ORIGINAL PAPER

An interspecific linkage map of SSR and intronic polymorphism markers in tomato

Kenta Shirasawa · Erika Asamizu · Hiroyuki Fukuoka · Akio Ohyama · Shusei Sato ·
Yasukazu Nakamura · Satoshi Tabata · Shigemi Sasamoto · Tsuyuko Wada · Yoshie Kishida ·
Hisano Tsuruoka · Tsunakazu Fujishiro · Manabu Yamada · Sachiko Isobe

Abstract Despite the collection and availability of abundant tomato genome sequences, PCR-based markers adapted to large scale analysis have not been developed in tomato species. Therefore, using public genome sequence data in tomato, we developed three types of DNA markers: expressed sequence tag (EST)-derived simple sequence repeat (SSR) markers (TES markers), genome-derived SSR markers (TGS markers) and EST-derived intronic polymorphism markers (TEI markers). A total of 2,047 TES, 3,510 TGS and 674 TEI markers were established and used in the polymorphic analysis of a cultivated tomato (*Solanum lycopersicum)* 'LA925' and its wild relative *Solanum pennellii* 'LA716', parents of the Tomato-EXPEN 2000 mapping population. The polymorphic ratios between parents revealed by the TES, TGS and TEI markers were 37.3, 22.6 and 80.0%, respectively. Those showing polymorphisms were used to genotype the Tomato-EXPEN 2000 mapping population, and a high-density genetic linkage map composed of 1,433 new and 683 existing marker loci was constructed on 12 chromosomes, covering 1,503.1 cM. In the present map, 48% of the mapped TGS loci were located within heterochromatic regions, while 18 and 21% of TES and TEI loci, respectively, were located in heterochromatin. The large number of SSR and SNP markers developed in this study provide easily handling genomic tools for molecular breeding in tomato. Information on the DNA markers developed in this study is available at http://www.kazusa.or.jp/tomato/.

Communicated by I. Paran.

K. Shirasawa · E. Asamizu · S. Sato · Y. Nakamura · S. Tabata ·
S. Sasamoto · T. Wada · Y. Kishida · H. Tsuruoka · T. Fujishiro ·
M. Yamada · S. Isobe (✉)
Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari,
Kisarazu, Chiba 292-0818, Japan
e-mail: sisobe@kazusa.or.jp

E. Asamizu
Gene Research Center, University of Tsukuba,
Ten-no dai 1-1-1, Tsukuba, Ibaraki 305-8572, Japan

H. Fukuoka · A. Ohyama
National Institute of Vegetable and Tea Science,
360 Kusawa, Ano, Tsu, Mie 514-2392, Japan

Y. Nakamura
National Institute of Genetics, 1111 Yata, Mishima,
Shizuoka 411-8540, Japan

Introduction

The tomato, *Solanum lycopersicum*, which originated in Latin America, is the second most important vegetable crop and is cultivated throughout the world (Foolad 2007). Tomato belongs to the family Solanaceae, which consists of approximately 100 genera and 2,500 species, including several plants of agronomic importance such as potato, eggplant, pepper and tobacco (Olmstead et al. 2008). Tomato has a relatively compact genome within the Solanaceae species, characterized by its diploidy ($2n = 2X = 24$). It is approximately 950 Mb in size, and is one of the most intensively characterized Solanaceae genomes (Arumuganathan and Earle 1991). The International Tomato Sequencing Project, established in 2004 with members from 10 countries, promotes structural genome analysis in tomato by sequencing the gene-rich regions of all 12 chromosomes through the generation of high quality sequences from bacterial artificial chromosomes (BAC). The data are released

immediately after completion, principally through the SOL Genomics Network (SGN) website (http://solgenomics.net/; Mueller et al. 2005). In addition, ESTs and full-length cDNAs of a miniature tomato cultivar (*S. lycopersicum*) named Micro-Tom have been published at MiBASE (http://www.kazusa.or.jp/jsol/microtom/) and KafTom (http://www.pgb.kazusa.or.jp/kaftom/) (Yamamoto et al. 2005). Currently, more than 320,000 ESTs and 461,000 BAC-end sequences have been registered in the databases.

Genetic diversity in the cultivated tomato is generally low, due to the occurrence of population bottlenecks during the domestication and the generation of modern varieties (Rick 1976). Therefore, during the past quarter century, several linkage maps of the tomato genome have been developed with more than 20 mapping populations derived from crosses between cultivated tomato (*S. lycopersicum*) and its wild relatives, such as *S. pennellii*, *S. pimpinellifolium*, *S. cheesmaniae*, *S. neorickii*, *S. chmielewskii*, *S. habrochaites* and *S. peruvianum* (reviewed by Foolad 2007). In particular, an F_2 mapping population named Tomato-EXPEN 1992, which was derived from a cross between *S. lycopersicum* cv. VF36 and the inbred accession of *S. pennellii* LA716, facilitated the development of a high-density restriction fragment length polymorphism (RFLP) linkage map with 1,030 loci (Tanksley et al. 1992). Subsequently, other types of DNA markers, such as random amplified polymorphic DNA (RAPD), simple sequence repeats (SSRs), amplified fragment length polymorphisms (AFLPs), cleaved amplified polymorphic sequence (CAPS) and single nucleotide polymorphisms (SNPs), were developed and mapped onto the Tomato-EXPEN 2000 mapping population derived from the cross between *S. lycopersicum* LA925 and *S. pennellii* LA716 (Frary et al. 2005; Fulton et al. 2002). Currently, the Tomato-EXPEN 2000 map comprises a total of 2,604 markers including 1,088 CAPS, 1,342 RFLP, 155 SSR and 19 SNP markers (http://solgenomics.net/cview/map.pl?map_id=9; Fulton et al. 2002). In addition, a total of 349 PCR-based markers are now available at Tomato Mapping Resource Database (http://www.tomatomap.net), and mapped on interspecific crosses between *S. pimpinellifolium* and *S. lycopersicum*.

