Biosynthesis of Steroidal Alkaloids Are Coordinately Regulated and Differ Among Tomatoes in the Red-Fruited Clade

Michael P. Dzakovich¹, David M. Francis², and Jessica L. Cooperstone^{1,3*}

¹Department of Horticulture and Crop Science, The Ohio State University, 2001 Fyffe Court, Columbus, OH 43210

²Department of Horticulture and Crop Science, The Ohio State University/Ohio Agricultural Research and Development Center, 1680 Madison Ave., Wooster, OH 44691

³Department of Food Science and Technology, The Ohio State University, 2015 Fyffe Court., Columbus, OH 43210

*Corresponding author (cooperstone.1@osu.edu)

Total word count: 5424

Introduction: 757

Materials and Methods: 1727

Results: 1600

Discussion: 1345

Number of figures: 5

Number of tables: 2

Supporting information: 3 supplemental figures, 9 supplemental tables.

32 Title: Biosynthesis of Steroidal Alkaloids Are Coordinately Regulated and Differ Among 33 Tomatoes in the Red-Fruited Clade

34

35 1.1 Summary

- 36 • We quantitatively profiled and genotyped two tomato populations representing
37 diversity in the red-fruited clade to address the lack of knowledge regarding the
38 chemical diversity, concentration, and genetic architecture controlling tomato
39 steroidal alkaloids.
- 40 • We grew 107 genetically diverse fresh market, processing, land-race, and wild
41 tomatoes in multiple environments. Nine steroidal alkaloids were quantified using
42 ultra-high performance liquid chromatography tandem mass spectrometry. The
43 diversity panel and a biparental population segregating for high alpha-tomatine,
44 were genotyped to identify and validate quantitative trait loci (QTL) associated
45 with steroidal alkaloids.
- 46 • Land-races and wild material exhibited higher alkaloid concentrations and more
47 chemical diversity. Average total content of steroidal alkaloids, often dominated
48 by lycoperside F/G/esculeoside A, ranged from 1.9 to 23.3 mg/100 g fresh
49 weight across accessions. Land-race and wild cherry accessions distinctly
50 clustered based on elevated early or late steroidal alkaloid concentrations.
51 Significant correlations were observed among early and late steroidal alkaloids in
52 a species-dependent manner. A QTL controlling multiple, early steroidal alkaloid
53 pathway intermediates on chromosome 3 was identified by genome wide
54 association (GWAS) and validated in a backcross population.
- 55 • Tomato steroidal alkaloids are diverse in the red-fruited tomato clade and their
56 biosynthesis is regulated in a coordinated manner.

57

58 Keywords:

59 Diversity panel, alpha-tomatine, lycopersides, esculeosides, heritability, GWAS, QTL
60 analysis, LC-MS/MS

61

62 1.2 Introduction

63 The tomato clade of Solanaceae produce a unique assortment of cholesterol-
 64 derived steroidal alkaloids (Cárdenas *et al.*, 2015). These secondary metabolites are
 65 thought to be produced for plant defense, due to their reported fungicidal and insecticidal
 66 properties (Irving *et al.*, 1945; Cipollini & Levey, 1997). Alpha-tomatine, the most well
 67 studied tomato steroidal alkaloid, was hypothesized to impart resistance to *Fusarium* in
 68 the mid 20th century (Gottlieb, 1943; Irving *et al.*, 1945) and was isolated and
 69 characterized shortly thereafter (Fontaine *et al.*, 1948). While steroidal alkaloids are
 70 frequently considered anti-nutritional compounds (Itkin *et al.*, 2013; Cárdenas *et al.*,
 71 2016; Ballester *et al.*, 2016), a growing body of literature suggests that tomato steroidal
 72 alkaloids may impart positive health benefits as part of the diet (Friedman *et al.*, 2000;
 73 Lee *et al.*, 2004; Choi *et al.*, 2012a; Dyle *et al.*, 2014).

74 Researchers have used association analysis, comparative transcriptomics, and
 75 genetic modification as well as high resolution mass spectrometry to describe the
 76 steroidal alkaloid biosynthetic pathway and its chemical constituents (Mintz-Oron *et al.*,
 77 2008; Itkin *et al.*, 2011; Iijima *et al.*, 2013; Schwahn *et al.*, 2014; Alseekh *et al.*, 2015;
 78 Ballester *et al.*, 2016; Abdelkareem *et al.*, 2017; Sonawane *et al.*, 2018; Zhu *et al.*, 2018;
 79 Bednarz *et al.*, 2019; Cárdenas *et al.*, 2019). These recent efforts to elucidate the steroidal
 80 alkaloid biosynthetic pathway and its regulatory mechanisms in tomato are bearing fruit
 81 (Itkin *et al.*, 2011, 2013; Ballester *et al.*, 2016; Sonawane *et al.*, 2018; Cárdenas *et al.*,
 82 2019; Yu *et al.*, 2020). GLYCOALKALOID METABOLISM (GAME) enzymes 4, 6, 7,
 83 8, 11, and 12 have been shown to catalyze a series of hydroxylation, oxidation, and
 84 transamination reactions on the aliphatic tail of cholesterol to generate the E and
 85 nitrogenous F rings characteristic of solanaceous steroidal alkaloids (Itkin *et al.*, 2013).
 86 This series of reactions results in the biosynthesis of dehydrotomatidine, the first steroidal
 87 alkaloid in the proposed pathway. Dehydrotomatidine can be converted into tomatidine
 88 by GAME25, SIS5 α R2 (a C5- α reductase), and SI3 β HSD1 (a C3-
 89 dehydrogenase/reductase) (Sonawane *et al.*, 2018; Akiyama *et al.*, 2019; Lee *et al.*,
 90 2019). It is been proposed that the pathway then branches into desaturated (derived from
 91 dehydrotomatidine) and saturated (derived from tomatidine) steroidal alkaloids which are
 92 biosynthesized in parallel using the same enzymes at each step. Dehydrotomatidine and

93 tomatidine are then converted to dehydrotomatine and alpha-tomatine, respectively, by a
 94 series of glycosylations catalyzed by GAME1, 17, 18, and 2 (Itkin *et al.*, 2011, 2013).
 95 These compounds can then be hydroxylated into hydroxy-dehydrotomatine and
 96 hydroxytomatine by 2-oxoglutarate-dependent dioxygenases (GAME31 and 32)
 97 (Cárdenas *et al.*, 2019). Based on chemical structure the next steps are presumed to
 98 follow the order hydroxytomatine, acetoxytomatine, lycoperside F/G/esculeoside A, and
 99 esculeoside B (and their desaturated counterparts). Recent evidence indicates that
 100 GAME5 (a uridine diphosphate glycosyltransferase) is critical for the enzymatic
 101 conversion of acetoxytomatine to esculeoside A (Szymański *et al.*, 2020). However, the
 102 structural and regulatory mechanisms responsible for the production of other steroidal
 103 alkaloids occurring in the later part of the pathway have yet to be described.

104 Alpha-tomatine is the primary steroidal alkaloid in mature green tomatoes and is
 105 converted into various downstream lycopersides and esculeosides as the fruit matures
 106 (Iijima *et al.*, 2009; Cárdenas *et al.*, 2015). Natural variation in this process has been
 107 observed in a small subset of *Solanum lycopersicum* var. *cerasiforme* Latin American
 108 cultivar accessions, which may be considered land-races, including LA2213, LA2256,
 109 and LA2262 (Rick *et al.*, 1994). Fruits from these accessions retain high levels of alpha-
 110 tomatine throughout ripening contrary to most members of the red-fruited clade, and are
 111 commonly consumed in the Alto Mayo region of Peru (Rick *et al.*, 1994). These
 112 accessions, along with other selections made based on known patterns of genetic
 113 diversity, provide variability that may offer insight into the regulation of the tomato
 114 steroidal alkaloid biosynthetic pathway in the red-fruited clade.

115 The purpose of our work is to describe the chemical diversity of tomato steroidal
 116 alkaloids through a detailed and quantitative analysis (Dzakovich *et al.*, 2020) of
 117 cultivated, land-race, and wild red-fruited relatives of tomato and identify quantitative
 118 trait loci (QTL) associated with these metabolites. Populations were selected based on
 119 known patterns of genomic variation and were cultivated in multiple locations and years
 120 to separate genetic and environmental effects. We quantified tomato steroidal alkaloids in
 121 a diversity panel of 107 red-fruit tomato accessions and determined their associated QTL
 122 by a genome-wide association analysis (GWAS). Finally, we estimated genetic effects
 123 and validated findings from the GWAS by generating, genotyping and phenotyping a bi-

parental population based on high alpha-tomatine germplasm crossed to a cultivated parent.

