Theor Appl Genet (2004) 108:619–627
DOI 10.1007/s00122-003-1467-x

ORIGINAL PAPER

J. M. Lee · S. H. Nahm · Y. M. Kim · B. D. Kim

Characterization and molecular genetic mapping of microsatellite loci in pepper

Received: 6 February 2003 / Accepted: 18 August 2003 / Published online: 29 November 2003

Abstract Microsatellites or simple sequence repeats are highly variable DNA sequences that can be used as informative markers for the genetic analysis of plants and animals. For the development of microsatellite markers in *Capsicum*, microsatellites were isolated from two small-insert genomic libraries and the GenBank database. Using five types of oligonucleotides, $(AT)_{15}$, $(GA)_{15}$, $(GT)_{15}$, $(ATT)_{10}$ and $(TTG)_{10}$, as probes, positive clones were isolated from the genomic libraries, and sequenced. Out of 130 positive clones, 77 clones showed microsatellite motifs, out of which 40 reliable microsatellite markers were developed. $(GA)_n$ and $(GT)_n$ sequences were found to occur most frequently in the pepper genome, followed by $(TTG)_n$ and $(AT)_n$. Additional 36 microsatellite primers were also developed from GenBank and other published data. To measure the information content of these markers, the polymorphism information contents (PICs) were calculated. *Capsicum* microsatellite markers from the genomic libraries have shown a high level of PIC value, 0.76, twice the value for markers from GenBank data. Forty six microsatellite loci were placed on the SNU-RFLP linkage map, which had been derived from the interspecific cross between *Capsicum annuum* "TF68" and *Capsicum chinense* "Habanero". The current "SNU2" pepper map with 333 markers in 15 linkage groups contains 46 SSR and 287 RFLP markers covering 1,761.5 cM with an average distance of 5.3 cM between markers.

Communicated by J. Dvorak

J. M. Lee · S. H. Nahm · Y. M. Kim · B. D. Kim (✉)
Department of Plant Science,
College of Agriculture and Life Sciences
and Center for Plant Molecular Genetics and Breeding Research,
Seoul National University,
Seoul 151-742, Korea
e-mail: kimbd@snu.ac.kr
Tel.: +82-2-880-4933
Fax: +82-2-813-5410

Present address:
S. H. Nahm, Biotechnology Center,
Nong Woo Bio Co., Ltd,
537-17 Jeongdan, Kanam, Kyonggi 469-880 Yeoju, Korea

Present address:
Y. M. Kim, Kumho Life and Environmental Science Laboratory,
1 Oryong-dong, Buk-gu, 500-712 Kwangju, Korea

Introduction

DNA markers are used extensively to create genetic and physical genome maps and for many basic and applied purposes. Microsatellite markers are particularly useful because of their high variability and ease of detection by the polymerase chain reaction (PCR).

Microsatellites are DNA sequences with short (less than 6 bp) repeated motifs (Litt and Luty 1989). Microsatellites are also called simple sequence repeats (SSRs) (Jacob et al. 1991) and short tandem repeats (STRs) (Edwards et al. 1991). They are abundant in most eukaryotic genomes and, generally, are uniformly distributed along a genome. SSRs are highly variable and co-dominant; therefore, their information value is high compared to other markers, such as RFLPs, AFLPs and RAPDs. The regions flanking SSRs are generally conserved among genotypes of the same species. PCR primers flanking the microsatellite regions are used to amplify the SSR-containing DNA fragments. Length polymorphism occurs when PCR products from different alleles vary in length as a result of variation in the number of repeat units in an SSR. This variation in repeat number is caused by slippage during DNA replication (Schlotterer and Tauz 1992) or unequal crossing-over between sister chromatids (Innan et al. 1997). Most SSR loci are defined by a unique pair of primers, which facilitates information exchange between laboratories.

Several studies have been conducted using molecular markers to assess the level of variation among *Capsicum* species (Prince et al. 1995; Paran et al. 1998; Rodriguez et al. 1999), to construct molecular linkage maps (Lefebvre et al. 1995; Livingstone et al. 1999), and to clone or tag genes related to useful traits (Tai et al. 1999; Huh et al.

2001). These studies employed RFLP, AFLP and RAPD molecular markers. Nagy et al. (1998) and Huang et al. (2000) reported the development of microsatellite markers from the pepper small-insert genomic library or EMBL database. A pepper molecular linkage map containing 150 RFLP and 430 AFLP markers was recently developed (Kang et al. 1997, 2001). Livingstone et al. (1999), previously reported a pepper map mostly utilizing tomato RFLP clones.

The objective of this study was to develop polymorphic *Capsicum* microsatellite markers, to characterize them, and integrate them into an existing molecular map of pepper.

Materials and methods

Plant materials and DNA extraction

For estimation of Polymorphic Information Content (PIC), 11 pepper lines were kindly provided by SeminisKorea. DNA was extracted from the following plants, *Capsicum baccatum*, *Capsicum frutescense*, *Capsicum chacoense*, *Capsicum pubescence*, *Capsicum chinense* cv PI159236, *Capsicum chinense* cv Habanero, *Capsicum annuum* cv TF68, *C. annuum* cv AC2258, *C. annuum* cv CM334, *C. annuum* cv Chilsungcho, and *C. annuum* cv ECW.

A total of 107 F_2 plants derived from the interspecific cross *C. annuum* cv TF68 × *C. chinense* cv Habanero were used as a mapping population. DNA was extracted from leaf tissues of each individual plant following a method described in Kang et al. (2001).

Library construction and screening

Two small-insert genomic libraries were constructed. Genomic DNA from *C. annuum* cv AC2258 was digested with *Hin*dIII, and restriction fragments (400–700 bp) were ligated into a *Hin*dIII-digested and de-phosphorylated plasmid vector, pBluescript-SK+ (Stratagene, USA).

Alternatively, genomic DNA from *C. annuum* cv TF68 was predigested with *Pst*I, electrophoresed on a 0.8% agarose gel and 2–10-kb-long restriction fragments were isolated using the QI-Aquick purification kit (Qiagen, Germany). The size-selected DNA was further digested with *Mbo*I and the restriction fragments with a mean insert size of 500 bp were cloned as described above.

The ligation of plant DNA and vector DNA was performed in a total volume of 20 μl, with 8 μl each of the gel-purified DNAs (approximately 80 ng of each DNA) and ten units of T4 DNA ligase (Promega, USA) at 16°C for 16 h. The ligation mix was used to transform MAX Efficiency DH5α competent cells (Life Technologies, USA), using an electroporator (Bio-Rad, USA) and plated on 150-mm culture plates containing LB broth, 1.8% agarose, ampicillin (50 mg/Ml), X-gal (20 mg/Ml) and IPTG (200 mg/Ml).

