

REPORT

PLANT SCIENCE

A chemical genetic roadmap to improved tomato flavor

Denise Tieman,^{1,2*} Guangtao Zhu,^{1,3*} Marcio F. R. Resende Jr.,⁴ Tao Lin,^{1,3} Cuong Nguyen,² Dawn Bies,² Jose Luis Rambla,⁵ Kristty Stephanie Ortiz Beltran,⁵ Mark Taylor,² Bo Zhang,² Hiroki Ikeda,² Zhongyuan Liu,² Josef Fisher,⁶ Itay Zemach,⁶ Antonio Monforte,⁵ Dani Zamir,⁶ Antonio Granell,⁵ Matias Kirst,⁷ Sanwen Huang,^{1,3†} Harry Klee^{1,2†}

Modern commercial tomato varieties are substantially less flavorful than heirloom varieties. To understand and ultimately correct this deficiency, we quantified flavor-associated chemicals in 398 modern, heirloom, and wild accessions. A subset of these accessions was evaluated in consumer panels, identifying the chemicals that made the most important contributions to flavor and consumer liking. We found that modern commercial varieties contain significantly lower amounts of many of these important flavor chemicals than older varieties. Whole-genome sequencing and a genome-wide association study permitted identification of genetic loci that affect most of the target flavor chemicals, including sugars, acids, and volatiles. Together, these results provide an understanding of the flavor deficiencies in modern commercial varieties and the information necessary for the recovery of good flavor through molecular breeding.

The tomato is the highest-value fruit and vegetable crop worldwide (1) and an important source of micronutrients in the human diet. Nonetheless, deterioration in flavor quality of the modern commercial tomato relative to heirloom varieties is a major cause of consumer complaint. To address this problem, we performed a comprehensive study of the chemistry and genetics of tomato flavor. The flavor of any food is the sum of interactions between taste and olfaction. For the tomato, sugars and acids activate taste receptors, while a diverse set of volatile compounds activate olfactory receptors (2–4). Volatiles, in particular, are essential for good flavor. For example, the removal of specific carotenoid-derived volatiles results in significant reductions in consumer liking (5). Refrigeration also selectively alters the volatile content of fruit without altering sugars and acids, resulting in reduced consumer liking (6). Flavor phenotyping is expensive, subject to environmental

variation, not amenable to high-throughput assays, and beyond the means of most breeding programs. Therefore, most breeders focus on yield, disease resistance, and firmness, which are essential for shipping, long-term storage, and external appearance rather than flavor quality. Flavor-associated volatiles are present at picomolar to nanomolar concentrations in fruits and are extremely difficult to quantify. Thus, these chemicals have largely been ignored, and an emphasis on production traits has inadvertently led to a decline in flavor quality. To reverse that decline, we must first identify the most important chemical contributions to consumer liking and then understand what has been lost from modern varieties and why. Because volatiles are active at such low concentrations, it should be possible to substantially increase their content in fruit with minimal impact on yield. Here we define the chemistry of consumer preferences for the tomato and identify genetic loci that provide a roadmap to improved flavor quality. To this end, we performed whole-genome sequencing and targeted metabolome quantification of sugars, acids, and volatiles that potentially contribute to flavor in 398 modern, heirloom, and wild accessions of the tomato (*Solanum lycopersicum*) and its closest relative, *S. pimpinellifolium* (table S1). The inclusion of wild accessions and the closest relative of the tomato provide a baseline for fruit chemical composition before human intervention. A total of 160 samples representing 101 different accessions were evaluated by a consumer panel that rated multiple sensory attributes, including overall liking and flavor intensity (table S2) (2). Sixty-six of these accessions were previously evaluated (2) and those data were included in this analysis. We identified 33 chemicals that correlated

with consumer liking and 37 that significantly correlated with flavor intensity (table S3), 28 of which were associated with both overall liking and flavor intensity. These chemicals provided a foundation for a genome-wide association study (GWAS).