1.3 Materials and Methods

1.3.1 Plant Material

A panel of 107 tomato accessions from public breeding programs, the C.M. Rick Tomato Genetics Resource Center (TGRC), and the United States Department of Agriculture (USDA) National Plant Germplasm System (NPGS) was assembled (Table S1). Selection criteria included known patterns of genetic diversity as determined by genotyping using a 7,720 SNP array (Sim *et al.*, 2012a; Blanca *et al.*, 2015), previous phenotypic information for alkaloid content (Rick *et al.*, 1994), and geographic origin (visualized in Fig. S1). To maximize genetic diversity within the red-fruited tomato species, selections included *S. pimpinellifolium*, *S. lycopersicum* var. *cerasiforme* (including wild, land-race, and cultivated cherry tomatoes), and *S. lycopersicum*. Sub-population sizes were chosen based on rarefaction analysis to capture at least of 85% of the genetic variation in previously defined sub-populations (Blanca *et al.*, 2015). This approach resulted in three species with 32 sub-populations of previously classified material: 25 accessions of *S. pimpinellifolium* (representing 11 sub-populations), 37 accessions of *S. lycopersicum* var. *cerasiforme* (representing 14 sub-populations; including six cultivated cherry varieties), and 39 accessions of *S. lycopersicum* (representing 7 sub-populations) (Blanca *et al.*, 2015). The *S. lycopersicum* selections included processing varieties representing the diversity present in this market class (Sim *et al.*, 2012a; Blanca *et al.*, 2015). Heinz 1706 (LA4345) and M82 (LA3475) were included as part of the cultivated processing tomatoes to serve as standard reference material due to their widespread and historical use in plant biology research. For subsequent analyses, we divide this population into five groups: cultivated processing, cultivated cherry, wide-cross hybrid, wild cherry (including land-races), and *S. pimpinellifolium* (Table S1). A list of all the material used in this diversity panel can be found in the supporting information (Table S1).

154 1.3.2 Field Trial

155 Plants were grown in three field environments during the summers of 2017 and
156 2018 at the North Central Agricultural Research Station in Fremont, OH. Summer, 2017
157 included the first field environment (planted 5/24/2017) while the remaining two (early
158 planting (5/29/2018); late planting (6/18/2018)) were grown during 2018. Environments
159 are differentiated by time (year and/or planting date) and geophysical conditions such as
160 soil composition. In all three environments, plants were grown in a randomized complete
161 block design with two blocks per environment (total n=642). Within each environment,
162 plants were grown in plots comprised of 6-10 plants and samples were an aggregate of
163 fruits from all plants within a plot excluding individuals at the ends of each plot. Red ripe
164 fruits were harvested over a three-week period from each environment and stored whole
165 at -40 ° C until analysis.

166

167 1.3.3 Inbred Backcross Population Development

168 Three accessions of tomatoes containing high levels of alpha-tomatine in red-ripe
169 fruits (LA2213, LA2256, and LA2262) as well as the cultivated processing accession
170 OH8243 (PI 601423) and the cultivated cherry accession Tainan (PI 647556) variety
171 were also cultivated under glasshouse conditions during Spring 2017 and crossed.
172 Seedlings from the resulting six sets of F₁ progeny as well as parental material were
173 included in the 2017 and 2018 field trials and profiled for steroidal alkaloids as described
174 below. F₁ individuals from OH8243×LA2213 were backcrossed to the cultivated
175 processing tomato recurrent parent (OH8243) and 200 individual BC₁ plants were grown
176 under glasshouse conditions during Spring 2018. Progeny were then field grown during
177 Summer 2018, and BC₁S₁ individuals were genotyped and profiled for steroidal alkaloids
178 as described below.

179

180 1.3.4 Steroidal Alkaloid Profiling

181 Steroidal alkaloids were extracted and quantified, as previously described, using a
182 rapid extraction and ultra-high performance liquid chromatography tandem mass

spectrometry (UHPLC-MS/MS) (Dzakovich *et al.*, 2020). Dehydrotomatidine, tomatidine, dehydrotomatine, alpha-tomatine, hydroxytomatine, acetoxytomatine, dehydrolycoperoside F/G/dehydroesculeoside A, lycoperoside F/G/esculeoside A, esculeoside B, and total steroidal alkaloids (the sum of all previous analytes) were quantified. Alkaloids that have the same mass (e.g., lycoperoside F/G/esculeoside A) were summed together if they could not be differentiated (Dzakovich *et al.*, 2020). Acetonitrile, formic acid, isopropanol, methanol, and water were of LC-MS grade and purchased from Fisher Scientific (Pittsburgh, PA). Alpha-tomatine ($\geq 90\%$ purity) and solanidine ($\geq 99\%$ purity) were purchased from Extrasynthese (Genay, France). Alpha-solanine ($\geq 95\%$ purity) and tomatidine ($\geq 95\%$ purity) were purchased from Sigma Aldrich (St. Louis, MO).

1.3.5 Genotyping

DNA was extracted using a hexadecyltrimethylammonium bromide (CTAB) procedure optimized for tomato and scaled to accommodate 96-tube racks (Sim *et al.*, 2015). Leaflet samples were homogenized with 60 μL of 5% Sarkosyl, 150 μL of CTAB lysis buffer (2 M NaCl, 0.2 M Tris, 0.05 M EDTA, and 2% CTAB maintained at a 7.5 pH), and 150 μL of extraction buffer (0.35 M sorbitol, 0.1 M Tris, 25 mM sodium bisulfite, and 5 mM EDTA maintained at pH 7.5) in 1.2 mL tubes. Two 4 mm metal beads were added to each tube prior to sealing and the rack was shaken at 200 rpm in a GenoGrinder (OPS Diagnostics, Lebanon, NJ) for two min. Racks were then placed in a 65 °C incubator for 20 min and then left to cool at room temperature for an additional 10 min. Each tube was then spiked with 350 μL of chloroform: isoamyl alcohol (24:1) and phase separation was induced by centrifuging plates for 10 min at 5,000 x g. Aqueous phases from each sample were transferred to a corresponding well on a fresh 96 well plate and DNA was precipitated by adding 110 μL of isopropyl alcohol to each well and centrifuging for an additional 15 min at 5000 x g. Plates were then uncovered and left to air dry for 30 min and DNA was resuspended in 100 μL of Tris-EDTA buffer (10 mM Tris and 0.1 mM EDTA).

Eighty five members of the diversity panel were previously genotyped for 7,720 SNPs using the SolCAP Illumina Infinium Array (Hamilton *et al.*, 2012; Sim *et al.*, 2012b,a; Blanca *et al.*, 2015). To genotype the BC₁S₁ population, a 384 single nucleotide polymorphism (SNP) marker panel created by the Solanaceae Coordinated Agricultural Project (SolCAP) was utilized (Sim *et al.*, 2015; Gill *et al.*, 2019). Genetic positions based on Sim and colleagues (Sim *et al.*, 2012b) and established patterns of linkage disequilibrium decay in *S. lycopersicum* (Robbins *et al.*, 2011) were used to select SNPs distributed across all 12 chromosomes in regions of high recombination. Gaps in genome coverage were filled in by selecting high PIC markers based on physical position. Genotyping was conducted on 156 BC₁S₁ individuals using the PlexSeq™ platform (Agriplex Genomics, Cleveland, OH).

1.3.6 Data Analysis

1.3.6.1 Linear modeling, heritability, and reliability

All statistical analyses and data visualization were conducted using R version 3.5.1 (R Development Core Team, 2018). A linear model based on our experimental design was used to estimate the effects of genetics and environment on steroidal alkaloid concentrations:

$$Y_{ijk} = \mu + G_i + E_j + B_k(E_j) + G_iE_j + \varepsilon_{ijk}$$

In the model above, Y_{ijk} represents the estimate for a given analyte, μ represents the population mean of a given analyte, G_i represents the contribution due to genetic factors, E_j represents the contribution due to environmental factors, $B_k(E_j)$ represents within environment effects, G_iE_j represents the interaction between genetic and environmental factors, and ε_{ijk} represents the residual error. Variance components were estimated using “ranef” function in the package lme4 (Bates *et al.*, 2015). Broad sense heritability (H^2) was calculated according to previously described methods (Cotterill, 1987):

$$H^2 = \frac{\sigma^2 G_i}{\sigma^2 G_i + \frac{\sigma^2 \varepsilon_{ijk}}{n * r} + \frac{\sigma^2 G_i E_j}{n}}$$

Where σ^2 represents the variance, n is equal to the number of environments (3), r is the number of reps within an environment (2), and all other variables are as described above. Reliability (i^2) was calculated with the previously defined model (Bernardo, 2020):

$$i^2 = \frac{\sigma^2 G_i}{\sigma^2 G_i + \sigma^2 E_j + \sigma^2 B_k(E_j) + \sigma^2 G_i E_j + \sigma^2 \varepsilon_{ijk}}$$

