

# Complex Organization and Evolution of the Tomato Pericentromeric Region at the *FER* Gene Locus<sup>1[w]</sup>

Romain Guyot<sup>2</sup>, Xudong Cheng<sup>2</sup>, Yan Su, Zhukuan Cheng, Edith Schlagenhauf, Beat Keller, and Hong-Qing Ling\*

Institute of Plant Biology, University of Zurich, 8008 Zurich, Switzerland (R.G., E.S., B.K., H.-Q.L.); and Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Chaoyang District, Beijing 100101, China (X.C., Y.S., Z.C., H.-Q.L.)

Tomato (*Lycopersicon esculentum*) is a model species for molecular biology research and a candidate for large-scale genome sequencing. Pericentromeric heterochromatin constitutes a large portion of the tomato chromosomes. However, the knowledge of the structure, organization, and evolution of such regions remains very limited. Here, we report the analysis of a 198-kb sequence near the *FER* gene, located in a distal part of pericentromeric heterochromatin on the long arm of tomato chromosome 6. Nine genes, one pseudogene, and 55 transposable elements (TEs) were identified, showing a low gene density (19.8 kb/gene) and a high content of transposable elements (>45% of the sequence). Six genes (*56B23\_g3*, *g5*, *g7*, *g8*, *g9*, and *g10*) have perfect matches (>98% identity) with tomato expressed sequence tags. Two genes (*56B23\_g1* and *g6*), which share <98% sequence identity with expressed sequence tags, were confirmed for transcriptional activity by reverse transcription-PCR. The genes were not uniformly distributed along the sequence and grouped into gene islands separated by stretches of retrotransposons, forming a pattern similar to that found in the gene-rich regions of the large genomes of maize (*Zea mays*) and Triticeae. Long terminal repeat retrotransposons account for 60% of the TE sequence length. Sixteen of 55 TEs were completely new and remain unclassified. Surprisingly, five of the seven identified DNA transposons were closely associated with coding regions. The action of transposable elements and DNA rearrangements form the molecular basis of the dynamic genome evolution at the *FER* locus. Multiple rounds of genome duplication in Arabidopsis (*Arabidopsis thaliana*) and subsequent gene loss have generated a mosaic pattern of conservation between tomato and Arabidopsis orthologous sequences. Our data show that the distal parts of pericentromeric heterochromatin may contain many valuable genes and that these regions form an evolutionary active part of the tomato genome.

The chromosomes of higher eukaryotes are cytologically characterized by the presence of lightly and strongly staining material called euchromatin and heterochromatin, respectively. In plant genomes, euchromatic DNA contains most of the active genes and heterochromatin was generally thought to be transcriptionally silent or devoid of genes. Heterochromatic DNA is mainly confined to centromeric regions and to variable lengths of DNA flanking centromeres of each chromosome, called pericentromeric regions. While centromeric regions are essential for correct segregation of chromosomes during mitosis and meiosis, the function of pericentromeric regions in plants remains unknown. In the Arabidopsis (*Arabidopsis thaliana*) and rice (*Oryza sativa*) genomes, centromere regions consist of specific tandemly repeated satellite DNA, occasionally interrupted by the insertion of

*Gypsy*-type long terminal repeat (LTR) retrotransposons (Copenhaver et al., 1999; Wu et al., 2004; Zhang et al., 2004), while pericentromeric regions are rich in middle repetitive elements, including transposons, retrotransposons, and pseudogenes (Copenhaver et al., 1999). Surprisingly, more than 200 expressed genes in the pericentromeric regions of the Arabidopsis genome (Copenhaver et al., 1999) and several active genes within the rice chromosome 8 centromere (Nagaki et al., 2004) have been discovered, indicating that the transcription of these genes is not affected by the heterochromatic environment.

The Solanaceae family contains many economically important species such as tomato (*Lycopersicon esculentum*), potato (*Solanum tuberosum*), tobacco (*Nicotiana tabacum*), eggplant (*Solanum melongena*), pepper (*Capiscum* spp.), and petunia species. Tomato is the model species in Solanaceae as it has a diploid genome ( $2n = 24$ ), the smallest genome size (953 Mb) among Solanaceae (Arumuganathan and Earle, 1991), a high-density genetic map (Tanksley et al., 1992), and a large mutant collection (<http://tgrc.ucdavis.edu>). Cytologically, tomato pachytene chromosomes show a typical morphology. Wide chromatic regions were observed on each side of the achromatic structure of the centromere, while distal regions of chromosome arms stain very lightly (Brown, 1949; Barton, 1950). In tomato, chromatic regions are pericentromeric heterochromatin consisting of a mosaic of highly and lightly

<sup>1</sup> This work was supported by the Chinese Ministry of Science and Technology (grant nos. 2004AA222061 to H.-Q.L. and 2002AA225011 to Z.C.), by the Chinese National Science Foundation (Talented Young Scientist grant no. 30225029 to H.-Q.L. and no. 30325008 to Z.C.), and by the Swiss National Science Foundation (grant nos. 3100-65114 to B.K. and 31-55288.98. to H.-Q.L.).

<sup>2</sup> These authors contributed equally to the paper.

\* Corresponding author; e-mail [hqling@genetics.ac.cn](mailto:hqling@genetics.ac.cn); fax 86-10-64860377.

[w] The online version of this article contains Web-only data. [www.plantphysiol.org/cgi/doi/10.1104/pp.104.058099](http://www.plantphysiol.org/cgi/doi/10.1104/pp.104.058099).

chromosome-specific chromatic zones. In contrast to Arabidopsis and rice, tomato chromosomes have wide blocks of pericentromeric heterochromatin accounting for 77% of the nuclear DNA (Peterson et al., 1996), while euchromatin is limited to small contiguous stretches in the distal parts of the chromosome arms. A similar dominance of heterochromatin in genome organization was observed in *Medicago truncatula*, a model genome for legumes, for which pericentromeric heterochromatin represents 60% of the whole genome (Kulikova et al., 2001). Very little is known about the composition and organization of heterochromatin regions in the tomato genome. In Arabidopsis and rice as well as in numerous other plant species, heterochromatin DNA is mainly composed of repetitive DNA sequences. Considering cytological observations, a large amount of repetitive DNA could be expected in the tomato genome. In contrast, DNA reassociation experiments indicated that most of the tomato genome (73%) is composed of single-copy sequences, suggesting that many of them must reside in the heterochromatin regions (Peterson et al., 1998). Analysis of 1,205 bacterial artificial chromosome (BAC) end sequences have confirmed that the tomato genome is composed of 70% unique or uncharacterized sequences and that repetitive sequence is around 12% (Budiman et al., 2000). Initial deletion mapping experiments indicated that pericentromeric heterochromatin regions might contain few genes. However, there is increasing evidence that these regions contain important genes, such as the root-knot nematode resistance gene *Mi* (van Daelen et al., 1993), tobacco mosaic virus resistance gene *Tm-2A* (Motoyoshi et al., 1996), and the *Jointless-2* gene (Budiman et al., 2004). Altogether, these results suggest an atypical composition and structure of pericentromeric heterochromatin in tomato.

