Dear Emmanuelle Porcher,

We revised our preprint submitted to PCI Evol Biol entitled "Potential adaptive divergence between subspecies and populations of snapdragon plants inferred from Q_{ST} – F_{ST} comparisons". My apologies for the time it took. Reviewers gave many constructive comments. We compared snapdragon subspecies and found differences. We therefore rerun all the analyses for each subspecies separately. We also rewrote several parts of the text to accommodate all the reviewers' comments, which improved greatly the quality of the manuscript. We hope that you – and the reviewers – will agree to follow up on this revised manuscript submission and that you will find this revised version of the manuscript more suitable for recommendation.

Please find below a point by point response to your comments and the comments by Sophie Karrenberg and Santiago Gonzalez (text of our response is in blue).

Dear Sara Marin,

I would first like to apologize for the time needed to return this decision, which is partly attributable to the difficulty of finding available reviewers in the middle of summer. Your preprint has now been read carefully by two experts (including Sophie Karrenberg (SK), who signed her review) and by me. We all agree that your manuscript addresses and important evolutionary question (patterns of adaptation to altitude), and that the experimental design was appropriate to address this question, such that I would ultimately be willing to recommend this preprint provided you can address the issues raised by the reviewers. Here are the most critical issues I identified, some of which were raised independently by both reviewers:

(1) Definition of local adaptation: both reviewers and I were lost with, or at the minimum confused by, your distinction between local adaptation (the meaning of which seems to vary throughout the paper, see comments by SK) and adaptation to latitudinal gradients, which involves local adaptation in the broad sense, such that individuals with the highest fitness in one location are on average those originating from this location. It is really important that this point is clarified. This may include for example specifying whether you consider local adaptation as a pattern or as a process or both. Another option for clarification would be to follow the first suggestion of SK, which is to focus more on biology and the question of adaptation to altitude, and how to disentangle it from adaptation to other environmental variables, which may vary in space at a finer grain than altitude (or not, see comment 7 by reviewer #2)

We now start the manuscript by defining local adaptation. We defined it as a process: "the evolutionary response to selection that makes populations fitter in their own local habitat than in other populations' local habitats". We modified the terms that we use throughout the text following SK and SG recommendations to avoid confusion between the local adaptation to the environmental conditions of their local sites of origin and to the altitude. Following SK comments, we gave more weight to the divergence between the two subspecies differ and to the biology rather than the methods in terms of adaptation.

(2) SK is worried that pooling the two subspecies might bias your analyses. There again, it is central to provide a solid justification for this pooling

We separated the two subspecies. This led to statistical power limitations for the study of some traits. We dropped these traits and kept those that were reliable.

(3) Methodological / statistical issues: both reviewers had several additional questions and suggestions regarding the statistical analyses, which I will not fully list here (although they all deserve attention), but which include for example: (i) How could you estimate the within population additive genetic variance, when the number of individuals and the number of families seem to be similar (i.e. about one individual per family)? (ii) How was the significance of the relationship between Fst (or Qst) and altitudinal differences tested in Figure 3? (iii) On the same figure, can the statistical significance of the differences in slopes be tested? More generally, the two reviewers often requested more details on the methodology (e.g. Qst-Fst comparisons, quantitative genetics analyses…), and I agree with them that some key information is sometimes missing.

(i) This was a misunderstanding caused by awkward writing. Sorry about that. We modified the text to clarify this point. The text now states precisely: "Nine to 42 seed families from each of the 13 study populations were grown outdoor in spring 2014 in a common garden at ENSFEA (Toulouse, France). Two plants per family were grown. Some plants died before measurements were done, which resulted in some families being represented by only one plant". We added the detail of the number of families and the total number of plants grown for each population in Table S1.

(ii) We had conducted a regression analysis for each relationship with altitude (not appropriate because the data was not independent) and a mantel test. The presentation of the results and the figure was unclear and confusing in the previous version. We have removed the regression results, clarified what was tested by using a mantel test, and conducted partial mantel tests following the reviewer's comments.

(iii) Unfortunately, it is not possible to estimate the statistical significance of slope differences for pairwise population differentiation data. The partial mantel tests, which we recognise are not perfect, are the best approach we can use for this type of data. It provides results close to those you were asking for. We agree that the information on methods and results was often confusing and that some information was missing. We have modified the text throughout the manuscript and hope that it will be clear.

(4) While the "Methods" section could be expanded, the two reviewers and I agreed that the text could also be shortened at places, particularly in the discussion.

We have shortened the text of the discussion.

Thank you very much for submitting your preprint for recommendation in PCI Evolutionary Biology. I hope you will find this feedback, and particularly the very thorough reviews of the experts, useful to revise your manuscript.

We appreciated the thorough reviews. They improved the quality of the manuscript. It was very pertinent to look into the subspecies divergence. It is generally acknowledged by the snapdragon plants research community that the two subspecies do not differ but for the flower colour and we are glad that the reviewers pushed us to investigate subspecies differences. The results of our indirect approach suggest that adaptation might have proceeded differently with

the phenotypes of the two subspecies, which opens the way for testing several hypotheses by using direct tests in the future.

Best regards Emmanuelle

Kind regards,

for the authors, Benoit Pujol


Sophie Karrenberg

Your manuscript on altitudinal divergence in snapdragon plants treats an interesting and timely question mostly with appropriate methods. However, I do have several difficulties to evaluate the quality of this research because important information appears to be missing from the manuscript. This concerns: I), most importantly, whether there were genetic or phenotypic differences between two subspecies that were apparently pooled here, according to Material and Methods and II) the experimental design such as spatial arrangement in the garden (how plants were arranged in how many containers) that should be taken account of as a random effect in the analyses, as well as crossing design and number of individuals per family used (see comments below).

Dear Sophie Karrenberg, thank you for your thorough review of our manuscript. We have taken into account all your comments, which improved the quality of our manuscript. The research community working on snapdragon plants generally considers that snapdragon subspecies do not differ but for the flower colour without really testing for it. Following your comment, we have compared the subspecies and found differences. We therefore analysed subspecies data separately. We had to drop a few traits in the process because the statistical power was not sufficient for them. Our results suggest that adaptation shaped differentially the phenotype of the two subspecies. This was a game changer for this paper. We modified the text and reran the analyses to accommodate this perspective. We also put the biology and the effect of altitude forward in the revised version of the manuscript. We took into account the blocks – containers – where plants were located randomly in the analyses, and clarified the crossing design. As you noticed, there was some confusion in the presentation of the numbers. More than one progeny was obviously present per family (we now give the number of plants and families for each population in Table S1). We are sorry that the presentation of many aspects was awkward in the methods and results. We have attended to it in this new version. We reduced the size of the discussion and gave clearer explanations in the methods and results. We have modified the text in many places. We hope that these modifications will have sorted out the problems that you mentioned. Please find below a more detailed, point by point response to your comments.

Regarding the subspecies, an explicit analysis would be most helpful (phenotypic and genetic differentiation, for example PCoA or structure analysis, potentially presented in the supplementary material). Pooling of the subspecies would probably only be appropriate if there are no considerable phenotypic differences or genetic differentiation between them. Fst in these analyses is used to reflect putatively neutral genetic differentiation -- this might be difficult

assumption if the dataset includes subspecies that are partially reproductively isolated from each other. The description of the subspecies in the method section as well as the citations do fuel this concern. Differences in habitat and flower color between subspecies are described — even pollinator-mediated reproductive isolation could be present in this system. Very careful and convincing analyses and explanations would be needed to motivate the pooling. And —if there are no differences between subspecies, would that not suggest that there might just be a flower color polymorphism? It is important to note that, without this potentially troublesome pooling of subspecies, the number of populations would probably be too low for the present analysis and this is why this issue is so critical. Provided that pooling can be justified, and the design clarified, this manuscript could become a good one, however, further improvements in the analysis and presentation are also needed.

We conducted a PCA analysis, see graphs below. No contrast could be observed between subspecies based on the graphs. We nevertheless conducted statistical analyses to test for phenotypic differences between subspecies. Hierarchical generalized linear models were conducted with population nested in subspecies. We found differences. This was surprising but very interesting to find out. The subspecies mostly share the same ecological niche. They are interfertile but are reproductively isolated by pollinators. Based on this information, we decided to consider them separately in all the following local adaptation analyses. By doing so, we lost statistical power and some traits were not reliable any longer for QST-FST comparisons. We dropped these traits from the study. The traits that were fine to analyse were all vegetative traits.
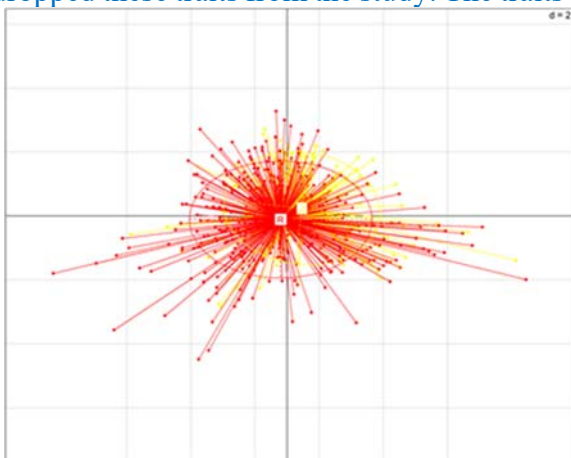
PCA on vegetative traits. Yellow = *A. m. striatum* Red = *A. m. pseudomajus*
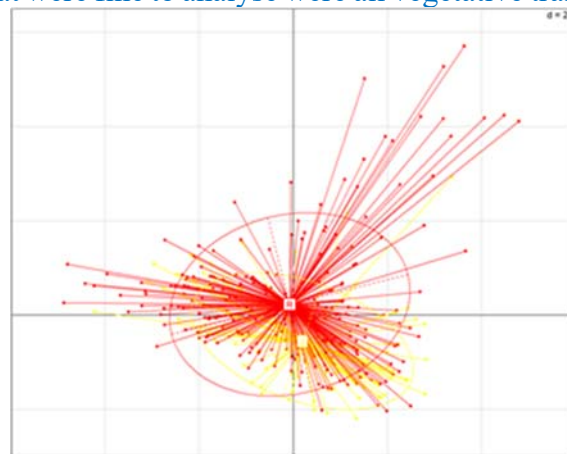
PCA on reproductive traits. Yellow = *A. m. striatum* Red = *A. m. pseudomajus*

The main changes that I believe are necessary are: 1) Focus on biology. The paper would be more interesting if focus was shifted from methods and general interpretations to the biology of altitudinal adaptation. This would be facilitated by presenting the trait-altitude correlations first. These results are interesting for all traits, currently they are only presented for 3 out of 13 traits. From that you could move on to more complex analyses to identify the underlying causes of such trait clines. In the current version, there is overall rather little on HOW the traits change along altitude, for example, there are no details on leaf traits, even though there is a lot of literature of variation in SLA (see citations in literature mentioned below). It would also be very good if explicit expectations for all traits or for trait groups, based on the literature, could be added in the introduction. I would also appreciate an introduction ending in specific questions.

We agree that QST-FST approaches are well-known and made changes accordingly. We now focus more on the biology. We reduced the text on the QST-FST approaches. These approaches are now presented more briefly. We present their limits but removed the unnecessary details on

the method itself and any text that could read as if it praised their use because they are far from perfect, and there is no novelty there. We now present the biological expectations in terms of trait adaptive change with altitude in the introduction and discus the results that we found in *A. m. striatum* in the light of these expectations.

2) In the methods part, you write that you used the between family variance component to estimate the within population additive genetic variance for a trait, as is commonly done. However, Table S1 states, for reproductive traits, the total number of families is 372 and the number of individuals is 380, such that you cannot have had more than one individual per family in most cases, unless you used a subset of families with more individuals (I could not find information on the number of individuals per family). With 1 individual per family it should be very difficult or impossible to estimate the between family variance component. In Table S3 you do not give the within family variance component so I wonder whether what you state as the within population variance component really is the between family variance component. This issue is in need of clarification. It would also be good if you could explain why you have fewer measurements for reproductive traits than for vegetative traits.

We agree that the presentation of the setting was awkward and misleading. As mentioned in our response to Emmanuelle Porcher: << We modified the text to clarify this point. The text now states precisely: "Nine to 42 seed families from each of the 13 study populations were grown outdoor in spring 2014 in a common garden at ENSFEA (Toulouse, France). Two plants per family were grown. Some plants died before measurements were done, which resulted in some families being represented by only one plant". We added the detail of the number of families and the total number of plants grown for each population in Table S1.>>. We had fewer reproductive measures because all plants grew but not all plants produced flowers, which is why reproductive traits harboured less statistical power. Consequently, when we separated the data set in two parts, one for each subspecies, reproductive traits could not reliably be analysed and we had to remove them from the study.

3) Methods, QST - FST. I found the description of the QST - FST comparisons, using expected distributions, unclear and in part contradictory (see comments in the text). In particular, it is not clear to me how you arrived at the confidence intervals for QST on Figure 2. It would fit better if you presented the distribution of expected values for QST together with the observed values, this is the method you stated in the methods section. Alternatively, you could change the methods section.

Our presentation was indeed incomplete. In particular, the presentation of the two methods that we used independently was confusing. We clarified the text to explain each method separately: "Comparisons between F_{ST} and overall Q_{ST} values were performed for each trait based on two methods: i) a comparison of confidence intervals, the Q_{ST} is considered non significantly different from neutral differentiation when the confidence interval of the overall Q_{ST} for a trait overlaps the mean F_{ST} value, ii) a bootstrapping method developed by Whitlock and Guillaume (2009). This last approach aims at comparing the observed difference between the overall Q_{ST} and the F_{ST} values with the expected simulated distribution of this difference under a scenario of neutral evolution. We generated 100 000 bootstrap replicates of the expected Q_{ST} - F_{ST} difference under the neutrality hypothesis for each trait, and built the corresponding distribution. In this approach, P values were estimated by assessing whether the observed value of the Q_{ST} - F_{ST} difference overlapped its expected distribution under neutrality. We used the modification by Lind *et al.* (2011) of the approach of Whitlock and Guillaume (2009) to estimate the variance components of the simulated values of the Q_{ST} - F_{ST} difference".

"Confidence intervals of Q_{ST} values were calculated following a parametric bootstrap method adapted from O'Hara and Merilä (2005)."