Although postharvest handling can negatively affect flavor (6), our results indicate that representative modern cultivars, such as Florida 47 and Flora-Dade, are not well liked even when grown using commercial practices and harvested when fully ripe (table S2). To address the question of why modern tomato cultivars lack the flavor of older accessions (heirlooms), we examined the flavor-associated chemical composition of 48 modern cultivars relative to 236 older *S. lycopersicum* accessions. Discriminant analysis of principal components (DAPC) using single-nucleotide polymorphism (SNP) data identified five classes, including one that contains all of the known modern elite inbreds and hybrids (fig. S1). A total of 13 flavor-associated volatiles were significantly reduced in modern varieties relative to heirloom varieties (fig. S2). Volatile chemicals define the unique flavor of a tomato and are essential for consumer liking (2, 4). Thus, poor flavor of modern varieties can largely be attributed to the dilution of many flavor volatiles that positively influence liking. This dilution of flavor chemicals should be correctable by reintroducing superior alleles of genes controlling their synthesis.

To identify superior alleles of genes controlling accumulation of flavor chemicals in tomato fruit, we performed a GWAS on modern and heirloom varieties, as well as wild *S. lycopersicum* variant (var) *cerasiforme* and the most closely related species, *S. pimpinellifolium*. A total of 27 volatiles, total soluble solids, glucose, fructose, citric acid, and malic acid were quantified in ripe fruits (fig. S3). Genome sequencing of the 398 chemically profiled accessions grown in Florida identified a total of 2,014,488 common SNPs [minor allele frequency (MAF) > 5% and missing rate < 10%]. A total of 251 association signals ($P < 1.24 \times 10^{-7}$) were detected for 20 traits, including 4 nonvolatile and 15 volatile flavor chemicals (figs. S4 to S22 and tables S4 and S5). We also evaluated an F_2 population with 197 individuals derived from a cross between the small-fruited, high-flavor Maglia Rosa Cherry line and the large-fruited, modern inbred line FLA 8059. Each plant was subjected to genotyping by sequencing, and fruits from each plant were chemically profiled. Loci associated with significant effects [logarithm of the odds ratio for linkage (LOD) > 3] on 4 nonvolatile and 23 volatile chemicals, as well as fruit weight, were identified (fig. S23 and table S6). Finally, a second, overlapping population was grown in Israel. This population consisted of 352 varieties, 262 of which overlapped with the set of 398 profiled accessions (figs. S14, S16, and S17 and table S7). Significant common associations included geranylacetone (chromosome 3), 6-methyl-5-hepten-2-one (MHO, chromosomes 3 and 9), and guaiacol (chromosomes 0 and 9).

Analysis of loci impacting sugar content provides a cautionary tale regarding crop domestication and improvement. We identified two loci

¹Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, No. 7, Pengfei Road, Dapeng District, Shenzhen 518124, China. ²Horticultural Sciences, Plant Innovation Center, University of Florida, Post Office Box 110690, Gainesville, FL 32611, USA. ³Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Horticultural Crops of the Ministry of Agriculture, Sino-Dutch Joint Laboratory of Horticultural Genomics, Beijing 100081, China. ⁴RAPiD Genomics, 747 SW 2nd Avenue, Gainesville, FL 32601, USA. ⁵Instituto de Biología Molecular y Celular de Plantas (Consejo Superior de Investigaciones Científicas–Universitat Politècnica de València), València, Spain. ⁶Faculty of Agriculture, Hebrew University of Jerusalem, Rehovot, Israel. ⁷School of Forest Resources and Conservation, Genetics Institute, University of Florida, Gainesville, FL 32611, USA.