Most of the loci mapped on the two Tomato-EXPEN mapping populations consist of RFLP and CAPS markers, which result in a higher cost for the identification of polymorphisms due to their laborious nature and the need for restriction enzymes. The rapid progress in genome analysis during the past several years has enabled large scale segregation analysis in molecular genetics. We therefore considered it essential to develop PCR-based markers designed to adapt large scale genotyping systems, such as SSR and SNP markers, for the promotion of genetics and genomics in tomato species. Recently, Ohyama et al. (2009) developed SSR markers using BAC-end and cDNA sequences in tomato and mapped a total of 148 SSR loci onto the EXPEN 2000 map. They investigated a total of 89,824 cDNA and 310,583 BAC-end sequences to generate SSR markers. Their results suggested that abundant sequence resources would allow the generation of a larger number of PCR-based markers in tomato. In this study, we developed three types of DNA markers, namely EST-derived SSR markers (TES markers), genome-derived SSR markers (TGS markers) and EST-derived intronic polymorphism markers (TEI markers), using the public genome sequence data. As Ohyama et al. (2009) reported that most of the SSR markers derived from BAC-end sequences mapped to the centromeric regions, we used different polymorphic sources for marker generation. The developed markers were mapped onto the Tomato-EXPEN 2000 mapping population to fill in the gaps in knowledge within and between tomato genomics and genetics.

Materials and methods

Plant material

A previously reported F_2 mapping population, which was derived from a cross between *S. lycopersicum* LA925 and the inbred accession *S. pennellii* LA716 (Fulton et al. 2002), was used for linkage map construction. The mapping parents and 83 F_2 individuals were kindly provided by Prof. S. Tanksley from Cornell University. Total DNA was extracted from the leaves of each plant using the DNeasy Plant Mini kit (Qiagen, Germany).

Development of TES (*t*omato *E*ST-*S*SR) and TGS (*t*omato *G*enome-*S*SR) markers

Microsatellite or SSR regions were identified in tomato ESTs and BAC-end sequences registered in public databases, namely MiBASE (http://www.kazusa.or.jp/jsol/microtom/), KafTom (http://www.pgb.kazusa.or.jp/kaftom/) and SGN (http://sgn.cornell.edu/). SSRs longer than 14 bases, which contained all possible combinations of dinucleotide (NN), trinucleotide (NNN) and tetranucleotide (NNNN) repeats, were identified using the FINDPATTERNS module in the GCG software package (Accelrys Inc., USA). Oligonucleotides for PCR primers were designed based on the flanking regions of the identified SSRs using the Primer3 program (Rozen and Skaletsky 2000) in such a way that the amplified products ranged between 90 and 300 bp in length. Markers corresponding to those previously developed by Fulton et al. (2002) and Ohyama et al. (2009) were identified based on the associated nucleotide sequences and excluded from our collection

of EST-derived SSR (TES) and BAC-end sequences-derived SSR (TGS) markers.

PCR was performed using 0.5 ng tomato genomic DNA in a 5 µl reaction mix containing 1× PCR buffer (BIO-LINE, UK), 3 mM MgCl$_2$, 0.04 U BIOTAQ^{TM} DNA polymerase (BIOLINE, UK), 0.2 mM dNTPs and 0.8 µM of each primer. A modified 'touchdown PCR' protocol was used, as described previously (Sato et al. 2005). PCR products were separated either on a 10% polyacrylamide gel with TBE buffer according to the standard protocol or in a Type 3730 DNA fragment analyzer (Applied Biosystems, USA). In the latter case, the data were analyzed using GeneMapper software (Applied Biosystems, USA).

Development of TEI (*tomato E*ST-derived *i*ntronic polymorphism) markers

The positions of introns were predicted by comparisons between KTU2 EST unigene sequences in MiBASE and *Arabidopsis thaliana* genomic sequences (http://www.arabidopsis.org/) using the GAP2 program (Huang 1994). Oligonucleotides for PCR primers were designed based on the flanking regions of the predicted intron positions in the tomato unigenes using the Primer3 program (Rozen and Skaletsky 2000). PCR was performed as described above. Polymorphic PCR amplicons were identified by high-resolution melting (HRM) analysis (Palais et al. 2005) using the LightScanner system (Idaho Technology, USA) with 0.5 µl of 10× LCGreen Plus + Melting Dye (Idaho Technology) as the fluorescent dye and 0.05 ng of genomic DNA of *S. pennellii* 'LA716' as the control DNA.

