Theor Appl Genet (2006) 114:113–130
DOI 10.1007/s00122-006-0415-y

ORIGINAL PAPER

Exploitation of pepper EST–SSRs and an SSR-based linkage map

Gibum Yi · Je Min Lee · Sanghyeob Lee · Doil Choi · Byung-Dong Kim

Received: 29 May 2006 / Accepted: 23 September 2006 / Published online: 18 October 2006

Abstract As genome and cDNA sequencing projects progress, a tremendous amount of sequence information is becoming publicly available. These sequence resources can be exploited for gene discovery and marker development. Simple sequence repeat (SSR) markers are among the most useful because of their great variability, abundance, and ease of analysis. By in silico analysis of 10,232 non-redundant expressed sequence tags (ESTs) in pepper as a source of SSR markers, 1,201 SSRs were found, corresponding to one SSR in every 3.8 kb of the ESTs. Eighteen percent of the SSR–ESTs were dinucleotide repeats, 66.0% were trinucleotide, 7.7% tetranucleotide, and 8.2% pentanucleotide; AAG (14%) and AG (12.4%) motifs were the most abundant repeat types. Based on the flanking sequences of these 1,201 SSRs, 812 primer pairs that satisfied melting temperature conditions and PCR product sizes were designed. 513 SSRs (63.1%) were successfully amplified and 150 of them (29.2%) showed polymorphism between *Capsicum annuum* 'TF68' and *C. chinense* 'Habanero'. Dinucleotide SSRs and EST–SSR markers containing AC-motifs were the most polymorphic. Polymorphism increased with repeat length and repeat number. The polymorphic EST–SSRs were mapped onto the previously generated pepper linkage map, using 107 F_2 individuals from an interspecific cross of TF68 × Habanero. One-hundred and thirtynine EST–SSRs were located on the linkage map in addition to 41 previous SSRs and 63 RFLP markers, forming 14 linkage groups (LGs) and spanning 2,201.5 cM. The EST–SSR markers were distributed over all the LGs. This SSR-based map will be useful as a reference map in *Capsicum* and should facilitate the use of molecular markers in pepper breeding.

Communicated by I. Paran.

Gibum Yi and Je Min Lee equally contributed to this work.

G. Yi · J. M. Lee · 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 151-921, South Korea
e-mail: kimbd@snu.ac.kr

S. Lee · D. Choi
Plant Genomics Laboratory, Genome Research Center,
Korea Research Institute of Bioscience and Biotechnology,
Daejon 305-333, South Korea

Present Address:
D. Choi
Department of Plant Science,
College of Agriculture and Life Sciences
and Center for Plant Molecular Genetics
and Breeding Research, Seoul National University,
Seoul 151-921, Korea

Introduction

DNA sequence polymorphisms or variations are of great importance in studies of molecular genetics. DNA markers, such as RFLPs, AFLPs, RAPDs, and SSRs have been valuable for analyzing genetic variation, and have helped to facilitate crop breeding (Kumar 1999). Since SSR markers display hypervariability, codominant inheritance, multiallelism, reproducibility, and good genome coverage, they are becoming the preferred marker system (Powell et al. 1996). SSRs, first referred to as microsatellites, are a class of DNA sequences consisting of simple motifs of

1–6 nucleotides that are repeated in tandem up to a few dozen times per site (Litt and Luty 1989). Replication slippage and unequal crossing over are the main driving forces for the insertion or deletion of tandem repeats, and lead to frequent variability of the SSRs (Ellegren 2004). SSR markers have been useful in a variety of applications such as linkage map construction, gene tagging, and studies of genetic diversity and evolution (Powell et al. 1996; Varshney et al. 2005).

To develop SSR markers, it is necessary to generate a small-insert genomic library, perform hybridization with SSR oligonucleotides and sequence candidate clones. This is quite time-consuming, costly, and labor-intensive. Alternatively, SSRs can be identified in sequence databases and used for SSR marker development (Morgante et al. 1993). When the volume of sequence data available for SSR analysis was still limited, this strategy was successfully used in tomato (Smulder et al. 1997), rice (Miyao et al. 1996), *Arabidopsis* (Bell and Ecker 1994), and soybean (Akkaya et al. 1992). The subsequent rapid increase in volume of sequence data in several species has facilitated the identification of a large number of SSRs in silico, using computer programs such as Sputnik, SSRFinder, MIcroSAtellite (MISA) and Tandem Repeat Finder (TRF) (summarized by Varshney et al. 2005).

Single-pass partial sequencing of the 5′ or 3′ ends of cDNA clones corresponding to mRNAs, yielding what are known as expressed sequence tags (ESTs), is a fast and efficient way to analyze the transcribed portion of the genome. ESTs are typically a few hundred base pairs in length and provide a robust sequence resource that can be exploited for gene discovery, genome annotation, and comparative genomics. Over three million sequences, from approximately 200 plant species, have been deposited in the publicly available plant EST sequence databases (Rudd 2003). These EST sequence resources have been shown to have a wide range of applications in the field of genome as well as transcriptome analysis. Contemporary DNA marker types using EST data include single nucleotide polymorphisms (SNPs) (Picoult-Newberg et al. 1999), conserved orthologue sets (COSs) (Fulton et al. 2002), and EST-derived SSRs (EST–SSRs) (Cordeiro et al. 2001).

SSR-containing ESTs (SSR–ESTs) have been investigated in plants such as *Arabidopsis*, maize, soybean, rice, and wheat (Cardle et al. 2000; Tóth et al. 2000; Morgante et al. 2002). They are currently used for genetic mapping, comparative mapping, and analysis of genetic and functional diversity (reviewed by Varshney et al. 2005). Although genomic SSRs are more polymorphic than genic SSRs (Cho et al. 2000; Lee et al. 2004a), the latter have advantages. First, the large amount of DNA sequence data available reduces the cost of SSR development. Second, the estimated frequency of genic SSRs is higher than of genomic SSRs (Morgante et al. 2002). Third, genic markers are functional markers that can be transferred between species since they derive from putative candidate genes related to traits of interest that are relatively well conserved among taxa (Gupta and Rustgi 2004). For these reasons, SSR markers have been developed from the sequence databases of many crops and used in marker-assisted breeding programs. EST–SSR markers provide a set of easily shared markers that can be used to unify different genetic maps and establish consensus maps.

The various pepper genome maps constructed to date, have a few simple functional markers in common (Lefebvre et al. 2002; Livingstone et al. 1999; Paran et al. 2004). Although SSR markers have been introduced into the pepper molecular map (Lee et al. 2004a), only a limited number of SSR markers are yet publicly available (Huang et al. 2000; Tam et al. 2005). Hence, it would be desirable to develop more SSR markers to construct saturated consensus genetic maps, as well as for marker-assisted breeding and germplasm identification. To generate highly informative and genome-wide polymorphic markers, we developed EST–SSR markers by analyzing a pepper EST database (http://www.plant.pdrc.re.kr/ks200201/pepper.html) (Lee et al. 2004b). These new markers were incorporated into our pepper maps consisting of SSR and RFLP markers (Kang et al. 2001; Lee et al. 2004a). Evenly distributed and informative SSR markers promise to promote the use of linkage maps in pepper genetics and breeding.

Materials and methods

Pepper EST–SSR analysis

A total of 10,232 non-redundant pepper EST sequences (http://www.plant.pdrc.re.kr/ks200201/pepper.html) (Lee et al. 2004b) was used for SSR analysis. The ESTs were generated from seven different cDNA libraries: KS01 (leaves infected with *Xanthomonas campestris* pv. glycines); KS07 (flower buds); KS08 (anthers); KS09 (young fruits); KS10 (hairy root); KS11 (early root); KS12 (green fruit placenta). We analyzed EST sequence data involving 25,819 sequences using an SGI Origin 3200 Unix machine (SGI Korea, South Korea). The ABI-formatted chromatogram sequences were fed into PHRAP (Ewing and Gree 1998; Ewing et al. 1998) and those that contained 97% or more unambiguous bases and exceeded 100 bp were further analyzed.

Sequences such as vectors, linker and polyA tail of the selected ESTs were trimmed. After trimming, the pre-processed ESTs were clustered into consensus sequences using StackPack (kindly provided by SANBI, http://www.sanbi.ac.za) to find non-redundant sequences. An EST that contained only one sequence was classified as a singleton. Sputnik (http://www.abajian.net), which finds SSRs of length over 11 bp with 2–5 bp motifs, was employed for SSR mining. In this study, the 11 bp length of SSRs were deleted.

ESTs containing polymorphic SSRs were searched against the GenBank non-redundant (nr) database using the TBLASTX algorithm (http://www.ncbi.nlm.nih.gov/BLAST). The putative functions of ESTs with expected value $<10^{-6}$ by TBLASTX were assigned.

Marker development

We attempted to design primers for all the EST–SSRs using Primer3 (http://www.-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). It proved to be difficult or impossible to design primers for 388 of the 1,201 EST–SSRs, since they were located at the border of the EST sequences or surrounded by AT-rich sequences. All of the primers were obtained from Bioneer Corporation (Daejon, Korea) and may be ordered from Bioneer (http://www.bioneer.com) by marker name. A total of 813 primer pairs were designed using parameters of product size 100–350 bp, primer length 20–24 bp, and melting temperature 60–68. All the primer pairs were used in the test for polymorphism between *Capsicum annuum* 'TF68' and *C. chinense* 'Habanero'. PCR amplifications and gel electrophoresis were performed as described (Lee et al. 2004a). All SSR amplifications were conducted in a PTC 200 DNA Engine Thermal Cycler (MJ Research, USA). After 3 min at 94, 35 cycles were performed with each cycle consisting of 30 s at 94, 30 s at 55 (except for HpmsE004 and -E016 at 60), 60 s at 72 and a final extension of 10 min at 72. Radioisotope-labeled PCR products were electrophoresed on 6% acrylamide gels and exposed to X-ray film. Two SSRs (HpmsE134, HpmsE136) could be analyzed on a single 2% agarose gel because of the considerable length polymorphism in the parental lines.

