A genome for Cissus illustrates features underlying the evolutionary success in dry savannas

Haiping Xin[1], Yi Wang[2], Qingyun Li[1], Tao Wan[1], Yujun Hou[1], Yuanshuang Liu[1], Duncan Gichuki[1], Huimin Zhou[1], Zhenfei Zhu[1], Chen Xu[1], Yadong Zhou[1], Zhiming Liu[1], Rongjun Li[1], Bing Liu[3], Limin Lu[2], Hongsheng Jiang[1], Jisen Zhang[4], Jun-Nan Wan[1], Rishi Aryal[5], Guangwan Hu[1], Zhi-Duan Chen[2], Robert Gituru[6], Zhenchang Liang[3], Jun Wen[7], and Qingfeng Wang[1]

[1]Chinese Academy of Sciences Wuhan Botanical Garden
[2]Institute of Botany Chinese Academy of Sciences
[3]Sino-Africa Joint Research Center Chinese Academy of Sciences
[4]Fujian Agriculture and Forestry University
[5]North Carolina State University at Raleigh
[6]Jomo Kenyatta University of Agriculture and Technology
[7]Smithsonian National Museum of Natural History

July 20, 2022

Abstract

Cissus is the largest genus in Vitaceae and is mainly distributed in the tropics and subtropics. Crassulacean acid metabolism (CAM), a photosynthetic adaptation for the occurrence of succulent leaves or stems, indicates that convergent evolution occurred in response to drought stress during species radiation. Here, we provided the chromosomal level assembly of Cissus rotundifolia (an endemic species in Eastern Africa) and genome-wide comparison with grape to understand genome divergence within an ancient eudicot family. Extensive transcriptome data were produced to illustrate the genetics underpinning C. rotundifolia's ecological adaption to seasonal aridity. The modern karyotype and smaller genome of C. rotundifolia (n = 12, 350.69 Mb/1C), which lack further whole-genome duplication, were mainly derived from gross chromosomal rearrangement such as fusions and segmental duplications, whilst sculpted by a very recent burst of retrotransposons activity. Bias on local gene amplification contributed to its remarkable functional divergence with grape and the specific proliferated genes associated with abiotic and biotic responses (e.g., HSP-20, NBS-LRR) enabled C. rotundifolia to survive in a hostile environment. Re-organization of existing enzymes of CAM characterized as diurnal expression patterns of relevant genes further confer to its present thriver in dry savannas.

Introduction

The plant family Vitaceae is well-known for its economically important fruit crop, the grape (*Vitis vinifera*). It comprises 16 genera with over 950 species and is classified into five tribes (Wen et al., 2018b). Many species in the family are dominant climbers in tropical/temperate forests, savannas, and mountains (Kubitzki et al., 2007), representing one of the earliest diverged lineages in the major Rosid clade of eudicot plants (Zeng et al., 2017; Zhang et al., 2016) (Figure 1a and 1b). The grapevine (PN40024) was the first fruit crop whose genome was decoded (Jaillon et al., 2007).

Cissus L. is the largest genus in Vitaceae, comprising over 300 species (Wen et al., 2018), and the only genus of the tribe Cisseae Rchb. Unlike grapevines which are mostly distributed in temperate regions, *Cissus*

mainly occur in the seasonal arid regions of the tropics and subtropics (DeSanto and Bartoli, 1996). Species in this genus exhibit considerable variations in both chromosome number ($2n$ = 24–66) and genome size (1C = 0.38–1.03 pg) (Chu et al., 2018). Morphological modifications such as succulent leaves or stems have arisen in some *Cissus* species in the face of drought stress (Figure 1a) (Griffiths and Males, 2017). Therefore, these groups provide an opportunity to investigate the strategies of plant adaptive evolution on drought tolerance. Crassulacean acid metabolism (CAM) is a water use efficient adaptation of photosynthesis that has evolved independently many times in diverse lineages of flowering plants (Niechayev et al., 2019). The genomes of CAM plants, including pineapple (*Ananas comosus*), orchid (*Phalaenopsis equestris*),*Kalanchoe fedtschenkoi* , *Dendrobium catenatum* , *D. officinale,* and *Sedum album,* was available (Cai et al., 2015; Ming et al., 2015; Wai et al., 2019; Yan et al., 2015; Yang et al., 2017; Zhang et al., 2016). Comparative analyses between *Kalanchoë fedtschenkoi* and non-CAM species identified the convergence in protein sequence in nocturnal CO_2 fixation and carbohydrate metabolism (Yang et al., 2017). CAM is also widespread in *Cissus*(DeSanto and Bartoli, 1996; Olivares et al., 1984; Sayed, 2001; Ting et al., 1983), enabling us to dissect the convergent evolution of CAM in the plant kingdom.

Cissus rotundifolia Lam. are mainly distributed in the tropical savannas of Eastern Africa. Their leaves are consumed as a local traditional food (Al-Bukhaiti et al., 2019). It has a relatively small genome with 1C = 0.38 pg (Chu et al. , 2018). To understand the adaptive strategies of the genus *Cissus* in the harsh climate, we generated and compared the draft genome of *C. rotundifolia* with that of *V. vinifera* to uncover how the genome evolved and genes regulated underpinning its arid adaptation. Further, we conducted extensive transcriptome comparisons to characterize the evolution of CAM in *C. rotundifolia* .

Materials and Methods

Plant materials. Stem cuttings of *C. rotundifolia* were collected from Endau hill, Kitui County, Kenya (01°17'49" S, 038°31'59" E). Voucher specimens (JAJIT-MU0128) were deposited in the Wuhan Botanical Garden herbarium (HIB). Young leaves of *C. rotundifolia*were collected for genome size evaluation and DNA isolation. Root tips were used for chromosome number determination. Matured leaves, stems, and young roots were collected for RNA isolation and tissue-specific transcriptome analysis.

C. rotundifolia individuals were grown in a glasshouse under artificial conditions (16 h light / 8 h dark, 25 degC) in 3 L pots. After well watering, plants were cultivated for one week and transferred to an incubator with 12 h light (from 6:00 to 18:00) at 30 degC and 12 h dark (18:00 to 6:00) at 20 degC (30 % humidity). After three days, leaves were collected from 18:00 on 9 January 2020 to 18:00 on 10 January 2020 at three-hour intervals. We selected nine time points including 18:00, 21:00, 24:00, 3:00, 6:00, 9:00, 12:00, 15:00, and 18:00 for the study. Three biological replicates were collected for each sample and immediately frozen in liquid nitrogen. Samples were stored at -80 degC for RNA extraction and titratable acid measurement.

Estimation of genome size. About 28.31 Gb Illumina reads were used to evaluate the genome size of *C. rotundifolia* by*K-mer* analysis. The *K-mer* frequency distribution was calculated using Jellyfish (v2.2.6) (Marcais and Kingsford, 2011) with the *K-mer* length of 19. Genomic heterozygosity was estimated by GenomeScope (v2.0) (Ranallo-Benavidez et al., 2020). Flow cytometry was used to determine the nuclear DNA content using hand-chopped materials as described by Galbraith et al. (1983), with minor modifications. Woody plant buffer (WPB) was used instead of Tris-$MgCl_2$ buffer to isolate the nuclei (Loureiro et al., 2007). *Raphanus sativus*cv. Saxa was used as the reference standard. Young leaves of *C. rotundifolia* and *R. sativus* were collected, and the protocols were described in detail by Gichuki et al., 2019.

Chromosome count. Root tips were treated with a saturated solution of 1-Bromonaphthalene for 3 h at room temperature (25–28 degC) to halt cell division. Microscopic slides were prepared from the treated root tips using the protocol developed by Kirov et al. (2014) and Gichuki et al. (2019). The prepared microscopic slides were stained with 4', -6-diamidino-2-phenylindole (DAPI), and images were captured using a fluorescence microscope (Leica DMi8) fitted with a camera (Leica DFC 550).

Genome sequencing. Genomic DNA of *C. rotundifolia* was extracted using CTAB-based protocol as described by Doyle & Doyle (1987), with minor modifications. Briefly, a washing step was included before

CTAB extraction to exclude secondary metabolites. The washing buffer contained 50 mM Tris-HCL, 5 mM EDTA-2Na, 0.35 M D-sorbitol, 1% (W/V) polyvinyl pyrrolidone (PVP-K 30), and 1 % 2-hydroxy-1-ethanethiol. Illumina libraries with 450 bp insertions were constructed according to the Illumina standard protocol, and the paired-end libraries were sequenced to about 77 x coverage on the Illumina Hiseq platform. Approximate 20 kb SMRTbell libraries were prepared and sequenced on the PacBio Sequel system following standard protocols (Berry Genomics Corporation, Beijing, China).

Hi-C library construction and sequencing. Young leaves from the same *C. rotundifolia* plant used for genome sequencing were collected for Hi-C analysis. The Hi-C library construction and sequencing for chromosome-level assembly were implemented by Biomarker Technologies Corporation (Beijing, China) as a previously published method (Xie et al., 2015). Briefly, the shoot tips of *C. rotundifolia* plants were covered with a black box, and the etiolated leaves were fixed with formaldehyde and lysed. The cross-linked DNA was digested overnight using HindIII. Digested fragments were biotinylated and ligated to form chimeric junctions that were enriched, sheared, and processed. Then the libraries were produced on the Illumina Hiseq platform. The paired-end Hi-C reads were uniquely mapped onto the contigs using Juicer (Durand et al., 2016b), and the non-duplicate mapped results were used as the input for the 3D-DNA pipeline (Dudchenko et al., 2017) to construct the genome sequence. Two rounds (-r 2) provided the best results with an assembly of 12 pseudo-chromosomes and N50 of ˜28 Mb. Fine-tuning assembled genome in a graphic and inter-action matrix was drawn in Juicebox (Durand et al., 2016a).