*These authors contributed equally to this work. †Corresponding author. Email: huangsanwen@caas.cn (S.H.); hjklee@ufl.edu (H.K.)

from the GWAS that have significant associations with glucose and fructose content on chromosomes 9 and 11 (Fig. 1A and table S4). Both of these loci are located within regions previously identified as being within both domestication and improvement sweeps, indicating early and continued selection for larger fruit (Fig. 1, C and D) (7). Most (47 out of 48) modern cultivars contain the reference alleles at these two loci. This combination of alleles results in significantly lower sugar content in fruit than all of the other allele combinations (Fig. 1E). This strong selection is likely due to a negative correlation between fruit size and sugar content. To further examine this potential negative correlation, we evaluated the combined effects of sugar content and fruit weight in the segregating F_2 population. Four significant associations with sugar content were identified on chromosomes 2, 3, 5, and 9 (Fig. 1B). The chromosome 9 locus colocalizes with the locus on chromosome 9 identified by GWAS and corresponds to the previously identified extracellular invertase, *Lin5* (8, 9). Although it did not show up as significant in the F_2 analysis, the population is also segregating for the chromosome 11 sugar locus identified by the GWAS (*SSC11.1*). Analysis of fruit from the F_2 population, by segregating for each

allele at the five loci, showed a strong negative correlation between fruit weight and sugar content (Fig. 1F). We further examined the molecular basis for the chromosome 9 sugar SNP. The *Lin5* SNP causes an Asn to Asp change at position 366. To validate *Lin5* as the causative gene for the chromosome 9 locus, transgenic plants overexpressing either the reference or alternate enzyme were produced. On average, fruit overexpressing the alternate enzyme contained significantly more sugar than the parent fruit or fruit expressing the reference enzyme, which is consistent with a role for the alternate form of *Lin5* in causing higher fruit sugar content (Fig. 1G).

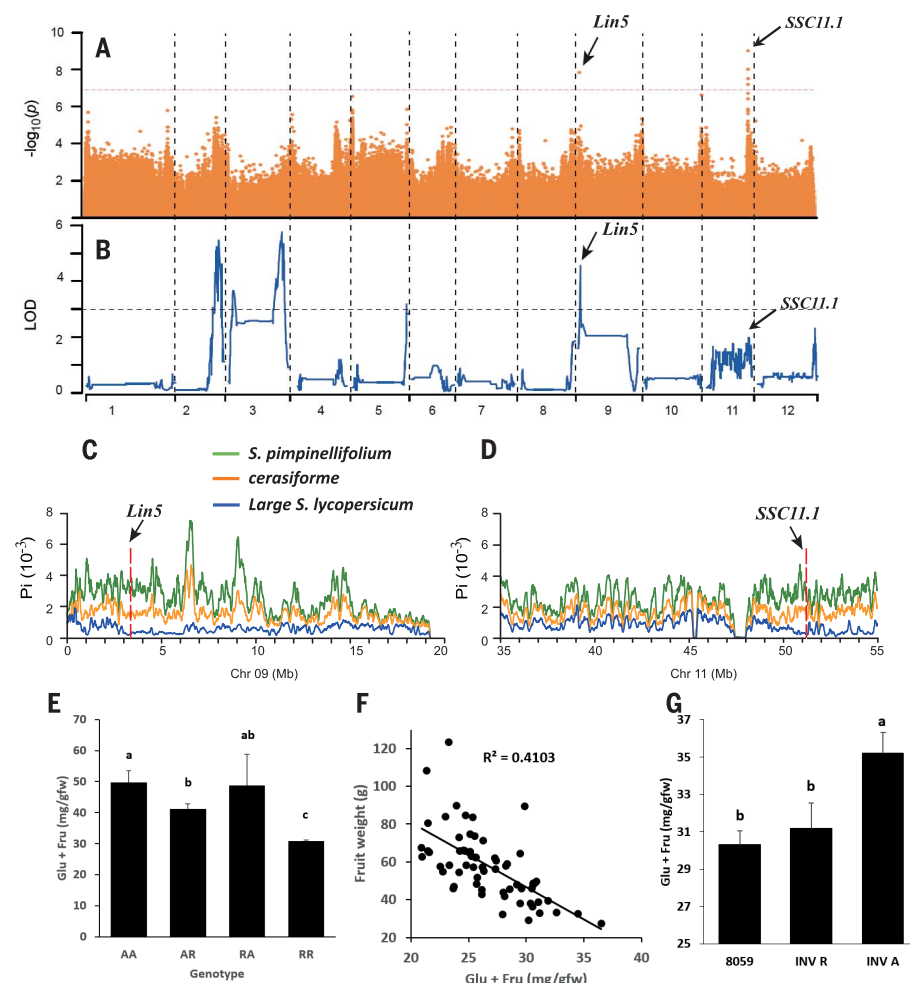
Taken together, the negative correlation between fruit weight and sugar content in *S. lycopersicum* is likely associated with the loss of the high-sugar alleles during domestication and improvement as ever-larger fruits were selected.

