Theor Appl Genet (2006) 112: 1248–1257
DOI 10.1007/s00122-006-0226-1

ORIGINAL PAPER

**Chunxian Chen · Ping Zhou · Young A Choi
Shu Huang · Fred G. Gmitter Jr**

Mining and characterizing microsatellites from citrus ESTs

Received: 15 July 2005 / Accepted: 15 January 2006 / Published online: 11 February 2006
© Springer-Verlag 2006

Abstract Freely available computer programs were arranged in a pipeline to extract microsatellites from public citrus EST sequences, retrieved from the NCBI. In total, 3,278 bi- to hexa-type SSR-containing sequences were identified from 56,199 citrus ESTs. On an average, one SSR was found per 5.2 kb of EST sequence, with the tri-nucleotide motifs as the most abundant. Primer sequences flanking SSR motifs were successfully identified from 2,295 citrus ESTs. Among those, a subset (100 pairs) were synthesized and tested to determine polymorphism and heterozygosity between/within two genera, sweet orange (*C. sinensis*) and *Poncirus* (*P. trifoliata*), which are the parents of the citrus core mapping population selected for an international citrus genomics effort. Eighty-seven pairs of primers gave PCR amplification to the anticipated SSRs, of which 52 and 35 appear to be homozygous and heterozygous, respectively, in sweet orange, and 67 and 20, respectively, in *Poncirus*. By pairing the loci between the two intergeneric species, it was found that 40 are heterozygous in at least one species with two alleles (9), three alleles (28), or four alleles (3), and the remaining 47 are homozygous in both species with either one allele (31) or two alleles (16). These EST-derived SSRs can be a resource used for understanding of the citrus SSR distribution and frequency, and development of citrus EST-SSR genetic and physical maps. These SSR primer sequences are available upon request.

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-006-0226-1 and is accessible for authorized users.

Communicated by S. J. Knapp

C. Chen · P. Zhou · Y. A Choi · S. Huang · F. G. Gmitter Jr (✉)
Institute of Food and Agricultural Sciences,
Citrus Research and Education Center, University of Florida,
700 Experiment Station Road, Lake Alfred, FL 33850, USA
E-mail: fgg@crec.ifas.ufl.edu
Tel.: +1-863-9561151
Fax: +1-863-9564631

Introduction

Microsatellites, also called simple sequence repeats (SSRs) or short tandem repeats (STRs), are a group of tandem repeated sequences simply comprised of mono-, di-, tri-, tetra-, penta-, or hexa-nucleotide units. Compared to other molecular markers, SSRs are uniquely characterized by their simplicity, abundance, ubiquity, variation, co-dominance, and multi-alleles among genomes (Powell et al. 1996). The polymorphism, mainly resulting from the number of repeat units, can easily be detected by PCR using primers flanking the SSR motif. SSRs have become a common tool broadly used in aspects of genetic mapping, molecular evolution, and systematic taxonomy in most genomes since they were first described in the 1980s (Hamada and Kakunaga 1982; Tautz and Renz 1984).

In general, SSRs are identified from either genomic DNA or cDNA sequences. Genomic SSR sequences are usually acquired by screening of small-insert genomic DNA libraries, either probing them with radioactively labeled probes or trapping them with biotinylated SSR motifs (Paniego et al. 2002; Lowe et al. 2004). These processes are very time-consuming and labor-intensive; furthermore, the SSRs acquired are limited to those probed SSR motifs, which in most cases are di- or tri-types. Submissions of DNA sequences, including a large portion of cDNAs, have been increasing exponentially over the past few years to public databases such as GenBank. An alternative strategy to develop SSR markers is to search for EST-based SSRs from cDNAs deposited in the public databases, using data mining pipelines composed primarily of SSR search and primer design programs. This EST-based approach has been successfully used in species such as barley (Thiel et al. 2003), maize (Sharopova et al. 2002), durum wheat (Eujayl et al. 2002), rye (Hackauf and Wehling 2002), sugarcane (Cordeiro et al. 2001) and grape (Scott et al. 2000). EST-SSR search programs, such as misa (Thiel et al. 2003), and ssr finder (Sharopova et al. 2002), have

been made available for public use; other SSR finding programs were also reported (Castelo et al. 2002; Robinson et al. 2004). Unlike screening a genomic library with predefined SSR motifs, SSR search programs can identify any SSR motifs within EST sequences, generating an overall view on the distribution and frequency of SSRs in the entire genome. The most important external program used in these data mining pipelines is Primer3, which can process the primer design in a batch mode (Rozen and Skaletsky 2000).

Some citrus SSRs, mostly di- and tri-nucleotide, were cloned from a genomic library for citrus SSR evaluation and characterization, cultivar identification, and genetic mapping (Kijas et al. 1995, 1997; Ahmad et al. 2003). Citrus genetic maps developed in various labs worldwide have primarily been populated with RAPD or/and AFLP markers in the entire genome (Cai et al. 1994; Luro et al. 1996; Simone et al. 1998; Ling et al. 1999), or focused on specific gene regions (Gmitter et al. 1996; Fang et al. 1998; Ling et al. 2000). Recently a collaborative project has aimed at construction of high density EST maps for sweet orange (*Citrus sinensis* L. Osb.) and *Poncirus trifoliata* L. Raf., which is an important component of a global citrus genome plan. Increasing numbers of citrus cDNAs are being released to the public sequence database, reaching nearly 60,000 as of July 7, 2004 (the time this work was initiated) and now more than 100,000 as of June 2005. Those EST–SSR search programs developed for other species with advanced genomic tool resources, together with the increasing number of citrus EST sequences available, will provide a shortcut for citrus geneticists to maximize the potential for the development of SSR and SNP based EST maps, and hasten the implementation of other SSRs in other genetic studies.

Typically, amplified SSR products are either separated and scored on sequencing polyacrylamide gels (PAG) using regular unlabeled (Thiel et al. 2003; Paniego et al. 2002) or isotope-labeled (Cregan et al. 2001) primers, or on an automatic capillary sequencer using fluorescently-labeled primers and subsequently analyzing by computer programs (Eujayl et al. 2002). Mass spectrometry has also been used for SSR genotyping (Paris and Jones 2002). Because of the inconvenience in preparation of and scoring from PAG, and the high cost of dye labeling each forward primer for separation on sequencers, alternative approaches have been tested. Instead of PAG, a new type of GTG or Metaphor Agarose was used to separate regular SSR products (Chani et al. 2002). Rather than dye-labeling each primer for SSR separation and scoring on automated sequencers, a common M13 forward primer sequence that can be labeled with different dies has been added to the 5′ terminal of one SSR primer (forward primer is usually used). It is used as a third primer, together with M13-tailed forward and regular reverse primers, for amplification of fluorescent products (Oetting et al. 1995); this approach can potentially provide substantial cost savings by eliminating the need to synthesize multiple labeled primer sequences for all loci.

