Received: 8 May 2024 | Revised: 17 July 2024 | Accepted: 22 July 2024

DOI: 10.1111/mec.17510

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Repeatome evolution across space and time: Unravelling repeats dynamics in the plant genus *Erythrostemon* Klotzsch (Leguminosae Juss)

Natália Castro[1] | Bruno Vilela[2] | Yennifer Mata-Sucre[1] | André Marques[3] | Edeline Gagnon[4] | Gwilym P. Lewis[5] | Lucas Costa[1] | Gustavo Souza[1]

[1]Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Biosciences Center, Federal University of Pernambuco, Recife, Brazil

[2]Institute of Biology, Federal University of Bahia, Salvador, Bahia, Brazil

[3]Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany

[4]Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada

[5]Accelerated Taxonomy Department, Royal Botanic Gardens, Kew, Richmond, UK

Correspondence
Gustavo Souza, Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Biosciences Centre, Federal University of Pernambuco, Recife-PE, 50670-901, Brazil.
Email: luiz.rodrigessouza@ufpe.br

Funding information
Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: PQ-312852/2021-5; Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco, Grant/Award Number: APQ-0970-2.03/15 and IBPG-1762-2.03/21; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: Estágio Pós Doutoral 88881.119479/2016-01

Handling Editor: Mitchell Cruzan

Abstract

Fluctuations in genomic repetitive fractions (repeatome) are known to impact several facets of evolution, such as ecological adaptation and speciation processes. Therefore, investigating the divergence of repetitive elements can provide insights into an important evolutionary force. However, it is not clear how the different repetitive element clades are impacted by the different factors such as ecological changes and/or phylogeny. To discuss this, we used the Neotropical legume genus *Erythrostemon* (Caesalpinioideae) as a model, given its ancient origin (~33 Mya), lineage-specific niche conservatism, macroecological heterogeneity, and disjunct distribution in Meso- and South American (MA and SA respectively) lineages. We performed a comparative repeatomic analysis of 18 *Erythrostemon* species to test the impact of environmental variables over repeats diversification. Overall, repeatome composition was diverse, with high abundances of satDNAs and Ty3/gypsy-Tekay transposable elements, predominantly in the MA and SA lineages respectively. However, unexpected repeatome profiles unrelated to the phylogeny/biogeography were found in a few MA (*E. coccineus*, *E. pannosus* and *E. placidus*) and SA (*E. calycinus*) species, related to reticulate evolution and incongruence between nuclear and plastid topology, suggesting ancient hybridizations. The plesiomorphic Tekay and satDNA pattern was altered in the MA-sensu stricto subclade with a striking genomic differentiation (expansion of satDNA and retraction of Tekay) associated with the colonization of a new environment in Central America around 20 Mya. Our data reveal that the current species-specific Tekay pool was the result of two bursts of amplification probably in the Miocene, with distinct patterns for the MA and SA repeatomes. This suggests a strong role of the Tekay elements as modulators of the genome–environment interaction in *Erythrostemon*, providing macroevolutionary insights about mechanisms of repeatome differentiation and plant diversification across space and time.

KEYWORDS
Caesalpinieae, ecological variables, Fabaceae, genome ecology, satDNA, transposable elements

1 | INTRODUCTION

Speciation and divergence in genomic repetitive fractions (repeatome) are closely related concepts in evolutionary biology and genomics. Different populations or species may accumulate distinct repetitive element profiles or exhibit different patterns of repetitive sequences due to the accumulation of mutations and genomic rearrangements over time. However, the impact of repeatome divergence on species divergence is complex and multifaceted. Studies have demonstrated a relationship between differences in the repetitive genomic fraction (repeatome) and diversification rates (Oliver et al., 2013; Ricci et al., 2018). It has been demonstrated that plant genera with a high frequency of natural hybridization often present a high similarity in the repetitive genomic fraction (Barros e Silva et al., 2010; Castro et al., 2024). However, many questions regarding the dynamics of repeatome evolution remain unknown.

The interaction between genome features (e.g. genome size, genes and mobile elements composition) and ecological variables (e.g. soil, temperature and precipitation) is an intriguing subject in evolutionary biology (Chumová et al., 2022; Moraes et al., 2022). For instance, this interaction has been applied to understand the relationship between total nuclear DNA content (genome size), highly variable in eukaryotes and geographical distribution/ecological conditions (Grotkopp et al., 2004; Hultgren et al., 2018; Kang et al., 2014; Lysák et al., 2000;Wang et al., 2011). Notably, in angiosperms, genome size increases from the equator towards the poles in both the Northern and Southern Hemispheres. However, specifically in the Northern Hemisphere, genome size increases towards temperate zones and then decreases again in Arctic regions, forming an 'S-shaped pattern' attributed to a larger temperature gradient in the north (Bureš et al., 2024). This genome size variation in plants accounts for approximately 2000-fold difference between the smallest and largest genomes reported (Gregory, 2005) and is mainly caused by TEs accumulation (Lwin et al., 2017).

DNA sequences found in the eukaryotic genome can be subdivided into two major classes: the single-copy coding DNA sequences (mainly composed of coding sequences) and the repetitive, mainly non-coding, DNA sequences (Bourque et al., 2018; Šatović-Vukšić & Plohl, 2023). Previous research has revealed genomic adaptation to the environment, highlighting correlations between gene polymorphisms and different environmental conditions (Pluess et al., 2016; Sun et al., 2019). Nevertheless, little is known about the effect of ecological variables on repetitive elements, although flowering plants stand out for having large amounts of repetitive DNA (repeats) making up their genome (Hannan, 2018; Jouffroy et al., 2016; Pita et al., 2019), with this being the primary cause of the remarkable genome size variation in the group (Pellicer et al., 2018).

The repeats are classified as tandemly (e.g. satellite DNAs) or disperse distributed in the genomes, and transposable elements (TE) are the major group within the dispersed counterparts (Galindo-González et al., 2017). TEs can be classified into two classes distinguished by their transposition mechanism: retrotransposons (RT) and DNA transposons (Bourque et al., 2018; Wicker et al., 2007). Among RTs, long terminal repeats (LTR-RTs) are one of the most abundant elements of plant repeatomes (Wendel et al., 2016). LTR-RTs are classified into superfamilies, with Ty1/copia and Ty3/gypsy standing out as the most notable LTR representatives, that are further subdivided into a diversity of lineages (Neumann et al., 2019; Wicker et al., 2007).

Recently, some studies demonstrated that repeat transposition can be affected by the environment, often resulting in gene activation/silencing and changes in the abundance/composition of repetitive DNA (Canapa et al., 2020; Guo et al., 2022; Kanazawa et al., 2009). Nonetheless, it is still unclear whether specific TE lineages are strongly affected by ecological conditions or not. In this regard, only a few studies have properly investigated how repeatome composition (in terms of repeat clades) might be affected by contrasting ecological conditions. For instance, in barley (*Hordeum spontaneum*), stress-induced activation of the retrotransposon BARE-1 is involved in adaptive selection across varying microclimates at Evolution Canyon, Israel, and is associated with an increase in genome size through its amplification (Kalendar et al., 2000). More recently, Schley et al. (2022) have shown differential patterns of expansion/contraction of specific Ty1/copia repeats under environmental stress. Taken together, these studies suggest that the genome–environment interactions could be better understood by analysing at fine genomic scale.

It has been reported that fluctuations in the abundance and composition of both satDNA and TEs are involved in speciation processes (Feiner, 2016; Garrido-Ramos, 2015; Pons & Gillespie, 2004). One of the possible causes is the rapid evolution of this genome fraction, which can sometimes become a significant source of biological innovations (Richardson et al., 2015). For instance, differences in repetitive sequences between populations can contribute to reproductive isolation by affecting chromosomal pairing during meiosis or leading to hybrid incompatibilities (Garrido-Ramos, 2015; Levy, 2013). Additionally, some previous studies demonstrate regulatory and maintenance functions of these elements in different organisms (Cordaux & Batzer, 2009; Goodier & Kazazian, 2008). Therefore, the repetitive fraction can be one of the key pieces during the evolutionary process of a biological group. Due to the complex interaction of repeats among themselves and to other components of the host genome, it has been proposed that repeat clades that compose the genome can be analogous to species in a hypothetical ecological community (Schley et al., 2022; Venner et al., 2009). This intriguing parallel is referred to as 'genome ecology' (Brookfield, 2005), a concept that allows us to better understand the repeatome diversity using ecological metrics and study repeats from an ecological perspective (Schley et al., 2022). Therefore, analysis commonly applied in ecology studies can be used to understand, for instance, whether environmental conditions impact the evolution of genomic traits (Schley et al., 2022).

The Caesalpinieae Rchb. tribe (Gagnon et al., 2024) is an interesting model to investigate genome–environment interactions due to the significant correlation between genome size and latitude/temperature observed for some species of the group (Souza et al., 2019). Furthermore, variation in heterochromatin amount and distribution on chromosomes was also linked to ecological niche and geographical distribution (Mata-Sucre, Costa, et al., 2020; Van-Lume et al., 2017). Within the tribe, *Erythrostemon* is noteworthy for its ecological diversity, and disjunct distribution. The genus is distributed disjunctly in three main centers: (i) Southern USA, Mexico, Central America and the Caribbean (23 species); (ii) Caatinga vegetation in Northeast Brazil (with one species: *Erythrostemon calycinus* (Benth.) L.P. Queiroz) and (iii) Argentina, Bolivia, Chile and Paraguay (seven species) (Gagnon et al., 2016, 2019, 2024), in habitats ranging from different Seasonally Dry Tropical Forest (SDTF sensu Lopez-Toledo et al., 2024) environments (e.g. lowland SDTFs, Caatinga, dry forest, deserts, Yungas-Puna transition zones and Chaco-transition forests) (Figure 1a).

Erythrostemon has a high diversity of heterochromatin distribution, despite its strong stability in chromosome number ($2n = 24$) and absence of polyploidy, reported for both Meso and South American species (Mata-Sucre, Costa, et al., 2020; Van-Lume et al., 2017). Moreover, the Mexican *E. hughesii* has a heterochromatin composed mainly by satellite DNAs (satDNA) in contrast to the Ty3/gypsy-Tekay-rich heterochromatin of some South American species occurring in northeast Brazil (in the Caesalpinia group genera *Cenostigma*, *Libidibida* and *Paubrasilia*). This points to a possible geographical differentiation of the repeatomes in the Caesalpinieae tribe (Castro et al., 2024; Mata-Sucre, Sader, et al., 2020; Van-Lume et al., 2019).

Given the previous findings for the Caesalpinieae, we here looked for putative repeats that may be under the influence of environmental variables. We took advantage of genome skimming from several species of the genus *Erythrostemon* coupled with information about environmental and geographic distribution to investigate the genome–environmental interaction in a fine genomic scale. By comparing the repeatome of 18 *Erythrostemon* species, covering all clades and more than 50% of the species diversity of the genus, we aimed to address three key questions: (1) Can *Erythrostemon* clades be distinguished by a differential repeat composition and diversity, given its 33 Mya origin? (2) Do ecological variables and/or phylogeny impact repeat diversity in the genus? (3) Are there specific repeat clades that are strongly associated with ecological traits?