Colonies were blotted onto nylon membranes (Hybond N+, Amersham Pharmacia Biotech, UK), which were soaked in 10% SDS for 3 min, denaturated with 0.5 N of NaOH, 1.5 M of NaCl for 5 min, neutralized with 1.5 M of NaCl, 0.5 M of Tris-HCl (pH 7.4) for 5 min, fixed in 2×SSC for 5 min and dried on 3 MM paper for 30 min. DNA on the membrane was immobilized by UV irradiation for 2.5 min. Membranes were washed with 5×SSC, 0.5% SDS and 1 mM of EDTA (pH 8.0) for 30 min at 50°C, and cellular debris was removed with Kimwipes (Sambrook et al. 1989).

Probes were labelled in a total volume of 20.5 μl with 1 μl of 10 pmol/μl oligonucleotides, 5 μl of [γ-^{32}P]ATP (3000 Ci/mmol, Amersham Pharmacia Biotech, UK) and ten units of T4 polynucleotide kinase (New England Biolabs, USA) at 37°C for 1 h. Oligonucleotides $(AT)_{15}$, $(GT)_{15}$, $(GA)_{15}$, $(ATT)_{10}$ and $(TTG)_{10}$ were used as probes. Hybridization was performed according to Sambrook et al. (1989) and membranes were washed to a stringency in 0.5×SSC, 0.1% SDS at 65°C for GT, GA and TTG or at 60°C for AT and ATT.

DNA sequencing and sequence analysis

Positive clones were sequenced using a ThermoSequenase 2.0 kit (Amersham Pharmacia Biotech, UK) on an ABI 377 DNA sequencer (Applied Biosystems, USA). Homology search was performed in the EMBL database.

Pepper sequences in the GenBank database were searched for the presence of microsatellite sequences longer than 12 bp. Primers were selected using PRIMER 3.0 (http://-genome.wi.mit.edu/, Whitehead Institute for Biomedical Research, USA). These primers for flanking region of microsatellite sequences were synthesized commercially (Bioneer, Korea).

PCR amplification

PCR amplifications were performed in a 25-μl volume containing 10–25 ng of genomic DNA as templates, microsatellite primers, 1 unit of *Taq* polymerase (Takara, Japan) and 10× buffer solution. Forward and reverse primers of 10 pmol/μl were labelled with [γ-^{32}P]ATP using T4 polynucleotide kinase.

All SSR amplifications were performed in thermocycler (Hybaid, UK). After 3 min at 94°C, 35 cycles were performed for 1 min at 94°C, 1 min at either 50 or 55°C, 2 min at 72°C and a final extension step for 10 min at 72°C.

PCR products were analyzed on 6% denaturing sequencing gel in 1×TBE buffer. After electrophoresis, the polyacrylamide gel was dried at 80°C on Whatman 3 MM paper in a gel dryer (Bio-Rad, USA) for 2 h and autoradiographed.

Estimation of polymorphic information content

The SSR allele composition of each analyzed pepper line was determined, and the polymorphic information content (PIC) of each microsatellite marker was calculated according to the formula

$$\mathrm{PIC} = 1 - \sum_{i=1}^{k} pi^2 ,$$

where pi is the frequency of the ith allele, and k is the total number of different alleles at the locus.

Map construction

Pepper, tomato and tobacco clones were used as probes for the construction of an RFLP map. Pepper clones were described in Kang et al. (2001) and CDI clones were kindly provided by D. Choi (KRIBB, Korea). Tomato clones were supplied by S. Tanksley and M. Jahn (Cornell University, USA), and tobacco clones were provided by D. Choi. Pepper defence-related genes (*CaPR*, *CaLTP* and *N32*) were kindly provided by K.-H. Paek (Korea University, Korea).

Linkage analysis of RFLP and SSR loci was performed using MAPMAKER 3.0/EXP (Lander et al. 1987). To identify linkage groups, pairwise comparisons and the grouping of markers were performed using the "Group" command at a maximum recombination fraction of 20 cM and a minimum LOD score above 4.0. To establish the most likely order within each linkage group, the "order" command was used and the remaining markers were added into a frame map using the "try" command. The order of markers was confirmed using the "ripple" command. Alternatively, a framework of markers was generated using the "compare" command and the best order was confirmed. Recombination fractions were converted to map distances in centiMorgans (cM) using the Kosambi mapping function (Kosambi 1944).

Results

Isolation and characterization of pepper microsatellites

About 20,000 clones from two small-insert genomic libraries with a mean insert size of 500 bp were screened by colony hybridization for the presence of $(AT)_n$, $(GA)_n$, $(GT)_n$, $(ATT)_n$ and $(TTG)_n$ motifs. Out of 130 positive clones isolated, sequence data identified 95 clones (73%) having microsatellite motifs, among which 18 clones were duplicates. Thus, a total of 77 non-redundant microsatellite clones were isolated. The average repeat length was 17.6 ranging from 6 to 36. Thirty eight clones had homogeneous repeats and 39 clones had compound repeats. Based upon sequence data, the 77 microsatellite clones could be categorized into 25 (27%) $(GT)_n$, 24 (25%) $(TTG)_n$, 21 (22%) $(GA)_n$, 20 (21%) $(AT)_n$ and 5 (5%) $(ATT)_n$ motifs, resulting in a total of 95 microsatellite motifs. Increase in the number of motifs was caused by double-counting of the heterogeneous repeats such as $(TG)_{23}(AG)_9$ of *Hpms*1-62.

The frequency of 95 microsatellite motifs found in these two small-insert libraries suggests that there are 25,700 microsatellites in the whole pepper genome or one microsatellite every 95 kb of the pepper genome. This calculation is based upon the fact that the two libraries of 20,000 clones with an average size of 500 bp per clone represent about 0.37% of the pepper genome, which is estimated to be about 2,702 Mb (Arumuganathan and Earle 1991).

To determine whether sequences of 77 microsatellite clones had any biological significance, the sequences were also used to search the EMBL database. Eleven (14%) of the microsatellite flanking sequences showed significant homology with repetitive elements at least at one end (data not shown). Homology was found to retrotransposon, transposon and the short interspersed repetitive element (SINE).

Capsicum microsatellite sequences in GenBank

GenBank database was searched for DNA sequences of *Capsicum*. A search of 213,061 bp from 312 *Capsicum* entries (until 16th October 1999) yielded 32 microsatellite motifs, but the majority (90.6%) were less than 20 bp long (data not shown). The cut-off was set as low as six repeat units for di- and tri-nucleotide repeats, and four repeat units for the tetranucleotide repeats (Smulders et al. 1997). GenBank sequence data suggest that microsatellites occur at the rate of 1 per 6.7 kb and that there are about 405,818 microsatellite sites in the *Capsicum* genome. (A)*n*, (AT)*n* and (AG)*n* sequences are the most frequent microsatellites, occurring at 1 per 21.3 kb, 1 per 53.3 kb and 1 per 53.3 kb, respectively, in the *Capsicum* genome.

