A cytochrome P450 regulates a domestication trait in cultivated tomato

Manohar Chakrabarti, Na Zhang, Christopher Sauvage, Stephane Munos, José Blanca, Joaquin Canizares, María José Díez, Rhiannon Schneider, Michael Mazourek, Jammi Mcclead, et al.

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

Title: A cytochrome P450 regulates a domestication trait in cultivated tomato

Manohar Chakrabarti[a,1], Na Zhang[a,1,2], Christopher Sauvage[b], Stéphane Muños [c], Jose Blanca[d], Joaquin Cañizares[d], Maria Jose Diez[d], Rhiannon Schneider[a], Michael Mazourek[e], Jammi McClead[a], Mathilde Causse[b] & Esther van der Knaap[a,3]

[a] Department of Horticulture and Crop Science, The Ohio State University/OARDC, Wooster, Ohio, USA. [b] INRA, Unite de Génétique et Amélioration des Fruits et Légumes, Domaine Saint-Maurice, Allée des Chênes, CS 60094, F-84143 Montfavet Cedex, France. [c] Laboratoire des Interactions Plantes Micro-Organismes (LIPM), CNRS-INRA, UMR 2594/441, 31326 Castanet-Tolosan Cedex, France. [d] Institute for the Conservation and Improvement of Agricultural Biodiversity, Polytechnic University of Valencia, Valencia, Spain. [e] Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York, USA.

Key words: Tomato, fruit mass, QTL, *SlKLUH*, domestication

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

Abstract

Domestication of crop plants had profound effects on human lifestyle and agriculture. However, little is known about the underlying molecular mechanisms accompanying the changes in fruit appearance as a consequence of selection by early farmers. We report the fine mapping and cloning of a tomato (*Solanum lycopersicum*) fruit mass gene encoding the ortholog of KLUH, a P450 enzyme of the CYP78A subfamily. The increase in fruit mass is predominantly the result of enlarged pericarp and septum tissues caused by increased cell number in the large fruited lines. *SlKLUH* also modulates plant architecture by regulating number and length of the side shoots, and ripening time, and these effects are particularly strong in plants that transgenically down regulate *SlKLUH* expression carrying fruits of a dramatically reduced mass. Association mapping followed by segregation analyses revealed that a single nucleotide polymorphism in the promoter of the gene is highly associated with fruit mass. This single polymorphism may potentially underlie a regulatory mutation resulting in increased *SlKLUH* expression concomitant with increased fruit mass. Our findings suggest that the allele giving rise to large fruit arose in the early domesticates of tomato and becoming progressively more abundant upon further selections. We also detected association of fruit weight with *CaKLUH* in chile pepper (*Capsicum annuum*) suggesting that selection of the orthologous gene may have occurred independently in a separate domestication event. Altogether, our findings shed new light on the molecular basis of fruit mass, a key domestication trait in tomato and other fruit and vegetable crops.

Significance

Our study offers novel insights into the molecular basis of the regulation of tomato fruit mass, a critical trait in the domestication of fruit and vegetable crops.

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

\body

Plant domestication and artificial selection led to improved agricultural production resulting from dramatic increases in fruit and seed weight (1). At the start of the Neolithic era approximately 10,000 years ago, domestication of animals and plants accompanied the change in lifestyle from hunter-gatherer to a farming routine (2, 3). Driven by the selection of alleles from wild relatives and those that arose after the initial domestication events, characters associated with the domestication syndrome such as larger fruit and seed are typically differentiating wild from cultivated forms (4). Genome-wide genetic diversity analyses support the notion that the initial domestication of tomato was situated in Northern Peru and Ecuador (5). Selections from the red-fruited wild relative *S. pimpinellifolium* L. evolved into the semi-domesticated *S. lycopersicum* L. var *cerasiforme* bearing fruit of small to medium weight. *S. l. cerasiforme* was further domesticated in Mexico giving rise to the large fruited tomato *S. lycopersicum* var *lycopersicum* which to date is cultivated throughout the world (5).

Tomato is an agriculturally important vegetable crop and is used as model for fruit development including ripening and morphological studies (6-8). Genetic studies have identified several QTLs associated with fruit mass in tomato, of which six loci (*fruit weight1.1* [*fw1.1*], *fw2.2, fw2.3, fw3.1/fw3.2, fw4.1 and fw9.1*) are postulated to be major QTLs (9). The only cloned fruit mass gene from vegetable or fruit crops is *FW2.2/Cell Number Regulator (CNR)* (10). The gene encodes a negative regulator of cell division and controls tomato fruit mass as well as organ size in other species, e.g. maize (10, 11) and nitrogen-fixing nodule number (12). *fw3.2* is the second major tomato fruit mass QTL, which explained 19 % of the phenotypic variance in a F_2 population derived from a cross between cultivated tomato accession 'Yellow Stuffer' and wild tomato accession 'LA1589' (13, 14). Despite the importance of fruit mass in the evolution of

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

fruit and vegetable plants, and the numerous genetic loci that underlie the trait (7, 9, 15-19), cloning of domestication genes of fruit and vegetable crops has lagged behind that of the cereal crops. Therefore, insights into the molecular mechanisms that led to the transition of the fruit from small to large remains largely unknown.