4) Analysis of pairwise data: the Methods section describes the commonly used and suitable analyses for correlation tests among matrices (pairwise data), Mantel tests and partial Mantel tests, controlling for the effect of neutral genetic divergence. In the results however, a regression analysis is presented (Fig. 3) that is not suitable for this type of non-independent data points. The text states that the Mantel test was used to assess significance of the regression — this is not possible. The actual data points are also not displayed, this is a bit worrisome due to the subspecies issue. My suggestion is to re-work this analysis according to the methods described and to display the data points on Fig. 3.

Sorry about that. Awkward wording lead to some confusion about which of these methods we used in the paper. We modified the text and approach accordingly to your comment. The regression was indeed not adapted and we removed it. We have clarified the text and now present results for the mantel and partial mantel tests only.

5) (Local) adaptation, interpretation. The interpretation of what constitutes local adaption is highly variable throughout the manuscript. At some places, the common cautious interpretation of QST - FST analyses, that they can provide hypotheses on traits that may have adaptively diverged due to selection (see for example review by Leinonen et al. 2013, Nature Reviews Genetics) is used. At quite some other places, however (abstract, introduction, discussion, marked in detailed comments), it appears that local adaptation is inferred whenever trait divergence cannot be attributed to either the altitudinal gradient or neutral divergence. I cannot be sure that the text is really meant that way, of course, but I testify here that it reads that way. This interpretation would be incorrect (genetic drift and many other processes could be responsible for trait divergence instead). Any text that can potentially be misunderstood is probably better changed.

We have modified the text accordingly to restore the coherence of our interpretation of what constitutes local adaptation throughout the manuscript. Right away from the start of the manuscript, we define local adaptation as a process: "the evolutionary response to selection that makes populations fitter in their own local habitat than in other populations' local habitats". We modified the text accordingly to tone down the conclusions that we obtained from QST-FST comparisons. We now say throughout the text that significant results do not confirm the presence of local adaptation but rather suggest local adaptation, reflect a potential for local adaptation, or identify candidate traits for local adaptation. We hope that this will read right. Please do not hesitate to point out specific sentences that would still be a problem.

6) Interpretation of the results on germination. The corresponding paragraph in the discussion unfortunately appears contradictory and the two types of results, overall QST - FST and pairwise QST - FST have not been compared (see comments in the text). Moreover, the text also appears as if it was not clear to the authors that a lack of quantitative genetic differentiation in traits does not exclude phenotypic plasticity (i.e., this is not really an alternative interpretation). It seems that this part needs adjustment.

We clarified the text to avoid any confusion. Plasticity is not expected to drive the overestimation of QST because the experimental environment is similar between populations. If populations express a convergent phenotypic response through plasticity to a similar

environment, then a lower QST than expected might be found. This latter mechanism might be an alternative hypothesis to standardizing selection when explaining QST<FST.

7) Interpretation of genetic divergence. You invoke reproductive isolation, when it appears that isolation by distance would be a sufficient interpretation, unless you refer to the subspecies differences here (see comments above and in the text). However, this cannot be evaluated from the data currently presented, as details on genetic differentiation are missing (see above). I suggest that this interpretation is re-worked one a more detailed analysis is presented.

We have clarified these aspects in the text. We now present the corresponding results to accommodate this comment in the paper. In brief, we did not find any isolation by distance. Mountains were a plausible ecological barrier to gene flow between populations. Subspecies are differentiated and pollination is an acknowledged pre-zygotic reproductive barrier between subspecies. For more details, please have a look at our previous paper on the ecological factors influencing the spatial genetic structure of snapdragon populations (Pujol et al, 2017 Botany Letters, which we cite in the paper).

8) Literature. The manuscript would much benefit by adding more literature on the analysis of altitudinal gradients and transplant experiments, for example: Luo, Y., A. Widmer, and S. Karrenberg. 2015. The roles of genetic drift and natural selection in quantitative trait divergence along an altitudinal gradient in Arabidopsis thaliana. Heredity 114:220. Halbritter, A. H., S. Fior, I. Keller, R. Billeter, P. Edwards, R. Holderegger, S. Karrenberg, A. R. Pluess, A. Widmer, and J. M. Alexander. 2018. Trait differentiation and adaptation of plants along elevation gradients. J. Evol. Biol. [Disclaimer: these are papers I am an author on-- but they really are related here... as are many of the citations in them].

We now cite Halbritter et al. 2018.

9) Methods are not really new. Throughout the intro and the discussion, the paper reads as if this mainly is a method advancement. This appears somewhat exaggerated, as similar methods have been used before and since a longer time (>5 yrs). I feel that the work of others should be acknowledged much more here, only few citations on such studies that are able to identify selective agents are given (for example, Hangartner et al).

We modified the text accordingly. We toned down the originality and the interest of the approach (altitudinal gradient effect on QST-FST comparisons). We also made sure that the work of others is better acknowledged by adding references (for example, Halbitter et al. 2018).

10) Shortening. I find that the text could likely benefit from shortening, in particular in the introduction and discussion, probably to at least to 2/3 of its present length. I have indicated several parts that appear repetitive in the detailed comments.

We have provided more details for the sake of clarity in the Material and methods and the results sections. We reduced the size of the text in the introduction and the discussion by removing redundancy, shortening the text about the methodological approach to draw a better focus on the biology.

11) More comments. I have made comments similar to the above comments as well as many further detailed comment directly on the text using pdf commenting tools in Adobe Acrobat. I

am not repeating these detailed comments here. Such a list of details would be very difficult to write and read, because the text did not include any line numbers to refer to.

Thank you for taking the time to comment directly on the text. We have followed all your suggestions.

I hope that you will find my comments clear and helpful for further revisions of your work.

Your recommendations were particularly useful and much appreciated to improve the quality of the manuscript.

With best wishes, Sophie Karrenberg

Kind regards,

for the authors, Benoit Pujol

Santiago Gonzalez
Evolutionary inference from Qst-Fst comparisons: disentangling local adaptation from altitudinal gradient selection in snapdragon plants
This manuscript provides new and relevant insights on local adaptation of snapdragons based on quantitative genetic data from a common garden and previously published nuSSR data. However, I found confusing the way the authors constructed the paper, contrasting what they called 'local adaptation' (i.e. populations standing alone) versus 'adaptation to altitudinal gradients'. From my perspective, adaptation to an environmental gradient involves also local adaptation (of each of the population of the gradient). In addition, other traits with Qst > Fst but not associated with altitudinal gradients (and thus part of the 'local adaptation' group) may still be associated to other environmental gradients (that have not been measured in this study, e.g. for soils).

We clarified the text accordingly. As mentioned above in response to EP's comments: <<We now start the manuscript by defining local adaptation. We defined it as a process: "the evolutionary response to selection that makes populations fitter in their own local habitat than in other populations' local habitats". We modified the terms that we use throughout the text following SK and SG recommendations to avoid confusion between the local adaptation to the environmental conditions of their local sites of origin and to the altitude.>>

In addition, the strongest evidence of local adaptation along the altitudinal gradient does not come from the overall Qst-Fst comparison but from the pairwise analyses and the simple correlations between phenotypic means and altitude (Figures 3 and 4).

We now discuss the biological significance of the correlation between phenotypic mean and altitude in accordance with hypotheses formulated in the introduction. This reinforces our discussion of the biology rather than discussing the pros and cons of QST FST methods that already widely acknowledged in the literature. The number of traits that we discussed had to be modified after following up on the suggestion to look into subspecies divergence. Since we

found divergence, we conducted QST-FST comparisons in each subspecies separately. Reproductive traits (measured in less plants than vegetative traits because not all plants flowered) had to be dropped from the paper for lack of statistical power when the analysis is conducted in each species separately. We found evidence for potential adaptation to altitude in only one subspecies (*Antirrhinum majus striatum*).

Finally, please, have a careful read of the manuscript to correct errata, uncomplete sentences, unfinished citations, etc.

We agree that the text was unclear in many places and could lead to some confusion, in particular with the methods. We have clarified the text throughout the manuscript, detailed the methods and results for the sake of clarity and synthesized the intro and the discussion to put biology forward and tone down the interest of this type of approach because it is already well described in the literature.

Some more specific comments follow:
Introduction
1. The Qst-Fst method, as applied here, has also important flaws compared to reciprocal transplantation approaches, which I think should be presented and briefly discussed in the Introduction. Perhaps the most important one is the bias produced by the common garden testing environment. Parameters like heritability or Qst may differ greatly when computed in different environments, and in particular field estimates are very different from those obtained in common gardens. This should be acknowledged.

We agree and now acknowledge this clearly in the text : "In plants, reciprocal transplants directly comparing fitness between the native habitat and the foreign habitats are often preferred to QST-FST approaches conducted in common gardens because they allow to evaluate the effect of environmental conditions (Etterson, 2004; Angert and Schemske, 2005; Kim and Donohue, 2013). When the conditions for the reciprocal transplant cannot be easily met, QST-FST comparisons represent an opportunity for exploring local adaptation hypotheses."

2. Despite being common to use adaptation to altitude as a first approach to understand adaptation to climate change, there are some fundamental differences that make the two processes very different, in particular for traits related to photoperiod where, for example, day length is also relevant. These differences should be acknowledged and discussed in the Introduction and in the Discussion.

We agree that using the environmental conditions associated with altitude to assess the potential effect of climate change is far from perfect. In the revised version of the manuscript, we do not mention climate change. We only mention that climate variables are associated with altitudinal differences amongst other environmental variables: "we investigated whether quantitative genetic differentiation increased with altitudinal difference, with the hypothesis that environmental changes associated with altitude, which include a suite of climatic variables, drove adaptive responses. Other environmental variables (e.g., atmospheric pressure) can also change with altitude. Our study also ultimately participate in evaluating whether Q_{ST}-F_{ST} comparisons can be used as a tool to identify candidate traits involved with the potential adaptation of populations to altitudinal gradients, and thereby climate differences."

3. I like frogs, but still, why to provide examples of frog adaptation to altitude in the Introduction? There is a rich literature in plants that could be used instead, to illustrate patterns

of genetic adaptation and plasticity along altitudinal (and latitudinal, which are perhaps more relevant to discuss climate change) gradients.

We agree and made sure to cite more references that are specific to plants adaptive changes with altitude (e.g., Halbritter et al. 2018, in JEB, which reviews this field of research)

Material and Methods
4. Key information is missing about the quantitative genetics models used by the authors, in particular about sample sizes. Also, it would be good to write down the model itself, as the experimental design is not very clear. For example, did the experiment follow any block design? How many families? How many full-sibs per family? I have seen Table S1, but number of samples seem similar to number of families, which I don't understand…

The text was indeed confusing. We modified the text to clarify this aspect and now give the exact number of plants and sib-families for each population in a table: "Nine to 42 seed families from each of the 13 study populations were grown outdoor in spring 2014 in a common garden at ENSFEA (Toulouse, France). Two plants per family were grown. Some plants died before measurements were done, which resulted in some families being represented by only one plant (Table S1, Supporting Information)." We also precise that plants were located randomly into containers and that the potential effect of these containers was taken into account into the statistical analyses.

5. In my opinion, the most original test presented in the manuscript is the pairwise analysis of Qst versus altitudinal differences (Figure 3). I am aware that the authors designed the study to include different altitudinal gradients so that dependence with distance is broken, but still some distance effects may remain and I think pairwise distance between populations should be included in these analyses (by means of a partial Mantel test or by using it as covariate).

We have modified this approach in the revised manuscript. Regressions used in the previous version were not suitable for this data (pairwise population differentiation because data points are not independent). We now use partial mantel tests. We do not have enough statistical power to fit pairwise distance in these analyses. We are nevertheless confident with our approach because (i) no isolation by distance was found on the basis of putatively neutral markers (we now present this finding in the paper) and (ii) as you mention, our sampling scheme (populations from different valleys and mountains) aimed at avoiding such spatial distance effect.

6. "Potential effects on population genetic differentiation… for by using AMOVA". Not clear how the authors did this… and also, should not Qst estimates be also corrected in the same way? I find a bit weird, for example, that different subspecies are pooled to estimate heritability and Qst.

Following this comment and comments by SK, we have investigated the phenotypic divergence between subspecies and found some. We therefore ran analyses in each subspecies separately and used different types of analyses: comparison between QST CI and mean FST, QST-FST difference simulation tests, mantel tests, mantel partial tests. Text, table and figure modifications related to these changes appear throughout the paper. We modified the text accordingly to incorporate the question on subspecies. As mentioned above in our response to EP: "It was very pertinent to look into the subspecies divergence. It is generally acknowledged by the snapdragon plants research community that the two subspecies do not differ but for the

flower colour and we are glad that the reviewers pushed us to investigate subspecies differences. The results of our indirect approach suggest that adaptation might have proceeded differently with the phenotypes of the two subspecies, which opens the way for testing several hypotheses by using direct tests in the future."

Results
7. The manuscript focuses on the power of the Qst-Fst method to detect adaptation to gradients, but still most of the traits with Qst > Fst did not correlate with altitudinal variation but with other, unknown factors. The story the paper tells is, of course, the author's own business, but still it makes me wonder whether these other traits with Qst > Fst do not deserve more attention (for example in the Discussion) and a substantial change of the main argumentation line.

We agree. QST FST methods and their interest and limits are thoroughly documented in the literature. We therefore removed text throughout the paper on methodological aspects to keep what was only necessary and focus more on the biological dimension of the study.

8. Figure 3. Differences in slope are pretty clear indeed, but I still think that they should be tested formally, using a test for differences of regression slopes.

As mentioned earlier, the regression approach was not adapted and was removed. We now use partial mantel tests.

Discussion
9. Overall I found the discussion a bit repetitive and unbalanced towards supporting the Qst-Fst approach as a means to detect adaptation along gradients. In my opinion, the advantages of the approach are not so well supported by the results. I appreciate the pairwise approach (Figure 3), which is relatively original, but still, as commented before, it does not take advantage of sampling along the gradients to produce more robust estimates.

We agree and modified the text accordingly. We removed redundancy and synthesized the content of the discussion to go straighter to the point. As mentioned earlier, we now focus on the biology rather than the approach so that our revised text should not read anymore as unbalanced towards supporting this type of approach.

