Dear editor and reviewers,

We thank you very much for taking the time to consider our manuscript for recommendation by PCI in Ecology and for the constructive review process. We have changed the manuscript according to the reviewers' comments and underlined in yellow the changes in the manuscript. Please find below more detailed answers about the referees' comments and explanations on how we addressed them. We have tried to stay consistent in the wording that describes apple and aphid material, we have also added information about some limits of our experimental design and corrected our statistical models. We hope that the revised version will be satisfactory for recommendation by PCI in Ecology.

Thanks a lot,

Sincerely,

Sergio Gabriel Olvera Vazquez, on behalf of all co-authors.

##

Reviewer 1 - *Alex Stemmelen* ##################

##

General comments

The authors should be careful when referring to an effect of the climate, when they actually test for an effect of the site. If aphids from Spain show higher fitness in the Spain's common garden, that can be a sign of adaptation to climate but also soil type, predation by natural enemies, etc.

> We agree with the reviewer's comment and we have added a sentence in the manuscript explaining that the test of local adaptation will not be restricted to bioclimatic parameters of the common gardens, in lines 77 to 79.

More broadly, the term climate is not mentioned anymore in the models written later, while it is present in the title of the manuscript. Maybe climate will be accounted for through the use of local climatic variables as covariates, but this information was missing.

> We agreed with the editor and the reviewers. We have modified our title "Experimental test for local adaptation of the rosy apple aphid (*Dysaphis plantaginea*) during the recent rapid colonization of cultivated apple (*Malus domestica*) in Europe."**.** We have also explained in lines 77 to 83 that we will test for the site effect, which includes biotic/abiotic effects. We have also added information in the statistical model section lines 702 to 706. We will record temperature with a Near-infrared Spectroscopy NIRS to estimate the microclimate of each leaf before and after aphid infestation. We will first run a statistical model per site, removing the fixed site effect, but adding as an effect the temperature per leaf before (one fixed effect) and after (another effect) infestations. If there is an effect of temperature, we will then integrate the temperature as a variance-covariance matrix of the difference in temperature among each apple tree before and after the infestations on a random site effect, Table 3). This will be a way to consider the effect of climate on aphid fitness. Note that we prefer to test the effect of climate first within the site, as adding this effect directly in the full model will add complexity to the already-complex equations.

It could be helpful to add some clarification/definition of confusing terms such as genotype/variety/cultivar, mentioned several times in the protocol, in order not to confuse the reader.

> We have followed the reviewer's suggestions and we have defined genotype as all the apple material that we will use in our experiment.

Authors presents in the question 1 to 3 a null hypothesis H0 and multiple alternate hypotheses. This is surely correct and of course it is appreciable that the authors have anticipated what could be the response(s) of their studied system. However, usually only one alternate hypothesis H1 is formulated, based on current knowledge of aphid biology and aphid-host interaction. Please note that this remark

is not a criticism of the planned work, just a general comment on what could be the last paragraph of the introduction of the paper to come.

-> We thank the reviewer for this useful comment, and we will take it into account when writing the introduction of the paper. We have modified the null (H0) and alternative hypotheses (H1), lines 89 to 169.

Question 4 mentioned wild and cultivated apple trees. Do we have information on how the domestication of apple tree could have altered resistance to aphid infestation? For example, plant domestication can lead to a decrease in constitutive and induced chemical defences, which could increase the growth and performance of aphid on cultivated apple tree (see Moreira et al. 2018, Scientific Reports for example). Such information could help refine or precise the hypotheses made in the question 4.

> Rootstock, hybridization, and introgression have been reported to impact several apple agronomic traits, including pest resistance (Rom et al., 1990; Cornille et al., 2019). However, a specific study on the impact of domestication on apple tree resistance to aphid infestation is lacking. However, we have suggested a plausible hypothesis according to the known evolutionary histories of the European wild apple and the cultivated apple, lines 142 to 155.