Recently, the *FER* gene involved in the iron uptake of roots has been identified and isolated from the tomato BAC clone 56B23, which was mapped close to the centromere of chromosome 6 (Ling et al., 2002). Here, we report fluorescence in situ hybridization (FISH) experiments that show that 56B23 resides in the pericentromeric heterochromatin. Sequence analysis of 198 kb of the BAC 56B23 revealed a complex structure of active genes embedded in a transposon-rich environment. DNA rearrangements in tomato and multiple duplications in Arabidopsis were the mechanisms responsible for the mosaic organization of conservation between tomato and Arabidopsis.

## RESULTS

### Pericentromeric Localization of BAC 56B23 on Chromosome 6 by FISH

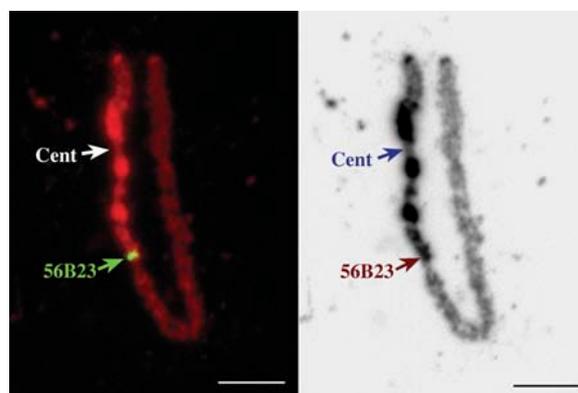
The tomato BAC clone 56B23 that carries the *FER* gene previously had been mapped to a genetic interval between markers TG590 and TG118 at a distance of approximately 10 cM from the chromosome 6 centromere (Tanksley et al., 1992; Ling et al., 1996).

To detect its physical location on the chromosome, we labeled the DNA isolated from 56B23 and hybridized it to pachytene chromosomes. Cytologically, 56B23 resides in a deeply staining region of the distal part of the pericentromeric heterochromatic region on the long arm of chromosome 6 (Fig. 1). The length from 56B23 to the centromere is about one-sixth of the total length of chromosome 6.

### Sequence Organization of BAC 56B23 from Tomato

The 198 kb of the BAC 56B23 was completely sequenced (accession no. AY678298). The overall GC content of 33.8% and a coding region GC content of 42.5% were similar to tomato genomic sequences analyzed in previous studies (Messeguer et al., 1991; Ku et al., 2000; Mao et al., 2001; Van der Hoeven et al., 2002). While coding regions accounted for 4.5% of the total sequence, putative transposable elements (TEs) and nonannotated sequences represented 45.5% and 50% of the whole BAC sequence, respectively.

Nine predicted genes and one pseudogene were identified, giving a gene density of 19.8 kb/gene (Table I; Fig. 2). While two predicted genes (*56B23\_g6* and *g10*) are intron free, seven genes (*56B23\_g1*, *g2*, *g3*, *g5*, *g7*, *g8*, and *g9*) have multiple exons, ranging from two to nine exons (Table I). All the predicted genes were found highly conserved in the *Lycopersicon* or *Solanum* expressed sequence tag (EST) databases (Table I), supporting gene prediction. Six of them (*g3*, *g5*, *g7*, *g8*, *g9*, and *g10*) have perfect matches with tomato ESTs (>98% sequence identity). The genes *g1* and *g6* share 91% and 95% identity with tomato ESTs (Table I), while the best EST hit of *56B23\_g2* was found in potato (CK256320, 94.6% identity). The three genes with lower identities to ESTs were further analyzed by reverse transcription (RT)-PCR for confirmation of



**Figure 1.** Physical mapping of BAC 56B23 by pachytene chromosome FISH. A, The BAC 56B23 (green signal) is mapped to the distal part of the heterochromatic region on the long arm of chromosome 6; Cent indicates centromere. Chromosomes were stained with 4',6-diamidino-phenylindole and pseudocolored as red. B, Inverted greyscale image of the same chromosome in A. The black portion represents the heterochromatic region of the chromosome. Bars = 5  $\mu$ m.

**Table 1.** Predicted genes in the BAC clone 56B23Le, Tomato; Ns, *Nicotiana glauca*; St, potato; id., nucleotide identity.

Gene Name	Predicted Function	Strand and Positions	Exon No.	BLASTN EST Best Hit	BLASTX Best Hit
56B23_g1 <sup>a,b</sup>	Putative protein	(+) 831–1,222 <i>bp</i>	2	Le, AI484069, E = e <sup>-122</sup> 95% id.	NP_915349 (rice) hypothetical protein, E = e <sup>-36</sup>
56B23_g2	Putative protein Kinase	(-) 35,141–37,493	3	St, CK256320, E = 0, 94.6% id.	NP_191154 (Arabidopsis) protein kinase E = e <sup>-107</sup>
56B23_g3	<i>Fer</i> protein <sup>c</sup>	(-) 50,573–52,222	4	Le, BF098042, E = 0, 100% id.	AAN39037 (tomato) bHLH transcriptional regulator E = e <sup>-150</sup>
56B23_g4 <sup>d</sup>	1,4-Benzoquinone reductase	(+) 66,102–66,236	1	Ns, BP748895, E = e <sup>-30</sup> , 88.1% id.	T05751 (Arabidopsis) hypothetical protein E = e <sup>-12</sup>
56B23_g5	Putative protein	(-) 71,693–88,596	8	Le, AI485185, E = 0, 99.2% id.	BAB09660 (Arabidopsis) unnamed protein product E = 0
56B23_g6 <sup>a</sup>	Putative ethylene receptor	(+) 156,222–156,722	1	Le, CK715332, E = 0, 91.2% id. <sup>§</sup>	AAF20093 (tobacco) putative ethylene receptor E = e <sup>-65</sup>
56B23_g7	Putative phospholipase C	(-) 163,090–169,745	9	Le, BI931651, E = 0, 98.4% id.	CAA63893 (potato) phospholipase C E = 0
56B23_g8	Putative phospholipase C	(-) 170,773–173,873	9	Le, BF096534, E = 0, 98.5% id.	CAA65127 ( <i>Nicotiana rustica</i> ) phospholipase C E = 0
56B23_g9	Putative cyclophilin	(+) 183,488–187,788	8	Le, BI931759, E = 0, 99.6% id.	NP_567029 (Arabidopsis) putative cyclophilin E = e <sup>-69</sup>
56B23_g10	Putative protein	(-) 188,283–188,645	1	Le, BG130993, E = 0, 100% id.	AAQ73526 ( <i>Mesembryanthemum crystallinum</i> ) EARLY flowering 4 E = e <sup>-23</sup>

