Peer Community In Infections

How do multiple host plants and virus species challenge aphid molecular machinery?

Sebastien Massart *based on peer reviews by* ***Michelle Heck*** *and* ***Juan José Lopez Moya***

Quentin Chesnais, Victor Golyaev, Amandine Velt, Camille Rustenholz, Maxime Verdier, Véronique Brault, Mikhail M. Pooggin, Martin Drucker (2022) Transcriptome responses of the aphid vector *Myzus persicae* are shaped by identities of the host plant and the virus. bioRxiv, ver. 6, peer-reviewed and recommended by Peer Community in Infections.
https://doi.org/10.1101/2022.07.18.500449

Submitted: 19 July 2022, Recommended: 14 December 2022

Cite this recommendation as:
Massart, S. (2022) How do multiple host plants and virus species challenge aphid molecular machinery?. *Peer Community in Infections*, 100006. https://doi.org/10.24072/pci.infections.100006

Published: 14 December 2022

The impact of virus infection of a plant on an aphid's behaviour has been observed in many studies [1]. Indeed, virus infection can alter plant biochemistry through the emission of volatile organic compounds and plant tissue content modification. These alterations can further impact the interactions between plants and aphids. However, although it is a well-known phenomenon, very few studies have explored the consequences of plant virus infection on the gene expression of aphids to understand better the aphid's manipulation by the plant virus. In this context, the recommended study [2] reports a comprehensive transcriptomic analysis of the genes expressed by one aphid species, *Myzus persicae*, a vector of several plant viruses, when feeding on plants. Michelle Heck underlined how significant this study is for comprehending the molecular bases of aphid-vector manipulation by plant viruses (see below).

Interestingly, the study design has integrated several factors that might influence the gene expression of *M. persicae* when feeding on the plant. Indeed, the authors investigated the effect of two plant species (*Arabidopsis thaliana* and *Camelia sativa*) and two virus species [turnip yellows virus (TuYV) and cauliflower mosaic virus (CaMV)]. Noteworthy, the transmission mode of TuYV is circulative and persistent, while CaMV is transmitted by a semi-persistent non-circulative mode. As Juan José Lopez Moya mentioned, multiple comparisons allowed the identification of the different responses of aphids in front of different host plants infected or not by different viruses (see below). This publication is complementary to a previous publication from the same team focusing on plant transcriptome analysis [3].

Thanks to their experimental design, the authors identified genes commonly deregulated by both viruses and/or both plant species and deregulated genes by a single virus or a single plant. Figure 4 nicely summarizes the number of deregulated genes. A thorough discussion on the putative role of deregulated genes in different conditions gave a comprehensive follow-up of the results and their impact on the current knowledge of plant-virus-vector interactions.

This study has now opened the gate to promising research focusing on the functional validation of the identified genes while also narrowing the study from the body to the tissue level.

References:

1. Carr JP, Tungadi T, Donnelly R, Bravo-Cazar A, Rhee S-J, Watt LG, Mutuku JM, Wamonje FO, Murphy AM, Arinaitwe W, Pate AE, Cunniffe NJ, Gilligan CA (2020) Modelling and manipulation of aphid-mediated spread of non-persistently transmitted viruses. Virus Research, 277, 197845. https://doi.org/10.1016/j.virusres.2019.197845

2. Chesnais Q, Golyaev V, Velt A, Rustenholz C, Verdier M, Brault V, Pooggin MM, Drucker M (2022) Transcriptome responses of the aphid vector Myzus persicae are shaped by identities of the host plant and the virus. bioRxiv , 2022.07.18.500449, ver. 5 peer-reviewed and recommended by Peer Community in Infections. https://doi.org/10.1101/2022.07.18.500449

3. Chesnais Q, Golyaev V, Velt A, Rustenholz C, Brault V, Pooggin MM, Drucker M (2022) Comparative Plant Transcriptome Profiling of Arabidopsis thaliana Col-0 and Camelina sativa var. Celine Infested with Myzus persicae Aphids Acquiring Circulative and Noncirculative Viruses Reveals Virus- and Plant-Specific Alterations Relevant to Aphid Feeding Behavior and Transmission. Microbiology Spectrum, 10, e00136-22. https://doi.org/10.1128/spectrum.00136-22

Reviews

Evaluation round #2

DOI or URL of the preprint: https://doi.org/10.1101/2022.07.18.500449
Version of the preprint: 4

Authors' reply, 07 December 2022

Download author's reply
Download tracked changes file

Decision by Sebastien Massart, posted 06 December 2022, validated 06 December 2022

Mansucript improved and few minor revision before recommendation

Dear Authors,

Thank you for sending the revised version of the publication together with the point-by-point response to the comments and suggestions made by the two reviewers and the recommenders.

