



Metabolic fingerprints reveal how an insect metabolome is affected by different larval host plant species

A. C. Riach¹ · M. V. L. Perera² · H. V. Florance^{2,3} · L. A. Robinson⁴ · S. D. Penfield⁵ · J. K. Hill¹

Received: 24 March 2016 / Accepted: 4 January 2019 / Published online: 22 February 2019
© Springer Nature B.V. 2019

Abstract

Oligophagous insects can consume a wide range of different host plant species, but how these host plants vary in their metabolite compositions and the extent to which this variation affects the biochemistry of the insect herbivores is largely unknown. An understanding of how defensive metabolites from plants are processed by insects may help us develop more effective pesticides. We studied the interactions between the oligophagous insect herbivore *Pieris rapae* (Lepidoptera: Pieridae) and five species of its larval host plants (family Brassicaceae and Cleomaceae) by examining untargeted metabolic fingerprints of the plants and the larval herbivores feeding on them. Visualisation of the metabolic fingerprints of the different host plant species showed highly distinctive clusters in the PCA-X score plots. Larvae could also be distinguished based on the species of host plant they fed on but clusters overlapped to a greater extent. The fingerprints of larvae feeding on *Cleome spinosa* plants were most distinctive due to a large group of abundant metabolites also found in high abundance in *C. spinosa*, but not in the other host plants examined. We conclude that host plants influence the biochemistry of their larval herbivores, and that some metabolites are conserved from one trophic level to the next.

Keywords Metabolic fingerprinting · Metabolome · Host plant · *Pieris rapae*

Introduction

Plants are used as a food source by nearly half of all insect species (Wu and Baldwin 2010), and there is a large body of research examining factors that affect these

herbivore–plant interactions (Schoonhoven et al. 2005). When insect herbivores consume host plants, they take in nutritious primary plant metabolites as well as defensive secondary metabolites which can be repellent or even toxic. The number of metabolites an insect consumes is generally unknown although a single species of plant is estimated to contain thousands of metabolites (Davies et al. 2010). Due to this large number of metabolites and the difficulties in measuring all these small compounds within an organism (Allwood et al. 2008) our knowledge is incomplete as to which metabolites occur within which plants, the concentrations of them and how they relate to the metabolites in the herbivores consuming the plants. Thus, wider examination of the composition and abundance of metabolites in plants and their insect herbivores could lead to a better understanding of insect–plant interactions. Such knowledge could help in the development of effective pesticides that are not circumvented by insects; assess if insects are affected by the genetic modification of host plants or measure the effects upon insects of increases in defensive metabolites caused by agricultural practices. We investigated the effects of host plant metabolite composition on the oligophagous butterfly *Pieris rapae*

Handling editors: Guy Smagghe and Heikki Hokkanen.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11829-019-09671-6>) contains supplementary material, which is available to authorized users.

✉ A. C. Riach
Anna.Riach@york.ac.uk

¹ Department of Biology, University of York, Heslington, York YO10 5DD, UK

² Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QD, UK

³ School of Biological Life Sciences, University of Edinburgh, Edinburgh EH9 3FF, UK

⁴ School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK

⁵ John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

(Lepidoptera; small white butterfly). Many larval host plants used by *P. rapae* belong to the plant family Brassicaceae (Asher et al. 2001; Beilstein et al. 2008; Stevens 2001 onwards). This family contains glucosinolates which are toxic to some herbivore species (Li et al. 2000; Rohr et al. 2009) but are used by other herbivores as ovipositional cues and phagostimulants (Huang and Renwick 1994; Renwick and Lopez 1999). Glucosinolates are well studied (Kroymann 2011; Rask et al. 2000), however, the focus on one group of metabolites may have led to less familiar groups of compounds being ignored despite evidence that other important compounds in the Brassicaceae family exist (Schroeder et al. 2006). Therefore, widening the range of metabolites examined in these host plants may bring new, important compounds to our attention and metabolic fingerprinting, which can measure the abundances of hundreds of metabolites simultaneously (Fiehn 2001), has the potential to do this.

Assessment of the herbivore in an insect–plant relationship usually considers performance factors such as growth rate, mortality and fecundity, whereas quantifying the metabolome of an insect (the metabolome being all the metabolites within an organism; Fiehn 2001) has been less common (Jansen et al. 2009; Snart et al. 2015). As with plant metabolomes (Bidart-Bouzat and Imeh-Nathaniel 2008), insect metabolomes are affected by abiotic factors, such as temperature (Colinet et al. 2012; Malmendal et al. 2006; Michaud et al. 2008; Verberk et al. 2013; Williams et al. 2014). However, little is known about the effect of diet on insect metabolomes, specifically whether insect metabolomes are affected by the species of host plant the insect feeds on. Evidence that insect metabolomes may be changed by insect diet comes from studies which have focused on single plant metabolites or specific metabolite groups (Opitz and Müller 2009), but studies that investigate a broader spectrum of metabolites from the insect metabolome are lacking. For example, once a wider range of metabolites are considered, the metabolome of an organism might be chiefly constant irrespective of diet.

In this paper, we use untargeted metabolomics to measure the abundances of a large number of polar metabolites, but without identifying these metabolites by name, to address two questions. First, we confirm that there are differences in the metabolic fingerprints of five species of Brassicales host plants before examining whether or not there are differences in the metabolic fingerprints of *P. rapae* raised on those host plants. Second, we examine differences among the five groups of *P. rapae* insects by differentiating the large number of metabolites we detect and cross-referencing the metabolites among these groups. This determines the number of metabolites found in all larvae irrespective of host plant and the number of metabolites that are characteristic to insects fed a specific host plant.