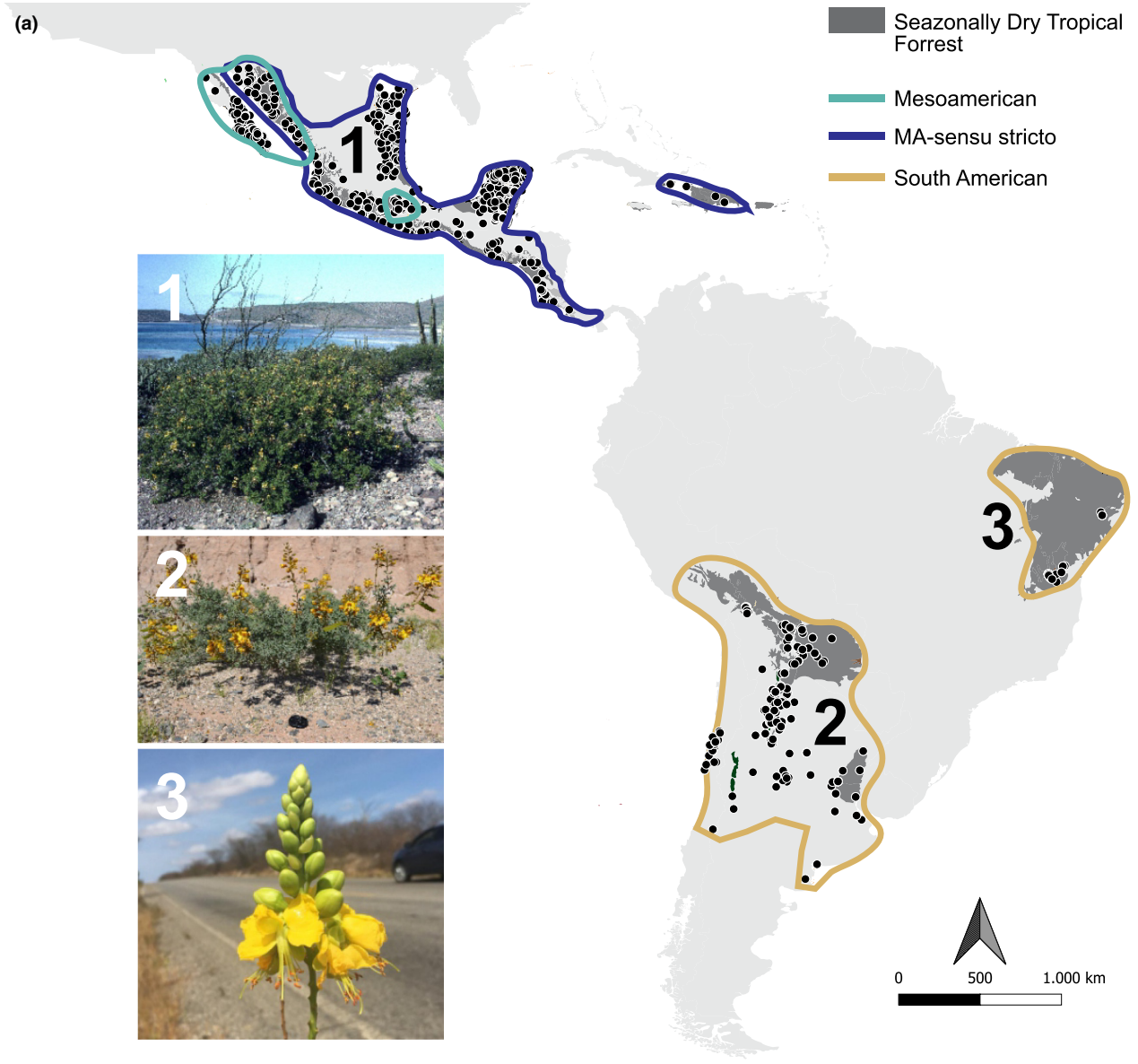
2 | MATERIALS AND METHODS

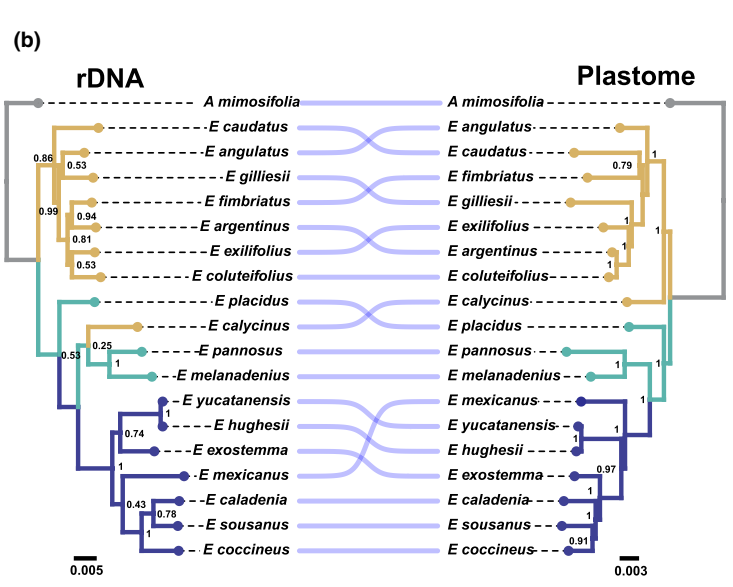
2.1 | Plant material, DNA extraction and genome sequencing

Eighteen species of the genus *Erythrostemon* were analysed, with eight representatives from the South American (SA) clade and 10 from the Mesoamerican (MA) clade. The seeds were obtained from the Millennium Seed Bank (Royal Botanic Gardens, Kew, UK), collected either from natural populations or obtained from cultivated plants (Table S1). Total genomic DNA was obtained from seeds/leaves using the CTAB extraction protocol of Ferreira and Grattapaglia (1995), slightly modified (omitting preheating of the extraction buffer before adding the sample) or using the Nucleopsin II kit (Macherey-Nagel) following manufacturer instructions. The genomes were then sequenced using the Illumina HiSeq 2000 platform at the Max Planck-Genome-Center, generating paired-end reads of 150 or 250 bp using the genome skimming approach. All of the sequenced genomes are available at the NCBI database (Project number: PRJNA739461).

2.2 | Phylogenetic analysis and molecular dating

We used full plastomes and nuclear rDNA to reconstruct phylogenetic relationships. Raw Illumina reads were mapped against the reference plastome of *E. calycinus*, (NCBI accession NC047361.1) and full 18S + ITS1 + 5.8S + ITS2 + 26S rDNA cistron of *Sena siamae* (NCBI accession ERX386821) using Geneious (v. 9.1.8) then the genomes were aligned using MAFFT (Katoh & Toh, 2008). For simplification, we used the most general model of DNA substitution GTR + I + G (Abadi et al., 2019) on the full plastome and nuclear rDNA alignments. Furthermore, the alignments were manually trimmed. Phylogenetic relationships were inferred using maximum likelihood (ML) with 1000 replicates in Geneious (v. 9.1.8) using the FastTree plugin (Price et al., 2009). Considering that the monophyletic status of *Erythrostemon* is well established (Gagnon et al., 2016), here we sought to characterize internal relationships based on our subsampling. For this, we selected the Andean species *Arquita mimosifolia* (Griseb.) Gagnon, G.P. Lewis & C.E. Hughes as an outgroup (Gagnon et al., 2019). To check for incongruences among the different topologies, by matching the tips of the inputted trees, the *Cophylo* package in software R was used (Revell, 2012). Finally, the network (Holland & Moulton, 2003) was constructed for full plastome and nuclear rDNA dataset using the program SplitsTree4 (Huson & Bryant, 2006). Divergence times were estimated using BEAST version 1.8.3 (Drummond & Rambaut, 2007) with the ML analysis result as the fixed plastidial topology. For calibration, we used the crown age of *Erythrostemon* (33.32 ± 4.32 Mya) and the crown ages of the SA (29.48 ± 6.02 Mya) and MA (30.97 ± 5.32 Mya) clades determined by Gagnon et al. (2019). An uncorrelated relaxed lognormal clock (Drummond & Rambaut, 2007) and Birth-Death speciation model (Gernhard, 2008) were applied. Two independent runs of 100,000,000 generations were performed, sampling every 10,000 generations. After removing 25% of samples as burn-in, the independent runs were combined and a maximum clade credibility tree was constructed using TreeAnnotator version 1.8.2 (Rambaut & Drummond, 2013). In order to verify the effective sampling of all parameters and assess convergence of independent chains, we examined their posterior distributions in TRACER. The MCMC sampling was considered sufficient at effective sampling sizes equal to or higher than 200.





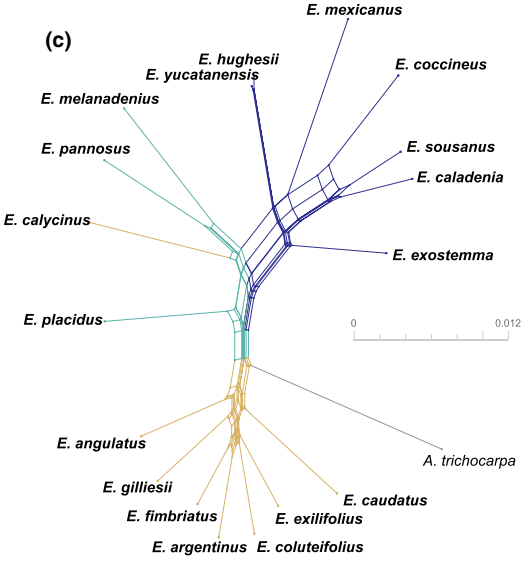


FIGURE 1 (a) Ecological diversity in *Erythrostemon*. Occurrences of *Erythrostemon* are highlighted on the map of Mesoamerica and South America, black points are the genus distribution. (1) *E. placidus* (Mesoamerica), (2) *E. exilifolius* (South America) and (3) *E. calycinus* (Northeastern Brazil). (b) Comparative topologies comparison obtained from plastome (right) and the rDNA tree (left). The number above the branches represents the bootstrap values. (c) The SplitsTreesnetwork based nuclear rDNA. The colours of the branches represent the geographic distribution of the species.

2.3 | Repeatome identification and characterization

For the individual analysis of the repetitive fraction of the species, genomic next-generation sequencing data was uploaded to the Galaxy/RepeatExplorer2 platform (Novák et al., 2020) to identify the most abundant repetitive elements, grouping them based on similarity and generating clusters for the different families of repetitive DNA. The number of reads for the comparative analysis were filtered by quality with default settings (95% of bases equal to or above the quality cut-off value of 10), then adjusted to obtain ~0.04× coverage for each species and interlaced (Table S2). Due to the difficulty in obtaining appropriate tissues for measuring genome size and since most of genome size variation within *Erythrostemon* clades is generally low (Souza et al., 2019), for species lacking genome size data, we extrapolated from the mean genome size of its respective clade (Table S2). Clustering was performed with default settings of 90% similarity over a 55% minimum sequence overlap. Only clusters with abundance above 0.01% were considered. Additionally, in order to compare the repeat composition observed in all 18 species, we employed a comparative analysis (Novák et al., 2020). The individual and comparative proportions of repeat clades sensu Neumann et al., 2019 (henceforth, whenever we mention a specific repeat clade or TE clade, it is referring to the total of all repeat units classified within that category), identified in each species genome, were calculated using the number of reads grouped versus the total amount of reads used for analysis, after excluding chloroplast and mitochondrial reads.

2.4 | Environmental data acquisition

To gather occurrence information of the 18 *Erythrostemon* species, data were downloaded from the Canadensys database (Gagnon et al., 2018) selecting only our subsample of *Erythrostemon* species. Additionally, from the coordinates of each species, we extracted values for the 19 climatic variables available in the WorldClim 1.4 (5 min) generic grid format (Hijmans et al., 2005) utilizing the package *raster* 2.6-7 (Hijmans, 2023) implemented in R.

After extracting the environmental variables, to avoid the selection of highly collinear variables, we checked for collinearity by using Pearson's correlation coefficient ($r > .75$). Finally, we selected non-collinear WorldClim variables which represented the finest temporal resolution, resulting in the following eight variables: Bio_2—Mean Diurnal Range, Bio_3—Isothermality (Bio_2/Bio_7) (×100), Bio_5—Max Temperature of Warmest Month, Bio_6—Min Temperature of Coldest Month, Bio_{13}—Precipitation of Wettest Month, Bio_{14}—Precipitation of Driest Month, Bio_{19}—Precipitation of Coldest Quarter and Elev—elevation data. In order to have typical variable values for each species, we calculated the species median of those eight selected environmental variables (Table S3).

2.5 | Environmental model selection and phylogenetic signal

To assess how each repeat clades in *Erythrostemon* genomes were influenced by different ecological variable predictors (non-collinear WorldClim variables), we evaluated which set of ecological variables best explained repeat lineage abundances. For this, we employed generalized linear models with Quasi-Binomial Regression. We performed an exhaustive model selection approach using glmulti (Calcagno, 2010) package implemented in R (Bartoń, 2023). The best minimum adequate models (model with less predictors) with lower Akaiake information criterion for regression models (Akaiake, 1973; Hurvich & Tsai, 1989) values were selected.

2.6 | Association between TEs variation and environment/phylogeny

We performed a Generalized Dissimilarity Modelling (GDM) (Ferrier et al., 2007), a distance-based, nonlinear extension of matrix regression implemented in the R package *gdm* (Fitzpatrick et al., 2022). For that, a distance matrix using the Bray–Curtis dissimilarity metric (Bray & Curtis, 1957) was created. The Bray–Curtis dissimilarity metric is based on the sum of the lesser abundant TEs shared between different species genomes in relation to the total TE abundances across the analysed species: $C = 1 - (2p_{jk})/(p_j + p_k)$, where p_{jk} is the sum of the lesser values for the species found in each site, while p_j and p_k are the sums of species values for samples j and k (Beals, 1984). Therefore, GDM analyses seek to establish relationships based on input data, here consisting of the distance dissimilarity matrix (Bray-Curtis) of repeatome composition against environmental and phylogenetic predictors. We used this set of variables for variance partitioning with the gdm.partition.deviance function. This analysis indicates the total amount of deviance explained by each variable set individually as well as in combination (Gattoni et al., 2022).

2.7 | Repeatome ecology and evolution

To investigate the repetitive DNA abundances between the SA and MA clades, we performed a Principal Coordinate Analysis (PCoA) using the *vegan* package (Oksanen et al., 2022) in R software (R Core Team, 2023). Furthermore, to test if the two *Erythrostemon*

clades form distinct groups in relation to their repeatome composition, we carried out a Permutation Multivariate Analysis of Variance (PERMANOVA, Anderson, 2005) also using a Bray–Curtis distance and 1000 permutations with the R package *vegan*.

