Early and late flowering gene expression patterns in maize

Tanja Pyhäjärvi based on reviews by Laura Shannon and 2 anonymous reviewers

A recommendation of:

Artificial selection experiments are key experiments in evolutionary biology. The demonstration that application of selective pressure across multiple generations results in heritable phenotypic changes is a tangible and reproducible proof of the evolution by natural selection. Artificial selection experiments are used to evaluate the joint effects of selection on multiple traits, their genetic covariances and differences in responses in different environments. Most studies on artificial selection experiments report and base their analyses on phenotypic changes [1]. More recently, changes in allele frequency and other patterns of molecular genetic diversity have been used to identify genomic locations where selection has had an effect. However, so far the changes in gene expression have not been in the focus of artificial selection experiment studies (see [2] for an example though). In plants, one of the most famous artificial selection experiments is

the Illinois Corn Experiment where maize (Zea mays) is selected for oil and protein content [3], but in addition, similar experiments have been conducted also for other traits in maize. In Saclay divergent selection experiment [4] two maize inbred lines (F252 and MBS847) have been selected for early and late flowering for 13 generations, resulting in two week difference in flowering time.

In "Transcriptomic response to divergent selection for flowering time in maize reveals convergence and key players of the underlying gene regulatory network " [5] Maud Tenaillon and her coworkers study the gene expression differences among these two independently selected maize populations. Their experiments cover two years in field conditions and they use samples of shoot apical meristem at three different developmental stages: vegetative, transitioning and reproductive. They use RNA-seq transcriptome level differences and qRT-PCR for gene expression pattern investigation. The work is continuation to earlier genetic and phenotypic studies on the same material [4, 6]. The reviewers and I agree that dataset is unique and its major benefit is that it has been obtained from field conditions similar to those that species may face under natural setting during selection. Their tissue sampling is supported by flowering time phenotypic observations and covers the developmental transition stage, making a good effort to identify key transcriptional and phenotypic changes and their timing affected by selection. Tenaillon et al. [5] identify more than 2000 genes that are differentially expressed among early and late flowering populations. Expectedly, they are enriched for known flowering time genes. As they point out, differential expression of thousands of genes does not mean that they all were independently affected by selection, but rather that the whole transcriptional network has shifted, possibly due to just few upstream or hub-genes. Also, the year-to-year variation had smaller effect in gene expression compared to developmental stage or genetic background, possibly indicating selection for stability across environmental fluctuation for such an important phenotype as flowering time.

Another noteworthy observation is that they find convergent patterns of transcriptional changes among the two selected lines. 115 genes expression patterns are shifted due to selection in both genetic backgrounds. This convergent pattern can be a result of either selection on standing variation or de novo mutations. The data does not allow testing which process is underlying the observed convergence. However, their results show that this is an interesting future question that can be addressed using genotype and gene expression data from the same ancestral and derived material and possibly their hybrids.

References

[1] Hill, W. G., & Caballero, A. (1992). Artificial selection experiments. Annual Review of Ecology and Systematics, 23(1), 287-310. doi: [10.1146/annurev.es.23.110192.001443](https://doi.org/10.1146/annurev.es.23.110192.001443) [2] Konczal, M., Babik, W., Radwan, J., Sadowska, E. T., & Koteja, P. (2015). Initial molecular-level response to artificial selection for increased aerobic metabolism occurs primarily through changes in gene expression. Molecular biology and evolution, 32(6), 1461-1473. doi: [10.1093/molbev/msv038](https://doi.org/10.1093/molbev/msv038) [3] Moose, S. P., Dudley, J. W., & Rocheford, T. R. (2004). Maize selection passes the century mark: a unique resource for 21st century genomics. Trends in plant science, 9(7), 358-364. doi: [10.1016/j.tplants.2004.05.005](https://doi.org/10.1016/j.tplants.2004.05.005) [4] Durand, E., Tenaillon, M. I., Ridel, C., Coubriche, D., Jamin, P., Jouanne, S., Ressayre, A., Charcosset, A. and Dillmann, C. (2010). Standing variation and new mutations both contribute to a fast response to selection for flowering time in maize inbreds. BMC evolutionary biology, 10(1), 2. doi: [10.1186/1471-2148-10-2](https://doi.org/10.1186/1471-2148-10-2) [5] Tenaillon, M. I., Seddiki, K., Mollion, M., Le Guilloux, M., Marchadier, E., Ressayre, A. and Dillmann C. (2019). Transcriptomic response to divergent selection for flowering time in maize reveals convergence and key players of the underlying gene regulatory network. BioRxiv, 461947 ver. 5 peer-reviewed and recommended by PCI Evolutionary Biology. doi: [10.1101/461947](https://doi.org/10.1101/461947)