10. "…trait homogenisation caused during the experiment by phenotypic plasticity might be another plausible explanation." I really don't see this point… I would say that is the other way around, that the differences you see in the field are the ones that are caused by phenotypic plasticity (please, notice that substantial additive variance and differences among families are still present in the common garden, as shown by high heritability for time to flowering and germination rate).

The same point was raised by SK. We have modified the text to clarify this aspect. "Plasticity is not expected to drive the overestimation of QST because the experimental environment is similar between populations. If populations express a convergent phenotypic response through plasticity to this similar environment, then a lower QST than expected might be expected. This latter mechanism might be an alternative hypothesis to standardizing selection to explain that QST<FST."

11. I also found weird that germination rate does not have Qst > Fst overall but it is significant in the pairwise test, i.e. Qst > Fst along altitude for pairs of populations.

The regression results were removed as they were not suitable to pairwise population data. Our results – that are now separated between subspecies – show a Germination date QST<FST in *A. m. striatum* but not in *A. m. pseudomajus*. Germination date QST did not show a significant relationship with altitudinal differences when using the partial mantel test approach.

Tables
12. TABLE S3. Please add the standard error to the heritability estimates.

We agree that this is important information. We added the 95% confidence intervals though, rather than SE, because we find them a lot more informative and conservative when it comes to assessing the statistical reliability of heritability estimates.

Several of your comments overlapped with comments by SK. Please read our response to her comments, as they might be complementary. We thank you for having thoroughly reviewed our paper and for your constructive comments that proved useful for improving the quality of this manuscript.

Kind regards,

for the authors,

Benoit Pujol

1   **Potential adaptive divergence between subspecies and populations of snapdragon plants**

2   **inferred from Q$_{ST}$ – F$_{ST}$ comparisons**

3   Sara Marin[1,2§], Anaïs Gibert[1§], Juliette Archambeau[3], Vincent Bonhomme[4], Mylène Lascoste[2]

4   and Benoit Pujol[1,2]*

5

6   [1] PSL Université Paris: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Université de Perpignan,

7   52 Avenue Paul Alduy, 66860, 66360 Perpignan Cedex, France.

8   [2] Laboratoire Évolution & Diversité Biologique (EDB UMR 5174), Université Fédérale de

9   Toulouse Midi-Pyrénées, CNRS, IRD, UPS, Toulouse, France

10   [3] BIOGECO, INRA, University of Bordeaux, Pessac, France

11   [4] Institut des Sciences de l'Évolution (ISEM), équipe "Dynamique de la biodiversité, anthropo-

12   écologie", UMR 5554, Université de Montpellier, CNRS, IRD, EPHE, Place Eugène Bataillon,

13   Cc 065, 34095 Montpellier cedex 05, France

14

15   § Both authors share first authorship of this work

16   * Author for correspondence: benoit.pujol@univ-perp.fr

17

18   **ABSTRACT**

19   Phenotypic divergence among natural populations can be explained by natural selection or by

20   neutral processes such as drift. Many examples in the literature compare putatively neutral (F_{ST})

21   and quantitative genetic (Q_{ST}) differentiation in multiple populations to assess their potential

22   evolutionary signature and identify potential traits involved with local adaptation. Investigating

23   these signatures in closely related or recently diversified species has the potential to shed light

24   on the potential divergence processes acting at the interspecific level. Here, we conducted this

25   comparison in two subspecies of snapdragon plants (eight populations of *Antirrhinum majus*

26   *pseudomajus* and five populations of *A. m. striatum*) in a common garden experiment. We also

27   tested whether altitude was potentially involved with population phenotypic divergence. Our

28   results identified candidate phenological and morphological traits involved with local

29   adaptation. Most of these traits were identified in one subspecies but not the other. Phenotypic

30   divergence increased with altitude for a few biomass-related traits, but only in *A. m. striatum*.

31   These traits therefore potentially reflect *A. m. striatum* adaptation to altitude. Our findings imply

32   that adaptive processes potentially differ at the scale of *A. majus* subspecies.

33

36

37

2

INTRODUCTION

Local adaptation - the evolutionary response to selection that makes populations fitter in their own local habitat than in other populations' local habitats - is widespread in both plant and animal species (Kawecki and Ebert, 2004; Leinonen et al., 2013; Halbritter et al., 2018). There is evidence for its role in the adaptive divergence of plant species (Leimu and Fischer, 2008; Hereford, 2009; Halbritter et al., 2018). For example, empirical studies have demonstrated differential adaptation in plant sister species or hybridizing species, for instance between pairs of *Silene* species (Favre et al., 2017), *Senecio* species (Abbott and Brennan, 2014), *Mimulus* species (Angert and Schemske, 2005). These studies compared local adaptation for sister species confronted to different ecological requirements. Different species may also respond similarly to a same type of environmental gradient. Recently, Halbritter et al. (2018) combined studies of multiple plant species along elevation gradients. They found significant evidence for adaptation to different elevations in terms of survival and biomass. Their results also showed variation across species in plant responses to elevation. The study of local adaptation in populations of closely related taxa exposed to environmental gradients, e.g. altitude, is an opportunity to investigate the conditions promoting or impeding the consistency of adaptive responses.

An indirect approach to investigate whether local adaptation might potentially be involved in the phenotypic divergence of populations is the Q_{ST}-F_{ST} comparison (Spitze, 1993; Merilä and Crnokrak, 2001; McKay and Latta, 2002). The comparison of population genetic differentiation estimated for putatively neutral molecular markers with the population quantitative genetic differentiation estimated for phenotypic traits can be used to identify potential candidate traits playing a role in local adaptation (Whitlock, 2008). This is done by estimating whether trait quantitative genetic differentiation among populations is more likely the potential result of

63   divergent selection ($Q_{ST}>F_{ST}$), stabilizing selection ($Q_{ST}<F_{ST}$), or neutral evolutionary

64   divergence ($Q_{ST}=F_{ST}$, e.g., as a result of drift). Some debate around the accuracy of Q_{ST}-F_{ST}

65   comparisons resulted in a variety of methodological adjustments (Whitlock, 2008; Edelaar et

66   al., 2011; Ovaskainen et al., 2011; Whitlock and Gilbert, 2012). In plants, reciprocal transplants

67   directly comparing fitness between the native habitat and the foreign habitats are often preferred

68   to Q_{ST}-F_{ST} approaches conducted in common gardens because they allow to evaluate the effect

69   of environmental conditions (Etterson, 2004; Angert and Schemske, 2005; Kim and Donohue,

70   2013). When the conditions for the reciprocal transplant cannot be easily met, Q_{ST}-F_{ST}

71   comparisons represent an opportunity for exploring local adaptation hypotheses.

72

73   In our study, we investigated patterns of potential local adaptation in two closely related plant

74   subspecies by using Q_{ST}-F_{ST} comparisons estimated in a common garden experiment, and

75   evaluated whether altitudinal gradients might play a role in the potential adaptive divergence of

76   populations. This evaluation was conducted in snapdragon plants (*Antirrhinum majus* L.,

77   Plantaginaceae). We studied eight populations of magenta-flowered *A. m. pseudomajus* and five

78   populations of yellow-flowered *A. m. striatum* sampled along altitudinal gradients. These two

79   species are interfertile (Andalo et al., 2010). They are distributed parapatrically, with the

80   geographic range of *A. m. striatum* surrounded by the range of *A. m. pseudomajus*, and come

81   frequently into contact at the margins of their ranges where there is evidence for gene exchanges

82   (Khimoun et al., 2011; Ringbauer et al., 2018). Their geographic separation is not explained by

83   actual climatic differences, as illustrated by the substantial overlap of environmental conditions

84   between the two species (Khimoun et al., 2013). This system is therefore promising to explore

85   potential differential adaptive responses between closely related subspecies, in particular

86   regarding the potential role played by altitude in their adaptive divergence.

87

88   There is poor support in the literature for adaptive changes in reproductive traits along

89   altitudinal gradients (Halbritter et al., 2018). In contrast, adaptive differentiation along

90   altitudinal gradients is expected for biomass-related traits and height, with a trend toward

91   smaller plants for populations from high altitude compared to plants from lowland sites

92   (Halbritter et al., 2018). We tested this hypothesis for five morphological traits (the basal stem

93   diameter, the number of branches on the plant, the number of vegetative nodes on the main

94   stem, and the total height of the plant). We also studied three additional traits: a phenological

95   trait (the germination date), a developmental trait (the average internode length) and a

96   functional trait (specific leaf area, SLA). We expected populations from higher altitudes to

97   germinate later, over a shorter period (Gimenez-Benavides et al., 2006; Donohue et al., 2010).

98   This is because germinating later over shorter periods allows plants to track the late arrival and

99   the shorter-term availability of suitable climatic conditions for growth at higher altitudes

100   (Körner, 1999). Because the internode length is a trait related to both plant height and growth

101   rates, we had no clear expectations. Finally, SLA refers to leaf construction cost and captures

102   information about leaf economic strategies (Wright et al., 2004); low SLA suggests high leaf

103   construction cost and high stress tolerance. Selective pressures associated with lower

104   temperatures at higher elevations are expected to promote leaf trait syndromes associated with

105   superior stress tolerance but inferior competitiveness (Read et al., 2014). These relationships

106   are generally stronger among species than among populations of the same species (Read et al.,

107   2014). Therefore, we expected no correlation or a negative correlation between SLA and

108   elevation among populations.

109

110   In this study, we estimated neutral genetic differentiation (F_{ST}), and quantitative genetic

111   differentiation (Q_{ST}) based on trait heritability (h^2) in *A. majus*. Previous studies of genetic

112   differentiation between populations and subspecies at putatively neutral microsatellite markers

113 brought evidence that gene flow was limited between populations (Debout et al., 2012; Pujol et

114 al., 2017), which sets the stage for local adaptation. We then tested for the hypothesis that traits

115 were potentially involved with local adaptation by comparing Q_{ST} and F_{ST}. Finally, we

116 investigated whether quantitative genetic differentiation increased with altitudinal difference,

117 with the hypothesis that environmental changes associated with altitude, which include a suite

118 of climatic variables, drove adaptive responses. Other environmental variables (e.g.,

119 atmospheric pressure) can also change with altitude. Our study also ultimately participates in

120 evaluating whether Q_{ST}-F_{ST} comparisons can be used as a tool to identify candidate traits

121 involved with the potential adaptation of populations to altitudinal gradients, and thereby

122 climate differences.

123

124 **MATERIAL AND METHODS**

125 **Study system**

126 *Antirrhinum majus* L. (Plantaginaceae) is a hermaphroditic, self-incompatible, short-lived

127 perennial species, characterized by a patchy distribution in southern Europe centred over the

128 Pyrenees Mountains (Khimoun et al., 2011). This species occurs from sea level to an altitude

129 of 1900 m (Andalo et al., 2010), on limestone or siliceous substrates and in habitats with

130 contrasted moisture regimes (rainfall 500-1000 mm per year), where it forms restricted patches

131 mostly in rocky outcrops and screes. *A. majus* thrives in disturbed habitats, and is especially

132 common along roadside and railway embankments (Khimoun et al., 2011).

133

134 **The subspecies level**

135 *A. majus* produces annual inflorescences with zygomorphic flowers. The colour of flowers is

136 either magenta or yellow and distinguishes two interfertile subspecies *A. m.* ssp. *pseudomajus*

137 and *A. m.* ssp. *striatum* respectively (Andalo et al., 2010). At the genetic level, ~1% genetic

138 differentiation was found between *A. m.* ssp. *pseudomajus* and *A. m.* ssp. *striatum* on the basis

139 of putatively neutral microsatellite loci, which was one order of magnitude lower than the ~10%

140 differentiation found among these populations (Pujol et al., 2017). There is evidence for gene

141 exchange between subspecies in multiple populations across contact zones (Khimoun et al.,

142 2011). Genome scans across a particular contact zone in the Pyrenees also revealed little to

143 negligible differentiation between the two subspecies, with the exception of loci underlying

144 flower colour differences between the two subspecies that were characterized by high

145 differentiation (Whibley, 2006; Tavares et al., 2018). At the environmental level, the separation

146 between the geographic distribution of *A. m.* ssp. *pseudomajus* and *A. m.* ssp. *striatum* is not

147 explained by habitat differences, as illustrated by the substantial overlap of environmental

148 conditions between the two species (Khimoun et al., 2013).

149

150 **Collection sites and plant material**

151 Thirteen wild populations of *A. majus* were sampled in 2011 across the geographic range

152 (between north-eastern Spain and south-western France) to represent the overall diversity of

153 the species, with eight populations of *A. m.* ssp. *pseudomajus* and five populations of *A. m.* ssp.

154 *striatum* (Figure 1, Table S1, Supporting Information). For each subspecies, we sampled

155 populations from low and high altitude habitats in different parts of the species geographic

156 range. Populations sampled along elevation gradients are likely to be confronted to contrasted

157 environmental conditions. Fifty-year averages (1950-2000) of mean annual temperature and

158 annual average rainfall were extracted from the WorldClim database (resolution 1km²,

159 www.worldclim.org, Hijmans et al., 2005). They ranged from 14.8°C and 52mm (BAN, 61m

160 above sea level) to 6.1°C and 94mm (MON, 1564m above sea level) (Figure S1, Supporting

161 Information). The sampling of populations in different valleys or on different summits limits

162   spatial autocorrelation in the data and shared phylogeographic history between populations

163   from similar altitudes.

164

165   In each wild population, seeds were randomly collected from plants across their entire spatial

166   distribution in October 2011. These seeds sampled in the wild were used to grow plants in 2012,

167   in a greenhouse at the CNRS Experimental Ecology Station in Moulis, France. Seeds were sown

168   in spring in individual pots ($9 \times 9 \times 10$ cm) filled with universal compost. Plants germinated

169   and grew with no nutrient addition under an average temperature from 15 to 28°C and weekly

170   watering. Mature plants were hand-pollinated during the summer 2012. These plants were not

171   measured. Crosses were conducted within populations where mates from different families

172   were assigned randomly. The seed collection of full sib families produced by these plants was

173   stored at room temperature, in the dark, under dry conditions until they were used to produce

174   the plants measured in our experiment. This intermediate generation of plants grown in

175   controlled conditions from seeds collected in the wild allowed us to reduce potential maternal

176   effects caused by natural habitat differences on trait measurements.

177

178   **Common garden experiment**

179   Nine to 42 seed families from each of the 13 study populations were grown outdoor in spring

180   2014 in a common garden at ENSFEA (Toulouse, France). Two plants per family were grown.

