Short structural variation fuelled CAM evolution within an explosive bromeliad radiation

Clara Groot Crego[1,2], Jaqueline Hess[1,3], Gil Yardeni[1,4], Marylaure de La Harpe[1,5], Francesca Beclin[1,2,6], Luiz A. Cauz-Santos[1], Sarah Saadain[1], Thelma Barbará[1], Eva M. Temsch[1], Hanna Weiss-Schneeweiss[1], Michael H.J. Barfuss[1], Walter Till[1], Karolina Heyduk[7,8], Christian Lexer[1], Ovidiu Paun[1*], Thibault Leroy[1,9*]

[1]Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria

[2]Vienna Graduate School of Population Genetics, Vienna, Austria

[3]Cambrium GmbH, Max-Urich-Str. 3, 13055 Berlin, Germany

[4]Institute of Computational Biology, Department of Biotechnology, University of Life Sciences and Natural Resources (BOKU), Muthgasse 18, 1190 Vienna, Austria.

[5]Office for Nature and Environment, Canton of Grisons, Chur, Switzerland

[6]Gregor Mendel Institute, Austrian Academy of Sciences, Vienna BioCenter, Vienna, Austria

[7]School of Life Sciences, University of Hawai‘i at Mānoa, Honolulu, HI USA

[8]Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT USA

[9]GenPhySE, INRAE, INP, ENVT, Université de Toulouse, Castanet-Tolosan, France

*Shared authorship

Correspondence to: clara.groot.crego@univie.ac.at, ORCID: 0000-0002-4547-3608

1. Abstract

Identifying the drivers of trait evolution and diversification is central to understanding plant diversity and evolution. The subgenus *Tillandsia* (Bromeliaceae) belongs to one of the fastest radiating clades in the plant kingdom and is characterised by the repeated evolution of the water-conserving Crassulacean Acid Metabolism (CAM). Despite its complex genetic basis, CAM has evolved independently across many plant families and over short timescales. By producing the first high-quality genome assemblies of a species pair representing a recent CAM/C3 shift, we were able to pinpoint the genomic drivers of trait evolution and diversification in *Tillandsia*. We combined genome-wide investigations of synteny, TE dynamics, sequence evolution, gene family evolution and differential expression to highlight the crucial role of rapid gene family expansion and transposable element activity associated with differentially expressed genes in fuelling CAM/C3 shifts in this vast plant radiation.

2. Introduction

The evolution of key innovation traits as a means to access novel niches has been described as an important mechanism that stimulates species diversification[1,2]. The water-conserving mode of photosynthesis known as Crassulacean Acid Metabolism (CAM) is an important trait for plant adaptation to arid environments and the epiphytic lifeform[3]. CAM functions as a carbon concentrating mechanism by shifting CO_2 assimilation to the night-time. This has the dual effect of enhancing the efficiency of both Rubisco, the first enzyme of the Calvin cycle, and of overall water use, as stomata remain closed during the day, which prevents evapotranspiration[4]. CAM has evolved repeatedly in at least 35 plant families[5], raising questions on the possible mechanisms

47 allowing this complex trait to continuously re-emerge throughout plant history. Achieving a better
48 understanding of the repeated evolution of CAM does not only contribute to our understanding of
49 complex traits but also provides opportunities to enhance the water use efficiency of agricultural
50 crops, a pressing matter in a future shrouded by drought and food insecurity linked to climate
51 change[4].

52 Though abundant ecological opportunity is traditionally regarded as the main driver of
53 rapid diversification in adaptive radiations, it is unclear what mechanisms can provide genetic
54 variation rapidly enough to be recruited over short timescales. Recent studies on adaptive radiation
55 have pointed at the role of large-scale rearrangements[6], indels[7], transposable element (TE)
56 dynamics[8], coding sequence evolution and regulatory divergence[8] as intrinsic drivers shaping
57 diversification, suggesting that lineages with elevated rates of genome evolution, also coined as
58 "genomic potential"[7], may be more prone than others to rapidly diversify and evolve key
59 innovation traits such as CAM.

60 TE dynamics can generate functional variation which has been shown to play a role in local
61 adaptation in *Arabidopsis*[9] and in the evolution of reproductive barriers in *Corvus*[10]. Chromosomal
62 fusions, inversions or translocations can increase linkage between co-adapted alleles and generate
63 reproductive barriers[11,12]. Both TE activity and chromosomal fusion have been associated with
64 elevated speciation rates[13,14]. Gene duplication occurs at higher rates than point mutation in many
65 lineages[15] and can lead to novel functional variation through dosage effects, neo-functionalization,
66 or subfunctionalization, as observed in teleost fish[16,17] and orchids[18].

67 The adaptive radiation of *Tillandsia* subgenus *Tillandsia* (Bromeliaceae) is part of one of
68 the fastest diversifying clades known in the plant kingdom (Tillandsioideae)[19] and is characterized
69 by multiple key innovation traits driving extraordinary diversity both on the taxonomic and

ecological level[20]. CAM has been described as an ecological driver of diversification in the subgenus *Tillandsia*[20], and across Bromeliaceae in general[19]. Though Bromeliaceae are regarded as a homoploid radiation with conserved chromosome counts and little genome size variation[21], more recent work has pointed at the high "genomic potential" of the subgenus *Tillandsia*, notably from elevated gene loss and duplication rates[22] and high transposable element dynamics (Neil McNair, personal communication).

In this study, we present and investigate the *de novo* assembled genomes of two ecologically divergent members of the subgenus *Tillandsia* to further our understanding of the drivers of this recent radiation. *Tillandsia fasciculata* (Fig. 1a). displays a set of phenotypes typically described as "grey" Tillandsia[23]: a dense layer of absorptive trichomes, CAM photosynthesis[22,24] and occurrence in arid places with high solar incidence and low rainfall. On the other hand, *T. leiboldiana* (Fig. 1b) is a typical "green" *Tillandsia*: a C3 plant[22,24] that displays tank formation, lacks a dense layer of trichomes and occurs in cooler, wetter regions. The two species belong to sister clades representing a clear CAM/C3 shift (Fig. 1c). This is the first study to produce high-quality bromelioid reference genomes representing a CAM/C3 shift at short evolutionary timescales. By investigating synteny, molecular evolution, gene family evolution and differential gene expression, we find evidence that CAM/C3 differences are largely regulatory, but have been aided by gene family expansion and transposon activity.

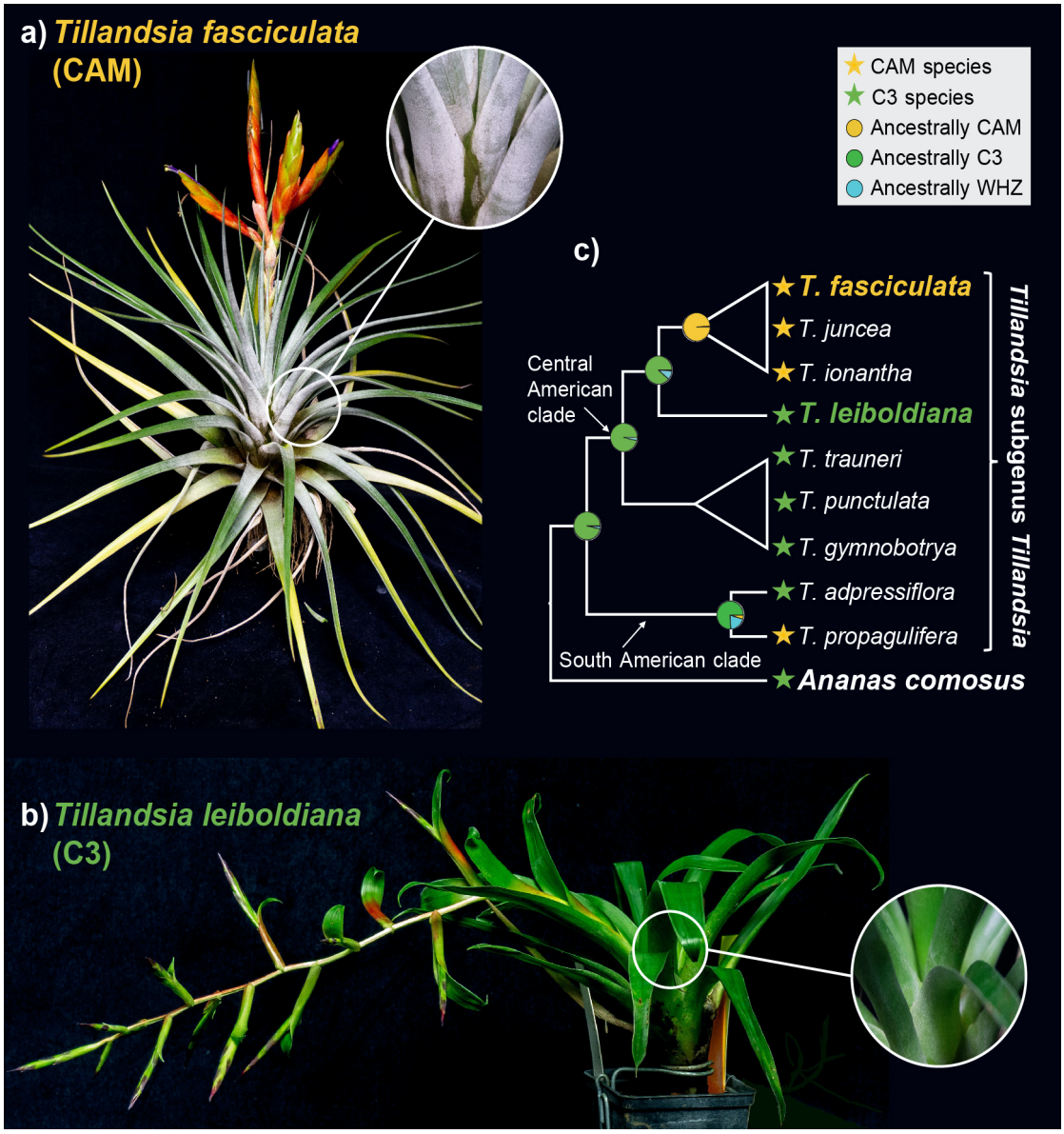


Figure 1: a) Tillandsia fasciculata*, a typical "grey"* Tillandsia *with a dense layer of trichomes and CAM photosynthesis.* ***b)*** Tillandsia leiboldiana*, a green* Tillandsia *with C3 photosynthesis, an impounding tank and few trichomes.* ***c)*** *Schematic representation of the evolutionary relationship between the two investigated species of* Tillandsia*. Stars indicate whether a species performs CAM or C3*[22]*. Pie charts at internal nodes show the ancestral state of photosynthetic metabolism as reported in* [22]*. WHZ stands for Winter-Holtum Zone and represents intermediate forms of the CAM/C3 spectrum.*

3. Results

3.1. Genome assembly and annotation

We constructed *de novo* haploid genome assemblies for both species (Table S1) using a combination of long-read (PacBio), short read (Illumina) and chromosome conformation capture (Hi-C) data. This resulted in assemblies of 838 Mb and 1198 Mb with an N50 of 23.6 and 43.3 Mb in *T. fasciculata* and *T. leiboldiana* respectively. The assembly sizes closely match the estimated genome size of each species based on flow cytometry and k-mer analysis (Table S2, SI Note 1,2). The 25 and respectively 26 longest scaffolds (hereafter referred to as 'main scaffolds') contain 72 % and 75.5 % of the full assembly, after which scaffold sizes steeply decline (SI Note 3, Fig. S1). This number of main scaffolds corresponds with the species karyotype in *T. fasciculata*, but deviates from the *T. leiboldiana* karyotype (SI Note 1), suggesting that a few fragmented chromosome sequences remain in this assembly.

Structural gene annotation resulted in a total of 34,886 and 38,180 gene models in *T. fasciculata* and *T. leiboldiana* respectively, of which 92.6 % and 71.9 % are considered robust based on additional curation (Methods, Section 5). Annotation completeness was evaluated with BUSCO using the liliopsida dataset resulting in a score of 89.7 % complete genes in *T. fasciculata* and 85.3 % in *T. leiboldiana* (Table S2).

3.2. Genic, repetitive and GC content

TE annotation performed with EDTA[25] revealed a total repetitive content of 65.5 % and 77.1 % in *T. fasciculata* and *T. leiboldiana* respectively. This closely matches estimates derived

123 from k-mer analyses (66 % and 75 %, SI Note 2). Compared to *T. fasciculata*, the repetitive content
124 in *T. leiboldiana* is enriched for Gypsy LTR retrotransposon and *Mutator* DNA transposon content,
125 with a 1.7-fold and 4.2-fold increase in genomic length, respectively (Table S3). *Mutator* DNA
126 transposons have high transposition rates and often insert in or near genes[26].

127 Repetitive content per scaffold is negatively correlated with gene count in both assemblies
128 (Kendall's correlation coefficient: -0.79 in *T. fasciculata*, -0.82 in *T. leiboldiana,* p-values < 2.2e-
129 [16]), with gene-rich regions primarily in distal positions (Fig. 2a, green track) and repetitive regions
130 in median positions (Fig. 2a, yellow track). This pattern is accentuated in *T. leiboldiana*: on
131 average, the repetitive-to-exonic content per scaffold is 1.6 times larger compared to *T. fasciculata*
132 (Mann Whitney U, p-values = 4.3×10^{-4}). The genome size difference between the two assemblies
133 is therefore mostly explained by differential accumulation of TE content in heterochromatic
134 regions.

