Supplementary Information

"Short structural variation fuelled CAM evolution within an explosive bromeliad radiation"

Groot Crego, C.; Hess, J.; Yardeni, G.; de La Harpe, M.; Beclin, F.; Cauz-Santos, L.A.; Saadain, S.; Barbará, T.; Temsch, EM; Weiss-Scheeweiss, H.; Barfuss, M.H.J.; Till, W.; Lexer, C.; Paun, O.; Leroy, T.

1. SUPPLEMENTARY FIGURES 2

2. SUPPLEMENTARY TABLES 14

3. SUPPLEMENTARY NOTES 21

NOTE 1: GENOME SIZE AND KARYOTYPE OF *T. FASCICULATA* AND *T. LEIBOLDIANA* 21
NOTE 2: PRE-ASSEMBLY ESTIMATION OF PER-ACCESSION HETEROZYGOSITY 21
NOTE 3: IDENTIFYING MAIN SCAFFOLDS IN DE NOVO ASSEMBLY 22
NOTE 4: ON THE SUCCESS OF *DE NOVO* ASSEMBLY OF HIGHLY REPETITIVE GENOMES 23
NOTE 5: ON THE SPATIAL DISTRIBUTION OF GC AND TE CONTENT IN BROMELIAD GENOMES 24
NOTE 6: IDENTIFYING LARGE-SCALE REARRANGEMENTS BETWEEN *T. FASCICULATA* AND *T. LEIBOLDIANA* 25
NOTE 7: CORRECTING MULTI-COPY GENE FAMILY SIZES 26
NOTE 8: SELECTING RAPIDLY EVOLVING GENE FAMILIES 27
NOTE 9: DETAILED DESCRIPTION OF CANDIDATE GENES FOR POSITIVE SELECTION 28
NOTE 10: DIFFERENTIAL GENE EXPRESSION USING THE *T. LEIBOLDIANA* ASSEMBLY 29
NOTE 11: DISTRIBUTION OF DE GENES ACROSS THE GENOME 30
NOTE 12: LIMITATIONS OF THIS STUDY AND FUTURE DIRECTIONS 30

4. REFERENCES 31

1. Supplementary Figures

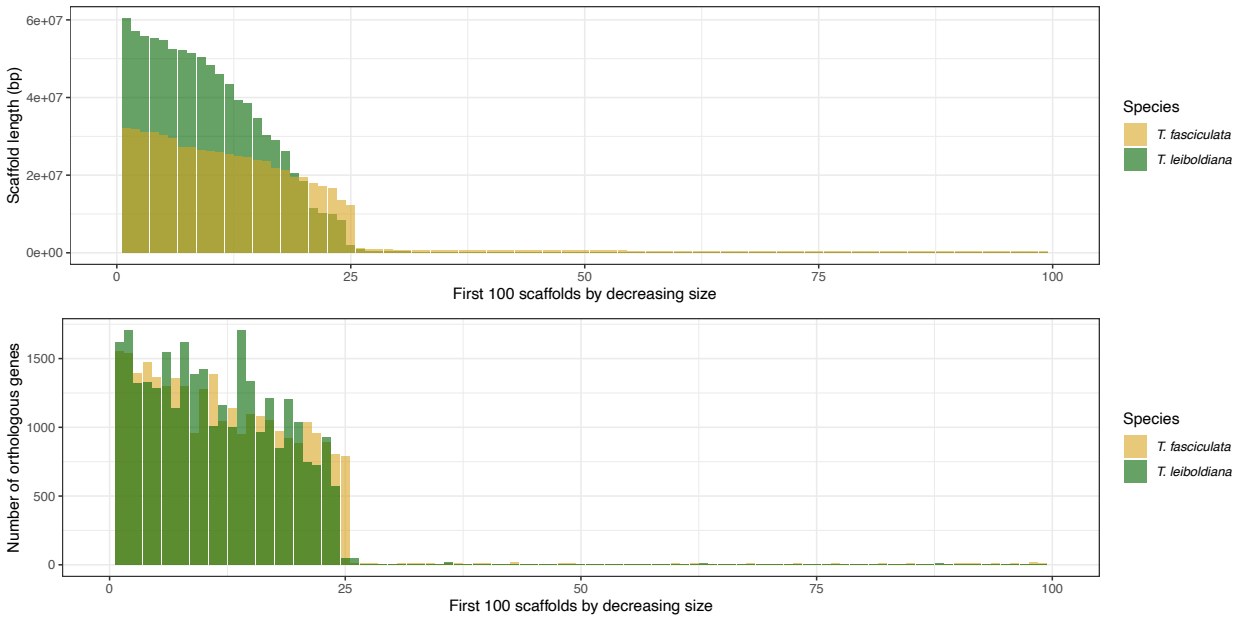


Figure S1: Distribution of scaffold sizes (top) and count of orthologous genes per scaffold (bottom) for the top 100 largest scaffolds of both assemblies.

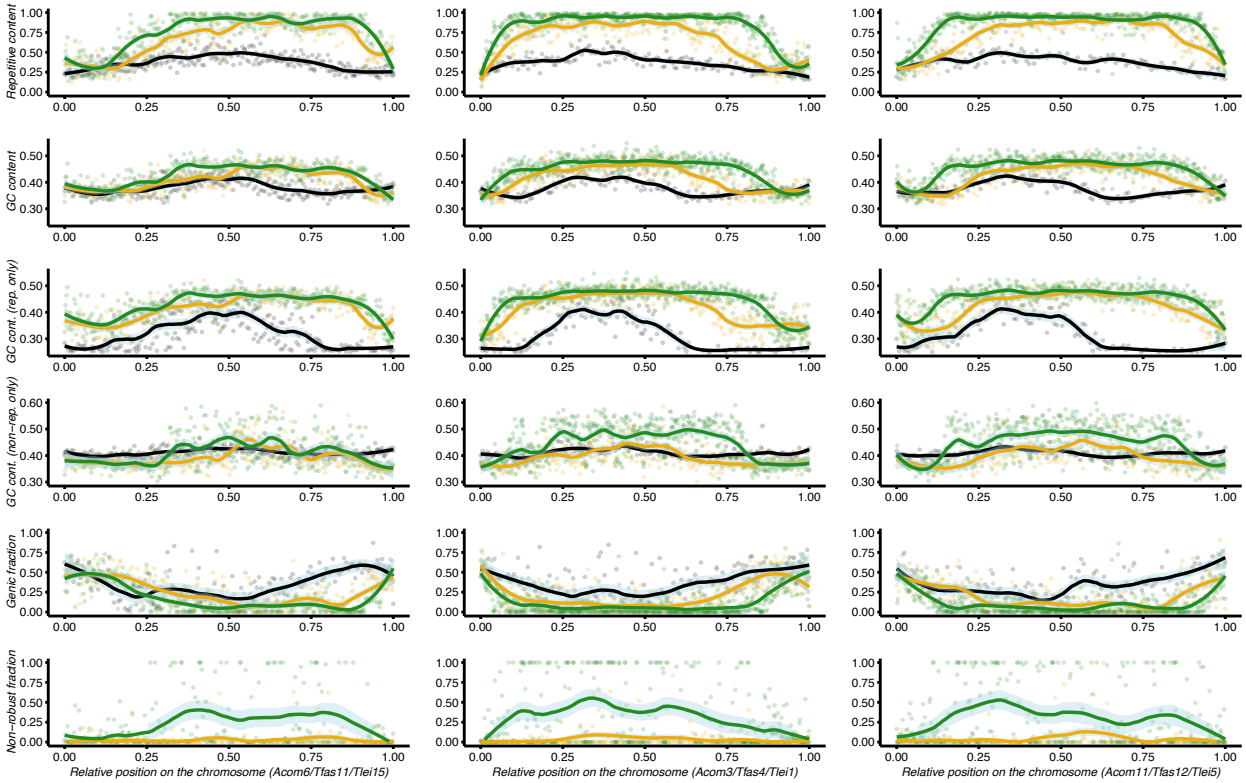
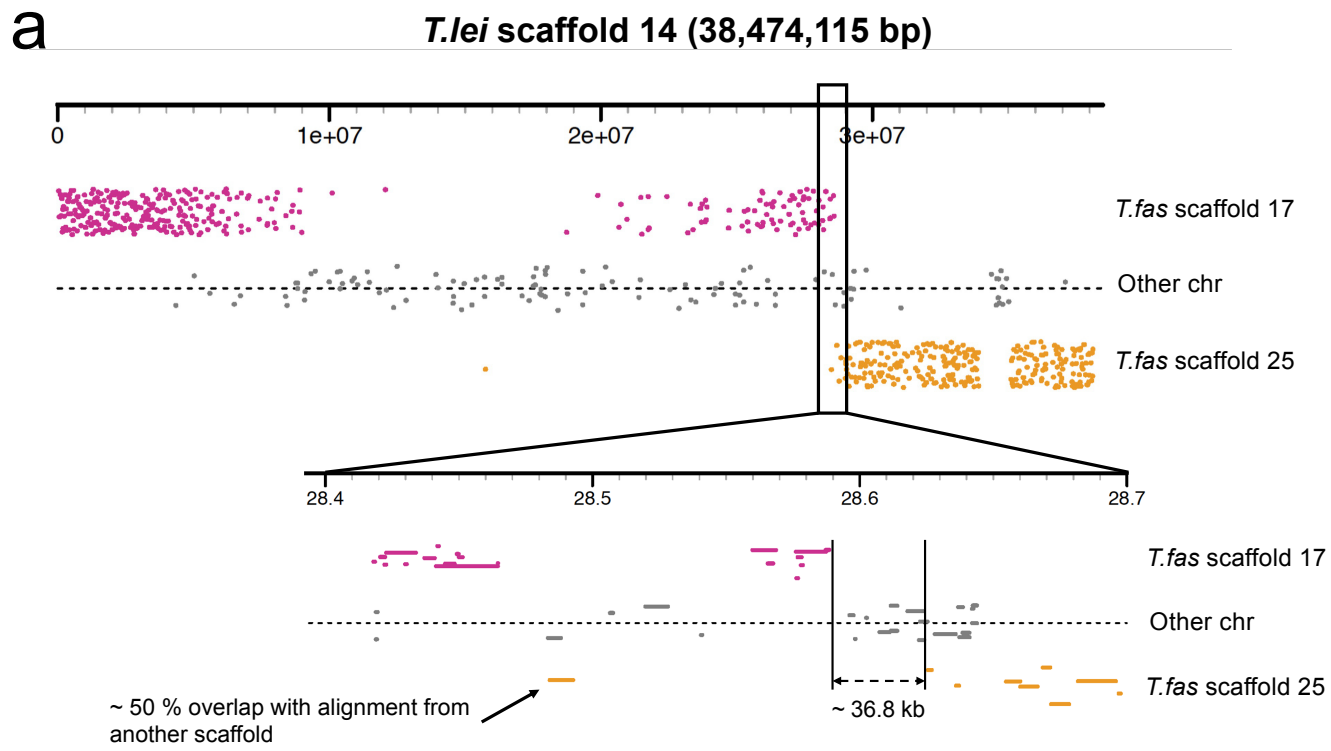
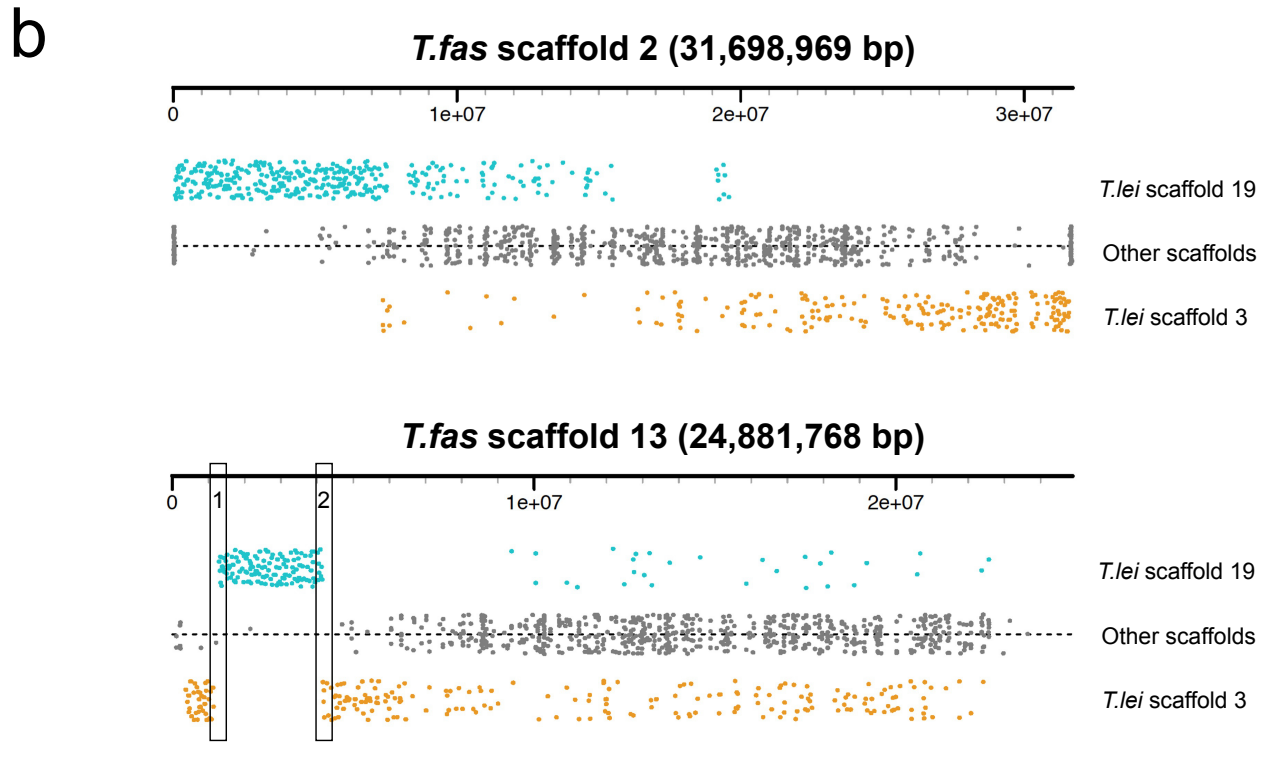


