Theor Appl Genet (2012) 125:47–56
DOI 10.1007/s00122-012-1815-9

ORIGINAL PAPER

Development of gene-based markers and construction of an integrated linkage map in eggplant by using *Solanum* orthologous (SOL) gene sets

Hiroyuki Fukuoka · Koji Miyatake · Tsukasa Nunome · Satomi Negoro · Kenta Shirasawa · Sachiko Isobe · Erika Asamizu · Hirotaka Yamaguchi · Akio Ohyama

Received: 15 December 2011 / Accepted: 31 January 2012 / Published online: 16 February 2012

Abstract We constructed an integrated DNA marker linkage map of eggplant (*Solanum melongena* L.) using DNA marker segregation data sets obtained from two independent intraspecific F_2 populations. The linkage map consisted of 12 linkage groups and encompassed 1,285.5 cM in total. We mapped 952 DNA markers, including 313 genomic SSR markers developed by random sequencing of simple sequence repeat (SSR)-enriched genomic libraries, and 623 single-nucleotide polymorphisms (SNP) and insertion/deletion polymorphisms (InDels) found in eggplant-expressed sequence tags (ESTs) and related genomic sequences [introns and untranslated regions (UTRs)]. Because of their co-dominant inheritance and their highly polymorphic and multi-allelic nature, the SSR markers may be more versatile than the SNP and InDel markers for map-based genetic analysis of any traits of interest using segregating populations derived from any intraspecific crosses of practical breeding materials. However, we found that the distribution of microsatellites in the genome was biased to some extent, and therefore a considerable part of the eggplant genome was first detected when gene-derived SNP and InDel markers were mapped. Of the 623 SNP and InDel markers mapped onto the eggplant integrated map, 469 were derived from eggplant unigenes contained within *Solanum* orthologous (SOL) gene sets (i.e., sets of orthologous unigenes from eggplant, tomato, and potato). Out of the 469 markers, 326 could also be mapped onto the tomato map. These common markers will be informative landmarks for the transfer of tomato's more saturated genomic information to eggplant and will also provide comparative information on the genome organization of the two solanaceous species. The data are available from the DNA marker database of vegetables, VegMarks (http://vegmarks.nivot.affrc.go.jp).

Communicated by I. Paran.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1815-9) contains supplementary material, which is available to authorized users.

H. Fukuoka (✉) · K. Miyatake · T. Nunome · S. Negoro · H. Yamaguchi · A. Ohyama
NARO Institute of Vegetable and Tea Science (NIVTS), National Agriculture and Food Research Organization, 360 Kusawa, Ano, Tsu, Mie 514-2392, Japan
e-mail: h.fukuoka@affrc.go.jp

K. Shirasawa · S. Isobe · E. Asamizu
Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

Present Address:
E. Asamizu
School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8572, Japan

Introduction

The Solanaceae family is a major plant group that includes several important species of practical value, including tomato (*Solanum lycopersicum* L.), potato (*Solanum tuberosum* L.), bell pepper (*Capsicum annuum* L.), and eggplant (*Solanum melongena* L.). Except for bell pepper, these species belong to the largest genus in the Solanaceae, *Solanum*, which includes more than 1,000 species (D'Arcy 1991). Tomato and potato have been used as model plants for the Solanaceae and for genus *Solanum*, and extensive genetic and genomic information has been accumulated for these species. High-density molecular marker linkage maps (Bonierbale et al. 1988; Shirasawa et al. 2010; Tanksley

et al. 1992) and comprehensive data sets of expressed sequence tags (ESTs) (Aoki et al. 2010; Flinn et al. 2005) have been developed for both species. Recently, a draft genome sequence has been published for potato (Potato Genome Sequencing Consortium 2011) and a draft sequence will soon be released for tomato.

In contrast, eggplant has been used less often in molecular genetics research, probably because it is produced and consumed less widely than tomato and potato. Eggplant and closely related *Solanum* species belonging to the subgenus *Leptostemonum* are, however, some of the most important vegetables in many countries in Asia, the Middle and Near East, Southern Europe, and Africa. From botanical and agronomical points of view, eggplant has many unique traits compared with the two *Solanum* model species, including larger fruit size, high temperature- and water-stress tolerance, parthenocarpy without negative pleiotropic effects, and stable *Verticillium* and bacterial wilt resistance. Furthermore, eggplant has a unique phylogenic aspect: it is endemic to the old world, whereas most solanaceous crops are believed to have originated in the Middle and South America (Daunay and Lester 1988). Therefore, the accumulation of genomic information about eggplant will not only facilitate genetics research and molecular breeding of eggplant itself, but will also make this species a valuable and unique member of the Solanaceae for comparative biological studies of the genetics, physiology, development, and evolution of this taxon.

Large-scale DNA marker development and the construction of a high-resolution linkage map would provide fundamental tools for the accumulation of genomic information. The first DNA marker linkage map in eggplant was reported by Nunome et al. (2001); the map was constructed using random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) markers developed for an intraspecific F_2 mapping population. The map was unsaturated and the estimated number of linkage groups in the map did not converge with the haploid chromosome number of eggplant ($n = 12$). Consecutively, more than 1,000 simple sequence repeat (SSR) markers were developed and an SSR-based linkage map was constructed using the same intraspecific F_2 population (Nunome et al. 2009). Since SSR markers are based on polymerase chain reaction (PCR), have co-dominant inheritance, and are highly polymorphic, this research remarkably advanced the marker resources and linkage map information available for eggplant.

