University of São Paulo
"Luiz de Queiroz" College of Agriculture

Introgression of natural genetic variation affecting trichome development and terpenes biosynthesis to obtain insect-resistant tomatoes

Rodrigo Therezan de Freitas

Thesis presented to obtain the degree of Doctor in Science. Area: Plant Physiology and Biochemistry

Piracicaba
2019

Rodrigo Therezan de Freitas
Bachelor in Agronomy

Introgression of natural genetic variation affecting trichome development and terpenes biosynthesis to obtain insect-resistant tomatoes

Advisor:
Prof. Dr. **LÁZARO EUSTÁQUIO PEREIRA PERES**

Thesis presented to obtain the degree of Doctor in Science. Area: Plant Physiology and Biochemistry

Piracicaba
2019

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Freitas, Rodrigo Therezan de

Introgression of natural genetic variation affecting trichome development and terpenes biosynthesis to obtain insect-resistant tomatoes/ Rodrigo Therezan de Freitas. - - Piracicaba, 2019.

96 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Tomate 2. Linhagem de introgressão 3. Sesquiterpenos 4. Resistência à inseto 5. Glandular tricomas I. Título

ACKNOWLEDGEMENTS

Foremost, I would like to express my gratitude to my advisor Prof. Dr. Lázaro Peres for allowing me to be part of their team, for their guidance, advise, involvement, and support during my research.

Besides my advisor, I am deeply grateful also to Dr. Petra Bleeker and Dr. Rob Schuurink, who provided me an opportunity to join their team as a visiting student at the University of Amsterdam, and for their important contribution to this project. A special thank should be given to Ruy Kortbeek, Michelle van der Gragt, Marc Galland, Saioa Legarrea and Pietro Zocca for kindly daily help in the lab.

My sincere thanks also goes to Prof. Severino Matias de Alencar who gave me access to his laboratory, for his collaboration and help during the experiments. I also need to thanks Adna Prado and Anna Paula de Souza for their support and assistance making possible the development of my experiments.

I also would like to thank my labmates, Karla Gasparini, Lilian Pino, Ariadne Felício, Mayara Carvalho, Diego Sevastian, Eloisa Vendemiatti, Maísa Pinto, Marcela Notini, Gabriela Prants, and Mateus Vicente for all knowledge shared and for all the fun we have had in the last years.

My sincere thanks also goes to Cassia Figueiredo and Francisco Vitti for their assistance, making possible the development of many experiments.

I would like to thank "Luiz de Queiroz" College of Agriculture for the opportunities granted to me along these years, and the Plant Physiology and Biochemistry Program, especially Maria Solizete Granziol Silva for the help, support and availability anytime.

I truly thank CAPES (Agency for the Support and Evaluation of Graduate Education) for the financial support throughout my doctorate studies, which made possible my studies in Brazil and my staying in the Netherlands.

Last but not least, I would like to thank my family, especially my parents Dulce and Osvaldo, for their love, prayers, caring and sacrifices for educating and preparing me for my future. I also express my thanks to my sister Evelyn and my cousin Daiane for their support and encouragement.

CONTENTS

RESUMO

Introgressão de variações genéticas naturais que afetam o desenvolvimento de tricomas e a biossíntese de terpenos para obtenção de tomates resistentes a insetos

Para lidar com insetos pragas, as espécies selvagens do tomateiro produzem uma variedade de compostos de defesas em seus tricomas glandulares do tipo-VI. Por outro lado, embora as espécies cultivadas (*Solanum lycopersicum* L.) também apresentem tricomas do tipo-VI, a glândula do tomate cultivado é muito menor, contendo principalmente monoterpenos e baixos níveis de sesquiterpenos derivados do citosol, que não têm efeito aparente contra os insetos pragas. No presente trabalho, dividido em dois capítulos, realizamos cruzamentos entre a espécie selvagem *S. habrochaites* e a cultivar de tomate Micro-Tom (MT), afim de se criar duas linhas de introgressão (ILs). No primeiro capítulo, transferimos com sucesso a via do sesquiterpeno derivado do plastídeo de *S. habrochaites* para tricomas do tipo-VI do tomate cultivado (cv. Micro-Tom). Os tricomas da linhagem introgredida denominada MT-*Sesquiterpeno-sintase 2* (MT-*Sst2*) apresentaram concentrações ainda maiored de α-santaleno, β-bergamoteno e α-bergamoteno em comparação com os tricomas glandulares do tipo-VI da espécie selvagem. Surpreendentemente, a presença de grandes quantidades de sesquiterpenos derivados do plastídio ainda assim não foi suficiente para conferir resistência as pragas específicas de tomate no MT-*Sst2*. O perfil de sesquiterpeno de MT-*Sst2* e LA1777 revela compostos derivados de sesquiterpeno encontrados apenas na espécie selvagen, o que aponta para a necessidade de etapas adicionais para obteção de tomate resistente a insetos. No segundo capítulo, testamos a hipótese de que o formato distinto do tricoma tipo-VI na espécie selvagen *S. habrochaites* é causado por uma combinação de genes que regulam o desenvolvimento e o metabolismo do tricoma. Utilizando uma linha quase isogênica chamada MT-*Pincushion-like* (MT-*Pik*), mostramos que as plantas MT-*Pik* carregam parte da variação alélica responsável pela morfologia diferencial desses trichomas. Também demonstramos que a transferência da capacidade de biossíntese de sesquiterpenos derivados de plastídeos de *S. habrochaites* para tricomas do tipo-VI do MT-*Pik* causa um aumento sinérgico na cavidade da glândula. Este trabalho abre caminho para o entendimento da morfologia e funcionalidade dos tricomas do tipo-VI e para a criação de tomates resistentes a insetos.

Palavras-chave: Tomate; Linhagem de introgressão; Tricoma glandular; Resistência à inseto; Sesquiterpeno.

ABSTRACT

Introgression of natural genetic variations affecting trichome development and terpenes biosynthesis to obtain insect-resistant tomatoes

To deal with insect pests the tomato wild relatives produce a variety of defensive compounds in their glandular trichomes type-VI. By contrast, although cultivated tomatoes (*Solanum lycopersicum* L.), also display type-VI trichomes, the gland in cultivated tomato is much smaller containing mainly monoterpenes and low levels of cytosolic-derived sesquiterpenes, which have no apparent effect against insect pests. In the present work, which was divided into two chapters, we carried out crosses between the wild species *S. habrochaites* and the tomato cultivar Micro-Tom (MT) in order to create two introgression lines (ILs). In the first chapter we successfully transferred the plastid-derived sesquiterpene pathway from *S. habrochaites* to type-VI trichomes of the cultivated tomato (cv. Micro-Tom). The trichomes of the introgressed line named MT-*Sesquiterpene synthase 2* (MT-*Sst2*) showed even higher concentration of α-santalene, β-bergamotene, and α-bergamotene compared to the wild species type-VI glandular trichomes. Surprisingly, the presence of high amounts of plastid-derived sesquiterpenes was not sufficient to confer resistance to specific tomato pests in MT-*Sst2*. The sesquiterpene profile of MT-*Sst2* and LA1777 unveils sesquiterpene-derived compounds only found in the wild species, which point for additional steps necessary to obtain insect-resistant tomatoes. In the second chapter, we tested the hypothesis that the distinct trichomes of the wild species are caused by a combination of genes regulating both trichome development and metabolism. Using a near isogenic line named MT-*Pincushion-like* (MT-*Pik*) we showed that MT-*Pik* plants carry part of the allelic variation responsible for the differential morphology of type-VI trichomes presented in the parent *S. habrochaites*. We further demonstrated that transferring plastid-derived sesquiterpenes biosynthesis capacity from *S. habrochaites* to type-VI trichomes of MT-*Pik* causes a synergistic increase in gland cavity. This work paves the way for both the understanding of the morphology and functionality of type-VI trichomes and for breeding insect resistant tomatoes.

Keywords: Tomato; Introgression line; Glandular trichomes; Insect resistance; Sesquiterpenes.

1. INTROGRESSION OF A NATURAL GENETIC VARIATION FOR PLASTID DERIVED SESQUITERPENE SYNTHESIS UNVEILS ADDITIONAL STEPS IN THE PURSUIT OF INSECT-RESISTANT TOMATOES

Abstract

To deal with insect pests the tomato wild relatives produce a variety of defensive compounds in their glandular trichomes. In *Solanum habrochaites* LA1777, a functional cluster of genes on chromosome 8 controls plastid-derived sesquiterpene synthesis not found in cultivated tomatoes. The main genes at the cluster are *Z-prenyltransferase* (*zFPS*) that produces *Z-Z*-farnesyl diphosphate (*Z,Z*-FPP), and *Santalene and Bergamotene Synthase* (*SBS*) that uses *Z,Z*-FPP to produce α-santalene, β-bergamotene, and α-bergamotene in type-VI glandular trichomes. Although both LA1777 and cultivated tomatoes display type-VI trichomes, the gland in cultivated tomato is much smaller containing mainly monoterpenes and low levels of cytosolic-derived sesquiterpenes, which have no apparent effect against insect pests. Herein, we successfully transferred the plastid-derived sesquiterpene pathway from LA1777 to type-VI trichomes of a cultivated tomato (cv. Micro-Tom) by a back-crossing approach. The trichomes of the introgressed line named MT-*Sesquiterpene synthase 2* (MT-*Sst2*) showed even higher concentration of α-santalene, β-bergamotene, and α-bergamotene compared to the wild species type-VI glandular trichomes. We also noticed that the type-VI trichome internal storage-cavity size increases in MT-*Sst2*, probably as an "inflated balloon" effect of the increased amount of sesquiterpenes. Surprisingly, the presence of high amounts of plastid-derived sesquiterpenes was not sufficient to confer resistance to specific tomato pests in MT-*Sst2*. The sesquiterpene profile of MT-*Sst2* and LA1777 unveils sesquiterpene-derived compounds only found in the wild species, which point for additional steps necessary to obtain insect-resistant tomatoes. Our results also provide for the understanding of the morphology of *S. habrochaites* type-VI trichomes.

Keywords: Tomato; Introgression line; Trichome; Terpenes; Santalene.

1.1 Introduction

Terpenes are the most abundant and diverse class of compounds produced by plants with a wide variety of biological functions (Dudareva et al., 2013). They are produced by multiples of isoprene (C5) units. Isoprenoids are essential for plant growth and development. They participate as precursors for several components of photosynthesis, respiration, cell cycle control (Estévez et al., 2001) and plant hormones such as gibberellins, abscisic acid, brassinosteroids and strigolactones (Falara et al., 2011). Isoprenoids also play an important role in the interactions of plants with the environment, including defense against herbivorous insects and pollinators attraction (Dudareva et al., 2013).

In plants, all terpenes originate from two distinct metabolic pathways: the mevalonate (MVA) located in the cytosol and the 2-C-methyl-D-erythritol 4-phosphate (MEP) located in plastids. Both pathways produce and use isopentenyl diphosphate (IPP) and dimethylallyl

diphosphate (DMAPP) to produce the isoprene building blocks used by terpene synthases (TPSs) to catalyze the formation of C10 monoterpenes, C15 sesquiterpenes or C20 diterpenes (Tholl, 2006).

In cultivated tomato (*Solanum lycopersicum*), sesquiterpene biosynthesis usually takes place in the cytosol through the MEV pathway. However, in glandular trichomes of the tomato-related wild species *S. habrochaites,* sesquiterpenes are also produced in the chloroplasts (Sallaud et al., 2009). The presence of plastid-derived sesquiterpenes in some wild species has been described to be responsible for the lower damage by insects, making these wild species naturally resistant to multiple pests such as lepidopterans (Eigenbrode et al., 1994; Eigenbrode et al., 1996), whiteflies (*Bemisia spp.*) (Bleeker et al., 2012) and also the herbivorous spider mites (Maluf et al., 2001).

Two independent loci have been associated with the biosynthesis of two different classes of sesquiterpenes in *S. habrochaites*. The *Sesquiterpene synthase 1* (*SsT1*) locus on chromosome 6 is responsible for the accumulation of class I cytosolic-derived sesquiterpenes. At this locus, the S. *lycopersicum TPS12* gene is associated with β-caryophyllene and α-humulene biosynthesis and *S. habrochaites TPS9* is associated with germacrene B production (Hoeven van der et al., 2000; Bleeker et al., 2011; Falara et al., 2011). The existence of a *S. lycopersicum TPS9 (SlTPS9)*, that makes germacrene C, was also reporter for the cv. VFNT Cherry (Colby et al., 1998), but it is worth noting that this cultivar has introgressions of the wild species *S. peruvianum* on chromosome 6. The second loci controlling sesquiterpenes is the *Sesquiterpene synthase 2* (*SsT2*) on chromosome 8. At this locus *S. habrochaites* encodes enzymes responsible for the accumulation of class II plastid-derived sesquiterpenes, including α-santalene, α-bergamotene, β-bergamotene and 7-epizingiberene. In cultivated tomato, a cluster of five functional TPS genes (*TPS18*, *TPS19*, *TPS20*, *TPS21*, and *TPS41*) is present in the equivalent locus on chromosome 8. In addition, this same tomato chromosomal region also contains the *NERYL DIPHOSPHATE SYNTHASE 1* (*SlNDPS1*) gene, which codes for an enzyme catalyzing the formation of neryl diphosphate (NPP). The NPP is used by tomato TPS20 to synthesize β-phellandrene and several other monoterpenes in the chloroplasts (Falara et al., 2011; Matsuba et al., 2013).

The *SsT2* locus from *S. habrochaites* has a cluster of three functional TPS genes (*TPS18, TPS20* and *TPS45*) and the *CIS-FARNESYL DIPHOSPHATE SYNTHASE* (*zFPS*) gene, which is homologous to *SlNDPS1* (Matsuba et al., 2013). Both *zFPS* and *TPS45* contain putative chloroplast targeting sequences and are associated with the biosynthesis of class II sesquiterpenes in this organelle. The *zFPS* codes for a Z-prenyltransferase that catalyzes the synthesis of *Z-Z*-farnesyl diphosphate (*Z,Z*-FPP) from IPP and DMAPP. The *TPS45* gene encodes a Santalene and Bergamotene Synthase (SBS) that uses *Z,Z*-FPP as a substrate to produce class II sesquiterpenes

(Sallaud et al., 2009; Matsuba et al., 2013). Both *zFPS* and *SBS* are expressed in type-VI trichomes (Sallaud et al., 2009) which are the most common and abundant glandular trichome found in several tomato-related species (Kang et al., 2010; Glas et al., 2012; Balcke et al., 2017). In cultivated tomato, type-VI trichomes contain a single basal cell connected to a short (~0.1 mm) unicellular stalk which is attached by an intermediate cell in the four-celled glandular head containing chloroplasts and other organelles (Bergau et al., 2015). In *S. habrochaites*, the type-VI glandular trichome has a distinct morphology with a longer stalk (~0.2 mm) and a round glandular head with a larger gland cavity (Besser et al., 2009).

In general, cultivated tomatoes are highly vulnerable to several arthropod pests, which include whiteflies, spider mites, and thrips. In heavy infestation, these pests can cause a reduction of plant vigor and yield which can lead to huge losses in productivity (Wakil et al., 2018). Consequently, to minimize the damage caused by pests, it has been necessary to apply high amount of pesticides (Silva et al., 2011). In this sense, an alternative to chemical pest control would be the use of commercial tomatoes carrying favorable genetic variations from tomato-related wild species. Herein, we investigated whether the introduction of the genetic pathway for class II plastid-derived sesquiterpenes into cultivated tomato could increase resistance to arthropod tomato pests. We show that the *Sst2* gene cluster that controls santalene and bergamotene in *S. habrochaites* LA1777 can be effectively transferred to cultivated tomato (cv. Micro-Tom) and function into its type-VI trichomes. We further demonstrated that, contrary to earlier works, the high production of "wild tomato sesquiterpenes" was not sufficient to confer resistance to the tomato-pests whitefly (*Bemisia tabaci*), spider mites (*Tetranychus urticae* and *T. evansi*) nor thrips (*Frankliniella occidentalis*). Conversely, the highly insect resistance presented by *S. habrochaites* LA1777 is likely to be due to sesquiterpene carboxylic acid derivatives that were detected in the wild species but not in the introgressed line. This means that the full set of genes responsible for the conversion of the terpenes to derived carboxylic acids lie outside the metabolic cluster on chromosome 8. Surprisingly, the type-VI trichome internal storage-cavity size increased in the MT line harboring the *Sst2* gene cluster, which is probably a pleiotropic effect of the *S. habrochaites* genes leading to the high accumulation of santalene and bergamotene. Our results provide for the understanding of type-VI trichome morphology of *S. habrochaites* and pave the way for new discoveries of specific enzymes of sesquiterpene metabolism which will assist in the further breeding and pyramiding of all loci necessary to obtain insect-resistant tomatoes.

1.2 Materials and Methods

1.2.1. Plant material

Seeds from Micro-Tom (MT) were donated by Dr. Avram Levy (Weizmann Institute of Science, Israel) in 1998 and maintained through self-pollination as a true-to-type cultivar since then. The *lutescent 1* mutation was introgressed into MT from its original background as described previously in Carvalho *et al.* (2011). Seeds from *Solanum habrochaites* LA1777 were obtained by the Tomato Genetics Resource Center (TGRC - University of California).

The sesquiterpene synthase 2 pathway from *S. habrochaites* LA1777 was introgressed into MT background by allelic substitution making use of the morphological marker MT-*lutescent 1*, which maps on the same arm of chromosome 8 (https://tgrc.ucdavis.edu/) (Fig.1). Briefly, pollen from *S. habrochaites* LA1777 was collected and used to fertilize emasculated MT-*lutescent 1* flowers. The F1 obtained was used as pollen donor for MT-*lutescent 1* plants and this procedure was repeated in the successive backcrossing (BCs). In each BC, we screened for reduced plant size (MT-like phenotype) and the absence of the *lutescent 1* phenotype (Fig. S1), which is an indicative of the presence of the LA1777 genes in the *SsT2* locus. After self-pollination in BC_6F_2 generation, we screened plants for the presence of the same sesquiterpenes compounds found in the wild parental species. The resulting homozygous MT-*Sst2* genotype was considered a near-isogenic line (NIL).

Plants were grown in a greenhouse with 30/26°C temperature day/night and 60–75% ambient relative humidity, 11.5 h/13 h (winter/summer) photoperiod, sunlight 250–350 μmol photons $m^{-2} s^{-1}$ PAR irradiance. Seeds were germinated in bulk in 350 mL pots with a 1:1 mixture of commercial potting mix Basaplant® (Base Agro, Artur Nogueira, SP, Brazil) and expanded vermiculite, and was supplemented with 1 g L^{-1} 10:10:10 NPK and 4 g L^{-1} dolomite limestone ($MgCO_3$ + $CaCO_3$). Upon the appearance of the first true leaf, seedlings of each genotype were individually transplanted to 150 mL pots containing the soil mix described above, except that NPK supplementation was increased to 8 g L^{-1}.

1.2.2. Genetic and Physical Mapping of the introgressed *Sst2* genes

Genomic DNA isolation was extracted from leaflets using the method described by Fulton, Chunwongse & Tanksley (1995) with minor modifications. Molecular mapping using Cleaved Amplified Polymorphic Sequence (CAPS) markers was performed as previously described by Shavrukov (2016). Details of tomato genetic maps and chromosome 8 molecular markers can be

accessed through the Solanaceous Genomics Network (http://solgenomics.net/). Primers and restriction enzymes yielding CAPS between tomato and *S. habrochaites* LA1777 are detailed in the Table 1.

Table 1. Oligonucleotide sequence used for CAPS markers.

Locus id[1]	Forward	Reverse	Enzyme	Fragment	
				MT	LA1777
Solyc08g005020	TTTGCGTGTACCT TTTGCAG	TCCTCCTCAAAAC CCTCTTCCTC	HinfI	536/512/ 239	1048/ 239
Solyc08g005640	GTTCTCATAGTTC CCACTATGTATCC	GACAAACTACTTT GTTGTCGGAGT	HindIII	693/155/ 322	693/477
Solyc08g006410	GATCCATTCCTTC CTTGGGCTGTTG	GATGGTATTGTT GGTCCAATTGTC	PacI	605	503/102
Solyc08g007130	GTCCTTGTGGTG AACTAAGATATCC	GAATATGCCTTCT GCTGATGCTAGG	BglII	986/490	1478
Solyc08g065740	CGGTGGTCTTAA GGATGAGAAC	CACAACTTCAAAA TAGGGTCTC	BanI	81/286/ 342	367/342
Solyc08g076820	GTACTACTACTCC CTTAGAGCAAC	GTATGCACTAGG GCTCATAATTCG	BspDI	536/125	661
Solyc08g083230	GATGTGGGTTGT TTTGCAGGA	CCAGACATGGAA AGTGTTAGCGA	TaqI	618/179/ 106	797/106

Locus according to Sol Genomics Network database (http://solgenomics.net/).

1.2.3. Trichome counts and phenotyping

Trichome counts were performed on leaflets taken from mature fifth leaves (counting from cotyledons) according to the methodology described by Vendemiatti *et al.* (2017). Both leaf surfaces were dissected along the longitudinal axis in 15 × 3 mm strips covering the middle section of the leaf blade (avoiding the primary veins). The strips were fixed on microscope slides using transparent nail polish (Revlon, Brazil). Five individuals per genotype were sampled, and four different strips were analyzed per plant. Images were taken using a Leica S8AP0 (Wetzlar, Germany) magnifying glass set to 80x magnification, coupled to a Leica DFC295 camera (Wetzlar, Germany). Counting of trichomes density (mm^2) was performed using the equipment analytical program.

The morphology of type-VI trichomes was examined under an EVOSfl (www.thermofisher.com) inverted microscope. Lateral leaflets strips were submerged in water under microscope slides and images of type-VI trichomes were taken. All trichome measurements

were performed on images of 5 plants per genotype using ImageJ software version 1.4.1. Gland volume and cavity volume were calculated using the volume of the prolate ellipsoid formula: $V = 4/3 \times \pi \times a$ (vertical axis) $\times b^2$ (horizontal axis).

1.2.4. GC-MS quantification

For GC-MS volatile terpene quantification 300 individual type-VI trichome glands were collected from leaves in adult vegetative phase (fifth leaf from the cotyledons) with a glass pulled Pasteur pipette under a Leica MZFLIII microscope (www.leica-microsystems.com). The terpene extraction was conducted according to the methodology described by Xu *et al.* (2018c). The collected glands were dissolved in 150 μL of hexane plus 0.5 ng μL of benzyl acetate (Sigma-Aldrich; www.sigmaaldrich.com) as an internal standard. Na_2CO_3 (Sigma-Aldrich) was used to remove water from the hexane. Volatiles were separated using an Agilent (www.agilent.com) 7890A gas chromatograph, attached to an Agilent 7200 accurate-mass quadrupole time-of-flight mass spectrometer. Here, 2μL of the sample was injected heated to 275°C in the injector port and separated on an HP-5ms column (0.25 mm in diameter, 30 m in length, with 0.25 μm film thickness) using Helium as carrier gas (flow rate 1 mL/min). The oven temperature was maintained at 40°C for 3 min and increased by 15°C per min until it reached 250°C and maintained for 3 min. Identification of the compounds was based on the retention time of the chromatographic peaks and their corresponding mass spectra, which were compared to terpene standards and data libraries. Quantification of peak areas was performed using Masshunter Qualitative Analysis software (Agilent). Peak areas were corrected for the internal standard and quantified using the available terpene standards. Terpene concentration was calculated per trichome gland (ng/gland) using the peak areas relative to the internal (benzyl acetate) and terpene standards available.

1.2.5. Whitefly bioassay

Bemisia tabaci (former biotype B; Middle East Asia Minor I-II (MEAM)) population was maintained in a climatized chamber (Snijders Tilburg; T 28°C, 16-h light, RH 75%) on cucumber plants prior to the experiment. For no-choice assay twenty adult whiteflies were randomly taken from the population, anesthetized with CO_2 and placed in a clipcage (2.5 cm diameter; Bioquip). Two clipcages were attached in two different leaflets per plant. Five plants per genotype were used. The plants were kept inside of a closed greenhouse compartment (28 °C, RH 65%) and after 5 days, the number of whiteflies alive was recorded.

1.2.6. Spider Mite bioassay

A non-choice performance assay was set- up using two species of spider mites. The two spotted spider mite *T. urticae* Koch Viçosa-1 and the red spider mite *T. evansi* Baker & Pritchard Viçosa-1 were initially collected from infested tomato plants (Sarmento et al., 2011). Before the experiments, *T. urticae* mites were maintained on detached leaves of *S. lycopersicum* cv. Santa Clara and *T. evansi* mites were maintained on detached leaves of *S. lycopersicum* cv. Castlemart following standard procedures (Ataide et al., 2016). The rearings were maintained in a climate room at 25 °C, a 16/8 h light regime with 300 μE m^{-2} s^{-1}, and 60% RH.

For each plant genotype, 15 leaf discs of 15 mm were made from the fifth leaf (counting from the cotyledons). The leaf discs were placed on 1.5% daishin agar (Duchefa Biochemie bv, Haarlem, The Netherlands) that was poured in small cups (3 cm diameter x 2 cm height) with their adaxial side facing up. On each leaf disk, a single 2-day-old adult female of *T. urticae* or *T. evansi* was placed using a soft paintbrush. Mites were confined into each cup and ventilation was assured by a 1 cm^2 opening on the lid that was covered with mite-proof mesh (pore size of 80 μm). The cups were maintained in a climate room at 25°C, a 16/8 h light regime with 300 μE m^{-2} s^{-1}, and 60% RH. Two days after infestation, spider mite survival was recorded and the average fecundity (number of eggs laid per female) was calculated using those spider mites that were alive after the two-day period for the calculations.

1.2.7. Thrips bioassay

A non-choice performance thrips bioassay was set up using the western flower thrips *Frankliniella occidentalis* (Pergande). The thrips colony was kept in the laboratory inside cages where bean pods were provided and supplemented with pollen as previously described (Muñoz-Cárdenas et al., 2017). For each plant genotype, 15 leaf discs of 15 mm were made. Similar to the spider mites set-up, the experimental arena consisted in cups (3 cm diameter x 2 cm height) filled with 1.5% daishin agar on which one leaflet was placed with the adaxial side up. Five adult females were collected from the colony with the help of a 1 ml pipette tip attached to a vacuum and were released inside each cup through a small opening on the side of the cup, that was otherwise sealed with parafilm. The cups were maintained in a climate room at 25°C, a 16/8 h light regime with 300 μE m^{-2} s^{-1}, and 60% RH. 72 hours after the release of the females, the adult thrips were removed and their survival (number of alive thrips) was scored with the help of a dissecting stereoscope. The number of larvae that emerged from the eggs laid during the experiment was assessed 7 after the beginning of the experiment.

1.2.8. RNA isolation and Quantitative RT-PCR

Total RNA was extracted from trichomes isolated by shaking stems in liquid nitrogen with a vortex mixer. Total RNA was isolated using Trizol reagent (Invitrogen) according to the manufacturer's instructions. RNA treated with TURBO DNase (Ambion; www.thermofisher.com) were reverse-transcribed to generate first-strand cDNA using RevertAid H Minus Reverse Transcriptase (Fermentas; www.thermofisher.com). cDNA was used as a template for quantitative RT-PCR (qRT-PCR). PCR reactions were performed using HOT FIREPol EvaGreen qPCR Mix Plus (Solis Biodyne; www.sbd.ee) and analyzed in an ABI 7500 Real-Time PCR System (Applied Biosystems; www.appliedbiosystems.com). Two technical replicates were analyzed for at less three biological samples, together with template free reactions as negative controls. Transcript abundances were normalized to *Rubisco conjugating enzyme 1* (*RCE1*) expression. Detailed primers information is described in the Table 2.

Table 2. Oligonucleotide sequence used for quantitative PCR analyses.

Primer name	Sequence	Gene symbol	References
SBS_QF	GCATTACAGAATGAGTTCGAGG	LOC101250138	-
SBS_QR	CTGATGGTAAATCAAGGCAGC		
ZFPS_QF	TTTCGAGGTCTGGAGTAAGAGTG	AHF95235	-
ZFPS_QR	TAGTGCAATCACAAGGTGAAGTC		
RCE1_QF	GATTCTCTCTCATCAATCAATTCG	TC153679	(Van Schie et
RCE1_QR	GAACGTAAATGTGCCACCCATA		al., 2007)

Gene symbol according to National Center for Biotechnology Information database (https://www.ncbi.nlm.nih.gov/).

1.2.9. Experimental Design and Statistical Analysis

Statistical analyses were done using SigmaPlot 11.0 for Windows. The experiments were arranged in a completely randomized design. All data were tested for normality and equal variance by Kolmogorov-Smirnov tests. The means were further analyzed by two-tailed Student's *t*-test ($P \leq 0.05$) or Fisher's LSD test ($P \leq 0.05$) after one-way ANOVA in multiple comparisons. For data that do not assume a specified variance or normality, we performed ranking tests Wilcoxon rank sum for pairwise comparisons and Kruskal–Wallis one-way analysis for multiple groups.

1.3. Results

1.3.1. Introgression of *SsT2* gene cluster from *Solanum habrochaites* LA1777 into *S. lycopersicum* cv. Micro-Tom (MT)

In order to introduce the class II plastid-derived sesquiterpenes pathway into the Micro-Tom (MT) cultivar, we crossed *S. habrochaites* LA1777 with a MT line harbouring the *lutescent 1* mutation (MT-*l*1) and used it as a recurrent parent (Fig.1A).

