IDENTIFICATION OF TOMATO CULTIVARS USING MICROSATELLITES

P. Arens, G. Bredemeijer, M.J.M. Smulders and B. Vosman
Centre for Plant Breeding and Reproduction Research (CPRO-DLO)
P.O. Box 16
6700 AA Wageningen
The Netherlands

Abstract

Microsatellite polymorphisms can be detected both by Southern hybridization and by Polymerase Chain Reactions. This paper describes the use of both applications of microsatellites for cultivar identification in tomato. The discriminating power of microsatellite hybridisation was demonstrated on two old introgression free cultivars. The potential of the Sequence Tagged Microsatellites Sites (STMS) application for cultivar identification was investigated using eight microsatellite containing loci found in the EMBL database. Only two loci were polymorphic in a test set of five tomato cultivars. For these two loci, i.e. LELE25 and LEEF1A, a screening of 16 tomato cultivars revealed three and six length alleles, respectively.

Key words: Cultivar identification, *Lycopersicon esculentum*, SSR, Simple sequence repeats, STMS, DNA-fingerprinting.

1. Introduction

For identification and registration of tomato cultivars until now morphological and physiological characteristics have been used. These characteristics have a number of drawbacks: they can be influenced by (a)biotic factors and it is necessary to study plants throughout the whole development. Molecular markers do not have these drawbacks and could be useful for fast cultivar identification and possibly registration procedures. Especially in a crop like *Lycopersicon esculentum*, where genetic diversity seems very limited (Miller and Tanksley, 1990) the molecular marker of choice must be very informative. Microsatellites appear suitable because of their highly polymorphic character. Microsatellites, or tandemly repeated simple sequences, are among the most common types of repetitive DNA. They contain a basic repeat unit of 2-8 base pairs (Hamada et al., 1982, 1984; Tautz and Rentz, 1984). The variation found in this type of satellite DNA has been suggested to be due to variations in the copy number of the basic repeat unit, possibly caused by slippage during replication (Schlötterer and Tautz, 1992).

Microsatellites can be used for cultivar identification in two different ways. Firstly in Southern hybridisations, where synthetic oligonucleotides complementary to microsatellites allow the simultaneous detection of several microsatellite loci, resulting in a multilocus DNA fingerprint (Epplen, 1992). Secondly, in a single locus approach via PCR, known as Sequence Tagged Microsatellite Sites (STMS; Beckmann and Soller, 1990). A primer pair flanking a microsatellite

enables amplification of this microsatellite, and any polymorphism in the length of the microsatellite between cultivars can thus be detected (Thomas and Scott, 1993). This paper describes the use of both applications of microsatellites for cultivar identification in tomato.

2. Material and methods

2.1. Plant material

All *Lycopersicon esculentum* cultivars used (legends figure 2) were obtained from the tomato collection of the Centre of Genetic Resources (CGN, part of CPRO-DLO, The Netherlands).

2.2. DNA extraction and digestion

Nuclear DNA was extracted from frozen leaves essentially as described by Bernatzky and Tanksley (1986) with some slight modifications (Vosman et al., 1992). From these DNA samples, 1 μg was digested with *Taq*I (Life Technologies). Fragments were size-separated through a 0.8% agarose gel and blotted onto Hybond-N+ filter (Amersham).

2.3. Southern hybridisation

GACA-containing microsatellites were detected using the plasmid pWVA16 as a probe. The plasmid pWVA16 and the labelling and hybridizations protocols were as described previously (Rus-Kortekaas et al., 1994; Arens et al., 1995).

2.4. Database search

Sequences from the genus *Lycopersicon* were selected from the EMBL (Version 41.0) and Genbank (Version 83.0) databases using SRS (Version 3.1), running at the CAOS-CAMM Centre in Nijmegen, The Netherlands. The 542 entries obtained were subsequently screened for homology to a number of di- tri- and tetranucleotide motifs using FASTA. For selected microsatellite sequences, primer pairs were designed using PRIMER (Version 0.5) so as to amplify microsatellite-containing DNA fragments of 100-250 bp in length. These primers were synthesised by Isogen (Amsterdam, the Netherlands).

2.5. Polymerase chain reactions

Amplification reactions contained, in 25 μl: 10 ng genomic DNA, 0.32 μM of each primer, 100 μM deoxyribonucleotides, 50 mM KCl, 20 mM Tris-HCl (pH 8.4), 1.5 mM $MgCl_2$, 0.05% (v/v) polyoxyethylene ether (W-1), and 0.5 U Taq DNA polymerase (Life Technologies). Amplifications were performed in tubes or microtiter plates using a Hybaid Omni Gene thermal cycler. Standard cycling conditions were: 1 cycle of 94°C for 3 min; 30 cycles of 50°C for 45 s, 72°C for 1 min 45 s, and 94°C for 45 s. After the final cycle, 1 cycle of 50°C for 45 s and 72°C for 3 min was added. In two cases the annealing temperature and the number of cycles were modified as described in the results.