[6] Durand, E., Tenaillon, M. I., Raffoux, X., Thépot, S., Falque, M., Jamin, P., Bourgais A., Ressayre, A. and Dillmann, C. (2015). Dearth of polymorphism associated with a sustained response to selection for flowering time in maize. BMC evolutionary biology, 15(1), 103. doi: [10.1186/s12862-015-0382-5](https://doi.org/10.1186/s12862-015-0382-5)

Revision round #2

2019-04-25
Dear Dr. Tenaillon,

the manuscript has improved considerably and almost all the reviewers comments have been taken into account. I especially like your explanation of the reponse and target gene expression and it is good that this is broght up in the text. Also the Figure 2 is very helpful. I still have a few requests that I ask you to do, before I can recommend this preprint.

Page 9: Please clarify what this means: "as part of the Selection category the subset of DE genes displaying differential expression between shoot apical meristem Status for FE but neither for FL nor for FVL, an reciprocally".

P 17. Why lower residual heterozygosity would lead to more DE genes? I would expect the opposite.

P. 21 Please, clarify the connection between unique stretches of heterozygosity and inbreeding depression. Why do you think it is unlikely to create patterns of convergence?

Did you take the expression level into account in the analysis of convergence? The total expression level has an effect on when differential expression can be deteceted. I am mostly concerned that FT_candidates in general have higher exrpession and that is a potential reason why they are enriched. Please check whether this suspicion is correct (I hope not!).

Page. 22 Could also hybrids between different selection lines help to differentiate between cis and trans changes? Whole genome sequencing to identify de novo mutations?

Page 24. Do I interpret correctly that there were altogether 3 biological replicates if two years are pooled together? One of the reviewers was asking about the blocks and randomization. Please answer to the question and dicuss statistical power accordingly.

Page 29. are the parenthesis referring to interaction or nesting? Please clarify or provide a reference to explain how multiple pairwise contrasts result in linear decomposition.

Please clarify the meaning of notation in Figure 2 [SelF] U [SelM] : 1130, [SelF] U[SelF]|[StatusProg] : 2120, [Sel] : 2451

Language issues

First sentence of abstract, replace 'bases' with 'basis'

Second last row of abstract, space missing.

Remove periods from the section titles.

Reconsider the use of the word 'progenitor'. At least I first thought that the word is referring to the founder lines and that made understanding the results difficult. Could selection 'line/lineage' be better?

First sentence of results, change the order of words

Add space between number and unit, e.g. 7 ml, not 7ml (SI standard)

Months should be capitalized.

Best regards, Tanja Pyhäjärvi

Preprint DOI: [10.1101/461947](10.1101/461947)

Author's reply:

Dear Recommender,

We thank you for your positive answer to our resubmission. Enclosed is a newly revised manuscript that incorporates the minor final revisions that you asked. We are detailing below our answers. Note also that the R code corresponding to our differential expression analyses is now available in Figshare together with the necessary data to run it (Supplementary tables). We hope that this version is now suitable for your recommendation in PCI Evolutionary Biology.

Sincerely,

Maud Tenaillon

(1) Page 9: Please clarify what this means: "as part of the Selection category the subset of DE genes displaying differential expression between shoot apical meristem Status for FE but neither for FL nor for FVL, an reciprocally". - We clarified by "Within Status x Progenitor interactions category, we also considered as part of the Selection category the subset of genes differentially expressed among Status for FE but neither for FL nor for FVL and reciprocally ☐ DE genes among status for FL or FVL but not for FE". (2) P 17. Why lower residual heterozygosity would lead to more DE genes? I would expect the opposite. - Thanks for pointing this mistake. The residual heterozygosity is higher in F252 and not lower as previously stated: "This is in line with overall higher level of residual heterozygosity detected in the former". (3) P. 21 Please, clarify the connection between unique stretches of heterozygosity and inbreeding depression. Why do you think it is unlikely to create patterns of convergence? - We erased the sentence making a connection between residual heterozygosity and inbreeding as we felt this was complex to explain and not the point here, and we clarified our statement regarding convergence: "Note that in maize, stretches of residual heterozygosity have been shown to be either unique or shared by very few lines

