

Genomic and ecological drivers of parallel arid adaptation in tree grapes (Vitaceae)

Corresponding Author: Professor Li-Min Lu

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

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This manuscript utilizes genomic data to explore the relationship between genome size, LTR diversity, and succulent traits in *Cyphostemma* plants. It integrates multiple data types, including genomics, phylogenetics, physiology, and ecology, presenting highly insightful perspectives. The entire paper features clear data presentation, figures, and writing. However, I found that some of the conclusions in the article are not rigorously supported by the relevant results, and some certain statistical findings require further examination of their reliability.

First, a large part of the paper is supported by the comparative genomic analysis of chromosome-level genome assemblies from 12 super-rosid species. Although 7 representative Vitaceae species are included, the selection of the other five outgroup species appears to lack representativeness or necessary justification. I noticed that these five species all seem to have relatively small genomes within their respective groups. Such deliberate selection could potentially introduce bias into the subsequent statistical conclusions. I would suggest that the authors reconsider whether to include these outgroups in the comparative analysis or provide sufficient justification for their selection strategy.

Second, whether a comparison between the two species, *C. dehongense* and *C. currorii*, can adequately represent the differences between succulent and non-succulent species was questionable. The two species originate from distinct regions and divergent phylogenetic lineages. Moreover, the lineage of *C. currorii* is not exclusively composed of succulent species. Therefore, the observed genomic differences between them may be more reflective of lineage-specific variations rather than distinctions related to the succulent trait. Based on the current comparative results of these two species, I believe the findings can only serve as a description of differences between these specific species. Extrapolating these differences to establish a link with the succulent trait requires extreme caution. I would suggest that the authors either select more closely related species for interspecific comparison, include additional species in the analysis, or temper the strength of claims derived from these results.

More detailed comments:

Figure 1b: This figure may cause ambiguity, as the purple bar plot could be misinterpreted as a stacked bar plot. It is recommended to use distinct legends to present the two datasets.

Line 139-140: I can't get such a conclusion from Fig 1e, so I would discourage citing this figure here. However, I note that no other sections in the main text describe or utilize the results from this figure. It is recommended to reconsider the organization of figures in the main text. Although the data in Supplementary Table 6 appear to strongly support this viewpoint, I still recommend conducting statistical tests and presenting the results. Furthermore, as I mentioned above, I suspect that the observed significant differences are closely tied to the selection of the outgroup. It would suggest the authors to verify the robustness of this conclusion across different outgroup choices.

Line 141-142: Based on my understanding, the curve Cyde-Cyde and the curve Ccur-Ccur in Supplementary Fig. 8 represent genome duplication events that occurred after the divergence of *Cyphostemma* from *Arabidopsis thaliana*. This appears to contradict the authors' statement "no ancient WGDs in Vitaceae".

Line 165-166: I don't think there are 77 species shown in Fig. 2a.

Line 166-168: I suspect this conclusion stems from Simpson's Paradox in statistics. As mentioned later in the main text, the African clade and the Malagasy clade exhibit "clade-specific LTR-RT dynamics", raising doubts about the validity of combining these two groups for statistical analysis. Based on the current results, if statistical analyses were conducted separately for each clade, it is likely that such correlations would not be possible to establish. Moreover, There are gaps between LTR diversity and massive proliferation of a limited number of LTR-RT families. Is it possible due to the deletion?

Line 168-171: Extended Data Fig. 2a can't support this point.

Line 171-174: I can't find any evidence to support younger insertion ages in the succulent species. I think there should be a statistical test here to support it. Furthermore, just basing on the current results here, it is hard to link the genomic data with ecological niche divergence.

Line 175-176: How to support the "elevated death rate"?

Extended Data Fig. 3: I think except for SIRE and Tekay, the number of samples is too small to give a boxplot for each gene family.

Line 177-178: There need more evidences to show the link between the elevated death rate in Tekay and ORF diversity.

Line 194: If over half of the total intronic LTR-RT length is from the Ale family in Ccur, it is hard for me to understand that the proportion of LTR-RT in intron is about 60% (Supplementary Table 6), considering the LTR-RT family composition in the whole genome shown in Extended Data Fig. 5.

Line 198-199: I don't think it is "the dual preferences", because the proportion of introns and intergenic regions in the whole genome are not equal. Thus, I think the statistical tests in Extended Data Fig. 2 between Ale intronic and Ale intergenic are nonsense.

Line 223: Here, it is necessary to provide justification for the correlation between extra-long introns and Ale insertions. Specifically, I note that the expression level in the moderate-length (250-4000) group appears to be the highest. Therefore, I suggest that the authors characterize the distribution of Ale elements across genes with different intron lengths to support the argument presented therein.

Figure 4c&d: I find it somewhat difficult to understand why the authors chose the length of intronic LTR of (Ccur-Cyde) as the y-axis variable at this point. Theoretically, under the neutral mutation hypothesis, longer sequences can accumulate a greater absolute number of mutations. Therefore, I am personally inclined toward using Ccur/Cyde or (Ccur-Cyde)/Cyde as the variable, as these ratios would provide a more meaningful basis for comparison.

Line 300: The yellow internal branch shown in both Figure 2a and Extended Data Fig. 8b is quite confusing for me. What mechanisms could have enabled the common ancestor of currently species with relatively small GS to reconstruct a substantially larger GS? Could the authors provide explanations for this finding from both methodological and biological perspectives?

Line 306: In this paragraph, I didn't get any evidence to support the link between LTR-RT compositions and environmental variables.

Fig. 5: It seems that this figure is as same as Extended Data Fig. 10.

Reviewer #2

(Remarks to the Author)

Yu et al. investigate whether LTR retrotransposons—especially intronic insertions—drive genome-size expansion in *Cyphostemma* (Vitaceae) and whether that expansion contributes to succulence. They assemble a *Cyphostemma* phylogeny, compile 139 genome-size records for Vitaceae, generate two chromosome-scale genomes (the non-succulent rainforest climber *C. dehongense* and the succulent desert shrub *C. currorii*), and augment these with shallow whole-genome sequencing across ~77 species to profile repeat composition, alongside RNA-seq, bisulfite-seq, leaf micromorphology, and climate/soil data. The study reports that Vitaceae—particularly *Cyphostemma*—show disproportionate intron expansion tied to LTR activity rather than recent whole-genome duplication; that a few LTR families (Tekay, Ogre, SIRE) dominate intergenic genome inflation while Ale elements are unusually enriched in introns; that genes with extra-long introns are down-regulated and more methylated; that cyclin/auxin modules retain intronic Ale and show expression divergence consistent with larger cells, lower stomatal density, and altered vasculature; and that succulence often precedes genome-size regime shifts, with later upsizing associated with arid, seasonal climates. Overall, it is an ambitious, multi-layered analysis linking TE ecology, genome architecture, trait anatomy, and environment; the intronic-LTR angle (Ale within introns) is distinctive and strengthened by the absence of recent WGD; clade-level "TE community" patterns (Tekay vs. Ogre) are clear within a thoughtful phylogenetic framework; and the leaf micromorphology provides a concrete phenotypic anchor.

However, the study is highly complex, and it is not always clear which datasets were used where, for what purpose, and how specific claims are supported. In particular, the manuscript advances a very specific hypothesis—that intronic LTRs (notably

Ale) help drive succulence—without clearly stating the priors that motivated this focus. Clarifying the origin of this hypothesis and providing a concise “data-to-claim” roadmap would make the narrative easier to follow. Below are my detailed comments.

The phylogeny using 1,367 single-copy genes across 151 taxa: single-copy genes were identified in 12 species and a few additional ones for which transcriptomes were assembled. This step where different annotations are used already introduces a bias. And then for the up to 151 species “shallow” read data was somehow mapped to these genes using the hybpipe pipeline to identify single copy genes in those taxa. Is that correct? How can it be ensured like that these genes are really single copy?