However, this map was also unsaturated and consisted of 12 provisional linkage groups that contained gaps of unknown genetic distance. Some linkage groups were much shorter than expected, suggesting that some parts of the eggplant genome were significantly under-represented. It has been reported that genomic SSRs exist preferentially in heterochromatic regions, and therefore it would be difficult to cover the whole genome solely using randomly isolated genomic SSR markers (Ohyama et al. 2009; Shirasawa et al. 2010). In addition, limited sequence homology information was available between SSR-flanking eggplant genomic sequences and tomato genome marker sequences (Nunome et al. 2009), leaving the correspondence between the eggplant linkage groups and the tomato chromosomes unclear. Recently, Barchi et al. (2010) also reported an AFLP-based linkage map for eggplant created using an intraspecific F_2 mapping population. Although their map consisted of 12 linkage groups and 238 markers, including 20 SSR markers common to the map reported by Nunome et al. (2009), the number of the common markers was too low to permit integration of the information in the two maps. On the other hand, Doganlar et al. (2002) adopted tomato restriction fragment length polymorphism (RFLP) markers to construct an eggplant map using an interspecific cross between eggplant and its wild ally *Solanum linnaeanum*. They successfully constructed a linkage map consisting of 12 converged linkage groups that spanned a total length of 1,480 cM. Their comparative analysis using tomato markers permitted inferences about syntenic relationships between the eggplant and tomato genomes. Recently, more than 100 additional markers developed from the conserved ortholog set (COSII), deduced by interspecific comparison of ESTs among higher plant species (Wu et al. 2006), were mapped to the reference RFLP map, and syntenic relationships between the two *Solanum* species were elucidated in more detail (Wu et al. 2009b). However, the markers used in the experiments of Doganlar et al. (2002) and Wu et al. (2009b) relied on interspecific DNA polymorphisms and, therefore, most of the markers were not directly applicable to the mapping population derived from intraspecific crosses.

In the present study, we performed large-scale screening of intraspecific DNA polymorphisms to focus on functional gene-related genomic sequences in eggplant. We used the results of this screening to construct eggplant linkage maps that would represent possible genomic regions that had been missed in our previous SSR-based map (Nunome et al. 2009). To select the eggplant genes that should be mapped, we performed comparative sequence analysis of three unigene data sets (eggplant, tomato, and potato) to build orthologous gene sets, and mapped more than 300 markers derived from the gene sets that were common to the eggplant and tomato maps. The results let us establish connections between the genomic and genetic information for the two species.

Materials and methods

Plant material

Two intraspecific F_2 mapping populations, LWF2 and ALF2, were used for construction of the eggplant map. The LWF2 population ($n = 90$) and the ALF2 population ($n = 93$) were derived from crosses between *S. melongena* LS1934 and *S. melongena* WCGR112-8 and between *S. melongena* AE-P03 and LS1934, respectively. An interspecific tomato F_2 population, EXPEN (with *S. lycopersicum* LA925 and *S. pennellii* LA716 as the parental lines), was used for mapping of tomato markers. Total genomic DNA samples were prepared from young leaves of each plant using the DNeasy Plant Mini Kit (Qiagen GmbH, Hilden, Germany).

Construction of *Solanum* orthologous (SOL) gene sets

As eggplant, tomato, and potato unigene data sets, we used 16k eggplant (16,245 sequences; Fukuoka et al. 2010), 47k tomato (DFCI Tomato Gene Index v. 12 (LeGI_v12); 46,849 sequences; Quackenbush et al. 2001), and 57k potato (Potato Gene Index v. 11 (StGI_v11); 56,712 sequences; Quackenbush et al. 2001). We constructed putative ortholog sets of ESTs from the three *Solanum* species as illustrated in Fig. 1. First, unigene data sets of the three species were reciprocally compared with each other using the Smith–Waterman algorithm, as implemented by the SSEARCH program (Pearson and Lipman 1988). If "reciprocal best hit" relationships existed (i.e., the first sequence finds the second sequence as its best hit in the second species and vice versa; Li et al. 2003) among all combinations of unigenes of the three species, the three unigenes (one unigene from each species) were presumed to comprise an ortholog group. Second, we applied an additional criterion to the tentative ortholog groups: we required that the ratio of the aligned homologous sequence length to the overlapping sequence length be 0.8 or higher in all comparisons. The putative ortholog groups that met both criteria were used for subsequent single-nucleotide polymorphism (SNP) and insertion/deletion (InDel) discovery. Gene ontology (GO)-based functional annotation of eggplant unigenes was done using the plant GO-Slim categories according to Fukuoka et al. (2010).

SNP and InDel discovery, marker development, and genotyping

Each eggplant and tomato unigene pair from each *Solanum* ortholog group was subjected to a BLASTX search against the *Arabidopsis thaliana* predicted proteome (TAIR8). The gene and cDNA sequences corresponding to the best-hit *Arabidopsis* protein and the eggplant and tomato query