I have analysed them carfelly. Before recommending the publication, there are still few additional comments and suggestions that arose from reading your responses.

Kind regards,

Sébastien Massart

Ln21: Instead of “genetic basis”, a most appropriate term would be related to transcriptome or gene expression

We prefer to keep « genetic bases » because it is more general than « transcriptome changes”

COMMENT: Genetics is the study of heredity, and more broadly of genes/genomes. Here, the impact is on the gene transcription (maybe there might be an epigenetic effect although it is not the focus of this study). Sensu lato, the sentence is understandable, sensu stricto it might not be the most appropriate term (but I welcome any reference to publications using this term for transcriptomic if maintained)

Ln210: why 4 genes (how did you decide this number and not 10) ?

We believe that statistically there is no difference whether you test 4 or 10 out of thousands of genes.

COMMENT: indeed, I agree, it is always difficult to define the number of genes (and 10 was given as an example among others) and there is no “standard” recommendation. To back up and give strength to the selection of 4 genes, could you add one or two references of publication having confirmed the differential expression by RT-qPCR on similar number of genes ?

Ln 213-215: please give one (or several) references for it

We do not believe that a reference is necessary to explain our reasoning. PCR is an exponential amplification process, meaning at each ‘ct’ (amplification cycle) the number of molecules is doubled (ct0=1, ct1=2, ct2=4, ct3=8, ct4=16,…., ct10=1024, ct11=2048, ct12=4096,…), and consequently its sensitivity increases but its power of discrimination between two values (accuracy) decreases! This means that at low ct values the method can discriminate between small changes (for example between 3 = ct2 and 8 = ct3), but at higher ct values it cannot (for example between 2048, 2060 and 3000 = all ct11)!

COMMENT: there is a misunderstanding, the request for reference corresponded to cases where the RT-PCR did not confirmed the differential expression because of its properties. It was not linked to the exponential properties of PCR themselves.

Figure 1b: M2 has been excluded because it did not clustered with M1 and M3 but it is important to see where it cluster actually (is it close to the virus infected datasets ?) as M1 and M3 are quite divergent on PC2 also.

We added a PCA graph showing the three Camelina mock samples in a Supplementary figure.

COMMENT: thanks for the update of information but this new graph in SupMat is somehow redundant with Figure 2. I would suggest to add also M2 in the figure 2 (and delete the supplementary figure) so the reader can directly observe its position while stating clearly in the legend that is has further not be taken into account. For example, you can use light green for it. It completes the discussion of the results in the text further on.

Ln 259: only 8 categories are mentioned but do they compare to the 11 or 25 or any other for Arabidopsis (it is not clear for me if these 8 can also be considered as top 25 enriched or not), please clarify

The Top 25 GO analysis identified only 8 (for TuYV) and 3 (for CaMV) significantly enriched GOs in Camelina. None of the 8 GO specific for aphids on infected Camelina were found for infected Arabidopsis. The paragraph was rewritten and we hope it is clearer now: “A different picture was found for Myzus on virus-infected Camelina (Figure 2c). In the case of TuYV infection, only 8 categories (2 BP, 3 CC and 3 MF) were identified by GO Top 25 analysis as being significantly enriched. Three of them (Figure 2d) were also identified in aphids from CaMV-infected Camelina, but none of them in aphids from infected Arabidopsis. The enriched processes included chitin-related processes (chitin binding, MF; chitin metabolic processes, BP; structural constituent of cuticle, MF), transcription (transcription factor complex, CC), oxidation reduction (oxidoreductase activity, MF) and plasma membrane-related processes (homophilic cell adhesion via plasma membrane, BP; plasma membrane, CC; extracellular region, CC). Although none of these GOs figured among the Arabidopsis Top 25

GO, there were three GO categories (related to oxidation/reduction and plasma membrane processes) that were similar to GOs identified in aphids fed on Arabidopsis."