Marker development

Isolated pepper microsatellites were further examined for marker development. Out of the 77 microsatellite clones, 29 clones were unsuitable for designing primer sequences because of partial sequence data (19 clones), the location of microsatellite sequences near the cloning site (eight clones) and a sub-optimal melting temperature detected during PCR primer selection (two clones). The remaining 48 clones, containing complete and suitable microsatellite sequences, produced products ranging from 100 to 300 bp. Primers that amplified products from 11 *Capsicum* accessions were given serial numbers following the designation *Hpms* (Table 1). Out of the 32 microsatellite motifs from GenBank accessions 26 primers were developed (Table 1). Out of the pepper gene clones from our laboratory a microsatellite motif was found in *CaSIG*19 (Lee 2000) at the intron, and in *hpMADS* (Yu 2001), at 3′ UTR, the polymorphism in these microsatellites could also be used for mapping and PIC calculation.

PIC values were computed for 76 markers including eight published markers (Huang et al. 2000). The average PIC value was 0.57 with a maximum of 0.91 for *Hpms*1-111 and a minimum of 0.36 for *Hpms*1-43. An average of 4.8 alleles were detected. The average number of allele-derived microsatellites was 6.3 in libraries and 3.2 in GenBank. The average PIC of microsatellites was 0.75 from libraries and 0.38 from GenBank.

Molecular linkage mapping

A total of 107 F_2 plants derived from an interspecific cross between *C. annuum* "TF68" and *C. chinense* "Habanero" (Nahm et al. 1997) was used to integrate those 76 SSR markers onto the existing pepper SNU-map (Kang et al. 2001). For efficient use of the gel, and time in the mapping process, three markers were analyzed together on one 6% acrylamide gel by loading the PCR products of 107 F_2 individuals in 10–30 min intervals (Fig. 1).

Of the 76 SSR markers, only 46 were successfully mapped (Fig. 2); 32 loci (from 29 primers) out of 40 markers (72.5%) from genomic libraries and 14 loci (from 13 primers) out of 36 markers (36.1%) from GenBank showed polymorphism between the parental lines. Except for three, each SSR marker was mapped at a single locus. *Hpms*2-26 was located on three loci. *Hpms*1-1 and AF244121 were located on two loci. In the case of *Hpms*CaSIG19, the RFLP marker was also available and mapped on the same locus as the SSR. The same was the case for *Hpms* hpMADS. SSR markers were relatively evenly distributed among linkage groups (LGs) compared to RFLP and AFLP markers, which tend to cluster. Some exceptions to this observation were noted; especially, 19 SSR markers were in one linkage group (LG1), whereas LGs 2, 12, 13, 14 and 15 did not contain any SSR markers. In the lower telomeric region of LG1 four SSR markers appeared in a cluster; however, intervals between these markers were more than 20 cM. Of the 46 SSRs, 41