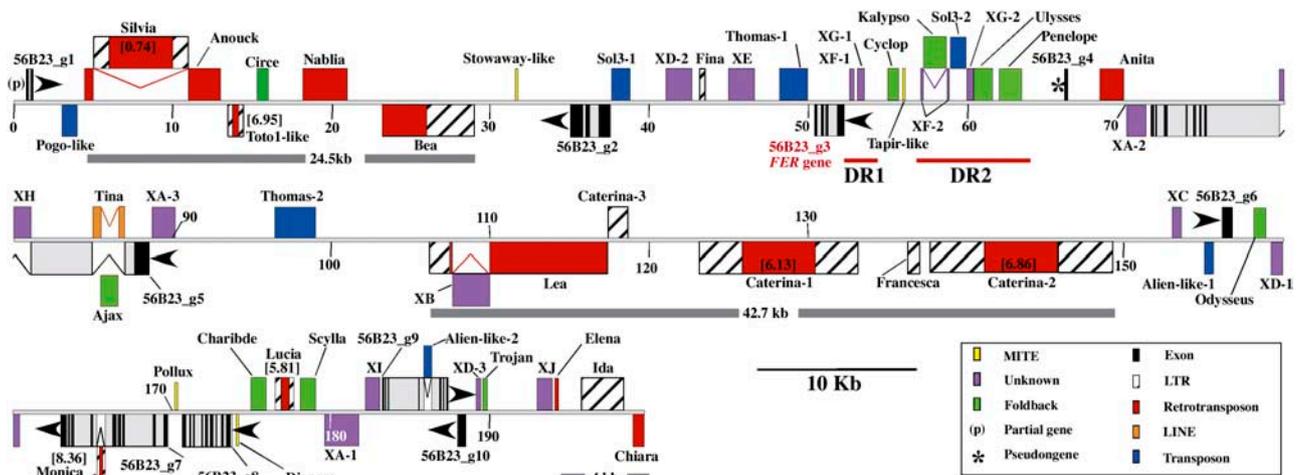
<sup>a</sup>Indicates that transcript of the gene was detected by RT-PCR.<sup>b</sup>Partial gene.<sup>c</sup>Ling et al. (2002).<sup>d</sup>Pseudogene.

active transcription. Transcripts of *56B23\_g1* and *56B23\_g6* were identified while the transcription of *56B23\_g2* was not observed under our experimental conditions (data not shown), indicating that the expression of *56B23\_g2* is possibly low, is induced under specific condition(s), or is limited to a specific tissue or cell type. These results suggest that the annotated genes or closely related paralogs are transcriptionally active except for *56B23\_g4*, which is predicted to be nonfunctional. This pseudogene contains the remnant of one exon partially similar to an internal part of an Arabidopsis gene (84% of amino acid identity of CAA19721, between positions 90–134 of 211). *56B23\_g5* spans over 16.9 kb of sequence. The unusually large size is mostly due to the size of the first intron (12.6 kb), which was caused by the insertion of three transposable elements (Fig. 2).

The predicted genes were mainly distributed in three locations within BAC 56B23. The proximal part of the BAC contained the partial *56B23\_g1* gene. The middle part included four genes (*56B23\_g2* to *g5*) and covered a region from 35 kb to 88.5 kb with a gene density of 13.3 kb/gene. The distal part grouped five genes (*56B23\_g6* to *g10*) in a sequence of approximately 32 kb and here the gene density reaches up to 6.4 kb/gene. These results show that the genes are nonuniformly distributed along this tomato BAC clone.

In addition to predicted genes, a high density of repetitive DNA was observed. In total, 55 TEs were identified, accounting for approximately 93 kb of sequence and representing 45.5% of the whole BAC sequence (Table II; Supplemental Table I). The majority of all identified TE sequences fell into the class I retrotransposon group. These elements represented approximately 28% of the whole BAC sequence and 59.9% of all the TE sequences identified in this analysis. The second most important family was the unclassified elements, representing 20.7% of all TE sequences. The class II group was the smallest with 21 elements divided into three types (transposons, foldbacks, and miniature inverted-repeat transposable elements [MITEs]). They contributed 9% (seven elements accounting for 8.4 kb), 9.6% (10 elements for 8.9 kb), and 0.8% (four elements for 0.7 kb) of all the annotated TE sequences, respectively (Table II).

The 18 identified retrotransposons can be further subdivided into LTR and non-LTR retrotransposons. LTR retrotransposons were classified into *Copia* and *Gypsy* types according to the specific position of integrase and transcriptase-RNaseH genes (Kumar and Bennetzen, 1999). Here, five members of each *Copia* and *Gypsy* type were found. Only three LTR retrotransposons, *Silvia* (*Copia*), *Caterina-1*, and *Caterina-2* (*Gypsy*), appeared to be complete elements (Fig. 2).



**Figure 2.** Physical map of the 198 kb sequence of BAC clone 56B23. Black and gray boxes represent exons and introns of predicted genes and arrowheads indicate the transcriptional orientation of genes. The *FER* gene is indicated in red. (p) and asterisk indicate partial gene and pseudogene, respectively. Colored boxes represent the different types of TEs according to data included in Supplemental Table I. Horizontal red lines indicate the local duplication in the upstream region of the *FER* gene and horizontal gray lines indicate stretches of LTR retrotransposons. Numbers in brackets indicate the estimated date of LTR retrotransposon insertions according to Table III.

Despite the identification of sequences similar to known plant retrotransposon genes, complete elements were all found to be defective. Accumulation of mutations was responsible for stop codons observed in reading frames as well as for numerous frame shift mutations. The other retrotransposons were found to be partial (conservation of one end of the element only), fragmented (conservation of internal sequences), or degenerated (conservation of small parts of the element). Three putative terminal repeat in miniature (TRIM) elements (Witte et al., 2001) and three putative SoloLTRs were also predicted.

A classification of retrotransposons into subfamilies was carried out using the nucleotide conservation of LTRs and internal parts of the elements (Supplemental Table I). With the exception of the three *Caterina* elements, each LTR retrotransposon belonged to an

independent subfamily. The *Toto1-like* TRIM element was found highly homologous to a complete and autonomous retrotransposon identified earlier (*Toto1*; AF220602) in genomic sequences from currant tomato (*Lycopersicon pimpinellifolium*). This result suggests that TRIM elements might be generated from complete elements of the same family by unequal recombination events and might coexist as partial and complete elements in the same species. The *Tina* element, a degenerated and partial long interspersed nuclear element (LINE) was found inserted in the first intron of the gene *56B23\_g5*. The *Tina* element itself was disrupted by the insertion of the foldback element *Ajax* (Fig. 2). Altogether, our data indicate a high diversity of retrotransposon families in tomato.

The 18 class I retrotransposons described in this study were mainly clustered within three intergenic

**Table II.** Summary of identified TEs in the tomato BAC clone 56B23

Class	Repeat Type	Subgroup	No.	Cumulated Length
I	LTR retrotransposons	<i>Copia-like type</i>	5	55.9 kb
		<i>Gypsy-like type</i>	5	
		TRIM	3	
		Putative SoloLTR	4	
		LINE	1	
II	Transposons	<i>POGO-like</i>	1	8.4 kb
		<i>Sol3</i>	2	
		<i>Tam3</i>	2	
		<i>Alien-like</i>	2	
		Foldbacks	10	
Other	Unclassified	MITEs	4	0.7 kb
			16	19.3 kb
Total			55	93.2 kb

regions. The proximal part of the BAC (between *56B23\_g1* and *g2*), the middle part (between *56B23\_g5* and *g6*), and the very distal part (downstream *56B23\_g10*) each contained five LTR retrotransposons within a distance of 24.5 kb, five LTR retrotransposons within a distance of 42.7 kb, and three LTR retrotransposons within a distance of 4 kb (Fig. 2). The remaining five retrotransposons were dispersed throughout the BAC clone. These results indicate that LTR retrotransposons are not uniformly distributed but clustered in long stretches of sequences, participating in the increase of the intergenic distances. The divergence of the two LTRs of the same element was used to estimate the relative age of LTR retrotransposon insertion (SanMiguel et al., 1998). Of the 17 LTR retrotransposons identified, only six had intact LTRs at both ends. The nucleotide identity of LTRs in these complete elements ranged from 89.2% to 98.9% for *Monica* and *Silvia*, respectively. Two different groups of insertion times were found. While five elements showed ancient insertion times from 5 to 8 million years ago (MYA), only one retrotransposon *Silvia* was estimated to have integrated recently (0.74 MYA; Table III). The full-length element *Silvia* is nested in the degenerated *Copia* LTR retrotransposon *Anouck*, demonstrating that the insertion of *Silvia* was a more recent event than the *Anouck* insertion. Despite the fact that the insertion time of degenerated and truncated elements cannot be estimated, we conclude that successive insertion of LTR retrotransposons has shaped this region of tomato chromosome 6.