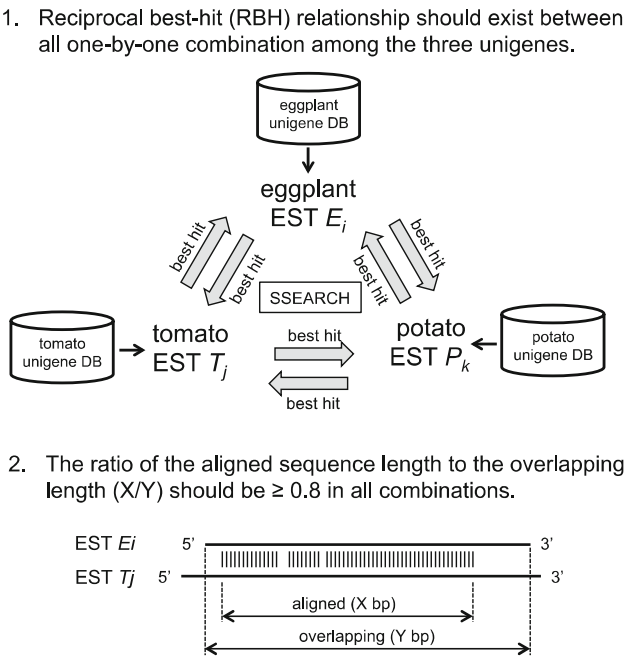

Fig. 1 The method used to construct the *Solanum* orthologous (SOL) gene sets

sequences were subjected to multiple alignment using the T-COFFEE program (Notredame et al. 2000) to predict intron positions and the 5′- and 3′-UTRs (untranslated regions) of each unigene. PCR primers were designed to amplify intron-containing genomic sequences or non-coding UTRs using the Primer3 software (Rozen and Skaletsky 2000). SNPs and InDels were screened by direct sequencing of amplified genomic DNA fragments using BigDye v3 sequencing premix and a 3730xl DNA sequencer (Life Technologies Corporation, Carlsbad, CA, USA). Sequence data were processed using the phred/phrap/cross_match package (Ewing et al. 1998). SNPs found between parental lines of mapping populations were mainly genotyped using the melting temperature (T_m)-shift PCR method (Wang et al. 2005), with some modifications (Fukuoka et al. 2008). When reliable allele specificity could not be achieved using T_m-shift PCR, such SNPs were genotyped by direct sequencing using the same primer sets used for SNP discovery. Genotyping of InDel markers was done using the GeneScan method with the 3730xl DNA sequencer, the GeneMapper software (Life Technologies Corporation), and post-PCR fluorescent labeling (Kukita and Hayashi 2002). Genotyping of SSR markers was performed using the procedure of Nunome et al. (2009).

Linkage analysis and map construction

Genotyping data obtained from the two eggplant mapping populations (LWF2 and ALF2) were used separately for

linkage grouping and marker ordering using the MAPMAKER/EXP 3.0 software (Lander et al. 1987), with LOD ≥6.0 (grouping) or LOD ≥3.0 (ordering) and a maximum distance ≤37.2 cM. The two mapping data sets were then combined using JoinMap 4.0 (Kyazma B.V., Wageningen, The Netherlands) to construct an integrated map (LWA2010) from the two linkage maps (LW2010 and AL2010) under the condition that the order of the common markers in each linkage group was fixed, and was determined in advance by MAPMAKER/EXP. For tomato, map construction was done using JoinMap 4.0, as described by Shirasawa et al. (2010).

Results

Construction of SOL gene sets and screening for intraspecific DNA polymorphisms

We performed reciprocal Smith–Waterman comparisons among the 16k eggplant, 47k tomato, and 57k potato unigene data sets using the SSEARCH program. In total, 4,754 putative *Solanum* orthologous gene groups were identified (Supplementary Table S1). Table 1 summarizes the classification of the unigenes used to build orthologous gene sets based on reciprocal best-hit (RBH) relationships defined by the comparative sequence analysis of the unigenes among the three *Solanum* species. Between the eggplant and tomato unigenes, RBH relationships were confirmed for 11,048 unigene pairs. Of these pairs, 8,540 met the second criterion (more than 80% of the overlapping sequence was aligned) for homology, which had the goal of excluding pairs in which only a limited part of the sequence contributed to the RBH relationship (e.g., short coding sequences for conserved functional motifs). In the same way, 8,258 and 16,830 putative orthologous unigene pairs were identified between eggplant and potato and between tomato and potato, respectively. Furthermore, 6,203 triple-gene sets were found that consisted of an eggplant unigene, a tomato unigene, and a potato unigene that shared RBH relationships with each other. Of these triplets, 4,754 gene sets met the 80% alignment criterion. We designated these 4,754 gene sets as the SOL gene set, and used them as the basis for DNA marker development. Among the 4,754 tomato unigenes included in the SOL gene set, 874 genes were identical to the genes in the COSII (Wu et al. 2006) gene set. Therefore, the other 3,880 SOL gene triplets were newly identified putative orthologs. The distribution of functional properties of the SOL genes based on GO-Slim classification showed no significant difference compared with that in the total eggplant 16k unigene set (Supplementary Fig. S1).

Table 1 Number of putative ortholog sets

Criteria	No. of gene set	Common to COSII
Eggplant unigene	16,245	–
Tomato unigene	46,849	2,527
Potato unigene	56,712	–
RBH between eggplant and tomato	11,048	1,340
Aligned ratio ≥0.8	8,540	1,099
RBH between eggplant and potato	10,860	–
Aligned ratio ≥0.8	8,258	–
RBH between tomato and potato	18,062	1,890
Aligned ratio ≥0.8	16,830	1,849
RBH among eggplant–tomato–potato	6,203	1,061
Aligned ratio ≥0.8	4,754	874

RBH Smith–Waterman reciprocal best-hit relationships. The alignment ratio equals the aligned sequence length (bp) divided by the overlapping sequence length (bp), and we used these ratios to exclude gene sets identified by the RBH analysis that were not strongly aligned

For non-coding-region-directed screening of SNPs, we predicted the positions of exon–intron junctions, 5′-UTRs, and 3′-UTRs by multiple alignments of the unigene sequences of the SOL members and known *Arabidopsis* genome sequences. We successfully designed PCR primer sets that would amplify the intron- and UTR-containing genomic regions for 2,489 eggplant SOL genes out of the 4,754 total SOL genes; the primer sequence information is available from the VegMarks database (http://vegmarks.nivot.affrc.go.jp), and is summarized in Supplementary Tables S2 and S4. SNPs and InDels were screened in the amplicons among the parental eggplant lines (AE-P03, LS-1934, and WCGR112-8). In total, 564 of the 2,489 SOL genes (23%) exhibited at least one SNP among the three lines. Table 2 summarizes the results of the SNP screening. The total length of the genomic DNA sequence that we examined was 2.87 Mbp; on average, we found

Table 2 Single-nucleotide polymorphisms (SNPs) found in the genic regions of the eggplant genome

	SNP position		Total
	Coding	Intron and UTR	
SNPs	99	1,224	1,323
Nucleotide examined (bp)	429,250	2,443,351	2,872,601
SNP/kb	0.231	0.501	0.461
Transition	58	715	773
Transversion	41	509	550
Codon first nucleotide	25	–	–
Codon second nucleotide	26	–	–
Codon third nucleotide	48	–	–
Synonymous	45	–	–
Non-synonymous	54	–	–

0.46 SNP/kbp. The SNP rate in introns and UTRs was roughly 2× that in the coding sequence (0.50 and 0.23 SNP/kbp, respectively). Transitions were more common than transversions (773/550 = 1.4×). For the SNPs found in exons, the mutation rate was highest in the third codon position, and synonymous variations ($n = 45$) were slightly less common than non-synonymous variations ($n = 54$). The InDel frequency was much lower than the SNP frequency; such mutations were only found in 28 of the SOL genes.