Linkage analysis

Linkage analysis was performed using genotypic data derived from the markers developed in this study along with data derived from a total of 683 published markers. These markers included 449 RFLPs, 61 CAPS markers and 173 SSR markers (Fulton et al. 2002; Ohyama et al. 2009), which were retrieved from the SGN (http://sgn.cornell.edu/) and VegMarks (http://vegmarks.nivot.affrc.go.jp/) databases. The genotypic data of the mapping population generated from the published markers were kindly provided by Prof. S. Tanksley at Cornell University and Dr. A. Ohyama at the National Institute of Vegetable and Tea Science of Japan (NIVTS). Henceforth, we refer to these as Cornell markers and NIVTS markers, respectively.

Linkage analysis was performed using the JoinMap® program version 4 (Van Ooijen 2006). The map positions of the Cornell markers reported by Fulton et al. (2002) were used as a frame for the linkage analysis. The genotyped markers were roughly classified into 12 linkage groups, which corresponded to the previous EXPEN 2000 map,

using the grouping module of JoinMap® and based on an LOD score of 4.0–10.0. Marker order and genetic distance were calculated using a regression mapping algorithm with the following parameters: Kosambi's mapping function, recombination frequency ≤ 0.35, LOD score ≥ 2.0. Genotype probabilities which show possible genotyping and data entry errors were investigate for all the mapped markers and presented as −log10(P). Euchromatic and heterochromatic regions in each linkage group were assumed based on the previously reported positions of markers anchored to heterochromatin (Frary et al. 2005; Ohyama et al. 2009; Tang et al. 2008; Wang et al. 2006).

Results

Development of SSR markers

A total of 7,599 SSR markers were generated by in silico data mining of 83,785 sequences, as described in the "Materials and methods", and designated TES (*T*omato *E*ST-*S*SR) markers. Of these generated markers, those corresponding to the Cornell and NIVTS markers were excluded from the TES marker group. Of the SSRs in the TES marker group, 6,043 (80%) were trinucleotide repeats, while 525 (7%) were dinucleotide repeats and 1,031 (13%) were tetranucleotide repeats (Table 1). The poly (AAG)$_n$ motif was the most abundant of the trinucleotide repeats (1,712 SSRs, 28%), followed by poly (ATC)$_n$ (919 SSRs, 15%), poly (AGC)$_n$ (708 SSRs, 12%) and poly (AAC)$_n$ (641 SSRs, 11%). While three types of dinucleotide repeats were observed, poly (AT)$_n$ and poly (AG)$_n$ were the most abundant and comprised 96% of the dinucleotide repeats. Among the tetranucleotide repeats, AT-rich motifs, namely poly (AAAT)$_n$, poly (AAAG)$_n$ and poly (AAAC)$_n$, were more frequently observed than other motifs and when combined, comprised up to 59% of the tetranucleotide repeats.

In addition to the TES markers, genome-derived SSR markers were generated by extracting SSRs from the 90,763 BAC-end sequences retrieved from the SGN database. A total of 13,501 primer pairs were designed to amplify the SSRs, which were subsequently named TGS (*T*omato *G*enome-*S*SR) markers. Among the TGS markers, 7,005 of the SSRs (52%) were trinucleotide repeats, while 2,338 (17%) were dinucleotide and 4,158 (31%) were tetranucleotide repeats (Table 1). Poly (AAT)$_n$ was the most abundant trinucleotide repeat (2,517 SSRs, 36%), followed by poly (AAG)$_n$ (1,771 SSRs, 25%), poly (AAC)$_n$ (1,012 SSRs, 14%) and poly (ATC)$_n$ (691 SSRs, 10%). Four types of dinucleotide repeats were observed in the SSRs of the BAC-end sequences, and poly (AT)$_n$ was the most frequently observed motif (71% of the identified dinucleotide repeats). Similar to the TES markers, AT-rich motifs

