

2023

Evidence of fruit syndromes in the recently diverged wild tomato clade opens new possibilities for the study of fleshy fruit evolution

Jacob R. Barnett

Rovin Sharma

Gina Buonauro

Ian M. Gillis

Maryam Rashidzade


See next page for additional authors

Authors

Jacob R. Barnett, Rovin Sharma, Gina Buonauro, Ian M. Gillis, Maryam Rashidzade, and Ana L. Caicedo

RESEARCH ARTICLE

Evidence of fruit syndromes in the recently diverged wild tomato clade opens new possibilities for the study of fleshy fruit evolution

Jacob R. Barnett¹  | Rovin Sharma²  | Gina Buonauro²  | Ian M. Gillis³  |
Maryam Rashidzade³  | Ana L. Caicedo² 

¹Graduate Program in Organismic and Evolutionary Biology, University of Massachusetts Amherst, Massachusetts, USA

²Department of Biology, University of Massachusetts Amherst, Massachusetts, USA

³Graduate Program in Plant Biology, University of Massachusetts Amherst, Massachusetts, USA

Correspondence

Jacob R. Barnett, Graduate Program in Organismic and Evolutionary Biology, University of Massachusetts Amherst, MA 01003, USA.

Email: jbarnett@umass.edu

Ana L. Caicedo, Department of Biology, University of Massachusetts Amherst, MA 01003, USA.

Email: caicedo@umass.edu

Funding information

UMass Natural History Collections Grant; UMass Graduate School Dissertation Research Grant; Lotta Crabtree Fellowship in Production Agriculture; National Science Foundation, Grant/Award Number: 1564366

Societal Impact Statement

Fleshy fruits provide humans with many flavorful and nutritious crops. Understanding the diversity of these plants is fundamental to managing agriculture and food security in a changing world. This study surveyed fruit trait variation across species of tomato wild relatives and explored associations among color, size, shape, sugars, and acids. These wild tomato species native to South America can be interbred with the economically important cultivated tomato. Beyond its application to tomatoes, deepening our knowledge of how fruit traits evolve together is valuable to crop improvement efforts aimed at breeding more nutritious and appealing varieties of fruits.

Summary

- Fleshy fruits display a striking diversity of traits, many of which are important for agriculture. The evolutionary drivers of this variation are not well understood, and most studies have relied on variation found in the wild. Few studies have explored this question on a fine-grained scale with a group of recently diverged species while controlling for environmental effects.
- We developed the tomato clade as a novel system for fruit trait evolution research by presenting the first common garden-based systematic survey of variation and phylogenetic signal in color, nutrition, and morphology traits across all 13 species of tomato wild relatives (*Solanum* sect. *Lycopersicon*). We laid the groundwork for further testing of potential evolutionary drivers by assessing patterns of clustering and correlation among disperser-relevant fruit traits as well as historical climate variables.
- We found evidence of two distinct clusters of associated fruit traits defined by color, sugar type, and malic acid concentration. We also observed correlations between a fruit's external appearance and internal nutrient content that could function as honest signals to dispersers. Analyses of historical climate and soil

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Plants, People, Planet* published by John Wiley & Sons Ltd on behalf of New Phytologist Foundation.

variables revealed an association between red/orange/yellow fruits and high annual average temperature.

- Our results establish the tomato clade as a promising system for testing hypotheses on the drivers of divergence behind early-stage fleshy fruit evolution, particularly selective pressure from frugivores.

KEYWORDS

crop wild relatives, dispersal syndrome hypothesis, fleshy fruit evolution, phylogenetic signal, *Solanum lycopersicum* (tomato), trait covariation

1 | INTRODUCTION

Flowering plants produce an astounding variety of fruits that have long fascinated biologists, yet questions remain as to how this diversity has arisen. Of particular interest are fleshy fruits, which enable seed dispersal through animal consumers and also provide great value to human society as agricultural food crops. Research across many species has established that fruit traits do not vary independently. Instead, certain suites of traits tend to occur together (for example small size, bright colors, and high sugar content) and often involve correlations between external appearance and internal chemical composition (Nevo et al., 2019; Schaefer et al., 2014; Sinnott-Armstrong et al., 2020; Sinnott-Armstrong et al., 2023; Valido et al., 2011). Repeatedly observed associations of certain traits are often referred to as fruit or dispersal syndromes (Rojas et al., 2022; Valenta & Nevo, 2020; van der Pijl, 1969).

Decades of research have explored factors that may underlie these patterns in fleshy fruit trait covariation. Frugivorous animal preferences have long been considered an important evolutionary force, since plants better at attracting seed-dispersing fruit consumers improve their reproductive success (Brodie, 2017; Janson, 1983; Nevo et al., 2018; Ridley, 1930; Valenta et al., 2018; van der Pijl, 1969). A conflicting force may be damage-inflicting seed predators and pathogens, in which case fruits that can deter or defend against these pests have a reproductive advantage (Mack, 2000; Nevo et al., 2017; Whitehead et al., 2016). Adaptation to abiotic factors such as openness of habitat (Bolmgren & Eriksson, 2005) or climatic variables (Zhao et al., 2018) may also underlie some fruit trait diversity. Additionally, fruit phenotypes may be constrained by mechanical (Valenta et al., 2022), chemical (Whitehead et al., 2016), developmental (Nevo et al., 2020), or other non-adaptive limitations due to shared genetic underpinnings from phylogenetic relatedness (Jordano, 1995).

In recent years, our understanding of fruit trait evolution has been advanced by a number of studies (Nevo & Ayasse, 2019; Valenta & Nevo, 2020), including analyses of large global datasets (do Nascimento et al., 2020; Onstein et al., 2020; Sinnott-Armstrong et al., 2018; Sinnott-Armstrong et al., 2021), but certain gaps have yet to be addressed. Most previous work has examined large sets of distantly related plant species. Some results from these broad datasets (e.g., Jordano, 1995) suggested that congeneric species may have limited potential for fruit trait evolutionary response due to shared

ancestry, but the fine-grained studies on closely related species needed to test this hypothesis are lacking. Some studies focus on only one habitat, such as a cloud forest (e.g., Rojas et al., 2022), where all species have adapted to similar abiotic conditions. These gaps could be addressed by examining closely related species that have recently diverged to occupy a range of environments. This type of system would also provide clearer insight into the early stages of fruit evolution than has been possible with previous approaches. Furthermore, few studies have used a controlled common garden setup in which all plants are grown under the same conditions to ensure that observed variations in fruit traits are mainly due to genetics rather than environment.

One system well-suited to this fine-grained approach to fruit diversity studies is the tomato clade, a group of 14 congeneric species (*Solanum* sect. *Lycopersicon*, Peralta et al., 2008) that includes the cultivated tomato (*Solanum lycopersicum* var. *lycopersicum*) and 13 species of wild relatives (Figure 1). Throughout this paper, we used the word “tomato” to refer to any of the 14 species in the group. The clade is a monophyletic group estimated to have diverged from a common ancestor ~2 to 2.5 million years ago based on fossil-calibrated molecular clock evidence (Pease et al., 2016; Särkinen et al., 2013), which enables consideration of phylogeny-related constraints at a finer time scale than many previous studies of fruit trait diversity. The 13 wild tomato species inhabit a range of environments across the coasts, deserts, and mountains of western South America, allowing us to account for abiotic factors as potential selective forces. The group displays interesting inter- and intra-species fruit differences such as colors ranging from red to orange to yellow to green—enough variation to make testing hypotheses of fruit trait evolution feasible. The tomato system has been well studied as a model of fruit development and genetics (Kimura & Sinha, 2008) as well as ecological and evolutionary genomics (Moyle, 2008), providing a stronger understanding of the molecular underpinnings of traits than is possible with most other fruit systems. Despite these advantages of the tomato clade as a study system, the extent of fruit trait diversity across all 13 wild species and the potential evolutionary drivers of this variation have yet to be assessed. By including all species in the clade, we establish a complete evolutionary framework so that assumptions do not have to be made about how often traits evolved independently.

In this study, we present the groundwork for a novel approach and scale to fruit trait evolution studies using the tomato system and

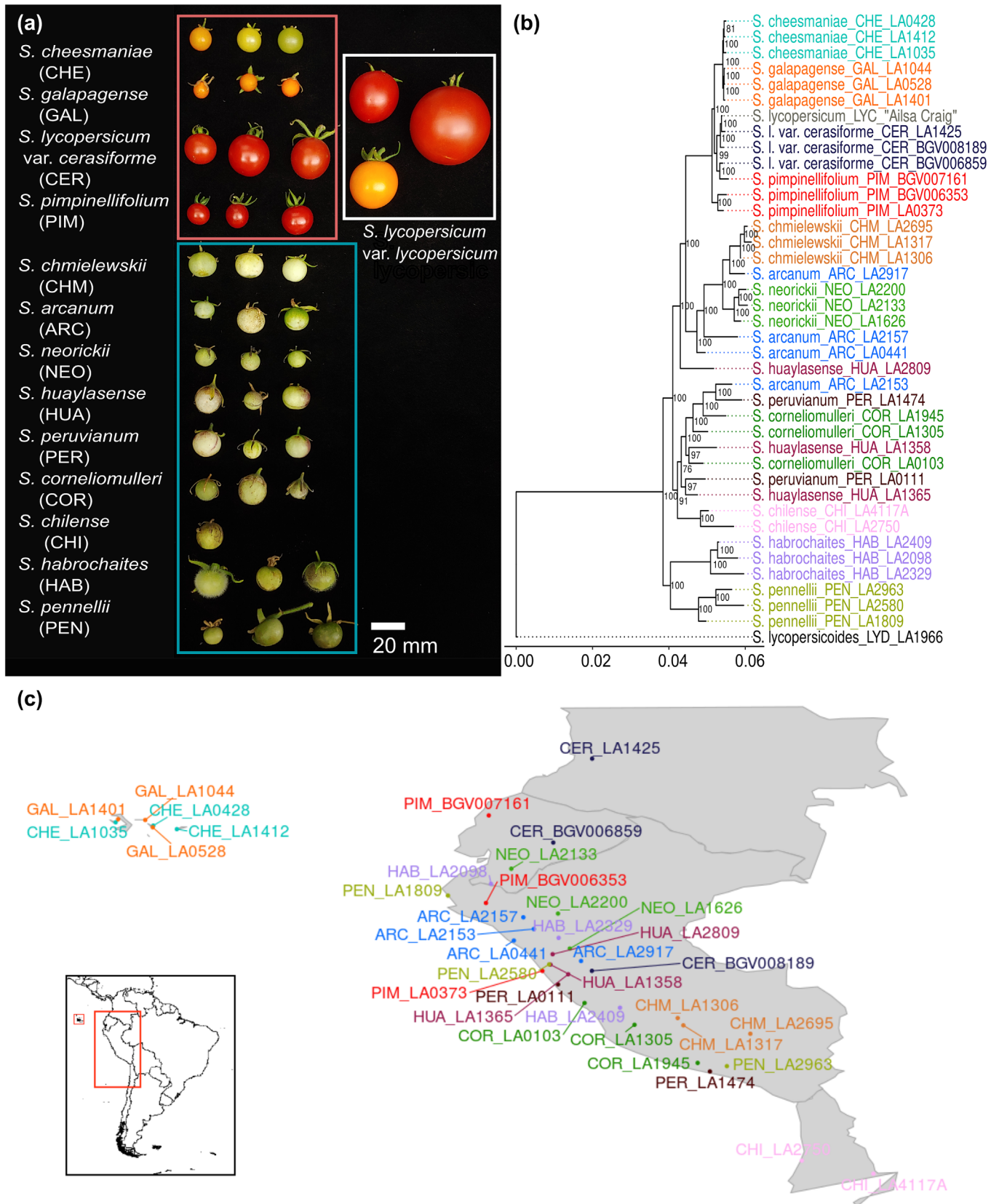


FIGURE 1 Legend on next page.