Mapping EST–SSRs

107 F_2 plants derived from the interspecific cross *C. annuum* 'TF68' and *C. chinense* 'Habanero' were used for linkage mapping (Nahm et al. 1997; Huh et al. 2001; Kang et al. 2001; Lee et al. 2004a).

Linkage analysis was performed using the software package MAPMAKER V3.0 (Lander et al. 1987). The pepper SSR-based linkage map was constructed with 243 molecular markers (180 SSR markers and 63 RFLP markers). To identify LGs using informative markers, pairwise comparisons and grouping of markers were performed, using the 'Group' command with a maximum recombination fraction of 25 cM and LOD score >5.0. The 'Order' command was used to construct a frame map (LOD >3, maximum distance <25 cM), and the ordered markers were confirmed using the 'Ripple' command. The rest of the markers were assigned to their positions with the 'Try' command. Map distances were calculated with the Kosambi mapping function.

Results

Frequency and distribution of SSRs in pepper EST sequences

By analyzing 10,232 non-redundant pepper ESTs comprising 4.56 Mb, we identified 1,201 SSRs in 1,042 ESTs (10.2%). Of the 1,201 SSR–ESTs, 129 (10.7%) contained more than one SSR. The average length of the SSRs was 17.1 bp and there was on average one SSR every 3.8 kb of the ESTs. The portion of the total length of EST sequence that consisted of SSRs, comprised about 0.45%. The 1,201 SSRs consisted of 66.0% trinucleotide, 18.0% dinucleotide, 7.7% tetranucleotide, and 8.2% pentanucleotide repeats (Table 1). Trinucleotide repeats were the predominant form both in frequency and length. The total length of the dinucleotide repeats made up 28.5% of all SSR sequences, whereas that of the trinucleotide repeats made up 56.8%. The average length of dinucleotide repeats was 27.1 bp, about twice that of trinucleotide repeats (14.7 bp). Of the total 10.2% SSRs were complex, which contains more than two repeat motives.

The frequencies of the individual SSR motifs are summarized in Table 2. Sixty-six types of motifs were recognized. The AAG motif was the most frequent (14.0%), followed by AG (12.4%), ACC (9.4%), AAC (7.3%), ACT (6.6%), AGT (6.0%), and AGG (5.8%). Other motifs made up less than 5% of the total.

Table 1 SSR frequencies in pepper ESTs

Repeat type	Number of SSRs	Total length (bp)	Average length (bp)
Di	216 (18.0)[a]	5,861 (28.5)	27.1
Tri	793 (66.0)	11,686 (56.8)	14.7
Tetra	93 (7.7)	1,485 (7.2)	16.0
Penta	99 (8.2)	1,537 (7.5)	15.5
Total	1201 (100.0)	20,569 (100.0)	17.1

[a] Numbers in parentheses are percentages

Table 2 Distribution of different types of SSRs in pepper ESTs

Repeat type	Motif	Number	Total length (bp)	Average length (bp)
Dinucleotide		216	5861	27.1
	AC	28	493	17.6
	AG	149	4601	30.9
	AT	39	767	19.7
Trinucleotide		793	11686	14.7
	AAC	88	1321	15.0
	AAG	168	2536	15.1
	AAT	50	808	16.2
	ACC	113	1639	14.5
	ACG	48	675	14.1
	ACT	79	1068	13.5
	AGC	50	766	15.3
	AGG	70	1006	14.4
	AGT	72	1086	15.1
	CCG	55	781	14.2
Tetranucleotide		93	1485	16.0
	AAAC	7	105	15.0
	AAAG	21	325	15.5
	AAAT	19	326	17.2
	AACC	2	26	13.0
	AACG	2	33	16.5
	AACT	1	20	20.0
	AAGG	10	143	14.3
	AAGT	2	26	13.0
	AATC	1	15	15.0
	AATG	1	22	22.0
	AATT	11	166	15.1
	ACAT	5	80	16.0
	ACCT	2	28	14.0
	ACGC	1	16	16.0
	ACTC	1	16	16.0
	AGCC	1	19	19.0
	AGCG	1	13	13.0
	AGCT	1	15	15.0
	AGGG	3	73	24.3
	CCCG	1	18	18.0
Pentanucleotide		99	1537	15.5
	AAAAC	6	93	15.5
	AAAAG	23	358	15.6
	AAAAT	8	113	14.1
	AAACG	1	14	14.0
	AAAGT	2	32	16.0
	AAATC	4	66	16.5
	AAATT	4	57	14.3
	AACAC	4	64	16.0
	AACCC	1	14	14.0
	AACCG	1	15	15.0
	AACGG	1	14	14.0
	AAGAC	1	14	14.0
	AAGAG	10	150	15.0
	AAGCC	1	14	14.0
	AAGGG	2	28	14.0
	AATAC	2	30	15.0
	AATAG	1	14	14.0
	AATAT	2	34	17.0
	AATGT	2	39	19.5
	ACAGG	1	14	14.0
	ACAGT	1	14	14.0
	ACATG	3	42	14.0
	ACCCC	4	80	20.0
	ACCCG	3	44	14.7
	ACCCT	1	14	14.0
	ACGGT	1	17	17.0
	ACTAT	1	18	18.0
	AGAGG	1	19	19.0
	AGAGT	1	16	16.0
	AGATC	2	29	14.5
	AGCCC	1	14	14.0
	AGGTC	2	32	16.0
	AGTAT	1	21	21.0
Total		1201	20569	17.1

The AG motif (12.4%) was the most abundant dinucleotide SSR, while the AT (3.2%) and AC (2.3%) motifs were much less frequent and we detected no GC motifs. Since the poly(A) sequences at the 3′ ends of the mRNAs were used to obtain the ESTs, the AA motif was excluded. The AG motif had the longest repeat length (30.9 bp), comprising 78.5% of the total length of the dinucleotide repeats. Ten different combinations of trinucleotide repeat motifs are possible and all the ten were detected in the pepper ESTs. AAG (14.0%), ACC (9.4%), and AAC (7.3%) were the most common. The lengths of the trinucleotide repeats ranged from 13.5 to 16.2 bp. There were 20 kinds of tetranucleotide repeat motif. Of these, nine were found just once and five were encountered two or three times. The most frequent tetranucleotide and pentanucleotide motifs were AAAG and AAAAG, respectively. AT-rich EST–SSRs were abundant among the tetra- and pentanucleotides (76.7%).

Marker development and polymorphism of the EST–SSRs

We designed SSR primers for 813 of the 1,201 EST–SSRs. Five-hundred and thirteen of the primer sets successfully amplified one or two PCR fragments from the genomic DNAs of 'TF68' and 'Habanero'. The primer sets that produced PCR products of the expected size from both genomes were used in further mapping. The products amplified from two EST–SSRs (HpmsE064 and -E071) were larger than expected. When null alleles (alleles that did not generate a PCR product) were detected in one or both genotypes, the corresponding SSRs were eliminated. The EST–SSR markers were designated HpmsE001–HpmsE150. The designation 'Hpms' stands for hot pepper microsatel-

lites; it was also used for SSR markers previously developed from genomic libraries (Lee et al. 2004a), and the letter 'E' next to 'Hpms' was added to indicate markers derived from pepper ESTs. We designed primers for 21 mononucleotide repeats and six hexanucleotide repeats. Four mononucleotide repeats (HpmsE006, -E009, -E010, and -E013) and one hexanucleotide (HpmsE005) repeat were found to be polymorphic.

The efficiency of marker development was examined for each repeat motif. To determine the rate of success of PCR amplification and the level of polymorphism of the EST–SSR markers for each SSR motif, several characteristics of marker development were calculated and are listed in Table 3. The average success rate of PCR amplification was 63.1%. The pentanucleotide motif had the highest rate (75.0%) of PCR amplification, followed by tri- (63.8%), di- (61.3%), and tetra- (50.8%) nucleotides. The EST–SSRs with AC, AT, AAC, AGC, ATC, and CCG motifs had success rates above 70%, with those containing AT-motif the most successful in terms of PCR amplification (84%).

EST–SSRs with dinucleotide repeats had higher levels of polymorphism (65.2%) than those with other repeat motifs (trinucleotide 26.7%, tetranucleotide 42.4%, and pentanucleotide 13.3%). Those containing AC, AG, AT, and AAT motifs had polymorphism levels in excess of 50%, with AC the most polymorphic (87%). The relationships among repeat numbers, SSR lengths, and polymorphism levels are shown in Fig. 1. There seems to be a correlation between polymorphism level and repeat number or length. Thirty-five percent of EST–SSRs with more than eight repeats were polymorphic compared with 14.7% of those with less than eight repeats. The polymorphism level for SSR lengths of more than 18 bp was 42.9%, whereas for SSR lengths of less than 18 bp, it was 17.7%.

