G. D. Esselink · M. J. M. Smulders · B. Vosman

Identification of cut rose (*Rosa hybrida*) and rootstock varieties using robust sequence tagged microsatellite site markers

Abstract In this study a DNA fingerprinting protocol was developed for the identification of rose varieties based on the variability of microsatellites. Microsatellites were isolated from *Rosa hybrida* L. using enriched small insert libraries. In total 24 polymorphic sequenced tagged microsatellite site (STMS) markers with easily scorable allele profiles, from six different linkage groups, were used to characterize 46 Hybrid Tea varieties and 30 rootstock varieties belonging to different species (*Rosa canina* L., *Rosa indica* Thory., *Rosa chinensis* Jacq., *Rosa rubiginosa* L., and *Rosa rubrifolia glauca* Pour.). Clones and known flower color mutants were identified as being identical, all other varieties were differentiated by a unique pattern with as few as three STMS markers. The high discriminating power of the loci suggests that a selection of the most-robust STMS markers may be able to differentiate any two varieties within rootstocks or Hybrid Teas except for mutants. The selected STMS markers will be useful as a tool for reference collection management, for assessing essential derivation of varieties and illegal propagation.

Keywords *Rosa* · STMS · STS · Molecular database · Allelic phenotype

Introduction

Hybrid Tea varieties of roses (*Rosa hybrida* L.) are among the most economically important cut-flower plants. The first Hybrid Tea rose was introduced in 1867, and since then more than 10,000 varieties have been released. Currently, at the Centre for Variety Research, the Netherlands, about 2,800 predominantly Hybrid Tea

G.D. Esselink · M.J.M. Smulders · B. Vosman (✉)
Department of Biodiversity and Identity,
Plant Research International, PO Box 16, 6700 AA Wageningen,
The Netherlands
e-mail: b.vosman@plant.wag-ur.nl

varieties have been submitted for Plant Breeders Rights and this number increases with, on average, 80 applications each year.

For modern rose varieties the registration and protection is based on morphological and physiological characteristics as described in the UPOV (Union Internationale pour la Protection des Obtentions Végétales) guidelines. These classical methods become less and less efficient as the number of varieties to be tested increases and the genetic distances between varieties becomes smaller. Side by side comparison of applicants with older varieties is difficult or even impossible as many varieties are not easily available, and maintaining large reference collections is far too expensive. Especially the identification of varieties of rootstocks based on morphological characteristics is difficult since there are only small differences between the varieties within species.

To fulfill the need for more-efficient, accurate and fast identification tools for rose, the application of several molecular marker systems has been investigated. The application of RFLPs (Rajapakse et al. 1992; Ballard et al. 1995), fingerprint analysis with mini- and micro-satellite probes (Ben-Meir and Vainstein 1994; Vainstein and Ben-Meir 1994) and RAPD (Torres et al. 1993; Cubero et al. 1996; Martin et al. 2001), revealed polymorphisms between a limited number (5–11) of varieties of *R. hybrida*. These marker systems have some major drawbacks for variety identification, i.e. they lack high levels of polymorphisms, are difficult to reproduce, are laborious or provide complex patterns inconvenient for database building (Vosman 1998). In contrast, microsatellites are highly polymorphic and have the advantage of providing a codominant marker system based on a PCR technology. As sequenced tagged microsatellite site (STMS) markers they provide simple banding patterns, especially suitable for automated and objective analysis and data can be stored easily in a database. New varieties or new markers can be easily added to an existing database. This STMS approach was successfully demonstrated in grapevine (Botta et al. 1995), soybean (Song et al. 1999), and in a collaborative study for the construc-

tion of databases for tomato (Vosman et al. 2001; Bredemeijer et al. 2002) and wheat (Röder et al. 2002).

In this paper we describe the development and characterisation of STMS markers from *R. hybrida* as the basis for a robust and reliable identification system for Hybrid Tea varieties. The usefullness of the markers was also tested on rootstock varieties.

Hybrid Tea varieties are the result of crosses within a very limited pool of genotypes and therefore a low degree of genetic variation may be expected (DeVries and Dubois 1996). However, based on RAPD analysis this lack of genetic variation within the modern varieties was doubted (Debener et al. 1996). Therefore, the microsatellite markers were also used to estimate the genetic variation and genetic relationships within and between Hybrid Tea and rootstock varieties.

Material and methods

Plant material

For the isolation of microsatellites, genomic DNA of *R. hybrida* variety Sonia was used. For building the STMS database 46 Hybrid Tea varieties (*R. hybrida* L.) were used, as well as 27 rootstock varieties belonging to *Rosa canina* L. (14 varieties), *Rosa indica* Thory. (8), *Rosa chinensis* Jacq.(1), *Rosa rubiginosa* L.(1), *Rosa rubrifolia glauca* Pour.(1) and *R. hybrida* L.(2). Both groups contained known duplicates or mutants from initial varieties. For estimation of diversity in a group of varieties belonging to different species, three hiprose (*Rosa carolina* L.) varieties were also included into the "rootstock" group. Young leaves of a single individual were harvested, immediately frozen in liquid nitrogen and stored at –80 °C until use.

DNA extraction

For the construction of a genomic library enriched for microsatellites, nuclear DNA of high quality was extracted from leaves of variety Sonia according to Vosman et al. (1992). For STMS analysis, DNA of rootstock and Hybrid Tea varieties was extracted from freeze-dried leaves using the Qiagen DNeasy Plant Mini Kit (Westburg, The Netherlands). Typically a yield of 30 μg of DNA was obtained per 20 mg of dried weight.

Microsatellite isolation

Microsatellites were isolated from enriched small-insert genomic libraries essentially as described by Karagyozov et al. (1993) and modified by Van de Wiel et al. (1999). This method has been applied successfully to a number of plant (Van der Schoot et al. 2000; Smulders et al. 2001) and animal (Arens et al. 2000) species. Sonicated DNA was used for enrichment of microsatellites using mixtures of synthetic oligonucleotides: $(GT)_{12}$, $(GA)_{12}$, and $(TGT)_9$ (pool1); $(TCT)_{10}$, $(TGTT)_8$, and $(GTAT)_8$ (pool 2); and $(GAG)_8$, $(GTG)_8$, and $(GATA)_8$ (pool 3). To increase the efficiency of the isolation of different tri- and tetra-nucleotides microsatellite repeats, a second enrichment procedure with DNA digested with *Alu*I, *Rsa*I or *Hae*III was performed using single tri- $(TCT)_{10}$, $(TGT)_9$, $(GAG)_8$, $(GTG)_8$, $(TGA)_9$, $(AGT)_{10}$, $(CTG)_8$, $(CGT)_8$ and tetra-nucleotide $(GATA)_8$, $(TGTT)_8$ and $(GTAT)_8$ microsatellite repeats.