Genome assembly. The contig-level assembly of *C. rotundifolia* genome was processed by combining ˜100 X PacBio long reads and ˜76 X Illumina short reads. The raw PacBio reads were initially corrected and trimmed by CANU-1.8 (Koren et al., 2017) with the parameters of genomeSize = 600m, useGrid = false, maxMemory = 200g, ovsMemory = 16G, ovbConcurrency = 15, ovsConcurrency = 15 and "batOptions = -dg 3 -db 3 -dr 1 -ca 500 -cp 50". Then, the corrected PacBio reads were assembled by two widely used PacBio assemblers, CANU-1.8 (Koren et al. , 2017), and WTDBG2 (Ruan and Li, 2019). We used N50 and the genome size of each assembler to inspect assembly quality. The Canu assembled results were adopted in the end. Illumina reads were used to further polish PacBio assembly using Pilon (Wang et al., 2014) program with parameters: -verbose -mindepth 4 -fix snps,indels -vcf. Then, the redundancy of the assembled sequences was removed to improve the continuity of the assembled contigs using Redundans (Pryszcz and Gabaldon, 2016) with parameters: -identity 0.5 -overlap 0.66 and Purge Haplotigs (Roach et al., 2018) with parameters: -l 15 -m 60 -h 190.

Evaluation of the genome assembly. The genome assembly quality of the *C. rotundifolia* was evaluated using two methods. Firstly, the error rate considering the homozygous mutation was estimated by mapping the 28.31 Gb whole genome sequence (WGS) reads onto this assembly by BWA software. Compared to eudicotyledons_odb10 database, the single copy orthologs in the assembled genome were identified and completeness of the assembly was evaluated using Benchmarking Universal Single-Copy Orthologs v2 (Simao et al., 2015) with the -long and default parameter, respectively.

Repeat sequences annotation. Repeat structures were analyzed by a combined strategy of the de novo prediction and homology-based prediction. A de novo repeat library of *C. rotundifolia* genome was built by RepeatModeler (v1.0.11, http://www.repeatmasker.org/RepeatModeler/) with the -engine ncbi parameter. Using this library, we processed repetitive sequences to annotate, classify, and mark by RepeatMasker (v4.0.7,*http://www.repeatmasker.org/*). Two built libraries were combined with Repbase (v20170127, https://www.girinst.org/) (Jurka et al., 2005) and Dfam (v20170127,*http://www.dfam.org/*) (Hubley et al., 2016) databases with default parameters. SSRs were identified using MISA (*http://pgrc.ipk-gatersleben.de/misa/misa.html*) (Thiel et al., 2003), with the unit lengths ranging from 1 to 7 and the min-length set to 10 bp. LTR were identified by LTR_retriever according to the method of Ou & Jiang (Ou and Jiang, 2018).

Gene model prediction. Three approaches were combined to annotate protein-coding genes. Firstly, Illumina RNA-seq data from three representative tissues were assembled with two different strategies (de novo or genome-guided assembly) by Trinity (v2.2.0) (Haas et al., 2013). The assembled RNA-seq data

were then aligned to the assembled genome and evidence-based prediction by PASA (v2.0.2) (Haas et al., 2003). Secondly, the ab initio methods, AUGUSTUS (v3.3) (Stanke et al., 2006), SNAP (Korf, 2004), and GeneMarkHMM (Lukashin and Borodovsky, 1998) (v 4.32) with default parameters were used to predict gene models with the training of the best candidate genes obtained from PASA (Haas et al., 2003). Thirdly, protein sequences from closely related species, including *V. vinifera* (PN40024) and other *Cissus*species downloaded from NCBI, were used to annotate protein homologs of*C. rotundifolia* by GenomeThreader (https://genomethreader.org/). Finally, the annotation results generated from evidence-based prediction, ab initio prediction, and homologous mapping were combined by EVM (v 1.1.1) (Haas et al., 2008) to integrate the consensus gene model, and genes were renamed according to their position in the genome sequence with the prefix of CRGY (*C. rotundifolia* genome).

Phylogenomic tree construction and gene family analyses. The protein-coding sequences from *C. rotundifolia* and 13 other representative species were used to identify orthologous groups, including those of *Arabidopsis thaliana* , *Oryza sativa* ,*Vitis vinifera* , *Actinidia chinensis* , *Coffea*arabica, *Solanum lycopersicum* , *Populous thichocarpa* ,*Theobroma cacao* , *Carica papaya* , *Citrus sinensis* ,*Fragaria ananassa* , *Malus domestica* , and *Prunus pesica* . All-vs-all BLASTP (Camacho et al., 2009) with an e-value cutoff of 1e-05 was performed, and orthologous genes were clustered using OrthoMCL (Li, 2003). Single copy genes were extracted from the clustering results and performed multiple sequence alignments using MUSCLE (v3.8.31) (Edgar, 2004). After removing low-quality alignment or divergent regions by Gblocks, high-quality aligned protein sequences remained. All aligned sequences were concatenated to one long sequence for each species, and these sequences were used to construct a phylogenetic tree by RAxML (v2.5.1) (Stamatakis, 2006) with PROTGAMMAJTTF model and bootstrap of 1000. MCMCtree (4.8a) from the PAML package (Yang, 2007) was adopted to estimate the species divergence time according to TimeTree (http://www.timetree.org). Four divergence times were used in this analysis, including *Coffea arabica* and*Solanum lycopersicum* , *Arabidopsis thaliana* and*Carica papaya* , *Prunus persica* and *Malus domestica* ,*Fragaria x ananassa* and *Prunus persica* . And the divergence times of *Vitis vinifera* and *C. rotundifolia* in Timetree were also referred in this study. The Markov chain Monte Carlo (MCMC) process analysis was set for 50,000 generations and 50,000 burn-in iterations. The OrthMCL results and time divergence tree were used as the input for CAFE (v3.1) program (Han et al., 2013), which was used to identify expansions and contractions of gene families across 15 plant genomes. The family expansion and contraction were analyzed by Count, and the methods and parameters were according to the study of the*Amborella* genome (Albert et al., 2013). Multi-species orthologous clusters with gene numbers greater than 0 in *V. vinifera* and*C. rotundifolia* were considered orthologous groups between these two species. Expanded orthogroups were defined according to their$p$ -value less than 0.05 and the gene number greater than the average value of multi-species. Dot plot representation of orthologous groups was performed with the R package ggplot2 (*http://ggplot2.org/*) (Wickham, 2011).

Tandem duplication analysis. Four additional typical succulent species, including *Ananas comosus, Hylocereus undatus, Kalanchoe fedtschenkoi,* and *Kalanchoe laxiflora* were added to the gene family clustering (Table S1). Multi-species orthologous clusters with a gene number greater than 0 in *V. vinifera* and*C. rotundifolia* were used to identify the lineage-specific expansion. Expanded orthogroups were identified with a p -value less than 0.05 and the gene number greater than the average value of multi-species. The tandem genes were identified by MCScanX (Wang et al., 2012), which was consistent with the method described in our WGD analysis. The gained TD genes of 18 species were obtained from the Count results and were further categorized into either co-expanded or lineage-specific expanded ones. The TD genes were GO termed by agriGO database (*http://systemsbiology.cau.edu.cn/agriGOv2/index.php*). Further, four succulent plants were annotated both by GO database and agriGO database.

Synteny analyses. All-vs-all BLASTP (Camacho et al., 2009) (e-value 1e-05) and MCScanX (Wang et al., 2012) was used to predict the collinear relationships and positional features between *C. rotundifolia* and *V. vinifera* (PN40024). Blocks larger than ten genes and gaps less than five genes were obtained. The synteny map and dotplot were processed by MCScan and drawn by the python scripts in MCScan packages (Tang et al., 2008).

The segment duplication events were predicted using self-vs-self BLASTP (Camacho et al., 2009) (e-value 1e-05) and MCScanX among the *C. rotundifolia* genome, requiring at least five genes per collinear block. Subsequently, the pairwise sequences from the synteny blocks and segment duplication pairs were processed by ParaAT (v2.0) (Zhang et al., 2012). The values of nonsynonymous mutation rate (*Ka*) and synonymous mutation rate (*Ks*) were calculated using the NG estimation method in Kaks_Calculator (v2.0) (Wang et al., 2010). The visualization plots of the *Ks* distribution were made using a custom R script. Additionally, whole-genome duplication (WGD) events were determined by the distribution of *Ks* of segment duplication pairs and identified by comparisons with the events of *V. vinifera*(PN40024). The AEK was inferred from the genomes of eudicot species with the smallest numbers of historical polyploidization events, including grape, cacao, and peach. Further, the AEK was refined as a post-τ AMK with ten protochromosomes and 13,916 ordered protogenes, a pre-τ AMK with five protochromosomes and 6,707 ordered protogenes (Murat et al., 2017). In the current study, the reconstruction of karyotype of the *V. vinifera* and *C. rotundifolia* were advised by a previous study by Florent Murat (Murat et al., 2017), and the genes and gene orders were used to construct the seven chromosomes and 21 chromosomes of AEK. To cover as many genes as possible, we used version 2.1 of the grape assembly, which anchored 32,424 coding genes (Table S1). MCScanX (Wang et al., 2012), in a BLASTP and dotplot-based approach, was used to detect the syntenic blocks between *C. rotundifolia* vs. AEK and *V. vinifera* vs. AEK with default parameters. The protein of the pre-γAEK and post-γAEK compared to *Cissus* and grape by BLASTP. The syntenic blocks were ordered according to the gene order of *C. rotundifolia* and *V. vinifera.* Some small syntenic blocks and small gaps were abandoned or closed to make the syntenic segments more complete. On the base of dotplot illustrations of the synteny between these two species, the karyotypic structures of the ancestral eudicots were explained by taking into account the fewest number of genomic rearrangements, which may have occurred between the AEK and modern eudicot genomes (Figure S5, Tables S2 and S3).

