GENETIC AND GENOMICS APPROACHES TO UNDERSTANDING THE BIOSYNTHESIS OF SPECIALIZED METABOLITES IN TRICHOMES OF THE CULTIVATED TOMATO AND ITS WILD RELATIVES

By

Jeongwoon Kim

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Plant Biology

2011

ABSTRACT

GENETIC AND GENOMICS APPROACHES TO UNDERSTANDING THE BIOSYNTHESIS OF SPECIALIZED METABOLITES IN TRICHOMES OF THE CULTIVATED TOMATO AND ITS WILD RELATIVES

By

Jeongwoon Kim

Trichomes are specialized epidermal appendages that cover the surface of plant tissues. Glandular secreting trichomes produce a variety of plant specialized metabolites. Trichome metabolites across the kingdom Plantae are extremely diverse and include those important for plant defense and therapeutic purposes. Due to these benefits, biosynthetic pathways leading to the production of trichome metabolites haven been intensively studied. Recent advances and development of analytical chemistry and genetic and genomic tools facilitate the study of trichome biochemistry. In this study, analytical chemistry, genetic and genomics approaches were used to understand biosynthetic pathways for the production of non-volatile metabolites in Solanum trichomes. Specifically, this study included three individual projects: (i) identification of tomato EMS mutants altered in biosynthetically diverse trichome non-volatile metabolites and phenotypic characterization of glycosylated flavonoid mutant (ii) identification and analysis of *Solanum lycopersicum* genetic variants altered in trichome methylated myricetin biochemistry, and (iii) chemical analysis of diverse trichome acylsugar profiles in 80 accessions from the wild species of cultivated tomato *S.*

habrochaites.

ACKNOWLEDGEMENTS

First of all I would like to thank Dr. Robert Last for his consistent support, encouragement and patience, which helped me explore and enjoy working on science despite great challenges from tough course work, exams, unexpected experimental results and endless learning during my graduate school years.

I would also like to thank my committee members Drs. Dan Jones, Gregg Howe and Dean DellaPenna for their constructive and supportive advice. Especially, I appreciate Dr. Dan Jones for teaching me mass spectrometry performance and data analysis.

In addition, it was wonderful to work with Last lab members and tomato trichome project members. Scientific discussion and everyday conversation were among the most pleasant parts of time in graduate school. I especially thank Dr. Tony Schilmiller and Ms. Amanda Charbonneau for their suggestion and help. I would also like to thank my friends at MSU for being great companies sharing the sense of achievement and concerns.

Last, I would like to thank my family, my parents and a brother, for their love and supports and being there. I would have not been able to make any achievement in my life without them. Thank you and love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction: understanding plant specialized metabolites using the Solanum trichome as a model system

Plant specialized (secondary) metabolites

A variety of low-weight molecules are produced as intermediates or final products from chemical reactions involved in plant metabolism. These compounds are commonly categorized to two groups, primary and secondary metabolites. *Primary metabolites* are defined as chemical compounds that are directly involved in plant growth and development (Aharoni and Galili 2011, Pichersky and Gang 2000, Schwab 2003). In contrast to primary metabolites, the term *secondary metabolite* refers to natural products that are traditionally thought to indirectly affect biological processes in plants (Aharoni and Galili 2011, Schwab 2003). To date, more than 200,000 plant secondary metabolites have been reported (Aharoni and Galili 2011). This term 'secondary metabolite' came from the traditional concept that these small molecules are not essential for plant survival. However, increasing numbers of studies revealed that these chemical compounds have critical impacts on fitness, for example, protecting plants against diverse environmental stresses (Baldwin *et al.* 2006, Kang *et al.* 2010a, Kang *et al.* 2010b) and promoting symbiotic interaction with beneficial organisms (Djordjevic *et al.* 1987, Zhang *et al.* 2009).

Despite the traditional view that plant metabolites can be divided to the categories of 'primary and secondary', there is a no solid rule to distinguish two groups of metabolites. One relatively simple and clear distinction is whether their biosynthesis is universal for

all plants or specific to different plant species. While the production of primary metabolites such as amino acids, sugars and lipids is common to all plant species (Aharoni and Galili 2011), secondary metabolite biosynthesis is taxon-specific (Pichersky and Gang 2000). For example, plants in the family *Brassicaceae* and genus *Catharanthus* produce glucosinolates (Wittstock and Halkier 2002) and indole alkaloids (Ziegler and Facchini 2008), respectively. Based on the taxon-specificity of secondary metabolite biosynthesis and discovery of increasing numbers of biological functions, the term 'specialized metabolite' has gained popularity in this field (Aharoni and Galili 2011, Pichersky and Gang 2000). This term will be used throughout the remainder of this dissertation.

The category of primary and specialized metabolites is determined based on the function of chemical compounds rather than on their structures. In fact, a number of specialized metabolites are derived from identical biosynthetic pathways using the same precursors as primary metabolites (Aharoni and Galili 2011, Frey *et al.* 1997). Therefore, some specialized metabolites are structurally similar to primary metabolites. For example, despite structural similarity (Buchana *et al.* 2002) (Figure 1.1), proline is categorized as a primary metabolite and the flowering-inducing molecule pipecolic acid as a specialized metabolite (Fujioka and Sakurai 1997).

Examples of the biosynthesis of plant specialized metabolites and their values

Among the various groups of plant specialized metabolites, this section will focus on two classes of compounds, flavonoids and acylsugars, because this dissertation work focuses

on these classes of compounds (See 'Chapters 2 and 3' for flavonoids and Chapter 4 for acylsugars, respectively). The biosynthetic pathways for terpenoids and alkaloids are also introduced briefly, because these are, along with flavonoids, two of the three major classes of plant specialized metabolites.

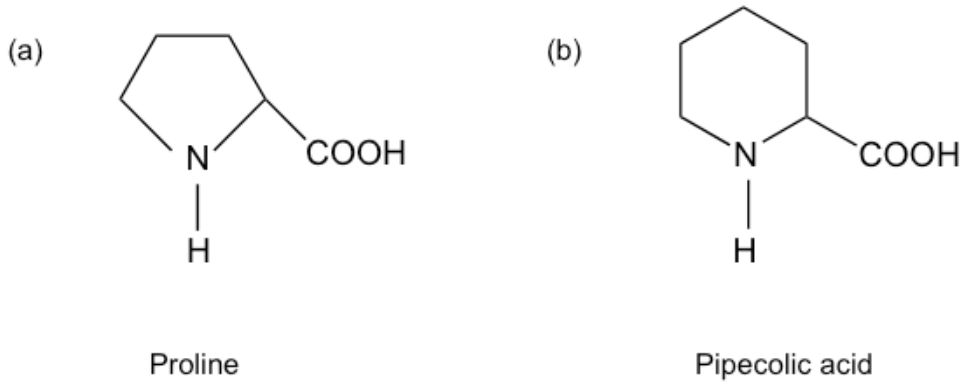


Figure 1.1. Structural similarity between the primary metabolite proline and specialized metabolite pipecolic acid.

Flavonoids

Flavonoids such as chalcones, flavonones, flavones, flavonols, isoflavones and anthocyanins are well-studied and widely distributed plant specialized metabolites (Boudet 2007, Buer *et al.* 2010, Dixon and Steele 1999). The biological roles of

flavonoids have been intensively investigated. Beside the impact on flower color (Mol *et al.* 1998), pollen viability (Taylor and Jorgensen 1992) and protecting plants against UV light (Li *et al.* 1993) and reactive oxygen species stress (Taylor and Grotewold 2005), flavonoids facilitate the interaction between plants and microorganisms such as in legume-bacterium symbiosis (Djordjevic *et al.* 1987) and nodule organogenesis in plant roots (Wasson *et al.* 2006, Zhang *et al.* 2009). In addition to the biological roles in plants, flavonoids were shown to have health benefits as anti-oxidants (Boudet 2007) and anti-microbials (Nowack and Schmitt 2008).

The biosynthesis of flavonoids begins with chalcone production by chalcone synthase using malonyl-CoA and coumaryl-CoA as substrates (Buchana *et al.* 2002) (Figure 1.2). Chalcone is then converted to structurally diverse flavonoid compounds such as flavanones, flavones, flavonols and isoflavonoids by reactions of oxidation, reduction and isomerization (Boudet 2007, Buchana *et al.* 2002, Buer *et al.* 2010, Dixon and Steele 1999) (Figure 1.3). The flavonoids are also modified by glycosylation and/or methylation, resulting in a wide diversity of chemical structures (Jones *et al.* 2003, Jorgensen *et al.* 2005, Matsuda *et al.* 2009, Schmidt *et al.* 2011).

A number of glycosylated flavonols have been identified from various plant tissues such as leaf tissues (Braca *et al.* 2001, Schmidt *et al.* 2011), stems and bark (Min *et al.* 2003), fruits (Reynertson *et al.* 2008), flowers (Modolo *et al.* 2009, Wu *et al.* 2008) and isolated glandular trichomes (Kang *et al.* 2010a). The structure of glycosylated flavonols vary depending on the types of flavonol aglycone and the types, numbers and positions of

sugars attached to flavonol ring structure (See Figure 1.2 for flavonoid ring structure and numbering system). The unmodified flavonol aglycone can be one of three flavonols, kaempferol, quercetin and myricetin, which differ by the number of hydroxyl groups on B ring (See Figure 1.2 for flavonoid ring structure and numbering system). While kaempferol has a hydroxyl group at C4', quercetin and myricetin have two and three hydroxyl groups at C3' and C4', and C3', C4' and C5' on the B ring, respectively. Many sugars, for instance, glucose, galactose, rhamnose, xylose or arabiose, are found in varying combinations and positions on glycosylated flavonols (Li *et al.* 2008).

The number of sugars in glycosylated kaempferol and - quercetin compounds is reported to range from one to five (Williams and Grayer 2004, Wu *et al.* 2008). When the flavonol is glycosylated with multiple numbers of sugars, individual sugars can be attached at the different positions or multiple sugars can be bonded to each other and then attached to a single position of flavonol. Glycosylation has been reported at the positions of C3, C7, C3' and C'4 in kaempferol or quercetin, among which C3 is the most often glycosylated position followed by C7 (Min *et al.* 2003, Modolo *et al.* 2009, Williams and Grayer 2004). The glycosyaltion of myricetin is also most commonly found at C3 (Braca *et al.* 2001, Williams and Grayer 2004). However, unlike the myriad of glycosylated derivates identified for kaempferol and quercetin, relatively few glycosylated myricetin compounds are known. In addition, while kaempferol and quercetin are often found glycosylated with

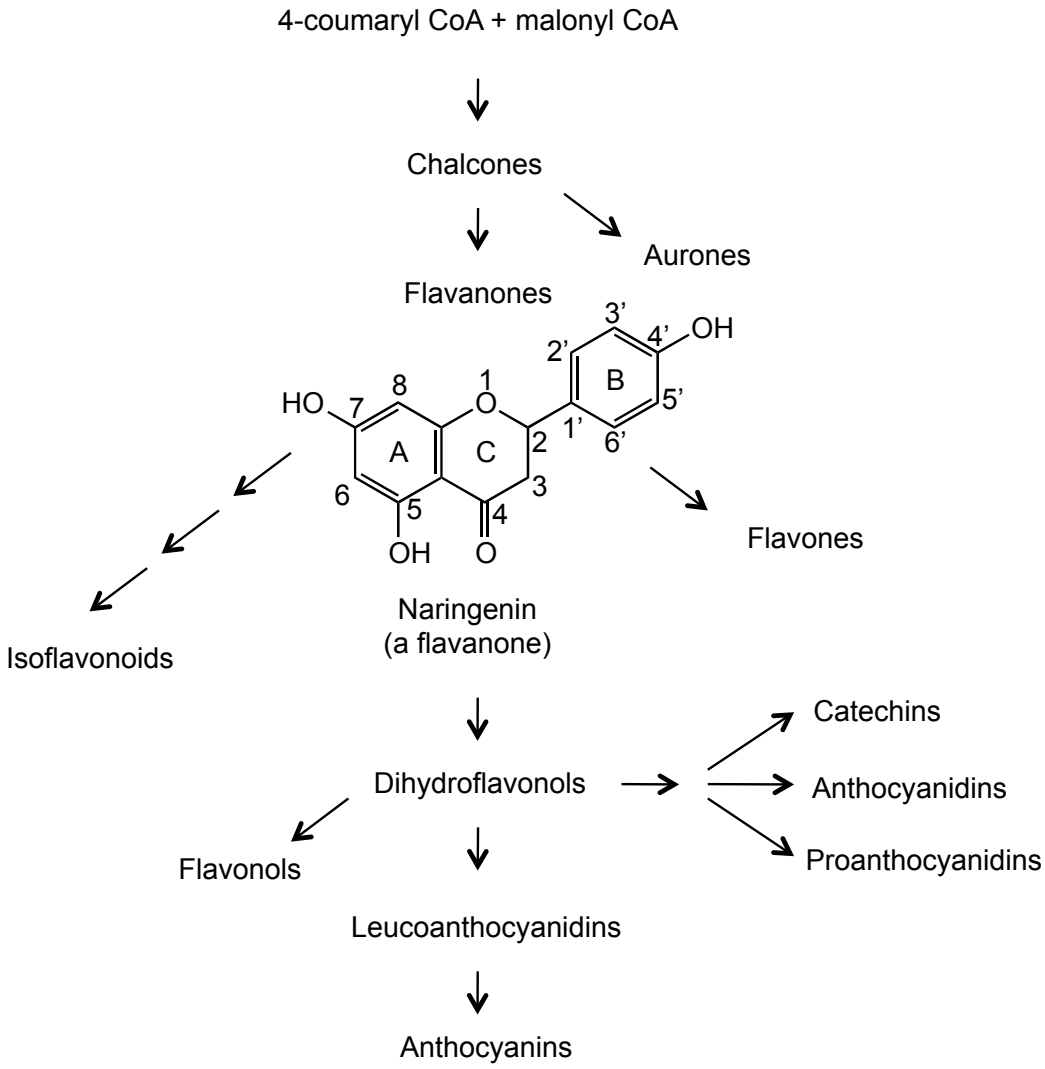


Figure 1.2. Biosynthetic pathways for flavonoid production.

Biosynthesis of flavonoid compounds starts with the production of chalcones. Chalcones are then converted to structurally diverse flavonoid compounds through a serious of chemical reactions. The chemical structure of naringenin is shown with numbering system for the ring structure. This biosynthetic pathway is modified from Buer *et al.* 2010.

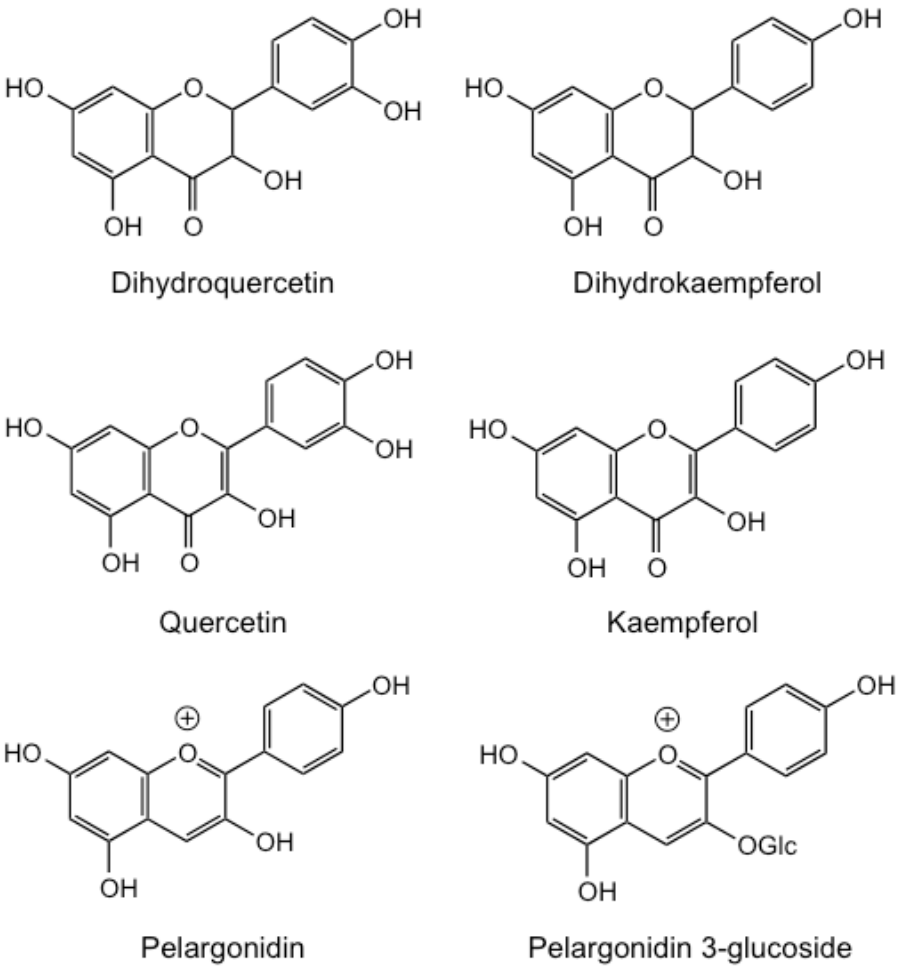


Figure 1.3. Examples of flavonoid and anthocyanin compounds.

polysaccharides, myricetin is mostly glycosylated with one or two sugars and highly glycosylated forms of myricetin are rare (Williams and Grayer 2004). Identification and characterization of a myricetin tetraglycoside compound is described in Chapter 2.

In addition, *O*-methylation of hydroxyl groups also contributes to the structural diversity of flavonol compounds. Structurally diverse methylated flavonols are described from various plant tissues including leaf (Kumari *et al.* 1984, Stevens *et al.* 1995), fruits (Brunet and Ibrahim 1986), flowers (Jay *et al.* 1982), roots (Ibrahim 2005) and isolated glandular trichomes (Roda *et al.* 2003, Schmidt *et al.* 2011, Stevens *et al.* 1995). Similar to glycosylated flavonols, the structures of methylated flavonols also vary by the types of flavonol aglycones and the numbers of methyl groups attached to flavonol backbone and their positions: C3, C5, C7 and available hydroxyl groups such as C3', C4' and C5' on B ring for the given flavonols (Ibrahim 2005, Kumari *et al.* 1984, Schmidt *et al.* 2011, Stevens *et al.* 1995) (Figure 1.2). Various plants accumulate polymethylated flavonols, of which *Chrysosplenium americanum* (Golden Saxifrage) is a well-studied example. The identification and characterization of *O*-methyltransferases involved in the production of polymethylated quercetins in *C. americanum* revealed that *O*-methylation at C3 position is the first committed step in the sequential methylation reactions (Collins *et al.* 1981, De Luca and Ibrahim 1982, De Luca and Ibrahim 1985, Ibrahim 2005).

In contrast to few glycosylated myricetin compounds known, a variety of methylated myricetin compounds were identified (Kang *et al.* 2010a, Schmidt *et al.* 2011, Stevens *et al.* 1995). For instance, previous study revealed that leaf trichomes from the wild relative

of cultivated tomato, *S. habrochaites* LA1777, accumulate polymethylated myricetins such as tri-, tetra- and pentamethylated myricetins, but not glycosylated myricetins (Schmidt *et al.* 2011). The lack of glycosylated myricetins raises a question whether myricetin tends to be methylated rather than glycosylated, whereas lack of methylation allows kaempferol and quercetin to be glycosylated. Especially, in that C3 of the flavonol is most likely the first position to be glycosylated or methylated (De Luca and Ibrahim 1982, De Luca and Ibrahim 1985, Ibrahim 2005), it is suspected that this position is critical to determine the modification type of myricetin in plants. In this dissertation study, various *S. lycopersicum* variants including an EMS mutant and *S. pennellii* introgression lines were identified as having altered trichome myricetin methylation phenotypes. Characterization of the altered methylated myricetin chemistry in these genetic variants is discussed in Chapter 3.

Acylsugars

Acylsugars are sugars esterified with acyl chains of various lengths at one or more hydroxyl groups and are found in plants of the family Solanaceae (Figure 1.4). Plants in the genus Solanum, including the cultivated tomato (*Solanum lycopersicum*) and its wild relatives, are known to accumulate acylsugars (Ghangas and Steffens 1993). Of particular note, the wild tomato plant *S. pennellii* LA0716 accumulates acylsugar exudates up to 20% of leaf dry weight, making the leaf surface sticky (Fobes *et al.* 1985, Goffreda *et al.* 1989). These sticky chemical compounds improve plant defense against herbivore insects by deterring insect feeding and oviposition (Hawthorne *et al.* 1992, Mutschler *et al.* 1996). The previous report that volatile branched chain aliphatic acids produced from the

metabolism of *Nicotiana attenuata* trichomes acylsugars by *Manduca sexta* caterpillars attract carnivorous ants is a fascinating example of indirect plant defense against the herbivore (Weinhold and Baldwin 2011). The diverse documented direct and indirect defensive roles of acylsugars suggest that these compounds are of widespread ecological importance.

Acylsugar compounds are structurally diverse, with differences in the sugar backbone (typically glucose or sucrose), and the number and length of individual acyl chains. In particular, chemical analysis of *S. pennellii* accessions revealed that differences in acylsugar chemistry are associated with the geographic distribution of accessions. For example, *S. pennellii* accessions from the southern part of the Andes mountains, including LA0716, predominately accumulate acylglucoses, whereas accessions from the northern range have higher ratios of acylsucroses to acylglucoses (Shapiro *et al.* 1994). This observation suggests that the diverse acylsugar profiles in *S. pennellii* accessions result from evolutionary adaptation. However, this question remains to be answered because relatively little is known about the biosynthetic and regulatory pathways leading to the production of chemically diverse acylsugar metabolites. Chemical diversity of acylsugar compounds and its correlation with the geography for 80 *S. habrochaites* accessions will be discussed in Chapter 4.

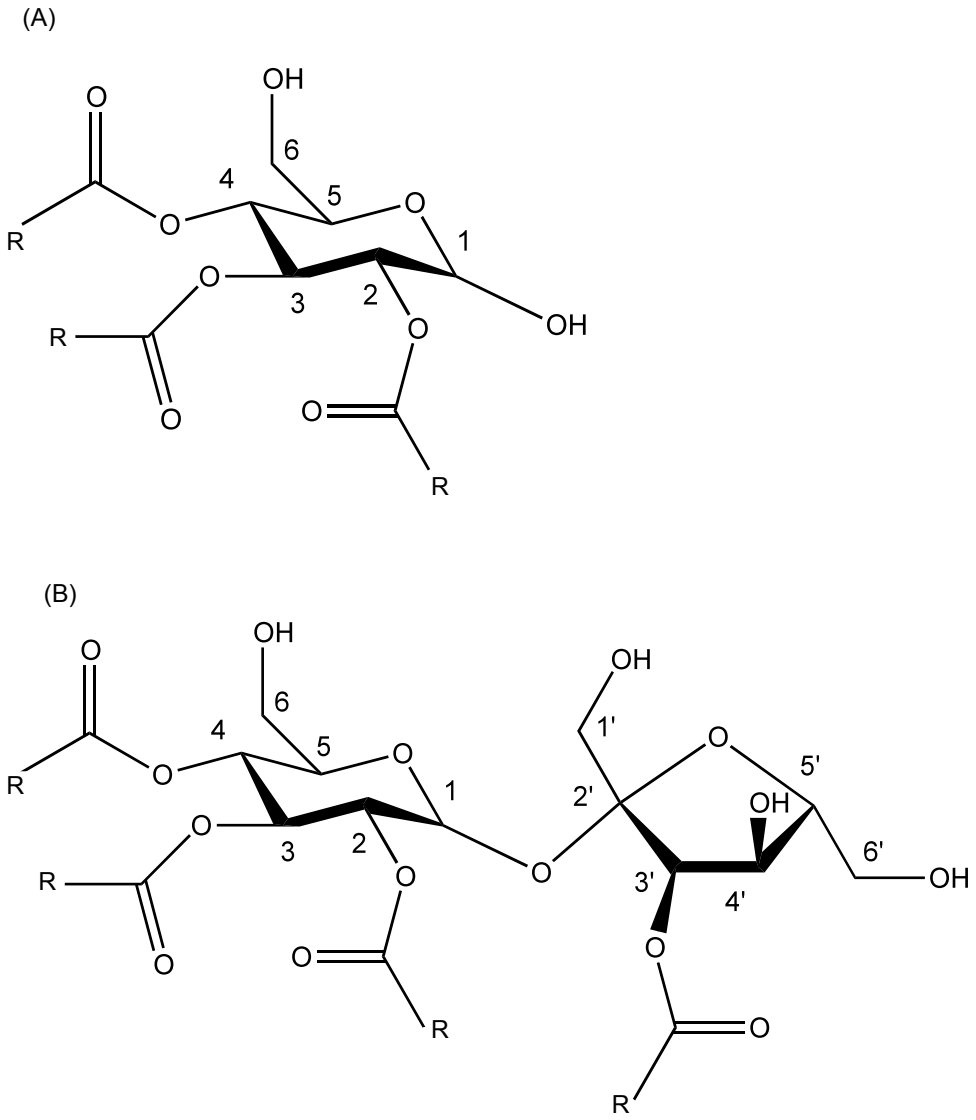


Figure 1.4. Chemical structures of acylsugars.

(a) Schematic structure of an acylglucose. This acylglucose consists of glucose esterified with three acyl chains. 'R' indicates various numbers of carbons in the individual acyl chains. Typical acyl chains from *S. pennellii* LA0716 range from C4 to C12.

(b) Schematic structure of an acylsucrose. The structure shown here depicts sucrose tetraester, a sucrose esterified with four acyl chains. Typical acyl chains from *S. lycopersicum M82* range from C2 to C12.

Terpenes

Terpenoids are structurally diverse compounds that include volatiles of commercial importance (Pichersky and Gershenzon 2002). Terpenes are defined based on the number of carbons in the terpene skeleton; for example, C_{10}, C_{15}, C_{20}, C_{30} and C_{40} for mono-, sesqui-, di-, tri and tetraterpenes, respectively (Buchana *et al.* 2002, Tholl 2006, Tholl and Lee 2011). Once terpene skeletons of various lengths are generated through the condensation of C_5 isoprene units, they are converted to structurally diverse terpene compounds through cyclization and modification such as glycosylation and oxidation (Pichersky and Gershenzon 2002, Tholl 2006) (Figures 1.5). Beside improvement of plant defense against herbivore insects (Baldwin *et al.* 2006), terpene compounds contribute to plant flavor and aroma (Pichersky and Gershenzon 2002, Schilmiller *et al.* 2008). Some terpenes are used as important drugs; for examples the antimalarial arteminisin (Liu *et al.* 2011) and anti-cancer agent taxol (Roberts 2007) (Figure 1.6). Collaborative work with Dr. Cornelius Barry's group in the Michigan State University Department of Horticulture revealed chemical diversity of terpene profiles in 80 *S. habrochaites* accessions and its correlation with geographic distribution of the individual accessions. This work is described in a manuscript entitled 'Evolution of TPS20-related terpene synthases influences chemical diversity in the glandular trichomes of the wild tomato relative *Solanum habrochaites*' submitted for publication in November, 2011.

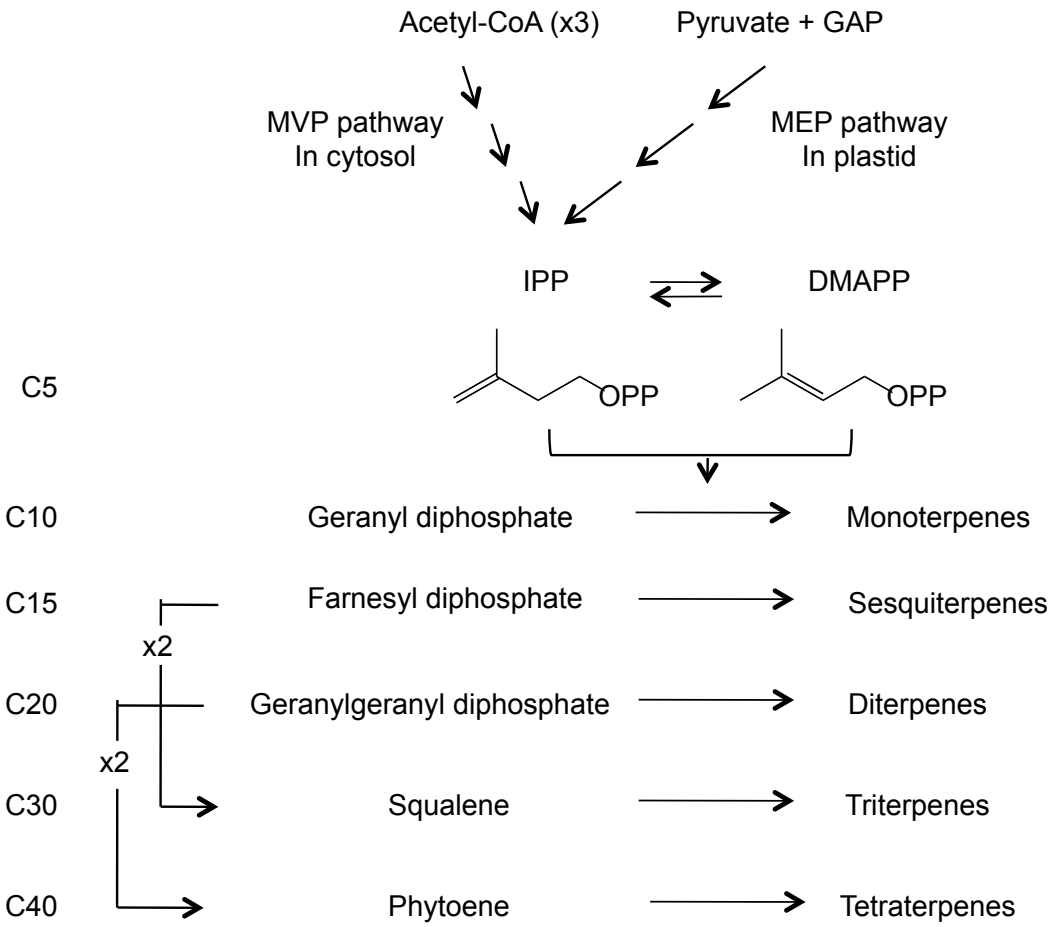


Figure 1.5. Overview of terpene biosynthesis.

Figure 1.5. Continued

The biosynthesis of terpene compounds begins with condensation of five carbon isoprene units, isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). IPP is synthesized by the mevalonate pathway (MVP) and methylerythritol phosphate pathway (MEP) operating in the cytosol and plastid respectively. 'GAP' indicates glyceraldehyde 3-phosphate. Rendering of the biosynthetic pathway is modified from Buchana *et al.* 2002. Previously *trans*-prenyldiphosphates such as E-geranyl diphosphate and E,E-farnesyl diphosphate were considered the universal intermediates for monoterpene and sesquiterpene biosynthesis, respectively. However, recent studies revealed that *S. lycopersicum* trichomes synthesize monoterpenes from the *cis*-prenyldiphosphate intermediate neryldiphosphate (Schilmiller *et al.* 2009), whereas the all *cis* Z,Z-farnesyl diphosphate intermediate is used to produce sesquiterpenes in *S. habrochaites* trichomes (Sallaud *et al.* 2009).

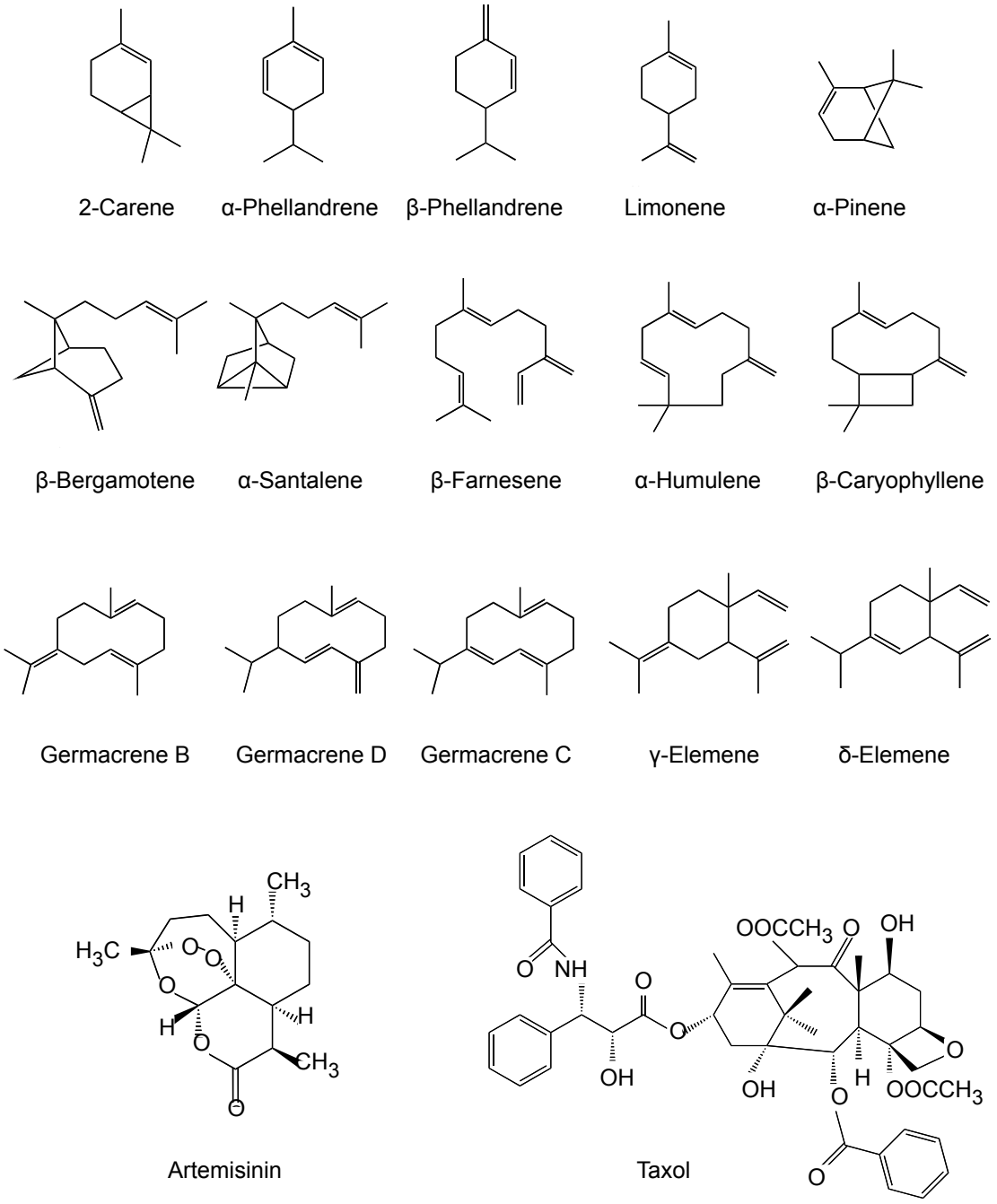


Figure 1.6. Examples of terpene compounds.

Alkaloids

Nitrogen-containing alkaloids include biosynthetically and structurally diverse chemical compounds (Facchini and St-Pierre 2005). In particular, biosynthetic pathways for the production of benzylisoquinoline- and monoterpenoid indole alkaloids have been intensively studied (Facchini and St-Pierre 2005, Ziegler and Facchini 2008) (Figure 1.7). Alkaloid compounds are also important pharmaceuticals; for example vinblastine is an anti-cancer drug and ajmaline is an efficient drug for antiarrythmic heart disorders (Figure 1.8).

Examination of metabolite profiles - metabolite profiling and metabolomics

Cataloguing the types of metabolites and their abundance is a prerequisite to understanding plant specialized metabolism. Metabolite cataloguing can be divided into two main categories: metabolite profiling and metabolomics (Fiehn 2002, Last *et al.* 2007). The main distinction between the two approaches is determined mainly by the number of metabolites in an analysis (Sweetlove *et al.* 2003). Metabolite profiling aims to detect and quantify relatively small numbers of metabolites, typically of similar chemical properties (Fiehn 2002, Last *et al.* 2007). As a result, metabolite profiling approaches typically are highly sensitive and accurate (Sweetlove *et al.* 2003). While metabolite profiling has strong advantages for studying compounds of known structure and physical properties, it is not well suited for more open-ended investigation of a large numbers of metabolites of diverse physical properties, including uncharacterized compounds. Metabolomics is an alternative approach where large numbers of structurally diverse compounds are analyzed (Fiehn 2002, Sweetlove *et al.* 2003). Although there are

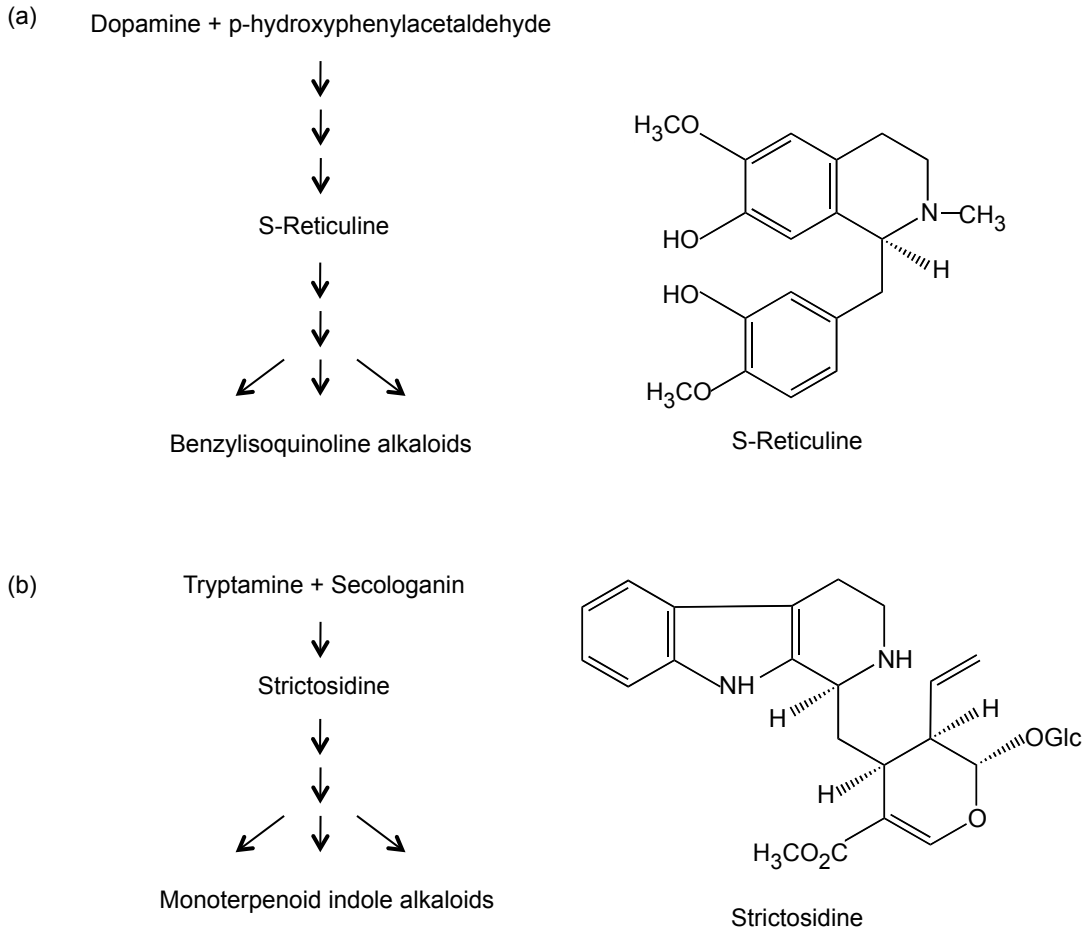


Figure 1.7. Biosynthetic pathways for alkaloid production.

(a) Benzylisoquinoline alkaloid production. S-reticuline is used to produce structurally diverse benzylisoquinoline alkaloids. The synthesis of S-reticuline is initiated with condensation of dopamine and *p*-hydroxyphenylacetaldehyde.

(b) Monoterpenoid indole alkaloid production. Strictosidine is produced by the condensation of tryptamine and secologanin and it serves as an intermediate for synthesis of diverse monoterpenoid indole alkaloid compounds.

The biosynthetic pathways (a-b) are modified from Facchini and St-Pierre 2005.

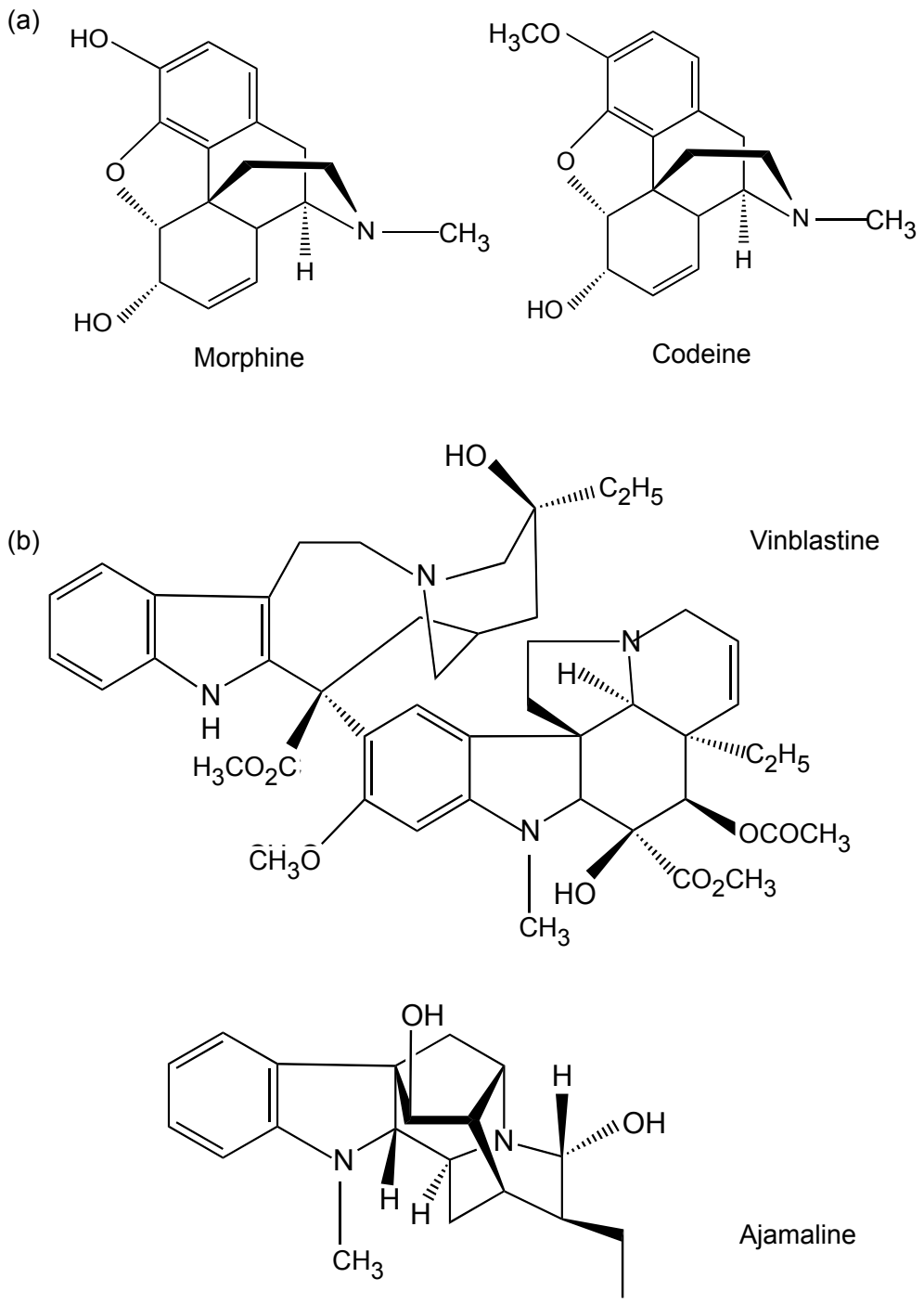


Figure 1.8. Examples of alkaloid compounds.

