Research

Fine mapping in tomato using microsynteny with the *Arabidopsis* genome: the *Diageotropica* (*Dgt*) locus

KwangChul Oh*, Kristine Hardeman*†, Maria G Ivanchenko*, Mary Ellard-Ivey*‡, Andreas Nebenführ*§, TJ White* and Terri L Lomax*

Addresses: *Department of Botany and Plant Pathology and the Center for Gene Research and Biotechnology, Oregon State University, Corvallis, OR 97331-2902, USA. †Monsanto-Mystic Research, 62 Maritime Drive, Mystic, CT 06355, USA. ‡Department of Biology, Pacific Lutheran University, Tacoma, WA 98447-0003, USA. §Department of Botany, University of Tennessee, Knoxville, TN 37996-1100, USA.

Correspondence: Terri L Lomax: lomaxt@bcc.orst.edu

Published: 28 August 2002

Genome ***Biology*** 2002, **3(9)**:research0049.1–0049.11

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2002/3/9/research/0049

(Print ISSN 1465-6906; Online ISSN 1465-6914)

Received: 18 December 2001
Revised: 28 May 2002
Accepted: 21 June 2002

Abstract

Background: The *Arabidopsis thaliana* genome sequence provides a catalog of reference genes applicable to comparative microsynteny analysis of other species, facilitating map-based cloning in economically important crops. We have applied such an analysis to the tomato expressed sequence tag (EST) database to expedite high-resolution mapping of the *Diageotropica* (*Dgt*) gene within the distal end of chromosome 1 in tomato (*Lycopersicon esculentum*).

Results: A BLAST search of the *Arabidopsis* database with nucleotide sequences of markers that flank the tomato *dgt* locus revealed regions of microsynteny between the distal end of chromosome 1 in tomato, two regions of *Arabidopsis* chromosome 4, and one on chromosome 2. Tomato ESTs homeologous to *Arabidopsis* gene sequences within those regions were converted into co-dominant molecular markers via cleaved amplified polymorphic sequence (CAPS) analysis and scored against an informative backcross mapping population. Six new microsyntenic EST (MEST) markers were rapidly identified in the *dgt* region, two of which further defined the placement of the *Dgt* gene and permitted the selection of a candidate tomato bacterial artificial chromosome clone for sequence analysis.

Conclusions: Microsynteny-based comparative mapping combined with CAPS analysis of recombinant plants rapidly and economically narrowed the *dgt* mapping region from 0.8 to 0.15 cM. This approach should contribute to developing high-density maps of molecular markers to target-specific regions for positional cloning and marker-assisted selection in a variety of plants.

Background

High-resolution mapping and chromosome walking, critical steps in the positional cloning of a mutant gene, may become problematic and tedious without high-density molecular markers. Although a number of molecular-marker maps are available for various species, further resolution of the target region is often required, as markers may be irregularly spaced along the chromosome owing to uneven rates of recombination. In addition, focusing the genetic interval reduces the time and resources necessary for chromosome walking. Comparative mapping is based on regions of microsynteny between two organisms and provides a powerful technique for enriching molecular markers in the region surrounding a gene of interest. A number of researchers

have suggested that map-based cloning in economically important crop species can be expedited by utilizing chromosomal microsynteny between the target and a model species [1-3]. The recently completed sequence of the entire genome of *Arabidopsis thaliana* [4] now provides a catalog of ordered reference genes immediately applicable to other higher plant species [5]. Conservation of synteny is well documented in closely related species within the same family: for example, *Arabidopsis* and *Brassica oleracea* [6]; rice and barley [7]; and tomato, potato and capsicum [8,9]. Recent comparative sequence analyses and mapping studies have indicated that microsynteny and macrosynteny are also well conserved between *Arabidopsis* and evolutionarily divergent species such as tomato [10,11] or soybean [12]. Thus, comparative mapping has the potential to rapidly identify additional molecular markers in a region of interest in those species.

The single-gene *diageotropica* (*dgt*) mutant of tomato (*Lycopersicon esculentum*) displays a pleiotropic phenotype that includes reduced auxin sensitivity [13]. A number of physiological and molecular studies suggest that a *Dgt*-dependent auxin signal transduction pathway regulates a subset of early auxin-response genes [14-17]; however, the nature of the *Dgt* gene is still unknown. We have been using a map-based cloning strategy to isolate the *Dgt* gene, previously mapped to the long arm of chromosome 1 [8]. On the basis of recent comparative sequence analyses showing well-conserved microsynteny between the tomato and *Arabidopsis* genomes within relatively small regions [10,11], we applied microsynteny-based comparative mapping to facilitate the positional cloning of the *Dgt* gene and successfully reduced the genetic interval with new molecular markers.

Results

Identification of microsyntenic regions in *Arabidopsis* chromosomes

Using restriction-length-fragment polymorphism (RFLP) and RFLP-derived cleaved amplified polymorphic sequence (CAPS) markers (Figure 1) based on previously published tomato genetic and RFLP maps [8], initial studies mapped the *dgt* locus to a region of around 0.8 cM near the bottom of the long arm of tomato chromosome 1. Of 1,308 backcross (BC_1) individuals screened, 10 plants were determined to be recombinant between markers TG269 and CT190, whereas no plants were found to be recombinant between TG389 and *dgt* (Figure 1). To identify additional genes within that region by finding microsyntenic regions in *Arabidopsis*, nucleotide sequences of the three RFLP markers most closely linked to the *dgt* locus (TG269, TG389, and CT190) were used to identify homeologous sequences in the *Arabidopsis* genome database [18]. BLASTN matches with an arbitrary threshold expect value (*E*-value) of less than 0.01 were investigated as significant matches.

Three putative microsyntenic regions (MSRs) were identified in the *Arabidopsis* genome (Figures 1,2). In MSR1, homeologs of all three tomato RFLP markers were found in the same linear order on two adjacent *Arabidopsis* bacterial artificial chromosome (BAC) clones from chromosome 4 (*Arabidopsis* accession numbers are in parentheses): F20M13 (AL035540) and T9A14 (AL035656). The *Arabidopsis* homeologs of TG269, TG389, and CT190 in MSR1 had corresponding *E*-values of $4e^{-3}$, $2e^{-7}$, and $3e^{-5}$, respectively. The *Arabidopsis* homeologs of markers TG269 (AT4g38850), TG389 (AT4g38730), and CT190 (AT4g38580) encode a small auxin-upregulated protein (SAUR), a hypothetical protein of unknown function (HP) and a farnesylated protein (FP), respectively. The second microsyntenic region (MSR2) spanned two *Arabidopsis* BAC clones, F11l11 (AL079347) and M4E13 (AL022023), and was also located on chromosome 4, but only yielded homeologs for tomato markers TG269 and CT190 (*E*-values of $1e^{-3}$ and $8e^{-3}$, respectively). The third microsyntenic region (MSR3) was located on *Arabidopsis* chromosome 2 and spanned *Arabidopsis* BACs F26H11 (AC006263) and F7O24 (AC007142). This region only contained a homeolog to TG389 (AT2g21120, *E*-value of $9e^{-4}$). However, additional auxin-regulated genes were identified in this region (AT2g21200 to AT2g21220), which made continued analysis potentially beneficial. In addition, this region had previously been reported to be syntenic to MSR1 [19]. Therefore, we also included MSR3 in the comparative analysis of genes in MSRs between tomato and *Arabidopsis*.