181   Some plants died before measurements were done, which resulted in some families being

182   represented by only one plant (Table S1, Supporting Information). Plants were grown in

183   individual pots ($9 \times 9 \times 10$ cm) filled with universal compost, with no nutrient addition, under

184   outdoor climatic conditions (average month temperatures ranging from 20.6 to 21.5°C and

185   cumulative monthly rainfall ranging from 28.3 to 73.4mm). Plants were arranged in a

186   randomized block design (40 plastic containers, $600 \times 400 \times 120$ mm) with each containing 24

187   randomly chosen plants. The bottom of each container was covered with an irrigation sheet

188   (400 g.m^{-2}) that allowed to regulate the moisture of the compost. Plants were supplied with

189   water in case of prolonged drought. Damage caused by herbivorous insects were contained by

190   using a wintering veil. This veil also limited pollination.

191

192   **Phenotypic data**

193   We measured several vegetative traits on each individual: a phenological trait (the germination

194   date), a functional trait (the specific leaf area, SLA), a developmental trait (the average

195   internode length) and multiple morphological traits. Morphological traits included the basal

196   stem diameter, the number of branches on the plant, the number of vegetative nodes on the main

197   stem, and the total height of the plant. The SLA refers to leaf construction cost and captures

198   information about species leaf economic strategies (Wright et al., 2004). It was calculated as

199   the ratio between the cumulated area of five mature but non-senescent fresh leaves and their

200   oven-dried mass (Pujol, Salager, et al., 2008; Pérez-Harguindeguy et al., 2016). Leaf area was

201   measured by using the R package Momocs v. 1.2.9 (Bonhomme et al., 2014).

202

203   **Molecular analyses**

204   To infer genetic diversity estimates in each population and to compute F_{ST}, we genotyped the

205   637 plants. DNA was extracted from silica gel dried leaf samples using the Biosprint 15 DNA

206   Plant kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Individuals

207   were genotyped for 23 putatively neutral microsatellite markers that were developed for

208   population genetic studies (Debout et al., 2012; Pujol et al., 2017). To compute F_{ST}, we used

209   population pairwise F_{ST} estimates and the overall F_{ST} estimate amongst populations from the

210   study by Pujol et al. (2017). We used the GenoDive 3.0 software (Meirmans and Van Tienderen,

211   2004) to compute the complementary parameters required for this study, e.g., the genetic

212   diversity at each locus.

213

214   **Statistical analysis**

215   All statistical analyses were performed using the R.3.5.0 software (R Core Team, 2018).

216

217   *Phenotypic traits*

218   First, to test for phenotypic differences between subspecies, hierarchical generalized linear

219   models were conducted with population nested in subspecies. Second, for each subspecies

220   linear mixed models were conducted to test for phenotypic differences among populations, with

221   population as a fixed effect and the plastic container as a random effect. Estimates of marginal

222   means for each trait in each population were extracted using the emmeans package (Lenth et

223   al., 2019). These linear mixed-effects models were implemented in R via the lme4 package

224   (Bates et al., 2015). Traits change with altitude was analysed by using a linear regression of the

225   marginal means by altitude. Finally, mean phenotypic traits were also generated, and provided

226   in the Supplementary materials (Fig S1).

227

228   *Calculation of h^2 and phenotypic differentiation indices (Q_{ST}).*

229   For each subspecies, narrow-sense heritabilities (h^2) were estimated for each phenotypic trait

230   across all populations using a model with population, family and plastic containers as random

231   factors as $h^2 = 2 V_w / (V_w + V_{res})$, with V_w is family variance component and V_{res} is the residual

232   variance component corresponding to the within-population variance component. We

233   multiplied V_w by two in the calculation of h^2 because we used a full-sib crossing design (Roff,

234   1997). Caution must be taken when considering h^2 values because its estimation can be biased

235   by the estimation of $2V_w$. Indeed, h^2 was calculated based on all the families, without

236   considering the differences of h^2 between different populations. Confidence intervals of h^2 were

237   calculated following a parametric bootstrap method adapted from O'Hara and Merilä (2005).

238

239   For each trait and each subspecies, quantitative trait divergence indices (Q_{ST}) were generated

240   among populations (overall Q_{ST}) and for each population pair (population pairwise Q_{ST}), based

241   on mixed model analyses. In these models, population, family and plastic containers were

242   random factors. Variance components were extracted from these analyses for each trait and

243   used for estimating Q_{ST} using the following formula (Spitze, 1993): $Q_{ST} = V_b / (V_b + 2\, h^2\, V_w)$

244   with V_b being the trait genetic variance among populations. h^2 was calculated based on all the

245   families and populations by subspecies. Here, no environmental sources of phenotypic variance

246   due to the ecological conditions of the location of origin of populations could in theory bias Q_{ST}

247   estimates because data was obtained from a common garden experiment (Pujol, Wilson, et al.,

248   2008). When a variance component was non-significant, it was considered as null in further

249   calculations. When necessary (as for population pairwise Q_{ST} calculation), data was linearized

250   by using a square root transformation. All variance components were estimated by using the

251   linear mixed model approach implemented in the R package lme4 v. 1.1.17 (Bates et al., 2015).

252   Confidence intervals of Q_{ST} values were calculated following a parametric bootstrap method

253   adapted from O'Hara and Merilä (2005).

254

255   *Overall Q_{ST}-F_{ST} comparisons.*

256   We compared overall Q_{ST} and F_{ST} for each trait to investigate if divergence was compatible

257   with a scenario of genetic drift (overall $Q_{ST} = F_{ST}$), or whether it was more likely explained by

258   directional selection (overall $Q_{ST} > F_{ST}$) or by stabilizing selection (overall $Q_{ST} < F_{ST}$).

259   Comparisons between overall Q_{ST} and F_{ST} values were performed for each trait based on two

260   methods: i) a comparison of confidence intervals, the Q_{ST} is considered non significantly

261   different from neutral differentiation when the confidence interval of the overall Q_{ST} for a trait

262   overlaps the mean F_{ST} value, ii) a bootstrapping method developed by Whitlock and Guillaume

263   (2009). This last approach aims at comparing the observed difference between the overall Q_{ST}

264   and the F_{ST} with the expected simulated distribution of this difference under a scenario of

265   neutral evolution. We generated 100 000 bootstrap replicates of the expected Q_{ST} - F_{ST}

266   difference under the neutrality hypothesis for each trait, and built the corresponding

267   distribution. In this approach, P values were estimated by assessing whether the observed value

268   of the Q_{ST} - F_{ST} difference overlapped its expected distribution under neutrality. We used the

269   modification by Lind *et al.* (2011) of the approach of Whitlock and Guillaume (2009) to

270   estimate the variance components of the simulated values of the Q_{ST} - F_{ST} difference.

271

272   *Mantel tests*

273   Mantel tests (Mantel, 1967) were used to analyse correlations between geographic distances,

274   environmental distances (altitudinal), neutral genetic differentiation (population pairwise F_{ST}),

275   and quantitative genetic differentiation (population pairwise Q_{ST}). First, a correlation test

276   between population pairwise F_{ST} and population pairwise geographic distance matrices was

277   performed to test for an isolation by distance relationship. Second, a correlation test between

278   population pairwise F_{ST} and population pairwise Q_{ST} was performed for each trait to test if

279   neutral genetic differentiation explained divergence in quantitative traits. Third, a correlation

280   test between population pairwise Q_{ST} and population pairwise altitudinal differences was

281   performed for each trait to test whether divergence in quantitative traits was related to altitudinal

282   differences. Finally, we conducted partial mantel tests to test for the association between

283   population pairwise Q_{ST} and population pairwise altitude differences while controlling for

284   neutral genetic differentiation (F_{ST}). All mantel and partial mantel tests were performed in R,

285   with a significance α=0.05, using the vegan package (Oksanen et al., 2009).

286 **RESULTS**

287 **Phenotypic differentiation between subspecies and populations**

288 The two subspecies - *A. m. pseudomajus* and *A. m. striatum* - were significantly differentiated

289 by several phenotypic traits (Table 1 a, Figure S2). When grown in a common garden, plants

290 of *A. m. pseudomajus* were on average taller, with more branches and nodes than plants from

291 *A. m striatum*. Their average germination date, internode length and SLA were however similar.

292 Differentiation between subspecies (c. 2.8 %) was lower than the differentiation estimated

293 among populations (c. 9.3 %, see mean R^2 in Table 1 a). Most of the traits showed a phenotypic

294 divergence among populations within each subspecies (see LRT in Table 1 b). Germination

295 date was the only trait that showed no significant difference among populations of *A. m.*

296 *pseudomajus* (see LRT in Table 1 b).

297

298 **Neutral genetic differentiation**

299 Population neutral genetic differentiation was low but significant. Overall F_{ST} among

300 populations of *A. m. pseudomajus* was 0.109 (P < 0.001), and ranged from 0.06 to 0.159 across

301 population pairs (see Table S2, and see Pujol et al., 2017 for more details on population pairwise

302 neutral genetic differentiation). F_{ST} among populations of *A. m. striatum* was 0.097 (P < 0.001),

303 and ranged from 0.055 to 0.131 (Table S2). There was no significant relationship between

304 population pairwise F_{ST} and population pairwise geographic distance for both subspecies

305 (Figure 2 a and b, *A. m. pseudomajus* Mantel r = 0.04, P = 0.405, *A. m. striatum* Mantel r = -

306 0.18, P = 0.595). Similarly, there was no significant relationship between population pairwise

307 F_{ST} and population pairwise altitude difference for both subspecies (Figure 2 c), although the

308 Mantel tests showed a relationship close to significance levels in *A. m. pseudomajus* (*A. m.*

309 *pseudomajus* Mantel r = 0.23, P = 0.052, *A. m. striatum* Mantel r = -0.3, P = 0.943).

310

13

Changes in phenotypic traits with altitude

We found significant correlations between trait values (i.e. population estimates of marginal means) and altitude for two traits across *A. m. striatum* populations. Plants from populations at low altitude had more nodes and branches than plants from populations at high altitude of *A. m. striatum* (Figure 3, see population arithmetic means in Figure S2 and population estimates of marginal means for other traits in Figure S3). No phenotypic changes associated with altitude were found significant in *A. m. pseudomajus*.

Inheritance of quantitative traits

Heritability estimates were comprised between 0.07 to 0.58 for *A. m. pseudomajus*, and 0.01 to 0.56 for *A. m. striatum* (Table S3, Supporting Information). The highest heritability estimates were found for the internode length in *A. m. pseudomajus* (0.58) and the number of nodes in *A. m. striatum* (0.56). Several traits had close to similar heritabilities between subspecies (stem diameter, number of nodes), as illustrated by their overlapping confidence intervals. However, other traits appeared to be different, with no CI overlap (germination date, number of branches, plant height, internode length, SLA, Table S3).

Q_{ST}-F_{ST} comparisons.

Overall Q_{ST} was higher than mean F_{ST} for one trait in *A. m. pseudomajus* (number of branches, Figure 4 a), and for three traits in *A. m. striatum* (number of branches, plant height and internode length, Figure 4 b), as illustrated by their non-overlapping confidence intervals. Overall Q_{ST} was lower than mean F_{ST} for the germination date in *A. m. pseudomajus* (Figure 4 a). We also rejected the hypothesis of neutrality for these two traits in *A. m. pseudomajus* (number of branches, germination date), and for three traits in *A. m. striatum* (number of branches, plant height and internode length) on the basis of the bootstrapping method developed by Whitlock

14

336   and Guillaume (2009). For these traits, observed values of overall Q_{ST} - F_{ST} differences were

337   either in the tail of the expected probability distribution under the hypothesis of neutrality, or

338   did not overlap with this distribution (Figure S4 et S5).

339   Mantel tests showed that population pairwise quantitative genetic differentiation (Q_{ST}) was not

340   correlated with population pairwise neutral genetic differentiation (F_{ST}) for all traits excepted

341   the germination date in *A. m. striatum* (Table 2). Population pairwise Q_{ST} for the germination

342   date was significantly correlated with population pairwise F_{ST}. They showed no support for the

343   overall $Q_{ST} < F_{ST}$ for the germination date in *A. m. pseudomajus*, almost certainly because of

344   three comparisons diverging abnormally from the overall pattern (Figure 5 b).

345

346   **Increased quantitative genetic differentiation with altitude difference**

347   Mantel tests showed a significant correlation between population pairwise Q_{ST} and population

348   pairwise altitudinal difference for two traits in *A. m. striatum*: the number of nodes and the

349   number of branches (Table 2, Figure 5). For both traits, the increase in pairwise population

350   differentiation associated with an increase in altitudinal difference was higher for the Q_{ST} than

351   for the F_{ST} (Figures 5 c and e). Partial mantel tests showed that population pairwise Q_{ST} was

352   significantly correlated with differences in altitude for the number of nodes (and marginally

353   significant for the number of branches, see Q_{ST} vs Alt. diff. / F_{ST} in Table 2) while controlling

354   for neutral genetic differentiation (F_{ST} matrix). This result is expected under the hypothesis that

355   the divergence among populations of *A. m. striatum* in the number of nodes is a result of

356   altitude-mediated divergent selection. In contrast, none of the seven traits showed a significant

357   correlation between population pairwise Q_{ST} and population pairwise altitude difference in *A.*

358   *m. pseudomajus*.

359

DISCUSSION

Our results support the hypothesis of differential adaptation between *A. m. pseudomajus* and *A. m. striatum* subspecies. We detected phenotypic differentiation in a common garden among the populations of *A. m pseudomajus*, among the populations of *A. m stratum*, and among subspecies. For both subspecies, local adaptation and neutral evolution explained the extent to which populations diverged over their geographic range, with slight differences between subspecies. Potential divergence along altitude was also detected, but only for one subspecies: *A. m. striatum.*

Our findings comforted the idea that Q_{ST}- F_{ST} comparisons are a good first step for exploring the potential roles of divergent natural selection and neutral evolutionary processes in phenotypic divergence (Whitlock, 2008; Edelaar et al., 2011; Ovaskainen et al., 2011; Whitlock and Gilbert, 2012). They highlighted how traits can be used to identify the potential ecological pressures underlying natural selection, with some traits potentially involved with *A. majus* adaptation to the conditions of populations' local sites of origin, and a subsample of these traits potentially playing a role in *A. m. striatum* adaptation to altitude.