In this paper, we report on the use of misa to exploit and characterize the microsatellites from citrus public EST sequences. We developed a Java program named "SeqFilter" to remove false EST sequences using keywords in the Fasta headlines. The amplification performances of M13-tailed and regular primers were also compared. Finally, an assessment was made of the usefulness of the derived SSRs for mapping in an F_1 intergeneric family of sweet orange (*C. sinensis*) and trifoliate orange (*P. trifoliata*).

Materials and methods

Plant materials

Five sweet oranges cultivars (*C. sinensis*, Sanford (Sa), Ridge Pineapple (RP), Succari (Su), Algerian (Al),

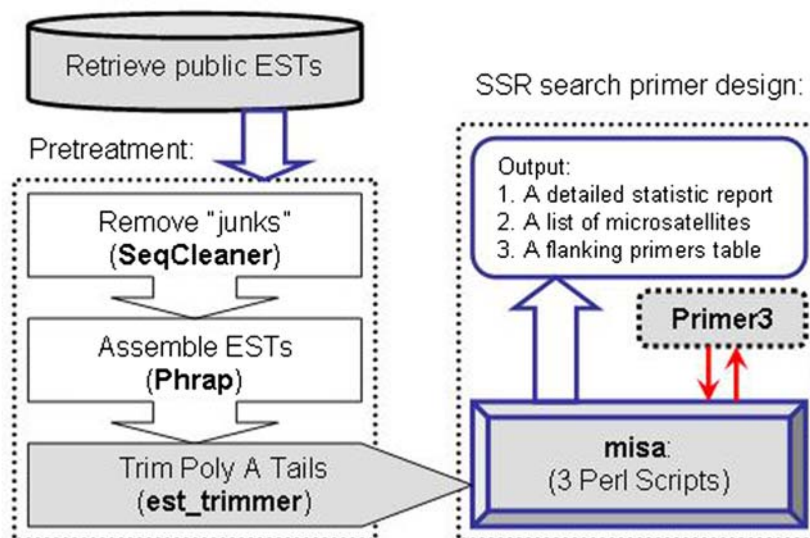


Fig. 1 Flow chart of the SSR mining procedure used with public citrus ESTs. The trimming step is optional if mono-type SSRs are excluded

and Ruby (Ru), three trifoliate orange biotypes (*P. trifoliata*), Argentina (Ar), Flying Dragon (FD), DPI0906 (Pt96), and eight F_1 offspring, were used to screen some of the SSRs identified from the pipeline output. These selections are parents of a so-called citrus core mapping population (CCMP), a composite of intergeneric F_1 families chosen by the International Citrus Genomics Consortium (ICGC) as the segregating family upon which to base international collaborative genomics efforts. Genomic DNA was isolated from tender leaves using the CTAB method described by Aldrich and Cullis (1993).

Retrieval of citrus EST sequences from NCBI

Using the keyword "Citrus mRNA", 61,976 putative citrus EST sequences were retrieved on 7 July 2004 from the National Center of Biotechnology Information (NCBI) for subsequent data mining process. We created a Java program named SeqFilter to remove all "junk" DNA sequences that are not true citrus ESTs. SeqFilter is a command-line program run under the Java run-time environment. Another function of this program is to extract only the sequence accession ID from the long FASTA headline, to simplify subsequent table-style SSR outputs. This program and its user guide can be downloaded from the SeqFilter page. The major contributors of these citrus EST sequences in NCBI were the University of California at Riverside (T. Close and M. Roose) and at Davis, USA (A. Dandekar), USDA-ARS at Fort Pierce, USA (M. Bausher and T. G. McCollum), National Institute of Fruit Tree Sciences, Japan (M. Omura), and CSIC-IATA, Spain (M. Sanchez-Ballesta). Most EST sequences were from sweet orange (*C. sinensis*, about 45,000), and others were mainly from four other *Citrus* species and two closely related genera, *Poncirus* and *Fortunella*.

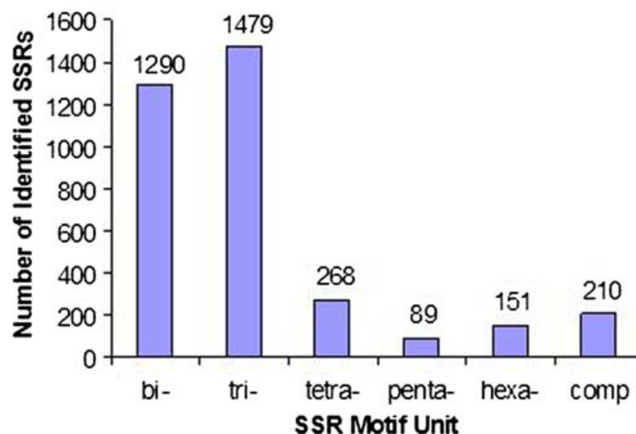


Fig. 2 Distribution of SSRs with different repeat unit size; *comp* compound type

Computer programs for data mining

All computer programs, Phred-Phrap-Consed (Gordon et al. 1998, 2001), Primer3 (Rozen and Skaletsky 2000), and misa (Thiel et al. 2003), were installed and/or compiled under Linux Red Hat 9.0. Red Hat was run inside a virtual machine, the VMWare workstation for Windows (VMWare Inc., Palo Alto, CA, USA), on a Dell Latitude D600 laptop. A flow chart summarizing the individual steps of the data-mining procedure was given in Fig. 1. The paired numbers representing SSR motif length and minimum repeat number in the misa configuration file were modified to 2–6, 3–5, 4–4, 5–4, and 6–4 in this study (mono-type excluded), and the parameters for Primer3 to design primers remained the same as described by Thiel et al. (2003).

Survey of SSR polymorphisms

One hundred pairs of SSR primers were selected for initial screening of two sweet orange varieties, Sanford

Table 1 Summary of SSR search in sequences assembled and categorized by "Phrap"

	Contigs (kb)	Singlets (kb)	Problems (kb)	Total (kb)
ESTs after assembled by "Phrap"	8,688 (8,365)	10,347 (5,517)	6,742 (3,347)	25,777 (17,229)
ESTs after trimmed by "est_trimmer"	8,678 (8,355)	10,067 (5,368)	6,623 (3,288)	25,368 (17,011)
Identified SSRs[a]	1,420	961	897	3,278
ESTs having SSRs	1,164	806	728	2,698
ESTs having more than 1 SSR	194	136	143	473
Compound SSRs	128	50	46	224
Bi-type	537	338	415	1,290
Tri-type	643	430	406	1,479
Tetra-type	120	109	39	268
Penta-type	47	32	10	89
Hexa-type	73	52	27	152
Subtotal of five types	1,420	961	897[b]	3,278