(a) Examples of benzylisoquinoline alkaloids

(b) Examples of monoterpenoid indole alkaloids

generally tradeoffs between the number of metabolites analyzed in a given method and the accuracy, reproducibility and selectivity of the assay, metabolomics is beneficial to understand a large number of metabolites even including poorly or uncharacterized metabolites in an open-ended analysis (Sumner *et al.* 2003).

Three major steps to understand plant specialized metabolite profiles

Investigation of the types of plant specialized metabolites and their abundance requires three major steps: 1. chemical extraction of the metabolites; 2. use of an analytical chemistry platform; 3. data analysis. Details of these individual steps are discussed below.

Chemical extraction of plant specialized metabolites

Because plants accumulate a myriad of specialized metabolites with varying chemical and physical properties, the choice of extraction system determines the types of metabolites harvested from plant materials (Kim *et al.* 2011, Last *et al.* 2007, Shin *et al.* 2010). Therefore, selecting an extraction system requires consideration of the chemical and physical characteristics of metabolites such as solubility and concentration range (Sweetlove *et al.* 2003). Furthermore, the accumulation of plant specialized metabolites is highly influenced by the environment. In addition, some classes of metabolites (for example, glucosinolates) are so highly reactive that special care must be taken to prevent their degradation or conversion during extraction or storage (Kim *et al.* 2011, Shin *et al.* 2010). Therefore, various methods for immediate quenching or cessation of all

metabolism are commonly used to prevent unwanted alteration of metabolite profiles during tissue harvest and chemical extraction (Kim *et al.* 2011).

Analytical chemistry platforms

Different analytical chemistry techniques have varying characteristics including sensitivity, selectivity and speed. Thus, understanding the benefits and technical limitations for each technique is essential in selecting suitable analytical instruments for specific experimental goals (Lei *et al.* 2011). Because liquid chromatography-mass spectrometry (LC-MS) was employed in this dissertation study, MS-based approaches are mainly discussed in this section.

MS is commonly used to investigate metabolite profiles due to three benefits. First, the high sensitivity of MS allows detection of low abundance metabolites down to the nanomolar concentration range (Last *et al.* 2007). Second, MS enables identification of a large number of metabolites based on characteristic mass to charge ratio (m/z) and quantification of specific metabolites by measuring signal intensity for the individual ion. Third, use of multidimensional or high mass accuracy MS can reveal information about the class of molecule being analyzed and guide structural elucidation of some chemical compounds (Lei *et al.* 2011, Sumner *et al.* 2003). The MS method is even more powerful when combined with chromatographic separation such as gas- or liquid chromatography. The availability of authentic standards enables unambiguous identification of compounds and heavy isotope-labeled standards permit accurate quantification of target metabolites (Gu *et al.* 2007).

Gas chromatography (GC)-MS is suitable for profiling volatile metabolites such as terpene compounds (Sallaud *et al.* 2009, Schilmiller *et al.* 2009). GC-MS is also used to analyze non-volatile metabolites after derivatization steps that increase the volatility of chemical compounds (Sumner *et al.* 2003). GC-MS has three major advantages: 1. complete ionization, minimizing ion suppression effects; 2. reproducible retention times, allowing comparisons of samples across time, instruments and laboratories; 3. availability of large mass spectral databases (Aprea *et al.* 2011, Kopka *et al.* 2005, Sawada *et al.* 2009, Sweetlove *et al.* 2003). These characteristics promote metabolite identification, data interpretation and cross comparisons of results from different studies and laboratories, making GC-MS an excellent tool to study plant specialized metabolites.

LC-MS techniques are suitable for analyzing non-volatile metabolites without derivatization and can be faster than GC-MS methods, makingLC-MS efficient for high-throughput analysis. These benefits make LC-MS one of the most popular analytical platforms despite existence of technical challenges such as matrix effects (i.e. ion suppression), lack of comparability of retention times and ionization characteristics and paucity of comprehensive mass spectral databases (Sumner *et al.* 2003, Sweetlove *et al.* 2003).

The combination of approaches such as GC- and LC-MS in the study of complex mixtures or elucidation of metabolite structures has advantages compared with the more rapid direct-infusion MS. For example, chromatographic separation provides additional

information such as retention time. Therefore, GC-MS or LC-MS can generate two-dimensional information indexes for the characteristic ions with m/z and retention time (referred to as 'analytical signals' in this dissertation). Furthermore, chromatographic separation prior to MS decreases the number of chemical compounds delivered to the MS at a given time, reducing ion suppression and allowing more accurate quantification and data analysis (Last *et al.* 2007, Lei *et al.* 2011).

Although MS-based approaches, especially combined with GC or LC, are efficient tools to study plant metabolites, MS analysis is not sufficient to unambiguously elucidate completely novel structures of chemical compounds and generally does not provide information about stereochemistry. This limitation can often be compensated by using nuclear magnetic resonance (NMR) methods (Kim *et al.* 2011, Sweetlove *et al.* 2003). NMR enables the identification and quantification of metabolites based on chemical shift, intensity, fine structure and magnetic relaxation properties. Because these signals are directly related to the chemical environment of nuclei in atoms in the metabolites, the NMR technique is suitable for elucidating the chemical structure of novel compounds, sometimes including stereochemistry (Krishnan *et al.* 2005). Disadvantages of these powerful methodologies are that they typically require relatively large amounts of a metabolite in a mixture for identification and quantification of known compounds and of relatively pure preparations of metabolites for structural elucidation of compounds for which standards are unavailable.

Targeted and non-targeted data analysis of large MS datasets

Because MS-based approaches generate large amounts of data, analysis and interpretation challenging. Analysis methods can be divided to two categories: targeted and non-targeted. The category is determined based on whether the identity and properties of metabolites to be analyzed is already known or not. Targeted data analysis aims to assess relatively small numbers of 'known' metabolites, whereas non-targeted method investigates datasets in an open-ended manner without previous knowledge or prediction of metabolite class or specific compound structures. Therefore, non-targeted data analysis is beneficial when studying poorly or uncharacterized metabolites.

Advances in bioinformatics facilitate the analysis of large MS datasets. For example, many software packages have been developed for noise reduction, analytical signal detection and normalization (Aprea *et al.* 2011, Tolstikov *et al.* 2003). In addition, computer-based statistical tests help to explore enormous datasets reliably and extract useful information. Multivariate analyses are often used to examine and interpret large datasets, especially for non-targeted data analysis (Kim *et al.* 2011, Sumner *et al.* 2003). For example, the application of the unsupervised principal component analysis (PCA) and hierarchical clustering analysis (HCA) methods assists data interpretation by reducing the dimensionality of datasets and efficiently visualizing them (Schilmiller *et al.* 2010a, Sumner *et al.* 2003). Supervised method such as partial least squares discriminant analysis (PLS-DA) is also useful to identify characteristic compounds that separate different sample groups, including metabolites of low abundance (Kim *et al.* 2011). The application of multivariate tests for non-targeted data analysis is discussed in Chapter 2.

Solanum trichomes as a model system to study plant specialized metabolites

Because the biosynthesis of specific types of plant specialized metabolites is often confined to taxa and or tissue types, selecting a proper model system is critical for their study (Pichersky and Gang 2000, Ziegler and Facchini 2008). Trichomes are specialized epidermal structures populating the surfaces of plant tissues. In particular, glandular trichomes or secreting glandular trichomes (SGTs) are 'chemical factories' where a number of specialized metabolites are produced and thus are useful as model systems to elucidate their biochemistry (Schilmiller *et al.* 2008, Wagner 1991). The genus Solanum includes commercially valuable crop species such as the cultivated tomato and their SGTs accumulate or exude high-levels of various specialized metabolites (Besser *et al.* 2009, Kandra *et al.* 1990, Slocombe *et al.* 2008).

In addition, morphological and chemical diversity in Solanum trichomes makes them especially appealing models for specialized metabolism studies (McDowell *et al.* 2011, Schilmiller *et al.* 2008). Seven different types of trichomes are documented in the cultivated and wild tomato species, four of which are known or proposed to be glandular (types I, IV, VI and VII in Figure 1.9). The types and density of trichomes vary across Solanum species and also according to their location on the plant (Luckwill 1943, McDowell *et al.* 2011). In addition to morphological diversity, SGTs have characteristic compositions of chemicals they produce, accumulate and/or secrete. Some specific types of SGTs in Solanum species have shown to accumulate certain classes of chemicals, for example, acylsugars are secreted by type 4 of *S. pennellii* (Fobes *et al.* 1985, McDowell *et al.* 2011) and methylketones are produced in type 6 of *S. habrochaites* (Fridman *et al.*

2005). The accumulation of gland-specific metabolites suggests that distinct biosynthetic pathways are present in particular types of SGTs. The great diversity in morphology and chemistry of trichomes in different Solanum species provides a good opportunity to identify trichome-synthesized metabolites and the major biosynthetic pathways operating in each type of trichomes.

Another advantage of this system compared with tissues and organs that produce other specialized metabolites is that purification of SGTs is relatively straightforward. Preparation of samples enriched with specific glands enables detailed chemical profiling,

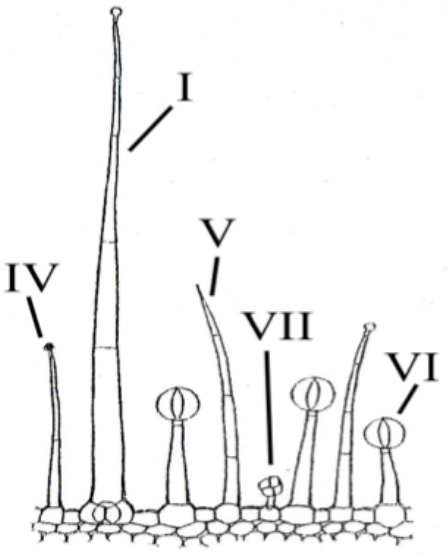


Figure 1.9. Seven types of Solanum trichomes.

Types I, IV, and VI are documented to be glandular trichomes, type VII was proposed to be glandular and the rest are non-glandular trichomes. Trichome types are drawn with modification of Luckwill 1943.

EST library construction and proteomics for individual types of SGTs (Schilmiller *et al.* 2008, Schilmiller *et al.* 2010b, Schmidt *et al.* 2011). The enriched sample preparation is beneficial for discovering novel genes and metabolites, especially when the accumulation of target transcripts and metabolites is at low levels (Schmidt *et al.* 2011).

The availability of various genetic and genomic resources also facilitates study of Solanum trichome metabolism. For example, an EMS (ethyl methane sulfonate) tomato mutant population is available (Menda *et al.* 2004) for forward genetics. Chromosomal substitution lines such as the *S. lycopersicum* M82 X *S. pennellii* introgression lines (ILs) (Eshed and Zamir 1994, Eshed and Zamir 1995) have proven to be highly useful resources for studies of trichome metabolism (Schilmiller *et al.* 2010a, Schilmiller *et al.* 2009). Details about the use of the tomato EMS mutant population and M82 X *S. pennellii* ILs are discussed in Chapters 2 and 3. In addition, accessions of wild species from different areas of the Andes Mountains and other areas are also available (http://tgrc.ucdavis.edu/). The use of natural variation in studies of trichome metabolites is discussed in Chapter 4.

Finally, Solanum trichomes are an increasingly powerful model system due to recent advances in genomic information and biological resources for the cultivated tomato and its wild relatives; these include whole genome sequence, ESTs and BAC clones and sequence (http://solgenomics.net/). The availability of genomic tools is beneficial in that they efficiently assist gene discovery, functional elucidation of genes and understanding evolutionary relationships of the genes (Buell and Last 2010, DellaPenna and Last 2008,

Falara *et al.* 2011). The application of tomato genomic resources will be discussed in Chapter 3.

As a part of this project aimed at elucidating biosynthetic pathways of trichome metabolites in the cultivated tomato and its wild relatives, three individual studies were performed for this dissertation work: (i) identification of tomato EMS mutants altered in trichome non-volatile metabolites and characterization of the glycosylated flavonoid phenotype in the JP314 EMS mutant (ii) identification and analysis of *S. lycopersicum* genetic variants such as the JP117 EMS mutant and *S. pennellii* IL6-3/4 for trichome methylated myricetin biochemistry, and (iii) chemical analysis of diverse trichome acylsugar profiles in 80 accessions from the wild tomato *S. habrochaites* species.

References

References

Aharoni, A. and Galili, G. (2011) Metabolic engineering of the plant primary-secondary metabolism interface. *Curr. Opin. Biotechnol.*, **22**, 239-244.

Aprea, E., Gika, H., Carlin, S., Theodoridis, G., Vrhovsek, U. and Mattivi, F. (2011) Metabolite profiling on apple volatile content based on solid phase microextraction and gas-chromatography time of flight mass spectrometry. *J. Chromatogr. A.*, **1218**, 4517-4524.

Baldwin, I.T., Halitschke, R., Paschold, A., von Dahl, C.C. and Preston, C.A. (2006) Volatile signaling in plant-plant interactions: "talking trees" in the genomics era. *Science*, **311**, 812-815.

Besser, K., Harper, A., Welsby, N., Schauvinhold, I., Slocombe, S., Li, Y., Dixon, R.A. and Broun, P. (2009) Divergent regulation of terpenoid metabolism in the trichomes of wild and cultivated tomato species. *Plant Physiol.*, **149**, 499-514.

Boudet, A.M. (2007) Evolution and current status of research in phenolic compounds. *Phytochemistry*, **68**, 2722-2735.

Braca, A., Bilia, A., Mendez, J. and Morelli, I. (2001) Myricetin glycosides from *Licania densiflora*. *Fitoterapia*, **72**, 182-185.

Brunet, G. and Ibrahim, R.K. (1986) O-methylation of flavonoids by cell-free extracts of calamondin orange. *Phytochemistry*, **19**, 741-746.

Buchana, B., Gruissem, W. and Jones, R. (2002) *Biochemistry & Molecular Biology of Plants*: Wiley.

Buell, C.R. and Last, R.L. (2010) Twenty-first century plant biology: impacts of the Arabidopsis genome on plant biology and agriculture. *Plant Physiol.*, **154**, 497-500.

Buer, C.S., Imin, N. and Djordjevic, M.A. (2010) Flavonoids: new roles for old molecules. *J. Integr. Plant Biol.*, **52**, 98-111.

Collins, F.W., De Luca, V., Ibrahim, R.K., Voirin, B. and Jay, M. (1981) Polymethylated flavonol synthesis in *Chrysosplenium americanum*: identification and enzymatic synthesis. *Z. Naturforsch*, **36c**, 730-736.

De Luca, V. and Ibrahim, R.K. (1982) Characterization of three distinct O-methyltransferases from *Chrysosplenium americanum*. *Phytochemistry*, **21**, 1537-1540.

De Luca, V. and Ibrahim, R.K. (1985) Enzymatic synthesis of polymethylated flavonols in *Chrysosplenium americanum.* I. Partial purification and some properties of S-adenosyl-L-methionine:flavonol 3-, 6-, 7- and 4'-O-methyltransferases. *Arch. Biochem. Biophys.*, **238**, 595-605.

DellaPenna, D. and Last, R.L. (2008) Genome-enabled approaches shed new light on plant metabolism. *Science*, **320**, 479-481.

Dixon, R.A. and Steele, C.L. (1999) Flavonoids and isoflavonoids - a gold mine for metabolic engineering. *Trends Plant Sci.*, **4**, 394-400.

Djordjevic, M.A., Redmond, J.W., Batley, M. and Rolfe, B.G. (1987) Clovers secrete specific phenolic compounds which either stimulate or repress nod gene expression in Rhizobium trifolii. *EMBO J.*, **6**, 1173-1179.

Eshed, Y. and Zamir, D. (1994) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: A tool for fine mapping of genes. *Euphytica*, **79**, 175-179.

Eshed, Y. and Zamir, D. (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, **141**, 1147-1162.

Facchini, P.J. and St-Pierre, B. (2005) Synthesis and trafficking of alkaloid biosynthetic enzymes. *Curr. Opin. Plant Biol.*, **8**, 657-666.

Falara, V., Akhtar, T.A., Nguyen, T.T., Spyropoulou, E.A., Bleeker, P.M., Schauvinhold, I., Matsuba, Y., Bonini, M.E., Schilmiller, A.L., Last, R.L., Schuurink, R.C. and Pichersky, E. (2011) The tomato terpene synthase gene family. *Plant Physiol.*, **157**, 770-789.

Fiehn, O. (2002) Metabolomics--the link between genotypes and phenotypes. *Plant Mol. Biol.*, **48**, 155-171.

Fobes, J.F., Mudd, J.B. and Marsden, M.P.F. (1985) Epicuticular lipid accumulation on the leaves of *Lycopersicon pennellii* (Corr.) D'Arcy and *Lycopersicon esculentum* Mill. *Plant Physiol.*, **77**, 567-570.

Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grun, S., Winklmair, A., Eisenreich, W., Bacher, A., Meeley, R.B., Briggs, S.P., Simcox, K. and Gierl, A. (1997) Analysis of a chemical plant defense mechanism in grasses. *Science*, **277**, 696-699.

Fridman, E., Wang, J., Iijima, Y., Froehlich, J.E., Gang, D.R., Ohlrogge, J. and Pichersky, E. (2005) Metabolic, genomic, and biochemical analyses of glandular

trichomes from the wild tomato species *Lycopersicon hirsutum* identify a key enzyme in the biosynthesis of methylketones. *Plant Cell*, **17**, 1252-1267.

Fujioka, S. and Sakurai, A. (1997) Conversion of lysine to L-pipecolic acid induces flowering in *Lemna paucicostata* 151. *Plant Cell Physiol.*, **38**, 1278-1280.

Ghangas, G.S. and Steffens, J.C. (1993) UDPglucose: fatty acid transglucosylation and transacylation in triacylglucose biosynthesis. *Proc. Nat. Acad. Sci. USA*, **90**, 9911-9915.

Goffreda, J.C., Ave, D.A., Mutschler, M.A., Steffens, J.C. and Tingey, W.M. (1989) Aphid deterrence by glucose esters in the glandular trichome exudate of the wild tomato, *L. pennellii*. *J. Chem. Ecol.*, **15**, 2135-2147.

Gu, L., Jones, A.D. and Last, R.L. (2007) LC-MS/MS assay for protein amino acids and metabolically related compounds for large-scale screening of metabolic phenotypes. *Anal Chem*, **79**, 8067-8075.

Hawthorne, D.J., Shapiro, J.A., Tingey, W.M. and Mutschler, M.A. (1992) Trichome-borne and artificially applied acylsugars of wild tomato deter feeding and oviposition of the leafminer *Liriomyza trifolii*. *Entomologia Experimentalis et Applicata*, **65**, 65-73.

Ibrahim, R. (2005) A forty-year journey in plant research: original contributions to flavonoid biochemistry. *Can. J. Bot.*, **83**, 433-450.

Jay, M., De Luca, V. and Ibrahim, R.K. (1982) *Meta*-methylation of rings A (8-) and B (3'-) is catalyzed by two distinct O-methyltransferases in *Lotus corniculatus*. *Z. Naturforsch*, **38**, 413-417.

Jones, P., Messner, B., Nakajima, J., Schaffner, A.R. and Saito, K. (2003) UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana*. *J. Biol. Chem.*, **278**, 43910-43918.

Jorgensen, K., Rasmussen, A.V., Morant, M., Nielsen, A.H., Bjarnholt, N., Zagrobelny, M., Bak, S. and Moller, B.L. (2005) Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.*, **8**, 280-291.

Kandra, G., Severson, R. and Wagner, G.J. (1990) Modified branched-chain amino acid pathways give rise to acyl acids of sucrose esters exuded from tobacco leaf trichomes. *Eur. J. Biochem.*, **188**, 385-391.

Kang, J.H., Liu, G., Shi, F., Jones, A.D., Beaudry, R.M. and Howe, G.A. (2010a) The tomato *odorless-2* mutant is defective in trichome-based production of diverse

specialized metabolites and broad-spectrum resistance to insect herbivores. *Plant Physiol.*, **154**, 262-272.

Kang, J.H., Shi, F., Jones, A.D., Marks, M.D. and Howe, G.A. (2010b) Distortion of trichome morphology by the *hairless* mutation of tomato affects leaf surface chemistry. *J. Exp. Bot.*, **61**, 1053-1064.

Kim, H.K., Choi, Y.H. and Verpoorte, R. (2011) NMR-based plant metabolomics: where do we stand, where do we go? *Trends Biotechnol.*, **29**, 267-275.

Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B. and Bergmuller, E. (2005) GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics*, **21**, 1635-1638.

Krishnan, P., Kruger, N.J. and Ratcliffe, R.G. (2005) Metabolite fingerprinting and profiling in plants using NMR. *J. Exp. Bot.*, **56**, 255-265.

Kumari, G., Rao, L. and Rao, N. (1984) Myricetin methyl ethers from Solanum pubescens. *Phytochemistry*, **23**, 2701-2702.

Last, R.L., Jones, A.D. and Shachar-Hill, Y. (2007) Towards the plant metabolome and beyond. *Nat. Rev. Mol. Cell. Biol.*, **8**, 167-174.

Lei, Z., Huhman, D.V. and Sumner, L.W. (2011) Mass spectrometry strategies in metabolomics. *J. Biol. Chem.*, **286**, 25435-25442.

Li, J., Ou-Lee, T.M., Raba, R., Amundson, R.G. and Last, R.L. (1993) Arabidopsis flavonoid mutants are hypersensitive to UV-B irradiation. *Plant Cell*, **5**, 171-179.

Li, Z., Ngojeh, G., DeWitt, P., Zheng, Z., Chen, M., Lainhart, B., Li, V. and Felpo, P. (2008) Synthesis of a library of glycosylated flavonols. *Tetrahedron Letters*, **49**, 7243-7245.

Liu, B., Wang, H., Du, Z., Li, G. and Ye, H. (2011) Metabolic engineering of artemisinin biosynthesis in *Artemisia annua L. Plant. Cell. Rep.*, **30**, 689-694.

Luckwill, L. (1943) *The genus Lycopersicon: A historical, biological, and taxonomic survey of the wild and cultivated tomatoes* Aberdeen, Scotland: Aberdeen University Press.

Matsuda, F., Yonekura-Sakakibara, K., Niida, R., Kuromori, T., Shinozaki, K. and Saito, K. (2009) MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant J.*, **57**, 555-577.

McDowell, E.T., Kapteyn, J., Schmidt, A., Li, C., Kang, J.H., Descour, A., Shi, F., Larson, M., Schilmiller, A., An, L., Jones, A.D., Pichersky, E., Soderlund,

C.A. and Gang, D.R. (2011) Comparative functional genomic analysis of Solanum glandular trichome types. *Plant Physiol.*, **155**, 524-539.

Menda, N., Semel, Y., Peled, D., Eshed, Y. and Zamir, D. (2004) *In silico* screening of a saturated mutation library of tomato. *Plant J.*, **38**, 861-872.

Min, B., Lee, S., Kim, J., Lee, J., Kim, T., Kim, D., Kim, Y., Joung, H., Lee, H. and Nakamura, N. (2003) Anti-complement activity of constituents from the stem-bark of *Juglans mandshurica*. *Biol. Pharm. Bull.*, **26**.1042-1044.

Modolo, L.V., Li, L., Pan, H., Blount, J.W., Dixon, R.A. and Wang, X. (2009) Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids. *J Mol Biol*, **392**, 1292-1302.

Mol, J., Grotewold, E. and Koes, R. (1998) How genes paint flowers and seeds. *Trends Plant Sci.*, **3**, 212-217.

Mutschler, M.A., Doerge, R.W., Liu, S.C., Kuai, J.P., Liedl, B.E. and Shapiro, J.A. (1996) QTL analysis of pest resistance in the wild tomato *Lycopersicon pennellii*: QTLs controlling acylsugar level and composition *Theor. Appl. Genet.* , **92**, 709-718.

Nowack, R. and Schmitt, W. (2008) Cranberry juice for prophylaxis of urinary tract infections – Conclusions from clinical experience and research. *Phytomedicine*, **15**, 653-667.

Pichersky, E. and Gang, D.R. (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.*, **5**, 439-445.

Pichersky, E. and Gershenzon, J. (2002) The formation and function of plant volatiles: perfumes for pollinator attraction and defense. *Curr. Opin. Plant Biol.*, **5**, 237-243.

Reynertson, K., Yang, H., Jiang, B., Basile, M. and Kennelly, E. (2008) Quantitative analysis of antiradical phenolic constituents from fourteen edible *Myrtaceae* fruits. *Food Chem.* , **109**, 883-890.

Roberts, S.C. (2007) Production and engineering of terpenoids in plant cell culture. *Nat. Chem. Biol.*, **3**, 387-395.

Roda, A.L., Oldham, N.J., Svatos, A. and Baldwin, I.T. (2003) Allometric analysis of the induced flavonols on the leaf surface of wild tobacco (Nicotiana attenuata). *Phytochemistry*, **62**, 527-536.

Sallaud, C., Rontein, D., Onillon, S., Jabes, F., Duffe, P., Giacalone, C., Thoraval, S., Escoffier, C., Herbette, G., Leonhardt, N., Causse, M. and Tissier, A. (2009)

A novel pathway for sesquiterpene biosynthesis from Z,Z-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. *Plant Cell*, **21**, 301-317.

Sawada, Y., Akiyama, K., Sakata, A., Kuwahara, A., Otsuki, H., Sakurai, T., Saito, K. and Hirai, M.Y. (2009) Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. *Plant Cell Physiol.*, **50**, 37-47.

Schilmiller, A., Shi, F., Kim, J., Charbonneau, A.L., Holmes, D., Daniel Jones, A. and Last, R.L. (2010a) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. *Plant J.*, **62**, 391-403.

Schilmiller, A.L., Last, R.L. and Pichersky, E. (2008) Harnessing plant trichome biochemistry for the production of useful compounds. *Plant J.*, **54**, 702-711.

Schilmiller, A.L., Miner, D.P., Larson, M., McDowell, E., Gang, D.R., Wilkerson, C. and Last, R.L. (2010b) Studies of a biochemical factory: tomato trichome deep expressed sequence tag sequencing and proteomics. *Plant Physiol.*, **153**, 1212-1223.

Schilmiller, A.L., Schauvinhold, I., Larson, M., Xu, R., Charbonneau, A.L., Schmidt, A., Wilkerson, C., Last, R.L. and Pichersky, E. (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc. Natl. Acad. Sci. U S A*, **106**, 10865-10870.

Schmidt, A., Li, C., Shi, F., Jones, A.D. and Pichersky, E. (2011) Polymethylated myricetin in trichomes of the wild tomato species Solanum habrochaites and characterization of trichome-specific 3'/5'- and 7/4'-myricetin O-methyltransferases. *Plant Physiol.*, **155**, 1999-2009.

Schwab, W. (2003) Metabolome diversity: too few genes, too many metabolites? *Phytochemistry*, **62**, 837-849.

Shapiro, J.A., Steffens, J.C. and Mutschler, M.A. (1994) Acylsugars of the wild tomato *Lycopersicon pennellii* in relation to geographic distribution of the species. *Biochem. Syst. Ecol.*, **22**, 545-561.

Shin, M.H., Lee do, Y., Liu, K.H., Fiehn, O. and Kim, K.H. (2010) Evaluation of sampling and extraction methodologies for the global metabolic profiling of *Saccharophagus degradans*. *Anal. Chem.*, **82**, 6660-6666.

Slocombe, S.P., Schauvinhold, I., McQuinn, R.P., Besser, K., Welsby, N.A., Harper, A., Aziz, N., Li, Y., Larson, T.R., Giovannoni, J., Dixon, R.A. and Broun, P. (2008) Transcriptomic and reverse genetic analyses of branched-chain fatty acid

and acyl sugar production in *Solanum pennellii* and *Nicotiana benthamiana*. *Plant Physiol.*, **148**, 1830-1846.

Stevens, J., Hart, H. and Wollenweber, E. (1995) The systematic and evolutionary significance of exudate flavonoids in *Aeonium Phytochemistry* **39**, 805-813.

Sumner, L.W., Mendes, P. and Dixon, R.A. (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*, **62**, 817-836.

Sweetlove, L.J., Last, R.L. and Fernie, A.R. (2003) Predictive metabolic engineering: a goal for systems biology. *Plant Physiol.*, **132**, 420-425.

Taylor, L. and Jorgensen, R. (1992) Conditional male-fertility in chalcone synthase-deficient Petunia. *J. Hered.*, **83**, 11-17.

Taylor, L.P. and Grotewold, E. (2005) Flavonoids as developmental regulators. *Curr. Opin. Plant Biol.*, **8**, 317-323.

Tholl, D. (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.*, **9**, 297-304.

Tholl, D. and Lee, S. (2011) *Terpene Specialized Metabolism in Arabidopsis thaliana*.

Tolstikov, V.V., Lommen, A., Nakanishi, K., Tanaka, N. and Fiehn, O. (2003) Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. *Anal. Chem.*, **75**, 6737-6740.

Wagner, G.J. (1991) Secreting glandular trichomes: more than just hairs. *Plant Physiol.*, **96**, 675-679.

Wasson, A.P., Pellerone, F.I. and Mathesius, U. (2006) Silencing the flavonoid pathway in *Medicago truncatula* inhibits root nodule formation and prevents auxin transport regulation by rhizobia. *Plant Cell*, **18**, 1617-1629.

Weinhold, A. and Baldwin, I.T. (2011) From the Cover: Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation. *Proc Natl Acad Sci U S A*, **108**, 7855-7859.

Williams, C. and Grayer, R. (2004) Anthocyanins and other flavonoids. *Nat. Prod. Rep.*, **21**, 539-573.

Wittstock, U. and Halkier, B.A. (2002) Glucosinolate research in the Arabidopsis era. *Trends Plant Sci.*, **7**, 263-270.

Wu, J.H., Huang, C.Y., Tung, Y.T. and Chang, S.T. (2008) Online RP-HPLC-DPPH screening method for detection of radical-scavenging phytochemicals from flowers of Acacia confusa. *J Agric Food Chem*, **56**, 328-332.

Zhang, J., Subramanian, S., Stacey, G. and Yu, O. (2009) Flavones and flavonols play distinct critical roles during nodulation of *Medicago truncatula* by *Sinorhizobium meliloti*. *Plant J.*, **57**, 171-183.

Ziegler, J. and Facchini, P.J. (2008) Alkaloid biosynthesis: metabolism and trafficking. *Annu. Rev. Plant Biol.*, **59**, 735-769.

Chapter 2. Chemical screening of tomato EMS mutants to identify altered trichome non-volatile metabolites

Abstract

EMS M2 family mutants of the cultivated M82 tomato were analyzed to identify altered trichome non-volatile metabolites using high-throughput LC-ToF-MS and targeted and non-targeted data analyses. Self-cross M3 progeny plants from M2 putative mutants were analyzed to test for inheritance of the phenotypes. Of 1,387 M2 mutant plants from 58 M2 families, 24 mutants were confirmed for altered trichome non-volatile chemistry. The mutant JP314 was identified as having decreased accumulation of a flavonol tetraglycoside with m/z 949.3 from targeted analysis. Non-targeted analysis validated the flavonol tetraglycoside mutant phenotype and revealed unannotated analytical signals highly correlated with the m/z 949.3. To further characterize these metabolites, tomato leaf extracts were chemically profiled in more detail. The results suggested that the flavonol tetraglycoside with m/z 949.3 is a myricetin derivative and the other unannotated analytical signals are adduct ions of this flavonol tetraglycoside. JP314 mutant also have decreased anthocyanins in leaf tissue. The phenotypes of reduced myricetin tetraglycoside and total anthocyanins lead to the hypothesis that flavonoid accumulation is impaired in this mutant. The JP314 mutant case study demonstrates the value of 'brute force' M2 mutant metabolite screening in uncovering mutations that affect specialized metabolism, improving metabolite annotation and generating testable hypothesis.

Introduction

Trichomes are specialized epidermal 'hairs' covering the surface of plant tissues (Schilmiller *et al.* 2008). Trichomes are categorized into non-glandular or glandular types based on the presence of one or more glandular cells on the top of a stalk (Luckwill 1943, Wagner 1991). In particular, secretory and glandular trichomes (SGTs) serve as 'chemical factories' where plant specialized metabolites are produced and accumulated. SGTs are often used as models to study plant specialized (also known as 'secondary') metabolism (Besser *et al.* 2009, McConkey *et al.* 2000, Schilmiller *et al.* 2008, Turner *et al.* 2000). For example, SGTs from plants in the genus Solanum including the cultivated tomato *Solanum lycopersicum* and its wild relatives have been studied to understand volatile metabolites such as terpenoids (Falara *et al.* 2011, Sallaud *et al.* 2009, Schilmiller *et al.* 2010b, Schilmiller *et al.* 2009) and methylketones (Fridman *et al.* 2005, Yu 2010). Previous studies also reported that Solanum SGTs accumulate a variety of specialized non-volatile metabolites such as flavonoids, alkaloids and acylsugars (Fobes *et al.* 1985, McDowell *et al.* 2011, Schilmiller *et al.* 2010a).

Forward and reverse genetics are efficient approaches to understand gene functions. Reverse genetics looks for phenotypes when the function of a specific gene or set of genes is deliberately altered. In contrast, forward genetics is used to screen large collections of genetic variants for altered phenotypes and then identify the allele responsible for a trait that differs from the wild type or reference organism. Mutagenesis with ethylmethane sulphonate (EMS) is commonly used to create genetic diversity for forward genetics in plants, microbes and animals (Jander *et al.* 2002, Peters *et al.* 2003).

EMS mutagenesis causes high densities of G:C to A:T point mutations at random chromosomal locations (Jander *et al.* 2003), allowing efficient sampling of changes in many genes in the genome by screening a relatively small number of mutants (Buttner *et al.* 2010, Marti *et al.* 2010). This makes it an excellent mutagen for generating genetic variants for mutant screens. Performing forward genetic screens on EMS mutant populations allows the discovery of genes that influence a phenotype without prior knowledge of gene function. This is beneficial for the investigation of plant specialized metabolism, especially when studying unknown or poorly understood biosynthetic pathways.

Careful choice of phenotypic assay is also important for forward genetic screens. These can range from a simple test looking for differences in plant morphology to technically complex chemical profiling to identify altered metabolites. For the identification of mutants altered in specialized metabolites, selecting the optimal analytical approach is an important prerequisite. Special challenges arise because plants accumulate complex mixtures of specialized metabolites varying in concentration range and chemical properties. To maximize the value of the screen, plant metabolite profiling requires analytical approaches with high sensitivity, accuracy and reproducibility. There are several reasons why mass spectrometry (MS) is a powerful tool to profile plant specialized metabolites. First, MS can detect metabolites down to the nanomolar concentration range (Last *et al.* 2007). The high sensitivity and selectivity of MS allows simplifying extraction methods (for instance, eliminating the need for two phase extraction), thus shortening sample preparation time while still allowing detection of low

abundance specialized metabolites. MS can be more powerful when combined with chromatographic separation such as liquid chromatography (LC) or gas chromatography (GC) (Last *et al.* 2007). The efficacy of MS combined with GC was revealed from chemical profiling of volatile terpenes in different tissue types of M82 cultivated tomato (Falara *et al.* 2011, Schilmiller *et al.* 2010b). LC-MS was used to investigate non-volatile metabolite profiles in *S. pennellii* introgression lines (Schilmiller *et al.* 2010a) and *S. habrochaites* SGTs (Schmidt *et al.* 2011).

Despite advances in detection and quantification of specialized metabolites, metabolite annotation still remains limited (Sumner *et al.* 2003, Wink 1999). One reason why structural characterization remains difficult is that there is enormous structural diversity and complexity of metabolites in plants. In addition, structural annotation is even harder due to low abundance of metabolites, lack of authentic standards and the formation of various types of adducts for a single metabolites in MS (Last *et al.* 2007). When metabolite annotation is not known, non-targeted data analysis (i.e. the open-ended data analysis) is useful. Unlike targeted analysis, non-targeted data analysis does not require previous knowledge of target metabolites. For instance, non-targeted analysis of LC-MS datasets revealed spatial distribution of metabolites in strawberry flowers without prior knowledge of metabolite composition (Hanhineva *et al.* 2008). Non-targeted data analysis also helped identify volatile metabolites in ripe tomato fruits (Tikunov *et al.* 2005). Non-targeted data analysis is advantageous to understand metabolite profiles, even including incompletely characterized metabolites (See 'Chapter 1. Introduction' for more details about targeted and non-targeted data analyses).

In this study, tomato EMS M2 family mutants were screened by high-throughput LC-Time of Flight (ToF)-MS analysis to identify altered non-volatile trichome chemistry. The workflow of mutant screening is described with the JP314 mutant as a case study. Another class of mutants from this screen is described in Chapter 3.

Results

The structure of the EMS M2 family mutant population and its use for mutant screening

A cultivated tomato (*Solanum lycopersicum*, cultivar M82) M2 mutant population was previously generated by EMS mutagenesis (Menda *et al.* 2004)Sol Genomics Network, http://solgenomics.net/) and provided by Dr. Dani Zamir, Hebrew University, Rehovot, Israel. This population was chosen to screen for trichome metabolites because it has a high mutation rate, and was previously shown to have mutants of many morphological phenotypes (Menda *et al.* 2004)The structure of this mutant population, with M2 families of seed from individual M1 plants (Figure 2.1), also helps identify siblings and independent alleles, supporting the discovery of heritable mutations (Krieger *et al.* 2010). Chemical extraction was performed by a single leaf dip method with a short incubation; this method was chosen to maximize trichome metabolite extraction, while minimizing the extraction of compounds from epidermal pavement cells (Schilmiller *et al.* 2010a). A negative mode LC-ToF-MS assay was employed to profile non-volatile metabolites from plant extracts enriched with trichome metabolites (Figure 2.2).

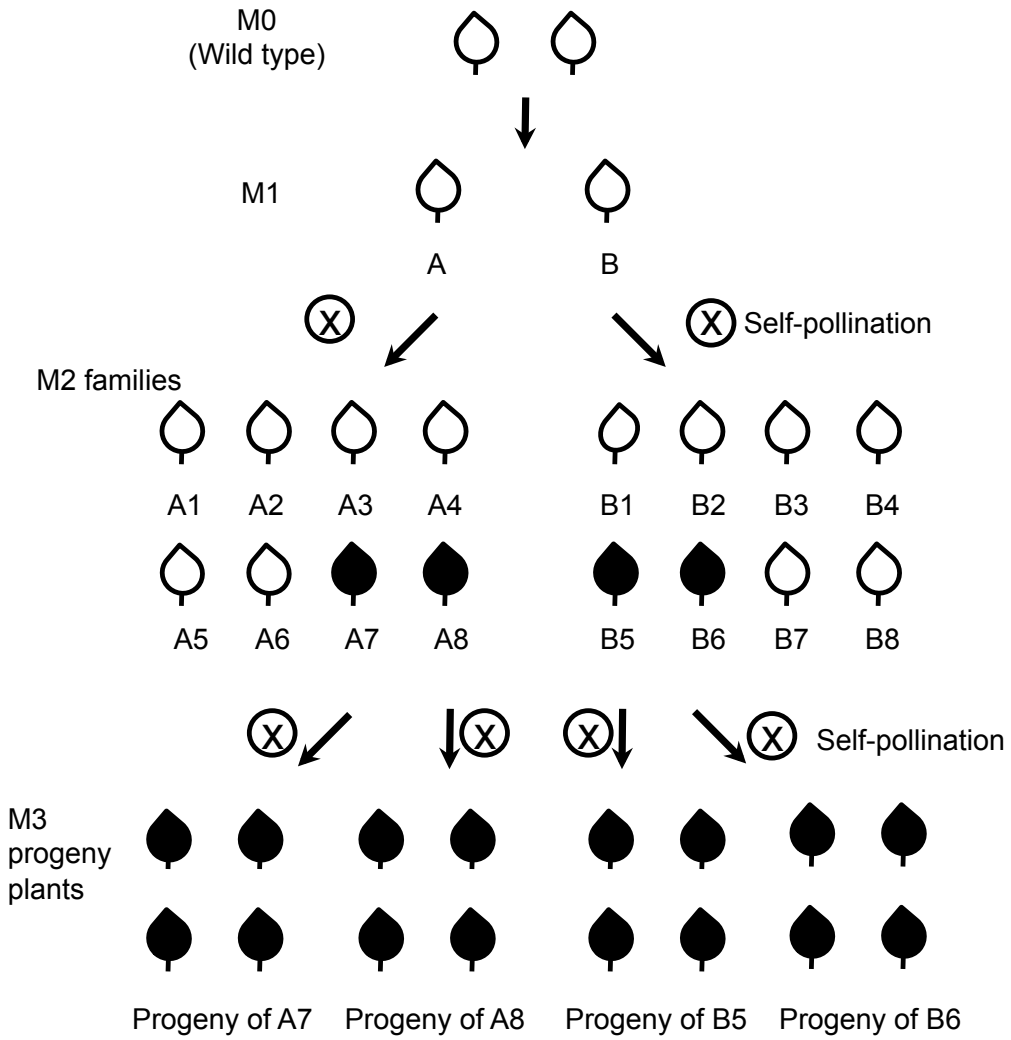


Figure 2.1. Structure of the mutant population used in this study.

Tomato EMS M2 family mutant seeds were previously generated (Menda *et al*. 2004) and kindly supplied by Dani Zamir, Hebrew University, Rehovot, Israel. The M2 plants were screened to identify altered trichome non-volatile metabolites. Once an M2 putative mutant was identified, the self-pollinated M3 progeny were analyzed to test for inheritance of the chemical phenotype. M2 plants derived from a single M1 plant are siblings and any common phenotype is presumed to result from the same original mutation.

M2 mutant plants were considered putative mutants when their chemical phenotypes varied from wild type controls by two serial statistical cutoffs. These putative mutants were self-pollinated to obtain M3 progeny. Because individual M2 mutant plants were assayed, putative mutants were confirmed by testing chemical phenotypes in M3 progeny plants. These progeny plants were analyzed with at least four biological replicates to test whether the M2 putative mutant phenotype was consistently observed in M3 progeny (Figures 2.1 and 2.2). I also took advantage of the mutant population structure in two ways (Menda *et al.* 2004). First, if two different M2 siblings from the same family showed identical phenotypes (for example, A7 and A8, or B5 and B6 in Figure 2.1), it is likely that they have shared mutations. Second, if similar phenotypes were found in M2 lines from two independent families (for example, A8 and B5, or A7 and B6 in Figure 2.1), it suggests a high chance that they are independent alleles for the same mutant genes.

The heritability of M2 putative mutant phenotypes was validated by chemical analysis of the self-pollinated M3 progeny plants. In addition to confirmation of the M2 phenotype, rescreening of four or more M3 plants increases the statistical power to detect other phenotypes (Figure 2.1).

