

# The molecular mechanisms that determine different degrees of polyphagy in the *Bemisia tabaci* species-complex

Osnat Malka<sup>1</sup>, Ester Feldmesser<sup>2</sup>, Sharon van Brunschot<sup>3</sup>, Diego Santos Garcia<sup>1</sup>, Susan Seal<sup>3</sup>, John Colvin<sup>3</sup>, and Shai Morin<sup>4</sup>

<sup>1</sup>Affiliation not available

<sup>2</sup>Weizmann Institute of Science

<sup>3</sup>University of Greenwich Natural Resources Institute

<sup>4</sup>Hebrew University of Jerusalem Robert H Smith Faculty of Agriculture Food and Environment

June 4, 2020

## Abstract

The whitefly *Bemisia tabaci* is a closely related group of > 35 cryptic species that feed on the phloem sap of a broad range of host plants. Species in the complex differ in their host-range breadth, but the mechanisms involved remain poorly understood. We investigated, therefore, how six different *B. tabaci* species cope with the environmental unpredictability presented by a set of four common and novel host plants. Behavioural studies indicated large differences in performances on the four hosts and putative specialization of one of the species to cassava plants. Transcriptomic analyses revealed two main insights. First, a large set of genes involved in metabolism (> 85%) showed differences in expression between the six species, and each species could be characterized by its own unique expression pattern of metabolic genes. However, within species, these genes were constitutively expressed, with a low level of environmental responsiveness (i.e., to host change). Second, within each species, sets of genes mainly associated with the super-pathways “environmental information processing” and “organismal systems”, responded to the host switching events. These included genes encoding for proteins involved in sugar homeostasis, signal transduction, membrane transport and immune, endocrine, sensory and digestive responses. Our findings suggested that the six *B. tabaci* species can be divided into four performance/transcriptomic “Types” and that polyphagy can be achieved in multiple ways. However, polyphagy level is determined by the specific identity of the metabolic genes/pathways that are enriched and overexpressed in each species (the species’ individual metabolic “tool kit”).

## 1. INTRODUCTION

Herbivorous insects exhibit a wide range of feeding-strategy interactions with their plant hosts, with the classic main division distinguishing between generalist and specialist habits. Species that can feed and adapt to plants from a diversity of botanical families or broad dietary niches are considered to be generalists, while specialist species feed on one specific family of plant hosts and utilize a narrow dietary niche (Ali & Agrawal, 2012; Birnbaum & Abbot, 2020; Raubenheimer & Simpson, 1999; Simpson & Raubenheimer, 1995). This dichotomic division is largely considered to be a good starting point, but its simplicity is continuously challenged. For example, the discovery of new complexes of cryptic species in the aphid (Hemiptera: Aphididae) and Tachinid (Diptera: Tachinidae) groups (Loxdale & Harvey, 2016; Loxdale, Lushai, & Harvey, 2011; Loxdale, Schöfl, et al., 2011), encouraged researchers to move beyond the traditional categories of generalists and specialists, placing herbivorous insects somewhere between the two extremes, with the generalist species being argued to be in many cases unrecognized complexes of more specialized species (Loxdale & Harvey, 2016; Charlery De La Masseliere et al. 2017; Loxdale, Balog, & Harvey, 2019). Yet, generalism without any

signal of division to host-specialist groups was shown to be highly prevalent in some evolutionary lineages, being not only the ‘probable’ but even the ‘inevitable’ evolutionary outcome in those lineages (Clarke, 2017). As it is agreed that generalism is quite rare (Jaenike, 1990), its presence within certain lineages can be hypothesized to associate with a unique set of largely unknown and restricted biological attributes (mechanisms) that promote the evolution of generalist lineages (Clarke, 2017).

It is widely believed, for example, that generalists and specialists insect herbivores differ in the strategies they use to cope with plants’ secondary defence compounds (Ali & Agrawal, 2012; Heidel-Fischer & Vogel, 2015; Heiko Vogel, Musser, & de la Paz Celorio-Mancera, 2018). Generalists that need to handle a wide range of such defences, are likely to rely on more generalized mechanisms such as avoidance (Cornell & Hawkins, 2003), or an arsenal of detoxification abilities that utilize detoxifying enzymes and xenobiotic transporters (Wannes Dermauw, Pym, Bass, Van Leeuwen, & Feyereisen, 2018a; Després, David, & Gallet, 2007). Specialists, on the other hand, mostly depend on effective sequestration, effectors secretion or specific detoxification mechanisms to cope or manipulate specific defence compounds present in their plant hosts (Engler, Spencer, & Gilbert, 2000; Ratzka, Vogel, Kliebenstein, Mitchell-Olds, & Kroymann, 2002; Sasabe, Wen, Berenbaum, & Schuler, 2004). Beside the challenge of handling a broad versus a narrow arsenal of plant defences, species that are significantly apart on the generalism-specialism spectrum, might also differ in their nutritional requirements. First, generalist are likely to experience greater heterogeneity in the contents and mixtures of nutrients found in their host plants than specialists (Raubenheimer & Simpson, 2003). Second, generalist and specialist can differ in the nutritional balancing of proteins and carbohydrates, the two main dietary nutrients acquired from plants (Behmer, 2009). When given a choice between protein- and carbohydrate-biased foods, generalist insect species, such as the lepidopterans *Spodoptera exigua*, *Heliothis virescens* and *Helicoverpa zea* select protein biased diets (Chiluwal, Roh, Kim, & Park, 2020; Deans et al., 2017; Kwang, Behmer, & Simpson, 2006; Merckx-Jacques, Despland, & Bede, 2008). In comparison, oligo- and monophagous species, such as *Manduca sexta*, *Heliothis subflexa* and *Spodoptera. exempta*, select diets with equal protein:carbohydrate contents or slightly carbohydrate biased, likely reflecting the nutrient content of their evolved host plants (Kimura, Yoshida-Kitahara, & Chiba, 1987; Kwang et al., 2006; Lee, Simpson, & Raubenheimer, 2004; Thompson & Redak, 2005).

An additional layer of complexity comes from the feeding behavior and the anatomy of the feeding apparatus (Bernays, 1998), as both generalist and specialist insect herbivores evolved capabilities that allow chewing, piercing-sucking, mining and boring of plant tissues (Bernays, 1998). For example, the importance of plant-defensive compounds in determining patterns of plant-insect interactions are well established in generalist/specialist chewing-insect herbivores, but more controversial in sap-feeders and pollinators (Ali & Agrawal, 2012; Wannes Dermauw, Pym, Bass, Van Leeuwen, & Feyereisen, 2018b; Kim & Jander, 2007; Stevenson, Nicolson, & Wright, 2017). Also, the nutrient intake space of both generalist and specialist phloem-feeding insects is very different from that described above, as these species need to cope with a poor diet that contains high levels of sugar with low levels of essential components such as amino acids (Douglas, 2006). Interestingly, phloem-feeding insects show an intake target (the amount and balance of nutrients that when ingested, result in maximum performance) not in favor of proteins, but heavily biased (8 : 1, 600 : 75 mM) in favor of carbohydrates (Abisgold, Simpson, & Douglas, 1994). Moreover, the high carbohydrate content of the phloem sap generates high osmotic pressure across the insect gut (Ashford, Smith, & Douglas, 2000). To reduce the osmotic pressure produced, phloem-feeding insects such as aphids and whiteflies utilize several mechanisms. Both aphids and whiteflies synthesize glucose-dominated oligosaccharides via a transglucosidase activity and excrete the oligosaccharides in the honeydew (Cristofolletti, Ribeiro, Deraison, Rahbe, & Terra, 2003; Price et al., 2007). At high sucrose concentrations, aphids and whiteflies also reduce osmotic pressure in the gut by isomerizing sucrose to the disaccharide melazitoze (Ashford et al., 2000) and trehalulose (Byrne & Miller, 1990), respectively. In addition, in order to compensate for their unbalanced diet and to obtain an optimal intake of energy and nutrients, phloem-feeding insects established mutualistic interactions with intracellular bacterial symbionts which provide their herbivore hosts with the required nutrients lacking in their diet, mainly essential amino acid, co-factors and vitamins (Hansen & Moran, 2014).

We focused here on the whitefly *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae), a phloem-feeding

cryptic species complex, containing at least 35 distinct species assigned to ~11 major clades (Barbosa et al., 2014; Hu et al., 2018). Many species in the complex were shown to be flexible in their diet breadth and capable of acquiring new plants hosts to their diet repertoire (De Barro, Liu, Boykin, & Dinsdale, 2011), and at the same time, to significantly differ in their documented host range (Malka et al., 2018). In a previous study, we selected six species within the complex, representing different phylogenetic groups and documented host-ranges, from narrow (~9 plant families) to extreme (~50 plant families) polyphagy (Malka et al., 2018). We tested if differences in the species expression profiles of detoxification genes (~300 genes) on common and novel hosts, were shaped more by their phylogenetic relationships or by their ability to utilize multiple hosts successfully. Our analyses indicated that *B. tabaci* species that differed in their ability to accept or utilize multiple plant hosts, also differed in their detoxification expression patterns. We also described a common detoxification “machinery”, shared between the more generalist species (Malka et al., 2018).

In this study, we extended our analyses to the full transcriptomes of the same six *B. tabaci* species, in order to understand the characteristics of carbohydrate, protein and lipid metabolism in the selected species and to identify the molecular mechanisms that might explain the huge variability in host adaptation among the species in the complex. For that, we utilized 87 transcriptomes (six species fed on four plant hosts and an additional control of 10% sucrose-only diet) and profiled the gene expression patterns of each species, identifying this way a suite of pathways that might be associated with the species level of polyphagy. Our analyses indicated that the majority of genes in each species (85%), were not affected by the insects’ diet. Moreover, plastic responses seemed not to be adaptive and mildly contributed to the performance of only one species. Within the 85% constitutively expressed genes, each species displayed its own unique expression pattern of metabolic genes. These findings suggest that polyphagy can be achieved in multiple ways in *B. tabaci*, but its level is determined by the specific identity of the metabolic genes/pathways that are enriched and overexpressed in each species (the species’ individual metabolic “tool kit”), together with the expression levels of non-metabolic genes in response to novel host plants.

