Received: 13 April 2023 | Accepted: 13 July 2023

DOI: 10.1002/ajb2.16223

American Journal of Botany

RESEARCH ARTICLE

Variation in ripe fruit volatiles across the tomato clade: An evolutionary framework for studying fruit scent diversity in a crop wild relative

Jacob R. Barnett[1] | **Denise M. Tieman**[2] | **Ana L. Caicedo**[3]

[1]Graduate Program in Organismic and Evolutionary Biology, University of Massachusetts Amherst, MA 01003, USA

[2]Horticultural Sciences Department, University of Florida, Gainesville, FL 32611, USA

[3]Department of Biology, University of Massachusetts Amherst, MA 01003, 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

Abstract

Premise: The scents of volatile organic compounds (VOCs) are an important component of ripe fleshy fruit attractiveness, yet their variation across closely related wild species is poorly understood. Phylogenetic patterns in these compounds and their biosynthetic pathways offer insight into the evolutionary drivers of fruit diversity, including whether scent can communicate an honest signal of nutrient content to animal dispersers. We assessed ripe fruit VOC content across the tomato clade (*Solanum* sect. *Lycopersicon*), with implications for crop improvement since these compounds are key components of tomato flavor.

Methods: We analyzed ripe fruit volatiles from 13 species of wild tomato grown in a common garden. Interspecific variations in 66 compounds and their biosynthetic pathways were assessed in 32 accessions, and an accession-level phylogeny was constructed to account for relatedness.

Results: Wild tomato species can be differentiated by their VOCs, with *Solanum pennellii* notably distinct. Phylogenetic conservatism exists to a limited extent. Major cladewide patterns corresponded to divergence of the five brightly colored-fruited species from the nine green-fruited species, particularly for nitrogen-containing compounds (higher in colored-fruited) and esters (higher in green-fruited), the latter appearing to signal a sugar reward.

Conclusions: We established a framework for fruit scent evolution studies in a crop wild relative system, showing that each species in the tomato clade has a unique VOC profile. Differences between color groups align with fruit syndromes that could be driven by selection from frugivores. The evolution of colored fruits was accompanied by changes in biosynthetic pathways for esters and nitrogen-containing compounds, volatiles important to tomato flavor.

KEYWORDS

chemical diversity, crop improvement, crop wild relatives, fleshy fruit evolution, fruit scent, interspecific variation, secondary metabolites, *Solanaceae*, *Solanum lycopersicum* (tomato), tomato flavor

Plants produce a vast array of secondary metabolites that perform many important functions. For fleshy-fruited plants, the scents of volatile organic compounds (VOCs) in ripe fruits are generally thought to help attract seed dispersers by making mature fruits noticeable and appealing to animal consumers, while also aiding in defense against pathogens and seed predators (Rodríguez et al., 2013; Nevo and Ayasse, 2019). Hundreds of different VOCs can be present in fleshy fruits, although this array of compounds is produced by relatively few main biosynthetic pathways (Jiang and Song, 2010). From the human perspective, VOCs are key components of fruit flavor in agricultural crops (Klee and Tieman, 2018). Assessing variation in these compounds across species is valuable to evolutionary biologists studying the drivers of fruit diversity and to plant breeders working to improve fruit flavor.

Researchers have been exploring the evolutionary forces behind fruit trait diversity for decades, but have only recently begun identifying and quantifying the VOCs comprising wild fruit scent (Nevo and Ayasse, 2019).

Am J Bot. 2023;110:e16223.
https://doi.org/10.1002/ajb2.16223

Fruit-eating animal preferences have long been considered an important driver because plants better at attracting fruit consumers improve their seed dispersal (Ridley, 1930; van der Pijl, 1969; Janson, 1983; Valenta and Nevo, 2020). In line with this hypothesis, VOCs are important for attracting nonhuman seed dispersers and signaling fruit ripeness (Nevo and Ayasse, 2019). Evidence that fruit scent chemicals evolved in tandem with consumption by animal dispersers has come from recent work on frugivorous bats (Hodgkison et al., 2013; Santana et al., 2021) and lemurs (Nevo et al., 2018). A mechanism underlying this fruit–frugivore coevolution could be that fruit scent communicates an honest signal of internal nutrient content to potential animal consumers (Nevo et al., 2019, 2020a). At the same time, it is important to consider the possibility that some fruit VOCs serve a defensive rather than attractive function (Nevo et al., 2017). Additionally, fruit scent may be constrained by non-adaptive chemical, developmental, or phylogenetic limitations on VOC production (Nevo et al., 2020b). The degree of phylogenetic constraint, or conservatism (i.e., similarity due to inheritance from common ancestors), appears to differ depending on the system, with some multi-genera community-based studies not finding much conservatism (Nevo et al., 2018, 2020b) and other narrower lineage-focused studies either finding it in *Ficus* (Hodgkison et al., 2013) or not finding it in *Piper* (Santana et al., 2021). Despite recent interest in wild fruit VOCs, the field is still young, and data only exists for a small number of non-agricultural systems.

The VOCs of agricultural fruits have received considerable attention due to the economic importance of fruit flavor (Klee and Tieman, 2018), but the wild relatives of these crop fruits have been far less studied and usually only in the context of domestication or crop improvement. A crop wild relative system has yet to be surveyed as a framework for studying wild fruit scent evolution, presenting an exciting opportunity to understand the evolutionary drivers of VOC composition in a system with direct relevance for agriculture. Furthermore, few studies of interspecies VOC variation have employed a common garden (Schwinning et al., 2022) in which all plants are grown in the same conditions to ensure that differences are due mainly to genetics rather than environment. This controlled approach is particularly powerful in a crop wild relative system because the genetic, biochemical, developmental underpinnings of fruit VOCs are better studied than in most wild systems.

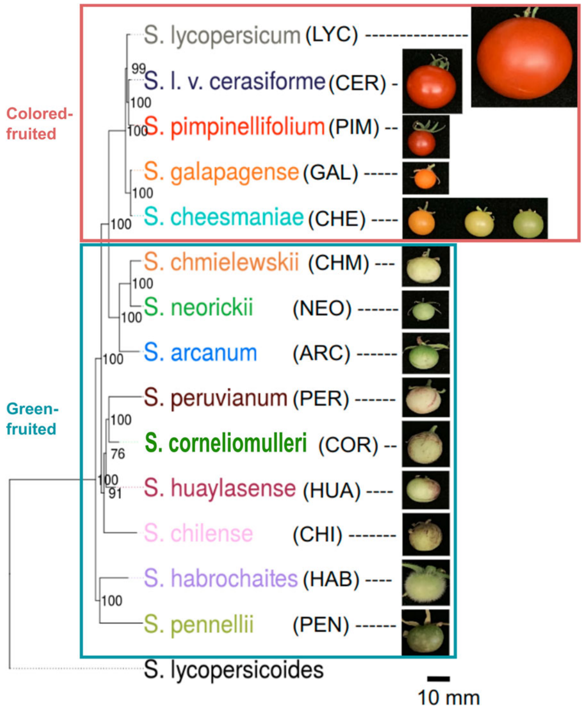


FIGURE 1 Phylogenetic relationships among species in the wild tomato clade (*Solanum* sect. *Lycopersicon*), with *Solanum lycopersicoides* included as outgroup. Maximum likelihood consensus tree based on 64,745 SNP markers obtained through genotyping by sequencing and constructed in IQ-TREE with 1000 ultrafast bootstraps using the TVM+F+ASC+R4 nucleotide substitution model selected by ModelFinder with ascertainment bias correction and SH-aLRT likelihood ratio test (Barnett et al., 2023). Numbers at nodes represent bootstrap percentages. Three-letter species abbreviations and species color-coding are used throughout this paper. Photographs show ripe fruits of each species; three photographs of *Solanum cheesmaniae* are included to show the different colors of the three accessions used in this study.

One system well-positioned to fill this gap in studies of the evolution of wild fruit scent is the tomato clade, a group of 14 species (*Solanum* sect. *Lycopersicon*, family *Solanaceae*, Peralta et al., 2008) composed of the cultivated tomato (*Solanum lycopersicum* var. *lycopersicum*) and 13 species of wild relatives (Figure 1). The clade is a monophyletic group estimated to have diverged from a common ancestor 2–2.5 million years ago based on fossil-calibrated molecular clock evidence (Särkinen et al., 2013; Pease et al., 2016). The 13 wild tomato species inhabit a range of environments in western South America (Appendix S1, Figure S1), from deserts to forests and coastlines to high mountains. Two informal subgroups of species within the clade are generally recognized based on ripe fruit color: (1) a monophyletic brightly colored-fruited group consisting of the red *Solanum pimpinellifolium*, red *Solanum lycopersicum* var. *cerasiforme*, orange *Solanum galapagense*, and yellowish/orange *Solanum cheesmaniae* and (2) a paraphyletic green-fruited group encompassing the remaining nine species. Wild tomatoes offer a valuable source of genetic diversity for crop improvement because most species can be interbred with each other and with cultivated tomatoes (Baek et al., 2015). The tomato system has received much attention as a model of fruit development and genetics (Kimura and Sinha, 2008), enabling a more in-depth consideration of the molecular basis of fruit VOCs than is possible with most other fruit systems.