a common garden. Given that wild tomatoes display multiple fruit colors and inhabit a range of environments, the system holds potential for testing the relative importance of animal preferences and abiotic conditions, two factors currently considered major drivers of fruit color syndromes (Sinnott-Armstrong et al., 2021). Our first specific objective was to establish whether the tomato clade holds enough genetically-driven variation in disperser-relevant traits to make future hypothesis testing tractable. We did this by assessing the extent of variation in color, nutrition, and morphology traits in a common garden and whether the traits vary independently of phylogeny. Our second specific objective was to explore whether patterns of covariation exist that could be tested for underlying evolutionary drivers. We did this by (a) assessing whether any groups of disperser-relevant fruit traits cluster together in distinct syndromes, (b) whether pairwise correlations exist between external appearance and internal composition, and (c) whether any fruit traits are correlated with historical climate variables. The presence of fruit syndromes in the tomato clade, particularly if they involve associations among external color and internal nutrients after correcting for phylogeny and accounting for climate, could be used to test the hypothesis that animal disperser preferences were an evolutionary driver of fruit trait divergence. This so-called dispersal syndrome hypothesis has been tested in a number of systems (Valenta & Nevo, 2020) but not yet across an entire clade of closely related species that have adapted to different environments on a relatively recent timescale.

2 | MATERIALS AND METHODS

2.1 | Plant material

Our sampling included 38 accessions chosen to span the known phylogenetic and geographic diversity across the tomato clade, consisting of 13 taxa: 12 species of wild tomato and the semi-wild *Solanum lycopersicum* var. *cerasiforme* (Figure 1). Three accessions of each taxa were selected, representing different parts of their geographic ranges when possible (Figure 1c). Seeds were obtained from the C. M. Rick

Tomato Genetics Resource Center at the University of California, Davis, USA (TGRC, <http://tgrc.ucdavis.edu>) and the Universitat Politècnica de València, València, Spain (accession IDs and collection locations listed in Dataset S1). Each accession is representative of an independently sampled population in the wild. For autogamous (self-fertilizing) self-compatible species (*Solanum cheesmaniae*, *Solanum galapagense*, some *Solanum pimpinellifolium*, *S. lycopersicum* var. *cerasiforme*, *Solanum neorickii*) each plant of an accession is likely genetically identical. For others, plants are not identical, as the facultative self-compatible or allogamous self-incompatible accessions are maintained through “mass sibling” pollination in germplasm centers.

Seeds were soaked in a 2.7% sodium hypochlorite solution (per TGRC protocol) and germinated in a greenhouse at the University of Massachusetts Amherst. After 8 weeks, seedlings were transplanted into a high tunnel greenhouse at the University of Massachusetts Crop and Animal Research and Education Farm in South Deerfield, MA, where they were irrigated twice a week and fertilized once a week via drip lines. For each accession, three separate plants were grown in randomized locations. The study plants grew throughout the summer farm season (May–September 2020). Self-incompatible accessions were hand-pollinated several times during the first few weeks of flowering to facilitate fruit set, otherwise flowers were accessible to natural pollinators.

2.2 | Fruit phenotyping

We chose to quantify 21 fruit traits representing morphology, color, and nutrition (Dataset S2), the three disperser-relevant categories commonly used in fruit studies (e.g., Rojas et al., 2022; Valido et al., 2011). We measured size and shape because they affect an animal's ability to handle and consume the fruit. Another trait of potential interest to dispersers is the proportion of a fruit's cross-section taken up by fleshy pericarp rather than the seed-containing central locules, known as pericarp area ratio, as this could affect the overall nutritional reward found in one fruit. We quantified color into three variables with the CIE L*C*h system (lightness, chroma, and hue) as

FIGURE 1 Overview of the tomato clade (*Solanum* sect. *Lycopersicon*) and the 38 accessions used in this study. (a) Representative ripe fruit from each of the three accessions phenotyped per species (except for *Solanum chilense* [CHI], for which only one of the two phenotyped accessions is shown). Red box denotes colored-fruited group, blue box green-fruited group, white box cultivated tomato. (b) Maximum likelihood consensus tree of accessions (color-coded by species, see below) based on 64,745 SNP markers obtained through genotyping-by-sequencing, constructed in IQ-TREE with 1000 ultrafast bootstraps using the TVM + F + ASC + R4 nucleotide substitution model selected by ModelFinder with ascertainment bias correction and SH-aLRT likelihood ratio test. Numbers at nodes represent bootstrap percentages; x-axis shows tree scale. The outgroup species *Solanum lycopersicoides* and the cultivated tomato *Solanum lycopersicum* “Ailsa Craig” are included in the tree for reference but were not phenotyped for this study. Accession LA0441, which we originally included in the study as an *Solanum peruvianum*, has now been reclassified as *Solanum arcanum* in the Tomato Genetics Resource Center (TGRC) database, consistent with its placement on our tree. Only two accessions of *S. chilense* are shown because the third accession we planted (LA1932) failed to produce any fruit. IDs correspond with those provided in Dataset S1. (c) Map of collection locations for the 38 accessions used in this study, color-coded by species. Countries shown are the southern part of Colombia, Ecuador, Peru, and the northern part of Chile. Colors used in (b) and (c) represent species, abbreviated as follows: ARC = *S. arcanum*, CER = *Solanum lycopersicum* var. *cerasiforme*, CHE = *Solanum cheesmaniae*, CHI = *S. chilense*, CHM = *Solanum chmielewskii*, COR = *Solanum corneliomulleri*, GAL = *Solanum galapagense*, HAB = *Solanum habrochaites*, HUA = *Solanum huaylasense*, NEO = *Solanum neorickii*, PEN = *Solanum pennellii*, PER = *S. peruvianum*, PIM = *Solanum pimpinellifolium*. ‘LA’ or ‘BGV’ numbers represent TGRC accession IDs as listed in Dataset S1

they provide unambiguous color measures commonly used in comparative studies (Valido et al., 2011). Because the main caloric reward of tomatoes is sugar (fruits consist of about 90% water and are low in protein and lipids; García-Alonso et al., 2020), we quantified glucose and fructose (i.e., hexoses) as well as sucrose, the three most abundant sugars in tomatoes (Beckles et al., 2012). We also measured citric acid and malic acid because these two organic acids are another major component of cultivated tomato flavor (Tieman et al., 2017) and thus may be relevant to attracting animal dispersers. Ratios between the types of sugars or acids, as well as the sugar to acid ratio, were considered distinct traits because these proportions have important effects on human taste preferences (Anthon et al., 2011; Malundo et al., 1995), and thus possibly on other animals' tastes.

We collected fruits as they turned fully ripe, which we determined qualitatively for each species (see Dataset S3). A total of five fruits per plant (and thus 15 fruits per accession) were phenotyped when possible. After measuring diameter (latitudinally across widest part) and length (longitudinally from stem end to blossom end) with digital calipers and fresh weight with an analytical balance, fruits were cut in half latitudinally and scanned at 400 dpi with a color document scanner. Then seeds and pulp were removed, and the remaining fruit pericarps were frozen and stored at -80°C until we later thawed and quantified their liquid extracts for concentrations of glucose, fructose, sucrose, citric acid, and malic acid with absorbance-based assay kits and Brix (total soluble solids) with a refractometer (Methods S1). Color, lobedness degree, and pericarp area ratio were calculated from scanned images of fruit cross-sections using Tomato Analyzer version 4.0 software (Darrigues et al., 2008) after manually adjusting the pericarp boundary. Color was calibrated with a ColorChecker Classic (X-Rite, Grand Rapids, MI, USA).

All data analyses were conducted in R v3.6.3 (R Core Team, 2020). Summary statistics for each of the 21 variables were computed with built-in R functions and plots produced with the R package GGPlot2 v3.3.3 (Wickham, 2016). Accession mean trait values were calculated (Dataset S4) for use in later analyses that required one value per accession. To account for uneven sampling, the means for all fruits from each plant (generally five fruits per plant) were computed first, then per-plant means (generally three plants per accession) were averaged to arrive at a per-accession mean value for each trait.

2.3 | DNA extraction, genotyping-by-sequencing (GBS), and phylogenetic tree construction

We constructed an original GBS phylogeny to provide a fine-grained, accession-level resolution of evolutionary relatedness. DNA was extracted from ground frozen tissue of young unexpanded leaves using the CTAB method (Porebski & Bailey, 1997). A reduced representation GBS library was prepared according to a protocol modified from (Elshire et al., 2011) using the ApeKI restriction enzyme (Methods S2), and sequenced on one lane of an Illumina HiSeq PE 150 (Illumina Inc., San Diego, CA, USA) by Novogene, Inc.

(Sacramento, CA, USA). Sequences were processed and aligned to the SL4.0 cultivated tomato reference genome (Hosmani et al., 2019) to produce a variant call file (VCF), from which a set of filtered SNPs was used to construct a maximum likelihood consensus tree (Methods S3). To quantify phylogenetic signal, we used the function *phyloSignal* from the R package PHYLOSIGNAL v1.3 (Keck et al., 2016) to calculate Blomberg's K (Blomberg et al., 2003) and Pagel's λ (Pagel, 1999) from accession mean values of each trait and the original phylogenetic tree described above. Trees with trait values were produced with the function *contMap* from the R package PHYTOOLS v0.7-70 (Revell, 2012).

2.4 | Trait cluster analysis

We performed a principal components analysis (PCA) on the three color, 11 nutrition, and seven morphology traits. We used centered and scaled raw trait values for 208 individual fruit samples in the function *prcomp* from the R package STATS. A correlation matrix was used to normalize the covariance to a constant scale since variables were measured in different units.

To test fruit trait differences between the two color groups, we produced phylogenetic generalized least-squares (PGLS) models that estimated lambda. We used the functions *comparative.data* and *ppls* from the R package CAPER v1.0.1 (Orme et al., 2013) with lambda = "ML" to simultaneously optimize phylogenetic signal and regression parameters as recommended by Revell (2009).