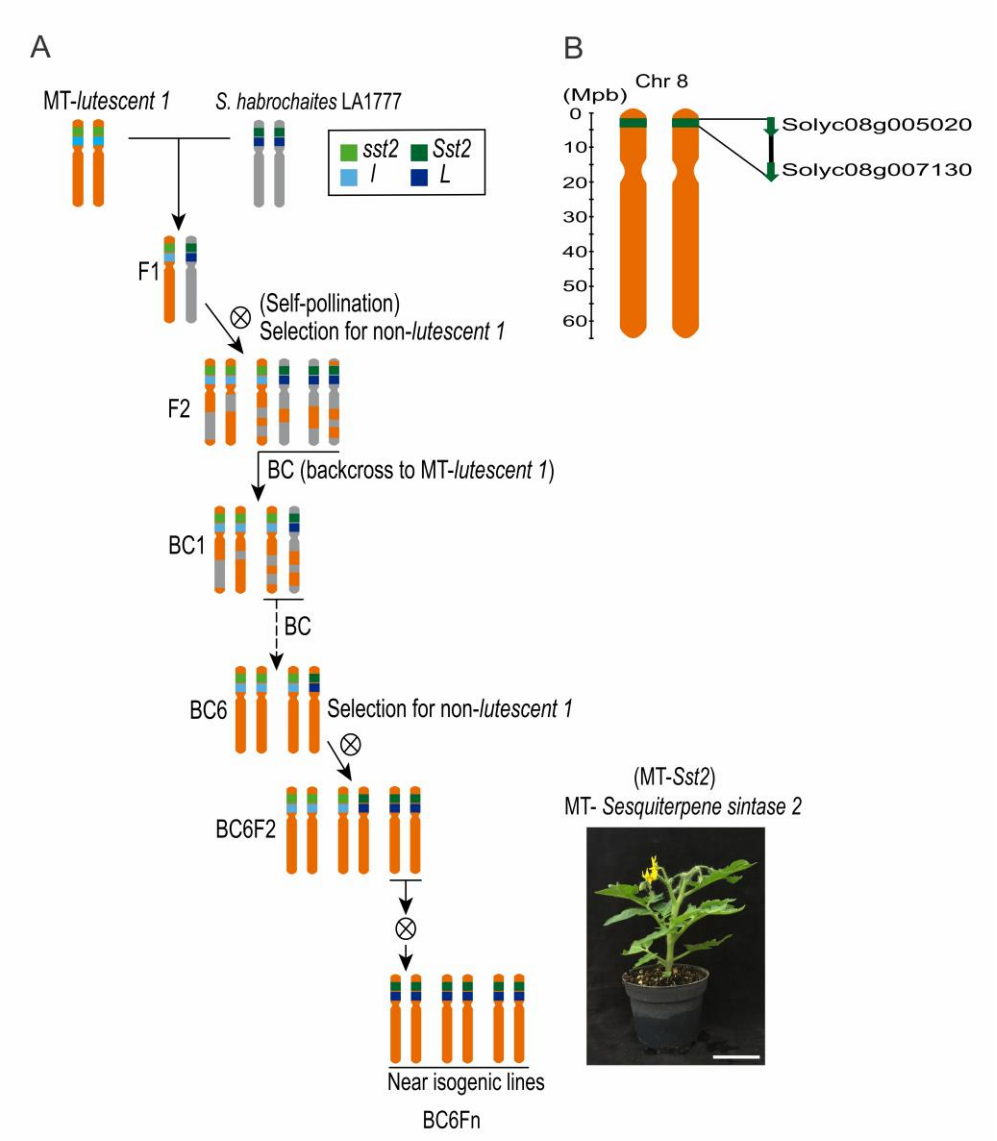

Figure. 1 Scheme of crossing and backcrossing (BC) to create a Micro-Tom (MT) near isogenic line (NIL) harboring the *Solanum habrochaites* LA1777 genes for the "*Sesquiterpene Synthase 2*" (*SsT2*) locus. **A** MT NIL bearing the *lutescent 1* mutation was used to assist the introgression process as a morphological marker (the absence of the *lutescent 1* phenotype was used as an indicative of the presence of the LA1777 genes in the *SsT2* locus). The presence of MT (*l* and *sst2*) or LA1777 (*L* and *Sst2*) variants is indicated in different colors. **B** The ID (Solyc) of the genetic markers used to determine the introgression borders are depicted.

Both the *lutescent 1* mutation and the *SsT2* locus map at the short arm of the chromosome 8 (Tanksley et al., 1992; Sallaud et al., 2009). This allowed us to introgress the *S. habrochaites SsT2* locus into MT by a relatively easy visual selection of progeny. The *lutescent 1* phenotype comprises a premature and progressive yellowing of the leaves due to impaired chlorophyll accumulation (starting from the base of the plant) (Fig S1 A and B), a lack of chlorophyll accumulation in the pistils (Fig S1 C) and whitish-yellow fruits (Barry et al., 2012). Since the *lutescent 1* mutation and the *S. habrochaites* alleles at the *Sst2* locus are in "repulsion phase", in each generation of introgression we selected for plants not presenting the *lutescent 1* phenotype.

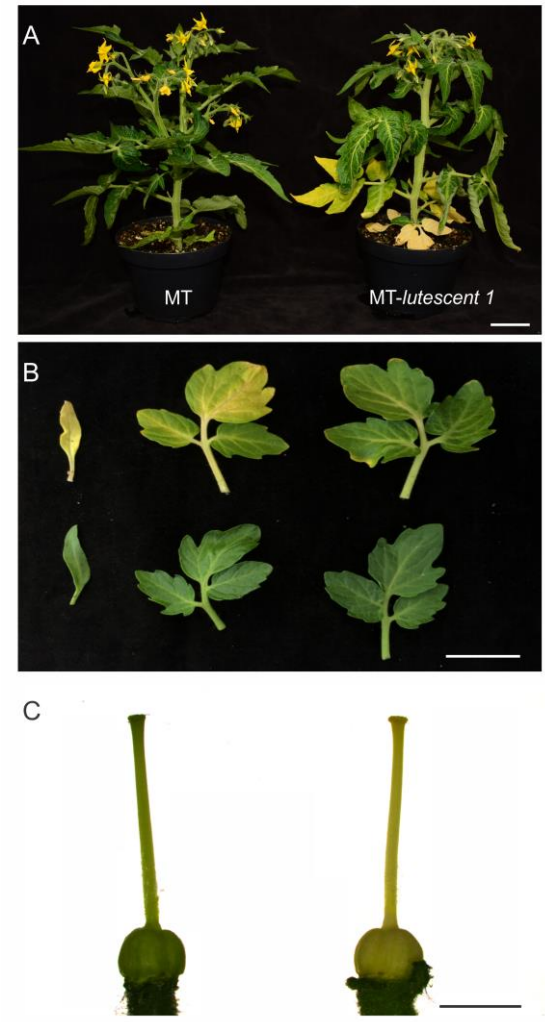

Figure. S1 A Representative MT and the MT near isogenic line (NIL) harboring the *lutescent 1* (l/l) mutation (MT-*lutescent 1*) which shows a premature and progressive chlorophyll loss that was used as a morphological marker for genetic introgression. Bar = 2 cm. **B.** MT-*lutescent 1* leaf phenotype (top) displaying premature and progressive chlorophyll loss on cotyledons and the first and second leaves, compared with MT (bottom), which do not present senescent-like leaves at this developmental stage. Bar = 3 cm. **C.** MT-*lutescent 1* pistil (right) with chlorophyll loss, compared with MT pistil (left). Bar = 2 mm. The photos were taken in 45-days old plants.

In the F_2 generation, plants not harboring the *lutescent 1* mutation (i.e. without impaired chlorophyll accumulation) were used for back-crossing (BC) with MT-*l1*. This procedure was repeated in the BC1 and subsequent generations. After six BC generations and self-pollination (BC_6F_2) using the visual marker, we employed CAPS markers designed based on single nucleotide polymorphisms (SNPs) for a genetic screen for homozygous plants harbouring the *SsT2* locus from the wild species. In the BC_6F_3 generation and generations thereafter (BC_6F_n), the obtained plants were considered a near isogenic line (NIL), no longer segregating for the presence of the wild sesquiterpenes pathway and other traits. The NIL was named MT-*Sesquiterpene synthase 2* (MT-*Sst2*).

We employed CAPS markers to determine the size of the fragment introgressed into the MT background. Genetic mapping positioned the introgressed region between Solyc08g005020 and Solyc08g007130 on the top of chromosome 8 (Fig. 1B), translating to roughly 2100 genes, based on the Heinz genome. The introgressed region coincides with the mapping position of *SsT2* previously reported by Sallaud *et al.* (2009), which also confirmed that both *zFPS* and *SBS* genes were mapped to this region. The introgressed region in MT-*Sst2* is also consistent with the *SsT2* locus found from the set of ILs between *S. lycopersicum* × *S. habrochaites* LA1777 that were able to produce class II sesquiterpenes (Hoeven van der et al., 2000).

1.3.2. Production of the sesquiterpenes santalene and bergamotene in the Micro-Tom line harboring the *S. habrochaites* genes in the *SsT2* locus

To confirm the presence of the *S. habrochaites* sesquiterpene pathway into MT-*Sst2* line, we performed gas chromatography and mass spectrometry (GC-MS) analysis. The type-VI trichomes of the MT-*Sst2* line not only accumulated class II sesquiterpenes, but also presented augmented amounts of these plastid-derived sesquiterpenes, when compared with parental *S. habrochaites* LA1777 (Fig. 2).

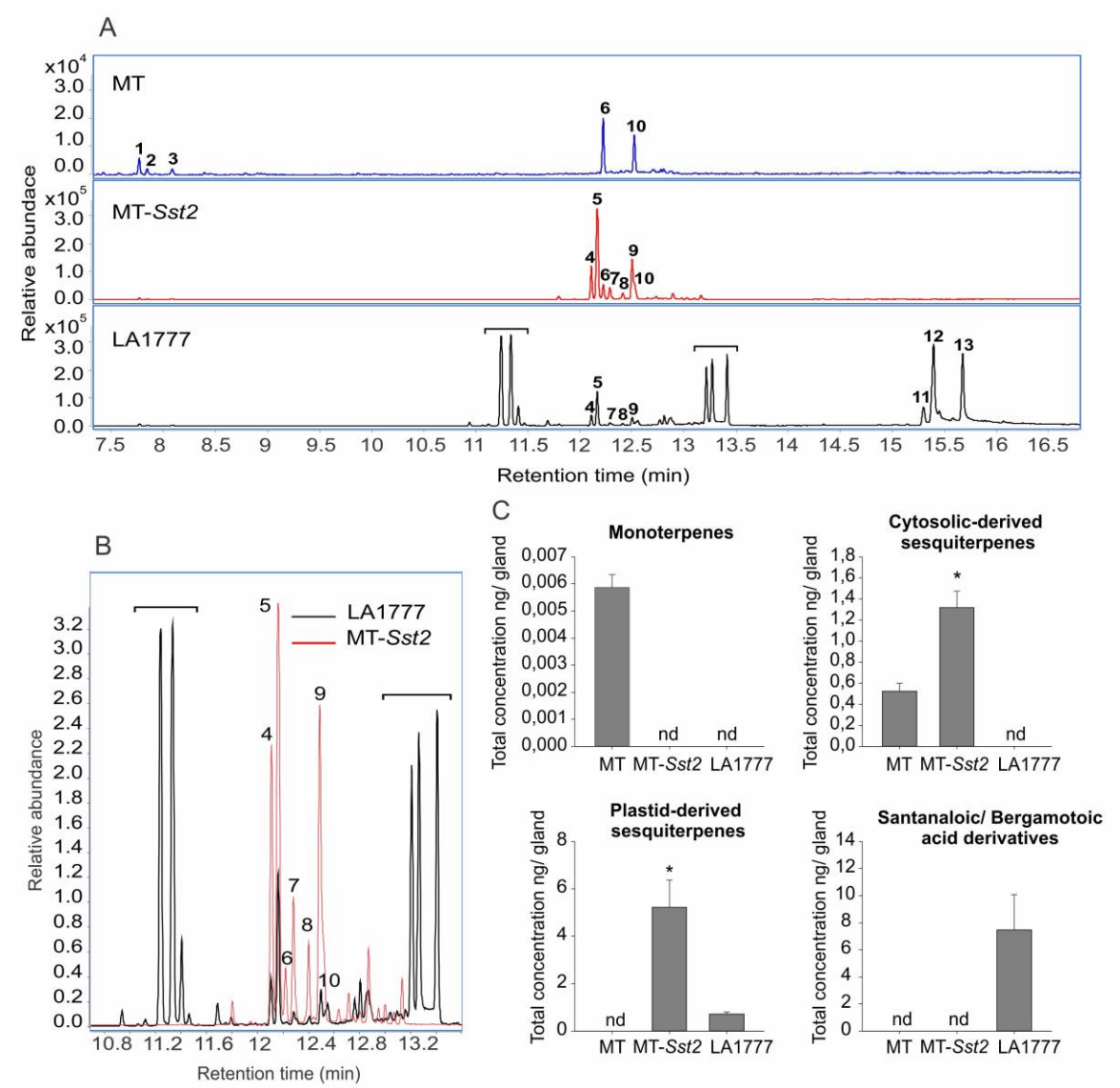

Figure 2 A. GC-MS chromatograms showing mono and sesquiterpenes found in type-VI trichomes from Micro-Tom (MT), MT-*Sst2,* and *Solanum habrochaites* LA1777. The indicated peak corresponds to the following compounds: (1) 2-carene, (2) α-phellandrene, (3) β-phellandrene/ D-limonene, (4) α-bergamotene, (5) α-santalene, (6) β-caryophyllene, (7) exo-α-bergamotene, (8) epi-β-santalene, (9) endo-β-bergamotene, (10) α-humulene, (11) α-bergamotoic acid, (12) α-santanaloic acid and (13) β-bergamotoic acid. The bracket indicates peaks related to germacrenes. The chromatogram shows the detector response for the terpene-specific ion mass 93. **B.** Gas chromatograms overlaying mono and sesquiterpenes found in type-VI trichomes from MT-*Sst2,* and *Solanum habrochaites* LA1777. The indicated peaks correspond to the following compounds: (4) α-bergamotene, (5) α-santalene, (6) β-caryophyllene, (7) exo-α-bergamotene, (8) epi-β-santalene, (9) endo-β-bergamotene, (10) α-humulene. The bracket indicates peaks related to germacrenes. The chromatogram shows the detector response for the terpene-specific ion mass 93. **C.** Total

amount of compounds present in type-VI glandular trichomes of each genotype. Total concentration of monoterpenes: (1) 2-carene, (2) α-phellandrene and (3) β-phellandrene/ D-limonene; Total concentration of cytosolic sesquiterpenes: (6) β-caryophyllene and (10) α-humulene; Total concentration of plastid-derived sesquiterpenes: (4) α-bergamotene, (5) α-santalene, (7) exo-α-bergamotene, (8) epi-β-santalene and (9) endo-β-bergamotene; Total concentration of santalanoic/ bergamotoic acid derivative: (11) α-bergamotoic acid, (12) α-santanaloic acid (13) β-bergamotoic. Each data point represents the mean + SE of five biological replicates. For each sample, 300 type-VI glandular trichomes were collected with a glass capillary before GC-MS. Bars indicated with an asterisk were significantly different according to t-test ($P \leq 0.05$). nd, Not detected.

We did not detect the monoterpenes found in MT background in the MT-*Sst2* line (Fig. 2, (peaks 1, 2 and 3; Fig S2)), which indicates that the introgression performed functioned was expected. Accordingly, since 2-carene, α-phellandrene and β-phellandrene/ D-limonene are encoded by *SlTPS20* genes on tomato chromosome 8, it was expected that these monoterpenes on homozygous MT-*Sst2* plants would be replaced by the sesquiterpenes encoded by the wild species alleles on the same chromosomal region. As a result, the MT-*Sst2* line, which has the *S. habrochaites SsT2* locus, presented increased levels of the plastid-derived (α-bergamotene, α-santalene, exo-α-bergamotene, epi-β-santalene, and endo-β-bergamotene) sesquiterpenes (Fig. 2C).

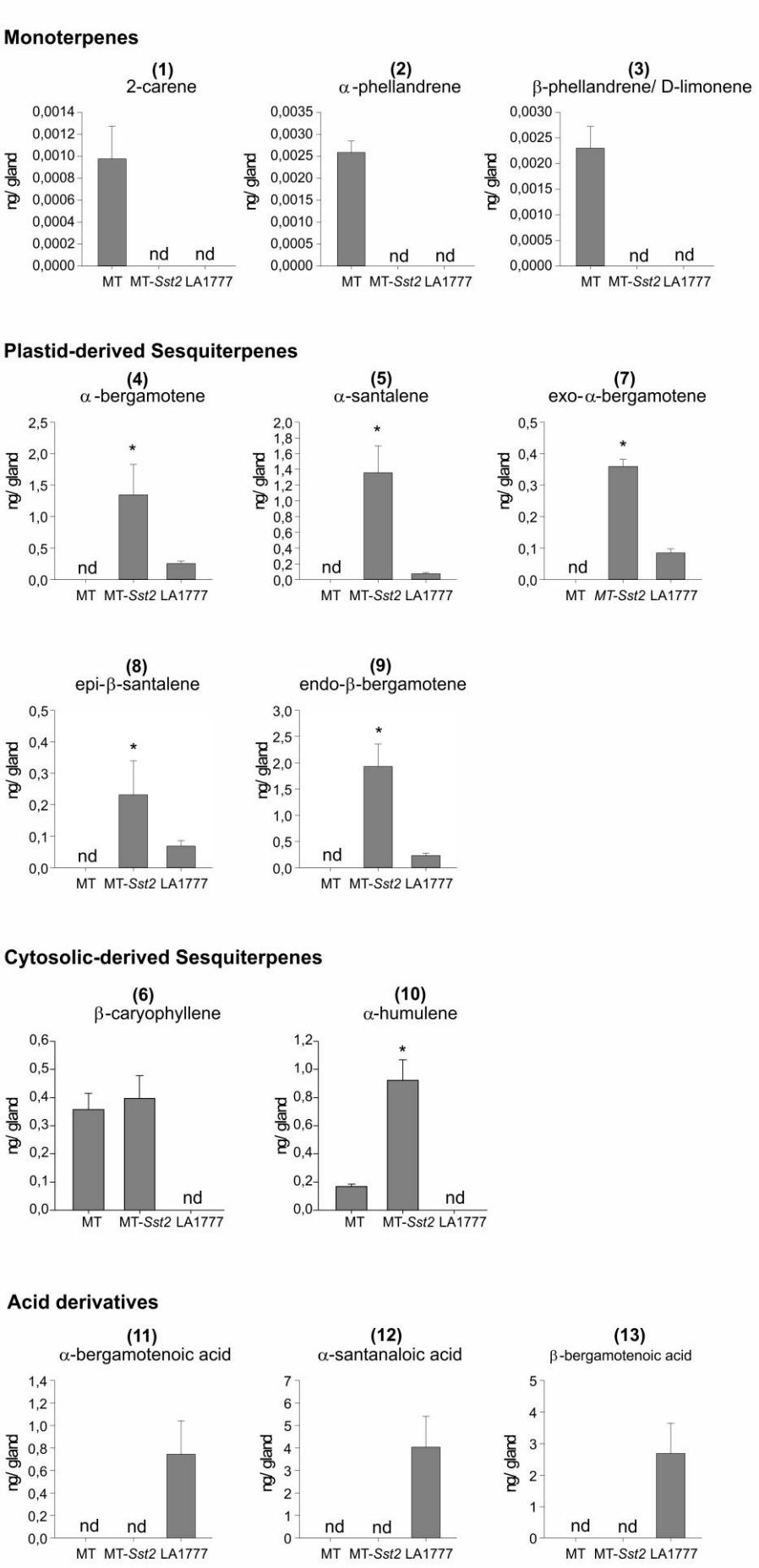
Monoterpenes
(1)
2-carene
(2)
α-phellandrene
(3)
β-phellandrene/ D-limonene
ng/ gland
nd
MT MT-Sst2 LA1777
Plastid-derived Sesquiterpenes
(4)
α-bergamotene
(5)
α-santalene
(7)
exo-α-bergamotene
(8)
epi-β-santalene
(9)
endo-β-bergamotene
Cytosolic-derived Sesquiterpenes
(6)
β-caryophyllene
(10)
α-humulene
Acid derivatives
(11)
α-bergamotenoic acid
(12)
α-santanaloic acid
(13)
β-bergamotenoic acid

Figure. S2 Volatile terpene levels in type-VI glandular trichomes from Micro-Tom (MT), MT-*Sst2* and the wild species *Solanum habrochaites* LA1777. The data show the amount of each compounds present in type-VI glandular trichomes. Each data point represents the mean + SE of five biological replicates. For each sample, 300 type-VI glandular trichomes were collected with a glass capillary before GC-MS. Bars indicated with an asterisk were significantly different according to t-test ($P \leq 0.05$). nd, Not detected.

As expected, no santalene or bergamotene are present in control MT trichomes (Fig. 2A, C and S2). However, not only was there a 7-fold increase in the plastid-derived sesquiterpenes of individual type-VI trichomes of MT-*Sst2* compared to LA1777, but unexpectedly, it appears that the MT-*Sst2* type-VI trichomes also produce substantially (2-fold) more cytosolic-sesquiterpenes β-caryophyllene and α-humulene when compared to MT (Fig. 2C).

Although we found substantial concentrations of plastid-derived sesquiterpenes in the MT-*Sst2* line, we did not detect any sesquiterpene carboxylic acid converted from santalene or bergamotene, as they did occur in the wild species (Fig. 2A, peaks 11, 12 and 13, and 2C). The non-formation of *α*-santalenoic and *α*- and *β*-bergamotenoic acid derivatives in the MT-*Sst2* line likely explain its increased amounts of santalene and bergamotene, when compared to the wild species.

1.3.3. Production of mono and sesquiterpenes in type-VI trichomes under different allelic dosages at the *SsT2* locus

In order to investigate in how far the allelic dosage at the *SsT2* locus could affect the abundance of mono and sesquiterpenes found in type-VI trichomes, we selected homozygous (*sst2*/*sst2*), (*Sst2*/*Sst2*) and heterozygous (*sst2*/*Sst2*) MT plants at this locus. The molecular markers used for selection were designed based on polymorphisms found in the genomic sequence from cultivated tomato and *S. habrochaites* LA1777 (Table 1). Note that there is no complete correspondence of the genes (synteny) present in the *SsT2* locus between *S. lycopersicum*, and *S. habrochaites*. Hence, at this locus, *S. lycopersicum* contains the functional genes *TPS18*, *TPS19*, *TPS20 (PHS)*, *TPS21*, *TPS41* and *SlNDPS1*, whereas *S. habrochaites* contains the functional genes *TPS18*, *TPS20*, *TPS45* (SBS) and zFPS (Matsuba et al., 2013). Thus, the *sst2*/*Sst2* plants can be better considered hemizygous for both set of genes from each parental.

The amount of the monoterpenes 2-carene, α-phellandrene and β-phellandrene/ D-limonene did not differ significantly between homozygous plants harboring MT genes (*sst2*/*sst2*) and hemizygous (*sst2*/*Sst2*) plants, while, as shown above, they were absent in plants homozygous for *S. habrochaites* genes (*Sst2*/*Sst2*) (Fig.3).

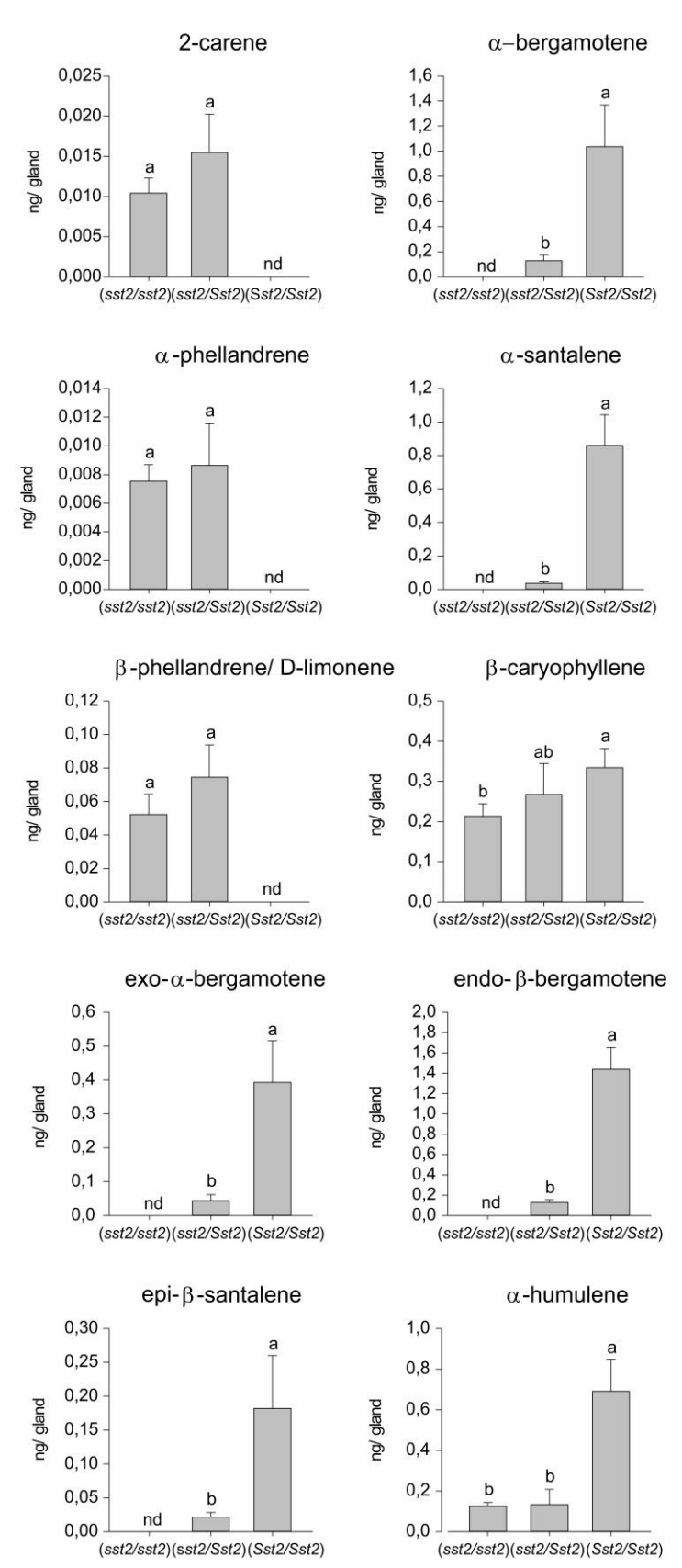

Figure 3 GC-MS chromatograms of type-VI glandular trichomes from homozygous Micro-Tom plants homozygous (*sst2/sst2*, *Sst2/Sst2*) and hemizygous (*sst2/Sst2)* at the *SsT2* locus. **A.** The data

show the amount of each compound present in type-VI glandular trichomes. Each data point represents the mean + SE of five biological replicates. For each sample, 300 type-VI glandular trichomes were collected with a glass capillary before GC-MS. Bars indicated with different letters were significantly different according to Fisher's LSD test (P≤ 0.05) after ANOVA. nd, Not detected.

The full displacement of precursor gene *SlNDPS1* and *SlTPS20* in *Sst2/Sst2* plants were expected to prevent the formation of monoterpenes in type-VI trichomes. On the other hand, there was no significant effect of gene dosage for *SlNDPS1-PHS* when comparing homozygous *sst2/sst2* and hemizygous *sst2/Sst2* plants. Regarding the concentration of cytosolic-derived sesquiterpenes β-caryophyllene and α-humulene, the hemizygous *sst2/Sst2* plants did not differ significantly from homozygous *sst2/sst2* plants, while the homozygous *Sst2/Sst2* did, presenting enhanced amounts of these compounds, especially humulene (Fig.3). Since the locus responsible for the production of theses cytosol-derived sesquiterpenes, the *SsT1* locus on chromosome 6, is the same for the three genotypes, it is unlikely to be the cause of the high amount of these compounds in the homozygous *Sst2/Sst2* plants. On the other hand, the absence of monoterpenes in the homozygous *Sst2/Sst2* plants might be related to low availability of precursor in their type-VI trichomes.

The concentrations of all plastid-derived sesquiterpenes α-bergamotene, α-santalene, exo-α-bergamotene, epi-β-santalene, and endo-β-bergamotene were higher in the homozygous (*Sst2/Sst2*) type-VI trichomes when compared to the trichomes of hemizygous (*sst2/Sst2*) plants. This can be explained by the higher gene dosage of both *zFPS* and *SBS* in homozygous *Sst2/Sst2* plants. It can also be caused by differential enzyme kinetics between the two enzymes. However, we cannot exclude the possibility that the presence of both SlNDPS1 and zFPS in the hemizygous *sst2/Sst2* plants can also influence negatively the amount of class II sesquiterpenes. Since both are acting in the same compartments, SlNDPS1 and zFPS enzymes might compete for a limited amount of IPP and DMAPP in type-VI trichomes (Dudareva et al., 2005; Besser et al., 2009; Schilmiller et al., 2009).

1.3.4. Trichome abundance, morphology and gene expression in Micro-Tom line harboring the *S. habrochaites* genes in the *SsT2* locus

We next verified if the locus substitution caused changes in abundance of different trichome types and the morphology of the class II sesquiterpene producing type-VI trichomes present in adult leaves of the MT-*Sst2* line. Figures 4A and 4B show that the densities of type-VI

trichomes were not altered by the insertion in MT-*Sst2* compared to MT, for both adaxial or abaxial leaf surfaces. Contrarily, the wild species showed significantly higher numbers of type-VI trichome on the adaxial and fewer numbers on the abaxial leaf surface (Fig. 4B).

