nature portfolio

Peer Review File

Lactuca super-pangenome provides insights into lettuce genome evolution and domestication

Corresponding Author: Dr Lisha Shen

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
I am satisfied with authors answers to my queries. In its current form the manuscript will make an interesting contribution to the existing knowledge base.

Reviewer #4

(Remarks to the Author)
I appreciate the authors' efforts in revising the manuscript. The manuscript has much improved, and their response convinces me that this study not only provides valuable genomic resources for Lactuca but also offers new insights into genome evolution. I have no further comments.
I also went through Reviewer 3's comments. Most of the concerns raised by Reviewer 3 have been adequately addressed, though one minor issue remains that, in my view, warrants further revision. Specifically, regarding Question (6), the reviewer expressed concern about the comparison of CMT2 or CHH methylation across all species. The authors merely stated in their response: "Other Lactuca species examined did not exhibit notable changes in CMT2 expression correlated with CHH methylation," without citing any figures or tables to support this conclusion. I recommend that the authors provide a figure to visually demonstrate these results.

Version 1:

Reviewer comments:

Reviewer #4

(Remarks to the Author)
Congratulations to the authors. They have thoroughly addressed all my concerns.

Response to Reviewers

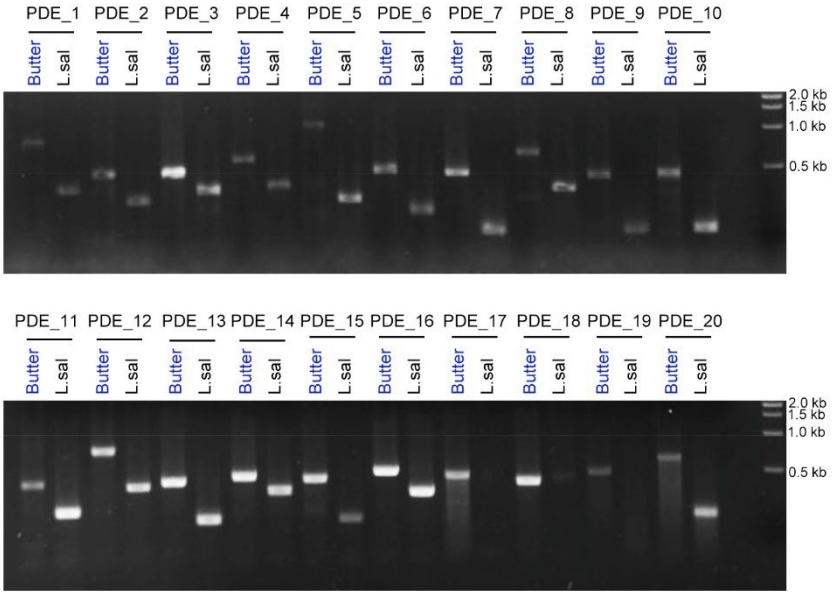
We would like to thank the Editor and reviewers for the time committed in reviewing this manuscript, and for the detailed suggestions on improving this manuscript. We have further revised the manuscript to fully address the concerns and criticisms.

Reviewer #2

The authors provided what I would consider partial responses to my queries:

>Reviewer:
(1) Authors: However, SyRI was less effective in identifying insertions (INS) and deletions (DEL),particularly in L. indica and L. virosa genomes, which exhibit significant divergence from the reference genome.
I am not sure how authors define 'less effective'. It looks like just raw numbers were compared and cuteSV identified more variants. Has it been proven that there are genuine variants identified by cuteSV but missed be SyRI?
>Author:
We thank the reviewer for this comment. To address this concern, we randomly chose 20 SVs identified by cuteSV but missed by SyRI, spanning across 9 chromosomes. Of these, 17 were confirmed through PCR analysis, supporting the reliability of the identified SVs and demonstrating the effectiveness of the cuteSV approach. This result has been incorporated into the revised manuscript (Supplementary Tables 4 and 6). Please also refer to the Relevant Figure below for the PCR validation results.

