



HAL
open science

Natural variation in wild tomato trichomes; selecting metabolites that contribute to insect resistance using a random forest approach

Ruy W J Kortbeek, Marc D Galland, Aleksandra Muras, Frans M van der Kloet, Bart André, Maurice Heilijgers, Sacha a F T van Hijum, Michel A Haring, Robert C Schuurink, Petra M Bleeker

► To cite this version:

Ruy W J Kortbeek, Marc D Galland, Aleksandra Muras, Frans M van der Kloet, Bart André, et al.. Natural variation in wild tomato trichomes; selecting metabolites that contribute to insect resistance using a random forest approach. *BMC Plant Biology*, 2021, 21, 10.1186/s12870-021-03070-x . hal-04229877

HAL Id: hal-04229877

<https://hal.inrae.fr/hal-04229877>

Submitted on 5 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH

Open Access



Natural variation in wild tomato trichomes; selecting metabolites that contribute to insect resistance using a random forest approach

Ruy W. J. Kortbeek^{1†}, Marc D. Galland^{1†}, Aleksandra Muras¹, Frans M. van der Kloet², Bart André³, Maurice Heilijgers¹, Sacha A. F. T. van Hijum⁴, Michel A. Haring¹, Robert C. Schuurink¹ and Petra M. Bleeker^{1*}

Abstract

Background: Plant-produced specialised metabolites are a powerful part of a plant's first line of defence against herbivorous insects, bacteria and fungi. Wild ancestors of present-day cultivated tomato produce a plethora of acylsugars in their type-I/IV trichomes and volatiles in their type-VI trichomes that have a potential role in plant resistance against insects. However, metabolic profiles are often complex mixtures making identification of the functionally interesting metabolites challenging. Here, we aimed to identify specialised metabolites from a wide range of wild tomato genotypes that could explain resistance to vector insects whitefly (*Bemisia tabaci*) and Western flower thrips (*Frankliniella occidentalis*). We evaluated plant resistance, determined trichome density and obtained metabolite profiles of the glandular trichomes by LC-MS (acylsugars) and GC-MS (volatiles). Using a customised Random Forest learning algorithm, we determined the contribution of specific specialised metabolites to the resistance phenotypes observed.

Results: The selected wild tomato accessions showed different levels of resistance to both whiteflies and thrips. Accessions resistant to one insect can be susceptible to another. Glandular trichome density is not necessarily a good predictor for plant resistance although the density of type-I/IV trichomes, related to the production of acylsugars, appears to correlate with whitefly resistance. For type VI-trichomes, however, it seems resistance is determined by the specific content of the glands. There is a strong qualitative and quantitative variation in the metabolite profiles between different accessions, even when they are from the same species. Out of 76 acylsugars found, the random forest algorithm linked two acylsugars (S3:15 and S3:21) to whitefly resistance, but none to thrips resistance. Out of 86 volatiles detected, the sesquiterpene α -humulene was linked to whitefly susceptible accessions instead. The algorithm did not link any specific metabolite to resistance against thrips, but monoterpenes α -phellandrene, α -terpinene and β -phellandrene/D-limonene were significantly associated with susceptible tomato accessions.

* Correspondence: P.M.Bleeker@uva.nl

Ruy W. J. Kortbeek and Marc Galland share first authorship

¹Green Life Science Research Cluster, Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands
Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions: Whiteflies and thrips are distinctly targeted by certain specialised metabolites found in wild tomatoes. The machine learning approach presented helped to identify features with efficacy toward the insect species studied. These acylsugar metabolites can be targets for breeding efforts towards the selection of insect-resistant cultivars.

Keywords: Tomato, Trichomes, Specialised metabolites, Insect resistance, Whitefly, Thrips, Random forest, Volatiles, Acylsugars

Background

Pests and diseases are a threat for food security with yearly losses of around 26% in primary yield, and even higher secondary yield losses up to 38% [1]. Herbivorous insects can damage plants directly through feeding, resulting in decreased photosynthetic capacity and in visible damage rendering crop plants unmarketable. However, they can also be vectors of diseases, typically transferring viruses from plant to plant, providing the virus mobility that can dramatically lower production. Insects like whiteflies (*Bemisia tabaci*; previously Sweet Potato- or Silverleaf Whitefly) and western flower thrips (*Frankliniella occidentalis*) are important, invasive herbivores found in agricultural production areas all over the world. Both insect species are regarded “supervectors” that can carry and transmit a multitude of viruses to a large range of host plants by feeding on the phloem (whitefly) or cell content (thrips) [2–5]. For both insects, plant-to-plant virus transmission predominantly takes place during their (mobile) adult life phase [6, 7]. For thrips, however, the virus is acquired exclusively in larval stage prior to developing into an infectious adult vector [8, 9]. Management of these pest vectors in agriculture is still largely based on chemical control and systemic insecticides that can provide a strong and lasting crop protection. However, with an increasing number of classical synthetic insecticides banned due to their negative off-target effects, understanding and deploying the insect resistance found in (wild) ancestral crop relatives is again gaining interest [10–13].

Plant resistance against (herbivorous) insects can be defined as a heritable ability to decrease insect attack in comparison to other plants without this ability thereby lowering inflicted damage [14, 15]. In nature, plants reduce or prevent insect colonisation through defence mechanisms such as anti-settling or anti-feeding properties, toxicity to the insect or interference in its development [16–18]. They range from physical barriers, surface coverage with sticky substances or the production of enzymes and specialised metabolites. Whereas non-glandular hairs physically impede the movement of certain herbivores on the plant surface, glandular tissues such as ducts, peltate cells or, in the case for tomato, glandular trichomes, produce and store specialised metabolites that act in defence against insects [19, 20].

Metabolites stored inside such glandular tissues are released via exudation or by disruption of the tissue through physical damage by the insect (e.g. chewing, penetration or touch). The exact mode of action of the phytochemicals is generally not fully elucidated. Besides physical impediments like stickiness or clogging of the insect’s mouthparts, specialised metabolites can act on the molecular level by receptor interference, disturbance of biological (e.g. enzymatic) processes, hormone balance or affect the insect’s nervous system by interfering in electron transport [21].

Tomato produces a plethora of specialised metabolites in its glandular trichomes which are mainly located on the green parts of the plant. Especially certain wild relatives of cultivated tomato (*Solanum lycopersicum*) display considerable resistance to agronomically important pest insects, which often can be traced back to trichome-borne specialised-metabolites [22–27]. Wild species like *S. pennellii*, *S. habrochaites* and *S. galapagense*, for example, exude a broad range of acylsugars from type I- and IV-glandular trichomes providing resistance against several pest insects [28–30]. These acylsugars consist of glucose- or sucrose moieties with acyl groups esterified to one- or more of its hydroxyl groups. The composition of acylsugars exudates can be highly diverse, as the complete blend is an assembly of acylsugars having different sugar moieties and varying acyl chain length- and positions [31–35]. Interestingly, insects appear to respond differently to particular acylsugar compositions, indicating that insect targeting is component specific [36, 37].

The (semi-)volatile mono- and sesquiterpenes that are, at least predominantly, produced and stored in type-VI trichomes serve as another important class of metabolites implicated in the defence against pests [24, 38–40]. The sesquiterpenes 7-epizingiberene and R-curcumene produced by wild tomato *S. habrochaites* PI127826 repel whiteflies and 7-epizingiberene alcohol derivatives have a repellent effect on the spider mite *Tetranychus urticae* [41, 42]. Also, 2,3-dihydrofarnesoic acid from *S. habrochaites* accessions LA1363 and LA1927 act on spider mites as a repellent [43] and type-VI trichomes of *S. habrochaites* accession LA1777 produce a mixture of bergamotonic- and santalonic acids that confers resistance against the tomato fruitworm (*Helicoverpa zea*) and

the beet armyworm (*Spodoptera exigua*) [44]. Type-VI trichomes store volatiles as concentrated oily substances in their internal storage-cavity, of which it is hypothesized that only a limited amount is released by (passive) diffusion [45]. Upon rupture, e.g. by touch, the insect receives a high dose of these essential oils which can be toxic by itself or creates toxic fumes [46, 47]. The biochemical (precursor-) pathways to produce these metabolites are rather conserved among tomato species, including cultivated tomato. As the leaves and stem surface of cultivated tomato also have type-VI and, although only in an early life stage, type-IV trichomes, the transfer of trichome-borne specialised metabolic-pathways from wild tomato to cultivars is feasible [19, 24, 48].

Although it is clear that metabolite-based insect-resistance traits are available from natural resources, it is still a challenge to identify those metabolites that target the pest insect of interest. The complexity in the blend of metabolites produced makes it difficult to identify the functional compound(s) causal to an observed phenotype (e.g. insect mortality or arrested development). Machine learning methodologies, including Random Forest (RF) classification, have emerged as powerful and accurate alternatives to the often-used dimension-reduction methods (e.g. Partial Least Squares regression) to determine variables discriminating (phenotypic) classes [49, 50]. Examples of successful RF applications in plant biology include the selection of metabolic markers for drought tolerance in potato [51] and the identification of DNA-methylation patterns in maize determining gene transcription-and translation [52]. The ability of RF to cope with large amounts of “omics” data derived from a relatively small number of samples makes RF also a promising tool to identify anti-insect metabolites from natural resources.

Here, we explore the natural variation in resistance to insects (i.e. whitefly and thrips) and production of specialised metabolites by glandular trichomes within a panel of wild tomato accessions. To this end we successfully quantified insect survival, density of different glandular trichome-types, and production of acylsugars and trichome-borne volatiles in the accessions. Next, we used a customised RF algorithm to predict which acylsugars and volatile molecules correlated with survival phenotypes of the insects. This resulted in a list of metabolites predicted to be involved in insect-specific resistance or susceptibility. Finally, the biological outcome, as well as the use of RF classification models on this type of metabolite-phenotype data, will be discussed.

Results

Differential insect responses on tomato accessions

Survival of whitefly was assessed on different tomato accessions in a no-choice set up. Survival ranged from zero

on *S. pennellii* accession LA0716, to 88% on *S. peruvianum* LA1278 (Fig. 1a). In addition, also the performance of thrips (*F. occidentalis*) on the same set of wild tomato was determined in a different survival assay using thrips larvae; the life stage in which thrips acquires viruses. In this experimental setup, the larval survival on each accession was recorded for 15 days on leaf discs whereafter the median survival time (i.e. when 50% of the thrips larvae deceased) was estimated using the Kaplan-Meier method giving the most conclusive proxy for the phenotype (Additional file: Figure S1). This resulted in a different ranking of the accessions than for whitefly survival, with median survival ranging from 1.5 days on *S. habrochaites* LYC4 to 14 days on *S. huaylense* LA1364 (Fig. 1b). These differences were further visualised when the survival parameters were plotted against each other (Fig. 1c). It showed that although some accessions were indeed a toxic environment to both insects (e.g. *S. habrochaites* LA1777), this was not the case for all. For example, whiteflies had a relative high survival rate on LA0407 and LA1278, whereas thrips did not. PI127826, on the other hand, was resistant to whiteflies but marginally more susceptible to thrips. As expected, the *S. lycopersicum* accessions Moneymaker (MM) and LA4024 benchmark susceptibility to both insects.

Trichome phenotype of the selected genotypes

Following the hypothesis that the antixenosis phenotype is facilitated by trichome-produced specialised metabolites, we first profiled the trichome density for each accession (illustrated in Fig. 2a, for all accessions see Additional file: Figure S2). Leaf-trichome density was determined for different trichome types: non-glandular (NG; including trichome-type II/III/V/VIII), glandular type I/IV and glandular type VI (Additional file: Figure S3). Comparing the trichome densities on the abaxial- and adaxial leaf surfaces shows that especially type I/IV trichomes prevail on the abaxial leaf surface. Type VI trichomes appear to be more or less equally distributed across adaxial- and abaxial leaf surfaces, with the exceptions of LA2695, LA1578 and LA1718 that have more type-VI trichomes on the adaxial-leaf surface. *S. pimpinellifolium* LA1578 is the hairiest accession in the panel with the majority of trichomes being NG.

