Evolution, 2023, **XX(XX)**, 1–15
https://doi.org/10.1093/evolut/qpac049
Original Article

Larval transcriptomes reflect the evolutionary history of plant–insect associations

Maria de la Paz Celorio-Mancera, PhD[1,*], Rachel A. Steward, PhD[1,*], Peter Pruisscher, PhD[2], Agata Smialowska, PhD[3], Mariana Pires Braga, PhD[4], Niklas Janz, PhD[1], Christopher W. Wheat, PhD[1], Sören Nylin, PhD[1]

[1]Department of Zoology, Stockholm University, Stockholm, Sweden
[2]Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden
[3]National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Stockholm University, Stockholm, Sweden
[4]Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden

Corresponding author: Department of Zoology, Stockholm University, S-106 91 Stockholm, Sweden. Email: soren.nylin@zoologi.su.se
*M.P.C.M. and R.A.S. should be considered joint first authors.

Abstract

In this study, we investigated whether patterns of gene expression in larvae feeding on different plants can explain important aspects of the evolution of insect–plant associations, such as phylogenetic conservatism of host use and re-colonization of ancestral hosts that have been lost from the host repertoire. To this end, we performed a phylogenetically informed study comparing the transcriptomes of 4 nymphalid butterfly species in *Polygonia* and the closely related genus *Nymphalis*. Larvae were reared on *Urtica dioica*, *Salix* spp., and *Ribes* spp. Plant-specific gene expression was found to be similar across butterfly species, even in the case of host plants that are no longer used by two of the butterfly species. These results suggest that plant-specific transcriptomes can be robust over evolutionary time. We propose that adaptations to particular larval food plants can profitably be understood as an evolved set of modules of co-expressed genes, promoting conservatism in host use and facilitating re-colonization. Moreover, we speculate that the degree of overlap between plant-specific transcriptomes may correlate with the strength of trade-offs between plants as resources and hence to the probability of colonizing hosts and complete host shifts.

Keywords: insect–plant associations, gene expression, genetic modules, trade-offs, host shifts, phenotypic plasticity

Interactions between phytophagous insects and their host plants are ubiquitous on the planet and, as such, have attracted considerable attention from both basic and applied scientists over the years (e.g., Bale et al., 2002; Dyer et al., 2007; Ehrlich & Raven, 1964; Erb & Reymond, 2019; Futuyma & Agrawal, 2009; Janz, 2011; Liebhold et al., 1995; Mitter et al., 1991). Consequently, the ecological and evolutionary patterns of insect–plant associations have been well documented in many respects. Insect–plant interactions are often very conservative over evolutionary time (Mitter et al., 1991; Nylin et al., 2014; Ronquist & Liljeblad, 2001), and specialization on a plant species or clade is the dominating pattern (Forister et al., 2015).

Somewhat paradoxically, given the general evolutionary conservatism of insect–plant associations, colonizations of new host plants are still commonly seen. This can result in a broadening of the host repertoire or a complete host shift (Forister et al., 2012; Joshi & Thompson, 1995; Nyman, 2010). Furthermore, phylogenetic recurrence (i.e., the same hosts apparently being colonized repeatedly in a clade over the course of its evolution) is often observed, suggesting that historical hosts that have been "lost" can be re-colonized relatively easily (Agosta & Klemens, 2009; Janz et al., 2001; Nylin et al., 2014). Such re-colonizations can contribute to the oscillations in host range (from relatively specialized to more generalized associations and vice versa) which seems to be another general feature of the evolution of insect-plant associations (Braga et al., 2018a; Janz & Nylin, 2008; Nosil, 2002). Ancestral hosts sometimes evidently will remain as potential hosts for the future even when they are temporarily not used (Braga et al., 2018b; Nylin & Wahlberg, 2008; Nylin et al., 2014), and the realized host range can thus fluctuate between including them or not. Very similar evolutionary patterns of specialization, phylogenetic conservatism, colonizations, re-colonizations, and oscillations in host range are observed in other parasite–host and even mutualistic species associations, broadening the perspective even further and implying consequences for invasive species and emerging infectious diseases (Agosta et al., 2010; Brooks et al., 2019; Nylin et al., 2018; Torres-Martinez et al., 2021).

The proximate mechanisms explaining these evolutionary patterns are much less well understood. In the case of insect–plant dynamics adaptations to deal with plant chemistry is often assumed to play an important role in constraining host use (Ehrlich & Raven, 1964; Heidel-Fischer & Vogel, 2015; Janz, 2011; Speed et al., 2015), and there is evidence that this

Received June 20, 2022; revisions received November 4, 2022; accepted November 25, 2022

is true in at least some insects (Becerra, 1997; Edger et al., 2015; van der Linden et al., 2021; Wahlberg, 2001). However, even in these taxa, it is not clear precisely how plant chemistry affects ecology and evolution. In order to better understand the causality behind such dynamics, we need to better understand what it actually means to be genetically "adapted" to use a specific plant as a resource. Recent genomic studies have provided much new information, but patterns are idiosyncratic and complex, meaning that generalizations are difficult to make regarding candidate genes (Chaturvedi et al., 2018; Vertacnik & Linnen, 2017). This is also true regarding entire plant-specific transcriptomes—i.e., the reaction norms of differentially expressed genes showing how the juvenile insect responds with phenotypic plasticity to a particular plant and its chemistry of secondary metabolites (Birnbaum & Abbot, 2020).

The study presented here is based on the hypothesis that such adaptations can be understood as sets of host plant-dependent *modules* of co-expressed genes and the corresponding phenotypes. Modularity is an old concept in evolutionary biology (see e.g., Schlosser & Wagner, 2004; West-Eberhard, 2003). It refers to the universal property of living things that although the parts of an organism are integrated, there is also always a degree of discreteness and dissociation among parts, and some integration within parts; this is in fact the reason why we can distinguish "traits" of organisms. In the early stages of the evolution of new modularity, we expect a low degree of discreteness and much overlap among modules, but over time selection may result in discrete co-functional modules (West-Eberhard, 2003). Recently, modularity has again emerged as a central concept in evolutionary biology, serving as a unified conceptual framework for genetics, developmental biology, systems biology, and all kinds of multivariate evolution (Melo et al., 2016; Schlosser & Wagner, 2004; Wagner et al., 2007), including the evolution of phenotypic plasticity and environment-specific gene expression (Snell-Rood et al., 2010). In genetics, the term "module" is most often applied to small sets of functionally related genes, typically members of the same biological pathway (Wagner et al., 2007). Here, we also use it in a broader sense to refer to sets of genes showing plant-specific expression, since co-expressed traits can potentially "evolve together as a coadapted set" (West-Eberhard, 2003). Such sets will in turn most likely contain smaller genetic modules related to specific functions and pathways.

Comparative studies of plant-specific gene expression in insects are still rare, even though they are an important step toward understanding the ecology and evolution of insect–plant associations (Birnbaum & Abbot, 2020). We theorize that ancestral modules of gene expression shared among related species could contribute to the evolutionary conservatism in insect–plant interactions and provide a possible mechanism for phylogenetic recurrence by functioning as a genetic "memory" of lost host plants (cf. Ho et al., 2020). To investigate this theory, we here use a comparative gene expression dataset from larval gut tissue of four related butterfly species reared on three different plant genera to test the following specific hypotheses:

1) Transcriptomes from larvae are plant-specific (i.e., shows broad-sense modularity).
2) Such plant-specific transcriptomes are shared between related species (i.e., reaction norms are conserved over evolutionary time).
3) Plant-specific transcriptomes are evolutionarily conserved even when the plant is no longer used as a host by ovipositing females.
4) Plant-specific transcriptomes show evidence of functional genetic modules, shared between related species (i.e., strict-sense modularity is conserved over evolutionary time).

The hypotheses were tested using butterflies in the nymphalid tribe Nymphalini. Previous research on this tribe has provided a wealth of information on the ecological and evolutionary dynamics of host utilization (e.g., Celorio-Mancera et al., 2013; Janz et al., 2001; Nylin, 1988; Nylin et al., 2015; Weingartner et al., 2006) and inspired the "oscillation hypothesis" (Janz & Nylin 2008). Notably, a limited set of plant families seem to be repeatedly colonized by butterflies in the clade (or never completely lost as hosts), and larvae of most species can survive and, in some cases, thrive on the ancestral host (the family Urticaceae and its relatives) even when it is no longer used by females (Janz et al., 2001; Nylin et al., 2015). Here, we show that conserved modularity in phenotypically plastic gene expression has likely facilitated these observed patterns–in this system as well as other species associations sharing similar features.

Materials and methods

Study organisms and phylogenetic context

The butterfly tribe Nymphalini (family Nymphalidae, subfamily Nymphalinae) has the plant family Urticaceae and/or other "urticalean rosids" (Ulmaceae and Cannabaceae) as sole hosts for all basal branches in the tribe and this is clearly the ancestral host association (Janz et al., 2001; Nylin & Wahlberg, 2008). Phylogenetic reconstructions suggest that the host repertoire was later widened to include other plant families such as Salicaceae, Betulaceae, and (in *Polygonia*) Grossulariaceae, in most species followed by re-specialization on the ancestral or novel hosts (Janz et al., 2001; Weingartner et al., 2006). Figure 1A shows part of the tribe, with the study species in *Nymphalis* and *Polygonia* and some phylogenetic context. The four species tested here (Figure 1A) were the following: Eurasian *Nymphalis xanthomelas*, a specialist on *Salix*; North American *Polygonia satyrus*, a specialist on *Urtica*; Eurasian *Polygonia c-album*, polyphagous on *Urtica* and the related *Ulmus* and *Humulus* ("urticalean rosids") as well as *Salix*, *Ribes*, *Betula,* and the related *Corylus*; and finally its sister species, North American *P. faunus*, polyphagous on *Salix*, *Ribes*, *Betula,* and the related *Alnus* as well as *Rhododendron*, but not using *Urtica* or other "urticalean rosids" as hosts in the field (Janz et al., 2001; Nylin et al., 2015; Weingartner et al., 2006).

The divergence between *Nymphalis* and *Polygonia* happened about 9–10 My ago, between *P. satyrus* and (*P. c-album* + *P. faunus*) about 5 My ago and between the *P. c-album* species complex (including the putative species *P. extensa*, *P. gongga*, and *P. interposita*) and *P. faunus* about 2 My ago (Figure 1A; divergence times from Chazot et al. [2021]).