Detection of significant expansion and contraction in succulent plants. To investigate the significant expansion or contraction of gene families, we divided 18 species into two categories, including five succulent plants (S5) and 13 non-succulent/others plants (O13). Five succulent plants, including *Cissus rotundifolia* , *Ananas comosus* , *Kalanchoe laxiflora* , *Hylocereus undatus* ,*Kalanchoe fedtschenkoi* and other 13 plants were described in Table S1. The average number of genes per orthogroup between two categories was available to evaluate the significant events. For S5 plants, a binomial test with a probability of success of p(W) = 5/18 was used. The criteria of significant expansion or contraction are as follows: 1, A statistically false discovery rate-adjusted p -value < 0.05 from the initial set of 97,344 orthogroups; 2, The minimal contribution of about three for S5 and seven for O13 species; 3, Contribute gene per orthogroup on average satisfied with (S5n/5) / (O13n/13) > 1. We found that 88 orthogroups were expanded (corresponding to 5,696 genes), and 178 were contracted in succulent plants relative to non-succulent plants.

Measurement of titratable acidity. The diurnal changes of titratable acid in leaves of *C. rotundifolia* were measured as described by Chen et al. (1983). The samples were collected as mentioned above. A total of 0.5 g leaves of each sample were cut into pieces, placed in centrifugal tubes, and boiled for 30 min after adding 10 mL CO_2-free distilled water. The supernatant after centrifugation was reserved. Additional 10 mL CO_2-free water was added to the pellet to extract and centrifuge again. Total supernatants obtained by the two-stage extraction process were titrated to pH 8.3 with 0.01 mol/L NaOH, and the acidity of the leaf was represented as μeq of acid per gram fresh weight (μeq g^{-1} FW).

RNA extraction and RNA-Seq library preparation. Total RNA was extracted from the samples using the Universal Plant Total RNA Fast Extraction Kit (BioTeke Corporation, Beijing, China). RNase-free DNase I was used to remove DNA from the extracted RNA. The purity and concentration of RNA were determined by a Nano Drop and Agilent 2100 bioanalyzer (Thermo Fisher Scientific, MA, USA). Subsequently, mRNA enriched by Oligo (dT) - attached magnetic beads was randomly fragmented into short pieces with an additional fragmentation buffer. Then, first-strand cDNA was synthesized by random hexamer-primed reverse transcription, followed by second-strand cDNA synthesis. A-Tailing Mix and RNA Index Adapters were added by incubating to end repair. The obtained cDNA fragments were amplified by PCR, and then products were purified by Ampure XP Beads. Agilent Technologies 2100 bioanalyzer was

used for quality control of products. Finally, the cDNA library was constructed, and the MGISEQ-2000 platform was used for paired-end sequencing (2 x 150 bp). Approximately 40 million bp were generated for each sample.

Transcriptome analysis. The quality of paired-end raw transcriptome data was checked by FastQC v0.11.8 and trimmed using Trimmomatic (v0.36) (Bolger et al., 2014). Then the trimmed reads were mapped onto *C. rotundifolia* latest assembled genome through TopHat (v2.1.1) (Trapnell et al., 2012). Using the gene model of*C. rotundifolia* , the expression levels of genes represented by FPKM (Fragments per Kilobase Million) for each sample were calculated by Cufflinks (v2.2.1) (Trapnell et al., 2012) with default parameters. The genes involved in the stomatal movement process and CAM pathway were picked to show their expression patterns by 'pheatmap' package in R.

Identification of CAM pathway and stomatal movement process-related genes. The genomes of *Ananas comosus* , *V. vinifera* (PN40024), *O. sativa* , *Zea mays* , and*Phalaenopsis equestris* were downloaded from Pineapple Genomics Database (PGD, http://pineapple.angiosperms.org/pineapple/html/index.html) (Xu et al., 2018), Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html) (Goodstein et al., 2012), Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/) (Kawahara et al., 2013), MaizeGDB (https://maizegdb.org/) (Portwood et al., 2019), and NCBI (https://www.ncbi.nlm.nih.gov/), respectively. Further, the CAM gene list of *Kalanchoë* was obtained from a supplemental table of its genome (Yang et al., 2017). The list of gene families, which included*carbonic anhydrase* (*CA*), *phosphoenolpyrvuate carboxylase* (*PEPC*), *phosphoenolpyruvate carboxylase kinase* (*PEPCK*), *malate dehydrogenase* (*MDH*),*malic enzyme* (*ME*), *phosphoenolpyruvate carboxykinase* (*PPCK*), and *pyruvate phosphate dikinase regulatory protein* (*PPDKRP*), was obtained from PGD (Xu et al., 2018). All given gene sequences in each family from pineapple, *O. sativa* , and *Z. mays* , *Phalaenopsis equestris* and*Kalanchoë* were used as queries to search corresponding family members in *C. rotundifolia* and *V. vinifera* (PN40024) by BLASTP. The genes with alignment length > 100bp and e-value < 1e-05 were considered as potential members. Then online software CD-search (*https://www.ncbi.nlm.nih.gov/cdd*) (Marchler-Bauer and Bryant, 2004) and PFAM (*https://pfam.xfam.org/*) (El-Gebali et al., 2019) were used to detect the specific domain. The genes without a unique domain of gene family were abandoned. Then the remaining genes were defined as candidate members and used for further analyses. Diel expression dataset of *Arabidopsis* C3 leaf (Mockler et al., 2007) and pineapple CAM leaf were used to compare with CAM-genes shown in Figure 4b in *C. rotundifolia* , whose orthologs were identified by BLASTP based on sequence similarity, and then, gene-pairs between two species were used to calculate their relationship (Pearson and Spearman) of transcript expression (Table S4). On the base of satisfying two correlation coefficients (Pearson and Spearman), genes ($R_{cr\text{-}at} < 0.5$) were determined as not correlative expression patterns during a day/night cycle between*Cissus* and *Arabidopsis* . Gene pairs ($R_{cr\text{-}at} < 0.5$ and $R_{cr\text{-}ac} > 0.8$) were defined strongly CAM genes. The genes for stomatal movement were identified using BLASTP with an Evalue cutoff of 1e-5 based on orthology in *Arabidopsis* as described by Chen et al. (2020).

Co-expression network and cluster analysis . Transcripts with average FPKM > 1 (calculated from three biological replicates) in at least one of the nine samples were used to construct a weighted gene co-expression network by R package WGCNA. The transcript expression was log2 transformed. Modules were constructed using the following parameters: power = 16, networkType = "signe", mergeCutHeight = 0.18, corType = "bicor", minModuleSize = 30. All the nine time point transcripts with three replicates were used to perform cluster analysis by maSigPro package. The parameters were as following: degree = 3, counts = F, MT.adjust = "BH". Transcripts were marked as influential by the T.fit() function. Genes with "non-flat" significantly changed across the nine time points. Nine clusters were displayed using the "see.genes" function with cluster.method="hclust", k=9 in maSigPro. The network of each cluster was constructed by ARACNE algorithm with 'Discovery' mode and 'Naive bayes' mutual information (MI) algorithm type in Cytoscape software. The p -value was calculated based on MI, in which, less than 0.05 were selected in each cluster. One percent of genes with at least ten edges in each network were selected by cytoHubba, and CAM genes also were chosen based on a minimum of ten directed edges.

***Cis*-element annotation and enrichment analysis of CAM related genes.** Promoter sequences in 2kb upstream of genes involved in CAM were extracted from the *C. rotundifolia* genome. Of all the promoter sequences, the *cis* -element enrichment of light, circadian, temperature, and drought in CAM-related and stomatal movement-related promoters were implemented by FIMO (Grant et al., 2011) program with p -value < 0.0002 in MEME. Enrichment analysis of about five known *cis* -elements including the morning element (CCACAC), the evening element (AAAATATCT), the CCA1-binding site (AAAAATCT), the TCP15 element (NGGNCCCAC), and the G-box element (G-box; CACGTG) (Michael and McClung, 2002) were performed by FIMO (Grant et al., 2011) program.

Results

Genome assembly, annotation, and repetitive sequences characterization

We assembled a highly heterozygous (1.19 %) genome of *C. rotundifolia* , by combining the 39.38 gigabases (Gb) of PacBio Sequel sequences (˜ 106 x) and 28.31 Gb of Illumina paired-end reads (˜ 77 x) (Figure S1, Table S5). We arranged 3,289 contigs (contig N50 = 186 Kb) based on the spatial relationship deduced from 130.44 Gb of Hi-C assay data (˜ 362 x) (Table S6). A total length of 350.69 Mb scaffolds was ordered and anchored onto 12 pseudo-chromosomes with scaffold N50 up to 27.6 Mb, covering 94.53 % of the assembled genome (Figure 1c, Figure S1, Table S7). We identified 169,723 homozygous mutation bases representing 0.045 % of assembled genomes (one error per 2.22 Kb).

A total of 30,824 protein-coding genes were predicted by using a combination of ab initio, transcript evidence, and homology-based methods. We used Swissport, NCBI, GO, KEGG, and eggNOG databases to annotate approximately 82.15 % of the coding genes (Table S8). Moreover, Benchmarking Universal Single-Copy Orthologs analysis suggested that 92.4 % of the genes could be recovered (Table S9). In addition, we identified 692 transfer RNAs, 128 microRNAs, 232 ribosomal RNAs (18S, 28S, 5.8S, and 5S), and 971 small nucleolar RNAs (Figure S2).

Repetitive sequences dominated 47.41 % of the genome, of which 31.07 % were long terminal repeat (LTR) elements (Table S10). Estimates of sequence divergence times between the adjacent 5' and 3' LTRs of the same retrotransposon suggested a very recent burst of activity in less than 90.77 thousand years ago (kya) and much severe invasion than in grape (Figure 1d, Table S10). Further, we found 584,679 (12.90 Mb) simple sequence repeats (SSRs) with six as the most abundance unit size, slightly less than that in *V. vinifera* (PN40024, 930,680, 23.05 Mb) (Table S11).

