CAM evolution is associated with gene family expansion in an explosive bromeliad radiation

Short title: CAM is associated with gene family expansion

Clara Groot Crego[1,2], Jaqueline Hess[1,3], Gil Yardeni[1,4], Marylaure de La Harpe[1,5], Clara Priemer[6], Francesca Beclin[1,2,7], Sarah Saadain[1,2], Luiz A. Cauz-Santos[1], Eva M. Temsch[1], Hanna Weiss-Schneeweiss[1], Michael H.J. Barfuss[1], Walter Till[1], Wolfram Weckwerth[6,8], Karolina Heyduk[9], Christian Lexer[1,†], Ovidiu Paun[1]*, Thibault Leroy[1,10]*

[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]Department of Functional and Evolutionary Ecology, Molecular Systems Biology (MOSYS), University of Vienna, Vienna, Austria
[7] Gregor Mendel Institute, Austrian Academy of Sciences, Vienna BioCenter, Vienna, Austria
[8] Vienna Metabolomics Center (VIME), University of Vienna, Vienna, Austria
[9]Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT USA
[10]GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet Tolosan, France

[†]deceased
*Joint last authors

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

The author(s) responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plcell/pages/General-Instructions) is (are): Clara Groot Crego (clara.groot.crego@univie.ac.at).

1. Abstract

The subgenus *Tillandsia* (Bromeliaceae) belongs to one of the fastest radiating clades in the plant kingdom and is characterised by the repeated evolution of Crassulacean Acid Metabolism (CAM). Despite its complex genetic basis, this water-conserving trait has evolved independently across many plant families and is regarded as a key innovation trait and driver of ecological diversification in Bromeliaceae. By producing high quality genome assemblies of a *Tillandsia* species pair displaying divergent photosynthetic phenotypes, and combining genome-wide investigations of synteny, TE dynamics, sequence evolution, gene family evolution and temporal differential expression, we were able to pinpoint the genomic drivers of CAM evolution in *Tillandsia*. Several large-scale rearrangements associated with karyotype changes between the two genomes and a highly dynamic TE landscape shaped the genomes of *Tillandsia*. However, our analyses show that rewiring of photosynthetic metabolism is mainly obtained through regulatory evolution rather than coding sequence evolution, as CAM-related genes are differentially expressed across a 24-hour cycle between the two species, but are no candidates of positive selection. Gene orthology analyses reveal that CAM-related gene families manifesting differential expression underwent accelerated gene family expansion in the constitutive CAM species, further supporting the view of gene family evolution as a driver of CAM evolution.

Word count for Abstract: **201**

Word count for Introduction: **1,166**

Word count for Results: **3,503**

Word count for Discussion: **1,474**

Total word count: 6,143

2. Introduction

Crassulacean Acid Metabolism (CAM) is a photosynthetic phenotype playing a major role in plant adaptation to arid environments and the epiphytic lifeform (Winter and Smith, 2012; Cushman, 2001; Silvera et al., 2010), and has been described as a key innovation trait driving plant diversification and speciation in several plant lineages (Quezada and Gianoli, 2011; Silvera et al., 2009; Ogburn and Edwards, 2009). CAM functions as a carbon concentrating mechanism by assimilating CO_2 overnight and storing it as malate in the vacuole, which greatly enhances the efficiency of Rubisco, the first enzyme of the Calvin cycle (Osmond, 1978). This also has the secondary effect of improving the plant's overall water use efficiency by reducing evapotranspiration, as stomata can remain closed during the day (Borland et al., 2014). Though often presented as a discrete trait, CAM actually encompasses a large spectrum of photosynthetic phenotypes including intermediate and facultative forms (Edwards, 2023). Phenotypes from this CAM continuum have evolved repeatedly in at least 37 plant families (Winter et al., 2021), yet the underlying evolutionary mechanisms allowing this complex and diverse trait to emerge multiple times throughout plant history are not fully understood.

Due to the sparse availability of CAM plant genomes, the majority of studies on CAM evolution have focussed on transcription levels and sequence evolution to understand its underlying genetic drivers. However, novel variation can be generated by other mechanisms which have not been investigated thoroughly in the context of CAM evolution. For example, several studies have pointed at the potential importance of gene family expansion as a driver of CAM evolution (Cai et al., 2015; Silvera et al., 2014). In C4 plants, duplicated gene copies tend to be more often retained compared to closely related C3 lineages (Hoang et al., 2023). Gene duplication occurs at higher rates than point mutation in many lineages (Katju and Bergthorsson, 2013) and can lead to novel functional variation through dosage effects, neo-functionalization, or subfunctionalization (Ohno, 1970), as observed in teleost fish (Arnegard et al., 2010; Moriyama et al., 2016) and orchids (Mondragón-Palomino and Theissen, 2009). Another form of structural variation that can contribute to the evolution of complex traits is TE insertion in and around genes, which has been shown to play a role in local adaptation, for example in *Arabidopsis* (Baduel et al., 2019). Large-scale rearrangements such as chromosomal fusions, inversions or translocations can increase linkage between co-adapted alleles and generate reproductive barriers (Luo et al., 2018; Lowry and Willis, 2010).

91 The adaptive radiation of *Tillandsia* subgenus *Tillandsia* (Bromeliaceae) is part of one of
92 the fastest diversifying clades known in the plant kingdom (Tillandsioideae) (Givnish et al., 2014)
93 and is characterised by a number of key innovation traits such as epiphytism, absorptive trichomes,
94 water-impounding tanks and photosynthetic metabolism driving extraordinary diversity both on
95 the taxonomic and ecological level (Barfuss et al., 2016). The group displays a broad range of
96 phenotypes of the CAM continuum, resulting in repeated evolution of constitutive CAM (Crayn
97 et al., 2015; De La Harpe et al., 2020). CAM evolution has been described as an ecological driver
98 of diversification in the subgenus *Tillandsia* (Barfuss et al., 2016; Crayn et al., 2004), and across
99 Bromeliaceae in general (Givnish et al., 2014; Crayn et al., 2004; Benzing and Bennett, 2000).
100 This renders the radiation a fascinating system both for studies on speciation and rapid adaptation
101 generally, and for studies on CAM evolution specifically, as comparative investigations between
102 recently diverged species with contrasting phenotypes prevent the overestimation of evolutionary
103 changes needed to evolve adaptations such as CAM (Heyduk et al., 2019a).
104 While Bromeliaceae is generally regarded as a homoploid radiation with conserved
105 chromosome counts and little genome size variation (Gitaí et al., 2014), more recent work has
106 pointed at a high "genomic potential" of the subgenus *Tillandsia*, notably from elevated gene loss
107 and duplication rates (De La Harpe et al., 2020), providing an exemplary system to study the role
108 of genome evolution and structural variation in CAM evolution. Not only are adaptive radiations
109 like *Tillandsia* characterised by repeated evolution of key innovation traits, the short timescales at
110 which novel variation arises in these systems challenges classical views of adaptive evolution,
111 stimulating a range of studies pointing at the potential importance of genome evolution as a
112 genomic driver of diversification (McGee et al., 2020; Brawand et al., 2015; Cicconardi et al.,
113 2021).
114 In this study, we comparatively investigated *de novo* assembled genomes of two
115 ecologically divergent members of the subgenus *Tillandsia* to further our understanding of genome
116 evolution in this recent radiation and its link to CAM evolution as a key innovation trait. *Tillandsia*
117 *fasciculata* (Fig. 1a) displays a set of phenotypes typically described as "grey" or "atmospheric"
118 *Tillandsia* (Benzing and Bennett, 2000): a dense layer of absorptive trichomes, CAM
119 photosynthesis and occurrence in arid places with high solar incidence and low rainfall. On the
120 other hand, *T. leiboldiana* (Fig. 1b) represents a typical "green" *Tillandsia* displaying tank
121 formation, C3-like leaf morphology, a lack of absorptive trichomes and occurrences in cooler,

122 wetter regions. While not sister species, *T. fasciculata* and *T. leiboldiana* belong to sister clades
123 displaying a shift in photosynthetic metabolism (Fig. 1c), and represent phenotypic extremes
124 within subgenus *Tillandsia*. Their photosynthetic metabolisms have been described with carbon
125 isotope ratios as strong CAM for *T. fasciculata* (carbon isotope ratio: $\delta^{13}C$ = -11.9 / -16.1, (Crayn
126 et al., 2015; De La Harpe et al., 2020)) and C3 for *T. leiboldiana* ($\delta^{13}C$ = -28.0 / -31.3; (Crayn et
127 al., 2015; De La Harpe et al., 2020), and are some of the most distinct values reported for the
128 subgenus on the CAM continuum.
129 However, due to the limited ability of carbon isotope measurements in capturing
130 intermediate CAM phenotypes and therefore representing the full CAM continuum (Messerschmid
131 et al., 2021; Pierce et al., 2002), the exact photosynthetic phenotypes of *T. leiboldiana* and *T.*
132 *fasciculata* need to be corroborated to fully understand what range of the CAM continuum is truly
133 encompassed in the subgenus. By characterising the photosynthetic metabolisms of *T. leiboldiana*
134 and *T. fasciculata* and investigating genomic variation between these two species on multiple
135 levels, from karyotype, chromosomal rearrangements, to molecular evolution, gene family
136 evolution and temporal differential gene expression, we thoroughly explore the degree of genomic
137 divergence found within this radiation and the link between this variation and the evolution of a
138 key innovation trait. We ascertain that the photosynthetic metabolisms of these species are clearly
139 distinct, with *T. fasciculata* at the late stages of CAM evolution (i.e., constitutive, strong CAM),
140 and *T. leiboldiana* likely at the very early stages (i.e., no night-time malate accumulation, but
141 CAM-like expression profiles of certain enzymes). We further document karyotype differences,
142 multiple chromosomal rearrangements, distinct TE landscapes and gene family evolution rates
143 between the two species. We further find evidence that molecular variation underlying the
144 difference in phenotype is largely found at the transcriptomic level. We also found a clear
145 association between CAM-related temporal gene expression differences and both gene family
146 expansion in the constitutive CAM plant and pre-existing gene duplications shared between both
147 species.
148

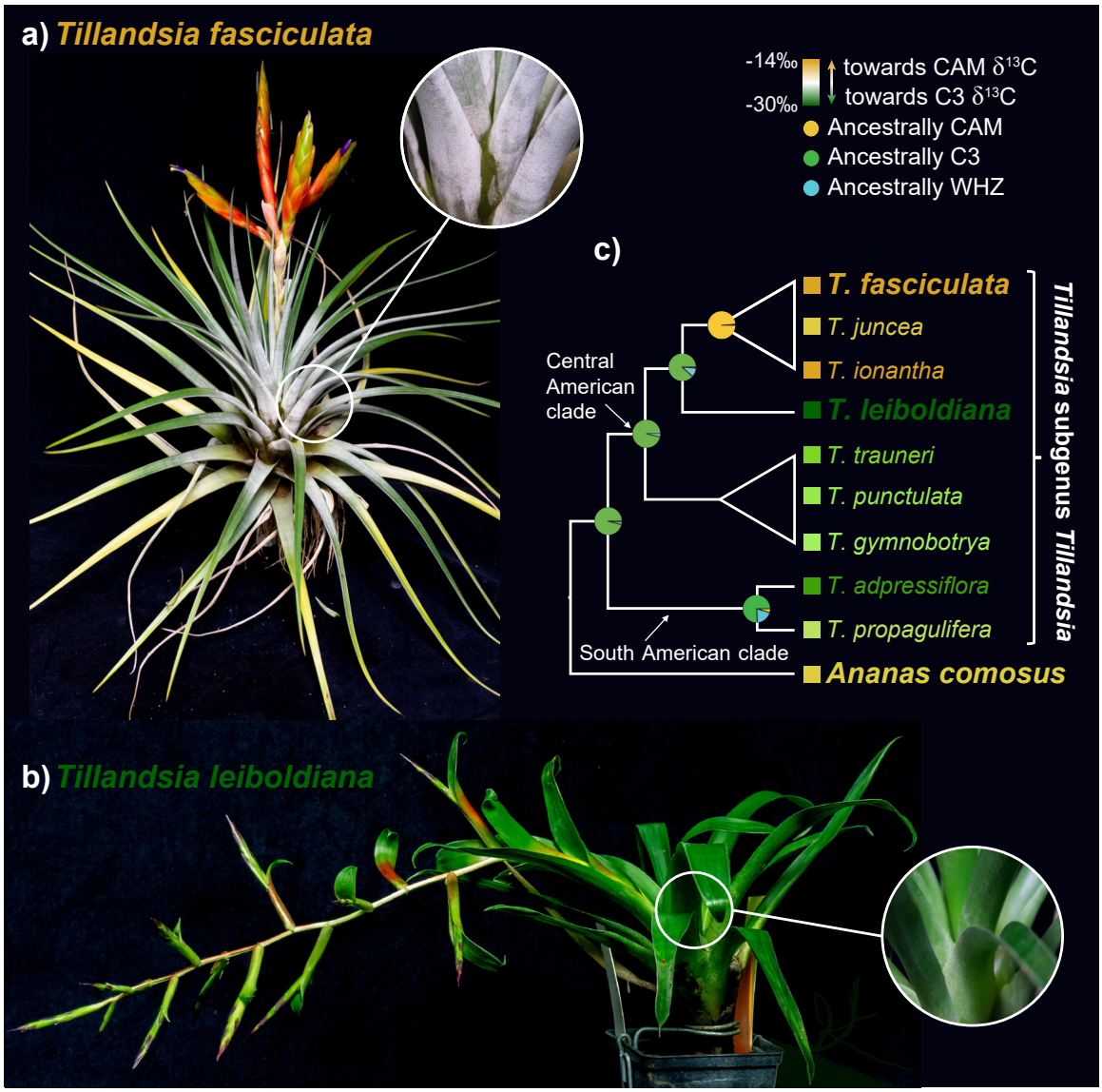


Figure 1: a) Tillandsia fasciculata*, a "grey" or "atmospheric"* Tillandsia *with a dense layer of trichomes, carbon isotope values within the CAM photosynthesis range, a lack of water-impounding tank and roots adapted to the epiphytic lifestyle.* ***b)*** Tillandsia leiboldiana*, a green* Tillandsia *with C3-like leaf morphology and carbon isotope values, an impounding tank and reduced trichome layer.* ***c)*** *Schematic representation of the evolutionary relationship between the two investigated species of* Tillandsia *within the subgenus. Colours indicate reported carbon isotope values* (De La Harpe et al., 2020; Crayn et al., 2015)*. The average was taken when multiple values have been reported for the same species. Pie charts at internal nodes show the ancestral state of photosynthetic metabolism as reported in De La Harpe et al. 2020. WHZ stands for Winter-Holtum Zone* (Males, 2018) *and represents intermediate forms of the* CAM continuum.

3. Results

3.1. Photosynthetic phenotypes of *T. fasciculata* and *T. leiboldiana*

To better understand the difference in photosynthetic metabolism between *T. fasciculata* and *T. leiboldiana*, we measured metabolite abundances with GC-MS for six samples per species at six timepoints across a 24-hour cycle. The overall composition of 49 metabolite abundances separate samples of both species within the first two principal components (Fig. 2a), which combined explain 41.5 % of variance. This suggests a pronounced general metabolic differentiation between *T. fasciculata* and *T. leiboldiana*. Besides amino acids, the organic acids malate, citrate and gluconic acid contribute most to this differentiation along PC1 (Fig. S1), which is a common pattern for species with diverging photosynthetic metabolism (De La Harpe et al., 2020; Benzing and Bennett, 2000; Popp et al., 2003). Sugars appear large contributors to differentiation along PC2 (Fig. S1).

Overnight malate accumulation is a core feature of CAM photosynthesis and therefore an indicator of the respective photosynthetic phenotypes of *T. fasciculata* and *T. leiboldiana*. Malate abundances in the leaf fluctuated strongly in *T. fasciculata* over 24 hours, with highest median abundances around midday (D+5) and lowest abundances in the early night (N+1), representing a 3.8-fold difference (Fig. 2b). In comparison, malate abundances were overall lower for *T. leiboldiana* and fluctuated less. The highest median abundance in the latter species was found at N+1, while the lowest was found at D+5, representing a 2.2-fold difference. Interestingly, the accumulation times of malate seem reversed in the two species, with the highest abundances in *T. fasciculata* found at the time of lowest abundances in *T. leiboldiana* and vice versa. The reversed timing of malate accumulation has been described as a key difference between C3 and CAM metabolisms ((Winter and Smith, 2022), but see also (Bräutigam et al., 2017)). The median accumulation in malate within 24-hours differs significantly between species (Wilcoxon rank-sum test p-value = 8.6xE-3, Fig. 2c), and is 3.2-fold higher in *T. fasciculata* than in *T. leiboldiana*.

