MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 4884–4895 doi: 10.1111/mec.12440

Mechanisms of macroevolution: polyphagous plasticity in butterfly larvae revealed by RNA-Seq

MARIA DE LA PAZ CELORIO-MANCERA,* CHRISTOPHER W. WHEAT,* HEIKO VOGEL,† LINA SÖDERLIND,* NIKLAS JANZ* and SÖREN NYLIN*

Department of Zoology Ecology, Stockholm University, Svante Arrheniusväg 18 B, 106 91 Stockholm, Sweden, †Department of Entomology, Max Planck Institute for Chemical Ecology, Beutenberg Campus, Hans-Knöll Straße 8 07745, Jena, Germany

Abstract

Transcriptome studies of insect herbivory are still rare, yet studies in model systems have uncovered patterns of transcript regulation that appear to provide insights into how insect herbivores attain polyphagy, such as a general increase in expression breadth and regulation of ribosomal, digestion- and detoxification-related genes. We investigated the potential generality of these emerging patterns, in the Swedish comma, *Polygonia c-album*, which is a polyphagous, widely-distributed butterfly. *Urtica dioica* and *Ribes uva-crispa* are hosts of *P. c-album*, but *Ribes* represents a recent evolutionary shift onto a very divergent host. Utilizing the assembled transcriptome for read mapping, we assessed gene expression finding that caterpillar life-history (i.e. 2nd vs. 4th-instar regulation) had a limited influence on gene expression plasticity. In contrast, differential expression in response to host-plant identified genes encoding serine-type endopeptidases, membrane-associated proteins and transporters. Differential regulation of genes involved in nucleic acid binding was also observed suggesting that polyphagy involves large scale transcriptional changes. Additionally, transcripts coding for structural constituents of the cuticle were differentially expressed in caterpillars in response to their diet indicating that the insect cuticle may be a target for plant defence. Our results state that emerging patterns of transcript regulation from model species appear relevant in species when placed in an evolutionary context.

Keywords: host range, insect-plant interaction, major facilitator superfamily, RNA-Seq, serine-type endopeptidases, transcriptomics

Received 27 March 2013; revision received 20 June 2013; accepted 26 June 2013

Background

Insect herbivores display a dynamic range of evolutionary relationships with their host plants, ranging from generalist- to specialist-feeding interactions (Schoonhoven *et al.* 2005). Butterflies have long been one of the primary examples of diversification via co-evolution (Ehrlich & Raven 1964), with some clades frequently recolonizing a small number of host plants used previously in ancestral lineages (Futuyma *et al.* 1995; Janz *et al.* 2001). These oscillatory patterns appear to have helped generate species diversity (Janz *et al.* 2006; Janz & Nylin 2008). To improve our understanding of how these patterns emerge in herbivores, a functional understanding of plant-matter detoxification and assimilation is needed (Janz 2011). However, almost nothing is known of the molecular mechanisms underlying such recurrent host-plant shifts.

In general, the underlying genetic mechanisms used by insects in response to their host plants still remain poorly understood (Dermauw *et al.* 2013). However, recent findings suggest that insect herbivores, particularly generalists, use transcriptional plasticity to their advantage, regulating three main categories of genes in response to their diet: ribosomal, digestive and detoxifying genes.

Genes coding for ribosomal proteins are known to be differentially regulated in response to a variety of host plants in different insect species. In the gut tissue of the

Correspondence: Sören Nylin, Fax: (+46) 08-16 77 15; E-mail: Soren.Nylin@zoologi.su.se

cotton bollworm, *Helicoverpa armigera*, these genes are both up- and down-regulated when the insects feed on unsuitable hosts (Celorio-Mancera *et al.* 2012). In two other generalist species, most ribosomal genes were up-regulated: when the hemipteran *Bemisia tabaci* responded to a genetically modified host that was producing higher amounts of secondary defence metabolites (Alon *et al.* 2012) and when the lepidopteran *Heliothis virescens* responded to a host silenced for its ability to elicit jasmonate-mediated defences (Govind *et al.* 2010). Because ribosomal proteins are mostly associated with their standard role in protein translation, the genes encoding them have been considered stably expressed 'housekeeping' genes, and therefore, it has been surprising to learn that these genes are in fact differentially regulated.

While changes in ribosomal gene expression may reflect the consequences of the observed large-scale changes in transcription, two additional dynamics may account for these observations. First, plants produce ribosome-inactivating proteins (RIPs) that inhibit protein synthesis in eukaryotes (Puri *et al.* 2012). Indeed, RIPs have been found to have insecticidal activity in aphids (Shahidi-Noghabi *et al.* 2009), beetles (Dowd *et al.* 2003), mosquitoes (Zhou *et al.* 2000) and moths in the family Noctuidae (Zhou *et al.* 2000; Shahidi-Noghabi *et al.* 2009). Therefore, the differential regulation of ribosomal genes in the insect might help counteract the presence of RIPs in the food source. Second, ribosomal proteins in different biological systems, including *Drosophila melanogaster,* have been shown to contribute to the regulation of genes via their interaction with epigenetic factors; this interaction leads to chromatin remodelling (Surovtseva *et al.* 2011; Coleno-Costes *et al.* 2012). In this case, the differential regulation of ribosomal genes might be reflecting an epigenetic mechanism for regulating gene expression.

Among the different genes involved in insect digestion, those coding proteases in particular have been studied extensively, perhaps because nitrogen acquisition is essential for insect nutrition (Mattson 1980; Brown *et al.* 2002). Insects must be able to efficiently digest plant proteins, as is demonstrated by the fact that protease inhibitors are a widespread defence mechanism against herbivory among plants (Ryan 1990). Therefore, special attention has been directed towards understanding how insects known to have a wide diet respond to specific plant protease inhibitors (Liu *et al.* 2004; Moon *et al.* 2004; Chi *et al.* 2009; Govind *et al.* 2010; Li *et al.* 2010). This research field has revealed that phytophagous insects react to protease inhibitors by either up-regulating those inhibitor-sensitive proteases or expressing an array of proteases that are not targets of the inhibitors (Vogel *et al.* in press).

The third class of genes commonly studied in response to specific plant-derived compounds are detoxifying genes; these include cytochrome P450 monooxygenases (P450s), glutathione S-transferases (GSTs), UDP-glycosyltransferases (UGTs) and carboxylesterases (CEs) (Teese *et al.* 2010; Ahn *et al.* 2012; Fang 2012; Feyereisen 2012; Montella *et al.* 2012). Specifically, enzymes coded by these genes help to chemically transform toxins entering the insect system into hydrophilic compounds that can be eliminated or sequestered for further use. Recent transcriptome-wide expression analyses in insects responding to secondary plant metabolites such as phenolic compounds have found these genes are differentially regulated (Celorio-Mancera *et al.* 2011; Alon *et al.* 2012; Tao *et al.* 2012). Only one study has compared the global transcriptional regulation between a lepidopteran generalist and specialist species in response to a common host (Govind *et al.* 2010). In response to certain plant defences, the generalist regulated more detoxifying transcripts than did its specialist counterpart (Govind *et al.* 2010). A similar conclusion was reached when researchers compared the total number of known genes coding for different detoxifying enzymes in a generalist and a specialist aphid species (Ramsey *et al.* 2010). Another group of genes indirectly involved in detoxification found to be differentially regulated in generalist insects are those encoding transporters (Govind *et al.* 2010; Dermauw *et al.* 2013). Rather than structurally modifying chemicals, transporters carry toxic substances or metabolites out of cells; notably, a member of the major facilitator super family of transporters has been found to confer drug resistance to a fungal strain (Henderson & Maiden 1990; Hayashi *et al.* 2002).

