nature genetics

Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication

Bin Cong[1,2], Luz S Barrero[1–3] & Steven D Tanksley[1,2]

Plant domestication represents an accelerated form of evolution, resulting in exaggerated changes in the tissues and organs of greatest interest to humans (for example, seeds, roots and tubers). One of the most extreme cases has been the evolution of tomato fruit. Cultivated tomato plants produce fruit as much as 1,000 times larger than those of their wild progenitors. Quantitative trait mapping studies have shown that a relatively small number of genes were involved in this dramatic transition, and these genes control two processes: cell cycle and organ number determination[1]. The key gene in the first process has been isolated and corresponds to *fw2.2*, a negative regulator of cell division[2,3]. However, until now, nothing was known about the molecular basis of the second process. Here, we show that the second major step in the evolution of extreme fruit size was the result of a regulatory change of a YABBY-like transcription factor (*fasciated*) that controls carpel number during flower and/or fruit development.

Most crops were domesticated from wild plants sometime in the past 10,000 years. These domestication events are notable for two reasons. First, they enabled humans to make a transition from hunter-gatherers to agriculturalists, ultimately resulting in the creation of modern civilizations. Second, although humans first domesticated plants without any knowledge of plant reproduction or genetics, domestication has rendered some of the most extraordinary and rapid evolutionary changes ever recorded. Unraveling the genetic and molecular pathways of domestication is thus important from both an anthropological and a biological perspective. Moreover, the knowledge gained from these enquiries may prove valuable in further genetic improvement of existing crops or in the domestication of new crop plants (for example, for use as biofuels). As a result, there has been a strong push by biologists to decipher the pathways of domestication[1,4,5].

Previous research has shown that one of the earliest steps in the evolution of larger tomato fruit involved a heterochronic regulatory mutation in a cell cycle–control gene, (*fw2.2*)[3]. However, it is also known that changes in *fw2.2* and other cell-cycle related genes cannot account for the extreme fruit size associated with the modern tomato[1,6]. Rather, the development of extreme fruit size has been traced to two QTLs, referred to as *locule-number* (chromosome 2) and *fasciated* (chromosome 11). Both affect organ (carpel) number rather than size, but the latter locus (*fasciated*) has the larger effect[6,7].

The number of carpels in a tomato flower determines the final number of locules (or compartments) in a mature fruit. Almost all wild tomatoes, as well as a number of smaller-fruited tomato cultivars, produce fruit with only 2–4 locules (**Fig. 1a**,**b**). However, most modern cultivars consumed today produce fruit with eight or more locules (**Fig. 1c**). This increase in locule number has been shown to cause as much as a 50% increase in fruit size and is believed

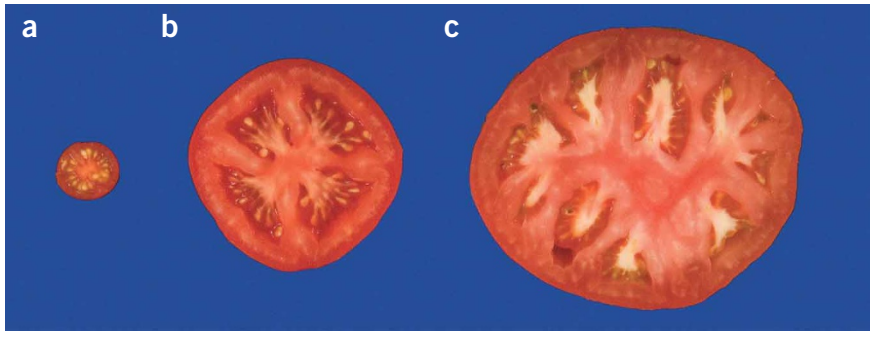

Figure 1 Progression of fruit size increase during tomato domestication. (**a**–**c**) Shown are fruit from a closely related wild species, *S. pimpinellifolium* (**a**), an intermediate-sized processing variety (E6203; **b**) and a large-sized fresh market variety (Jumbo Red; **c**). Progression in fruit size from **a** to **b** occurred mainly through increased cell division, but not changes in locule number. However, progression in fruit size from **b** to **c** was enabled by increases in locule number.