## 2. METHODS

### 2.1 *Bemisia tabaci* and host plant species

Six species of *B. tabaci* representing four different genetic groups (De Barro et al., 2011) were selected for analyses: SSA1-SG3 (Sub-Saharan Africa 1, sub-group 3, collected in Tanzania in 2013/maintained on *Manihot esculenta*), ASIA II-1 (Asia-II genetic group, species 1, collected in Pakistan in 2013/maintained on *Gossypium hirsutum*), New-World 2 (hence after NW2) (New-World genetic group, species 2, collected in Brazil in 2013/maintained on *Solanum lycopersicum*), ( and MEAM1 (Middle East-Asia Minor species 1), MED-Q1 (Mediterranean Q species 1) and UGANDA-MED-ASL (Mediterranean non-silverleafing sub-group from Uganda) (Africa/Middle East/Asia minor genetic group, collected in Peru in 2012/maintained on *Gossypium hirsutum*; France in 2011/maintained on *Capsicum annuum* and Uganda in 2012/maintained on *Ipomoea batatas*, respectively). The identity of the six species was verified using their mtCOI DNA sequences (deposited in study accession number PRJEB21948). At least 2 months (~3–4 generations) before starting the experiments, ~500 founders from each of the six colonies were transferred to eggplant, to allow them to establish on a common baseline host plant. Colonies were reared under standard conditions of 28 ± 2degC, 60% humidity and a 14:10-h light:dark cycle. The selection of the experimental host plants was made based on our previously published literature survey and host reconstruction analyses (Malka et al., 2018), which identified common host plants, shared by many *B. tabaci* species, and non-common novel host plants that are utilized by only few species. Based on this, four host plants were selected, eggplant, a common host (*Solanum melongena*, cv. Black Beauty, Solanaceae/ Solanales), and three novel host plants, also known to produce toxic phytotoxins (Malka et al., 2018): pepper (*Capsicum annuum*, cv. California Wonder; Solanaceae/Solanales), cassava (*Manihot esculenta*, cv. MCol22; Euphorbiaceae/Malpighiales) and kale (*Brassica oleracea*, var. *sabellica*, cv. Dwarf Green Curled; Brassicaceae/Brassicales). All experimental plants were grown in rearing rooms maintained at 28 ± 2degC, 60% humidity and a 14:10 -hr light:dark cycle.

### 2.2 Performance score

The reproductive success of each species on the four host plants was evaluated by scoring the  $F_1/F_0$  ratio. Adult pairs ( $n=20-50$ ) of *B. tabaci* from the colony of each species on eggplants were transferred in triplicates to eggplant, pepper, kale and cassava plants. The adult females were allowed to lay eggs for two weeks and were then removed from the plant. After 30-40 days, the performance score was calculated by estimation the ratio of the emerging  $F_1$  adults to the number of the transferred  $F_0$  adults. The scoring scale of  $F_1/F_0$  was between 1-5:  $F_1/F_0 = 0$  (1),  $F_1/F_0 < 0.1$  (2),  $0.1 \leq F_1/F_0 < 0.5$  (3),  $0.5 \leq F_1/F_0 < 1$  (4),  $F_1/F_0 \geq 1$  (5). (Table S1).

### 2.3 RNA isolation and Illumina sequencing

200 newly emerged adults from each species, grown on eggplants, were subjected to a feeding period of 72 hr on 10% sucrose diet, to obtain a standardized genes expression pattern, which was host-plant independent (plants metabolites and physiology). The sucrose diet lids were then replaced by clip cage lids and the adults were transferred, for a feeding period of 24 hr, to the four experimental host plants: eggplant, pepper, kale and cassava. The control adults were fed on 10% sucrose diet in new sucrose lids (the experimental setup is presented in Figure 1). Next, the surviving adults were collected for RNA extraction. About 50 adults were pooled for each RNA sample to obtain sufficient RNA. Three feeding experiments per host plant were performed for each *B. tabaci* species. Total RNA was extracted according to the manufacturer's instructions (Isolate II mini kit, Bioline). Library construction and sequencing were performed by the Centre for Genomic Technologies at the Hebrew University of Jerusalem, using a NextSeq 500 desktop sequencer, which produced approximately 27 million 75-bp single-end reads for each sample.

### 2.4 Gene expression analysis

The reads obtained were then subjected to quality control using the FASTQC software (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). For mapping and expression analysis, a reference backbone of 46,898 genes data set, established for MED-Q1, was used (provided by Prof. Xiao-Wei Wang, Zhejiang University, China). The dataset was manually curated to include 298 detoxification genes from six families: P450 monooxygenases (P450s), carboxylesterases (COEs), glutathione S-transferases (GSTs), sulfotransferases, UDP-glucosyltransferases (UDPGTs) and ATP-binding cassette transporters (ABC transporters) (for more details see (Malka et al., 2018)). The reads were mapped and quantified using RSEM (RNA-Seq by Expectation Maximization), v1.2.18. (B. Li & Dewey, 2011). The transcript reference was first prepared (rsem-prepare reference), followed by rsem-calculate-expression

with the parameter Bowtie2 (Langmead & Salzberg, 2012). The percentage of mapped reads ranged from 41 to 78. Genes that did not have at least 10 reads in 4% of the samples were filtered out. The RSEM gene quantification for all the remaining genes was used as input for the DESeq2 R package, version 1.10.1. The gene counts were normalized using DESeq2 defaults, taking into account the mapped read numbers of the different samples. Differential expression analysis was performed using a full two-factorial model. Pairwise comparisons were performed between plants within species and between species within plants. All pairwise comparisons were applied with the parameter "cooksCutoff = FALSE". False discovery rate (FDR) was corrected for all the 30,012 genes that were not filtered out. The 95% log<sub>2</sub>- converted fold-change range (2.5%–97.5% quantiles) was -3.65 to 4.14 between species and -1 to 1.2 within species. For visualizations by principal component analysis (PCA) and hierarchical clustering, DESeq2 rld values were used. Rld stands for regularized log transformation of the count data to a log<sub>2</sub> scale. PCA, hierarchical and K-means clustering were performed using the PARTEK GENOMICS SUITE software, version 6.6 (v6.6; St. Louis, MO, 2014). The correlation method was applied to calculate the dispersion matrix of the PCA, and the eigenvectors were normalized. For the hierarchical clustering, Pearson's dissimilarity and complete linkage were applied.

K-means clustering was applied to differentially expressed genes (fold change  $\geq 2$  and false discovery rate  $\leq 0.05$  between any two conditions). Clustering was calculated on standardized rld values using Euclidian distance. The Davies-Bouldin index was used to determine the optimal number of clusters that the transcriptomics data classified into. The number of clusters was chosen based on the minimal index value to ensure a good separation between clusters and cluster homogeneity. This minimal value was manually

inspected to ensure that the minimal index values reflect the biological variability in the data set.

## 2.5 Enrichment Analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations were obtained by submitting the sequence of each transcript to the KEGG Automatic Annotation Server (KAAS, <https://www.genome.jp/kegg/kaas/>) and receiving its KEGG Orthology (KO) number (Moriya, Itoh, Okuda, , & 2007). These were submitted to the KEGG Mapper – Reconstruct Pathway ([http://www.genome.jp/kegg/tool/map\\_pathway.html](http://www.genome.jp/kegg/tool/map_pathway.html)) for identifying the pathways. KEGG gene counts per term and/or cluster were performed using in-house Perl scripts. Enrichment  $P$  values and false discovery rates were calculated by applying the hypergeometric test in R.

## 3. RESULTS

### 3.1 Species performance on four different host plants

We estimated the reproductive success (an ordinal score of the  $F_1/F_0$  performance, see *materials & methods* for more details) of the six studied *B. tabaci* species (MED-Q1, UGANDA-MED-ASL, NW2, SSA1-SG3, MEAM1 and ASIA II-1) on the four host plants (eggplant, kale, pepper and cassava) (Figure 2, Table S1). The reproductive success score on eggplant was 3 for NW2 and 5 for all the other species. On kale, the reproductive success score was 5 for MEAM1, 3 for MED-Q1 and 2 for all the other species. On pepper, the reproductive success score was 2 for UGANDA-MED-ASL and NW2, 3 for MEAM1 and SSA1-SG3, 4 for MED-Q1 and 5 for ASIA II-1. On cassava, only the SSA1-SG3 species produced viable progeny in the  $F_1$  generation with a performance score of 5 versus the other five species tested having a score of 1.

Based on these findings, the six species were divided into four performance types (Figure 3): “Type I”, generalist species with high reproductive capacity on all hosts excluding cassava (a plant species which seems to be outside the “normal” host range of the *B. tabaci* species-complex). From the six species analyzed, only MEAM1 was considered a member of this type. “Type II”, generalist species showing an ability to utilize all host plants (excluding cassava), with reproductive success that can range from poor to good or very good. Both the MED-Q1 and ASIA II-1 species were considered members of this type. “Type III”, species with performance relatively similar to “Type II” species, but with unique specialization to a novel host. The SSA1-SG3 species with its special adaptation to cassava was considered a member of this type. “Type IV”, species that are members of this type have the ability to utilize all host plants, as the three types above, but perform poorly on most of them. The NW-2 and to a moderate extent the UGANDA-MED-ASL species were considered members of this type.