To investigate these emerging patterns of gene regulation in generalist herbivores, that is, the transcriptional plasticity of ribosomal, digestive and detoxifying genes, we focus on the well-studied butterflies of the tribe Nymphalini (Family Nymphalidae), which have an ancestral association with the plants in the nettle family (Urticaceae) of approximately 90 Ma (Janz *et al.* 2001; Weingartner *et al.* 2006; Wahlberg *et al.* 2009). While most Nymphalini species feed exclusively on Urticaceae, some feed on more than one host-plant order outside of this ancestral host taxon and a few have specialized on these novel hosts (Janz & Nylin 2008). The Swedish comma butterfly, *Polygonia c-album,* is one of those exceptions. Hosts of *P. c-album* include *Urtica dioica* (nettle) and host plants from different related orders within the Rosids group (Rosales, Malpighiales, Fagales). Various species of *Ribes* are also used, such as *R. uva-crispa* (gooseberry), which represents the use of a host from a distantly related order (Saxifragales), one that is outside of the Rosids (Nylin

1988). In fact, the ability to feed on *Ribes* is an evolutionary innovation in the genus *Polygonia* and is part of a series of host-range expansions in the *Nymphalis-Polygonia* clade of the Nymphalini (Janz *et al.* 2001; Weingartner *et al.* 2006). These expansions coincide with a large positive change in net speciation rate for this clade, reversing the lower speciation trend across Nymphalini (Janz & Nylin 2008). If the use of *Ribes* is considered a single colonization event, it can be dated around 15 Ma (Wahlberg *et al.* 2009), which makes this event a more recent colonization compared with the association with Urticaceae (~90 Ma). Moreover, *Ribes* is not used as a host by any other genus in the family Nymphalidae and is also a very rare host among the rest of the butterflies (Savela 2013; and references therein). Although the uniqueness of the colonization implies that there is something 'difficult' about *Ribes*, because only *Polygonia* species and some lycaenids can feed on them (Savela 2013), there is no detailed knowledge of the physiological impact of feeding on these plants.

It is possible that the challenge represented by *Ribes* resides in the quality of hydroxynitrile glucosides (including cyanogenic glucosides) and cyanolipids they produce. The hydroxynitrile glucosides produced by members of the order Saxifragales including *Ribes* are different from those produced by plants grouped in Rosales including *Urtica*. The difference arises from the type of amino acid used as a precursor for the synthesis of hydroxynitrile glucosides, with leucine used in Rosales and isoleucine in Saxifragales (Bjarnholt & Moller 2008). Hydroxynitrile glucosides can be classified in two groups: those that release toxic cyanide upon plant cell disruption – these are known as cyanogenic glucosides – and those that are noncyanogenic, namely β- and γ-hydroxynitrile glucosides. Cyanogenic glucosides may deter those insects that casually try to feed on plants that contain this type of hydroxynitrile glucosides but do little against insects that specialize on cyanogenic plants (Zagrobelny *et al.* 2004; Bjarnholt & Moller 2008). In fact, the production of noncyanogenic hydroxynitrile glucosides in plants may reduce their nutritious value to herbivores that have evolved the ability to exploit cyanide (Bjarnholt & Moller 2008). For example, larvae of *Zygaena filipendulae* performed better on a *Lotus* species containing only cyanogenic glucosides than on a *Lotus* containing both cyanogenic and noncyanogenic hydroxynitrile glucosides (Zagrobelny *et al.* 2007). These observed metabolic differences and the evolutionary distance between Rosales and Saxifragales indicate that a significant host-expansion adaptation has taken place in *P. c-album*.

Gene expression plasticity has been previously investigated in *P. c-album* caterpillars in response to different hosts. Heidel-Fischer *et al.* (2009) inspected whether plant growth form (tree vs. herb) and relatedness (Urticales vs. Malpighiales) among host plants affected gene expression in different caterpillar tissues. The authors observed the differential regulation of genes encoding ribosomal proteins, transporters and serine proteases depending on the host and tissue type. However, genes related to detoxification were not differentially regulated in the gut tissue, and in the rest of the body, only one P450 gene was up-regulated when the larvae fed on the leaves of trees (Heidel-Fischer *et al.* 2009). To our knowledge, global transcriptional changes have not been inspected in response to a novel host, distantly related to the other plants in the diet repertoire of *P. c-album* or any other polyphagous herbivore. In addition, the extent of differences in gene expression among developmental stages (instars) has not been investigated in this butterfly species. This is an important consideration because previous work has shown that different plants affect the growth rate of *P. c-album* larvae (Janz *et al.* 1994).

Here, we report our functional insights into host-plant adaptation and likely detoxification using RNA-Seq analysis of two different *P. c-album* instars feeding either on *U. dioica* or on *R. uva-crispa*. Our goal was to test whether, in response to the novel host, our focus species displays a pattern of gene expression consistent with patterns in generalist insects. We predicted that comparisons of gene expression profiles in response to the ancestral host (*U. dioica*) and the acquired host distantly related to *U. dioica* (*R. uva-crispa*) would show the differential regulation of the three categories of genes mentioned above. More genes related to ribosomes were found to be up-regulated when the larvae of *P. c-album* fed on *U. dioica* in comparison with when they fed on the other two plant diets tested in the gene expression analysis performed by Heidel-Fischer *et al.* (2009). Therefore, we predicted that this pattern will be maintained, meaning that a higher number of ribosome-related genes would be up-regulated in larvae feeding on *U. dioica* than in those larvae when feeding on *R. uva-crispa*. In terms of genes related to digestion – specifically, those encoding proteases – we limited our prediction to the differential expression of a higher number of genes in response to *R. uva-crispa* feeding. However, unlike the results obtained by Heidel-Fischer *et al.* (2009), we expected the qualitative and quantitative regulation of detoxification-related genes in *P. c-album* larvae in response to *R. uva-crispa* feeding to be broad. We also expected that a particular developmental stage uses different categories of genes depending on the host plant, not only genes related to detoxification and digestion but also genes related to growth and development. Finally, we hypothesized that the accumulation of plant material over time – that is,

the comparison between developmental stages feeding on a given host from the beginning of their development – mostly entails the regulation of genes involved in primary metabolism.

Experimental procedures

Insects and host plants

P. c-album females were caught in the surrounding areas of Stockholm, Sweden, during summer 2010 and used for egg production. Offspring from 5 female butterflies were reared from neonate stage to second or fourth instar on either *U. dioica* or *R. uva-crispa* according to a split-brood experimental design with two biological replicates at the unit level. Each experimental unit consisted of two individual larvae from each of the 5 families (i.e. $n = 10$ larvae total per unit). There were 2 units per larval instar (second and fourth) by host-plant feeding treatment (*U. dioica* and *R. uva-crispa*), for a total of 8 sampled units. This split-brood design ensured that two full-siblings were represented in each unit of experimental measurement, and this procedure was duplicated. The split-brood design was intended to account for possible variability due to the effect of family, as these could potentially confound the results. Larvae were reared under controlled conditions in a climate chamber (LD 12:12, 17 °C) and were snap-frozen in liquid nitrogen after they reached the desired developmental stage and stored at −76 °C until RNA isolation. Some larvae were thus killed on their first day as second-instar larvae, whereas others were frozen on their third day as that instar, because our purpose was to freeze as many individuals as possible on the same date. Development time up until freezing was recorded.