Methods

Plant and insect rearing

We studied *P. rapae* on five of its host plants (Asher et al. 2001; Braby and Trueman 2006; Eeles 2018). Four plant species were from the family Brassicaceae, namely *Arabidopsis thaliana* (thale cress; seeds from Penfield lab, John Innes Centre, UK), *Barbarea vulgaris* (yellow rocket; Thompson and Morgan, Suffolk, UK), *Brassica oleracea* (cabbage; Groves Nurseries, Dorset, UK) and *Lunaria annua* (honesty; Chiltern Seeds, Oxfordshire, UK) and one from the related family Cleomaceae, *Cleome spinosa* (spider flower; Chiltern Seeds, Oxfordshire, UK). The host plant species studied were chosen from those that *P. rapae* larvae had been reported to feed on, would grow readily in the greenhouse and that would span as wide a phylogenetic range of host plants as possible, allowing us to examine plants likely to have contrasting chemical compositions.

Plant seeds were sown in trays using Levington F2 + S seed and modular compost (added N:P:K 150:200:200 mg/l) over a 2-week period in an unheated glasshouse in July and August which varied in temperature between 10 and 30 °C. Plants were grown under natural daylight and watered daily. Four to six week old plant material was harvested on the same day between 9 am and 10.30 am and flash frozen for metabolomic analysis. Leaves were taken from 13 individuals per plant species to make 13 biological replicates per host plant species. Samples comprised the youngest leaves of the plants although the number of leaves had to vary to obtain sufficient plant material for analysis and account for variation in size of the plants. Therefore, the single youngest leaf of *B. oleracea* and *L. annua*, the three youngest leaves of *B. vulgaris* and *C. spinosa*, and the whole rosette of *A. thaliana* were sampled. These sampled leaves reflected the parts of the plants that the larvae were observed typically to consume.

P. rapae larvae were the F1 offspring of 16 adult female butterflies caught in York, UK (53°95'N, 1°08'W) in August. Female butterflies were kept individually in a glasshouse, provided with honey solution and potted *B. oleracea* plants for oviposition. Five days after hatching, second instar larvae were transferred using a paint brush and randomly assigned to different host plants. Larvae were placed in plastic boxes (175 × 116 × 52 mm³; 13 boxes per plant species) with fine netting taped over the front. Ten larvae, from a mixture of females, were placed in each box to ensure there were sufficient larvae for analysis despite mortality. Cut plant leaves, similar to those taken as samples for metabolite analyses, were placed in vials of water inside boxes, except for *A. thaliana* plants (which quickly wilted using this method) which were left

in soil in small pots. Plant material was replaced daily as required and thus larvae were provided with excess fresh leaves during the experiment. Larvae in boxes were kept in temperature-controlled cabinets (Sanyo, MLR 350) at 21 °C under a photoperiod of 16L:8D. When the majority of larvae in a box had developed to fifth-instar (between 13 and 22 days after hatching), four larvae from a box were randomly chosen, pooled and flash frozen for metabolomic analysis after the head and gut were removed to avoid analysing undigested plant material. Subsequently, the fat body was voided of haemolymph. A total of 13 pooled larval samples was analysed for each of the five host plant species.

Metabolic fingerprint analyses

Leaf and insect material was stored at -80 °C and then freeze dried for 16 h. Material was ground for 2 min at 20 rpm in a ball mill. 10 mg samples were extracted twice with 400 μ l of 80% methanol on ice. 1.98 μ g of umbelliferone per ml of the methanol extraction buffer was added as an internal standard for quality control (Perera 2011). Samples were sonicated, vortexed and the supernatant removed. The two supernatants were combined and filtered through a 0.4 μ m (PVDF) syringe filter.

Metabolite profiling of leaf and insect material was performed using a QToF 6520 mass spectrometer coupled to a 1200 series Rapid Resolution LC system. To analyse samples in a random order, they were randomized in the tray which was kept at 5 °C. 5 μ l of sample extract was loaded onto a Zorbax StableBond C18 1.8 μ m, 2.1×100 mm² reverse phase analytical column (LC/MS and column, Agilent Technologies, Palo Alto, USA). Features were detected in positive ionisation mode only. In retrospect, a negative ion mode would have given greater coverage of the metabolome. Mobile phase A comprised 5% acetonitrile with 0.1% formic acid in water, and mobile phase B was 95% acetonitrile with 0.1% formic acid in water. The following gradient was used: 0 min—0% B; 1 min—0% B; 5 min—20% B; 20 min—100% B; 30 min—100% B; 31 min—0% B; 7 min post time. The flow rate was 0.25 ml min⁻¹ and the column temperature was held at 35 °C for the duration of the chromatographic run. The source conditions for electrospray ionisation were as follows: gas temperature was 325 °C with a drying gas flow rate of 9 l min⁻¹ and a nebuliser pressure of 35 psig. The capillary voltage was 3.5 kV. Skimmer and fragmentor voltages were 115 V and 70 V respectively. The scan range used was 50–1700 m/z. Reference ions (positive ion mode: 121.0509 and 933.0098) were directly infused through a second nebulizer to enable a reference ion correction every scan. Calibration was corrected using the reference ions with every scan. This method could have been improved if pooled samples had been run intermittently among the samples.

Precursor ion scans (full scan mode) were followed by automated MS/MS scans to aid putative identification of metabolites of interest during post processing. For acquisition, four full scans per second were performed and the five most abundant precursor ions were selected for MS/MS. The collision induced dissociation (CID) values were automated and applied using the following equation: (slope) \times (m/z)/100 + offset where the slope was 3.5 V and the offset 5 V. The m/z refers to the precursor ion mass selected by the acquisition software during the full scan.

Metabolic fingerprint data pre-processing

The Molecular Feature Extractor (MFE) in MassHunter software (Agilent Technologies, Palo Alto, USA) identified features (potential metabolites) from peaks produced by the LC/MS. Features eluting within the first minute are contained within the 'dead' volume, and features post 27 min are within the re-equilibration period. These particular features were excluded. During Mass Hunter deconvolution, ion abundance was set to 100. During deconvolution, the software included sodium and potassium adducts, but other adducts such as ammonium ions and neutral losses such as water and phosphates were excluded to minimise errors in identifying unique features. For deconvolution the mass resolution was 25 ppm. The scope of the instrumentation relied on a tolerance of up to 20 ppm so further information regarding retention time and fragmentation was found to be of use when eliminating false hits.