Figure S2: TE, GC and gene content at three examples of syntenic chromosome triplets of A. comosus (black), T. fasciculata (yellow) and T. leiboldiana (green). Each column represents a separate chromosome triplet. Each dot corresponds to an estimate in a non-overlapping 100 kb window. The line corresponds to the local regression (loess). Row-wise, from top to bottom, the plots show: (1) per-window proportion of soft-masked position in the assemblies (repetitive content), (2) GC content at all non-N positions (soft-masked or not), (3) GC content at soft-masked positions only, (4) GC content at non-softmasked positions only, (5) per-window proportion of bases falling in genes (genic fraction) and (6) the proportion of the genic fraction corresponding to non-robust genes (i.e. 1 minus this fraction corresponds to "robust" gene regions). This latter information is only provided for our two reference assemblies.


a
T.lei scaffold 14 (38,474,115 bp)
0
1e+07
2e+07
3e+07
T.fas scaffold 17
Other chr
T.fas scaffold 25
28.4
28.5
28.6
28.7
T.fas scaffold 17
Other chr
T.fas scaffold 25
~ 50 % overlap with alignment from another scaffold
~ 36.8 kb


b
T.fas scaffold 2 (31,698,969 bp)
0
1e+07
2e+07
3e+07
T.lei scaffold 19
Other scaffolds
T.lei scaffold 3
T.fas scaffold 13 (24,881,768 bp)
0
1
2
1e+07
2e+07
T.lei scaffold 19
Other scaffolds
T.lei scaffold 3

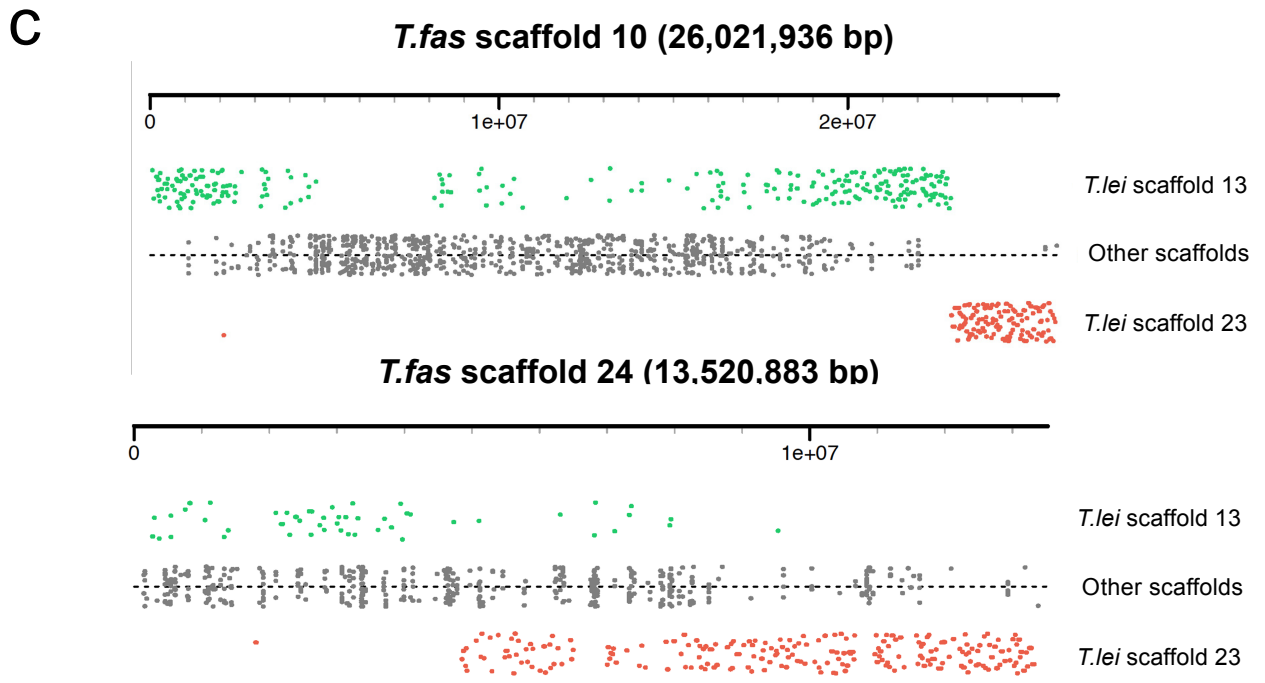
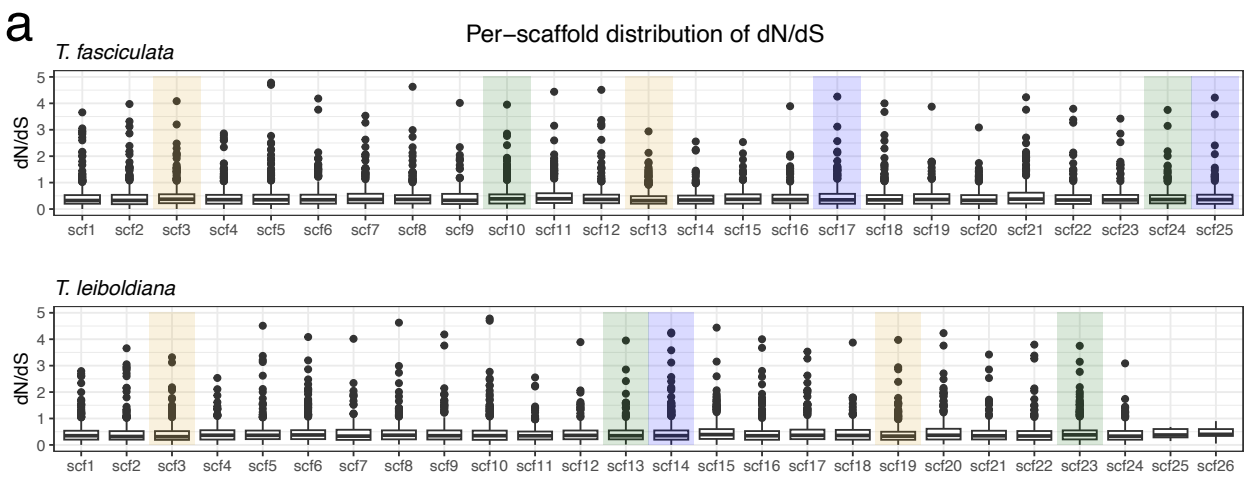
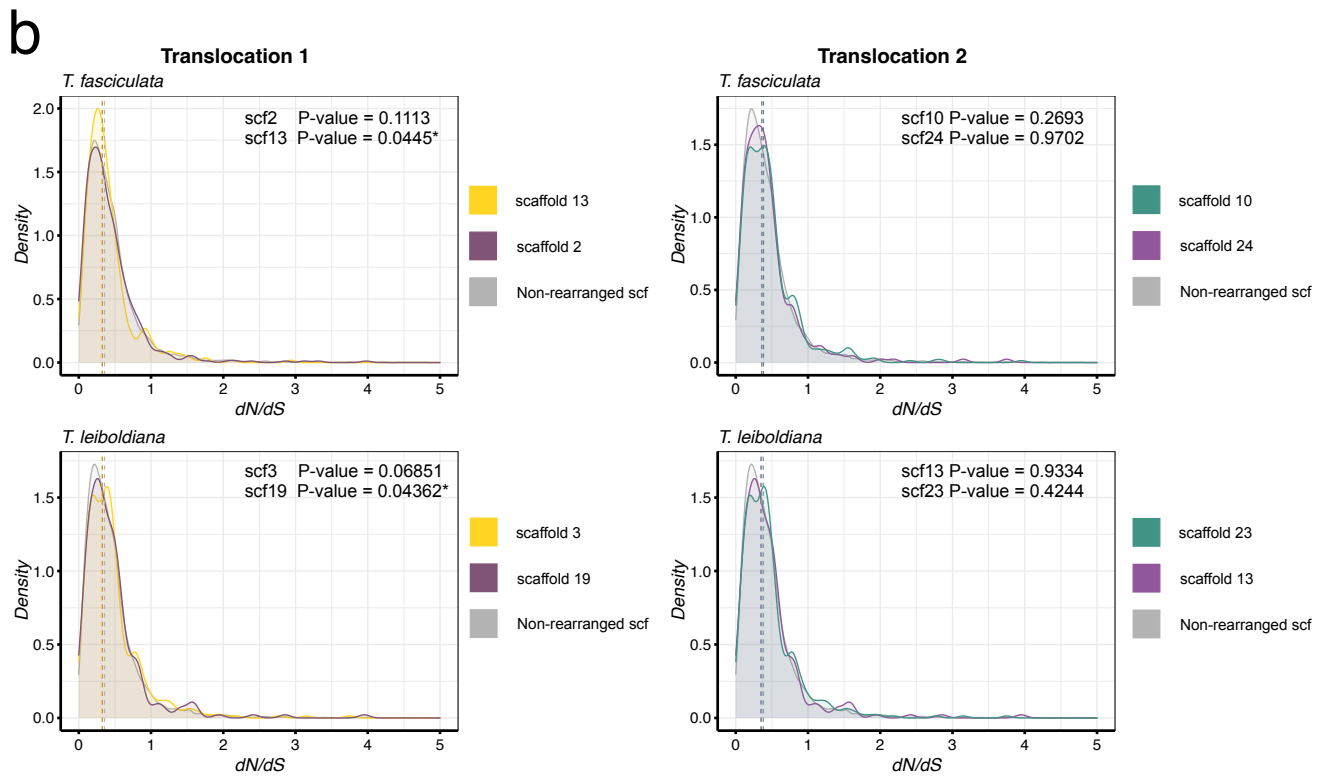


Figure S3: In-depth visualisation of large-scale rearrangements between T. fasciculata *and* T. leiboldiana *based on local alignments with less than a 90 % overlap with any other alignment. a) Potential fusion of scaffold 14 in* T. leiboldiana*, with enlargement of the breakpoint area. B) Translocation 1 – alignments were too sparse to determine a breakpoint on scaffold 2. Breakpoint 1 on scaffold 13 in* T.fasciculata *was not supported by raw PacBio alignments, however breakpoint 2 was. C) Translocation 2 – we find PacBio alignment support for the breakpoint on scaffold 10, but alignments were too sparse on scaffold 24 to determine a breakpoint. For more in-depth analysis and visualization of PacBio alignments, see the document 'Tfas_Tlei_rearrangements.pdf' on our github repository.*



58



59

60 *Figure S4: Genome-wide distribution of d_N/d_S values between single-copy orthologous genes. A) Boxplot*
61 *of d_N/d_S values in each scaffold of both assemblies. For ease of reading, the y-axis is cut-off at a d_N/d_S*
62 *value of five. Therefore, candidate genes with high values are not shown here. Scaffolds highlighted in*
63 *colors are involved in the three reported large-scale rearrangements: (1) chromosomal fusion in* T.
64 leiboldiana *(blue), (2) translocation 1 (yellow), and (3) translocation 2 (green) B) Distribution of d_N/d_S*
65 *values of all non-rearranged chromosomes versus chromosomes involved in translocations. P-values*
66 *were obtained through the Mann Whitney U test.*

67

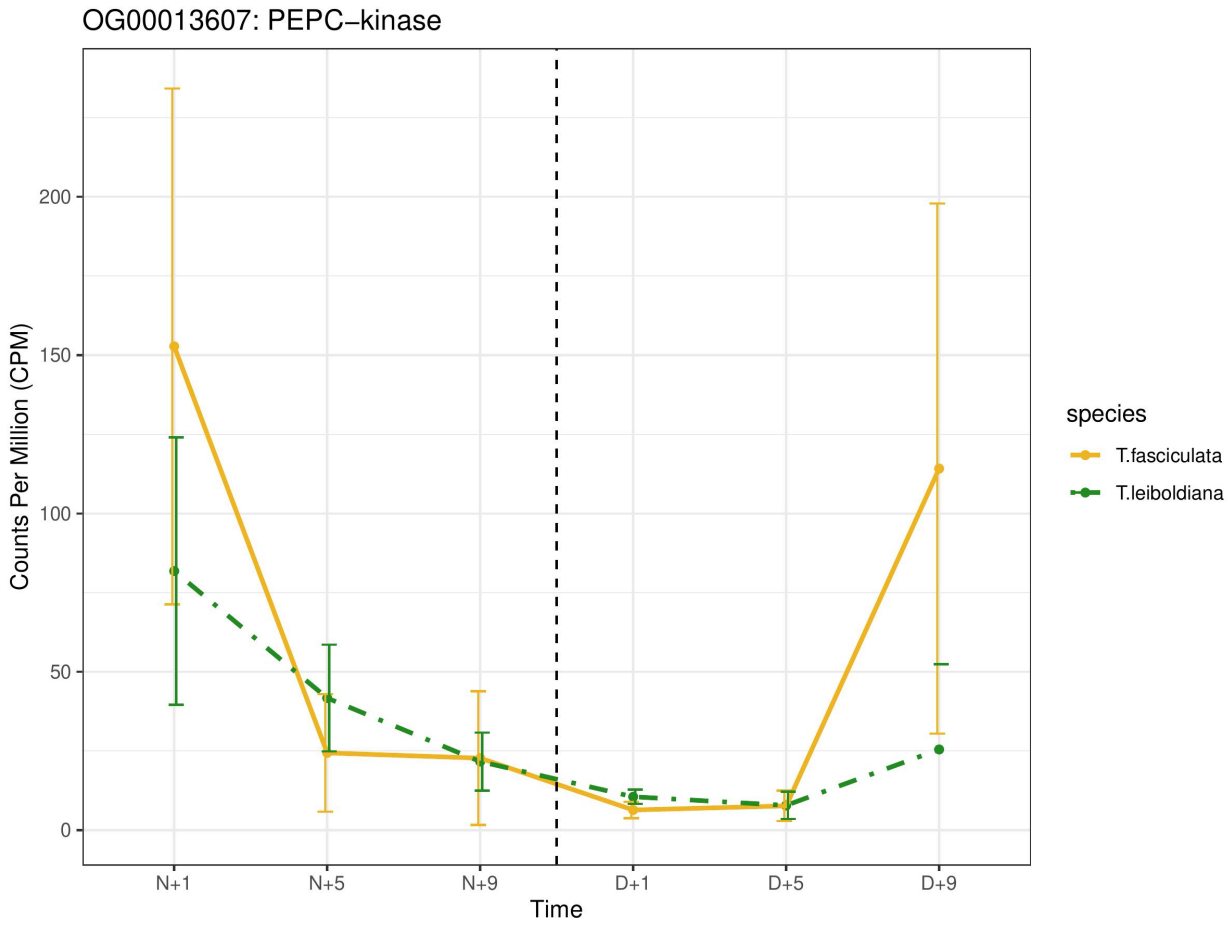


Figure S5: Average expression curve of PEPC kinase in T. fasciculata *and* T. leiboldiana *with standard deviation. The dashed vertical line marks the point where the light was switched on. Time is indicated in hours after the lights go off (N=Night) and after they go on (D=Day).*

