species such as *Malus* × *domestica* and *Pyrus communis*, the microsatellite repeat is no longer present. Moreover, allelic similarities between *Malus* and *Pyrus* cultivars lead us to propose that those alleles do not belong to the same locus but instead arose either from gene duplication or allopolyploidisation. Segregation analysis in *Malus* × *domestica* showed two monomorphic distinct loci. After sequencing, each allelic size class was compared between *Malus* and *Pyrus* accessions; one single deletion showed relatedness between one and the other fragment across the subgenera *Malus* and *Pyrus*. Previous study in the subfamily *Maloideae* suggested that the *Maloideae* genome ($x = 17$) was derived from allopolyploidy between primitive members of the subfamilies *Prunoideae* ($x = 8$) and *Spiroideae* ($x = 9$) (Stebbins 1958). Additional loci have frequently been observed for example in wheat, where 29% of primer pairs amplified more than one locus (Bryan et al. 1999), and in apple, 25% (Guilford et al. 1997). This is probably reflecting chromosomal duplication or allopolyploidy in those species. In our case, one distinct locus additional to PacB22 was amplified as confirmed by segregation analysis.

Therefore, the variability recorded within the flanking sequence of the PacB22 locus is sufficient to distinguish between two loci in the same genome, either duplicated or located on two homoeologous chromosomes. This could be particularly useful for differentiating homoeologous chromosomes in allopolyploid species. It also shows that EST-derived markers (RFLP, SNP, SSCP) would represent markers of choice, compared to microsatellite markers, for genome comparisons across genera in the *Rosaceae*.

Acknowledgements We thank Drs. C.-E. Durel and M. Tavaud for providing specimens or DNA of *Malus* × *domestica* and *P. avium* respectively, as well as Dr. F. Laigret for performing some of the sequencing reactions. The *P. communis* and *P. veitchii* accessions were kindly provided by Dr. Davies from the HRI, East Malling (UK), and maintained in in vitro culture thanks to M. Lansac and J.-P. Eyquard. We acknowledge excellent care of the plants by A. Bonnet, C. Hevin, B. Douens, J.-P. Petit and J.-P. Robert. The authors thank Drs. G. Moran and P. Butcher (CSIRO Forestry and Forest Products, Australia) for critically reading the manuscript. Our study was supported by grants from the Aquitaine region (CCRDT20000307010 and CCRDT20010307007) and from the CIVB (Comité Interprofessionnel des Vins de Bordeaux #23611/23612).

References

Arnold C, Rossetto M, McNally J, Henry RJ (2002) The application of SSRs characterized for grape (Vitis vinifera) to conservation studies in Vitaceae. Am J Bot 89:22–28

Aubourg S, Rouzé P (2001) Genome annotation. Plant Physiol Biochem 39:181–193

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germ plasm. Theor Appl Genet 94:773–781

Becker J, Heun M (1995) Barley microsatellites: allele variation and mapping. Plant Mol Biol 27:835–845

Bryan GJ, Stephenson P, Collins A, Kirby J, Smith JB, Gale MD (1999) Low levels of DNA sequence variation among adapted genotypes of hexaploid wheat. Theor Appl Genet 99:192–198

Butcher PA, Decroocq S, Gray Y, Moran GF (2000) Development, inheritance and cross-species amplification of microsatellite markers from Acacia mangium. Theor Appl Genet 101:1282–1290

Cipriani G, Lot G, Huang WG, Marrazzo MT, Peterlunger E, Testolin R (1999) AC/GT and AG/CT microsatellite repeats in peach [Prunus persica (L) Batsch]: isolation, characterization and cross-species amplification in Prunus. Theor Appl Genet 99:65–72

Chalhoub BA, Thibault S, Laucou V, Rameau C, Höfte H, Cousin R (1997) Silver staining and recovery of AFLP amplification products on large denaturing polyacrylamide gels. BioTechniques 22:216–220

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from Pine trees. Plant Mol Biol Rep 11:113–116