The alignment of features across samples, filtering out noise, and missing value imputation were performed using an in-house alignment algorithm, 'Kernel Feature Alignment'. This included normalisation of the data using a log₂ transformation followed by calculating the total volume using all the feature's areas from across the extract. The internal standard was also used in the alignment process (Perera 2011). Plant and insect data had to be aligned separately due to the large number of samples and metabolites examined. Thus, we could not directly match metabolites found in the insects with those recorded in the plants, and instead we carried out manual searches of the plant dataset using the mass and retention times of metabolites of interest found in insects. Features that were not detected in at least 7 out of the 13 replicates were excluded from the dataset, reducing the number of features detected in plants from 50,958 to 12,023 and in insects from 25,479 to 1481. Principal component analyses (PCAs) were performed on the datasets before and after the removal of these features to ensure their exclusion did not qualitatively alter the conclusions of the analyses, and we only present the more conservative analyses based on their removal. Prior to data analysis, missing value imputation (MVI) was applied in those cases where metabolites were detected in more than

seven but fewer than 13 replicates (Hrydziuszko and Viant 2012). The algorithm used for MVI combined the Gaussian Expectation Maximization (GEM) and *k*th nearest neighbour (kNN) methods (Perera 2011). Data were log₂ transformed and mean centred before multivariate analyses. While a large number of metabolites were measured, it is acknowledged that this will not be all of the metabolites in the metabolomes. While there is not one single method to extract and detect all metabolites, negative ionisation mode would have increased the number of metabolites measured in this experiment.

Visualisation of data

Multivariate analyses were carried out in Simca-P software (Umetrics UK Ltd, Windsor, UK). PCA was used to summarise the metabolomic data to compare patterns of metabolites among plants and among insects. Plant and insect data from each of the five plant species studied were analysed in two separate PCAs. Outliers falling outside Hotelling's T^2 using a 95% confidence level (a generalization of Student's *t*-distribution applied in multivariate situations; Prokhorov 2011) were identified and excluded to prevent their having a disproportionate influence on the analysis (one *B. vulgaris* plant sample and one *B. oleracea* insect sample were excluded). However, the exclusion of these two samples did not have a visible effect on the general sample patterning, as indicated by the PCA score plots. To visualise the data, plant and insect replicates were plotted according to their principal component (PC) scores.

In addition to the PCA, a Partial Least Squares Discriminant Analysis (PLS-DA) model was fitted to the insect metabolomic data (in Simca-P software) whereby this supervised model used the identity of samples (i.e., identity of larval host plant species) to maximize discrimination among groups. The latent variable scores produced are equivalent to principal component scores in PCA. To avoid over-fitting the data, the PLS-DA model was cross-validated using permutations by excluding a seventh of the data in turn and testing the predictive ability of a model fitted to the remaining data. If a PLS-DA model has low predictive ability, the model cannot be validated implying that there are no differences among the groups being examined. Simca-P software presents the predictive ability of the model as a Q^2 value, ranging between 0 and 1, with values closer to 1 indicating better predictive ability.

A further cross-referencing analysis considering only the qualitative presence of a metabolite in the different insect groups (opposed to presence as well as abundance considered in the multivariate analyses) determined the number of metabolites that were unique and shared among insects on different hosts. These metabolites were presented as

proportions of the total number of metabolites measured across all the insect groups.

Isolating important metabolites

To identify metabolites responsible for distinguishing larval samples feeding on a particular plant species, an Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) was performed on the insect metabolic fingerprint data. This supervised analysis explains the maximum amount of variation between two chosen groups of samples (in this case, insects on one host plant and insects on all other host plants). The OPLS-DA model was performed on a balanced dataset (13 replicates of larvae fed on one plant species vs. 13 replicates from a mixture of the other four insect groups) so that the larger dataset would not skew the analysis. This group of 13 larval replicates from the other four larval groups was chosen by performing a PCA on the data and selecting replicates which reflected the full range of PC 1 scores whilst ensuring each of the four host plant groups were represented.

To confirm the differences in individual metabolites, statistical differences in insect metabolite abundances were analysed using ANOVAs carried out in R (R Core Team 2013) with host plant species as a fixed factor, once parametric assumptions were met. The *P* values from each analysis were converted to *q* values to correct for false discovery rates following the large number (1481) of analyses performed (Benjamini and Hochberg 1995; similar to the more common Bonferroni correction). To compare how abundant metabolites of interest were in an insect feeding on one plant species, the average abundances of the 13 larvae samples and the larval group with the next highest average abundance were used to calculate the fold difference (number of times greater) in abundance. Fold differences in abundance were also calculated for equivalent metabolites found in the equivalent plant samples.

Results

Host plants and insects fed on them have distinct metabolic fingerprints

The metabolic fingerprints of the five host plant species were distinguished by the first two principal component (PC) scores in a PCA analysis (Fig. 1). All five species clustered separately although *C. spinosa* was completely separate from the other four plant species. The clusters of *B. oleracea* and *B. vulgaris* were adjacent to each other as were *L. annua* and *A. thaliana*. Together the first two PCA components explained 30.1% of the variation in the metabolite data. As a well-studied group of compounds, glucosinolates were