Rosy aphid colonies are reared and maintained on Jonagold cultivar before being synchronized on Golden Delicious cultivar, and before being transferred onto several cultivars in the different common gardens. While the synchronization steps is defined as being done with parthenogenic female - thus preventing adaptation to the Golden Delicious cultivar – could it be possible that the maintenance phase will lead to an adaptation of the aphid genotype to Jonagold cultivar? If so, could we expect an increased fitness in apple genotypes closer to Jonagold cultivar? In other word, could we anticipate that the results could be different should the synchronization phase be done on any other cultivar?

>We thank the referee for this interesting comment. Genetic adaptation is not possible as we only maintain asexually the colonies on the Jonagold cultivars, as we will do for the synchronization step. We have defined that information in lines 303 to 305. Acclimatization is a possibility, but we choose the apple genotype for aphid rearing (this case the Jonagold) to be different from any cultivars that are in the infestation experiment of the Spring of 2021. It will therefore avoid any aphid acclimatization to a specific apple genotype used in the cross-infestation experiment. Finally, we have the information on the genetic relatedness among the cultivars used in the study and will be able to test this effect on our fitness measurements. This information has been added lines 362-372.

The protocol mentioned the acquisition of several types of data on the local climate *(Temperature, humidity)* and apple tree physiology *(polyphenol content, chlorophyll)* but no information on how these data will be used is reported. While these data could probably be used *a posteriori* in the study (or in a companion paper), it is unclear how they will help answer the original question about local adaptation. Maybe it is not necessary to include them in the protocol if the authors don't plan on using them latter in there analyses.

>We totally agree with the suggestion of the reviewer. We have deleted the paragraphs about the measurements of the ecophysiological apple traits. However, we will record the temperature and humidity during the experiment with a local meteorological station available next to each common garden. We will also record the temperature of each leaf, before, during, and after the infestation with Near-infrared Spectroscopy (NIRS). We have added this information on lines 528-531. These temperature records will be used in statistical analyses.

Cross-infestation will begin in Spring 2021. Does the start of the experiment will be similar in each location? Climatic condition could be different between Spain and Belgium for example and could lead to different fitness between location. Will this be taken into account in the modelling approach?

>We will perform the infestation at stage E2 of the plant, in all the common gardens, so we will rather follow the apple phenology than running the experiments at the same time at each site. We specified the phenological stage of the apple genotype for infestation, see lines 348 to 350 and in Figure 4. We will also include the time of infestation (day and hour) in the model as a block effect (see Equation 1, line 666, Equation 2, line 711, and Equation 3, line 725).

Data analyses

Authors planned to make repeated measurements of aphid fitness in each apple tree in each common garden, as explain in the Fig. 4 below. If the authors conduct repeated measurements by using nine aphid genotypes on each tree, they have to make sure to include a parameter assessing the variability in aphid fitness attributable to the apple tree identity. The equation would then look like this:

$$Whijklmntt2z = \mu W + aphid\_originh + apple\_origini + sitej + sitej(blockk) + Ghl(leafm(Gpn)) + day\_of\_infestationt + hour\_of\_infestationt2 + Tree\_IDu + aphid\_originh*sitej + aphid\_originh*apple\_origini + aphid\_originh*apple\_origini*sitej + \varepsilon hijklmntt2z.$$

With Tree_IDu being random and a single ID given to each apple tree across all common garden, to assess the part of variability attributable to the apple tree identity.

>Thanks a lot for this remark, we have added this random effect in the equations (**+ ω_o)**. Note that we have also added a leaf effect (**+ μ_m)** in the equations, *i.e.*, whether the aphid genotype is infested in the upper, medium, or lower part of the tree, following another reviewer's remarks.

Representation of rosy apple aphid infestation on the different apple genotypes. a) Nine aphid genotypes from different origins (three from Belgium, three from France, and three from Spain) will be used to infest a cultivated apple tree. b) A single synchronized adult female aphid from each of the nine aphid genotypes will be randomly infested on nine leaves of a tree. c) Each infestation will be protected with a cellophane bag and sealed with a stapler. BE = Belgium, FR=France, SP=Spain.
Following this issue, it is said that "aphid *genotypen* is nested within *leaf IDm,* and *leaf IDm* is nested within apple genotype *Ghl,* and they were added to the models as random-effect terms."
However, each apple tree will be infested with 9 aphid genotypes on 9 different leaves or leaf clusters but there won't be more than one of each aphid genotype per apple tree, hence wouldn't the random

effect *leaf IDm* confounded by the aphid *genotypen* effect? You probably don't need to include the random *leaf IDm* effect, as it would lead to potential overparametrization of your model.