Our current study focuses on fine mapping of a tomato fruit mass locus *fw3.2* and cloning of the underlying gene. Association mapping, segregation analysis and transgenic studies led us to identify the putative molecular basis of fruit weight at this locus and a likely regulatory SNP in the promoter of the gene that is highly associated with fruit mass. Phenotypic evaluations demonstrated the cellular basis of increased fruit mass and pleiotropic effects associated with the locus. We also investigated the likely origin of the derived allele and the potential role of this gene in the regulation of fruit mass in other crop species.

Results

***fw3.2* regulates fruit mass, delays ripening and modulates plant architecture, but not total yield per plant.** To elucidate whether fruit mass differences were determined during flower or fruit development, we evaluated the ovary size at anthesis. Evaluations of the nearly isogenic lines (NIL) that differ for the allele at *fw3.2* showed that the ovary perimeter at anthesis was similar while the mature fruit perimeter was significantly different. Further analyses revealed that in particular the pericarp and septum areas were significantly larger in the large fruited NIL *fw3.2(ys)* (Table 1, Fig. S1A - B). The pericarp of the mature fruit carrying the large fruit allele *fw3.2(ys)* showed an increase in cell number whereas cell size remained the same (Table 1, Fig. S1C - D). The time from anthesis to ripe fruit was prolonged in the *fw3.2(ys)* NIL (Table 1). In addition to fruit, *fw3.2(ys)* increased seed mass, although seed number was not significantly

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

different (Table S1). Number of fruits were significantly higher in the lines carrying the small fruited *fw3.2(wt)* allele, while yield per plant was the same (Table 2). Additionally, *fw3.2(wt)* NIL contained more inflorescences and side shoots, including increased side shoot length (Table 2 and Table S1). The number of flowers per inflorescence remained the same. Retaining only 12 fruits per plant led to increased fruit weight in both NILs (Table S1) demonstrating that the increase in fruit mass controlled by *fw3.2* was not due to changes in source-sink relationships. Thus, the results demonstrate that increases in fruit mass coincide with a reduction of number of fruit per plant caused by a reduction in side shoot number and length yielding fewer inflorescences.

Fine mapping delimited the *fw3.2* locus to a 24.4 kb region. Previously, the *fw3.2* locus was fine mapped to a 51.4 kb region comprised of 7 candidate genes (14). An additional recombinant screen delimited the locus to a 24.4 kb region comprised of three candidate proteins: a cytochrome P450 (*ORF6*) belonging to CYP78A subfamily, an ABC transporter (*ORF7*) distantly related to PGP transporters, and a Kelch domain-containing protein of unknown function (*ORF8*). The fine mapping experiment excluded most of the *ORF8* gene to only include the promoter, first exon and first intron until the NDF9 marker (Fig. 1, Table S2). Gene action analysis revealed the additive nature of the *fw3.2* alleles suggesting that neither allele is a null (Table S2).

Association mapping and genetic segregation analysis identified a potential regulatory SNP in the promoter of *ORF6*. To further investigate which of the candidate genes may underlie the *fw3.2* locus, we sought to identify the underlying nucleotide polymorphisms that were correlated

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

with larger fruit. We conducted an association mapping study using a core collection of tomato genotypes demonstrated to have broad genetic diversity within red-fruited tomatoes and comprised of domesticated *S. l. lycopersicum*, the wild relative *S. pimpinellifolium* and semi-domesticated *S. l. cerasiforme* accessions (20). We sequenced approximately 7 kb of the *fw3.2* locus and genotyped 115 markers inside and outside the *fw3.2* region spanning 48.5 kb (Fig. 1 and Dataset S1). Accounting for kinship and population structure, we identified six SNPs that were significantly associated with fruit mass (Fig. 1 and Table S3). With the exception of one SNP, all were located in the upstream region of *ORF6* and *ORF7*. The two most significant SNPs were found in or near *ORF6*, 512 bp upstream of the start of transcription (M9) and 72 bp from the stop codon in the 3′UTR (M42) (Fig. 1). Successful association mapping typically requires rapid linkage disequilibrium (LD) decay. However, low rates of LD decay were found around the *fw3.2* locus, suggesting few natural recombination events in this region that would potentially compromise the interpretation of the association mapping results (21). Therefore, to determine the validity of the association of these SNPs to fruit mass, we performed intraspecific segregation analyses using populations derived from parents that showed few polymorphisms including M9 and three to five additional SNPs (Dataset S1 and Table S4). Genetic analyses using the F_2 and BC_1F_2 populations showed segregation for fruit mass with the allele of M9 in all families in either one or both populations (Table S4). Importantly, family 12S74 segregated for M9 but not for M42 and showed that fruit mass segregated with the allele of M9 (Dataset S1 and Table S4). This finding demonstrated that the M9 SNP was more critical in regulating fruit mass than M42, lending support for the notion that the *ORF6* promoter SNP may regulate fruit mass differences at *fw3.2*.