STMS development and amplification

Microsatellites were amplified by PCR in a 20-μl reaction volume containing 16 ng of genomic DNA, 4 pmol of each primer, 100 μM of each dNTP, 10 mM of Tris–HCL pH 9.0, 20 mM of $(NH_4)_2SO_4$, 0.01% Tween 20, 1.5 mM of $MgCl_2$ and 0.1 Unit of Goldstar *Taq* DNA polymerase (Eurogentec, Maastricht, The Netherlands). The optimized PCR conditions used for the database construction were 94 °C for 3 min followed by 94 °C for 30 s, at the calculated annealing temperature for 30 s, 72 °C for 45 s and a final extension at 72 °C for 3 min.

Unlabeled primers were obtained from Isogen (Maarssen, The Netherlands), and fluorescently labeled [Indodicarbocyanine(Cy5)phosphoramidite] primers from Amersham/Pharmacia (Roosendaal, The Netherlands).

Detection of STMS polymorphisms

For characterization of STMS primers the amplification products were separated on a 6% acrylamide gel and visualized with silver staining according to the Promega Silver sequence DNA sequencing system (Promega, Leiden, The Netherlands) as described by Van de Wiel et al. (1999). Fluorescent amplification products were detected using an ALFexpress DNA sequencer (Amersham/Pharmacia, Roosendaal, The Netherlands) as described by Bredemeijer et al. (1998).

Creating the STMS database

A selection of STMS markers with high quality patterns was used for the creation of the database. Screening of varieties in a first round revealed all existing alleles for each marker and allowed selection of a set of varieties representing all the alleles. These varieties were included in each following run and used as a reference for allele determination. In this way alleles were assigned a name based on an exact match to the length of the corresponding allele present in the reference variety and not as a specific length in base pairs. Allele codes (A, B, etc) were transferred to a numeric database in an Excel spreadsheet for storage and further analyses.

Data analysis

Co-dominant scoring of the markers in heterozygote samples, which is necessary to assess the complete genotype of a variety, is complicated by the polyploid nature of roses. Differences were detected in the amount of product for different alleles in a particular genotype. However, it turned out to be very difficult to use these differences in a reliable way to estimate whether a particular allele was present in one, two or three copies, and thus to deduce the actual genotype of a variety. Therefore only the presence of alleles were scored and recorded as a presence/absence matrix (0/1). For example, AABC, ABBC and ABCC genotypes would all be scored as ABC. For both variety groups the number of alleles per locus, the effective number of alleles and the number of allelic phenotypes was calculated. The effective number of alleles (n_e) is estimated as the reciprocal of $\sum p_i^2$, were P_i is the frequency of the ith allele in the unique varieties examined (Hartl and Clark 1997). When a locus was analyzed as one character, we refer to this as the "allelic phenotype" of a variety (after Becher et al. 2000). On the bases of individual alleles Jaccard similarities were calculated and the varieties were clustered using the unweighted pair group method using arithmetic averages (UPGMA) module of NTSYS, version 2.1.

Results

Microsatellite isolation from the rose genome

Two microsatellite-enriched genomic libraries were constructed. The first enrichment was performed using sonificated DNA of variety Sonia (*R. hybrida*) and pools of synthetic repeats. The enrichment efficiency ranged from 4% to 31%. Sequencing of 175 inserts revealed a high number of unique sequences but also duplicates were found. In pool 1 27 out of 63 clones contained singular dinucleotide repeats, 36 were compound repeats of which the majority consisted of GT and GA repeats. Pool 2 revealed mainly clones with TCT repeats of which a high number were duplicates. Very low numbers of positive clones were found for the tetranucleotide and GAG and GTG motives of pool 2 and 3. Overall 61 (45%) of the unique sequence STMS primer pairs could be developed, 37 for di, 24 for tri and none for tetra-nucleotide repeats.

In a new round of enrichment we tried to increase the efficiency of isolating different tri- and tetra-nucleotide microsatellite repeats. For this an enrichment procedure with DNA of variety Sonia, digested separately with *Alu*I, *Rsa*I or *Hae*III, was performed using single tri- and tetra-nucleotide microsatellite repeats. The highest numbers of positive clones were obtained with GTAT and TGA (13%). Unfortunately, for GTAT only 38% of the positive clones contained unique sequences. The remaining clones showed a compound motif consisting of 7 (GT) repeat units and 24, 26, 27 or 29 (GTAT) repeat units. Overall efficiency (28%) was lower than the result of the first enrichment. However, the procedure did obtain tetranucleotide repeats at a reasonable rate, whereas the same motifs had not been found in the mixed enrichment.

Selection of STMS markers

For 84 unique microsatellite sequences, primer pairs were designed. These primer pairs were tested for pattern quality and degree of polymorphism on a set of ten varieties differing widely in morphological characteristics. PAGE/silver staining pattern quality was scored according to Smulders et al. (1997). Thirty five primer pairs produced polymorphic banding patterns with quality 1 or 2 (i.e. no or moderate stutter bands, easy to