Relevant Figure

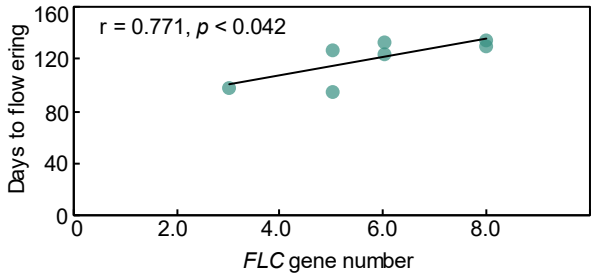


Legend: Verification of 20 randomly selected SVs identified by cuteSV but not by SyRI using PCR. Genomic DNA was extracted from butterhead lettuce (butter)

and one wild relative *L. saligna* (L.sal). Primers spanning the predicted deletions were designed based on SVs that indicate deletions in the wild relative compared to the butterhead reference genome, and primer sequences are listed in Supplementary Table S6. PCR assays successfully verified all selected SVs except PDE_17, PDE_18, and PDE_19, which failed to amplify any product from the genome of the wild relative.

>Reviewer:
(2) I don't feel like the initial concern 'Looking at Extended Data Fig. 5 days to flowering does not seem to correlate with copy number' was addressed. The authors explain how they chose data points (and what data was excluded), but the answer does not seem to answer the query.
>Author:
We thank the reviewer for this comment. To address the concern, we calculated the correlation between days to flowering and *FLC* copy number. Our analysis showed a correlation value of 0.771 with a *p*-value of < 0.042 (*Spearman*'s rank correlation coefficient), indicating a statistically significant relationship (See Relevant Figure below). We have included the relevant figure in Extended Data Fig. 5d in the revised manuscript.

Relevant Figure (Extended Data Fig. 5d)



Extended Data Fig. 5d Correlation between flowering time and *FLC* copy numbers (*Spearman*'s rank correlation coefficient).

Reviewer #4

The manuscript presents a study on the Lactuca super-pan genome, utilizing cutting-edge sequencing technologies to generate high-quality genome data and standardized analyses. While the data resources provided are valuable for lettuce breeding, the paper lacks significant innovation and does not substantially contribute to the current state of knowledge.

Major Concerns:

>Reviewer:
(1) As noted by Reviewer 2, the paper does not add much to the existing knowledge base. The analysis of biological processes using the super-pan genome, particularly the diploidization process, fails to yield new conclusions. The section titled “Similar diploidization effect across the Lactuca genus” discusses gene changes post-WGT

events but lacks a summarizing conclusion. The diploidization process has been extensively covered in previous literature, and this study does not provide new insights.
>Author:
We thank the reviewer for this comment. Although the diploidization process has been extensively studied, our work provides a novel perspective by leveraging the super-pan genome to analyze gene loss and retention patterns across multiple *Lactuca* species. This comparative approach offers new insights into how diploidization manifests at the pan-genome level, an aspect that has not been systematically explored in previous studies. Notably, our research provides novel insights by examining species within the *Lactuca* genus, which display remarkably diverse genome sizes, ranging from 2.1 Gb to 5.5 Gb. This level of variation is rare among other diploid crop species, such as maize, rice, cabbage, and even the diploid progenitors of wheat, which typically exhibit more uniform genome sizes (https://www.ncbi.nlm.nih.gov/datasets/genome/). Given this diversity, we aimed to investigate the role of diploidization in genome evolution post-WGT. Our analysis revealed a similar diploidization effect across the *Lactuca* genus, highlighting its contribution alongside long terminal repeat (LTR) expansion as a key driver of genome alterations during long-term evolution. In addition, we have revised the Results section and added the concluding sentence: "These results suggest that similarly strong diploidization has occurred across the *Lactuca* genus." according to the reviewer's suggestion.

Our subsequent analysis on transposon expansion showed that species with larger genomes experienced extensive LTR expansion—a well-documented phenomenon in genome evolution. Interestingly, we identified distinct, multiple insertion peaks of LTRs in *L. indica*, a pattern that has not been previously reported in genome studies. Moreover, while epigenetics is widely recognized as a key factor influencing transposon dynamics and genome size variation, few studies have provided concrete evolutionary examples. Our study further contributes to this field by identifying CMT2 as a potential regulator of LTR insertions, although further functional validation is required to fully establish its role in this process. We believe that these results provide new insights into genome evolution.

In addition, we identified abundant SVs related to domestication, which represent valuable resources for studying domestication-related traits in lettuce. Overall, we hope the reviewer agree that our work advances the understanding of *Lactuca* genome evolution by integrating a super-pan genome approach with comprehensive analyses of diploidization and transposon activity as well as identification of SVs related to domestication. The improved genome assemblies provided here also offer a robust resource for future studies on gene annotation, structural variation, and comparative genomics.