The data provide many interesting candidates for gene function. For example, a locus on chromosome 9 that has significant effects on the contents of two metabolically linked volatiles, guaiacol and methylsalicylate, was identified (figs. S17 and S20 and table S4). Methylsalicylate is derived from the stress-associated phytohormone salicylic acid (10). These two volatiles have been

described as medicinal and have been proposed to negatively contribute to consumer liking (11, 12). Varieties containing the reference allele have 4.8- and 5.5-fold more methylsalicylate and guaiacol, respectively, than varieties containing the alternate allele. Within the area of significantly correlated SNPs is *E8* (Soly009g089580), one of the most abundantly expressed, ripening-specific genes in tomato fruit (13). Three significantly associated SNPs are in the upstream intergenic region of *E8*, one is in an intron of *E8*, and another is in the downstream intergenic region of *E8*. We validated *E8* as the causative gene by quantifying volatile production in an antisense line (14) compared to its control cultivar, Ailsa Craig. The antisense line exhibited a pattern of volatile emissions that was similar to the average of all lines containing the alternate allele (Fig. 2). Loss of *E8* function modulates the levels of multiple volatiles but does not abolish their synthesis. Although the precise mechanism of *E8* action remains unknown, it may increase ethylene content, which in turn could affect expression of genes involved in volatile synthesis (14).

Multiple loci affecting carotenoid-derived volatiles (apocarotenoids) were identified by the GWAS (table S4). Two apocarotenoid volatiles,

Fig. 1. Genetic interactions between sugar content and weight of fruit. (A) Manhattan plot of SNPs associated with total soluble solids in the 398-member population from Florida. (B) LOD plot of SNPs associated with glucose and fructose concentration in a segregating F_2 population from a cross between lines FLA 8059 and Maglia Rosa Cherry. In (A) and (B), the x axis shows chromosomal position. (C) Distribution of nucleotide diversity of *S. pimpinellifolium*, *S. lycopersicum* var *cerasiforme*, and large *S. lycopersicum* varieties on chromosome 9. The position of *Lin5* is located in a domestication and improvement sweep. Pi, nucleotide diversity. (D) Distribution of nucleotide diversity of *S. pimpinellifolium*, *S. lycopersicum* var *cerasiforme*, and large *S. lycopersicum* varieties on chromosome 11. *SSC11.1* is located in a domestication and improvement sweep. (E) Sugar content [glucose (glu) and fructose (fru) measured in milligrams per gram of fresh weight (mg/gfw)] in each of the four combinations of reference (R) and alternate (A) alleles in chromosomes 9 (chr9:3477979) and 11 (chr11:51186147) in the F_2 population, respectively. Letters indicate separations as determined by pairwise Student's *t* test ($P < 0.05$), where bars with no shared letters indicate a statistically significant difference between the data sets. (F) Sugar content versus fruit weight in different chromosome allele (chr2:48533210, chr3:62073897, chr5:63443613, and chr9:3300595) combinations in the F_2 population. R^2 , coefficient of determination. (G) Average sugar content of transgenic lines overexpressing reference (InvR, $n = 6$) and alternate (InvA, $n = 17$) forms of cell wall invertase (*Inv*) relative to the control line FLA 8059 (8059). Letters indicate separations as determined by pairwise Student's *t* test ($P < 0.05$), where bars with no shared letters indicate a statistically significant difference between the data sets.



