Genome Analysis

Striking Natural Diversity in Glandular Trichome Acylsugar Composition Is Shaped by Variation at the Acryltransferase2 Locus in the Wild Tomato Solanum habrochaites\[W\][OA]

Jeongwoon Kim, Kiyoong Kang, Eliana Gonzales-Vigil*, Feng Shi, A. Daniel Jones, Cornelius S. Barry, and Robert L. Last*

Department of Plant Biology (J.K., R.L.L.), Department of Energy Plant Research Laboratory (J.K.), Department of Biochemistry and Molecular Biology (K.K., A.D.J., R.L.L.), Department of Horticulture (E.G.-V., C.S.B.), and Department of Chemistry (F.S., A.D.J.), Michigan State University, East Lansing, Michigan 48824

Acylsugars are polyesters of short- to medium-length acyl chains on sucrose or glucose backbones that are produced in secretory glandular trichomes of many solanaceous plants, including cultivated tomato (Solanum lycopersicum). Despite their roles in biotic stress adaptation and their wide taxonomic distribution, there is relatively little information about the diversity of these compounds and the genes responsible for their biosynthesis. In this study, acylsugar diversity was assessed for 80 accessions of the wild tomato species Solanum habrochaites from throughout the Andes Mountains. Trichome metabolites were analyzed by liquid chromatography-time of flight-mass spectrometry, revealing the presence of at least 34 structurally diverse acylsucroses and two acylglucoses. Distinct phenotypic classes were discovered that varied based on the presence of glucose or sucrose, the numbers and lengths of acyl chains, and the relative total amounts of acylsugars. The presence or absence of an acetyl chain on the acylsucrose hexose ring caused clustering of the accesses into two main groups. Analysis of the Acryltransferase2 gene (the apparent ortholog of Solyc01g105580) revealed differences in enzyme activity and gene expression correlated with polymorphism in S. habrochaites accessions that varied in acylsucrose acetylation. These results are consistent with the hypothesis that glandular trichome acylsugar acetylation is under selective pressure in some populations of S. habrochaites and that the gene mutates to inactivity in the absence of selection.

Trichomes are specialized epidermal cells that protrude from the surface of a variety of plant tissues. They are thought to protect against environmental stresses such as herbivory (Kang et al., 2010a; Weinhold and Baldwin, 2011), loss of water through transpiration, and UV irradiation (Zhou et al., 2007). In particular, secretory glandular trichomes (SGTs) serve as “chemical factories” where specialized metabolites are produced, stored, or volatilized (Wagner, 1991; Schilmiller et al., 2008, 2010a). In addition, SGTs produce and secrete proteins on the plant surface for insect protection (Yu et al., 1992; Thipyapong et al., 1997) and pathogen defense (Shepherd et al., 2005). SGTs also contribute to the taste and smell of plants by releasing volatile metabolites. For example, the distinctive aromas of many Mediterranean herbs of the Lamiaceae (mint family) derive from SGTs (Schilmiller et al., 2008), and compounds from the glands of hops (Humulus lupulus in the Cannabaceae) contribute to beer flavor and aroma (Wang et al., 2008). Furthermore, a number of SGT-borne metabolites are commercially valuable, especially for pharmaceutical purposes. For example, artemisinin, a widely used antimalarial, is a sesquiterpene lactone from the trichomes of Artemisia annua (Liu et al., 2011). In addition to their value in foods and medicines, trichomes provide excellent models for analyzing biosynthetic enzymes and pathways (Schilmiller et al., 2008, 2009, 2012b; Bohlmann and Gershenzon, 2009; Sallaud et al., 2009).

Plants in the genus Solanum include important crop species such as potato (Solanum tuberosum), eggplant (Solanum melongena), and tomato (Solanum lycopersicum). Previous studies reported that SGTs of cultivated tomato and its wild relatives accumulate high levels of exudates containing a variety of specialized metabolites, for example flavonoids, alkaloids, and terpenoids (Wagner, 1991; Schilmiller et al., 2008, 2010a; McDowell et al., 2011). Cultivated tomato and its wild relatives have
morphologically and chemically diverse trichomes. For example, Luckwill (1943) defined seven morphologically distinguishable types of trichomes in plants of this genus, including four glandular types (types 1, 4, 6, and 7; Supplemental Fig. S1; for more Solanum spp. trichome images, see Kang et al., 2010a, 2010b). The presence of specific types of trichomes and their densities vary across species and even within a single plant according to tissue types, developmental stages, and environmental conditions (Werker, 2000; Li et al., 2004). These morphologically distinct SGTs vary in the amounts and types of metabolites that they produce, accumulate, and/or secrete (Werker, 2000). For example, S. lycopersicum M82 leaf type 6 SGTs accumulate the sesquiterpenes β-caryophyllene and α-humulene, while the glands on the stem lack these metabolites (Schilmiller et al., 2010b). There are also species- and accession-specific differences in SGT metabolite profiles. For instance, methylketones accumulate in type 6 glands of a subset of Solanum habrochaites accessions (Fridman et al., 2005; Yu et al., 2010). Similarly, acyloglucoses are highly abundant in type 4 glands of Solanum pennelli LA0716, while acylsuccrases predominate in S. lycopersicum and S. habrochaites (Shapiro et al., 1994; McDowell et al., 2011). The chemical and morphological diversity of trichomes in different Solanum species and accessions makes the genus an attractive target for the identification of diverse trichome-borne metabolites and the major biosynthetic pathways responsible for their synthesis operating in each trichome type.

The value of the comparative metabolomics approach in trichomes was recently demonstrated in studies of Solanum spp. trichome monoterpene and sesquiterpene biosynthesis (Bohlmann and Gershenzon, 2009; Sallaud et al., 2009; Schilmiller et al., 2009). It was discovered that S. lycopersicum SGTs synthesize monoterpene biosynthesis from the cis-prenyldiphosphate intermediate nerylphosphate (Sallaud et al., 2009; Schilmiller et al., 2009). This is contrary to the previous paradigm, where the trans-prenyldiphosphate geranylphosphate was considered the universal intermediate for monoterpene biosynthesis. An analogous example of biosynthetic innovation was reported for SGTs of S. habrochaites LA1777 (Sallaud et al., 2009), shown to produce sesquiterpenes in the plastid using the all-cis-prenyldiphosphate substrate Z,Z-farnesylphosphate. This is counter to the commonly described cytosolic sesquiterpene pathway, which uses the all-trans-sesquiterpene synthase substrate E,E-farnesylphosphate. Furthermore, a recent study demonstrated chemical diversity of trichome terpenes in geographically distinct S. habrochaites accessions associated with the evolution of terpene synthases, revealing how the plasticity of biosynthetic enzymes contributes to chemical complexity and diversity (Gonzales-Vigil et al., 2012). These observations suggest that trichome specialized metabolism is evolutionarily plastic, perhaps due to selective pressure from insects or other environmental stress agents.

Acylogs are sticky exudates made in SGTs that are thought to physically or chemically improve plant defense (Mirnezhad et al., 2010; Weinhold and Baldwin, 2011). Results from the literature indicate strong acylsugar diversity in various Solanum spp. trichomes (Schilmiller et al., 2010a, 2010b; McDowell et al., 2011). Acylogs are categorized as either Suc or Glc esters based on the type of sugar core (Fig. 1), and they also have varying numbers and lengths of acyl chains decorating the sugar moiety. In particular, S. pennelli accumulates enormous amounts of acylsugars, up to 20% of leaf dry weight (Fobes et al., 1985). In addition, previously published data showed that total acylsugars in geographically distinct S. pennelli accessions vary in quantity, the proportion of Suc or Glc backbones, and the overall types of fatty acid esters (FAs) on the sugars (Shapiro et al., 1994). However, this study did not identify specific acylsugar types.