Table 3 Characteristics of pepper EST–SSRs and efficiency of marker development

Motif	No. of EST–SSRs	No. of designed primers	No. of amplified EST–SSRs[a] (%)	No. of polymorphic EST–SSRs[b] (%)
Di-	216	75	46 (61.3)	30 (65.2)
AC	28	10	7 (70.0)	6 (85.7)
AG	149	46	23 (50.0)	16 (69.6)
AT	39	19	16 (84.2)	8 (50.0)
Tri-	793	511	326 (63.8)	87 (26.7)
AAC	88	58	42 (72.4)	15 (35.7)
AAG	168	114	68 (59.6)	19 (27.9)
AAT	50	31	16 (51.6)	8 (50.0)
ACC	113	56	38 (67.9)	9 (23.7)
ACG	48	36	19 (52.8)	3 (15.8)
ACT	79	52	28 (53.8)	2 (7.1)
AGC	50	32	23 (71.9)	9 (39.1)
AGG	70	44	27 (61.4)	7 (25.9)
ATC	72	50	38 (76.0)	12 (31.6)
CCG	55	38	27 (71.1)	3 (11.1)
Tetra-	93	65	33 (50.8)	14 (42.4)
Penta-	99	60	45 (75.0)	6 (13.3)
Complex	122	102	63 (61.7)	13 (20.6)
Total	1201	813	513 (63.1)	150 (29.2)

[a] Percentage of successfully amplified EST–SSRs per designed primer pair

[b] Percentage of polymorphic markers per amplified primer pair

Production of a pepper SSR-based linkage map

One-hundred and fifty EST–SSRs were polymorphic between the parental lines. Of these 139 could be positioned on the pepper linkage map using 41 SSR and 63 RFLP markers selected from our first (Kang et al. 2001) and second (Lee et al. 2004a) SNU maps (Fig. 2). Eleven of the 150 EST–SSR markers failed to show linkage. Forty-one SSR and 63 RFLP markers were selected to serve as anchor markers on the basis of genetic information value, distal location and gap-filling value to ensure full coverage for mapping. Fourteen LGs were generated with 243 markers in LOD >5.0 and maximum distance <25 cM (Fig. 2). Most EST–SSRs were codominant except for HpmsE138, that mapped to a unique location.

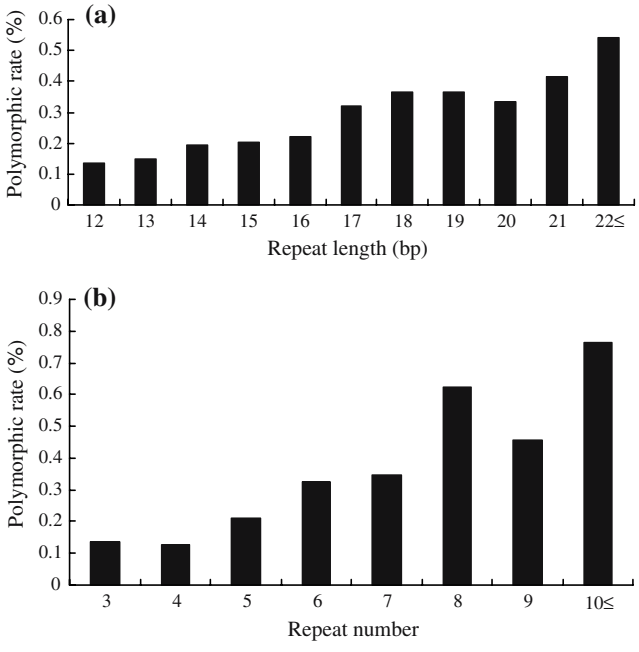

Fig. 1 Relationship between SSR length and polymorphism level **a**, and between repeat number and polymorphism level **b**. Polymorphism level was calculated as polymorphic markers per amplified primer

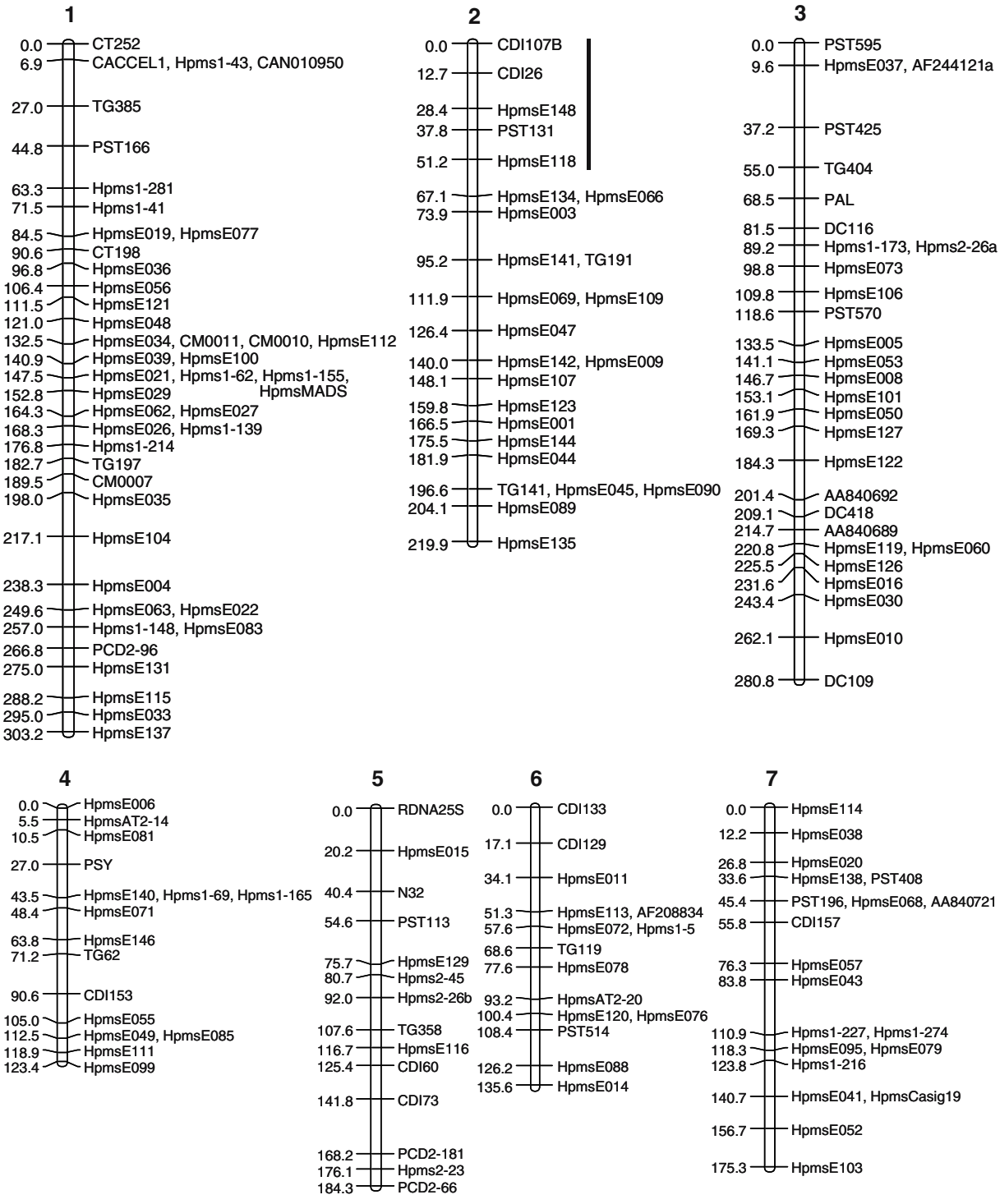

Fig. 2 An SSR-based pepper linkage map (SNU3). This map was constructed from 107 F_2 plants derived from an interspecific cross of *C. annuum* 'TF68' and *C. chinense* 'Habanero' using 139 new EST–SSR markers with the published anchors of 41 SSRs and 63 RFLPs (Lee et al. 2004a). Positions of loci are given in centiMorgan (cM). Multiple markers at the same location were within 5 cM of one another. There is no chromosome 8 in the linkage map since chromosomes 1 and 8 could not be separated in an interspecific map (Lee et al. 2004a). The previously designated LG 13 and LG 15 (Lee et al. 2004a) were merged by this study into chromosome 2 and chromosome 12, respectively, as indicated by the *vertical lines*. The previous LG 14 and LG 16 are renamed LG a and LG b, respectively. LG c was newly generated in this study solely from the EST–SSRs