geranylacetone and MHO, illustrate how a simple visual score, or lack thereof, can impact allele selection. Geranylacetone is generated by oxidative cleavage of phytoene, phytofluene, ζ -carotene, and neurosporene. These carotenoids are present in small amounts and do not make meaningful contributions to the color of a ripe fruit. In contrast, MHO is generated by oxidative cleavage of lycopene, the major color pigment in a ripe fruit. The amounts of apocarotenoid volatiles are proportional to their carotenoid precursor contents in fruits (5, 15). Thus, MHO content is directly proportional to lycopene content. We identified one

MHO locus and four geranylacetone loci, as well as two loci associated with both volatiles in the 398-member population. We determined the apocarotenoid volatile contents for different combinations of alleles in wild, heirloom, and modern accessions, as well as the frequencies of those allelic combinations in each group (Fig. 3). Allelic combinations for both volatiles were progressively lost during human selection. Every modern accession contains one of two combinations of the three MHO alleles, and most (39 of 41 with available sequence information) contain only two combinations of the four alleles

specific for geranylacetone. Because the color associated with lycopene content is readily apparent, breeders, in selecting deep-red fruits, have enriched for the two alleles associated with the highest MHO content. In contrast, where selection requires gas chromatography, the two most common geranylacetone allelic combinations are not the most desirable combinations. We also noted a strong association with both geranylacetone and MHO on chromosome 3 (figs. S14 and S16). The gene encoding the enzyme involved in the first step of carotenoid synthesis, phytoene synthase, is within this area of association (table S4).

Whole-genome sequencing and the GWAS resulted in the identification of candidate loci capable of altering 15 of the chemicals contributing to consumer liking and an additional 6 chemicals that contribute to overall flavor intensity of tomatoes. Although most breeding programs routinely assess fruit size as well as sugar and acid content, they lack the means to track volatile content. Thus, there has been strong selection for alleles favoring fruit weight (described above) and acid content (20 of 20 loci with the favorable allele) in the modern variety cluster. In contrast, only 13 of 26 volatile-associated loci contain the desirable allele. This distribution was not significantly different from a random set of 32 SNPs (data not shown), consistent with random loss of desirable alleles in the

Fig. 2. E8 is associated with a guaiacol-methylsalicylate locus on chromosome 9. Relative volatile contents of accessions containing alternate or reference alleles at chr9:69296937 and an antisense E8 (Soly09g089580) transgene relative to nontransgenic line Ailsa Craig (AC). Significantly different from control, as determined by pairwise Student's *t* test. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