The two polyphagous species *P. c-album* and *P. faunus* were collected and reared for gene expression studies on *Urtica dioica*, *Salix* spp., and *Ribes* spp. in Stockholm, Sweden and Edmonton, Canada, respectively. The stinging nettle *U. dioica* occurs on both continents and was used at both sites, but for the other two genera, we used pairs of similar species thought

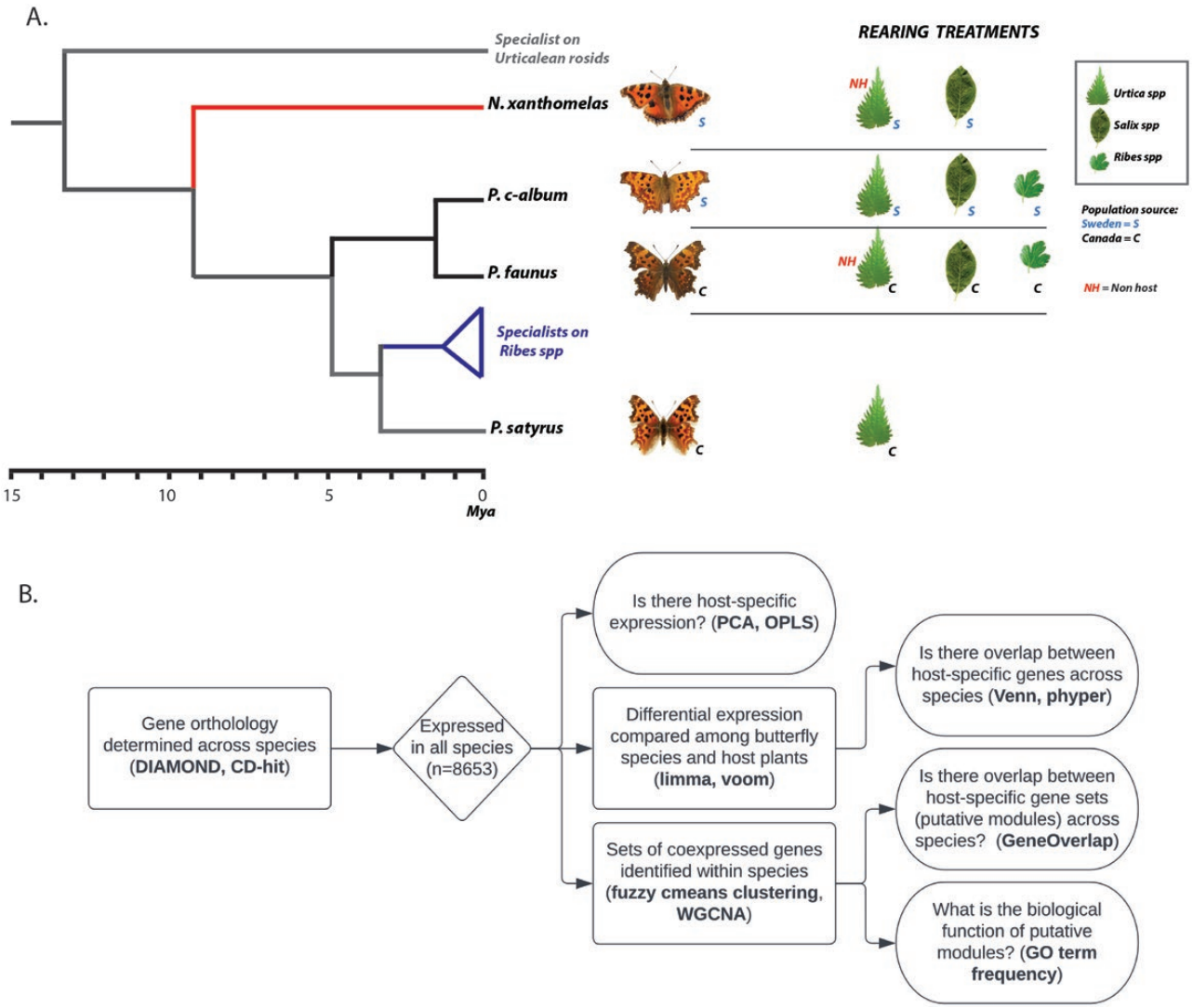

Figure 1. Project design. (A) Simplified phylogenetic context (left) for the study species in *Nymphalis* and *Polygonia* together with experimental design and geographical source population (right). Lineages reconstructed by parsimony as specialists on urticalean rosids in gray, *Salix* specialists in red, *Ribes* specialists in blue and polyphagous species in black. Species divergence times (bottom) from Chazot et al. (2021). Note that the ancestral host *Urtica* is a non-host for *N. xanthomelas* and *P. faunus* but viable for growth and could thus be tested. (B) Flowchart of analyses performed on transcriptome data in order to find plant-specific gene expression and putative host plant-related genetic modules shared among butterfly species. Analytical tools in bold.

to be closely related (*S. caprea* and *S. discolor*; *R. uva-crispa* and *R. oxycanthoides*; see also Nylin et al [2015]). *P. satyrus* was collected and reared in Edmonton on *U. dioica* (hatchlings did not survive to first molt on *Salix* or *Ribes*), and *N. xanthomelas* was collected and reared in Stockholm on *U. dioica* and *S. caprea* (hatchlings did not survive to first molt on *Ribes*).

The genera *Urtica, Salix,* and *Ribes* are quite distantly related (placed in the orders Rosales, Malpighiales, and Saxifragales, respectively) and thus can be expected to differ in plant chemistry. The chemical composition of the species used here has been studied individually, but unfortunately, not in a systematic manner allowing for easy comparison. However, the plants do differ in the chemistry of secondary metabolites. Known compounds in leaves of *U. dioica* include sterols, saponins, coumarins, alkaloids, flavonoles such as quercetin and rutin (Abdeltawab et al., 2012; Kavtaradze et al., 2001) and (in the stinging hairs) also, e.g., histamine and acetylcholine (Hegnauer, 1990). *S. caprea* (and presumably the close relative *S. discolor*) is low in the phenolic glucosides that characterize some other *Salix* species and deter many insects, but have high levels of condensed tannins (Hallgren et al., 2003). There are also unusual flavonoids such as salicaprene present, in addition to, e.g., quercetin and rutin (Hegnauer, 1990). *Ribes* is not used as a host by any other nymphalid butterflies outside of *Polygonia*, suggesting that it is challenging to colonize (Celorio-Mancera et al., 2013). One possible reason is that leaves contain a rare form of hydroxynitrile glucosides with isoleucine as a precursor, including cyanogenic glucosides in at least some species, like the two tested here (Bjarnholt & Moller, 2008; Celorio-Mancera et al., 2013; Hegnauer, 1990). Also present are more common compounds such as, e.g., quercetin and kaempferol (Hegnauer, 1990).

Sampling for gene expression

Larvae were reared individually from hatchlings, on cuttings of the respective plants. A split-brood design was possible for those species that could be reared on more than one plant species. For each of the nine butterfly species ×

plant treatments (Figure 1A), three larvae from three different families (n = 9) were sacrificed in the fourth instar, and the midgut was dissected out. This is the penultimate instar, where much of the feeding (in absolute terms) happens, but before any processes related to pupation are expected to take place. The peritrophic membrane containing the digested plant material was removed, and midguts were stored in RNAlater (Ambion, Austin, TX). Samples were stored at −80°C until they could be processed. Total RNA was isolated by first homogenizing the midgut tissue in TRIzol (Life Technologies Corporation, Carlsbad, CA) and later processed using the Direct-zol RNA miniprep (Zymoresearch, Irvine, CA) following the manufacturer's instructions. The quality of RNA was verified by gel-based electrophoresis (Experion, Bio-Rad, Hercules, CA) and the quantity by fluorometry (Qubit, Life Technologies Corporation, Carlsbad, CA).

The RNA from the midgut of individual larvae was sequenced at the National Genomics Infrastructure (NGI), Science for Life Laboratory, Stockholm, Sweden. The RNA libraries constructed with poly-A selection (Illumina TruSeq Stranded mRNA) were sequenced using paired read length of 2 × 125 bp on a HiSeq2500 platform. Four de novo transcriptomes were assembled, one for each butterfly species, using Trinity software 2.4.0. (Grabherr et al., 2011). The quality assessment of the transcriptomes and downstream analyses such as transcript quantification were performed following the protocol for using Trinity for de novo transcriptome assembly (Haas et al., 2013).

We obtained an average of 12 M reads of raw data per sample. After trimming and aligning the reads in pairs, we obtained 97%–99% of them represented by each species-specific transcriptome assembly. We also assessed the transcript contig length based on the set of transcripts representing 90% of the normalized expression data (E90N50). Detailed statistics of the transcriptomes assembled are summarized in Supplementary Table S1.

Orthology across butterfly species

In order to be able to assess the degree of similarity of plant-specific transcriptomes between all four butterfly species (enabling a test of Hypotheses 2–4; see a flowchart of analyses in Figure 1B), it was necessary to determine the orthology of genes across species. This was done by making use of the most closely related butterfly species with a well-annotated genome, *Heliconius melpomene* (version Hmel2; The Heliconius genome consortium, 2012; Davey et al., 2016). We searched for sequences in the four study-species with high similarity to genes in this species, providing us with a read-count table for genes identified as expressed in all butterfly–plant pairs.

The method to obtain this dataset used DIAMOND (Buchfink et al., 2015) to align the transcript sequences to the non-redundant (CD-hit collapsed) predicted gene set (PGS) of the *H. melpomene* genome at the amino acid level (Fu et al., 2012; Li & Godzik, 2006) available from Lepbase (Challis et al., 2016). In order to estimate transcript abundance, once the ortholog set was obtained, we implemented the RNA-Seq by Expectation-Maximization analysis (Haas et al., 2013). The CD-Hit collapsed PGS for *H. melpomene* resulted in 12,607 sequences and out of these 10,923 had orthologues in the study species. 8,653 genes were expressed in all four study species and included in the full dataset.

Patterns of gene expression

Unsupervised multivariate statistical analyses and principal component analyses (PCA) were applied to the ortholog dataset (8,653 genes) using Qlucore Omics Explorer 3.7 (Qlucore, Lund, Sweden). Gene expression count data for PCA were first filtered to remove genes with very low counts (<10 in >90% of samples), leaving 7,923 genes. Following the Qlucore default and recommended pipeline, log-transformed counts were set to a mean of zero and scaled to unit variance.

Since the number of variables in our study exceeds the number of samples and have highly correlated functions and expression patterns, we applied a supervised multivariate analysis to test whether the profiles of gene expression group according to plant used for rearing (Hypotheses 1–2). The multivariate analysis consisted of an orthogonal partial least squares (OPLS) method using the package ***ropls*** (v.1.22.1; Thevenot et al., 2015) in R (v. 4.0.5; R_Core_Team, 2018), available via Bioconductor (Huber et al., 2015) and the BiocManager package. We tested the predictability of our model with a training set containing the real association of the plant treatment to each sample. In order to test whether the predictability of our model was possible just by chance, we instead used a training set where the association between sample and plant was randomized. Both tests were cross-validated by using 1,000 iterations.

Differential gene expression among butterfly species and plant treatments was analyzed using limma + voom (packages "limma" [Ritchie et al., 2015] and "edgeR" [Robinson et al., 2010] in R) as implemented in the "extended statistics" available for the Qlucore program. To reduce the risk of false positives, we chose a q-value (i.e., p-value adjusted for false discovery rates) threshold of $q < 0.05$ for significance in these analyses.

Gene ontology terms were assigned to the dataset as described in Celorio-Mancera et al. (2016). Significance of overlap between genes expressed on the same plant in different species (of relevance to test Hypotheses 2–4) was investigated using the R package GeneOverlap (Shen & Sinai, 2021) or calculated using the *phyper* function in R, where the probability of the overlap is given by phyper(q, m, n, k, lower.tail = FALSE) and q = size of overlap-1, m = number of upregulated genes in species 1, n = total number of genes in dataset-m), and k = number of upregulated genes in species 2. The expected overlap from chance is $(m * k)/n$.

Co-expressed genes and modularity

We used two methods, fuzzy c-means clustering and WGCNA gene co-expression analysis (Langfelder & Horvath, 2008), to further investigate plant-specific co-expression of genes and putative plant-specific genetic modules (testing Hypothesis 4; Figure 1B). The two methods are different and complementary. We used fuzzy c-means clustering to find sets of differentially expressed genes (previously identified with limma + voom, see above) with similar expression profiles across plant treatments. In contrast, WGCNA evaluates normalized counts from all genes to find networks of genes (termed "modules") with correlated gene expression across samples, and only after that "modules" are tested for significant correlation with treatment (or any other variable describing the samples). The first method is more direct and intuitive, but the second should be more powerful in finding putative genetic modules.

C-means clusters

Clustering of differentially expressed genes across plants was performed separately for *P. c-album* and *P. faunus*. They were reared on three different plants, leading to a complex pattern of many small clusters when all genes were included. Since we were focusing on responses to plant treatments across species, we decided to reduce the complexity by analyzing only the set of genes that were differentially expressed according to plant treatment, in a dataset including only these two species (limma + voom at $q < 0.05$; 2,047 genes). The number of clusters present in the data was determined by SSE clustering using K-means (Jain & Dubes, 1988). The raw expression data of the genes were standardized using the variance stabilizing transformation (vst, DESeq2 package; Love et al., 2014), and subsequently centered and scaled (scale, edgeR package; Robinson et al., 2010). This was done in order to be able to compare between genes and to identify clusters with similar expression profiles independent of expression levels. After scaling, the replicates of each sample were averaged within each gene to one sample mean. These scaled mean values were used in the cluster estimation and subsequent clustering. Fuzzy c-means clustering was performed by estimating the "fuzzifier" needed for c-means clustering (Kumar & Futschik, 2007), and then using the cmeans function of the R package e1071 (Pal et al., 1996) to estimate "membership" or the degree to which a genes expression fit with the overall expression of each cluster. The combination of these methods allowed for the detection of supported clusters inherent in the data. Finally, in order to obtain a representative, non-overlapping group of genes with similar expression, we filtered clusters to contain only genes with membership scores > 0.51. This cutoff was chosen because it ensured that, of all clusters, this gene belongs most to this particular cluster. We further explored the potential biological processes of different and overlapping clusters using GO terms previously assigned to each gene. GO terms were considered if they appeared two or more times in a cluster.