[a]The number of mono-type, mostly A or T, was 1,547, generated in a separate process. They were not included in the total or statistic calculations, and they were not used for primer design

[b]The "problems" sequences that failed to assemble could be either overlapped or single, so the number of unique SSRs will be smaller. It was indicated by later primer sequence comparison

Table 2 Occurrence and number of repeats of the top 30 SSR motifs in citrus ESTs

Repeats	4	5	6	7	8	9	10	11	12	13	Above[a]	Total
AG/CT	–	–	172	79	46	73	39	37	27	22	70	565
AC/GT	–	–	86	32	178	113	22	6	3	7	7	454
AT/AT	–	–	64	79	57	29	19	21	7	17	50	343
AAG/CTT	–	118	79	45	40	9	10	6	1	2	1	311
AAT/ATT	–	115	71	24	42	11	2	5	5	1	7	283
AGC/GCT	–	115	141	16	4	3	1					280
ACG/CGT	–	73	102	13	5							193
AGT/ACT	–	49	19	4	8	5	1					86
ACC/GGT	–	48	17	8	1	1					3	78
AAC/GTT	–	42	9	10	6			2				69
AGG/CCT	–	45	14	5	5							69
ACT/AGT	–	30	18	2	5	4	1		1			61
AAAT/ATTT	43	13	1	1								58
CCG/CGG	–	33	10	7		1		1				52
AAAG/CTTT	20	12	7	4					2			45
AATT/AATT	16					1					1	18
AAGACG/CGTCTT	18											18
ACAT/ATGT	12	3	1	1								17
AAAAG/CTTTT	11	4	1									16
AATG/CATT	12	1										13
AAAC/GTTT	9	3										12
AAAAT/ATTTT	12											12
AAGCAG/CTGCTT	9											9
AGCT/AGCT	8											8
AAAAAG/CTTTTT	8											8
ACTC/GAGT	4	3										7
AGAT/ATCT	3	4										7
AACTC/GAGTT	2	1		2								5
AACT/AGTT	2	1	1									4
AAGT/ACTT	3	1										4
Other motifs[a]	128	29	9	5	1	1						173
Total	320	743	822	337	398	251	95	78	46	49	139	3278

[a]The total numbers of SSRs with more than 13 repeats were presented, as well as the totals of other SSR motifs

(Sa) and Ridge Pineapple (RP), and two *Poncirus* biotypes, Argentina (Ar) and Flying Dragon (FD), to determine the levels of polymorphism between the two parental species. The PCR products were resolved on GTG Agarose gels for the initial screening.

SSR analysis was performed according to Oetting et al. (1995) with a modification using a 20-bp long universal M13 forward primer sequence (GTT GTA AAA CGA CGG CCA GT, designated M13) in this study. M13 was added as a common tail to the 5′ end of all SSR forward primers. All SSR primers, including regular and M13-tailed forward primers, were synthesized by Operon Technologies (Huntsville, AL, USA). Four fluorescently labeled universal M13 primers, using 6FAM, VIC, NED, and PET, were synthesized by ABI (Applied Biosystems Inc., Foster City, CA, USA) and

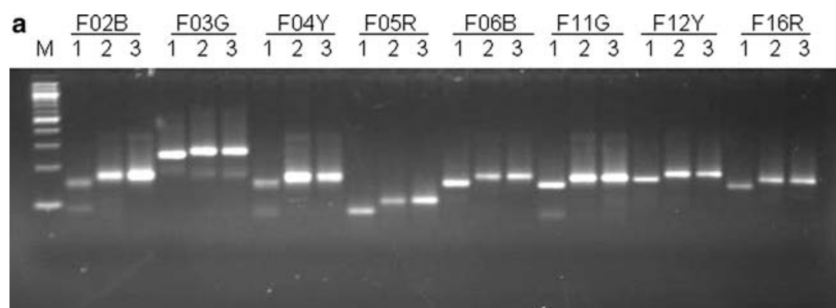
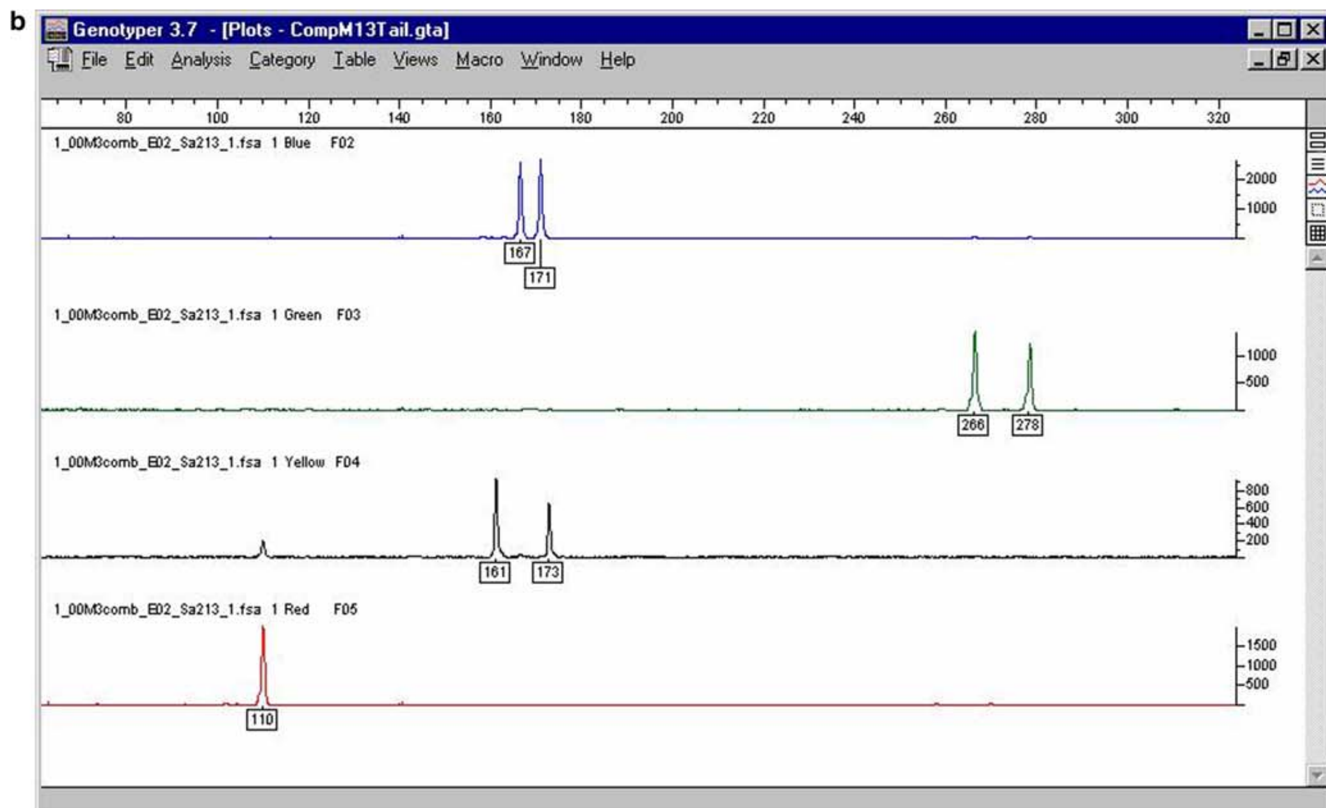