2.6. Detection of STMS

After PCR amplification, 7 µl aliquots of the reaction mixture were electrophoresed using 2% agarose gels with Tris-borate buffer (89 mM Tris-borate, 89 mM boric acid, and 2 mM EDTA, pH 8.0) to check for amplification. A 1 kb DNA ladder (Life Technologies) was used as marker to estimate the sizes of the amplification products.

Samples for PAGE electrophoresis were prepared by adding an equal volume of formamide, containing 10 mM NaOH and 0.05% bromphenol blue, to individual PCR reactions. After denaturation at 80°C for 5 min followed by quenching on ice, samples were analyzed on vertical gels (6% polyacrylamide, 8 M urea, Tris-borate buffer) using a Model S2 sequencing gel electrophoresis apparatus (Life Technologies). The DNA bands were visualised by silver staining according to the Silver sequence™ DNA sequencing system (technical manual 1993, Promega). The sizes of the PCR products were determined by electrophoresis of an accompanying sequence reaction using pGEM-3Zf(+) control DNA (Promega).

3. Results and discussion

3.1. Cultivar identification based on Southern hybridization using microsatellite-containing probes

Oligonucleotide probes complementary to microsatellite sequences can be used for identification of tomato cultivars. Even though genetic diversity within *L. esculentum* is low, it is possible to identify 15 different cultivars by a unique DNA fingerprint using either GATA- or GACA-containing probes (Vosman et al., 1992). These 15 cultivars representing the whole spectrum of different tomato types were mostly more recent cultivars with introgressed resistance genes. When comparing the two introgression-free, obsolete cultivars Moneymaker and Premier, Van der Beek et al. (1992) found only three polymorphic probe-enzyme combinations among 195 RFLP probes and six different restriction enzymes. We DNA fingerprinted these two cultivars using a GACA-containing probe (figure 1). Although the GACA-containing probe detected a lower number of polymorphic bands than the GATA-probe, there were five differences between the fingerprint patterns of the two cultivars.

3.2. Microsatellite containing loci from the EMBL database

Southern hybridization is a method that requires a well-equipped laboratory and is time-consuming. Detection of microsatellite-containing loci using PCR is faster, technically not so demanding, and allows the simultaneous detection of both alleles from a locus. Therefore, we set out to develop a STMS approach for tomato cultivar identification.

In the present study eight different microsatellite loci selected from the EMBL nucleic acid database were chosen for further analysis. To allow PCR amplification primers based on unique DNA sequences flanking the microsatellite array were designed (table 1). Four loci contained perfect repeats: three dinucleotide and one tetranucleotide

repeat; four loci had compound perfect repeat sequences, with two or more adjacent repeat runs.

3.3. PCR amplification and length variation of the microsatellites

Amplification was initially carried out on genomic DNA of a test set of five cultivars, including Moneymaker, Vision, Roma VF, Evita and Calypso, under standard PCR conditions. The PCR products were analyzed on 2% agarose gels. For two loci, i.e. LEGAST 1 and LEEF1A amplification conditions were modified by altering the annealing temperature (55°C) and the number of cycles (25). With the eight primer sets, shown in table 1, amplification was seen in all five cultivars. Six primer sets gave amplification products around the expected size; however, primers constructed to amplify parts of the LEPRP4 and LEEF1A loci generated products which exceeded the expected sizes. This could be due to deletions in the clone originally sequenced, or insertions in the cultivars used in the present study. To resolve this, amplification should be performed on the cultivars from which the sequences were derived.

Since agarose as a separating medium presents a number of limitations with respect to detection and high resolution of the amplification products (Allen et al., 1989) agarose gel electrophoresis was, in fact, used only to check for amplification (conditions and presence) and to estimate the amount of PCR reaction mixture for PAGE electrophoresis. Denaturing PAGE electrophoresis in combination with the silver staining was used for detection of polymorphism. Of the microsatellite loci studied, only LELE25 and LEEF1A exhibited polymorphism among the cultivars of the test set. Therefore these loci were used on 16 tomato cultivars.