To explore the detailed acylsugar chemotypes within accessions of one species, we focused on 80 accessions collected throughout the geographical range of S. habrochaites in Peru and Ecuador (Supplemental Table S1). We describe differences in sugar backbone as well as

---

**Figure 1. Structural classes of acylsugars in Solanum species.** A, Schematic structure of an acyloglucose. The structure shown depicts a Glc triester composed of Glc and three acyl chains with various numbers of carbons represented as R. B, Schematic structure of an acylsucrose. The proposed structure shows a Suc tetraester with three acyl chains on the Glc ring and one on the Fru ring. If the sugar moiety is decorated with three, four, or five acyl chains, it is referred to as a Suc triester, tetraester, or pentaester, respectively. The positions of the acyl chains are currently unknown, with the exception of the most abundant acyloglucose in cultivated tomato (M82) that was structurally characterized by NMR (Schilmiller et al., 2010a). In addition, a few acylsugars were isolated and reported from S. habrochaites and other species by King et al. (1990, 1993). Note the changes in nomenclature since these papers were published: Lycopersicum typicum LA1777 is now called S. habrochaites LA1777, and Lycopersicum hirsutum has been changed to S. habrochaites.
numbers and lengths of acyl chains, including the presence or absence of an acetyl group, which we found to be a major difference in accessions from the southern and northern Andes Mountains. The recent identification of the acyltransferase2 enzyme (SlAT2; encoded by Solyc01g105580), involved in acylsucrose biosynthesis in S. lycopersicum (Schilmiller et al., 2012a), permitted a test of the hypothesis that differences in expression or activity of this enzyme play an important role in the chemical diversity observed. The results extend previous evidence that Solanum spp. SGT chemistry is highly dynamic (Gonzales-Vigil et al., 2012) and show that the AT2 gene is surprisingly diverse across populations of S. habrochaites.

RESULTS

Overview of the Approach

Acylsugars are produced in SGTs of a variety of Solanaceous species, sometimes to such high levels that they make the surface of the plant sticky (Fobes et al., 1985; Wagner, 1991). Despite an interest in the possible antimicrobial or insecticidal roles of these metabolites, relatively little is known about their structural diversity or biosynthesis (Ghangas and Steffens, 1993). To document acylsugar diversity within a single species, we analyzed 80 S. habrochaites accessions collected from throughout the range of the species (Supplemental Table S1; Sifres et al., 2011) and M82 cultivated tomato (S. lycopersicum).

As outlined in Figure 2, extracts of trichome and surface metabolites were analyzed by liquid chromatography-time of-flight mass spectrometry (LC-TOF-MS; Schilmiller et al., 2010a). In this method, the combination of LC and high-mass-accuracy TOF-MS provides high metabolite selectivity in a relatively rapid assay. The leaf-dip sampling method was chosen to efficiently extract glandular trichome metabolites while minimizing contamination from epidermal pavement cells (Schilmiller et al., 2010a). Mass spectra were collected at multiple collision energies (multiplexed collision-induced dissociation; Gu et al., 2010; Schilmiller et al., 2010a), with spectra stored separately for each collision energy. The MS spectra obtained at low collision energies provide molecular mass information, and those generated at higher energies yield successively more fragment ions whose masses are useful for structural annotation (Supplemental Fig. S2). Each metabolite elutes from the HPLC column at a characteristic retention time, and the lowest collision energy yields ions of mass-to-charge ratio (m/z) indicative of its molecular mass. The combination of retention time and m/z is referred to as an “analytical signal.” The mass defect of the ions, the digits of the mass value that follow the decimal point, was also used to categorize the type of compound represented by each analytical signal. The relative mass defect, which is the defect normalized to the measured mass, largely reflects the hydrogen content of the ionized molecule (Last et al., 2007; Stagliano et al., 2010).

Thirty-six analytical signals attributed to acylsugars were characterized from one or more of the 80 S. habrochaites accessions, with 34 acylsucroses and two acylglucoses (Table I). Extracted ion chromatograms for formate adducts of acylsugars (labeled [M+HCOO]− m/z in Table I) at the lowest collision energy were used for quantification of the analytical signal for each acylsugar. The peak area for each extracted ion chromatogram was integrated and normalized to an internal standard and dry leaf weight. The normalized data were used for the quantitative comparison of acylsugar profiles in 80 S. habrochaites accessions.

Acylsugar Complexity and Diversity Revealed by Comparison of Three Solanum spp. Accessions

Examination of the acylsugar analytical signals (Table I) revealed highly diverse acylsugar profiles among...
S. lycopersicum M82 and the two S. habrochaites accessions LA1777 and LA2106, as shown in Figure 3. Acylsugar profiles similar to published results were found for S. lycopersicum M82 (Schilmiller et al., 2010a), with the Suc tetraester S4:17 (2,5,5,5) as the most abundant compound, as shown in Figure 3A. Note that in this nomenclature, S or G indicates a Suc or Glc backbone, respectively, 4 indicates the total number of acyl chains, 17 is the sum of the number of carbon molecules in the acyl chains, and (2,5,5,5) describes the length of each of the individual acyl chains. The next most abundant acylsugars in M82 are S4:16 (2,4,5,5) followed by S3:22 (5,5,12) and S4:24 (2,5,5,12).

It is important to note that these designations do not distinguish isomeric acylsugars that may differ in acyl group branching or in positions of acyl group substitution unless these isomers are resolved by liquid chromatographic separation. Integrated LC-MS peak areas, therefore, may represent a combination of multiple isomeric acylsugars and may underestimate the true diversity of acylsugar metabolites.

S. habrochaites LA1777 trichome acylsugar profiles were more complex than S. lycopersicum M82, with 10 major acylsucrose analytical signals detected (Fig. 3B). The most striking difference was the length of acyl-sugar chains. The major acylsugars in M82 consist