Table 1 Description of hot pepper microsatellite markers

Locus	Repeat	Primer		Size (bp)	Condition[a]	No. of alleles	PIC	LG[b]
Hpms 1–1[c]	$(CA)_{12}(TA)_4$	L[d]: tcaacccaatattaaggtcacttcc	R[e]: ccaggcggggattgtagatg	283	50, 30	10	0.89	1
Hpms 1–3	$(AT)_{10}(GT)_{12}$	L: tgggaaataggatgcgctaaacc	R: aactttaagactcaaaatccataacc	223	55, 30	6	0.66	9
Hpms 1–5	$(AT)_{11}(GT)_{17}$	L: ccaaacgaaccgatgaacactc	R: gacaatgttgaaaaaggtggaagac	311	55, 30	8	0.80	6
Hpms 1–6	$(AT)_2(GT)_4(AT)_8(GT)_{13}$	L: tccataacttcacccatgagtatga	R: gcaacacccacattcccttctc	197	55, 30	11	0.91	
Hpms 1–41	$(AT)_6(GT)_{32IMP}$[f]	L: gggtatcatccgttgaaagttagg	R: caagaggtatcacaacatgagagg	192	55, 30	3	0.66	1
Hpms 1–43	$(GT)_9T(TG)_7$	L: aaccagcaatcccatgaaaacc	R: gggctttggggagaatagtgtg	154	55, 30	3	0.36	1
Hpms 1–62	$(TG)_{23}(AG)_9$	L: catgaggtctcgcatgatttcac	R: ggagaaggaccatgtactgcagag	186	55, 30	7	0.81	1
Hpms 1–69	$(AC)_9(AT)_4$	L: cggtggcatgtagtttctggag	R: aagacatgaaatccacaagttttc	217	50, 35	5	0.64	4
Hpms 1–106	$(AAAAAT)_4$	L: tccaaactacaagcctgcctaacc	R: ttttgcattattgagtcccacagc	159	55, 35	4	0.55	
Hpms 1-111	$(AAT)_{11}$	L: aagcttatccctttcaaatataa	R: atatctcacgtattgcggattctt	159	50, 35	11	0.91	
Hpms 1–117	$(AT)_9(GT)_{14}$	L: acccaaatttgccttgttgat	R: aatccataaccttatcccataaa	189	50, 35	4	0.77	9
Hpms 1–139	$(CT)_2(AG)_{15}$	L: ccaacagtaggacccgaaaatcc	R: atgaaggctactgctgcgatcc	299	55, 30	6	0.76	1
Hpms 1–143	$(AG)_{12}$	L: aatgctgagctggcaaggaaag	R: tgaaggcagtaggtggggagtg	221	55, 30	7	0.88	
Hpms 1–145	$(CT)_{21}(AT)_{13}$	L: agcttgtgtcataatcttgaaaaactc	R: tgaaaagacgattttgtctaatgcg	150	50, 35	9	0.86	
Hpms 1–148	$(GA)_{14}$	L: ggcggagaagaactagacgattagc	R: ccacccaatccacatagacg	197	55, 30	7	0.83	1
Hpms 1–155	$(TA)_3CA(GA)_{21}$	L: acgaggcccaagctgttatgtc	R: ttgtcccgactctccattgacc	207	55, 30	6	0.66	1
Hpms 1–165	$(GA)_{13}$	L: ggctatttccgacaaaccctcag	R: ccattggtgttttcactgttgtg	213	55, 35	7	0.75	4
Hpms 1–166	$(GA)_{13}AA(AT)_2$	L: ctccttcggaaatgcatgatgg	R: cagcactttgattccagaacttcg	132	55.30	5	0.74	
Hpms 1–168	$(TA)_{17}(GA)_{12}$	L: gccccgatcaatgaatttcaac	R: tgatttttgggtggagagaaaacc	208	55, 30	9	0.86	16
Hpms 1–172	$(GA)_{15}$	L: gggtttgcatgatctaagcatttt	R: cgctggaatgcattgtcaaaga	344	55, 30	9	0.88	11
Hpms 1–173	$(GA)_{16}(TG)_2$	L: tgctgggaaagatctcaaaagg	R: atcaaggaagcaaaccaatgc	163	50, 35	6	0.86	3
Hpms 1–214	$(GTTT)_2(TTG)_9$	L: tgcgagtaccgagttctttctag	R: ggcagtcctgggacaactcg	100	50, 35	6	0.78	1
Hpms 1–216	$(TTG)_7...(TTG)_8$	L: tgcttgttgtttttaccctcagc	R: agtgaaaggtgggcaacagc	108	55, 30	5	0.76	7
Hpms 1–227	$(TTG)_7$	L: cgtggcttcaagtatggactgc	R: ggggcggaacttttcttatcc	237	55, 30	3	0.58	7
Hpms 1-230	$(AG)_2(TTG)_7(ATG)_2$	L: gaatgtgaatcgcccgtatgc	R: catccggcatcaatatgttagtagc	214	50, 35	n.d.[g]	n.d.	
Hpms 1–274	$(GTT)_7$	L: tcccagacccctcgtgatag	R: tcctgctccttccacaactg	174	55, 30	7	0.83	7
Hpms 1–281	$(TTG)_6$	L: tgaggcagtggtatggtctgc	R: cccgagttcgtctgccaatag	132	55, 30	6	0.7	1
Hpms 2–2[h]	$(GT)_9$	L: gcaaggatgcttagttgggtgtc	R: tcccaaaattaccttgcagcac	146	55, 35	7	0.83	11
Hpms 2-9	$(GT)_7$	L: cccgtatgtgattctaggatgg	R: cgttagcaggtactgaggataagg	189	50, 35	3	0.31	
Hpms 2–13	$(AC)_{12}(AT)_4$	L: tcacctcataagggcttatcaatc	R: tccttaaccttacgaaaccttgg	259	50, 35	8	0.84	1
Hpms 2–18	$(TTG)_{11}$	L: cctccccagacctcactttatgc	R: tcaaccaaacaagtcgaagtcagc	162	55, 35	n.d.	n.d.	
Hpms 2–21	$(AT)_{11}(AC)_9(ATAC)_{10}$	L: tttttcaattgatgcatgaccgata	R: catgtcattttgtcattgatttgg	295	50, 35	7	0.83	10
Hpms 2–23	$(TTG)_7(GT)_9$	L: ccctcggctcaggataaatacc	R: ccccagactcccactttgtg	126	50, 35	6	0.75	5
Hpms 2–24	$(CT)_{17}(CA)_5A_{21}$	L: tcgtattggcttgtgatttaccg	R: ttgaatcgaatacccgcaggag	205	55, 35	6	0.79	9
Hpms 2–26	$(TTG)_7$	L: gggatgtaggaacaaccctaacc	R: tgcatcttttcttcatccccttc	217	55, 35	5	0.74	1,3,5
Hpms 2–27	$(TTG)_7$	L: atggactgcgtacatcctaccc	R: tgtcttttatgcactagcttgagagg	144	55, 35	n.d.	n.d.	
Hpms 2–41	$(TTG)_7$	L: cttcccagacctcactttgtgg	R: tctttgcggttatgtcaagtgc	161	55, 35	n.d.	n.d.	
Hpms 2–45	$(TTG)_9$	L: cgaaaggtagttttgggcctttg	R: tgggcccaatatgcttaagagc	148	55, 35	4	0.63	5
*Hpms*AT2–14	$(AAT)_{16IMP}$	L: tttagggtttccaactcttcttcc	R: ctaaccccaccaagcaaaacac	174	50, 35	6	0.79	4
*Hpms*AT2–20	$(AAT)_{18}$	L: tgcactgtcttgtgttaaaatgacg	R: aaaattgcacaaatatggctgctg	148	55, 35	6	0.75	6
*Hpms*CaSIG19	$(CT)_6(AT)_8(GTAT)_5$	L: catgaatttcgtcttgaaggtccc	R: aagggtgtatcgtacgcagcctta	218	55, 35	7	0.89	7
*Hpms*hpMADS	$(AT)_{17}$	L: tgctttcaaaacaatttgcatgg	R: vgcgtctaatgcaaaacacacattac	210	55, 30	3	0.5	1
CACCEL1[i]	$(AT)_{16IMP}$	L: ctctaataggcaatagctcacatgc	R: gcagtctcccagaacgttgtcc	243	50, 35	4	0.69	1
AA840689	$(GAGGTC)_2(GAGGGC)_2$	L: gacaacataggcggacctttgg	R: tgctttaggtctacgtccttgcac	267	55, 35	3	0.58	3
AA840692	T_{20}	L: tggaagtgattactggaaaccatgc	R: ggggtttagtcatggaatcttttgc	202	55, 35	3	0.43	3
AA840749	$(AAG)_3(AAT)_2A(AC)_2$	L: acactctcggtggaaagcttgg	R: tcctcttttcgggattccttagc	181	50, 35	1	0	
CAN132623	$(TTTG)_2(TTG)_{10}$	L: ggatgcggaagatgaagacga	R: caacaacagcaacaacaaatcaaa	237	60, 35	nd	nd	
AA840721	A_{13}	L: cactttgatacgtgaacacttcc	R: agtttgcactggtcctgctc	112	55, 30	5	0.64	7
CAN130829	T_{16}	L: gctaattacttgctccgttttg	R: aatgggggagtttgttttgg	184	55, 30	8	0.86	
AF242731	T_{18IMP}	L: gggctgacggccattaagaac	R: cagacagctagaaagagaggaattctg	195	50, 35	5	0.55	16
AF242732	$(TTG)_{6IMP}$	L: tggtgatgctggcatggttaag	R: cagcactgttcacttccccttc	205	50, 35	2	0.17	
AF244121	$(TTG)_{5IMP}(AT)_{6IMP}(GT)_{3IMP}$	L: tacctcctcgccaatccttctg	R: ttgaaagttctttccatgacaacc	238	50, 35	9	0.88	1,3
CAN010950	$(TA)_9$	L: gattttggtggcagaagaattgg	R: tgcactttcgaagcaaacaaacc	254	55, 35	6	0.81	1
AF208834	T_7A_{11}	L: tgcaccaaggtccagtaaggttg	R: ccaaccaccatggttcatacaag	201	55, 35	5	0.76	6
AF121441	$(TAG)_{4IMP}(GTT)_3$	L: gcagacaatcttggttggtttgg	R: caccttgagggctaaaggaagg	234	55, 35	1	0	
AF130118	$(AT)_{11IMP}$	L: atggatgcttcgccatctcaac	R: gacacagcagccatatgtatacgc	176	55, 35	3	0.31	
AF208832	$(TC)_5(TTTC)_2$	L: ccaaggtgaagtgacgatgagg	R: gactcggactacgactcgaaaag	178	50, 35	1	0	
AF222989	$(GAA)_6$	L: actgctgggcattgagggttag	R: gcgacatccctcacgacataac	208	55, 35	1	0	
AF242730	$(ATT)_7$	L: aggccatggaaaaagacactgg	R: acatgaccatgaccaaccatcc	189	55, 35	2	0.46	
CAN011109	$(GT)_{5IMP}(CA)_3$	L: ggccttcaactcctcaaagacg	R: tactgcatcggttcctggattg	195	55, 35	3	0.31	
AA840739	$(TTTC)_4$	L: ccctctctcttcccttcacc	R: agtgctacgatggcttaggg	95	50, 35	3	0.31	
AF039662	T_{26IMP}	L: ccccctcgtctctctttattt	R: ttgcaaatcttttgtcaattttt	114	50, 35	6	0.79	
AA840677	$(GA)_{9IMP}$	L: tgccttgaaggtagccaagagg	R: caaacatcaattgaggaggata-ccg	154	55, 35	2	0.40	
AA840763	$(GAA)_{6imp}$	L: tgctattgatgaatacgctgtgg	R: cctgtcacaaggaagtcttgatcg	179	55, 35	1	0	
AA840773	$(AAG)_{7imp}$	L: gagaagaaggccgagggagttc	R: gcacctgcttcagcaccttg	154	55, 35	1	0	
AA842818	$(TA)_4(TC)_3$	L: tccactgcaccacaaccaatg	R: ccataagcaaccaacagaatta-ggg	202	55, 35	3	0.31	
AA842825	$(CCA)_8$	L: ccaccatgaccaccatgagg	R: ttcagcccctctcatcaacc	202	55, 35	1	0	
AF065616	$(AG)_5(AAG)_{4imp}$	L: tctgtcccctgtgaagatttgg	R: gctccaccagttctgtaccttgg	205	55, 35	1	0	
CM0004[j]	$(CT)_6$	L: acaaacatatctatagtgcaaatt	R: attgtgctctgtcaaaacaa	103	50, 35	1	0	
CM0005	$(CCA)_8$	L: catgaccaccatgaggata	R: gatagccacgagcatagtatt	160	50, 35	6	0.79	
CM0006	$(AG)_5$	L: agttaacaactttggtgctgt	R: taatatggtaagcacattcca	130	50, 35	1	0	
CM0007	$(AG)_6$	L: tgttcctctctcttcttatcg	R: ccggagataagatcttgataa	103	50, 35	4	0.55	1
CM0009	$(AT)_5$	L: tgaggtctgaaaaaggtaaag	R: ttagttgaacttgcagaatcc	112	50, 35	1	0	
CM0010	$(AT)_6$	L: ttggtttttgctactggtaat	R: aaactgtcatatatttgtgtgact	158	50, 35	4	0.61	1
CM0011	$(AC)_5(TA)_8$	L: tctgctttaaaaacacatacat	R: cattctaactgaaattgcatg	116	50, 35	5	0.74	1
CM0012	$(AAT)_{11}$	L: gggatttaataaggaacaatg	R: tcaaatatcgacattagcatg	200	50, 35	1	0	

