Extensive *Myzus cerasi* transcriptional changes associated with detoxification genes upon primary to secondary host alternation

Peter Thorpe[1], Carmen M. Escudero-Martinez[1,2], Sebastian Eves-van den Akker[3], Jorunn I.B. Bos[1,2]

[1]Cell and Molecular Sciences, The James Hutton Institute, Invergowrie, Dundee, DD2 5DA, UK

[2]Division of Plant Sciences, School of Life Sciences, University of Dundee, Dundee

[3]Department of Plant Sciences, University of Cambridge, CB2 3EA, UK

*Corresponding Author:

j.bos@dundee.ac.uk

Abstract

Background

Aphids are phloem-feeding insects that cause yield losses to crops globally. These insects feature complex life cycles, which in the case of many agriculturally important species involves the use of primary and secondary host plant species. The switching between winter or primary hosts, on which overwintering eggs are laid upon sexual reproduction, and the secondary hosts, on which aphids reproduce asexually by parthenogenesis, is called host alternation. Here, we used *Myzus cerasi* (black cherry aphid), which overwinters on cherry trees and in summer spreads to herbaceous plant species, to assess aphid transcriptional changes that occur upon host alternation.

Results

Adaptation experiments of *M. cerasi* collected from local cherry tress to reported secondary host species revealed low survival rates when aphids were moved to two secondary host species. Moreover, aphids were unable to survive on one of the reported hosts (Land cress) unless first adapted to another secondary host (cleavers). Transcriptome analyses of populations adapted to the primary host cherry and two secondary host species showed extensive transcriptional plasticity to host alternation, with predominantly genes involved in oxidation-reduction differentially regulated. Most of the differentially expressed genes across the *M. cerasi* populations from the different hosts were duplicated and we found evidence for differential exon usage. In contrast, we observed only limited transcriptional to secondary host switching.

Conclusion

Aphid host alternation between summer and winter host plant species is an intriguing feature of aphid life cycles that is not well understood, especially at the molecular level. Here we show that, under controlled conditions, *M. cerasi* adaptation from primary to secondary host species does not readily occur and involves extensive changes in aphid gene expression. Our data suggests that different sets of genes involved in detoxification are required to feed from primary versus secondary host species.

Background

Aphids are phloem-feeding insects that belong to the order Hemiptera. Insects within this order feature distinctive mouthparts, or stylets, that in the case of phytophagous insects, are used to pierce plant tissues and obtain nutrients from the plant phloem. One striking feature of the complex life cycle of about 10 % of aphid species is the seasonal host switching between unrelated primary (winter) and secondary (summer) host plants, also called host alternation or heteroecy [1]; [2]. Host alternating aphids predominantly use woody plants as their primary hosts, on which (overwintering) eggs are laid, from which the first generation of aphids, or fundatrices, emerge in spring. The fundatrices, and their offspring, reproduce by parthenogenesis (asexual reproduction), giving birth to live nymphs. Winged forms (alate) will migrate to secondary host plants over the summer months where the aphid populations will go through multiple parthenogenic generations. In autumn, sexual female and male aphids will reproduce sexually and overwintering eggs are laid on the primary host. Exceptions to this general life cycle exist, with some aphids for example having multi-year cycles [3].

Heteroecy in aphids has independently arisen in different aphid lineages throughout evolutionary history [4] with monoecy (with the entire life cycle taking place on one plant species) on trees thought to be the ancestral state. Many different hypotheses explain the maintenance of heteroecy and driving factors described include nutritional optimization, oviposition sites, natural enemies, temperature tolerance, and fundatrix specialization [4]. With many important agricultural crops being aphid secondary hosts, understanding how aphids are able to switch between their primary and secondary hosts will provide better insight into the mechanisms of crop infestation. It is likely that switching between host plant species requires aphids to adapt to differences in host nutritional status as well as potential differences in plant defense mechanisms against insects. Host plant specialization in the pea aphid species complex is associated with differences in genomic regions encompassing predicted salivary genes as well as olfactory receptors [5]. Moreover, adaptation of *M. persicae* to different secondary host plant species involves gene expression changes, including of genes predicted to encode for cuticular proteins, cathepsin B protease, UDP-glycosyltransferases and P450 monooxygenases [6]. Analyses of gene expression differences between *Hyalopterus persikonus* (mealy aphids) collected from primary versus secondary host plant species under field conditions showed that genes with predicted functions in detoxification, but also predicted effector genes,

required for parasitism, were differentially regulated [7]. However, a detailed understanding of how aphids respond through the process of host alternation remains unclear.

Myzus cerasi, or black cherry aphid, is a host-alternating aphid that uses mainly *Prunus cerasus* (Morello cherry) and *Prunus avium (*sweet cherry), but also other *Prunus* species as primary hosts and several herbaceous plants (*Gallium* spp., *Veronica* spp., and cruciferous species) as secondary hosts [8] [9]. Infestation can cause significant damage on cherry trees, due to leaf curling, and shoot deformation, pseudogall formation, as well as fruit damage. Recently, the genome of *M. cerasi* was sequenced, providing novel insights into the potential parasitism genes as well as genome evolution [10]. The increasing availability of genomics resources for aphids, including *M. cerasi,* facilitates further understanding of aphid biology, including the processes involves in host alternation.

In this study, we adapted *M. cerasi* aphids collected from local cherry trees (primary hosts) to secondary hosts *Galium aparine* (cleavers) and *Barbarea verna* (Land cress) and assessed differences in aphid transcriptomes upon adaptation. We found that aphids collected from their primary host differed in their ability to adapt to secondary host plant species. The adaptation from primary to secondary host plant species involved extensive transcriptional changes in *M. cerasi*, especially with regards to genes involved in detoxification. However, we only observed limited transcriptional changes between *M. cerasi* adapted to the two different secondary host plant species.

Results and discussion

Myzus cerasi **host alternation under laboratory conditions is associated with low survival rates**

To determine whether transcriptional plasticity plays a potential role during primary to secondary host alternation, we used the aphid species *M. cerasi*, which can be found on its primary host, cherry, in spring, and uses several herbaceous secondary host species over the summer [9]. When attempting to establish a colony of *M. cerasi* from populations occurring on local cherry trees, we observed differences in survival rates upon transfer to reported secondary host plant species. While aphids were unable to survive transfer from primary host cherry to Land cress (*Barbarea verna*), we observed a 10%-20% survival rate

upon transfer to cleavers (*Galium aparine*) (Fig. 1). However, once aphid populations were established on cleavers, individuals from this population were successfully transferred to cress plants. We performed primary to secondary host transfer experiments with aphids collected from cherry trees at two different locations in three independent replicates with similar results (Fig. 1). Our observation that *M. cerasi* is unable to readily infest reported secondary hosts likely reflects that these aphids need to adapt to a change in host environment.

Myzus cerasi shows extensive transcriptional plasticity to host alternation

We assessed the changes that take place at the transcriptional level in *M. cerasi* upon moving aphids from their primary host cherry to secondary host plant cleavers and upon subsequent transfer to secondary host plant cress. Specifically, we sequenced the transcriptomes of *M. cerasi* populations collected from cherry, and established over a 3-week period on cleavers or cress using RNAseq.

