

Description of Additional Supplementary Files

Supplementary Data 1 - Information on individuals of *Cyphostemma* used for phylogeny reconstruction, including voucher specimen, collection No., locality, and native distribution. “†” indicates sequences newly produced in this study. “WGS”, whole genome sequencing data.

Supplementary Data 2 - Sources of the whole genome sequencing (WGS) data of outgroup species used in phylogeny reconstruction and biogeography analyses. “†” indicates sequences produced in this study.

Supplementary Data 3 - Results of model test in BioGeoBEARS, based on the time tree of *Cyphostemma*. The best model is highlighted in bold.

Supplementary Data 4 - Genome size records of Vitaceae species used in this study. “ADCD” represents the Angiosperm DNA C-values Database.

Supplementary Data 5 - Results of model test for genome size evolution in Vitaceae and *Cyphostemma*, using R package OUwie with the assumption that genome size evolved differently in succulent and non-succulent species in models allowing multiple optimum values. The best model is highlighted in bold.

Supplementary Data 6 - Genome traits and detailed information of assemblies for selected superrosid species.

Supplementary Data 7 - Summary of the two-sided ANCOVA model testing the contributions of lineage, proportion of LTR-RT in genome, and genome size. Significant terms ($P < 0.05$) are indicated in bold.

Supplementary Data 8 - Detailed information of *Cyphostemma* species used in regression analyses investigating relationships between environment and genome size variation, including environment variables, habitats, morphological traits, and LTR-RT read counts.

Supplementary Data 9 - Information of intact LTR-RT in *C. dehongense* and *C. currorii*. For gene distance, exon distance, and 5utr distance, positive value means the LTR-RT is in an upstream position, negative value means a downstream position, and “0” means the LTR-RT is inserted in the gene. For internal ORFs, “1” means present, “0” means absent. “TPM” means “transcript per million”.

Supplementary Data 10 - Information of LTR-RT lineages and families in *C. dehongense* and *C. currorii*. “TPM” means “transcript per million”. “Inf” means an infinite value.

Supplementary Data 11 - Estimations of insert time, age distribution, death rate, and the hypothesis test of a constant insertion rate of lineages with > 50 elements in *C. dehongense* and *C. currorii*. “Es1” represents results estimated with a single negative binomial distribution fit, and “Es2” represents results estimated with mixed models. “KL” represents Kullback–Leibler divergence describing the goodness-of-fit. “NA” means this value is not available. The negative binomial fit and the evaluation of fitness were conducted in R package TE.

Supplementary Data 12 - Percentages of LTR-RT families in introns of *C. dehongense* and *C. currorii*. The chi-square tests are two-sided. LTR-RT groups showing significantly skewed enrichment in introns or intergenic regions are shown in bold.

Supplementary Data 13 - Information of used LTR-RT elements for time tree construction to investigate the origin of intron-preferring *Ale* lineages.

Supplementary Data 14 - Summary of general linear regression models estimating relationships between age and GC content of autonomous LTR-RT elements in each family and each lineage with > 30 autonomous elements. “Cyde” represents *C. dehongense*, and “Ccur” represents *C. currorii*. *P*-values < 0.05 are indicated in bold.

Supplementary Data 15 - Potential syntenic succulence-related genes collect from previous literature, including genes related to cyclin (CYC), cyclin-dependent kinases (CDK), target of rapamycin (TOR), ribosomal S6 kinases (S6K), abscisic acid (ABA), cytokinin, and auxin. “TPM” means “transcript per million”. The genes with GO term enriched in genes with elongated intron in *C. currorii* are shown in bold.

Supplementary Data 16 - The identified succulence-related genes with significantly different expression level in succulent and non-succulent *Cyphostemma* species. “log2FC” means log-based fold changes between succulent and non-succulent species, where positive value means succulent species have higher expression level, and negative value means non-succulent species have higher expression level. The FDR *P*-value was derived from likelihood-ratio test that assesses whether the observed difference is larger than expected.

Supplementary Data 17 - Measurements of guard cell length, epidermal cell area, and stomatal density of *Cyphostemma* species.

Supplementary Data 18 - Climatic variables used for generalized least square regressions in this study.

Supplementary Data 19 - Summary of the best model in PGLS analyses explaining variations in genome size with environmental factors, tested in phylogenetic generalized least square regressions. Significant terms ($P < 0.05$) are indicated in bold.

Supplementary Data 20 - Akaike information criterion (AIC) values of tested models in phylogenetic generalized least square regressions in continental African clade, involving different combinations of uncorrelated environmental variables. The AIC value for best models is indicated in bold.

Supplementary Data 21 - Akaike information criterion (AIC) values of tested models in phylogenetic generalized least square regressions in Malagasy clade, involving different combinations of uncorrelated environmental variables. The AIC value for best models is indicated in bold.

Supplementary Data 22 - Summary of the best model in PGLS analyses explaining variations in genome size (1C-value; pg) with repeats of the four most abundant LTR-RT families, tested in phylogenetic generalized least square regressions. Significant terms ($P < 0.05$) are indicated in bold.

Supplementary Data 23 - Akaike information criterion (AIC) values of tested models in phylogenetic generalized least square regressions in continental African clade, involving different combinations of the four most abundant LTR-RT families.

Supplementary Data 24 - Akaike information criterion (AIC) values of tested models in phylogenetic generalized least square regressions in Malagasy clade, involving different combinations of the four most abundant LTR-RT families.

Supplementary Data 25 - Summary of the models in PGLS analyses explaining variations in LTR-RT counts with environment factors in continental African clade, tested in phylogenetic generalized least square regressions. Significant terms ($P < 0.05$) are indicated in bold.

Supplementary Data 26 - Summary of the models in PGLS analyses explaining variations in LTR-RT counts with environment factors in Malagasy clade, tested in phylogenetic generalized least square regressions. Significant terms ($P < 0.05$) are indicated in bold.

Supplementary Data 27 - Summary of the best model in PGLS analyses explaining variations in genome size (1C-value; pg) with soil nitrogen and phosphorus content, tested in phylogenetic generalized least square regressions. Significant terms ($P < 0.05$) are indicated in bold.

Supplementary Data 28 - Akaike information criterion (AIC) values of tested models in phylogenetic generalized least square regressions in continental African clade and Malagasy clade, involving different combinations of soil nitrogen and phosphorus content. The AIC values for best models are indicated in bold.

Supplementary Data 29 - The information of genome sequencing data for *C. dehongense* and *C. currorii*.