In the cultivated tomato, more than 400 VOCs have been detected (Petro-Turza, 1986; Baldwin et al., 2000), although only about 20 of these compounds have been identified as major contributors to human flavor preferences (Martina et al., 2021). The four main categories of

VOCs in tomatoes are based on the precursor molecules lipids, carotenoids/terpenes, non-aromatic amino acids (leucine and isoleucine), and phenylalanine (Colantonio et al., 2022). These categories are consistent with those commonly used for fleshy fruit VOCs in general (Rodríguez et al., 2013; Nevo and Ayasse, 2019). The ~20 compounds considered to have a strong impact on cultivated tomato flavor come from across the four precursor categories and impart aromas including green/viney/earthy, nutty/stale, and fruity/floral/sweet/tropical (Baldwin et al., 2008). In terms of chemical classification, VOCs can also be grouped into esters, alcohols, aldehydes, ketones, lactones, and terpenoids (Jiang and Song, 2010). Among these chemical classes, esters are particularly important for attracting animal dispersers to wild fruits due to their fruity/floral aromas (Rodríguez et al., 2013), which are especially enticing to scent-oriented mammals (Peris et al., 2017) including lemurs (Nevo et al., 2018) and elephants (Nevo et al., 2020). Interestingly, while humans respond positively to ester aromas in certain crops such as apple, banana, citrus, and strawberry (Jiang and Song, 2010), esters have a negative impact on human flavor preference in cultivated tomatoes (Klee and Tieman, 2018). Previous work on three of the nine green-fruited wild tomato species showed they produced higher percentages of acetate esters than the two red-fruited species, with intermediate levels produced in the two yellow/orange species (Goulet et al., 2012). However, a survey of ripe fruit VOCs across all 13 wild species has not yet been carried out. Given this preliminary evidence of interspecies differences in disperser-relevant scent compounds, establishing a complete evolutionary framework of VOC variation for all species in the clade would enable hypothesis testing without making assumptions about the frequency of evolutionary changes.

In this study, we present a novel framework for wild fruit scent evolution research by surveying variation in ripe fruit volatiles across all 13 wild tomato species in a common garden. The combination of a recently diverged, phylogentically tight clade and a common growing environment distinguishes this study from many others that used a phylogenetically broader community approach and a wild setting. By quantifying variations in individual VOCs and their categories across the tomato clade, our objectives were to (1) examine which compounds and biosynthetic pathways are most important for differentiating species and determine the extent of phylogenetic conservatism in the clade, (2) assess whether fruit scent variation aligns with two previously hypothesized wild tomato fruit syndromes (colored-fruited vs. green-fruited) that could have been driven by disperser preferences (Barnett et al., 2023), and (3) explore whether scent could communicate an honest signal to animal dispersers by testing for an association between VOCs and fruit sugar content. We predicted each species would display a distinct VOC profile and that the colored-fruited group would differ from the green-fruited group, particularly for disperser-relevant ester compounds and VOCs important to human-perceived flavor, as could be expected if animal preferences were an evolutionary driver of fruit trait diversity.

MATERIALS AND METHODS

Study species and plant material

Our sampling included 32 accessions chosen to span the known phylogenetic and geographic diversity across the tomato clade (*Solanum* sect. *Lycopersicon*), which consists of 14 taxa: 12 species of fully wild relatives, the semi-wild *S. lycopersicum* var. *cerasiforme*, and the cultivated tomato variety Ailsa Craig (Figure 1; Appendix S2, Table S1). Three accessions of each taxon representing different parts of their geographic ranges were initially selected and planted (Appendix S1, Figure S1), but not all accessions produced enough fruit for VOC quantification; thus, for some taxa, VOCs for only one or two accessions were measured. 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 Appendix S2, Table S1). Each accession is representative of an independently sampled population in the wild. For autogamous (self-pollinating) self-compatible species (*S. cheesmaniae*, *S. galapagense*, some *S. pimpinellifolium*, *S. lycopersicum* var. *cerasiforme*, *Solanum neorickii*), each plant of an accession is likely genetically identical. For others, plants are not identical, because 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% v/v sodium hypochlorite solution (as per the 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, Massachusetts, where they were irrigated twice a week and fertilized once a week via drip lines. For each accession, three individual plants were grown in different randomized locations. Our study plants grew throughout the summer farm season (May–September 2020) and were pruned as needed. Self-incompatible accessions were hand-pollinated several times during the first few weeks of flowering to facilitate fruit set, and all flowers were accessible to natural pollinators.

Fruit volatile collection

Ripe fruits were collected on three separate days (3 and 31 August, 14 September, 2020). Fruit ripeness was determined qualitatively for each species based on a combination of fruit or calyx color change, softening, and/or falling off at the touch (Appendix S2, Table S2). Harvested fruits were packed in bubble wrap in ventilated plastic clamshell containers, and shipped overnight from Massachusetts to the University of Florida for VOC analysis. For most accessions, fruits from several plants were pooled to get as close as possible to 100 g of fruit per sample, although for some small-fruited taxa such as *S. galapagense*, the total mass was as low as 8 g. A total of 66 samples representing 14 taxa and 32 accessions were collected (Appendix S2, Table S3).

GC-MS quantification of VOCs

Samples were processed at the University of Florida and analyzed with gas chromatography–mass spectrometry (GC-MS) for compound identification as detailed by Tieman et al. (2006). They were also quantified using a gas chromatograph equipped with a flame ionization detector. Briefly, air that was filtered through a hydrocarbon trap was passed over chopped fruit in glass tubes for 1 h with the aid of a vacuum pump to collect volatiles on a Super Q column (Hayes Separations, Bandera, TX, USA). The trapped volatiles were eluted with methylene chloride after adding nonyl acetate as an internal standard and analyzed on an Agilent (Palo Alto, CA, USA) 6890 N gas chromatograph equipped with an Agilent 30 m DB-5 column. Peaks were quantified using Agilent GC ChemStation Rev. A.10.02 [1757] software based on retention times compared to known standards. Compound identity was confirmed for each accession via mass spectrometry using Agilent MSD Productivity ChemStation v.D.03.xx software and the Agilent NIST05 MS Library database/MS Search Program v.2.0d.

A total of 66 compounds with clear and consistent peaks in most samples were chosen to be quantified (Appendix S2, Table S4). Most of these compounds were known to be present in cultivated tomatoes, although we also included additional compounds not usually found in cultivated tomatoes that were identified in multiple wild species samples. Peak areas were adjusted relative to the nonyl acetate standard and converted to nanograms per grams of fruit weight per hour (ng $\text{gfw}^{-1}\text{h}^{-1}$).

Compounds were categorized based on their precursor molecules as described by Colantonio et al. (2022). For each category of compounds, content values for all the compounds within a category were added together to get a category sum. Category percentages were computed by dividing the category sum by the sum of all 66 compounds.

Data analyses

Analyses were carried out in R v4.2.2 (R Core Team, 2022) and RStudio Desktop v2022.12.0.353 (Posit Team, 2022), with plots produced using the R packages GGPLOT2 v3.4.0 (Wickham, 2016) and PHEATMAP v1.0.12 (Kolde, 2019).

Differentiating species and assessing phylogenetic patterns

We performed a principal component analysis (PCA) on centered and scaled accession mean values of the 66 VOC compounds using the function prcomp from the R package STATS, plotting the first two PC axes with GGPLOT2.

To determine which VOCs were most important for differentiating species, we quantified the random forest mean decrease accuracy for all 66 compounds using MetaboAnalyst 5.0 (Pang et al., 2021). Data were log (base 10)-transformed, but data scaling and sample normalization were not used. Only 61 of 66 samples (representing 11 of 14 species) were used for this analysis because *Solanum chilense*, *Solanum huaylasense*, and *Solanum peruvianum* had fewer than three samples. In addition to conducting the analysis with all 11 species included, the same quantifications were made for the five colored-fruited species only and the six green-fruited species only to determine which compounds were the best species differentiators within each of the two color groups.

We used an accession-level phylogenetic tree from Barnett et al. (2023) to place VOC variation in a phylogenetic context. To quantify the degree of phylogenetic conservatism in VOC compounds and categories, 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 compound or category.

To test the difference in ester content between colored-fruited and green-fruited accessions, phylogenetic generalized least squares (PGLS) models with color group as the predictor variable and ester sum (or percentage) as the response variable were produced with the functions comparative.data and pgls 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).

Associations between VOCs and fruit sugar content

We examined potential associations between VOCs and nutrient content using sugar data from Barnett et al. (2023). Briefly, concentrations (in mg/mL) of glucose, fructose, and sucrose were quantified from liquid extracts of ripe fruit pericarps using absorbance-based assay kits (Megazyme, Bray, Ireland); "total sugars" represents the sum of those three concentrations. Raw (i.e., nontransformed) accession mean values for VOC categories and sugars were used for analyses in R. Plots were produced with ggplot2, and Pearson's r was calculated with the built-in R function. Phylogenetic generalized least squares (PGLS) models, with VOC category (ester, terpene, or apocarotenoid) sum (or percentage) as the predictor variable and total sugars as the response variable, were produced with CAPER with lambda = "ML".

RESULTS

Extent of variation in VOCs across the tomato clade

To establish a framework for studying fruit scent evolution across the tomato clade, we quantified 66 of the VOCs found in fruit from *S. lycopersicum* var. *lycopersicum* and 13 related wild tomato species. To assess how different biosynthetic pathways affected patterns of variation in these compounds, we categorized VOCs by their biosynthetic precursors: 31 lipid-derived, 7 carotenoid/terpene-derived, 12 non-aromatic amino acid (leucine and isoleucine)-derived, 10 phenylalanine-derived, and 6 other (Appendix S2, Table S4). In addition to labeling the precursor category for each compound, we labeled 17 of the 66 compounds as esters since multiple precursor types (lipid, non-aromatic amino acid, phenylalanine, and other) can be esterified. Esters are a group of interest because they tend to have fruity or floral scents and are known to be attractive to vertebrate dispersers (Peris et al., 2017), so may be subject to selective pressure from animals. Also of note from a human perspective, fifteen of the compounds we quantified (Appendix S2, Table S5) are considered key components of cultivated tomato aroma (Martina et al., 2021); we did not quantify all 21 on that list because not all of those compounds had clear, consistently detectable peaks in our GC-MS samples. Compounds not usually found in cultivated tomatoes that showed clear peaks in multiple wild species samples were the non-aromatic, amino-acid-derived compounds acetoin and 2-propylthiazole; the terpene-derived compounds alpha-pinene, alpha-curcumene, and zingiberene; and the other-derived compound 2-pentylfuran.