>We plan to infest each apple genotype with nine different aphid genotypes (three different aphid genotypes from each of three different countries) on nine different leaves. Then, following your appreciable remark, we have divided the position of the infested leaves into three levels: upper, medium, and lower level (**+ μ_m)**. The aphid genotype will be nested in the leaf level and these nested in the apple genotype. We will also consider the "universal" effect of the leaf position on variability in aphid fitness in our equations. We have updated our equations in the "statistical model" section and Table 3.

During the cross-infestation experiment and if only one block is infected each day, caution should be taken that the random effect *day_of_infestationt* isn't confounded with the random *blockk* effect.
-> Thanks for this interesting remark. However, it will be technically impossible to infest one block per day, so we should not have a confounding effect of the *day_of_infestation*$_t$ and the *block* effect.

Data will be transformed to fit linear model assumption. We suggest trying Generalized Linear Mixed-Models with suitable family *(Poisson or Gaussian for example)*, depending on the data distribution, to avoid unnecessary transformation of the data. It is possible that some aphid colonies will not establish, such that aphid count can be zero (i.e., less than the originally n = 1 aphid per leaf). It is therefore possible that the dataset will contain a large amount of zeros, which may be a problem. Maybe this won't be the case, but if so it would be nice to know how the authors will deal with the data distribution. It is for instance possible to use zero-inflated poisson distrubution in GLM(M)s. Alternatively, should the number of zeros be large, one option is to use a two-steps modeling approach, starting with a binomial response (1 = presence of an aphid colony ; 0 = No colony), followed by the analysis of aphid count data limited to the subset of non-zero outcomes.

>We fully agree with the reviewer. We have added these alternative analyses in the text in lines 775 to 786.

##########

Reviewer 2 *anonymous reviewer*#######

##########

The experimental test addresses the question of the adaptation of an insect plant-parasite to a local host, both domestic and wild, and to its local climate. The experimental design aims to answer five questions related to the adaptation of the parasitic apple aphid Dysaphis plantaginea on apple tree species Malus sp. As a large-scale study with replication at three garden orchard sites and many individual replicates, this multi-genotypic cross-infection experiment has very promising potential. The whole set-up and preparation of the material for the experimental test scheduled for March 2021 already began in 2019 with the growth of wild or domesticated apple scions and the breeding of aphids. The plants were transferred in early 2020 to each site. Global planting was carried out with a randomization of the plots to avoid any plot combination effect as much as possible and will

be included in the model. I believe the study was well thought and measured at a large scale in order to optimize the repetability and homogeneity bewteen experimental sites, apple trees and aphid treatments and preparation. As it is an ambitious project, the work load is very important and the inoculating experiment will thus be conducted over 18 days. As this may as well cause important variability, the days will be integrated in the statistical model as a random effect.
I have though some minor comments / concerns about the experimental analyse.

Fitness variables. I believe authors should be more accurate about the fitness variable aquisition and measure. Authors specify: "We will estimate rosy apple aphid fitness as the growth rate of the colony (cumulative number of nymphs produced per surviving females produced on the infested plant; Warneys et al., 2018), and the insect life cycle (aphid stages (L1 to L5), apterous adults, nymphs, and winged forms (Angeli and Simoni, 2006))." First, it is not completely clear to me at what time points the fitness measurment is occuring or in other words are you following the cumulative number every days since inoculation and for how long?

> We, therefore, estimate the duration to recover the aphid colonies after the infestation on each apple tree to 12 days, but it will be adapted depending on the reference Golden Delicious genotype. Indeed, in the literature, the mean of the time for a female to start producing larvae that will become a full-grown female is 12 days (Warneys et al., 2018). This timing may change depending on the local conditions, so that we will use as reference the susceptible Golden Delicious genotype: once we will observe more than two big females on the GD genotype (meaning that the two funder females have produced full-grown females), we will recover the colony on other cultivars.