SNP and InDel marker development and linkage map construction

We performed linkage mapping for SNPs and InDels that we found between parental lines from each mapping population using the F_2 segregation data. In addition to the SNP and InDel markers based on the SOL gene set, we used 154 EST-based SNP and InDel markers for construction of the linkage map. These additional markers were initially developed as SOL markers in 2007, but were subsequently defined as non-SOL markers based on the results of re-calculation of RBH relationships using new unigene data obtained in 2009. These EST-based non-SOL markers were designated using the prefixes 'est_' and 'est_sl_' for eggplant and tomato markers, respectively. They were used separately for construction of the eggplant and tomato linkage maps, but not for comparative analysis of the two species. We also used 329 previously reported SSR markers in the experiments. Detailed information on the markers that we developed or used in this study is available in the VegMarks database and is summarized in Supplementary Table S2.

To overcome the problem of a low intraspecific polymorphism rate among the lines, we used two F_2 populations, LWF2 and ALF2, for construction of the linkage maps LW2010 and AL2010, respectively. Marker segregation data were obtained for 499 SNP and InDel markers and 212 SSR markers in LWF2, and for 310 SNPs and InDels and 263 SSRs in ALF2, and 328 of these markers were common to the two populations. The two segregation data sets were used separately for linkage grouping and marker ordering with the MAPMAKER/EXP 3.0 software. Segregation distortion was observed for 4.6 and 6.5% of the mapped marker loci for the population ALF2 and LWF2, respectively (Supplementary Table S2). We then combined the two maps to generate an integrated map using JoinMap 4.0 (Fig. 2, Supplementary Table S2). The linkage map LWA2010 spanned a total genetic distance of 1,285.5 cM and consisted of 12 linkage groups (Table 3), which corresponds to the haploid chromosome number of eggplant. In total, we mapped 952 loci in the integrated map, with an average interval between markers of 1.4 cM (Table 3).

Based on the common SSR markers, we compared LWA2010 with our previous linkage map (EW2009; Nunome et al. 2009) using genomic SSR and RAPD markers. Although the linkage grouping and marker order of the two maps were generally comparable, EW2009 was missing several large genomic regions that were included in LWA2010 (Fig. 2). In the most striking examples (E05 and E12), more than 70% of the genomic region had not been explored in EW2009 and was first recognized by mapping of EST-derived SOL markers in the present study. In total, LWA2010 covered approximately 1.5 times the genomic region covered by EW2009.

Mapping of SOL markers to the tomato linkage map and macro-syntenic relationships between eggplant and tomato

SNPs and InDels in the tomato SOL genes were screened between LA925 and LA716, the parental lines of the tomato mapping population EXPEN; in total, 288 tomato SOL genes were successfully mapped in the EXPEN tomato linkage map (EXPEN-NIVTS-2010) by means of T_m-shift PCR genotyping or direct sequencing. Detailed information on the markers is available from the VegMarks database, and is summarized in Supplementary Table S3. Of these 288 newly mapped tomato SOL genes, 255 were common to the eggplant LWA2010 map. In addition, 67 COSII markers and four other EST-based markers that were already mapped in EXPEN (Shirasawa et al. 2010) were found to be derived from the SOL gene sets defined in the present study, corresponding to the eggplant SOL markers mapped in LWA2010. In total, 326 markers were mapped in both the eggplant LWA2010 and tomato EXPEN-NIVTS-2010 maps.

Predicted macro-syntenic relationships between the genomes of the two species are shown in Fig. 3, and detailed information on the correspondence between markers is provided in Supplementary Table S4. The eggplant linkage groups E01, E02, E03, E06, E07, E08, and E09 showed good correspondence with T01, T02, T03, T06, T07, T08, and T09, respectively, but inversions of partial chromosomal segments in these groups were predicted in E02–T02 and E09–T09. More complex relationships appear to exist among the other chromosomes. T04 and T10 correspond to the lower half of E04 and the lower half of E11 (inverted) and to the upper half of E04 (inverted) and lower half of E10 (partly inverted), respectively. In addition, a correspondence between short segments of E12 and T10 was suggested by three closely linked SOL markers (SOL1056, SOL1218, and SOL2050; shown in Supplementary Table S4). In T11, the upper half corresponds to the upper half of E11 (inverted) and the lower half corresponds to the lower half of E12 (with