The *56B23* BAC clone contains 21 elements related to class II TEs, subclassified into transposons, foldback, and MITE repeat types. Transposons were identified by sequence similarities with known plant transposon nucleotide sequences and transposase proteins. In addition to sequence similarity, putative foldbacks and MITEs were identified on the basis of the absence of coding capacities, the presence of well conserved terminal inverted repeat (TIR) structures at both ends of putative TEs, the presence of target site duplications flanking the TIRs, and/or the presence of multiple copies in public sequence databases. According to

their length, these TEs were subclassified into foldback (>300 bp) and MITE (<300 bp) types.

Seven transposons were classified into the four subgroups *Sol3* (Oosumi and Belknap, 1997), *POGO-like* (Feschotte and Mouches, 2000), *Alien-like* (Pozueta-Romero et al., 1996), and *Tam3-like* (Rubin et al., 2001), ranging in size from 440 to 2,152 bp (Supplemental Table I). Two of the seven transposons identified (*Thomas-1* and *Thomas-2*) were considered as complete elements. Although their TIR and target site duplications could not be clearly identified, these two transposons carried transposase-like genes. Five of the seven identified transposons were located at less than 1.4 kb of the 5' or the 3' of genes (Fig. 2). Interestingly, the *Sol3* nonautonomous transposon *Sol3-1* was found in the promoter region of *56B23\_g2* at a distance of 0.3 kb from the predicted transcription initiation site. These results support the observations of a close association between class II TEs and coding regions (Wessler, 1998). Ten different types of foldback elements were classified. All of the foldback elements are short, ranging in size from 0.3 kb for *Trojan* to 1.5 kb for *Kalypso*. A notable accumulation of four elements (*Cyclop*, *Kalypso*, *Ulysses*, and *Penelope*) was observed within an interval of 8.3 kb in the upstream region of the *FER* gene (*56B23\_g3*, Fig. 2). Four MITEs, ranging in size from 44 to 273 bp, were also identified.

The tomato BAC clone also contains 16 putative TEs showing neither coding capacities nor common structural features of TEs classified previously. The unclassified elements have been identified by the presence of multiple complete or partial copies in available Solanaceae genomic sequences (Supplemental Table I). In some cases, the observation of the disruption of known nucleotide structures led to the identification of unclassified elements (*XB*, Fig. 2).

## DNA Rearrangements

We have found evidence of DNA rearrangements linked to TEs in the tomato sequence studied. Of the 17 LTR retrotransposons identified, only six (including TRIM elements) are intact elements. The presence of truncated elements and soloLTRs in the remaining LTR retrotransposons indicates multiple DNA rearrangements such as unequal and illegitimate recombination similar to those observed in Arabidopsis, rice, maize (*Zea mays*), and Triticeae genomes (Devos et al., 2002; Ma et al., 2004).

DNA rearrangements were also observed outside TEs. The inactive *56B23\_g4* gene showed deletions in coding regions in both 5' and 3' sequences and a duplication of a segment of approximately 2 kb occurred in the 5' region of the *FER* gene (*56B23\_g3*, Fig. 2). The duplicated segments, called DR1 and DR2, were in direct orientation and separated by 2.3 kb. While DR1 showed a size of approximately 2 kb, DR2 covered a distance of 7 kb, representing a size expansion of 3.5-fold. Nucleotide alignments between DR1 and DR2 revealed that the DR2 region was fragmented

**Table III.** Estimated times of retrotransposon insertion

LTRs	% Identity <sup>a</sup>	K (±SE) <sup>b</sup>	Time MYA (±SE) <sup>c</sup>
<i>Monica</i>	89.256	0.1164 (0.0332)	8.36 (2.38)
<i>Toto1-like</i>	90.972	0.0968 (0.0159)	6.95 (1.14)
<i>Caterina-2</i>	91.044	0.0956 (0.0064)	6.86 (0.45)
<i>Caterina-1</i>	91.943	0.0854 (0.0068)	6.13 (0.48)
<i>Lucia</i>	92.417	0.0806 (0.0146)	5.81 (1.04)
<i>Silvia</i>	98.968	0.0104 (0.0037)	0.74 (0.26)

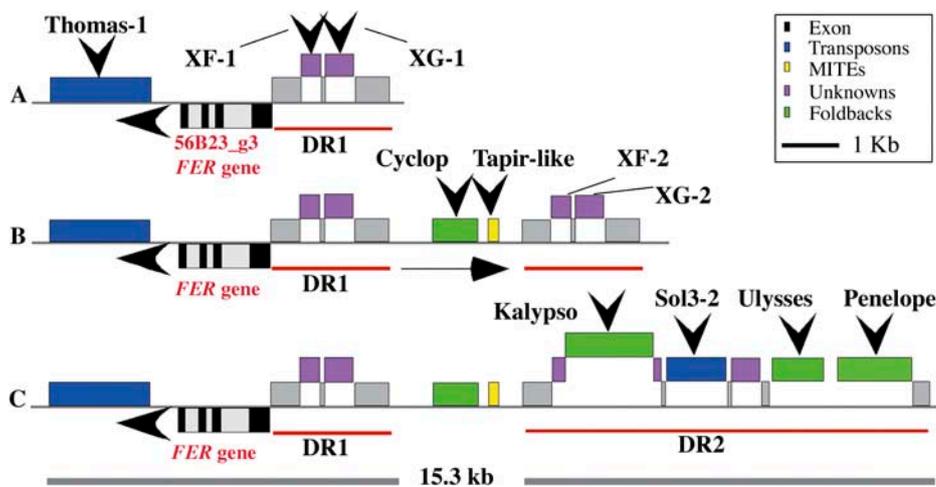
<sup>a</sup>Percent identity calculated with GAP (GCG). <sup>b</sup>K estimated number of substitutions per nucleotide site and its SE, calculated with MEGA v2 (Kumar et al., 2001) and based on the K2P model. <sup>c</sup>Time of divergence, calculated on the average substitution rate (*r*) of  $6.96 \times 10^{-9}$  substitutions per synonymous site per year for tomato (Moniz de Sa and Drouin, 1996).

into four well-conserved segments (88.6% to 93.1% of nucleotide identity). DR2 segments were separated by the insertion of transposable elements that were not present in the DR1 counterpart. Comparative analysis of the two duplicated segments allowed us to make a model of genome evolution in the *FER* gene region (Fig. 3). The ancestor locus of the *FER* gene has undergone at least two insertions before the duplication. Two putative unclassified elements (*XF-1* and *XG-1*) were nested within the ancestor of the DR1 segment (Fig. 3A). *Thomas-1* was also inserted at a distance of 0.5 kb of the 3' of the *FER* gene. In the second step, the duplication of DR1 generated the DR2 segment (Fig. 3B). DR2 was then disrupted by four distinct TEs *Sol3-2* (*Sol3* transposon), *Kalypso*, *Ulysses*, and *Penelope* (foldbacks), increasing the size of the segment to 7 kb. Altogether, the sequence analysis revealed the presence of 11 complete, partial, or disrupted elements within a distance of 15.3 kb surrounding the *FER* gene.