Gene co-expression networks

We used the WGCNA R-package (Langfelder & Horvath, 2008) to investigate gene co-expression networks and detect putative genetic modules in all four butterfly species. The dataset was filtered to exclude genes where > 90% of samples had counts < 10, as suggested by the WGCNA FAQ. The data was subsequently TMM-normalized (edgeR package) and log-transformed. We chose a soft-clustering power of 18 as suggested by the authors for signed networks with < 20 samples. We specified signed networks using biweight midcorrelation and Pearson correlations as a fallback, per package recommendations, with a minimum of 30 genes per module. The same settings were used for all analyses, except the mergeCutHeight parameter (the threshold for merging of modules), which was adjusted to produce a similar number of modules in each species (*P. c-album* 0.1; *P. faunus* 0.2; *P. satyrus* 0.2; *N. xanthomelas* 0.3). This was deemed necessary because the number of modules identified by WGCNA is somewhat arbitrary and dependent on settings as well as on the exact patterns of correlated gene expression in each dataset. Choosing the same setting for all species resulted in one very large module or many very small modules identified in one or more species, making comparisons between species impractical. For the species reared on more than one plant, we tested for a correlation between module eigenvalues and plant used for rearing. In the case of *P. c-album* and *P. faunus*, reared on three plants, we performed completely separate analyses including only data from two plants at a time (*Urtica* and *Ribes* or *Urtica* and *Salix*), i.e., with the nominal variable "plant treatment" coded as a binary trait. The analysis on the *Urtica* specialist *P. satyrus* was performed without the plant treatment variable, in order to test whether similar modules still emerge when all data is from the same plant treatment.

The R package GeneOverlap (Shen & Sinai, 2021) was used to test the significance of gene overlap between different gene sets. The size of the total gene set for analysis was set to 7,923 for all analyses except for comparing clusters identified in *P. c-album* and *P. faunus* (where only the 2,047 differentially expressed genes in these two species were included); this is the number of genes expressed in all four species and excluding genes with consistently low expression as above.

For *P. c-album*, the 10 genes identified by WGCNA as having the highest probability of belonging to each module, as well as (for significantly plant-associated modules) the 10 genes in each module being most significantly correlated with the plant variable (in the *Urtica* vs. *Ribes* analysis only) were manually investigated for probable function in the specific context of the midgut. This was typically done using the STRING database (Jensen et al., 2009) to find homologs to the *H. melpomene* gene in *Drosophila melanogaster* followed by scrutiny of the record in FlyBase (Larkin et al., 2021). If no relatively close homolog in *D. melanogaster* was found, the UniProt database (The UniProt Consortium, 2021) was used to find probable functions of closer homologs in other species. A few of the *H. melpomene* genes were not found in the STRING database. In such cases, the sequence was downloaded from Lepbase (Challis et al., 2016) and annotated homologs were searched for using the Blastx procedure at NCBI (Bethesda, MD). The same procedure was followed for modules in *P. faunus* and *N. xanthomelas*, but only for the two modules most significantly positively and negatively correlated with the plant variable. For the *Urtica* specialist *P. satyrus*, we used a similar approach to investigate gene function in two modules (the "Psa_m1" and "Psa_m12" modules) showing strong gene overlap with the most strongly *Urtica*- and *Ribes*-associated modules, respectively, in *P. c-album* and *P. faunus*.

Results

Gene expression is predicted by species and plant treatment

We compared gene expression among 81 different samples (nine treatments with nine replicates each: *P. c-album* and *P. faunus* each × 3 plant treatments; *P. satyrus* × 1 plant; *N. xanthomelas* × 2 plants). Samples clustered in PCA space both according to butterfly species and plant genus (the latter supporting Hypothesis 1 and potentially 2–3). This was seen regardless of whether all 10,923 genes were included (Supplementary Figure S1A,B), or only genes expressed in all four species and filtered to exclude genes with consistently low counts (7,923 genes; Supplementary Figure S1C,D).

A supervised multivariate analysis of the dataset (OPLS) confirmed that gene expression is significantly predicted by plant treatment across butterfly species, despite the large species effect (Figure 2; supporting Hypothesis 1 and potentially 2–3). The model assigned all *Urtica* samples correctly to plant

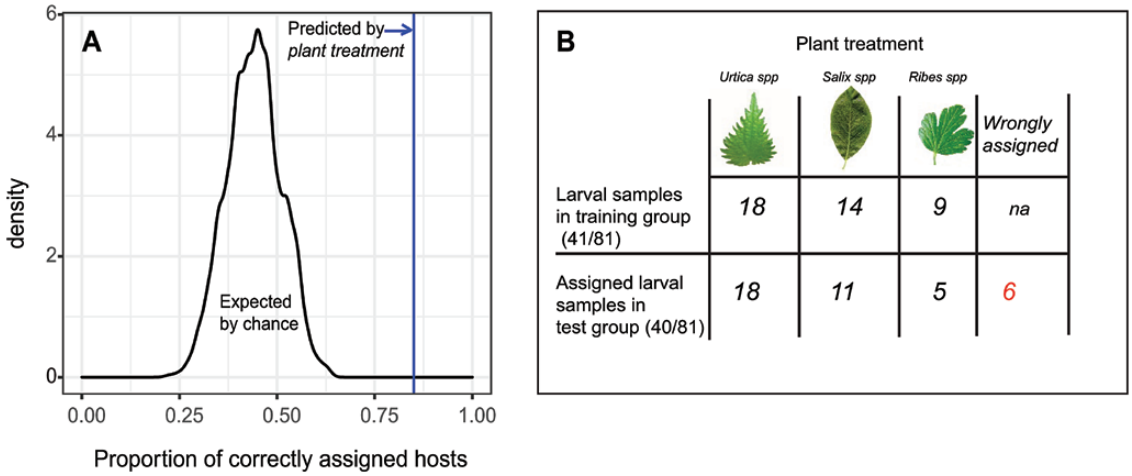

Figure 2. Plant feeding treatment significantly predicts gene expression profiles in orthogonal partial least squares (OPLS) model. (A) A model using plant treatment as the explanatory variable and trained using about half of the samples from each plant in the experiment (n = 41) assigned the remaining samples (n = 40) to the correct plant treatment with 85% accuracy (blue line). This was significantly higher (p < .001) than model predictions when plant treatment was randomized for each of the test samples (1,000 iterations). (B) Number of samples per plant genus that were used to train the OPLS model, and number of samples correctly and wrongly assigned by the model.

treatment, but performed less well with in particular *Ribes* (used only by two species, and thus not well represented in the training set). The proportion of plant treatments correctly assigned by the model using the gene expression dataset (85%) was much higher than when the model was run on randomized data (Figure 2; permutation test p < .001).

Plant-specific gene expression is shared between species

We found a total of 7,578 genes with significantly different gene expression among butterfly species and 3,553 genes differing among plant treatments (limma + voom at q < 0.05). The strongest plant treatment expression differences were found between *Urtica* and *Ribes* in both *Polygonia c-album* and *P. faunus* (limma + voom at q < 0.05; Supplementary Table S2). Of these genes, many were similarly up or down regulated in both butterfly species, representing 63.3% and 37.9% of differentially expressed genes in *P. c-album* and *P. faunus*, respectively (Supplementary Figures S2A and S3A). The overlap between the sets of genes upregulated on either *Urtica* or *Ribes*, respectively, is much higher than expected from chance (phyper p < .001; Supplementary Table S2). We note that this similarity in expression is driven by plant treatment (supporting Hypotheses 1–3), as there is in contrast very low overlap between genes upregulated on different plants in the two species (Supplementary Figure S2A). However, the two *Polygonia* species differed considerably in their gene expression on *Salix* (Supplementary Table S2).

Gene expression in *Nymphalis xanthomelas* showed a degree of overlap with *Polygonia* concerning upregulation on *Urtica* or *Salix* (Supplementary Table S2; Supplementary Figure S2B,C). Many of the significantly upregulated genes in *N. xanthomelas* on *Urtica* (limma + voom at q < 0.05) were the same as the genes upregulated in one or both of the two *Polygonia* species when reared on *Urtica* (vs. *Salix*), in both cases higher than chance would predict (supporting Hypotheses 2–3; phyper, p < .001; Supplementary Table S2; Supplementary Figure S2B,C). The overlap between *N. xanthomelas* and *P. c-album* is however not statistically significant when it comes to upregulation on *Salix*, whereas the overlap with *P. faunus* is significant (phyper p < .001; Supplementary Table S2; Supplementary Figure S2B,C).

Shared genes also tended to have the largest fold changes between plant treatments (Supplementary Figure S3). When differentially expressed genes were filtered to only the 20 most upregulated (highest fold change), there was a high degree of overlap between the two *Polygonia* species, and also with *N. xanthomelas* (supporting Hypotheses 2–3; Supplementary Table S3). On *Urtica*, these shared genes included hemicentin-1; an attacin-like protein; a carboxypeptidase; two peptidoglygan recognition proteins and an antibacterial peptide. On *Ribes* (and *Salix* in the case of *N. xanthomelas*) a prostaglandin reductase-like gene and two facilitated trehalose transporter tret1-like genes were the most prominently shared upregulated genes, besides several non-annotated genes (Supplementary Table S3).

Plant-specific gene expression shows shared broad-sense modularity

Genes that were differentially expressed among plant treatments were further evaluated using PCAs with individual genes as data points, revealing distinct and almost mutually exclusive sets of genes with high or low expression on *Urtica*, and the reverse on *Ribes*, in both *P. c-album* and *P. faunus* (Figure 3). These plant-specific transcriptomes indicate similar broad-sense genetic modules (sets of co-expressed genes) in the two species, supporting Hypotheses 2–3. This is less true for *Salix*, where gene expression patterns are much less distinctly clustered in PCA space and overlap with the other two plant treatments (Figure 3). *N. xanthomelas* could only be reared on *Urtica* and *Salix* and, similarly to *Polygonia*, there was a lack of very distinct clusters of co-expressed plant-specific genes. However, a tendency for genes with high expression on *Urtica* and low expression on *Salix* can be seen at negative values of the PC1 axis (Figure 3). Results are similar when including all genes except consistently low-expressed genes (Supplementary Figure S4). Finally, *P. satyrus* could only be reared on *Urtica*. It is nonetheless interesting to note that genes with particularly high expression in this *Urtica* specialist (relative to the other species) are again mainly seen at