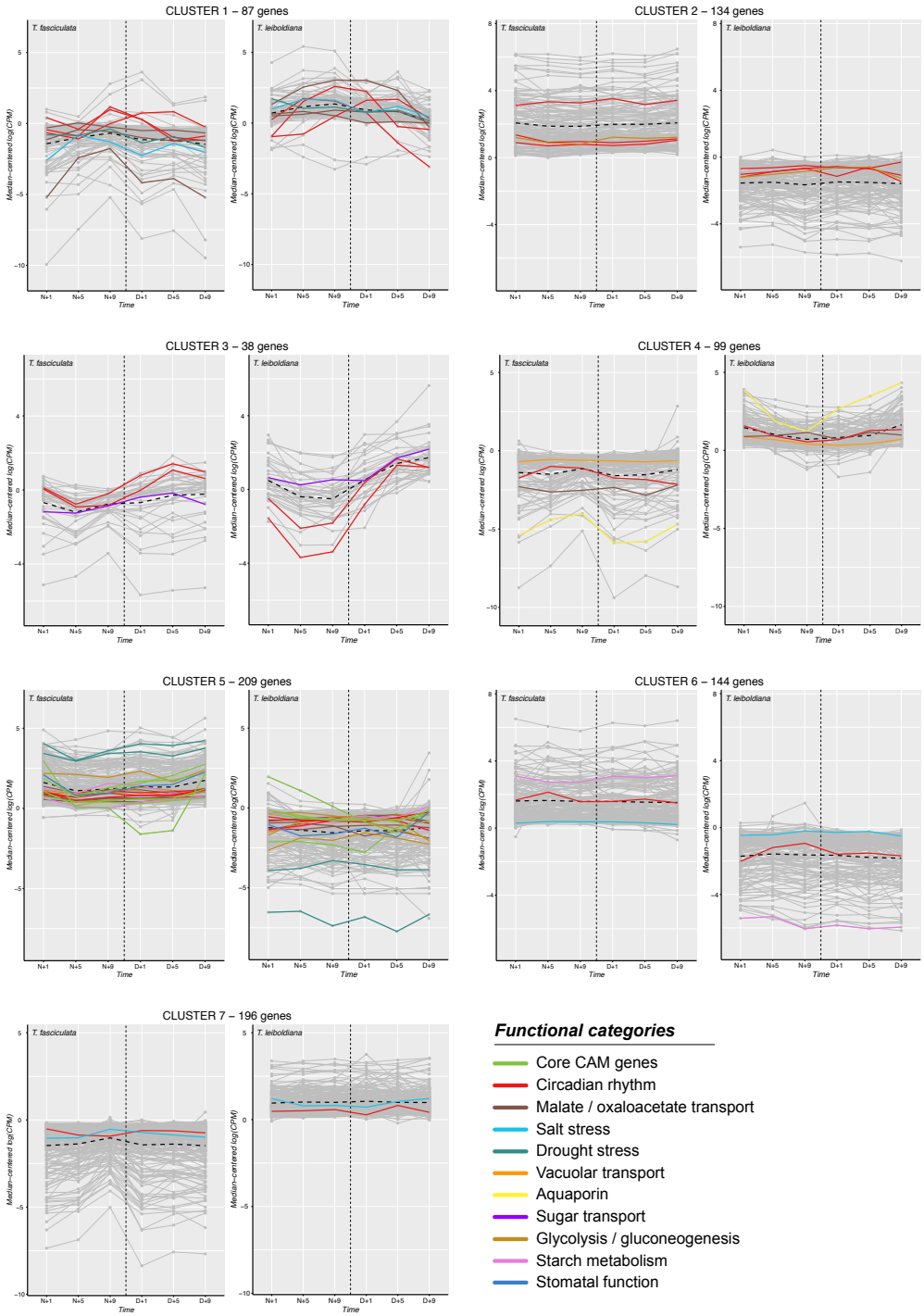


Figure S6: Per-gene expression curves of all differentially expressed genes, spread over 7 co-expression clusters inferred with MaSigPro (and T. fasciculata *as reference genome). The dashed vertical line marks the point where the light was switched on. Time is indicated in hours after the lights go off (N=Night) and after they go on (D=Day). Highlighted expression curves represent candidate genes underlying CAM-related functions. The colors correspond to specific subfunctions, laid out in the legend below.*

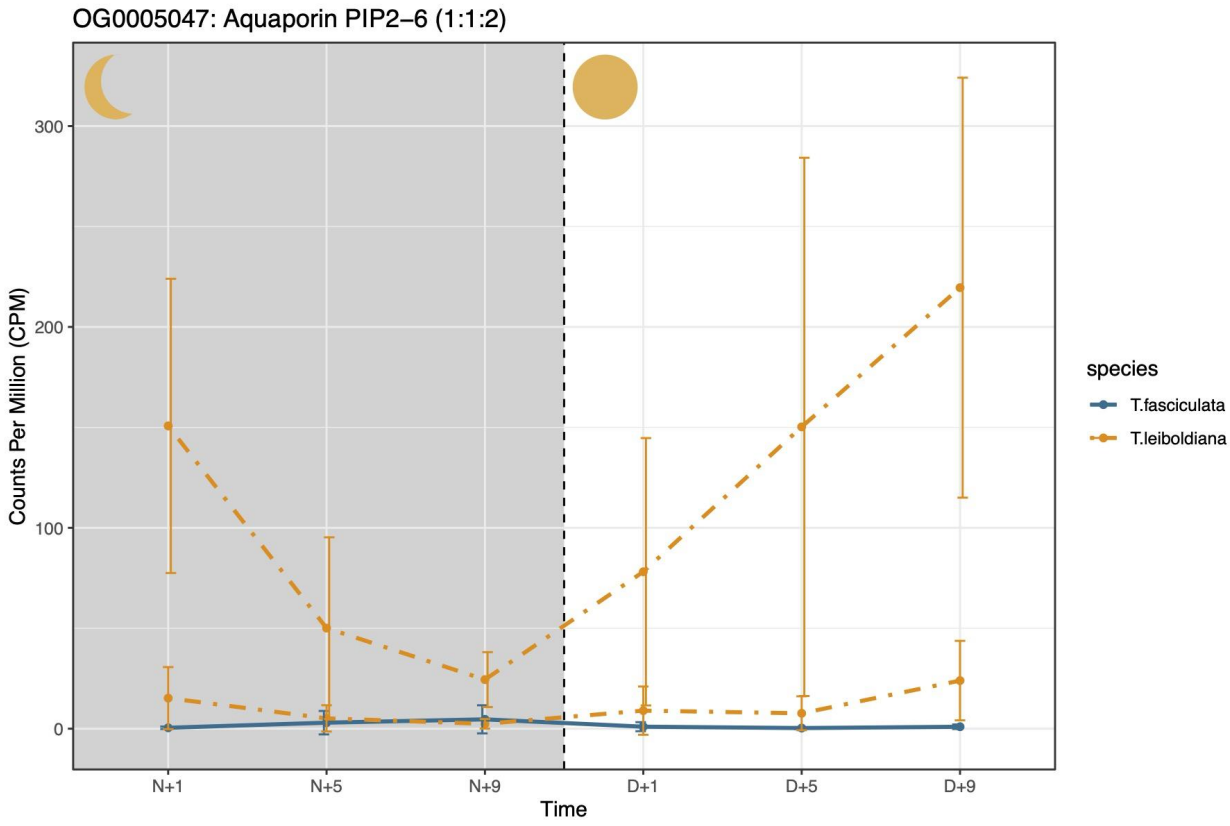


Figure S7: Average expression curve of Aquaporin PIP2-6 in T. fasciculata *and* T. leiboldiana *with standard deviation. The dashed vertical line marks the point where the light was switched on. Time is indicated in hours after the lights go off (N=Night) and after they go on (D=Day).*

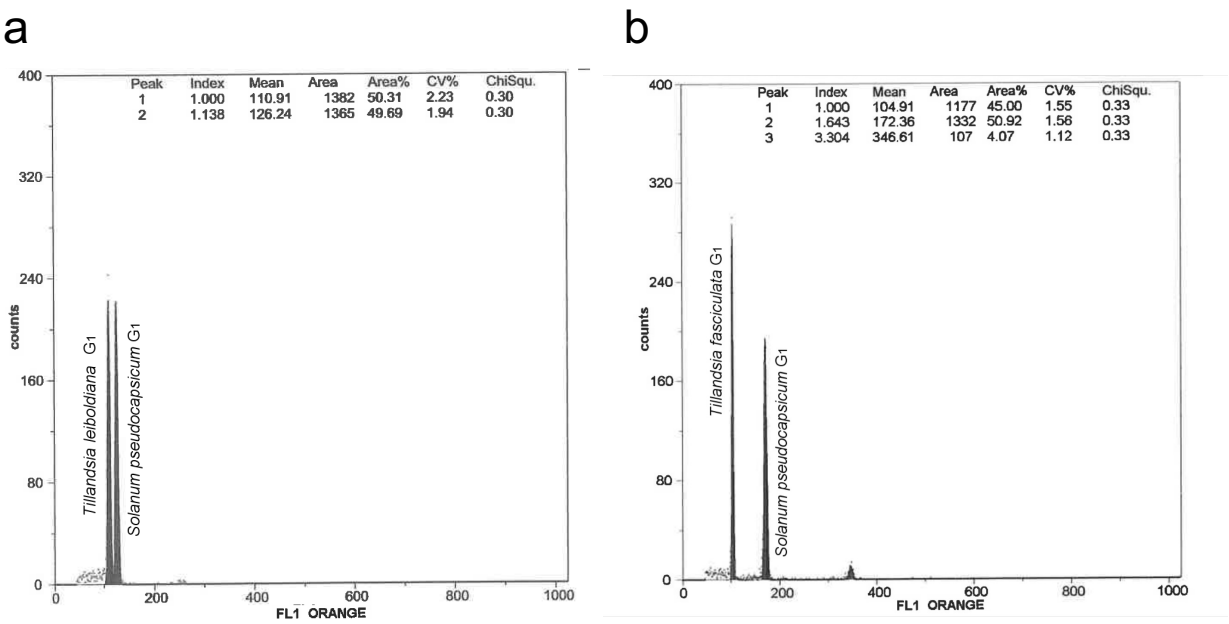


Figure S8: Genome size measurement histograms of each one exemplary run of Tillandsia fasciculata *and* T. leiboldiana *showing the mean G_1 nuclei peak positions on the x-axis (fluorescence intensity) of the samples and the standard organism (Solanum pseudocapsicum, 1.295pg/1C).*

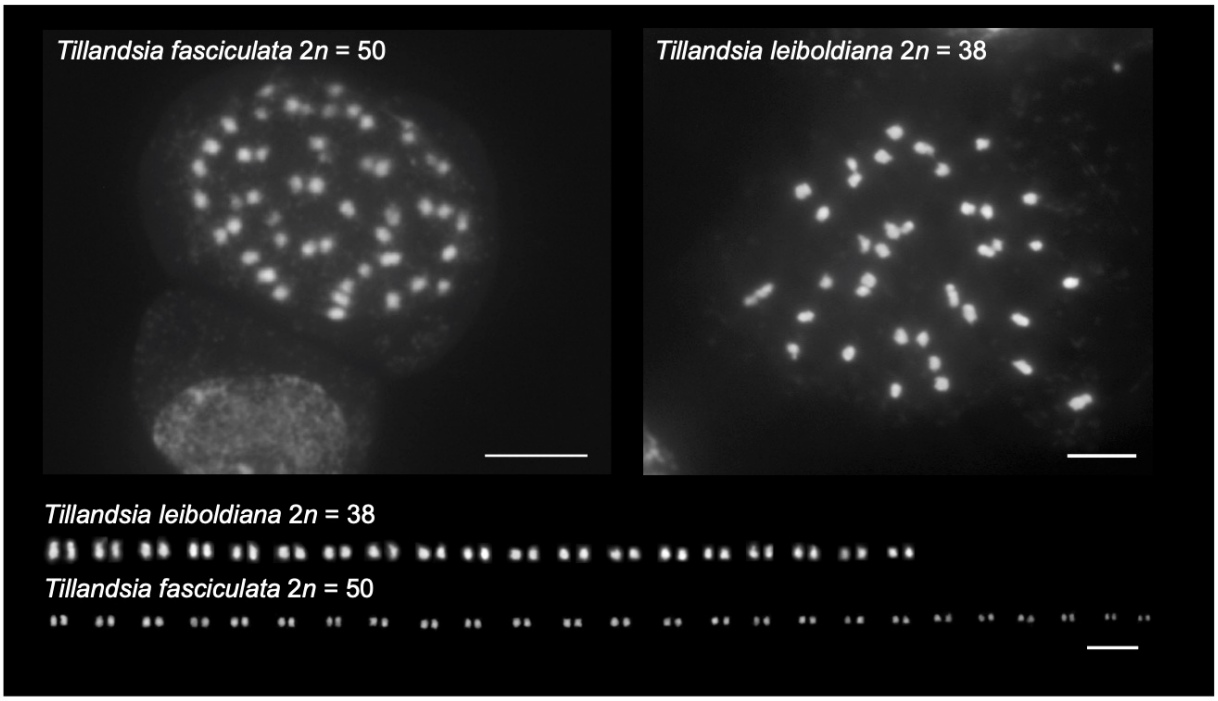


Figure S9: Mitotic metaphase chromosomes and karyotypes of Tillandsia fasciculata *and* Tillandsia leiboldiana*. Scale bar, 5µm.*