---

**Table 1. List of analytical signals characterized as acylsugars in LC-TOF-MS analysis**

<table>
<thead>
<tr>
<th>Metabolite Annotation</th>
<th>Retention Time</th>
<th>[M+HCOO]$^-$</th>
<th>Fragment Ions</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3:13 (4,4,5)</td>
<td>16.5</td>
<td>611</td>
<td>579, 495, 425, 411, 341, 101, 87</td>
</tr>
<tr>
<td>S3:14 (4,5,5)</td>
<td>18.1</td>
<td>625</td>
<td>593, 425, 407, 341, 323, 179, 101</td>
</tr>
<tr>
<td>S3:15 (5,5,5)</td>
<td>20.5</td>
<td>639</td>
<td>593, 551, 87</td>
</tr>
<tr>
<td>S4:14 (2,4,4,4)</td>
<td>16.5</td>
<td>639</td>
<td>607, 565, 101, 87</td>
</tr>
<tr>
<td>S4:15 (2,4,4,5)</td>
<td>18.2</td>
<td>653</td>
<td>621, 579, 425, 407, 341, 323, 179, 101</td>
</tr>
<tr>
<td>S4:16 (2,4,5,5)</td>
<td>20.1</td>
<td>667</td>
<td>635, 551, 425, 407, 341, 323, 179, 101</td>
</tr>
<tr>
<td>S3:18 (5,5,8)</td>
<td>27.5</td>
<td>681</td>
<td>649, 565, 509, 425, 341, 323, 157, 101</td>
</tr>
<tr>
<td>S4:17 (2,5,5,5)</td>
<td>22.5</td>
<td>681</td>
<td>656, 593, 425, 407, 341, 323, 157, 101</td>
</tr>
<tr>
<td>S3:19 (5,5,9)</td>
<td>29.1</td>
<td>695</td>
<td>565, 495, 425, 407, 341, 171, 101</td>
</tr>
<tr>
<td>S3:20 (5,5,10)</td>
<td>30.0</td>
<td>695</td>
<td>649, 579, 509, 425, 341</td>
</tr>
<tr>
<td>S4:18 (4,4,5,5)</td>
<td>24.7</td>
<td>695</td>
<td>663, 579, 495, 425, 341</td>
</tr>
<tr>
<td>S3:21 (4,5,10)</td>
<td>20.1</td>
<td>667</td>
<td>635, 551, 425, 407, 341, 323, 179, 101</td>
</tr>
<tr>
<td>S3:22 (5,5,11)</td>
<td>33.5</td>
<td>723</td>
<td>667, 593, 509, 425, 341, 323, 157, 101</td>
</tr>
<tr>
<td>S3:23 (4,5,12)</td>
<td>34.4</td>
<td>723</td>
<td>677, 635, 481, 411, 341, 323, 101</td>
</tr>
<tr>
<td>S4:20 (2,4,4,10)</td>
<td>28.1</td>
<td>723</td>
<td>677, 593, 509, 425, 341, 323, 101</td>
</tr>
<tr>
<td>S4:20 (5,5,5,5)</td>
<td>27.1</td>
<td>723</td>
<td>677, 593, 425, 407, 341, 323, 157, 101</td>
</tr>
<tr>
<td>S4:21 (2,4,5,10)</td>
<td>30.6</td>
<td>723</td>
<td>691, 607, 509, 425, 341, 323, 179, 101</td>
</tr>
<tr>
<td>S4:23 (5,5,5,11)</td>
<td>36.3</td>
<td>751</td>
<td>691, 649, 495, 393, 341, 323, 171, 101</td>
</tr>
<tr>
<td>S4:22 (2,4,4,12)</td>
<td>31.6</td>
<td>751</td>
<td>709, 663, 481, 393, 341, 323, 305, 101</td>
</tr>
<tr>
<td>S4:22 (2,5,5,10)</td>
<td>32.3</td>
<td>751</td>
<td>705, 621, 579, 425, 341, 323, 199, 101</td>
</tr>
<tr>
<td>S3:24 (6,6,12)</td>
<td>38.1</td>
<td>765</td>
<td>719, 537, 349, 341, 323, 199, 115</td>
</tr>
<tr>
<td>S4:23 (2,4,5,12)</td>
<td>33.7</td>
<td>765</td>
<td>719, 677, 495, 411, 341, 323, 199, 101</td>
</tr>
<tr>
<td>S4:23 (2,5,5,11)</td>
<td>33.4</td>
<td>765</td>
<td>719, 677, 509, 425, 341, 323, 199, 101</td>
</tr>
<tr>
<td>S4:24 (2,5,5,12)</td>
<td>35.5</td>
<td>779</td>
<td>733, 691, 509, 425, 341, 323, 199, 101</td>
</tr>
<tr>
<td>S4:24 (4,4,5,10)</td>
<td>33.9</td>
<td>779</td>
<td>677, 593, 509, 425, 341, 323, 179, 101</td>
</tr>
<tr>
<td>S4:25 (5,5,5,10)</td>
<td>30.9</td>
<td>793</td>
<td>663, 579, 395, 199, 101, 115</td>
</tr>
<tr>
<td>S4:25 (2,5,6,12)$^a$</td>
<td>33.6</td>
<td>793</td>
<td>761, 677, 593, 509, 425, 341, 323, 101</td>
</tr>
<tr>
<td>S5:25 (5,5,5,5,$^a$)</td>
<td>32.1</td>
<td>807</td>
<td>761, 677, 593, 509, 425, 341, 323, 101</td>
</tr>
<tr>
<td>S4:26 (5,5,5,11)</td>
<td>35.6</td>
<td>807</td>
<td>775, 691, 607, 509, 425, 341, 323, 199, 101</td>
</tr>
<tr>
<td>S4:27 (5,5,12)</td>
<td>37.2</td>
<td>821</td>
<td>375, 329, 227, 143, 125, 101</td>
</tr>
<tr>
<td>G3:15 (5,5)</td>
<td>23.5</td>
<td>477</td>
<td>375, 329, 227, 199, 143, 101</td>
</tr>
<tr>
<td>G3:22 (5,5,12)</td>
<td>35.3</td>
<td>575</td>
<td>375, 329, 227, 199, 143, 101</td>
</tr>
</tbody>
</table>

---

<sup>a</sup> Fragment Ions $^a$ were obtained from the nonfragmenting condition (aperture 1 voltage of 10 V) and fragment ions from fragmenting environment (aperture 1 voltage of 55 V) in the mass spectrometer transit lens. S4:25 (2,5,6,12)$^a$ was the most likely annotation for the corresponding analytical signal. This was also distinct from the other isomer of m/z 793 in LC retention time. However, the annotation is of relatively low confidence due to its low abundance and the stronger signal intensity of m/z 723, S3:21 (5,5,11), whose retention time was very close to that of S4:25 (2,5,6,12)$^a$.

It is important to note that these designations do not distinguish isomeric acylsugars that may differ in acyl group branching or in positions of acyl group substitution unless these isomers are resolved by liquid chromatographic separation. Integrated LC-MS peak areas, therefore, may represent a combination of multiple isomeric acylsugars and may underestimate the true diversity of acylsugar metabolites.

S. habrochaites LA1777 trichome acylsugar profiles were more complex than S. lycopersicum M82, with 10 major acylsucrose analytical signals detected (Fig. 3B). The most striking difference was the length of acyl-sugar chains. The major acylsugars in M82 consist of...
predominantly of shorter chain FAs, especially C2 and C5. In addition to these short-chain FAs, LA1777 accumulated acylsugars with FA chains whose length ranged from C8 to C12 (Fig. 3, A and B). M82 and LA1777 even had differences in types of short-chain FA lengths: while C4 and C5 acyl chains were of similar abundance in LA1777, M82 mostly accumulated C5. LA1777 extracts included acylsugars with acyl chain lengths not detected in M82; for example, S3:18 appears to have a C8 acyl chain (5,5,8), and three have C10 acyl chains: S4:20 (2,4,4,10), S4:21 (2,4,5,10), and S4:22 (2,5,5,10). LA1777 not only produced additional major acylsugars not abundant in M82, but there were also differences in the relative ratio of acylsugars identified in both species. For instance, both S4:16 (2,4,5,5) and S4:17 (2,5,5,5) were found in M82 and LA1777, but their relative abundance was opposite in the two accessions. In contrast, S3:22 (5,5,12) and S4:24 (2,5,5,12) were abundant in M82 but not in LA1777. This comparative analysis revealed that acylsugars of two accessions from different species, *S. lycopersicum* M82 and *S. habrochaites* LA1777, vary in the length and numbers of acyl chains and in their relative abundance.

Comparison of trichome metabolites in the two *S. habrochaites* accessions LA1777 and LA2106 also revealed dramatic intraspecific differences in acylsugar composition (Fig. 3, B and C). In fact, no overlap in major acylsugars was found between these two accessions. This was in large part due to the low abundance of acylsugars containing C2 chains in LA2106. Instead, S3:21 (5,5,11), S4:20 (5,5,5,5), and S3:15 (5,5,5) were the most abundant acylsugars. In addition, LA2106 contained acylsugars with C9 and C11 acyl chains, which were not detected in LA1777 or M82. Taken together, these results show major differences in trichome acylsugar accumulation both between and within species. This observation led us to consider acylsugar accumulation in a larger set of *S. habrochaites* accessions.

**Systematic Comparison of Acylsugar Composition in 80 *S. habrochaites* Accessions**

To explore *S. habrochaites* metabolite diversity more fully, the complete set of normalized analytical signals from trichome extracts of 80 accessions was analyzed using hierarchical clustering analysis (HCA; Fig. 4).
and principal component analysis (PCA; Figs. 5 and 6). HCA provides an efficient approach to summarize the detailed chemical phenotypic relationships among accessions and information about the influence of specific acylsugars on the phenotypic clustering. PCA makes these trends easier to visualize.

In this analysis, the fractional peak area (analogous to mol %) for the 36 individual analytical signals was used rather than the normalized peak areas for two reasons. First, plant total specialized metabolite accumulation is sensitive to the environment. Furthermore, because we do not have authentic standards for the target acylsugar analytes, it is not possible to control for differences in relative detector response to each of the 36 metabolites or changes in MS detector performance over time. Despite the inherent challenge of performing absolute quantification with LC-MS, the ratio of individual analytical signals to the total detector response was consistent across identical samples run at different times (Supplemental Fig. S3B). In addition, the fold differences in total acylsugars in S. habrochaites accessions and M82 tomato were modest (Supplemental Fig. S4) compared with two previously characterized nearly isogenic S. pennellii × M82 introgression lines (lines 5-3 and 11-5) that accumulate approximately 15-fold reduced amounts of total acylsugars than that of M82 (Schlimmer et al., 2010a; for details, see below). Taken together, comparative analysis with fractional peak area of each acylsugar is an efficient approach to understanding diverse acylsugar profiles across accessions.