Once we will have recovered the colony in a tube filled with ethanol, we will then count the number of aphids, and if possible, the different stages (*i.e.*, aphid larvae (L1 to L5), apterous adults, and winged forms (Angeli and Simoni, 2006)). This will be done by scaling the individuals into three categories: big (apterous females), small (larvae), winged) to include this information in our statistical models.

To estimate the fitness, we will subtract the number of aphids at the end of the infestation minus the number of aphids that were infested at the beginning of the infestation, and divide this number with the total number of days of infestation.

We have added this information in lines 524 to 526.

Second, it is not described how experimentators will combine growth number (nymph/viable mother) and cycle development (aphid stages). They should indicate if they will be using both data independently for modelling and what they expect for their distribution. One particular question is :how do you account for the 5 aphid stages in the model? Do you model for 5 dependant variables? Or do you run multivariate models?

>For now, we plan to analyze them separately. We will run our model using different absolute fitness values, first the colony growth rate, later the aphid sizes for the three categories of aphid (big, little, winged). We have added this information in lines 700 to 703.

Apple trees description. Since Chen at al is not available yet, authors should precise how the choice of wild apple tree was done for this approach. It seems to me according to authors' previous data (Cornille et al. 2015) that Malus sylvestris shows about so five distinct European genetic clusters among which only three are present in Spain/France/Belgium. Are all the selected wild genotype among those three groups?

>The choice of the will apple genotypes was based on previous studies that showed that Spanish and Belgian wild apples belonged to genetically differentiated populations in Europe (Cornille et al. 2013, 2015). Note that the 28 apple genotypes used in this experiment have been genetically characterized using 13 microsatellite markers and by Illumina sequencing data (data not shown) which will be analyzed in 2020 and 2021. We have tried to be clearer in lines 269-272.

Also, while it is not the main focus of the study, is there any other easy accessible genomic data that could be utilised regarding resistance such as known R genes / QTLs presence on those tree accounting for parasite fitness?

>All apple genotypes in this experiment have their genomes sequenced and the scan for R genes will be done in parallel to this study to potentially detect regions associated with resistance to aphids.

Multigenotype test per single trees and systemic response. Authors should consider the systemic response within individuals and maybe to a less extent between individuals (suppositedly through metabolite exudation as reported in some studies about rhizosphere plant/plant communication). Testing 9 genotypes per tree will induce systemic responses and possibly decrease quite some variability that could be perceived testing single inoculation per tree. As much as I understand the setup of the study to maximise the number of combinaisons with minimal number of tree, it has to be specified and accounted for.

>We agreed with the reviewers; we could expect a systemic response of apple trees that can impact the fitness of a given aphid genotype within each apple tree. However, each aphid genotype will be randomly infested on each leaf of each tree, at least eight times, we will therefore control for this systemic infestation. Besides, we have now added a "leaf level" effect that will also allow considering a specific systemic effect depending on the level of the leaf in the tree used for the infestation. We described this equation in detail in the "statistical models" section, lines 680 and 681, and from line 689 to 691. We explained the rationale in the "sample size rationale" lines 450-455.

Tree ecophysiological traits to be included in the model. Authors should indicate how they will account for tree development and ecophysiological traits (measurements on modality 2). It does not seemed to be appearing in any of the tested models. Could it be added in the model as nested(?) covariates for each apple tree genotype?

>We have removed the description of ecophysiological data because we will not use that information to address our main objective, *i.e.*, for testing aphid local adaptation.

Overall this project seems very promising and will certainly result in important and inovative findings regarding parasite/plant adaptation and speciation in wild and domesticated trees. I am looking forward to seeing the results. Also, as findings may result in large datasets that could be further decomposed in several publication, autors will have to indicate any change in their expected publication scheme.