Overall, the malate accumulation curves suggest distinct photosynthetic phenotypes for *T. fasciculata* and *T. leiboldiana*. *T. fasciculata* appears to behave as a constitutive CAM plant in normal conditions, accumulating malate from the early night until the early day, while *T. leiboldiana*'s flux is more C3-like, without a clear accumulation overnight.

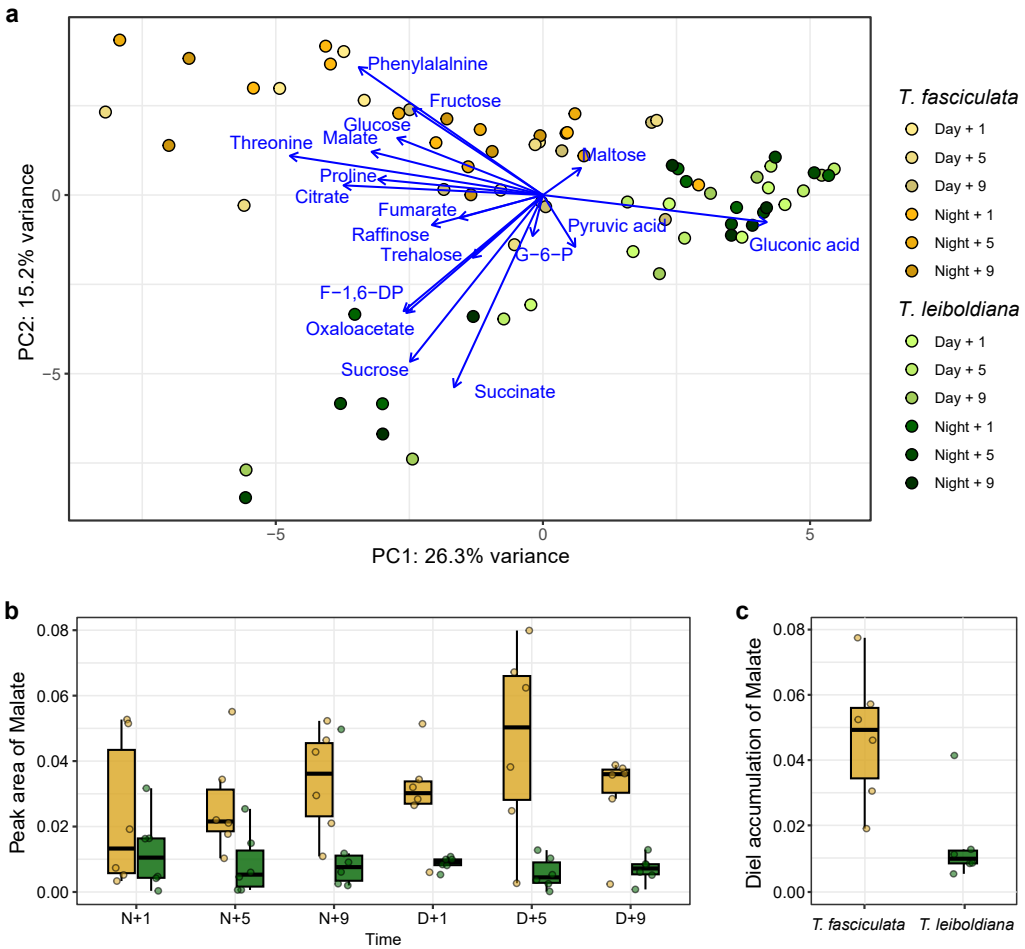


***Figure 2**: Metabolomic analyses of* T. fasciculata *and* T. leiboldiana *leaf material throughout a 24-hour cycle. Abundances of individual metabolites were measured with GC-MS and normalised against the Main Total Ion Count (MTIC). **a)** Principal component analysis of metabolic composition of 72 leaf samples based on 77 metabolic compounds including soluble sugars, amino acids and organic acids. The first and second principal components are displayed. Blue arrows show the loadings of a subset of metabolites relevant for photosynthetic metabolism. **b)** Malate abundance in leaf material of* T. fasciculata *and* T. leiboldiana *at six timepoints across a 24-hour cycle. Dots represent individual observations across timepoints. Timepoints are noted as hours into the day (D) or into the night (N). **c)** Distribution of per-individual accumulation of malate per species over a 24-hour cycle. Accumulations were obtained by taking the difference in malate abundance between the highest and lowest reported abundances across time for each accession.*

202 ### 3.2. Genome assembly and annotation

203 We constructed *de novo* haploid genome assemblies for both species (Table S1) using a
204 combination of long-read (PacBio), short read (Illumina) and chromosome conformation capture
205 (Hi-C) data. This resulted in assemblies of 838 Mb and 1,198 Mb with an N50 of 23.6 and 43.3
206 Mb in *T. fasciculata* and *T. leiboldiana* respectively. The assembly sizes closely match the
207 estimated genome size of each species based on flow cytometry and k-mer analysis (Table S2, SI
208 Note 1,2). The 25 and respectively 26 longest scaffolds (hereafter referred to as 'main scaffolds')
209 contain 72 % and 75.5 % of the full assembly, after which scaffold sizes steeply decline (SI Note
210 3, Fig. S2). The number of main scaffolds corresponds with the species karyotype in *T. fasciculata*,
211 but deviates from the *T. leiboldiana* karyotype (SI Note 1), suggesting that a few fragmented
212 chromosome sequences remain in the latter assembly.

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

218

219 ### 3.3. Genic, repetitive and GC content

220 TE annotation performed with EDTA (Ou et al., 2019) revealed a total repetitive content
221 of 65.5 % and 77.1 % in *T. fasciculata* and *T. leiboldiana* respectively. This closely matches
222 estimates derived from k-mer analyses (66 % and 75 %, SI Note 2). Compared to *T. fasciculata*,
223 the repetitive content in *T. leiboldiana* is enriched for Gypsy LTR retrotransposon and *Mutator*
224 DNA transposon content, with a 1.7-fold and 4.2-fold increase in total covered genomic length,
225 respectively (Table S3).

226 Repetitive content per scaffold is negatively correlated with gene count in both assemblies
227 (Kendall's correlation coefficient: -0.79 in *T. fasciculata*, -0.82 in *T. leiboldiana,* p-values < 2.2E-
228 16), with gene-rich regions in distal positions (Fig. 3a, green track) and repetitive regions primarily
229 in median positions (Fig. 3a, yellow track). This pattern is accentuated in *T. leiboldiana*: on
230 average, the repetitive-to-exonic content per scaffold is 1.6 times larger compared to *T. fasciculata*
231 (Mann Whitney U, p-value = 4.3E-04). The genome size difference between the two assemblies is

therefore mostly explained by differential accumulation of TE content, mostly in heterochromatic regions.

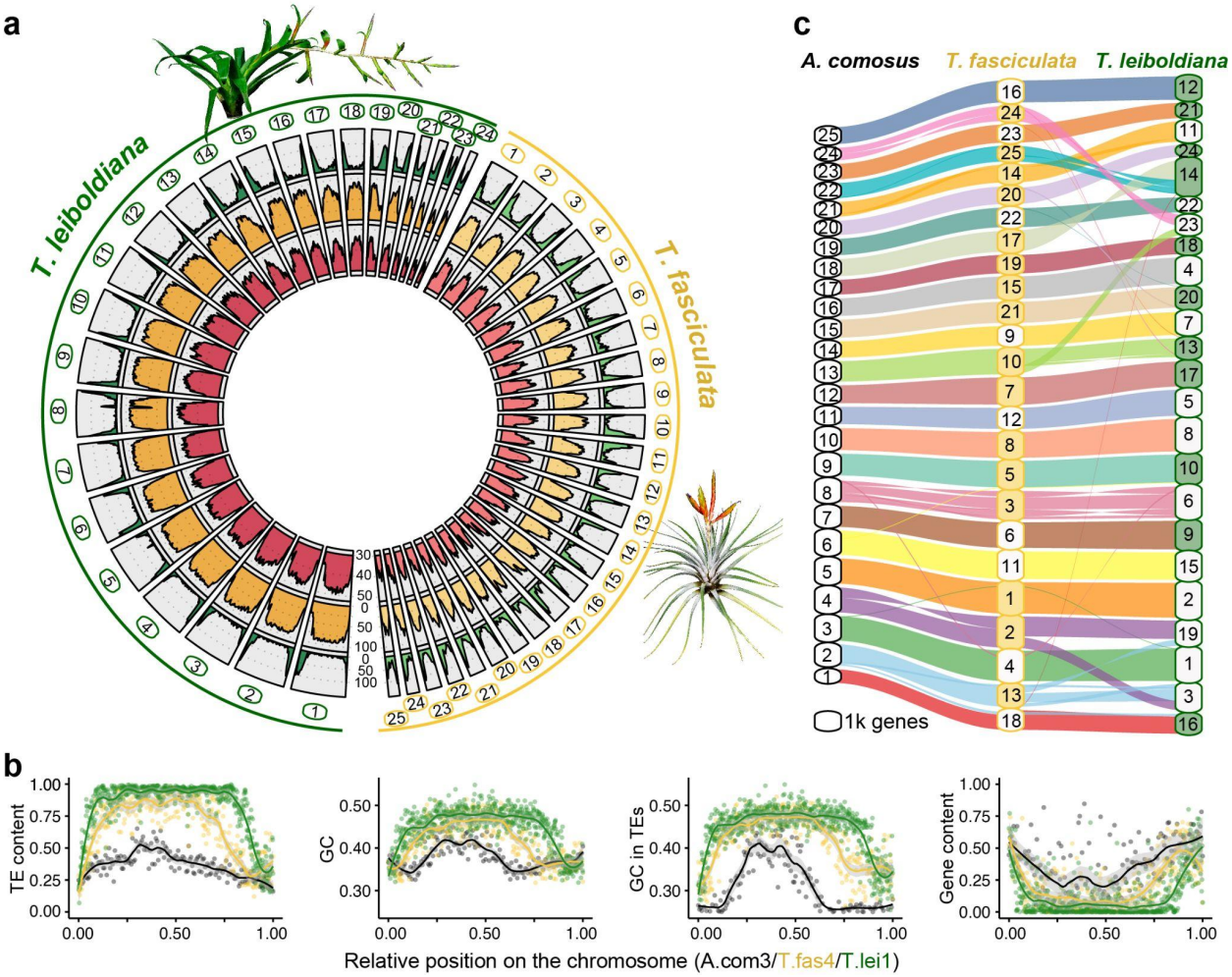
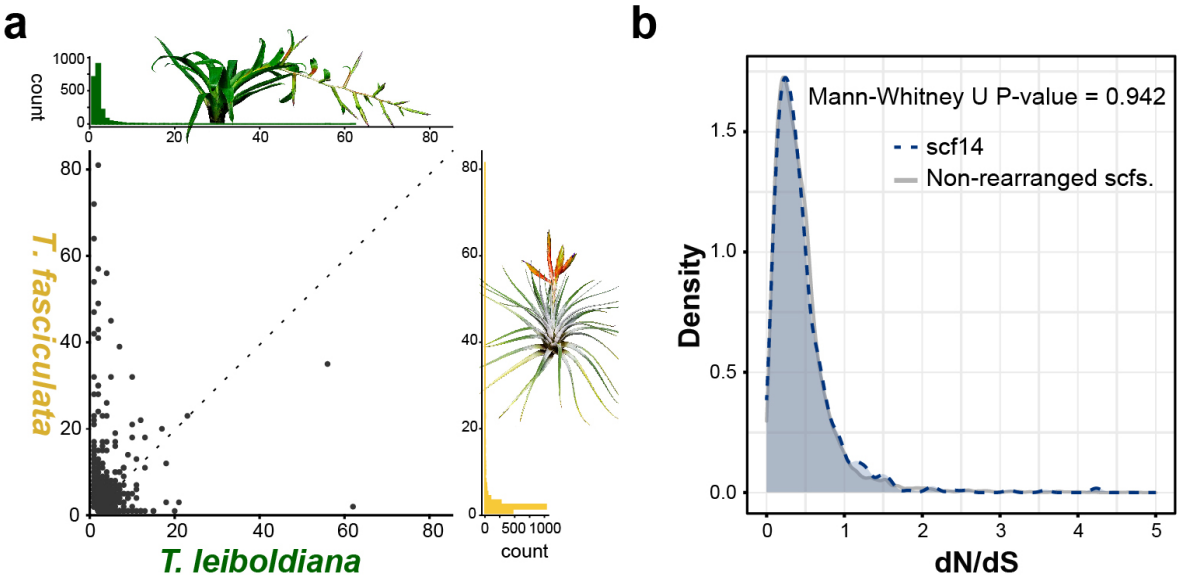


Figure 3: a) *Circular overview of the main scaffolds of the* T. fasciculata *(right) and* T. leiboldiana *(left) genome assemblies. Scaffolds 25 and 26 of* T. leiboldiana *are not shown due to their reduced size. Going inwards, the tracks show: (1, green) gene count; (2, yellow) proportion of repetitive content; (3, red), and GC content per 1-Mb windows.* ***b)*** *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. S3 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 proportional to genic content and therefore does not represent the true scaffold size. Colour-filled boxes indicate scaffolds with reversed coordinates as compared to the sequences in* A. comosus*.*

245 Contrary to the general expectation, GC content is negatively correlated with gene content
246 in both species (Kendall's correlation coefficient: -0.68 in *T. fasciculata*, -0.71 in *T. leiboldiana,*
247 p-values < 2.2E-16, red track in Fig. 3a, detailed in Fig. 3b). By visualising GC and TE content
248 across a syntenic chromosome triplet of *A. comosus*, *T. fasciculata* and *T. leiboldiana,* we show
249 that this unusual relationship can be mostly explained by elevated GC content in repetitive regions
250 (Fig. 3b). TE-rich regions indeed exhibit a much higher GC content than TE-poor regions, a pattern
251 which is exacerbated as the overall TE content per species increases (Fig. 3b, SI Note 4).
252
253 **3.4. Synteny and chromosomal evolution**
254 Cytogenetic karyotyping (SI Note 1) revealed a difference of six chromosome pairs
255 between *T. fasciculata* (2*n* = 50) and *T. leiboldiana* (2*n* = 38), which is atypical in this largely
256 homoploid clade with generally constant karyotype (Gitaí et al., 2014; Brown and Gilmartin,
257 1989). To investigate orthology and synteny, we inferred orthogroups between protein sequences
258 of *Ananas comosus* (Ming et al., 2015) (pineapple), *T. fasciculata* and *T. leiboldiana* using
259 Orthofinder (Emms and Kelly, 2019). This resulted in 21,045 (78 %), 26,325 (87.5 %) and 23,584
260 (75 %) gene models assigned to orthogroups respectively, of which 10,021 were single-copy
261 orthologs between all three species (Table S4).
262 Syntenic blocks were then defined across all three assemblies using Genespace (Lovell et
263 al., 2022) (Fig. 3c). The observed karyotype difference between species could arise from
264 chromosomal loss in *T. leiboldiana*. However, our Genespace analysis reveals highly conserved
265 synteny between the two assemblies without major orphan regions in *T. leiboldiana*, consistent
266 with a scenario of chromosomal fusion, rather than loss. We found clear evidence of such a fusion
267 on scaffold 14 in *T. leiboldiana* (Fig. 3c, Fig. S4a), which was confirmed with in-depth analyses
268 of potential breakpoints (SI Note 5). However, chromosomal rearrangements are not limited to
269 fusions, since we also detected two major reciprocal translocations (Fig. 3c, hereafter referred as
270 Translocations 1 and 2, Fig. S4b-c).
271
272 **3.5. Gene family evolution**
273 6,261 genes in *T. fasciculata* and 4,693 genes in *T. leiboldiana* were assigned to non-unique
274 gene families with multiple gene copies in at least one species, after correcting gene family sizes

275 (Table S4, SI Note 6). On average, the multicopy gene family size is 1.3x larger in *T. fasciculata*
276 than in *T. leiboldiana* (Mann Whitney U, p-value: 8.8E-16, Fig. 4a).
277 To investigate the role of expanded gene families in CAM evolution, we combined gene
278 ontology (GO) enrichment tests on multicopy orthogroups (SI Note 7) with a targeted search of
279 known genes involved in the CAM pathway. This highlighted 25 multi-copy gene families with
280 functions putatively related to CAM (Table S5), of which 17 have expanded in *T. fasciculata* and
281 eight in *T. leiboldiana.* Families expanded in *T. fasciculata* included a malate dehydrogenase
282 (MDH) and a β-carbonic anhydrase (CA), which are putatively involved in the carbon fixation
283 module of CAM photosynthesis, and subunits of the two vacuolar pumps (V-ATPase and V-
284 PPiase) known to energise the night-time transport of malate in pineapple (McRae et al., 2002)
285 (Table S5, Fig. 6). Additionally, two enolase families (members of the glycolysis pathway), a
286 vacuolar acid invertase putatively involved in day-time soluble sugar accumulation in the vacuole
287 (McRae et al., 2002; Holtum et al., 2005), and a pyrophosphate-dependent phosphofructokinase
288 associated with night-time conversion of soluble sugars through the glycolysis to PEP (Carnal and
289 Black, 1989) were expanded in *T. fasciculata*. Two subunits of succinate dehydrogenase, a
290 member of the tricarboxylic acid cycle and the electron transport chain which also plays a role in
291 stomatal opening regulation, a relevant aspect of CAM photosynthesis (Araújo et al., 2011), were
292 also expanded in *T. fasciculata* (Fig. 6). XAP5 CIRCADIAN TIMEKEEPER (XCT), a regulator
293 of circadian rhythm and disease resistance (Liu et al., 2022) which was previously identified as
294 undergoing rapid gene family evolution in *Tillandsia* (De La Harpe et al., 2020) is also expanded
295 in *T. fasciculata*.
296 Gene families expanded in *T. leiboldiana* contained three members of the glycolysis, two
297 aquaporins, a member of the tricarboxylic acid cycle and a regulator of stomatal opening (Table
298 S5).



