Tracing the evolutionary and genetic footprints of atmospheric tillandsioids transition from land to air

Lyu *et al.*

Supplementary Method 1. Speciation rate estimation

We employed BAMM v.2.5.0 [1] to study the diversity dynamics within Tillandsioideae. The data selected were processed using the 'setBAMMpriors' function and BAMM tools v.2.1.7 [2] within R [3]. This methodology enabled a thorough examination of diversity dynamics within the Tillandsioideae subfamily. The 'PlotRateThroughTime' function was utilized to generate rate-time graphs for the entire sunfamily and five specific phylogenetic branches and three genera. Additionally, TESS v.2.1 [4] complemented the analysis by identifying significant changes in speciation and extinction rates through R scripts [5]. The "getTipRates" function was employed to determine the species formation rate within Tillandsioideae, incorporating lineage-specific diversification models such as BAMM [6,7] and semiparametric species-level lineage DR to estimate the rate of species formation [8]. BAMM tip rates and DR results were fitted into a linear model using phylogenetic generalized least squares (PGLS) under a Brownian motion model in APE v. 5.5 [9] to assess the correlation between the two methodologies.

Supplementary Method 2. Genome assembly

Genome size, repeat sequence ratio, and heterozygosity were assessed were analyzed by k-mer (k=17) using the jellyfish2.2.7 (parameter: -G 2 -m 17 -C -o kmercount, kmercount -o 17merFreq, kmercount -o jelly.log) [10]. On account of the expected value of the Poisson distribution (depth=75), the *T. duratii* genome was estimated to be 1,030.11 Mb, of which the heterozygous rate was 1.6% and the repeat rate was 76.03%. On account of the expected value of the Poisson distribution (depth=87), the *V. erythrodactylon* genome was estimated to be 456.84 Mb, of which the heterozygous rate was 1.73% and the repeat rate was 63.14%. Because of the high heterozygosity (1%) of the two genomes, we used hifiasm v0.16.1 with default parameters for assembly. Hifiasm (https://github.com/chhylp123/hifiasm) [11] is a fast build haplotype - *de novo* assembly program for PacBio Hifi reads. The assembled contigs/scaffolds sequences were then arranged into pseudochromosomes using ALLHIC v 0.9.8 [12] (parameter: allhic extract group.clean.bam group. fasta --RE GATC allhic partition --pairsfile group.clean.pairs.txt --contigfile group.clean.counts_GATC.txt -K 24 --minREs 50 --maxlinkdensity 3 --NonInformativeRabio 0) (https://github.com/tangerzhang/ALLHiC) based one Hi-C sequencing data. Following this, manual correction based on chromosome interaction strength was visualized using JuiceBox v1.11.08 [13].

Supplementary Method 3. Genome assembly quality assessment

The completeness, coverage rate and average depth, and consensus quality were evaluated by several complementary methods. First, the completeness of the assembled genomes were assessed based on conserved plant genes in BUSCO v5.2.2 (Benchmarking Universal Single-Copy Orthologs：http://busco.ezlab.org/, parameter: -l embryophyta_odb10 -m genome) and CEGMA v2.5 (Core Eukaryotic Genes Mapping Approach ： http://korflab.ucdavis.edu/datasets/cegma/, default parameter) database. For the assessment of coverage rate and average depth, and sequencing uniformity, short fragment libraries of reads were selected and aligned to the assembled genome by BWA v0.7.8 (parameter: bwa mem -k

32 -w 10 -B 3 -O 11 -E 4 -t 16) [14]. Subsequently, Samtools v0.1.19 (https://github.com/samtools/samtools, parameter: samtools flagstat bam; samtools depth -a -q 0 -Q 0 bam) was utilized to sort the alignment results based on chromosome coordinates, remove duplicate reads, conduct SNP calling, and filter raw data. The assembled genome sequences were assessed in 10k windows for GC content and average depth. Minor deviations were noted in regions with low GC content, and contamination checks confirmed that all alignments were consistent with plants, indicating the absence of non-plant foreign sources in the assembled genome.

Supplementary Method 4. Genome annotation

For repeat annotation, the repeat annotation pipeline utilized a dual approach combining sequence similarity and *de novo* search to identify repetitive sequences across the entire genome. For homology-based predictions, the Repbase database (http://www.girinst.org/repbase) was used in conjunction with RepeatMasker v4.05 (http://www.repeatmasker.org/) and RepeatProteinMask v4.07 (https://github.com/rmhubley/RepeatMasker/blob/master/RepeatProteinMask) to extract repeat regions. RepeatModeler v1.0.5 (http://www.repeatmasker.org/RepeatModeler.html), RepeatScout v1.0.5 [15], Tandem Repeats Finder v4.09 [16] and LTR_FINDER v4.09 [17] were used to conduct *de novo* predictions of novel repetitive elements. All identified repeat sequences longer than 100bp, with less than 5% 'N' gap content, were assembled into the initial transposable element (TE) library. This library combined Repbase entries with generated TE sequences, processed using uclust to ensure non-redundancy. Finally, this comprehensive library was employed by RepeatMasker v4.05 to identify DNA-level repeats throughout the genome.

For gene annotation, the prediction of protein-encoding genes involved ab initio prediction, homology-based prediction, and RNA-Seq-assisted prediction to annotate gene models. Augustus v3.2.3 (https://github.com/Gaius-Augustus/Augustus), Geneid v1.4 (https://genome.crg.cat/software/geneid/index.html), Genescan v1.0 (http://genes.mit.edu/GENSCAN.html), GlimmerHMM v3.0.2 (https://ccb.jhu.edu/software/glimmerhmm/), and SNAP v2013.11.29 (https://snap.stanford.edu/) were employed in the ab initio- prediction. Sequences from closely related species including *Puya raimondii*, *Ananas comosus* (CB5 and F153), *Kobresia littledalei* [18], *Oryza sativa* [19], *Musa acuminata* [20], and *Elaeis guineensis* [21] were downloaded and aligned with the genome using TBLASTN v2.2.26 [22] (E-value≤1e−5). Subsequently, these matched protein sequences were further aligned with sequences from *V. erythrodactylon* and *T. duratii* to achieve accurate spliced alignments. This process was facilitated by GeneWise v2.4.1 [23] software, which accurately predicted the gene structures within each protein region. For RNA-seq data. Total RNA from roots, stem, leaves and flowers were extracted and sequenced for annotation. Trinity v2.8.5 was used for genome annotation to generate transcriptome reads assemblies. To improve accuracy, HISAT v2.2.1 (https://daehwankimlab.github.io/hisat2/download/) was employed in for the alignment of RNA-Seq reads to the genome sequence, identifying exon regions and splice sites. The

alignment results were subsequently fed into Stringtie v2.2.1 (https://github.com/gpertea/stringtie) using default parameters for genome-based transcript assembly. A non-redundant reference gene set was compiled by integrating genes predicted from three methods using EvidenceModeler v1.1.1 (https://github.com/EVidenceModeler/EVidenceModeler/releases), incorporating PASA terminal exon support and including masked transposable elements in the gene prediction process. For non-coding RNA, the tRNAscan-SE v1.4 (http://lowelab.ucsc.edu/tRNAscan-SE/) was used to predict tRNAs. The rRNA sequences of six closely species [18-21] were used as references, which are highly conserved. Blast was used to predict rRNA sequences according to the above references. The Rfam database was searched by infernal v1.1.2 (http://infernal.janelia.org/), identifying miRNAs, snRNAs and other non-coding RNAs (ncRNAs).

For gene functional annotations, Blastp (E-value≤1e−5) was employed in aligning protein sequences against the Swiss-Prot database to determine gene functions. InterProScan70 v5.39 [24] was employed in annotating motifs and domains through various databases ProDom, PRINTS, Pfam, SMART, PANTHER, and PROSITE. Gene Ontology (GO) IDs were assigned based on corresponding InterPro entries. Transferring annotations based on the closest BLAST hit (E-value < 10-5) in the Swissprot database and DIAMOND v0.8.22/ BLAST hit (E-value < 10-5) in the NR database [25], were used to predict protein functions. Additionally, gene sets were mapped to KEGG pathways to identify the best pathway match for each gene.

Supplementary Method 5. Identification of chromosomal rearrangements.

To figure out the linkage and conservation of homologous genes, genome sequences of *V.erythrodactylon* and *A. comosus* were aligned to the reference *T.duratii* genome. Collinearity analysis was performed utilizing MuMmer (http://mummer.sourceforge.net/), LASTZ (https://lastz.github.io/lastz/) and MCscanX v1.1.11 [26].

Supplementary Method 6. *De novo* identification of LTR-RTs

The identified repetitive sequences in genome annotation may manifest as partial fragments. To identify repeat sequences with complete structures, LTRharvest v1.07 [27] and LTRfinder v4.09 [17] were employed to detect the complete length of LTR-RTs. Following the elimination of overlapping outcomes pinpointed by LTRfinder, these results were merged with those from LTRharvest to exclude elements with intersecting positions.

Supplementary Method 7. Annotation of LTR-RTs

Using tRNAscan-SE v1.4 [28] to predict tRNA sequences in the genome, which are used to predict the PBS (primer binding site) of LTR-RTs (long terminal repeat retrotransposons). The PBS, an approximately 18 bp segment at the 5' end of the LTR, can complementarily bind to the 3' end of specific tRNAs, serving as the initial step in reverse transcription. Conserved domain models (HMM) for the Pol gene (comprising AP, IN, RT, and RH domains), gag gene, and env gene were acquired from GyDB (http://gydb.org/) [29] for forecasting the presence or absence of protein domains in LTR-RTs, along with their primary location. Subsequently,

LTRdigest v1.5.8 [30] was utilized to annotate the structure of LTR-RTs, followed by filtration through SILIX v1.2.11 [31].

Supplementary Method 8. Construction of phylogenetic trees of LTR-RTs

RT sequences were extracted from intact LTR retrotransposon elements. MUSCLE v3.8.31 was employed to align RTs lacking premature termination codons [32]. Subsequently, a rootless tree was generated using the default TreeBeST parameter via neighbor-joining (NJ) phylogenetic analysis [33]. This approach facilitated the categorization of LTR-RTs into distinct lineages and clades based on phylogenetic analysis [34].

Supplementary Method 9. Identification of solo-LTRs

Initially, the identified LTR-RT sequences in the genome are masked, followed by utilizing BLAST to align these sequences back to the masked genome. In the alignment coverage region, if only the 5'-LTR or 3'-LTR sequences align without any association with LTR-RT-related proteins, that specific region is recognized as the site of a solo-LTR. Subsequently, all 5'-LTR and 3'-LTR sequences of LTR-RTs are aligned back to the masked genome, and the highest scoring alignment position within the solo-LTR region determines the sequence and location of the solo-LTR.