We performed differential gene expression analysis (LOG fold change >2, False Discovery Rate (FDR) $p<0.001$) between the different aphid populations to identify gene sets associated with specific host plant species. Cluster analyses of the aphid transcriptional responses from this and previous work reporting on differential aphid gene expression in head versus body tissues [11] revealed that the overall expression profiles could be distinguished based on the aphid tissue used for sample preparation as well as the host species that *M. cerasi* was collected from (Additional File 1: Fig. S1a). Indeed, principal component analyses showed a clear separation between aphid transcriptomes associated with usage of different primary and secondary host species (Additional File 1: Fig. S1b). Overall, we identified 934 differentially expressed genes by comparing the different datasets for each of the aphid populations (Fig. 2a, Additional File 2: Table S1). A heat map of these 934 genes shows that gene expression profiles from aphids on their secondary hosts (cleavers and cress) are more similar to each other than to the gene expression profiles of aphids on their primary host (cherry) (Fig. 2a). Co-expression analyses reveals six main clusters of differentially expressed genes, two of which (A and E) contain the majority of genes (Fig. 2a and 2b). Cluster A contains 493 genes, which show higher expression in aphids on secondary host plants versus those on primary host cherry, and cluster B contained 342 genes showing an opposite profile of being more highly expressed in aphids on the primary host versus secondary hosts. GO annotation

165 revealed over-representation of terms associated with oxido-reductase activity in both
166 clusters, as well as several terms associated with carotenoid/tetrapenoid biosynthesis in
167 case of cluster B (Additional file 3: Table S2).
168
169 To assess differential expression of *M. cerasi* genes across the host plant-aphid
170 interactions we also analyzed pairwise comparisons for differentially expressed gene sets.
171 The largest set of differentially expressed genes (736) was found in comparisons between
172 aphids from cherry and cress, with 443 genes more highly expressed in aphids from cress,
173 and 293 more highly expressed in aphids from cherry (Fig. 3a). A total of 733 differentially
174 expressed genes were found in comparisons of aphids from cherry versus cleavers, with
175 367 genes more highly expressed in aphids collected from cherry and 366 genes more
176 highly expressed in aphids collected from cleavers (Fig. 3a). The observation that more
177 genes are highly expressed in aphids from cress compared to the other host species may
178 reflect the need for specific gene products to cope with adaptation to cress.
179
180 A relatively small number of genes were differentially expressed between aphids collected
181 from the two secondary hosts cleavers and cress, with only 5 genes more highly
182 expressed in aphids from cleavers, and 74 genes more highly expressed in aphids form
183 cress (Fig. 3a). This suggests that *M. cerasi* shows limited transcriptional plasticity to a
184 switch in secondary host environment. This is in line with our previous observation that
185 only a relatively small set of genes is differentially expressed in *M. persicae* and *R. padi*
186 when exposed to different host or non-/poor-host plants [10] as well as the relatively small
187 number of transcriptional changes when *M. persicae* is adapted to different secondary
188 hosts [6].
189
190 GO enrichment analyses of the 443 genes more highly expressed in aphids collected from
191 cress compared to those collected from cherry shows overrepresentation of genes
192 predicted to be involved various processes, including in heme binding (GO:0020037),
193 tetrapyrrole binding (GO:0046906), monooxygenase activity (GO:0004497),
194 oxidoreductase activity (GO:0016705), iron ion binding (GO:0005506), and hydrolase
195 activity (GO:0016787) (Additional file 4: Table S3). This set of 443 genes contains 282 of
196 the 366 genes that are also more highly expressed in aphids from the other secondary
197 host plant species, cleavers, with similar GO annotations (Fig. 3b; Additional file 5: Table
198 S4). The 293 genes more highly expressed in aphids collected from cherry than those
199 from cress shows over-representation of genes predicted in oxidoreductase activity

200 (GO:0016620, GO:0016903, GO:0055114, GO:001649) as well as other processes such
201 as fatty-acyl-CoA reductase (alcohol-forming) activity (GO:0080019), interspecies
202 interaction between organisms (GO:0044419), and symbiosis (GO:0044403) (Additional
203 file 4: Table S3). For the gene sets differentially expressed between aphids collected from
204 cherry and cleavers, GO enrichment analyses reveal that in reciprocal comparisons genes
205 predicted to function in oxidation-reduction are also over-represented (Additional file 4:
206 Table S3). Interestingly, we detected an overrepresentation of different GO terms
207 associated with oxidation reduction in gene sets of aphids collected from primary and
208 secondary host species, which suggest that *M. cerasi* may require different gene sets for
209 detoxification on primary versus secondary hosts.
210
211 Interestingly, among the 367 genes more highly expressed in aphids collected from cherry
212 compared to those collected from cleavers, we found that the majority of GO terms
213 identified through enrichment analyses correspond to metabolic processes biosynthesis
214 (Additional file 4: Table S3). Of these 367 transcripts 268 show similar expression
215 differences in aphids collected from cherry versus those collected from cress, whereas 98
216 are specific to the comparison of aphids collected from cherry versus cleavers (Fig. 3c).
217 Whilst GO enrichment analyses showed over-representation of genes involved in oxidation
218 reduction in the set of 268 overlapping transcripts, the 98 transcripts specifically up-
219 regulated in aphids collected from cherry versus cleavers show over-representation in
220 metabolic processes, and especially those associated with terpenoid/carotenoid
221 biosynthesis which are involved in aphid pigmentation (Additional file 6: Table S5) [12].
222 Possibly this observation reflects that *M. cerasi* requires specific gene sets for
223 pigmentation and feeding on primary host cherry compared to secondary host plants.
224 Notably, we did not observe any noticeable change in aphid color upon switching from
225 primary to secondary hosts, with aphids being a dark brown to black color on all plant
226 species tested (not shown), suggesting the differential regulation of carotenoid genes is
227 not associated with aphid color in this case but other unknown physiological functions.
228
229 To independently test whether select *M. cerasi* genes were differentially expressed in
230 aphids collected from primary and secondary host plants, we repeated the collection of
231 aphids from local cherry trees (separate site, location 2) and performed adaptation
232 experiments to cleavers and cress. We selected 10 genes for independent validation of
233 expression profiles by qRT-PCR. Five of these 10 genes were selected based on
234 enhanced expressed in aphids from cherry compared to aphids from secondary host

plants, and another 5 genes for being more highly expressed in aphids from secondary host plants compared to aphids from cherry. The genes selected based on higher expression in aphids from cherry showed similarity to genes predicted to encode a peroxidase, RNA-binding protein 14-like, hybrid sensor histidine kinase response regulator, maltase isoform a, and a lactase-phlorizin hydrolase. The genes selected based on higher expression in aphids from secondary hosts showed similarity to genes predicted to encode an unknown protein, a venom-like protease, a thaumatin-like protein, protein kintoun, and a cytochrome P450. Except for the gene with similarity to a venom-like protease, all genes showed a similar gene expression profile in both samples used for the RNAseq experiments and in the independently collected and adapted aphids from a different site, indicating that this gene set is consistently differentially expressed when *M. cerasi* switches from primary to secondary host plants (Additional file 7: Fig. S2). Most of these genes have predicted functions in detoxification, in line with our hypothesis that aphids require different sets of genes to deal with potential defensive plant compounds to be able to feed on different primary or secondary host plants. In *H. persikonus* collected from primary and secondary host plant species in the field, a similar observation was made in that an extensive gene set associated with detoxification was differentially regulated [7].

Single Nucleotide Polymorphism (SNP) analyses suggest that an aphid sub-population is able to switch from the primary to secondary hosts

We used the transcriptome dataset we generated here to compare the level of sequence polymorphisms between the aphid populations from the different primary and secondary host plants species. Variants/SNPs were predicted by mapping the RNAseq dataset for each aphid population (cherry, cleavers and cress) to the *M. cerasi* reference genome for each condition, with only unique mapping being allowed. The number of SNPs within each 10Kb window was calculated. The *M. cerasi* population from cherry has significantly more SNPs per 10Kb than the populations from both cleavers and cress when mapping reads back to the reference genome ($p<0.001$, Kruskal-Wallis with Bonferroni post hoc correction). In contrast, the aphid populations from cleavers and cress showed no significant difference in the number of SNPs per 10Kb ($p=0.29$, Kruskal-Wallis with Bonferroni post hoc correction). These results are not surprising considering that the *M. cerasi* reference genome was generated using a clonal line adapted to the secondary host cress and suggest that the population on cherry may be more genetically diverse than the populations on secondary host plants. Moreover, based on these findings we propose that

only a subpopulation of the primary host population may switch to secondary host plant species. It should be noted that these data are based on RNAseq, and do not rule out the possibility of allele-specific expression across the different host interactions. Hence, further characterization of the *M. cerasi* (sub)populations using DNAseq will be required to gain further insight into adaptation of this aphid species to its hosts.