High-throughput mRNA sequencing

Whole larvae were pooled per experimental unit ($n = 10$, two from each family) and homogenized under liquid nitrogen, with total RNA extracted from a fraction of the pulverized tissue with TRIsure (Bioline GmbH, Luckenwalde, Germany) following the manufacturer's protocol. DNA traces were removed from the samples by DNase treatment following the manufacturer's protocol (Turbo DNase; Ambion, Darmstadt, Germany). Additional RNA purification, quantification and quality control were performed following the protocols described previously (Celorio-Mancera *et al.* 2012). Instar and host-specific transcriptome sequencing of the 8 mRNA pools was performed using RNA-Seq assays on an Illumina HiSeq 2000 Genome Analyzer platform, using 100-bp single-end sequencing with RNA fragmented to an average of 150-bp size, with sequencing performed by a commercial service provider (Fasteris, Switzerland).

Gene expression analysis

P. c-album de novo transcriptome assembly (TA) as reference. Quality control measures, including the filtering of high-quality reads based on the score value given in 'fastq' files, removal of reads containing primer/adaptor sequences and trimming of read length, were taken using the CLC Genomics Workbench software 5.5.1 (http://www.clcbio.com). The *de novo* TA was performed using the CLC Genomics Workbench software by comparing an assembly with standard settings and two additional CLC-based assemblies with different parameters. Default settings were nucleotide mismatch cost = 3; insertion = deletion costs = 3; length fraction = 0.3; similarity = 0.80; second assembly settings were nucleotide mismatch cost = 2; insertion = deletion costs = 2; length fraction = 0.4; similarity = 0.85. Third assembly settings were nucleotide mismatch cost = 2; insertion = deletion costs = 3; length fraction = 0.3; similarity = 0.9; conflicts among individual bases were resolved in all assemblies by voting for the base with the highest frequency. In all assemblies, word size and bubble size of the graph parameters were automatically calculated. Contigs shorter than 250 bp were removed from the final analysis. The three assemblies were subsequently compared according to quality criteria such as N50 contig size, total number of contigs and the number of sequence reads not included in the contig assembly. For each assembly, the 100 largest contigs were manually inspected for chimeric sequences. The presumed optimal consensus transcriptome was selected among the three different assemblies based on the criteria of highest N50 contig size, lowest total number of contigs and lowest number of chimeric sequences in the 100 largest contigs found in the contig assembly with the following parameters: nucleotide mismatch cost = 2; insertion = deletion costs = 3; length fraction = 0.3; similarity = 0.9. The selected TA was used as a reference for mapping the reads of the individual *P. c-album* samples.

Annotation of the consensus transcriptome used BLAST searches conducted on a local server against the National Center for Biotechnology Information (NCBI) using the NCBI BLAST. Homology searches (BLASTx and BLASTn) of unique sequences and functional annotation by gene ontology (GO) terms (www.geneontology.org) used the BLAST2GO software suite with an *e*-value cut-off $\leq 10^{-3}$ to obtain the best homolog (Gotz *et al.* 2008). The sequence data generated in this study have been deposited at NCBI in the Short Read Archive database.

Quality assessment of the RNA-Seq data used principal component analysis and hierarchical clustering on

CLC Genomic Workbench software (CLC Bio, Aarhus, Denmark). Analyses of both methods were conducted on normalized 'reads per kilobase of exon model per million mapped reads' (RPKM). These relative transcript levels were generated from each mapped sample condition against the optimal consensus TA using the CLC Genome workbench RNA-Seq function (minimum length fraction = 0.7, maximum mismatches = 2). We performed one analysis grouping data per host, instar and host–instar interaction comparison in order to discover differential expressed genes for each of these effects, and in the process, the corresponding normalized RPKM data were analysed using the count-based gene expression analysis Baggerly's test [false discovery rate (FDR) *P*-value <0.05]. The Baggerly's test (beta-binomial) is a weighted t-type test statistic that compares the proportions of counts in a group of samples against those of another group, giving them different weights depending on their sizes (total counts) (Baggerly *et al*. 2003). Additional gene set enrichment analyses (GSEAs) were performed on the corresponding lists obtained from these statistical tests per comparison (minimum number of genes per ontology annotation = 10). The differential gene expression and GSEAs were performed using the CLC Genomic Workbench program. This gene expression analysis based on TA-contig mapping and its annotation obtained via BLAST2GO search is denoted from this point forward 'TA-contig expression.'

Quantitative whole-gene expression. The *de novo* TA-contigs used for the RNA-Seq mapping were assigned to a database of predicted proteins from *D. plexippus* (monarch butterfly) using BLASTx. Each TA-contig was assigned to its putative *D. plexippus* ortholog using the following cut-off levels: bit score ≥50, *e*-value ≤ 1×10^{-6}, over a length of 33 amino acids. The database of predicted *D. plexippus* proteins was produced by a self-blast of the complete predicted gene set from the *D. plexippus* genome (obtained from MonarchBase http://monarchbase.umassmed.edu/). For proteins having > 90% amino acid identity in the self-blast over a region 33 peptides long, the shorter of the two was discarded to filter out recent gene duplicates and alternative splicing isoforms. The functional annotation of the *D. plexippus* proteins was based on GO terms, which were obtained as described above for the TA-contig expression. We then used this relational table between the TA-contigs and their putative *D. plexippus* orthologs to generate whole-gene expression values for the *P. c-album* raw reads. For each *P. c-album* TA-contig hitting the same *D. plexippus* protein, the raw reads mapping to those contigs were summed, and this was done independently for each of the sequenced pools. RPKM expression levels were calculated using the nucleotide length for the corresponding *D. plexippus* predicted gene. This gene expression analysis is referred to henceforth as 'whole-gene expression', and the quantitative gene expression levels comparing host, instars and all host–instar contrasts were analysed using normalized RPKM data by Baggerly's and GSEA tests under the parameters described for the TA-contig expression analysis.

Results

Development time

It was considered that the treatment termination date – that is, the day the larvae was frozen – indicated the development time of larvae on a given host. A t-test comparison between the development time up until freezing between the two different host plants for each termination instars (second and fourth) revealed that *P. c-album* larvae grew faster on *U. dioica* than on *R. uva-crispa* (second instar: df = 70; $t = 2.099$; $P = 0.039$, fourth instar: df=57; $t = 2.79$; $P = 0.007$)(Fig. S1, Supporting information).