Amplification of the $(TA)_n$-microsatellite of the LELE25 locus gave amplification products in all cultivars (figure 2a). No cultivar exhibited more than two intense bands which is consistent with a locus-specific amplification (Ellegren, 1992). The LELE25 locus showed three length alleles in the cultivars studied of 215, 217 and 219 bp respectively, values close to the expected size of 225 bp (table 1). Three cultivars were heterozygous at this locus, these cultivars were F_1 hybrids.

In contrast with LELE25, amplification of the $(TA)_n(ATA)_n$-microsatellite of the LEEF1A locus generated an inferior image, with several stuttering bands of nearly equal intensity (figure 2b). An image is considered optimal when the true length alleles are easily distinguished from stuttering bands and artifacts on the basis of product intensity (Schwengel et al., 1994). This was not the case with the primer Bset designed to amplify the microsatellite of the LEEF1A locus. Attempts to optimize the PCR conditions for this locus by testing higher annealing temperatures (57 and 59°C) or lower number of cycli (23) failed. Possibly, the use of non-denaturing gels (Neilan et al., 1994) or the designing of another primer set for the LEEF1A locus may eliminate the nonspecific amplification products. It is also possible that this problem is directly related with the type of locus. The LEEF1A locus is an elongation factor which could well be a member

of a larger gene family. Although it is difficult to establish the actual length alleles on the basis of figure 2b, the results do indicate polymorphism of this microsatellite among tomato cultivars. Probably, the $(TA)_n(ATA)_n$-microsatellite has six alleles in the 16 samples that were amplified, ranging from about 193 to 213 bp in size. In the experiment performed for photographing the gel no amplification of the LEEF1A locus was observed in cv. Moneymaker using the optimised PCR conditions for this locus. However, when using DNA from a new DNA isolation of cv. Moneymaker, amplification did take place and yielded a banding pattern similar as observed for cv. Mirabell.

4. Conclusions

Microsatellite motifs used as probes in Southern hybridizations are useful for identification of tomato cultivars. Both modern and introgression free cultivars can be distinguished by a unique fingerprint pattern.

The present data show that two of eight STMS loci studied in tomato are polymorphic and thus may be useful for identifying cultivars. Optimisation of the PCR conditions was important. The optimum conditions were not the same for the various primer pairs. In addition, the optimum conditions for a primer pair can differ between cultivars, as observed for different *Arabidopsis* strains (Bell and Ecker, 1994). Finally, design of other primer sets for a certain locus may be necessary to eliminate nonspecific amplification products when optimisation of PCR conditions has no effect, like in the case of the LEEF1A locus.

Combining the genetic variation observed for these two loci does not produce a unique STMS-combined genotype for all 16 tomato cultivars shown in figure 2. Therefore, combination of more polymorphic microsatellite loci will be necessary to allow identification of all cultivars, as shown for grapevine cultivars (Thomas and Scott, 1993). As the microsatellites we used were among the longest perfect repeats found, the currently available database information is likely to be insufficient for cultivar identification of the whole range of tomato cultivars. Therefore isolation of new microsatellite loci from highly enriched DNA libraries, appears a necessity for cultivar identification in crops with low genetic diversity.

Acknowledgements

The database searches were done using the bioinformatics facilities of the CAOS-CAMM Center in Nijmegen, The Netherlands.

Research is funded by the Ministry of Agriculture, Nature Management and Fisheries and the EG Biotechnology program B102CT-930295.

References

Allen, R.C., Graves, G. and Budowle, B., 1989. Polymerase chain reaction amplification products separated on rehydratable polyacrylamide gels and stained with silver. BioTechniques 7:736-744.

Arens, P.F.P., Odinot, P., Van Heusden, A.W., Lindhout, P. and Vosman, B., 1995. GATA- and GACA-repeats are not evenly distributed throughout the tomato genome. Genome 38:84-90.

Beckmann, J.S. and Soller, M., 1990. Towards a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. Bio/technology 8:930-932.

Bell, C.J. and Ecker, J.R., 1994. Assignment of 30 microsatellite loci to the linkage map of Arabidopsis. Genomics 19:137-144.

Bernatzky; R. and Tanksley, S.D. 1986 Genetics of actin-related sequences in tomato. Theor Appl Genet 72:314-321

Ellegren, H., 1992. Polymerase-chain-reaction (PCR) analysis of microsatellites: A new approach to studies of genetic relationships in birds. The Auk 109:886-895.

Epplen, J.T. 1992. The methodology of multilocus DNA fingerprinting using radioactive or nonradioactive oligonucleotide probes specific for simple repeat motifs. In: Chrambach A, Dunn MJ, Radola BJ (eds) Advances in Electrophoresis Volume 5. VCH, New York, Weinheim, pp. 62-112.

Hamada, H., Petrino, M.G. and Kakunaga, T., 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci., 79:6465-6469.

