Génétique Quantitative et Evolution – Le Moulon
Ferme du Moulon
91190, Gif-sur-Yvette
France
Tel: +33(0)1 69 33 23 34
E-mail:tenaillon@moulon.inra.fr
http://moulon.inra.fr/~mtenaillon/

April 19th 2019

Dear Recommender,

We thank you for handling our manuscript. Enclosed is our revised manuscript entitled "Transcriptomic response to divergent selection for flowering time in maize reveals convergence and key players of the underlying gene regulatory network". We have received three very thorough and insightful reviews that highlighted the interest of reviewers for our work, and also suggested a number of modifications. Following yours and the reviewers' suggestions (among many other modifications detailed below): (1) we reoriented our ms towards the patterns of convergence notably by rewriting the first paragraph of the introduction; (2) we replaced the MDS plot by a PCA plot to clarify our approach; (3) we present now only three genes (instead of five) for which patterns were assessed by qRT-PCR (they are well-known flowering time gene and we corroborate some previous findings demonstrating the reliability of our approach); (4) we replaced one of the tables by a figure that illustrates the contrasts and corresponding number of differentially-expressed genes; (5) we detailed the methods (and mapping rate). Also note that sequencing data have been submitted to the SRA archive. All substantial changes are indicated in orange in the text. Altogether the reviews have helped us improving our manuscript. We hope that this revised version adequately addresses the reviewers' comments and will be suitable for your recommendation in PCI Evolutionary Biology.

Sincerely,

Maud Tenaillon

#REVIEWER 1

(1) Tenaillon et al. present results from a transcriptomic analysis of an artificial selection experiment for flowering time in maize. The authors perform RNA-seq of the shoot apical meristem of artificially selected progenies of two selection lines. The authors then identify differentially expressed genes by using variance partitioning and contrasts, and validate their RNA-seq results using RT-qPCR. Overall, the experiment provides quite a unique material for asking interesting evolutionary questions regarding divergence in gene expression as a result of artificial selection, but I found that the authors could take more advantage of this experimental setting in the discussion of their results. Overall, I enjoyed reading the manuscript and have only a single rather technical major comment. I found that the manuscript is overall well written, but that sometimes the main message is lost in the numerous acronyms and technical terms used in describing the results (e.g. "The Status category combined DE genes whose expression varied between two SAM status either in the F252 ([StatusF]) or MBS ([StatusM]) genetic backgrounds. " , or in general terms like "[SelM][StatusProg] category"). I would encourage the authors to try minimizing technical expressions and bring out more of the interesting biological and evolutionary results.

- We simplified technical terms over the ms and minimized the number of acronyms (for instance we got rid of SAM, FT, PC and DSE throughout the text). Also the Figure 2 is now replacing Table 2 and helps visualize the contrasts ([StatusF], [StatusProg], ..).

(2) p. 7: After trimming, filtering and mapping steps, we recovered between 29.87% and 47.22% of the reads (Table S1) that were used to estimate gene expression. Question: I'm surprised about the low percentage of retained reads after mapping and trimming. I would like to see a summary of the different criteria that account for reads to be not mapped/filtered out (e.g. are most of the reads of bad quality or otherwise unmappable?). I would in addition like to see the authors address the potential problems that such low percentage of usable reads pose: Are transcript models well-covered by sequencing reads, or is mapping coverage "patchy" over gene models? The concern here is that if coverage is unequal over gene models it's unclear how biologically relevant the observed "expression differences" actually are.