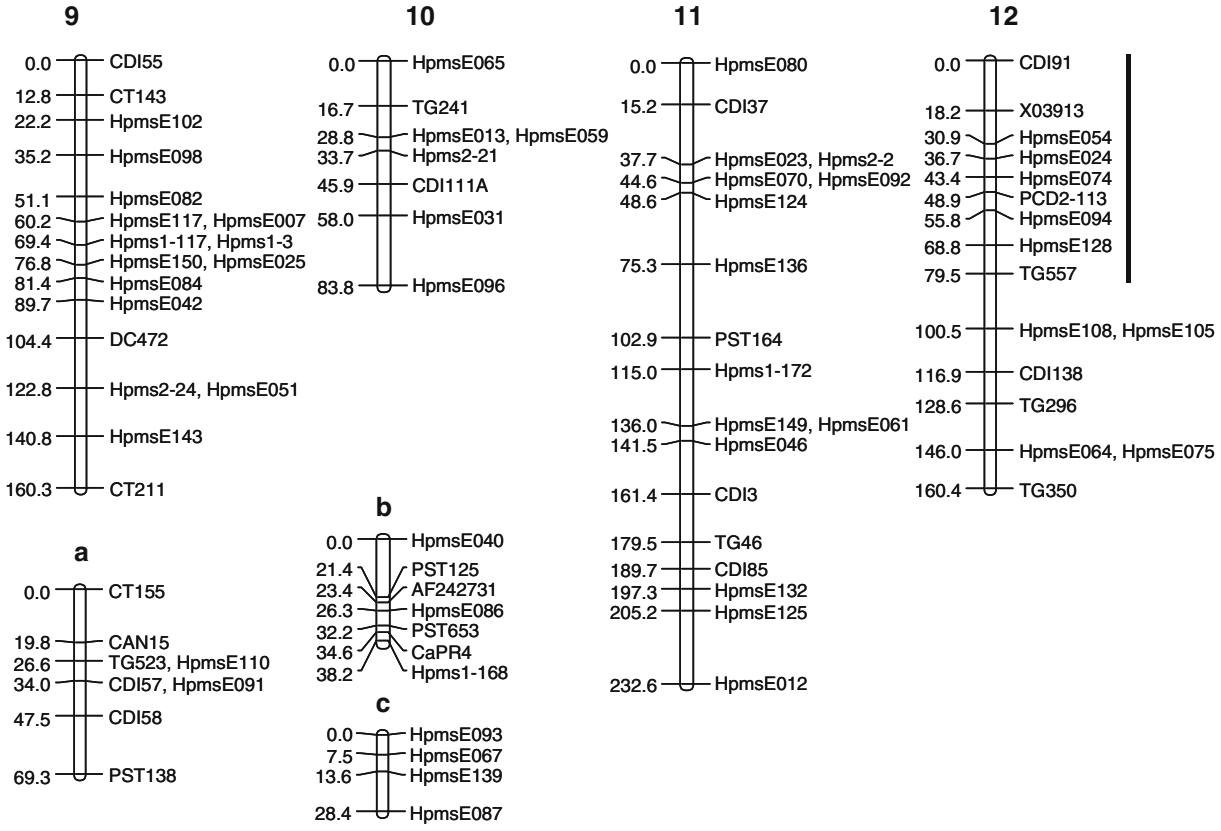

Fig. 2 continued

The 180 SSR and 63 RFLP markers were assigned to 14 LGs consisting of 10 large (123–303 cM) and 4 small LGs (28–83 cM), covering altogether 2,201.5 cM, slightly more than other pepper maps (Lefebvre et al. 2002; Livingstone et al. 1999; Paran et al. 2004). The average distance between markers was 9.1 cM and the EST–SSR markers were relatively evenly distributed throughout the genome. Every LG contained more than one EST–SSR marker, and the number of SSRs per LG ranged from 2 to 39. There was no chromosome 8 in the linkage map since chromosomes 1 and 8 could not be separated in an interspecific map because of pseudolinkage due to reciprocal translocation (Livingstone et al. 1999; Lee et al. 2004a). The previously designated LG 13 and LG 15 (Lee et al. 2004a) were merged with chromosomes 2 and 12, respectively in this study. The previous LG 14 and LG 16 were renamed LG a and LG b, respectively. A new LG containing only EST–SSRs was generated and named LG c.

Fifty (33.3%) of the 150 EST–SSRs deviated from the expected F_2 ratio of 1:2:1 or 3:1 at $P < 0.01$. 21 were skewed toward the 'TF68 parent', 20 toward 'Habanero', and 2 towards the heterozygote. One dominant EST–SSR was skewed toward 'Habanero'. Six of the skewed EST–SSR markers showed no linkage. The distorted markers were clustered or scattered on chromosomes 1, 2, 3, 7, 9, 11, 12, and LG a. The regions containing distorted markers were similar to those found in other interspecific mapping populations (Livingstone et al. 1999) except for the regions of chromosome 9 and LG a.

The EST–SSRs were obtained from a pepper-EST database, generated from a variety of tissues; leaves infected with *Xanthomonas campestris* pv. glycines, flower buds, anthers, young fruit, hairy root, early root, and green fruit placenta. The EST–SSR markers for which sequence homology has been established by BLAST search are listed in Table 4 (Altschul et al. 1997). Of the 150 EST–SSRs, 100 showed homology with putative gene sequences in 20 different organisms, and 71 of them corresponded to genes with known functions. Sequence similarity searches revealed defense-related genes and regulatory factors, as well as structural genes and genes involved in primary and secondary metabolism and signal transduction. The putative functions of 79 markers could not be established.

Table 4 Description of EST–SSR markers polymorphic between *C. annuum* 'TF 68' and *C. chinense* 'Habanero' and their functional annotation by tBLASTx

Marker name	Motif	Forward primer	Reverse primer	Product size (bp)	BLAST	Accession No.	E-value	Organism	LG
HpmsE001	(ATT)5	TGCCACCCATAAAATTCTTAAACCA	TGCAAGATCCCAAATTGAAATGA	207	-	-	-	-	2
HpmsE002	(CT)19	GCACGGCTCTATATCTCTCTATCTT	GGAACGAAATCGATCAACATCT	177	fructokinase	AAB51108	3e-172	*Solanum lycopersicon*	U
HpmsE003	(TTC)6	TTTCTGCAATTCCCCTTGTTCA	CAGCAGAGCCTTCAGTAGCAGC	164	expressed protein	NP_190719	1.9e-22	*A. thaliana*	2
HpmsE004	(CAT)13	TGGGAAGAGAAATTGTGAAAGCA	CAATGCCAACAATGGCATCCTA	137	glyoxalase family protein	NP_029429	1.5e-11	*A. thaliana*	1
HpmsE005	(CACCAT)3	TGCCTCAGTTTCCCAACCCT	ACCAACACCGTAACGCACCC	204	auxin-induced protein family	NP_567196	9.3e-27	*A. thaliana*	3
HpmsE006	(T)22 imp	GCTGACCGTTTTCGTTTTGGG	CAAAATTCAACCGCACCAACA	243	-	-	-	-	4
HpmsE007	(AT)12	CCCCATTTCCCCTTCCCATA	GAGGGGTCATGTTGAAGGCAA	134	-	-	-	-	9
HpmsE008	T14(AGT) 5(GTT)3	CCCCTTAACTTTTAATTCTAGATCTGC	TCGTTGTTCCTCCATCACC TCA	230	Unknown	AAM61476	2.7e-09	*A. thaliana*	3
HpmsE009	(A)23	TGCACAAACATCATACACCTCA	CCCATGACTGATAGTCCGGGTC	214	hypothetical protein	NP_849966	3.1e-48	*A. thaliana*	2
HpmsE010	(C)12	CTGTTTGCCAATCACCATCAGG	GCTATTTTCCGGCGTGTGAGAG	198	Unknown	AAM61098	2.4e-18	*A. thaliana*	3
HpmsE011	(GAA)5	GCAGAAGACCAAAGCCCTAGCA	TGGTTTCCATTGTCACTGTATGC	151	expressed protein	NP_188010	5.6e-24	*A. thaliana*	6
HpmsE012	(ATT) 10imp	AAACGCTGAAAAAGGCGTTGAC	TGCACCAACTTCTTCCATGCAC	208	U- box domain-containing protein	NP_566402	1.7e-36	*A. thaliana*	11
HpmsE013	(T)20	GCGCCAAGTGAGTTGAATTGAT	CACCAATCCGCTTGCTGTTGTA	256	methionine synthase	AAF74983	0	*Solanum tuberosum*	10
HpmsE014	(CGT)7	CTTTGGAACATTTCTTTGGGGG	GCGGACGTAGCAGTAGGTTTGG	106	fiber protein Fb19	AAR07598	2.4e-57	*Gossypium barbadense*	6
HpmsE015	(GCA)8	TTGTGAGGGTTTGACACTGGGA	CCGAGCTCGATGAGGATGAACT	146	AG- motif binding protein	BAC98495	1.6e-55	*Nicotiana tabacum*	5
HpmsE016	(TACA)3	CCAAGTTCAGGCCCAGGAGTAA	TGCAGAGAAGACTCACCAGTCC	187	lipoxygenase	T07065	4.1e-83	*S.tuberosum*	3
HpmsE017	(GCA)6	CTCGGAAATCAGATGCAGCAAA	GCATTGCATTGTTGGCAGTAGG	199	-	-	-	-	U
HpmsE018	(CAA) 5...A16	TGCCCAAATATCAGTCAAGCTCA	CAGCCAATTATGGCTTCT GCTG	267	bZIP transcription factor	AAN03468	2.4e-36	*Glycine max*	U
HpmsE019	(AT)9	AAGTCATCAGCTGCAAAGACCA	TTCAACATGCATCCAGCTTCTT	232	-	-	-	-	1
HpmsE020	(TTC)4	CCCCCGAGAGGAACAGAATCAT	TTCCATTTTGGTCCAGCTACCA	200	NHL repeat-containing protein	NP_196993	3e-06	*A. thaliana*	7

Table 4 continued

Marker name	Motif	Forward primer	Reverse primer	Product size (bp)	BLAST	Accession No.	E-value	Organism	LG
HpmsE021	(AAAT)4	CACACTAAGCATTCTGCTTTCACA	GGAGGGAATAGTAGCGGTTTGGA	250	N- hydroxycinnamoyl-CoA:tyramine N-hydroxycinnamoyl transferase	AAL99191	1e-116	*S. lycopersicon*	1
HpmsE022	(AGC)4	GCACCAGCATCAACATCAGCAT	CAGCAGGTGAAGGACTTGCAGA	206	-	-	-	-	1
HpmsE023	(AAT)8	TTTAACACCTCTCTAACCGTCACC	GCGATTTCAGCCCATCAACAAT	206	woun d-induced protein	P20144	2.6e-06	*S.tuberosum*	11
HpmsE024	(AAC)4	CGAGCCTAACCACCCAAATCAG	AAGGGAACGGAGGGACGACTAC	212	-	-	-	-	12
HpmsE025	(TG)9	TGAGCATCCCGTTATCTCAAATCA	CCCAATTCTTCAGGCAATCTCC	213	-	-	-	-	9
HpmsE026	(AGG)5	CCAAAGTCCATCGACGTCTCAA	ATCAAATGGCAAACCAGGAGGA	222	rubredoxin family protein	NP_175852	1.3e-46	*A. thaliana*	1
HpmsE027	(AAAC)5	TGGAGAATTGGTGTTACATGAAGG	TTCGGACCCTTCTCCATCACTT	230	phospholipid-transporting ATPase 1/aminophospholipid flippase 1/magnesium-ATPase 1	NP_568146	1.3e-14	*A. thaliana*	1
HpmsE028	(AAC)4	AGCCTGCATTTCTAATTGTTCACAT	CCTCATGGCATTTTGTCTTTGC	231	-	-	-	-	U
HpmsE029	(TCG)4	GATGGAGAAGATCGCCGACAAG	TACATCAGCAGGTTTGCCTCCA	202	reticulon family protein (RTNLB5)	NP_566065	2.6e-74	*A.thaliana*	1
HpmsE030	(GATT)5	GAAGCAGGGGCCAGAGCTAGA	GCCCCCAATTCTCAAACAGAGA	227	Avr9/C f-9 rapidly elicited protein 132	AAG43550	2.1e-18	*N. tabacum*	3
HpmsE031	(TC)6	CCCTAAATCAACCCCAAATTCAA	CCCCCATTACCTGACTGCAAAA	167	putative chromaffin granule ATPase II	NP_921451	9e-57	*Oryza sativa*	10
HpmsE032	(TG)6	ATGCGCAAAGGGAGAAAATTCA	CGAACTAACCGTTCATGGTGGA	231	-	-	-	-	U
HpmsE033	(AAATT)3	TGGATCCTCCTTTCTACTTCAACA	AAGGGTGGTGAAAAGGGGATTT	249	S- adenosyl-methionine-sterol-C-methyltransferase	AAB62808	2e-16	*N. tabacum*	1
HpmsE034	(ATC)6	AAACCGCTTGGTAATAAAGGTGCT	CCACTCCAGTTTTGATCATCTCCA	202	shock protein SRC2 homolog	T51602	5.7e-07	*A. thaliana*	1
HpmsE035	(AAG)7	GCTCAATCCATGATCGAGGACA	CTGCAGCTTGGAGTTCCTCTCC	226	nucleolar protein	CAA10127	8e-120	*Cicer arietinum*	1
HpmsE036	(CA)7	CCCACTCACACACACACACTGC	TTGAACCATCATTGAATCAAACAA	261	expre ssed protein	NP_565960	6.3e-21	*A. thaliana*	1
HpmsE039	(TGG)5	TCCCAAGTCAAGGGCAAGGATA	GGGGAATGGGGGATTTCATACA	252	60S acidic ribosomal pretein P2	P41099	5e-25	*Parthenium argentatum*	1
HpmsE040	(CCT)4	GGGAAAGAGAAGCCTTGGGTTG	CAAGTGGAGTCTGGGGAGGGTA	245	-	-	-	-	b