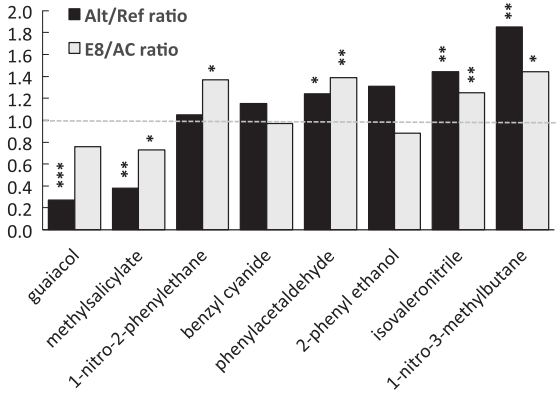
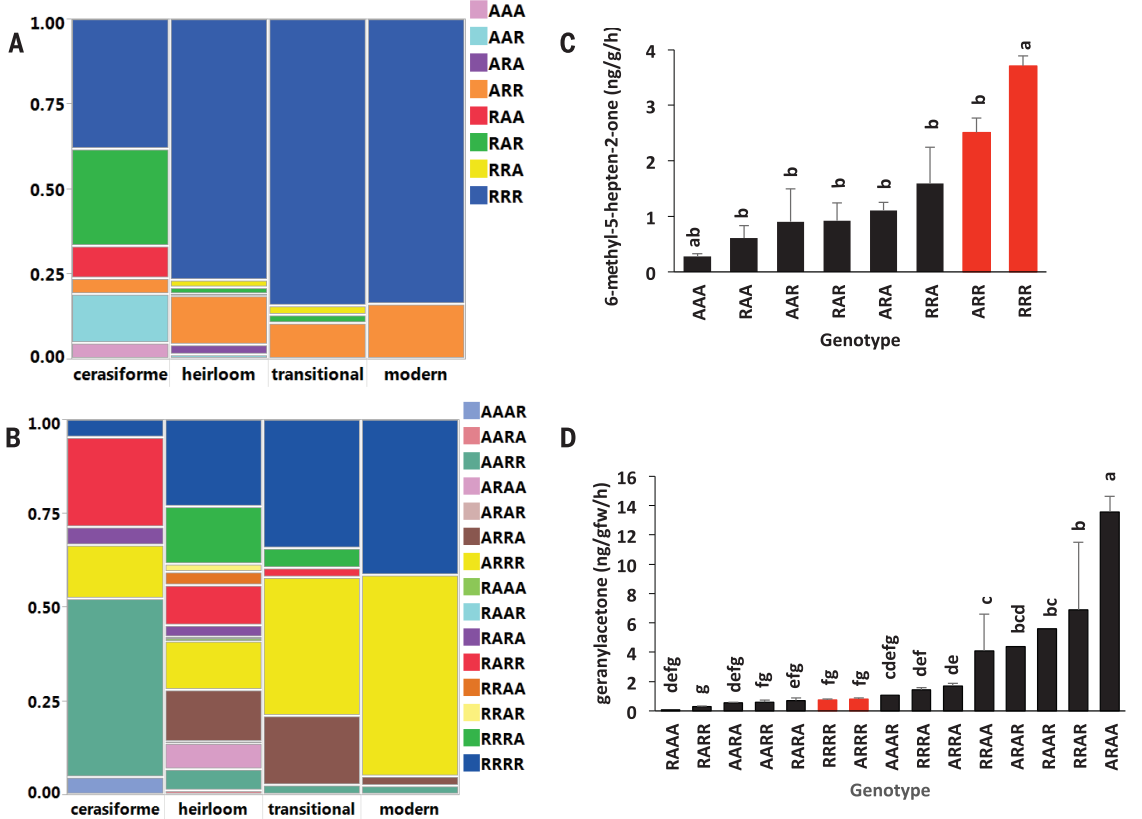


Fig. 3. Combinations of 6-methyl-5-hepten-2-one (MHO) and geranylacetone alleles found in *S. lycopersicum* var *cerasiforme*, heirloom, transitional, and modern tomato varieties. In all panels, *R* is the reference (Heinz 1706) allele, and *A* is the alternate allele. (A) Allele distribution of MHO loci at positions chr2:40378276, chr3:4024959, and chr9:66583920 in *S. lycopersicum* var *cerasiforme*, heirloom, transitional, and modern varieties.

(B) Allele distribution in varieties containing *R* or *A* allele combinations associated with geranylacetone content. Four loci located at chr2:40883244, chr4:60434091, chr10:63081709, and chr11:7652084 were chosen. (C) Mean (\pm SE) MHO content in varieties containing the indicated allele combinations. (D) Mean (\pm SE) geranylacetone content in the indicated allele combinations. The two most abundant allele combinations in modern varieties are indicated in (C) and (D) in red. Letters in (C) and (D) indicate separations as determined by pairwise Student's *t* test, where bars with no shared letters indicate a statistically significant difference in volatile contents between the data sets.



absence of any selection. Reductions in flavor volatiles due to the loss of a single superior allele may not be apparent to an individual making selections, but cumulative loss of superior alleles affecting many volatiles has resulted in overall flavor deterioration over many breeding cycles. However, content of desirable chemicals can be increased, and content of undesirable chemicals decreased, by utilizing molecular markers associated with superior alleles.