2.5 | Correlation analyses for pairs of traits

To assess correlations between pairs of variables while taking phylogeny into account, as in Valido et al. (2011) and Rojas et al. (2022), we calculated phylogenetically independent contrasts (PICs) for each trait with the function *pic* from the R package APE v5.6-2 (Paradis & Schliep, 2019), using centered and scaled accession mean values as input. We then calculated Pearson's *r* for these PIC trait values using the function *corr.test* in the R package PSYCH v2.2.9 (Revelle, 2022) with *p*-values adjusted for multiple testing via the false discovery rate method (Benjamini & Hochberg, 1995).

2.6 | Abiotic variables analyses

Historical climate and soil data (36 variables total) for each accession's collection coordinates were obtained from the WORLDCLIM and Soil-Grids databases (Methods S4, Dataset S1). We examined Pearson's correlation coefficients for pairs of variables (Dataset S5; calculated from centered and scaled values using the *corr.test* function in R with false discovery rate *p*-value adjustment) and in cases where $r > |.65|$, we kept only one variable, resulting in these eight uncorrelated variables: temperature annual average, precipitation annual total, temperature seasonality, precipitation seasonality, solar radiation annual

average, wind speed coefficient of variation (CV), soil bulk density, and soil coarse fragment volume.

We produced PGLS models with one fruit trait as the response variable and the eight uncorrelated climate variables as predictors using the *pgls* function with $\lambda = \text{"ML"}$. All variables were centered and scaled prior to analysis. Top models based on AIC scores were compared with the *dredge* function, and each predictor's sum of Akaike weights over all models was calculated with the *importance* function, both from the R package MuMIn v1.43.1 (Barton, 2022).

3 | RESULTS

3.1 | Extent of variation in individual fruit traits

Because the extent of fruit trait diversity across all wild tomato species had not previously been surveyed, we first quantified the ranges

of variation across all fruits for each trait separately. We used CV to compare variability across traits, and found that, in general, nutrition traits showed greater variability than color and morphology (Figures 1a, 2; Dataset S6). Total sugars ranged in concentration from 4.1 (*S. neorickii*) to 168.0 (*Solanum arcanum*) mg/ml, and total acids from 0.24 (*Solanum huaylasense*) to 30.4 (*Solanum habrochaites*) mg/ml (Dataset S2). However, the most variable nutrition trait in terms of CV was sucrose/hexose ratio (Dataset S6). Color variations included light green, dark green, yellow, orange, red, and some purple streaking on a few otherwise green fruits (Figure 2). Chroma, which measures how vivid colors are, had the greatest CV, followed closely by hue (color of the human visual spectrum; Dataset S6). In general, greenish fruits tended to be less pure/vivid and more white (lightness) than red/orange/yellow fruits.

Tomato fruit sizes mainly fell between 10 and 20 mm in diameter and 1–3 g in fresh weight, ranging from minimums of 7.64 mm/0.24 g in *S. galapagense* to maximums of 28.29 mm/11.43 g in

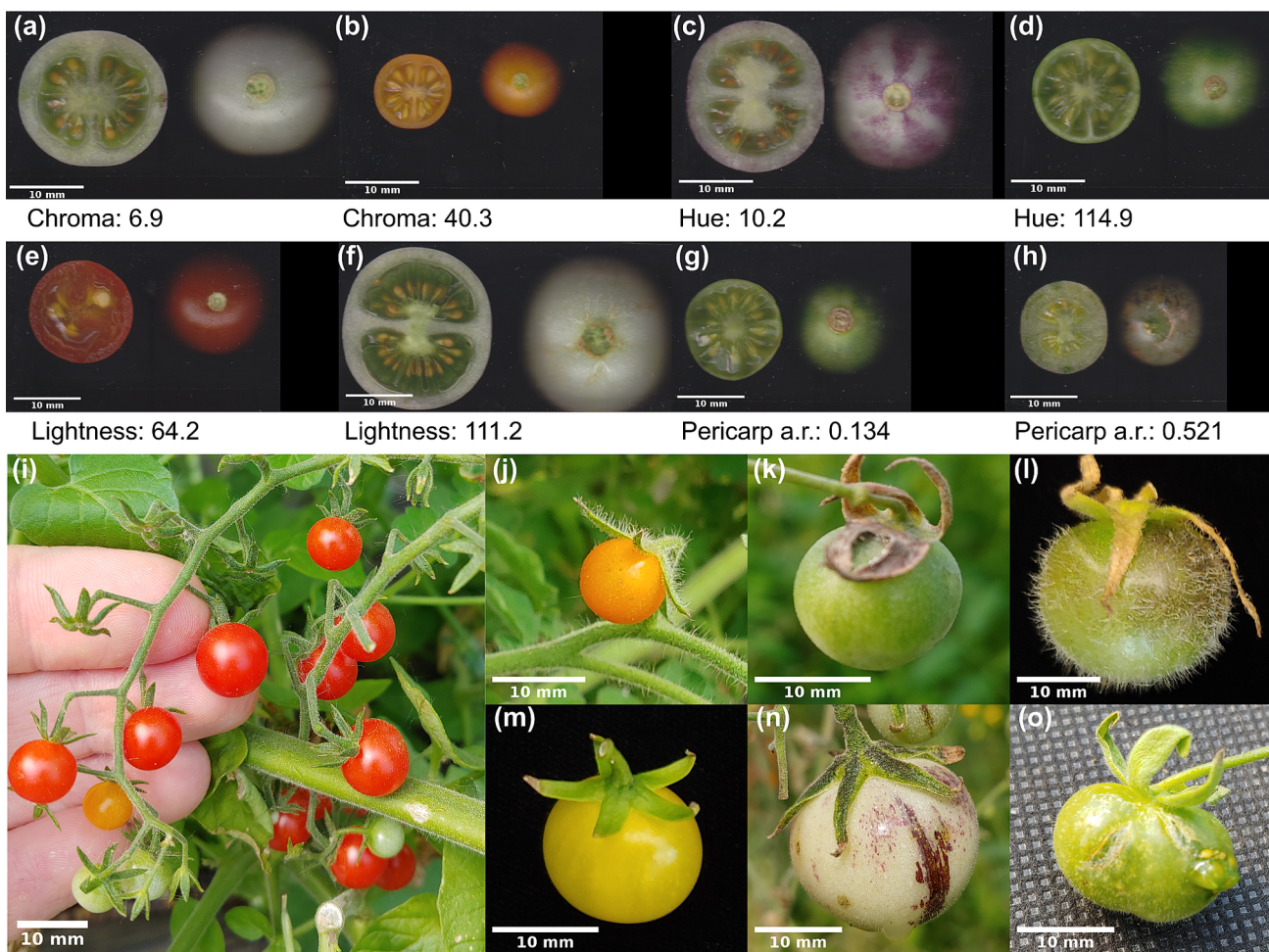


FIGURE 2 Phenotypic diversity in ripe wild tomato fruits. (a–h) Fruit cross-section scans showing fruits with the minimum and maximum values for chroma, hue, lightness, and pericarp area ratio: (a) *Solanum arcanum* LA2153, (b) *Solanum galapagense* LA0528, (c) *Solanum peruvianum* LA0111, (d) *S. arcanum* LA2157, (e) *Solanum lycopersicum* var *cerasiforme* BGV008189, (f) *S. peruvianum* LA1474, (g) *S. arcanum* LA2157, (h) *Solanum corneliomulleri* LA1945. (i–o) Select photos highlighting the range of variation across the clade: (i) *Solanum pimpinellifolium* LA0373, (j) *S. galapagense* LA0528, (k) *S. arcanum* LA2157, (l) *Solanum habrochaites* LA2329, (m) *Solanum cheesmaniae* LA0428, (n) *S. peruvianum* LA1474, (o) *Solanum pennellii* LA2963. Scale bars represent 10 mm.

S. lycopersicum var. *cerasiforme*. Shapes were quite spherical for most species (diameter/length ratio 1.0–1.2), except for *Solanum pennellii*, which tended to be more oval with a ratio as high as 1.5. The greatest CV for morphology traits was observed for fresh weight and seed count (Dataset S6).

3.2 | Trait variation relative to our GBS phylogeny

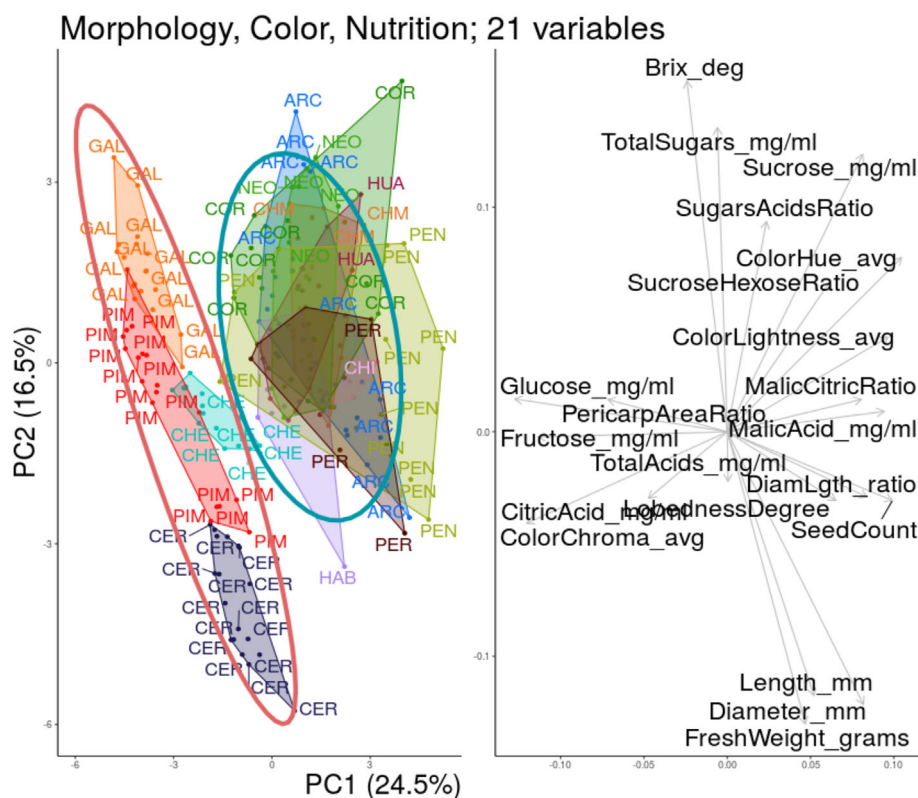
We established a fine-grained phylogenetic framework by creating an original phylogeny of the accessions phenotyped in this study. This enabled us to account for the statistical nonindependence of data points from related taxa due to shared ancestry (Felsenstein, 1985; Huey et al., 2019), as well as explore how patterns in fruit variation aligned with evolutionary relatedness. A total of 64,745 filtered SNPs were used to produce the maximum likelihood consensus tree, with TVM + F + ASC + R4 selected as the best-fit nucleotide substitution model.