### 3.2 Patterns of gene expression of the six species on the different diets

We performed RNA-seq analyses on the six species after switching them from eggplant to a 24 h feeding period on one of the four plant hosts or a 10% sucrose-only artificial diet (Figure 1). Hierarchical clustering analysis of all the differentially expressed genes in at least one comparison (27417 contigs), demonstrated that the samples group according to their level of phylogenetic relatedness and not according to their performance (“Types”): MEAM1, MED-Q1 and Uganda-MED-ASL in one group, ASIA II-1, SSA1-SG3 and NW2, in the other (Figure 4a). In addition, no consistent grouping pattern of the five diets was found within each species. PCA analyses confirmed the associations determined by the hierarchical clustering (Figure 4b). Next, we identified specific gene expression patterns and groups of co-regulated genes, by performing K-means clustering on the list of 27417 differentially expressed genes. The list clustered into 8 co-expression groups (Figure 5a, Table S2), allowing the identification of groups that are uniquely upregulated mainly in one species and sometimes to some extent in additional one/s. Clusters 5, 6 and 8 were found to be uniquely overexpressed in the NW2, MED-Q1 and the SSA1-SG3 species, respectively. Genes in cluster 1 were overexpressed mainly in MEAM1 and to a certain level in MED-Q1. Cluster 2 was mainly overexpressed in ASIA II-1 and to some degree in NW2 and SSA1-SG3. Cluster 3 was highly overexpressed in UGANDA-MED-ASL and moderately overexpressed in MED-Q1. Cluster 7 grouped genes that were only downregulated in SSA1-SG3. Interestingly, within-species plastic expression related to the different diets (host-plants or

sucrose) was mostly found in cluster 4.

### 3.3 Functional analysis of enriched biological terms

Functional analysis of enriched biological terms was performed for each cluster using KEGG annotations (Figure 5b). Initial enrichment cutoff was set to  $P$  [?] 0.05. Clusters 1, 2, 3, 5, 6 and 8 harbored genes involved in “metabolism”, which were specifically upregulated in one or more species (see above). Genes involved in “genetic information processing” were enriched mainly in clusters 2 and 7. Those involved in “cellular processes” were enriched in clusters 4 and 7. Cluster 4 was also enriched in genes involved in “organismal systems” and “environmental information processing”.

The KEGG database grouped the genes involved in “metabolism” into 12 functional “subcategories” or “super-pathways” (<https://www.genome.jp/kegg/pathway.html>), including among others those responsible for the production/metabolism of amino acids, secondary metabolites and xenobiotics, carbohydrates, energy, lipids, nucleotides, glycans and co-factors/vitamins. Each “subcategory” contained several biochemical pathways, and 39 of them (Figure S1 and Table S3) were enriched in different clusters using a cutoff value of  $P$  [?] 0.05. Moreover, some pathways were enriched in more than one cluster but with different gene sets. In some cases, genes in the same KEGG pathway were up-regulated in one species and down regulated in another. The xenobiotics pathways: 00982, 00980 and 00983 were enriched in five clusters (Figure S1). Genes identified in these pathways mainly belonged to the *UDPGT* and *GST* gene families and were expressed differently between species. For example, a specific set of *UDPGT* and *GST* genes in cluster 8 was upregulated in SSA1-SG3 but downregulated in all other species. A different set of *UDPGT* and *GST* genes in cluster 6 was upregulated in MED-Q1 and downregulated in all other species. The amino acid metabolic pathways were nearly unique in each cluster (00300, 00330, 00220, 00260, 00340, 00280 and 00290). The only exception was the 00250 pathway (“alanine, aspartate and glutamate metabolism”) which was enriched in two clusters. Genes that are involved in carbohydrates metabolisms were enriched in five different KEGG pathways: 00053 (“ascorbate and aldarate metabolism”) and 00040 (“pentose and glucuronate interconversions”) were enriched in clusters 5, 6 and 8, 00052 (“galactose metabolism”) was enriched in cluster 1 and 6, 00500 (“starch and sucrose metabolism”) was enriched only in cluster 1 and 00520 (“amino sugar and nucleotide sugar metabolism”) only in cluster 5. The genes identified in these pathways were mainly those coding for UTP-glucose-1-phosphate uridylyltransferase, maltases and phosphatidylinositol and glycoside hydrolases. Other pathways that showed enrichment belonged to the “co-factors and vitamins metabolism” and lipid metabolism” super-pathways. KEGG pathway 00860 (“porphyrin and chlorophyll metabolism”) was enriched in clusters 2, 5, 6 and 8, while 00830 (“retinol metabolism”) was enriched in cluster 5, 6 and 8. KEGG pathways 00565 (“ether lipid metabolism”) and 00100 (“steroid biosynthesis”) were enriched in cluster 8, 00561 (“glycerolipid metabolism”) was enriched in clusters 1 and 3, while 00140 (“steroid hormone biosynthesis”) was enriched in clusters 5, 6 and 8.

### 3.4 Correlation between performance and expression plasticity

As mentioned above, Cluster 4 was the only one to group genes responding to the different diets (host plants or sucrose). The plastic genes showed enrichment in KEGG pathways from the super-pathways “organismal system” – “circulatory”, “development”, “digestive”, “endocrine”, “excretory”, “immune”, “nervous”, “sensory”, “aging”, “environmental adaptation”, “environmental information processing” – “membrane transport”, “signal transduction”, “signaling molecules and interaction”, “cellular processes” – “cellular community (eukaryotes)”. Spearman correlations between the mean expression levels of the enriched pathways in Cluster 4 (Figure 6) and the reproductive success score of the species/“Types” on each of the four host plants, indicated the existence of two distinctive correlation patterns. “Type I” and “II” species, capable of performing well on multiple hosts, largely showed constitutive expression (very low level of plasticity), which did not correlate with the species performance on the different hosts (-0.10 [?] Spearman’s  $r$  [?] 0.11,  $P$  [?] 0.79, Figure 6a and 6b). In contrast, “Type III” and “Type IV” species largely showed a plastic response, which seemed to be adaptive only to some extent in “Type III” and non-adaptive in “Type IV” (Figures 6c and 6d). This observation remains non-conclusive for the “Type III” species as the correlation between the mean expression level of the pathways and performance was not significant (Spearman’s  $r$  = -0.89,  $P$  = 0.10). On the contrary,

the mean expression level of the pathways was significantly negatively correlated with performance in the “Type IV” species (Spearman’s  $r = -0.80$ ,  $P = 0.016$ ).

Following this, we tested the enrichment in Cluster 4, of eight gene families, known to play a role in the utilization of ingested sugars as a nutritional resource, the maintenance of osmotic balance and the detoxification of phytotoxins. Among these eight groups, only two, sugar transporters and ABC transporters, showed a significant enrichment (25 out of a total of 76 sugar transporters and 24 out of a total of 47 ABC transporters, Table S4). Again, there was clear difference in the expression pattern of the two gene groups between the outlined four performance “Types”. “Type I” and “II” species constitutively overexpressed (standardized rld values [?] 1) ~10 sugar transporter and 1-2 ABC transporter genes. The “Type III” species SSA1-SG3 was similar to the “Type I” and “II” species by constitutively overexpressing high number of sugar and ABC transporter genes on three and four hosts (19 and 15, respectively) (Table S5), but differed from them by displaying also an inducible pattern of gene expression, mainly in the ABC transporters group (12 genes) in response to host switching to pepper and kale (Table S5). “Type IV” species were different from “Types I-III” and constitutively overexpressed only two sugar transporters and no ABC transporter genes (Table S5). Again, these species mainly displayed an inducible pattern of gene expression, as 19 and 13 sugar transporter and 18 and 15 ABC transporter genes were found to be overexpressed in NW-2 and UGANDA-MED-ASL, respectively, in response to a specific plant host (mainly cassava) (Table S5).

### 3.5 The unique metabolic tool box of each specie/performance type

To identify the unique metabolic tool box of each species/performance “Type”, we focused on enriched pathways that showed a conservative cutoff of  $P$  [?] 0.01. As indicated above, the phylogenetically (geographically) related species from the Mediterranean group presented similarity to some extent but it was clear that each species/“Type” has its unique metabolic toolbox (Figure 7). The “starch and sucrose metabolism” pathway was found to be overexpressed and uniquely enriched only in the “Type I” species MEAM1, suggesting enhanced production of metabolites (glucose, fructose, trehalose, trehalulose and oligosaccharides) that can be used in different metabolic pathways such as energy supply and storage but also in osmoregulation. MEAM1 also uniquely overexpressed genes involved in “glutathione metabolism”. Most of the overexpressed pathways, uniquely enriched in “Type II” species were found to be involved in amino acid biosynthesis and recycling of ammonia (“purine metabolism”, “pyrimidine metabolism”, “arginine biosynthesis” and “nitrogen metabolism”). These pathways provide an additional sources of energy production through purine and pyrimidine metabolism or the production of fumarate (intermediate in the citric acid cycle) during “arginine biosynthesis”. Species that are members of the “Type III” and “Type IV” groups, uniquely overexpressed gene clusters that were enriched in nine and ten pathways, respectively. Both “Types” showed enrichment in two carbohydrate metabolic pathways that are central to the conversion of glucose. The first, “ascorbate and aldarate metabolism” pathway forms UDP-glucose, which is then oxidized to UDP-glucuronic acid, and mainly used for the detoxification of toxic compounds through conjugation (Chen, Xia, Shang, Song, & Gao, 2019; Jiang et al., 2019). The second is the “pentose and glucuronate interconversions” pathway which generates NADPH and pentoses, mainly for maintaining/regenerating the cellular detoxifying and antioxidative defence systems (Agedal, Niere, & Ziegler, 2010) and the synthesis of nucleotides and nucleic acids (Ceddia, Bikopoulos, Hilliker, & Sweeney, 2003), respectively. The “Type III” species SSA1-SG3, was the only one showing enrichment of over-expressed genes involved in the metabolism of cyano amino acids, amino acid derivatives that contain a cyanide group. As indicated above, SSA1-SG3 showed specific adaptation to cassava plants that produce cyanogenic glucosides (Alves, 2002). “Type IV” species uniquely showed enrichment of overexpressed genes in the “glycine, serine and threonine metabolism” pathway, which are essential for energy homeostasis and amino acid metabolism through the citric acid cycle. This pathway was previously shown to be involved in cold and hypoxia tolerance in insects (Cui, Wang, Qiu, Liu, & Geng, 2017; Y. Li et al., 2015). Species from “Types II, III and IV” showed enrichment of overexpressed genes in several pathways within the “xenobiotics biodegradation and metabolism” super-pathway. These pathways are mostly involved in the detoxifying of phytotoxins and/or free radicals.