Intron length analysis and genome annotations: are the assemblies/ annotation of the 12 species comparable? All long read, de novo annotated with external evidence, high N50 and BUSCO scores, etc. ?

What is the analysis of the longest intron aiming at? These might also be fused gene models/ annotation artefacts.

The authors claim that it is remarkable that LTR-RT proliferation is the primary driver of genome and intron expansion as “LTR-RTs are typically enriched in intergenic regions and rarely generate extra-long introns in angiosperms, suggesting a unique evolutionary mechanism favoring intronic retrotransposon colonization in Cayratieae.” Reference #26 does mention TEs within introns in other crops. In the two mentioned references, I don't find proof that LTR-RTs are rarely within introns. Please make sure this claim is accurate.

The central inference that Ale preferentially colonizes introns in *Cyphostemma* is supported by two chromosome-level assemblies (*C. dehongense*, *C. currorii*) only. The ~77 additional species are profiled with shallow, reference-free repeat clustering, which estimates family abundances but cannot localize insertions (intronic vs intergenic). Maybe I'm misunderstanding but if not, I think it should be made more clear that all the claims of the manuscript are about these two assemblies and not a general analyses of the genus. I think right now, the manuscript makes it look like it it.

What does shallow sequencing mean? How deep?

Enrichment analysis of genes with introns elongated by LTR-RTs: how was the functional enrichment done? What was the background? All genes or genes without long introns? It must be ensured here that the genes with long introns are intact gene models and not annotation artefacts before concluding about expression

It seems like the authors compiled a list of “succulence genes” from the literature. Is that correct? If so, this should be provided as a supplementary table. But also this feels a little like cherry picking. I think better would be to pull all the genes that deviate from the trend where genes with longer LTR-RT insertions in *C. dehongense* tend to exhibit contraction or less elongation in *C. currorii* and do a functional enrichment of those and see what they are enriched for instead of the top-down approach.

Data availability: the link to figshare is missing

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I would like to thank the authors for well responding to my suggestions and confusions in the previous round of review. I believe that the current manuscript has significantly improved in the rationality of data presentation and the reliability of conclusion. However, there are still a few issues that I hope the authors can consider.

1.Fig. 2b: Is the statistical data in the figure based on the entire lineages or representative species within it? If based on the entire lineages, further explanation is needed on how this proportion value was obtained; if based on representative species, please specify the species names.

2.Line 199-201: Fig. 2c can't support this sentence.

3.Line 202-204: In Fig. 3b, I noticed that the Tekay data from *C. dehongense* is too small to compare with *C. currorii*. How did the authors reach this conclusion regarding ORF? Moreover, for the subsequently emphasized Ale, the differences in ORF characteristics between the two are not significant. Does this imply that the subsequent impact of intronic LTR on expression is not correlated with methylation?

4.Line 210-211: As it said, LTR patterns “remain evolutionarily conserved”. However, significant diversity in succulence exists within both the continental African clade and the Malagasy clade. Does this suggest that the succulence is not caused by LTRs? I find this the most perplexing aspect of the current manuscript and hope the authors can provide a more plausible explanation.

5.In the section, “Larger genomes selected by arid habitats”, most conclusions are based on the comparison of succulence within lineages, yet this aspect has not been addressed in the previous sections and lacks associated conclusions. How can the authors prove that the GS differences within lineages indeed reflect variations in LTRs, especially intronic LTRs? As noted in the previous comment, the authors also require more robust evidence or clarification concerning the role of LTR diversity within lineages. Additionally, I noticed that the author's team recently published related research in JIPB, and the work in this section of the two articles appears very similar. Can that article provide any additional support for the conclusions in this section?

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

Thank you for revising the manuscript. I believe all my comments were addressed well and improvement were incorporated into the manuscript and supplementary material.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have comprehensively addressed all major concerns and suggestions from the previous rounds. I find no significant additional modifications necessary at this stage.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Manuscript No.: NCOMMS-25-60992-A

Title: Genomic and ecological drivers of parallel arid adaptation in tree grapes

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This manuscript utilizes genomic data to explore the relationship between genome size, LTR diversity, and succulent traits in *Cyphostemma* plants. It integrates multiple data types, including genomics, phylogenetics, physiology, and ecology, presenting highly insightful perspectives. The entire paper features clear data presentation, figures, and writing. However, I found that some of the conclusions in the article are not rigorously supported by the relevant results, and some certain statistical findings require further examination of their reliability.

First, a large part of the paper is supported by the comparative genomic analysis of chromosome-level genome assemblies from 12 super-rosid species. Although 7 representative Vitaceae species are included, the selection of the other five outgroup species appears to lack representativeness or necessary justification. I noticed that these five species all seem to have relatively small genomes within their respective groups. Such deliberate selection could potentially introduce bias into the subsequent statistical conclusions. I would suggest that the authors reconsider whether to include these outgroups in the comparative analysis or provide sufficient justification for their selection strategy.

Second, whether a comparison between the two species, *C. dehongense* and *C. currorii*, can adequately represent the differences between succulent and non-succulent species was questionable. The two species originate from distinct regions and divergent phylogenetic lineages. Moreover, the lineage of *C. currorii* is not exclusively composed of succulent species. Therefore, the observed genomic differences between them may be more reflective of lineage-specific variations rather than distinctions related to the succulent trait. Based on the current comparative results of these two species, I believe the findings can only serve as a description of differences between these specific species. Extrapolating these differences to establish a link with the succulent trait requires extreme caution. I would suggest that the authors either select more closely related species for interspecific comparison, include additional species in the analysis, or temper the strength of claims derived from these results.

[Response: Thank you for your insightful feedback and constructive suggestions. We appreciate your positive comments on our efforts to integrate multiple datasets and the clarity of data presentation. Regarding the concerns you raised, we would like to address them as follows. Please note that the line numbers indicated below refer to the track-changed version of the](#)

manuscript. Additionally, we have submitted a clean version of the main text for your convenience.

First, we understand your concern that the five outgroup species with relatively small genomes might affect the robustness of our findings. In response, we have expanded the dataset to include additional high-quality genome assemblies from super-rosids as outgroups (please see details in the point-to-point response) to ensure a more reasonable comparison across the super-rosid species. In addition, we have clarified and justified our outgroup selection in the revised manuscript (Lines 932–935). Importantly, we re-evaluated our statistical analyses using the expanded dataset. The Wilcoxon test consistently indicates that Vitaceae species have a significantly higher intron/exon ratio relative to non-Vitaceae super-rosids (Fig. 1b). Furthermore, the ANCOVA test shows that the variation in intron length/exon length ratio is primarily associated with taxonomic affiliation (Vitaceae) and the length of intronic LTR-RTs, while showing little dependence on overall genome size (Supplementary Table 7).

We appreciate your comment regarding the comparison between *C. dehongense* and *C. currorii*. We agree that the genomic differences between the two species may, at least in part, reflect lineage-specific variation rather than differences strictly attributable to the succulent trait. Due to current sampling constraints, it is unfortunately not feasible to include additional closely related non-succulent species at this stage. The non-succulent species sister to the lineage of *C. currorii* is restricted to tropical Africa, where obtaining fresh material suitable for high-quality genome or transcriptome sequencing remains challenging. In light of this limitation, we have revised the manuscript to temper the strength of our conclusions derived from this interspecific comparison and have avoided over-interpreting these results as general patterns associated with the evolution of succulence (e.g., Line 36 added “likely”; Lines 411–412 “elucidates” changed to “provides insight into”, “has” changed to “may have”; Line 415 “the effects of” changed to “a potential role for”). Additionally, we now emphasize that broader taxon sampling within *Cyphostemma*, especially including succulent and non-succulent sister lineages from both continental Africa and Madagascar will be essential for robustly testing hypotheses related to the genomic basis of succulence. This perspective is explicitly highlighted as an important direction for future research (Lines 428–429).