Our tree recovered the current paradigm of relationships among wild tomato species (Figure 1b). The four colored-fruited species (*S. cheesmaniae*, *S. galapagense*, *S. lycopersicum* var. *cerasiforme*, and *S. pimpinellifolium*) fell within a well-resolved monophyletic clade, with *S. cheesmaniae* and *S. galapagense*, two species endemic to the Galápagos, as sister species. Sister to that clade was a monophyletic group consisting of *S. neorickii*, *Solanum chmielewskii*, and *S. arcanum*. Of note was the placement of *S. arcanum* LA2153 within the *Solanum peruvianum* clade, suggesting that the taxonomic designation of this accession may need revision. Our tree was consistent with other evidence

that genetic identity can conflict with taxonomic designation for some accessions of *S. arcanum*, *Solanum corneliomulleri*, *S. huaylasense*, and *S. peruvianum* (Labate et al., 2014; Pease et al., 2016). Our *Solanum chilense* accessions formed a distinct group within the larger “peruvianum complex” clade, consistent with evidence that *S. chilense* evolved from *S. peruvianum* ancestors (Stam et al., 2019). Finally, *S. habrochaites* and *S. pennellii* appeared as well-differentiated sister species that grouped together in a clade sister to all the other tomato species.

The trait phylogenetic signal results based on this tree confirmed the need to incorporate phylogeny into our analyses when possible, as 13 of 21 traits had significant values for both Blomberg's K and Pagel's lambda ($p < .01$; 10 of those 13 traits were also significant at a Bonferroni-corrected cutoff of $p < .002$; Dataset S6). These patterns also provided insight into which traits were labile (able to change over short evolutionary timescales) versus which were potentially more constrained. Traits with low phylogenetic signal values ($K < 0.8$, $\lambda < 0.65$) indicative of lability were the nutrition variables of total sugars, total acids, citric acid, Brix, and sugars/acids ratio, as well as the size variables of diameter, length, and fresh weight. For those nutrition traits, high and low values were spread randomly when plotted across the tree (Figure S1n,o,q,r,s). For the size traits, the low phylogenetic signal was mainly due to the smallest- and largest-fruited species (*S. galapagense* and *S. lycopersicum* var. *cerasiforme*) being closely related (Figure S1a,b,d). The remaining traits displayed high phylogenetic signal values ($K > 0.8$, $\lambda > 0.65$), although this was due to several different distribution patterns across the tree depending on the trait (Figure S1). For the three color variables, the high signal was

FIGURE 3 Wild tomato fruit traits cluster into two putative syndromes mainly defined by color, sugar type, and malic acid concentration, as shown by a non-phylogenetic principal components analysis (PCA) of 21 phenotypic traits for 208 fruits representing 38 accessions. All variables were centered and scaled. In left-hand plot: points = individual fruits; three-letter abbreviations denote species, abbreviated as follows: ARC = *Solanum arcanum*, CER = *Solanum lycopersicum* var. *cerasiforme*, CHE = *Solanum cheesmaniae*, CHI = *Solanum chilense*, CHM = *Solanum chmielewskii*, COR = *Solanum corneliomulleri*, GAL = *Solanum galapagense*, HAB = *Solanum habrochaites*, HUA = *Solanum huaylasense*, NEO = *Solanum neorickii*, PEN = *Solanum pennellii*, PER = *Solanum peruvianum*, PIM = *Solanum pimpinellifolium*; colored polygons = species. Red circles surround the colored-fruited species, blue circles the green-fruited species. Right-hand plot shows PCA factor loadings of the variables (details in Dataset S7).



largely due to all of the color-fruited accessions belonging to the same monophyletic group. A similar distribution was seen for the three sugars and malic acid (as well as their associated ratios), with most or all accessions within the colored-fruited clade showing high glucose (Figure S1k) and fructose values but low sucrose and malic acid values relative to the green-fruited accessions. Interestingly, the green-fruited *S. pennellii* was an exception to this pattern as the LA1809 and LA2580 accessions were relatively high in glucose and fructose but low in sucrose (although they were high in malic acid like the rest of the green-fruited accessions). In contrast, the high signal in the case of diameter/length ratio (Figure S1c), lobedness degree, and seed count was likely due to only the closely related *S. pennellii* accessions displaying the highest trait values, as low trait values were more scattered throughout the tree.

3.3 | Trait clustering

We looked for evidence of fruit syndromes through PCA, to determine how fruits clustered together based on phenotypic similarity and which traits contributed most to this positioning. A PCA with all

208 individual fruits measured for all 21 traits as points (Figure 3) showed two distinct clusters differentiated primarily along PC1, and factor loadings (Dataset S7) provided insight into which traits covaried together. One group consisted of fruits primarily green in hue, white in lightness, and high in malic acid/sucrose/seed count, whereas the other contained fruits reddish in hue, more vivid in chroma, high in glucose/fructose, and lower in seed count. There was no obvious distinction along PC2, suggesting that the size and total sugar concentration traits most influential to that axis were not part of the two observed clusters. With the caveat that analyzing subsets of data may introduce bias, clustering patterns in PCAs conducted separately for each trait category (Figure S2) also supported the existence of two fruit trait clusters differentiated primarily by color, sugar type, malic acid, and possibly seed count and shape.

Because these two clusters appeared to correspond to the informal colored- and green-fruited species sub-groupings within the clade, we assessed the strength of these two putative syndromes by testing differences between the two color groups in sugar types, malic acid, seed count, and shape with PGLS models, which account for phylogenetic relatedness (Figure 4). There were statistically significant ($p < .008$ after Bonferroni correction) differences in glucose

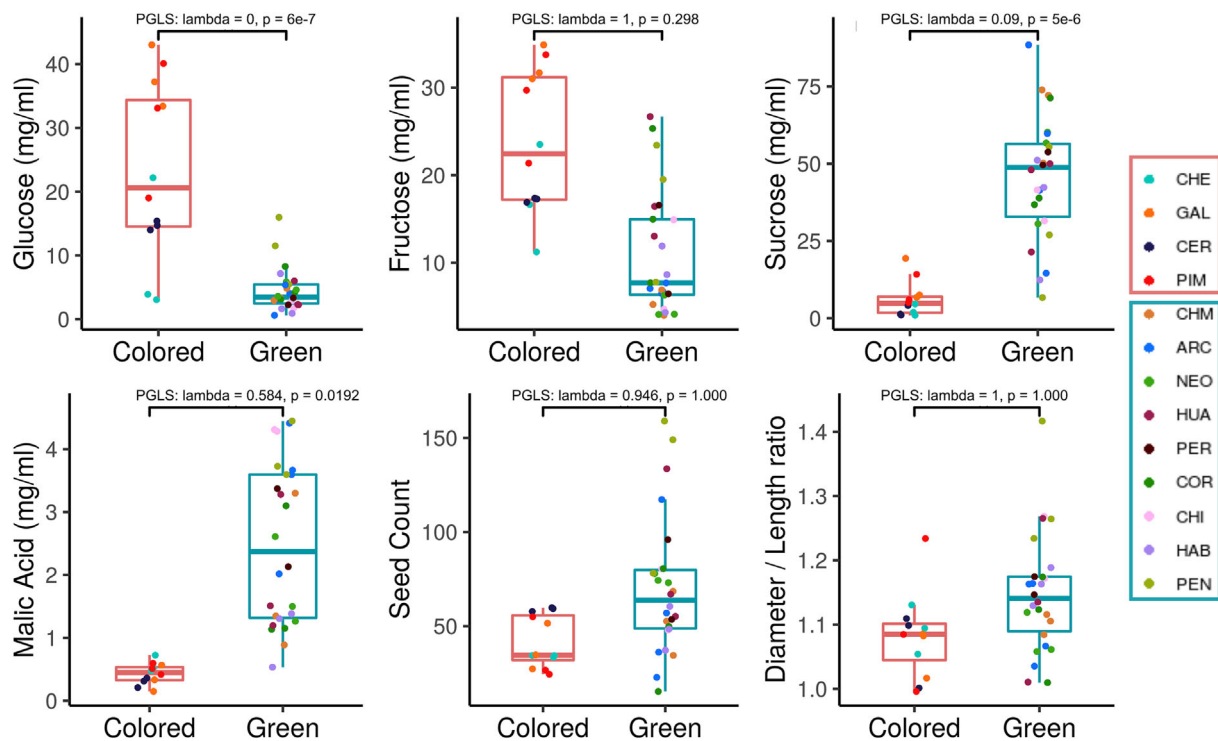


FIGURE 4 To assess the strength of the two fruit syndromes suggested by principal components analysis (PCA) (Figure 3), these boxplots test differences between colored- and green-fruited species groups for the nutrition and morphology traits that were the strongest differentiators in the PCA. Colored points represent accessions and are color-coded by species, abbreviated as follows: ARC = *Solanum arcanum*, CER = *Solanum lycopersicum* var. *cerasiforme*, CHE = *Solanum cheesmaniae*, CHI = *Solanum chilense*, CHM = *Solanum chmielewskii*, COR = *Solanum corneliomulleri*, GAL = *Solanum galapagense*, HAB = *Solanum habrochaites*, HUA = *Solanum huaylasense*, NEO = *Solanum neorickii*, PEN = *Solanum pennellii*, PER = *Solanum peruvianum*, PIM = *Solanum pimpinellifolium*. Red boxplots denote colored-fruited group, blue boxplots green-fruited group. Diameter/length ratio was chosen to represent shape because it was highly correlated ($r = .81^{***}$; Dataset S8) with lobedness degree, the one other shape-related variable we measured. Raw (untransformed) accession mean values were used; sample sizes were 12 for the colored group and 26 for the green group. Lambda and p -values are from phylogenetically controlled generalized least-squares (PGLS) models with lambda = “ML”

($\lambda = 0$, $p = 1e-7$), sucrose ($\lambda = 0.09$, $p = 8.1e-7$), and malic acid ($\lambda = 0.584$, $p = .0032$) but not fructose ($\lambda = 1$, $p = .0496$), seed count ($\lambda = 0.946$, $p = .346$), and diameter/length ratio ($\lambda = 1$, $p = .475$). In summary, sugar type and malic acid concentration were the key traits along with color that defined the two broad fruit clusters within the clade, with high glucose and low malic acid associated with colored fruits, while high sucrose and high malic acid were associated with green fruits.

Looking at how species were distributed within the two broad PCA clusters, the colored-fruited group showed more species-specific differentiation than the green-fruited group when all traits were considered together (Figure 3). Different fruit trait types appeared to vary in their effects on species differentiation when PCAs were conducted separately (see caveat above) for each trait category (Figure S2). Noteworthy species with unique traits were *S. lycopersicum* var. *cerasiforme* (large size), *S. galapagense* (small size and high glucose/fructose), *S. pimpinellifolium* (high glucose/fructose), and *S. pennellii* (more oval shape and high seed count). Interestingly, the various shades of green displayed by the green-fruited species are not distinct enough to differentiate the species.