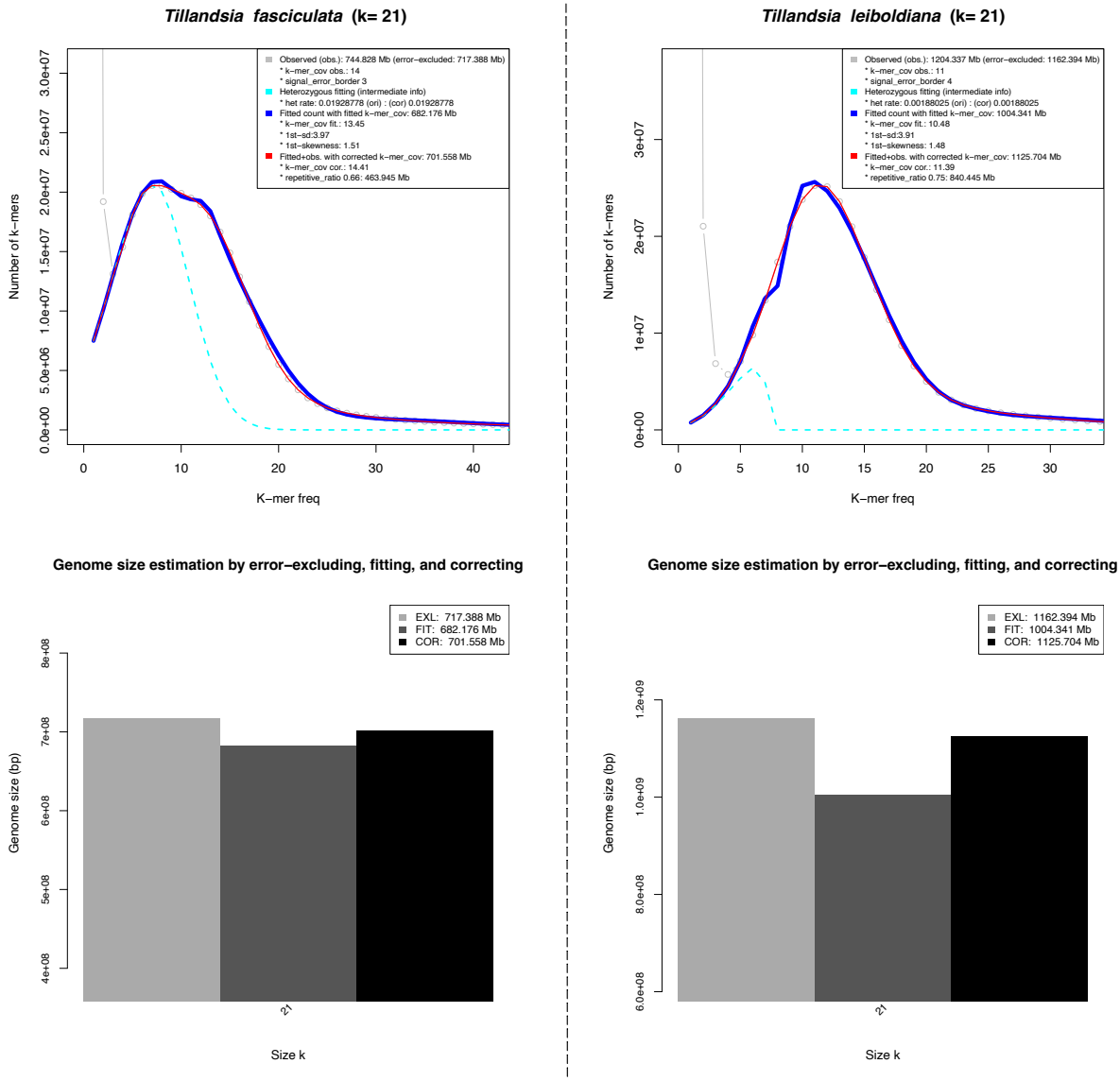


Figure S10: Heterozygosity and genome size estimation with a k-mer based approach implemented in findGSE for Tillandsia fasciculata *(left)* and T. leiboldiana *(right).*

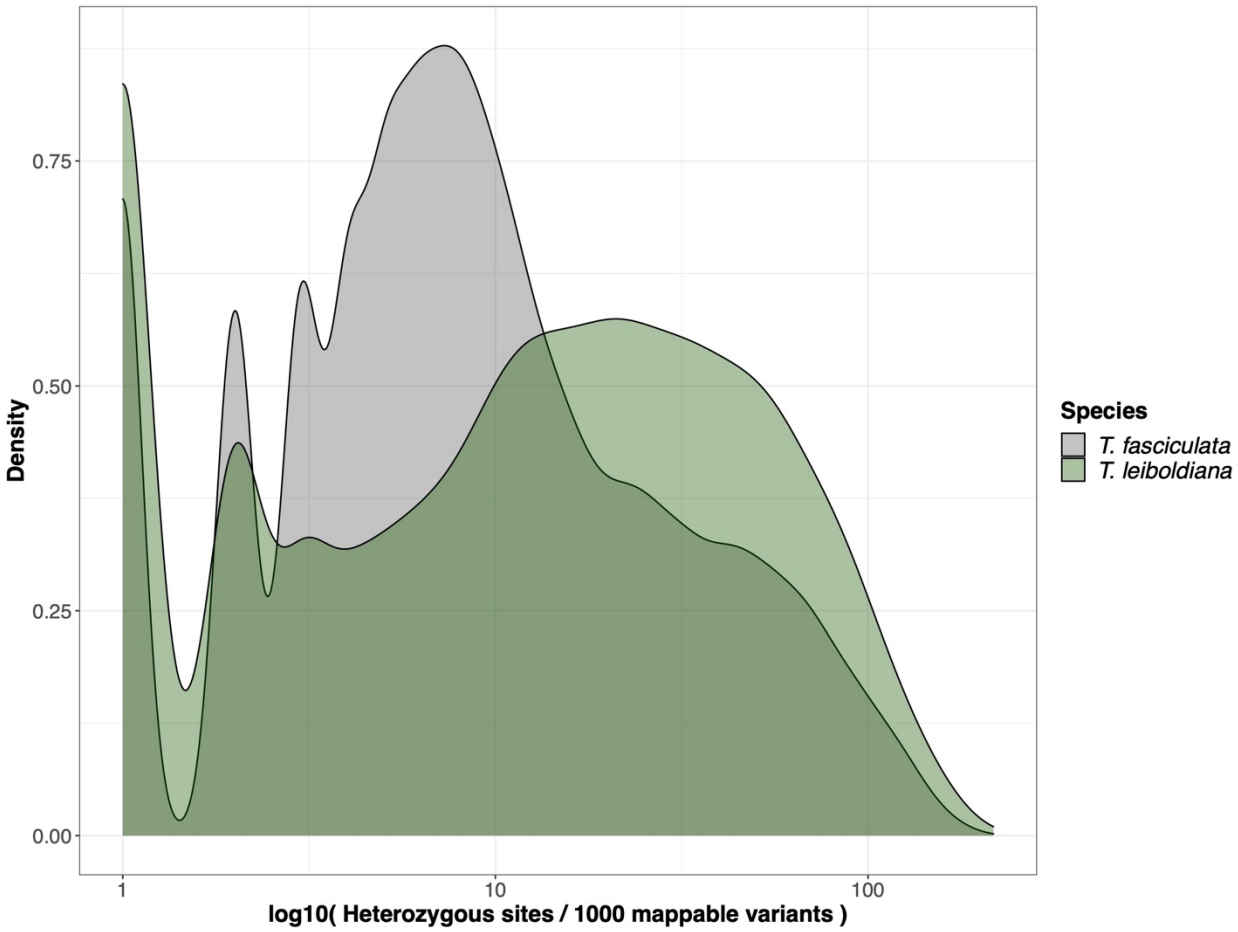


Figure S11: Distribution of heterozygous sites per 1000 mappable variants on a logarithmic scale.

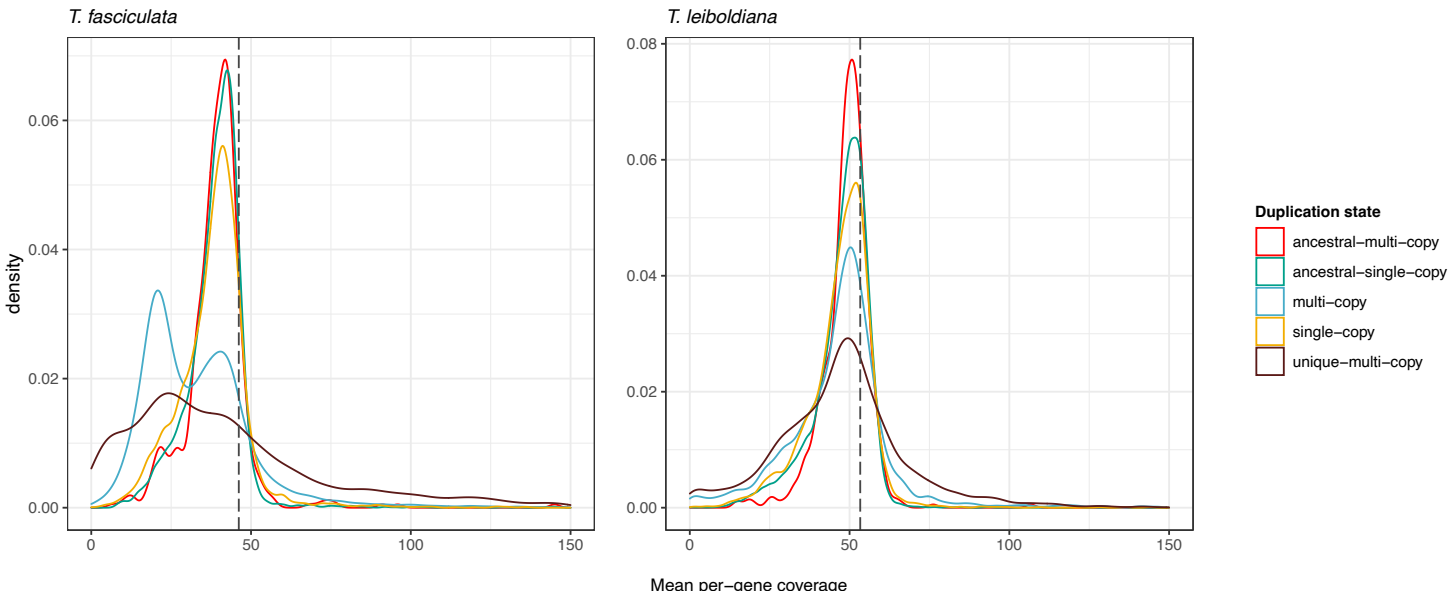


Figure S12: Mean per-gene coverage distribution across different gene family categories in T. fasciculata *(left) and* T. leiboldiana *(right). For further explanation of the different categories, see SI Note 7. Grey dashed lines indicate the whole-genome mean coverage.*

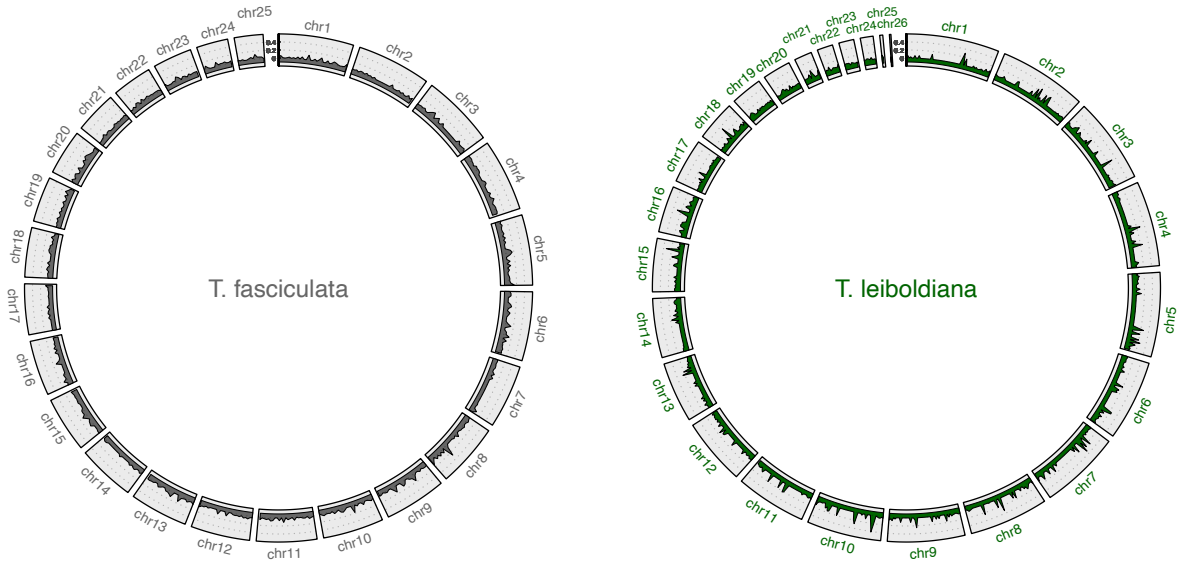


Figure S13: Proportion of genes per 1 Mb window that are differentially expressed across the T. fasciculata *(left) and* T. leiboldiana *(right) assembly.*

2. Supplementary tables

Table S1: Information on sampling and collection of accessions used in this study.