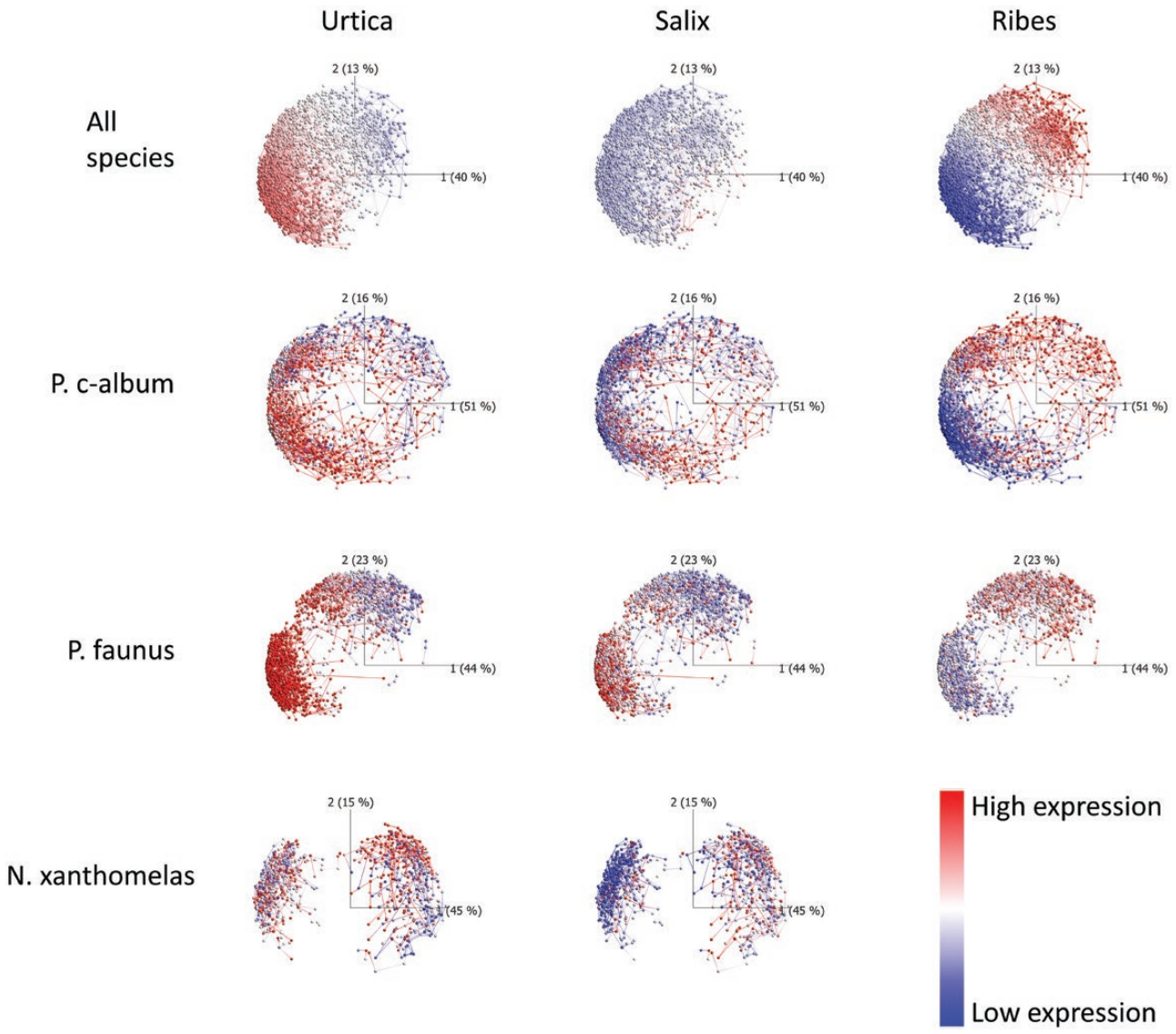

Figure 3. Genes clustered by expression on *Urtica* and *Ribes* when compared across all butterfly species. PCA plots of the first and second principal components showing high (in red) or low (in blue) gene expression on the respective plant when larvae were reared on *Urtica, Salix* and *Ribes* plants. Each plot point represents a single gene, and genes showing similar expression patterns (i.e., nearest neighbors in total PCA space, including additional principal components not visible in this rotation) are joined with lines. Top row with all four study species included; following rows with only data from *Polygonia c-album, P. faunus* or *Nymphalis xanthomelas*, respectively. Genes were filtered to those differing significantly by plant treatment, leaving 3,553, 2,355, 2,571, and 975 genes, respectively.

negative values of the PC1 axis, in the shared PCA space of the "all species" dataset (Supplementary Figure S5).

To summarize, samples differed between plant treatments in a similar manner across butterfly species, supporting Hypotheses 2–3. This is well illustrated by plotting the gene expression reaction norms based on the position of each sample on the PC1 axis in the PCA space including all four species. In both generalist and specialist species, *Urtica* samples are consistently found toward negative values of the PC1 axis (regardless of whether *Urtica* is an actual host for the species), *Ribes* samples more at the positive end, and *Salix* samples in between (Figure 4).

Co-expression of genes suggest shared plant-specific functional modules

Gene clusters

We used two different approaches to search for putative functional genetic modules involved in adaptation to specific plants (Hypothesis 4). The first approach involved clustering of differentially expressed genes according to their expression levels across plant treatments. Six such clusters were identified in *P. c-album* (containing 44–287 genes each) and six in *P. faunus* (79–221 genes each; Figure 5A and B; Supplementary Table S4).

In line with results reported above, no cluster shows clearly specific upregulation on *Salix* in any of the two polyphagous *Polygonia* species (Figure 5A and B). Rather, there tends to be intermediate expression on this plant for all genes, or upregulation on both *Salix* and another plant. Overlooking *Salix* for simplicity, in *P. c-album*, there are two clusters with genes upregulated on *Urtica* (referred to below as "*Urtica* up") and four clusters with genes upregulated on *Ribes* ("*Ribes* up"; Figure 5A and C). In *P. faunus*, there are four "*Urtica* up" clusters and two "*Ribes* up" clusters (Figure 5B and C).

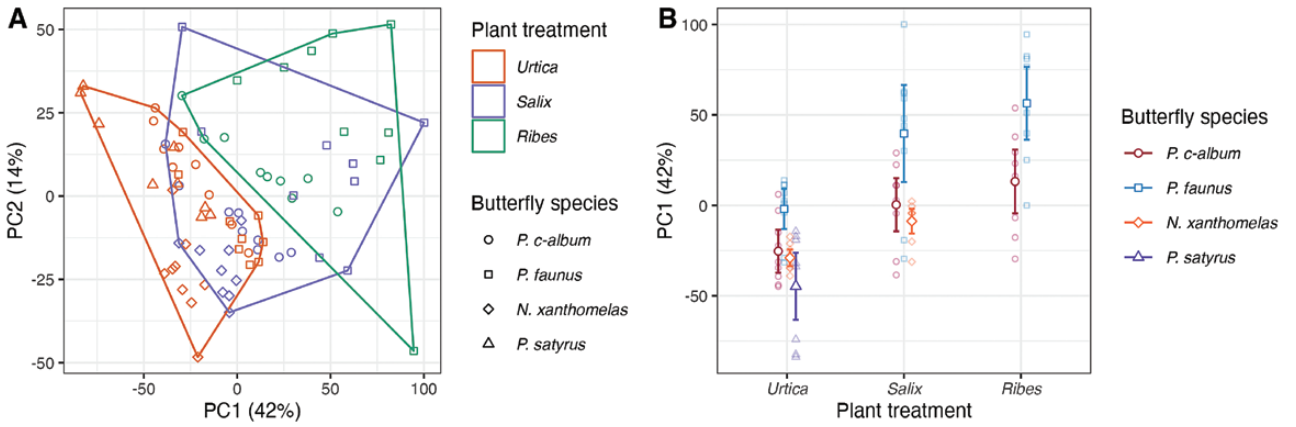

Figure 4. (A) PCA plot of all samples, based on the gene expression dataset filtered to 3,553 genes that were differentially expressed according to plant treatment, showing how samples from the same plant treatment tended to cluster across butterfly species. (B) Reaction norms for gene expression according to plant treatment (position of each sample on the PC1 axis of Figure 4A; error bars show 95% confidence intervals around the means) in larvae of the butterfly species *Polygonia c-album*, *P. faunus*, *P. satyrus*, and *Nymphalis xanthomelas* when reared on the plants *Urtica dioica*, *Salix* spp., or *Ribes* spp. Not all species could be reared on all plants.

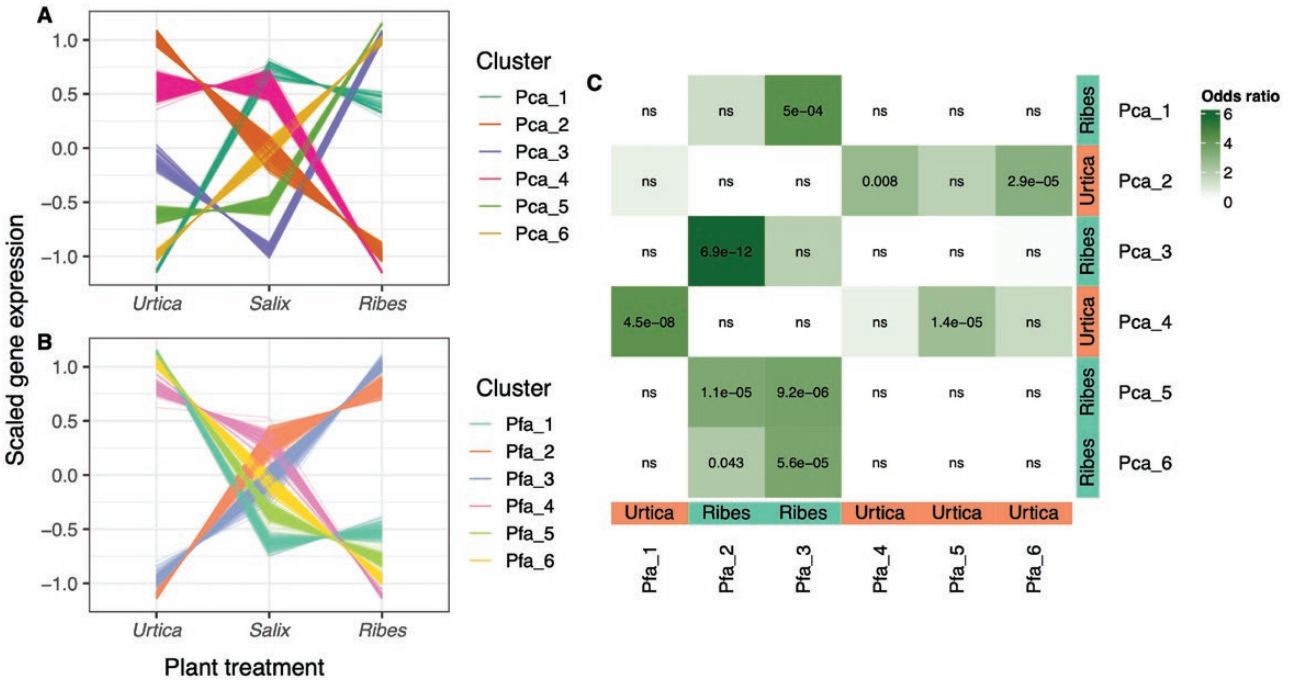

Figure 5. Patterns of gene expression in *Polygonia c-album* and *P. faunus* across three plant treatments. (A,B) Clusters of co-expressed genes across the plant treatments *Urtica, Salix* and *Ribes* in (A) *P. c-album* and (B) *P. faunus*. Lines represent individual genes and darker lines indicate stronger membership in the cluster. (C) Statistical significance of overlap between the gene clusters identified in *P. c-album* and *P. faunus*, out of 2,047 total differentially expressed genes in these two species. Darker green background indicates higher odds ratios. Tiles are labeled with *p*-values (Fisher's exact test, Benjamini–Hochberg correction for multiple comparisons, $p > .05$ labeled as n.s.). Clusters are labeled by row and column, and according to whether they show more upregulation on *Urtica* or *Ribes*. Clusters with significant overlap all show similar patterns of upregulation on *Urtica* or *Ribes*.