Identification and characterization of trichome non-volatile metabolite mutants by high-throughput LC-ToF-MS with targeted and non-targeted data analysis

The chemical profiles of EMS M2 mutant plants were analyzed and compared with control M82 wild type plants and a set of control *S. pennellii* introgression lines (ILs)

previously found to have quantitative changes in trichome chemistry (Schilmiller *et al.* 2010a) (Figure 2.2, See 'Materials and Methods' for details about plant growth). The rapid LC-ToF-MS chemical analysis generated a list of thousands of 'analytical signals', which is defined as the combination of retention time in LC and mass-to-charge ratio (m/z) with high mass accuracy for corresponding ions detected in MS (Gu *et al.* 2010, Last *et al.* 2007, Schilmiller *et al.* 2010a). The analytical signals were evaluated by targeted and non-targeted data analysis (Figure 2.2).

Thirty-five analytical signals were tested for targeted data analysis (Table 2.1). Target metabolites were chosen based on previously characterized annotation (Schilmiller *et al.* 2010a, Table 2.1 and Figure 2.2). Unannotated analytical signals were also included when reproducibly detected (Schilmiller *et al.* 2010a). Two serial cutoffs were applied to identify M2 putative mutants in the targeted data analysis: 1. median +/- 5 median absolute deviation (MAD); 2. > 3-fold change compared to the median (Figure 2.2). When the alteration of trichome metabolites in M2 mutant plants met two serial cutoffs, they were considered putative mutants in M2 screening (Figure 2.2). Non-targeted data analysis such as principal component analysis (PCA) was performed to identify M2 putative mutants that targeted data analysis might have failed to reveal (Figure 2.2).

The same LC-ToF-MS assay was used to analyze M3 progeny of putative mutants. Targeted data analysis was performed to identify M3 progeny altered in the 35 analytical signals using two serial cutoffs: 1. > 3-fold change in the mean between mutant and M82

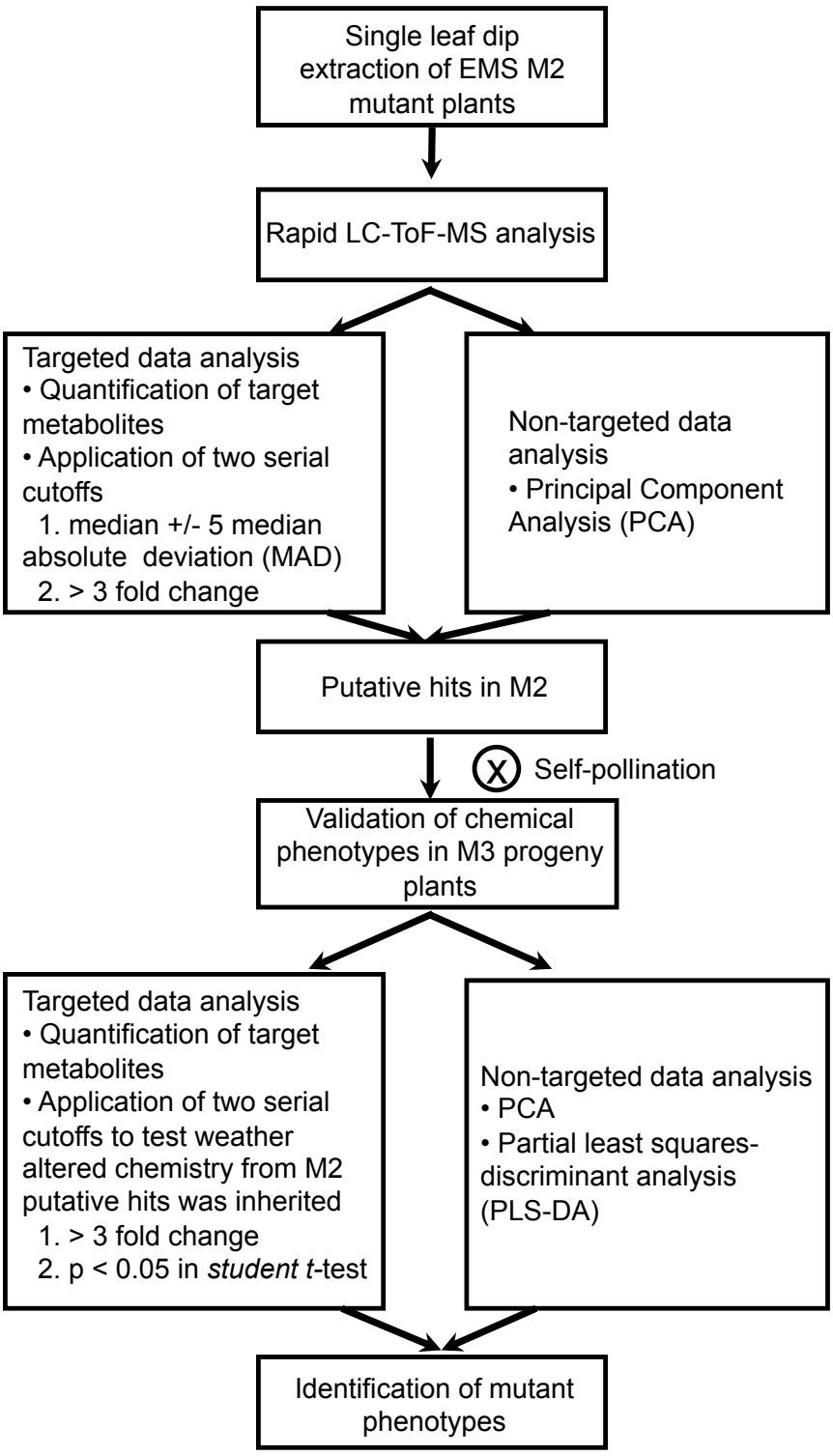


Figure 2.2. EMS mutant screening workflow.

Figure 2.2. Continued

LC-ToF-MS (negative ion mode) analysis of single leaf dip extracts was used to profile trichome non-volatile metabolites. Targeted and non-targeted data analyses were performed to investigate LC-MS datasets. Two serial cutoffs were used in targeted data analysis to identify M2 putative mutants and validate the phenotypes in M3. Multivariate analyses such as PCA and PLS-DA was performed for non-targeted data analysis.

wild type; 2. statistically significant alteration corresponding to $p < 0.05$ from *Student's t*-test (Figure 2.2). In addition to PCA, the partial least squares discriminant analysis (PLS-DA) supervised method was also employed for non-targeted data analysis in M3 progeny. PLS-DA allows identification of minor factors (i.e. analytical signals) contributing to the separation between mutant and wild-type plants and searching for the correlation among these analytical signals (Westerhuis *et al.* 2007).

Twenty-four mutants with altered trichome non-volatile metabolites

This mutant screening tested 1,387 M2 mutants from 58 families. Of 1,387 plants, 223 (16%) were identified as M2 putative mutants and 175 yielded M3 seed for rescreening. In total, 24 mutants had consistent phenotypes in M3 progeny (Table 2.2). The 24 identified mutants revealed altered chemistry in diverse classes of trichome non-volatile metabolites ranging from methylated flavonoids to acylsugars (Table 2.3). Notably, multiple mutants were discovered for five types of trichome chemistry, methylated myricetins, flavonol tetraglycoside with m/z 949.3, kaempferol 3-O-rutinoside, acylsugar S4:15 and an unannotated analytical signal with m/z 493.3. Based on the M2 mutant population structure, they are likely to be either siblings (from the same M2 family, for example, JP1913 and JP1938 for flavonol glycoside phenotype) or independently derived alleles (from different M2 families, for example, JP857 and JP2883 for acylsugar S4:15 (2,4,4,5) phenotype) (Table 2.3). Seven mutants from three M2 families were identified with methylated myricetin phenotypes; thus this collection contains at least three independent alleles for this phenotype (see Chapter 3 for detailed analysis of one such

Table 2.1. The list of 35 analytical signals tested for targeted data analysis.

Thirty-five analytical signals tested for targeted data analysis are listed. The metabolite annotation, LC retention time, m/z and ion type is from Schilmiller *et al.* 2010a. Unannotated or unknown analytical signals were also included when reproducibly detected from M82 plants. Acylsugar nomenclature includes type of sugar, number of acyl chains, total number of carbons in acyl chains and expected carbon number for each acyl chain; for example, S4:17 (2,5,5,5) indicates sucrose esterified with 4 acyl chains, one acyl chain with 2 carbons and three with 5 carbons respectively, for a total of 17 carbons. The abundance of metabolites was estimated based on peak areas for the corresponding analytical signals. Relative abundance was obtained by comparing the peak areas of the individual analytical signals with that of S4:17 (2,5,5,5), the most abundant acylsugar in M82 cultivated tomato (Schilmiller *et al,* 2010) and indicated as follows; < 1% (---), 1~3 (--), 3~5 (-), 5~10 (+), 10~20 (++), 20~35 (+++), 35~50 (++++) and 50~100 (+++++).

Table 2.1. Continued

Metabolite Annotation	Retention time (min)	m/z (Da)	Ion Type	Relative Abundance
Glucose	0.4	215.0	$[M+Cl]^-$	-
Sucrose	0.4	377.1	$[M+Cl]^-$	--
Malic acid	0.5	133.0	$[M-H]^-$	-
Quinic acid	0.5	191.0	$[M-H]^-$	+
Chlorogenic acid	1.7	353.1	$[M-H]^-$	-
Flavonol tetraglycoside	2.2	949.3	$[M-H]^-$	---
Quercetin 3-O-glucosylglucoside	2.3	625.2	$[M-H]^-$	---
Flavonol glycoside	2.4	771.2	$[M-H]^-$	---
Quercetin 3-O-rutinoside	2.5	609.2	$[M-H]^-$	+++
Kaempferol 3-O-rutinoside	2.6	593.2	$[M-H]^-$	-
Mono-O-methylmyricetin	2.7	331.1	$[M-H]^-$	---
Dehydrotomatine	2.8	1076.5	$[M+formate]^-$	+
Tomatine	2.9	1078.5	$[M+formate]^-$	-
Di-O-methylmyricetin	2.9	345.1	$[M-H]^-$	--
Tri-O-methylmyricetin	3.2	359.1	$[M-H]^-$	---
Acylsugar S4:15 (2,4,4,5)	3.3	653.3	$[M+formate]^-$	---
Acylsugar S3:15 (5,5,5)	3.3	639.3	$[M+formate]^-$	++
Acylsugar S4:16 (2,4,5,5)	3.3	667.3	$[M+formate]^-$	++
Acylsugar S4:17 (2,5,5,5)	3.4	681.3	$[M+formate]^-$	+++++
Acylsugar S4:18 (4,4,5,5)	3.5	695.3	$[M+formate]^-$	--
Acylsugar S4:20 (2,4,4,10 or 5,5,5,5)	3.5	723.4	$[M+formate]^-$	---
Acylsugar S3:20 (5,5,10 or 4,5,11)	3.7	709.4	$[M+formate]^-$	--
Acylsugar S4:22 (2,4,4,12 or 2,5,5,10)	3.8	751.4	$[M+formate]^-$	-
Acylsugar S3:21 (5,5,11 or 4,5,12)	3.8	723.4	$[M+formate]^-$	---
Acylsugar S4:23 (2,4,5,12 or 2,5,5,11)	3.8	765.4	$[M+formate]^-$	---
Acylsugar S3:22 (5,5,12)	3.8	737.4	$[M+formate]^-$	+++
Acylsugar S3:23 (5,6,12)	3.9	751.4	$[M+formate]^-$	---
Acylsugar S4:24 (2,5,5,12)	3.9	779.4	$[M+formate]^-$	++++

Table 2.1. Continued

Metabolite Annotation	Retention time (min)	m/z (Da)	Ion Type	Relative Abundance
Unknown	0.5	371.1		---
Unknown	1.4	369.1		---
Unknown	2.2	693.4		---
Unknown	2.2	963.3		---
Unknown	2.2	1009.3		---
Unknown	3.0	493.3		--
Unknown	3.1	485.3		--
n-Propyl-4-hydroxybenzoate (Internal standard)	3.0	179.1	$[M-H]^-$	+

Table 2.2. Summary of mutant screen results

1,387 M2 mutant plants from 58 families were screened by LC-ToF-MS for altered trichome metabolite levels (See Figures 2.1 and 2.2 for details). 223 of 1,387 plants were identified as M2 putative mutants. Self-pollination yielded viable progeny for 175 of the 223 putative mutants, and 24 of these lines revealed consistent phenotypes upon rescreening (See Table 2.3 for details of chemical phenotypes in the 24 identified mutants).

The stage of screening	**No. of plants**	**Description**
Screening M2 individual mutant plants	1387	From 58 M2 families
Identification of putative mutants in M2	223	
Validation of chemical phenotypes in M3 progeny	175	No seed (39), No germination (9) for M3 progeny
Identification of mutants	24	From 14 M2 families

Table 2.3. Chemical phenotypes of the 24 mutants identified from this screening.

Twenty-four mutants were identified for altered trichome non-volatile metabolites. SGN M2 family ID (http://zamir.sgn.cornell.edu/mutants/) is listed with mutant names assigned in this study.

Mutant chemical phenotypes	Mutant Name	SGN M2 Family ID
Increased mono-O-methylmyricetin and decreased di-O-methylmyricetin and tri-O-methylmyricetin	JP117, JP130, JP334	e0004b
	JP256, JP446	e0007b
	JP3045, JP3060	e0139b
Decreased flavonol tetraglycoside (m/z 949.3, RT 2.2) and unknown (m/z1009.3, RT 2.2)	JP314	e0004b
Increased flavonol glycoside (m/z 949.3, RT 2.2) and unknown (m/z1009.3, RT 2.2)	JP1913, JP1938	e0124b
Increased kaempferol 3-O-rutinoside	JP761	e0019b
	JP416	e0014b
	JP3427	e0142b
	JP3123	e0139b
Decreased kaempferol 3-O-rutinoside	JP1229	e0116b
Decreased tomatine	JP1885	e0117b
Increased total acylsugars	JP269	e0005b
Decreased acylsugars S4:20 (2,4,4,10 or 5,5,5,5) and S4:18 (4,4,5,5)	JP389	e0009b
Increased acylsugar S4:23 (2,4,5,12 or 2,5,5,11)	JP757	e0019b
Increased acylsugar S4:15 (2,4,4,5)	JP857	e0016b
	JP2883	e0131b
Increased unknowns (m/z 693.4, RT 2.2; m/z 371.1, RT 0.5)	JP316	e0005b
Increased unknown (m/z 493.3, RT 3.0)	JP691, JP612	e0118b

mutant). In summary, this screen identified 24 mutants for trichome non-volatile chemistry, providing valuable genetic variants for further experiments.

A Case Study: Identification of JP314 as a mutant altered in flavonoid glycosides

Identification of JP314 as having increased kaempferol 3-O-rutinoside

Targeted analysis of LC-ToF-MS assay data revealed higher accumulation of kaempferol 3-O-rutinoside in the M2 mutant JP314. This putative mutant was identified using two serial cutoffs, 5 MAD and > 3 fold-change compared to the median of cohort plants (Figure 2.2). In this screen, thirteen cohort M2 plants were grown in the same flat and screened together with JP314 and one plant for each M82 wild type and IL11-3 (See 'Materials and Methods' for details of plant growth). As previously reported IL11-3 revealed decreased total acylsugar accumulation (Schilmiller *et al.* 2010a), with no alteration in the remaining 22 analytical signals compared to M82 (Table 2.1), supporting the reproducibility of chemical assay and data analysis in this mutant screening (Figure 2.3).

Non-targeted data analysis with PCA did not identify any putative mutants in M2 screening (Figure 2.4), which might be because PCA is often insensitive to subtle differences, especially when the metabolites causing the phenotypes are low abundance compounds. PCA with all 13 cohort plants distinguished IL11-3 control plant from others (Figure 2.4ab). The strong chemical phenotypes in IL11-3 (Schilmiller *et al.* 2010a) led to the hypothesis that the presence of IL11-3 prevents PCA from distinguishing potential mutants with subtle phenotypes, especially for low abundance compounds. Therefore,

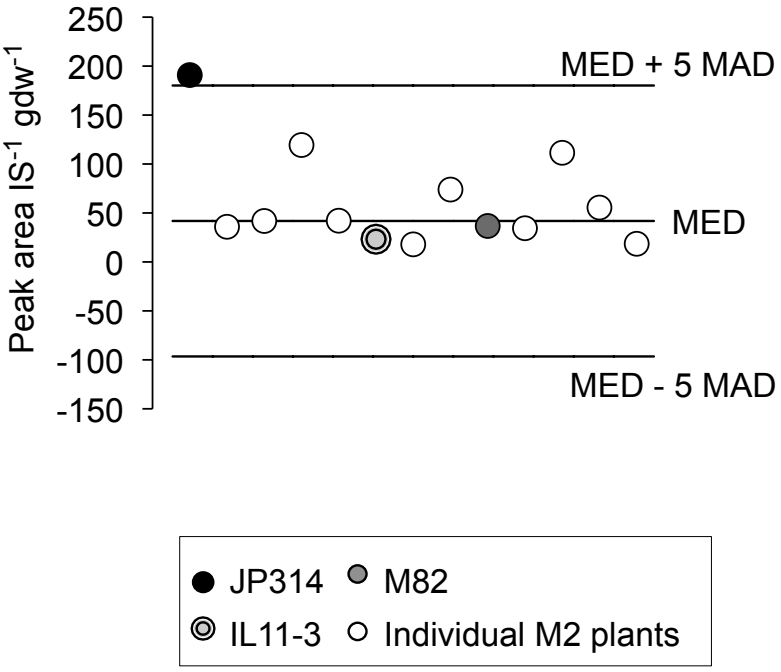


Figure 2.3. Targeted data analysis identified an M2 putative mutant JP314 as having increased kaempferol 3-O-rutinoside.

The altered accumulation of kaempferol 3-O-rutinoside in JP314 met the two serial cutoffs used for the targeted data analysis in M2 screening: median + 5 MAD and > 3 fold change compared to the median.

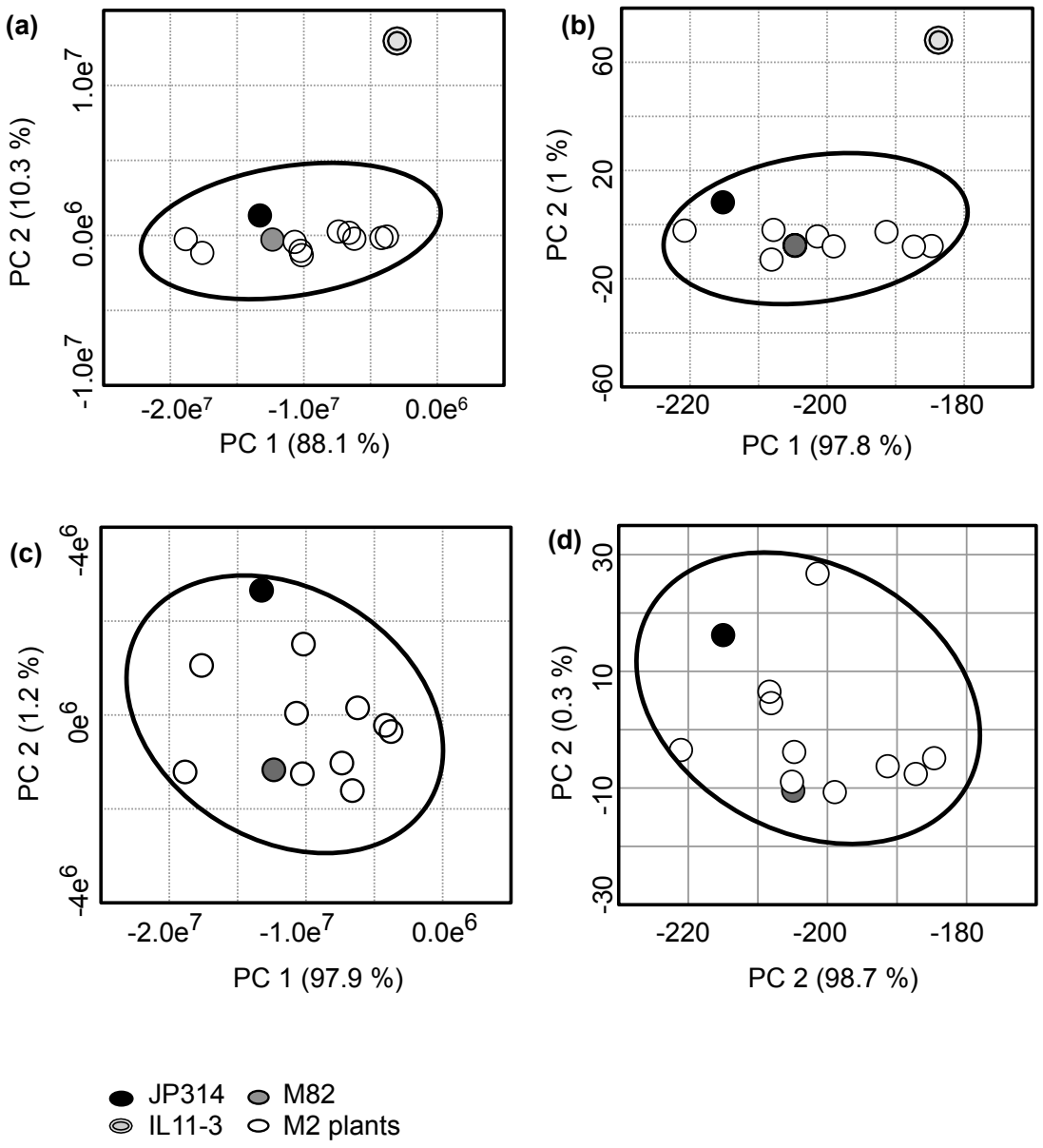


Figure 2.4. Non-targeted data analysis could not identify M2 putative mutant, not even JP314.

Figure 2.4. Continued

Non-targeted data analysis with PCA was not a sensitive approach to identify M2 putative mutants. Scores plot is shown with PC1 and PC2. The 95% confidence region is marked with a circle. Although all 13 plants (12 plants for c and d) were included in this analysis, some plants are almost superimposed with others due to their close position in the scores plot. Filtering was performed by median intensity and scaling by log transformation using MetaboAnalyst webtools (Xia *et al.* 2009). See 'Materials and Methods' for details.

(a) PCA with IL11-3, with neither filtering nor scaling

(b) PCA with IL11-3, with filtering and scaling

(c) PCA without IL11-3, with neither filtering nor scaling

(d) PCA without IL11-3, with filtering and scaling

PCA was performed without IL11-3, but did not identify any outlier M2 plants in this cohort, despite the fact that the targeted data analysis identified JP314 (Figure 2.4cd). The failure could be due to two reasons: 1. the lack of replicate plants in M2 screening; 2. subtle differences in low abundance metabolites.

Identification of novel metabolite changes in M3 progeny of JP314

Eighteen JP314 M3 progeny plants were analyzed by LC-ToF-MS to test the heritability of the phenotype identified from M2 screening. Surprisingly, the altered kaempferol 3-O-rutinoside phentoype was not seen in the M3 progeny, indicating that this trait was not inherited (Figure 2.5). In addition, the M3 progeny showed a phenotype not noted in the M2 sample: reduction in a flavonol tetraglycoside with m/z 949.3 and two previously unannotated analytical signals with m/z 963.3 and 1009.3 (Figure 2.5). These are low abundance ions whose signal intensity is less than 1% of S4:17 (2,5,5,5), the most abundant acylsugar in the cultivated tomato (Table 2.1). These chemical alterations met two serial cutoffs, > 3-fold change and p-value < 0.05 in *Student's t*-test (Figure 2.2). To test whether these chemical phenotypes were consistently seen in the JP314 M3 progeny, another set of nine M3 plants were grown and assayed at different time. The phenotype was consistent in the second set of M3 progeny test (data not shown). These results indicated that the JP314 M3 progeny are true breeding for a mutation that caused the reduction in the flavonol tetraglycoside with m/z 949.3 and unannotated signals with m/z 963.3 and 1009.3.

Non-targeted data analysis to search for additional phenotypes in JP314 M3 progeny

Non-targeted data analysis with PCA and PLS-DA was performed to identify additional analytical signals associated with JP314 M3 plants, and to search for potential correlation

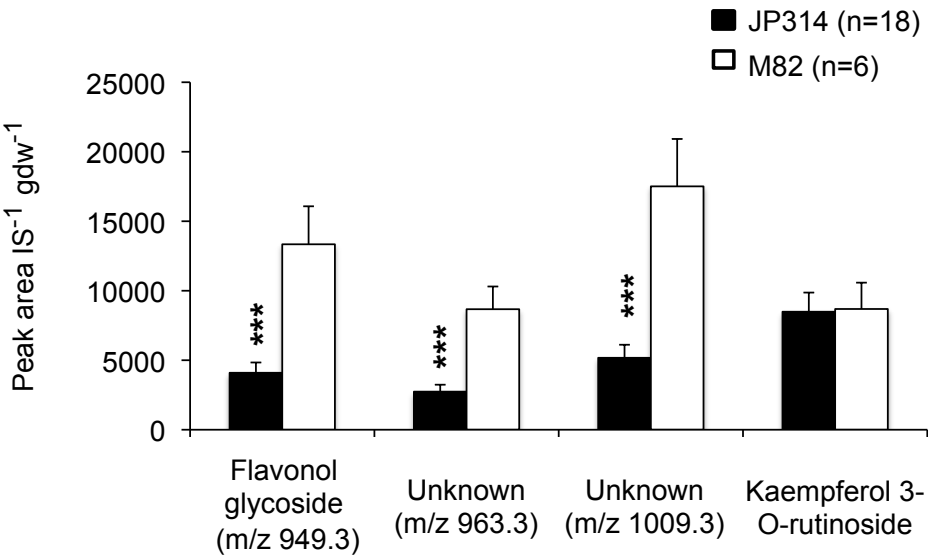
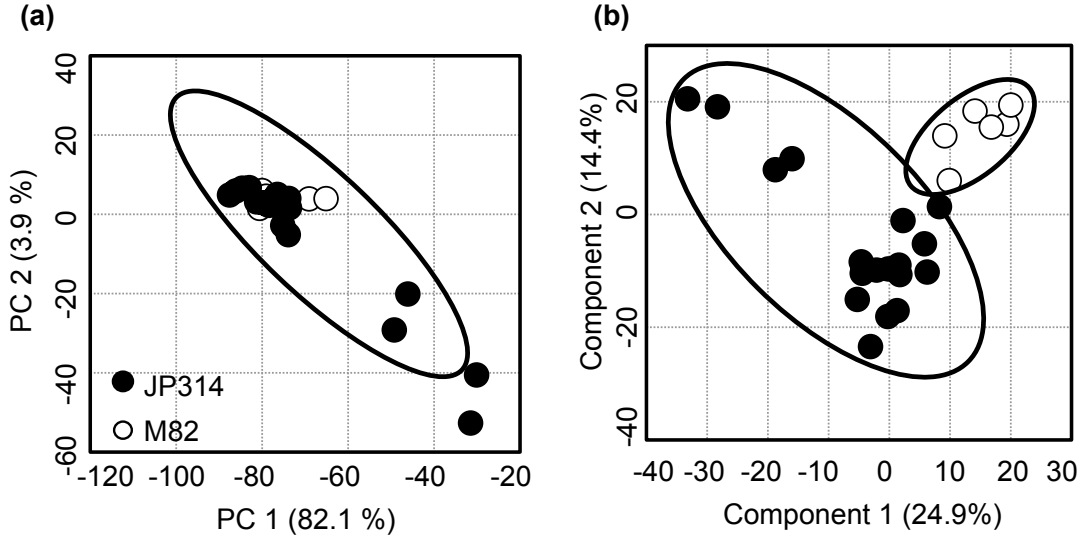


Figure 2.5. Targeted data analysis of JP314 M3 progeny identified chemical phenotypes not seen in M2 screening.

JP314 M3 progeny testing did not confirm the kaempferol 3-O-rutinoside phenotype identified from M2 screening (Figure 2.3). Surprisingly, targeted data analysis of M3 progeny identified novel phenotypes, reduced accumulation of three analytical signals, a flavonol tetraglycoside with m/z 949.3 and two unannotated ions (m/z 963.3, 1009.3). *** indicates p-value < 0.001 from *Student's t*-test. Pearson and Spearman correlation test revealed that the three analytical signals are highly correlated to each other ($r > 0.99$ from both tests).

among the signals (Figure 2.6). PCA did not distinguish JP314 M3 progeny from M82 wild type (Figure 2.6a). This is presumably because PCA works by identifying major components driving the separation of groups; it is not uncommon that this approach to be insensitive to differences in low abundance metabolites (Schilmiller *et al.* 2010a). To identify the influence of low abundance metabolites being altered in the mutant, the supervised method PLS-DA was used (Figure 2.6b-c). The PLS-DA separated the JP314 M3 progeny from M82 (Figure 2.6b) and revealed a list of 30 'Variables Important for Projections' (VIPs), which are analytical signals that contribute to the separation between JP314 M3 progeny and M82 (Figure 2.6c). The VIPs included the three analytical signals previously identified by targeted data analysis (m/z 949.3, 963.3 and 1009. 3, See 'Figure 2.5').

Of 30, 5 VIPs were shown to be highly correlated with the flavonol tetraglycoside with m/z 949.3 ($r > 0.7$ from Pearson and Spearman correlation tests respectively), including m/z 949.3 itself, 963.3, 1009.3 and two other signals: m/z 995.3 and 981.3 and (Figure 2.6c). These five analytical signals had the same LC retention time when run using the 5 min method (Figure 2.6c; Schilmiller *et al.* 2010a), leading to the hypothesis that they are structurally related to the flavonoid glycoside with m/z 949.3. In fact, m/z 995.3 was reported to be a formate adduct of the m/z 949.3 flavonol tetraglycoside (Figure 2.6c, Table 2.1, Schilmiller *et al.* 2010a).



(c)

Rank	Metabolite (m/z_RT)	VIP mean	Rank	Metabolite (m/z_RT)	VIP mean
1	837.3_ 3.6	2.0979	16	793.4_3.9	1.2956
2	1023.3_ 2.5	2.0106	17	657.2_3.3	1.2866
3	583.3_ 3.8	1.9896	18	753.2_3.3	1.2824
4	653.3_3.3 (acylsugar S4:15*)	1.9619	19	387.1_0.3	1.2714
5	**995.3_ 2.2**	1.9375	20	977.3_2.3	1.2674
6	779.4_3.9 (acylsugar S4:24*)	1.9337	21	741.2_2.3	1.2345
7	819.4_3.6	1.9140	22	865.4_3.7	1.2247
8	1289.6_3.3	1.7945	23	725.2_2.4	1.2222
9	1073.3_2.5	1.7904	24	681.3_3.4 (acylsugar S4:17*)	1.2211
10	1261.6_3.3	1.6268	25	707.2_1.7	1.2163
11	**949.3_2.2 (flavonol glycoside*)**	1.4579	26	639.3_3.3 (acylsugar S3:15*)	1.1932
12	**963.3_2.2 (unknown*)**	1.4309	27	1076.5_2.8 (dehydrotomatine*)	1.1837
13	**981.3_2.2**	1.4197	28	**1009.3_2.2 (unknown*)**	1.1528
14	751.4_3.9 (acylsugar S3:23*)	1.3081	29	469.0_0.8	1.1512
15	1497.8_3.8	1.3032	30	851.3_3.6	1.1365

Figure 2.6.

Figure 2.6. Non-targeted data analysis of JP314 M3 progeny validated the flavonol tetraglycoside (m/z 949.3) phenotype and identified unannotated analytical signals highly correlated with the m/z 949.3.

(a) PCA did not distinguish JP314 M3 progeny from M82 wild type plants. Scores plot is shown with PC1 and PC2. The 95% confidence region is marked with a circle.

(b) PLS-DA distinguished JP314 M3 progeny from M82 wild type plants. Scores plot is shown with component 1 and component 2. The 95% confidence region is marked with a circle.

(c) Top 30 Variables Important for Projection (VIPs) from PLS-DA are listed with rank, m/z, LC retention time (RT) and VIP mean. This list includes the flavonol tetraglycoside with m/z 949.3 and four unannotated analytical signals highly correlated with m/z 949.3 (shown in bold). Analytical signals tested with targeted data analysis are labeled with * (See 'Table 2.1' for the details).

Structural characterization of the flavonol tetraglycoside with m/z 949.3

Previous study structurally characterized m/z 949.3 as a flavonol tetraglycoside but did not provide insight into the nature of the flavonol backbone or glycosylation (Schilmiller *et al.* 2010a). Detailed structural characterization of the flavonol tetraglycoside with m/z 949.3 and the other four signals correlated with the flavonol tetraglycoside is challenging because of their low abundance (Table 2.1) and LC co-elution (Figure 2.6c). Several steps were taken to attempt to overcome these obstacles including the analysis of concentrated plant extracts and longer LC (20 minutes) for better separation (See 'Materials and Methods' for the details).

Twenty-minute LC-ToF-MS with multiplexed collision-induced dissociation (MUX-CID) analysis provided the evidence that the flavonol tetraglycoside with m/z 949.3 is a myricetin derivative. Under LC-ToF-MS with MUX-CID analysis, mass spectra are collected at multiple collision energies and saved for each energy condition. MS spectra obtained at low collision energies provide molecular mass information and those generated at higher energies yield successively more fragment ions ((Gu *et al.* 2010, McDowell *et al.* 2011). To test whether the fragmentation of the flavonol tetraglycoside with m/z 949.3 yields the ion of m/z for a flavonol aglycone, MS spectra were searched under 10eV and 55eV (Figure 2.7). Molecular ion for the flavonol tetraglycoside with m/z 949.3 was detected from extracted ion chromatogram (XIC) at low energy conditions (Figure 2.7a) and fragment ions with m/z 317.1, 301.1 and 285.1 corresponding to flavonol aglycones myricetin, quercetin and kaempferol respectively at high energy

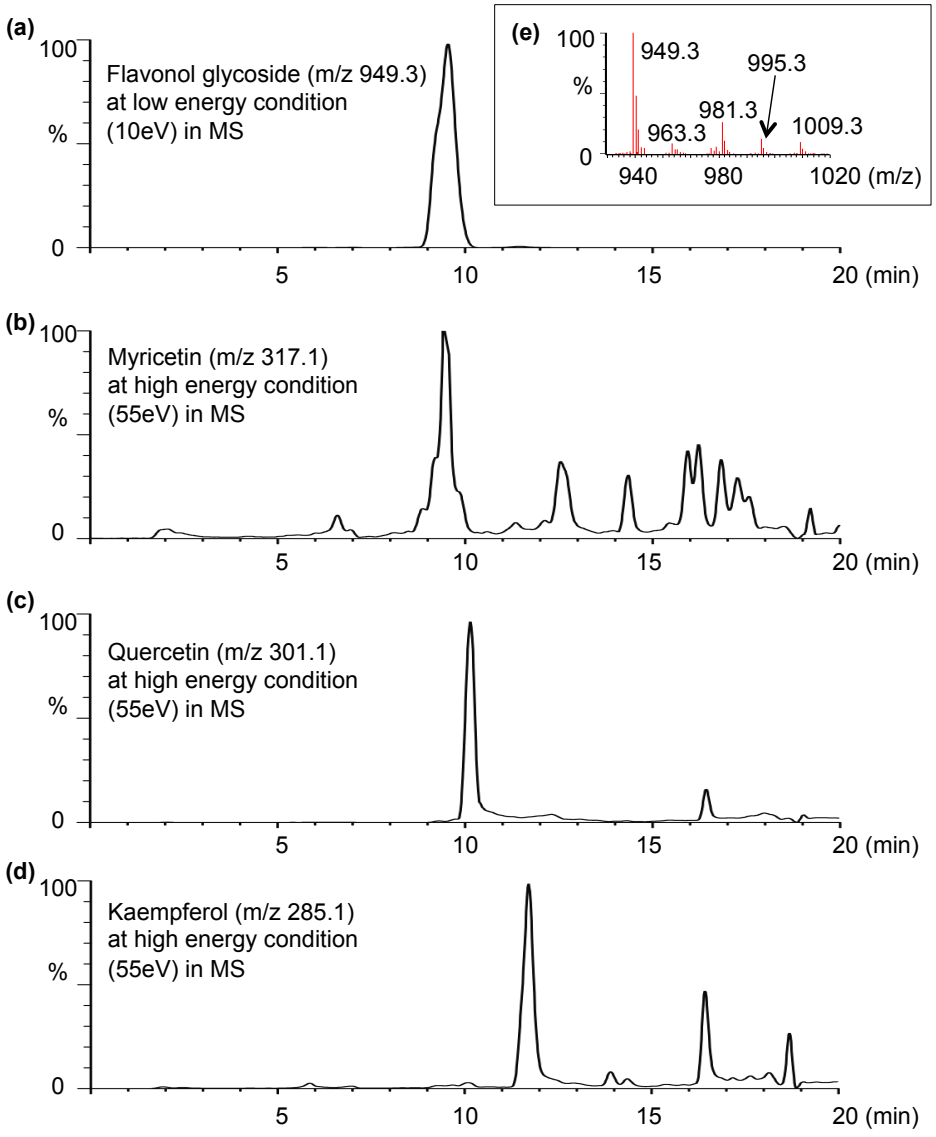

(a)
100
%
0
Flavonol glycoside (m/z 949.3)
at low energy condition
(10eV) in MS
5
10
15
20 (min)
(e)
100
%
0
949.3
995.3
981.3
963.3
1009.3
940
980
1020 (m/z)
(b)
100
%
0
Myricetin (m/z 317.1)
at high energy condition
(55eV) in MS
5
10
15
20 (min)
(c)
100
%
0
Quercetin (m/z 301.1)
at high energy condition
(55eV) in MS
5
10
15
20 (min)
(d)
100
%
0
Kaempferol (m/z 285.1)
at high energy condition
(55eV) in MS
5
10
15
20 (min)

Figure 2.7.

Figure 2.7. LC-ToF-MS with MUX-CID analysis suggested that the flavonol tetraglycoside with m/z 949.3 is a myricetin derivative and revealed adduct ions of this myricetin tetraglycoside.

(a)-(b). Myricetin was shown to result from fragmentation of the flavonol tetraglycoside with m/z 949.3.

(c)-(d). Fragment ions of the flavonol tetraglycoside (m/z 949.3) did not yield ions of m/z 301.1 or 285.1 corresponding to quercetin (c) or kaempferol (d) respectively.

(e) All four ions m/z 963.3, 981.3, 995.3 and 1009.3, shown to be highly correlated with m/z 949.3 in Figure 2.6c, were detected under the peak of m/z 949.3 at 10eV.

condition (Figure 2.7bcd). The LC retention time (RT) for the major peak of m/z 317.1 was identical with that of m/z 949.3. However, neither peak of m/z 301.1 nor 285.1 revealed an identical RT with m/z 949.3. These results suggest that the flavonol tetraglycoside has a myricetin aglycone. The broad chromatographic peak of m/z 949.3 was presumably due to overloading of analytes from the highly concentrated plant extract on the column. The XICs for all three flavonols revealed several chromatographically separable peaks, possibly due to different types of modified flavonols such as methylated and/or glycosylated conjugates (Figure 2.7bcd). All four analytical signals m/z 963.3, 981.3, 995.3 and 1009.3, shown to be highly correlated with the flavonoid glycoside of m/z 949.3 (Figure 2.6c), were detected under the peak of m/z 949.3, suggesting that these four signals are adduct ions of m/z 949.3 (Figure 2.7e). The four individual signals, m/z 963.3, 981.3, 995.3 and 1009.3, are proposed as methylated, oxidized, formate and acetate adduct ion of m/z 949.3 respectively.

Pale leaf color due to reduced total anthocyanins in JP314 mutant

In addition to the trichome myricetin tetraglycoside phenotype, the JP314 mutant has a brighter leaf color than M82 due to the decreased accumulation of total anthocyanins (Figure 2.8a-d). The reduction of total anthocyanins in JP314, only 40% of wild type levels, was estimated by measuring light absorbance of crude leaf extracts (Figure 2.8e).

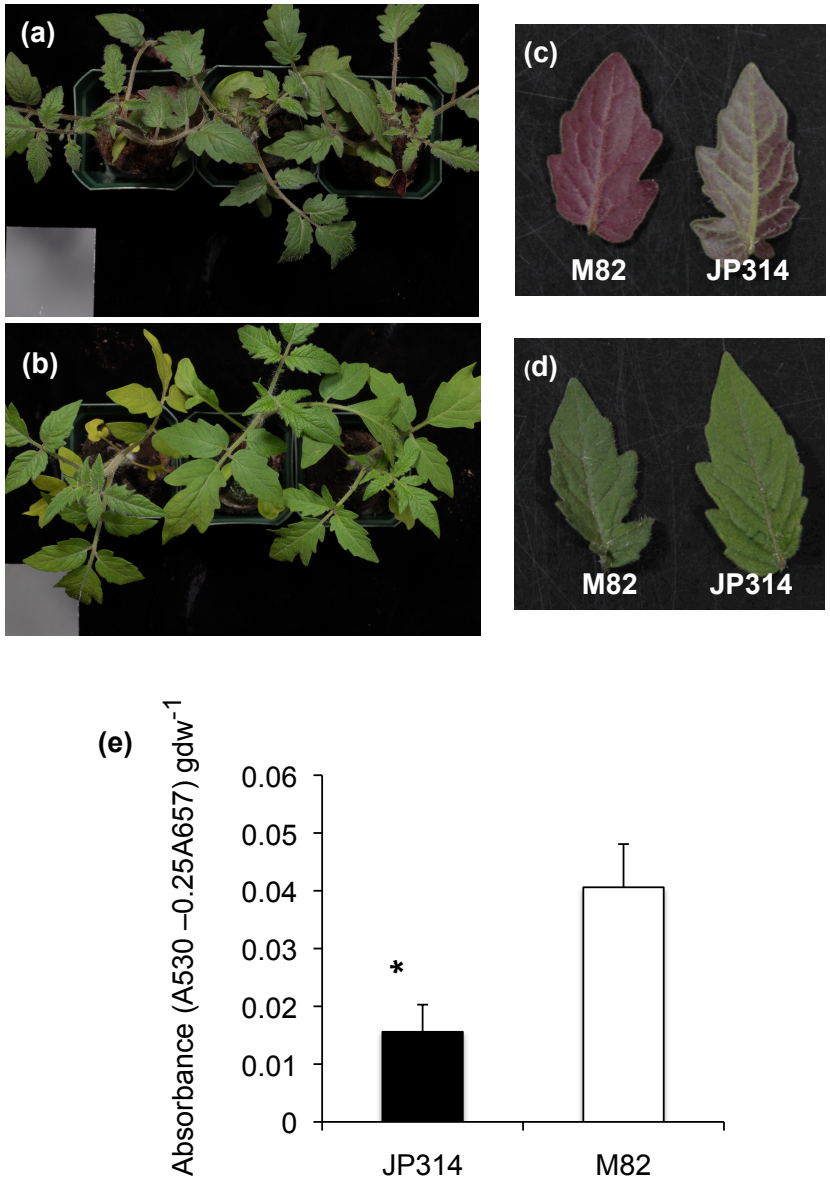


Figure 2.8. Pale leaf color reveals reduced total anthocyanin accumulation in JP314.

Figure 2.8. Continued

(a)-(b) Differences in the color of stems and leaves of M82 wild type (a) and EMS mutant JP314 (b) are visible in three different 3-week-old plants. Reduced purple pigment is very noticeable in the abaxial (c) and adaxial (d) surfaces of JP314 leaves.

(e) Light absorbance, indicative of total anthocyanin amounts, was significantly lower in leaf extracts of JP314 than M82 ($p < 0.05$ in *Student's t*-test). The bar graphs are presented with mean and standard error (n=4 for each genotype).