## 4. DISCUSSION

The development of a large set of genomic approaches in the last decade, has enabled a remarkable growth in our understanding of the molecular mechanisms that facilitates host plant use by generalist and specialist herbivorous insects and how these mechanisms facilitate or constrain host shifts (Birnbaum & Abbot, 2020). Generalists are considered to harbor "general purpose genomes", with variation that is persistently 'tested by selection' on alternative plants. Expression plasticity may be abundant, and generally adaptive when challenged by new stressors (Birnbaum, Rinker, Gerardo, & Abbot, 2017; Schweizer, Heidel-Fischer, Vogel, & Reymond, 2017), facilitating broad niche occupancy. Specialists, on the other hand, may have "specialized genomes", characterized by genomes that are 'untested' in alternative environments (Huang, Tran, & Agrawal, 2016). Most genes are expected to be constitutively expressed, and because the specialist can be considered 'naive' to new environments, initially non-adaptive patterns of plastic expression is expected when new stressors are encountered (Birnbaum & Abbot, 2020).

It is quite clear however, that these observations/predictions are likely to apply more to generalists chewing and mesophyll feeders that show large transcriptional changes during host switching (e.g., De La Paz Celorio-Mancera et al., 2013; Dermauw et al., 2013; Govind et al., 2010; Grbić et al., 2011; Roy et al., 2016; Vogel, Badapanda, Knorr, & Vilcinskas, 2014; Zhurov et al., 2014). In contrast, the early transcriptomic responses (24 h) to diet shifts of the six *B. tabaci* species analysed here, were mainly characterized by a constitutive "non-changed" expression. Less than 15% of the genes showed overall plasticity even when the insects experienced a novel host plant like cassava. Moreover, the gene expression patterns obtained in cluster 4 (the only cluster showing significant within-species plastic expression related to the different host plants) supported the above predictions only to a certain degree. "Type I" and "II" *B. tabaci* species, capable of performing well on multiple hosts, largely showed constitutive expression also in cluster 4, suggesting the presence of non-plastic "general purpose genomes" that provide in most cases sufficient standing ability to utilize new hosts. In contrast, "Type III" and "Type IV" species largely showed a plastic response, which seemed to be adaptive only to some extent in "Type III" and non-adaptive in "Type IV". This raises the possibility that "Type IV" species harbor "more specialized" genomes that present general plasticity that was not tested by selection, making it in many cases non-adaptive. Limited transcriptional responses, after transfer for 3 h and 24 h to a host plant, a non-host plant or an artificial diet, were also reported in the generalist aphid *M. persicae* and the cereal aphid *Rhopalosiphum padi* (Thorpe, Escudero-Martinez, Cock, Eves-van den Akker, & Bos, 2018). In addition, these findings reinforce similar ones obtained in a previous study on the generalist aphid *M. persicae*, where colonies that were reared in parallel for one year on *Brassica rapa* or *Nicotiana benthamiana* were compared. The colonies' transcriptomes differed by only 171 differentially expressed genes from a total of >18,000 genes (<1%) (Mathers et al., 2017). Taken together, these data suggest that phloem-feeding insects might differ from other herbivorous insects feeding guilds both in their early and late transcriptional responses to new hosts. The main differences might be related to the different challenges faced upon host switching. As already indicated, the involvement of plant defence compounds in determining the fate of insect-plant interactions in generalist sap-feeders are more questioned and controversial than in generalist chewers. For example, a wide variety of plant defence compounds are stored as glucoside conjugates. Only after the sugar is hydrolyzed by  $\beta$ -glucosidase enzymes upon tissue damage, the released active aglucones are further modified to form substances that are toxic or deterrent to herbivores (Halkier & Gershenzon, 2006; Morant et al., 2008). Our knowledge of how sap-feeding insects circumvent the activation or toxic effects of activated compounds is limited but it is generally believed that these insects do not damage the tissue extensively enough to cause the mixing of plant  $\beta$ -glucosidases and glucosylated defence compounds, and hence the formation of the active aglucones (Walling, 2008).

On the other hand, phloem-feeding insects face a unique challenge when feeding on phloem sap that contains an abundance of simple sugars, few essential amino acids as well as other nutrients and relatively low concentration of plant secondary metabolites (Douglas, 2003). Therefore, it is not surprising that many of the metabolic pathways that were enriched and overexpressed in the analyzed *B. tabaci* species relate to the metabolism of carbohydrates and amino acids. Moreover, although the concentration of sugars and the osmotic pressure of the phloem sap varies among plants, the osmotic pressure is always higher in the phloem sap than in the phloem-feeding insect (Douglas, 2003). As a result, it can be assumed that that

the mechanisms that phloem feeders are using for sugars homeostasis have an effect on their host plant acceptance. For example, it can be hypothesized that phloem-feeding species that are able to achieve perfect homeostasis in sugar concentrations, independently of the external sugar concentrations that are present in a range of host plants, are the ones that should be considered as more generalist species. Our division of the *B. tabaci* species complex into four performance/gene expression “Types” supports this idea, as the number of constitutively overexpressed sugar transporters was 10, 10 and 19 in “Type I, II and III” species and only two in the “Type IV” species. Moreover, in the “Type IV” species (NW-2 and MED-ASL), the sugar homeostasis was more affected by the identity of the host plant (more than 90% of the sugar transporters were induced in response to the novel host plant compared to less than 50% in species belonging to “Types I, II and III”). Transporter-mediated uptake of dietary sugar was previously shown to be essential in the biology of phloem-feeding insects. For example, their silencing in the brown planthopper (BPH) *Nilaparvata lugens*, significantly interfered with the insects’ longevity, reproduction, development and viability (Ge et al., 2015; Kikuta, Hagiwara-Komoda, Noda, & Kikawada, 2012). In addition, the *Acyrtosiphon pisum* genome was found to contain a relatively large number of genes encoding predicted sugar transporters, likely resulting from recent gene duplications that allow the efficient handling of high sugar concentrations in the insects’ gut (D. R. G. Price et al., 2010). Support to the possible link between variation in sugar transporters expression and diet acceptance also comes from parallel experimental systems. A recent study on *Drosophila melanogaster*, which feeds on rotting fruits that contain sugar and yeast (Markow, 2015), brought evidence that whole body depletion of a sugar transporter highly expressed in the fly midgut, results in lethality after three days on a high sugar diet, and is accompanied by a characteristic starvation phenotype (Francis et al., 2020) The authors conclude that sugar transporters mediate their effects on starvation resistance by regulating glucose metabolism, raising the possibility that expression levels of sugar transporters might also be involved in determining if a plant will be considered as a host or non-host (causing starvation) by phloem-feeding species.

It appears that our division of the *B. tabaci* species-complex into four performance/gene expression “Types” can be generalized and applied to other hemipteran systems, taking the well-studied aphid superfamily as our main comparative model. For example, the aphid species *M. persicae*, which is globally distributed and presents a host range of > 400 plant species (Blackman & Eastop, 2000), and *R. padi*, which is also distributed world-wide and feeds on a wide range of grasses (Dixon, 1971), are likely to be aphid representatives of “Type I” and “Type II” species. As already indicated above, both species showed extremely limited transcriptional response (< 10 genes) in the first 24 h upon transfer to a host, nonhost plant or artificial diet treatments (Thorpe et al., 2018) (Thorpe et al. 2018). The scale *Paratachardina pseudolobata* might be another good example of a hemipteran “Type I” generalist, as it has been recorded on more than 300 host plant species and was reported to be invasive widespread. Gene expression analysis of *P. pseudolobata* on three different host plants revealed that only 2.5% of the genes were differentially expressed across all comparisons. Moreover, of the 1196 putative detoxification genes that were expressed, only 23 (1.9%) were significantly overexpressed on any one host (Christodoulides et al. 2017).

Species belonging to “Type III” show transcriptomic characteristics with similarities both to “Types I and II” and to “Type IV” species, and are unique in showing specialization to a novel and well defended host plant. Within the *B. tabaci* complex, the best example is the adaptation of species within the SSA group to cassava (Berry et al., 2004). Outside the *B. tabaci* complex, the best hemipteran example might be the adaptation of *M. persicae* to tobacco (*Nicotiana tabacum*) that led to a formation of a sub-species, *M. persicae nicotianae* (Blackman, 1987). Adaptation to tobacco is driven by constitutive overexpression of the P450 gene *CYP6CY3*, which allows the tobacco-adapted insects to efficiently detoxify nicotine (Bass et al., 2013). Recently, nicotine tolerance in *M. persicae nicotianae* was also associated with overexpression of *UDPGT* genes (Pan, Xu, Zeng, Liu, & Shang, 2019). It is important to note that both the cassava-adapted species of *B. tabaci* and the tobacco-adapted species of *M. persicae* can still utilize other plants as host, but seem to show reduced performance levels (Malka et al. 2018, Nikolakakis, Margaritopoulos, & Tsitsipis, 2003). One possible explanation to our findings that “Type III” species share transcriptomic and performance similarities with both “Types I and II” and “Type IV” species, is the possibility that “Type III”

species are in a transformation phase of acquiring a novel host into their normal host range, a process which seems important/advantageous to the general fitness of the species (Nylin & Janz, 2009). As a result, these species are likely undergoing genetic accommodation and the acquisition of quantitative genetic changes that can either increase or decrease the expression level and environmental responsiveness of various gene families, which is likely to affect their performance on some hosts (Levis & Pfennig, 2016).