COMMENT: thanks, it clarifies indeed, could you simply state the meaning of the abbreviation (BP,CC...) when they appear for the first time in the paragraph

Ln 337: Why this homolog analysis is described/carried out here and not for upregulated genes ?

Actually, we used the same reasoning for up- and down-regulated genes, but for the upregulated ones we found no homologs. For more clarity, the explicative paragraph was moved to the beginning of the section and reads now like this: "We extracted in this analysis genes differentially up- or downregulated under all conditions. In the case of downregulated but not of upregulated genes, we found some genes homologs where one homolog was downregulated for one virus and another one for the other virus (Table 1). For example, we identified two potentially secreted homologous cathepsin B-like proteases (g8486 for aphids infesting TuYV-infected plants and g24532 for aphids infesting CaMV-infected plants). These homologs were included in the analysis. The rationale was that one specific host or infection condition might deregulate a specific gene but that the overall effect on plant aphid interactions might be the same or very similar for both genes (in this case the two cathepsin Bs might have a similar role as saliva effectors)."

COMMENT: reorganization of first sentence suggested: "This analysis was carried out on genes differentially up- or downregulated under all conditions. No homolog was identified for up-regulated genes. In the case of downregulated genes, we found some genes homologs where one homolog was downregulated for one virus and another one for the other virus (Table 1)."

Evaluation round #1

DOI or URL of the preprint: https://doi.org/10.1101/2022.07.18.500449
Version of the preprint: 3

Authors' reply, 30 November 2022

Download author's reply
Download tracked changes file

Decision by Sebastien Massart, posted 21 November 2022, validated 22 November 2022

Preprints merits a revision

Dear Authors,

After careful reading of your preprint by both reveiwers and myself, we would like to communicate that your preprint is a valuable scientific work that merits revisions before being recommended.

The comments and suggestions of the reviewers are in attachment to this message and I am sharing with you my own observations. Please check the use of past tense throughout the text.

Abstract :

- Ln 19 : what do you mean by performance ?
- Ln21: proposal of modification: ....vector manipulation caused by the virus. It is the organism that cause the manipulation through its adaptation. Please clarify the exact meaning of the sentence (also for line 39)
- Ln21: Instead of "genetic basis", a most appropriate term would be related to transcriptome or gene expression
- Ln 26: what is a "player" ? We can guess it is a gene but it can be stated more clearly

- Ln 26: proposal of change: “revealed a substantial proportion of commonly deregulated genes, revealing general players in plant-virus-aphid interactions” -> “revealed a substantial proportion of commonly deregulated genes, among which general players in plant-virus-aphid interactions”

Material and methods

- Overall: please check that all reagents have their proper provider mentioned (for example not stated in lines 167 and 168 for chloroform and isopropanol respectively…)
- There is no quality control of the extracted RNA before library preparation ?
- Is there a DNAse treatment?
- Ln 138: has this clone already been used in other published experiments ? If so, please add reference. If not, what is its origin (The Netherlands? Which year ?)
- Ln 140: precise what “form” means ?
- Ln 148: could you clarify the link between this protocol and the one described just before for Aphids ?
- Ln 165: reminding that aphids are the larvae 5 days old
- Ln 165: how were frozen the larvae before storage ? Liquid nitrogen or directly -80°C ?
- Ln 172: please refer explicitly to the kit used so tracing back the protocol is possible
- Ln 173: the 6 conditions are mentioned for the first time and it is not clear what they are ? Maybe stating them clearly in the section
- Ln 180 : you can clarify the recommended protocol as there is only missing the volume of Master Mix and of water
- Ln 182: if different temperatures have been used for different primers, please indicate the corresponding temperature in SupTable 2
- Ln 183: adding one (or several) publications where EF1 has been used
- Ln 183: indicating that the primers of the targeted genes are also listed in this table (in the table, please add also full name of the selected genes in addition to their internal reference)
- Bioinformatic analyses: could you specify which parameters have been used for each step (with only the software, there is not enough information to reproduce the analysis). Done for STAR (default parameters) but not the other ones ?
- Ln 188: number of reads can be trasnfered in results section
- Ln 195 to 199: giving percentages can be trasnfered in results section

