Title

Leaf shape is a predictor of fruit quality and cultivar performance in tomato

Journal

New Phytologist, 226(3)

ISSN

0028-646X

Authors

Rowland, Steven D
Zumstein, Kristina
Nakayama, Hokuto
et al.

Publication Date

2020-05-01

DOI

10.1111/nph.16403

Peer reviewed

New Phytologist

Research

Leaf shape is a predictor of fruit quality and cultivar performance in tomato

Steven D. Rowland[1], Kristina Zumstein[1], Hokuto Nakayama[1,2], Zizhang Cheng[3], Amber M. Flores[1], Daniel H. Chitwood[1,4], Julin N. Maloof[1] and Neelima R. Sinha[1]

[1]Department of Plant Biology, University of California, Davis, CA 95616, USA; [2]Gradute School of Science, University of Tokyo, Hongo Bunkyo-ku, Tokyo 113-0033, Japan; [3]College of Science, Sichuan Agriculture University, Yaan, Sichuan Province 625014, China; [4]Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA

Author for correspondence:
Neelima R. Sinha
Tel: +1 530 7548441
Email: nrsinha@ucdavis.edu

Received: *12 August 2019*
Accepted: *14 December 2019*

New Phytologist (2020) **226:** 851–865
doi: 10.1111/nph.16403

Key words: biomass, fruit quality, heirloom, leaf shape, partial least squares path modeling (PLS-PM), photosynthesis, *Solanum lycopersicum* (tomato).

Summary

- Commercial tomato (*Solanum lycopersicum*) is one of the most widely grown vegetable crops worldwide. Heirloom tomatoes retain extensive genetic diversity and a considerable range of fruit quality and leaf morphological traits.
- Here the role of leaf morphology was investigated for its impact on fruit quality. Heirloom cultivars were grown in field conditions, and BRIX by yield (BY) and other traits were measured over a 14-wk period. The complex relationships among these morphological and physiological traits were evaluated using partial least-squares path modeling, and a consensus model was developed.
- Photosynthesis contributed strongly to vegetative biomass and sugar content of fruits but had a negative impact on yield. Conversely leaf shape, specifically rounder leaves, had a strong positive impact on both fruit sugar content and yield. Cultivars such as Stupice and Glacier, with very round leaves, had the highest performance in both fruit sugar and yield. Our model accurately predicted BY for two commercial cultivars using leaf shape data as input.
- This study revealed the importance of leaf shape to fruit quality in tomato, with rounder leaves having significantly improved fruit quality. This correlation was maintained across a range of diverse genetic backgrounds and shows the importance of leaf morphology in tomato crop improvement.

Introduction

The rise of agriculture *c.* 7000 BC ensured a stable food supply, allowing human civilizations to develop and populations to grow (Barker, 2006). The challenge of feeding a growing population is exacerbated by climate unpredictability, with drought and temperature increases, leading to decreased crop yield (Matiu *et al.*, 2017). Tomato (*Solanum lycopersicum*) is by far the most widely grown vegetable crop worldwide (Bauchet & Causse, 2012). The narrow genetic base of most crops, combined with selection for performance under optimal conditions, has reduced the genetic variability in environmental stress responses, and the modern cultivars of tomato are no exception (Bai & Lindhout, 2007; Bauchet & Causse, 2012; Bergougnoux 2014) . The wild relatives of tomato have the genetic ability to adapt to extreme habitats, and many heirloom cultivars also retain this ability as a result of directed breeding with wild species, and less selection for commercially valuable traits (Sim *et al.*, 2012; Lin *et al.*, 2014; Blanca *et al.*, 2015; RodrÍGuez-Burruezo *et al.*, 2005) . Heirloom tomatoes are defined as varieties, which have been passed down through multiple generations of a family (Tomato Fest, https://www.tomatofest.com/what-is-heirloom-tomato.html).

Improvement in tomato has focused on flowering, fruit traits, and disease resistance probably as a result of a perceived negative correlation between fruit size and sugar content (Tieman *et al.*, 2017). Thus, potential impacts of other factors on yield and fruit quality are relatively ignored (Grandillo *et al.*, 1999; RodrÍGuez-Burruezo *et al.*, 2005; Passam *et al.*, 2007; Bauchet & Causse, 2012; Bergougnoux 2014; Tieman *et al.*, 2017; Zhu *et al.*, 2018).

In a previous study by Chitwood *et al.* (2014), a meta-analysis on a set of introgression lines linked leaf complexity and leaflet shape in tomato to fruit sugar content measured on the same lines by other researchers (Baxter *et al.*, 2005). This correlation showed that plants with complex and rounder leaflets also had increased fruit sugar content (Chitwood *et al.*, 2014). Because leaves are the primary site of photosynthesis, it is possible that leaf shape changes may impact photosynthetic capacity and therefore result in different sugar content (BRIX) and yield in fruits. In addition to photosynthesis, sugar transport, and distribution to sinks are other potential sites of regulation in leaf function as source tissue. While sugar transport in plants is well described, distribution among different sink tissues is not fully understood (Lemoine *et al.*, 2013).

We analyzed tomato cultivars with varied yield and fruit quality, photosynthetic capacity, leaflet shape, and other vegetative traits and found that leaflet shape was strongly correlated to overall fruit quality assessed as a composite measure of BRIX and yield (BY; Eshed and Zamir, 1995), with rounder leaflets positively correlated with higher BY values. Photosynthesis, on the other hand, had a negative correlation with yield. Based on our analysis, leaf shape seems to play an important role in the distribution of photoassimilates. Additionally, we performed phylogenetic network analysis on 23 cultivars, including eight identified as having the rounder Potato Leaf Morph (and included in our analysis), known to be caused by a mutation in the *C-locus* (Busch *et al.*, 2011), to determine their breeding histories and identify any potential selection for this trait.

Materials and Methods

Plant material and growth conditions

Eighteen heirloom tomato varieties identified as having a range of fruit types, including cherry and beefsteak tomatoes, and several intermediate types, were analyzed. These tomato varieties also differed in fruit production timing from early to late, and the type of leaf morphology. These cultivars were selected based on leaf shape as described in Tatiana's TOMATObase and The Heirloom Tomato (http://tatianastomatobase.com/wiki/Main_Page; Goldman, 2008; Supporting Information Fig. S1).

Tomato seeds were treated, germinated, and field planted as previously described (Chitwood *et al.*, 2014). In both the 2014 and 2015 seasons, plants were laid out in a randomized block design and were planted (late May) and grown in soil, with furrow irrigation once weekly.

Gas exchange and intercepted PAR measurements

Gas exchange measurements were done in the field on attached leaves after the plants had recovered from transplanting. Measurements were made weekly from week 10 to week 15 (vegetative growth), on week 17 (initiation of flowering), and weeks 18–21 (fruiting stages), on *c.* 60 plants each week, on three plants per cultivar wk^{-1}. Measurements were made on leaves from the upper and lower portions of the plants to eliminate positional bias within the plant, and measured for three leaves per plant. The A (photosynthesis), g_{st} (stomatal conductance), transpiration, and ϕ_{PS2} (amount of photons entering photosystem II) of a 6 cm^2 area of the leaflet were measured using the LI-6400 XT infrared gas exchange system (Li-Cor, Lincoln, NE, USA), and a fluorescence head (6400-40; Li-Cor). The chamber was positioned on terminal leaflets such that the midvein was not within the measured area. Light within the chamber was provided by the fluorescence head at 1500 $\mu mol\ m^{-2}\ s^{-1}$ photosynthetically active radiation (PAR), and the chamber air flow volume was 400 $\mu mols\ s^{-1}$ with the chamber atmosphere mixed by a fan. CO_2 concentration within the chamber was set at 400 $\mu mols\ mol^{-1}$ (average atmospheric concentration). Humidity, leaf and chamber temperature were allowed to adjust to ambient conditions; however, the chamber block temperature was not allowed to exceed 36°C. Measured leaflets were allowed to equilibrate for 2–3 min before measurements were taken, allowing sufficient time for photosynthetic rates to stabilize with only marginal variation.

The amount of intercepted PAR (PAR_i) was measured in four orientations per plant and an average PAR_i calculated. PAR_i was measured by placing a Line Quantum Sensor (LQA-2857; Li-Cor) onto a base made from ¼" PVC piping, and a Quantum Sensor (LI-190R; Li-Cor) approximately 1 m above the plant on the PVC rig. Measurements from both sensors were taken simultaneously for each sample using a Light Sensor Logger (LI-1500; Li-Cor). This allowed variation in overall light intensities such as cloud movement to be measured and accounted for in the total PAR_i.

Harvest measurements

After gas exchange measurements, three plants per cultivar were destructively harvested each week. The final yield (weight of all fruit per plant) and fresh vegetative weight of each plant harvested was measured using a hanging scale (TL 440; American Weigh Scale Inc., Norcross, GA, USA) in the field. Five leaves were collected at random from the bottom and top of the plant to capture all canopy levels, and approximately nine fruit were collected for BRIX measurements. FW was used owing to the large number of plants and measurements being done *in situ* in the field setting. All measurements were made in kg. To measure the BRIX value of the tomatoes, the collected fruit was taken to the laboratory where the juice was collected and measured on a refractometer (HI 96801 Refractometer; Hanna Instruments, Woonsocket, RI, USA). The yield and BRIX for each plant were multiplied together to get the BRIX × yield index (BY), which gives an overall fruit quality measure, accounting for variations and extreme values in either measurement. It should be noted that while BRIX is used as a standard quality measure, BY is a composite value that folds in yield to assess weight (kg) of soluble solids per plant and is being used to measure commercial (grower) quality and not consumer (taste) quality (Eshed and Zamir, 1995). BY measurements were done for both the 2014 (preliminary field) and the more detailed 2015 fields. These data were compared to test for reproducibility of results (Fig. S2).

Leaf morphology analysis

The leaf complexity measures included all leaflets present on the leaf. Subsequently, primary leaflets were used for imaging and analysis of shape and size as previously described (Chitwood *et al.*, 2014), and the images then processed in ImageJ (Schneider *et al.*, 2012). The images were cropped to individual leaflets maintaining the exact pixel ratio of the original image, and then cropped again to only include the single leaflet using a custom Java script written for Fiji (Schindelin *et al.*, 2012). Single leaflet images converted to a binary image as black on a white background, and smoothed to allow for the exclusion of any particulates in the image were then processed in R using momocs

(Bonhomme *et al.*, 2014), a shape analysis package. Leaflet images were imported and then aligned along their axes so that all images faced the same direction. They were then processed using elliptical Fourier (eFourier) analysis based on the calculated number of harmonics from the MOMOCS package. Principal component analysis was performed on the resulting eFourier analysis and the principal components (PCs) were used for subsequent analysis. Traditional shape measures such as leaflet area, circularity, solidity, and roundness were done with the area measurement based on pixel density. These measures were compared with the PCs to determine the characteristics captured by each PC. The PC values were used for all subsequent leaflet shape and size analyses. Total leaf area for each plant was measured by imaging the whole plant and a 4 cm^2 red square and then processed in the EASY LEAF AREA software (Easlon & Bloom, 2014; Fig. 2b).