To visualise the relationship between glandular-trichome density and insect response, we plotted the type-I/IV and type-VI trichome density of each accession against the insect survival rates (Fig. 2b, c). A higher density of type I/IV trichomes, typically found on *S. pennellii* LA0716 and *S. habrochaites* LA1777, LYC4, PI127826 and PI134418 accessions (Additional file: Figure S3), appeared to correlate to lower whitefly survival (Fig. 2b, top panel; $r^2 = 0.897$, $p > 0.001$). In contrast, we did not find a significant linear relationship

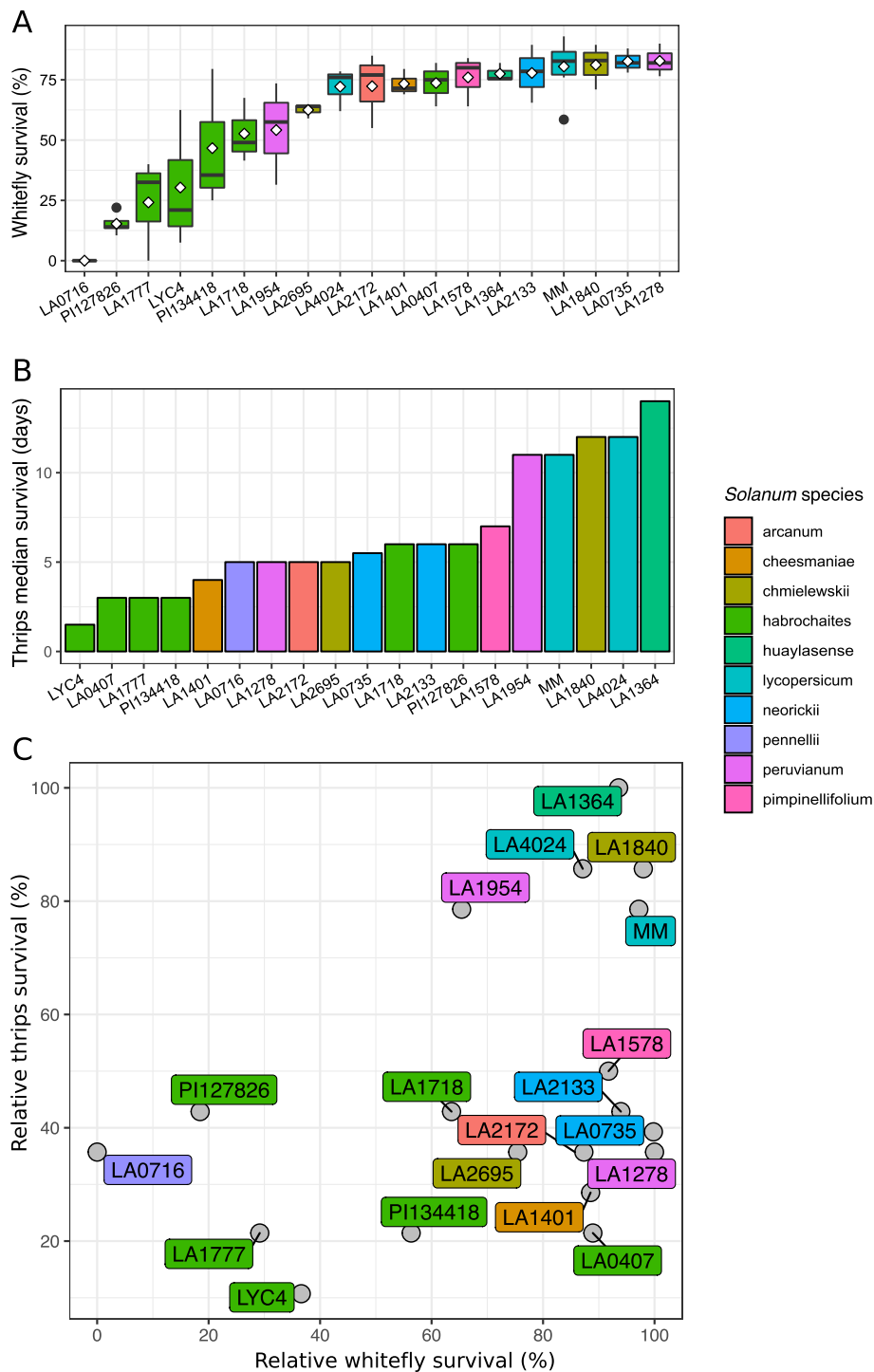


Fig. 1 Insect response to selected *Solanum* accessions **(a)** Adult whitefly (*B. tabaci*) survival after 5 days in clip-cages attached to leaflets of the fourth fully expanded leaf from the top of six-weeks-old plants ($n = 3-8$). Boxplots indicate mean (diamonds) and median (black horizontal bars) survival rates. Accessions are ordered by ascending mean survival rates. **(b)** Median survival time (days) of thrips L1 larvae (*F. occidentalis*) placed on a leaf disc ($n = 24-36$) for a maximum of 19 days. Leaf discs were made from the fourth fully expanded leaf from the top of six-week-old plants. Accessions are ordered by ascending median survival. **(c)** Relative survival scores of whiteflies and thrips plotted against one another. Relative survival score was calculated by setting the highest mean (whitefly) or median (thrips) survival rate to 100%. Accessions are colour-coded by their *Solanum* species

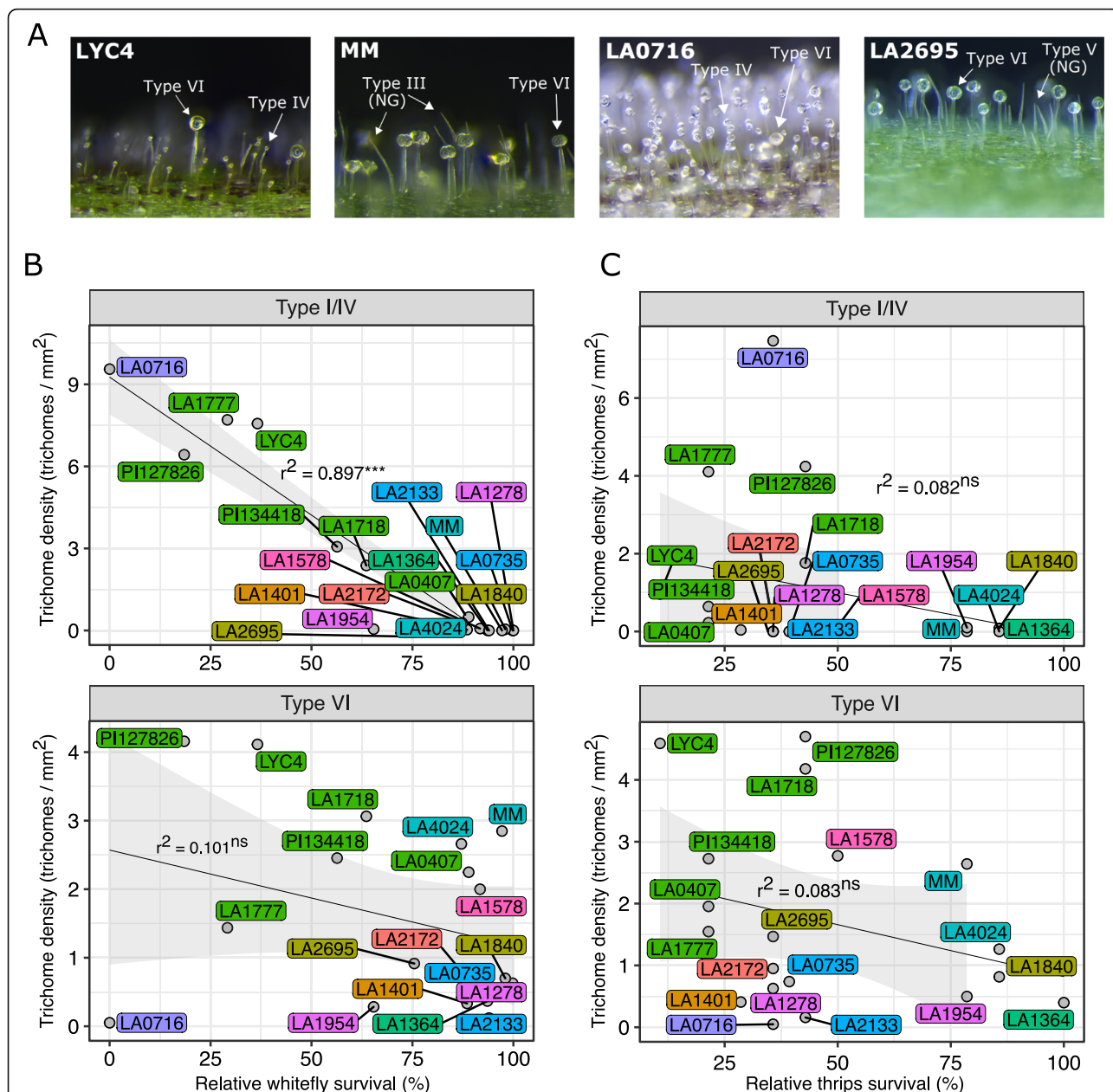


Fig. 2 Insect survival plotted against glandular-trichome densities on leaves of selected accessions. **a** Photographs of *S. habrochaites* LYC4, *S. lycopersicum* cv. MoneyMaker (MM), *S. pennellii* LA0716 and *S. chmielewskii* LA2695 to visualise the diversity in trichome-types. Images of all accessions can be found in Additional file: Figure S2. **b** Mean glandular trichome density per mm² leaf surface of type I/IV trichomes (top panel) and type VI trichomes (lower panel) against relative whitefly survival rates on the different accessions and **c** against thrips (*F. occidentalis*) relative survival. The black regression lines show the linear relationship, with the 95% confidence interval in light grey, between the trichome densities and the survival phenotypes. The r^2 is given together with asterisks for the significance of the relationships with p -values < 0.05*, < 0.01**, < 0.001***, ns: non-significant. Leaf material analysed originated from the fourth fully expanded leaf from the top of six-week-old plants

between type-I/IV density and the survival of the thrips larvae (Fig. 2c top panel; $r^2 = 0.082$, $p = 0.044$). Also, the linear regression models did not indicate a linear relationship between type-VI trichome density and the survival rates of either whiteflies or thrips (Fig. 2b, c, lower panels; $p > 0.05$). For example, with approximately 300 type-VI trichomes per cm², *S. lycopersicum* cultivars

MM and LA4024 were amongst the most densely type-VI haired accessions (Additional file: Figure S3) but displayed high insect survival rates. Additional analysis on NG trichomes, and all trichome types combined, also did not yield a significant relationship between densities and the survival rates of either of the insects (Additional file: Figure S3).

Specialised metabolite composition

Although glandular trichomes can contain other specialised metabolites, we focussed here on the most abundant molecules in type I/IV and type VI glandular trichomes: i.e. acylsugars and volatile molecules respectively. Acylsugars were obtained by leaf washes and analysed by liquid-chromatography mass-spectrometry (UHPLC-MS). Untargeted analysis resulted in a list of 76 acylsugar moieties across the 19 tomato accessions that were annotated according to their sucrose (S) or glucose (G) backbone, followed by the number of acyl chains plus the total number of carbon atoms distributed over the acyl chains. Reflecting the type-I/IV trichome density, accessions most resistant to whiteflies contain high amounts of acylsugars, but this was not necessarily the case for the accessions resistant to thrips (Additional file: Figure S4a-c). Although the majority of the detected acylsugars consist of acylsucroses, there was no indication that the type of backbone, i.e. sucrose or glucose, correlated to the levels of resistance. However, acylsugar exudates of whitefly resistant accessions consisted mainly of tri-acylsugars (i.e. G3 and S3) whereas the more susceptible accessions, albeit in low levels, contained mostly tetra- and penta-acyl sugars (Additional File: Figure S4d). The total number of carbon atoms distributed over the acyl chains predominantly ranged from 14 to 21 in resistant accessions, and from 21 to 25 in susceptible plants (Additional File: Figure S4e).

To visualise individual acylsugars, we created a heatmap and clustered the acylsugars based on their abundance in different accessions ranked by whitefly survival (Fig. 3a). From the appearing clusters, cluster 1 covers a selection of mainly acylglucoses that occurred only in a few resistant accessions. In this cluster, *S. pennellii* accession LA0716 contained a unique subset of acylglucoses and sucroses, while two *S. habrochaites* accessions contained another unique subset also comprising mixed set of sugar backbones. Cluster 2 compounds, mainly tri-acylsucroses (S3), were overrepresented in resistant *S. habrochaites* accessions. In susceptible accessions, acylsugars from cluster 3 were predominantly present, which mainly consisted of tetra-acylsucroses (S4) and tetra-acylglucoses (G4).