As shown in Figure 4, HCA revealed two major clusters of trichome acylsugar chemistry (“superclusters” 1 and 2) that consist of two and three main subgroups, respectively (subgroups A and B and subgroups C, D, and E). Supercluster 1 includes M82 tomato and S. habrochaites LA1777 and is characterized by accessions with tetracylsucroses containing C2 chains (acylsugar group I listed on the bottom of Fig. 4; for LC-TOF extracted ion chromatograms of M82 and LA1777, see Fig. 3, A and B, respectively). PCA also revealed that this supercluster accumulates tetracylsucroses (subgroups A and B in Fig. 5) with C2 FA chains in relatively high abundance (Fig. 6). In contrast, supercluster 2 is chemically more diverse, and a distinguishing feature is the accumulation of triacylsucroses without C2 chains in most of these lines; this is readily seen in the extracted ion chromatogram for LA2106 (Fig. 3C) and more broadly for accessions in this supercluster in Figures 5C and 6C. The general trend observed progressing from subgroup A to E is that the proportion of tetracylsucroses decreases and triacylsucroses increases (Fig. 5).

Many of the dominant differences in trichome metabolites driving the subgroupings within each supercluster are visible in the HCA (Fig. 4). The two subgroups within the tetracylsucrose-dominant supercluster 1 are largely distinguished by the presence (subgroup A) or absence (subgroup B) of one to two C4 acyl chains in most of the abundant molecules. These trends in acyl chain lengths are also quite evident by PCA, as seen in Figure 6.

Another distinguishing feature of major acylsugars in subgroup B accessions is the abundance of short-chain esters, typically C2, C4, and C5. In contrast, a broader size range of esters (C2, C4, C5, C10, and C12) is found in subgroup A with abundant S4:14 (2,4,4,4), S4:15 (2,4,4,5), S4:20 (2,4,4,10), S4:21 (2,4,5,10), and S4:22 (2,4,4,12).

Tetraacylsucrose-dominated accessions in supercluster 2 split into three subgroups (C–E) based upon various acyl chain characteristics. For example, plants in the small subgroup C accumulated characteristic acylsugars including S3:24 (6,6,12), while the C6 FA chain was not commonly observed in other S. habrochaites accessions. Subgroup C accessions LA1285, LA1624, and LA1625 also contained G3:15 (5,5,5) in relatively high abundance (for the trend of G3 accumulation in PCA and its proportion in each accession, see Fig. 5). Subgroup D showed a shift toward high relative abundance of acylsugars with longer acyl chains ranging from C6 to C12. Although the accessions in this subgroup still accumulated minor amounts of acylsugars containing C2, major acylsugars consisted of longer FA chains without C2 chains: S3:20 (4,5,11), S3:21 (5,5,11), S3:23 (5,6,12), and S4:26 (5,5,5,11). Subgroup E showed some variation within the group. For example, LA2155 and LA2156 from subgroup E accumulated the pentaacylsucrose S5:25 (5,5,5,5,5), and three accessions (LA2100, LA2864, and LA2869) accumulated high quantities of the C6 FA containing acylsucrose S3:21 (5,6,10) (Figs. 5 and 6).

Variation of Total Acylsugar Levels in 80 S. habrochaites Accessions

The accumulation of total acylsugars was analyzed to assess variation across accessions. S. habrochaites accessions accumulated acylsucroses in high abundance and relatively low or undetectable amounts of acylglucoses (Fig. 5; Supplemental Fig. S4). This analysis also revealed the variation of total acylsugar levels within this species, with the total quantity obtained by summing the normalized peak areas for the 36 analytical signals in Table I as described in “Materials and Methods” (Supplemental Fig. S4). With the exception of LA2101, this analysis showed a maximum 8-fold range of total acylsugars across all accessions, ranging from one-half to 4-fold that of M82.

LA2101 was an exceptionally low accumulator, with approximately one-tenth as much acylsugar detected as M82. This led us to hypothesize that LA2101 had dramatically reduced numbers of SGTs or abnormal trichome development. To test this hypothesis, trichomes from three S. habrochaites accessions, LA2101, LA1777, and LA2975, were analyzed by light microscopy (Supplemental Fig. S1). These three lines were chosen as accessions that accumulate total acylsugars in relatively low, intermediate, and high levels, respectively (Supplemental Fig. S4). This analysis did not reveal abnormally low SGT density or aberrant morphology in...
Figure 4. HCA of acylsugar chemistry in Solanum spp. accessions. Normalized signal intensities were used to calculate the percentage composition of 36 individual acylsugars, and the data set was analyzed by HCA. Ward’s minimum variance method.
LA2101 compared with the intermediate and high acylsugar accessions. This suggests that differences in acylsugar biosynthesis, storage, or turnover, rather than relative numbers of SGTs, are responsible for the reduced metabolite accumulation in LA2101.

The accumulation of total acylsugars was also assessed to test variation among accessions in different HCA clusters (Fig. 4). Visual inspection of the data in Supplemental Figure S4 suggested that lines in supercluster 1 (subgroups A and B) have higher total acylsugars than supercluster 2, with statistical analysis confirming that accessions from supercluster 1 accumulate approximately 2-fold higher levels of acylsugars than those of supercluster 2 ($P < 10^{-7}$, Student’s $t$ test).

The Correlation between Geographic Distribution and Acylsugar Chemistry in *S. habrochaites*

We investigated whether collection sites of *S. habrochaites* accessions are related to the diversity of acylsugar chemistry (Fig. 7; Supplemental Table S1). In general, the C2 Suc tetraester-containing supercluster 1 plants (subgroups A and B) were collected from

---

**Figure 4.** Types of Glc and Suc esters in *Solanum* spp. accessions. PCA was performed based upon the percentage composition of Glc triesters (3G) as well as Suc triesters (3S), tetraesters (4S), and pentaesters (5S) and compared with the acylsugar composition of individual accessions, shown in C. A, Score plots. B, Loading plots. C, Average percentage compositions of four types of acylsugars plotted with SE in individual accessions. The accessions are listed in the identical order as in Figure 4.
Genetic Variation of AT2 Correlates with Acyl Suc Acetylation in S. habrochaites Accessions

The S. lycopersicum BAHD superfamily member SIAT2 protein product recently was shown to acetylate triacylsucroses, converting them to tetraacylsucroses (Schilmiller et al., 2012a). This discovery led us to hypothesize that differential activity of AT2 is associated with the diversity of acylsugar chemistry in S. habrochaites accessions that differ by absence (supercluster 2) or presence (supercluster 1) of the C2 chain. To test this hypothesis, we attempted to amplify AT2 genomic DNA (gDNA) from 15 S. habrochaites accessions that represent the different chemical subgroups and obtained sequence from 13 (Supplemental Fig. S5; note that we find no evidence for introns in S. lycopersicum or S. habrochaites AT2). AT2 is part of a cluster of three paralogous genes sharing approximately 70% amino acid identity in the cultivated tomato genome sequence assembly (Schilmiller et al., 2012a), raising the concern that one or more of these diverse sequences might be from the neighboring AT1 or AT3 gene. Consistent with the hypothesis that we amplified AT2 sequences from S. habrochaites, inspection of the sequence revealed clear differences between AT2 sequences from the various species and the paralogous genes AT1 and AT3 (Supplemental Fig. S5). This hypothesis was strengthened by phylogenetic analysis of the acyltransferase sequences (Fig. 8): the AT2 sequences from S. habrochaites, S. lycopersicum, and Solanum pimpinellifolium cluster together, distinct from AT1 or AT3 from these species.