RNA-Seq and transcriptome assembly

A total of 80 million, 100-bp single-end reads were obtained. Approximately 20 million reads were generated, on average, for each group of second and fourth instars feeding on either *U. dioica* or *R. uva-crispa*. The optimal *de novo* TA resulted in 45 649 contigs (N50 = 628) against which approximately 75% of the reads could be remapped. Almost 98% of the TA-contigs had reads corresponding to a response to *U. dioica* feeding remapped onto them, regardless of larval instar. In turn, 61 and 70% of the TA-contigs contained remapped reads generated, respectively, from second and fourth instars feeding on *R. uva-crispa* (Table 1). GO annotations were obtained for 45% of the total number of TA-contigs, with most contigs identified as gene products active in catabolic processes, associated with protein complexes and/or involved in nucleotide binding (Fig. S2, Supporting information). Another 3592 contigs hit a gene in the databases encoding a predicted or hypothetical protein with no GO annotation. Most of the predicted genes in the TA were identified as homologs of species within Insecta (78.5%), while the remaining predicted genes corresponded to other species of eukaryotes. 3.3% belonged to eubacteria, which is to be expected as we were sequencing whole larvae. Only 0.4% corresponded to viruses. The highest number of predicted genes, representing approximately 70% of those identified (with and without GO annotation), had the best homolog from the *D. plexippus* coding sequence data set (see species distribution graph, Fig. S3, Supporting information). Estimating the total number of genes in a typical butterfly genome

Table 1 *P. c-album* caterpillar sequence reads used for the consensus transcriptome assembly (TA) and mapping for gene expression analysis. Host-feeding treatments: *U. dioica* and *R. uva-crispa*. Two biological replicates for each developmental stage of *P. c-album* larvae (second and fourth)

	U. dioica				*R. uva-crispa*			
	Second		Fourth		Second		Fourth	
	1	2	1	2	1	2	1	2
Total number of reads	7 149 006	11 968 888	11 001 464	8 788 272	9 801 293	7 476 496	7 639 101	16 668 961
Number of reads after trimming	7 122 862	11 925 681	10 961 911	8 757 589	9 761 390	7 449 965	7 609 906	16 603 647
Number of reads used for TA	80 192 951							
Number of TA-contigs	45 649							
N50 (bp)	628							
Longest contig (bp)	24 053							
Percentage of reads mapped to TA	78.5	80.4	79.9	81.1	80.2	71.9	74.7	80.9
Percentage of TA-contigs with mapped reads	96.7		96.7		70		61	

covered by our TA, we found that the *P. c-album* TA hit 66% of the total number of *D. plexippus* genes, with 77% of these *D. plexippus* genes having a GO annotation (Table S1, Supporting information).

TA-contig expression

Host effect. Principal component analysis showed that the samples clearly grouped together by host-feeding treatment. However, there is no clear grouping of samples by instar within each host-feeding treatment (Fig. 1). Individual gene analyses identified 8% (FDR corrected *P*-value <0.05) of the total number of contigs as differentially expressed in response to host (Table 2). GSEAs detected that most of these differentially expressed genes, independent of confidence level, were involved in transmembrane transport (many of them monocarboxylate transporters), encoding ribosomal proteins and/or proteins which interact in a selective, noncovalent way with other molecules. This gene list included transcription factors and RNA-binding proteins (Table 3; Table S2, Supporting information). Additionally, a smaller number of genes predicted to encode proteins integral to membranes, serine-type endopeptidases and the structural constituents of the insect cuticle were also differentially expressed in the caterpillars in response to host regardless of the p-value cut-off. Nucleic acid and protein binding predicted products along with hydrolases (mostly carboxylesterases) were also differentially regulated in response to the host only at the least conservative corrected *P*-value (Table 3; Table S2, Supporting information).

There was no GO term enriched among the fraction of genes (0.3%) that was differentially expressed between second and fourth instar from the total number of contigs in the TA.

Effect of the host–instar interaction

The analysis of the effect of the host–instar interaction allowed us to test (i) whether different genes were used by a particular developmental stage depending on the host plant, (ii) whether feeding on a particular host over time might result in different gene regulation in later instars and (iii) gene regulation differences between different instars feeding on different hosts.

Genes were differentially expressed in all host–instar comparisons, and enriched GO categories were found prevalent on all lists of statistically filtered genes (Table 2). The highest number of enriched GO categories was found in the comparisons between fourth-instar larvae reared on *U. dioica* and fourth-instar larvae reared on *R. uva-crispa*, and between the fourth-instar larvae reared on *U. dioica* and second-instar larvae reared on *R. uva-crispa* (Fig. 2; comparisons 3 and 4), reflecting mostly the effect of the host on primary metabolism and on both genetic and environmental information processing in the caterpillars. Chitin and lipid metabolic processes along with transmembrane transport were affected with up-regulated transcripts in response to *U. dioica*, particularly in fourth-instar larvae. A higher number of ribosome-related proteins remained up-regulated when *U. dioica* was a host, but when different developmental stages feeding on *R. uva-crispa* were compared, all differentially expressed genes were up-regulated exclusively in the second instar. A similar pattern was observed for the general category of binding,

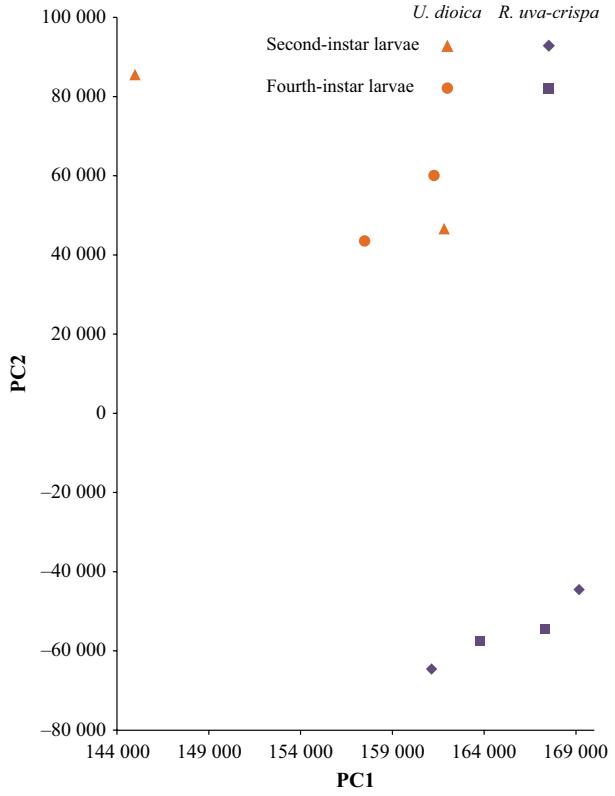

Fig. 1 Quality control of *P. c-album* RNA-Seq data using consensus transcriptome as a reference. Principal component analysis on normalized RNA-Seq reads mapped onto consensus transcriptome. Second-instar larvae reared on *U. dioica* (triangle); fourth-instar larvae reared on *U. dioica* (circle); second-instar larvae reared on *R. uva-crispa* (rhombus); fourth-instar larvae feeding on *R. uva-crispa* (square).

although in this case even more proteins in this category were up-regulated in the oldest larval stage when it fed on *U. dioica*. More hydrolase-encoding genes were transcribed in the fourth-instar larvae regardless of host (Fig. 2).