Fig. 3 Comparison of amplification from Sanford sweet orange by three combinations of eight primers. The three primer combinations are SSR reverse primer plus forward (*1*), M13-forward (*2*), and M13-forward with dye-labeled M13 added (*3*), respectively. *B*, *G*, *Y*, *R* indicate the dye 6FAM (*blue*), VIC (*green*), NED (*yellow*), and PET (*red*) labeled M13 forward primer used. All products were resolved on a SeaPlaque GTG Agarose gel (**a**), and the dye-labeled products (all of *lane 3*) were separated on an ABI 3100-*Avant* Genetic Analyzer (**b**, **c**). Primers F02, F03, F04, and F06 are heterozygous, but they appear as a single band on the GTG Agarose gel. M is 100-bp ladder (Biolabs, Beverly, MA, USA). The products in *lanes 2* and *3* are larger, because they include the 20-base M13 forward primer sequence

b



c

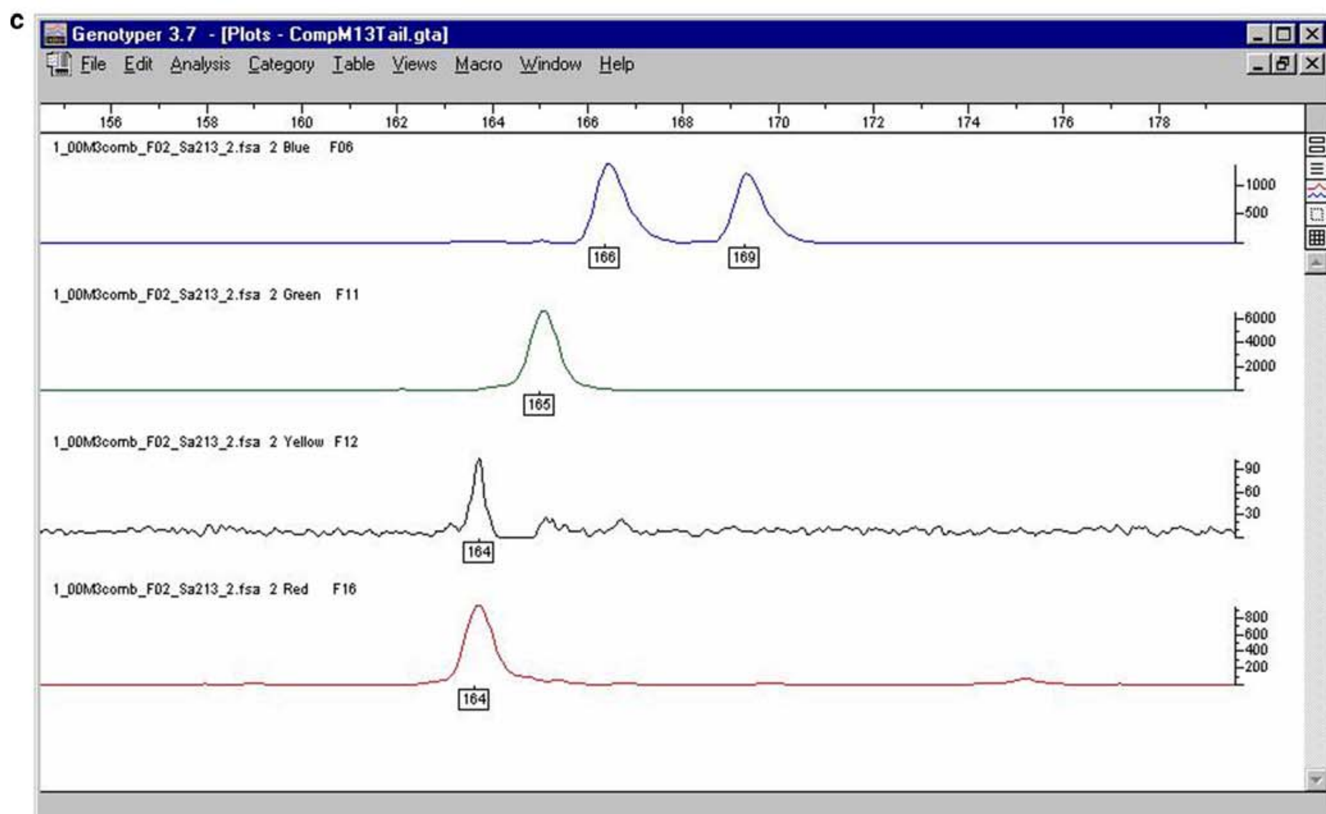


Fig. 3 (Contd.)

Table 3 Comparison of genotypes between sweet orange (SO) and *Poncirus* (Po)

Genotype	SO	Po	Ar	FD	Pt96
HM	52	67	67	70	70
HT	35	20	20	17	17

There were no differences detected among five sweet oranges, but there were among three *Poncirus* selections. Two *null* alleles are included in the HM number in *Poncirus*. HM, homozygous; HT, heterozygous

Table 4 Distribution of alleles in sweet orange (SO) and *Poncirus* (Po)

Alleles	SO–Po	Number	Pattern	Mappable in F_1
1	aa–aa	31	HM–HM	No
2	aa–bb	16	HM–HM	No
2	aa–ab	0	HM–HT	Yes
2	ab–aa	5	HT–HM	Yes
2	ab–ab	4	HT–HT	Yes
3	aa–bc	5	HM–HT	Yes
3	ab–cc	15	HT–HM	Yes
3	ab–ac	8	HT–HT	Yes
4	ab–cd	3	HT–HT	Yes
Failed		13		

HM homozygous, *HT* heterozygous

used for ABI G5 high throughput genotyping analysis on an ABI 3100-*Avant* Genetic Analyzer.