135 Surprisingly, GC content is negatively correlated with gene content in both species
136 (Kendall's correlation coefficient: -0.68 in *T. fasciculata*, -0.71 in *T. leiboldiana,* p-values < 2.2e-
137 [16,] Fig. 2a, red track, Fig. 2b). By visualizing GC and TE content across a syntenic chromosome
138 triplet of *A. comosus*, *T. fasciculata* and *T. leiboldiana,* we show that this unusual relationship can
139 be mostly explained by elevated GC content in repetitive regions (Fig. 2b). TE-rich regions indeed
140 exhibit a much higher GC content than TE-poor regions, a pattern which is exacerbated as the
141 overall TE content per species increases (Fig. 2b, SI Note 5).
142

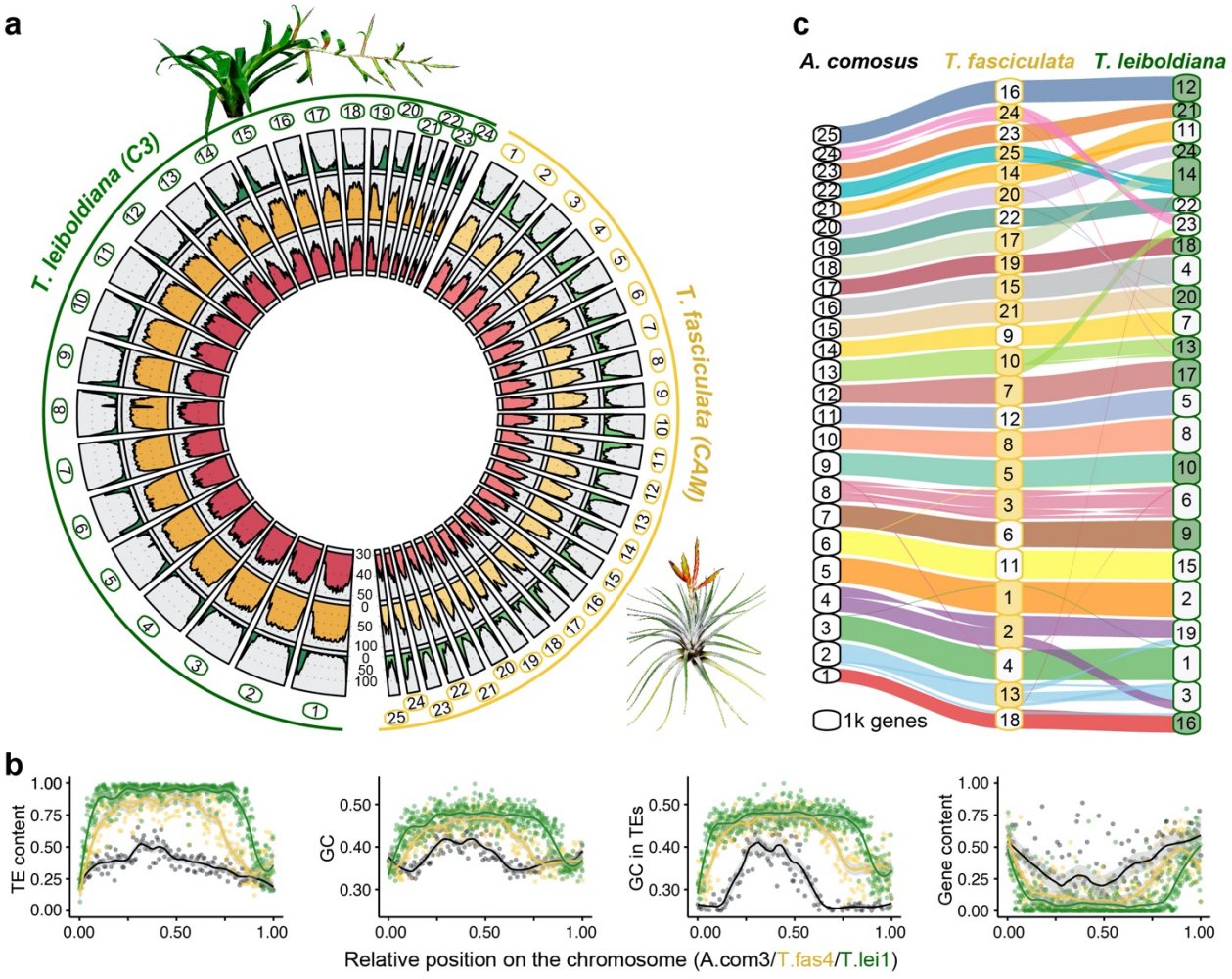


Figure 2: a) *Circular overview of the main scaffolds of the* T. fasciculata *(right) and* T. leiboldiana *(left) genome assemblies. Scaffolds 25 and 26 of* T. leiboldiana *are not shown due to their small size. In inward direction, the tracks show: (1, green) Gene count, (2, yellow) proportion of repetitive content, (3, red), and GC content per 1-Mb windows.* ***b)*** *Distribution of TE and GC content, GC content exclusively in TEs, and genic content in a triplet of syntenic scaffolds between* Ananas comosus *(LG3, black),* T. fasciculata *(scaffold 4, grey) and* T. leiboldiana *(scaffold 1, green; see Fig. S2 for other syntenic chromosomes).* ***c)*** *Syntenic plot linking blocks of orthologous genes between* A. comosus*,* T. fasciculata *and* T. leiboldiana*. The size of each scaffold on the y-axis is determined by genic content and therefore doesn't represent the true size of the scaffold. Color-filled boxes indicate scaffolds with reversed coordinates as compared to the sequences in the reference.*

3.2. Synteny and chromosomal evolution

Cytogenetic karyotyping revealed a difference of six chromosome pairs between *T. fasciculata* (2n = 50) and *T. leiboldiana* (2n = 38), an unexpected finding for a clade that was believed to be largely homoploid with constant karyotype[21,27] (See SI Note 1). To investigate orthology and synteny, we inferred orthogroups between protein sequences of *Ananas comosus*[28] (pineapple), *T. fasciculata* and *T. leiboldiana* using Orthofinder[29]. This resulted in 21,045 (78 %), 26,325 (87.5 %) and 23,584 (75 %) gene models assigned to orthogroups respectively, of which 10,021 were single-copy orthologs between all three species (Table S4).

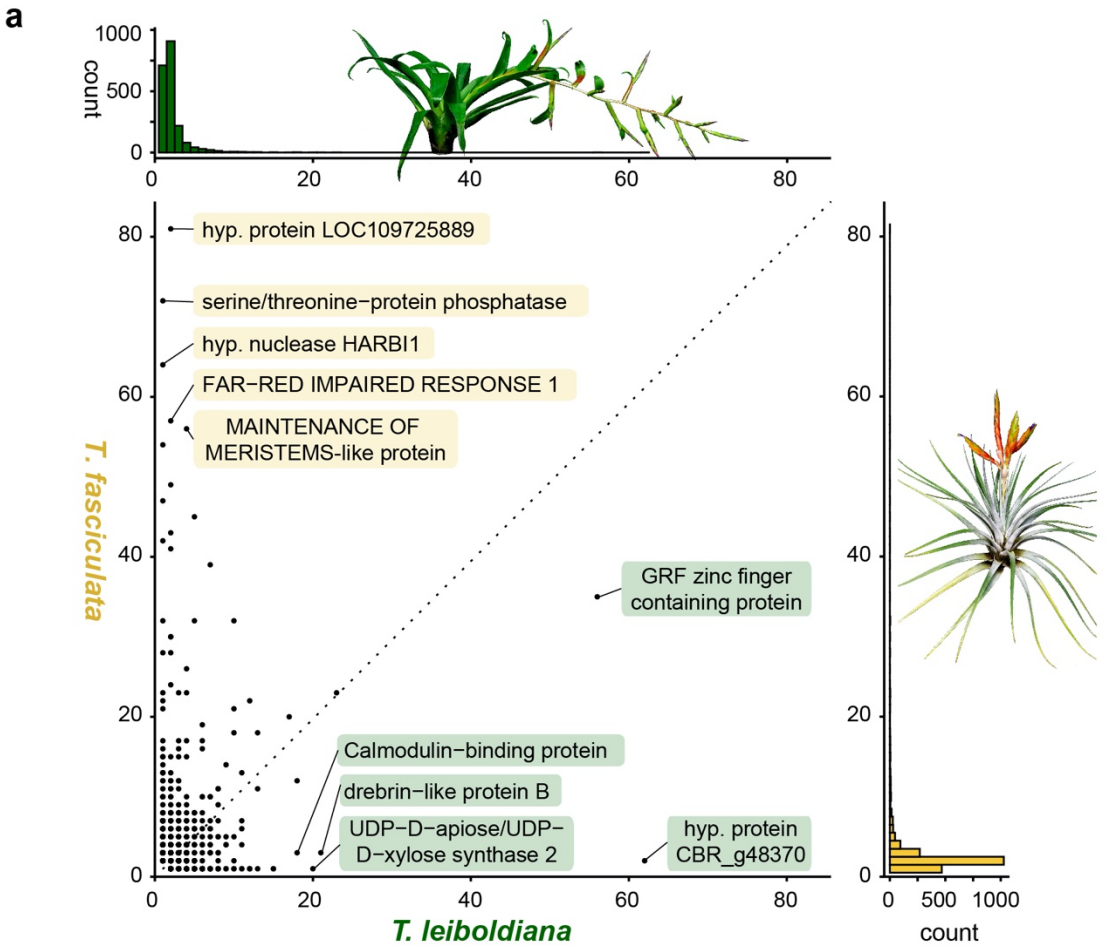
Syntenic blocks were then defined across all three assemblies using Genespace[30] (Fig. 2c). Despite the observed karyotype difference, these blocks reveal highly conserved synteny between the two assemblies, consistent with a scenario of chromosomal fusions in *T. leiboldiana*. We found clear evidence of such a fusion on scaffold 14 in *T. leiboldiana* (Fig. 2c, Fig. S3a), which was confirmed with in-depth analyses of potential breakpoints (SI Note 6). We also detected two major reciprocal translocations (Fig. 2c, hereafter referred as Translocation 1 and 2, Fig. S3b and Fig. S3c).

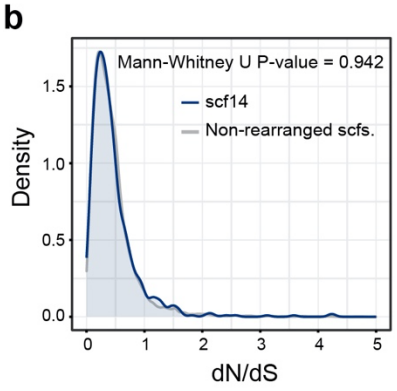
3.3. Gene family evolution

6,261 genes in *T. fasciculata* and 4,693 genes in *T. leiboldiana* were assigned to non-unique gene families with multiple gene copies in at least one species, after correcting gene family sizes (Table S4, SI Note 7). On average, the multicopy gene family size is 1.3x larger in *T. fasciculata* than in *T. leiboldiana* (Mann Whitney U, p-value: $8.8e^{-16}$, Fig. 3a).

To investigate the role of expanded gene families in CAM evolution, we performed gene ontology (GO) term enrichment tests on multicopy orthogroups (SI Note 8). This highlighted several multi-copy gene families with functions putatively related to CAM (Table S5), such as two malate dehydrogenase families (MDH) which reversibly convert malate to oxaloacetate, two families of enolases, which catalyse the penultimate step in the glycolysis resulting in PEP, and two subunits of succinate dehydrogenase, a protein complex that is both part of the tricarboxylic acid cycle and the electron transport chain (see Fig. 5). Interestingly, succinate dehydrogenase also plays a role in stomatal regulation, which is relevant for diel cycling in CAM[31]. Another salient gene family in this list is XAP5 CIRCADIAN TIMEKEEPER (XCT), a regulator of circadian rhythm and disease resistance[32] which was previously identified as undergoing rapid gene family evolution in *Tillandsia*[32].

For many of these gene families, we also observe circadian- and species-related differential gene expression (see Results, section 3.5). In these CAM-analogous gene families, evidence for gene copy number increases in both *T. fasciculata* and *T. leiboldiana* suggest that loss or gain of gene copies could be contributing to the shift of C3/CAM metabolisms.





c

Orthogroup	dN/dS	adj-P	Function
OG0002972	∞	0.00162	jacalin-related lectin 3-like
OG0005000	∞	0.00628	cucumber peeling cupredoxin-like
OG0006253	∞	0.00066	mitochondrial prohibitin-3
OG0009278	∞	0.00155	chloroplastic Peroxiredoxin-2E-2
OG0012770	∞	0.01374	metallo-hydrolase/oxidoreductase superfamily protein
OG0011786	9.3	0.00008	U-box_domain-containing_protein
OG0015603	8.1	0.01575	anaphase-promoting complex subunit CDC27
OG0009004	3.1	0.01148	Hydroquinone glycosyltransferase
OG0008977	2.6	0.00512	glutamate receptor 2.8-like
OG0010014	2.0	0.00603	Glycerophosphodiester phosphodiesterase GDPDL7

Figure 3: a) *Scatterplot: composition of per-species gene counts among orthogroups. Labels indicate the functions of the top 5 largest orthogroups in each species. Upper histogram: distribution of per-orthogroup gene count in* T. fasciculata*. Lower histogram: distribution of per-orthogroup gene count in* T. leiboldiana*.* ***b)*** *Distribution of d_N/d_S values of one-to-one orthologs across non-rearranged scaffolds (grey profile) and scaffold 14 in* T. leiboldiana *(blue profile), which is the result of a fusion.* ***c)*** *Top 10 single-copy orthogroups with highest significant d_N/d_S values and their functions. Infinite d_N/d_S values characterize genes which accumulated no synonymous substitutions due to the young divergence between the two species. Further explanation about the biological significance of these functions can be found in SI Note 9.*

3.4. Adaptive sequence evolution

Adaptive sequence evolution was evaluated in 9,077 orthologous gene pairs using the non-synonymous to synonymous substitution ratio ($\omega = d_N/d_S$). Little among-scaffold variation in d_N/d_S was observed, with per-scaffold median d_N/d_S values ranging from 0.32 to 0.39 in *T. fasciculata* and 0.31 to 0.4 in *T. leiboldiana* (Fig. S4a). Regions of large chromosomal rearrangement such as the fused scaffold 14 in *T. leiboldiana* do not exhibit strong signatures of fast coding sequence evolution (Fig. 3b), though for Translocation 1, d_N/d_S values are slightly yet significantly lower for scaffold 13 in *T. fasciculata* and scaffold 19 in *T. leiboldiana* (Fig. S4b, SI Note 6).