Across all 66 samples and all 66 compounds, amounts varied from not detected (a total of 53 compounds had a 0 value in at least one sample) to 688.54 ng gfw^{-1}h^{-1} (zingiberene) (Appendix S2, Table S3). Note that for zingiberene, *Solanum habrochaites* LA2329 was the only accession with such high levels; the second highest accession was *S. cheesmaniae* LA0428 with 24.1 ng gfw^{-1}h^{-1}. The compounds with the next three highest maximum values were 3-methyl-1-butanol (614.73 ng gfw^{-1}h^{-1} in *S. galapagense* LA0528), 2-methyl-1-butanol (532.72 ng gfw^{-1}h^{-1} in *S. neorickii* LA1626), and hexyl alcohol (410.74 ng gfw^{-1}h^{-1} in *S. neorickii* LA2200); all three of these compounds have odors described as fusel, alcoholic, and sweet (Appendix S2, Table S4). On the other end of the spectrum, the compounds with the three lowest maximum values were 2-ethyl-1-hexanol (0.669 ng gfw^{-1}h^{-1} in *Solanum arcanum* LA0441), phenethyl acetate (1.489 ng gfw^{-1}h^{-1} in *S. galapagense* LA0528), and 3-methyl-2-butenal (1.773 ng gfw^{-1}h^{-1} in *S. peruvianum* LA1474).

Only 13 of the 66 compounds were detected in all 66 samples; 12 of them were either derived from lipids or non-aromatic amino acids, and none were esters. Included in these 12 were the four least variable compounds in terms of coefficient of variation (CV): 3-pentanone (mean 12.565 ng gfw^{-1}h^{-1}, 0.622 CV), 2-methylbutyraldehyde (mean 21.941 ng gfw^{-1}h^{-1}, 0.669 CV), 1-penten-3-ol (mean 9.362 ng gfw^{-1}h^{-1}, 0.734 CV), and hexanal (mean 66.965 ng gfw^{-1}h^{-1}, 0.921 CV). Compound summary statistics are listed in Appendix S2, Table S4.

Compounds and biosynthetic pathways differentiating species and the extent of phylogenetic conservatism

To visualize how well species could be differentiated by levels of the 66 compounds, we performed a PCA with sample averages for the 32 accessions as points, drawing color-coded polygons around accessions from the same species (Figure 2; Appendix S1, Figure S2). Most species occupied a distinct area of the PCA space, with some overlap among the green-fruited *Solanum chmielewskii*, *S. neorickii*, *S. arcanum*, *Solanum corneliomulleri*, and *S. chilense*. Noticeably separate from all other species was *Solanum pennellii*.

To determine which individual VOCs were most important for differentiating species, we quantified the random forest mean decrease accuracy for all 66 compounds. A total of 17 compounds had scores above 0.01 (Figure 3; Appendix S2, Table S6) and were thus considered major differentiators among species; these compounds included representatives from all five precursor molecule categories, seven of them were esters, and one was nitrogen-containing.

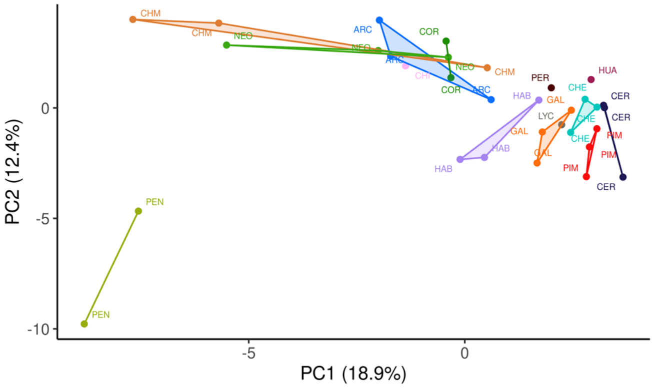


FIGURE 2 Wild tomato species are largely differentiated by their volatile organic compound (VOC) profiles, as shown by a PCA of centered and scaled raw values (in ng gfw^{-1}h^{-1}) for all 66 compounds. Produced with prcomp function in R. Points represent accession averages (32 total), color-coded by species (14 total). ARC = *Solanum arcanum*, CER = *S. lycopersicum* var. *cerasiforme*, CHE = *S. cheesmaniae*, CHI = *S. chilense*, CHM = *S. chmielewskii*, COR = *S. corneliomulleri*, GAL = *S. galapagense*, HAB = *S. habrochaites*, HUA = *S. huaylasense*, LYC = *S. lycopersicum* var. *lycopersicum*, NEO = *S. neorickii*, PEN = *S. pennellii*, PER = *S. peruvianum*, PIM = *S. pimpinellifolium*. Points from the same species are connected to facilitate viewing of species differentiation. See Figure S2 in Appendix S1 for a version of this figure with accession labels for each point.

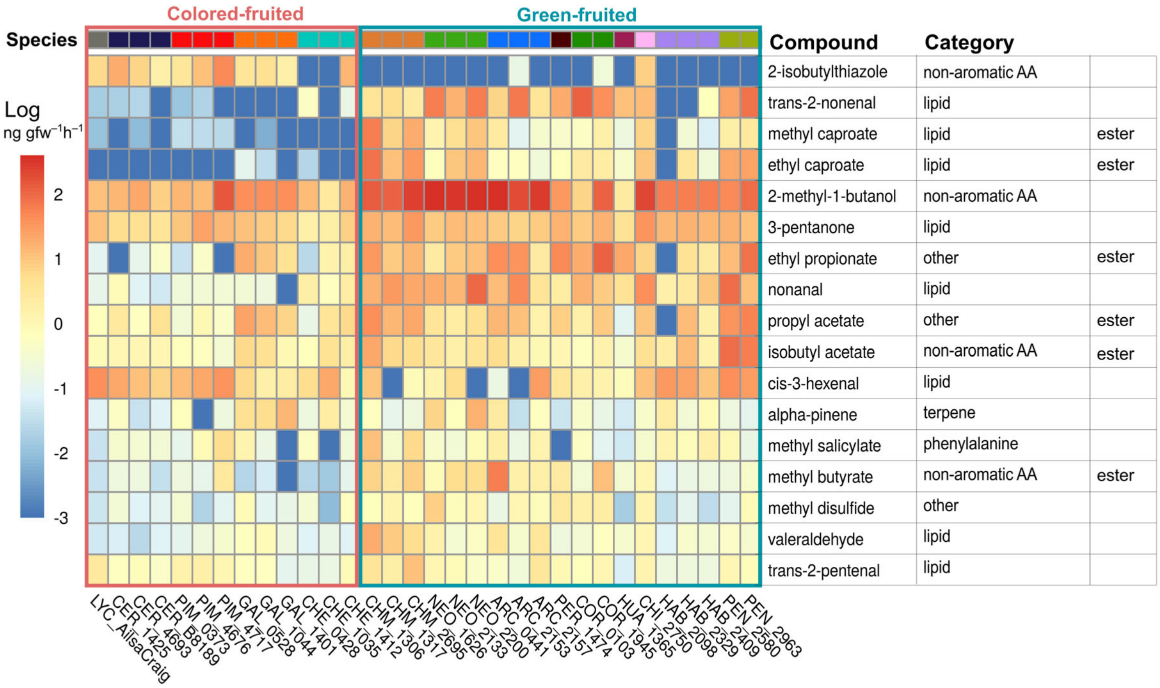


FIGURE 3 Heatmap of the 17 compounds with the greatest contribution to differentiating species based on random forest mean decrease accuracy values greater than 0.01 (Appendix S2, Table S6). See Table S4 in Appendix S2 for details on the compounds. Categories are based on the compound's precursor molecule; esters are also indicated. Values were log (base 10)-transformed but not scaled. Columns represent accession averages, with color-coding along the top denoting species (see Figure 1). Accessions are arranged phylogenetically based on a tree from Barnett et al. (2023), with brightly colored-fruited species on the left and green-fruited species on the right. ARC = *Solanum arcanum*, CER = *S. lycopersicum* var. *cerasiforme*, CHE = *S. cheesmaniae*, CHI = *S. chilense*, CHM = *S. chmielewskii*, COR = *S. corneliomulleri*, GAL = *S. galapagense*, HAB = *S. habrochaites*, HUA = *S. huaylasense*, LYC = *S. lycopersicum* var. *lycopersicum*, NEO = *S. neorickii*, PEN = *S. pennellii*, PER = *S. peruvianum*, PIM = *S. pimpinellifolium*. Numbers after 3-letter species abbreviations are Tomato Genetics Resource Center (TGRC) "LA" accession numbers, except for CER_B8189 (full accession ID BGV008189), which does not have an "LA" number.

When the sum of all 66 compound levels was calculated for each species, *S. neorickii* and *S. pennellii* had the two highest totals (Figure 4); interestingly, these two species are not closely related. On the other end of the spectrum, the more closely related *S. huaylasense* and *S. peruvianum* produced the lowest sum totals. All four of these species are green-fruited; the five monophyletic colored-fruited species all had intermediate sums.

