

 Open access • Posted Content • DOI:10.1101/366450

## **Extensive transcriptional changes in the aphid species *Myzus cerasi* under different host environments associated with detoxification genes** — [Source link](#)

Peter Thorpe, Carmen Escudero-Martinez, Carmen Escudero-Martinez, Sebastian Eves-van den Akker ...+2 more authors

**Institutions:** James Hutton Institute, University of Dundee, University of Cambridge

**Published on:** 30 May 2019 - bioRxiv (Cold Spring Harbor Laboratory)

**Topics:** Myzus cerasi, Intermediate host, Host (biology) and Aphid

Related papers:

- [Transcriptional changes in the aphid species \*Myzus cerasi\* under different host and environmental conditions](#)
- [Extensive \*Myzus cerasi\* transcriptional changes associated with detoxification genes upon primary to secondary host alternation](#)
- [Expression differences in \*Aphidius ervi\* \(Hymenoptera: Braconidae\) females reared on different aphid host species](#)
- [Plant resistance in different cell layers affects aphid probing and feeding behaviour during non-host and poor-host interactions.](#)
- [The chemical ecology of aphid host alternation: How do return migrants find the primary host plant?](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/extensive-transcriptional-changes-in-the-aphid-species-myzus-q9o50ol4uo>

1 **Extensive transcriptional changes in the aphid species *Myzus cerasi* under different**  
2 **host environments associated with detoxification genes**

3

4 Peter Thorpe<sup>1</sup>, Carmen M. Escudero-Martinez<sup>1,2</sup>, Sebastian Eves-van den Akker<sup>3</sup>, Jorunn  
5 I.B. Bos<sup>1,2</sup>

6

7

8 <sup>1</sup>Cell and Molecular Sciences, The James Hutton Institute, Invergowrie, Dundee, DD2  
9 5DA, UK

10 <sup>2</sup>Division of Plant Sciences, School of Life Sciences, University of Dundee, Dundee

11 <sup>3</sup>Department of Plant Sciences, University of Cambridge, CB2 3EA, UK

12

13

14 \*Corresponding Author:

15 Jorunn I.B. Bos, PhD

16 Division of Plant Sciences, School of Life Sciences, University of Dundee

17 Cell and Molecular Sciences, James Hutton Institute

18 Invergowrie, Dundee

19 DD2 5DA

20 United Kingdom

21

22 Phone: +44 (0)3449285428

23 Email: [j.bos@dundee.ac.uk](mailto:j.bos@dundee.ac.uk)

24

25

26

27

28

29

30 **Abstract**

31

32 Aphids are phloem-feeding insects that cause yield losses to crops globally. These insects  
33 feature complex life cycles, which in the case of many agriculturally important species  
34 involves the use of primary and secondary host plant species. Whilst host alternation  
35 between primary and secondary host can occur in the field depending on host availability  
36 and the environment, aphid populations maintained as laboratory stocks generally are kept  
37 under conditions that allow asexual reproduction by parthenogenesis on secondary hosts.  
38 Here, we used *Myzus cerasi* (black cherry aphid) to assess aphid transcriptional  
39 differences between populations collected from primary hosts in the field and those  
40 adapted to secondary hosts under controlled environment conditions. Adaptation  
41 experiments of *M. cerasi* collected from local cherry trees to reported secondary host  
42 species resulted in low survival rates. Moreover, aphids were unable to survive on  
43 secondary host Land cress, unless first adapted to another secondary host, cleavers.  
44 Transcriptome analyses of populations collected from primary host cherry in the field and  
45 the two secondary host plant species in a controlled environment showed extensive  
46 transcriptional plasticity to a change in host environment, with predominantly genes  
47 involved in redox reactions differentially regulated. Most of the differentially expressed  
48 genes across the *M. cerasi* populations from the different host environments were  
49 duplicated and we found evidence for differential exon usage. In contrast, we observed  
50 only limited transcriptional to a change in secondary host plant species.

51

52 **Keywords: aphid host adaptation, laboratory environment, RNAseq, transcriptional**  
53 **plasticity, detoxification**

54

55 **Introduction**

56

57 Aphids are phloem-feeding insects that belong to the order Hemiptera. Insects within this  
58 order feature distinctive mouthparts, or stylets, that in the case of phytophagous insects,  
59 are used to pierce plant tissues and obtain nutrients from the plant phloem. One striking  
60 feature of the complex life cycle of about 10 % of aphid species is the seasonal host  
61 switching between unrelated primary (winter) and secondary (summer) host plants, also  
62 called host alternation or heteroecy (Mordvilko 1928); (Williams 2007). Host alternating  
63 aphids predominantly use woody plants as their primary hosts, on which (overwintering)  
64 eggs are laid, from which the first generation of aphids, or fundatrices, emerge in spring.  
65 The fundatrices, and their offspring, reproduce by parthenogenesis (asexual reproduction),  
66 giving birth to live nymphs. Winged forms (alate) will migrate to secondary host plants over  
67 the summer months where the aphid populations will go through multiple parthenogenic  
68 generations. In autumn, sexual female and male aphids will reproduce sexually and  
69 overwintering eggs are laid on the primary host. Exceptions to this general life cycle exist,  
70 with some aphids for example having multi-year cycles (Kennedy 1959).

71

72 Heteroecy in aphids has independently arisen in different aphid lineages throughout  
73 evolutionary history (Moran 1988) with monoecy (with the entire life cycle taking place on  
74 one plant species) on trees thought to be the ancestral state. Many different hypotheses  
75 explain the maintenance of heteroecy and driving factors described include nutritional  
76 optimization, oviposition sites, natural enemies, temperature tolerance, and fundatrix

77 specialization (Moran 1988). It is likely that switching between host plant species requires  
78 aphids to adapt to differences in host nutritional status as well as potential differences in  
79 plant defense mechanisms against insects. Host plant specialization in the pea aphid  
80 species complex is associated with differences in genomic regions encompassing  
81 predicted salivary genes as well as olfactory receptors (Jaquiéry et al. 2012). Moreover,  
82 adaptation of *M. persicae* to different secondary host plant species involves gene  
83 expression changes, including of genes predicted to encode for cuticular proteins,  
84 cathepsin B protease, UDP-glycosyltransferases and P450 monooxygenases (Mathers et  
85 al. 2017). Aphid secondary hosts include many important agricultural crops and are  
86 generally more suitable for maintaining clonal (asexual) aphid laboratory stocks used for  
87 research experiments. To what extent aphid gene expression is affected upon collecting  
88 aphids from the field and adapting them to select secondary host plants in a laboratory  
89 environment remains unclear.

90  
91 *Myzus cerasi*, or black cherry aphid, uses mainly *Prunus cerasus* (Morello cherry) and  
92 *Prunus avium* (sweet cherry), but also other *Prunus* species as primary hosts and several  
93 herbaceous plants (*Gallium* spp., *Veronica* spp., and cruciferous species) as secondary  
94 hosts (Barbagallo et al. 2017) (Blackman and Eastop 2000). Infestation can cause  
95 significant damage on cherry trees, due to leaf curling, and shoot deformation, pseudogall  
96 formation, as well as fruit damage. Recently, we generated a draft genome for *M. cerasi*,  
97 providing novel insights into potential parasitism genes as well as genome evolution  
98 (Thorpe et al. 2018). The increasing availability of genomics resources for aphids,  
99 including *M. cerasi*, facilitates further understanding of aphid biology, including the  
100 processes involved adapting to a change in host environment. In this study, we collected  
101 *M. cerasi* from primary hosts (cherry trees) in the field with the aim to adapt these to  
102 different reported secondary hosts, *Galium aparine* (cleavers) and *Barbarea verna* (Land

103 cress) under laboratory conditions for comparative transcriptome analyses. We found that  
104 aphids collected from their primary host in the field differed in their ability to adapt to the  
105 secondary host plant species in a controlled environment, with no aphids surviving transfer  
106 to *Barbarea verna* (Land cress). We compared the transcriptomes of *M. cerasi* aphids  
107 adapted under laboratory conditions to secondary hosts *Galium aparine* (cleavers) and  
108 *Barbarea verna* (Land cress) and only observed limited transcriptional changes. However,  
109 when comparing the transcriptomes of these adapted aphids to field collected aphids from  
110 primary hosts, we noted extensive transcriptional changes, especially with regards to  
111 predicted detoxification genes. The majority of differentially expressed genes were  
112 duplicated, implicating multigene families in aphid adaptation to host environments.