Among the 9,077 orthologous gene pairs, 13 candidates (0.21%) exhibit a significant $\omega > 1$ (Fig. 3c, Table S6, SI Note 9). Notably, we recover a significant signal in a glycerophosphodiester phosphodiesterase (GDPDL-7). GDPDL's are involved in cell wall cellulose accumulation and pectin linking, and play a role in trichome development[33], a main trait differentiating the two species and more broadly, green and grey *Tillandsia*.

A glutamate receptor (GLR) 2.8-like also exhibits a significant $\omega > 1$. By mediating Ca2+ fluxes, GLRs act as signalling proteins and mediate a number of physiological and developmental processes in plants[34], including stomatal movement[35]. Although it is associated with drought-stress response in *Medicago trunculata*[36], the specific function of GLR2.8 still remains unknown.

3.5. Co-expression analyses

To study gene expression differences linked to CAM/C3 shifts, we performed a time-series RNA-seq experiment using six plants of each species (Table S1), sampled every four hours in a circadian cycle. We recovered 907 genes with a differential expression (DE) profile between *T.*

fasciculata and *T. leiboldiana* across time points. GO term enrichment revealed many CAM-related functions such as malate and oxaloacetate transport, circadian rhythm, light response, water and proton pumps, sucrose and maltose transport and starch metabolism (Table S7). Nine of 22 genes reported by De La Harpe et al. 2020 as candidates for adaptive sequence evolution during C3 / CAM transitions in *Tillandsia* were recovered in this subset (Table S7). Core CAM enzymes phosphoenolpyruvate carboxylase (PEPC) and phosphoenolpyruvate carboxylase kinase (PEPC kinase) display clear temporal expression cycling in *T. fasciculata* (Fig. 4c, S5). However, PEPC kinase also shows an increase in expression in *T. leiboldiana* (Fig. S5), a phenomenon that has been documented before in C3 *Tillandsia*[36] and also in other plant systems with a recent shift to CAM[37].

Clustering analysis distributed DE genes across seven clusters with sizes ranging from 209 to 38 genes (Table S7). CAM-related genes were distributed across six of seven clusters, highlighting the diversity of expression profiles associated with CAM (Fig. S6). While core CAM genes (see Fig. 5) are mainly found in cluster 5, we find malate transporters in cluster 1, circadian regulators in clusters 2 and 3, sugar transport in clusters 3 and 6, and vacuolar transport regulators in clusters 2, 4 and 6. Cluster 7, though not containing any core CAM candidate genes, was enriched for salt and heat stress response and contains a mitochondrial isocitrate dehydrogenase, which is predicted to increase in activity in CAM plants[38].

The expression curves of the respective clusters (Fig. S6), demonstrate a complex web of expression changes between CAM and C3. Generally, we find the following patterns: (i) an overall increase in expression across all timepoints in the CAM species compared to the C3 species (stomatal regulation, starch metabolism, drought stress response - clusters 2,5 and 6), (ii) an overall increase in expression in the C3 species (malate transmembrane transport, aquaporins, vacuolar

transport regulators - clusters 4 and 7), (iii) switches from a linear time signal to a circadian pattern (PEPC), and (iv) amplifications of circadian patterns in one species compared to the other (PEPC Kinase, cluster 3).

3.5.1. Differentially expressed genes have more TE insertions

To investigate whether TE activity and differential gene expression are associated in *Tillandsia*, we tested whether intronic TE insertions are significantly enriched in DE genes in both species. Genic TE insertions are generally more common in *T. leiboldiana* than in *T. fasciculata*, further highlighting the increased TE dynamics in this species observed across the entire genome (See Results section 3.2.). While DE genes in both genomes contain a similar proportion of genes with one or more transposable element insertions in intronic regions compared to the full gene set, the average number of TE insertions per gene is significantly higher in DE than in non-DE genes (Table 1). An elevated intronic TE insertion rate in DE genes, both in the CAM and C3 species, points at the potential role of TE dynamics providing expression changes that lead to more CAM- or C3-like circadian profiles.

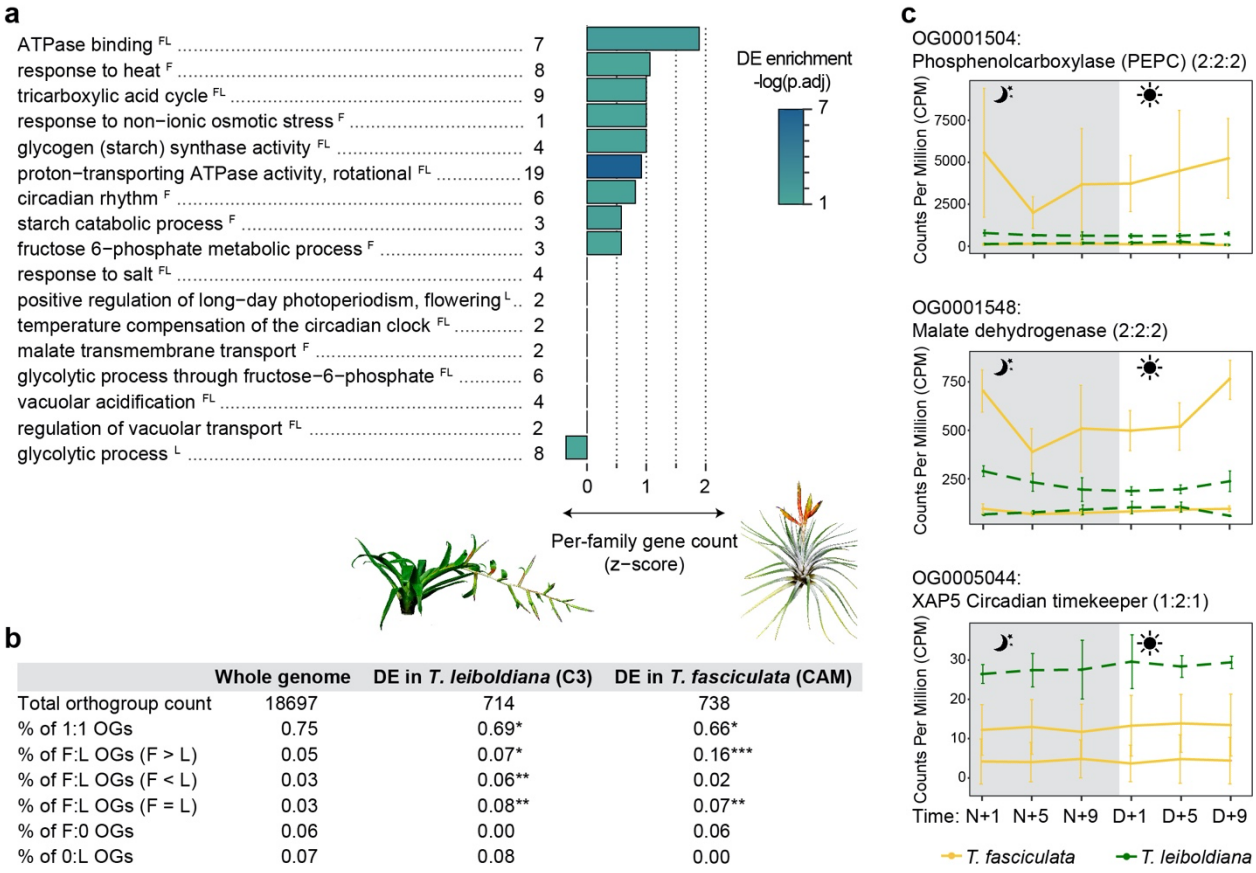
Table 1: Statistical test results on TE insertions in DE versus non-DE genes

Presence of TE insertions			
	DE genes with a TE insertion	Total number of genes with a TE insertion	Chi-square p-value
In *T. fasciculata*	473 (52 %)	15844 (50 %)	0.2324
In *T. leiboldiana*	472 (54.6 %)	18251 (54.6 %)	0.9645
Average TE insertion counts per gene			
	in DE genes	in non-DE genes	Mann-Whitney U p-value
In *T. fasciculata*	3.66	2.88	0.0179
In *T. leiboldiana*	4.2	3.27	0.0159

3.5.2. Differentially expressed genes belong more often to multi-copy orthogroups

To investigate the consequences of gene family evolution on gene expression, we tested whether the proportion of multi-copy orthogroups underlying DE genes was significantly elevated to that of the whole-genome set of orthogroups in both species (Fig. 4b, SI Note 10). The 907 DE genes in *T. fasciculata* are found in 738 orthogroups containing a total of 2,141 and 910 genes in *T. fasciculata* and *T. leiboldian*a, respectively. Genes from multi-copy orthogroups are more likely to be differentially expressed: while multi-copy orthogroups account for 24 % of all orthogroups in the genome, they represent 31 % of DE genes. This difference is primarily explained by a 3.2x larger proportion of multi-copy orthogroups with a larger family size in *T. fasciculata* than in *T. leiboldiana* in the DE subset compared to the whole genome (Chi-square $P = 1.59e^{-66}$).

Reciprocally, the DE analysis on the *T. leiboldiana* genome (See SI Note 10) resulted in 836 DE genes belonging to 714 orthogroups, of which 489 overlap with the DE orthogroups resulting from the analysis on the *T. fasciculata* genome. As in the analysis on the *T. fasciculata* genome, we find that orthogroups with a larger family size in *T. leiboldiana* are enriched among differentially expressed genes. This suggests that gene families undergoing expansion or contraction play a role in time-specific gene expression, in part related to photosynthetic metabolism, in both species. Additionally, both analyses point at a significant enrichment for multi-copy orthogroups with equal family sizes in both species, suggesting that also older duplications preceding the split of *T. fasciculata* and *T. leiboldiana* play a role in day-night regulatory evolution. This highlights the importance not only of novel, but also ancient variation in fuelling trait evolution in *Tillandsia*.



b

	Whole genome	DE in *T. leiboldiana* (C3)	DE in *T. fasciculata* (CAM)
Total orthogroup count	18697	714	738
% of 1:1 OGs	0.75	0.69*	0.66*
% of F:L OGs (F > L)	0.05	0.07*	0.16***
% of F:L OGs (F < L)	0.03	0.06**	0.02
% of F:L OGs (F = L)	0.03	0.08**	0.07**
% of F:0 OGs	0.06	0.00	0.06
% of 0:L OGs	0.07	0.08	0.00

Figure 4: a) *CAM-related enriched GO terms among differentially expressed (DE) genes between* T. fasciculata *and* T. leiboldiana*. The genome in which a GO term has been found to be enriched among DE genes is shown by* ^{F} *and* ^{L} *for T. fascuculata and T. leiboldiana, respectively. The family size difference for the underlying orthogroups is represented as a Z-score*: a *negative score indicates a tendency towards gene families with larger size in* T. leiboldiana *than in* T. fasciculata*, and vice versa. The p-value displayed represents the significance of the GO-term enrichment among DE genes in* T. fasciculata*, unless the term was only enriched in* T. leiboldiana*. The number of DE genes underlying each function is shown next to the GO-term name.* ***b)*** *Proportion of orthogroup types by gene family size relationship between* T. fasciculata (F) *and* T. leiboldiana (L) *across the whole genome and in DE orthogroups in each species. A chi-square test was applied to all categories between the whole genome and each DE subset. Contribution of each category to the total Chi-square score is indicated as follows: *2-10 %, **10-50%, ***50-100%* ***c)*** *Examples of the circadian expression of individual CAM-related gene families (PEPC, MDH and XCT) displayed at the orthogroup level. We show two families with older duplications preceding the split of* T. fasciculata *and* T. leiboldiana *(PEPC and MDH) and one gene family with a recent duplication in* T. fasciculata *(XCT).*

Certain CAM-related biological functions appear associated with gene family expansion or contraction (Fig. 4a). Most notable is the unequal number of functions tending to larger gene

311 families in *T. fasciculata* than in *T. leiboldiana* (nine versus one). Functions associated with V-
312 ATPase proton pumps especially tend to have larger gene family size in *T. fasciculata* than in *T.*
313 *leiboldiana* (ATPase binding, proton-transporting ATPase activity).

314 Examples of CAM-related, differentially expressed genes that belong to multi-copy
315 orthogroups are (i) XAP5 CIRCADIAN TIMEKEEPER (XCT), which has an extra copy in *T.*
316 *fasciculata*, (ii) a family of malate dehydrogenase (MDH) with two copies in both species, and
317 (iii) core CAM enzyme Phosphoenolpyruvate carboxylase (PEPC), which shares an ancient
318 duplication among monocots[39] (Fig. 4c). A candidate gene family with larger gene family size in
319 *T. leiboldiana* is a probable aquaporin PIP2-6 (OG0005047, Fig. S7), which is involved in water
320 regulation and follows a circadian pattern in pineapple (A. comosus)[40].