Notably, we found that the sets of genes in *P. c-album* and *P. faunus* clusters overlapped significantly in gene identity only when the clusters showed the same plant-specificity (supporting Hypotheses 2–3 and potentially 4; i.e., "*Urtica* up" or "*Ribes* up" in both species; see Figure 5C). Several of the "*Urtica* up" clusters in these two *Polygonia* species also overlapped significantly with the set of genes upregulated on *Urtica* in *N. xanthomelas*, whereas all "*Ribes* up" clusters in *Polygonia* overlapped with genes upregulated on *Salix* in *N. xanthomelas* (Supplementary Figure S6). It is thus of interest to further characterize the biological processes associated with these gene sets (Supplementary Table S5). All "*Urtica* up" clusters in both polyphagous *Polygonia* species had a high frequency of genes associated with the biological process "*regulation of transcription, DNA-dependent*." Other commonly shared GO terms between "*Urtica* up" clusters in *P. c-album* and *P. faunus* included "*regulation of GTPase activity*," "*mRNA splicing, via spliceosome*," "*microtubule-based movement*," "*intracellular signal transduction*," "*protein ubiquitination*," and "*serine family amino acid metabolic process*." Genes in the "*Ribes* up" clusters most commonly shared the processes "*oxidation-reduction process*," "*ribosome*

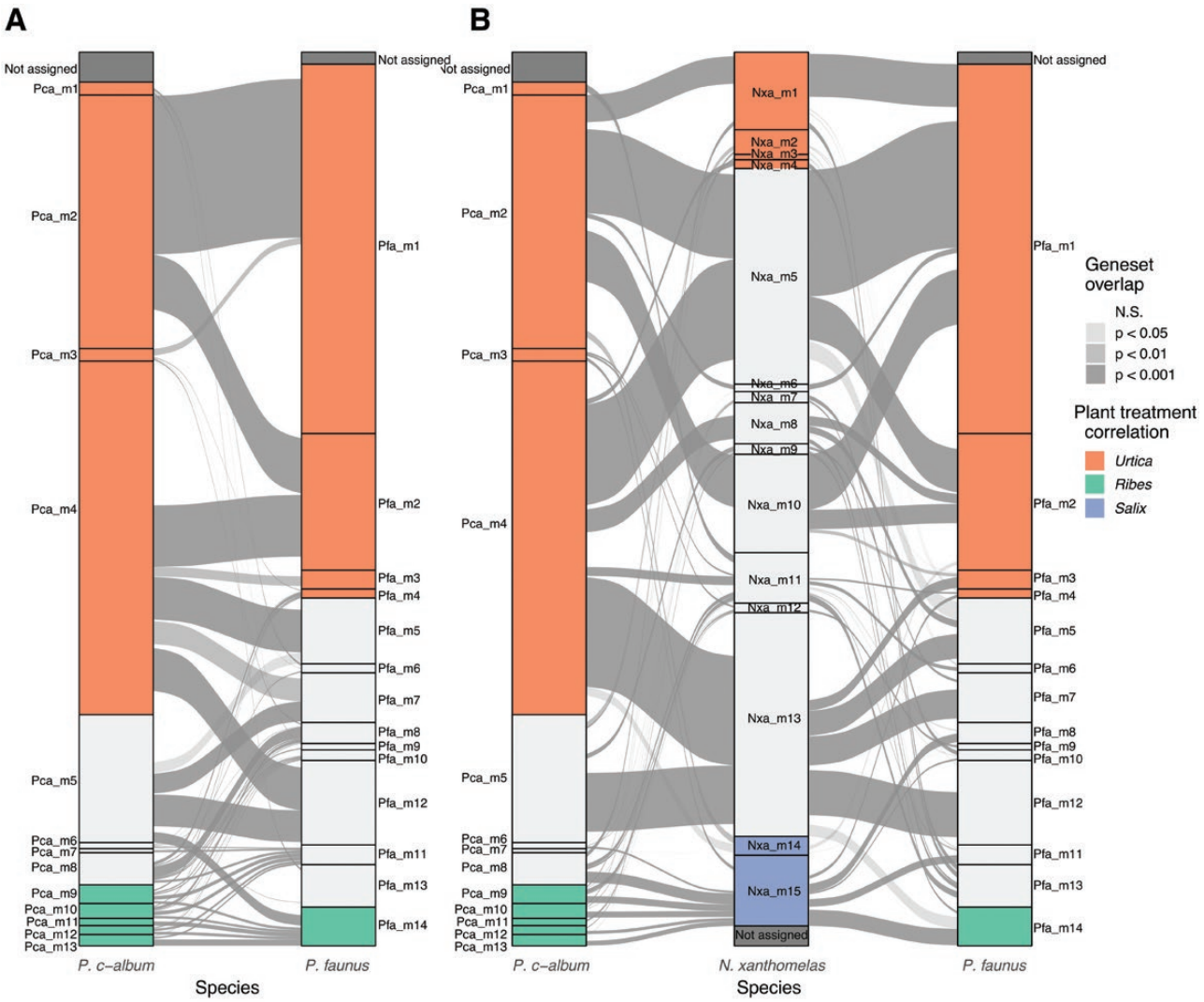

Figure 6. Overlap of genes in modules of co-expressed gene networks associated with plant treatments in *Polygonia c-album*, *P. faunus*, and *Nymphalis xanthomelas*. (A) Modules significantly correlated with *Urtica* (orange) or *Ribes* (green) tended to share genes between *P. c-album* and *P. faunus*, whereas modules correlated with different plant treatments showed little overlap between species. (B) Modules significantly correlated with *Urtica* in *N. xanthomelas* primarily shared genes with the most *Urtica*-correlated modules in both *Polygonia* species, while modules correlated with *Salix* (blue) shared genes with *Ribes*-correlated modules. Modules are ordered from strongest correlation with upregulation on *Urtica* at the top to most upregulation on the alternative plant at the bottom. Nonsignificant correlations are colored in light gray. Overlaps of gene sets were calculated using Fisher's exact tests implemented in GeneOverlap and are represented by gray bars connecting labeled modules. Bar width scales with the number of shared genes and darker colors indicate more significant overlaps. Nonsignificant overlaps are not shown.

biogenesis," "*transmembrane transport*," "*transport*," "*proton transport*," and "*protein folding*." Other processes were not as generally shared, but more characterized single clusters in both species, suggestive of more strict-sense modularity (Hypothesis 4). The clearest example is "*ATP hydrolysis coupled proton transport*" (high frequencies in the highly overlapping clusters Pca_c6 and Pfa_c3 but rare or absent in the other "*Ribes* up" clusters).

Regarding differentially expressed genes according to plant treatment in *N. xanthomelas*, similarly to *Polygonia*, the term "*regulation of transcription, DNA-dependent*" was more common among the "*Urtica* up" genes, and "*oxidation-reduction process*," "*ribosome biogenesis*," "*transmembrane transport*," "*proton transport*," and "*protein folding*" were more common among upregulated genes in the "alternative host" (*Salix*) treatment (Supplementary Table S5).

Gene co-expression networks

The second approach to gene co-expression was to use the WGCNA R package to find networks of genes with correlated expression patterns, termed "modules" in the package (putative functional genetic modules in the sense of the present paper). With the settings employed 13 such modules were identified in *P. c-album*, 14 in *P. faunus*, 15 in *N. xanthomelas*, and 12 in *P. satyrus*. Four of the modules in *P. c-album*, totaling 5,615 genes, were significantly negatively correlated with the plant treatment variable (i.e., upregulated on *Urtica*; Figure 6A) and five were significantly positively correlated (i.e., upregulated on *Ribes*; 542 genes). In *P. faunus*, four modules were significantly upregulated on *Urtica* (4,740 genes) while only one module was significantly associated with *Ribes* (344 genes; Figure 6A). Significantly plant-correlated modules in the two species showed strong gene identity overlap when the

same plant treatment was involved, less when the plant treatments differed (supporting Hypotheses 2–4). This is depicted graphically in Figure 6A, with details given in Supplementary Figure S7. Considering only modules significantly correlated with upregulation on a given plant in one of these two species, and only modules showing significant gene overlap with such a module in the other species, the same plant treatment was generally involved (11 out of 14 cases; Supplementary Figure S7). Moreover, all three exceptions involved the Pfa_m4 module in *P. faunus*, the module showing the weakest (significant) correlation with *Urtica* up-regulation. The single significant "*Ribes* up" module in *P. faunus* (Pfa_m14) shows significant gene overlap with all of the five "*Ribes* up" modules in *P. c-album* (Supplementary Figure S7 and cf. Figure 6A).

For *N. xanthomelas*, there was significant overlap between several of the "*Urtica* up" modules in *N. xanthomelas* and *Urtica*-correlated modules in the two *Polygonia* species, as well as some overlaps conflicting with plant treatments (Supplementary Figure S8 and cf. Figure 6B). The module most correlated with upregulation in the alternative plant treatment (here *Salix*; Nxa_m15 module) shows significant gene overlap with most of the *Ribes*-associated modules in the two *Polygonia* species (Supplementary Figure S8 and cf. Figure 6B). Setting the "alternative plant" to *Salix*, rather than *Ribes*, for module detection and plant treatment correlations also in the two *Polygonia* species shows much less overlap between plant-specific modules in the three species (Supplementary Figures S9 and S10). This is consistent with gene expression on *Salix* being intermediate to expression on the other two plants in *Polygonia* (cf. Figures 3 and 4).

Polygonia satyrus was only reared on *Urtica*, and similar plant treatment correlations thus cannot be formally tested. However, it is interesting to note that the major *Urtica*-correlated modules in *P. c-album* (Pca_m2, Pca_m4) and *P. faunus* (Pfa_m1, Pfa_m2), for both species show strong and significant gene overlap with the same *P. satyrus* module (Psa_m1; Supplementary Figure S11). Similarly, the *Ribes*-correlated modules in the two polyphagous species show significant gene overlap with the same modules in *P. satyrus* (in particular Psa_m12; Supplementary Figure S11 and cf. Supplementary Figure S12). Other modules in *P. c-album* and *P. faunus* that overlap strongly with each other also typically overlap with the same module in *P. satyrus* (Supplementary Figure S11 and cf. Supplementary Figure S12). This suggests that similar gene networks are present also in the separate analysis of the specialist, meaning that they are not a product of including two plant treatments in the data for WGCNA analysis but rather may be related to actual functional gene modules shared among these related butterflies (supporting Hypothesis 4).

Gene modularity

Comparing the gene clusters found in *P. c-album* and *P. faunus*, and the plant-specific differentially expressed genes in *N. xanthomelas*, with the WGCNA "modules" (i.e., gene co-expression networks) confirms that these plant-specific gene sets overlap significantly (Fisher's exact test) with the gene sets in the corresponding plant-specific modules both within and across species (supporting Hypotheses 2–4; Supplementary Figures S13 and S14). This was true especially for the modules most strongly correlated to upregulation in a plant treatment.

Thus, the biological processes described above for genes in the clusters should also broadly apply to the gene modules. This is supported by a closer manual inspection of the genes with the highest scores for membership in the module or highest correlation to the host variable (Supplementary Table S6A–D). First, the main *Urtica*-correlated modules in *P. c-album* and *P. faunus* are both characterized by genes with the biological functions found for the "*Urtica* up" clusters (see above), in particular regulation of transcription and mRNA splicing (Supplementary Table S6A–B). They are also to some extent characterized by defense/immune functions. The single significant *Ribes*-correlated module in *P. faunus* (Pfa_m14) is characterized by genes with functions related to respiration and stress responses (Supplementary Table S6B), but the module clearly also contains many genes related to metabolism, transport, and translation/ribosome biogenesis, overall similar to functions found above for the "*Ribes* up" clusters. The evidence for this is the significant gene overlap with all five *Ribes*-correlated modules in *P. c-album* where these functions are seen in the inspected genes (Supplementary Figure S7; Supplementary Table S6A and below).

There are some finer details on gene modularity which are not captured by these broad patterns of plant-specificity. First, not all *Urtica*-correlated or *Ribes*-correlated modules are characterized by the same gene functions. This is particularly true for the five *Ribes*-correlated modules in *P. c-album*. Whereas three of the modules are similar in containing top genes related to metabolism, respiration, and transport, one (Pca_m11) is strongly characterized by ribosome biosynthesis and another (Pca_m12) by genes coding for vacuolar ATPase subunits (Supplementary Table S6A). These patterns are also reflected by similar modularity in the other two butterfly species. For instance, Pca_m11 in *P. c-album* overlaps strongly with Nxa_m9 of *N. xanthomelas* and with Psa_m9 of *P. satyrus* (Supplementary Figures S8A and S11A).

Other putative functional modules do not align with the broad plant specificity of gene expression shared among butterfly species. One clear example is the Pca_m1 module in *P. c-album*, which is the one most strongly correlated with *Urtica* in this species. We tentatively interpret this module as related to the development and function of the visceral muscles of the midgut (Supplementary Table S6). It shows strong gene overlap with Pfa_m6 in *P. faunus* and Nxa_m11 in *N. xanthomelas* (none of which is significantly correlated to plant treatment) as well as with Psa_m3 in *P. satyrus* (Supplementary Figures S7, S8A, S11A and cf. Figure 6; Supplementary Figure S12).