Hamada, H., Petrino, M.G., Kakunaga, T., Seidman, M. and Stollar, B.D., 1984. Characterization of genomic poly(dT-dG) poly(dC-dA) sequences: structure, organization and conformation. Mol. Cell Biol., 4:2610-2621.

Miller, J.C. and Tanksley, S.D., 1990. RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet., 80:437-448.

Neilan, B.A., Leigh, D.A., Raply, E., McDonald, B.L. 1994. Microsatellite genome screening: rapid non-denaturing, non-isotopic dinucleotide repeat analysis. BioTechniques 17:708-712.

Rus-Kortekaas, W., Smulders, M.J.M., Arens, P. and Vosman, B., 1994. Direct comparison of levels of genetic variation in tomato detected by a GACA-containing microsatellite probe and by random amplified polymorphic DNA. Genome 37:375-381.

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

Schwengel, D.A., Jedlicka, A.E., Nanthakumar, E.J., Weber, J.L. and Levitt, R.C., 1994. Comparison of fluorescence-based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. Genomics 22:46-54.

Tanksley, S.D., Young, N.D., Paterson, A.H. and Bonierbale, M.W., 1989. RFLP mapping in plant breeding: new tools for an old science. Biotechnology, 7:257-264.

Tautz, D. and Renz, M., 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res., 12:4127-4138.

Thomas, M.R. and Scott, N.S., 1993. Microsatellite repeats in grapevine reveal DNA polymorphisms when analyzed as sequence-tagged sites (STSs). Theor. Appl. Genet. 86:985-990.

Van der Beek, H., Verkerk, R., Zabel, P. and Lindhout, P. 1992. Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: CF9 (resistance to *Cladosporium fulvum*) on chromosome 1. Theor. Appl. Genet., 84:106-112.

Vosman, B., Arens, P., Rus-Kortekaas, W. and Smulders, M.J.M., 1992. Identification of highly polymorphic DNA regions in tomato. Theor. Appl. Genet., 85:239-244.

Table 1 - *Lycopersicon* microsatellite loci.

Primer set	Locus	Repeat	Primer sequences	Fragment size (bp)
1	LELE25	(TA)11	5'-TTCTTCCGTATGAGTGAGT-3'	225
			5'-CTCTATTACTTATTATTATCG-3'	
2	LEGAST1	(TA)12	5'-ATCTCTATTGTTTTCGACTCG-3'	143
			5'-TCTGTTGTTGCTGCTGCTC-3'	
3	LEGTOM5	(TA)10	5'-AAAGATAAAGCATGAAATGAA-3'	181
			5'-GGAGTTGAGATAAAGTGAAGA-3'	
4	LEEF1A	(TA)8(ATA)9	5'-AAATAATTAGCTTGCCAATTG-3'	131
			5'-CTGAAAGCAGCAACAGTATTT-3'	
5	LEILV1B	(T)8(TA)10(T)5	5'-GATCGACACATTTGAATTTGT-3'	143
			5'-GGTCACTAATTAATTGATTCC-3'	
6	LEPRP4	(TAT)3(TGT)5	5'-TTCATTTCTTGCAACTACGAT-3'	200
			5'-CATACTAGCAACATCAAAGGG-3'	
7	LESSRPSPG	(TATT)5	5'-GAATATATCGGGGACAATCTC-3'	219
			5'-AACGAAATCTTTGTTCAGTGA-3'	
8	LECAB9	(TA)6(CA)3	5'-TTTATTATCCCAGAAGCCTTC-3'	118
			5'-CCTCACATTTAAACAAATTGC-3'	