[a] PCR anealing temperature in °C and cycle numbers, [b] Linkage group, [c] *Hpms*1: hot pepper microsatellite from *Hin*dIII library, [d] Left primer, [e] Right primer, [f] Imperfect repeat, [g] Not determined, [h] *Hpms*2: hot pepper microsatellite from *Pst*I/*Mbo*I library, [i] GenBank accession number, [j] Reported by Huang et al. (2000)

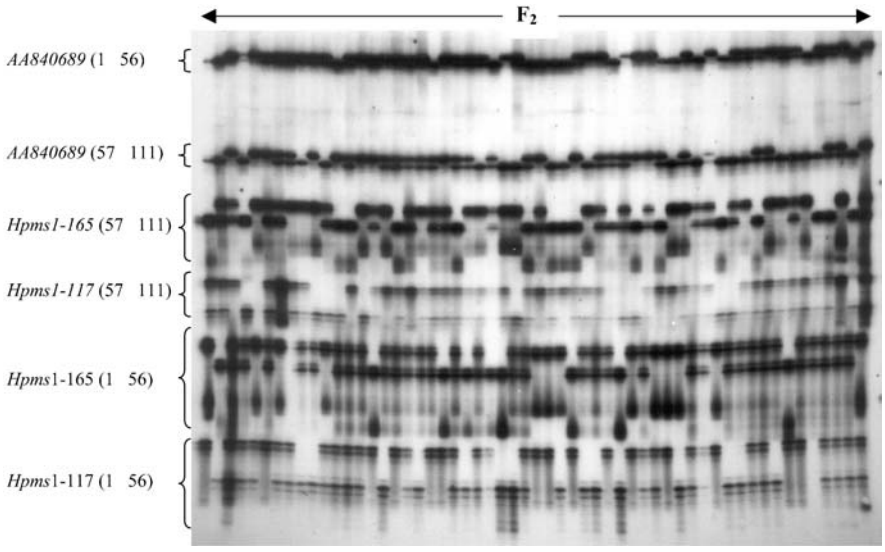

Fig. 1 Multiple loading of the 6% acryamide gel showing three different microsatellite markers segregating in 107 F_2 plants. Three different PCR products for individual F_2 plants were loaded sequentially in one gel-lane at 10 to 30 min intervals. *Hpms*1-117, *Hpms*1-165 and *AA840689* indicates primer pairs and microsatellite loci. The number in parenthesis identifies the individual F_2 plant used for PCR analysis

markers segregated in a co-dominant manner. Five dominant markers were *Hpms*1-1-a, *Hpms*1-1-b, *Hpms*2-13, *Hpms*2-23 and *Hpms*2-26-c.

The up-dated pepper map, designated "SNU2", consists of 333 molecular markers including 46 SSRs and 287 RFLPs (146 new loci and 141 from the original "SNU" map) (Fig. 2). The 333 markers were distributed among the 15 linkage groups covering 1,761.5 cM, and provided an average marker density of 1 marker per 5.3 cM.

Linkage groups in the "SNU2" map were designated to be consistent with the Cornell pepper map (Livingstone et al. 1999), the Volcani pepper map (Ben Chaim et al. 2001) and the Cornell tomato map (Tanksley et al. 1992). The SNU LGs were re-designated in the SNU2 map as follows: 1 and 16 to 1; 3 to 2 and 13; 2 and 12 to 3; 7 to 4; 10 and 15 to 5; 4 to 6; 6 to 7; 5 to 9; 13 to 10; 9 to 11; 8 to 12 and 15; 14 to 14; and 11 to 16. LG8 is missing as is the case of the Cornell pepper map according to the finding of Ben Chaim et al. (2001), who reported that the upper region of LG1 and the unlinked group A in the Cornell pepper map were joined into LG8 of the Volcani pepper map.

Discussion

Pepper microsatellites

The number, length, composition and chromosomal distribution of microsatellite sequences in a genome can vary drastically among taxa. The prevalence of AT repeats seems to be a general feature of the plant genome (Lagercrantz et al. 1993; Morgante and Olivieri 1993). CA/GT repeats, which is an abundant and variable class of microsatellites in mammalian genomes, is generally less frequent and less variable in plant genomes (Powell et al. 1996). In *Arabidopsis*, the CA/GT microsatellite is poorly represented (Depeiges et al. 1995) and shows a very low level of variability (Bell and Ecker 1994). In the sugar beet, GT-containing microsatellite sequences are part of a more complex repeating element which is present in multiple copies near the centromere and thus have limited potential for mapping purposes (Schmidt and Heslop-Harrison 1996). However, many informative microsatellite markers have been successfully developed based on the CA/GT microsatellite motif in maize (Chin et al. 1996), wheat (Roder et al. 1998) and rice (Temnykh et al. 2000). The CA/GT microsatellite was found to be the most abundant motif in pepper but less polymorphic than other motifs. Of the 30 microsatellite markers of pepper reported by Nagy et al. (1998), 16 microsatellite markers contained the CA/GT motif and had a low level of polymorphism, which is consistent with our results.