To explore whether particular biosynthetic pathways were important drivers of VOC differences across the clade, we compared each compound category's sum and percentage (category sum divided by sum of all compounds) among species. Most species showed a similar partitioning of VOC categories in terms of sums and percentages (Figure 4), with lipid-derived and non-aromatic amino-acid-derived being the two major categories in all species. *Solanum habrochaites* was unique for its high level of carotenoid/terpene compounds, mainly driven by very high amounts of the sesquiterpenes alpha-curcumene and zingiberene in one accession (LA2329).

In terms of esters (a category that includes compounds from multiple precursor categories), *S. huaylasense* stood out for having the highest percentage of esters, despite producing the lowest total VOC sum of any species (Figure 4). *Solanum pennellii* was notable for high values of both sum and percentage of esters. Interestingly, *S. neorickii* had the lowest percentage of esters for a green-fruited species, while its sister species *S. chmielewskii* had one of the highest percentages. Among colored-fruited species, the two Galápagos Islands endemics *S. galapagense* and *S. cheesmaniae* produced more esters than the red-fruited mainland species.

The similarities in distantly related species and differences in closely related species suggest that ripe fruit scent does not show much phylogenetic conservatism. Phylogenetic signal results also support the conclusion that conservatism is limited, because only five of 66 compounds had strong (>0.7) and 12 of 66 compounds had moderate (>0.4 and <0.7) values for both Blomberg's *K* and Pagel's lambda (Appendix S2, Table S4). Interestingly, four of the five high-signal and five of the 12 moderate-signal compounds were esters. Of the seven compound categories tested, only "ester" and "other" had at least moderate (>0.4) values for both *K* and lambda.

Associations between VOCs and putative dispersal syndromes

There were stark differences between the colored-fruited and green-fruited species groups for a subset of the

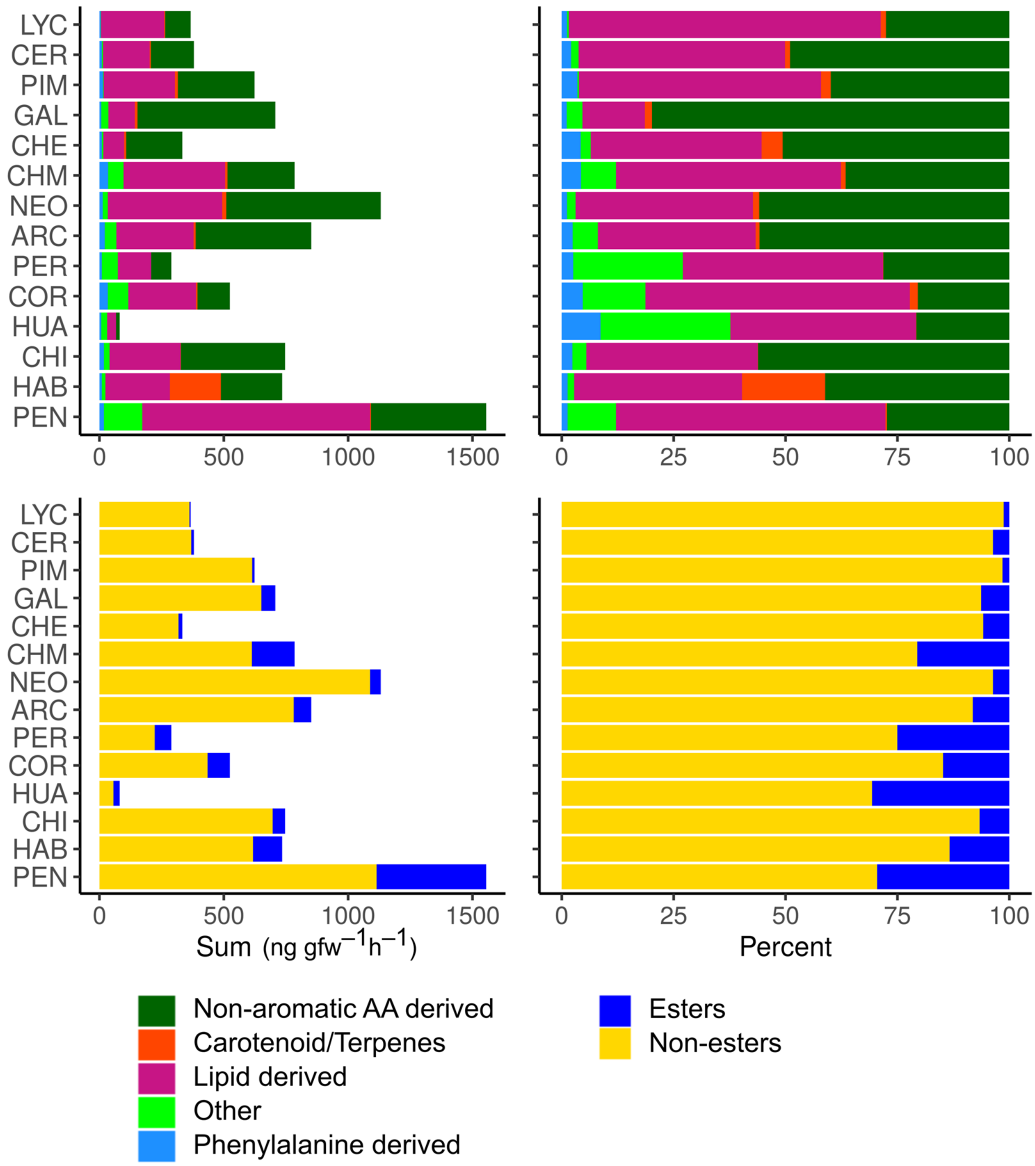


FIGURE 4 Biosynthetic pathways drive differences in volatile organic compounds (VOCs) across the wild tomato clade, as shown by stacked bar plots of compound categories based on either precursor molecules (top row) or as esters/non-esters (bottom row) for 14 species. ARC = *Solanum arcanum*, CER = *S. lycopersicum* var. *cerasiforme*, CHE = *S. cheesmaniae*, CHI = *S. chilense*, CHM = *S. chmielewskii*, COR = *S. corneliomulleri*, GAL = *S. galapagense*, HAB = *S. habrochaites*, HUA = *S. huaylasense*, LYC = *S. lycopersicum* var. *lycopersicum* NEO = *S. neorickii*, PEN = *S. pennellii*, PER = *S. peruvianum*, PIM = *S. pimpinellifolium*. Species are arranged phylogenetically along the *y*-axis with cultivated tomato at the top (see Figure 1). Species values were computed by first calculating accession averages from 66 samples and then averaging those accession averages (32 total). Left plots show category sums of compound values. Right plots show category percentages of the sum of all 66 compounds.

species-differentiating compounds described above, in line with the putative fruit syndromes known to distinguish these two groups for other fruit traits (Barnett et al., 2023). In particular, the nitrogen-containing 2-isobutylthiazole was high in colored-fruited but low in green-fruited species, while the esters methyl caproate and ethyl caproate and the lipid-derived *trans*-2-nonenal were low in colored-fruited but high in green-fruited species (Figure 3).

We quantified random forest mean decrease accuracy separately for just the colored-fruited and just the green-fruited species to determine whether certain compounds were more important for differentiating species within each group. These additional analyses revealed that valeraldehyde (lipid-derived), *cis*-3-hexenal (lipid-derived), and methyl salicylate (phenylalanine-derived) were the top three differentiators among colored-fruited species only, while 1-pentanol (lipid-derived), *trans*-2-nonenal (lipid-derived), and 2-methyl-1-butanol (non-aromatic amino acid-derived) were the top three for green-fruited species only (Appendix S2, Table S6). The only compound that was a major differentiator within both color groups was 3-pentanone (lipid-derived). Among the colored-fruited species, *S. galapagense* stood out with high levels of alpha-pinene, ethyl propionate, and propyl acetate. Notable green-fruited species were *S. chmielewskii* with high levels of *trans*-2-pentenal, methyl caproate, ethyl caproate, methyl butyrate, and valeraldehyde; *S. neorickii* with high levels of methyl caproate, ethyl caproate, *trans*-2-nonenal, methyl disulfide, methyl butyrate, 2-methyl-1-butanol, and alpha-pinene; and *S. pennellii* with high levels of ethyl caproate, *trans*-2-nonenal, isobutyl acetate, and propyl acetate.

For a finer-grained look at interspecies patterns among biochemical precursor categories, we also examined how individual compounds within each category or biosynthetic pathway varied across accessions (Appendix S1, Figure S3). Apocarotenoids were generally higher in colored-fruited accessions, while terpenes were higher in green-fruited accessions (Appendix S1, Figure S3a). Among non-aromatic amino acid-derived compounds, the nitrogen-containing 2-propylthiazole and 2-isobutylthiazole were mostly absent from green-fruited accessions (Appendix S1, Figure S3b); these two compounds are also largely absent from fruits other than tomatoes (Liscombe et al., 2022). The most abundant lipid-derived compounds across the clade contained five or six carbons, while longer chain compounds such as octanal, nonanal, and *trans*-2-nonenal were present in higher levels in green-fruited species (Appendix S1, Figure S3d). Most phenylalanine-derived compounds were not very abundant across the clade, with guaiacol and salicylaldehyde completely absent from the green-fruited species (Appendix S1, Figure S3e). Esters were generally highest in green-fruited accessions, with some nuance for compounds such as *cis*-3-hexenyl acetate, phenethyl acetate, ethyl caprylate, and hexyl hexanoate—these were absent from most green-fruited species but high in two or three green-fruited species (Appendix S1, Figure S3f).