321

322 **4. Discussion**

323

324 The sources of variation fuelling trait evolution in rapid radiations have been a long-
325 standing topic in evolutionary biology[41]. By integrating comparative genomics using *de novo*
326 assembly and in-depth gene expression analyses of two *Tillandsia* species representing a CAM/C3
327 shift, we found support for TE dynamics, gene family expansion and adaptive sequence evolution
328 as drivers of trait evolution (Fig. 5).

329 Differences between the two genomes related to a CAM/C3 shift can be primarily found
330 on the regulatory level, with DE genes between species across a circadian cycle significantly
331 enriched for many CAM-related functions. These reveal a complex web of underlying expression
332 changes (Fig. S6). Genes underlying the same function can show multiple types of expression
333 changes: salt stress response genes both increase and decrease in expression in the CAM species,

334 while circadian rhythm regulators show all possible pattern changes and are widely distributed
335 across co-expression clusters. These findings emphasize the complexity of CAM regulation,
336 lacking both a master regulator and a clear direction of expression changes[42,43].

337 Though we observe a karyotype difference of 6 chromosome pairs between *T. fasciculata*
338 and *T. leiboldiana* and identified one fusion in the *T. leiboldiana* assembly, along with two
339 reciprocal translocations, we did not find detectable consequences of large-scale rearrangements
340 for either functional diversification or adaptation in *Tillandsia* unlike other studies[5,43] (Fig. 3b, S4,
341 S6, SI Note 10, but see SI Note 6 and 11).

342 We did identify relevant associations, however, between shorter structural variants and trait
343 evolution in *Tillandsia*. While TE insertions in genic regions are overall more common in *T.*
344 *leiboldiana*, both species showed significantly elevated transposition rates in differentially
345 expressed genes, suggesting a potential role of TEs in modifying gene regulation towards the
346 evolution of a CAM/C3 shift.

347 Gene family expansion has been previously witnessed in CAM lineages[5,44] and suggested
348 as a driver of CAM evolution[45]. Strikingly, the subset of differentially expressed genes was
349 significantly enriched for multicopy gene families. Several enriched CAM-related functions show
350 a bias towards expanded gene families in *T. fasciculata* (circadian rhythm, vacuolar ATPase
351 activity, tricarboxylic acid cycle and starch metabolism), and glycolysis showed a bias towards
352 expansion in *T. leiboldiana* (Fig. 4a). More ancient duplications, preceding the split of *T.*
353 *fasciculata* and *T. leiboldiana* are also significantly associated with day-night expression changes
354 (Fig. 4b).

355

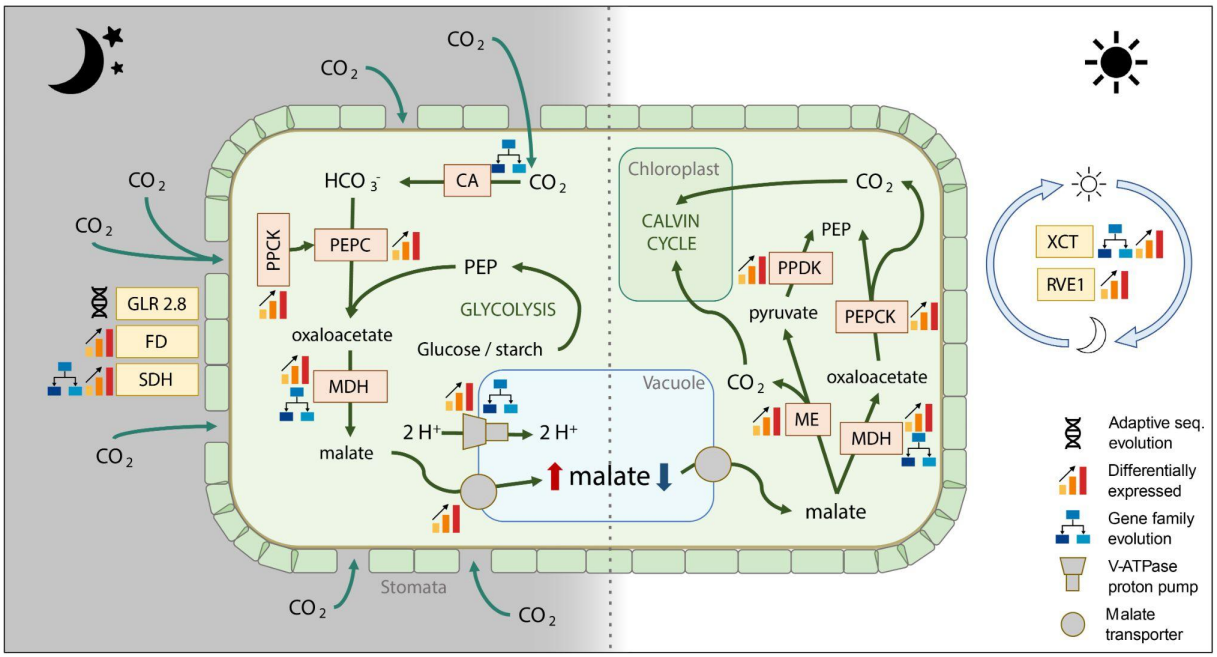


Figure 5: *Pathway of Crassulacean Acid Metabolism (CAM), highlighting underlying genes detected in this study as differentially expressed, with gene family expansion and/or with signature of adaptive sequence evolution. CO_2 is absorbed at night and first converted to HCO_3^- by carbonic anhydrase (CA). Then, it is converted to malate by carboxylating phosphoenol pyruvate (PEP), a key component of the glycolysis. In a first step, PEP carboxylase (PEPC) converts PEP to oxaloacetate, after being activated by PEPC kinase. In a second step, malate dehydrogenase (MDH) converts oxaloacetate to malate. Malate is then transported into the vacuole by two possible transporters, either a tonoplast dicarboxylate transporter or an aluminum-activated malate transporter, which are assisted by V-ATPase proton pumps. During the day, the accumulated malate becomes the main source of CO_2 for photosynthesis. This allows the stomata to remain closed, which greatly enhances the water use efficiency (WUE) of the plant. Malate is again transported out of the vacuole and reconverted to PEP by two possible mechanisms. One possibility is that malate is converted to oxaloacetate by MDH, and then decarboxylated to PEP and CO_2 by PEP carboxykinase (PEPCK). Another option is that malate is decarboxylated into pyruvate and CO_2 by malic enzyme (ME). Pyruvate is then reconverted into PEP by pyruvate orthophosphate dikinase (PPDK). The CO_2 will cycle through the Calvin cycle and generate sugars. Though it is known that in* A. comosus *the main decarboxylase is PEPCK, we recovered circadian differential gene expression in both decarboxylases, and therefore display both possible pathways. While enzymes that are part of the CAM core pathway are highlighted in orange boxes, regulators of stomatal movement and circadian clock are highlighted in yellow boxes.*

The expression curves of DE multicopy gene families with a potential link to CAM reveal a multitude of expression behaviours (Fig. 4c), which supports that complex regulatory evolution on the transcriptional level underlie CAM evolution. Our findings suggest that gene family

evolution played a significant role in modulating regulatory changes underlying a C3 to CAM shift in *Tillandsia*. As gene family expansion leads to increased redundancy, selection on the individual gene copies and their expression relaxes, facilitating changes in expression leading to the evolution of a more CAM or C3-like expression profile[46].

Candidate genes under positive selection underlie a broad array of functions and had no immediate link to CAM photosynthesis, except for glutamate receptor 2.8-like (OG0008977), a potential regulator of stomatal movement[35]. The lack of overlap between regulatory and adaptive sequence evolution is in line with previously proposed mechanisms of CAM evolution largely relying on regulatory changes[39] (but see SI Note 10).

The two *de novo* assemblies presented in this study are the first tillandsioid and third bromeliad genomes published so far. To our knowledge, these are also the first CAM/C3 species pair assembled so far at such short evolutionary timescales. Despite both genomes exhibiting one of the highest TE contents reported to date for a non-polyploid plant species[47], the joint use of long-read sequencing and chromatin conformation capture successfully led to highly contiguous assemblies with high-quality gene sets (SI Note 4). Along with other recently developed resources for Bromeliaceae[48,49], these genomes will be crucial in future investigations of this highly diverse and species-rich plant family, and in further studies of CAM evolution.

Our analyses reveal genomic changes of all scales between two members of an adaptive radiation representing a recent CAM/C3 shift (but see SI Note 11). Large scale rearrangements observed so far seem unlinked from functional divergence, more likely affecting reproductive isolation[50,51]. We however find a clear link between one of the fundamental key innovation traits of this radiation, CAM, and regulatory changes, which have potentially been driven by smaller-scale structural variants[52] such as gene family expansion and transposable element activity (Fig.

403 5). Our findings support a crucial role of small-scale genome evolution in shaping novel variation
404 that fuels trait evolution in adaptive radiation.
405
406 ## 5. Online Methods
407
408 ### 5.1. Flow cytometry and cytogenetic experiments
409
410 #### 5.1.1. Genome size measurements

411 Approximately 25 mg of fresh leaf material was co-chopped according to the chopping method of
412 Galbraith et al. (1983)[53] together with an appropriate reference standard (*Solanum*
413 *pseudocapsicum*, 1.295 pg/1C)[54,55] in Otto´s I buffer[56]. After filtration through a 30 µm nylon mesh
414 (Saatilene Hitech, Sericol GmbH, Germany) and incubation with RNase A (0.15mg/ml, Sigma-
415 Aldrich, USA) at 37°C, Otto´s II buffer[56] including propidium iodide (PI, 50mg/L, AppliChem,
416 Germany) was added. Staining took place in the refrigerator for at least one hour or up to over-
417 night. Measurement was conducted on a CyFlow ML or a CyFlow Space flow cytometer
418 (Partec/Sysmex, Germany) both equipped with a green laser (532nm, 100mW, Cobolt AB,
419 Sweden). The fluorescence intensity (FI) of 10,000 particles were measured per preparation and
420 the 1C-value calculation for each sample followed the equation: $1C_{Obj}$ = (FI peak $mean_{G1\ Obj}$ / FI
421 peak $mean_{G1\ Std}$) × $1C_{Std}$
422
423 #### 5.1.2. Karyotyping

424 Actively growing root meristems of genome assembly accessions (see Table S1) were harvested
425 and pretreated with 8-hydroxyquinoline for 2 hrs at room temperature and 2 hrs at 4°C. The roots

were then fixed in Carnoy's fixative (3 : 1 ethanol : glacial acetic acid) for 24 hours at room temperature and stored –20 °C until use. Chromosome preparations were made after enzymatic digestion of fixed root meristems as described in Jang and Weiss-Schneeweiss (2015)[57]. Chromosomes and nuclei were stained with 2 ng/µl DAPI (4',6-diamidino-2-2phenylindole) in Vectashield antifade medium (Vector Laboratories, Burlingame, CA, USA). Preparations were analyzed with an Axiolmager M2 epifluorescent microscope (Carl Zeiss) and images were captured with a CCD camera using AxioVision 4.8 software (Carl Zeiss). Chromosome number was established based on analyses of several preparations and at least five intact chromosome spreads. Selected images were contrasted using Corel PhotoPaint X8 with only those functions that applied equally to all pixels of the image and were then used to prepare karyotypes.

5.2. Genome Assembly

5.2.1. Plant material selection and sequencing

Genome assemblies were constructed from the plant material of one accession per species (see Table S1). The accessions were placed in a dark room for a week to minimize chloroplast activity and recruitment, after which the youngest leaves were collected and flash frozen with liquid nitrogen. High molecular weight extraction for ultra-long reads, SMRTbell library preparation and PacBio Sequel sequencing was performed by Dovetail Genomics™ (now Cantata Bio). Dovetail Genomics™ also prepared Chicago[58] and Hi-C[59] libraries which were sequenced as paired-end 150bp reads on an Illumina HiSeq X instrument. Additional DNA libraries were

448 prepared for polishing purposes using Illumina's TruSeq PCR-free kit, which were sequenced on
449 a HiSeq2500 as paired-end 125 bp reads at the Vienna BioCenter Core Facilities (VBCF), Austria.
450 RNA-seq data of *T. fasciculata* used for gene annotation was sampled, sequenced and
451 analyzed in De La Harpe et al. 2020 under SRA BioProject PRJNA649109. For gene annotation
452 of *T. leiboldiana*, we made use of RNA-seq data obtained during a similar experiment, where
453 plants were kept under greenhouse conditions and sampled every 12 hours in a 24-hour cycle.
454 Importantly, while the *T. fasciculata* RNA-seq dataset contained three different genotypes, only
455 clonal accessions were used in the *T. leiboldiana* experiment. For *T. leiboldiana*, total RNA was
456 extracted using a QIAGEN RNeasy® Mini Kit, and poly-A capture was performed at the Vienna
457 Biocenter Core Facilities (VBCF) using a NEBNext kit to produce a stranded mRNA library. This
458 library was sequenced on a NovaSeq SP as 150 bp paired end reads.
459 For both species, sequencing data from different time points and accessions were merged
460 into one file for the purpose of gene annotation. Before mapping, the data was quality-trimmed
461 using AdapterRemoval[60] with default options (--trimns, --trimqualities). We allowed for
462 overlapping pairs to be collapsed into longer reads.
463
464 **5.2.2. First draft assembly and polishing**
465
466 We constructed a draft assembly using long-read PacBio data with CANU v1.8[61] for both
467 species. To mitigate the effects of a relatively low average PacBio coverage (33x), we ran two
468 rounds of read error correction with high sensitivity settings (corMhapSensitivity=high
469 corMinCoverage=0 corOutCoverage=200) for *T. fasciculata.* Additionally, we applied high
470 heterozygosity (correctedErrorRate=0.105) settings, since K-mer based analyses pointed at an

471 elevated heterozygosity in this species (See SI Note 2, Figures S10-11), and memory optimization
472 settings (corMhapFilterThreshold=0.0000000002 corMhapOptions=" --repeat-idf-scale 50"
473 mhapMemory=60g mhapBlockSize=500).