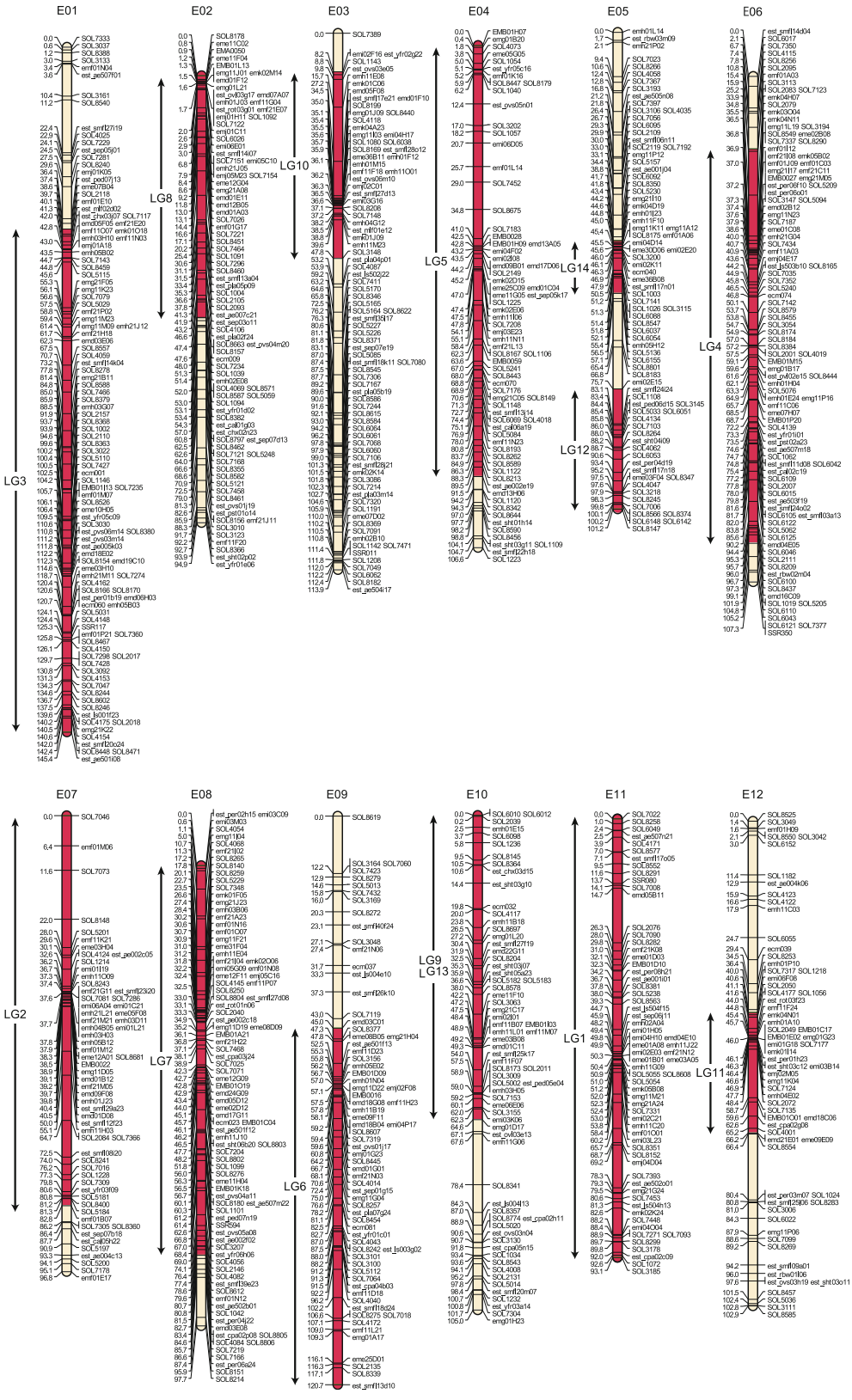

Fig. 2 The integrated eggplant DNA marker linkage map (LWA2010) developed in the present study. Eggplant linkage groups are designated as E01–E12 corresponding to Doganlar et al. (2002). The map consists of 952 DNA markers and spans a total genetic distance of 1,285.5 cM in 12 linkage groups. *Red regions* in the linkage groups and lines with *double arrowheads* indicate the genomic regions corresponding to our previous SSR-based linkage map (EW2009; Nunome et al. 2009)

Table 3 Markers in the integrated eggplant linkage map LWA2010

Linkage group	Length (cM)	Number of markers				Marker interval (cM)	
		SOL[a]	Other gene-based	Genomic SSR	Total	Average	Max
E01	145.4	60 (42)	18	33	111	1.3	11.2
E02	94.9	41 (27)	18	27	86	1.1	6.5
E03	113.9	47 (31)	17	22	86	1.3	11.5
E04	106.6	35 (24)	10	26	71	1.5	6.3
E05	101.2	50 (38)	11	18	79	1.3	8.9
E06	107.3	54 (41)	16	31	101	1.1	9.3
E07	96.8	24 (17)	9	28	61	1.6	10.4
E08	97.7	34 (18)	20	36	90	1.1	8.5
E09	120.7	31 (21)	14	25	70	1.7	12.2
E10	105.0	32 (26)	15	20	67	1.6	10.8
E11	93.1	29 (18)	10	26	65	1.4	11.6
E12	102.9	32 (23)	12	21	65	1.6	14.0
Total	1285.5	469 (326)	170	313	952	1.4	14.0

[a] Numbers in parentheses represent the number of markers also mapped in the tomato EXPEN-NIVTS-2010 map

several rearrangements). The upper half of T05 mainly corresponds to the upper half of E10, but short (in genetic length) segments of E03 are also present, and the lower half of T05 corresponds to the upper half of E05 (inverted). Lastly, the upper half of T12 corresponds to the upper half of E12 and, in part, the upper half of E10, whereas the lower half corresponds to the lower half of E05. In T04, T05, T11, and T12, small regions that were not assigned to any eggplant genome remained (shown in gray) because the junctions of the neighboring regions corresponding to the different eggplant linkage groups could not be objectively determined.