Di Gaspero G, Peterlunger E, Testolin R, Edwards KJ, Cipriani G (2000) Conservation of microsatellite loci within the genus Vitis. Theor Appl Genet 101:301–308

Dirlewanger E, Cosson P, Tavaud M, Aranzana MJ, Poizat C, Zanetto A, Arús P, Laigret F (2002) Development of microsatellite markers in peach [Prunus persica (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (Prunus avium L.). Theor Appl Genet 105:127–138

Domon E, Fuijita M, Ishikawa N (2002) The insertion/deletion polymorphisms in the waxy gene of barley genetic resources from East Asia. Theor Appl Genet 104:132–138

Downey SL, Iezzoni AF (2000) Polymorphic DNA Markers in Black Cherry (Prunus serotina) are identified using sequences from Sweet Cherry, Peach and Sour Cherry. J Am Soc Hort Sci 125:76–80

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques 20:758–760

Erpelding JE, Blake NK, Blake TK, Talbert LE (1996) Transfer of sequence tagged site PCR markers between wheat and barley. Genome 39:802–810

FitzSimmons NN, Moritz C, Moore SS (1995) Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. Mol Biol Evol 12:432–440

Galet P (1967) Recherches sur les méthodes d'identification et de classification des Vitacées des zones tempérées. PhD dissertation, Université de Montpellier, Faculté des Sciences, Montpellier, France

Grimaldi MC, Crouau-Roy B (1997) Microsatellite allelic homoplasy due to variable flanking sequences. J Mol Evol 44:336–340

Guilford P, Prakash S, Zhu JM, Rikkerink E, Gardiner S, Bassett H, Forster R (1997) Microsatellites in Malus × domestica (apple): abundance, polymorphism and cultivar identification. Theor Appl Genet 94:249–254

Karhu A, Dieterich JH, Savolainen O (2000) Rapid expansion of microsatellite sequences in pines. Mol Biol Evol 17:259–265

Lefort F, Douglas GC (1999) An efficient micro-method of DNA isolation from mature leaves of four hardwood tree species Acer, Fraxinus, Prunus and Quercus. Ann For Sci 56:259–263

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

Orti G, Pearse DE, Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. Proc Natl Acad Sci USA 94:10,745–10,749

Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross species amplification of Soybean (Glycine max) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. Mol Biol Evol 15:1275–1287

Powell W, Gordon CM, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Rehder A (1949) Bibliography of cultivated trees and shrubs hardy in the cooler temperate regions of the Northern hemisphere. The Arnold Arboretum of Harvard University, Jamaica Plain, Massachussets, USA

Rossetto M (2001) Sourcing of SSR markers from related plant species. In: Henry R (ed) Plant genotyping – the DNA fingerprinting of plants. CAB International, Wallingford, UK, pp 211–224

Rossetto M, McNally J, Henry RJ (2002) Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae. Theor Appl Genet 104:61–66

Sosinski B, Gannavarapu M, Hager LD, Beck LE, King GJ, Ryder CD, Rajapakse S, Baird WV, Ballard RE, Abbott AG (2000) Characterisation of microsatellite markers in peach [Prunus persica (L) Batsch]. Theor Appl Genet 101:421–428

Stebbins GL (1958) On the hybrid origins of the Angiosperms. Evol 12:267–270

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

Washington JM, Box A, Karakousis A, Barr AR (2000) Developing waxy barley cultivars for food, feed and malt. In: Logue S (ed) Barley genetics VIII. Adelaide University, Adelaide, 3:303–305

Westman AL, Kresovich S (1998) The potential for cross-taxa simple-sequence repeat (SSR) amplification between Arabidopsis thaliana L. and crop brassicas. Theor Appl Genet 96:272–281

Whitton J, Rieseberg LH, Ungerer MC (1997) Microsatellite loci are not conserved across the Asteraceae. Theor Appl Genet 101:1282–1290

Yamamoto T, Kimura T, Sawamura Y, Kotobuki K, Ban Y, Hayashi T, Matsuta N (2001) SSRs isolated from apple can identify polymorphism and genetic diversity in pear. Theor Appl Genet 102:865–870