Species belonging to “Type IV” differ from the three “Types” above mainly by showing a much “weaker” generalist habit. They are still likely to be capable of utilizing many plants as hosts but perform poorly on most of them. The only *B. tabaci* species we analyzed that perfectly matched this description is NW-2. UGANDA-MED-ASL showed many of the outlined characteristics, but seemed to be capable of performing well on some hosts, suggesting larger variation in the definition of this performance “Type” (Vyskočilová, Seal, & Colvin, 2019). In a broader sense, these two species might be the *B. tabaci* versions of host-specialized species described mainly in the aphid lineage. For example, the cotton-melon aphid, *Aphis gossypii*, might be considered as a “Type IV” species. It has a wide host range of hundreds of plant species belonging to various families such as Cucurbitaceae, Malvaceae, Solanaceae, Rutaceae, and Asteraceae (Blackman & Eastop, 2000), but many *A. gossypii* populations were found to form more specialized strains that utilize only a subset of host plant species in their recorded host range (Ma et al., 2019). Comparative analysis of gene expression of three populations of *A. gossypii*: cotton-specialized, cucurbit-specialized, and a cucurbit-specialized reared on cowpea, revealed two very different expression patterns. Only 3941 genes from 38,398 (10.3%) were differentially expressed between the cotton- and cucurbit-specialized populations feeding on their home plant, mostly associated with sugar metabolism, immune, antioxidative and detoxification systems and salivary secretions. However, more than 15,000 genes (> 40%) were found to be differentially expressed when these two populations were compared to the population feeding on the novel host cowpea. This time, the differentially expressed genes were mostly enriched in the KEGG super-pathways “organismal systems” and “cellular processes”, indicating a significant stress-response (Zhang, Lei, Miao, & Liu, 2017). It is interesting to note that this response was quite similar in its characteristics to the non-adaptive plastic response that “Type IV” *B. tabaci* species presented when switching to feed on a novel/non-suitable host plants (Figure 6d), raising the possibility for the existence of an hemipteran common “stress-response tool-box”.

In conclusion, analysis of 87 transcriptomes belonging to six species of *B. tabaci* feeding on four plant hosts and a sucrose-only artificial diet, clearly indicated that many of the enriched and overexpressed pathways that differ between the species relate to the transport and metabolism of sugars and the synthesis of a range of amino acids, and less to pathways involved in the detoxification of phytotoxins. Still, our current knowledge is quite limited. Many more comparative studies, integrating transcriptomic and fitness metrics in different hemipteran systems are required, before any general micro- and/or macro-evolutionary argument can be made on the evolution of diet breadth in generalist and specialist phloem-feeding species. Primary metabolism constrains and osmoregulation/dehydration challenges might play a significant role in the guild interaction with its host plant. We encourage future research to focus on the interface between primary and secondary metabolism as a promising essential tool for achieving better understanding of the regulatory networks that controls carbon flux in phloem-feeding insects, likely uncovering the lineage unique diversification processes.

## ACKNOWLEDGEMENTS

This work was supported by the Israel Science Foundation grant 1039/12 to SM and the Natural Resources Institute, University of Greenwich from a grant provided by the Bill & Melinda Gates Foundation (Grant Agreement Number OPP1058938).

## DATA ACCESSIBILITY

Mitochondrial cytochrome oxidase I (mtCOI) DNA sequences of the six species and transcriptome sequence reads of the 72 samples will be deposited in the European Nucleotide Archive under the study accession number PRJEB21948.

## AUTHORS’ CONTRIBUTIONS

The authors contributed in the following manner: conceptualization (OM, EF, SS, JC, SM); provision of insects, genetic materials and facilities (OM, SVB, EF, SS, JC, SM); funding acquisition (SM, JC); data acquisition and formal analysis (OM, EF, SVP, DSG, SM); method development (OM, JC, SM); OM and SM drafted the manuscript with input from EF, JC, SS, SVB.

## REFERENCES

- Abisgold, J. D., Simpson, S. J., & Douglas, A. E. (1994). Nutrient regulation in the pea aphid *Acyrtosiphon pisum* : application of a novel geometric framework to sugar and amino acid consumption. *Physiological Entomology* , 19 (2), 95–102.
- Agledal, L., Niere, M., & Ziegler, M. (2010). The phosphate makes a difference: cellular functions of NADP. *Redox Report* , 15 (1) 2–10.
- Ali, J. G., & Agrawal, A. A. (2012). Specialist versus generalist insect herbivores and plant defense. *Trends in Plant Science* , 17 (5), 293–302.
- Barbosa, L. F., Marubayashi, J. M., De Marchi, B. R., Yuki, V. A., Pavan, M. A., Moriones, E., ... Krause-Sakate, R. (2014). Indigenous American species of the *Bemisia tabaci* complex are still widespread in the Americas. *Pest Management Science* , 70 (10), 1440–1445.
- Bass, C., Zimmer, C. T., Riveron, J. M., Wilding, C. S., Wondji, C. S., Kausmann, M., ... Nauen, R. (2013). Gene amplification and microsatellite polymorphism underlie a recent insect host shift. *Proceedings of the National Academy of Sciences of the United States of America* , 110 (48), 19460–19465.
- Behmer, S. T. (2009). Insect Herbivore Nutrient Regulation. *Annual Review of Entomology* , 54 (1), 165–187.
- Bernays, E. A. (1998). Evolution of insect feeding behavior in herbivores: success seen as different ways to eat without being eaten. *BioScience* , 48 (1), 35–44.
- Berry, S. D., Fondong, V. N., Rey, C., Rogan, D., Fauquet, C. M., & Brown, J. K. (2004). Molecular Evidence for Five Distinct *Bemisia tabaci* (Homoptera: Aleyrodidae) Geographic Haplotypes Associated with Cassava Plants in Sub-Saharan Africa. *Annals of the Entomological Society of America* , 97, 852–859.
- Birnbaum, S. S. L., & Abbot, P. (2020, March 1). Gene expression and diet breadth in plant-feeding insects: summarizing trends. *Trends in Ecology and Evolution* , Vol. 35, pp. 259–277.
- Birnbaum, S. S. L., Rinker, D. C., Gerardo, N. M., & Abbot, P. (2017). Transcriptional profile and differential fitness in a specialist milkweed insect across host plants varying in toxicity. *Molecular Ecology* , 26 (23), 6742–6761.
- Blackman, R. L. (1987). Morphological discrimination of a tobacco-feeding form from *Myzus persicae* (Sulzer) (Homoptera: Aphididae), and a key to New World *Myzus* (Nectarosiphori) species. *Bulletin of Entomological Research* , 77 (4), 713–730.
- Blackman, R. L., & Eastop, V. F. (2000). *Aphids on the World's Crops, An Identification and Information Guide* , 2nd edn. John Wiley & Sons, Chichester, UK.
- Ceddia, R. B., Bikopoulos, G. J., Hilliker, A. J., & Sweeney, G. (2003). Insulin stimulates glucose metabolism via the pentose phosphate pathway in *Drosophila* Kc cells. *FEBS Letter* , 555, 307–310.
- Chen, X., Xia, J., Shang, Q., Song, D., & Gao, X. (2019). UDP-glucosyltransferases potentially contribute to imidacloprid resistance in *Aphis gossypii* glover based on transcriptomic and proteomic analyses. *Pesticide Biochemistry and Physiology* , 159 , 98–106.
- Chiluwal, K., Roh, G. H., Kim, J., & Park, C. G. (2020). Acaricidal activity of the aggregation pheromone of Japanese pine sawyer against two-spotted spider mite. *Journal of Asia-Pacific Entomology* , 23 (1), 86–90.
- Clarke, A. R. (2017). Why so many polyphagous fruit flies (Diptera: Tephritidae)? A further contribution to the “generalism” debate. *Biological Journal of the Linnean Society* , 120, 245–257.