1.3.6.2 GWAS

Association analysis was conducted using the R package rrBLUP with GWAS utilities (Endelman, 2011). SNP calls were converted to numerical scoring where 1 was homozygous for the common allele of the reference variety OH8245 (Berry *et al.*, 1991). Homozygous for the alternative allele were scored as -1 and heterozygous calls were scored as 0. We used the unified mixed model to test for marker-trait associations (Yu *et al.*, 2006). This model contains a matrix for kinship (K), considered a random effect, and fixed effects for structure (Q). The K matrix was obtained from the “a.mat” function in rrBLUP. The Q matrix was developed from principal components analysis (PCA), as described (Price *et al.*, 2006). Briefly, a covariate matrix was calculated from the standardized $n \times M$ (number of lines $\times$ number of markers) matrix (N) by multiplying it by its transformation ($N \times N^{-1}$). The resulting $n \times n$ covariate matrix was then used as an input for PCA. We examined the effects of using 1 to 3 principal components (PC) by comparing Bayesian information criterion (BIC) for marker-trait models with different numbers of PC. For example, the naïve model $Y = \text{Marker } (M)$ was compared to $Y = PC1 + M$ and similarly the model with PC1 was compared to $Y = PC1 + PC2 + M$ and $Y = PC1 + PC2 + PC3 + M$ to test the effect of adding additional PC to the model. Benjamini-Hochberg corrected P values (Benjamini & Hochberg, 1995) were generated to control the overall false discovery rate (FDR; $P < 0.05$), account for multiple testing and provide balanced protection against type-I and -II error using the function “p.adjust.” The results of genome wide association analysis with QK models were visualized by plotting $-\log_{10}$ FDR corrected P-value vs genome position using ggplot2 (Wickham,

270 2016). The minor allele frequency (MAF) was set to 10% with alleles below the MAF
271 threshold disregarded.

272

273 1.3.6.3 QTL Analysis

274

275 Analysis of BC₁S₁ progeny were analyzed using ANOVA with the following
276 model:

$$277 Y = \mu + M + \varepsilon$$

278 Where Y is the vector of steroidal alkaloid concentrations and M represents the effect of a
279 given marker. Significant marker-trait associations were defined at a FDR-corrected $P <$
280 0.05 (Benjamini & Hochberg, 1995). P -values, the proportion of variance explained (r^2),
281 and the effects of allele substitution were summarized as a description of the QTL
282 identified.

283

284 1.3.6.4 Visualization of Steroidal Glycoalkaloid Variation

285

286 Least squares means were extracted from the fixed effects linear models for
287 steroidal alkaloid concentrations for every genotype and used to generate covariance
288 matrices used for principal components analysis (PCA) and correlation analyses (Table
289 **S3**). Population structure and phenotypic variation was visualized using PCA with the
290 packages FactoMineR and Factoextra using a correlation matrix with each variable scaled
291 to a mean of zero and standard deviation of one as the input data (Lê *et al.*, 2008).
292 Boxplots and maps displaying the geographic origin of germplasm studied were
293 generated using the packages ggplot2, ggmap, and maptools (Wickham, 2009; Kahle &
294 Wickham, 2013; Bivand & Lewin-Koh, 2018). Correlations among analytes were
295 visualized using the package corrplot (Wei & Simko, 2017).

296

297 1.4 Results

298 1.4.1 Land-races and wild tomato species exhibit steroidal alkaloid chemical diversity

299 We applied a recently developed method that extracts and quantifies 9 members
300 of the proposed tomato steroidal alkaloid pathway (Dzakovich *et al.*, 2020) to a

population of 107 (Table **S1**) tomato accessions grown in three environments. We defined early and late pathway intermediates based on their position in the proposed steroidal alkaloid biosynthesis pathway, with dehydrotomatidine, tomatidine, dehydrotomatine, alpha-tomatine, hydroxytomatine and acetoxytomatine considered early and dehydrolycoperoside F/G/dehydroesculeoside A, lycoperoside F/G/esculeoside A, and esculeoside B as late. This definition of early and late pathway intermediates is reinforced by clustering patterns and correlation based on concentrations observed within tomato germplasm.

In the case of hydroxytomatine, acetoxytomatine, dehydrolycoperoside F/G/dehydroesculeoside A, lycoperoside F/G/esculeoside A, and esculeoside B, multiple peaks are observed, confirming the presence of structural isomers (Manabe *et al.*, 2013; Nohara *et al.*, 2015; Hövelmann *et al.*, 2019; Steinert *et al.*, 2020). For example, lycoperoside F, G, and esculeoside A ($C_{58}H_{95}NO_{29}$, monoisotopic mass = 1269.5990 Da), are isobaric (Yahara *et al.*, 2004) and exhibit similar fragmentation patterns since their molecular differences are limited to subtle changes in stereochemistry on the F ring (Dzakovich *et al.*, 2020). Therefore, values reported here are the sum of all isomers within a given mass, and compounds are annotated as precisely as their identities are known. We report the sum of all steroidal alkaloids in a sample as means plus or minus standard deviation for each alkaloid. Genotype means and standard deviations for all steroidal alkaloids and their individual isomers can be found in the supporting information (Table **S2**). Differences of up to four orders of magnitude in steroidal alkaloid concentration were observed across the accessions selected. Box and whisker plots with log₁₀ scaled y-axes detailing concentrations of each tomato steroidal alkaloid separated into five classes of tomato germplasm present in this collection can be found in Fig. **1**. Each dot within the box and whisker plots represents an individual observation. Means for all alkaloids were higher in land-races and wild species accessions when compared to cultivated material. A multi-modal distribution can be observed for concentrations of lycoperoside F/G/esculeoside A within the wild cherry tomato accession group (which also contains land-races). For all steroidal alkaloids measured, concentrations in cultivated material were lower relative to land-race and wild accessions.

331 The predominant steroidal alkaloid in 85 of the 107 accessions was lycoperside
 332 F/G/esculeoside A, a late pathway intermediate. Exceptions to this trend included high
 333 alpha-tomatine, hydroxytomatine, or acetoxytomatine land-race or wild accessions (Table
 334 **S1,2**). Lycoperside F/G/esculeoside A comprised 82% of the total steroidal alkaloids in
 335 *S. pimpinellifolium* accessions, with similar proportions in the wide-cross hybrids (72%)
 336 and cultivated cherry (72%). The steroidal alkaloid profile of land-race and wild cherry
 337 and cultivated processing tomatoes was comprised of 37% and 48% lycoperside
 338 F/G/esculeoside A, respectively. In land-race and wild cherry tomatoes, acetoxytomatine
 339 was the second most abundant steroidal alkaloid (28%) on average. Alpha-tomatine was
 340 the second most abundant (23%) steroidal alkaloid in cultivated processing tomatoes.
 341 However, alpha-tomatine only comprised 3%, 5%, and 20% of total steroidal alkaloids in
 342 *S. pimpinellifolium*, cultivated cherry, and wild cherry accessions. These generalities are
 343 not always true for each individual accession given the diversity that exists within each
 344 tomato class. Median concentrations of individual steroidal alkaloids were between 0.1
 345 µg/100g and 1 mg/100 g fresh weight (Fig. **1**). For total steroidal alkaloids, median
 346 values were between 1 and 5 mg/100g fresh weight for cultivated processing and
 347 cultivated cherry, respectively. *S. pimpinellifolium*, land-race, and wild cherry had
 348 median total steroidal alkaloid values around 23 mg/100g fresh weight.

349

350 1.4.2 Early or late pathway steroidal alkaloids drive separation in land-race and wild
 351 cherry tomatoes

352 Differences in the overall profile of steroidal alkaloids from ripe fruit in
 353 accessions from the diversity panel were visualized using principal components analysis
 354 (PCA, Fig. **2**). Least squares means were calculated for each genotype, and these data
 355 were used as inputs for the PCA (Table **S3**). The first PC explained 30.1% of the
 356 phenotypic variance and the second PC, 26.8%. Cultivated material clusters strongly
 357 together. Two sub-groups of wild cherry fruits are apparent, both of which separate from
 358 cultivated cherry (Fig. **2A**). Hierarchical clustering using Ward's D2 minimum variance
 359 linkage method confirmed that these individuals form unique clusters (Fig. **S2**). A
 360 loadings plot (Fig. **2B**) of the PCA indicates separation along PC1 is driven by total

steroidal alkaloid content. Location in the tomato steroidal alkaloid biosynthesis pathway drives separation along PC2, with early pathway intermediates towards the bottom, and later pathway products at the top.