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

Down regulation of *ORF6* leads to reduced fruit mass. To further investigate the role of the candidate genes in regulating fruit mass, we down-regulated transcript accumulation of *ORF6* and *ORF7*. Since expression of *ORF6* and *ORF7* was lower in the lines that carried small fruit (Fig. 2A), the constructs were transformed into the *fw3.2(ys)* NILs typically yielding large fruit. Down regulation of *ORF6* transcript levels dramatically reduced fruit and seed mass (Fig. 2B - E and Table S5). Similar to the NILs, down regulation of *ORF6* led to a higher number of side shoots and faster ripening (Table S5). The transgenic lines showed additional defects compared to the NILs, e.g. reduced plant height, smaller leaves and leaflets, and severely reduced seed number indicating that extensive down regulation of *ORF6* led to phenotypes affecting the entire plant (Table S5). The RNAi-2 lines targeting the coding region led to less severe phenotypes, including slightly larger fruit with viable seeds. Even though overall seed number was significantly reduced, fruit with 50-100 seeds showed reduced fruit weight compared to the control with the same number of seed (Fig. S1H). Thus, when accounting for seed number, fruit mass is still reduced in lines that down regulate *ORF6*. Contrary to *ORF6*, down regulation of *ORF7* using four amiRNA constructs did not show a correlation between reduction in fruit mass and transcript level (Table S7 and Fig. S1I), and the plants were indistinguishable from the wild type *fw3.2(ys)* NIL. Thus, it is unlikely that *ORF7* underlies the *fw3.2* locus. In light of these findings, we assumed that *ORF6* underlies the *fw3.2* locus and renamed it *SlKLUH* after the founding member of this subfamily of P450s (22). Expression of *SlKLUH* was the highest in vegetative meristems and in the young flower buds (Fig. 2F). Compared to *SlKLUH*, the expression of *SlFW2.2/CNR* was much lower in most tissues examined and differed in tissue specificity. Also, expression of *SlKLUH* was very high in developing seed and not in the growing pericarp, contrary to *SlFW2.2/CNR* (Fig. 2G).

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

Molecular diversity and phylogenetic analysis of *SlKLUH*. To examine the evolutionary history of the *fw3.2* locus, DNA sequence variation of the segments spanning part of the coding region (fragment A) and promoter (fragment B) of *SlKLUH* were investigated. The *SlKLUH* region showed reduced nucleotide diversity (π) in *S. lycopersicum* (0 and 0.000195 for fragment A and B, respectively) compared to *S. pimpinellifolium* (0.002874 and 0.004962) and *S. l. cerasiforme* (0.003203 and 0.004364) (Fig. 3A). The ratio of nucleotide diversity of *S. l. lycopersicum* to *S. pimpinellifolium* at *SlKLUH* was significantly lower compared to the average diversity for chromosome 3 [percentage of $\pi_{S.\ lycopersicum}/\pi_{S.pimpinellifolium}$ for fragment A, B and chromosome 3 are 0, 3.9 and 56.9 %, respectively]. Tajima's D analysis of all accessions combined showed significant values for parts of the two sequenced fragments, ranging from -2.0851 to -1.5187 and from -2.1922 to -1.9117 for fragment A and B, respectively as compared to the average Tajima's D for chromosome 3 (-1.3247) (Fig. 3A). The findings of the molecular diversity analysis were also supported by the results from the phylogenetic analysis using the *SlKLUH* gene sequences. One cluster consisted of all *S. pimpinellifolium* and several *S. l. cerasiforme* accessions and displayed relatively high sequence diversity. The other cluster is comprised of the *S. l. lycopersicum* and the remaining *S. l. cerasiforme* accessions and showed limited genetic diversity (Fig. 3B). The derived allele designated by the M9 SNP was only found in a single subclade with some *S. l. cerasiforme* and all the cultivated tomato accessions, suggesting the mutation arose in the *S. l. cerasiforme* background.

Possible origin of the M9 mutation and identification of fruit mass QTL overlapping with *fw3.2* in chile pepper. To further explore where the derived allele arose, we genotyped the M9

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

SNP in a tomato collection that covers the proposed trajectory of tomato domestication (5). The M9 SNP associated with increased fruit mass was already present at approximately 30% frequency in the Ecuadorian and Northern Peruvian *S. l. cerasiforme* that were closely related to the wild progenitor *S. pimpinellifolium*. The allele frequency increased slightly in the Mesoamerican *S. l. lycopersicum* accessions, and increased further in the *S. l. lycopersicum* landraces from Europe and elsewhere to become practically fixed in modern accessions (Fig. 3C and Table S6). This finding suggested that even though the derived allele arose early, selections for larger fruit weight may not have been as critical or that this allele was not selected early and only became relevant in recent times.

Based on coinciding QTLs, it has been proposed that orthologous genes may have been selected independently in different crops (7, 23). To evaluate whether chile pepper (*C. annuum*) may harbor a fruit mass QTL at a region that is syntenic to tomato *fw3.2*, we determined whether the pepper ortholog *CaKLUH* was associated with fruit mass in this species. A population of 106 F9 RILs derived from a cross between a large-fruited pepper and a small-fruited wild relative *C. frutescens* showed fruit mass segregated significantly with the large fruited allele of *KLUH* from the *C. annuum* parent [dry fruit weight (mean±SD, g) for groups carrying the *C. frutescens* and the *C. annuum* allele of *fw3.2* are 3.21 ± 0.36 and 4.99 ± 0.34, respectively; P value < 0.0005]. These findings demonstrate that *fw3.2* QTL is found in both tomato and chile pepper, and that *KLUH* may regulate fruit mass in two independently domesticated species.