Supplementary Method 10. Calculation of LTR-RTs insertion time

The 5'-LTR and 3'-LTR sequences of LTR-RTs are extracted, followed by a muscle alignment to compute the nucleic acid difference λ (λ>0.75, considered invalid). Subsequently, the genetic distance K is calculated using the formula $K=-0.75\ln(1-4\lambda/3)$. Finally, the insertion time of LTR-RTs is determined by

$$T=K/2r \quad (1)$$

where r denotes the natural mutation rate; for instance, the mutation rate of rice is 1.3e-8 bp-1·year-1) [34].

Supplementary Method 11. Identification of one-to-one orthologous genes

To explore the adaptive evolution of tillandsioids, a comparative genomic analysis was conducted involving 14 plant species from various plant lineages. All annotated genes were categorized into different gene families using OrthoFinder v2.3.7 (https://github.com/davidemms/OrthoFinder) [35]. Protein sequences belonging to single-copy gene families were aligned using MUSCLE v3.8.31 (https://github.com/cran/muscle) [36], trimmed with TRIMAL v1.2 (https://vicfero.github.io/trimal/), and then a maximum likelihood phylogenetic tree by concatenating gene sequences through RAxML v8.2.1 were constructed (https://github.com/stamatak/standard-RAxML) [37].

Supplementary Method 12. Comparative genomic analysis

To elucidate genome variations during evolution, the clustering outcomes of gene families were employed for expansion/contraction analysis using CAFE v2.1 (https://github.com/hahnlab/CAFE/blob/master/README.md) [38]. Following the alignment of protein sequences from different species, genomic collinearity was assessed through MCScanX

v1.1.11 (https://github.com/wyp1125/MCScanx). Subsequently, Ks, Ka, and Ka/Ks ratios were computed utilizing PAML 4.9i (https://github.com/abacus-gene/paml), and a graphical representation of the Ks distribution to identify whole-genome duplication events was generated using ggplot2 v2.2.1 [39].

Supplementary Method 13. Trajectory and RNA velocity analysis

The trajectory analysis of all leaf cell types in both early and late stages were performed using Monocle3 tools [40] and the order_cells and learn_graph functions with default parameters were used to acquire a trajectory and set up a node in vascular clusters (vascular-m and vascular-y) for inferring pseudo-time. RNA velocity of root cells was analyzed using Velocyto v0.17.16 [41]. The spatial velocity graph was generated using velocity embedding stream function with spatial coordinates.

Supplementary Method 14. Differential gene expression analysis

Differentially Expressed Genes (DEGs) were identified using the edgeR package in R software [42], with a false discovery rate of ≤0.05 and an absolute value of log2 fold change ≥1 (|log2 FC|≥1) utilized as the threshold to ascertain statistically significant differences in gene expression [43].

Supplementary Method 15. GO enrichment analysis

Utilizing known gene functions and gene ontology biological processes in *Arabidopsis* (TAIR10), homologs of Arabidopsis genes in T. *duratii* or *V. erythrodactylon* were identified through BLASTP [44] analysis. Subsequently, GO enrichment analysis was conducted using the R package clusterProfiler [45], with TAIR10 annotation serving as the background for the analysis.

Supplementary Method 16. Operational taxonomic units (OTUs) Identification

We then utilized the original FastQ files, applying Trimmomatic v0.25 [46] and FLASH v1.2.11 [47] for quality filtering and merging, respectively. The UPARSE v7.1 (http://drive5.com/uparse/) tool [48] was employed to cluster the processed sequences into OTUs with a 97% similarity cut-off, while UCHIME v2.4.2 [49] was used to identify and eliminate chimeric sequences. Plastids (e.i. mitochondria and chloroplast) were filtered out after annotation. Further classification of the OTUs was conducted using the RDP Classifier algorithm (http://rdp.cme.msu.edu/) against the Silva database (https://www.arb-silva.de/), with a confidence threshold set at 70%.

Supplementary Method 17. Metagenome assembly

After sequencing, the raw reads underwent quality trimming, where sequences with a quality score below 20 and a length less than 50 bp were removed. The remaining clean reads were then assembled into scaffolds using MEGAHIT v1.0.4 (https://github.com/voutcn/megahit), which were subsequently fragmented at N junctions to produce scaftigs (i.e., continuous sequences within scaffolds). Following quality control, the clean reads from each sample were aligned to the assembled scaftigs using Bowtie2 v2.2.4 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) to isolate unused paired-end reads. These unused

reads from each sample were pooled and subjected to mixed assembly. The resulting mixed-assembled scaffolds were fragmented at N junctions to obtain N-free scaftigs. Finally, scaftigs shorter than 500 bp were filtered out to generate contigs for subsequent prediction and annotation.

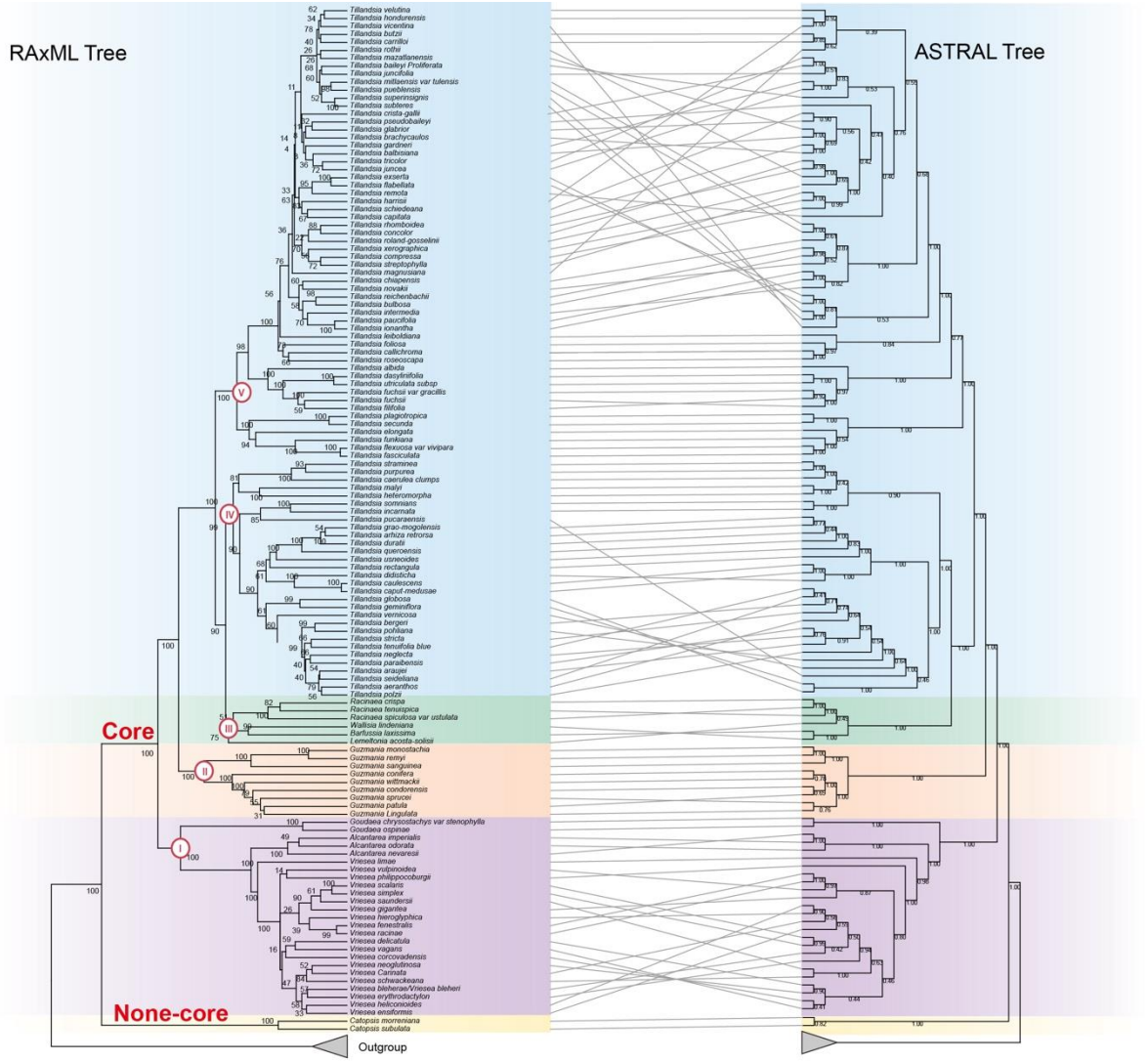
Supplementary Method 18. Metagenome gene prediction and annotation

MetaGeneMark v2.10 (http://exon.gatech.edu/GeneMark/metagenome/Prediction) [50] was used for Open Reading Frame (ORF) prediction, and sequences shorter than 100nt were filtered out. Subsequently, ORF predictions from individual samples and mixed assemblies were processed using CD-HITv4.5.8 (http://www.bioinformatics.org/cd-hit/) to remove redundancy, obtaining a non-redundant initial gene catalogue clustered at 95% identity and 90% coverage. The longest sequence from each cluster was selected as the representative sequence. Clean reads from each sample were aligned to the initial gene catalogue using Bowtie2 to calculate the number of reads mapped to each gene in every sample. Genes with reads ≤ 2 in all samples were filtered out to obtain the final gene catalogue (Unigenes) for subsequent analysis. The ORFs were annotated with the cluster of orthologous groups of proteins (COG) information obtained from the eggNOG v4.5 database through BLASTP v2.2.28 [42] with an e-value cutoff of 1e-5. Additionally, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation was conducted by performing a BLAST search against the KEGG database (http://www.genome.jp) with an optimized e-value cutoff of 1e-5.