Gross chromosomal shuffling reassembled the *C. rotundifolia* genome

We collected a total of 342 single copy genes (61,639 homologous amino acids) among 13 representative angiosperms to clarify the divergence of*C. rotundifolia* (hereafter *Cissus*) and *V. vinifera*(hereafter grape) (Figure S3, Table S1). Reconstruction of the phylogeny indicated these two species had separated as early as 60.19–84.68 million years ago (mya) (Figure 1b), coincided with the distribution pattern of synonymous substitutions per synonymous site (Ks) (a peak of distribution as $Ks = 0.33$) (Figure 1e). The subsequentKs analysis of all paralogous genes in the genomes of*Cissus* and grape and syntenic regions support a shared whole-genome triplication, namely WGT-γ, the 'pivot' palaeo-hexaploidy event that occurred in the most recent common ancestors (MRCA) of all eudicots (Albert et al., 2013; Jaillon et al., 2007) (Figure 1b and 1e). No more signatures of whole-genome duplication (WGD) were observed in*Cissus* and grape genomes. Nevertheless, in *Cissus* , there is another small peak of duplicated genes with $Ks =$ ˜ 0.1, and the majority of the paired genes were devoid of inter-chromosomal regions (Figure 1e, Figure S4). Approximately 236 duplication events occurred inside their chromosome and were characterized as segmental duplications. Such recent local gene cluster duplication burst finally accounted for 8.31 % of *Cissus*functional genome profiles (4.75 % in grape, Table S12).

Interestingly, gene ontology analysis (GO) of the segmental duplications gene clusters in two species revealed similar function enrichments mostly associated with basic biological processes such as phosphorus metabolic process and cellular protein metabolic process. Meanwhile, specific biochemical pathways (e.g., brassinos-

teroid homeostasis) and chromosome dynamics (e.g., meiotic chromosome condensation and meiotic sister chromatid cohesion) are only enriched in *Cissus* (Table S12). It would be worthy of further investigation on the role of the 'connected' gene cluster as a module of function during speciation and the retention of duplicated segments with gene dosage relationship preserved (Freeling, 2009).

Considerable high collinearity was observed between *Cissus* and grape chromosomes, presenting a pattern as a combination of each two of the 19 chromosomes in grape often correspond to one chromosome in *Cissus* , leading to fewer monoploid chromosome numbers in the latter (n=12) (Figure 2a). To search for genomic features that might contribute to *Cissus* ' modern 12 chromosomes, we compared the ancestral eudicot karyotype (AEK) reconstructed from an integration of the Vitales (grape), Malvales (cacao) and Rosales (peach) major subfamilies to uncover that at least five fusions occurred in *Cissus* after inheriting 21 AEK post-γ chromosomes from the MRCA of eudicots (Murat et al., 2017) (Figure 2b). Specifically, pairwise comparisons among *Cissus* , grape and AEK post-γ revealed that 82.59 % of grape genomic regions were linked to AEK post-γ, higher than the number (71.24 %) of *Cissus* (Table S2). This may partly be due to the ancestral reference derived from a comparison of grape-cacao-peach, particularly when the grape preserved more ancestral genomic organizations. Alternatively, the *Cissus* genome may have lost more ancestral gene arrays than the grape, probably attributed to a higher frequency of chromosomal rearrangement and recombination (Wan et al., 2021; Xiong et al., 2011). Beyond that, 17.7 % of grape genes (4,757 genes) were embedded in syntenic blocks with a 3:1 relationship to each *Amborella trichopoda* (Albert et al., 2013) region that resulted from WGT-γ, higher than syntenic block genes in *Cissus* genome (13.7 %, 3,687 genes) (Figure 2c and 2d, Table S3). Likewise, the number of syntenic genes was lower in *Cissus* (19.3 %, 4,188 genes) than in grape (23.4 %, 5,089 genes) when aligned with *Aristolochia fimbriata* (Qin et al., 2021) (A species that is similar to *Amborella* lack of further WGDs since the origin of extant angiosperms). Moreover, 2,419 genes harbored in chromosome 1 were identified as 76 %, of which were specific to *Cissus* (Figure 2b, Figure S5, Table S13). Together with the above, it would imply a more diverged genome of *Cissus* reshaped after long-term separation with *Vitis* .

Functional divergence enabled *Cissus* adaptation to aridity

In *Cissus* , 675 orthogroups were remarkably expanded ($P < 0.05$) and 232 were diminished ($P < 0.05$) compared to other representative eudicots (Figure S3). The expanded orthogroups were mainly enriched in the abiotic/biotic stress-responsive pathways, metabolism of carbohydrates, and hormone biosynthesis (Figure 3a). *Cytochrome P450* , found in dramatic proliferation (Table S14), could contribute to the foliar wax deposition in *Cissus* (Shepherd and Wynne Griffiths, 2006). Some polysaccharide-related genes, such as pectate lyase, pectinesterase, and polysaccharide biosynthesis genes, also displayed an increased paralogous number, probably attributable to the succulent leaf formation through the modification of pectin and other polysaccharides in cells (Griffiths and Males, 2017). Apart from the genes that directly contributed to the leaf character, transcription factors like *MYB* , *WRKY* , *AP2/ERF* , *GRAS* , and *LEA* were strongly expanded (Figure 3a), suggesting that the probable enhanced abiotic stress resistance and secondary metabolism in *Cissus* (Dubos et al., 2010; Gao and Lan, 2016; Jiang and Rao, 2020). To further address the functional divergence of *Cissus* referring to adaptation, we compared its gene repertoire to grape which indicated that selective amplification of genes belonging to plant immunity had occurred in these two species (Figure S6, Table S15). Among respective orthogroups, nucleotide-binding site leucine-rich repeat (*NBS-LRR*) genes were found in favor of expansion in both species but showed preference for different subclasses (e.g., orthogroup 12 in the grape; orthogroup 4,7 in *Cissus*). The co-abundance of *R* genes would represent the basic objective of an organism to protecting itself against the surging threats from microbial pathogens (Plomion et al., 2018; Tobias and Guest, 2014). The significant copy number variation of paralogous genes (orthogroup 13: terpenoid cyclase, orthogroup 2: TMV resistance protein N-like) likely suggested the different responses to pathogens induction (Mestre and Baulcombe, 2006; Warren et al., 2015). Additionally, small heat shock proteins (sHSPs), *HSP20-like* were found particularly amplified in *Cissus* and upregulated in its shoot and leaves compared with root (Figure S6, Table S16). This would fairly reflect the increased ability of *Cissus* 's vegetative organs to deal with heat shock and promote resistance to environmental stress factors (Bondino et al., 2012; Guo et al., 2020). The enrichment pattern of the

gene family in *Cissus* led us to investigate if a similar preference for gene proliferation occurred in other succulent species. To this end, we took another four typical succulent plants (*Ananas comosus* , *Hylocereus undatus* ,*Kalanchoe fedtschenkoi,* and *Kalanchoe laxiflora*) into account on the gene family comparison. We found that 88 of the 97,335 orthogroups demonstrated as succulent-specific expansion, which significantly (P -adjust < 0.05) enriched in '*terpene synthase* ', '*HSP20* ' (Figure 3b, Tables S17-S19). However, 178 orthogroups GO termed mainly as serine/threonine-protein kinase receptor precursor (*SKR*), cysteine-rich receptor-like protein kinase (*CRK*), wall-associated receptor kinase (*RLK*) were observed in co-expansion in the other 13 non-succulent plant genomes investigated (Figure 3b, Tables S17 and S18, Table S20). The diverged preference for functional gene families would reflect a specialized convergent mechanism in succulent plants dealing with high temperatures and water deficiency (Griffiths and Males, 2017). On the other hand, we identified 1,878 tandemly duplicate (TD) arrays of two or more genes in *Cissus* , and the total number of genes in such arrays is 4,746, slightly higher than 3,958 genes in 1,524 TD arrays in grape (Table S21). There are 2,582 TD genes shared in two species, whose functional classification is mainly enriched in 134 GO terms (e.g., oxidoreductase activity, oxidation-reduction process, and response to auxin), and a total of 2,164 TD genes are species-specific in*Cissus* (Table S22). Functional bias in TD retention was observed encompassing different periods of evolution in *Cissus* (Figure 3c). An overrepresented number of genes in the *Cissus* lineage were enriched in cell wall-related pathways (e.g., cell wall modification, cell wall organization and xyloglucan metabolic activity), probably, conferred to its succulent leaves or stems (Ahl et al., 2019). In contrast, functional categories specific to grapes were mainly associated with stress responses (Figure 3c, Table S23). The result is consistent with the notion that TD genes would have a lineage-specific selection (Freeling, 2009). Nevertheless, earlier studies in*Arabidopsis* and rice demonstrated that an elevated probability of retention of stress-responsive TD is preferential for adaptive evolution after speciation (Hanada et al., 2008; Rizzon et al., 2006). The functional bias of TD in *Cissus* indicates genes referred to as morphological innovation for adaptation might be particularly selected and expanded via local duplication. It would be interesting to check if a similar profile of lineage-specific TD is exhibited in other morphology-specialized plants. We found that lineage-TD genes categorized as 'cellular component-related' and 'resistance' were overrepresented in gross tandem duplicated genes in succulent species, in contrast to the discrete pattern that occurred in the other non-succulent plants (P = 0.03 and P = 0.007) (Figure 3d, Table S24). Local gene amplification with a high frequency of gene birth/death plays a critical role in plants' adaptive responses to environmental stimuli and is mostly attributable to gene copy number and allelic variation within a population (Hanada et al., 2008; Jiang and Rao, 2020). The succulent fashion of TD expansion observed here would suggest another pattern of functional bias in TD retention during seed plant evolution. We speculated that the intense environment change provided multi-options for plants on morphological innovation and rapid expansion of resistance genes.

CAM photosynthesis in *C.* rotundifolia

CAM photosynthesis is a recurrently evolved strategy for high water use efficiency (WUE), enabling plants to survive in water-limited environments (Silvera et al., 2010). In CAM plants, the carbon dioxide (CO_2) is fixed in the cytosol and stored as malic acid in the vacuole during the night (Figure 4c). The stomata remained closed during the daytime to decrease water loss by evapotranspiration, and the stored malic acid is decarboxylated to release CO_2 that could be re-fixed through the Calvin-Benson cycle (Borland et al., 2014). Such a feature of CO_2 uptake was ubiquitous in*Cissus* lineage, which may have facilitated the spread of the genus from wet into arid tropics (DeSanto and Bartoli, 1996).