Besides acylsugars, trichome-derived volatiles were obtained by leaf washes and analysed by gas-chromatography mass-spectrometry (GC/MS-TOF). We detected 86 different volatiles in total and deduced their identity by comparison of their mass spectrum combined with their Kovats Retention Index (KI) to available libraries (Additional file: Table S3). The accessions displayed a diverse set of volatile profiles in which methyl ketones, monoterpenes, sesquiterpenes and other aromatic hydrocarbons were most abundant (Additional file: Fig. S5a). We did not find one of these structural

classes to be predominantly present in the volatile profile of whitefly-resistant accessions, although the profile of susceptible accessions encompassed mainly monoterpenes (Additional File: Fig. S5a). Summing all ion-abundances per sample illustrates that neither whitefly nor thrips survival rates could be explained by the total level of volatiles detected in each accession (Additional File: Fig. S5b, c).

Plotting the individual metabolites resulted in a heatmap with dispersed abundances, visualising the rather distinct volatile profiles between accessions (Fig. 3b). Accessions with high whitefly survival rates appeared to contain volatile metabolites annotated in cluster 1 and 5, i.e. predominantly mono- and sesquiterpenes, although these compounds were also found in accessions with low whitefly survival. The three most resistant accessions all had a unique set of metabolites. *S. pennellii* LA0716 for instance produced a distinct set of volatile compounds, annotated in cluster 3, constituting of cyclic- and acyclic (oxygenated) hydrocarbons whereas *S. habrochaites* PI127826 and LA1777 each display a distinct set of sesquiterpenes, annotated in cluster 2 and 4 respectively. Other *S. habrochaites* accessions, LYC4, PI134418 and LA1718, produced methyl ketones as indicated in cluster 2.

Linking specialised metabolites to whitefly resistance

To predict which specific, individual metabolites could be linked to the observed resistance phenotypes of the accessions, a Random Forest machine learning approach was applied. The RF model requires a binary classification of the accessions as an input variable and we therefore first fitted a generalised linear model (GLM) on the whitefly and thrips survival data creating two phenotypic classifications per insect proxy (Methods; Additional file: Table S1). The group of accessions with the lowest survival rates were hence labelled “resistant” while the remaining accessions were labelled “susceptible” (Table 1). As illustrated in Fig. 4, the RF algorithm was used to compute the importance of each metabolite (“metabolite feature-importance”) to reconstruct the phenotypic classification of the accessions as predefined by the GLM (see also: Methods). We first implemented the model on the acylsugars and the whitefly resistance classification of the accessions. The model exhibited a total accuracy score of 79%. We opted to apply a stringent p -value ($p < 0.01$) to the model resulting in two specific acylsugars, S3:15 and S3:21 (Table 2), that were identified as metabolites with significantly higher feature importances compared to the randomly permuted model (Additional file: Figure S6a, b), indicating their importance to separate the insect-phenotypic classes. Plotting the abundances of these two compounds indicates their predominantly presence in accessions

(See figure on previous page.)

Fig. 3 Heatmaps showing the detected metabolites in the different accessions. Accessions are ordered according to the mean whitefly survival rates and metabolites are clustered by complete-linkage clustering based on their abundance in the accessions. **a** Acylsugars detected by UHPLC-MS ($n = 6$) labelled using the following nomenclature: Sugar moiety backbone (G for glucose; S for sucrose) followed by the numbers of esterified acyl groups and the number of carbon atoms distributed over the acyl groups. In case of structural isomers, a hyphen followed by the isomer number is added. The panel on top of the heatmap indicates the sugar-moiety constituting the acylsugar backbone. **b** Volatiles detected by GC/MS-TOF ($n = 3-4$). Volatiles are labelled according to mass spectral matches to available libraries. The panel on top indicates the structural classification of the respective metabolite. Leaf material analysed originated from the fourth fully expanded leaf from the apex of six-week-old plants

classified as whitefly resistant. Acylsugar S3:15 occurred in *S. habrochaites* accessions PI127826, LYC4, PI134418, LA1718 while acylsugar S3:21 was also present in *S. pennellii* LA0716 (Fig. 5a, b). One susceptible *S. habrochaites* accession, LA0407, however also appeared to also produce both S3:15 and S3:21.

The RF model next linking volatiles to whitefly resistance classifications had an accuracy score of 63% and indicated only one out of the 86 volatiles as a feature of significant importance ($p < 0.01$; Additional file: Figure S6c). The significant sesquiterpene, α -humulene (Table 2) was prevalent in whitefly-susceptible accessions (Fig. 5c).

Linking specialised metabolites to thrips resistance

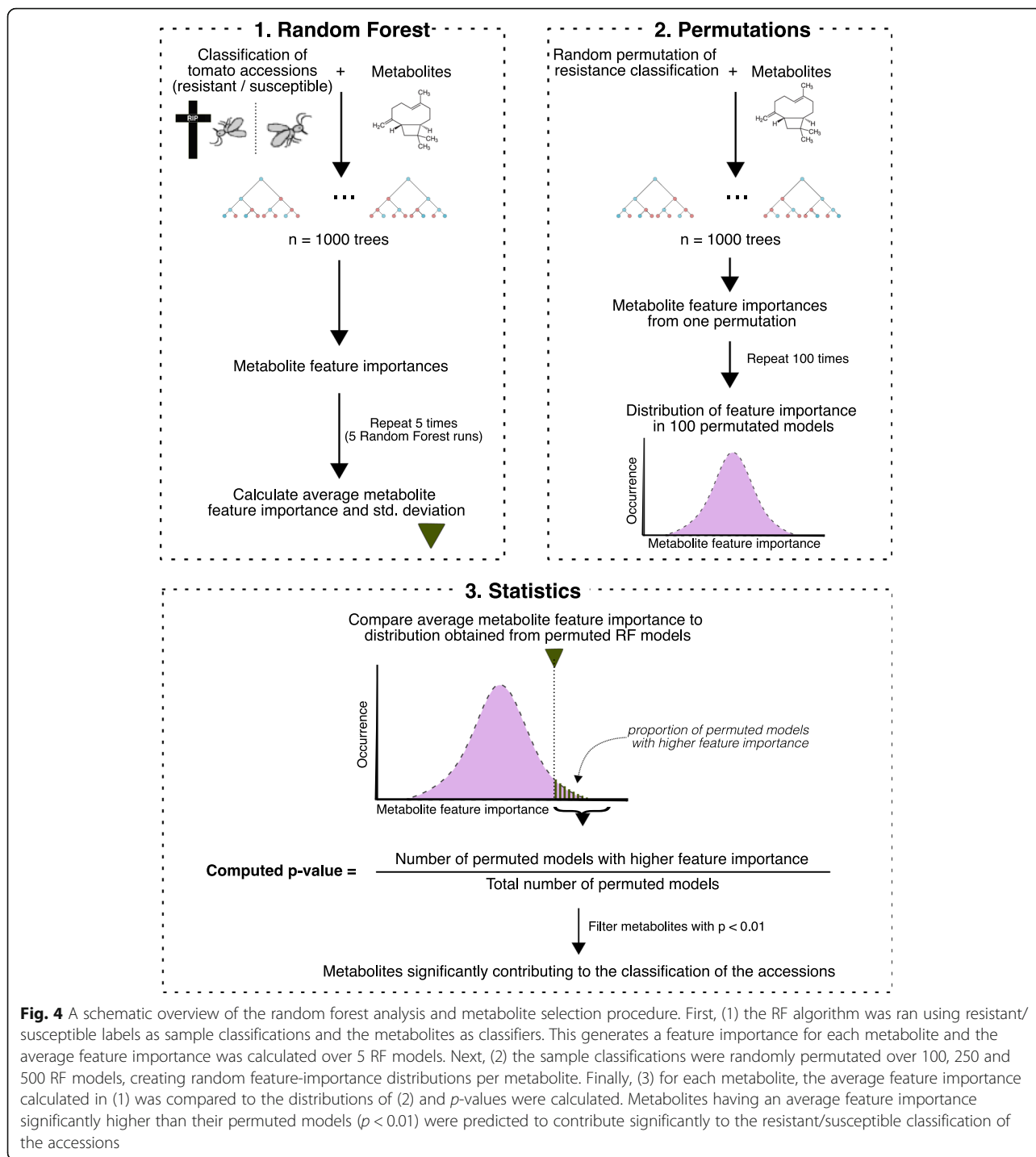
As noted earlier, thrips survival on the different genotypes differed from that of whitefly. Therefore, the

accessions were again classified into two distinct groups, i.e. resistant-or susceptible to thrips (Table 1) after which the importance of individual features was computed. The thrips-acylsugar RF model exhibited a relatively low accuracy score of 58% and did not predict any acylsugar with a feature importance that significantly deviated from the randomly permuted models. The model based on volatile data exhibited a better accuracy score of 85% and resulted in three significant volatile peaks, all monoterpenes, including α -phellandrene, α -terpinene and β -phellandrene/D-limonene (Table 2; Additional file: Figure S7 and Additional file: Figure S8). Plotting their abundances showed that the monoterpenes α -phellandrene and α -terpinene were both predominantly present in thrips susceptible plants (Fig. 5d-e). The other volatile peak, β -phellandrene/D-limonene also occurred in some of the thrips resistant accessions (Fig. 5f).

Table 1 List of 19 selected *Solanum* accessions and their classification

Accession	Species	Whitefly classification	Thrips classification
LA2172	<i>S. arcanum</i>	susceptible	resistant
LA1401 ^a	<i>S. cheesmaniae</i>	susceptible	resistant
LA1840	<i>S. chmielewskii</i>	susceptible	susceptible
LA2695	<i>S. chmielewskii</i>	resistant	resistant
LA0407	<i>S. habrochaites</i>	susceptible	resistant
LA1777	<i>S. habrochaites</i>	resistant	resistant
PI134418	<i>S. habrochaites</i>	resistant	resistant
LYC4	<i>S. habrochaites</i>	resistant	resistant
LA1718	<i>S. habrochaites</i>	resistant	susceptible
PI127826	<i>S. habrochaites</i>	resistant	susceptible
LA1364	<i>S. huaylasense</i>	susceptible	susceptible
MoneyMaker	<i>S. lycopersicum</i>	susceptible	susceptible
LA4024	<i>S. lycopersicum</i>	susceptible	susceptible
LA2133	<i>S. neorickii</i>	susceptible	susceptible
LA0735	<i>S. neorickii</i>	susceptible	susceptible
LA0716	<i>S. pennellii</i>	resistant	resistant
LA1278	<i>S. peruvianum</i>	susceptible	resistant
LA1954	<i>S. peruvianum</i>	resistant	susceptible
LA1578	<i>S. pimpinellifolium</i>	susceptible	susceptible

Regression analysis on the no-choice survival data with generalised linear model for whitefly data; $p < 0.05$ and Cox proportional hazards coefficients for thrips; $p = 0.01$, divides the tomato accessions in either "resistant" or "susceptible" environments for the insects. ^a LA1401 morphotype *S. cheesmaniae* [25, 53].