Correlations between metabolites support tradeoffs between of early and late pathway intermediates (Fig. 3). All steroidal alkaloids associated strongly with adjacent pathway metabolites (e.g., concentrations of lycoperside F/G/esculeoside A are highly correlated with esculeoside B) and all analytes were positively correlated with total steroidal alkaloids. Steroidal alkaloids with the same mass were also highly, positively correlated with one another when each isomer was considered separately (Fig. S3). Negative, but statistically non-significant correlations were observed among some early and late pathway steroidal alkaloids including dehydrotomatine, alpha-tomatine, and acetoxytomatine with dehydrolycoperoside F/G/dehydroesculeoside A, lycoperside F/G/esculeoside A, and esculeoside B, when considering the full population. To understand how alkaloid correlations differ among tomato type, correlation matrices were generated for each of the five tomato classes. Wild cherry/land-race accessions (Fig. 3c) exhibited the strongest significant negative correlations among early and late steroidal alkaloid pathway intermediates; consistent with multiple subgroups observable by eye in boxplots (Fig. 1), PCA (Fig. 2), and confirmed heuristically using hierarchical clustering (Fig. S2).

1.4.3 Diversity in tomato steroidal alkaloids is under strong genetic control

Multiple growing environments allowed us to partition effects due to genetics, environment, and their interaction on tomato steroidal alkaloid concentrations. Variance partitioning demonstrated that genetic sources of variation were the major contributor to concentrations of most steroidal alkaloids (Table 1). For alpha-tomatine, hydroxytomatine, acetoxytomatine, and lycoperside F/G/esculeoside A, broad sense heritability ranged from 0.53 to 0.96 and reliability ranged from 0.46 to 0.68 (Table 1). Dehydrotomatidine and tomatidine, which represent on average 0.01 and 0.05% of total steroidal alkaloids, respectively, exhibited relatively low proportions of variance explained by “genotype”. The strongest effects were attributed to variation within a

391 single field location. As such, estimates of broad sense heritability and reliability
392 decreased accordingly (Table 1).

393 To determine the gene action underlying high alpha-tomatine concentration in
394 ripe fruits, LA2213, LA2256, and LA2262 were each crossed to OH8243 (PI 601423; a
395 processing tomato cultivar) and Tainan (PI 647556; a fresh market grape-cherry tomato).
396 F₁ individuals were not statistically different from the cultivated parent in terms of alpha-
397 tomatine content (Table S4). While LA2213, LA2256, and LA2262 were selected for
398 crosses because of their high alpha-tomatine content (Rick *et al.*, 1994), other steroidal
399 alkaloids were also profiled in parental and F₁ individuals. Patterns of inheritance for
400 other steroidal alkaloids varied for each analyte. For example, hydroxytomatine
401 concentrations among cultivated parents and F₁ individuals were statistically similar for
402 all crosses (Table S4). OH8243 and Tainan were numerically higher in the late pathway
403 intermediates dehydrolycoperoside F/G/dehydroesculeoside A and lycoperoside
404 F/G/esculeoside A, though these differences were not statistically significant.

405 Genome wide association analysis using the QK model, adjusted for structure
406 with different numbers of PC for Q and a kinship matrix (K) were explored. Including a
407 single PC produced the best fit for all models based on BIC. GWAS with a single PC
408 identified 301 putative associations between the ten steroidal alkaloid traits (each
409 individual alkaloid, as well as total steroidal alkaloids) at an FDR-corrected P < 0.05.
410 Phenotypic data, SNP calls, outputs, and summarized results of the GWAS are available
411 in Table S5-7. Significant associations were based on 234 unique SNPs spread across all
412 12 chromosomes. For 278 of the 301 FDR-corrected P<0.05 marker-trait associations, the
413 allele most frequently identified in wild germplasm was associated with an increase in
414 concentration of steroidal alkaloids.

415
416 Fig. 4 displays Manhattan plots of steroidal alkaloids, highlighting different
417 genetic associations among early and late pathway intermediates and condensed the large
418 number of single-marker trait associations into a more defined set of loci based on
419 linkage. Thirty-six associations for alpha-tomatine reduce to 10 loci on 9 chromosomes.
420 Associations for early pathway intermediates tended to be stronger and more frequent
421 than those of late pathway intermediates. For example, loci were detected by multiple

linked markers for alpha-tomatine on 8 of 12 chromosomes while only four loci were identified for dehydrolycoperoside F/G/dehydroesculeoside A on chromosome 1, 9 and 12. Markers associated with multiple steroidal alkaloids tended to be associated with neighboring metabolites (e.g., the association of solcap_snp_sl_100848 with hydroxytomatine and acetoxytomatine).

1.4.4 Association determined by GWAS are validated in a biparental population

The BC₁S₁ material generated from crossing OH8243 (cultivated processing; low alpha-tomatine in ripe fruit) with LA2213 (land-race cherry; high alpha-tomatine in ripe fruit) provided a separate, independent population to confirm metabolic QTL (mQTL) found in the GWAS. Phenotypic data, SNP calls, and summarized results are available in Table S8,9. Analysis of backcross progeny confirmed a role for a locus on chromosome 3 in controlling multiple early pathway steroidal alkaloids as well as total concentration in tomato (Table 2). These associations are also visualized with Manhattan plots seen in Fig. 5. SNPs on chromosome 3 (solcap_snp_sl_7942, solcap_snp_sl_7939, solcap_snp_sl_7919, solcap_snp_sl_5761, and solcap_snp_sl_5656) were strongly associated with tomato steroidal alkaloids from the first half of the pathway (dehydrotomatine, alpha-tomatine, hydroxytomatine, and acetoxytomatine), as well as total tomato steroidal alkaloids. Later pathway metabolites (dehydrolycoperoside F/G/dehydroesculeoside A, lycoperside F/G/esculeoside A, and esculeoside B) did not exhibit this association with loci on chromosome 3. Markers on chromosomes 10 (solcap_snp_sl_13202 and solcap_snp_sl_46386) and 11 (SL10890_654) were associated with the late pathway steroidal alkaloid lycoperside F/G/esculeoside A.

1.5 Discussion

1.5.1 Diversity for steroidal alkaloids is greatest in land-race and wild species

Steroidal alkaloids were found in higher concentrations in land-race and wild accessions compared to cultivated cherry and processing classes (Fig. 1). Alpha-tomatine was a minor component of steroidal alkaloid profiles except for cultivated processing and

451 wild cherry material; the latter being inflated due to five high alpha-tomatine land-races
 452 (LA1654, LA2213, LA2256, LA2262, and LA2308; Table **S1,2**). This finding agrees
 453 with the current model that early pathway intermediates are converted to late pathway
 454 during ripening (Yamanaka *et al.*, 2009; Itkin *et al.*, 2011). Average values for individual
 455 alkaloids ranged between 0.1 and 7.6 mg/100 g fresh weight (Table **S1,2**), demonstrating
 456 tomato steroidal alkaloids exist in concentrations higher than previously thought
 457 (Friedman, 2002; Caprioli *et al.*, 2014; Baldina *et al.*, 2016), and consistent with more
 458 recent reports (Steinert *et al.*, 2020). Prominent tomato phytochemical classes like
 459 carotenoids are often found between 0.09 to 9.93 mg/100 g fresh weight (Dzakovich *et al.*, 2019), suggesting steroidal alkaloids can be found in comparable concentrations.
 460 Exceptional levels were observed. For example, lycoperside F/G/esculeoside A were
 461 quantified in concentrations of up to 60 mg/100g fresh weight in wild accessions (PI-
 462 155372; Table **S1, 2**), suggesting substantial diversity for concentrations of steroidal
 463 alkaloids in wild tomato. Lower average concentrations and diversity of steroidal
 464 alkaloids in cultivated processing material fit with the hypothesis that steroidal alkaloids
 465 were under strong negative selection during domestication (Zhu *et al.*, 2018).

467

468 1.5.2 Land-race and wild cherry accessions can be differentiated by early or late pathway
 469 steroidal alkaloid concentrations

470 A noticeable bifurcation in PCA scores plot of the diversity panel alkaloid
 471 concentrations can be seen in Fig. **2a**. Separation on PC1 is driven by total steroidal
 472 alkaloid concentrations, thus differentiating between cultivated and land-race and wild
 473 accessions. Individuals separated on PC2 appeared to be distinguished based on the
 474 abundance of specific steroidal alkaloids. The loadings plot in Fig. **2b** visualizes the
 475 separation of early and late steroidal alkaloid biosynthetic pathway intermediates and
 476 confirms that these profiles are driving the separation along PC2 (Fig. **2a**). Ten *S.*
 477 *lycopersicum* var. *cerasiforme* individuals appear separated in the PCA scores plot from
 478 other *S. lycopersicum* var. *cerasiforme* and *S. pimpinellifolium* accessions, and this
 479 relationship was confirmed heuristically by hierarchical clustering analysis (Fig. **S2**).
 480 These *S. lycopersicum* var. *cerasiforme* individuals include the high alpha tomatine

481 accessions LA2213, LA2256, and LA2262 previously described (Rick *et al.*, 1994) as
 482 well as several other accessions from the same region of Peru. Correlation analyses
 483 among tomato steroidal alkaloids reflect a differential pattern of pathway intermediate
 484 accumulation among germplasm classes as well (Fig. 3). Late-pathway tomato steroidal
 485 alkaloids tend to be negatively correlated with early-pathway metabolites in the land-race
 486 and wild cherry accessions but not other germplasm groups (Fig. 3c). This finding
 487 suggests differential regulation of early and late parts of the pathway in some land-race
 488 and wild cherry selections. Exploring tomato germplasm within the red-fruited clade may
 489 therefore offer insight into the genetic regulation of steroidal alkaloids.