- We agree that we applied stringent criteria to retain reads. A large proportion of reads corresponded to rRNAs. Extracting total RNA from meristems was quite challenging. We applied a single round of purification to diminish the loss of RNA, and two rounds may have been better. We now provide in more details the number of reads recovered from the different steps (Table S1). As now stated in the ms: P.8: "Hence, less than half of the original reads were used which was primarily due to "contamination" of rRNAs – around 60% of the reads corresponded to rRNAs – and to a lesser extent to sequencing quality – 10% of the reads were discarded (Table S1)". Note that P.8 "The distributions of the normalized counts indicated lowest and highest quartiles comprised between ~30 and ~500, with a median coverage around 200 (Figure S3)". This means that the median coverage per transcript was 200. Because we found (1) that the genes from the flowering time pathway are over-represented among our differentially expressed genes, and (2) convergence of gene expression between the two inbred, we believe that our results are biologically relevant. Besides as now stated P. 8 "After filtering and normalization, we recovered a subset of 21,488 genes (55%) for which there was at least one count per million reads in half of our libraries. This estimate is close to what was found in the maize gene expression atlas for comparable organs (pooled samples of stem and shoot apical meristem) at two vegetative stages (Stelpflug et al., 2016)."

(3) page 3: One important observation from these experiments is that the response to selection is generally steady over generations …Question: It's unclear what's meant by "steady".

- We replaced "steady" by "continuous".

(4) p. 4: This is particularly intriguing for the Saclay's DSEs that started from inbred lines with limited standing variation (<1.9%) and evolved under very small population size. Question: It's unclear what's referred to by "<1.9%".

- We modified the sentence to clarify “This is particularly intriguing for the two Saclay divergent selection experiments that were conducted independently in two genetic backgrounds (two inbred lines) with limited standing variation (<1.9% residual heterozygosity) and very small population size”.

(5) p. 4: The combination of both new mutations and standing variation is consistent with the known complexity of flowering time determinism in maize, and a high mutation target, i.e. >100 loci (Buckler et al., 2009). Question: I understand from this sentence that the assumption is that changes in the expression of most of the DE genes are due to new mutations. In this context what's the authors take in the proportion of DE genes showing selection effect - is this more than expected by chance, suggesting convergence? Or is expression divergence mutation-limited in this system?

- What we have shown is that both new mutations AND standing variation (residual heterozygosity) have contributed to the response. The expression divergence should not be mutation-limited (the response is not, it is continuous). I don’t think we can make prediction about how many genes DE genes should display selection effect because a lot of DE genes are response genes and not directly targeted by selection. Convergence should be more common with standing variation. We changed the text P.4 to clarify “The contribution of both, new mutations and standing variants, to the response to selection to flowering time is consistent with the known complexity of this trait – its high mutation target, i.e. >100 loci (Buckler, 2009)”. And also P.21 in the discussion “While convergence between inbred lines is more likely to happen from shared standing genetic variants, previous results have shown that *de novo* mutations have contributed to our observed response to selection”. And finally we added P.22: “The relative contribution of standing expression variants in ancestral genotypes versus *de novo* expression variants acquired independently during the course of the experiment, to patterns of convergence remains to be established. This could be achieved by investigating allele specific expression in the ancestors of the subset of 115 DE genes that displayed convergence of expression between the two inbreds.”

(6) p. 5: Second paragraph about candidate genes. Question: I wonder could the authors come back to these genes and mechanisms in the discussion - I have trouble to follow how these genes were relevant in the light of the results.

- We have now reduced to three genes (getting rid of two potential candidates and focusing on the three well-described genes). We justified more cleary our choice P. 11: “In order to validate our material and methodology, we employed all RNA samples (Table 1) including those used to build RNAseq libraries to investigate via qRT-PCR, the expression of three genes whose interactions and effects on floral transition has been established. Those three genes are the florigen *ZCN8* (Meng et al., 2011) and the negative regulator of flowering *RAP2.7*, that both modulate the expression of the floral meristem identity integrator *ZMM4* (Danilevskaya,et al. 2008) (Figure S5). In addition to the shoot apical meristem, we examined expression in three additional organs: the immature part, mature part and the sheath of the last visible leaf (Figure S6)”.