113

## 114 **Results and discussion**

115

### 116 ***Myzus cerasi* host adaptation under controlled laboratory conditions is associated** 117 **with low survival rates**

118

119 When attempting to establish a colony of *M. cerasi* from populations occurring on local  
120 cherry trees, we observed differences in survival rates upon transfer to reported secondary  
121 host plant species under controlled plant growth conditions. While aphids were unable to  
122 survive transfer from primary host cherry to Land cress (*Barbarea verna*), we observed a  
123 10%-20% survival rate upon transfer to cleavers (*Galium aparine*) (Fig. 1). However, once  
124 aphid populations were established on cleavers, individuals from this population were  
125 successfully transferred to cress plants. We performed similar field to lab host transfer  
126 experiments with aphids collected from cherry trees at two different locations in three  
127 independent replicates with similar results (Fig. 1). Our observation that *M. cerasi*  
128 collected from primary hosts in the field shows a difference in ability to infest two reported

129 secondary hosts likely reflects these aphids face a major hurdle represented by a change  
130 in host environment. In addition, our finding raises the question whether aphids may  
131 expand their secondary host range once adapted to a preferred secondary host species.

132

133 ***Myzus cerasi* shows extensive transcriptional plasticity to a change in host**  
134 **environment**

135

136 We assessed the changes that take place at the transcriptional level in *M. cerasi* when  
137 adapting the field-collected aphids from cherry to secondary hosts in a controlled plant  
138 growth environment. Specifically, we sequenced the transcriptomes of *M. cerasi*  
139 populations collected from cherry (field conditions), and of aphids established over a 3-  
140 week period on cleavers or cress (controlled environment) using RNAseq.

141

142 We performed differential gene expression analysis (LOG fold change >2, False Discovery  
143 Rate (FDR)  $p < 0.001$ ) between the different aphid populations to identify gene sets  
144 associated with the different host plant environments. Cluster analyses of the aphid  
145 transcriptional responses from this and previous work reporting on differential aphid gene  
146 expression in head versus body tissues (Thorpe et al. 2016) revealed that the overall  
147 expression profiles could be distinguished based on the aphid tissue used for sample  
148 preparation as well as the host environment (Fig. S1A). Indeed, principal component  
149 analyses showed a clear separation between aphid transcriptomes associated with  
150 primary host (field conditions) and the different secondary hosts (controlled conditions)  
151 (Fig. S1B). Overall, we identified 934 differentially expressed genes by comparing the  
152 different datasets for each of the aphid populations (Fig. 2A, Table S1). A heat map of  
153 these 934 genes shows that gene expression profiles from aphids adapted to secondary  
154 hosts (cleavers and cress) and maintained in a controlled environment are more similar to

155 each other than to the gene expression profiles of aphids collected from primary hosts in  
156 the field (Fig. 2A). Co-expression analyses reveals six main clusters of differentially  
157 expressed genes, two of which (A and E) contain the majority of genes (Fig. 2A and 2B).  
158 Cluster A contains 493 genes, which show higher expression in aphids maintained on  
159 secondary hosts (controlled environment) versus those collected from primary hosts (field),  
160 and cluster B contained 342 genes showing an opposite profile. GO annotation revealed  
161 over-representation of terms associated with oxido-reductase activity in both clusters, as  
162 well as several terms associated with carotenoid/tetrapenoid biosynthesis in case of  
163 cluster B (Table S2).

164

165 To assess differential expression of *M. cerasi* genes across the host environment-aphid  
166 interactions we also analyzed pairwise comparisons for differentially expressed gene sets.  
167 The largest set of differentially expressed genes (736) was found in comparisons between  
168 aphids from cherry (primary host, field) and cress (secondary host, controlled  
169 environment), with 443 genes more highly expressed in aphids from cress, and 293 more  
170 highly expressed in aphids from cherry (Fig. 3A). A total of 733 differentially expressed  
171 genes were found in comparisons of aphids from cherry (primary host, field) versus  
172 cleavers (secondary host, controlled environment, with 367 genes more highly expressed  
173 in aphids collected from cherry and 366 genes more highly expressed in aphids collected  
174 from cleavers (Fig. 3A). The higher number of genes up-regulated in the cress-cherry  
175 comparison may reflect the difficulties in adapting to this secondary host species, with *M.*  
176 *cerasi* unable to infest cress when collected from cherry (field). Our differential gene  
177 expression data not only is impacted by a change in host plant species, but also a change  
178 in plant growth environment to which the aphids were exposed. To what extent both these  
179 factors contribute the observed gene expression changes remains an open question.  
180 Nevertheless, aphids cannot make the transition to cress in the controlled environment



181 unless they first parasitise cleavers in the controlled environment. This suggests that the  
182 transition to cleavers enables subsequent transitions through a yet unidentified  
183 mechanism.

184

185 A relatively small number of genes were differentially expressed between aphids collected  
186 from the two secondary hosts cleavers and cress (both grown under controlled conditions),  
187 with only 5 genes more highly expressed in aphids from cleavers, and 74 genes more  
188 highly expressed in aphids from cress (Fig. 3A). This suggests that *M. cerasi* shows limited  
189 transcriptional plasticity to a switch in secondary host environment, once adapted. This is  
190 in line with our previous observation that only a relatively small set of genes is differentially  
191 expressed in *M. persicae* and *R. padi* when exposed to different host or non-/poor-host  
192 plants (Thorpe et al. 2018) as well as the relatively small number of transcriptional  
193 changes when *M. persicae* is adapted to different secondary hosts (Mathers et al. 2017).

194

195 GO enrichment analyses of the 443 genes more highly expressed in aphids collected from  
196 cress (controlled environment) compared to those collected from cherry (field) shows  
197 overrepresentation of genes predicted to be involved various processes, including in heme  
198 binding (GO:0020037), tetrapyrrole binding (GO:0046906), monooxygenase activity  
199 (GO:0004497), oxidoreductase activity (GO:0016705), iron ion binding (GO:0005506), and  
200 hydrolase activity (GO:0016787) (Table S3). This set of 443 genes contains 282 of the 366  
201 genes that are also more highly expressed in aphids from the other secondary host plant  
202 species, cleavers, with similar GO annotations (Fig. 3B; Table S4). The 293 genes more  
203 highly expressed in aphids collected from cherry (field) than those from cress (controlled  
204 environment) shows over-representation of genes predicted in oxidoreductase activity  
205 (GO:0016620, GO:0016903, GO:0055114, GO:001649) as well as other processes such  
206 as fatty-acyl-CoA reductase (alcohol-forming) activity (GO:0080019), interspecies

207 interaction between organisms (GO:0044419), and symbiosis (GO:0044403) (Table S3).  
208 For the gene sets differentially expressed between aphids collected from cherry (field) and  
209 cleavers (controlled environment), GO enrichment analyses reveal that in reciprocal  
210 comparisons genes predicted to function in redox reactions are also over-represented  
211 (Table S3). The over-representation of differentially expressed genes involved in redox  
212 across the different *M. cerasi* populations likely reflects different requirements for aphids  
213 under different host environment conditions.

214

215 Interestingly, among the 367 genes more highly expressed in aphids collected from cherry  
216 (field) compared to those collected from cleavers (controlled environment), we found that  
217 the majority of GO terms identified through enrichment analyses correspond to metabolic  
218 processes (Table S3). Of these 367 transcripts 268 show similar expression differences in  
219 aphids collected from cherry (field) versus those collected from cress (controlled  
220 environment), whereas 98 are specific to the comparison of aphids collected from cherry  
221 (field) versus cleavers (controlled environment) (Fig. 3C). Whilst GO enrichment analyses  
222 showed over-representation of genes involved in redox reactions in the set of 268  
223 overlapping transcripts, the 98 transcripts specifically up-regulated in aphids collected from  
224 cherry (field) versus cleavers (controlled environment) show over-representation in  
225 metabolic processes, and especially those associated with terpenoid/carotenoid  
226 biosynthesis which are involved in aphid pigmentation (Table S5) (Moran and Jarvik  
227 2010). Possibly this observation reflects that *M. cerasi* requires specific gene sets for  
228 pigmentation and feeding under specific host environmental conditions. Notably, we did  
229 not observe any noticeable change in aphid color upon adapting aphids from primary  
230 hosts in the field to secondary hosts in the lab. Aphids featured a dark brown to black color  
231 on all plant species tested (not shown), suggesting the differential regulation of carotenoid

232 genes is not associated with aphid color in this case but with other unknown physiological  
233 functions.