490

491 1.5.3 Tomato steroidal alkaloids are heritable

492 Heritability and reliability estimates demonstrate that production of tomato
 493 steroidal alkaloids such as alpha-tomatine, hydroxytomatine, acetoxymatine, and
 494 lycoperside F/G/esculeoside A are under strong genetic control (Table 1). Two previous
 495 reports provide estimates of broad sense heritability between 0.35 and 0.95, with most
 496 being above 0.50 (Alseekh *et al.*, 2015; Zhu *et al.*, 2018). These data are consistent with
 497 our findings. All crosses between high alpha-tomatine land-race tomatoes (LA2213,
 498 LA2256, and LA2262) with low alpha-tomatine contemporary cultivated tomatoes show
 499 that low alpha-tomatine in ripe fruits is a dominant trait (Table S4). The typical pattern in
 500 contemporary cultivated tomato is for the early pathway metabolites such as alpha-
 501 tomatine to be converted to late-pathway intermediates such as lycoperside
 502 F/G/esculeoside A during ripening (Iijima *et al.*, 2009; Cárdenas *et al.*, 2019; Szymański
 503 *et al.*, 2020). The recessive nature of high alpha-tomatine suggests that elevated levels in
 504 ripe fruit occur due to a loss of function and a block in the shift from early to late
 505 pathway steroidal alkaloids. These results are consistent with previous findings where
 506 LA2213, LA2256, and LA2262 were initially characterized and crossed with low alpha-
 507 tomatine varieties (Rick *et al.*, 1994). The trade-off between late and early pathway
 508 metabolites and strong heritability are consistent with changes in pathway regulation.
 509

510 1.5.4 A QTL on chromosome 3 controls multiple early pathway steroidal alkaloids

511 GWAS illuminated hundreds of putative associations between molecular markers
512 and steroidal alkaloids with 85% of these associations occurring with tomatidine,
513 dehydrotomatine, alpha-tomatine, hydroxytomatine, and acetoxytomatine. These analytes
514 appear to be driving the separation visualized along PC2 (Fig. 2). Associations on
515 chromosome 3 for dehydrotomatine, tomatine, hydroxytomatine, acetoxytomatine, and
516 total steroidal alkaloids were validated in a biparental backcross population. Associations
517 for the late pathway intermediate lycoperside F/G/esculeoside A detected in the bi-
518 parental backcross population on chromosomes 10 and 11 were not observed in the
519 diversity panel GWAS. The identification of QTL associated with the concentration of
520 early metabolites on chromosome 3 (Table 2, Fig. 5.) is a novel finding and drives the
521 separation of germplasm based on metabolite loading (Fig. 2b). These data suggest that
522 the coordinate regulation of early and late tomato steroidal alkaloid pathway steps is
523 under distinct genetic control.

524 Previous studies seeking to map genes involved in the biosynthesis of tomato
525 steroidal alkaloids have used wild tomato species with an emphasis on *S. pennellii*
526 LA0716. These studies have found QTL on all chromosomes of the tomato genome
527 (Alseekh *et al.*, 2015), but the majority of reports highlight chromosomes 2, 3, 7, 10, and
528 12 (Itkin *et al.*, 2013; Baldina *et al.*, 2016; Ballester *et al.*, 2016; Zhu *et al.*, 2018;
529 Szymański *et al.*, 2020). The GAME genes have been found on chromosomes 1, 2, 7, and
530 12 (Itkin *et al.*, 2013). A previous mQTL study identified multiple candidate genes in a
531 sweep region at the end of chromosome 3 (Zhu *et al.*, 2018). Among these candidates
532 were two ethylene-responsive transcription factors. Many of the steps in the steroidal
533 alkaloid biosynthetic pathway that have been elucidated are modulated by hormones,
534 such as ethylene, during ripening (Iijima *et al.*, 2008, 2009; Itkin *et al.*, 2011). A
535 transcription factor that can affect multiple steps in the biosynthetic pathway may be
536 responsible given that multiple steroidal alkaloids in our population appear to be affected
537 by the same QTL on chromosome 3. This QTL has not been reported before and may be
538 unique to the germplasm we selected.

539 Quantitative profiling of red-fruited species demonstrates that tomato steroidal
540 alkaloids are an abundant and diverse class of secondary metabolites in ripe fruit. Much

541 of the focus on elucidating the genetic architecture of these compounds has been within
542 the context of their potential bitterness or perceived negative effects on human health
543 (Rick *et al.*, 1994; Itkin *et al.*, 2013; Cárdenas *et al.*, 2016, 2019; Ballester *et al.*, 2016),
544 despite numerous reports indicating that steroidal alkaloids may be associated with
545 positive health outcomes (Cayen, 1971; Lee *et al.*, 2004; Choi *et al.*, 2012b; Dyle *et al.*,
546 2014), and accumulate in the body after consumption (Cichon *et al.*, 2017; Cooperstone
547 *et al.*, 2017; Hövelmann *et al.*, 2019, 2020). By purposefully creating germplasm to study
548 the control of steroidal alkaloid biosynthesis in the red-fruited clade, we generated
549 quantitative evidence of a metabolic tradeoff between early and late pathway
550 intermediates. This pattern was observable through phenotypic data, unsupervised
551 learning, correlation among pathway intermediates, and genetic analysis. Our results
552 confirm that the previous model hypothesizing a degradation of tomato steroidal
553 alkaloids during ripening (Rick *et al.*, 1994), does not accurately reflect the shift in
554 metabolites that occurs in all of the red-fruited clade. We identified a previously
555 unreported QTL on chromosome 3 associated with the accumulation of dehydrotomatine,
556 alpha-tomatine, hydroxytomatine, acetoxytomatine, and total steroidal alkaloids in ripe
557 tomato fruits. This association was derived from wild cherry (including land-race)
558 accessions from San Martin and La Libertad regions of Peru and strongly supported in
559 the biparental population derived from LA2213. Fine mapping experiments to identify
560 gene(s) underlying this QTL will allow for the exploitation of steroidal alkaloids to better
561 understand their role in defense, flavor, and human health.
562

Acknowledgments

We thank the crew at the North Central Agriculture Research Station of OSU, particular Matt Hoffelich and Frank Thayer for excellent tomato care. We would also like to thank Troy Aldrich and Jiheun Cho for helping coordinate planting and seed saving.

Financial support for this work was provided by an Ohio Agricultural Research and Development Center Early Career Investigator Award, USDA National Needs Fellowship (2014-38420-21844) and Hatch funds (OHO01470, OHO01405), Foundation for Food and Agricultural Research New Innovator Award, and Foods for Health, a focus area of the Discovery Themes at OSU.

The authors have no conflicts of interest to disclose.

577 Table 1. Percentage of total variance due to the contribution of genetics and environment for steroidal alkaloid content.

Model Parameter ^a	Dehydro- tomatidine	Tomatidine	Dehydro- tomatine	Alpha- tomatine	Hydroxy- tomatine	Acetoxy- tomatine	Dehydro- lycoperoside F/G/Dehydro- esculeoside A	Lycoperoside F/G/ esculeoside A	Esculeoside B	Total
Genotype	9.0%	3.0%	25.0%	67.7%	52.2%	46.5%	33.7%	66.6%	42.1%	59.1%
Environment	2.0%	3.3%	0.1%	0.2%	1.5%	0.3%	0.1%	0.0%	2.4%	0.4%
Within Environment	3.0%	26.9%	69.6%	16.3%	23.4%	34.3%	56.1%	16.9%	26.1%	15.0%
Genotype by Environment	43.3%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%
Residual	42.7%	66.8%	5.3%	15.8%	22.9%	18.9%	10.1%	16.5%	29.3%	25.5%
Heritability ^b	0.21	0.29	0.97	0.96	0.93	0.94	0.97	0.96	0.90	0.93
Reliability	0.03	0.09	0.25	0.68	0.52	0.46	0.34	0.67	0.42	0.59

578 ^aModel parameter refers to components of random effects models used to calculate variance for each steroidal alkaloid. A full model
579 statement and details of its execution are enumerated in the data analysis section of the materials and methods.

580 ^bBroad sense heritability calculated as previously described (Cotterill, 1987).
581

582 Table 2. Markers associated (FDR-corrected $P < 0.05$) with tomato steroidal alkaloids in the BC₁S₁ validation population. Marker
583 means reported in $\mu\text{g}/100\text{ g}$ fresh weight.