Close inspection of the microsyntenic regions detected by BLASTN analysis suggested that the genomic microstructure was highly conserved in all three MSRs. Eighteen of 45 genes (40%) identified in MSR1 also have homologs in MSR2 and/or MSR3 (Figure 2). The order of the microsyntenic genes was highly conserved in the three regions with the colinear pattern of three genes - phospholipase-like protein (pEARLI 4), cyclophilin-type peptidyl-prolyl *cis/trans* isomerase (PPI), and a cluster of SAUR and auxin-induced proteins (COSAP) - serving as a common footprint for these MSRs. At the same time, variation in syntenic genes by gene duplication and/or translocation was also evident in all three MSRs. For example, the sequences of PPI (AT4g38740), phosphatidylinositol-specific phospholipase C precursor (PPLC, AT4g38690), and disease-resistance response protein (DRRP, AT4g38700) of MSR1 are duplicated only in either MSR2 or MSR3. PPLCs (AT4g34920/AT4g34930) and DRRPs (AT2g21110/AT2g21120) remain tandemly duplicated in MSR2 and MSR3, respectively, whereas PPI (AT4g34960) appears to have been translocated after gene duplication in MSR2.

High-resolution mapping of the *dgt* locus by MEST markers

Because *Arabidopsis* MSR1 contained homeologs in common with three RFLP markers in the *dgt* region of

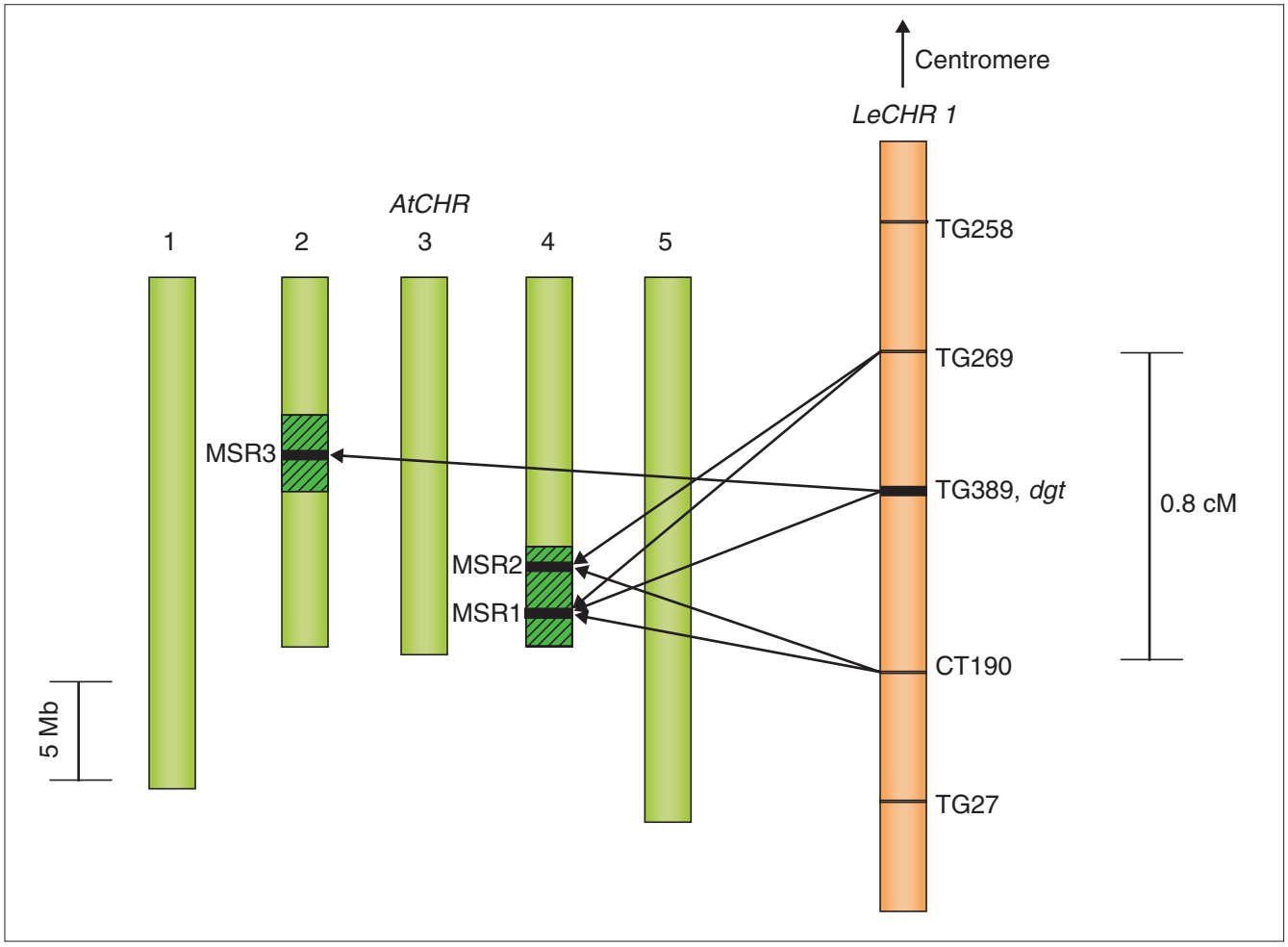

Figure 1
Three microsyntenic regions identified by BLASTN on *Arabidopsis* chromosomes. Arrows indicate *Arabidopsis* BAC clones with BLASTN matches for the tomato RFLP marker. MSR, microsyntenic region. Solid segments indicate the MSRs defined here. The default values for TAIR BLAST and the Blosum 62 scoring matrix were used for the BLASTN parameter options. Hatched segments on *Arabidopsis* chromosomes (*AtCHR*) 2 and 4 represent previously reported duplicated chromosomal segments [19]. *LeCHR1*, *L. esculentum* chromosome 1.

tomato, we used it as our main target region to search for corresponding tomato homeologs. MSR1 contained 25 intervening genes between the homeologs of TG269 and CT190. TBLASTN analysis of each intervening *Arabidopsis* gene against the Tomato Gene Index [20] generated 260 TBLASTN hits that were chosen for further investigation on the basis of the original annotation of the *Arabidopsis* gene and a low expect value (< e^{-20}). From those, 40 robust PCR products were obtained and selected for development into new CAPS markers. The original 40 products provided 30 new CAPS markers (Table 1) that were used for PCR-based genotyping against the informative mapping population (Figure 3). The 10 previously identified informative recombinant plants, as well as an equal number of non-informative plants, were used to analyze the new microsyntenic EST (MEST) markers. Of the 30 CAPS markers tested, six MEST markers mapped in the *dgt* region. Two MESTs, BG643476 (putative protein, PP) and TC85079 (EF-hand containing protein-like protein, EFP), identified a crossover between TG269 and TG389/*dgt*, and one MEST, TC98260 (glycine-rich protein, GRP), revealed a crossover between TG389/*dgt* and CT190. These three new markers, which flank *dgt*, narrowed the target region from 0.8 cM to 0.15 cM in our small mapping population. Three MEST markers, TC89380 (PPI), TC92082 (PPLC), and AW624844 (DRRP), co-segregated with TG389 and *dgt*. Most of the MEST markers derived from MSR1 have paralogs in other MSRs, but BG643476 (PP, AT4g34830) and TC85079 (EFP, AT4g38810) are unique to MSR1 and MSR2, respectively (Figure 2).