**Microcolinearity between the Tomato BAC 56B23 and the Arabidopsis Genome**

Nine genes and one pseudogene annotated in the tomato BAC 56B23 were used as queries for BLASTP searches against a local database composed of all putative annotated proteins from Arabidopsis. The number of Arabidopsis homologs ranged from one for *56B23\_g1* to more than 250 for *56B23\_g2* (Supplemental Table II). The genomic location of homologs led to the identification of three Arabidopsis BAC clones from chromosomes 2, 3, and 5. None of them were located close to Arabidopsis centromeres. Each carried several coding regions that are homologous to anno-

tated genes present on tomato BAC 56B23. Homologs of *56B23\_g9* and *g10* were found conserved in the same orientation between tomato and the Arabidopsis chromosome 2. In tomato, the two genes were separated by a distance of 5.16 kb on 56B23, whereas the Arabidopsis homologs At2g29960 and At2g29950 were separated by a distance of 2.78 kb (Supplemental Table II). This represented a limited intergenic expansion in tomato by a factor of 1.8. Five tomato genes (*56B23\_g1*, *g2*, *g7*, *g8*, and *g9*) separated by a distance of approximately 187 kb on 56B23 showed homology to four Arabidopsis genes (At3g55960, At3g55950, At3g55940, and At3g55920) in a region of approximately 19 kb on the chromosome 3 (Supplemental Tables II and III). Comparison of distances indicates an expansion by a factor of approximately 10 in tomato. Although the gene order was strictly conserved, the transcriptional orientation of the tomato gene *56B23\_g2* was reversed compared to its Arabidopsis homolog (At3g55950). Moreover, a gene duplication has occurred in tomato (*56B23\_g7* and *g8*, phosphoinositide-phospholipase), whereas only one such gene was present in the Arabidopsis At3g55940 locus. Three tomato genes (*56B23\_g7*, *g8*, and *g9*) separated by a distance of approximately 25 kb showed homology to four Arabidopsis genes (At5g58700, At5g58690, At5g58670, and At5g58710) in a region of 15 kb on the chromosome 5 (Supplemental Tables II and III). This represents an expansion of distance by a factor of 1.6 in tomato. Here, three Arabidopsis genes were found similar to the tomato duplicated genes *56B23\_g7* and *g8*. The remaining tomato genes that did not show any homologs located in the three regions on Arabidopsis chromosomes 2 (At2g29960-At2g29950), 3 (At3g55960-At3g55920) and 5 (At5g58670-At5g58710) have



**Figure 3.** Model for the evolution of the upstream region of the *FER* gene. A, Step 1, two unclassified elements *XF-1* and *XG-1* inserted within the promoter region of the *FER* gene. B, Step 2, a fragment of 2 kb located in the promoter region of the *FER* gene was duplicated in direct orientation, creating a paralogous segment carrying copies of the *XF-1* and *XG-1* unclassified transposable elements (*XF-2* and *XG-2*, respectively). Two additional transposable elements (*Cyclop* and *Tapir-like*) have been inserted between the two paralogous segments. C, Step 3, four TEs (*Kalypso*, *Sol3-2*, *Ulysses*, and *Penelope*) inserted within the duplicated segment *DR2* increasing the size of the fragment from 2 kb to 7 kb.

homologs located elsewhere in the Arabidopsis genome (Supplemental Table II). However, no additional colinearity was found between Arabidopsis and the tomato BAC 56B23. Our results indicate conservation of microcolinearity between the distal and proximal part of the tomato BAC clone and several Arabidopsis genomic segments located on different chromosomes. The central part of the tomato BAC, encompassing four predicted genes (*56B23\_g3* to *g6*) of which belongs the *FER* gene (*56B23\_g3*), was not found colinear with any Arabidopsis segments.

To investigate the mechanisms of genome evolution at the origin of the mosaic conservation of the colinearity, the three colinear Arabidopsis segments were subjected to a detailed analysis. Two of the three Arabidopsis regions that are colinear with tomato fall into the “recent” duplication of the Arabidopsis genome that has occurred before the Arabidopsis/*Brassica rapa* split, 24 to 40 MYA (Blanc et al., 2003). Whereas the chromosome 2 At2g29960-At2g29950 region was not found duplicated, the chromosome 5 (At5g58670-At5g58710) and 3 (At3g55960-At3g55920) regions were found duplicated in chromosomes 3 and 2, respectively (Supplemental Fig. 4). To characterize the duplicated segments, 100 predicted genes of Arabidopsis were extracted from duplicated blocks surrounding At5g58670-At5g58710 and At3g55960-At3g55920 regions. Genes were translated into proteins and were compared with themselves using BLASTP algorithm. Results confirmed the segmental duplications between the chromosomes 2 and 3, surrounding the At3g55960-At3g55920 and between the chromosomes 3 and 5 surrounding At5g58670-At5g58710 (Supplemental Fig. 4). Only two genes located on the paralogous segment of around At5g58670-At5g58710 (At3g47220 and At3g47290, Supplemental Table II) were found similar to genes in the tomato BAC 56B23 (*56B23\_g7* and *g8*). These results indicate that few colinear genes between tomato and Arabidopsis At5g58670-At5g58710 and At3g55960-At3g55920 regions were found conserved on their paralogous segments through the large duplication as observed at the ovate locus on the tomato chromosome 2 (Ku et al., 2000). These observations suggest an extensive gene loss in the duplicated segments of Arabidopsis, which is mainly responsible for the mosaic conservation of colinear genes observed between tomato and Arabidopsis.

## DISCUSSION

### Low Gene Density and High TE Content Characterize the *FER* Gene Region

The tomato BAC clone 56B23 carrying the *FER* gene genetically mapped about 10 cM from the centromere of chromosome 6 (Ling et al., 1996, 2002). Cytologically, it is located in the distal part of the pericentromeric heterochromatin of the chromosome 6 long arm. Sequence analysis of the 198 kb of the 56B23 BAC

identified nine genes and one pseudogene, giving a gene density of 19.8 kb/gene. This is low compared to previous genomic sequencing studies in tomato, in which the gene density ranged from 5 to 17 kb/gene (Rossberg et al., 2001; Van der Hoeven et al., 2002). Compared to Arabidopsis, it is similar to the gene density (20 kb/gene) in the centromeric region of chromosome 2 and lower than the gene density (11–14 kb/gene) in the pericentromeric region (Copenhaver et al., 1999). In contrast to low gene density, a high density of TEs was identified, with TEs accounting for 45.5% of the whole sequence of the BAC clone. This is much higher than estimates based on the complete sequences of gene-containing BAC clones and the random sequence of BAC ends (Budiman et al., 2000). Similarly, TEs were found significantly more abundant in pericentromeric heterochromatin regions than on chromosome arms (Copenhaver et al., 1999). Because BAC 56B23 was located by FISH mapping in the distal part of the pericentromeric region, we suggest that this segment might be a transition region physically separating the gene-rich euchromatin from the centromeric heterochromatin regions of the chromosome 6.