Table 1 Numbers of SSR motifs in the TES and TGS markers

Motif		TES		TGS	
		Designed (%)	Mapped (%)	Designed (%)	Mapped (%)
Dinucleotide	AC	61 (0.8)	4 (0.6)	272 (2.0)	27 (4.3)
	AG	229 (3.0)	39 (6.0)	408 (3.0)	35 (5.5)
	AT	235 (3.1)	44 (6.8)	1,653 (12.2)	192 (30.3)
	GC	0 (0.0)	0 (0.0)	5 (0.0)	0 (0.0)
Trinucleotide	AAC	641 (8.4)	56 (8.6)	1,012 (7.5)	38 (6.0)
	AAG	1,712 (22.5)	147 (22.7)	1,771 (13.1)	82 (12.9)
	AAT	539 (7.1)	84 (13.0)	2,517 (18.6)	117 (18.5)
	ACG	121 (1.6)	9 (1.4)	41 (0.3)	[1] (0.2)
	ACT	188 (2.5)	19 (2.9)	248 (1.8)	11 (1.7)
	AGC	708 (9.3)	34 (5.2)	199 (1.5)	5 (0.8)
	ATC	919 (12.1)	59 (9.1)	691 (5.1)	17 (2.7)
	GGA	491 (6.5)	22 (3.4)	253 (1.9)	5 (0.8)
	GGC	230 (3.0)	19 (2.9)	44 (0.3)	1 (0.2)
	GGT	494 (6.5)	45 (6.9)	229 (1.7)	[5] (0.8)
Tetranucleotide	AAAC	160 (2.1)	11 (1.7)	403 (3.0)	9 (1.4)
	AAAG	274 (3.6)	12 (1.9)	690 (5.1)	10 (1.6)
	AAAT	272 (3.6)	20 (3.1)	1,892 (14.0)	45 (7.1)
	AACG	3 (0.0)	0 (0.0)	11 (0.1)	0 (0.0)
	AAGC	40 (0.5)	1 (0.2)	44 (0.3)	3 (0.5)
	AATC	56 (0.7)	5 (0.8)	165 (1.2)	4 (0.6)
	AATG	47 (0.6)	1 (0.2)	194 (1.4)	0 (0.0)
	AATT	71 (0.9)	8 (1.2)	607 (4.5)	23 (3.6)
	AGGC	2 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)
	AGGT	2 (0.0)	0 (0.0)	14 (0.1)	0 (0.0)
	GACG	2 (0.0)	0 (0.0)	4 (0.0)	0 (0.0)
	GACT	5 (0.1)	1 (0.2)	11 (0.1)	0 (0.0)
	GAGC	8 (0.1)	0 (0.0)	6 (0.0)	0 (0.0)
	GATC	9 (0.1)	2 (0.3)	4 (0.0)	0 (0.0)
	GGAC	2 (0.0)	0 (0.0)	4 (0.0)	0 (0.0)
	GGAT	27 (0.4)	0 (0.0)	25 (0.2)	1 (0.2)
	GGCA	5 (0.1)	0 (0.0)	6 (0.0)	0 (0.0)
	GGCC	2 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)
	GGCT	6 (0.1)	1 (0.2)	3 (0.0)	0 (0.0)
	GGGA	22 (0.3)	2 (0.3)	36 (0.3)	1 (0.2)
	GGGC	0 (0.0)	0 (0.0)	3 (0.0)	1 (0.2)
	GGGT	16 (0.2)	3 (0.5)	34 (0.3)	1 (0.2)
Total		7,599 (100)	648 (100)	13,501 (100)	634 (100)

were more frequently observed in the tetranucleotide repeats of the TGS markers. The combination of poly $(AAAT)_n$, poly $(AAAG)_n$ and poly $(AATT)_n$ motifs comprised 86% of the tetranucleotide repeats.

From all the generated SSR markers, 2,047 TES and 3,510 TGS markers were examined for polymorphisms between the parents of the EXPEN mapping population. A total of 451 TES markers (22.0% of all tested markers) and 229 TGS markers (6.5%) showed co-dominant polymorphisms between the parents. In addition, 313

TES markers (15.3% of all tested) and 566 TGS markers (16.1%) showed dominant polymorphisms. Therefore, 41% [313/(451 + 313)] of all TES markers and 71% [566/(229 + 566)] of all TGS markers were dominant. In most of the dominant markers, null alleles were observed in *S. pennellii* (LA716). The polymorphic ratios for the di-, tri- and tetranucleotide repeats among the TES markers were 13.5, 75.7 and 10.9%, respectively, while those for the TGS markers were 40.8, 43.8 and 15.5%, respectively.

Development of TEI (*tomato EST-derived Intronic* polymorphism) markers

A total of 589 unigenes retrieved from a KTU2 unigene set in the MiBASE database were aligned with *A. thaliana* genome sequences using the GAP2 program (Huang 1994). By taking the GU-AG rule into account, the positions of 1,073 introns were predicted for 206 of the unigene sequences. The mean number of predicted introns per single unigene was 5.2. Primer pairs were designed using sequences of the flanking regions of the predicted intron positions to generate DNA markers named TEI (*tomato EST-derived intronic* polymorphism).

A total of 674 TEI markers were tested for polymorphisms using HRM analysis, as described in the "Materials and methods", and 537 of them (80%) showed polymorphisms between the mapping parents. These 537 TEI markers represented 166 independent unigenes. Therefore, 166 non-redundant TEI markers were selected for subsequent linkage analysis. Of these selected TEI markers, 148 and 18 showed co-dominant and dominant polymorphisms between the parents, respectively. Though the SNPs were identified by HRM, it is sometimes not possible to identify polymorphisms between heterozygous and one parental specific homozygous SNPs, due to the flanking sequences of the SNPs. These kinds of SNPs were scored as dominant SNPs.

Construction of a linkage map

Segregation data were generated for a total of 1,725 markers (764 TES, 795 TGS and 166 TEI markers) in the Tomato-EXPEN 2000 mapping population. These data were combined with data derived from 547 Cornell and 136 NIVTS markers and subjected to linkage analysis. The map positions of the 547 Cornell markers reported by Fulton et al. (2002) were used as a frame for the construction of the linkage map. Subsequently, 2,116 loci, including 634 TGS, 648 TES, 151 TEI, 36 NIVTS and 547 Cornell loci, were mapped onto 12 linkage groups corresponding to the 12 chromosomes (Chrs), while 292 loci, including 161 TGS, 116 TES and 15 TEI loci, were excluded from the analysis. The total length of the linkage groups was 1,503.1 cM, as shown in Table 2, Fig. 1 and the Supplementary Table. The total number of newly mapped loci was 1,433, ranging from 105 to 160 in each linkage group. The average distance between two loci was 0.71 cM ranging from 0.59 cM (Chr 9) to 0.88 cM (Chr 5).