Limited differential expression of predicted *M. cerasi* effectors during host alternation

We assessed whether predicted *M. cerasi* effectors are differentially expressed when aphids were raised on the primary or secondary hosts plants. The 224 predicted *M. cerasi* effectors we previously identified [10] show a wide range of expression levels across different interactions, with most expression variation in aphids collected from cherry (Fig. 4a; Additional file 8: Table S6). However, when assessing expression of a random non-effector set of similar size, this expression variation in aphids from cherry was less pronounced (Fig. 4a). Despite the observed variation in expression patterns, we only found a small number of differentially expressed candidate effectors, mainly when comparing aphids collected from the primary versus secondary hosts. Specifically, 13 candidate effectors are more highly expressed in aphids from both secondary host species compared to aphids from the primary host cherry, with one additional candidate effector more highly expressed in the case of aphids from cleavers compared to cherry only (Mca17157|adenylate kinase 9-like) (Additional file 8: Table S6). Although these candidate effectors were mainly of unknown function, several show similarity to thaumatin-like proteins and a venom protease. Interestingly, the candidate effector with similarity to the venom protease, Mca05785 (upregulated in secondary hosts), is member of a venom protease gene family cluster that consists of four members (3 are tandem duplications, 1 is a proximal duplication). Three of these are predicted to encode secreted proteins, and all members show higher expression levels in aphids from secondary hosts compared to aphids from primary host, but this variation is not statistically significantly different (Additional file 8: Table S6). In addition, 1 candidate effector (similar to RNA-binding protein 14) was differentially expressed when comparing aphids from the two secondary host plants, and 5 candidate effectors (Mca07285, Mca07514, Mca16980, Mca07516, Mca09259) were more highly expressed in aphids collected from cherry compared to aphids from cleavers and/or cress (Additional file 8: Table S6).

305 **Break-down of aphid effector co-regulation associated with the primary host**
306 **interaction**
307
308 We previously showed that expression of aphid effector genes, required for parasitism,
309 was tightly co-regulated pointing to a mechanism of shared transcriptional control [10]. To
310 test whether this also applies to *M. cerasi* adapted to primary versus secondary hosts, we
311 assessed co-expression patterns of Mc1 and Me10-like, an effector pair that is physically
312 linked across aphid genomes and tightly co-regulated, together with all other genes. A
313 total of 35 genes showed a high level of co-expression in *M. cerasi* during interaction with
314 the three different host plant species and across different aphid tissues (Fig. 4c). This
315 number is much smaller compared to the set of co-regulated genes in *R. padi* (213) and
316 *M. persicae* (114), which could be due to differences the quantity and quality of the
317 RNAseq datasets we used for these analyses [10, 13]. However, the pattern of co-
318 regulation observed in aphids collected from secondary host plants is not apparent in
319 aphids collected from the primary host cherry (Fig. 4c), which affects the overall accuracy
320 and ability to predict co-regulated genes. Possibly, shared transcriptional control of
321 effector genes is more relevant during aphid parasitism on secondary host plant species
322 rather than on primary hosts.
323
324 **Differential exon usage in *M. cerasi* upon host alternation**
325
326 We also found evidence for differential exon usage when comparing the different aphid
327 transcript datasets. Overall, 263 genes show significant differential exon usage when
328 comparing aphid datasets associated with the different primary and secondary hosts
329 (Additional file 9: Table S7). These 263 genes contain 2551 exons, of which 443 show
330 differential expression between aphids from primary host cherry versus the secondary
331 hosts. No significant GO annotation is associated with these 263 genes. One example of
332 differential exon usage in *M. cerasi* is peroxidase gene Mca06436, which contains 5
333 exons, 2 of which are significantly more highly expressed in aphids collected from the
334 primary host cherry compared to aphids from secondary host cleavers (Fig. 5). This
335 suggests that alternative splicing may be associated with adaptation to primary versus
336 secondary host species.
337
338 **The majority of *M. cerasi* genes differentially expressed across different hosts are**
339 **duplicated**

340

341 For genes differently expression in *M. cerasi* when grown on different hosts (cherry,
342 cleavers and cress) it is interesting to observe that the majority of these are duplicated (not
343 single copy). For the genes upregulated in *M. cerasi* when grown on cress versus cherry
344 only 14% are single copy, which is significantly lower than the percentage of single copy
345 genes in a randomly selected set of genes ($p<0.001$, Mann-Whitney U test). Moreover, for
346 all sets of differentially expressed genes, the differentially expressed genes were more
347 likely to be duplicated when compared to a background random gene set ($p<0.001$)
348 (Additional file 10: Table S8). To assess the categories of gene duplication within the
349 differential expressed gene sets, 100 iterations of randomly selecting 100 genes were
350 conducted to obtain a background population. This yielded a mean and standard deviation
351 for each duplication type from the parent gene population (normally distribution). A
352 probability calculator (Genstat) was used to determine how likely the observed counts
353 were to occur at random. This showed that most of the duplicated differentially expressed
354 genes were within the "dispersed duplication" category ($p<0.001$) and that there was no
355 significant difference in the occurrence of tandem or proximal gene duplications ($p>0.05$)
356 (Additional file 10: Table S8). In contrast to predicted *M. cerasi* effectors, the differentially
357 expressed genes identified in this study were not significantly further away from their
358 neighbor in the 3' direction ($p=0.163$, Mann-Whitney U Wilcoxon rank-sum test), or their 5'
359 neighbor gene ($p=0.140$, Mann-Whitney U Wilcoxon rank-sum test) when compared to an
360 equal sized random population (Additional file 11: Figure S3). Altogether our data suggest
361 that *M. cerasi* multi-gene families may play an important role on host alternation between
362 primary and secondary host plant species. This is in line with Mathers et al, [6] who
363 showed that duplicated genes play a role in adaptation of *M. persi*cae to different
364 secondary host species.

365

366 **Conclusion**

367

368 Aphid host alternation between summer and winter host plant species is an intriguing
369 feature of aphid life cycles that is not well understood, especially at the molecular level.
370 Here we show that, under controlled conditions, *M. cerasi* adaptation from primary to
371 secondary host species does not readily occur, with only 10-20% aphid survival, and
372 involves extensive changes in aphid gene expression. Our data suggests that different
373 sets of genes involved in detoxification are required to feed from primary versus secondary
374 host species. Many of these genes are members of multi-gene families, and changes in

375 aphid transcriptomes can also be associated with differential exon usage. In contrast, we
376 find only limited transcriptional plasticity to secondary host switching.
377
378 **Methods**
379

380 **Aphid collection and adaptation**

381 *M. cerasi* was collected in July 2013 from two separate locations in Dundee, United
382 Kingdom. Mixed age aphids from branches of an infested cherry tree were flash frozen in
383 liquid nitrogen upon collection (3 replicates of 50 aphids per location). For adaptation to
384 secondary host plants, 50 aphids of mixed age, were transferred to *Galium aparine*
385 (cleavers) or *Barbarae verna* (Land cress) detached branches placed in 3 replicate cup
386 cultures per location. Aphid survival was assessed after 1 week. Then, 5 aphids of the
387 surviving population on cleavers, were moved to a fresh cup culture containing detached
388 cleavers branches. Fresh plant material was added to the cups after 2 weeks. One week
389 later 50 mixed-age aphids per cup were flash frozen (aphids adapted to cleavers for
390 RNAseq) and fresh cleavers branches together with Land cress branches were added to
391 the cups. One week after adding the Land cress plant material, all cleavers material was
392 removed and fresh cress branches were added and fresh plant material was regularly
393 provided. Three weeks later 50 mixed-age aphids were collected per cup culture and flash
394 frozen (aphids adapted to cress for RNAseq). Aphids were maintained in cup cultures in
395 controlled environment cabinets at 18°C with a 16 hour light and 8 hour dark period.

396 **RNA sample preparation and sequencing**
397 Aphid samples were ground to a fine powder and total RNA was extracted using a plant
398 RNA extraction kit (Sigma-Aldrich), following the manufacturer's instructions. We prepared
399 three biological replicates for *M. cerasi* collected from each host. RNA quality was
400 assessed using a Bioanalyzer (Agilent Technologies) and a Nanodrop (Thermo Scientific).
401 RNA sequencing libraries were constructed with an insert size of 250bp according to the
402 TruSeq RNA protocol (Illumina), and sequenced at the previous Genome Sequencing Unit
403 at the University of Dundee using Illumina-HiSeq 100bp paired end sequencing. All raw
404 data are available under accession number PRJEB24338.