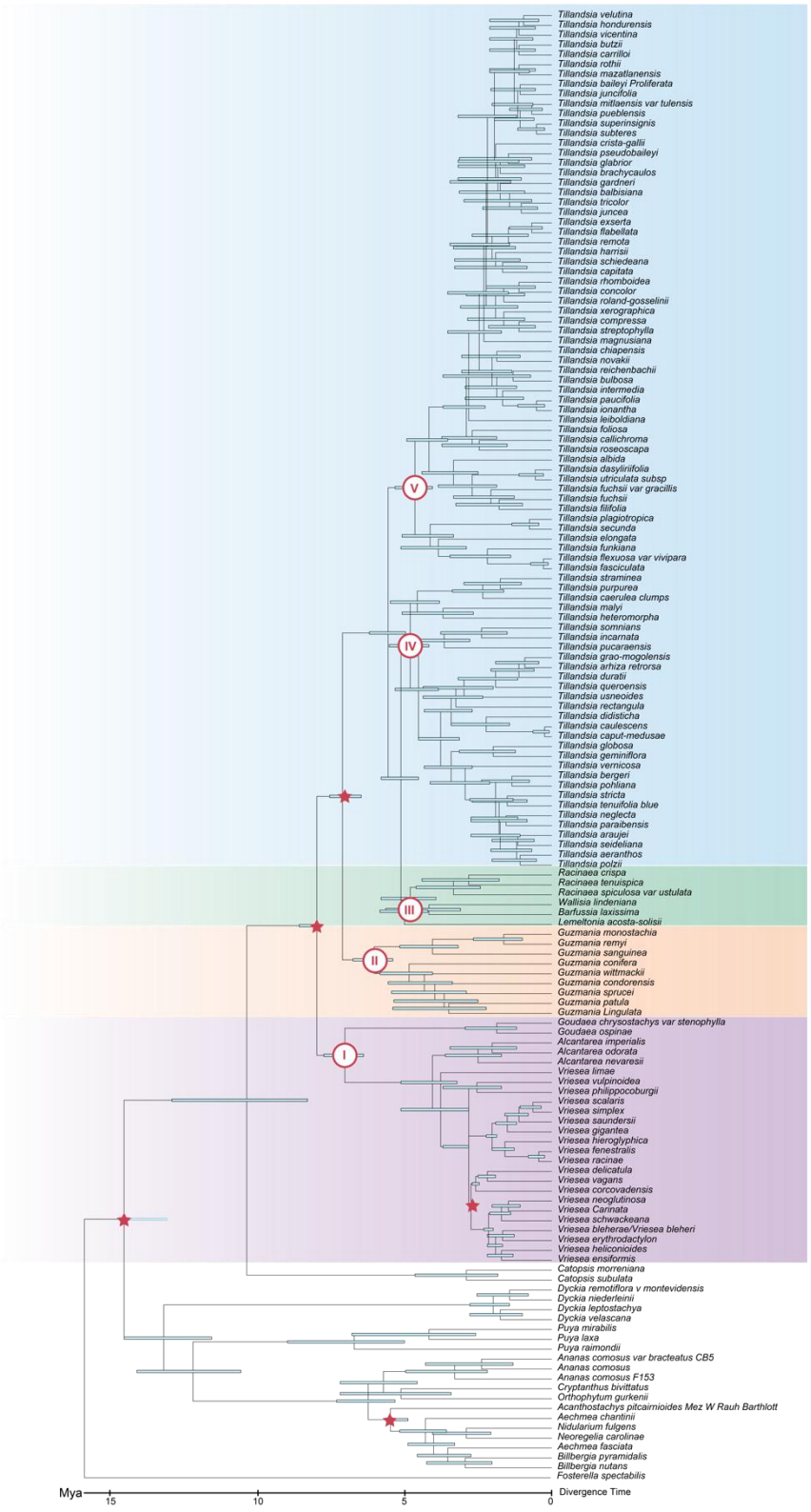
Supplementary Method 19. Metagenome species annotation

Unigenes were aligned against bacterial, fungal, archaeal, and viral sequences extracted from NCBI's NR database (Version: 2018.01) using DIAMOND (blastp, evalue ≤ 1e-5). For the blast result of each sequence, select alignments where the e-value is ≤ the minimum e-value*10 for subsequent analysis. After filtering, each sequence may have multiple alignment results with different species classification information. To ensure biological relevance, the Lowest Common Ancestor (LCA) algorithm (implemented in MEGAN for taxonomic classification) is employed. It assigns the taxonomic annotation of the first branching level before divergence as the species annotation information for each sequence.

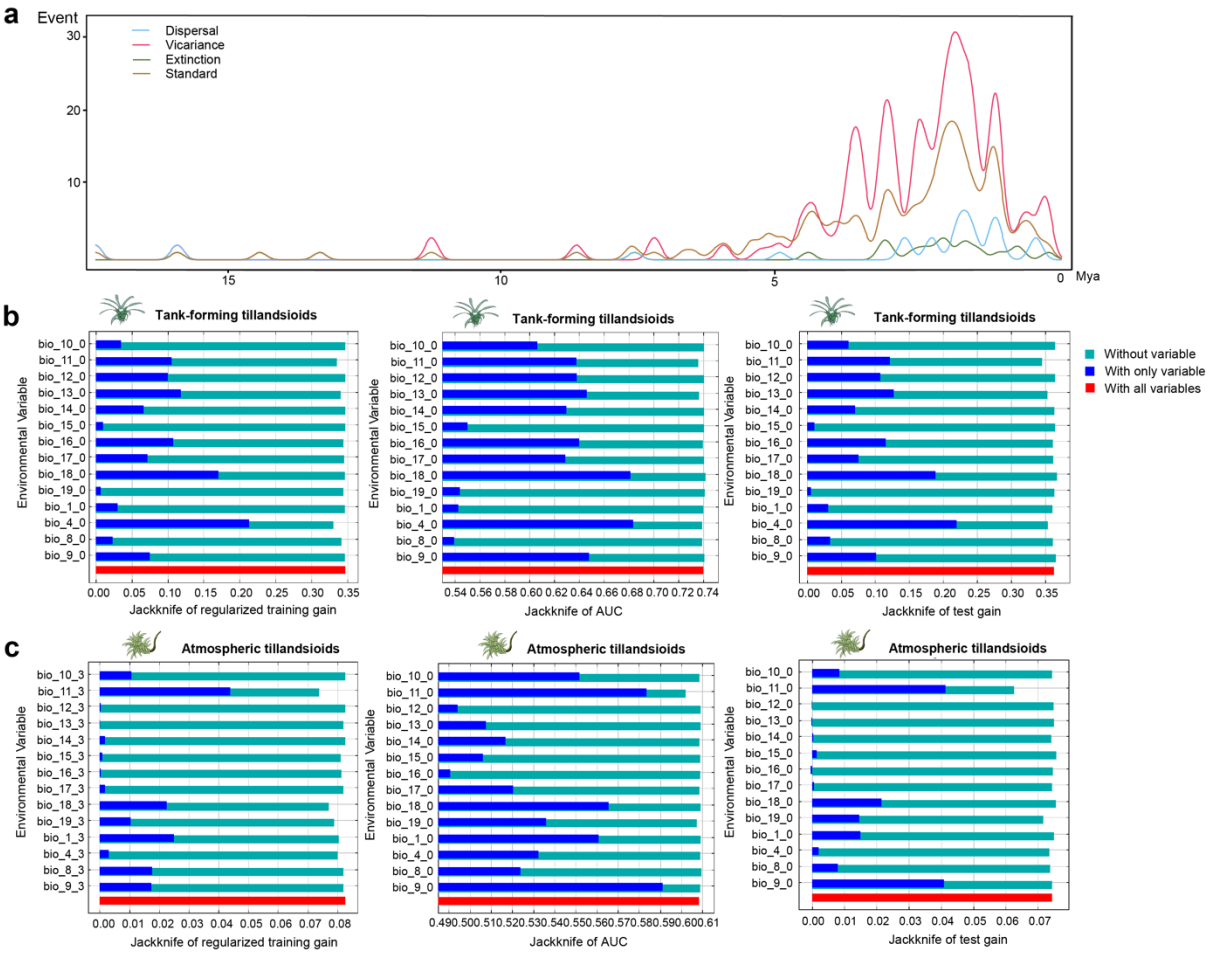
Supplementary Fig. 1. Morphologies of two types of tillandsioids. (a-h) Tank-forming tillandsioids. Scale bars, 2cm; **(i-p)** Atmospheric tillandsioids. Scale bars, 5cm.



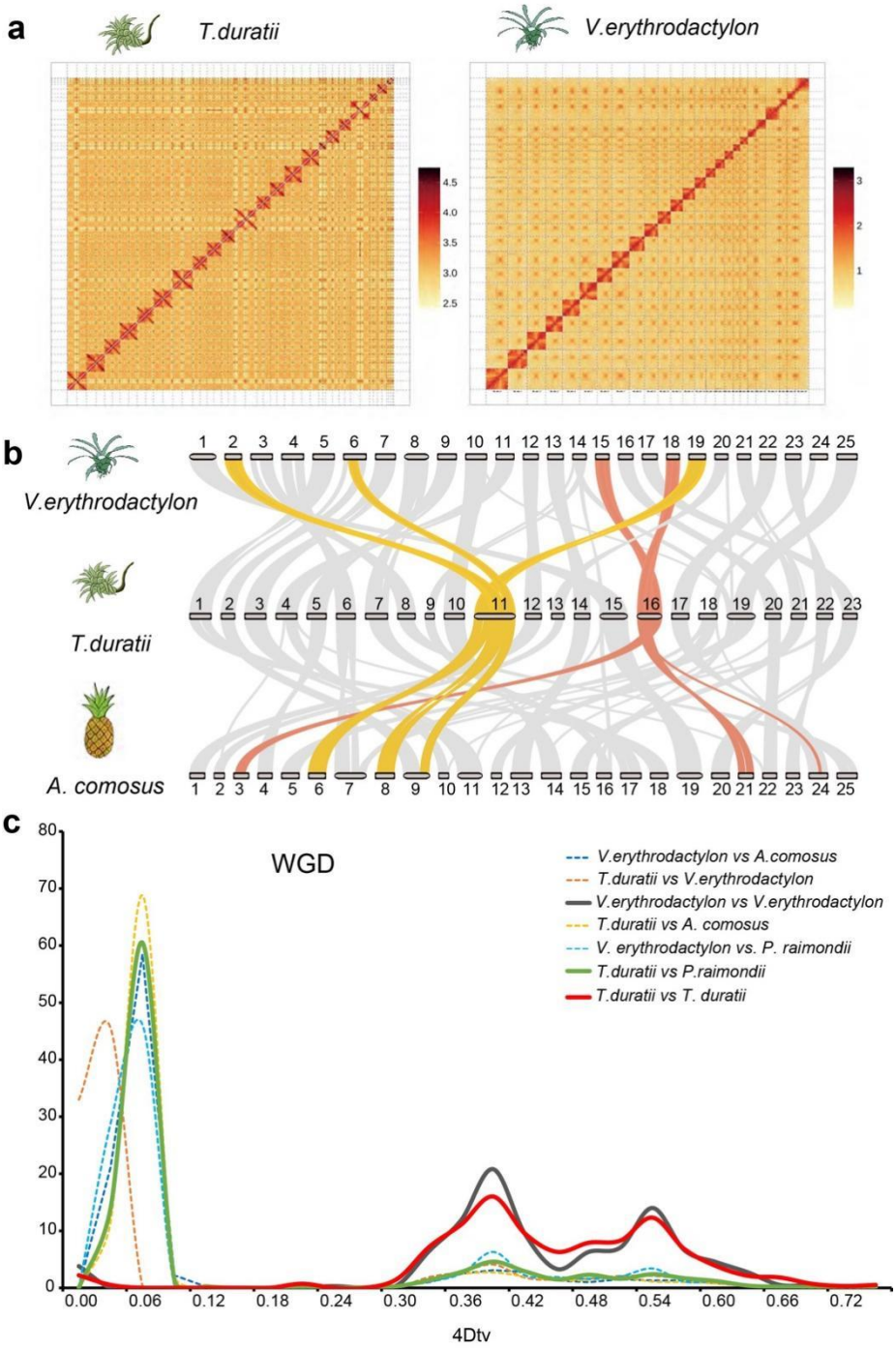
Supplementary Fig. 2. Tanglegram of the RAxML (left) and ASTRAL (right) phylogenies of Tillandsioideae. Numbers one the branch are bootstrap values.