[1]Department of Plant Breeding and Genetics and [2]Department of Plant Biology, Cornell University, Ithaca, New York 14853, USA. [3]Present address: Corporación Colombiana de Investigación Agropecuaria, Bogotá, Colombia. Correspondence should be addressed to S.D.T. (sdt4@cornell.edu).

Received 8 January; accepted 17 March; published online 11 May 2008; doi:10.1038/ng.144

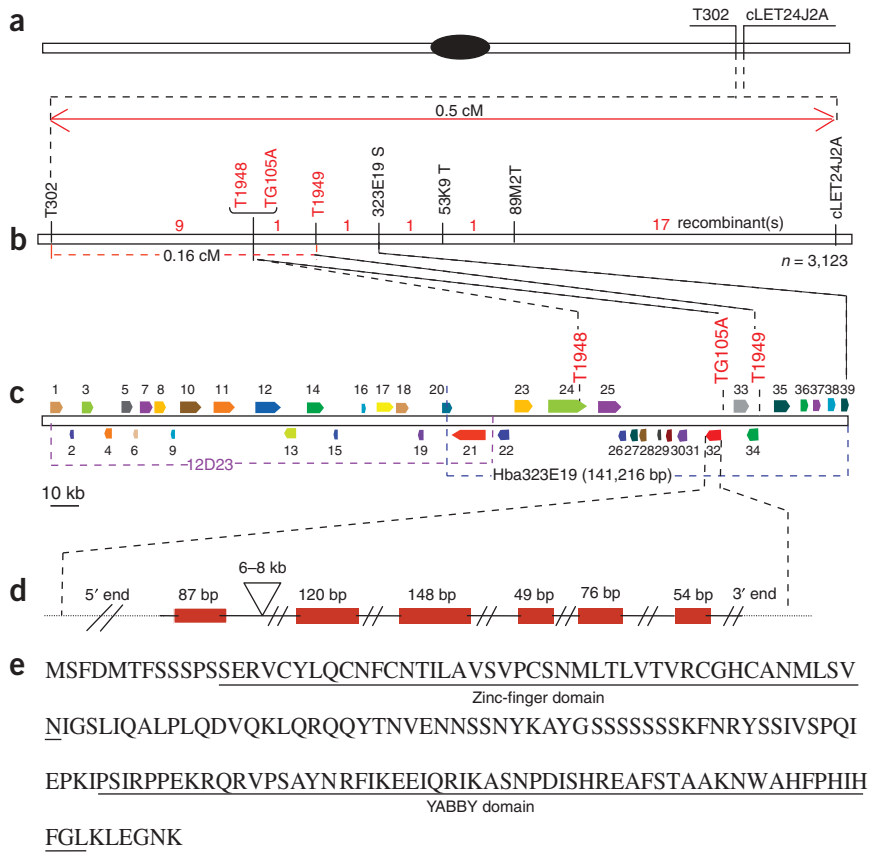

Figure 2 High-resolution mapping of the *fasciated* (*fas*) locus. (**a**) T302-cLET24J2A interval containing *fas* on tomato chromosome 11. (**b**) High-resolution mapping in which *fas* was found to cosegregate with T1948 and TG105A. (**c**) Annotated BAC contig encompassing the *fas* region. Numbers indicate annotated genes (**Supplementary Table 2**). (**d**) Exons and introns of the *fasciated* gene. Reverse triangle indicates 6- to 8-kb insertion found in the first intron of mutant lines. (**e**) Amino acid sequence of *fas* gene with the conserved domains underlined.

to represent a late step in the marked increase in tomato fruit size during domestication[1,6].