To investigate the temporal dynamics of Tekay elements in *Erythrostemon*, we calculated the distributions of pairwise divergence between Illumina reads mapped to the reverse transcriptase (RT) domain (Mascagni et al., 2020; Usai et al., 2017). We extracted the reverse-transcriptase sequence using the Protein Domains Filter function of RepeatExplorer (Neumann et al., 2019). Then, we mapped the raw reads of the analysed species to the corresponding reverse-transcriptase domain sequence using the Low Sensitivity preset (five iterations) of the Geneious Read Mapper v. 6.0.3 plugin implemented in Geneious v. 7.1.9 (Kearse et al., 2012). From the mapped reads, we randomly selected 100 reads and calculated pairwise divergence using the MAFFT plugin (Katoh, 2002) under the Kimura two-parameter model of sequence evolution (Kimura, 1980). The pairwise distances were then converted to millions of years using the rice substitution rate of 4.9×10^{-9} substitutions/site/year (Mascagni et al., 2020; Usai et al., 2017). The resulting values were used to build density plots using the ggplot2 package (Wickham, 2016) in R software, where peaks indicate bursts of Tekay insertion.

To investigate what could cause repeat differences among the clades, we calculated the β-diversity (Whittaker, 1960) which enabled comparative understanding of the repeat composition of *Erythrostemon* genomes. That is, we investigated if the differences in TEs clades abundances across the genomes were linked to turnover (which implies the replacement of some TEs 'species' by others) or/and nestedness (i.e. *Erythrostemon* genomes with smaller numbers of TEs 'species' are subsets of the *Erythrostemon* genomes with higher TEs richness) (Baselga, 2010).

In order to evaluate alfa diversity of repetitive elements in *Erythrostemon* species, we used two metrics: 'TEs clades richness', that is, the number of distinct TEs clades present in each *Erythrostemon* repeatome and Pielou's evenness (Pielou, 1966), which assesses the uniformity of TE clades abundances in each *Erythrostemon* representative. We specifically chose Pielou's evenness to investigate whether certain TEs clades are highly dominant compared to others in our sampling. Additionally, in order to test if these metrics had a phylogenetic significance, we measured the phylogenetic signal λ (Pagel, 1999) using the *phytools* package (Revell, 2012) implemented in R software.

In order to test the phylogenetic significance of repeat abundances, we calculated the phylogenetic signal λ (Pagel, 1999) across all species using the results from the individual clustering analysis. For this, we used the *phylosig* function from the *phytools* package, implemented in the R software. With the aim of exploring changes in satDNA and Tekay abundance over time, these data were reconstructed along the plastid phylogeny using the *contmap* function implemented in the *phytools* package in R software. Finally, to investigate relationships between LTR TEs and satDNA, Tekay abundance, genome size (Souza et al., 2019) and latitude, we performed Pearson's correlation analyses. Additionally, we test the Pearson correlation of the repetitive elements with each other.

3 | RESULTS

3.1 | Revisiting *Erythrostemon* phylogeny and dating

The plastid alignment (155,846 bp) showed 125,612 identical sites (80.6%) and 5230 phylogenetically informative sites (3.35%). This phylogenetic reconstruction of the 18 analysed *Erythrostemon* species, was congruent with previous phylogenetic analyses (Gagnon et al., 2016, 2019). We observed the presence of two well-supported clades (BS = 1; Figure 2). The SA clade, is composed of Andean/Chaco species (*E. angulatus*, *E. argentinus*, *E. coluteifolius*, *E. exilifolius*, *E. fimbriatus* and *E. gilliesii*) and one NE Brazilian Caatinga species (*E. calycinus*). Only one species distributed in the Mesoamerican region (*E. caudatus*) was positioned in the SA clade. The MA clade comprised *E. caladenia*, *E. coccineus*, *E. exostemma*, *E. hughesii*, *E. melanadenius*, *E. mexicanus*, *E. pannosus*, *E. placidus*, *E. sousanus* and *E. yucatanensis* (BS = 1; Figure 2). According to our molecular clock analysis, the estimated crown age of the analysed species was 33.6 million years (My; Figure 2), with subsequent diversification of the MA and SA clades at around the same time (~30 Mya), similar to the times described by Gagnon et al. (2019).

For the rDNA tree, the rDNA alignment (5873 bp) showed 5156 identical sites (87.8%) and 349 phylogenetically informative sites (5.94%). The positions of *E. calycinus*, *E. mexicanus* and all species of the SA clade were incongruent when comparing the plastid and rDNA topologies (Figure 1b). In addition to the incongruence, the nuclear rDNA topology revealed low support and poorly resolved relationships, which gave rise to the hypothesis of reticulate evolution. The network analysis that was performed revealed a highly reticulated pattern in the backbone of the network, especially in the low-resolution parts of the tree (Figure 1c). On the other hand, the network analysis of the full plastome dataset did not show signs of reticulate evolution (data not shown).

3.2 | Repetitive DNA composition in species of *Erythrostemon*

The number of analysed reads of our individual RepeatExplorer analysis ranged from 170,563 reads in *E. hughesii* to 741,976 in *E. coluteifolius* (Table S2). The analysis revealed a heterogeneity in abundance of the repetitive fraction varying from 21.0% (*E. hughesii*) to 74.0% (*E. fimbriatus*) (Figure 2; Table S2).

Notably, both SA and MA species exhibited high proportions of repetitive elements ranging from 20.9% (*E. hughesii*) to 74.0% (*E. fimbriatus*). Ty3/gypsy-Tekay was especially abundant in the SA species (mean value = 29.4%) while appearing in smaller proportions in most MA species (mean value = 5.92%) (Figure 2; Table S2).

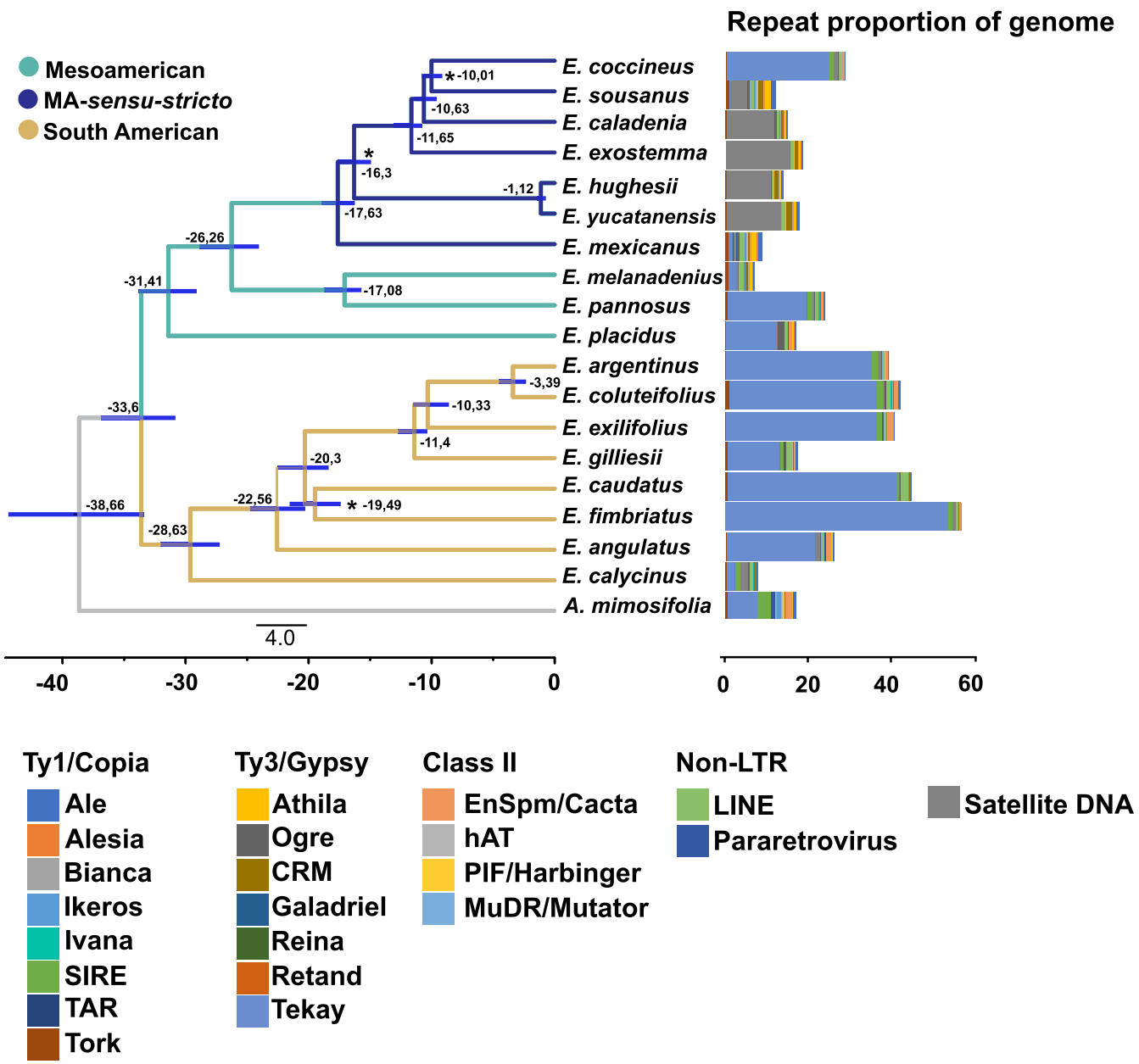


FIGURE 2 Dated plastid phylogenetic tree of 18 *Erythrostemon* species and individual repeat proportion (unclassified abundances are omitted) of each species including the outgroup *Arquita mimosifolia* (grey branch) analysed by Repeatexplorer. The colours of the branches represent the geographic distribution of the species. Bars on the right of the tips are coloured based on repetitive DNA; the *x*-axis represents the proportion of repeats expressed as a percentage for each genome. MA = Mesoamerican. Nodes marked with an asterisk (*) indicate bootstrap values (BS) lower than 1.

Remarkably, the MA species *E. melanadenius*, *E. pannosus* and *E. placidus* conserved higher Tekay abundances similar to the SA species (Tekay mean value = 11.03%). This pattern was strongly altered in the subclade comprising *E. coccineus*, *E. caladenia*, *E. exostemma*, *E. hughesii*, *E. mexicanus*, *E. sousanus* and *E. yucatanensis*, referred to here as the 'MA-*sensu stricto* subclade'. Within this group, Tekay accounted for less than 2% of the repetitive fraction in most species (mean value = 3.73%), except for *E. coccineus* (24.5%) which has most of its repetitive fraction composed of this TE lineage (Figure 2; Table S2). On the other hand, satDNAs proportions were relatively low in the SA clade, ranging from 0.01% in *E. exilifolius* to 1.9% in *E. calycinus* (mean value = 0.65%), but significantly higher in the MA-sensu stricto subclade (mean value = 7.97%).

For the comparative analysis, a total of 1,034,936 Ilumina reads from all 18 species were analysed, with a final coverage of ~0.01×. Of these, 777,831 reads were grouped into 122,168 clusters, while 257,105 remained unclustered. Considering only clusters with at least 0.01% genome abundance, 287 clusters were identified as repetitive elements (Figure S1). Our results of the species-specific and comparative repeatome analysis revealed contrasting satDNA and TEs patterns between the two phylogenetic clades in *Erythrostemon*.

3.3 | Composition and biogeographic patterns of transposable elements in *Erythrostemon*

To check whether the differential repeat accumulation found across *Erythrostemon* species can help to differentiate the taxonomic clades, we performed a PCoA analysis. The analysis revealed that the first and second principal components explained 74% and 12% of the variation in repeat abundances, respectively (Figure S2). These axes mainly separated the MA-sensu stricto subclade and the SA clade, with few incongruencies. Curiously, the MA species *E. coccineus*, *E. pannosus* and *E. placidus* exhibited a similarity with South American species in the PCoA. Furthermore, the unique Northeastern Brazilian representative (*E. calycinus*), appeared more similar to Mesoamerican species. Additionally, the PERMANOVA analysis confirmed that the South and MA sensu stricto clades can be differentially separated by repeat lineage compositions ($F_{1,14} = 11.14$; $p = .0037$).

TEs richness varied from 6 (*E. fimbriatus*) to 18 (*E. mexicanus*) (Figure 3) and did not differentiate the clades, suggesting a heterogeneity in repeatome composition in the genus. The Pielou evenness was higher in the MA clade (mean value = 0.65; standard deviation = 0.25) compared to the SA clade (mean value = 0.30; standard deviation = 0.21). Within the remaining Mesoamerican species (*E. melanadenius*, *E. pannosus* and *E. placidus*) Pielou evenness was intermediate (mean value = 0.50; standard deviation = 0.24) (Figure 3). The unique Northeastern Brazil representative (*E. calycinus*) exhibited higher Pielou evenness values, similar to those of the Mesoamerican clade (Figure 3). The Pielou index also presented a high phylogenetic signal λ (Table S4).