Results

- Please clarify the wording used for reads (either 32 million of paired reads of 64 million of reads paired) throughout the text.
- Ln 210: how were the 4 genes selected ? Please state the selection criteria to understand why these genes and how they are relevant for validating the RNA-Seq data generated
- Ln210: to which process/pathway belong these 4 genes ?
- Ln210: why 4 genes (how did you decide this number and not 10 for example) ?
- Ln 213-215: please give one (or several) references for it
- Figure 1b: M2 has been excluded because it did not clustered with M1 and M3 but it is important to see where it cluster actually (is it close to the virus infected datasets ?) as M1 and M3 are quite divergent on PC2 also.
- Ln 229-232: there are twice more aphids DEG on Arabidopsis than on Camelina but were there similar number of sequenced genes in total (genes with mapped reads) ?
- Ln 245: “Aphid processes” might not be the most appropriate term -> metabolic pathway ? pathway ?
- Ln 259: only 8 categories are mentioned but do they compare to the 11 or 25 or any other for Arabidopsis (it is not clear for me if these 8 can also be considered as top 25 enriched or not), please clarify
- Ln269-270: precise that it is the feeding on virus-infected Arabidopsis that has impact on aphids gene expression (it is obvious but it is more precise)

- Ln275: the sentence is confusing as it seems that the modifications have occurred in plant host while you are analysing the aphid. Please check for proper wording here but also in other locations of the text.
- Ln 278: Arabidopsis OR camelina

Discussion

- As the discussion is very well structured between the different cases, a global figure of the results based on the same structure would be welcome in result section (Venn diagram ?). The idea is to be able to observe the number of genes that are DEG in each case, only depending on plant or on virus… So all the qualitative information provided in the discussion can be view quantitatively in a single graph
- Ln 293: might be more appropriate to avoid using "We". The same comment can be applied for the other parts of the document
- Ln 297: or by one host species
- Ln 304-305: might not be useful as it somehow repeats previous paragraph
- Ln 337: Why this homolog analysis is described/carried out here and not for upregulated genes ?
- Ln 398: "DEGs deregulated" it is a repetition, if they are DEG, they are deregulated
- Ln 463-465: "Since for Arabidopsis the total number of such aphid DEGs was 380, we applied a cut-off of logFC (fold changes) > 0.5 for upregulated genes and < -0.5 for downregulated genes to limit the number to 90 genes.". How did you manage the possibility that one of the DEG eliminated for Arabidopsis was an homolog of the 22 Camelina gene ? This is in link with comment on Ln 337: how did you manage homologues globally ?
- Ln 484: which kind of experiment would be needed ? Why only stating this sentence for this specific case and not for the other ones ? Maybe just let (or extend) last sentence in the conclusion (Ln 592-594)
- Ln 486 and 491: could you give the number of Myzus DEG ?
- Ln 505: TuYV, being circulative, … what do you mean by delicately ? Is it a usual term for this meaning ?

Minor comments:

- Ln 18 : document -> have documented
- Ln 30: name of genera in italics
- Ln 32: name of genus in italics
- Ln 33: the first sentence is too long and can be split, please use comma instead of hyphen
- Ln 45: delete "for example"
- Ln 47: "… of cells of the…" -> "… of cells in the …"
- Ln 62: … on the virus' mode…
- Ln 68: adding "through the hemolymph" at the end of the sentence
- Ln 100: most work -> most studies (and adapting the verb – without s)
- Ln 104-105: deleted "For example"
- Ln 108-109: replace hyphens by comma or point
- Ln 119: considering other verb than "accomplished" ?
- Ln 130-131: past tense for the verb
- Ln134: selected instead fo chose ?
- Ln 198: do not start a sentence with a number, sentence should be adapted or number written
- Ln 269: correct typo
- Ln 293: in the following section…
- Ln 315: the brackets should be for the reference only
- Ln 352: TuYV or CaMV
- Ln 360: …conditions is coding for….
- Ln 378: … and, possibly, ….
- Ln 510: … here only the subset of XX most strongly deregulated …