We screened a BAC library with TG389 and obtained two BAC clones, 52M1 and 93O2. Each tomato BAC clone was estimated to be approximately 120 kb in length by pulsed-field gel electrophoresis (data not shown), but neither

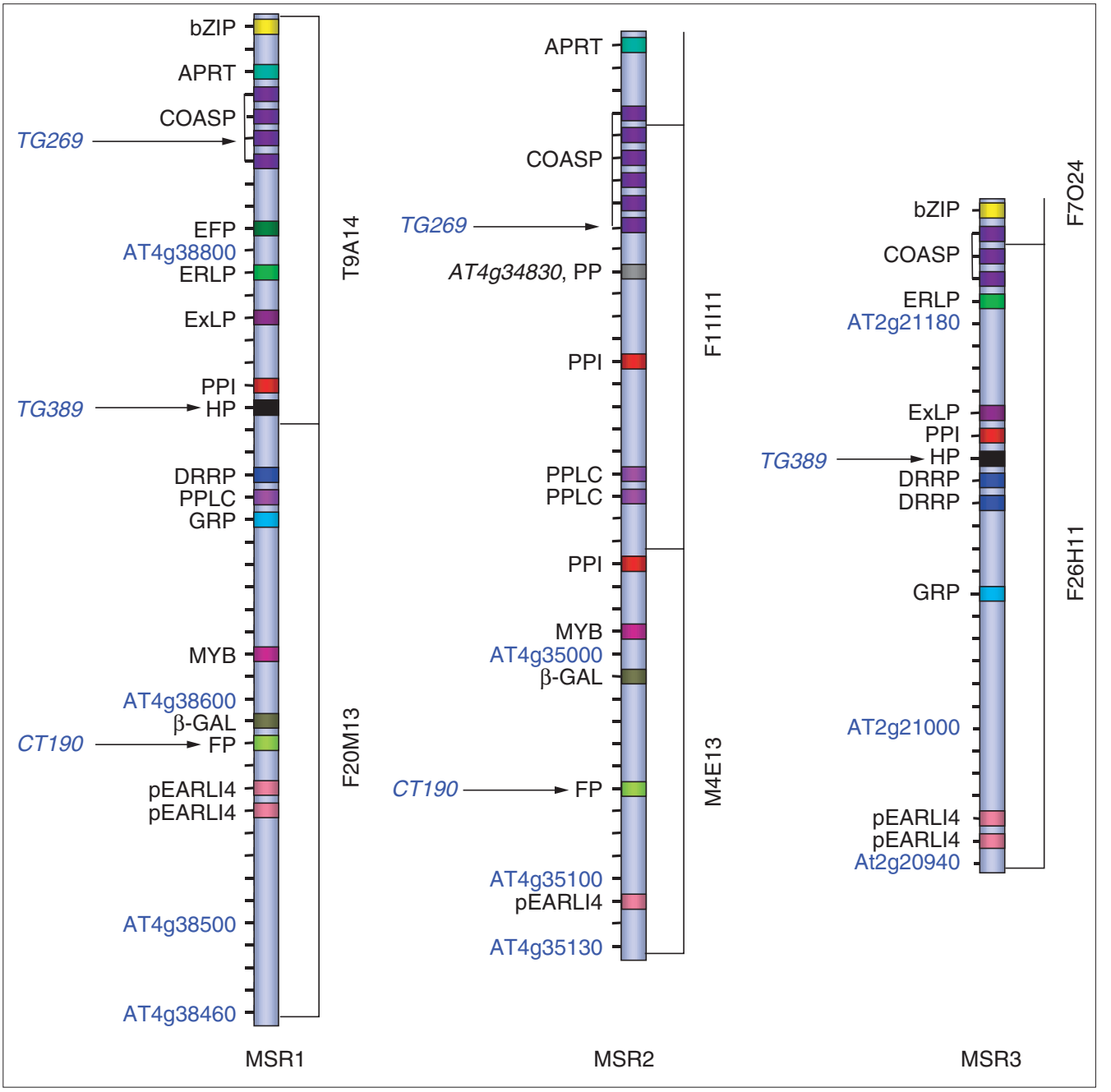

Figure 2
Three MSRs have conserved content and order of genes. Each MSR is aligned by the orientation of tomato RFLP markers. The *Arabidopsis* homeologs of tomato RFLP markers closely linked to the *Dgt* gene are indicated by arrows. Syntenic genes are represented by the same symbol and abbreviated name. EFP (AT4g38810) and PP (AT4g34830) are unique genes in MSR1 and MSR2, respectively, that were used for genotyping (see text). APRT, amidophosphoribosyl transferase; bZIP, basic leucine zipper transcription factor; COSAP, cluster of SAUR and auxin-induced proteins; DRRP, disease-resistance response protein; EFP, EF-hand containing protein-like protein; ERLP, endoplasmic reticulum lumen protein-retaining receptor; ExLP, extensin-like protein; FP, farnesylated protein; GRP, glycine-rich protein; HP, hypothetical protein; MYB, Myb transcription factor; pEARLI4, phospholipase-like protein; PPLC, phosphatidylinositol-specific phospholipase C precursor; PP, putative protein; PPI, cyclophilin-type peptidyl-prolyl isomerase; β-GAL, β-galactosidase-like protein. Bars to the right of the MSRs represent the corresponding annotated *Arabidopsis* BAC clones.

contained marker TG269 or CT190 sequences (data not shown) to allow confirmation that the *dgt* locus was located on either BAC clone. When both BAC clones were probed with the newly identified MEST molecular markers, TC98260 (the intervening marker between TG389 and CT190), was detected only on BAC 52M1, whereas

Table 1

CAPS markers converted from RFLP markers and microsyntenic ESTs*

Arabidopsis gene†	C/P/H‡	Molecular markers mapped in the *dgt* region§	Type	Expect value¶	Primer sequences#	Restriction enzyme
AT4g38580	-	CT190¥	RFLP	-	5′-TTCTCGTCGCTAAAGGCAGT-3′ 5′-TCACACAAAACAATGGGTGTTCTT-3′	*Hin*fI
AT4g38620	2 / 2 / 40					
AT4g38630	1 / 1 / 1					
AT4g38660	1 / 2 / 16					
AT4g38670	1 / 2 / 15					
AT4g38680	1 / 1 / 1	TC98260 (GRP)	EST	$1.7e^{-40}$	5′-GTGCCTCACAATCAAAGGGTTTTA-3′ 5′-CTCCATAACCACGATTTCCTCCTC-3′	*Rsa*I
AT4g38690	4 / 4 / 5	TC92082 (PPLC)	EST	$1.4e^{-84}$	5′-TGGTTGAGCTGATTTTCTTGGTTT-3′ 5′-CCTGGTTCTGATTATCGCTCAGAT-3′	*Hin*fI
AT4g38700	1 / 2 / 5	AW624844 (DRRP)	EST	$1.2e^{-28}$	5′-AAACGTCATGGGCTAAGAGAGTTG-3′ 5′-TCTAGATGCAATGGCTTGTTTTCC-3′	*Apo*I
AT4g38730	-	TG389	RFLP	-	5′-TCACTAGCTCAAGGGAGTCATCTG-3′ 5′-ACCACTTTGACCATCATCGCAAGC-3′	*Hin*fI
AT4g38740	1 / 3 / 10	TC89380 (PPI)	EST	$5.2e^{-76}$	5′-CAAATCCAAAGGTTTTCTTTGACC-3′ 5′-CTGGTAGAAGCAACACAACAACCA-3′	*Hae*III
AT4g38790	1 / 2 / 4					
AT4g38810	1 / 2 / 2	TC85079 (EFP)	EST	$4.0e^{-129}$	5′-CGAAACTGGCTTCCCTTCTA-3′ 5′-AGTCAGGTGATGGACGGTTC-3′	*Ban*I
AT4g38830	10 / 12 /146					
AT4g38840	1 / 1 / 6					
AT4g38850	-	TG269	RFLP	-	5′-CAAATTCTTCCTCAGCTTGACT-3′ 5′-TGATCTCACATCTTGCTTGCG-3′	*Dde*I
AT4g38880	1 / 1 / 1	TC87150 (APRT)	EST	$3.8e^{-90}$	5′-CAGAAAAATGACTTGGAGGGAGAG-3′ 5′-CCAAGATTGTGAGGCTGTTAAAGG-3′	*Rsa*I
AT4g38900	1 / 1 / 3	TC47447 (bZIP)	EST	$1.3e^{-98}$	5′-AACTTGGAAGCGTCTGCACT-3′ 5′-GGACGACCTGTTTTCTGCAT-3′	*Rsa*I
AT4g34830	1 / 1 / 1	BG643476	EST	$5.3e^{-86}$	5′-GGTTGATGGACTGCATAAAAATCC -3′ 5′-TGCAAATTCCCAATTTACCATTTT -3′	*Hha*I
AT4g35050	2 / 3 / 4					