Larvae of different developmental stages did not differ greatly in their response to either *U. dioica* or *R. uva-crispa*, because both the total number of differentially expressed genes and that of enriched GO annotations were the fewest among the comparisons (Table 2; Fig. 2; comparisons 5 and 6). When feeding on *U. dioica*, second- and fourth-instar larvae produced different levels of those transcripts encoding serine-type endopeptidases. Fourth-instar larvae up-regulated many more of these enzymes in response to both hosts. More genes encoding the structural constituents of the cuticle were up-regulated in the fourth-instar larvae when they fed on *R. uva-crispa*.

Whole-gene expression vs. TA-contig expression

Both expression analyses were consistent with the finding that the regulation of caterpillar genes involved in environmental (proteins integral to the membrane) and genetic (nucleic acid binding) information processing depends on the type of host, even at the most conserva-

Table 2 Percentage of *P. c-album* TA-contigs or *D. plexippus* orthologous genes showing differential expression in response to host, instar and their interaction (Baggerly's test, FDR *P*-value correction <0.05). Only the absence of GO category significantly enriched in the corresponding fraction of differentially expressed genes is noted (No GSEA). Host-plant feeding treatment by instar comparisons: 1: *U. dioica*/second vs. *R. uva-crispa*/second; 2: *U. dioica*/second vs. *R. uva-crispa*/fourth; 3: *U. dioica*/fourth vs. *R. uva-crispa*/fourth: 4: *U. dioica*/fourth vs. *R. uva-crispa*/second; 5: *U. dioica*/second *vs. U. dioica*/fourth; 6: *R. uva-crispa*/second vs. *R. uva-crispa*/fourth

Expression analysis	Effect/comparison		% putative genes by FDR *P*-value 0.05
TA-contig	Host		8.0
	Instar		0.3/No GSEA
	Host–instar interaction	1	4.2
		2	7.1
		3	7.0
		4	4.6
		5	1.1
		6	1.7
Whole gene	Host		10.5
	Instar		0.6/No GSEA
	Host–instar interaction	1	6.4
		2	8.8
		3	8.5
		4	7.3
		5	1.8/No GSEA
		6	2.0

Table 3 Gene set enrichment analysis on differentially expressed predicted genes detected in response to host-feeding treatment using TA-contig and whole-gene expression analyses. (Baggerly's test, FDR *P*-value correction <0.05). Information within parentheses corresponds to the results obtained using the whole-gene expression analysis

GO domain	GO term	No. of putative genes
Biological process	Transmembrane transport	32 (11)
Cellular component	Ribosome	36
	Integral to membrane	24 (23)
Molecular Function	Binding	46
	Serine-type endopeptidase activity	40
	Structural constituent of cuticle	20 (15)
	Hydrolase activity	13 (15)
	Nucleic acid binding	10 (22)
	Protein binding	9

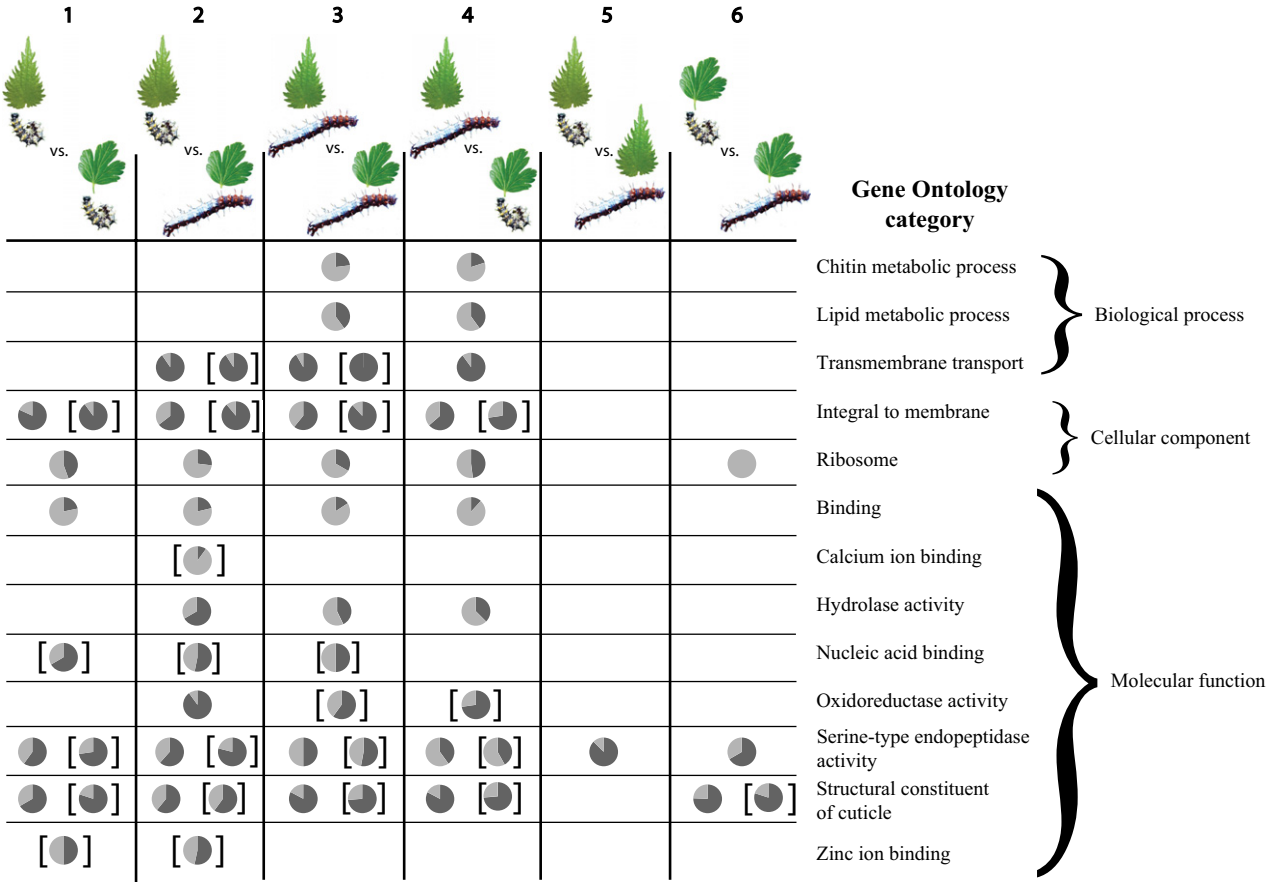

Fig. 2 Effect of host–instar interaction on transcriptional regulation in *P. c-album* caterpillars analysed by GSEA. Charts depict the proportion of up-regulated predicted genes in each condition per comparison within the most enriched and significant categories found by GSEA (light grey = uppermost/first condition in the comparison). Host-plant feeding treatment by instar comparisons: 1 = *U. dioica*/second vs. *R. uva-crispa*/second; 2 = *U. dioica*/second vs. *R. uva-crispa*/fourth; 3 = *U. dioica*/fourth vs. *R. uva-crispa*/fourth: 4 = *U. dioica*/fourth vs. *R. uva-crispa*/second; 5 = *U. dioica*/second vs. *U. dioica*/fourth; 6 = *R. uva-crispa*/second vs. *R. uva-crispa*/fourth. Results within parentheses correspond to the whole-gene expression analysis.

tive confidence level (Table 3). Neither analysis found any enriched gene categories in response to instar (Table 2). However, while the TA-contig analysis found a differential expression of certain groups of genes in response to host (e.g. ribosome-related proteins or serine-type endopeptidases), such groups were not detected in the whole-gene expression analysis (Table 3).