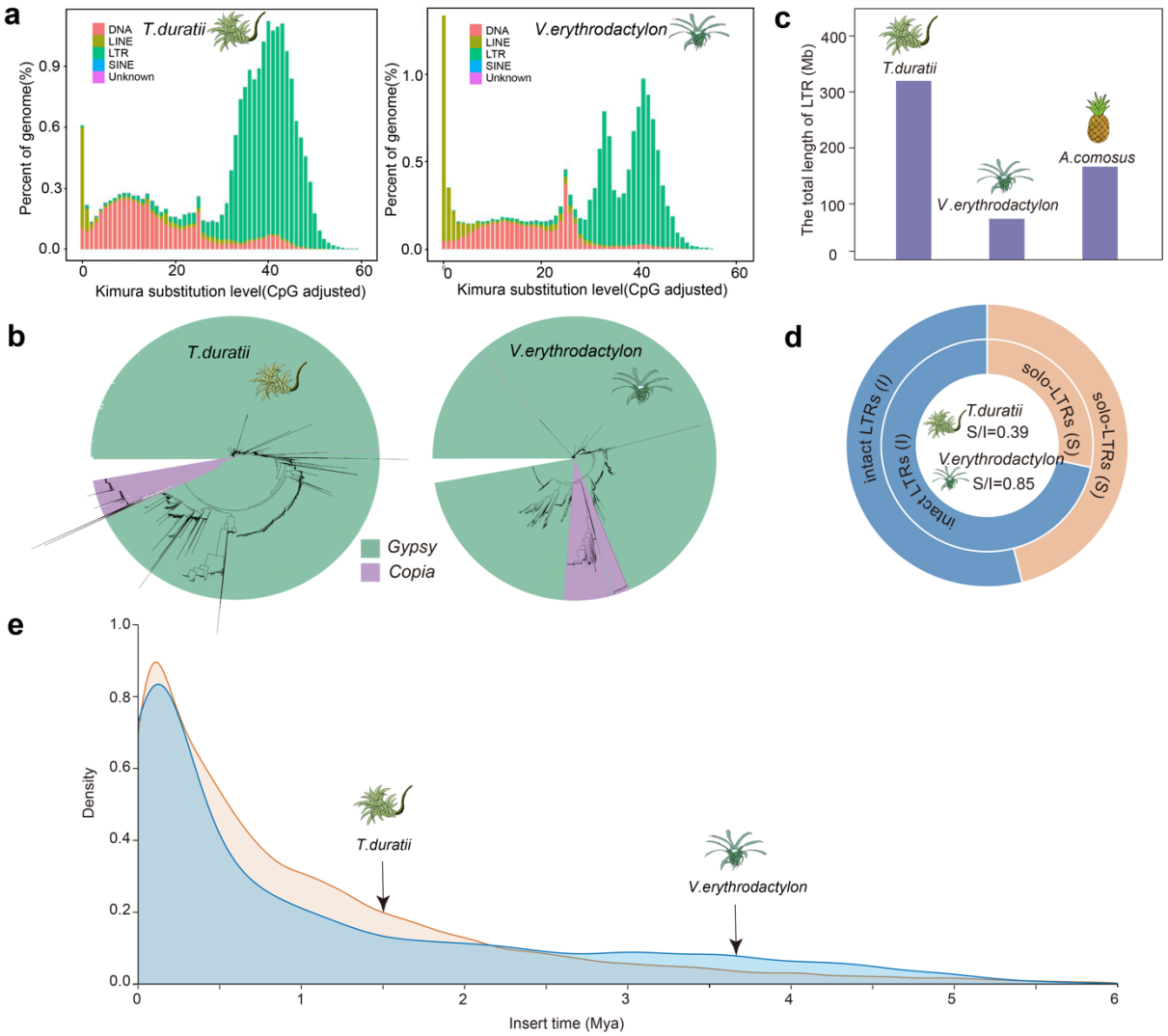
Supplementary Fig. 3. Chronogram of Tillandsioideae Estimated Using a Bayesian Relaxed Molecular Clock. This time tree is the result of MEGA 11 analysis using the ML tree from the concatenated 91 genes. The red stars on the nods indicate the existence of fossil evidence that supported the estimated time. Light-blue bars indicate 95% credibility intervals of the divergence times. The geological timescale at the bottom is in million years.



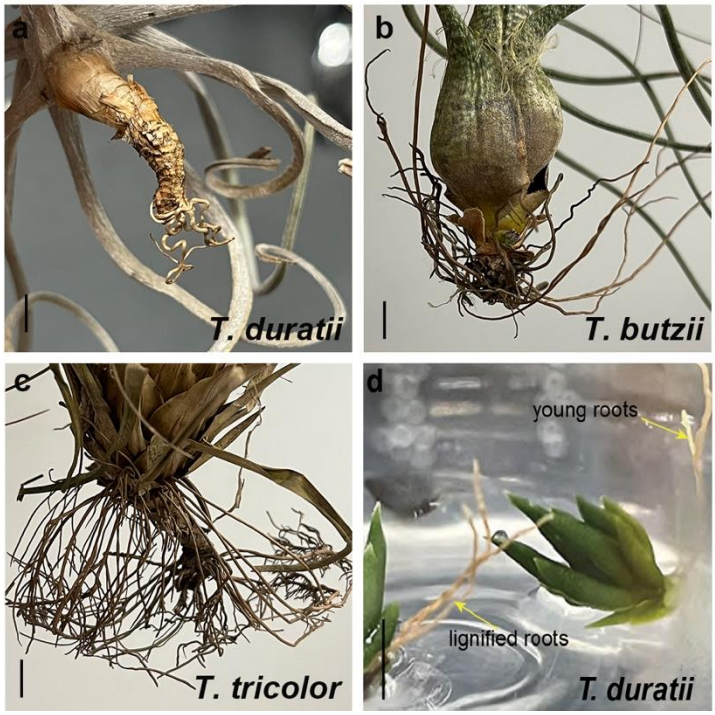
Supplementary Fig. 4. Evolutionary events and environmental factors derived the evolution of tillandsioids. **(a)** The dispersal, vicariance and extinction events during tillandsioids evolution. **(b-c)** Environmental factors derived the tachylalic evolution of tillandsioids. Jackknife test of individual environmental variable importance in the development of the MaxEnt model relative to all environmental variables (hachured bars) for each predictor variable alone (blue bars), and the drop in training gain when the variable is removed from the full model (green bars). The environmental variable of tank-forming tillandsioids with highest gain when used in isolation is bio_4_3, indicating the temperature seasonality as the most important environmental factor driving the tank-forming tillandsioids diversification **(b)**. While the environmental variable of atmospheric tillandsioids with highest gain when used in isolation is bio_11_3, indicating that mean temperature of the coldest quarter is the most important environmental factor driving the tank-forming tillandsioids diversification **(c)**.



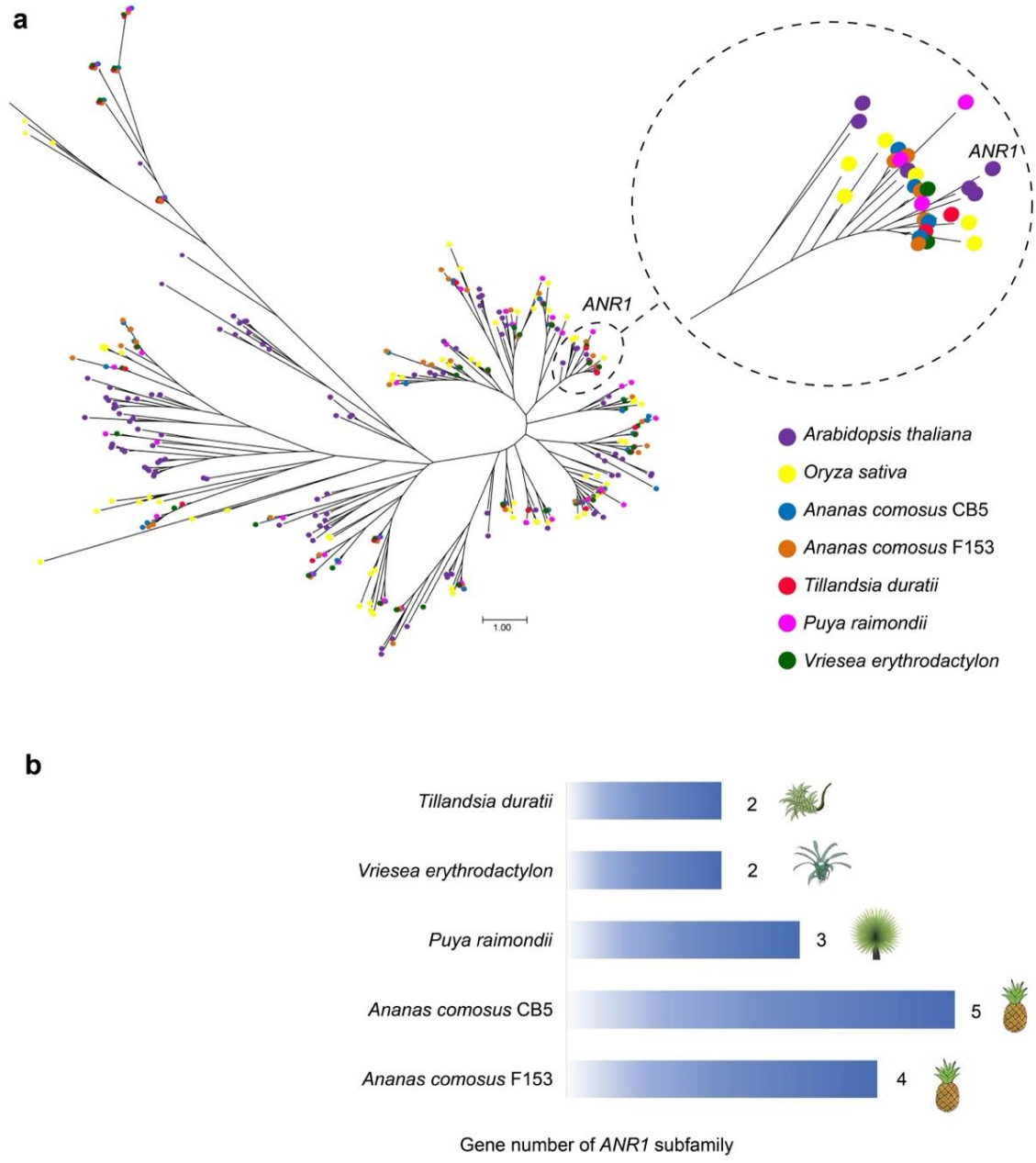
Supplementary Fig. 5. HiC mapping, Genomic collinearity and WGD event analysis of two tillandsioids. (a) Summary of Hi-C mapping of *T. duratii* and *V. erythrodactylon*. **(b)** Genomic collinearity between *T. duratii*, *V. erythrodactylon* and *A. comosus*. Gray wedges connect matching gene pairs, with two sets highlighted in light and dark orange showing the origin of two fused chromosomes in *T. duratii*. **(c)** Dating of WGD events in bromeliads.