Reviewed by Juan José Lopez Moya, 30 September 2022

The manuscript by Chesnais and co-authors is a well written and interesting account of the findings after comparison of transcriptomes of Myzus persicae aphids that were feeding on two different plant species (Arabidopsis thaliana and Camellina sativa), and under two conditions: uninfected or infected with either TuYV or CaMV, two taxonomically very different viruses (polerovirus and caulimovirus). These two viruses diverged in mode of transmission (circulative and non-circulative), among many other differences, but the focus of the work deals with the relationship of virus-vector-plant in each pathosystem considered. The design of the experiment allows multiple comparisons to identify responses in the same insect vector species after feeding on the two different plant host species, and also the responses when the different viruses were or were not infecting the plant hosts. The study is rather complete, and provides novel information that can be valuable to understand the system, including the complexities of the ecology of two pathogenic viruses transmitted by the same insect vector to two different host plants.

Despite this general good impression, some components of the article might be modified in order to clarify the message and enhance the accuracy of the scientific contents. In particular the title could be more precise if some of the particular details of the study are mentioned, in its current form it is too general. Perhaps it would be advisable to mention the only aphid species being considered, and avoid to focus excesively in the ”mode of transmission” of the two viruses when many other important differences (starting with genome composition, distribution in plant tissues and so on) could have a potential impact in the host plants and indirectly on the infesting insects. Indeed, the broadness of the work is much better explained in the abstract, where the suggestion about alterations with possible conducive effects depending on the transmission mode are less strong that the title with the statement ”are shaped”. In other words, in my opinion the possible correlation between the observed responses and the mode of transmission is insufficiently supported by data coming from only two divergent viruses. Softening that idea instead of going too far will not reduce the impact of the results, but can help to make the general message more credible by readers.

Regarding the contents of the manuscript, the introduction is rather comprehensive, providing a clear picture of the antecedents about manipulative strategies, and the interest of exploring potential mechanisms of action. The opportunity and interest of the study is fully justified.

Also the methods are sufficiently well described.

Regarding the long Results and Discussion section, it is very well structured in sub-sections with descriptive titles. In a first sub-section the quality of the data is addressed, and in this regard it will be interesting to inform about the origin (in case it is possible) of the 15% and 19% of reads not matching with aphid genome (see p. 6, l. 207-208). Also the reason to eliminate one of the replicates corresponding to mock-Camellina might require a better explanation (p. 7, l. 224 to 227). Since the same experimental setup was used in a recent work by the same team (as declared on p. 15, l. 576; reference Chesnais et al 2022a, which appears incomplete in the bibliography but it is accessible in bioarxiv), it would be interesting to know if the discarded aphid sample corresponded also to a divergent (or not) plant sample: would it be possible to track the origin of each one of the biological replicates in both works, and derive information about the out layer? An additional explanation will be adequate here.

The rest of the sections are very well organized and structured to deliver a clear message. It is particularly valuable the discussion of the findings under different comparative perspectives, which allows a rational and well structured follow up of the results and their implications.

Only a few minor points might require attention to clarify some aspects of the discussions intercalated with the descriptions of results:

- For instance, the final sentence in p. 7, l. 242 appears to give similar or less importance to the virus species than to the transmission mode, which is again uncertain in my opinion; indeed, this is better expressed later (p. 8, l. 291), and even also providing examples (p. 10, l. 365).

- Another minor terminological aspect that deserves clarification to avoid confussion is the naming of pancreatic lipase-related like protein (p. 12, l. 447), considering the absence of pancreas in insects.

- The interesting discussion about calcium-binding products (p. 13, l. 516) requires to better identify if the suggested targets are in the plant or the insect. Do authors propose a delivery of the proteins from the vector to the plant, or another mechanism to exert influence in the host? This result is unclear with the actual wording of the sentences, and how it might affect the proposed hypotheses in the last part of the paragraph.