234

235 To independently test whether select *M. cerasi* genes were differentially expressed in  
236 aphids collected from primary (field) and secondary (controlled environment) host plants,  
237 we repeated the collection of aphids from local cherry trees (separate site, location 2) and  
238 performed adaptation experiments to cleavers and cress. We selected 10 genes for  
239 independent validation of expression profiles by qRT-PCR. Five of these 10 genes were  
240 selected based on enhanced expressed in aphids from cherry (field) compared to aphids  
241 from secondary (controlled environment) host plants, and another 5 genes for being more  
242 highly expressed in aphids from secondary (controlled environment) host plants compared  
243 to aphids from cherry (field). The genes selected based on higher expression in aphids  
244 from cherry (field) showed similarity to genes predicted to encode a peroxidase, RNA-  
245 binding protein 14-like, hybrid sensor histidine kinase response regulator, maltase isoform  
246 a, and a lactase-phlorizin hydrolase. The genes selected based on higher expression in  
247 aphids from secondary (controlled environment) hosts showed similarity to genes  
248 predicted to encode an unknown protein, a venom-like protease, a thaumatin-like protein,  
249 protein kintoun, and a cytochrome P450. Except for the gene with similarity to a venom-  
250 like protease, all genes showed a similar gene expression profile in both samples used for  
251 the RNAseq experiments and in the independently collected and adapted aphids from a  
252 different site, indicating that this gene set is consistently differentially expressed when *M.*  
253 *cerasi* was adapted from primary hosts in the field to secondary hosts in a controlled  
254 environment (Fig. S2). Most of these genes have predicted functions in detoxification, in  
255 line with our hypothesis that aphids require different sets of genes to deal with potential  
256 defensive plant compounds associated with different host environments. To what extent  
257 our observations are associated with primary versus secondary host factors or field versus

258 controlled environment factors is not clear. Notably, in *H. persikonus* collected from  
259 primary and secondary host plant species in the field, a similar observation was made in  
260 that an extensive gene set associated with detoxification was differentially regulated (Cui  
261 et al. 2017).

262

263 **Single Nucleotide Polymorphism (SNP) analyses suggest that an aphid sub-**  
264 **population is able to adapt from primary hosts in the field to secondary hosts in a**  
265 **controlled environment**

266

267 We used the transcriptome dataset we generated here to compare the level of sequence  
268 polymorphisms between the aphid populations from the different primary (field) and  
269 secondary (controlled environment) host plants species. Variants/SNPs were predicted by  
270 mapping the RNAseq dataset for each aphid population (cherry, cleavers and cress) to the  
271 *M. cerasi* reference genome for each condition, with only unique mapping being allowed.  
272 The number of SNPs within each 10Kb window was calculated. The *M. cerasi* population  
273 from cherry has significantly more SNPs per 10Kb than the populations from both cleavers  
274 and cress when mapping reads back to the reference genome ( $p < 0.001$ , Kruskal-Wallis  
275 with Bonferroni post hoc correction). In contrast, the aphid populations from cleavers and  
276 cress showed no significant difference in the number of SNPs per 10Kb ( $p = 0.29$ , Kruskal-  
277 Wallis with Bonferroni post hoc correction). These results are consistent with the  
278 observation that the population of *M. cerasi* went through a bottle neck during the transfer  
279 from cherry to cleavers, but not cleavers to cress. Based on these findings we propose  
280 that only a subpopulation of the primary host population may switch to secondary host  
281 plant species. It should be noted that these data are based on RNAseq, and do not rule  
282 out the possibility of allele-specific expression across the different host interactions.

283 Hence, further characterization of the *M. cerasi* (sub)populations using DNaseq will be  
284 required to gain further insight into adaptation of this aphid species to its hosts.

285

286 **Limited differential expression of predicted *M. cerasi* effectors across field and**  
287 **adapted populations**

288

289 We assessed whether predicted *M. cerasi* effectors are differentially expressed in aphid  
290 population collected from a primary host (field) or secondary hosts plants (controlled  
291 environment). The 224 predicted *M. cerasi* effectors we previously identified (Thorpe et al.  
292 2018) show a wide range of expression levels across different interactions, with most  
293 expression variation in aphids collected from cherry under field conditions (Fig. 4A; Table  
294 S6). However, when assessing expression of a random non-effector set of similar size, this  
295 expression variation in aphids was less pronounced (Fig. 4A). Despite the observed  
296 variation in expression patterns, we only found a small number of differentially expressed  
297 candidate effectors, mainly when comparing aphids collected from the primary (field)  
298 versus secondary hosts (controlled environment). Specifically, 13 candidate effectors are  
299 more highly expressed in aphids from both secondary host species (controlled  
300 environment) compared to aphids from the primary host cherry (field), with one additional  
301 candidate effector more highly expressed in the case of aphids from cleavers compared to  
302 cherry only (Mca17157|adenylate kinase 9-like) (Table S6). Although these candidate  
303 effectors were mainly of unknown function, several show similarity to thaumatin-like  
304 proteins and a venom protease. Interestingly, the candidate effector with similarity to the  
305 venom protease, Mca05785 (upregulated in secondary hosts (controlled environment)), is  
306 member of a venom protease gene family cluster that consists of four members (3 are  
307 tandem duplications, 1 is a proximal duplication). Three of these are predicted to encode  
308 secreted proteins, and all members show higher expression levels in aphids from

309 secondary hosts under controlled conditions compared to aphids from primary host in the  
310 field, but this variation was below the LOG2 fold change cut-off (Table S6). In addition, 1  
311 candidate effector (similar to RNA-binding protein 14) was differentially expressed when  
312 comparing aphids from the two secondary host plants (controlled environment), and 5  
313 candidate effectors (Mca07285, Mca07514, Mca16980, Mca07516, Mca09259) were more  
314 highly expressed in aphids collected from cherry (field) compared to aphids from cleavers  
315 and/or cress (controlled environment) (Table S6).

316

317 **Break-down of aphid effector co-regulation in the aphid population collected from**  
318 **primary hosts under field conditions.**

319

320 We previously showed that expression of aphid effector genes, required for parasitism,  
321 was tightly co-regulated pointing to a mechanism of shared transcriptional control (Thorpe  
322 et al. 2018). To assess co-regulation in the different *M. cerasi* populations, we analyzed  
323 the co-expression patterns of Mc1 and Me10-like, an effector pair that is physically linked  
324 across aphid genomes and tightly co-regulated, together with all other genes. A total of 35  
325 genes showed a high level of co-expression in the three different *M. cerasi* populations  
326 and across different aphid tissues (Fig. 4C). This number is much smaller compared to the  
327 set of co-regulated genes in *R. padi* (213) and *M. persicae* (114), which could be due to  
328 differences the quantity and quality of the RNAseq datasets we used for these analyses  
329 (Thorpe et al. 2018). However, the pattern of co-regulation observed in aphids collected  
330 from secondary host plants (controlled environment) is not apparent in aphids collected  
331 from the primary host cherry (field) (Fig. 4C), which affects the overall accuracy and ability  
332 to predict co-regulated genes. Possibly, the diversity of the *M. cerasi* population collected  
333 from primary hosts in the field underlies the observed break-down in co-regulation.

334

### 335 **Differential exon usage in *M. cerasi* populations**

336 We also found evidence for differential exon usage when comparing the different aphid  
337 transcript datasets. Overall, 263 genes show significant differential exon usage when  
338 comparing aphid datasets associated with the different primary (field) and secondary hosts  
339 (controlled environment) (Table S7). These 263 genes contain 2551 exons, of which 443  
340 show differential expression between aphid populations from primary hosts in the field  
341 versus populations from secondary hosts in a controlled environment. No significant GO  
342 annotation is associated with these 263 genes. One example of differential exon usage in  
343 *M. cerasi* is peroxidase gene Mca06436, which contains 5 exons, 2 of which are  
344 significantly more highly expressed in aphids collected from primary hosts in the field (Fig.  
345 5). This suggests that alternative splicing may be associated with the adaptation of the  
346 field population to the secondary hosts in a controlled environment.