Adaptive evolution of *A. m. striatum* populations along the altitudinal gradient

Our results brought indirect evidence supporting potential adaptive divergence between *A. m. pseudomajus* and *A. m. striatum*. They imply that the quantitative genetic basis of two of the seven traits under study (number of nodes, and marginally significant for the number of branches) was shaped by divergent selection between populations from different altitudes in *A. m. striatum* but not in *A. m. pseudomajus*. Most studies on plant adaptation to altitude report the selection of smaller plants at higher altitudes (Körner, 1999; Halbritter et al., 2018). In agreement with this expectation, we found that *A. m. striatum* plants at higher altitudes had less

16

385   branches and less nodes. There is also evidence for changes in leaf traits with elevation (Read

386   et al., 2014; Halbritter et al., 2018), with a decrease of SLA with elevation gradients. Our results

387   did not support a potential scenario of selection based on SLA at play in *A. m. striatum*.

388

389   **Support for different subspecies scenarios of adaptation to local sites of origin**

390   Our results showed that quantitative genetic differentiation was higher than what could be

391   explained by neutral evolutionary divergence among *A. m. pseudomajus* populations for one of

392   the seven studied traits (number of branches), and among *A. m. striatum* populations for three

393   of the seven traits (number of branches, plant height and internode length). They imply that

394   adaptation to local sites of origin potentially shaped the phenotypic diversity of populations for

395   both subspecies across their geographic range, with potentially different intensities reflected by

396   different signatures between subspecies. We used classical overall Q_{ST} - F_{ST} comparisons to

397   detect potential adaptation to local sites conditions (Leinonen et al., 2008) and also more recent

398   methods to insure that our findings were robust against a range of neutral evolution scenarios

399   for these traits (Whitlock, 2008). Furthermore, our approach minimized the possibility that

400   phenotypic differences between populations were generated by environmental effects by using

401   a common garden experiment, and including trait heritability estimates in Q_{ST} calculations

402   (Spitze, 1993; Pujol, Wilson, et al., 2008). In contrast, four of the seven studied traits

403   (germination date, diameter, number of nodes and SLA) did not show departure from plausible

404   baseline scenarios of neutral evolutionary divergence, using overall Q_{ST} - F_{ST} comparisons. One

405   particular trait (germination date) was in fact more similar among populations than expected

406   under neutrality in *A. m. pseudomajus*. A scenario of stabilizing selection is classically

407   extrapolated in the case of similar results (Lamy et al., 2012) but another plausible explanation

408   is that population similarity might have been caused by convergent phenotypic responses to the

409   common garden environmental similarity. Caution must be taken when interpreting different

410   Q_{ST}- F_{ST} patterns between subspecies as the signature of different adaptive processes. Here we

411   found different patterns between subspecies, which supports the hypothesis of their potential

412   adaptive divergence. Our results cannot be interpreted as direct proof for their adaptive

413   divergence, but only as evidence that this hypothesis has some potential.

414

415   **The ecological significance of adaptation to local sites of origin in _A. majus_**

416   In the absence of environmental measures included in the overall Q_{ST}-F_{ST} analysis, it is

417   impossible to identify the potential environmental agents of local selection that shape the

418   quantitative genetic variation of traits. The functions behind the traits that have diverged can

419   nevertheless be used to discuss plausible evolutionary scenarios of natural selection. Our results

420   imply that adaptation to local sites of origin has potentially shaped the vegetative architecture

421   of plants that is specific to each _A. majus_ population. The quantitative genetic variation of

422   several phenotypic traits characterising the vegetative growth and development of plants (plant

423   height, internode length, number of branches) has likely diverged among populations as a result

424   of adaptation to local sites of origin. Divergence in the genetic variation underlying the shape

425   and size of plants was already found at the level of _Antirrhinum_ species but its adaptive

426   significance was not tested for (Langlade et al., 2005). In southern France and northern Spain,

427   under the Mediterranean climate, dryer locations are expected to select for plants with a bushier

428   vegetative architecture, i.e. plants with smaller leaves and more branches that have a better

429   water use efficiency and resilience to drought stress (Langlade et al., 2005). It is difficult to

430   identify exactly which environmental pressures underlay selection at local sites because several

431   combinations of environmental parameters (vegetation cover, wind, disturbance, temperature,

432   water availability, etc.) can interact to affect phenotypic traits.

433

434   **Gene flow, ecological and reproductive isolation**

18

435   Our findings imply that the most likely evolutionary scenario applying to *A. majus* requires

436   invoking a history of adaptation to local sites in a complex background of gene flow, ecological

437   heterogeneity and reproductive isolation. Pyrenees mountains are widely acknowledged to

438   constitute a heterogeneous landscape promoting complex patterns of population connectivity

439   and prone to generate local adaptation (Alberto et al., 2010). Q_{ST}-F_{ST} comparisons reflected a

440   potential scenario of population divergent adaptation to contrasting environmental conditions

441   between their local sites of origins. Our findings also suggested that evolutionary signatures of

442   local adaptation differed between *A. m. pseudomajus* and *A. m striatum*, which includes the

443   potential adaptation to altitude of *A.m. striatum* populations. One might speculate that this

444   divergence might be related to the distribution of *A. m. striatum* populations across a narrower

445   range of climatic conditions, even if both subspecies share to a large extent the same ecological

446   niche (Khimoun et al., 2013). However, caution must be taken with this explanation because

447   the state of the environment in the past, when divergence might have occurred, is unknown and

448   might have differed. Contrasting hypotheses might be interesting to consider, e.g., different

449   evolutionary potentials in the presence of similar environmental pressures. These scenarios are

450   not exclusive and can reinforce each other through a feedback loop between reproductive

451   isolation, neutral divergence and selection.

452

453   Restricted gene flow or strong selection pressures are required for evolutionary divergence.

454   Genetic drift, or foundation events by different gene pools, might have shaped differentially the

455   genetic background of *A. majus* populations and to some extent subspecies at the scale of their

456   global geographic range. There is evidence for the genetic signature of restricted gene

457   exchanges in *A. majus* (Pujol et al., 2017). No genetic isolation by distance was found but

458   ecological barriers characterizing the mountain landscape of the Pyrenees likely participate to

459   isolate populations (Pujol et al., 2017). At first sight, *A. majus* subspecies divergence might not

19

460 be expected because both subspecies are interfertile (Andalo et al., 2010), and no genome wide

461 barrier to gene flow was found between them at the scale of a hybrid zone across c. 2km in the

462 Pyrenees (Ringbauer et al., 2018). There is also evidence for gene exchanges between the two

463 subspecies in several contact zone locations across at the periphery of their geographic ranges

464 (Khimoun et al., 2011). Yet, subspecies flower color differences attest that flower color genes

465 are under frequency dependent selection and generate reproductive isolation between

466 subspecies (Tastard et al., 2012; Ringbauer et al., 2018). This reproductive isolation might

467 participate to the subspecies phenotypic divergence of other traits that we detected here.

468

469 **CONCLUSION**

470 Our findings corroborate the utility of Q_{ST} - F_{ST} approaches conducted in common garden

471 experiments to explore potential adaptive evolutionary divergence among populations and

472 between subspecies in plants. They also illustrate the limit of this approach that identifies traits

473 that might be involved with local adaptation but does not bring direct evidence for their

474 response to selection. Here, our common garden results for *A. m. pseudomajus* and A. *m.*

475 *striatum* populations identified vegetative traits that might play a role in the local adaptation

476 and the differential adaptation of *A. m. pseudomajus* and *A. m. striatum* along altitudinal

477 gradients. They suggest that the adaptation to climate variables of otherwise interfertile

478 subspecies might differ as a result of reproductive isolation.

479

489

490 TABLE 1

491 Effects of subspecies and populations on phenotypic traits. a) R² and P-value from hierarchical

492 generalized linear models (GLM) with subspecies alone and populations nested in subspecies,

493 implemented in JMP® (Version X ,SAS Institute Inc., Cary, NC, 1989-2019). b) Likelihood

494 Ratio Tests (LRT) comparing the maximum-likelihood fit between a model where populations

495 were pooled and a model estimating the effect of the population of origin. A significant P- value

496 means the model including populations effect fitted the data better than the null model.

497 Significant results (P-value < 0.05) are in bold.

a)	Subspecies		Populations in subspecies	
	R^2	P-value	R^2	P-value
Germination date	0.0005	0.587	0.02	0.260
Diameter	**0.007**	**0.00028**	**0.05**	**0.00068**
Nodes	**0.06**	**<0.0001**	**0.13**	**<0.0001**
Branches	**0.03**	**<0.0001**	**0.06**	**0.00001**
Plant height	**0.09**	**<0.0001**	**0.21**	**<0.0001**
Internode length	0.0002	0.708	**0.13**	**0**
SLA	0.005	0.066	**0.05**	**0.0003**
Mean	0.028		**0.093**	

b)	*A. m. pseudmomajus*		*A. m. striatum*	
	LRT	P-value	LRT	P-value
Germination date	6	0.570	12	**0.021**
Diameter	18	**0.001**	23	**<0.0001**
Nodes	30	**<0.0001**	20	**<0.0001**
Branches	70	**<0.0001**	26	**<0.0001**
Plant height	32	**<0.0001**	81	**<0.0001**
Internode length	64	**<0.0001**	37	**<0.0001**
SLA	21	**0.004**	15	**0.004**

498

499

500 TABLE 2

501 Mantel tests and partial Mantel tests on pairwise Q_{ST} vs F_{ST} and Q_{ST} vs difference in altitude of

502 origin (Alt. diff.), as well as partial Mantel tests on Q_{ST} vs Alt. diff. controlled for F_{ST}, for

503 phenology traits in a) eight populations of *A. m. pseudomajus* and b) five populations of *A. m.*

504 *striatum*, that were grown in a common garden. Significant values are indicated in bold.

a)	*A majus pseudomajus*					
	Q_{ST} vs F_{ST}		Q_{ST} vs Alt. diff.		Q_{ST} vs Alt. diff. / F_{ST}	
Traits	Mantel r	P-value	Mantel r	P-value	Mantel r	P-value
Germination date	-0.37	0.931	-0.13	0.737	-0.06	0.593
Diameter	-0.09	0.636	-0.15	0.812	-0.13	0.780
Nodes	0.07	0.426	-0.16	0.820	-0.18	0.888
Branches	0.07	0.329	-0.14	0.750	-0.17	0.820
Height	0.23	0.191	-0.13	0.751	-0.19	0.911
Internode length	0.24	0.184	0.03	0.335	-0.02	0.442
SLA	0.23	0.229	0.02	0.379	-0.04	0.529

| b)	*A majus striatum*					
	Q_{ST} vs F_{ST}		Q_{ST} vs Alt. diff.		Q_{ST} vs Alt. diff. / F_{ST}	
	Mantel r	P-value	Mantel r	P-value	Mantel r	P-value
Germination date	**0.53**	**0.042**	0.05	0.333	0.26	0.267
Diameter	-0.01	0.508	-0.02	0.458	-0.03	0.517
Nodes	-0.3	0.842	**0.95**	**0.008**	**0.94**	**0.008**
Branches	-0.18	0.750	**0.89**	**0.033**	0.90	0.058
Height	-0.58	0.883	0.08	0.283	-0.12	0.6
Internode length	0.34	0.267	-0.02	0.492	0.09	0.258
SLA	-0.69	0.883	0.07	0.367	-0.2	0.858

505

506

507　**FIGURE**
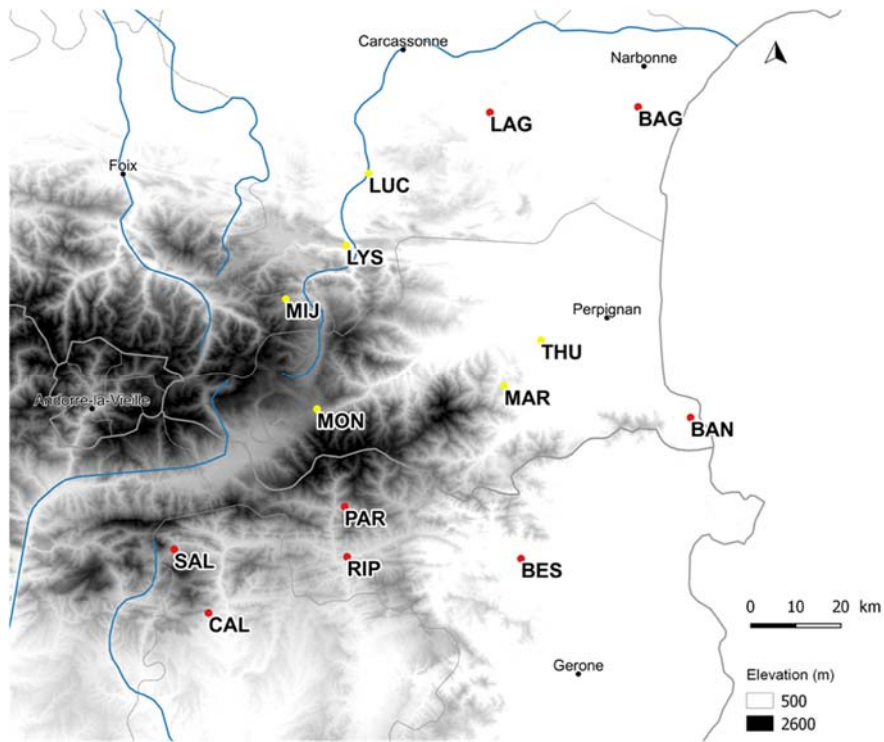
508

509　FIGURE 1

510　Map of *Antirrhinum majus* populations that were sampled across the geographic range of the

511　species in Southern France. Red dots represent *A. m. pseudomajus* populations, yellow dots

512　represent *A. m striatum* populations. Population names and description can be found in Table