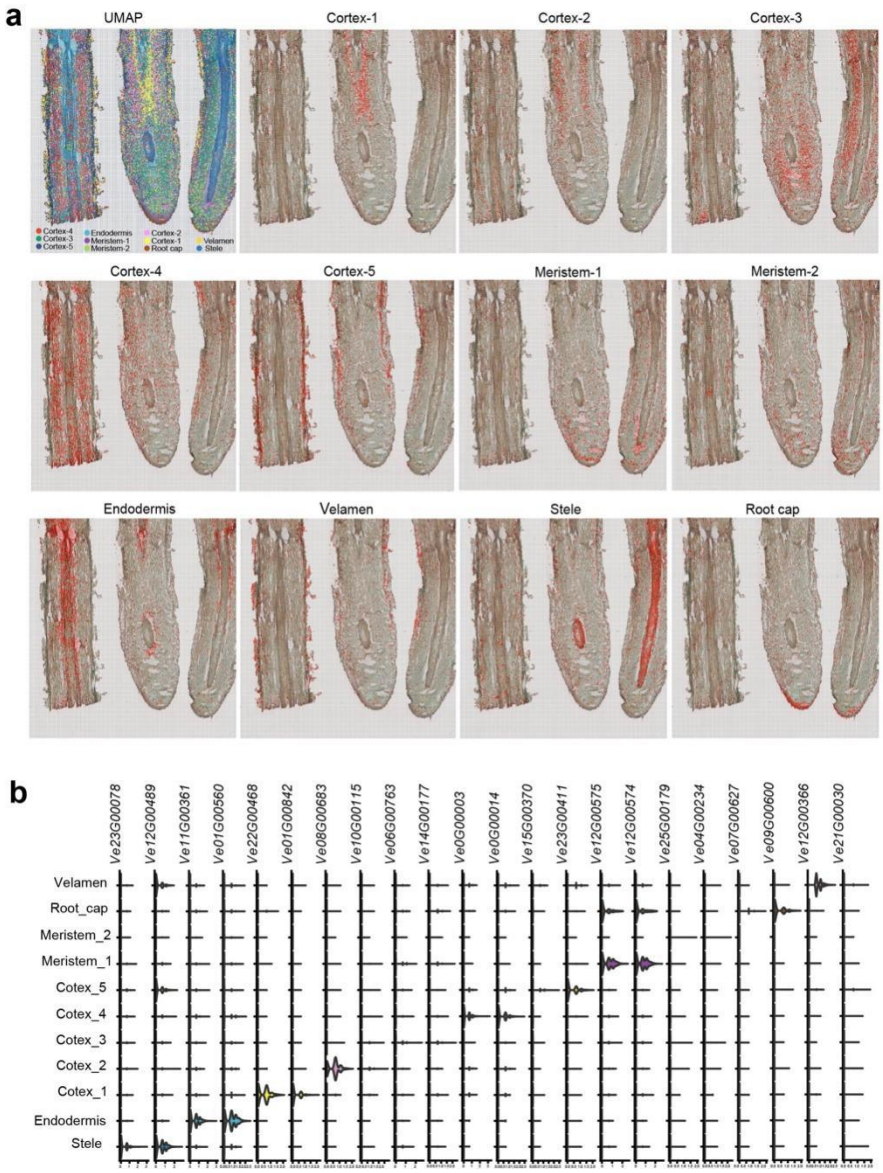
Supplementary Fig. 6. Repetitive DNA sequence analysis of two tillandsioids genomes. (a) Kimura substitution level (%) for each copy against its consensus sequence used as proxy for expansion history of the transposable elements. LINE, long interspersed nuclear elements; LTR, long terminal repeat; SINE, short interspersed nuclear element. **(b)** The phylogenetic relationships of Gypsy and Copia LTR-retrotransposons identified in *T. duratii* (left) and *V. erythrodactylon* (right) genomes. **(c)** The number of LTR-retrotransposons in *T. duratii*, *V. erythrodactylon* and *A. comosus*. **(d)** The proportion of solo-LTRs in two genomes. **(e)** Distribution of insertion times of LTR-retrotransposons in two genomes.



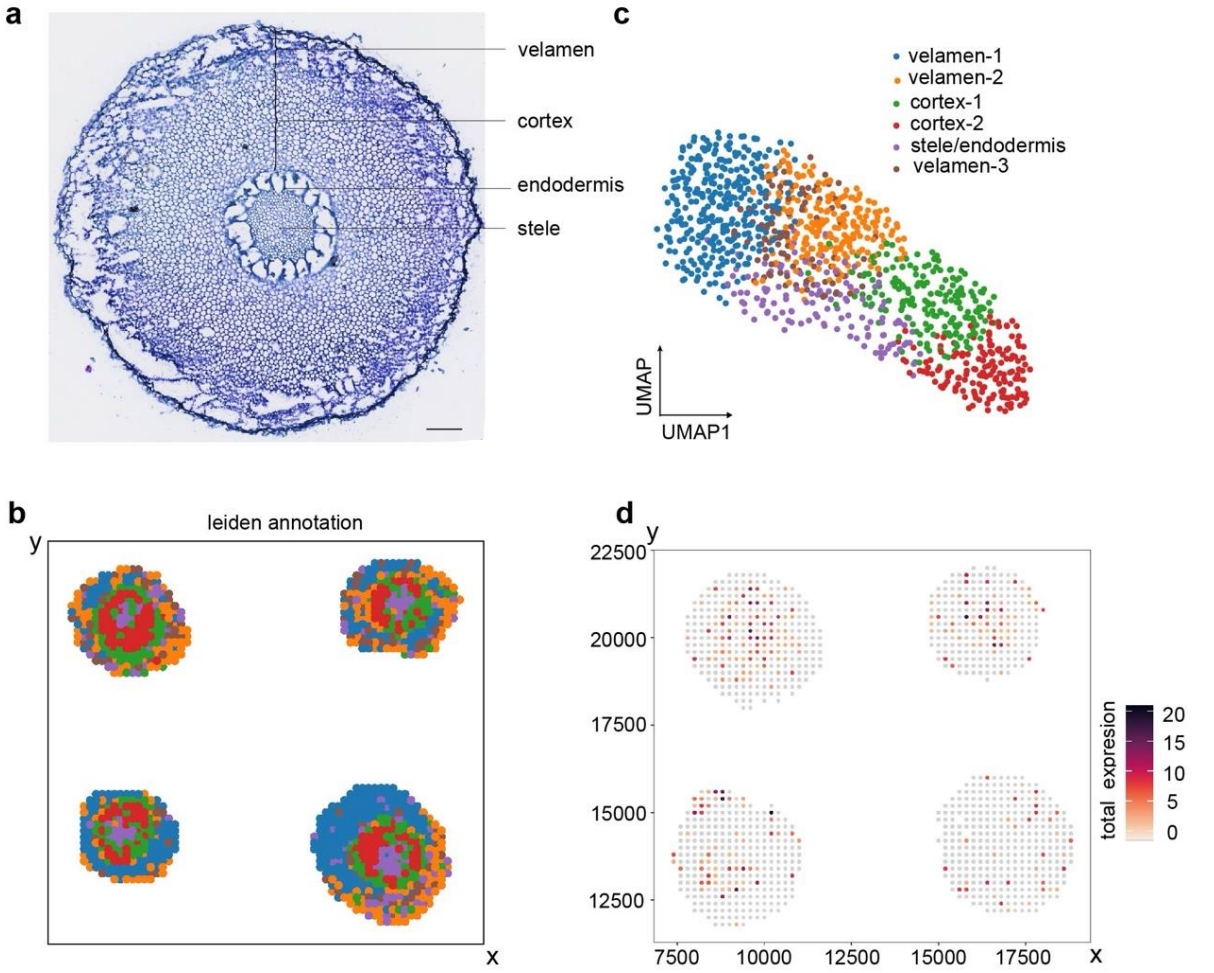
Supplementary Fig. 7. Photos of mature roots. *T. duratii* **(a)**, *T. butzii* **(b)**, *T. tricolor* **(c)**, and newly emerged roots of *T. duratii* growing on culture substrate **(d)**. Scale bars, 1 cm.