Table 4 continued

Marker name	Motif	Forward primer	Reverse primer	Product size (bp)	BLAST	Accession No.	E-value	Organism	LG
HpmsE045	(AAATT)3	GCTTGTTCAAGGAGCATTACGACA	TGTCCTTCCTTGCATTTGCATC	240	unknown protein	AAO22748	2.4e-10	*A. thaliana*	2
HpmsE046	(TC)8	TCCTCAAGATTGTTGTCATCATCA	AAAGGCATTTTCATTTTCGACTTT	277	-	-	-	-	11
HpmsE047	(GAA)5	AACCCGTGTTCAATCCCCAAAT	TGGCCATACCACCAGCAGTAGA	260	putative zinc finger protein	AAM62833	3.7e-59	*A. thaliana*	2
HpmsE048	(GAAA)3	CAAAGCTGAAGTTGGGCAAAGG	AGCATCATCATCAGGTCCAGCA	245	plectin-related protein-like	BAD68172	2e-32	*O. sativa*	1
HpmsE049	(AGC)13	CACTCCAACAGCAGCAGCAAAC	CCTTGCCGATGTTGAAGCTTTT	247	-	-	-	-	4
HpmsE050	(TC)9	CCCCACCTTCCATCATCAGCTA	TGGTAATTCTGCGGTCGATTCC	247	unknown protein		4.2e-09	*A. thaliana*	3
HpmsE051	(AT)12	TGGCCAGCTTCACACAGAGGTA	TGTCACAATATTGGAGGCCAGAA	262	-	-	-	-	9
HpmsE052	(AT)17	TCTCCAGGCTTTTCCTAACCAAGA	TCAACTGGCCTCAGTAAATTCTGC	221	-	-	-	-	7
HpmsE053	(GAA)6	TTCAAAGAATCCAGAGACTTCACA	TTCATGCAATTCCAAAGTCTCCA	197	4- coumarate:CoA ligase	BAA08366	2.7e-50	*Lithospermum erythrorhizon*	3
HpmsE054	(CT)6	GCCACCCCTCACCTCTCTCTCT	GTTGTTCGCTGGGCTCTTTCTC	219	-	-	-	-	12
HpmsE055	(AAC)6	GGCCGCTTAAAGTTGTTCAAGG	TGTGGCTAGCGGTGTTATGCAC	275	-	-	-	-	4
HpmsE056	(GAA)5	TCCGATTCAATCACTCCCAACA	GTATCGGCAATACGGGCAGAAG	214	expressed protein	NP_194884	1.3e-13	*A. thaliana*	1
HpmsE057	(TTTA)4	ACCCACTCCCTCTCCTCTTTGG	GCAGTGGAAAAACAGTCCTGTGG	211	-	-	-	-	7
HpmsE058	(CT)6	ATCACCAGCATCAATGGCTGTC	GGCTGAAACGGTTGTTTTACCG	202	-	-	-	-	U
HpmsE059	(ACT)4(AGG)5	GCAAGGACGCAGTCGTTAGACA	CCGCCTGTGCTGAATTGTTTAG	235	-	-	-	-	10
HpmsE060	(CT)7	GACAGGCAGATACCTCTCTCCAA	CGGAAAATGTTCGCTGATCACA	206	GAI- like protein 1	AAM19210	5.4e-58	*Vitis vinifera*	3
HpmsE061	(AAG)6	CCCAAGAAAGAAGTTGGGAATGG	TTCGACGAGCTTGGAAGGTGAT	249	PHP domain-containing protein	NP_565360	2.9e-29	*A. thaliana*	11
HpmsE062	(CCG)4	ATTTTTGCCCCAAATGTCAACG	AAACGTTGGGGGTGGAGATTG	191	-	-	-	-	1
HpmsE063	(AAC)7	CCAAATCATCATCAATGCGAAGA	CGGAGTAGATTTGGCGGAGAAA	209	-	-	-	-	1
HpmsE064	(AAT)4	CCCTCCTTTTACCTCGTCAAAAA	ATGCCAAGGAGCAATGAGAACC	221	unknown protein	AAN60250	6.6e-40	*A. thaliana*	12
HpmsE065	(TACA)3	TGAAATAGGCCAATCCCTTTGC	ATTCCCTGGGATTCCTGCATTA	199	rhizo me secoisolariciresinol dehydrogenase	AAK38664	2.3e-42	*Podophyllum peltatum*	10
HpmsE066	(AATT)4	AAGCGCGGTCACTGGAACTAAC	CGCCGGTTTTTCATCAATTACA	241	hypothetical protein F26O13.160	T45783	2.1e-39	*A. thaliana*	2
HpmsE067	(ATC)5	TCCACCAAAATCCCATTTCACA	ACATTGCCCACACCAATCACTG	212	L- lactate dehydrogenase	CAA71612	2.7e-45	*S. lycopersicon*	c
HpmsE068	(GTA)5	TGTTCCTTTTGTTGTTACCTTTTG	CGTCTAGGAATGGAAGAAGAGC	232	leucine-rich repeat family protein	NP_188391	1.5e-27	*A. thaliana*	7

Table 4 continued

Marker name	Motif	Forward primer	Reverse primer	Product size (bp)	BLAST	Accession No.	E-value	Organism	LG
HpmsE069	(CCA)4(ACC)10	CCTCAAGAATTTCACAACCCATGT	GATGGGTATTGGTGGTGGGGTA	2 16	hypothetical protein F3H9.7	A86409	6.3e-06		2
HpmsE070	(ATT)6	CACTCGTTATATTTTTCTGTCTCG	GTGAATATATCCGACCCTGTTT	215	expressed protein	NP_564266	3.3e-09	*A. thaliana*	11
HpmsE071	(ACC)5	CCCCTTCTCCTCCCTCATAAGC	TTCCATGATGTTACCGGAGCAA	188	-	-	-	-	4
HpmsE072	(AAC)7	GCTCATCAACCCACCTTCATCA	ATGCGTTGTCCGAGTAGGGAAG	199	expressed protein	NP_199491	6.9e-26	*A. thaliana*	6
HpmsE073	(TTC)5	TTATTCAGGCCCACTTATCGAA	CAGCAGCCAAATTCTTGATTTC	220	CaCBF1 A	AAQ88399	3.5e-34	*Capsicum annuum*	3
HpmsE074	(CAA)4	CGCCAGTTAACTTCCGACCTGT	GGAGCGGGTTGAGAGAGTTGAG	230	callus- expressing factor	AAP40022	7.2e-06	*N. tabacum*	12
HpmsE075	(ACC)6	GCGGCTCAGCAGAAAGAGAGAG	TGCCACAGCTGGAGAACGTAAA	205	zinc finger (AN1-like) family protein	NP_566429	6.5e-14	*A. thaliana*	12
HpmsE076	(AAC)4	CCATTCTTCTCCCCCAACAACA	CCATCACCAGCACCCTTCTCTT	239	-	-	-	-	6
HpmsE077	(AAC)9	TGGAGCTTGGATTAACCTTAGGA	CCATTATTTGTCCCTTCTTCACA	235	-	-	-	-	1
HpmsE078	(GAG)6	TTTGTGAAGAAGCAACCGGTGA	TGTGAGGAAGAAAGTGCGAAGG	203	-	-	-	-	6
HpmsE079	(AG)8	GGGCCTTATGCAACTAGTGTGTG	ATGGGCTGCAGGTTAGATGTGC	197	-	-	-	-	7
HpmsE080	(CCT)5	TGAGGGTGAGAACAAGTGTGGA	GTTATGGTGGACGGGGTACGAA	254	-	-	-	-	11
HpmsE081	(CA)8	TTGGCTATTGCCTCTCCAGACC	CATGTTGGGGAGAATGTGAACG	185	-	-	-	-	4
HpmsE082	(ATC)5	TTTTTCCCACTTTGCCCTTTCC	CAACCCAAGAAAACCCATTGGA	232	adenylate kinase, chloroplast precursor (ATP-AMP transphosphorylase)	Q8HSW1	4.8e-69	*A. thaliana*	9
HpmsE083	(CT)6	CTTCTGCAAGCACGCCTAATGA	TGCACCCATGAGACCATCTACC	209	OSJNBa0 086B14.17	CAD40846	1.5e-41	*O. sativa*	1
HpmsE084	(AAAAC)3	GCCAGAAGATCCATACTCTCATCA	GGAATGAGCAAAAACAAGAGTCC	220	-	-	-	-	9
HpmsE085	(ACC)4	TGCCCAAATATCAGTCAAGCTCA	TGGTTGTTGTTCTCATGGTGGTG	204	bZIP transcription factor ATB2	AAN03468	2.4e-36	*G. max*	4
HpmsE086	(TAG)6	GCAAGTCATGTGGTGACGGAAG	CCCTCGTGGCTCTCACTTGTTT	221	hypothetical protein	NP_193752	4e-06	*A. thaliana*	b
HpmsE087	(AT)10	CGATGGCTAAGGGAACGAGATG	TGCAATTTTCGATGATTTGTTTGA	247	unknown protein	AAN60317	6.4e-23	*A. thaliana*	c
HpmsE088	(GAG)6	GCAAATGGTTCCCTAAACTGCTT	GCTCTCCGTTTCCGATGTGATT	199	-	-	-	-	6
HpmsE089	(CTT)5	CCCCGTCAGTCTCTTCTTCCAA	GGCCATGTGAGTGAGACCACAG	240	phosphate transporter	AAL66293	1e-152	*G. max*	2
HpmsE090	(AAAG)4	CACCGGCCGTTTAAACTTTCAC	GCTGGAGTTTCCAAAGGACAAA	184	hypothetical protein	S16323	5.4e-09	*A. thaliana*	2
HpmsE091	(CCT)5	GCGATTACCAAACGGCGATCTA	CGGAACGGAAAACACCGTAATC	194	tyrosine specific protein phosphatase family protein	NP_171993	2.1e-28	*A. thaliana*	a
HpmsE092	(AT)10	CTCTGGCCCTTTTGTTCTTTCTTG	ACGCCTATTGCGAATTTCAGGA	180	expressed protein	NP_849712	8.4e-21	*A. thaliana*	11