405 **Quality control, RNAseq assembly and differential expression**
406 The raw reads were assessed for quality before and after trimming using FastQC [14].
407 Raw reads were quality trimmed using Trimmomatic (Q22) [15], then assembled using

408 genome-guided Trinity (version r20140717) [16]. Transrate was run twice to filter out low
409 supported transcripts [17].
410
411 RNAseq assembly and annotation is available at DOI: 10.5281/zenodo.1254453. For
412 differential gene expression, reads were mapped to the *Myzus cerasi* genome [10], per
413 condition using STAR [31]. Gene counts were generated using Bedtools [32]. Differential
414 gene expression analysis was performed using EdgeR [19], using LOG fold change >2,
415 FDR p<0.001 threshold. GO enrichment analysis was performed using BLAST2GO
416 (version 2.8, database September 2015) [33] using FDR 0.05. The genome annotations
417 were formatted using GenomeTools [34] and subsequently HTseq [35] was used to
418 quantify exon usage. Differential exon expression was performed using DEXSEQ FDR
419 p<0.001 [36]. Heatmaps were drawn as described in Thorpe et al. [10].
420
421 Gene duplication categories were used from Thorpe et al. [10]. From these data, a random
422 population was generated by running one hundred iterations on a set of 100 randomly
423 selected genes and their duplication types for subsequence statistical analyses. The script
424 to generate random mean and standard deviation counts is available on Github
425 (https://github.com/peterthorpe5/Myzus.cerasi_hosts.methods). Statistical analysis was
426 performed using Probability Calculator in Genstat (17th edition). The obtained value from
427 the gene set of interest (differentially express genes across aphid populations) was
428 compared to the distribution of the random test set. Datasets identified as being
429 significantly different from the random population did not significantly deviate from a
430 normal distribution, thus the data was normally distributed. To assess the distances from
431 one gene to the next, an equal sized population (1020) of random genes and their values
432 for distance to their neighboring gene in a 3' and 5' direction was generated. The real
433 value and random values were not normally distributed and were analyzed in Genstat (17th
434 edition) using a non-parametric Mann-Whitney U Wilcoxon rank-sum test.
435
436 For SNP identification, RNAseq data was mapped back to the reference genome using
437 STAR (2.5.1b) with --outSAMmapqUnique 255, to allow only unique mapping [31]. SNPs
438 were identified using Freebayes [37]. VCFtools (0.1.15) [38] was used on the resulting vcf
439 files to identify SNPs per 10Kb.
440
441 **Validation of expression profiles by qRT-PCR**

442 Validation of the RNA-Seq experiment was completed with the Universal Probe Library
443 (UPL) RT-qPCR system (Roche Diagnostics ©). RNA samples analyzed were those used
444 for RNAseq analyses (aphid collections from location 1) as well as samples from aphids
445 collected at a separate location (location 2) and adapted to secondary hosts (3 biological
446 replicates) . For all experiments, aphid RNA was extracted using RNeasy Plant Mini Kit
447 (Qiagen). RNA samples were DNAse treated with Ambion® TURBO DNA-free™.
448 SuperScript® III Reverse Transcriptase (Invitrogen) and random primers were used to
449 prepare cDNA. Primers and probes were designed using the predicted genes sequences
450 generated in the RNA-seq data analysis and the Assay Design Center from Roche,
451 selecting "Other organism" (https://lifescience.roche.com/en_gb/brands/universal-probe-
452 library.html). Primers were computationally checked to assess if they would amplify one
453 single product using Emboss PrimerSearch. Primers and probes were validated for
454 efficiency (86-108 %) before gene expression quantification; five dilutions of threefolds for
455 each primer pair-probe were used for generating the standard curve. The 1:10 dilution of
456 cDNA was selected as optimal for RT-qPCR using the UPL system. Reactions were
457 prepared using 25µl of total volume, 12.5µl of FastStart TaqMan Probe Master Mix
458 (containing ROX reference dye), 0.25µl of gene-specific primers (0.2mM) and probes
459 (0.1mM). Step-One thermocycler (Applied Biosystems by Life Technology©) was set up as
460 follows: 10 min of denaturation at 95°C, followed by 40 cycles of 15 s at 94°C and 60 s at
461 60°C. Relative expression was calculated with the method ΔCt (Delta Cycle threshold)
462 with primer efficiency consideration. Three technical replicates were run per sample.
463 Reference genes for normalization of the cycle threshold values were selected base on
464 constant expression across different conditions in the RNA-Seq experiment. The reference
465 genes were CDC42-Kinase (Mca01274), actin (Mca10020) and tubulin (Mca04511). The
466 fold change calculations were done by ΔΔ Ct method (Delta Cycle threshold) and primer
467 efficiency was taken into consideration.
468
469
470
471
472
473
474
475
476

Acknowledgement

We thank Brian Fenton for help with adapting *Myzus cerasi* collected from cherry trees to secondary hosts, and Melanie Febrer for advice on the RNAseq experiment and sample processing.

Funding

This work was supported by the Biotechnology and Biological Sciences (BB/R011311/1 to SEvdA), European Research Council (310190-APHIDHOST to JIBB), and Royal Society of Edinburgh (fellowship to JIBB).

Availability of data and materials

All data are available under accession numbers PRJEB24338. *Myzus cerasi* genome and annotation was downloaded from http://bipaa.genouest.org/is/aphidbase/ and DOI:10.5281/zenodo.1252934. All custom python scripts used to analyse the data use Biopython [13], as well as details on how they where applied for data analyses are available on https://github.com/peterthorpe5/Myzus.cerasi_hosts.methods, and DOI: 10.5281/zenodo.1254453.

Authors' contributions

PT and JIBB conceived the experiments, PT and CEM performed sample preparation and qRT-PCR analyses, PT, CEM, SEvsA and JIBB analyzed data, PT and JB wrote the manuscript, with input from all authors. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Figures

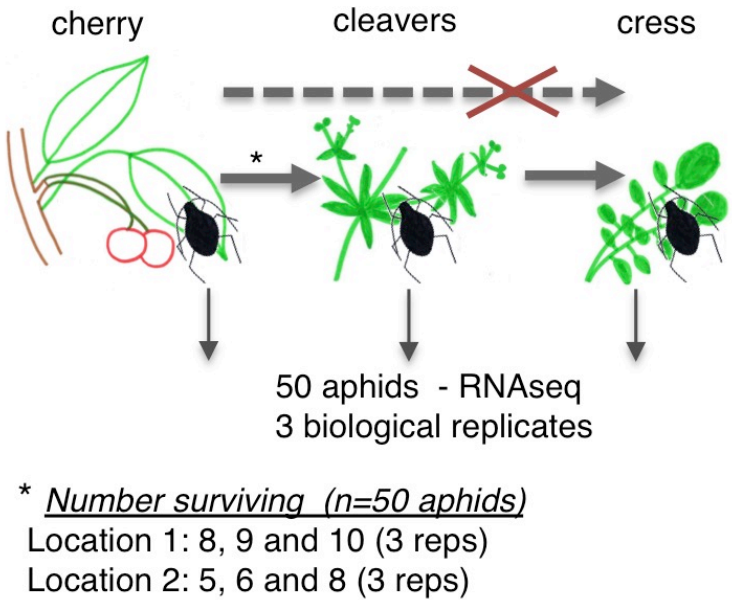


Fig. 1

Schematic overview of host switch experiments and aphid survival rates. *Myzus cerasi* aphids were collected from cherry trees at two separate locations. None of the aphids collected from cherry was able to survive directly on *Barbarae verna* (Land cress) plants. However, a 10-20% survival rate was recorded when aphids were moved onto *Galium aparine* (cleavers). The host adaptation experiments were performed in 3 biological replicates.