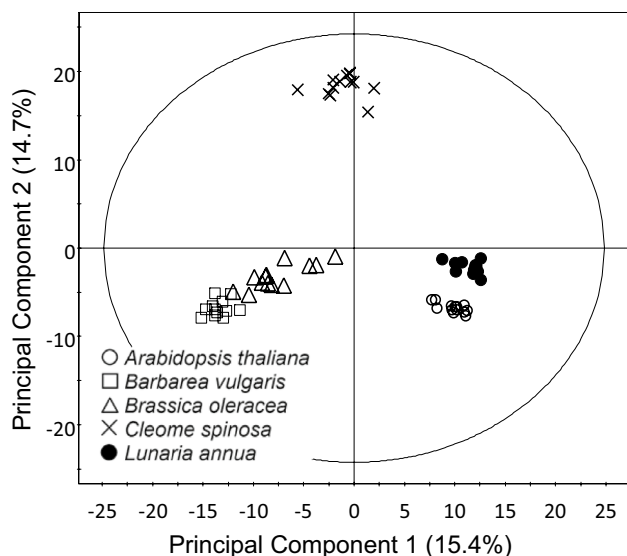


Fig. 1 Scores from a PCA model fitted to the metabolic fingerprints of five host plant species demonstrate that species can be distinguished by their metabolic fingerprints. A total of 51.5% of the variation in the plant metabolic fingerprints was explained by five axes of a PCA fitted to the data. The variation in the data explained by the first two PC scores (30.1%) is shown in brackets on the axes

suspected to account for the differences among the plant species. Thus, an estimate was made of the relative importance of glucosinolates compared with other metabolites in discriminating plant species based on their metabolic fingerprints. We obtained the molecular weights calculated

(m/z) for around 120 glucosinolate compounds from existing databases (Fahey et al. 2002, KEGG, <http://www.genome.jp/kegg/ligand>; METLIN, metlin.scripps.edu/metabo_search_alt2; KNApSACk, http://www.kanaya.naist.jp/knapsack_jsp/top) and 19 metabolites in our plant samples were putatively identified as possible glucosinolates using this molecular weight. To visualise the influence of these metabolites that are likely to be glucosinolates, relative to the other measured metabolites, we highlight their PCA loadings in plots from the PCA analysis performed on the plant metabolic fingerprints, but none of the 19 metabolites had high loadings (Fig. 2). Thus, we conclude that variation in glucosinolates was unlikely to be important in distinguishing among host plant species based on their metabolic fingerprints.

Of the 12,023 metabolites measured in the plants, only 6.89% were found in all five of the plant species and 68.56% were unique to only one of the plants. The largest number of unique metabolites were found in *C. spinosa* (1573 metabolites).

Next, we analysed the metabolite composition of the insect fat bodies. Compared to the distinctiveness of plant metabolic fingerprints (Fig. 1), PCA revealed less differentiation in metabolic fingerprints amongst larvae raised on different host plants (Fig. 3). This demonstrates that differences among the insect metabolomes are smaller than among the plant metabolomes. In the PCA analysis, only those larvae that fed on *C. spinosa* could be distinguished from other larvae (Fig. 3), and when these larvae were excluded from the analysis, we found no differentiation

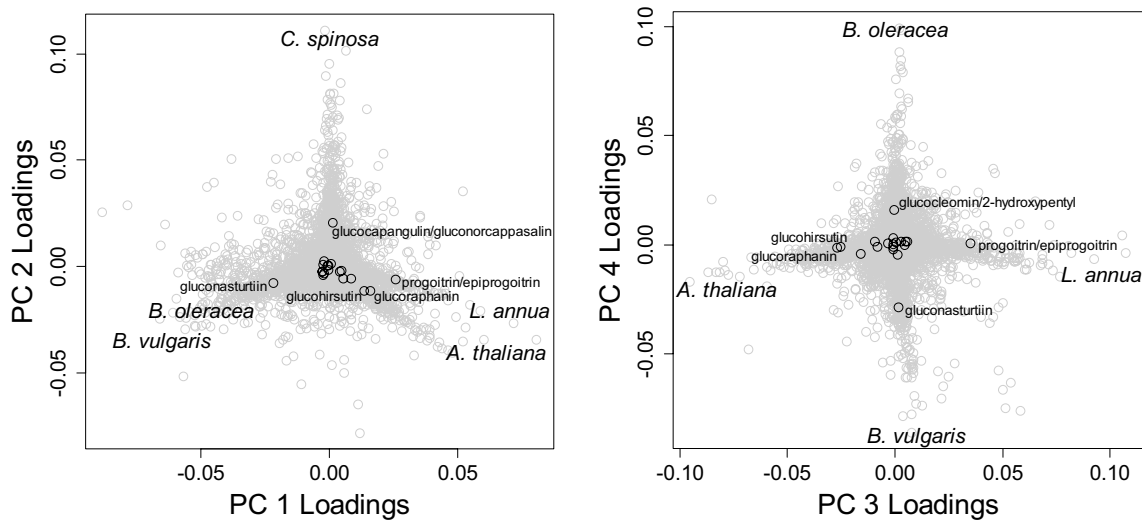


Fig. 2 Loading plots of a PCA analysing the metabolic fingerprints of five host plant species. The 19 metabolites that matched the molecular weights of 29 glucosinolates (i.e., including glucosinolate isomers) have their loadings highlighted in black. For some metabolites the putative glucosinolate identification is shown (based on the molecular mass of the metabolite). The loadings of other metabolites

are shown in grey. The approximate positions of the five host plants in the corresponding PCA score plots are indicated by the names of the plant species except for *C. spinosa* which is in the centre of the PC 3 and 4 loadings plot. Compared with other metabolites, the metabolites putatively identified as glucosinolates do not have a large influence on discriminating the different host plant species