Most of the identified genes of the *FER* locus were shown to be active or potentially active, despite a heterochromatic environment. The presence of such heterochromatic genes has been shown in several organisms including *Drosophila melanogaster*, Arabidopsis, and rice. In *D. melanogaster* and Arabidopsis heterochromatin regions, 453 annotated genes and 200 expressed genes have been identified, respectively (Copenhaver et al., 1999; Hoskins et al., 2002). In *D. melanogaster*, where heterochromatic genes have been initially studied, it appears that a heterochromatin environment is necessary for the expression of heterochromatic genes (Weiler and Wakimoto, 1995). It is not known if such gene adaptation occurred for heterochromatic genes of the tomato genome.

The structure and composition of transcriptionally active heterochromatin regions remain poorly understood in eukaryotes and particularly in plants. In tomato, previous analyses identified several valuable genes in the pericentromeric regions of chromosomes 6, 9, and 12 (van Daelen et al., 1993; Motoyoshi et al., 1996; Budiman et al., 2004). The root-knot nematode resistance locus *Mi* resides in the pericentromeric region of the chromosome 6 short arm (van Daelen et al., 1993). A short genomic sequence of 52 kb spanning the locus was isolated and sequenced (Milligan et al., 1998) and five open reading frames were discovered. Three of them are conserved resistance genes organized in a cluster. A survey of the TE content revealed the presence of several transposons and a degenerated retrotransposon (Milligan et al., 1998; R. Guyot, X. Cheng, B. Keller, and H.-Q. Ling, unpublished data). Unfortunately, the small size of the sequenced locus and the absence of any data on the physical location on pachytene chromosome limit the comparison with the *FER* locus.

The organization of the tomato genome in large blocks of heterochromatin, the presence of active genes and transposable elements in the distal part of these regions and an ongoing project of large-scale genome sequencing (the International Solanaceae Genome Initiative, <http://www.sgn.cornell.edu>) indicate that the tomato genome is an attractive model to explore in detail the composition and organization of the pericentromeric heterochromatin regions.

### Genome Organization and Evolution at the *FER* Locus

The detailed characterization of the BAC clone 56B23 revealed a complex organization of genes and TEs. Genes were not uniformly distributed along the BAC, but were clustered in gene islands separated by stretches of retrotransposons. This class of elements accounts for approximately 28% of the BAC sequence. Successive insertion events were evident from the identification of a nested retrotransposon structure. The estimation of the relative LTR retrotransposon insertions suggested that these elements were active from 8 MYA until recently (0.74 MYA). The accumulation of retrotransposons is responsible for the observed distribution of genes and possibly for the increase of intergenic distances between gene islands.

In total, 55 potential TEs were identified. The majority of them were not described before in the tomato genome. TEs were classified into 44 different types demonstrating a high diversity of classes and types. Heterogeneity of retrotransposon families has been described based on PCR experiments using degenerated primers for both *Gypsy* and *Copia* retrotransposon families in the tomato genome. Four active and highly heterogeneous families of *Copia* retrotransposons have been found in the *Lycopersicon chilense* genome (Yanez et al., 1998) and six families of *Gypsy* were reported from *L. esculentum* (Su and Brown, 1997). LTR retrotransposon structure, organization, and evolution have been actively studied in euchromatin of cereals in which they contribute 50% to 80% to the genome size (SanMiguel et al., 1998). Similar patterns of retrotransposon insertions as well as recombination events removing such elements were found in the tomato *FER* region. LTR retrotransposons acting as target site for insertion of additional elements participate to the formation of long stretches of nested elements as observed in gene-rich regions of maize, wheat, and barley (SanMiguel et al., 1998; Vicient et al., 1999; Wicker et al., 2001). Mechanisms of LTR retrotransposon removal were also identified in tomato, similar to mechanisms described in Arabidopsis, rice, and Triticeae genomes (Devos et al., 2002; Wicker et al., 2003; Ma et al., 2004).

Beside retrotransposons, a high variety of Class II elements was also identified in this region. Class II elements represent 12% of the TE number and were found closely associated with genes (Wessler, 1998). *Sol3-1*, a nonautonomous transposon, was found inserted within the promoter of *56B23\_g2*. A promoter location of *Sol3* was previously also observed in the

tomato 1-aminocyclopropane-1-carboxylate synthase (Shiu et al., 1998) and polygalacturonase genes (Oosumi et al., 1995), and a function as regulatory sequences was proposed for these elements (Shiu et al., 1998). The association between *Sol3* and promoter regions as well as more generally association between transposons and genes in Solanaceae will make them useful for future development of targeted tagging tools in these genomes. Unclassified elements account for about 10% of the BAC clone. Probably, these elements are remnants of deletion and recombination explaining why no common structure with known classes of transposable elements was found. Large-scale sequencing of the tomato genome will allow the identification of the full-length elements and then determination of their complete structures.

The dynamic evolution of the tomato genome is based on DNA rearrangements and can be easily studied near the *FER* gene. There, a local duplication of a promoter fragment of 2 kb predated four TE insertions in one of the two duplicated copies. In total, 11 TEs were identified, increasing the local TE density to 1.3 kb/TE. TE insertions and a local duplication made the region a hot spot of genome evolution. Transposable elements were reported to have an important impact in the evolution and expression of genes in large genome size species such as maize and Triticeae. Here, the accumulation of diverse types of TE upstream and downstream the *FER* gene raises the question if they affect *FER* gene activity.

DNA rearrangements and accumulation of TE demonstrate that the pericentromeric transition region around the *FER* gene is an evolutionary dynamic part of the tomato genome. Furthermore, the recent activity of LTR retrotransposons and the presence of active genes indicate that the organization of the *FER* locus is evolutionarily young. Altogether, these observations suggest that the *FER* region is at a dynamic evolutionary transition stage from gene-rich euchromatin to centromeric heterochromatin regions.

### Mosaic Pattern of Microcolinearity Conservation between the Tomato *FER* Gene Transition Region and the Arabidopsis Genome

Our results indicate a complex relationship of microcolinearity between the tomato pericentromeric BAC 56B23 and three distinct regions on Arabidopsis chromosomes 2, 3, and 5. While the colinear region on chromosome 2 was not duplicated in Arabidopsis, the regions on chromosome 3 and 5 represent recently duplicated blocks of the Arabidopsis genome, estimated to have occurred 24 to 40 MYA (Blanc et al., 2003). However, very few colinear genes were found conserved in the Arabidopsis paralogous segments, indicating an extensive process of selective gene loss or gene movement subsequent to the Arabidopsis duplication. Similar observations were described in the tomato chromosome 2 ovate region (Ku et al., 2000) where the Arabidopsis colinear segments were found

duplicated prior to gene loss. Close examination of the tomato colinear region on Arabidopsis chromosomes 2, 3, and 5 revealed intragenomic colinearity. The conservation of gene order among the three segments indicated successive ancestral duplications of the regions, prior to the “recent” duplications. Ancestral large-scale duplications in Arabidopsis have been described in detail and there is good evidence that Arabidopsis has undergone three rounds of genome duplications (Simillion et al., 2002). We made a relative dating of Arabidopsis and tomato segments by phylogenetic analysis using the cyclophilin genes that are the only common genes in all the segments studied (R. Guyot, X. Cheng, B. Keller, and H.-Q. Ling, unpublished data). Our results suggest that the ancestral duplication between At5g58670-At5g58710 and At2g29960-At2g29950 regions occurred after the divergence of tomato and Arabidopsis. This ancestral duplication must have happened between 90 to 150 MYA and the recent duplication in Arabidopsis 24 to 40 MYA (Blanc et al., 2003).