>Reviewer:
(2) The investigation into the impact of transposons on genome size and structural variations (SVs) is largely descriptive and mirrors traditional pan-genome studies without offering novel perspectives.
>Author:
Please refer to our response to Comment No. (1).

>Reviewer:

(3) While the identification of domestication-related PAVs is an interesting aspect, the methodology used is unclear. It appears that the authors rely on previously published selective swept regions to classify PAVs as domestication-related. This approach is rather simplistic. I recommend that the authors consider identifying domesticated PAVs based on relative allele frequencies, as domesticated PAVs should exhibit lower allele frequencies in cultivated populations compared to wild populations.

>Author:

In our study, domestication-associated PAVs were identified as follows. PAVs that are present or absent in each cultivated type compared to the wild progenitor were considered as domestication-associated. Based on allele frequencies, these domestication-related PAVs were categorized as private (present in only one accession), dispensable (present in 2-7 accessions), and core PAVs (present in all accessions) (Fig. 6b). We specifically focused on core PAVs shared across all analyzed horticultural types of cultivated lettuce, ensuring that the identified domestication-related PAVs are robust across diverse cultivated morphotypes. Our approach aligns with the pan-genome concept for constructing these domestication-related PAVs. We have added a detailed explanation of this approach in the "Method" section.

We thank the reviewer for the suggestion regarding the identification of domesticated PAVs based on relative allele frequencies and agree that further validation using larger and more diverse populations is necessary in future studies. We have addressed this limitation in the "Discussion" section, emphasizing the need for population-level analyses involving a broader set of genomes. Despite this limitation, our work provides a valuable starting point for identifying candidate PAVs associated with domestication and lays the groundwork for future studies.

>Reviewer:

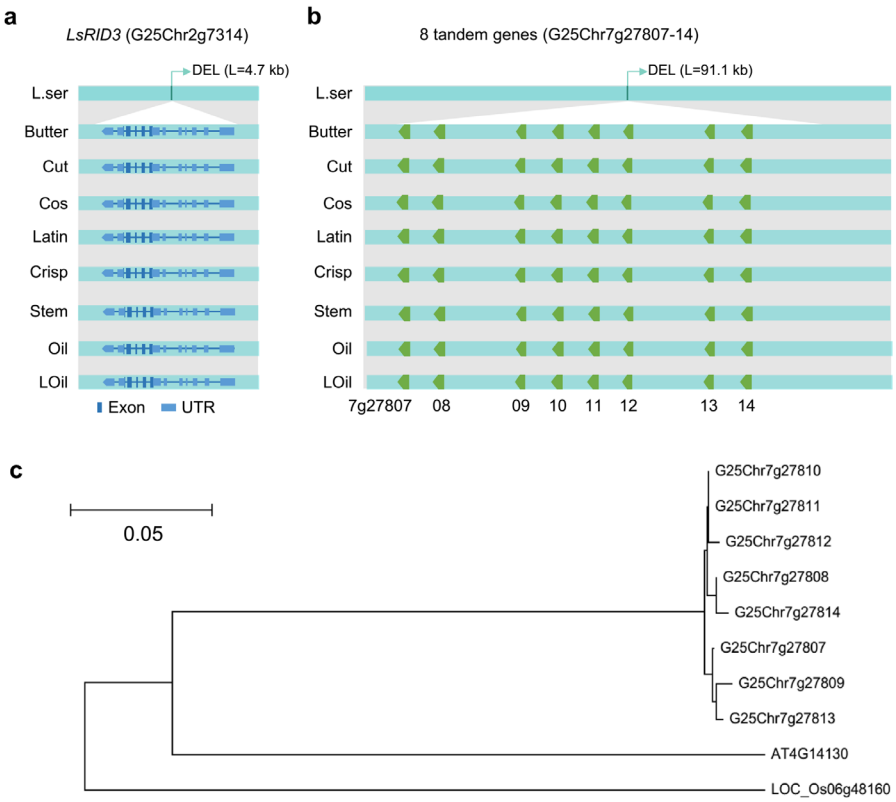
(4) The choice of the FLC gene as an example for discussing domestication-related PAVs is not compelling. Although FLC is functionally important, it has been extensively described in numerous genomic studies. I suggest the authors replace this example with another significant functional gene that has not been as thoroughly explored.