*The conversion rate of the amplicons, generated from ESTs in the target region, to CAPS was 75% (30/40) and 20% of CAPS markers (6/30) were successfully mapped in the *dgt* region. †Genes in MSR1 with exceptions of AT4g34830 and AT4g35050 in MSR2. ‡C/P/H, the number of CAPS markers developed/the number of PCR products investigated/the number of TBLASTN matches of interest (< e^{-20}). §Abbreviations in parentheses: GRP, glycine-rich protein; PPLC, phosphatidylinositol-specific phospholipase C precursor; DRRP, disease-resistance response protein; PPI, cyclophilin-type peptidyl-prolyl isomerase; EFP, EF-hand containing protein-like protein; APRT, amidophosphoribosyl transferase; bZIP; basic leucine zipper transcription factor. ¶Each value represents the *E*-value of a TBLASTN search using the *Arabidopsis* microsyntenic gene against the Tomato Gene Index using the Blosum 62 scoring matrix and default parameters. The *E*-value of a BLASTN search using tomato RFLP markers is not described in this table (see text). #Oligonucleotide sequences are only indicated for each MEST marker. ¥Amplified by modified PCR conditions: 3 min 30 sec for elongation and 2.5 mM $MgCl_2$.

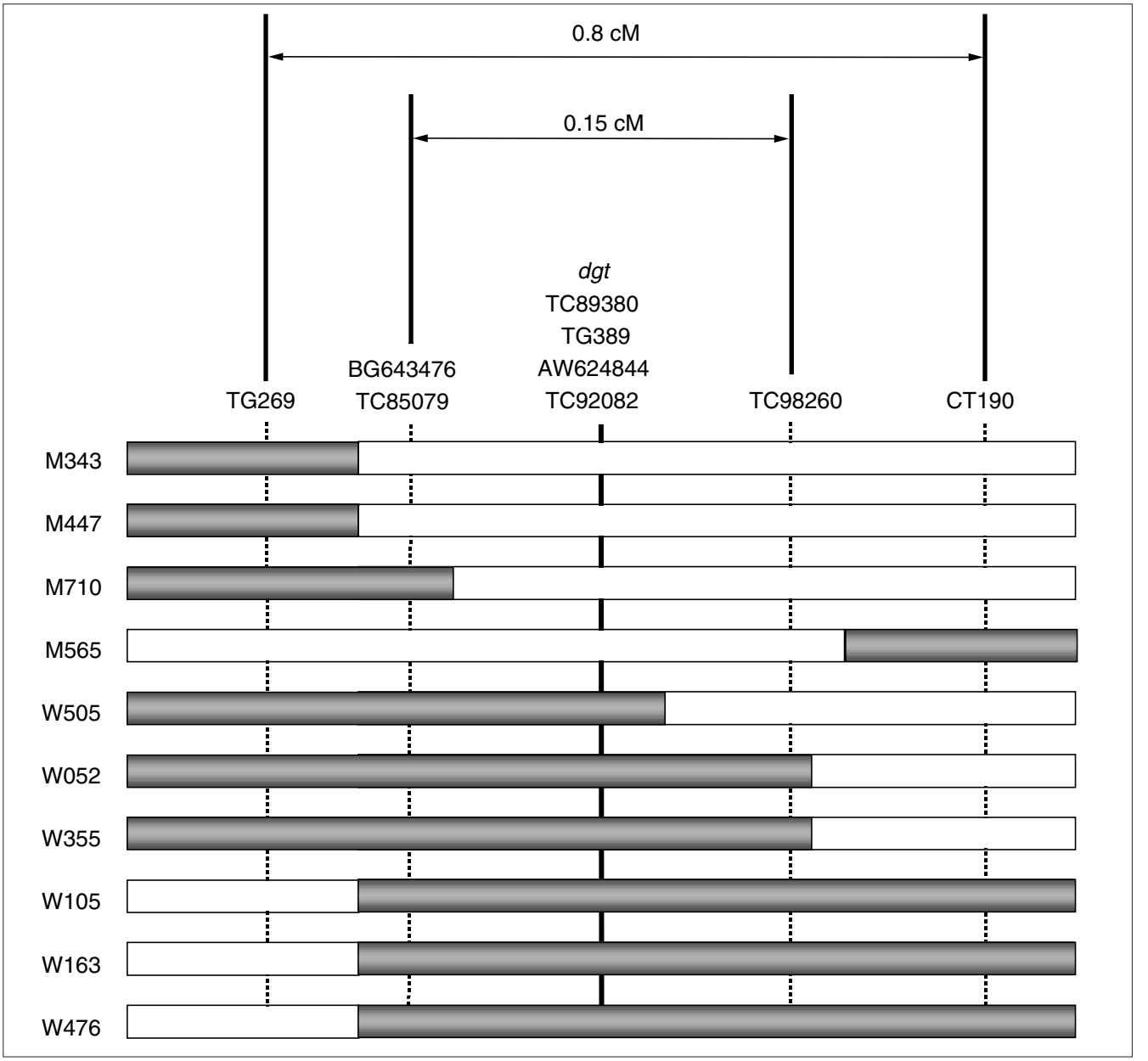

Figure 3
MEST markers substantially narrow the *Dgt* mapping region. Each bar shows the putative breakpoint of recombination in an informative BC_1 individual. Closed and open bars represent the chromosomal fraction of *L. pennellii* (wild-type phenotype) and *L. esculentum* (mutant phenotype), respectively. Mutant (M) or wild-type (W) phenotype is indicated for each individual plant and identifying number. On the basis of CAPS analysis, three MEST markers, BG643476, TC85079, and TC98260, were identified as new intervening markers between TG389 and the flanking RFLP markers, TG269 and CT190, with new recombination events identified in BC_1 individuals M710 and W505, respectively. TC89380, AW624844, and TC92082 co-segregated with TG389 and, thus are tightly linked to the *Dgt* gene. Genetic distances between molecular markers are indicated at the top.