To isolate the *fasciated* locus, we used a positional cloning approach. Previous studies had localized *fasciated* to the long arm of chromosome 11 (refs. 6,7; **Fig. 2a**). The tomato cultivar LA2371, which produces an extremely high number of locules, was crossed with a segmental chromosome substitution line (IL11-3 and IL11-4) carrying the low-locule-number *fasciated* allele[7,8]. An F_2 population of 3,123 plants was screened with two PCR-based markers, T302 and cLET24J2A, which flank the 0.5-cM interval known to encompass the *fasciated* locus (**Fig. 2a**). We thus identified 30 plants with crossovers in the interval from which F_3 progeny could be derived (**Fig. 2b**). From each F_3 progeny, we isolated a set of homozygous recombinant and nonrecombinant lines using marker-assisted selection. The carpel and/or locule number was recorded for each plant, and statistical contrasts were made between sets of recombinant and nonrecombinant F_3 progeny in order to deduce the position of the *fasciated* locus on the high-resolution map (**Supplementary Table 1** online). The results indicate that the *fasciated* locus cosegregates with markers T1948 and TG105A (**Fig. 2b**).

We then isolated two overlapping BAC clones encompassing the T1948-TG105A region of chromosome 11. The resulting ~300-kb contig was subjected to annotation and determined to contain 39 putative genes (**Fig. 2c** and **Supplementary Table 2** online). On the basis of functional annotation, we considered two of these genes as possible candidates for the *fasciated* locus. One encodes a YABBY-like transcription factor (**Fig. 2d,e**), and the other encodes a TOUSLED-like protein kinase—both implicated as controllers of organ number from studies in other species[9–11].

To determine whether either of these candidate genes might correspond to the *fasciated* locus, we carried out a set of transformation and complementation experiments—starting with the gene encoding the YABBY-like protein, because initial RT-PCR and sequence comparison between mutants and wild type for TOUSLED-like protein kinase showed no differences (**Supplementary Figs. 1** and **2** online). A genomic region encompassing the coding, upstream and downstream regions of this gene from a low-locule-number genotype (TA492) was cloned into a binary vector. We then used the construct to transform a high-locule-number genotype (LA2371). As the low-locule-number allele is known from prior genetic studies to be dominant to the high-locule-number allele, we expected to observe a reduction in fruit locule number if the YABBY-like gene was indeed the *fasciated* gene. Multiple independent T_0 plants as well as derived T_1 segregants indicated that complementation had occurred, as all transformants produced fruit with significantly fewer locules than nontransformed controls (**Supplementary Table 3** and **Supplementary Fig. 3** online). We thus concluded that the *fasciated* gene encodes a YABBY-like transcription factor (hereafter, we refer to this gene as *fasciated* (*fas*); **Fig. 2e**).

We next turned our attention to determining what genetic change(s) in the *fas* gene might cause an increase in locule number. Sequence comparisons of the low-locule-number allele (from TA209 and TA492) and the high-locule-number allele (from LA2371 and LA0767) showed no differences in the coding region. Thus, the high-locule-number phenotype cannot be caused by a change of function of the FAS protein. However, an examination of noncoding

Table 1 Association of insertion in the first intron of *fas* gene with downregulation of *fas* expression across multiple tomato accessions

	Number of accessions	330 bp upstream	Insertion in first intron	Fruit phenotype	RT-PCR
Cultivated tomatoes (*S. lycopersicum*)	25	7-bp deletion	6- to 8-kb insertion	*fas*	–
	3	No	No	Wild type	+
	9	7-bp insertion	No	Wild type	+
	1	7-bp deletion	No	Wild type	+

See **Supplementary Table 4** for details.