Source	Accession code	Herbarium accession	DNA №	Collector	Species	Country	Locality	Comment
Genome Assembly								
Botanical Garden of the University of Vienna	HBV 0024657 (B179/91)	WU 0013642	MHJB-B1840	W. & S. Till 7116	Tillandsia fasciculata	Costa Rica	Prov. Puntarenas, SW declivities of Cordillera de Tilaran, along the road from Sta. Elena to Rancho Grande.	
Botanical Garden of the University of Vienna	HBV 0024715 (B82/91)	WU 0003058	MHJB-B1842	W. & S. Till 7112	Tillandsia leiboldiana	Costa Rica	Prov. Alajuela, 2 km N San Ramon	
RNA-SEQ for Gene Annotation								
Botanical Garden of the University of Vienna	HBV 0026950 (B103/94)	WU 0006321	MHJB-B1839	G. Noller 9106	Tillandsia fasciculata	Guatemala	Depto. Sololá, Lago de Atitlán	Originally sequenced for De La Harpe et. al. (2020)
Botanical Garden of the University of Vienna	HBV 0024656 (B108/94)	WU 0002124	MHJB-B360	K.-D. & R. Ehlers EM890701	Tillandsia fasciculata	Mexico	Estdo. Chiapas, Sumidero Cañon bei Tuxtla Gutierrez	Originally sequenced for De La Harpe et. al. (2020)
Botanical Garden of the University of Vienna	HBV 0025322 (B100/91)	WU 0013632	MHJB-B2295	W. & S. Till 7005	Tillandsia fasciculata	Costa Rica	Prov. Heredia, Barva north of Heredia	Originally sequenced for De La Harpe et. al. (2020)
Botanical Garden of the University of Vienna	HBV 0000663 (B84/91, BRO000613)	WU 0001725, WU 0003008	MHJB-B323	W. & S. Till 7043	Tillandsia leiboldiana	Costa Rica	Prov. Cartago, Turriaba, Centro Agronomico Tropical de Investigadon y Enseñanza (CATIE)	
RNA-SEQ for Time Course Experiment								
Botanical Garden of the University of Vienna	HBV 0025322 (B100/91)	s.d.	MHJB-B2295	W. & S. Till 7005	Tillandsia fasciculata	Costa Rica	Prov. Heredia, Barva north of Heredia	Tfas_E in RNA-seq analysis
Botanical Garden of the University of Vienna	HBV 0025194 (B293/96)	s.d.	MHJB-B2296	S. Schatzl 77/59	Tillandsia fasciculata	Mexico	Estdo. Jalisco, ca. 20 km S of Porto Vallarta alc	Tfas_B in RNA-seq analysis
Botanical Garden of the University of Vienna	HBV 0025334 (B99B53-1)	WU 0008562, WU 0013708	MHJB-B1838	E. Kamm s.n.	Tillandsia fasciculata	Honduras	s.d.	Tfas_C in RNA-seq analysis
Botanical Garden of the University of Vienna	HBV 0024655 (B99/91) (90/91)	WU 0013632, WU 0013754	MHJB-B1841	W. & S. Till 7004	Tillandsia fasciculata	Costa Rica	Prov. Heredia, Barva north of Heredia	Tfas_D in RNA-seq analysis
Botanical Garden of the University of Vienna	HBV 0025326 (B90/91)	s.d.	MHJB-B2297	W. & S. Till 7006 (7004)	Tillandsia fasciculata	Costa Rica	Prov. Heredia, Barva north of Heredia	Tfas_A in RNA-seq analysis
Botanical Garden of the University of Vienna	HBV 0024657 (B179/91)	WU 0013642	MHJB-B1840	W. & S. Till 7116	Tillandsia fasciculata	Costa Rica	Prov. Puntarenas, SW declivities of Cordillera de Tilaran, along the road from Sta. Elena to Rancho Grande.	Tfas_F in RNA-seq analysis
Botanical Garden of the University of Vienna (Com.Bak BV)	HBV 0032437 (Bak 126)	s.d.	MHJB-B1960	s.d. (DBG Sept. 2011)	Tillandsia leiboldiana	Costa Rica	above Carteso	Tlei_A in RNA-seq analysis
Botanical Garden of the University of Vienna (Com.Bak BV)	HBV 0032433 (Bak 27)	s.d.	MHJB-B1957	P. Bak s.n.	Tillandsia leiboldiana	Costa Rica	s.d.	Tlei_C in RNA-seq analysis
Botanical Garden of the University of Vienna	HBV 0024715 (B82/91)	WU 0003058	MHJB-B1842	W. & S. Till 7112	Tillandsia leiboldiana	Costa Rica	Prov. Alajuela, 2 km N San Ramon	Tlei_D in RNA-seq analysis
Botanical Garden of the University of Vienna (Com.Bak BV)	HBV 0032436 (Bak 119)	s.d.	MHJB-B1959	s.d. (DBG Sept. 2011)	Tillandsia leiboldiana	Mexico	Estdo. Puebla, Amixtlán	Tlei_E in RNA-seq analysis
Botanical Garden of the University of Vienna (Com.Bak BV)	HBV 0032434 (Bak 37)	s.d.	MHJB-B1958	P. Bak s.n.	Tillandsia leiboldiana	Costa Rica	s.d.	Tlei_F in RNA-seq analysis
Botanical Garden of the University of Vienna (Com.Bak BV)	HBV 0032435 (Bak 45)	s.d.	MHJB-B1956	P. Bak s.n.	Tillandsia leiboldiana	Honduras	s.d.	Tlei_G in RNA-seq analysis
Abbreviations in Table:								
WU = Institutional Code (Wien Universität)								
HBV = Hortus Botanicus Vindobonensis (Botanical Garden of the University of Vienna)								
DBG = Deutsche Bromelien-Gesellschaft (German Bromeliad Society)								
s.n. = sin nomero (without number)								
s.d. = sin datos (without data)								

Table S2: Summary statistics of the *T. fasciculata* and *T. leiboldiana* assembly and gene annotation

Assembly statistics	*T. fasciculata*	*T. leiboldiana*
Total length (bp)	837,577,910	1,198,225,148
Total scaffold count (> 1 kb)	2,321	10,433
Total contig count	8,625	20,447
N50 (Mb)	23,642	43,365
N90 (Kb)	145,438	27,985
L50	16	12
L90	565	2,898
GC content	42.81 %	44.73 %
Uniquely mapping RNA-seq reads	69.13 %	92.37 %
Complete BUSCO genes	91.8 %	88.1 %
Duplicated BUSCO genes	6.2 %	1.9 %
Fragmented BUSCO genes	5.2 %	5.4 %
Gene model statistics	***T. fasciculata***	***T. leiboldiana***
Gene model count	34,886	38,180
Average length (bp)	4,090	4,225
Complete BUSCO genes	89.7 %	85.3 %
Duplicated BUSCO genes	11.6 %	6.5 %
Fragmented BUSCO genes	5.2 %	7.9 %
Uniquely mapping RNA-seq reads	64.61 %	84.76 %
Gene models with AED-score > 0.5	93 %	89.9 %
Scaffolds containing gene models	1,191	2,621
Statistics of functional annotation	***T. fasciculata***	***T. leiboldiana***
Gene models with BLAST	31,883	33,971
Gene models with GO terms	26,505	27,148
Gene models with Blast2Go annotation	24,319	24,633

Table S3: Abundances of LTR, TIR and Helitron classes in main contigs of *T. fasciculata* and *T. leiboldiana*

		Tillandsia fasciculata			*Tillandsia leiboldiana*		
Class	Type	Element count	Total length (bp)	Proportion of genome	Element count	Total length (bp)	Proportion of genome
	Total	692,254	392,992,866	65.52 %	1,268,380	697,969,745	77.07 %
LTR							
	Copia	85,077	72,241,681	12.04 %	226,304	117,239,901	12.95 %
	Gypsy	136,750	130,003,938	21.67 %	243,366	215,847,337	23.84 %
	Unknown	179,371	110,637,301	18.45 %	357,701	222,072,889	24.52 %
	Total	401,198	312,882,920	52.16 %	827,371	555,160,127	61.31 %
TIR							
	CACTA	26,027	6,822,973	1.14 %	22,788	6,785,096	0.75 %
	Mutator	78,405	20,912,187	3.49 %	227,557	87,286,075	9.64 %
	PIF_Harbinger	13,059	3,335,951	0.56 %	14,124	3,580,678	0.4 %
	Tc1_Mariner	3,435	648,896	0.11 %	1,589	372,253	0.04 %
	hAT	31,871	8,365,061	1.39 %	21,951	6,587,992	0.73 %
	Total	152,797	40,085,068	6.69 %	288,009	104,612,094	11.56 %
Helitron							
	Helitron	118,069	35,165,625	5.86 %	130,790	33,461,880	3.7 %

Table S4: Orthology statistics before and after curation (removal of plastid genes and size correction). Size correction was not applied to *A. comosus*, however, unique genes to this species were removed from the dataset.

	Before curation			After curation		
General and single-copy statistics	*A. comosus*	*T. fasciculata*	*T. leiboldiana*	*A. comosus*	*T. fasciculata*	*T. leiboldiana*
Number of genes assigned to orthogroups	21,045	26,325	23,584	20,416	24,397	22,968
Proportion of input sequences assigned	78 %	87.5 %	75 %	-	-	-
Number of single-copy genes in a given species	12,794	14,311	15,537	12,602	15,236	15,479
Proportion	52.5 %	54.36 %	65.88 %	61.73 %	62.45 %	67.39 %
Number of single-copy orthologues (1:1:1)	10,012	10,012	10,012	10,707	10,707	10,707
Proportion	47.57 %	38.03 %	42.45 %	52.44 %	43.89 %	46.62 %
Number of single-copy orthologues in *Tillandsia*	-	13,128	13,128	-	14,086	14,086
Proportion	-	49.87 %	55.66 %	-	57.74 %	61.33 %
Multi-copy statistics						
Number of multi-copy genes in a given species	8,260	12,014	8,011	7,809	9,161	7,489
Proportion	39.25 %	45.64 %	33.97	38.25 %	37.55 %	32.6 %
Number of genes in orthogroups with family size *T. fas* > *T. lei*	2,594	6,976	2,709	1,178	4,170	1,356
Proportion	12.33 %	26.50 %	11.49 %	5.77 %	17.09 %	5.9 %
Number of genes in orthogroups with family size T. fas < T. lei	817	905	2,222	725	833	2,079
Proportion	3.88 %	3.44 %	9.42 %	35.51 %	34.14 %	9.05 %
Number of genes in orthogroups with family size T. fas = T. lei	1,244	1,265	1,265	1,291	1,258	1,258
Proportion	5.91 %	4.81 %	5,36 %	6.32 %	4.6 %	4.6 %
Unique gene statistics						
Number of unique genes	-	3,101	3,192	-	3,051	3,154
Proportion	-	14.74 %	13.53 %	-	12.51 %	13.73 %

131
132

Table S5: List of potentially CAM-related expanded gene families in either *T. fasciculata* or *T. leiboldiana*.

Orthogroup	Conformation of gene family size (*A.comosus* : *T. fasciculata* : *T. leiboldiana*)	Function	Description	Differentially expressed
OG0005285	1 : 1 : 2	Malate dehydrogenase, cytoplasmic	Catalyses the conversion of malate to oxaloacetate bidirectionally in the cytoplasm	No
OG0002059	2 : 2 : 1	NAD-dependent malate dehydrogenase, mitochondrial	Catalyses the conversion of malate to oxaloacetate bidirectionally in the mitochondrial matrix	No
OG0000469	0 : 15 :1	Cytosolic enolase 3	Catalyses the conversion of 2-phosphoglycerate (2-PG) to phosphoenolpyruvate (PEP)	No
OG0000555	0 : 7 : 1	enolase	Catalyses the conversion of 2-phosphoglycerate (2-PG) to phosphoenolpyruvate (PEP)	No
OG0005044	1 : 2 : 1	Protein XAP5 CIRCADIAN TIMEKEEPER	Involved in the regulation of light response[1], the circadian clock[2], and disease resistance[3].	Yes
OG0000539, OG0004427	2 : 8 : 3 , 1 : 2 : 1	Vacuolar-type proton ATPase subunit H	Subunit of a proton pump involved in the acidification of intracellular organelles. Subunit H has a regulatory role in the activity of the ATPase, not in the assembly.	Yes, No
OG0001440	2 : 2 : 1	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (SDHA)	Subunit of the Succinate-ubiquinone oxidoreductase complex (complex II), which is simultaneously a member of the mitochondrial respiratory chain and of the tricarboxylic acid cycle. SDHA converts succinate to fumarate and FAD to FADH2, therefore playing a role in both pathways. Succinate dehydrogenase has been linked to photosynthetic activity and regulation of stomatal opening in *Solanum* [4].	No
OG0003437	1 : 2 : 1	succinate dehydrogenase subunit 6, mitochondrial	Plant-specific [5] subunit of the succinate dehydrogenase complex involved in anchoring the complex to the membrane[6].	Yes
OG0005172	1 : 2 : 1	isocitrate dehydrogenase [NAD] catalytic subunit 5, mitochondrial	Member of the tricarboxylic acid cycle	Yes
OG0000601	3 : 3 : 2	V-type_proton_ATPase_16_kDa_proteolipid_subunit_c1	Vacuolar proton pump potentially linked to CAM through circadian rhythm regulation and/or malate transport (See Fig. 5)	Yes
OG0002114	1 : 1 : 3	Regulator of V-ATPase in vacuolar membrane protein 1	Regulator of vacuolar proton pump	No
OG0000320	4 : 4 : 5	glyceraldehyde-3-phosphate dehydrogenase 2, cytosolic	Member of the glycolysis	No
OG0001933	2 : 1 : 2	phosphoglycerate kinase, chloroplastic	Member of the glycolysis	No
OG0000507	7 : 1 : 2	Isocitrate_dehydrogenase_NADP_1.1.1.42	Member of the tricarboxylic acid cycle	No

Citations

[1] Ellen L. Martin-Tryon, Stacey L. Harmer, XAP5 CIRCADIAN TIMEKEEPER Coordinates Light Signals for Proper Timing of Photomorphogenesis and the Circadian Clock in Arabidopsis , The Plant Cell, Volume 20, Issue 5, May 2008, Pages 1244–1259, https://doi.org/10.1105/tpc.107.056655

[2] Liu, Lei, Xiaoyun Li, Li Yuan, Guofang Zhang, Hui Gao, Xiaodong Xu, and Hongtao Zhao. "XAP5 CIRCADIAN TIMEKEEPER specifically modulates 3'splice site recognition and is important for circadian clock regulation partly by alternative splicing of LHY and TIC." Plant Physiology and Biochemistry 172 (2022): 151-157.