Of the 2,116 loci mapped in this study, 1,481 and 635 loci were assumed to be located in euchromatic and heterochromatic regions, respectively, according to the position of markers anchored to heterochromatin reported in previous studies (Frary et al. 2005; Ohyama et al. 2009; Tang et al. 2008; Wang et al. 2006). The markers used as anchor markers of euchromatic and heterochromatic regions are listed in the Supplemental Table. In the present map, 48% (305) of the mapped TGS loci were predicted to be present in the heterochromatic regions, while only 18% (119) and 21% (32) of the TES and the TEI loci were in the same regions.

Segregation distortion was observed for 38.2% of the mapped marker loci (Fig. 1, Table 2). The distortion ratios varied from chromosome to chromosome; Chr 4, 5, 8, 9 and 12 showed distortions for less than 10% of the mapped loci, while Chr 1, 10 and 11 showed distortions for more than 70% of the loci. Heterochromatin-specific segregation distortion was observed for Chr 2 and 7. The 'LA716' (*S. pennellii*) genotypes were more frequently observed in the distorted loci mapped to Chr 1, 2, 3, 6, 10 and 11, while more 'LA925' (*S. lycopersicum*) genotypes were identified in the distorted loci mapped to Chr 7.

Discussion

In this study, we developed three types of DNA markers designated TES, TGS and TEI. The TES markers that contained trinucleotide repeats exhibited a higher ratio of polymorphisms than those containing di- and tetranucleotide repeats. On the other hand, a significant difference was not observed in the frequency of polymorphisms between di- and trinucleotide repeats in TGS markers. Dinucleotide repeats in coding regions often causes critical changes, such as frame-shift mutations. The decreased number of polymorphisms present in the dinucleotide repeats in TES markers in comparison to TGS markers suggested that the coding region contained higher sequence conservation than the intergenic regions of *S. lycopersicum* and *S. pennellii*. Among the polymorphic markers, the percentage of dominant TES markers was 41% [313/(451 + 313)], while that of dominant TGS markers was 71% [566/(229 + 566)]. In most of the dominant markers, null alleles were observed in *S. pennellii*. As the primers were designed from the *S. lycopersicum* genome and EST sequences, sequence divergence between the two species might cause poor annealing of the primers to the *S. pennellii* genomic DNA. The results also indicate that sequence conservation is higher in the intragenic regions than the intergenic regions.

A segregation distortion ratio was observed for more than 70% of loci mapped to Chr 1, 10 and 11, while in less than 10% of loci mapped to Chr 4, 5, 8, 9, and 12. Segregation distortion or transmission ratio distortion (TRD) in a hybrid population can result from various factors, such as hybrid lethality or sterility of gametophytic competition (Harushima et al. 2001). The bias of segregation distortion can be explained by the presence of more TRD factors on
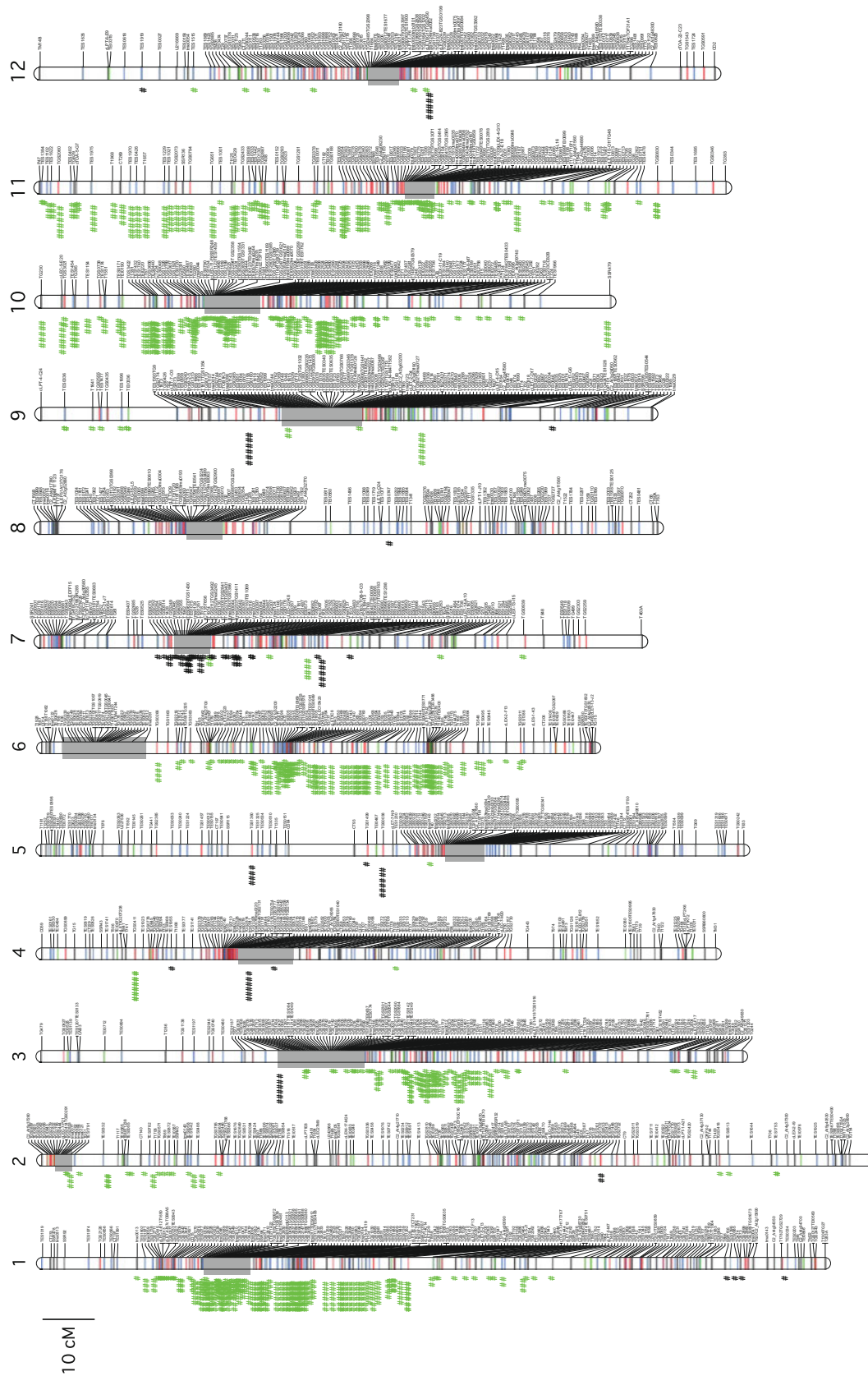
Fig. 1 Genetic linkage map of the tomato genome. Twelve linkage groups were constructed using 634 TGS markers (*red lines*) and 683 reported markers (*black lines*). The gray box in each linkage group indicates the estimated heterochromatic region. The distorted loci are indicated by *hatch symbols* in *green* (increases in *S. pennellii* alleles) and *black* (increases in *S. lycopersicum* alleles), as determined using the Chi-square test (#$P < 0.05$, ##$P < 0.01$, ###$P < 0.001$, ####$P < 0.005$, #####$P < 0.0005$ and ######$P < 0.0001$). *Scale bar* 10 cM