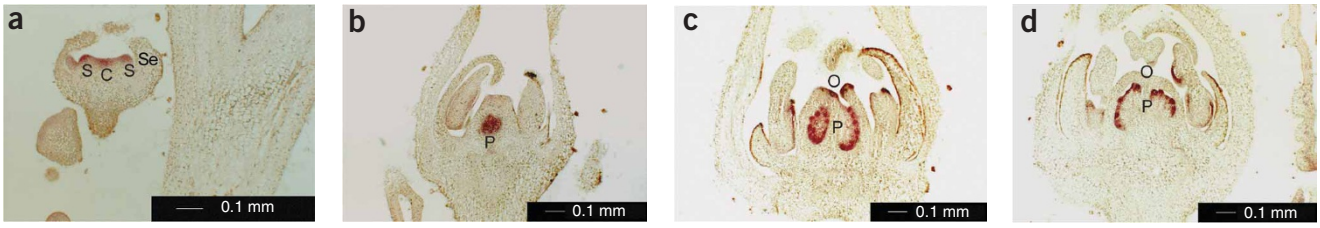

Figure 3 *In situ* hybridization with *fas* gene during flower development. (**a**) Early development before initiation of carpel primordia. (**b**) Mid-stage development shortly after carpel initiation. (**c**,**d**) Late-stage development. All flower buds are from low-locule-number accession TA209. C, carpel primordia; O, ovary; P, placenta; S, stamen primordia; Se, sepal.

regions (including 3.5 kb upstream and 1.5 kb downstream of the *fas* gene) revealed two notable differences. The high-locule-number allele contains a 7-bp deletion 330 bp upstream of the start codon, as well as a 6- to 8-kb insertion in the first intron (**Supplementary Table 4** and **Supplementary Fig. 4** online). To determine whether either of these sequence differences might be causal to the increased-locule-number phenotype, we assayed 25 cultivated tomato accessions known from prior studies to carry the high-locule-number allele and 13 cultivated tomato accessions as well as 6 wild species of tomato known to carry the low-locule-number allele for presence or absence of the 7-bp deletion and intron insertion (**Supplementary Table 4**). All high-locule-number stocks carried the intron insertion, which was absent from all low-locule-number stocks. However, one low-locule-number stock (TA3754) also carried the 7-bp deletion. On the basis of these results, we conclude that the 7-bp deletion is not likely to be the cause for the high-locule-number phenotype. This leaves the large insertion in the first intron as the most probable cause of the high-locule-number phenotype (**Table 1**). However, we cannot rule out the possibility that other, more subtle, sequence changes beyond the 5′ and 3′ regions examined may also contribute to the downregulation of the *fas* gene.

As no changes were observed in the *fas* coding region, the increase in locule number might be due to a change in regulation of the *fas* gene. To test this hypothesis and to better understand the temporal and spatial expression of the *fas* gene, we conducted a set of RNA expression and *in situ* hybridization experiments. The *in situ* studies showed that *fas* RNA is localized to stamen and carpel primordia in early flower development (**Fig. 3a**), and in placental tissue and the tissue surrounding ovules at later stages—results consistent with the *fas* gene exercising control in determining carpel number (**Fig. 3b–d**). In addition, *fas* is also strongly expressed in leaves, but not in roots or seedlings (**Fig. 4a**). A comparison of *fas* expression in a low-locule-number genotype versus a high-locule-number genotype showed that expression is dramatically reduced in developing flower buds in the latter (**Fig. 4a**,**b**). Further, expression analysis, conducted in an F_2 population segregating for the *fas* locus, confirmed that downregulation of *fas* cosegregates with the higher-locule-number allele (**Fig. 4c**). Thus, we conclude that the cause of the high-locule-number phenotype is downregulation of the *fas* gene during flower development and that this downregulation is likely caused by the large insertion in the first intron. We note that *fas* expression is not entirely eliminated in the high-locule-number genotype, as low levels of transcript can still be detected in some stages (**Fig. 4**). It is possible that complete loss of expression would be lethal.

Although the experiments described above establish that a regulatory mutation in the *fas* gene is the cause of the high-locule-number phenotype, they leave unanswered a key question: when and where during domestication did the high-locule-number allele appear? To investigate this question, we surveyed a representative panel of 30 accessions from *Solanum lycopersicum* cv. cerasiforme for sequence variation in and around the *fas* gene (**Supplementary Table 4**). Cerasiforme is a wild form of tomato widely distributed throughout Central and South America and is thought to represent an admixture of the remnant ancestral gene pool from which tomato was domesticated[12]. The results indicate that none of the cerasiforme accessions carry the large insertion in the first intron. However, three accessions (all with low locule number) possessed the 7-bp upstream deletion—further confirming that this mutation is not causal to high locule number. One of the cerasiforme accessions (LA2845) does produce fruit with a higher locule number, but it does not possess the insertion in the first intron. Further sequencing revealed that this accession contains a 1-bp insertion in the 3′ portion of the *fas* coding region which would lead to a truncated protein (118 amino acids versus 177 amino acids) (**Supplementary Fig. 5** online). It seems likely that the shortened FAS protein would have altered or reduced function and hence could lead to the observed increased-locule-number phenotype.