(*7) p. 7: FT occurred at the same plant developmental stage in Early (FE) and Late (FL) genotypes in F252 (8 visible leaves), but occurred at an earlier plant developmental stage (9 visible leaves) in Early (ME) than in Late (ML) MBS progenitors (10 visible leaves). Question: This is an interesting observation and I wonder if this reflects different mechanisms for response to selection? How this difference could be reflected in the different DE genes under selection in the two lines?*

- This is indeed an interesting observation. We have characterized very finely the development of those evolved genotypes during two years of fields experiments. We are still analysing the data and will be precisely answering this question of the different developmental routes to early flowering. This is beyond the scope of this paper.

(8) p. 8: Altogether, our methodology revealed repeatable patterns and a visible signal of differential gene expression. Question: What is meant by “visible” signal?

- We erased the sentence.

(9) p. 8: We performed 27 contrasts to detect DE genes (Table 2). Question: It would be helpful for the reader if these contrasts would be summarized in some sort of schematic form. In addition, what is the contrast in SAM status (floral vs vegetative?)?

- We replaced Table 2 by Figure 2 that provides a schematic representation of the contrasts that we performed. That should answer your question.

(10) p. 9: For MBS, we found 446 DE genes within the Selection category. For F252, we found 2,120 in the Selection category, that comprised 748 DE genes between Early and Late or Very Late F252 progenitors ([SelF]). Considering both F252 and MBS, there were 2,451 DE genes falling into the Selection category (Table 2 & Table S4). Question: I would like to see the authors discuss why there is so big difference in the number of Selection-category gene between the two lines (e.g. because expression divergence is mutation-limited? Because of single big trans-acting mutations in one line but not the other?).

- This is discussed P17: "Overall, F252 displayed more DE genes within the Selection category than MBS (2120 vs 446). This is in line with overall lower level of residual heterozygosity detected in the former (Durand, et al. 2015). However, if we considered only the Early and Late progenitors of each Line (discarding the VeryLate Progenitor of F252), we obtained the inverse trend, with less DE genes for F252 (346) than for MBS (446) (Figure 2, Figure S4). This pattern was consistent with a lack of phenotypic response in the Late F252 population after seven generations of selection but a continuation in the Late MBS population until G13 (Durand, et al. 2015). Such lack may be explained by the strong selection operated in the VeryLate F252 population that in turn relaxed selection in the Late F252 population (Figure 1)."

(11) p. 9: We performed a principal component analysis on the set of 7,370 DE genes, and attributed each one of them to a principal component axis based on correlation coefficient values (Table S4). Question: What is the logic/advantages to use correlation instead of simply the PC loadings of the individual genes for this analysis?

- Correlation coefficients between variables and PCs are fully related to PC loadings, but account both for the inertia (accounted for by the PC) and the trait variance. Besides, they vary between 0 and 1. For both reasons, their interpretation is easier. In addition, statistical tests may be performed with correlation coefficients (FactoMine R library). This approach also is the one proposed in the Mixomics package. We added a few sentences in the Method section to clarify P.30 : "To reduce the complexity, we used an approach similar to the one proposed in the MixOmics package (Rohart, et al. 2017). We calculated the Pearson correlation coefficient of each gene to the first 13 Principal Components and, based on the absolute value of the greatest correlation coefficient among 13, we attributed each gene to a single principal component axis. Correlation coefficients were preferred to PCA loadings because they accounted for both the variance of the PC axis and the trait variance."

(12) p. 10: Out of 2,451 DE genes of the Selection category, the PC1 exhibited the greatest proportion of all PCs (46.3%, Table 3) followed by PC2 (30.1%, Table 3). Question: I have troubles following the logic of this analysis attributing DE genes of certain category to PC's. I'm assuming I'm missing some key point here.

- We agree that Table 3 (now Table 2) is quite confusing and does not add much to the discussion. We now simply used the attribution of DE genes to PCs to draw the heatmaps (one heatmap per PC). We do not interpret further the attributions to DE genes /certain category to PCs and erased Table 3. We have shortened the text accordingly.

(13) p. 11: Out of 70 and 984 of the FT_candidates and GWA_candidates, 54 and 294 respectively displayed differential gene expression (Table 3). Question: Is this difference in proportions statistically significant?