Two fewer GO categories were found to be enriched by whole-gene expression (Fig. 2). In addition, narrower functional categories, such as zinc, calcium and nucleic acid binding, were detected by whole-gene expression in the different host–instar comparisons (Fig. 2). The differential expression of genes involved in the primary metabolism and encoding of ribosome-related proteins in response to the host–instar interaction effect was observed only when using TA-contig expression. The rest of the categories were consistently observed, at least in one comparison, and the direction of gene regulation was also in agreement between analyses.

Discussion

Differentially regulated genes

Ribosomal genes. According to our predictions, genes coding for ribosomal proteins were up-regulated in *P. c-album* larvae reared on *U. dioica*. Nevertheless, such differential regulation was only observed when using the TA-contig expression analysis. On the other hand, we consider important to note the differential expression of transcription factors and a few other gene products, such as RNA-binding proteins, when the larvae fed on the novel host. It is well established that transcription factors can either directly or indirectly regulate the expression of a target gene (Spitz & Furlong 2012) and that RNA-binding proteins play an important role in eukaryotic post-transcriptional regulation (Glisovic *et al.* 2008). As ribosomal proteins may have a role in the regulation of gene expression, it might be worth

investigating their possible interaction with these other genes involved in transcript regulation. Stricter functional analyses are, however, necessary to confirm the relevance to insect herbivory and host adaptation of the candidate genes found in this study.

Digestion-related genes. Regarding digestive enzymes, protease activity in particular has been suggested to potentiate polyphagy in *Helicoverpa armigera*. Chikate *et al.* (2013) chose a set of trypsin and chymotrypsin transcripts and inspected their pattern of expression across different developmental stages of this lepidopteran pest species in response to diets with different nutritional qualities. The authors observed a dynamic regulation of protease expression in response to diet and developmental stage. On the basis of gene expression profiling, substrate and inhibitor specificity of gene products, it was concluded that serine proteases are key enzymes allowing larvae to adapt to many different diets and therefore augmenting polyphagy (Chikate *et al.* 2013). Here, coming from an unbiased global transcriptome-wide analysis, we reached similar conclusions. *P. c-album* is able to regulate the expression of serine-type endopeptidase transcripts in an instar- and host-specific manner. We expected a higher number of these genes to be differentially regulated in response to *R. uva-crispa*. However, the results from this study indicate that the instar effect is a stronger determinant of expression of serine-type endopeptidase transcripts; older larvae can up-regulate more of these genes regardless of host plant. It remains to be investigated whether this ability to regulate protease-encoding genes made it possible to include *R. uva-crispa* in the diet repertoire of *P. c-album*.

Detoxification-related genes. According to our prediction, we found that detoxification-related genes, specifically CEs, were differentially regulated when comparing host-feeding treatments. Nevertheless, the number of up-regulated genes with hydrolase activity, including CES, was also determined by the developmental stage of larvae. Again, this differential regulation was only observed when using the TA-contig expression analysis. Our findings also agree with those reported in a recent study inspecting the global transcriptional response in the extreme generalist species *Tetranychus urticae*. When Dermauw *et al.* (2013) examined the adaptation of this spider mite to host-plant transfer and pesticides, they found differential regulation characterized in genes encoding not only detoxifying enzymes but also membrane-binding proteins and transporters. The authors conclude that membrane-binding proteins and transporters belonging to the 'major facilitator super family' (MFS) are key players facilitating insects' adaptation and survival in response to a new diets and toxic compounds such as pesticides (Dermauw *et al.* 2013). We also observed these two types of proteins in our examination of *P. c-album's* transcriptional response to host plant. Specifically, we detected a strong up-regulation of genes coding for transporter and membrane-binding proteins regardless of larval instar in response to *R. uva-crispa*. Therefore, we consider the dynamic regulation of membrane-associated proteins and transporters (including MFS and monocarboxylate types) to be an emerging trend, which may facilitate polyphagy in herbivores.

Genes coding for structural proteins. Particularly intriguing is the differential expression of transcripts coding for structural constituents of the cuticle observed in our analysis. Unsurprisingly, as it is such a vital element of the insect body, the cuticle may be a target for plant defence. A microarray-based study revealed the differential expression of cuticular proteins when insecticide-resistant and insecticide-susceptible strains were compared in a generalist aphid species (Puinean *et al.* 2010). It was suggested that cuticular proteins may help thicken the cuticle and, as a consequence, reduce the penetration of insecticides (Puinean *et al.* 2010). Therefore, in *P. c-album*, it is possible that a strengthened cuticular component in the insect gut may prevent material and toxic plant metabolites from entering the insect body. Alternatively, plant chemicals may interfere with chitin metabolism and cuticular protein interactions, compelling the insect to regulate the production of those important structural constituents.

Whole-gene vs. TA-contig expression analyses. Functional insights into host-plant utilization from GSEAs were partially concordant in the two different expression analyses implemented (whole gene vs. TA-contig). The inconsistencies between these methods may be attributed to two factors. First, the whole-gene expression analysis is expected to be less sensitive to the fragmented nature of transcriptome assemblies (several contigs in an assembly can represent one gene) and therefore is more robust (Hornett & Wheat 2012). However, in the case of recent gene families, this approach offers a much more conservative analysis than the TA-contig approach. Second, the whole-gene expression analysis used the predicted gene set of a specialist insect species. Therefore, using *D. plexippus* as our proxy for a full transcriptome could have biased some of our results. We expect then that results found in common for both analyses are robust. Thus, the differential regulation of digestive enzymes, cuticular proteins, membrane proteins, transcription factors and transporters in *P. c-album* caterpillars detected by the TA-contig expression was consistent with the results

observed for the whole-gene expression analysis for most host–instar comparisons.

How do insect herbivores attain polyphagy?. Considering that polyphagy is rarely found among phytophagous insects, in contrast to the relative abundance of oligophagy, which is a more intermediate level of host specificity (Bernays & Graham 1988), and that host-range expansions seem to promote species diversity (Janz *et al.* 2006; Weingartner *et al.* 2006), it is useful to study generalist species under a phylogenetic framework for host-plant use. Because knowledge about the evolutionary history of host use for *P. c-album* and its lineage is available, we considered this species a good candidate for the inspection of transcriptional plasticity in response to plant feeding. It has been suggested that *P. c-album* derives from an ancestor able to make use of a wide host repertoire, including *Ribes* spp, envisioning the *Nymphalis-Polygonia* clade as being pre-adapted for host shifts that facilitated parallel colonization attempts of novel host plants (Weingartner *et al.* 2006). This pre-adaptation may indicate how induced biochemical pathway(s) confer the ability to feed on the novel host plant, in this case *R. uva-crispa* (West-Eberhard 2005; Weingartner *et al.* 2006; Agosta & Klemens 2008). While some *Polygonia* species have either become respecialized on hosts in the Urticaceae family of Rosales or a novel host (e.g. *P. egea* and *P. gracilis*, respectively) (Janz *et al.* 2001), the polyphagous state of *P. c-album* includes ancestral and novel hosts and thus resembles the possible state of the *Nymphalis-Polygonia* ancestor at the 'evolutionary moment' of expanded host-plant use.