To investigate the CAM evolution in *Cissus* , we determined the pattern of diurnal oscillation of titratable acidity in the leaves of*C. rotundifolia* in growth chambers with a climate close to the dry seasons in Kenya (Figure 4a). The amount of titratable acid reached its maximum (150 μeq g^{-1} FW) early in the dawn (~ 6:00 a.m.) and dropped to its minimum (20 μeq g^{-1} FW) later in the day (~ 6:00 p.m.), which qualified *C. rotundifolia* as a CAM species (Nelson and Sage, 2008; Sayed, 2001). We identified 47 candidate CAM pathway genes based on their orthologs in pineapple (*Ananas comosus* L. Merr., CAM plant) (Ming et al., 2015), maize (*Zea mays* L., C4 plant) (Schnable et al., 2009), rice (*Oryza sativa* L., C3 plant),*Kalanchoe fedtschenkoi* (CAM plant) (Yang et al., 2017), and*Phalaenopsis equestris* (CAM plant) (Cai et al., 2015).

Further, these genes were well categorized into nine gene families that characterized the core network of carboxylation and decarboxylation pathways (Ming et al., 2015) (Table S25). These gene families showed no significant expansions in *C. rotundifolia* compared with other plants, as shown in Table S26, implying that CAM photosynthesis might evolve through the re-organization of existing enzymes (Chen et al., 2020).

The diurnal expression patterns of these CAM genes were interrogated by transcriptome comparison of leaves during 3-hour intervals over a 24-hour period. In general, the expression of 21 genes showed typical circadian patterns as defined via a polynomial regression (Figure 4b). The transcripts of enzymes involved in carbon assimilation such as*carbonic anhydrase* (*CA*), *phosphoenolpyruvate carboxylase kinase* (*PPCK*), and *malate dehydrogenase*(*MDH*) were highly accumulated at night. As the core gene involved in CO_2 fixation, four PEPC genes of *Cissus* have extremely high expression in the daytime rather than at nighttime (Figure S7). Similar expression patterns were also found in other CAM plants, such as *Kaladp0095s0055.1* in *Kalanchoe* and*Sal_001109* in *Sedum album* (Abraham et al., 2020; Wai et al., 2019; Yang et al., 2017; Zhang et al., 2016). Correspondingly, the enzymes that participate in decarboxylation processes, such as*MDH* , *ME-NADP* , and *phosphoenolpyruvate carboxykinase*(*PEPCK*) were highly expressed during the day (Figure 4b). Interestingly, as a major protein for carbon fixation, previous studies in pineapple have shown that only βCA subfamily is expressed at nighttime and early morning in green leaf tissues (Ming et al., 2015). We observed all five *CAs* including α (1), β (3), and γ (1) expressed at night in *C. rotundifolia* (Figure 4b). The expression of β*CA1* in *Cissus* and pineapple ($R_{cr\_ac} > 0.8$) increased during the night, and a peak occurred at 9:00 in the morning. While its orthologs in *Arabidopsis*($R_{cr\_at} < 0.5$) showed stable and lower expression during the diurnal cycle (Figure S8, Table S4) (Mockler et al., 2007). Beyond that, members of the *MDH* also showed diverged expression patterns as *MDH2* was more active at night while the other four *MDHs* were upregulated during the day, consistent with other CAM plants (Ming et al., 2015; Yang et al., 2017; Wickell et al., 2021) (Figure 4b). This may be associated with their different roles in decarboxylation processes since *MDH* catalyzes the reversible reaction between oxaloacetic acid and malic acid.

We constructed the gene co-expression network based on the transcriptome data from nine mature leaf samples collected every three hours over a 24-hour period. Among 27 modules identified, MEbrown2 (2,020 genes that were highly expressed during the night) was significantly ($P< 0.05$) related to the night period (Figure S9). We found thatβ*CA2* , β*CA3,* and γCA were also found in this MEbrown2 module. Pathways such as response to organonitrogen compound and root meristem growth in this module were significantly enriched in this module (Table S27). MEdarkorange2 module (311 genes that were highly expressed in the day) was found to be significantly associated with the day period. We found *PEPCK* , *PPDK* , *MDH6,*and *ALMT* s in this module. Biological processes such as response to abiotic stimulus were enriched in this module (Table S27).

Moreover, transcripts in leaf with time-course diel expression patterns were classified into 9 clusters (Figure S10, Table S28). The highly connected hubs genes identified by network construction for each cluster were associated with CAM genes. For example, Cluster 4 contained*PPCK2* (CRGY0218762) and γCA (CRGY0214246) and had patatin-like phospholipase as the hub (Figure 4d, Table S29). Heat shock protein, which played important roles during stress responses in many plants, was the hub in Cluster5 and connected with*PEPC1* and *PEPC5* (Figure 4e, Table S29).

The promoters of the diurnally expressed photosynthetic genes were enriched in circadian clock-related *cis* -elements (Chen et al., 2020; Michael et al., 2008) (Figure 4b). Comparative analysis between*Cissus* , pineapple, rice, maize, and sorghum showed that onlyβ*CA1* with typical circadian patterns in *Cissus* had one EE (Table S30) (Ming et al., 2015), suggesting its contribution to CO_2 fixation via combination EE motif during nighttime (Wai et al., 2019). Additional comparison within CAM genes indicated that EE and G-box elements were mainly enriched in the subgroups of highly expressed genes at night (Table S30).

The higher WUE in CAM plants relied on the appropriate control of stomatal movement during day and night. We identified the stomata open/close related genes in the *C. rotundifolia* genome based on their homologs in *Arabidopsis* (Chen et al., 2020) (Table S31). A subset of genes that are responsible for the stomata opening or closing were uniquely expressed either at night or during the daytime, which implied

the coincidental organization of stomata movement and CAM genes (Figure S11). The expression patterns of stomata movement genes were compared to their orthologs in *Arabidopsis* (Table S4). We identified 86 out of 141 stomata movement genes with diurnal expression patterns in *Cissus* (Table S31). The diurnal expression of 64 genes showed a low correlation with its orthologs in *Arabidopsis*(Figure S11), suggesting their putative roles during stomatal movement in *Cissus* . *OST1* (Stomatal opening factor1), which plays a vital role in abscisic acid (ABA) triggered stomatal closure (Mustilli et al., 2002), was found to be highly expressed at 9:00 (Figure S11), compared with accumulated transcription of its orthologs at night in*Arabidopsis* . The result was also consistent with diel expression patterns of OST1 in *A. americana* , *K. laxiflora,* and*Kalanchoë* (Boxall et al., 2020**;** Abraham et al., 2016; Abraham et al., 2020). Interestingly, the MOE and G-box motifs were enriched in the promoter of *OST1* in *C. rotundifolia* but not in *V. vinifera* and *Arabidopsis* (Table S31). These results indicate coordinated transcriptional regulation of circadian rhythm and stomatal movement-related genes with evolved CAM in *C. rotundifolia* .

Discussion

Vitaceae is a sister to most of the Rosids in the highly diverse Rosid clade of the flowering plants (Soltis et al., 2000). Grape (*V. vinifera*) was believed as one of the most slowly evolved species representing a more conserved ancestral structure of the genome that can be used to unravel the evolution and genome duplication history of other eudicots (Murat et al., 2017). Here, we present another genome in Vitaceae, *C. rotundifolia* , to show probable diverse evolutionary history considering distinct ecological niches. It is exciting to observe the evidence of the palaeo-hexaploidy event shared by*Cissus* and grape. Moreover, the lack of any other WGDs suggests*Cissus* also might hold a relatively ancestral state of genome organization after divergence from their common ancestors (Chanderbali et al., 2022; Qin et al., 2021; Van de Peer et al., 2009). This would be reflected by ˜13.7 % of the total genes in*Cissus* (˜17.7 % of grape) belonging to the 3:1 duplicated region to *Amborella* and considerable high collinearity of chromosomes to grape. Besides, we proposed that more frequent chromosome shuffling, including at least five independent fusion events, might occur in *Cissus* after its origin (three fusions in grape). The recent segmental duplications probably further accounted for the increased genetic and biological complexity (Bondino et al., 2012), together with chromosome fusions serving as a prelude to the modern karyotype configuration of Vitaceae.

The very recent burst of activity in LTRs (90.77 kya) detected in*Cissus* could probably be explained by the severe climate transition from arid to wet that occurred in the past 100 kya in eastern Africa, where substantial ecological habitats turnover was recorded (Figure 1d) (Alzohairy et al., 2012). However, such server invasion of retrotransposons left a smaller genome (350.69 Mb) than grape (475 Mb). Indeed, we found much higher gene density in *Cissus* than in grape, indicating the lower frequency of repeats in intergenic regions where LTR insertion is usually preferred (Table S32). Moreover,*Cissus* probably had experienced a fair loss of ancient genomic arrays compared to grape. Together, transposable element removal and sequence elimination accompanied by chromosomal rearrangement (e.g., chromosomal fusion and recombination) could contribute to the selected, size-reduced genome of *Cissus* (Wan et al., 2021; Xiong et al., 2011). Nevertheless, a small genome size could be particularly advantageous for plants to enhance water use efficiency through increased stomatal responsiveness of smaller cells (Drake et al., 2013; Roddy et al., 2020).

Seasonal drought is one of the biggest challenges for agriculture in East Africa. The evolution of the water storage tissue of plants is the most common adaptive strategy in arid and semi-arid regions (Eggli and Nyffeler, 2009). The leaves of *C. rotundifolia* are succulent, which exemplifies a convergent evolution with plants from dry habitats like Agave (Newton & Chan, 1998). We found gene families of enzymes responsible for the polysaccharide synthesis, such as pectate lyase and pectinesterase, were remarkably expanded in *Cissus* . Therefore, modified pectin and other polysaccharides in cells may confer to the occurrence of succulent leaves (Morse, 1990). Signature of the noticeable proliferation of gene families associated with biotic and abiotic responses (i.e., *P450* , *LEA* , and *LRR*) would play key roles in the objective arms race against pathogens and unfavorable environment (Hundertmark and Hincha, 2008; Plomion et al. , 2018; Rai et al., 2015).