3.4 | Correlations among pairs of traits

We further examined associations among fruit traits in a phylogenetic context by conducting Pearson correlation tests on each trait's PICs. To assess whether correlation patterns differed within color groups, calculations were made separately for: (1) all 38 accessions, (2) only the 12 colored-fruited accessions, and (3) only the 26 green-fruited accessions (Dataset S8). Stars following correlation coefficients denote significance levels (** $p < .001$, * $p < .01$, $p < .05$) after false discovery rate adjustment for multiple testing.

Variables that were very highly correlated (Pearson's $r > .8^{***}$) in all three groupings revealed some measurement redundancy among the 21 traits. The size variables of diameter, length, and fresh weight all had r values $> .9^{***}$. The total sugars-Brix correlation ($r > .87^{***}$) suggested that glucose, fructose, and sucrose make up the majority of soluble solids in wild tomatoes, while total acids-citric acid ($r > .86^{***}$) showed malic acid to be the substantially less abundant of the two acids quantified.

We next looked at pairs of color or size and sugar/acid traits to explore whether a fruit's external appearance may provide an honest signal of its nutrient content. The only pairs of color-nutrition traits that were consistently correlated ($r > .4$) across all three groupings (Dataset S9) were chroma-total sugar ($r = .48^*$, $.70$, $.54$ for all, colored-fruited, and green-fruited groups, respectively) and lightness-sugars/acids ratio ($r = .48^*$, $.43$, $.49$). These associations suggested that throughout the clade, more vivid fruits tend to have higher sugar concentrations and more whitish fruits tend to have more sugars relative to acids. The only morphology-nutrition pairing consistently correlated ($r > |.38|$) across all three groupings was diameter-glucose ($r = -.49^*$, $-.66$, $-.39$), showing that smaller fruits tend to have higher glucose concentrations, particularly within the colored-fruited group.

Other pairs of traits showed correlation patterns that differed between the two color groups (Dataset S9). Associations that were present only in the colored-fruited group were related to color and sugar, most notably chroma-glucose, chroma-fructose, hue-glucose, and hue-fructose ($r > |.45|^*$). Additionally, chroma-hue were correlated in the colored-fruited group ($r = -.70^*$) but not the green-fruited group ($r = .19$). In contrast, an association only present in the green-fruited group was lobedness degree-malic/citric ratio ($r = .57$). Interestingly, size and sugar were correlated in both groups but in opposite directions: diameter-sucrose ($r = -.68$) and diameter-total sugars ($r = -.71$) were negative within colored-fruited and positive within green-fruited ($r = .48$ and $r = .41$, respectively), suggesting that larger fruits tend to have lower sugar concentrations in the colored-fruited group but higher sugar concentrations in the green-fruited group. These associations between a fruit's external appearance and internal nutrient content could potentially function as honest signals to animal dispersers.

3.5 | Abiotic factors associated with fruit traits

To explore whether the fruit trait covariations seen in our clustering and correlation analyses may have been related to adaptation to environmental niches, we assessed which historical climate variables were most associated with particular fruit traits via two methods.

We first conducted Pearson correlation tests for eight uncorrelated climate variables (described in Section 2) and 12 fruit variables (chosen because they were important in clustering and correlation results). A non-phylogenetic correlation chart of these 20 variables (centered and scaled) with all 38 accessions included (Dataset S10) showed that seven climate-fruit trait pairs had $r \geq |.4|$, of which two had $r \geq |.5|$. Stars following correlation coefficients denote significance levels (** $p < .01$, $p < .05$) after false discovery rate adjustment for multiple testing. Temperature annual average was correlated ($r > |.40|^*$) with lightness, chroma, glucose, fructose, and sucrose. The only other climate-fruit trait pairs with $r \geq |.4|$ were temperature seasonality-fructose ($r = .40^*$) and precipitation annual total-diameter ($r = .42^{**}$). Before concluding whether a pair of traits was associated, we also examined PGLS models in which one of the 12 syndrome-relevant fruit traits was the response variable and the eight climate variables were predictors. For each model, predictor variables with sum of weights $> .6$ were considered strongly associated with the fruit trait response variable. Six of the 12 response variables had one or more predictor climate variables above the cutoff (Dataset S11).

After comparing results from the two methods, only three fruit-climate associations (Figure S3) were above our cutoffs in both the non-phylogenetic Pearson correlation tests and the PGLS models: (1) diameter-precipitation annual total (Pearson $r = .42^{**}$, PGLS sum of weights $.87$), (2) chroma-temperature annual average (Pearson $r = .50^{**}$, PGLS sum of weights $.67$), and (3) fructose-temperature seasonality (Pearson $r = .40^*$, PGLS sum of weights $.63$). Two of those three associations involved color or sugar variables that were part of the fruit trait clusters identified above. Chroma-temperature annual average was the stronger of the two, leading us to hypothesize that

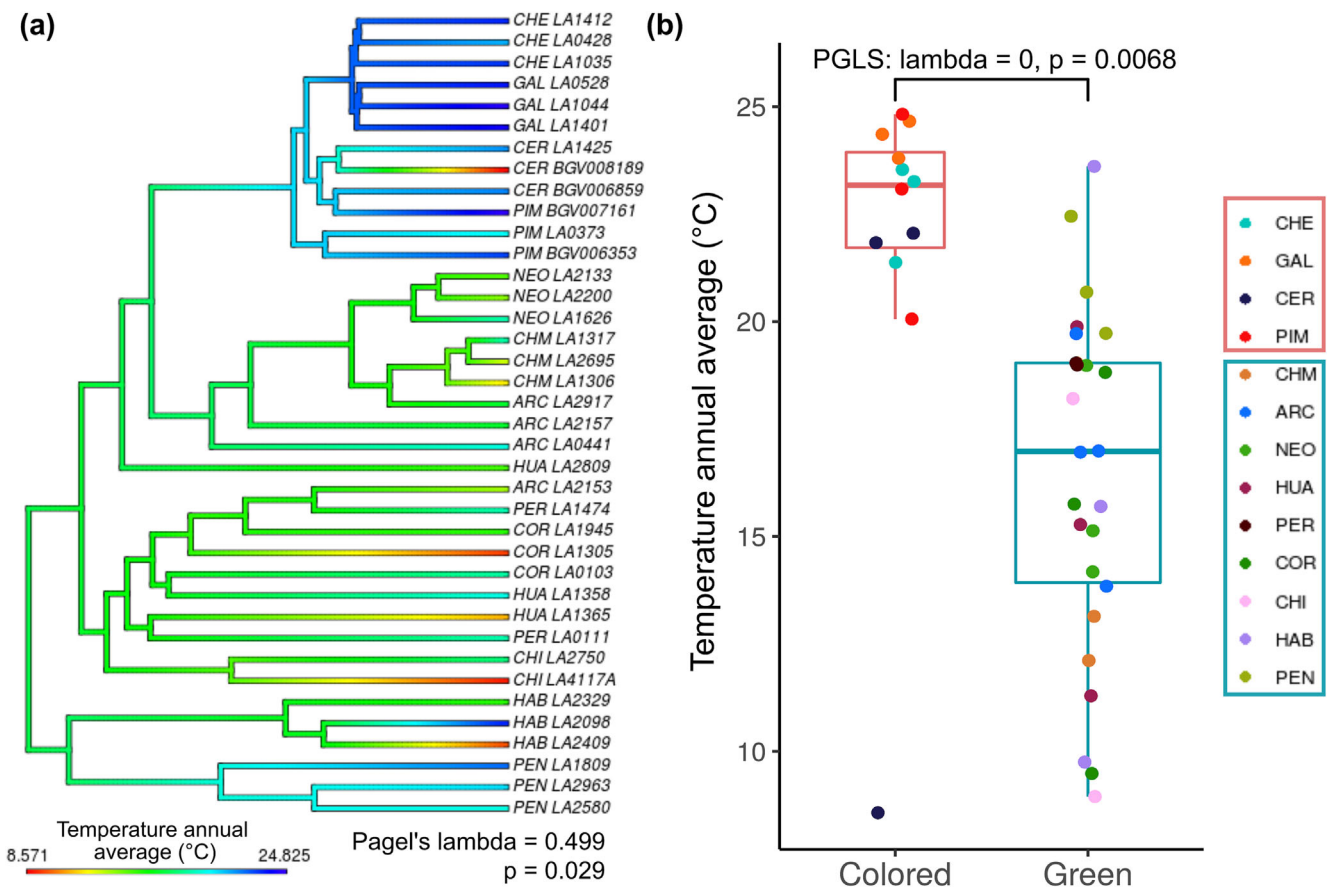


FIGURE 5 Temperature annual average and fruit color showed the strongest association between an historical climate variable and a syndrome-related fruit trait, leading us to hypothesize that the evolution of colored fruits may be related to warmer environments. (a) Plot showing a heatmap of temperature annual average raw values at accession collection sites (in degrees Celsius) distributed across our phylogenetic tree. Pagel's lambda phylogenetic signal test results (with p -value) for temperature annual average are shown below the tree. (b) Boxplot showing the significant difference in temperature annual average between colored-fruited (12 accessions, denoted by red box) and green-fruited (26 accessions, denoted by blue box) species groups. Colored points represent accessions and are color-coded by species, abbreviated as follows: ARC = *Solanum arcanum*, CER = *Solanum lycopersicum* var. *cerasiforme*, CHE = *Solanum cheesmaniae*, CHI = *Solanum chilense*, CHM = *Solanum chmielewskii*, COR = *Solanum corneliomulleri*, GAL = *Solanum galapagense*, HAB = *Solanum habrochaites*, HUA = *Solanum huaylasense*, NEO = *Solanum neorickii*, PEN = *Solanum pennellii*, PER = *Solanum peruvianum*, PIM = *Solanum pimpinellifolium*. Pagel's lambda and p -values are from a phylogenetically controlled generalized least-squares (PGLS) model with lambda = "ML". The majority of accessions in the monophyletic colored-fruited clade were collected from warmer locations than the green-fruited accessions, although there were three accessions that deviated from this trend: one *S. lycopersicum* var. *cerasiforme* (BGV008189) collected at an unusually cold site for the species, as well as one *S. habrochaites* (LA2098), and one *S. pennellii* (LA1809) collected at warm sites near the equator at the northern edge of these species' ranges

the evolution of the colored-fruited cluster may be related to warmer climates. To visualize whether closely related accessions inhabit similar environments in terms of temperature annual average, we plotted these temperature values on our phylogeny (Figure 5a). The majority of accessions in the monophyletic colored-fruited clade were collected from warmer locations than the green-fruited accessions, with a significant difference (PGLS $p < .001$) between the color groups (Figure 5b).

4 | DISCUSSION

Fleshy fruits display a striking diversity of traits whose evolutionary drivers are not well understood despite their relevance to agriculture.