- We replaced P.13 by "Out of 70 and 984 of the FT_candidates and GWA_candidates, 54 (representing 77.1%) and 294 (representing 29.9%) respectively displayed differential gene expression. Comparatively, FT_candidates therefore presented a greater enrichment of DE genes than GWA_candidates (P-value=7.5 10^{-7})".

(14) p. 12: We focused on four organs: the immature part, mature part and the sheath of the last visible leaf, and the SAM. Question: What is meant by "part"?
- We added a Supplementary Figure to show which organ we sampled (Figure S6).
(15) p. 18: We found an overall significant overdispersion (dispersion parameter=1.17, P-value=1.04 10-7), albeit with noticeable differences among chromosomes (P-value=0.0006). Chromosomes 1, 6 and 10 were significantly enriched for DE genes of the Selection category, while chromosomes 4, 7, 9 were significantly depleted. Question: I think this is an interesting analysis, could be given more weight and would better fit the results section on its own.
- While agree this is an interesting observation, it is difficult to speculate on the number of flowering time genes/QTLs bear by each chromosome. We therefore decided not to discuss this point.
(16) p. 18: …the majority of these genes (91.1%) belonged to the 4 first PCs… Question: What does "belong" here mean?
- we changed for "were attributed to".
(17) p. 21: Genetic convergence between inbred lines is detected. Question: I like the discussion of possible cis vs trans sources of transcriptomic convergence - but I can't follow which types of changes the authors think are more prevalent.
- Both sources can result in transcriptomic convergence. Again, it's difficult to speculate on which one is most prevalent. We actually erased a couple of sentences P.20 as we agree it was quite confusing.

#Reviewer 2
Tenaillon et al. used transcriptomics on inbred maize lines under divergent selection regimes to compare the response to selection in terms of gene expression. They chose two inbred lines to undergo 13 generations of selection for early and late flowering time. Transcriptomics data was taken from shoot apical meristems at multiple time points defined by developmental stage. They performed 27 contrasts to identify genes that were differentially expressed under these conditions. They found differential expression by time point, genotype, and selection. The most interesting of these are the genes that are differentially expressed across multiple contrasts. Specifically, there are a subset of the genes that were differentially expressed across the selection contrast in both inbreds pointing to a convergent response to selection. This subset included many genes already known to play a role in flowering time in maize.
Overall this is an interesting study and worthy of recommendation. My comments are mostly confined to things that could be done to improve clarity. Many of the conclusions seem to be dependent on a PCA analysis described in table 3.

(1) Table 3 is a little confusing, and I think it might be beneficial to have a figure of the PCA.
- We agree and replaced Table 2 by Figure 2 that provides a schematic representation of the contrasts that we performed. We also replaced the MDS plot by a PCA.
(2) You mention in the first paragraph of the discussion that you demonstrated that the timing of floral transition is linked to flowering time, but you also mention that floral transition drives flowering time in the introduction as part of the explanation for your work. So I think you need to highlight the specific part of this that's a new discovery.
- To clarify, we wote P. 14: "Here, we showed that by selecting for flowering time difference, we indirectly selected for the timing of floral transition, which occurred earlier in early progenitors as compared with late progenitors".
(3) I thought the most exciting part of this study was the discussion section "genetic convergence between inbred lines is detected". But I also felt like the first time it was brought up was on page 21. I think these findings could be more clearly highlighted in the results section.
- The genetic convergence is mentioned in the title, the abstract at the end of the introduction. However other reviewers also felt that the introduction should highlight better this aspect and we consequently rewrote the first paragraph of the introduction P. 3: "Artificial selection experiments are designed to investigate evolutionary responses of complex traits. These responses inform us

about the limits to the evolution of ecologically relevant phenotypes as well as their genetic architecture, determined by interactions among a multitude of environmentally sensitive genes. Replication of such experiments across distinct genetic backgrounds provides a unique opportunity to test whether convergent evolution at the phenotypic level recruits similar molecular solutions. […]. These studies have collectively recovered genetic predictability at different levels, SNPs, genes and functional units. This predictability however depends on the origins of mutations that contribute to the response to selection. Hence, standing genetic variants shared between genetic backgrounds – inherited from a common ancestor – are more likely to generate convergence than unshared *de novo* mutations (Graves, et al. 2017).