Overall, green-fruited species produced higher total sums and percentages of esters compared to the colored-fruited species (Figure 5). Phylogenetic generalized least squares (PGLS) models that tested the difference between colored-fruited and green-fruited groups did not yield significant *P*-values when all species were included ($P = 0.29$, lambda = 0.822 for esters sum; $P = 0.17$, lambda = 0.767 for esters percent; 30 df). However, when the two accessions of *S. pennellii* were removed, the *P*-values became highly significant ($P = 0.009$, lambda = 0 for esters sum; $P = 0.002$, lambda = 0 for esters percent; 28 df); this difference in *P*-values was due to *S. pennellii*'s high ester levels relative to its sister species *S. habrochaites*.

Five of the compounds we quantified were nitrogen-containing (Appendix S1, Figure S3g) and were mainly present only in colored-fruited (particularly red) accessions, with a few interesting exceptions. 3-Methylbutanenitrile (isovaleronitrile) was present at low levels in almost all green-fruited accessions, but the other four compounds were mostly absent from the green-fruited accessions. Interestingly, the green-fruited *S. corneliomulleri* LA1945 produced appreciable levels of four of the five nitrogen-containing compounds and was the only green-fruited accession in which we detected 1-nitro-2-phenylethane. *S. chilense* LA2750 was also notable for being the only green-fruited accession to produce fairly high levels of both 2-isobutylthiazole and 2-propylthiazole. Also noteworthy is the lack of 2-propylthiazole, 1-nitro-2-phenylethane, and

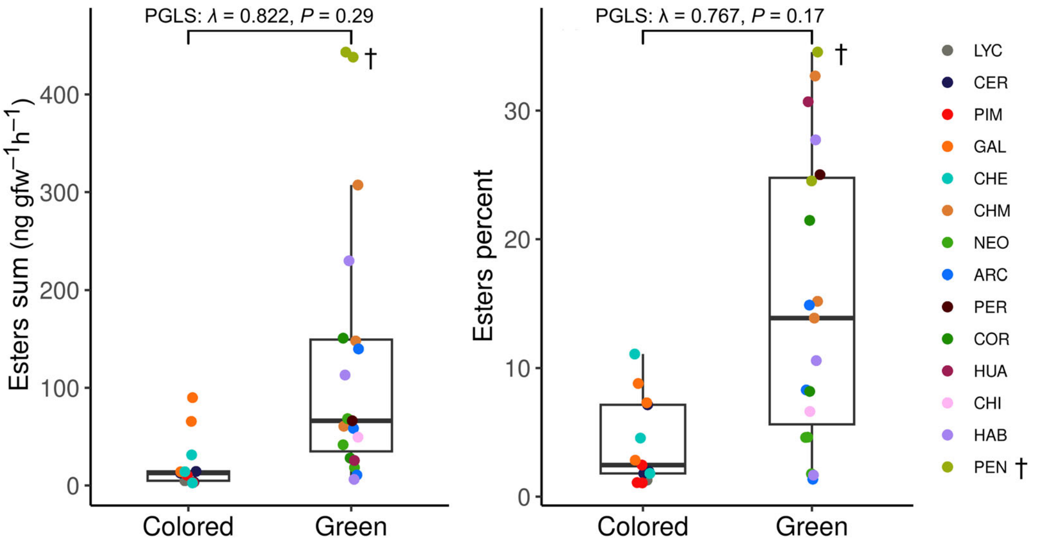


FIGURE 5 Boxplots showing differences between brightly colored-fruited and green-fruited species groups in the amount of ester volatile organic compounds (VOCs), displayed as total sum of the 17 ester compounds quantified (left), and percentage of the total sum of all 66 compounds (right), based on raw/nontransformed accession mean values. Colored points represent accessions and are color-coded by species. ARC = *Solanum arcanum*, CER = *S. lycopersicum* var. *cerasiforme*, CHE = *S. cheesmaniae*, CHI = *S. chilense*, CHM = *S. chmielewskii*, COR = *S. corneliomulleri*, GAL = *S. galapagense*, HAB = *S. habrochaites*, HUA = *S. huaylasense*, LYC = *S. lycopersicum* var. *lycopersicum* NEO = *S. neorickii*, PEN = *S. pennellii*, PER = *S. peruvianum*, PIM = *S. pimpinellifolium*. All 32 accessions were included ($N = 13$ for the colored group, $N = 19$ for the green group). Phylogenetic generalized least squares (PGLS) models were produced from raw/nontransformed accession mean values, with lambda = "ML", *P*-values are based on 30 df. †Note that *S. pennellii* (PEN) had a strong effect on the PGLS models due to its high values; with the two *S. pennellii* accessions (olive-green points) removed, the PGLS *P*-values became 0.009 for esters sum and 0.002 for esters percentage; both had lambda = 0 and 28 df.

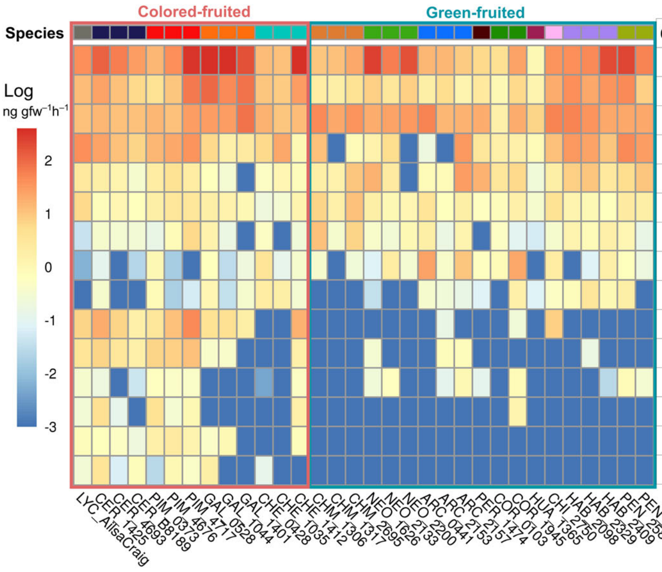
benzyl cyanide in the orange/yellow-fruited *S. galapagense* and *S. cheesmaniae* relative to their red-fruited sister species, suggesting that the loss of red fruit color on the Galápagos Islands may be correlated with decreased production of those three nitrogen-containing compounds.

Inter-accession patterns of variation in the 15 compounds important to cultivated tomato human flavor preference (Appendix S2, Table S5) are shown in Figure 6. The carotenoid-derived compounds geranylacetone and 6-methyl-5-hepten-2-one are more abundant in the colored-fruited species compared to the green-fruited species, which is not surprising since red/orange/yellow fruit color is due to carotenoid accumulation. Other compounds more abundant in colored-fruited relative to green-fruited species were 2-isobutylthiazole, 1-nitro-2-phenylethane, and guaiacol. The compound *cis*-3-hexenal (the most abundant VOC in cultivated tomato) showed an interesting pattern of being highest in the three red-fruited species as well as *S. habrochaites* and *S. pennellii*, but low in the other green-fruited species (Appendix S1, Figure S4).

Taken together, these phylogenetic patterns revealed that many of the major differences in VOCs and their biosynthetic pathways across the tomato clade are associated with the divergence of the five monophyletic colored-fruited species from the remaining nine green-fruited species, with *S. pennellii* particularly distinct from all other species. Interestingly, the divergence of the two color groups is also associated with differences in fruit sugar type and malic acid levels that may represent dispersal syndromes (Barnett et al., 2023). In general, the color-fruited group has evolved lower levels of esters and higher levels of nitrogen-containing volatiles relative to the rest of the clade, although there are a few interesting exceptions. Lower ester production appears to have evolved independently at least once within the green-fruited species in *S. neorickii*. Higher nitrogen-containing VOC production may have arisen independently in the green-fruited *S. corneliomulleri* and *S. chilense*, while the production of some nitrogen-containing compounds appears to have been lost in the yellow/orange-fruited *S. galapagense* and *S. cheesmaniae*.

Associations between VOCs and fruit sugar content

To explore whether fruit scent may communicate an honest signal about fruit nutrient content to animal dispersers, we examined the association between VOCs and sugars. For scent, we focused on the ester, terpene, and apocarotenoid VOC categories because all three can be important for attracting wild animal dispersers (Rodríguez et al., 2013; Nevo and Ayasse, 2019). We quantified nutrient content as total sugars, that is, the summed concentrations of glucose, fructose, and sucrose, the three main sugar types in tomato (Beckles et al., 2012). These sugars comprise the fruit's main caloric reward since tomatoes are low in protein and lipids (García-Alonso et al., 2020). We hypothesized that scent-based signaling would be more evident in green-fruited



Compound	Category	Flavor Effect
3-methyl-1-butanol	non-aromatic AA	NSS
isovaleraldehyde	non-aromatic AA	NSS
2-methylbutyraldehyde	non-aromatic AA	NSS
cis-3-hexenal	lipid	NSS
trans-2-hexenal	lipid	NSS
1-penten-3-one	lipid	positive
methyl salicylate	phenylalanine	NSS
2-phenyl ethanol	phenylalanine	positive
phenylacetaldehyde	phenylalanine	positive
2-isobutylthiazole	non-aromatic AA	positive
6-methyl-5-hepten-2-one	carotenoid	positive
trans-2-heptenal	lipid	positive
1-nitro-2-phenylethane	phenylalanine	positive
geranylacetone	carotenoid	positive
guaiacol	phenylalanine	negative

FIGURE 6 Heatmap of the 15 compounds we quantified that are on the list of the 21 most important to cultivated tomato flavor according to Martina et al. (2021). Category is based on precursor molecules. Flavor effect comes from Klee and Tieman (2018), NSS = not statistically significantly correlated with consumer liking. See Appendix S2, Table S5 for further details on the compounds. Data values were log (base 10)-transformed but not scaled. Columns show accession averages; color-coding along the top denotes species, arranged phylogenetically based on a tree from Barnett et al. (2023) with brightly colored-fruited species (red box) on the left and green-fruited species (blue box) on the right. ARC = *Solanum arcanum*, CER = *S. lycopersicum* var. *cerasiforme*, CHE = *S. cheesmaniae*, CHI = *S. chilense*, CHM = *S. chmielewskii*, COR = *S. corneliomulleri*, GAL = *S. galapagense*, HAB = *S. habrochaites*, HUA = *S. huaylasense*, LYC = *S. lycopersicum* var. *lycopersicum*, NEO = *S. neorickii*, PEN = *S. pennellii*, PER = *S. peruvianum*, PIM = *S. pimpinellifolium*. Numbers after species abbreviations are Tomato Genetics Resource Center (TGRC) "LA" accession numbers, except for CER_B8189 (full accession ID BGV008189), which does not have an "LA" number.

tomato species compared to colored-fruited ones, because the former may depend on smell to attract animal dispersers to their visually inconspicuous fruits. Other plants with dull-colored fruits known to attract animals via scent include some species of *Piper* (e.g., Santana et al., 2021) and *Ficus* (e.g., Lomáscolo et al., 2010).