Using a common garden and a fine-grained phylogenetic framework, we developed a novel approach and scale for fruit evolution research by conducting the first clade-wide systematic survey of disperser-relevant fruit traits across wild tomato species. Contrary to expectations that fruit traits have low evolutionary potential due to constraints of shared ancestry, as suggested by taxonomically broad studies (e.g., Jordano, 1995), we found substantial variation in color and nutrition traits across the recently diverged tomato clade. Morphology traits did not vary as widely across wild species, which is notable given the many shapes and sizes seen in cultivated tomato varieties. Total sugars, total acids, and size varied independently of phylogeny, while differences in color, sugar type, and malic acid concentration were aligned with the evolutionary split between the four colored-fruited species and the nine green-fruited species. After

correcting for nonindependence due to shared ancestry, the differences in glucose, sucrose, and malic acid between those two groups of species were statistically significant, suggesting that the differences have been maintained by selective pressure. Furthermore, these trait differences clustered into syndromes involving covariations among external color and internal nutrients that could function as honest signals to dispersers, a pattern in line with the hypothesis that animal preferences are an underlying evolutionary driver. Historical climate and soil variables were not strongly correlated with fruit traits, so selective pressure from frugivores may have been a more important evolutionary force than adaptation to abiotic conditions. We did find one notable association between colored-fruited species and high annual average temperature that should be considered when trying to tease apart the biotic and abiotic factors underlying the two wild tomato fruit syndromes.

To our knowledge, there is currently no systematic data on which animals eat wild tomato fruits, although there are some anecdotal reports. The TGRC database contains collection notes from over 1700 total accessions of species we studied, out of which we found only 35 records that mention the words “eaten,” “disperse,” “animal,” “bird,” or “rodent” (Dataset S12), providing evidence of inferred or reported fruit consumption by humans, grackles, other birds, tortoises, rodents, and “other animals.” These TGRC notes do not reveal any clear patterns in disperser preferences: birds, mammals, and reptiles appear to consume both colored- and green-fruited species but the prevalence of any given animal cannot be determined from these limited anecdotal reports. Beyond collection notes, studies of captive animals showed that passage through the guts of Galápagos tortoises (Rick & Bowman, 1961) and mockingbirds (Rick, 1964) can improve germination of *S. cheesmaniae* seeds, although dispersal in the wild is still unstudied (Heleno et al., 2011). There has been one recent video captured of a Galápagos mockingbird eating a yellow wild tomato fruit (Matthew Gibson, 2019, personal communication). Recent field observations of *S. chilense* plants in Chile noted that ripe fruits always hang low to the ground and may be dispersed by *Microlophus* lizards (Remco Stam and Edeline Gagnon, 2023, personal communication). One field study in northern Chile (Chetelat et al., 2009) found partially eaten fruit remains from *Solanum sitiens*, a green-fruited relative of wild tomatoes, in a rodent burrow. Beyond that, we are not aware of any published data of potential tomato seed dispersers. Future studies that employ cameras to monitor animal visitation at plants with ripe fruits (e.g., Levey et al., 2006) would be a valuable contribution to the limited data currently available.

The two distinct trait clusters we observed involve associations among color and nutrient content in line with those found in other fruit systems where selective pressure from animal dispersers (i.e., the dispersal syndrome hypothesis) has been inferred (Sinnott-Armstrong et al., 2020; Valenta & Nevo, 2020; Valido et al., 2011). The colored and green tomato fruit groupings correspond to distinct balances of sugar and acid types, which are generally consistent with conventional syndromes (Janson, 1983) based on bird versus mammal preferences. One observed cluster consists of fruits that are red/orange/yellow, high in glucose and fructose, and low in sucrose and malic acid—traits

thought to be preferred by birds, which are expected to be the main dispersers of the colored-fruited tomatoes because most birds have excellent color vision that they use to locate food (Lomáscolo & Schaefer, 2010), and some birds are unable to digest sucrose (Martinez del Rio & Stevens, 1989). The other cluster includes fruits greenish and lighter in color, high in sucrose and malic acid but low in glucose and fructose—traits generally thought to be more in line with fruits primarily dispersed by mammals. Despite the difference in sugar types, both groups of fruits produce similar total sugar concentrations (Figure S4), suggesting both syndromes offer comparable overall caloric rewards. The fact that green-fruited species devote energy into sugary fruits could mean that they benefit from attracting animal dispersers, a hypothesis contrary to speculations that green-ripe wild tomato fruits are not consumed at all (e.g., Kamiyoshihara et al., 2020).

Our correlation analyses show that a wild tomato fruit's external appearance could also provide reliable information about its internal nutrient content on a finer, more quantitative scale, with a few different nuances for the colored versus green species groups. Regardless of species grouping, high chroma fruits tend to have greater total sugar concentrations and high lightness fruits tend to have more sugars relative to acids. Within each color group, however, hue and diameter appear to signal different nutritional traits: for the colored-fruited species only, less reddish, smaller fruits tend to be higher in glucose, fructose and total sugar concentration; while for the green-fruited species only, less greenish, larger fruits tend to be higher in total sugar and lower in citric acid concentration. Interestingly, the colored-fruited species showing higher sugar concentration in smaller fruits is consistent with cultivated tomatoes (Levin & Schaffer, 2013), but the green-fruited species displayed the opposite trend with sugar concentration tending to be higher in larger fruits. These different nuances in signaling between species groups merit further testing as to whether they may have been influenced by distinct pressures from dispersers. Additionally, it would be valuable to assess whether signal-reward correlations exist for fruits at the intraspecific level because that scale of variation may be more ecologically relevant to frugivores (Nevo et al., 2022).

While some correlations between color and nutrition could be due to shared biochemistry rather than natural selection from animals (Cazetta et al., 2012), this seems unlikely to be the case for tomato fruits because color, sugar, and acid are controlled by distinct biochemical pathways. Red/orange/yellow colors are produced via the carotenoid pathway from isoprenoid precursor molecules (Sun & Li, 2020), sugar levels are altered by sucrose and starch metabolism (Beckles et al., 2012), and citric and malic acid are intermediates in the tricarboxylic acid cycle (Zhang & Fernie, 2018). Molecular mapping has revealed only a few genomic regions affecting both sugar and acid content (Grandillo et al., 2013). Humans have taken advantage of simply inherited variations in these disparate pathways to produce different color, sugar, and acid combinations in domesticated cultivars (Levin & Schaffer, 2013), demonstrating that changes in only a few loci can result in large fruit trait differences. For example, altered expression levels of the *TIV1* invertase gene cause tomato fruits to

shift between sucrose or hexose accumulation (Moy et al., 2007). Given that the biochemical basis of color, sugar, and acid traits allows for evolutionary lability, the persistence of two clusters of associated traits with high phylogenetic signal within the clade suggests a stabilizing selective force.

The correlations among color, sugar type, and malic acid traits may be linked to closely related species adapting to environments with similar selective pressures. Our results suggest that colored-fruited accessions evolved in locations with warmer annual average temperature (which correlates with lower elevation and higher annual average vapor pressure) than green-fruited accessions, consistent with previous work showing the mean annual temperature across species distributions was higher for *S. cheesmaniae*, *S. galapagense*, *S. lycopersicum* var. *cerasiforme*, and *S. pimpinellifolium* relative to the green-fruited species (Nakazato et al., 2010; Ramírez-Ojeda et al., 2021). Thus something about warmer locales may have been necessary for the evolution of colored fruits. Perhaps warmer environments enable fruits to complete a more energetically expensive developmental color sequence due to a longer growing season (Sinnott-Armstrong et al., 2018), which could offset the lost photosynthetic capability in fruits that become brightly colored upon ripening (Cipollini & Levey, 1991). Warmer climates could also have the indirect effect of unique biotic pressures, such as a greater proportion of animal dispersers that prefer colored fruits (Willson & Whelan, 1990).

It is also worth considering what evolutionary forces may be behind some of the more unique species within the clade. *S. pennellii* fruits stand out for their oval shape and high seed count. The species inhabits environments with relatively high annual average temperature (compared to other green-fruited species) and very low annual precipitation (second lowest in the clade behind *S. chilense*) (Nakazato et al., 2010; Ramírez-Ojeda et al., 2021); perhaps this unique combination plays host to a different assemblage of animal dispersers. *S. lycopersicum* var. *cerasiforme* produces the largest fruits in the clade and is found in locales with the highest annual precipitation; this correlation could be due to greater water availability being necessary for the production of larger fruits (tomato fresh weight is ~90% water), or perhaps locations with more rainfall have different resident dispersers that are able to consume larger fruits. The two Galápagos species are notable for stark differences in fruit traits between very closely related species—*S. cheesmaniae* fruits have low chroma, light orange/yellow/greenish hue, and low total sugar concentrations, while *S. galapagense* fruits have high chroma, darker orange hue, small size, and high total sugar concentration. Although genetic drift may be a powerful force in these small island populations, whether selection from dispersers has influenced the differences is an open question.

Our evidence of an association among color, sugar type, and malic acid traits in wild tomato fruits lays the groundwork for future tests of the dispersal syndrome hypothesis. More data on wild tomato frugivores and their preferences would enable the hypothesis to be tested on a more fine-grained phylogenetic scale than has yet been used. Furthermore, data on disperser preferences and a more thorough quantification of intraspecific fruit trait variation could test whether

the correlations among color, nutrition, and size we observed do in fact serve as honest signals linking a fruit's external appearance to its internal content. Some abiotic factors such as temperature annual average may be related to the trait associations, although the influence of climate is difficult to disentangle from the effects of unique assemblages of frugivores inhabiting different environments. Another pressure we did not explore but would be worth investigating is defense, as pathogens or seed predators may have influenced variation in traits such as trichomes, acylsugars, and glycoalkaloids.

Cultivated tomatoes have long been considered the predominant model organism for fleshy-fruited plants, leading to many ongoing discoveries of the genetic mechanisms and developmental processes controlling fleshy fruit traits. Our survey shows that the wild tomato clade harbors a rich diversity of fruit traits, offering an exciting opportunity to explore the evolutionary drivers that made the eventual domestication of this beloved fruit crop possible. Further expanding the study of model systems to incorporate wild relatives, particularly through ecological field studies of biotic interactions, as is now possible in tomatoes, can provide unparalleled opportunities to understand the genetic, developmental, and ecological factors that have shaped fleshy fruit evolution.

AUTHOR CONTRIBUTIONS

Jacob R. Barnett and Ana L. Caicedo designed and led the study. Jacob R. Barnett grew the plants and phenotyped the fruits, prepared the GBS sequencing library, and performed analyses. Rovin Sharma assisted with growing plants and phenotyping fruits, particularly measurements from scanned images. Gina Buonauro extracted DNA and assisted with fruit phenotyping and GBS sequencing library prep. Ian M. Gillis helped develop protocols for growing plants and measuring phenotypes. Maryam Rashidzade extracted DNA and assisted with GBS sequencing library prep. Ana L. Caicedo supervised the research. Jacob R. Barnett and Ana L. Caicedo wrote the manuscript, and all authors contributed to revising the manuscript.