Leaflet sugars

Five plants per line were used to analyze leaflet sugar content. The plants were grown under the same conditions as field plants with the following exceptions. Plants remained in the glasshouse after transfer to 1 gallon (4.546 l) pots. All plants were watered with nutrient solution and grown until mature leaves could be sampled. Using a hole punch, a disk with an area of 0.28 cm^2 was taken from the leaflets and extracted from the disks using a modified extraction method from the Ainsworth laboratory (Bishop *et al.*, 2018). Leaf disks were placed in 2 mM HEPES (Affymetrix Inc., Santa Clara, CA, USA) in 80% EtOH (Sigma-Aldrich) and heated to 80°C for 20 min and the liquid collected and stored at −20°C. The entire process was repeated twice. They were then placed in 2 mM HEPES in 50% EtOH and heated, collecting the liquid and storing at −20°C followed by another 2 mM HEPES in 80% treatment. The collected liquid was then used to measure the amount of sugar present per area of disk.

To measure leaf sugar content a working solution of 100 mM HEPES (Affymetrix Inc.), 6.3 mM $MgCl_2$ (Sigma-Aldrich), and 3 mM ATP (Sigma-Aldrich) and NADP (Sigma-Aldrich) at pH 7 was prepared. From the working solution, an assay buffer was made adding 50 U of glucose-6-phosphate dehydrogenase (G6PDH; Sigma-Aldrich), and 295 or 280 µl of the working solution was added to a 96-well plate (Costar, Corning, NY, USA) for sucrose standards or samples, respectively. Standards were added at a 60-fold dilution and samples were added at a 15-fold dilution. Then 0.5 U of hexokinase (Sigma-Aldrich), 0.21 U of phosphoglucoisomerase (PGI; Sigma-Aldrich), and 20 U of invertase (Sigma-Aldrich) were added to each well and the plates allowed to sit overnight to reach equilibrium. The plates were measured on a UV spectrometer (Molecular Devices SPECTRAmax 340, San Jose, CA, USA) at 340 nm, followed by analysis in JMP (JMP Pro 14.0.0, 2018; SAS Institute Inc., Cary, NC, USA).

Statistical analysis

All statistical analyses were performed using JMP (JMP Pro 14.0.0, 2018) software. To determine statistical significance, measurements were modeled using general linear regression model and tested by a one-way ANOVA followed by Tukey's honestly significant difference, if necessary. These modeled data for all measured values were compiled into a table and used to create a model using partial least-squares path modeling (PLS-PM) in SMARTPLS 3.0 (Ringle *et al.*, 2015). Modeled data were used for the statistical analyses as many measurement types varied in number of data points, and therefore a set of generated predicted values of equal size was used to make an equal data matrix (Table S1).

Partial least squares-PM was used to explore the cause-and-effect relationships between the measured variables through latent values. PLS-PM is effective in both exploring unknown relationships and combining large-scale data, such as field, physiological, and morphological data, that otherwise are not well described together (Barberán *et al.*, 2014). In addition to running the PLS-PM, 1000 bootstraps were performed to obtain statistical significance and confidence intervals of the path coefficients and the R^2 values of each latent variable. The path coefficients are the standardized partial regression coefficients (Barberán *et al.*, 2014), and represent the direction and strength of causal relationships of direct effects. Indirect effects are the multiplied coefficients between the predictor variable and the response variable of all possible paths other than the direct effect (Barberán *et al.*, 2014). To determine the best path model, the latent variables (LVs) were combined using our best understanding of biological relationships, and a general model using all data was generated. The paths between LVs were altered until a best-fit model was found. PLS-Predict was then used on the dataset to ensure that the model did not over or under fit the data, and for predictive performance of each manifest variable (MV). This structural model, and not the fit values, was retained for use in predictive modeling of a separate dataset.

PLSPREDICT, with the structural model developed as described earlier, was used on a separate dataset to determine the efficacy of the model. Two commercial cultivars, M82 and Lukullus, were used and only the leaf shape values were entered as exogenous variables. The predicted values for each output variable (yield, BRIX, and vegetative mass) were compared with the actual measured values to determine how well the model predicted these variables.

DNA library construction and sequencing

DNA was extracted using a GeneJET Plant Genomic DNA Purification Mini Kit (Thermo Scientific, Waltham, MA, USA) from plants grown for a month, and DNA-Seq libraries were prepared based on BrAD-seq (Townsley *et al.*, 2015) with the following modifications. After DNA fragmentation with Covaris E220 (Covaris, Inc., Woburn, MA, USA), the fragmented DNA was end-repaired, A-tailed, and adapter-ligated with Y-adapter. Enrichment PCR was then performed with the adapter-ligated product as previously described (Townsley *et al.*, 2015). After final library clean-up with AMPure beads (Beckman Coulter, Brea, CA, USA), DNA-Seq libraries were sequenced at Novogene (Sacramento, CA, USA) using the HiSeq 2500 platform (Illumina Inc., San Diego, CA, USA). All data are deposited in sequence read archive (accession no. PRJNA527863).

Phylogenetic analyses

To perform phylogenetic analysis, all single nucleotide polymorphisms (SNPs) detected by CLC Genomics Workbench 11.0 (CLC Bio; a Qiagen Company, Aarhus, Denmark) from whole-genome sequencing (WGS) were exported as a vcf file. The SNPRELATE package for R (Zheng *et al.*, 2012) was used to determine the variant positions that overlapped between cultivars and then all sequences combined into a single gds file (Table S2). This file was run through SNPhylo (Lee *et al.*, 2014) with the following parameters: the linkage disequilibrium was set to 1.0, as we wanted to exclude as few variants as possible based on this factor, the minor allele frequency was set to 0.05, and the missing rate was set to 0.1. In all, 1000 bootstraps were performed for confidence intervals and significance. *Solanum pimpinellifolium* was used as the outgroup. The bootstrapped output tree was displayed in MEGA7 (Kumar *et al.*, 2016) . Analysis of *c* gene flow was performed using PHYLONETWORKS (Solís-Lemus *et al.*, 2017). All common SNPs from chromosome 6 were run through the TICR pipeline (Stenz *et al.*, 2015) and then analyzed using PHYLONETWORKS with default settings, except for the number of runs which was set to 20. After the hybrid network for chromosome 6 was obtained, bootstrap analysis was done in PHYLONETWORKS using default settings with the following exceptions: ftolRel was set to 0.01, ftolAbs was set to 0.001, liktolAbs was set to 0.0001, and Nfail was set to 5. These adjustments were made to decrease processing time. The bootstrapped tree was output in DENDROSCOPE (Huson & Scornavacca, 2012).

Results

Fruit and vegetative traits

Yield, and fruit BRIX (soluble solids content) were measured over 14 wk of the growing season (Fig. S3). For most cultivars the yield remained at or below 5 kg of fruit per plant at the time of harvest (Fig. S3a). The exceptions to this were Bloody Butcher, Glacier, Brandywine, Prudens Purple, and Stupice. All had a yield higher than 5 kg per plant, with Stupice having the highest yield, at *c.* 20 kg per plant by final harvest (week 23; Fig. S3a). Fruit BRIX remained nearly constant for all cultivars across the growing season (Fig. S3b), with the exception of ABC Potato Leaf, which had a large increase at time of harvest.

To better quantify fruit quality, BRIX and yield were multiplied to obtain a BY value (Grandillo *et al.*, 1999) for each measurement (Fig. 1). Bloody Butcher, Glacier, and Stupice all had a BY value > 60, with Stupice reaching near 100 at terminal harvest (week 23; Fig. 1). The average BY value for harvest weeks (17–23) was 16.39, while Bloody Butcher, Brandywine, Glacier, and Prudens Purple had an average of *c.* 23 (Table 1). Stupice showed the highest deviation with a mean BY value of 37.86 (Table 1), setting it apart as the highest fruit quality cultivar tested in this study. Stupice maintained a stable BRIX content in its fruit despite the large increase in yield (Fig. S3), which resulted in the large increase in overall fruit quality compared with other cultivars. Vegetative traits such as total biomass and leaf area were measured for the growing season as well (Fig. 2). Fig. 2(a) shows the vegetative biomass and leaf area over the course of the growing season, which remain stably linked, indicating that overall leaf area increase contributed to increased biomass of the plant. This trend appears common in heirloom tomatoes but is different in commercial tomatoes, which have determinate growth (Pnueli, *et al.*, 1998).

Photosynthesis

As photosynthesis is the primary source of sugar production in plants, a time course for photosynthesis, stomatal conductance (g_{st}; Table S3), PAR_i and ϕ_{PS2} was performed on all cultivars using a LI-6400XT (Fig. 3; Li-Cor). Additionally, we analyzed leaf sugar and vasculature for these lines in glasshouse conditions (Table S4; Dengler & Kang, 2001). Fig. 3(a) shows photosynthesis by g_{st} and the trend is similar among all cultivars, with photosynthesis reaching a maximum rate after 0.6–0.8 g_{st}, which is a standard response curve (Gilbert *et al.*, 2011). Optimal photosynthetic performance regardless of light conditions has been observed in a forest tree species (Ostria-Gallardo *et al.*, 2018), and we saw the same at different levels of the canopy. Fig. 3(b,c) shows the PAR_i and ϕ_{PS2}. ϕ_{PS2} had an overall downward trend across the season, as the amount of light used in photosystem II decreased with age. This corresponds well with the increase in vegetative biomass (Fig. 2a), and the increased PAR_i (Fig. 3). Individual leaf contribution to overall photosynthesis, and therefore photons used in PSII, decreases as the leaf area of the plant increases.

Because of this trend, we calculated the whole-plant photosynthesis, as photosynthesis/area, corrected for the green area visible in overhead images, and g_{st} to capture the total rates and not just specifically measured leaves (Fig. S4). The trend is linear for photosynthesis vs g_{st} when the whole plant-exposed green area is incorporated, compared with Fig. 3(a) where the trend is more logarithmic This corresponds to our previous observation that individual leaf contribution decreases as the total vegetative mass increases.