BG643476 and TC85079 (the intervening markers between TG269 and TG389), were present only on BAC 93O2 (Figure 4). The three MEST markers that co-segregated with TG389 were detected on both BAC clones. These results permitted partial ordering of the newly identified MEST markers and demonstrated that the *dgt* locus was present within the two BAC clones.

Subsequent BAC end-sequencing and BLAST searches of the Tomato Gene Index identified BAC 93O2 and 52M1 insert DNA ends nearest the telomere as containing genomic nucleotide sequence for two previously identified MESTs - TC92082 and TC98260, respectively. This information placed TC92082 between TC98260 and the MEST markers that co-segregated with TG389 (Figures 3,5). The BAC

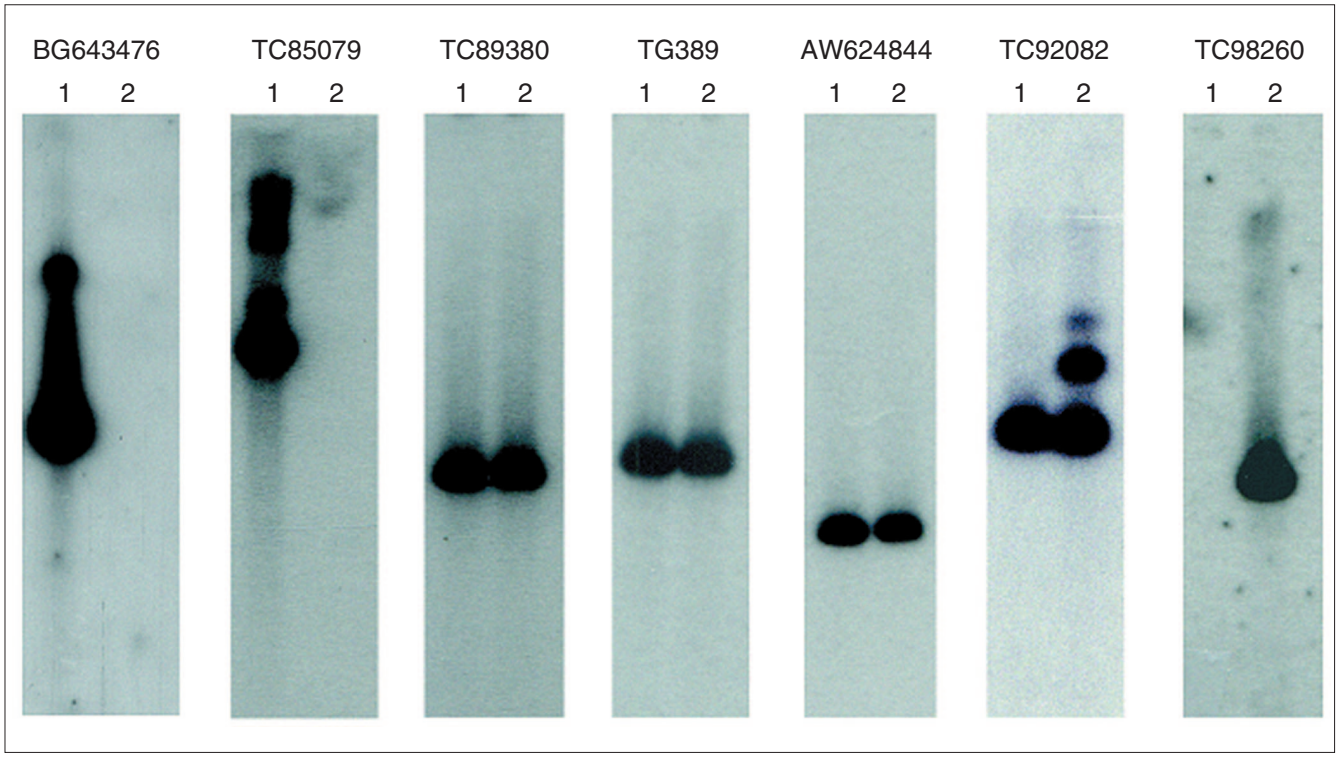

Figure 4
Southern hybridization of TG389-containing BAC clones identifies MEST gene order. DNA samples of BAC clones were digested with *Hin*dIII. The separated fragments were blotted and probed with MEST markers indicated at the top of each panel. The left and right lanes of each panel are 93O2 (1) and 52M1 (2), respectively.

end-sequences nearest to the centromere for BACs 93O2 and 52M1 were identified as BG643476 (PP) and TC87127 (chalcone synthase-like protein, CHSP), respectively. BG643476 was confirmed as a MEST when the sequence was later posted to the tomato EST database. The BAC end-location placed BG643476 between markers TG269 and TC85079. Taken together, these results strongly suggest that TC85079 and TC98260 are the closest flanking molecular markers to *dgt* present in the two overlapping BAC clones containing TG389. BLAST searches of the BAC end-sequences against the *Arabidopsis* Information Resource (TAIR) database further confirmed microsynteny between the *dgt* region of tomato chromosome 1 and the three MSRs in *Arabidopsis* (Figure 5).

As would be expected if microsynteny were maintained, two additional markers, TC87150 (amidophosphoribosyl transferase, APRT) and TC47447 (basic leucine zipper transcription factor, bZIP), were determined to be non-informative MEST markers with respect to our positional cloning strategy because they fell outside TG269. However, their sequences maintained microsyntenic alignment and co-segregated with tomato RFLP marker TG258 (data not shown), which is proximal to TG269 on chromosome 1 (see Figure 1). Given no rearrangement of tomato chromosome 1 and that microsynteny between tomato and *Arabidopsis* genomes remains firm, the predicted order of the new MEST and BAC end-generated markers in tomato would be as follows: TC47447 → TC87150 → TG269 → BG643476 → TC85079 → TC87127 → TC89380 → TG389 → AW624844 → TC92082 → TC98260 → CT190 (centromere to telomere, Figure 5). The exact order of these markers and the position of *dgt* will be determined by complete sequencing of tomato BACs 93O2 and 52M1.

Discussion

With the sequencing of the entire *Arabidopsis* genome, comparative mapping and homeology-based gene cloning is now available in other species via microsyntenic alignment of molecular markers and genes against the sequenced reference genome [21]. Although exceptions to the blanket application of this approach have been noted [22], our study successfully applied microsynteny analysis between the tomato and *Arabidopsis* genomes to facilitate positional cloning of the *Dgt* gene in tomato. Using three sequential tomato RFLP markers from our target region, we searched the *Arabidopsis* genome database and found three MSRs in the *Arabidopsis* genome that enabled us to construct a detailed molecular map of the target area. Comparison of the