A clear pattern of selective amplification of immunity genes in*Cissus* and *Vitis* was present, indicating a

potential functional divergence related to adaptations. Further, succulent-specific expansion in a certain gene family (e.g.,*terpene synthase* ', '*HSP20-like* ') suggested a convergent mechanism in such a morphologically modified group. Interestingly, the succulent fashion of TD retention was also correlated to morphological innovation, which might unveil another functional bias pattern of TD content in the face of rapid and intense environmental change during seed plant evolution.

The innovation of the CAM photosynthetic pathway in *Cissus*further contributes to its adaptation in the dry savannas by enhancing the WUE (Bloom and Troughton, 1979). The decarboxylation in*Cissus* is likely induced in two ways: one is driven by ME and PPDK enzymes, and another is catalyzed by PEPCK enzyme (Figure 4c, Figure S8). For decarboxylase process, ME and PPDK enzymes were used in*K. fedtschenkoi* and PEPCK enzyme was utilized in pineapple (Ming et al., 2015; Yang et al., 2017). The genes had undergone convergent evolution in *Kalanchoe fedtschenkoi* , which included PEPC, nucleosome assembly protein 1-like 4 (NAP1L4), transcription factor hy5-like protein (HY5), and chloroplast-localized glucose-6-phosphate isomerase (GPI). However, no amino acids showed convergent evolution patterns for CAM and stoma-related genes in *Cissus* by a similar analysis (Yang et al., 2017), suggesting that the evolution of the characteristics may be derived from multiple modifications. In*Sedum album* , the number variation of *cis* -elements between C3 and CAM-cycling status showed a phase shift during the daytime (Wai et al., 2019). While *cis* -elements of CAM cycling genes in*Isoetes howellii* are not strongly associated with transcript expression, additionally lacking ME and G-box on promoters of CAM genes (Wickell et al., 2021). The EE is over-presented in the promoter of evening-phased genes (Huang et al., 2016). Compared to other plants (Ming et al., 2015), *βˆA1* with one EE in *Cissus* maybe contribute to CO_2 fixation during nighttime (Table S30). The identification of *cis* -regulatory elements in the promoter of CAM genes in *C. rotundifolia* would help to explain the evolution of CAM from C3 plants and provide valuable information for breeding drought-tolerant crops.

Acknowledgements

This work was supported by the National Science Foundation of China (31961143026) and Scientific Research Program of Sino-Africa Joint Research Center (SAJC201614 and SAJL201607). We would like to thank the Institute of Experimental Botany, Czech Republic for kindly provide the seeds of *Raphanus sativus* cv. Saxa as standard for flow cytometry.

References

Abraham, P. E., Castano, N. H., Cowan-Turner, D., Barnes, J., Poudel, S., Hettich, R., Flutsch, S., Santelia, D., & Borland, A. M. (2020). Peeling back the layers of crassulacean acid metabolism: functional differentiation between *Kalanchoë fedtschenkoi* epidermis and mesophyll proteomes. *The Plant Journal* , *103* (2), 869-888.

Abraham, P. E., Yin, H., Borland, A. M., Weighill, D., Lim, S. D., De Paoli, H. C., Engle, N., Jones, P. C., Agh, R., Weston, D. J., et al. (2016). Transcript, protein and metabolite temporal dynamics in the CAM plant Agave. *Nat. Plants* , *2* , 16178.

Ahl, L. I., Mravec, J., Jorgensen, B., Rudall, P. J., Ronsted, N., & Grace, O. M. (2019). Dynamics of intracellular mannan and cell wall folding in the drought responses of succulent *Aloe* species.*Plant Cell and Environment* , *42* , 2458-2471.

Al-Bukhaiti, W. Q., Noman, A., Mahdi, A. A., Abed, S. M., Ali, A. H., Mohamed, J. K., & Wang, H. (2019). Proximate composition, nutritional evaluation and functional properties of a promising food: Arabian wax Cissus (*Cissus rotundifolia* Forssk) leaves. *J. Food Sci. Technol.* , *56* , 4844-4854.

Albert, V. A., Barbazuk, W. B., dePamphilis, C. W., Der, J. P., Leebens-Mack, J., Ma, H., Palmer, J. D., Rounsley, S., Sankoff, D., Schuster, S. C., et al. (2013). The *Amborella* genome and the evolution of flowering plants. *Science* , *342* , 1241089.

Alzohairy, A. M., Yousef, M. A., Edris, S., Kerti, B., Gyulai, G., & Bahieldin, A. (2012). Detection of LTR retrotransposons reactivation induced by *in vitro* environmental stresses in barley (*Hordeum vulgare*) via

RT-qPCR. *Life Sci. J.* , *9* , 5019-5026.

Bloom, A. J., & Troughton, J. H. (1979). High productivity and photosynthetic flexibility in a CAM plant. *Oecologia* , *38* , 35-43.

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

Bondino, H. G., Valle, E. M., & Ten Have, A. (2012). Evolution and functional diversification of the small heat shock protein/α-crystallin family in higher plants. Planta 235:1299-1313.

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

Boxall, S. F., Kadu, N., Dever, L. V., Knerova, J., Waller, J. L., Gould, P. J. D., & Hartwell, J. (2020). *Kalanchoë* PPC1 is essential for crassulacean acid metabolism and the regulation of core circadian clock and guard cell signaling genes. *The Plant Cell* ,*32* (4), 1136-1160.

Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W. C., Liu, K. W., Chen, L. J., He, Y., Xu, Q., Bian, C., et al. (2015). The genome sequence of the orchid *Phalaenopsis equestris* . *Nat Genet.* ,*47* , 65-72.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* , *10* , 21.

Chanderbali, A. S., Jin, L. L., Xu, Q. J., Zhang, Y., Zhang, J. B., Jian, S. G., Carroll, E., Sankoff, D., Albert, V. A., Howarth, D. G., et al. (2022). Buxus and Tetracentron genomes help resolve eudicot genome history. *Nature communications* , *13* (1), 1-10.

Chen, L. Y., Xin, Y., Wai, C. M., Liu, J., & Ming, R. (2020). The role of *cis* -elements in the evolution of crassulacean acid metabolism photosynthesis. *Hortic. Re* s., *7* , 5.

Chen, S. S., & Black, C. C. (1983). Diurnal changes in volume and specific tissue weight of crassulacean acid metabolism plants.*Plant Physiol.* , *71* , 373-378.

Chu, Z. F., Wen, J., Yang, Y. P., Nie, Z. L., & Meng, Y. (2018). Genome size variation and evolution in the grape family Vitaceae. *J. Syst. Evol.* , *56* , 273-282.

DeSanto, A. V., & Bartoli, G. (1996). Crassulacean acid metabolism in leaves and stems of *Cissus quadrangularis* . *Ecol. Stu. An.* ,*114* , 216-229.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Bull.* , *19* , 11-15.

Drake, P. L., Froend, R. H., & Franks, P. J. (2013). Smaller, faster stomata: scaling of stomatal size, rate of response, and stomatal conductance. *J. Exp. Bot.* , *64* , 495-505.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis* .*Trends Plant Sci* ., *15* , 573-581.

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., et al. (2017). *De novo* assembly of the *Aedes aegypti*genome using Hi-C yields chromosome-length scaffolds. *Science* ,*356* , 92-95.

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016a). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom.*Cell Systems* , *3* , 99-101.

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016b). Juicer provides a One-Click system for analyzing loop-resolution Hi-C experiments. *Cell Systems* ,

3 , 95-98.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* , *32* , 1792-1797.

Eggli, U., & Nyffeler, R. (2009). Living under temporarily arid conditions - succulence as an adaptive strategy. *Bradleya* ,*27* , 13-36.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., et al. (2019). The Pfam protein families database in 2019.*Nucleic Acids Res.* , *47* , D427-D432.

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu. *Rev. Plant Biol.* , *60* , 433-453.

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

Gao, J., & Lan, T. (2016). Functional characterization of the late embryogenesis abundant (LEA) protein gene family from *Pinus tabuliformis* (Pinaceae) in *Escherichia coli* . *Sci. Rep.* ,*6* , 19467.

Gichuki, D. K., Ma, L., Zhu, Z., Du, C., Li, Q., Hu, G., Zhong, Z., Li, H., Wang, Q., & Xin, H. (2019). Genome size, chromosome number determination, and analysis of the repetitive elements in Cissus quadrangularis. *Peerj* , *7* , e8201.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., et al. (2012). Phytozome: a comparative platform for green plant genomics.*Nucleic Acids Res.* , *40* , D1178-1186.

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* , *27* , 1017-1018.

Griffiths, H., & Males, J. (2017). Succulent plants. *Curr. Biol.* , *27* , R890-R896.

Guo, L. M., Li, J., He, J., Liu, H., & Zhang, H. M. (2020). A class I cytosolic HSP20 of rice enhances heat and salt tolerance in different organisms. *Sci. Rep-Uk.* , *10* , 1383.

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology* , *9* , R7.

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Jr., Hannick, L. I., Maiti, R., Ronning, C. M., Rusch, D. B., Town, C. D., et al. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* ,*31* , 5654-5666.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.*Nature Protocols* , *8* , 1494-1512.

Han, M. V., Thomas, G. W., Lugo-Martinez, J., & Hahn, M. W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* ,*30* , 1987-1997.

Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K., & Shiu, S. H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli.*Plant Physiology* , *148* , 993-1003.

Huang, H., & Nusinow, D. A. (2016). Into the evening: Complex interactions in the *Arabidopsis* circadian clock. *Trends in Genetics* , *32* (10), 674-686.

Hubley, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F., & Wheeler, T. J. (2016). The Dfam database of repetitive DNA families. *Nucleic Acids Res.* , *44* , D81-89.

Hundertmark, M., & Hincha, D. K. (2008). LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana* . *BMC Genomics* , *9* , 118.

Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* , *449* , 463-467.