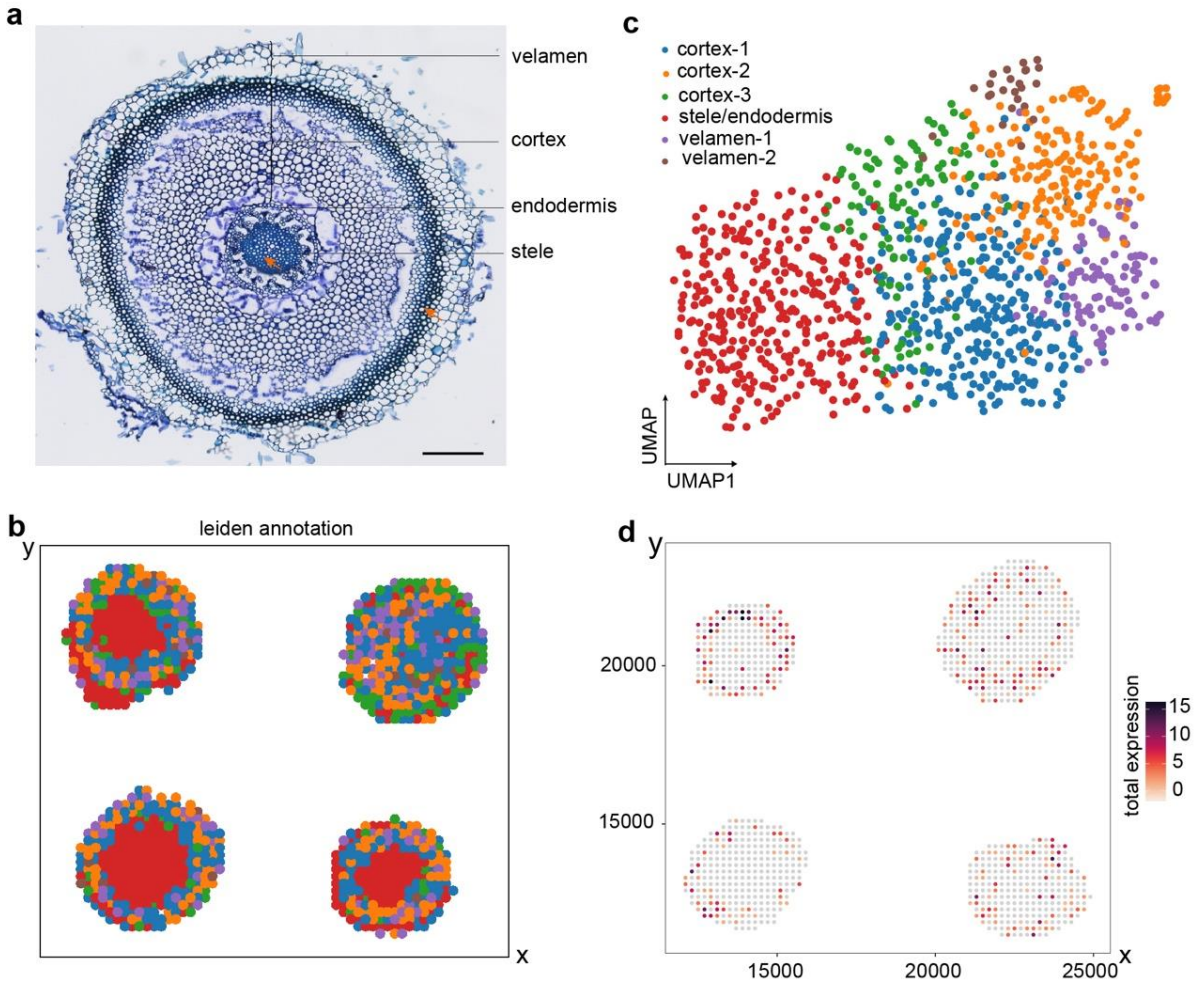
Supplementary Fig. 8. Reduced *ANR1* genes involved in development failure of lateral root in epiphytic bromeliads. (a) Phylogenetic analysis of MADS-box genes among *Tillandsia duratii*, *Vriesea erythrodactylon*, *Ananas comosus* (CB5 and F153), *Puya raimondii*. *O. sativa* and *Arabidopsis*. The *ANR1* subfamily are marked by dashed box. **(b)** Gene number of *ANR1* subfamily in different Bromeliaceae species.



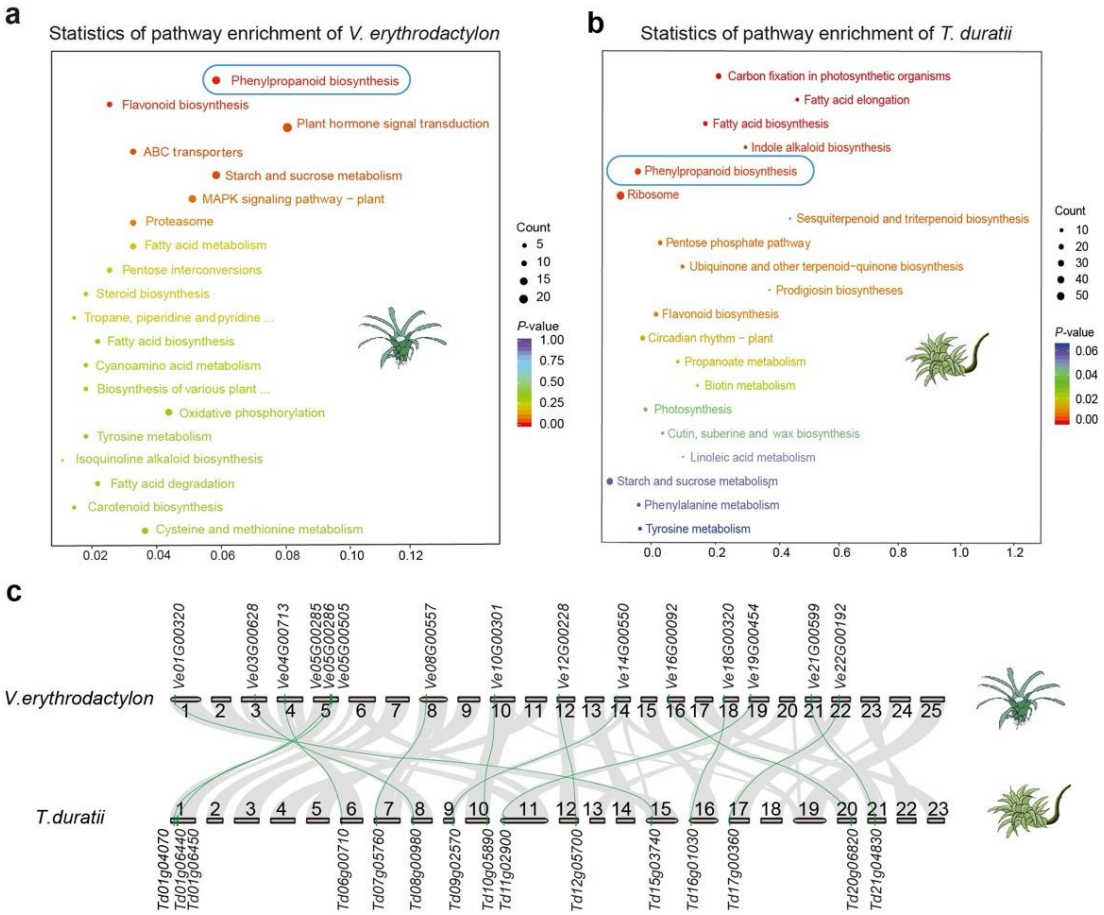
Supplementary Fig. 9. Spatial maps of different spot clusters over a representative single strand DNA staining image. (a) and the Expression patterns of top-2 marker genes in different clusters **(b)** of *V. erythrodactylon* roots.



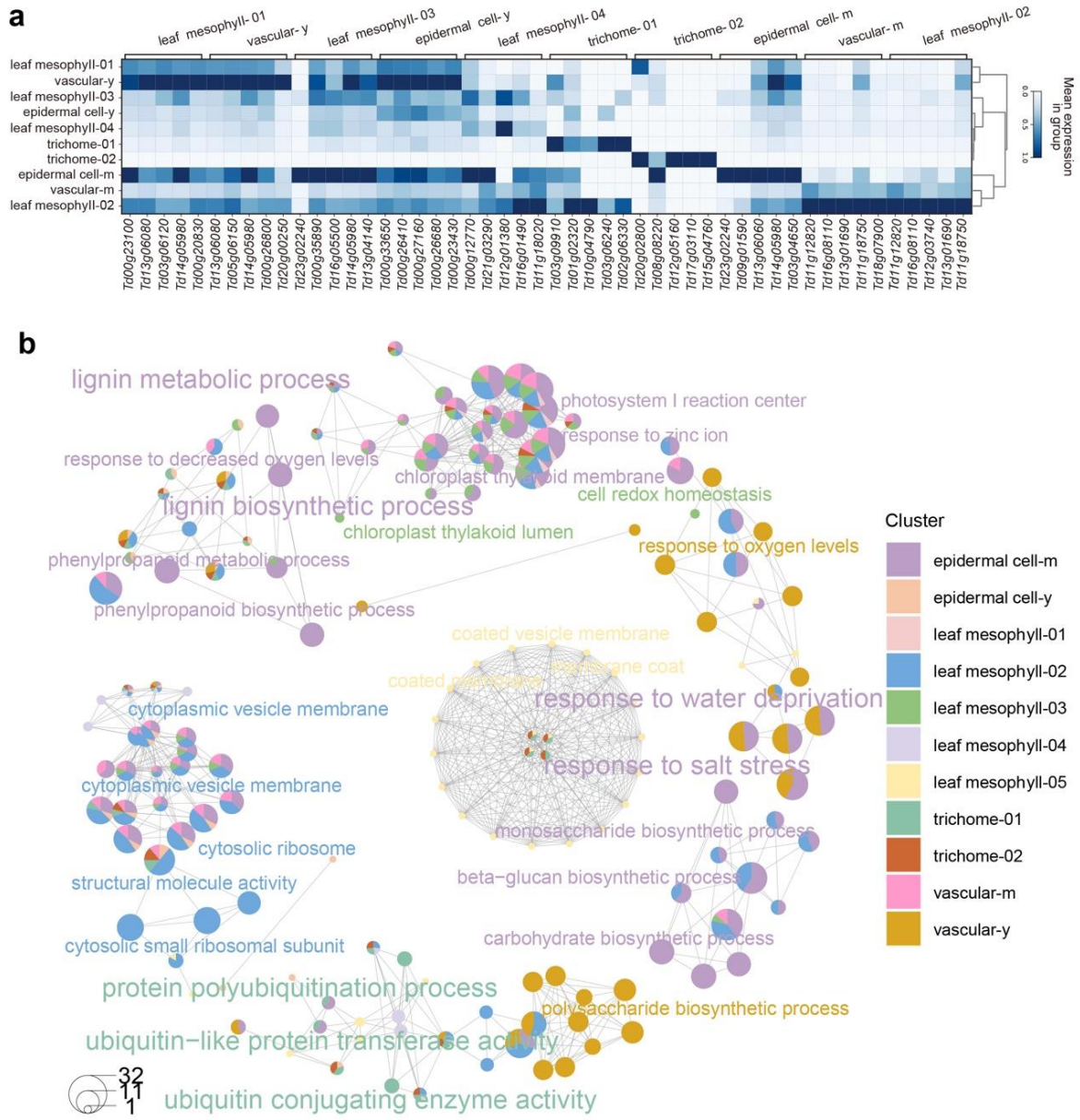
Supplementary Fig. 10. Spatial RNA-seq analysis reveals spatial distribution of gene expression in new root development of *V. erythrodactylon*. **(a)** Cross section of *V. erythrodactylon* new root. Scale bar, 500 μm. **(b-c)** Cell clustering and cell-type identification in *V. erythrodactylon* new root from cross section based on spatial transcriptomics. Visualization and annotation of identified cell types in roots. Cell types are marked by different colors. **(d)** The spatial visualization of the expression of marker gene modules (involved in plant secondary cell wall biosynthesis) specific to marker genes identified in cotex-5 (Fig. 5f). The detailed gene names of each module are listed in the right border.