347

### 348 **The majority of *M. cerasi* genes differentially expressed across different host** 349 **environments are duplicated**

350

351 Interestingly, the majority of genes differently expression across the three *M. cerasi*  
352 populations are duplicated (not single copy). For the genes upregulated in *M. cerasi* from  
353 cress (controlled environment) versus cherry (field) only 14% are single copy, which is  
354 significantly lower than the percentage of single copy genes in a randomly selected set of  
355 genes ( $p < 0.001$ , Mann-Whitney U test). Moreover, for all sets of differentially expressed  
356 genes, the differentially expressed genes were more likely to be duplicated when  
357 compared to a background random gene set ( $p < 0.001$ ) (Table S8). To assess the  
358 categories of gene duplication within the differential expressed gene sets, 100 iterations of  
359 randomly selecting 100 genes were conducted to obtain a background population. This  
360 yielded a mean and standard deviation for each duplication type from the parent gene



361 population (normally distributed). A probability calculator (Genstat) was used to determine  
362 how likely the observed counts were to occur at random. This showed that most of the  
363 duplicated differentially expressed genes were within the “dispersed duplication” category  
364 ( $p < 0.001$ ) and that there was no significant difference in the occurrence of tandem or  
365 proximal gene duplications ( $p > 0.05$ ) (Table S8). In contrast to predicted *M. cerasi*  
366 effectors, the differentially expressed genes identified in this study were not significantly  
367 further away from their neighbor in the 3' direction ( $p = 0.163$ , Mann-Whitney U Wilcoxon  
368 rank-sum test), or their 5' neighbor gene ( $p = 0.140$ , Mann-Whitney U Wilcoxon rank-sum  
369 test) when compared to an equal sized random population (Figure S3). Altogether our data  
370 suggest that *M. cerasi* multi-gene families may play an important role adaptation to host  
371 environments. This is in line with Mathers et al, (Mathers et al. 2017) who showed that  
372 duplicated genes play a role in adaptation of *M. persicae* to different secondary host  
373 species.

374

## 375 **Conclusion**

376

377 Aphids feature complex life cycles, which in some cases involve alternation between  
378 summer and winter host plant species. Here we show that, under controlled conditions, *M.*  
379 *cerasi* adaptation from primary to secondary host species does not readily occur, with only  
380 10-20% aphid survival. Comprehensive gene expression analyses of *M. cerasi* populations  
381 collected from primary hosts in the field and adapted to secondary hosts under controlled  
382 conditions revealed sets of detoxification genes that are differentially regulated and  
383 differential exon usage associated with a change in host environment. Many of the  
384 differentially expressed genes are members of multi-gene families. In contrast, we find  
385 only limited transcriptional plasticity to secondary host switching under controlled  
386 conditions.



## 387 **Methods**

388

### 389 **Aphid collection and adaptation**

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

### 405 **RNA sample preparation and sequencing**

406 Aphid samples were ground to a fine powder and total RNA was extracted using a plant  
407 RNA extraction kit (Sigma-Aldrich), following the manufacturer's instructions. We prepared  
408 three biological replicates for *M. cerasi* collected from each host. RNA quality was  
409 assessed using a Bioanalyzer (Agilent Technologies) and a Nanodrop (Thermo Scientific).  
410 RNA sequencing libraries were constructed with an insert size of 250bp according to the  
411 TruSeq RNA protocol (Illumina), and sequenced at the previous Genome Sequencing Unit

412 at the University of Dundee using Illumina-HiSeq 100bp paired end sequencing. All raw  
413 data are available under accession number PRJEB24338.

#### 414 **Quality control, RNAseq assembly and differential expression**

415 The raw reads were assessed for quality before and after trimming using FastQC  
416 (Andrews 2010). Raw reads were quality trimmed using Trimmomatic (Q22) (Bolger and  
417 Giorgi), then assembled using genome-guided Trinity (version r20140717) (Grabherr et al.  
418 2011). Transrate was run twice to filter out low supported transcripts (Smith-Unna et al.  
419 2015).

420

421 RNAseq assembly and annotation is available at DOI: 10.5281/zenodo.1254453. For  
422 differential gene expression, reads were mapped to the *Myzus cerasi* genome (Thorpe et  
423 al. 2018), per condition using STAR (Dobin et al. 2013). Gene counts were generated  
424 using Bedtools (Quinlan and Hall 2010). Differential gene expression analysis was  
425 performed using EdgeR (Robinson et al. 2010), using LOG fold change >2, FDR p<0.001  
426 threshold. GO enrichment analysis was performed using BLAST2GO (version 2.8,  
427 database September 2015) (Conesa et al. 2005) using FDR 0.05. The genome  
428 annotations were formatted using GenomeTools (Gremme et al. 2013) and subsequently  
429 HTseq (Anders et al. 2015) was used to quantify exon usage. Differential exon expression  
430 was performed using DEXSEQ FDR p<0.001 (Anders et al. 2012). Heatmaps were drawn  
431 as described in Thorpe et al. (Thorpe et al. 2018).

432

433 Gene duplication categories were used from Thorpe et al. (Thorpe et al. 2018). From these  
434 data, a random population was generated by running one hundred iterations on a set of  
435 100 randomly selected genes and their duplication types for subsequence statistical  
436 analyses. The script to generate random mean and standard deviation counts is available  
437 on Github ([https://github.com/peterthorpe5/Myzus.cerasi\\_hosts.methods](https://github.com/peterthorpe5/Myzus.cerasi_hosts.methods)). Statistical

438 analysis was performed using Probability Calculator in Genstat (17<sup>th</sup> edition). The obtained  
439 value from the gene set of interest (differentially express genes across aphid populations)  
440 was compared to the distribution of the random test set. Datasets identified as being  
441 significantly different from the random population did not significantly deviate from a  
442 normal distribution, thus the data was normally distributed. To assess the distances from  
443 one gene to the next, an equal sized population (1020) of random genes and their values  
444 for distance to their neighboring gene in a 3' and 5' direction was generated. The real  
445 value and random values were not normally distributed and were analyzed in Genstat (17<sup>th</sup>  
446 edition) using a non-parametric Mann-Whitney U Wilcoxon rank-sum test.

447

448 For SNP identification, RNAseq data was mapped back to the reference genome using  
449 STAR (2.5.1b) with --outSAMmapqUnique 255, to allow only unique mapping (Dobin et al.  
450 2013). SNPs were identified using Freebayes (Garrison and Marth 2012). VCFtools  
451 (0.1.15) (Danecek et al. 2011) was used on the resulting vcf files to identify SNPs per  
452 10Kb.

453

#### 454 **Validation of expression profiles by qRT-PCR**

455 Validation of the RNA-Seq experiment was completed with the Universal Probe Library  
456 (UPL) RT-qPCR system (Roche Diagnostics ©). RNA samples analyzed were those used  
457 for RNAseq analyses (aphid collections from location 1) as well as samples from aphids  
458 collected at a separate location (location 2) and adapted to secondary hosts (3 biological  
459 replicates) . For all experiments, aphid RNA was extracted using RNeasy Plant Mini Kit  
460 (Qiagen). RNA samples were DNase treated with Ambion® TURBO DNA-free™.  
461 SuperScript® III Reverse Transcriptase (Invitrogen) and random primers were used to  
462 prepare cDNA. Primers and probes were designed using the predicted genes sequences  
463 generated in the RNA-seq data analysis and the Assay Design Center from Roche,

464 selecting “Other organism” ([https://lifescience.roche.com/en\\_gb/brands/universal-probe-](https://lifescience.roche.com/en_gb/brands/universal-probe-library.html)  
465 [library.html](https://lifescience.roche.com/en_gb/brands/universal-probe-library.html)). Primers were computationally checked to assess if they would amplify one  
466 single product using Emboss PrimerSearch. Primers and probes were validated for  
467 efficiency (86-108 %) before gene expression quantification; five dilutions of threefolds for  
468 each primer pair-probe were used for generating the standard curve. The 1:10 dilution of  
469 cDNA was selected as optimal for RT-qPCR using the UPL system. Reactions were  
470 prepared using 25µl of total volume, 12.5µl of FastStart TaqMan Probe Master Mix  
471 (containing ROX reference dye), 0.25µl of gene-specific primers (0.2mM) and probes  
472 (0.1mM). Step-One thermocycler (Applied Biosystems by Life Technology©) was set up as  
473 follows: 10 min of denaturation at 95°C, followed by 40 cycles of 15 s at 94°C and 60 s at  
474 60°C. Relative expression was calculated with the method  $\Delta\text{Ct}$  (Delta Cycle threshold)  
475 with primer efficiency consideration. Three technical replicates were run per sample.  
476 Reference genes for normalization of the cycle threshold values were selected base on  
477 constant expression across different conditions in the RNA-Seq experiment. The reference  
478 genes were CDC42-Kinase (Mca01274), actin (Mca10020) and tubulin (Mca04511). The  
479 fold change calculations were done by  $\Delta\Delta\text{Ct}$  method (Delta Cycle threshold) and primer  
480 efficiency was taken into consideration.