Leaf shape and *C-locus* mutations

Leaf shape was shown to be strongly correlated with fruit BRIX and sugar accumulation in a meta-analysis of an introgression population (Chitwood *et al.*, 2014). How leaf shape contributed to fruit BRIX was unclear, as shape and size of leaves do impact photosynthesis directly (Sarlikioti *et al.*, 2011), but direct links between leaf shape and fruit quality appear lacking. Here, the heirloom population used displayed a wide array of leaf shapes, from very large and narrow to small and round. To understand if this range of leaf shapes had any impact on the overall fruit quality we measured leaflet shape and size for *c.* 3733 leaflets. Fig. 4 shows the resultant PCs of all primary leaflets (terminal and lateral) measured and their relationship to traditional shape measures. PC1 contributes 78% of all variation found in the population and is tightly correlated with leaflet size ($R^2 = 0.99$), indicating that size was the largest source of variation among the

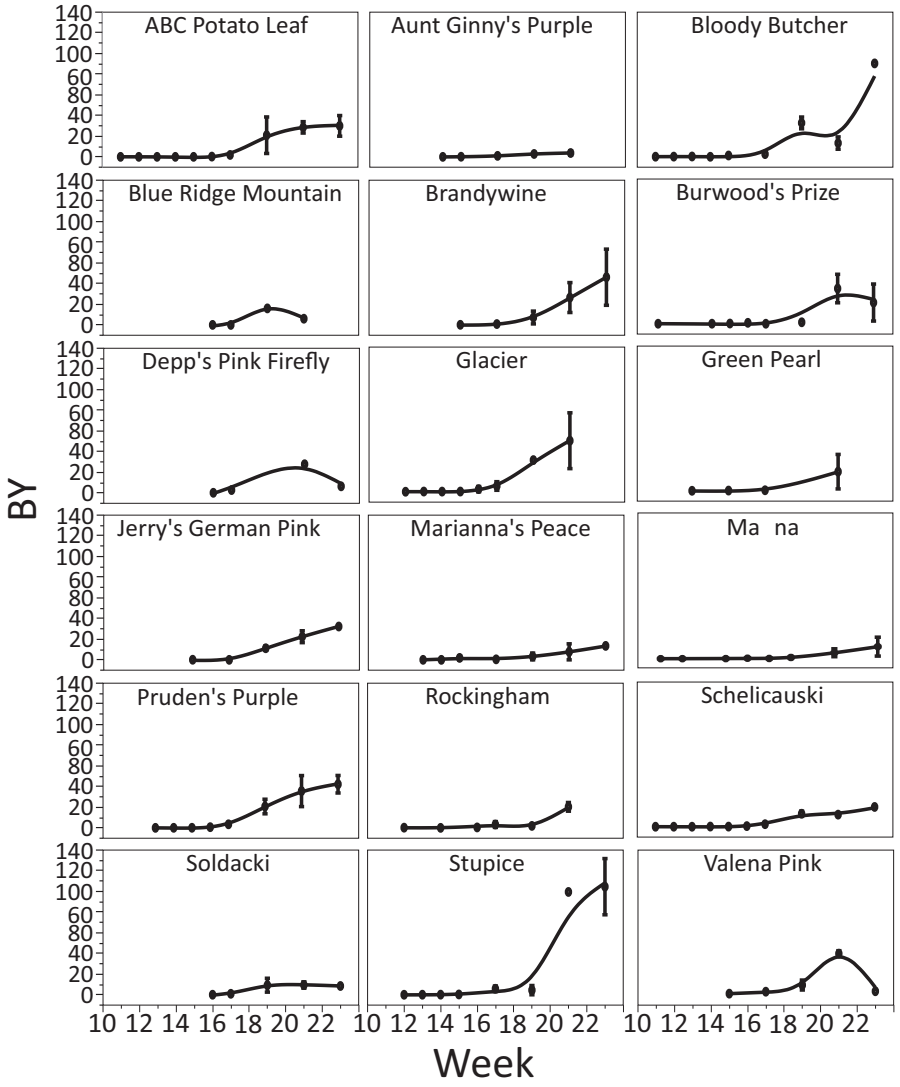

Fig. 1 BRIX (soluble solids) by yield index of 18 heirloom cultivars. Potato Leaf Morph heirloom tomato cultivars were grown in the field and the fruit BRIX and yield were measured over a 14-wk growing period. The fruit BRIX and yield were then multiplied together to obtain the BY value, giving a better indicator of overall fruit quality. Stand-out cultivars were Bloody Butcher, Brandywine, Glacier, Prudens Purple, and Stupice, all of which obtained a BY value > 20. The average BY value during harvest weeks (weeks 17–23) was 16.39. The values shown are the mean BY, and error bars represent ± SE.

heirloom leaflets (Fig. 4). PC1 was also correlated with solidity (lobing/serration; $R^2 = 0.48$), contributing to the slight shape changes seen in this PC (Fig. 4). PC2 and PC4, while having no traditional shape measure correlation, indicate the left- and right-handedness of the lateral primary leaflets, as these leaflets are mirror images of each other and therefore this measure describes the overall variation in leaf symmetry (Fig. 4a; Chitwood *et al.*, 2014). PC3 accounts for 3.8% of all variation, but has a strong correlation with aspect ratio, or the width divided by the length of the leaflet, with an R^2 of 0.8 (Fig. 4). PC3 therefore represents the roundness or narrowness of the leaflets, one aspect previously shown to be linked to fruit quality (Chitwood *et al.*, 2014).

The heirloom cultivars analyzed here were described as 'potato leaf', having broader, smoother leaves and typically lack the serration and lobes seen in other tomato varieties (Goldman, 2008; http://tatianastomatobase.com/wiki/Main_Page). However, despite this they had a wide range of leaf shape and size as illustrated in the leaf shape analysis (Fig. 4; Table S5). The classical *potato leaf* mutation (abbreviated to *c*) is caused by a 5 kb transposable element (TE) (RIDER; Jiang *et al.*, 2009; Jiang *et al.*, 2012) inserted into the third exon of the *C* locus (Solyc06g074910; Busch *et al.*, 2011). To determine if this locus harbored mutations in the selected lines, a subset of the higher-performing cultivars were selected for WGS analysis. Other mutations at the *C* locus have been described, and cause varied leaf shape (Busch *et al.*, 2011). Fig. 5(a) shows the location of the mutations found in the *C* locus in these select lines. While the full Rider insertion could not be directly determined as the reference genome lacks this insertion, overhangs on reads in the third exon matched the Rider TE sequence (Figs 5a, S5). It is possible that different sizes and fragments of Rider are present in different cultivars, as the length and sequence of the overhangs varied (Fig. S5). The identified Rider sequences were present in all but two of the sequenced lines, Prudens Purple and Glacier. No mutations were found in Glacier despite it having a rounder leaflets, although these were smaller in size with higher overall leaf complexity (Figs 5, S9). Prudens Purple had a novel single-base-pair substitution in the first exon outside the MYB/SANT conserved domain which results in the amino acid change P42R (Fig. 5a). We analyzed this mutation using the PROTEIN VARIANCE

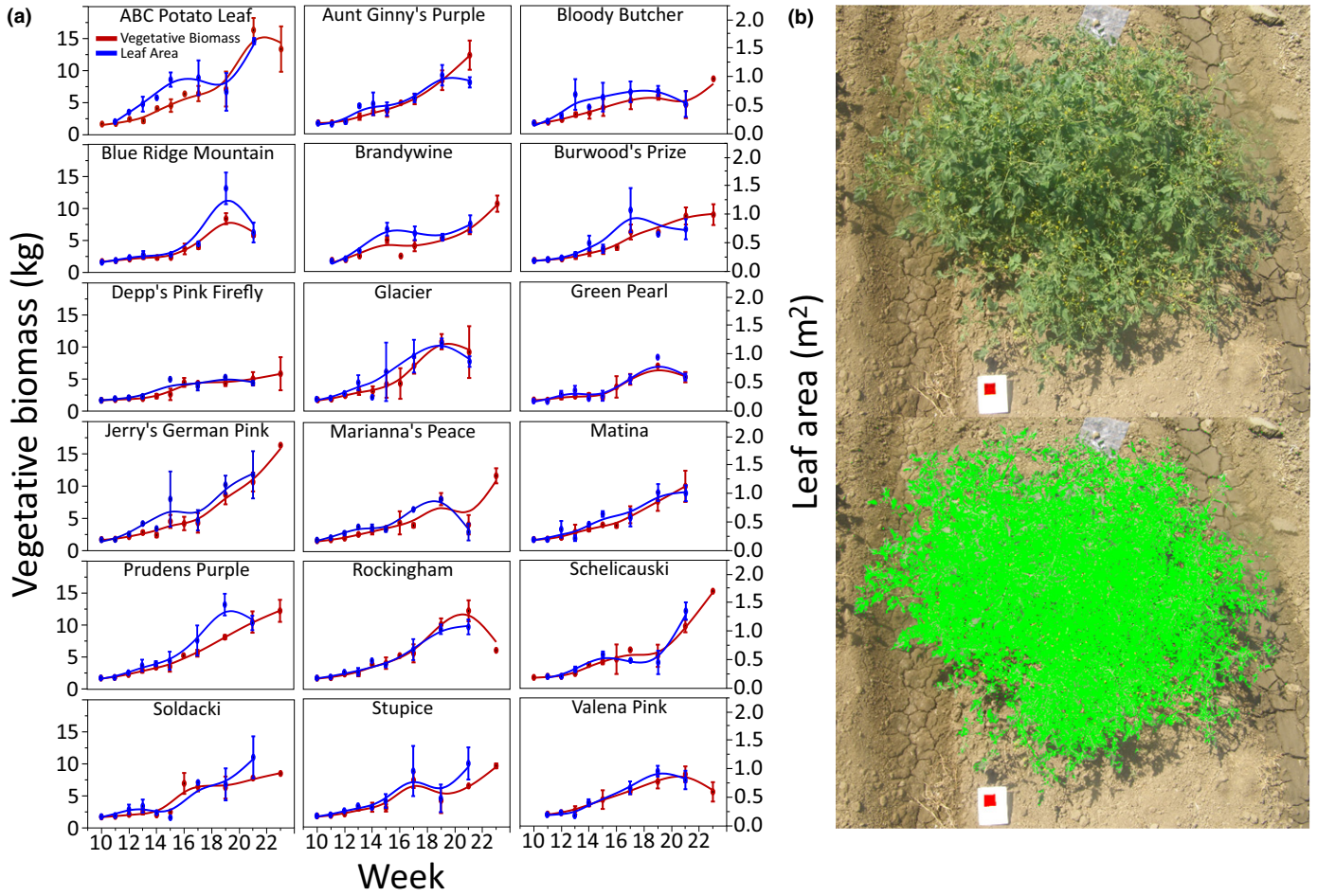

Fig. 2 Vegetative biomass and leaf area of heirloom tomatoes. Over the 14-wk growing period the total vegetative biomass (FW) and leaf area were measured. (a) The mean measurements for both total biomass and leaf area for each cultivar. Error bars represent SE. Leaf area mostly followed vegetative biomass, increasing incrementally with biomass; however, in Burwood's Prize, the leaf area levels out *c.* week 17 and does not increase with increase biomass. (b) Method for obtaining leaf area. Overhead photographs were taken of each plant, with a red square of known size, and then using (Easy Leaf Area software) the green area exposed to sunlight was calculated.

