Round #2

Please note that the most recent revised version on biorXiv is v5 (The PCI site seems to think that v5 is the previous one).

Please find our responses in black as in the previous revision round.

by Sebastien Massart, 06 Dec 2022 11:27
Manuscript: **https://doi.org/10.1101/2022.07.18.500449** version 4

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 carefully. 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)

> You are write, “*genetic bases*” is not the appropriate term. We replaced the sentence in question with the following sentence: “*However, until now, the **gene expression changes** correlating with these effects and indicative of modified vector pathways and mechanisms are mostly unknown.*”

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 ?

> While many publications validate a varying number of gene expressions by RT-qPCR, there are also many instances where no such effort was made at all. For example and citing only work on aphid transcriptomes, Xia et al. 2014 validated 14 genes by RT-qPCR (https://doi.org/10.1186/1471-2164-15-1050). Like in the present study, four genes were validated by Liu et al. 2012 (https://doi.org/10.1371/journal.pone.0045161), but no genes were validated by Boulain et al. 2019 (https://doi.org/10.3389/fpls.2019.01301), Matsuda et al. 2020 (https://doi.org/10.1016/j.cbd.2020.100740) or by Parker et al. 2021 (https://doi.org/10.1111/evo.14174).
> We changed the text in the manuscript and cite the paper by Liu et al.: "*Exemplarily, a similar trend of gene deregulation was confirmed by RT-qPCR for four Myzus genes with different levels of deregulation and expression, but the same trend in both infection conditions. We screened only four genes as in a previous work on aphid transcriptomics (Liu et al., 2012). Three genes showed the same trend of downregulation in RNA-seq and RT-qPCR experiments, while the forth (g15329) was found to be upregulated in all RNA-seq and RT-qPCR experiments, except for RT-qPCR on TuYV-infected plants (Supplementary Figure S1). The discrepancy in the results for g15329 expression was likely due its weak expression changes that in general are difficult to detect by RT-qPCR because of the exponential amplification kinetics of this technique.*"

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.

> We had the same result (discrepancy between RNAseq and RT-qPCR for genes with low expression changes) in a plant transcriptomic study (Chesnais et al. 2022 https://doi.org/10.1128/spectrum.00136-22). We cite it now in the same paragraph as above: "*Exemplarily, a similar trend of gene deregulation was confirmed by RT-qPCR for four Myzus genes with different levels of deregulation and expression, but the same trend in both infection conditions. We screened only four genes as in a previous work on aphid transcriptomics (Liu et al., 2012). Three genes showed the same trend of downregulation in RNA-seq and RT-qPCR experiments, while the forth (g15329) was found to be upregulated in all RNA-seq and RT-qPCR experiments, except for RT-qPCR on TuYV-infected plants (Supplementary Figure S1). The discrepancy in the results for g15329 expression was likely due its weak expression changes that in general are difficult to detect by RT-qPCR because of the exponential amplification kinetics of this technique.* ***We observed the same phenomenon in a previous validation experiment (Chesnais et al., 2022a)***."

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.

> We changed the figure. This changes slightly the axes, but it is acceptable.

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

> The basic categories are now spelled out in the text: "*A different picture was found for Myzus on virus-infected Camelina (Figure 2c). In the case of TuYV infection, only 8 categories [2 in biological processes (BP), 3 in cellular components (CC) and 3 in molecular functions (MF)] were identified by GO Top 25 analysis as being significantly enriched.*"

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)."

Thanks for the suggestion. It really makes the point clear!