Discussion

Wild tomatoes can produce specific metabolic compounds that improve resistance to a variety of insects [24, 25, 54–57]. Plant resistance is generally a broadly used term, first delineated by Painter in 1951 [14] and could be described as “a heritable trait that minimises the damage experienced by the plant” [15]. Insect response to a plant can be recorded in numerous ways

and the choice of measure can greatly impact the conclusions drawn [15]. Here we investigated a short-term antibiosis response for whitefly and thrips providing a measure of direct constitutive defence. The level and specificity of such resistance varies greatly among the selected accessions (Fig. 1). Our work corroborates that plant resistance properties against herbivorous insects cannot be generalised, but rather are defined by a

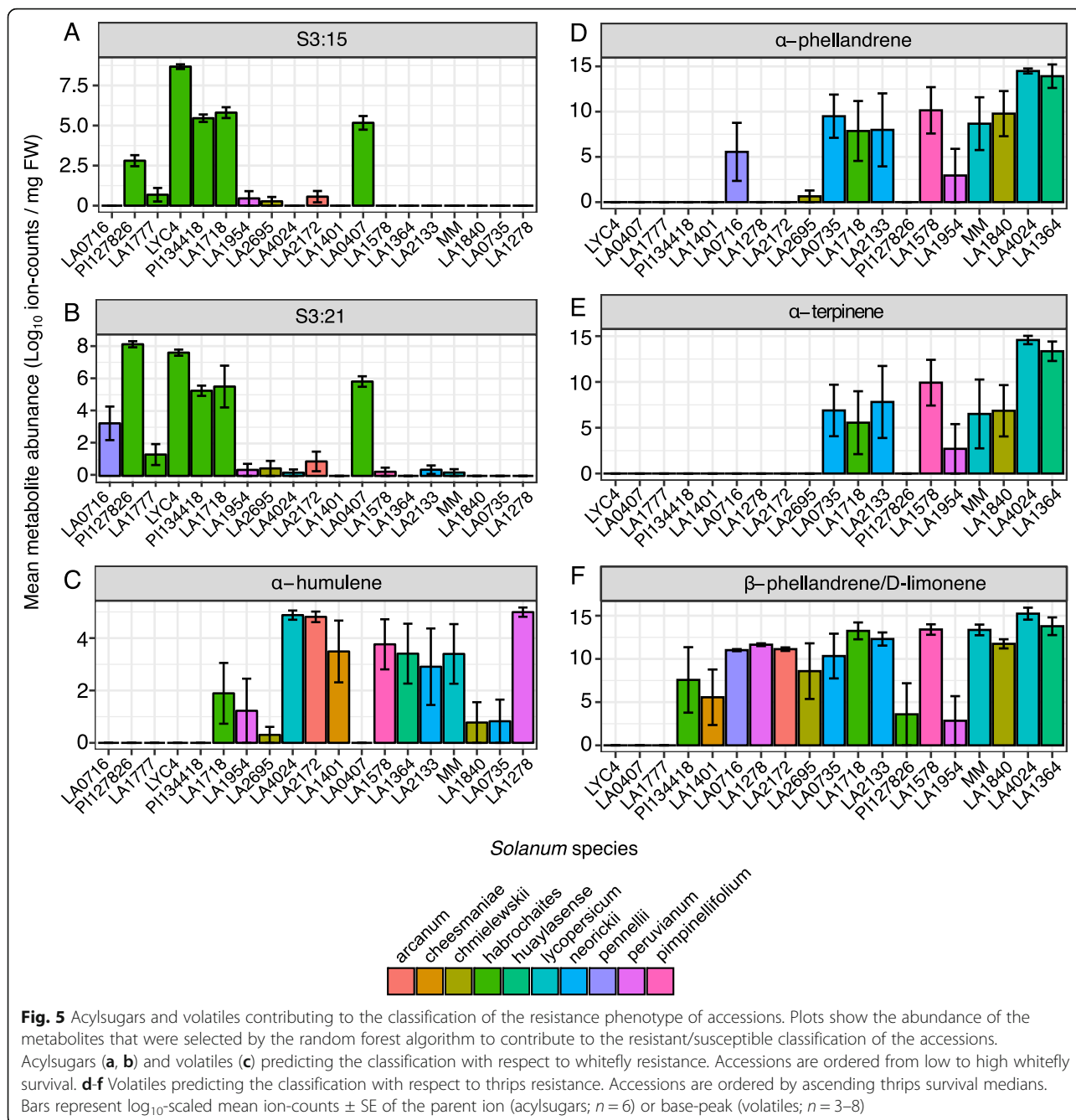
Table 2 Significant metabolites according to the random forest models

Metabolite	m/z	Experimental KI	Theoretical KI	Insect	Classification	P-value
S3:15	617 [M + Na] ⁺	–	–	whitefly	resistance	$p < 0.001$
S3:21	715 [M + Na] ⁺	–	–	whitefly	resistance	$p < 0.001$
α -humulene	204.2	1457	1454	whitefly	susceptibility	$p < 0.001$
α -phellandrene ^a	136.1	1000	1002	thrips	susceptibility	$p < 0.001$
α -terpinene ^a	136.1	1013	1017	thrips	susceptibility	$p < 0.001$
β -phellandrene/D-limonene	136.1	1027	1029	thrips	susceptibility	$p < 0.001$

Acylsugars are annotated by their sucrose (S) or glucose (G) backbone followed by the number of acyl chains and the number of carbon atoms distributed over the acyl chains. Volatiles were annotated using the Kovats Retention Index (KI) and fragmentation-pattern comparison to the Adams [46] and NIST libraries. The table provides the resistance/susceptibility classification for the different insects by the RF model and the p -value indicates whether the calculated feature importance significantly deviates from a randomly permuted model (threshold: $p < 0.05$). ^a Metabolite annotation confirmed by an analytical standard. The β -phellandrene/D-limonene peak could not be separated due to their strong co-elution. Mass spectra of metabolites not verified by an analytical standard are given in Additional file: Figure S8.

specific (biochemical) environment in combination with the insect [25, 29, 56]. Indeed, this does not only hold true for the insect species but also for their life stages and it depends on the parameters chosen e.g. development, oviposition, survival, repellence [15]. Considering the resistance levels observed here, it suggests that tomato accessions harbour particular defence metabolites targeting adult whiteflies and thrips larvae. Although the mode of action of specialised metabolites on insects is not always fully understood, effects on cellular levels including respiration, membrane integrity, cellular transport and receptors have been described [21, 58, 59]. By choosing for survival assays, alternative defence mechanisms acting through repellence or distorting development were not included here. Hence wild-tomato metabolites we identified earlier as repellents, i.e. the volatile terpenes 7-*epi*-zingiberene and *R*-curcumene from PI127826, would not be discerned here to be associated to survival as they turned out not to be toxic [24, 60]. In addition, there are defence metabolites that are not produced by trichomes as well as enzymes and cuticle structures that could be involved in insect survival [61–63]. The choice for trichome-produced defence metabolites is however interesting in light of breeding for plant-produced specialised metabolites, as an alternative for the application of synthetic pesticides. Tomato cultivars can develop different types of trichomes, including type IV and VI on the green tissues, and while trichomes can store high levels of (auto)toxic metabolites, they are barely present on the mature tomato fruit [64]. The introgression of trichome-specific traits from wild species to cultivars, especially a combination of repellent and toxic metabolites, might therefore be a feasible option to improve resistance. The complexity of metabolite mixtures found trichomes often makes it hard to determine the effective individuals and we therefore aimed for a strategy that included metabolite types (i.e. acylsugars and volatiles) that have been implicated in resistance before.

The general relationship between (glandular) trichome density and the impact on herbivores is well-described in tomato [47, 65–67]. This is not always a negative relation though. In fact, whiteflies prefer to oviposit on hirsute leaves as - especially non-glandular - trichomes offer protection from natural enemies, such as parasitoids, to the nymphs [68–70]. Here we found indications that the impact of trichomes on insects appears attributable to its chemical content and that glandular trichome density is not necessarily a reliable indicator of resistance (Fig. 2). Type-VI trichome density did not correlate with the survival rates of either whitefly or thrips and the results imply that for this set of tomatoes type-VI trichome-related resistance depends more on the nature of metabolites than on the dose. Indeed, insects can be resilient to plants producing a relatively high quantity of terpenoids (Additional file: Figure S5). However, we did observe a clear negative correlation between type-I/IV trichome density and adult whitefly survival (Fig. 2) in line with previous findings [28, 71, 72], indicating there is probably a dose-dependent effect of acylsugars on whitefly mortality. The resistant tomato accessions in the panel do indeed exhibit much higher acylsugar levels (Additional file: Figure S4). Recently it was indicated that the effect of acylsugar exudates on insect resistance depends on the type of sugar backbone and the composition of the acyl chains [28, 29, 59]. Leckie et al. [36] concluded that acylsugar exudates enriched in acylsucroses and 3-methylbutanoic (iC5) fatty acids would be most effective in reducing whitefly oviposition. We also found the majority of the resistant accessions to accumulate mainly tri-acylsucroses (Additional file: Figure S4). The model did not significantly distinguish the resistant-and susceptible accessions based on the total amount of tri-acylsucroses or sucrose backbones (Additional file: Figure S4), but rather predicted S3:15 and S3:21 specifically to impair adult whitefly survival (Table 2). Both compounds are in cluster 2 of Fig. 3A, which comprises compounds predominantly produced in resistant



accessions. A study on greenhouse whitefly (*Trialeurodes vaporariorum*) resistance by Vosman et al. [56] showed that the renowned whitefly-resistant tomato accession *S. galapagense* LA1401 also contained high levels of acylsugars, in which several isoforms of S3:21 are abundant. Although we could not determine the exact acylsugar configuration here due to methodological constraints, we can deduce the composition of the acyl chains as these compounds were previously elucidated in tomato. Ghosh et al. (2014) reported that the S3:21 isomers in LA1777 all included an iC5, a 2-methylbutanoate (aiC5)

and a 9-methyldecanoate (iC11) chain, hence S3:21 (5,5, 11). S3:15 found in, amongst others, *S. habrochaites* LA1718 and LA0407 were determined to contain three 3-methylbutanoate groups hence S3:15 (5,5,5) [73]. Besides the validation of the individual compounds, it would be valuable to see if indeed the esterification of iC5 acyl chains is particularly responsible for the resistance effect as noted by Leckie et al. [36].

Survival of the thrips larvae on the other hand, did not appear to be linked to any of the acylsugars present in the tomatoes studied. Firstly, survival did not correlate

to the density of type-I/IV trichomes to the extent seen for whiteflies (Fig. 2c). Secondly, the RF model did not indicate any of the detected acylsugars present in this panel to be involved in separating thrips-resistant from thrips-susceptible accession. However, based on *S. pennellii* (LA0716) introgression material, Ben Mahmoud et al. [37] and Leckie et al. [36] did find indications that both acylsugar dose and the fatty-acid composition affected thrips oviposition rates. The protective effect of acylsugars against thrips could thus very well depend on the thrips' developmental stage. We used the thrips larvae in our bioassays, which are perhaps less likely to come in contact with the glandular head-cells, thereby receiving a lower dose of acylsugars compared to the bigger and much more mobile adults.

Although the evidence for the importance of specific acylsugars, or their structural features, in plant resistance starts to accumulate, there are also indications for synergistic effects of acylsugars mixtures. Hence, a specific combination of acylsugars can cause a stronger effect on the insect than the sum of individual compounds [36, 37]. Ideally next, validation experiments with individual and purified acylsugars S3:15 (5,5,5) and S3:21 (5,5,11) would be performed. However, such efforts remain challenging. Isolation, or synthesis, of single acylsugars is extremely difficult as they are complex molecules, with a high molecular similarity, generally present in varying mixtures in plants (Fig. 3a) [74]. Alternatively, creating a segregating population from one of the resistant accessions identified here, or making use of tomato introgression lines with different acylsugar profiles such as done by, amongst others, Leckie et al. [74], Smeda et al. [75] and Ben-Mahmoud et al. [37], could help reveal a causal link between a specific metabolite (combination) and the effect on different types of pests.

The sesquiterpene α -humulene was the only volatile predicted to distinguish whitefly resistant from whitefly-susceptible accessions. This compound is present in susceptible lines and was previously reported to potentially serve as a host attraction cue [76, 77], a response not tested here. The effect observed appears to be compound specific though, as the total volatile quantity (based on peak area) or structural metabolic group, did not discriminate resistant from susceptible accessions after RF modelling (Additional file: Figure S5). The model predicted three monoterpene peaks, α -phellandrene, α -terpinene and β -phellandrene/D-limonene to separate thrips susceptible from resistant accessions. These compounds were particularly prevalent in susceptible accessions and do not affect thrips survival. However, upon activation of the Jasmonate signalling-pathway in *S. lycopersicum* cultivars, the abundance of monoterpenes, including the ones described above, increases resulting in a decreased preference response of

adult thrips [78]. So, although our results suggest that these monoterpenes do not have a toxic effect on the thrips larvae, they could play a role in the repellence of the adults. Also, here short-term performance assays were done on naïve plants, not primed by earlier damage, focussing on constitutive and direct defence, but both thrips and whitefly infestations have been shown to induce specific chemical defence in accordance with their feeding style and can even manipulate host response in neighbouring plants [79, 80], highlighting the complexity of this dynamic trait.

While resistance against thrips is noticeably different between accessions (Fig. 1b; Additional file: Figure S2) this may involve trichome-independent metabolites such as (glyco)alkaloids, as was recently suggested for thrips feeding behaviour [81]. Alternatively, it might still be that volatile metabolites measured here, may be involved but were not selected because of sparsity in the dataset. This is supported by the observation that a relatively large portion of the metabolites were exclusively detected in a single accession (Fig. 3). The metabolic diversity in the set of tomatoes used here makes statistical analysis challenging, especially since the number of resistant accessions with overlapping metabolites is small, particularly in case of the volatiles. As noted above, it might mean that a uniquely appearing metabolite, despite having biological activity, was not selected only due to a lack of statistical power. For example, we recently identified the 7-*epi*-zingiberene derivative 9-hydroxy-10, 11-epozyzingiberene (9H10epoZ) to be toxic to whiteflies [60]. This compound is produced by, and specific to, *S. habrochaites* PI127826 (Fig. 3b), and was therefore probably not marked as a significant discriminant by the RF model. Another example is santalonic acid which, as mentioned before, is toxic to *H. zea* and *S. exigua* larvae [44] and potentially to other insects as well [82]. This metabolite was also not selected by our model, as it was exclusively detected in *S. habrochaites* accession LA1777. These examples indicate the limitations of a model when metabolic diversity is broad with too few overlapping accessions. The use of more closely related chemotypes (e.g. segregating populations or accessions of the same species) displaying contrasting resistance phenotypes would therefore be ideal for RF analyses.