Effect Analyzer (Provean; Choi, 2012; Choi *et al.*, 2012; Choi and Chan, 2015), and found that it is predicted to be deleterious to the protein with a value of −8.454 (threshold set at −2.5), predicted to result in either a nonfunctional or partially functional protein (Choi, 2012; Choi *et al.*, 2012; Choi and Chan, 2015). Based on leaf shape analysis, Prudens Purple shows a Potato Leaf like phenotype (Fig. S6), although it differs slightly from the classical Potato Leaf shape seen in the reference allele and is reminiscent of the other mutations in *C* that have varying leaf shapes (Busch *et al.*, 2011). These data demonstrate that different mutations in *C*, coupled with genetic background differences, may give rise to a range of leaf shapes seen among some of these cultivars.

Phylogeny and phylogenetic networks

Pedigrees would probably inform the overall leaf shape in addition to the source of the *C-locus* mutations, but were not readily available. To elucidate relationships among these cultivars we used the WGS data from the select cultivars as well as WGS data obtained from the 150 Genomes Project (Aflitos *et al.*, 2014) to assemble a phylogeny and perform phylogenetic network analysis (Fig. 5b,c). The phylogeny includes several commercial cultivars, commercial heirloom cultivars, *Solanum pimpinellifolium,* and *Solanum lycopersicum* var. *cerasiforme*. ABC Potato Leaf does not appear to cluster with other Potato Leaf heirloom cultivars analyzed here (Fig. 5b). Stupice, Glacier, and Bloody Butcher are closely related in this phylogeny, corresponding to their often being listed as closely related in popular literature (Goldman, 2008), and congruent with phenotypic similarities they exhibit in fruit size and leaf shape. Bloody Butcher and Stupice both have the Rider insertion in the third exon at the *C* locus, while Glacier does not (Fig. 5a), suggesting the presence of other modifiers to leaf shape, which may have been selected for during the breeding of Glacier. A similar situation is seen in Prudens Purple (Fig. 5b), which is closely related to Jerry's German Pink and Green Pearl. While Jerry's German Pink and Green Pearl carry the Rider insertion at C, a novel single-base-pair substitution in the first exon leading to a deleterious effect on protein function is seen in Pruden's Purple. Included in the clade is Silvery Fir Tree, a

Table 1 Mean BRIX by yield (BY) values for the 18 heirloom cultivars with SE, number of plants sampled, and the final row containing the mean of all lines measured during the harvest time period.

Genotype	Mean BY	SE	*n*
ABC Potato Leaf	17.99	5.47	12.00
Aunt Ginny's Purple	2.34	0.68	10.00
Bloody Butcher	21.78	8.08	11.00
Blue Ridge Mountain	10.79	3.33	5.00
Brandywine	23.40	10.42	12.00
Burwood's Prize	15.79	7.62	7.00
Depp's Pink Firefly	12.35	7.79	3.00
Glacier	24.39	11.42	6.00
Green Pearl	12.51	11.35	3.00
Jerry's German Pink	16.62	4.07	8.00
Marianna's Peace	6.01	2.88	6.00
Matina	9.78	5.04	6.00
Prudens Purple	24.59	5.83	11.00
Rockingham	8.53	3.31	9.00
Schelicauski	10.39	2.35	7.00
Soldacki	6.82	2.22	7.00
Stupice	37.86	16.93	9.00
Valena Pink	14.84	5.09	11.00
Weeks 17–23 (mean)	16.39	1.91	143.00

nonPotato Leaf heirloom with very distinct leaf morphology. These cultivars come from a similar region of eastern Europe (Goldman, 2008), and our WGS phylogeny supports a region-specific breeding history. The relationships between the Potato Leaf and nonPotato Leaf heirlooms are not well resolved in our WGS-based phylogeny, probably as a result of close relationships between the cultivars and interbreeding. To identify any breeding history specifically related to the Potato Leaf Morph, we performed PhyloNetworks analysis using the WGS SNPs (Fig. 5c; Solís-Lemus *et al.*, 2017). We identified four hybridization events, relating to *C* mutants (blue squares in Fig. 5c). It is noteworthy that Prudens Purple with a unique mutation at the *C-locus* is not part of this series of hybridization events (Fig. 5c). These hybridization events suggest a breeding effort for desirable traits associated with this morphology. In addition we also analyzed chromosomes 1, 6 and 12 and found unique hybridizations for all of these chromosomes (Fig. S7). These data suggest that analyzing a much larger group of tomato cultivars for hybridization history could be very fruitful.

Partial least-squares path modeling

When doing large-scale field studies, it is difficult to understand how all the collected data points relate to each other, and what the causative relationships are (Granier and Vile, 2014). We performed several key correlation tests between measured traits (Fig. S8), but to test all traits we would need to perform 91 independent correlation tests. As such, to decipher how all the physiological and morphological traits measured related to each other, we performed PLS-PM using SmartPLS3.0 (Ringle *et al.*, 2015), which gives weighted causative paths with bootstrapping for confidence and significance values. In PLS-PM, each LV (such as leaf shape) is a composite value of its associated MVs (determined through correlations) and forms an outer model (Table S6). The inner model consists of the connections between LVs, with R^2 values indicating the degree to which each endogenous LV is described by the connections to it (Table S7). Here, the only exogenous LV is leaf shape, as it has only its associated MVs and is descriptive of other LVs. Some LVs (photosynthesis) are described by other LVs within the model (such as gas exchange, and light input in the case of photosynthesis). When the value of a causative LV (such as leaf shape) increases, the corresponding connected LVs change in accordance with their relationship with the causative variable. Similarly, in the outer model, changes in MVs reflect a change in their LV, and thus connect the outer model with MVs to the inner model of LVs. For instance, the MV PC3 has a negative correlation with the LV leaf shape (Table S6; Fig. S9), so that as the value of PC3 decreases, it reflects as a corresponding increase in LV leaf shape (Fig. S9). This change is represented as an increase in the roundness of the leaf. This then corresponds to a positive change in yield (LV), which is in turn a reflection of fruit biomass (MV) (Fig. S9).

The model indicates that photosynthesis has a strong positive influence on both fruit BRIX and vegetative biomass but has a negative impact on fruit yield. As photosynthetic rates increase (along with light capture and gas exchange), fruit BRIX increases, but at the sacrifice of yield, an inverse relationship which has long been known (Fig. 6; Eshed and Zamir, 1995; Zanor *et al.*, 2009; Chitwood *et al.*, 2014; Fridman *et al.*, 2000; Osorio *et al.*, 2014; Lytovchenko *et al.*, 2011). Leaf shape has a negative relationship with vegetative biomass, which corresponds to the decreased leaf complexity with the Potato Leaf Morph (Figs 5, 6, S3). However, leaf shape has a strong positive influence on both fruit BRIX and yield (Fig. 6), suggesting that leaf shape influences fruit quality as seen previously by Chitwood *et al.* (2014). The effect of leaf shape on fruit quality does not work through leaf sugar, as this correlation was not significant. Our leaf sugar measurements were completed in the glasshouse, owing to the complexity of the chemical analyses required, and as such the model was tested without leaf sugar. No significant causative relationship changes occurred in the model upon omitting the leaf sugar values. While our work does not implicitly study mechanisms, the negative relationship between leaf sugar and fruit BRIX is of interest, and may provide some avenues for future research into the mechanisms underlying impact of leaf shapes on fruit quality in tomato.

Fig. 6(b) displays the effect of each trait on the overall output of the plants (fruit BRIX, yield, and vegetative biomass). Leaf shape has no strong contribution to vegetative biomass. Although shape shows a negative relationship with biomass, this influence is minimal when compared with photosynthesis (Fig. 6b). However, leaf shape shows the largest influence on both yield and fruit BRIX, with photosynthesis second, and is the only positive contributor to yield (Fig. 6b). This positive correlation is from rounder, Potato Leaf Morph-like leaves, while narrower leaves have the opposite effect (Fig. 6a) based on the PC contributions to leaf shape. The negative effect of photosynthesis on tomato fruit yield and the strong contribution of leaf shape to yield and BRIX are novel findings that run counter to the interpretation of