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 4: a) *Scatterplot: composition of per-species gene counts among orthogroups. 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)*** *Single-copy orthogroups with significant* d_N/d_S *values and their functions. Three uncharacterized genes that are excluded here are detailed in Table S6. Infinite* d_N/d_S *values correspond to genes with* $d_S=0$ *(no synonymous substitutions), an expected situation considering the low divergence of the two species. Further explanation about the biological significance of these functions can be found in SI Note 8.*

310 **3.6. Adaptive sequence evolution**

311 Adaptive sequence evolution was evaluated in 9,077 orthologous gene pairs using the non-
312 synonymous to synonymous substitution ratio ($\omega = d_N/d_S$). Little among-scaffold variation in d_N/d_S
313 was observed, with per-scaffold median d_N/d_S values ranging from 0.32 to 0.39 in *T. fasciculata*
314 and 0.31 to 0.4 in *T. leiboldiana* (Fig. S5a). Regions of large chromosomal rearrangement such as
315 the fused scaffold 14 in *T. leiboldiana* do not exhibit strong signatures of fast coding sequence
316 evolution (Fig. 4b), though for Translocation 1, d_N/d_S values are slightly, yet significantly, lower
317 for scaffold 13 in *T. fasciculata* and scaffold 19 in *T. leiboldiana* (p-value < 0.05, Fig. S5b, SI
318 Note 5).

319 Among the 9,077 orthologous gene pairs, 13 candidates (0.21%) exhibit a significant d_N/d_S
320 > 1 (adjusted p-value < 0.05, Fig. 4c, Table S6, SI Note 8). Notably, we recover a significant signal
321 in a type B glycerophosphodiester phosphodiesterase (GDPDL-7). GDPDL's are involved in cell
322 wall cellulose accumulation and pectin linking, and play a role in trichome development (Hayashi
323 et al., 2008), a main trait differentiating the two species and more broadly, green and grey
324 *Tillandsia*. Additionally, GDPDL-7 may be involved in response to drought and salt stress (Cheng
325 et al., 2011).

326 A glutamate receptor (GLR) 2.8-like also exhibits a significant d_N/d_S > 1. By mediating
327 Ca2+ fluxes, GLRs act as signalling proteins and mediate a number of physiological and
328 developmental processes in plants (Weiland et al., 2015), including stomatal movement (Kong et
329 al., 2016). Although it is associated with drought-stress response in *Medicago trunculata* (Philippe
330 et al., 2019), the specific function of GLR2.8 still remains unknown.

331

332 **3.7. Gene expression analyses**

333 To study gene expression differences linked to distinct photosynthetic phenotypes, we
334 performed a time-series RNA-seq experiment using six plants of each species (Table S1, SI Note
335 9), sampled every four hours in a 24-hour period. We recovered 907 genes with a differential
336 temporal expression (DE) profile between *T. fasciculata* and *T. leiboldiana*. Among them are 46
337 known CAM-related genes and 22 genes associated with starch metabolism and
338 glycolysis/gluconeogenesis. GO term enrichment of the 907 DE genes revealed many CAM-
339 related functions such as malate and oxaloacetate transport, circadian rhythm, light response, water
340 and proton pumps, sucrose and maltose transport and starch metabolism (Table S7; Fig. 5a). While

341 none of the candidate genes for adaptive sequence evolution recovered in this study were
342 differentially expressed, nine of 22 genes reported by De La Harpe et al. 2020 as candidates for
343 adaptive sequence evolution during transitions to constitutive CAM in the wider context of the
344 genus *Tillandsia* were also differentially expressed in this study (Table S7).

345 Core CAM enzymes phosphoenolpyruvate carboxylase (PEPC) and phosphoenolpyruvate
346 carboxylase kinase (PEPC kinase, PPCK) display clear temporal expression cycling in *T.*
347 *fasciculata* (Fig. 5c, S5). PPCK also shows a night-time increase in expression in *T. leiboldiana*
348 (Fig. S6), albeit with a milder temporal effect, a phenomenon that has been documented before in
349 C3-assigned *Tillandsia* (De La Harpe et al., 2020) and also in other C3-like species belonging to
350 CAM- and C4-evolving lineages (Heyduk et al., 2019b, 2019a). Clustering analysis distributed DE
351 genes across seven clusters with sizes ranging from 209 to 38 genes (Table S7). CAM-related
352 genes were distributed across six of seven clusters, highlighting the diversity of expression profiles
353 associated with CAM (Fig. S7). While core CAM genes (see Fig. 6) are mainly found in cluster 5,
354 we find malate transporters in cluster 1, circadian regulators in clusters 2 and 3, sugar transport in
355 clusters 3 and 6, and vacuolar transport regulators in clusters 2, 4 and 6. Cluster 7, though not
356 containing any core CAM candidate genes, was enriched for salt and heat stress response and
357 contains a mitochondrial isocitrate dehydrogenase, which has been proposed as an alternative
358 carbon fixator in CAM plants (Tay et al., 2021; Töpfer et al., 2020).

359 The expression curves of the respective clusters (Fig. S7) demonstrate a complex web of
360 expression changes between photosynthetic phenotypes. The most common expression change
361 pattern among CAM-related genes is an overall increase in expression in the strong CAM plant
362 (*T. fasciculata*), paired with increased diel cycling peaking in the early night (clusters 2, 5 and 6).
363 This involved members of the night-time carbon fixating module of CAM such as PEPCK, PPCK
364 and MDH, enzymes involved in malate transport as V-ATPase and several glycolysis enzymes
365 such as glucose-6-phosphate isomerase, aldolase, Ppi-dependent phosphofructokinase (PFK) and
366 enolase (Fig. 6, S6). Enzymes of both soluble sugar transport (SUT2, ERD6, cluster 6) and starch
367 metabolism (Starch synthase I, α- and β-amylase and glucose-6-phosphate/phosphate translocator
368 (GPT), cluster 5 and 6) are overall up-regulated and show cycling expression curves in *T.*
369 *fasciculata* compared to *T. leiboldiana,* all showing highest activity in the late day. While the
370 increased night-time expression of Ppi-dependent PFK suggests a primary role of soluble sugars
371 as night-time source for PEP (Carnal and Black, 1989) in CAM *Tillandsia*, the simultaneously

372 cycling expression patterns in starch metabolic enzymes point also at transitory starch as a
373 potential source. A handful of CAM-related genes show increased expression in *T. leiboldiana*:
374 aluminium-activated malate transporter (ALMT), secondary vacuolar proton pump AVP1, which
375 also displays a phase shift peaking later in the night in *T. fasciculata*, and three circadian clock
376 regulators (LHY, GI and RVE1, cluster 1), which all show similar but reduced cycling patterns in
377 *T. fasciculata* compared to *T. leiboldiana*. We also see a phase shift in succinate dehydrogenase
378 peaking earlier in the night in *T fasciculata* versus *T. leiboldiana*'s early morning peak.
379 Overall, the majority of CAM-related DE gene expression profiles align with the view that
380 *T. fasciculata* is a constitutive CAM plant while *T. leiboldiana* performs a C3-like metabolism in
381 normal conditions, though showing signs of very early CAM evolution. The difference in
382 metabolism between both plants seems to be largely attained through regulatory rewiring of
383 functional enzymes.
384
385 **3.8. Genomic features of DE and CAM-related genes**
386
387 **3.8.1. Differentially expressed genes have more TE insertions**
388 To investigate whether transposable element (TE) activity and differential gene expression
389 are associated in *Tillandsia*, we tested whether intronic TE insertions are significantly enriched in
390 DE genes in both species. Both the presence of one or more TE insertions in a gene, as well as the
391 average number of TE insertions per gene is higher across all genes of *T. leiboldiana* compared to
392 the *T. fasciculata* gene set, which was expected given its larger proportion of repetitive content
393 (See Results section 3.2.).
394 While in both genomes, the proportion of DE genes with one or more TE insertions in
395 intronic regions is not significantly different to that of the full gene set, the average number of TE
396 insertions per gene is significantly higher in DE compared to non-DE genes (Table 1).
397 On the other hand, DE genes related to CAM, starch metabolism and glycolysis /
398 gluconeogenesis display on average reduced, but non-significant TE insertion rates in both
399 genomes (Table S8). However, the proportion of CAM-related DE genes with a TE insertion is
400 larger in *T. fasciculata* (41.2 %) than in *T. leiboldiana* (36.2 %), despite *T. leiboldiana*'s generally
401 elevated intronic TE insertion rate.
402

Table 1: Statistical test results on TE insertions in DE versus non-DE genes, and in DE genes previously described as underlying CAM, glycolysis or starch metabolism, versus all other genes.

Total number of genes with one or more TE insertions			
	All genes	DE genes	Chi-square p-value
In *T. fasciculata*	15844 (50 %)	473 (52 %)	0.2324
In *T. leiboldiana*	18251 (54.6 %)	472 (54.6 %)	0.9645
	All genes	DE CAM genes	Chi-square p-value
In *T. fasciculata*	15844 (50 %)	28 (41.2 %)	0.178
In *T. leiboldiana*	18251 (54.6 %)	21 (36.2 %)	0.008*
Average and (median) TE insertion counts per gene			
	in DE genes	in non-DE genes	Mann-Whitney U p-value
In *T. fasciculata*	3.66 (1)	2.88 (1)	0.0179
In *T. leiboldiana*	4.2 (1)	3.27 (1)	0.0159
	in CAM DE genes	in all other genes	Mann-Whitney U p-value
In *T. fasciculata*	2.618 (0)	2.906 (1)	0.375
In *T. leiboldiana*	2.862 (0)	3.291 (1)	0.083

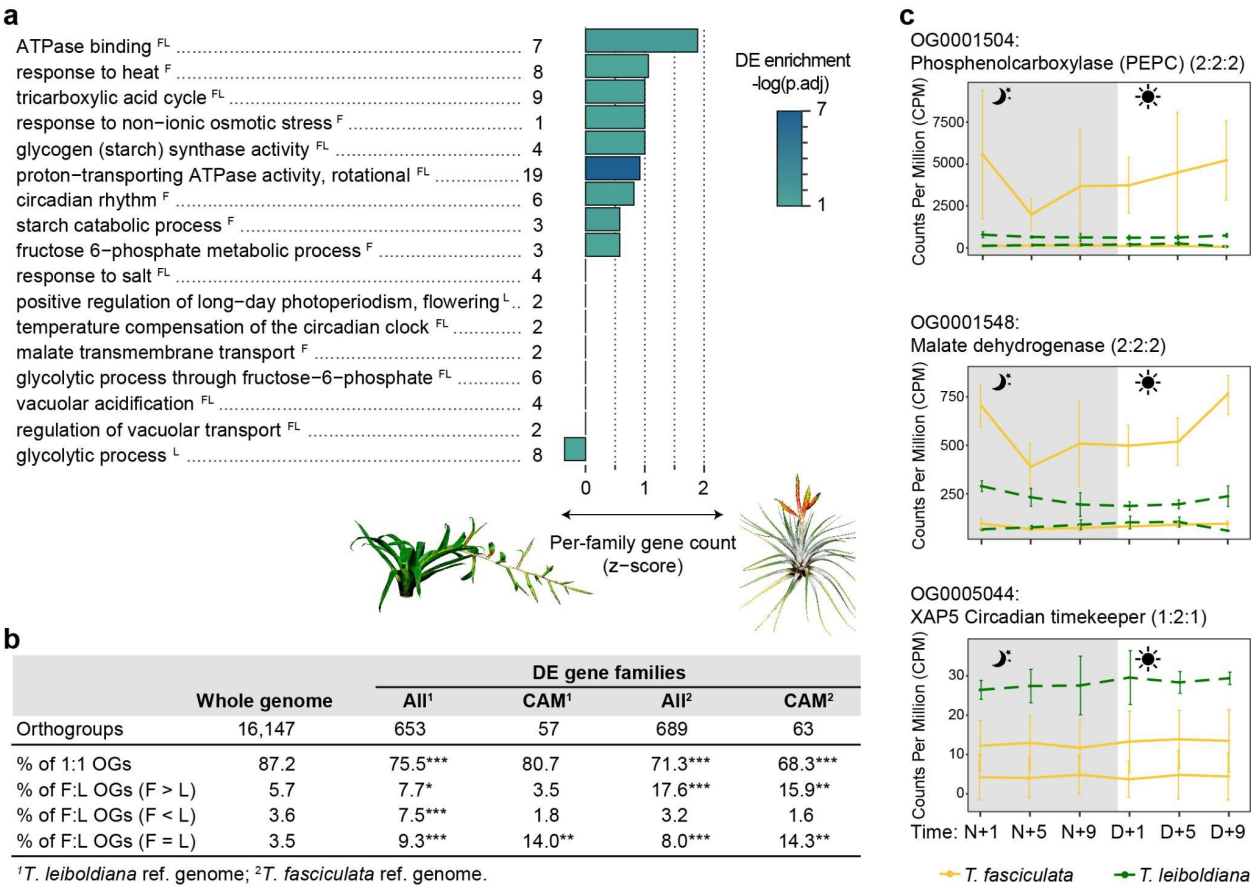
403

404 Nine DE genes related to CAM, starch and gluconeogenesis display more than twice the
405 number of TE insertions as the genome-wide average in *T. fasciculata*. This includes a V-type
406 proton ATPase subunit H (vacuolar transport and acidification), an aluminium-activated malate
407 transporter, a ABC transporter C family member 4 (stomatal opening and circadian rhythm) and a
408 mitochondrial isocitrate dehydrogenase subunit (Table S9), which all had more TE insertions than
409 their orthologs in *T. leiboldiana*. On the other hand, six DE genes of interest had elevated TE
410 insertion rates in *T. leiboldiana,* including a vacuolar acid invertase (sugar metabolism) and a sugar
411 transporter ERD6-like. Four of these genes showed high amounts of TE insertions in both
412 genomes, such as a glucose-1-phosphate adenylyltransferase subunit (starch synthesis), a V-type
413 proton ATPase subunit C, and circadian clock regulator GIGANTEA (Dalchau et al., 2011), which
414 also plays a role in stomatal opening (Ando et al., 2013).