Table 1 Characteristics of the 24 selected microsatellite markers for rose

STMS[a] marker	Repeat sequence	Linkage group (Debener et al. 2001)	Locus scorability[b]	Rootstock varieties[c]			Hybrid Tea varieties[c]		
				No. of alleles	No. of eff. alleles	No. of allelic phenotypes	No. of alleles	No. of eff. alleles	No. of allelic phenotypes
RhP518	$(CAT)_4CAATT(CAT)_6$ $CAATT(CAT)_6$	5	2	6	4.6	7	4	1.9	5
RhAB22	$(GT)_{13}(GA)_{13}$	6	1	11	5.0	13	4	2.0	5
RhBK4	$(AT)_5(GT)_{12}$	n.d.[d]	2.3	7	1.6	4	4	2.1	6
RhAB13	$(GA)_{11-1}(GT)_8$	4	2	13	6.2	9	4	2.2	4
RhE2a	$(TGT)_{33-12}$	n.d.[d]	1	5	3.4	5	4	2.2	5
RhE3	$(TGT)_{21}$	n.d.[d]	2	8	2.7	6	5	2.2	6
RhD206	$(TCT)_{14}$	2	1	17	9.2	17	4	2.4	5
RhE2b	$(TGT)_{20-6}$	n.d.[d]	1	9	6.2	13	6	2.4	11
RhI402	$(GTG)_{11-2}$	3	1	6	3.4	7	5	2.5	7
RhAB1	$(GT)_9(GA)_{10}$	n.d.[d]	1	6	3.1	8	3	2.6	7
RhB19	$(GT)_{10}$	n.d.[d]	2	8	4.7	12	3	2.8	8
RhJ404	$(GAG)_6$	4	2,3	7	3.0	8	4	3.0	7
RhD221	$(TCT)_{21-1}$	4	1	7	3.9	11	5	3.0	10
RhEO506	$(CAG)_6(CAA)_{18-7}(CAG)_6$	2	2	18	6.4	15	6	3.0	14
RhD201	$(TCT)_{33}$	n.d.[d]	1	13	7.1	19	4	3.3	10
RhP519	$(TGA)_{11-1}$	n.d.[d]	2	8	4.1	9	5	3.3	10
RhO517	$(GAC)_7$	1	1	4	3.3	8	4	3.8	10
RhAB15	$(GT)_{19-2}(GA)_{16}$	2	1	16	4.3	17	5	3.9	10
RhM405	$(TCTGAT)_5$	n.d.[d]	1	6	3.6	10	4	3.9	6
RhAB40	$(TC)_{14}(AC)_{11-1}$	4	3	16	8.3	18	8	4.1	16
RhP507	$(TGA)_{15}(TAA)_5(TGA)_5$ $(TAA)_5(TGA)_{14}$	4	1	16	8.1	17	10	4.6	17
RhB303	$(GA)_{11}$	n.d.[d]	1	9	4.7	10	6	4.7	18
RhAB26	$(GT)_{18-2}(GA)_{17}$	n.d.[d]	2,3	16	5.9	16	7	4.8	20
RhP524	$(AAT)_5(GAT)_8(GAC)_7$ $(AAT)_{9-2}(GAT)_{9-1}$	n.d.[d]	2,3	6	3.1	8	9	6.1	31

[a] Primers are available for academic research upon request to the last author
[b] Scorability 1 = unambiguously, 2 = some alleles show differences of 1–2 bp between samples and could not be scored accurately, both peaks were assigned to the same allele, 3 = contains alleles showing differential amplification lower than 20% of the main amplified allele
[c] Clones have been excluded
[d] n.d. = not determined

score). These were selected for further analysis with fluorescent primers on an automated fluorescent DNA sequencer (ALF). The same set of ten varieties was used for pattern quality assessment. Although promising in the first selection with PAGE/silver staining, seven primer pairs were discarded, since the patterns turned out to be too complex for unambiguous scoring of alleles, due to artificial bands, stutter bands from one allele overlapping the next allele or highly differential amplification of alleles. In four cases the amplification of alleles with the labeled primer failed systematically in all or in a large number of varieties. In Table 1 the STMS markers with unambiguously scorable alleles are presented that have been used for fingerprinting the set of 76 varieties. These 24 STMSs represent two perfect and seven compound repeats of dinucleotide loci, four perfect, six imperfect and four compound repeats of trinucleotide microsatellite loci, as well as one hexanucleotide locus.

Marker optimization

Screening of the test set revealed locus-specific characteristics for each STMS as summarized in Table 1. The selected markers showed low to low/moderate stutter bands. High stutter markers had been efficiently selected against in the first round of selection on PAGE/silver stained gels. The observed stutter characteristics did not interfere with scorability and allele assignment for any of these selected STMS markers. Complete absence of stutter bands was found in two trinucleotide markers and in the hexanucleotide marker only. Non-template addition of a single nucleotide by *Taq* DNA polymerase (Clark 1988) was observed for the majority of the primer pairs. The scoring problems related to this can be circumvented by using PIG-tail primers (Brownstein et al. 1996). Results (data not shown) indicate that the scorability of most of the selected rose STMS markers clearly increased with primers containing a PIG-tail.

For most of the STMS markers the allele sizes differed by one or multiple repeat units from each other, i.e. separated by steps of 2 bp for dinucleotide repeats or 3 bp for trinucleotide repeats. Sometimes this relation was not so obvious for all alleles, suggesting that other types of sequence variation may also be involved in allele diversity as well. This was not further investigated. Whenever allele allocation was complicated by this, the scorability was marked as 2 (example in Fig. 1, panel C)

We observed the occurrence of differences in peak intensity between the alleles in heterozygous varieties. This is expected in a polyploid species with a maximum ratio of 1:3 in tetraploid (Hybrid Teas) and of 1:4 for most of the (pentaploid) rootstock varieties. Although differences existed among loci, for most loci the alleles could be accurately identified since the relative peak area of the least intense peak was well over 20% of the most intense peak. However, in three cases (RhBK4, RhAB26, RhAB40) the peak areas of the least amplified peaks were between 15 and 20%, and in two cases
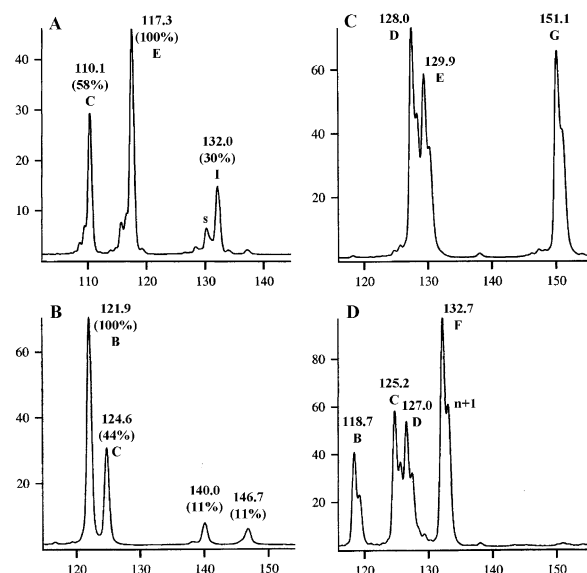