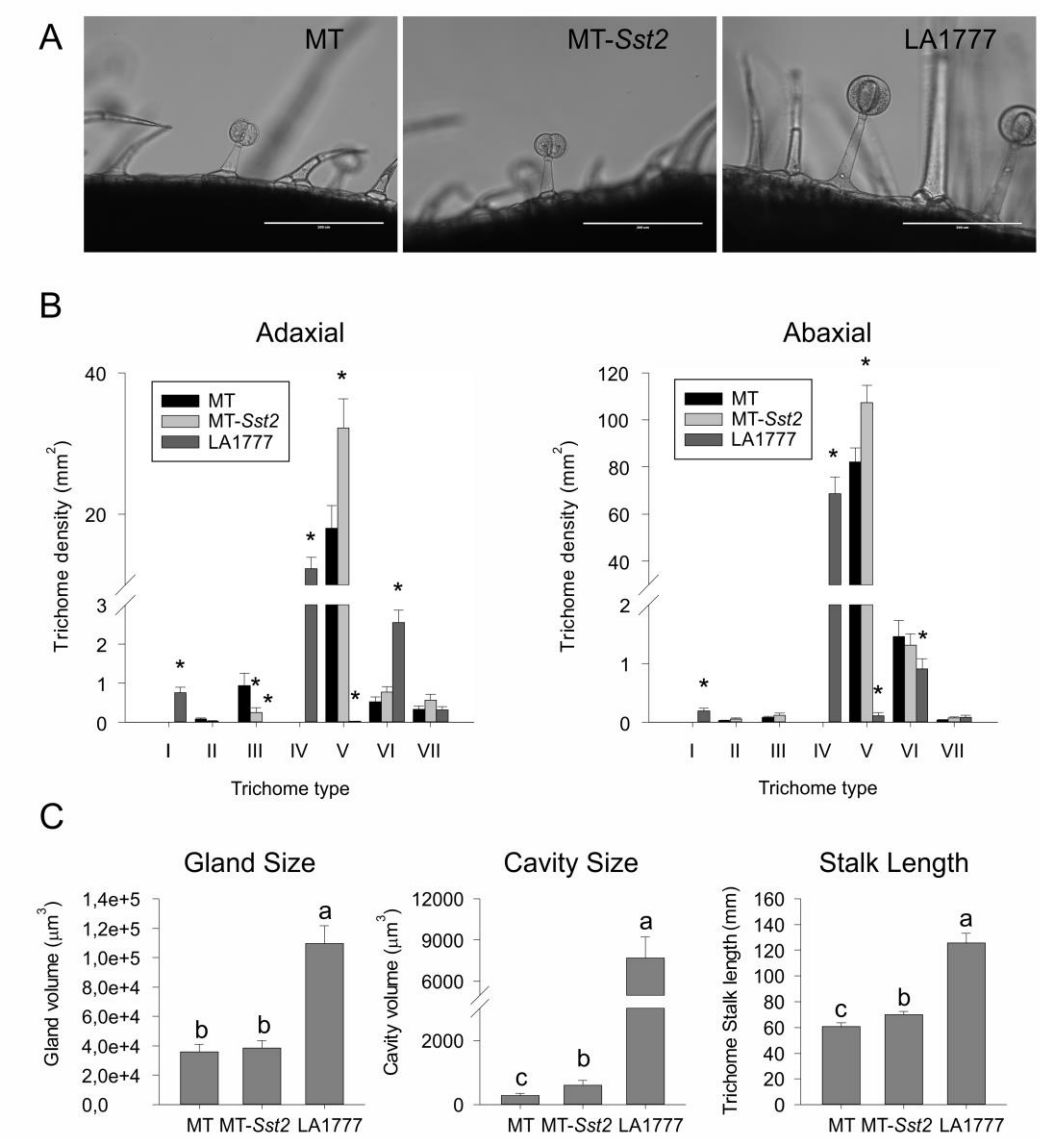

Figure 4. A. Bright field microscopy of trichomes on the leaf surface of representative 45-days old plants of Micro-Tom (MT), MT-*Sst2*, and *Solanum habrochaites* LA1777. Scale bar = 200 µm. **B.** Density (mm^2) of trichome types on adaxial and abaxial leaf surfaces. Data are mean (n=40) for each surface. Asterisks indicate mean significantly different from the control MT, according to Student t-test ($P \leq 0.05$). **C.** Trichome gland size, cavity volume and stalk length of type-VI trichomes. Data are mean (±SE) of 20 trichomes of 2 replicate leaves of five plants. Bars indicated with different letters were significantly different according to Fisher's LSD test ($P \leq 0.05$) after ANOVA.

We also observed that *S. habrochaites* LA1777 displayed an expressive number of type-IV trichomes that was not found on leaves of MT or MT-*Sst2* (Fig. 4B). Type-IV trichomes were previously associated with increased production of acylsugars providing resistance to insect pests in wild species (Simmons and Gurr, 2005). The lack of trichome type-IV in adult leaves of MT-*Sst2* indicates that the introgressed segment on chromosome 8 from *S. habrochaites* LA1777 is not involved in controlling the presence of this type of trichomes. Except for the wild species, which has more glandular type-IV trichomes than the other types, the majority of glandular trichomes found in MT and MT-*Sst2* was type-VI trichomes. For non-glandular trichomes, type-V is the most abundant type found in MT and MT-*Sst2*. Interestingly, we found an increased number of (non-glandular) type-V trichomes on both leaf surfaces in the introgressed line. Approximately 2-fold more type-V trichomes were observed on the adaxial leaf surface of MT-*Sst2* when compared with MT background (Fig. 4B).

We next analysed the gland and internal cavity size of the type-VI glandular trichomes (Fig. 4A and C). There is no difference between MT and MT-*Sst2* gland size. However, the internal cavity size is increased in MT-*Sst2* compared to MT. The introgression line also appears to exhibit a (subtle but significant) increase in stalk length compared to MT, but still much smaller than that of type-VI trichomes found in LA1777 (Fig. 4 A and C). Recently (Xu et al., 2018c), demonstrated that the downregulation of the *SlMYC1*, a bHLH transcript factor, produces plants with reduced type-VI trichome stalks. Since *SlMYC1* (Solyc08g005050) is also located on top of chromosome 8, we looked specifically for this transcript factor into our introgressed line. Hence, the presence of the wild *ShMYC1* allele inside the region introgressed was confirmed by CAPS marks (Fig 1B and S3).

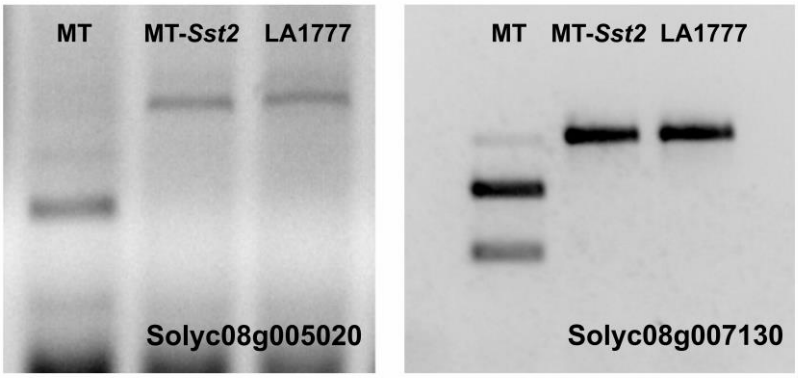

Figure. S3 Electrophoresis gels showing the positive genetic markers used to differentiate MT-*Sst2* from MT.

The relative transcript levels of *zFPS* and *SBS* were analysed by quantitative RT-PCR in trichomes of MT-*Sst2* and *S. habrochaites* LA1777. Transcripts levels of both genes were significantly lower in MT-*Sst2* when compared to the wild parent (Fig. 5).

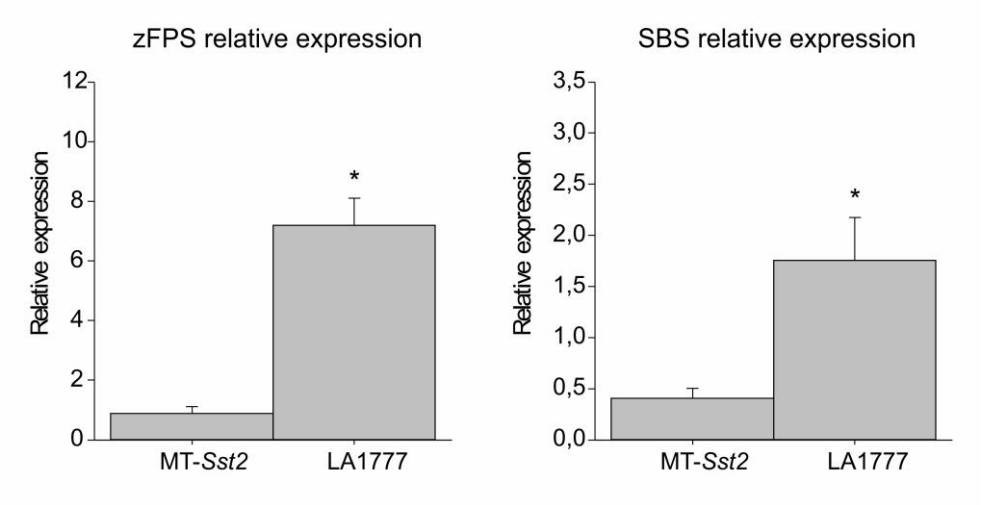

Figure 5. Transcript accumulation of the *SsT2* locus-derived genes *cis-farnesyl diphosphate synthase* (*zFPS*) and *santalene and bergamotene synthase* (*SBS*) in trichomes from MT-*Sst2* and *Solanum habrochaites* LA1777. Mean values of 4 biological replicates are shown. Transcript levels were normalized to *Rubisco conjugating enzyme 1* (RCE1) expression. Bars indicated with different letters were significantly different according to Fisher's LSD test ($P \leq 0.05$) after ANOVA. gels showing the positive genetic markers used to differentiate MT-*Sst2* from MT.

Since MT-*Sst2* and LA1777 share the same chromosomal segment comprising the *SsT2* locus, it was expected that they have the same cis-regulatory elements controlling the expression of the *zFPS* and *SBS* genes. Therefore, the lower expression of these genes in MT-*Sst2* may suggest the involvement of a set of yet unknown trans-regulatory elements (e.g. transcription factors and other regulatory elements in different chromosomal regions) necessary to increase the expression of the genes present in the *SsT2* locus.

1.3.5. Insect resistance of the Micro-Tom line with increased amounts of santalene and bergamotene

In order to verify if the increase in class II sesquiterpenes found into MT-*Sst2's* type-VI trichomes would result in improved resistance to herbivores pests, we conducted a number of no-

choice bioassays using four important herbivore pests in tomato. In the survival whitefly bioassay, MT-*Sst2* plants did not differ from MT, while the LA1777 displayed a high reduction in the survival of this pest. Almost 80% of the whiteflies survived in MT-*Sst2*, whereas less than 40% survived on the wild species (Fig. 6A).

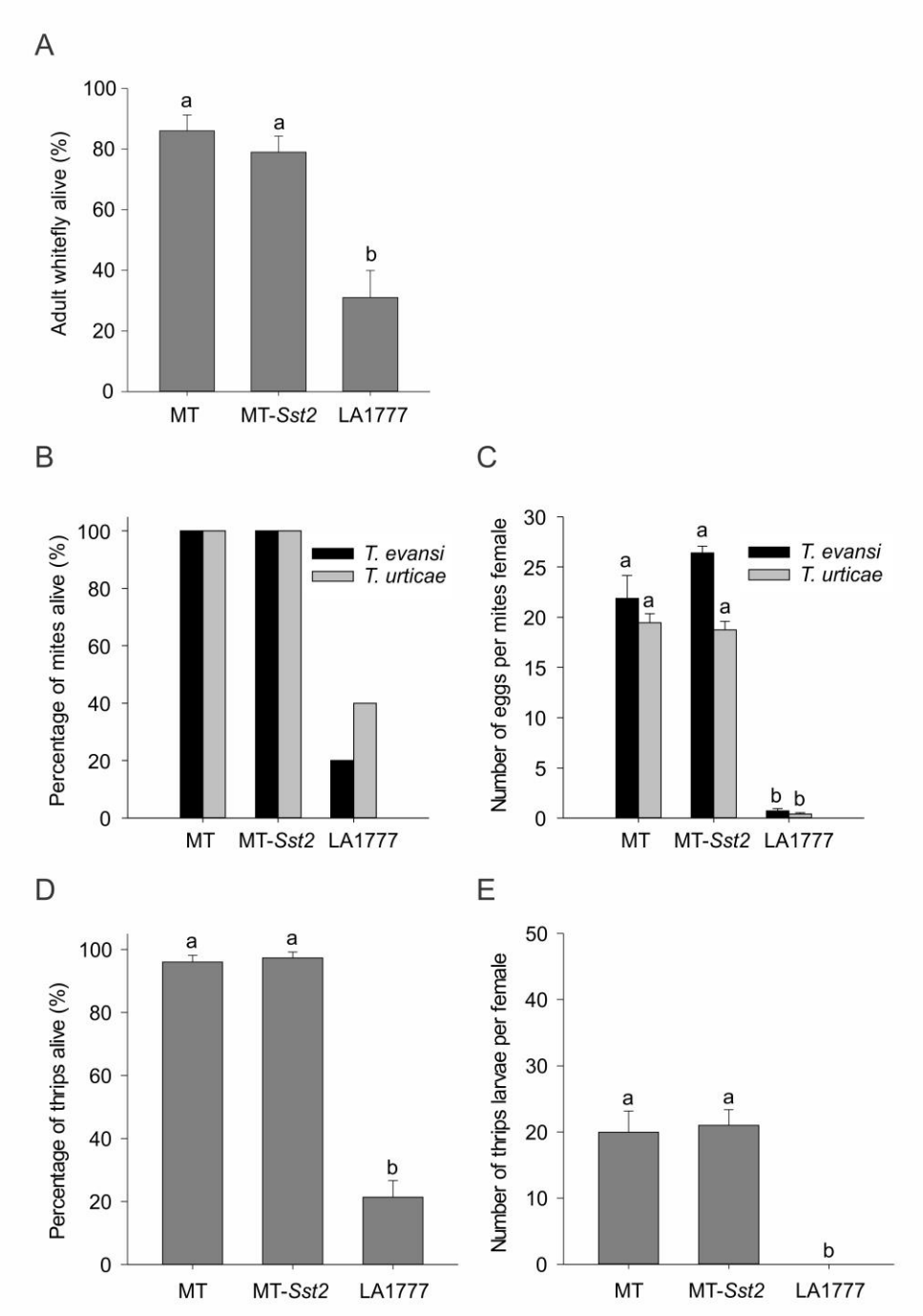

Figure 6. Herbivory tests performed on Micro-Tom (MT), MT-*Sst2* and *Solanum habrochaites* LA1777 genotypes. **A.** Percentage of adult whitefly *Bemisia tabaci* alive after five days of feeding on leaves. Data are means (±SE) of five plants, each with two cages. Bars indicated with different letters were significantly different according to Fisher's LSD test (P≤ 0.05) after ANOVA. **B** and

C. Female spider mite (*Tetranychus evansi* and *Tetranychus urticae)* survival and number of eggs after two days of feeding on plants. Bars indicated with different letters were significantly different according to Fisher's LSD test (P≤ 0.05) after ANOVA. **D** and **E.** Percentage of adult thrips alive and number of thrips larvae per female that emerged after two days on leaf discs. Bars indicated with different letters were significantly different according to Fisher's LSD test (P≤ 0.05) after ANOVA.

Next, a bioassay using the defense-suppressing spider mite *Tetranychus evansi* and the defense-inducing spider mite *Tetranychus urticae* was performed. Both MT and MT-*Sst2* showed 100% of both spider mites species surviving, while on *S. habrochaites* LA1777 only approximately 20% of *Tetranychus evansi* and 40% of *Tetranychus urticae* survived after 2 days (Fig. 6B). Oviposition rates of both spider mite species were unaltered between MT-*Sst2* and MT, whereas the number of eggs produced was strongly reduced in the wild species (Fig. 6C).

Finally, we conducted a bioassay with the western flower thrips comparing adult survival on MT, MT-*Sst2* and the wild species. There were no significant differences in survival or egg hatching observed as a result of the introgression (Fig. 6D and E). Female thrips were still able to lay eggs in both genotypes and the larvae hatched from the eggs reaching the larval stage. However again, significant reductions in the number of surviving adults and the number of emerged larvae was observed for the wild species.

1.4. Discussion

1.4.1. *Solanum habrochaites* plastid-derived sesquiterpene synthesis can be transferred to type-VI trichomes of cultivated tomato

We successfully introgressed the *SsT2* gene cluster responsible for the biosynthesis of class II plastid-derived sesquiterpene pathway from *S. habrochaites* LA1777 into the genetic model system cv. Micro-Tom (MT) (Fig 1A). The genes transferred correctly expressed into the tomato type-VI trichomes (Fig. 5) and produced high amounts of plastid-derived sesquiterpenes (Fig. 2C). We chose LA1777 as a parental line since this species produces a mixture of sesquiterpenes previously associated with insect resistance, which is not present in cultivated tomato (Frelichowski and Juvik, 2001). In cultivated tomato, type-VI trichomes accumulate mainly monoterpenes and low levels of cytosol-derived sesquiterpenes (Besser et al., 2009).

Despite the fact that the introgression line MT-*Sst2* contains significantly higher concentrations of santalene and bergamotene in isolated type-VI glands compared to LA1777 (Fig 2), we did not find back any sesquiterpene derivative in MT-*Sst2* as present in the wild species

(Fig.2A and C). In *S. habrochaites* LA1777, santalene and bergamotene are converted to derivatives identified as sesquiterpene carboxylic acids, *a*-santalenoic, and *a*- and *β*-bergamotenoic acids (Coates et al., 1988), that were found acutely toxic to caterpillars (Frelichowski and Juvik, 2001). Notably, high sesquiterpene abundance in MT-*Sst2* compared to LA1777 evidences that santalene and bergamotene are used as precursors for further metabolism into corresponding alcohol and acids derivatives (Frelichowski and Juvik, 2005; Besser et al., 2009; Gonzales-Vigil et al., 2012).

The relative transcript levels of *zFPS* and *SBS* were higher in LA1777 compared to the introgressed line (Fig. 5). It has been shown that poor terpenoids producer genotypes have also drastically reduced transcript levels for key steps in the terpenoid pathway (Tissier, 2012). Since it is unlikely that this observation is due to differences in the cis-elements controlling the *zFPS* and *SBS* genes, our results point at a role for additional genetic components such as transcription factors or other trans regulatory-elements that have not been introgressed. However, we can also not rule out feedback inhibition caused by accumulation of the terpene products. Up to now, only a few transcription factors are known to be involved in the regulation of terpenoid pathways in tomato (Spyropoulou et al., 2014; Xu et al., 2018c), but these were not implicated as positive regulators of *zFPS* or *SBS* specifically.

1.4.2. Introgression of Sst2 appears to affect type-VI trichome morphology

The MT-*Sst2* line with augmented contents of class II sesquiterpenes in type-VI trichomes displayed an increased glandular cavity volume (Fig 4C). A possible explanation for this is that the boost in the total amount of terpenes could result in a physical pressure in the cavity wall, forcing the internal cavity to expand like a balloon (Ben-Israel et al., 2009). However, the subtle increase in the internal cavity of MT-*Sst2* was not paired with an altered external gland shape. Modification in external gland shape depends both on genes related to cell wall remodelling (Bennewitz et al., 2018) and on synthesis and accumulation of very high levels of compounds into the gland (Ben-Israel et al., 2009). Thus, a combination of genes controlling the high flux of metabolites with genes controlling the cell wall remodeling might push the gland to expand, creating the characteristic round type-VI trichome of *S. habrochaites* (Ben-Israel et al., 2009). Further studies combining the MT-*Sst2* line with other introgressed line related to cell wall remodeling could shed light on the *S. habrochaites* type-VI trichome morphology and functionality.

Although LA1777 harbors type-IV glandular trichomes, which are the major site for acyl sugar biosynthesis (Schilmiller et al., 2015), this trichome type was not found on adult leaves of MT-*Sst2*, just like MT (Fig. 4B). The lack of type-IV trichomes in cultivated tomato was recently linked to the progression from the juvenility to the adult phase (Vendemiatti et al., 2017). The absence of this trichome type in MT-*Sst2* likely means that the region introgressed is not involved

in heterochrony and type-IV trichome development. The non-glandular trichome type-V on the other hand, was the most abundant trichome in both MT and MT-*Sst2* contrary to the wild parent (Fig. 4B), which might be attributed to the negative correlation between the presence of trichomes types-IV and V, previously described by Vendemiatti *et al.* (2017).

The increased type-VI trichome stalk length observed in MT-*Sst2* (Fig. 4C) is unlikely the result of the effect of the enzymes encoded on the *SsT2* locus introgressed. However, we found that the introgressed region contained the *Solanum habrochaites* allele of transcription factor *MYC1 (*Solyc08g005050*)* (Fig. 1B and S3). The *ShMYC1* variant of the basic helix-loop-helix (bHLH) transcription factor, previously linked to trichome stalk length (Xu et al., 2018c) replaced cultivated tomato *SlMYC*. Stalk length is one of the morphologic characteristics used to identify the different types of trichomes in *Solanum* (Simmons and Gurr, 2005). In general, *S. habrochaites* species exhibit a longer type-VI trichome stalk compared to cultivated tomatoes (Simmons and Gurr, 2005; Bergau et al., 2015). The biological role for a higher stalk, or a taller trichome in the wild species is not described.

1.4.3. Enzymes involved in anti-insect sesquiterpene carboxylic acids (SCA) terpenoids are not localized to a single metabolic locus

Even though MT-*Sst2* produced relatively high levels of santalene and bergamotene, compared to *S. habrochaites* LA1777, this did not confer resistance to any of the herbivores tested here (Fig. 6). Herbivore resistance observed in the wild species is therefore likely due to the sesquiterpene carboxylic acids (SCA) derivatives absent in MT-*Sst2*. It has indeed been shown before that the presence of these derivatives in LA1777 were responsible for larval feeding behavior and survival of two tomato insect pests (Frelichowski and Juvik, 2005). However, transgenic *S. lycopersicum* producing the *S. habrochaites* sesquiterpene 7-epizingiberene, product of the expression of *zFPS* and *ShZIS*, probably an allelic variant of *ShSBS*, did display a clear toxicity phenotype against spider mites (Bleeker et al., 2012).

Up to now, little is known about the enzymes catalyzing the formation of SCAs from sesquiterpenes in LA1777. It can be hypothesized that cytochrome P450 enzymes (which often hydroxylate terpenes) play a role. In *Santalum album* (Santalaceae) santalenes (α-, β- and *epi*-β-santalene) and α-*exo*-bergamotene are further metabolized into sesquiterpene alcohols α-, β-, and *epi*-β-santalol and α-*exo*-bergamotol by a CYP76F cytochrome P450 (Diaz-Chavez et al., 2013). Nevertheless, co-expression of *S. album* santalene/bergamotene oxidase *SaCYP76F39v1* in transgenic tobacco plants did not hydroxylate santalene or bergamotene (Yin and Wong, 2019). In *Tanacetum cinerariifolium (*Asteraceae) two oxidation reactions convert trans-chrysanthemol into trans-chrysanthemic acid (Xu et al., 2018b). Using tomato transgenic lines, Xu *et al.* (2018a) also

showed that expressing chrysanthemyl diphosphate synthase from *Tanacetum cinerariifolium* together with two genes from the tomato accession *S. habrochaites* LA1777 was sufficient for the transgenic lines to produce trans-chrysanthemic acid.

Although the introgressed region in line MT-*Sst2* contains both alcohol dehydrogenase (ADH) and P450s from *S. habrochaites* LA1777 (Appendix 1), it seems that the genes required for the conversion of the terpenes to derived alcohols or carboxylic acids lie outside the metabolic cluster on chromosome 8. Peripheral pathway genes located outside the core metabolic cluster in different chromosomes has been described before for tomato and other species (Nützmann et al., 2016). The fact that we could not find any sesquiterpene derivative in MT-*Sst2* at least indicates that the alcohol dehydrogenases and cytochrome P450s introgressed must have different functions.

The results presented here suggest that in order to create insect resistant cultivated tomato, additional genomic segments need to be introgressed, containing the genes that convert sesquiterpenes into acid derivatives. Given the specificity of the metabolic pathway, present in the gland of type-VI trichomes, a promising approach to identify additional biosynthesis genes, would be the use of MT-*Sst2* in crosses with *S. habrochaites* for a second round of introgression coupled with a GC-MS selection for the presence of *α*-santalenoic and *α*- and *β*-bergamotenoic acid in each generation of backcross. If successful, this would not only contribute to fundamental science (*e.g.* the discovery of new genetic components in sesquiterpene metabolism) but also applied molecular breeding in the pursuit of insect-resistant cultivated tomatoes.

1.5. Conclusion

Altogether, this study demonstrates that the plastid-derived sesquiterpene synthesis from *S. habrochaites* can be transferred to type-IV trichomes of cultivated tomato. However, additional genetic components from the wild species should be transferred to acquire herbivore resistance in cultivated tomato. The results presented here indicate that at least part of these genetic components is likely to encode: i) enzymes that convert santalene and bergamotene into their carboxylic acids, ii) transcription factors modulating the expression *zFPS* and *SBS* and iii) enzymes and possible regulatory genes for cell wall remodeling and enlargement of the internal cavity of the trichome gland. The introgressed line presented here, in the model system Micro-Tom, can provide for rapid introgression and transgenic manipulation of the additional genetic components involved in sesquiterpene metabolism and type-VI trichome development. Besides the obvious practical application of such knowledge for molecular breeding of insect-resistant commercial tomatoes, it will also improve our understanding of glandular trichome biology and the evolution of the tomato and its related wild species.

REFERENCES

Ataide LMS, Pappas ML, Schimmel BCJ, Lopez-Orenes A, Alba JM, Duarte MVA, Pallini A, Schuurink RC, Kant MR (2016) Induced plant-defenses suppress herbivore reproduction but also constrain predation of their offspring. Plant Sci 252: 300–310

Balcke GU, Bennewitz S, Bergau N, Athmer B, Henning A, Majovsky P, Jiménez-Gómez JM, Hoehenwarter W, Tissier A (2017) Multi-Omics of tomato glandular trichomes reveals distinct features of central carbon metabolism supporting high productivity of specialized metabolites. Plant Cell 29: 960–983

Barry CS, Aldridge GM, Herzog G, Ma Q, McQuinn RP, Hirschberg J, Giovannoni JJ (2012) Altered Chloroplast Development and Delayed Fruit Ripening Caused by Mutations in a Zinc Metalloprotease at the lutescent2 Locus of Tomato. Plant Physiol 159: 1086–1098

Ben-Israel I, Yu G, Austin MB, Bhuiyan N, Auldridge M, Nguyen T, Schauvinhold I, Noel JP, Pichersky E, Fridman E (2009) Multiple biochemical and morphological factors underlie the production of methylketones in tomato trichomes. Plant Physiol 151: 1952–1964

Bennewitz S, Bergau N, Tissier A (2018) QTL mapping of the shape of type VI glandular trichomes in tomato. Front Plant Sci 9: 1421

Bergau N, Bennewitz S, Syrowatka F, Hause G, Tissier A (2015) The development of type VI glandular trichomes in the cultivated tomato Solanum lycopersicum and a related wild species S. habrochaites. BMC Plant Biol 15: 289

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

Bleeker PM, Mirabella R, Diergaarde PJ, VanDoorn A, Tissier A, Kant MR, Prins M, de Vos M, Haring MA, Schuurink RC (2012) Improved herbivore resistance in cultivated tomato with the sesquiterpene biosynthetic pathway from a wild relative. Proc Natl Acad Sci 109: 20124–20129

Bleeker PM, Spyropoulou EA, Diergaarde PJ, Volpin H, De Both MTJ, Zerbe P, Bohlmann J, Falara V, Matsuba Y, Pichersky E, et al (2011) RNA-seq discovery, functional characterization, and comparison of sesquiterpene synthases from Solanum lycopersicum and Solanum habrochaites trichomes. Plant Mol Biol 75: 323–336

Carvalho RF, Campos ML, Pino LE, Crestana SL, Zsögön A, Lima JE, Benedito VA, Peres LE (2011) Convergence of developmental mutants into a single tomato model system: "Micro-Tom" as an effective toolkit for plant development research. Plant Methods 7: 18

Coates RM, Denissen JF, Juvik JA, Babka BA (1988) Identification of α-Santalenoic and Endo β-Bergamotenoic Acids as Moth Oviposition Stimulants from Wild Tomato Leaves. J Org Chem 53: 2186–2192

Colby SM, Crock J, Dowdle-Rizzo B, Lemaux PG, Croteau R (1998) Germacrene C synthase from Lycopersicon esculentum cv. VFNT cherry tomato: cDNA isolation, characterization, and bacterial expression of the multiple product sesquiterpene cyclase. Proc Natl Acad Sci U S A 95: 2216–21

Diaz-Chavez ML, Moniodis J, Madilao LL, Jancsik S, Keeling CI, Barbour EL, Ghisalberti EL, Plummer JA, Jones CG, Bohlmann J (2013) Biosynthesis of Sandalwood Oil: Santalum album CYP76F Cytochromes P450 Produce Santalols and Bergamotol. PLoS One 8: e75053

Dudareva N, Andersson S, Orlova I, Gatto N, Reichelt M, Rhodes D, Boland W, Gershenzon J (2005) The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. Proc Natl Acad Sci U S A 102: 933–938

Dudareva N, Klempien A, Muhlemann JK, Kaplan I (2013) Biosynthesis, function and metabolic engineering of plant volatile organic compounds. New Phytol 198: 16–32

Eigenbrode SD, Trumble JT, Millar JG, White KK (1994) Topical Toxicity of Tomato Sesquiterpenes to the Beet Armyworm and the Role of These Compounds in Resistance Derived from an Accession of Lycopersicon hirsutum f. typicum. J Agric Food Chem 42: 807–810

Eigenbrode SD, Trumble JT, White KK (1996) Trichome exudates and resistance to beet armyworm (Lepidoptera: Noctuidae) in Lycopersicon hirsutum f. typicum accessions. Environ Entomol 25: 90–95

Estévez JM, Cantero A, Reindl A, Reichler S, León P (2001) 1-Deoxy-D-xylulose-5-phosphate Synthase, a Limiting Enzyme for Plastidic Isoprenoid Biosynthesis in Plants. J Biol Chem 276: 22901–22909

Falara V, Akhtar TA, Nguyen TTH, Spyropoulou EA, Bleeker PM, Schauvinhold I, Matsuba Y, Bonini ME, Schilmiller AL, Last RL, et al (2011) The tomato terpene synthase gene family. Plant Physiol 157: 770–89

Frelichowski JE, Juvik JA (2005) Inheritance of sesquiterpene carboxylic acid synthesis in crosses of Lycopersicon hirsutum with insect-susceptible tomatoes. Plant Breed 124: 277–281

Frelichowski JE, Juvik JA (2001) Sesquiterpene Carboxylic Acids from a Wild Tomato Species Affect Larval Feeding Behavior and Survival of Helicoverpa zea and Spodoptera exigua (Lepidoptera: Noctuidae). J Econ Entomol 94: 1249–1259