481

## 482 **Acknowledgement**

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

486

487

488

489

490 **Funding**

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

494

495 **Availability of data and materials**

496

497 All data are available under accession numbers PRJEB24338. *Myzus cerasi* genome and  
498 annotation was downloaded from <http://bipaa.genouest.org/is/aphidbase/> and  
499 [DOI:10.5281/zenodo.1252934](https://doi.org/10.5281/zenodo.1252934). All custom python scripts used to analyse the data use  
500 Biopython (Cock et al. 2009), as well as details on how they were applied for data  
501 analyses are available on [https://github.com/peterthorpe5/Myzus.cerasi\\_hosts.methods](https://github.com/peterthorpe5/Myzus.cerasi_hosts.methods),  
502 and DOI: 10.5281/zenodo.1254453.

503

504 **Authors' contributions**

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

508

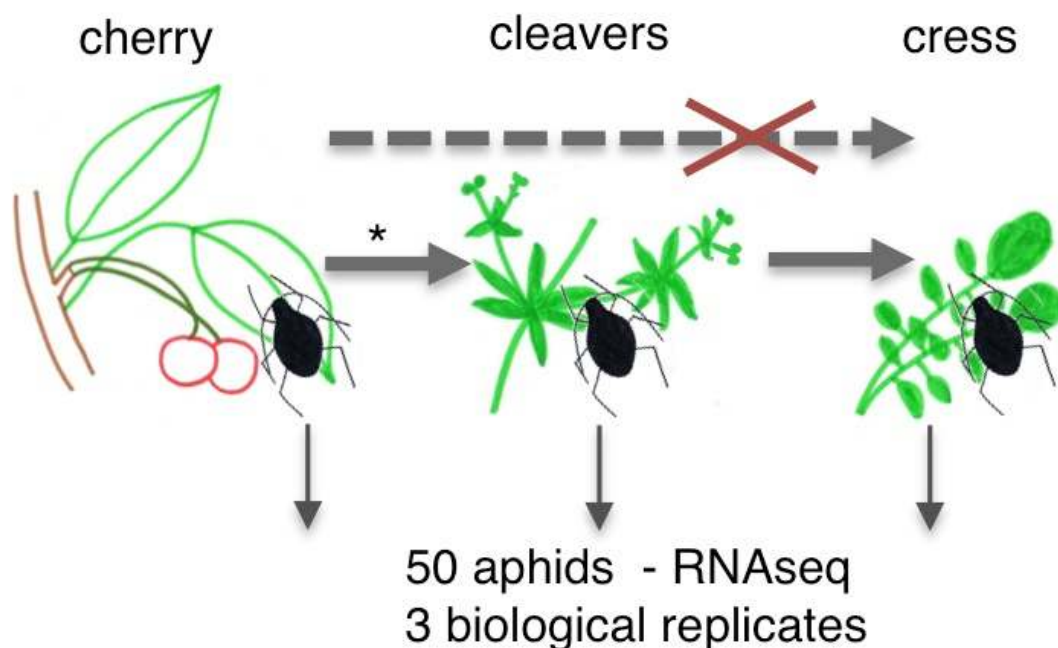
509 **Competing interests**

510 The authors declare that they have no competing interests.

511

512

## Figures



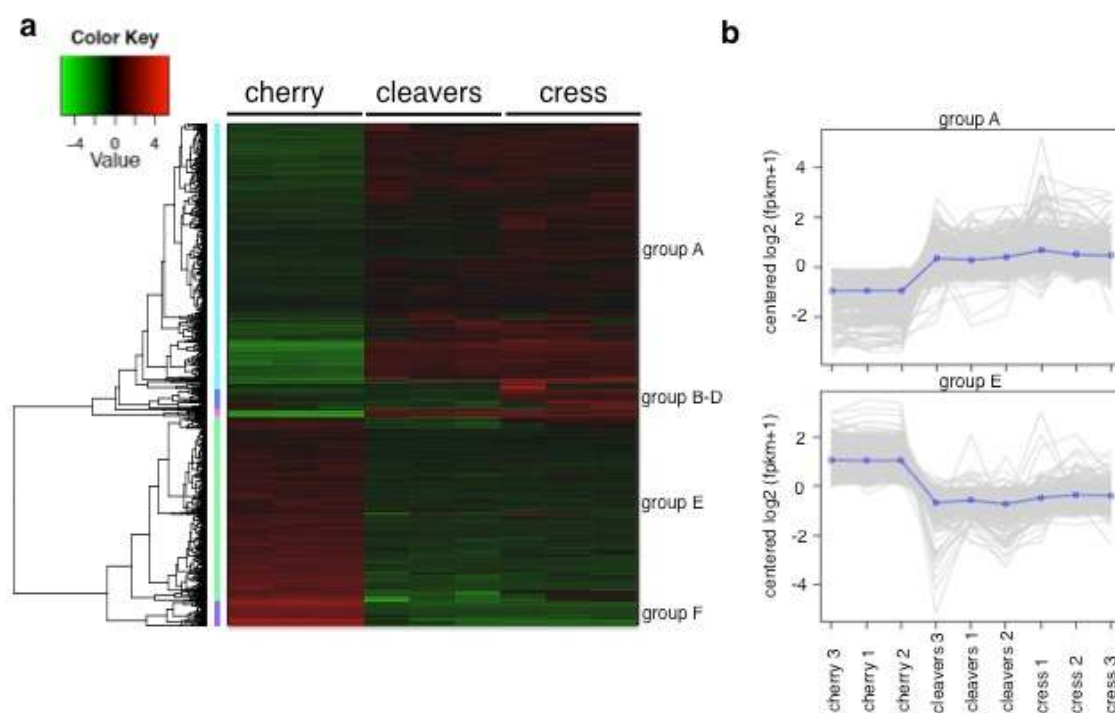
\* Number surviving (n=50 aphids)

Location 1: 8, 9 and 10 (3 reps)

Location 2: 5, 6 and 8 (3 reps)

### Fig. 1

Schematic overview of host environment adaptation experiments and aphid survival rates. *Myzus cerasi* aphids were collected from cherry trees at two separate field locations. None of the aphids collected from cherry in the field were 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) in a controlled environment. The host adaptation experiments were performed in 3 biological replicates.



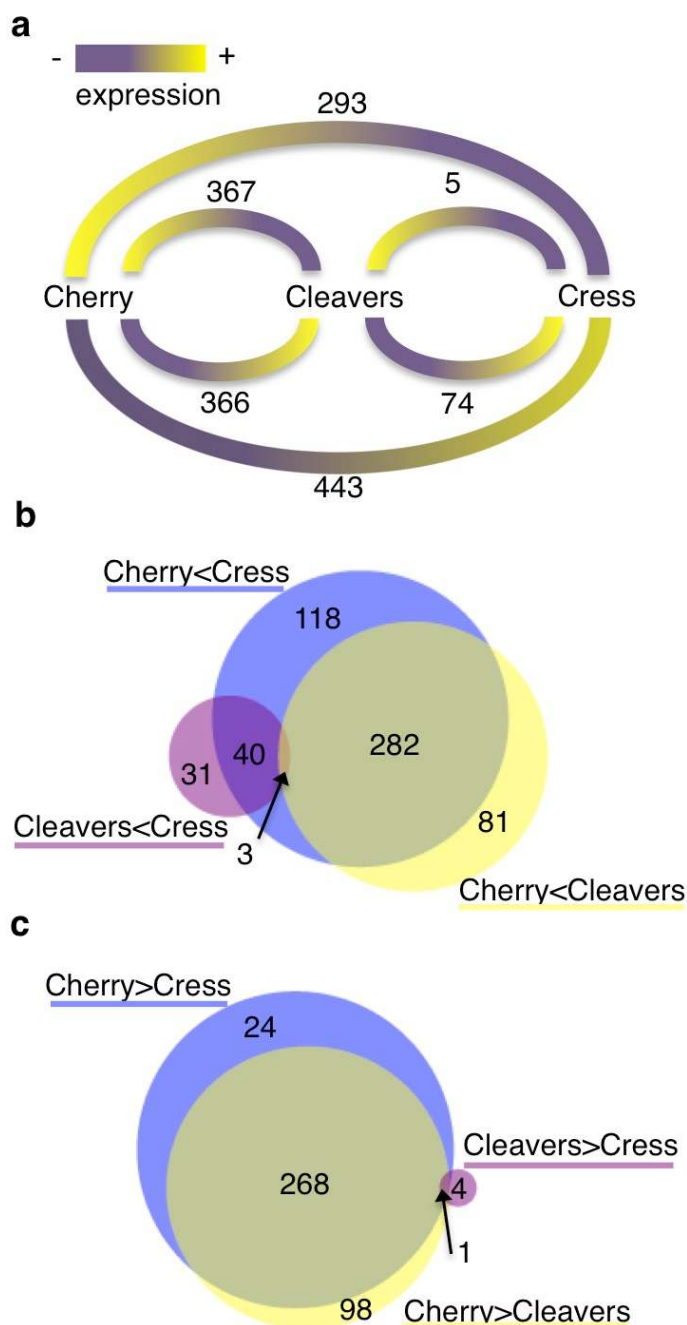
**Fig. 2**

Clustering of differentially expressed genes across *Myzus cerasi* populations from primary (field) and secondary (controlled environment) hosts.