Fig. 1A–D Electropherograms of three rose STMS markers differing in scorability. **A** Marker RhB303, which is unambiguously scorable (scorability 1). **B** RhJ404 showing differential allele amplification, with the least amplifying peak lower than 20% of the main peak area (scorability 3). **C–D** RhP518 showing a 1-bp difference in peaks between two varieties but assigned the same allele code D (scorability 2). Each panel shows the relative fluorescence intensity (Y-axis) and the size of the fragments in base pairs (X-axis). Each peak that was recognized as an allele was assigned a *letter value*. The relative peak area (%) is shown between brackets; baselines of peaks are not shown, s = stutter peak, n+1 = fragment with non-template added nucleotide by *Taq* polymerase

(RhJ404, RhP524), about 10% of the main peak. Scoring of these alleles (scorability 3) was only reliable under optimal conditions, i.e. the maximum fluorescence signal and optimal separation (example in Fig. 1, panel B).

Database construction

For the construction of the database, 30 rootstock varieties, representing various species, and 46 Hybrid Tea varieties were analyzed in duplicate with these 24 markers. After the first analysis a set of varieties representing all known alleles were included during the following electrophoretic runs and were used for allele designation. Although the use of a sophisticated DNA sequencer in combination with internal sizers allowed accurate fragment sizing, differences in fragment sizes of the same alleles were observed between different electrophoretic runs (differences increased up to 2 bp when gels were reloaded). The advantage of using an allelic ladder is that: (1) genotyping of varieties becomes independent of experimental variations in size determination, (2) allele assignment is based on fragments with comparable sequence, and (3) new alleles can be readily identified
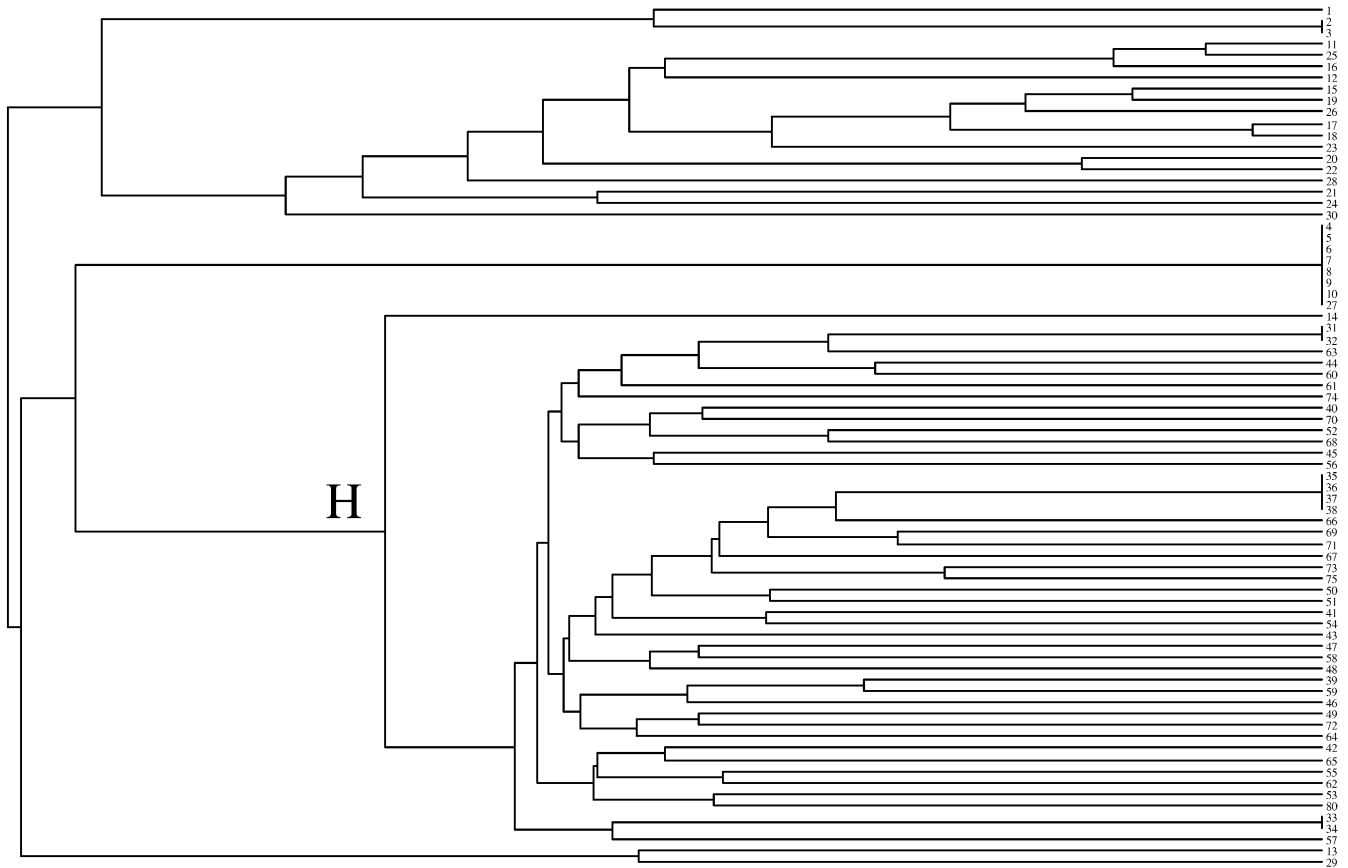
Fig. 2 Cluster analysis (UPGMA) of rose varieties. Numbers 1–3 are hip roses, 4–30 are rootstocks, 31–80 are Hybrid Tea varieties. *H* indicates the branch for Hybrid Teas

(Sprecher et al. 1996). In addition, STMS profiles generated by different equipment and/or laboratories can be accurately compared (Vosman et al. 2001). For a given microsatellite locus between two and nine reference varieties were necessary to produce all alleles for the ladder. Each STMS marker amplified five to 18 different alleles. A total of 260 different alleles were amplified of which 137 were unique for rootstock and 22 for Hybrid Tea varieties (Table 1). On average ten alleles per locus were detected among the varieties of the rootstock group compared to five alleles per locus among the 46 varieties of Hybrid Tea. In total 11 varieties failed in amplifying alleles for a specific STMS and were classified as having null alleles.