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Report 13: 207–209

Glas JJ, Schimmel BCJ, Alba JM, Escobar-Bravo R, Schuurink RC, Kant MR (2012) Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores. Int J Mol Sci 13: 17077–17103

Gonzales-Vigil E, Hufnagel DE, Kim J, Last RL, Barry CS (2012) Evolution of TPS20-related terpene synthases influences chemical diversity in the glandular trichomes of the wild tomato relative Solanum habrochaites. Plant J cell Mol Biol 71: 921–935

Hoeven van der RS, Monforte AJ, Breeden D, Tanksley SD, Steffens JC (2000) Genetic control and evolution of sesquiterpene biosynthesis in Lycopersicon esculentum and L. hirsutum. Plant Cell 12: 2283–2294

Kang J-H, Shi F, Jones AD, Marks MD, Howe GA (2010) Distortion of trichome morphology by the hairless mutation of tomato affects leaf surface chemistry. J Exp Bot 61: 1053–1064

Maluf WR, Campos GA, Das Graças Cardoso M (2001) Relationships between trichome types and spider mite (Tetranychus evansi) repellence in tomatoes with respect to foliar zingiberene contents. Euphytica 121: 73–80

Matsuba Y, Nguyen TTH, Wiegert K, Falara V, Gonzales-Vigil E, Leong B, Schäfer P, Kudrna D, Wing RA, Bolger AM, et al (2013) Evolution of a complex locus for terpene biosynthesis in Solanum. Plant Cell 25: 2022–2036

Muñoz-Cárdenas K, Ersin F, Pijnakker J, van Houten Y, Hoogerbrugge H, Leman A, Pappas ML, Duarte MVA, Messelink GJ, Sabelis MW, et al (2017) Supplying high-quality alternative prey in the litter increases control of an above-ground plant pest by a generalist predator. Biol Control 105: 19–26

Nützmann HW, Huang A, Osbourn A (2016) Plant metabolic clusters – from genetics to genomics. New Phytol 211: 771–789

Sallaud WC, Rontein D, Onillon S, Oise Jabè F, Duffé P, Cile Giacalone C, Thoraval S, Escoffier C, Herbette GT, Leonhardt N, et al (2009) A Novel Pathway for Sesquiterpene Biosynthesis from Z,Z-Farnesyl Pyrophosphate in the Wild Tomato Solanum habrochaites. Plant Cell 21: 301–317

Sarmento RA, Lemos F, Bleeker PM, Schuurink RC, Pallini A, Oliveira MGA, Lima ER, Kant M, Sabelis MW, Janssen A (2011) A herbivore that manipulates plant defence. Ecol Lett 14: 229–236

Schilmiller AL, Moghe GD, Fan P, Ghosh B, Ning J, Jones AD, Last RL (2015) Functionally Divergent Alleles and Duplicated Loci Encoding an Acyltransferase Contribute to Acylsugar Metabolite Diversity in *Solanum* Trichomes. Plant Cell 27: 1002–1017

Schilmiller AL, Schauvinhold I, Larson M, Xu R, Charbonneau AL, Schmidt A, Wilkerson C, Last RL, 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

Shavrukov YN (2016) CAPS markers in plant biology. Russ J Genet Appl Res 6: 279–287

Silva GA, Picanço MC, Bacci L, Crespo ALB, Rosado JF, Guedes RNC (2011) Control failure likelihood and spatial dependence of insecticide resistance in the tomato pinworm, Tuta absoluta. Pest Manag Sci 67: 913–920

Simmons AT, Gurr GM (2005) Trichomes of Lycopersicon species and their hybrids: Effects on pests and natural enemies. Agric For Entomol 7: 265–276

Spyropoulou EA, Haring MA, Schuurink RC (2014) RNA sequencing on Solanum lycopersicum trichomes identifies transcription factors that activate terpene synthase promoters. BMC Genomics 15: 1–16

Tanksley SD, Ganal MW, Prince JP, De Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, et al (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132: 1141–1160

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

Tissier A (2012) Glandular trichomes: what comes after expressed sequence tags?. Plant J 70: 51–68

Van Schie CCN, Haring MA, Schuurink RC (2007) Tomato linalool synthase is induced in trichomes by jasmonic acid. Plant Mol Biol 64: 251–63

Vendemiatti E, Zsögön A, Silva GFF e, de Jesus FA, Cutri L, Figueiredo CRF, Tanaka FAO, Nogueira FTS, Peres LEP (2017) Loss of type-IV glandular trichomes is a heterochronic trait in tomato and can be reverted by promoting juvenility. Plant Sci 259: 35–47

Wakil W, Brust GE, Perring TM (2018) Tomato and Management of Associated Arthropod Pests: Past, Present, and Future. Sustain. Manag. Arthropod Pests Tomato. pp 3–12

Xu H, Lybrand D, Bennewitz S, Tissier A, Last RL, Pichersky E (2018a) Production of trans-chrysanthemic acid, the monoterpene acid moiety of natural pyrethrin insecticides, in tomato fruit. Metab Eng 47: 271–278

Xu H, Moghe GD, Wiegert-Rininger K, Schilmiller AL, Barry CS, Last RL, Pichersky E (2018b) Coexpression Analysis Identifies Two Oxidoreductases Involved in the Biosynthesis of the Monoterpene Acid Moiety of Natural Pyrethrin Insecticides in Tanacetum cinerariifolium. Plant Physiol 176: 524–537

Xu J, van Herwijnen ZO, Dräger DB, Sui C, Haring MA, Schuurink RC (2018c) SlMYC1 Regulates type VI glandular trichome formation and terpene biosynthesis in tomato glandular cells. Plant Cell 30: 2988–3005

Yin JL, Wong WS (2019) Production of santalenes and bergamotene in Nicotiana tabacum plants. PLoS One 14: e0203249

APPENDIX

Appendix 1. List of genes introgressed in MT-*Sst2* from *S. habrochaites* LA1777.

Gene	Annotation
Solyc08g005020.2.1	Unknown Protein (AHRD V1)
Solyc08g005030.1.1	Unknown Protein (AHRD V1)
Solyc08g005040.1.1	Unknown Protein (AHRD V1)
Solyc08g005050.2.1	Transcription factor (AHRD V1 *--* Q9M4A8_MAIZE); contains Interpro domain(s) IPR001092 Basic helix-loop-helix dimerisation region bHLH
Solyc08g005060.2.1	SnRK1-interacting protein 1 (AHRD V1 ***-Q9LEH5_HORVU)
Solyc08g005070.2.1	ADP-ribosylation factor GTPase-activating protein 3 (AHRD V1 **** B6TII3_MAIZE); contains Interpro domain(s) IPR001164 Arf GTPase activating protein
Solyc08g005080.1.1	Ubiquitin carboxyl-terminal hydrolase (AHRD V1 **** B9SNV7_RICCO); contains Interpro domain(s) IPR001394 Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2
Solyc08g005090.1.1	Unknown Protein (AHRD V1)
Solyc08g005100.2.1	Zinc-binding family protein (AHRD V1 **-* D7MDD4_ARALY); contains Interpro domain(s) IPR006734 Protein of unknown function DUF597
Solyc08g005110.1.1	Ring H2 finger protein (AHRD V1 *-*- D9ZHD8_HYPPE); contains Interpro domain(s) IPR018957 Zinc finger, C3HC4 RING-type
Solyc08g005120.2.1	Cinnamoyl-CoA reductase-like protein (AHRD V1 *-*-Q9M0B3_ARATH); contains Interpro domain(s) IPR016040 NAD(P)-binding domain
Solyc08g005130.1.1	Inositol-tetrakisphosphate 1-kinase 1 (AHRD V1 **** B6TJ44_MAIZE); contains Interpro domain(s) IPR017427 Inositol-tetrakisphosphate 1-kinase
Solyc08g005140.2.1	Serine/threonine-protein kinase BUD32 (AHRD V1 **** B0WXE7_CULQU); contains Interpro domain(s) IPR002290 Serine/threonine protein kinase
Solyc08g005150.2.1	Ubiquitin ligase (AHRD V1 **-* B3G3Y7_ADIVA); contains Interpro domain(s) IPR006575 RWD

Solyc08g005160.1.1	Unknown Protein (AHRD V1)
Solyc08g005170.2.1	Heat stress transcription factor A3 (AHRD V1 *-*- D1M7W9_SOLLC); contains Interpro domain(s) IPR000232 Heat shock factor (HSF)-type, DNA-binding
Solyc08g005180.1.1	Cathepsin L (AHRD V1 ***- D6WR25_TRICA); contains Interpro domain(s) IPR013128 Peptidase C1A, papain
Solyc08g005190.2.1	Pre-mRNA-splicing factor cwc22 (AHRD V1 *-*- B0W4Z5_CULQU); contains Interpro domain(s) IPR003890 MIF4G-like, type 3
Solyc08g005200.2.1	Syntaxin-like protein (AHRD V1 ***- Q3HRZ4_SOLTU); contains Interpro domain(s) IPR010989 t-SNARE
Solyc08g005210.2.1	Dolichol phosphate-mannose biosynthesis regulatory protein (AHRD V1 **** C0H8B0_SALSA); contains Interpro domain(s) IPR009914 Dolichol phosphate-mannose biosynthesis regulatory
Solyc08g005220.2.1	Plastid lipid-associated protein 3, chloroplastic (AHRD V1 **-- PAP3_BRACM); contains Interpro domain(s) IPR006843 PAP fibrillin
Solyc08g005230.2.1	BZIP transcription factor bZIP108 (Fragment) (AHRD V1 *--- Q0GPJ1_SOYBN); contains Interpro domain(s) IPR012458 Protein of unknown function DUF1664
Solyc08g005240.1.1	Lysine-specific demethylase 5B (AHRD V1 *--* KDM5B_HUMAN); contains Interpro domain(s) IPR013129 Transcription factor jumonji
Solyc08g005250.2.1	Jumonji domain protein (AHRD V1 ***- B9MU66_POPTR)
Solyc08g005260.1.1	Myb family transcription factor-like protein (AHRD V1 *--- Q84UP8_ORYSJ); contains Interpro domain(s) IPR006447 Myb-like DNA-binding region, SHAQKYF class
Solyc08g005270.2.1	Poly polymerase catalytic domain containing protein expressed (AHRD V1 **-- Q84T80_ORYSJ); contains Interpro domain(s) IPR012317 Poly(ADP-ribose) polymerase, catalytic region
Solyc08g005280.1.1	Cellulose synthase-like protein (AHRD V1 **** B9IPJ4_POPTR); contains Interpro domain(s) IPR005150 Cellulose synthase
Solyc08g005290.2.1	BZIP transcription factor 3 (AHRD V1 **** Q94KA7_PHAVU); contains Interpro domain(s) IPR012900 G-box binding, MFMR
Solyc08g005300.1.1	Chaperone DnaJ (AHRD V1 *-*- A8URH8_9AQUI); contains Interpro domain(s) IPR003095 Heat shock protein DnaJ

Solyc08g005310.2.1	Guanine nucleotide-binding protein alpha-1 subunit (AHRD V1 *-*- B6TWS6_MAIZE); contains Interpro domain(s) IPR001019 Guanine nucleotide binding protein (G-protein), alpha subunit
Solyc08g005320.2.1	F-box family protein (AHRD V1 ***- B9GFH4_POPTR); contains Interpro domain(s) IPR006566 FBD-like
Solyc08g005330.1.1	F-box/LRR-repeat protein At3g03360 (AHRD V1 ***- FBL42_ARATH); contains Interpro domain(s) IPR001810 Cyclin-like F-box
Solyc08g005340.1.1	F-box protein At5g03100 (AHRD V1 ***- FB250_ARATH)
Solyc08g005350.2.1	F-box family protein (AHRD V1 ***- B9GFH4_POPTR); contains Interpro domain(s) IPR001810 Cyclin-like F-box
Solyc08g005360.1.1	F-box family protein (AHRD V1 ***- B9GFH4_POPTR); contains Interpro domain(s) IPR001810 Cyclin-like F-box
Solyc08g005370.1.1	F-box family protein (AHRD V1 *-*- B9GFH4_POPTR); contains Interpro domain(s) IPR001810 Cyclin-like F-box
Solyc08g005380.1.1	Unknown Protein (AHRD V1)
Solyc08g005390.1.1	Pentatricopeptide repeat-containing protein (AHRD V1 ***- D7LE02_ARALY); contains Interpro domain(s) IPR002885 Pentatricopeptide repeat
Solyc08g005400.2.1	Cytosine-specific methyltransferase (AHRD V1 ***- B9S1U5_RICCO); contains Interpro domain(s) IPR001525 C-5 cytosine-specific DNA methylase
Solyc08g005410.2.1	Kinase-START 1 (Fragment) (AHRD V1 *--- B9UY89_LOPEL); contains Interpro domain(s) IPR009769 Protein of unknown function DUF1336
Solyc08g005420.2.1	WD repeat-containing protein 5 homolog (AHRD V1 *--- WDR5_DICDI); contains Interpro domain(s) IPR017986 WD40 repeat, region
Solyc08g005430.2.1	Growth-regulating factor 4 (AHRD V1 **-- Q6ZIK5_ORYSJ); contains Interpro domain(s) IPR014977 WRC
Solyc08g005440.2.1	Cc-nbs-lrr, resistance protein
Solyc08g005450.1.1	CAX-interacting protein 4 (AHRD V1 *--- CXIP4_ARATH); contains Interpro domain(s) IPR001878 Zinc finger, CCHC-type
Solyc08g005460.2.1	Protein BREVIS RADIX (AHRD V1 ***- BRX_ARATH); contains Interpro domain(s) IPR013591 Disease resistance/zinc finger/chromosome condensation-like region

Solyc08g005470.2.1	Cell division protein kinase 7 (AHRD V1 ***- CDK7_HUMAN); contains Interpro domain(s) IPR002290 Serine/threonine protein kinase
Solyc08g005480.2.1	Unknown Protein (AHRD V1)
Solyc08g005490.2.1	Dehydration-responsive protein-like (AHRD V1 **--Q5VRG9_ORYSJ); contains Interpro domain(s) IPR004159 Protein of unknown function DUF248, methyltransferase putative
Solyc08g005500.2.1	Cc-nbs-lrr, resistance protein
Solyc08g005510.1.1	Tir-nbs-lrr, resistance protein
Solyc08g005520.2.1	Ankyrin repeat domain protein (AHRD V1 *-*-B3CPL5_WOLPP); contains Interpro domain(s) IPR002110 Ankyrin
Solyc08g005530.2.1	Unknown Protein (AHRD V1)
Solyc08g005540.1.1	Amino acid permease-like protein (Fragment) (AHRD V1 *-*-B2ZME9_9ASTR); contains Interpro domain(s) IPR002293 Amino acid/polyamine transporter I
Solyc08g005550.2.1	Alternative oxidase (AHRD V1 **** C1I1U5_NICGU); contains Interpro domain(s) IPR002680 Alternative oxidase
Solyc08g005560.2.1	Binding protein (AHRD V1 ***- D7MGU4_ARALY)
Solyc08g005570.2.1	DUF647 domain-containing protein (AHRD V1 *---C5GE92_AJEDR); contains Interpro domain(s) IPR006968 Protein of unknown function DUF647
Solyc08g005580.2.1	ABC transporter G family member 14 (AHRD V1 **** AB14G_ARATH); contains Interpro domain(s) IPR013525 ABC-2 type transporter
Solyc08g005590.2.1	SWIB/MDM2 domain protein (AHRD V1 ***-B1LUR3_METRJ); contains Interpro domain(s) IPR019835 SWIB domain
Solyc08g005600.2.1	Unknown Protein (AHRD V1)
Solyc08g005610.2.1	Cytochrome P450
Solyc08g005620.2.1	Gibberellin-regulated family protein (AHRD V1 **--D7L2S3_ARALY); contains Interpro domain(s) IPR003854 Gibberellin regulated protein
Solyc08g005630.2.1	Glucose-methanol-choline oxidoreductase (AHRD V1 ***-D6TF66_9CHLR); contains Interpro domain(s) IPR012400 Long-chain fatty alcohol dehydrogenase

Solyc08g005640.2.1	Ent-kaurene synthase-like protein 1 (AHRD V1 **** Q673F9_HORVD); contains Interpro domain(s) IPR005630 Terpene synthase, metal-binding domain
Solyc08g005650.2.1	Cytochrome P450
Solyc08g005660.1.1	Undecaprenyl pyrophosphate synthase (AHRD V1 *-*- B7GG85_ANOFW); contains Interpro domain(s) IPR001441 Di-trans-poly-cis-decaprenylcistransferase-like
Solyc08g005670.2.1	Ent-kaurene synthase (AHRD V1 ***- B9HI37_POPTR); contains Interpro domain(s) IPR005630 Terpene synthase, metal-binding domain
Solyc08g005680.2.1	Undecaprenyl pyrophosphate synthase (AHRD V1 **** B7GG85_ANOFW); contains Interpro domain(s) IPR001441 Di-trans-poly-cis-decaprenylcistransferase-like
Solyc08g005690.1.1	Unknown Protein (AHRD V1)
Solyc08g005700.1.1	Unknown Protein (AHRD V1)
Solyc08g005710.2.1	Ent-copalyl diphosphate synthase (Fragment) (AHRD V1 *-** B3F2U6_9ORYZ); contains Interpro domain(s) IPR001906 Terpene synthase-like
Solyc08g005720.2.1	Ent-kaurene synthase-like protein 1 (AHRD V1 **** Q673F9_HORVD); contains Interpro domain(s) IPR005630 Terpene synthase, metal-binding domain
Solyc08g005750.1.1	Alcohol acyl transferase (AHRD V1 **-- Q6QLX4_SOLLC)
Solyc08g005760.1.1	Alcohol acetyltransferase (AHRD V1 **-* B1A9J8_CUCME); contains Interpro domain(s) IPR003480 Transferase
Solyc08g005770.2.1	Alcohol acetyltransferase (AHRD V1 **-* B1A9J8_CUCME); contains Interpro domain(s) IPR003480 Transferase
Solyc08g005780.2.1	Beta-amylase (AHRD V1 **** B6SVZ0_MAIZE); contains Interpro domain(s) IPR013781 Glycoside hydrolase, subgroup, catalytic core
Solyc08g005790.2.1	Pectinacetylesterase (Fragment) (AHRD V1 *--- C6JT68_9MYRT); contains Interpro domain(s) IPR004963 Pectinacetylesterase
Solyc08g005800.2.1	Pectinacetylesterase like protein (Fragment) (AHRD V1 **-- Q56WP8_ARATH); contains Interpro domain(s) IPR004963 Pectinacetylesterase

Solyc08g005810.1.1	Aspartic proteinase nepenthesin I (AHRD V1 **-- A9ZMF9_NEPAL); contains Interpro domain(s) IPR001461 Peptidase A1
Solyc08g005820.1.1	Aspartic proteinase nepenthesin I (AHRD V1 **-- A9ZMF9_NEPAL); contains Interpro domain(s) IPR001461 Peptidase A1
Solyc08g005830.1.1	Aspartic proteinase nepenthesin I (AHRD V1 **-- A9ZMF9_NEPAL); contains Interpro domain(s) IPR001461 Peptidase A1
Solyc08g005840.1.1	Aspartic proteinase nepenthesin I (AHRD V1 **-- A9ZMF9_NEPAL); contains Interpro domain(s) IPR001461 Peptidase A1
Solyc08g005850.1.1	Aspartic proteinase nepenthesin I (AHRD V1 **-- A9ZMF9_NEPAL); contains Interpro domain(s) IPR001461 Peptidase A1
Solyc08g005860.2.1	Putrescine-binding periplasmic protein (AHRD V1 *-*- A8FVZ2_SHESH); contains Interpro domain(s) IPR001188 Bacterial periplasmic spermidine/putrescine-binding protein
Solyc08g005870.1.1	MYB transcription factor (AHRD V1 *--- D0VYJ9_DIOKA); contains Interpro domain(s) IPR015495 Myb transcription factor
Solyc08g005880.2.1	Multidrug resistance protein mdtK (AHRD V1 **-- MDTK_PROMH); contains Interpro domain(s) IPR002528 Multi antimicrobial extrusion protein MatE
Solyc08g005890.2.1	Hydroxycinnamoyl transferase (AHRD V1 **-* D2XJ64_9MAGN); contains Interpro domain(s) IPR003480 Transferase
Solyc08g005900.2.1	Zinc finger and SCAN domain containing 29 (Predicted) (AHRD V1 *--- A9X1B1_PAPAN); contains Interpro domain(s) IPR017877 MYB-like
Solyc08g005910.2.1	Ankyrin repeat domain protein (AHRD V1 *-*- C0R3F1_WOLWR); contains Interpro domain(s) IPR002110 Ankyrin
Solyc08g005920.2.1	Unknown Protein (AHRD V1)
Solyc08g005930.1.1	F-box family protein (AHRD V1 ***- D7KL20_ARALY); contains Interpro domain(s) IPR015915 Kelch-type beta propeller
Solyc08g005940.1.1	Trypsin proteinase inhibitor (AHRD V1 ***- Q1WL38_NICAC)
Solyc08g005950.2.1	G protein gamma subunit 2 (AHRD V1 ***- D0VE83_ORYSI)

Solyc08g005960.1.1	Cortical cell-delineating protein (AHRD V1 *--- B6U436_MAIZE); contains Interpro domain(s) IPR013770 Plant lipid transfer protein and hydrophobic protein, helical
Solyc08g005970.2.1	Protein arginine N-methyltransferase (AHRD V1 **** A9T4F9_PHYPA); contains Interpro domain(s) IPR007857 Skb1 methyltransferase
Solyc08g005980.2.1	AT4G35080-like protein (Fragment) (AHRD V1 *--- D6PRD4_9BRAS); contains Interpro domain(s) IPR011541 Nickel/cobalt transporter, high-affinity
Solyc08g005990.2.1	AT4G35080-like protein (Fragment) (AHRD V1 *--- D6PRD4_9BRAS); contains Interpro domain(s) IPR011541 Nickel/cobalt transporter, high-affinity AT4G35080-like protein (Fragment)
Solyc08g006000.2.1	Unknown Protein (AHRD V1); contains Interpro domain(s) IPR009606 Protein of unknown function DUF1218
Solyc08g006010.2.1	WD-repeat domain phosphoinositide-interacting protein 3 (AHRD V1 **-- B6SUF7_MAIZE); contains Interpro domain(s) IPR011046 WD40 repeat-like
Solyc08g006020.2.1	NAC domain transcription factor (AHRD V1 *-** B6U2D4_MAIZE); contains Interpro domain(s) IPR003441 No apical meristem (NAM) protein
Solyc08g006030.2.1	Unknown Protein (AHRD V1)
Solyc08g006040.2.1	40S ribosomal protein S6 (AHRD V1 ***- B6T1H0_MAIZE); contains Interpro domain(s) IPR014401 Ribosomal protein S6, eukaryotic
Solyc08g006050.1.1	Unknown Protein (AHRD V1)
Solyc08g006060.2.1	Protein phosphatase 2C (AHRD V1 ***- C9E7A0_9CILI); contains Interpro domain(s) IPR015655 Protein phosphatase 2C
Solyc08g006070.2.1	AIG2-like protein (AHRD V1 ***- B4FVZ7_MAIZE); contains Interpro domain(s) IPR013024 Butirosin biosynthesis, BtrG-like
Solyc08g006080.1.1	Exostosin (AHRD V1 *-*- Q98SV5_XENLA); contains Interpro domain(s) IPR015338 EXTL2, alpha-1,4-N-acetylhexosaminyltransferase
Solyc08g006090.2.1	Peptidyl-prolyl cis-trans isomerase (AHRD V1 ***- Q9LPC7_ARATH); contains Interpro domain(s) IPR002130 Peptidyl-prolyl cis-trans isomerase, cyclophilin-type

Solyc08g006100.2.1	DUF599 family protein (AHRD V1 ***- Q8T1D5_DICDI); contains Interpro domain(s) IPR006747 Protein of unknown function DUF599
Solyc08g006110.2.1	Opaque 2 (Fragment) (AHRD V1 *--- Q84Y96_ZEAMM); contains Interpro domain(s) IPR011616 bZIP transcription factor, bZIP-1
Solyc08g006120.2.1	Mitochondrial import inner membrane translocase subunit TIM16 (AHRD V1 ***- B6UGA4_MAIZE); contains Interpro domain(s) IPR005341 Protein Transporter, Pam16
Solyc08g006130.1.1	B3 domain-containing protein At1g05920 (AHRD V1 *-*- Y1592_ARATH); contains Interpro domain(s) IPR005508 Protein of unknown function DUF313
Solyc08g006140.1.1	Receptor-like protein kinase (AHRD V1 ***- Q9FZP2_ARATH); contains Interpro domain(s) IPR002290 Serine/threonine protein kinase
Solyc08g006150.2.1	ChaC cation transport regulator-like 1 (AHRD V1 ***- A8KBI8_DANRE); contains Interpro domain(s) IPR006840 ChaC-like protein
Solyc08g006160.2.1	Plastid fibrillin 3 (Fragment) (AHRD V1 **-- A6XBK9_COFCA); contains Interpro domain(s) IPR006843 PAP fibrillin
Solyc08g006170.1.1	Pentatricopeptide repeat-containing protein (AHRD V1 ***- D7KJV7_ARALY); contains Interpro domain(s) IPR002885 Pentatricopeptide repeat
Solyc08g006180.2.1	F-box family protein (AHRD V1 ***- D7MV05_ARALY); contains Interpro domain(s) IPR001810 Cyclin-like F-box
Solyc08g006190.1.1	B3 domain-containing protein Os03g0212300 (AHRD V1 *-*- Y3123_ORYSJ); contains Interpro domain(s) IPR003340 Transcriptional factor B3
Solyc08g006200.1.1	B3 domain-containing transcription factor VRN1 (AHRD V1 *-** VRN1_ARATH); contains Interpro domain(s) IPR003340 Transcriptional factor B3
Solyc08g006210.1.1	B3 domain-containing protein Os03g0212300 (AHRD V1 *-*- Y3123_ORYSJ); contains Interpro domain(s) IPR003340 Transcriptional factor B3
Solyc08g006220.2.1	B3 domain-containing protein Os03g0212300 (AHRD V1 ***- Y3123_ORYSJ); contains Interpro domain(s) IPR003340 Transcriptional factor B3

Solyc08g006230.2.1	B3 domain-containing transcription factor VRN1 (AHRD V1 *-** VRN1_ARATH); contains Interpro domain(s) IPR003340 Transcriptional factor B3
Solyc08g006240.1.1	B3 domain-containing protein LOC_Os12g40080 (AHRD V1 ***- Y1208_ORYSJ); contains Interpro domain(s) IPR003340 Transcriptional factor B3
Solyc08g006250.1.1	Copper transporter (AHRD V1 **** A9PEN3_POPTR); contains Interpro domain(s) IPR007274 Ctr copper transporter
Solyc08g006260.1.1	B3 domain-containing protein Os03g0212300 (AHRD V1 ***-Y3123_ORYSJ); contains Interpro domain(s) IPR003340 Transcriptional factor B3
Solyc08g006270.1.1	B3 domain-containing protein REM1 (AHRD V1 *-*-REM1_BRAOB); contains Interpro domain(s) IPR003340 Transcriptional factor B3
Solyc08g006280.1.1	Transcriptional factor B3 family protein (AHRD V1 *-*-D7LIB2_ARALY); contains Interpro domain(s) IPR003340 Transcriptional factor B3
Solyc08g006290.1.1	Unknown Protein (AHRD V1)
Solyc08g006300.2.1	Fasciclin-like arabinogalactan protein 17 (AHRD V1 *-*-A9XTM2_GOSHI); contains Interpro domain(s) IPR000782 FAS1 domain
Solyc08g006310.2.1	Cellulose synthase-like C1-2 glycosyltransferase family 2 protein (AHRD V1 **-* D8T077_SELML); contains Interpro domain(s) IPR001173 Glycosyl transferase, family 2
Solyc08g006320.2.1	WRKY transcription factor 3 (AHRD V1 ***-A7UGD0_SOLTU); contains Interpro domain(s) IPR003657 DNA-binding WRKY
Solyc08g006330.2.1	UDP-glucose salicylic acid glucosyltransferase (AHRD V1 **** Q9M6E7_TOBAC); contains Interpro domain(s) IPR002213 UDP-glucuronosyl/UDP-glucosyltransferase
Solyc08g006350.2.1	UDP-glucose glucosyltransferase (AHRD V1 **** B6EWX8_LYCBA); contains Interpro domain(s) IPR002213 UDP-glucuronosyl/UDP-glucosyltransferase
Solyc08g006360.1.1	UDP-glucose glucosyltransferase (AHRD V1 **** B6EWX8_LYCBA); contains Interpro domain(s) IPR002213 UDP-glucuronosyl/UDP-glucosyltransferase

Solyc08g006370.1.1	UDP-glucosyltransferase family 1 protein (AHRD V1 ***-C6KI44_CITSI); contains Interpro domain(s) IPR002213 UDP-glucuronosyl/UDP-glucosyltransferase
Solyc08g006380.2.1	UDP-glucosyltransferase family 1 protein (AHRD V1 **** C6KI44_CITSI); contains Interpro domain(s) IPR002213 UDP-glucuronosyl/UDP-glucosyltransferase
Solyc08g006390.1.1	UDP-glucosyltransferase family 1 protein (AHRD V1 ***-C6KI44_CITSI); contains Interpro domain(s) IPR002213 UDP-glucuronosyl/UDP-glucosyltransferase
Solyc08g006400.1.1	UDP-glucosyltransferase family 1 protein (AHRD V1 ***-C6KI44_CITSI); contains Interpro domain(s) IPR002213 UDP-glucuronosyl/UDP-glucosyltransferase
Solyc08g006410.2.1	UDP-glucose glucosyltransferase (AHRD V1 **** B6EWX8_LYCBA); contains Interpro domain(s) IPR002213 UDP-glucuronosyl/UDP-glucosyltransferase
Solyc08g006420.2.1	Myosin-like protein (Fragment) (AHRD V1 *-*-Q84VD2_ORYSJ)
Solyc08g006430.2.1	N-methyl-L-tryptophan oxidase (AHRD V1 *-*-D2ZAX1_9ENTR); contains Interpro domain(s) IPR006281 Sarcosine oxidase, monomeric
Solyc08g006440.2.1	Genomic DNA chromosome 5 P1 clone MXM12 (AHRD V1 *-*- Q9SDA2_ARATH); contains Interpro domain(s) IPR012417 Calmodulin-binding, plant
Solyc08g006450.1.1	Unknown Protein (AHRD V1)
Solyc08g006460.2.1	RING-finger protein like (AHRD V1 *-*- B6SVN2_MAIZE); contains Interpro domain(s) IPR018957 Zinc finger, C3HC4 RING-type
Solyc08g006470.2.1	Zinc finger family protein (AHRD V1 ***- D7MB67_ARALY); contains Interpro domain(s) IPR007087 Zinc finger, C2H2-type
Solyc08g006480.2.1	Unknown Protein (AHRD V1)
Solyc08g006490.1.1	B3 domain-containing protein Os03g0212300 (AHRD V1 ***-Y3123_ORYSJ); contains Interpro domain(s) IPR003340 Transcriptional factor B3
Solyc08g006500.2.1	Glutamate-gated kainate-type ion channel receptor subunit GluR5 (AHRD V1 **** B9HB97_POPTR); contains Interpro domain(s) IPR017103 Ionotropic glutamate-like receptor, plant