For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

Discussion

This M2 screen identified 24 heritable mutants altered in various trichome non-volatile metabolites (Table 2.3). Surprisingly, qualitative chemical phenotypes such as the presence or absence of specific metabolites in mutants were not found. In addition, no mutant was identified for relatively high abundance compounds, for example, acylsugars S4:17, S4:24, S3:22 or quercetin 3-O-rutinoside (Table 2.1). The lack of these phenotypes could be due to three reasons. First, a relatively small size of mutant population was screened. While 6,000 mutants were generated for a saturated EMS M2 mutant population (Menda *et al.* 2004), this screening tested 1,387 mutants. Second, despite efforts to 'cast a wide net', cutoffs in the M2 screening might have been so stringent that they missed putative mutants. Third, altered trichome metabolites might have been masked by genetic or

metabolic redundancy (Liljegren *et al.* 2000, Sasaki *et al.* 2002, Vision *et al.* 2000) or promiscuous enzymatic activity (Khersonsky *et al.* 2006, Schmidt *et al.* 2011). Due to these factors, identification of mutant phenotypes might have been limited.

Furthermore, technical challenges in analytical chemistry make the identification of altered trichome metabolites even harder. For instance, structural annotations are available only for a limited number of plant specialized metabolites. Metabolite identification and quantification is also challenging due to the lack of authentic standards and LC-MS database and the inherent issues in LC-MS including matrix effects and

retention time shifts (Last *et al.* 2007, Niessen *et al.* 2006, Wong *et al.* 2008) (See 'Chapter 1. Introduction' for the details).

Despite the challenges in analytical chemistry, LC-ToF-MS was efficient to identify altered accumulation of myricetin tetraglycoside, a low abundance metabolite, whose ion peak is less than 1% of S4:17, the most abundant acylsugar in the cultivated tomato (Table 2.1, Schilmiller *et al.* 2010a). Targeted and non-targeted data analyses identified myricetin tetraglycoside mutant phenotypes and unannotated analytical signals highly correlated with the metabolite. Detailed structural analysis improved metabolite annotation of a flavonol tetraglycoside with m/z 949.3 as myricetin tetraglycoside and identified chemical relationships between this analytical signal and adduct ions.

Although a variety of methylated myricetin compounds have been identified from various plants, relatively few glycosylated myricetins are known. In addition, unlike quercetin and kaempferol flavonols, which are often found as glycosylated with oligosaccharides up to pentasaccharides, examples of myricetins decorated with multiple sugar residues are rare (Kang *et al.* 2010, McDowell *et al.* 2011, Schilmiller *et al.* 2010a, Schmidt *et al.* 2011, Stevens *et al.* 1995, Williams and Grayer 2004). Therefore, it is interesting that this study identified differences in myricetin tetraglycoside accumulation in the trichome exudates from the cultivated tomato and EMS mutants. Specifically, the JP314 mutant has reduced accumulation of the myricetin tetraglycoside. Furthermore, in depth structural characterization of the myricetin tetraglycoside, including sugar types and their

attached positions on the myricetin aglycone backbone, will help us understand trichome myricetin tetraglycoside biochemistry.

Beside the decreased trichome myricetin tetraglycoside, the JP314 mutant has reduced total anthocyanins in crude leaf extracts. Given that the flavonoid and anthocyanin biosynthetic pathways share common precursors such as dihydroflavonols, the mutant phenotypes led to the hypothesis that the JP314 mutation alters the level of dihydromyricetin, a precursor of myricetin and its modified conjugates and anthocyanins structurally related to myricetin (Figure 2.9). Previous study reported that the white color mutant of the Bilberry fruit (*Vaccinium myrtillus*) lacks myricetin and myricetin-derived anthocyanin accumulation, whereas the wild type fruit highly accumulates anthocyanins and myricetin flavonol (Jaakola *et al.* 2002). Jaakola *et al.* 2002 also described that flavonoid 3',5'-hydroxylase, a gene converting dihydrokaempferol to dihydromyricetin, is involved in the production of myricetin and myricetin-derived anthocyanins in Bilberry fruits. Given the JP314 mutant phenotypes, it will be useful to analyze anthocyanin profiles to test whether the reduced total anthocyanin accumulation in the JP314 mutant is due to the lack of myricetin-derived anthocyanins. If this hypothesis is supported, examining expression and sequence of the flavonoid 3',5'-hydroxylase gene will be helpful to test the hypothesis that this gene is altered in JP314 mutant.

In summary, the combination of analytical chemistry, forward genetics and statistical tests identified tomato EMS mutants altered in trichome non-volatile metabolites.

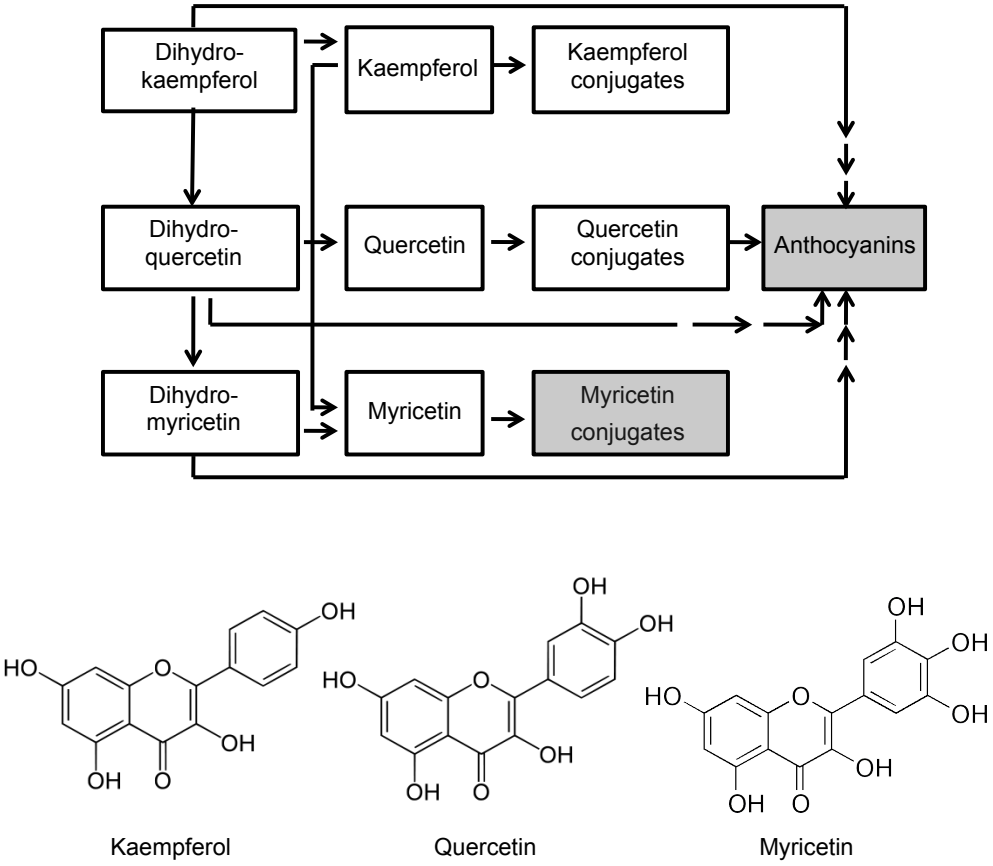


Figure 2.9. Overview of flavonoid and anthocyanin biosynthetic pathways.

Schematic flavonoid and anthocyanin biosynthetic pathways are shown with grey boxes indicating reduced amounts in JP314 mutant. Structures are presented for the flavonols kamepferol, quercetin and myricetin.

Furthermore, this chemical analysis of mutant plants improved structural annotation of incompletely characterized metabolites and led to the generation of testable hypotheses.

Materials and methods

Plant growth conditions

M82 cultivated tomato seeds were obtained from the Tomato Genetic Resource Center (http://tgrc.ucdavis.edu/). Tomato EMS M2 mutant population was generously provided from Dr. Dani Zamir group (The Hebrew Univeristy of Jerusalem, Faculty of Agriculture, Rehobot, Israel). Thirty-two plants including one or two of each M82 wild type and IL control plants were grown in a single flat for 3 weeks. The plants were randomly located within the flat. Then, chemical assays were conducted on leaves from the plants except ones having poor growth. Identical growth conditions and chemical assay methods were applied as described in (Schilmiller *et al.* 2010a)

Chemical extraction method and LC-ToF-MS analysis

A single leaf dip extraction method and rapid LC-ToF-MS analysis with MUX-CID was performed for mutant screening as described in (Schilmiller *et al.* 2010a).

Highly concentrated trichome extracts were analyzed by LC-ToF-MS with 20 min LC performance for structural characterization of the flavonol tetraglycoside (m/z 949.3). Because the accumulation of the metabolite was higher in M82 than JP314 mutant, M82 leaves were used for chemical extraction. Ten of the youngest leaves from a single M82 plant were collected and they were placed into 10 ml of extraction solvent used for a

single leaf dip method. The incubation was performed for 1 min with a gentle rocking. The same procedure was repeated in 10 M82 plant batches, for a total of 100 leaves in 100 ml of solvent. The solvent from plant extracts was brought to 10 ml by rotary evaporation at room temperature. The 10 ml of concentrated plant extract was transferred to a glass test tube, evaporated to dryness under a liquid nitrogen stream and re-dissolved in 100 μl of extraction solvent. This highly concentrated plant extract was analyzed by 20 min LC method. The gradient of mobile phase was also optimized for better separation of flavonoids. Two solvents (solvent A – 0.15% formic acid in MilliQ water, solvent B – methanol) were applied as mobile phases in LC with 0.4 ml/min flow rate. Gradient profile of mobile phase was as follows: 1% B for the first 0.5 min; linear gradient to 60% B to 12 min; 85% B to 15 min; 100% B maintained to 18 min; 1% B to 20 min.

Measurement of total anthocyanin levels

Crude leaf extraction was performed with 0.3% hydrochloric acid in methanol to extract total anthocyanins (Kong *et al.* 2003). A single leaflet was collected from the identical position where plant tissues were obtained for trichome non-volatile metabolite profiling. The leaf was placed into 1 ml of anthocyanin extraction solvent and incubated for 24 hrs with gentle rocking. The amount of total anthocyanins was estimated by measuring light absorbance of the plant extracts at the specific wavelength. The light absorbance was measured at 530 nm (A530) and 657 nm (A657) to evaluate the accumulation of total anthocyanins and chlorophyll degradation products, respectively. Finally, A530 - 0.25 x A657 (Spitzer-Rimon *et al.* 2010) was calculated as the indicative of total anthocyanin levels.

Targeted data analysis

Peak areas were integrated for the individual 35 analytical signals using QuanLynx software (Waters, http://www.waters.com/) and normalized to internal standard and leaf dry weight. The normalized dataset was used for targeted data analysis.

Median and median absolute deviation (MAD) were calculated as described in (Lu *et al.* 2008), to look for outliers altered in the 35 analytical signals in M2 screening. Every comparative analysis and statistical test was performed with plants grown in a same flat to control for flat difference. The median and MAD were used instead of mean and standard deviation, respectively, because the median is less likely to be affected by extreme values, for example, outliers, compared to mean. In M2 screening, two serial cutoffs were applied to identify putative mutants. First, median +/- 5 MAD were used to identify outliers. Once outliers were selected, it was tested whether the amount of specific analytical signal in the outliers has > 3-fold change compared to the median. When the outliers met two serial cutoffs, median +/- 5 MAD and > 3-fold change, they were considered M2 putative mutants.

Non-targeted data analysis

The signal intensity of all MS spectra ranging from 50 to 1,500 Da was obtained using Markerlynx software (Waters, http://www.waters.com/) and normalized to internal standard and leaf dry weight. The normalized dataset was used to perform statistical tests including PCA, PLS-DA and correlation analysis, using MetaboAnalyst webtools (Xia *et al.* 2009). PCA and PLS-DA were performed with various conditions such as

with/without filtering and/or scaling. Median intensity and log transformation was used for filtering and scaling respectively when applied. Variable Importance in Projection (VIP) was ranked based on VIP mean, the indicative of the significance of the variables for the separation in PLS-DA (Xia *et al.* 2009). Pearson and Spearman correlation analyses were also performed to search for analytical signals whose behavior was associated with the flavonol tetraglycoside with m/z 949.3.

Chemical analysis of M3 progeny plants

Seeds were collected from the self-pollinated M3 progeny of the individual M2 putative mutants. Identical growth conditions and chemical assay methods were applied to M3 progeny test as M2 mutant screening.

References

References

Besser, K., Harper, A., Welsby, N., Schauvinhold, I., Slocombe, S., Li, Y., Dixon, R.A. and Broun, P. (2009) Divergent regulation of terpenoid metabolism in the trichomes of wild and cultivated tomato species. *Plant Physiol*, **149,** 499-514.

Buttner, B., Abou-Elwafa, S.F., Zhang, W., Jung, C. and Muller, A.E. (2010) A survey of EMS-induced biennial *Beta vulgaris* mutants reveals a novel bolting locus which is unlinked to the bolting gene B. *Theor. Appl. Genet.*, **121,** 1117-1131.

Falara, V., Akhtar, T.A., Nguyen, T.T., Spyropoulou, E.A., Bleeker, P.M., Schauvinhold, I., Matsuba, Y., Bonini, M.E., Schilmiller, A.L., Last, R.L., Schuurink, R.C. and Pichersky, E. (2011) The tomato terpene synthase gene family. *Plant Physiol*, **157,** 770-789.

Fobes, J.F., Mudd, J.B. and Marsden, M.P.F. (1985) Epicuticular lipid accumulation on the leaves of *Lycopersicon pennellii* (Corr.) D'Arcy and *Lycopersicon esculentum* Mill. *Plant Physiol.*, **77,** 567-570.

Fridman, E., Wang, J., Iijima, Y., Froehlich, J.E., Gang, D.R., Ohlrogge, J. and Pichersky, E. (2005) Metabolic, genomic, and biochemical analyses of glandular trichomes from the wild tomato species *Lycopersicon hirsutum* identify a key enzyme in the biosynthesis of methylketones. *Plant Cell*, **17,** 1252-1267.

Gu, L., Jones, A.D. and Last, R.L. (2010) Broad connections in the Arabidopsis seed metabolic network revealed by metabolite profiling of an amino acid catabolism mutant. *Plant J.*, **61,** 579-590.

Hanhineva, K., Rogachev, I., Kokko, H., Mintz-Oron, S., Venger, I., Karenlampi, S. and Aharoni, A. (2008) Non-targeted analysis of spatial metabolite composition in strawberry (*Fragaria* x *ananassa*) flowers. *Phytochemistry*, **69,** 2463-2481.

Jaakola, L., Maatta, K., Pirttila, A.M., Torronen, R., Karenlampi, S. and Hohtola, A. (2002) Expression of genes involved in anthocyanin biosynthesis in relation to anthocyanin, proanthocyanidin, and flavonol levels during bilberry fruit development. *Plant Physiol*, **130,** 729-739.

Jander, G., Baerson, S.R., Hudak, J.A., Gonzalez, K.A., Gruys, K.J. and Last, R.L. (2003) Ethylmethanesulfonate saturation mutagenesis in Arabidopsis to determine frequency of herbicide resistance. *Plant Physiol*, **131,** 139-146.

Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M. and Last, R.L. (2002) *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol.*, **129,** 440-450.

Kang, J.H., Liu, G., Shi, F., Jones, A.D., Beaudry, R.M. and Howe, G.A. (2010) The tomato *odorless-2* mutant is defective in trichome-based production of diverse specialized metabolites and broad-spectrum resistance to insect herbivores. *Plant Physiol.*, **154,** 262-272.

Khersonsky, O., Roodveldt, C. and Tawfik, D.S. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.*, **10,** 498-508.

Kong, J.M., Chia, L.S., Goh, N.K., Chia, T.F. and Brouillard, R. (2003) Analysis and biological activities of anthocyanins. *Phytochemistry*, **64,** 923-933.

Krieger, U., Lippman, Z.B. and Zamir, D. (2010) The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nat. Genet.*, **42,** 459-463.

Last, R.L., Jones, A.D. and Shachar-Hill, Y. (2007) Towards the plant metabolome and beyond. *Nat. Rev. Mol. Cell. Biol.*, **8,** 167-174.

Liljegren, S.J., Ditta, G.S., Eshed, Y., Savidge, B., Bowman, J.L. and Yanofsky, M.F. (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature*, **404,** 766-770.

Lu, Y., Savage, L.J., Ajjawi, I., Imre, K.M., Yoder, D.W., Benning, C., Dellapenna, D., Ohlrogge, J.B., Osteryoung, K.W., Weber, A.P., Wilkerson, C.G. and Last, R.L. (2008) New connections across pathways and cellular processes: industrialized mutant screening reveals novel associations between diverse phenotypes in *Arabidopsis*. *Plant Physiol.*, **146,** 1482-1500.

Luckwill, L. (1943) *The genus Lycopersicon: A historical, biological, and taxonomic survey of the wild and cultivated tomatoes* Aberdeen, Scotland: Aberdeen University Press.

Marti, L., Stefano, G., Tamura, K., Hawes, C., Renna, L., Held, M.A. and Brandizzi, F. (2010) A missense mutation in the vacuolar protein GOLD36 causes organizational defects in the ER and aberrant protein trafficking in the plant secretory pathway. *Plant J.*, **63,** 901-913.

McConkey, M.E., Gershenzon, J. and Croteau, R.B. (2000) Developmental regulation of monoterpene biosynthesis in the glandular trichomes of peppermint. *Plant Physiol*, **122,** 215-224.

McDowell, E.T., Kapteyn, J., Schmidt, A., Li, C., Kang, J.H., Descour, A., Shi, F., Larson, M., Schilmiller, A., An, L., Jones, A.D., Pichersky, E., Soderlund, C.A. and Gang, D.R. (2011) Comparative functional genomic analysis of Solanum glandular trichome types. *Plant Physiol.*, **155,** 524-539.

Menda, N., Semel, Y., Peled, D., Eshed, Y. and Zamir, D. (2004) In silico screening of a saturated mutation library of tomato. *Plant J.*, **38,** 861-872.

Niessen, W.M., Manini, P. and Andreoli, R. (2006) Matrix effects in quantitative pesticide analysis using liquid chromatography-mass spectrometry. *Mass. Spectrom. Rev.*, **25,** 881-899.

Peters, J.L., Cnudde, F. and Gerats, T. (2003) Forward genetics and map-based cloning approaches. *Trends. Plant. Sci.*, **8,** 484-491.

Sallaud, C., Rontein, D., Onillon, S., Jabes, F., Duffe, P., Giacalone, C., Thoraval, S., Escoffier, C., Herbette, G., Leonhardt, N., Causse, M. and Tissier, A. (2009) A novel pathway for sesquiterpene biosynthesis from Z,Z-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. *Plant Cell*, **21,** 301-317.

Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y., Antonio, B.A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K., Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K., Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y., Idonuma, A., Iijima, M., Ikeda, M., Ikeno, M., Ito, S., Ito, T., Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S., Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T., Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M., Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M., Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T., Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K., Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K., Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J.H., Kim, H.I., Eun, M.Y., Yano, M., Jiang, J. and Gojobori, T. (2002) The genome sequence and structure of rice chromosome 1. *Nature*, **420,** 312-316.

Schilmiller, A., Shi, F., Kim, J., Charbonneau, A.L., Holmes, D., Daniel Jones, A. and Last, R.L. (2010a) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. *Plant J.*, **62,** 391-403.

Schilmiller, A.L., Last, R.L. and Pichersky, E. (2008) Harnessing plant trichome biochemistry for the production of useful compounds. *Plant J.*, **54,** 702-711.

Schilmiller, A.L., Miner, D.P., Larson, M., McDowell, E., Gang, D.R., Wilkerson, C. and Last, R.L. (2010b) Studies of a biochemical factory: tomato trichome deep expressed sequence tag sequencing and proteomics. *Plant Physiol.*, **153,** 1212-1223.

Schilmiller, A.L., Schauvinhold, I., Larson, M., Xu, R., Charbonneau, A.L., Schmidt, A., Wilkerson, C., Last, R.L. and Pichersky, E. (2009) Monoterpenes

in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc. Natl. Acad. Sci. U S A*, **106,** 10865-10870.

Schmidt, A., Li, C., Shi, F., Jones, A.D. and Pichersky, E. (2011) Polymethylated myricetin in trichomes of the wild tomato species *Solanum habrochaites* and characterization of trichome-specific 3'/5'- and 7/4'-myricetin O-methyltransferases. *Plant Physiol.*, **155,** 1999-2009.

Spitzer-Rimon, B., Marhevka, E., Barkai, O., Marton, I., Edelbaum, O., Masci, T., Prathapani, N.K., Shklarman, E., Ovadis, M. and Vainstein, A. (2010) *EOBII*, a gene encoding a flower-specific regulator of phenylpropanoid volatiles' biosynthesis in Petunia. *Plant Cell*, **22,** 1961-1976.

Stevens, J., Hart, H. and Wollenweber, E. (1995) The systematic and evolutionary significance of exudate flavonoids in *Aeonium. Phytochemistry,* **39,** 805-813.

Sumner, L.W., Mendes, P. and Dixon, R.A. (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*, **62,** 817-836.

Tikunov, Y., Lommen, A., de Vos, C.H., Verhoeven, H.A., Bino, R.J., Hall, R.D. and Bovy, A.G. (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiol.*, **139,** 1125-1137.

Turner, G.W., Gershenzon, J. and Croteau, R.B. (2000) Development of peltate glandular trichomes of peppermint. *Plant Physiol.*, **124,** 665-680.

Vision, T.J., Brown, D.G. and Tanksley, S.D. (2000) The origins of genomic duplications in *Arabidopsis*. *Science*, **290,** 2114-2117.

Wagner, G.J. (1991) Secreting glandular trichomes: more than just hairs. *Plant Physiol.*, **96,** 675-679.

Westerhuis, J.A., Hoefsloot, H.C.J., Smit, S., Vis, D.J., Smilde, A.K., van Velzen, E.J.J., van Duijnhoven, J.P.M. and van Dorsten, F.A. (2007) Assessment of PLSDA cross validation. *Metabolomics*, **4,** 81-89.

Williams, C. and Grayer, R. (2004) Anthocyanins and other flavonoids. *Nat. Prod. Rep.*, **21,** 539-573.

Wink, M. (1999) *Biochemistry of plant secondary metabolism* edn. Sheffield, England: Sheffield Academic Press.

Wong, M.C., Lee, W.T., Wong, J.S., Frost, G. and Lodge, J. (2008) An approach towards method development for untargeted urinary metabolite profiling in metabonomic research using UPLC/QToF MS. *J. Chromatogr. B.*, **871,** 341-348.

Xia, J., Psychogios, N., Young, N. and Wishart, D.S. (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic. Acids. Res.*, **37,** W652-660.

Yu, G., Nguyen, T.T., Guo, Y., Schauvinhold, I., Auldridge, M.E., Bhuiyan, N., Ben-Israel, I., Iijima, Y., Fridman, E., Noel, J.P. and Pichersky, E. (2010) Enzymatic functions of wild tomato methylketone synthases 1 and 2. *Plant Physiol.*, **154,** 67-77.

Chapter 3. Analysis of genetic variants of *Solanum lycopersicum* with altered trichome methylated flavonoids

Abstract

Flavonoids are ubiquitous plant aromatic specialized metabolites found in a variety of cell types and organs. Glycosylated and methylated flavonoids are found in various Solanum plants. For example, *Solanum lycopersicum* M82 secreting glandular trichomes (SGTs) accumulate both glycosylated and methylated myricetin flavonol compounds. As part of a genetic dissection of SGT specialized metabolites in M82, random EMS mutants and a collection of *S. lycopersicum* X *S. pennellii* (LA0716) introgression lines (ILs) were screened for changes in flavonoid types. The JP117 EMS mutant was identified as having reduced amounts of di- and tri-methylated myricetin, and increased mono-methylated myricetin. The recessive mutation in JP117 was mapped to the bottom of tomato chromosome 6. In contrast, introduction of the region of *S. pennellii* chromosome 6 defined by introgression lines (ILs) 6-3 and 6-4 caused a shift toward more highly methylated myricetin. The overlapping introgressed region of the *S. pennellii* chromosome in 6-3 and 6-4 is completely contained within the chromosomal region for the JP117 mutant gene. Within the targeted chromosomal region, two *S. lycopersicum* homologous genes for myricetin *O*-methyltransferases were identified, slMOMT1a and 1b, and are candidate genes for the production of Me-Ms. From insect feeding assays in the field and green house, JP117 mutant showed higher susceptibility and IL6-4 more resistance than M82 plants against *Manduca sexta*, tobacco hornworm. This trend was consistently discovered against the flea beetle *Epitrix cucumeris* from an independent

field experiment, suggesting an impact of altered Me-M profiles on plant defense against insect herbivores.

Introduction

Secretory glandular trichomes accumulate a wide variety of plant specialized metabolites (Schilmiller *et al.* 2008, Wagner 1991). These chemical compounds protect plants against UV-B radiation, water loss, heat stress (Wagner *et al.* 2004) and attack from herbivorous insects (Kang *et al.* 2010a, Kang *et al.* 2010b, Weinhold and Baldwin 2011). Furthermore, trichome metabolites contribute to food flavor and aroma (Gang 2005, Wang *et al.* 2008) and are used as therapeutic drugs (Liu *et al.* 2010, van Der Heijden *et al.* 2004). Due to these benefits, trichome metabolites have been intensively studied in various plants (Schilmiller *et al.* 2008).

Flavonoids and structurally related compounds are examples of plant non-volatile specialized metabolites, collectively including chalcones, flavonones, flavones, flavonols and anthocyanins (Buer *et al.* 2010). The first step in flavonoid biosynthesis is chalcone production by chalcone synthase using acetate-derived malonyl-CoA and p-coumaryl-CoA as substrates (Figure 1.2 from Chapter 1. Introduction). Chalcone is used as a precursor to produce diverse flavonoids by a series of reactions such as oxidation, reduction and isomerization (Buchana *et al.* 2002).

Among various flavonoids, flavonols such as kaempferol, quercetin and myricetin are relatively abundant in plants (Heim *et al.* 2002, Stevens *et al.* 1995). These flavonols are

typically found as modified compounds rather than unmodified flavonol molecules in plants (Stevens *et al.* 1995). For example, glycosylated and/or methylated flavonols were identified from various plant cell types, including trichomes of the cultivated tomato and its wild relatives (Braca *et al.* 2001, McDowell *et al.* 2011, Schilmiller *et al.* 2010a, Schmidt *et al.* 2011, Stevens *et al.* 1995). Chemical modifications are known to affect the chemical properties of flavonoids. For instance, *O*-methylation at hydroxyl group of flavonoids makes the compounds more hydrophobic and changes their reactivity and accessibility (Heim *et al.* 2002).

Previous studies identified plant flavonoid *O*-methyltransferases and characterized their enzymatic activities (Kim *et al.* 2005, Muzac *et al.* 2000). These results commonly showed regio-specificity of the enzymes (Kim *et al.* 2005, Muzac *et al.* 2000). Two myricetin *O*-methyltransferases recently characterized from the wild species of tomato *S. habrochaites* (LA1777) also exhibited regio-specificity (Schmidt *et al.* 2011). The regio-specificity of flavonoid *O*-methyltransferase provides a potential explanation of how methylated flavonoids are produced with methyl groups at different positions. This shows that the identification and characterization of flavonoid *O*-methyltransferases enzymatic activities facilitates understanding of methylated flavonoid production.

The availability of various genetic resources makes tomato plants appealing model systems for studies of many biological processes. For example, with EMS mutant population (See 'Chapter 2' for the details), and chromosomal substitution lines such as introgression lines (ILs) are highly useful plant resources. The commonly used *S.*

pennellii LA0716 ILs replace single regions of the *S. lycopersicum* M82 genome with a homologous region from *S. pennellii* LA0716. When phenotypes are found in specific ILs, locus identification can be efficient because individual *S. pennellii* ILs carry a genetically characterized fragment of the *S. pennellii* chromosome. The identification of a novel trichome monoterpene biosynthetic pathway is a recent example of the utility of these ILs in gene discovery (Schilmiller *et al.* 2009).

Physiological and ecological studies revealed that plant specialized metabolites are important for plant defense against herbivore insects. For instance, volatile trichome terpenes were shown to improve plant defense against insects in local plants (Weinhold and Baldwin 2011) and alarm neighbor plants in the community (Baldwin *et al.* 2006). Flavonoids were also discovered to affect insect feeding (Onyilagha *et al.* 2004). Specifically, flavonols inhibit insect cytochrome P450 dependent steroid hydroxylase activity in a concentration-dependent and species-specific manner (Mitchell *et al.* 1993).

In this study, EMS mutants and *S. lycopersicum* X *S. pennellii* 0716 ILs with altered degree of myricetin methylation were identified and characterized. While di-methylated myricetin (di-Me-M) was the most abundant in the M82 cultivated tomato, the JP117 mutant and IL6-3/4 accumulated predominantly mono-Me-M and tri-Me-M, respectively. Genetic analysis revealed that JP117 behaves as a single recessive mutation and the *S. pennellii* allele is dominant in IL6-3/4. Genetic mapping placed the JP117 locus on the bottom of chromosome 6, in the same region of the *S. pennellii* chromosome in IL6-3/4.

Insect feeding suggested an impact of altered Me-Ms on plant resistance against the insects *Manduca sexta* and *Epitrix cucumeris.*

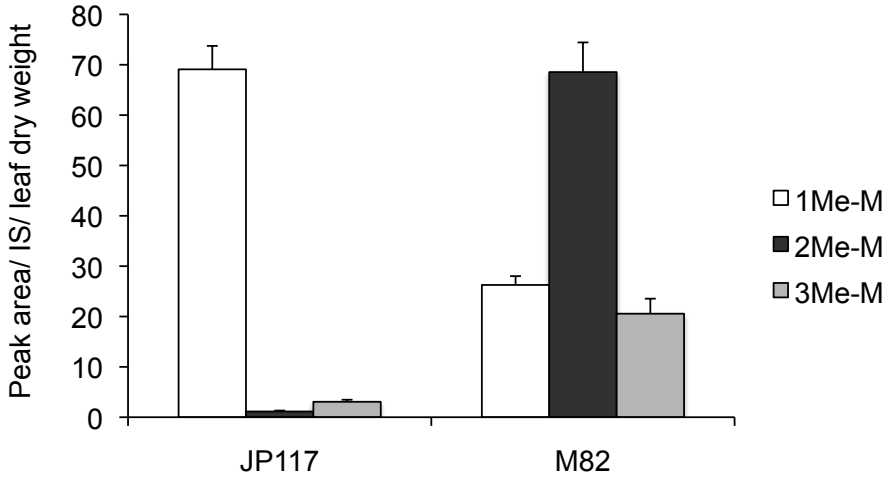
Results

Phenotypic characterization of altered Me-M profiles in the JP117 EMS mutant

Seven mutants with increased ratios of mono- to di-Me-M (JP117, 130, 256, 334, 446, 3045 and 3060) were identified from the tomato EMS mutant screen described in Chapter 2 (See Table 2.3 from Chapter 2). Three types of Me-Ms, mono-, di- and tri-Me-M, were detected in M82 cultivated tomato analysis of single leaf dip trichome extracts using liquid chromatography-time of flight-mass spectrometry (LC-ToF-MS) in negative ion mode (Figure 3.1). While the normalized ion peak area for di-Me-M was the most abundant in M82, mono- and tri-Me-M were also detected. In contrast, the relative proportions were shifted in the JP117 mutant such that mono-Me-M was predominant, with lower amount of di- and tri-Me-Ms (Figure 3.1a). The total accumulation of Me-Ms was also significantly decreased in the JP117 mutant (approximately 64% of M82, p-value < 0.0005 from *Student's t*-test) (Figure 3.1b).

Negative ion mode LC-ToF-MS analysis of single leaf dip extracts permits rapid analysis of mono-, di- and tri-Me-Ms. However, this method provides limited molecular structure information and might extract non-trichome epidermal cell compounds. Chao Li, Daniel Jones and I used the complementary LC-MS/MS positive ion mode analysis of individual trichome types. This approach revealed several similarities to the LC-ToF-MS method

(a)



(b)

Genotypes	Total Me-Ms	p-value
JP117	73.3 ± 4.8	0.000213
M82	115.4 ± 9.1	

Figure 3.1. Increased mono-Me-M and decreased di- and tri-Me-M accumulation in the JP117 EMS mutant identified from LC-ToF-MS analysis of single leaf dip extracts.

Figure 3.1. Continued

(a) Compared to the M82 wild type plant, the JP117 mutant (M3 plant) accumulated an increased amount of mono-Me-M with decreased di- and tri-Me-M levels. Average of peak areas normalized to internal standard and leaf dry weight are presented with standard error of the mean (n=24 for JP117, 21 for M82). 1Me-M, 2Me-M and 3Me-M indicates mono-, di- and tri-Me-M respectively. An identical system is used in subsequent figures describing LC-ToF results, unless specified otherwise.

(b) The total amount of Me-Ms is decreased in the JP117 mutant to approximately 64% that of M82 (p-value < 0.0005 from *Student's t*-test). The total Me-M level was calculated by summing normalized peak areas of all three Me-Ms from samples used in (a).

and to published LC-MS/MS data for *S. habrochaites* LA1777 (Figure 3.3; Schmidt *et al.* 2011). The JP117 mutant trichomes accumulated a higher amount of mono-Me-M having a methyl group at position 3 [mono-Me-M (3)], than was detected in M82 (See Figure 3.2 for the flavonoid numbering system). In contrast, di-Me-M (3,3') levels were approximately 14-fold higher in M82 trichomes than in JP117 (Figure 3.3). Thus, LC-MS/MS positive ion mode analysis of isolated trichomes revealed the same increased ratio of mono- to di-Me-M in JP117 trichomes compared with M82 as originally identified by negative ion mode LC-ToF-MS analysis of single leaf dip extracts (Figure 3.1 and 3.3). Furthermore, higher accumulation of Me-Ms was found in type 1 glands than type 6 for the JP117 mutant and M82 (Figure 3.3), consistent with recently published data that Me-Ms are enriched in type 1 glands of *S. habrochaites* LA1777 (Schmidt *et al.* 2011).

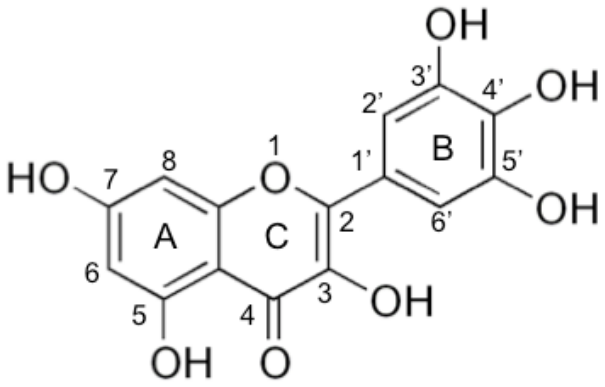


Figure 3.2. The chemical structure of unmodified myricetin.

The structure of the flavonol myricetin is shown with the numbering system indicated. The hydroxyl groups of myricetin are typically modified *in vivo*, for example, by methylation or glycosylation.

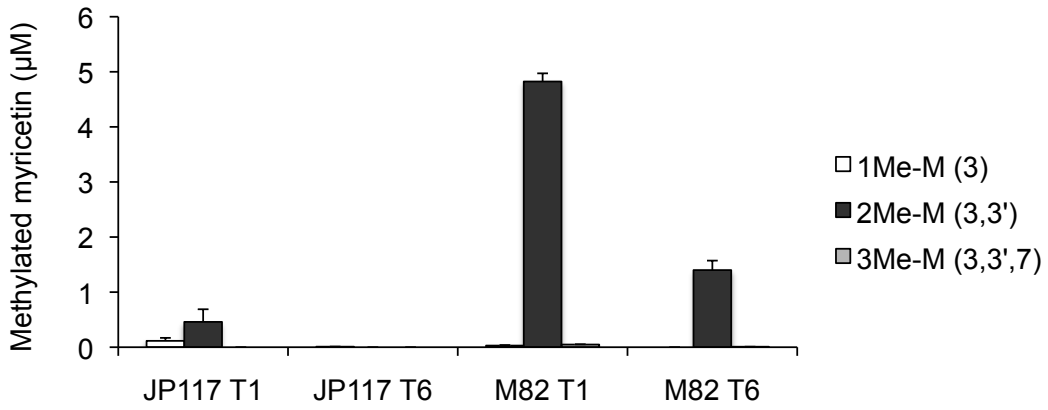


Figure 3.3. The Me-M levels measured in isolated glands of types 1 and 6 trichomes from JP117 mutant and M82 plants.

LC-MS/MS analysis of trichome extracts revealed higher accumulation of Me-Ms in type 1 (T1) glands than type 6 (T6) from both the JP117 mutant and M82 wild type plants. While 1Me-M levels were higher in JP117 mutant trichomes than M82, 2Me-M was higher in M82 trichomes than JP117. LC-MS/MS also revealed the position of methyl groups for individual Me-Ms. For instance, 3Me-M (3,3',7) has methyl groups on at the 3, 3' and 7 positions (See Figure 3.2 for the numbering system). The amount of methylated myricetin (μM) is presented with the average and standard error (See 'Materials and Methods' for metabolite detection and identification). Fifty glands of specific types of trichomes were collected from each of three biological replicate plants.

In addition to validating previously identified chemical profiles, the LC-MS/MS analysis expands our understanding about Me-M phenotypes in two ways. First, it provides information, regarding the position of methyl groups on Me-Ms. Second, the use of commercially available methylated myricetins as a flavonoid calibration standard (See 'Materials and Methods' for the details) provided corroboration for the relative concentrations of Me-Ms in M82 and JP117 compared with quantititave Me-M profiles obtained by normalized peak areas under LC-ToF-MS method.

Genetic analysis of JP117

To test whether JP117 behaves as a recessive monogenic mutation, it was backcrossed to the parental M82 wild type. Backcross F1 progeny plants (BC F1) were analyzed by LC-ToF-MS and their Me-M profiles compared with those of JP117 and M82. The Me-M profiles in BC F1 were similar to those of M82 in that di-Me-M was most abundant in every F1 plant (Figure 3.4a). These results are consistent with the hypothesis that the JP117 mutation is recessive to wild type. However, while di-Me-M is restored to wild type levels, the overall F1 phenotype differs from M82 in that it accumulated mono-Me-M intermediate between that of JP117 and M82 (Figure 3.4a). This gene dosage effect suggests a defect in a structure gene(s) rather than a regulatory gene(s) in JP117.

BC F2 progeny were analyzed to test whether the JP117 mutation segregates as a monogenic Mendelian trait. The ratio of di- to mono-Me-M (di-Me-M/ mono-Me-M) was used as a diagnostic phenotype. Of 23 BC F2 plants, 19 (82.6%) showed similar profiles

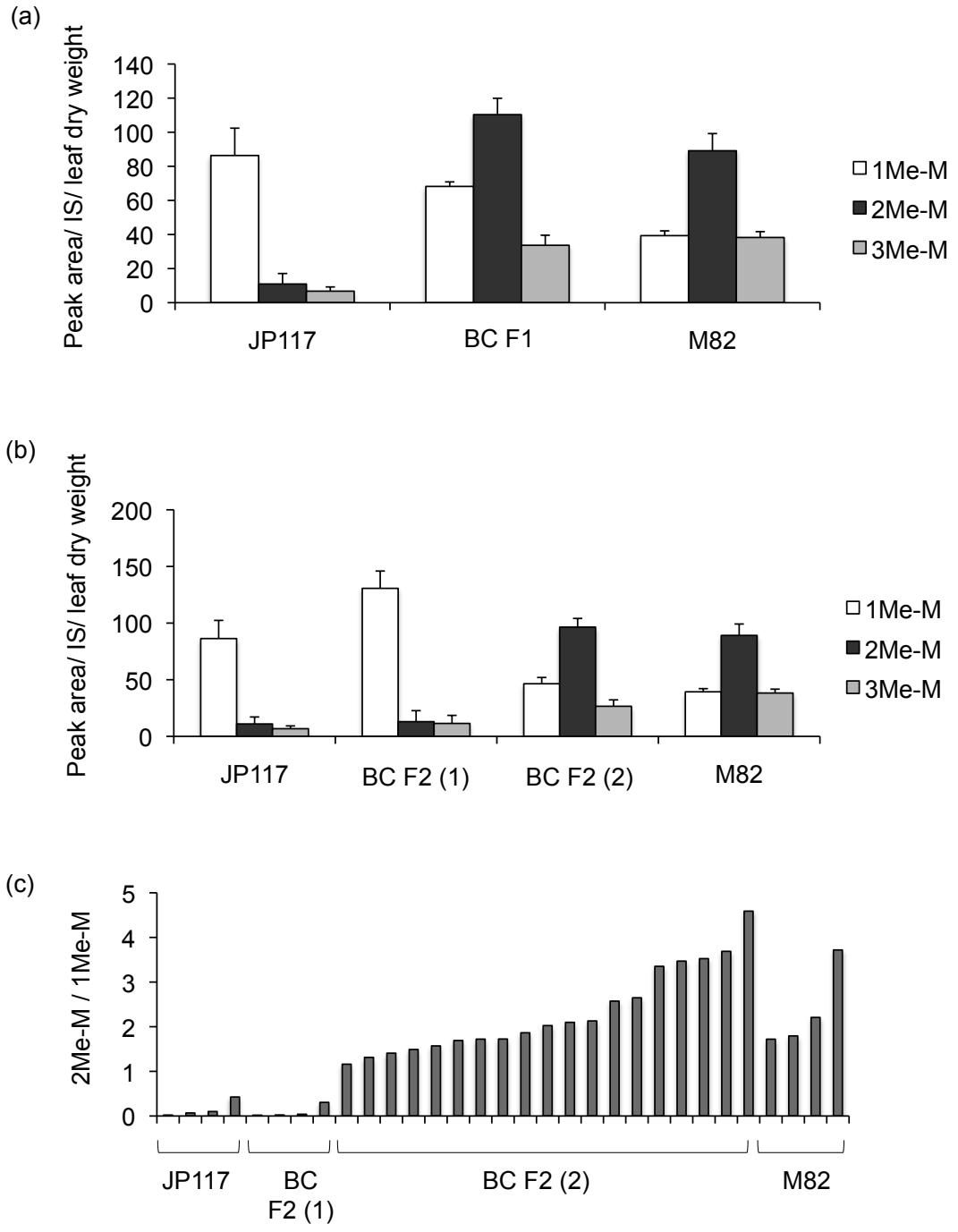


Figure 3.4. Evidence that a single recessive mutation is responsible for the JP117 mutant phenotype.

Figure 3.4. Contined

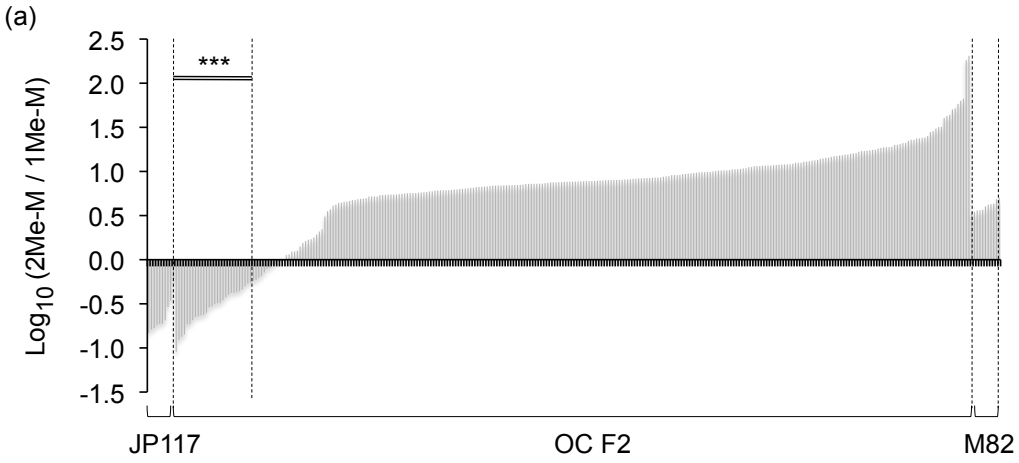
(a) Me-M profiles in trichomes of M82 X JP117 BC F1 plants are generally similar to those of M82, with levels of 1Me-M intermediate between M82 and JP117 mutant. Four replicate plants were assayed for each genotype.