Variety identification

The 24 markers unequivocally identified all varieties with a unique allelic phenotype, except for one group of eight, one group of four and three groups of two varieties. In all these cases the varieties within these groups were known duplicates or mutants (sports) of initial varieties. A unique genotype for all rootstock varieties was

already obtained by fingerprinting with as few as two microsatellites (e.g. RhAB40 + RhEO506 or RhP507 + RhD201). For Hybrid Tea varieties all except two varieties were uniquely identified by RhEO506 + RhD201. Addition of the data, for example of RhP517, differentiated the remaining two varieties. Several other selected combinations of three STMS loci were capable of identifying all the varieties tested.

Figure 2 shows the UPGMA tree obtained with the data on all occuring alleles at 24 STMS marker loci. As expected, the rootstocks are clearly seperated from the Hybrid Tea varieties (separation marked by "H"). As the dendrogram is constructed on the basis of tetra- and penta-ploid roses it can only be used to a very limited extent. Separate clusters can be identified for *R. carolina* (1–3), *R. canina* (11,12,15–20,22–26,29) and *R. indica* (4–10 and 27), although most clusters contain some varieties of other species, suggesting close relatedness.

In the dendrogram the identical genotypes mentioned before can readily be identified. All other varieties were well-differentiated. For example, the most related varieties within the Hybrid Tea varieties, no. 69 and no. 71, differed at 15 alleles (at 12 loci). Within the *R. canina*, varieties no. 11 and no. 25 differed on seven alleles (at four loci) and no. 17 and no. 18 on three alleles (at two loci).

In order to quantify the discriminating power of the STMS markers, the number of effective alleles (n_e) for each marker was calculated (Hartl and Clark 1997) and the number of allelic phenotypes was counted for each

282

Table 2 Total numbers of different alleles and allelic phenotypes generated by 24 selected STMS markers

Item	No. of different alleles	No. of different allelic phenotypes
Rootstocks (n=22)	238	266
of which *R canina* (n=14)	155	141
Hybrid Tea (n=41)	123	248
Total	260	463
Average/marker	11	19

group of varieties (Table 1). The information brought forth by the primer pairs was different for each variety group. Locus RhP524 is very discriminating for Hybrid Teas with 31 allelic phenotypes among the 41 unique genotypes and with the highest number of effective alleles, but is among the least informative loci within the rootstocks. The reverse situation was observed for locus RhAB13.

Diversity in Hybrid Tea and rootstock varieties

The total number of amplified alleles was twice as high for the rootstock varieties as for the Hybrid Teas even though the number of varieties was only half as many (Table 2). On the other hand, the number of allelic phenotypes observed for both variety groups were comparable. The number of varieties for most of the rootstock species was too small to determine whether there were larger differences among loci for these species. A more-detailed analysis was possible for the *R. canina* group (Table 2). This group of 14 varieties contained more different alleles than in the 46 Hybrid Tea varieties. The difference in allele richness can only partly be explained by the difference in ploidy level (*R. canina* is pentaploid, Hybrid Teas are tetraploid). The number of allelic phenotypes of this group is 141, which is lower than for the other rootstock varieties. A random set of 14 Hybrid Tea varieties showed a comparable 151 allelic phenotypes. Therefore, the Hybrid Tea varieties have many more allelic phenotypes than the rootstocks in proportion to the number of alleles.

Discussion

Enrichment efficiency

To our knowledge this is the first report on the isolation and characterization of microsatellite loci for rose. In this study we used a microsatellite-enrichment protocol (Van de Wiel et al. 1999) based on hybridization of genomic DNA fragments to filters containing synthetic oligonucleotide repeats. The highest number of isolated microsatellites were dinucleotides, mostly perfect GT or GT/GA compound repeats. In comparable isolation pro-

tocols for several species (Edwards et al. 1996; Van de Wiel et al. 1999; Cordeiro et al. 2000; Rae et al. 2000; Van der Schoot et al. 2000), more GT containing clones were obtained than GA containing clones, but the reverse has also been observed (Lagercrantz et al. 1993; Morgante and Olivieri 1993; Guilford et al. 1997; Taramino et al. 1997; Sosinski et al. 2000). This discrepancy could be the result of variation of the genome structure between species or the result of the selectivity of the isolation procedure (Van de Wiel et al. 1999). The overall frequency of isolated tri- and tetra-nucleotide containing clones was high, but this was almost exclusively due to the isolation of TCT repeats. To obtain STMS markers for a broader range of microsatellite motifs, especially tri- and tetra-nucleotides, the isolation procedure was modified by using filters with single motifs and by hybridizing them with a combined genomic DNA pool after separate digestion with three different restriction enzymes. A comparison of multiplex and single oligonucleotide enrichment procedures shows that screening for separate motives increased the number for TGT, GTG and GTAT repeats. Amongst the trinucleotides, the highest frequency of 12% was represented by a TGA motif, which was not present in the mixed oligonucleotide isolation method. Although the number of clones sequenced was limited, this may nevertheless be a real representation of the rose genome since only one duplicate clone was found. The TGA motif was used in a few other species. In *Arabidopsis thaliana* (Depeiges et al. 1995), sugarcane (Cordeiro et al. 2000) and black poplar (Van der Schoot et al. 2000) the frequencies of isolated clones were all below 3%. Also the frequency of GTAT clones was relatively high, but here the efficiency of obtaining unique STMS markers was very low since the majority of the clones contained the same microsatellite repeat. This could be the result of selective amplification during the PCR steps, but could also reflect the occurrence of the GTAT repeat within a repeat family as described for TGT in lettuce (Van de Wiel et al. 1999) and for GA in sugarbeet (Mörchen et al. 1996).