(4) Some of the results of table 2 are discussed in detail, but I was most surprised by the ones with very low numbers of differentially expressed genes, which seems to be an unexpected result. I think it would be helpful if the authors mentioned what drives particularly low numbers of differentially expressed genes. Is it at these time points turned out not to be meaningful?

- We agree that in some contrasts we detected a low number of DE genes. These contrasts involve the transitioning and reproductive meristems for the MBS line, and the VeryLate genotype of F252 (Figure 2). This is likely due to the fact that we had only the reproductive stage for one of the two years as shown in Figure S2. We now explicitly propose this explanation in the text P.8-9 "Note that because we were able to obtain reproductive meristems only for one of the two years for MBS and VeryLaye F252 genotypes (Figure S2), we detected a correspondingly low number of DE genes (<4) in contrasts involving vegetative and transitioning meristems".

(5) Finally, I have a copy editing comment: on page 2 paragraph 2 the phrase "we obtained" is repeated.

- Thanks!

(6) Also, as someone who is not directly involved in this field, the sheer number of different acronyms made this paper very hard to follow. I recognize that if I had worked on maize or flowering time more recently these would probably be clearer. For example, in table 1 there are 3 meristem statuses V,T, and R and you do define them very early in the paper but I lost track of what they were by the time I got to the table. It would be helpful if you re-defined your acronyms in figure and table legends.

- cf. Reviewer 1: We simplified technical terms over the ms and minimized the number of acronyms (for instance we got rid of SAM, FT, PC and DSE throughout the text).

#Reviewer 3

Review of Tenaillon *et al.*: Transcriptomic response to divergent selection for flowering time in maize reveals convergence and key players of the underlying gene regulatory network

In the manuscript the authors study early and late flowering maize inbred lines that have been obtained from selection experiments. The authors aim at understanding the responses to phenotypic selection at the transcriptome level, at adding new information to the gene regulatory network of flowering in maize and at testing for convergence of selection at the level of individual genes between different lines. The authors conducted an RNAseq study in agronomical field conditions on shoot apical meristems in different developmental stages. Expression of some genes were also measured by qRT-PCR.

Timing of flowering is economically important trait in many crop species. The genetic basis of flowering time variation has been and still is widely studied. Traditionally these studies focused on the function of individual genes, but during the last years RNAseq studies have become an important method for identifying gene groups that are differentially expressed between selected lines.

The strength of the paper is that the experiment is done in the field conditions and in my opinion this is the aspect that the authors should emphasize more. In general, the paper reads well, is nicely structured and it is easy to follow the story. However, there are several issues that should be addressed to improve figures and the readability of the paper:

(1) In the manuscript "FT" is used for abbreviation of "floral transition". FT is commonly used for FLOWERING LOCUS T. In the context of flowering and its genetic basis, I would not use FT for anything else, to avoid misunderstandings and confusion. For instance, the sentence "ZCN8 encodes a florigen protein that migrates through the phloem from the leaf to SAM triggering via its accumulation, the reprogramming of the SAM to FT" in the introduction can be very confusing.

- We replaced throughout the text by floral transition, we just keptthe worf "FT-candidates". More generally, we got rid of a number of acronyms (cf. Reviewers 1 & 2).

(2) Include the full gene name when the gene is mentioned the first time; both in the abstract and in the main text.

- Done.

(3) The material used in the study is obtained from the selection experiments called Saclay's DSE. Authors' role in those experiments should be made more clear.