- Cornell, H. V., & Hawkins, B. A. (2003). Herbivore responses to plant secondary compounds: A test of phytochemical coevolution theory. *American Naturalist* , 161 (4), 507–522.
- Cui, S., Wang, L., Qiu, J., Liu, Z., & Geng, X. (2017). Comparative metabolomics analysis of *Callosobruchus chinensis* larvae under hypoxia, hypoxia/hypercapnia and normoxia. *Pest Management Science* , 73 (6), 1267–1276.
- De Barro, P. J., Liu, S.-S., Boykin, L. M., & Dinsdale, A. B. (2011). *Bemisia tabaci* : A Statement of Species Status . *Annual Review of Entomology* , 56 (1), 1–19.
- De La Paz Celorio-Mancera, M., Wheat, C. W., Vogel, H., Soderlind, L., Janz, N., & Nylin, S. (2013). Mechanisms of macroevolution: Polyphagous plasticity in butterfly larvae revealed by RNA-Seq. *Molecular Ecology* , 22 (19), 4884–4895.
- Deans, C. A., Behmer, S. T., Tessnow, A. E., Tamez-Guerra, P., Pusztai-Carey, M., & Sword, G. A. (2017). Nutrition affects insect susceptibility to Bt toxins. *Scientific Reports* , 7 .
- Dermauw, W., Wybouw, N., Rombauts, S., Menten, B., Vontas, J., Grbic, M., ... Van Leeuwen, T. (2013). A link between host plant adaptation and pesticide resistance in the polyphagous spider mite *Tetranychus urticae*. *Proceedings of the National Academy of Sciences* , 110 (2), E113–E122.
- Dermauw, Wannes, Pym, A., Bass, C., Van Leeuwen, T., & Feyereisen, R. (2018). Does host plant adaptation lead to pesticide resistance in generalist herbivores? *Current Opinion in Insect Science* , 26 , 25–33. doi: 10.1016/j.cois.2018.01.001
- Despres, L., David, J.P., & Gallet, C. (2007). The evolutionary ecology of insect resistance to plant chemicals. *Trends in Ecology & Evolution (Personal Edition)* , 22 (6), 298–307.
- Dixon, A. F. G. (1971). The life-cycle and host preferences of the bird cherry-oat aphid, *Rhopalosiphum padi* L. , and their bearing on the theories of host alternation in aphids. *Annals of Applied Biology* , 68 (2), 135–147.
- Douglas, A. E. (2003). The Nutritional Physiology of Aphids. *Advances in Insect Physiology* , Vol. 31, pp. 73–140.
- Engler, H. S., Spencer, K. C., & Gilbert, L. E. (2000). Insect metabolism: Preventing cyanide release from leaves. *Nature* , 406 (6792), 144–145.
- Francis, D., Ghazanfar, S., Havula, E., Krycer, J. R., Senior, A., Minard, A. Y., ... James, D. E. (2020). Genome wide analysis in *Drosophila* reveals diet by gene interactions and uncovers diet-responsive genes. *BioRxiv* , 718304. doi: 10.1101/718304
- Ge, L.-Q., Jiang, Y.-P., Xia, T., Song, Q.-S., Stanley, D., Kuai, P., ... Wu, J.-C. (2015). Silencing a sugar transporter gene reduces growth and fecundity in the brown planthopper, *Nilaparvata lugens* (Stal) (Hemiptera: Delphacidae). *Scientific 729 reports*, 5 , 12194.
- Govind, G., Mittapalli, O., Griebel, T., Allmann, S., Bocker, S., & Baldwin, I. T. (2010). Unbiased transcriptional comparisons of generalist and specialist herbivores feeding on progressively defenseless *Nicotiana attenuata* plants. *PLoS ONE* , 5 (1), e8735.
- Grbić, M., Van Leeuwen, T., Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., ... Van De Peer, Y. (2011). The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* , 479 (7374), 487–492.
- Halkier, B. A., & Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annual Review of Plant Biology* , 57 , 303–333.
- Heidel-Fischer, H. M., & Vogel, H. (2015, April 1). Molecular mechanisms of insect adaptation to plant secondary compounds. *Current Opinion in Insect Science* , 8, 8–14.

- Hu, J., Zhang, X., Jiang, Z., Zhang, F., Liu, Y., Li, Z., ... Zhang, Z. (2018). New putative cryptic species detection and genetic network analysis of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in China based on mitochondrial COI sequences *Mitochondrial DNA A DNA Mapp Seq Anal* . 29(3), 474–484.
- Huang, Y., Tran, I., & Agrawal, A. F. (2016). Does genetic variation maintained by environmental heterogeneity facilitate adaptation to novel selection? *American Naturalist* , 188 (1), 27–37.
- Jaenike, J. (1990). Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics* , 21 (1), 243–273.
- Jiang, T., Chen, Y., Tan, Z., Li, J., Qian, P., Tang, S., & Shen, X. (2019). Expression analysis and functional identification of several genes related to diapause in *Bombyx mori*. *Development Growth and Differentiation* , 61 (2), 150–157.
- Kikuta, S., Hagiwara-Komoda, Y., Noda, H., & Kikawada, T. (2012). A novel member of the trehalose transporter family functions as an H<sup>+</sup>-dependent trehalose transporter in the reabsorption of trehalose in Malpighian tubules. *Frontiers in Physiology* , 3 , 1–9.
- Kikuta, S., Kikawada, T., Hagiwara-Komoda, Y., Nakashima, N., & Noda, H. (2010). Sugar transporter genes of the brown planthopper, *Nilaparvata lugens* : A facilitated glucose/fructose transporter. *Insect Biochemistry and Molecular Biology* , 40 (11), 805–813.
- Kim, J. H., & Jander, G. (2007). *Myzus persicae* (green peach aphid) feeding on *Arabidopsis* induces the formation of a deterrent indole glucosinolate. *Plant Journal* , 49 , 1008–1019.
- Kimura, A., Yoshida-Kitahara, F., & Chiba, S. (1987). Characteristics of transglucosylation of honeybee  $\alpha$ -Glucosidase I. *Agricultural and Biological Chemistry* , 51 (7), 1859–1864.
- Kwang, P. L., Behmer, S. T., & Simpson, S. J. (2006). Nutrient regulation in relation to diet breadth: A comparison of *Heliothis* sister species and a hybrid. *Journal of Experimental Biology* , 209 (11), 2078–2084.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* , 9 (4), 357–359.
- Lee, K. P., Simpson, S. J., & Raubenheimer, D. (2004). A comparison of nutrient regulation between solitary and gregarious phases of the specialist caterpillar, *Spodoptera exempta* (Walker). *Journal of Insect Physiology* , 50 (12), 1171–1180.
- Levis, N. A., & Pfennig, D. W. (2016). Evaluating “plasticity-first” evolution in nature: key criteria and empirical approaches. *Trends in Ecology and Evolution* , 31, 563–574.
- Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* , 12 (1), 323.
- Li, Y., Zhang, L., Chen, H., Košťál, V., Simek, P., Moos, M., & Denlinger, D. L. (2015). Shifts in metabolomic profiles of the parasitoid *Nasonia vitripennis* associated with elevated cold tolerance induced by the parasitoid’s diapause, host diapause and host diet augmented with proline. *Insect Biochemistry and Molecular Biology* , 63 , 34–46.
- Loxdale, H. D., & Harvey, J. A. (2016). The ‘generalism’ debate: misinterpreting the term in the empirical literature focusing on dietary breadth in insects. *Biological Journal of the Linnean Society* , 119 (2), 265–282.
- Loxdale, H. D., Lushai, G., & Harvey, J. A. (2011). The evolutionary improbability of “generalism” in nature, with special reference to insects. *Biological Journal of the Linnean Society* , 103 (1), 1–18.
- Loxdale, H. D., Schöfl, G., Wiesner, K. R., Nyabuga, F. N., Heckel, D. G., & Weisser, W. W. (2011). Stay at home aphids: Comparative spatial and seasonal metapopulation structure and dynamics of two specialist tansy aphid species studied using microsatellite markers. *Biological Journal of the Linnean Society* , 104 (4), 838–865.

- Ma, L., Li, M. Y., Chang, C. Y., Chen, F. F., Hu, Y., & Liu, X. D. (2019). The host range of *Aphis gossypii* is dependent on aphid genetic background and feeding experience. *PeerJ* , 2019 (9), e7774.
- Malka, O., Santos-Garcia, D., Feldmesser, E., Sharon, E., Krause-Sakate, R., Delatte, H., . . . Morin, S. (2018). Species-complex diversification and host-plant associations in *Bemisia tabaci*: A plant-defence, detoxification perspective revealed by RNA-Seq analyses. *Molecular Ecology* , 27 (21), 4241–4256.
- Markow, T. A. (2015). The secret lives of *Drosophila* flies. *ELife* , 4 (6), 273–289.
- Mathers, T. C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S. T., Baa-Puyoulet, P., . . . Hogenhout, S. A. (2017). Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biology* , 18 (1), 1–20.
- Merkx-Jacques, M., Despland, E., & Bede, J. C. (2008). Nutrient utilization by caterpillars of the generalist beet armyworm, *Spodoptera exigua* . *Physiological Entomology* , 33 (1), 51–61.
- Morant, A. V., Jørgensen, K., Jørgensen, C., Paquette, S. M., Sánchez-Pérez, R., Møller, B. L., & Bak, S. (2008).  $\beta$ -Glucosidases as detonators of plant chemical defense. *Phytochemistry* , 69, 1795–1813.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* , 35, W182– W185.
- Nikolakakis, N. N., Margaritopoulos, J. T., & Tsitsipis, J. A. (2003). Performance of *Myzus persicae* (Hemiptera: Aphididae) clones on different host-plants and their host preference . *Bulletin of Entomological Research* , 93 (3), 235–242.
- Nylin, S., & Janz, N. (2009). Butterfly host plant range: An example of plasticity as a promoter of speciation? *Evolutionary Ecology* , 23, 137– 146.
- Pan, Y., Xu, P., Zeng, X., Liu, X., & Shang, Q. (2019). Characterization of UDP-Glucuronosyltransferases and the Potential Contribution to Nicotine Tolerance in *Myzus persicae*. *International Journal of Molecular Sciences* , 20 (15), 3637.
- Price, D. R. G., Tibbles, K., Shigenobu, S., Smertenko, A., Russell, C. W., Douglas, A. E., . . . Gatehouse, J. A. (2010). Sugar transporters of the major facilitator superfamily in aphids; from gene prediction to functional characterization. *Insect Molecular Biology* , 19 , 97–112.
- Price, Daniel R.G., Wilkinson, H. S., & Gatehouse, J. A. (2007). Functional expression and characterisation of a gut facilitative glucose transporter, NIHT1, from the phloem-feeding insect *Nilaparvata lugens* (rice brown planthopper). *Insect Biochemistry and Molecular Biology* , 37(11), 1138–1148
- Ratzka, A., Vogel, H., Kliebenstein, D. J., Mitchell-Olds, T., & Kroymann, J. (2002). Disarming the mustard oil bomb. *Proceedings of the National Academy of Sciences of the United States of America* ,99 (17), 11223–11228.
- Raubenheimer, D., & Simpson, S. J. (1999). Integrating nutrition: A geometrical approach. *Entomologia Experimentalis et Applicata* ,91 (1), 67–82.
- Raubenheimer, D., & Simpson, S. J. (2003). Nutrient balancing in grasshoppers: Behavioural and physiological correlates of dietary breadth. *Journal of Experimental Biology* , 206 (10), 1669–1681.
- Roy, A., Walker, W. B., Vogel, H., Chattington, S., Larsson, M. C., Anderson, P., . . . Schlyter, F. (2016). Diet dependent metabolic responses in three generalist insect herbivores *Spodoptera* spp. *Insect Biochemistry and Molecular Biology* , 71 , 91–105.
- Sasabe, M., Wen, Z., Berenbaum, M. R., & Schuler, M. A. (2004). Molecular analysis of CYP321A1, a novel cytochrome P450 involved in metabolism of plant allelochemicals (furanocoumarins) and insecticides (cypermethrin) in *Helicoverpa zea*. *Gene* , 338(2), 163–75.