STMS characteristics and scoring in polyploids

Markers must fulfill several criteria to be useful for variety identification and subsequent database building. First, only STMS markers that demonstrate a high degree of polymorphism between varieties should be selected. Second, the amplified products must be easily distinguishable from each other, in particular when genotyping of varieties is being automated. Third, the markers should be evenly distributed over the genome. To achieve this in rose we applied a strict selection of markers. A first screening, based on silver-stained profiles using a small subset of varieties of different genotypes, allowed efficient selection of polymorphic STMS markers of quality 1 or 2 according to criteria developed by Smulders et al. (1997). A second selection of markers was achieved using electrophoretic patterns on the ALF

automatic sequencer. These patterns showed clearly several characteristics of the primer pairs, which are known to affect the discrete generation of STMS, markers, i.e. stutter bands and differential amplification of alleles (Bredemeijer et al. 1998). Stutter characteristics are a problem for genotyping. Although stutter patterns can be reduced and sometimes eliminated by optimizing PCR conditions (Schwengel et al. 1994) or a two- and three-stage linear amplification technique (Odelberg and White 1993), these solutions are not very practical in high-throughput genotyping. Also mathematical models have been developed for allele assignments in stutter patterns (Perlin et al. 1995; Miller and Yuan 1997). However, it is our experience that a most-straightforward approach is to select only those microsatellites that have stutter characteristics that do not disturb allele assignment. In general stutter bands are less of a problem in tri- and tetra-nucleotide repeats, which was confirmed in our study, but dinucleotide repeats with acceptable stutter characteristics were also identified and selected.

The scorability of a STMS marker is also determined by the differences in peak area between alleles in a heterozygous variety. Considering a tetraploid variety for which two alleles are detected, at the most a 3:1 ratio is expected in an amplified product. In that case the peak area of the least-amplified allele will be 30% of the main allele. Unfortunately, some microsatellites show preferential amplification of one allele over the other in heterozygous samples. This was suggested to be the result of point mutation in the primer-binding site (Lavi et al. 1994; Devos et al. 1995) or in the sequence downstream from a primer (Liu et al. 1997), or an effect of the repeat structure itself. The differential amplification may make objective genotyping difficult. This was more pronounced for dinucleotide repeats, as 33% (3 out of 9) of the selected dinucleotide repeat markers showed scoring difficulties compared to only 15% (2 out of 14) of the trinucleotide repeats. Some dinucleotide repeats (RhAB26 and RhAB40) showed differential amplification for alleles with the greatest length, suggesting that differential amplification at these loci is related to the repeat length. Based on the electropherograms of 76 varieties, the peak area of the least-amplified allele was at least 20% of the main allele in 22 of the 24 selected markers. Using this value as a threshold for allele assignment, the present/absent allelic phenotypes could easily be established.

Finally, an ideal set of markers used for variety identification is evenly distributed over the genome. A number of the developed STMS markers described here were mapped on a diploid *Rosa multiflora* Thunb. map (Debener et al. 2001). Six out of seven linkage groups, present in the rose genome, were covered by one or more microsatellites (see also Table 1). Unfortunately, half of the STMS markers could not be mapped in this cross.

Identification of rootstock varieties

All selected STMS markers, which were derived from *R. hybrida*, amplified polymorphic alleles in seven other *Rosa* species tested. This may indicate that these microsatellite loci are conserved within the genus *Rosa*, but it may also mean that genomes of the species are highly homologous.

Using the data of 24 STMS markers the 30 varieties were separated into 22 genoypes. Twenty varieties showed a unique genotype, one group consisted of a duplicate sample and a group with eight varieties consisted of clones of one variety but vegetatively multiplied at different trials. Results of the latter clearly indicate the repeatability of the STMS analysis. This distinction among the varieties could also be achieved by fingerprinting with subsets of two or three selected STMS markers.

Based on the number of unique alleles in rootstocks and assuming independent inheritance of an allele, theoretically very high numbers of allelic phenotypes can be expected. However, the number of allelic phenotypes detected was about the same as the number of observed alleles. This may be related, at least partially, to the peculiar *canina* meiosis as is described for all species in *Rosa* sect. *Caninae* (Täckholm 1920; Blackburn and Heslop-Harrison 1921) to which the majority of the rootstock varieties belong. Analysis of progenies from interspecific crosses with the same selected STMS markers as described here suggested that, during meiosis, bivalent formation preferentially takes place between chromosomes that consistently share the same microsatellite alleles (Nybom et al. 2002). Consequently, allele recombination is quite limited and allelic phenotypes appear very homologous.

The UPGMA dendrogram illustrates the STMS distance between hiprose, rootstock and Hybrid Tea varieties. Within the rootstocks, the distance between varieties originating from different species is relatively high and varieties within a species are well clustered, although these clusters do contain some varieties belonging to different species. Distinct clustering of accessions belonging to one *Rosa* species was also observed with RAPD markers (Debener et al. 1996; Millan et al. 1996; Reynders-Aloisi and Bollereau 1996; Jan et al. 1999; Martin et al. 2001). The molecular differentiation contrasts to the fact that it is difficult to distinguish all those varieties using morphological measurements. It will be interesting to test a large set of varieties from different species to assess the STMS genealogy between *Rosa* species more clearly, and compare this with morphological classifications and known pedigree information.

Identification of Hybrid Tea varieties

All Hybrid Tea varieties, except for the known color mutants, were uniquely identified with the microsatellite markers. In fact, only three selected STMS loci were suf-

ficient to distinguish between all unique Hybrid Tea varieties. Color mutants are due to a mutation in one of the genes involved in the synthesis of the color components. The chance of detecting the mutations with molecular markers is very small (Weising et al. 1995). No mutation-related polymorphisms were observed with microsatellites in *Pelagonium* (Becher et al. 2000) and peach (Testolin et al. 2000), nor in rose with RAPD and AFLP markers (Debener et al. 2000). Therefore, it can be expected that mutants have the same microsatellite fingerprint as the original variety and other mutants derived from the same variety. This implies that this set of highly polymorphic microsatellite markers is very suited to relate mutants to the original variety. As long as the original variety or other mutants of that variety are present, the mutant will be detected by the fact that it is part of a cluster of identical genotypes.

The high levels of polymorphisms generated by microsatellite markers in rose, have also been found in diploid *Rosaceae* species like *Malus × domestica* (Guilford et al. 1997; Gianfranschi et al. 1998; Hokanson et al. 1998) and *Prunus persica* (Sosinski et al. 2000), although in the latter case eight STMS markers were needed to identify 28 varieties. For roses it means that for constructing a molecular database for all existing and new varieties, STMS markers may be selected which fulfill the criteria for use in variety identification, i.e. polymorphic (the highest numbers of allelic phenotypes and effective alleles), with the best scorability and evenly distributed over the seven chromosomes. In pairwise comparison of varieties, even STMS markers with a lower scorability will be useful.

The high level of variation detected with the STMS markers may be related to both outcrossing and the polyploid nature of the rose varieties, which is confirmed by the high number of heterozygotes (data not shown) observed for the majority of the selected markers (average heterozygosity = 0.74).