Marker	Physical Position ^b	Chr ^c	Steroidal Alkaloid	Q-value ^d	r ²	Marker Mean ($\mu\text{g}/100\text{ g}$ fresh weight) ^a			
						Recurrent Parent (OH8243)	Het. ^e	Donor (LA2213)	Donor Allele
solcap_snp_sl_5656	46.04	3	Dehydrotomatine	4.16E-2	.23	29	110	112	G
solcap_snp_sl_7942	55.98	3	Dehydrotomatine	5.32E-3	.26	37	66	177	C
solcap_snp_sl_7939	56.01	3	Dehydrotomatine	1.94E-3	.27	36	66	194	A
solcap_snp_sl_7919	56.44	3	Dehydrotomatine	1.94E-3	.27	377	63	193	A
solcap_snp_sl_5761	45.59	3	Alpha-tomatine	3.22E-2	.23	464	1831	1637	C
solcap_snp_sl_5656	46.04	3	Alpha-tomatine	2.49E-3	.26	380	1998	1789	G
solcap_snp_sl_7942	55.98	3	Alpha-tomatine	2.49E-3	.26	541	1258	2825	C
solcap_snp_sl_7939	56.01	3	Alpha-tomatine	2.49E-3	.27	601	1085	3103	A
solcap_snp_sl_7919	56.44	3	Alpha-tomatine	2.49E-3	.27	616	1029	3103	A
solcap_snp_sl_7919	56.44	3	Hydroxytomatine	3.79E-2	.59	213	217	428	A
solcap_snp_sl_5761	45.59	3	Acetoxytomatine	4.49E-2	.22	97	401	317	C
solcap_snp_sl_5656	46.04	3	Acetoxytomatine	1.53E-2	.24	87	422	350	G
solcap_snp_sl_7942	55.98	3	Acetoxytomatine	2.60E-2	.23	136	223	589	C
solcap_snp_sl_7939	56.01	3	Acetoxytomatine	1.53E-2	.24	149	190	636	A
solcap_snp_sl_7919	56.44	3	Acetoxytomatine	1.53E-2	.24	49	179	641	A
solcap_snp_sl_13202	1.16	10	Lycoperside F/G/esculeoside A	4.52E-2	.66	596	690	1115	C
solcap_snp_sl_46386	1.73	10	Lycoperside F/G/esculeoside A	4.39E-2	.67	571	718	1060	C
SL10890_654	51.61	11	Lycoperside F/G/esculeoside A	1.62E-2	.67	640	539	1145	C
solcap_snp_sl_5761	45.59	3	Total	3.46E-2	.43	1567	3256	3310	C
solcap_snp_sl_5656	46.04	3	Total	3.79E-3	.45	1472	3448	3466	G
solcap_snp_sl_7942	55.98	3	Total	7.20E-4	.46	1681	2412	4887	C
solcap_snp_sl_7939	56.01	3	Total	6.78E-4	.47	1748	2274	5154	A
solcap_snp_sl_7919	56.44	3	Total	6.78E-4	.47	1770	2165	5236	A

584 ^aAverage value obtained for individuals containing marker alleles associated with the recurrent parent (OH8243), donor parent
585 (LA2213), or heterozygotes. All quantities shown are measured in $\mu\text{g}/100\text{ g}$ fresh weight.

586 ^bThe physical location (measured in megabases (Mb)) of a marker

587 ^cChromosome number

588 ^dFDR-corrected p-value

589 ^eHeterozygote

590

591

592

References

- 593 **Abdelkareem A, Thagun C, Nakayasu M, Mizutani M, Hashimoto T, Shoji T. 2017.** Jasmonate-
594 induced biosynthesis of steroidal glycoalkaloids depends on CO1 proteins in tomato.
595 *Biochemical and Biophysical Research Communications* **94**.
- 596 **Akiyama R, Lee HJ, Nakayasu M, Osakabe K, Osakabe Y, Umemoto N, Saito K, Muranaka T,**
597 **Sugimoto Y, Mizutani M. 2019.** Characterization of steroid 5 α -reductase involved in α -tomatine
598 biosynthesis in tomatoes. *Plant Biotechnology*.
- 599 **Alseekh S, Tohge T, Wendenberg R, Scossa F, Omranian N, Li J, Kleessen S, Giavalisco P, Pleban**
600 **T, Mueller-Roeber B, et al. 2015.** Identification and mode of inheritance of quantitative trait loci
601 for secondary metabolite abundance in tomato. *The Plant cell* **27**: 485–512.
- 602 **Baldina S, Picarella ME, Troise AD, Pucci A, Ruggieri V, Ferracane R, Barone A, Fogliano V,**
603 **Mazzucato A. 2016.** Metabolite Profiling of Italian Tomato Landraces with Different Fruit Types.
604 *Frontiers in Plant Science* **7**: 664.
- 605 **Ballester A-R, Tikunov Y, Molthoff J, Grandillo S, Viquez-Zamora M, de Vos R, de Maagd RA,**
606 **van Heusden S, Bovy AG. 2016.** Identification of Loci Affecting Accumulation of Secondary
607 Metabolites in Tomato Fruit of a *Solanum lycopersicum* $\times$ *Solanum chmielewskii* Introgression
608 Line Population. *Frontiers in Plant Science* **7**: 1428.
- 609 **Bates D, Mächler M, Bolker B, Walker S. 2015.** Fitting Linear Mixed-Effects Models Using **lme4**.
610 *Journal of Statistical Software* **67**: 1–48.
- 611 **Bednarz H, Roloff N, Niehaus K. 2019.** Mass Spectrometry Imaging of the Spatial and Temporal
612 Localization of Alkaloids in Nightshades. *Journal of Agricultural and Food Chemistry*.
- 613 **Benjamini Y, Hochberg Y. 1995.** Controlling the false discovery rate: a practical and powerful
614 approach to multiple testing. *Journal of the Royal Statistical Society. Series B* **57**: 289–300.
- 615 **Bernardo R. 2020.** Reinventing quantitative genetics for plant breeding: something old,
616 something new, something borrowed, something BLUE. *Heredity*.
- 617 **Berry SZ, Gould WA, Wiese KL. 1991.** ‘Ohio 8245’ Processing Tomato. *HortScience* **26**: 1093.
- 618 **Bivand R, Lewin-Koh N. 2018.** maptools: Tools for Handling Spatial Objects.
- 619 **Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Díez MJ, Francis D, Causse M, van der**
620 **Knaap E, Cañizares J. 2015.** Genomic variation in tomato, from wild ancestors to contemporary
621 breeding accessions. *BMC genomics* **16**: 257.

622 **Caprioli G, Cahill MG, James KJ. 2014.** Mass Fragmentation Studies of α -Tomatine and
623 Validation of a Liquid Chromatography LTQ Orbitrap Mass Spectrometry Method for Its
624 Quantification in Tomatoes. *Food Analytical Methods* **7**: 1565–1571.

625 **Cárdenas PD, Sonawane PD, Heinig U, Bocobza SE, Burdman S, Aharoni A. 2015.** The bitter side
626 of the nightshades: Genomics drives discovery in Solanaceae steroidal alkaloid metabolism.
627 *Phytochemistry* **113**: 24–32.

628 **Cárdenas PD, Sonawane PD, Heinig U, Jozwiak A, Panda S, Abebie B, Kazachkova Y, Pliner M,
629 Unger T, Wolf D, et al. 2019.** Pathways to defense metabolites and evading fruit bitterness in
630 genus *Solanum* evolved through 2-oxoglutarate-dependent dioxygenases. *Nature*
631 *Communications* **10**: 5169.

632 **Cárdenas PD, Sonawane PD, Pollier J, Vanden Bossche R, Dewangan V, Weithorn E, Tal L, Meir
633 S, Rogachev I, Malitsky S, et al. 2016.** GAME9 regulates the biosynthesis of steroidal alkaloids
634 and upstream isoprenoids in the plant mevalonate pathway. *Nature Communications* **7**: 10654.

635 **Cayen M. 1971.** Effect of dietary tomatine on cholesterol metabolism in the rat. *Journal of Lipid*
636 *Research* **12**: 482–490.

637 **Choi SH, Ahn J-BB, Kozukue N, Kim H-JJ, Nishitani Y, Zhang L, Mizuno M, Levin CE, Friedman M.
638 2012a.** Structure–Activity Relationships of α -, β 1-, γ -, and δ -Tomatine and Tomatidine against
639 Human Breast (MDA-MB-231), Gastric (KATO-III), and Prostate (PC3) Cancer Cells. *Journal of*
640 *Agricultural and Food Chemistry* **60**: 3891–3899.