Tables could benefit by a further edition. In tables 1, 2, 3 and 4 separation of lanes are unclear, with more gene names (column 4) than functional categories (column 1). Apparently, sometimes more than one gene can be included in a given category, but it is not clear to which one. This can be solved by grouping categories and using horizontal separation lines.

Figures are fine, and perhaps the supplemental ones could be incorporated to the main text (no need to separate them, since they illustrate some aspects of the results).

To summarize, the manuscript has many strenghts derived from a good experimental design, generating results that can shed light into poorly addressed aspects of the virus-vector-host relationships. The work incorporates a good analysis of the data, and it is noteworthy the organization of the results into sections that really facilitates the reading.

Concerning weaknesses, no major flaws are found, and most of the aspects mentioned above that need attention could be easily solved in a revised version.

Reviewed by Michelle Heck, 09 November 2022

The authors present very nice work looking at the transcriptomics response of aphids to encountering plant infection with two different viruses that have unique transmission modes on two different host plants. The study is a significant contribution to the field because very little is known about the molecular mechanisms of aphid-vector manipulation by plant viruses. An understanding of the molecular pathways involved in aphid-virus interaction are necessary for the development of new biological control tactics that block or reduce transmission within a crop. Globally, Myzus persicae is one of the most prolific vectors for plant viruses, transmitting more than 100 known plant viruses in a range of economically important crops.

I have a few comments for the authors to consider in their revised version.

1. Research has shown that host switching changes the protein and transcript expression levels and sub-cellular localization of proteins involved in virus transmission (Pinheiro and Wilson et al. 2017). The authors should discuss how rearing the aphids on Chinese cabbage and then transferring to new hosts might result in transcriptomic changes in dependent of viral infection.

2. Unless I missed it, none of the observed transcriptome changes were more than 2-fold different, up or down. This is extremely interesting to see such a muted effect on the aphid transcriptome! I have a few thoughts on this.

A. Using whole insects (in contrast to excised guts or salivary glands) masks larger gene expression changes that occur at the tissue specific level where the viruses would normally directly interact with aphid cells. This first hypothesis could be argued against to some extent in part because these viruses do not replicate in aphid cells and because a prior study using PEMV and the pea aphid showed a similar muted effect on gene expression changes in the aphid gut (Brault et al. 2010). But on the contrary, it is definitely possible to have under-estimated the changes you observed because whole insects were used in this study and not excised tissues. A study from our lab on the Asian citrus psyllid, Diaphorina citri, showed that the transcriptomic profiles of each insect organ was unique and even more different than the effect of the phloem colonizing citrus greening pathogen (Mann et al. 2022 https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giac035/6575386). Thus, when we pool insects, we lose the tissue-level resolution for individual genes and this is undoubtedly going to have a major impact on data interpretation.

B. The second hypothesis for the observed muted effect on the aphid transcriptome could be that interactions between aphids and infected plants are occurring at the metabolite or protein level, and not transcriptional. Their finding supports the idea that physiological changes in the aphid in response to viral infection are plastic and transcriptional regulation is not a major component to the molecular mechanisms of how aphids respond to viral infection. Both of these ideas could party explain the results shown here. How might these possibilities influence the design of follow-up functional studies? The very last paragraph of the paper nicely introduces the next step in the research and this might be a good place to include some of this speculation/discussion.

3. More details are required in the methods section for the parameters used for DEG analysis, specifically with DESeq2 and SMARTools.

4. I'm wondering if the authors might reconsider their selection of the word "deregulated" to describe differentially abundant transcripts. Aphids encounter virus infected plants in the landscape as a normal part of their ecology. De-regulated would not be the word I would selected to describe transcripts whose abundance changes in response to viral infection. Differentially expressed (but we actually don't know if these are real mRNA expression differences or due to mRNA stability, silencing and so on), differentially abundant, or change in abundance are options the team might consider.

In conclusion, this is a wonderful, insightful, carefully written paper. The authors go to great lengths to discuss their transcriptomic changes in the context of what is known about these genes from other aphid studies and other insects. The changes I suggest are relatively minor and I am looking forward to being able to share this work with my students.