513　S1.
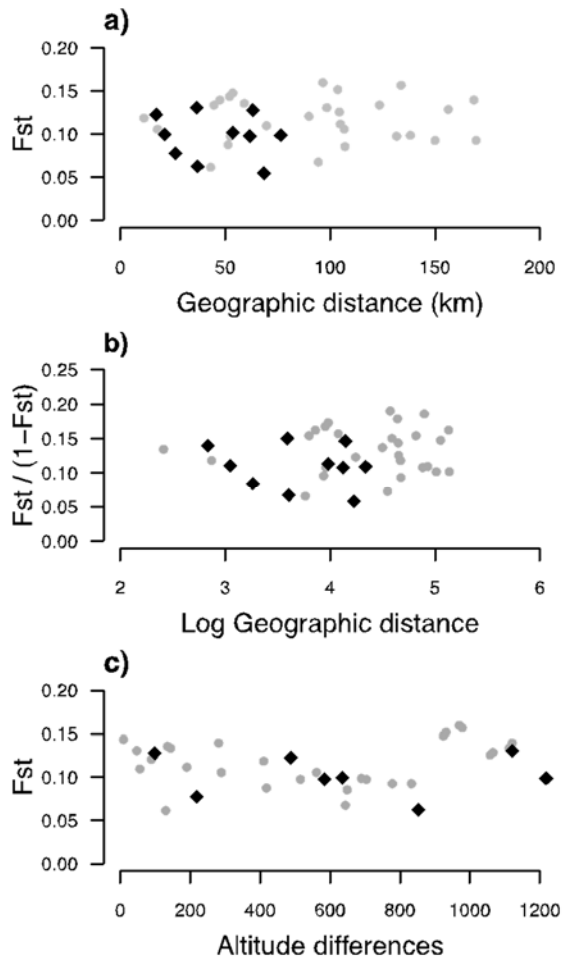
514

515 FIGURE 2

516 Pairwise neutral genetic differentiation F_{ST} plotted against pairwise geographic distances or

517 altitudinal differences among eight *Antirrhinum majus pseudomajus* populations pairs (grey

518 dots), and five *A. m. striatum* populations pairs (black diamonds). There were non-significant

519 linear regression between a) F_{ST} and geographic distance in *A. m. pseudomajus* (y= 1e-05 x +

520 0.116, P-value = 0.92 ns) and in *A. m striatum* (y= -0.00019 x + 0.106, P = 0.67 ns), b) F_{ST} / (1-

521 F_{ST}) and the log of geographic distance in *A. m. pseudomajus* (y= 0.0022 x + 0.124, P = 0.83

522 ns) and in *A. m striatum* (y= -0.0105 x + 0.148, P = 0.63 ns), c) F_{ST} and altitude differences in

523 *A. m. pseudomajus* (y= 2.e-05 x + 0.109, P = 0.26 ns) and in *A. m striatum* (y= -2e-05 x + 0.111,

524 P = 0.39 ns).

525

25

a)

b)

y= −0.00214 x + 13 (p−value = 0.001 **)

y= −0.00316 x + 17 (p−value = 0.008 **)
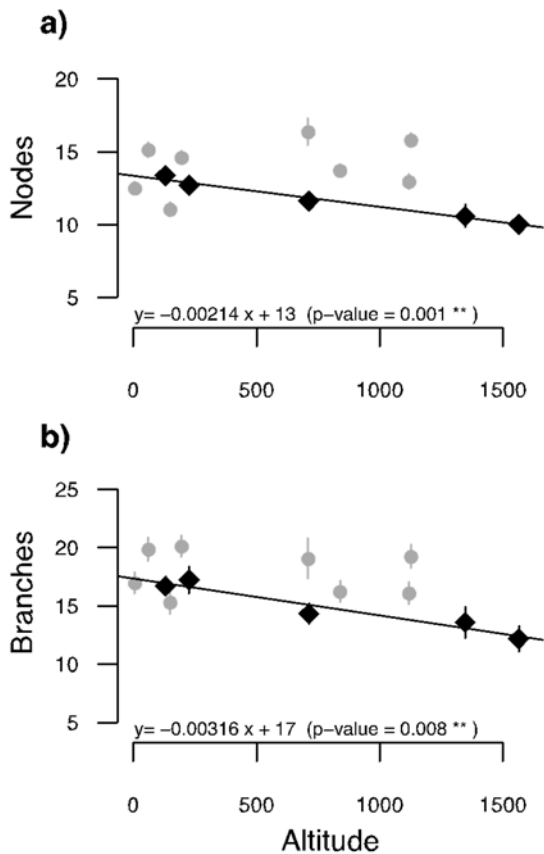
526

527   FIGURE 3

528   Population estimates of marginal means with standard errors of two phenotypic traits (a. number

529   of nodes, b. number of branches) in populations of two subspecies of *Antirrhinum majus* grown

530   in a common garden. Means are plotted against altitude of origin. Lines refer to the linear

531   regression between trait mean estimates and altitude. Grey dots represent *A. m. pseudomajus*

532   populations, black diamonds represent *A. m. striatum* populations. Equation of non-significant

533   linear regressions were a) y= 0.00125 x +12 (P = 0.43 ns) and b), y= -0.00048 x +18 (P = 0.78

534   ns) for *A. m. pseudomajus*.

26
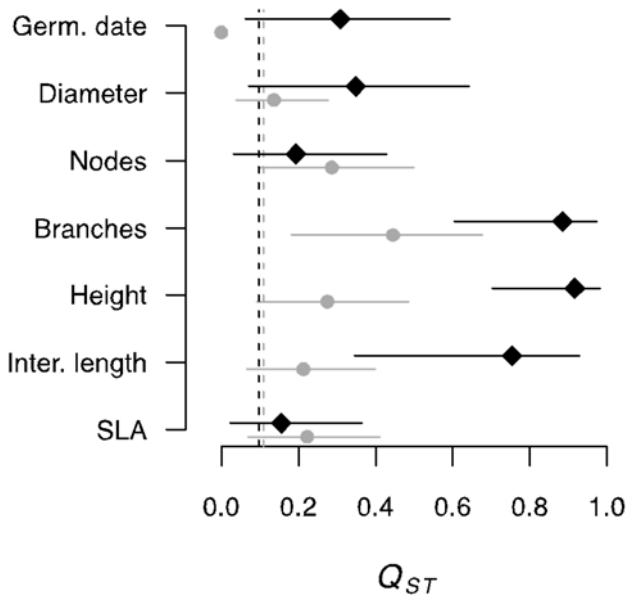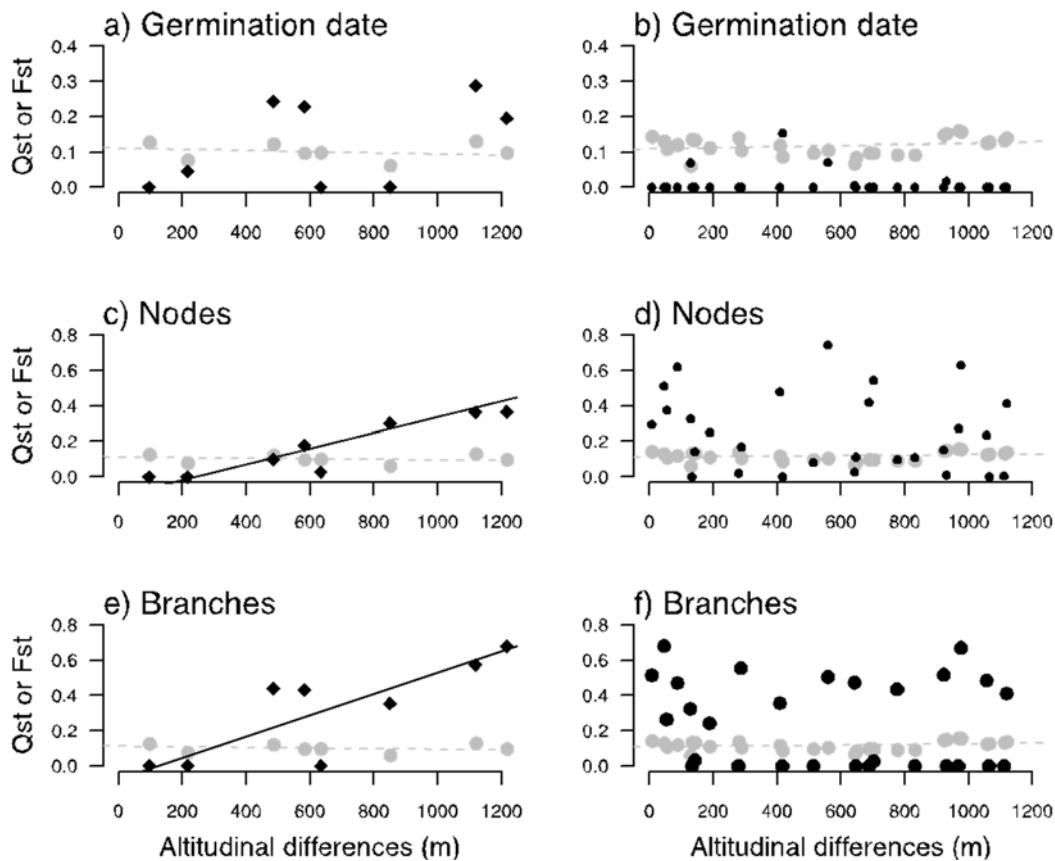
FIGURE 4

Overall Q_{ST} estimates with their 95% CI characterizing seven phenotypic traits in eight *Antirrhinum majus pseudomajus* populations (grey dots) and five *A. m. striatum* populations (black diamonds) that were grown in a common garden. Average population F_{ST} is represented by the dashed grey line for *A. m. pseudomajus*, and the dashed black line for *A. m. striatum*. Germ.date = germination date, Diameter= stem diameter, Nodes = number of nodes, Branches = number of branches, Height= plant height, Inter. Length= internodes length, SLA= specific leaf area.

546

547   FIGURE 5

548   Population pairwise quantitative trait differentiation (Q_{ST}) for the germination date, the number

549   of branches and the number of nodes in *Antirrhinum majus striatum* (a, c and e, black diamonds)

550   and *A. m. pseudomajus* (b, d and f, black dots). Dashed line indicates the trend and *P* the level

551   of significance of the linear regression between the population pairwise Q_{ST} and population

552   pairwise altitudinal differences (m). Grey dots and dashed line refer to population neutral

553   genetic differentiation (F_{ST}). Equation of linear regressions were a) y= 1e-05 x +0.09 (P = 0.9

554   ns), b) 1e-05 x +0.018 (P = 0.5 ns), c) y=0.00044 x - 0.107 (P = 0 ***), d) y= -9e-05 x + 0.303

555   (P =0.42 ns), e) y= 6e-04 x – 0.076 (P =5e-04***), f) y=-1e-04 x+0.302 (P = 0.46 ns).

556

557

558

Reference List

Abbott, R. J., and A. C. Brennan. 2014. Altitudinal gradients, plant hybrid zones and evolutionary novelty. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 20130346.

Alberto, F., J. Niort, J. Derory, O. Lepais, R. Vitalis, D. Galop, and A. Kremer. 2010. Population differentiation of sessile oak at the altitudinal front of migration in the French Pyrenees: genetic diversity of oaks in the Pyrenees. *Molecular Ecology* 19: 2626–2639.

Andalo, C., M. B. Cruzan, C. Cazettes, B. Pujol, M. Burrus, and C. Thébaud. 2010. Post-pollination barriers do not explain the persistence of two distinct *Antirrhinum* subspecies with parapatric distribution. *Plant Systematics and Evolution* 286: 223–234.

Angert, A. L., and D. W. Schemske. 2005. The evolution of species' distributions: reciprocal transplants across the elevation ranges of *Mimulus cardinalis* and *M. lewisii*. *Evolution* 59: 1671–1684.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software* 67.

Bonhomme, V., S. Picq, C. Gaucherel, and J. Claude. 2014. **Momocs**: outline analysis using *R*. *Journal of Statistical Software* 56.

Debout, G. D. G., E. Lhuillier, P.-J. G. Malé, B. Pujol, and C. Thébaud. 2012. Development and characterization of 24 polymorphic microsatellite loci in two *Antirrhinum majus* subspecies (Plantaginaceae) using pyrosequencing technology. *Conservation Genetics Resources* 4: 75–79.

Donohue, K., R. Rubio de Casas, L. Burghardt, K. Kovach, and C. G. Willis. 2010. Germination, postgermination adaptation, and species ecological ranges. *Annual Review of Ecology, Evolution, and Systematics* 41: 293–319.

Edelaar, P., P. Burraco, and I. Gomez-Mestre. 2011. Comparisons between Q_{ST} and F_{ST}-how wrong have we been? *Molecular Ecology* 20: 4830–4839.

Etterson, J. R. 2004. Evolutionary potential of *Chamaecrista fasciculata* in relation to climate change. II. Genetic architecture of three populations reciprocally planted along an environmental gradient in the great plains. *Evolution* 58: 1459–1471.

Gimenez-Benavides, L., A. Escudero, and J. M. Iriondo. 2006. Local adaptation enhances seedling recruitment along an altitudinal gradient in a high mountain mediterranean plant. *Annals of Botany* 99: 723–734.

Halbritter, A. H., S. Fior, I. Keller, R. Billeter, P. J. Edwards, R. Holderegger, S. Karrenberg,

et al. 2018. Trait differentiation and adaptation of plants along elevation gradients. *Journal of Evolutionary Biology* 31: 784–800.

Hereford, J. 2009. A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist* 173: 579–588.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. *Ecology Letters* 7: 1225–1241.

Khimoun, A., M. Burrus, C. Andalo, Z.-L. Liu, C. Vicédo-Cazettes, C. Thébaud, and B. Pujol. 2011. Locally asymmetric introgressions between subspecies suggest circular range expansion at the *Antirrhinum majus* global scale. *Journal of Evolutionary Biology* 24: 1433–1441.

Khimoun, A., J. Cornuault, M. Burrus, B. Pujol, C. Thebaud, and C. Andalo. 2013. Ecology predicts parapatric distributions in two closely related *Antirrhinum majus* subspecies. *Evolutionary Ecology* 27: 51–64.

Kim, E., and K. Donohue. 2013. Local adaptation and plasticity of *Erysimum capitatum* to altitude: its implications for responses to climate change. *Journal of Ecology* 101: 796–805.

Körner, C. 1999. Alpine plant life. Springer Berlin Heidelberg, Berlin, Heidelberg.

Lamy, J.-B., C. Plomion, A. Kremer, and S. Delzon. 2012. $Q_{ST} < F_{ST}$ as a signature of canalization. *Molecular Ecology* 21: 5646–5655.