Finally, as noted above genes with defense functions are seen in many "*Urtica* up" clusters and *Urtica*-correlated networks, and are also among the very most upregulated genes on this plant. None of the analysis methods well captures a strict-sense functional module related to defense, but this may rather be a shortcoming of the methods. After all, among the genes observed to be upregulated are several actors in the defense-related Toll pathway (Valanne et al., 2011), such as peptidoglycan recognition proteins, Toll-activating genes, and several antibacterial peptides.

Discussion

Going back to the hypotheses outlined in the Introduction, we found evidence of plant-specific transcriptome plasticity (broad-sense modularity; Hypothesis 1), presumably largely because of differences in plant chemistry among the distantly related plants used as food treatments. The plastic responses were to a large extent shared among related

butterfly species (Hypothesis 2), and this was true even for species where the plant in question is no longer a host in the wild (Hypothesis 3). We also found some evidence of shared strict-sense genetic modules in response to plant treatments (Hypothesis 4).

We found distinctive non-overlapping modules for *Urtica* and *Ribes* gene expression, but not in the case of *Salix*. This result is not compatible with a simple positive correlation between time since colonization and degree of modularity, since of the three hosts *Ribes* is the one most recently colonized (Celorio-Mancera et al., 2013; Gamberale-Stille et al., 2019). However, even the first colonization of *Ribes* apparently happened several millions of years ago, as *Ribes* is used as a host by several species of *Polygonia* in separate clades, although not by other nymphalid butterflies (cf. Gamberale-Stille et al., 2019; dating from Chazot et al. [2021]). Thus, there has been plenty of time to evolve host-specific modularity in all of the cases explored here.

The shared responses among species mean that these plastic responses to specific plants have been evolutionarily conserved at least to some extent over millions of years. It is striking that this similarity in responses was robust enough to be observed in both specialists and generalist butterflies, and even though rearings were performed in two different labs in Sweden and Canada, at different temperatures, and for two of the plant treatments using congeneric plants rather than the exact same species. Moreover, the similar transcriptomes on *Urtica dioica* were seen even in two butterfly species that no longer use this ancestral host in the field (*P. faunus* and *N. xanthomelas*). Alternative predictions could well have been that in these two species the evolved response to *Urtica* would have been lost to drift and mutation, or perhaps that this non-host would elicit a generalized stress response or a universal genetic mechanism for coping with novel or suboptimal plants (Celorio-Mancera et al., 2016; Mathieu-Bégné et al., 2022). In any case, this would have been a response very different from that observed in the two species for which it is the preferred host. Instead, we observed shared plant-specific patterns of gene expression in all four species, providing a possible mechanism for relatively easy re-colonization of *Urtica* or relatives, particularly in *P. faunus,* where hatchlings have been shown to readily accept *Urtica* as a resource in the laboratory (Nylin et al., 2015).

It has been a long-standing enigma how phytophagous insects and other organisms with symbiotic lifestyles evidently can retain a repertoire of hosts over long evolutionary time spans, on occasion losing them as actual hosts but still keeping them as potential hosts that can be re-colonized at a later time. This ability can be seen in establishment tests with juveniles on non-hosts (Janz et al., 2001; Larose et al., 2019; Lehnert & Scriber, 2012; Nylin et al., 2015; Scriber et al., 2008). In phylogenetic reconstructions, it is reflected by shifts back to ancestral hosts, and/or repeated colonizations, losses and recolonizations of the same limited set of host taxa (e.g., Agosta & Klemens, 2009; Janz et al., 2001; Nylin & Wahlberg, 2008; Nyman et al., 2019; Stigenberg et al., 2015; Weingartner et al., 2006; Wilcox et al., 2021). The shared broad-sense modularity in gene expression among related species shown here could potentially explain both phylogenetic conservatism in host use and phylogenetic recurrence of hosts.

It should be noted that our analysis is restricted to genes with identified orthologues across all four species. Furthermore, larvae were reared on the same plant from hatchling until they were sacrificed in the fourth instar, meaning that the responses would have included any downstream effects of the host affecting growth and development, not only direct responses to plant chemistry. This means that the observed patterns of gene expression are not likely to be due to genes with highly plant-specific functions, including members of important rapidly evolving gene families such as detoxifying cytochrome p450s (Calla et al., 2017; Scott & Wen, 2001) or chemosensory genes (Briscoe et al., 2013). The reported transcriptomes rather reflect how the overall conserved genetic "machinery" is affected by exposure of juveniles to a particular plant resource, changing gene expression in an attempt to sustain homeostasis (Petre et al., 2020). This is in line with recent findings that host plant adaptation is likely to involve many different genes and pathways (Celorio-Mancera et al., 2013; Chaturvedi et al., 2018). Similarly, transcriptomic adjustments involving core functions such as cellular machinery or energetic metabolism permit host shifts in a crustacean ectoparasite on fish (Mathieu-Bégné et al., 2022).

While the aforementioned caveats are important, they do not take away from the fact that the ability to respond similarly—and presumably adaptively—to a particular plant resource seems to be conserved among related species and remain long after it has been lost as an actual host, provided that the insect lineage had enough time to evolve the modular response in the first place (Hypotheses 2–3). Rather, the nature of the dataset suggests an explanation for how plant-specific modules of gene expression can be kept even when juveniles are no longer exposed to the host. Modularity may help adaptive plasticity to evolve but do not as such prevent these genetic adaptations from being lost to mutation accumulation (Snell-Rood et al., 2010). However, although genes with highly plant-specific functions might, under such circumstances, be lost (Edger et al., 2015), core genes with multiple functions would be likely to remain, and evolving modularity can facilitate such co-option of existing gene activity (Espinosa-Soto & Wagner, 2010). The situation would then be an additional example of when adaptive reaction norms are conserved over long time spans so that the appropriate phenotypes can be expressed via an upstream developmental switch according to the environment. Well-known examples of such developmental plasticity shared across species in a clade—but not always expressed in a given species—include temperature-dependent sex determination in reptiles (Merchant-Larios & Diaz-Hernandez, 2013), temperature- and/or photoperiod-dependent seasonal forms in butterflies (Nylin et al., 2005; van Bergen et al., 2017), and nutrition-dependent determination of caste in social insects (West-Eberhard, 2003; Maleszka, 2008; Rajakumar et al., 2012; West-see also Lafuente and Beldade [2019] for a review).

One relevant question is to what extent plant-specific transcriptomes are analogous to the developmental plasticity in the literature cited above, which mainly concerns morphological phenotypes (albeit with a physiological background). This was discussed by Nylin and Janz (2009), who argued that the multivariate nature of host plant utilization certainly makes for a more complex phenotype, but that this rather increases the scope for plasticity and developmental switches since single genes will rarely if ever determine host use (cf. Chaturvedi et al., 2018). Similarly, the concept of modularity in biology was originally applied to morphological traits but has been extended to functional traits such as protein–protein

interactions and gene regulation (Wagner et al., 2007). Indeed, West-Eberhard (2003) defines a modular trait as any subunit of the phenotype determined by a decision point in development, whether the endpoint is morphology, physiology, or behavior, and it would seem likely that such decision points are involved in shaping plant-specific transcriptomes.

We found not only a tight correlation of gene expression patterns in PCA space across species at least on *Urtica* and *Ribes* but also that many of the exact same genes are strongly upregulated on a given plant. This is a strong indication that not only broad-sense modularity is involved, but that genetic modules in the stricter sense are switched on or off depending on the plant resource (Hypothesis 4). Identification and description of these modules is mostly beyond the scope of the present manuscript. This is due not the least to the limitations of the dataset, which is from a single time-point and tissue and thus cannot distinguish between up- and downstream patterns. It is possible that a few high-level regulatory factors could act as "developmental switches" driving the plastic responses, but this cannot be determined from the present data.

It, however, seems clear that one or more of the genetic modules upregulated on *Urtica* is involved in the regulation of transcription, mRNA splicing, and development, whereas modules upregulated on alternative plant resources (*Salix* or *Ribes*) are more related to metabolism, detoxification, stress responses, and respiration. The consistency of the latter responses across species is somewhat unexpected given that these "alternative hosts" are the preferred hosts (over *Urtica*) in two of the species. It is, however, in agreement with the interpretation that no strong barrier to feeding on the ancestral host has evolved, whereas *Salix* and *Ribes* do present some challenges as a diet, for all of the species. We also found evidence that modules involved in defense against microorganisms are upregulated on *Urtica*, a pattern also found in other comparative transcriptome studies (Hou et al., 2021; Petre et al., 2020). Since *Urtica* is a shared host with several other related butterflies, sharing also pathogens, it may well be an environment where infection risk is high. In contrast, a prostaglandin reductase is strongly upregulated in the other plant treatments, which may serve to dampen the immune response (Knight, 2020). It is not clear whether the upregulation of defense genes on *Urtica* is a direct response to the presence of pathogens or an example of immune anticipation (Zhong et al., 2013).

We posit that the distinct and opposing patterns of gene expression on *Urtica* and *Ribes* in *P. c-album* and *P. faunus* (see also Celorio-Mancera et al., 2013) suggest alternative developmental pathways being followed by larvae reared on the respective plants. It is tempting to further speculate that this may result in trade-offs preventing optimal use of both plants as hosts, which could be a factor in explaining why only one nymphalid species (*P. c-album*) normally use both plants as hosts. One clade of *Polygonia* species in the Nearctic has shifted more or less completely to *Ribes* (Figure 1A), at least in some species even to the extent that they no longer are able to feed on *Urtica* (Janz et al., 2001). In contrast, the weak modularity seen on *Salix*, and the high degree of overlap with gene expression on *Urtica* and *Ribes*, indicates that *Salix* may have functioned as an evolutionary "bridge" among the hosts of *Polygonia* butterflies, from the ancestral *Urtica* and relatives, over *Salix*, to the more challenging *Ribes*. Selection experiments to test this trade-off hypothesis are currently being carried out.

In conclusion, the study of plant-specific modularity in gene expression shows great promise as a tool in explaining insect–plant association evolutionary dynamics, and by extension other symbiotic species associations. Genetic adaptation to a host may be best understood not in terms of single genes but as having evolved genetic modules of co-expressed genes capable of dealing well with the plant as a resource, after an initial period where it can be used only poorly, as a byproduct of existing adaptations. In other words, initial colonization of a novel host happens through "ecological fitting" (Janzen, 1985) without genetic change; such change comes later as a result of selection for improved performance. If the new host is retained in the actual repertoire for long enough, this may result in a suite of functional genetic modules adapted "for" feeding on this particular host. Conserved modules of this kind may help explain the host plant "memory" of larvae that allow recurrence of ancestral host use, and strongly opposing modularity may indicate a degree of incompatibility among plant resources resulting in trade-offs, with evolutionary consequences for the likelihood of colonizations and host shifts.

Supplementary material

Supplementary material is available online at *Evolution* (https://academic.oup.com/evolut/qpac049)

Data availability

Relevant data associated with this manuscript (assembled transcriptomes and gene expression count data) is publicly available at the Harvard Dataverse (https://doi.org/10.7910/DVN/OMIPPK). The data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB55562 (study ERP140477) (https://www.ebi.ac.uk/ena/browser/view/PRJEB55562)

Author contributions

M.P.C.M., N.J., C.W.W., and S.N. designed the study. M.P.C.M., N.J., and S.N. collected butterflies and data. M.P.C.M., P.P., A.S., R.S., M.P.B., and S.N. carried out the analyses. M.P.C.M., R.S., and S.N. wrote this article with help from all other authors.

Funding statement

Financial support was from grants from the Swedish Research Council to S.N. (2015-04218 and 2019-03441) and to C.W.W. (2017-04386).

Conflict of interest: The authors have no conflict of interest

Acknowledgments

The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. Thanks to Lina Söderlind who conducted the rearings in Edmonton, Canada, and to Felix Sperling and colleagues for hosting and helping her. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b2016162.