- We rephrased P. 4: "In these experiments, by applying divergent selection over 16 generations, **we generated** considerable phenotypic response with up to three-weeks difference between early and late flowering populations (Durand, et al. 2015)".

(4) The focus of this paper is on flowering time variation. Were there any other phenotypes co-varying with flowering time? Moreover, to me the difference between early and late lines is not that big (based on Figure 1, only about 10 days in the MBS line, about 16 days in the F252 line). Please discuss this.

- Yes, there are other phenotypes co-varying. We have done intense phenotyping of those genotypes but have not published them yet. We think this is out of the scope of this paper. We disagree that 10-15 days is not that big. We clarified this point P.6 at the beginning of the Results section: "We created a unique material by applying divergent selection for flowering time independently to two maize inbred lines, F252 and MBS (Durand, et al. 2015). This material is used to investigate evolutionary responses to selection of a complex trait and to dissect its genetic architecture. After 13 generations of selection, phenotypic responses revealed roughly two weeks difference between Early and VeryLate F252 populations, and between Early and Late MBS populations (Figure 1). Starting from nearly-fixed commercial inbreds, this difference is objectively striking. It compares to what is observed across multiple European inbreds evaluated across contrasted environments (Lehermeier, et al. 2014). Our selection experiments therefore maximized the phenotypic differences between evolved lines in nearly-identical genetic backgrounds".

(5) Although gene expression studies (especially RNAseq) in the field conditions are still quite rare, some of them have been published. They should be cited and discussed in the paper. Moreover, are there other RNAseq studies on maize? If they exist, also they should be discussed. Explain why this experiment was done in the field conditions.

- We are now comparing our results to RNA-seq studies previously published in maize P. 8: "This estimate is close to what was found in the maize gene expression atlas for comparable organs (pooled samples of stem and shoot apical meristem) at two vegetative stages (Stelpflug, et al. 2016)". P12-13: "But *ZCN8* displayed no detectable expression in RNAseq consistently with previous results detecting no expression in the meristem (Stelpflug, et al. 2016)". P. 19: "Consistently, a previous study detected very low/no expression (fragments per kilobase exon model per million mapped fragments [FPKM]<2) using RNAseq from meristem and leave tissues (Stelpflug, et al. 2016)". We are also referring to a transcriptomic field experiment in Arabidopsis publication P.16: "Interestingly in *Arabidopsis thaliana*, shoots sampled in the field at 3-days intervals over a growing season in two accessions revealed that temperature and precipitation captured a small proportion of the transcriptional variance relative to flowering status (Richards, et al. 2012)".

(6) *Describe why the used genes were selected for qRT-PCR.*

- cf. Comment (6) Reviewer 1.

(7) *Results, First paragraph: please make it more clear that the used flowering time data is from another study, not from the one presented here.*

- we added: P.6 :"(data extracted from (Durand, et al. 2015), Figure 1)".

(8) If the Material and Methods section will stay at the end of the manuscript, more details of the field experiment are needed in the results section. For instance, the years are not explained, location of the field experiment is not mentioned etc. I would personally put the Material and Methods between Introduction and Results to avoid redundancy.

- P.21: It is written "All 5 progenitors were selfed to produce seeds. The resulting progenies were grown in the field at Gif-sur-Yvette France during summer 2012 and 2013.", we now added P.23 "These experiments were held in the field at Université Paris-Saclay (Gif-sur-Yvette, France)". And also P. 24 "Precipitations over the growing period totalized 117 mm and 60.5 mm and the average daily temperature reached 16.51°C and 18.41 °C for 2012 and 2013, respectively. The second year of experiment was therefore hotter and drier than the first year". We decided to keep Material and Method as the last section.

(9) "We determined the timing of FT of each progenitor as the earliest stage at which we observed a majority of transitioning SAMs". What does the majority mean in this case? Please be more clear. It would be better to use the proportion of meristems being vegetative/transitioning/being reproductive (also in Table 1).

- we rephrased: P.7 "We determined the timing of floral transition of each progenitor as the earliest stage at which the proportion of transitioning shoot apical meristems was the highest (Figure S2, Table 1)".