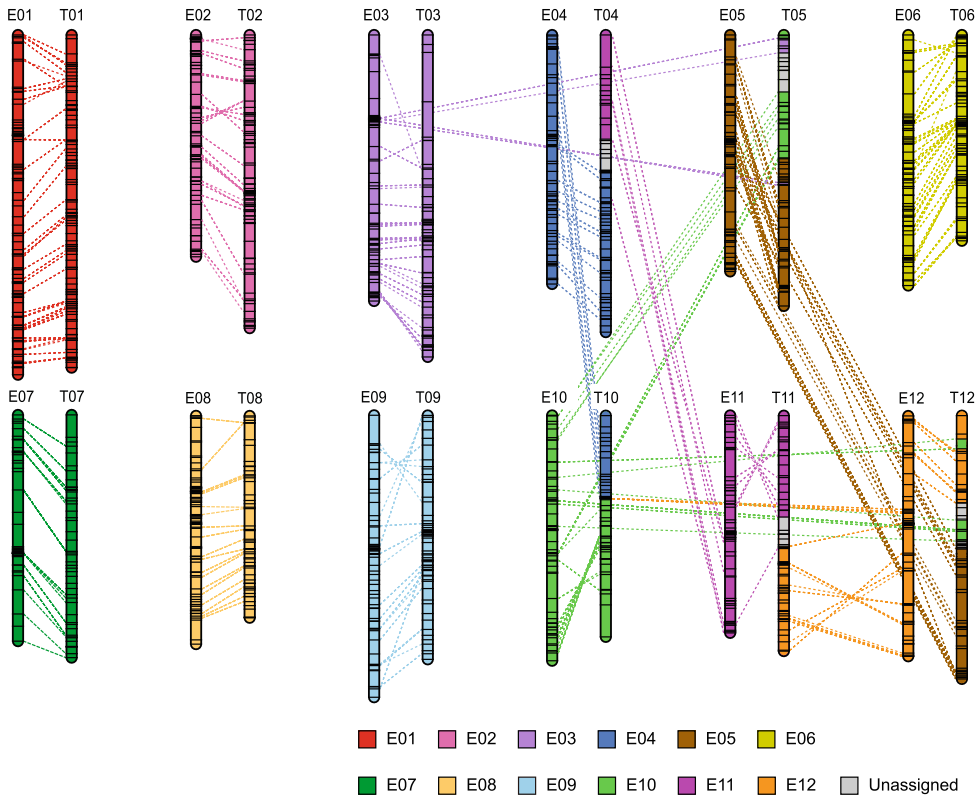

Fig. 3 Macro-syntenic relationships between the eggplant (E01–E12) and tomato (T01–T12) genomes deduced using the SOL and COSII markers developed in the present study. Each eggplant linkage group and its corresponding tomato chromosomal regions are assigned a different color. Each marker pair that is common between the two maps is connected by a dotted line whose color is that of the linkage group in which the eggplant marker is located

Discussion

To date, most comprehensive molecular marker-based linkage maps of eggplant have been constructed using mapping populations derived from interspecific crosses (e.g., *Solanum melongena* × *S. linnaeanum*) because of the low frequency of intraspecific DNA polymorphisms (Doganlar et al. 2002; Wu et al. 2009b). Therefore, most markers in those maps were not directly applicable for intraspecific genetic analysis to support practical eggplant breeding. The marker resource developed in this study was highly polymorphic and transferable among eggplant cultivars, and it would therefore be quite useful for further applied studies to support eggplant molecular breeding. In addition, segregation data sets obtained from F_2 populations derived from intraspecific crosses would have advantage in ordering and mapping of markers in terms of possible segregating distortion. Shirasawa et al. (2010) reported that segregation distortion was observed for 38.2% of the mapped marker loci in tomato interspecific mapping population EXPEN. In this study, only 4.6 and 6.5% of the mapped marker loci exhibited significant ($P < 0.05$) segregation distortion in the two interspecific populations, ALF2 and LWF2, respectively, which would contextually support the reliability of the linkage maps reported in this study.

It is known that genomic SSRs tend to be unusually abundant in the pericentromeric heterochromatin regions; as a result, large chromosomal segments would be under-represented by SSR markers. In the present study, a set of 623 SNP and InDel markers (469 SOLs and 154 other EST-based markers) was newly mapped to construct the integrated eggplant linkage map LWA2010 using two F_2 populations derived from intraspecific crosses. Comparative analysis of LWA2010 with EW2009 revealed that LWA2010 covered 1.5 times the genomic region covered by EW2009; thus, approximately one-third of the eggplant genome was under-represented in the latter map, which was constructed using genomic SSR markers that had been isolated randomly from SSR-enriched genomic libraries. This suggests that the genomic SSRs were not evenly distributed throughout the whole genome and, therefore, marker development based on nucleotide sequences obtained from various experimental sources (e.g., enriched genomic libraries and cDNA libraries) would be important for map construction covering the whole genome of an inadequately explored species. As shown in Table 2, the SNP frequency in non-coding sequences was 0.501 SNP/kbp, which was comparable to a previous result (0.8 SNP/kbp) detected by means of large-scale sequencing of reaction site-associated DNA (RAD) tags (Barchi et al. 2011). In coding regions, the SNP frequency was approximately half of that in non-coding regions (0.23 SNP/kbp); this is similar to the results obtained in rice (Yamamoto et al. 2010), for which about half of the SNPs (54/99) were non-synonymous variations. These results are not robust because they might be greatly influenced by the genetic relationships among the cultivars used in each experiment. However, the data suggest that it would be practically feasible to screen for SNPs in protein-coding genes and their adjacent non-coding introns and UTRs for DNA marker development in eggplant, and probably in similar autogamous crop species with low levels of intraspecific DNA polymorphism.

The integrated map LWA2010 converged into 12 linkage groups with a total map length of 1,285.5 cM. In LWA2010, 639 gene-based markers (469 SOL and 170 other) and 313 genomic SSR markers were mapped with an average marker interval of 1.4 cM and the largest gap equaling 14.0 cM. The total map length of LWA2010 was comparable to that of the interspecific linkage maps reported by Doganlar et al. (2002) and Wu et al. (2009b). Our comparison of LWA2010 with the tomato reference linkage map EXPEN2000 showed that the whole eggplant genome generally corresponded to the reference tomato genome, though with several rearrangements, inversions, and gaps. This suggests that LWA2010 covered most of the eggplant genome. Therefore, LWA2010 is currently, to our knowledge, one of the most informative, comprehensive, and versatile eggplant linkage maps based on intraspecific crosses and consisting of PCR-based and sequence-tagged markers. Recently, Barchi et al. (2011) reported a large-scale effort to discover SNPs between the parental lines of their eggplant intraspecific mapping population, with more than 2000 SNPs detected that are likely to be mappable by means of bead-array technology. When those SNPs have been mapped in their intraspecific linkage maps, it should be feasible to integrate their map with LWA2010 to improve our knowledge of the eggplant genome.