(A) Cluster analyses of the 934 genes differentially expressed in *M. cerasi* populations from different host environments.

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





**Fig. 3**

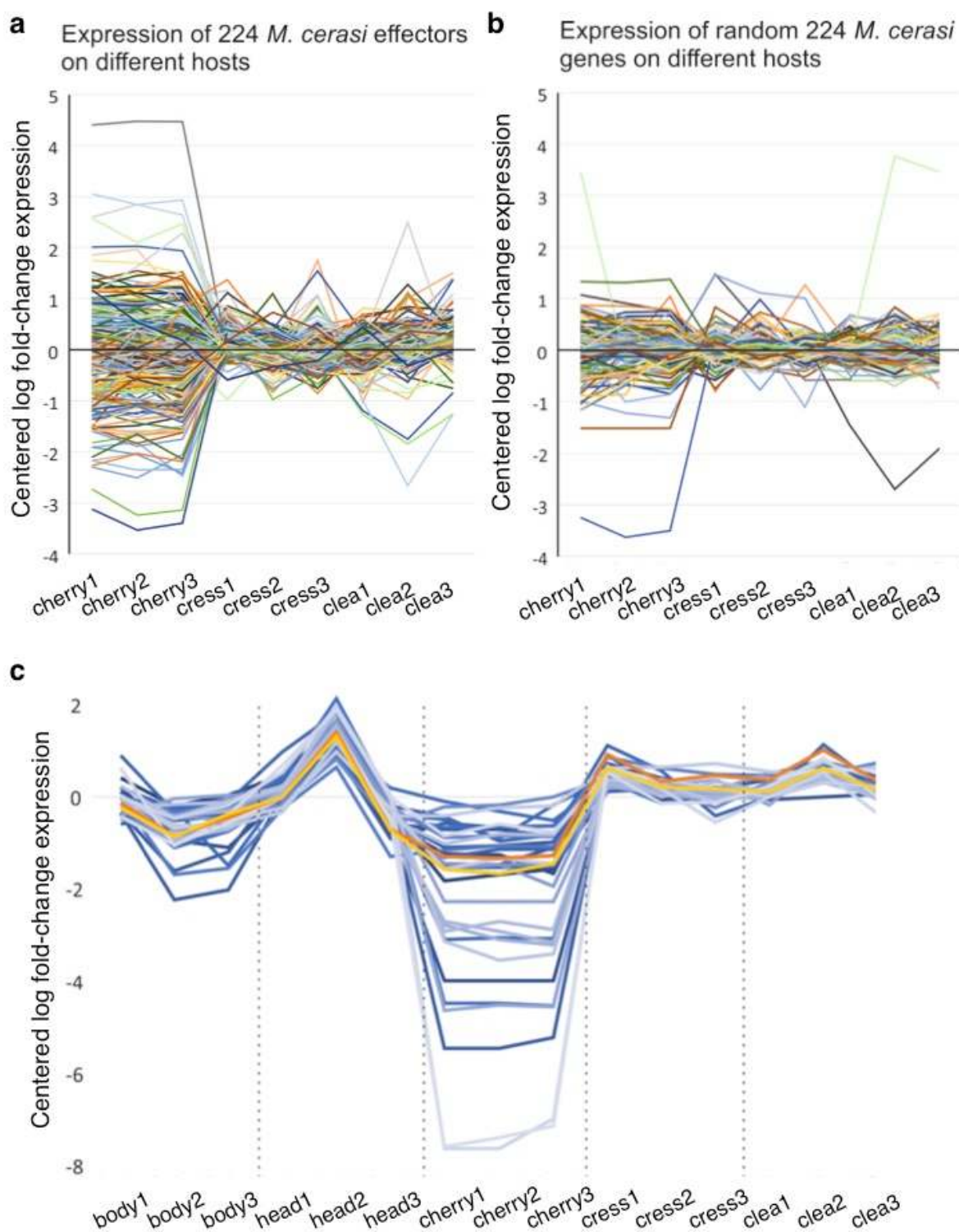
Differentially expressed genes in pairwise comparisons between the different *Myzus cerasi* populations.

(A) Numbers of genes for each pairwise comparison between aphids collected from the different host species, cherry (field), cleavers (controlled environment) and cress (controlled environment). 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 (field) compared to those collected from secondary hosts cleavers and cress (both in a controlled environment), 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 (field) compared to those collected from secondary hosts cleavers and cress (both in a controlled environment), and also higher expressed in aphids from cleavers than those from cress.



**Fig. 4**

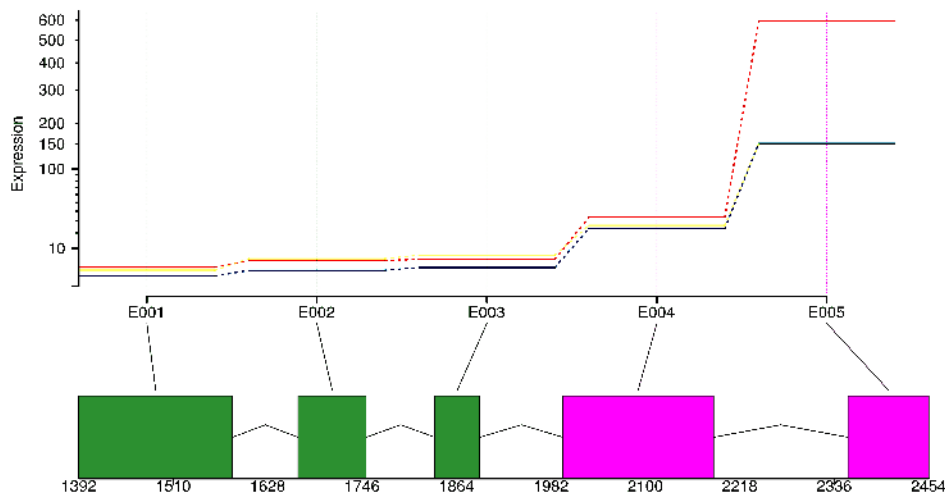
*Myzus cerasi* effector gene expression profiles across populations from different host environments.

(A) Mean centered log fold-change expression of 224 *M. cerasi* putative effectors across aphid populations from different host environments, including primary host cherry, under

field conditions, (cherry1-3), as well as cress (cress1-3) and cleavers (clea1-3), both in a controlled environment.