##################################

Reviewer 3 *Sharon Zytynska* ########

##################################

I have read and evaluated this pre-registration of an experiment planned for 2021. The experimental idea is interesting and should provide sufficient data to test the hypotheses presented. There has obviously been a lot of thought put into the design of the experiment with a clear timeline for experimental setup and a clear design at the field level. I do not have any major comments to add, just a couple of points to consider for the set up and data collection. I also added these to the submitted pdf but will highlight the main points here, in case you cannot access the attached file.

I like the idea to infest different leaves of the apple trees to the nine different aphid genotypes. However, there is no mention of the potential pitfalls of using cellophane bags outside for the duration of the season. Some bags we have previously used do not stand up well to water, and so rain periods could affect the integrity of the bags. Securing the bags effectively is also important to secure the aphids inside and keep natural enemies out. Elastic tends to degrade quickly in sunlight and break while thread/wool can loosen over time. My comment is to just consider this if you haven't already. Will you also cover the control tree leaves with cellophane bags? While this may have no impact, it is possible that any increased temperature/humidity inside the bags (small but unlikely to be zero) might impact pathogen growth leading to stress increase for the plant other than aphid infestation. Covering the leaves could also induce some plant signalling that may also influence plant responses. It is worth considering if you haven't already.

>We agree this is an important point to consider; we have performed additional trials since the submission of the proof for the experiment and have decided to use clip cages instead of cellophane bags based on infestation trials. We have updated our manuscript. We now present the clip cage design in Figure 4 of the manuscript (lines 414-416).

I think the counting of aphid adults and nymphs will be extremely time consuming after the first generation, so perhaps once the new offspring develop to adults it is reasonable to just count the numbers of winged vs unwinged aphids. Perhaps a small subset of replicates could be chosen to do more in depth aphid life cycle observations. With only three people per common garden, it might be worthwhile simplifying the data collection to maintain reliability.

>We agree with the reviewer. We have more precisely explained that lines 534-536. We will, if possible, count the different insect life stages (*i.e.*, aphid larvae (L1 to L5), apterous adults, and winged forms (Angeli and Simoni, 2006)). This will be done by scaling the individuals into three categories: big (apterous females), small (larvae), winged. We are also currently performing trials of infestation to identify the easiest way to record information about the aphid stages.

Other minor comments: Keep consistency with what you call a cultivar versus a genotype throughout the hypotheses and figures etc Question 5, all other subheadings for hypothesis have been in bold text Question 5, Hypothesis 1 (the 1 is missing) Throughout just check you write susceptible rather than sensitive (correct in figures but I noticed two places on page 6 and Table 1).

>We have updated the manuscript following these observations. We really appreciate all these remarks.

References

Angeli, G. I. N. O., & Simoni, S. (2006). Apple cultivars acceptance by *Dysaphis plantaginea* Passerini (Homoptera: Aphididae). *Journal of Pest Science*, *79*(3), 175-179.

Cornille, A., Gladieux, P., & Giraud, T. (2013). Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *Evolutionary Applications*, *6*(5), 737-748.

Cornille, A., Feurtey, A., Gélin, U., Ropars, J., Misvanderbrugge, K., Gladieux, P., & Giraud, T. (2015). Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: a basis for conservation and breeding programs in apples. *Evolutionary applications*, *8*(4), 373-384.

Cornille, A., Antolín, F., Garcia, E., Vernesi, C., Fietta, A., Brinkkemper, O., ... & Roldán-Ruiz, I. (2019). A multifaceted overview of apple tree domestication. *Trends in plant science*, *24*(8), 770-782.

Rom, C. R., Rom, R. C., & Stasiak, M. J. (1990). Size controlling apple rootstocks affect growth, spur quality, foliar nutrition and productivity. *Size controlling apple rootstocks affect growth, spur quality, foliar nutrition and productivity.*, *23*, 17-21.

Warneys, R., Gaucher, M., Robert, P., Aligon, S., Anton, S., Aubourg, S., ... & Heintz, C. (2018). Acibenzolar-S-Methyl Reprograms Apple Transcriptome Toward Resistance to Rosy Apple Aphid. *Frontiers in plant science*, *9*, 17