415 **3.8.2. Differentially expressed genes belong more often to multi-copy**
416 **orthogroups**

417 To investigate the consequences of gene family evolution on gene expression, we tested
418 whether the proportion of multi-copy orthogroups underlying DE genes was significantly elevated
419 to that of the whole-genome set of orthogroups in both species (Fig. 5b, SI Note 10). The 907 DE
420 genes in *T. fasciculata* are found in 738 orthogroups (hereafter called DE orthogroups) containing
421 a total of 2,141 and 910 genes in *T. fasciculata* and *T. leiboldian*a, respectively. Genes from multi-
422 copy orthogroups are more likely to be differentially expressed: while multi-copy orthogroups
423 account for 24 % of all orthogroups in the genome, they represent 31 % of DE orthogroups. This
424 difference is primarily explained by a 3.2x larger proportion of multi-copy orthogroups with a
425 larger family size in *T. fasciculata* than in *T. leiboldiana* in the DE subset compared to the whole
426 genome (Chi-square $P = 1.59e^{-66}$).

427 Reciprocally, the DE analysis on the *T. leiboldiana* genome (SI Note 10) resulted in 836
428 DE genes belonging to 714 orthogroups, of which 489 overlap with the DE orthogroups resulting
429 from the analysis on the *T. fasciculata* genome. As in the analysis on the *T. fasciculata* genome,
430 we find that orthogroups with a larger family size in *T. leiboldiana* are enriched among DE genes.
431 Additionally, both analyses point at a significant enrichment for multi-copy orthogroups with equal
432 family sizes in both species, suggesting that also older duplications preceding the split of *T.*
433 *fasciculata* and *T. leiboldiana* play a role in day-night regulatory evolution. This highlights the
434 importance not only of novel, but also ancient variation in fuelling trait evolution in *Tillandsia*.



b

	Whole genome	DE gene families			
		All[1]	CAM[1]	All[2]	CAM[2]
Orthogroups	16,147	653	57	689	63
% of 1:1 OGs	87.2	75.5***	80.7	71.3***	68.3***
% of F:L OGs (F > L)	5.7	7.7*	3.5	17.6***	15.9**
% of F:L OGs (F < L)	3.6	7.5***	1.8	3.2	1.6
% of F:L OGs (F = L)	3.5	9.3***	14.0**	8.0***	14.3**

[1]*T. leiboldiana* ref. genome; [2]*T. fasciculata* ref. genome.

Figure 5: 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)*** *Composition of gene families by relative size between T. fasciculata (F) and T. leiboldiana (L) for three subsets of gene families (whole genome, DE and CAM-related DE). Species-specific orthogroups are not included in this analysis. A chi-square test of independence was applied to test the significance of composition changes in 2x2 contingency tables for each category when testing the entire DE gene subset. For CAM DE genes, the Fisher's exact test was applied. Significant p-values of both tests are reported as: *0.05-0.01, **0.01-0.0001, ***0.0001-0.* ***c)*** *Expression profiles in a 24-hour period of exemplary CAM-related gene families (PEPC, MDH and XCT) displayed at the*

451 *orthogroup level. For each gene copy and timepoint, the average read count (in Counts Per*
452 *Million), and the standard deviation across accessions are displayed. Read counts of each*
453 *ortholog are obtained by mapping conspecific accessions to their conspecific reference genome.*
454 *We show two families with older duplications preceding the split of* T. fasciculata *and* T.
455 leiboldiana *(PEPC and MDH) and one gene family with a recent duplication in* T. fasciculata
456 *(XCT).*
457
458 Multi-copy gene families are also enriched in a restricted subset of DE orthogroups related
459 to CAM, starch metabolism and gluconeogenesis (68 genes in 67 orthogroups), especially gene
460 families with equal copy number and with copy number expansion in *T. fasciculata* (Fig. 5b, Table
461 S5). Importantly, expanded gene families in *T. leiboldiana* are not significantly enriched in this
462 functional subset, showing that while the full set of DE orthogroups exhibit increased gene family
463 dynamics in both lineages, CAM-related gene family expansion is only associated with *T.*
464 *fasciculata* (constitutive CAM). This pattern is also reflected on the GO-term level, where enriched
465 CAM-related biological functions appear disproportionally associated with gene family expansion
466 in *T. fasciculata* (nine functions) than in *T. leiboldiana* (one function). Functions associated with
467 V-ATPase proton pumps especially tend to have larger gene family size in *T. fasciculata* than in
468 *T. leiboldiana* (ATPase binding, proton-transporting ATPase activity).
469 CAM-related expanded gene families often show one highly expressed copy that performs
470 diel cycling, while the other copies are lowly or not expressed (e.g. SDH, glyceraldehyde-3-
471 phosphate dehydrogenase (GAPDH), V-ATPase subunit H, SI Note 11), however, several gene
472 families show diel and/or elevated expression in *T. fasciculata* in two or more gene copies (starch
473 synthase, Ppi-dependent PFK and enolase, SI Note 11). Both copies of XCT are expressed in *T.*
474 *fasciculata,* though showing no diel cycling or increased expression compared to *T. leiboldiana*
475 (Fig. 5c), making its role in CAM photosynthesis unclear. V-ATPase subunit H has eight copies
476 in *T. fasciculata* and three in *T. leiboldiana*. While in both species only one copy is highly
477 expressed (with diel cycling peaking at night in *T. fasciculata*), the copies implemented for
478 elevated expression in either species are not each other's homologs (SI Note 11), suggesting that
479 different copies are recruited for the distinct photosynthetic phenotypes of these species. The other
480 copies are lowly or not expressed. A candidate gene family with larger gene family size in *T.*
481 *leiboldiana* is a probable aquaporin PIP2-6 (OG0005047, Fig. S8), which is involved in water

482 regulation and follows a diel pattern in pineapple (*A. comosus*) (Zhu and Ming, 2019). While lowly
483 expressed in *T. fasciculata*, one of the two gene copies shows strong diel expression in *T.*
484 *leiboldiana*, with highest expression in the early night. This is another indication that an early,
485 latent CAM cycle may be present in *T. leiboldiana*.

486 CAM-related gene families with duplications preceding the divergence of *T. leiboldiana*
487 and *T. fasciculata* include a malate dehydrogenase (MDH) with two copies in both species, where
488 only one copy is highly expressed and cycling in *T. fasciculata,* and the core CAM enzyme
489 Phosphoenolpyruvate carboxylase (PEPC), which shares an ancient duplication among monocots
490 (Deng et al., 2016) (Fig. 5c). The widely varying expression patterns of multicopy DE CAM-
491 related families suggest a variety of mechanisms possibly contributing to CAM regulatory
492 evolution: dosage changes ("more of the same"), subfunctionalization and neofunctionalization.
493

494 **4. Discussion**

495 The sources of variation fuelling trait evolution in rapid radiations have been a long-
496 standing topic in evolutionary biology (Simpson, 1953), and our understanding of how complex
497 traits such as CAM evolve repeatedly is still incomplete. By showcasing a broad range of
498 photosynthetic phenotypes and repeated evolution of constitutive CAM, the subgenus *Tillandsia*
499 provides an excellent opportunity to study CAM evolution. The recent divergence between
500 members of *Tillandsia* allows to pinpoint the necessary evolutionary changes to evolve a
501 constitutive CAM phenotype. By integrating comparative genomics using *de novo* assemblies and
502 in-depth gene expression analyses of two closely related *Tillandsia* species representing one of the
503 most distinct photosynthetic phenotypes within the clade, we found support for regulatory
504 evolution and gene family expansion as major features of CAM evolution (Fig. 6).

505 Our metabolic analyses of night-time malate accumulation and gene expression analyses
506 provide a much more detailed understanding of the photosynthetic phenotypes present in the
507 subgenus than the previously reported carbon isotope measurements, which do not accurately
508 reflect all stages of the CAM continuum. For example, weak CAM phenotypes have been reported
509 in Bromeliaceae for species with carbon isotope ratios falling in the C3 range (-26.5) (Pierce et al.,
510 2002), indicating that the photosynthetic metabolism of *T. leiboldiana* could be different from C3
511 *sensu stricto* as its carbon isotope measurements suggest. Our timewise malate measurements
512 show distinct fluxes for *T. fasciculata* and *T. leiboldiana*, with constitutive accumulation of night-

513 time malate in *T. fasciculata* which is absent in *T. leiboldiana*. On the other hand, *T. leiboldiana*
514 does display CAM-like temporal expression profiles for certain enzymes, such as PEPC kinase
515 and Aquaporin PIP-6. It has been suggested that repeated evolution of CAM (similar to C4) may
516 be facilitated in lineages where C3 species already display increased or CAM-like expression of
517 core genes (Heyduk et al., 2019a, 2019b; Kajala et al., 2012). However, while a CAM cycle is
518 seemingly not being expressed in *T. leiboldiana* under normal circumstances, we can not exclude
519 that a latent CAM cycle could become activated under certain conditions, for example under
520 drought stress. In that case, *T. leiboldiana* would rather be at the very early stages of CAM
521 evolution than a "pre-adapted" C3 plant (De La Harpe et al., 2020). We hope that future studies
522 will study the drought response of *T. leiboldiana* to better understand its exact position in the CAM
523 continuum.

524 On the other hand, even if *T. leiboldiana* and potentially all subgenus *Tillandsia* species
525 previously labelled as C3 represent in fact the early stages of the CAM continuum, our analyses
526 show widely distinct photosynthetic phenotypes within the radiation which required divergent
527 evolution. Therefore, while this study may underestimate the total number of evolutionary changes
528 needed to establish constitutive CAM from C3 *sensu stricto*, it highlights the evolutionary drivers
529 underlying the least understood section of the CAM continuum: from early CAM to constitutive
530 CAM.

531 Differences between the two genomes related to CAM evolution can be primarily found
532 on the regulatory level, with CAM-related genes showing temporal differential expression between
533 species across a 24-hour period. These reveal a complex web of underlying expression changes,
534 as they are distributed over all inferred co-expression clusters (Fig. S7), emphasising the lack of a
535 master regulator and a clear overall direction of expression changes (Wickell et al., 2021; Heyduk
536 et al., 2022).

537 Gene family expansion has been previously observed in CAM lineages (Cai et al., 2015;
538 Silvera et al., 2010) and suggested as a driver of CAM evolution (Silvera et al., 2014). Strikingly,
539 both the total subset of differentially expressed genes and a more stringent group of CAM-related
540 DE genes was significantly enriched for gene families that have expanded in *T. fasciculata*
541 (constitutive CAM). CAM-related functions that were enriched in DE genes show a
542 disproportionate bias towards gene family expansion in *T. fasciculata* (circadian rhythm, vacuolar
543 ATPase activity, tricarboxylic acid cycle and starch metabolism) compared to *T. leiboldiana*

544 (glycolysis) (Fig. 5a). Gene duplications, preceding the split of *T. fasciculata* and *T. leiboldiana*
545 are also significantly associated with day-night expression differences in CAM-related genes (Fig.
546 5b), suggesting that older, already existing gene duplications may also be recruited in CAM
547 evolution, alongside novel duplications.

548 The expression curves of DE multicopy gene families with a potential link to CAM reveal
549 a multitude of expression behaviours (e.g. Fig. 5c), which supports that complex regulatory
550 evolution on the transcriptional level underlies CAM evolution. Our findings suggest that gene
551 family evolution played a significant role in modulating regulatory changes underlying the
552 evolution towards constitutive CAM in *Tillandsia*. As gene family expansion leads to increased
553 redundancy, selection on individual gene copies and their expression relaxes, facilitating the
554 assimilation of a constitutively expressed CAM expression profile (Ohno, 1970).

555 Another potential driver of trait evolution is TE insertion, though its role in CAM evolution
556 in *Tillandsia* remains unclear. TE insertions are overall less common in CAM-related DE genes
557 compared to all genes in both genomes, suggesting a selection pressure against TE insertions
558 around these genes. However, the proportion of CAM-related DE genes with TE insertions is
559 greater in *T. fasciculata* than in *T. leiboldiana,* despite the overall higher genic TE insertion rate
560 in *T. leiboldiana*. This suggests that the pressure to maintain CAM-related genes TE-free is
561 reduced in the constitutive CAM lineage relative to *T. leiboldiana*. We detect nine and six CAM-
562 related DE genes with more than twice the whole-genome average TE insertion count in *T.*
563 *fasciculata* and *T. leiboldiana* respectively, of which four are single-copy orthologs shared
564 between species. The high degree of sharedness of TE-rich DE genes between species rather
565 suggests that TE insertions are not a major driver of CAM-related gene expression changes. In
566 fact, genes with exceptionally high insertion rates in *T. fasciculata* tend to show reduced
567 expression (ALMT, V-ATPase subunit H copy Tfasc_v1.24696, ABCC4, SI Note 11). Instead, the
568 larger proportion of CAM-related DE genes with one or more TE insertions in *T. fasciculata* may
569 be a consequence of higher rates of gene family expansion and eventual pseudogenization of
570 redundant copies.

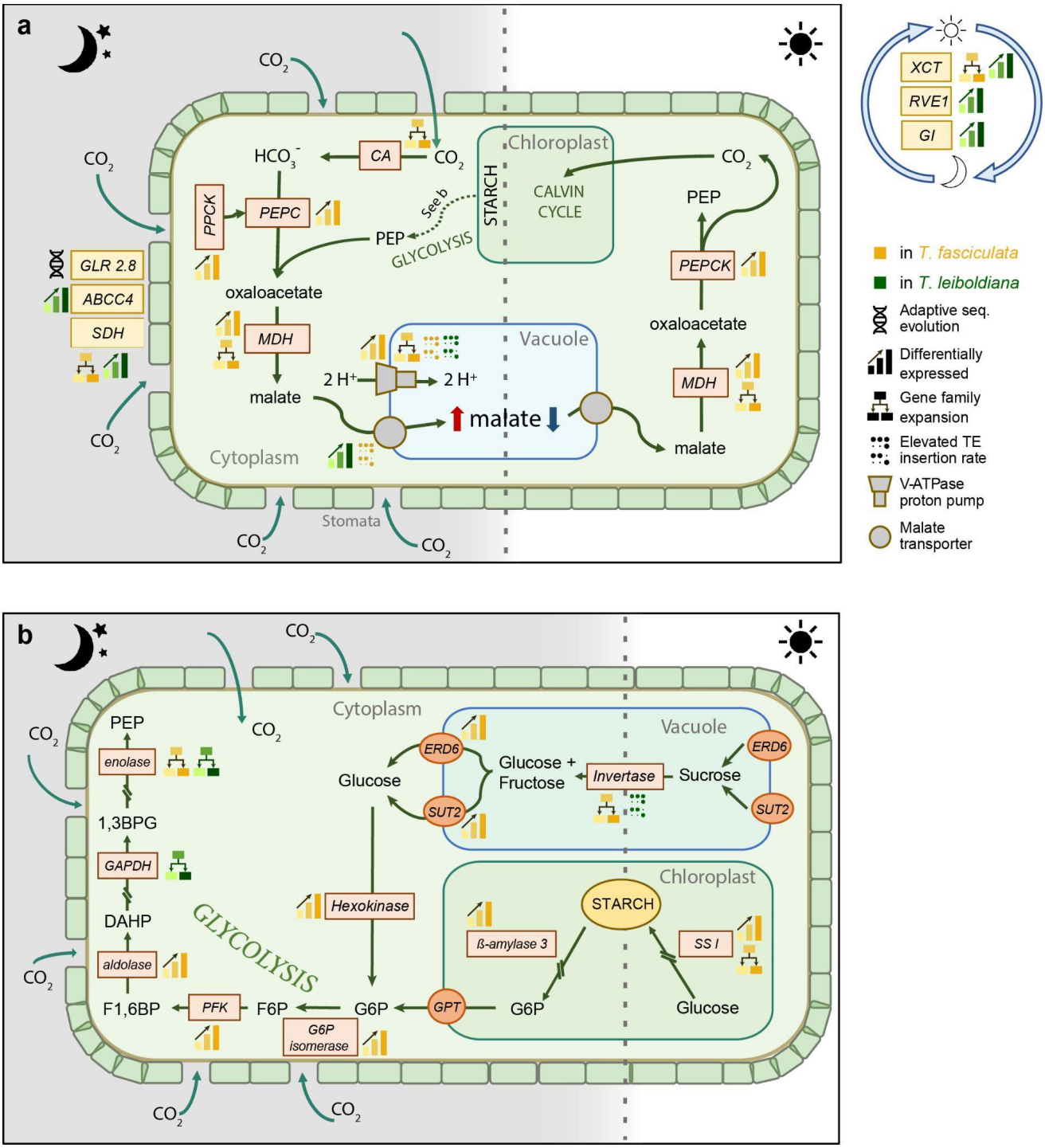


Figure 6: *Pathway of Crassulacean Acid Metabolism (CAM), highlighting underlying genes detected in this study as differentially expressed, with gene family expansion, with signature of adaptive sequence evolution or elevated TE insertion counts. The colour of the process indicates the species showing diel or elevated expression, increased gene family size or increased TE insertion rate. Enzymes are shown in coloured boxes, while pathway products are shown in bold*