We believe these revisions substantially improve the rigor, transparency, and interpretative balance of the manuscript. The following are the point-to-point responses.

More detailed comments:

Figure 1b: This figure may cause ambiguity, as the purple bar plot could be misinterpreted as a stacked bar plot. It is recommended to use distinct legends to present the two datasets.

Response: Thank you for this helpful suggestion. We agree that the original presentation of

Figure 1b could potentially cause ambiguity. In the revised manuscript, after incorporating additional super-rosid outgroup species, we have redesigned Figure 1b to present the differences between Vitaceae and outgroup species in intron length/exon length ratio, average intron length, and intronic LTR-RT length using separate boxplots (Fig. 1b). This revised visualization improves clarity and avoids potential misinterpretation.

Line 139-140: I can't get such a conclusion from Fig 1e, so I would discourage citing this figure here. However, I note that no other sections in the main text describe or utilize the results from this figure. It is recommended to reconsider the organization of figures in the main text. Although the data in Supplementary Table 6 appear to strongly support this viewpoint, I still recommend conducting statistical tests and presenting the results. Furthermore, as I mentioned above, I suspect that the observed significant differences are closely tied to the selection of the outgroup. It would suggest the authors to verify the robustness of this conclusion across different outgroup choices.

Response: Thank you for this thoughtful and constructive comment. We agree that this conclusion was not sufficiently supported by the original Fig. 1e. In response, we have redesigned the figure and revised the main text accordingly. Specifically, we have removed Fig. 1e and added boxplots comparing intron length/exon length ratio, average intron length, and intronic LTR-RT length to Fig. 1 to better support the key conclusions in the manuscript.

Following your recommendation, we have also verified the robustness of our conclusions by adding additional outgroup species. The revised outgroup dataset now includes all available super-rosid genomes from Phytozome, NCBI, and Rosaceae genome database that meet our strict quality criteria: chromosome-level assemblies using long-read sequencing data; sequencing depth > 30× (except for *Vicia faba* with a huge genome of 11.9 Gb); assembly scaffold N50 > 5 Mb; complete BUSCO > 90%; gene annotation corrected with external evidence; repeat annotations validated by multiple lines of evidence. Several originally included outgroup genomes were excluded because they didn't meet these criteria. Detailed information on the revised outgroup genomes has been added to Supplementary Table 6.

With a total of 55 outgroup species included in the updated analysis, the Wilcoxon test continues to show significant differences in intron length/exon length ratio between Vitaceae and non-Vitaceae species (Fig. 1b). Furthermore, the ANCOVA test indicates that this pattern represents a Vitaceae-specific feature and is significantly associated with the length of intronic LTR-RTs (Supplementary Table 7). These results are now explicitly presented and discussed in the revised manuscript (Lines 154–158).

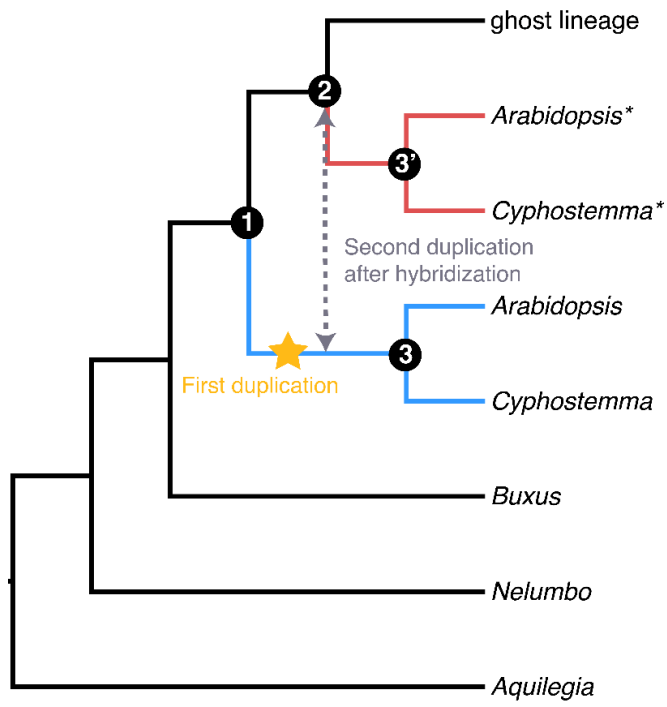
Line 141-142: Based on my understanding, the curve Cyde-Cyde and the curve Ccur-Ccur in Supplementary Fig. 8 represent genome duplication events that occurred after the divergence

of *Cyphostemma* from *Arabidopsis thaliana*. This appears to contradict the authors' statement “no ancient WGDs in Vitaceae”.

Response: Thank you for pointing out this important issue and for helping us clarify the interpretation of the results. We agree that the Cyde-Cyde and Ccur-Ccur curves in Supplementary Fig. 8 do suggest a genome duplication event (WGD) that occurred after the divergence of *Cyphostemma* from *Arabidopsis thaliana*.

Importantly, based on our syntenic analysis and a previous comparative genomic study (Zhu et al., 2023), we can confirm that no additional WGDs occurred after the divergence of *Vitis* and Cayratieae. We therefore maintain the inference that the genomic expansion and intron expansion observed in Cayratieae are not attributable to a lineage-specific WGD event, but are instead driven by other mechanisms, such as transposable element dynamics.

We would like to further discuss the potential “duplication event that occurred after the divergence of *Cyphostemma* from *Arabidopsis thaliana*”. Previous study suggested that the only WGD event that *Vitis* has experienced was the γ triplication event, a well-established event shared by all core eudicots, with no subsequent WGDs following the divergence of *Vitis* and Cayratieae. Therefore, we propose that the peaks of Cyde-Cyde and Ccur-Ccur curves observed in the 4DTv distributions are most parsimoniously interpreted as corresponding to the γ event. The suspicious position of the peaks may reflect the two-step process of the γ event, in which an initial WGD was followed by a second polyploidization involving a third genome, potentially through hybridization with a ghost lineage that branched off the core eudicot ancestral line before the initial tetraploid event (Chanderbali et al., 2022). Under this scenario, *Arabidopsis* and *Cyphostemma* diverged at node 1, and the second step of the γ event happened after their divergence (see Fig. R1 below, which has been added to our manuscript as Supplementary Fig. 8c). This would result in gene duplications after node 1 and consequently generate a peak in 4DTv distributions in Supplementary Fig. 8b without invoking a post-divergence WGD in Vitaceae. To further clarify this interpretation, we have added Supplementary Note 1 to explicitly discuss the two-step triplication γ event, integrating evidence from both our analyses and previous studies.



[Fig. R1 Our illustration for the two-step process of the \$\gamma\$ event shared by core eudicots. “*” indicates the paralogous gene lineages caused by WGD.](#)

[References:](#)

[Zhu, S. et al. Chromosome-level reference genome of *Tetrastigma hemsleyanum* \(Vitaceae\) provides insights into genomic evolution and the biosynthesis of phenylpropanoids and flavonoids. *Plant J.* 114, 805–823 \(2023\).](#)

[Chanderbali, A. S. et al. *Buxus* and *Tetracentron* genomes help resolve eudicot genome history. *Nat. Commun.* 13, 643 \(2022\).](#)

Line 165-166: I don’t think there are 77 species shown in Fig. 2a.