The discovery of polymorphic AT2 sequences led us to hypothesize that differences in AT2 gene function were responsible for the presence or absence of C2 chains in S. habrochaites accessions. Consistent with this hypothesis, AT2 proteins from tetraacylsucrose-accumulating supercluster 1 accessions LA1777 (subgroup A) and LA1731 (subgroup B) produced acetylated tetraacylsucrose products when incubated with a mixture of triacylsucroses and acetyl-CoA (Fig. 9A; Supplemental Fig. S6). In contrast, the AT2 sequences of supercluster 2 subgroup D accessions LA0407, LA1253, LA2106,
and LA2107 all had a single nucleotide deletion at position 1,045 bp of the protein-coding region (Supplemental Fig. S5). This frameshift mutation is predicted to cause deletion of the last 96 amino acids of AT2, including a DFGWG domain that is highly conserved in previously characterized BAHD enzyme superfams (Fig. 9B; Ma et al., 2005; Schilmiller et al., 2012a). The hypothesis that these truncated proteins would be enzymatically inactive was tested by in vitro assays of LA2106 protein expressed in Escherichia coli. Indeed, in contrast to AT2 from the supercluster 1 accessions LA1777 and LA1731, the truncated LA2106 AT2 protein failed to acetylate triacylsucroses (Fig. 9; Supplemental Fig. S6). We did not find AT2 RNA-seq reads from LA2106 (Table II; compare LA2106 with the supercluster 1 line LA1777), consistent with known examples of nonsense alleles causing mRNA destabilization (Chang et al., 2007). Thus, the lack of acylsucrose acetylation in these four supercluster 2 subgroup D accessions is correlated with an inability to produce full-length and active AT2 enzyme.

In contrast, subgroup E AT2 gene sequences show no polymorphisms predicted to cause defective enzyme activities (Supplemental Fig. S5). In vitro enzyme assays confirmed that AT2 from LA2098 expressed in the yeast Pichia pastoris is indeed capable of acetylating triacylsucrose substrates (Fig. 9; Supplemental Fig. S6). This presents a quandary: if the AT2 enzyme is active in vitro, why do subgroup E accessions accumulate little acetylated acylsucroses? The first line of evidence was obtained from RNA-seq analysis, where very few...
reads correspond to AT2 transcript in LA2098 (subgroup E) compared with the higher transcript abundance found in supercluster 1 accession LA1777 (Table II). Reverse transcription (RT)-PCR analysis both validated the reduced expression of the AT2 gene in LA2098 and extended this observation to the subgroup E accessions LA1352 and LA2204 (Fig. 10; compare with supercluster 1 accession LA1777). These results are consistent with the hypothesis that these three subgroup E accessions accumulate relatively low amounts of AT2 transcript, causing reduced acylsugar acetylation.

The three supercluster 2 subgroup C accessions represent two additional types of loss-of-function AT2 alleles. Evidence that two of these AT2 coding regions, LA1624 and LA1625, had strongly diverged was first obtained when no PCR product was observed using a pair of primers that amplify the full AT2 coding region sequence in all other accessions tested (Fig. 11, A and C; primers a and f in Fig. 11B, and primer DNA sequences in red in Supplemental Fig. S5). In contrast, the third subgroup C accession (Fig. 4), LA1265, yielded a full-length PCR product (Fig. 11C). These results raised the possibility that the negative results from the two accessions were due to a local divergence in the region of primer a or f. To address this question, three primers downstream of a (b–d) were tested with primer e, which is approximately 300 nucleotides upstream from primer f, for all three subgroup C accessions (Fig. 11C). Although each of these primer pairs amplified AT2 from subgroup E positive control LA2098 as well as subgroup C accession LA1265, no PCR product was obtained from LA1624 or LA1625. The possibility that poor-quality gDNA prevented PCR amplification was addressed by successfully amplifying a control that spans an intron of Elongation factor1a (EF1a) from the gDNA template (Fig. 11C). Taken together, these results are consistent with the hypothesis that AT2 from the accessions LA1624 and LA1625, collected from populations geographically close to one another (Supplemental Table S1), have diverged to the point where multiple PCR primers failed to amplify their gDNA. In contrast, gDNA sequence from the third

Figure 9. Results of in vitro AT2 enzyme assays and comparison of predicted AT2 amino acid sequences of variants. A, Accessions LA1777, LA1731, and LA2098 AT2 are able to use acetyl-CoA to acetylate triacylsucroses, with the major products corresponding to S4:17 (2,5,5,5) and S4:24 (2,5,5,12). The truncated LA2106 AT2 protein was inactive under these conditions. The substrate used to assay for AT2 triacylsucrose acetylation was purified from IL1-3 trichomes as described (Schilmiller et al., 2012a). B, A part of the predicted translation products of S. lycopersicum M82 and representatives of each acylsugar phenotypic subgroups are shown. LA2106 AT2 is predicted to encode a truncated protein due to the introduction of a premature stop codon by frameshift caused by the deletion of a single nucleotide, as highlighted in the box. The DFGWG motif conserved among the BAHD enzyme superfamilies (Ma et al., 2005) is underlined and highlighted with a box. Chemical subgroup is represented as an extension to the accession number (e.g. LA1777_A for subgroup A).
subgroup C accession, LA1265, revealed deletion of four amino acids and the nonconservative mutation of Asp, the first amino acid of the DFGWG domain conserved in BAHD enzyme superfamily members (Ma et al., 2005), to Gly. These results are consistent with the hypothesis that accession LA1265 encodes an inactive AT2 enzyme (Fig. 11D).

Taken together, the analysis of variation in S. habrochaites AT2 genes revealed that supercluster 1 accessions encode functional AT2 proteins, whereas AT2 variants from supercluster 2 accessions were nonfunctional due to mutations that affect the production of active protein (four accessions from subgroup D and three accessions from subgroup C) or because transcript accumulation was very low (three accessions in subgroup E). These findings are consistent with the hypothesis that the presence or absence of abundant acetylated acylsucroses in S. habrochaites is associated with expression or activity of the AT2 enzyme.

DISCUSSION

Evolution of specialized metabolism is quite dynamic compared with older and more highly conserved central metabolic pathways (Milo and Last, 2012), and there is increasing evidence that the Solanum spp. SGT specialized metabolic pathways are especially evolutionarily plastic (Sallaud et al., 2009; Schilmiller et al., 2009; Gonzales-Vigil et al., 2012). This study revealed that S. habrochaites acylsugars are an unusually diverse group of Glc- and Suc-based polyesters with short- to medium-length (C2–C12) acyl chains. These metabolites are produced in the SGTs of plants in the Solanaceae, sometimes exuded in large enough amounts to cause the surface of the tissue to become sticky. Understanding the chemodiversity and biosynthetic pathways of these and other glandular trichome-produced molecules is of interest because they have documented roles in plant defense (Kang et al., 2010a; Mirnezhad et al., 2010; Weinhold and Baldwin, 2011).