(10) "FT occurred later in MBS than in F252 consistently with the flowering time difference between these two inbreds." Is this difference statistically significant?

- P. 7 "floral transition occurred later in MBS than in F252 consistently with the flowering time difference between these two inbreds (P-value<2.2 10^{-16} and <2.35 10^{-12} in 2012 and 2013 respectively);".

(11) Section "Genome-wide patterns of gene expression as determined by RNA-seq". Were the same individuals used for both phenotyping and RNAseq? Please name the samples used in the pairwise correlations so that they can easily be identified in Table 1. Some data (and statistics) on clustering in Figure 1 should be included.

- cf (7) above, we now made clearer in the text and legend of Figure 1 that data were extracted from a previous publication. P.27, we wrote: "We used a subset of the RNA samples used in the qRT-PCR to perform RNA-Seq. This subset contained shoot apical meristem samples from 2012 and one of the 2013 replicate". The individuals used for phenotyping and RNAseq were different (the procedure is stated P.23 with selfing of selected progenitors).

(12) Section "DE genes and targets of selection". I find the two first paragraphs very list-like. The second paragraph repeats the numbers of Table 2. Is there a way to make it more readable?

- We combined the two first paragraphs and shortened them as well to make it more readable. Changes were not tracked but they correspond to the first paragraph of that section.

(13) Section "Expression of 5 genes as determined by qRT-PCR and correlation with RNA-Seq measures". Mention the names of the studied genes already at the beginning of the section. Please explained why those five genes were selected for qRT-PCR. Connect to the previous section ("Functional relevance of DE genes"). Were these genes among the DE genes in the RNAseq study?

- cf (6) #Reviewer 1 (P. 11). We modified the text to explain/connect/name the genes. We discarded two of them. P. 12: "Interestingly, we noticed differences in expression between Early and Late or VeryLate Progenitors in F252 for *RAP2.7*. Note that we found good agreement between qRT-PCR and RNAseq for *ZMM4* and *RAP2.7*, with significant correlations between levels of expression determined by the two methods (Figure S8). But *ZCN8* displayed no detectable expression in RNAseq consistently with previous results detecting no expression in the meristem (Stelpflug, et al. 2016)".

(14) The organs used for qRT-PCR are not clear for those who do not work with maize. What are the immature and mature parts of leaves? Pictures would help.

- We added Figure S6 with pictures to show the organs.

(15) The sentence "We also incorporated with the amplification of each gene control samples for F252 and MBS for which all organs were pooled" is not clear.

- We erased that sentence.

(16) "While we were able to detect ZCN8 expression in the SAM via qRT-PCR (Figure 5)": Figure 5B - Figure 4B

- We do not refer to that Figure anymore.

(17) Discussion. In general, I think the Discussion is a bit too long. Maybe it could be shortened and made more compact.

- We shortened slightly the discussion.

(18) There is a conflict between sentences "As expected, expression varied more between lines than among developmental stages" (page 15) and "We have shown that the meristem developmental status is the main source of differential expression" (page 22).

- Corrected P.22: "is the second main source of differential gene expression".

(19) Material and Methods. Some environmental data on the field site during the field experiment should be added. The coordinates of the field site would also be informative. When were the seeds sown and/or the plants planted? Were the plants let to germinate in the field conditions or somewhere else? If somewhere else, the conditions for the pre-growing should be described. How many plants were planted? How was the experimental set up – were the plants randomized, were there blocks?

- We added P.23: "These experiments were held in the field at Université Paris-Saclay (Gif-sur-Yvette, France)." and later on P.23 "The resulting progenies were grown in the field at Université Paris-Saclay (Gif-sur-Yvette, France) during summer 2012 and 2013. Precipitations over the growing period totalized 117 mm and 60.5 mm and the average daily temperature reached 16.51°C and 18.41 °C for 2012 and 2013, respectively. The second year of experiment was therefore hotter and drier than the first year".