Discussion

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

***SlKLUH* is most likely underlying *fw3.2* and a critical SNP in its promoter is proposed to regulate fruit mass.** *ORF6* encodes a protein with the highest similarity to the Arabidopsis KLUH/CYP78A5 protein controlling vegetative, reproductive organ size and also seed size (22, 24). Other members of this subfamily of cytochrome P450 also control organ and plant size, e.g. CYP78A9, CYP78A6 in Arabidopsis and CYP78A11 in rice (25-27). Two additional members of this subfamily CYP78A27 and CYP78A28 control colony size in moss (28). This highly conserved function of CYP78A subfamily in regulating organ size led us to postulate that *ORF6* was the most likely candidate gene for the *fw3.2* QTL.

The first experiment to test this hypothesis was to identify SNPs that were significantly associated with fruit mass. Our expectations were that these SNPs would be found near the most likely candidate gene. The most significant SNPs identified from the association mapping were the M9 SNP located 512 bp upstream of the transcription start site and the M42 SNP located in the 3′ UTR at 72 bp downstream from the stop codon of *ORF6*. However, LD decay rate was low (21), which was likely due to the lack of outcrossing and recombination events during domestication of tomato. Yet, when the M9 polymorphism arose, its effect on fruit mass could have been of functional significance. This pattern of few SNPs in large LD blocks has also been found in human (29-32) and may be common in plants as well. The low LD yet significant association of the SNP with fruit mass led us to conduct segregation analyses which excluded the importance of M42 leaving M9 as the most significantly associated SNP of potential functional relevance, i.e. to regulate fruit mass. Interestingly, the M9 SNP was located within the third repeat of four 30 bp tandem repeats, with each repeat sharing 83 to 100% homology with its consensus sequence. The tandem repeat was located 440 bp to 559 bp upstream from the *SlKLUH* transcription start site. Moreover, M9 mutation is located within a putative *cis* element,

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

related to the organ-specific element (OSE) found in nodulin and leghemoglobin genes in soybean and *Vicia faba* (33-35). In the future, it will be interesting to assess the role of this putative *cis*-element in gene expression and to explore whether the tandem repeat has any bearing on the modulation of *SlKLUH* expression. Even though the M9 SNP appears most promising to regulate fruit mass, we can't exclude the roles of three other SNPs, M60, M74 and M82 in the regulation of fruit mass while additional SNPs that may be found in the remaining regions that were not sequenced cannot be excluded either.

The second experiment was to transgenically down regulate the expression of the candidate genes, *ORF6* and *ORF7*. Our expectation was that expression levels positively correlate with fruit mass since the *fw3.2(wt)* NIL exhibit smaller fruit and lower transcript levels. Reduced expression of *ORF6*, and not of *ORF7*, had pronounced effects on fruit mass, seed mass, side shoot number and fruit ripening which were traits that were also significantly altered in the NILs. The severity of these phenotypes followed a gradient from *fw3.2(ys)*, *fw3.2(wt)* to the RNAi line. Thus, fine mapping, association mapping and functional segregation analyses of a highly significant SNP in the promoter of *ORF6* combined with the plant transformation results and phenotypic evaluations strongly implied that *ORF6* underlies the *fw3.2* locus. The expression of *ORF6*, which we renamed *SlKLUH*, was particularly high in the developing seeds. However, the effect of increased fruit mass was found in the pericarp of the maternal tissues. Thus, it is plausible that a seed-derived signal might play a role in regulation of fruit mass by *SlKLUH*. A similar hypothesis has been proposed to explain the action of Arabidopsis KLUH (22).

The role of *SlKLUH* in regulating fruit mass, ripening and plant architecture. Fruit development starts with the fertilization of the ovules in the ovary at the time of anthesis (flower

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

opening) leading to the initiation of seed development. Immediately following fertilization, the fruit undergoes a short period of cell division which is followed by a longer period of cell expansion until the final dimensions are reached (36-39). The increase in fruit mass was due to growth processes taking place after fertilization, particularly of the pericarp and septum tissues. These findings showed that the extra cell divisions led to enlarged fruit and a concomitant delay in ripening. Therefore, the delay in ripening was likely the result of an extension of the cell division stage resulting from increased expression. *SlFW2.2/CNR* also controls fruit mass by extending the cell division period (40). Thus, the increases in fruit mass resulting from domestication may operate on similar processes during fruit development yet through different mechanisms based on putative gene function. The lengths of other plant organs were not affected in the NILs, suggesting a specific role of the natural alleles of *SlKLUH* in the regulation of fruit and seed mass. Our results also indicated that plant architecture was affected implying a pleiotropic effect of *SlKLUH* on plant growth and development.

Diversity of *SlKLUH* and origin of M9 mutation. Our findings indicate reduced nucleotide diversity in *S. l. lycopersicum* in the *SlKLUH* region compared to *S. pimpinellifolium*, and an overall reduced diversity compared to the entire chromosome. Parts of two sequenced regions around *SlKLUH* showed significant Tajima's D values, which may suggest an excess of low frequency SNPs in certain regions of the gene in tomato which is potentially the result of a population expansion after a bottleneck or a selective sweep. Taken together, these results support a selective pressure around the *fw3.2* region but this selection is likely to have predated domestication because reduction is found in both *S. pimpinellifolium* and *S. l. lycopersicum*. Exploration of possible origin of M9 mutation using a wide array of tomato accessions showed

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

that the increase in the derived allele frequency followed the same path as the proposed domestication of tomato and was consistent with the selection for a larger fruit along this process (5). However, the allele did not become fixed early during domestication of tomato, further supporting the notion that this region may not have been under strong selection during domestication. Even though the derived M9 allele was detected in low frequency in the Peruvian *S. pimpinellifolium*, without further study we cannot ascertain the validity of this finding and whether the allele may have originated in a wild relative.