To assess the utility of the labeled M13 tail primer approach to SSR analysis on an automated sequencer in comparison with gel-based methods, three SSR primer combinations were compared; these combinations were based on using the same reverse primer respectively combined with the regular forward primer, the M13-tailed forward primer, and M13-tailed forward primer plus dye-labeled M13. PCR was performed in Bio-Rad (Hercules, CA, USA) iCyclers in 15 μl volume consisting of 1× PCR buffer, 0.2 mM dNTP, 2 mM $MgCl_2$, 0.3 μM forward primer (regular or M13-tailed), 0.3 μM reverse primer, 0.75 U *Taq* DNA polymerase (Promega, Madison, WI, USA), and 20 ng genomic DNA templates. An additional 0.05 μM dye-labeled M13 was added to the above PCR mix containing the M13-tailed forward primer, to amplify fluorescent products. A touchdown PCR program was run with an initial denaturation temperature of 94°C for 3 min, followed by 10 cycles of denaturation at 94°C for 30 s, annealing at 66°C for 30 s with a 1°C decrement each cycle, and extension at 72 °C for 45 s; this was followed by 30 additional cycles with a constant annealing temperature of 56°C (other parameters were the same), plus a final extension at 72°C for 30 min. The annealing temperatures were lowered 2–4° for re-screening the primer pairs that resulted in unstable amplification. The amplified fragments from this comparative experiment were separated on 2.5% NuSieve GTG (Cambrex Bio Science, East Rutherford, NJ, USA) or Metaphor Agarose gels (FMC Bioproducts, Rockland, ME, USA) in 1×TBE buffer for 4 h at 3 V/cm. The initial screening of all SSR primers was performed in the same manner on gels. The dye-labeled PCR products were also run on an ABI 3100-*Avant* Genetic Analyzer. GeneScan 3.7 NT and Genotyper 3.7 NT were used to extract the trace data and generate the microsatellite alleles table, respectively.

Results and discussion

Frequency and distribution of SSRs in citrus ESTs

As shown in Table 1, a total of 3,278 SSRs were identified from 2,698 EST sequences, indicating that 10.6% EST sequences had at least one SSR. Of those SSR containing ESTs, 473 (about 21.5%) had more than one SSR motif, 224 of which were considered the compound type according to the predefined criteria. Twenty-two compound SSRs were found where two SSRs were immediately adjacent to each other. On average, at least one SSR was found per 5.2 kb in the 17.0 Mb EST sequences that were searched.

The occurrences of different repeat unit size SSRs were tri- (45.1%), di- (39.4%), tetra- (8.2%), hexa- (4.6%), and penta-nucleotide (2.7%). The bi- and tri-nucleotide SSRs, a total of 2769, account for 84.5% (Fig. 2). The top 30 SSR motifs (any two complementary sequences considered one motif) were listed in Table 2, representing 65% of the SSR-containing sequences. AG/CT, AAG/CTT, AAAG/CTTT, AAAAG/CTTTT, and ACTCCG/CGGAGT were dominant in their unit length groups respectively (Table 2).

Primers were successfully designed for 2295 ESTs, but failed in the remaining 546 suitable sequences. By sorting and comparing primer sequences in Microsoft EXCEL, it was determined that these EST sequences from the contigs and singlets (considered putative unigenes) did not share any primers, except for those sequences containing more than one SSR from which the primers could be the same; therefore the corresponding SSR loci likely are unique in the citrus genome. However, those problems sequences that failed to assemble could be either overlapped or single ESTs. Primer sequences for this group were compared and found to be associated with at least 2 and up to 23 sequences (on average, 9), accounting for nearly 60% of a total of 629 problem assemblies for which primers were successfully designed. These primers were removed thus reducing the total number of SSR primers to 1960 (See the supplementary material 1).

Development of SSR markers for genetic mapping

The initial screening of 100 primer pairs (See the supplementary material 2) for assessment of the polymorphism among the four parents revealed that 65 of 94