Solyc08g006510.2.1	NAD dependent epimerase/dehydratase family protein (AHRD V1 **-- B6TVJ7_MAIZE); contains Interpro domain(s) IPR016040 NAD(P)-binding domain
Solyc08g006520.1.1	GDU1 (AHRD V1 *-*- B6TNE4_MAIZE)
Solyc08g006530.2.1	CONSTANS-like protein (AHRD V1 **-- Q0MQL9_SOLTU); contains Interpro domain(s) IPR010402 CCT domain
Solyc08g006540.2.1	Peptidyl-prolyl cis-trans isomerase (AHRD V1 *-** B1XHQ5_SYNP2); contains Interpro domain(s) IPR001179 Peptidyl-prolyl cis-trans isomerase, FKBP-type
Solyc08g006550.2.1	Rop guanine nucleotide exchange factor 1 (AHRD V1 ***-ROGF1_ARATH); contains Interpro domain(s) IPR005512 Rop nucleotide exchanger, PRONE
Solyc08g006560.2.1	3-oxoacyl- (AHRD V1 ***- A8YCB5_MICAE); contains Interpro domain(s) IPR004655 Beta-ketoacyl-acyl carrier protein synthase III (FabH)
Solyc08g006570.2.1	Unknown Protein (AHRD V1)
Solyc08g006580.2.1	Unknown Protein (AHRD V1)
Solyc08g006590.1.1	Os10g0479800 protein (Fragment) (AHRD V1 *-*-Q0IWY0_ORYSJ); contains Interpro domain(s) IPR009769 Protein of unknown function DUF1336
Solyc08g006600.2.1	Dynamin-2 (AHRD V1 ***- D0MQ53_PHYIN); contains Interpro domain(s) IPR000375 Dynamin central region
Solyc08g006610.2.1	Os06g0571300 protein (Fragment) (AHRD V1 *-*-Q0DBE4_ORYSJ)
Solyc08g006620.2.1	Tsi1-interacting protein TSIP1 (AHRD V1 **-- Q9ZPJ0_TOBAC); contains Interpro domain(s) IPR001305 Heat shock protein DnaJ, cysteine-rich region
Solyc08g006630.2.1	Spindle and kinetochore-associated protein 2 (AHRD V1 ***-SKA2_BOVIN)
Solyc08g006640.2.1	Processive diacylglycerol glucosyltransferase (AHRD V1 **-* C5NXP9_9BACL); contains Interpro domain(s) IPR009695 Monogalactosyldiacylglycerol synthase
Solyc08g006650.2.1	Transmembrane protein 85 (AHRD V1 **-- B2W047_PYRTR); contains Interpro domain(s) IPR009445 Protein of unknown function DUF1077

Solyc08g006660.1.1	Conserved oligomeric Golgi complex subunit 2 (AHRD V1 ***-COG2_HUMAN); contains Interpro domain(s) IPR009316 COG complex component, COG2
Solyc08g006670.2.1	Unknown Protein (AHRD V1)
Solyc08g006680.2.1	Folate/biopterin transporter (AHRD V1 **** B4B3T5_9CHRO); contains Interpro domain(s) IPR004324 Biopterin transport-related protein BT1
Solyc08g006690.1.1	Pectinesterase (AHRD V1 **** D7L333_ARALY); contains Interpro domain(s) IPR006501 Pectinesterase inhibitor
Solyc08g006700.2.1	Enhancer of rudimentary homolog (AHRD V1 ***-C6TKU9_SOYBN); contains Interpro domain(s) IPR000781 Enhancer of rudimentary
Solyc08g006710.2.1	Unknown Protein (AHRD V1)
Solyc08g006720.2.1	Glutathione peroxidase (AHRD V1 ***- D6BR59_9ROSI); contains Interpro domain(s) IPR000889 Glutathione peroxidase
Solyc08g006730.1.1	N-acetyltransferase (AHRD V1 ***- B6SUK9_MAIZE); contains Interpro domain(s) IPR000182 GCN5-related N-acetyltransferase
Solyc08g006740.2.1	Decarboxylase family protein (AHRD V1 ***- B1ILJ6_CLOBK); contains Interpro domain(s) IPR002129 Pyridoxal phosphate-dependent decarboxylase
Solyc08g006750.2.1	Decarboxylase family protein (AHRD V1 ***- B1ILJ6_CLOBK); contains Interpro domain(s) IPR002129 Pyridoxal phosphate-dependent decarboxylase
Solyc08g006760.2.1	Unknown Protein (AHRD V1)
Solyc08g006770.2.1	Anthocyanidin synthase (Fragment) (AHRD V1 **-* P93120_DIACA); contains Interpro domain(s) IPR005123 Oxoglutarate and iron-dependent oxygenase
Solyc08g006780.2.1	Complex interacting protein 9 (AHRD V1 ***-Q8LDX3_ARATH); contains Interpro domain(s) IPR004401 Uncharacterised protein family UPF0133
Solyc08g006790.2.1	Early nodulin-55-1 (Fragment) (AHRD V1 *--- NO551_SOYBN); contains Interpro domain(s) IPR003245 Plastocyanin-like
Solyc08g006800.1.1	Unknown Protein (AHRD V1)
Solyc08g006810.1.1	Fasciclin-like arabinogalactan protein 17 (AHRD V1 *-*-A9XTM2_GOSHI); contains Interpro domain(s) IPR000782 FAS1 domain

Solyc08g006820.2.1	Transmembrane 9 superfamily protein member 4 (AHRD V1 ***-B6SXZ2_MAIZE); contains Interpro domain(s) IPR004240 Nonaspanin (TM9SF)
Solyc08g006830.2.1	Caffeoyl-CoA O-methyltransferase 1 (AHRD V1 **** B6TR72_MAIZE); contains Interpro domain(s) IPR002935 O-methyltransferase, family 3
Solyc08g006840.2.1	30S ribosomal protein S15 (AHRD V1 *-*- A9TLD1_PHYPA); contains Interpro domain(s) IPR009068 S15/NS1, RNA-binding
Solyc08g006850.2.1	Patatin-like phospholipase family protein (AHRD V1 **** Q2QNW3_ORYSJ); contains Interpro domain(s) IPR002641 Patatin
Solyc08g006860.2.1	Patatin (AHRD V1 ***- Q7DMP8_SOLBR); contains Interpro domain(s) IPR002641 Patatin
Solyc08g006870.2.1	BZIP domain class transcription factor (AHRD V1 **--D9ZIQ8_MALDO); contains Interpro domain(s) IPR004249 NPH3
Solyc08g006880.2.1	ABC transporter C family member 2 (AHRD V1 **--AB2C_ARATH); contains Interpro domain(s) IPR001140 ABC transporter, transmembrane region
Solyc08g006890.2.1	Tubulin alpha-3 chain (AHRD V1 ***- B6SPX4_MAIZE); contains Interpro domain(s) IPR002452 Alpha tubulin
Solyc08g006900.2.1	Ribosomal protein L32 (AHRD V1 ***- Q45NI6_MEDSA); contains Interpro domain(s) IPR001515 Ribosomal protein L32e
Solyc08g006910.2.1	WD-40 repeat family protein (AHRD V1 **-- D7LF16_ARALY); contains Interpro domain(s) IPR020472 G-protein beta WD-40 repeat, region
Solyc08g006920.1.1	F-box/LRR-repeat protein 4 (AHRD V1 ***- FBL4_ARATH); contains Interpro domain(s) IPR013101 Leucine-rich repeat 2
Solyc08g006930.2.1	Photosystem I reaction center subunit X psaK (AHRD V1 ***-Q84QE6_TOBAC); contains Interpro domain(s) IPR017493 Photosystem I reaction center, PsaK, plant
Solyc08g006940.2.1	F-box family protein (AHRD V1 ***- D7M2J7_ARALY); contains Interpro domain(s) IPR001810 Cyclin-like F-box
Solyc08g006950.2.1	AT-hook motif nuclear localized protein 2 (AHRD V1 ***-A1A6F0_ORYSJ); contains Interpro domain(s) IPR005175 Protein of unknown function DUF296

Solyc08g006960.2.1	AP-2 complex subunit mu (AHRD V1 ***- B6TIQ4_MAIZE); contains Interpro domain(s) IPR015629 Clathrin coat associated protein AP-50
Solyc08g006970.2.1	Lrr, resistance protein fragment
Solyc08g006980.2.1	Baculoviral IAP repeat-containing protein 3 (AHRD V1 *---B0XJ37_CULQU); contains Interpro domain(s) IPR001841 Zinc finger, RING-type
Solyc08g006990.1.1	Aluminum-activated malate transporter (Fragment) (AHRD V1 *--* C6EP55_SECCE); contains Interpro domain(s) IPR006214 Uncharacterised protein family UPF0005
Solyc08g007000.2.1	Protein phosphatase 2C containing protein (AHRD V1 *-** B6T998_MAIZE); contains Interpro domain(s) IPR015655 Protein phosphatase 2C
Solyc08g007010.2.1	FeS assembly protein SufD (AHRD V1 ***- Q3MG34_ANAVT); contains Interpro domain(s) IPR011542 SUF system FeS cluster assembly, SufD
Solyc08g007020.1.1	NAC domain protein IPR003441 (AHRD V1 *-*-B9N2P7_POPTR); contains Interpro domain(s) IPR003441 No apical meristem (NAM) protein
Solyc08g007030.2.1	Ubiquitin family protein (AHRD V1 ***- D7LHV6_ARALY); contains Interpro domain(s) IPR019387 Domain of unknown function SAYSvFN
Solyc08g007040.2.1	Glycine cleavage system H protein 1 (AHRD V1 ***-B6UHJ7_MAIZE); contains Interpro domain(s) IPR017453 Glycine cleavage H-protein, subgroup
Solyc08g007050.1.1	Cytochrome P450
Solyc08g007060.2.1	Peptide transporter-like protein (AHRD V1 **-* Q9SVS9_ARATH); contains Interpro domain(s) IPR000109 TGF-beta receptor, type I/II extracellular region
Solyc08g007070.2.1	Unknown Protein (AHRD V1)
Solyc08g007080.2.1	Inositol 1 4 5-trisphosphate 5-phosphatase (AHRD V1 **** Q712G2_ARATH); contains Interpro domain(s) IPR000300 Inositol polyphosphate related phosphatase
Solyc08g007090.1.1	Expansin-like protein (AHRD V1 ***- A7X331_SOLLC); contains Interpro domain(s) IPR007117 Pollen allergen/expansin, C-terminal

Solyc08g007100.2.1	HAD-superfamily hydrolase subfamily IA variant 3 (AHRD V1 *--- B2IYD2_NOSP7); contains Interpro domain(s) IPR005834 Haloacid dehalogenase-like hydrolase
Solyc08g007110.2.1	ATP dependent helicase (AHRD V1 **-* A8NEE6_COPC7); contains Interpro domain(s) IPR018999 RNA helicase UPF1, UPF2-interacting domain
Solyc08g007120.2.1	Inhibitor of apoptosis-like protein (AHRD V1 *--- B6T3H2_MAIZE); contains Interpro domain(s) IPR017066 S-ribonuclease binding protein, SBP1, pollen
Solyc08g007130.2.1	Beta-amylase 8 (AHRD V1 **** D7MC27_ARALY); contains Interpro domain(s) IPR001554 Glycoside hydrolase, family 14

List according to Sol Genomics Network database (http://solgenomics.net/) ITAG 2.3.

2. NATURAL GENETIC VARIATION CONTROLLING BOTH TYPE-VI TRICHOME MORPHOLOGY AND PLASTID-DERIVED SESQUITERPENE PRODUCTION SYNERGISTICALLY DETERMINE THE GLAND STORAGE CAPACITY IN TOMATO

Abstract

Glandular trichomes can accumulate and release defensive compounds against herbivores pests. In the wild tomato *Solanum habrochaites*, type-VI glandular trichomes have a round-shape morphology, differing from the mushroom-shaped trichomes of the cultivated tomato. The wild species morphology is associated with a large intercellular gland cavity, which can accumulate high levels of plastid-derived sesquiterpene insecticides. By contrast, type-VI trichome glands of cultivated tomato, which lack the plastid-derived sesquiterpene pathway, contain a small intercellular space. For this reason, the transference of the genetic determinants of the wild species' type-IV trichome morphology and functionality to the cultivated tomato is of great interest. Despite this, so far, not much is known about the molecular mechanism underlying the development of the gland, its cavity size and the consequent capacity to store insecticides compounds. Here, we tested the hypothesis that the distinct trichomes of the wild species are caused by a combination of genes regulating both trichome development and metabolism. Using a near isogenic line named MT-*Pincushion-like* (MT-*Pik*) we showed that MT-*Pik* plants carry part of the allelic variation responsible for the differential morphology of type-VI trichomes presented in the parent *S. habrochaites*. Using genotyping by sequencing (GBS) and RNAseq analysis we mapped the *Pik* locus and the associated candidate genes. We further demonstrated that transferring plastid-derived sesquiterpenes biosynthesis capacity from *S. habrochaites* to type-VI trichomes of MT-*Pik* causes a synergistic increase in gland cavity. This work paves the way for both the understanding of the morphology and functionality of type-VI trichomes and for breeding insect resistant tomatoes.

Keywords: Tomato; Introgression line; Glandular trichome; Terpenes.

2.1. Introduction

Trichomes are epidermal outgrowths present in leaves, stems and floral organs of many higher plants. They represent physical and chemical barriers and are generally classified in glandular or non-glandular (Wagner *et al.*, 2004). Glandular trichomes have in common the capacity to synthesize and secrete different amounts of secondary metabolites (Glas *et al.*, 2012). As a consequence, glandular trichomes have become commercially important (Schilmiller *et al.*, 2008; Tissier, 2012).

In tomato and related wild species, different glandular types of trichomes were identified. According to Luckwill (1943) trichome types I, IV, VI and VII were classified as glandular trichomes. The presence of these structures differs in number, type and size according to the species analyzed (Bergau *et al.*, 2015). Moreover, there are differences within the same type, as in type-VI trichome. This type of glandular trichome is one of the most outspread in several tomato-

related species (Kang *et al.*, 2010; Glas *et al.*, 2012; Tian *et al.*, 2012; Balcke *et al.*, 2017). However, although type-VI trichomes are present in both wild and cultivated tomatoes, in the wild species type-VI glandular trichome has a distinct round-shaped morphology, contrasting to the mushroom-shaped trichomes of cultivated tomatoes (Bergau *et al.*, 2015). Both species contains type-VI trichomes with the glandular head formed by four glandular cells, however, the wild species has a large intercellular cavity which accumulate high levels of compounds, as consequence, type-VI trichomes in *S. habrochaites* display almost a perfect circular shape (Bennewitz *et al.*, 2018). In contrast, the cultivated tomato, type-VI trichomes contain a very small intercellular space to accumulate the compounds (Bergau *et al.*, 2015).

For this reason, the presence of the round-shaped type-VI trichome is of great interest. This type of trichome is associated with the production and accumulation of methyl ketones in *Solanum habrochaites glabratum* (Fridman *et al.*, 2005) and mono and sesquiterpenes in *Solanum habrochaites* species (Hoeven van der *et al.*, 2000; Falara *et al.*, 2011; Matsuba *et al.*, 2013). Both substances have a known effect on the control of various insect and herbivores pest (Magalhães *et al.*, 2001; Simmons and Gurr, 2005; Bleeker *et al.*, 2012; Antonious and Snyder, 2015).

However, so far, the focus of most studies in trichomes were to elucidate the biochemical pathways that led to the biosynthesis of compounds inside of the glandular cavity (Schilmiller *et al.*, 2008, 2009; Falara *et al.*, 2011). Conversely, not much is known about the genetic mechanism underlying the development of the gland, its cavity size and the consequent capacity to store compounds. Just a few genes were identified related to multicellular trichome development, but hitherto, most of the genes described are not specifically related to the gland morphology specification and its related functionality. Instead they are related to the presence or abundance of multicellular trichomes (Yang *et al.*, 2011; Tian *et al.*, 2012; Chang *et al.*, 2018; Xu *et al.*, 2018).

Regarding the development of the trichome type-VI gland, Ben-Israel *et al.*, (2009) previously suggested that the intercellular storage compartment could arise due to the high flux of metabolites accumulated inside the glandular cavity. However, subsequently, Bennewitz *et al.*, (2018) showed that this does not appear to be the major contributing factor. The shape of glandular trichome type-VI from *S. habrochaites* is a quantifiable trait with several QTLs contributing to the spherical trichome head. Additionally, the strongest QTLs found by Bennewitz *et al.*, (2018) located on chromosome 1 and 7 do not contain genes involved in biosynthesis of the same sesquiterpenes found in the wild species and still showed the top-scoring for the round-shape trichome type-VI. Nevertheless, the mechanism and the genes required to create the characteristic round type-VI trichome of *S. habrochaites* is still unknown.

Here, we tested the hypothesis that the round trichome glandular shape are built by a combination of genes regulating the production of metabolites and genes controlling the trichome development. Using a near isogenic line named MT-*Pincushion-like* (MT-*Pik*) we showed that MT-*Pik* plants carry at least part of the allelic variation responsible for the differential morphology of type-VI trichomes presented in the wild parent *S. habrochaites* PI127826. We further demonstrated that transferring plastid-derived sesquiterpenes from *S. habrochaites* LA1777 to trichomes type-VI of MT-*Pik* plants causes an increase in gland and cavity cell size. Subsequently, using genotyping by sequencing (GBS) and CAPS molecular markers we found that *Pik* is mapped along the chromosome 1. Finally, using RNAseq database we point out a potentially list of interesting candidates related to cell wall modification, which are in process of being analyzed for their function in the development of glandular type-VI trichomes.

2.2. Materials and Methods

2.2.1 Plant Material and Breeding

Seeds of *Solanum habrochaites* PI127826 were obtained by the Tomato Genetics Resource Center (TGRC - University of California) and Micro-Tom (MT) seeds were donated from Dr. Avram Levy (Weizmann Institute of Science, Israel) and maintained through self-pollination as a true-to-type cultivar since 1998. The near-isogenic line MT-*Sst2*, which contains the plastid-derived the sesquiterpene biosynthesis pathway from *S. habrochaites* LA1777 was introgressed into MT background in a previous study (See Chapter 1).

In order to introgress the genetic determinants of the round type-VI trichome from *S. habrochaites* PI127826 into MT, pollen from the wild species was used to fertilize emasculated MT flowers, creating F1 plants that were further backcrossed (BC) with MT. This process was repeated until the sixth generation of backcross (BC_6). In each BC, plants recombining the MT phenotype with the presence of type-VI trichomes morphologically resembling that of *S. habrochaites* were selected. The BC_6 plants were allowed to self-pollination for six generation and the true-to-type BC_6F_6 line was considered a near isogenic line and named MT-*Pincushion-like* (MT-*Pik*).

Plants were grown in a greenhouse compartment with 30/26°C temperature day/night and 60–75% ambient relative humidity, 11.5 h/13 h (winter/summer) photoperiod, sunlight 250–350 μmol photons m^{-2} s^{-1} PAR irradiance. Seeds were germinated in bulk in 350 mL pots with a 1:1 mixture of commercial potting mix Basaplant® (Base Agro, Artur Nogueira, SP, Brazil) and expanded vermiculite, and was supplemented with 1 g L^{-1} 10:10:10 NPK and 4 g L^{-1} dolomite limestone ($MgCO_3$ + $CaCO_3$). Upon the appearance of the first true leaf, seedlings of each

genotype were individually transplanted in pots containing the soil mix described above, except that NPK supplementation was increased to 8 g L^{-1}.

2.2.2. Trichome counts and phenotyping

Trichome counts were performed as described by Vendemiatti *et al.*, (2017). Both leaf surfaces were dissected along the longitudinal axis in 15 × 3 mm strips covering the middle section of the leaf blade (avoiding the primary veins). Five individuals per genotype were sampled, and four different strips were analyzed per plant. Images were taken using a Leica S8AP0 (Wetzlar, Germany) magnifying glass set to 80x magnification, coupled to a Leica DFC295 camera (Wetzlar, Germany). Counting of trichomes type-VI density (mm^2) was performed using the equipment analytical program.

The morphology of type-VI trichomes was examined under an EVOSfl (www.thermofisher.com) inverted microscope. Lateral leaflets strips were submerged in water under microscope slides and images of type-VI trichomes were taken. All trichome measurements were performed on images of 5 plants per genotype using ImageJ software version 1.4.1. Gland volume and cavity volume were calculated using the volume of the prolate ellipsoid formula: $V = 4/3 \times \pi \times a$ (vertical axis) $\times b^2$ (horizontal axis).

2.2.3. Gas Chromatography-Mass Spectrometry quantification

GC-MS volatile terpene quantification was performed by either collecting individual type-VI trichomes or by dipping leaves in hexane. For the first method, 300 type-VI trichome glands from leaves in the adult vegetative phase (fifth leaf from the cotyledons) were collected with a glass pulled Pasteur pipette under a Leica MZFLIII microscope (www.leica-microsystems.com).

The terpene extraction for the collected type-VI trichome glands was conducted according to the methodology described by Xu *et al.*, (2018). The glands were dissolved in 150 μL of hexane plus 0.5 ng μL of benzyl acetate (Sigma-Aldrich; www.sigmaaldrich.com) as an internal standard. Na_2CO_3 (Sigma-Aldrich) was used to remove water from the hexane. The volatiles were separated using an Agilent (www.agilent.com) 7890A gas chromatograph, attached to an Agilent 7200 accurate-mass quadrupole time-of-flight mass spectrometer. 2μL of the sample was injected heated to 275°C in the injector port and separated on an HP-5ms column (0.25 mm in diameter, 30 m in length, with 0.25 μm film thickness) using Helium as carrier gas (flow rate 1 mL/min). The oven temperature was maintained at 40°C for 3 min and increased by 15°C per min until it reached 250°C and maintained for 3 min. Identification of the compounds was based on the retention time of the chromatographic peaks and their corresponding mass spectra, which were compared to

terpene standards and data libraries. Quantification of peak areas was performed using Masshunter Qualitative Analysis software (Agilent). Peak areas were corrected for the internal standard and quantified using the available terpene standards. Terpene concentration was calculated per trichome gland (ng/gland) using the peak areas relative to the internal (benzyl acetate) and terpene standards available.

For leaf dip methodology, the terpene extraction was performed in leaflets from leaves in the adult vegetative phase (fifth leaf from the cotyledons). One gram of fresh leaflets was dipped in five mL of hexane and shacked on vortex mixer for one min. Next, one mL was collected and added the internal standard y-terpinene (Sigma-Aldrich). The volatiles were separated using a GC-2010 gas chromatography (Shimadzu Corp., Kyoto, Japan) attached to a QP 2010 Plus mass spectrometer (Shimadzu Corp., Kyoto, Japan), with Helium as the charging gas. 1μL of the sample was injected heated to 250°C in the injector port and separated on a Rxi-5ms column (0.25 mm in diameter, 30 m in length, with 0.25 μm film thickness). The data obtained were analyzed using software version 2.5 Lab Solutions-GCMS (Shimadzu Corp., Kyoto, Japan). Identification of the compounds was based on the retention time of the chromatographic peaks and fragments of the mass spectrometer, which were compared to the available standards and data libraries. Terpene concentration was calculated per mg/g of fresh weight (FW) using the peak area relative to the internal standard.

2.2.4. GBS library preparation and Sequencing

GBS library preparation and sequencing was conducted by LGC Genomics (Germany). Briefly, young leaf samples were freeze-dried and powdered in a Star-Beater (VWR, UK) with 5 mm acid-rinsed (0.2M HCl) soda-glass balls in a 2 mL Eppendorf microcentrifuge tube. DNA was extracted from ~50 mg samples with an E.Z.N.A® Plant DNA Kit (Omega Bio-Tek, USA). DNA samples were digested with the restriction enzyme MslI and fragments were ligated to sequencing adapters and sample-specific barcodes. Sequencing of 150 base-pair (bp) paired-end reads was conducted on an Illumina NextSeq 500 V2.

The FASTQ files where pre-processed by LGC Genomics and involved demultiplexing of library groups into individual samples, removal of the sequencing adapters and filtering of reads containing a restriction enzyme site at 5` ends. The resulting processed FASTQ files were aligned to the *Solanum lycopersicum* Heinz 1706 reference genome (SL2.50) with BWA (v0.7.15). BWA can sometimes leave unusual flag information on SAM records, therefore the SAM files were processed with Samtools Fixmate (v1.3.1) to clean up read pairing information and flags. Additionally, InDels

were realigned with GATK's IndelRealigner (v3.8-0) before variant calling with Samtools Mpileup (v1.3.1) and Bcftools Call (v1.3).

The resulting raw VCF files, along with the raw VCF files of the 40x resequenced MT (recurring parent) as well as the VCF files from the resequencing projects of both the 360 Resequencing Project (Lin *et al.*, 2014) and 150 Tomato Genome Resequencing Project (Aflitos *et al.*, 2014) were combined into an index with Tersect (Kurowski and Mohareb, 2019). Tersect was also used to determine which variants were shared between the near isogenic line (NIL) and the corresponding wild parent but not shared with the recurring MT parent. Once the exact accession used as the wild parent in the MT-*Pik* was not in the re-sequenced datasets, a union of all *habrochaites* accessions (total 8 accessions) variants of that species was used instead.

The resulting variants output from Tersect were then filtered as follows, all variants with a quality score less than 20, a mapping quality score below 40 and both a raw read depth below 10 and above 200 were removed. The variant density of the filtered variants over a 10 kb window were plotted across all 12 chromosomes with ggplot2 within R.

2.2.5. Genetic and Physical Mapping

Genomic DNA was extracted from leaflets using the method described by Fulton *et al.*, (1995). Molecular mapping using CAPS markers based on Single Nucleotides Polymorphisms (SNPs) were designed comparing the genomic sequence from *S. lycopersicum* cv. Heinz and *S. habrochaites* LA1777. PCR with specific primers was performed using DreamTaq DNA polymerase (Thermo Scientific) followed by a digest with the respective restriction enzyme and separation on an agarose gel. Primers, restriction enzymes and PCR product sizes are detailed in Table 1.

Table S1. Oligonucleotide sequence used for CAPS markers.

Locus id[1]	Forward and reverse sequence	Enzyme	Fragment	
			MT	PI 127826
Solyc01g088630	TTCGTAACCGTCATCCACAA GTGTATCCTACGGACGCAGA	HincII	89/127	216
Solyc01g096900	TGTGGACTAGAGGTTTGTTTGG TCGGACCAAGGAATCATAGAC	EcoRV	229/122	363
Solyc01g103690	TTTCACCAGCACGAGACAAT ATCCTCACACCACCGAAGAT	NcoI/StyI	462	253/210

[1]Locus according to Sol Genomics Network database (http://solgenomics.net/).

2.2.6. RNA-Sequencing (RNA-Seq) analysis

Total RNA was isolated using the RNeasy Plant Mini Kit (QIAGEN) from 50 to 150mg of trichome material. RNA concentration and quality were assessed using the Agilent 2100 Bioanalyzer instrument (Agilent). mRNA sequencing was done using the Proton Ion system using PI chips and yielded between 25 and 33 million reads per sample. Analysis of mRNA-Seq reads was done using a custom pipeline (https://github.com/BleekerLab/rnaseq-analysis-kallisto-sleuth/releases/tag/v0.1.0) based on Kallisto (version 0.43.1) (Bray *et al.*, 2016) and Sleuth (version 0.29.0) (Pimentel *et al.*, 2017). Normalization was made based on DESeq2. Filtered and normalized read counts were generated for visualization and analysis using Pheatmap package within R. Statistical analysis was performed using "Benjamini-Hochberg" method (Benjamini and Hochberg, 1995) with p-value ≤ 0.05. Genes and samples (1-4) were hierarchically clustered using the clustering metric Euclidean.

2.2.7. Gene ontology enrichment analysis and in silico analysis.

Gene ontology (GO) enrichment analyses were conducted as described by Ashburner *et al.*, (2000) and Carbon *et al.*, (2019). The genes IDs were inserted into the search field in the Panther classification system (http://www.geneontology.org/). GO terms for molecular function, biological process and cellular component were collected.

Gene expression data in tomato tissues were obtained from the Tomato eFP Browser (http://bar.utoronto.ca/efp_tomato/cgi-bin/efpWeb.cgi).

2.2.8. Experimental design and Statistical analysis

Statistical analyses were done using SigmaPlot 11.0 for Windows. The experiments were arranged in a completely randomized design. All data were tested for normality and equal variance by Kolmogorov-Smirnov tests and then the means were further analyzed by two-tailed Student's *t*-test ($P \leq 0.05$) or Fisher's LSD test ($P \leq 0.05$) after one-way ANOVA in multiple comparisons. For data that do not assume a specified variance or normality, we performed Kruskal-Wallis one-way analysis for non-parametric data.

2.3 Results and Discussion

2.3.1 Introgression of *Pik (Pincushion-like)* from *Solanum habrochaites* PI127826 into *S. lycopersicum* cv. Micro-Tom (MT).