Nevertheless, the RF method chosen here resulted in the selection of specialised metabolites from complex mixtures of acylsugars and volatiles in tomato glandular trichomes that could discriminate resistant from susceptible genotypes. Regression-based classification-methods like PLS-DA, widely applied in metabolomics and plant-insect interaction studies in particular [83–86], could not be applied here, as a consequence of the sparsity, or zero-valued elements, and the high feature to sample ratio [87, 88]. Even sparse PLS-DA [89] led to impaired

model convergence. The RF approach is regarded to be a powerful classification method, especially in case of sparse data [50, 90] and a low sample-to-feature ratio, common in multi-omics analyses [91]. In addition, the comparison of the RF results to a permuted dataset allowed statistical discrimination of the importance of individual metabolites. Despite its promise only a few examples of the use of RF for metabolite selection in plant science are available, including metabolite selection for potato-tuber discolouration [92], fingerprinting of primary metabolism in *Arabidopsis* mutants [93] and sex-related volatiles emitted by the cones of *Ficus* species [94]. Here we pose that RF can predict specific metabolites with potential in defence from a challenging dataset, making it a promising tool in future identification of specialised metabolites with potential biological significance from complex mixtures often found in plants.

Breeding for metabolite-based resistance is challenging and requires the introgression of several wild-species QTLs to obtain effective concentrations of herbivore-specific metabolites. Advances towards insect resistance in tomato have been made by creating pre-breeding material, introgression lines containing metabolite-QTLs that can be combined for insect-specific resistance, especially for acylsugars [28, 30, 36, 37, 74, 95]. The approach presented here can be complementary. The prediction of specific bioactive compounds that are causal to the desired phenotype enables a more targeted approach for elucidation of the genetic components underlying the metabolite composition, and a more efficient incorporation of insect resistance into breeding material [31, 96–98]. For agriculture, losses caused by herbivorous insects can be largely attributed to their vectoring of viruses. Therefore, treatment of the vector insects will impact also the pathogens they carry. The two candidate acylsugars we predicted to negatively affect whitefly survival are expected to additionally decrease the efficient spread of virus by this vector [30].

Conclusions

While there are many reports on the insecticidal properties of essential oils and plant extracts, there is only limited information about individual biologically active compounds [99]. Here we saw that diverse accessions that vary in metabolite profile can have different effects on different vector insects, and we put forward a set of specialised metabolites with the potential to act as natural insecticides. In the material analysed two specific acylsugars appear to have potential for chemical defence against whiteflies, however not for thrips larvae. Breeding for metabolite-based resistance poses a challenge that requires a clear definition of the desired trait and incorporation of multiple QTLs. By the use of RF we

were able to assign specific specialised metabolites from complex mixes to distinct phenotypic traits. The model can be improved further using accessions that display more overlap in metabolic profiles, such as in more closely related species or in segregating populations with distinct resistance profiles.

Methods

Plant and insect materials

The panel of tomato accessions was aimed to be diverse in specialised-metabolite profiles and contrasting in insect resistance. Based on previous SNP annotations [100] and in combination with seed availability we selected 19 accessions from 10 different tomato species. A relatively large number of accessions belong to the species *Solanum habrochaites* (previous known as *S. typicum* f. *glabratum* or *S. typicum* f. *hirsutum*), as these are well known for herbivore resistance and specialised metabolite production [22–24, 27, 54, 73, 101–106]. Tomato accessions were obtained from Enza Zaden, Enkhuizen according to the appropriate guidelines and licences for plant material. Plants were grown in an enclosed greenhouse compartment under controlled conditions (22–25 °C, 16/8 h light/dark regime). All leaf material used in the experiments originates from leaflets of similar developmental stage and plant age. Unless otherwise indicated, the first pair of leaflets from the fourth fully expanded leaf from the apex of a whole plant was used. For each experiment (i.e. whitefly bioassay, thrips bioassay, trichome density, acylsugar measurements, volatile measurements), a separate batch of plants was grown from seed while randomly distributed over the same greenhouse compartment.

The *B. tabaci* population (MEAM1) was reared in a climate cabinet (Snijders, Tilburg; 16/8 h light/dark; 150 $\mu\text{E m}^{-2} \text{s}^{-1}$ at 28 °C) on fresh cucumber plants. The Western Flower Thrips (*F. occidentalis*) population was reared on runner bean supplemented with *Typha latifolia* pollen in an ECO2box plastic container (Duchefa Biochemie, Haarlem, The Netherlands), located in a climate cabinet (16/8 h light/dark; 150 $\mu\text{E m}^{-2} \text{s}^{-1}$ at 24–20 °C).

Insect bioassays and statistics

Bioassays were tailored to the specific insect to provide the most reliable read out. The resistance of 19 different tomato accessions to whitefly (*B. tabaci*; MEAM1) were determined using a no-choice bioassay to obtain the mean percentage of whitefly survival. Whitefly bioassays were performed using 6-weeks old plants randomly localised over an enclosed greenhouse compartment (22–25 °C; 16/8 h light/dark). Per tomato accession, 3 individual plants were used except for PI127826 ($n = 5$) and MM ($n = 8$) used to monitor block- and positional

effects. Each plant harboured two clip cages (\varnothing 2.5 cm, Bioquip, cat #1458) placed on the fourth fully expanded leaf from the apex. The cages were filled with \sim 15 adult whiteflies and after 5 days the number of dead and alive whiteflies per clip cage was recorded using a stereomicroscope (Euromex StereoBlue Zoom SB.1902). Whitefly survival (%) was first averaged over the two clip cages to obtain an average survival per plant before calculating the accession's average survival and standard error. To create two phenotypic groups as input for the RF modelling, a generalised linear model was fitted using the *glm* and the *glmer* function of R package *lme4* [107, 108] taking the variable "accession" as a fixed effect and using *S. lycopersicum* accession Moneymaker, functioning as a sensitive standard, as the intercept. Whitefly survival data from *S. pennellii* LA0716 was excluded from this analysis; as the survival was 0%, the coefficient could not be calculated. Accessions with statistically significant ($p < 0.05$) coefficients were labelled as "resistant" and others as "susceptible".

For the thrips assays we assayed larval performance, as thrips become virus vectors at this life stage, hence establishing the median survival time as survival parameter. For this, \varnothing 15 mm leaf discs were placed in a 12-well plate (Greiner-bio one #665102) on top of a \varnothing 20 mm filter paper (Whatman #28413904) with 100 μ l of tap water to keep the leaf discs hydrated. Per well (24 to 36 wells per accession) one thrips larva (L1 stage) from a synchronised population was placed on the adaxial leaf surface of a leaf disc. Plates were closed, sealed with tape and placed back into the climate cabinet after which the survival status, 1 (dead) or 0 (alive), of each individual was documented daily for 19 days. Kaplan-Meier survival curves (Additional file: Figure S1) were drawn in R [108] using the *survfit* function [109] and the *ggsurvplot* function of the *survminer* package version 0.4.0 [110]. A Cox proportional hazard model, as previously used by Macel et al. [111], was fitted using the *coxph* function of the *survival* [109] package with the accession as explanatory variable (Additional file: Table S2). For each tomato accession, the fitted Cox proportional hazard regression model estimated the probability coefficient for a thrips L1 larva to die when exposed to a leaf disc, compared to the baseline (susceptible standard *S. lycopersicum* Moneymaker). For instance, *S. habrochaites* LYC4 has an exponentiated coefficient of 215 meaning that the chance of dying on a LYC4 leaf disc is 215 times greater compared to MM. Accessions with a statistically significant ($p < 0.01$) coefficient higher than 1 were labelled as "resistant" and otherwise as "susceptible".

Trichome density and microscopy images

Leaflets were taken from 3 to 5 individual plants per accessions, from which four leaf discs (\varnothing 4 mm) were taken;

two leaf discs were used to count trichomes on the abaxial side and two for the adaxial side. For all, type II/III/V/VIII ("non-glandular"), glandular type I/IV and type VI were counted using a stereomicroscope (Euromex StereoBlue Zoom SB.1902). Type I and IV were taken together as one group as the two types are related and contain similar acylsugar compositions in the glands. Moreover, though type I should have a taller stalk than type IV [112] this criterium was impossible to distinguish varying in the wild accessions. In the absence of a statistically significant leaf disc effect ($p > 0.05$), trichome density (trichomes/mm²) was averaged over the two leaf discs from an individual leaflet (plotted in Additional file: Figure S3), whereafter the average trichome density per accession was calculated. As the setup of the insect-bioassay allowed whiteflies to move freely between both sides of the leaflet, trichome densities were averaged over the abaxial-and adaxial surface to analyse the effect of trichome density on whitefly survival (Fig. 2b; Additional file: Figure S3c). Thrips larvae were restricted to reside only on the adaxial surface and the survival rates were therefore compared to trichome densities of the adaxial surface only (Fig. 2c; Additional file: Figure S3d). Linear modelling of the trichome densities and insect-survival rates was done in R using the simple linear modelling *lm* function.

Pictures of wild- and cultivated tomato trichomes were obtained using a Leica MZFLIII stereomicroscope connected to a Nikon DS-Fi2 digital camera. Tissue was placed under the microscope and photos were taken in layers of \sim 10 nm under bright light. Layers (20 to 60 photos) were stacked into a single image using layer stacking in Adobe Photoshop (CC 2019, Adobe Systems).

Metabolic profiling

Metabolite extractions

Volatiles were extracted from one leaflet ($n = 3-4$), that was weighed and briefly (\sim 5 s) immersed in 500 μ l *n*-hexane spiked with 0.5 ng/ μ l benzyl acetate as internal standard. The short extraction time served to avoid extraction additional cuticle components, which was subsequently monitored. Next, 10 mg Na₂SO₄ (s) was added to the extract to remove water and the sample was vortexed for \sim 10 s. Extracts were then centrifuged for 5 min at 13.9 k rcf whereafter the hexane layer was removed and stored in glass vials under N₂ (g) at -20° C prior to the analysis. For acylsugar extractions, two leaflets ($n = 6$) were collected, weighed and immersed in 2 mL dichloromethane while gentle rocking for 30 s. The extracts were air-dried overnight, stored at -20° C, and dissolved in 300 μ l MeOH for analysis.

Data acquisition and identification

Volatile analysis was done using an Agilent 7890A gas chromatograph coupled to a 7200 accurate mass time-

of-flight (TOF) mass spectrometer. One μL of extract was injected, immediately heated to 275°C and separated on a HP-5 ms column ($30\text{ m} \times 250\ \mu\text{m}$; $0.25\ \mu\text{m}$ film thickness; Agilent) using helium as a carrier gas (7.0699 psi; flow rate: $1\text{ mL}/\text{min}$). The column was heated for 3 min at 40°C thereafter the temperature increased with $5^\circ\text{C}/\text{min}$ to 140°C , then $10^\circ\text{C}/\text{min}$ to 250°C which was held for 5 min. Samples were ionised at 70 eV under vacuum using EI. After a solvent delay of 4.1 min, ions in the range of 30–350 mu were detected at 50 scans/second. Peak detection was done using MassHunter Qualitative Analysis software package (Agilent). After chromatogram deconvolution, peaks were picked with 50 ppm accuracy when having a minimum ion count of 1% compared to the highest peak in the chromatogram and a S/N ratio > 10 . Metabolite identification was done by mass spectrum comparison to the NIST and Adams (2007) libraries [113, 114] in combination with the Kovats Retention Index (KI) [115], calibrated by running a C8-C20 alkane standard (Sigma Aldrich #04070-5ML). Identifications were verified by analytical standards when available and peaks with a library hit of low confidence were labelled as “unknown” (Additional File: Table S3). Due to their strong co-elution β -phellandrene and D-limonene peaks could not be separated and were treated as one. Metabolite peaks were integrated by their base-peak ion and peak areas were corrected by the internal standard and sample dilution and normalised by the fresh-leaf weight.