577 *outside boxes. Enzymatic members of CAM metabolism pathways are shown in orange boxes,*
578 *while stomatal and circadian regulators are highlighted in yellow. Stomatal regulators are shown*
579 *on the left outside the cell and circadian regulators on the right.* ***a)*** *CO_2 is absorbed at night and*
580 *first converted to HCO_3^- by carbonic anhydrase (CA). Then, it is converted to malate by*
581 *carboxylating phosphoenol pyruvate (PEP), a key component of the glycolysis. In a first step, PEP*
582 *carboxylase (PEPC) converts PEP to oxaloacetate, after being activated by PEPC kinase (PPCK).*
583 *In a second step, malate dehydrogenase (MDH) converts oxaloacetate to malate. Malate is then*
584 *transported into the vacuole by two possible transporters, either a tonoplast dicarboxylate*
585 *transporter or an aluminium-activated malate transporter, which are assisted by V-ATPase proton*
586 *pumps. During the day, the accumulated malate becomes the main source of CO_2 for*
587 *photosynthesis. This allows the stomata to remain closed, which greatly enhances the water use*
588 *efficiency of the plant. Malate is again transported out of the vacuole and reconverted to*
589 *oxaloacetate by MDH, and then decarboxylated to PEP and CO_2 by PEP carboxykinase (PEPCK).*
590 *The CO_2 will cycle through the Calvin cycle and generate sugars. Abbreviations: GLR2.8 –*
591 *Glutamate receptor 2.8, ABCC4 - ABC transporter C 4, SDH – Succinate Dehydrogenase, XCT –*
592 *XAP5 CIRCADIAN TIMEKEEPER, GI – protein GIGANTEA, RVE1 – REVEILLE 1.* ***b)*** *Glycolysis,*
593 *transitory starch and sugar metabolism are tightly linked with the core CAM pathway as providers*
594 *of starting materials such as PEP. During the day, CAM plants can store starch in the chloroplast*
595 *and hexoses in the vacuole. In Bromeliaceae, the relative importance of soluble sugars versus*
596 *starch as a source for PEP is variable across species*[56]*. At night, the stored starch and/or sugars*
597 *are transported into the cytoplasm, converted to glucose or fructose and broken down via the*
598 *glycolysis to PEP. Abbreviations: G6P - glucose-6-phosphate; PGM – phosphoglucomutase; GPT*
599 *- glucose-6-phosphate/phosphate translocator; PFK – phosphofructokinase; GAPDH -*
600 *Glyceraldehyde-3-phosphate dehydrogenase; F6P – fructose-6-phosphate; F1,6BP – fructose-*
601 *1,6-biphosphate; SSI – starch synthase I; 1,3BPG - 1,3-Bisphosphoglyceric acid; PEP –*
602 *Phosphoenol pyruvate; SUT2 – Sucrose transporter 2; ERD6 - EARLY RESPONSE TO*
603 *DEHYDRATION 6. For a detailed description and accompanying per-gene expression profiles,*
604 *see SI Note 11.*
605
606 Candidate genes under positive selection underlie a broad array of functions, but had no
607 immediate link to CAM photosynthesis. The lack of overlap between regulatory and adaptive

608 sequence evolution is in line with previously proposed mechanisms of CAM evolution largely
609 relying on regulatory changes in other systems (Deng et al., 2016). A small number of cases of
610 convergent and adaptive sequence evolution between distantly related CAM and C3 species have
611 been described (Yang et al., 2017), though no overlap was found between convergence in
612 expression and sequence evolution. Our study suggests that while on larger evolutionary scales
613 adaptive sequence evolution may play an important role, distinct photosynthetic phenotypes
614 between closely related species may be achieved primarily with gene expression changes (but see
615 SI Note 14), or may be especially relevant between the transition of C3 *sensu stricto* to the CAM
616 continuum.

617 Though we observe a karyotype difference of six chromosome pairs between *T. fasciculata*
618 and *T. leiboldiana* and we identified one fusion in the *T. leiboldiana* assembly, along with two
619 reciprocal translocations, we did not find detectable consequences of large-scale rearrangements
620 for either functional diversification or adaptation in *Tillandsia*, unlike other studies (Davey et al.,
621 2016; Cicconardi et al., 2021) (Fig. 4b, S4, S6, SI Note 12, but see SI Note 5 and 14). The role of
622 these large-scale rearrangements in the evolution of species barriers and/or the evolution of other
623 key innovation traits in *Tillandsia* will be an exciting area of study for future research.

624 The two *de novo* assemblies presented in this study are the first tillandsioid and fourth
625 bromeliad genomes published so far. Despite both genomes exhibiting one of the highest TE
626 contents reported to date for a non-polyploid plant species (Pedro et al., 2021), the joint use of
627 long-read sequencing and chromatin conformation capture successfully led to highly contiguous
628 assemblies with high-quality gene sets (SI Note 13). Along with other recently developed
629 resources for Bromeliaceae (Yardeni et al., 2021; Liu et al., 2021), these genomes will be crucial
630 in future investigations of this highly diverse and species-rich plant family, and in further studies
631 of CAM evolution.

632 Our analyses reveal genomic changes of all scales between two members of an adaptive
633 radiation representing a recent shift to constitutive CAM. However, in this recent shift between
634 closely related species, differences in photosynthetic metabolism are brought about largely by
635 temporal expression changes enabled by both existing and *de novo* gene duplication, rather than
636 adaptive sequence evolution of existing gene copies, which may play a role at later stages of
637 divergence (but see SI Note 14). Large scale rearrangements observed so far seem unlinked from
638 functional divergence, more likely affecting reproductive isolation (de Vos et al., 2020; Faria and

639 Navarro, 2010), and need further study. Our findings support an important role for gene family
640 expansion in generating novel variation that fuels the evolution of the CAM continuum.
641
642 **5. Materials & Methods**
643
644 **5.1. Flow cytometry and cytogenetic experiments**
645
646 **5.1.1. Genome size measurements**

647 Approximately 25 mg of fresh leaf material was co-chopped according to the chopping method of
648 (Galbraith et al., 1983) together with an appropriate reference standard (*Solanum pseudocapsicum*,
649 1.295 pg/1C) (Temsch, 2010; Temsch et al., 2022) in Otto´s I buffer (Otto et al., 1981). After
650 filtration through a 30 µm nylon mesh (Saatilene Hitech, Sericol GmbH, Germany) and incubation
651 with RNase A (0.15mg/ml, Sigma-Aldrich, USA) at 37°C, Otto´s II buffer (Otto et al., 1981)
652 including propidium iodide (PI, 50mg/L, AppliChem, Germany) was added. Staining took place
653 in the refrigerator for at least one hour or up to over-night. Measurement was conducted on a
654 CyFlow ML or a CyFlow Space flow cytometer (Partec/Sysmex, Germany) both equipped with a
655 green laser (532nm, 100mW, Cobolt AB, Sweden). The fluorescence intensity (FI) of 10,000
656 particles were measured per preparation and the 1C-value calculation for each sample followed
657 the equation: $1C_{Obj} = (\text{FI peak mean}_{G1\ Obj} / \text{FI peak mean}_{G1\ Std}) \times 1C_{Std}$
658
659 **5.1.2. Karyotyping**

660 Actively growing root meristems of genome assembly accessions (see Table S1) were harvested
661 and pretreated with 8-hydroxyquinoline for 2 hrs at room temperature and 2 hrs at 4°C. The roots
662 were then fixed in Carnoy's fixative (3 : 1 ethanol : glacial acetic acid) for 24 hours at room
663 temperature and stored –20 °C until use. Chromosome preparations were made after enzymatic
664 digestion of fixed root meristems as described in (Jang and Weiss-Schneeweiss, 2015).
665 Chromosomes and nuclei were stained with 2 ng/µl DAPI (4',6-diamidino-2-2phenylindole) in
666 Vectashield antifade medium (Vector Laboratories, Burlingame, CA, USA). Preparations were
667 analyzed with an Axiolmager M2 epifluorescent microscope (Carl Zeiss) and images were
668 captured with a CCD camera using AxioVision 4.8 software (Carl Zeiss). Chromosome number

669 was established based on analyses of several preparations and at least five intact chromosome
670 spreads. Selected images were contrasted using Corel PhotoPaint X8 with only those functions
671 that applied equally to all pixels of the image and were then used to prepare karyotypes.
672
673 **5.2. Genome Assembly**
674
675 **5.2.1. Plant material selection and sequencing**
676 Genome assemblies were constructed from the plant material of one accession per species
677 (see Table S1). The accessions were placed in a dark room for a week to minimise chloroplast
678 activity and recruitment, after which the youngest leaves were collected and flash frozen with
679 liquid nitrogen. High molecular weight extraction for ultra-long reads, SMRTbell library
680 preparation and PacBio Sequel sequencing was performed by Dovetail Genomics™ (now Cantata
681 Bio). Dovetail Genomics™ also prepared Chicago (Putnam et al., 2016) and Hi-C (Lieberman-
682 Aiden et al., 2009) libraries which were sequenced as paired-end 150bp reads on an Illumina HiSeq
683 X instrument. Additional DNA libraries were prepared for polishing purposes using Illumina's
684 TruSeq PCR-free kit, which were sequenced on a HiSeq2500 as paired-end 125 bp reads at the
685 Vienna BioCenter Core Facilities (VBCF), Austria.
686 RNA-seq data of *T. fasciculata* used for gene annotation was sampled, sequenced and
687 analyzed in De La Harpe et al. 2020 under SRA BioProject PRJNA649109. For gene annotation
688 of *T. leiboldiana*, we made use of RNA-seq data obtained during a similar experiment, where
689 plants were kept under greenhouse conditions and sampled every 12 hours in a 24-hour cycle.
690 Importantly, while the *T. fasciculata* RNA-seq dataset contained three different genotypes, only
691 clonal accessions were used in the *T. leiboldiana* experiment. For *T. leiboldiana*, total RNA was
692 extracted using a QIAGEN RNeasy® Mini Kit, and poly-A capture was performed at the Vienna
693 Biocenter Core Facilities (VBCF) using a NEBNext kit to produce a stranded mRNA library. This
694 library was sequenced on a NovaSeq SP as 150 bp paired end reads.
695 For both species, sequencing data from different time points and accessions were merged
696 into one file for the purpose of gene annotation. Before mapping, the data was quality-trimmed
697 using AdapterRemoval (Schubert et al., 2016) with default options (--trimns, --trimqualities). We
698 allowed for overlapping pairs to be collapsed into longer reads.
699

700 **5.2.2. First draft assembly and polishing**

701 We constructed a draft assembly using long-read PacBio data with CANU v1.8 (Koren et
702 al., 2017) for both species. To mitigate the effects of a relatively low average PacBio coverage
703 (33x), we ran two rounds of read error correction with high sensitivity settings
704 (corMhapSensitivity=high corMinCoverage=0 corOutCoverage=200) for *T. fasciculata.*
705 Additionally, we applied high heterozygosity (correctedErrorRate=0.105) settings, since K-mer
706 based analyses pointed at an elevated heterozygosity in this species (See SI Note 2), and memory
707 optimization settings (corMhapFilterThreshold=0.0000000002 corMhapOptions=" --repeat-idf-
708 scale 50" mhapMemory=60g mhapBlockSize=500).

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

712 To minimise the retention of heterozygous sequences as haplotigs in *T. fasciculata* (see SI
713 Note 2), we reassigned allelic contigs using the pipeline Purge Haplotigs (Roach et al., 2018). Raw
714 PacBio data was mapped to the draft assembly produced in the previous step with minimap2 (Li,
715 2018), before using the Purge Haplotigs pipeline.

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

720

721 **5.2.3. Scaffolding and final polishing**

722 Scaffolding of both assemblies was performed in-house by Dovetail GenomicsTM using
723 Chicago and Hi-C data and the HiRise scaffolding pipeline (Putnam et al., 2016). To increase base
724 quality and correct indel errors, we ran additional rounds of polishing with high-coverage Illumina
725 data (See above, section 2.1.) using Pilon v1.22 (Walker et al., 2014). The Illumina data was
726 aligned to the scaffolded assembly using BWA-MEM (Li, 2013), and then Pilon was run on these
727 alignments. We evaluated the result of each round using BUSCO v.3 (Waterhouse et al., 2018)
728 with the Liliopsida odb9 library and proceeded with the best version. For *T. fasciculata*, polishing
729 was performed twice, fixing SNPs and indels. We did not fix small structural variation in this
730 genome due to the relatively low coverage (35x) of Illumina data. For *T. leiboldiana*, one round

731 of polishing on all fixes (SNPs, indels and small structural variants) resulted in the highest BUSCO
732 scores.
733
734 ## 5.3. Annotation
735
736 ### 5.3.1. TE annotation and repeat masking
737 *De novo* TE annotation of both genome assemblies was performed with EDTA v.1.8.5 (Ou
738 et al., 2019) with option –sensitive. To filter out genes that have been wrongly assigned as TEs, *A.*
739 *comosus* (pineapple) coding sequences (Ming et al., 2015) were used in the final steps of EDTA.
740 Using the species-specific TE library obtained from EDTA, we masked both genomes
741 using RepeatMasker v.4.0.7 (Smit et al., 2013-2015). Importantly, we excluded all TE annotations
742 marked as "unknown" for masking to prevent potentially genic regions flagged as TEs to be
743 masked during annotation. The search engine was set to NCBI (-e ncbi) and simple and low-
744 complexity repeats were left unmasked (-nolow). We produced both hard-masked and soft-masked
745 (--xsmall) genomes.
746
747 ### 5.3.2. Transcriptome assembly
748 We constructed transcriptome assemblies for both species using Trinity *de novo* assembler
749 v.2.4.8. (Grabherr et al., 2011) using default parameters starting from the raw mRNA-seq data.
750 These were evaluated with BUSCO. Additionally, before feeding the transcriptome assemblies to
751 the gene annotation pipeline, we ran a round of masking of interspersed repeats to avoid an
752 overestimation of gene models due to the presence of active transposases in the RNA-seq data.
753
754 ### 5.3.3. Gene prediction and functional annotation
755 Gene models were constructed using a combination of BRAKER v.2.1.5 (Hoff et al., 2019)
756 and MAKER2 v.2.31.11 (Campbell et al., 2014). Starting with BRAKER, we obtained splicing
757 information from RNA-seq alignments to the masked genome as extrinsic evidence using the
758 *bam2hints* script of AUGUSTUS v.3.3.3 (Stanke et al., 2008). A second source of extrinsic
759 evidence for BRAKER were single-copy protein sequences predicted by BUSCO when run on the
760 masked genomes in genome mode with option --long. Predictions made by BRAKER were
761 evaluated with BUSCO and with RNA-seq alignments.