Table 4 continued

Marker name	Motif	Forward primer	Reverse primer	Product size (bp)	BLAST	Accession No.	E-value	Organism	LG
HpmsE093	(CTT)4	TCTCTCTCCAATCCTTCCCTTCC	GAAGAAGGAGGTAGGGTTGTTTCG	207	-	-	-	-	c
HpmsE094	(TGG)5	CCAGTTGAGAGCTGCTGCAAAA	CACCAACAAAACAAAGGCCACA	241	fiber protein Fb14	AAP34361	2.8e-35	*G. barbadense*	12
HpmsE095	(TAGG)4	TGCTTAAACCCACTTGCGTCTTG	TCTGCACAGCACAAGACATTCG	184	zinc finger protein constans-like 7	Q9SYM2	6.9e-34	*A. thaliana*	7
HpmsE096	(TTC)4	CGGGTCAAACAAAAACCGAAGT	GCTTGTGGTTGAGCTCGCTCTT	237	P0468B 07.17	NP_915921	6.2e-12	*O. sativa*	10
HpmsE097	(CAA)5	GCTAAAACCTCCGCCGTAAACC	GCAATGGTGGCTGAGAAAATGA	250	-	-	-	-	U
HpmsE098	(ATC)4	CCACAAACACTCCCTAAACCCACT	CAAACCCCCTTCAACTTCCTCA	198	sing le-stranded DNA binding protein precursor	AAL39067	1.2e-24	*S. tuberosum*	9
HpmsE099	(TTC)5	CAATCATTGCCACCTTATTTTTGC	TCACAAGGGGTTGATGGAAATG	163	-	-	-	-	4
HpmsE100	(GCA)5	AGCAGGAAACAATGACCGGAAA	CCTACAGGAGCAATGCCTTTCG	220	putative FtsH protease	AAP32310	2.7e-87	*S. lycopersicon*	1
HpmsE101	(TAC)4	CCCTTACCCCTCCCTCCTTCTT	TCCGCTTGTGACGGACTTTGTA	167	mitoge n-activated protein kinase kinase	NP_177492	9.3e-10	*A. thaliana*	3
HpmsE102	(TTG)4	AAGGTGGTGGTATGGACTGCGTA	TTTAAGCGTTCAGTTGGCGAAA	163	-	-	-	-	9
HpmsE103	(AC)7	ATTGTGACCCGACTCCTCCAT	TGCTAATGGTGCTAATGCGGTA	177	thioredoxin family protein	NP_567765	3.5e-09	*A. thaliana*	7
HpmsE104	(GGC)4	TCCAATCATCGCTCCTTTACGG	CGAGCAACATGAGACGGTGAAT	212	-	-	-	-	1
HpmsE105	(AAC)5	CAACCAAATTGAAACCCCTCCA	CGTCCACCTAATAATGCGACCA	238	-	-	-	-	12
HpmsE106	(AAT)9	AGGGATCAGCTCCGAGGTTCTT	ATGA++GCAGCGTTGCCATGATTA	168	-	-	-	-	3
HpmsE107	(GCA)6	TCCCAATTCCACTTTCCTCCAA	TCCACGGTGTTTGGTGGTGTTA	184	-	-	-	-	2
HpmsE108	(TC)16	GGCCGCGTCAACCCTAAAAA	TGGGAGGGAAAAGTAAGCGAAA	182	poly(A)- binding protein	AAF63202	3.1e-65	*Cucumis sativus*	12
HpmsE109	(AAC)4	TGTTGCTGAGGACAACACTAACGA	TCACCGCTGTAGAACTGGTTGC	226	expressed protein	NP_174162	1e-05	*A. thaliana*	2
HpmsE110	(CTT)4	TGCTCGAAATGACAAAGTTTGGTT	AGCAATGTGTTGGACAGGCTGA	191	-	-	-	-	a
HpmsE111	(CT)7	CCATCATTTCTCCCCAATTCCA	GAGAGCAGAAGAAGGGGTGGTG	219	-	-	-	-	4
HpmsE112	(GTA)6	GGCATCGTCCGGGATACTACAA	TCGTTGTTGTTGTCGTCATCGTT	206	BEL1- related homeotic protein 29	AAN03626	1.4e-41	*S. tuberosum*	1
HpmsE113	(GCA)4	CCCTAAAGCTCGAGAAATTGAAGC	GAATGCTGTTGCTGGGGTTGTT	225	-	-	-	-	6
HpmsE114	(AAG)6	GGTGAGGGAGGTGTGAGCAAA	GATCCACATACGCCATCACTGC	190	express ed protein	NP_565449	1.9e-39	*A. thaliana*	7
HpmsE115	(ATTA)5	TCATCTCATAGCCTGCCCCCTA	CCACTTGAAGAAGCCATGACCA	216	bZIP DNA-binding protein	AAN61914	1e-74	*Capsicum chinense*	1
HpmsE116	(CTT)12	CATCTCTCCGTTGAATCTATTTCC	ACGGTCATCCATTAGAACCGTA	189	coronatine-insensitive 1/ COI1 (FBL2)	NP_565919	7.5e-19	*A. thaliana*	5
HpmsE117	(CA)6	GCAAATGCTGCCACTCACTTTC	ACAAGGGGTGTTTCCTCATCCA	198	expressed protein	NP_187558	7.8e-06	*A. thaliana*	9
HpmsE118	(CCA)7	CAGGCGATAGAGGCAAAATTGG	TCCGGAACAGTTTCACATGGAG	193	-	-	-	-	2
HpmsE119	(CT)11	TGTTCTGGACTGCTGCTCTTCG	TGACCAACAACCTCATCATCCAA	241	unknown protein	BAB02808	8.2e-15	*A. thaliana*	3
HpmsE120	(CCA)6	GGGGGAGGAAGAGAAGAAGTCG	CCGGACTTTACGAGCACAACCT	209	ATFP3	AAD09507	6.8e-28	*A. thaliana*	6