[Response: Thank you for pointing this out. We apologize for the lack of clarity in the original text. For improved visualization, the phylogeny shown in Fig. 2a was pruned, and therefore includes fewer than the total number of species analyzed. The complete dataset used for the analysis of LTR-RT composition comprises 74 *Cyphostemma* species \(see Supplementary Table 8\) and a phylogeny showing all these species is provided in Extended Data Fig. 8. We have revised the manuscript to explicitly clarify this point and to cite Supplementary Table 8 and Extended Data Fig. 2 in relevant statement \(Lines 173–175\). We have retained Fig. 2a because, despite the pruning, it effectively illustrates the overall patterns of LTR-RT composition across *Cyphostemma*.](#)

Line 166-168: I suspect this conclusion stems from Simpson's Paradox in statistics. As mentioned later in the main text, the African clade and the Malagasy clade exhibit “clade-specific LTR-RT dynamics”, raising doubts about the validity of combining these two groups for statistical analysis. Based on the current results, if statistical analyses were conducted separately for each clade, it is likely that such correlations would not be possible to establish. Moreover, There are gaps between LTR diversity and massive proliferation of a limited number of LTR-RT families. Is it possible due to the deletion?

Response: Thank you for this insightful comment. We agree that combining the continental African and Malagasy clades in a single statistical analysis may lead to misleading conclusions due to their clade-specific LTR-RT dynamics. Accordingly, in the revised manuscript we have removed the conclusion regarding the correlation between LTR-RT diversity and genome size (Lines 180–182).

Regarding gaps between LTR-RT diversity and the massive proliferation of a limited number of LTR-RT families, we agree that deletion processes may partly influence the observed diversity patterns. However, we consider the predominant driver of these patterns to be the recent or ongoing proliferation of a small number of LTR-RT families. In plant genomes, LTR-RTs are primarily removed through unequal recombination (UR) and illegitimate recombination (IR) (Ma et al., 2004). These processes typically generate solo LTRs (in UR) or incomplete elements (in IR), both of which remain detectable and are included in our analysis using RepeatExplorer2.

Our results indicate that a small number of LTR-RT families account for the majority of the LTR-RT content (Supplementary Table 8). If additional families had undergone extensive proliferation in the past, it is unlikely that deletion alone, even if substantial, would have reduced their contribution to such a minor proportion of the genome without leaving detectable signatures. Therefore, while deletion may contribute to shaping LTR-RT diversity, our data support the conclusion that the dominance of a few highly proliferative families is the primary factor underlying the observed patterns. We have revised the manuscript accordingly (Lines 183–190).

Reference:

Ma, J., Devos, K. M. & Bennetzen, J. L. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res. 14, 860–869 (2004).

Line 168-171: Extended Data Fig. 2a can't support this point.

Response: Thank you for pointing this out. We apologize for the unclear reference of figures

and tables in the original manuscript. We have now clarified the citations to ensure that each statement is properly supported by the corresponding data. Extended Data Fig. 2a is cited to illustrate the greater length of these elements, while Supplementary Table 8 is cited to demonstrate their higher genomic abundance. We have updated the corresponding sentences to clearly distinguish the evidence provided by the figure and the table (Lines 174–175).

Line 171-174: I can't find any evidence to support younger insertion ages in the succulent species. I think there should be a statistical test here to support it. Furthermore, just basing on the current results here, it is hard to link the genomic data with ecological niche divergence.

Response: Thank you for this valuable comment. We agree that the original statement regarding younger insertion ages in succulent species was not sufficiently supported. We have added a Wilcoxon test in the revised manuscript to more rigorously assess differences in overall insertion ages between succulent and non-succulent species (Fig. 3a).

We also agree that linking genomic data with ecological niche divergence at this point appeared too early and not sufficiently supported by the current analyses. Accordingly, we have removed this statement to avoid potential misunderstanding (Lines 212–213).

Line 175-176: How to support the “elevated death rate”?

Response: Thank you for raising this point. In the revised manuscript, we have added statistical analyses to explicitly test differences in LTR-RT death rates, the results of which are presented in Extended Data Fig. 3. Because sample sizes for most LTR-RT families are insufficient to support robust statistical comparisons, we limited this analysis to two families with adequate representation, *SIRE* and *Tekay* in *C. currorii*. Under this framework, *Tekay* shows a statistically higher death rate than *SIRE*. We have revised the text accordingly (Lines 216–217).

Extended Data Fig. 3: I think except for *SIRE* and *Tekay*, the number of samples is too small to give a boxplot for each gene family.

Response: Thank you for this comment. We agree that boxplots are not appropriate for LTR-RT families with small sample sizes. In the revised manuscript, we have replaced boxplots with individual value plots for all families with small samples, while retaining boxplots only where sample sizes are sufficient (Fig. 3a, Extended Data Figs. 3 and 5).

Line 177-178: There need more evidences to show the link between the elevated death rate in *Tekay* and ORF diversity.

Response: Thank you for this feedback. We agree that our current analyses do not provide sufficient evidence to support the link between the elevated death rate and non-autonomous

elements. To avoid overinterpretation, we have tempered the related statements and revised the subsequent discussion accordingly (Lines 220–221).

Line 194: If over half of the total intronic LTR-RT length is from the *Ale* family in *Ccur*, it is hard for me to understand that the proportion of LTR-RT in intron is about 60% (Supplementary Table 6), considering the LTR-RT family composition in the whole genome shown in Extended Data Fig. 5.

Response: Thank you for this comment and for pointing out the potential source of confusion. We apologize for the unclear terminology in the original Supplementary Table 6. In that table, “Proportions of LTR-RT in intron” was intended to represent the proportion of intron sequences that consist of LTR-RTs in total intron length. To avoid further misunderstanding, we have changed the label to “Proportion of intron occupied by LTR-RTs” in Supplementary Table 6.

In *C. currorii*, the total length of LTR-RT in the genome is 1,240.36 Mb, of which 142.32 Mb is intronic LTR-RT (Supplementary Table 12). Among these, intronic *Ale* LTR-RTs occupy 77.78 Mb, representing 54.65% of the intronic LTR-RT content. The total intron length is 233.22 Mb, meaning that about 61.0% of the intron length consists of LTR-RTs. The lengths of each LTR-RT family in intergenic and intronic regions are provided in Supplementary Table 12.

Line 198-199: I don’t think it is “the dual preferences”, because the proportion of introns and intergenic regions in the whole genome are not equal. Thus, I think the statistical tests in Extended Data Fig. 2 between *Ale* intronic and *Ale* intergenic are nonsense.

Response: Thank you for this insightful comment. We agree that the original statement of “the dual preferences” of *Ale* was potentially misleading. A larger proportion of intergenic regions can indeed lead to a greater accumulation of intergenic LTR-RT elements.

To address this, we used the Chi-square test to explicitly examine the bias of LTR-RT insertions in intronic and intergenic regions. This analysis revealed a significant enrichment of *Ale* insertions in introns, whereas other abundant families generally showed a bias toward intergenic regions (Supplementary Table 12). Moreover, we found no evidence of large-scale deletions of *Ale* insertions driving this pattern, as *Ale* lineages did not show significantly higher solo/intact ratios or low fraction/intact ratios in intergenic regions (Supplementary Table 10). Together, these results suggest that the observed bias is more likely due to lineage-specific insertion preferences rather than post-insertion deletions.

To avoid confusion, we have revised the manuscript to remove the term “dual preferences” and instead state that certain *Ale* lineages are more likely to accumulate in introns (Lines 242–243; Lines 249–250).

Regarding the original Extended Data Fig. 2 (Extended Data Fig. 5 in the revised manuscript), we agree that separating intronic and intergenic *Ale* elements was unnecessary and potentially confusing in this context. We have therefore merged the data for intronic and intergenic *Ale* elements and present a combined boxplot, which more clearly illustrates the differences across families.