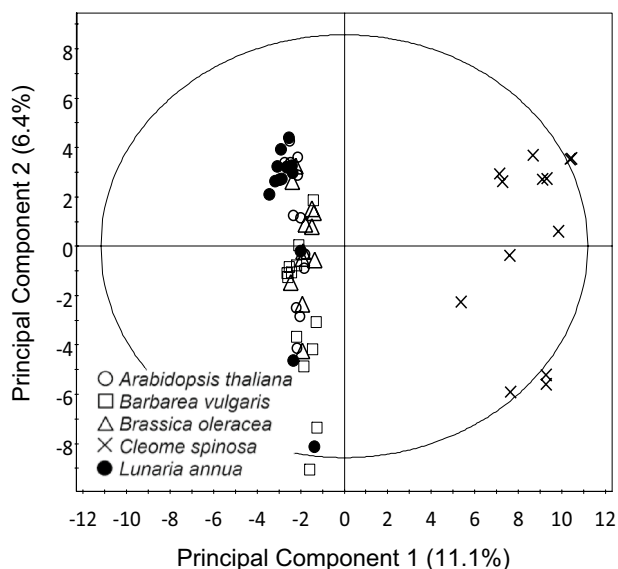


Fig. 3 Scores from PCs 1 and 2 from a PCA model fitted to *P. rapae* larval metabolic fingerprints. Larvae fed on *C. spinosa* are distinguished from larvae feeding on other host plants. The PCA was fitted to four axes which together explained 25.9% of the total variation in the dataset. The variation explained by the first two PCs (17.5%) is shown in brackets on the axes

among larvae fed on the other four host plants. A supervised PLS-DA confirmed the findings of the PCA in that larvae fed *C. spinosa* could be distinguished from other larvae (first latent variable of the PLS-DA; ESM. 1). However, where the unsupervised PCA had failed to distinguish, the next three latent variables of the PLS-DA were

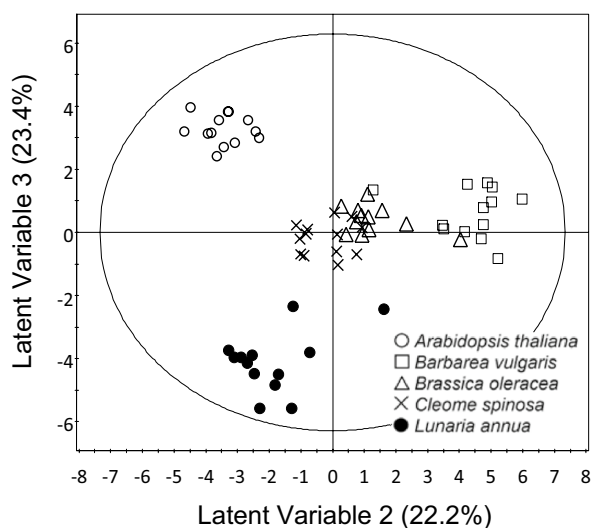
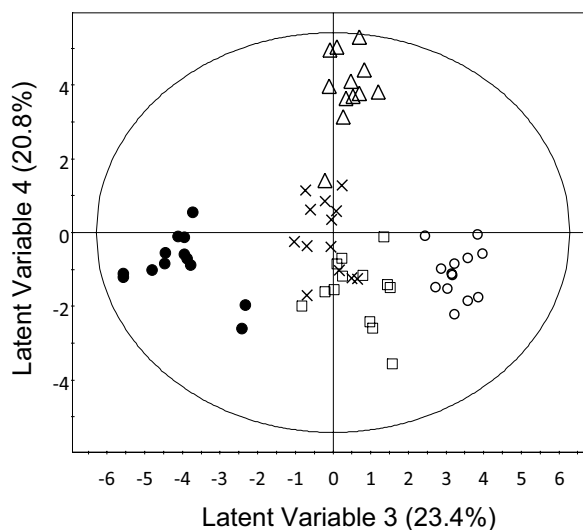


Fig. 4 Latent variables 2, 3 and 4 score plots from a PLS-DA model fitted to larval metabolic fingerprints emphasize the differences among larvae feeding on different host plants. The % of the modelled

able to discriminate among larvae reared on the other host plants (Fig. 4). Thus, host plants influence the metabolic fingerprints of *P. rapae* larvae and these changes are especially large in larvae feeding on *C. spinosa*.

To further investigate metabolite differences among insects feeding on the five host plant species, the number of metabolites that were unique to insects on each host plant species was calculated. The number of metabolites that were found in common across insects on all five plant species was also recorded (Fig. 5). Of the 1481 metabolites measured in insects, 37% were unique to only one of the host plant groups. A large number of unique metabolites were found in those insects that had eaten *C. spinosa* (199 metabolites) reflecting the large number of unique metabolites found in the *C. spinosa* plant. (*A. thaliana* (130) had the second largest number of unique metabolites compared with insects that had eaten the other host plants (*B. oleracea* 64, *B. vulgaris* 85 and *L. annua* 66). We found that 28% of all insect metabolites recorded in the study were found in all five insect groups, indicating that only about a quarter of metabolites were present in *P. rapae* irrespective of larval host plant, and the remaining three quarters of metabolites depended on larval diet.

Of these unique metabolites found in the insects a larger proportion of those in the *C. spinosa* fed insect group (62 out of 199, 31%) were found to match the mass and retention times of metabolites' presence in the *C. spinosa* plant. Fewer mass and retention time matches were found between the other insect groups and their respective plants (*A. thaliana* 8 out of the 130, 6%, *B. oleracea* 5 out of 64, 8%, *B. vulgaris* 9 out of 85, 11% and *L. annua* 5 out of 66, 8%).