References

Abdeltawab, A. A., Ullah, Z., Al-Othman, A. M., Ullah, R., Hussain, I., Ahmad, S., & Talha, M. (2012). Evaluation of the chemical composition and element analysis of *Urtica dioca*. *African Journal of Pharmacy and Pharmacology*, 6, 1555–1558. https://doi.org/10.5897/ajpp12.268

Agosta, S. J., Janz, N., & Brooks, D. R. (2010). How specialists can be generalists: Resolving the "parasite paradox" and implications for emerging infectious disease. *Zoologia*, 27, 151–162. https://doi.org/10.1590/S1984-46702010000200001

Agosta, S. J., & Klemens, J. A. (2009). Resource specialization in a phytophagous insect: No evidence for genetically based performance trade-offs across hosts in the field or laboratory. *Journal of Evolutionary Biology*, 22, 907–912. https://doi.org/10.1111/j.1420-9101.2009.01694.x

Bale, J. S., Masters, G. J., Hodkinson, I. D., Awmack, C., Bezemer, T. M., Brown, V. K., Butterfield, J., Buse, A., Coulson, J. C., Farrar, J., Good, J. E. G., Harrington, R., Hartley, S., Jones, T. H., Lindroth, R. L., Press, M. C., Symrnioudis, I., Watt, A. D., & Whittaker, J. B. (2002). Herbivory in global climate change research: Direct effects of rising temperature on insect herbivores. *Global Change Biology*, 8, 1–16. https://doi.org/10.1046/j.1365-2486.2002.00451.x

Becerra, J. X. (1997). Insects on plants: Macroevolutionary chemical trends in host use. *Science*, 276, 253–256. https://doi.org/10.1126/science.276.5310.253

Birnbaum, S. S. L., & Abbot, P. (2020). Gene expression and diet breadth in plant-feeding insects: Summarizing trends. *Trends in Ecology & Evolution*, 35, 259–277. https://doi.org/10.1016/j.tree.2019.10.014

Bjarnholt, N., & Moller, B. L. (2008). Hydroxynitrile glucosides. *Phytochemistry*, 69, 1947–1961. https://doi.org/10.1016/j.phytochem.2008.04.018

Braga, M. P., Araujo, S. B. L., Agosta, S., Brooks, D., Hoberg, E., Nylin, S., Janz, N., & Boeger, W. A. (2018a). Host use dynamics in a heterogeneous fitness landscape generates oscillations in host range and diversification. *Evolution*, 72, 1773–1783. https://doi.org/10.1111/evo.13557

Braga, M. P., Guimaraes, P. R., Wheat, C. W., Nylin, S., & Janz, N. (2018b). Unifying host-associated diversification processes using butterfly-plant networks. *Nature Communications*, 9. https://doi.org/10.1038/s41467-018-07677-x

Briscoe, A. D., Macias-Munoz, A., Kozak, K. M., Walters, J. R., Yuan, F. R., Jamie, G. A., Martin, S. H., Dasmahapatra, K. K., Ferguson, L. C., Mallet, J., Jacquin-Joly, E., & Jiggins, C. D. (2013). Female behaviour drives expression and evolution of gustatory receptors in butterflies. *PLoS Genetics*, 9. https://doi.org/10.1371/journal.pgen.1003620

Brooks, D. R., Hoberg, E. P. & Boeger, W. A. (2019). *The Stockholm paradigm: Climate change and emerging disease*. The University of Chicago Press.

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12, 59–60. https://doi.org/10.1038/nmeth.3176

Calla, B., Noble, K., Johnson, R. M., Walden, K. K. O., Schuler, M. A., Robertson, H. M., & Berenbaum, M. R. (2017). Cytochrome P450 diversification and hostplant utilization patterns in specialist and generalist moths: Birth, death and adaptation. *Molecular Ecology*, 26, 6021–6035. https://doi.org/10.1111/mec.14348

Celorio-Mancera, M. D., Wheat, C. W., Huss, M., Vezzi, F., Neethiraj, R., Reimegard, J., Nylin, S., & Janz, N. (2016). Evolutionary history of host use, rather than plant phylogeny, determines gene expression in a generalist butterfly. *BMC Evolutionary Biology*, 16. https://doi.org/10.1186/s12862-016-0627-y

Celorio-Mancera, M. D., 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, 4884–4895. https://doi.org/10.1111/mec.12440

Challis, R. J., Kumar, S., Dasmahapatra, K. K., Jiggins, C. D., & Blaxter, M. (2016). Lepbase: The Lepidopteran genome database. *bioRxiv*, 056994. https://doi.org/10.1101/056994

Chaturvedi, S., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Gompert, Z. (2018). The predictability of genomic changes underlying a recent host shift in Melissa blue butterflies. *Molecular Ecology*, 27, 2651–2666. https://doi.org/10.1111/mec.14578

Chazot, N., Condamine, F. L., Dudas, G., Pena, C., Kodandaramaiah, U., Matos-Maravi, P., Aduse-Poku, K., Elias, M., Warren, A. D., Lohman, D. J., Penz, C. M., DeVries, P., Fric, Z. F., Nylin, S., Muller, C., Kawahara, A. Y., Silva-Brandao, K. L., Lamas, G., Kleckova, I., … Wahlberg, N. (2021). Conserved ancestral tropical niche but different continental histories explain the latitudinal diversity gradient in brush-footed butterflies. *Nature Communications*, 12. https://doi.org/10.1038/s41467-021-25906-8

Davey, J. W., Chouteau, M., Barker, S. L., Maroja, L., Baxter, S. W., Simpson, F., Joron, M., Mallet, J., Dasmahapatra, K. K., & Jiggins, C. D. (2016). Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. *G3: Genes, Genomes, Genetics*, 6, 695–708. https://doi.org/10.1534/g3.115.023655

Dyer, L. A., Singer, M. S., Lill, J. T., Stireman, J. O., Gentry, G. L., Marquis, R. J., Ricklefs, R. E., Greeney, H. F., Wagner, D. L., Morais, H. C., Diniz, I. R., Kursar, T. A., & Coley, P. D. (2007). Host specificity of Lepidoptera in tropical and temperate forests. *Nature*, 448, 696–699. https://doi.org/10.1038/nature05884

Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Gloeckner, G., Platts, A. E., Heckel, D. G., Der, J. P., Wafula, E. K., Tang, M., Hofberger, J. A., Smithson, A., Hall, J. C., Blanchette, M., Bureau, T. E., Wright, S. I., dePamphilis, C. W., Schranz, M. E., Barker, M. S., … Wheat, C. W. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 8362–8366. https://doi.org/10.1073/pnas.1503926112

Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and plants: A study in coevolution. *Evolution*, 18, 586–608. https://doi.org/10.2307/2406212

Erb, M., & Reymond, P. (2019). Molecular interactions between plants and insect herbivores. *Annual Review of Plant Biology*, 70, 527–557. https://doi.org/10.1146/annurev-arplant-050718-095910

Espinosa-Soto, C., & Wagner, A. (2010). Specialization can drive the evolution of modularity. *PLoS Computational Biology*, 6, e1000719. https://doi.org/10.1371/journal.pcbi.1000719

Forister, M. L., Dyer, L. A., Singer, M. S., Stireman, J. O., & Lill, J. T. (2012). Revisiting the evolution of ecological specialization, with emphasis on insect-plant interactions. *Ecology*, 93, 981–991. https://doi.org/10.1890/11-0650.1

Forister, M. L., Novotny, V., Panorska, A. K., Baje, L., Basset, Y., Butterill, P. T., Cizek, L., Coley, P. D., Dem, F., Diniz, I. R., Drozd, P., Fox, M., Glassmire, A. E., Hazen, R., Hrcek, J., Jahner, J. P., Kaman, O., Kozubowski, T. J., Kursar, T. A., … Dyer, L. A. (2015). The global distribution of diet breadth in insect herbivores. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 442–447. https://doi.org/10.1073/pnas.1423042112

Fu, L. M., Niu, B. F., Zhu, Z. W., Wu, S. T., & Li, W. Z. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28, 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Futuyma, D. J., & Agrawal, A. A. (2009). Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 18054–18061. https://doi.org/10.1073/pnas.0904106106

Gamberale-Stille, G., Schäpers A., Janz, N., & Nylin, S. (2019). Selective attention by priming in host search behavior of 2 generalist butterflies. *Behavioral Ecology*, 30, 142–149. https://doi.org/10.1093/beheco/ary146

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. D., Chen, Z. H., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & Regev, A. (2011). Full-length transcriptome assembly from

RNA-Seq data without a reference genome. *Nature Biotechnology* 29:644-U130. https://doi.org/10.1038/nbt.1883

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hallgren, P., Ikonen, A., Hjalten, J., & Roininen, H. (2003). Inheritance patterns of phenolics in F1, F2, and back-cross hybrids of willows: Implications for herbivore responses to hybrid plants. *Journal of Chemical Ecology*, 29, 1143–1158. https://doi.org/10.1023/A:1023829506473

Hegnauer, R. (1990). *Chemotaxonomie der Pflanzen: eine Übersicht über die Verbreitung und die systematische Bedeutung der Pflanzenstoffe. Bd 9, Nachträge zu Band 5 und Band 6 (Magnoliaceae bis Zygophyllaceae)*. Birkhauser.

Heidel-Fischer, H. M., & Vogel, H. (2015). Molecular mechanisms of insect adaptation to plant secondary compounds. *Current Opinion in Insect Science*, 8, 8–14. https://doi.org/10.1016/j.cois.2015.02.004

Ho, W. C., Li, D. Y., Zhu, D., & Zhang, J. Z. (2020). Phenotypic plasticity as a long-term memory easing readaptations to ancestral environments. *Science Advances*, 6. https://doi.org/10.1126/sciadv.aba3388

Hou, Z. H., Shi, F. M., Ge, S. X., Tao, J., Ren, L. L., Wu, H., & Zong, S. X. (2021). Comparative transcriptome analysis of the newly discovered insect vector of the pine wood nematode in China, revealing putative genes related to host plant adaptation. *BMC Genomics*, 22. https://doi.org/10.1186/s12864-021-07498-1

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oles, A. K., … Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, 12, 115–121. https://doi.org/10.1038/NMETH.3252

Jain, A., & Dubes, R. (1988). *Algorithms for clustering data*. Prentice Hall.

Janz, N. 2011. Ehrlich and Raven revisited: Mechanisms underlying codiversification of plants and enemies. In D. J. Futuyma, H. B. Shaffer, & D. Simberloff (Eds.), *Annual review of ecology, evolution, and systematics* (Vol 42, pp. 71–89). Annual Reviews. https://doi.org/10.1146/annurev-ecolsys-102710-145024

Janz, N., Nyblom, K., & Nylin, S. (2001). Evolutionary dynamics of host-plant specialization: A case study of the tribe Nymphalini. *Evolution*, *55*, 783–796. https://doi.org/10.1554/0014-3820(2001)055[0783:edohps]2.0.co;2

Janz, N. & Nylin S. (2008). The oscillation hypothesis of host-plant range and speciation. In K. J. Tilmon (Ed.), *Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects* (pp. 203–215). University of California Press.