Conclusion. To date, only one gene underlying a fruit mass QTL is known (10). Our study identified *SlKLUH* as the second major gene controlling fruit mass in a vegetable or fruit crop, shedding light on the molecular regulation of this trait in plants. Additionally, previous studies had shown an overlap of the *fw2.2/cnr* QTL in chile pepper, albeit that the effect of the pepper QTL was minor (7, 41), whereas the existence of a pepper *fw3.2* QTL was less certain. The findings from this study raise the possibility that *KLUH* may regulate fruit mass in chile pepper and potentially other crops.

Materials and methods

Plant materials and phenotypic evaluation. Plants were grown under field and greenhouse condition at OARDC, Wooster, Ohio in 2010 - 2012, and a subset in Gainesville, Florida in 2011. Most experiments were conducted with 10 plants for *fw3.2(ys)* and *fw3.2(wt)* NILs and multiple independent lines for the *ORF6*_RNAi and *ORF7*_amiRNA constructs. Fruit and seed mass analyses in *fw3.2* NILs, *ORF6*_RNAi and *ORF7*_amiRNA lines were performed with 20 ripe fruits per plant, unless fruit set was impaired. Side shoot number and lengths were measured

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

at 56 and 70 days after sowing, whereas plant height and number of nodes were recorded at 35 and 56 days after sowing. Leaf attributes were measured on 12th and 13th leaves. Floral organ measurements were performed with five flowers per plant. For cell measurement of the ovaries at anthesis, three ovaries per plant and three sections per ovary were analyzed. For cell measurements at mature fruit stage, two fruits per plant and two sections per fruit were analyzed. To test source-sink relationship, a total of 12 fruit plant were kept and fruit weight from these plants was compared with control plants with no fruit removal. All phenotypic evaluations were performed independently at least twice, except the fruit removal experiment which was done once.

Whole fruit, pericarp and cell measurements. Mature fruit perimeter, area of pericarp, septum, columella and placenta were analyzed using Tomato Analyzer (version 2.3) (42). For the cell measurements in the pericarp, thick sections were made with a razor blade, stained with 0.5% Toluidine Blue in 0.1% Sodium Carbonate solution (SPI, Electron Microscopy Supplies, West Chester, PA), photographed with a Leica MZFLIII dissecting microscope coupled to a digital camera (SPOT RT KE, Diagnostic Instruments, Inc. USA), followed by analysis with ImageJ software. For cell measurements of the ovary at anthesis, the tissue was fixed in 0.1M phosphate buffer (pH7.2) containing 3% glutaraldehyde and 2% paraformaldehyde. Samples were dehydrated in 25%-90% ethanol series, followed by infiltration in 50%-100% LR White resin (Electron Microscopy Supplies, Hatfield, PA) gradient at room temperature. Samples were embedded in 100% LR White for 24 hours at 55°C. For each ovary, 5µm transverse sections were stained and photographed and cell measurements were similar as described for the pericarp cell measurements.

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

Fine mapping, gene action and genetic segregation analysis. To identify additional recombinants, 178 seedlings of 09S69-68 line were screened, resulting in recombinant line 10S187-82. The progeny testing of the other line was from a previous study (Zhang et al, 2012). For the genetic segregation analysis of the M9 SNP, we developed five F_2 and BC_1F_2 populations each that were derived from parents from the core collection showing few polymorphisms in addition to M9. Alleles for the other known fruit weight/shape genes that were segregating in the populations were fixed such that only *fw3.2* was segregating. For the progeny testing and segregation analysis, 10-13 plants per genotype [genotypes carrying homozygous *fw3.2(ys)* and homozygous *fw3.2(wt)* alleles] were grown and 20 ripe fruits representing the average per plant were weighed. The average fruit mass of each plant was contrasted between the genotypes and the significance of fruit mass segregation in each family was determined using student's t-test. For gene action analysis, 10 plants each for *fw3.2(ys), fw3.2(wt)* and heterozygous for *fw3.2* were analyzed. Gene action was represented as D/A, where D = Aa – (AA+aa)/2 and A = (AA-aa)/2. Gene action experiments were repeated with the same NIL family grown in two separate fields.

Gene expression analysis. Total RNA was extracted using Trizol® (Invitrogen Inc. USA) as recommended by the manufacturer. Northern blots were performed as described previously (43). For analyzing *ORF7* transcript accumulation, total RNA was isolated from mature leaves, genomic DNA was removed using TURBO DNA-free^{TM} kit (Invitrogen^{TM}), followed by first strand cDNA synthesis and qPCR analysis using iQ^{TM}SYBR Green Supermix (Bio-Rad, USA). For Northern blot and qPCR analyses, the expression of eIF4α and CAC were used as internal control (for primers see Table S7). For RNA seq analysis, strand-specific libraries of approximately 250 bp fragments were prepared using 10 µg total RNA. Four replicates were generated for each tissue type and time point. Single end sequences of 51 bp were generated on

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

an Illumina HiSeq2000 at Weill Medical College (44). Datasets for the digital gene expression and transcriptome analysis of developing fruit using RNA seq can be accessed using NCBI SRA accession number SRA061767, SRA068200 and SRA091611 (45). Analyzed data can also be accessed at the Tomato Functional Genomics Database (http://ted.bti.cornell.edu/cgi-bin/TFGD).