Based on a natural genetic variation found in *Solanum habrochaites* PI127826 which has a characteristic round type-VI trichome, we performed genetic introgression aiming to transfer the round shaped type-VI trichome using the tomato cultivar Micro-Tom (MT) as the recurrent parental. To transfer the characteristic round-shaped type-VI trichome from the wild species, we collected pollen from *S. habrochaites* PI127826 and fertilize emasculated MT flowers. The F1 plants were backcrossed (BCs) using MT as a recurrent parental. In each generation, we screened plants for the presence of round-shaped type-VI glandular trichome, using microscopy. We also selected plants that showed longer stalk length, since we notice that this trait segregated with the rounder gland. After six generation of backcrossing and self-pollinations (BC_6F_6), the resulting homozygous genotype was considered a near isogenic line and named MT-*Pincushion-like* (MT-*Pik*). The introgression scheme is shown in Figure 1.

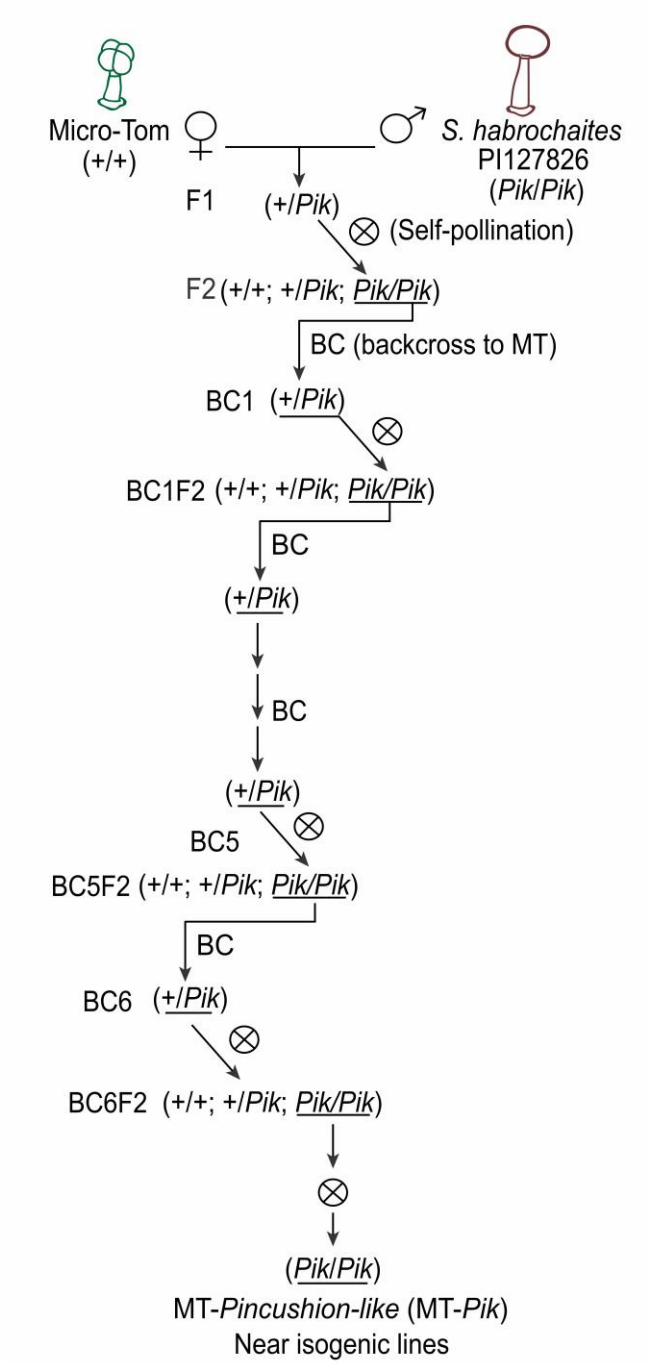

Figure. 1 Scheme of crossing and backcrossing (BC) to create a Micro-Tom (MT) near-isogenic line (NIL) harboring the *Solanum habrochaites* PI127826 alleles for the round-shape trichome type-VI. The NIL BC6Fn was denominated "*Pincunshion-like*" (MT-*Pik*).

The isogenic line MT-*Pik*, had type-VI trichomes with a long stalk, when compared to MT, but still smaller than that of type-VI trichomes found in the wild species (Fig. 2a and b). MT-*Pik* has a small basal cell but did not differ from MT for the intermediate cell (Fig. 2a and b). The density of trichome type-VI also did not differ for both adaxial or abaxial leaf surfaces in MT-*Pik*

compared to MT (Fig. 2c). Contrarily, the wild species showed significantly higher numbers of type-VI trichomes on both leaf surface.

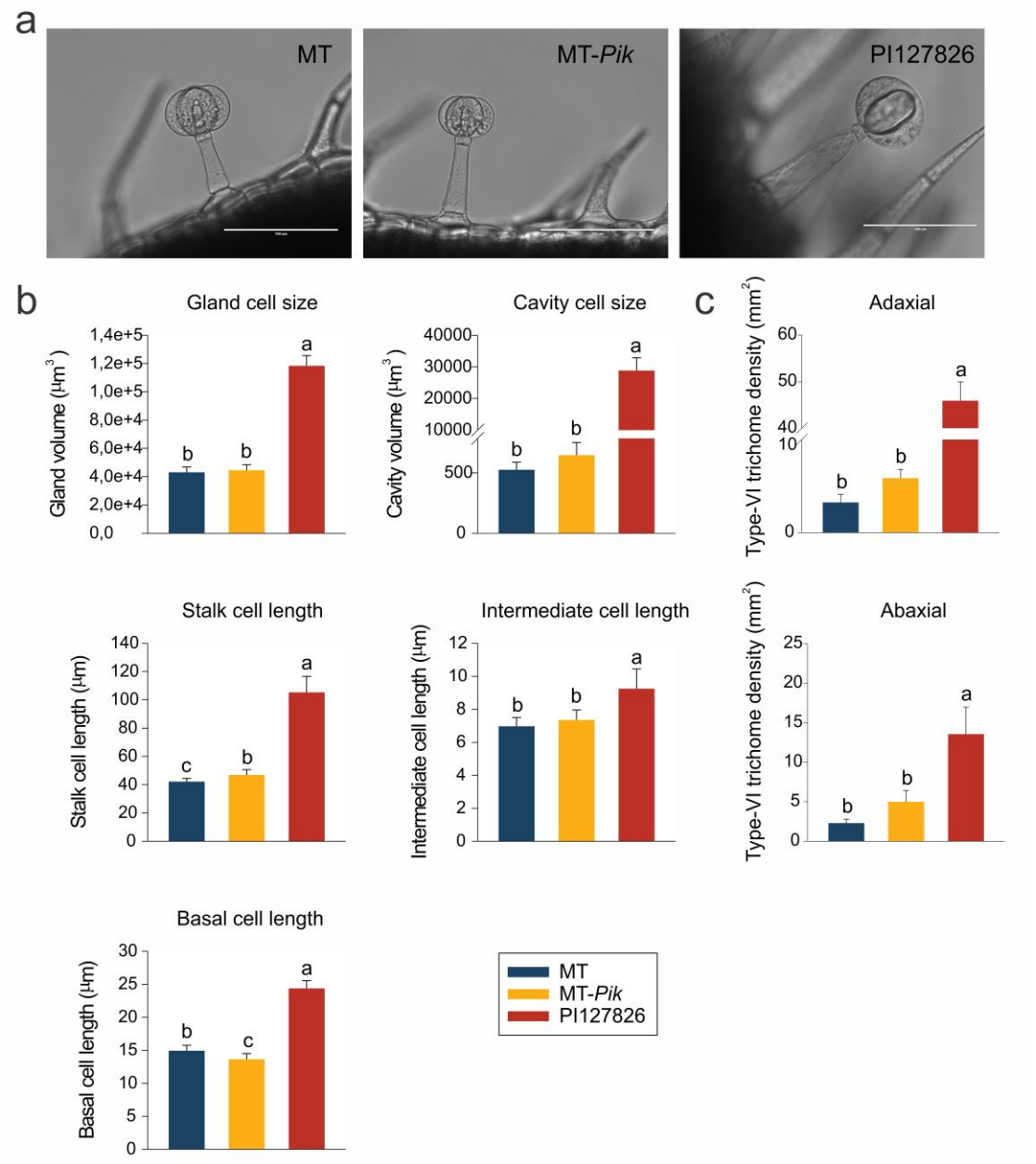

Figure. 2 a. Bright-field microscopy of trichome type-VI on the leaves surface of representative plants of MT, MT-*Pik* and *S. habrochaites* PI127826. Scale bar = 100 µm. **b.** Trichome glandular cell size, cavity cell size, stalk cell, intermediate cell and basal cell length of type-VI trichomes. Data are mean (±SE) of 50 trichomes of 2 replicate leaflets of five plants. **c.** Density (mm^2) of trichome type-VI on adaxial and abaxial surfaces. Data are mean (n=40) for each surface. Bars indicated with different letters were significantly different according to Fisher's LSD test ($P \leq 0.05$) after ANOVA

We next analyzed the gland and internal cavity size of the type-VI glandular trichomes. Although we selected visually plants that showed rounder trichome type-VI using a microscope, we did not find statistical difference in gland and cavity size between MT and MT-*Pik* after applied the prolate ellipsoid formula. Oppositely *S. habrochaites* PI127826 showed bigger value for both gland size and internal cavity size, when compared to MT and the introgressed line MT-*Pik* (Fig. 2a and b).

2.3.2 *Pincushion-like* gene mapping

Next, we used genotyping by sequencing (GBS) approach for the identification of interspecific Single Nucleotides Polymorphisms (SNPs) between the *S. habrochaites* PI127826 and *S. lycopersicum* cv. MT genomes.

The physical map constructed with the identified SNPs showed that the main SNPs were mapped along the chromosome 1 (Fig 3).

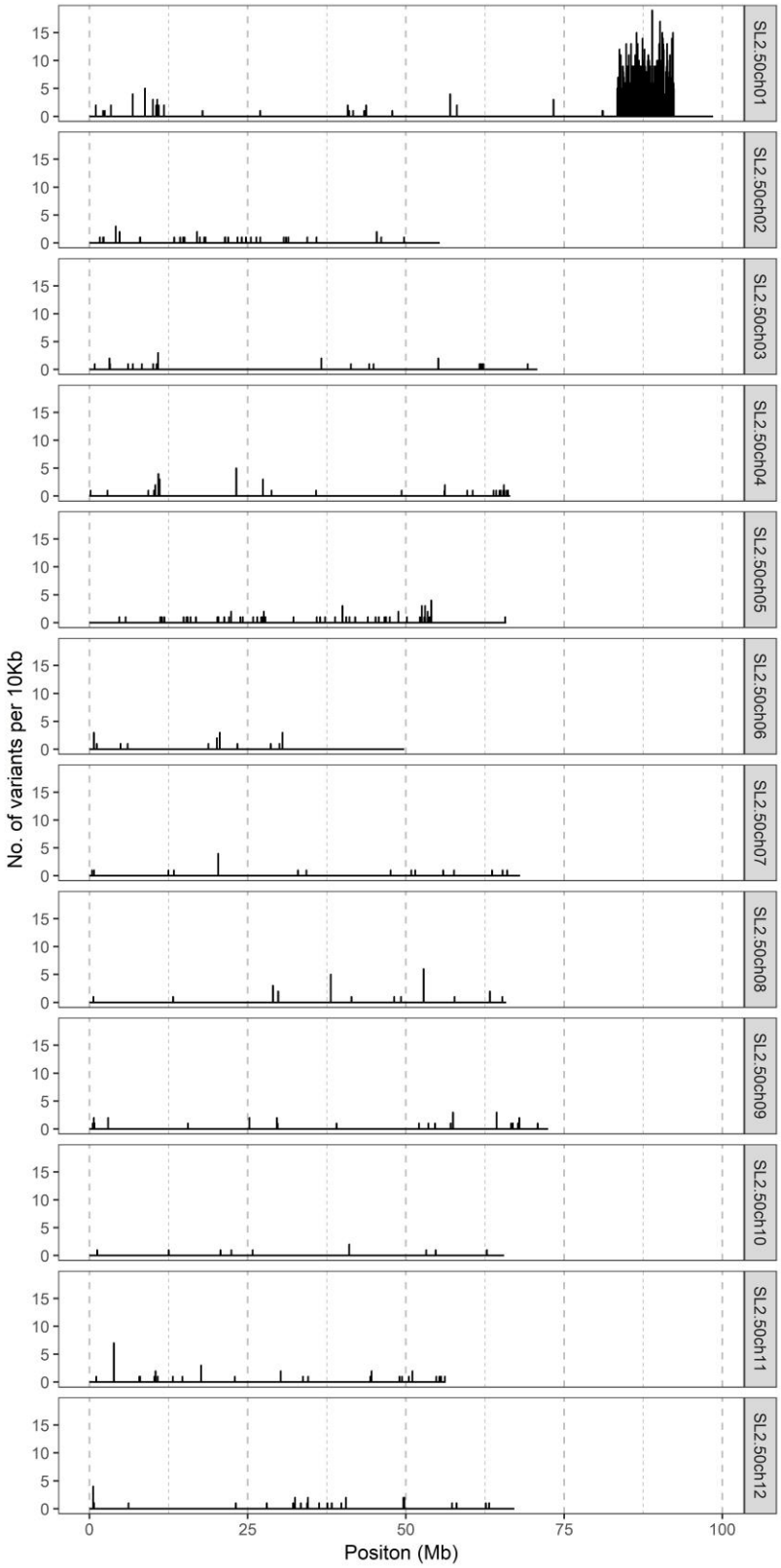

Figure. 3 Distribution of raw read densities for each *S. lycopersicum* chromosome region, showing the region introgressed from *S. habrochaites* PI127826 into MT background.

Thus, the introgressed region spans a genetic interval of approximately 8,91 Mb, between the Solyc01g088640 and Solyc01g103800 markers, translating to roughly 1,112 genes, based on the Heinz genome (SL2.50). Using three CAPS molecular markers designed on SNPs found between *S. lycopersicum* and *S. habrochaites* we provided further evidence that a segment from the chromosome 1 of *S. habrochaites* was introgressed into the MT background and that the alleles of the wild species are in the homozygous form (Table 1).

The region found on chromosome 1 coincides with one of the six QTLs associated with the round shape type-VI trichome previously described by Bennewitz *et al.*, (2018). The six QTLs were distributed in four different chromosomes (1,7, 8 and 11) in a segregating population from a cross between the *S. lycopersicum* var. *cerasiforme* and *S. habrochaites* LA1777. The strongest QTL described for the distinct spherical trichome type-VI was located on chromosome 1 and match with the region introgressed in MT-*Pik*.

2.3.3 Production of terpenes in type-VI trichomes of the *Pincushion-like* line

Taking into account that Ben-Israel *et al.*, (2009) previously suggested that the high flux of metabolites inside the glandular cavity could contribute to the increase in the gland size, we hypothesized that MT-*Pik* might have lost the compounds needed to completely fill the trichome internal cavity during the introgression process. In other words, the lack of genes controlling the high flux of metabolites or the lack of specific pathways would affect the gland to expand like a balloon (See Chapter 1). To confirm this hypothesis, we performed gas chromatography and mass spectrometry (GC-MS) analysis (Fig 4).

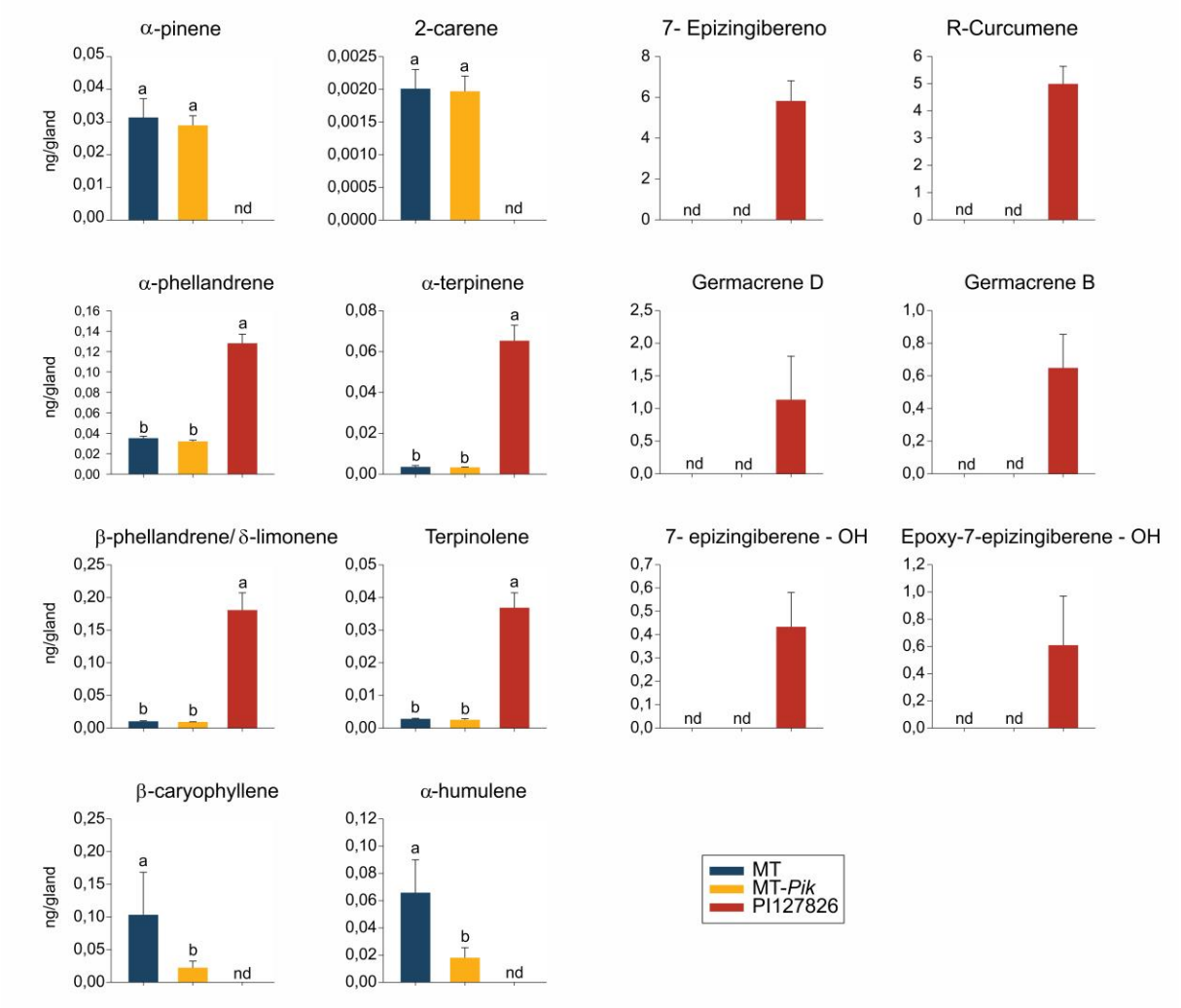

Figure. 4 Volatile terpene levels in type-VI trichomes from MT, MT-*Pik* and the wild species *Solanum habrochaites* PI127826. The data show the amount of each compound present in type-VI trichomes. Each data point represents the mean + SE of five biological replicates. For each sample, 300 type-VI glandular trichomes were collected with a glass capillary before GC-MS. Bars indicated with different letters were significantly different according to Fisher's LSD test ($P \leq 0.05$) after ANOVA. nd, Not detected.

Type-VI trichomes of the MT-*Pik* line accumulate the same monoterpenes (α-pinene, 2-carene, α-phellandrene, α-terpinene, β-phellandrene/ δ-limonene and terpinolene) found in the MT cultivar, whose amounts were statistical equal (Fig 4). However, MT-*Pik* showed fewer amounts of cytosolic sesquiterpenes β-caryophyllene, and α-humulene than MT. These sesquiterpenes are encoded by *SlTPS12* located on chromosome 6 and use *e,e*-farnesyl diphosphate (*e,e*-FPP) as a substrate in the cytosol (Falara *et al.*, 2011). Since the genetic mapping shows that MT-*Pik* has no introgressions from PI127826 at chromosome 6 (Fig 3), the reduced levels of cytosolic

sesquiterpenes in MT-*Pik* might be due to an unknown pleiotropic effect of *Pik*, which primarily controls trichome shape, or due to other genes associated to *Pik* by linkage drag. The β-caryophyllene and α-humulene were not found in the wild parent PI127826. The lack of sesquiterpenes derived from the cytosol in the wild species was expected, since the biosynthesis of sesquiterpenes occurs in the chloroplast in *S. habrochaites* species (Sallaud *et al.*, 2009). At the chromosome 6, *S. habrochaites* TPS9 allele encodes a terpene synthase that produces mainly germacrene B (Bleeker *et al.*, 2011*b*). As expected, we found in the wild species germacrene B and D, however, we did not detect any germacrene in MT or MT-*Pik* (Fig 4).

The wild parental showed high amounts of their characteristic 7-epizingiberene, which is a plastid-derived sesquiterpene whose 7-epizingiberene synthase (ShZIS) enzyme acts in the chloroplast and it is encode by the *ShTPS20* gene in *SsT2* locus on chromosome 8 (Bleeker *et al.*, 2011*a*). We did not detect the same plastid-sesquiterpenes in collected type-VI trichomes of the MT-*Pik* line. This result was already expected, since *ShTPS20* gene was not introgressed (Fig. 3). At this syntenic position on chromosome 8, cultivated tomato has the *SlTPS20* gene, which is responsible for the production of monoterpenes already mentioned. The wild species PI127826 also shows 7-epizingiberene alcohol's derivatives which were not found in MT-*Pik* (Fig 4). The genes involved in the production of these derivatives are currently unknown but they are probably not in the region introgressed comprising the *Pik* locus.

Together, the GC-MS results indicate that the lack of specific biosynthetic pathways in MT-*Pik* precludes its type-VI trichome glands to accumulate the high content of terpenes seen in the *S. habrochaites*. The lack of such compounds might also affect the distinct spherical trichome type-VI expected for the *Pik* genotype.

2.3.4 Synergistic effects of the trichome shape genes and genes involved in the biosynthesis of sesquiterpenes

In order to test if the presence of specific sesquiterpene biosynthesis pathway could contribute for the morphology of type-VI trichomes in MT-*Pik*, we decided to cross MT and MT harboring the *Pik* locus with *S. habrochaites* LA1777 (Fig 5a).

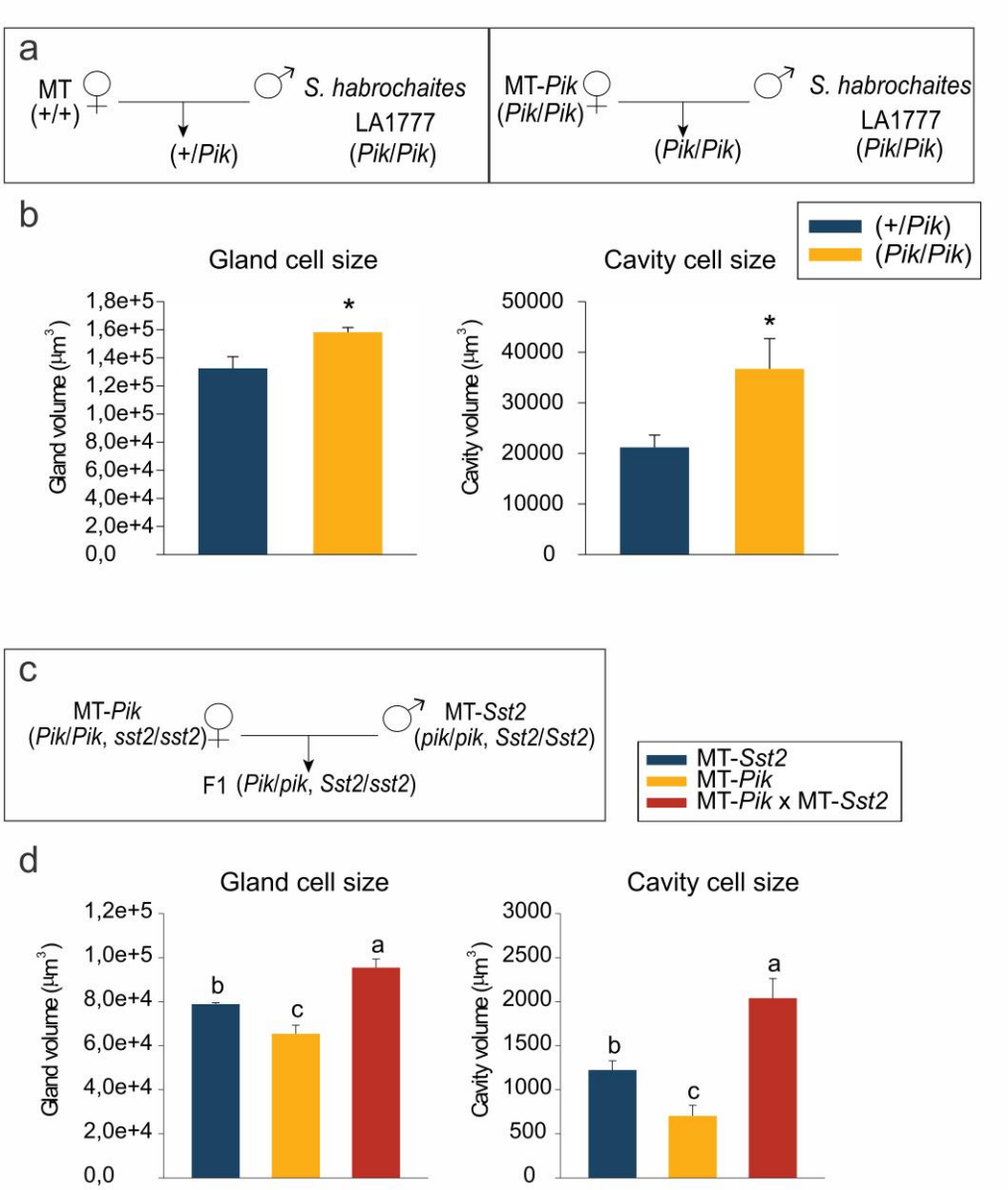

Figure. 5 a. Schematic crossing to create MT and MT-*Pik* hybrid (F1) using *Solanum habrochaites* LA1777 as parental. **b.** Trichome glandular cell size and cavity cell size of type-VI trichomes. Data are mean (±SE) of 50 trichomes of 2 replicate leaflets of five plants. Bars indicated with an asterisk were significantly different according to t-test (P≤ 0.05). **c.** Schematic crossing to create a double mutant between MT-*Pik* and MT-*Sst2*. **d.** Trichome glandular cell and cavity cell size of type-VI trichomes. Data are mean (±SE) of 50 trichomes of 2 replicate leaflets of five plants. Bars indicated with different letters were significantly different according to Fisher's LSD test (P≤ 0.05) after ANOVA.

The wild species *S. habrochaites* LA1777 is closely related to the cultivated tomato and it has been described as naturally resistant to multiple tomato pests. Part of this insect resistance is credit to the presence of the *Sesquiterpene synthase 2* (*SsT2*) locus on chromosome 8, which encodes enzymes responsible for the accumulation of class II plastid-derived sesquiterpenes (α-bergamotene, α-santalene, exo-α-bergamotene, epi-β-santalene, and endo-β-bergamotene) and, in a lesser stent, for the presence of the *Sesquiterpene synthase 1* (*SsT1*) locus on chromosome 6, which encodes enzymes responsible for the accumulation of class I cytosolic-derived sesquiterpenes (mainly germacrene B) in type-VI trichomes (Hoeven van der *et al.*, 2000; Falara *et al.*, 2011).

Given that both *S. habrochaites* LA1777 and PI127826 display the same round-shaped type-VI glandular trichomes (Bergau *et al.*, 2015), they probably have alleles with the same effects at the *Pik* locus. If so, the F1 from the cross MT-*Pik* x LA1777 would harbor two alleles with the same effect for *Pik* while the F1 MT x LA1777 would have just one allele from LA1777 (Fig 5a). For all the other genetic loci, both the F1 from MT and the F1 from MT-*Pik* would harbor just one allele from LA1777. Therefore, all phenotypic differences, when comparing these two sets of F1 plants, should be attributes to the gene dosage at the *Pik* locus, which becomes a proxy to further analyze the *Pik*-related phenotype.

The F1 MT-*Pik* x LA1777 type-VI trichomes showed an increase in glandular and internal cavity size compared to those from the F1 MT x LA1777 (Fig 5b). The cavity cell size from the F1 MT-*Pik* x LA1777 type-VI trichomes is comparable to the one of the wild parental (Fig. 5b and Fig. 2b). This result evidences that MT-*Pik* plants carry at least part of the allelic variation responsible for the differential morphology of type-VI trichomes presented in the wild parent used in the introgression.

Next, we used a near isogenic line (NIL) harboring the alleles of *S. habrochaites* LA1777 at the *SsT2* locus as one attempt to increase the content of sesquiterpenes of the MT-*Pik* plants and to verify the impact of the increased sesquiterpene production capacity on the type-VI trichome morphology (Fig 5c). The NIL named MT-*Sst2* was previously introgressed and accumulates augmented amounts of plastid-derived sesquiterpenes (α-bergamotene, α-santalene, exo-α-bergamotene, epi-β-santalene, and endo-β-bergamotene) found in the will parent. The heterozygous F1 plants harboring both *Pik*/*pik* and *Sst2*/*sst2* alleles showed an increased gland and cavity cell size compared to MT-*Pik* and MT-*Sst2* plants. The increase in the cell cavity volume in the F1 plants combining both *Sst2* and *Pik* alleles was 2.9-fold higher compared to MT-*Pik* (Fig 5d).

Interestingly, among the QTLs controlling the round type-VI trichome found by Bennewitz *et al.*, (2018), one of them coincides with the region present in MT-*Sst2* controlling

sesquiterpene production. As commented before, the other main QTL found in this study also coincides with the *Pik* region. These reinforce the idea that the round type-VI trichome of *S. habrochaites* depends on both genes controlling sesquiterpene metabolism and genes controlling trichome morphology (development) itself.

Concomitantly, we performed GC-MS analysis in MT-*Pik*, MT-*Sst2*, and MT-*Pik* x MT-*Sst2* (F1) plants. The chromatogram showed that the F1 plants had an increasing amount of the plastid-derived sesquiterpenes produced (Fig 6a, peaks 1, 2, 4, 5, and 6). The increase in plastid-derived sesquiterpenes in F1 MT-*Pik* x MT-*Sst2* plants was approximately 1.85-fold higher when compared with MT-*Sst2* (Fig 6b).