762 Subsequently, we built our final gene predictions using MAKER2. As evidence, we used
763 (1) the gene models predicted by BRAKER, (2) a transcriptome assembly of each respective
764 species (see above section 3.2.), (3) a protein sequence database containing proteins of *Ananas*
765 *comosus comosus* (F153) (Ming et al., 2015) and *Ananas comosus bracteatus* (CB5) (Chen et al.,
766 2019) and manually curated swissprot proteins from monocot species (64,748 sequences in total)
767 and (4) a GFF file of complex repeats obtained from the masked genome (see above section 3.1.)
768 and an extended repeat library containing both the EDTA-produced *Tillandsia*-specific repeats and
769 the monocot repeat library from RepBase (7,857 sequences in total). By only providing masking
770 information of complex repeats and setting the model organism to "simple" in the repeat masking
771 options, hardmaksing in MAKER2 was limited to complex repeats while simple repeats were soft-
772 masked, which makes these available for gene prediction. MAKER2 predicts genes both *ab initio*
773 and based on the given evidence using AUGUSTUS.
774 We evaluated the resulting set of predicted gene models by mapping the RNA-seq data
775 (section 2.1.) back to both the transcript and full gene model sequences and running BUSCO in
776 transcriptome mode. We also calculated the proportion of masked content in these gene models to
777 ascertain that MAKER2 hadn't predicted TEs as genes. A second run of MAKER, which included
778 training AUGUSTUS based on the predicted models from the first round, resulted in lower
779 BUSCO scores and was not further used. We functionally annotated the final set of gene models
780 in Blast2Go v.5.2.5 (Götz et al., 2008) using the Viridiplantae database.
781
782 **5.4. Inferring gene orthology**
783 Orthology between gene models of *T. fasciculata*, *T. leiboldiana* and *Ananas comosus* was
784 inferred using Orthofinder v.2.4.0 (Emms and Kelly, 2019). Protein sequences produced by
785 MAKER2 of inferred gene models were used for *T. fasciculata* and *T. leiboldiana*. For *A. comosus*,
786 the publicly available gene models of F153 were used. The full Orthofinder pipeline was run
787 without additional settings. Counts per orthogroup and the individual genes belonging to each
788 orthogroup were extracted from the output file Phylogenetic_Hierarchical_Orthogroups/N0.tsv.
789 Orthofinder was run a second time on gene models present only on main contigs (See
790 Results). For each gene model, the longest isoform was selected, and gene models with protein
791 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.

5.5. Gene model assessment and curation

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

5.6. Analyzing TE class abundances

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

5.7. Spatial distribution of repetitive, genic and GC content

The spatial distribution of genes, transposable elements and GC content as shown in Fig. 3a, 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 (Gu et al., 2014). 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

822 for significance using the Kendall Rank Correlation Coefficient, after testing for normality using
823 the Shapiro-Wilk test.
824 Repetitive, GC and gene content as shown in Fig. 3b was estimated directly from the soft-
825 masked reference genomes using 100 kb non-overlapping sliding windows as described in (Leroy
826 et al., 2021). TE content corresponds to the proportion of soft-masked positions per window, using
827 the curated TE library (see above, section 6.) as a basis for soft-masking in RepeatMasker. As
828 compared to the version of Leroy et al. 2021, this script was modified to estimate GC content in
829 repetitive regions only. In addition to this, we estimated the genic fraction by considering the total
830 number of genomic positions falling in genes based on the GFF files (feature = "gene") divided
831 by the size of the window (100 kb). This estimate was derived for the same window boundaries as
832 used for GC and TE content to be able to compare all statistics. The relative per-window proportion
833 of genic bases corresponding to non-robust genes (see above, section 5) was also estimated by
834 dividing the number of non-robust gene positions with the total number of gene positions.
835
836 **5.8. Synteny between *T. fasciculata* and *T. leiboldiana***
837 Synteny was inferred with GENESPACE (Lovell et al., 2022), using orthology information
838 obtained with Orthofinder of the gene models from *A. comosus*, *T. fasciculata* and *T. leiboldiana*.
839 This provided a first, visual graphical to detect large-scale rearrangements. We used GENESPACE
840 with default parameters, except that we generated the syntenic map (riparian plot) using
841 minGenes2plot=200. Other methods have also been used to confirm the chromosomal
842 rearrangements and to identify the genomic breakpoints more precisely (see SI Note 5).
843
844 **5.9. Gene family evolution**
845 Gene family counts were corrected for multi-copy orthogroups due to unusual coverage
846 distribution, especially in *T. fasciculata* (see SI Note 6). The distribution of gene counts per
847 multicopy orthogroup was compared between *T. fasciculata* and *T. leiboldiana* with a non-
848 parametric test (Mann-Whitney U). Using the log-ratio of per-species gene count, we investigated
849 which gene families experienced large changes in gene count compared to the background (SI
850 Note 7).
851 Functional characterization of multicopy families was done with a GO term enrichment
852 analysis of the underlying genes using the Fisher's exact test in TopGo (Alexa and Rahnenführer,

853 2009). Enrichment analyses were done on all genes belonging to multicopy orthogroups, on a
854 subset of genes belonging to families that are larger in *T. fasciculata* and on a subset of genes
855 belonging to families that are larger in *T. leiboldiana*. The top 100 significantly enriched GO terms
856 were then evaluated. GO terms putatively associated with key innovation traits were used to list
857 multicopy gene families of interest.

858 Additionally, we searched for specific genes that are known to underlie CAM evolution in
859 these multi-copy gene families. The IDs of candidate pineapple genes for CAM were obtained
860 from (Yardeni et al., 2021), who compiled extensive lists of genes from a diverse set of studies.
861 For CAM, we considered all genes listed in Table S1 in this study under the categories
862 "Differentially expressed in CAM / C3 experiment" (186 genes) (De La Harpe et al., 2020),
863 "Positive selection in CAM / C3 shifts" (22) (De La Harpe et al., 2020), Gene families associated
864 with CAM/C3" (79) (De La Harpe et al., 2020), "CAM-related *A. comosus*" (29) (Ming et al.,
865 2015), "stomatal function" (48) (Christin et al., 2014), "aquaporin regulation" (24) (Vera-Estrella
866 et al., 2012), "drought resistance" (61) (Xiao et al., 2007), "circadian metabolism" (47) (Wai et al.,
867 2017), "malate transferase" (28) (Cosentino et al., 2013) and "circadian clock" (3) (McClung,
868 2006), resulting in a total of 527 genes. A separate list was made for gluconeogenesis and starch
869 metabolism genes (288 genes) (Cushman et al., 2008). After obtaining these lists of pineapple gene
870 IDs, we searched for their orthologs in *T. fasciculata* and *T. leiboldiana*, and investigated their
871 presence in multi-copy gene families.

872

873 **5.10. d_N/d_S analysis**

874

875 **5.10.1. On single-copy orthologous pairs**

876 One-to-one orthologous genes were subjected to a test of positive selection using the non-
877 synonymous to synonymous substitution ratio ($\omega = d_N/d_S$). Gene pairs where both genes were
878 incomplete (missing start and/or stop codon) or where the difference in total length was more than
879 20 % of the length of either gene were removed to avoid misalignments. We performed codon-
880 aware alignments using the alignSequences program from MACSE v.2.05 (Ranwez et al., 2018)
881 with options -local_realign_init 1 -local_realign_dec 1 for optimization. Pairwise d_N/d_S ratios were
882 estimated with the codeML function of PAML v.4.9. (Yang, 2007). Using a single-ratio model
883 across sites and branches (Nssites = 0, model = 0), we tested for a fixed $\omega = 1$ as null hypothesis,

884 against an unfixed ω as the alternative hypothesis. Automization of codeML was achieved with a
885 modified script from AlignmentProcessor
886 (https://github.com/WilsonSayresLab/AlignmentProcessor/). The results of codeML under both
887 the null and alternative model were compiled and significance of the result was calculated with the
888 likelihood-ratio test (Wong et al., 2004). Multiple-testing correction was applied with the
889 Benjamini-Hochberg method and an FDR threshold of 0.05. Orthologous gene pairs with a d_N/d_S
890 ratio larger than one and an adjusted p-value under 0.05 were considered candidate genes under
891 divergent selection.

892 The d_N/d_S values of all orthologous gene pairs with five or more variant sites in the MACSE
893 alignment were used to obtain per-scaffold distributions of d_N/d_S values in both genomes. We
894 visualized d_N/d_S distributions of all main scaffolds in both assemblies with boxplots and used
895 density plots to visualize the d_N/d_S distribution in rearranged chromosomes compared to all non-
896 rearranged chromosomes. To test whether these distributions were significantly different, we ran
897 a non-parametric test (Mann-Whitney U) between the distribution of each single rearranged
898 chromosome and that of all non-rearranged chromosomes in each assembly.

899

900 #### 5.10.2. On duplicated orthogroups

901 We also performed tests of selection using d_N/d_S on all orthogroups that were consisted of
902 a single gene in *A. comosus* and a duplicated gene in either T. *leiboldiana* (1:1:2), or *T. fasciculata*
903 (1:2:1). Only orthogroups that maintained this conformation after size correction (SI Note 6) were
904 used in this analysis. Pairwise alignments were performed between the ortholog of one species and
905 either paralog of the other species using MACSE. Then, ω was estimated in the same way as
906 mentioned above.

907

908 ### 5.11. RNA-seq experiment capturing photosynthetic phenotypes and expression

909

910 #### 5.11.1. Experiment set-up and sampling

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

915 by fluorescent lamps with a spectrum ranging from 400 to 700 nm. The light intensity was set at
916 124 umol/m2s. The plants acclimated to these conditions for 4 weeks prior to sampling, during
917 which they were watered every second day.
918 Leaf material from each plant was sampled every 4 hours in a 24-hour cycle starting one
919 hour after lights went off. One leaf was pulled out of the base at each time-point without cutting.
920 The base and tip of the leaf were then removed, and the middle of the leaf immediately placed in
921 liquid nitrogen, then stored at -80 °C.
922
923 **5.11.2. Targeted metabolite analyses**
924 To corroborate the photosynthetic phenotypes of *T. fasciculata* and *T. leiboldiana*, we
925 measured malate abundances in the leaf throughout a 24-hour cycle. An approximate amount of
926 20 mg of frozen leaf material collected at six timepoints during the above-mentioned experiment
927 was collected and ground to a powder with a TissueLyser and metal beads. Subsequent steps were
928 performed at the Vienna Metabolomics Center (VIME, Department of Ecogenomics and Systems
929 Biology, Vienna, Austria).
930 Polar metabolites were extracted by modifying the procedure of Weckwerth et al., 2004. A
931 weighed amount of deep frozen and ground plant tissues was combined with 750 µL of ice-cold
932 extraction solvent, consisting of methanol (LC-MS grade, Merck), chloroform (anhydrous >99 %,
933 Sigma Aldrich), and water (MilliQ) in a ratio of 2.5:1:0.5 (v/v). Additionally, 7 µL of a solution
934 of 10 mmol of pentaerythritol (PE) and 10 mM phenyl-β-d-glucopyranoside (PGP) respectively in
935 water (MilliQ) were added as an internal standard mix. After ultrasonication at 4 °C for 20 minutes
936 and centrifugation (4 min, 4 °C, 14,000 g) the supernatant was transferred to a new 1.5 mL tube
937 (polypropylene). Another 250 µL of extraction solvent was added to the remaining pellet and after
938 another cycle of ultrasonication and centrifugation as described before, the supernatant was
939 combined with the previous. To induce phase separation 350 µL of water (MilliQ) was added.
940 After thorough mixing and consecutive centrifugation (4 min, 4 °C, 14,000 g) 900 µL of the upper
941 phase were transferred to a new 1.5 mL tube. Approximately 100 µL of the remaining polar phase
942 of all samples were combined. The 900 µL aliquotes of this mixed sample were used as quality
943 control during measurements. The polar phases and the aliquotes of the sample mix were dried in
944 a vacuum centrifuge for 5 hours at 30 °C and 0.1 mbar.

945 The dried extracts were derivatized as described earlier (Doerfler et al., 2013) by dissolving
946 the metabolite pellet carefully in 20 µL of 40 mg of methoxyamine hydrochloride (Sigma Aldrich)
947 in 1 mL pyridine (anhydrous >99,8 %, Sigma Aldrich). After incubation at 30 °C and 700 rpm for
948 1.5 hours on a thermoshaker, 80 µL of *N*-methyl-*N*-trimethylsilyl-trifluoroacetamid (Macherey-
949 Nagel) were added. The samples were incubated for 30 min at 37 °C and 750 rpm and
950 consecutively centrifuged for 4 min at room temperature and 14,000 g.
951 Metabolite analysis was performed on an Agilent 7890B gas chromatograph equipped with
952 a LECO Pegasus® BT-TOF mass spectrometer (LECO Corporation). Derivatized metabolites
953 were injected through a Split/Splitless inlet equipped with an ultra inert single tapered glass liner
954 with deactivated glass wool (5910-2293, Agilent Technologies), a split ratio of 1:25 was used and
955 the temperature was set to 230 °C. Components were separated with helium as carrier gas on a
956 Restek Rxi-5Sil MS column (length: 30 m, diameter: 0.25 mm, thickness of film: 0.25 µm). The
957 initial oven temperature was set to 70 °C held for 1 minute and ramped with a rate of 9 °C per
958 minute until reaching 340 °C help for 10 minutes. Collection of spectra started after an acquisition
959 delay of 280 seconds with a detector voltage of 1692.5 V, a rate of 15 spectra per second and a
960 mass range of 50−500 *m/z*. Retention indices were calculated based on the retention times of the
961 alkane mixture C_{10}-C_{40} run within each of the 2 batches. Samples were measured in randomized
962 order and randomly distributed across the batches. Within each batch, a mixture of standard
963 metabolites was measured for MSI level I identification of metabolites. Deconvolution, annotation,
964 and processing of chromatograms was performed according to (Zhang et al., 2023) using
965 ChromaTOF® (Version 5.55.29.0.1187, LECO Cooperation) and MS-DIAL, version 4.7
966 (Tsugawa et al., 2015). Areas of derivatization products of single metabolite were summed and
967 normalized by the main targeted ion content of each sample.
968
969 **5.11.3. RNA extraction and sequencing**
970 Using the same sampled leaf material as for targeted metabolite analyses, total RNA was
971 extracted using the QIAGEN RNeasy® Mini Kit in an RNAse free laboratory. Samples were
972 digested using the kit's RLT buffer with 1 % Beta-mercaptoethanol. Elution was done in two steps.
973 The purity and concentration of the extractions was measured using Nanodrop, and RIN and
974 fragmentation profiles were obtained with a Fragment Analyzer™ system. RNA libraries were

975 prepared by the Vienna Biocenter Core Facilities (VBCF) using a NEBNext stranded mRNA kit
976 before sequencing 150-bp paired-end reads on one lane of Illumina NovaSeq S4.
977
978 **5.11.4. RNA-seq data processing**
979 The raw RNA-seq data was evaluated with FastQC
980 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC (Ewels et al., 2016),
981 then quality trimmed using AdapterRemoval v.2.3.1 (Schubert et al., 2016) with settings --trimns
982 --trimqualities --minquality 20 --trimwindows 12 --minlength 36. The trimmed data was then
983 aligned to both the *T. fasciculata* and *T. leiboldiana* genomes using STAR v.2.7.9 (Dobin et al.,
984 2013) using GFF files to specify exonic regions. Because mapping bias was lowest when mapping
985 to *T. fasciculata* (see SI Note 9, Fig. S9, S10), our main analyses have been performed on the reads
986 mapped to this genome. However, the alignments to *T. leiboldiana* were used for verification or
987 expansion of the main analysis (SI Note 10).
988
989 **5.11.5. Differential Gene Expression analysis**
990 We quantified read counts per exon using FeatureCounts from the Subread package v.2.0.3.
991 (Liao et al., 2014) for paired-end and reversely stranded reads (-p -s 2). The counts were then
992 summed up across exons per gene to obtain gene-level counts. The composition of the count data
993 was investigated with PCA in EdgeR (Robinson et al., 2009). Then, counts were normalised using
994 the TMM method in EdgeR, and every gene with a mean cpm < 1 was removed. We ran a
995 differential gene expression (DE) analysis in maSigPro (Conesa et al., 2006), which detects genes
996 with differential diurnal expression profiles between species using a regression approach.
997 *Tillandsia leiboldiana* was used as the baseline in this analysis. Significant DE genes were then
998 clustered using the hclust algorithm into modules, with the number of modules being determined
999 with the K-means algorithm. Expression curves were plotted by taking the average expression
1000 across all replicates per species at each time point with a custom R script. Expression curves for
1001 entire clusters (Fig. S7) were plotted by median-centering the log(CPM) of each gene and time
1002 point against the median of all genes at each time point, while expression curves for individual
1003 genes or gene families (Fig. 5c, S5, S7) report average CPM.