Jiang, C. K., & Rao, G. Y. (2020). Insights into the diversification and evolution of R2R3-MYB transcription factors in plants. *Plant Physiology* , *183* , 637-655.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* , *110* , 462-467.

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., Schwartz, D. C., Tanaka, T., Wu, J. Z., Zhou, S. G., et al. (2013). Improvement of the *Oryza sativa*Nipponbare reference genome using next generation sequence and optical map data. *Rice* , *6* , 4.

Kirov, I., Divashuk, M., Van Laere, K., Soloviev, A., & Khrustaleva, L. (2014). An easy "SteamDrop" method for high quality plant chromosome preparation. *Mol. Cytogenet.* , *7* , 21.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptivek-mer weighting and repeat separation. *Genome Research* , *27* , 722-736.

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* , *5* , 59.

Kubitzki, K., Bayer, C., & Stevens, P. F. (2007). *Flowering plants: Eudicots; Berberidopsidales, Buxales, Crossosomatales, Fabales p.p., Geraniales, Gunnerales, Myrtales p.p., Proteales, Saxifragales, Vitales, Zygophyllales, Clusiaceae Alliance, Passifloraceae Alliance, Dilleniaceae, Huaceae, Picramniaceae, Sabiaceae* (Springer, Berlin, New York), pp. 467-479.

Li, L. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* , *13* , 2178-2189.

Loureiro, J., Rodriguez, E., Dolezel, J., & Santos, C. (2007). Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Ann. Bot.* , *100* , 875-888.

Lukashin, A. V., & Borodovsky, M. (1998). GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* , *26* , 1107-1115.

Marcais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers.*Bioinformatics* , *27* , 764-770.

Marchler-Bauer, A., & Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* , *32* , W327-331.

Mestre, P., & Baulcombe, D. C. (2006). Elicitor-mediated oligomerization of the tobacco N disease resistance protein. *Plant Cell* , *18* , 491-501.

Michael, T. P., & McClung, C. R. (2002). Phase-specific circadian clock regulatory elements in *Arabidopsis* . *Plant Physiol.* ,*130* , 627-638.

Michael, T. P., Mockler, T. C., Breton, G., McEntee, C., Byer, A., Trout, J. D., Hazen, S. P., Shen, R., Priest, H. D., Sullivan, C. M., et al. (2008). Network discovery pipeline elucidates conserved time-of-day-specific *cis* -regulatory modules. *PLoS Genet.* ,*4* , e14.

Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., Lyons, E., Wang, M. L., Chen, J., Biggers, E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* , *47* , 1435-1442.

Mockler, T. C., Michael, T. P., Priest, H. D., Shen, R., Sullivan, C. M., Givan, S. A., McEntee, C., Kay, S. A., & Chory, J. (2007). The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harb. Symp. Quant. Biol.* , *72* , 353-363.

Morse, S. R. (1990). Water-balance in hemizonia-luzulifolia - the role of extracellular polysaccharides. *Plant Cell and Environment* ,*13* , 39-48.

Murat, F., Armero, A., Pont, C., Klopp, C., & Salse, J. (2017). Reconstructing the genome of the most recent common ancestor of flowering plants. *Nature Genetics* , *49* , 490-496.

Mustilli, A. C., Merlot, S., Vavasseur, A., Fenzi, F., & Giraudat, J. (2002). *Arabidopsis* OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. *Plant Cell* , *14* , 3089-3099.

Nelson, E. A., & Sage, R. F. (2008). Functional constraints of CAM leaf anatomy: tight cell packing is associated with increased CAM function across a gradient of CAM expression. *Journal of Experimental Botany* , *59* , 1841-1850.

Newton, D. J., & Chan, J. (1998). South Africa's trade in southern African succulent plants. Traffic east/southern Africa.

Niechayev, N. A., Paula N. P., & John, C. C. (2019). Understanding trait diversity associated with crassulacean acid metabolism (CAM).*Current Opinion in Plant Biology* , *49* , 74-85.

Olivares, E., Urich, R., Montes, G., Coronel, I., & Herrera, A. (1984). Occurrence of crassulacean acid metabolism in *Cissus Trifoliata* L (Vitaceae). *Oecologia* , *61* , 358-362.

Ou, S., & Jiang, N. (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* , *176* , 1410-1422.

Plomion, C., Aury, J. M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., Faye, S., Francillonne, N., Labadie, K., Le Provost, G., et al. (2018). Oak genome reveals facets of long lifespan. *Nature Plants* , *4* , 440-452.

Portwood, J. L., Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., Walsh, J. R., Sen, T. Z., Cho, K. T., Schott, D. A., et al. (2019). MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Research* ,*47* , D1146-D1154.

Pryszcz, L. P., & Gabaldon, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research* ,*44* , e113-e113.

Qin, L. Y., Hu, Y. H., Wang, J. P., Wang, X. L., Zhao, R., Shan, H. Y., Li, K. P., Xu, P., Wu, H. Y., Yan, X. Q., et al. (2021). Insights into angiosperm evolution, floral development and chemical biosynthesis from the *Aristolochia fimbriata* genome. *Nature Plants* ,*7* , 1239-1253.

Rai, A., Singh, R., Shirke, P. A., Tripathi, R. D., Trivedi, P. K., & Chakrabarty, D. (2015). Expression of rice CYP450-like gene (*Os08g01480*) in *Arabidopsis* modulates regulatory network leading to heavy metal and other abiotic stress tolerance. *PLoS One* , *10* , e0138574.

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* , *11* , 1432.

Rizzon, C., Ponger, L., & Gaut, B. S. (2006). Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis*and rice. *PLoS Comput. Biol.* , *2* , 989-1000.

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

Roddy, A. B., Theroux-Rancourt, G., Abbo, T., Benedetti, J. W., Brodersen, C. R., Castro, M., Castro, S., Gilbride, A. B., Jensen, B., & Jiang, G. F. (2020). The scaling of genome size and cell size limits maximum rates of photosynthesis with implications for ecological strategies. *Int. J. Plant Sci.* , *181* , 75-87.

Ruan, J., & Li, H. (2019). Fast and accurate long-read assembly with wtdbg2. *Nature Methods* , *17* , 155-158.

Sayed, O. H. (2001). Crassulacean acid metabolism 1975-2000, a check list. *Photosynthetica* , *39* , 339-352.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T. A., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics.*Science* , *326* , 1112-1115.

Sharma, T., Dreyer, I., Kochian, L., & Pineros, M. A. (2016). The ALMT family of organic acid transporters in plants and their involvement in detoxification and nutrient security. *Frontiers in Plant Science* ,*7* , 1488.

Shepherd, T., & Wynne Griffiths, D. (2006). The effects of stress on plant cuticular waxes. *New Phytologist* , *171* , 469-499.

Silvera, K., Neubig, K. M., Whitten, W. M., Williams, N. H., Winter, K., & Cushman, J. C. (2010). Evolution along the crassulacean acid metabolism continuum. *Funct. Plant Biol.* , *37* , 995-1010.

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* ,*31* , 3210-3212.

Soltis, D. E., Soltis, P. S., Chase, M. W., Mort, M. E., Albach, D. C., Zanis, M., Savolainen, V., Hahn, W. H., Hoot, S. B., Fay, M. F., et al. (2000). Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Bot. J. Linn Soc.* , *133* , 381-461.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.*Bioinformatics* , *22* , 2688-2690.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research* , *34* , W435-W439.

Tang, H., Wang, X., Bowers, J. E., Ming, R., Alam, M., & Paterson, A. H. (2008). Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* , *18* , 1944-1954.

Thiel, T., Michalek, W., Varshney, R. K., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.).*Theor. Appl. Genet.* , *106* , 411-422.

Ting, I. P., Sternberg, L. O., & Deniro, M. J. (1983). Variable photosynthetic metabolism in leaves and stems of *Cissus quadrangularis* L. *Plant Physiology* , *71* , 677-679.

Tobias, P. A., & Guest, D. I. (2014). Tree immunity: growing old without antibodies. Trends in *Plant Science* , *19* , 367-370.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* , *7* , 562-578.

Van de Peer, Y., Maere, S., & Meyer, A. (2009). OPINION The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* , *10* , 725-732.

Wai C. M., Weise S. E., Ozersky P., Mockler T. C., & Michael T. P., VanBuren R. (2019). Time of day and network reprogramming during drought induced CAM photosynthesis in *Sedum album* . *PLoS Genet.* ,*15* (6), e1008209.

Wan, T., Liu, Z. M., Leitch, I. J., Xin, H. P., Maggs-Kolling, G., Gong, Y. B., Li, Z., Marais, E., Liao, Y. Y., Dai, C., et al. (2021). The*Welwitschia* genome reveals a unique biology underpinning extreme longevity in deserts. *Nature Communications* , *12* , 4247.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., & Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* ,*8* , 77-80.

Wang, J., Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* , *9* , e112963.

Wang, Y. P., Tang, H. B., DeBarry, J. D., Tan, X., Li, J. P., Wang, X. Y., Lee, T. H., Jin, H. Z., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* , *40* , e49.

Warren, R. L., Keeling, C. I., Yuen, M. M., Raymond, A., Taylor, G. A., Vandervalk, B. P., Mohamadi, H., Paulino, D., Chiu, R., Jackman, S. D., et al. (2015). Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *The Plant Journal* , *83* , 189-212.

Wen, J., Lu, L. M., Nie, Z. L., Liu, X. Q., Zhang, N., Ickert-Bond, S., Gerrath, J., Manchester, S. R., Boggan, J., & Chen, Z. D. (2018). A new phylogenetic tribal classification of the grape family (Vitaceae).*J. Syst. Evol.* , *56* , 262-272.

Wickell, D., Kuo, L. Y., Yang, H. P., Ashok, A. D., Irisarri, I., Dadras, A., de Vries, S., de Vries, J., Huang, Y. M., Li, Z., et al. (2021). Underwater CAM photosynthesis elucidated by *Isoetes*genome. *Nat Commun.* , *12* , 6348.

Wickham, H. (2011). ggplot2. Wiley interdisciplinary reviews: computational statistics. *3* (2), 180-185.