Association of *CaKLUH* with fruit weight in chile pepper. To identify nucleotide polymorphisms, the 5´ upstream, intron, 3´downstream regions of the *CaKLUH* ortholog in pepper were sequenced in *C. annuum* cv. 'NuMex RNaky' and the *C. frutescens* accession BG 2814-6 (46). Genomic DNA was extracted from 106 individuals from a F_9 RIL population derived from a cross between these parents. The intron SNP was developed into a CAPS marker using the primer pair *fw3.2*_pepper_F and *fw3.2*_pepper_R and restriction enzyme digestion with *Hae*III. Ten fruit were dried and weighed from each line.

Development of transgenic RNA interference lines for *ORF6* and *ORF7*. Two RNAi hairpin constructs for *ORF6*, pMC2 corresponding to RNAi-3 transgenic lines and pMC3 corresponding to RNAi-2 transgenic lines were generated. RNAi-2 targets the coding sequence of *SlKLUH*, whereas RNAi-3 targets mainly the 3´UTR (Fig. 2B). Sense and antisense arms were cloned in the *Hind*III-*Xho*I and *Sac*I-*Xba*I sites of a modified pKYLX80 vector, where *Xho*I and *Sac*I sites are separated by a 150 bp ω-3 fatty acid desaturase intron. Expression cassettes between *Hind*III-*Xba*I were subcloned into binary vector pKYLX71 under a 35S promoter with duplicated enhancer (47). Four artificial miRNAs specific for *ORF7* were designed using WMD3- Web MicroRNA Designer (48) and were cloned into pKYLX71 binary vector. Primers for the RNAi and sequences for amiRNAs are shown in Table S7. All constructs were introduced in the *fw3.2(ys)* NIL 08S591-10 carrying an introgression of ~130 kb. Transformations were carried out at the Plant Transformation Core Research Facility at the University of Nebraska-Lincoln.

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

Association mapping. For the association mapping, approximately 7 kb of the *fw3.2* locus was sequenced and additional markers were used to genotype the region in 86 accessions of the core collection (20). Genomic positions and scores of markers used in association analysis are presented in Table S3. Association analysis was done using MLM model of TASSEL2.1 software (49). Q and K matrix were generated with STRUCTURE 2.2 (50) and SPAGeDi (51), respectively. To generate Q and K matrix, 20 EST-SSR markers distributed throughout the genome were used and the data for these markers in the core collection was obtained from a previous study (20).

Molecular diversity, phylogenetic analysis and estimation of M9 allele frequency. For the molecular diversity analysis, fragment A (637bp) and fragment B (507bp) were sequenced in 157 (15 *S. pimpinellifolium*; 20 *S. l. lycopersicum* and 122 *S. l. cerasiforme*) and 172 lines (16 *S. pimpinellifolium*; 24 *S. l. lycopersicum* and 132 *S. l. cerasiforme*), respectively (20). Fragment A and B span genomic positions 58850848 to 58851481 and 58852273 to 58852778, respectively (positions are based on the Tomato WGS Chromosomes SL2.40). The molecular diversity analysis π and Tajima's D were computed using DnaSP5.0 (52) with a sliding window length 100 and step size 25. Nucleotide diversity across chromosome 3 was calculated using 10 accessions each of *S. l. lycopersicum* and *S. pimpinellifolium*.

A contig of 2528 bp comprised of five fragments spanning the 4382 bp region (corresponding to 58848391-58852772 in Tomato WGS chromosome SL2.40ch03) around *SlKLUH* was obtained from the 86 accessions of the association mapping collection (Dataset S1) and used for the phylogenetic analysis. All sequences were aligned using ClustalX2.1 with default multiple sequence alignment parameters and DNA weight matrix ClustalW (1.6). The alignment file was imported to MEGA5.05 (53) and converted into mega (.meg) file format. The phylogeny was

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

reconstructed using the Neighbor-Joining statistical method with 1000 bootstrap replications and Maximum Composite Likelihood model. For computing phylogenetic distance, a cut off value of 50% bootstrap value was used. Frequency of the derived and ancestral alleles of M9 was analyzed in the following sub-populations: *S. p*_Per (N=30), *S. p*_Ecu (N=13), *S. l. c*_Ecu (N=24), *S. l. c*_N. Per (N=13), *S. l. c*_Mes (N=32), *S. l. l*_Mes (N=16), *S. l. l*_other (N=47) and *S. l. l*_Imp (N=21).

Acknowledgements. We thank Dr. D. Choi (Seoul National University) for providing DNA sequence of chile pepper *CaKLUH* region; and Molecular and Cellular Imaging Center, OARDC, The Ohio State University for assistance with microscopy. This work is supported by the National Science Foundation Grant (IOS-0922661) to EVDK.