As reviewed by DeVries and Dubois (1996), the genetic variability within the modern Hybrid Tea varieties is limited since most of the varieties share several common progenitors. In contrast, most of the rootstock varieties are taxonomically classified to several wild *Rosa* species (Leemans and Van der Laar 1977). This is consistent with the fact that many microsatellite alleles observed at high frequencies in Hybrid Tea varieties are not observed, or only at very low frequencies, in rootstock varieties. Furthermore, also the total number of different alleles was much higher in the rootstock varieties as compared to the Hybrid Tea varieties. Still, some of the rootstock varieties were relatively similar to others, which we did not observe in the Hybrid Tea varieties (Fig. 1). Lower similarities between some of the wild species as compared to modern rose varieties, including Hybrid Teas, were also observed with RAPD markers (Debener et al. 1996). As discussed before this may be the result of the *Canina* meiosis (Nybom et al. 2002). Remarkably, the number of different allelic phenotypes was relatively much higher in Hybrid Tea varieties. This

is likely to be the result of high breeding activities within a very limited pool of genotypes.

Using codominant markers, genetic relationships between varieties and/or species can be accurately described. However, it is not always possible to assess the exact STMS genotype of polyploid outbreeding species (Provan et al. 1996; Falque et al. 1998; Becher et al. 2000; Mengoni et al. 2000). This limits the discriminative power of a codominant marker and, as a consequence, scoring was done in a dominant way (allele absence/presence). Nevertheless, UPGMA clustering of Jaccard similarities clearly demonstrated distinct distances between groups, varieties and mutants, and proved to be a powerful tool for identification of mutants.

In conclusion, we have isolated and characterized a set of highly polymorphic and well-scorable microsatellite markers, which are very effective in generating unique DNA profiles of rose varieties. These profiles can easily be converted into a numeric database. As a result these robust PCR-based markers can be used as a tool for reference collection management and will be very useful in Plant Breeders Right research, assessing essential derived varieties, illegal propagation and in future Distinctness, Uniformity and Stability (DUS) testing. In all cases the existence of a molecular database will strengthen the applicability of these markers.

References

Arens P, Van 't Westende W, Bugter R, Smulders MJM, Vosman B (2000) Microsatellite markers for the European tree frog *Hyla arborea*. Mol Ecol 9:1944–1946

Ballard R, Rajapakse S, Abbott A, Byrne D (1995) DNA markers in rose and their use for cultivar identification and genome mapping. Acta Hort 424:265–268

Becher SA, Steinmetz K, Weising K, Boury S, Peltier D, Renou J-P, Kahl G, Wolff K (2000) Microsatellites for cultivar identification in *Pelargonium*. Theor Appl Genet 101:643–651

Ben-Meir H, Vainstein A (1994) Assessment of genetic relatedness in roses by DNA fingerprint analysis. Sci Hort 58:115–121

Blackburn K, Heslop-Harrison JW (1921) The status of the British rose forms as determined by their cytological behavior. Ann Bot 35:159–188

Botta R, Scott NS, Eynard I, Thomas MR (1995) Evaluation of microsatellite sequence-tagged site markers for characterizing *Vitis vinifera* cultivars. Vitis 34:99–102

Bredemeijer GMM, Arens P, Wouters D, Visser D, Vosman B (1998) The use of semi-automated fluorescent microsatellite analysis for tomato cultivar identification. Theor Appl Genet 97:584–590

Bredemeijer GMM, Cooke RJ, Ganal MW, Peeters R, Isaac P, Noordijk Y, Rendell S, Jackson J, Röder MS, Wendehake K, Dijcks M, Amelaine M, Wickaert V, Bertrand L, Vosman B (2002) Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theor Appl Genet (in press) DOI 10.1007/s00122-002-1038-6

Brownstein ML, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. BioTechniques 20:1004–1010

Cairns T (2000) Modern Roses XI, The World Encyclopedia of Roses. Academic Press

Clark JM (1988) Novel non-template nucleotide addition reactions, catalyzed by prokaryotic and eucaryotic DNA polymerases. Nucleic Acids Res 16:9677–9686

Cordeiro GM, Taylor GO, Henry RJ (2000) Characterization of microsatellite markers from sugarcane (*Saccharum* sp.) a highly polyploid species. Plant Sci 155:161–168

Cubero JI, Millan T, Osuna F, Torres AM, Cobos S (1996) Varietal identification in *Rosa* by using isozyme and RAPD markers. Acta Hort 424:261–264

Debener TL, Bartels C, Mattiesch L (1996) RAPD analysis of genetic variation between a group of rose cultivars and selected wild rose species. Mol Breed 119:71–74

Debener TL, Janakiram T, Mattiesch L (2000) Sports and seedlings of rose varieties analyzed with molecular markers. Plant Breed 119:71–74

Debener TL, Mattiesch L, Vosman B (2001) A molecular map for roses. Acta Hort 547:283–287

Depeiges A, Goubely C, Lenoir A, Cocherel S, Picard G, Raynal M, Grellet F, Delseny M (1995) Identification of the most represented repeated motifs in *Arabidopsis thaliana* microsatellite loci. Theor Appl Genet 91:160–168

De Vries DP, Dubois LAM (1996) Rose breeding: past, present, prospects. Acta Hort 424:241–248

Devos KM, Bryan GJ, Collins AJ, Stephenson P, Gale MD (1995) Application of two microsatellite sequences in wheat storage proteins as molecular markers. Theor Appl Genet 90:247–252

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

Falque M, Keurentjes J, Bakx-Schotman JMT, Dijk PJ (1998) Development and characterization of microsatellite markers in the sexual-apomictic complex *Taraxacum officinale* (dandelion). Theor Appl Genet 97:283–292

Gianfranceschi L, Seglias N, Tarchini R, Komjane M, Gessler C (1998) Simple sequence repeats for the genetic analysis of apple. Theor Appl Genet 96:1069–1076

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

Hartl L, Clark AG (1997) Principles of population genetics. Sinaur Associates Inc, Sunderland, Massachusetts

Hokanson SC, Szewc-McFadden AK, Lamboy WF, McFerson JR (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus × domestica* Borkh. core subset collection. Theor Appl Genet 97:671–683