We found that accessions very high in esters were moderately high in total sugars, while accessions low in esters can range from very low total sugars to very high total sugars (Figure 7; Appendix S1, Figure S5). Thus, a fruit with a strong ester scent does appear to signal the presence of a sugar reward. In contrast, a fruit without a strong ester scent does not seem to provide any information about sugar content because low ester accessions can either be high or low in sugar. Consistent with our hypothesis, all of the accessions high in esters (>15%) were from green-fruited species, particularly *S. chmielewskii* and *S. pennellii*. Interestingly, within-species variation in ester levels was notable for most green-fruited species other than *S. pennellii*; for example, *S. habrochaites* had one high-, one medium-, and one low-ester accession.

We did not find any clear association between terpenes and total sugars (Appendix S1, Figure S6). The accession *S. habrochaites* LA2329 was a lone outlier with very high levels of terpenes. With this accession excluded, there was still no clear trend, because high and low total sugar levels were found among the remaining accessions that were high in terpenes.

We did not find evidence of an association between apocarotenoids and sugars (Appendix S1, Figure S7). We were only able to quantify two carotenoid-derived compounds for this study, and these were almost exclusively detected in colored-fruited species, so only those 12 accessions were included in this analysis. The red-fruited *S. pimpinellifolium* accessions were generally high in both apocarotenoids and total sugars. *S. galapagense* and *S. cheesmaniae* were low in apocarotenoids, but the first was high in sugars, and the last was low in sugars.

DISCUSSION

Our results represent the first cladewide survey of ripe fruit volatiles across cultivated tomato and its 13 species of wild relatives, establishing an evolutionary context for the patterns of variation in these aroma-producing compounds and their biosynthetic pathways. By growing all plants in a common garden, we ensured that interspecies differences were mainly due to genetics rather than environment. We found that species can be differentiated by their ripe fruit VOC profiles and that phylogenetic conservatism only exists to a limited degree, with many of the major cladewide patterns corresponding to the divergence of the five monophyletic brightly colored-fruited species from the remaining nine green-fruited species, and *S. pennellii* quite distinct from all other species.

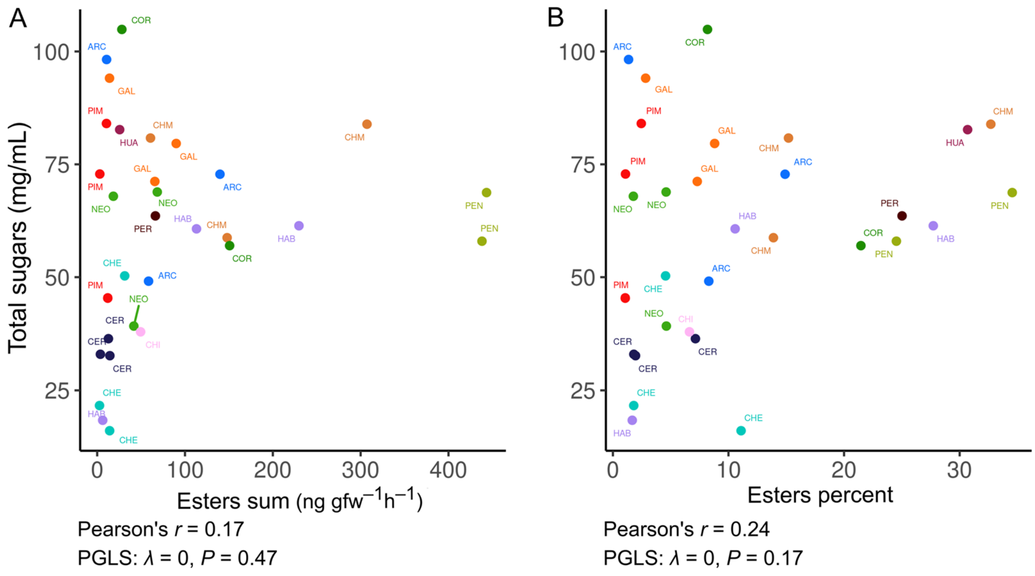


FIGURE 7 Association between total sugar concentration (glucose+fructose+sucrose) and ester levels, displayed as (A) total sum of the 17 ester compounds quantified, and (B) percentage of the total sum of all 66 compounds, based on raw/nontransformed accession mean values. Colored points represent accessions and are color-coded by species. ARC = *Solanum arcanum*, CER = *S. lycopersicum* var. *cerasiforme*, CHE = *S. cheesmaniae*, CHI = *S. chilense*, CHM = *S. chmielewskii*, COR = *S. corneliomulleri*, GAL = *S. galapagense*, HAB = *S. habrochaites*, HUA = *S. huaylasense*, LYC = *S. lycopersicum* var. *lycopersicum* NEO = *S. neorickii*, PEN = *S. pennellii*, PER = *S. peruvianum*, PIM = *S. pimpinellifolium* (see Appendix S1, Figure S5 for a version of this plot with accession numbers labeled). For this analysis, 31 accessions were included; the cultivated variety Ailsa Craig was removed because it was influenced by human selection. Pearson's *r* and phylogenetic generalized least squares (PGLS) models were produced from raw/nontransformed accession mean values, with 29 df and lambda = "ML" for the PGLS models.

Phylogenetic conservatism in fruit scent only exists to a limited extent within the wild tomato clade, with a number of similarities in distantly related species and differences in closely related species. The VOCs that do show some evidence of conservatism are mainly associated with the split of the monophyletic colored-fruited group from the paraphyletic green-fruited group. The low levels of esters and high levels of nitrogen-containing compounds in the color-fruited group could be explained by shared genetic changes inherited from a common ancestor; however, it is difficult to disentangle this phylogenetic inertia from stabilizing selection due to similar ecological conditions such as similar disperser preferences. There does not seem to be as much phylogenetic conservatism within the green-fruited group; for example, *S. chmielewskii* has high ester levels consistent with many other green-fruited species, while its sister species *S. neorickii* has quite low ester levels (Figure 4).

The differences between colored-fruited and green-fruited species support the existence of two wild tomato fruit syndromes that could be linked to disperser preferences (Barnett et al., 2023). The colored-fruited and green-fruited species had particularly notable contrasts in esters and nitrogen-containing compounds, suggesting that the evolution of colored fruits was accompanied by changes in those two biosynthetic pathways. For example, nitrogen-containing 2-isobutylthiazole was mainly produced only in colored-fruited species, while the levels of the ester methyl caproate were high only in green-fruited species. There were a few interesting exceptions to these trends, however, and compounds from all categories of molecules varied in ways that contributed to the unique aroma profile of each species. In addition, high ester levels in green-fruited species appeared to signal the presence of a sugar reward, although low ester levels were found in fruits with both high and low sugar content.

Our findings that green-fruited species can be differentiated by VOC profiles and that many of these species produce high levels of esters are noteworthy because it suggests these species can use scent to attract animal dispersers to their visually inconspicuous fruits. According to the dispersal syndrome hypothesis (Valenta and Nevo, 2020), green fruits are thought to be mainly dispersed by mammals rather than birds because many mammals have an acute sense of smell and poor color vision. Ester compounds have been shown to attract scent-oriented mammals (Peris et al., 2017), which could be the selective force behind the high levels of esters in some wild tomato species. We confirmed that colored-fruited tomato species produced significantly lower percentages of esters than in green-fruited species, consistent with the hypothesis that brighter-colored fruits evolved to primarily attract birds that have excellent color vision but a poor sense of smell. A possible genetic basis of this difference in ester production between colored- and green-fruited species has been identified; a retrotransposon insertion in the promoter of the *SlCXE1* carboxylesterase gene is found in all four colored-fruited species, but none of the green-fruited species that have thus far been sequenced (*S. chmielewskii*, *S. neorickii*, *S. habrochaites*, and *S. pennellii*) (Goulet et al., 2012). That change in the promoter was shown to increase gene expression, leading to higher levels of esterase enzyme and thus more breakdown of esters. Apart from limited anecdotal reports, however, little is known about the animal dispersers of wild tomatoes in their native ranges in South America (Barnett et al., 2023), so ecological field studies of these species would be an exciting opportunity to test the hypothesis that animal preferences may have driven fleshy fruit evolution.