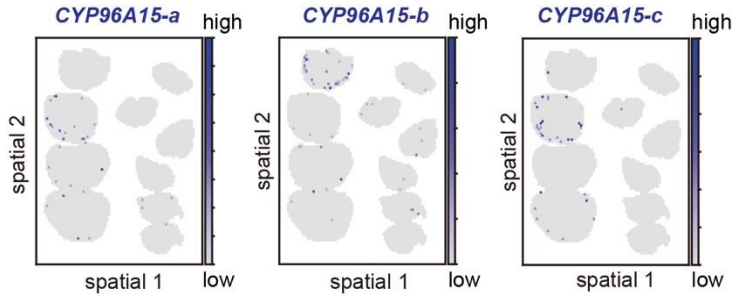
Supplementary Fig. 11. Spatial RNA-seq analysis reveals spatial distribution of gene expression in lignified root development of *V. erythrodactylon*. **(a)** Cross section of *V. erythrodactylon* lignified root. The red arrows refer to SCWs. Scale bar, 500 μm. **(b-c)** Cell clustering and cell-type identification in *V. erythrodactylon* lignified root from cross section based on spatial transcriptomics. Visualization and annotation of identified cell types in roots. Cell types are marked by different colors. **(d)** The spatial visualization of the expression of marker gene modules (involved in plant secondary cell wall biosynthesis) specific to marker genes identified in cotex-5 (Figure 5f). The detailed gene names of each module are listed in the right border.



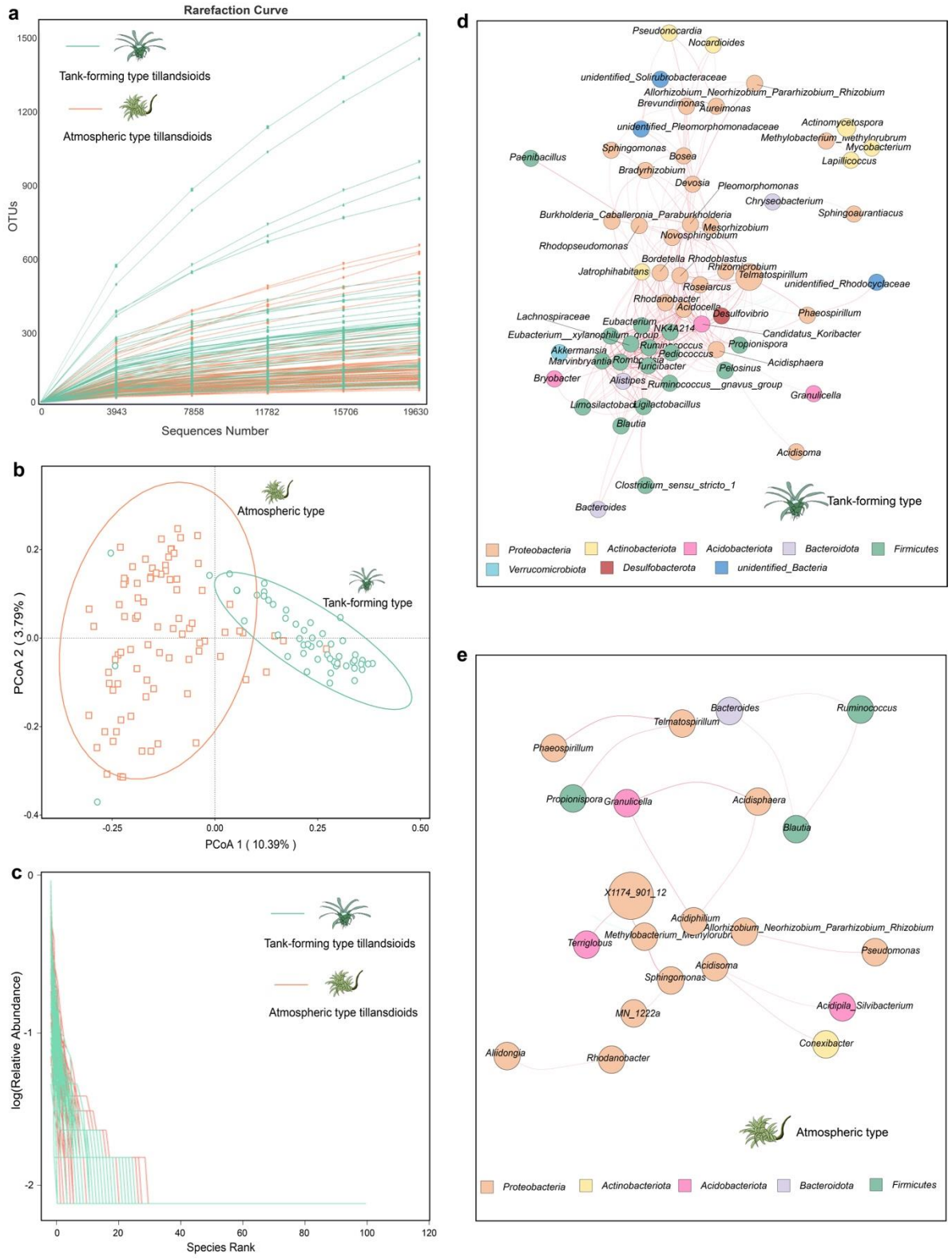
Supplementary Fig. 12. Gene expression in lignified roots of tillandsioids. (a-b) The functional enrichment of up-regulated genes in the elongation zone (lignified) of *T. duratii* and *V. erythrodactylon* root. Bonferroni multiple comparison Hypergeometric Test was performed for *P* values. **(c)** The relationship of the DEGs involved in phenylpropanoid biosynthesis. The genes connected by green lines are orthologous.



Supplementary Fig. 13. Expression profiles of top 5 marker genes across cell clusters. (a) and network analysis revealing enriched GO terms/pathways of DEGs in leaf clusters identified via spatial RNA-seq of T. *duratii* **(b)**.

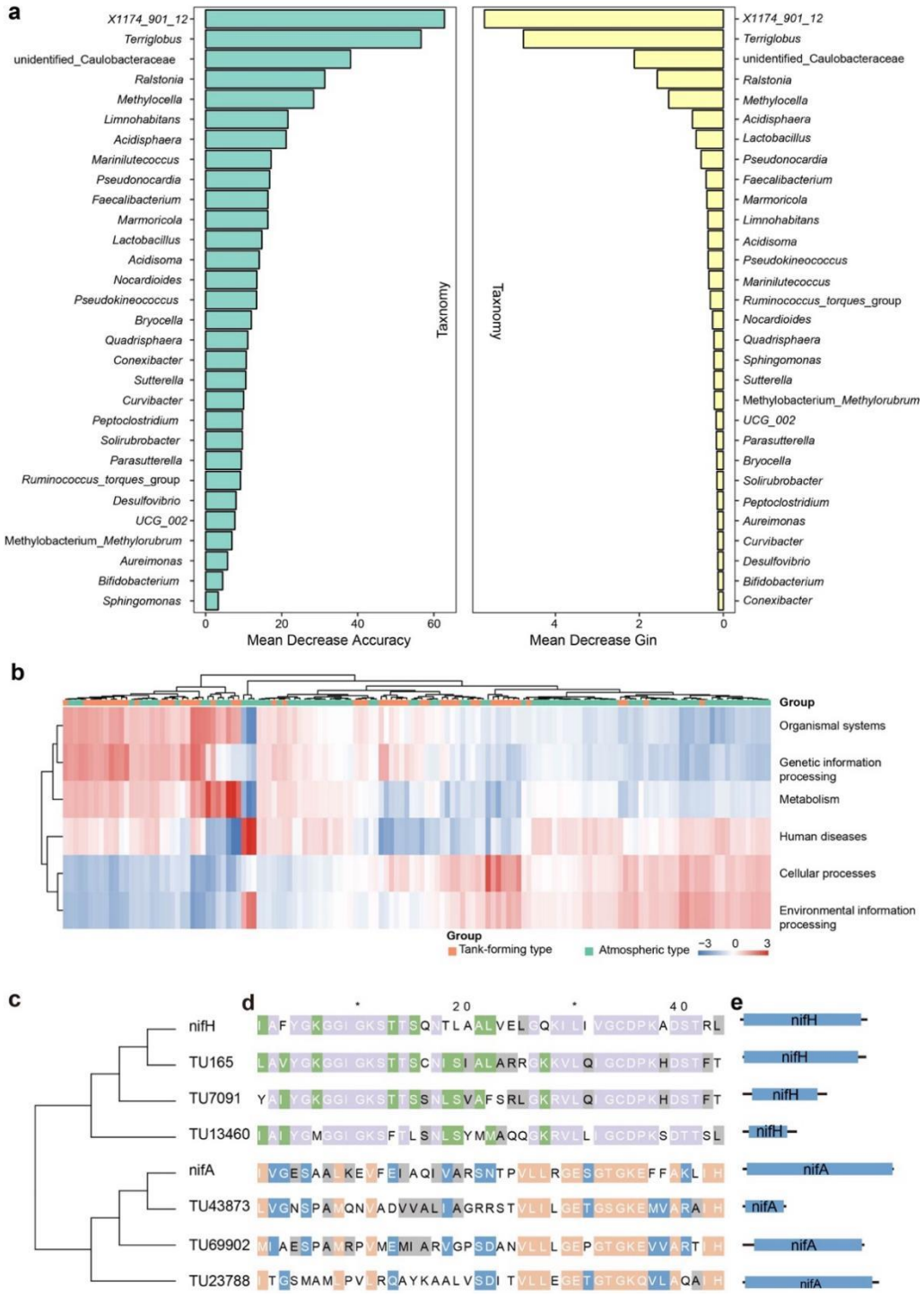


Supplementary Fig. 14. The spatial visualization of the expression of the three tandem repeat *CYP96A15* genes specific to trichome cell region.