Line 223: Here, it is necessary to provide justification for the correlation between extra-long introns and *Ale* insertions. Specifically, I note that the expression level in the moderate-length (250-4000) group appears to be the highest. Therefore, I suggest that the authors characterize the distribution of *Ale* elements across genes with different intron lengths to support the argument presented therein.

Response: Thank you for this insightful feedback. We agree that additional justification is needed to support the proposed link between extra-long introns and *Ale* insertions. To address this, we have calculated the LTR-RT compositions across genes grouped by intron lengths ($\leq$ 250bp, 250-4,000bp, and $>$ 4,000 bp).

This analysis shows high percentages of *Ale* in all the intronic LTR-RT fragments for all three gene groups. Notably, genes with extra-long introns ($>$ 4,000 bp) exhibit the highest percentage of *Ale* insertions among intronic LTR-RTs (Supplementary Fig. 9). These results provide direct support for our statement that *Ale* elements are preferentially associated with genes containing longer introns, thereby strengthening the interpretation presented in this section. The manuscript has been revised accordingly (Lines 238–240).

Figure 4c&d: I find it somewhat difficult to understand why the authors chose the length of intronic LTR of (Ccur-Cyde) as the y-axis variable at this point. Theoretically, under the neutral mutation hypothesis, longer sequences can accumulate a greater absolute number of mutations. Therefore, I am personally inclined toward using Ccur/Cyde or (Ccur-Cyde)/Cyde as the variable, as these ratios would provide a more meaningful basis for comparison.

Response: Thank you for raising this important point. We agree that, under the neutral mutation hypothesis, comparisons based on absolute length of LTR-RT elements may be confounded, as longer sequences are expected to accumulate more mutations simply due to their size. Following your constructive suggestion, we have revised the analysis by using Ccur/Cyde as the y-axis variable in the updated version of Fig. 4. The new results still show a positive and significant correlation between the length of intronic LTR-RT insertions in *C. dehongense* and *C. currorii*, which supports our original conclusion while providing a more appropriate basis for comparison.

Line 300: The yellow internal branch shown in both Figure 2a and Extended Data Fig. 8b is

quite confusing for me. What mechanisms could have enabled the common ancestor of currently species with relatively small GS to reconstruct a substantially larger GS? Could the authors provide explanations for this finding from both methodological and biological perspectives?

Response: We appreciate your feedback, and we apologize for the confusion. The yellow internal branch in Fig. 2a and Extended Data Fig. 8b does not represent the actual genome size of the reconstructed common ancestor. Instead, it indicates the evolutionary direction of genome size based on the Ornstein-Uhlenbeck (OU) model, which infers the evolutionary trend towards an optimal genome size under selection.

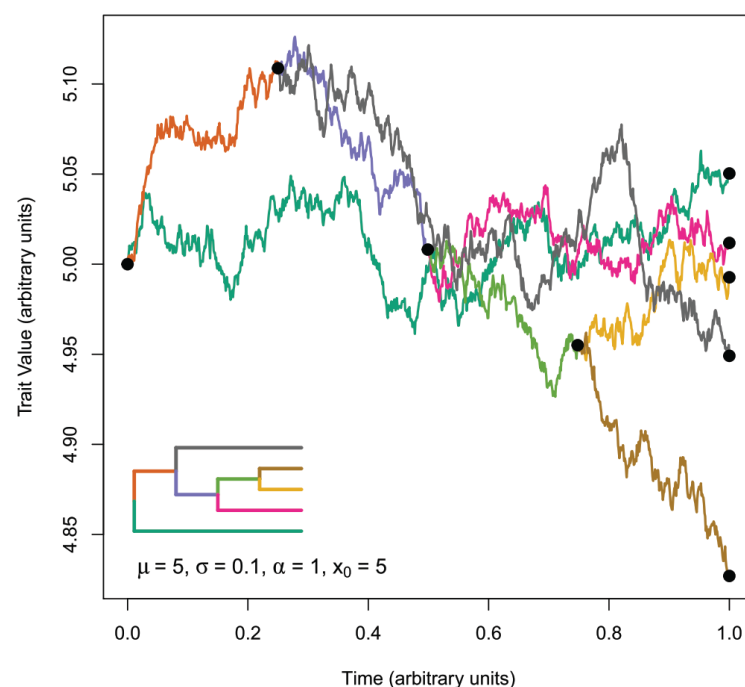


Fig. R2 Ornstein-Uhlenbeck evolution along a species tree, adapted from Blomberg et al., 2020. μ is the OU process mean (equal to the parameter θ in the R package BAYOU we used for OU modelling), σ is the strength of the restraining force, and α is the diffusion coefficient. x_0 is the starting value for the process.

From the methodological perspective, the OU model estimates the genome size evolution across phylogenies under adaptive selection, reflecting pressures and trends that influenced genome size changes over time. Among the parameters of the model, θ is used to show the OU process mean that the trait evolves towards (Fig. R2). Biologically, θ suggests the “fittest” value of genome size during the evolutionary process shown by the branch. We hope this can clarify the meaning of the yellow branch in our figures.

To prevent further misunderstanding, we have clarified this interpretation in the Methods (Lines 946–950) and legends of Fig. 2a and Extended Data Fig. 2 explicitly stating that the branch color reflects the estimated optimal GS that the lineages evolve towards, rather than the reconstructed ancestral genome sizes.

References:

Blomberg, S. P., Rathnayake, S. I. & Moreau, C. M. Beyond Brownian motion and the Ornstein-Uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. *Am. Nat.* 195, 145–165 (2020).

Line 306: In this paragraph, I didn't get any evidence to support the link between LTR-RT compositions and environmental variables.

Response: Thank you for this insightful comment. We agree that the original paragraph did not provide explicit evidence supporting a link between LTR-RT compositions and environmental variables. In fact, we have conducted relevant analyses, but the results were not included in the original manuscript because of a lack of significant correlation between the expansion of specific LTR-RT components and the environmental variables examined.

In the revised manuscript, we have now added the results of PGLS analyses testing relationships between environmental variables and LTR-RT compositions (Supplementary Tables 22–26). These results indicate that, despite the important role of LTR-RT insertions in genome expansion, variation in LTR-RT composition is not significantly associated with the environmental factors analyzed. We have revised the text accordingly to clarify this point and to provide a more comprehensive and balanced interpretation of genome size evolution (Lines 377–389; Lines 1035–1042).

Fig. 5: It seems that this figure is as same as Extended Data Fig. 10.

Response: We apologize for this oversight and thank the reviewer for pointing it out. Fig. 5 was inadvertently duplicated from Extended Data Fig. 10 in the original submission. We have now replaced Fig. 5 with the correct figure in the revised manuscript and verified that all figure references and captions are consistent.

Reviewer #2 (Remarks to the Author):

Yu et al. investigate whether LTR retrotransposons—especially intronic insertions—drive

genome-size expansion in *Cyphostemma* (Vitaceae) and whether that expansion contributes to succulence. They assemble a *Cyphostemma* phylogeny, compile 139 genome-size records for Vitaceae, generate two chromosome-scale genomes (the non-succulent rainforest climber *C. dehongense* and the succulent desert shrub *C. currorii*), and augment these with shallow whole-genome sequencing across ~77 species to profile repeat composition, alongside RNA-seq, bisulfite-seq, leaf micromorphology, and climate/soil data. The study reports that Vitaceae—particularly *Cyphostemma*—show disproportionate intron expansion tied to LTR activity rather than recent whole-genome duplication; that a few LTR families (Tekay, Ogre, SIRE) dominate intergenic genome inflation while Ale elements are unusually enriched in introns; that genes with extra-long introns are down-regulated and more methylated; that cyclin/auxin modules retain intronic Ale and show expression divergence consistent with larger cells, lower stomatal density, and altered vasculature; and that succulence often precedes genome-size regime shifts, with later upsizing associated with arid, seasonal climates. Overall, it is an ambitious, multi-layered analysis linking TE ecology, genome architecture, trait anatomy, and environment; the intronic-LTR angle (Ale within introns) is distinctive and strengthened by the absence of recent WGD; clade-level “TE community” patterns (Tekay vs. Ogre) are clear within a thoughtful phylogenetic framework; and the leaf micromorphology provides a concrete phenotypic anchor.