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

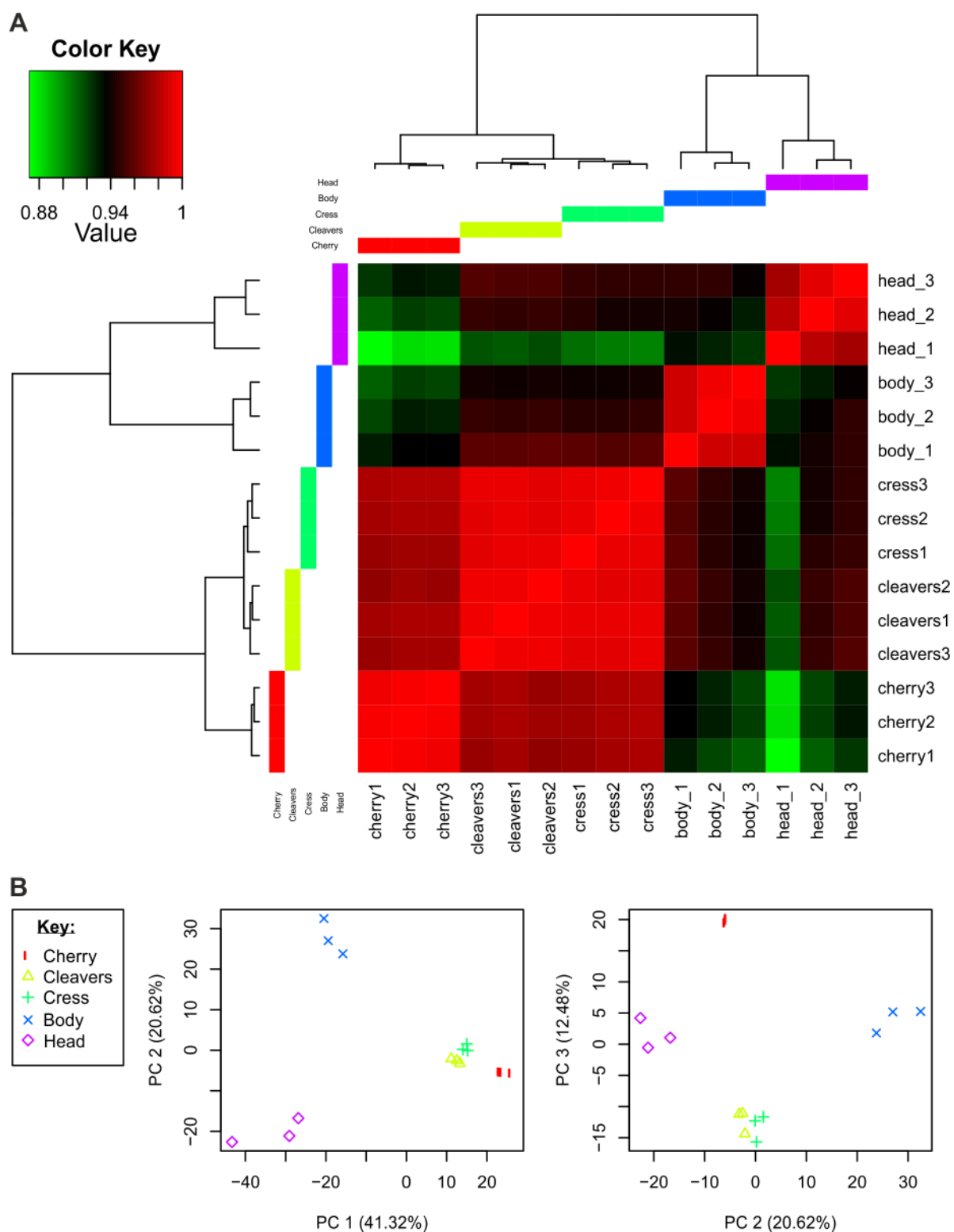
(C) Identification of all other genes in the *M. cerasi* genome that are co-regulated with the Mp1:Me10-like pair based on a >90% Pearson's correlation across different populations (blue, n = 35). Mp1-like is indicated in orange and Me10-like in yellow.



**Fig. 5**

Graphical representation of differential exon usage observed in gene Mca06436|peroxidase-like in the transcriptome of *Myzus cerasi* populations from different host environments, including primary host cherry, in the field, (red line), as well as cress (yellow line) and cleavers (blue line), both in a controlled environment. 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 in the field compared to those from cleavers or cress, both in a controlled environment.

## Supplementary data



**Fig. S1.** Transcriptome differences between *Myzus cerasi* populations from different host environments

Genome-wide analysis of *M. cerasi* transcriptional responses to interaction primary host cherry (field) or secondary hosts cress and cleavers (both in controlled environment), and comparison to previously published tissue-specific transcriptome of dissected heads and bodies (Thorpe et al. 2016).

(A) Clustering of transcriptional responses reveals that *M. cerasi* gene expression is different in populations from the different host environments 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.

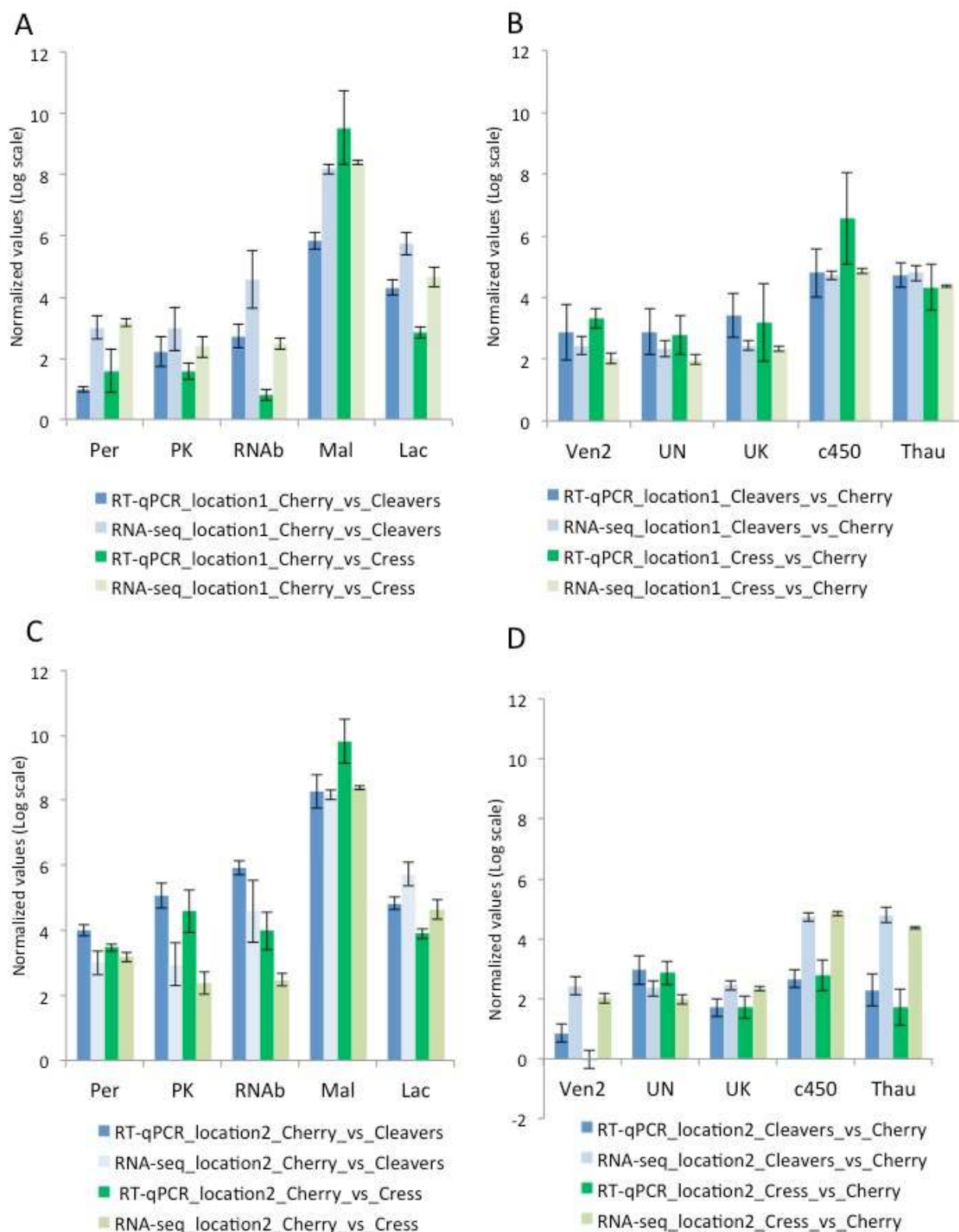
**Table S1.** List of 934 differentially *Myzus cerasi* genes across different host environments.

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

**Table S3.** List of significant GO-terms associated with genes differentially expressed across different *Myzus cerasi* host environments, corresponding to Fig. 3A.

**Table S4.** List of significant GO-terms associated with genes differentially expressed across different *Myzus cerasi* host environments, corresponding to Fig. 3B.

**Table S5.** List of significant GO-terms associated with genes differentially expressed across different *Myzus cerasi* host environments, corresponding to Fig. 3C.



**Fig.S2.** Validation of differential gene expression by qRT-PCR

(A) Genes up-regulated during cherry (field) versus cleavers/cress (controlled environment) interactions in the *Myzus cerasi* population collected from location 1.

(B) Genes up-regulated during the cleavers/cress (controlled environment) versus the cherry (field) interactions in the *M. cerasi* population collected from location 1.

(C) Genes up-regulated during cherry (field) versus cleavers/cress (controlled environment) interactions in the *M. cerasi* population collected from location 2.

(D) Genes up-regulated during the cleavers/cress (controlled environment) versus the cherry (field) interactions in the *M. cerasi* population collected from location 2.

The validated genes up-regulated during the cherry 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 cress/cleavers 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.