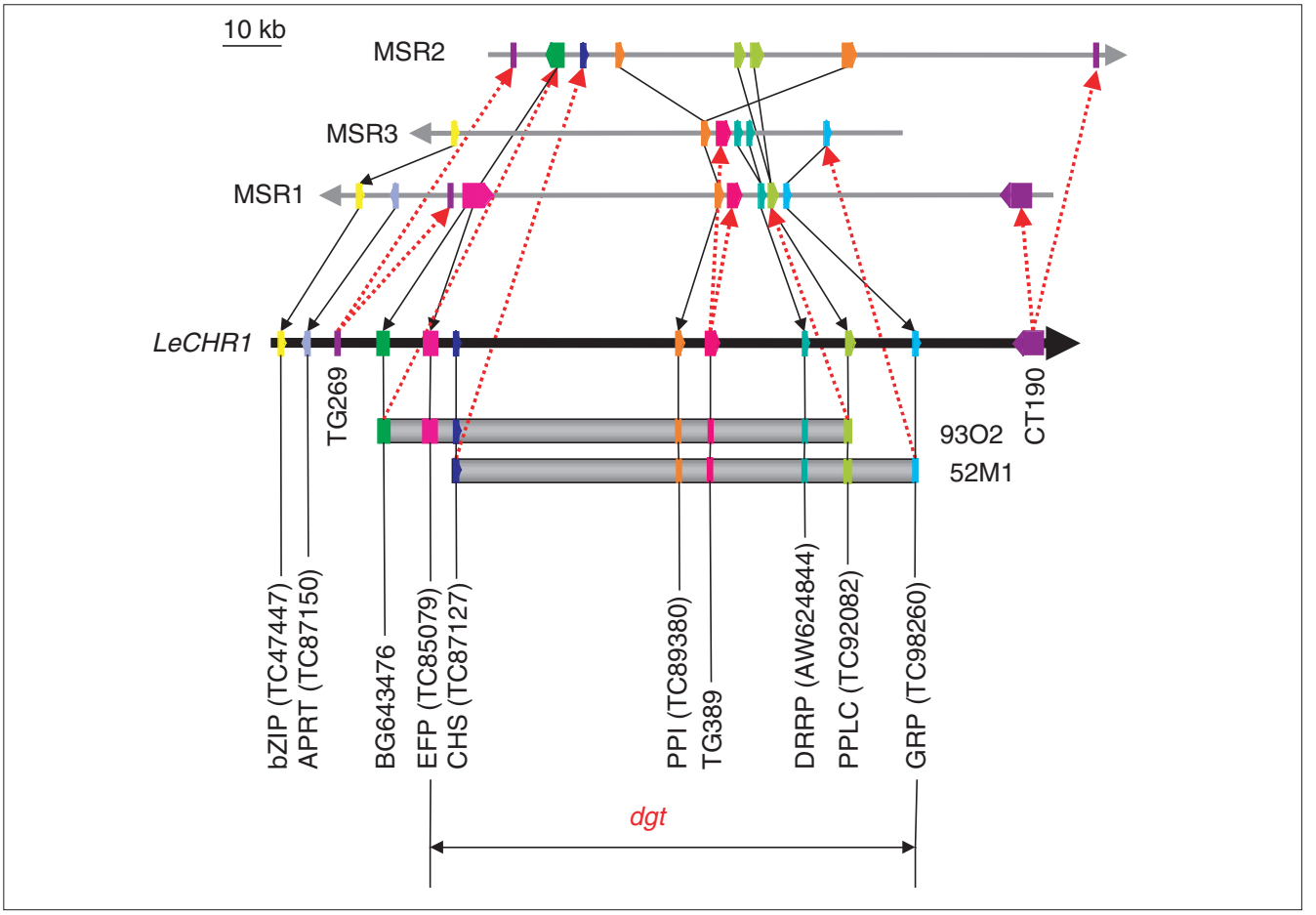

Figure 5
The *Dgt* locus is contained within two overlapping BAC clones. The comparative map shows MEST positions on tomato chromosome 1 (*LeCHR1*) based on microsynteny. Orientation is denoted by arrowheads on each chromosome. Arrows from *LeCHR1* to the *Arabidopsis* MSRs represent BLASTN matches of tomato RFLP markers and BAC insert ends, whereas arrows from MSR1 to *LeCHR1* indicate MESTs confirmed by CAPS analysis. See the legends for Table 1 and Figure 2 for MEST abbreviations. The order of MESTs is predicted by the order of *Arabidopsis* genes in the MSRs, DNA blot analysis, and sequence analysis (see Figure 4). On the basis of this comparative map, the *Dgt* gene would be located between TC85079 and TC98260 on BAC 93O2 and/or 52M1. Note that the scale bar represents *Arabidopsis* chromosome physical length.

three MSRs from the *Arabidopsis* genome shows that the content and order of the genes are well conserved within these regions. MSR1 and MSR3 exhibit reverse polarity compared to orientation of tomato RFLP markers, whereas the chromosomal segment MSR2 has the same polarity as tomato. Using information from the three *Arabidopsis* MSRs, we identified eight new molecular markers on tomato chromosome 1, two of which narrowed the genetic distance to the *dgt* locus from 0.8 cM to less than 0.2 cM and provided the necessary data to confirm the location of the *dgt* locus on two overlapping BAC clones, thus avoiding the need for a chromosomal walk.

Our strategy for microsynteny-based comparative mapping was straightforward and simple. First, the BLASTN program at TAIR was used to identify microsyntenic regions in *Arabidopsis* containing homeologs of tester markers (tomato RFLPs) in similar order within a relatively small physical interval (1-2 cM). Several published reports supported our strategy of using simple computer-based comparisons between tester and reference nucleotide sequences. Two recent comparative sequence analyses clearly presented microsynteny of a 105 kb BAC DNA insert [10] and a 57 kb cosmid DNA insert [11] from tomato chromosomes 2 and 7, respectively, to their homeologous counterparts in the *Arabidopsis* genome. Paterson *et al.* [23] suggested that 43% to 58% of chromosomal segments of less than 3 cM remain colinear in distantly related species. Direct nucleotide sequence comparisons (BLASTN) were used for each search as this provides a more stringent test of homology between tomato and *Arabidopsis* sequences than do the conceptual translations of DNA sequences (TBLASTX) that can be used for less stringent comparisons between evolutionarily divergent species [12]. In our study, BLASTN searches using default

parameters enabled us to find a high level of microsynteny without encountering any major differences between the highest-scoring matches in either TBLASTX or BLASTN searches (data not shown). However, the *E*-value for each BLAST hit was quite different between TBLASTX and BLASTN, suggesting that an arbitrary threshold *E*-value is critical to determine acceptance of BLAST matches between evolutionarily distant nucleotide sequences and to place the homeologs on *Arabidopsis* chromosomes.

The occurrence of multiple MSRs as a result of segmental duplications on the *Arabidopsis* genome has been demonstrated by other comparative sequence analyses [10,12]. Most MEST markers identified in this study were conserved in at least two MSRs, suggesting that the use of genes conserved in multiple MSRs increases the probability of obtaining microsyntenic markers that are conserved between two species. Using gene annotations in the *Arabidopsis* genome database, this simple computer-based analysis not only avoided the time and cost required for hybridization-based experimental methods [24-26], but also allowed us to make informed decisions as to the location of unidentified genes of interest, as well as intervening genes between test markers in the target area.