Table 2 Markers in the euchromatic, heterochromatic and entire chromosomal regions of the tomato EXPEN 2000 map

Chromosome	Length (cM)	Number of loci						Average distance between two loci (cM)	Segregation distortion ratio (%)[a]
		TGS	TES	TEI	Cornell	NIVTS	Total		
Entire region									
chr01	153.7	75	72	13	77	16	253	0.61	70.8
chr02	162.6	36	56	18	71	6	187	0.87	18.7
chr03	137.5	65	73	12	34	12	196	0.70	43.4
chr04	132.3	47	49	9	56	10	171	0.77	8.8
chr05	137.9	50	54	9	31	12	156	0.88	5.8
chr06	108.5	46	49	18	56	3	172	0.63	63.4
chr07	107.3	50	50	11	43	11	165	0.65	40.0
chr08	110.1	41	53	11	43	11	159	0.69	1.3
chr09	109.0	58	56	15	39	17	185	0.59	8.6
chr10	101.5	57	41	13	36	11	158	0.64	94.9
chr11	122.3	49	48	12	29	12	150	0.82	84.7
chr12	120.4	60	47	10	32	15	164	0.73	9.8
Total	1,503.1	634	648	151	547	136	2,116	0.71	38.2
Euchromatic region[b]									
chr01	144.4 (94)	44 (59)	62 (86)	12 (92)	71 (77)	5 (31)	194 (77)	0.74	61.9
chr02	159.9 (98)	33 (92)	56 (100)	18 (100)	70 (96)	2 (33)	179 (96)	0.89	16.2
chr03	120.5 (88)	25 (38)	46 (63)	7 (58)	33 (58)	2 (17)	113 (58)	1.07	55.8
chr04	118.7 (90)	22 (47)	38 (78)	8 (89)	50 (70)	1 (10)	119 (70)	1.00	10.1
chr05	128.5 (93)	17 (34)	47 (87)	6 (67)	22 (61)	3 (25)	95 (61)	1.35	9.5
chr06	90.1 (83)	30 (65)	42 (86)	14 (78)	49 (79)	1 (33)	136 (79)	0.66	79.4
chr07	100.1 (94)	28 (56)	46 (92)	10 (91)	37 (75)	3 (27)	124 (75)	0.81	24.2
chr08	103.8 (94)	25 (61)	50 (94)	7 (64)	36 (79)	7 (64)	125 (79)	0.83	1.6
chr09	94.8 (87)	19 (33)	33 (59)	11 (73)	25 (50)	4 (24)	92 (50)	1.03	15.2
chr10	90.1 (89)	21 (37)	29 (71)	7 (54)	25 (53)	1 (9)	83 (53)	1.09	92.8
chr11	116.7 (95)	30 (61)	37 (77)	9 (75)	26 (68)	0 (0)	102 (68)	1.14	79.4
chr12	116.4 (97)	35 (58)	43 (91)	10 (100)	27 (73)	4 (27)	119 (73)	0.98	12.6
Total	1,384.8 (92)	329 (52)	529 (82)	119 (79)	471 (70)	33 (24)	1,481 (70)	0.94	37.8
Heterochromatic region[b]									
chr01	9.3 (6)	31 (41)	10 (14)	1 (8)	6 (23)	11 (69)	59 (23)	0.16	100.0
chr02	2.7 (2)	3 (8)	0 (0)	0 (0)	1 (4)	4 (67)	8 (4)	0.34	75.0
chr03	17 (12)	40 (62)	27 (37)	5 (42)	1 (42)	10 (83)	83 (42)	0.20	26.5
chr04	13.6 (10)	25 (53)	11 (22)	1 (11)	6 (30)	9 (90)	52 (30)	0.26	5.8
chr05	9.4 (7)	33 (66)	7 (13)	3 (33)	9 (39)	9 (75)	61 (39)	0.15	0.0
chr06	18.5 (17)	16 (35)	7 (14)	4 (22)	7 (21)	2 (67)	36 (21)	0.51	2.8
chr07	6.3 (6)	22 (44)	4 (8)	1 (9)	6 (25)	8 (73)	41 (25)	0.15	87.8
chr08	6.3 (6)	16 (39)	3 (6)	4 (36)	7 (21)	4 (36)	34 (21)	0.19	0.0
chr09	14.2 (13)	39 (67)	23 (41)	4 (27)	14 (50)	13 (76)	93 (50)	0.15	2.2
chr10	11.4 (11)	36 (63)	12 (29)	6 (46)	11 (47)	10 (91)	75 (47)	0.15	97.3
chr11	5.6 (5)	19 (39)	11 (23)	3 (25)	3 (32)	12 (100)	48 (32)	0.12	95.8
chr12	4 (3)	25 (42)	4 (9)	0 (0)	5 (27)	11 (73)	45 (27)	0.09	2.2
Total	118.3 (8)	305 (48)	119 (18)	32 (21)	76 (30)	103 (76)	635 (30)	0.19	39.2