The fact that the large insertion in the first intron was only observed in modern cultivated tomatoes (and not in undomesticated forms) suggests that this mutation occurred relatively recently in tomato domestication. Further, as all high-locule-number varieties surveyed thus far contain an identical intronic insertion, it seems likely that this allele occurred relatively recently and rapidly spread throughout modern tomatoes as a result of selection for extreme fruit size.

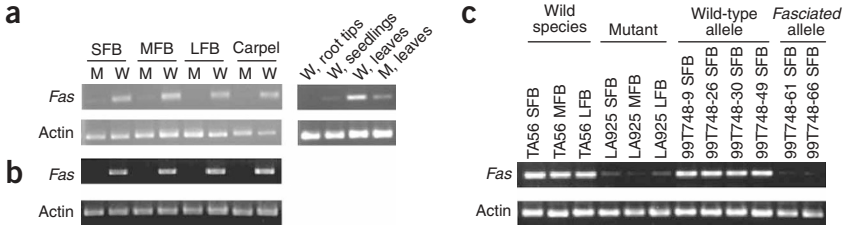

Figure 4 Expression analysis of the *fasciated* gene using RT-PCR. (**a**) W, wild type (low-locule-number, accession TA209); M, mutant (high-locule-number, accession LA2371). (**b**) Repeat experiment using different high-locule-number (LA0767) and low-locule-number (TA496) accessions. (**c**) Co-segregation of high-locule-number genotype (*fas*) and reduced *fas* mRNA expression in the F_2 population derived from a cross between *S. lycopersicum* (LA0925, high locule number) and *S. pennellii* (LA0716, wild species with low locule number). Note that all segregants shown were homozygous for either the *fasciated* allele or the wild-type allele. SFB, small flower buds; MFB, medium flower buds; LFB, large flower buds. Actin (*Tom52*) was used as internal control.

METHODS

Plant material, phenotyping and genotyping. All accessions used in this study, as well as their sources, are listed in **Supplementary Table 4**. All plants were grown in the greenhouse in 2000–2006. We evaluated carpel number at anthesis or mature fruit stage. For high-resolution mapping, we derived an F_2 population of 3,123 plants from a cross between *Solanum lycopersicum* LA2371 and the introgression line IL11-3 or IL11-4, which carries the distal portion of chromosome 11 from the *S. pennellii* genome in an otherwise *S. lycopersicum* background[8]. F_2 plants were screened with markers T302 and cLET24J2 flanking the *fasciated* region. Recombinant F_2 progeny were self pollinated to obtain F_3. We subjected selected F_3 progeny to phenotypic analysis for carpel/locule number, using a minimum of five anthesis ovaries and five fruit mature per plant. Ovaries and fruit were dissected transversally and digitally imaged using a Zeiss dissecting microscope and SCION software (Scion) and a computer scanner and VISTASCAN software (UMAX Technologies) for carpels and fruit, respectively. For genotyping, plant DNA was extracted by the procedure described previously[13] and subjected to marker analyses as summarized in **Supplementary Table 5** online.

BAC isolation and sequencing. A BAC library derived from *S. lycopersicum* cv. Heinz 1706 (TA492) (Clemson University Genomics Institute) was screened with the markers T1948, T1949 and T302. A single positive clone (Hba323E19) was subjected to 10× shotgun sequencing and subsequently assembled into a single 141-kb contiguous molecule. This was further extended by screening a second BAC library (derived from the related species *Solanum bulbocastanum* cv. PT29) using the markers 119J5 COS R, 122H3 SP6 and 262J23 T7 from the left terminus of the BAC. We used this second library, as no positive clones were obtained when screening the previous library with these markers. From this second screen, a single positive BAC clone (12D23) was subjected to 6.7× shotgun sequencing and assembly. These two overlapping BACs thus comprised an approximately 300-kb contig (**Fig. 2c**).