[3] Xu, Yong-Ju, Yang Lei, Ran Li, Ling-Li Zhang, Zhi-Xue Zhao, Jing-Hao Zhao, Jing Fan et al. "XAP5 CIRCADIAN TIMEKEEPER Positively Regulates RESISTANCE TO POWDERY MILDEW8. 1–Mediated Immunity in Arabidopsis." Frontiers in Plant Science 8 (2017): 2044.

[4] Araújo WL, Nunes-Nesi A, Osorio S, et al. Antisense inhibition of the iron-sulphur subunit of succinate dehydrogenase enhances photosynthesis and growth in tomato via an organic acid-mediated effect on stomatal aperture. Plant Cell. 2011;23(2):600-627. doi:10.1105/tpc.110.081224

[5] Millar AH, Eubel H, Jänsch L, Kruft V, Heazlewood JL, Braun HP. Mitochondrial cytochrome c oxidase and succinate dehydrogenase complexes contain plant specific subunits. Plant Mol Biol. 2004;56(1):77-90. doi:10.1007/s11103-004-2316-2

[6] Christine Schikowsky, Jennifer Senkler, Hans-Peter Braun, SDH6 and SDH7 Contribute to Anchoring Succinate Dehydrogenase to the Inner Mitochondrial Membrane in Arabidopsis thaliana , Plant Physiology, Volume 173, Issue 2, February 2017, Pages 1094–1108, https://doi.org/10.1104/pp.16.01675

133

Table S6: Full list of candidate genes for adaptive sequence evolution.

Orthogroup	Genes	dN/dS	adj p-value	Function
One-to-one orthologues				
OG0002972	Tfasc_v1.01130-RA, Tlei_v1.02341-RA	Inf	0,00162	jacalin-related lectin 3-like
OG0005000	Tfasc_v1.24851-RA, Tlei_v1.26186-RA	Inf	0,00628	cucumber peeling cupredoxin-like
OG0006253	Tfasc_v1.06027-RA, Tlei_v1.01410-RA	Inf	0,00066	mitochondrial prohibitin-3
OG0008124	Tfasc_v1.15280-RB, Tlei_v1.22017-RA	3,1536	0,01368	isoform X1
OG0008528	Tfasc_v1.04577-RA, Tlei_v1.08412-RA	8,17	0,01292	hypothetical protein ACMD2_08159
OG0008977	Tfasc_v1.16327-RA, Tlei_v1.06894-RA	2,6204	0,00512	glutamate receptor 2.8-like
OG0009004	Tfasc_v1.16390-RA, Tlei_v1.06962-RA	3,1379	0,01148	Hydroquinone glycosyltransferase
OG0009278	Tfasc_v1.29917-RA, Tlei_v1.20972-RA	Inf	0,00155	chloroplastic Peroxiredoxin-2E-2
OG0010014	Tfasc_v1.26649-RA, Tlei_v1.28282-RA	2,0042	0,00603	GDPDL7
OG0011786	Tfasc_v1.12421-RA, Tlei_v1.10008-RA	9,2985	0,00008	U-box_domain-containing_protein
OG0012770	Tfasc_v1.15761-RA, Tlei_v1.22494-RA	Inf	0,01374	superfamily protein
OG0015603	Tfasc_v1.14595-RA, Tlei_v1.21364-RA	8,0697	0,01575	CDC27
OG0019176	Tfasc_v1.13382-RA, Tlei_v1.18537-RA	2,8534	0,00561	uncharacterized protein LOC109718296
1:1:2 orthologues				
OG0003849	Tfasc_v1.03397-RA, Tlei_v1.09369-RA	Inf	0,00697	kinesin-like protein KIN-10C
	Tfasc_v1.03397-RA, Tlei_v1.09368-RA	4,1684	0,09607	
1:2:1 orthologues				
OG0003795	Tfasc_v1.16138-RA, Tlei_v1.06694-RA	6,2573	0,02611	Ubiquitin-conjugating enzyme E2
	Tfasc_v1.16140-RA, Tlei_v1.06694-RA	0,2438	0,03386	
OG0004404	Tfasc_v1.08653-RA, Tlei_v1.17093-RA	3,9913	0,02787	protein ECERIFERUM 1-like
	Tfasc_v1.21655-RA, Tlei_v1.17093-RA	0,1733	0,00059	

Table S7: Description of gene clusters inferred in co-expression analyses from maSigPro

Cluster	Number of genes	Genes of interest	GO terms of interest
1	87	Tfasc_v1.23066: protein LNK1-like; Tfasc_v1.19779: Dicarboxylate transporter 1, chloroplastic; Tfasc_v1.08739: mitochondrial uncoupling protein 5-like; Tfasc_v1.00158: ABC transporter G family member 5-like; Tfasc_v1.01823: Protein REVEILLE 1 [1]; Tfasc_v1.03880: Transcription factor PCF2 [2]; Tfasc_v1.06881*: protein LHY-like isoform X1 [2]; Tfasc_v1.09028: aluminum-activated malate transporter 9-like isoform X1 [1]	GO:1902356: oxaloacetate(2-) transmembrane transport; GO:0071423: malate transmembrane transport; GO:1902074: response to salt; GO:0071472: cellular response to salt stress; GO:1902584: positive regulation of response to water deprivation; GO:1901002: positive regulation of response to salt stress; GO:0015131: oxaloacetate transmembrane transporter activity; GO:0015140: malate transmembrane transporter activity
2	134	Tfasc_v1.25154: V-type proton ATPase catalytic subunit A [3]; Tfasc_v1.11797: protein XAP5 CIRCADIAN TIMEKEEPER [3]; Tfasc_v1.21051: V-type proton ATPase subunit e1; Tfasc_v1.15469*: Acyl-coenzyme A thioesterase [2]; Tfasc_v1.09221: Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha; Tfasc_v1.12690*: pentatricopeptide repeat-containing protein At1g09900-like [2];	GO:0033179: proton-transporting V-type ATPase, V0 domain; GO:0047334: diphosphate-fructose-6-phosphate 1-phosphotransferase activity
3	38	Tfasc_v1.09150: protein HOMOLOG OF MAMMALIAN LYST-INTERACTING PROTEIN 5; Tfasc_v1.03774: probable aquaporin PIP2-6 [4]; Tfasc_v1.14176*: long chain acyl-CoA synthetase 4-like [2]; Tfasc_v1.24696: V-type proton ATPase subunit H-like [5]; Tfasc_v1.25341: pyrophosphate-energized vacuolar membrane proton pump [3]; Tfasc_v1.28097: F-box/kelch-repeat protein SKIP25 [2]	GO:0007623: circadian rhythm; GO:0010378: temperature compensation of the circadian clock; GO:0046323: glucose import; GO:0009637: response to blue light; GO:0048578: positive regulation of long-day photoperiodism, flowering; GO:0071482: cellular response to light stimulus
4	99	Tfasc_v1.09150: protein HOMOLOG OF MAMMALIAN LYST-INTERACTING PROTEIN 5; Tfasc_v1.03774: probable aquaporin PIP2-6 [4]; Tfasc_v1.14176: long chain acyl-CoA synthetase 4-like [2]; Tfasc_v1.24696: V-type proton ATPase subunit H-like [5]; Tfasc_v1.25341: pyrophosphate-energized vacuolar membrane proton pump [3]; Tfasc_v1.28097: F-box/kelch-repeat protein SKIP25 [2]	GO:1903335: regulation of vacuolar transport; GO:0000221: vacuolar proton-transporting V-type ATPase, V1 domain
5	209	Tfasc_v1.01733: V-type proton ATPase 16 kDa proteolipid subunit [2]; Tfasc_v1.16595: soluble starch synthase I; Tfasc_v1.21461: Pyruvate kinase, cytosolic isozyme [2][6]; Tfasc_v1.16311: Phosphoenolpyruvate carboxylase (PEPC) [7]; Tfasc_v1.03126: malate dehydrogenase [7]; Tfasc_v1.03128: PEPC kinase [7]; Tfasc_v1.06749: phosphoglucan phosphatase DSP4, amyloplastic isoform X1; Tfasc_v1.09378: V-type proton ATPase subunit d2 [2][5]; Tfasc_v1.20370: ATP-dependent 6-phosphofructokinase 5, chloroplastic; Tfasc_v1.24086: ATP-dependent 6-phosphofructokinase 3; Tfasc_v1.17156: aconitate hydratase; Tfasc_v1.04514: glucose-6-phosphate isomerase, cytosolic; Tfasc_v1.07899: phosphoenolpyruvate carboxykinase (ATP); Tfasc_v1.26280: protein MAEA homolog; Tfasc_v1.27863: pyruvate decarboxylase 1; Tfasc_v1.03307: V-type proton ATPase subunit a3 [5]; Tfasc_v1.00598: nuclear pore complex protein NUP50A-like [2]; Tfasc_v1.01511: late embryogenesis abundant protein Lea5-like [8]; Tfasc_v1.03613*: probable pyruvate, phosphate dikinase regulatory protein, chloroplastic [2][7]; Tfasc_v1.14299: late embryogenesis abundant protein group 8 protein [8]; Tfasc_v1.14724: V-type proton ATPase subunit B 2 [3]; Tfasc_v1.15525: protein YLS7-like [2]; Tfasc_v1.16080: V-type proton ATPase subunit c"2 [2][5]; Tfasc_v1.17712: V-type proton ATPase subunit C [3]; Tfasc_v1.18209: V-type proton ATPase subunit G 1-like [3]; Tfasc_v1.19718: F-box protein At1g55000 [2]; Tfasc_v1.19742: V-type proton ATPase subunit H [5] Tfasc_v1.26860: gamma-aminobutyrate transaminase 1, mitochondrial [2]; Tfasc_v1.28537: protein FD [9]; Tfasc_v1.28610: uncharacterized protein LOC109727440 [9]	GO:0005983: starch catabolic process; GO:0061615: glycolytic process through fructose-6-phosphate; GO:0006099: tricarboxylic acid cycle; GO:0006094: gluconeogenesis; GO:0006002: fructose 6-phosphate metabolic process; GO:0045721: negative regulation of gluconeogenesis; GO:0007035: vacuolar acidification; GO:0006107: oxaloacetate metabolic process; GO:0004737: pyruvate decarboxylase activity; GO:0004612: phosphoenolpyruvate carboxykinase (ATP) activity; GO:0008964: phosphoenolpyruvate carboxylase activity; GO:0046961: proton-transporting ATPase activity, rotational mechanism; GO:0047780: citrate dehydratase activity; GO:0004347: glucose-6-phosphate isomerase activity
6	144	Tfasc_v1.20354: sucrose transport protein SUT1-like [3]; Tfasc_v1.25107: VMA21-like domain-containing protein; Tfasc_v1.06152*: Protein EDS1L [2]; Tfasc_v1.08627: WAT1-related protein [2]; Tfasc_v1.16433: arabinogalactan peptide 16-like [2]; Tfasc_v1.24231*: replication stress response regulator SDE2 [2]; Tfasc_v1.30364*: protein FLX-like 4 [2]; Tfasc_v1.25107: VMA21-like domain-containing protein; Tfasc_v1.26899: obg-like ATPase 1; Tfasc_v1.25528: granule-bound starch synthase	GO:0015770: sucrose transport; GO:0015768: maltose transport; GO:0070072: vacuolar proton-transporting V-type ATPase complex assembly; GO:1901001: negative regulation of response to salt stress; GO:0005364: maltose:proton symporter activity; GO:0003985: acetyl-CoA C-acetyltransferase activity; GO:0004373: glycogen (starch) synthase activity
7	196	Tfasc_v1.30467: isocitrate dehydrogenase [NAD] catalytic subunit 5, mitochondrial; Tfasc_v1.21339: catalase isozyme 1; Tfasc_v1.12645: flavonoid 3',5'-hydroxylase 2-like [2]; Tfasc_v1.25055*: berberine bridge enzyme-like 18 [2]; Tfasc_v1.29220: ABC transporter C family member 14-like [3]	GO:1902074: response to salt; GO:1900034: regulation of cellular response to heat; GO:0004449: isocitrate dehydrogenase (NAD+) activity
	*	Candidate gene for adaptive sequence evolution in CAM/C3 shifts reported in [2]	

Citations

[1]	Wai, Ching M. et al. 2017. "Temporal and Spatial Transcriptomic and MicroRNA Dynamics of CAM Photosynthesis in Pineapple." The Plant Journal 92(1): 19–30.
[2]	De La Harpe, M. et al. 2020. "Genomic footprints of repeated evolution of CAM photosynthesis in a Neotropical species radiation." Plant, Cell and Environment 43(12): 2987-3001.
[3]	McClung, C. Robertson. 2006. "Plant Circadian Rhythms." The Plant Cell 18(4): 792–803.
[4]	Vera-Estrella, R. et al. 2012. "Day/Night Regulation of Aquaporins during the CAM Cycle in Mesembryanthemum Crystallinum." Plant, Cell & Environment 35(3): 485–501.
[5]	Cosentino, Cristian et al. 2013. "Proteomic Analysis of Mesembryanthemum Crystallinum Leaf Microsomal Fractions Finds an Imbalance in V-ATPase Stoichiometry during the Salt-Induced Transition from C3 to CAM." Biochemical Journal 450(2): 407–15.
[6]	Cushman, John C. et al. 2008. "Large-Scale MRNA Expression Profiling in the Common Ice Plant, Mesembryanthemum Crystallinum, Performing C3 Photosynthesis and Crassulacean Acid Metabolism (CAM)." Journal of Experimental Botany 59(7): 1875–94.
[7]	Ming, Ray et al. 2015. "The Pineapple Genome and the Evolution of CAM Photosynthesis." Nature Genetics 47(12): 1435–42.
[8]	Xiao, B., et al. 2007. "Over-Expression of a LEA Gene in Rice Improves Drought Resistance under the Field Conditions." Theoretical and Applied Genetics 115(1): 35–46.
[9]	Christin, Pascal-Antoine et al. 2014. "Shared Origins of a Key Enzyme during the Evolution of C4 and CAM Metabolism." Journal of Experimental Botany 65(13): 3609–21.