In general, the β-diversity revealed intermediate values, that is, most of the repeat clades are shared among the species. When the repeat sharing did not occur, it was mostly by turnover, with fewer cases of nestedness, especially in *E. fimbriatus* (Figure S3). Furthermore, there was no clear evolutionary nestedness direction related to a

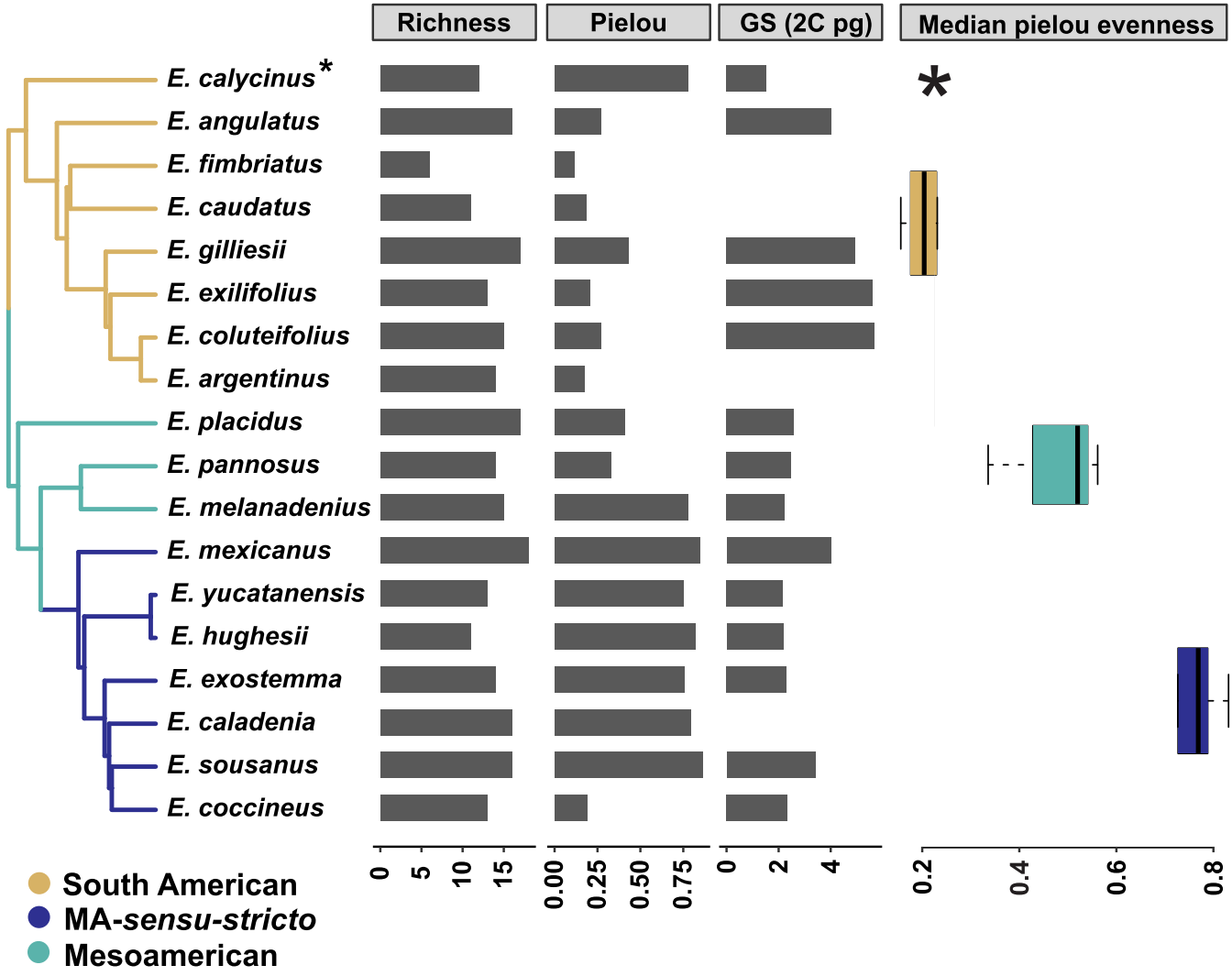


FIGURE 3 Quantitative diversity metrics (richness and Pielou's evenness), genome size variation and Pielou's evenness index for each clade represented as a boxplot for all *Erythrostemon* coloured by phylogenetic relationships. The phylogeny of the plastome shows the grouping of species in South America, Mesoamerica and MA-sensu stricto. The single species from northeast Brazil is marked with an asterisk (*) in the boxplot.

loss of TEs among different species (Figure S3). Interestingly, *E. exostemma*, *E. hughesii* and *E. yucatanensis*, species from the MA-sensu stricto subclade tended to have higher β-diversity related to turnover events compared with the SA clade and other Mesoamerican species (Figure S3).

We investigated the dynamics of Tekay insertion events in the *Erythrostemon* genomes. In general, two bursts of Tekay amplification were identified: (1) 11–5 Mya in the Miocene and (2) <5 Mya in the Pliocene–Pleistocene. Most of the SA species presented both bursts 1 and 2 (Figure S4). As with the previous results, the MA species *E. pannosus* and *E. placidus* exhibited a similar pattern to the South American representatives (Figure S4). On the other hand, within the MA- sensu stricto clade, *E. caladenia*, *E. coccineus*, *E. mexicanus* and *E. sousanus* revealed a more recent expansion scenario of Tekay elements (burst 2) (Figure S4). Remarkably, our results reveal several synchronized independent Tekay bursts across the genus *Erythrostemon* in the Neotropics and further hint at an environmental effect on differential TE accumulation.

3.4 | Factors that impact repeatome composition in *Erythrostemon*

To check whether environmental variables have an effect on differential repeat accumulation across *Erythrostemon* species we performed model selection analysis. One best fit model was yielded for each repeat lineage abundance, with most of the Ty3/gypsy repeats presenting at least one ecological variable as the best explanatory variable, except for the Athila element. Of note was that the Ty3/gypsy-Tekay element was the most strongly influenced by Bio_6. Additionally, other Ty3/gypsy, Ty1/copia, Class II and Non-LTR elements were also influenced by ecological variables, although showing weaker relationships (Figure 4a; Table S5).

In the GDM analysis, environmental variables alone explained 6.68% of the compositional dissimilarity across all *Erythrostemon*. Due to the strong correlation between the phylogeny and distribution of the species (which, in turn, is correlated with environmental variation), phylogeny alone did not account for any variation. This indicates that all of the percentage explained by phylogeny was also captured by environmental factors (15.63%) (Figure 4b). The sum of all analysed factors and their possible interactions contributed to explain 22.31% of total variation in *Erythrostemon* repeatomes (Figure 4b). To investigate if TEs abundance reflected phylogenetic relationships, we assessed Pagel's λ of the TEs across the species genomes. We found a high phylogenetic signal for six repeat clades (Table S4). Furthermore, Pearson's correlation analysis within all TE clades in *Erythrostemon* revealed positive correlations between different TEs. Notably, Tekay and SIRE exhibited significantly negative correlations with some of the remaining repeats (Figure S5).

Given Tekay and satDNA abundances were highly different among the two biogeographic clades, we investigated how it changed within *Erythrostemon* evolution across space and time (Figure 5a,b). Interestingly, the species with higher satDNA abundance present less LTR-TEs, corroborated by inverse correlation among these different repetitive DNA groups (Figure 5c). Additionally, the abundances of Tekay were positively associated with latitude and genome size (Figure 5c). Our results suggest that both ecological and phylogenetic factors are shaping repeatome composition within *Erythrostemon*, with the environment exerting a particularly strong influence, especially on the TE Tekay.

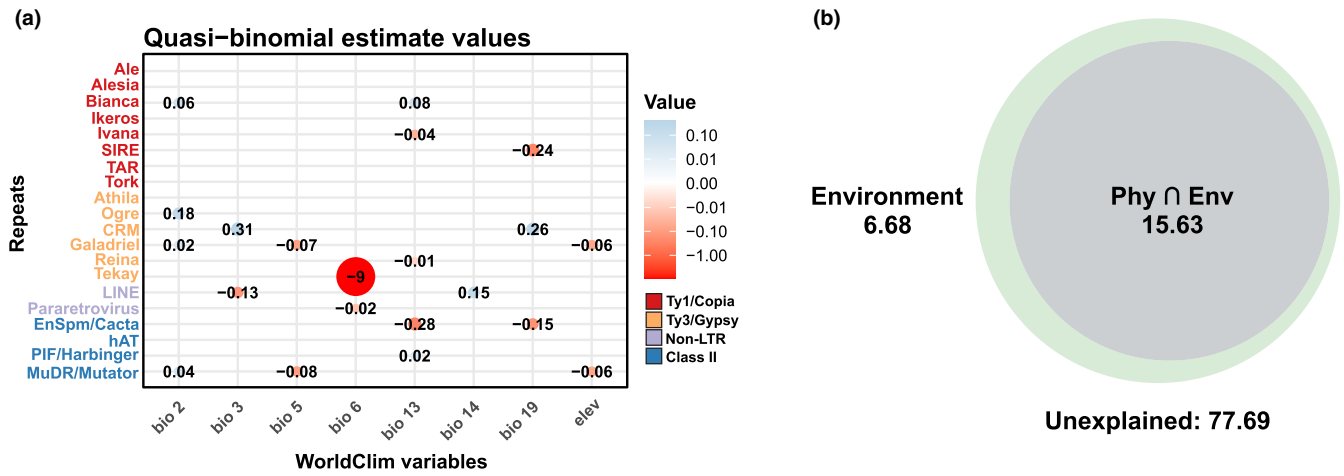


FIGURE 4 Factors influencing repeatome composition in *Erythrostemon*. (a) The estimate values of the best model selected for each TE lineage abundance in *Erythrostemon* species. The Y-axis is TE clades; colours of repeat names are based on their superfamily or class. The *X*-axis is the WorldClim variables: Bio2—Mean Diurnal Range, Bio3 Isothermality (Bio2/Bio7) (×100), Bio5—Max Temperature of Warmest Month, Bio6—Min Temperature of Coldest Month, Bio13—Precipitation of Wettest Month, Bio14—Precipitation of Driest Month, Bio19—Precipitation of Coldest Quarter and Elev—elevation data. (b) Venn diagram of the deviance partitioning analysis obtained to assess the relative importance of environment (green) variables and phylogeny (grey) in all TE abundances within the *Erythrostemon* species.