Most of the SNP and InDel markers developed in this study were based on the genic sequences of hypothetical orthologs within the genus *Solanum* (SOL genes). Wu et al. (2006) proposed an approach similar to ours, and their concept of "conserved ortholog sets" (COS) has contributed greatly to the comparative genetics of several solanaceous species (Wu et al. 2009a, 2010; Wu and Tanksley 2010) and species in other plant taxonomic groups (Cabrera et al. 2009; Fregene and Castelblanco 2006; Krutovsky et al. 2006) In our study, 4,754 putative SOL gene sets were identified by comparative sequence analysis among unigene sets of eggplant, tomato, and potato. Of the 2,527 tomato unigenes in the COSII gene sets built by Wu et al. (2006), only 874 genes (35%) were common to the tomato unigenes in the SOL gene sets. When the COSII gene set was constructed, a limited number of eggplant ESTs were available and, therefore, an eggplant data set was not involved in

computational screening of the orthologous gene set in their study. On the other hand, species such as *Arabidopsis thaliana* (Brassicaceae), coffee (*Coffea arabica*, Rubiaceae), and sunflower (*Helianthus annuus*, Asteraceae) that are more distantly related to the Solanaceae were involved. The difference in the taxonomic scope of the calculations could explain the discrepancy between our results and those of Wu et al. (2006). As Wu et al. (2006) noted, incomplete genomic information would lead to fallacious inferences about orthologous and paralogous relationships. The eggplant data set used in the present study consisted of 16,245 unigenes, which is likely to be much less than the number of transcripts encoded by the eggplant genome. The incompleteness of the eggplant unigene set used in our study might also explain the discrepancy. Recently, 86% of the potato genome sequence has been sequenced, and 39,031 protein-coding genes have been predicted (Potato Genome Sequencing Consortium 2011). Even though 14% of the potato genome must still be explored, the size of the potato unigene set (56,712 unigenes, StGI_v11) used in our study might over-represent the whole transcriptome and, therefore, potential data redundancy might also confuse inferences about orthologous and paralogous relationships. In addition, we based our nucleotide sequence comparisons on the Smith–Waterman algorithm implemented by the SSEARCH program, whereas the COSII gene set was constructed based on the results of BLASTN/BLASTX/TBLASTN comparisons, and the difference in calculation methods would also contribute to the discrepancy.

Even though the DNA marker set developed based on putative orthologous genes might contain some fallacious inferences about orthologous and paralogous relationships, it remains a powerful tool for the recognition of genome-wide syntenic relationships among the genomes of related solanaceous species. Our GO-based comparison (Supplemental Fig. S1) showed that no obvious difference existed between the distribution of functional annotations of the SOL genes and that of the whole eggplant unigene set, suggesting that selection of a gene set based on RBH relationships would not severely bias the selection of nucleotide sequences for genome-wide marker development. Comparative analysis of the organization of the tomato and eggplant genomes was first reported by Doganlar et al. (2002) using RFLP-based tomato COS markers, to which more detailed information based on PCR-based and sequence-tagged markers (COSII markers) was appended by Wu et al. (2009b). In the present study, the corresponding chromosomal regions in the eggplant and tomato genomes were identified by mapping of 326 common markers, including 255 SOL markers, 67 COSII markers, and 4 other EST-derived tomato markers in the SOL gene sets. The overall systemic relationships between eggplant and tomato deduced in this study agreed well with those described by Wu et al. (2009b), except that a small segment of eggplant linkage group 12 shared three orthologous markers (SOL1056, SOL2050, and SOL1218) mapped in a possible pericentromeric heterochromatin region of tomato chromosome 10. On the other hand, 14 other SOL markers (shown in Supplementary Table S4) corresponded to unexpected (less plausible) tomato genomic regions; because these correspondences were not supported by two or more neighboring markers, we considered them to be unreliable and ignored them in this study. Genomic information with much higher resolution would be required to elucidate the possible mid- to microscale synteny between the eggplant and tomato genomes.

The integrated intraspecific eggplant linkage map developed in this study contains a total of 952 markers. All markers were PCR based and sequence tagged and, therefore, offer a highly versatile tool for genetic analysis of various parental combinations in eggplant. Once traits of interest have been successfully mapped using the SOL markers, the corresponding tomato genome information (contigs, scaffolds, and pseudomolecule sequences) will be a good source for additional marker development to support the next investigative steps, such as fine mapping and map-based cloning. In addition, genetic information on the basis of useful eggplant-specific characteristics such as resistance to soil-borne diseases (Liu et al. 2009), waterlogging tolerance (Lin et al. 2004), and parthenocarpy without pleiotropic defects (Saito et al. 2009) would become transferrable to molecular breeding of tomato and potato through comparative genetic analyses. The SOL markers reported here provide powerful genetic tools to connect eggplant, an old world *Solanum* species, with tomato and potato, its new world allies.

Acknowledgments We thank Prof. S. Tanksley (Cornell University) for kindly providing plants of the tomato mapping population and genotyping data for their markers, and Dr. T. Saito (NARO Institute of Vegetable and Tea Science, NIVTS) for providing the eggplant materials. We are also grateful to Ms. Y. Kitamura (NIVTS) for her skillful technical assistance. This work was supported by grants-in-aid from the Ministry of Agriculture, Forestry, and Fisheries of Japan ('Genomics for Agricultural Innovation', DD-4010/SGE-1001; 'Development of mitigation and adaptation techniques to global warming in the sectors of agriculture, forestry, and fisheries', C-3-1010).