The second aspect of the comparative microsynteny approach was identification of tomato ESTs homeologous for the coding sequence of each *Arabidopsis* gene in the most closely related MSR. PCR primer sets were designed from sequences of high-scoring ESTs and robust PCR products were obtained from 15% of the ESTs. Thirty of the PCR products were then converted to co-dominant CAPS markers and used to genotype a small number of informative recombinant plants. The use of ESTs in comparative mapping has been successfully applied to *Brassica* species [24,25] and maize [27]. However, the success of EST-derived CAPS markers depends on a well-established EST database. The Tomato Gene Index of The Institute for Genomic Research (TIGR) is the third largest sequence database for plants and provided 131,988 EST sequences at the time of this study [20,28]. The tentative consensus (TC) sequences of assembled ESTs can be used for integration of complex mapping data and identification of orthologous genes between divergent species [28]. Expressed gene sequences were used in this study as they can be converted to molecular markers more consistently than non-coding regions when comparing distantly related species [12]. Moreover, two recent comparative sequence analyses using monocots versus *Arabidopsis* [29] and dicots versus *Arabidopsis* [11] clearly showed that exon sizes are well conserved in homeologs even between monocots and dicots, whereas intron length varied in rice versus barley and tomato versus *Arabidopsis*. In BLAST searches using either the processed nucleotide sequence or the predicted protein sequence of an *Arabidopsis* gene, the score of each BLAST match was clearly higher and the *E*-value was significantly lower when compared to the use of either non-processed nucleotide sequence or intergenic nucleotide sequence. Presumably, use of the predicted protein sequence of an *Arabidopsis* gene increases the probability of identifying the microsyntenic EST in comparative mapping of distantly related species, considering the average substitution rate, 6×10^{-9}/nucleotide site/year, of nucleotides in plants [30] and the separation, 112 million years ago, of tomato and *Arabidopsis* [10].

The MEST-derived comparative map indicating that three regions of the *Arabidopsis* genome are related to each other (Figures 2,5) supports the hypothesis that at least two rounds of duplication occurred in the *Arabidopsis* genome followed by selective deletion of genes and/or minor rearrangements [10]. Minor rearrangements of microsyntenic genes could potentially impede high-resolution mapping if placement of the syntenic markers within a relatively large segment indicated a missing sequence fragment, as reported for the comparative mapping of maize and sorghum [31]. Given that several genes exist between the homeologs of the two closest flanking MESTs, TC85079 (EFP) and TC98260 (GRP), and neither a deletion nor local rearrangement of the *Dgt* counterpart has occurred in the MSRs, the *Dgt* counterpart could be one of the annotated genes between the homeologs of the two closest flanking MESTs. It remains to be seen, however, whether additional tomato genes are present in the region.

Our results indicate that comparative microsynteny-based mapping can facilitate positional cloning of a target gene when information on genomic location is limited. Ku *et al.* [32] recently utilized microsynteny-based comparative mapping to add new molecular markers in a 0.067 cM region defined by a previously determined 100 kb BAC clone containing the *ovate* locus in tomato. Here, microsynteny-based comparative mapping was used to define the position of the *Dgt* gene within a much larger region (0.8 cM) of the tomato genome by contributing six intervening MEST markers, initially derived from *Arabidopsis* gene sequences, between TG269 and CT190. The approach proved to be less technically complicated than other fingerprinting methods and points to several co-segregating genes for further investigation as potential *Dgt* candidates. We anticipate that this general approach will contribute significantly to the development of dense molecular marker maps for a variety of higher plant species to expedite map-based cloning.

Materials and methods

Mapping population

A mapping population consisting of 1,308 backcross (BC_1) plants derived from a backcross between *L. esculentum* cv. Ailsa Craig (*dgt*/*dgt*) x F_1 [*L. esculentum* cv. Ailsa Craig (*dgt*/*dgt*) x *L. pennellii* (*Dgt*/*Dgt*)] were initially screened with RFLP markers localized to the distal end of chromosome 1 [8]. RFLP analyses were performed to identify

markers closely linked to the *Dgt* gene (Figure 1). Two RFLP markers, TG269 and CT190, were converted to CAPS markers [33] and established as flanking markers to *Dgt*, approximately 0.8 cM apart. Ten plants were identified as recombinant within this interval and were designated as the informative recombinant population in this study. Ten non-recombinant plants were randomly selected for use as the non-informative population.

Comparative sequence analysis

Arabidopsis genome database searches were performed with BLASTN software in TAIR [18]. Tomato EST database searches were performed with TBLASTN software in the Tomato Gene Index at TIGR [27]. The resulting tentative consensus (TC) sequences or high-scoring singleton EST sequences were analyzed and used to design PCR primers. Further sequence analysis was carried out with the Genetics Computer Group 10 (GCG; Madison, WI) program. Alignments of conserved regions within a multigene family were made using PILEUP and adjusted manually to design gene-specific primers. All PCR primers used in this research were designed using Primer3 [34] software and are listed in Table 1.

CAPS analysis of molecular markers

Tomato genomic DNA was extracted from leaf tissue as described by Dellaporta *et al.* [35]. RFLP markers and ESTs were converted to co-dominant PCR-based molecular markers (CAPS) as described by Konieczny and Ausubel [33]. Amplification reactions consisted of a 25 μl reaction containing 100 ng genomic DNA or 20 ng BAC DNA, 200 μM dNTPs, 200 nM each forward and reverse primer, 1x reaction buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl), 1.5 mM $MgCl_2$, and 1 U Maxi Taq polymerase [36]. Standard temperature parameters for amplifying ESTs from genomic DNA in this study were as follows: initial denaturation at 94°C for 3 min; 40 cycles of 94°C for 45 sec, 58°C for 1 min, and 72°C for 3 min; final elongation at 72°C for 5 min. All PCR experiments were performed in a RoboCycler 96 Gradient with Hot Top Assembly (Stratagene, La Jolla, CA). Following PCR, products were digested with the indicated restriction enzyme (Table 1) to yield co-dominant markers. The digested PCR products were resolved in either 2% or 2.5% agarose electrophoretic gels in 1x TAE buffer (40 mM Tris-acetate, 1 mM EDTA pH 8.0) and visualized by ethidium bromide staining.

BAC library screening and analysis

TG389, an RFLP marker tightly linked to *dgt* and situated between markers TG269 and CT190, was used to screen a tomato (*L. esculentum* cv. Heinz 1706) BAC library [37] as described by the Clemson University Genome Institute [38]. Two tomato BAC clones, 52M1 and 93O2, were isolated by hybridization to the TG389 probe. BAC DNA was isolated by alkaline lysis [39] modified for a 200 ml LB broth culture volume containing 17 mg/l chloramphenicol. BAC DNA samples were digested with *Not*I to liberate tomato genomic insert DNA then digested with *Hin*dIII and separated by electrophoresis in 1% agarose gels. Fractionated DNAs were transferred to Hybond-XL membrane (Amersham Pharmacia Biotech, Piscataway, NJ) as described by Sambrook *et al.* [39]. Membranes were prehybridized for 1 to 2 h in Church buffer [40] before adding denatured probe in fresh Church buffer. Labeled probe was synthesized using a High Prime random priming kit, according to the manufacturer's instructions (Roche Molecular Biochemicals, Indianapolis, IN), from 100 ng of genomic PCR product and 50 μCi [α-^{32}P]dCTP. After a 30 min incubation at 37°C, labeled probes were purified on Sephadex G50 spin columns. Membranes were hybridized for 16 h at 65°C and briefly rinsed in low-stringency buffer containing 40 mM Na_2HPO_4 (pH 7.2), 1 mM EDTA (pH 8.0) and 5% SDS at room temperature, followed by washing in low-stringency buffer at 65°C for 30 min. Membranes were washed twice at 65°C for 30 min in high-stringency buffer containing 40 mM Na_2HPO_4 (pH 7.2), 1 mM EDTA (pH 8.0) and 1% SDS, before exposure to autoradiographic film.