Langlade, N. B., X. Feng, T. Dransfield, L. Copsey, A. I. Hanna, C. Thebaud, A. Bangham, et al. 2005. Evolution through genetically controlled allometry space. *Proceedings of the National Academy of Sciences* 102: 10221–10226.

Leimu, R., and M. Fischer. 2008. A meta-analysis of local adaptation in plants. *PLoS ONE* 3: e4010.

Leinonen, T., R. J. S. McCairns, R. B. O'Hara, and J. Merilä. 2013. Q_{ST} - F_{ST} comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics* 14: 179–190.

Leinonen, T., R. B. O'Hara, J. M. Cano, and J. Merilä. 2008. Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis: Q_{ST} - F_{ST} meta-analysis. *Journal of Evolutionary Biology* 21: 1–17.

Lenth, R., H. Singmann, J. Love, P. Buerkner, and M. Herve. 2019. Package 'emmeans': Estimated marginal means, aka least-squares means. Version 1.4.3.01. *CRAN*.

Lind, M. I., P. K. Ingvarsson, H. Johansson, D. Hall, and F. Johansson. 2011. Gene flow and selection on phenotypic plasticity in an island system of *Rana temporaria*. *Evolution* 65: 684–697.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach.

Cancer Research 27: 209–220.

McKay, J. K., and R. G. Latta. 2002. Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution* 17: 285–291.

Meirmans, P.G., and P.H. Van Tienderen. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms, *Molecular Ecology Notes* 4: 792-794.

Merilä, J., and P. Crnokrak. 2001. Comparison of genetic differentiation at marker loci and quantitative traits: Natural selection and genetic differentiation. *Journal of Evolutionary Biology* 14: 892–903.

O'Hara, R. B., and J. Merilä. 2005. Bias and precision in Q_{ST} estimates: problems and some solutions. *Genetics* 171: 1331–1339.

Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, G. Simpson, P. Solymos, M. Stevens, and H. Wagner. 2009. The VEGAN Package: community ecology package.

Ovaskainen, O., M. Karhunen, C. Zheng, J. M. C. Arias, and J. Merilä. 2011. A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. *Genetics* 189: 621–632.

Pérez-Harguindeguy, N., S. Díaz, E. Garnier, S. Lavorel, H. Poorter, P. Jaureguiberry, M. S. Bret-Harte, et al. 2016. Corrigendum to: New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany* 64: 715.

Pujol, B., J. Archambeau, A. Bontemps, M. Lascoste, S. Marin, and A. Meunier. 2017. Mountain landscape connectivity and subspecies appurtenance shape genetic differentiation in natural plant populations of the snapdragon (*Antirrhinum majus* L.). *Botany Letters* 164: 111–119.

Pujol, B., J.-L. Salager, M. Beltran, S. Bousquet, and D. McKey. 2008. Photosynthesis and leaf structure in domesticated Cassava (Euphorbiaceae) and a close wild relative: have leaf photosynthetic parameters evolved under domestication? *Biotropica* 40: 305–312.

Pujol, B., A. J. Wilson, R. I. C. Ross, and J. R. Pannell. 2008. Are Q_{ST} - F_{ST} comparisons for natural populations meaningful? *Molecular Ecology* 17: 4782–4785.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Read, Q. D., L. C. Moorhead, N. G. Swenson, J. K. Bailey, and N. J. Sanders. 2014. Convergent effects of elevation on functional leaf traits within and among species. *Functional Ecology* 28: 37–45.

Ringbauer, H., A. Kolesnikov, D. L. Field, and N. H. Barton. 2018. Estimating barriers to gene flow from distorted isolation-by-distance patterns. *Genetics* 208: 1231–1245.

660   Roff, D. A. 1997. Evolutionary Quantitative Genetics. Springer US, Boston, MA.

661   Spitze, K. 1993. Population structure in *Daphnia obtusa*: quantitative genetic and allozymic

662   variation. *Genetics* 135: 367–374.

663   Tastard, E., J.-B. Ferdy, M. Burrus, C. Thébaud, and C. Andalo. 2012. Patterns of floral colour

664   neighbourhood and their effects on female reproductive success in an *Antirrhinum* hybrid zone:

665   flower colour and selection in a hybrid zone. *Journal of Evolutionary Biology* 25: 388–399.

666   Tavares, H., A. Whibley, D. L. Field, D. Bradley, M. Couchman, L. Copsey, J. Elleouet, et al.

667   2018. Selection and gene flow shape genomic islands that control floral guides. *Proceedings of*

668   *the National Academy of Sciences* 115: 11006–11011.

669   Whibley, A. C. 2006. Evolutionary paths underlying flower color variation in Antirrhinum.

670   *Science* 313: 963–966.

671   Whitlock, M. C. 2008. Evolutionary inference from Q_{ST}. *Molecular Ecology* 17: 1885–1896.

672   Whitlock, M. C., and K. J. Gilbert. 2012. Q_{ST} in a hierarchically structured population.

673   *Molecular Ecology Resources* 12: 481–483.

674   Whitlock, M. C., and F. Guillaume. 2009. Testing for spatially divergent selection: comparing

675   Q_{ST} to F_{ST}. *Genetics* 183: 1055–1063.

676   Wright, I. J., P. B. Reich, M. Westoby, D. D. Ackerly, Z. Baruch, F. Bongers, J. Cavender-

677   Bares, et al. 2004. The worldwide leaf economics spectrum. *Nature* 428: 821–827.

678

679

Supplementary material

TABLES

682 TABLE S1 Description of *Anthirinum majus* populations grown in the common garden

683 experiment. *Nfam*= number of families, *N*= number of plants

Acronym	Latitude	Longitude	Location	Elevation (m)	Subspecies	Description	*Nfam*	*N*
BAG	43.10	2.98	Bages	6	*pseudomajus*	Dunes on seaside (rocky / herbaceous)	40	67
BAN	42.49	3.12	Banyuls-sur-Mer	61	*pseudomajus*	Rockside bank (rocky)	32	54
THU	42.64	2.72	Thuir	130	*striatum*	Roadside bank (herbaceous)	34	60
LAG	43.09	2.58	Lagrasse	149	*pseudomajus*	Roadside bank (rocky / herbaceous)	32	55
BES	42.21	2.67	Besalú	195	*pseudomajus*	Stone walls in village	39	69
LUC	42.97	2.26	Luc-sur-Aude	227	*striatum*	Roadside bank and river-side bank (rocky)	19	29
RIP	42.21	2.20	Ripoll	709	*pseudomajus*	Roadside bank (herbaceous)	9	16
LYS	42.83	2.20	'Pierre-Lys' gorge	713	*striatum*	Roadside bank (rocky / herbaceous)	32	53
CAL	42.10	1.83	Berga	838	*pseudomajus*	Roadside bank (herbaceous)	42	69
PAR	42.31	2.20	Pardines	1118	*pseudomajus*	Roadside bank (herbaceous)	32	58
SAL	42.23	1.74	Saldes	1126	*pseudomajus*	Banks in pasture (herbaceous)	30	55
MIJ	42.73	2.04	Mijanès	1347	*striatum*	Roadside bank (herbaceous)	10	18
MON	42.51	2.12	Mont-Louis citadelle	1564	*striatum*	Stone walls on fortifications	21	34
						All populations	*372*	*637*

684   TABLE S2

685   Population pairwise F_{ST} for a) *Anthirinum majus pseudomajus* and b) *A. m. striatum.*

a) *Anthirinum majus pseudomajus*

Obs.	BAG	BAN	BES	CAL	LAG	PAR	RIP	SAL
BAG	0.000000	0.109661	0.111560	0.093282	0.134302	0.134404	0.097873	0.140489
BAN	0.109661	0.000000	0.136375	0.093343	0.120889	0.125510	0.086353	0.128994
BES	0.111560	0.136375	0.000000	0.068186	0.131472	0.147507	0.098488	0.152424
CAL	0.093282	0.093343	0.068186	0.000000	0.099317	0.139897	0.062174	0.106419
LAG	0.134302	0.120889	0.131472	0.099317	0.000000	0.159517	0.106148	0.156761
PAR	0.134404	0.125510	0.147507	0.139897	0.159517	0.000000	0.119122	0.144271
RIP	0.097873	0.086353	0.098488	0.062174	0.106148	0.119122	0.000000	0.088219
SAL	0.140489	0.128994	0.152424	0.106419	0.156761	0.144271	0.088219	0.000000

b) *Anthirinum majus striatum*

| Obs	LUC	LYS	MIJ	MON	THU
LUC	0.000	0.123	0.131	0.102	0.128
LYS	0.123	0.000	0.100	0.063	0.098
MIJ	0.131	0.100	0.000	0.078	0.099
MON	0.102	0.063	0.078	0.000	0.055
THU	0.128	0.098	0.099	0.055	0.000

686

687

688 TABLE S3

689 Quantitative genetics parameters for phenotypic traits among eight populations of *Anthirinum*

690 *majus pseudomajus* and five populations of *Anthirinum majus striatum* grown in a common

691 garden. Values for trait heritability (h^2), family variance (Vw), among-population variance (Vb),

692 residual variance corresponding to the within-population variance (Vres). The degrees of

693 freedom used in the bootstrapping procedures are seven for the among-population component

694 (Vb) for *A.m. pseudomajus* and four for *A. m. striatum*. Degrees of freedom are given in this

695 table for the within-population component (df$V$w).

| Traits	h^2	h^2.CI	Vw	Vb	Vres	df$V$w
a) *A. m. pseudomajus*						
Germination date	0.35	0.3; 0.42	14.05	0	65.13	246
Diameter	0.17	0.14; 0.20	0.06	0.02	0.69	248
Nodes	0.45	0.39; 0.52	3.03	2.62	10.37	248
Branches	0.07	0.06; 0.09	1.54	2.83	42.11	248
Plant height	0.32	0.26; 0.38	19.71	16.1	105.45	184
Internode length	0.58	0.51; 0.67	0.08	0.05	0.19	248
SLA	0.12	0.10; 0.15	128.26	78.01	1973.77	248
b) *A. m. striatum*						
Germination date	0.11	0.09; 0.15	4.37	4.49	71.98	110
Diameter	0.20	0.15; 0.26	0.06	0.08	0.57	111
Nodes	0.56	0.46; 0.69	2.94	1.55	7.6	111
Branches	0.01	0.01; 0.01	0.14	3.52	26.14	111
Plant height	0.06	0.04; 0.08	2.48	92.94	84	82
Internode length	0.05	0.04; 0.06	0.01	0.05	0.23	111
SLA	0.34	0.27; 0.43	397.54	155.9	1942.69	111

35

696 **Figures**
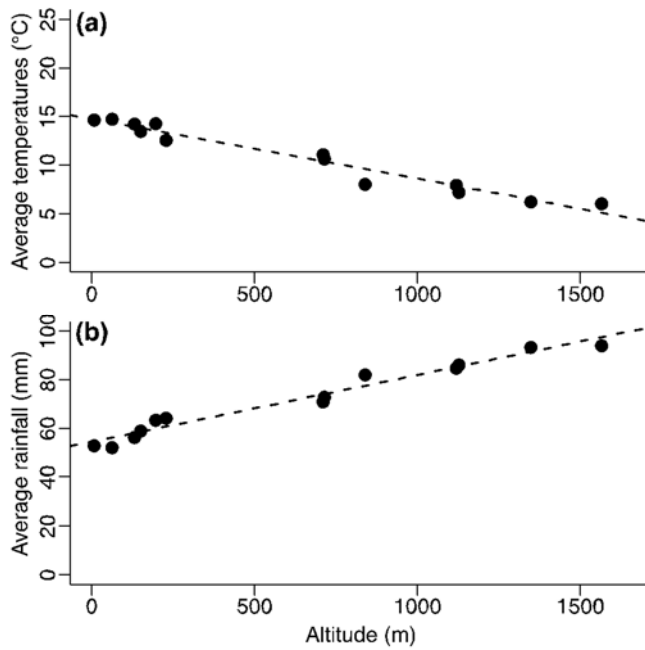
697

698 FIGURE S1.

699 Annual average temperatures and rainfall of 13 *Antirrhinum majus* populations from the

700 Southern France. Population average temperature (a) and average rainfall (b) as a function of

701 altitude. Bioclimatic data was extracted from the *WorldClim* database *(*www.worldclim.org).

702
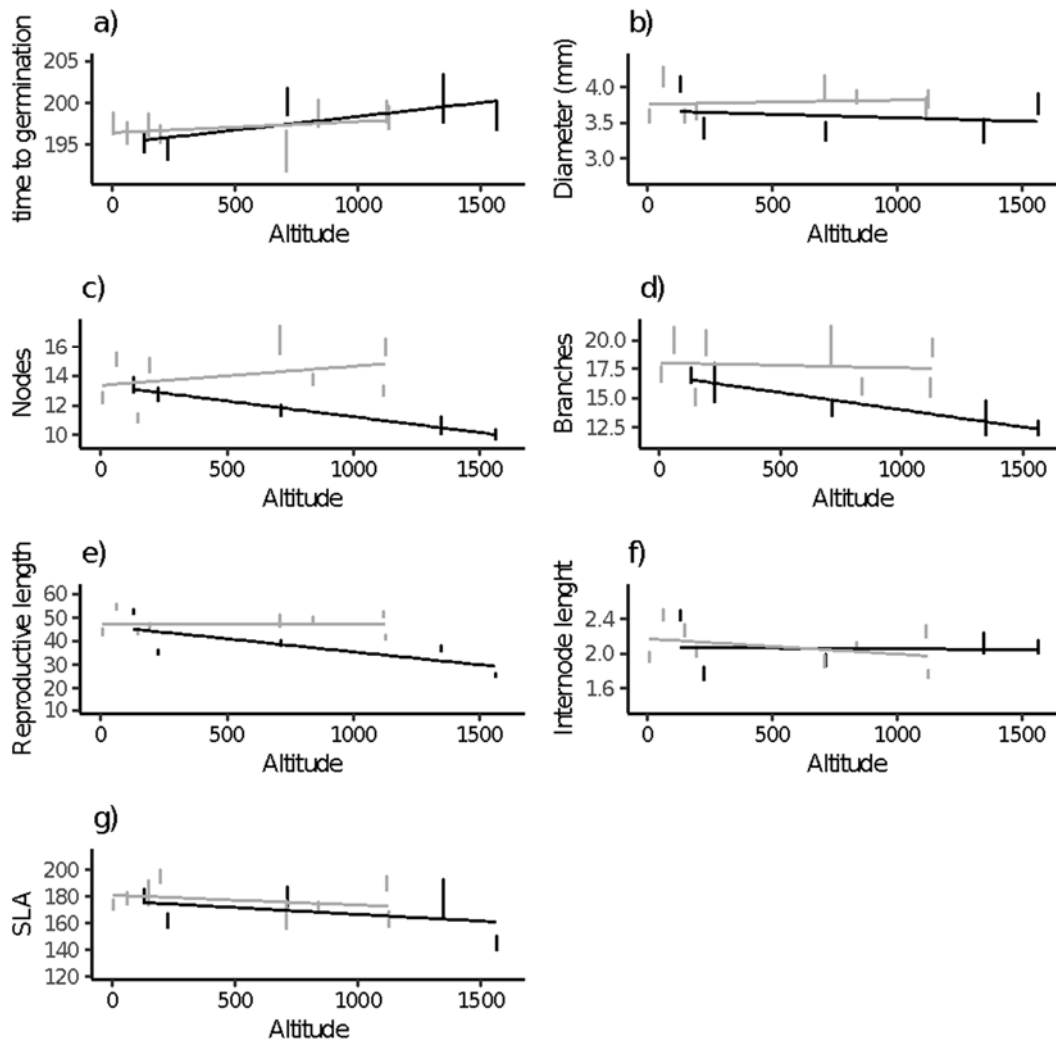
703

704 FIGURE S2.

705 Population arithmetic means with standard errors of seven phenotypic traits in populations of

706 two subspecies of *Antirrhinum majus* grown in a common garden. Means are plotted against