Table 4 continued

Marker name	Motif	Forward primer	Reverse primer	Product size (bp)	BLAST	Accession No.	E-value	Organism	LG
HpmsE121	(TTC)7	TGAGAGTTGGATCCCTCTTCTTCG	CTTGTTCGGGCTTGAAAGAGGA	198	carbon-nitrogen hydrolase family protein	NP_196765	1.3e-40	*A. thaliana*	1
HpmsE122	(ACG)4	GCAATGGCTCAGGTCTCCATCT	TGTCGCCCTTTAATGCAAAACC	231	-	-	-	-	3
HpmsE123	(TTTC)6	CGGTAGCTGTCGTGTTCTTGAG	TCAATACAGGAGGCAAACATGG	180	putative mitogen-activated protein kinase	NP_179409	2.5e-24	*Prunus armeniaca*	2
HpmsE124	(CAA)10	TCGTTAGCAGCAACAACAACAA	CTGCTTCCTGTACATGGCTGTC	227	60S ribosomal protein L24	O65743	1.2e-59	*C. arietinum*	11
HpmsE125	(GA)6	AGGGAGTTCACGCCATTTTTGA	GCGATAACCCACCGGAGAAAAT	204	-	-	-	-	11
HpmsE126	(ATC)4	TGGGTATTTCCTTGCTGGAGAA	CCTTTTCAGGCACCACTTTGAT	192	lipoxygenase (EC 1.13.11.12) loxD	T07409	3.8e-48	*S. lycopersicon*	3
HpmsE127	(GAG)4	CCACCCCATGTAAGCCCATCT	AATTTCGCGTCCTCAGGAGGTT	277	28 kDa ribonucleoprotein, chloroplast precursor (28RNP)	P19682	2e-63	*Nicotiana sylvestris*	3
HpmsE128	(ATC)5	TGGATCCCAAAAGACTCAGAACA	TATTTCCCTCAGTCGAGGTCGT	226	calcium/calmodulin protein kinase 1	AAN71903	9.7e-41	*N. tabacum*	12
HpmsE129	(CAA)4	CACCGACCACAACTTCAGCAAA	TCCCAAATCCTCCTCCCTTCTC	233	-	-	-	-	5
HpmsE130	(AT)11	ATTTGGGGACTTCACGTCTCTC	CCCATCTTCTTCTTTCTCCACAA	221	MADHp ms-box protein 9	AAQ72497	1.5e-61	*Petunia x hybrida*	U
HpmsE131	(TCA)9	TGGATTATTTATGGAGCCATGC	GGATCATATCCATTTTTGTTGGAT	246	serine carboxypeptidase II-3 precursor (CP-MII.3)	P52711	4.6e-47	*Hordeum vulgare*	1
HpmsE132	(AAAAC)3	ACATCCACAGCAAAAGGAAAAA	GTGGAAAGTTTGGTGGATCAGA	197	major intrinsic protein 2	CAB46351	3e-127	*S. tuberosum*	11
HpmsE133	(ATG)4	CGACGAAGTTTCCGAGCTCAA	GACACGGCGCTTCTTTCCTC	205	transcriptional coactivator p15 (PC4) family protein (KELP)	NP_192830	8.2e-47	*A. thaliana*	U
HpmsE134	(ATTT)3	TCCTGCGATTACTCCTCCTTCG	CGAATTCATAACAACCGAAGCAAA	235	-	-	-	-	2
HpmsE135	(CTG)4	TCGGCCCATTTAACCAAAACAC	TGGATCGGTTCGTTTAGGAGGA	209	zinc finger (DHHC type) family protein	NP_850638	1.1e-35	*A. thaliana*	2
HpmsE136	(TTG)4	TTTTGGCCTCTCGCTTTCAAAC	GGCAACTGACCAAAAACTGAAAAA	294	-	-	-	-	11
HpmsE137	(GGT)4	GAGGTGGTAGCCGTGGTTATGG	TGCAGTTTGTGGCAAAGTGTCC	189	glycine-rich protein 2	P27484	2.7e-13	*N. sylvestris*	1
HpmsE138	(GAC)4... (TGA)4	AAGAAGGGGGAGATGATGACGA	TCCTCAAGCCATGATA GAGAAGCA	250	-	-	-	-	7
HpmsE139	(GAC)10	CGACGAAGGTGATAGCGACGAT	CTCGCCATTTTCTTCCGGTTC	160	expressed protein	NP_564517	3.8e-08	*A. thaliana*	c
HpmsE140	(CTT)4	GGCTCTGCCTCTCGTCTCCTC	AGGATCAGAAGCAGCGCATTTC	225	thylakoid-bound ascorbate peroxidase	BAA78552	4e-17	*N. tabacum*	4

Table 4 continued

Marker name	Motif	Forward primer	Reverse primer	Product size (bp)	BLAST	Accession No.	E-value	Organism	LG
HpmsE141	(ATC)6	TCCCAACAACTCAAATGGCTTC	TGGAGGTGCCCTTCTGGTAAAG	206	urease accessory protein G	CAC33001	3.6e-68	*S. tuberosum*	2
HpmsE142	(CCG)6	AGATTCCGGCAGTCCGGTTACT	TCCGGTAAGCTGCCTTGATCTC	214	DnaJ protein homolog	T11855	1.4e-32	*Phaseolus vulgaris*	2
HpmsE143	(AAAAT)3	CCATTCAGCTAGGGTTCAGTCCA	CGACCAAATCGAATCTTCGTGA	107	transducin family protein/ WD-40 repeat family protein	NP_194712	3.2e-14	*A. thaliana*	9
HpmsE144	(GAA)4...(CTT)4... (ATG)4	CCGAACCAGAAAATGGCTCATC	GACTCTATCGCTCCCCGACTCA	236	-	-	-	-	2
HpmsE145	(TTA)4	TGAGGGTATTTTCGTCATTTCAC	GAAAGCGGAAAACATTAAGAGTCA	222	-	-	-	-	U
HpmsE146	(AAG)4...(AAT)4	CCCTTCTTCCTTTCCACCATCA	AGGCGTGAAAGGGTTA TGAGGA	227	senescence-associated protein-related	NP_565167	7.5e-22	*A. thaliana*	4
HpmsE147	(GA)6	ACCAAAAAACGCATGCACACATC	CGGTCAATCTCCGATGGATTTC	178	-	-	-	-	U
HpmsE148	(CT)6	AAACATCCCGCGAAGGCAACT	CAGGGAGCCAAATTACGAGGTG	205	photosystem II stability/ assembly factor HCF136	NP_197703	1.1e-25	*A. thaliana*	2
HpmsE149	(TC)15(AC)18	CGGAAACTAAACACACTTTCTCTC	GACTGGACGCCAGTTTGATT	198	-	-	-	-	11
HpmsE150	(CTT)7	CCCTCTTCCCCGACTCTCTCTT	AAGCCAATGACTGCATGACCAC	219	mRNA binding protein precursor	T52071	2.8e-53	*S. lycopersicon*	9

imp imperfect repeat - not significantly found U unlinked

Discussion

We have developed SSR markers derived from a pepper EST database by in silico mining, and constructed an SSR-based pepper linkage map. A similar approach has been applied in various plant species (Varshney et al. 2005). Although genic SSR markers are generally less polymorphic than genomic ones, this disadvantage is balanced by other advantages. The value of EST–SSRs compared to genomic SSRs is enhanced by their transferability across taxa, and their potential as functional markers in defining genes affecting traits of interest. In silico mining of sequence databases allows EST–SSR marker development at a relatively low cost, and requires limited time and labor. Therefore, EST databases may be considered rich and valuable resources for SSR marker and map development.

From our data mining, we found 1,042 SSR–ESTs (10.2%) by screening 10,232 non-redundant pepper ESTs (4.56 Mb), with a frequency of one SSR every 3.8 kb. This frequency is higher than our previous result from GenBank sequence analysis, which yielded one genic SSR every 6.7 kb (Lee et al. 2004a). It is consistent with the finding of Morgante et al. (2002) that the frequency of SSRs was higher (1.5–2-fold) in ESTs than in genomic DNA across all species despite the relatively small size of the EST sequences. Previous reports and the availability of a pepper EST database encouraged us to pursue the development of EST–SSRs in pepper.

In SSR analysis from a sequence database the first step is to define the SSRs. Since the criteria for SSRs depend on the parameters of the SSR search algorithm (for example, minimum length of SSR, gap, and mismatch penalties), exact comparison of SSR characteristics between different plants is complicated. Despite the different search criteria reported for SSR mining in EST databases, the percentage of EST–SSRs found in this study was similar to that (ranging from ~2 to ~16%) observed in other species (Gao et al. 2003; Morgante et al. 2002; Temnykh et al. 2000).

Kantety et al. (2002) analyzed SSR distributions in EST databases of barley, maize, rice, sorghum, and wheat. They identified SSRs with a minimum length of 18 (di, tri) to 20 (tetra) bp. In the pepper EST database, 241 SSRs satisfied this criterion. This corresponded to 2.4% of the ESTs, a lower frequency than the average in those five crops (3.2%). The percentages of each motif were 42% dinucleotide, 52% trinucleotide, and 6% tetranucleotide. The proportion of dinucleotide repeat was larger than in the other five crops.

Cardle et al. (2000) screened SSRs from several plants with length of dinucleotide repeat >15 bp, tri- >15, tetra- >16, and penta- >20. Only 3% of Arabidopsis ESTs contained SSRs. Three-hundred and sixty-four SSRs (3.6%) satisfied this criterion in pepper, made up of 130 dinucleotide, 204 trinucleotide, 23 tetranucleotide, and 7 pentanucleotide repeats. There was one SSR per every 12.52 kb in the pepper EST database and one SSR per every 13.83 kb in the *Arabidopsis* EST database. The abundant motifs were AG, AAG, ACC, AAC, and ATC in that order in pepper, compared with AAG, AG, ATC, and AAC in *Arabidopsis*. The numbers of SSRs with the different trinucleotide motifs were similar in pepper, whereas, AAG and ATC motifs were more abundant than the others in *Arabidopsis*.

Kumpatla and Mukhopadhyay (2005) mined and surveyed EST–SSRs in 49 dicot species as well as in pepper. In that study, the criterion for SSRs was less stringent: mononucletide SSRs and 5 dinucleotide repeats were included. Therefore, mononucleotides were most frequent, followed by dinucleotides and trinucleotides, in contrast with the results of our survey of pepper EST–SSRs. Even though their criteria for SSRs differed from ours, several outcomes are almost identical. First, 9.64% of the ESTs contained SSRs. Second, the AG motif was the most abundant in dinucleotides and AAG in trinucleotides. These characteristics represent a general trend in the EST–SSRs of dicot plants, whereas CCG is the most abundant trinucleotide in monocot plants (Temnykh et al. 2001). To avoid incorrect mining of poly(A) mRNA and to enhance the level of polymorphism, our stringent criteria for SSR will be suitable for EST–SSR marker development, since the probability of polymorphism increases with increasing length of SSR (Cho et al. 2000; Temnykh et al. 2001).

The mining of pepper SSR-EST data in this study revealed that 66% of SSR–ESTs were trinucleotide repeats. This is in close agreement with other observations in monocot and dicot plants (Kantety et al. 2002; Morgante et al. 2002) and has been attributed to negative selection against frameshift mutations in coding regions (Metzgar et al. 2000), and possibly to positive selection for stretches of particular amino acids (Morgante et al. 2002). In contrast, the non-coding regions of eukaryotic genomes have been found to contain mainly dinucleotide repeats (Tóth et al. 2000). Morgante et al. (2002) and Varshney et al. (2005) demonstrated that the AG motif was the most frequent in dinucleotide SSRs in plant ESTs, and our results are in total agreement with theirs. Although the functions of the SSRs in ESTs are still not clear and some characteristics of the EST–SSRs are quite divergent in different plant species, our results demonstrate that the general trends of EST–SSR characteristics are very similar in plants.