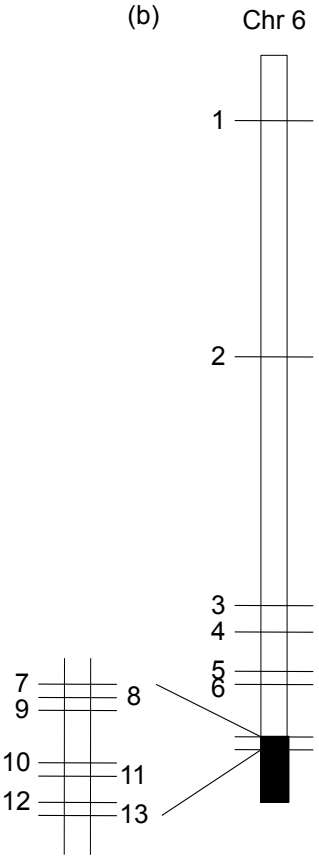
(b) Averages of classes of F2 BC plants. Of 23 BC F2 plants, four labeled as BC F2 (1) predominately accumulated 1Me-M, while 19 labeled as BC F2 (2) 2Me-M. The grouping of BC F2 plants was based on the individual chemical profiles from (c). Four replicate plants were tested for JP117 and M82 respectively.

(c) The ratio of di- to mono-Me-M (2Me-M/1Me-M) is shown for individual plants tested in (b). Although BC F2 plants showed continuous values for the ratio, the 2Me-M/1Me-M ratios of the 4 plants from BC F2 (1) were consistently as low as the JP117 mutant and the ratios of 19 plants from BC F2 (2) were as high as M82.

as M82, while four samples (17.4%) resembled JP117 mutant (Figure 3.4bc). Although the total number of progeny tested is low, Chi-square test revealed that these numbers fit the segregating ratio expected for monogenic Mendelian trait at p-value of 0.05. These results suggested that the JP117 mutation is recessive and segregates as a monogenic recessive trait.

To test whether JP117 behaves as a single locus, genetic mapping was performed (Jander *et al.* 2002). Because the genetic diversity of *S. lycopersicum* is relatively low (Rodriguez-Saona *et al.* 2011), a polymorphic mapping population was obtained by crossing JP117 to the wild species *S. pennellii* LA0716. *S. pennellii* LA0716 was chosen because it is sexually compatible with the cultivated tomato and there are large numbers of DNA polymorphisms between the two species. Chemical profiles were analyzed for 289 outcross (OC) F2 plants by LC-ToF-MS. Because no Me-Ms were detected in LA0716 (data not shown), Me-M profiles of OC F2 were compared with those of JP117 and M82. The ratio of di- to mono-Me-M was highly variable in the OC F2 mapping population, ranging from 0.09 to 203.97 (corresponding to -1.05 and 2.31 respectively on a $\log_{10}$ scale, as presented in Figure 3.5a). The ratio was noticeably lower in the JP117 mutant than M82 despite the value varied among replicate plants for each genotype. The ratio of di-Me-M/mono-Me-M ranged from 0.15 to 0.60 in JP117 mutant and 3.45 to 5.01 in M82 (-0.82 to -0.22 and 0.54 to 0.70 respectively with $\log_{10}$ scale in Figure 3.5a), revealing that even the highest ratio from JP117 was lower than the lowest from M82. The highest ratio among the JP117 mutant replicates 0.60 (-0.22 with log10 scale in





	[1]Marker	[2]Genetic Location (cM)	[3]Physical Location (Mbp)	[4]No. of plants with *S. lycopersicum* homozygous allele
1	U231369	12.7	-	2
2	At1g22850	46.0	-	8
3	B	74.5	-	19
4	At1g29900	80.7	-	21
5	sc06019	-	44.24	21
6	sc03622-1	-	44.30	21
7	sc03622-2	-	44.52	27
8	sc03622-3	-	44.52	28
9	***sc03622-4***	-	***44.58***	***28***
10	***sc03622-5***	-	***44.83***	***30***
11	***sc03622-6***	-	***44.83***	***30***
12	***sc03622-7***	-	***45.00***	***30***
13	***U146140***	***97.2***	***45.08***	***30***

Figure 3.5. The JP117 mutation maps to the bottom of chromosome 6.

Figure 3.5. Contined

JP117 was crossed to *S. pennellii* LA0716 and the F2 progeny phenotypically and genotypically analyzed.

(a) The ratio of 2Me-M/1Me-M is shown for 289 individual F2 plants, as well as the JP117 mutant and M82 controls. Due to the great variation seen in the interspecific cross, the ratio is presented as a log scale. Using a cutoff of -0.22, the highest value among the replicates for the JP117 mutant, identified 30 OC F2 with the JP117 mutant phenotype: these are labeled with *** and used for genetic mapping in (b).

(b) Results of genotyping of the 30 OC F2 plants selected from (a) localizes the JP117 mutation within the 1.5Mb intervals from the bottom of chromosome 6. Thirteen genetic markers for chromosome 6 used for genetic mapping[1] (See 'Materials and Methods' for the details). [2]Genetic locations of the markers were obtained from the Sol Genomics Network (SGN; http://solgenomics.net). [3]Physical locations were determined based on tomato whole genome sequence (http://solgenomics.net/, SL2.31 chromosome sequence). '-' indicates no information was available at SGN. [4]The number of plants with *S. lycopersicum* homolozygous alleles (LL) was also shown for the individual markers. All 30 plants showed LL genotype when tested with markers 10 - 13.

Figure 3.5a) was used as a cutoff to identify OC F2 plants having the mutant phenotypes and 30 F2 plants were selected.

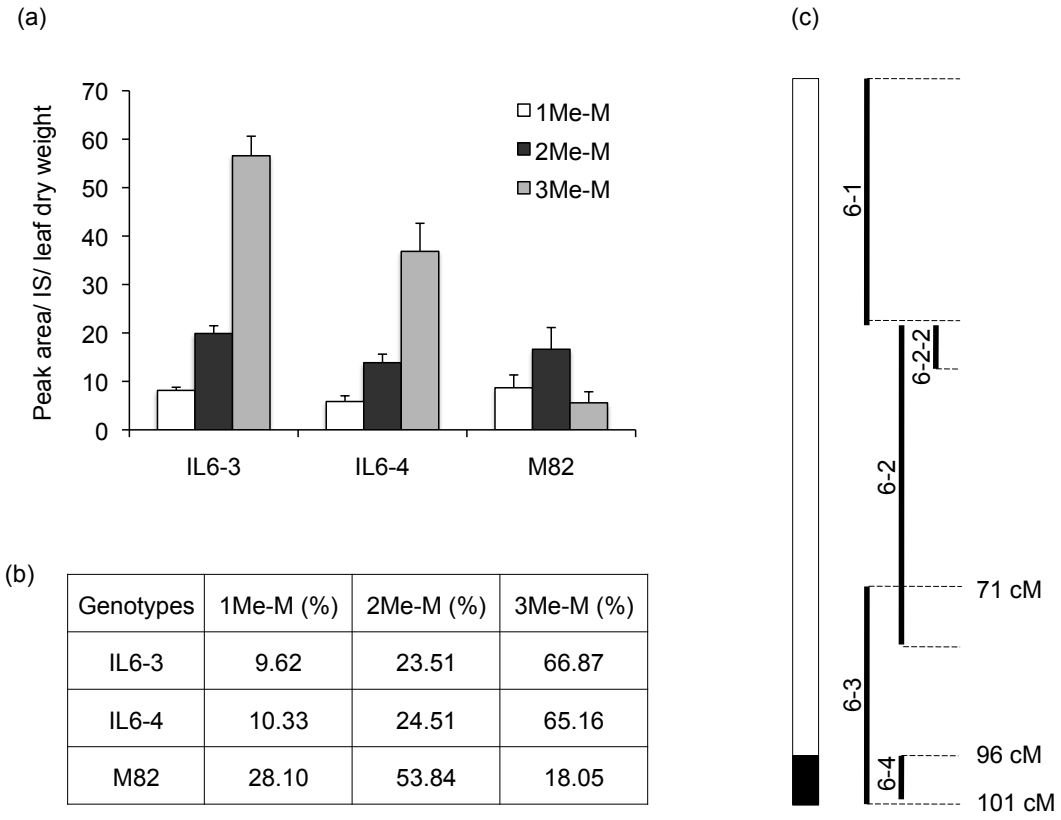
Thirty OC F2 plants that met the criteria for having the JP117 mutant phenotype were used for genetic mapping (Figure 3.5). First-pass mapping results with genetic markers throughout the Solanum genome (the SGN, http://solgenomics.net/) suggested that the JP117 mutation is located on chromosome 6 (data not shown). Fine mapping was performed with polymorphic genetic markers throughout chromosome 6 from the SGN and others generated in this study (Figure 3.5b). Genotyping with markers closely linked to the mutation are expected to show a preponderance of *S. lycopersicum* homozygous alleles (LL), rather than heterozygous (LP) or *S. penellii* homozygous alleles (PP). In fact, the number of plants with LL alleles increased as genotyping was performed with genetic markers progressing toward the bottom of chromosome 6. All 30 F2 plants revealed LL genotype with markers 10 ~ 13, indicating that the JP117 mutant gene is closely linked to these markers (Figure 3.5b, Table 3.1). The physical locations of markers 8 and 9 and the end of chromosome 6 are known to be 44.52, 44.58 and 46.04 Mbp, respectively (http://solgenomics.net, Figure 3.5b). Taken together, these genetic mapping results indicate that the JP117 gene is located within the last 1.5 Mbp at the bottom of chromosome 6.

Phenotypic characterization of altered Me-M profiles in IL6-3/4

As a parallel project aimed at identification of genes involved in SGT metabolism, LC-ToF-MS analysis of *S. pennellii* introgressions (ILs) was performed to identify regions of

Table 3.1. The list of genetic markers generated in this project for JP117 genetic mapping. Primer sequences and restriction enzymes are shown for individual markers used for JP117 mutant genetic mapping. See 'Materials and Methods' for details about genotyping procedures.

Marker Name	Primer Type	Primer Sequence (5'-3')	Restriction Enzyme	Expected product size (bp)	
				L	P
sc06019	F	AAAGGAAAAGGAGC AGAAGTTAAAGC	EcoRV	320, 386	706
	R	TTGGGTCAGATTTAA CATTTTTCTCACG			
sc03622-1	F	TCCGAGTGAATGGA ACGC	PsiI	131, 203	334
	R	TTAGTAGAAAAGTAC AGTGTAAAATAGTTG			
sc03622-2	F	ATGAGCTTAGTTGGA GCATTATGG	MspI	1079	538, 541
	R	GAAGGAAGAAGACG ACCAAAGAAG			
sc03622-3	F	TATAATTTGATTTGA CATTTGTGTTGCTT	ApoI	196, 266	102, 164, 196
	R	CTATGGTCCATTTAA ACATAAACGTTC			
sc03622-4	F	ATGACCGTGGTGTC CAGG	ApoI	161, 398	559
	R	CCAAAATAAGCAGA GATGGCTTG			
sc03622-5	F	ACCTCACTTGTCATT TGCCAC	HaeIII	297, 878	1175
	R	TTCTGATTGTCGCCG GAAAC			
sc03622-6	F	GAAACTGAACATCTT GATCCAAAAGATT	HaeIII	282, 783	1065
	R	CACCTATACCTTCTC CTCCAGCC			
sc03622-7	F	ATGGATAATGACTGC TTCTCTAATGG	PvuII	428, 445	873
	R	ATCATCTTCATTGCCC TTTTGTTC			



(b)

Genotypes	1Me-M (%)	2Me-M (%)	3Me-M (%)
IL6-3	9.62	23.51	66.87
IL6-4	10.33	24.51	65.16
M82	28.10	53.84	18.05

Figure 3.6. Higher accumulation of 3Me-M in IL6-3/4 plants and a genetic map of chromosome 6 ILs.

(a) IL6-3 and 6-4 consistently revealed increased accumulation of 3Me-M. Four replicate plants were tested for each genotype.

(b) The percentage composition of individual Me-Ms to total Me-Ms was calculated based on LC-ToF peak area. While 2Me-M was the most abundant signal, representing 54% of total Me-Ms in M82, IL6-3/4 predominately accumulated 3Me-M with up to 65~67% of total Me-M peak area.

(c) The genetic map of IL6-3/4 (Eshed and Zamir 1994, Eshed and Zamir 1995) reveals that IL6-4 is contained within the IL6-3 region.

the *S. pennellii* genome that alter *S. lycopersicum* M82 metabolite profiles (Schilmiller *et al.* 2010a). ILs 6-3 and 6-4 caused strong increases in tri-Me-M accumulation compared with M82 (Figure 3.6a). These two ILs have chromosomal substitutions at the bottom of chromosome 6 (Figure 3.6c), with IL6-4 completely contained within IL6-3 (Eshed and Zamir 1994, Eshed and Zamir 1995). In contrast to M82, which accumulated only up to 18% of total Me-Ms as tri-Me-M, IL6-3/4 trichomes accumulated tri-Me-M up to 65~67% of total (Figure 3.6b). The consistent phenotypes in the two overlapping ILs indicates that gene(s) responsible for differences in trichome myricetin methylation are located between 96-101 cM at the bottom of chromosome 6 (Figure 3.6c).

To complement the negative ion mode LC-ToF-MS analysis of single leaf dip extracts, positive ion mode LC-MS/MS analysis of individual trichome types was employed. The LC-MS/MS approach also revealed higher accumulation of tri-Me-Ms in IL6-3/4 trichomes than M82 confirming IL6-3/4 phenotypes identified by the LC-ToF-MS method (Figures 3.6 and 3.7). Furthermore, the LC-MS/MS approach expands our understanding of IL6-3/4 Me-M profiles beyond what was learned from the LC-ToF-MS analysis. For example, this analysis distinguished two isomers of tri-Me-M: the more abundant species (3,3',7) is seen in all genotypes while a minor one (3,3',5') is found in IL6-3/4 samples. Furthermore, the tetra-Me-Ms (3,3',5',7) was detected in the IL6-3/4 samples but not in M82 trichomes, or in any sample under negative ion mode LC-ToF-MS analysis of single leaf dip extracts.

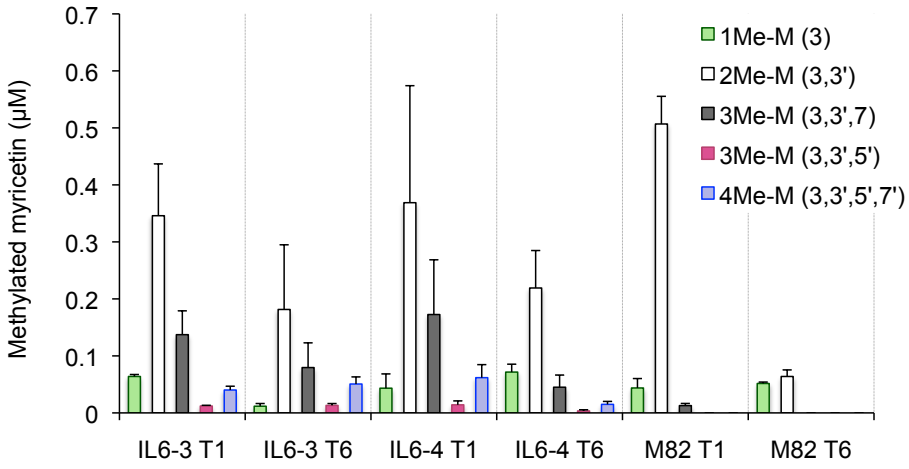


Figure 3.7. The Me-M levels measured in isolated glands of types 1 and 6 trichomes from IL6-3/4 and M82 plants.

LC-MS/MS analysis with trichome extracts revealed Me-M profiles including the position of methyl groups for individual Me-Ms. In contrast to M82, which shows the enrichment of Me-Ms in type 1 glands, IL6-3/4 reveals relatively high accumulation of Me-Ms in type 6 glands as well as in type 1. While three Me-Ms were detected from M82 including 1Me-M, 2Me-M and 3Me-M (3,3',7), two additional Me-Ms, 3Me-M (3,3',5') and 4Me-M (3,3'7,5') were also found in IL6-3/4 glands. See Figure 3.2 for the numbering system. Compared to M82, IL6-3/4 accumulate higher amount of polymethylated Me-Ms such as 3Me-Ms and 4Me-M. The amount of methylated myricetin (µM) is presented with the average and standard error. A sample of fifty glands of specific types of trichomes was assayed from three independent plants.

Genetic analysis of IL6-4

To ask whether the *S. pennellii* IL6-4 locus conditions increased myricetin methylation, M82 X IL6-4 F1 and F2 progeny plants were analyzed. The three F1 progeny exhibited increased tri-Me-M reminiscent of the IL6-4 parent, consistent with the hypothesis that the *S. pennellii* allele is dominant to M82 (Figure 3.8a). F2 progeny analysis gave results consistent with the hypothesis that the *S. pennelli* IL6-4 high tri-MeM is caused by a single dominant gene or multiple linked genes; nine of 12 plants showed the same Me-M profiles as IL6-4, and the other three F2 plants resembled M82 (Figure 3.8b and c).

The search for candidate genes influencing myricetin methylation in JP117 and IL6-4

Analysis of the JP117 EMS mutant and IL6-3/4 ILs indicated the presence of genes that influence myricetin methylation at the bottom of chromosome 6 in *S. lycopersicum* M82 and *S. pennellii* LA0716 (Figure 3.9ab). A simple two-part hypothesis is that: 1. JP117 is mutated in an *O*-methyltransferase (OMT) that normally functions in conversion of mono-Me-M to di-Me-M; 2. the *S. pennellii* IL6-4 region contains an OMT that efficiently converts di-Me-M to tri-Me-M and even tetra-Me-M, which is absent from or non-functional in M82 (Figure 3.9).

The map positions of JP117 and IL6-4 are consistent with the hypothesis that there are two or more OMTs near the bottom of chromosome 6. During the time that this project was being conducted, progress was made in sequencing and assembly of the *S. lycopersicum* Heinz 1706 variety genome (http://solgenomics.net, SL2.31 chromosome

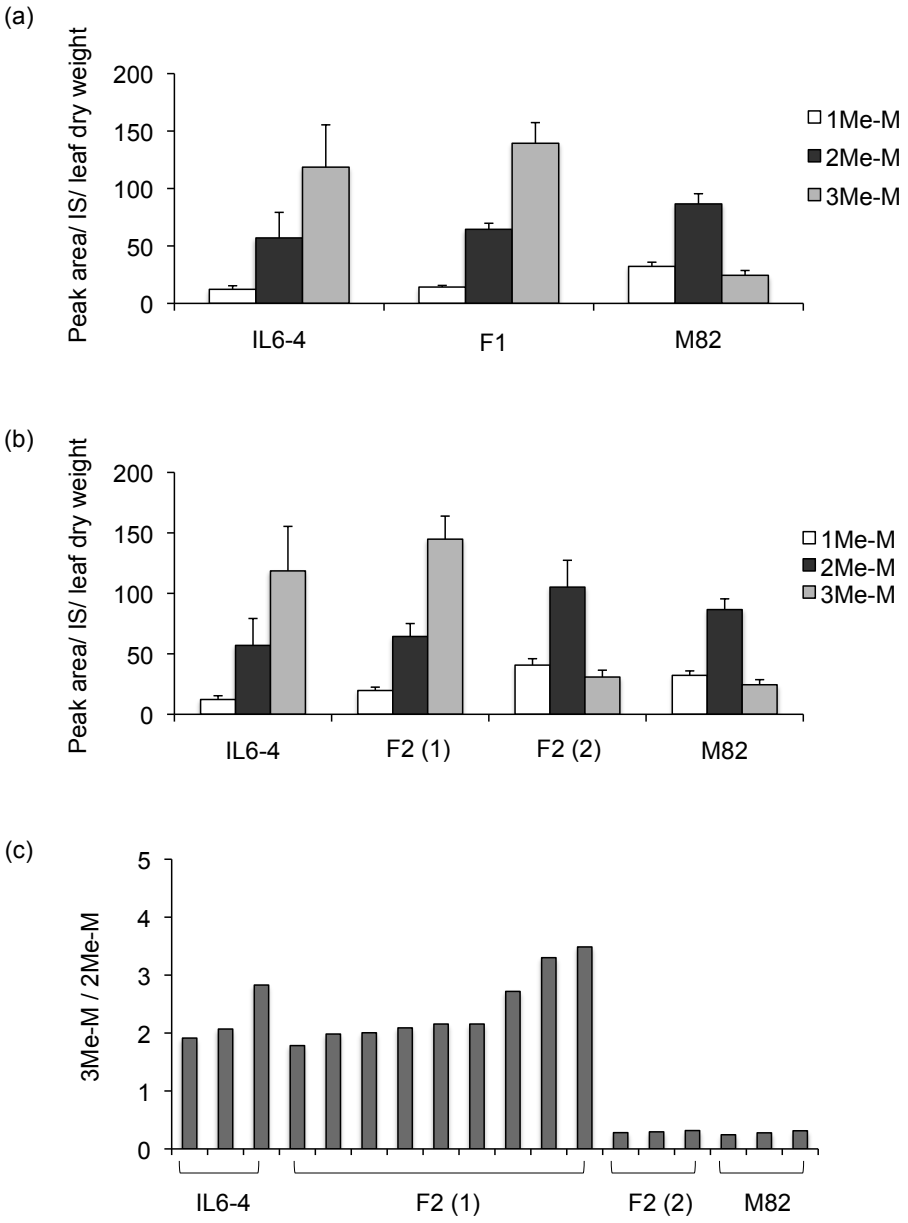

(a)
Peak area/ IS/ leaf dry weight
200
150
100
50
0
1Me-M
2Me-M
3Me-M
IL6-4
F1
M82
(b)
Peak area/ IS/ leaf dry weight
200
150
100
50
0
1Me-M
2Me-M
3Me-M
IL6-4
F2 (1)
F2 (2)
M82
(c)
3Me-M / 2Me-M
5
4
3
2
1
0
IL6-4
F2 (1)
F2 (2)
M82

Figure 3.8.

Figure 3.8. Chemical profiles of IL6-4 genetic cross plants suggest that the *S. pennellii* allele is dominant for IL6-4 chemical phentoypes and that the phenotypes segregate genetically.

(a) Me-M profiles in M82 X IL6-4 F1 plant leaf dips are similar to those of IL6-4. Three replicate plants were tested for each genotype.

(b) The IL6-4 chemical phenotypes segregate in F2 plants. Of 12 F2 plants, 9 labeled as F2 (1) revealed the IL6-4 phenotype, while 3 labeled as F2 (2) show the M82 phenotype. The grouping of BC F2 plants is based on the individual chemical profiles from (c). Three replicate plants were tested for JP117 and M82.

(c) The ratio of tri- to di-Me-M (3Me-M/2Me-M) is shown for individual plants tested in (b). The 3Me-M/2Me-M ratios of the nine plants from F2 (1) were similarly high as IL6-4 and those of the three plants from F2 (2) are as low as M82.

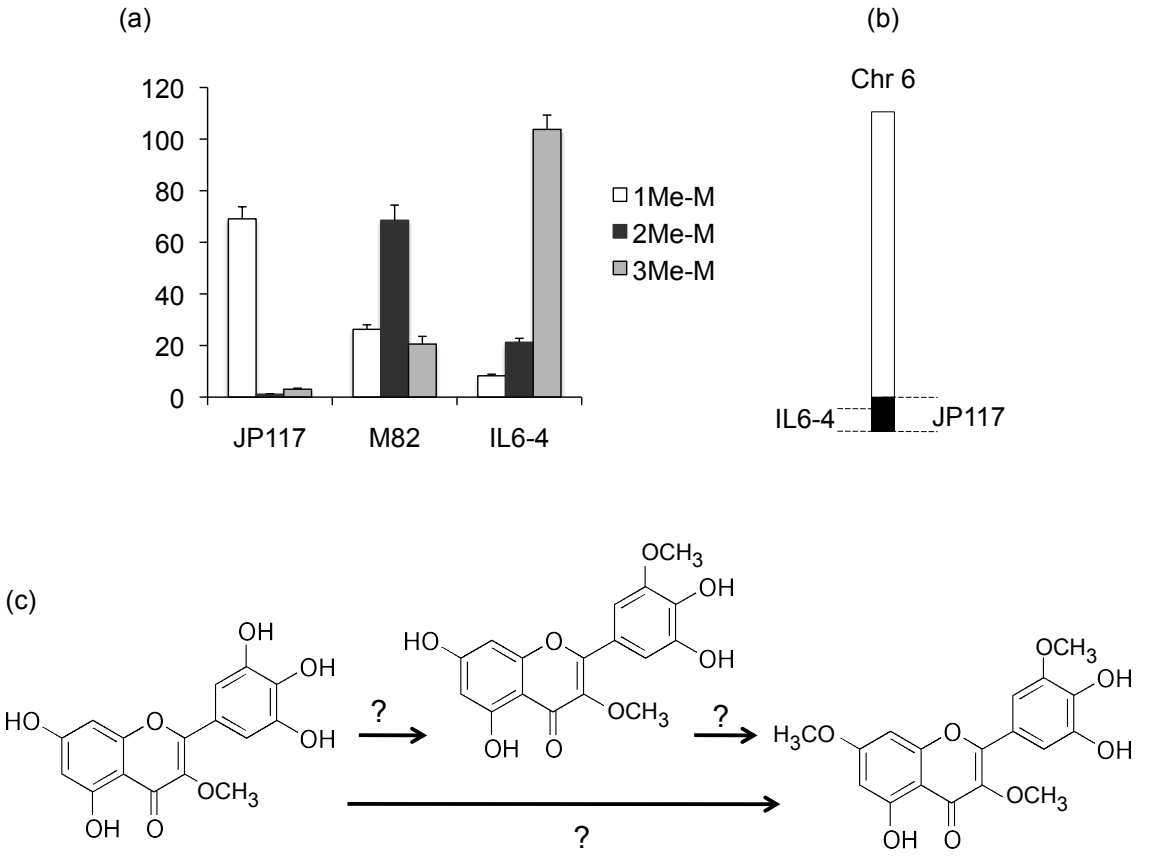


Figure 3.9. Summary of the chemical phenotypes and chromosomal locations of the JP117 mutant and the IL6-4 region.

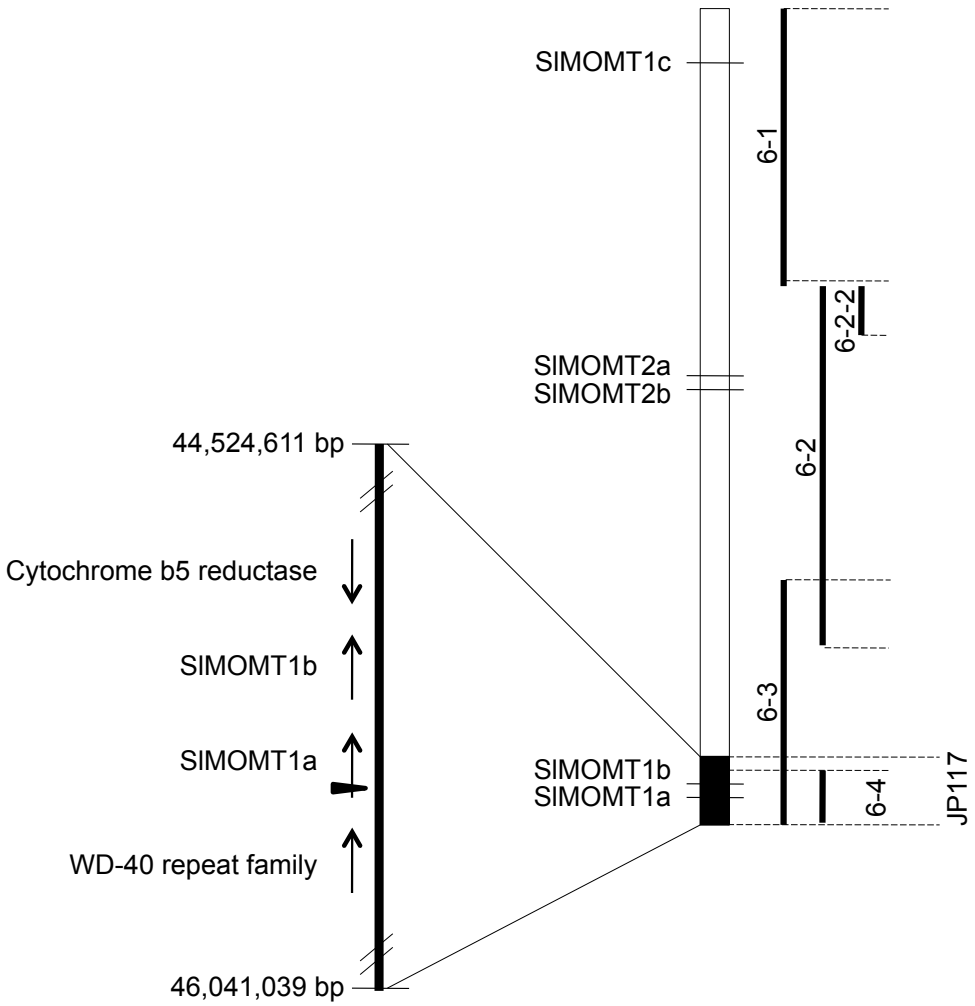
(a) While the JP117 mutant revealed higher accumulation of 1Me-M and reduced total Me-M amounts, IL6-4 plants showed increased tri-Me-M accumulation.

(b) The regions associated with JP117 mutant and IL6-4 genes overlap at the bottom of chromosome 6.

(c) Proposed pathways for the production of Me-Ms in *S. lycopersicum* variants. The indicated positions of methyl groups on Me-Ms are based on the LC-MS/MS analysis performed by Chao Li. The structure of 3Me-M (3,3',7) is shown as it is the most abundant isomer of 3Me-M (See Figure 3.7 for the details).

sequence) and shotgun sequence data became available for *S. pennellii* LA0716. These assemblies were searched for sequences homologous to two trichome-specific myricetin OMTs from *S. habrochaites* LA1777 (shMOMT1 and 2) (Pearson and Lipman 1988, Schmidt *et al.* 2011, Vingron and Waterman 1994). Five *S. lycopercisum* MOMTs were discovered to have high sequence identity to shMOMTs from chromosome 6, including three genes for shMOMT1 and two for shMOMT2; these were named slMOMT1a, 1b and 1c and slMOMT2a and 2b, respectively (Figure 3.10). *S. pennellii* OMT genes were also identified by BLAST analysis from shotgun DNA sequence using shMOMTs and slMOMTs sequences. Despite the relatively short genomic sequence assemblies for this species, six homologous genes were identified: spMOMT1a, 1b, 1c, 2a, 2b and an additional gene spMOMT1aa (Figures 3.11 and 3.12). All sl- and spMOMTs contained characteristic plant OMT domains (Ibrahim *et al.* 1998).

The specific position of each slMOMT genes on chromosome 6 was identified using genome assembly and confirmed by genetic mapping. The *S. lycopersicum* genome assembly (http://solgenomics.net/, SL2.31 chromosome sequences) contained slMOMT1a and 1b at the bottom of chromosome 6, in the region containing the JP117 mutation and IL6-4 (Figure 3.10). In contrast, slMOMT1c, slMOMT2a and 2b were located on the top and middle of chromosome 6 respectively. Because the tomato genome sequence was found to have mistakes for the terpene synthase gene family, including genes annotated on incorrect chromosomal regions and hybrid assemblies (Falara *et al.* 2011), each slMOMT gene was genetically mapped. This was done by identifying a polymorphic sequence for each gene and looking for the presence of the *S. pennellii*



Gene ID	Chromosome Location (bp)
SlMOMT1c	1,816,555 - 1,820,094
SlMOMT2a	36,549,479 - 36,551,598
SlMOMT2b	36,559,878 - 36,561,067
Cytochrome b5 reductase	45,134,673 - 45,136,744
SlMOMT1b	45,137,707 - 45,140,902
SlMOMT1a	45,144,339 - 45,145,525
WD-40 repeat family	45,159,823 - 45,158,037

Figure 3.10.

Figure 3.10. Search for candidate *O*-methyltransferases (OMTs) based on sequence homology.

Five myricetin OMT (MOMT) homologous genes were identified on chromosome 6 of *S. lycopersicum* and named as slMOMT1a, 1b, 1c, 2a and 2b. Two slMOMTs, slMOMT1a and 1b, were identified within the target region for JP117 and IL6-4. The chromosomal locations of five slMOMTs were obtained from the tomato whole genome sequence (http://solgenomics.net/, SL2.31 chromosome sequence) and confirmed by genetic mapping. Sequence truncation was found at 5' end of slMOMTa (See Figure 3.11 for details).

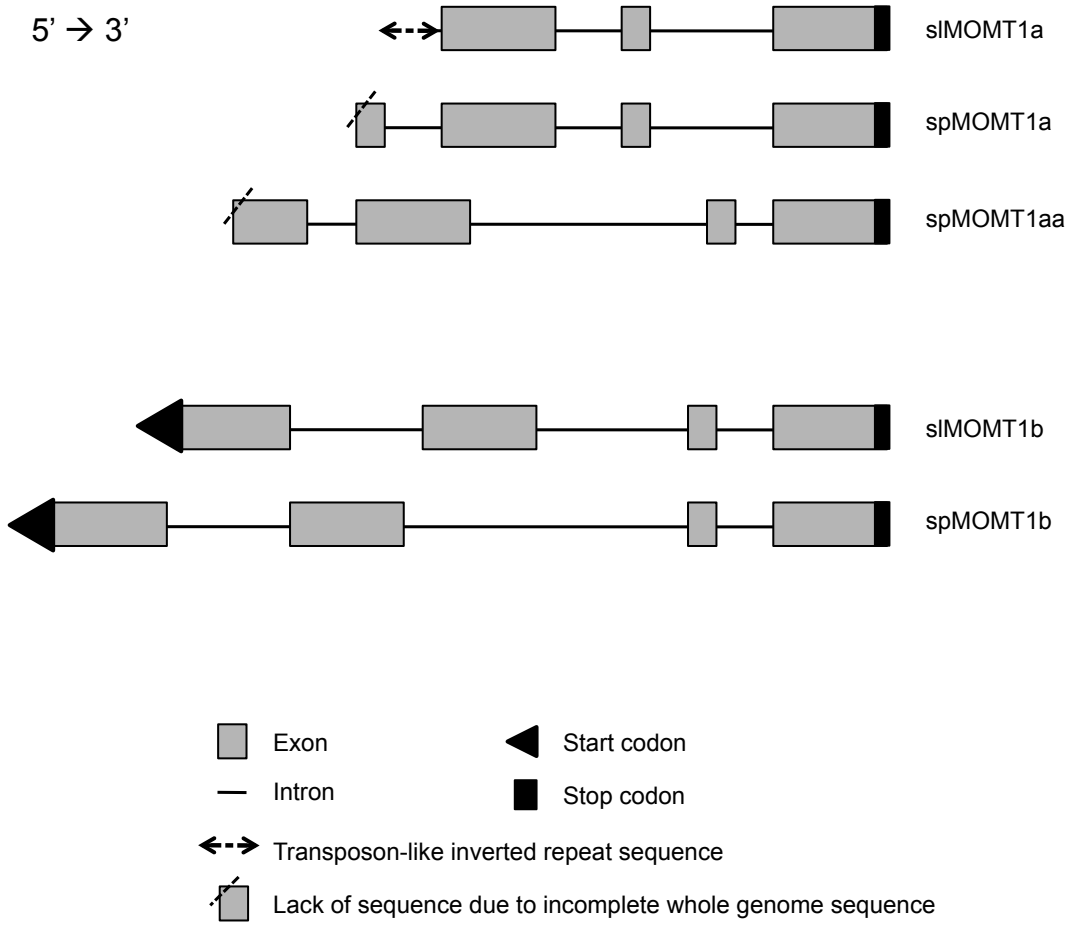


Figure 3.11. Predicted gene structures of slMOMT1a and 1b and the corresponding *S. pennellii* homologs.

Gene models were predicted for slMOMT1a, 1b and *S. pennellii* homologous genes using the Softberry computer program (www.softberry.com). slMOMT1a, spMOMT1a and spMOMT1aa are predicted to have similar gene structures, except for a truncation at the 5' end of slMOMT1a. BLAST analysis with the upstream sequence of this truncated region revealed transposon-like sequences at the 5' end.

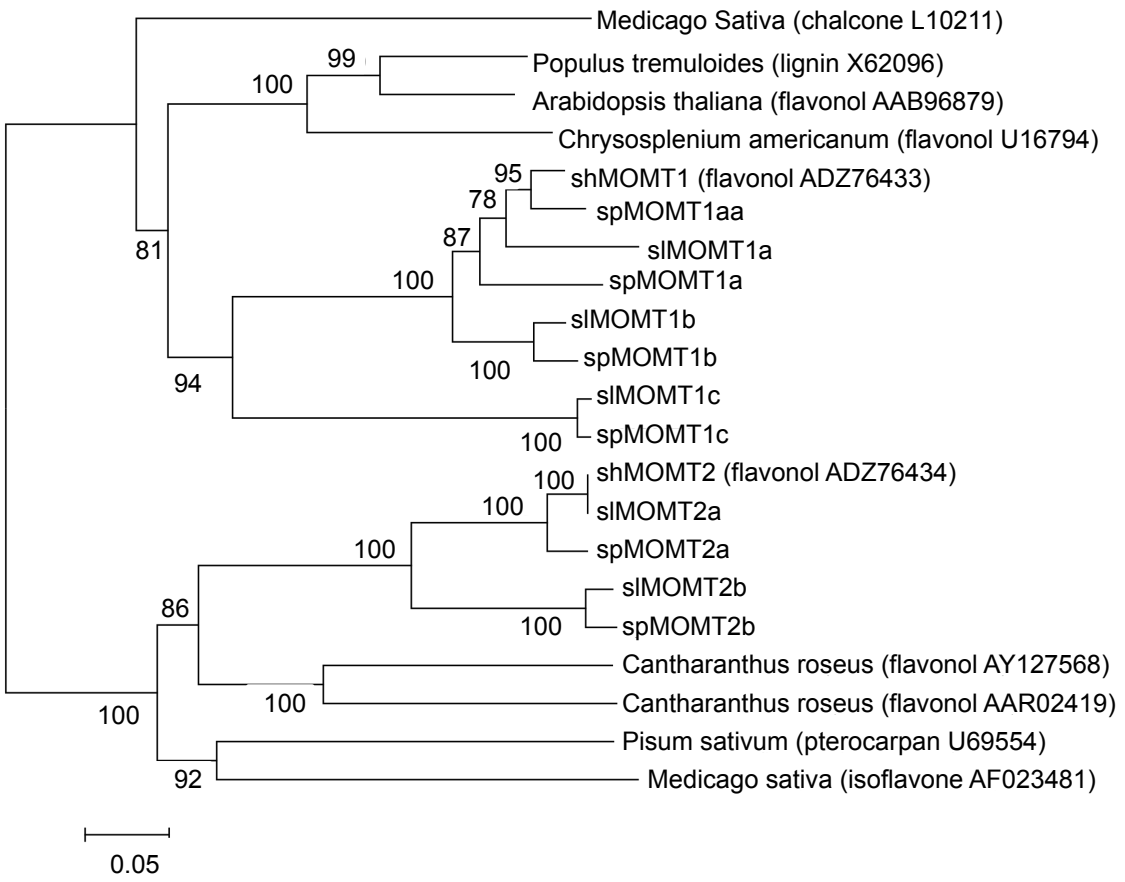


Figure 3.12. Evolutionary relationships among various plant OMTs.

Protein sequence was obtained for previously characterized plant OMTs. The detailed information is presented for individual OMTs including the name of plant species from which the OMT was characterized, their preferred substrate identified from biochemical experiments and accession number of the protein sequence. The phylogeny tree was generated by the neighbor-joining method (Saitou and Nei 1987). Bootstrap testing (Felsenstein 1985) was performed to validate the reliability of the tree with n=500 and the result is shown next to each branch. MEGA5 was used to generate trees and for statistical tests, (Tamura et al. 2011). As expected, slMOMTs and spMOMTs cluster together with shMOMTs. They were also closely related with other OMTs that use flavonols as preferred substrates.

LA0716 polymorphism on the four ILs covering chromosome 6 (IL6-1, 2, 3 and 4 in Figure 3.10). In each case the *S. pennellii* polymorphism was found associated with the expected IL (Table 3.2, Figure 3.10).

Gene structures were predicted for slMOMT1a, 1b and the corresponding to *S. pennellii* homologs using genomic DNA sequence obtained from the BLAST analysis (Figure 3.11). slMOMT1a, spMOMT1a and spMOMT1aa are predicted to have high similarity with the exception that slMOMT1a has a 5' end deletion. This truncation would result in the complete loss of the first exon, while spMOMT1a and 1aa are predicted to encode functional proteins. BLAST analysis was performed with 10 kb upstream sequence from the beginning of the second exon of slMOMT1a and a transposon-like repeated inverted sequence was found at the 5' end of this gene (Figure 3.11).

To test the relationship of slMOMTs with previously characterized OMTs, a phylogenetic tree was generated with plant OMT protein sequences (Figure 3.12). As expected, slMOMTs were closely related to shMOMTs and spMOMTs. In addition, slMOMT genes showed closer relationships with OMTs preferentially using flavonol substrates.

Testing the roles of Me-Ms in plant-herbivorous insect interactions

Trichome metabolites are known to improve plant defense against herbivorous insects by affecting insect feeding or attracting predators of herbivore insects (Kang *et al.* 2010a, Weinhold and Baldwin 2011). To test the hypothesis that altered trichome Me-M profiles affect plant-insect interactions, an insect feeding assay was performed with JP117

mutant, M82 and IL6-4. The *odorless-2* (*od-2*) mutant was used as a control because it lacks the major trichome metabolites, including mono-Me-M, and is more susceptible to insects compared to its isogenic wild type CastleMart (CM) (Kang *et al.* 2010a).

Because we did not have predictions about which, if any, insects are affected by Me-Ms, Field experiments were chosen to ask whether herbivores found in mid-Michigan would respond differentially to the genotypes tested. A Solanaceous specialist *Manduca sexta*, tobacco hornworm, was found on plants grown in the field in early September 2010 (Figure 3.13a). The JP117 mutant showed strong damage by the tobacco hornworm *Manduca sexta*, whereas IL6-4 was least attacked by *M. sexta* with M82 damaged at a level intermediate between JP117 and IL6-4 (Figure 3.13b). The number of insects on plants was measured at a single time point. The largest number of *M. sexta* individuals was found on the JP117 mutant and the lowest on IL6-4 (Figure 3.13c). This result suggests that the degree of myricetin methylation is inversely correlated with hornworm herbivory. The plant damage by *M. sexta* was higher in the *od-2* mutant compared to the CM wild type, consistent with published results (Kang *et al.* 2010a). No insects were found on *od-2* or CM possibly due to the small population (n=2 for both genotypes).