[a] Significant at $P < 0.05$

[b] Numbers in parentheses indicate percentage of the entire region

some chromosomes than on others, and the number of markers that were linked to those TRD. In addition, genotyping errors may have resulted in segregation distortion on some markers. The segregation distortion of loci is often reported in interspecific crosses between cultivated tomato and its wild relatives. In a cross between *S. lycopersicum* and its close relative *S. pimpinellifolium*, 8% of loci indicated segregation distortion (Grandillo and Tanksley 1996), while in wider crosses such as *S. lycopersicum* × *S. cheesmaniae* and *S. lycopersicum* × *S. neorickii,* 51 and 69% of loci showed segregation distortion, respectively (Paterson et al. 1988, 1991). Rieseberg et al. (1995) reported that TRD in hybrid populations potentially represents some level of reproductive isolation due to chromosome rearrangements or genetic interactions. The results of our study can therefore be considered indicative of genetic diversity between *S. lycopersicum* and *S. pennellii* that is wider in Chr 1, 10 and 11, and narrower in Chr 4, 5, 8, 9 and 12. The total length of the linkage map developed in this study was 1,503 cM, which was slightly longer than the 1,460 cM of the Tomato-EXPEN 2000 map reported by Fulton et al. (2002). However, the difference in the total length was considered to be within the error range of the linkage analysis, and the map was well saturated because most of the end markers mapped on each linkage were Cornell markers. A total of 292 markers were excluded from the analysis and not mapped. We speculated that this exclusion might be due to genotyping error, for example, if markers producing non-specific bands were screened as polymorphic markers. The average physical distance per cM was calculated to be 632 kb, deduced from the entire genome size of 950 Mb (Arumuganathan and Earle 1991). The length of the euchromatic and heterochromatic regions were 1,384 and 118 cM, respectively, and the physical distance per cM in both regions was estimated to be 172 and 6,042 kb, respectively, assuming a proportion of euchromatin and heterochromatin of 25% (238 Mb) and 75% (713 Mb), respectively, of the entire genome (Peterson et al. 1998). These results suggest that the physical distance per cM in the heterochromatin region is approximately 35 times longer than in the euchromatin region.

The average distance between two loci in euchromatin was 0.94 cM, which was five times longer than in heterochromatin, which measured 0.19 cM. In the present map, 48% of the mapped TGS loci were located within heterochromatic regions, while 18 and 21% of TES and TEI loci, respectively, in heterochromatin regions. These results agree with previous reports (Frary et al. 2005; Ohyama et al. 2009) showing that genome-derived SSR markers are clustered within heterochromatin, while EST-derived SSR markers tend to be randomly distributed across the chromosomes in tomato. The heterochromatin region comprises 75% of the tomato genome with the gene density 10–100 times lower

than that of euchromatin (Wang et al. 2006). Moreover, recombination events are suppressed in the heterochromatic regions in tomato (Sherman and Stack 1995; Peters et al. 2009). These characteristics of the tomato genome could explain the clustering of TGS markers into such short genetic distances. We therefore conclude that EST-derived markers are superior in their ability to cover the entire genome compared to markers derived from randomly selected genomic regions, as described by Ohyama et al. (2009).

In this study, we performed a large scale PCR-based marker development and newly mapped a total of 1,433 SSR and intronic polymorphism loci on the Tomato-EXPEN 2000 mapping population. Information on these DNA markers is available at http://www.kazusa.or.jp/tomato/. The DNA markers developed in this study are scheduled to be applied to anchoring scaffold sequences in an ongoing whole genome shotgun sequencing project (Mueller et al. 2005). The large number of SSR and intronic polymorphic markers developed in this study provide easily handled and abundant seed points for tomato genome sequencing.