Table 5 Allelic details of the 56 mappable SSR primers

Name	Accession ID[a]	Forward primer	Reverse primer	Size[b]	SO product	Po product	Allele	HT/HM difference
F02	BQ623425	AACAGTGTAGCATCGCACTTTCAC	GATACAAGGGACTTGCCCATCTC	155	167–171	173	ab–cc	
F03	DR912164	CGAGGATGACTCAAGTGATGAAGA	TCTTGGTCTTTGGCTTTTTCTCAG	123	266–278	272	ac–bb	
F04	CN181804	AGTGAACTGTCCATTGGATTTTCG	GTGTTGAATCCCGACCTTCTACC	155	161–173	155	bc–aa	
F06	CK933007	TTCATTTGGAACAAAACCCAATTC	GCTGCTAATCACAGCATCAAGAGA	151	166–169	161	bc–ab	Ar: HT (161–167)
F07	DR908935	CTGTTACCGTTGAGGAAACCAAAG	CTCTTCAGCTGGTTTCTCTTCCTG	98	102–114	102–108	ac–ab	Pt96: HM (108)
F09	CF417302	AGCAGTTGTCATCTTCGGTCAGTT	ATAACCAGCAGGACCAGTCACAAT	155	157–174	162–174	ac–bc	
F10	DN958537	CCTCAGCTCTAGCAAAAGCACATT	AGAGGCTATAGATCGTGGATGCAG	160	171–177	171–177	ab–ab	Ar: HT (171–183)
F11	DN795635	AAAAATCGAAATCGAGCACCC	GAAGTAACGGAGAATTCCGATGAG	152	164	170	aa–bc	Ar: HT (170–176)
F13	CK933215	AAACCCAAGTCATAAACGTCAGGA	ATCTTCAATGCTTTTGGAGCAAAC	158	172–178	178	ab–bb	
F14	CX296884	GCTCCTCGAATGAGAATGAAATGA	TGGTTGTGCGAAAATGAAGAGATA	109	99–105	122–128	ab–cd	Ar: HM (128)
F15	CO912711	TGTGTCAAACCATCACCCTG	TGCTGTCACGTTGGGTGTAT	138	159–171	165	ac–bb	Pt96: HT (159–165)
F16	DR909803	GTCTTCACCCTCTCCATCTTCATC	GGGACTATGGCAACAATAACTCCA	149	161–173	167	ac–bb	Pt96: HT (161–167)
F17	CV710821	GATACAAATTAGCATTTGATTGAATGGA	ATCGGGACTCGCATTAGGGT	116	131–146	128–131	bc–ab	Ar: HM (131)
F18	CF830946	GTCTTCAACGAAGTTGCAGGCT	TACTATTTCGAGAGAGCAGCAGCA	141	153–159	158	ac–bb	
F19	DR404147	ATTCTCATGTCATGCGTACCTCG	TGAATCGTGAGAGACGAGTTGAAG	150	149–167	149	ab–aa	
F20	DN617970	CTACATCTCAGACTCCTGCACCAA	ACACATCCGATACTGAAGAGGCTG	147	162	168	aa–bb	
F21	DN621499	CTACAAGTTCCCCAGTTATCCCG	ACTTGACCCGCTCTAGGAGTGAC	140	145–151	151–158	ab–bc	Pt96: HM (151)
F23	CF417259	TTTTAGGAAATTTCAATCACCCGA	TCATTGTCATCACCATCACCATC	160	174–179	161	bc–aa	
F25	CN186475	TGGAGGTTGCTCCAAAATTAGTGT	GATGGTAGCAATCCTACTCCCTTG	154	173	161	aa–bb	
F27	CX544885	AGGAGCTATCATCTGAAGGGGACT	CATCGTCTTCCTCTTCACCATTCT	147	162–174	168	ac–bb	Pt96: HT (162–168)
F28	CD575748	ACAACATTTCAGGAATCCCACAAC	ATTCGGAGTGTTGCCAGAGAGTAG	105	102	120	aa–bb	
F29	CX664921	TTCACCACAAACGAAGACTCAGAC	CTGTAATCCACTCGGTAATCCGAC	131	150–153	145	bc–aa	
F30	CF417515	TAGGTTCTCTTTCAACCCCCTTTC	CTGCTTCGGCTGTAATTGTGATT	85	92–104	98	ac–bb	
F31	CF507648	CGAAGAAGAATTGAAAGAGCCAGA	CAACAGATTTGTTACTGGAAGGGG	137	154	141–148	cc–ab	
F32	CF506971	ACCCTCTCTTAGTGGGCAGTCTTT	ACAAAGAAAGGGATTGTGACGGTA	146	162–168	163	ac–bb	
F33	CX637978	ATTCGGAGTGTTGCCAGAGAGTAG	ACAACATTTCAGGAATCCCACAAC	105	105	123	aa–bb	
F34	CD574567	ACTTATTTTGTTCTTGCCCTCGGT	CGTAAGCATTATTGTTGACGGGTT	102	142	118	aa–bb	
F38	CN188007	TAGCTGCTGCCATAATTTTTCACA	AAAAGGGTTGATTTTTGTTCCCTC	130	138–149	138	ab–aa	
F39	CV885396	TTTGTTTGTTACAAGGCAAGGCA	TACAAGTTGTGAGGGGGTGAATTT	143	154–164	159	ac–bb	
F40	CV716690	TGTTTGTTACAAGGCAAGGCAAG	CTGCAGAGGAGCTTCCTTACTCAA	81	86–96	91	ac–bb	
F41	CD576138	TCTCTTGTTCTCTTCTTTCCCCAA	AACTGGCTTGGAGAGAATCCATTA	160	195	172–177	cc–ab	Pt96: HM (177)
F42	CD576630	CCCATTTAAATCTCAAAATTCGCA	TTAAGACTGGCAAGAATCGTTGGT	153	164–169	164–169	ab–ab	
F43	CN181994	ACAAAAACAAAACAGGAGAGCACA	ATTTCTCCACAGTGAAAGGAGCTG	107	112	122	aa–bb	
F45	CK939690	AATAACCCCCTTCAACCTCAAAAA	ATTAGTTCTTCGTCAGTTGGGTGG	148	165	142	aa–bb	
F46	CD575672	TACATCCCCTTGCACCAAGACTAT	AGCTAAAGAATGGACCCTTTTCCA	145	157	162	aa–bc	Ar: HT (162–172)
F49	CN253989	GGGGAACTCAAACTTCTCAAGGG	AGCTCATACTCAAGCTCCCGTCTT	94	109	114	aa–bb	
F50	CN183579	AGGACGATTTCAGGTTCAGACTTG	GAGCAAAAATCGAAGAGATCAAGC	142	158–163	null	ab–cc	
F51	DR911460	AAGAATGGACCCTTTTCCAGTAGG	TACATCCCCTTGCACCAAGACTAT	140	158	163	aa–bc	Ar: HT (163–173)
F52	BE205735	CGTATAATAAGCATATCAGCAGCGG	ATCGTACGACATGTTGTTGTTGCT	145	93–101	93–101	ab–ab	
F53	CX078531	CTCCATTACCACAACACTCACTCC	GACGAAGAAGAAAGAAAGGGAAGG	151	157–167	167–172	ab–bc	
F56	CD575772	CCATTAACGAGAAAACCAAACACTG	GGGGTTGCAAAGAAATGTAAAGAA	160	166–176	180–185	ab–cd	
F57	CF828502	CCTCGCCAATGACCTTTGTATTTA	CAATACGTTTGGGTTCTAGTTCCG	143	152	161	aa–bb	
F58	CX298927	GGGTTCTTCGTTTCTGTTTTGTTG	TCCCTTTGGTAGCATACGACTCTT	120	136	146	aa–bb	
F61	CD574034	TCCGTCCTTGTTTTAACTTGGAAA	GCTCCGTAGGTTGAAGAAAGATGA	93	109	112	aa–bb	
F65	CF420759	CCAGAGCAAAACATAGAAAGAGCG	ACTCTAGCTCTGCTCCCTTCATCA	129	145–148	136–144	cd–ab	
F67	BQ623531	CAGAAATTGATTAAACAGAGTGAGATCC	CGAAGAAGCTAAAACCAAGAGAGC	117	124–133	124	ab–aa	
F72	CF831383	TAACTGAGAAAACAAATGGCCACA	ACTCCTTGCCACTGCTTTATCAAC	160	179–183	179–183	ab–ab	FD: HM (179)
F77	CK933430	CGTAATTGTAATTGCGAAACACCG	CTTGAAGAGATCCGTGTTCGAGAG	100	111–119	111–123	ab–ac	
F79	BQ624534	GTGGCTGCAAGACTTCTTGTGTAG	TATAGATGTGGCAAGCAGAGCAAG	153	171	175	aa–bb	
F81	CK939975	CCCGTTAAAATTGTAAACCCACAG	TTTATAATACGGAACGTTGGGAGG	85	100–103	87	bc–aa	

Table 5 (Contd.)

Name	Accession ID[a]	Forward primer	Reverse primer	Size[b]	SO product	Po product	Allele	HT/HM difference
F87	DR909369	ATGAAGGCTTTTAGAGCCGAGTT	ATAATAGGGGCCCACTTGACTTG	121	137–142	142–146	ab–bc	Pt96: HM (142)
F88	CN189762	GTTCGCTCCACGCGATTTAT	TGTGAAGAGAAAGATTTGGTGGGTTT	85	104–108	104	ab–aa	
F89	CK665531	AAACACACGGAGAGTTTTGGAGAA	ATCCACCACCTGTACCTCCTGTAA	90	108	104	aa–bb	
F90	CB292130	CACCGCCAGCGGTATGTAAT	GTCCAACCCAACCCAATAAAATC	157	175	null	aa–bb	
F97	CK933638	CTTCTTCTCCTGCTCCTCCTC	AGTGAGAAGCCAAAACACCAAAC	93	113	103	aa–bb	
F98	CX545856	ATATAGCCCTAGGCTCCCTATC	GAGTAACCATGGGAGGAGAAAGGA	159	176	180	aa–bb	