>Author:

As suggested, we have included two additional examples of domestication-related PAVs (See the Relevant Figure below) in the revised manuscript. First, a PAV contains *G25Chr2g7314*, a homolog of *AT3G49180* (*RID3*) (Relevant Figure a), which is involved in cell division regulation and heat response (Tamaki et al., 2009; Ohbayashi et al., 2017). Second, a deletion spanning over 90 kb includes eight tandemly repeated genes (Relevant Figure b,c), which are homologs of AT4G14130 involved in stress response (Mahalingam et al., 2003; Larkindale and Vierling, 2008). This figure has been included in Extended Data Fig. 6 in the revised manuscript.

These examples illustrate that domestication-related PAVs are prevalent across our pan-genome assembly. Although *FLC* has been extensively described in other genomic studies, its potential function and copy number variation in relation to flowering time in lettuce remain unexplored. Given its relevance to our study, we believe that retaining the analysis of *FLC* copy number and its association with flowering time across different lettuce species is important for the scope of this

manuscript. Further exploration of other domestication-related PAVs will be valuable to pursue in future studies.
Relevant Figure (Extended Data Fig. 6)



Extended Data Fig. 6 Additional examples of genes related to core domestication-associated PAVs. a, SVs in *G25Chr2g7314* between wild (*L. serriola*) and cultivated lettuce. b, A > 90 kb deletion encompassing 8 genes identified in wild (*L. serriola*) compared to cultivated lettuce. c, Phylogenetic tree based on protein sequences encoded by the 8 tandemly repeated lettuce genes of lettuce, along with homologs from *Arabidopsis* and rice.

Relevant reference:
Larkindale, J., and Vierling, E. (2008). Core genome responses involved in acclimation to high temperature. Plant Physiol 146, 748-761.
Mahalingam, R., Gomez-Buitrago, A., Eckardt, N., Shah, N., Guevara-Garcia, A., Day, P., Raina, R., and Fedoroff, N.V. (2003). Characterizing the stress/defense transcriptome of Arabidopsis. Genome Biol 4, R20.
Ohbayashi, I., Lin, C.Y., Shinohara, N., Matsumura, Y., Machida, Y., Horiguchi, G., Tsukaya, H., and Sugiyama, M. (2017). Evidence for a role of ANAC082 as a ribosomal stress response mediator leading to growth defects and developmental alterations in Arabidopsis. Plant Cell 29, 2644-2660.
Tamaki, H., Konishi, M., Daimon, Y., Aida, M., Tasaka, M., and Sugiyama, M. (2009). Identification of novel meristem factors involved in shoot regeneration through the analysis of temperature-sensitive mutants of Arabidopsis. Plant J 57, 1027-1039.

Reviewer #3

The paper “Lactuca super-pangenome provides insights into lettuce genome evolution and domestication” represents a valuable pangenome study of the Lactuca genus where authors compared 5 cultivated lectica sativa morphotypes and one landrace to the wild progenitor species Lactuca serriola as well as 4 others wild Lactuca species.
Cutting-edge methodologies for high-quality genome sequencing, assembly and genome analysis were used. Results and conclusion are adequately presented here, with a graph-based reference genomes that could be impactful for capturing genetic diversity across the different Lactuca species (and accessions), illustrating genetic variation.

I suggest that authors should consider or address the following points and concrns to make the paper definitely acceptable for publication:
>Author:
We appreciate the positive comments from the reviewer and insightful suggestions on improving this manuscript.

>Reviewer:
(1) Only a genome sequence from one single accession of the 5 wild Lactuca species is sequenced and analyzed here. While I am not asking that authors sequence genomes from further accessions of each wild Lactuca species, and enough sequence data are presented here, author should take this fact carefully in interpretations, discussions dans conclusions.
For example, page 14, lines 282- 291 and related figure & and Extended data, authors found more complete FLC copies in the domesticated lettuce morphotypes (diverging also in copy number between each other) than the solely compared wild progenitor accession of L. serriola. Moreover, in the later species it is argued that missing (truncated) FLC copies are more due to insertions/deletions in FLC genes, than rather new duplicated FLC gene copies in domesticated morphotypes. Nevertheless, one could conclude that “truncation” of FLC genes in the wild L. serriola may have happened after domestication or the most-likely explanation is that L. serriolia accessions (not compared here) diverge in copy number of FLC genes, among which the domesticated types have been selected.
Linking late or no flowering in the domesticated morphotypes to copy number variation of FLC genes is valid here (as it was also observed when comparing Chinese type oilseed rape (Brassica napus) to winter European type for example)).
>Author:
We thank the review for this suggestion. As suggested, we have been more cautious in interpreting our results and drawing conclusions, considering the limitations of using genome sequences from single accessions for each wild *Lactuca* species.