Previous studies identified a glucosyltransferase from S. pennellii LA0716 thought to catalyze early steps of acylglucose biosynthesis (Ghangas and Steffens, 1993). Published data also revealed that acyl chains of C4 and larger are derived from branched-chain amino acid metabolic intermediates (Kandra et al., 1990; Walters and Steffens, 1990). More recently, reverse genetics approaches provided evidence that species-specific pathways for the elongation of acyl chains operate in S. pennellii and tobacco (Nicotiana tabacum) SGTs (Slocombe et al., 2008). Nevertheless, detailed

Table II. Summary of differences in gene expression and enzyme activity of S. habrochaites AT2 alleles

<table>
<thead>
<tr>
<th>Supercluster</th>
<th>Subgroup</th>
<th>Accession</th>
<th>DNA Sequence</th>
<th>Transcript Abundance</th>
<th>Enzyme Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>LA1777</td>
<td>Full</td>
<td>47.7</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>LA1731</td>
<td>Full</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>LA1624</td>
<td>Not amplified</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>LA2106</td>
<td>Full with premature stop codon</td>
<td>0.00</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>LA2098</td>
<td>Full</td>
<td>0.14</td>
<td>+</td>
</tr>
</tbody>
</table>

Figure 10. Lower mRNA accumulation of the AT2 gene in subgroup E accessions. RT-PCR results revealed lower expression of AT2 in subgroup E accessions LA1352, LA2098, and LA2204 than in supercluster 1 accession LA1777. The subgroup D LA2106 frameshift and early termination allele also exhibited little accumulation of AT2 transcript, consistent with RNA-seq data from total trichome RNA (Table II; for AT2 FPKM values from seven S. habrochaites accessions including LA1777 from subgroup A, LA2106 from subgroup D, and LA1352 and LA2098 from subgroup E, see “Materials and Methods”). The chemical subgroup for each accession is indicated next to the accession number. Elongation factor1a (EF1a) was used as a control. Z,Z-Farnesyl diphasphate synthase (zFPS) was also included to independently compare semiquantitative RT-PCR and quantitative RNA-seq data because our trichome RNA-seq analysis had revealed that the transcript abundance of zFPS was approximately 150-fold lower in LA1352 than in LA1777, LA2106, and LA2098. Because AT2 lacks introns, it was not possible to distinguish between signal from low-abundance RNA and small amounts of contaminating gDNA contamination. gDNA contamination was assessed by using primers designed based on intron sequences of EF1a and zFPS. Amplification of gDNA suggests that the minor amount of AT2 PCR products from supercluster 2 accessions is from gDNA rather than complementary DNA.
Figure 11. PCR analysis suggests that subgroup C accessions have nonfunctional AT2 alleles. A, PCR amplification of the AT2 coding region from representative accessions of each chemical subgroup. Note that no amplification was obtained from the subgroup C accession LA1624. Accession name and chemical subgroup are presented. The PCR analysis was performed using primers [a] and [f] (for primer locations, see B). The PCR products were confirmed as AT2 by sequence analysis. B, Locations of primers used to amplify AT2. [a] to [d] and [e] to [f] are forward and reverse primers, respectively. For the primer sequences, see Supplemental Figure S5. C, Results of PCR analysis for three subgroup C accessions (1, LA1265; 2, LA1624; 3, LA1625) with positive control subgroup E accession LA2098 (4). LA1624 and LA1625 revealed no PCR products with various combinations of primers for AT2, whereas LA1265 amplified the full length. The combinations of primers employed are presented in brackets. Positive control EF1a was used to test the quality of gDNA in this analysis. D, Comparison of AT2 protein sequences from LA1265 and functional LA1777. LA1265 AT2 protein is predicted to have a deletion of four amino acids as well as a point mutation in a protein motif conserved across acyltransferases (Ma et al., 2005); the conserved region is underlined. 

information is lacking regarding the biosynthetic pathway for any individual acylsugar.

We are taking a multifaceted approach to understand the biosynthesis and chemical diversity of SGT specialized metabolites in tomato and wild relatives (http://www.trichome.msu.edu). Previous work with a limited number of accessions of several Solanum species revealed differences in the sugar backbone and overall types of acyl chains (Schilmiller et al., 2010a, 2012a). Analysis of acyl sugar diversity in S. lycopersicum M82 × S. pennellii LA0716 introgression lines led to the identification of AT2, a BAHD acyltransferase that uses acetyl-CoA to acylate tracylsugars (Schilmiller et al., 2010a, 2012a). We extended this work by intensive study of S. habrochaites because of the availability of dozens of accessions, with geographic information about collection sites for many of these lines (http://trgrc.ucdavis.edu/). Despite some experimental shortcomings of working with this species, including self-incompatibility and lack of a reference genome sequence, the comparative chemical approach was fruitful in revealing patterns of trichome terpene diversity and the underlying biochemical mechanisms (Gonzales-Vigil et al., 2012).

While authentic standards are lacking for all but one of the acylsugars being studied, reverse-phase LC separation coupled with high-mass-accuracy negative-ion mode TOF-MS provides information about the sugar backbone as well as acyl chain number and lengths for 36 SGT acylsugars detected in this study (Fig. 1; Table I; Supplemental Fig. S2). All but two are Suc esters; the Glc triesters, G3:15 (5,5,5) and G3:22 (5,5,12), found in a total of five accessions in subgroups C and E, represented 7% or less of total acylsugars in these lines (Figs. 4 and 5). Total acylsugar amount was estimated to vary by 8-fold across the 80 accessions (Supplemental Fig. S4), and diversity in numbers and lengths of acyl chains was found that was well beyond expectations for accessions of plants in the same species, suggesting considerable plasticity in the biosynthesis of these specialized metabolites (Figs. 4–6).

The Influence of AT2 on Acylsugar Chemotypes

Based on HCA and PCA of the acylsugar profiles, the accessions define two superclusters and at least five subgroups, and these results are helpful in revealing patterns in the complex data set. One general observation is that the M82 cultivated tomato had less diverse acylsugar profiles than most of the wild species accessions (Figs. 3 and 4). This presumably is due to genetic bottlenecks introduced during domestication and breeding for S. lycopersicum varieties with desirable agronomic properties (Miller and Tanksley, 1990; Frary et al., 2000; Powell et al., 2012). Our analysis revealed several major qualitative differences in SGT acylsugar profiles. A main distinction is the presence or absence of C2 acyl chains in predominant acylsugars, which is a key feature separating superclusters 1 and 2 (Figs. 4–6). Within the tetraacylsucrose-dominated supercluster 1, subgroup B accessions mainly have C2, C4, and C5 acyl chains, whereas subgroup A acylsugars have a more varied range of acyl chain lengths. In addition to the general lack of C2 esters, supercluster 2 has a diverse array of acylsugars, including pentaacylsucrose S5:25 (5,5,5,5,5), C6-containing esters, and the Glc triesters G3:15 (5,5,5) and G3:22 (5,5,12) (Figs. 4–6).

Interestingly, four of six of the S. lycopersicum M82 acylsugars detected contain C2 acyl chains, in common with the chemotypes of nearly 30 S. habrochaites supercluster 1 accessions. This is in contrast to the supercluster 2 accessions, which either do not have
detectable C2-containing acylsugars or only accumulate measurable amounts of one C2-containing metabolite, S4:25 (2,5,6,12), based on its current annotation (Table I). This makes the acylsugar chemotype of the supercluster 1 S. habrochaites accessions more similar to that of S. lycopersicum M82 than to other conspecific accessions. This observation led to the hypothesis that production of acetylated acylsugars predates the divergence event that led to these extant species.

The recent identification of AT2 as the enzyme that acetylates triacylsugars in the cultivated tomato S. lycopersicum M82 provided a tool to test this hypothesis. Our study with S. habrochaites provides evidence that the AT2 gene existed in a common ancestor and was lost more than once since S. habrochaites species establishment through loss-of-function mutations in protein coding sequences or altered gene expression. These apparently independent changes include two haplotypes with deleterious changes in amino acid sequence. Three subgroup D lines have the same single nucleotide deletion predicted to create a truncated protein and also causes reduced mRNA accumulation in LA2106 (Figs. 9 and 10; Supplemental Fig. S5). The LA1265 subgroup C accession is predicted to encode a mutant AT2 protein with a four-amino-acid acid deletion as well as an Asp-to-Gly substitution in a region highly conserved in BAHD acyltransferases (Fig. 11D). The discovery of subgroup E accessions encoding an enzymatically active AT2 that is expressed at low levels compared with tetraacylsucrose-producing supercluster 1 accessions is another type of change leading to a reduction in acetylated acylsugars. Although our data do not allow us to determine whether the reduced mRNA expression of AT2 in subgroup E accessions is caused by mutation at the AT2 locus or a change in a trans-acting factor (for instance, a transcription factor or protein that is involved in mRNA stability), these alleles presumably evolved independently of the subgroup D alleles and LA1265. The third type of supercluster 2 AT2 allele (LA1624 and LA1625 in subgroup C) has diverged to the point where we were unable to amplify it using multiple pairs of ShAT2 PCR primers (Fig. 11). Because DNA sequence was not obtained for these alleles, we cannot distinguish between the possibility that they evolved from degeneration of a subgroup D or E allele or represent a fourth independent loss of enzyme function.