Replacement of undesirable alleles should have a strong positive effect on consumer liking. Since volatiles are active at picomolar to nanomolar concentrations, substantial increases in their contents can be achieved with minimal metabolic penalty and less yield drag. However, sugars, which correlate highly with consumer liking, are present in millimolar concentrations, so increasing their content will only be achievable at the cost of reduced fruit size. We identified multiple loci with potential for increasing fruit sugar content. For example, reintroduction of the alternate *Lin5* should result in higher sugar content. Although in most cases, higher sugar content is associated with a reduction in the size of a fruit, consumers do not prefer large fruit and are very willing to purchase minimally smaller fruit with superior taste (16). One possible solution to this trade-off between higher sugar and smaller size is based on our observations that certain fruit volatiles significantly enhance the percep-

tion of sweetness (2, 5). Notably, large effects on perceived sweetness are conferred by apocarotenoid volatiles. The selection of alleles associated with higher apocarotenoid content that are identified here can potentially improve perceived sweetness without negatively affecting fruit size.

Modern commercial tomatoes do not have the flavor of older varieties. Although postharvest practices, such as refrigeration, can irreversibly damage flavor, making improvements in flavor-associated chemicals is the first step to restoring the potential of widely grown commercial varieties. Our results provide a roadmap for improvement of flavor. The genes and pathways identified here in the tomato almost certainly point to pathways worth investigating for improvement of flavor quality in other fruit crops.

REFERENCES AND NOTES

- Food and Agriculture Organization of the United Nations; <http://faostat.fao.org/site/339/default.aspx>.
- D. Tieman *et al.*, *Curr. Biol.* **22**, 1035–1039 (2012).
- R. G. Buttery, R. Teranishi, R. A. Flath, L. C. Ling, *Am. Chem. Soc. Symp.* **388**, 213–222 (1987).
- E. A. Baldwin, J. W. Scott, C. K. Shewmaker, W. Schuch, *HortScience* **35**, 1013–1022 (2000).
- J. Vogel *et al.*, *J. Sci. Food Agric.* **90**, 2233–2240 (2010).
- B. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12580–12585 (2016).
- T. Lin *et al.*, *Nat. Genet.* **46**, 1220–1226 (2014).
- E. Fridman, F. Carrari, Y.-S. Liu, A. R. Fernie, D. Zamir, *Science* **305**, 1786–1789 (2004).
- M. I. Zanor *et al.*, *Plant Physiol.* **150**, 1204–1218 (2009).
- D. Tieman *et al.*, *Plant J.* **62**, 113–123 (2010).
- M. I. Zanor *et al.*, *J. Exp. Bot.* **60**, 2139–2154 (2009).
- B. Zierler, B. Siegmund, W. Pfannhauser, *Anal. Chim. Acta* **520**, 3–11 (2004).
- J. Deikman, R. Kline, R. L. Fischer, *Plant Physiol.* **100**, 2013–2017 (1992).
- L. Peñarrubia, M. Aguilar, L. Margossian, R. L. Fischer, *Plant Cell* **4**, 681–687 (1992).
- E. Lewinsohn *et al.*, *J. Agric. Food Chem.* **53**, 3142–3148 (2005).
- A. E. Oltman, S. M. Jervis, M. A. Drake, *J. Food Sci.* **79**, S2091–S2097 (2014).

ACKNOWLEDGMENTS

This work was supported by the NSF (grant IOS-0923312 to H.K.), the China National Key Research and Development Program for Crop Breeding (grant 2016YFD0100307 to S.H.), the Leading Talents of Guangdong Province Program (grant 00201515 to S.H.), the National Natural Science Foundation of China (grant 31601756 to G.Z.), the European Research Council (grant ERC-2011-AdG 294691 YIELD to D.Z.), and the European Commission Horizon 2020 program (TRADITOM grant 634561 to A.G. and D.Z.) This work was also supported by the Chinese Academy of Agricultural Science (ASTIP-CAAS) and the Shenzhen municipal and Dapeng district governments. We acknowledge the assistance of L. Kates in fieldwork and volatile, sugar, and acid quantification.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/355/6323/391/suppl/DC1
Materials and Methods
Figs. S1 to S25
Tables S1 to S8
References (17–35)
6 October 2016; accepted 22 December 2016
10.1126/science.aal1556