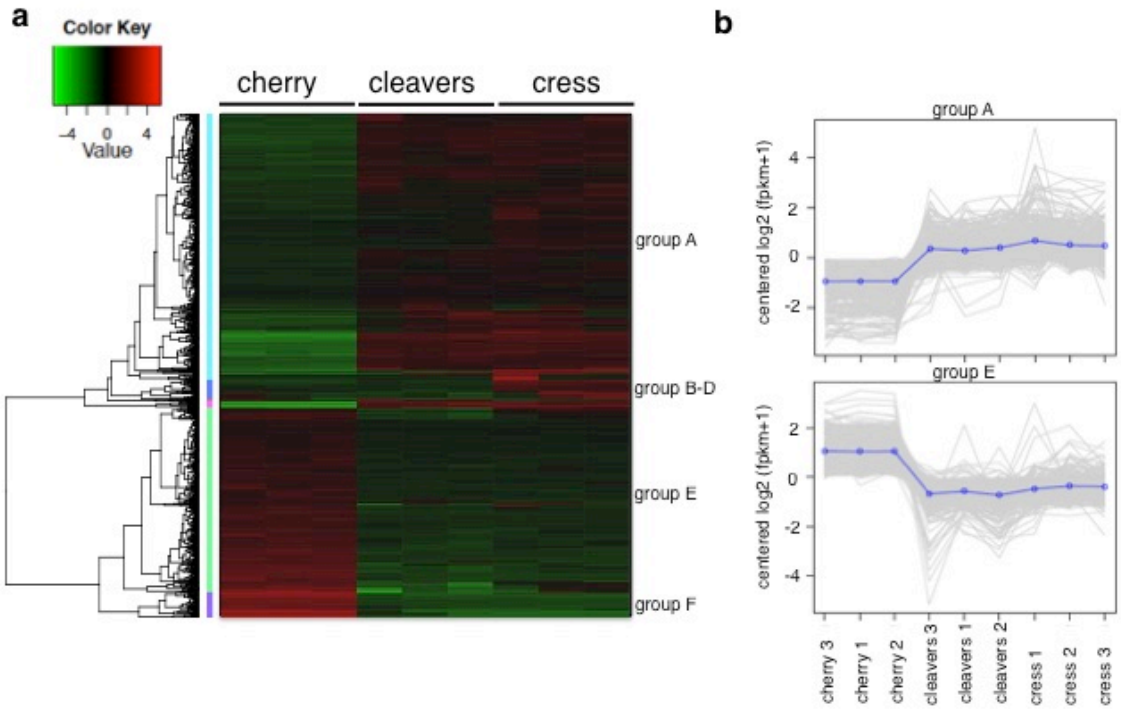


Fig. 2

Clustering of differentially expressed genes across *Myzus cerasi* populations from primary and secondary hosts.

a. Cluster analyses of the 934 genes differentially expressed in *M. cerasi* populations from different host species.

b. Expression profiles of the 493 co-regulated genes in cluster A and of the 342 co-regulated genes in cluster E.

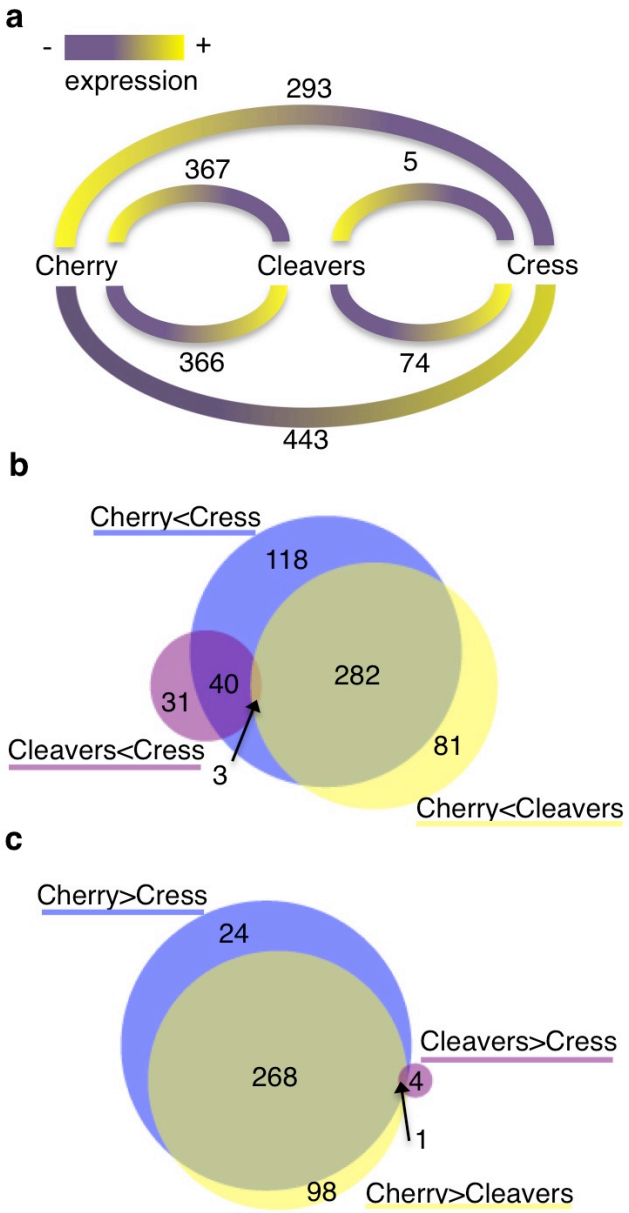


Fig. 3

Differentially expressed genes in pairwise comparisons between the different *Myzus cerasi* populations.

a. Numbers of genes for each pairwise comparison between aphids collected form the different primary and secondary host species, cherry, cleavers and cress. Yellow color indicates high level of expression, whereas purple color indicates low expression in the different pairwise comparisons.

b. Venn diagram showing the overlap in differentially expressed genes sets that are lower expressed in the aphids from primary host cherry compared to those collected from

secondary hosts cleavers and cress, and also lower expressed in aphids from cleavers than those from cress.

c. Venn diagram showing the overlap in differentially expressed genes sets that are higher expressed in the aphids from primary host cherry compared to those collected from secondary hosts cleavers and cress, and also higher expressed in aphids from cleavers than those from cress.