Interestingly, the TTG repeat was the next most abundant class in the pepper genome, whereas it was rare in other plants (Morgante et al. 2002). Homology search revealed that out of 24 pepper TTG repeats nine TTG clones had significant homology with the tobacco SINE element (Yoshioka et al. 1993) containing $(TTG)_7$ microsatellites, and two TTG clones exhibited homology with transposable elements from rice and pepper. It might be interesting to see whether the TTG repeat is also abundant in other members of the *Solanaceae*. Microsatellite markers of the *Solanaceae* have been reported (Broun and Tanksley 1996; Milbourne et al. 1998), however, there have been no studies using TTG as a probe for library screening. Ramsay et al. (1999) reported that a high percentage of SSRs is associated with both cereal retrotransposon-like elements and dispersed repetitive elements in barley.

Public databases used as a source for SSR development become increasingly valuable as the amounts of sequence data increase. Markers could easily be developed using sequence information on SSRs and their flanking sequences in the databases. Utilizing a database can obviate the time-consuming work of constructing a genomic library and screening it for microsatellite sequences; and this strategy has been successfully used

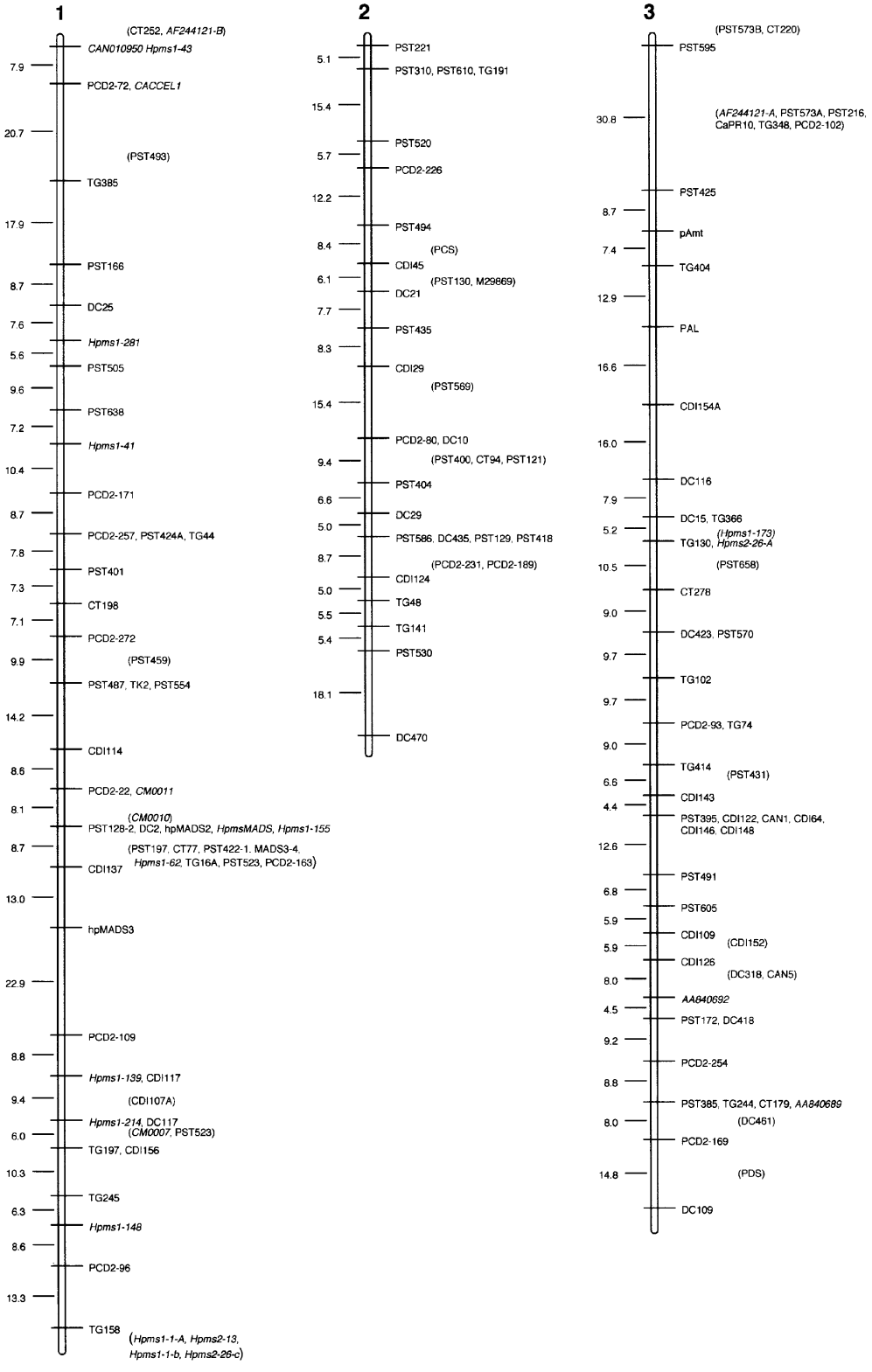

Fig. 2 The SNU2 pepper map consists of 46 SSRs and 287 RFLPs. The linkage groups (1–12) were labeled according to the synteny of tomato markers described in Livingstone et al. (1999) and Ben Chaim et al. (2001), and other groups were arbitrarily labeled according to the total map distance of each linkage group. On the left of the vertical double lines are map distances in cM calculated by the Kosambi function, and on the right are DNA markers by identification numbers and names. Markers by tick marks are framework markers ordered at LOD>3, and multiple markers at a locus mark were closely linked within 5 cM. Markers in parenthesis were placed between framework markers at LOD<3 or located using "try" command. Markers in *italic* type correspond to SSRs (Table 1). Marker types and designations are as follows: pepper SSRs (*Hpms*, *CM*, and *GenBank accession number*); pepper genomic RFLP (PST); pepper cDNA RFLP (PCD, DC and CDI); pepper secondary metabolite biosynthesis genes RFLP (PSY, GPS, PDS, LCY, CCS, TK2, CRTHYD, PFTF, COMT, PAL, Ca4H, PCS and pAMT); pepper EST RFLP (CAN, CFR and CLF); pepper MADS genes RFLP (hpMADS and MADSP10); pepper defence-related genes RFLP (DD19, CaPR, CaLTP and N32); pepper rDNA RFLP (RDNA); tomato genomic RFLP (TG); tomato cDNA RFLP (CT and CD); tobacco cDNA RFLP (X03913 and M29869). Uppercase letters at the end of the marker names indicate that the marker is one of at least two segregating loci detected by a single assay

for a number of species (Bell and Ecker 1994; Smulders et al. 1997; Cho et al. 2000).