Janzen, D. H. (1985). On ecological fitting. *Oikos*, 45, 308–310. https://doi.org/10.2307/3565565

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., & von Mering, C. (2009). STRING 8-a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research*, 37, D412–D416. https://doi.org/10.1093/nar/gkn760

Joshi, A., & Thompson, J. N. (1995). Trade-offs and the evolution of host specialization. *Evolutionary Ecology*, 9, 82–92. https://doi.org/10.1007/BF01237699

Kavtaradze, N. S., Alaniya, M. D., & Aneli, J. N. (2001). Chemical components of *Urtica dioica* growing in Georgia. *Chemistry of Natural Compounds*, 37, 287–287. https://doi.org/10.1023/A:1012594713302

Knight, K. (2020). Beet armyworms keep prostaglandins under control with dehydrogenase and reductase. *Journal of Experimental Biology*, 223, 3. https://doi.org/10.1242/jeb.238394

Kumar, L., & Futschik, M. (2007). Mfuzz: A software package for soft clustering of microarray data. *Bioinformation*, 2, 5–7. https://doi.org/10.6026/97320630002005

Lafuente, E., & Beldade, P. (2019). Genomics of developmental plasticity in animals. *Frontiers in Genetics*, 1. https://doi.org/10.3389/fgene.2019.00720

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559. https://doi.org/10.1186/1471-2105-9-559

Larkin, A., Marygold, S. J., Antonazzol, G., Attrill, H., dos Santos, G., Garapati, P. V., Goodman, J. L., Gramates, L. S., Millburn, G., Strelets, V. B., Tabone, C. J., Thurmond, J., & FlyBase, C. (2021). FlyBase: Updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Research*, 49, D899–D907. https://doi.org/10.1093/nar/gkaa1026

Larose, C., Rasmann, S., & Schwander, T. (2019). Evolutionary dynamics of specialisation in herbivorous stick insects. *Ecology Letters*, 22, 354–364. https://doi.org/10.1111/ele.13197

Lehnert, M. S., & Scriber, J. M. (2012). Salicaceae detoxification abilities in Florida tiger swallowtail butterflies (*Papilio glaucus maynardi* Gauthier): Novel ability or Pleistocene holdover? *Insect Science*, 19, 337–345. https://doi.org/10.1111/j.1744-7917.2011.01459.x

Li, W. Z., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Liebhold, A. M., Macdonald, W. L., Bergdahl, D., & Maestro, V. C. (1995). Invasion by exotic forest pests-a threat to forest ecosystems. *Forest Science*, 41, 1–49. https://doi.org/10.1093/forestscience/41.s1.a0001

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Maleszka, R. (2008). Epigenetic integration of environmental and genomic signals in honey bees. *Epigenetics*, 3, 188–192. https://doi.org/10.4161/epi.3.4.6697

Mathieu-Bégné, E., Blanchet, S., Mitta, G., Le Potier, C., Loot, G., & Rey, O. (2022). Transcriptomic adjustments in a freshwater ectoparasite reveal the role of molecular plasticity for parasite host shift. *Genes*, 13, 525. https://doi.org/10.3390/genes13030525

Melo, D., Porto, A., Cheverud, J. M., & Marroig, G. (2016). Modularity: Genes, development and evolution. *Annual Review of Ecology, Evolution, and Systematics*, 47, 463–486. https://doi.org/10.1146/annurev-ecolsys-121415-032409

Merchant-Larios, H., & Diaz-Hernandez, V. (2013). Environmental sex determination mechanisms in reptiles. *Sexual Development*, 7, 95–103. https://doi.org/10.1159/000341936

Mitter, C., Farrell, B., & Futuyma, D. J. (1991). Phylogenetic studies of insect-plant interactions-insights into the genesis of diversity. *Trends in Ecology & Evolution*, 6, 290–293. https://doi.org/10.1016/0169-5347(91)90007-K

Nosil, P. (2002). Transition rates between specialization and generalization in phytophagous insects. *Evolution*, 56, 1701–1706. https://doi.org/10.1111/j.0014-3820.2002.tb01482.x

Nylin, S. (1988). Host plant specialization and seasonality in a polyphagous butterfly, *Polygonia c-album* (Nymphalidae). *Oikos*, 53, 381–386. https://doi.org/10.2307/3565539

Nylin, S., Agosta, S., Bensch, S., Boeger, W. A., Braga, M. P., Brooks, D. R., Forister, M. L., Hamback, P. A., Hoberg, E. P., Nyman, T., Schapers, A., Stigall, A. L., Wheat, C. W., Osterling, M., & Janz, N. (2018). Embracing colonizations: A new paradigm for species association dynamics. *Trends in Ecology & Evolution*, 33, 4–14. https://doi.org/10.1016/j.tree.2017.10.005

Nylin, S., Gotthard, K., & Nygren, G. H. (2005). Seasonal plasticity, host plants, and the origin of butterfly biodiversity. In M. D. E. Fellowes, G. J. Holloway, & J. Rolff (Eds.), *Insect evolutionary ecology* (pp. 111–137). CABI Publishing.

Nylin, S., & Janz, N. (2009). Butterfly host plant range: An example of plasticity as a promoter of speciation? *Evolutionary Ecology*, 23, 137–146. https://doi.org/10.1007/s10682-007-9205-5

Nylin, S., Slove, J., & Janz, N. (2014). Host plant utilization, host range oscillations and diversification in nymphalid butterflies: A phylogenetic investigation. *Evolution*, 68, 105–124. https://doi.org/10.1111/evo.12227

Nylin, S., Söderlind, L., Gamberale-Stille, G., Audusseau, H., Celorio-Mancera, M. D. L. P., Janz, N., & Sperling, F. A. H. (2015). Vestiges of an ancestral host plant: Preference and performance in the butterfly *Polygonia faunus* and its sister species *P. c-album*. *Ecological Entomology*, 40, 307–315. https://doi.org/10.1111/een.12187

Nylin, S., & Wahlberg, N. (2008). Does plasticity drive speciation? Host plant shifts and diversification in nymphaline butterflies (Lepidoptera: Nymphalidae) during the tertiary. *Biological Journal of the Linnean Society*, 94, 115–130. https://doi.org/10.1111/j.1095-8312.2008.00964.x

Nyman, T. (2010). To speciate, or not to speciate? Resource heterogeneity, the subjectivity of similarity, and the macroevolutionary consequences of niche-width shifts in plant-feeding insects. *Biological Reviews*, 85, 393–411. https://doi.org/10.1111/j.1469-185X.2009.00109.x

Nyman, T., Onstein, R. E., Silvestro, D., Wutke, S., Taeger, A., Wahlberg, N., Blank, S. M., & Malm, T. (2019). The early wasp plucks the flower: Disparate extant diversity of sawfly superfamilies (Hymenoptera: "Symphyta') may reflect asynchronous switching to angiosperm hosts. *Biological Journal of the Linnean Society*, 128, 1–19. https://doi.org/10.1093/biolinnean/blz071

Pal, N. R., Bezdek, J. C., & Hathaway, R. J. (1996). Sequential competitive learning and the fuzzy c-means clustering algorithms. *Neural Networks*, 9, 787–796. https://doi.org/10.1016/0893-6080(95)00094-1

Petre, B., Lorrain, C., Stukenbrock, E. H., & Duplessis, S. (2020). Host-specialized transcriptome of plant-associated organisms. *Current Opinion in Plant Biology*, 56, 81–88. https://doi.org/10.1016/j.pbi.2020.04.007

Rajakumar, R., San Mauro, D., Dijkstra, M. B., Huang, M. H., Wheeler, D. E., Hiou-Tim, F., Khila, A., Cournoyea, M., & Abouheif, E. (2012). Ancestral developmental potential facilitates parallel evolution in ants. *Science*, 335, 79–82. https://doi.org/10.1126/science.1211451

R_Core_Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ritchie, M., Phipson, B., Wu, D., Hu, Y. L., Law, C., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43, e47. https://doi.org/10.1093/nar/gkv007

Robinson, M., McCarthy, D., & Smyth, G. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Ronquist, F., & Liljeblad, J. (2001). Evolution of the gall wasp-host plant association. *Evolution*, 55, 2503–2522. https://doi.org/10.1111/j.0014-3820.2001.tb00765.x

Schlosser, G. & Wagner G. P. (Eds.). 2004. *Modularity in development and evolution*. University of Chicago Press.

Scott, J. G., & Wen, Z. M. (2001). Cytochromes P450 of insects: The tip of the iceberg. *Pest Management Science*, 57, 958–967. https://doi.org/10.1002/ps.354

Scriber, J. M., Larsen, M. L., Allen, G. R., Walker, P. W., & Zalucki, M. P. (2008). Interactions between Papilionidae and ancient Australian Angiosperms: Evolutionary specialization or ecological monophagy? *Entomologia Experimentalis et Applicata*, 128, 230–239. https://doi.org/10.1111/j.1570-7458.2008.00691.x

Shen, L. (2021). GeneOverlap: Test and visualize gene overlaps. R package version 1.28.0. http://shenlab-sinai.github.io/shenlab-sinai/.

Snell-Rood, E. C., Van Dyken, J. D., Cruickshank, T., Wade, M. J., & Moczek, A. P. (2010). Toward a population genetic framework of developmental evolution: The costs, limits, and consequences of phenotypic plasticity. *Bioessays*, 32, 71–81. https://doi.org/10.1002/bies.200900132

Speed, M. P., Fenton, A., Jones, M. G., Ruxton, G. D., & Brockhurst, M. A. (2015). Coevolution can explain defensive secondary metabolite diversity in plants. *New Phytologist*, 208, 1251–1263. https://doi.org/10.1111/nph.13560

Stigenberg, J., Boring, C. A., & Ronquist, F. (2015). Phylogeny of the parasitic wasp subfamily Euphorinae (Braconidae) and evolution of its host preferences. *Systematic Entomology*, 40, 570–591. https://doi.org/10.1111/syen.12122

The Heliconius Genome Consortium. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487, 94–98. https://doi.org/10.1038/nature11041

The UniProt Consortium. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49, D480–D489. https://doi.org/10.1093/nar/gkaa1100

Thevenot, E. A., Roux, A., Xu, Y., Ezan, E., & Junot, C. (2015). Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *Journal of Proteome Research*, 14, 3322–3335. https://doi.org/10.1021/acs.jproteome.5b00354

Torres-Martinez, L., Porter, S. S., Wendlandt, C., Purcell, J., Ortiz-Barbosa, G., Rothschild, J., Lampe, M., Warisha, F., Le, T., Weisberg, A. J., Chang, J. H., & Sachs, J. L. (2021). Evolution of specialization in a plant-microbial mutualism is explained by the oscillation theory of speciation. *Evolution*, 75, 1070–1086. https://doi.org/10.1111/evo.14222

Valanne, S., Wang, J. -H., & Rämet M. (2011). The *Drosophila* toll signaling pathway. *The Journal of Immunology*, 186, 649–656. https://doi.org/10.4049/jimmunol.1002302

van Bergen, E., Osbaldeston, D., Kodandaramaiah, U., Brattstrom, O., Aduse-Poku, K., & Brakefield, P. M. (2017). Conserved patterns of integrated developmental plasticity in a group of polyphenic tropical butterflies. *BMC Evolutionary Biology*, 17. https://doi.org/10.1186/s12862-017-0907-1

van der Linden, C. F. H., WallisDeVries, M. F., & Simon, S. (2021). Great chemistry between us: The link between plant chemical defenses and butterfly evolution. *Ecology and Evolution*, 11, 8595–8613. https://doi.org/10.1002/ece3.7673

Vertacnik, K. L., & Linnen, C. R. (2017). Evolutionary genetics of host shifts in herbivorous insects: Insights from the age of genomics. *Annals of the New York Academy of Sciences*, 1389, 186–212. https://doi.org/10.1111/nyas.13311

Wagner, G. P., Pavlicev, M., & Cheverud, J. M. (2007). The road to modularity. *Nature Reviews Genetics*, 8, 921–931. https://doi.org/10.1038/nrg2267

Wahlberg, N. (2001). The phylogenetics and biochemistry of host-plant specialization in Melitaeine butterflies (Lepidoptera: Nymphalidae). *Evolution*, 55, 522–537. https://doi.org/10.1554/0014-3820(2001)055[0522:TPABOH]2.0.CO;2

Weingartner, E., Wahlberg, N., & Nylin, S. (2006). Dynamics of host plant use and species diversity in *Polygonia* butterflies (Nymphalidae). *Journal of Evolutionary Biology*, 19, 483–491. https://doi.org/10.1111/j.1420-9101.2005.01009.x

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.

Wilcox, J. J. S., Lopez-Cotto, J. J., & Hollocher, H. (2021). Historical contingency, geography, and anthropogenic patterns of exposure drive the evolution of host-switching in the *Blastocystis* species-complex. *Parasitology*, 148, 1–37. https://doi.org/10.1017/S003118202100055X

Zhong, W. H., McClure, C. D., Evans, C. R., Mlynski, D. T., Immonen, E., Ritchie, M. G., & Priest, N. K. (2013). Immune anticipation of mating in *Drosophila*: Turandot M promotes immunity against sexually transmitted fungal infections. *Proceedings of the Royal Society B-Biological Sciences*, 280. https://doi.org/10.1098/rspb.2013.2018