Statistical analyses. We used F_2 recombinants to construct a high-resolution map of the *fasciated* region of chromosome 11 using Mapmaker version 2.0 software and the Kosambi function (**Fig. 2b**)[14,15]. The average carpel and locule number of homozygous recombinants and homozygous nonrecombinant F_3 individuals were compared, from each recombinant F_2 individual, via the unpaired *t*-tests using the statistical program StatView (SAS; **Supplementary Table 1**).

Sequencing of *fasciated* alleles. We used PCR to amplify the genomic region containing the *fasciated* locus, in overlapping fragments, from DNA of various tomato stocks using primer pairs listed in **Supplementary Table 6** online. A full-length cDNA from the *fasciated* gene was synthesized and sequenced from RNA isolated from young flower buds from TA209. First-strand cDNA was reverse transcribed using B_{25}, then amplified by primer FP1 and B_{26} (**Supplementary Table 7** online).

Transgenic complementation. BAC 323E19 DNA was digested by double restriction enzymes *Eco*RI and *Bam*HI, and the resulting products were subject to gel electrophoresis. The fragment containing the *fasciated* gene (low-locule-number allele), plus 5′ (3,302 bp) and 3′ (2,014 bp) regions was gel isolated and cloned into the pCLD04541 Agrobacterium transformation vector[16]. After sequencing for confirmation, we transformed the 11,502-bp insert into multiple plants of the high-locule-number genotype LA2371.

Semi-quantitative RT-PCR. We used the *fasciated* mutants and wild type (**Supplementary Table 4**) for the *fasciated* gene expression studies. Plant materials were grown in the greenhouse in spring 2006. Small (<5 mm), medium (5–10 mm) and large (>10 mm) flower buds, as well as anthesis carpels, from both high-locule-number lines and low-locule-number lines were collected from 3–4 plants with the same genotypes and immediately frozen in liquid nitrogen. Three independent sets of flower buds and carpels (for all stages) were collected and used for independent RT-PCR experiments. In addition, small flower buds from the F_2 population derived from the cross *S. lycopersicum* LA0925 (carrying the high-locule-number *fasciated* allele) × *S. pennellii* LA0716 (carrying the low-locule-number *fasciated* allele) were collected for another set of expression studies.

Frozen flower buds and carpels were ground to a fine powder in liquid nitrogen, and total RNA was isolated using Trizol reagent (Invitrogen). The concentration of total RNA from each sample was determined from 100× diluted solution using spectrophotometry. We treated 1 μg of total RNAs from each sample with RNase-free DNaseI (amplification grade, Invitrogen). First-strand cDNA was synthesized from total RNA (1 μg) by reverse transcription with oligo$(dT)_{17}$ following the manufacturer's protocol (Invitrogen). We used gene-specific primers FP3 and Ex6R for the *fasciated* gene for semiquantitative RT-PCR and actin (TOM52) as a quantitative internal control with forward primer and reverse primer (**Supplementary Table 7**).

RNA *in situ* hybridization. To generate an RNA probe of the *fasciated* gene for *in situ* hybridization, we cloned a fragment amplified from *fasciated* cDNA using primers 3′ UTR F and 3′ UTR R (**Supplementary Table 7**) into the *pCR2.1* cloning vector. The orientation of the cDNA fragment was determined by sequencing. The plasmid was linearized by digestion with *Hin*dIII and served as a template for synthesis of antisense and sense strand RNA probes using T7 RNA polymerase, with the Digoxigenin (DIG) RNA Labeling Kit (Roche). Probe concentrations were estimated by dot blot comparisons with serial dilutions of standards in the DIG-RNA Kit (Roche). Materials used for the *in situ* hybridization study were the same as those grown for the expression study. We harvested at least ten young tomato flower buds, at different developmental stages, from each of ten plants of wild-type TA209 and mutant LA2371. Collected samples were fixed, dehydrated and embedded in paraplast (Sigma); detailed procedures of hybridization and signal detection followed the protocol described previously[17]. After hybridization and washing, DIG-labeled RNA was detected by using antibody to DIG conjugated to alkaline phosphatase (Roche). Alkaline phosphatase activity was measured by using nitro blue tetrazolium (Roche) and bromo-chloro-indolyl phosphate (Roche). Chromogen production was then recorded with a digital camera (Polaroid DMC) mounted on a fluorescence and light microscope (Olympus model AH2-RFL).