641 **Choi SH, Ahn J-BB, Kozukue N, Kim H-JJ, Nishitani Y, Zhang L, Mizuno M, Levin CE, Friedman M.
642 2012b.** Structure–Activity Relationships of α -, β 1-, γ -, and δ -Tomatine and Tomatidine against
643 Human Breast (MDA-MB-231), Gastric (KATO-III), and Prostate (PC3) Cancer Cells. *Journal of*
644 *Agricultural and Food Chemistry* **60**: 3891–3899.

645 **Cichon MJ, Riedl KM, Wan L, Thomas-Ahner JM, Francis DM, Clinton SK, Schwartz SJ. 2017.**
646 Plasma Metabolomics Reveals Steroidal Alkaloids as Novel Biomarkers of Tomato Intake in Mice.
647 *Molecular Nutrition & Food Research* **61**: 1700241.

648 **Cipollini ML, Levey DJ. 1997.** Antifungal activity of *Solanum* fruit glycoalkaloids: Implications for
649 frugivory and seed dispersal. *Ecology* **78**: 799–809.

650 **Cooperstone JL, Tober KL, Riedl KM, Teegarden MD, Cichon MJ, Francis DM, Schwartz SJ,
651 Oberszyn TM. 2017.** Tomatoes protect against development of UV-induced keratinocyte
652 carcinoma via metabolomic alterations. *Scientific Reports* **7**: 5106.

653 **Cotterill P. 1987.** Short note: on estimating heritability according to practical applications. *Silvae*
654 *genetica* **36**: 46–48.

655 **Dyle MC, Ebert SM, Cook DP, Kunkel SD, Fox DK, Bongers KS, Bullard S a., Dierdorff JM, Adams
656 CM. 2014.** Systems-based discovery of tomatidine as a natural small molecule inhibitor of
657 skeletal muscle atrophy. *Journal of Biological Chemistry* **289**: 14913–14924.

658 **Dzakovich MP, Gas-Pascual E, Orchard CJ, Sari EN, Riedl KM, Schwartz SJ, Francis DM,**
659 **Cooperstone JL. 2019.** Analysis of Tomato Carotenoids: Comparing Extraction and
660 Chromatographic Methods. *Journal of AOAC International* **102**: 1069–1079.

661 **Dzakovich MP, Hartman JL, Cooperstone JL. 2020.** A High-Throughput Extraction and Analysis
662 Method for Steroidal Glycoalkaloids in Tomato. *Frontiers in Plant Science* **11**: 767.

663 **Endelman JB. 2011.** Ridge Regression and Other Kernels for Genomic Selection with R Package
664 rrBLUP. *The Plant Genome* **4**: 250–255.

665 **Fontaine TD, Irving GW, Ma R, Poole JB, Doolittle SP. 1948.** Isolation and partial
666 characterization of crystalline tomatine, an antibiotic agent from the tomato plant. *Archives of*
667 *biochemistry* **18**: 467–75.

668 **Friedman M. 2002.** Tomato glycoalkaloids: role in the plant and in the diet. *Journal of*
669 *Agricultural and Food Chemistry* **50**: 5751–5780.

670 **Friedman M, Fitch TE TE, Levin CCE, Yokoyama WHW, Qualities N, Friedman M, Fitch TE TE,**
671 **Levin CCE, Yokoyama WHW. 2000.** Feeding Tomatoes to Hamsters Reduces their Plasma Low-
672 density Lipoprotein Cholesterol and Triglycerides. *Sensory and Nutritive Qualities of Food* **65**:
673 897–900.

674 **Gill U, Scott JW, Shekasteband R, Ogundiwin E, Schuit C, Francis DM, Sim S-C, Smith H, Hutton**
675 **SF. 2019.** Ty-6, a major begomovirus resistance gene on chromosome 10, is effective against
676 Tomato yellow leaf curl virus and Tomato mottle virus. *Theoretical and Applied Genetics* **132**:
677 1543–1554.

678 **Gottlieb D. 1943.** Expressed sap of tomato plants in relation to wilt resistance. *Phytopathology*
679 **33**: 1111.

680 **Hamilton JP, Sim S-C, Stoffel K, Van Deynze A, Buell CR, Francis DM. 2012.** Single Nucleotide
681 Polymorphism Discovery in Cultivated Tomato via Sequencing by Synthesis. *The Plant Genome*.

682 **Hövelmann Y, Jagels A, Schmid R, Hübner F, Humpf H-U. 2019.** Identification of potential
683 human urinary biomarkers for tomato juice intake by mass spectrometry-based metabolomics.
684 *European Journal of Nutrition*: 1–13.

685 **Hövelmann Y, Lewin L, Steinert K, Hübner F, Humpf H. 2020.** Mass Spectrometry–Based
686 Analysis of Urinary Biomarkers for Dietary Tomato Intake. *Molecular Nutrition & Food Research*
687 **64**: 2000011.

688 **Iijima Y, Fujiwara Y, Tokita T, Ikeda T, Nohara T, Aoki K, Shibata D. 2009.** Involvement of
689 Ethylene in the Accumulation of Esculeoside A during Fruit Ripening of Tomato (*Solanum*
690 *lycopersicum*). *Journal of Agricultural and Food Chemistry* **57**: 3247–3252.

691 **Iijima Y, Nakamura Y, Ogata Y, Tanaka K, Sakurai N, Suda K, Suzuki T, Suzuki H, Okazaki K,**
692 **Kitayama M, et al. 2008.** Metabolite annotations based on the integration of mass spectral
693 information. *The Plant journal : for cell and molecular biology* **54**: 949–62.

694 Iijima Y, Watanabe B, Sasaki R, Takenaka M, Ono H, Sakurai N, Umemoto N, Suzuki H, Shibata
695 D, Aoki K. 2013. Steroidal glycoalkaloid profiling and structures of glycoalkaloids in wild tomato
696 fruit. *Phytochemistry* **95**: 145–157.

697 Irving GW, Fontaine TD, Doolittle SP. 1945. Lycopersicin, a fungistatic agent from the tomato
698 plant. *Science (New York, N.Y.)* **102**: 9–11.

699 Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S,
700 Finkers R, *et al.* 2013. Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated
701 by clustered genes. *Science (New York, N.Y.)* **341**: 175–9.

702 Itkin M, Rogachev I, Alkan N, Rosenberg T, Malitsky S, Masini L, Meir S, Iijima Y, Aoki K, de Vos
703 R, *et al.* 2011. GLYCOALKALOID METABOLISM1 Is Required for Steroidal Alkaloid Glycosylation
704 and Prevention of Phytotoxicity in Tomato. *The Plant Cell* **23**: 4507–4525.

705 Kahle D, Wickham H. 2013. ggmap: Spatial Visualization with ggplot2. *The R Journal* **5**: 144–161.

706 Lê S, Josse J, Husson F. 2008. FactoMineR : An R Package for Multivariate Analysis. *Journal of*
707 *Statistical Software* **25**: 1–18.

708 Lee K-R, Kozukue N, Han J-S, Park J-H, Chang E, Baek E-J, Chang J-S, Friedman M. 2004.
709 Glycoalkaloids and Metabolites Inhibit the Growth of Human Colon (HT29) and Liver (HepG2)
710 Cancer Cells. *Journal of Agricultural and Food Chemistry* **52**: 2832–2839.

711 Lee HJ, Nakayasu M, Akiyama R, Kobayashi M, Miyachi H, Sugimoto Y, Umemoto N, Saito K,
712 Muranaka T, Mizutani M. 2019. Identification of a 3b-hydroxysteroid dehydrogenase/ 3-
713 ketosteroid reductase involved in a-tomatine biosynthesis in tomato. *Plant and Cell Physiology*
714 **60**: 1304–1315.

715 Manabe H, Fujiwara Y, Ikeda T, Ono M, Murakami K, Zhou JR, Yokomizo K, Nohara T. 2013.
716 Saponins esculeosides B-1 and B-2 in Italian canned tomatoes. *Chemical and Pharmaceutical*
717 *Bulletin* **61**: 764–767.

718 Mintz-Oron S, Mandel T, Rogachev I, Feldberg L, Lotan O, Yativ M, Wang Z, Jetter R, Venger I,
719 Adato A, *et al.* 2008. Gene Expression and Metabolism in Tomato Fruit Surface Tissues. *Plant*
720 *Physiology* **147**: 823–851.

721 Nohara T, Fujiwara Y, Zhou JR, Urata J, Ikeda T, Murakami K, El-Aasr M, Ono M. 2015.
722 Saponins, esculeosides B-1 and B-2, in tomato juice and sapogenol, esculeogenin B 1. *Chemical*
723 *and Pharmaceutical Bulletin* **63**: 848–850.

724 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal
725 components analysis corrects for stratification in genome-wide association studies. *Nature*
726 *Genetics* **38**: 904–909.