Jan CH, Byrne DH, Manhart J, Wilson H (1999) Rose germplasm analysis with RAPD markers. Hort Sci 34:341–345

Karagyozov L, Kalcheva ID, Chapman VM (1993) Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. Nucleic Acids Res 21:3911–3912

Lagercrantz U, Ellegran H, Anderson L (1993) The abundancy of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21:1111–1115

Lavi U, Akkaya M, Bhagwat A, Lahav E, Cegan PB (1994) Methodology of generation and characteristics of simple sequence repeat DNA makers in avocado (*Persea americana* M.). Euphytica 80:171–177

Leemans JA, Van der Laar HJ (1977) Rootstocks for roses. Characteristics and cultural value. Stichting Plant Publicity Holland Boskoop, The Netherlands

Liu Q, Thorland EC, Sommer SS (1997) Inhibition of PCR amplification by a point mutation downstream of a primer. Biotechniques 22:292–300

Martin M, Piola F, Chessel Jay M, Heizmann P (2001) The domestication process of the Modern Rose: genetic structure and allelic composition of the rose complex. Theor Appl Genet 102:398–404

Mengoni A, Gori A, Bazzicalupo M (2000) Use of RAPD and microsatellite (SSR) variation to assess genetic relationships among populations of tetraploid alfalfa, *Medigo sativa*. Plant Breed 119:311–317

Millan T, Osuna F, Cobos S, Torres M, Cubero JI (1996) Using RAPDs to study phylogenetic relationships in Rosa. Theor Appl Genet 92:273–277

Miller MJ, Yuan B (1997) Semiautomated resolution of overlapping stutter patterns in genomic microsatellite analysis. Anal Biochem 251:50–56

Mörchen M, Cuguen J, Michaelis G, Hänni C, Saumitou-Laprade P (1996) Abundance and length polymorphism of microsatellite repeats in *Beta vulgaris*. Theor Appl Genet 92:326–333

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Nybom H, Esselink DG, Werlermark G, Vosman B (2002) Microsatellite DNA marker inheritance indicates preferential pairing between two highly homologous genomes in polyploid and hemisexual dog roses, *Rosa* L. sect. *Caninae*. Heredity (in press)

Odelberg SJ, White R (1993) A method for accurate amplification of polymorphic CA-repeat sequences. PCR Methods Appl 3:7–12

Perlin MW, Lancia G, Ng SK (1995) Towards fully automated genotyping: genotyping microsatellite markers by deconvolution. Am J Hum Genet 57:1199–1210

Provan J, Powell W, Waugh R (1996) Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). Theor Appl Genet 92:1078–1084

Rae SJ, Aldam C, Dominguez I, Hoebrechts SR, Edwards KJ (2000) Development and incorporation of microsatellite markers into the linkage map of sugar beet (*Beta vulgaris* spp.). Theor Appl Genet 100:1240–1248

Rajapakse S, Hubbard M, Kelly JW, Abbott A, Ballard R (1992) Identification of rose culivars by restriction fragment length polymorphism. Sci Hort 52:237–245

Reynders-Aloisi S, Bollereau P (1996) Characterization of genetic diversity in genus *Rosa* by randomly amplified polymorphic DNA. Acta Hort 424:253–260

Röder MS, Wendehake K, Korzun V, Bredemeijer G, Loborie D, Bertranel L, Isaac P, Rendell S, Jackson J, Cooke RJ, Vosman B, Ganall MW (2002) Construction and analysis of a microsatellite-based database of European wheal varieties. Theor Appl Genet (in press) DOI 10.1007/s00122-002-1061-7

Schwengel DA, Jedlicka AE, Nanthakumar EJ, Weber JL, Levitt RC (1994) Comparison of fluorescence-based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. Genomics 22:46–54

Smulders MJM, Bredemeijer G, Rus-Kortekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor Appl Genet 97:264–272

Smulders MJM, van der Schoot J, Arens P, Vosman B (2001) Trinucleotide repeat microsatellite markers for Black Poplar (*Populus nigra* L.). Mol Ecol Notes 1:188–190

Song QJ, Quigley CV, Nelson RL, Carter TE, Boerma HR, Strachan JL, Cregan PB (1999) A selected set of trinucleotide simple sequence repeat markers for soybean cultivar identification. Plant Var Seeds 12:207–220

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

Sprecher CJ, Puers C, Lins AM, Schumm JW (1996) General approach to analysis of polymorphic short tandem repeat loci. Biotechniques 20:266–276

Täckholm G (1920) On the cytology of the genus *Rosa*. Svensk Bot Tidskrift 14:300–311

Taramino G, Tarchini R, Ferrario S, Lee M, Pe ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorgum bicolor*. Theor Appl Genet 95:66–72

Testolin R, Marrazzo T, Cipriani G, Quarta R, Verde I, Dettori MT, Pancaldi M, Sansavini S (2000) Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. Genome 43:512–520

Torres AM, Millán T, Cubero JI (1993) Identifying Rose cultivars using random amplified polymorphic DNA markers. Hort Sci 28:333–334

Vainstein A, Ben-Meir H (1994) DNA fingerprint analysis of Roses. J Am Soc Hort Sci 119:1099–1103

Van de Wiel C, Arens P, Vosman B (1999) Microsatellite retrieval in lettuce (*Lactuca sativa* L.). Genome 42:139–149

Van der Schoot J, Pospíkova M, Vosman B, Smulders MJM (2000) Development and characterization of microsatellite markers in black popular (*Populus nigra* L.). Theor Appl Genet 101:317–322

Vosman B (1998) The use of molecular markers for the identification of tomato cultivars. In: Karp A, Isaac PG, Ingram DS (eds) Molecular tools for screening biodiversity, pp 382–387

Vosman B, Arens P, Rus-Kortekaas W, Smulders MJM (1992) Identification of highly polymorphic DNA regions in tomato. Theor Appl Genet 85:239–244

Vosman B, Cooke RJ, Ganal M, Peeters R, Isaac P, Bredemeijer G (2001) Standardization and application of microsatellite markers for variety identification in tomato and wheat In: Proc Int Symp on Molecular Markers for Characterizing Genotypes and Identifying Cultivars in Horticulture. Acta Hort 546:307–316

Weising K, Nybom H, Wolff K, Meyer W (1995) DNA fingerprinting in plants and fungi. CRC Press, Bocta Raton, Florida