Another example of AT2 gene loss was described in studies of introgression lines IL1-3 and IL1-4, which contain the S. pennellii LA0716 AT2 region introgressed into the M82 cultivated tomato genome (Schilmiller et al., 2012a). The substitution of an inactive AT2 allele causes the production of nonacetylated triacylsugars in place of the normally abundant tetraacylsugars in these introgression lines. Presumably, the loss of AT2 function in S. pennellii LA0716 is related to the observation that this accession accumulates abundant acylglucoses rather than acylsugars. There are interesting open questions regarding the importance of AT2 activity outside of cultivated tomato and S. habrochaites.

Does this enzyme play a role in acylsugar acetylation beyond tomato and its close relatives? Do AT2 enzyme variants play roles in adding acyl chains of other lengths, on different sugar substrates, or at multiple positions in additional Solanum species or more distant relatives in the Solanaceae?

Geographical Context

The data showing geographical differences in S. habrochaites acylsugar profiles across Peru and Ecuador extend published studies of S. pennellii accessions (Shapiro et al., 1994). Shapiro and coworkers (1994) reported that accessions from the southern part of the range, including the widely studied LA0716, predominantly accumulate Glc esters with C4 to C12 chains. As S. pennellii accessions were sampled going from south to north of their range in the Andes Mountains, an increase in the ratio of Suc to Glc esters was observed.

There is a strong prevalence of C2-containing tetraacylsugars in southern S. habrochaites accessions (supercluster 1), whereas supercluster 2 accessions predominate from northern Peru and Ecuador (Fig. 7). The collection sites of supercluster 1 accessions have relatively higher elevation and lower precipitation and temperature than supercluster 2 accessions (Supplemental Table S1). These findings raise the question of whether the geographic clustering of plants with similar acylsugar chemistry results primarily from incomplete population dispersal, or is the result of adaptation to the local abiotic or biotic environments, or a combination of the two. Our discovery of multiple examples of AT2 activity loss is unusual (Chan et al., 2010) and consistent with the idea that specific acylsucrose structures provide a selective advantage in a local environment (Weinhold and Baldwin, 2011). In one scenario, if a population is established by movement to a new environment where acetylation provides no selective advantage, the loss of AT2 activity would occur without penalty. Additionally, if there is a fitness cost associated with the production of acetylated acylsucrose, inactivation of the AT2 gene might provide a selective advantage. The possibility of achieving a detailed understanding of the nature of the selective agent(s) impacted by tetraacylsugars or triacylsugars is exciting and will require study in the native habitats of these plants (Prasad et al., 2012). Identification of more enzymes of acylsugar biosynthesis will also allow a deeper understanding of how the large numbers of chemotypes are generated and whether the strong diversification seen for AT2 is the exception or the rule in S. habrochaites.

CONCLUSION

LC-TOF-MS screening of S. habrochaites SGT acylsugars revealed diverse acysugar chemotypes beyond the presence and absence of the acetyl group. The varied lengths of acyl chains and combinations of
numbers and types of esters in the 36 acylsugar structures suggest the interplay of a varied set of enzymes in *S. habrochaites* acylsugar biosynthesis. Knowing which metabolites accumulate in specific accessions, for instance, the presence or absence of longer chain fatty acid esters, the accumulation of unusual length acyl chains such as C6 or C8, or the synthesis of pentaacysucrose S5:25 (5,5,5,5,5), will inform the choice of accessions to study the biosynthesis and functions of acylsugars.

Another approach that should be of value in selecting accessions for study is to consider phenotypic relationships revealed by HCA (Fig. 4). Subgroup A accession LA1777 shows different clusters of abundant acylsugars that illustrate this point. For instance, S4:22 (2,5,5,10) and S4:23 (2,5,5,11) could reflect their synthesis from the common biosynthetic intermediate S3:12 (2,5,5), as diagrammed on the right side of Supplemental Figure S7. Similarly, tight clustering of four major acylsugars in the HCA, S4:14 (2,4,4,4), S4:15 (2,4,4,5), S4:20 (2,4,4,10), and S4:21 (2,4,5,10), suggests that their production involves common biosynthetic enzymes (Supplemental Fig. S7, left side). Solving the structures of these compounds by NMR, coupled with analysis of mRNA or proteomics data from the relevant accessions, should provide candidate enzymes for detailed analysis of these biosynthetic pathways. Once the genes for these enzymes are known, it ought to be possible to infer the changes in gene sequences, gene expression, and enzyme activities that lead to some or all of the metabolic diversity. Furthermore, because there is information about the collection sites of many of these accessions, it should become possible to place the chemical diversity into an ecological context, including identifying abiotic and biotic factors that drove their evolution. This information will be helpful for suggesting breeding or transgenic plant strategies to improve stress tolerance in solanaceous plants.

**MATERIALS AND METHODS**

**Plant Growth, Metabolite Extraction, and Identification**

Germplasm for M82 cultivated tomato (*Solanum lycopersicum*) and 80 *Solanum habrochaites* accessions were obtained from the Tomato Genetic Resource Center (http://tgrc.ucdavis.edu/). Plants were grown and nonvolatile leaf surface and trichome metabolites were extracted as described previously (Schilmiller et al., 2010a). Two to three biological replicates (individual plants grown in the same chamber) were prepared for most *S. habrochaites* accessions, and eight replicates for M82 cultivated tomato were analyzed; only one plant was screened for accessions of *S. habrochaites* that had poor germination and growth (n = 1 for LA1253, LA1764, LA2812, and LA2976).

LC-TOF-MS (Shimadzu LC-20AD pumps, Waters LCT Premier mass spectrometer) was used for the chemical analysis of nonvolatile metabolites by a modification of published methods (Gu et al., 2010; Schilmiller et al., 2010a). Analytical signals obtained from negative-ion-mode LC-TOF-MS analysis were investigated to identify acylsugar-associated signals by using a series of criteria: relative mass defect, LC retention time, and m/z under low ionization conditions. Fragment ions produced from high collision energies in MS were investigated to identify acylsugar-associated signals by using a series of fragment ion masses generated using elevated collision energies. However, when a number of peaks with the same m/z were seen within a relatively small retention time window, minor peaks whose abundance is nearly above detection level were not chosen for peak integration. This approach resulted in 36 distinct acylsugars for this chemical analysis. The integrated peak area for each specific signal was then normalized to internal standard (propyl-4-hydroxybenzoate; Sigma-Aldrich; http://www.sigmaaldrich.com/) and leaf dry weight (Supplemental Table S2).

An estimate of the amount of total acylsugars was obtained by summing the normalized peak areas for 36 analytical signals (Table I). As the chemical analysis for 80 accessions was performed in three different batches with separate plant growth, harvests, and LC-MS performances, the value was further normalized to make a batch-to-batch comparison possible (Supplemental Fig. S4). Every batch included M82 cultivated tomato plants as a control (n = 8 in total; three, two, and three per planting and analysis, respectively), and their profiles were qualitatively similar with minor variation of quantity (Supplemental Fig. S3). Thus, the average of all eight M82 plants was used to normalize values from the individual batches. All statistical tests were performed by JMP 8.0 software (SAS Institute; http://www.jmp.com/). For HCA, the percentage composition of 36 analytical signals was computed individually for every plant extract, and the average was obtained for the individual accessions with available replicates. The averaged value was standardized by JMP 8.0 software before HCA so that all variables would have equal impact in the clustering process. Ward’s minimum variance method was used for HCA (Milligan, 1980).

For the analysis of Suc and Glc ester composition by PCA, the same data set was used as for HCA. The normalized average peak areas for 36 analytical signals were categorized to acylsucrases or acylglucoses. Acylsucrases were further grouped to 3S, 4S, or 5S categories according to the number of acyl chains. 3G was the only type of acylglucose identified in this study. The peak areas for each category were summed, their percentage composition was calculated, and PCA was performed (Fig. 5).