ACKNOWLEDGMENTS

For advice and assistance growing the plants, we are grateful to the UMass Vegetable Extension's Katie Campbell-Nelson, Susan Scheufele, and Genevieve Higgins; UMass Greenhouse staff Chris Joyner, Dave O'Neil, Chris Phillips, and Dan Jones; and UMass South Deerfield Research Farm staff Bob Skalbite, Neal Woodard, and Zach Zenke. For assistance with plant care, phenotyping, and sequencing prep, we thank Caicedo Lab members Xiang Li, Sherin Perera, Jacob Scott, Carter Hissam, Dan Lowey, Jessica Lessard, Fabbih Hossain, and Amanda Semple. Denise Tieman provided invaluable advice on sugar/acid assay protocols. We also thank Ravi Ranjan in the UMass Genomics Resource Laboratory at the Institute for Applied Life Sciences for performing the BluePippin size selection during the genotyping-by-sequencing (GBS) library prep. For advice with statistical analysis, we are indebted to the UMass Quantitative Sciences Group, Jason Kamilar, and João Menezes. Lynn Adler provided helpful feedback on the manuscript. We thank Roger Chetelat for the suggestion to search the Tomato Genetics Resource Center collection notes

database for mentions of animal dispersers. This research was supported by the UMass Natural History Collections Grant to JB in 2020, UMass Graduate School Dissertation Research Grant to JB in 2020, Lotta Crabtree Fellowship in Production Agriculture to JB in 2019–2022, and National Science Foundation Grant No 1564366 to AC.

CONFLICT OF INTEREST STATEMENT

The authors confirm there are no conflicts of interest.

DATA AVAILABILITY STATEMENT

Accession details and raw data for phenotypic and climate variables can be found in Datasets S1, S2, and S4. DNA sequence data produced for GBS is deposited in the NCBI SRA database under BioProject ID PRJNA878823 at <https://www.ncbi.nlm.nih.gov/sra/PRJNA878823>.

ORCID

Jacob R. Barnett  <https://orcid.org/0000-0002-4213-4010>

Rovin Sharma  <https://orcid.org/0009-0005-0820-7692>

Gina Buonauro  <https://orcid.org/0009-0008-3483-3461>

Ian M. Gillis  <https://orcid.org/0009-0003-6631-853X>

Maryam Rashidzade  <https://orcid.org/0009-0004-6374-0871>

Ana L. Caicedo  <https://orcid.org/0000-0002-0378-6374>

REFERENCES

- Anthon, G. E., LeStrange, M., & Barrett, D. M. (2011). Changes in pH, acids, sugars and other quality parameters during extended vine holding of ripe processing tomatoes. *Journal of the Science of Food and Agriculture*, 91, 1175–1181. <https://doi.org/10.1002/jsfa.4312>
- Barton, K. (2022). MuMIn-package: multi-model inference R Package Version 1.46.0. <https://cran.r-project.org/package=MuMIn>
- Beckles, D. M., Hong, N., Stamova, L., & Luengwilai, K. (2012). Biochemical factors contributing to tomato fruit sugar content: A review. *Fruits*, 67, 49–64. <https://doi.org/10.1051/fruits/2011066>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Blomberg, S. P., Garland, T. Jr., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution; International Journal of Organic Evolution*, 57, 717–745. <https://doi.org/10.1111/j.0014-3820.2003.tb00285.x>
- Bolmgren, K., & Eriksson, O. (2005). Fleshy fruits - origins, niche shifts, and diversification. *Oikos*, 109, 255–272. <https://doi.org/10.1111/j.0030-1299.2005.12663.x>
- Brodie, J. F. (2017). Evolutionary cascades induced by large frugivores. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 11998–12002. <https://doi.org/10.1073/pnas.1710172114>
- Cazetta, E., Galetti, M., Rezende, E. L., & Schaefer, H. M. (2012). On the reliability of visual communication in vertebrate-dispersed fruits. *The Journal of Ecology*, 100, 277–286. <https://doi.org/10.1111/j.1365-2745.2011.01901.x>
- Chetelat, R. T., Pertuzé, R. A., Faúndez, L., Graham, E. B., & Jones, C. M. (2009). Distribution, ecology and reproductive biology of wild tomatoes and related nightshades from the Atacama Desert region of northern Chile. *Euphytica/Netherlands Journal of Plant Breeding*, 167, 77–93. <https://doi.org/10.1007/s10681-008-9863-6>
- Cipollini, M. L., & Levey, D. J. (1991). Why some fruits are green when they are ripe: Carbon balance in fleshy fruits. *Oecologia*, 88, 371–377. <https://doi.org/10.1007/BF00317581>
- Darrigues, A., Hall, J., van der Knaap, E., Francis, D. M., Dujmovic, N., & Gray, S. (2008). Tomato analyzer-color test: A new tool for efficient digital phenotyping. *Journal of the American Society for Horticultural Science. American Society for Horticultural Science*, 133, 579–586. <https://doi.org/10.21273/JASHS.133.4.579>
- do Nascimento, L. F., Guimarães, P. R., Onstein, R. E., Kissling, W. D., & Pires, M. M. (2020). Associated evolution of fruit size, fruit colour and spines in Neotropical palms. *Journal of Evolutionary Biology*, 33, 858–868. <https://doi.org/10.1111/jeb.13619>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6, e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125, 1–15. <https://doi.org/10.1086/284325>
- García-Alonso, F.-J., García-Valverde, V., Navarro-González, I., Martín-Pozuelo, G., González-Barrio, R., & Periago, M. J. (2020). Chapter 15 - Tomato. In A. K. Jaiswal (Ed.), *Nutritional composition and antioxidant properties of fruits and vegetables* (pp. 255–271). Academic Press. <https://doi.org/10.1016/B978-0-12-812780-3.00015-5>
- Grandillo, S., Termolino, P., & van der Knaap, E. (2013). Molecular Mapping of Complex Traits in Tomato. In *Genetics, genomics, and breeding of tomato* (pp. 190–267). CRC Press.
- Heleno, R., Blake, S., Jaramillo, P., Traveset, A., Vargas, P., & Nogales, M. (2011). Frugivory and seed dispersal in the Galápagos: What is the state of the art? *Integrative Zoology*, 6, 110–129. <https://doi.org/10.1111/j.1749-4877.2011.00236.x>
- Hosmani, P. S., Flores-Gonzalez, M., van de Geest, H., Maumus, F., Bakker, L. V., Schijlen, E., van Haarst, J., Cordewener, J., Sanchez-Perez, G., Peters, S., Fei, Z., Giovannoni, J. J., Mueller, L. A., & Saha, S. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *bioRxiv*: 767764. <https://doi.org/10.1101/767764>
- Huey, R. B., Garland, T. Jr., & Turelli, M. (2019). Revisiting a key innovation in evolutionary biology: Felsenstein's 'phylogenies and the comparative method'. *The American Naturalist*, 193, 755–772. <https://doi.org/10.1086/703055>
- Janson, C. H. (1983). Adaptation of fruit morphology to dispersal agents in a neotropical forest. *Science*, 219, 187–189. <https://doi.org/10.1126/science.219.4581.187>
- Jordano, P. (1995). Angiosperm fleshy fruits and seed dispersers: A comparative analysis of adaptation and constraints in plant-animal interactions. *The American Naturalist*, 145, 163–191. <https://doi.org/10.1086/285735>
- Kamiyoshihara, Y., Miyajima, S., Miyagawa, Y., Moriyama, K., Mizuno, S., Goulet, C., Klee, H., & Tateishi, A. (2020). Functional divergence of principal alcohol o-acyltransferase for biosynthesis of volatile acetate esters among tomato wild species (*Solanum* sect. *Lycopersicon*). *Plant Science: an International Journal of Experimental Plant Biology*, 300, 110612. <https://doi.org/10.1016/j.plantsci.2020.110612>
- Keck, F., Rimet, F., Bouchez, A., & Franc, A. (2016). PhyloSignal: An R package to measure, test, and explore the phylogenetic signal. *Ecology and Evolution*, 6, 2774–2780. <https://doi.org/10.1002/ece3.2051>
- Kimura, S., & Sinha, N. (2008). Tomato (*Solanum lycopersicum*): A model fruit-bearing crop. *Cold Spring Harbor Protocols*, 2008(11), pdb.emo105. <https://doi.org/10.1101/pdb.emo105>
- Labate, J. A., Robertson, L. D., Strickler, S. R., & Mueller, L. A. (2014). Genetic structure of the four wild tomato species in the *Solanum peruvianum* s.l. species complex. *Genome/National Research Council Canada = Genome/Conseil National de Recherches Canada*, 57, 169–180. <https://doi.org/10.1139/gen-2014-0003>