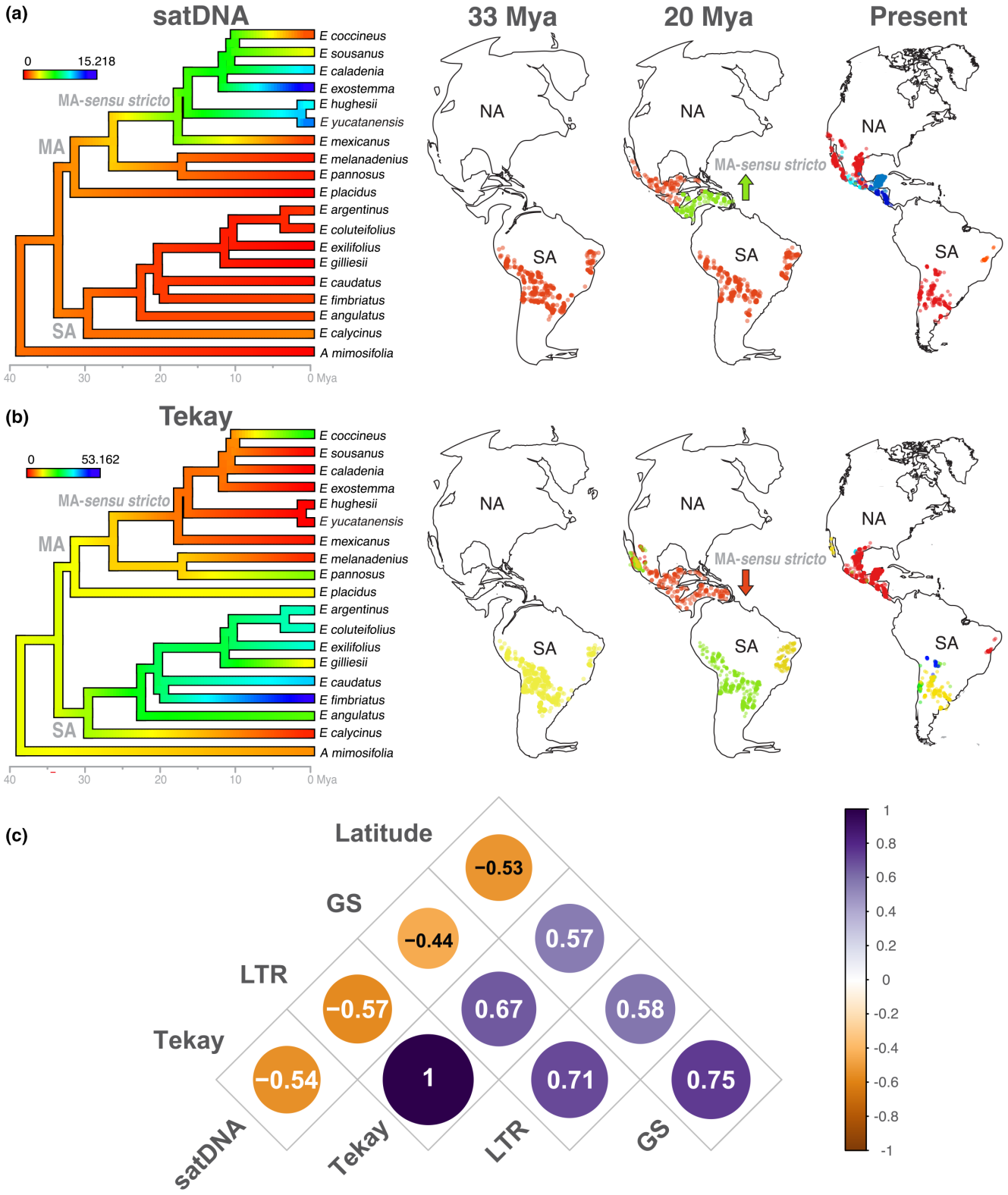


FIGURE 5 The interpretation of *Erythrostemon* repeatome evolution across space and time. (a) satDNA and (b) Ty3/gypsy Tekay reconstruction and their evolution throughout *Erythrostemon* diversification. The maps on the right of the reconstructions show the hypothesized *Erythrostemon* distribution over 30, 20 Mya and at the present, together with the abundance of satDNA and Tekay element throughout the genus diversification. The maps next to the reconstructions shows the hypothesized *Erythrostemon* distribution over 30, 20 Mya and at the present, respectively, together with the abundance of satDNA and Tekay element throughout the genus diversification. (c) Correlogram of latitude, nuclear genome size (GS) and genome traits: Proportion of satDNA, LTRs and Tekay elements. Shades of purple indicate increasing positive correlation coefficient; shades of orange indicate increasing negative correlation coefficient. Correlation values are within each ball, white values represent significant correlation ($p < 0.05$), black values represent non-significant correlations ($p < 0.05$).

4 | DISCUSSION

4.1 | Multi-trait influences shape repeatome diversity in *Erythrostemon*

Here we report a heterogeneity in the repeat 'community' composition within the genomes of *Erythrostemon* species across their geographical range in the Neotropics, which can be explained, in part, by environmental variables with less influence of the phylogeny. Although the environmental factors analysed here did not account for the majority of repeat variation in *Erythrostemon*, they substantially contributed to repeat's composition compared to phylogeny, which alone lacks explanatory power but shows significance when coupled with environmental factors. Despite existing studies investigating the phylogenetic significance of repeat abundances (Dodsworth et al., 2015; Vitales et al., 2020), there is still limited exploration into dissecting and differentiate the contribution of the phylogeny and other factors that may affect repeatome composition across different species.

The environmental diversity in the Neotropics could potentially contribute to the differentiation of repetitive elements, since correlations between genomic and ecological variables have been widely reported in plants (Bilinski et al., 2018; Lyu et al., 2018; Mata-Sucre, Costa, et al., 2020; Souza et al., 2019; Van-Lume et al., 2017). Different ecological variables, such as temperature, humidity, and biotic interactions, can shape the genetic patterns of populations and species (Negi et al., 2016). In *Erythrostemon*, our regression model analyses revealed that precipitation and temperature variables are ecological factors responsible for a fraction of the variation in some TE abundances, especially for Ty3/gypsy-Tekay element. Considering that repeats have been shown to undergo fast evolution (Macas et al., 2015), combined with the age of *Erythrostemon* (~33 Mya) and the high ecological specialization with different species showing niche conservatism (Gagnon et al., 2019), might help to explain the genomic differentiation described here. We suggest that this genome–environment interaction (genome size variation and consequently repeat composition) has been one of the factors shaping *Erythrostemon* genomes since their origin 33 Mya, aligning with the evolutionary trend previously reported for other members of tribe Caesalpinieae (Souza et al., 2019).

Previous analysis of the Mesoamerican *E. hughesii* had revealed a contrasting repeatome composition compared to other Caesalpinieae taxa genomes occurring in Northeastern Brazil (e.g. *Cenostigma* spp., *Libidibia ferrea* (Mart. ex Tul.) L.P.Queiroz and *Paubrasilia echinata* (Lam.) Gagnon, H.C.Lima & G.P.Lewis) (Castro et al., 2024; Mata-Sucre, Sader, et al., 2020; Van-Lume et al., 2019). Our results show that this may be related to a broader pattern, since within *Erythrostemon* the SA and MA sensu stricto clades formed separate groups (with few exceptions) based on their TE clades abundances, with a few lineages even revealing high phylogenetic signal. This repeatome differentiation between clades is mainly mediated by satDNA and Tekay abundance variation. The inverse correlation between LTR repeats and satDNA is well documented in other angiosperms (Ambrožová, et al., 2010; Chumová et al., 2022; Mata-Sucre et al., 2023). This pattern may be elucidated by the concept of genome "carrying capacity" (Schley et al., 2022), suggesting that, like species that inhabit an ecosystem, the genome also has an optimal threshold for accommodating repetitive DNAs and when the genome exceeds its 'optimal size', there is a selective pressure that may trigger repeat elimination (Schley et al., 2022).

The high number of repetitive elements with significant abundance correlation with each other (Figure S5) suggests that the repeats may have a regulatory dynamic between themselves. These interaction between repeats can be analysed using genome ecology concepts (Brookfield, 2005; Venner et al., 2009), which establish relationships such as commensalism (relationship between units of two repetitions in which one copy obtains benefits from the other without harming or benefiting the latter), competition (one repeat is negatively affected by the presence of another) and cooperation (both elements benefit in the same genomic niche) within repetitive elements (Le rouzic et al., 2007; Venner et al., 2009). In this regard, the negative correlation between Tekay and other repeats (ALE, Alesia, Athila, Bianca, CRM, Harbringer and satDNA), along with lower Pielou's evenness associated with Tekay dominance in SA genomes, suggests that interactions among the different TE clades may also modulate repetitive DNA composition in *Erythrostemon*.

Nevertheless, our results suggest that stressful environmental conditions encountered during the colonization of Mesoamerica by *Erythrostemon* may have led to selective stress-mediated Tekay elimination and opportunistic amplification of satDNAs. This differentiated their genomes from the plesiomorphic Tekay-rich condition of the SA repeatomes. In support of this hypothesis, we found a positive correlation between Tekay abundance and genome size. Repetitive DNAs, especially LTRs, are a key driver of genome size variation in plants (Bennetzen et al., 2005; Bennetzen & Wang, 2014; Bureš et al., 2024; Lisch, 2013; Tenaillon et al., 2011). Genome size variations can impact phenotypic traits, leading to selection for smaller genomes, a phenomenon known as the 'nucleotypic effect' (Bennett, 1972). In plants, larger genomes often face selection pressure due to their negative impact on anatomy and physiology leading to reduced fitness (Knight et al., 2005; Vinogradov, 1995, 2003). In contrast, smaller genome sizes promote features such as faster rates of cell division (Francis et al., 2008; Šímová & Herben, 2012), and lower nutrient demands (Peng et al., 2022; Šmarda et al., 2013), which can be important in adaptation to certain environmental conditions (Bureš et al., 2024; Schley et al., 2022). In this context, Tekay reduction and consequently, genome size reduction is likely maintained in the genome due to nucleotypic effects, contributing to the pattern observed in *Erythrostemon* species. These findings align with the latitudinal and environmental trends previously described for genome size in angiosperms that follows an S-shaped distribution towards the globe (Bureš et al., 2024).

4.2 | Historical drivers of repeatome diversification in *Erythrostemon*

Our analysis of the temporal dynamics of Tekay, revealed two bursts of Tekay amplification, around 10 and 5 Mya (Figure S4). These

bursts were observed, with only a few exceptions, across most SA species, and in the earliest diverging MA species (*E. pannosus* and *E. placidus*). We observed a significant positive correlation ($\rho = 0.32$, $p = 2.2e{-}16$) between the age of the Tekay insertions and their abundance in the genome, suggesting that SA lineages with a greater proportion of this element have been accumulating these TEs for a longer time. The Late Miocene-Early Pliocene (ca. 11–3 Ma) was a period characterized by the emergence of new habitats in South America, including the uplift of the Patagonian Andean Cordillera (Pascual et al., 1996). During this period, temperatures declined, and seasonality became more pronounced compared to the middle Miocene (20 Ma) (Donato, 2006; Ortiz-Jaureguizar & Cladera, 2006; Pascual et al., 1996). Thus, we suggest that such substantial changes in environmental conditions likely influenced the dynamics of Tekay in *Erythrostemon*.

Interestingly, *E. coccineus*, the unique MA sensu stricto representative with Tekay-rich repeatome similar to species in the SA clade, exhibited an independent burst of Tekay insertions around 5 Mya. This could be attributed to a major Pliocenic event, such as the closure of the Panama Isthmus, coinciding with the northern hemisphere glaciations in the Late Pliocene (Bartoli et al., 2005). Due to the similar ages of diversification of the Mesoamerican and South American clades (~30 Mya), and the widespread distribution of species in the Americas/Caribbean for other closely related Caesalpinieae genera (e.g. *Cenostigma* and *Libidibia*) (Gagnon et al., 2019) we hypothesize that *Erythrostemon* diversification began in South America. By extension, the common ancestor of the genus would have been widely distributed in South America. This scenario finds support in geological and paleontological evidence which indicates an arid period with expansion and establishment of the Succulent Biome (Arakaki et al., 2011). Moreover, this is corroborated by our results that show that the Tekay and satDNA abundances of the probable ancestor would most likely resemble SA repeatomes (a predominance of Tekay instead of satDNAs) (Figure 5a,b).

An unexpected result was the observation of a species from South America (*E. calycinus*) with a repeatome characteristics similar to the Mesoamerican species, such as low abundance of Tekay. Additionally, some Mesoamerican species (*E. coccineus*, *E. pannosus* and *E. placidus*) presented a repeatome profile similar to South American species. This incongruence could be related to hybridization events in the genus *Erythrostemon*. Interspecific hybridization, leading to reticulate evolution, can cause low resolution in phylogenies, resulting in clades with low support and incongruence between nuclear and plastid trees as observed here (Stefanović & Costea, 2008; Stull et al., 2020). Lewis (1995), based in morphological characters, suggested the existence of natural on-going hybridization among *Erythrostemon* species. This also raises the possibility of hybridization events occurring during the evolution of the genus. Moreover, hybridization can lead to the mobilization of TEs due to potential disruptions in the regulation of TE movement within hybrids after the 'genomic shock' (Smukowski et al., 2021; McClintock, 1984). Interesting, *E. coccineus* exhibits a notable increase in Tekay abundance, a highly divergent repeatome compared to others species in the MA sensu stricto clade. This divergence could be related to a hybrid origin, with activation/silencing of repeats generating an unexpected profile. Therefore, we cannot rule out the hypothesis that ancient hybridization events during *Erythrostemon* diversification could also explain some of the inconsistencies in the repeatome composition among the clades reported in our PCoA analysis. On the other hand, it is important to note that tree discordances in groups that underwent rapid ancient radiation can also occur due to incomplete lineage sorting (ILS), producing patterns similar to those caused by hybridization and introgression (Wang et al., 2014), thus, those phylogenetic incongruences should be analysed with caution (Morales-Briones et al., 2021). Additionally, the absence of data for additional Mesoamerican species should be considered, as these non-analysed species may also exhibit distinct repeatomic patterns that could enhance our understanding of the repetitive DNA variation observed in *Erythrostemon*. Hence, we have not disregarded other potential explanations, beyond hybridization, for the phylogenetic incongruencies and differential repeatomic patterns reported here.