1004 GO term enrichments were performed for each cluster using the R package TopGO (Alexa
1005 and Rahnenführer, 2009). Separately, known candidate genes underlying CAM and starch
1006 metabolism (See Methods, Section 5.9.) were searched among differentially expressed genes.
1007
1008
1009 **5.12. Intersecting findings of gene family evolution, TE insertion ad differential**
1010 **gene expression**
1011
1012 **5.12.1. Transposable element insertions and differential gene expression**
1013 Intronic TE insertions were obtained using *bedtools intersect* on the GFF files of the TE
1014 and gene annotations of both species. We used the full transcript length of a gene (feature =
1015 "mRNA" in GFF file) for this analysis, and only applied "known" TE annotations. This resulted
1016 in a dataset reporting the number of TE insertions per gene. We then performed two tests on the
1017 resulting TE counts per gene: (1) whether the proportion of genes with one or more TE insertions
1018 is elevated in DE genes compared to the full gene set (chi-square test), and (2) whether the rate of
1019 TE insertions per gene measured, as the total count of intersections for each gene annotation with
1020 a TE annotation, is elevated in DE genes compared to non-DE genes (Mann-Whitney U test).
1021 The same test was also applied to a restricted set of DE genes previously described as
1022 CAM-related, or involved in starch metabolism and gluconeogenesis. Then, genes of interest with
1023 a TE insertion rate higher than twice the genome-wide average were selected and the difference in
1024 number of TE insertions between orthologs of *T. leiboldiana* and *T. fasciculata* was taken in case
1025 of a one-to-one relationship.
1026
1027 **5.12.2. Gene family evolution and differential gene expression**
1028 Orthogroups were split based on relative family size in *T. fasciculata* (F) versus *T.*
1029 *leiboldiana* (L) in the following categories: Single-copy orthogroups (F = 1 : L = 1), orthogroups
1030 with family size larger in *T. fasciculata* (F > L), orthogroups with family size smaller in *T.*
1031 *fasciculata* (F < L), orthogroups with equal family sizes that are larger than 1 (F = L). Orthogroups
1032 unique to one species (F:0 or 0:L) were not considered in this analysis. We counted the number of
1033 orthogroups belonging to each category for the full orthogroup set, for the subset of orthogroups
1034 containing DE genes (DE orthogroups) and for the subset of orthogroups containing DE genes that

1035 have been previously described as CAM-related (CAM DE orthogroups). We then tested whether
1036 counts in each orthogroup category were enriched in DE orthogroups and DE CAM orthogroups
1037 compared to all orthogroups. For comparisons of all orthogroups versus DE orthogroups, we used
1038 the chi-square test of independence in R. For the comparison of CAM DE orthogroups versus all
1039 orthogroups, Fisher's exact test on 2x2 contingency tables of each orthogroup category versus all
1040 other categories, and DE genes versus non-DE genes was used due to small sample sizes. To study
1041 the effect of the reference genome used on our findings on gene family evolution in DE genes, we
1042 performed the same analysis on read counts obtained from mapping to *T. leiboldiana* (SI Note 10).
1043

1044 **6. Accession Numbers**

1045 The genome assemblies and raw data used in this study are available at NCBI-SRA under
1046 BioProject PRJNA927306. Specifically, the *T.fasciculata* genome assembly TFas_v1 can be
1047 downloaded here: https://ncbi.nlm.nih.gov/datasets/genome/GCA_029168755.1/. The *T.*
1048 *leiboldiana* genome assembly TLei_v1 can be found at:
1049 https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_029204045.1/. The annotation of both
1050 genomes is available on the github repository at:
1051 https://github.com/cgrootcrego/Tillandsia_Genomes, together with the list of orthogroups, counts
1052 table used for RNA-seq analyses, full GO term enrichment results, and all scripts written for this
1053 manuscript. The *A. comosus* sequences used in this study stem from BioProject PRJNA371634
1054 (F153): https://www.ncbi.nlm.nih.gov/bioproject/PRJNA371634/ and BioProject PRJNA747096
1055 (CB5): https://www.ncbi.nlm.nih.gov/bioproject/747096. Information on all accessions used in
1056 this study can be found in Supplemental Table 1.
1057

1058 **7. Supplemental data**

1059 Supplemental Figure S1. Per-metabolite loadings of PC1 and PC2
1060 Supplemental Figure S2. Distribution of scaffold sizes and orthologous gene counts for the top
1061 100 largest scaffolds of both assemblies.
1062 Supplemental Figure S3. TE, GC and gene content at three examples of syntenic chromosome
1063 triplets.
1064 Supplemental Figure S4. In-depth visualisation of large-scale rearrangements between *T.*
1065 *fasciculata* and *T. leiboldiana.*

1066 Supplemental Figure S5. Genome-wide distribution of d_N/d_S values between single-copy
1067 orthologous genes.
1068 Supplemental Figure S6. Average expression curve of PEPC kinase (PPCK) in *T. fasciculata* and
1069 *T. leiboldiana.*
1070 Supplemental Figure S7. Per-gene expression curves of all differentially expressed genes.
1071 Supplemental Figure S8. Average expression curve of Aquaporin 2-6 in *T. fasciculata* and *T.*
1072 *leiboldiana.*
1073 Supplemental Figure S9. Unique mapping rates of RNA-seq reads to three different reference
1074 genomes.
1075 Supplemental Figure S10. Multi-mapping rates of RNA-seq reads to three different reference
1076 genomes.
1077 Supplemental Figure S11. Genome size measurement histograms.
1078 Supplemental Figure S12. Mitotic metaphase chromosomes and karyotypes of *Tillandsia*
1079 *fasciculata* and *Tillandsia leiboldiana.*
1080 Supplemental Figure S13. Heterozygosity and genome size estimation with a k-mer based
1081 approach.
1082 Supplemental Figure S14. Distribution of heterozygous sites per 1000 mappable variants.
1083 Supplemental Figure S15. Mean per-gene coverage distribution across different gene family
1084 categories.
1085 Supplemental Figure S16. Distribution of DE genes across the genome.
1086 Supplemental Table S1. List of accessions used in this study.
1087 Supplemental Table S2. Assembly statistics.
1088 Supplemental Table S3. TE Abundances.
1089 Supplemental Table S4. Orthology statistics.
1090 Supplemental Table S5. List of CAM-related expanded orthogroups.
1091 Supplemental Table S6. Full list of candidate genes under positive selection.
1092 Supplemental Table S7. Full list of co-expression modules.
1093 Supplemental Table S8. List of DE genes related to CAM, starch metabolism and gluconeogenesis
1094 with a high TE insertion rate.
1095 Supplemental File 1. Supplementary Notes.
1096

8. Acknowledgments

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

9. 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, 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.

10. References

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

Ando, E., Ohnishi, M., Wang, Y., Matsushita, T., Watanabe, A., Hayashi, Y., Fujii, M., Ma, J.F., Inoue, S.-I., and Kinoshita, T. (2013). TWIN SISTER OF FT, GIGANTEA, and CONSTANS have a positive but indirect effect on blue light-induced stomatal opening in Arabidopsis. Plant Physiol. 162: 1529–1538.

Araújo, W.L. et al. (2011). Antisense inhibition of the iron-sulphur subunit of succinate

1129 dehydrogenase enhances photosynthesis and growth in tomato via an organic acid-
1130 mediated effect on stomatal aperture. Plant Cell 23: 600–627.

1131 Arnegard, M.E., Zwickl, D.J., Lu, Y., and Zakon, H.H. (2010). Old gene duplication facilitates
1132 origin and diversification of an innovative communication system—twice. Proceedings of
1133 the National Academy of Sciences 107: 22172–22177.

1134 Baduel, P., Quadrana, L., Hunter, B., Bomblies, K., and Colot, V. (2019). Relaxed purifying
1135 selection in autopolyploids drives transposable element over-accumulation which provides
1136 variants for local adaptation. Nat. Commun. 10: 5818.

1137 Barfuss, M.H.J., Till, W., Leme, E.M.C., Pinzón, J.P., Manzanares, J.M., Halbritter, H., Samuel,
1138 R., and Brown, G.K. (2016). Taxonomic revision of bromeliaceae subfam. Tillandsioideae
1139 based on a multi-locus DNA sequence phylogeny and morphology. Phytotaxa 279: 1–97.

1140 Benzing, D.H. and Bennett, B. (2000). Bromeliaceae: Profile of an Adaptive Radiation
1141 (Cambridge University Press).

1142 Borland, A.M., Hartwell, J., Weston, D.J., Schlauch, K.A., Tschaplinski, T.J., Tuskan, G.A., Yang,
1143 X., and Cushman, J.C. (2014). Engineering crassulacean acid metabolism to improve
1144 water-use efficiency. Trends Plant Sci. 19: 327–338.

1145 Bräutigam, A., Schlüter, U., Eisenhut, M., and Gowik, U. (2017). On the Evolutionary Origin of
1146 CAM Photosynthesis. Plant Physiol. 174: 473–477.

1147 Brawand, D. et al. (2015). The genomic substrate for adaptive radiation in African cichlid fish.
1148 Nature 513: 375–381.

1149 Brown, G.K. and Gilmartin, A.J. (1989). Chromosome Numbers in Bromeliaceae. Am. J. Bot. 76:
1150 657–665.

1151 Cai, J. et al. (2015). The genome sequence of the orchid Phalaenopsis equestris. Nat. Genet. 47:
1152 65–72.

1153 Campbell, M.S., Holt, C., Moore, B., and Yandell, M. (2014). Genome Annotation and Curation
1154 Using MAKER and MAKER-P. Curr. Protoc. Bioinformatics 2014: 4.11.1-4.11.39.

1155 Carnal, N.W. and Black, C.C. (1989). Soluble Sugars as the Carbohydrate Reserve for CAM in
1156 Pineapple Leaves : Implications for the Role of Pyrophosphate:6-Phosphofructokinase in
1157 Glycolysis. Plant Physiol. 90: 91–100.

1158 Chen, L.-Y. et al. (2019). The bracteatus pineapple genome and domestication of clonally
1159 propagated crops. Nat. Genet. 51: 1549–1558.

1160 Cheng, Y., Zhou, W., El Sheery, N.I., Peters, C., Li, M., Wang, X., and Huang, J. (2011).
1161 Characterization of the Arabidopsis glycerophosphodiester phosphodiesterase (GDPD)
1162 family reveals a role of the plastid-localized AtGDPD1 in maintaining cellular phosphate
1163 homeostasis under phosphate starvation. Plant J. 66: 781–795.

1164 Christin, P.-A., Arakaki, M., Osborne, C.P., Bräutigam, A., Sage, R.F., Hibberd, J.M., Kelly, S.,
1165 Covshoff, S., Wong, G.K.-S., Hancock, L., and Edwards, E.J. (2014). Shared origins of a
1166 key enzyme during the evolution of C4 and CAM metabolism. J. Exp. Bot. 65: 3609–3621.

1167 Cicconardi, F., Lewis, J.J., Martin, S.H., Reed, R.D., Danko, C.G., and Montgomery, S.H. (2021).
1168 Chromosome Fusion Affects Genetic Diversity and Evolutionary Turnover of Functional
1169 Loci but Consistently Depends on Chromosome Size. Mol. Biol. Evol. 38: 4449–4462.

1170 Conesa, A., Nueda, M.J., Ferrer, A., and Talón, M. (2006). maSigPro: a method to identify
1171 significantly differential expression profiles in time-course microarray experiments.
1172 Bioinformatics 22: 1096–1102.

1173 Cosentino, C., Di Silvestre, D., Fischer-Schliebs, E., Homann, U., De Palma, A., Comunian, C.,
1174 Mauri, P.L., and Thiel, G. (2013). Proteomic analysis of Mesembryanthemum crystallinum
1175 leaf microsomal fractions finds an imbalance in V-ATPase stoichiometry during the salt-
1176 induced transition from C3 to CAM. Biochem. J. 450: 407–415.

1177 Crayn, D.M., Winter, K., Schulte, K., and Smith, J.A.C. (2015). Photosynthetic pathways in
1178 Bromeliaceae: Phylogenetic and ecological significance of CAM and C3 based on carbon
1179 isotope ratios for 1893 species. Bot. J. Linn. Soc. 178: 169–221.

1180 Crayn, D.M., Winter, K., and Smith, J.A.C. (2004). Multiple origins of crassulacean acid
1181 metabolism and the epiphytic habit in the Neotropical family Bromeliaceae. Proceedings
1182 of the National Academy of Sciences 101: 3703–3708.

1183 Cushman, J.C. (2001). Crassulacean acid metabolism. A plastic photosynthetic adaptation to arid
1184 environments. Plant Physiol. 127: 1439–1448.

1185 Cushman, J.C., Tillett, R.L., Wood, J.A., Branco, J.M., and Schlauch, K.A. (2008). Large-scale
1186 mRNA expression profiling in the common ice plant, Mesembryanthemum crystallinum,
1187 performing C3 photosynthesis and Crassulacean acid metabolism (CAM). J. Exp. Bot. 59:
1188 1875–1894.

1189 Dalchau, N., Baek, S.J., Briggs, H.M., Robertson, F.C., Dodd, A.N., Gardner, M.J., Stancombe,
1190 M.A., Haydon, M.J., Stan, G.-B., Gonçalves, J.M., and Webb, A.A.R. (2011). The
1191 circadian oscillator gene *GIGANTEA* mediates a long-term response of the *Arabidopsis*
1192 *thaliana* circadian clock to sucrose. Proc. Natl. Acad. Sci. U. S. A. 108: 5104–5109.

1193 Davey, J.W., Chouteau, M., Barker, S.L., Maroja, L., Baxter, S.W., Simpson, F., Merrill, R.M.,
1194 Joron, M., Mallet, J., Dasmahapatra, K.K., and Jiggins, C.D. (2016). Major Improvements
1195 to the Heliconius melpomene Genome Assembly Used to Confirm 10 Chromosome Fusion
1196 Events in 6 Million Years of Butterfly Evolution. G3 6: 695–708.

1197 De La Harpe, M. et al. (2020). Genomic footprints of repeated evolution of CAM photosynthesis
1198 in a Neotropical species radiation. Plant Cell Environ. 43: 2987–3001.

1199 Deng, H., Zhang, L.-S., Zhang, G.-Q., Zheng, B.-Q., Liu, Z.-J., and Wang, Y. (2016). Evolutionary
1200 history of PEPC genes in green plants: Implications for the evolution of CAM in orchids.
1201 Mol. Phylogenet. Evol. 94: 559–564.

1202 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M.,
1203 and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics
1204 29: 15–21.

1205 Doerfler, H., Lyon, D., Nägele, T., Sun, X., Fragner, L., Hadacek, F., Egelhofer, V., and
1206 Weckwerth, W. (2013). Granger causality in integrated GC–MS and LC–MS
1207 metabolomics data reveals the interface of primary and secondary metabolism.
1208 Metabolomics 9: 564–574.

1209 Edwards, E.J. (2023). Reconciling continuous and discrete models of C4 and CAM evolution.
1210 Ann. Bot.

1211 Emms, D.M. and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative
1212 genomics. Genome Biol. 20: 238.

1213 Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results
1214 for multiple tools and samples in a single report. Bioinformatics 32: 3047–3048.

1215 Faria, R. and Navarro, A. (2010). Chromosomal speciation revisited: rearranging theory with
1216 pieces of evidence. Trends Ecol. Evol. 25: 660–669.

1217 Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P., and Firoozabady, E.
1218 (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Science 220:
1219 1049–1051.

1220 Gitaí, J., Paule, J., Zizka, G., Schulte, K., and Benko-Iseppon, A.M. (2014). Chromosome numbers
1221 and DNA content in Bromeliaceae: Additional data and critical review. Bot. J. Linn. Soc.
1222 176: 349–368.

1223 Givnish, T.J. et al. (2014). Adaptive radiation, correlated and contingent evolution, and net species
1224 diversification in Bromeliaceae. Mol. Phylogenet. Evol. 71: 55–78.