We take results presented here to be initial evidence in support of our prediction that plasticity of gene expression is mostly responsible for the ability of larvae of the comma butterfly to eat *R. uva-crispa*. If we assume that *P. c-album* represent a species at the evolutionary moment of expanded host-plant use, we can speculate that the ancestor of the *Nymphalis-Polygonia* clade resembled *P. c-album* in terms of host repertoire and that it may have shaped similar transcriptional regulatory switches to use novel plants. In this light, our observations suggest that future work should compare insects' transcriptional profiles in response to sustained exposure to either host (this study) and those in response to immediate and reciprocal host switches in a particular time point in larval development. This comparison will allow us to assess another level of genetic plasticity and test whether the expression of host-specific genes or transcript levels is readily induced, given that the polyphagous insect species has more than one host plant available at a particular place and time (ecological polyphagy). In addition, attention should be focused on gene function in the specialist species that have derived from polyphagous and 'plastic ancestors' similar to *P. c-album*, in accordance with the 'oscillation hypothesis' of insect–plant interactions (Janz & Nylin 2008). Only by studying such transitional species can we build up enough empirical insights to attain a general understanding of the mechanisms of polyphagy and their role in evolution.

Conclusions

Recently, gene expression profiling in generalist insect herbivores has indicated their ample ability to regulate the transcription of ribosome-related, detoxifying and digestive enzymes in response to diet. We tested whether these emerging transcriptional patterns will be observed in caterpillars of the polyphagous species *P. c-album* when feeding on a host plant that has been recently added to their diet repertoire. We hypothesized that those gene expression signatures found in *P. c-album* will be observed independently of the larval stage. In this study, we observed that *P. c-album* caterpillars primarily regulate transcripts coding for detoxification and digestive enzymes. Additionally, *P. c-album* larvae reconfigure the expression of membrane proteins and transporters depending on their diet just as other generalist species. Moreover, when we contrast two caterpillar developmental instars, the direct effects of host-plant chemistry seem to overshadow those due to life history experience on a particular host plant. The differential regulation of cuticular proteins, membrane proteins and transporters in *P. c-album* caterpillars in response to the host plants reveals new candidates for understanding the molecular dynamics in the plant–insect interaction. The transcription factors that were found differentially regulated deserve special attention for future functional investigations, because they are potentially responsible for the transcriptional plasticity observed in other gene groups responding to the host-feeding treatments. In order to understand the macroevolutionary mechanisms that may allow insect species to achieve polyphagy, we must investigate plasticity in insect species within an evolutionary framework of host-plant use.

Acknowledgments

We thank Alexander Schäpers for help during summer 2010. This work was jointly supported by the Max Planck Society and a strategic grant awarded to Niklas Janz by the Stockholm University Faculty of Science. We thank Emily Wheeler for editorial assistance. We also acknowledge support from a Swedish Research Council grant to Sören Nylin (621-2011-5636).

References

Agosta SJ, Klemens JA (2008) Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. *Ecology Letters*, **11**, 1123–1134.

Ahn SJ, Vogel H, Heckel DG (2012) Comparative analysis of the UDP-glycosyltransferase multigene family in insects. *Insect Biochemistry and Molecular Biology*, **42**, 133–147.

Alon M, Elbaz M, Ben-Zvi MM *et al.* (2012) Insights into the transcriptomics of polyphagy: Bemisia tabaci adaptability to phenylpropanoids involves coordinated expression of defense and metabolic genes. *Insect Biochemistry and Molecular Biology*, **42**, 251–263.

Baggerly KA, Deng L, Morris JS, Aldaz CM (2003) Differential expression in SAGE: accounting for normal between-library variation. *Bioinformatics*, **19**, 1477–1483.

Bernays E, Graham M (1988) On the evolution of host specificity in phytophagous arthropods. *Ecology*, **69**, 886–892.

Bjarnholt N, Moller BL (2008) Hydroxynitrile glucosides. *Phytochemistry*, **69**, 1947–1961.

Brown AS, Simmonds MJ, Blaney W (2002) Relationship between nutritional composition of plant species and infestation levels of thrips. *Journal of Chemical Ecology*, **28**, 2399–2409.

Celorio-Mancera MP, Ahn SJ, Vogel H, Heckel DG (2011) Transcriptional responses underlying the hormetic and detrimental effects of the plant secondary metabolite gossypol on the generalist herbivore Helicoverpa armigera. *BMC Genomics*, **12**, 575.

Celorio-Mancera MP, Heckel DG, Vogel H (2012) Transcriptional analysis of physiological pathways in a generalist herbivore: responses to different host plants and plant structures by the cotton bollworm, Helicoverpa armigera. *Entomologia Experimentalis Et Applicata*, **144**, 123–133.

Chi YH, Salzman RA, Balfe S *et al.* (2009) Cowpea bruchid midgut transcriptome response to a soybean cystatin - costs and benefits of counter-defence. *Insect Molecular Biology*, **18**, 97–110.

Chikate YR, Tamhane VA, Joshi RS, Gupta VS, Giri AP (2013) Differential protease activity augments polyphagy in Helicoverpa armigera. *Insect Molecular Biology*, **22**, 258–272.

Coleno-Costes A, Jang SM, de Vanssay A *et al.* (2012) New partners in regulation of gene expression: the enhancer of Trithorax and Polycomb Corto interacts with methylated ribosomal protein L12 via its chromodomain. *Plos Genetics*, **8**, 10.

Dermauw W, Wybouw N, Rombauts S *et al.* (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**, E113–E122.

Dowd PF, Zuo WN, Gillikin JW, Johnson ET, Boston RS (2003) Enhanced resistance to Helicoverpa zea in tobacco expressing an activated form of maize ribosome-inactivating protein. *Journal of Agricultural and Food Chemistry*, **51**, 3568–3574.

Ehrlich P, Raven P (1964) Butterflies and plants: a study in coevolution. *Evolution*, **18**, 586–608.

Fang SM (2012) Insect glutathione S-transferase: a review of comparative genomic studies and response to xenobiotics. *Bulletin of Insectology*, **65**, 265–271.

Feyereisen R (2012) 8 - Insect CYP Genes and P450 Enzymes. In: *Insect Molecular Biology and Biochemistry* (ed Lawrence I. G.), pp. 236–316. Academic Press, San Diego.

Futuyma DJ, Keese MC, Funk DJ (1995) Genetic constraints on macroevolution - The evolution of host affiliation in the leaf beetle genus *Ophraella*. *Evolution*, **49**, 797–809.

Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. *Febs Letters*, **582**, 1977–1986.

Gotz S, Garcia-Gomez JM, Terol J *et al.* (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, **36**, 3420–3435.

Govind G, Mittapalli O, Griebel T *et al.* (2010) Unbiased transcriptional comparisons of generalist and specialist herbivores feeding on progressively defenseless nicotiana attenuata plants. *PLoS One*, **5**, 1.

Hayashi K, Schoonbeek HJ, De Waard MA (2002) Bcmfs1, a novel major facilitator superfamily transporter from Botrytis cinerea, provides tolerance towards the natural toxic compounds camptothecin and cercosporin and towards fungicides. *Applied and Environmental Microbiology*, **68**, 4996–5004.