Disperser attraction might explain the particularly unique VOC profile of *S. pennellii* compared to the rest of the clade. This species appears to produce the strongest scent of any wild tomato species, with the highest total sum of compounds. Many of these compounds are fruity-smelling esters; *S. pennellii* displayed the highest sum and second highest percentage for the ester category, with particularly high levels of isobutyl acetate and propyl acetate compared to other species. Interestingly, it produced more *cis*–3-hexenal (the most abundant VOC in cultivated tomato, with a green/grassy odor) than any other wild species. The environments inhabited by *S. pennellii* are distinct from the rest of the clade, with relatively high annual average temperature (the highest among 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 dispersers are more spread out in these desert habitats and need to be attracted from great distances, while relatively warm conditions enable the production of strong scents in part because biosynthetic enzyme kinetics are faster at higher temperatures, as shown for floral VOCs (Farré-Armengol et al., 2014). Volatile compounds would also be released more quickly at high temperatures.

It is worth noting that our VOCs were collected from chopped fruits, which likely have different VOC emissions than intact fruits as shown for cultivated tomatoes (Baldassarree et al., 2015). While future studies measuring VOCs from intact wild tomato fruits (as well as the leaves) would provide a more complete picture of the scent profiles experienced by frugivores in nature, we think chopped fruits are still relevant ecologically for two reasons. First, animals that do not swallow the fruit whole would break the fruit when chewing/tasting it and presumably then decide whether to eat more and disperse the seeds. Second, ripe or overripe fruits may drop off the plant and break open, releasing volatiles that attract animals to the plant. Anecdotally, we observed fallen fruits under the plants of many wild species (particularly *S. arcanum*, *S. chilense*, *S. chmielewskii*, *S. galapagense*, *S. habrochaites*, *S. neorickii*, and *S. pennellii*) and noticed distinctive scents from most of the wild species while collecting fruits.

In regard to whether scent provides an honest signal of nutrient content, our finding that accessions with high ester levels are also high in sugars is consistent with previous studies of animal-dispersed fleshy fruits in Madagascar (Nevo et al., 2019, 2022) that found a positive correlation between ester and sugar levels. This correlation may be due to a yet undetermined biochemical link between fruit sugars and ester production. Others have proposed that alcohols, which are necessary for ester synthesis, may accumulate in fruits with high sugar. For example, ethanol can be produced by microbial fermentation of sugars (Nevo et al., 2019). However, fermentation seems unlikely to explain the high ester emissions of wild tomatoes because the fruits we measured were not overripe and several of the most abundant esters (e.g., isopentyl acetate, propyl acetate, isobutyl acetate) are not derived from ethanol. While wild tomato fruits with a strong ester scent do appear to be an honest signal of high sugar reward, it is worth noting that fruits with scents less driven by esters can have a wide range of sugar content. Whether these low-ester fruits with high sugar contents have a different way of signaling to dispersers is an intriguing question.

Our results that the majority of green-fruited species produce almost no nitrogenous volatiles were consistent with expectations (Liscombe et al., 2022), although we were surprised to discover that some activity in the pathway may have independently evolved in *S. corneliomulleri* LA1945 and *S. chilense* LA2750. These accessions may be worth investigating for variations in the *SlTNH1* gene recently shown to encode an enzyme required for the biosynthesis of many of the nitrogen-containing VOCs in cultivated tomato fruits (Liscombe et al., 2022). The potential ecological function of nitrogen-containing VOCs in wild tomato fruits also merits further study; while herbivore damage has been shown to elicit production of nitrogenous volatiles in leaves of some plants (van den Boom et al., 2004), these compounds are not commonly emitted by fruits (Nevo and Ayasse, 2019).

Interestingly, humans have selected against esters and in favor of nitrogenous volatiles during domestication. The fruity/floral aromas of esters are negatively associated with cultivated tomato flavor (Klee and Tieman, 2018), while several nitrogen-containing volatiles contribute positively to flavor (Liscombe et al., 2022). This trend toward lower ester and higher nitrogenous VOC levels in red-fruited species appears to have begun before humans started eating tomatoes, however, because our findings show this pattern present in the wild red-fruited *S. pimpinellifolium*. Intriguingly, the two yellow/orange-fruited Galápagos species may be reverting back toward the green-fruited norm; *S. galapagense* and *S. cheesmaniae* produce more esters and fewer nitrogenous volatiles than their red-fruited sister taxa do. These results are consistent with previous work showing low ester levels in *S. pimpinellifolium* and intermediate ester levels in the two Galápagos species (Goulet et al., 2012). Whether the decrease in esters is connected to the opposing trend in nitrogenous volatiles is an interesting question. It is possible these changes in colored-fruited VOCs were a byproduct of early red-fruited plants evolving to attract bird dispersers, which likely rely on vision (Lomáscolo and Schaefer, 2010) rather than scent to locate food. Whatever the original cause, human selection seems to have accelerated the red-fruited trend toward lower ester and higher nitrogenous volatile production.

From a modern human perspective, our VOC measurements reveal that several wild species could be worth exploring as potential germplasm sources for crop improvement due to their high levels of compounds positively associated with flavor: notable examples were *S. arcanum* and *S. corneliomulleri* for 2-phenyl ethanol; *S. cheesmaniae*, *S. habrochaites*, and *S. pennellii* LA2580 for phenylacetaldehyde; and *S. chmielewskii* for 1-penten-3-one. In addition, we tasted fruits from many of the accessions used in this study and compiled anecdotal notes on their flavors and aromas (Appendix S2, Table S7).

CONCLUSIONS

The cladewide patterns reported here provide a framework for future studies of wild tomato fruit VOCs in a crop relative system. Our survey identifies compounds underlying the rich variety of scents produced by these species and provides clues as to the evolutionary context in which this variation arose. The wild tomato system is particularly powerful for investigating fruit evolution because cultivated tomatoes have been well studied as a model system for fleshy fruit development and genetics. Complementing this knowledge with field studies in the native ranges of these species would be particularly informative because the ecological context in which these fruits produce their scents is largely unknown. In addition, furthering our understanding of wild relatives would enhance breeding efforts to improve fruit flavor.

AUTHOR CONTRIBUTIONS

J.B. and A.C. designed the study. J.B. grew the plants and collected the fruits, quantified GC-MS peaks, analyzed data, and wrote the manuscript. D.T. ran the samples on the GC-MS and assisted with quantifications and analyses. A.C. supervised the research and helped write the manuscript. 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; UMass South Deerfield Research Farm staff Bob Skalbite, Neal Woodard, and Zach Zenke; and Caicedo Lab members Rovin Sharma, Gina Buonauro, Xiang Li, Maryam Rashidzade, and Sherin Perera. We thank Seanne Clemente and Charles Goulet for helpful suggestions with data analysis. Lynn Adler and Seanne Clemente provided valuable feedback on the manuscript. We also thank two anonymous reviewers for thoughtful and constructive comments on the manuscript. This work was supported by the UMass Natural History Collections Grant to J.B. in 2020, UMass Graduate School Dissertation Research Grant to J.B. in 2020, Lotta Crabtree Fellowship in Production Agriculture to J.B. in 2019–2022, and National Science Foundation Grant #1564366 to A.C. and D.T.

DATA AVAILABILITY STATEMENT

Data can be found in Appendix S2 (Table S3).

ORCID

Jacob R. Barnett http://orcid.org/0000-0002-4213-4010
Denise M. Tieman http://orcid.org/0000-0003-0259-2583
Ana L. Caicedo http://orcid.org/0000-0002-0378-6374

REFERENCES

Baek, Y. S., P. A. Covey, J. J. Petersen, R. T. Chetelat, B. McClure, and P. A. Bedinger. 2015. Testing the SI × SC rule: Pollen–pistil interactions in interspecific crosses between members of the tomato clade (*Solanum* section *Lycopersicon*, Solanaceae). *American Journal of Botany* 102: 302–311.

Baldassarre, V., G. Cabassi, N. D. Spadafora, A. Aprile, C. T. Müller, H. J. Rogers, and A. Ferrante. 2015. Wounding tomato fruit elicits ripening-stage specific changes in gene expression and production of volatile compounds. *Journal of Experimental Botany* 66: 1511–1526.

Baldwin, E. A., K. Goodner, and A. Plotto. 2008. Interaction of volatiles, sugars, and acids on perception of tomato aroma and flavor descriptors. *Journal of Food Science* 73: S294–S307.

Baldwin, E. A., J. W. Scott, C. K. Shewmaker, and W. Schuch. 2000. Flavor trivia and tomato aroma: biochemistry and possible mechanisms for control of important aroma components. *HortScience* 35: 1013–1022.

Barnett, J. R., R. Sharma, G. Buonauro, I. M. Gillis, M. Rashidzade, and A. L. Caicedo. 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* ppp3.10399.

Beckles, D. M., N. Hong, L. Stamova, and K. Luengwilai. 2012. Biochemical factors contributing to tomato fruit sugar content: a review. *Fruits* 67: 49–64.

Blomberg, S. P., T. Garland Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57: 717–745.

Colantonio, V., L. F. V. Ferrão, D. M. Tieman, N. Bliznyuk, C. Sims, H. J. Klee, P. Munoz, and M. F. R. Resende Jr. 2022. Metabolomic selection for enhanced fruit flavor. *Proceedings of the National Academy of Sciences, USA* 119: e2115865119.

Farré-Armengol, G., I. Filella, J. Llusià, U. Niinemets, and J. Peñuelas. 2014. Changes in floral bouquets from compound-specific responses to increasing temperatures. *Global Change Biology* 20: 3660–3669.

García-Alonso, F.-J., V. García-Valverde, I. Navarro-González, G. Martín-Pozuelo, R. González-Barrio, and M. J. Periago. 2020. Chapter 15 - Tomato. *In* A. K. Jaiswal [ed.], Nutritional composition and antioxidant properties of fruits and vegetables, 255–271. Academic Press, London, UK.

Goulet, C., M. H. Mageroy, N. B. Lam, A. Floystad, D. M. Tieman, and H. J. Klee. 2012. Role of an esterase in flavor volatile variation within the tomato clade. *Proceedings of the National Academy of Sciences, USA* 109: 19009–19014.