(Brandenburg, et al. 2017). Therefore, except for shared streches between F252 and MBS, sorting of pre-existing alleles by differential selection between early and late populations should not translate into patterns of convergence between inbred lines". (4) Did you take the expression level into account in the analysis of convergence? The total expression level has an effect on when differential expression can be deteceted. I am mostly concerned that FT*candidates in general have higher exrpession and that is a potential reason why they are enriched. Please check whether this suspicion is correct (I hope not!). - We indeed verified as now stated in the text P.13: "Because the level of expression may affect our power to detect DE genes, we verified that FT*candidates were not expressed at a higher level than all transcripts taken together (P-value=0.615)". (5) Page. 22 Could also hybrids between different selection lines help to differentiate between cis and trans changes? Whole genome sequencing to identify de novo mutations? - Those are all interesting perspectives that we have now included at the end of the discussion P.22 with an additional reference: "In addition, allele specific expression in hybrids created from crosses between evolved genotypes would bring insights into cis- versus trans-regulation of gene expression (de Meaux, et al. 2005); and whole genome sequencing of ancestral and derived genotypes would allow identifying the origin of mutations and their fate through generations of selection. (6) Page 24. Do I interpret correctly that there were altogether 3 biological replicates if two years are pooled together? One of the reviewers was asking about the blocks and randomization. Please answer to the question and dicuss statistical power accordingly. - We went back to the reviewer's comments on the first round of reviews: "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?". This information was indeed missing and we are now providing more details about our experimental setting P.23: "The resulting progenies were sown and grown in the field at Université Paris-Saclay (Gif-sur-Yvette, France) during summer 2012 and 2013. The experimental design contained rows of 25 plants from the same progenitor. For each line, each of the progenitors was represented by nine rows that were randomized in three blocks". And P. 24: "we collected plants from the different blocks on a daily basis early morning (between 8:00 and 9:00 am). We randomly chose plants at the same developmental stage for a given progenitor". (7) Page 29. are the parenthesis referring to interaction or nesting? Please clarify or provide a reference to explain how multiple pairwise contrasts result in linear decomposition. - There is neither interaction nor nesting. All are independent contrasts as now repeatedly stated

P.29 (independent comparisons, independent contrasts) as well as in the legend of Figure 2. (8) Please clarify the meaning of notation in Figure 2 [SelF] U [SelM] : 1130, [SelF] U[SelF]|[StatusProg] : 2120, [Sel] : 2451 - This is explained in the legend: "Contrasts categories and total number of detected DE genes by category (calculated from the union ($\cup$) of all contrasts for that category) are shown in shaded boxes". (9) All language issues were corrected. Thanks for pointing them out. We maintain the use of the word 'progenitor' because that is the word we used in previous publications describing these experiments. Each progenitor was used to derived progenies by selfing, which we used in our experiments (cf. P6).

Revision round #1

2019-01-28

Tenaillon et al. (https://doi.org/10.1101/461947) present a study on gene expression changes in flowering time experimental selection lines of maize and provide evidence of convergent patterns of expression divergence. The overall question of expression changes and related convergent patterns are interesting and the material is suitable for further detailed studies on the genetic basis of the observed expression pattern changes. The experiment is conducted in field conditions, is replicated across years and using material from two selection experiments. Three reviewers have provided careful feedback and comments on the preprint manuscript. They find the manuscript unique, interesting, generally well written and worth recommendation. They also point that manuscript would benefit from clarifying the experimental setup and simplifying the presentation. In many ways the current version is very demanding to read for people who are not familiar with the experimental design, flowering time genetics and maize anatomy.

For the revision, please take into account all the reviewers comments, most importantly

1) Clarify the main results of the paper, especially the convergent patterns that multiple reviewers found to be the most interesting part of the work   2) Justify the PCA analysis or replace with a more suitable method   3) Justify the choice of genes for the qRT-PCR and change the order of RNASeq and qRT-PCR methods. If the choice of the genes is not justified, consider omitting the qRT-PCR results altogether   4) Explain the experimental design in more detail: at least sample

collection timing, germination procedure, number of plants, and design (random blocks?)   5) Cut the technical acronyms, omit the usage of FT   6) Allow access to Supplemental Tables   7) Regarding the low mapping rate, please explain why the mapping percent is low. Just as a suggestion, consider using methods for RNAseq transcript quantification that can handle multi-mapping reads (like Salmon)

Even though the questions presented in the first paragraph of the introduction are important, it would help reader to concentrate on the questions that are addressed in this manuscript. For example, can you quantify here the relative importance of gene expression and amino-acid sequence changes?

I am excited to see the revised version of the manuscript.

Tanja Pyhäjärvi

Preprint DOI: [10.1101/461947](10.1101/461947)

Reviewed by anonymous reviewer, 2018-12-13 17:10

[Download the review (PDF file)](#)
Reviewed by [Laura Shannon](#), 2018-12-13 20:20

[Download the review (PDF file)](#)
Reviewed by anonymous reviewer, 2018-12-13 20:21

[Download the review (PDF file)](#)
Author's reply:

[Download author's reply (PDF file)](#)