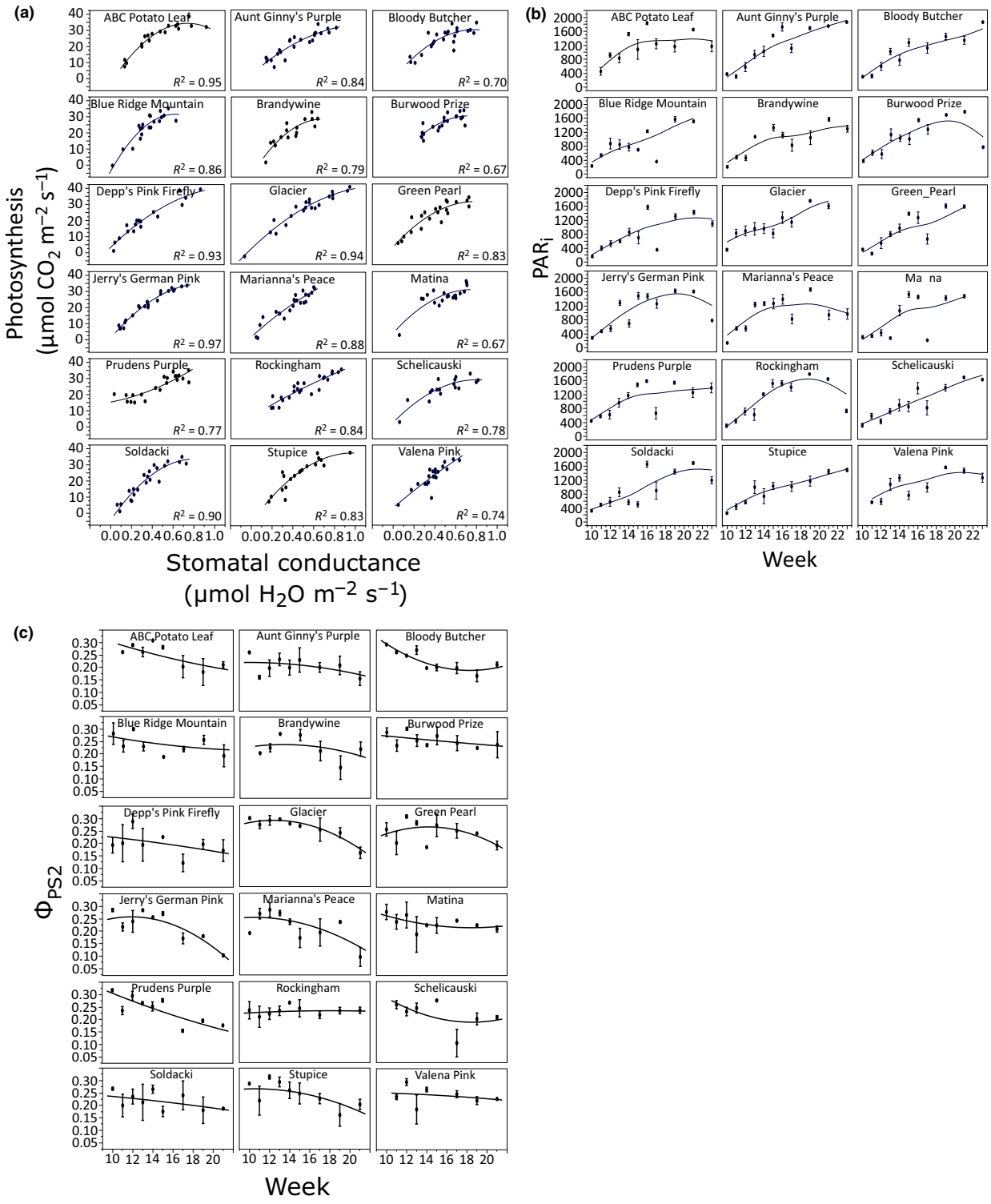

Fig. 3 Li-Cor LI-6400XT and LQA 2857 measurements (Li-Cor, Lincoln, NE, USA). (a) Photosynthetic rate (*A*) and stomatal conductance (g_{st}) were measured using the LI-6400XT for two leaves per plant (bottom and top of plant). The correlation between *A* and g_{st} is shown, with the majority of lines having a logarithmic curve. Rockingham and Valena Pink have a more linear relationship, and Prudens Purple shows a near-exponential increase. (b) Intercepted photosynthetically active radiation (PAR_i) was measured using the LQA-2857 and (LI-190R quantum sensor) by placing the line sensor beneath the plant and the quantum sensor above and taking four measurements per plant and averaging the result. All plants show a steady increase in PAR_i over time, corresponding to increased vegetative biomass and leaf area. (ϕ_{PS2} was measured with the LI-6400XT (6400-40 fluorescence head). Values are the mean measurements over time and error bars represent ± SE. (c) The overall ϕ_{PS2} decreases over time, with the only exception being Rockingham, which remains stable over the entire growing season.

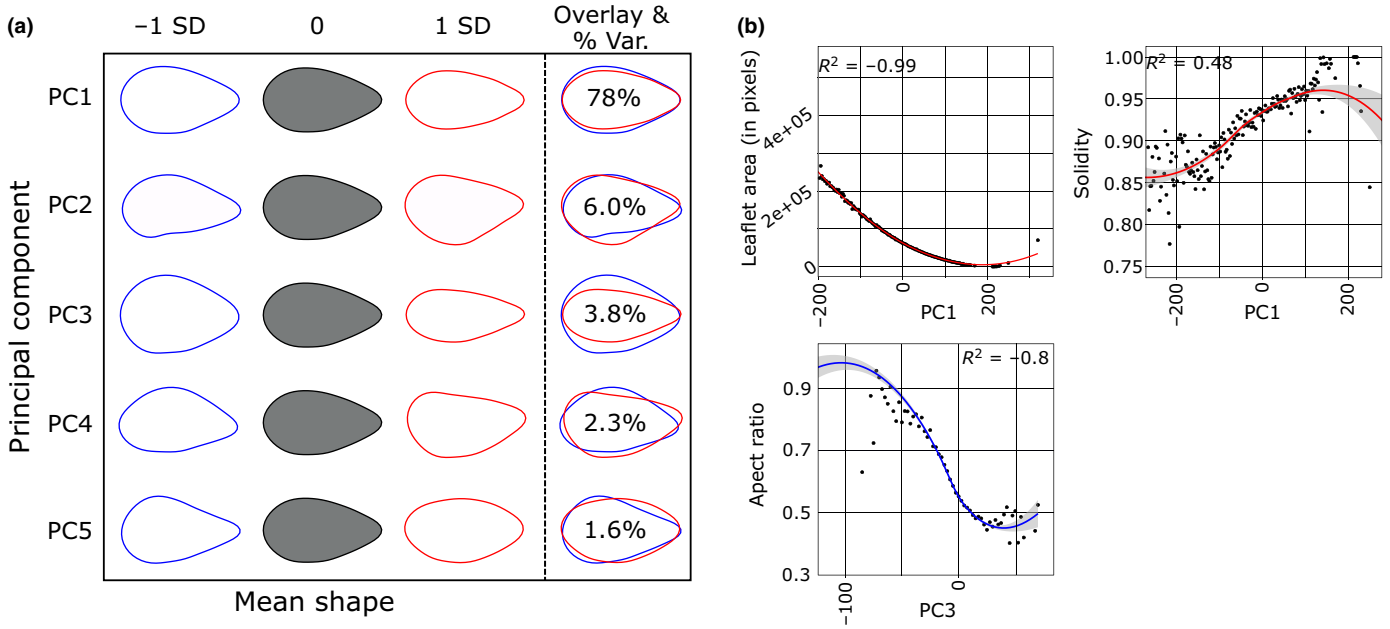

Fig. 4 Leaf shape analysis of heirloom tomatoes. (a) The calculated principal component (PC) values for 3733 leaflets are shown. PC1, PC3 and PC5 account for 83.4% of all variance found in the leaves, with PC1 having 78% of the variance. (b) Correlations of PC values to traditional measurements. PC1 correlates with leaflet size ($R^2 = 0.99$) and solidity ($R^2 = 0.48$), which is an indicator of how serrated or lobed a leaflet is. Higher values of solidity indicate a smoother, less lobed leaflet. Size is represented here in total number of pixels for correlation purposes. PC5 does not have a traditional shape measure associated with it, but does represent the tip to base ratio of the leaflet. PC2 and PC4 are inversions of the lateral primary leaflets which are mirror images of each other represented by these PCs. Red lines indicate correlations with PC1 and the blue line correlation with PC3.

fruit quality improvement, as increased photoassimilate should result in more available sucrose to stronger sinks such as fruit (Osorio *et al.*, 2014).

To test the model performance we used PLSPREDICT on the entire heirloom dataset used to build the structural model. Table S8 shows the mean absolute percentage error (MAPE) and Q^2 value for the complete model. We also used part of the dataset that included ABC Potato Leaf and Aunt Ginny's Purple in a similar analysis (Table S8). The complete model has *c.* 20–30% error for each LV, which is expected given the diversity of genotypes in the dataset, with fruit weight giving the highest MAPE, at 93.2% (Table S8). The Q^2 value for most variables is positive and shows that they have relevance in the predictive performance, with the exception of leaf sugar, which is slightly negative (Table S8). In the case of ABC Potato Leaf and Aunt Ginny's Purple, two lines selected randomly to test the model on individual cultivars, a significant increase in Q^2 and decrease in MAPE is seen for all LVs except leaf sugar (Table S8). This indicates that the model is substantially stronger in predictive performance for individual cultivars, but also predicts well with the complete model.

To evaluate the predictive performance of our model on additional datasets, we used data from two other cultivars grown in the same field, M82 and Lukullus, that were not used to construct the model. PLSPREDICT was used in SMARTPLS 3.0, along with the structural model constructed using the heirloom cultivars, to test the model performance by use of training sets and hold out samples, both taken from the M82/Lukullus dataset. By using the leaf shape PC values, we were able to compare the predicted mean values for the remaining MVs, or the predicted measured values, against the actual measured values and evaluate the relative performance of the model. Tables 2 and 3 show the results for M82 and Lukullus, respectively. PC values for leaf shape are not included as they are input variables and used for predicting the other values. For M82 the predicted median values compared with the actual median values showed under 1% difference for all except leaf complexity, which had a percentage difference of −8.42% (Table 2). This indicates that the model was underpredicting the leaf complexity of M82 by *c.* 8%. Lukullus-predicted values were also under 1% different, except for leaf complexity and stomatal conductance which varied by −2.56% and 1.31%, respectively (Table 3). In addition to the predicted values PLSPREDICT also tests the model performance and reports the root mean square error, mean absolute error, and MAPE for each of the MVs tested (Tables 2, 3). The MAPE shows the accuracy of the predictions, with lower percentages representing better performance. Leaf complexity for both cultivars showed the largest MAPE values, 201.2% and 26.5% in M82 and Lukullus, respectively (Tables 2, 3). The M82 MAPE indicates that the model does not predict leaf complexity well for mid-level complexities such as 18 but does improve at high-end leaf complexities near 40 (Tables 2, 3). Most heirloom cultivars had low leaf complexities (Fig. S10), potentially explaining the poor performance in predicting leaf complexity for M82. Contrary to previous findings (Chitwood *et al.*, 2014), we found that leaf complexity does not impact yield or BRIX, and only impacts vegetative biomass, so this inaccuracy