The investigation of 312 database entries from *Capsicum* sequences yielded 32 microsatellites and this shows one microsatellite in every 6.7 kb sequence per database. This is comparable to the previous data showing one microsatellite every 8.4 kb sequence in *Lycopersicon* (Smulders et al. 1997) and every 13.8 kb sequence in *Arabidopsis* (Bell and Ecker 1994). The frequency of microsatellites in GenBank was higher than in the

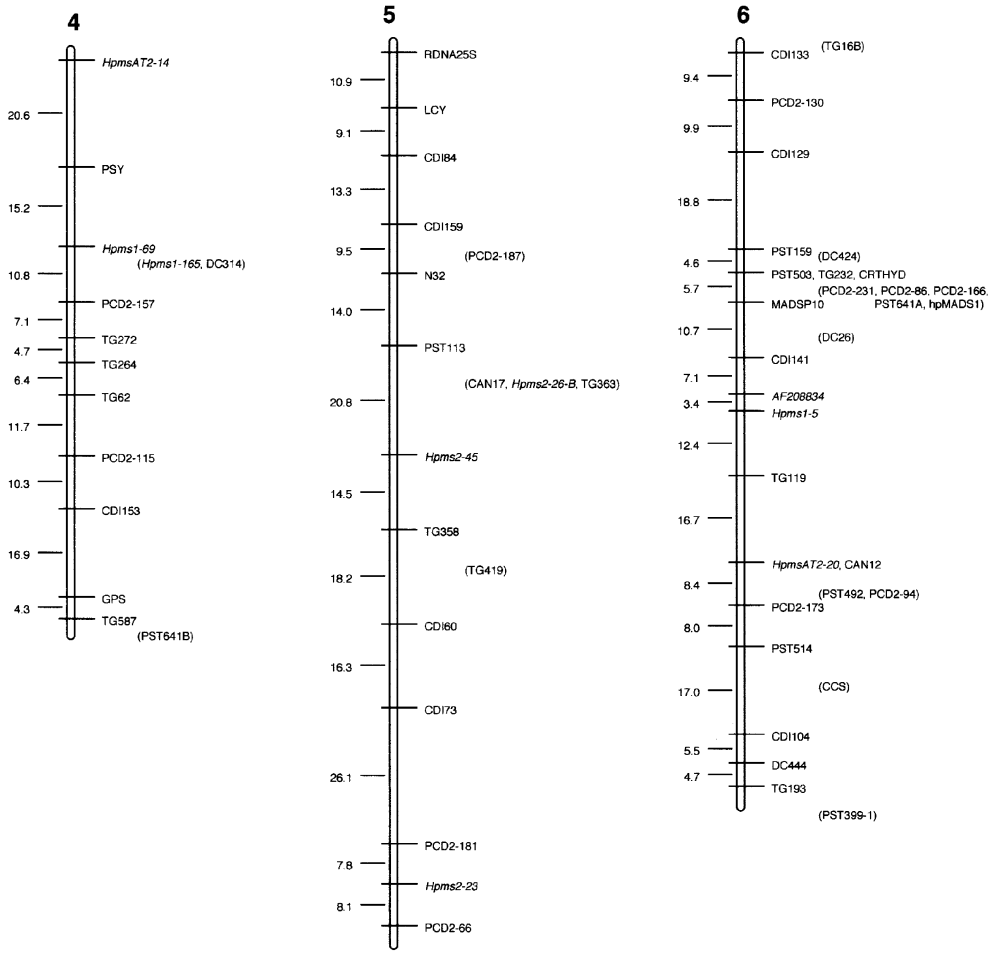

Fig. 2 (continued)

genomic libraries, indicating that the microsatellite frequency of the intragenic region was higher than the intergenic region (Morgante et al. 2002).

Among the 32 *Capsicum* microsatellites derived from GenBank, 15 microsatellites were found in ESTs. Of the ten primer pairs developed by EST-microsatellites, four pairs showed polymorphism between *C. annuum* cv TF68 and *C. chinense* cv Habanero. These data indicate that many microsatellite sequences and markers can be developed from the pepper EST project.

Microsatellites derived from GenBank had lower PIC values than those developed from the genomic libraries, which is consistent with studies on rice (Cho et al. 2000). Although there is no direct comparison among pepper marker types, those SSR markers would be likely to show a higher level of PIC than that of other markers, such as RFLPs, AFLPs and RAPDs (Prince et al. 1995; Paran et al. 1998; Rodriguez et al. 1999).

Pepper genome mapping

The first 46 polymorphic microsatellite loci have been assigned to the *Capsicum* RFLP linkage map. Of these, 13 SSR markers were obtained from GenBank and 29 SSR markers from the genomic DNA libraries. Detection of discrete loci, segregation in a Mendelian fashion, and co-dominance of these SSR markers made them ideal genetic markers.

In addition to the SSR markers, 146 RFLP markers were added onto our previous SNU map containing 141 RFLP markers (Kang et al. 2001). New additional RFLPs included pepper genomic clones, tomato clones, defence-related pepper clones, pungency related clones and other pepper genes. In the SNU2 map, 51 markers (15.3%) are from tomato and the rest from pepper, which would serve as a good reference for comparative pepper genome studies.

The SNU2 map consists of 15 linkage groups even though pepper has 12 sets of chromosomes. However, more recent data and an additional intraspecific map we are developing, suggest that LG2 and LG15 are syntenic and LG12 and LG13 are syntenic; thus reducing the