Direct BAC end sequencing

For sequencing, BAC DNA was isolated from a 100 ml LB culture containing 17 mg/l chloramphenicol using a QIAGEN Plasmid Midi Kit (Valencia, CA), following the manufacturer's instruction. The DNA sample was then subjected to digestion of co-eluted residual genomic DNA with Plasmid-Safe™ ATP-dependent DNase (Epicentre, Madison, WI). BAC insert ends were sequenced on an ABI377 automated sequencer (Applied Biosystems, Foster City, CA) using the ABI PRISM® BigDye™ terminator cycle sequencing kit with T7 and SP6 sequencing primers. Sequencing was performed by the Central Services Laboratory of the Center for Gene Research and Biotechnology at Oregon State University.

Acknowledgements

This research was supported by a grant from the National Science Foundation Integrative Plant Biology Program. We thank S.D. Tanksley (Cornell University) for providing the RFLP probes and J.J. Giovannoni, S. Knapp, and G.B. Martin for critical review of this manuscript.

References

1. Bennetzen JL, Freeling M: **Grasses as a single genetic system: genome composition, colinearity, and compatibility.** *Trends Genet* 1993, **9:**259-261.
2. Moore G, Devos KM, Wang Z, Gale MD: **Cereal genome evolution. Grasses, line up and form a circle.** *Curr Biol* 1995, **5:**737-739.
3. Schmidt R: **Synteny: recent advances and future prospects.** *Curr Opin Plant Biol* 2000, **3:**97-102.
4. The *Arabidopsis* Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408:**796-815.
5. Mysore KS, Tuori RP, Martin GB: ***Arabidopsis* genome sequence as a tool for functional genomics in tomato.** *Genome Biol* 2001, **2:**1003.1-1003.4.
6. Kowalski SP, Lan T-H, Feldmann KA, Paterson AH: **Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea***

chromosomes reveals islands of conserved organization. *Genetics* 1994, **138:**499-510.
7. Kilian A, Chen J, Han F, Steffenson B, Kleinhofs A: **Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle.** *Plant Mol Biol* 1997, **35:**187-195.
8. Tanksley SD, Ganal MW, Prince JP, de Vivente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB: **High density molecular linkage maps of the tomato and potato genomes.** *Genetics* 1992, **132:**1141-1160.
9. Livingstone KD, Lackney VK, Blauth JR, van Wijk R, Jahn MK: **Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae.** *Genetics* 1999, **152:**1183-1202.
10. Ku H-M, Vision T, Liu J, Tanksley SD: **Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny.** *Proc Natl Acad Sci USA* 2000, **97:** 9121-9126.
11. Rossberg M, Theres K, Acarkana A, Herrero R, Schmitt T, Schumacher K, Schmitz G, Schmidt R: **Comparative sequence analysis reveals extensive microcolinearity in the *Lateral suppressor* regions of the tomato, *Arabidopsis*, and *Capsella* genomes.** *Plant Cell* 2001, **13:**979-988.
12. Grant D, Cregan P, Shoemaker RC: **Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*.** *Proc Natl Acad Sci USA* 2000, **97:**4168-4173.
13. Kelly MO, Bradford KJ: **Insensitivity of the *diageotropica* tomato mutant to auxin.** *Plant Physiol* 1986, **82:**713-717.
14. Coenen C, Lomax TL: **The *diageotropica* gene differentially affects auxin and cytokinin responses throughout development in tomato.** *Plant Physiol* 1998, **117:**63-72.
15. Mito N, Bennett AB: **The *diageotropica* mutation and synthetic auxins differentially affect the expression of auxin-regulated genes in tomato.** *Plant Physiol* 1995, **109:**293-297.
16. Nebenführ A, White TJ, Lomax TL: **The *diageotropica* mutation alters auxin induction of a subset of the *AUX/IAA* gene family in tomato.** *Plant Mol Biol* 2000, **44:**73-84.
17. Rice MS, Lomax TL: **The auxin-resistant *diageotropica* mutant of tomato responds to gravity via an auxin-mediated pathway.** *Planta* 2000, **210:**906-913.
18. **The *Arabidopsis* Information Resource** [http://www.Arabidopsis.org/Blast].
19. Blanc G, Barakat A, Guyot R, Cooke R, Delseney M: **Extensive duplication and reshuffling in the *Arabidopsis* genome.** *Plant Cell* 2000, **12:**1093-1101.
20. **Tomato Gene Index** [http://www.tigr.org/tdb/lgi/].
21. Bancroft I: **Duplicate and diverge: the evolution of plant genome microstructure.** *Trends Genet* 2001, **17:**89-93.
22. Bennetzen JL: **Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions.** *Plant Cell* 2000, **12:**1021-1029.
23. Paterson AH, Lan TH, Reischmann KP, Chang C, Lin YR: **Toward a unified map of higher plant chromosomes, transcending the monocot-dicot divergence.** *Nat Genet* 1996, **14:**380-382.
24. Conner JA, Conner P, Nasrallah ME, Nasrallah JB: **Comparative mapping of the *Brassica* S locus region and its homeolog in *Arabidopsis*: implications for the evolution of mating systems in the Brassicaceae.** *Plant Cell* 1998, **10:**801-812.
25. Lan TH, DelMonte TA, Reischmann KP, Hyman J, Kowalski SP, McFerson J, Kresovich S, Paterson A: **An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana.*** *Genome Res* 2000, **10:**776-788.
26. Pan Q, Liu Y-S, Budai-Hadrian O, Selan M, Carmel-Goren L, Zamir D, Fluhr R: **Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and *Arabidopsis*.** *Genetics* 2000, **155:**309-322.
27. Davis GL, McMullen MD, Baysdorfer C, Musket T, Grant D, Staebell M, Xu G, Polacco M, Koster L, Melia-Hancock S, *et al.*: **A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map.** *Genetics* 1999, **152:**1137-1172.
28. Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J: **The TIGR gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species.** *Nucleic Acid Res* 2001, **29:**159-164.
29. Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Liuling Y, Shiloff BA, Bennetzen JL: **Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes.** *Plant Physiol* 2001, **125:**1342-1353.
30. Wolfe KH, Li W-H, Sharp PM: **Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs.** *Proc Natl Acad Sci USA* 1987, **84:**9054-9058.
31. Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A: **The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4.** *Plant Cell* 2000, **12:**381-392.
32. Ku HM, Liu J, Doganlar S, Tanksley SD: **Exploitation of *Arabidopsis*-tomato synteny to construct a high-resolution map of the *ovate*-containing region in tomato chromosome 2.** *Genome* 2001, **44:**470-475.
33. Konieczny A, Ausubel FM: **A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype specific PCR-based markers.** *Plant J* 1993, **4:**403-410.
34. **Primer3** [http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi]
35. Dellaporta SL, Wood J, Hicks JB: **A plant DNA minipreparation: version II.** *Plant Mol Biol Rep* 1983, **1:**19-21.
36. Pluthero FG: **Rapid purification of high-activity *Taq* DNA polymerase.** *Nucleic Acid Res* 1993, **21:**4850-4851.
37. Budiman MA, Mao L, Wood TC, Wing RA: **A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing.** *Genome Res* 2000, **10:**129-136.
38. **Clemson University Genome Institute** [http://www.genome.clemson.edu]
39. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
40. Church GM, Gilbert W: **Genomic sequencing.** *Proc Natl Acad Sci USA* 1984, **81:**1991-1995.