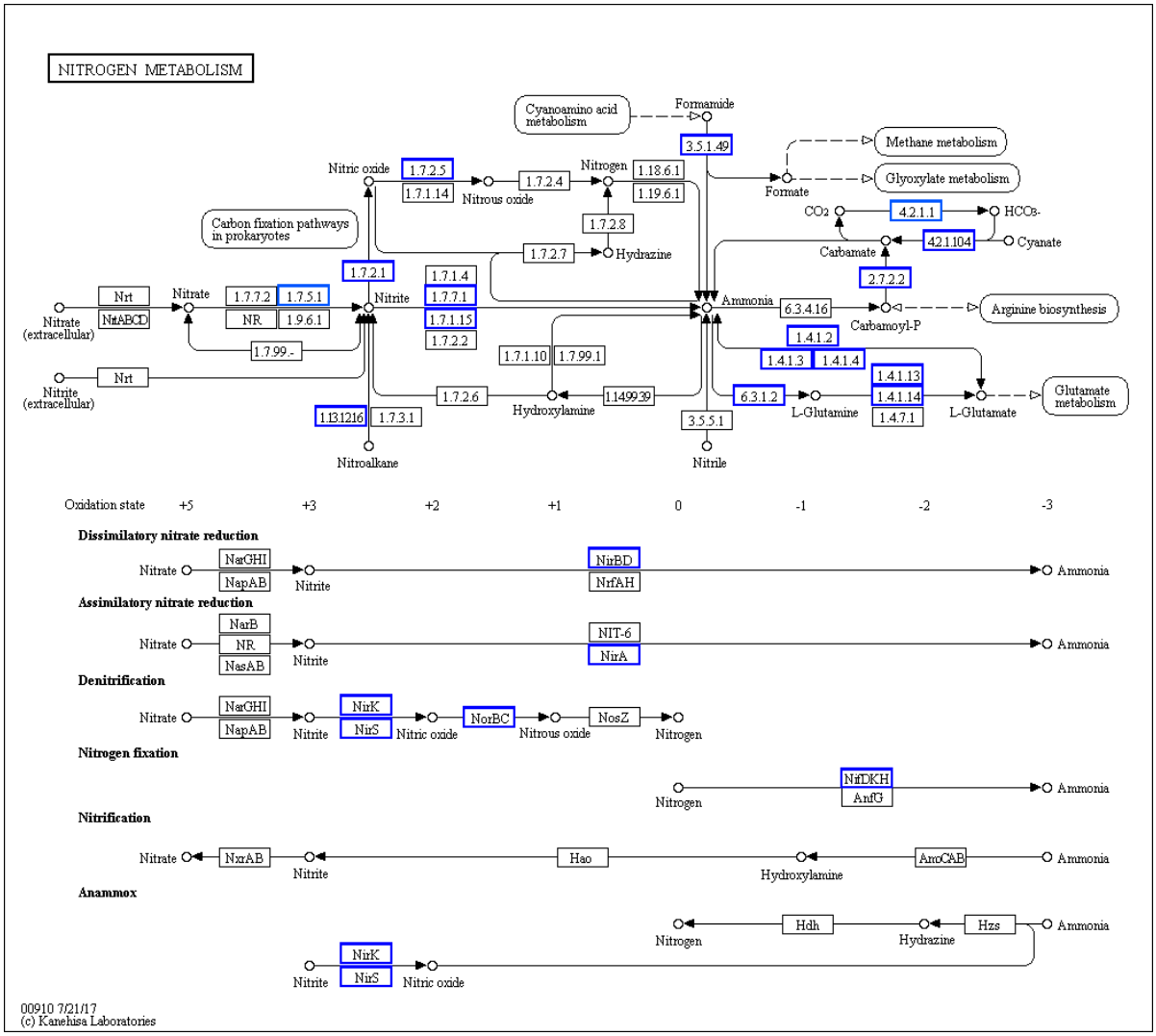


Supplementary Fig. 15. Comparative analysis of phyllospheric bacteria composition and functional predictions in two types of tillandisoids. (a) Rarefaction curves of phyllospheric bacteria in two types of tillandisoids. **(b)** PCoA describing species variation in the composition of phyllospheric bacterial communities. Species in same types are presented using same color makers. **(c)** Relative abundance of phyllospheric bacteria in two types of tillandisoids. **(d-e)** Network analysis of the phyllospheric bacteria of tank-forming **(d)** and atmospheric **(e)**

tillandisoids in genus level. Each node represents a distinct genus, where the size reflects the average relative abundance of that genus. Nodes belonging to the same phylum are color-coded similarly. The thickness of the lines connecting nodes correlates positively with the absolute value of the correlation coefficient for species interactions. Line colors indicate positive (red) or negative (blue) correlations.



Supplementary Fig. 16. Random Forest analysis and functional prediction of phyllospheric bacteria in tillandsioids. (a) Random Forest analysis of the top-30 genera by Mean Decrease Accuracy and Mean Decrease Gin. **(b)** Function prediction heatmap by Tax4Fun of the phyllospheric bacteria in two types of tillandisoids. **(c-e)** Six nitrogenases found in *T. usneoides* phyllospheric bacterial communities by metagenome sequencing. **(c)** The phylogenetic relationship of nifA and nifH from *Bradyrhizobium japonicum*, and 6 nitrogenases. **(d)** Protein structure of nitrogenases. Blue box indicates the domain region. **(e)** The conserved amino acid sequence of nitrogenases.



Supplementary Fig. 17. Enzymes enriched in nitrogen metabolism pathways in *T. usneoides* phyllospheric bacterial communities. Enzymes detected in metagenome are marked by blue boxes.

Supplementary Table 1. The names of the species pictured in Figure 1.

Number	Species
1	*Catopsis subulata*
2	*Catopsis floribunda*
3	*Goudaea ospinae*
4	*Alcantarea nevaresii*
5	*Vriesea corcovadensis*
6	*Vriesea scalaris*
7	*Vriesea fenestralis*
8	*Vriesea vagans*
9	*Guzmania sanguinea*
10	*Guzmania sprucei*
11	*Barfussia laxissima*
12	*Wallisia lindeniana*
13	*Lemeltonia acosta-solisii*
14	*Racinaea tenuispica*
15	*Racinaea spiculosa var. ustulata*
16	*Tillandsia magnusiana*
17	*Tillandsia polystachia*
18	*Tillandsia montana*
19	*Tillandsia bulbosa*
20	*Tillandsia straminea*
21	*Tillandsia mitlaensis var. tulensis*
22	*Tillandsia paraibensis*
23	*Tillandsia remota*
24	*Tillandsia duratii*
25	*Tillandsia tectorum*
26	*Tillandsia malyi*
27	*Tillandsia incarnata*
28	*Tillandsia albida*
29	*Tillandsia crista-gallii*
30	*Tillandsia globosa*
31	*Tillandsia caput-medusae*
32	*Tillandsia rectangula*
33	*Tillandsia usneoides*

Supplementary Table 2. Summary of the sequencing data for genome assembly.

Species		***V. erythrodactylon***	***T. duratii***
Genome size		417.89Mbp	1030.11Mbp
Coding genes number		20, 415	26,185
Hi-C mapping		98.41%	95.80%
contig N50		1.32Mbp	1.92Mbp
scaffold N50		1.32Mbp	1.92Mbp
Assembly assessment	**Complete BUSCOs**	92.50%	97.80%
	CEGMA	97.98%	97.58%
	Illumina read mapping	92.70%	99.33%

Supplementary Table 3. **Summary of the sequencing data for genome assembly.**

Type	Repeat size（% of genome）	
	V. erythrodactylon	***T. duratii***
TRF	5.64%	8.41%
Repeat masker	60%	73.33%
Repeat protein mask	9.76%	16.42%
Total	62.56%	76.93%

Supplementary Table 4. List of RNA probes for *in situ* hybridization.

Gene name	Gene ID	Probe sequences
CYP96A150-a	*Td08g03120*	GCGGAGTGGAGGTAGACTAACTTAC
CYP96A151-a	*Td08g03140*	TGACGAATTCTATGCCGGAGAGCCA
CYP96A152-a	*Td08g03150*	GGTGCGAGCAACGACGATACAATAG
TUA6	*Td01g02320*	ATGGTGCGCTTGGTCTTGATAGTGG
PMEI	*Td15g12100*	ATGATGAACCTCGTCGCGCTGTTGT
S6K2	*Td22g03620*	CTCGATTTCCTGTCTCAAGTTGTGC
LTP3	*Td10g09010*	GCCTGTTTGTCCGCCGTGCTCTTTG
GALT6	*Td03g05520*	TTCGCTTCTATCTCCTCCTCGCTGT

Supplementary references

1. Rabosky, D.L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* ***9***, e89543. (2014).
2. Rabosky, D.L. *et al.* BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* **5**, 701-707 (2014).
3. Team, R.C. R: A language and environment for statistical computing. (2018).
4. Höhna, S., May, M.R. & Moore, B.R. TESS: an R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. *Bioinformatics* **32**, 789-791 (2016).
5. Condamine, F.L., Rolland, J., Höhna, S., Sperling, F.A.H. & Sanmartín, I. Testing the role of the red queen and court jester as drivers of the macroevolution of apollo butterflies. *Syst Biol* **67**, 940-964 (2018).
6. Moore, B.R., Höhna, S., May, M.R., Rannala, B. & Huelsenbeck, J.P. Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proc Natl Acad Sci U S A* **113**, 9569-9574 (2016).
7. Rabosky, D.L., Mitchell, J.S. & Chang, J. Is BAMM Flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. *Syst Biol* **66**, 477-498 (2017).
8. Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. The global diversity of birds in space and time. *Nature* **491**, 444-448 (2012).
9. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526-528 (2019).
10. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764-770 (2011).
11. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170-175 (2021).
12. Zhang, X., Zhang, S., Zhao, Q., Ming, R. & Tang, H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat Plants* **5**, 833-845 (2019).
13. Durand, N.C. *et al.* Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst* **3**, 99-101 (2016).
14. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
15. Price, A.L., Jones, N.C. & Pevzner, P.A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21 Suppl 1**, i351-358 (2005).
16. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573-80 (1999).
17. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**, W265-268 (2007).
18. Can, M. *et al.* Genome sequence of *Kobresia littledalei*, the first chromosome-level genome in the family Cyperaceae. *Sci Data* **7**, 175 (2020).
19. Kawahara, Y. *et al.* Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 4 (2013).
20. D'Hont, A. *et al.* The banana (*Musa acuminata*) genome and the evolution of

monocotyledonous plants. *Nature* **488**, 213-217 (2012).
21. Singh, R. *et al.* Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* **500**, 335-339 (2013).
22. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997).
23. Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *GENOME RESEARCH* **14**, 988-995 (2004).
24. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).
25. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* **28**, 45-48 (2000).
26. Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49 (2012).
27. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).
28. Lowe, T.M. & Eddy, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-964 (1997).
29. Llorens, C. *et al.* The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res* **39**, D70-74 (2011).
30. Steinbiss, S., Willhoeft, U., Gremme, G. & Kurtz, S. Fine-grained annotation and classification of *de novo* predicted LTR retrotransposons. *Nucleic Acids Res* **37**, 7002-7013 (2009).
31. Miele, V., Penel, S. & Duret, L. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* **12**, 116 (2011).
32. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).
33. Vilella, A.J. *et al.* EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19**, 327-335 (2009).
34. Macas, J., Neumann, P. & Navrátilová, A. Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and Medicago truncatula. *BMC Genomics* **8**, 427 (2007).
35. Emms, D.M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).
36. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797 (2004).
37. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
38. De Bie, T., Cristianini, N., Demuth, J.P. & Hahn, M.W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269-1271 (2006).
39. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, (Springer Cham, 2016).
40. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).
42. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. clusterProfiler: an R package for comparing

biological themes among gene clusters. *Omics* **16**, 284-287 (2012).
43. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
44. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
45. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
45. Reiner, A., Yekutieli, D. & Benjamini, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**, 368-375 (2003).
46. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
47. Magoč, T. & Salzberg, S.L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957-2963 (2011).
48. Edgar, R.C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**, 996-8 (2013).
49. Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194-2200 (2011).
50. Noguchi, H., Park, J. & Takagi, T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Research* **34**, 5623-5630 (2006).