The EXPEN 2000 mapping population was generated from cross between a cultivated tomato (*S. lycopersicum*) and its wild relative, *S. pennellii* (Fulton et al. 2002). As genetic diversity in *S. lycopersicum* is low (Miller and Tanksley, 1990), saturated linkage maps have not been constructed in intraspecific mapping populations of *S. lycopersicum* (reviewed by Foolad 2007) and required more DNA markers. The large number of PCR-based markers developed in this study should be a useful genomic resource for interspecific genetic analysis in *S. lycopersicum*, although there is a possibility that some markers might not contain polymorphisms in intraspecific mapping. In addition, the existence of the dense linkage map allows map-based cloning of genes from wild relatives, such as abiotic stress tolerance and disease resistance genes, with the utilization of whole genome sequence that will be available soon. Once the utilizing genes are identified, the large amount of available markers should facilitate selection in introgressive breeding of *S. lycopersicum* and its wild relatives. The DNA markers and linkage map developed in this study are expected to enhance breeding by allowing a better understanding of whole genome sequences in tomato.

References

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

Foolad MR (2007) Genome mapping and molecular breeding of tomato. Int J Plant Genomics 2007:64358

Frary A, Xu Y, Liu J, Mitchell S, Tedeschi E, Tanksley S (2005) Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their usefulness for genetics and breeding experiments. Theor Appl Genet 111:291–312

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

Grandillo S, Tanksley SD (1996) Genetic analysis of RFLPs, GATA microsatellites and RAPDs in a cross between *L. esculentum* and *L. pimpinellifolium*. Theor Appl Genet 92:957–965

Harushima Y, Nakagahra M, Yano M, Sasaki T, Kurata N (2001) A genome-wide survey of reproductive barriers in an intraspecific hybrid. Genetics 159:883–892

Huang X (1994) On global sequence alignment. Comput Appl Biosci 10:227–235

Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. Theor Appl Genet 80:437–448

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD (2005) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. Plant Physiol 138:1310–1317

Ohyama A, Asamizu E, Negoro S, Miyatake K, Yamaguchi H, Tabata S, Fukuoka H (2009) Characterization of tomato SSR markers developed using BAC-end and cDNA sequences from genome databases. Mol Breed 23:685–691

Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia V, Collier SM (2008) A molecular phylogeny of the Solanaceae. Taxon 57:1159–1181

Palais RA, Liew MA, Wittwer CT (2005) Quantitative heteroduplex analysis for single nucleotide polymorphism genotyping. Anal Biochem 346:167–175

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335(6192):721–726

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127:181–197

Peters SA, Datema E, Szinay D, van Staveren MJ, Schijlen EG, van Haarst JC, Hesselink T, Abma-Henkens MH, Bai Y, de Jong H, Stiekema WJ, Klein Lankhorst RM, van Ham RC (2009) *Solanum lycopersicum* cv. Heinz 1706 chromosome 6: distribution and abundance of genes and retrotransposable elements. Plant J 58:857–869

Peterson DG, Pearson WR, Stack SM (1998) Characterization of the tomato (*Lycopersicon esculentum*) genome using in vitro and in situ DNA reassociation. Genome 41:346–356

Rick CM (1976) Tomato, *Lycopersicon esculentum* (Solanaceae). In: Simmonds NW (ed) Evolution of crop plants. Longman, London, pp 268–273

Rieseberg LH, Linder CR, Seiler GJ (1995) Chromosomal and genic barriers to introgression in Helianthus. Genetics 141:1163–1171

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Sato S, Isobe S, Asamizu E, Ohmido N, Kataoka R, Nakamura Y, Kaneko T, Sakurai N, Okumura K, Klimenko I, Sasamoto S, Wada T, Watanabe A, Kohara M, Fujishiro T, Tabata S (2005) Comprehensive structural analysis of the genome of red clover (*Trifolium pratense* L.). DNA Res 12:301–364

Sherman JD, Stack SM (1995) Two-dimensional spreads of synaptonemal complexes from Solanaceous plants. VI. High-resolution recombination nodule map for tomato (*Lycopersicon esculentum*). Genetics 141:683–708

Tang X, Szinay D, Lang C, Ramanna MS, van der Vossen EA, Datema E, Lankhorst RK, de Boer J, Peters SA, Bachem C, Stiekema W, Visser RG, de Jong H, Bai Y (2008) Cross-species bacterial artificial chromosome-fluorescence in situ hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements. Genetics 180:1319–1328

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. Genetics 132:1141–1160

Van Ooijen JW (2006) JoinMAP® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V, Wageningen

Wang Y, Tang X, Cheng Z, Mueller L, Giovannoni J, Tanksley SD (2006) Euchromatin and pericentromeric heterochromatin: comparative composition in the tomato genome. Genetics 172:2529–2540

Yamamoto N, Tsugane T, Watanabe M, Yano K, Maeda F, Kuwata C, Torki M, Ban Y, Nishimura S, Shibata D (2005) Expressed sequence tags from the laboratory-grown miniature tomato (*Lycopersicon esculentum*) cultivar Micro-Tom and mining for single nucleotide polymorphisms and insertions/deletions in tomato cultivars. Gene 356:127–134