Given that *M. sexta* feeding varied on JP117, M82 and IL6-4 from the field experiment, the feeding assay was repeated under green house conditions. The weight of *M. sexta* larvae fed on different genotypes was measured. Of 60 larvae fed on JP117, M82 and IL6-4, 17, 11 and 9 larvae survived, respectively (Figure 3.14a). Also, insect weight was approximately 2 fold higher in JP117 than M82. The larvae weight was slightly higher in

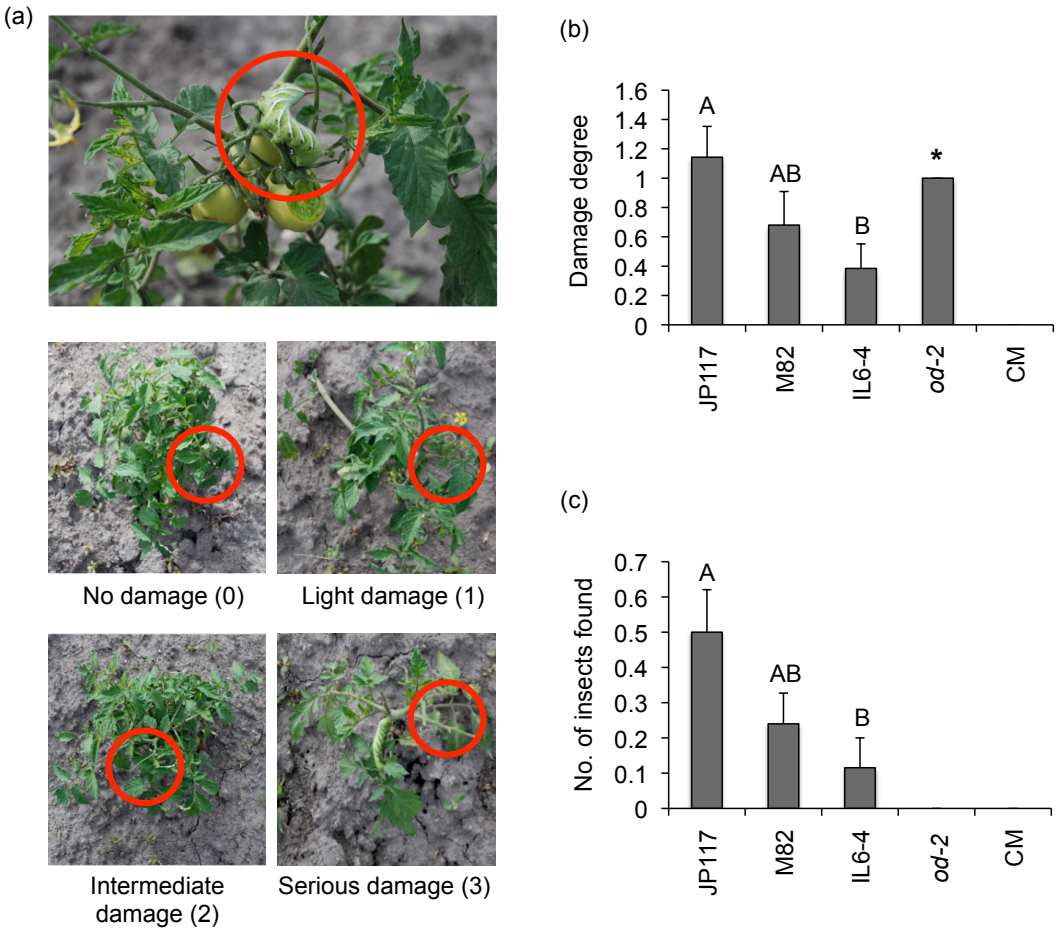

(a)
No damage (0)
Light damage (1)
Intermediate damage (2)
Serious damage (3)
(b)
Damage degree
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
A
AB
B
*
JP117
M82
IL6-4
od-2
CM
(c)
No. of insects found
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
A
AB
B
JP117
M82
IL6-4
od-2
CM

Figure 3.13

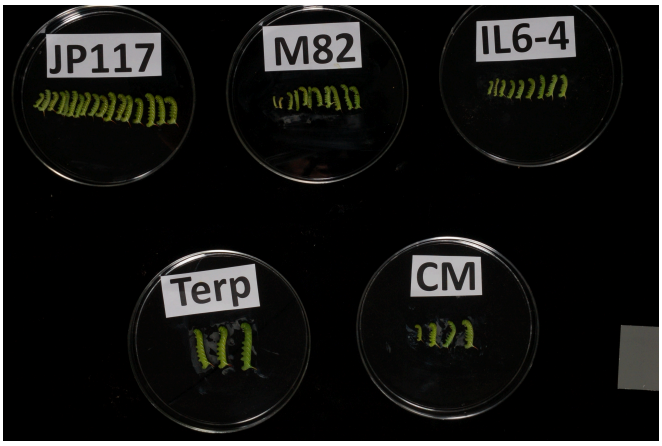
Figure 3.13. Field experiments revealed altered *Manduca sexta* feeding on *S. lycopersicum* variants.

(a) Representative photographs showing that *M. sexta* caused various degrees of plant damage to JP117 mutant, M82 and IL6-4 plants. Plant damage was scored from values of 0 to 3, depending on how much plant material was lost to insect herbivory. See 'Materials and Methods' for details.

(b) *M. sexta* feeding caused the most serious damage on the JP117 mutant and the least on IL6-4. One-way ANOVA revealed that the variation of plant damage among JP117, M82 and IL6-4 is statistically significant (p-value = 0.013). Different lower case letters in the bar graph indicate statistically different groups. Because the *odorless-2* (*od-2*) mutant and its wild type Castlemart (CM) control plants are not in the M82 genetic background, they were not included in the ANOVA test. The identical system is used in subsequent figures describing the ANOVA test results, unless specified otherwise. As previously reported (Kang et al. 2010a), *od-2* mutant plants were significantly more susceptible than CM (p-value from student's t-test < 1 x 10-4). Plant damage is presented with average and standard error (n = 28, 25, 26, 2 and 2 for JP117, M82, IL6-4, ordorless-2 and CM, respectively).

(c) A single timepoint describing the number of *M. sexta* visiting plants. This analysis was performed with the same plants monitored in (b). The largest number of insects was observed on the JP117 mutant and the lowest on IL6-4 (p-value = 0.024 from one-way ANOVA test). No *M. sexta* was found on the *od-2* mutant or CM plants, possibly due to the small population number of plants and single assay time.

(a)



(b)

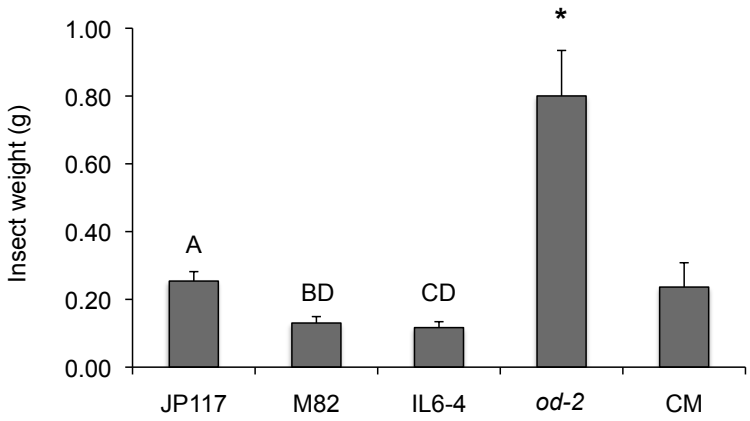


Figure 3.14.

Figure 3.14. Greenhouse analysis of *M. sexta* feeding.

(a) Sixty *M. sexta* larvae were placed on each genotype and living larvae collected at the end of 2 weeks from JP117, M82 and IL6-4 respectively. Larvae from JP117 mutant were bigger than those from M82 and IL6-4 plants. In addition, *M. sexta* from *od-2* (= *terp*) mutant were bigger than ones from CM as previously reported (Kang *et al.* 2010a).

(b) Insect weight for *M. sexta* larvae shown in (a). Larvae that fed on JP117 mutant were heaviest and on IL6-4 the lightest (p-value = 0.0005 from one-way ANOVA test). *M. sexta* from *od-2* mutant was significantly heavier than larvae from CM (p-value = 0.03 from *Student's t*-test), consistent with previously published data (Kang *et al.* 2010a). Larval weight is presented with average and standard error.

M82 compared to IL6-4 (Figure 3.14b). Larvae from the control plant *od-2* were approximately 3-fold heavier than those from CM, as previously reported (Kang *et al.* 2010a) (Figure 3.14b). The green house assay results were consistent with the observation from the field test and reinforce the idea that hornworm feeding is affected by altered Me-M profiles.

Hydroperiod and seasonality affect insect community structure by varying insect types and their density (Fontanarrosa *et al.* 2009, Stamp and Yang 1996) and the impact of plant specialized metabolites on herbivore insects is species-specific (Mitchell *et al.* 1993). To look for further evidence that insects native to mid-Michigan respond to differences in trichome myricetin methylation patterns, another independent field experiment was performed on plants transplanted to the field in early August 2011. In this field test, the flea beetle *Epitrix cucumeris,* was found to attack *S. lycopersicum* variants. As the flea beetle makes holes on plant leaves, the number of holes was counted to estimate plant damage (Figure 3.15a). The number of holes was approximately 2-fold higher in JP117 than M82, and M82 had a slightly larger number of damage sites than IL6-4 (Figure 3.15b). As expected (Kang *et al.* 2010a), a large difference in damage site numbers was seen between the control plants *od-2* and CM (Figure 3.15b). *E. cucumeris* feeding followed the similar trend as *M. sexta,* supporting the hypothesis that trichome Me-Ms composition has an impact on tomato defense against herbivore insects.

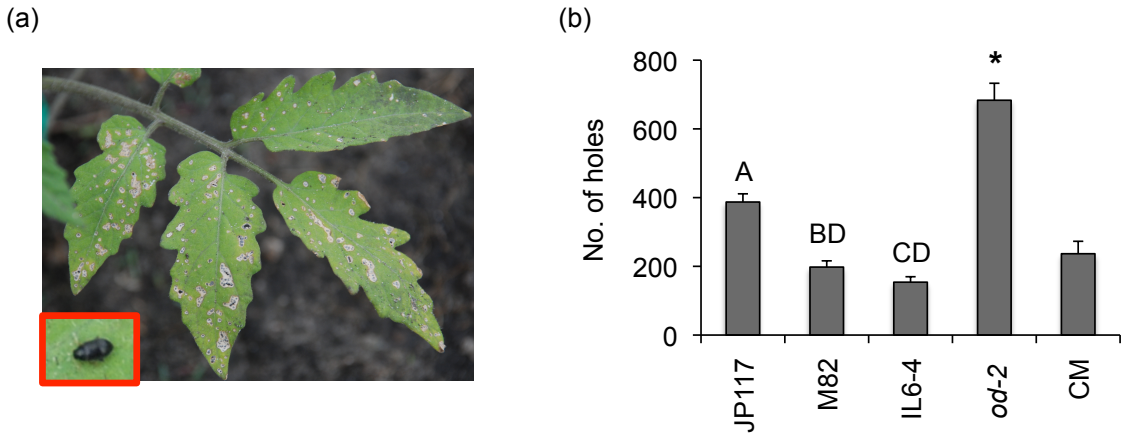


Figure 3.15. Field experiments reveal genotype-dependent differential *Epitrix cucumeris* damage.

(a) *E. cucumeris*, flea beetle, feeding left holes on leaf tissues.

(b) The number of holes was highest in JP117, followed by M82 and IL6-4 (p-value < 0.0001 from one-way ANOVA test). Control plants *od-2* mutant and CM showed dramatic difference (p-value from *Student's t*-test < 1 x 10-4), consistent with published data (Kang et al. 2010a). The number of holes is presented with the average and standard error (n = 28, 23, 18, 6 and 6 for JP117, M82, IL6-4, *od-2* and CM, respectively).

Discussion

The identification of altered Me-M phenotypes in the JP117 mutant and IL6-3/4 benefited from two factors: 1. high sensitivity of the LC-ToF-MS method, 2. the use of the ratios of differentially methylated myricetins (for examples, tri:dimethylated or di:monomethylated) as robust diagnostic phenotypes. Despite the low abundance of Me-Ms - < 1% of the most dominant acylsugar in M82 cultivated tomato (Schilmiller *et al.* 2010a) (Table 2.1 in Chapter 2) - the LC-ToF-MS method reproducibly measures Me-M levels.

The combination of negative mode LC-ToF-MS analysis and the single leaf dip extraction method was efficient for a fast chemical screen to identify genetic variants altered in trichome metabolites. However, positive mode LC-MS/MS analysis of individual trichomes was a useful complementary method for three reasons. First, the different ionization modes (i.e, positive and negative mode) provide distinct information. For glycosylated flavonols as an example, negative ion mode can identify the flavonol aglycone and the types of sugars attached to this flavonol core while positive ion mode MS can reveal information about the position of sugars on the flavonol ring structure (Tian *et al.* 2002). Second, LC-MS/MS analysis allows distinguishing positional isomers with identical m/z. Third, the analysis of isolated trichomes confirms that Me-Ms are trichome metabolites. As preparations of both type 1 and 6 trichomes were previously shown to contain Me-Ms (Schmidt *et al.* 2011), these two trichome types were tested in this study. Consequently, positive mode LC-MS/MS analysis of isolated trichomes characterized Me-Ms as trichome metabolites and generally validated JP117 and IL6-3/4

phenotypes identified from negative LC-ToF-MS analysis of single leaf extracts (Figures 3.1, 3.3, 3.6 and 3.7). Furthermore, LC-MS/MS analysis extended the LC-ToF-MS analysis by distinguishing two isomers of tri-Me-M and detecting a tetra-Me-M. LC-ToF-MS analysis of single leaf extracts is effective for fast screening, whereas LC-MS/MS analysis of the individual trichomes provides more in depth structural information such as the confirmation of isomers. Therefore, the combination of two complementary approaches was advantageous to reveal the Me-M phenotypes.

The comprehensive understanding of JP117 and IL6-3/4 Me-M phenotypes using two complementary analytical approaches leads to a hypothesis on the causes of altered Me-M profiles. In contrast to the JP117 mutant, which has decreased total Me-M levels as well as a higher ratio of mono- to di-Me-M compared to M82 (Figures 3.1 and 3.3), IL6-3/4 shows a strong shift toward higher accumulation of polymethylated myricetins such as tri-Me-Ms and tetra-Me-Ms without altering total Me-M level (Figures 3.6 and 3.7). In LC-MS/MS analysis of individual trichomes, total Me-M levels were only approximately 1.5-fold higher in IL6-3/4 than M82 and this was not statistically significant (p-value > 0.05 from *Student's t*-test) (Figure 3.7). LC-ToF-MS analysis of single leaf dip extracts also identified no significant alteration of total Me-Ms between IL6-3/4 and M82. In the first analysis (Figure 3.6), total Me-M levels were 2~3 fold higher in IL6-3/4 than M82 and the alteration was shown to be statistically significant only in IL6-3 (p-value = 0.005 and 0.09 from *Student's t*-test for IL6-3 and IL6-4 respectively) (Figure 3.6). Therefore, other independent sets of IL6-3/4 were analyzed and the alteration was not consistent (Data not shown). Higher accumulation of tri-Me-Ms in IL6-3/4 could result from the

activity of *S. pennellii* OMT allele(s) introgressed into M82 genetic background, which shifted Me-M profiles toward polymethylated myricetins without altering total Me-M levels. The appearance of tetra-Me-M (3,3',7,5') in LC-MS/MS analysis of IL6-3/4 samples could also be caused by conversion of the abundant tri-Me-Ms by an activity encoded on chromosome 6 or elsewhere in the genome. Furthermore, while M82 and JP117 revealed enrichment of Me-Ms in type 1 glands, IL6-3/4 accumulated high amount of Me-Ms in isolated type 6 glands as well as in type 1 (Figure 3.7). Although this could be from cross-contamination during harvesting individual types of trichomes due to rupture or leakage of glands, an exciting hypothesis is that it might represent a real shift in synthesis site for Me-M production in IL6-3/4.

Recent advance of tomato genomics assisted genetic approaches, allowing identification of two candidate genes, slMOMT1a and 1b, for the production of Me-Ms in tomato trichomes. Searching for the genes responsible for the phenotypes in JP117 and IL6-3/4 relying on only genetic mapping is challenging for two reasons. First, phenotypic and genetic analyses of altered trichome metabolites are extremely time- and labor-intensive. In particular, only a small portion of the mapping population revealed JP117 mutant phenotypes and could be used for further genetic analysis (For example, 30 of 289 OC F2 plants as shown in Figure 3.5). This low yield might be because of the interspecific cross used for mapping. LA0716 was shown to accumulate methylated kaempferol and -quercetin compounds instead of Me-Ms (personal communication with Feng Shi and Chao Li). These chemical profiles suggest that LA0716 does not have functional OMTs or myricetin substrate for the production of Me-Ms. In addition, epistatic interactions

between LA0716 and JP117 mutant backgrounds might have changed biochemistry in OC F2 in unpredicted manners. Another challenge in genetic mapping is that genetic markers are not saturated for the cultivated tomato and its wild relatives. Given the circumstance, although still incomplete, the genomic sequence of *S. lycopersicum* and *S. pennellii* guided to the identification of candidate genes by enabling gene homology search and providing polymorphic sequences based on which genetic markers can be generated.

SlMOMT1a and 1b genes were found to be closely linked within the last 1.5Mb region of chromosome 6. Previous studies reported examples where specialized metabolite biosynthesis genes are clustered. Clustering of genes for terpene production is a well-known example (Falara *et al.* 2011). These clustered gene families are shown to merge during gene duplication and obtain novel functions (Pichersky and Gang 2000, Sasaki *et al.* 2002, Vision *et al.* 2000).

In addition to the biochemistry of trichome metabolites, understanding their biological roles is also important. Although the impact of various flavonoids on insect feeding (Onyilagha *et al.* 2004) and insect enzyme activities (Mitchell *et al.* 1993) were studied, little is known about the role of Me-Ms in plant-herbivore insect interactions. Field and green house experiments revealed the higher plant damage in JP117 than M82 and IL6-4 against *M. sexta* and *E. cucumeris* insects. Because plants produce and release cocktails of a wide range of chemical compounds (Baldwin 2001), it is hard to test the role of specific metabolites in plant defense. Two potential reasons are speculated to explain

variable plant defense among the genetic variants, JP117, M82 and IL6-4. First, given that the impact of specialized metabolites on plant defense is concentration-dependent (Mitchell *et al.* 1993), reduced total Me-Ms in JP117 can be thought to impair plant resistance. Otherwise, based on the reports that chemical modification of hydroxyl groups on flavonoids alters their chemical activities (Cao *et al.* 1997, Heim *et al.* 2002), differential accumulation of Me-Ms can be considered to vary plant defense in JP117, M82 and IL6-4.

Despite intensive efforts, our knowledge of trichome methylated myricetin biochemistry is still limited. This study revealed that two or more MOMTs on the bottom of chromosome 6 are involved in the production of trichome Me-Ms in *S. lycopersicum* cultivated tomato and suggested the impact of Me-Ms on plant defense against herbivore insects. In that little is known about trichome metabolites, especially about non-volatile compounds, this study helps improve our understanding about trichome metabolism.

Materials and methods

Plant growth conditions

M82 cultivated tomato seeds were obtained from the Tomato Genetic Resource Center (http://tgrc.ucdavis.edu/). Tomato EMS M2 mutant population and *S. pennellii* ILs were provided from Dr. Dani Zamir (The Hebrew Univeristy of Jerusalem, Faculty of Agriculture, Rehobot, Israel). Plants were grown for 3 weeks under the growth conditions as described in Schilmiller *et al.* 2010a.

A single leaf dip extraction and negative mode LC-ToF-MS analysis

For chemical extraction of non-volatile trichome metabolites, a single leaf dip method was used as described in Schilmiller *et al.* 2010a. Negative mode LC-ToF-MS analysis was performed following previous studies (Schilmiller *et al.* 2010a). To quantify the accumulation of Me-Ms, individual peak areas of mono-, di- and tri-Me-M were obtained from LC-ToF-MS analysis and normalized to internal standard and leaf dry weight. Total Me-M amount was calculated by summing the peak areas of all three Me-Ms.

Harvesting types 1 and 6 trichome extracts and positive mode LC-MS/MS analysis

A total of 50 glands from each type of trichomes were collected with micropipettes and extracted in 50 μl of the extraction solvent used for LC-ToF-MS analysis. Types 1 and 6 trichomes were harvested from 3 week-old plants grown under the identical growth conditions as used for LC-ToF-MS analysis. LC-MS/MS analysis and data interpretation including metabolite annotation and quantification was performed by Chao Li as described in (Schmidt *et al.* 2011).

Generation of genetic cross progeny for JP117 mutant and IL6-3/4

JP117 mutant was backcrossed to M82 wild type plant in a reciprocal manner, resulting in M82 x JP117 and JP117 x M82 plants. F1 progeny plants from both reciprocal BC were chemically analyzed and they revealed identical Me-M profiles. BC F1 and F2 progeny from a M82 x JP117 cross were used for the genetic studies.

IL6-3/4 plants were crossed to M82 in a reciprocal manner. Consistent Me-M profiles were observed from all F1 progeny from various crosses, M82 x IL6-3, IL6-3 x M82, M82 x IL6-4 and IL6-4 x M82. BC F1 and F2 progeny from a M82 x IL6-4 cross were used for the genetic studies.

Genetic mapping of JP117 mutant

JP117 mutant was outcrossed to *S. pennellii* (LA0716) to generate mapping population, OC F2 progeny. Publicly available markers (markers 1~4 and 13 from Figure 3.5) were obtained from Sol Genomics Network (http://solgenomics.net/). When markers were not available for the location of interest, CAPS markers were generated in this study (markers 5~12 from Figure 3.5). The polymorphic sequences between *S. lycopersicum* and *S. pennellii* were identified based on genomic sequence and trichome ESTs for two species, *S. lycopersicum* (http://solgenomics.net/, SL2.31 chromosome sequences) and *S. pennellii* (http://mapman.mpimp-golm.mpg.de/blast/blast.html) (Schilmiller *et al.* 2010b). CAPS markers were generated by CAPS designer program (http://solgenomics.net/tools/caps_designer/caps_input.pl/). Primer sequences and the types of restriction enzymes for individual markers are listed in Table 3.1.

Genotyping was performed by genomic DNA (gDNA) PCR followed by digestion with suitable restriction enzymes. For PCR, 0.5 μg of gDNA was used with 10 μl of RedTaq mix (Sigma-Aldrich, http://www.sigmaaldrich.com/) and 5 μl of forward and reverse primers (2pmol/μl), respectively. The thermocycling conditions were as follows: 40 times for 45 s at 94°C, 45 s at 55°C and 1 min at 72 °C. The PCR products were incubated with

restriction enzymes for 4 hrs at 37°C. These restriction enzymes were chosen to distinguish *S. lycopersicum* or *S. pennellii* genetic background by recognizing polymorphic sequence between two different species *S. lycopersicum* and *S. pennelli* and cutting DNA fragment amplified from one species not from the other.

Mapping the location of MOMTs on chromosome 6

The generation of CAPS markers and genotypes was performed as described above. Primer sequences and the types of restriction enzymes for individual markers are listed in Table 3.2.

Search for candidate OMT genes

BLAST analysis was performed to obtain genomic DNA sequence for MOMT homologs in *S. lycopersicum* (http://solgenomics.net/, SL2.31 chromosome sequences) and *S. pennellii* (http://mapman.mpimp-golm.mpg.de/blast/blast.html) using shMOMT gene sequences (Schmidt *et al.* 2011). The domain characteristic for plant OMT was searched (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). Gene structures were predicted by Softberry program (www.softberry.com).

Generation of phylogeny tree and validation of its reliability

Phylogeny tree was generated by Neighbor joining method (Saitou and Nei 1987) and its reliability was tested by bootstrap test (Felsenstein 1985) with n=500. Phylogeny tree generation and statistical test was performed by MEGA5 program (Tamura *et al.* 2011).

The field - and green house insect feeding assays

Plants were grown for 3 weeks under growth chamber conditions as discussed above. Three week-old plants were transferred to a green house and grown for another week. The 4 week-old plants were used for the field - and green house insect feeding assays. Different genetic variants were placed with randomized orders.

Field experiments were performed at the Department of Plant Pathology Research Farm, in the summer of 2010 and 2011. Plants were monitored once a week to observe herbivore insects and plant damage. The monitoring was performed twice a week once plant damage occurred by insects. To estimate plant damage by *M. sexta*, the damage degree was assigned to the scale of 0 to 3 based on how much plant tissues were lost. For example, when the loss was 0~10% compared to whole plant mass, value '0' was scored. Values from 1 to 3 were assigned when the loss was 10~30, 30~60 and 60~100% respectively. Damage scoring was performed with a blind approach (i.e. the genotype of plants were not provided to observers when plant damage was scored). To estimate plant damage by *E. cucumeris,* the number of holes was counted.

Green house test was performed to repeat *M. sexta* feeding assay. The eggs were obtained from the Department of Entomology, North Carolina State Universtiy in Raleigh and hatched at 26°C following supplier's recommendation. Two newly hatched larvae were placed on the consistent positions of leaf tissue for individual plants. This aims to minimize the variation of insect feeding due to different plant developmental stages. Thirty replicate plants were tested for each genotype of JP117, M82 and IL6-4. Four

Table 3.2. The list of co-dominant amplified polymorphic sequence (CAPS) markers used to map MOMTs on chromosome 6.

Primer sequences and restriction enzymes are shown for individual CAPS markers used to map the location of MOMTs on chromosome 6. See 'Materials and Methods' for details about genotyping procedures.

MOMT Name	Primer Type	Primer Sequence (5'-3')	Restriction Enzyme	Expected Product Size (bp)	
				L	P
1a	F	AGGTGGACTTGGAA TATCATTAGCT	EcoRV	200, 500	700
	R	TGTATTTGTTCAATT ACAACTACTTTACC ATCA			
1b	F	GAACTTGACCTATT TGAGATTATAGCA	ApoI	1146	366, 780
	R	AATATTCTAAGTAA GGTGAGTGTGAAAT TTC			
1c	F	ATGACTTCTTTTCAT CCGTTAA	KpnI	341	117, 224
	R	TTATTGTTAATTTTC TAAGTTA			
2a	F	ATGCCAACGCAATC TTACTCAAGG	None	1300	1100
	R	TAAAACCAGCCTCC AAAAAGAGTTTTTC			
2b	F	CAACATGTATGCTA ATGAGCTAATAG	DraI	1210	177, 1033
	R	TTAAGGATAAACTT CCATTAGAGACC			

replicate plants were tested for control plants *od-2* and CM respectively. Living *M. sexta* larvae were collected after 14 days of feeding on the individual genetic variants and insect weight was measured.

One-way ANOVA test was performed to validate the variation of insect feeding in JP117 mutant, M82 and IL6-4 by using JMP8 program (http://www.jmp.com/software/jmp8/).

Acknowledgement

I thank Chao Li, a former graduate student in Dr. Jones' group, and Dr. Jones for LC-MS/MS performance and data analysis including metabolite identification and quantification for LC-MS/MS analysis of the isolated types 1 and 6 trichome extracts.

References

References

Baldwin, I.T. (2001) An ecologically motivated analysis of plant-herbivore interactions in native tobacco. *Plant Physiol.*, **127,** 1449-1458.

Baldwin, I.T., Halitschke, R., Paschold, A., von Dahl, C.C. and Preston, C.A. (2006) Volatile signaling in plant-plant interactions: "talking trees" in the genomics era. *Science*, **311,** 812-815.

Braca, A., Bilia, A.R., Mendez, J. and Morelli, I. (2001) Myricetin glycosides from *Licania densiflora*. *Fitoterapia*, **72,** 182-185.

Buchana, B., Gruissem, W. and Jones, R. (2002) *Biochemistry & Molecular Biology of Plants*: Wiley.

Buer, C.S., Imin, N. and Djordjevic, M.A. (2010) Flavonoids: new roles for old molecules. *J. Integr. Plant. Biol*, **52,** 98-111.

Cao, G., Sofica, E. and Priora, R.L. (1997) Antioxidant and prooxidant behavior of flavonoids: structure-activity relationships. *Free Radical Biology and Medicine*, **22,** 749-760.

Eshed, Y. and Zamir, D. (1994) A genomic library of *Lycopersicon pennellii* in *Lycopersicon esculentum* - A tool for fine mapping of genes *EUPHYTICA*, **79,** 175-179.

Eshed, Y. and Zamir, D. (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, **141,** 1147-1162.

Falara, V., Akhtar, T., Nguyen, T.T., Spyropoulou, E.A., Bleeker, P.M., Schauvinhold, I., Matsuba, Y., Bonini, M.E., Schilmiller, A.L., Last, R.L., Schuurink, R.C. and Pichersky, E. (2011) The tomato (*Solanum lycopersicum*) terpene synthase gene family. *Plant Physiol.,* **157,** 770-789.

Felsenstein, J. (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, **39,** 783-791.

Fontanarrosa, M.S., Collantes, M.B. and Bachmann, A.O. (2009) Seasonal patterns of the insect community structure in urban rain pools of temperate Argentina. *J. Insect. Sci.*, **9,** 10.

Gang, D.R. (2005) Evolution of flavors and scents. *Annu. Rev. Plant. Biol.*, **56,** 301-325.

Heim, K.E., Tagliaferro, A.R. and Dennis J Bobilyaa, A. (2002) Flavonoid antioxidants: chemistry, metabolism and structure-activity relationships. *The Journal of Nutritional Biochemistry*, **13,** 572-584.

Ibrahim, R.K., Bruneau, A. and Bantignies, B. (1998) Plant *O*-methyltransferases: molecular analysis, common signature and classification. *Plant Molecular Biology*, **36,** 1-10.

Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M. and Last, R.L. (2002) *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol.*, **129,** 440-450.

Kang, J.H., Liu, G., Shi, F., Jones, A.D., Beaudry, R.M. and Howe, G.A. (2010a) The tomato *odorless-2* mutant is defective in trichome-based production of diverse specialized metabolites and broad-spectrum resistance to insect herbivores. *Plant Physiol.*, **154,** 262-272.

Kang, J.H., Shi, F., Jones, A.D., Marks, M.D. and Howe, G.A. (2010b) Distortion of trichome morphology by the *hairless* mutation of tomato affects leaf surface chemistry. *J. Exp. Bot.*, **61,** 1053-1064.

Kim, D.H., Kim, B.G., Lee, Y., Ryu, J.Y., Lim, Y., Hur, H.G. and Ahn, J.H. (2005) Regiospecific methylation of naringenin to ponciretin by soybean *O*-methyltransferase expressed in *Escherichia coli*. *J. Biotechnol.*, **119,** 155-162.

Liu, B., Wang, H., Du, Z., Li, G. and Ye, H. (2010) Metabolic engineering of artemisinin biosynthesis in *Artemisia annua L*. *Plant. Cell. Rep.*

McDowell, E.T., Kapteyn, J., Schmidt, A., Li, C., Kang, J.H., Descour, A., Shi, F., Larson, M., Schilmiller, A., An, L., Jones, A.D., Pichersky, E., Soderlund, C.A. and Gang, D.R. (2011) Comparative functional genomic analysis of Solanum glandular trichome types. *Plant Physiol.*, **155,** 524-539.

Mitchell, M.J., Keogh, D.P., Crooks, J.R. and Smith, S.L. (1993) Effects of plant flavonoids and other allelochemicals on insect cytochrome P-450 dependent steroid hydroxylase activity. *Insect Biochem. Mol. Biol.*, **23,** 65-71.

Muzac, I., Wang, J., Anzellotti, D., Zhang, H. and Ibrahim, R.K. (2000) Functional expression of an *Arabidopsis* cDNA clone encoding a flavonol 3'-*O*-methyltransferase and characterization of the gene product. *Arch. Biochem. Biophys.*, **375,** 385-388.

Onyilagha, J.C., Lazorko, J., Gruber, M.Y., Soroka, J.J. and Erlandson, M.A. (2004) Effect of flavonoids on feeding preference and development of the crucifer pest *Mamestra configurata* Walker. *J. Chem. Ecol.*, **30,** 109-124.

Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A*, **85,** 2444-2448.

Pichersky, E. and Gang, D.R. (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends. Plant. Sci.*, **5,** 439-445.

Rodriguez-Saona, C., Vorsa, N., Singh, A.P., Johnson-Cicalese, J., Szendrei, Z., Mescher, M.C. and Frost, C.J. (2011) Tracing the history of plant traits under domestication in cranberries: potential consequences on anti-herbivore defences. *J Exp Bot*, **62,** 2633-2644.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4,** 406-425.

Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y., Antonio, B.A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K., Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K., Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y., Idonuma, A., Iijima, M., Ikeda, M., Ikeno, M., Ito, S., Ito, T., Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S., Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T., Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M., Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M., Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T., Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K., Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K., Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J.H., Kim, H.I., Eun, M.Y., Yano, M., Jiang, J. and Gojobori, T. (2002) The genome sequence and structure of rice chromosome 1. *Nature*, **420,** 312-316.

Schilmiller, A., Shi, F., Kim, J., Charbonneau, A.L., Holmes, D., Daniel Jones, A. and Last, R.L. (2010a) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. *Plant J.*, **62,** 391-403.

Schilmiller, A.L., Last, R.L. and Pichersky, E. (2008) Harnessing plant trichome biochemistry for the production of useful compounds. *Plant J.*, **54,** 702-711.

Schilmiller, A.L., Miner, D.P., Larson, M., McDowell, E., Gang, D.R., Wilkerson, C. and Last, R.L. (2010b) Studies of a biochemical factory: tomato trichome deep expressed sequence tag sequencing and proteomics. *Plant Physiol.*, **153,** 1212-1223.

Schilmiller, A.L., Schauvinhold, I., Larson, M., Xu, R., Charbonneau, A.L., Schmidt, A., Wilkerson, C., Last, R.L. and Pichersky, E. (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate

precursor rather than geranyl diphosphate. *Proc. Natl. Acad. Sci. U. S. A.*, **106,** 10865-10870.

Schmidt, A., Li, C., Shi, F., Jones, A.D. and Pichersky, E. (2011) Polymethylated myricetin in trichomes of the wild tomato species *Solanum habrochaites* and characterization of trichome-specific 3'/5'- and 7/4'-myricetin *O*-methyltransferases. *Plant Physiol.*, **155,** 1999-2009.

Stamp, N.E. and Yang, Y. (1996) Response of Insect Herbivores to Multiple Allelochemicals Under Different Thermal Regimes. *Ecology*, **77,** 1088-1102.

Stevens, J., Hart, H. and Wollenweber, E. (1995) The systematic and evolutionary significance of exudate flavonoids in Aeonium. *Phytochemistry*, **39,** 805-813.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*

Tian, Q., Li, D. and Patil, B.S. (2002) Identification and determination of flavonoids in buckwheat (Fagopyrum esculentum Moench, Polygonaceae) by high-performance liquid chromatography with electrospray ionisation mass spectrometry and photodiode array ultraviolet detection. *Phytochem Anal*, **13,** 251-256.

van Der Heijden, R., Jacobs, D.I., Snoeijer, W., Hallard, D. and Verpoorte, R. (2004) The *Catharanthus* alkaloids: pharmacognosy and biotechnology. *Curr. Med. Chem.*, **11,** 607-628.

Vingron, M. and Waterman, M.S. (1994) Sequence alignment and penalty choice. Review of concepts, case studies and implications. *J Mol Biol*, **235,** 1-12.

Vision, T.J., Brown, D.G. and Tanksley, S.D. (2000) The origins of genomic duplications in *Arabidopsis*. *Science*, **290,** 2114-2117.

Wagner, G. (1991) Secreting glandular trichomes: more than just hairs. *Plant Physiol.*, **96,** 675-679.

Wagner, G.J., Wang, E. and Shepherd, R.W. (2004) New approaches for studying and exploiting an old protuberance, the plant trichome. *Ann. Bot.*, **93,** 3-11.

Wang, G., Tian, L., Aziz, N., Broun, P., Dai, X., He, J., King, A., Zhao, P.X. and Dixon, R.A. (2008) Terpene biosynthesis in glandular trichomes of hop. *Plant Physiol.*, **148,** 1254-1266.

Weinhold, A. and Baldwin, I.T. (2011) Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation. *Proc. Natl. Acad. Sci. U. S. A*, **108,** 7855-7859.

Chapter 4. Mass spectrometry reveals strong diversity in glandular trichome acylsugars of the wild tomato *Solanum habrochaites*

Abstract

Acylsugars are polyesters of short to medium length acyl chains on sucrose or glucose produced in secretory glandular trichomes of many solanaceous plants, including cultivated tomato. Despite their abundance and wide taxonomic distribution, there is relatively little information about the diversity of these compounds and nearly nothing is known about 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. Trichome acylsugar profiles were compared to each other and to that of the cultivated tomato (*S. lycopersicum* M82) using hierarchical clustering and principal component analyses. These approaches revealed distinct phenotypic classes that varied based on the presence of glucose or sucrose, the numbers and lengths of acyl chains and relative total amounts of acylsugars. Chemically similar accessions clustered geographically. The finding of such strong *S. habrochaites* acylsugar diversity suggests the existence of a relatively large repertoire of biosynthetic enzymes in the acylsugar metabolic network.

Introduction

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, secreting glandular trichomes (SGTs) serve as 'chemical factories' where specialized metabolites are produced, stored or volatized (Schilmiller *et al.* 2010a, Schilmiller *et al.* 2008, Wagner 1991). In addition, SGTs produce and secrete proteins on the plant surface for insect protection (Thipyapong *et al.* 1997, Yu *et al.* 1992) 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 L.* 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* L. (Liu *et al.* 2011). In addition to their value in foods and medicines, trichomes provide excellent models for analyzing biosynthetic enzymes and pathways (Bohlmann and Gershenzon 2009, Sallaud *et al.* 2009, Schilmiller *et al.* 2008, Schilmiller *et al.* 2009).

Plants in the genus Solanum include important crop species such as potato, eggplant and tomato. Previous studies reported that SGTs of cultivated tomato (*Solanum lycopersicum*)

and its wild relatives accumulate high levels of exudates containing a variety of specialized metabolites, for example flavonoids, alkaloids and terpenoids (McDowell *et al.* 2011, Schilmiller *et al.* 2010a, Schilmiller *et al.* 2008, Wagner 1991).

Cultivated tomato and its wild relatives have morphologically and chemically diverse trichomes. For example, Luckwill (Luckwill 1943) defined seven morphologically distinguishable types of trichomes in plants of this genus including four glandular types (type 1, 4, 6, and 7) (Figure 4.1; for more examples of Solanum trichome images, see Figures 3 and 1 from Kang *et al.* 2010a and Kang *et al.* 2010b, respectively). The presence of specific types of trichomes and their densities vary across species and even within a single plant according to tissue types and developmental stages (Werker 2000). 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 *S. habrochaites* accessions (Fridman *et al.* 2005, Yu *et al.* 2010). Similarly, acylglucoses are highly abundant in type 4 glands of *S. pennellii* LA0716 while acylsucroses predominate in *S. lycopersicum* and *S. habrochaites* (Kennedy 2003, McDowell *et al.* 2011). The chemical and morphological diversity of trichomes in different Solanum species and accessions makes the genus an attractive target for identification of diverse trichome-borne metabolites and the major biosynthetic pathways responsible for their synthesis operating in each trichome type.

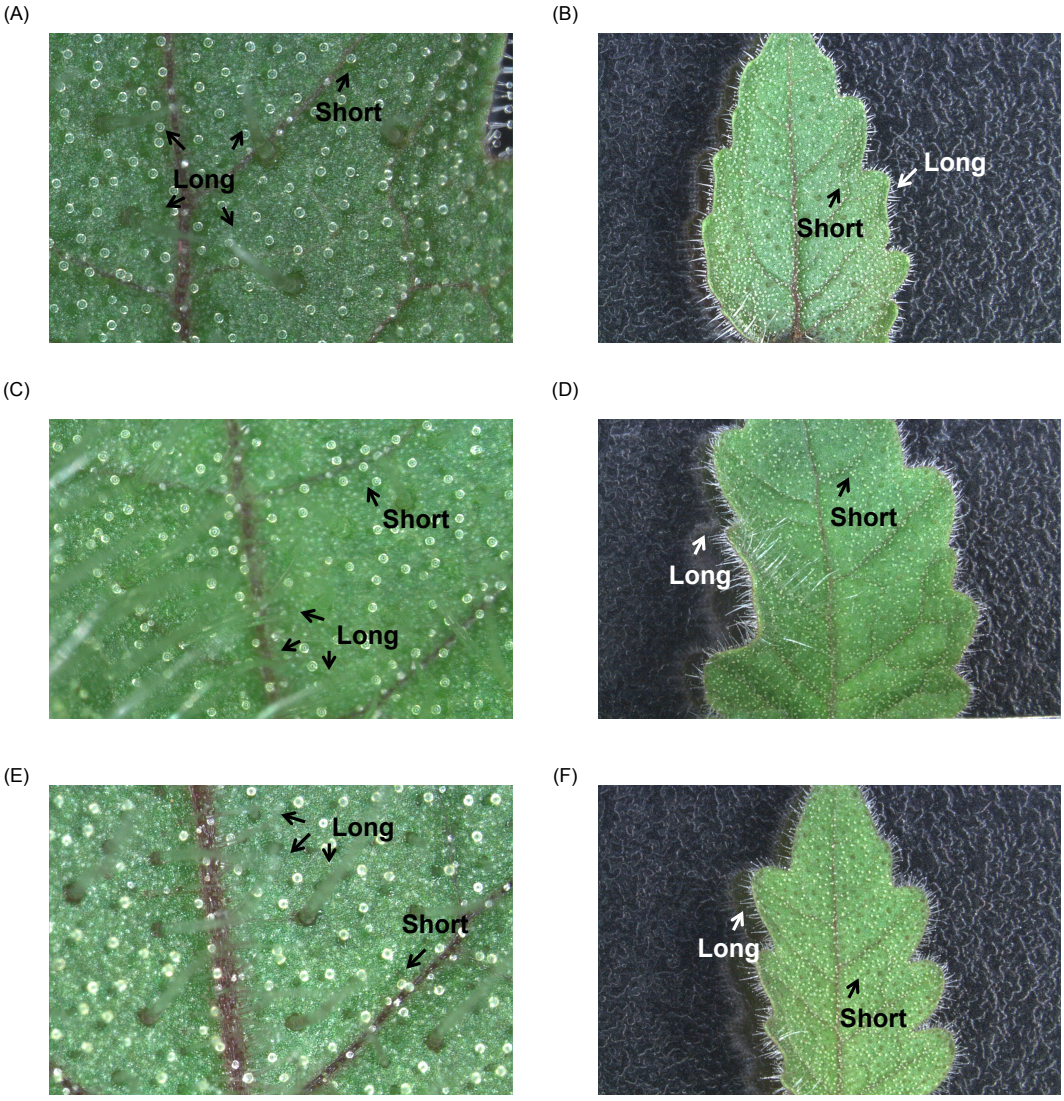


Figure 4.1. Light microscope images for leaf trichomes and whole leaf from three *S. habrochaites* accessions.

A-B. LA2101; C-D. LA1777; E-F. LA2975. Images were taken with three week-old plants. 'Long' indicates long SGTs, either type 1 or 4 and 'short' short SGTs such as type 6 or 7.