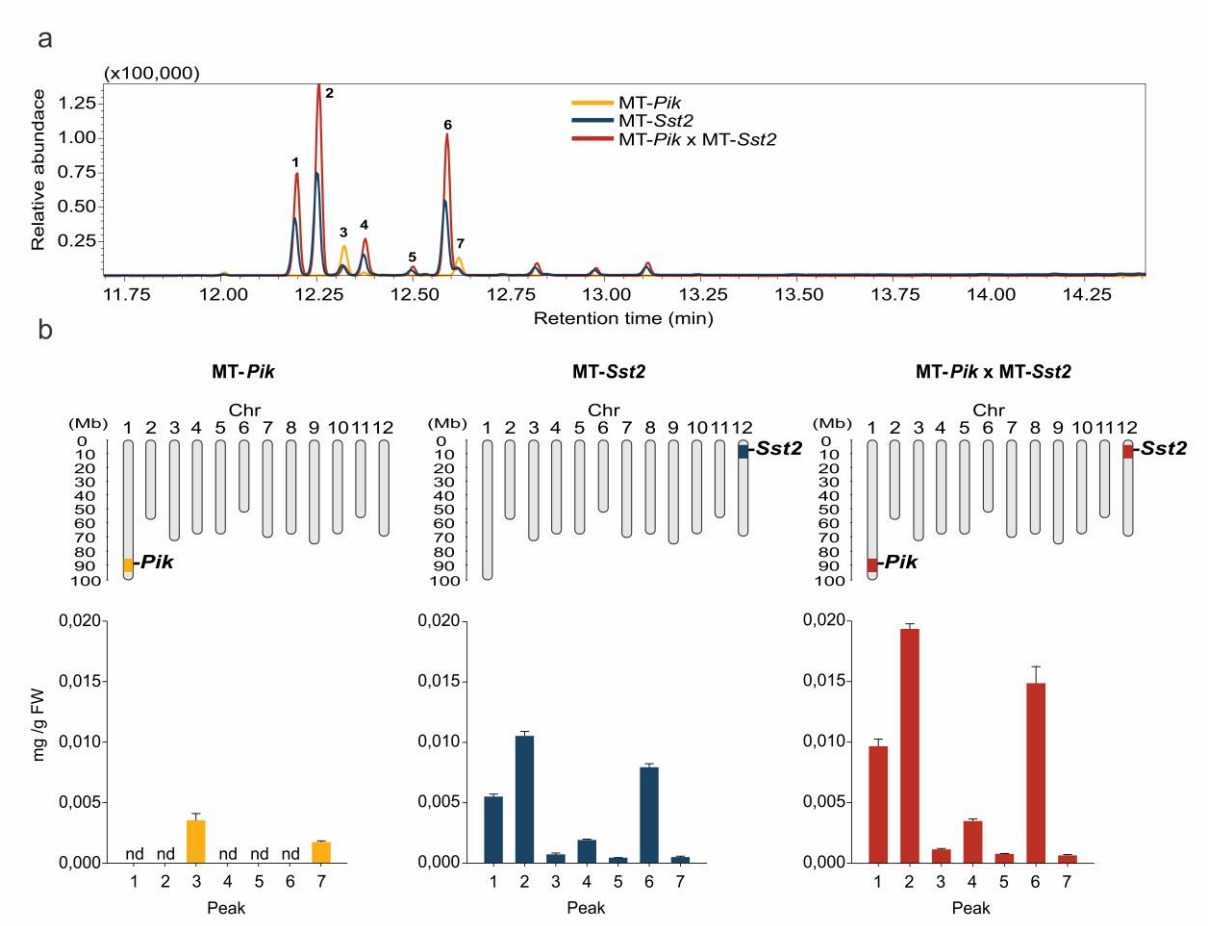

Figure. 6 a. Gas chromatograms showing sesquiterpenes found in fresh leaves. The indicated peaks correspond to the following compounds: (1) α-bergamotene, (2) α-santalene, (3) β-caryophyllene, (4) exo-α-bergamotene, (5) epi-β-santalene, (6) endo-β-bergamotene and (7) α-humulene. The chromatogram shows the detector response for the terpene-specific ion mass 93 **c.** Volatile terpene levels in fresh leaves from MT-*Sst2*, MT-*Pik,* and the double mutant. The data show the amount of each compound (peak). Each data point represents the mean + SE of four biological replicates. **b.** Chromosomal localization of the introgressed region for MT-*Sst2*, MT-P*ik*

and the double heterozygous F1 between MT-*Pik* x MT-*Sst2*. The chromosome numbers are indicated on top of chromosomes. Leaf surface terpenes from MT, MT-*Pik* and the wild species *Solanum habrochaites* PI127826. Each data point represents the mean + SE of five biological replicates. Peak numbers represent (1) α-bergamotene, (2) α-santalene, (3) β-caryophyllene, (4) exo-α-bergamotene, (5) epi-β-santalene, (6) endo-β-bergamotene and (7) α-humulene. nd, Not detected.

Collectively, these results indicate that *Pik* is involved with the morphological determinants of type-VI trichome gland cavity size and that this parameter is also dependent on the capacity to accumulate high amounts of sesquiterpenes.

2.3.5 Candidate genes involved in glandular trichome type-VI cavity size.

Starting from the idea that genes responsible for the distinct spherical gland phenotype of type-VI trichome would be more strongly expressed in the wild species compared to the cultivated tomato, we used RNA sequencing data to reveal the transcriptome changes between trichomes of *S. lycopersicum* and *S. habrochaites* PI127826. From 1,112 genes introgressed into MT background, a total of 873 genes were expressed in trichomes of both species. Of these, 160 genes were statistically more expressed in *S. habrochaites* PI127826 compared to *S. lycopersicum*. Detailed information on gene expression and annotation are presented in the appendix 1.

One of the strongest regulated gene, a lipoxygenase (Solyc01g099160) had more than four-fold higher expression in trichomes of the wild parental. The lipoxygenases (LOXs) act catalyzing the insertion of oxygen into polyunsaturated fatty acids. From this reaction, lipoxygenases can produce active compounds called oxylipins involved in the biosynthesis of jasmonate, associated with a vast number of functions including trichome formation (Huang *et al.*, 2017). This gene is a good candidate since the lipoxygenases are also associated with cell wall regulation. The mutant for the closest *Arabidopsis* homolog AtLOX1 (At1g55020) shows higher cell wall extensibility compared to the wild-type (Mabuchi *et al.*, 2016). AtLOX1 was also detected in guard cells, and a defect in *AtLOX1* compromises the ability of plants to close stomata (Montillet *et al.*, 2013).

Three other lipoxygenases were also highly regulated in trichomes Solyc01g099150, Solyc01g099180 and Solyc01g099190. Except for Solyc01g099190, the other lipoxygenases were previously described on the trichome RNA sequencing data published by Balcke *et al.*, (2017). Solyc01g099150 with the closest *Arabidopsis* homolog AtLOX5 (At3g22400) has been described acting together with AtLOX1 in plant defense, the double mutant *lox1 lox5*, lacks 9-LOX activity and shows susceptibility after bacterial infection (Lõpez *et al.*, 2011).

The other strongest regulated gene found in trichomes was a pectinesterase (Solyc01g091050). The pectinesterase is involved in the deesterification of pectin promoting cell

wall loosening in many processes (Micheli, 2001). A positive staining with ruthenium red indicates that the Arabidopsis trichomes cell wall contains a large portion of pectin (Marks *et al.*, 2008) and labeling cell wall components Bergau *et al.*, (2015) shows a strong signal for esterified pectin particularly during the formation of the inter-cellular trichome cavity, making this gene a potential interesting candidate.

Several transcription factors were highly regulated in the wild species trichomes, including a member of the WRKY transcription factor gene family implicated in multiple regulations of plant development. The WRKY family contains a total of 81 members (Huang *et al.*, 2012). In *Arabidopsis*, *Transparent Testa Glabra 2* (*TTG2*)/*WRKY44* (Solyc10g084380) controls the early development of trichomes (Johnson *et al.*, 2002). *TTG2* was the first member of the WRKY family with a role in trichome development described (Bakshi and Oelmüller, 2014).

Other *SlWRKYs* also have been previously described with expression in tomato trichomes *SlWRKY22* (Solyc01g095100), *SlWRKY28* (Solyc12g011200), *SlWRKY73* (Solyc03g113120) and *SlWRKY78* (Solyc07g055280), and all four were up-regulated by the treatment with jasmonic acid (Spyropoulou *et al.*, 2014). In our, RNA sequencing data the *SlWRKY22* showed two-fold higher expression in the wild parental compared to *S. lycopersicum*. *SlWRKY22* has a highly conserved heptapeptide stretch WRKYGQK at its N-terminus followed by a zinc-finger motif (Huang *et al.*, 2012). The *wrky22 Arabidopsis* mutant shows down-regulation of pectin lyases and expansins suggesting a less degradation of pectin and less loosening of the cell wall (Kloth *et al.*, 2016).

Four other transcription factors containing zinc finger domain(s) were also found highly regulated in the wild trichome species (Solyc01g099340, Solyc01g100330, Solyc01g103320, and Solyc01g103340), whose function is still unclear. Zinc finger TFs constitute one of the largest families of regulatory proteins involved in a wide range of functions (Englbrecht *et al.*, 2004; Pattanaik *et al.*, 2014). GLABROUS INFLORESCENCE STEMS (GIS), GIS2, ZINC FINGER PROTEIN 8 (ZFP8), ZFP5 and ZFP6 were previously described related to trichome development in *Arabidopsis* (Pattanaik *et al.*, 2014). In tomato, the *Hair absent* (*h*), also encodes a single C2H2 zinc-finger transcription factor (Chang *et al.*, 2018).

Another highly expressed transcription factor is a putative cell wall protein SlTFR88 (Solyc01g095170), with the closest *Arabidopsis* homolog Late embryogenesis abundant protein (LEA) (At2g46150). Usually, LEA proteins accumulate in vegetative plant organs exposed to dehydration and might act as stabilizers, hydration buffers, membrane protectants, antioxidants, organic glass formers and/or ion chelators (Tunnacliffe and Wise, 2007; Amara *et al.*, 2014). In Spearmint (*Mentha spicata*) this transcription factor was significantly more abundant in peltate glandular trichomes when compared to leaf without trichomes (Jin *et al.*, 2014). Although LEA

proteins have been found highly expressed in trichomes their precise function in this appendage remains unclear, which could be also related to trichome shape but given their main role in dehydration tolerance it is likely to be the same function in trichomes.

To narrow down the list of genes and identify candidate genes likely involved in trichome glandular shape, we explored the molecular and biological functions of these 160 differentially expressed genes through GO (Gene Ontology) enrichment analysis (Fig.7a).

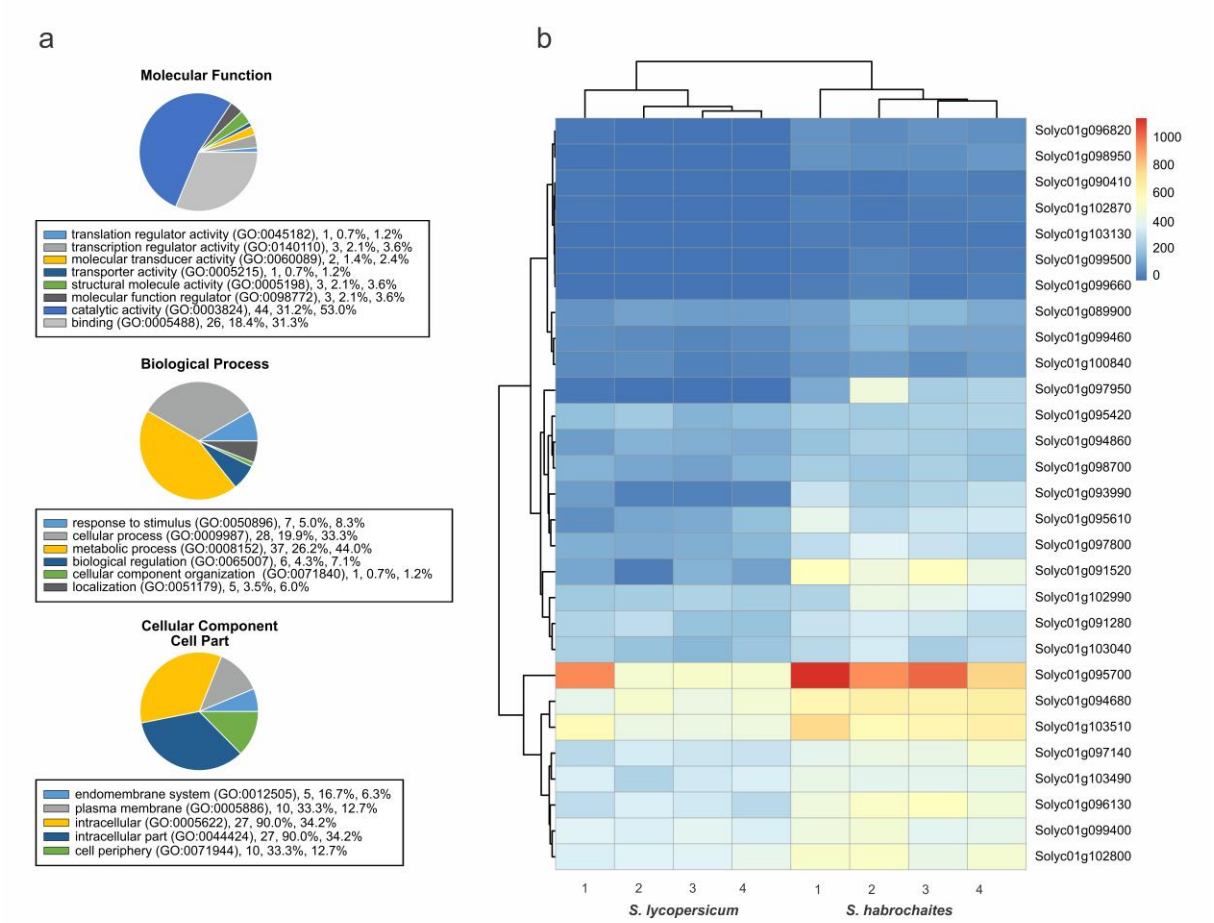

Figure. 7 a GO (Gene Ontology) enrichment analysis of the differentially expressed genes (DEGs) *in S. habrochaites* PI127628. The genes IDs were inserted into the search field in the Panther classification system using *Solanum lycopersicum* as organism. DEGs were grouped into three major functional categories: molecular function, biological process, and cellular component. The values in the box represent: category name (accession), numbers of genes, percent of gene hit against total genes and percent of gene hit against total function hits. **b** Heatmap of RNA-Seq transcriptome analysis for selected genes related to cellular component from gene ontology enrichment analysis. Scale: red indicates high expression and blue is the low expression. Genes and samples (1-4) were hierarchically clustered using the clustering metric Euclidean.

The major GO categories assigned within molecular function term, were catalytic activity (53%) and binding (31,3%). Among the biological functions, GO enrichment for the metabolic process was the major category (44%). This category includes genes with predicted function for lipoxygenases, acyltransferases, isomerases, and proteases. The second major category based on fold enrichment was cellular process (33,3%). Terms related to response to stimulus, biological regulation, cellular component organization and localization were also assigned.

Given that MT-*Pik* might be related to the expansion of the gland cavity and assuming that genes responsible for the distinct spherical trichome type-VI should be related to GO terms for cellular component, we created a heatmap showing those genes included in these terms (Fig.7b).

Although we found a list of potentially interesting candidate genes highly expressed in trichomes, after analyzing the heatmap one of the genes stood out from the list. The potential candidate Solyc01g095610 showed more than two-fold higher expression in collected trichomes in the wild parental PI127826 compared to *S. lycopersicum*. The gene candidate previously annotated as leaf senescence protein-like, with unknown function, has a strong similarity to the proteins Trichome birefringence-like 25 (TBL25) from Arabidopsis (At1g01430). Trichome birefringence belongs to a gene family (*TBR-like* [*TBL*]) with 46 members in Arabidopsis which contain two conserved domains, the DUF231 domain and the TBL (Trichome Birefringence-Like) domain (Bischoff *et al.*, 2010*a*,*b*).

In silico analysis of transcript abundance in tomato showed that the candidate gene is highly expressed in fully expanded leaves and stems (http://bar.utoronto.ca/efp_tomato/cgi-bin/efpWeb.cgi). Another member of the TBL gene family *At5g06700* is preferentially expressed in isolated trichomes showing a high level of expression, 5.11-fold higher compared to shoots without trichomes (Marks *et al.*, 2009). GUS activity was also prominent in leaf and stem trichomes for *At5g06700* (Bischoff *et al.*, 2010*a*).

The *tbr At5g06700* mutant was previously reported as a mutant that is completely impaired in their ability to synthesize secondary wall cellulose (Potikha and Delmer, 1995). Loss of TBR increase pectin methylesterase activity, reduce pectin esterification, and decrease cellulose deposition in trichomes (Bischoff *et al.*, 2010*a*). Recently, four TBLs including the homologous of our gene candidate TBL25 (At1g01430) were found to be capable of transferring acetyl groups onto the mannohexaose acceptor, indicating that TBL25, renamed mannan O-acetyltransferases 3 (MOAT3) is a mannan O-acetyltransferase (Zhong *et al.*, 2018).

Mannose is an abundant constituent of the secondary cell wall in *Arabidopsis* trichomes (Marks *et al.*, 2008). Mutants defective in cell wall acetylation display fragile and collapsed trichomes lacking structural integrity (Suo *et al.*, 2013; Nafisi *et al.*, 2015). Based on the importance of the cell

wall to constrain and regulate cell expansion (Bashline *et al.*, 2014), we hypothesized that this gene could also be potentially interesting candidate determining the glandular trichome shape, however we cannot exclude the existence of others highly expressed genes described here, which also could contribute as an additional effect for the distinct spherical trichome type-VI shape.

2.4. Concluding Remarks

In conclusion, our results reinforce that the distinct spherical trichome type-VI from *Solanum habrochaites* species are built by a combination of genes regulating the production of metabolites and genes controlling trichome development, which might be involved in cell modification. Hence, our results suggest that the turgor pressure caused by the rise of metabolites push the internal cavity to expand, which is likely to depend on genes related to cell wall modifications. Independent of the genetic determinants of the *Pik* phenotype, the results presented here paves the way for both the understanding of the morphology and functionality of type-VI trichomes and for breeding insect resistant tomatoes.

REFERENCES

Aflitos S, Schijlen E, De Jong H, *et al.* 2014. Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. Plant Journal 80, 136–148.

Amara I, Zaidi I, Masmoudi K, Ludevid MD, Pagès M, Goday A, Brini F. 2014. Insights into Late Embryogenesis Abundant (LEA) proteins in plants: From structure to the functions. American Journal of Plant Sciences 5, 3440–3455.

Antonious GF, Snyder JC. 2015. Repellency and oviposition deterrence of wild tomato leaf extracts to spider mites, Tetranychus urticae Koch. Journal of environmental science and health. 50, 667–73.

Ashburner M, Ball CA, Blake JA, *et al.* 2000. Gene ontology: Tool for the unification of biology. Nature Genetics 25, 25–29.

Bakshi M, Oelmüller R. 2014. Wrky transcription factors jack of many trades in plants. Plant Signaling and Behavior 9, e27700.

Balcke GU, Bennewitz S, Bergau N, Athmer B, Henning A, Majovsky P, Jiménez-Gómez JM, Hoehenwarter W, Tissier A. 2017. Multi-Omics of tomato glandular trichomes reveals distinct features of central carbon metabolism supporting high productivity of specialized metabolites. The Plant Cell 29, 960–983.

Bashline L, Lei L, Li S, Gu Y. 2014. Cell wall, cytoskeleton, and cell expansion in higher plants. Molecular Plant 7, 586–600.

Ben-Israel I, Yu G, Austin MB, Bhuiyan N, Auldridge M, Nguyen T, Schauvinhold I, Noel JP, Pichersky E, Fridman E. 2009. Multiple biochemical and morphological factors underlie the production of methylketones in tomato trichomes. Plant physiology 151, 1952–1964.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57, 289–300.

Bennewitz S, Bergau N, Tissier A. 2018. QTL mapping of the shape of type VI glandular trichomes in tomato. Frontiers in Plant Science 9, 1421.

Bergau N, Bennewitz S, Syrowatka F, Hause G, Tissier A. 2015. The development of type VI glandular trichomes in the cultivated tomato Solanum lycopersicum and a related wild species S. habrochaites. BMC Plant Biology 15, 289.

Bischoff V, Nita S, Neumetzler L, Schindelasch D, Urbain A, Eshed R, Persson S, Delmer D, Scheible WR. 2010*a*. TRICHOME BIREFRINGENCE and its homolog AT5G01360 encode plant-specific DUF231 proteins required for cellulose biosynthesis in arabidopsis. Plant Physiology 153, 590–602.

Bischoff V, Selbig J, Scheible WR. 2010*b*. Involvement of TBL/DUF231 proteins into cell wall biology. Plant Signaling and Behavior 5, 1057–1059.

Bleeker PM, Diergaarde PJ, Ament K, *et al.* 2011*a*. Tomato-produced 7-epizingiberene and R-curcumene act as repellents to whiteflies. Phytochemistry 72, 68–73.

Bleeker PM, Mirabella R, Diergaarde PJ, VanDoorn A, Tissier A, Kant MR, Prins M, de Vos M, Haring MA, Schuurink RC. 2012. Improved herbivore resistance in cultivated tomato with the sesquiterpene biosynthetic pathway from a wild relative. Proceedings of the National Academy of Sciences 109, 20124–20129.

Bleeker PM, Spyropoulou EA, Diergaarde PJ, *et al.* 2011*b*. RNA-seq discovery, functional characterization, and comparison of sesquiterpene synthases from Solanum lycopersicum and Solanum habrochaites trichomes. Plant Molecular Biology 75, 323–336.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology 34, 525–527.

Carbon S, Douglass E, Dunn N, *et al.* 2019. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Research 47, D330–D338.

Chang J, Yu T, Yang Q, *et al.* 2018. Hair, encoding a single C2H2 zinc-finger protein, regulates multicellular trichome formation in tomato. Plant Journal 96, 90–120.

Englbrecht CC, Schoof H, Böhm S. 2004. Conservation, diversification and expansion of C2H2 zinc finger proteins in the Arabidopsis thaliana genome. BMC Genomics 5, 1–17.

Falara V, Akhtar TA, Nguyen TTH, *et al.* 2011. The tomato terpene synthase gene family. Plant physiology 157, 770–89.

Fridman E, Wang J, Iijima Y, Froehlich JE, Gang DR, Ohlrogge J, 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. The Plant cell 17, 1252–67.

Fulton TM, Chunwongse J, Tanksley SD. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Molecular Biology Reporter 13, 207–209.

Glas JJ, Schimmel BCJ, Alba JM, Escobar-Bravo R, Schuurink RC, Kant MR. 2012. Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores. International Journal of Molecular Sciences 13, 17077–17103.

Hoeven van der RS, Monforte AJ, Breeden D, Tanksley SD, Steffens JC. 2000. Genetic control and evolution of sesquiterpene biosynthesis in Lycopersicon esculentum and L. hirsutum. The Plant cell 12, 2283–2294.

Huang S, Gao Y, Liu J, Peng X, Niu X, Fei Z, Cao S, Liu Y. 2012. Genome-wide analysis of WRKY transcription factors in Solanum lycopersicum. Mol Genet Genomics 287, 495–513.

Huang H, Liu B, Liu L, Song S. 2017. Jasmonate action in plant growth and development. Journal of Experimental Botany 68, 1349–1359.

Jin J, Panicker D, Wang Q, Kim MJ, Liu J, Yin JL, Wong L, Jang IC, Chua NH, Sarojam R. 2014. Next generation sequencing unravels the biosynthetic ability of Spearmint (Mentha spicata) peltate glandular trichomes through comparative transcriptomics. BMC Plant Biology 14, 2–14.

Johnson CS, Kolevski B, Smyth DR. 2002. Transparent Testa Glabra2, a trichome and seed coat development gene of arabidopsis, encodes a WRKY transcription factor. Plant Cell 14, 1359–1375.

Kang J-H, Shi F, Jones AD, Marks MD, Howe GA. 2010. Distortion of trichome morphology by the hairless mutation of tomato affects leaf surface chemistry. Journal of experimental botany 61, 1053–1064.

Kloth KJ, Wiegers GL, Busscher-Lange J, Van Haarst JC, Kruijer W, Bouwmeester HJ, Dicke M, Jongsma MA. 2016. AtWRKY22 promotes susceptibility to aphids and modulates salicylic acid and jasmonic acid signalling. Journal of Experimental Botany 67, 3383–3396.

Kurowski TJ, Mohareb F. 2019. Tersect: a set theoretical utility for exploring sequence variant data (B Berger, Ed.). Bioinformatics, btz634.

Lin T, Zhu G, Zhang J, *et al.* 2014. Genomic analyses provide insights into the history of tomato breeding. Nature Genetics 46, 1220–1226.

Lõpez MA, Vicente J, Kulasekaran S, Vellosillo T, Martínez M, Irigoyen ML, Cascõn T, Bannenberg G, Hamberg M, Castresana C. 2011. Antagonistic role of 9-lipoxygenase-derived oxylipins and ethylene in the control of oxidative stress, lipid peroxidation and plant defence. Plant Journal 67, 447–458.

Luckwill LC. The genus Lycopersicon: historical, biological, and taxonomic survey of the wild and cultived tomatoes. Aberdeen: Aberdeen University Press, 1943. 44 p.

Mabuchi A, Soga K, Wakabayashi K, Hoson T. 2016. Phenotypic screening of Arabidopsis T-DNA insertion lines for cell wall mechanical properties revealed ANTHOCYANINLESS2, a cell wall-related gene. Journal of Plant Physiology 191, 29–35.

Magalhães STV de, Jham GN, Picanço MC, Magalhães G. 2001. Mortality of second-instar larvae of Tuta absoluta produced by the hexane extract of Lycopersicon hirsutum f. glabratum (PI 134417) leaves. Agricultural and Forest Entomology 3, 297–303.

Marks MD, Betancur L, Gilding E, Chen F, Bauer S, Wenger JP, Dixon RA, Haigler CH. 2008. A new method for isolating large quantities of Arabidopsis trichomes for transcriptome, cell wall and other types of analyses. Plant Journal 56, 483–492.

Marks MD, Wenger JP, Gilding E, Jilk R, Dixon RA. 2009. Transcriptome analysis of arabidopsis wild-type and gl3-sst sim trichomes identifies four additional genes required for trichome development. Molecular Plant 2, 803–822.

Matsuba Y, Nguyen TTH, Wiegert K, *et al.* 2013. Evolution of a complex locus for terpene biosynthesis in Solanum. The Plant Cell 25, 2022–2036.

Micheli F. 2001. Pectin methylesterases: Cell wall enzymes with important roles in plant physiology. Trends in Plant Science 6, 414–419.

Montillet JL, Leonhardt N, Mondy S, *et al.* 2013. An abscisic acid-independent oxylipin pathway controls stomatal closure and immune defense in Arabidopsis. PLoS Biology 11, e1001513.

Nafisi M, Stranne M, Fimognari L, *et al.* 2015. Acetylation of cell wall is required for structural integrity of the leaf surface and exerts a global impact on plant stress responses. Frontiers in Plant Science 6, 550.

Pattanaik S, Patra B, Singh SK, Yuan L. 2014. An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. Frontiers in Plant Science 5, 1–8.

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. Nature Methods 14, 687–690.

Potikha T, Delmer DP. 1995. A mutant of Arabidopsis thaliana displaying altered patterns of cellulose deposition. The Plant Journal 7, 453–460.

Sallaud WC, Rontein D, Onillon S, *et al.* 2009. A Novel Pathway for Sesquiterpene Biosynthesis from Z,Z-Farnesyl Pyrophosphate in the Wild Tomato Solanum habrochaites. The Plant Cell 21, 301–317.

Schilmiller AL, Last RL, Pichersky E. 2008. Harnessing plant trichome biochemistry for the production of useful compounds. The Plant Journal 54, 702–711.

Schilmiller AL, Schauvinhold I, Larson M, Xu R, Charbonneau AL, Schmidt A, Wilkerson C, Last RL, Pichersky E. 2009. Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. Proceedings of the National Academy of Sciences of the United States of America 106, 10865–10870.

Simmons AT, Gurr GM. 2005. Trichomes of Lycopersicon species and their hybrids: Effects on pests and natural enemies. Agricultural and Forest Entomology 7, 265–276.

Spyropoulou EA, Haring MA, Schuurink RC. 2014. RNA sequencing on Solanum lycopersicum trichomes identifies transcription factors that activate terpene synthase promoters. BMC Genomics 15, 1–16.

Suo B, Seifert S, Kirik V. 2013. Arabidopsis glassy hair genes promote trichome papillae development. Journal of Experimental Botany 64, 4981–4991.

Tian D, Tooker J, Peiffer M, Chung SH, Felton GW. 2012. Role of trichomes in defense against herbivores: comparison of herbivore response to woolly and hairless trichome mutants in tomato (Solanum lycopersicum). Planta 236, 1053–1066.

Tissier A. 2012. Glandular trichomes: what comes after expressed sequence tags?. The Plant Journal 70, 51–68.

Tunnacliffe A, Wise MJ. 2007. The continuing conundrum of the LEA proteins. Naturwissenschaften 94, 791–812.

Vendemiatti E, Zsögön A, Silva GFF e, de Jesus FA, Cutri L, Figueiredo CRF, Tanaka FAO, Nogueira FTS, Peres LEP. 2017. Loss of type-IV glandular trichomes is a heterochronic trait in tomato and can be reverted by promoting juvenility. Plant Science 259, 35–47.

Wagner GJ, Wang E, Shepherd RW. 2004. New approaches for studying and exploiting an old protuberance, the plant trichome. Annals of Botany 93, 3–11.

Xu J, van Herwijnen ZO, Dräger DB, Sui C, Haring MA, Schuurink RC. 2018. SlMYC1 Regulates type VI glandular trichome formation and terpene biosynthesis in tomato glandular cells. The Plant Cell 30, 2988–3005.