We amplified 513 EST–SSRs in both parental lines in reactions using 813 EST–SSR primer pairs (63.1%). Rates of SSR amplification have generally ranged from 60 to 90% in plants (Varshney et al. 2005). The amplification rate of pepper EST–SSRs is very similar to that of barley EST-SSTs (64%) (Thiel et al. 2003) and somewhat lower than the 73% reported in rice (Temnykh et al. 2001) and the 83% reported in tomato (He et al. 2003). The lower number of EST–SSR primer sets showing efficient amplification in this study may be due to the loss of priming sites at intron–exon splice sites, the use of two genetically distant *Capsicum* species as the mapping population, limitations of the primer design software, or primer development from sequences of low quality. The polymorphism level in the parental lines is 29.2%. Although we used an interspecific cross between *C. annuum* and *C. chinense* to generate the mapping population, the polymorphism level was lower than what might be expected. A low level of polymorphism (19.8%) in an interspecific mapping population was, however, also reported in cotton EST–SSRs (Park et al. 2005). The slightly lower level of polymorphism may be due to self-pollination of domesticated pepper crops and sequence conservation of genic regions. The level of polymorphism of SSRs often increases with increasing SSR length and number of repeat units (Ellegren 2004; Sharapova et al. 2002; Temnykh et al. 2001). The polymorphism level in rice decreased in SSRs of less than 18 bp (Cho et al. 2000). Similarly, in this study, EST–SSRs of more than 18 bp had higher polymorphism levels than those with less than 18 bp (42.9% vs. 17.7%). Dinucleotides were more polymorphic than tri-, tetra-, and pentanucleotides. This is in agreement with previously reported results from other plant and animal species, which have reported low variability of most trinucleotide SSR loci (Chakraborty et al. 1997; Schug et al. 1998). In general, EST–SSR markers have been reported to show lower polymorphism than SSR markers derived from genomic libraries (Cho et al. 2000; Eujayl et al. 2002). Our previous data indicated that genomic pepper SSRs (75%) were twofold more polymorphic than genic SSRs (36%) (Lee et al. 2004a). In the present study, the polymorphism level of the EST–SSRs was slightly lower than that of the previous genic SSRs. Although these EST–SSRs are less polymorphic than genomic SSRs, several EST–SSRs were found to be highly polymorphic in a study of variety identification in pepper (Kwon et al. 2005).

We observed an expansion of map length. Maps of pepper range from about 1,300 to 1,800 cM (Lee et al. 2004a; Lefebvre et al. 2002; Livingstone et al. 1999; Paran et al. 2004). The linkage map of the EST–SSR markers is larger than these, since it consists mainly of genic markers, and recombination may be more frequent in gene-rich regions than in non-coding regions. In pepper, as in many other SSR-based maps in plants, the SSR markers are distributed relatively evenly over the genome, and provide good coverage of all linkage groups. There are no obvious biases in terms of the chromosomal locations of SSR markers with different motifs. Nevertheless, there are some regions of the pepper genome where EST–SSRs are more frequent. These regions seem to be gene-rich or euchromatic regions. The frequency and map positions of the distorted EST–SSR markers in this study are quite similar to those of previous RFLP and SSR markers (Kang et al. 2001; Lee et al. 2004a). In general, distorted markers that grouped together were skewed in the same direction. Segregation distortion has been described in many plant DNA marker-based linkage maps, and has been attributed to causes such as segregation distortion, deleterious recessive alleles, self-incompatibility alleles, structural rearrangements, and differences in DNA content (Moretzsohn et al. 2005; Taylor and Ingvarsson 2003). Two small linkage groups in the previous study were merged into their respective chromosomes. A small linkage group was formed anew in this study. There were three small linkage groups that could not be assigned to chromosomes. More markers, which will be developed in the future would allow these small LGs to find their respective chromosomes. Direct hybridization of the markers as probes to chromosome by FISH technique could also be helpful, which we plan to pursue. Public or private organizations pursuing pepper breeding program are interested in dissecting agronomically important traits such as fungal and viral pathogen resistance, yield, and fruit quality. To understand complex traits especially at the molecular level, segregating populations and DNA marker systems are essential. In this respect, the cDNA-based markers will not only provide codominant markers, but may also include candidate genes that are linked to, or could be targeted to, the trait of interest (Gupta and Rustgi 2004). Thus, the mapping using EST–SSRs may provide useful information for understanding monogenic or polygenic traits and could be utilized in candidate gene approaches (Pflieger et al. 2001). The availability of large numbers of EST–SSR or cDNA-based markers should allow the construction of a high-resolution and marker-dense transcriptional map in the near future.

Currently, there are about 30,000 unigene sets in the pepper EST database (http://www.genepool.pdrc.re.kr). Mining SSRs in smaller sets of database derived from individual ESTs may produce some

degree of redundancy. Redundancy arises because several ESTs could be transcribed from one mRNA. To avoid this, researchers need to identify SSRs in consensus sequences that represent unique ESTs. In developing new EST–SSR markers in pepper, therefore, nonredundant EST datasets should be used after clustering newly developed ESTs with the ESTs obtained in this study.

Publicly available markers such as SSRs can facilitate the development of a consensus map, so that map information can be readily shared without comprehensive mapping experiments. The SSR markers in our previous SNU2 map (Lee et al. 2004a) were used to help assign pepper chromosome numbers to unknown linkage groups in a separate pepper map (Ogundiwin et al. 2005). These additional EST–SSR markers may greatly accelerate mapping projects that are under way in pepper as well as in other Solanaceae. The availability of these markers will also benefit other applications such as cultivar identification (Kwon et al. 2005) and marker-assisted selection in pepper.

Acknowledgements This work was supported by grants from the Agricultural R & D Promotion Center (ARPC), and from the Center for Plant Molecular Genetics and Breeding Research (CPMGBR) through the Korea Science and Engineering Foundation (KOSEF) and the Korea Ministry of Science and Technology (MOST).

References

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

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

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci USA 94:1041–1046

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

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

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435–445

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

Ewing B, Gree P (1998b) Basecalling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186–194

Ewing B, Hiller L, Wendle MC, Green P (1998a) Base-calling of automated sequencer traces using Phred. I Accuracy assesement. Genome Res 8:175–185

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467

Gao L, Tang J, Li H, Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 12:245–261

Gupta PK, Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct Integr Genomics 4:139–162

He C, Poysa V, Yu K (2003) Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. Theor Appl Genet 106:363–373

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

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

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

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Kumar LS (1999) DNA markers in plant improvement: an overview. Biotechnol Adv 17:143–182

Kumpatla SP, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48:985–998

Kwon YS, Lee JM, Yi GB, Yi SI, Kim KM, Soh EH, Bae KM, Park EK, Song IH, Kim BD (2005) Use of SSR markers to complement tests of distinctiveness, uniformity, and stability (DUS) of pepper (*Capsicum annuum* L.) varieties. Mol Cells 19:428–435

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

Lee JM, Nahm SH, Kim YM, Kim BD (2004a) Characterization and molecular genetic mapping of microsatellite loci in pepper. Theor Appl Genet 108:619–627

Lee S, Kim SY, Chung E, Joung YH, Pai HS, Hur CG, Choi D (2004b) EST and microarray analyses of pathogen-responsive genes in hot pepper (*Capsicum annuum* L.) non-host resistance against soybean pustule pathogen (*Xanthomonas axonopodis* pv. glycines). Funct Integr Genomics 4:196–205

Lefebvre V, Pflieger S, Thabuis A, Caranta C, Blattes A, Chauvet JC, Daubeze AM, Palloix A (2002) Towards the saturation of the pepper linkage map by alignment of three intraspecific maps including known-function genes. Genome 45:839–854

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

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

Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10:72–80

Miyao A, Zhong HS, Monna L, Yano M, Yamamoto K, Havukkala I, Minobe Y, Sasaki T (1996) Characterization and genetic mapping of simple sequence repeats in the rice genome. DNA Res 3:233–238

Moretzsohn MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SC, Gimenes MA, Martins WS, Valls JF, Grattapaglia D, Bertioli DJ (2005) A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). Theor Appl Genet 111:1060–1071

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 nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Nahm SH, Yu JW, Kang BC, Kim BD (1997) Selection of parental lines for hot pepper mapping population using RFLP and AFLP analysis. J Kor Soc Hortic Sci 38:693–697

Ogundiwin EA, Berke TF, Massoudi M, Black LL, Huestis G, Choi D, Lee S, Prince JP (2005) Construction of 2 intraspecific linkage maps and identification of resistance QTLs for *Phytophthora capsici* root-rot and foliar-blight diseases of pepper (*Capsicum annuum* L.). Genome 48:698–711

Paran I, van der Voort JR, Lefebvre V, Jahn M, Landry L, van Schriek M, Tanyolac B, Caranta C, Ben-Chaim A, Livingstone K, Palloix A, Peleman J (2004) An integrated genetic linkage map of pepper (*Capsicum* spp.). Mol Breed 13:251–261

Park YH, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, el-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol Genet Genomics 274:428–441

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breed 7:275–291

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST databases. Genome Res 9:167–174

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

Rudd S (2003) Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci 8:321–329

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TF, Aquadro CF (1998) The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. Mol Biol Evol 15:1751–1760

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 Jr (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48:463–481

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

Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce SR, Grandbastien MA (2005) Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. Theor Appl Genet 110:819–831

Taylor DR, Ingvarsson PK (2003) Common features of segregation distortion in plants and animals. Genetica 117:27–35

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

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

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

Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10:967–981

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55