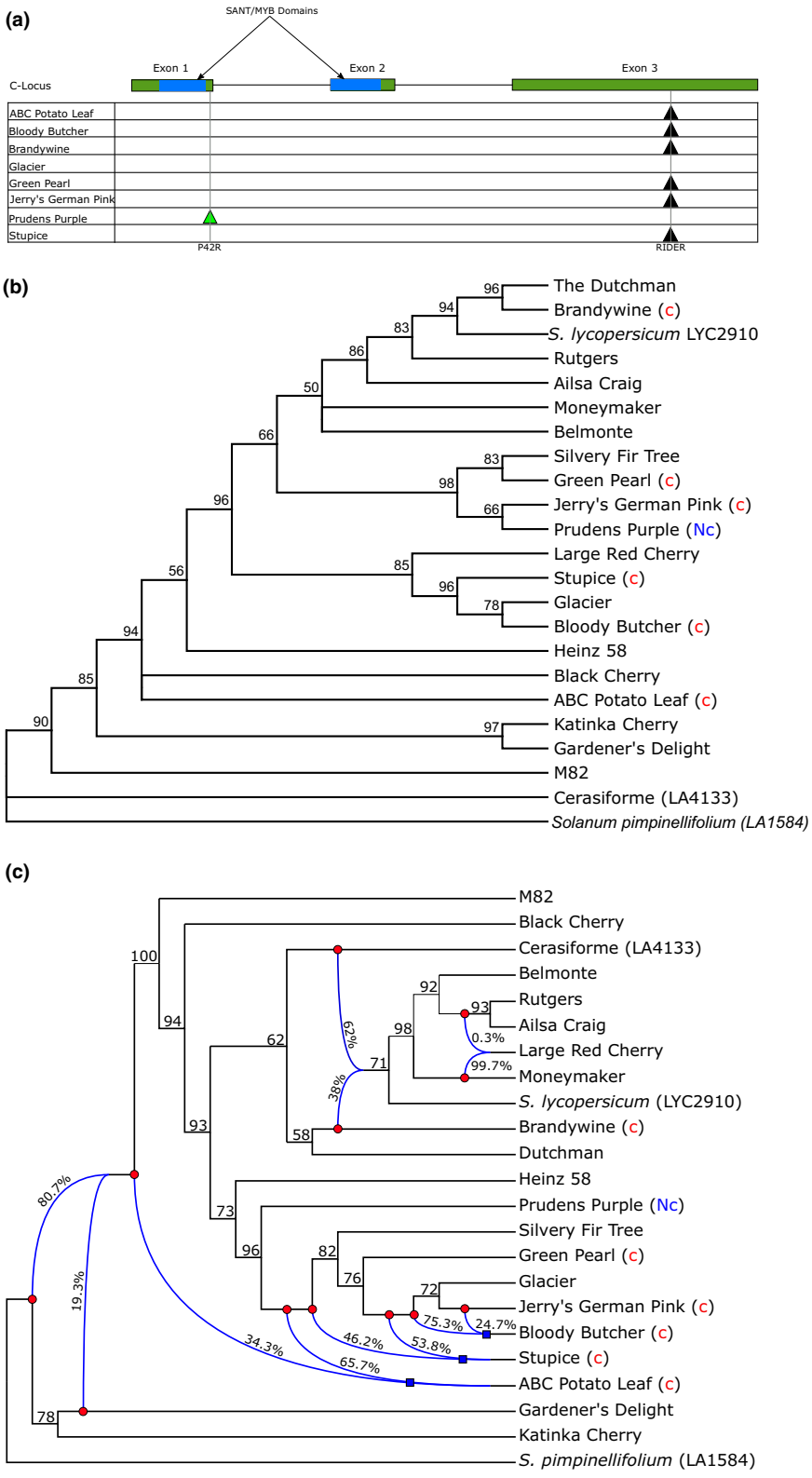

Fig. 5 *C-locus* mapping, whole-genome sequencing (WGS) and PhyloNetworks based network. (a) Eight heirloom cultivars were sequenced using WGS and the mutations found in the *C-locus* are shown. Black arrows indicate the position of the Rider transposable element insert, and the green arrow indicates the single-base-pair mutation and its resultant amino acid swap. Glacier had no mutations found in the *C-locus*. (b) Using variants from the WGS sequencing, a phylogeny was generated for the eight heirloom cultivars and 14 additional lines. M82 was sequenced by WGS (by us), while the remaining sequences were obtained from the 150 Genomes Project (Aflitos *et al.*, 2014). (c) PhyloNetworks analysis of whole genome single nucleotide polymorphisms (SNPs) shows six hybridization events among these 23 cultivars. Percentage numbers represent the gene flow from each 'parent' cultivar to the hybridization event. Red 'c's represent *C-locus* mutants, which have the Potato Leaf Morph, while 'Nc' represents the novel *C-locus* mutation found here. Bootstrap values > 50% (except for for the hybridization events) are shown.

would only impact vegetative output predictions by the model. Lukullus has indeterminate growth like the heirlooms analyzed here, but M82 is determinate; however, the predictive accuracy of the model was still good, indicating its usefulness in assessing field performance of other tomato cultivars.

Discussion

The primary focus of crop improvement has been on fruit traits (sink) and photosynthesis (source), with some studies focusing on how sugars are moved from source to sink. Despite heirloom

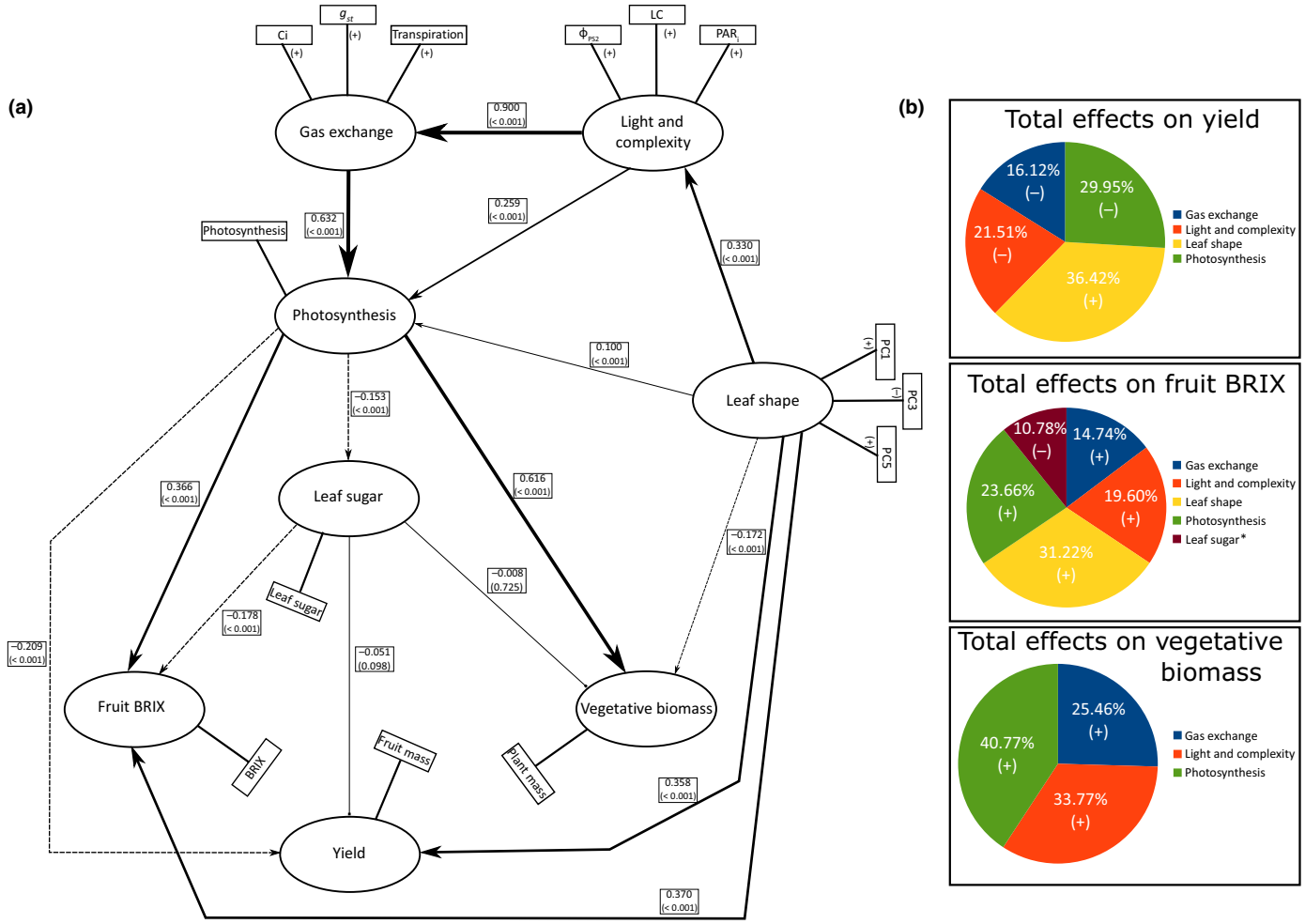

Fig. 6 Partial least-squares path modeling (PLS-PM) of all collected physiological and morphological data. (a) The finalized version of the PLS-PM using SmartPLS3.0 is shown. Traits within the circles are latent variables (LVs) which represent the measurements in rectangles, the manifest variables (MVs). The (+) and (−) next to each MV represent its relationship to its LV. PC1 and PC5 are positively correlated with the LV leaf shape, while PC3 is negatively correlated. This corresponds to a rounder leaflet shape (−PC3, +PC5). Solid lines indicate a positive correlation, and dashed lines indicate a negative correlation between LVs. The size of the arrow indicates path weight, which is listed next to the path along with the *P*-value. Blunted lines indicate nonsignificant connections. Photosynthesis has a strong positive correlation with fruit BRIX (soluble solids) and vegetative biomass but has a negative correlation with yield. Leaf sugar content has no significant impact on vegetative biomass or yield, but is negatively correlated with fruit BRIX, indicating that a lower leaf sugar content corresponds to an increased fruit BRIX. Leaf shape has a positive correlation with both yield and fruit BRIX, indicating that leaf shape may play a role in distribution of photosynthate. (b) Total effects of different LVs on the outputs of the PLS-PM: yield, fruit BRIX, and vegetative biomass. For both yield and fruit BRIX, leaf shape has the strongest effect, followed by photosynthesis. (+) indicates a positive effect while (−) indicates a negative effect. Percentages are the proportion of path weights contributing to each output.

varieties with the Potato Leaf Morph being prized for fruit quality by the gardening community, vegetative traits such as leaf shape have been relatively ignored in breeding efforts. In this study we investigated the role of leaf shape on fruit quality by measuring both input traits (such as photosynthesis, leaf shape and leaf complexity) and output traits (such as yield, BRIX, and vegetative biomass) for 18 heirloom cultivars. All these cultivars were classified as Potato Leaf, but varied greatly in their leaf shapes, development, and fruit quality (Figs 1, 2, 4). We found that these lines do not vary significantly in overall photosynthetic capacity, or their usage of light when available (Fig. 3), suggesting that the variation in BY (Fig. 1) among these cultivars was not a result of improved/decreased photosynthetic capacity. While our measurements for photosynthesis do not show significant difference when PAR is available, the PAR_i differed between cultivars based on their growth patterns (Figs 2b, 3b). All cultivars exceeded 1200 µmols m^{-2} s^{-1} of PAR_i but varied in the later weeks between 1200 and 2000 µmols m^{-2} s^{-1} (Fig. 3b).

Combining multiple complex physiological and morphological measurements into informative relationships has proven difficult and has limited our understanding of how these different traits impact each other (Granier and Vile, 2014). Focusing on any one part, such as photosynthesis or fruit sink strength, while providing improvements (Zanor *et al.*, 2009), occurs at the expense of a comprehensive understanding of the overall relationships between these traits. Analyzing the individual PCs revealed significant differences in leaf shape among the heirloom cultivars, with several having stronger Potato Leaf Morphs (Figs 4, S5) and

Table 2 Predicted and actual median values for M82 of manifest variables in the partial least-squares path model, with error rates for model accuracy.