Yang C, Li H, Zhang J, *et al.* 2011. A regulatory gene induces trichome formation and embryo lethality in tomato. Proceedings of the National Academy of Sciences of the United States of America 108, 11836–11841.

Zhong R, Cui D, Ye ZH. 2018. Members of the DUF231 Family are O-Acetyltransferases Catalyzing 2-O- and 3-O-Acetylation of Mannan. Plant & cell physiology 59, 2339–2349.

APPENDIX

Appendix 1 List of the genes significatively more expressed in trichomes of *S. habrochaites* PI127826 compared to *S. Lycopersicum* cv. Moneymaker.

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g089900	108,08	160,21	Alpha-mannosyltransferase (AHRD V1 **** B7FU21_PHATR); contains Interpro domain(s) IPR005599 Alg9-like mannosyltransferase
Solyc01g090040	9,39	45,35	Unknown Protein (AHRD V1)
Solyc01g090050	6,89	35,11	Unknown Protein (AHRD V1)
Solyc01g090190	2.215,26	3.019,92	Nuclear RNA binding protein (AHRD V1 ***- C7TPG6_SOLTU); contains Interpro domain(s) IPR006861 Hyaluronan/mRNA binding protein
Solyc01g090210	31,09	50,57	Genomic DNA chromosome 5 P1 clone MCA23 (AHRD V1 ***- Q9FIK5_ARATH)
Solyc01g090220	163,43	265,53	Lipase (Fragment) (AHRD V1 *--- Q9ZTW1_DIACA); contains Interpro domain(s) IPR002921 Lipase, class 3
Solyc01g090380	55,14	96,76	Os02g0655100 protein (Fragment) (AHRD V1 ***- Q0DZ09_ORYSJ); contains Interpro domain(s) IPR018881 Uncharacterised protein family UPF0565
Solyc01g090410	5,95	25,60	Palmitoyl protein thioesterase family protein (AHRD V1 **** D7LRP2_ARALY); contains Interpro domain(s) IPR002472 Palmitoyl protein thioesterase
Solyc01g090890	1,48	64,44	Xenotropic and polytropic retrovirus receptor (AHRD V1 **-- B2GU54_XENTR); contains Interpro domain(s) IPR004331 SPX, N-terminal

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g090920	8,72	32,48	EF hand family protein (AHRD V1 ***-B6TBW8_MAIZE); contains Interpro domain(s) IPR011992 EF-Hand type
Solyc01g091030	49,04	147,40	Auxin-responsive family protein (AHRD V1 ***- D7LRT5_ARALY); contains Interpro domain(s) IPR003676 Auxin responsive SAUR protein
Solyc01g091050	5.750,45	21.115,32	Pectinesterase (AHRD V1 ***-B9RD90_RICCO); contains Interpro domain(s) IPR000070 Pectinesterase, catalytic
Solyc01g091170	113,01	728,41	Agmatinase (AHRD V1 ***-Q1IPT1_ACIBL); contains Interpro domain(s) IPR016160 Aldehyde dehydrogenase, conserved site IPR006035 Ureohydrolase
Solyc01g091280	259,74	344,54	Palmitoyltransferase PFA4 (AHRD V1 **** A8N304_COPC7); contains Interpro domain(s) IPR001594 Zinc finger, DHHC-type
Solyc01g091330	79,76	141,33	Maleylacetoacetate isomerase/glutathione S-transferase (AHRD V1 ***-C9Q4X6_9VIBR); contains Interpro domain(s) IPR005955 Maleylacetoacetate isomerase
Solyc01g091350	111,46	392,19	ATP-dependent DNA helicase 2 subunit KU80 (AHRD V1 **-* KU80_ARATH); contains Interpro domain(s) IPR016194 Spen Paralogue and Orthologue C-terminal-like
Solyc01g091380	40,77	61,48	Microtubule-associated protein MAP65-1a (AHRD V1 **** Q9FEV9_TOBAC); contains

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
			Interpro domain(s) IPR007145 MAP65/ASE1
Solyc01g091440	2,17	11,49	ATP-dependent DNA helicase 2 subunit KU80 (AHRD V1 ***- KU80_ORYSJ)
Solyc01g091480	178,11	220,34	Kinesin (AHRD V1 *-*-D2VUW4_NAEGR); contains Interpro domain(s) IPR001752 Kinesin, motor region
Solyc01g091520	113,91	534,56	Enoyl-CoA hydratase/isomerase family protein (AHRD V1 **** D7KSC7_ARALY); contains Interpro domain(s) IPR001753 Crotonase, core
Solyc01g091710	119,75	159,75	F-box family protein (AHRD V1 ***-B9IC68_POPTR)
Solyc01g091720	96,97	142,66	F-box family protein (AHRD V1 ***-B9IC68_POPTR)
Solyc01g091920	16,91	88,94	Subtilisin-like protease (AHRD V1 **--A9XG40_TOBAC); contains Interpro domain(s) IPR015500 Peptidase S8, subtilisin-related
Solyc01g093970	274,82	449,95	Glycosyltransferase (AHRD V1 **--B9IC41_POPTR); contains Interpro domain(s) IPR002495 Glycosyl transferase, family 8
Solyc01g093990	61,87	296,49	Glyoxylate reductase/hydroxypyruvate reductase (AHRD V1 ***-GRHPR_HUMAN); contains Interpro domain(s) IPR006140 D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding
Solyc01g094010	1.810,28	3.455,42	CXE carboxylesterase (AHRD V1 **--Q0ZPW6_9ROSA); contains Interpro domain(s) IPR013094 Alpha/beta hydrolase fold-3

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g094140	5,08	145,48	Cytochrome P450
Solyc01g094240	10,38	24,63	Short-chain dehydrogenase/reductase (AHRD V1 ***- B0KZQ5_CAPAN); contains Interpro domain(s) IPR002347 Glucose/ribitol dehydrogenase
Solyc01g094250	92,23	252,32	Short-chain dehydrogenase/reductase (AHRD V1 ***- B0KZQ5_CAPAN); contains Interpro domain(s) IPR002347 Glucose/ribitol dehydrogenase
Solyc01g094260	4,17	87,63	Short-chain dehydrogenase/reductase (AHRD V1 ***- B0KZQ5_CAPAN); contains Interpro domain(s) IPR002347 Glucose/ribitol dehydrogenase
Solyc01g094270	1,63	8,80	Carbonyl reductase 1 (AHRD V1 **** B6U607_MAIZE); contains Interpro domain(s) IPR002347 Glucose/ribitol dehydrogenase
Solyc01g094320	5,98	22,92	Cysteine and glycine-rich protein 3 (AHRD V1 **-- D2SQP3_PIG); contains Interpro domain(s) IPR001781 Zinc finger, LIM-type
Solyc01g094340	157,79	566,27	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha (AHRD V1 *-*- A0YRC5_LYNSP); contains Interpro domain(s) IPR001095 Acetyl-CoA carboxylase, alpha subunit
Solyc01g094410	587,25	1.507,57	Phosphoribosylanthranilate transferase (Fragment) (AHRD V1 *--- Q43085_PEA); contains Interpro domain(s) IPR013583 Phosphoribosyltransferase C-terminal, plant
Solyc01g094640	923,28	1.264,34	(related) (AHRD V1 ***- Q2HTJ8_MEDTR)
Solyc01g094660	551,57	855,94	Receptor-like protein kinase (AHRD V1 *-*- Q9FZP2_ARATH); contains Interpro

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
			domain(s) IPR002290 Serine/threonine protein kinase
Solyc01g094680	490,31	671,92	Signal peptide peptidase (AHRD V1 **** Q9MA44_ARATH); contains Interpro domain(s) IPR007369 Peptidase A22B, signal peptide peptidase
Solyc01g094810	140,83	474,13	Ubiquitin-conjugating enzyme (AHRD V1 **** D2VN46_NAEGR); contains Interpro domain(s) IPR015581 Ubiquitin-conjugating enzyme
Solyc01g094850	8,77	18,13	Pentatricopeptide repeat-containing protein (AHRD V1 ***- D7LDV7_ARALY); contains Interpro domain(s) IPR002885 Pentatricopeptide repeat
Solyc01g094860	145,14	241,23	Transmembrane protein 8B (AHRD V1 *-*-TMM8B_HUMAN)
Solyc01g094870	6,54	40,82	MRNA clone RAFL21-92-I07 (AHRD V1 *---- Q681W9_ARATH); contains Interpro domain(s) IPR006461 Protein of unknown function Cys-rich
Solyc01g094920	497,82	769,46	Receptor-like kinase (AHRD V1 ***-C6ZRN8_SOYBN)
Solyc01g094940	272,54	392,13	Receptor-like protein kinase (AHRD V1 ***-Q9LYS5_ARATH); contains Interpro domain(s) IPR002290 Serine/threonine protein kinase
Solyc01g094950	2.746,78	3.657,72	Protein BPS1, chloroplastic (AHRD V1 **--BPS1_ARATH)
Solyc01g095020	581,32	812,54	cDNA clone J023132J12 full insert sequence (AHRD V1 ***- B7EL27_ORYSJ)

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g095100	62,96	144,42	WRKY transcription factor 23 (AHRD V1 ***- C9DI12_9ROSI); contains Interpro domain(s) IPR003657 DNA-binding WRKY
Solyc01g095170	201,75	460,39	Harpin-induced protein (AHRD V1 ***- B6SP51_MAIZE); contains Interpro domain(s) IPR010847 Harpin-induced 1
Solyc01g095240	1,16	5,14	Unknown Protein (AHRD V1)
Solyc01g095300	687,60	1.257,13	Glycosyl transferase group 1 (AHRD V1 *-** D2LJY9_RHOVA); contains Interpro domain(s) IPR001296 Glycosyl transferase, group 1
Solyc01g095390	26,94	101,24	SKIP interacting protein 7 (Fragment) (AHRD V1 *-*- B8Q893_ORYSI)
Solyc01g095410	882,38	5.267,00	Eukaryotic translation initiation factor 1A (AHRD V1 **** Q7Y1V3_ORYSJ); contains Interpro domain(s) IPR001253 Translation initiation factor 1A (eIF-1A)
Solyc01g095420	204,46	260,49	Unknown Protein (AHRD V1)
Solyc01g095490	206,37	261,85	Ubiquitin-conjugating enzyme E2 8 (AHRD V1 **** UBC8_ARATH); contains Interpro domain(s) IPR000608 Ubiquitin-conjugating enzyme, E2
Solyc01g095580	375,36	670,69	GH3 family protein (AHRD V1 ***- B9I8V1_POPTR); contains Interpro domain(s) IPR004993 GH3 auxin-responsive promoter
Solyc01g095610	144,37	362,05	Leaf senescence protein-like (AHRD V1 ***- Q654T8_ORYSJ); contains Interpro domain(s) IPR004253 Protein of unknown function DUF231, plant

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g095700	648,84	1.005,02	Abscisic acid receptor PYL8 (AHRD V1 **-* PYL8_ARATH); contains Interpro domain(s) IPR019587 Polyketide cyclase/dehydrase
Solyc01g095720	192,93	609,90	Lipase (AHRD V1 *-*- C1FIH0_9CHLO); contains Interpro domain(s) IPR002921 Lipase, class 3
Solyc01g096050	123,01	276,40	Transcription factor (AHRD V1 *--* Q9M4A8_MAIZE); contains Interpro domain(s) IPR011598 Helix-loop-helix DNA-binding
Solyc01g096070	839,07	1.319,95	Auxin response factor 9 (AHRD V1 **** D7R605_SOLLC); contains Interpro domain(s) IPR010525 Auxin response factor
Solyc01g096080	188,49	401,80	Unknown Protein (AHRD V1)
Solyc01g096130	340,31	535,98	ORM1-like protein 2 (AHRD V1 ***- B4FEN2_MAIZE); contains Interpro domain(s) IPR007203 ORMDL
Solyc01g096230	177,33	488,64	ATPase AAA family protein expressed (AHRD V1 **-- Q8GZX0_ORYSJ); contains Interpro domain(s) IPR003959 ATPase, AAA-type, core
Solyc01g096380	1,18	15,86	Unknown Protein (AHRD V1)
Solyc01g096470	107,23	296,15	Transcription factor (AHRD V1 *--* D7LUT4_ARALY)
Solyc01g096710	705,77	822,78	Nuclear transcription factor Y subunit C-1 (AHRD V1 *--- B6SWV5_MAIZE); contains Interpro domain(s) IPR003958 Transcription factor CBF/NF-Y/archaeal histone
Solyc01g096730	106,79	222,25	Major facilitator superfamily transporter (AHRD V1 **-- A8NUY3_COPC7); contains Interpro domain(s) IPR016196 Major

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
			facilitator superfamily, general substrate transporter
Solyc01g096820	6,73	79,26	Calcium-dependent protein kinase 1 (AHRD V1 ***- Q3YAS9_PETIN); contains Interpro domain(s) IPR002290 Serine/threonine protein kinase
Solyc01g096890	123,54	227,06	Genomic DNA chromosome 5 P1 clone MXI22 (AHRD V1 *-*- Q9FK37_ARATH)
Solyc01g097040	138,70	282,68	IFA-binding protein-like (AHRD V1 *-*- Q5ZC01_ORYSJ); contains Interpro domain(s) IPR007656 Protein of unknown function DUF593
Solyc01g097080	1,11	11,01	Unknown Protein (AHRD V1)
Solyc01g097130	91,88	114,26	Pentatricopeptide repeat-containing protein (AHRD V1 ***- D7LT70_ARALY); contains Interpro domain(s) IPR002885 Pentatricopeptide repeat
Solyc01g097140	343,51	472,19	Dual-specificity tyrosine-phosphatase CDC25 (AHRD V1 **** Q10SX6_ORYSJ); contains Interpro domain(s) IPR001763 Rhodanese-like
Solyc01g097220	3,12	21,32	Ankyrin repeat protein-like (AHRD V1 ***- Q6K268_ORYSJ); contains Interpro domain(s) IPR002110 Ankyrin
Solyc01g097240	303,19	1.072,78	Pathogenesis-related protein 4B (Fragment) (AHRD V1 **-- Q6LBM4_TOBAC); contains Interpro domain(s) IPR018226 Barwin, conserved site IPR001153 Barwin
Solyc01g097340	942,51	2.525,51	NAD-dependent epimerase/dehydratase family protein-like protein (AHRD V1 ***- Q2XPW6_SOLTU); contains Interpro

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
			domain(s) IPR016040 NAD(P)-binding domain
Solyc01g097470	352,34	1.376,04	Neurogenic locus notch protein-like (AHRD V1 ***- B6SSE6_MAIZE)
Solyc01g097500	485,23	724,71	Serine/threonine kinase (AHRD V1 *-** C4QNN2_SCHMA); contains Interpro domain(s) IPR002290 Serine/threonine protein kinase
Solyc01g097510	433,45	640,71	Annexin 3 (AHRD V1 ***-A9X4R3_BRAJU); contains Interpro domain(s) IPR015472 Annexin like protein IPR001464 Annexin
Solyc01g097520	1.475,46	3.482,64	Annexin 11 (AHRD V1 ***-D6QX28_SOYBN); contains Interpro domain(s) IPR009118 Annexin, type plant
Solyc01g097600	69,85	114,73	tRNA pseudouridine synthase D (AHRD V1 ***- D3BIM4_POLPA); contains Interpro domain(s) IPR001656 Pseudouridine synthase, TruD
Solyc01g097800	163,38	339,37	Galactose-6-phosphate isomerase subunit lacB (AHRD V1 *-*- LACB_CLOAB); contains Interpro domain(s) IPR012100 DNA-damage-repair/toleration protein, DRT102
Solyc01g097950	5,85	293,96	Cysteine synthase (AHRD V1 ***-Q9FSF5_TOBAC); contains Interpro domain(s) IPR005859 Cysteine synthase A
Solyc01g097980	502,34	631,40	Kinase family protein (AHRD V1 *-** D7M4D4_ARALY); contains Interpro domain(s) IPR002290 Serine/threonine protein kinase
Solyc01g098110	127,93	193,22	Hydrolase alpha/beta fold family protein (AHRD V1 **-- D7L9I6_ARALY); contains

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g098200	314,89	518,84	Interpro domain(s) IPR007130 Diacylglycerol acyltransferase Vesicle-associated membrane protein-associated protein A (AHRD V1 **--B6UGD8_MAIZE); contains Interpro domain(s) IPR016763 Vesicle-associated membrane protein
Solyc01g098230	206,98	298,07	Mitochondrial import inner membrane translocase subunit TIM44 (AHRD V1 ***-D7L3P3_ARALY); contains Interpro domain(s) IPR007379 Mitochondrial inner membrane translocase complex, subunit Tim44-related
Solyc01g098240	2,37	271,79	RNA polymerase Rpb1 C-terminal repeat domain-containing protein (AHRD V1 *---C5GU31_AJEDR); contains Interpro domain(s) IPR012474 Frigida-like
Solyc01g098370	4,66	155,52	LRR receptor-like serine/threonine-protein kinase, RLP
Solyc01g098700	151,37	241,85	Branched-chain amino acid aminotransferase-like protein (AHRD V1 ***-Q84L60_CICAR); contains Interpro domain(s) IPR001544 Aminotransferase, class IV
Solyc01g098710	103,36	187,72	Aspartyl protease family protein (AHRD V1 **-- D7LT22_ARALY); contains Interpro domain(s) IPR001461 Peptidase A1
Solyc01g098730	253,29	468,01	Unknown Protein (AHRD V1)
Solyc01g098750	164,93	203,51	RING finger protein B (AHRD V1 *---RNGB_DICDI); contains Interpro domain(s) IPR018957 Zinc finger, C3HC4 RING-type

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g098950	3,34	86,72	Glyceraldehyde-3-phosphate dehydrogenase (AHRD V1 ***- Q9XG67_TOBAC); contains Interpro domain(s) IPR000173 Glyceraldehyde 3-phosphate dehydrogenase
Solyc01g099100	280,90	737,63	Long-chain-fatty-acid coa ligase (AHRD V1 ***- Q16WF9_AEDAE); contains Interpro domain(s) IPR000873 AMP-dependent synthetase and ligase
Solyc01g099120	52,28	124,02	Protein AUXIN RESPONSE 4 (AHRD V1 ***- AXR4_ARATH)
Solyc01g099130	355,06	501,12	LysM domain containing protein (AHRD V1 ***- B6TTS5_MAIZE); contains Interpro domain(s) IPR018392 Peptidoglycan-binding lysin domain
Solyc01g099150	39,13	629,18	Lipoxygenase (AHRD V1 **** Q9FT17_SOLLC); contains Interpro domain(s) IPR001246 Lipoxygenase, plant
Solyc01g099160	2.047,93	9.358,01	Lipoxygenase (AHRD V1 **** Q9FT17_SOLLC); contains Interpro domain(s) IPR001246 Lipoxygenase, plant
Solyc01g099180	11,01	79,35	Lipoxygenase (AHRD V1 **** Q9FT17_SOLLC); contains Interpro domain(s) IPR001246 Lipoxygenase, plant
Solyc01g099190	0,00	51,09	Lipoxygenase (AHRD V1 **** Q42873_SOLLC); contains Interpro domain(s) IPR001246 Lipoxygenase, plant
Solyc01g099310	34,82	179,70	RNA polymerase II holoenzyme cyclin-like subunit (AHRD V1 *-** SSN8_GIBMO); contains Interpro domain(s) IPR015432 Cyclin H
Solyc01g099340	47,48	110,18	Zinc finger protein (AHRD V1 *--- Q764N7_MALDO); contains Interpro

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
			domain(s) IPR007087 Zinc finger, C2H2-type
Solyc01g099380	42,07	136,53	U4/U6.U5 tri-snRNP-associated protein 1 (AHRD V1 *--- B0WPY3_CULQU); contains Interpro domain(s) IPR005011 SART-1 protein
Solyc01g099400	398,52	463,76	WD-repeat domain phosphoinositide-interacting protein 3 (AHRD V1 **--B6SUF7_MAIZE); contains Interpro domain(s) IPR017986 WD40 repeat, region
Solyc01g099450	1,72	4,32	Unknown Protein (AHRD V1)
Solyc01g099460	62,68	133,46	ATP-dependent DNA helicase (AHRD V1 *-*- D8G648_9CYAN); contains Interpro domain(s) IPR004589 DNA helicase, ATP-dependent, RecQ type
Solyc01g099470	15,99	38,96	ATP-dependent DNA helicase Q-like SIM (AHRD V1 *-*- RQSIM_ARATH)
Solyc01g099500	4,71	32,61	ATP-dependent DNA helicase RecQ (AHRD V1 *-*- C6MFV5_9PROT); contains Interpro domain(s) IPR004589 DNA helicase, ATP-dependent, RecQ type
Solyc01g099560	0,00	11,64	Carbonyl reductase 3 (AHRD V1 **** B6TRS7_MAIZE); contains Interpro domain(s) IPR002347 Glucose/ribitol dehydrogenase
Solyc01g099650	100,76	268,51	Formin 3 (AHRD V1 *-*-D0QAN4_ARATH); contains Interpro domain(s) IPR015425 Actin-binding FH2
Solyc01g099660	0,79	31,71	Heat shock protein (AHRD V1 ***-Q84KP8_CYAME); contains Interpro domain(s) IPR013126 Heat shock protein 70

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g099700	360,35	418,68	Os07g0673000 protein (Fragment) (AHRD V1 ***- Q0D3Q5_ORYSJ)
Solyc01g099720	254,35	324,04	Alpha subunit of F-actin capping protein (AHRD V1 **** D7L3V5_ARALY); contains Interpro domain(s) IPR018315 F-actin capping protein, alpha subunit, actin binding
Solyc01g099780	115,16	2.897,95	Translationally-controlled tumor protein homolog (AHRD V1 ***- B5XDL0_SALSA); contains Interpro domain(s) IPR001983 Translationally controlled tumour-associated TCTP
Solyc01g099840	2.078,02	5.145,94	Auxin-repressed protein (AHRD V1 ***-B4FA62_MAIZE); contains Interpro domain(s) IPR008406 Dormancyauxin associated
Solyc01g099850	320,03	511,00	Multi-pass transmembrane protein (AHRD V1 ***- Q5CLX7_CRYHO); contains Interpro domain(s) IPR018143 Folate receptor, conserved region
Solyc01g099980	20,60	81,19	Glycine-rich protein (AHRD V1 *---D7M9Z2_ARALY); contains Interpro domain(s) IPR009836 Protein of unknown function DUF1399
Solyc01g100040	191,88	291,83	Integrin-linked kinase-associated serine/threonine phosphatase 2C (AHRD V1 **** ILKAP_HUMAN); contains Interpro domain(s) IPR015655 Protein phosphatase 2C
Solyc01g100050	2.318,61	3.039,80	WD-40 repeat protein (AHRD V1 *---Q8Z020_ANASP); contains Interpro domain(s) IPR017986 WD40 repeat, region

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g100120	4,51	29,70	Kinesin (AHRD V1 ***-Q5MNW6_GOSHI); contains Interpro domain(s) IPR001752 Kinesin, motor region
Solyc01g100140	494,05	744,32	Protein DEHYDRATION-INDUCED 19 homolog 3 (AHRD V1 **-- DI193_ARATH); contains Interpro domain(s) IPR008598 Drought induced 19
Solyc01g100160	31,13	88,73	Protein DEHYDRATION-INDUCED 19 homolog 3 (AHRD V1 ***- DI193_ARATH); contains Interpro domain(s) IPR008598 Drought induced 19
Solyc01g100330	259,06	316,71	Zinc finger matrin type 2 (AHRD V1 ***-Q6DEN8_XENTR); contains Interpro domain(s) IPR003604 Zinc finger, U1-type
Solyc01g100450	30,32	52,70	Pentatricopeptide repeat-containing protein (AHRD V1 ***- D7KJU9_ARALY); contains Interpro domain(s) IPR002885 Pentatricopeptide repeat
Solyc01g100510	65,39	152,93	Knotted-like homeobox protein (AHRD V1 ***- Q49RB7_POPTO); contains Interpro domain(s) IPR001356 Homeobox IPR005539 ELK IPR017970 Homeobox, conserved site
Solyc01g100610	335,19	396,21	Solute carrier family 40 member 1 (AHRD V1 **-- S40A1_DANRE); contains Interpro domain(s) IPR016196 Major facilitator superfamily, general substrate transporter
Solyc01g100760	930,85	1.621,96	Susceptibility homeodomain transcription factor (Fragment) (AHRD V1 *-*-Q8SAA7_ORYSA); contains Interpro domain(s) IPR007493 Protein of unknown function DUF538

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g100840	59,73	97,78	Trafficking protein particle complex subunit 4 (AHRD V1 ***- B6U8Z1_MAIZE); contains Interpro domain(s) IPR007233 Sybindin-like protein
Solyc01g101240	4.259,78	6.418,88	Aspartic proteinase (AHRD V1 **** ASPRX_ORYSJ); contains Interpro domain(s) IPR001461 Peptidase A1
Solyc01g102290	410,77	785,68	Stress-induced hydrophobic peptide (AHRD V1 **-- B9N409_POPTR); contains Interpro domain(s) IPR000612 Uncharacterised protein family UPF0057
Solyc01g102440	3,49	40,30	Organic solute transporter-like (AHRD V1 *-*- Q8H5Q5_ORYSJ); contains Interpro domain(s) IPR005178 Protein of unknown function DUF300
Solyc01g102480	278,90	333,73	Transmembrane protein 34 (AHRD V1 ***- C0H9K2_SALSA); contains Interpro domain(s) IPR005178 Protein of unknown function DUF300
Solyc01g102490	119,82	217,18	NADPH adrenodoxin oxidoreductase fprA (NADPH-ferredoxin reductase) (AHRD V1 **-* A2VNR6_MYCTU); contains Interpro domain(s) IPR000759 Adrenodoxin reductase
Solyc01g102680	6,59	37,16	Receptor like kinase, RLK
Solyc01g102720	335,30	998,81	START domain containing 10 (Fragment) (AHRD V1 **-- Q52LA1_MOUSE); contains Interpro domain(s) IPR002913 Lipid-binding START
Solyc01g102760	645,81	781,52	PHD finger family protein (AHRD V1 ***- D7KBK9_ARALY)

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g102800	408,30	522,48	Histidyl-tRNA synthetase (AHRD V1 *-*-Q52NV4_HUMAN); contains Interpro domain(s) IPR015807 Histidyl-tRNA synthetase, class IIa, subgroup
Solyc01g102850	6,50	33,51	Tir-nbs-lrr, resistance protein
Solyc01g102870	4,81	32,02	LRR receptor-like serine/threonine-protein kinase, RLP
Solyc01g102880	71,77	202,48	Tir-nbs-lrr, resistance protein
Solyc01g102950	87,74	213,00	FAD dependent oxidoreductase (AHRD V1 **-- B5VX05_SPIMA)
Solyc01g102970	469,29	625,05	RNA-binding protein (AHRD V1 *--* Q5CPB6_CRYHO); contains Interpro domain(s) IPR012677 Nucleotide-binding, alpha-beta plait
Solyc01g102990	257,56	397,01	ATP-dependent Clp protease ATP-binding subunit clpX (AHRD V1 ***-B6SSC5_MAIZE); contains Interpro domain(s) IPR004487 ClpX, ATPase regulatory subunit
Solyc01g103040	226,35	307,66	Mitotic checkpoint protein MAD1 (AHRD V1 *-*- A8JF64_CHLRE); contains Interpro domain(s) IPR008672 Mitotic checkpoint
Solyc01g103110	269,85	470,10	Unknown Protein (AHRD V1); contains Interpro domain(s) IPR007434 Protein of unknown function DUF482
Solyc01g103120	1,55	25,46	Dynamin-2A (AHRD V1 ***-B6UEQ3_MAIZE); contains Interpro domain(s) IPR011993 Pleckstrin homology-type
Solyc01g103130	2,79	24,37	Dynamin-2A (AHRD V1 ***-B6UEQ3_MAIZE); contains Interpro

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
			domain(s) IPR001401 Dynamin, GTPase region
Solyc01g103170	164,19	251,51	Lysine ketoglutarate reductase trans-splicing related 1-like (AHRD V1 ***- Q5JLN0_ORYSJ); contains Interpro domain(s) IPR007877 Protein of unknown function DUF707
Solyc01g103230	517,74	1.065,73	Non-lysosomal glucosylceramidase (AHRD V1 ***- B9SS77_RICCO); contains Interpro domain(s) IPR014551 Beta-glucosidase, GBA2 type
Solyc01g103290	2,85	22,73	Unknown Protein (AHRD V1)
Solyc01g103300	0,97	28,14	SWIB complex BAF60b domain-containing protein (AHRD V1 *-*- D7M1C3_ARALY)
Solyc01g103310	1,89	44,27	Unknown Protein (AHRD V1)
Solyc01g103320	61,02	278,51	Zinc finger CCCH domain-containing protein 44 (AHRD V1 *--- C3H44_ARATH); contains Interpro domain(s) IPR018144 Plus-3 domain, subgroup
Solyc01g103330	6,99	59,45	Unknown Protein (AHRD V1)
Solyc01g103340	90,18	304,97	Zinc finger CCCH domain-containing protein 44 (AHRD V1 **-- C3H44_ARATH); contains Interpro domain(s) IPR018144 Plus-3 domain, subgroup
Solyc01g103400	183,14	285,13	Unknown Protein (AHRD V1)
Solyc01g103470	10,52	24,48	Unknown Protein (AHRD V1)
Solyc01g103490	351,76	437,35	Necrotic spotted lesions 1 (Fragment) (AHRD V1 *-*- B6ZA21_HELAN); contains Interpro domain(s) IPR001862 Membrane attack complex component/perforin/complement C9

Gene	Expression average		Annotation
	S. Lycopersicum	*S. habrochaites*	
Solyc01g103510	509,21	683,95	Ribosomal protein L3-like (AHRD V1 ***- Q2VCJ2_SOLTU); contains Interpro domain(s) IPR000597 Ribosomal protein L3
Solyc01g103560	81,22	105,48	tRNA pseudouridine synthase A (AHRD V1 ***- D6ZZF1_STAND); contains Interpro domain(s) IPR001406 Pseudouridine synthase I, TruA