474 Given that the coverage of *T. leiboldiana* PacBio averaged 40x, we limited error correction
475 for this species to only one round. CANU was run with additional settings accommodating for
476 high frequency repeats (ovlMerThreshold=500) and high sensitivity settings as mentioned above.

477 To minimize the retainment of heterozygous sequences as haplotigs in *T. fasciculata* (see
478 SI Note 2), we reassigned allelic contigs using the pipeline Purge Haplotigs[62]. Raw PacBio data
479 was mapped to the draft assembly produced in the previous step with minimap2[63], before using
480 the Purge Haplotigs pipeline.

481 Since the size of the *T. leiboldiana* draft assembly indicates, together with previous
482 analyses, that this species is largely homozygous (SI Note 2, Figures S10-11), we did not include
483 a purge_haplotigs step. However, we did make use of the higher average coverage of the *T.*
484 *leiboldiana* PacBio data to polish the assembly with two rounds of PBMM v.1.0 and Arrow v2.3.3
485 (Pacific Biosciences).

486

487 **5.2.3. Scaffolding and final polishing**

488

489 Scaffolding of both assemblies was performed in-house by Dovetail Genomics™ using
490 Chicago and Hi-C data and the HiRise scaffolding pipeline[58]. To increase base quality and correct
491 indel errors, we ran additional rounds of polishing with high-coverage Illumina data (See above,
492 section 2.1.) using Pilon v1.22[64]. The Illumina data was aligned to the scaffolded assembly using
493 BWA-MEM[65], and then Pilon was run on these alignments. We evaluated the result of each round

494 using BUSCO v.3[66] with the liliopsida library and proceeded with the best version. For *T.*
495 *fasciculata*, polishing was performed twice, fixing SNPs and indels. We did not fix small structural
496 variation in this genome due to the relatively low coverage (35x) of Illumina data. For *T.*
497 *leiboldiana*, one round of polishing on all fixes (SNPs, indels and small structural variants) resulted
498 in the highest BUSCO scores.
499
500 **5.3. Annotation**
501
502 **5.3.1. TE annotation and repeat masking**
503
504 *De novo* TE annotation of both genome assemblies was performed with EDTA v.1.8.5[25]
505 with option –sensitive. To filter out genes that have been wrongly assigned as TEs, *A. comosus*
506 (pineapple) coding sequences[28] were used in the final steps of EDTA.
507 Using the species-specific TE library obtained from EDTA, we masked both genomes
508 using RepeatMasker v.4.0.7[69]. Importantly, we excluded all TE annotations marked as "unknown"
509 for masking to prevent potentially genic regions flagged as TEs to be masked during annotation.
510 The search engine was set to NCBI (-e ncbi) and simple and low-complexity repeats were left
511 unmasked (-nolow). We produced both hard-masked and soft-masked (--xsmall) genomes.
512
513 **5.3.2. Transcriptome assembly**
514
515 We constructed transcriptome assemblies for both species using Trinity *de novo* assembler
516 v.2.4.8.[70] using default parameters starting from the raw mRNA-seq data. These were evaluated
517 with BUSCO. Additionally, before feeding the transcriptome assemblies to the gene annotation

pipeline, we ran a round of masking of interspersed repeats to avoid an overestimation of gene models due to the presence of active transposases in the RNA-seq data.

5.3.3. Gene prediction and functional annotation

Gene models were constructed using a combination of BRAKER[71] v.2.1.5 and MAKER2[72] v.2.31.11. Starting with BRAKER, we obtained splicing information from RNA-seq alignments to the masked genome as extrinsic evidence using the *bam2hints* script of AUGUSTUS v.3.3.3[73]. A second source of extrinsic evidence for BRAKER were single-copy protein sequences predicted by BUSCO when run on the masked genomes in genome mode with option --long. Predictions made by BRAKER were evaluated with BUSCO and with RNA-seq alignments.

Subsequently, we built our final gene predictions using MAKER2. As evidence, we used (1) the gene models predicted by BRAKER, (2) a transcriptome assembly of each respective species (see above section 3.2.), (3) a protein sequence database containing proteins of *Ananas comosus comosus* (F135)[28] and *Ananas comosus bracteatus* (CB5)[75] and manually curated swissprot proteins from monocot species (64,748 sequences in total) and (4) a GFF file of complex repeats obtained from the masked genome (see above section 3.1.) and an extended repeat library containing both the EDTA-produced *Tillandsia*-specific repeats and the monocot repeat library from RepBase (7,857 sequences in total). By only providing masking information of complex repeats and setting the model organism to “simple” in the repeat masking options, hardmaksing in MAKER2 was limited to complex repeats while simple repeats were soft-masked, which makes these available for gene prediction. MAKER2 predicts genes both *ab initio* and based on the given evidence using AUGUSTUS.

We evaluated the resulting set of predicted gene models by mapping the RNA-seq data (section 2.1.) back to both the transcript and full gene model sequences and running BUSCO in transcriptome mode. We also calculated the proportion of masked content in these gene models to ascertain that MAKER2 hadn't predicted TEs as genes. A second run of MAKER, which included training AUGUSTUS based on the predicted models from the first round, resulted in lower BUSCO scores and was not further used.

We functionally annotated the final set of gene models in Blast2Go v.5.2.5[76] using the Viridiplantae database.

5.4. Inferring gene orthology

Orthology between gene models of *T. fasciculata*, *T. leiboldiana* and *Ananas comosus* was inferred using Orthofinder v.2.4.0[29]. Protein sequences produced by MAKER2 of inferred gene models were used for *T. fasciculata* and *T. leiboldiana*. For *A. comosus*, the publicly available gene models of F153 were used (SRA BioProject PRJNA371634). The full Orthofinder pipeline was run without additional settings. Counts per orthogroup and the individual genes belonging to each orthogroup were extracted from the output file Phylogenetic_Hierarchical_Orthogroups/N0.tsv.

Orthofinder was run a second time on gene models present only on main contigs (See Results). For each gene model, the longest isoform was selected, and gene models with protein sequences shorter than 40 amino acids were removed. This resulted in 27,024, 30,091 and 31,194 input sequences for *A.* comosus, *T. fasciculata* and *T. leiboldiana* respectively. Then, the steps mentioned above were repeated.

564

565 **5.5. Gene model assessment and curation**

566

567 Gene model sets were assessed and curated using several criteria. Gene models with
568 annotations indicating a repetitive nature (transposons and viral sequences) together with all their
569 orthologs were marked with "NO_ORTHOLOGY" in the GFF file and excluded from downstream
570 analyses. Using the per-exon expression data obtained in our mRNA-seq experiment (see below,
571 section 10.) and information gathered on the length of the CDS and the presence / absence of a
572 start and stop codon, we further classified our gene models into ROBUST and NOT-ROBUST
573 categories. A gene model was considered ROBUST (i) if all exons are expressed or, (ii) if both
574 start and stop codons are present and the CDS has a minimum length of 50 amino-acids.

575

576 **5.6. Analyzing TE class abundances**

577

578 By rerunning EDTA with step --anno, we obtained TE abundances and detailed annotation
579 of repetitive content for the whole assembly. Per-contig abundances of each class were calculated
580 with a custom python script (available at https://github.com/cgrootcrego/). Using this curated TE
581 library, the assemblies were masked again with RepeatMasker for downstream analyses. The
582 resulting TE class abundances reported by RepeatMasker were then compared between species
583 and reported.

584

585 **5.7. Spatial distribution of repetitive, genic and GC content**

586

The spatial distribution of genes, transposable elements and GC content as shown in Fig. 2a, was analysed on a per-window basis, using windows of 1 Mb. Gene counts were quantified as the number of genes starting in every window, based on genes with assigned orthology, including both single and multicopy gene models. Repetitive content was measured as the proportion of masked bases in each window, stemming from the hard-masked assembly using the curated TE library. Per-window gene counts and proportion of repetitive bases was then visualized using the R package circlize[78]. GC content was calculated as the proportion of G and C bases per 1 Mb windows. Correlation between genic, repetitive and GC content was calculated and tested for significance using the Kendall Rank Correlation Coefficient, after testing for normality using the Shapiro-Wilk test.

Repetitive, GC and gene content as shown in Fig. 2b was estimated directly from the soft-masked reference genomes using 100 kb non-overlapping sliding windows as described in Leroy et al. 2021[79]. TE content corresponds to the proportion of soft-masked positions per window, using the curated TE library (see above, section 6.) as basis for soft-masking in RepeatMasker. As compared to the version of Leroy et al. 2021, this script was modified to estimate GC content in repetitive regions only. In addition to this, we estimated the genic fraction by considering the total number of genomic positions falling in genes based on the GFF files (feature = “gene”) divided by the size of the window (100 kb). This estimate was derived for the same window boundaries as used for GC and TE content to be able to compare all statistics. The relative per-window proportion of genic bases corresponding to non-robust genes (see above, section 5) was also estimated by dividing the number of non-robust gene positions with the total number of gene positions.

609

610 **5.8. Synteny between *T. fasciculata* and *T. leiboldiana***

611

612 Synteny was inferred with GENESPACE[30], using orthology information obtained with
613 Orthofinder of the gene models from *A. comosus*, *T. fasciculata* and *T. leiboldiana*. This provided
614 a first, visual graphical to detect large-scale rearrangements. We used GENESPACE with default
615 parameters, except that we generated the syntenic map (riparian plot) using minGenes2plot=200.
616 Other methods have also been used to confirm the chromosomal rearrangements and to identify
617 the genomic breakpoints more precisely (see SI Note 6).

618

619 **5.9. Gene family evolution**

620

621 Gene family counts were corrected for multi-copy orthogroups due to unusual coverage
622 distribution, especially in *T. fasciculata* (see SI Note 7). The distribution of gene counts per
623 multicopy orthogroup was compared between *T. fasciculata* and *T. leiboldiana* with a non-
624 parametric test (Mann-Whitney U). Using the log-ratio of per-species gene count, we investigated
625 which gene families experienced large changes in gene count compared to the background (see SI
626 Note 8).

627 Functional characterization of multicopy families was done with a GO term enrichment
628 analysis of the underlying genes using the Fisher's exact test in TopGo[81]. Enrichment analyses
629 were done on all genes belonging to multicopy orthogroups, on a subset of genes belonging to
630 families that are larger in *T. fasciculata* and on a subset of genes belonging to families that are
631 larger in *T. leiboldiana*. The top 100 significantly enriched GO terms were then evaluated. GO

terms putatively associated with key innovation traits were used to list multicopy gene families of interest.

5.10 d_N/d_S analysis

5.10.1. On single-copy orthologous pairs

One-to-one orthologous genes were subjected to a test of positive selection using the non-synonymous to synonymous substitution ratio ($\omega = d_N/d_S$). Gene pairs where both genes were incomplete (missing start and/or stop codon) or where the difference in total length was more than 20 % of the length of either gene were removed to avoid misalignments. We performed codon-aware alignments using the alignSequences program from MACSE v.2.05[82] with options -local_realign_init 1 -local_realign_dec 1 for optimization. Pairwise d_N/d_S ratios were estimated with the codeML function of PAML v.4.9.[60]. Using a single-ratio model across sites and branches (Nssites = 0, model = 0), we tested for a fixed $\omega = 1$ as null hypothesis, against an unfixed ω as the alternative hypothesis. Automization of codeML was achieved with a modified script from AlignmentProcessor[83]. The results of codeML under both the null and alternative model were compiled and significance of the result was calculated with the likelihood-ratio test[84]. Multiple-testing correction was applied with the Benjamini-Hochberg method and an FDR threshold of 0.05. Orthologous gene pairs with a d_N/d_S ratio larger than one and an adjusted p-value under 0.05 were considered candidate genes under divergent selection.

The d_N/d_S values of all orthologous gene pairs with five or more variant sites in the MACSE alignment were used to obtain per-scaffold distributions of d_N/d_S values in both genomes. We

visualized d_N/d_S distributions of all main scaffolds in both assemblies with boxplots and used density plots to visualize the d_N/d_S distribution in rearranged chromosomes compared to all non-rearranged chromosomes. To test whether these distributions were significantly different, we ran a non-parametric test (Mann-Whitney U) between the distribution of each single rearranged chromosome and that of all non-rearranged chromosomes in each assembly.

5.10.2. On duplicated orthogroups

We also performed tests of selection using d_N/d_S on all orthogroups that were consisted of a single gene in *A. comosus* and a duplicated gene in either T. *leiboldiana* (1:1:2), or *T. fasciculata* (1:2:1). Only orthogroups that maintained this conformation after size correction (SI Note 7) were used in this analysis. Pairwise alignments were performed between the ortholog of one species and either paralog of the other species using MACSE. Then, ω was estimated in the same way as mentioned above.