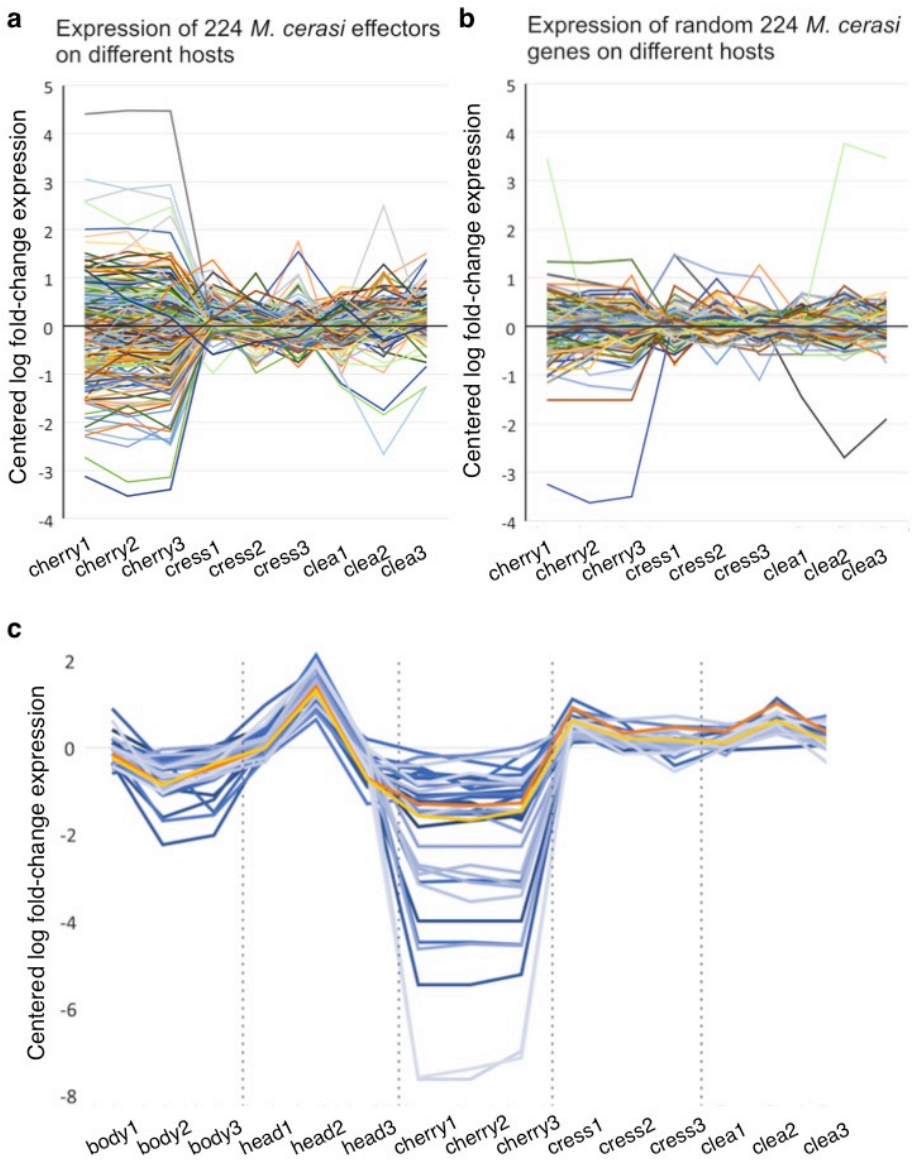


Fig. 4

Myzus cerasi effector gene expression profiles across populations from different host species.

a. Mean centered log fold-change expression of 224 *M. cerasi* putative effectors across aphid populations from different host plant species, including cherry (cherry1-3), cress (cress1-3) and cleavers (clea1-3).

b. Mean centered log fold-change expression of 224 *M. cerasi* randomly selected genes across aphid populations from different host plant species, including cherry (cherry1-3), cress (cress1-3) and cleavers (clea1-3).

c. Identification of all other genes in the *M. cerasi* genome that are co-regulated with the Mp1:Me10-like pair based on a >90% Pearson's correlation across different populations (blue, n = 35). Mp1-like is indicated in orange and Me10-like in yellow.

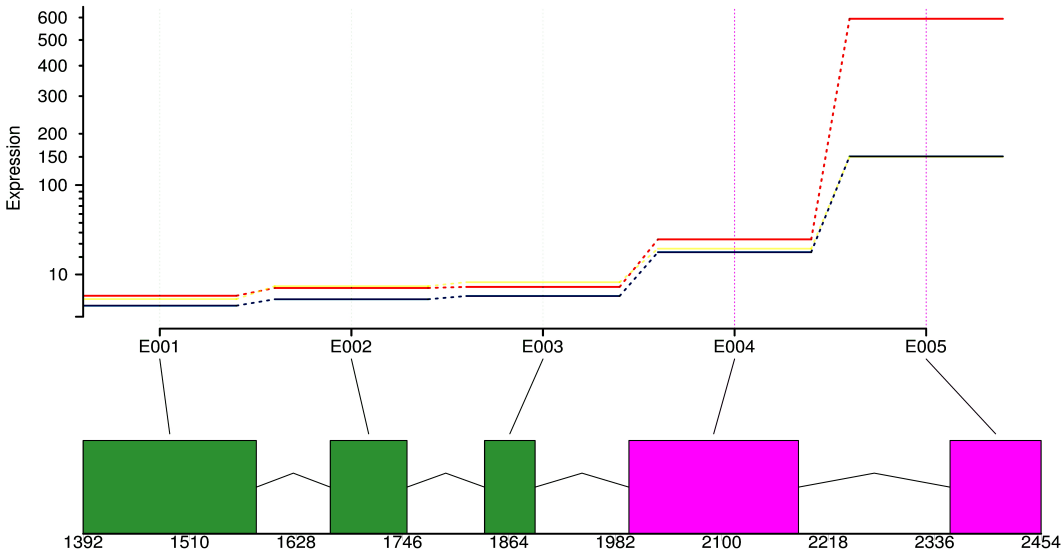
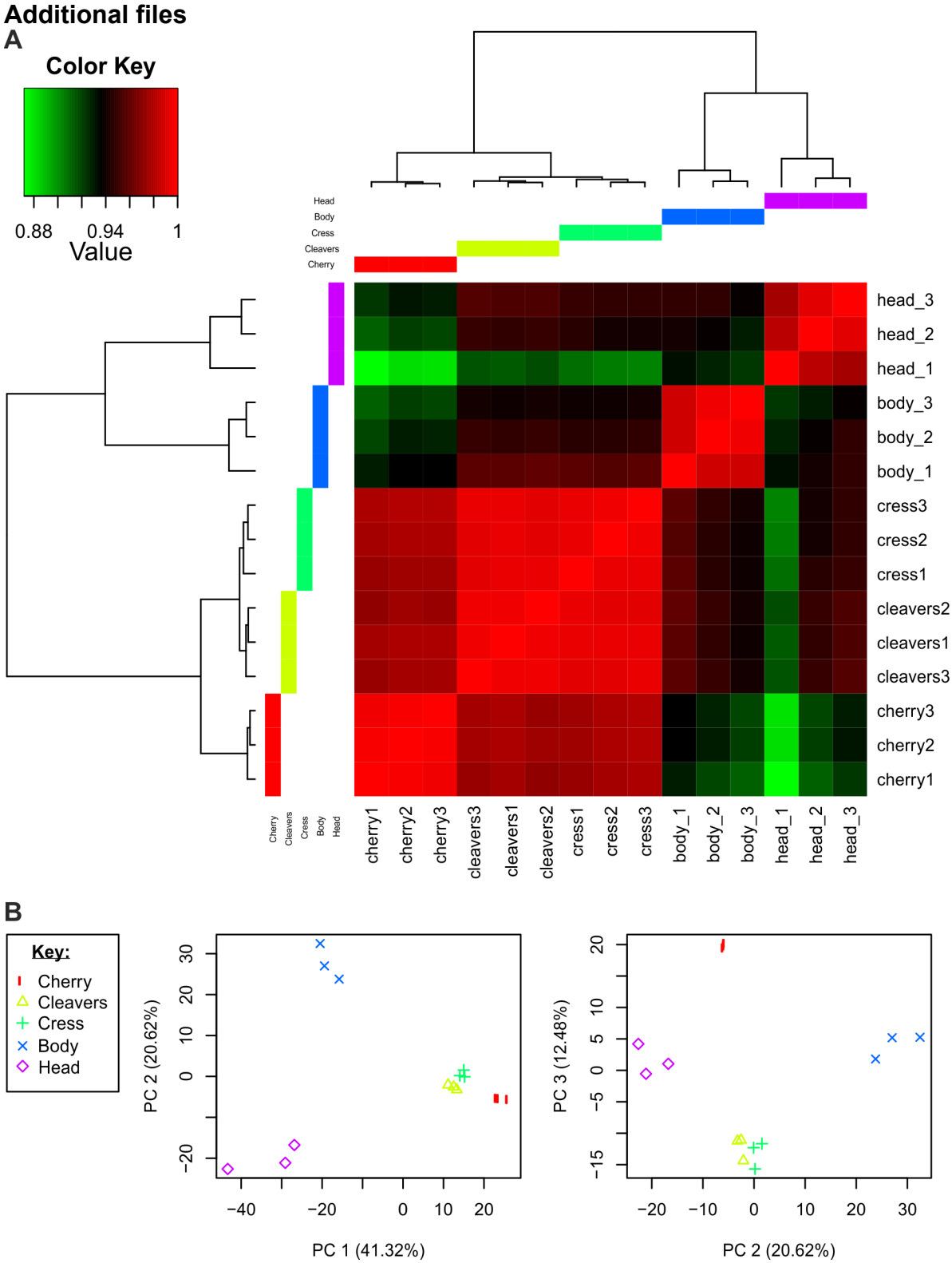


Fig. 5

Graphical representation of differential exon usage observed in gene Mca06436|peroxidase-like in the transcriptome of *Myzus cerasi* populations from different host species, including cherry (red line), cress (yellow line) and cleavers (blue line). The five different exons are indicated by E001-E005 and exons displaying significant differential expression are coloured pink. Numbers indicate nucleotide start and end positions of the different exons. The last exon shows 4 times greater expression in aphids collected from cherry compared to those from cleavers or cress.