Hodgkison, R., M. Ayasse, C. Häberlein, S. Schulz, A. Zubaid, W. A. W. Mustapha, T. H. Kunz, and E. K. V. Kalko. 2013. Fruit bats and bat fruits: the evolution of fruit scent in relation to the foraging behaviour of bats in the New and Old World tropics. *Functional Ecology* 27: 1075–1084.

Janson, C. H. 1983. Adaptation of fruit morphology to dispersal agents in a neotropical forest. *Science* 219: 187–189.

Jiang, Y., and J. Song. 2010. Fruits and fruit flavor: classification and biological characterization. *In* Y. H. Hui [ed.], Handbook of fruit and vegetable flavors, 1–23. John Wiley & Sons, Hoboken, NJ, USA.

Keck F., F. Rimet, A. Bouchez, and A. Franc. 2016. phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecology and Evolution* 6: 2774–2780.

Kimura, S., and N. Sinha. (2008). Tomato (Solanum lycopersicum): A Model Fruit-Bearing Crop. *Cold Spring Harbor Protocols* 2008: pdb.emo105.

Klee, H. J., and D. M. Tieman. 2018. The genetics of fruit flavour preferences. *Nature Reviews. Genetics* 19: 347–356.

Kolde, R. 2019. pheatmap: Pretty heatmaps. Comprehensive R Archive Network (CRAN). Website: https://cran.r-project.org/web/packages/pheatmap/index.html [accessed 18 October 2022].

Liscombe, D. K., Y. Kamiyoshihara, J. Ghironzi, C. J. Kempthorne, K. Hooton, B. Bulot, V. Kanellis, et al. 2022. A flavin-dependent monooxygenase produces nitrogenous tomato aroma volatiles using cysteine as a nitrogen source. *Proceedings of the National Academy of Sciences, USA* 119: e2118676119.

Lomáscolo, S.B., D. J. Levey, R. T. Kimball, B. M. Bolker, and H. T. Alborn. 2010. Dispersers shape fruit diversity in *Ficus* (Moraceae). *Proceedings of the National Academy of Sciences, USA* 107: 14668–14672.

Lomáscolo, S. B., and H. M. Schaefer. 2010. Signal convergence in fruits: a result of selection by frugivores? *Journal of Evolutionary Biology* 23: 614–624.

Martina, M., Y. Tikunov, E. Portis, and A. G. Bovy. 2021. The genetic basis of tomato aroma. *Genes* 12: 226.

Nakazato, T., D. L. Warren, and L. C. Moyle. 2010. Ecological and geographic modes of species divergence in wild tomatoes. *American Journal of Botany* 97: 680–693.

Nevo, O., and M. Ayasse. 2019. Fruit scent: biochemistry, ecological function, and evolution. In J.-M. Merillon and K. G. Ramawat [eds.], Co-evolution of secondary metabolites, 1–23. Springer International, Cham, Switzerland.

Nevo, O., D. Razafimandimby, J. A. J. Jeffrey, S. Schulz, and M. Ayasse. 2018. Fruit scent as an evolved signal to primate seed dispersal. *Science Advances* 4: eaat4871.

Nevo, O., D. Razafimandimby, K. Valenta, J. A. J. Jeffrey, C. Reisdorff, C. A. Chapman, J. U. Ganzhorn, and M. Ayasse. 2019. Signal and reward in wild fleshy fruits: Does fruit scent predict nutrient content? *Ecology and Evolution* 9: 10534–10543.

Nevo, O., M. H. Schmitt, M. Ayasse, and K. Valenta. 2020a. Sweet tooth: Elephants detect fruit sugar levels based on scent alone. *Ecology and Evolution* 10: 11399–11407.

Nevo, O., K. Valenta, A. Helman, J. U. Ganzhorn, and M. Ayasse. 2022. Fruit scent as an honest signal for fruit quality. *BMC Ecology and Evolution* 22: 139.

Nevo O., K, Valenta, A. Kleiner, D. Razafimandimby, J. A. J. Jeffrey, C. A. Chapman, and M. Ayasse. 2020b. The evolution of fruit scent: phylogenetic and developmental constraints. *BMC Evolutionary Biology* 20: 138.

Nevo, O., K. Valenta, A. G. Tevlin, P. Omeja, S. A. Styler, D. J. Jackson, C. A. Chapman, and M. Ayasse. 2017. Fruit defence syndromes: the independent evolution of mechanical and chemical defences. *Evolutionary Ecology* 31: 913–923.

Orme, D., R. Freckleton, G. Thomas, T. Petzoldt, S. Fritz, N. Isaac, and W. Pearse. 2013. The caper package: comparative analysis of phylogenetics and evolution in R. R package version 5: 1–36. Website: https://cran.r-project.org/web/packages/caper/index.html

Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401: 877–884.

Pang, Z., J. Chong, G. Zhou, D. A. de Lima Morais, L. Chang, M. Barrette, C. Gauthier, et al. 2021. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research* 49: W388–W396.

Pease, J. B., D. C. Haak, M. W. Hahn, and L. C. Moyle. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PloS Biology* 14: e1002379.

Peralta, I. E., D. M. Spooner, and S. Knapp. 2008. Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia, sect. Lycopersicon; Solanaceae). Systematic Botany Monographs* 84: 1–186.

Peris, J. E., A. Rodríguez, L. Peña, and J. M. Fedriani. 2017. Fungal infestation boosts fruit aroma and fruit removal by mammals and birds. *Scientific Reports* 7: 5646.

Petro-Turza, M. 1986. Flavor of tomato and tomato products. *Food Reviews International* 2: 309–351.

Posit Team. 2022. RStudio: integrated development environment for R. Posit Software, PBC, Boston, MA, USA. Website: https://posit.co/ [accessed 2 January 2023].

Ramírez-Ojeda, G., I. E. Peralta, E. Rodríguez-Guzmán, J. Sahagún-Castellanos, J. L. Chávez-Servia, T. C. Medina-Hinostroza, J. R. Rijalba-Vela, et al. 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.

R Core Team. 2022. R: a language and environment for statistical computing. The R Foundation for Statistical Computing, Vienna, Austria. Website: https://www.r-project.org/ [accessed 2 January 2023].

Revell, L. J. 2009. Size-correction and principal components for interspecific comparative studies. *Evolution* 63: 3258–3268.

Ridley, H. 1930. The dispersal of plants throughout the world. Reeve and Ashford, Kent, UK.

Rodríguez, A., B. Alquézar, and L. Peña. 2013. Fruit aromas in mature fleshy fruits as signals of readiness for predation and seed dispersal. *New Phytologist* 197: 36–48.

Santana, S. E., Z. A. Kaliszewska, L. B. Leiser-Miller, M. E. Lauterbur, J. H. Arbour, L. M. Dávalos, and J. A. Riffell. 2021. Fruit odorants mediate co-specialization in a multispecies plant–animal mutualism. *Proceedings of the Royal Society, B, Biological Sciences* 288: 20210312.

Särkinen, T., L. Bohs, R. G. Olmstead, and S. Knapp. 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology* 13: 214.

Schwinning, S., C. J. Lortie, T. C. Esque, and L. A. DeFalco. 2022. What common-garden experiments tell us about climate responses in plants. *Journal of Ecology* 110: 986–996.

Tieman, D. M., M. Zeigler, E. A. Schmelz, M. G. Taylor, P. Bliss, M. Kirst, and H. J. Klee. 2006. Identification of loci affecting flavour volatile emissions in tomato fruits. *Journal of Experimental Botany* 57: 887–896.

Valenta, K., and O. Nevo. 2020. The dispersal syndrome hypothesis: How animals shaped fruit traits, and how they did not. *Functional Ecology* 9: 1.

Van Den Boom, C. E. M., T. A. Van Beek, M. A. Posthumus, A. De Groot, and M. Dicke. 2004. Qualitative and quantitative variation among volatile profiles induced by *Tetranychus urticae* feeding on plants from various families. *Journal of Chemical Ecology* 30: 69–89.

van der Pijl, L. 1969. Principles of dispersal in higher plants. Springer-Verlag, Berlin, Germany.

Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag, NY, NY, USA.

SUPPORTING INFORMATION

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

Appendix S1.

Figure S1. Map of collection locations for 38 wild accessions of *Solanum* species initially planted for this study.

Figure S2. PCA of centered and scaled raw values for all 66 compounds.

Figure S3. Heatmaps of compounds grouped by biosynthetic pathway.

Figure S4. Levels of *cis*-3-hexenal in 14 *Solanum* species.

Figure S5. Association between total sugar concentration (glucose+fructose+sucrose) and ester level.

Figure S6. Association between total sugar concentration (glucose+fructose+sucrose) and terpene levels.

Figure S7. Association between total sugar concentration (glucose+fructose+sucrose) and apocarotenoid levels.

Appendix S2.

Table S1. Accession IDs, germplasm collection locations, and VOC averages.

Table S2. Notes on how fruit ripeness was determined.

Table S3. VOC data for each of the 66 fruit samples.

Table S4. Summary statistics, biochemical categories, and odor descriptors for each VOC.

Table S5. Compounds considered key components of cultivated tomato aroma.

Table S6. Random forest mean decrease accuracy for VOCs differentiating species.

Table S7. Tasting notes on wild tomato fruits.

How to cite this article: Barnett, J. R., D. M. Tieman, and A. L. Caicedo. 2023. Variation in ripe fruit volatiles across the tomato clade: An evolutionary framework for studying fruit scent diversity in a crop wild relative. *American Journal of Botany* 110(9): e16223. https://doi.org/10.1002/ajb2.16223