The accumulation of large-scale duplications and subsequent gene loss obviously is of great significance for the conservation of colinearity between distantly related plants species. The central part of the tomato BAC 56B23, comprising four genes, was not found colinear with any Arabidopsis segment. However, homologs of the four genes were found conserved, but dispersed throughout the Arabidopsis genome. This colinearity disruption can be generated by gene movements either in Arabidopsis and/or in tomato. In Arabidopsis, extensive gene loss on duplicated segments generated hidden duplications, leading to an apparent gene movement and generating perturbations in the colinearity between Arabidopsis and other plant species. In tomato, the disruption of colinearity is accompanied by a large variation in size compared to the Arabidopsis chromosome 3 colinear segments. Expansion is due to the presence of noncolinear genes and insertions of numerous transposable elements. Gene movements can also be the result of the dynamic evolution of this part of the genome through the action of transposable elements. Further detailed and extensive analysis of pericentromeric heterochromatin transition regions is needed to provide a more global view on the structure and evolution of this highly interesting part of the tomato genome.

## MATERIALS AND METHODS

### FISH Mapping

Immature tomato (*Lycopersicon esculentum*) flower buds of about 3.0 mm were harvested and fixed in Carnoy's solution (ethanol:glacial acetic = 3:1). Microsporocytes at meiosis were squashed in acetocarmine solution according to Wu (1967). Slides were frozen in liquid nitrogen. After removing the coverslips, they were dehydrated through an ethanol series (70%, 90%, and 100%) prior to use in FISH. The FISH procedure used for chromosomes was performed according to Jiang et al. (1995). BAC DNA of the tomato BAC clone 56B23 was isolated using a standard alkaline extraction and labeled by nick translation with digoxigenin-16-dUTP (Roche Diagnostics, Indianapolis). Probes were detected with fluorescein isothiocyanate-conjugated sheep-anti-

digoxigenin (Roche Diagnostics). Chromosomes were counterstained with 4',6-diamidino-phenylindole in an antifade solution (Vector Laboratories, Berlingame, CA). Chromosomes and FISH signal images were captured under the Olympus BX61 fluorescence microscope in conjunction with a microCCD camera. Greyscale images were captured for each color channel and then merged using the software of Image-Pro Plus.

### RT-PCR Analysis

Total RNA was extracted from roots and shoots of tomato with Trizol reagent (Invitrogen, Carlsbad, CA). After elimination of genomic DNA contamination by treatment with DNAase I (Promega, Madison, WI) at 37°C for half an hour, 3 µg of total RNA was used for RT with an anchored oligo(dT) primer and 200 units of Moloney murine leukemia virus reverse transcriptase (Invitrogen) in a reaction volume of 20 µL according to the manufacturer's instructions. RT-PCR analysis was performed following the protocol described by Li et al. (2004) with the gene-specific primers (5'-GTACCGAG-AACATGTGAAAG-3' and 5-TTCCTCGCTGTCAATGCAG-3' for 56B23\_g1, 5-CCCAAACGACTGTATTTCAG-3' and 5-GTACCAGCAGCCTTCATTGG-3' for 56B23\_g2, 5-GAGATAGTGGAGTAGTTGC-3' and 5-CTCTTCAT-CACCCATCAC-3' for 56B23\_g6). The PCR reaction conditions were 60 s at 94°C, 60 s at 53°C, 60 s at 72°C for 35 cycles and adding 6 min incubation at 72°C for extension. The PCR products were then cloned into pGEM T-easy vector (Promega, Madison, WI) and sequenced.

### BAC Sequencing and Analysis

BAC DNA was isolated using a large-construct kit (Qiagen, Valencia, CA) and the subcloning library was constructed using a TOPO shotgun cloning kit from Invitrogen. After shotgun sequencing, the sequences were assembled using the PRED/PHRAP software (Ewing et al., 1998). A total of 2,639 reads were produced giving an average coverage of 8.4 x. Remaining gaps between sequence contigs were filled by PCR. The sequence was deposited in GenBank under the accession number AY678298.

### Sequence Analysis and Gene Annotation Method

DNA sequences were analyzed using BLAST algorithms (Altschul et al., 1997) against public and local nucleotide databases. Detailed analysis was performed with the GCG Sequence Analysis Software Package version 10.1 (Madison, WI), the EMBOSS package (Rice et al., 2000), and by dot plot (DOTTER; Sonnhammer and Durbin, 1995). Putative genes were determined by a combination of coding region prediction software (GENSCAN, FGENESH, and MZEF with Arabidopsis and/or monocot matrix, available through the RiceGAAS Web site, <http://ricegaas.dna.affrc.go.jp>) and similarity searches against transcript and protein sequence databases. Gene predictions as well as amino acid and nucleotide alignments were manually evaluated to improve automatic predictions. Final annotation was performed with the Artemis tool (Rutherford et al., 2000).

### Annotation and Classification of Transposable Elements

Putative elements were annotated by amino acid and nucleotide similarity searches against local databases of plant TEs using an E-value of  $E < 10^{-10}$ . Ends of putative TEs were carefully investigated using dot-plot alignment of the query sequence against itself or against public Solanaceae genomic sequences. Novel elements were predicted according to the structural features of TEs. Dating experiments were carried out using full-length LTR retrotransposons and TRIM elements. 5' and 3' LTRs were aligned using GAP (GCG) and the Kimura-2 parameter distance (K) was calculated by MEGA-2 (Kumar et al., 2001) and the average substitution rate of  $6.96 \times 10^{-9}$  substitutions per synonymous site per year (Moniz de Sa and Drouin, 1996) was used to estimate the relative age of LTR retrotransposon integration as in (SanMiguel et al., 1998). The putative elements were classified according to their mobility mechanisms and named according to rules established for the Triticeae repeat database (TREP, <http://wheat.pw.usda.gov/ITMI/Repeats/>).

### Colinearity Analysis with the Arabidopsis Genome

BLASTP searches were conducted with the tomato genes as query with a cutoff of  $E < 10^{-10}$ , against a local database composed of Arabidopsis

(*Arabidopsis thaliana*) sequences, downloaded from TAIR (<http://www.arabidopsis.org>). The Arabidopsis duplicated blocks and colinearity between tomato and Arabidopsis segments were displayed and analyzed using the Genome Pixelizer package (Cannon et al., 2003).

Sequence data from this article have been deposited with the EMBL/GenBank data libraries under accession number AY678298.

## ACKNOWLEDGMENTS

We thank two anonymous reviewers and the editor for their valuable comments.

Received December 11, 2004; revised April 12, 2005; accepted April 13, 2005; published July 11, 2005.