Accession codes. GenBank: EU557673 (*S. lycopersicum* cv. Heinz 1706 genomic DNA); EU557674 (*fasciated* cDNA); EU557675 (*LeTousled*); EU557676 (*fasciated-1* allele); EU557677 (LA2371 *fasciated* promoter).

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the Tomato Genetic Resource Center at University of California Davis for providing wild tomato accessions and A. Frary for critical review of the manuscript. We thank Y. Wang, I. Philips and N. Van Eck for technical assistance. Support for this research was provided by a grant from the US Department of Agriculture National Research Initiative to S.D.T.

AUTHOR CONTRIBUTIONS

S.D.T. conceived and oversaw the research. L.S.B. and B.C. contributed to the positional cloning of the *fasciated* gene. B.C. conducted sequence analysis, the RT-PCR analysis, *in situ* hybridization, generation and characterization of the transgenic plants and association studies. L.S.B. generated the mapping populations. B.C. and S.D.T. jointly prepared the manuscript.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Tanksley, S.D. The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell* **16** (Suppl.), S181–S189 (2004).
2. Frary, A. *et al.* fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85–88 (2000).
3. Cong, B., Liu, J. & Tanksley, S.D. Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proc. Natl. Acad. Sci. USA* **99**, 13606–13611 (2002).
4. Doebley, J.F., Gaut, B.S. & Smith, B.D. The molecular genetics of crop domestication. *Cell* **127**, 1309–1321 (2006).
5. Konishi, S. *et al.* An SNP caused loss of seed shattering during rice domestication. *Science* **312**, 1392–1396 (2006).
6. Lippman, Z. & Tanksley, S.D. Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species Lycopersicon

pimpinellifolium and L. esculentum var. Giant Heirloom. *Genetics* **158**, 413–422 (2001).
7. Barrero, L.S. & Tanksley, S.D. Evaluating the genetic basis of multiple-locule fruit in a broad cross section of tomato cultivars. *Theor. Appl. Genet.* **109**, 669–679 (2004).
8. Eshed, Y. & Zamir, D. An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**, 1147–1162 (1995).
9. Jang, S. *et al.* Ectopic expression of OsYAB1 causes extra stamens and carpels in rice. *Plant Mol. Biol.* **56**, 133–143 (2004).
10. Fourquin, C., Vinauger-Douard, M., Fogliani, B., Dumas, C. & Scutt, C.P. Evidence that CRABS CLAW and TOUSLED have conserved their roles in carpel development since the ancestor of the extant angiosperms. *Proc. Natl. Acad. Sci. USA* **102**, 4649–4654 (2005).
11. Roe, J.L., Rivin, C.J., Sessions, R.A., Feldmann, K.A. & Zambryski, P.C. The Tousled gene in A. thaliana encodes a protein kinase homolog that is required for leaf and flower development. *Cell* **75**, 939–950 (1993).
12. Nesbitt, T.C. & Tanksley, S.D. Comparative sequencing in the genus Lycopersicon: implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics* **162**, 365–379 (2002).
13. Fulton, T.M., Chunwongse, J. & Tanksley, S.D. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol. Biol. Rep.* **13**, 207–209 (1995).
14. Lander, E.S. *et al.* MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181 (1987).
15. Kosambi, D.D. The estimation of map distance from recombination values. *Ann. Eugen.* **12**, 172–175 (1944).
16. Jones, J.D. *et al.* Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants. *Transgenic Res.* **1**, 285–297 (1992).
17. Coen, E.S. *et al.* floricaula: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **63**, 1311–1322 (1990).