Heidel-Fischer HM, Freitak D, Janz N *et al.* (2009) Phylogenetic relatedness and host plant growth form influence gene expression of the polyphagous comma butterfly (Polygonia c-album). *BMC Genomics*, **10**, 506.

Henderson PJF, Maiden MCJ (1990) Homologous sugar-transport proteins in *Escherichia coli* and their relatives in both prokaryotes and eukaryotes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **326**, 391–410.

Hornett E, Wheat C (2012) Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *BMC Genomics*, **13**, 361.

Janz N (2011) Ehrlich and raven revisited: mechanisms underlying codiversification of plants and enemies. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 71–89.

Janz N, Nylin S (2008) The oscillation hypothesis of host-plant range and speciation. In: *Specialization, Speciation and Radiation: the Evolutionary Biology of Herbivorous Insects*, pp. 203–215.

Janz N, Nylin S, Wedell N (1994) Host-plant utilization in the comma butterfly-Sources of variation and evolutionary implications. *Oecologia*, **99**, 132–140.

Janz N, Nyblom K, Nylin S (2001) Evolutionary dynamics of host-plant specialization: A case study of the tribe Nymphalini. *Evolution*, **55**, 783–796.

Janz N, Nylin S, Wahlberg N (2006) Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evolutionary Biology*, **6**, 4.

Li HM, Sun L, Mittapalli O *et al.* (2010) Bowman-Birk inhibitor affects pathways associated with energy metabolism in Drosophila melanogaster. *Insect Molecular Biology*, **19**, 303–313.

Liu YL, Salzman RA, Pankiw T, Zhu-Salzman K (2004) Transcriptional regulation in southern corn rootworm larvae challenged by soyacystatin N. *Insect Biochemistry and Molecular Biology*, **34**, 1069–1077.

Mattson WJ (1980) Herbivory in relation to plant nitrogen-content. *Annual Review of Ecology and Systematics*, **11**, 119–161.

Montella IR, Schama R, Valle D (2012) The classification of esterases: an important gene family involved in insecticide resistance - A Review. *Memorias Do Instituto Oswaldo Cruz*, **107**, 437–449.

Moon J, Salzman RA, Ahn JE, Koiwa H, Zhu-Salzman K (2004) Transcriptional regulation in cowpea bruchid guts during adaptation to a plant defence protease inhibitor. *Insect Molecular Biology*, **13**, 283–291.

Nylin S (1988) Host plant specialization and seasonality in a polyphagous butterfly, Polygonia c-album (Nymphalidae). *Oikos*, **53**, 381–386.

Puinean AM, Foster SP, Oliphant L *et al.* (2010) Amplification of a Cytochrome P450 gene is associated with resistance to Neonicotinoid insecticides in the Aphid Myzus persicae. *Plos Genetics*, **6**, 11.

Puri M, Kaur I, Perugini MA, Gupta RC (2012) Ribosome-inactivating proteins: current status and biomedical applications. *Drug Discovery Today*, **17**, 774–783.

Ramsey JS, Rider DS, Walsh TK *et al.* (2010) Comparative analysis of detoxification enzymes in Acyrthosiphon pisum and Myzus persicae. *Insect Molecular Biology*, **19**, 155–164.

Ryan CA (1990) Protease inhibitors in plants - genes for improving defenses against insects and pathogens. *Annual Review of Phytopathology*, **28**, 425–449.

Savela M (2013) *Lepidoptera and some other life forms*. Available from http://ftp.funet.fi/pub/sci/bio/life/intro.html.

Schoonhoven L, van Loon J, Dicke M (2005) *Insect-Plant Biology*. Oxford University Press, Oxford.

Shahidi-Noghabi S, Damme EJM, Smagghe G (2009) Expression of Sambucus nigra agglutinin (SNA-I') from elderberry bark in transgenic tobacco plants results in enhanced resistance to different insect species. *Transgenic Research*, **18**, 249–259.

Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, **13**, 613–626.

Surovtseva YV, Shutt TE, Cotney J *et al.* (2011) Mitochondrial Ribosomal Protein L12 selectively associates with human mitochondrial RNA polymerase to activate transcription. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 17921–17926.

Tao XY, Xue XY, Huang YP, Chen XY, Mao YB (2012) Gossypol-enhanced P450 gene pool contributes to cotton bollworm tolerance to a pyrethroid insecticide. *Molecular Ecology*, **21**, 4371–4385.

Teese MG, Campbell PM, Scott C *et al.* (2010) Gene identification and proteomic analysis of the esterases of the cotton bollworm, Helicoverpa armigera. *Insect Biochemistry and Molecular Biology*, **40**, 1–16.

Vogel H, Musser RO, Celorio-Mancera MP (in press) Transcriptome responses in herbivorous insects towards host plant and toxin feeding. In: *Annual Plant Reviews Volume 47: Plant Insect Interactions* (eds. Voelckel C, Jander G). John Wiley & Sons, Ltd., Chichester, West Sussex, UK.

Wahlberg N, Leneveu J, Kodandaramaiah U *et al.* (2009) Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary boundary. *Proceedings of the Royal Society B-Biological Sciences*, **276**, 4295–4302.

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.

West-Eberhard MJ (2005) Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6543–6549.

Zagrobelny M, Bak S, Rasmussen AV *et al.* (2004) Cyanogenic glucosides and plant-insect interactions. *Phytochemistry*, **65**, 293–306.

Zagrobelny M, Bak S, Ekstrom CT, Olsen CE, Moller BL (2007) The cyanogenic glucoside composition of Zygaena filipendulae (Lepidoptera: Zygaenidae) as effected by feeding on wild-type and transgenic lotus populations with variable cyanogenic glucoside profiles. *Insect Biochemistry and Molecular Biology*, **37**, 10–18.

Zhou X, Li XD, Yuan JZ, Tang ZH, Liu WY (2000) Toxicity of cinnamomin - a new type II ribosome-inactivating protein to bollworm and mosquito. *Insect Biochemistry and Molecular Biology*, **30**, 259–264.

N.J., L.S., H.V., S.N. and M.P.C.M. contributed in the conception and design of the study. L.S. conducted the treatments for insect rearing, collected and analyzed development time data. M.P.C.M. performed the molecular laboratory work previous to mRNA sequencing. H.V. conducted the de novo transcriptome assembly. C.W.W. and M.P.C.M. performed the gene expression analyses. M.P.C.M. drafted the manuscript and together with the rest of the authors contributed to the interpretation and the iterative refinement of the article. All authors have read and approved the submitted version.

Data accessibility

RNA-Seq sequence reads: European Nucleotide Archive Accession no.: ERP002606. Final *P. c-album* transcriptome assembly, TA-contig and whole-gene expression data: Dryad doi:10.5061/dryad.8t6c0.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 List of *D. plexippus* predicted proteins which had a *P. c-album* orthologous sequence(s) in the TA.

Table S2 List of TA-predicted genes differentially expressed in response to *R. uva-crispa* feeding compared to *U. dioica* in *P. c-album* larvae (Baggerly's test FDR corrected *P*-value cut-off 0.05).

Fig. S1 Effect of host plant on development time.

Fig. S2 Distribution of *P. c-album* annotated contigs in the *de novo* transcriptome assembly across gene ontology categories. A) biological process; B) cellular component and C) molecular function.

Fig. S3 Distribution of contigs across top-hit species.