3. Supplementary Notes

Note 1: Genome size and karyotype of *T. fasciculata* and *T. leiboldiana*

The genome size of both specimens used for *de novo* assembly was measured with flow cytometry (See Methods, section 1.1.), estimating genome sizes of approximately 790 and 1,130 Mb for *Tillandsia fasciculata* and *T. leiboldiana* respectively (Fig. S8). These estimates are slightly higher than those obtained computationally with a kmer-based approach implemented in findGSE[1] (k = 21 ,701 and 1,125 Mb respectively, Fig. S10); but deviations of genome size in computational approaches have been reported frequently[1–3]. We also obtained a karyotype for both species using root material (See Methods, section 5.1.2.). We observe a change in karyotype between the two species resulting in a reduction by six chromosome pairs in *T. leiboldiana* compared to *T. fasciculata*, which carries the base karyotype of 2n = 50 encountered in *Tillandsioideae*[4] (Fig. S9). This is in accordance with chromosome counts reported by Brown and Gilmartin (1989)[4].

Note 2: Pre-assembly estimation of per-accession heterozygosity

Heterozygosity estimates of the chosen accessions along with several other candidate accessions from the Botanical Gardens of the University of Vienna were obtained with short-read Illumina data before *de novo* assembly, with the aim to select accessions with lowest heterozygosity and to make adjustments during *de novo* assembly to account for potentially elevated rates of heterozygosity. This short-read data was later used for polishing purposes and the sequencing details can be found in the Methods section *Plant material selection and sequencing*. Heterozygosity estimates were obtained by two approaches: a k-mer based approach and a reference-based approach. For the k-mer based approach, findGSE[5] was used with a k = 21 to obtain k-mer peaks of heterozygosity. For the reference-based approach, reads were trimmed with TrimmOmatic[6] and mapped with GSNAP[7] to the *Tillandsia adpressiflora* pseudoreference built by De la Harpe and colleagues[8]. After filtering for low mapping quality and marking duplicates, variants were called for all accessions using *freebayes*[9]. Variants with an individual depth under 5 and above 45 were removed and no missing data was retained. Then, heterozygous sites per 1000 mappable sites were counted with a custom-made python script, filtering for allele balance

between 0.25 and 0.75. This yielded between 51,000 and 58,000 windows, translating to roughly a 50 Mb portion of the genome. We expect these windows to be enriched for genic and other conserved regions, therefore resulting in an underestimate, though for relative comparisons, we regard this approach as valid. Both the k-mer (Fig. S10) and reference-based approach (Fig. S11) showed that heterozygosity is elevated in *T. fasciculata* compared to *T. leiboldiana*. This result is consistent with the nucleotide diversity estimates reported by Yardeni et al.[10] based on sequencing of 1776 targeted loci in several Tillandsia species, including *T. leiboldiana* (π_S=5.7x10^{-3}) and *T. fasciculata* (π_S=8.1x10^{-3}). The relatively moderate levels of diversity in these species helped us to obtain assemblies with such a remarkable contiguity (Table S2), despite their very high repetitive content (see Supplementary Note 4).

Note 3: Identifying main scaffolds in de novo assembly

After scaffolding with Hi-C data, the resulting *de novo* assemblies contained a total of 2,321 and 10,443 scaffolds (> 1 kb) for *T. fasciculata* and *T. leiboldiana* respectively. However, more than 99 % of one-to-one orthologous gene pairs are located on the 25 and 26 largest scaffolds of both assemblies respectively, while 90.7 % and 87.6 % of all gene models are on these scaffolds. Therefore, the remaining scaffolds mainly consist of repetitive content, virtually corresponding to short, duplicated regions in the assembly. The mean proportion of repetitive content in the remaining scaffolds is indeed much higher than in main scaffolds (94.9 % and 91 % in small scaffolds of *T. fasciculata* and *T. leiboldiana*, versus 65.5 and 77.1% in the "main" scaffolds (see main text)). In addition, these "main" scaffolds contain the vast majority of the assembly (72 % and 75.5 % of the total assembly length). While the sizes of the longest 25 and 24 scaffolds are over 1 Mb, other scaffold sizes steeply decline afterwards (Fig. S1). Given this, we decided to regard these scaffolds as representative for the respective genomes and excluded all secondary scaffolds from downstream analyses from this point onwards. Though scaffolds 26 and 25 in *T. leiboldiana* are smaller than 1 Mb, they contain a substantial number of orthologous genes (Fig. S1) and were therefore maintained in all analyses.

Note 4: On the success of *de novo* assembly of highly repetitive genomes

The highly repetitive content observed in many plant genomes often causes fragmented genome assemblies. Despite considerable progress thanks to long-read sequencing technologies[13], little to no genomic resources are available yet for some plant clades, particularly for species with the remaining challenge of a highly repetitive content[10]. The availability of long-read sequencing and chromatin conformation capture technologies have now enabled the assembly of particularly complex genomes with little fragmentation, in the best case at the chromosome-level. Our project was therefore launched and made possible because of these recent technological advances.

The kmer-based approach implemented in findGSE[1] (k = 21, see also SI Note 1 and 2) estimates the TE content directly from raw reads, *i.e.,* prior to generating *de novo* assemblies. We estimated that the repetitive content of *T. fasciculata* and *T. leiboldiana* was around 66 % and 75 %, respectively. Based on *de novo* generated assemblies and TE annotations performed with EDTA, we observed remarkably consistent estimates (65.5 % and 77.1 % in *T. fasciculata* and *T. leiboldiana*, Table S3) on the main scaffolds (SI Note 3). Such values are one of the most elevated among plant genomes assembled at chromosome scale[11].

Additionally, our analyses of spatial distribution of TE content highlight the extreme local levels of repetitive content in these genomes, especially in *T. leiboldiana*. In non-telomeric regions, the observed repetitive content most often reaches values above 80% for *T. fasciculata* and nearly 100% for *T. leiboldiana* (Fig. 2b, Fig. S2).

Considering the extremely high repetitive content in centromeric regions, especially for *T. leiboldiana*, the limited fragmentation of our assembly can be considered as a success and therefore represent empirical evidence of the progress made in plant genomics, only twenty years after the release of the first plant genome. At the age of long-read sequencing and chromatin conformation capture technologies, the *de novo* sequencing of plant species associated with a highly repeated genomic content is becoming more and more feasible.

Note 5: On the spatial distribution of GC and TE content in bromeliad genomes

After finding that GC and genic content were negatively correlated in both *Tillandsia* genomes (See Results), we decided to study the link between GC content and repetitive content in all bromeliad genomes available to us at the time (*A. comosus*, *T. fasciculata* and *T. leiboldiana*) Using softmasked versions for all three genomes (see Materials and Methods), we computed the proportion of soft-masked bases across 100 kb windows. We also computed the overall GC content (considering both softmasked and non-softmasked positions), the GC content for soft-masked bases only, and the GC content for non-softmasked bases in the same windows. Based on this, the average TE content was estimated to be of 36.7 %, 67.9 % and 79.1% for *A. comosus*, *T. fasciculata* and *T. leiboldiana*, respectively. These three genomes therefore represent a gradient regarding the amount of TEs. After having reported the relatively well conserved synteny (Fig. 2c), we were able to estimate the evolution of the repetitive, GC and genic chromosomal landscapes across syntenic chromosomes. We selected three examples of syntenic triplets considering scaffolds with no main chromosomal rearrangements: triplets A.com 3 / T.fas 4 / T.lei 1 (Fig. 2b), A.com 6 / T.fas 11 / Tlei. 15 and A.com 11 / T.fas 12 / T.lei 5 (Fig. S2). We then considered a relative position of each window on the scaffold (window position/scaffold length) to account for the difference in length of the syntenic chromosomes in the three assemblies. From this visualization, it became clear that the GC content landscape is largely shaped by TE dynamics in the three genomes, since the GC% at non-repetitive content show no to little variation across the scaffold, except in *T. leiboldiana*. This contributed to building large GC-rich isochores in centromeric regions. Note here that our *de novo* TE libraries are not necessarily exhaustive and therefore a part of the repetitive content may have remained non-softmasked, which could partly or completely explain the pattern observed for GC at non-softmasked position in *T. leiboldiana*. Since this pattern is observed in all three species, regardless of the difference in repetitive content, this may be a family-wide phenomenon.

Note 6: Identifying large-scale rearrangements between *T. fasciculata* and *T. leiboldiana*

Large-scale rearrangements were first perceived in the synteny analysis of the two assemblies. These were further investigated by whole-genome alignment of the two assemblies to each other, and to *A. comosus*, using nucmer[12] and visualised with Dot (https://github.com/dnanexus/dot).

Large rearrangements that were identified between *T. fasciculata* and *T. leiboldiana* were further investigated by performing LastZ alignments[13] of soft-masked genomes. LastZ was run with the following settings: --notransition, --step=10, --gapped, --chain, --gfextend, --format=maf. Local alignments were filtered by the 95th identity and length quartile as implemented in Leroy et al (2021)[14]. Additionally, alignments were filtered by uniqueness with a custom-made python script, by removing all alignments with more than a 90 % overlap. Final local alignments were visualised for each scaffold as in Leroy et al (2021). All scripts are available at: https://shorturl.at/xLS15.

The breakpoint area of confirmed rearrangements was defined as the region between the last alignment of a given scaffold and the first alignment of another scaffold. Rearrangements were then finally confirmed by investigating the alignment of long-read PacBio data to the assembly of the same species. Whenever no clear break could be identified in the long-read alignment within the breakpoint area, the rearrangement was considered as confirmed.

With these methods, we described three potential large-scale rearrangements. Scaffold 14 in *T. leiboldiana* could be a fusion of scaffolds 17 and 25 in *T. fasciculata*, or scaffolds 17 and 25 could be the result of a fission in a reversed scenario (Fig. S3a). We also detected two potential translocations (Fig. S3b for Translocation 1 and Fig. S3b for Translocation 2). All breakpoints were confirmed by alignment of raw PacBio reads, except for breakpoint 1 on scaffold 13 of *T. fasciculata* of Translocation 1 (Fig. S3b). However, breakpoint 2 on this scaffold was confirmed, which led us to maintain the translocation as a candidate rearrangement. For several scaffolds, local alignments were too sparse to determine a clear breakpoint (See Fig. S3). To see the PacBio alignment at each breakpoint, see the supplementary PDF file "Tfas_Tlei_rearrangements.pdf" on our github repository.

We studied the effects of large-scale rearrangement on the genomic distribution of d_N/d_S values and, sperarately, on DE genes (See SI Note 11) to understand if there is a link between

chromosomal and functional evolution in *Tillandsia*. We did this by testing whether the distribution of d_N/d_S values in any of the rearranged chromosomes deviated from that of non-rearranged chromosomes (See Methods, section 9.1). Of the nine scaffolds involved in the three reported rearrangements, only two had a d_N/d_S distribution significantly deviating from that of non-rearranged chromosomes (scaffold 13 in *T. fasciculata* and scaffold 19 in *T. leiboldiana*, see Fig. S4b). The d_N/d_S values in these scaffolds, which are both involved in Translocation 1, were ever so slightly reduced compared to non-rearranged scaffolds (median scaffold 13 = 0.3197, other scaffolds in *T. fasciculata* = 0.3569; median scaffold 19 = 0.3326, other scaffolds in *T. leiboldiana* = 0.3591). An overall reduction in chromosome-wide d_N/d_S values can be expected in rearranged chromosomes due to an increase in linkage disequilibrium resulting from recombination suppression, which in turn increases background selection[15,16]. This can have important implications for both adaptation and speciation, as functional loci may be under increased selective constraint, and selection against introgression may also become stronger[16]. However, the significance of this reduction in d_N/d_S is very slight and only visible in two of nine rearranged scaffolds. Combined with our results on DE gene distribution across the genome, which show no signal of rearrangements playing a role in spatial distribution of ecologically relevant genes (See SI Note 11), we are cautious in heralding large-scale rearrangement as a key driving force of ecological diversification in *Tillandsia* until additional supporting evidence becomes available (See SI Note 12).

Note 7: Correcting multi-copy gene family sizes

Gene counts per orthogroup were evaluated using per-gene mean coverage to detect potential haplotig gene sequences that may have escaped Purge Haplotigs in the assembly step. To do this, whole-genome Illumina reads of both species (See Methods, section 2.1.) were aligned to their respective assemblies using Bowtie2[17] with the very-sensitive-local option. Bowtie2 specifically assigns multi-mapping reads randomly, allowing the detection of artificial gene models thanks to a decreased overall coverage across the orthogroup, as reads from one biological copy are randomly distributed over two or more locations in the genome. Per-base coverage in genic regions was calculated using samtools depth and a bedfile specifying all locations of orthologous genes. We then calculated the average coverage per orthologous gene.

The distribution of per-gene mean coverage in each species' gene model set was then visualized using ggplot2[18] for different categories of genes: single-copy (only one gene model assigned to the orthogroup in the species investigated), multi-copy (more than one gene assigned to the orthogroup in the species investigated), ancestral single-copy (only one gene model assigned to the orthogroup in all species used in the orthology analysis) and ancestral multi-copy (multiple gene model assigned to the orthogroup in all species used in the orthology analysis and the number of gene models assigned is equal across species). This revealed that, while most categories of genes had a unimodal distribution centered around the average coverage across the genome, multi-copy and unique multi-copy families showed a bimodal or expanded distribution, especially in *T. fasciculata* (Fig. S12). This points at the presence of false gene copies in the annotation.

Gene count sizes per orthogroup and species were therefore corrected by the ratio of the total coverage across all genes of one species in the orthogroup and the expected coverage, which was calculated as the product of the total number of genes in the orthogroup and the average coverage of single-copy genes in that species.

Size corrections were only applied on orthogroups containing multicopy genes. Plastid and mitochondrial genes were excluded from this analysis. We detected plastid genes with BLASTn against the *A. comosus* chloroplast sequence and the *Oryza* IRSGP-1 mitochondrial sequence. Additionally, all genes annotated as "ribosomal" were also excluded from the downstream gene family evolution analyses.