Schweizer, F., Vogel, H., & Reymond, P. (2017). Arabidopsis glucosinolates trigger a contrasting transcriptional response in a generalist and a specialist herbivore. *Insect Biochemistry and Molecular Biology* , 85 , 21– 31.

Simpson, S. J., & Raubenheimer, D. (1995). The geometric analysis of feeding and nutrition: a user's guide. *Journal of Insect Physiology* , 41(7) , 545–553

Stevenson, P. C., Nicolson, S. W., & Wright, G. A. (2017). Plant secondary metabolites in nectar: impacts on pollinators and ecological functions. *Functional Ecology* , 31 (1), 65–75.

Thompson, S. N., & Redak, R. A. (2005). Feeding behaviour and nutrient selection in an insect *Manduca sexta* L . and alterations induced by parasitism. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* ,191 (10), 909–923.

Thorpe, P., Escudero-Martinez, C. M., Cock, P. J. A., Eves-van den Akker, S., & Bos, J. I. B. (2018). Shared Transcriptional Control and Disparate Gain and Loss of Aphid Parasitism Genes. *Genome Biology and Evolution* , 10 (10), 2716–2733.

Vogel, H., Badapanda, C., Knorr, E., & Vilcinskas, A. (2014). RNA-sequencing analysis reveals abundant developmental stage-specific and immunity-related genes in the pollen beetle *Meligethes aeneus* . *Insect Molecular Biology* , 23 (1), 98–112.

Vogel, Heiko, Musser, R. O., & de la Paz Celorio-Mancera, M. (2014). Transcriptome responses in herbivorous insects towards host plant and toxin feeding. *Annual Plant Reviews online*, 47, 197–233.

Vyskočilová, S., Seal, S., & Colvin, J. (2019). Relative polyphagy of “Mediterranean” cryptic Bemisia tabaci whitefly species and global pest status implications. *Journal of Pest Science* , 92 (3), 1071–1088.

Walling, L. L. (2008). Avoiding effective defenses: strategies employed by phloem-feeding insects. *Plant Physiology* , 146 (3), 859–866.

Zhang, Y.-C., Lei, H.-X., Miao, N.-H., & Liu, X.-D. (2017). Comparative transcriptional analysis of the host-specialized aphids *Aphis gossypii* (Hemiptera: Aphididae). *Journal of Economic Entomology* ,110 (2), 702–710.

Zhurov, V., Navarro, M., Bruinsma, K. A., Arbona, V., Estrella Santamaria, M., Cazaux, M., ... Grbić, V. (2014). Reciprocal responses in the interaction between arabidopsis and the cell-content-feeding chelicerate herbivore spider mite. *Plant Physiology* , 164 (1), 384–399.

## SUPPORTING INFORMATION

**Table S1** . The reproductive success of each species on the four host plants.

**Table S2**. Standardized rld values for each gene in each sample. Genes were divided into eight clusters by K-means clustering according to their expression pattern.

**Table S3**. Functional analysis of enriched KEGG pathways categories in eight co-expressed clusters. The clusters are presented in Figures 5a and S1.

**Table S4**. The distribution and standardized rld values of ABC transporters and sugar transporters (STs) across the eight co-expressed clusters (rld data from Table S2).

**Table S5**. The overexpression patterns of sugar transporters and ABC transporters across the four performance types.

**Figure S1**. Significantly enriched biological functions (hypergeometric test,  $P < 0.05$ ) related to ‘metabolism’ in the eight gene clusters displayed in Figure 5a. Color-cluster associations are the same as in Figure 5a and 5b.

## Figure captions

**Figure 1.** Experimental setup. Two-hundred newly emerged adults from each of the six species, grown on eggplant, were subjected to a feeding period of 72 hr on 10% sucrose diet, to obtain a standardized genes expression pattern. The sucrose diet lids were then replaced by clip cage lids and the adults were transferred, for a feeding period of 24 hr, to the four experimental host plants. Control adults were subjected to additional 24 hr of feeding on 10% sucrose diet in new sucrose lids. Adults were then collected for RNASeq analysis.

**Figure 2.** The estimated reproductive success of the six species on the four plant hosts. The  $F_1/F_0$  scoring scale represents the percentage-ratio between the number of emerging adults in the  $F_1$  generation and the number of initiating  $F_0$  adults that were transferred to the plants.

**Figure 3.** The size and variety of plant bouquets illustrate the “polyphagy level” of the four performance “Types”, based on the reproductive success results presented in Figure 2. MEAM1 (“Type I”), MED-Q1 & Asia II-1 (“Type II”), SSA1-SG3 (“Type III”) and Uganda MED-ASL & NW2 (“Type IV”).

**Figure 4.** Differential expression of 27417 genes in adults of the six analyzed *Bemisia tabaci* species after feeding for 24 hr on 10% sucrose, the common host plant eggplant or the novel host plants: kale, cassava and pepper. (a) Principal components analysis (PCA) was performed on the data from the 27417 differentially expressed genes across the 30 insect species — diet combinations. The graph shows principal component 1 vs. principal component 2 and 3 values for each combination (three biological replicates) in the study. (b) Hierarchical clustering, genes are represented in rows and samples are shown in columns. Red represents upregulation and blue represents downregulation (standardization was made on rld values for each gene across all 30 species and diet combinations). The PCA and Hierarchical clustering analyses were performed using Partek Genomics Suite software with Pearson’s dissimilarity distance and complete linkage methods.

**Figure 5.** Gene Expression Profiles of six species of *B. tabaci* (ASIA II-1, MEAM1, NW2, MED-Q1, SSA1-SG3, UGANDA-MED-ASL) on four host plants (cassava, eggplant, kale, pepper) and 10% sucrose diet. (a) K-means clustering. Red represents upregulation and blue represents downregulation (standardization was made on rld values for each gene across all 30 species and diet combinations). (b) Significantly enriched KEGG pathways (hypergeometric test,  $P [?] 0.05$ ) in the eight gene clusters displayed in Figure 5a. Color-cluster associations are the same as in Figure 5a.

**Figure 6.** Correlation (Spearman’s  $r$ , global significance  $P [?] 0.05$ ) between the performance score on the four host plants and the mean standardized rld expression values of the genes in cluster 4 that were associated with the following enriched KEGG super-pathways: “organismal system”, “environmental information processing” and “cellular processes” (see more details in Table S3). (a) “Type I” (MEAM1), (b) “Type II” (Asia II-1 & MED-Q1), (c) “Type III” (SSA1-SG3), (d) “Type IV” (UGANDA MED-ASL & NW2). Eggplant (\*), kale (), pepper () and cassava(). Color – cluster associations are the same as in Figure 5a.

**Figure 7.** The unique metabolic “tool-box” of each specie/performance “Type”. Significantly enriched KEGG pathways (hypergeometric test,  $P [?] 0.01$ ) in clusters associated with the “metabolism” category in Figure 5a. Green, yellow, blue and red colors indicate enriched pathways in “Type I”, “Type II”, “Type III” and “Type IV” performance groups, respectively.

**Figure 1.**

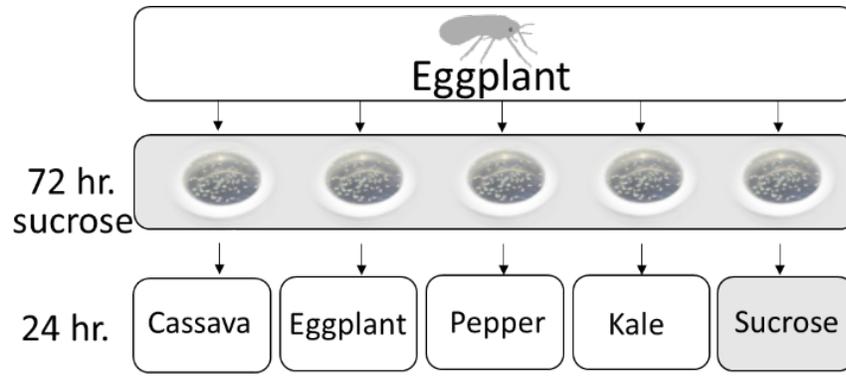


Figure 2.