Additional files



Additional file 1: Fig. S1. Transcriptome differences between *Myzus cerasi* populations from different host species

Genome-wide analysis of *M. cerasi* transcriptional responses to interaction primary host cherry or secondary hosts cress and cleavers, and comparison to previously published tissue-specific transcriptome of dissected heads and bodies [11].

a. Clustering of transcriptional responses reveals that *M. cerasi* gene expression is different in populations from the different host species and also that expression in head and body tissues can be separated based on these analyses

b. Principle component analysis. The top 3 most informative principle components describe approximately 75% of the variation, and separate the both the host species interaction data as well as tissue-specific data well.

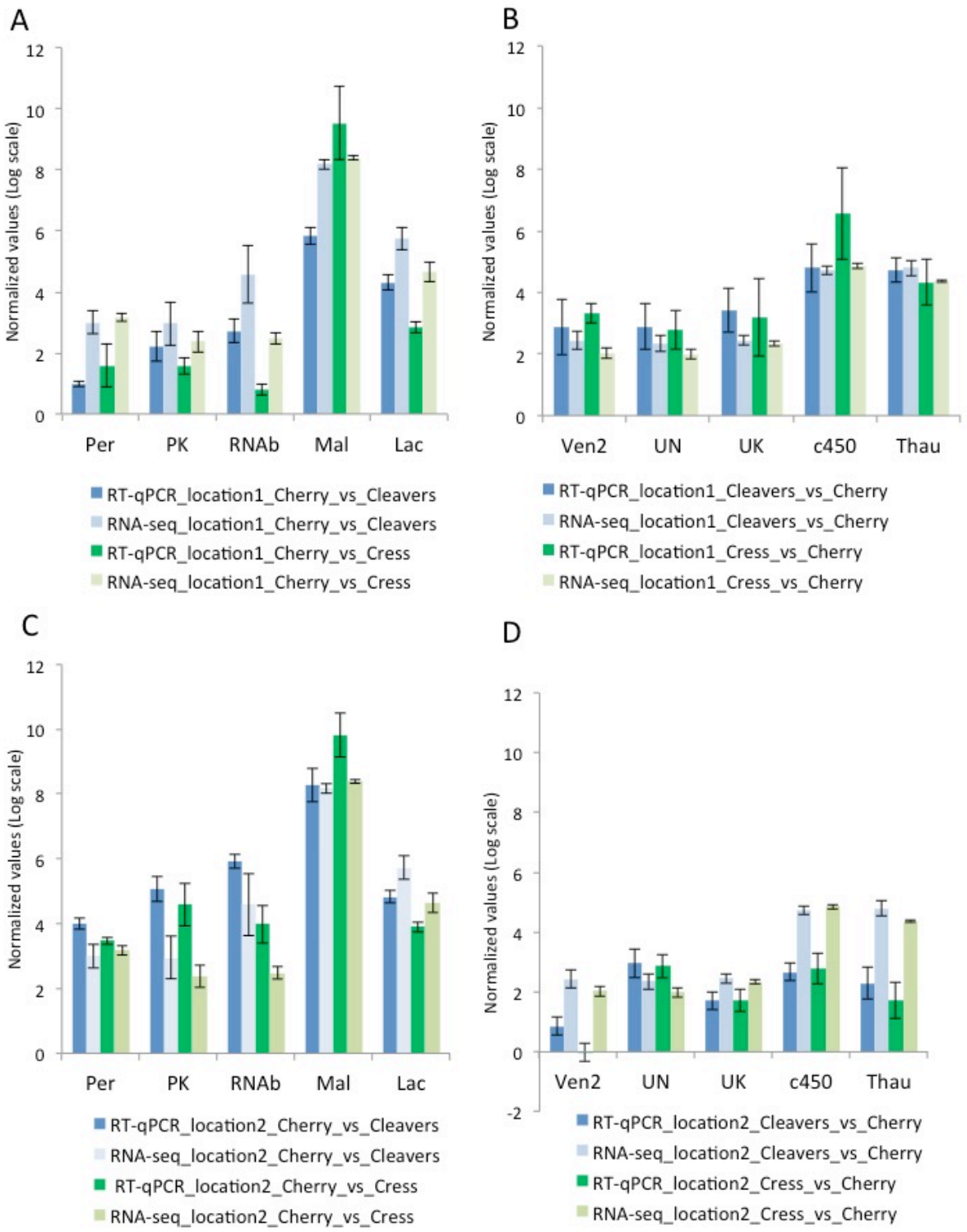
Additional file 2: Table S1. List of 934 differentially *Myzus cerasi* genes across different host species interactions

Additional file 3: Table S2. List of significant GO-terms associated with genes differentially expressed in cluster A and E (Fig. 2).

Additional file 4: Table S3. List of significant GO-terms associated with genes differentially expressed across different *Myzus cerasi* host interactions, corresponding to Fig. 3a.

Additional file 5: Table S4. List of significant GO-terms associated with genes differentially expressed across different *Myzus cerasi* host interactions, corresponding to Fig. 3b.

Additional file 6: Table S5. List of significant GO-terms associated with genes differentially expressed across different *Myzus cerasi* host interactions, corresponding to Fig. 3c.



Additional file 7: Fig.S2. Validation of differential gene expression by qRT-PCR

a. Genes up-regulated during primary host (cherry) versus secondary host interactions (cleavers and cress) in the *Myzus cerasi* population collected from location 1.

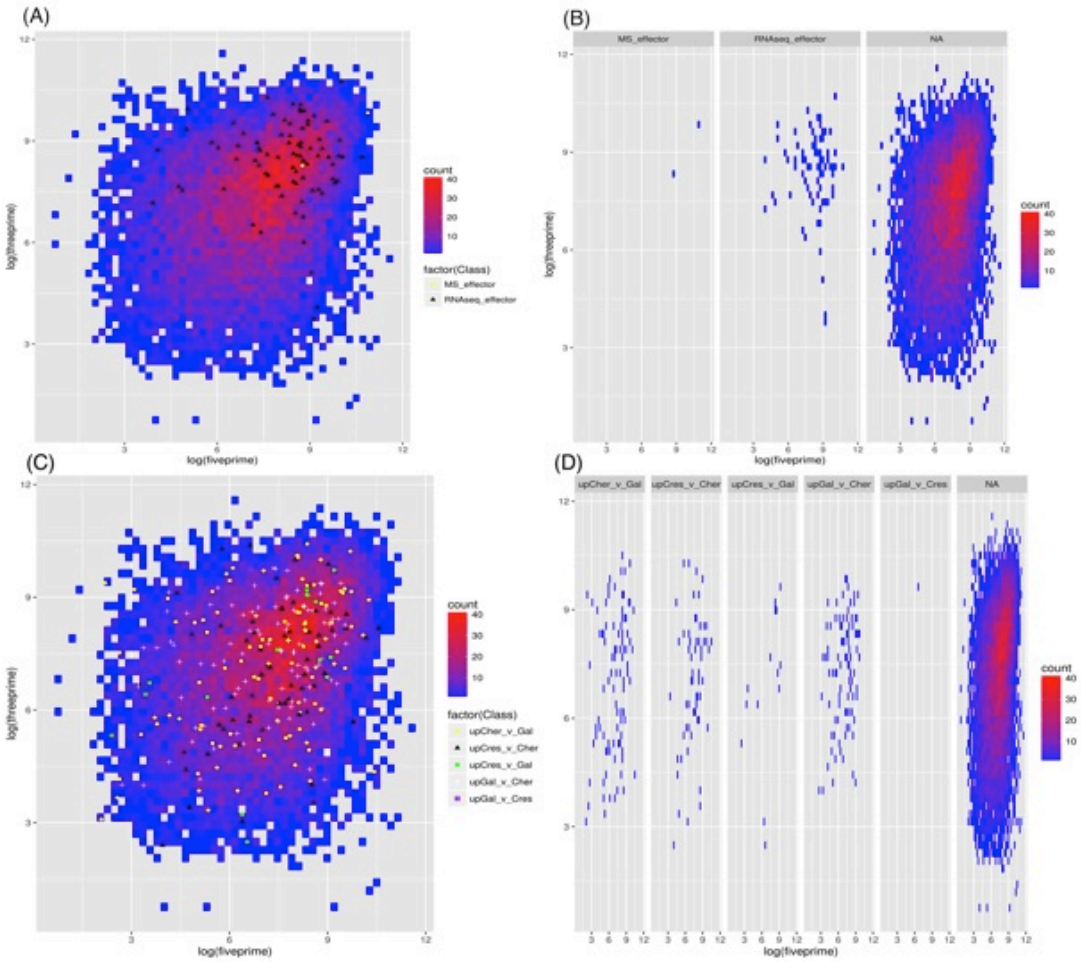
b. Genes up-regulated during the secondary host versus the primary host interactions in the *M. cerasi* population collected from location 1.