727 R Development Core Team. 2018. R: A language and environment for statistical computing.

728 **Rick CM, Uhlig JW, Jones AD. 1994.** High alpha-tomatine content in ripe fruit of Andean
729 *Lycopersicon esculentum* var. *cerasiforme*: developmental and genetic aspects. *Proceedings of*
730 *the National Academy of Sciences of the United States of America* **91**: 12877–81.

731 **Robbins MD, Sim S-Chur, Yang Wencai, Van Deynze A, van der Knaap E, Joobeur T, Francis DM.**
732 **2011.** Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that
733 detect polymorphism in cultivated tomato. *J Exp Bot* **62**: 1831–1845.

734 **Schwahn K, de Souza LP, Fernie AR, Tohge T. 2014.** Metabolomics-assisted refinement of the
735 pathways of steroidal glycoalkaloid biosynthesis in the tomato clade. *Journal of Integrative Plant*
736 *Biology* **56**: 864–875.

737 **Sim S-C, Van Deynze A, Stoffel K, Douches DS, Zarka D, Ganai MW, Chetelat RT, Hutton SF,**
738 **Scott JW, Gardner RG, et al. 2012a.** High-Density SNP Genotyping of Tomato (*Solanum*
739 *lycopersicum* L.) Reveals Patterns of Genetic Variation Due to Breeding (S-B Wu, Ed.). *PLoS ONE*
740 **7**: e45520.

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

744 **Sim SC, Robbins MD, Wijeratne S, Wang H, Yang W, Francis DM. 2015.** Association analysis for
745 bacterial spot resistance in a directionally selected complex breeding population of tomato.
746 *Phytopathology*.

747 **Sonawane PD, Heinig U, Panda S, Gilboa NS, Yona M, Pradeep Kumar S, Alkan N, Unger T,**
748 **Bocobza S, Pliner M, et al. 2018.** Short-chain dehydrogenase/reductase governs steroidal
749 specialized metabolites structural diversity and toxicity in the genus *Solanum*. *Proceedings of the*
750 *National Academy of Sciences of the United States of America* **115**: E5419–E5428.

751 **Steinert K, Hövelmann Y, Hübner F, Humpf H-U. 2020.** Identification of Novel Iso -Esculeoside B
752 from Tomato Fruits and LC–MS/MS-Based Food Screening for Major Dietary Steroidal Alkaloids
753 Focused on Esculeosides . *Journal of Agricultural and Food Chemistry* **68**: 14492–14501.

754 **Szymański J, Bocobza S, Panda S, Sonawane P, Cárdenas PD, Lashbrooke J, Kamble A, Shahaf**
755 **N, Meir S, Bovy A, et al. 2020.** Analysis of wild tomato introgression lines elucidates the genetic
756 basis of transcriptome and metabolome variation underlying fruit traits and pathogen response.
757 *Nature Genetics* **52**: 1111–1121.

758 **Wei T, Simko V. 2017.** R package ‘corrplot’: Visualization of a Correlation Matrix.

759 **Wickham H. 2009.** *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.

760 **Wickham H. 2016.** *ggplot2 Elegant Graphics for Data Analysis (Use R!)*. New York, NY: Springer-

761 Verlag.

762 **Yahara S, Uda N, Yoshio E, Yae E. 2004.** Steroidal Alkaloid Glycosides from Tomato (
763 *Lycopersicon esculentum*). *Journal of Natural Products* **67**: 500–502.

764 **Yamanaka T, Vincken J-P, Zuilhof H, Legger A, Takada N, Gruppen H. 2009.** C22 Isomerization in
765 α -Tomatine-to-Esculeoside A Conversion during Tomato Ripening Is Driven by C27 Hydroxylation
766 of Triterpenoidal Skeleton. *Journal of Agricultural and Food Chemistry* **57**: 3786–3791.

767 **Yu G, Li C, Zhang L, Zhu G, Munir S, Shi C, Zhang H, Ai G, Gao S, Zhang Y, et al. 2020.** An allelic
768 variant in GAME9 determines its binding capacity to the GAME17 promoter in the regulation of
769 steroidal glycoalkaloid biosynthesis in tomato. *Journal of Experimental Botany* **9**: 2527–2536.

770 **Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen**
771 **DM, Holland JB, et al. 2006.** A unified mixed-model method for association mapping that
772 accounts for multiple levels of relatedness. *Nature Genetics* **38**: 203–208.

773 **Zhu G, Wang S, Huang Z, Zhang S, Liao Q, Zhang C, Lin T, Qin M, Peng M, Yang C, et al. 2018.**
774 Rewiring of the Fruit Metabolome in Tomato Breeding. *Cell* **172**: 249-261.e12.

775

776 Figure Legends

777 Fig. **1.** Box and whisker plots of steroidal alkaloid concentrations across all germplasm,
778 separated by class. Each dot represents an individual observation. The y-axis was log
779 transformed to visually condense the large amount of variation observed in the
780 concentrations of all tomato steroidal alkaloids measured in this study. Distinct sub-
781 groups of wild cherry tomatoes can be observed for lycoperside F/G/esculeoside A

782
783 Fig. **2.** Principal components analysis scores plot (A) and corresponding loading plots (B)
784 for 107 genotypes represented in the diversity panel. Loadings represent vectors of
785 steroidal alkaloids phenotyped in the population and their direction/magnitude indicate
786 their influence on a given principal component. Wild accessions (i.e., wild cherry and *S.*
787 *pimpinellifolium*) exhibit greater diversity in steroidal alkaloids relative to cultivated
788 accessions. Two distinct subgroups of wild cherry tomatoes appeared to separate based
789 on location in the tomato steroidal alkaloid pathway.

790
791 Fig. **3.** Correlation matrix of all tomato steroidal alkaloids quantified in diversity panel
792 and sub-populations including a) wild cherry, b) cultivated processing, c) cultivated
793 cherry, d) *Solanum pimpinellifolium*, and e) the full diversity panel. Size and darkness of
794 circle indicate intensity of correlation coefficient (see legend on right) and *, **, and ***
795 indicate statistical significance at $P < 0.05$, 0.01, and 0.001, respectively. Cells with no
796 significance indicator were found to be $P > 0.05$. Pathway intermediates tended to
797 correlate strongly with neighboring metabolites in the proposed biosynthetic pathway. All
798 analytes correlated with “total” tomato steroidal alkaloids to varying degrees.

799
800 Fig. **4.** Manhattan plots of steroidal alkaloids used in GWAS. The dashed line indicates
801 SNPs above a significance threshold of $-\log_{10}$ Benjamini-Hochberg corrected P-values of

802 1.3 ($P < 0.05$).

803

804 Fig. **5**. Manhattan plots of steroidal alkaloids used in QTL analysis on the BC₁S₁
805 validation population. The dashed line indicates SNPs above a significance threshold of -
806 $\log_{10}$ Benjamini-Hochberg corrected P-values of 1.3 ($P < 0.05$).

807

808 Supporting Information Figure and Table Legends

809

810 Fig. **S1**. Map of Central and South America indicating collection location of wild
811 germplasm.

812

813 Fig. **S2**. Principal components analysis scores plot of wild cherry tomato accessions
814 overlaid with hierarchical clustering assignments.

815

816 Fig. **S3**. Correlation plot of all steroidal alkaloids and isomers measured in this study.

817

818 Table **S1**. Raw data and metadata of the diversity germplasm. Data are presented in
819 $\mu\text{g}/100\text{g}$ FW

820

821 Table **S2**. Genotypic means and standard deviations of each steroidal alkaloid quantified.

822

823 Table **S3**. Genotypic least squares means of each steroidal alkaloid quantified.

824

825 Table **S4**. Means and standard deviations of each steroidal alkaloid of parental and F₁
826 material. Letters represent significance groups as determined by a Tukey-Kramer post-
827 hoc test.

828

829 Table **S5**. Genotypic means of each steroidal alkaloid and SNP calls for diversity panel
830 germplasm members that were previously genotyped using the SolCAP Illumina Infinium
831 Array (7,720 SNPs).

832

833 Table **S6**. Marker-trait associations generated by GWAS of the diversity panel with a
834 minor allele frequency cutoff of 10%

835

836 Table **S7**. Summary of significant marker trait associations generated by GWAS of the
837 tomato diversity panel.

838

839 Table **S8**. Genotypic means of each steroidal alkaloid and SNP calls for the BC₁S₁
840 biparental mapping population using the 384 marker SolCAP array.

841

842 Table **S9**. Marker-trait associations generated by QTL analysis of the BC₁S₁ validation
843 population with a minor allele frequency cutoff of 10%

844

845

Steroidal Alkaloid Concentrations in Tomato

■ Cultivated Processing ■ Cultivated Cherry ■ Wide Cross Hybrid ■ Wild Cherry ■ *S. pimpinellifolium*

mg/100 g fresh weight

A

B

A

Cultivated Processing

B

Cultivated Cherry

C

Wild Cherry

D

S. pimpinellifolium

E

Full Population

F