The value of the comparative metabolomics approach in trichomes was recently demonstrated in the studies of Solanum trichome mono- and sesquiterpene biosynthesis (Bohlmann and Gershenzon 2009, Sallaud *et al.* 2009, Schilmiller *et al.* 2009). It was discovered that *S. lycopersicum* trichomes synthesize monoterpenes from the *cis*-prenyldiphosphate intermediate neryl diphosphate (Schilmiller *et al.* 2009). This is contrary to the previous paradigm where the *trans*-prenyldiphosphate geranyldiphosphate 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*-farnesyldiphosphate (*Z,Z*-FPP). This is counter to the commonly described cytosolic sesquiterpene pathway, which uses the all *trans* sesquiterpene synthase substrate *E,E*-farnesyldiphosphate. These observations suggest that trichome specialized metabolism is evolutionarily plastic, perhaps due to selective pressure from insects or other environmental stress agents.

Acylsugars are sticky exudates that are thought to physically or chemically improve plant defense (Mirnezhad *et al.* 2010). Results from the literature indicate strong acylsugar diversity in various Solanum trichomes (McDowell *et al.* 2011, Schilmiller *et al.* 2010a, Schilmiller *et al.* 2010b). Acylsugars are categorized as either sucrose- or glucose esters based on the type of sugar core (Figure 4.2) and they also have varying number of acyl chains decorating the sugar moiety. In particular, *S. pennellii* 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.*

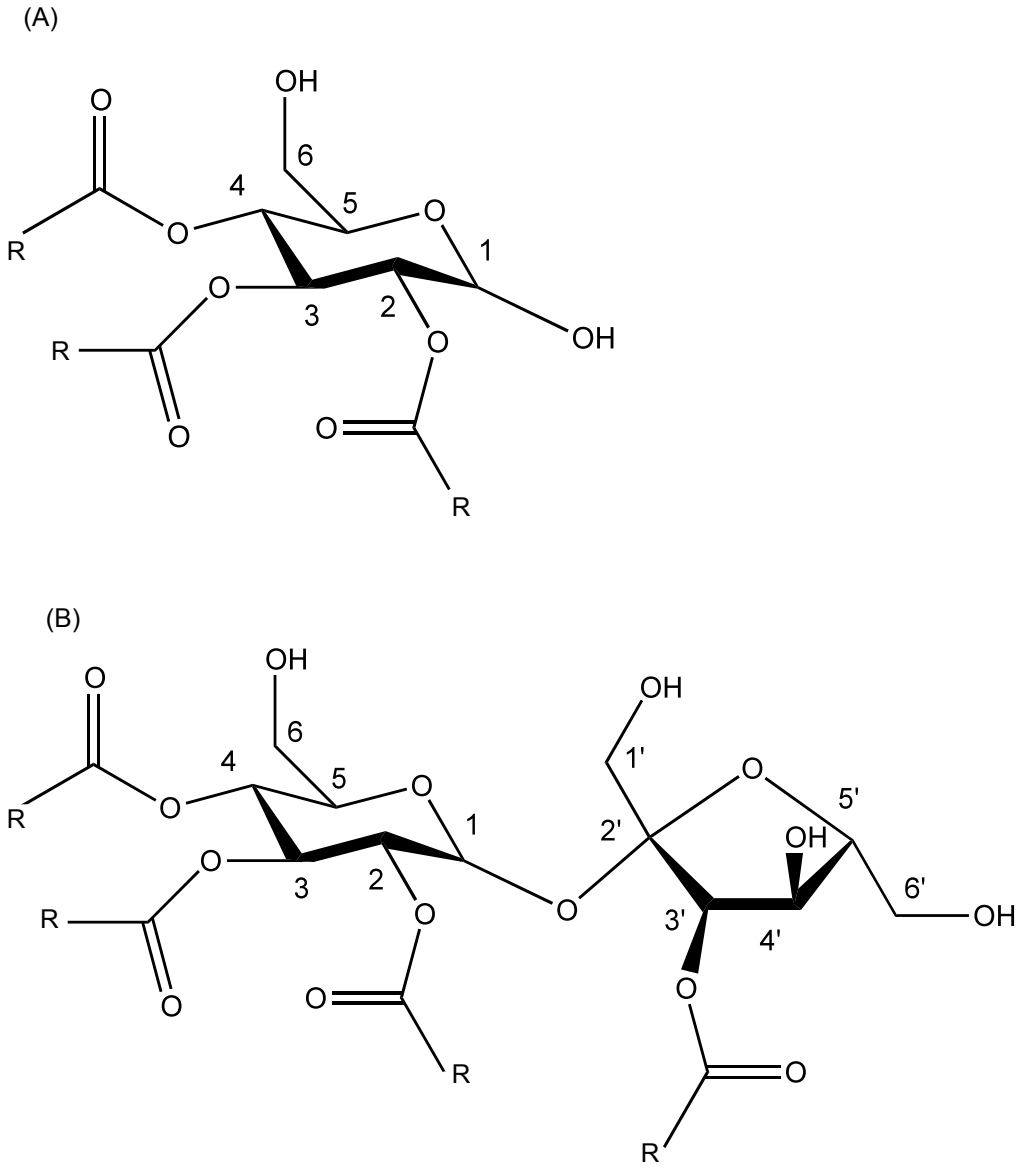


Figure 4.2. Structural classes of acylsugars in Solanum species.

(A) Schematic structure of an acylglucose. The structure shown depicts a glucose triester comprised of glucose and 3 acyl chains with various numbers of carbons represented as 'R'.

Figure 4.2. Continued

(B) The schematic structure of an acylsucrose. The proposed structure of a sucrose tetraester with three acyl chains on the glucose ring and one on the fructose ring. If the sugar moiety is decorated with either three or five acyl chains, it is referred to as either a sucrose triester or pentaester, respectively. The positions of the acyl chains are currently unknown, with the exception of the most abundant acylsugar in cultivated tomato (M82) that was structurally characterized by NMR (Schilmiller *et al.* 2010a).

pennellii accessions vary in quantity, the proportion of sucrose or glucose 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 (Table 4.1).

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 *Solanum habrochaites* accessions collected from throughout the range of the species (Table 4.1 and Sifres *et al*. 2011) and M82 cultivated tomato (*Solanum lycopersicum*).

As diagrammed in Figure 4.3, 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 good metabolite selectivity in a relatively rapid assay. The leaf dip sampling method was chosen to maximize extraction of glandular trichome metabolites, while minimizing contamination by epidermal pavement cells (Schilmiller *et al.* 2010a). Mass

Table 4.1. The location of original collection site for *S. habrochaites* accessions.

The geographic information was obtained from TGRC (Tomato Genetics Resource Center, http://tgrc.ucdavis.edu/). Accessions were listed in geographical order from north to south. Accessions lacking latitude and longitude information in the TGRC database are listed in the order of accession numbers. See Figure 4.6 for more details about HCA result.

Table 4.1.

Accession	Site	Dept./ Prov	Country	Lati-tude	Longi-tude	Eleva-tion	HCA
LA1625	S of Jipijapa	Manabi	Ecuador	-1.33	-80.58	-	C
LA1624	Jipijapa	Manabi	Ecuador	-1.33	-80.57	-	C
LA1266	Pallata-nga	Chimbo-razo	Ecuador	-1.98	-78.45	1,000	D
LA1223	Alausi	Chimbo-razo	Ecuador	-2.2	-78.83	2,200	D
LA0407	Mirador Guayaq-uil	Guayas	Ecuador	-2.98	-79.77	40	D
LA2119	Saraguro	Loja	Ecuador	-3.6	-79.22	2,600	D
LA2861	Las Juntas	Loja	Ecuador	-3.82	-79.27	2,090	D
LA2128	Zumbi	Zamora Chinchi-pe	Ecuador	-3.89	-78.78	1,000	D
LA2105	Jardin Botanico, Loja	Loja	Ecuador	-4	-79.22	2,200	D
LA1253	Pueblo Nuevo-Landang-ue	Loja	Ecuador	-4.14	-79.2	2,000	D
LA2106	Yambra	Loja	Ecuador	-4.22	-79.23	1,700	D
LA3863	Sozora-nga	Loja	Ecuador	-4.33	-79.78	-	D
LA2100	Sozora-ngo	Loja	Ecuador	-4.33	-79.78	1,200	E
LA2864	Sozora-ngo	Loja	Ecuador	-4.33	-79.78	1,650	E
LA2101	Cariam-anga	Loja	Ecuador	-4.33	-79.55	1,800	D
LA2860	Cariam-anga	Loja	Ecuador	-4.33	-79.55	1,700	D
LA2098	Sabianga	Loja	Ecuador	-4.37	-79.8	700	E
LA2110	Yangana #2	Loja	Ecuador	-4.38	-79.18	2,000	D

Table 4.1. Continued

LA3864	Yangana	Loja	Ecuador	-4.38	-79.18	-	D
LA2109	Yangana #1	Loja	Ecuador	-4.38	-79.18	2,000	D
LA2650	Ayabaca	Piura	Peru	-4.63	-79.95	800	B
LA2175	Timbar-uca	Cajam-arca	Peru	-5.14	-79.01	1,150	E
LA1718	Huanca-bamba	Piura	Peru	-5.24	-79.45	-	D
LA2204	Balsa-pata	Amaz-onas	Peru	-5.77	-77.87	-	E
LA2196	Caclic	Amaz-onas	Peru	-6.18	-77.91	1,600	E
LA2314	San Francisco	Amaz-onas	Peru	-6.42	-77.87	1,650	E
LA2156	Ingenio Montan	Cajam-arca	Peru	-6.53	-78.79	1,700	E
LA2155	Maydas-bamba	Cajam-arca	Peru	-6.65	-78.54	2,600	E
LA2812	Lambay-eque	Lamba-yeque	Peru	-6.7	-79.92	-	E
LA0386	Cajama-rca	Cajam-arca	Peru	-7.16	-78.5	2,500	B
LA4137	Barrio Delta, Cajama-rca	Caja-marca	Peru	-7.17	-78.35	2,900	B
LA2167	Cemen-terio Cajam-arca	Cajam-arca	Peru	-7.18	-78.52	2,300	B
LA1352	Rupe	Cajam-arca	Peru	-7.3	-78.81	2,100	E
LA3854	Llaguen	Chicam a	Peru	-7.72	-78.73	1,800	B
LA2329	Aricapam pa	La Libert-ad	Peru	-7.81	-77.7	2,450	B
LA1393	Caraz	Ancash	Peru	-9.05	-77.81	2,280	A
LA1362	Chacc-han	Ancash	Peru	-9.31	-77.98	2,100-2,300	A
LA1775	Rio Casma	Ancash	Peru	-9.52	-77.88	900-1,800	A
LA2574	Cullas-pungro	Ancash	Peru	-9.54	-77.8	2,000	A

Table 4.1. Continued

LA1361	Pariacoto	Ancash	Peru	-9.55	-77.89	1,490	A
LA2975	Coltao	Ancash	Peru	-9.55	-77.68	3,200	A
LA1777	Rio Casma	Ancash	Peru	-9.55	-77.59	3,150	A
LA2976	Huangra	Ancash	Peru	-9.56	-77.7	3,000	A
LA1392	Huaraz to Casma	Ancash	Peru	-9.56	-77.7	3,120	A
LA1779	Rio Casma	Ancash	Peru	-9.56	-77.66	2,600	A
LA1978	Colca	Ancash	Peru	-10.11	-77.48	2,450	B
LA3796	Anca, Marca	Ancash	Peru	-10.12	-77.49	2,810	B
LA3794	Alta Fortaleza	Ancash	Peru	-10.15	-77.36	-	B
LA1557	Rio Huara	Lima	Peru	-11.22	-76.63	-	A
LA1764	W of Canta	Lima	Peru	-11.47	-76.62	2,400	B
LA0094	Canta-Yangas	Lima	Peru	-11.52	-76.68	2,100	B
LA1559	Desvio-Huaman-tanga	Lima	Peru	-11.53	-76.69	-	B
LA1648	Above-Yaso	Lima	Peru	-11.55	-76.72	1,500	B
LA0361	Canta	Lima	Peru	-11.57	-76.73	1,600	B
LA1298	Yaso	Lima	Peru	-11.58	-76.75	1,600	B
LA1295	Surco	Lima	Peru	-11.88	-76.44	2,100	B
LA1753	Surco	Lima	Peru	-11.88	-76.44	2,000	B
LA1560	Matu-cana	Lima	Peru	-11.89	-76.44	-	B
LA2409	Miraflo-res	Lima	Peru	-12.27	-75.83	3,000	D
LA1691	Yauyos	Lima	Peru	-12.45	-75.93	2,074	B
LA2722	Puente Auca	Lima	Peru	-12.59	-75.95	2,500	B
LA1681	Mushka	Lima	Peru	-12.75	-75.85	2,450	B
LA1695	Cacach-uhuasin, Canete	Lima	Peru	-12.81	-75.78	-	B
LA1696	Huanchu y-Cacra	Lima	Peru	-12.82	-75.79	2,100	B

Table 4.1. Continued

LA1721	Ticrapo Viejo	Huanca velica	Peru	-13.38	-75.44	2,100	B
LA1918	Llauta	Ica	Peru	-14.27	-74.92	2,600	B
LA1928	Ocana	Ica	Peru	-14.38	-74.82	2,660	B
LA1927	Ocoba-mba	Ica	Peru	-14.47	-74.81	2,540	B
LA1033	Hacien-da Taulis	Lamb-ayeque	Peru	-	-	-	E
LA1255	Pedistal	Loja	Ecuador	-	-	-	D
LA1265	Rio Chimbo	Chim-borazo	Ecuador	-	-	-	C
LA1347	Empalme Otusco	La Liber-tad	Peru	-	-	-	A
LA1391	Bagua to Olmos	Cajam-arca	Peru	-	-	-	E
LA1731	Rio San Juan	Huanca velica	Peru	-	-	-	B
LA2099	Sabianga to Sozora-ngo	Loja	Ecuador	-	-	-	D
LA2104	Pena Negra	Loja	Ecuador	-	-	-	D
LA2107	Los Lirios	Loja	Ecuador	-	-	-	D
LA2144	Chanc-han	Chimb-orazo	Ecuador	-	-	-	D
LA2552	Las Flores	Cajama rca	Peru	-	-	-	B
LA2869	Matola-La Toma	Loja	Ecuador	-	-	-	E

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 (for an example see Figure 4.4). 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 'analytical signals'. The mass defect of the ions, which is the digit of the mass value that follows the decimal place, 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 were characterized as associated with acylsugars from one or more of the 80 *S. habrochaites* accessions analyzed, with 34 acylsucroses and two acylglucoses (Table 4.2). Extracted ion chromatograms for formate adducts of acylsugars (labeled '$[M+HCOO]^-$ m/z' in Table 4.2) 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 to control for variation in injection volume. The value was further normalized to leaf dry weight to account for differences in the size of the plant tissues extracted. These normalized peak areas were used as a measure of acylsugar quantity (Figure 4.3).

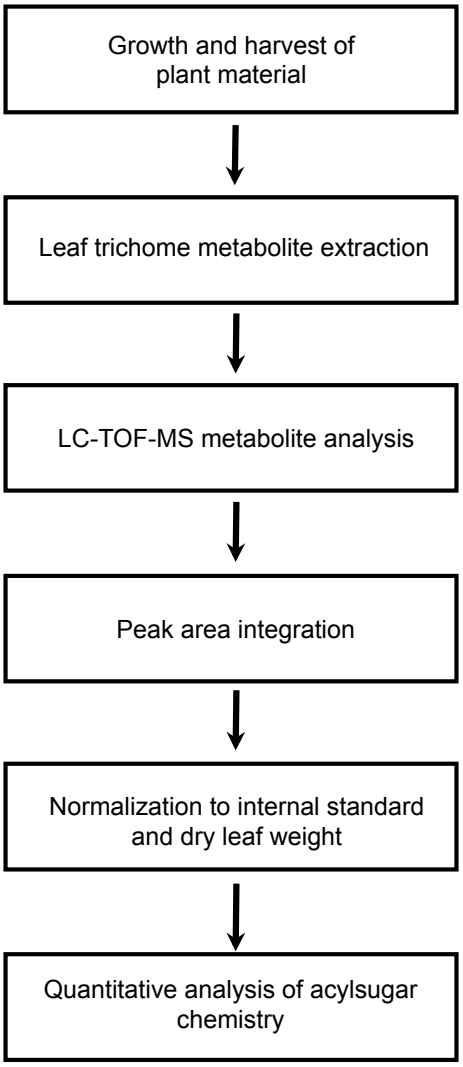


Figure 4.3. Acylsugar analysis workflow.

Plant extracts from the single leaf dip method were analyzed by LC-TOF-MS (Schilmiller *et al.* 2010a). Peak areas for extracted ion chromatogram corresponding to 36 acylsugar analytical signals were integrated using QuanLynx software. These signals were 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.

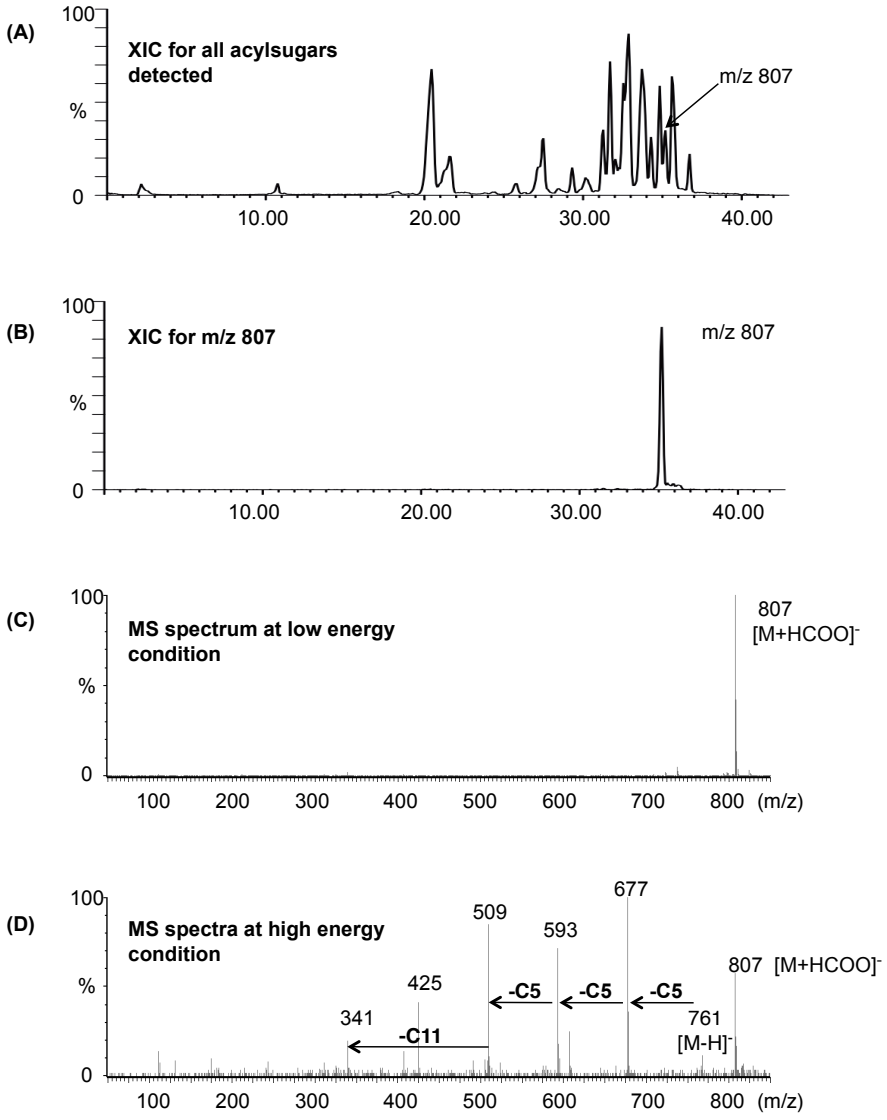

(A)
XIC for all acylsugars detected
m/z 807
100
%
0
10.00
20.00
30.00
40.00
(B)
XIC for m/z 807
m/z 807
100
%
0
10.00
20.00
30.00
40.00
(C)
MS spectrum at low energy condition
807
[M+HCOO]-
100
%
0
100
200
300
400
500
600
700
800
(m/z)
(D)
MS spectra at high energy condition
677
509
593
807 [M+HCOO]-
425
-C5
-C5
-C5
341
761
[M-H]-
-C11
100
%
0
100
200
300
400
500
600
700
800
(m/z)

Figure 4.4

Figure 4.4. An example of the use of LC-TOF with multiplexed collision-induced dissociation to annotate acylsugar S4:26 (5,5,5,11).

Extracted ion chromatograms (XICs) and mass spectra (MS) from *S. habrochaites* accession LA2869 were shown to describe how metabolite annotation was characterized by using MS spectra from different energy conditions. This analytical signal with m/z 807 and RT 35.6 was annotated as S4:26 (5,5,5,11) based on MS spectra obtained under fragmenting and non-fragmenting energy conditions (Table 4.2).

(A) XIC for all acylsugars detected in accession LA2869 under low collision energy (aperture 1 voltage of 10 V) reveals a complex population of related metabolites. A peak with nominal m/z 807 eluting at 35.6 min. is indicated.

(B) XIC for m/z 807 shows a single peak at 35.6 min elution time (aperture 1 voltage of 10 V).

(C) For the peak of m/z 807, the MS spectrum was obtained under low energy non-fragmenting conditions (aperture 1 voltage of 10 V), revealing a single major molecular formate adduct ion $[M+HCOO]^-$.

(D) For the peak of m/z 807, MS spectra were obtained from conditions that promote fragmentation (aperture 1 voltage of 55 V). The results revealed major peaks due to loss of three C5 and one C11 group from m/z 761 $[M-H]^-$, consistent with the annotation of S4:26 (5,5,5,11) for this peak. The m/z 425 peak is interpreted as loss of C11 from m/z 593 peak, S2:16 (5,11), to yield S1:5 (5).

Acylsugar complexity and diversity revealed by comparison of three Solanum accessions

Examination of the acylsugar analytical signals (Table 4.2) revealed highly diverse acylsugar profiles among *S. lycopersicum* M82 and the two *S. habrochaites* accessions LA1777 and LA2106 as shown in Figure 4.5. Acylsugar profiles similar to published results were found for *S. lycopersicum* M82 (Schilmiller *et al.* 2010a), with the sucrose tetraester S4:17 (2,5,5,5) as the most abundant compound, as shown in Figure 4.5a. Note that in this nomenclature, 'S' or 'G' indicates a sucrose or glucose 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 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).

S. habrochaites LA1777 trichome acylsugar profiles were more complex than *S. lycopersicum* M82, with 10 major acylsucrose analytical signals detected (Figure 4.5b). The most striking difference was the length of acylsugar chains. The major acylsugars in M82 consist 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 (Figure 4.5a,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 found at all in M82; for example, S3:18 appears to have a C8 acyl chain

Table 4.2. The list of analytical signals characterized as acylsugars in LC-TOF-MS analysis.

36 analytical signals corresponding to acylsugars were selected based on retention time in LC and accurate mass obtained from low energy condition in negative ion mode MS. Fragment ions generated from high energy condition were used to distinguish isomers that possess the same mass-to-charge ratio (m/z), resulting in identification of 34 acylsucroses and 2 acylglucoses. The metabolite annotation includes the type of sugar backbone, the number of acyl chains, total number of carbons in acyl chains and expected length for each acyl chain. For example, S3:19 (4,5,10) indicates an acylsucrose with 3 acyl chains of 4,5 and 10 carbons respectively, for a total of 19 carbons. $[M+HCOOH]^-$ (m/z) was obtained from the non-fragmenting condition (aperture 1 voltage of 10V) and fragment ions from fragmenting environment (aperture 1 voltage of 55V) in MS. S4:25 (2,5,6,12)* 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)*.

Table 4.2. Continued

Metabolite annotation	Retention time (min)	$[M+HCOO]^-$ (m/z)	Fragment ions (m/z)
S3:13 (4,4,5)	16.5	611	341, 323, 143
S3:14 (4,5,5)	18.1	625	579, 495, 425, 411, 341, 101, 87
S3:15 (5,5,5)	20.5	639	593, 425, 407, 341, 323, 179, 101
S4:14 (2,4,4,4)	16.5	639	593, 551, 87
S4:15 (2,4,4,5)	18.2	653	607, 565, 101, 87
S4:16 (2,4,5,5)	20.1	667	621, 579, 537, 495, 341, 323, 101, 87
S3:18 (5,5,8)	27.5	681	635, 551, 425, 407, 341, 179, 143, 101
S4:17 (2,5,5,5)	22.5	681	635, 551, 509, 425, 407, 341, 323, 101
S3:19 (5,5,9)	29.1	695	649, 565, 509, 425, 341, 323, 157, 101
S3:19 (4,5,10)	30.0	695	565, 495, 425, 407, 341, 171, 101
S4:18 (4,4,5,5)	24.7	695	649, 579, 509, 425, 341
S3:20 (5,5,10)	31.4	709	663, 579, 509, 425, 407, 341, 323, 171, 101
S3:20 (4,5,11)	32.5	709	663, 579, 495, 411, 341, 323, 185, 101, 87
S4:19 (4,5,5,5)	25.5	709	663, 579, 495, 425, 341
S3:21 (5,6,10)	32.2	723	667, 593, 579, 509, 425, 407, 341, 323, 171, 115, 101
S3:21 (5,5,11)	33.5	723	593, 509, 425, 407, 341, 323, 185, 101
S3:21 (4,5,12)	34.4	723	723, 677, 593, 523, 411, 325,199,101, 87
S4:20 (2,4,4,10)	28.1	723	677, 635, 481, 411, 393, 341, 323, 101, 87
S4:20 (5,5,5,5)	27.1	723	677, 593, 509, 425, 341, 101
S3:22 (5,5,12)	34.4	737	691, 607, 509, 425, 407, 341, 323, 199, 179, 101
S4:21 (2,4,5,10)	30.6	737	691, 649, 495, 393, 341, 323, 171, 101, 87

Table 4.2. Continued

Metabolite annotation	Retention time (min)	$[M+HCOO]^-$ (m/z)	Fragment ions (m/z)
S3:23 (5,6,12)	36.3	751	705, 621, 607, 523, 439, 421, 341, 323, 199, 115, 101
S4:22 (2,4,4,12)	31.6	751	709, 663, 481, 393, 341, 323, 305, 199, 87
S4:22 (2,5,5,10)	32.3	751	705, 621, 579, 425, 407, 341, 323, 171, 101
S3:24 (6,6,12)	38.1	765	719, 537, 439, 341, 323, 199, 115
S4:23 (2,4,5,12)	33.7	765	719, 677, 495, 411, 341, 323, 199, 101, 87
S4:23 (2,5,5,11)	33.4	765	719, 677, 509, 425, 185, 101
S4:24 (2,5,5,12)	35.5	779	733, 691, 509, 425, 407, 341, 323, 199, 101
S4:24 (4,5,5,10)	33.9	779	677, 593, 509, 425, 407, 341, 171, 101
S4:25 (5,5,5,10)	30.9	793	663, 579, 425, 341, 171, 101
S4:25 (2,5,6,12)*	33.6	793	663, 579, 395, 199, 101, 115
S5:25 (5,5,5,5,5)	32.1	807	761, 677, 593, 509, 425, 407, 341, 323, 101
S4:26 (5,5,5,11)	35.6	807	761, 677, 593, 509, 425, 341, 185, 101
S4:27 (5,5,5,12)	37.2	821	775, 691, 607, 509, 425, 341, 323, 199, 101
G3:15 (5,5,5)	23.5	477	375, 329, 227, 143, 125, 101
G3:22 (5,5,12)	35.3	575	375, 329, 227, 199, 143, 101

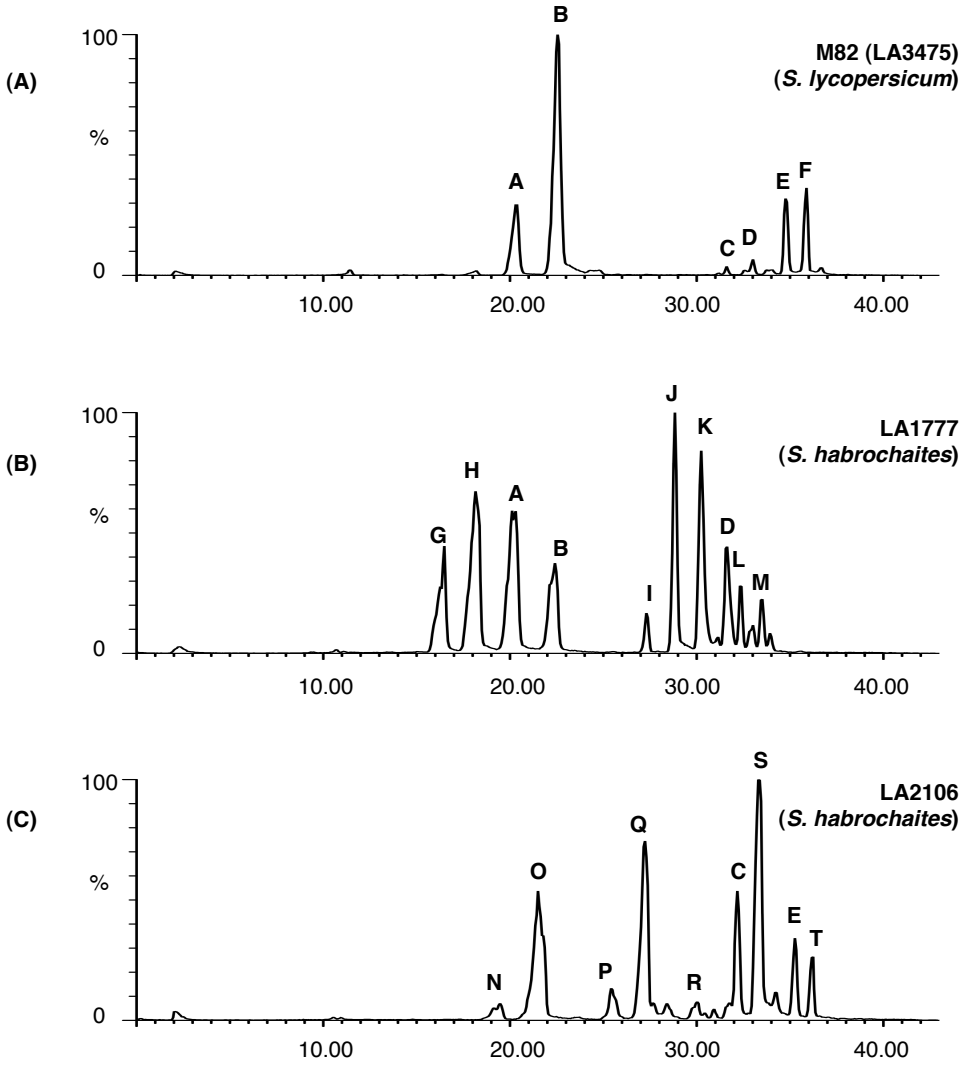

(A)
100
%
0
B
A
E
F
C
D
10.00
20.00
30.00
40.00
M82 (LA3475)
(S. lycopersicum)
(B)
100
%
0
J
K
H
A
G
B
D
L
M
I
10.00
20.00
30.00
40.00
LA1777
(S. habrochaites)
(C)
100
%
0
S
Q
O
C
E
T
N
P
R
10.00
20.00
30.00
40.00
LA2106
(S. habrochaites)

Figure 4.5

Figure 4.5. Extracted ion chromatograms of three accessions illustrating chemical diversity in acylsugar profiles.

Extracted ion chromatograms (XICs) were obtained from leaf dip extraction of M82 (A), LA1777 (B) and LA2106 (C) for acylsugar-associated analytical signals (Table 4.2). The metabolite annotations for the peaks detected with relatively high abundance are listed. A, S4:16 (2,4,5,5); B, S4:17 (2,5,5,5); C, S3:20 (5,5,10); D, S4:22 (2,5,5,10); E, S3:22 (5,5,12); F, S4:24 (2,5,5,12); G, S4:14 (2,4,4,4); H, S4:15 (2,4,4,5); I, S3:18 (5,5,8); J, S4:20 (2,4,4,10); K, S4:21 (2,4,5,10); L, S4:22 (2,4,4,12); M, S4:23 (2,4,5,12); N, S3:14 (4,5,5); O, S3:15 (5,5,5); P, S4:19 (4,5,5,5); Q, S4:20 (5,5,5,5); R, S3:19 (5,5,9); S, S3:21 (5,5,11); T, S3:23 (5,6,12). In M82, early-eluting high intensity peaks correspond to acylsugars with short FA chains. LA2106 has high abundance late-eluting peaks associated with longer FA chain-containing acylsugars. LA1777 accumulated both types of acylsugars.

(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 detected 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 differences in acylsugar composition despite the fact that they are in the same species (compare Figure 4.5b 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 metabolites. 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; Figure 4.6) and principal component analysis (PCA; Figures 4.7 and 4.8). 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 helps make 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 to perform 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 (Figure 4.9b). In addition, the fold differences in total acylsugars in *S. habrochaites* accessions and M82 tomato were modest (Figure 4.10) compared to two previously characterized nearly isogenic *S. pennellii* x M82 introgression lines (ILs 5-3 and 11-3) that accumulate approximately 15-fold reduced amount of total acylsugars than that of M82 (Schilmiller *et al.* 2010a; see 'Variation of total acylsugar levels in 80 *S. habrochaites* accessions' below for details). Taken together, comparative analysis with

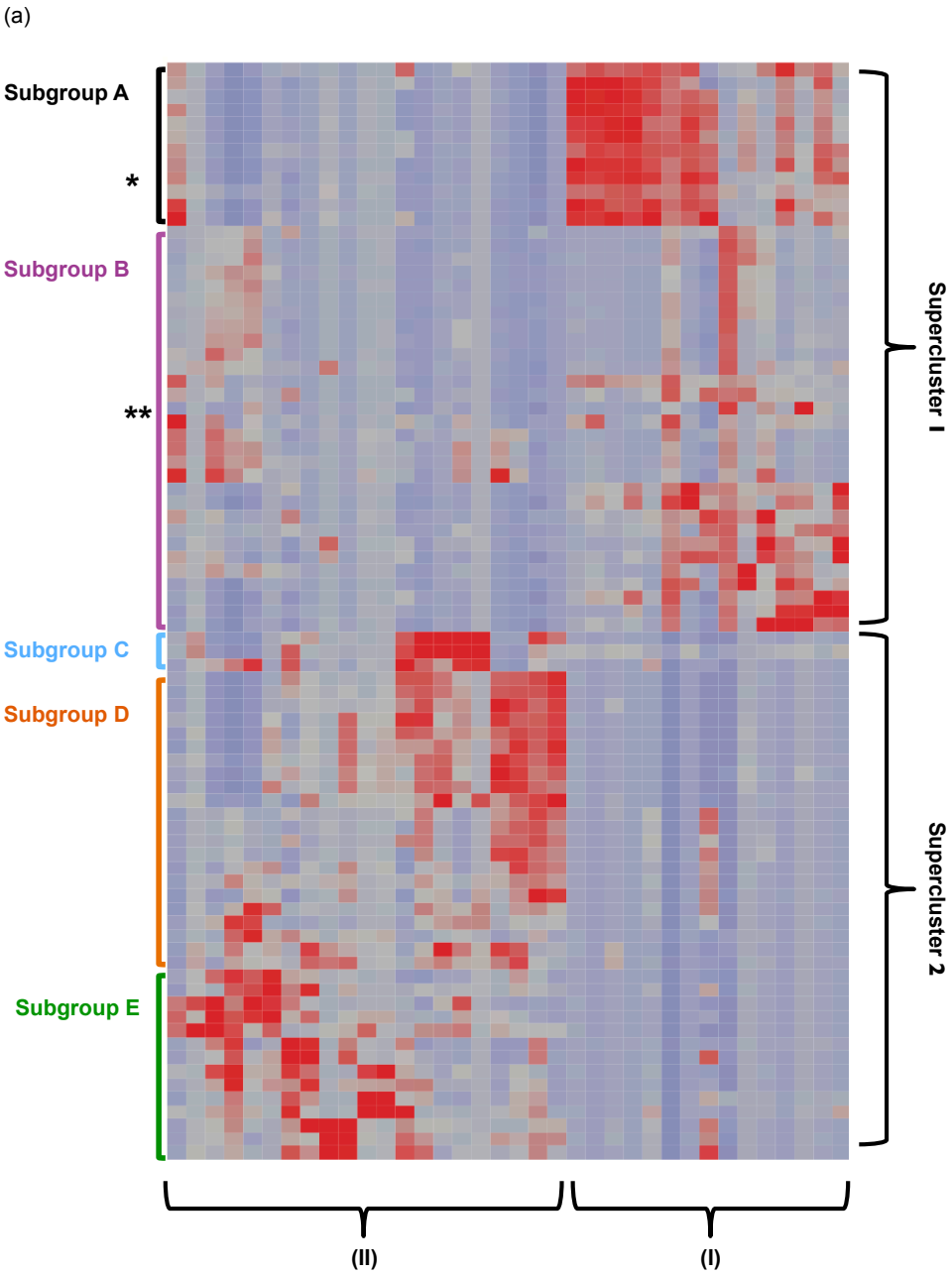

(a)
Subgroup A
*
Subgroup B
**
Subgroup C
Subgroup D
Subgroup E
Supercluster I
Supercluster 2
(II)
(I)

Figure 4.6

Figure 4.6. Continued

(b)

Accession	Sub-group	Accession	Sub-group	Accession	Sub-group	Accession	Sub-group
LA1347	A	LA1731	B	LA1265	C	LA1266	D
LA1361	A	LA2722	B	LA1625	C	LA1718	D
LA1362	A	LA1559	B	LA1624	C	LA2099	D
LA1557	A	LA1648	B	LA1253	D	LA2409	D
LA1393	A	M82	B	LA1255	D	LA1033	E
LA1779	A	LA0361	B	LA2110	D	LA2098	E
LA1392	A	LA1721	B	LA2106	D	LA1391	E
LA1775	A	LA1918	B	LA2107	D	LA2175	E
LA1777	A	LA1928	B	LA2104	D	LA2812	E
LA2574	A	LA1927	B	LA2861	D	LA1352	E
LA2976	A	LA0386	B	LA3864	D	LA2314	E
LA2975	A	LA2552	B	LA3863	D	LA2196	E
LA0094	B	LA1978	B	LA2101	D	LA2204	E
LA1681	B	LA3796	B	LA2105	D	LA2155	E
LA1691	B	LA3794	B	LA2109	D	LA2156	E
LA1295	B	LA4137	B	LA2128	D	LA2100	E
LA1753	B	LA1298	B	LA2119	D	LA2864	E
LA1695	B	LA1764	B	LA2144	D	LA2869	E
LA1696	B	LA2167	B	LA1223	D		
LA1560	B	LA2329	B	LA2860	D		
LA2650	B	LA3854	B	LA0407	D		

Figure 4.6. Continued

(c)

Acylsugar annotation	Group	Acylsugar annotation	Group
S3:13 (4,4,5)	II	S4:14 (2,4,4,4)	I
G3:22 (5,5,12)	II	S4:15 (2,4,4,5)	I
S3:14 (4,5,5)	II	S4:20 (2,4,4,10)	I
S3:15 (5,5,5)	II	S4:21 (2,4,5,10)	I
S3:22 (5,5,12)	II	S3:18 (5,5,8)	I
S4:27 (5,5,5,12)	II	S4:16 (2,4,5,5)	I
S3:20 (5,5,10)	II	S4:22 (2,4,4,12)	I
S4:25	II	S3:19 (5,5,9)	I
S3:21 (5,6,10)	II	S4:17 (2,5,5,5)	I
S4:26 (5,5,5,11)	II	S4:24 (4,5,5,10)	I
S4:25 (5,5,5,10)	II	S4:18 (4,4,5,5)	I
S5:25 (5,5,5,5,5)	II	S4:23 (2,4,5,12)	I
S3:19 (4,5,10)	II	S4:24 (2,5,5,12)	I
S3:20 (4,5,11)	II	S4:22 (2,5,5,10)	I
S3:24 (6,6,12)	II	S4:23 (2,5,5,11)	I
S3:21 (4,5,12)	II		
G3:15 (5,5,5)	II		
S4:19 (4,5,5,5)	II		
S4:20 (5,5,5,5)	II		
S3:21 (5,5,11)	II		
S3:23 (5,6,12)	II		

Figure 4.6. Hierarchical clustering analysis (HCA) of acylsugar chemistry in Solanum accessions.

(a) Normalized signal intensities were used to calculate the percentage composition of 36 individual acylsugars and the dataset analyzed by HCA. Ward's minimum variance method was used with standardization. A red square means the % composition of a specific acylsugar in the given accession was higher than the average for entire samples, a gray square equal level as the average, and a blue square lower than the average. The relative intensities of color indicate the degree of difference between the average signal intensity for entire samples and intensity in the specific accession for the acylsugar of interest. That is, the more dramatic difference exists, the darker color is presented in the square. HCA resulted in 2 superclusters that consist of 2 and 3 subgroups, respectively (subgroups A and B, and subgroups C - E). LA1777 (*) and M82 (**), two relatively well-studied accessions, were marked. See 'Results' for details.

(b) Accessions numbers are listed with subgroup name. Accessions from the top of (a) are listed first followed by ones at the bottom. For example, LA1347 is the first accession on the top and LA2869 is the last one at the bottom in (a).

(c) Acylsugar annotations are listed following the order from left to right in (a). That is, S3:13(4,4,5) and S4:23(2,5,5,11) are the very left and right respectively in (a).

fractional peak area of each acylsugars is an efficient approach to understanding diverse acylsugar profiles across accessions.

As shown in Figure 4.6, HCA revealed two major clusters of trichome acylsugar chemistry ('superclusters' 1 and 2) that consist of 2 and 3 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 tetraacylsucroses containing C2 chains (acylsugar group I listed on the bottom of Figure 4.6; see Figure 4.5a-b for LC-TOF extracted ion chromatograms of M82 and LA1777, respectively). PCA also revealed that this supercluster accumulates tetraacylsucroses (subgroups A and B in Figure 4.7) with C2 FA chains in relatively high abundance (Figure 4.8). In contrast, supercluster 2 is chemically more diverse, and a distinguishing feature is accumulation of triacylsucroses without C2-chains in most of these lines; this is readily seen in the extracted ion chromatogram for LA2106 (Figure 4.5c), and more broadly for accessions in this supercluster in Figures 4.7c and 4.8c. The general trend observed progressing from subgroup A to E is that the proportion of 4S decreases and 3S increases (Figure 4.7).