[a]Only one accession number in each contig was listed for reference

[b]The size is expected by Primer3. The actual size of the amplified allele products from SO and Po are indicated in the next two columns

SO sweet orange, *Po Poncirus*, *FD* Flying Dragon, *Ar* Argentina, *Pt96* DPI 0906, *HM* homozygous, *HT* heterozygous

well-amplified fragments presented polymorphisms of length or/and heterozygosity, and 43 appeared heterozygous in at least one parent. Heterozygous and homozygous differences in the products of eight primers were found between Flying Dragon and Argentina, but no differences were found between the two sweet oranges. The amplified products of three primers, F03, F47, and F99, were much larger than expected, suggesting that an intron is inside these amplicons. Unfortunately, it was found too difficult to score segregation reliably and accurately in the population on GTG Agarose gels. The size difference among the alleles at a locus usually was less than 10 bp, not large enough to yield sufficient and unambiguous resolution on the Agarose gels. A subsequent comparison indicated that any "single band" wider and brighter than usual possibly consisted of two unresolved heterozygous fragments. For instance in Fig. 3, the products of primers F02, F03, F04 and F06 were resolved as a single band on a SeaPlaque GTG Agarose gel (Fig. 3a), but they were truly determined as heterozygous in individuals when tested on the ABI 3100-*Avant* Genetic Analyzer (Fig. 3b, c). The added M13 tail appeared to have little influence on PCR performance of most primers (Fig. 3a). However, multiple peaks (more than two) were amplified from 7 of the M13-tailed primers, from which the true alleles were not easily determined. These seven SSRs were clearly amplified and easily scored when using their regular primers. A locus having more than two peaks amplified by the M13-tailed primers may belong to a gene family whose multiple members possess high identity; the M13 tail might interfere with the precise annealing of the primer to the true target.

The results from 87 primers used with eight sweet orange and *Poncirus* parents and eight F_1 offspring, when run on the ABI 3100-*Avant* Genetic Analyzer, revealed 52 homozygous (HM) and 35 heterozygous (HT) loci in sweet orange, and 67 HM and 20 HT in *Poncirus*, respectively. Calculated from the 87 successful primers, 40% of the sweet orange loci were heterozygous and 23% were heterozygous in *Poncirus*; the greater frequency of heterozygosity in the former is consistent with the presumed hybrid origin of sweet orange (Table 3). Pairwise comparisons of the loci between two intergeneric species showed that 47 were homozygous in both species with either one allele (the number of primers is 31) or two (16); 40 loci were heterozygous in at least one species with two alleles (9), three alleles (28), or four alleles (3) (Table 4). In addition, fourteen homozygous/heterozygous differences were found among three *Poncirus* biotypes, Argentina (Ar), Flying Dragon (FD), and DPI0906 (Pt96), but none were found among five sweet orange parents (Table 5). There are three pairs of alleles, from primers F18, F32, and F65, that had only one base difference presented in the polymorphisms. Such two alleles, if presented in heterozygous individuals, cannot be labeled automatically and scored correctly by Genotyper; the unlabeled allele has to be manually input (Fig. 4).

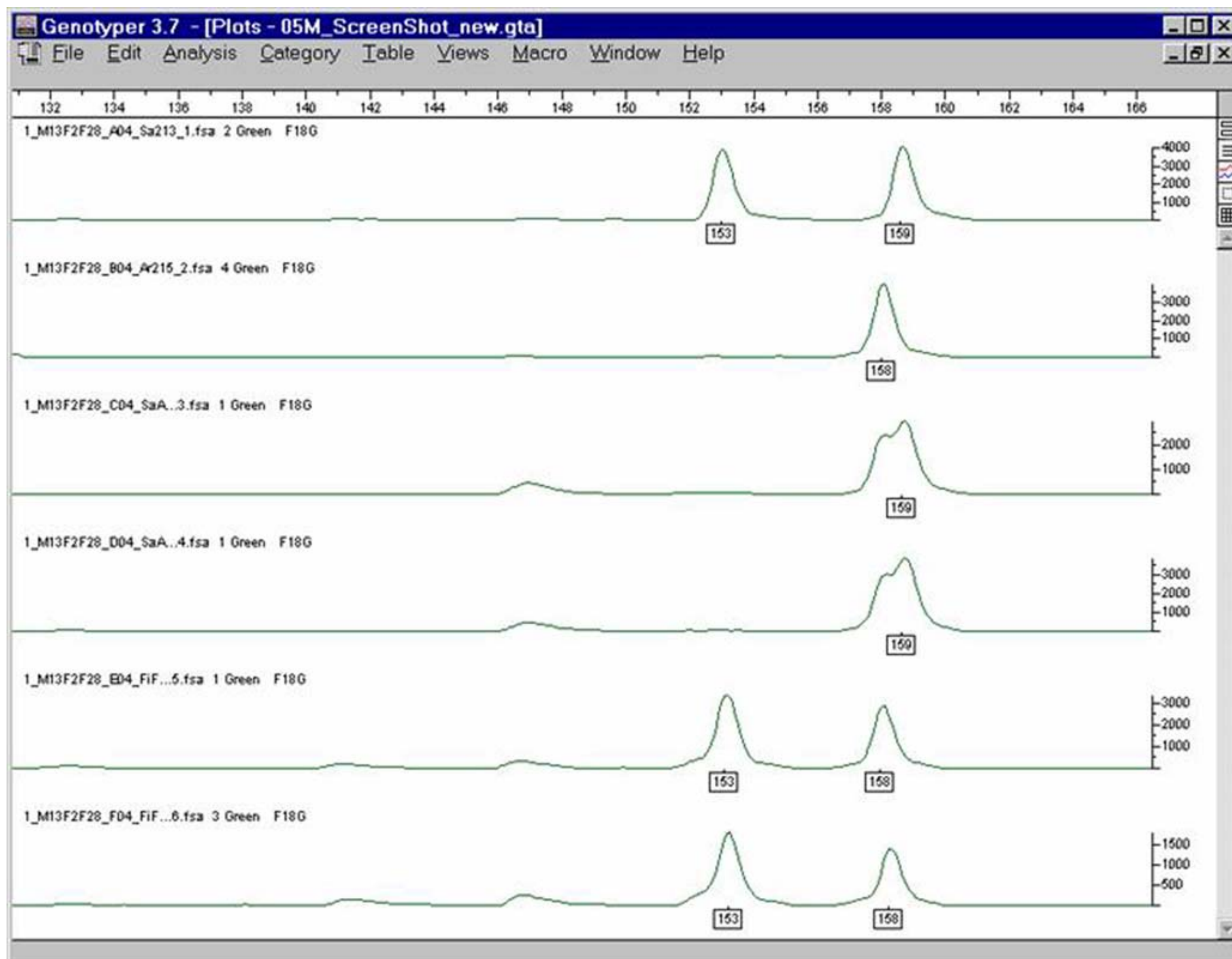


Fig. 4 One SSR locus, F18, showing two alleles with a one-base difference. The samples were Sanford sweet orange, Argentina trifoliate orange, and four F_1 individuals. The segregation of alleles was obvious in the F_1 individuals, but Genotyper could not distinguish or correctly label the two peaks with one-base difference