We agree with your observation regarding the presence of more intact *FLC* copies in domesticated lettuce morphotypes compared to the wild progenitor *L. serriola*. To address this, we have revised the relevant sentences to describe these findings more accurately, replacing "expansion of the *FLC* gene family" with "*FLC* copies." Additionally, we have added a discussion emphasizing the need for future studies involving multiple *L. serriola* accessions to explore this further.

>Reviewer:
(2) Page 7 lines 127-128:
The sentence: “Interestingly, gene family losses were significantly higher than gene family gains among the Lactuca genus”. is not valid as this super Pangenome analysis

does not address gene family gain. At best, these consist of gene retention (meaning deleted in one genome species and not the other considered one).

>Author:

As suggested, we have revised these sentences to use "gene retention" instead of "gene gain".

>Reviewer:

(3) Page 7 lines 141 to 144

a. Please do not use the term "gene gain" as there is no proof of new gene emergence and use instead "gene retention".

>Author:

As suggested, we have replaced "gene gain" with "gene retention" in the revised manuscript.

>Reviewer:

b. Moreover the gene losses (353) as compared to gene retention (139) estimated here for L. sativa (represented here by 5 morphotypes and one landrace) as compared to wild relatives (combining apparently 4 wild relatives) is not precise and may lead to wrong estimations. Instead, pairwise species comparisons of retained-lost genes, including pairs of wild Lactuca species, and mainly L. sativa/L. serriola would be of better impact and will lead to better classification-designation of the domestication related lost or retained genes

>Author:

As suggested, we have reperformed the analysis using pairwise comparisons between the main cultivated lettuce morphotypes (*L. sativa*: butterhead, cutting, cos, Latin, stem, and crisp lettuce) and the wild progenitor (*L. serriola*). Through this analysis, we identified 491 lost genes and 158 retained genes shared among the cultivated morphotypes. We have updated Figures 2d and 2e, and revised the relevant text in the manuscript.

>Reviewer:

(4) The chapter "Similar diploidization effect across the Lactuca genus", (pages 8 and 9) clearly argue that diploidization, occurring after the last palaeopolyploidization, clearly occurred before the divergence between the Lactuca species. This combined to the chapter that follows (about differential transposon expansion) between Lactuca species should be stressed in the discussion part. For example, how and why this has no or low impact on gene loss/retention in comparison to other species comparision like the different wheat species (Triticum and Aegilops) or the Brassica species (example B. rapa and B. oleracea).

>Author:

We thank the reviewer for this thoughtful comment. As suggested, we have expanded the Discussion section to discuss the impact of diploidization and transposon dynamics on gene loss/retention.

>Reviewer:

(5) In the Chapter "transposon expansion contributing to genome size" (pages 9 and 10)

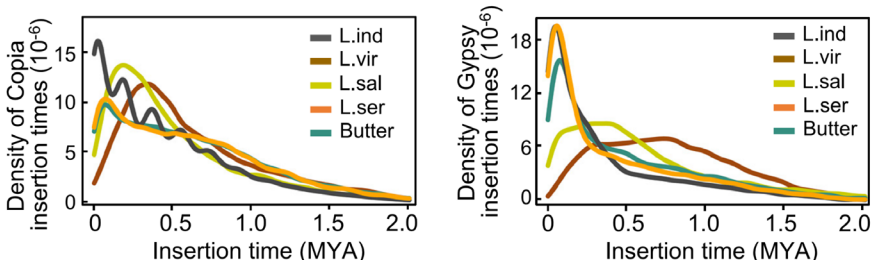
a. Since LTR-retrotransposons account for 82.6% of the total repetitive sequences, it is very important that author analyze and compare bursts of these LTR transposon expansions by comparing left and right LTR sequence divergence and inferring insertion dates. This is important and critical as it will tell us about period of high

activity leading to these differential TE spaces between species. Example of methods to infer in Charles et al. 2008, https://doi.org/10.1534/genetics.108.092304)

>Author:

As suggested, we used the recommended analysis to investigate the proliferation periods of retrotransposon superfamilies in *Lactuca* species. Our analysis revealed that Copia and Gypsy retrotransposons proliferated more recently in the *L. indica* genome, whereas they proliferated earlier in the genomes of *L. virosa* and *L. saligna* genome (See Relevant Figure below). We have included the relevant figure in Extended Data Fig. 3a in the revised manuscript.