However, the study is highly complex, and it is not always clear which datasets were used where, for what purpose, and how specific claims are supported. In particular, the manuscript advances a very specific hypothesis—that intronic LTRs (notably Ale) help drive succulence—without clearly stating the priors that motivated this focus. Clarifying the origin of this hypothesis and providing a concise “data-to-claim” roadmap would make the narrative easier to follow. Below are my detailed comments.

[Response: Thank you for your thorough summary and thoughtful evaluation of our study. We appreciate your recognition of the integrative and multi-layered nature of the work, as well as your positive assessment of the intronic LTR-RT focus, the phylogenetic framework, and the linkage between genome architecture and leaf micromorphology.](#)

[We agree that the complexity of the datasets and analyses may have obscured how specific data support individual claims. In response, we have revised the manuscript to more explicitly clarify the corresponding datasets and analyses used for specific claims \(e.g., Lines 115–117 state that the trend of genome expansion in Vitaceae is detected from ancestral state reconstruction based on genome size data of 139 Vitaceae and outgroup species; Lines 208–210 and Lines 233–234 state that the dynamics of LTR-RTs are inferred based on genome assemblies of *C. dehongense* and *C. currorii*\).](#)

[We also appreciate your comment regarding the motivation for our focus on intronic LTR, particularly *Ale* elements. To address this, we have expanded the Introduction to better articulate](#)

[the background and prior evidence motivating this hypothesis \(Lines 63–65; Lines 74–78\).](#)

[We hope that these revisions improve the overall coherence of the narrative and make the logical connections between data, analyses, and conclusions more transparent to readers. Below, we provide detailed point-to-point responses to each of your comments. Please note that the line numbers indicated below refer to the track-changed version of the manuscript. Additionally, we have submitted a clean version of the main text for your convenience.](#)

The phylogeny using 1,367 single-copy genes across 151 taxa: single-copy genes were identified in 12 species and a few additional ones for which transcriptomes were assembled. This step where different annotations are used already introduces a bias. And then for the up to 151 species “shallow” read data was somehow mapped to these genes using the hybpipe pipeline to identify single copy genes in those taxa. Is that correct? How can it be ensured like that these genes are really single copy?

[Response: Thank you for raising these points regarding potential bias and the identification of single-copy genes. The single-copy ortholog genes used for phylogeny inference were identified from a reference set of six Vitaceae species, including four genomes \(*Cyphostemma currorii*, *C. dehongense*, *Tetrastigma hemsleyanum*, and *Cissus rotundifolia*\) and two transcriptomes \(*Causonis ciliifera* and *Pseudocayratia speciosa*\). The transcriptomes were assembled in our previous study, so the consistency in the annotation process can be ensured. For the downloaded genomes, we have reviewed their annotation pipeline and found them to be aligned with our annotation methods, minimizing potential methodological bias.](#)

[To ensure the accuracy of single-copy genes, we filtered the gene set based on sequence similarity, gene length, gene tree branch lengths, and genomic synteny \(for genome data\). Although historical gene duplication and loss preclude absolute single-copy status across all taxa, we believe that the rigorous filtering process has resulted in a reliable set of orthologs suitable for phylogenetic reconstruction. We have clarified these procedures in the revised manuscript \(Supplementary Note 5\).](#)

Intron length analysis and genome annotations: are the assemblies/ annotation of the 12 species comparable? All long read, de novo annotated with external evidence, high N50 and BUSCO scores, etc. ?

[Response: Thank you for your question regarding the comparability of genome assemblies and annotations used in the intron-length analyses. In the revised manuscript, we have increased our outgroup to 55 species for a more comprehensive representation of genome characteristics across the super-rosids \(Line 146\). All the genomes in our comparative analyses are of high quality and broadly comparable. Specifically, all assemblies were generated using long-read](#)

[sequencing data with coverage exceeding 30× \(with the exception of *Vicia faba*, which has an exceptionally large genome of 11.9 Gb\), and all represent chromosome-level assemblies with high scaffold N50 values \(> 5 Mb\) and BUSCO scores \(> 90%\). Gene annotations for all genomes were corrected with external evidence, and repeat annotations were validated using multiple lines of evidence. Detailed assembly and annotation statistics for all genomes used in this analysis are provided in Supplementary Table 6.](#)

What is the analysis of the longest intron aiming at? These might also be fused gene models/ annotation artefacts.

[Response: Thank you for this insightful comment. We apologize for the lack of clarity in our original presentation. The distribution of intron length in Fig. 1 is for all introns, rather than only the longest intron per gene, which was incorrectly stated in the figure legend. We have corrected this in the revised figure legend.](#)

[In our analyses, we only use the longest intron as the criterion to classify genes into three groups \(genes with introns \$\leq\$ 250bp, 250-4,000 bp, and > 4,000 bp\). The extra-long intron can affect gene function by increasing transcriptional burden and the risk of splicing errors. Consequently, genes containing exceptionally long introns are more likely to experience relaxed selection \(Cai et al., 2021\). We therefore suggest that the length of the longest intron can be a good measurement for selection force, as the presence of an extremely long intron can substantially impair gene function. Moreover, as many long introns arise from LTR-RT insertions, grouping genes by their longest intron length can show the effect of LTR-RT insertion more directly. This rationale has been added to the revised manuscript \(Lines 982–983\).](#)

[We acknowledge that excessively long introns may, in some cases, reflect fused gene models or annotation artifacts. However, the gene annotations of genome assemblies used in our comparative analyses were validated using transcriptome data \(Lines 932–935\), substantially reducing the likelihood of widespread annotation errors. Although a small number of artefactual cases cannot be entirely excluded, they are unlikely to affect the overall trends observed in our study.](#)

[References:](#)

[Cai, L. et al. Deeply altered genome architecture in the endoparasitic flowering plant *Sapria himalayana* Griff. \(Rafflesiaceae\). Curr. Biol. 31, 1002-1011.e9 \(2021\).](#)

The authors claim that it is remarkable that LTR-RT proliferation is the primary driver of genome and intron expansion as “LTR-RTs are typically enriched in intergenic regions and

rarely generate extra-long introns in angiosperms, suggesting a unique evolutionary mechanism favoring intronic retrotransposon colonization in Cayratieae.” Reference #26 does mention TEs within introns in other crops. In the two mentioned references, I don’t find proof that LTR-RTs are rarely within introns. Please make sure this claim is accurate.

Response: Thank you for this helpful comment. We have revised the statement to avoid overgeneralization and to more precisely define the term “extra-long introns”. In the revised manuscript, “extra-long introns” explicitly refers to introns exceeding 4,000 bp, consistent with the criteria applied throughout our analyses.

Reference #26 documents the presence of TEs within introns in several crop species, but emphasizes that the majority of TEs are concentrated in gene-poor, intergenic regions. Reports of very long introns in crops predominantly involve monocots, and Table S1 of Reference #27 further indicates that extra-long introns occur much less frequently in eudicots. In light of this distinction, we have replaced “angiosperms” with “eudicots” to improve taxonomic accuracy (Line 36; Lines 51–52; Line 167).