MV	Predicted median	Actual median	Percentage difference	RMSE	MAE	MAPE
Stomatal conductance	0.152	0.152	0.14%	0.014	0.010	6.92
Internal CO_2	72.098	71.623	0.66%	7.038	5.897	8.21
Transpiration (mmol)	2.720	2.736	−0.60%	0.231	0.189	7.07
ϕ_{PS2}	0.071	0.070	0.96%	0.006	0.004	6.08
PAR_i	952.922	956.761	−0.40%	15.947	12.462	1.31
Photosynthesis	7.575	7.531	0.59%	0.607	0.457	6.12
Leaf sugar	2.723	2.725	−0.08%	0.092	0.074	2.75
Complexity	16.310	17.809	−8.42%	9.719	7.806	201.28
Plant weight	2.864	2.888	−0.82%	0.427	0.339	12.25
Fruit weight	1.246	1.236	0.76%	0.209	0.163	13.21
BRIX	4.134	4.136	−0.04%	0.038	0.032	0.76

RMSE, root mean square error; MAE, mean absolute error; MAPE, mean absolute percentage error; ϕ_{PS2}, amount of photons entering photosystem II; PAR_i, amount of intercepted photosynthetically active radiation.

Table 3 Predicted and actual median values for Lukullus of manifest variables in the partial least-squares path model, with error rates for model performance evaluation.

MV	Predicted median	Actual median	Percentage difference	RMSE	MAE	MAPE
Stomatal conductance	0.116	0.114	1.31%	0.008	0.006	5.42
Internal CO_2	77.646	77.348	0.39%	6.193	5.008	6.46
Transpiration (mmol)	2.207	2.228	−0.91%	0.175	0.139	6.45
ϕ_{PS2}	0.048	0.048	0.68%	0.003	0.003	5.20
PAR_i	1166.423	1169.253	−0.24%	71.522	56.135	4.83
Photosynthesis	5.142	5.154	−0.24%	0.376	0.308	6.05
Leaf sugar	2.111	2.110	0.05%	0.051	0.042	2.01
Complexity	41.737	42.834	−2.56%	12.012	9.547	26.50
Plant weight	2.092	2.088	0.20%	0.223	0.178	8.72
Fruit weight	1.466	1.462	0.28%	0.258	0.203	14.20
BRIX	4.407	4.406	0.02%	0.069	0.055	1.25

RMSE, root mean square error; MAE, mean absolute error; MAPE, mean absolute percentage error; ϕ_{PS2}, amount of photons entering photosystem II; PAR_i, amount of intercepted photosynthetically active radiation.

higher BY values (Fig. 1), with some correlation between these traits. Potential epidermal shape changes that could arise from leaf shape changes and that could influence yield would relate to stomatal number. Our Li-Cor data measured stomatal conductance and showed no significant differences (Table S3). A previous study in 2002 analyzed several tomato cultivars developmentally and histologically (Kessler *et al.* 2002) and found no real differences between these cultivars. This and another study in 2010 (Kang & Sinha, 2010) suggest that there are no gross anatomical differences between these tomato cultivars.

We used PLS-PM to combine all these measured traits, using the modeled final harvest data as input to find causative relationships (Fig. 6a). Strong relationships among gas exchange, light, and photosynthesis (photosynthesis per plant) were expected, along with a strong positive effect of photosynthesis on vegetative biomass (Fig. 6a,b). Photosynthesis has a strong positive effect on fruit BRIX, both directly and indirectly (Fig. 6a). Increased photosynthesis results in lowered leaf sugar content, and a concomitant increase in fruit BRIX. It is possible that increased sugar production from photosynthesis results in higher rates of transport of sugars out of the leaves and into sinks. The mechanisms that regulate source–sink relations and sugar distribution are still not fully understood on a whole-plant physiological level (Osorio *et al.*, 2014); however, based on our model, increased photosynthesis negatively impacts total yield (Figs 6a,b, 7). While photosynthesis does lead to increased sugar production and is shown in our model to drive higher sugar content within existing fruit, it does not provide a means to increase yield. Leaf shape, specifically rounder, less lobed leaves, has a positive effect on both fruit BRIX and yield (Figs 6a,b, 7). Of all the factors measured here, only leaf shape positively influenced yield, with other paths having negative influences (Fig. 6b). Rounder leaves still drive slightly increased photosynthesis indicated by the thin arrow (Fig. 7a), which results in increased fruit BRIX. This path should also result in decreased yield. However, leaf shape has a strong positive and direct correlation with yield that overcomes the negative impact of photosynthesis and leads to increased yield as well as BRIX (Fig. 7a). Conversely, with narrow leaflets there is a small negative impact on photosynthesis which should result in increased yield, but narrow leaves have a direct negative impact on yield which is stronger than the photosynthetic pathway (Fig. 7b). The strong causative relationship among leaf shape, fruit BRIX, and yield suggests that leaf shape impacts both high

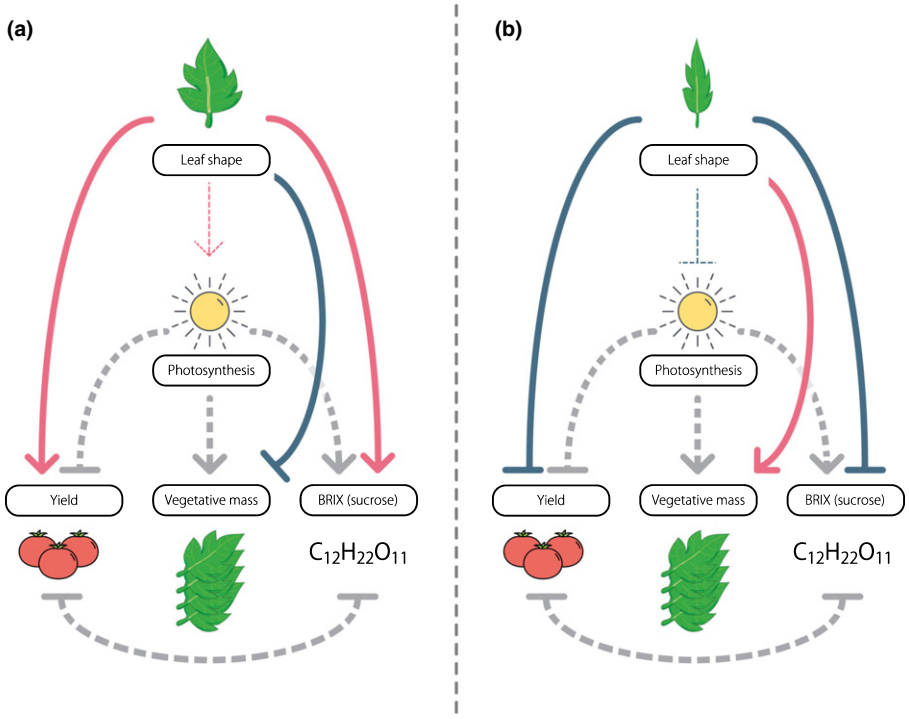

Fig. 7 Composite model for leaf shape effects on fruit quality. The model was derived from the partial least-squares path modeling (PLS-PM) analysis. (a) Effects of round leaves on fruit quality and photosynthesis. (b) Effects of narrow leaves on fruit quality and photosynthesis. Red lines indicate a positive interaction while blue lines indicate a negative interaction. Gray dashed lines indicate that the relationship does not change between the two leaf shapes. Colored dashed lines indicate a significant but weak relationship between the two connected traits.

fruit BRIX and increased number of fruits, probably by modulating sugar distribution, therefore bypassing the direct impacts of photosynthesis itself (Fig. 7). How leaf shape affects this distribution is unclear, as it does not act directly through leaf sugar content, or through strong regulation of photosynthesis to improve yield (Fig. 7). A recent study looked at the diversity of leaf shape in sweet potato (*Ipomea batatas*; Gupta *et al.*, 2020). Any correlations between leaf shape and yield traits in this species would be of interest and help to establish general principles.

The whole-genome phylogenetic analysis of 23 tomato cultivars showed many of the Potato Leaf Morph cultivars were closely related to each other, with the exception of Brandywine, though it did not show the origin of the *C-locus* mutation (Fig. S7a). To address this and identify if this morphology was selected for in breeding, we performed PhyloNetworks analysis (Solís-Lemus *et al.*, 2017). This analysis showed several hybridizations between Potato Leaf Morph and nonPotato Leaf Morph cultivars, and probably a unique incidence of the *C-locus* mutation in Prudens Purple (Fig. 5b). PhyloNetworks analysis of chromosome 1, 6, and 12 specific common SNPs each showed unique hybridization events, distinct from those seen in the WGS analysis (Fig. S7). The PhyloNetworks analysis suggests multiple hybridization events with Potato Leaf Morph-containing cultivars. Potato leaf cultivars have been suggested to increase disease resistance compared with regular leaf varieties (Male, 1999) and may have been selected for this reason or for other as-yet-unknown benefits present.

We have shown that leaf shape strongly impacts the overall fruit quality in tomato, with rounder, less lobed leaves giving rise to higher yield and higher fruit BRIX. Photosynthesis, surprisingly, has a negative impact on yield while still positively contributing to fruit BRIX. Using data from cultivars not included in making our path model, we also showed that the model has a strong predictive performance for linking leaf shape to BY and could be used to potentially predict the outputs of a cultivar using leaf shape data (Tables 2, 3). Our work shows the importance of leaf shape to yield and BRIX across a wide array of genetic backgrounds, implicating leaf morphology in playing a significant and previously unidentified role in tomato fruit quality.

Acknowledgements

We acknowledge help from Mary Lee, Kirsten Brand, Gabriel Luis Moreira, Kristina Khuu, Eduardo Ramirez, Divya Kumaria, Jason Kao, and Surbhi Chopla in field sample, leaf and BRIX data collection, whole-plant imaging, and seed collection. We also thank Siyu Li, Diane Beckles, Donnelly West, and Jessica Budke for advice on images and data analysis. Support from the USDA-NIFA (grant no. 2014-67013-21700) is gratefully acknowledged.

Author contributions

NRS, JNM and DHC conceived of the research idea; JNM gave advice on research and analysis methods; ZC and HN did the whole-genome sequencing; AMF helped design the sugar assays and collect data; KZ organized the field study, and helped with data collection and all leaf analyses; SDR led the research effort, collected and analyzed data and wrote the paper with input from all authors; NRS supervised the research and helped to write the manuscript.

ORCID

Daniel H. Chitwood https://orcid.org/0000-0003-4875-1447
Amber M. Flores https://orcid.org/0000-0002-4050-2174
Julin N. Maloof https://orcid.org/0000-0002-9623-2599
Hokuto Nakayama https://orcid.org/0000-0002-5724-4861
Steven D. Rowland https://orcid.org/0000-0002-8633-1710
Neelima R. Sinha https://orcid.org/0000-0002-1494-7065
Kristina Zumstein https://orcid.org/0000-0001-9516-8081

References

Aflitos S, Schijlen E, de Jong H, de Ridder D, Smit S, Finkers R, Wang J, Zhang G, Li N, Mao L *et al.* 2014. Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *The Plant Journal* **80**: 136–148.

Bai Y, Lindhout P. 2007. Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Annals of Botany* **100**: 1085–1094.

Barberán A, Ramirez KS, Leff JW, Bradford MA, Wall DH, Fierer N. 2014. Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecology Letters* **17**: 794–802.