Originally, 9,210 genes in *T. fasciculata* and 6,257 genes in *T. leiboldiana* were assigned to orthogroups with multiple gene copies in at least one species. After correcting orthogroup sizes by coverage, we retained 6,261 and 4,693 gene models, respectively (Table S4).

Note 8: Selecting rapidly evolving gene families

To better understand the distribution of gene size differences, the log-ratio was taken of *T. fasciculata* to *T. leiboldiana* gene counts, and the overall mean log-ratio was subtracted to correct for background rates of gene loss or duplication. Orthogroups were ranked by corrected log-ratios and the top and bottom 2 % were then selected for further analysis. Due to the relatively large proportion of one-to-one relationships (79 %) among orthogroups, all orthogroups with a family size change between *T. fasciculata* and *T. leiboldiana* were included in the top 2 % of gene changes

and therefore selected for GO term enrichment, which was performed separately for orthogroups with gene count larger in *T. fasciculata* (916 orthogroups) and larger in *T. leiboldiana* (583 orthogroups).

Note 9: Detailed description of candidate genes for positive selection

Our d_N/d_S calculations pointed at 13 single-copy and 3 multi-copy genes exhibiting signatures of divergent selection between *T. fasciculata* and *T. leiboldiana*. The most relevant genes have been described in the main text and in Table 2, but here we provide more information on all candidate genes, which could be interesting for future work to investigate speciation genes in *Tillandsia*.

Among single-copy candidates, we found a Jacalin-related lectin (JLR, OG0002972), which are often associated with biotic and abiotic stimuli, though their biological function is largely unknown. In wheat, a mannose-specific JLR has been identified as a component of the defence system[19]. In rice, a JLR has been described as playing a role in salt stress response[20].

Orthogroup OG0005000 codes for a cupredoxin (cupredoxin cucumber peeling-like). Cupredoxins are small proteins containing a copper centre which function as electron transfer shuttles between redox partners, but their more specific biological function is largely unknown[21]. However, they tend to play a role in respiration, photosynthesis and metabolism[22], and therefore represent another interesting candidate for further investigations.

Another candidate for adaptive sequence evolution was a mitochondrial prohibitin-3 (OG0006253), a subunit of the prohibitin complex. While the exact mechanism of prohibitin is unknown, it has been associated with mitochondrial biogenesis in *Nicotiana benthamiana*[23] and more specifically protection against salt stress in *Arabidopsis thaliana*[24].

We recovered a hydroquinone glycosyltransferase (OG0009004), a broad-spectrum glycosyltransferase involved in the secondary metabolism of many phenolic compounds and xenobiotics[25].

Orthogroup OG0009278 codes for chloroplastic Peroxiredoxin-2E-2, which is a member of the thiol peroxidase family. These enzymes play an important role in regulating Reactive Oxygen Species (ROS) by reducing hydroperoxides. Peroxiredoxin-2E-2 is present in

chloroplasts, especially in reproductive tissues, and its expression is sensitive to light and salt levels[26].

The remaining single-copy orthogroups that are candidates for adaptive sequence evolution are involved in cell replication (OG0015603) or members of a broad gene superfamily (OG0012770).

In addition to single copy genes, we tested for adaptive sequence evolution in orthogroups with a 1:1:2 or 1:2:1 relationship, *i.e.* a single gene in *A. comosus* and a duplicated gene either in *T. leiboldiana* (1:1:2 , 108 genes), or *T. fasciculata* (1:2:1, 190). We recovered a gene family in 1:2:1 conformation coding for a protein ECIFERUM 1-like, where one of two copies in *T. fasciculata* had $\omega > 1$. In *A. thaliana,* protein ECIFERUM-1 is involved in the biosynthesis of alkanes, which form hydrophobic cuticular waxes that protect the plant from desiccation. It has been shown that changes in expression of ECIFERUM-1 affect susceptibility to water stress and pathogens, therefore linking the protein with responses to biotic and abiotic stress[27].

We also recovered one candidate orthogroup in 1:1:2 conformation coding for a kinesin-like protein KIN-10C. Members of the kinesin superfamily are molecular motors playing important roles in intracellular transport of vesicles and organelles, spindle formation and elongation, chromosome segregation, morphogenesis, and signal transduction[28]. Both *T. leiboldiana* gene copies showed elevated d_N/d_S ratios (See Table 1), though only one ratio was significant, suggesting that both gene copies have undergone significant evolution in *T. leiboldiana*.

Another candidate orthogroup in 1:2:1 conformation codes for a Ubiquitin-conjugating (UBC) enzyme E2, which plays an important role in the targeting of proteins by ubiquination for the proteasome. In mung bean, a UBC E2 enhances osmotic stress tolerance[29], and in Arabidopsis the overexpression of a soy bean[30] and peanut[31] UBC E2 protein increases drought and salt tolerance.

Note 10: Differential gene expression using the *T. leiboldiana* assembly

In addition to the DE analysis using the *T. fasciculata* genome as reference, we performed a second DE analysis with the *T. leiboldiana* genome as reference to test whether the one-directional enrichment of multi-copy gene families is the result of a technical bias when using the *T. fasciculata* genome for DE analysis. It is indeed possible that differential gene expression in

additional copies of *T. leiboldiana* are missed, as these are not present in the *T. fasciculata* genome and may be too divergent to map onto a different copy. We indeed find enrichment for gene families with gene counts higher in *T. leiboldiana*, which occur twice as much compared to the whole genome (Chi-square P = $1.011568e^{-33}$, Table 4). We also find a small increase of multicopy families with higher gene counts in *T. fasciculata* compared to the whole genome when using mapped reads to *T. leiboldiana* (See Results).

Note 11: Distribution of DE genes across the genome

Using only robust gene annotations, we calculated the relative density of DE genes in 1 Mb windows across each genome, by dividing the DE gene count in each window by the total gene count. The result was then visualised with the R package circlize[32] (Fig. S13). Across the *T. fasciculata* genome, differentially expressed genes follow a similar distribution as all other genes, and we do not detect any peaks of high DE gene density. In *T. leiboldiana*, regions with elevated DE gene density can be seen on most chromosomes, though these are in interior regions where the total gene count in a window is generally low. Therefore, these peaks appear inflated by low sample size. We don't see any clear signal of increased DE density in rearranged chromosomes.

Note 12: Limitations of this study and future directions

A few limitations of this study should be mentioned. First, a pairwise framework limits the array of methods available, and especially restricts us from using methods that rely on a phylogeny. Especially the tests of adaptive sequence evolution would benefit from further investigations with methods relying on a phylogenetic framework, to better understand which species experienced positive selection, and whether these cases of selection are repeated across CAM/C3 shifts. Despite the availability of the *A. comosus* reference genome, this species is too divergent from *Tillandsia*, making alignments of a notable portion of the genes unreliable. Additional genomic resources will be needed, but our current work paves the way towards achieving this goal. Interestingly, though we found no overlap between DE genes and genes undergoing adaptive sequence evolution, almost half of the genes identified as candidates of positive selection in CAM lineages after a CAM/C3

446 shift in a previous study using a phylogenetic framework were differentially expressed in this
447 study[8].
448 Secondly, our investigations on large-scale rearrangements were limited due to the available
449 resources and remaining fragmentation in one of our *de novo* assemblies. While the *T. fasciculata*
450 *de novo* assembly recovered all chromosomes in individual scaffolds, this is not the case for *T.*
451 *leiboldiana*, which still has 26 main scaffolds despite counting 19 haploid chromosomes. This
452 remaining fragmentation limited our study of rearrangement in *T. leiboldiana*, especially regarding
453 the karyotype differences between both species, despite the detection of one fused scaffold. To
454 further investigate the impact of these large-scale rearrangements on *Tillandsia* evolution,
455 improvement of the *T. leiboldiana* assembly with additional genomic data will be necessary.
456 Additionally, our analyses on the effect of large-scale rearrangement on functional sites (SI Note
457 6) could also be expanded by a population-level study, which would allow for measurement of
458 per-scaffold nuclear diversity and recombination rates, which may point at the role of
459 rearrangement in shaping chromosome-wide recombination and selection landscapes. This would
460 provide additional evidence that large-scale rearrangements played a role in the *Tillandsia*
461 radiation, something that has already been reported in other radiating lineages, but that we can so
462 far not confidently state for here. Since the cost of genome assemblies and sequencing is rapidly
463 decreasing, a future large-scale study of rearrangements across *Tillandsia* including more species
464 and accessions is becoming feasible.
465

466 **4. References**

467 1. Al-Qurainy, F. *et al.* Estimation of Genome Size in the Endemic Species Reseda pentagyna
468 and the Locally Rare Species Reseda lutea Using comparative Analyses of Flow Cytometry
469 and K-Mer Approaches. *Plants* **10**, (2021).

470 2. Pflug, J. M., Holmes, V. R., Burrus, C., Johnston, J. S. & Maddison, D. R. Measuring Genome
471 Sizes Using Read-Depth, k-mers, and Flow Cytometry: Methodological Comparisons in
472 Beetles (Coleoptera). *G3* **10**, 3047–3060 (2020).

473 3. Elliott, T. A. & Gregory, T. R. What's in a genome? The C-value enigma and the evolution
474 of eukaryotic genome content. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140331 (2015).

4. Brown, G. K. & Gilmartin, A. J. CHROMOSOME NUMBERS IN BROMELIACEAE. *Am. J. Bot.* **76**, 657–665 (1989).

5. Sun, H., Ding, J., Piednoël, M. & Schneeberger, K. findGSE: estimating genome size variation within human and Arabidopsis using k-mer frequencies. *Bioinformatics* **34**, 550–557 (2018).

6. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

7. Wu, T. D., Reeder, J., Lawrence, M., Becker, G. & Brauer, M. J. GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. in *Statistical Genomics: Methods and Protocols* (eds. Mathé, E. & Davis, S.) 283–334 (Springer New York, 2016).

8. De La Harpe, M. *et al.* Genomic footprints of repeated evolution of CAM photosynthesis in a Neotropical species radiation. *Plant Cell Environ.* ce.13847 (2020).

9. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv [q-bio.GN]* (2012).

10. Marks, R. A., Hotaling, S., Frandsen, P. B. & VanBuren, R. Representation and participation across 20 years of plant genome sequencing. *Nat Plants* **7**, 1571–1578 (2021).

11. Pedro, D. L. F. *et al.* An Atlas of Plant Transposable Elements. *F1000Res.* **10**, 1194 (2021).

12. Delcher, A. L., Phillippy, A., Carlton, J. & Salzberg, S. L. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* **30**, 2478–2483 (2002).

13. Harris, R. S. Improved pairwise alignment of genomic DNA. (The Pennsylvania State University, 2007).

14. Leroy, T. *et al.* A bird's white-eye view on avian sex chromosome evolution. Preprint at https://doi.org/10.24072/pcjournal.70 (2021).

15. Mérot, C., Oomen, R. A., Tigano, A. & Wellenreuther, M. A Roadmap for Understanding the

Evolutionary Significance of Structural Genomic Variation. *Trends Ecol. Evol.* **35**, 561–572 (2020).

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

17. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

18. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).

19. Xiang, Y. *et al.* A jacalin-related lectin-like gene in wheat is a component of the plant defence system. *J. Exp. Bot.* **62**, 5471–5483 (2011).

20. Zhang, W. *et al.* Isolation and characterization of a jacalin-related mannose-binding lectin from salt-stressed rice (Oryza sativa) plants. *Planta* **210**, 970–978 (2000).

21. Guss, J. M., Merritt, E. A., Phizackerley, R. P. & Freeman, H. C. The structure of a phytocyanin, the basic blue protein from cucumber, refined at 1.8 A resolution. *J. Mol. Biol.* **262**, 686–705 (1996).

22. Choi, M. & Davidson, V. L. Cupredoxins--a study of how proteins may evolve to use metals for bioenergetic processes. *Metallomics* **3**, 140–151 (2011).

23. Ahn, C. S., Lee, J. H., Reum Hwang, A., Kim, W. T. & Pai, H.-S. Prohibitin is involved in mitochondrial biogenesis in plants. *Plant J.* **46**, 658–667 (2006).

24. Wang, Y., Ries, A., Wu, K., Yang, A. & Crawford, N. M. The Arabidopsis Prohibitin Gene PHB3 Functions in Nitric Oxide–Mediated Responses and in Hydrogen Peroxide–Induced Nitric Oxide Accumulation. *Plant Cell* **22**, 249–259 (2012).

25. Hefner, T., Arend, J., Warzecha, H., Siems, K. & Stöckigt, J. Arbutin synthase, a novel member of the NRD1beta glycosyltransferase family, is a unique multifunctional enzyme converting various natural products and xenobiotics. *Bioorg. Med. Chem.* **10**, 1731–1741

(2002).

26. Rouhier, N. & Jacquot, J.-P. The plant multigenic family of thiol peroxidases. *Free Radic. Biol. Med.* **38**, 1413–1421 (2005).

27. Bourdenx, B. *et al.* Overexpression of Arabidopsis ECERIFERUM1 promotes wax very-long-chain alkane biosynthesis and influences plant response to biotic and abiotic stresses. *Plant Physiol.* **156**, 29–45 (2011).

28. Li, J., Xu, Y. & Chong, K. The novel functions of kinesin motor proteins in plants. *Protoplasma* **249 Suppl 2**, S95-100 (2012).

29. Chung, E. *et al.* Overexpression of VrUBC1, a Mung Bean E2 Ubiquitin-Conjugating Enzyme, Enhances Osmotic Stress Tolerance in Arabidopsis. *PLoS One* **8**, e66056 (2013).

30. Zhou, G.-A., Chang, R.-Z. & Qiu, L.-J. Overexpression of soybean ubiquitin-conjugating enzyme gene GmUBC2 confers enhanced drought and salt tolerance through modulating abiotic stress-responsive gene expression in Arabidopsis. *Plant Mol. Biol.* **72**, 357–367 (2010).

31. Wan, X., Mo, A., Liu, S., Yang, L. & Li, L. Constitutive expression of a peanut ubiquitin-conjugating enzyme gene in Arabidopsis confers improved water-stress tolerance through regulation of stress-responsive gene expression. *J. Biosci. Bioeng.* **111**, 478–484 (2011).

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