Accordingly, our findings in Cayratieae do not imply that intronic LTR-RT insertions are absent in other plant lineages. Rather, they indicate that the contribution of LTR-RTs to intron expansion in Cayratieae is unusually pronounced relative to other eudicots examined to date. We have rephrased the relevant statements (Lines 165–170) to emphasize that the preferential accumulation of LTR-RTs within introns in Cayratieae represents an underexplored aspect of genome evolution in this lineage.

The central inference that *Ale* preferentially colonizes introns in *Cyphostemma* is supported by two chromosome-level assemblies (*C. dehongense*, *C. currorii*) only. The ~77 additional species are profiled with shallow, reference-free repeat clustering, which estimates family abundances but cannot localize insertions (intronic vs intergenic). Maybe I’m misunderstanding but if not, I think it should be made more clear that all the claims of the manuscript are about these two assemblies and not a general analyses of the genus. I think right now, the manuscript makes it look like it.

Response: Thank you for this important comment. We agree that only the two chromosome-level assemblies of *C. dehongense* and *C. currorii* allow precise localization of *Ale* insertions within intronic versus intergenic regions. We have revised the manuscript to clearly state that conclusions regarding insertion preferences are derived from analyses of these two chromosome-level genome assemblies, and not on the reference-free repeat clustering performed for the additional species (Line 233).

What does shallow sequencing mean? How deep?

Response: In this study, two different sequencing depths were used for distinct purposes. The whole genome sequencing data used for assembling chromosome-level genomes had a high coverage of more than 100×. In contrast, the “shallow sequencing” data for comparative analyses across the genus had a much lower coverage, approximately 0.2–20×. For analyses of repeat composition using RepeatExplorer2, the recommend sequencing depth is only 0.1–0.5×. Therefore, the sequencing depth employed in our study is sufficient and appropriate for this purpose. We have clarified this in our revised Methods section (Lines 955–956). Detailed sequencing depths for each species are provided in Supplementary Tables 1 and 2.

Enrichment analysis of genes with introns elongated by LTR-RTs: how was the functional enrichment done? What was the background? All genes or genes without long introns? It must be ensured here that the genes with long introns are intact gene models and not annotation artefacts before concluding about expression

Response: Thank you for raising this important point. For the functional enrichment analysis, we used all annotated genes in the *C. currorii* genome as the background. Restricting the background to genes without long introns would not adequately capture the overall functional landscape of the genome and could introduce bias. We have clarified these details in the Methods section of the revised manuscript (Lines 992–993).

To ensure that genes with long introns are not annotation artefacts, we applied rigorous quality controls during gene annotation. Specifically, genes with long introns have been validated by transcript data (Lines 854–856). We further confirmed that all genes included in this analysis are well-annotated, and gene annotations with fragmented ORFs, lacking RNA-seq support, detectable homology, or conserved protein domains were excluded.

It seems like the authors compiled a list of “succulence genes” from the literature. Is that correct? If so, this should be provided as a supplementary table. But also this feels a little like cherry picking. I think better would be to pull all the genes that deviate from the trend where genes with longer LTR-RT insertions in *C. dehongense* tend to exhibit contraction or less elongation in *C. currorii* and do a functional enrichment of those and see what they are enriched for instead of the top-down approach.

Response: Thank you for your valuable suggestions. Yes, the set of “succulence genes” was compiled from the literature (Heyduk, 2021), and we have now provided the gene list in Supplementary Table 15.

We appreciate the concern regarding potential bias associated with a top-down, candidate-gene approach. To address this, we conducted an additional, data-driven enrichment analyses. Specifically, we identified genes that deviate from the overall trend by selecting *C. currorii*

genes with extra-long introns (> 4,000 bp) elongated by LTR-RTs compared with their syntenic counterparts in *C. dehongense*. This analysis revealed that some succulent-related genes have GO terms enriched in the genes deviated from the general trend, as they are involved in mitosis process. These results are consistent with, and provide independent support for, our original conclusions. The enrichment results are provided in Supplementary Table 15 and the corresponding description has been added to the revised manuscript (Line 324).

References:

Heyduk, K. The genetic control of succulent leaf development. Curr. Opin. Plant Biol. 59, 101978 (2021).

Data availability: the link to figshare is missing

Response: Thank you for pointing this out. We apologize for the missing link. We have updated the manuscript to include the correct figshare link (<https://figshare.com/s/71cf4e8d081c59fe52d7>) and ensure that all datasets are accessible to reviewers. A public figshare link will be activated upon publication of the manuscript.

Manuscript No.: NCOMMS-25-60992B

Title: Genomic and ecological drivers of parallel arid adaptation in tree grapes (Vitaceae)

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

I would like to thank the authors for well responding to my suggestions and confusions in the previous round of review. I believe that the current manuscript has significantly improved in the rationality of data presentation and the reliability of conclusion. However, there are still a few issues that I hope the authors can consider.

[Response: We sincerely thank the reviewer for the positive feedback and for acknowledging the improvements made to our manuscript. In response to the remaining concerns, we have further refined the manuscript accordingly. Our point-by-point responses to the remaining issues are provided below. Please note that all referenced line numbers correspond to the change-tracked version of the revised manuscript.](#)

1.Fig. 2b: Is the statistical data in the figure based on the entire lineages or representative species within it? If based on the entire lineages, further explanation is needed on how this proportion value was obtained; if based on representative species, please specify the species names.

[Response: Thank you for raising this important question. We agree that this was not clearly stated in the original manuscript. The proportions shown in the pie charts represent the mean proportions calculated across all sampled species within each clade. To clarify this, we have revised the legend of Fig. 2b and described how the mean proportions were calculated for each clade \(Lines 966 - 968\).](#)

2.Line 199-201: Fig. 2c can't support this sentence.

[Response: Thank you for pointing this out. We have corrected the citation by replacing Fig. 2 with Fig. 3, which more appropriately supports this statement \(Line 204\).](#)

3.Line 202-204: In Fig. 3b, I noticed that the Tekay data from *C. dehongense* is too small to compare with *C. currorii*. How did the authors reach this conclusion regarding ORF? Moreover, for the subsequently emphasized Ale, the differences in ORF characteristics between the two are not significant. Does this imply that the subsequent impact of intronic LTR on expression

is not correlated with methylation?

Response: Thank you for your careful observation and for raising these important questions. Regarding *Tekay*, we acknowledge that the number of *Tekay* elements in *C. dehongense* is relatively low, which limits direct quantitative comparison with *C. currorii*. Our inference is therefore not based solely on element abundance, but rather on the overall methylation context and ORF characteristics. Specifically, *C. currorii* exhibits higher overall CHH methylation levels than *C. dehongense*, suggesting a more efficient or timely silencing of LTR-RTs. However, elements lacking complete ORFs (non-autonomous elements) show relatively lower CHH levels (Extended Data Fig. 4), indicating that these elements may be less strongly targeted by CHH methylation. Notably, the *Tekay* family contains a relatively high proportion of non-autonomous elements. We therefore speculate that with the globally elevated CHH methylation, *Tekay* may gain a relative survival advantage among LTR-RT families. We have revised the manuscript to clarify that this interpretation is speculative and not based on a direct statistical comparison of *Tekay* abundance between the two species.

With respect to *Ale* elements, we agree that the differences in ORF characteristics between the two species are not significant. Previous studies have shown that CHH methylation has a stronger effect on elements from the *Gypsy* superfamily (such as *Tekay*), whereas *Ale* elements from the *Copia* superfamily are generally less influenced (Wang et al., 2018). In contrast, CG and CHG methylation are typically more stable and less sensitive to LTR-RT activity. Therefore, the expression patterns of intronic *Ale* elements are unlikely to be explained solely by differences in ORF structure or CHH methylation levels. Instead, their elevated expression may be due to proximity to exons, which increases the possibility of co-transcription with host genes. However, the precise mechanisms need further examination by investigation into the structure and function of reverse transcriptase and integrase in LTR-RT elements. We have revised the relevant text to clarify these interpretations (Lines 203 – 217).