5.11. RNA-seq experiment capturing CAM and C3 expression differences

5.11.1. Experiment set-up and sampling

To capture gene expression patterns related to CAM, we designed an RNA-seq experiment where individuals of *T. fasciculata* (CAM) and *T. leiboldiana* (C3) were sampled at six time points throughout a 24-hour cycle. Six plants of each species were placed in a PERCIVAL climatic cabinet at 22 °C and a relative humidity (rH) of 68 % for 4 weeks, with a 12-hour light cycle. Light

678 was provided by fluorescent lamps with a spectrum ranging from 400 to 700 nm. The light intensity
679 was set at 124 umol/m2s. The plants acclimatised to these conditions for 4 weeks prior to sampling,
680 during which they were watered every second day.
681 Leaf material from each plant was sampled every 4 hours in a 24-hour cycle starting one
682 hour after lights went off. One leaf was pulled out of the base at each time-point without cutting.
683 The base and tip of the leaf were then removed, and the middle of the leaf immediately placed in
684 liquid nitrogen, then stored at -80 °C.
685
686 **5.11.2. RNA extraction and sequencing**
687
688 Total RNA was extracted using the QIAGEN RNeasy® Mini Kit in an RNAse free
689 laboratory. Samples were digested using the kit's RLT buffer with 1 % Beta-mercaptethanol.
690 Elution was done in two steps. The purity and concentration of the extractions was measured using
691 Nanodrop, and RIN and fragmentation profiles were obtained with a Fragment Analyzer™ system.
692 RNA libraries were prepared by the Vienna Biocenter Core Facilities (VBCF) using a NEBNext
693 stranded mRNA kit before sequencing 150-bp paired-end reads on one lane of Illumina NovaSeq
694 S4.
695
696 **5.11.3. RNA-seq data processing**
697
698 The raw RNA-seq data was evaluated with FastQC[85] and MultiQC[86], then quality trimmed
699 using AdapterRemoval v.2.3.1[76] with settings --trimns --trimqualities --minquality 20 --
700 trimwindows 12 --minlength 36. The trimmed data was then aligned to both the *T. fasciculata* and

T. leiboldiana genomes using STAR v.2.7.9[18] using GFF files to specify exonic regions. Because mapping bias was lowest when mapping to *T. fasciculata*, our main analyses have been performed on the reads mapped to this genome. However, the alignments to *T. leiboldiana* were used for verification or expansion of the main analysis (SI Note 10).

5.11.4. Co-expression analysis

We quantified read counts per exon using FeatureCounts from the Subread package v.2.0.3.[36] for paired-end and reversely stranded reads (-p -s 2). The counts were then summed up across exons per gene to obtain gene-level counts. The composition of the count data was investigated with PCA in EdgeR[37]. Then, counts were normalized using the TMM method in EdgeR, and every gene with a mean cpm < 1 was removed. We ran a differential gene expression (DE) analysis between species and timepoints in maSigPro[38], with *T. leiboldiana* (C3) as the baseline. Significant DE genes were then clustered using the hclust algorithm into modules, with the number of modules being determined with the K-means algorithm. GO term enrichments were performed for each cluster using the R package TopGO[29] and expression curves were plotted by taking the average expression across all replicates per species at each time point with a custom R script. Expression curves for entire clusters (Fig. S6) were plotted by median-centering the log(CPM) of each gene and timepoint against the median of all genes at each time point, while expression curves for individual genes or gene families (Fig. 4c, S5, S7) report average CPM.

5.12. Intersecting findings of gene family evolution, TE insertion ad differential gene expression

5.12.1 Transposable element insertions and differential gene expression

Intronic TE insertions were obtained using *bedtools intersect* on the GFF files of the TE and gene annotations of both species. We used the full transcript length of a gene (feature = "mRNA" in GFF file) for this analyses, and only applied "known" TE annotations. This resulted in a dataset reporting the number of TE insertions per gene. We then performed two tests on the resulting TE counts per gene: (1) whether the proportion of genes with one or more TE insertions is elevated in DE genes compared to the full gene set (chi-square test), and (2) whether the rate of TE insertions per gene measured, as the total count of intersections for each gene annotation with a TE annotation, is elevated in DE genes compared to non-DE genes (Mann-Whitney U test).

5.12.2. Gene family evolution and differential gene expression

Orthogroups were split based on relative family size in *T. fasciculata* (F) versus *T. leiboldiana* (L) in the following categories: Single-copy orthogroups (F = 1 : L = 1), orthogroups with family size larger in *T. fasciculata* (F > L), orthogroups with family size smaller in *T. fasciculata* (F < L), orthogroups with equal family sizes that are larger than 1 (F = L), and orthogroups unique to one species (F:0 or 0:L). We counted the number of orthogroups belonging to each category both for the full orthogroup set and for the subset of orthogroups containing DE genes (DE orthogroups). We then tested whether counts in each orthogroup category were enriched in DE

746 orthogroups compared to all orthogroups using the chi-square test of independence in R. The
747 contribution of each category to the total Chi-square score was calculated as $\frac{r^2}{\chi^2}$, with r the
748 respective residual, and then converted to percentage.
749 To study the effect of the reference genome used on our findings on gene family evolution in
750 DE genes, we performed the same analysis on read counts obtained from mapping to *T.*
751 *leiboldiana* (SI Note 10) and combined these findings in resulting statistics and figures.
752

753 **6. Acknowledgments**

754

755 *In memoriam of Christian Lexer – we will treasure your enthusiasm, guidance and memory*
756 *always*.
757 This work was financially supported by the Austrian Science Fund (FWF) through the doctoral
758 programme (DK) grant W1225-B20 to a faculty team including O.P., FWF grant P35275 to O.P.,
759 and by the professorship start-up grant of Christian Lexer at the University of Vienna BE772002.
760 We thank Joachim Hermisson, Magnus Nordborg, Andrew Clark, Nicholas Barton, Virginie
761 Courtier-Orgogozo, John Parsch, Andreas Futschik, Aglaia Szukala, Florian Schwarz, Marta
762 Pelizzolla and Ahmad Muhammad for insightful discussions, advice,and feedback. We thank Gert
763 Bachmann and Eline de Vos for help with the setup of the RNA-seq experiment; Peter Bak,
764 Andreas Franzke, Nils Köster and Helmut & Lieselotte Hromadnik for their generous donations
765 of *Tillandsia* accessions. We also thank Barbara Knickmann, Viktor Vagovics and Manfred
766 Speckmaier for the care of *Tillandsia* plants at the Botanical Garden of the University of Vienna.
767 Computational resources were provided by the Life Science Computer Cluster (LiSC) of the
768 University of Vienna and the Vienna Scientific Cluster.

7. Author Contributions

This study was conceived by CL, JH, OP, TL and CGC. Sampling was conducted by MHJB, WT, GY and MDH. Laboratory work was conducted by MHJB, SS, TB, LACS and CGC. Cytogenetic work was performed by HWS and EMT. The RNA-Seq experiment and DE analysis was conducted under the guidance of KH and OP. Analyses were performed by CGC, JH, GY, TL and FB. The manuscript was primarily written by CGC and amended following the dedicated reading and feedback of all co-authors, especially KH, TL and OP.

8. Data Availability

The genome assemblies, annotations and raw PacBio and Illumina sequences are available at NCBI-SRA under BioProject PRJNA927306. Hi-C and Chicago data is accessible on request. The list of orthogroups, counts table used for RNA-seq analyses and full GO term enrichment results, along with all scripts written for this manuscript are available on a github repository at https://github.com/cgrootcrego/Tillandsia_Genomes.

9. References

1. Miller, A. H. Some ecologic and morphologic considerations in the evolution of higher taxonomic categories. *Ornithologie als biologische Wissenschaft* **28**, 84–88 (1949).
2. Hunter, J. P. Key innovations and the ecology of macroevolution. *Trends Ecol. Evol.* **13**, 31–36 (1998).

3. Winter, K. & Smith, J. A. C. *Crassulacean Acid Metabolism: Biochemistry, Ecophysiology and Evolution*. (Springer, 2012).

4. Borland, A. M. *et al.* Engineering crassulacean acid metabolism to improve water-use efficiency. *Trends Plant Sci.* **19**, 327–338 (2014).

5. Silvera, K. *et al.* Evolution along the crassulacean acid metabolism continuum. *Funct. Plant Biol.* **37**, 995–1010 (2010).

6. Cicconardi, F. *et al.* Chromosome Fusion Affects Genetic Diversity and Evolutionary Turnover of Functional Loci but Consistently Depends on Chromosome Size. *Mol. Biol. Evol.* **38**, 4449–4462 (2021).

7. McGee, M. D. *et al.* The ecological and genomic basis of explosive adaptive radiation. *Nature* **586**, 75–79 (2020).

8. Brawand, D. *et al.* The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381 (2015).

9. Baduel, P., Quadrana, L., Hunter, B., Bomblies, K. & Colot, V. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat. Commun.* **10**, 5818 (2019).

10. Weissensteiner, M. H. *et al.* Discovery and population genomics of structural variation in a songbird genus. *Nat. Commun.* **11**, 3403 (2020).

11. Luo, J., Sun, X., Cormack, B. P. & Boeke, J. D. Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* **560**, 392–396 (2018).

12. Lowry, D. B. & Willis, J. H. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**, (2010).

13. Ricci, M., Peona, V., Guichard, E., Taccioli, C. & Boattini, A. Transposable Elements Activity is Positively Related to Rate of Speciation in Mammals. *J. Mol. Evol.* **86**, 303–310 (2018).
14. Davey, J. W. *et al.* Major Improvements to the Heliconius melpomene Genome Assembly Used to Confirm 10 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. *G3* **6**, 695–708 (2016).
15. Katju, V. & Bergthorsson, U. Copy-number changes in evolution: rates, fitness effects and adaptive significance. *Front. Genet.* **4**, 273 (2013).
16. Arnegard, M. E., Zwickl, D. J., Lu, Y. & Zakon, H. H. Old gene duplication facilitates origin and diversification of an innovative communication system—twice. *Proceedings of the National Academy of Sciences* **107**, 22172–22177 (2010).
17. Moriyama, Y. *et al.* Evolution of the fish heart by sub/neofunctionalization of an elastin gene. *Nat. Commun.* **7**, 10397 (2016).
18. Mondragón-Palomino, M. & Theissen, G. Why are orchid flowers so diverse? Reduction of evolutionary constraints by paralogues of class B floral homeotic genes. *Ann. Bot.* **104**, 583–594 (2009).
19. Givnish, T. J. *et al.* Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Mol. Phylogenet. Evol.* **71**, 55–78 (2014).
20. Barfuss, M. H. J. *et al.* Taxonomic revision of bromeliaceae subfam. Tillandsioideae based on a multi-locus DNA sequence phylogeny and morphology. *Phytotaxa* **279**, 1–97 (2016).
21. Gitaí, J., Paule, J., Zizka, G., Schulte, K. & Benko-Iseppon, A. M. Chromosome numbers and DNA content in Bromeliaceae: Additional data and critical review. *Bot. J. Linn. Soc.* **176**, 349–368 (2014).

22. De La Harpe, M. *et al.* Genomic footprints of repeated evolution of CAM photosynthesis in a Neotropical species radiation. *Plant Cell Environ.* **43**, 2987–3001 (2020).

23. Benzing, D. H. & Bennett, B. *Bromeliaceae: Profile of an Adaptive Radiation*. (Cambridge University Press, 2000).

24. Crayn, D. M., Winter, K., Schulte, K. & Smith, J. A. C. Photosynthetic pathways in Bromeliaceae: Phylogenetic and ecological significance of CAM and C3 based on carbon isotope ratios for 1893 species. *Bot. J. Linn. Soc.* **178**, 169–221 (2015).

25. Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).

26. Cresse, A. D., Hulbert, S. H., Brown, W. E., Lucas, J. R. & Bennetzen, J. L. Mu1-related transposable elements of maize preferentially insert into low copy number DNA. *Genetics* **140**, 315–324 (1995).

27. Brown, G. K. & Gilmartin, A. J. Chromosome Numbers in Bromeliaceae. *Am. J. Bot.* **76**, 657–665 (1989).

28. Ming, R. *et al.* The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* **47**, 1435–1442 (2015).

29. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

30. Lovell, J. T. *et al.* GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife* **11**, e78526 (2022).

31. Araújo, W. L. *et al.* Antisense inhibition of the iron-sulphur subunit of succinate dehydrogenase enhances photosynthesis and growth in tomato via an organic acid-mediated effect on stomatal aperture. *Plant Cell* **23**, 600–627 (2011).

32. Liu, L. *et al.* XAP5 CIRCADIAN TIMEKEEPER specifically modulates 3' splice site recognition and is important for circadian clock regulation partly by alternative splicing of LHY and TIC. *Plant Physiol. Biochem.* **172**, 151–157 (2022).
33. Hayashi, S. *et al.* The glycerophosphoryl diester phosphodiesterase-like proteins SHV3 and its homologs play important roles in cell wall organization. *Plant Cell Physiol.* **49**, 1522–1535 (2008).
34. Weiland, M., Mancuso, S. & Baluska, F. Signalling via glutamate and GLRs in Arabidopsis thaliana. *Funct. Plant Biol.* **43**, 1–25 (2015).
35. Kong, D. *et al.* L-Met Activates Arabidopsis GLR Ca2+ Channels Upstream of ROS Production and Regulates Stomatal Movement. *Cell Rep.* **17**, 2553–2561 (2016).
36. Philippe, F., Verdu, I., Morère-Le Paven, M.-C., Limami, A. M. & Planchet, E. Involvement of Medicago truncatula glutamate receptor-like channels in nitric oxide production under short-term water deficit stress. *J. Plant Physiol.* **236**, 1–6 (2019).
37. Heyduk, K. *et al.* Shared expression of crassulacean acid metabolism (CAM) genes pre-dates the origin of CAM in the genus Yucca. *J. Exp. Bot.* **70**, 6597–6609 (2019).
38. Tay, I. Y. Y., Odang, K. B. & Cheung, C. Y. M. Metabolic Modeling of the C3-CAM Continuum Revealed the Establishment of a Starch/Sugar-Malate Cycle in CAM Evolution. *Front. Plant Sci.* **11**, 2221 (2021).
39. Deng, H. *et al.* Evolutionary history of PEPC genes in green plants: Implications for the evolution of CAM in orchids. *Mol. Phylogenet. Evol.* **94**, 559–564 (2016).
40. Zhu, F. & Ming, R. Global identification and expression analysis of pineapple aquaporins revealed their roles in CAM photosynthesis, boron uptake and fruit domestication. *Euphytica* **215**, 132 (2019).