Acylsugar extracts were analysed using an Agilent 1290 Infinity II UHPLC coupled to an Agilent 6230 TOF mass spectrometer. One μL of extract was injected and separated on an Agilent ZORBAX RRHD Eclipse Plus C18 column at 50°C with a mobile-phase flow-rate of $0.3\text{ mL}/\text{min}$. The mobile phase consisted of water + 0.1% formic acid (A) and acetonitrile + 0.1% formic acid (B) in the following A:B gradient; from 60:40 to 45:55 in 6 min to 10:90 in 8 min to 60:40 in 3 min. Molecules were ionised at 325 eV in positive mode and ions were detected in the range of 50–1500 mu at 1 spectrum/sec. Acylsugars were identified using MassHunter Quantitative Analysis software (Agilent) by calculating the molecular formulas of the parent ion constituting the chromatographic peaks. Here, molecular formulas were constrained by allowing carbon, hydrogen and oxygen atoms to reconstruct the parent ion, in combination with H^+ , Na^+ and K^+ and formate adducts to appear, plus a double bond equivalent (DBE) range of 1–10. The generated molecular formula, in combination with the DBE, allows extrapolation of the acylsugar molecular structure; the backbone moiety, number of esterified acyl chains and the total number of carbon atoms constituting the acyl chains. Structural isomers were labelled by a hyphen followed by a number.

Random forest

As input for the RF classification algorithm, metabolite levels of the accessions were averaged across biological replicates and the accessions' classification (i.e. “resistant” or “susceptible”) was taken from the regression analysis. Every RF model was run by a six-fold cross validation scheme to build training-and test sets using 1000 decision trees, yielding a feature importance for every metabolite. The model was run 5 times with random initialisation parameters producing an average feature importance and standard deviation for each metabolite (Fig. 4, step 1). Next, the significance of each metabolite's average feature importance was evaluated by comparison to feature-importance distributions obtained from 100, 250 and 500 RF models using randomly-permuted datasets (Fig. 4, step 2). A metabolite was regarded significantly contributing to phenotypic classification when the calculated average feature importance was higher than in 95% of the 100 permuted RF models ($p < 0.01$; Fig. 4, step 3). The modelling was performed using the Python Machine Learning library *scikit-learn* implementation [116] in the Python Pandas data structure [117]. The full script is available as Jupyter notebook in the companion GitHub repository (<http://www.github.com/BleekerLab>).

Methods and plant material described above complied with relevant institutional, national and international guidelines and legislation.

Abbreviations

DBE: Double bond equivalent; GC-MS: Gas chromatography-mass spectrometry; KI: Kovats retention index; MM: MoneyMaker; NG: Non glandular; PLS: Partial least squares; PLS-DA: Partial least squares discriminant analysis; Q-TOF: Quadrupole time-of-flight; RF: Random forest; UHPLC: Ultra high-performance liquid chromatography; iC5: 3-methyl-butanoate; aiC5: 2-methyl-butanoate; iC11: 9-methyl-decanoate

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-021-03070-x>.

Additional file 1: Figure S1. Kaplan-Meier survival curves of the thrips bioassay. The data shows the survival of the thrips larvae over time on leaf discs of the accessions used in this study.

Additional file 2: Figure S2. Trichome photos of all 19 accessions. The photos illustrate the diverse trichome landscape that can be found over the accessions.

Additional file 3: Figure S3. Trichome densities per trichome-type and leaf surface. Boxplots showing the individual data points used to calculate the average trichome densities and the scatterplots show the relationship between trichomes densities and insect survival rates that were not displayed in the main figures.

Additional file 4: Figure S4. Overview of the acylsugar composition of the accessions. The plots display the contribution of the different structural features of the detected acylsugars measured in each accession and the relationship between the total ion-counts and insect survival rates.

Additional file 5: Figure S5. Overview of the structural composition of volatile metabolites. The plots display the contribution of each structural

class to the total blend of volatiles measured in each accession and the relation between the total ion-counts and insect survival.

Additional file 6: Figure S6. Feature importance of metabolites as computed by the whitefly-metabolite RF-models versus their distribution in permuted models. Only metabolites with a feature importance significantly higher compared to the permuted models are shown.

Additional file 7: Figure S7. Feature importance of metabolites as computed by the thrips-volatiles RF-models versus their distribution in permuted models. Only metabolites with a feature importance significantly higher compared to the permuted models are shown.

Additional file 8: Figure S8. Mass spectra of metabolites with a significant feature importance that could not be verified by an authentic standard. Given spectra were used for identification by library comparison in combination with their Kovats Retention Index.

Additional file 9: Table S1. Generalised Linear Model fitted on the whitefly survival data. The output of the model was used to classify accessions as either resistant or susceptible to whiteflies which served as an input for the random forest algorithm.

Additional file 10: Table S2. Cox proportional hazard model fitted on the thrips survival data. The output of the model was used to classify accessions as either resistant or susceptible to whiteflies which served as an input for the random forest algorithm.

Additional file 11: Table S3. Normalised peak-area of the volatiles and acylsugars detected over the samples. The file contains the data that was used as input material for Fig. 3, Fig. 5 and the random forest models and includes the experimental-and theoretical Kovats Retention Indices used for identification of the volatiles.

Acknowledgments

We thank Laura Riedijk for her help in the thrips phenotyping assays and Sanne Ypenburg and Michelle van der Gragt for their help with the trichome counting. Jurre Bleeker is acknowledged for the acylsugar measurements and Johan Westerhuis and Yipeng Song for their input on the data and statistical analysis.

Authors' contributions

PB, RK conceived the study; RK and PB wrote the manuscript. RK, MG, AM, BA performed the experiments; MH provided the photographs and PB, MG and RK drew the image in Fig. 4. MG conducted statistical analyses, MG, SH and FK performed the Random Forest modelling; RS and MH revised the manuscript. All of the authors discussed the results and commented on the manuscript, and approved its submission for potential publication.

Funding

This work was supported by a Marie-Curie Individual Fellowship (grant number 655656, DURETO) for MG, covering full time salary of the Postdoctoral researcher and by an NWO VIDI-grant from the Dutch Organisation for Scientific Research (NWO; grant number 12988) obtained by PB, covering all project cost and the salary of RK and AM. The funding bodies were not involved in the design of the study, nor in the collection, analysis, and interpretation of the data or writing of the manuscript.

Availability of data and materials

All data generated during this study are included in this published article and its supplementary information files. We provide the underlying datasets, and code as R-scripts and Jupyter notebooks, through a linked GitHub repository (<https://github.com/BleekerLab/natural-insecticides-thrips-whiteflies>) and at the Zenodo archive (<https://zenodo.org/>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Green Life Science Research Cluster, Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands. ²Data Analysis Group, Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands. ³Enza Zaden Research & Development B.V, Haling 1E, 1602 DB Enkhuizen, The Netherlands. ⁴Radboud University Medical Center, Bacterial Genomics Group, Geert Grooteplein Zuid 26-28, 6525 GA Nijmegen, The Netherlands.

Received: 15 February 2021 Accepted: 20 May 2021

Published online: 02 July 2021

References

- Cerda R, Avelino J, Gary C, Tixier P, Lechevallier E, Allinne C. Primary and secondary yield losses caused by pests and diseases: assessment and modeling in coffee. *PLoS One*. 2017;12(1):e0169133. <https://doi.org/10.1371/journal.pone.0169133>.
- Fereres A, Moreno A. Behavioural aspects influencing plant virus transmission by homopteran insects. *Virus Res*. 2009;141(2):158–68. <https://doi.org/10.1016/j.virusres.2008.10.020>.
- Liu B, Preisser EL, Chu D, Pan H, Xie W, Wang S, et al. Multiple forms of vector manipulation by a plant-infecting virus: Bemisia tabaci and tomato yellow leaf curl virus. *J Virol*. 2013;87:03571–12.
- Shi X, Chen G, Pan H, Xie W, Wu Q, Wang S, et al. Plants pre-infested with Viruliferous MED/Q cryptic species promotes subsequent Bemisia tabaci infestation. *Front Microbiol*. 2018;9. <https://doi.org/10.3389/fmicb.2018.01404>.
- Stafford CA, Walker GP, Ullman DE. Hitching a ride: vector feeding and virus transmission. *Communicative Integr Biol*. 2012;5(1):43–9. <https://doi.org/10.4161/cib.18640>.
- Brown JK, Czosnek H. Whitefly transmission of plant viruses. *Adv Plant Pathol*. 2002;36:65–100.
- Rotenberg D, Jacobson AL, Schneewis DJ, Whitfield AE. Thrips transmission of tospoviruses. *Curr Opin Virol*. 2015;15:80–9. <https://doi.org/10.1016/j.coviro.2015.08.003>.
- Van De Wetering F, Goldbach R, Peters D. Tomato spotted wilt tospovirus ingestion by first instar larvae of Frankliniella occidentalis is a prerequisite for transmission. *PHYTOPATHOLOGY-NEW YORK AND BALTIMORE THEN ST PAUL*. 1996;86:900–5.
- Moritz G, Kumm S, Mound L. Tospovirus transmission depends on thrips ontogeny. *Virus Res*. 2004;100(1):143–9. <https://doi.org/10.1016/j.virusres.2003.12.022>.
- Carreck NL. A beekeeper's perspective on the neonicotinoid ban. *Pest Manag Sci*. 2017;73(7):1295–8. <https://doi.org/10.1002/ps.4489>.
- Huseth AS, Chappell TM, Chitturi A, Jacobson AL, Kennedy GG. Insecticide resistance signals negative consequences of widespread neonicotinoid use on multiple field crops in the US Cotton Belt. *Environ Sci Technol*. 2018; 52(4):2314–22. <https://doi.org/10.1021/acs.est.7b06015>.
- Kliot A, Kontsedalov S, Ramsey JS, Jander G, Ghanim M. Adaptation to nicotine in the facultative tobacco-feeding hemipteran Bemisia tabaci. *Pest Manag Sci*. 2014;70(10):1595–603. <https://doi.org/10.1002/ps.3739>.
- Rundlöf M, Andersson GKS, Bommarco R, Fries I, Hederström V, Herbertsson L, et al. Seed coating with a neonicotinoid insecticide negatively affects wild bees. *Nature*. 2015;521(7550):77–80. <https://doi.org/10.1038/nature14420>.
- Painter RH. Insect resistance in crop plants. The Macmillan Company; 1951.
- Stenberg JA, Muola A. How should plant resistance to herbivores be measured? *Front Plant Sci*. 2017;8:663. <https://doi.org/10.3389/fpls.2017.00663>.
- Kortbeek R, van der Gragt M, Bleeker PM. Endogenous plant metabolites against insects. *Eur J Plant Pathol*. 2019;154(1):67–90. <https://doi.org/10.1007/s10658-018-1540-6>.
- Mithöfer A, Boland W. Plant defense against herbivores: chemical aspects. *Annu Rev Plant Biol*. 2012;63(1):431–50. <https://doi.org/10.1146/annurev-arplant-042110-103854>.
- War AR, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, et al. Mechanisms of plant defense against insect herbivores. *Plant Signal Behav*. 2012;7(10):1306–20. <https://doi.org/10.4161/psb.21663>.