1225 Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M.,
1226 Talón, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and
1227 data mining with the Blast2GO suite. Nucleic Acids Res. 36: 3420–3435.

1228 Grabherr, M.G. et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a
1229 reference genome. Nat. Biotechnol. 29: 644–652.

1230 Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). Circlize implements and enhances
1231 circular visualization in R. Bioinformatics 30: 2811–2812.

1232 Hayashi, S., Ishii, T., Matsunaga, T., Tominaga, R., Kuromori, T., Wada, T., Shinozaki, K., and
1233 Hirayama, T. (2008). The glycerophosphoryl diester phosphodiesterase-like proteins
1234 SHV3 and its homologs play important roles in cell wall organization. Plant Cell Physiol.
1235 49: 1522–1535.

1236 Heyduk, K., McAssey, E.V., and Leebens-Mack, J. (2022). Differential timing of gene expression
1237 and recruitment in independent origins of CAM in the Agavoideae (Asparagaceae). New

1238 Phytol. 235: 2111–2126.

1239 Heyduk, K., Moreno-Villena, J.J., Gilman, I.S., Christin, P.-A., and Edwards, E.J. (2019a). The
1240 genetics of convergent evolution: insights from plant photosynthesis. Nat. Rev. Genet. 20:
1241 485–493.

1242 Heyduk, K., Ray, J.N., Ayyampalayam, S., Moledina, N., Borland, A., Harding, S.A., Tsai, C.-J.,
1243 and Leebens-Mack, J. (2019b). Shared expression of crassulacean acid metabolism (CAM)
1244 genes pre-dates the origin of CAM in the genus Yucca. J. Exp. Bot. 70: 6597–6609.

1245 Hoang, N.V. et al. (2023). The Gynandropsis gynandra genome provides insights into whole-
1246 genome duplications and the evolution of C4 photosynthesis in Cleomaceae. Plant Cell 35:
1247 1334–1359.

1248 Hoff, K.J., Lomsadze, A., Borodovsky, M., and Stanke, M. (2019). Whole-genome annotation
1249 with BRAKER. In Gene prediction (Humana: New York), pp. 65–95.

1250 Holtum, J.A.M., Smith, J.A.C., and Neuhaus, H.E. (2005). Intracellular transport and pathways of
1251 carbon flow in plants with crassulacean acid metabolism. Funct. Plant Biol. 32: 429–449.

1252 Jang, T.-S. and Weiss-Schneeweiss, H. (2015). Formamide-Free Genomic in situ Hybridization
1253 Allows Unambiguous Discrimination of Highly Similar Parental Genomes in Diploid
1254 Hybrids and Allopolyploids. Cytogenet. Genome Res. 146: 325–331.

1255 Kajala, K., Brown, N.J., Williams, B.P., Borrill, P., Taylor, L.E., and Hibberd, J.M. (2012).
1256 Multiple Arabidopsis genes primed for recruitment into C_4 photosynthesis. Plant J. 69: 47–
1257 56.

1258 Katju, V. and Bergthorsson, U. (2013). Copy-number changes in evolution: rates, fitness effects
1259 and adaptive significance. Front. Genet. 4: 273.

1260 Kong, D. et al. (2016). L-Met Activates Arabidopsis GLR Ca2+ Channels Upstream of ROS
1261 Production and Regulates Stomatal Movement. Cell Rep. 17: 2553–2561.

1262 Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017).
1263 Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat
1264 separation. Genome Res. 27: 722–736.

1265 Leroy, T., Anselmetti, Y., Tilak, M.-K., Bérard, S., Csukonyi, L., Gabrielli, M., Scornavacca, C.,
1266 Milá, B., Thébaud, C., and Nabholz, B. (2021). A bird's white-eye view on avian sex
1267 chromosome evolution. Peer Community Journal 1.

1268 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1269 arXiv [q-bio.GN].

1270 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094–
1271 3100.

1272 Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program

1273 for assigning sequence reads to genomic features. Bioinformatics 30: 923–930.

1274 Lieberman-Aiden, E. et al. (2009). Comprehensive mapping of long-range interactions reveals
1275 folding principles of the human genome. Science 326: 289–293.

1276 Liu, L., Li, X., Yuan, L., Zhang, G., Gao, H., Xu, X., and Zhao, H. (2022). XAP5 CIRCADIAN
1277 TIMEKEEPER specifically modulates 3' splice site recognition and is important for
1278 circadian clock regulation partly by alternative splicing of LHY and TIC. Plant Physiol.
1279 Biochem. 172: 151–157.

1280 Liu, L., Tumi, L., Suni, M.L., Arakaki, M., Wang, Z.-F., and Ge, X.-J. (2021). Draft genome of
1281 Puya raimondii (Bromeliaceae), the Queen of the Andes. Genomics 113: 2537–2546.

1282 Lovell, J.T., Sreedasyam, A., Schranz, M.E., Wilson, M., Carlson, J.W., Harkess, A., Emms, D.,
1283 Goodstein, D.M., and Schmutz, J. (2022). GENESPACE tracks regions of interest and gene
1284 copy number variation across multiple genomes. Elife 11: e78526.

1285 Lowry, D.B. and Willis, J.H. (2010). A widespread chromosomal inversion polymorphism
1286 contributes to a major life-history transition, local adaptation, and reproductive isolation.
1287 PLoS Biol. 8.

1288 Luo, J., Sun, X., Cormack, B.P., and Boeke, J.D. (2018). Karyotype engineering by chromosome
1289 fusion leads to reproductive isolation in yeast. Nature 560: 392–396.

1290 Males, J. (2018). Concerted anatomical change associated with crassulacean acid metabolism in
1291 the Bromeliaceae. Funct. Plant Biol. 45: 681–695.

1292 McClung, C.R. (2006). Plant circadian rhythms. Plant Cell 18: 792–803.

1293 McGee, M.D., Borstein, S.R., Meier, J.I., Marques, D.A., Mwaiko, S., Taabu, A., Kishe, M.A.,
1294 O'Meara, B., Bruggmann, R., Excoffier, L., and Seehausen, O. (2020). The ecological and
1295 genomic basis of explosive adaptive radiation. Nature 586: 75–79.

1296 McRae, S.R., Christopher, J.T., Smith, J.A.C., and Holtum, J.A.M. (2002). Sucrose transport
1297 across the vacuolar membrane of Ananas comosus. Funct. Plant Biol. 29: 717–724.

1298 Messerschmid, T.F.E., Wehling, J., Bobon, N., Kahmen, A., Klak, C., Los, J.A., Nelson, D.B., dos
1299 Santos, P., de Vos, J.M., and Kadereit, G. (2021). Carbon isotope composition of plant
1300 photosynthetic tissues reflects a Crassulacean Acid Metabolism (CAM) continuum in the
1301 majority of CAM lineages. Perspect. Plant Ecol. Evol. Syst. 51: 125619.

1302 Ming, R. et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. Nat.
1303 Genet. 47: 1435–1442.

1304 Mondragón-Palomino, M. and Theissen, G. (2009). Why are orchid flowers so diverse? Reduction
1305 of evolutionary constraints by paralogues of class B floral homeotic genes. Ann. Bot. 104:
1306 583–594.

1307 Moriyama, Y., Ito, F., Takeda, H., Yano, T., Okabe, M., Kuraku, S., Keeley, F.W., and Koshiba-

1308 Takeuchi, K. (2016). Evolution of the fish heart by sub/neofunctionalization of an elastin
1309 gene. Nat. Commun. 7: 10397.

1310 Ogburn, R.M. and Edwards, E.J. (2009). Anatomical variation in Cactaceae and relatives: Trait
1311 lability and evolutionary innovation. Am. J. Bot. 96: 391–408.

1312 Ohno, S. (1970). Evolution by gene duplication (Springer Berlin: Heidelberg).

1313 Osmond, C.B. (1978). Crassulacean Acid Metabolism: A Curiosity in Context. Annu. Rev. Plant
1314 Physiol. 29: 379–414.

1315 Otto, F.J., Oldiges, H., Göhde, W., and Jain, V.K. (1981). Flow cytometric measurement of nuclear
1316 DNA content variations as a potential in vivo mutagenicity test. Cytometry 2: 189–191.

1317 Ou, S. et al. (2019). Benchmarking transposable element annotation methods for creation of a
1318 streamlined, comprehensive pipeline. Genome Biol. 20: 275.

1319 Pedro, D.L.F., Amorim, T.S., Varani, A., Guyot, R., Domingues, D.S., and Paschoal, A.R. (2021).
1320 An Atlas of Plant Transposable Elements. F1000Res. 10: 1194.

1321 Philippe, F., Verdu, I., Morère-Le Paven, M.-C., Limami, A.M., and Planchet, E. (2019).
1322 Involvement of Medicago truncatula glutamate receptor-like channels in nitric oxide
1323 production under short-term water deficit stress. J. Plant Physiol. 236: 1–6.

1324 Pierce, S., Winter, K., and Griffiths, H. (2002). Carbon isotope ratio and the extent of daily CAM
1325 use by Bromeliaceae. New Phytol. 156: 75–83.

1326 Popp, M., Janett, H.-P., Lüttge, U., and Medina, E. (2003). Metabolite gradients and carbohydrate
1327 translocation in rosette leaves of CAM and C3 bromeliads. New Phytol. 157: 649–656.

1328 Putnam, N.H. et al. (2016). Chromosome-scale shotgun assembly using an in vitro method for
1329 long-range linkage. Genome Res. 26: 342–350.

1330 Quezada, I.M. and Gianoli, E. (2011). Crassulacean acid metabolism photosynthesis in
1331 Bromeliaceae: an evolutionary key innovation. Biol. J. Linn. Soc. Lond. 104: 480–486.

1332 Ranwez, V., Douzery, E.J.P., Cambon, C., Chantret, N., and Delsuc, F. (2018). MACSE v2:
1333 Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop
1334 Codons. Mol. Biol. Evol. 35: 2582–2584.

1335 Roach, M.J., Schmidt, S.A., and Borneman, A.R. (2018). Purge Haplotigs: allelic contig
1336 reassignment for third-gen diploid genome assemblies. BMC Bioinformatics 19: 460.

1337 Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2009). edgeR: A Bioconductor package for
1338 differential expression analysis of digital gene expression data. Bioinformatics 26: 139–
1339 140.

1340 Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming,
1341 identification, and read merging. BMC Res. Notes 9: 88.

1342 Silvera, K., Neubig, K.M., Mark Whitten, W., Williams, N.H., Winter, K., and Cushman, J.C.
1343 (2010). Evolution along the crassulacean acid metabolism continuum. Funct. Plant Biol.
1344 37: 995–1010.

1345 Silvera, K., Santiago, L.S., Cushman, J.C., and Winter, K. (2009). Crassulacean acid metabolism
1346 and epiphytism linked to adaptive radiations in the Orchidaceae. Plant Physiol. 149: 1838–
1347 1847.

1348 Silvera, K., Winter, K., Rodriguez, B.L., Albion, R.L., and Cushman, J.C. (2014). Multiple
1349 isoforms of phosphoenolpyruvate carboxylase in the Orchidaceae (subtribe Oncidiinae):
1350 implications for the evolution of crassulacean acid metabolism. J. Exp. Bot. 65: 3623–
1351 3636.

1352 Simpson, G.G. (1953). The Major Features of Evolution (Columbia University Press).

1353 Smit, A.F.A., Hubley, R., and Green, P. (2013-2015). RepeatMasker Open-4.0.

1354 Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically
1355 mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24: 637–644.

1356 Tay, I.Y.Y., Odang, K.B., and Cheung, C.Y.M. (2021). Metabolic Modeling of the C3-CAM
1357 Continuum Revealed the Establishment of a Starch/Sugar-Malate Cycle in CAM
1358 Evolution. Front. Plant Sci. 11: 2221.

1359 Temsch, E.M. (2010). Genome size in liverworts. Preslia 82: 63–80.

1360 Temsch, E.M., Koutecký, P., Urfus, T., Šmarda, P., and Doležel, J. (2022). Reference standards
1361 for flow cytometric estimation of absolute nuclear DNA content in plants. Cytometry A
1362 101: 710–724.

1363 Töpfer, N., Braam, T., Shameer, S., Ratcliffe, R.G., and Sweetlove, L.J. (2020). Alternative
1364 Crassulacean Acid Metabolism Modes Provide Environment-Specific Water-Saving
1365 Benefits in a Leaf Metabolic Model. Plant Cell 32: 3689–3705.

1366 Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst,
1367 J., Fiehn, O., and Arita, M. (2015). MS-DIAL: data-independent MS/MS deconvolution
1368 for comprehensive metabolome analysis. Nat. Methods 12: 523–526.

1369 Vera-Estrella, R., Barkla, B.J., Amezcua-Romero, J.C., and Pantoja, O. (2012). Day/night
1370 regulation of aquaporins during the CAM cycle in Mesembryanthemum crystallinum. Plant
1371 Cell Environ. 35: 485–501.

1372 de Vos, J.M., Augustijnen, H., Bätscher, L., and Lucek, K. (2020). Speciation through
1373 chromosomal fusion and fission in Lepidoptera. Philos. Trans. R. Soc. Lond. B Biol. Sci.
1374 375: 20190539.

1375 Wai, C.M. et al. (2017). Temporal and spatial transcriptomic and microRNA dynamics of CAM
1376 photosynthesis in pineapple. Plant J. 92: 19–30.

1377 Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng,
1378 Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for
1379 comprehensive microbial variant detection and genome assembly improvement. PLoS One
1380 9: e112963.

1381 Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G.,
1382 Kriventseva, E.V., and Zdobnov, E.M. (2018). BUSCO Applications from Quality
1383 Assessments to Gene Prediction and Phylogenomics. Mol. Biol. Evol. 35: 543–548.

1384 Weckwerth, W., Wenzel, K., and Fiehn, O. (2004). Process for the integrated extraction,
1385 identification and quantification of metabolites, proteins and RNA to reveal their co-
1386 regulation in biochemical networks. Proteomics 4: 78–83.

1387 Weiland, M., Mancuso, S., and Baluska, F. (2015). Signalling via glutamate and GLRs in
1388 Arabidopsis thaliana. Funct. Plant Biol. 43: 1–25.

1389 Wickell, D. et al. (2021). Underwater CAM photosynthesis elucidated by Isoetes genome. Nature
1390 Communications 2021 12:1 12: 1–13.

1391 Winter, K., Garcia, M., Virgo, A., and Smith, J.A.C. (2021). Low-level CAM photosynthesis in a
1392 succulent-leaved member of the Urticaceae, Pilea peperomioides. Funct. Plant Biol. 48:
1393 683–690.

1394 Winter, K. and Smith, J.A.C. (2022). CAM photosynthesis: the acid test. New Phytol. 233: 599–
1395 609.

1396 Winter, K. and Smith, J.A.C. (2012). Crassulacean Acid Metabolism: Biochemistry,
1397 Ecophysiology and Evolution (Springer: Berlin, Germany).

1398 Wong, W.S.W., Yang, Z., Goldman, N., and Nielsen, R. (2004). Accuracy and power of statistical
1399 methods for detecting adaptive evolution in protein coding sequences and for identifying
1400 positively selected sites. Genetics 168: 1041–1051.

1401 Xiao, B., Huang, Y., Tang, N., and Xiong, L. (2007). Over-expression of a LEA gene in rice
1402 improves drought resistance under the field conditions. Züchter Genet. Breed. Res. 115:
1403 35–46.

1404 Yang, X. et al. (2017). The Kalanchoë genome provides insights into convergent evolution and
1405 building blocks of crassulacean acid metabolism. Nat. Commun. 8: 1899.

1406 Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24:
1407 1586–1591.

1408 Yardeni, G. et al. (2021). Taxon-specific or universal? Using target capture to study the
1409 evolutionary history of rapid radiations. Mol. Ecol. Resour.

1410 Zhang, S., Ghatak, A., and Bazargani, M. (2023). Cell-type proteomic and metabolomic resolution
1411 of early and late grain filling stages of wheat endosperm. Plant Biotechnology Journal.

1412 Zhu, F. and Ming, R. (2019). Global identification and expression analysis of pineapple aquaporins
1413 revealed their roles in CAM photosynthesis, boron uptake and fruit domestication.
1414 Euphytica 215: 132.

1415