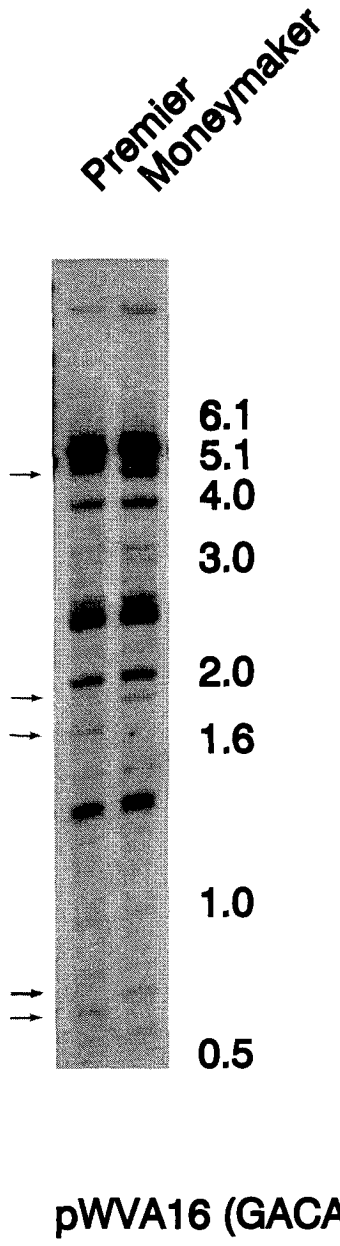

Figure 1 - DNA-fingerprint analysis of the cultivars Premier and Moneymaker with pWVA16 as probe. DNA was digested with *TaqI*. Fragment sizes in kb are indicated in the right margin, arrows indicate polymorphic bands.

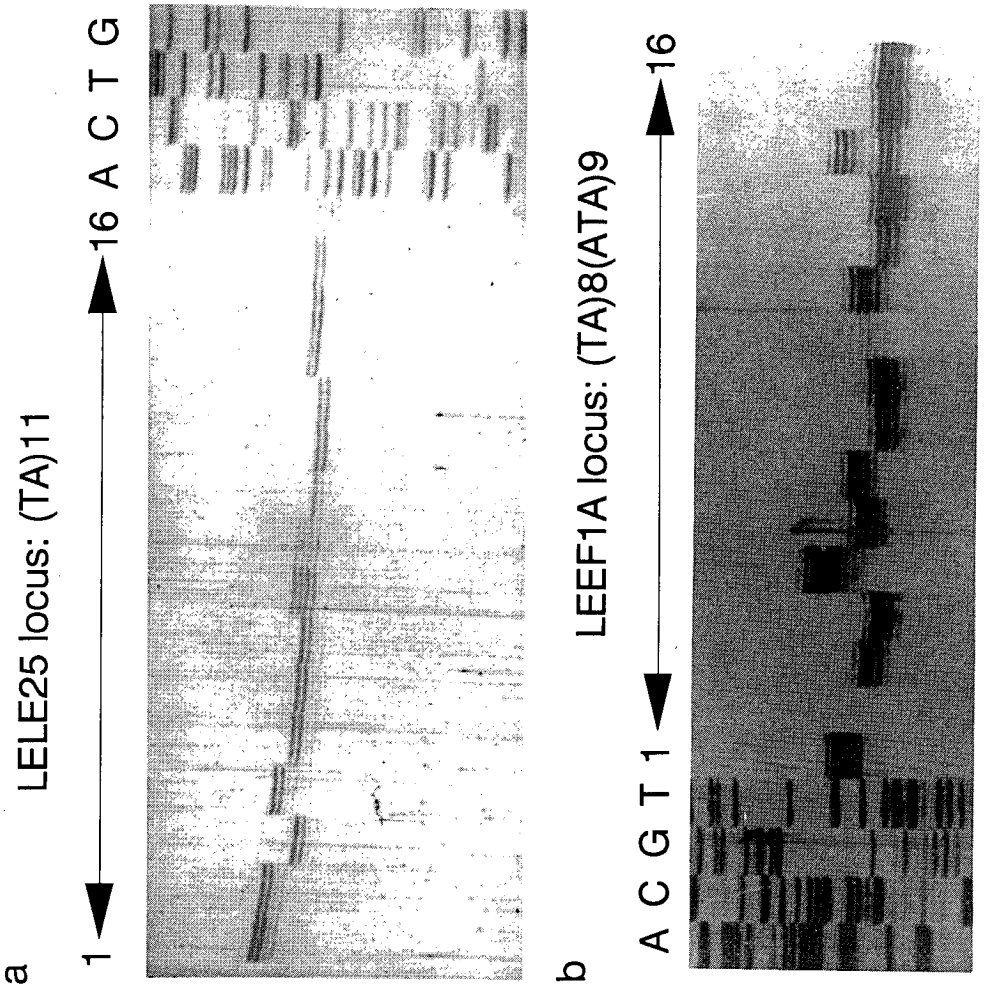

Figure 2 - PCR amplification of the LELE25 (a) and the LEEF1A (b) microsatellite loci in 16 cultivars of *Lycopersicon esculentum*. Lane 1, Premier; lane 2, Moneymaker; lane 3, Mirabell; lane 4, Pipo; lane 5, San Marzino Lampadone; lane 6, San Marzano; lane 7, Marmande; lane 8, Dombito; lane 9, Vision; lane 10, Roma VF; lane 11, Carma; lane 12, Trend; lane 13, Liberto; lane 14, Evita; lane 15, Blizzard; lane 16, Calypso. A,C,G,T: sequence ladders.