Y variation is shown in brackets on the axes labels. 6 latent variables explained 97.2% of the Y variation in the model. The Q^2 value was 0.836 implying very good predictive ability of the model

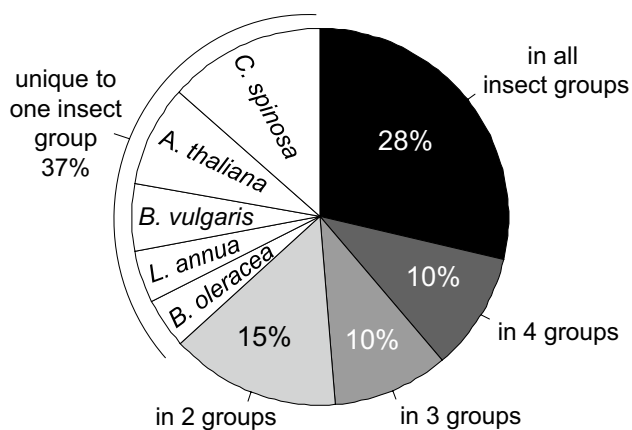


Fig. 5 The total number (1481) of metabolites measured in insects and the proportions found in only one group of insects (white area; i.e., unique to larvae feeding on a particular host plant species) or in multiple groups of insects (grey and black areas). The species of host plant the larvae fed from are in italics

Identification of abundant metabolites in larvae feeding on *C. spinosa*

To understand why the metabolic fingerprints of *C. spinosa*-fed larvae were different, those metabolites that distinguished these larvae from other larvae were found by performing an OPLS-DA model (ESM. 2). The OPLS-DA produced an ‘S-plot’ which was used to visualise the relative importance of different metabolites and to highlight metabolites that were highly abundant in those insects feeding on *C. spinosa*. We focused on metabolites that were at least four times more abundant in *C. spinosa*-fed larvae than in other larvae, to ensure that any selected metabolites represented a major difference in abundance compared to other larvae. Using this criterion, 44 metabolites were found to be highly abundant in *C. spinosa*-fed larvae. Putatively annotated chemical identities for some of these metabolites were obtained using the molecular masses of metabolites as determined by the LC-MS and from molecular masses in metabolic databases (KEGG, METLIN, and KNApSack). Putative matches were found for six out of the 44 metabolites, with some metabolites matching multiple isomers (ESM. 3), and the tandem MS spectra for these metabolites were compared with reference spectra in databases using the Kernel Feature Alignment software (Perera 2011). For one metabolite with mass of 456.35 Da at retention time 22.36 min the tandem MS fragments were similar to the spectrum for oleanolic acid. However, a fragmented standard run through the same machine used to analyse the insect samples failed to confirm this and so none of the metabolites characterising the insect fingerprints raised on *C. spinosa* were definitively identified.

To establish whether the metabolites that were abundant in larvae feeding on *C. spinosa* originated from the host plant, the plant metabolic fingerprints were examined for metabolites with the same masses and retention times as those in insects. Of the 44 metabolites most influential in distinguishing the metabolic fingerprints of larvae fed *C. spinosa* from other larvae, 42 matched metabolites in the plant fingerprint data based on equivalent mass and retention times. Furthermore, the abundances of these metabolites were many times higher in *C. spinosa* plants compared with other plant species (fold difference minimum 4.4 and maximum 176; ESM. 3). Thus, we conclude that the majority (95.5%) of metabolites detected in larvae feeding on *C. spinosa* plants, which distinguished them from other larvae, originated directly from the *C. spinosa* host plants. The other two distinguishing metabolites are hypothesised to be metabolisation products.

Discussion

The advantage of untargeted metabolomics is that a wide range of metabolites are examined, thereby providing a more comprehensive overview of changes to metabolomes. We note that the coverage of the metabolome would be improved if the samples were analysed in negative ionisation mode as well as positive. Nevertheless, by applying this untargeted metabolomics approach to insect herbivores as well as their host plants we could evaluate the effect of host plant species on the insect metabolome. Our discovery of a group of metabolites likely to have transferred from the host plant *C. spinosa* to *P. rapae* larvae was possible because an untargeted approach was used rather than focusing primarily on a single group of known metabolites.

One of our aims in this study was to investigate if the metabolome of an insect herbivore is determined by the host plant it feeds on. While the differences among insects raised on different host plants were less than differences among the plants themselves (Fig. 1), they were still apparent. An unsupervised PCA could only separate those insects raised on *C. spinosa* (Fig. 3), but a supervised PLS-DA could discriminate among the other four groups of larvae (Fig. 4). Thus, host plants influence the metabolite composition in the fat body of *P. rapae* larvae and these changes are especially large in larvae feeding on *C. spinosa*.

Two other targeted studies support the idea that larval diet affects the metabolites found within insects: the concentration of cardenolide compounds in dogbane tiger moths (*Cycnia tenera*) are dependent on the species of *Asclepias* plant larvae feed on (Cohen and Brower 1983), and *P. rapae* larvae were found to contain pinorelinol at the end of their glandular hairs if they fed on *B. oleracea* (Schroeder et al.

2006). Our study is the first to examine the wider insect metabolome to demonstrate dietary effects.

Larvae raised on *C. spinosa* and *A. thaliana* had the highest number of unique metabolites found only in their metabolomes (Fig. 5). *C. spinosa* is listed as a host plant for the Asian subspecies *P. rapae crucivora* but not for European *P. rapae* used in this study (Robinson et al. 2011). Similarly *A. thaliana* grows at a different time of the year to when *P. rapae* larvae feed in the UK and is, therefore, not a natural host plant. This may suggest why the insects feeding on these unfamiliar plants showed more distinctive metabolomes. It is possible that plants not usually encountered by a population are more likely to contain metabolites that the insect cannot metabolise.