## LITERATURE CITED

- Altschul SE, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402
- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* **9**: 208–218
- Barton DW (1950) Pachytene morphology of the tomato chromosome complement. *Am J Bot* **37**: 639–643
- Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. *Genome Res* **13**: 137–144
- Brown SW (1949) The structure and meiotic behavior of the differentiated chromosomes of tomato. *Genetics* **34**: 437–461
- Budiman MA, Chang SB, Lee S, Yang TJ, Zhang HB, de Jong H, Wing RA (2004) Localization of jointless-2 gene in the centromeric region of tomato chromosome 12 based on high resolution genetic and physical mapping. *Theor Appl Genet* **108**: 190–196
- Budiman MA, Mao L, Wood TC, Wing RA (2000) A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. *Genome Res* **10**: 129–136
- Cannon SB, Kozik A, Chan B, Michelmore R, Young ND (2003) DiagenHunter and GenoPix2D: programs for genomic comparisons, large-scale homology discovery and visualization. *Genome Biol* **4**: R68
- Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD, et al (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* **286**: 2468–2474
- Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. *Genome Res* **12**: 1075–1079
- 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
- Feschotte C, Mouches C (2000) Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the Arabidopsis thaliana genome has arisen from a pogo-like DNA transposon. *Mol Biol Evol* **17**: 730–737
- Hoskins RA, Smith CD, Carlson JW, Carvalho AB, Halpern A, Kaminker JS, Kennedy C, Mungall CJ, Sullivan BA, Sutton GG, et al (2002) Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. *Genome Biol* **3**: RESEARCH0085
- Jiang J, Gill BS, Wang GL, Ronald PC, Ward DC (1995) Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. *Proc Natl Acad Sci USA* **92**: 4487–4491
- Ku HM, Vision T, Liu J, Tanksley SD (2000) 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* **97**: 9121–9126
- Kulikova O, Gualtieri G, Geurts R, Kim DJ, Cook D, Huguet T, de Jong JH, Fransz PF, Bisseling T (2001) Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. *Plant J* **27**: 49–58
- Kumar A, Bennetzen JL (1999) Plant retrotransposons. *Annu Rev Genet* **33**: 479–532
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**: 1244–1245
- Li L, Cheng X, Ling H-Q (2004) Isolation and characterization of Fe(III)-chelate reductase gene *LeFRO1* in tomato. *Plant Mol Biol* **54**: 125–136
- Ling H-Q, Bauer P, Berczky Z, Keller B, Ganai M (2002) The tomato fer gene encoding a bHLH protein controls iron uptake responses in roots. *Proc Natl Acad Sci USA* **99**: 13938–13943
- Ling H-Q, Pich A, Scholz G, Ganai MW (1996) Genetic analysis of two tomato mutants affected in the regulation of iron metabolism. *Mol Gen Genet* **252**: 87–92
- Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res* **14**: 860–869
- Mao L, Begum D, Goff SA, Wing RA (2001) Sequence and analysis of the tomato JOINTLESS locus. *Plant Physiol* **126**: 1331–1340
- Messeguer R, Ganai MW, Steffens JC, Tanksley SD (1991) Characterization of the level, target sites and inheritance of cytosine methylation in tomato nuclear DNA. *Plant Mol Biol* **16**: 753–770
- Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P, Williamson VM (1998) The root knot nematode resistance gene Mi from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *Plant Cell* **10**: 1307–1319
- Moniz de Sa M, Drouin G (1996) Phylogeny and substitution rates of angiosperm actin genes. *Mol Biol Evol* **13**: 1198–1212
- Motoyoshi F, Ohmori T, Murata M (1996) Molecular characterization of heterochromatic regions around the Tm-2 locus in chromosome 9 of tomato. *Symp Soc Exp Biol* **50**: 65–70
- Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. *Nat Genet* **6**: 138–145
- Oosumi T, Belknap WR (1997) Characterization of the Sol3 family of nonautonomous transposable elements in tomato and potato. *J Mol Evol* **45**: 137–144
- Oosumi T, Garlick B, Belknap WR (1995) Identification and characterization of putative transposable DNA elements in solanaceous plants and *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* **92**: 8886–8890
- 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
- Peterson DG, Price HJ, Johnston JS, Stack SM (1996) DNA of content of heterochromatin and euchromatin in tomato (*Lycopersicon esculentum*) pachytene chromosomes. *Genome* **39**: 77–82
- Pozueta-Romero J, Houlne G, Schantz R (1996) Nonautonomous inverted repeat Alien transposable elements are associated with genes of both monocotyledonous and dicotyledonous plants. *Gene* **171**: 147–153
- Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276–277
- Rossberg M, Theres K, Acarkan A, Herrero R, Schmitt T, Schumacher K, Schmitz G, Schmidt R (2001) Comparative sequence analysis reveals extensive microcolinearity in the lateral suppressor regions of the tomato, Arabidopsis, and capsella genomes. *Plant Cell* **13**: 979–988
- Rubin E, Lithwick G, Levy AA (2001) Structure and evolution of the hAT transposon superfamily. *Genetics* **158**: 949–957
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–945
- SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. *Nat Genet* **20**: 43–45
- Shiu OY, Oetiker JH, Yip WK, Yang SF (1998) Promoter of LE-ACS7 an early flooding-induced 1-aminocyclopropane-1-carboxylate synthase gene of the tomato is tagged by a Sol3 transposon. *Proc Natl Acad Sci USA* **95**: 10334–10339
- Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y (2002) The hidden duplication past of Arabidopsis thaliana. *Proc Natl Acad Sci USA* **99**: 13627–13632
- Sonnhammer EL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**: GC1–10

- Su PY, Brown TA** (1997) Ty3/gypsy-like retrotransposon sequences in tomato. *Plasmid* **38**: 148–157
- Tanksley SD, Ganai 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
- van Daelen RA, Gerbens F, van Ruissen F, Aarts J, Hontelez J, Zabel P** (1993) Long-range physical maps of two loci (*Aps-1* and *GP79*) flanking the root-knot nematode resistance gene (*Mi*) near the centromere of tomato chromosome 6. *Plant Mol Biol* **23**: 185–192
- Van der Hoeven R, Ronning C, Giovannoni J, Martin G, Tanksley S** (2002) Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *Plant Cell* **14**: 1441–1456
- Vicient CM, Suoniemi A, Anamthawat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH** (1999) Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* **11**: 1769–1784
- Weiler KS, Wakimoto BT** (1995) Heterochromatin and gene expression in *Drosophila*. *Annu Rev Genet* **29**: 577–605
- Wessler SR** (1998) Transposable elements associated with normal plant genes. *Physiol Plant* **103**: 581–586
- Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B** (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*T. monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant J* **26**: 307–316
- Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B** (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat. *Plant Cell* **15**: 1186–1197
- Witte CP, Le QH, Bureau T, Kumar A** (2001) Terminal-repeat retrotransposons in miniature (*TRIM*) are involved in restructuring plant genomes. *Proc Natl Acad Sci USA* **98**: 13778–13783
- Wu HK** (1967) Note on the preparing of pachytene chromosomes by double mordant. *Scientific Agriculture* **15**: 40–44
- Wu J, Yamagata H, Hayashi-Tsugane M, Hijishita S, Fujisawa M, Shibata M, Ito Y, Nakamura M, Sakaguchi M, Yoshihara R, et al** (2004) Composition and structure of the centromeric region of rice chromosome 8. *Plant Cell* **16**: 967–976
- Yanez M, Verdugo I, Rodriguez M, Prat S, Ruiz-Lara S** (1998) Highly heterogeneous families of Ty1/copia retrotransposons in the *Lycopersicon chilense* genome. *Gene* **222**: 223–228
- Zhang Y, Huang Y, Zhang L, Li Y, Lu T, Lu Y, Feng Q, Zhao Q, Cheng Z, Xue Y, et al** (2004) Structural features of the rice chromosome 4 centromere. *Nucleic Acids Res* **32**: 2023–2030