Check that all parenthesis are opened and closed in the "qRT-PCR assays and statistical procedure" section. Add a comma between the R primer of ZCN8 and the RAP2.7 gene.

- Corrected.

In expression studies the timing of sample collection is important. Although this study focuses on developmental genes which are often either "on" or "off", some of them might have circadian expression patterns. It should be mentioned what time of the day the samples were collected and how this issue was considered.

- P24: "collected early morning (between 8:00 and 9:00 am) on a daily basis

(20) Figures. Fig. 1: How were days to flowering measured? In which conditions? Days from where? Add the abbreviations used in the text to the figure (FE, FL etc.).

- Done (see legend + figure)

Fig. 2: Specify which sample is which (has to be comparable to Table 1, at least it has to be clear which samples are from early/late lines). Are the symbols correct? Conflicts with Table 1: e.g. in Table 1 there are 7 F252 samples for RNAseq, but in Figure 2 only 6.

- Thanks for pointing that out, we have now added the progenitor information on Figure 2 (now Figure 3) and modified the text P.9: "We performed a principal component analysis on the mean normalized expression of the set of 7,370 DE genes. It revealed a separation between lines and among shoot apical meristem status in F252 along axis 1, and among progenitors and status along axis 2 (Figure 3)".

Fig. 3: please add the axis names, meaning of colors and explain the diagrams.

- Done.

Fig. 4: explain the colors used in the figure.

- cf. Legend.

Fig. 5: this figure should be cited in the results section, not only in the discussion. Check that the used colors correspond to the ones mentioned in the figure legend. Add A-F in the figure. B and C figures are swapped, to my understanding.

- We think it will make the text unclear to describe the expression patterns at those genes without discussing simultaneously their patterns of expression. We therefore only cited the figure in the discussion. We have corrected the legend.

Fig. S1: adding FE, FL, FVL, ME and ML next to the red circles of their progenitors in the figure would be informative.

- We modified the S1 figure to add the names of the progenitors and we also now only present the first 13 generations of selection. We simplified the description of the methods accordingly.

Fig. S2: the y-axis is missing. The x-axis in the first figure is missing. Different leaves are shown for different lines and it makes it difficult to compare the figures. To me it is not completely sure how the figure should be read. Does it mean that for instance in FL line (year 1), in all plants with 6 leaves the SAM is vegetative, and in ~20% of plants with 9 leaves the SAM is transitioning and in remaining ~80 the SAM is reproductive? The number of pooled samples should be mentioned in the figure legend.

- We modified the legend to clarify.

Fig. S3: specify which sample is which (see Table 1).

- We used the names of the libraries given in Table S1 as now specified in the legend.

Fig. S5: "When known relationships" -> "When known, relationships"; Are the explanations of black dots and arrows correct? Explain GRN and GDB.

- We corrected the legend of the figure.

In Figure 1 the flowering time is shown as days and in Figure S2 as developmental stage. How are these connected to each other? Can you compare those phenotypes?

- Flowering time is never estimated as developmental stage but in days. It is the floral transition that is shown as developmental stage. Now that we don't use the FT acronym anymore, it should clarify this confusion.

(Microscopic) pictures of plants in different developmental stages would be informative for people not familiar with phenotyping in maize.

- Sure, this is why we illustrated the developmental stages of the meristem Figure S5.

(21) Tables.

Table 1: Explain V, T and R. The reproductive stage (year 1) is not reached for FVL, ME and ML. Why?

- V, T, R are now explained in the legend, and clarified P.8 "Note that because we were able to obtain reproductive meristems only for one of the two years for MBS and VeryLaye F252 genotypes (Figure S2), we detected a correspondingly low number of DE genes (■) in contrasts involving vegetative and transitioning meristems".

I could not access the Supplementary Tables and those materials are not reviewed. Please make sure they are available.

- Supplementary tables are now accessible at the link provided in the text.

(22) We took into account all small text edits of this reviewer.