41. Simpson, G. G. *The Major Features of Evolution*. (Columbia University Press, 1953).
42. Wickell, D. *et al.* Underwater CAM photosynthesis elucidated by Isoetes genome. *Nature Communications 2021 12:1* **12**, 1–13 (2021).
43. Heyduk, K., McAssey, E. V. & Leebens-Mack, J. Differential timing of gene expression and recruitment in independent origins of CAM in the Agavoideae (Asparagaceae). *New Phytol.* **235**, 2111–2126 (2022).
44. Cai, J. *et al.* The genome sequence of the orchid Phalaenopsis equestris. *Nat. Genet.* **47**, 65–72 (2015).
45. Silvera, K., Winter, K., Rodriguez, B. L., Albion, R. L. & Cushman, J. C. Multiple isoforms of phosphoenolpyruvate carboxylase in the Orchidaceae (subtribe Oncidiinae): implications for the evolution of crassulacean acid metabolism. *J. Exp. Bot.* **65**, 3623–3636 (2014).
46. Ohno, S. *Evolution by gene duplication*. (Springer Berlin, 1970).
47. Pedro, D. L. F. *et al.* An Atlas of Plant Transposable Elements. *F1000Res.* **10**, 1194 (2021).
48. Yardeni, G. *et al.* Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. *Mol. Ecol. Resour.* (2021) doi:10.1111/1755-0998.13523.
49. Liu, L. *et al.* Draft genome of Puya raimondii (Bromeliaceae), the Queen of the Andes. *Genomics* **113**, 2537–2546 (2021).
50. de Vos, J. M., Augustijnen, H., Bätscher, L. & Lucek, K. Speciation through chromosomal fusion and fission in Lepidoptera. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190539 (2020).
51. Faria, R. & Navarro, A. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* **25**, 660–669 (2010).

52. Mérot, C., Oomen, R. A., Tigano, A. & Wellenreuther, M. A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends Ecol. Evol.* **35**, 561–572 (2020).

53. Galbraith, D. W. *et al.* Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* **220**, 1049–1051 (1983).

54. Temsch, E. M. Genome size in liverworts. *Preslia* **82**, 63–80 (2010).

55. Temsch, E. M., Koutecký, P., Urfus, T., Šmarda, P. & Doležel, J. Reference standards for flow cytometric estimation of absolute nuclear DNA content in plants. *Cytometry A* **101**, 710–724 (2022).

56. Otto, F. J., Oldiges, H., Göhde, W. & Jain, V. K. Flow cytometric measurement of nuclear DNA content variations as a potential in vivo mutagenicity test. *Cytometry* **2**, 189–191 (1981).

57. Jang, T.-S. & Weiss-Schneeweiss, H. Formamide-Free Genomic in situ Hybridization Allows Unambiguous Discrimination of Highly Similar Parental Genomes in Diploid Hybrids and Allopolyploids. *Cytogenet. Genome Res.* **146**, 325–331 (2015).

58. Putnam, N. H. *et al.* Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).

59. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).

60. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).

61. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).

62. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460 (2018).

63. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

64. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).

65. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).

66. Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).

67. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-4.0. http://www.repeatmasker.org (2013-2015).

68. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

69. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-genome annotation with BRAKER. in *Gene prediction* 65–95 (Humana, 2019).

70. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome Annotation and Curation Using MAKER and MAKER-P. *Curr. Protoc. Bioinformatics* **2014**, 4.11.1-4.11.39 (2014).

71. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).

72. Chen, L.-Y. *et al.* The bracteatus pineapple genome and domestication of clonally propagated crops. *Nat. Genet.* **51**, 1549–1558 (2019).

73. Götz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435 (2008).

74. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).

75. Leroy, T. *et al.* A bird's white-eye view on avian sex chromosome evolution. *Peer Community Journal* **1**, (2021).

76. Alexa, A. & Rahnenführer, J. Gene set enrichment analysis with topGO. *Bioconductor Improv* **27**, 1–26 (2009).

77. Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N. & Delsuc, F. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Mol. Biol. Evol.* **35**, 2582–2584 (2018).

78. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

79. *AlignmentProcessor: Prepares CDS fasta Alignments for use with several programs*. (Github).

80. Wong, W. S. W., Yang, Z., Goldman, N. & Nielsen, R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**, 1041–1051 (2004).

81. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

82. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

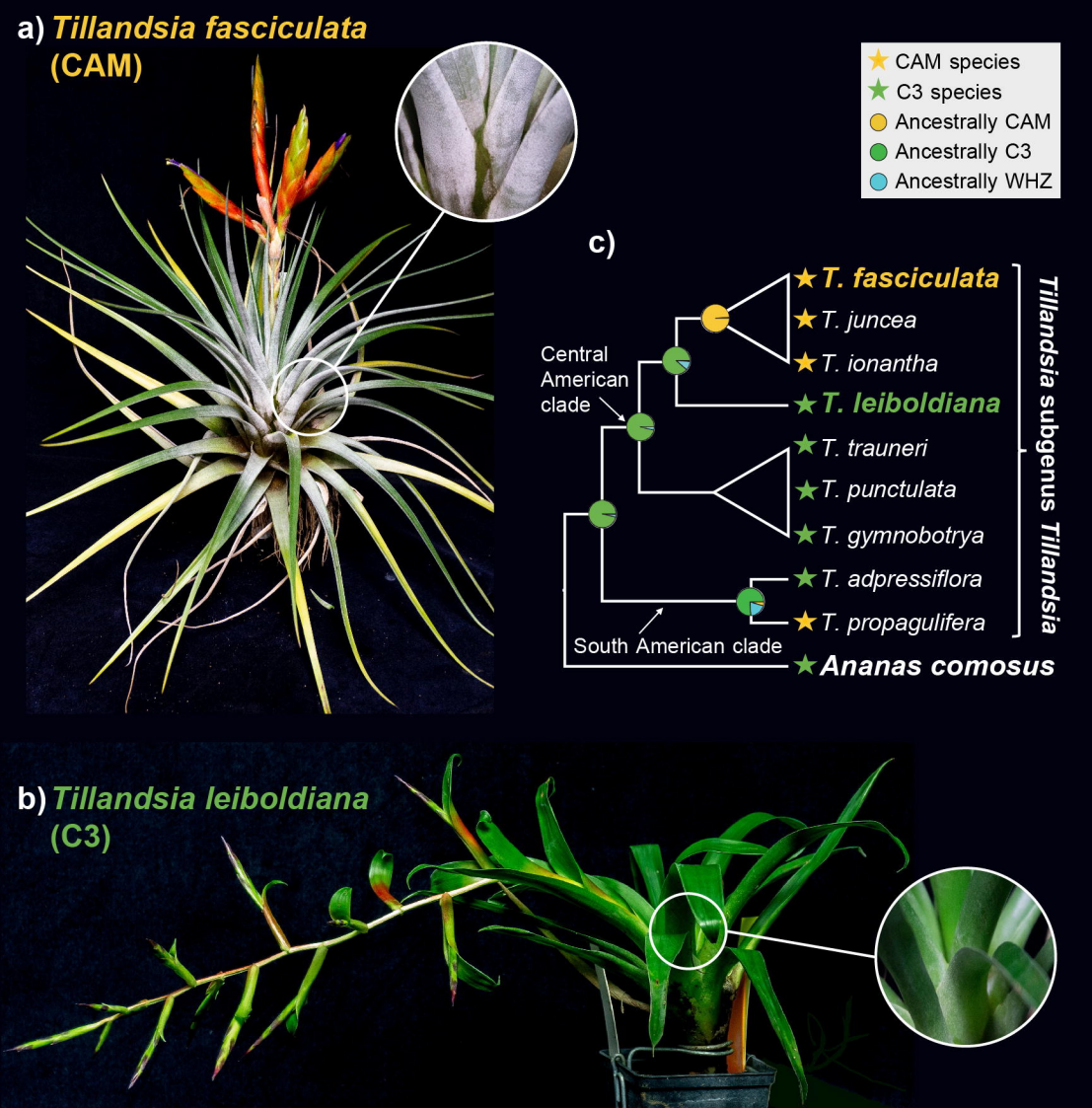
83. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21

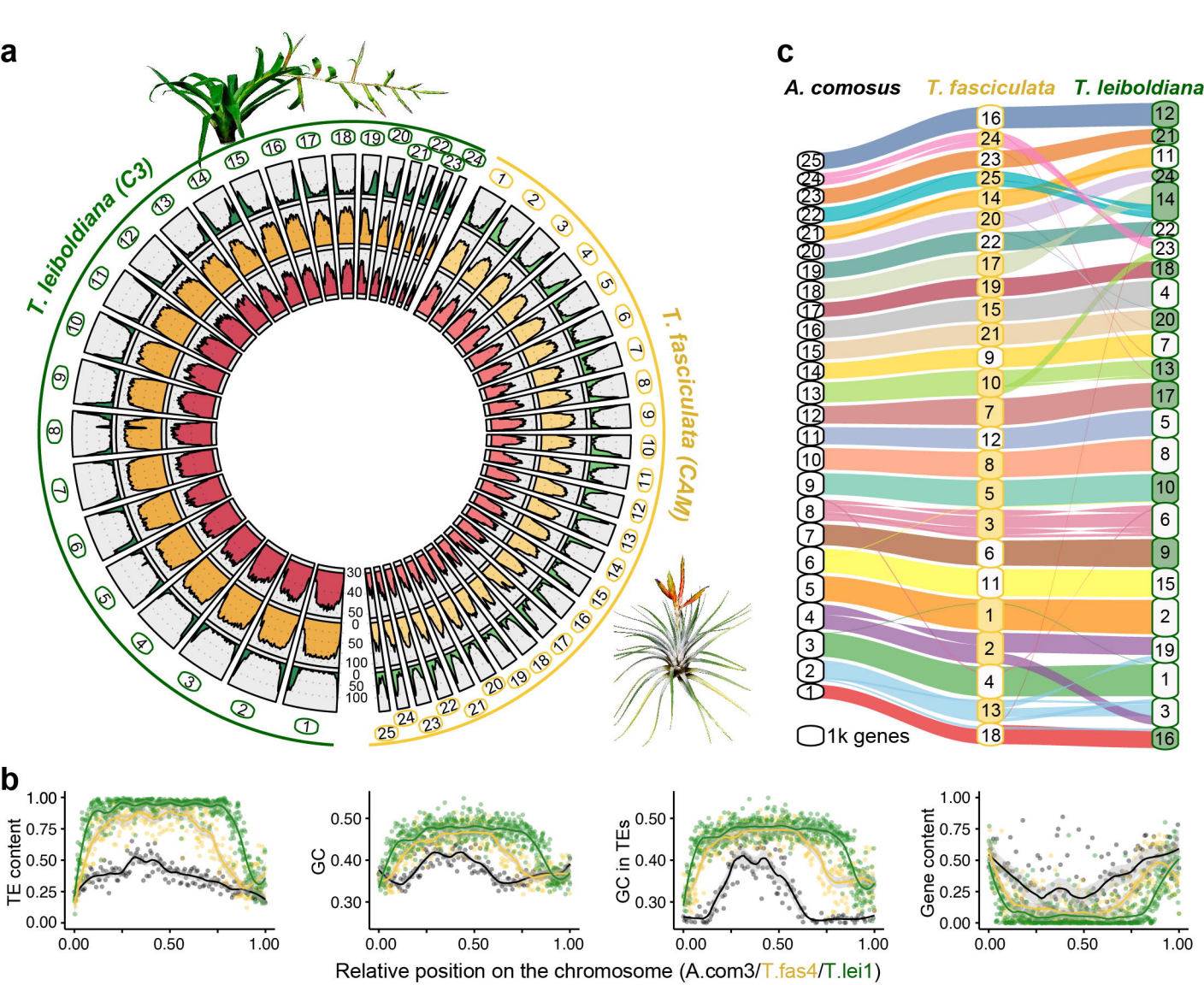
(2013).