Relevant Figure



Legend: Proliferation periods of the main retrotransposon superfamilies of Copia (left) and Gypsy (right) in *Lactuca* species.

>Reviewer:

(6) Page 9, lines 201 to 213

a. The decrease in CHH methylation in L. indica as compared to that in other species is obvious. Inferring the low CHH methylation to high transpositions and subsequently at the end of the page to lower expression of the CMT2 gene in L. indica are good indication for an explanation but may not represent the whole explanation or the only cause of the observed differences in TE content between species. Note that all of the species differ in genome size and the authors didn't show here comparison of the CMT2 or CHH methylations between all species. Authors can also present and argue about these other possible comparisons as well.

>Author:

We thank the reviewer for this suggestion. CHH methylation is one of many factors influencing TEs expansion and genomic size. While *L. indica* shows low *CMT2* expression, low CHH methylation, and high TE activities, other *Lactuca* species examined did not exhibit notable changes of *CMT2* expression correlated with CHH methylation. However, our current results do not exclude the effect of methylation on TE expansion and genomic stability, as complex mechanisms are involved in genomic stability, including chromosome rearrangements, the S-phase checkpoint, and double-strand-break repair machinery (Aguilera and Gómez-González, 2008), which may also be regulated directly or indirectly by DNA methylation (Zhou and Robertson, 2016). We have added this relevant discussion to the revised manuscript.

Reference:

Aguilera, A. & Gomez-Gonzalez, B. Genome instability: a mechanistic view of its causes and consequences. Nat Rev Genet 9, 204-17 (2008).

Zhou, D. & Robertson, K.D. Chapter 24 - Role of DNA methylation in genome

stability. in Genome Stability (eds. Kovalchuk, I. & Kovalchuk, O.) 409-424 (Academic Press, Boston, 2016).

>Reviewer:
(7) Page 14, lines 282-291
Please supply here 8 and in related discussion) comparison with FLC copy number and structural variations done in other species like Brassica species (Sun et al 2017, https://doi.org/10.1111/tpj.13669)
>Author:
As suggested, we have included a comparison between *FLC* copy number and structural variations in *Lactuca* species and relevant findings from *Brassica* species (Sun et al., 2017; Belser et al., 2018) and *Arabidopsis* species (Jiang et al., 2021), in the revised manuscript.

Reference:
Sun, F. et al. The high-quality genome of Brassica napus cultivar 'ZS11' reveals the introgression history in semi-winter morphotype. Plant J. 92, 452-468 (2017).
Belser, C. et al. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat. Plants 4, 879-887 (2018).
Jiang, X., Song, Q., Ye, W. & Chen, Z.J. Concerted genomic and epigenomic changes accompany stabilization of Arabidopsis allopolyploids. Nat. Ecol. Evol. 5, 1382-1393 (2021).

>Reviewer:
(8) Page 19, lines 306-308,
Please rephrase the sentence: “We reveal a similar diploidization effect across the Lactuca genus, and an expansion of transposable elements in contributing to their genome size variations” to: We reveal that diploidization that followed the last palaeopolyploidization has occurred much earlier than divergence between the Lactuca species whereas differential activity of transposable elements is the cause of the important genome size variation.
>Author:
We have revised the sentence following your suggestion.

>Reviewer:
(9) Page 16 line 332: misspelling “reference”
>Author:
We have revised it in the manuscript.

>Reviewer:
(10) Page 16 lines 346 to 349. Please take into account point number (1), as we didn’t analyze of representative accessions of L. serriola (only one accession is presented here) and whereas you argue initially for more truncation of FLC copies in L. serriola.
>Author:
We have revised this sentence as suggested in point number (1) to: “Consistently, the observed higher number of *FLC* copies in cultivated leafy-type lettuce, compared to wild lettuce, could be associated with their delayed flowering traits (Extended Data Fig. 5c), a desired characteristic for many leafy vegetable crops that has arisen through domestication.”.