Several metabolites that occurred in high abundance were responsible for observed differences between the metabolic fingerprints of larvae-fed *C. spinosa* and larvae fed on other plants. It is highly likely that these chemicals originated from *C. spinosa* plants because metabolites with similar molecular mass and retention times were also found in high abundance in *C. spinosa* plants (ESM. 3). Unfortunately, we were unable to identify these metabolites, although we demonstrated that some metabolites can transfer from *C. spinosa* plants into larvae of *P. rapae* with unchanged chemical structures. Other studies have shown examples of sequestration of metabolites by insects from their host plants (Opitz and Müller 2009) including the cardenolide metabolites from milkweed host plants (Asclepiadaceae) sequestered for defence by monarch butterflies (*Danaus plexippus*; Brower et al. 1967). Therefore, active sequestration of *C. spinosa* metabolites by *P. rapae* larvae is one possibility. Alternatively, these metabolites may have no function in the insects but be passively absorbed and bioaccumulate because larvae are unable to digest, break down or excrete the compounds. There is evidence from other studies revealing the toxicity of *Cleome* sp. to lepidoptera larvae, for example, concentrated extracts of *Cleome arabica* were toxic to larvae of *Spodoptera littoralis* and *Cleome droserifolia* caused mortality in the first instar larvae of *Phthorimaea operculella* (Ladhari et al. 2013; Soliman 2012). However, no mortality of larvae-fed *C. spinosa* was observed in our study, and other feeding trials have not shown increased mortality or reduced growth rates in first instar *P. rapae* fed *C. spinosa* (Riach unpublished data).

The host plants of oligophagous insect herbivores have not previously been examined in relation to their metabolic fingerprints. We showed that host plants of *P. rapae* have very different metabolomes (Fig. 1) which is to be expected since the metabolome is essentially the end product of the genome (Sumner et al. 2003) and species are genetically unique. Much research has focused on differences between Brassicaceae host plants in terms of glucosinolates (Hasapis et al. 1981; Koritsas et al. 1991; Rodman and Chew 1980),

and glucosinolates might be expected to account for the differences in plant species metabolic fingerprints detected in this study. However, our fingerprint data suggests that among Brassicaceae host plant species, there are many other metabolites that have larger variations in abundance than glucosinolates (Fig. 2). The study of insect–Brassicaceae interactions may benefit from considering other metabolites in addition to glucosinolates. This consideration of a larger range of metabolites could more accurately reflect how insects experience difference host plants.

In summary, we have provided new perspectives on interactions between plants and insects and discovered metabolites in the plant *C. spinosa* that are likely to be transferring between trophic levels. It was not possible in this study to identify these metabolites, however, as databases of profiled metabolites improve this could be easier in the future. This study raises the possibility of identifying chemical properties of metabolites that prevent insects from breaking down the chemical which could aid in the development of effective pesticides.

Author contributions ACR, SDP and JKH designed the study, ACR, HVF, LAR and MVLP analysed the data. ACR performed the experiments and instigated the writing of the paper and all authors commented on the paper.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

References

- Allwood JW, Ellis DI, Goodacre R (2008) Metabolomic technologies and their application to the study of plants and plant–host interactions. *Physiol Plantarum* 132:117–135
- Asher J, Warren M, Fox R, Harding P, Jeffcoate G, Jeffcoate S (2001) *The Millennium Atlas of butterflies in Britain and Ireland*. Oxford University Press Inc., New York
- Beilstein MA, Al-Shehbaz IA, Mathews S, Kellogg EA (2008) Brassicaceae phylogeny inferred from phytochrome A and *ndhF* sequence data: tribes and trichomes revisited. *Am J Bot* 95:1307–1327. <https://doi.org/10.3732/ajb.0800065>
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 57:289–300
- Bidart-Bouzat MG, Imeh-Nathaniel A (2008) Global change effects on plant chemical defenses against insect herbivores. *J Integr Plant Biol* 50:1339–1354. <https://doi.org/10.1111/j.1744-7909.2008.00751.x>
- Braby MF, Trueman JWH (2006) Evolution of larval host plant associations and adaptive radiation in Pierid butterflies. *J Evol Bio* 19:1677–1690. <https://doi.org/10.1111/j.1420-9101.2006.01109.x>
- Brower LP, Brower JV, Corvino JM (1967) Plant poisons in a terrestrial food chain. *Proc Natl Acad Sci USA* 57:893–898. <https://doi.org/10.1073/pnas.57.4.893>
- Cohen JA, Brower LP (1983) Cardenolide sequestration by the dogbane tiger moth (*Cycnia tenera*, Arctiidae). *J Chem Ecol* 9:521–532. <https://doi.org/10.1007/bf00990224>