It has been proposed that Central and South America were connected during a pre-Pliocene period (late Eocene and early Oligocene) by a land bridge known as GAARlandia (Greater Antilles Rise-Aves Ridge) (Agnolin et al., 2019; Iturralde-Vinent & MacPhee, 1999, 2023) which may have enabled colonization of Central America during the Oligocene (~20 Mya). Remarkably, the Caribbean species *E. glandulous* originated around 32 Mya (Gagnon et al., 2019), reinforcing the relative ancient distribution of *Erythrostemon*. During the mid-late Eocene (~38–34 Mya), a tendency towards drier seasons initiated (Ortiz-Jaureguizar & Cladera, 2006). This environmental shift would have provided a favourable niche for the ancestral *Erythrostemon* to diversify into and thus colonize the Americas, especially given that the genus has a strong predilection for niche conservatism (Gagnon et al., 2019). In this context, we hypothesize, that the colonization of Mesoamerica, culminating in the specialization of the *MA* sensu stricto subclade may have led to stress-mediated elimination of Tekay and amplification of satDNAs throughout the heterochromatin, resulting in the observed pattern described here (Figure 5). Furthermore, the disjunct distribution of South American species in Chaco/Andes and Northeastern Brazil could be explained by an ancient Chaco/Caatinga biogeographic connection known as the South American Dry Diagonal, an open vegetation belt extending from Northeastern Brazil to northern Argentina that includes several xeric biomes (~19 Mya) (Collevatti et al., 2020; Prado & Gibbs, 1993; Rull & Carnaval, 2020). After this initial expansion throughout Mesoamerica and South America, increasing humidity and subsequent re-expansion of evergreen forests in South America (Prado & Gibbs, 1993; Furley & Ratter, 1988) may have led to the extinction of several highly specialized *Erythrostemon* lineages, which led to the present relict disjunct distribution among seasonally dry biomes. We hypothesize that these historical processes were accompanied by changes in repeat evolution, so that

evolutionary shifts are reflected in the currently contrasting repeatome patterns, marked by an expansion in Tekay abundance in Chaco/Andean species and Tekay removal in the MA-sensu stricto subclade (Figure 5).

The expansion of specific TEs can occur gradually or in bursts (Naito et al., 2009). It has been proposed that bursts of transpositions can occur in response to stressful conditions (McClintock, 1984). Some of these changes in TE content have been shown to be associated with species adaptation to new environments (Schrader & Schmitz, 2019; Stapley et al., 2015). Therefore, considering the relation of Tekay content with ecological variables in *Erythrostemon*, as well as its high abundance in the SA representatives, this element could be involved with the genus adaptation and diversification throughout the Americas. Likewise, the expansion of satDNA in the *MA*-sensu stricto clade is noteworthy. The satDNA is one of the most dynamic type of sequences of the genome, due to its fast evolution rate associated with rapid changes in both structure and abundance (Alix et al., 2017; Burgess et al., 2014; Čertner et al., 2017; Jiao et al., 2011). In this context, these sequences can be a strong source of genetic variation, which can result in reproductive isolation in various ways (Ferree & Prasad, 2012), increase genomic instability and generate infertile progeny (McClintock, 1984), phenomena that can accompany speciation (Levy, 2013). Hence, several studies have revealed differences in satDNA between closely related species in short periods of time (Ahmad et al., 2020; Belyayev et al., 2019; Camacho et al., 2022). This peculiar way of satDNA evolution may have boosted diversification of MA-sensu stricto clade throughout its evolution, promoting rapid differentiation from ancient *Erythrostemon* lineages. Although, previous analysis revealed a constant diversification rate for the tribe Caesalpinieae (Gagnon et al., 2019), interestingly *Erythrostemon* genus stands out for its diversity (31 spp.) when compared to other genera within the tribe. Remarkably, in addition to being the most diverse genus of the tribe, *Erythrostemon* showed the highest intrageneric heterogeneity in the repetitive genomic fraction (Castro et al., 2024; Mata-Sucre, Sader, et al., 2020; Van-Lume et al., 2019), suggesting a relationship between repeatome dynamics and biological diversity.

Overall, our data show that phylogenetic and environmental factors should be carefully considered when examining specific repeat lineages abundance in *Erythrostemon* genomes. Furthermore, intrinsic factors such as reticulate evolution and interactions between different repeat clades may also be taken into account. When considering the plasticity of Ty3/gypsy Tekay and its correlation with latitude and genome size, we suggest that this particular repeat might be key to understand *Erythrostemon* diversification and genome–environment correlations previously described in the tribe Caesalpinieae. This suggests a strong role of the Tekay elements as modulators of the genome–environment interaction, providing macroevolutionary insights about mechanisms of repeatome differentiation and plant diversification across space and time. Additional fine-scale genomics will enable deeper investigation of repeats evolution patterns on a broader biogeographical scale.

AUTHOR CONTRIBUTIONS

N.C. carried out the analysis and data curation, wrote the first draft, and reviewed and edited subsequent drafts. L.C. contributed to data curation, writing—reviewing and editing, and project supervision. B.V. contributed to data curation, resource provision, writing—reviewing and editing. Y.M.-S. contributed to data curation, resource provision, reviewing and editing. A.M. contributed to resource provision, reviewing and editing. E.G. contributed to resource provision, reviewing and editing. G.P.L. contributed to resource provision, reviewing and editing and G.S. contributed to project conceptualization, resource provision, writing—reviewing and editing, project supervision, visualization, and funding acquisition.

ACKNOWLEDGEMENTS

The authors thank the Brazilian agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq number 04/2021) and Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE-APQ-0970-2.03/15; IBPG-1762-2.03/21) for financial support. G.S. thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for a post-doctoral grant (Estágio Pós Doutoral 88881.119479/2016-01). A.M. is financially supported by the Max Planck Society. This study was financed in part by the CAPES (Brazil)—Finance Code 001. G.S. received a productivity fellowship from CNPq (process numbers PQ-312852/2021-5). We also thank the ELIXIR-CZ Research Infrastructure Project (LM2015047) for providing computational resources for RepeatExplorer analysis.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All sequencing data used in this study were submitted to the NCBI under the Bioproject no. PRJNA739461. All other data needed to support the conclusions presented in the paper are provided in the paper and/or the Supplementary Materials. Additional data relating to this study may be requested from the authors.

ORCID

Natália Castro https://orcid.org/0009-0006-7221-0412
Gustavo Souza https://orcid.org/0000-0002-5700-6097

REFERENCES

Abadi, S., Azouri, D., Pupko, T., & Mayrose, I. (2019). Model selection may not be a mandatory step for phylogeny reconstruction. *Nature Communications*, *10*(1), 1. https://doi.org/10.1038/s41467-019-08822-w

Agnolin, F. L., Chimento, N. R., Lucero, S. O., Agnolin, F. L., Chimento, N. R., & Lucero, S. O. (2019). Pre-GABI biotic connections between the Americas: An alternative model to explain the "less-splendid isolation" of South America. *Revista Geológica de América Central*, *61*, 91–106. https://doi.org/10.15517/rgac.v61i0.40089

Ahmad, S. F., Singchat, W., Jehangir, M., Suntronpong, A., Panthum, T., Malaivijitnond, S., & Srikulnath, K. (2020). Dark matter of primate genomes: Satellite DNA repeats and their evolutionary dynamics. *Cells*, *9*(12), 2714. https://doi.org/10.3390/cells9122714

Akaiake, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, *60*(2), 255–265. https://doi.org/10.1093/biomet/60.2.255

Alix, K., Gérard, P. R., Schwarzacher, T., & Heslop-Harrison, J. S. P. (2017). Polyploidy and interspecific hybridization: Partners for adaptation, speciation and evolution in plants. *Annals of Botany*, *120*, 183–194. https://doi.org/10.1093/aob/mcx079

Ambrožová, K., Mandáková, T., Bureš, P., Neumann, P., Leitch, I. J., Koblížková, A., Macas, J., & Lysak, M. A. (2010). Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria* lilies. *Annals of Botany*, *107*(2), 255–268. https://doi.org/10.1093/aob/mcq235

Anderson, M. J. (2005). *Permutational multivariate analysis of variance* (Vol. *26*, pp. 32–46). Department of Statistics, University of Auckland.

Arakaki, M., Christin, P.-A., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R. M., Spriggs, E., Moore, M. J., & Edwards, E. J. (2011). Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proceedings of the National Academy of Sciences*, *108*(20), 8379–8384. https://doi.org/10.1073/pnas.1100628108

Barros e Silva, A. E., Marques, A., dos Santos, K. G., & Guerra, M. (2010). The evolution of CMA bands in *citrus* and related genera. *Chromosome Research*, *18*, 503–514. https://doi.org/10.1007/s10577-010-9130-2

Bartoń, K. (2023). MuMIn: Multi-Model Inference. R package version 1.47.5. https://CRAN.R-project.org/package=MuMIn

Bartoli, G., Sarnthein, M., Weinelt, M., Erlenkeuser, H., Garbe-Schönberg, D., & Lea, D. W. (2005). Final closure of Panama and the onset of northern hemisphere glaciation. *Earth and Planetary Science Letters*, *237*(1–2), 33–44. https://doi.org/10.1016/j.epsl.2005.06.020

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, *19*(1), 134–143. https://doi.org/10.1111/j.1466-8238.2009.00490.x

Beals, E. W. (1984). Bray-Curtis ordination: An effective strategy for analysis of multivariate ecological data. In A. MacFadyen & E. D. Ford (Eds.), *Advances in ecological research* (Vol. *14*, pp. 1–55). Academic Press. https://doi.org/10.1016/S0065-2504(08)60168-3

Belyayev, A., Josefiová, J., Jandová, M., Kalendar, R., Krak, K., & Mandák, B. (2019). Natural history of a satellite DNA family: From the ancestral genome component to species-specific sequences, concerted and non-concerted evolution. *International Journal of Molecular Sciences*, *20*(5), 1201. https://doi.org/10.3390/ijms20051201

Bennett, M. D. (1972). Nuclear DNA content and minimum generation time in herbaceous plants. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *181*(1063), 109–135. https://doi.org/10.1098/rspb.1972.0042

Bennetzen, J. L., Ma, J., & Devos, K. M. (2005). Mechanisms of recent genome size variation in flowering plants. *Annals of Botany*, *95*(1), 127–132. https://doi.org/10.1093/aob/mci008

Bennetzen, J. L., & Wang, H. (2014). The Contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annual Review of Plant Biology*, *65*(1), 505–530. https://doi.org/10.1146/annurev-arplant-050213-035811

Bilinski, P., Albert, P. S., Berg, J. J., Birchler, J. A., Grote, M. N., Lorant, A., Quezada, J., Swarts, K., Yang, J., & Ross-Ibarra, J. (2018). Parallel altitudinal clines reveal trends in adaptive evolution of genome size in *Zea mays*. *PLoS Genetics*, *14*(5), e1007162. https://doi.org/10.1371/journal.pgen.1007162

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology*, *19*(1), 199. https://doi.org/10.1186/s13059-018-1577-z

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, *27*(4), 326–349. https://doi.org/10.2307/1942268

Brookfield, J. F. Y. (2005). The ecology of the genome—Mobile DNA elements and their hosts. *Nature Reviews Genetics*, *6*(2), 128–136. https://doi.org/10.1038/nrg1524

Bureš, P., Elliott, T. L., Veselý, P., Šmarda, P., Forest, F., Leitch, I. J., Nic Lughadha, E., Soto Gomez, M., Pironon, S., Brown, M. J. M., Šmerda, J., & Zedek, F. (2024). The global distribution of angiosperm genome size is shaped by climate. *New Phytologist*, *242*(2), 744–759. https://doi.org/10.1111/nph.1954

Burgess, M. B., Cushman, K. R., Doucette, E. T., Talent, N., Frye, C. T., & Campbell, C. S. (2014). Effects of apomixis and polyploidy on diversification and geographic distribution in *Amelanchier* (Rosaceae). *American Journal of Botany*, *101*(8), 1375–1387. https://doi.org/10.3732/ajb.1400113

Calcagno, V., & de Mazancourt, C. (2010). glmulti: An R package for easy automated model selection with (generalized) linear models. *Journal of Statistical Software*, *34*(12), 1–29. https://doi.org/10.18637/jss.v034.i12

Camacho, J. P. M., Cabrero, J., López-León, M. D., Martín-Peciña, M., Perfectti, F., Garrido-Ramos, M. A., & Ruiz-Ruano, F. J. (2022). Satellitome comparison of two oedipodine grasshoppers highlights the contingent nature of satellite DNA evolution. *BMC Biology*, *20*(1), 36. https://doi.org/10.1186/s12915-022-01260-z

Canapa, A., Biscotti, M. A., Barucca, M., Carducci, F., Carotti, E., & Olmo, E. (2020). Shedding light upon the complex net of genome size, genome composition and environment in chordates. *The European Zoological Journal*, *87*(1), 192–202. https://doi.org/10.1080/24750263.2020.1747558

Castro, N., Mata-Sucre, Y., Carvalho-Sobrinho, J., Marques, A., de Queiroz, R. T., & Souza, G. (2024). Genomic stability in Cenostigma Tul., (Caesalpinioideae, Fabaceae): Causes and consequences. *Botanical Journal of the Linnean Society*, *204*(2), 137–151. https://doi.org/10.1093/botlinnean/boad043

Čertner, M., Fenclová, E., Kúr, P., Kolář, F., Koutecký, P., Krahulcová, A., & Suda, J. (2017). Evolutionary dynamics of mixed-ploidy populations in an annual herb: Dispersal, local persistence and recurrent origins of polyploids. *Annals of Botany*, *120*(2), 303–315. https://doi.org/10.1093/aob/mcx032

Chumová, Z., Belyayev, A., Mandáková, T., Zeisek, V., Hodková, E., Šemberová, K., Euston-Brown, D., & Trávníček, P. (2022). The relationship between transposable elements and ecological niches in the greater cape floristic region: A study on the genus *Pteronia* (Asteraceae). *Frontiers in Plant Science*, *13*, 982852. https://doi.org/10.3389/fpls.2022.982852

Collevatti, R. G., Lima, N. E., & Vitorino, L. C. (2020). The diversification of extant angiosperms in the South America dry diagonal. In V. Rull & A. C. Carnaval (Eds.), *Neotropical diversification: Patterns and processes* (pp. 547–568). Springer International Publishing. https://doi.org/10.1007/978-3-030-31167-4_21

Cordaux, R., & Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics*, *10*(10), 691–703. https://doi.org/10.1038/nrg2640

Dodsworth, S., Chase, M. W., Kelly, L. J., Leitch, I. J., Macas, J., Novak, P., Piednoël, M., Weiss-Schneeweiss, H., & Leitch, A. R. (2015). Genomic repeat abundances contain phylogenetic signal. *Systematic Biology*, *64*(1), 112–126.