References

1. Darwin C (1868) *The variation of animals and plants under domestication* (John Murray, Albemarle Street, London).
2. Diamond J (1997) *Guns, Germs, and Steel.* (WW Norton & Co., New York, NY).
3. Smith BD (1995) Seed plant domestication in eastern North America. *Last Hunters-First Farmers* 354:193-213.
4. Pickersgill B (2007) Domestication of plants in the Americas: Insights from mendelian and molecular genetics. *Ann Bot* 100(5):925-940.
5. Blanca J, et al. (2012) Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. *Plos One* 7(10):e48198.
6. Klee HJ & Giovannoni JJ (2011) Genetics and control of tomato fruit ripening and quality attributes. *Annu Rev Genet* 45:41-59.
7. Paran I, van der Knaap E (2007) Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J Exp Bot* 58(14):3841-3852.
8. Tanksley SD (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell* 16 Suppl:S181-189.
9. Grandillo S, Ku HM, Tanksley SD (1999) Identifying loci responsible for natural variation in fruit size and shape in tomato. *Theor Appl Genet* 99(6):978-987.
10. Frary A, et al. (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289(5476):85-88.

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

11. Guo M, et al. (2010) *Cell Number Regulator1* affects plant and organ size in maize: implications for crop yield enhancement and heterosis. *Plant Cell* 22(4):1057-1073.
12. Libault M, et al. (2010) A member of the highly conserved *FWL* (tomato *FW2.2-like*) gene family is essential for soybean nodule organogenesis. *Plant J* 62(5):852-864.
13. van der Knaap E, Tanksley SD (2003) The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in Yellow Stuffer tomato. *Theor Appl Genet* 107(1):139-147.
14. Zhang N, Brewer MT, & van der Knaap E (2012) Fine mapping of *fw3.2* controlling fruit weight in tomato. *Theor Appl Genet* 125(2):273-284.
15. Blas AL, et al. (2012) Genetic mapping of quantitative trait loci controlling fruit size and shape in papaya. *Mol Breeding* 29(2):457-466.
16. Costantini L, Battilana J, Lamaj F, Fanizza G, Grando MS (2008) Berry and phenology-related traits in grapevine (*Vitis vinifera* L.): from quantitative trait loci to underlying genes. *BMC Plant Biol* 8:38.
17. Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002) Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics* 161(4):1713-1726.
18. Eduardo I, et al. (2011) QTL analysis of fruit quality traits in two peach intraspecific populations and importance of maturity date pleiotropic effect. *Tree Genet Genomes* 7(2):323-335.
19. Zhang GR, et al. (2010) Fruit size QTL analysis of an F-1 population derived from a cross between a domesticated sweet cherry cultivar and a wild forest sweet cherry. *Tree Genet Genomes* 6(1):25-36.
20. Ranc N, Munos S, Santoni S, Causse M (2008) A clarified position for *Solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (Solanaceae). *BMC Plant Biol* 8:130.
21. Zhang N (2012) Fine mapping and characterization of *fw3.2*, one of the major QTL controlling fruit size in tomato. *PhD dissertation, The Ohio State University*:1-152.
22. Anastasiou E, et al. (2007) Control of plant organ size by *KLUH/CYP78A5*-dependent intercellular signaling. *Dev Cell* 13(6):843-856.
23. Paterson AH, et al. (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* 269(5231):1714-1718.
24. Adamski NM, Anastasiou E, Eriksson S, O'Neill CM, Lenhard M (2009) Local maternal control of seed size by *KLUH/CYP78A5*-dependent growth signaling. *Proc Natl Acad Sci U S A* 106(47):20115-20120.
25. Fang W, Wang Z, Cui R, Li J, Li Y (2012) Maternal control of seed size by *EOD3/CYP78A6* in *Arabidopsis thaliana*. *Plant J* 70(6):929-939.
26. Ito T, Meyerowitz EM (2000) Overexpression of a gene encoding a cytochrome P450, *CYP78A9*, induces large and seedless fruit in arabidopsis. *Plant Cell* 12(9):1541-1550.
27. Miyoshi K, et al. (2004) *PLASTOCHRON1*, a timekeeper of leaf initiation in rice, encodes cytochrome P450. *Proc Natl Acad Sci U S A* 101(3):875-880.
28. Katsumata T, et al. (2011) Involvement of the CYP78A subfamily of cytochrome P450 monooxygenases in protonema growth and gametophore formation in the moss *Physcomitrella patens*. *Biosci Biotechnol Biochem* 75(2):331-336.
29. Meyer KB, et al. (2011) A functional variant at a prostate cancer predisposition locus at 8q24 is associated with PVT1 expression. *PLoS Genet* 7(7):e1002165.