707 altitude of origin. Grey dots represent *A. m.* ssp. *pseudomajus* populations, black diamonds

708 represent *A. m.* ssp. *striatum* populations.
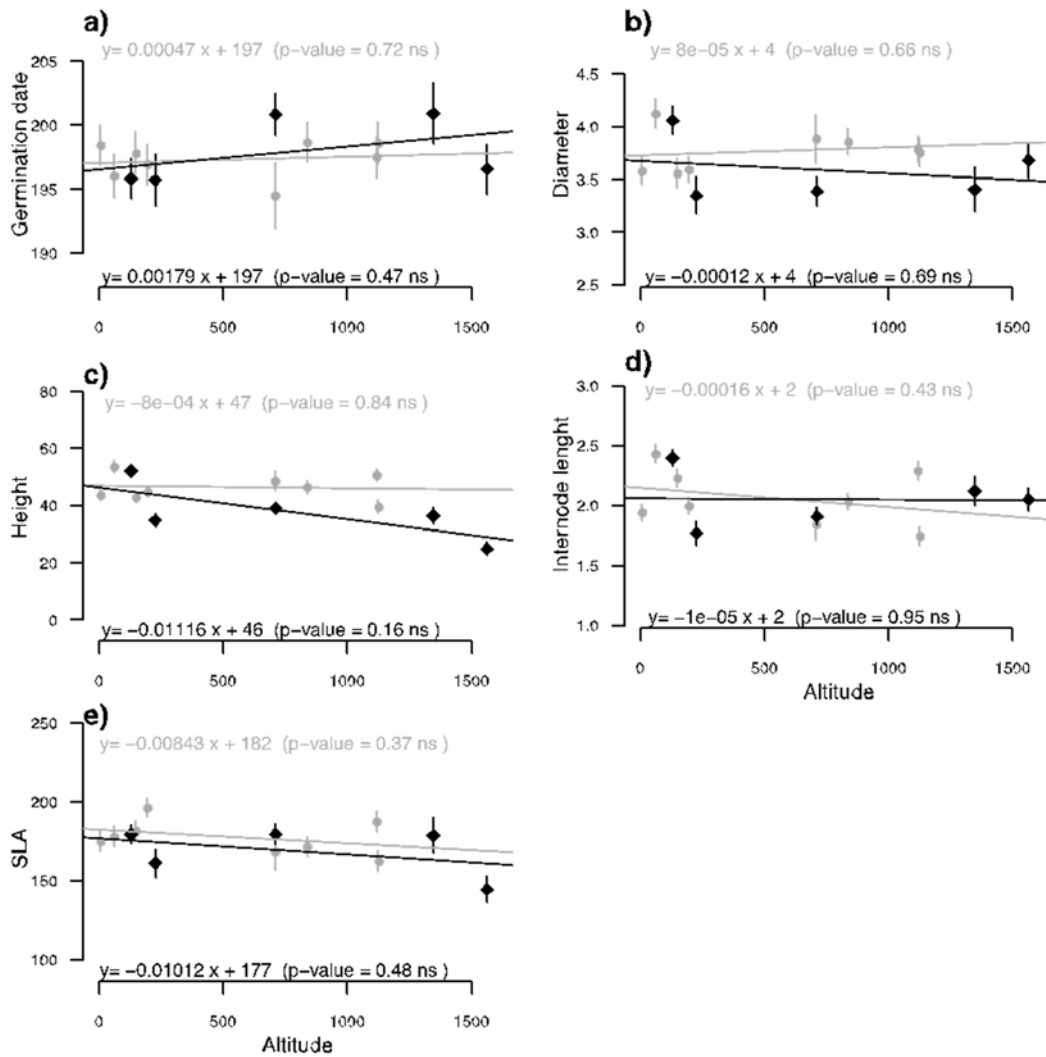
FIGURE S3. Population estimates of marginal means with standard errors of five phenotypic traits in populations of two subspecies of *Antirrhinum majus* grown in a common garden. Means are plotted against altitude of origin. Lines refer to the linear regression between traits means estimates and altitude. Grey dots and lines represent *A. m. pseudomajus* populations, black diamonds and lines represent *A. m.* ssp. *striatum* populations.

a) Germination date
b) Diameter
c) Nodes
d) Branches
e) Height
f) Internode lenght
g) SLA
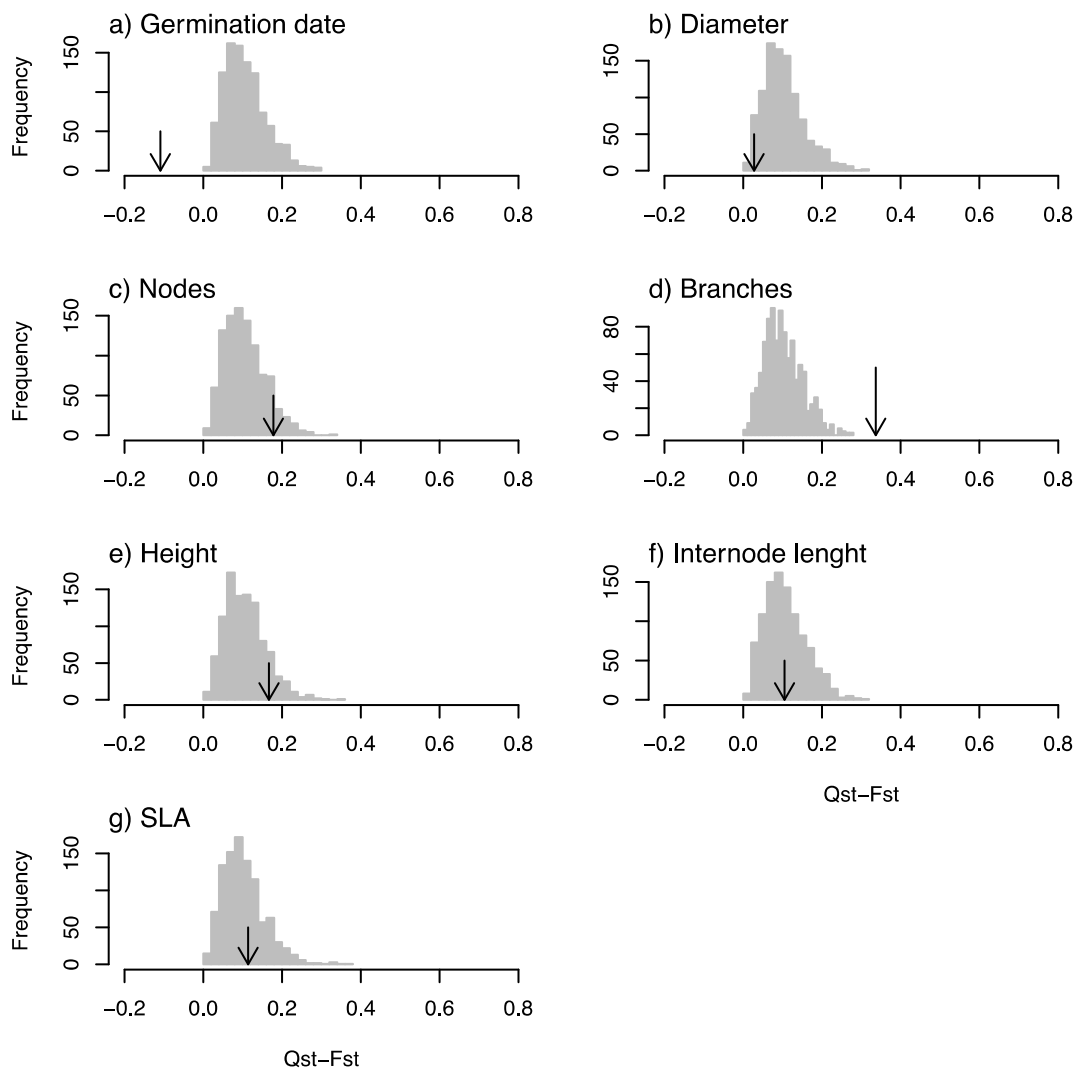
FIGURE S4. The simulated distribution of Q_{ST} - F_{ST} for a neutral trait, and the observed point estimates of Q_{ST} - F_{ST} differences in seven phenotypic traits measured for the eight *Antirrhinum majus pseudomajus* populations from the Southern France. The distribution of Q_{ST} - F_{ST} differences for a neutrally evolving trait was simulated following Whitlock and Guillaume (2009) based upon the observed population differentiation in neutral markers (F_{ST}) and the within-population variance in each trait. The arrow indicates the observed Q_{ST} - F_{ST}.
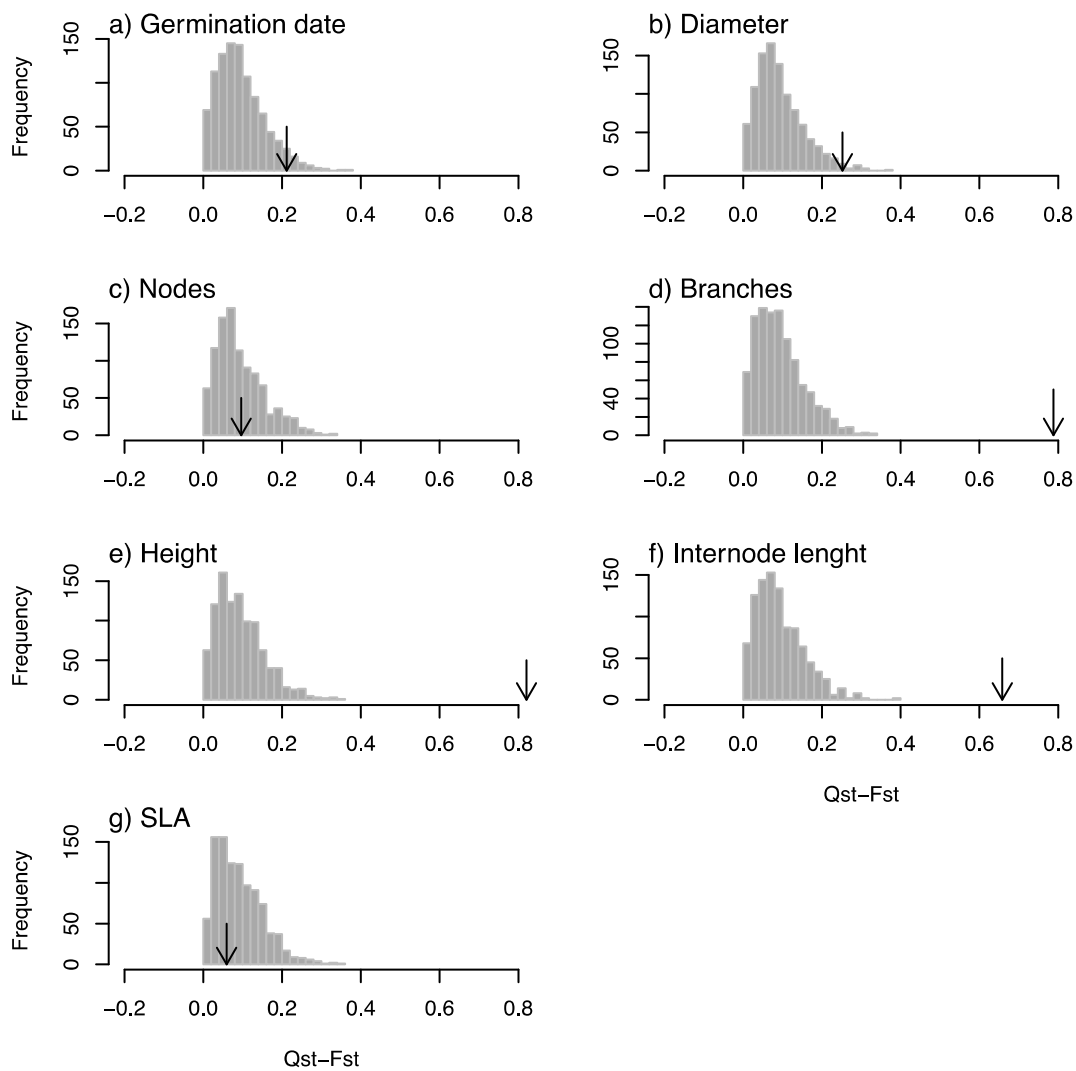
FIGURE S5. The simulated distribution of Q_{ST} - F_{ST} for a neutral trait, and the observed point estimates of Q_{ST} - F_{ST} differences in seven phenotypic traits measured for the five *Antirrhinum majus striatum* populations from the Southern France. The distribution of Q_{ST} - F_{ST} differences for a neutrally evolving trait was simulated following Whitlock and Guillaume (2009) based upon the observed population differentiation in neutral markers (F_{ST}) and the within-population variance in each trait. The arrow indicates the observed Q_{ST} - F_{ST}.