According to the results from these 100 primers, the percentage of polymorphism between sweet orange and *Poncirus* among the EST SSR loci is about 65%. However, those polymorphic SSRs that are homozygous in both species cannot be mapped in citrus F_1 populations that are frequently used for genetic mapping of highly heterozygous woody trees, although they should be useful for mapping in F_2 or backcross populations. Therefore the estimated number of SSRs that can be mapped in F_1 populations between sweet orange and *Poncirus* is reduced to about 40% of the total of 1960 primers, which equals 784 EST SSR markers. Compared with previous approaches to SSR discovery used for citrus, the increased number of SSRs found from ESTs will greatly accelerate the citrus genetic mapping project that is under the way, and the availability of this resource will also be of substantial benefit to other genetic applications and analyses, such as germplasm characterization, screening of zygotic seedlings among nucellar seedling populations, and developing markers for marker-assisted selection (MAS) in citrus breeding programs.

Acknowledgment The authors would like to thank Dr. Thomas Thiel and Dr. Lisa Davis for their kind answers to the questions about using the misa Perl scripts for SSR search, and implementing microsatellite analysis on ABI 3100 machine, Mr. Allan Burrage and Mr. Gary Wilhite for their enthusiastic assistance to setup Linux in the VMWare virtual machine, and Mrs. Margie K. Wendell for her excellent technical support. This research was supported in part under Project Number 0110-03I of the Florida Citrus Production Research Advisory Council, and by a grant from USDA-CSREES-NCRI 2002-34399-12765. Florida Agricultural Experiment Station Journal Series Number is R-11027.

References

Ahmad R, Struss D, Southwick SM (2003) Development and characterization of microsatellite markers in citrus. J Am Soc Hortic Sci 128:584–590

Aldrich KJ, Cullis CA (1993) RAPD analysis in flax: optimization of yield and reproducibility using KlenTaq1 DNA polymerase, chelex 100 and gel purification of genomic DNA. Plant Mol Biol Rep 11:128–141

Cai Q, Guy CL, Moore GA (1994) Extension of the linkage map in citrus using random amplified polymorphic DNA (RAPD) markers and RFLP mapping of cold-acclimation-responsive loci. Theor Appl Genet 89:606–614

Castelo AT, Martins W, Gao GR (2002) TROLL—tandem repeat occurrence locator. Bioinformatics 18:634–636

Chani E, Ashkeni V, Hillel J, Veilleux RE (2002) Microsatellite marker analysis of an anther-derived potato family: skewed segregation and gene-centromere mapping. Genome 45:236–242

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci 160:1115–1123

Cregan PB, Kollipara KP, Xu SJ, Singh RJ, Fogarty SE, Hymowitz T (2001) Primary trisomics and SSR markers as tools to associate chromosomes with linkage group in soybean. Crop Sci 41:1262–1267

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Fang DQ, Federici CT, Roose ML (1998) A high-resolution linkage map of the citrus tristeza virus resistance gene region in *Poncirus trifoliata* (L.) Raf. Genetics 150:883–890

Gmitter Jr FG, Xiao SY, Huang S, Hu XL, Garnsey SM, Deng Z (1996) A localized linkage map of the citrus tristeza virus resistance gene region. Theor Appl Genet 92:688–695

Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8:195–202

Gordon D, Desmarais C, Green P (2001) Automated finishing with Autofinish. Genome Res 11:614–625

Hackauf B, Wehling P (2002) Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breeding 121:17–25

Hamada H, Kakunaga T (1982) Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature 298:396–398

Kijas JMH, Fowler JCS, Thomas MR, Scott NS (1995) An evaluation of sequence tagged microsatellite site markers for genetic analysis within citrus and related species. Genome 38:349–355

Kijas JMH, Thomas MR, Fowler JCS, Roose ML (1997) Integration of trinucleotide microsatellites into a linkage map of citrus. Theor Appl Genet 94:701–706

Ling P, Yu C, Deng Z, Chen C, Huang S, Wendell MK, Gmitter Jr FG (1999) Citrus genome mapping with AFLP markers. Plant & Animal Genome VII, San Diego, CA, Poster 189, 17–21 January 1999

Ling P, Duncan LW, Deng Z, Dunn D, Xu X, Huang S, Gmitter FG (2000) Inheritance of citrus nematode resistance and its linkage with molecular markers. Theor Appl Genet 101:1010–1017

Lowe AJ, Moule C, Trick M, Edwards KJ (2004) Efficient large-scale development of SSRs for marker and mapping applications in *Brassica* crop species. Theor Appl Genet 108:1103–1112

Luro F, Laigret F, Lorieux M, Ollitrault P (1996) Citrus genome mapping with molecular markers: two maps obtained by segregation analysis of progeny of one intergeneric cross. Proc Int Soc Citricult 2:862–866

Oetting WS, Lee HK, Flanders DJ, Wiesner GL, Sellers TA, King RA (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. Genomics 30:450–458

Paniego N, Echaide M, Muñz M, Fernádez L, Torales S, Faccio P, Fuxan I, Carrera M, Zandomeni R, Suáez EY, Hopp HE (2002) Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L.). Genome 45:34–43

Paris M, Jones MGK (2002) Microsatellite genotyping by primer extension and MALDI-ToF mass spectrometry. Plant Mol Bio Rep 20:259–263

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

Robinson AJ, Love CG, Batley J, Barker G, Edwards D (2004) Simple sequence repeat loci discovery using SSR Primer. Bioinformatics 20:1475–1476

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S and Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386. The source code is available at http://www.fokker.wi.mit.edu/primer3/

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

Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia-Hancock S, Musket T, Duru N, Polacco M, Edwards K, Ruff T, Register JC, Brouwer C, Thompson R, Velasco R, Chin E, Lee M, Woodman-Clikeman W, Long MJ, Liscum E, Cone K, Davis G, Coe EH (2002) Development and mapping of SSR markers for maize. Plant Mol Bio 48:463–481

Simone M de, Russo MP, Puelo G, Marsan PA, Lorenzoni C, Marocco A, Recupero GR (1998) Construction of genetic maps for Citrus aurantium and C. latipes based on AFLP, RAPD and RFLP markers. Fruits 53:383–390

Supplementary material 1. 1960 primer-successful citrus EST SSR motifs and accession numbers

Supplementary material 2. 100 tested EST SSR primers

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

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422. The source code is available at http://www.pgrc.ipk-gatersleben.de/misa/misa.html