Donato, M. (2006). Historical biogeography of the family Tristiridae (Orthoptera: Acridomorpha) applying dispersal–vicariance analysis. *Journal of Arid Environments*, *66*(3), 421–434. https://doi.org/10.1016/j.jaridenv.2006.01.006

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, *7*(1), 214. https://doi.org/10.1186/1471-2148-7-214

Feiner, N. (2016). Accumulation of transposable elements in Hox gene clusters during adaptive radiation of Anolis lizards. *Proceedings*

of the Royal Society B: Biological Sciences, *283*(1840), 20161555. https://doi.org/10.1098/rspb.2016.1555

Ferree, P. M., & Prasad, S. (2012). How can satellite DNA divergence cause reproductive isolation? Let us count the chromosomal ways. *Genetics Research International*, *2012*, 1–11. https://doi.org/10.1155/2012/430136

Ferreira, M. E., & Grattapaglia, D. (1995). *Introdução ao uso de marcadores RAPD e RFLP em análise genética*. Embrapa-Cenargen Brasília.

Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, *13*(3), 252–264. https://doi.org/10.1111/j.1472-4642.2007.00341.x

Fitzpatrick, M., Mokany, K., Manion, G., Nieto-Lugilde, D., & Ferrier, S. (2022). gdm: Generalized Dissimilarity Modeling. https://mfitzpatrick.al.umces.edu/gdm/, https://github.com/fitzLab-AL/gdm/

Francis, D., Davies, M. S., & Barlow, P. W. (2008). A strong nucleotypic effect on the cellcycle regardless of ploidy level. *Annals of Botany*, *101*, 747–757. https://doi.org/10.1093/aob/mcn038

Furley, P. A., & Ratter, J. A. (1988). Soil resources and plant communities of the Central Brazilian Cerrado and their development. *Journal of Biogeography*, *15*(1), 97. https://doi.org/10.2307/2845050

Gagnon, E., Bruneau, A., Hughes, C. E., de Queiroz, L., & Lewis, G. P. (2016). A new generic system for the pantropical Caesalpinia group (Leguminosae). *PhytoKeys*, *71*, 1–160. https://doi.org/10.3897/phytokeys.71.9203

Gagnon, E., Clark, R. P., Ringelberg, J. J., & Lewis, G. P. (2024). 6. Tribe Caesalpinieae. In A. Bruneau, L. P. Queiroz, & J. J. Ringelberg (Eds.), *Advances in legume systematics 14. Classification of Caesalpinioideae. Part 2: Higher-level classification* (Vol. *240*, pp. 103–145). PhytoKeys. https://doi.org/10.3897/phytokeys.240.101716

Gagnon, E., Ringelberg, J., Lewis, G. P., Bruneau, A., & Hughes, C. E. (2018). Species occurrence dataset of the pantropical Caesalpinia Group. v1. Université de Montréal Biodiversity Centre. Dataset/Occurrence. http://data.canadensys.net/ipt/resource?r=caesalpinia-gagnon&v=1.0

Gagnon, E., Ringelberg, J. J., Bruneau, A., Lewis, G. P., & Hughes, C. E. (2019). Global succulent biome phylogenetic conservatism across the pantropical Caesalpinia group (Leguminosae). *New Phytologist*, *222*(4), 1994–2008. https://doi.org/10.1111/nph.15633

Galindo-González, L., Mhiri, C., Deyholos, M. K., & Grandbastien, M.-A. (2017). LTR-retrotransposons in plants: Engines of evolution. *Gene*, *626*, 14–25. https://doi.org/10.1016/j.gene.2017.04.051

Garrido-Ramos, M. A. (2015). Satellite DNA in plants: More than just rubbish. *Cytogenetic and Genome Research*, *146*(2), 153–170. https://doi.org/10.1159/000437008

Gattoni, K., Gendron, E. M. S., Borgmeier, A., McQueen, J. P., Mullin, P. G., Powers, K., Powers, T. O., & Porazinska, D. L. (2022). Context-dependent role of abiotic and biotic factors structuring nematode communities along two environmental gradients. *Molecular Ecology*, *31*(14), 3903–3916. https://doi.org/10.1111/mec.16541

Gernhard, T. (2008). The conditioned reconstructed process. *Journal of Theoretical Biology*, *253*(4), 769–778. https://doi.org/10.1016/j.jtbi.2008.04.005

Goodier, J. L., & Kazazian, H. H. (2008). Retrotransposons revisited: The restraint and rehabilitation of parasites. *Cell*, *135*(1), 23–35. https://doi.org/10.1016/j.cell.2008.09.022

Gregory, T. R. (2005). The C-value enigma in plants and animals: A review of parallels and an appeal for partnership. *Annals of Botany*, *95*(1), 133–146. https://doi.org/10.1093/aob/mci009

Grotkopp, E., Rejmánek, M., Sanderson, M. J., & Rost, T. L. (2004). Evolution of genome size in pines (*Pinus*) and its life-history correlates: Supertree analyses. *Evolution*, *58*(8), 1705–1729. https://doi.org/10.1111/j.0014-3820.2004.tb00456.x

Guo, L. Y., Bian, J., Davis, A. E., Liu, P., Kempton, H. R., Zhang, X., Chemparathy, A., Gu, B., Lin, X., Rane, D. A., Xu, X., Jamiolkowski, R. M., Hu, Y., Wang, S., & Qi, L. S. (2022). Multiplexed genome regulation in vivo with hyper-efficient Cas12a. *Nature Cell Biology*, *24*(4), 590–600. https://doi.org/10.1038/s41556-022-00870-7

Hannan, A. J. (2018). Tandem repeats mediating genetic plasticity in health and disease. *Nature Reviews Genetics*, *19*(5), 286–298. https://doi.org/10.1038/nrg.2017.115

Hijmans, R. (2023). raster: Geographic Data Analysis and Modeling: R package version 3.6-26. https://CRAN.R-project.org/package=raster

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Holland, B., & Moulton, V. (2003). Consensus networks: A method for visualising incompatibilities in collections of trees. In *International workshop on algorithms in bioinformatics* (pp. 165–176). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-39763-2_13

Hultgren, K. M., Jeffery, N. W., Moran, A., & Gregory, T. R. (2018). Latitudinal variation in genome size in crustaceans. *Biological Journal of the Linnean Society*, *123*(2), 348–359. https://doi.org/10.1093/biolinnean/blx153

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. https://doi.org/10.1093/molbev/msj030

Iturralde-Vinent, M., & MacPhee, R. D. (1999). Paleogeography of the Caribbean region: Implications for Cenozoic biogeography. *Bulletin of the AMNH*, *238*, 1–95.

Iturralde-Vinent, M. A., & MacPhee, R. D. (2023). New evidence for late Eocene-early Oligocene uplift of Aves ridge and paleogeography of GAARlandia. *Geologica Acta*, *21*, 1–10.

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., & Depamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, *473*(7345), 97–100. https://doi.org/10.5061/dryad.8546

Jouffroy, O., Saha, S., Mueller, L., Quesneville, H., & Maumus, F. (2016). Comprehensive repeatome annotation reveals strong potential impact of repetitive elements on tomato ripening. *BMC Genomics*, *17*(1), 624. https://doi.org/10.1186/s12864-016-2980-z

Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., & Schulman, A. H. (2000). Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proceedings of the National Academy of Sciences*, *97*(12), 6603–6607. https://doi.org/10.4073/pnas.110587497

Kanazawa, A., Liu, B., Kong, F., Arase, S., & Abe, J. (2009). Adaptive evolution involving gene duplication and insertion of a novel Ty1/copia-like retrotransposon in soybean. *Journal of Molecular Evolution*, *69*(2), 164–175. https://doi.org/10.1007/s00239-009-9262-1

Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B.-K., Jun, T. H., Hwang, W. J., Lee, T., Lee, J., Shim, S., Yoon, M. Y., Jang, Y. E., Han, K. S., Taeprayoon, P., Yoon, N., Somta, P., Tanya, P., Kim, K. S., … Lee, S.-H. (2014). Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nature Communications*, *5*(1), 1. https://doi.org/10.1038/ncomms6443

Katoh, K. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Katoh, K., & Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, *9*(4), 286–298. https://doi.org/10.1093/bib/bbn013

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120. https://doi.org/10.1007/bf01731581

Knight, C. A., Molinari, N. A., & Petrov, D. A. (2005). The large genome constraint hypothesis: Evolution, ecology and phenotype. *Annals of Botany*, *95*(1), 177–190. https://doi.org/10.1093/aob/mci011

Le Rouzic, A., Dupas, S., & Capy, P. (2007). Genome ecosystem and transposable elements species. *Gene*, *390*(1–2), 214–220. https://doi.org/10.1016/j.gene.2006.09.023

Levy, A. A. (2013). Transposons in plant speciation. In N. V. Fedoroff (Ed.), *Plant transposons and genome dynamics in evolution* (pp. 164–179). John Wiley and Sons Inc. https://doi.org/10.1002/9781118500156.ch9

Lewis, G. P. (1995). Systematic studies in neotropical 'Caesalpina L.' (Leguminosae:S2 Caesalpinioideae), including a revision of the 'Poincianella-Erythrostemon' group. University of St. Andrews (United Kingdom).

Lisch, D. (2013). How important are transposons for plant evolution? *Nature Reviews Genetics*, *14*(1), 49–61. https://doi.org/10.1038/nrg3374

Lopez-Toledo, L., Prieto-Torres, D. A., Barros, F. D. V., Ribeiro, N., & Pennington, R. T. (2024). Seasonally dry tropical forests: New insights for their knowledge and conservation. *Frontiers in Forests and Global Change*, *6*, 1350375. https://doi.org/10.3389/ffgc.2023.1350375

Lwin, A. K., Bertolini, E., Pè, M. E., & Zuccolo, A. (2017). Genomic skimming for identification of medium/highly abundant transposable elements in *Arundo donax* and *Arundo plinii*. *Molecular Genetics and Genomics*, *292*, 157–171. https://doi.org/10.1007/s00438-016-1263-3

Lysák, M. A., Rostková, A., Dixon, J. M., Rossi, G., & Doležel, J. (2000). Limited genome size variation in *Sesleria albicans*. *Annals of Botany*, *86*(2), 399–403. https://doi.org/10.1006/anbo.2000.1200

Lyu, H., He, Z., Wu, C.-I., & Shi, S. (2018). Convergent adaptive evolution in marginal environments: Unloading transposable elements as a common strategy among mangrove genomes. *New Phytologist*, *217*(1), 428–438. https://doi.org/10.1111/nph.14784

Macas, J., Novák, P., Pellicer, J., Čížková, J., Koblížková, A., Neumann, P., Fuková, I., Doležel, J., Kelly, L. J., & Leitch, I. J. (2015). In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. *PLoS One*, *10*(11), e0143424. https://doi.org/10.1371/journal.pone.0143424

Mascagni, F., Vangelisti, A., Giordani, T., Cavallini, A., & Natali, L. (2020). A computational comparative study of the repetitive DNA in the genus *Quercus* L. *Tree Genetics & Genomes*, *16*, 1–11. https://doi.org/10.1007/s11295-019-1401-2

Mata-Sucre, Y., Costa, L., Gagnon, E., Lewis, G. P., Leitch, I. J., & Souza, G. (2020). Revisiting the cytomolecular evolution of the Caesalpinia group (Leguminosae): A broad sampling reveals new correlations between cytogenetic and environmental variables. *Plant Systematics and Evolution*, *306*(2), 48. https://doi.org/10.1007/s00606-020-01674-8