Many of the dominant differences in trichome metabolites driving the subgroupings within each supercluster are visible in the HCA (Figure 4.6). The two subgroups within the tetraacylsucrose-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

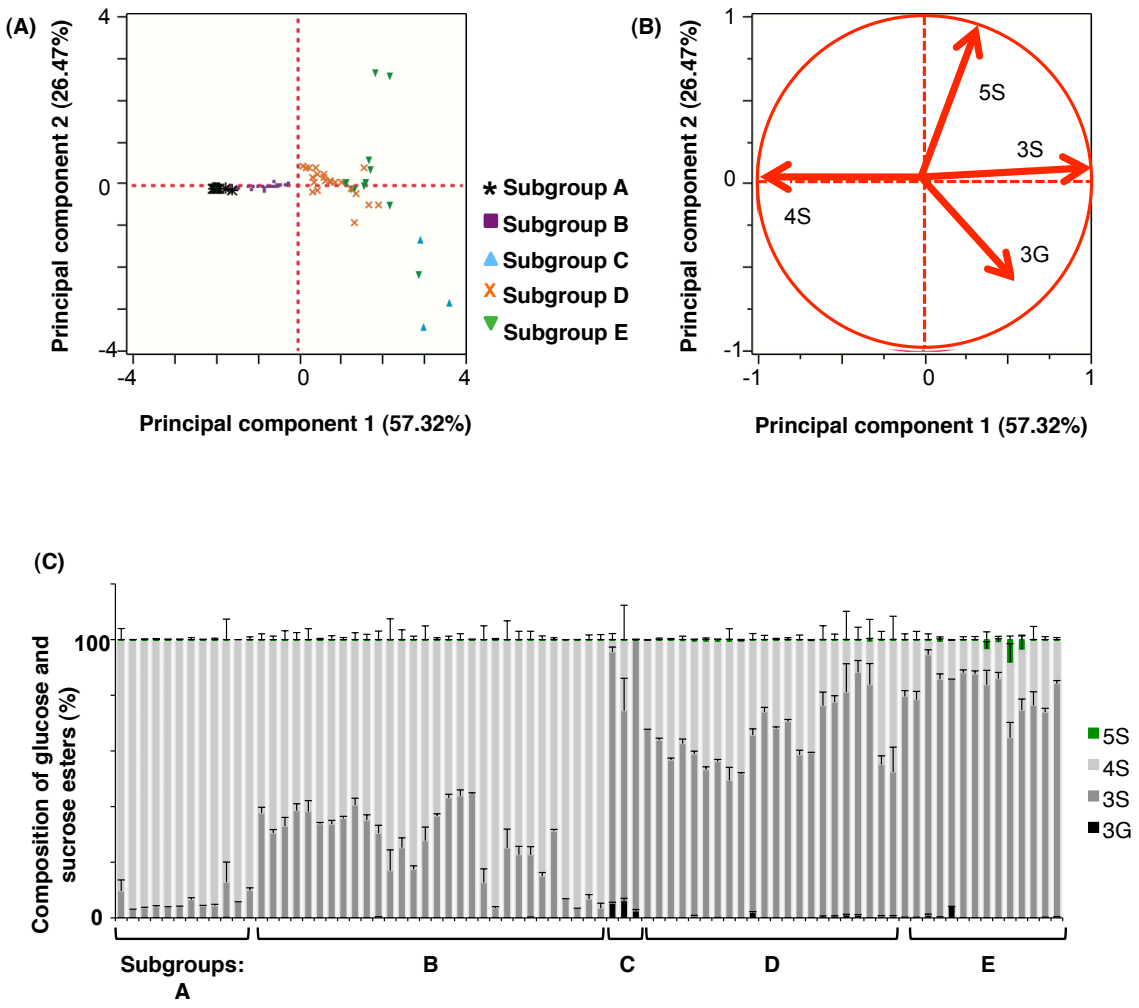


Figure 4.7. Types of glucose and sucrose esters in Solanum accessions.

Figure 4.7. Continued

A-B. Principal component analysis (PCA) was performed based upon the percentage composition of glucose triesters (3G) as well as sucrose triesters (3S), tetraesters (4S) and pentaester (5S), and compared with the acylsugar composition of individual accessions, shown in part C. A. Score plots; B. Loading plots; C. The average % compositions of 4 types of acylsugars are plotted with standard error in individual accessions. The accessions were listed in the identical order as in HCA (Figure 4.6).

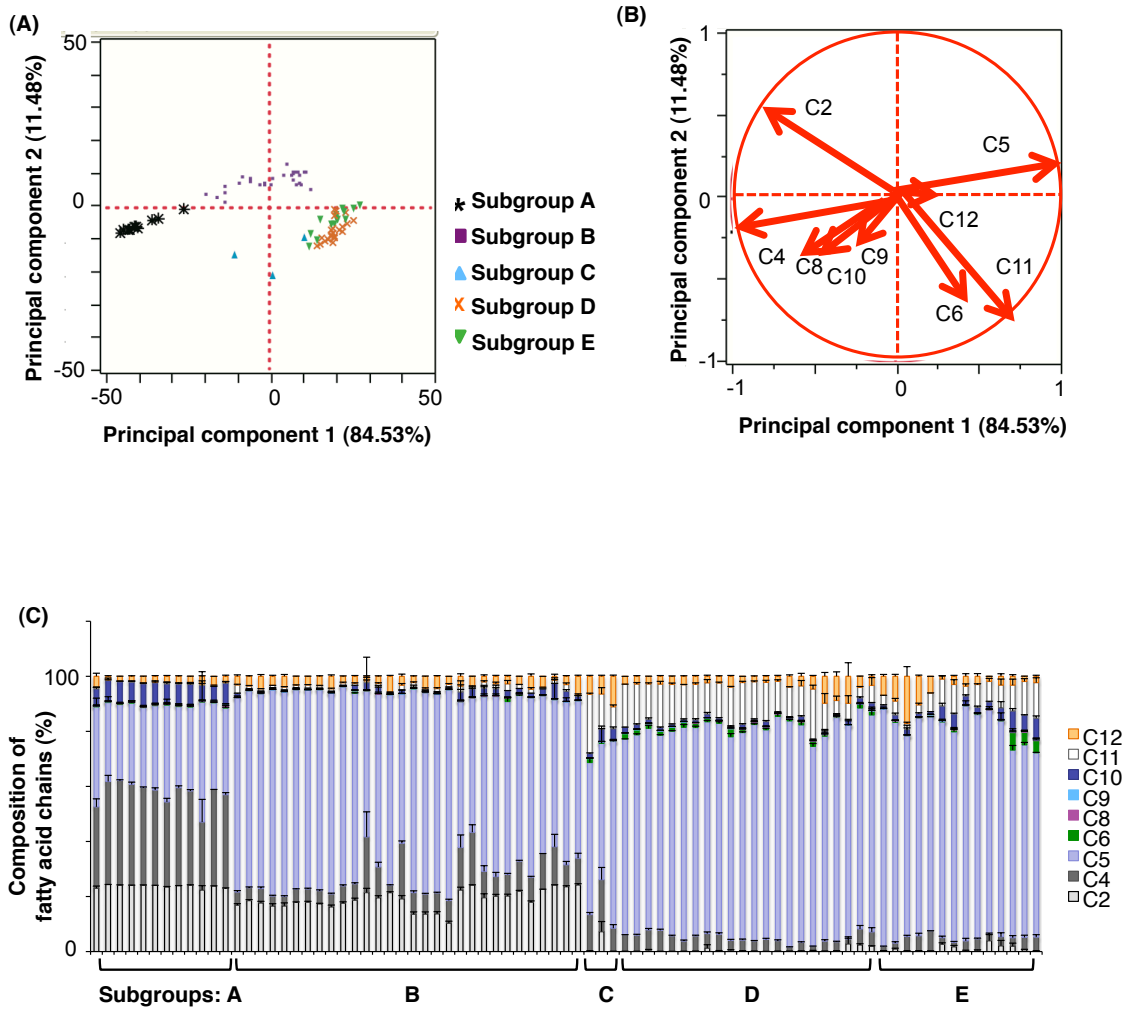


Figure 4.8. The composition of fatty acids in trichome acylsugars.

Figure 4.8. Continued

The percentage composition of the 9 acyl chains, C2, 4, 5, 6, 8, 9, 10, 11 and 12, was calculated from acylsugar profiles and used for PCA. A. Score plots; B. Loading plots; C. The averaged composition of each acyl chain in individual accessions is shown with standard error. The accessions were listed in the identical order as in HCA (Figure 4.6). S4:25 (2,5,6,12) was not included in this analysis due to its relatively low confident annotation (Table 4.2), but it only altered the second place of decimal points compared to the calculation including S4:25 (2,5,6,12) as source of corresponding FAs.

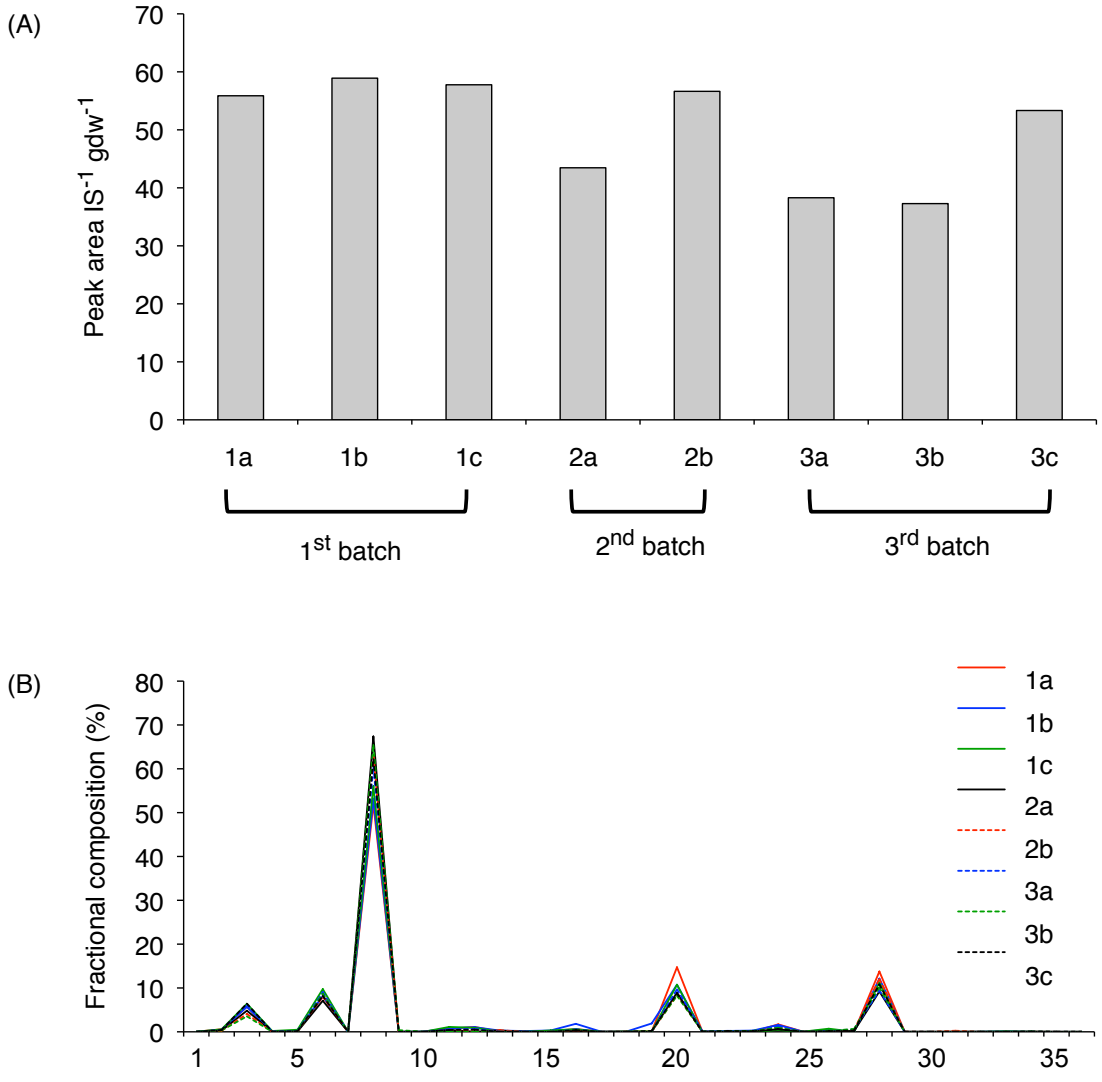


Figure 4.9. Test of reproducibility in acylsugar quantity and composition in eight M82 replicate plants.

Figure 4.9. Continued

Eight M82 replicate plants were grown and chemically analyzed in 3 separate batches with n = 3, 2, 3, respectively.

(A) Total acylsugar amounts in individual M82 plants. The amount of total acylsugars was obtained by summing normalized peak areas for 36 acylsugars. Even within replicate plants from identical batch showed variation in total acylsugar amounts, for example, M82 - 3a, 3b and 3c in the third batch.

(B) Percentage composition of 36 individual acylsugars in each M82 plants. The fractional composition of 36 acylsugars was very similar among 8 replicate plants.

seen in Figure 4.8. 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, 4, 5, 10 and 12) are 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).

Triacylsucrose-dominated accessions in supercluster 2 split into three subgroups (C to E) based upon various acyl chain characteristics. For example, plants in the small group C accumulated characteristic acylsugars including S3:24 (6,6,12), while the C6 FA chain was not commonly observed in other *S. habrochaites* accessions. Group C accessions, LA1285, LA1624 and LA1625, also contained G3:15 (5,5,5) in relatively high abundance (See Figure 4.7 for the trend of G3 accumulation in PCA and its proportion in each accession). Group D showed a shift toward high relative abundance of acylsugars with longer acyl chains ranging from C6 to C12. Although the accessions in this group 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). Group E showed some variation within the group. For example, LA2155 and LA2156 from group E accumulated a pentaacylsucrose (Figure 4.7).

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 (Figures 4.7 and 4.10). This analysis also

revealed the variation of total acylsugar levels across this species, with the total quantity obtained by summing the normalized peak areas for the 36 analytical signals in Table 4.2 as described in the methods section (Figure 4.10). With the exception of LA2101, this analysis showed a maximum eight-fold range of total acylsugars across all accessions, ranging from half to four-fold that of M82.

LA2101 was an exceptionally low accumulator, with approximately one-tenth as much acylsugar detected as M82. This led to the hypothesis 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 (Figure 4.1). These three lines were chosen as accessions that accumulate total acylsugars in relatively low, intermediate and high levels, respectively (Figure 4.10). This analysis did not reveal abnormally low SGT density or aberrant morphology in LA2101 compared to the intermediate and high acylsugar level accessions. This suggests that differences in acylsugar metabolism or storage, 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 (Figure 4.6). Visual inspection of the data in Figure 4.10 suggested that lines in supercluster 1 (subgroups A and B) have higher total acylsugars than supercluster 2. In fact, this analysis revealed that the mean accumulation of total

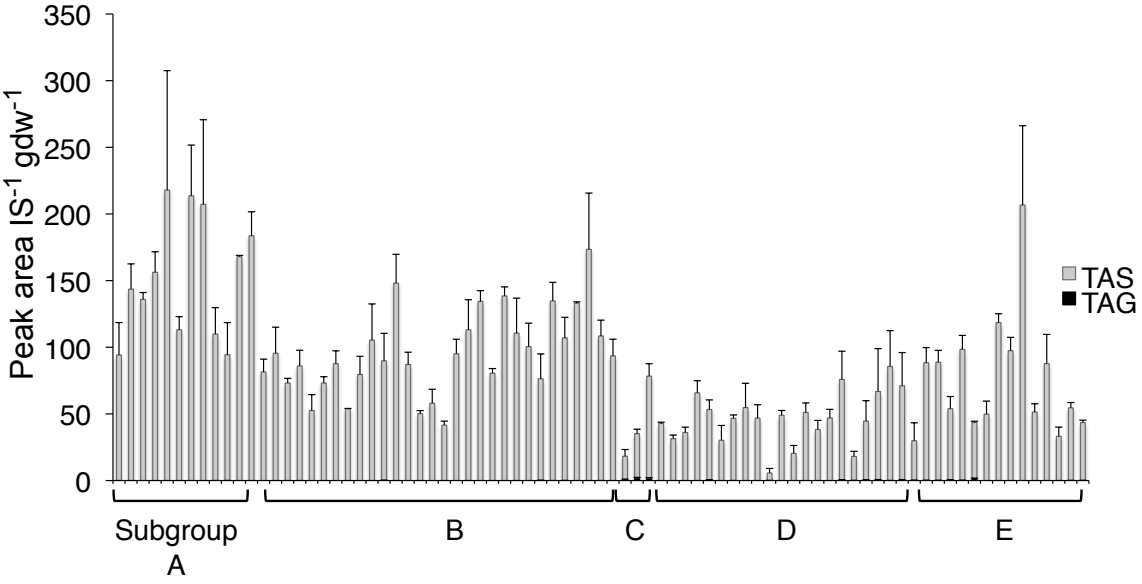


Figure 4.10. Normalized total acylsugar levels in *S. habrochaites* accessions.

The amount of total acylsugars was obtained by summing the normalized peak areas for 34 acylsucroses and 2 acylglucoses. The mean level was calculated from replicates to obtain total acylsucrose (TAS) and total acylglucose (TAG) results, with standard error [*n*=2-3 for individual accessions (except *n*=1 for LA1253, LA1764, LA2812 and LA2976 and *n*=8 for M82)]. Then, the values were further normalized so that total acylsugar amounts can be compared across accessions including those grown and analyzed on different batches. See 'Materials and Methods' for the details. The plants are listed in the same order as in Figure 4.6.

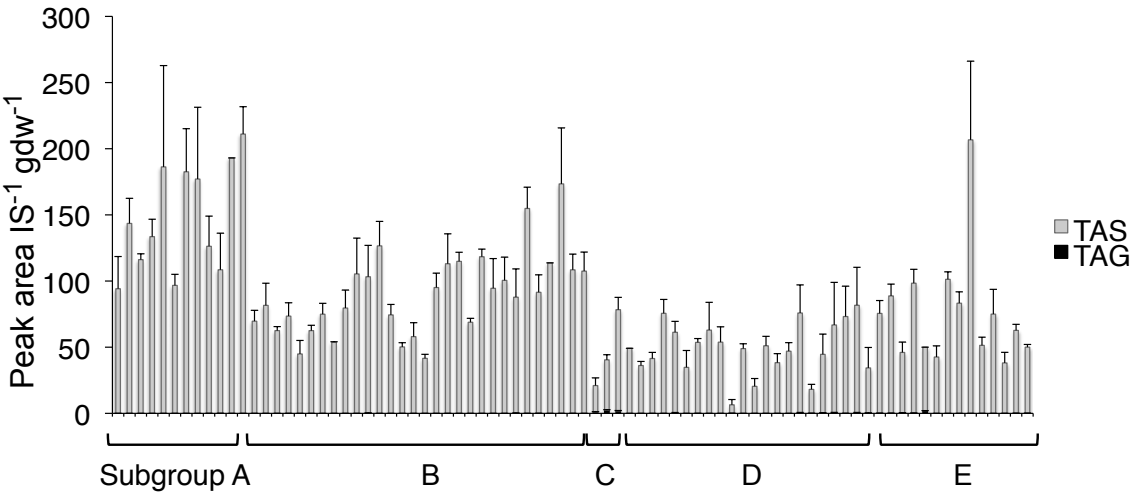


Figure 4.11. Total acylsugar levels in *S. habrochaites* accessions without additional normalization for batch-to-batch comparison.

Total acylsugar levels were calculated using identical set of data used for Figure 4.10. However, while the values for Figure 4.10 were further normalized for batch-to-batch comparison, the additional normalization was not applied in this analysis. However, two analyses, with and without additional normalization, mean results were within ±15%.

acylsugars in accessions from supercluster 1 was approximately 2-fold higher 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 acylsugar chemistry (Figure 4.12). In fact, the C2 sucrose tetraester-containing supercluster1 plants (subgroups A and B) were collected from southern Peru, and supercluster 2 plants (subgroups C, D and E) were distributed in northern Peru and Ecuador.

Discussion

This study revealed that *S. habrochaites* acylsugars are a structurally diverse group of glucose and sucrose-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 biosynthesis and diversity of these molecules is of interest because they have documented roles in plant defense (Kang *et al.* 2010a, Mirnezhad *et al.* 2010, Weinhold and Baldwin 2011), yet little is known about their biosynthesis.

We are taking a multi-faceted 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 (McDowell *et al.* 2011, Schilmiller *et al.* 2010a). More recently we chose *S. habrochaites* for intensive study because of the availability of dozens of accessions, with geographic information about collection sites for many of these lines (http://tgrc.ucdavis.edu/).

Despite the lack of availability of authentic standards 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 (Figure 4.2, Table 4.2, Figure 4.4). All but two are sucrose esters; the glucose triesters, G3:15 (5,5,5) and G3:22 (5,5,12), found in a total of 5 accessions in groups C and E represented ≤ 7% of total acylsugars in these lines (Figures 4.6 and 4.7). Total acylsugar amount was estimated to vary by 8-fold across the 80 accessions (Figures 4.10 and 4.11) and diversity in numbers and lengths of acyl chains was found that was beyond our expectations for accessions of plants in the same species (Figures 4.6-8).

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 dataset. One general observation is that the M82 cultivated tomato had less diverse acylsugar profiles than most of the wild species accessions (see

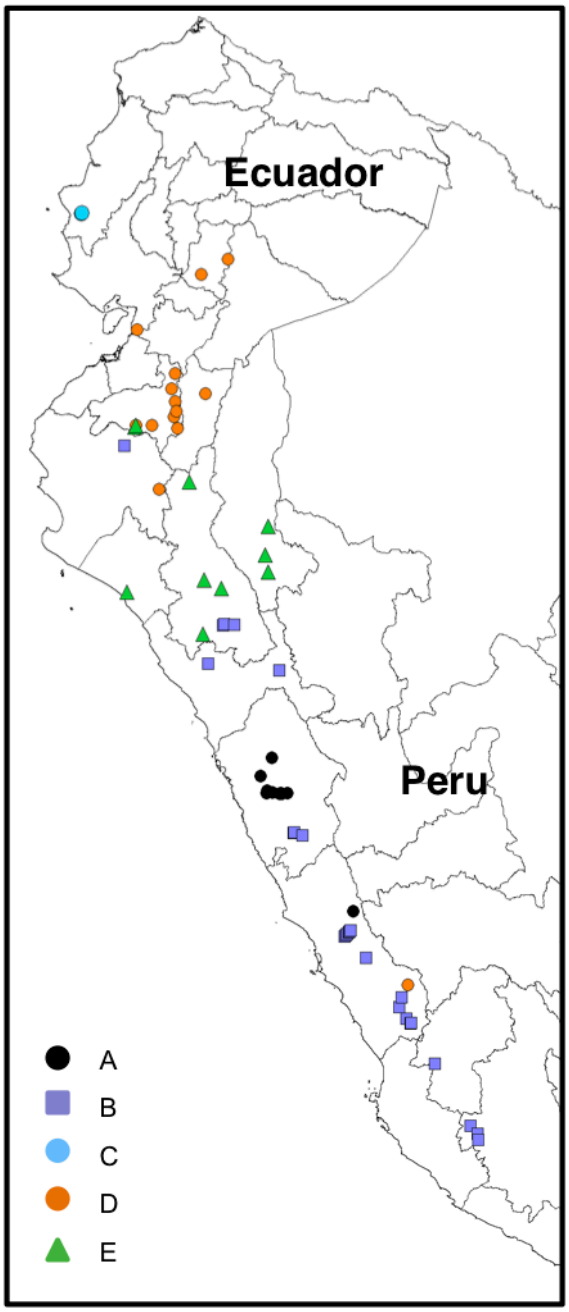


Figure 4.12. Correlation between geographic distribution and acylsugar chemistry in *S. habrochaites* accessions.

Figure 4.12. Continued

Accessions are coded based upon subgroups indentified by HCA (Figure 4). There is little overlap between the two superclusters: Supercluster 1 plants (subgroups A and B), which have C2-containing sucrose tetraesters, were collected in southern Peru, and supercluster 2 plants (subgroups C, D and E) were distributed in northern Peru and Ecuador. Although all accessions whose collection site information was available (Table 1) were plotted, some accessions are almost superimposed with others in this map due to their close geographic location.

Figures 4.5 and 4.6). This is presumably due to genetic bottlenecks introduced during domestication and breeding for *S. lycopersicum* varieties with desirable agronomic properties (Rodriguez-Saona *et al.* 2011). Our analysis revealed several major qualitative differences in SGT acylsugar profiles. A main distinction is the presence or absence of acylsugars containing C2 acyl chains, which is a key feature separating superclusters 1 and 2 (Figures 4.6-8). Within the tetraacylsucrose-dominated supercluster 1, subgroup B accessions mainly have C2, C4 and C5 acylchains, whereas subgroup A acylsugars have a more varied range of acylchain 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 glucose triesters G3:15 (5,5,5) and G3:22 (5,5,12) (Figures 4.6-8).

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 4.2). 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.

Genetic differences in presence or absence of acylsucrose C2 chains were previously found in a comparison between *S. lycopersicum* M82 and the overlapping nearly isogenic chromosome substitution lines IL1-3 and IL1-4. The ILs, accumulated S3:15 (5,5,5) in

high abundance while S4:17 (2,5,5,5) was predominant in M82 parent. This result is consistent with the hypothesis that an acetyltransferase active in *S. lycopersicum* is defective or modified in the *S. pennellii* IL1-3 region, causing lack of enzymatic activity or altered substrate specificity. It is interesting to speculate that the differences in acylsucrose acetylation in *S. habrochaites* supercluster 1 and 2 accessions may be caused by presence or absence of the same acetyltransferase(s). In this scenario, SGT expression of the enzyme responsible for adding the C2 chain to sucrose esters would predate the split leading to *S. habrochaites* and *S. lycopersicum*, and have been lost in some *S. habrochaites* lines.

Furthermore, HCA results helped us speculate possible biosynthetic pathways based on relationships of particular acylsugars in clustering (Figure 4.6). For instance, four major acylsugars, 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), were closely clustered in LA1777. Two acylsugars of high abundance, S4:22 (2,5,5,10) and S4:23 (2,5,5,11), are another example for close chemical relationships (acylsugar group I in Figure 4.6). Based on their close relationships in the clusters, these compounds are thought to behave in a similar manner in LA1777. The results suggest one possibility that these acylsugars share common substrates and/or enzymes for the production of the compounds in LA1777 as proposed in Figure 4.13. Although characterization of enzymes and substrate intermediates is required to test the proposed biosynthetic pathways, our acylsugar profiling study helped us generate hypotheses for further elucidation of acylsugar biochemistry.

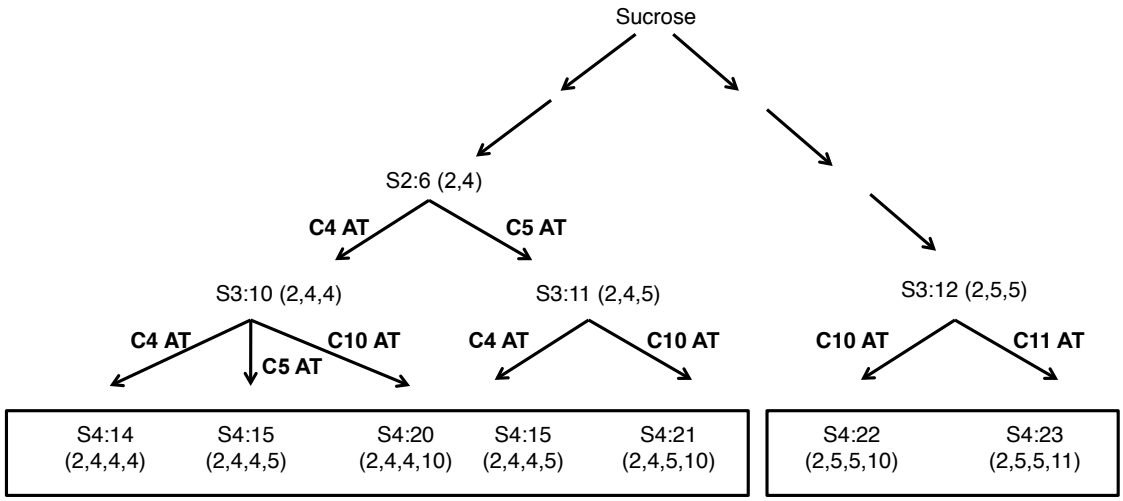


Figure 4.13. Hypothetical biosynthetic pathways to produce a subset of major acylsugars in LA1777.

HCA revealed chemical relationship of acylsugars in *S. habrochaites* accessions (Figure 4.6). In particular, two subsets of highly abundant acylsugars in LA1777 showed close relationships in chemical clustering, 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), and S4:22 (2,5,5,10) and S4:23 (2,5,5,11), respectively (group I in Figure 4.6). Their close relationships in the clustering led to a hypothesis that they share common substrates and/or enzymes for the production. Hypothetical biosynthetic pathways leading to the production of these acylsugars in LA1777 were proposed, assuming the reactions are sequential. For chemical compounds, identical nomenclature was used as Table 4.2. 'AT' indicates acyltransferase.

Geographical Context

Our results 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 reported that accessions from the southern part of the range, including the widely-studied LA0716, predominately accumulate glucose esters with C4-C12 chains. As *S. pennellii* accessions were sampled going from south to north of their range in the Andes Mountains, they found an increase in the ratio of sucrose to glucose esters. Our results show that there is a strong prevalence of C2-containing tetraacylsucroses from southern *S. habrochaites* accessions (supercluster 1), whereas supercluster 2 accessions predominated from northern Peru and Ecuador. 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.

Summary

LC-TOF-MS screening of *S. habrochaites* SGT acylsugars revealed diverse acylsugar chemotypes. The varied lengths of acyl-chains and combinations of numbers and types of esters in the 36 acylsugar structures suggests the interplay of a varied set of enzymes in *S. habrochaites* acylsugar biosynthesis. Knowing which metabolites accumulate in specific accessions − for instance presence or absence of C2 esters, accumulation of unusual length acyl chains such as C6 or C8 or synthesis of penta-acylsucrose S5:25 (5,5,5,5,5) − will inform selection of accessions to study the biosynthesis and functions of acylsugars. Once the genes for these enzymes are known, it should 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.

Materials and methods

Plant growth, metabolite extraction and identification.

Germplasms for M82 cultivated tomato and 80 *S. habrochaites* accessions were obtained from the Tomato Genetic Resource Center (http://tgrc.ucdavis.edu/). Plant seedlings were grown and non-volatile leaf surface and trichome metabolites were extracted as previously described in Schilmiller *et al.* 2010a. Two to three biological replicates (individual plants grown in the same growth chamber) were prepared for most *S. habrochaites* accessions and 8 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).

Chemical analysis by LC-TOF-MS

LC-TOF-MS (Shimadzu LC-20AD pumps; Waters LCT Premier mass spectrometer) was used for the chemical analysis of non-volatile metabolites by modification of published methods (Gu *et al.* 2010, Schilmiller *et al.* 2010a). Ten microliters of extract was injected into a BetaBasic C18 HPLC column (1 x 150 mm with 3.0 µm particles, Thermo Scientific, http://www.thermoscientific.com/) maintained at 35°C. Two solvents were used as mobile phase (solvent A – 0.15% formic acid in MilliQ water and solvent B –

methanol), with a 0.05ml/min flow rate. The gradient profile was as follows: 5% B for initial step; linear gradient to 50% B at 5 min; linear gradient to 95% B at 33 min; 100% B from 33-38 min; 5% B at 43 min. TOF-MS with multiplexed collision-induced dissociation was performed by switching Aperture 1 voltages (10, 25, 40, 55 and 80 V) for quasi-simultaneous production of spectra without and with fragment ions. Lowest energy spectra were used for quantification of analytical signals, while spectra generated at 55V were used for metabolite structural information.

Characterization of analytical signals for acylsugars and their metabolite annotation

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 mass-to-charge ratio (m/z) under low ionization conditions. Fragment ions produced from high collision energies in MS were used for metabolite annotation.

Specifically, the analytical signals were obtained from LC-TOF-MS analysis with 43min LC separation and m/z 50-1,500 MS detection range. The signals were collected using MarkerLynx software (Waters, http://www.waters.com/). The relative mass defect of each resulting signal was calculated to identify acylsugar metabolites. Mass defect was computed with accurate mass information obtained from TOF following the equation, [(experimental m/z – nominal value of m/z) ÷ experimental m/z] x 10^6 (ppm). As relative mass defect is correlated with the number of hydrogen atoms in the compound, it is helpful in categorizing the class of chemical compounds (Last *et al.* 2007, Stagliano *et al.*

2010). In this study compounds with relative mass defect between 200-700ppm, as an inclusive cut-off, were used to exclude analytical signals that are not associated with acylsugars. Known acylsugars from the Solanacea have relative mass defects in the range of 300-600 ppm, whereas earlier eluting compounds with < 300ppm often correspond to flavonoids that have lower hydrogen content. In contrast, membrane lipids characteristically have signals with relative mass defects > 600ppm.

LC retention times were used in addition to accurate mass to define analytical signals. Acylsugars eluted after glycosylated alkaloid tomatine from the reverse-phase analysis in HPLC due to lower aqueous solubility of acylsugars. Furthermore, a previous study reported a list of m/z for acylsugar-associated analytical signals detected in relatively high abundance in M82 cultivated tomato (Schilmiller *et al.* 2010). The combination of LC retention time and m/z list enabled identification of 36 analytical signals that correspond to acylsugars in *S. habrochaites* accessions. From the previous study, m/z 821 was the largest value found for acylsugar-associated signals in M82 cultivated tomato. Thus, analytical signals with successive increase of m/z 14 (corresponding to methylene groups) were searched from m/z 821 to query whether there are any potential signals for novel acylsugars detected in *S. habrochaites* accessions. Signals with incremental decrease of m/z 14 were also searched from the minimal value reported for cultivated tomato. However, neither approach found additional acylsugar-associated analytical signals.

For the 36 analytical signals identified as acylsugars, metabolite annotation was obtained based on co-eluting fragment ions generated at aperture 1 higher voltage conditions. Information obtained from this analysis included the number of acyl chains and lengths of individual acyl chains. As a result, analytical signals for 34 acylsucroses and 2 acylglucoses were identified (Table 4.2).

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 be identical compounds by assessing fragment ion masses generated using elevated collision energies. However, when a number of peaks with same m/z were seen within 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.

An estimate of the amount of total acylsugars was obtained by summing the normalized peak areas for 36 analytical signals (Table 4.2). As the chemical analysis for 80 accessions was performed in 3 different batches with separate plant growth, harvests and LC-MS performances, the value was further normalized to make a batch-to-batch comparison possible (Figure 4.10). Every batch included M82 cultivated tomato plants as

control (n=8 in total; 3, 2, 3 per planting and analysis, respectively), and their profiles were qualitatively similar with minor variation of quantity (Figure 4.9). Thus, the average of all 8 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 percent 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 sucrose- and glucose 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 acylsucroses or acylglucoses. Acylsucroses were further grouped to 3S, 4S or 5S categories according to the number of acyl chains. 3G was only type of acylglucose identified in this study. The peak areas for each category were summed, their percent composition was calculated and PCA performed (Figure 4.7).

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 level of C2, 4 and 10 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 12, and then the percent of each FAs was calculated. Then, the composition was analyzed by PCA (Figure 4.8).

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 (Table 4.1). Locations were mapped using DIVA-GIS software (http://www.diva-gis.org/).

References

References

Bohlmann, J. and Gershenzon, J. (2009) Old substrates for new enzymes of terpenoid biosynthesis. *Proc. Natl. Acad. Sci. U S A*, **106**, 10402-10403.

Fobes, J.F., Mudd, J.B. and Marsden, M.P.F. (1985) Epicuticular lipid accumulation on the leaves of *Lycopersicon pennellii* (Corr.) D'Arcy and *Lycopersicon esculentum* Mill. *Plant Physiol.*, **77**, 567-570.

Fridman, E., Wang, J., Iijima, Y., Froehlich, J.E., Gang, D.R., Ohlrogge, J. and Pichersky, E. (2005) Metabolic, genomic, and biochemical analyses of glandular trichomes from the wild tomato species *Lycopersicon hirsutum* identify a key enzyme in the biosynthesis of methylketones. *Plant Cell*, **17**, 1252-1267.

Ghangas, G.S. and Steffens, J.C. (1993) UDPglucose: fatty acid transglucosylation and transacylation in triacylglucose biosynthesis. *Proc. Natl. Acad. Sci. U S A*, **90**, 9911-9915.

Gu, L., Jones, A.D. and Last, R.L. (2010) Broad connections in the Arabidopsis seed metabolic network revealed by metabolite profiling of an amino acid catabolism mutant. *Plant J.*, **61**, 579-590.

Kandra, G., Severson, R. and Wagner, G.J. (1990) Modified branched-chain amino acid pathways give rise to acyl acids of sucrose esters exuded from tobacco leaf trichomes. *Eur. J. Biochem.*, **188**, 385-391.

Kang, J.H., Liu, G., Shi, F., Jones, A.D., Beaudry, R.M. and Howe, G.A. (2010a) The tomato *odorless-2* mutant is defective in trichome-based production of diverse specialized metabolites and broad-spectrum resistance to insect herbivores. *Plant Physiol.*, **154**, 262-272.

Kang, J.H., Shi, F., Jones, A.D., Marks, M.D. and Howe, G.A. (2010b) Distortion of trichome morphology by the *hairless* mutation of tomato affects leaf surface chemistry. *J. Exp. Bot.*, **61**, 1053-1064.

Kennedy, G.G. (2003) Tomato, pests, parasitoids, and predators: tritrophic interactions involving the genus Lycopersicon. *Annu. Rev. Entomol.*, **48**, 51-72.

Last, R.L., Jones, A.D. and Shachar-Hill, Y. (2007) Towards the plant metabolome and beyond. *Nat. Rev. Mol. Cell. Biol.*, **8**, 167-174.

Liu, B., Wang, H., Du, Z., Li, G. and Ye, H. (2011) Metabolic engineering of artemisinin biosynthesis in *Artemisia annua L. . Plant. Cell. Rep.*, **30**, 689-694.

Luckwill, L. (1943) *The genus Lycopersicon: A historical, biological, and taxonomic survey of the wild and cultivated tomatoes* Aberdeen, Scotland: Aberdeen University Press.

McDowell, E.T., Kapteyn, J., Schmidt, A., Li, C., Kang, J.H., Descour, A., Shi, F., Larson, M., Schilmiller, A., An, L., Jones, A.D., Pichersky, E., Soderlund, C.A. and Gang, D.R. (2011) Comparative functional genomic analysis of Solanum glandular trichome types. *Plant Physiol.*, **155**, 524-539.

Milligan, G.W. (1980) An examination of the effect of 6 types of error perturbation on 15 clustering algorithms. *Psychometrika*, **45**.

Mirnezhad, M., Romero-Gonzalez, R.R., Leiss, K.A., Choi, Y.H., Verpoorte, R. and Klinkhamer, P.G. (2010) Metabolomic analysis of host plant resistance to thrips in wild and cultivated tomatoes. *Phytochem. Anal.*, **21**, 110-117.

Rodriguez-Saona, C., Vorsa, N., Singh, A.P., Johnson-Cicalese, J., Szendrei, Z., Mescher, M.C. and Frost, C.J. (2011) Tracing the history of plant traits under domestication in cranberries: potential consequences on anti-herbivore defences. *J. Exp. Bot.*, **62**, 2633-2644.

Sallaud, C., Rontein, D., Onillon, S., Jabes, F., Duffe, P., Giacalone, C., Thoraval, S., Escoffier, C., Herbette, G., Leonhardt, N., Causse, M. and Tissier, A. (2009) A novel pathway for sesquiterpene biosynthesis from Z,Z-farnesyl pyrophosphate in the wild tomato Solanum habrochaites. *Plant Cell*, **21**, 301-317.

Schilmiller, A., Shi, F., Kim, J., Charbonneau, A.L., Holmes, D., Daniel Jones, A. and Last, R.L. (2010a) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. *Plant J.*, **62**, 391-403.

Schilmiller, A.L., Last, R.L. and Pichersky, E. (2008) Harnessing plant trichome biochemistry for the production of useful compounds. *Plant J.*, **54**, 702-711.

Schilmiller, A.L., Miner, D.P., Larson, M., McDowell, E., Gang, D.R., Wilkerson, C. and Last, R.L. (2010b) Studies of a biochemical factory: tomato trichome deep expressed sequence tag sequencing and proteomics. *Plant Physiol.*, **153**, 1212-1223.

Schilmiller, A.L., Schauvinhold, I., Larson, M., Xu, R., Charbonneau, A.L., Schmidt, A., Wilkerson, C., Last, R.L. and Pichersky, E. (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc. Natl. Acad. Sci. U S A*, **106**, 10865-10870.

Shapiro, J.A., Steffens, J.C. and Mutschler, M.A. (1994) Acylsugars of the wild tomato *Lycopersicon pennellii* in relation to geographic distribution of the species. *Biochem. Syst. Ecol.*, **22**, 545-561.

Shepherd, R.W., Bass, W.T., Houtz, R.L. and Wagner, G.J. (2005) Phylloplanins of tobacco are defensive proteins deployed on aerial surfaces by short glandular trichomes. *Plant Cell*, **17**, 1851-1861.

Sifres, A., Blanca, J. and Nuez, F. (2011) Pattern of genetic variablity of *Solanum habrochaites* in its natural area of distribution. *Genet. Resour. Crop. Ev.,* **58**, 347-360.

Slocombe, S.P., Schauvinhold, I., McQuinn, R.P., Besser, K., Welsby, N.A., Harper, A., Aziz, N., Li, Y., Larson, T.R., Giovannoni, J., Dixon, R.A. and Broun, P. (2008) Transcriptomic and reverse genetic analyses of branched-chain fatty acid and acyl sugar production in *Solanum pennellii* and *Nicotiana benthamiana*. *Plant Physiol.*, **148**, 1830-1846.

Stagliano, M.C., DeKeyser, J. G., Omiecinski, C. J., and Jones, A. D. (2010) Bioassay directed fractionation for discovery of bioactive neutral lipids guided by relative mass defect filtering and multiplexed collision induced dissociation. . *Rapid Commun. Mass Spectrom.*, **24**, 3578-3584.

Thipyapong, P., Joel, D.M. and Steffens, J.C. (1997) Differential expression and turnover of the tomato polyphenol oxidase gene family during vegetative and reproductive development. *Plant Physiol.*, **113**, 707-718.

Wagner, G.J. (1991) Secreting glandular trichomes: more than just hairs. *Plant Physiol.*, **96**, 675-679.

Walters, D.S. and Steffens, J.C. (1990) Branched chain amino acid metabolism in the biosynthesis of *Lycopersicon pennellii* glucose esters. *Plant Physiol.*, **93**, 1544-1551.

Wang, G., Tian, L., Aziz, N., Broun, P., Dai, X., He, J., King, A., Zhao, P.X. and Dixon, R.A. (2008) Terpene biosynthesis in glandular trichomes of hop. *Plant Physiol.*, **148**, 1254-1266.

Weinhold, A. and Baldwin, I.T. (2011) From the cover: trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation. *Proc. Natl. Acad. Sci. U S A*, **108**, 7855-7859.

Werker, E. (2000) Trichome diversity and development. In *Advances in Botanical Research Incorporating Advances in Plant Pathology*, pp. 1-35.

Yu, G., Nguyen, T.T., Guo, Y., Schauvinhold, I., Auldridge, M.E., Bhuiyan, N., Ben-Israel, I., Iijima, Y., Fridman, E., Noel, J.P. and Pichersky, E. (2010) Enzymatic functions of wild tomato methylketone synthases 1 and 2. *Plant Physiol.*, **154**, 67-77.

Yu, H., Kowalski, S.P. and Steffens, J.C. (1992) Comparison of polyphenol oxidase expression in glandular trichomes of Solanum and Lycopersicon species. *Plant Physiol.*, **100**, 1885-1890.

Zhou, B., Li, Y., Xu, Z., Yan, H., Homma, S. and Kawabata, S. (2007) Ultraviolet A-specific induction of anthocyanin biosynthesis in the swollen hypocotyls of turnip (*Brassica rapa*). *J. Exp. Bot.*, **58**, 1771-1781.