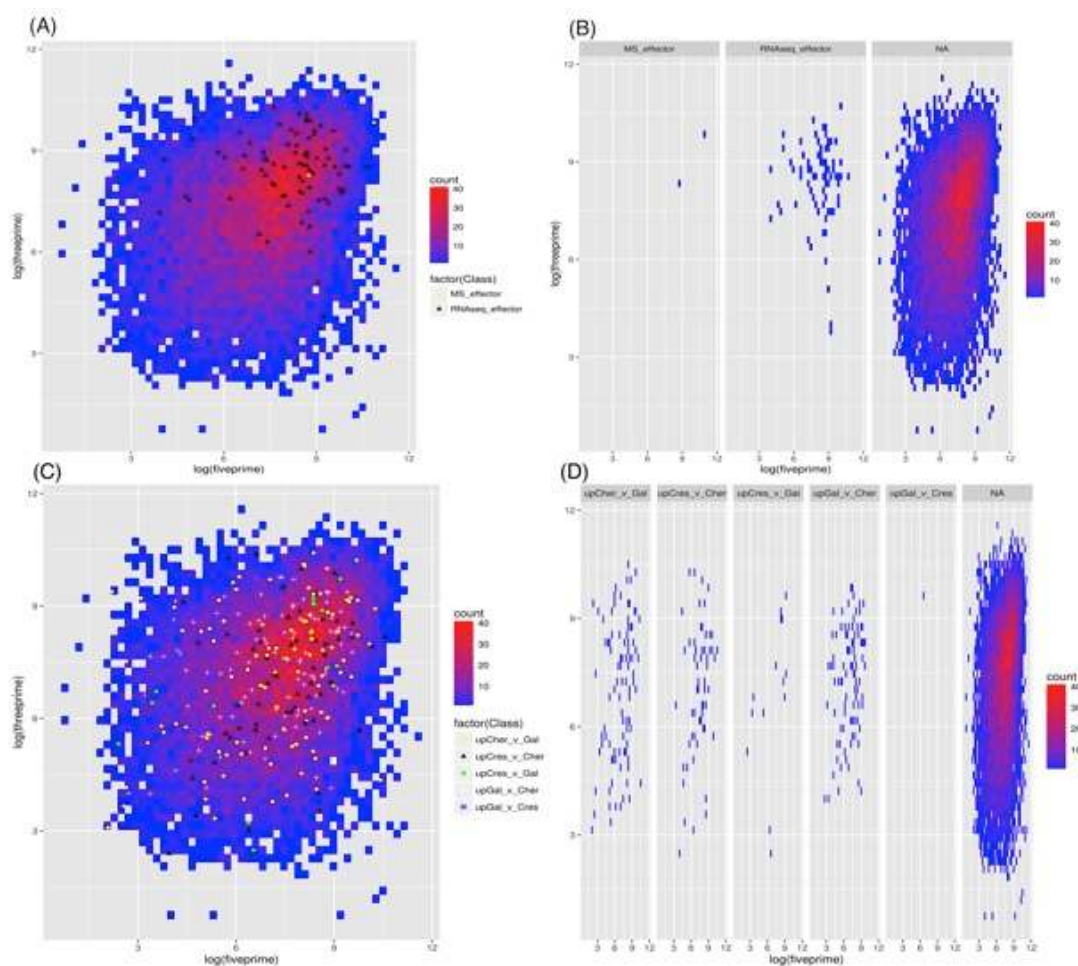
Error bars indicate standard error.

**Table S6.** List of 224 *Myzus cerasi* putative effectors and their expression levels across different host environments.

**Table S7.** List of *Myzus cerasi* genes showing differential exon usage across different host environments.

**Table S8.** Gene duplication types in *Myzus cerasi* genes differentially expressed across different host environments.





**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) Genic distance heat map for predicted effectors, which were significantly further away from their neighboring genes and thus in gene sparse regions (Thorpe et al. 2018). (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

- Anders, S, Pyl, PT and Huber, W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166-9.
- Anders, S, Reyes, A and Huber, W (2012) Detecting differential usage of exons from RNA-seq data. *Genome research* **22**: 2008-2017.
- Andrews, S (2010) FastQC: A quality control tool for high throughput sequence data. *Reference Source*.
- Barbagallo, S, Cocuzza, GEM, Cravedi, P and Shinkichi, K (2017) *IP Case Studies: Decidious Fruit Tree Aphids. Aphids as Crop Pests 2nd Edition*. CABI International, Wallingford.
- Blackman, R and Eastop, V (2000) Aphids on the World's Crops: An Identification and Information Guide. In: *Book Aphids on the World's Crops: An Identification and Information Guide* (Editor, ed.^eds.). Vol., pp. Wiley-Blackwell, City.
- Bolger, A and Giorgi, F Trimmomatic: A Flexible Read Trimming Tool for Illumina NGS Data. URL <http://www.usadellab.org/cms/index.php>.
- Cock, PJ, Antao, T, Chang, JT, Chapman, BA, Cox, CJ, Dalke, A, Friedberg, I, Hamelryck, T, Kauff, F and Wilczynski, B (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422-1423.
- Conesa, A, Götz, S, García-Gómez, JM, Terol, J, Talón, M and Robles, M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674-3676.
- Cui, N, Yang, PC, Guo, K, Kang, L and Cui, F (2017) Large-scale gene expression reveals different adaptations of *Hyalopterus persikonus* to winter and summer host plants. *Insect Sci* **24**: 431-442.
- Danecek, P, Auton, A, Abecasis, G, Albers, CA, Banks, E, DePristo, MA, Handsaker, RE, Lunter, G, Marth, GT, Sherry, ST, McVean, G, Durbin, R and Group, GPA (2011) The variant call format and VCFtools. *Bioinformatics* **27**: 2156-8.
- Dobin, A, Davis, CA, Schlesinger, F, Drenkow, J, Zaleski, C, Jha, S, Batut, P, Chaisson, M and Gingeras, TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.
- Garrison, E and Marth, G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv.1207.3907*.
- Grabherr, MG, Haas, BJ, Yassour, M, Levin, JZ, Thompson, DA, Amit, I, Adiconis, X, Fan, L, Raychowdhury, R and Zeng, Q (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**: 644-652.
- Gremme, G, Steinbiss, S and Kurtz, S (2013) GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on* **10**: 645-656.
- Jaquiéry, J, Stoeckel, S, Nouhaud, P, Mieuzet, L, Mahéo, F, Legeai, F, Bernard, N, Bonvoisin, A, Vitalis, R and Simon, JC (2012) Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex. *Mol Ecol* **21**: 5251-64.
- Kennedy, JS, Stroyan, H. L. G. (1959) Biology of Aphids. *Annu. Rev. Entomol.* **4**: 139-160.
- Mathers, TC, Chen, Y, Kaithakottil, G, Legeai, F, Mugford, ST, Baa-Puyoulet, P, Bretaudeau, A, Clavijo, B, Colella, S, Collin, O, Dalmay, T, Derrien, T, Feng, H, Gabaldón, T, Jordan, A, Julca, I, Kettles, GJ, Kowitwanich, K, Lavenier, D, Lenzi, P, Lopez-Gomollon, S, Loska, D, Mapleson, D, Maumus, F, Moxon, S, Price, DR, Sugio, A, van Munster, M, Uzest, M, Waite, D, Jander, G, Tagu, D, Wilson, AC, van Oosterhout, C, Swarbreck, D and Hogenhout, SA (2017) Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol* **18**: 27.

- Moran, NA (1988) The evolution of host-plant alternation in aphids: evidence for specialization as a dead end. *Am. Naturalist*: 681-706.
- Moran, NA and Jarvik, T (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**: 624-7.
- Mordvilko, AK (1928) The evolution of cycles and the evolution of heteroecy (migration) in plant lice *Annu. Mag. Nat. Hist.* **2**: 570-582.
- Quinlan, AR and Hall, IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.
- Robinson, MD, McCarthy, DJ and Smyth, GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.
- Smith-Unna, RD, Bournsnel, C, Patro, R, Hibberd, JM and Kelly, S (2015) TransRate: reference free quality assessment of de-novo transcriptome assemblies. *BioRxiv*: 021626.
- Thorpe, P, Cock, PJ and Bos, J (2016) Comparative transcriptomics and proteomics of three different aphid species identifies core and diverse effector sets. *BMC genomics* **17**: 1.
- Thorpe, P, Escudero-Martinez, C, Cock, P, Laetsch, D, Eves-van den Akker, S and Bos, J (2018) Shared transcriptional control and disparate gain and loss of aphid parasitism genes and loci acquired via horizontal gene transfer. *bioRxiv*.
- Williams, ISaD, A.F.G. (2007) *Life cycles and polymorphism. Aphids as crop pests*. CABI International, Wallingford. 69-87 pp.