Mata-Sucre, Y., Matzenauer, W., Castro, N., Huettel, B., Pedrosa-Harand, A., Marques, A., & Souza, G. (2023). Repeat-based phylogenomics shed light on unclear relationships in the monocentric genus *Juncus* L. (Juncaceae). *Molecular Phylogenetics and Evolution*, *189*, 107930. https://doi.org/10.1016/j.ympev.2023.107930

Mata-Sucre, Y., Sader, M., Van-Lume, B., Gagnon, E., Pedrosa-Harand, A., Leitch, I. J., Lewis, G. P., & Souza, G. (2020). How diverse is heterochromatin in the Caesalpinia group? Cytogenomic characterization of *Erythrostemon hughesii* Gagnon & G.P. Lewis (Leguminosae: Caesalpinioideae). *Planta*, *252*(4), 49. https://doi.org/10.1007/s00425-020-03453-8

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science*, *226*, 792–801. https://doi.org/10.1126/science.15739260

Moraes, A. P., Engel, T. B. J., Forni-Martins, E. R., de Barros, F., Felix, L. P., & Cabral, J. S. (2022). Are chromosome number and genome size associated with habit and environmental niche variables? Insights from the Neotropical orchids. *Annals of Botany*, *130*(1), 11–25. https://doi.org/10.1093/aob/mcac021

Morales-Briones, D. F., Kadereit, G., Tefarikis, D. T., Moore, M. J., Smith, S. A., Brockington, S. F., Timoneda, A., Yim, W. C., & Yang, Y. (2021). Disentangling sources of gene tree discordance in phylogenomic data sets: Testing ancient hybridizations in Amaranthaceae sl. *Systematic Biology*, *70*(2), 219–235. https://doi.org/10.1093/sysbio/syaa066

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C. N., Richardson, A. O., Okumoto, Y., Tanisaka, T., & Wessler, S. R. (2009). Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*, *461*(7267), 1130–1134. https://doi.org/10.1038/nature08479

Negi, P., Rai, A. N., & Suprasanna, P. (2016). Moving through the stressed genome: Emerging regulatory roles for transposons in plant stress response. *Frontiers in Plant Science*, *7*, 1448. https://doi.org/10.3389/fpls.2016.01448

Neumann, P., Novák, P., Hoštáková, N., & Macas, J. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA*, *10*(1), 1. https://doi.org/10.1186/s13100-018-0144-1

Novák, P., Neumann, P., & Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nature Protocols*, *15*(11), 3745–3776. https://doi.org/10.1038/s41596-020-0400-y

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., … Weedon, J. (2022). vegan: Community Ecology: R package version 2.6-4. https://CRAN.R-project.org/package=vegan

Oliver, K. R., McComb, J. A., & Greene, W. K. (2013). Transposable elements: Powerful contributors to angiosperm evolution and diversity. *Genome Biology and Evolution*, *5*(10), 1886–1901. https://doi.org/10.1093/gbe/evt141

Ortiz-Jaureguizar, E., & Cladera, G. A. (2006). Paleoenvironmental evolution of southern South America during the Cenozoic. *Journal of Arid Environments*, *66*(3), 498–532. https://doi.org/10.1016/j.jaridenv.2006.01.007

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884. https://doi.org/10.1038/44766

Pascual, R., Ortiz-Jaureguizar, E., & Prado, J. L. (1996). Land mammals: Paradigm for Cenozoic South American geobiotic evolution. *Münchner Geowissenschaftliche Abhandlungen*, *30*(A), 265–319.

Pellicer, J., Hidalgo, O., Dodsworth, S., & Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes*, *9*(2), 2. https://doi.org/10.3390/genes9020088

Peng, Y., Yang, J., Leitch, I. J., Guignard, M. S., Seabloom, E. W., Cao, D., Zhao, F., Li, H., Xingguo, H., Jiang, Y., Leitch, A. R., & Wei, C. (2022). Plant genome size modulates grassland community responses

to multi-nutrient additions. *New Phytologist*, *236*(6), 2091–2102. https://doi.org/10.1111/nph.18496

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, *13*, 131–144. https://doi.org/10.1016/0022-5193(66)90013-0

Pita, S., Díaz-Viraqué, F., Iraola, G., & Robello, C. (2019). The Tritryps comparative Repeatome: Insights on repetitive element evolution in Trypanosomatid pathogens. *Genome Biology and Evolution*, *11*(2), 546–551. https://doi.org/10.1093/gbe/evz017

Pluess, A. R., Frank, A., Heiri, C., Lalagüe, H., Vendramin, G. G., & Oddou-Muratorio, S. (2016). Genome–environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytologist*, *210*(2), 589–601. https://doi.org/10.1111/nph.13809

Pons, J., & Gillespie, R. G. (2004). Evolution of satellite DNAs in a radiation of endemic Hawaiian spiders: Does concerted evolution of highly repetitive sequences reflect evolutionary history? *Journal of Molecular Evolution*, *59*, 632–641. https://doi.org/10.1007/s00239-004-2655-2

Prado, D. E., & Gibbs, P. E. (1993). Patterns of species distributions in the dry seasonal forests of South America. *Annals of the Missouri Botanical Garden*, *80*(4), 902–927. https://doi.org/10.2307/2399937

Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, *26*(7), 1641–1650. https://doi.org/10.1093/molbev/msp077

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rambaut, A., & Drummond, A. J. (2013). TreeAnnotator v1. 7.0. http://beast.bio.ed.ac.uk

Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Ricci, M., Peona, V., Guichard, E., Taccioli, C., & Boattini, A. (2018). Transposable elements activity is positively related to rate of speciation in mammals. *Journal of Molecular Evolution*, *86*(5), 303–310. https://doi.org/10.1007/s00239-018-9847-7

Richardson, S. R., Doucet, A. J., Kopera, H. C., Moldovan, J. B., Garcia-Perez, J. L., & Moran, J. V. (2015). The influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Mobile DNA*, *III*, 1165–1208. https://doi.org/10.1128/9781555819217.ch51

Rull, V., & Carnaval, A. C. (Eds.). (2020). *Neotropical diversification: Patterns and processes*. Springer International Publishing. https://doi.org/10.1007/978-3-030-31167-4

Šatović-Vukšić, E., & Plohl, M. (2023). Satellite DNAs—From localized to highly dispersed genome components. *Genes*, *14*(3), 742. https://doi.org/10.3390/gene14030742

Schley, R. J., Pellicer, J., Ge, X.-J., Barrett, C., Bellot, S., Guignard, M. S., Novák, P., Suda, J., Fraser, D., Baker, W. J., Dodsworth, S., Macas, J., Leitch, A. R., & Leitch, I. J. (2022). The ecology of palm genomes: Repeat-associated genome size expansion is constrained by aridity. *New Phytologist*, *236*(2), 433–446. https://doi.org/10.1111/nph.18323

Schrader, L., & Schmitz, J. (2019). The impact of transposable elements in adaptive evolution. *Molecular Ecology*, *28*(6), 1537–1549. https://doi.org/10.1111/mec.14794

Šímová, I., & Herben, T. (2012). Geometrical constraints in the scaling relationships between genome size, cell size and cell cycle length in herbaceous plants. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1730), 867–875. https://doi.org/10.1098/rspb.2011.1284

Šmarda, P., Hejcman, M., Březinová, A., Horová, L., Steigerová, H., Zedek, F., Bureš, P., Hejcmanová, P., & Schellberg, J. (2013). Effect of phosphorus availability on the selection of species with different ploidy levels and genome sizes in a long-term grassland fertilization experiment. *New Phytologist*, *200*(3), 911–921. https://doi.org/10.1111/nph.12399

Smukowski Heil, C., Patterson, K., Hickey, A. S. M., Alcantara, E., & Dunham, M. J. (2021). Transposable element mobilization in interspecific yeast hybrids. *Genome Biology and Evolution*, *13*(3), evab033. https://doi.org/10.1093/gbe/evab033

Souza, G., Costa, L., Guignard, M. S., Van-Lume, B., Pellicer, J., Gagnon, E., Leitch, I. J., & Lewis, G. P. (2019). Do tropical plants have smaller genomes? Correlation between genome size and climatic variables in the Caesalpinia group (Caesalpinioideae, Leguminosae). *Perspectives in Plant Ecology, Evolution and Systematics*, *38*, 13–23. https://doi.org/10.1016/j.ppees.2019.03.002

Stapley, J., Santure, A. W., & Dennis, S. R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular Ecology*, *24*(9), 2241–2252. https://doi.org/10.1111/mec.13089

Stefanović, S., & Costea, M. (2008). Reticulate evolution in the parasitic genus Cuscuta (Convolvulaceae): Over and over again. *Botany*, *86*(8), 791–808. https://doi.org/10.1139/B08-033

Stull, G. W., Soltis, P. S., Soltis, D. E., Gitzendanner, M. A., & Smith, S. A. (2020). Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *American Journal of Botany*, *107*(5), 790–805. https://doi.org/10.1002/ajb2.1468

Sun, D., Jeannot, K., Xiao, Y., & Knapp, C. W. (2019). Editorial: Horizontal gene transfer mediated bacterial antibiotic resistance. *Frontiers in Microbiology*, *10*, 1933. https://doi.org/10.3389/fmicb.2019.01933

Tenaillon, M. I., Hufford, M. B., Gaut, B. S., & Ross-Ibarra, J. (2011). Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. *Genome Biology and Evolution*, *3*, 219–229. https://doi.org/10.1093/gbe/evr008

Van-Lume, B., Esposito, T., Diniz-Filho, J. A. F., Gagnon, E., Lewis, G. P., & Souza, G. (2017). Heterochromatic and cytomolecular diversification in the Caesalpinia group (Leguminosae): Relationships between phylogenetic and cytogeographical data. *Perspectives in Plant Ecology, Evolution and Systematics*, *29*, 51–63. https://doi.org/10.1016/j.ppees.2017.11.004

Usai, G., Mascagni, F., Natali, L., Giordani, T., & Cavallini, A. (2017). Comparative genome-wide analysis of repetitive DNA in the genus *Populus* L. *Tree Genetics & Genomes*, *13*(5), 96. https://doi.org/10.1007/s11295-017-1181-5

Van-Lume, B., Mata-Sucre, Y., Báez, M., Ribeiro, T., Huettel, B., Gagnon, E., Leitch, I. J., Pedrosa-Harand, A., Lewis, G. P., & Souza, G. (2019). Evolutionary convergence or homology? Comparative cytogenomics of Caesalpinia group species (Leguminosae) reveals diversification in the pericentromeric heterochromatic composition. *Planta*, *250*(6), 2173–2186. https://doi.org/10.1007/s00425-019-03287-z

Venner, S., Feschotte, C., & Biémont, C. (2009). Dynamics of transposable elements: Towards a community ecology of the genome. *Trends in Genetics*, *25*(7), 317–323. https://doi.org/10.1016/j.tig.2009.05.003

Vinogradov, A. E. (1995). Nucleotypic effect in Homeotherms: Body-mass-corrected basal metabolic rate of mammals is related to genome size. *Evolution*, *49*(6), 1249–1259. https://doi.org/10.1111/j.1558-5646.1995.tb04451.x

Vinogradov, A. E. (2003). Selfish DNA is maladaptive: Evidence from the plant red list. *Trends in Genetics*, *19*(11), 609–614. https://doi.org/10.1016/j.tig.2003.09.010

Vitales, D., Garcia, S., & Dodsworth, S. (2020). Reconstructing phylogenetic relationships based on repeat sequence similarities. *Molecular*

Phylogenetics and Evolution, *147*, 106766. https://doi.org/10.1016/j.ympev.2020.106766

Wang, W., Kerstetter, R. A., & Michael, T. P. (2011). Evolution of genome size in duckweeds (*Lemnaceae*). *Journal of Botany*, *2011*, 1–9. https://doi.org/10.1155/2011/570319

Wang, W., Li, H., & Chen, Z. (2014). Analysis of plastid and nuclear DNA data in plant phylogenetics—Evaluation and improvement. *Science China Life Sciences*, *57*, 280–286. https://doi.org/10.1007/s11427-014-4620-7

Wendel, J. F., Jackson, S. A., Meyers, B. C., & Wing, R. A. (2016). Evolution of plant genome architecture. *Genome Biology*, *17*(1), 37. https://doi.org/10.1186/s13059-016-0908-1

Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, *30*(3), 279–338. https://doi.org/10.2307/1943563

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973–982. https://doi.org/10.1038/nrg2165

Wickham, H. (2016). Getting started with ggplot2. In *Ggplot2* (pp. 11–31). Springer. https://doi.org/10.1007/978-3-319-24277-4_2

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Castro, N., Vilela, B., Mata-Sucre, Y., Marques, A., Gagnon, E., Lewis, G. P., Costa, L., & Souza, G. (2024). Repeatome evolution across space and time: Unravelling repeats dynamics in the plant genus *Erythrostemon* Klotzsch (Leguminosae Juss). *Molecular Ecology*, *00*, e17510. https://doi.org/10.1111/mec.17510