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

30. Nakki A, et al. (2010) Allelic variants of *IL1R1* gene associate with severe hand osteoarthritis. *BMC Med Genet* 11:50.
31. Nischwitz S, et al. (2011) More CLEC16A gene variants associated with multiple sclerosis. *Acta Neurol Scand* 123(6):400-406.
32. Uno S, et al. (2010) A genome-wide association study identifies genetic variants in the *CDKN2BAS* locus associated with endometriosis in Japanese. *Nat Genet* 42(8):707-710.
33. Sandal NN, Bojsen K, Marcker KA (1987) A small family of nodule specific genes from soybean. *Nucl Acids Res* 15(4):1507-1519.
34. Stougaard J, Jorgensen JE, Christensen T, Kuhle A, Marcker KA (1990) Interdependence and nodule specificity of cis-acting regulatory elements in the soybean leghemoglobin-*Lbc3* and *N23* gene promoters. *Mol Gen Genet* 220(3):353-360.
35. Vieweg MF, et al. (2004) The promoter of the *Vicia faba* L. leghemoglobin gene *VfLb29* is specifically activated in the infected cells of root nodules and in the arbuscule-containing cells of mycorrhizal roots from different legume and nonlegume plants. *Mol Plant-Microbe Inter* 17(1):62-69.
36. Gillaspy G, Ben-David H, Gruissem W (1993) Fruits: a developmental perspective. *Plant Cell* 5(10):1439-1451.
37. Iglesias DJ, et al. (2007) Physiology of citrus fruiting. *Brazil J Plant Physiol* 19(4):333-362.
38. Marcelis L, Hofman-Eijer L (1993) Cell division and expansion in the cucumber fruit. *J Hort Sci* 68(5):665-671.
39. Xiao H, et al. (2009) Integration of tomato reproductive developmental landmarks and expression profiles, and the effect of *SUN* on fruit shape. *BMC Plant Biol* 9:49.
40. Cong B, Liu J, Tanksley SD (2002) Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proc Natl Acad Sci U S A* 99(21):13606-13611.
41. Rao GU, Ben Chaim A, Borovsky Y, Paran I (2003) Mapping of yield-related QTLs in pepper in an interspecific cross of *Capsicum annuum* and *C. frutescens*. *Theor Appl Genet* 106(8):1457-1466.
42. Brewer MT, et al. (2006) Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species. *Plant Physiol* 141(1):15-25.
43. Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319(5869):1527-1530.
44. Zhong S, et al. (2011) High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb Protoc* 2011(8):940-949.
45. Huang ZJ, Van Houten J, Gonzalez G, Xiao H, van der Knaap E (2013) Genome-wide identification, phylogeny and expression analysis of *SUN*, *OFP* and *YABBY* gene family in tomato. *Mol Genet Genomes* 288(3-4):111-129.
46. Ben-Chaim A, et al. (2006) QTL analysis for capsaicinoid content in *Capsicum*. *Theor Appl Genet* 113(8):1481-1490.
47. Schardl CL, et al. (1987) Design and construction of a versatile system for the expression of foreign genes in plants. *Gene* 61(1):1-11.
48. Ossowski S, Schwab R, Weigel D (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J* 53(4):674-690.

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

49. Bradbury PJ, et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633-2635.
50. Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7(4):574-578.
51. Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2(4):618-620.
52. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451-1452.
53. Tamura K, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731-2739.

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

Figure legends

Fig. 1. Genome structure of the *fw3.2* locus and association mapping of the polymorphisms. UTRs, exons, introns and intergenic regions are represented with red boxes, black boxes, black lines and green lines, respectively. The direction of transcription is denoted with red arrows. Blue vertical lines represent polymorphisms not associated with fruit mass; red vertical lines represent polymorphisms significantly associated with fruit mass. Two highly signficantly SNPs are represented with an asterisk (*). Red horizontal lines depict the sequenced regions. Two markers P450-2 and NDF9, which delimited the *fw3.2* locus to 24.4 kb region are shown with green vertical lines.

Fig. 2. Transcript accumulation and effect of *SlKLUH* knock down on fruit and seed mass. A. Expression of *ORF6* (*SlKLUH*) and *ORF7* in *fw3.2(ys)* and *fw3.2(wt)* NILs. B. Regions of *SlKLUH* targeted by two RNAi constructs. C and D. Effect of *SlKLUH* knock down on fruit and seed mass, respectively. E. Transcript accumulation of *SlKLUH* in *fw3.2(ys)* and *fw3.2(wt)* NILs and *SlKLUH* knock down lines. F. Average transcript accumulation of *SlKLUH* and *SlFW2.2/CNR* in different tissues of *S. pimpinellifolium* LA1589. G. Average transcript accumulation of *SlKLUH* and *SlFW2.2/CNR* in *S. l. lycopersicum* pericarp and developing seeds at different fruit development stages. Veg. meristem, vegetative meristem; dpa, days post anthesis; P, pericarp; S, seed.

Fig. 3. Molecular diversity and phylogenetic analysis of *SlKLUH*, and origin of M9 mutation. A. Nucleotide diversity (π) and Tajima's D were calculated in fragment A and B spanning part of the coding region and promoter of *SlKLUH* (Fig.1). The Tajima's D is calculated from the *SlKLUH* sequences of *S. p*, *S. l. c* and *S. l. l* combined. B. Phylogenetic analysis of *SlKLUH*. *S. l.*

Proceedings of the National Academy of Sciences of the United States of America (2013), Vol. 110, N°.42, p. 17125-17130, DOI: 10.1073/pnas.1307313110
Journal homepage: www.pnas.org/cgi/doi/10.1073/pnas.1307313110

l, S. l. c, *S. p* and *S. pennellii* are depicted in red, black, green and blue, respectively. Large (>60 g), medium (>10g and <60g) and small (<10g) fruit-bearing lines are represented as large, medium and no dots after their name, respectively. Accessions carrying the M9 mutation are depicted with an asterisk (*). C. Frequency of derived and ancestral M9 SNP allele in tomato subpopulations. *S. p, S. pimpinellifolium*; *S. l. c, S. l. cerasiforme*; *S. l. l, S .l. lycopersicum*; Per, Peru; Ecu, Ecuador; N.Per, Northern Peru; Mes, Mesoamerica; Imp, Improved accessions.