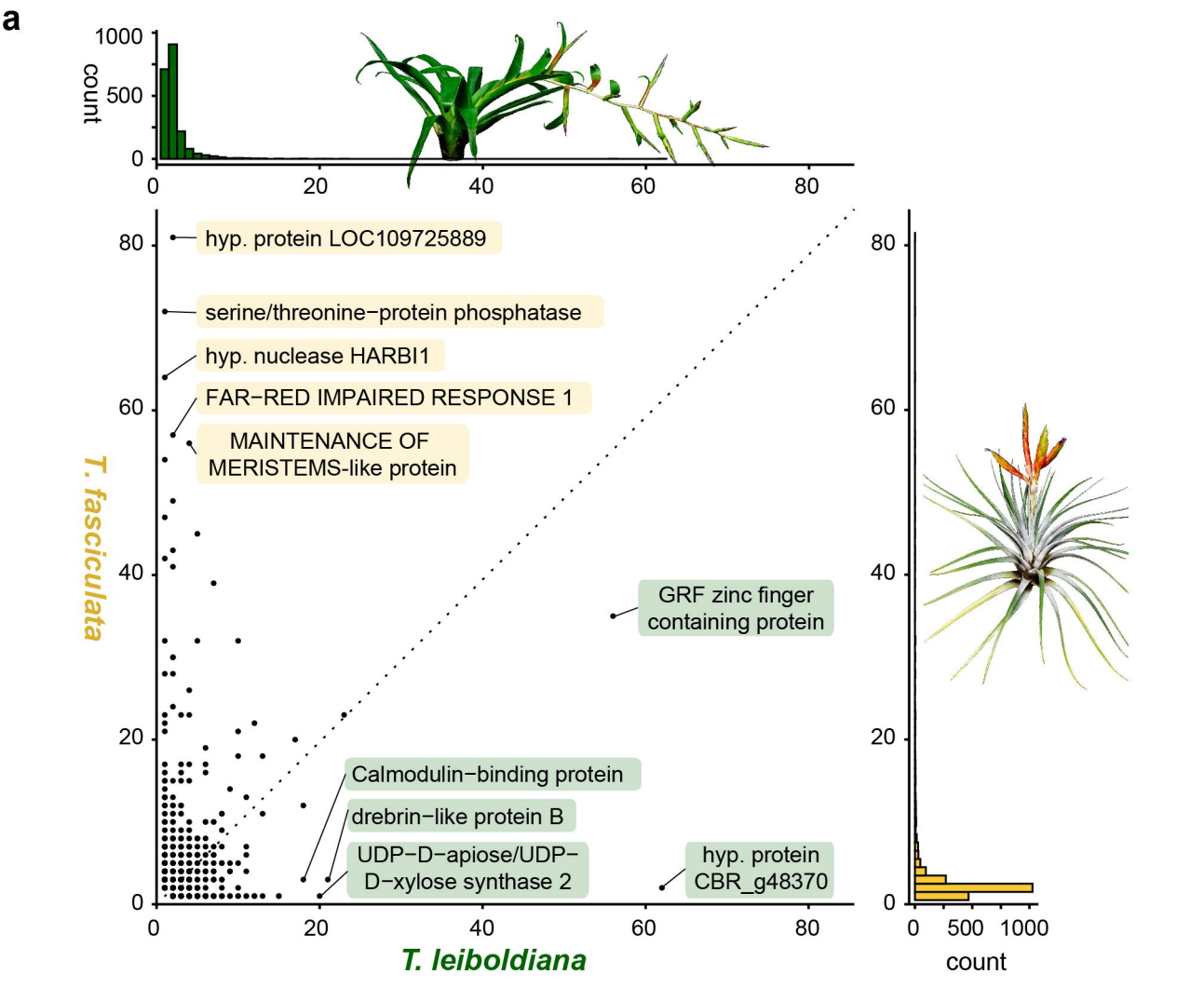
84. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

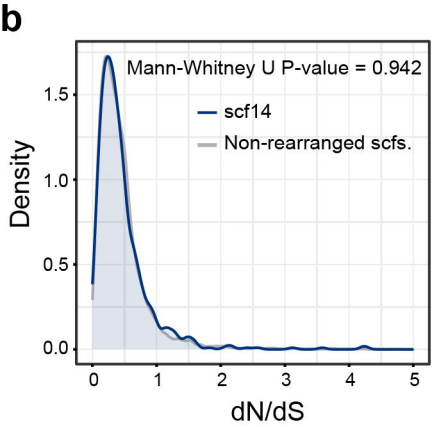
85. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).

86. Conesa, A., Nueda, M. J., Ferrer, A. & Talón, M. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* **22**, 1096–1102 (2006).


a) Tillandsia fasciculata
(CAM)
CAM species
C3 species
Ancestrally CAM
Ancestrally C3
Ancestrally WHZ
c)
T. fasciculata
T. juncea
T. ionantha
T. leiboldiana
T. trauneri
T. punctulata
T. gymnobotrya
T. adpressiflora
T. propagulifera
Ananas comosus
Central American clade
South American clade
Tillandsia subgenus Tillandsia
b) Tillandsia leiboldiana
(C3)


a
T. leiboldiana (C3)
T. fasciculata (CAM)
c
A. comosus
T. fasciculata
T. leiboldiana
1k genes
b
TE content
GC
GC in TEs
Gene content
Relative position on the chromosome (A.com3/T.fas4/T.lei1)

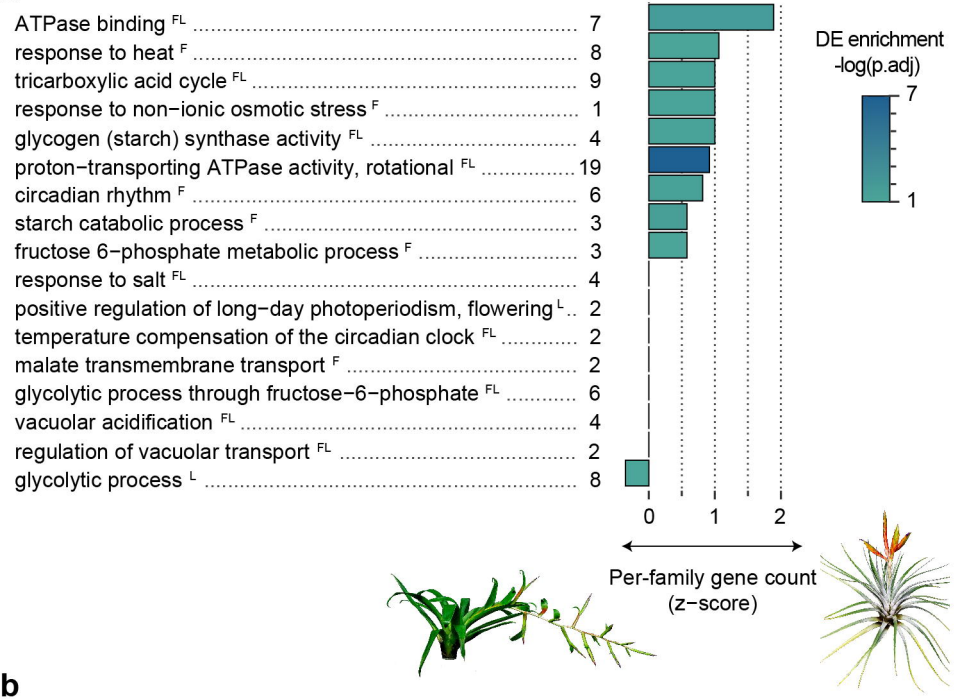




c

Orthogroup	dN/dS	adj-P	Function
OG0002972	∞	0.00162	jacalin-related lectin 3-like
OG0005000	∞	0.00628	cucumber peeling cupredoxin-like
OG0006253	∞	0.00066	mitochondrial prohibitin-3
OG0009278	∞	0.00155	chloroplastic Peroxiredoxin-2E-2
OG0012770	∞	0.01374	metallo-hydrolase/oxidoreductase superfamily protein
OG0011786	9.3	0.00008	U-box_domain-containing_protein
OG0015603	8.1	0.01575	anaphase-promoting complex subunit CDC27
OG0009004	3.1	0.01148	Hydroquinone glycosyltransferase
OG0008977	2.6	0.00512	glutamate receptor 2.8-like
OG0010014	2.0	0.00603	Glycerophosphodiester phosphodiesterase GDPDL7

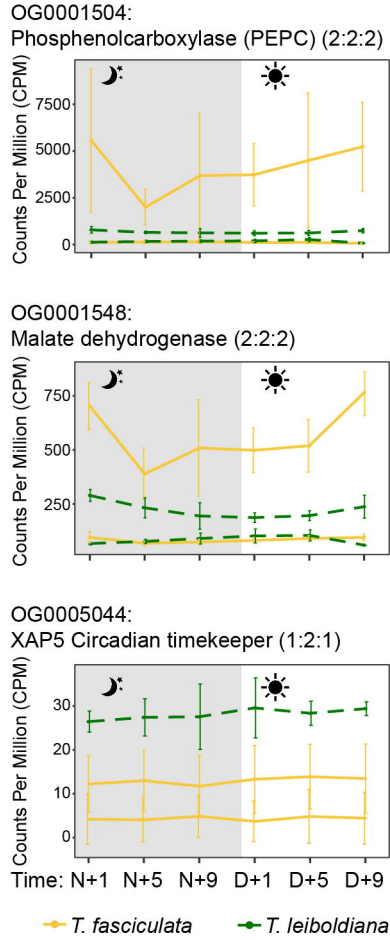
a

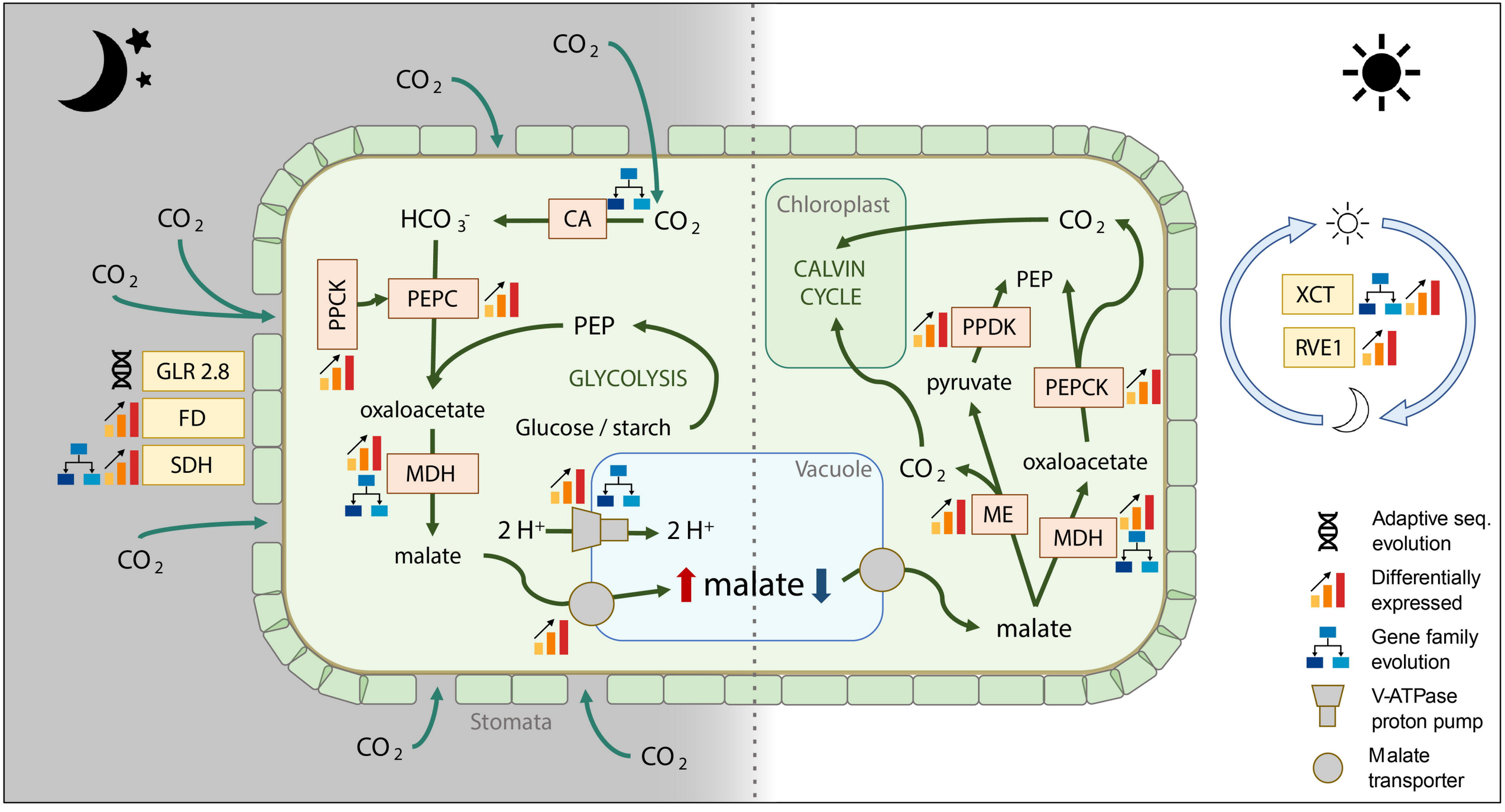


b

	Whole genome	DE in *T. leiboldiana* (C3)	DE in *T. fasciculata* (CAM)
Total orthogroup count	18697	714	738
% of 1:1 OGs	0.75	0.69*	0.66*
% of F:L OGs (F > L)	0.05	0.07*	0.16***
% of F:L OGs (F < L)	0.03	0.06**	0.02
% of F:L OGs (F = L)	0.03	0.08**	0.07**
% of F:0 OGs	0.06	0.00	0.06
% of 0:L OGs	0.07	0.08	0.00

c




CO2
CO2
CO2
CO2
CO2
CO2
CO2
CO2
CA
HCO3-
PPCK
PEPC
PEP
GLYCOLYSIS
Glucose / starch
oxaloacetate
MDH
malate
GLR 2.8
FD
SDH
2 H+
2 H+
malate
Vacuole
Chloroplast
CALVIN CYCLE
PPDK
pyruvate
PEP
PEPCK
oxaloacetate
ME
MDH
malate
Stomata
XCT
RVE1
Adaptive seq. evolution
Differentially expressed
Gene family evolution
V-ATPase proton pump
Malate transporter