- Levey, D. J., Tewksbury, J. J., Cipollini, M. L., & Carlo, T. A. (2006). A field test of the directed deterrence hypothesis in two species of wild chili. *Oecologia*, 150, 61–68. <https://doi.org/10.1007/s00442-006-0496-y>
- Levin, I., & Schaffer, A. A. (2013). Mapping and Tagging of Simply Inherited Traits. In *Genetics, genomics, and breeding of tomato* (pp. 149–189). CRC Press.
- Lomáscolo, S. B., & Schaefer, H. M. (2010). Signal convergence in fruits: A result of selection by frugivores? *Journal of Evolutionary Biology*, 23, 614–624. <https://doi.org/10.1111/j.1420-9101.2010.01931.x>
- Mack, A. L. (2000). Did fleshy fruit pulp evolve as a defence against seed loss rather than as a dispersal mechanism? *Journal of Biosciences*, 25, 93–97. <https://doi.org/10.1007/BF02985186>
- Malundo, T. M. M., Shewfelt, R. L., & Scott, J. W. (1995). Flavor quality of fresh tomato (*Lycopersicon esculentum* mill.) as affected by sugar and acid levels. *Postharvest Biology and Technology*, 6, 103–110. [https://doi.org/10.1016/0925-5214\(94\)00052-T](https://doi.org/10.1016/0925-5214(94)00052-T)
- Martinez del Rio, C., & Stevens, B. R. (1989). Physiological constraint on feeding behavior: Intestinal membrane disaccharidases of the starling. *Science*, 243, 794–796. <https://doi.org/10.1126/science.2916126>
- Moy, M., Dai, N., Cohen, S., Hadas, R., Granot, D., Petrikov, M., Yeselson, Y., Shen, S., & Schaf, A. A. (2007). The presence of a retrotransposon in the promoter region of the TIV gene encoding for soluble acid invertase distinguishes between the sucrose and hexose accumulating species of *lycopersicon*. *Acta Horticulturae*, 429–436. <https://doi.org/10.17660/ActaHortic.2007.745.28>
- Moyle, L. C. (2008). Ecological and evolutionary genomics in the wild tomatoes (SOLANUM SECT. LYCOPERSICON). *Evolution; International Journal of Organic Evolution*, 62, 2995–3013. <https://doi.org/10.1111/j.1558-5646.2008.00487.x>
- Nakazato, T., Warren, D. L., & Moyle, L. C. (2010). Ecological and geographic modes of species divergence in wild tomatoes. *American Journal of Botany*, 97, 680–693. <https://doi.org/10.3732/ajb.0900216>
- Nevo, O., & Ayasse, M. (2019). Fruit Scent: Biochemistry, Ecological Function, and Evolution. In J.-M. Merillon & K. G. Ramawat (Eds.), *Co-evolution of secondary metabolites* (pp. 1–23). Springer International Publishing. https://doi.org/10.1007/978-3-319-76887-8_33-1
- Nevo, O., Razafimandimby, D., Valenta, K., Jeffrey, J. A. J., Reisdorff, C., Chapman, C. A., Ganzhorn, J. U., & Ayasse, M. (2019). Signal and reward in wild fleshy fruits: Does fruit scent predict nutrient content? *Ecology and Evolution*, 9, 10534–10543. <https://doi.org/10.1002/ece3.5573>
- Nevo, O., Valenta, K., Helman, A., Ganzhorn, J. U., & Ayasse, M. (2022). Fruit scent as an honest signal for fruit quality. *BMC Ecology and Evolution*, 22, 139. <https://doi.org/10.1186/s12862-022-02064-z>
- Nevo, O., Valenta, K., Kleiner, A., Razafimandimby, D., Jeffrey, J. A. J., Chapman, C. A., & Ayasse, M. (2020). The evolution of fruit scent: Phylogenetic and developmental constraints. *BMC Evolutionary Biology*, 20, 138. <https://doi.org/10.1186/s12862-020-01708-2>
- Nevo, O., Valenta, K., Razafimandimby, D., Melin, A. D., Ayasse, M., & Chapman, C. A. (2018). Frugivores and the evolution of fruit colour. *Biology Letters*, 14, 20180377. <https://doi.org/10.1098/rsbl.2018.0377>
- Nevo, O., Valenta, K., Tevlin, A. G., Omeja, P., Styler, S. A., Jackson, D. J., Chapman, C. A., & Ayasse, M. (2017). Fruit defence syndromes: The independent evolution of mechanical and chemical defences. *Evolutionary Ecology*, 31, 913–923. <https://doi.org/10.1007/s10682-017-9919-y>
- Onstein, R. E., Vink, D. N., Veen, J., Barratt, C. D., Flantua, S. G. A., Wich, S. A., & Kissling, W. D. (2020). Palm fruit colours are linked to the broad-scale distribution and diversification of primate colour vision systems. *Proceedings of the Royal Society B*, 287, 20192731. <https://doi.org/10.1098/rspb.2019.2731>
- Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., & Pearse, W. (2013). The caper package: comparative analysis of phylogenetics and evolution in R. *R package version 5*, 1–36. <https://cran.r-project.org/package=caper>
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401, 877–884. <https://doi.org/10.1038/44766>
- Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Pease, J. B., Haak, D. C., Hahn, M. W., & Moyle, L. C. (2016). Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biology*, 14, e1002379. <https://doi.org/10.1371/journal.pbio.1002379>
- Peralta, I. E., Spooner, D. M., & Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; Solanaceae). *Systematic botany monographs: monographic series of the American Society of Plant Taxonomists*, 84. <https://www.jstor.org/stable/25027972>
- Porebski, S., & Bailey, L. G. (1997). Modification of a CTAB DNA extraction protocol for plants. *Plant Molecular Biology Reporter / ISPMB*, 15, 8–15. <https://doi.org/10.1007/BF02772108>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Ramírez-Ojeda, G., Peralta, I. E., Rodríguez-Guzmán, E., Sahagún-Castellanos, J., Chávez-Servía, J. L., Medina-Hinostroza, T. C., Rijalba-Vela, J. R., Vásquez-Núñez, L. P., & Rodríguez-Pérez, J. E. (2021). Edaphoclimatic descriptors of wild tomato species (*Solanum* sect. *Lycopersicon*) and closely related species (*Solanum* sect. *Juglandifolia* and sect. *Lycopersicoides*) in South America. *Frontiers in Genetics*, 12, 748979. <https://doi.org/10.3389/fgene.2021.748979>
- Revell, L. J. (2009). Size-correction and principal components for interspecific comparative studies. *Evolution; International Journal of Organic Evolution*, 63, 3258–3268. <https://doi.org/10.1111/j.1558-5646.2009.00804.x>
- Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things): Phytools: R package. *Methods in Ecology and Evolution / British Ecological Society*, 3, 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.2.9, <https://CRAN.R-project.org/package=psych>
- Rick, C. M. (1964). Some plant–animal relations on the Galápagos Islands. In R. Bowman (Ed.), *The Galápagos: Proceedings of the symposia of the Galápagos international scientific project* (pp. 215–224). University of California Press.
- Rick, C. M., & Bowman, R. I. (1961). Galápagos Tomatoes and Tortoises. *Evolution; International Journal of Organic Evolution*, 15, 407–417. <https://doi.org/10.2307/2406309>
- Ridley, H. (1930). *The dispersal of plants throughout the world*. Reeve and Ashford.
- Rojas, T. N., Zampini, I. C., Isla, M. I., & Blendinger, P. G. (2022). Fleshy fruit traits and seed dispersers: Which traits define syndromes? *Annals of Botany*, 129, 831–838. <https://doi.org/10.1093/aob/mcab150>
- Schaefer, H. M., Valido, A., & Jordano, P. (2014). Birds see the true colours of fruits to live off the fat of the land. *Proceedings. Biological Sciences / the Royal Society*, 281, 20132516. <https://doi.org/10.1098/rspb.2013.2516>
- Sinnott-Armstrong, M. A., Donoghue, M. J., & Jetz, W. J. (2021). Dispersers and environment drive global variation in fruit colour syndromes. *Ecology Letters*, 24, 1387–1399. <https://doi.org/10.1111/ele.13753>
- Sinnott-Armstrong, M. A., Downie, A. E., Federman, S., Valido, A., Jordano, P., & Donoghue, M. J. (2018). Global geographic patterns in the colours and sizes of animal-dispersed fruits. *Global Ecology and Biogeography: a Journal of Macroecology*, 27, 1339–1351. <https://doi.org/10.1111/geb.12801>

- Sinnott-Armstrong, M. A., Lee, C., Clement, W. L., & Donoghue, M. J. (2020). Fruit syndromes in viburnum: Correlated evolution of color, nutritional content, and morphology in bird-dispersed fleshy fruits. *BMC Evolutionary Biology*, 20, 7. <https://doi.org/10.1186/s12862-019-1546-5>
- Sinnott-Armstrong, M. A., Middleton, R., Ogawa, Y., Jacucci, G., Moyroud, E., Glover, B. J., Rudall, P. J., Vignolini, S., & Donoghue, M. J. (2023). Multiple origins of lipid-based structural colors contribute to a gradient of fruit colors in viburnum (Adoxaceae). *The New Phytologist*, 237, 643–655. <https://doi.org/10.1111/nph.18538>
- Stam, R., Nosenko, T., Hörger, A. C., Stephan, W., Seidel, M., Kuhn, J. M. M., Haberer, G., & Tellier, A. (2019). Thede NovoReference genome and transcriptome assemblies of the wild tomato *Solanum chilense* Highlights birth and death of NLR genes between tomato species. *G3: Genes, Genomes, Genetics*, 9, 3933–3941. <https://doi.org/10.1534/g3.119.400529>
- Sun, T., & Li, L. (2020). Toward the 'golden' era: The status in uncovering the regulatory control of carotenoid accumulation in plants. *Plant Science: an International Journal of Experimental Plant Biology*, 290, 110331. <https://doi.org/10.1016/j.plantsci.2019.110331>
- Särkinen, T., Bohs, L., Olmstead, R. G., & Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree. *BMC Evolutionary Biology*, 13, 214. <https://doi.org/10.1186/1471-2148-13-214>
- Tieman, D., Zhu, G., Resende, M. F. R. Jr., Lin, T., Nguyen, C., Bies, D., Rambla, J. L., Beltran, K. S. O., Taylor, M., Zhang, B., Zhu, G., Resende, M. F. R. Jr., Lin, T., Nguyen, C., Bies, D., Rambla, J. L., Beltran, K. S. O., Taylor, M., Zhang, B., ... Klee, H. (2017). A chemical genetic roadmap to improved tomato flavor. *Science*, 355, 391–394. <https://doi.org/10.1126/science.aal1556>
- Valenta, K., & Nevo, O. (2020). The dispersal syndrome hypothesis: How animals shaped fruit traits, and how they did not (M Schleuning, Ed.). *Functional Ecology*, 9, 1. <https://doi.org/10.1111/1365-2435.13564>
- Valenta, K., Bhramdat, H. D., Calhoun, G. V., Daegling, D. J., & Nevo, O. (2022). Variation in ripe fruit hardness: A mechanical constraint? *Oikos*, 2022. <https://doi.org/10.1111/oik.08074>
- Valenta, K., Kalbitzer, U., Razafimandimby, D., Omeja, P., Ayasse, M., Chapman, C. A., & Nevo, O. (2018). The evolution of fruit colour: Phylogeny, abiotic factors and the role of mutualists. *Scientific Reports*, 8, 14302. <https://doi.org/10.1038/s41598-018-32604-x>
- Valido, A., Schaefer, H. M., & Jordano, P. (2011). Colour, design and reward: Phenotypic integration of fleshy fruit displays. *Journal of Evolutionary Biology*, 24, 751–760. <https://doi.org/10.1111/j.1420-9101.2010.02206.x>
- van der Pijl, L. (1969). *Principles of dispersal in higher plants*. Springer-Verlag.
- Whitehead, S. R., Quesada, M. F. O., & Bowers, M. D. (2016). Chemical tradeoffs in seed dispersal: Defensive metabolites in fruits deter consumption by mutualist bats. *Oikos*, 125, 927–937. <https://doi.org/10.1111/oik.02210>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer. <https://doi.org/10.1007/978-3-319-24277-4>
- Willson, M. F., & Whelan, C. J. (1990). The evolution of fruit color in fleshy-fruited plants. *The American Naturalist*, 136, 790–809. <https://doi.org/10.1086/285132>
- Zhang, Y., & Fernie, A. R. (2018). On the role of the tricarboxylic acid cycle in plant productivity. *Journal of Integrative Plant Biology*, 60, 1199–1216. <https://doi.org/10.1111/jipb.12690>
- Zhao, Y., Cao, H., Xu, W., Chen, G., Lian, J., Du, Y., & Ma, K. (2018). Contributions of precipitation and temperature to the large scale geographic distribution of fleshy-fruited plant species: Growth form matters. *Scientific Reports*, 8, 17017. <https://doi.org/10.1038/s41598-018-35436-x>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Barnett, J. R., Sharma, R., Buonauro, G., Gillis, I. M., Rashidzade, M., & Caicedo, A. L. (2023). Evidence of fruit syndromes in the recently diverged wild tomato clade opens new possibilities for the study of fleshy fruit evolution. *Plants, People, Planet*, 1–15. <https://doi.org/10.1002/ppp3.10399>