The normalized average peak areas were also used to calculate the composition of FA chains. The level of FAs in acylsugars was calculated based on metabolite annotation for each analytical signal. For instance, if the normalized peak area of S4:20 (2,4,4,10) is 1, the levels of C2, C4, and C10 were weighted 1, 2, and 1-fold, respectively. In this way, the level of FAs was estimated from the individual analytical signals. The resulting values were summed for individual FAs, ranging from C2 to C12, and then the percentage of each FA was calculated. Then, the composition was visualized by PCA (Fig. 6).

**The Geographic Distribution of *S. habrochaites* Accessions**

For the original collection site of *S. habrochaites* accessions, latitude and longitude information was obtained from the Tomato Genetic Resource Center, except for the accessions lacking the information in the database (Supplemental Table S1). Locations were mapped using DIVA-GIS software (http://www.diva-gis.org/). The information of elevation, annual precipitation, and annual temperature was obtained from WorldClim version 1.3 at 5 min of resolution (Supplemental Table S1; http://www.worldclim.org/).

**The Analysis of Natural Variation in ShAT2 Genes**

The full-length and different parts of the protein coding sequence for *ShAT2* were amplified from gDNA (for primer sequences, see Supplemental Table S3). gDNA was isolated from leaf tissues of 3-week-old plants using the Qiagen DNA Isolation Kit. PCR was performed with Pfx DNA polymerase (Invitrogen), and PCR products were verified by DNA sequencing, or a modification of published methods (Gu et al., 2010; Schilmiller et al., 2010a). Analytical signals obtained from negative-ion-mode LC-TOF-MS analysis were investigated to identify acylsugar-associated signals by using a series of criteria: relative mass defect, LC retention time, and m/z under low ionization conditions. Fragment ions produced from high collision energies in MS were used for metabolite annotation. Details are given in Supplemental Text S1.

**Quantitative Data Analysis and Statistical Tests**

For quantitative analysis, peak areas were integrated for extracted ion chromatograms for the individual analytical signals using QuanLynx software (Waters; http://www.waters.com/). All peaks that have the same m/z and retention time were confirmed to share common acyl group lengths by assessing fragment ion masses generated using elevated collision energies. However, when a number of peaks with the same m/z were seen within a relatively small retention time window, minor peaks whose abundance is nearly above detection level were not chosen for peak integration. This approach resulted in 36 distinct acylsugars for this chemical analysis. The integrated peak area for each specific signal was then normalized to internal standard (propyl-4-hydroxybenzoate; Sigma-Aldrich; http://www.sigmaaldrich.com/) and leaf dry weight (Supplemental Table S2).

The full-length AT2 open reading frame was cloned from LA1731 and LA2098 (for primer sequences, see Supplemental Table S3). The PCR products

---

Kim et al.
were digested with corresponding restriction enzymes KpnI and NolI and ligated into pPCZC vector (Invitrogen). The construct was transformed into *Pichia pastoris* strain X-33 cells (Invitrogen). A colony carrying the correct construct was induced following the instructions of the manufacturer. Isolated cultures were lysed at 4°C in 50 mM sodium phosphate, pH 7.4, 1 mM EDTA, 5% glycerol, and 5 mM β-mercaptoethanol by vortexing with 0.5-mm glass beads. The recombinant protein was purified using nickel-affinity chromatography including an elution step with 200 mM imidazole followed by a washing step with 20 mM imidazole.

The full-length AT2 open reading frame was cloned from LA1777 and LA2106 (for primer sequences, see Supplemental Table S3). The PCR products were digested with restriction enzymes NdeI and XhoI and ligated into pET28a vector (Novagen). The construct was transformed into BL21 (Rosetta) (Novagen). A colony containing the correct construct was grown in Luria-Bertani medium at 37°C to an optical density at 600 nm of 0.8. Then, cultures were induced with 0.05 mM isopropyl-β-D-thiogalactopyranoside and incubated at 15°C for 12 h. The induced cultures were lysed by sonication in the identical lysis buffer used for *Pichia* spp. cultures, and the recombinant protein was purified as described previously. Ten microliters of each fraction was loaded onto a 10% SDS-PAGE gel and stained for protein using Coomassie blue.

In Vitro Enzyme Assays of ShAT2

Five micrograms of the purified recombinant protein was assayed in 30 μL of lysis buffer containing 100 mM acetyl-CoA and triacylsugar substrates. The triacylsugar substrates were collected from *Solanum pennelli* I-L4 plant extracts as described (Schilmiller et al., 2012a). After 30 min of incubation at 30°C, the reaction was terminated by adding 60 μL of acetonitrile/isopropanol/formic acid solution (1:1:0.01, v/v/v). Ten microliters of the reaction mixture was subjected to LC-MS analysis following previously published methods (Schilmiller et al., 2010a).

RNA-seq Analysis

Total trichomes were collected from stems and petioles of two to three fully grown plants of individual *S. habrochaites* accessions. Trichomes were harvested by freezing the plant tissues with liquid nitrogen and scraping off the trichomes from the frozen tissues. Total RNA was isolated from the collected trichomes using the RNAeasy Plant Mini Kit (Qiagen). RT was performed using SuperScript II reverse transcriptase (Invitrogen) following the manufacturer’s recommendation. RT-PCR analysis was performed under the conditions of 45 s at 94°C, 45 s at 55°C, and 1 min at 72°C for 22, 30, and 35 cycles (for primer sequences, see Supplemental Table S3).

Sequence data from this article can be found in the GenBank/EMBL data libraries under the following accession numbers: JX148287 to JX148289, D8571836, SRX159744, SRX159100, SRX159765, SRX159766, SRX159767, SRX159768, and SRX159769.

**Supplemental Data**

The following materials are available in the online version of this article.

**Supplemental Figure S1.** Light microscope images for leaf trichomes and whole leaf from three *S. habrochaites* accessions.

**Supplemental Figure S2.** An example of the use of LC-TOF-MS with multiplexed collision-induced dissociation to annotate acylsugar S426 (5,5,11).

**Supplemental Figure S3.** Test of reproducibility in acylsugar quantitation and composition in eight M82 replicate plants.

**Supplemental Figure S4.** Normalized total acylsugar levels in *S. habrochaites* accessions.

**Supplemental Figure S5.** Comparison of *S. habrochaites* AT2 alleles and its homologous genes.

**Supplemental Figure S6.** Expression and purification of *S. habrochaites* AT2 proteins.

**Supplemental Figure S7.** Hypothetical biosynthetic pathways to produce a subset of major acylsugars in LA1777.

**Supplemental Table S1.** Geographic and climate information on the collection sites for the *S. habrochaites* accessions used in this study.

**Supplemental Table S2.** Normalized peak areas for acylsugar-associated analytical signals.

**Supplemental Table S3.** Sequences of primers used in this study.

**Supplemental Text S1.** Chemical analysis by LC-TOF-MS.

**AKNOWLEDGMENTS**

We thank Amanda Charbonneau for technical assistance and Dr. Tony Schilmiller and other past and present members of the *Solanum* Trichome Project (www.trichome.msu.edu) for helpful discussions. We thank Drs. Robin Buell, Kevin Childs, and Elsa Gongora and members of the Buell laboratory for providing J.K. with training for RNA-seq data analysis. We thank the C.M. Rick Tomato Genetics Resource Center at the University of California, Davis, for *S. habrochaites* and *S. pennelli* seed stocks. R.L.L. thanks the Lady Davis Foundation of the Hebrew University in Jerusalem, Israel, and the Weizmann Institute of Science in Rehovot, Israel, for generous sabbatical support.

Received August 3, 2012; accepted October 6, 2012; published October 9, 2012.

**LITERATURE CITED**


Schilmiller AL, Pichersky E, Last RL (2012b) Taming the hydra of specialized metabolism: how systems biology and comparative approaches are revolutionizing plant biochemistry. Curr Opin Plant Biol 15: 338–344


Weinhold A, Baldwin IT (2011) Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation. Proc Natl Acad Sci USA 108: 7855–7859