- Colinet H, Larvor V, Laparie M, Renault D (2012) Exploring the plastic response to cold acclimation through metabolomics. *Funct Ecol* 26:711–722. <https://doi.org/10.1111/j.1365-2435.2012.01985.x>
- Davies HV, Shepherd LVT, Stewart D, Frank T, Rohlig RM, Engel KH (2010) Metabolome variability in crop plant species—when, where, how much and so what? *Regul Toxicol Pharm* 58:S54–S61. <https://doi.org/10.1016/j.yrtph.2010.07.004>
- Eeles P (2018) Species Listings: Small White <http://www.ukbutterflies.co.uk/species.php?species=rapae>
- Fahey JW, Zalcmann AT, Talalay P (2002) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants (vol 56, p. 5, 2001). *Phytochemistry* 59:237–237. [https://doi.org/10.1016/s0031-9422\(01\)00419-8](https://doi.org/10.1016/s0031-9422(01)00419-8)
- Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genomics* 2:155–168. <https://doi.org/10.1002/cfg.82>
- Hasapis X, Macleod AJ, Moreau M (1981) Glucosinolates of 9 Cruciferae and 2 Capparaceae species. *Phytochemistry* 20:2355–2358. [https://doi.org/10.1016/s0031-9422\(00\)82665-5](https://doi.org/10.1016/s0031-9422(00)82665-5)
- Hrydziusko O, Viant MR (2012) Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. *Metabolomics* 8:S161–S174. <https://doi.org/10.1007/s11306-011-0366-4>
- Huang XP, Renwick JAA (1994) Relative activities of glucosinolates as oviposition stimulants for *Pieris rapae* and *P. napi oleracea*. *J Chem Ecol* 20:1025–1037. <https://doi.org/10.1007/bf02059739>
- Jansen JJ, Allwood JW, Marsden-Edwards E, van der Putten WH, Goodacre R, van Dam NM (2009) Metabolomic analysis of the interaction between plants and herbivores. *Metabolomics* 5:150–161. <https://doi.org/10.1007/s11306-008-0124-4>
- Koritsas VM, Lewis JA, Fenwick GR (1991) Glucosinolate responses of oilseed rape, mustard and kale to mechanical wounding and infestation by cabbage stem flea beetle (*Psylliodes chrysocephala*). *Ann Appl Biol* 118:209–221. <https://doi.org/10.1111/j.1744-7348.1991.tb06099.x>
- Kroymann J (2011) Natural diversity and adaptation in plant secondary metabolism. *Curr Opin Plant Biol* 14:246–251. <https://doi.org/10.1016/j.pbi.2011.03.021>
- Ladhari A, Laarif A, Om ezzine F, Haouala R (2013) Effect of the extracts of the spiderflower, *Cleome arabica*, on feeding and survival of larvae of the cotton leafworm, *Spodoptera littoralis*. *J Insect Sci* 13:1–14
- Li Q, Eigenbrode SD, Stringham GR, Thiagarajah MR (2000) Feeding and growth of *Plutella xylostella* and *Spodoptera eridania* on *Brassica juncea* with varying glucosinolate concentrations and myrosinase activities. *J Chem Ecol* 26:2401–2419
- Malmendal A, Overgaard J, Bundy JG, Sorensen JG, Nielsen NC, Loeschcke V, Holmstrup M (2006) Metabolomic profiling of heat stress: hardening and recovery of homeostasis in *Drosophila*. *Am J Physiol Regul Integr Comp Physiol* 291:R205–R212. <https://doi.org/10.1152/ajpregu.00867.2005>
- Michaud MR, Benoit JB, Lopez-Martinez G, Elnitsky MA, Lee RE, Denlinger DL (2008) Metabolomics reveals unique and shared metabolic changes in response to heat shock, freezing and desiccation in the Antarctic midge, *Belgica antarctica*. *J Insect Physiol* 54:645–655. <https://doi.org/10.1016/j.jinsphys.2008.01.003>
- Opitz SEW, Müller C (2009) Plant chemistry and insect sequestration. *Chemoecology* 19:117–154. <https://doi.org/10.1007/s00049-009-0018-6>
- Perera MVL (2011) Metabolic profiling of plant disease: from data alignment to pathway predictions. Dissertation, University of Exeter
- Prokhorov AV (2011) Hotelling's T^2 distribution. *Encyclopedia of Mathematics*: <http://www.encyclopediaofmath.org/index.php?title=Hotelling-T%5E2-distribution&oldid=12614>. Accessed 17 February 2016
- R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Rask L, Andreasson E, Ekblom B, Eriksson S, Pontoppidan B, Meijer J (2000) Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Mol Biol* 42:93–113. <https://doi.org/10.1023/a:1006380021658>
- Renwick JAA, Lopez K (1999) Experience-based food consumption by larvae of *Pieris rapae*: addiction to glucosinolates? *Entomol Exp Appl* 91:51–58. <https://doi.org/10.1046/j.1570-7458.1999.00465.x>
- Robinson G, Ackery P, Kitching I, Beccaloni G, Hernández L (2001) Hostplants of the moth and butterfly caterpillars of the Oriental Region. Natural History Museum in association with Southdene Sdn Bhd, Malaysia
- Rodman JE, Chew FS (1980) Phytochemical correlates of herbivory in a community of native and naturalized Cruciferae. *Biochem Syst Ecol* 8:43–50. [https://doi.org/10.1016/0305-1978\(80\)90024-1](https://doi.org/10.1016/0305-1978(80)90024-1)
- Rohr F, Ulrichs C, Mewis I (2009) Variability of aliphatic glucosinolates in *Arabidopsis thaliana* (L.)—impact on glucosinolate profile and insect resistance. *J Appl Bot Food Qual* 82:131–135
- Schoonhoven L, van Loon J, Dicke M (2005) Insect–plant biology. Second edition. Oxford University Press, Oxford
- Schroeder FC, del Campo ML, Grant JB, Weibel DB, Smedley SR, Bolton KL, Meinwald J, Eisner T (2006) Pinosinol: a lignol of plant origin serving for defense in a caterpillar. *Proc Nat Acad Sci USA* 103:15497–15501. <https://doi.org/10.1073/pnas.0605921103>
- Snart CJP, Hardy ICW, Barrett DA (2015) Entometabolomics: applications of modern analytical techniques to insect studies. *Entomol Exp Appl* 155:1–17. <https://doi.org/10.1111/eea.12281>
- Soliman MMM (2012) Toxicity and biological effects of some wild plant extracts against potato tuber moth, *Phthorimaea operculella* (Zeller) (Lepidoptera: Gelechiidae). *Egypt J Biol Pest Control* 22:19–25
- Stevens PF (2001) Angiosperm Phylogeny Website. <http://www.mobot.org/MOBOT/research/APweb/>. Accessed 17 February 2016
- Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*. [https://doi.org/10.1016/s0031-9422\(02\)00708-2](https://doi.org/10.1016/s0031-9422(02)00708-2)
- Verberk WCEP, Sommer U, Davidson RL, Viant MR (2013) Anaerobic metabolism at thermal extremes: A metabolomic test of the oxygen limitation hypothesis in an aquatic insect. *Integr Comp Biol* 53:609–619. <https://doi.org/10.1093/icb/ict015>
- Williams CM, Waranabe M, Guarracino MR, Ferraro MB, Edison AS, Morgan TJ, Boroujerdi AFB, Hahn DA (2014) Cold adaptation shapes the robustness of metabolic networks in *Drosophila melanogaster*. *Evolution* 68:3505–3523
- Wu J, Baldwin IT (2010) New Insights into plant responses to the attack from insect herbivores. *Annu Rev Genet* 44:1–24. <https://doi.org/10.1146/annurev-genet-102209-163500>