c. Genes up-regulated during primary host (cherry) versus secondary host interactions (cleavers and cress) in the *M. cerasi* population collected from location 2.
d. Genes up-regulated during the secondary host versus the primary host interactions in the *M. cerasi* population collected from location 2.
The validated genes up-regulated during the primary host interactions were peroxidase (Mca14094-Per), protein kinase (Mca07516-PK), RNA binding (Mca07514-RNAb), maltase (Mca25862-Mal) and lactase (Mca19306-Lac). Validated genes up-regulated during the secondary host interactions were venom protein (Mca05785-Ven), uncharacterized protein (Mca06816-UN), unknown protein (Mca06864-UK), cytochrome 450 (Mca22662-c450) and thaumatin (Mca12232-Thau). Blue and green series represent RT-qPCR validation results and pale blue and pale green represent RNA-seq results.
Error bars indicate standard error.

Additional file 8: Table S6. List of 224 *Myzus cerasi* putative effectors and their expression levels across different host species interactions.

Additional file 9: Table S7. List of *Myzus cerasi* genes showing differential exon usage across different host species interactions.

Additional file 10: Table S8. Gene duplication types in *Myzus cerasi* genes differentially expressed across different host species interactions.



Additional file 11: Figure S3. Heat maps graphically representing the LOG nucleotide distance from one gene to its neighboring genes in a 3'- and 5' -direction. Various gene categories are colored and coded in the relevant keys. (A) and (B) Geneic distance heat map for predicted effectors, which were significantly further away from their neighboring genes and thus in gene sparse regions [10]. (C) and (D) Genic distances for differentially expressed genes identified in this study. These are not significantly further away from their neighboring genes in either 3'- or 5'-direction.

References

1. Mordvilko AK: **The evolution of cycles and the evolution of heteroecy (migration) in plant lice** *Annu Mag Nat Hist* 1928, **2**:570-582.
2. Williams ISaD, A.F.G.: **Life cycles and polymorphism. Aphids as crop pests**: CABI International, Wallingford; 2007.
3. Kennedy JS, Stroyan, H. L. G.: **Biology of Aphids**. *Annu Rev Entomol* 1959, **4**:139-160.
4. Moran NA: **The evolution of host-plant alternation in aphids: evidence for specialization as a dead end.** *Am Naturalist* 1988:681-706.

5. Jaquiéry J, Stoeckel S, Nouhaud P, Mieuzet L, Mahéo F, Legeai F, Bernard N, Bonvoisin A, Vitalis R, Simon JC: **Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex**. *Mol Ecol* 2012, **21**(21):5251-5264.
6. Mathers TC, Chen Y, Kaithakottil G, Legeai F, Mugford ST, Baa-Puyoulet P, Bretaudeau A, Clavijo B, Colella S, Collin O *et al*: **Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species**. *Genome Biol* 2017, **18**(1):27.
7. Cui N, Yang PC, Guo K, Kang L, Cui F: **Large-scale gene expression reveals different adaptations of Hyalopterus persikonus to winter and summer host plants**. *Insect Sci* 2017, **24**(3):431-442.
8. Barbagallo S, Cocuzza GEM, Cravedi P, Shinkichi K: **IP Case Studies: Decidious Fruit Tree Aphids. *Aphids as Crop Pests 2nd Edition***: CABI International, Wallingford; 2017.
9. Blackman R, Eastop V: **Aphids on the World's Crops: An Identification and Information Guide**. In.: Wiley-Blackwell; 2000.
10. Thorpe P, Escudero-Martinez C, Cock P, Laetsch D, Eves-van den Akker S, Bos J: **Shared transcriptional control and disparate gain and loss of aphid parasitism genes and loci acquired via horizontal gene transfer**. *bioRxiv* 2018.
11. Thorpe P, Cock PJ, Bos J: **Comparative transcriptomics and proteomics of three different aphid species identifies core and diverse effector sets**. *BMC genomics* 2016, **17**(1):1.
12. Moran NA, Jarvik T: **Lateral transfer of genes from fungi underlies carotenoid production in aphids**. *Science* 2010, **328**(5978):624-627.
13. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B: **Biopython: freely available Python tools for computational molecular biology and bioinformatics**. *Bioinformatics* 2009, **25**(11):1422-1423.
14. Andrews S: **FastQC: A quality control tool for high throughput sequence data**. *Reference Source* 2010.
15. Bolger A, Giorgi F: **Trimmomatic: A Flexible Read Trimming Tool for Illumina NGS Data**. *URL* http://www *usadellab org/cms/index php*.
16. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q: **Full-length transcriptome assembly from RNA-Seq data without a reference genome**. *Nature biotechnology* 2011, **29**(7):644-652.
17. Smith-Unna RD, Boursnell C, Patro R, Hibberd JM, Kelly S: **TransRate: reference free quality assessment of de-novo transcriptome assemblies**. *BioRxiv* 2015:021626.
18. Langmead B: **Aligning short sequencing reads with Bowtie**. *Current protocols in bioinformatics* 2010:11.17. 11-11.17. 14.
19. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics* 2010, **26**(1):139-140.
20. Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, Wrobel N, Gharbi K, Simpson GG, Owen-Hughes T: **How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?** *RNA* 2016, **22**(6):839-851.
21. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M: **De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis**. *Nature protocols* 2013, **8**(8):1494-1512.

22. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching**. *Nucleic acids research* 2011, **39**(suppl 2):W29-W37.
23. Yang Y, Smith SA: **Optimizing de novo assembly of short-read RNA-seq data for phylogenomics**. *BMC genomics* 2013, **14**(1):328.
24. Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0**. *Journal of molecular biology* 2004, **340**(4):783-795.
25. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes**. *Journal of molecular biology* 2001, **305**(3):567-580.
26. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications**. *BMC bioinformatics* 2009, **10**(1):421.
27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene Ontology: tool for the unification of biology**. *Nature genetics* 2000, **25**(1):25-29.
28. Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, Arnold R, Rattei T, Letunic I, Doerks T: **eggNOG v3. 0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges**. *Nucleic acids research* 2012, **40**(D1):D284-D289.
29. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW: **RNAmmer: consistent and rapid annotation of ribosomal RNA genes**. *Nucleic acids research* 2007, **35**(9):3100-3108.
30. Buchfink B, Xie C, Huson DH: **Fast and sensitive protein alignment using DIAMOND**. *Nature methods* 2015, **12**(1):59-60.
31. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: **STAR: ultrafast universal RNA-seq aligner**. *Bioinformatics* 2013, **29**(1):15-21.
32. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics* 2010, **26**(6):841-842.
33. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research**. *Bioinformatics* 2005, **21**(18):3674-3676.
34. Gremme G, Steinbiss S, Kurtz S: **GenomeTools: a comprehensive software library for efficient processing of structured genome annotations**. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on* 2013, **10**(3):645-656.
35. Anders S, Pyl PT, Huber W: **HTSeq--a Python framework to work with high-throughput sequencing data**. *Bioinformatics* 2015, **31**(2):166-169.
36. Anders S, Reyes A, Huber W: **Detecting differential usage of exons from RNA-seq data**. *Genome research* 2012, **22**(10):2008-2017.
37. Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** *arXiv12073907* 2012.
38. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al*: **The variant call format and VCFtools**. *Bioinformatics* 2011, **27**(15):2156-2158.