19. Glas JJ, Schimmel BCJ, Alba JM, Escobar-Bravo R, Schuurink RC, Kant MR. Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores. *Int J Mol Sci*. 2012;13(12):17077–103. <https://doi.org/10.3390/ijms131217077>.
20. Huchelmann A, Boutry M, Hachez C. Plant glandular trichomes: natural cell factories of high biotechnological interest. *Plant Physiol*. 2017;175(1):6–22. <https://doi.org/10.1104/pp.17.00727>.
21. Rattan RS. Mechanism of action of insecticidal secondary metabolites of plant origin. *Crop Prot*. 2010;29(9):913–20. <https://doi.org/10.1016/j.cropro.2010.05.008>.
22. Freitas JA, Maluf WR, das Graças Cardoso M, Gomes LAA, Bearzotti E. Inheritance of foliar zingiberene contents and their relationship to trichome densities and whitefly resistance in tomatoes. *Euphytica*. 2002;127(2):275–87. <https://doi.org/10.1023/A:1020239512598>.
23. Antonious GF, Dahlman DL, Hawkins LM. Insecticidal and acaricidal performance of methyl ketones in wild tomato leaves. *Bull Environ Contam Toxicol*. 2003;71(2):400–7. <https://doi.org/10.1007/s00128-003-0178-y>.
24. Bleeker PM, Mirabella R, Diergaarde PJ, VanDoorn A, Tissier A, Kant MR, et al. Improved herbivore resistance in cultivated tomato with the sesquiterpene biosynthetic pathway from a wild relative. *Proc Natl Acad Sci*. 2012;109(49):20124–9. <https://doi.org/10.1073/pnas.1208756109>.
25. Lucatti AF, van Heusden AW, de Vos RC, Visser RG, Vosman B. Differences in insect resistance between tomato species endemic to the Galapagos Islands. *BMC Evol Biol*. 2013;13(1):175. <https://doi.org/10.1186/1471-2148-13-175>.
26. Lucatti AF, Meijer-Dekens F, Mumm R, Visser RGF, Vosman B, van Heusden S. Normal adult survival but reduced *Bemisia tabaci* oviposition rate on tomato lines carrying an introgression from *S. habrochaites*. *BMC Genet*. 2014;15(1):142.
27. Levy J, Tamborindeguy C. *Solanum habrochaites*, a potential source of resistance against *Bactericera cockerelli* (Hemiptera: Trioziidae) and "Candidatus *Liberibacter solanacearum*". *J Econ Entomol*. 2014;107(3):1187–93. <https://doi.org/10.1603/EC13295>.
28. Firdaus S, van Heusden AW, Hidayati N, Supena EDJ, Mumm R, de Vos RC, et al. Identification and QTL mapping of whitefly resistance components in *Solanum galapagense*. *Theor Appl Genet*. 2013;126(6):1487–501. <https://doi.org/10.1007/s00122-013-2067-z>.
29. Rakha M, Hanson P, Ramasamy S. Identification of resistance to *Bemisia tabaci* Genn. In closely related wild relatives of cultivated tomato based on trichome type analysis and choice and no-choice assays. *Genet Resour Crop Evol*. 2017;64(2):247–60. <https://doi.org/10.1007/s10722-015-0347-y>.
30. Marchant WG, Legarrea S, Smeda JR, Mutschler MA, Srinivasan R. Evaluating Acylsugars-mediated resistance in tomato against *Bemisia tabaci* and transmission of tomato yellow leaf curl virus. *Insects*. 2020;11(12):842. <https://doi.org/10.3390/insects11120842>.
31. Fan P, Leong BJ, Last RL. Tip of the trichome: evolution of acylsugar metabolic diversity in Solanaceae. *Curr Opin Plant Biol*. 2019;49:8–16. <https://doi.org/10.1016/j.pbi.2019.03.005>.
32. Fan P, Miller AM, Liu X, Jones AD, Last RL. Evolution of a flipped pathway creates metabolic innovation in tomato trichomes through BAH1D enzyme promiscuity. *Nat Commun*. 2017;8(1):2080. <https://doi.org/10.1038/s41467-017-02045-7>.
33. Leong BJ, Lybrand DB, Lou Y, Fan P, Schillmiller AL, Last RL. Evolution of metabolic novelty: a trichome-expressed invertase creates specialized metabolic diversity in wild tomato. *Sci Adv*. 2019;5(4):eaaw3754.
34. Schillmiller AL, Charbonneau AL, Last RL. Identification of a BAH1D acetyltransferase that produces protective acyl sugars in tomato trichomes. *Proc Natl Acad Sci*. 2012;109(40):16377–82. <https://doi.org/10.1073/pnas.1207906109>.
35. Lybrand DB, Anthony TM, Jones AD, Last RL. An integrated analytical approach reveals Trichome Acylsugar metabolite diversity in the wild tomato *Solanum pennellii*. *Metabolites*. 2020;10(10):401. <https://doi.org/10.3390/metabo10100401>.
36. Leckie BM, D'Ambrosio DA, Chappell TM, Halitschke R, De Jong DM, Kessler A, et al. Differential and synergistic functionality of acylsugars in suppressing oviposition by insect herbivores. *PLoS One*. 2016;11(4):e0153345.
37. Ben-Mahmoud S, Smeda JR, Chappell TM, Stafford-Banks C, Kaplinsky CH, Anderson T, et al. Acylsugar amount and fatty acid profile differentially suppress oviposition by western flower thrips, *Frankliniella occidentalis*, on tomato and interspecific hybrid flowers. *PLoS One*. 2018;13(7):e0201583. <https://doi.org/10.1371/journal.pone.0201583>.
38. Kang J, Liu G, Shi F, Jones AD, Beaudry RM, Howe GA. The tomato odorless-2 mutant is defective in trichome-based production of diverse specialized metabolites and broad-spectrum resistance to insect herbivores. *Plant Physiol*. 2010;154(1):262–72. <https://doi.org/10.1104/pp.110.160192>.
39. Bosch M, Berger S, Schaller A, Stintzi A. Jasmonate-dependent induction of polyphenol oxidase activity in tomato foliage is important for defense against *Spodoptera exigua* but not against *Manduca sexta*. *BMC Plant Biol*. 2014;14(1):257. <https://doi.org/10.1186/s12870-014-0257-8>.
40. Ekanayaka EP, Li C, Jones AD. Sesquiterpenoid glycosides from glandular trichomes of the wild tomato relative *Solanum habrochaites*. *Phytochemistry*. 2014;98:223–31. <https://doi.org/10.1016/j.phytochem.2013.11.011>.
41. Bleeker PM, Diergaarde PJ, Ament K, Schütz S, John B, Dijkink J, et al. Tomato-produced 7-epizingiberene and R-curcumene act as repellents to whiteflies. *Phytochemistry*. 2011;72(1):68–73. <https://doi.org/10.1016/j.phytochem.2010.10.014>.
42. Dawood MH, Snyder JC. The alcohol and epoxy alcohol of Zingiberene, produced in Trichomes of wild tomato, Are More Repellent to Spider Mites Than Zingiberene. *Front Plant Sci*. 2020;11:35.
43. Snyder JC, Guo Z, Thacker R, Goodman JP, Pyrek JS. 2, 3-Dihydrofarnesoic acid, a unique terpene from trichomes of *Lycopersicon hirsutum*, repels spider mites. *J Chem Ecol*. 1993;19(12):2981–97. <https://doi.org/10.1007/BF00980597>.
44. Frelichowski JE Jr, Juvik JA. 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*. 2001;94(5):1249–59. <https://doi.org/10.1603/0022-0493-94.5.1249>.
45. Tissier A, Morgan JA, Dudareva N. Plant volatiles: going 'in' but not 'out' of trichome cavities. *Trends Plant Sci*. 2017;22(11):930–8. <https://doi.org/10.1016/j.tplants.2017.09.001>.
46. Kennedy GG, Dimock MB. 2-Tridecanone: A natural toxicant in a wild tomato responsible for insect resistance. In *Natural Products*: Anonymous Elsevier; 1983. p. 123–8.
47. Dimock MB, Kennedy GG. The role of glandular trichomes in the resistance of *Lycopersicon hirsutum* f. *glabratum* to *Heliothis* zea. *Entomol Exp Appl*. 1983;33(3):263–8. <https://doi.org/10.1111/j.1570-7458.1983.tb03267.x>.
48. Kortbeek R, Xu J, Ramirez A, Spyropoulou E, Diergaarde P, Otten-Bruggeman I, et al. Engineering of tomato glandular Trichomes for the production of specialized metabolites. *Meth Enzymol*. 2016;576:305–31. <https://doi.org/10.1016/b.s.mie.2016.02.014>.
49. Breiman L. Random forests. *Mach Learning*. 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>.
50. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? *J Machine Learn Res*. 2014;15(1):3133–81.
51. Sprenger H, Erban A, Seddig S, Rudack K, Thalhammer A, Le MQ, et al. Metabolite and transcript markers for the prediction of potato drought tolerance. *Plant Biotechnol J*. 2018;16(4):939–50. <https://doi.org/10.1111/pbi.12840>.
52. Sartor RC, Noshay J, Springer NM, Briggs SP. Identification of the expressome by machine learning on omics data. *Proc Natl Acad Sci*. 2019;116(36):18119–25. <https://doi.org/10.1073/pnas.1813645116>.
53. Nuez F, Prohens J, Blanca JM. Relationships, origin, and diversity of Galapagos tomatoes: implications for the conservation of natural populations. *Am J Bot*. 2004;91(1):86–99. <https://doi.org/10.3732/ajb.91.1.86>.
54. Antonious S. JC: natural products: repellency and toxicity of wild tomato leaf extracts to the two-spotted spider mite, *Tetranychus urticae* Koch. *J Environ Sci Health B*. 2006;41(1):43–55. <https://doi.org/10.1080/03601230500234893>.
55. Mirnezhad M, Romero-González RR, Leiss KA, Choi YH, Verpoorte R, Klinkhamer PG. Metabolomic analysis of host plant resistance to thrips in wild and cultivated tomatoes. *Phytochem Anal*. 2010;21(1):110–7. <https://doi.org/10.1002/pca.1182>.
56. Vosman B, van't Westende WP, Henken B, van Eekelen HD, de Vos RC, Voorrips RE. Broad spectrum insect resistance and metabolites in close relatives of the cultivated tomato. *Euphytica*. 2018;214(3):46. <https://doi.org/10.1007/s10681-018-2124-4>.
57. Wang F, Park Y, Gutensohn M. Glandular trichome-derived sesquiterpenes of wild tomato accessions (*Solanum habrochaites*) affect aphid performance