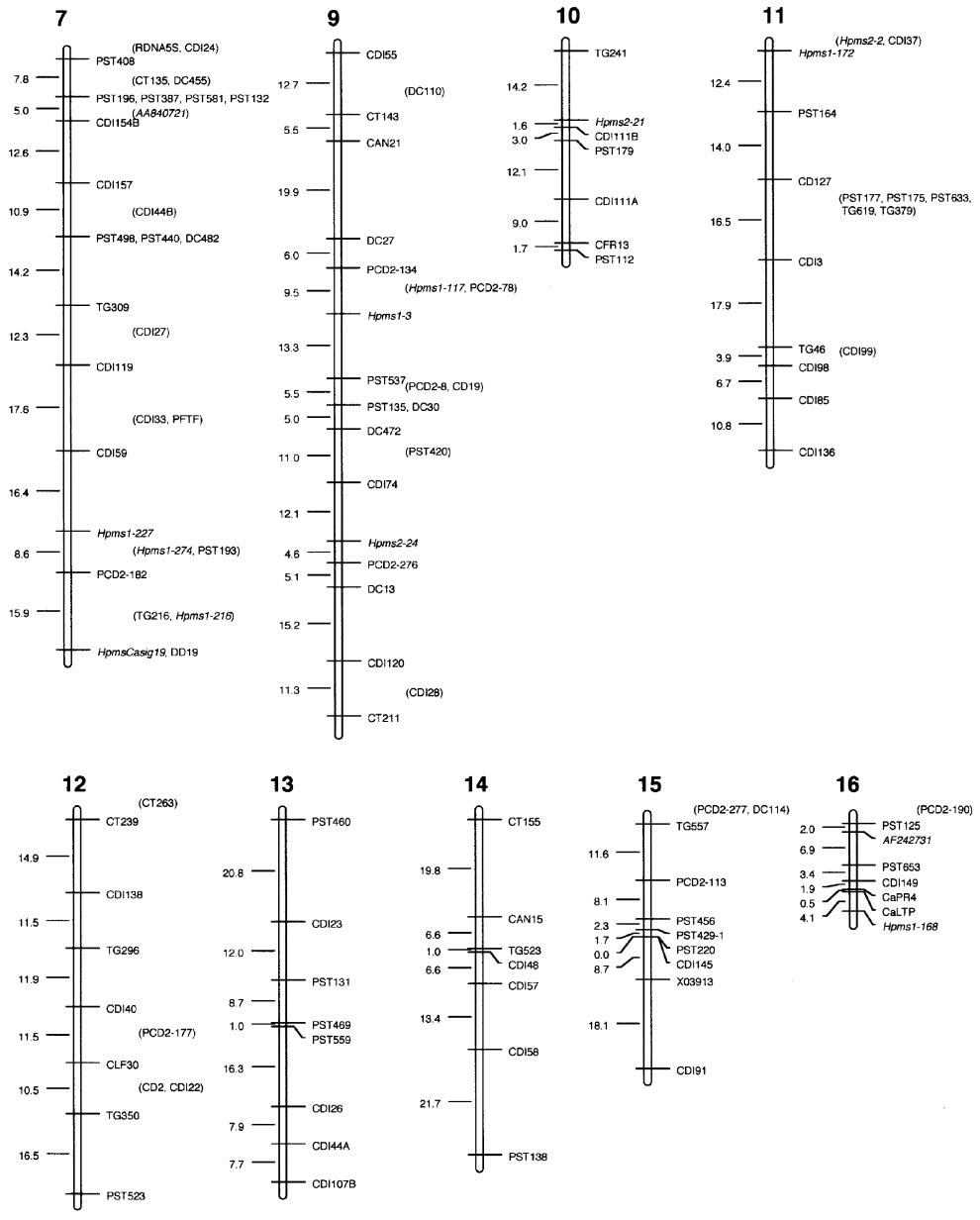

Fig. 2 (continued)

number of linkage groups into 13 (unpublished data). The numbering of LGs in SNU2 is consistent with other pepper and tomato maps.

Acknowledgement This work was supported by a grant from the R & D Promotion Center for Agriculture and Forestry, and from the KOSEF through the Center for Plant Molecular Genetics and Breeding Research.

References

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137–144

Ben Chaim A, Paran I, Grube RC, Jahn M, van Wijk R, Peleman J (2001) QTL mapping of fruit-related traits in pepper (*Capsicum annuum*). Theor Appl Genet 102:1016–1028

Broun P, Tanksley SD (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol Gen Genet 250:39–49

Chin ECL, Senior ML, Shu H, Smith JSC (1996) Maize simple repetitive DNA sequences: abundance and allele variation. Genome 39:866–873

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

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

Edwards A, Civitello H, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with a trimeric tetrameric tandem repeat. Am J Hum Genet 49:746–756

Huh JH, Kang BC, Nahm SH, Kim S, Ha KS, Lee MH, Kim BD (2001) A candidate gene approach identified phytoene synthase as the locus for mature fruit color in red pepper (*Capsicum* spp.). Theor Appl Genet 102:524–530

Huang S, Zhang B, Milbourne D, Cardle L, Yang G, Guo J (2000) Development of pepper SSR markers from sequence databases. Euphytica 117:163–167

Innan H, Terauchi R, Miyashita NT (1997) Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. Genetics 146:1441–52

Jacob HJ, Lindpaintner K, Lincodpaintner K, Lincoln SE, Kusumi K, Bunker RK, Mao YP, Ganten D, Dzau VJ, Lander ES (1991) Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensile rat. Cell 67:213–224

Kang, BC, Yu JW, Lee MH, Kim BD (1997) Applicability of AFLP on hot pepper genetic analysis. J Korean Soc Hort Sci 38:698–703

Kang BC, Nahm SH, Huh JH, Yoo HS, Yu JW, Lee MH, Kim BD (2001) An interspecific (*Capsicum annuum* × *C. chinese*) F_2 linkage map in pepper using RFLP and AFLP markers. Theor Appl Genet 102:531–539

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugenics 12:172–175

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

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

Lee SJ (2000) Isolation and molecular characterization of two SA-inducible genes from hot pepper plants (*Capsicum annuum* L.). PhD thesis, Seoul National University

Lefebvre V, Palloix A, Caranta C, Pochard E (1995) Construction of an intraspecific integrated linkage map of pepper using molecular markers and doubled-haploid progenies. Genome 38:112–121

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplication of a dinucleotide repeat within the cardiac muscle action gene. Am J Hum Genet 44:388–396

Livingstone KD, Lackney VK, Blauth JR, Van Wijk R, Jahn MK (1999) Genome mapping in Capsicum and evolution of genome structure in the Solanaceae. Genetics 152:1183–1202

Milbourne D, Meyer RC, Collins AJ, Ramsay LD, Gebhardt C, Waugh R (1998) Isolation, characterisation and mapping of simple sequence repeat loci in potato. Mol Gen Genet 259:233–45

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

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with non-repetitive DNA in plant genomes. Nature Genet 30:194–200

Nagy I, Polley A, Ganal M (1998) Development and characterization of microsatellte markers in pepper. Xth Meeting on Genetics and Breeding of *Capsicum* and Eggplant, pp 235–237

Nahm SH, Yu JW, Kang BC, Kim BD (1997) Election of parental lines for the hot pepper mapping population using RFLP and AFLP analyses. J Korean Soc Hort Sci 38:693–697

Paran I, Aftergoot E, Shifriss C (1998) Variation in *Capsicum annuum* revealed by RAPD and AFLP markers. Euphytica 99:167–173

Powell W, Morgante M. Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Prince JP, Lackney VK, Angeles C, Blauth JR, Kyle MM (1995) A survey of DNA polymorphism within the genus *Capsicum* and the fingerprinting of pepper cultivar. Genome 36:404–417

Ramsay L, Macaulay M, Cardle L, Morgante M, Ivanissevich S, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J 17:415–25

Roder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–23

Rodriguez JM, Berke T, Engle L, Nienhuis J (1999) Variation among and within *Capsicum* species revealed by RAPD markers. Theor Appl Genet 99:147–156

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

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

Schmidt T, Heslop-Harrison JS (1996) The physical and genomic organization of microsatellites in sugar beet. Proc Natl Acad Sci USA 93:8761–8765

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 94:264–272

Tai TH, Dahlbeck D, Clark ET, Gajiwala P, Pasion R, Whalen MC, Stall RE, Staskawicz BJ (1999) Expression of the *Bs*2 pepper gene confers resistance to bacterial spot disease in tomato. Proc Natl Acad Sci USA 23:14153–14158

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes; Biological inferences and practical applications. Genetics 132:1141–1160

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch S (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Yoshioka Y, Matsumoto S, Kojima S, Ohshima K, Okada N, Machida Y (1993) Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific transfer RNAs. Proc Natl Acad Sci USA 90:6562–6566

Yu JW (2001) Isolation and molecular characterization of three MADS genes from the hot pepper plant (*Capsicum annuum* L.). PhD thesis, Seoul National University