References:

Wang, Y., Liang, W. & Tang, T. Constant conflict between *Gypsy* LTR retrotransposons and CHH methylation within a stress-adapted mangrove genome. *New Phytol.* **220**, 922–935 (2018).

4.Line 210-211: As it said, LTR patterns “remain evolutionarily conserved”. However, significant diversity in succulence exists within both the continental African clade and the Malagasy clade. Does this suggest that the succulence is not caused by LTRs? I find this the most perplexing aspect of the current manuscript and hope the authors can provide a more plausible explanation.

Response: Thank you for highlighting this important conceptual issue. Indeed, the composition of LTR-RTs within each clade is relatively conserved, even though succulent traits have evolved multiple times independently. We agree this requires further clarification.

In our study, the statement that LTR patterns “remain evolutionarily conserved” refers to the relative proportions of major LTR-RT families at the clade level. We do not propose that large-scale shifts in overall LTR-RT composition directly cause the origin of succulence. Rather, our hypothesis focuses on the functional impact of specific LTR-RT insertions. In particular, we identify intronic LTR-RT insertions in genes associated with succulence development, which are mainly cyclin- and auxin-related genes involved in cell cycle regulation (Fig. 4 and Lines 311 – 332). Such LTR-RT insertions, mostly from the *Ale* family (Supplementary Fig. 9), may influence cell cycle regulation and promote cell expansion (e.g., via promoting endoreduplication that can increase cell volume and water storage), thereby contributing to the emergence of succulence (Lines 323 – 332). Therefore, the emergence of succulence does not necessarily require substantial changes in overall LTR-RT abundance or composition; instead, regulatory effects of insertions at specific cell cycle-related genes may be sufficient.

Importantly, conservation of overall LTR-RT composition does not preclude functional diversification at finer scales. Even in species that have experienced recent LTR-RT bursts, the relative compositions of major LTR-RT families often remain stable. Similar patterns have been reported in other plant lineages (e.g., Neumann et al., 2021; Schley et al., 2022). We suggest that the pronounced changes in LTR-RT compositions appear to be more commonly associated with large-scale genomic rearrangements or chromosomal restructuring, as observed in Neumann et al. (2021). However, whether such shifts represent a general principle in plant genome evolution remains to be tested with additional high-quality genome assemblies and cytogenetic data.

We regret that this distinction between overall LTR-RT composition and locus-specific insertion effects was not clearly articulated in our original manuscript. To address this, we have revised the relevant sentences in discussions (Lines 343 – 348) and added a conceptual summary diagram to clarify the proposed relationships among genome size dynamics, LTR-RT insertions, succulence evolution, and habitat transitions (Supplementary Fig. 14, also see below).

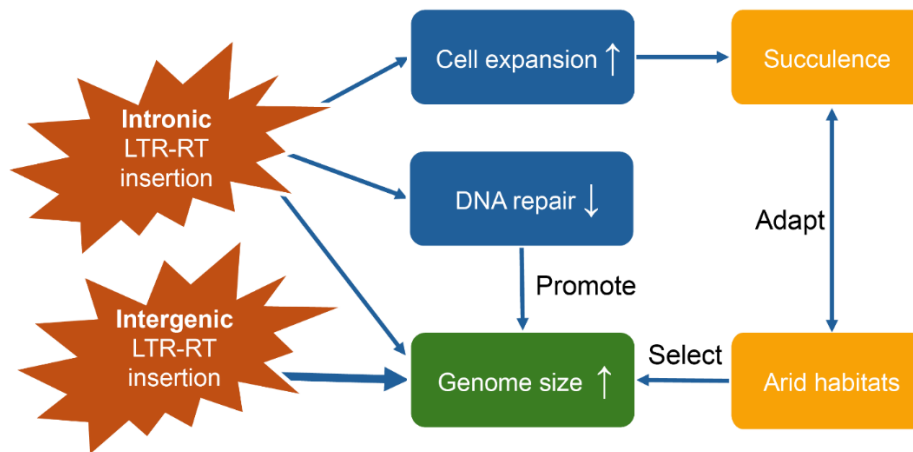


Fig. S14 | Conceptual diagram summarizing the main findings of this study. The thicker arrow from “intergenic LTR-RT insertion” indicates that these insertions represent the primary driver of genome expansion.

References:

Neumann, P. *et al.* Impact of parasitic lifestyle and different types of centromere organization on chromosome and genome evolution in the plant genus *Cuscuta*. *New Phytol.* **229**, 2365–2377 (2021).

Schley, R. J. *et al.* The ecology of palm genomes: repeat-associated genome size expansion is constrained by aridity. *New Phytol.* **236**, 433–446 (2022).

5. In the section, “Larger genomes selected by arid habitats”, most conclusions are based on the comparison of succulence within lineages, yet this aspect has not been addressed in the previous sections and lacks associated conclusions. How can the authors prove that the GS differences within lineages indeed reflect variations in LTRs, especially intronic LTRs? As noted in the previous comment, the authors also require more robust evidence or clarification concerning the role of LTR diversity within lineages. Additionally, I noticed that the author's team recently published related research in JIPB, and the work in this section of the two articles appears very similar. Can that article provide any additional support for the conclusions in this section?

Response: Thank you for raising these important concerns. As clarified in our previous response, we do not propose that succulence is driven by the changes in overall LTR-RT abundance or compositions. Rather, we suggest that intronic LTR-RT insertions at specific genes associated with succulence may contribute to the phenotypic divergence. However, the genome expansions in *Cyphostemma* are driven by selective forces, as suggested by our best-fitting Ornstein–Uhlenbeck model for genome evolution (Fig. 2a), which suggests evolution toward

an optimal genome size. Therefore, the section entitled “Larger genomes selected by arid habitats” is included in our manuscript to examine the potential ecological source of this selective pressure (Lines 339 – 341).

Our analysis of genome size and environmental factors shows that larger genome sizes are significantly associated with arid habitats (Fig. 4). Importantly, we find that this genome expansion is not attributable to the disproportionate proliferation of any specific LTR-RT lineages (Lines 364 – 366). Instead, our results suggest that the overall genome expansion, rather than the proliferation of particular LTR-RT families, may provide a survival advantage under arid conditions (Lines 366 – 368). We did not indicate the direct correlation between intronic LTR-RT insertions and environmental selection in this section.

Regarding the JIPB study, that work was initiated prior to the current genomic investigation and primarily focused on macroevolutionary relationships among traits, habitat, and genome size. It demonstrated that succulent *Cyphostemma* species in arid habitats tend to have larger genomes and extended this association to eudicots more broadly (Ranaivoson et al., 2026). The present study integrates genome assemblies to investigate the genomic mechanisms underlying genome size variation and adaptive trait evolution, with particular emphasis on LTR-RT proliferation and insertion patterns. The diversification analyses in the JIPB study showed that larger genomes are associated with higher net diversification rates, providing independent macroevolutionary support for the hypothesis that increased genome sizes may have ecological or evolutionary advantages.

To address your concerns, we have now explicitly cited the JIPB study and revised related discussions (Lines 351 – 356) to improve connection with previous sections of the manuscript. Additionally, we have added a conceptual summary diagram (Supplementary Fig. 14) to clarify the relationships among LTR-RT dynamics, genome size evolution, succulence, and habitat transitions.

References:

Ranaivoson, R. M. *et al.* Plastid and nuclear phylogenomics of *Cyphostemma* (Vitaceae) provide new insights into genome size evolution across sub-Saharan Africa. *J. Integr. Plant Biol.* **In Press**, (2026).