Barker G. 2006. *The agricultural revolution in prehistory: why did foragers become farmers?* New York, NY, USA: Oxford University Press.

Bauchet G, Causse M. 2012. Genetic diversity in tomato (*Solanum lycopersicum*) and its wild relatives. In: Muhammed A, ed. *Genetic diversity in plants, vol. 133*. Berlin, Germany: Springer Science & Business Media, 162.

Baxter CJ, Carrari F, Bauke A, Overy S, Hill SA, Quick PW, Fernie AR, Sweetlove LJ. 2005. Fruit carbohydrate metabolism in an introgression line of tomato with increased fruit soluble solids. *Plant and Cell Physiology* **46**: 425–437.

Bergougnoux V. 2014. The history of tomato: from domestication to biopharming. *Biotechnology Advances* **32**: 170–189.

Bonhomme V, Picq S, Gaucherel C, Claude J. 2014. Momocs: outline analysis using R. *Journal of Statistical Software* **56**: 1–24.

Bishop KA, Lemonnier P, Quebedeaux JC, Montes CM, Leakey ADB, Ainsworth EA. 2018. Similar photosynthetic response to elevated carbon dioxide concentration in species with different phloem loading strategies. *Photosynthesis Research* **137**: 453–464.

Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Díez MJ, Francis D, Causse M, van der Knaap E, Cañizares J. 2015. Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics* **16**: 257.

Busch BL, Schmitz G, Rossmann S, Piron F, Ding J, Bendahmane A, Theres K. 2011. Shoot branching and leaf dissection in tomato are regulated by homologous gene modules. *Plant Cell* **23**: 3595–3609.

Chitwood DH, Ranjan A, Kumar R, Ichihashi Y, Zumstein K, Headland LR, Ostria-Gallardo E, Aguilar-Martinez JA, Bush S, Carriedo L *et al.* 2014. Resolving distinct genetic regulators of tomato leaf shape within a heteroblastic and ontogenetic context. *Plant Cell* **26**: 3616–3629.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **7**: e46688.

Choi Y. 2012. A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. In: *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine - BCB' 12*. New York, NY, USA: ACM Press, 414–417.

Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**: 2745–2747.

Dengler N, Kang J. 2001. Vascular patterning and leaf shape. *Current Opinion in Plant Biology* **4**: 50–56.

Easlon HM, Bloom AJ. 2014. Easy leaf area: automated digital image analysis for rapid and accurate measurement of leaf area. *Applications in Plant Sciences* **2**: 1400033.

Eshed Y, Zamir D. 1995. 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.

Fridman E, Pleban T, Zamir D. 2000. A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proceedings of the National Academy of Sciences, USA* **97**: 4718–4723.

Gilbert ME, Holbrook NM, Zwieniecki MA, Sadok W, Sinclair TR. 2011. Field confirmation of genetic variation in soybean transpiration response to vapor pressure deficit and photosynthetic compensation. *Field Crops Research* **124**: 85–92.

Goldman A. 2008. *The Heirloom tomato*. New York, NY, USA: Bloomsbury.

Grandillo S, Zamir D, Tanksley SD. 1999. Genetic improvement of processing tomatoes: A 20 years perspective. *Euphytica* **110**: 85–97.

Granier C, Vile D. 2014. Phenotyping and beyond: Modelling the relationships between traits. *Current Opinion in Plant Biology* **18**: 96–102.

Gupta S, Rosenthal DM, Stinchcombe JR, Baucom RS. 2020. The remarkable morphological diversity of leaf shape in sweetpotato (*Ipomoea batatas*): the influence of genetics, environment, and G×E. *New Phytologist* **225**: 2183–2195.

Huson DH, Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology* **61**: 1061–1067.

Jiang N, Gao D, Xiao H, van der Knaap E. 2009. Genome organization of the tomato sun locus and characterization of the unusual retrotransposon Rider. *The Plant Journal* **60**: 181–193.

Jiang N, Visa S, Wu S, Van Der Knaap E. 2012. Rider transposon insertion and phenotypic change in tomato. *Topics in Current Genetics* **24**: 297–312.

Kang J, Sinha NR. 2010. Leaflet initiation is temporally and spatially separated in simple and complex tomato (*Solanum lycopersicum*) leaf mutants: a developmental analysis. *Botany* **88**: 710–724.

Kessler S, Kim M, Pham T, Weber N, Sinha N. 2002. Mutations altering leaf morphology in tomato. *International Journal of Plant Sciences* **162**: 475–492.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**: 1870–1874.

Lee T-H, Guo H, Wang X, Kim C, Paterson AH. 2014. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* **15**: 162.

Lemoine R, La Camera S, Atanassova R, Dédaldéchamp F, Allario T, Pourtau N, Bonnemain J-L, Laloi M, Coutos-Thévenot P, Maurousset L *et al.* 2013. Source-to-sink transport of sugar and regulation by environmental factors. *Frontiers in Plant Science* **4**: 272.

Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, Zhang Z, Lun Y, Li S, Wang X *et al.* 2014. Genomic analyses provide insights into the history of tomato breeding. *Nature Genetics* **46**: 1220–1226.

Lytovchenko A, Eickmeier I, Pons C, Osorio S, Szecowka M, Lehmberg K, Arrivault S, Tohge T, Pineda B, Anton MT *et al.* 2011. Tomato fruit photosynthesis is seemingly unimportant in primary metabolism and ripening but plays a considerable role in seed development. *Plant Physiology* **157**: 1650–1663.

Male CJ. 1999. *100 Heirloom tomatoes for the American garden*. New York, NY, USA: Workman Publishing.

Matiu M, Ankerst DP, Menzel A. 2017. Interactions between temperature and drought in global and regional crop yield variability during 1961–2014. *PLoS ONE* **12**: e0178339.

Osorio S, Ruan Y-L, Fernie AR. 2014. An update on source-to-sink carbon partitioning in tomato. *Frontiers in Plant Science* **5**: 516.

Ostria-Gallardo E, Ranjan A, Ichihashi Y, Corcuera LJ, Sinha NR. 2018. Decoding the gene coexpression network underlying the ability of *Gevuina avellana* to live in diverse light conditions. *New Phytologist* **220**: 278–287.

Passam HC, Karapanos IC, Bebeli PJ, Savvas D. 2007. A review of recent research on tomato nutrition, breeding and post-harvest technology with reference to fruit quality. *The European Journal of Plant Science and Biotechnology* **1**: 1–21.

Pnueli L, Carmel-Goren L, Hareven D, Gutfinger T, Alvarez J, Ganal M, Zamir D, Lifschitz E. 1998. The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. *Development* **125**: 1979–1989.

Ringle CM, Wende S, Becker J-M. 2015. *SmartPLS 3*. Boenningstedt, Germany: SmartPLS GmbH [WWW document] URL http://www.smartpls.com.

RodrÍGuez-Burruezo S, Prohens J, RosellÓ J, Nuez F. 2005. "Heirloom" varieties as sources of variation for the improvement of fruit quality in greenhouse-grown tomatoes. *The Journal of Horticultural Science and Biotechnology* **80**: 453–460.

Sarlikioti V, de Visser PHB, Buck-Sorlin GH, Marcelis LFM. 2011. How plant architecture affects light absorption and photosynthesis in tomato: towards an ideotype for plant architecture using a functional–structural plant model. *Annals of Botany* **108**: 1065–1073.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**: 671–675.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B *et al.* 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**: 676–682.

Sim SC, Durstewitz G, Plieske J, Wieseke R, Ganal MW, van Deynze A, Hamilton JP, Buell CR, Causse M, Wijeratne S *et al.* 2012. Development of a large snp genotyping array and generation of high-density genetic maps in tomato. *PLoS ONE* **7**: e40563.

Solís-Lemus C, Bastide P, Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. *Molecular Biology and Evolution* **34**: 3292–3298.

Stenz NWM, Larget B, Baum DA, Ané C. 2015. Exploring tree-like and non-tree-like patterns using genome sequences: an example using the inbreeding plant species *Arabidopsis thaliana* (L.) heynh. *Systematic Biology* **64**: 809–823.

Tieman D, Zhu G, Resende MFR, Lin T, Nguyen C, Bies D, Rambla JL, Beltran KSO, Taylor M, Zhang B *et al.* 2017. A chemical genetic roadmap to improved tomato flavor. *Science* **355**: 391–394.

Townsley BT, Covington MF, Ichihashi Y, Zumstein K, Sinha NR. 2015. BrAD-seq: Breath Adapter Directional sequencing: a streamlined, ultra-simple and fast library preparation protocol for strand specific mRNA library construction. *Frontiers in Plant Science* **6**: 366.

Zanor MI, Osorio S, Nunes-Nesi A, Carrari F, Lohse M, Usadel B, Kühn C, Bleiss W, Giavalisco P, Willmitzer L *et al.* 2009. RNA interference of LIN5 in tomato confirms its role in controlling brix content, uncovers the influence of sugars on the levels of fruit hormones, and demonstrates the importance of sucrose cleavage for normal fruit development and fertility. *Plant Physiology* **150**: 1204–1218.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**: 3326–3328.

Zhu G, Wang S, Huang Z, Zhang S, Liao Q, Zhang C, Lin T, Qin M, Peng M, Yang C *et al.* 2018. Rewiring of the fruit metabolome in tomato breeding. *Cell* **172**: 249–261.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Representative leaf images for the 18 heirloom cultivars.

Fig. S2 Comparison of 2014 and 2015 field seasons.

Fig. S3 Yield and fruit BRIX whole season measurements.

Fig. S4 Whole-plant-adjusted A and g_{st}.

Fig. S5 Alignment of sequence overhangs in the third exon of the *C-locus* with the Rider transposable element sequence.

Fig. S6 Mean shape from eFourier and PCA analysis of eight heirloom cultivars.

Fig. S7 PhyloNetworks analysis and resultant networks from chromosome 1, 6 and 12 SNPs.

Fig. S8 Individual correlations for physiological traits not included in the PLS-Path Model.

Fig. S9 PLS-PM inner and outer model representation and their correlations.

Fig. S10 Mean leaf complexity of all 18 heirloom cultivars.

Table S1 Linear modeled field data used as input for the PLS-PM.

Table S2 Total number of SNPs and unique SNPs per cultivar. Total shared represents the SNPs shared among all 23 cultivars taken to SNPhylo for phylogenetic analysis.

Table S3 Stomatal conductance and statistical comparison of the 18 heirloom cultivars grown in the field.

Table S4 Vascular density of 18 heirloom cultivars ordered from the highest to the lowest fruit BRIX.

Table S5 PCA scores for the 3733 leaflets analyzed.

Table S6 Outer model loadings and significance.

Table S7 R2 values for each latent variable in the inner model.

Table S8 Predictive performance and model fit for heirloom cultivars in PLS-PM.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.