- and feeding behavior. *Phytochemistry*. 2020;180:112532. <https://doi.org/10.1016/j.phytochem.2020.112532>.
58. Wink M. Interference of alkaloids with neuroreceptors and ion channels. In *Studies in natural products chemistry*. Volume 21: Anonymous Elsevier; 2000. p. 3–122.
 59. Gershenzon J, Dudareva N. The function of terpene natural products in the natural world. *Nat Chem Biol*. 2007;3(7):408–14. <https://doi.org/10.1038/nchembio.2007.5>.
 60. Zabel S, Brandt W, Porzel A, Athmer B, Bennewitz S, Schäfer P, et al. Tissier, a: a single cytochrome P450 oxidase from *Solanum habrochaites* sequentially oxidizes 7-epi-zingerone to derivatives toxic to whiteflies and various microorganisms. *Plant J*. 2021;105(5):1309–25. <https://doi.org/10.1111/tpj.15113>.
 61. Constabel CP, Barbehenn R. Defensive roles of polyphenol oxidase in plants. In *Induced plant resistance to herbivory*. Anonymous Springer; 2008. p. 253–70.
 62. Vandendorre G, Smaghe G, Van Damme EJ. Plant lectins as defense proteins against phytophagous insects. *Phytochemistry*. 2011;72(13):1538–50. <https://doi.org/10.1016/j.phytochem.2011.02.024>.
 63. Gorb EV, Gorb SN. Anti-adhesive effects of plant wax coverage on insect attachment. *J Exp Bot*. 2017;68(19):5323–37. <https://doi.org/10.1093/jxb/erx271>.
 64. Fich EA, Fisher J, Zamir D, Rose JK. Transpiration from tomato fruit occurs primarily via trichome-associated transcuticular polar pores. *Plant Physiol*. 2020;184(4):1840–52. <https://doi.org/10.1104/pp.20.01105>.
 65. Weston PA, Johnson DA, Burton HT, Snyder JC. Trichome secretion composition, trichome densities, and spider mite resistance of ten accessions of *Lycopersicon hirsutum*. *J Am Soc Hortic Sci*. 1989;114(3):492–8.
 66. Snyder JC, Carter CD. Leaf trichomes and resistance of *Lycopersicon hirsutum* and *L. esculentum* to spider mites. *J Am Soc Hortic Sci*. 1989;109(6):837–43.
 67. Goffreda JC, Mutschler MA, Tingey WM. Feeding behavior of potato aphid affected by glandular trichomes of wild tomato. *Entomol Exp Appl*. 1988;48(2):101–7. <https://doi.org/10.1111/j.1570-7458.1988.tb01152.x>.
 68. Neal JW Jr, Bentz J. Evidence for the stage inducing phenotypic plasticity in pupae of the polyphagous whiteflies *Trialeurodes vaporariorum* and *Bemisia argentifolii* (Homoptera: Aleyrodidae) and the reason d'être. *Ann Entomol Soc Am*. 1999;92(6):774–87. <https://doi.org/10.1093/aesa/92.6.774>.
 69. Tian D, Tooker J, Peiffer M, Chung SH, Felton GW. Role of trichomes in defense against herbivores: comparison of herbivore response to woolly and hairless trichome mutants in tomato (*Solanum lycopersicum*). *Planta*. 2012;236(4):1053–66. <https://doi.org/10.1007/s00425-012-1651-9>.
 70. Firdaus S, Van Heusden A, Harpenas A, Supena ED, Visser RG, Vosman B. Identification of silverleaf whitefly resistance in pepper. *Plant Breed*. 2011;130(6):708–14. <https://doi.org/10.1111/j.1439-0523.2011.01894.x>.
 71. Muigai SG, Bassett MJ, Schuster DJ, Scott JW. Greenhouse and field screening of wild *Lycopersicon* germplasm for resistance to the whitefly *Bemisia argentifolii*. *Phytoparasitica*. 2003;31(1):27–38. <https://doi.org/10.1007/BF02979764>.
 72. Alba JM, Montserrat M, Fernández-Muñoz R. Resistance to the two-spotted spider mite (*Tetranychus urticae*) by acylsucroses of wild tomato (*Solanum pimpinellifolium*) trichomes studied in a recombinant inbred line population. *Exp Appl Acarol*. 2009;47(1):35–47. <https://doi.org/10.1007/s10493-008-9192-4>.
 73. Kim J, Kang K, Gonzales-Vigil E, Shi F, Jones AD, Barry CS, et al. Striking natural diversity in glandular trichome acylsugar composition is shaped by variation at the *Acytransferase2* locus in the wild tomato *Solanum habrochaites*. *Plant Physiol*. 2012;160(4):1854–70. <https://doi.org/10.1104/pp.112.204735>.
 74. Leckie BM, De Jong DM, Mutschler MA. Quantitative trait loci increasing acylsugars in tomato breeding lines and their impacts on silverleaf whiteflies. *Mol Breed*. 2012;30(4):1621–34. <https://doi.org/10.1007/s11032-012-9746-3>.
 75. Smeda JR, Schillmiller AL, Last RL, Mutschler MA. Introgression of acylsugar chemistry QTL modifies the composition and structure of acylsugars produced by high-accumulating tomato lines. *Mol Breed*. 2016;36(12):160. <https://doi.org/10.1007/s11032-016-0584-6>.
 76. Bruce TJ, Wadhams LJ, Woodcock CM. Insect host location: a volatile situation. *Trends Plant Sci*. 2005;10(6):269–74. <https://doi.org/10.1016/j.tplants.2005.04.003>.
 77. Darshane HL, Ren H, Ahmed N, Zhang Z, Liu Y, Liu T. Volatile-mediated attraction of greenhouse whitefly *Trialeurodes vaporariorum* to tomato and eggplant. *Front Plant Sci*. 2017;8:1285. <https://doi.org/10.3389/fpls.2017.01012.85>.
 78. Escobar-Bravo R, Klinkhamer PG, Leiss KA. Induction of jasmonic acid-associated defenses by thrips alters host suitability for conspecifics and correlates with increased trichome densities in tomato. *Plant Cell Physiol*. 2017;58(3):622–34. <https://doi.org/10.1093/pcp/pcx014>.
 79. Chen G, Klinkhamer PG, Escobar-Bravo R, Leiss KA. Type VI glandular trichome density and their derived volatiles are differently induced by jasmonic acid in developing and fully developed tomato leaves: implications for thrips resistance. *Plant Sci*. 2018;276:87–98. <https://doi.org/10.1016/j.plantsci.2018.08.007>.
 80. Zhang P, Wei J, Zhao C, Zhang Y, Li C, Liu S, et al. Airborne host–plant manipulation by whiteflies via an inducible blend of plant volatiles. *Proc Natl Acad Sci*. 2019;116(15):7387–96. <https://doi.org/10.1073/pnas.1818599116>.
 81. Bac-Molenaar JA, Mol S, Verlaan MG, van Elven J, Kim HK, Klinkhamer PG, et al. Trichome independent resistance against Western flower Thrips in tomato. *Plant Cell Physiol*. 2019;60(5):1011–24. <https://doi.org/10.1093/pcp/pcz018>.
 82. Juvik JA, Berlinger MJ, Ben-David T, Rudich J. Resistance among accessions of the generally lycopersicon and solanum to four of the main insect pests of tomato in Israel. *Phytoparasitica*. 1982;10(3):145–56. <https://doi.org/10.1007/BF02994524>.
 83. Undas AK, Weihrauch F, Lutz A, van Tol R, Delatte T, Verstappen F, et al. The use of metabolomics to elucidate resistance markers against damson-hop aphid. *J Chem Ecol*. 2018;44(7–8):711–26. <https://doi.org/10.1007/s10886-018-0980-y>.
 84. Macel M, Visschers IG, Peters JL, Kappers IF, de Vos RC, van Dam NM. Metabolomics of Thrips resistance in pepper (*Capsicum spp.*) reveals monomer and dimer acyclic Diterpene glycosides as potential chemical defenses. *J Chem Ecol*. 2019;45(5):490–501. <https://doi.org/10.1007/s10886-019-01074-4>.
 85. Riach AC, Perera M, Florance HV, Robinson LA, Penfield SD, Hill JK. Metabolic fingerprints reveal how an insect metabolome is affected by different larval host plant species. *Arthropod Plant Interact*. 2019;13(4):571–9. <https://doi.org/10.1007/s11829-019-09671-6>.
 86. Sanchez-Arcos C, Kai M, Svatoš A, Gershenzon J, Kunert G. Untargeted metabolomics approach reveals differences in host plant chemistry before and after infestation with different pea aphid host races. *Front Plant Sci*. 2019;10:188. <https://doi.org/10.3389/fpls.2019.00188>.
 87. Nelson PRC, Taylor PA, MacGregor JF. Missing data methods in PCA and PLS: score calculations with incomplete observations. *Chemom Intell Lab Syst*. 1996;35(1):45–65. [https://doi.org/10.1016/S0169-7439\(96\)00007-X](https://doi.org/10.1016/S0169-7439(96)00007-X).
 88. Gromski PS, Muhamadali H, Ellis DI, Xu Y, Correa E, Turner ML, et al. A tutorial review: metabolomics and partial least squares-discriminant analysis—a marriage of convenience or a shotgun wedding. *Anal Chim Acta*. 2015;879:10–23. <https://doi.org/10.1016/j.aca.2015.02.012>.
 89. Lê Cao K, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics*. 2011;12(1):253. <https://doi.org/10.1186/1471-2105-12-253>.
 90. Kokla M, Virtanen J, Kolehmainen M, Paananen J, Hanhineva K. Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: a comparative study. *BMC Bioinformatics*. 2019;20(1):1–11.
 91. Mirza B, Wang W, Wang J, Choi H, Chung NC, Ping P. Machine learning and integrative analysis of biomedical big data. *Genes*. 2019;10(2):87. <https://doi.org/10.3390/genes10020087>.
 92. Acharjee A, Kloosterman B, de Vos RC, Werij JS, Bachem CW, Visser RG, et al. Data integration and network reconstruction with omics data using random Forest regression in potato. *Anal Chim Acta*. 2011;705(1–2):56–63. <https://doi.org/10.1016/j.aca.2011.03.050>.
 93. Scott IM, Vermeer CP, Liakata M, Corol DI, Ward JL, Lin W, et al. Enhancement of plant metabolite fingerprinting by machine learning. *Plant Physiol*. 2010;153(4):1506–20. <https://doi.org/10.1104/pp.109.150524>.
 94. Ranganathan Y, Borges RM. Reducing the babel in plant volatile communication: using the forest to see the trees. *Plant Biol*. 2010;12(5):735–42. <https://doi.org/10.1111/j.1438-8677.2009.00278.x>.
 95. Ben-Mahmoud S, Anderson T, Chappell TM, Smeda JR, Mutschler MA, Kennedy GG, et al. A thrips vector of tomato spotted wilt virus responds to

- tomato acylsugar chemical diversity with reduced oviposition and virus inoculation. *Sci Rep*. 2019;9(1):1–11.
96. Zhou F, Pichersky E. The complete functional characterisation of the terpene synthase family in tomato. *New Phytol*. 2020;226(5):1341–60. <https://doi.org/10.1111/nph.16431>.
 97. Rodríguez-López MJ, Moriones E, Fernández-Muñoz R. An Acylsucrose-producing tomato line derived from the wild species *Solanum pimpinellifolium* decreases fitness of the whitefly *Trialeurodes vaporariorum*. *Insects*. 2020;11(9):616. <https://doi.org/10.3390/insects11090616>.
 98. Mihaylova-Kroumova AB, Artiouchine I, Korenkov VD, Wagner GJ. Patterns of inheritance of acylsugar acyl groups in selected interspecific hybrids of genus *Nicotiana*. *J Plant Res*. 2020;133(4):509–23. <https://doi.org/10.1007/s10265-020-01188-x>.
 99. Isman MB, Grieneisen ML. Botanical insecticide research: many publications, limited useful data. *Trends Plant Sci*. 2014;19(3):140–5. <https://doi.org/10.1016/j.tplants.2013.11.005>.
 100. Aflitos S, Schijlen E, de Jong H, de Ridder D, Smit S, Finkers R, et al. Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *Plant J*. 2014;80(1):136–48.
 101. Vallejo C, Restrepo S, Lobo A. Genetic introgression of the resistance to fruit borer present in *Solanum habrochaites* var *glabratum* into *Solanum lycopersicum* 'unapal-Maravilla'. *Acta Horticulturae*. 2009;821:89–94.
 102. Ghosh B, Westbrook TC, Jones AD. Comparative structural profiling of trichome specialized metabolites in tomato (*Solanum lycopersicum*) and *S. habrochaites*: acylsugar profiles revealed by UHPLC/MS and NMR. *Metabolomics*. 2014;10(3):496–507. <https://doi.org/10.1007/s11306-013-0585-y>.
 103. Gonzales-Vigil E, Hufnagel DE, Kim J, Last RL, Barry CS. Evolution of TPS20-related terpene synthases influences chemical diversity in the glandular trichomes of the wild tomato relative *Solanum habrochaites*. *Plant J*. 2012;71(6):921–35. <https://doi.org/10.1111/j.1365-3113.2012.05040.x>.
 104. Momotaz A, Scott JW, Schuster DJ. Identification of quantitative trait loci conferring resistance to *Bemisia tabaci* in an F2 population of *Solanum lycopersicum* × *Solanum habrochaites* accession LA1777. *J Am Soc Hortic Sci*. 2010;135(2):134–42. <https://doi.org/10.21273/JASHS.135.2.134>.
 105. Li C, Schmidt A, Pichersky E, Shi F, Jones AD. Identification of methylated flavonoid regioisomeric metabolites using enzymatic semisynthesis and liquid chromatography-tandem mass spectrometry. *Metabolomics*. 2013;9(1):92–101. <https://doi.org/10.1007/s11306-012-0451-3>.
 106. Schwahn K, de Souza LP, Fernie AR, Tohge T. Metabolomics-assisted refinement of the pathways of steroidal glycoalkaloid biosynthesis in the tomato clade. *J Integr Plant Biol*. 2014;56(9):864–75. <https://doi.org/10.1111/jipb.12274>.
 107. Bates D, Mächler M, Bolker B, Walker S. "Fitting Linear Mixed-Effects Models Using lme4." *J Stat Softw*. 2015;67(1):1–48. <https://doi.org/10.18637/jss.v067.i01>.
 108. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012. URL <http://www.R-project.org> 2018
 109. Therneau TM. A Package for Survival Analysis in R. R package version 3.2-11, 2021. <https://CRAN.R-project.org/package=survival>.
 110. Kassambara A, Kosinski M. *Survminer: Drawing Survival Curves using 'ggplot2'*; 2018.
 111. Macel M, Bruinsma M, Dijkstra SM, Ooijendijk T, Niemeyer HM, Klinkhamer PG. Differences in effects of pyrrolizidine alkaloids on five generalist insect herbivore species. *J Chem Ecol*. 2005;31(7):1493–508. <https://doi.org/10.1007/s10886-005-5793-0>.
 112. McDowell ET, Kapteyn J, Schmidt A, Li C, Kang J, Descour A, et al. Comparative functional genomic analysis of *Solanum* glandular trichome types. *Plant Physiol*. 2011;155(1):524–39. <https://doi.org/10.1104/pp.110.1.67114>.
 113. Shen VK, Siderius DW, Krekelberg WP, Hatch HW. NIST standard reference database number 173. National Institute of Standards and Technology; 2006.
 114. Adams RP. Identification of essential oil components by gas chromatography/mass spectrometry: allured publishing corporation; 2007.
 115. van Den Dool H, Dec. Kratz P: a generalization of the retention index system including linear temperature programmed gas—liquid partition chromatography. *J Chromatogr A*. 1963;11:463–71. [https://doi.org/10.1016/S0021-9673\(01\)80947-X](https://doi.org/10.1016/S0021-9673(01)80947-X).
 116. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12(Oct):2825–30.
 117. The Pandas Development Team, pandas-dev/pandas. Pandas, 2020, Zenodo. <https://doi.org/10.5281/zenodo.3509134>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