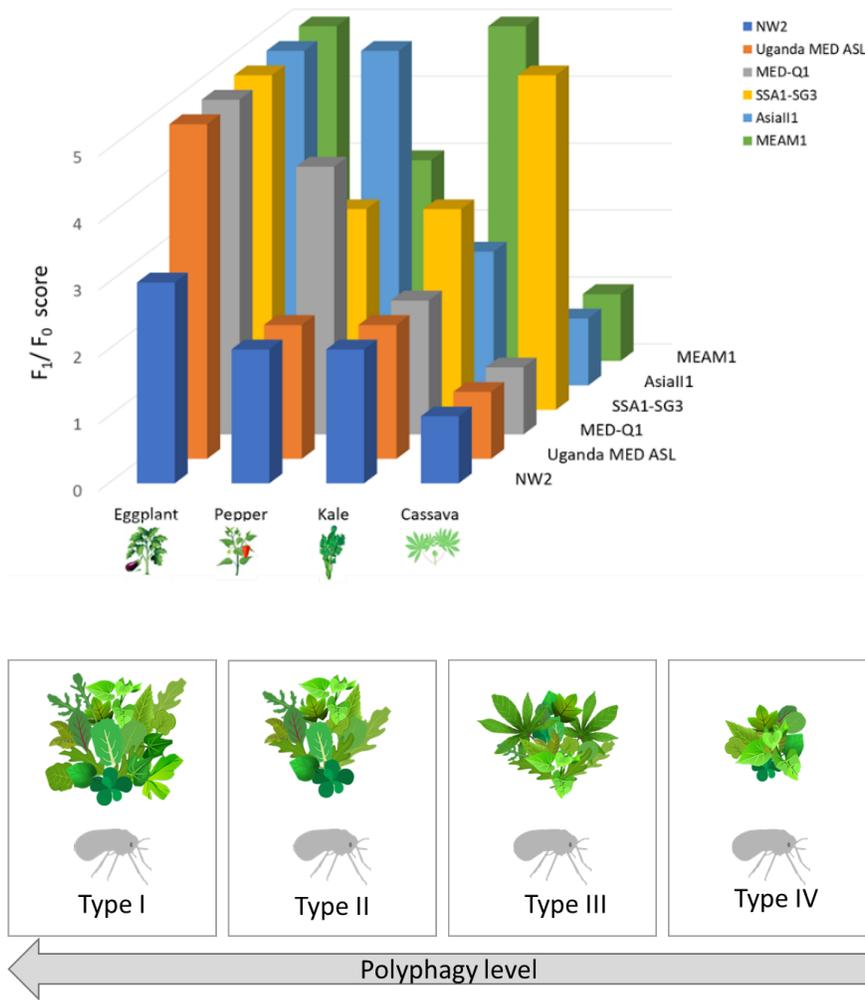


Figure 3.

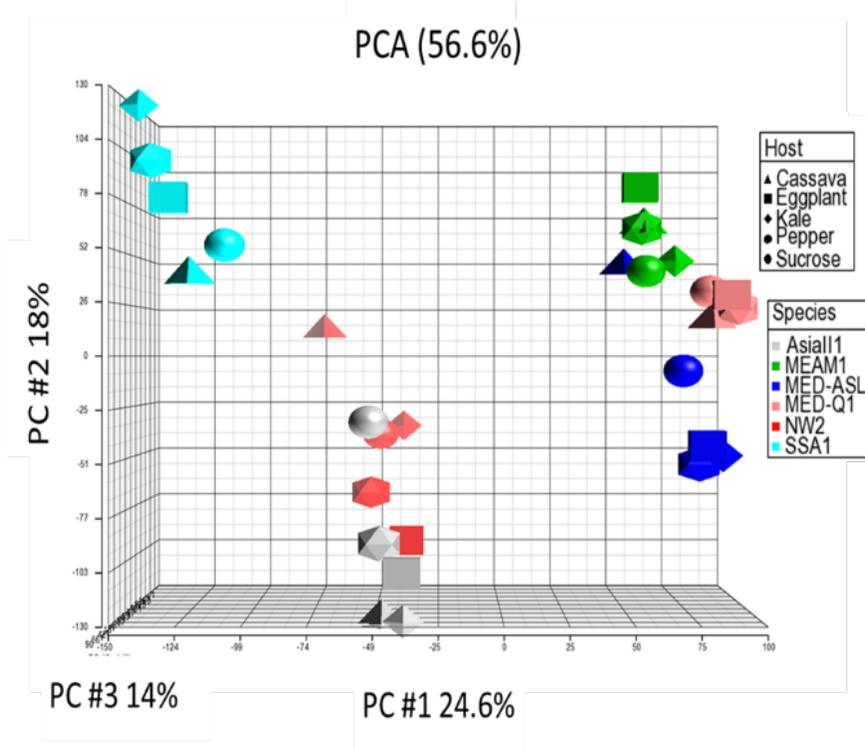


Figure 4a.

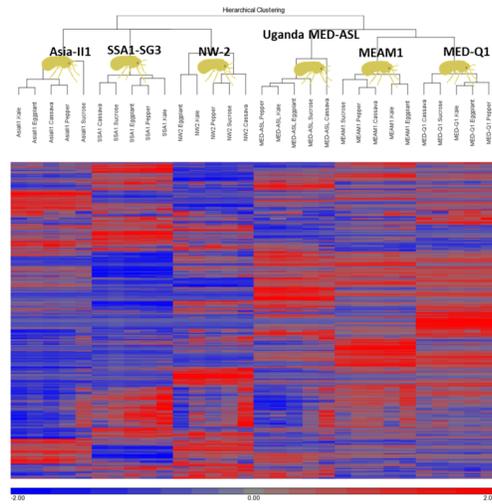
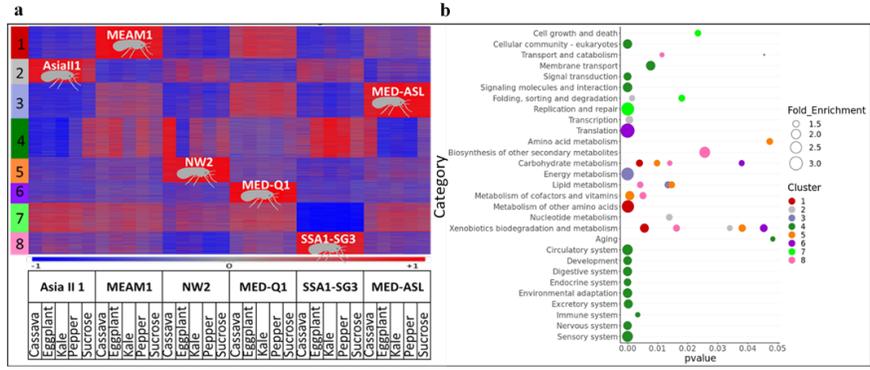


Figure 4b



**Figure 5.**

**Figure 6.**

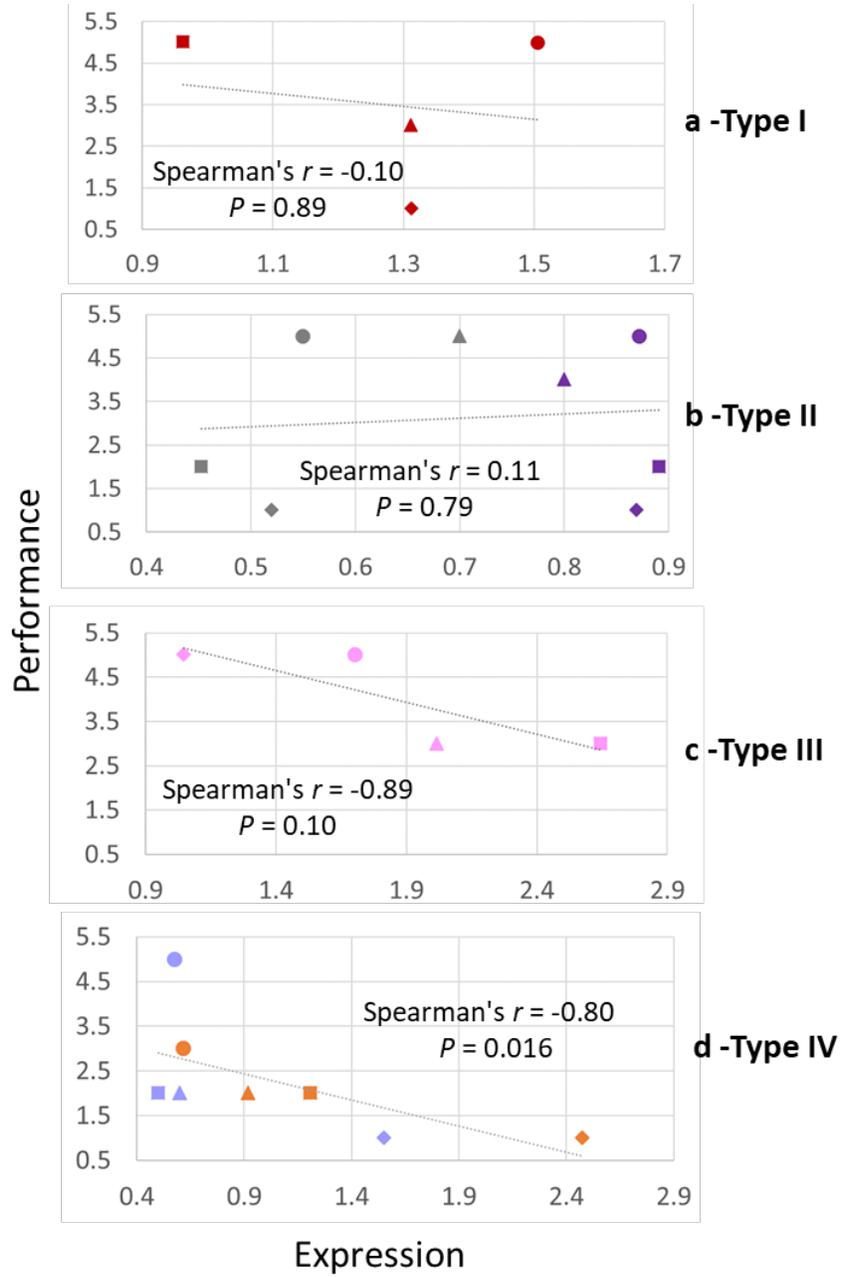


Figure 7.

Metabolic payhway (p<0.01)	MEAM1	MED-Q1	ASIAII1	SSA1-SG3	NW-2	MED-ASL
Starch and sucrose metabolism						
Glutathione metabolism						
Arginine biosynthesis						
Nitrogen metabolism						
Drug metabolism-cytochrome P450						
Metabolism of xenobiotics by cytochrome P450						
Purine metabolism						
Pyrimidine metabolism						
Ascorbate and aldarate metabolism						
Pentose and glucuronate interconversions						
Steroid hormone biosynthesis						
Porphyrin and chlorophyll metabolism						
Retinol metabolism						
Drug metabolism-other enzymes						
Cyanoamino acid metabolism						
Glycine, serine and threonine metabolism						
Thiamine metabolism						
Oxidative phosphorylation						
Caprolactam degradation						