References

Aoki K, Yano K, Suzuki A, Kawamura S, Sakurai N, Suda K, Kurabayashi A, Suzuki T, Tsugane T, Watanabe M, Ooga K, Torii M, Narita T, Shin-i T, Kohara Y, Yamamoto N, Takahashi H, Watanabe Y, Egusa M, Kodama M, Ichinose Y, Kikuchi M, Fukushima S, Okabe A, Arie T, Sato Y, Yazawa K, Satoh S, Omura T, Ezura H, Shibata D (2010) Large-scale analysis of

full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. BMC Genomics 11:210

Barchi L, Lanteri S, Portis E, Stagel A, Vale G, Toppino L, Rotino GL (2010) Segregation distortion and linkage analysis in eggplant (*Solanum melongena* L.). Genome 53:805–815

Barchi L, Lanteri S, Portis E, Acquadro A, Vale G, Toppino L, Rotino GL (2011) Identification of SNP and SSR markers in eggplant using RAD tag sequencing. BMC Genomics 12:304

Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLp maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095–1103

Cabrera A, Kozik A, Howad W, Arus P, Iezzoni AF, van der Knaap E (2009) Development and bin mapping of a Rosaceae Conserved Ortholog Set (COS) of markers. BMC Genomics 10:562

D'Arcy WG (1991) The Solanaceae since 1976 with a review of its biogeography. In: Hawkes JG, Lester RN, Nee M, Estrada N (eds) Solanaceae III. Royal Botanic Gardens/Kew, London, pp 75–138

Daunay MC, Lester RN (1988) The usefulness of taxonomy for *Solanaceae* breeders, with special reference to the genus *Solanum* and to *Solanum melongena* L. (eggplant). Capsicum Newslett 7:70–79

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002) A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. Genetics 161:1697–1711

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185

Flinn B, Rothwell C, Griffiths R, Lague M, DeKoeyer D, Sardana R, Audy P, Goyer C, Li XQ, Wang-Pruski G, Regan S (2005) Potato expressed sequence tag generation and analysis using standard and unique cDNA libraries. Plant Mol Biol 59:407–433

Fregene M, Castelblanco W (2006) SSCP-SNP-based conserved ortholog set (COS) markers for comparative genomics in cassava (*Manihot esculenta* Crantz). Plant Mol Biol Rep 24:229–236

Fukuoka H, Miyatake K, Negoro S, Nunome T, Ohyama A, Yamaguchi H (2008) Development of a routine procedure for single nucleotide polymorphism marker design based on the T_m-shift genotyping method. Breed Sci 58:461–464

Fukuoka H, Yamaguchi H, Nunome T, Negoro S, Miyatake K, Ohyama A (2010) Accumulation, functional annotation, and comparative analysis of expressed sequence tags in eggplant (*Solanum melongena* L.), the third pole of the genus *Solanum* species after tomato and potato. Gene 450:76–84

Krutovsky K, Elsik C, Matvienko M, Kozik A, Neale D (2006) Conserved ortholog sets in forest trees. Tree Genet Genomes 3:61–70

Kukita Y, Hayashi K (2002) Multicolor post-PCR labeling of DNA fragments with fluorescent ddNTPs. Biotechniques 33:502–506

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

Li L, Stoeckert CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189

Lin KHR, Weng CC, Lo HF, Chen JT (2004) Study of the root antioxidative system of tomatoes and eggplants under waterlogged conditions. Plant Sci 167:355–365

Liu N, Zhou BL, Zhao X, Lu B, Li YX, Hao J (2009) Grafting eggplant onto tomato rootstock to suppress *Verticillium dahliae* infection: the effect of root exudates. HortScience 44:2058–2062

Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217

Nunome T, Ishiguro K, Yoshida T, Hirai M (2001) Mapping of fruit shape and color development traits in eggplant (*Solanum melongena* L.) based on RAPD and AFLP markers. Breed Sci 51:19–26

Nunome T, Negoro S, Kono I, Kanamori H, Miyatake K, Yamaguchi H, Ohyama A, Fukuoka H (2009) Development of SSR markers derived from SSR-enriched genomic library of eggplant (*Solanum melongena* L.). Theor Appl Genet 119:1143–1153

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

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448

Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J (2001) The *TIGR* gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Res 29:159–164

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

Saito T, Yoshida T, Monma S, Matsunaga H, Sato T, Saito A, Yamada T (2009) Development of the parthenocarpic eggplant cultivar 'Anominori'. Jpn Agric Res Q 43(2):123–127

Shirasawa K, Asamizu E, Fukuoka H, Ohyama A, Sato S, Nakamura Y, Tabata S, Sasamoto S, Wada T, Kishida Y, Tsuruoka H, Fujishiro T, Yamada M, Isobe S (2010) An interspecific linkage map of SSR and intronic polymorphism markers in tomato. Theor Appl Genet 121:731–739

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB et al (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

Wang J, Chuang K, Ahluwalia M, Patel S, Umblas N, Mirel D, Higuchi R, Germer S (2005) High-throughput SNP genotyping by single-tube PCR with Tm-shift primers. Biotechniques 39:885–893

Wu F, Tanksley SD (2010) Chromosomal evolution in the plant family Solanaceae. BMC Genomics 11:182

Wu F, Mueller LA, Crouzillat D, Petiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics 174:1407–1420

Wu F, Eannetta NT, Xu Y, Durrett R, Mazourek M, Jahn MM, Tanksley SD (2009a) A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. Theor Appl Genet 118:1279–1293

Wu F, Eannetta NT, Xu Y, Tanksley SD (2009b) A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. Theor Appl Genet 118:927–935

Wu F, Eannetta NT, Xu Y, Plieske J, Ganal M, Pozzi C, Bakaher N, Tanksley SD (2010) COSII genetic maps of two diploid *Nicotiana* species provide a detailed picture of synteny with tomato and insights into chromosome evolution in tetraploid *N. tabacum*. Theor Appl Genet 120:809–827

Yamamoto T, Nagasaki H, Yonemaru J, Ebana K, Nakajima M, Shibaya T, Yano M (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11:267