Xie, T., Zheng, J. F., Liu, S., Peng, C., Zhou, Y. M., Yang, Q. Y., & Zhang, H. Y. (2015). *De novo* plant genome assembly based on chromatin interactions: a case study of *Arabidopsis thaliana* .*Mol. Plant* , *8* , 489-492.

Xiong, Z. Y., Gaeta, R. T., & Pires, J. C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus* . *P. Natl. Acad. Sci. USA.* , *108* , 7908-7913.

Xu, H., Yu, Q., Shi, Y., Hua, X., Tang, H., Yang, L., Ming, R., & Zhang, J. (2018). PGD: pineapple genomics database. *Hortic. Res.* ,*5* , 66.

Yan, L., Wang, X., Liu, H., Tian, Y., Lian, J. M., Yang, R. J., Hao, S. M., Wang, X. J., Yang, S. C., Li, Q. Y.,et al. (2015). The genome of*Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb. *Mol. plant* , *8* (6), 922-34.

Yang, X. H., Hu, R. B., Yin, H. F., Jenkins, J., Shu, S. Q., Tang, H. B., Liu, D. G., Weighill, D. A., Yim, W. C., Ha, J. M., et al. (2017). The *Kalanchoe* genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nat Commun.* ,*8* , 1899.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood.*Molecular Biology and Evolution* , *24* , 1586-1591.

Zeng, L. P., Zhang, N., Zhang, Q. A., Endress, P. K., Huang, J., & Ma, H. (2017). Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytologist* , *214* , 1338-1354.

Zhang, G. Q., Xu, Q., Bian, C., Tsai, W. C., Yeh, C. M., Liu, K. W., Yoshida, K., Zhang, L. S., Chang, S. B., Chen, F., et al. (2016). The*Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution.*Scientific Reports* , *6* , 19029.

Zhang, L. S., Chen, F., Zhang, G. Q., Zhang, Y. Q., Niu, S., Xiong, J. S., Lin, Z. G., Cheng, Z. M., & Liu, Z. J. (2016). Origin and mechanism of crassulacean acid metabolism in orchids as implied by comparative transcriptomics and genomics of the carbon fixation pathway. *The Plant Journal* , *86* (2), 175-185.

Zhang, N., Wen, J., & Zimmer, E. A. (2016). Another look at the phylogenetic position of the grape order Vitales: chloroplast phylogenomics with an expanded sampling of key lineages. *Molecular Phylogenetics and Evolution* , *101* , 216-223.

Zhang, Z., Xiao, J., Wu, J., Zhang, H., Liu, G., Wang, X., & Dai, L. (2012). ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* , *419* , 779-781.

Data Accessibility Statement

All data used and generated in this study have been deposited on National Genomics Data Center (NGDC, https://ngdc.cncb.ac.cn/) with the project numbers PRJCA005006. The final assembled genome and annotation files were also deposited in the GWH with the accession number WGS019100. All data is available from the corresponding author upon reasonable request.

Competing interests

The authors declare no competing interests.

Author contributions

H.P.X. and Q.F.W. initiated the study of *Cissus rotundifolia*genome sequencing project. H.P.X, Y.W., Q.Y.L and T.W. are joint first authors. D.K.G and Z.F.Z confirmed the genome size and chromosome numbers with help of B.L. D.K.G and H.M.Z. isolated DNA. Y.W. and Q.Y.L. carried out the genome assembly, annotation and transcriptome analysis with the help of J.S.Z., Y.W. carried out the phylogenomic analyses with the help of Y.D.Z., B.L., T.W. and Z.D.C., Y.J.H., Y.S.L. and C.X. detected the diel acid fluctuation in the leaves of *C. rotundifolia* . Y.S.L., R.J.L, Z.M.L and Q.Y.L. identified the CAM pathway related genes. Y.J.H and Q.Y.L isolated the total RNAs and performed the expression pattern analysis of CAM pathway related genes with the help of H.S.J., H.P.X., Y.W., Q.Y.L., J.W. and Q.F.W. wrote the initial manuscript. J.W., J.N.W., Z.D.C., Z.C.L., L.M.L., G.W.H., R.J.L., R.A. and R.W.G. contributed the discussion of project at different stages. All authors revised and contributed to the final version of the text.

Supplementary Information is available in the online version of the paper.

Additional information

Information on reprints and permissions is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to H.P.X (xinhaiping@wbgcas.cn) or Q.F.W (*qfwang@wbgcas.cn*).

List of Figures

Figure 1 Morphological features and genome evolution of *C. rotundifolia* . (a) The succulent leaf, leaf abaxial surface, flower (without petals), and fruit of *C. rotundifolia* (top). Correspondingly, the vegetative and reproductive organs of grape (down). (b) Divergency history between *Cissus* and grape within the phylogeny of flowering plants. Age estimates of each node are based on 342 single copy genes from 13 representative plant species. The WGD or WGT were indicated on the corresponding branches. The number of gene family expansion and contraction was indicated along the related branches. (c) Distribution of genomic features of *Cissus* genome. Each track shows the GC content, repetitive sequences distribution, gene density, and gene expression profile in different tissues from outside to inside. (d) Estimation of LTR activity shows a very recent burst event in *Cissus* in less than 90.77 kya and a much more severe invasion

of LTR than grape. (e) Distribution of *Ks* for the whole paranome of *Cissus* and cross-comparison between *Cissus* and grape. Right corner of the image showed the segmental duplications within the chromosomes.

Figure 2 Gross chromosomal rearrangement underlying *C. rotundifolia* modern 12 chromosomes. (a) Macrosynteny patterns between *C. rotundifolia* and *V. vinifera* , the *C. rotundifolia* chromosomes were numbered according to their physical length from long to short. (b) Seven colored codes were used according to the earlier prediction of ancient seven chromosomes of AEK pre-γ, the schematic representation of paralogous regions derived from a grape–cacao–peach comparison (Freeling, 2009). The karyotypes of *C. rotundifolia* and *V. vinifera* were derived from syntenic comparison with AEK pre-γ and were defined by the occurrence of the syntenic regions as linked clusters in AEK pre-γ, independently of intrachromosomal rearrangements. The evolutionary events were predicted according to the more parsimonious model of evolution. (c) The syntenic relationship among *Cissus* , grape, and *Amborella* . Each *A. trichopoda* scaffold region aligns with up to three regions in either *Cissus* or grape, which were highlighted in red. Shades represent matching gene pairs. (d) Statistic of syntenic regions among *A. trichopoda* , *V. vinifera* and *C. rotundifolia* . The subset of the proportion of genes in syntenic blocks to the whole genome was indicated on the histograms.

Figure 3 Evolution history of functional profiles of *C. rotundifolia* genome. (a) Heatmap shows categorized orthogroups that have significantly increased paralogous numbers in *Cissus* compared with other angiosperms analyzed. (b) The satter plot displayed the expanded orthogroups in five succulent plants and 13 other non-succulent plants. Numbers in square brackets associated with circle sizes stand for -log (P -adjust), where P -adjust is the p -value of the binomial test adjusted for multiple testing. 1–18 are terpene synthase, plant self-incompatibility protein S1, hsp20/alpha crystallin family protein, aspartic proteinase nepenthesin-1 precursor, eukaryotic aspartyl protease family protein, leucine rich repeat protein, F-box domain and LRR containing protein, MATE efflux family protein, nuclear transcription factor Y subunit, UDP-glycosyltransferase, serine/threonine-protein kinase receptor precursor, retrotransposon protein, disease resistance RPP13-like protein 1, cysteine-rich receptor-like protein kinase, leucine-rich repeat receptor-like protein kinase family protein, retrotransposon protein, wall-associated receptor kinase, and F-box family protein. (c) GO categories with an overrepresented number of tandemly duplicated genes in expanded orthogroups encompassing different evolutionary periods of *Cissus* (upper) and the functional bias of tandem duplicate genes retention in grape (lower). The number of TD events was indicated on the branches. (d) Percentage of GO categories from expanded lineage-specific TD in succulent plants and other non-succulent plants. Cellular component and resistance categories in two subgroups were tested by a two-sample t-test.

Figure 4 The CAM pathway in *C. rotundifolia* . (a) The diurnal variation of titratable acidity in *C. rotundifolia* leaves. (b) Expression patterns and *cis* -regulatory elements of CAM-related genes across the diurnal variation. The expression level of each gene was shown in the log10-transformed method. The numbers of five circadian clock-related motifs, including G-box element, evening element (EE), morning element (MOM), CIRCADIAN CLOCK ASSOCIATED 1 (ACC1) binding site, and TCP15 were shown in the 2-kb promoter region of each gene. (c) The overview of the CAM pathway. The carboxylation process (dark period) was shown in the left part, and the decarboxylation process (light period) was shown in the right part. The enzymes were marked in blue and green, respectively, with corresponding expression profiles. A network was constructed for Cluster 4 (d) and Cluster5 (e) using ARACNE. The top 1 % of each network was highlighted by yellow circle, and blue nodes with greater than 10 edges were CAM related genes in *C. rotundifolia* . The yellow and blue nodes were annotated in Table S29.

List of Supplementary Tables

Table S1 Data sources used in this study.

Table S2 Statistic of genomic region linked to AEK post-γ.

Table S3 Synenic genes number and rate with each *Amborella trichopoda* genes.

Table S4 The genes expression with typical circadian patterns in *Cissus rotundifolia* and corresponding orthologies in *Arabidopsis* and *Ananas comosus* during diel cycle and the correlation between them (Mockler

et al., 2007).

Table S5 Information of raw data used in *C. rotundifolia* genome.

Table S6 The detail information of each step of the assembly.

Table S7 The length of *C. rotundifolia* chromosomes.

Table S8 The number of predicted coding genes annotated in public databases.

Table S9 BUSCO analysis of the assembled *C. rotundifolia* genome.

Table S10 The repetitive sequences in *C. rotundifolia* and *V. vinifera* genome.

Table S11 The number of SSRs in *C. rotundifolia* and *V. vinifera* genomes.

Table S12 Segmental duplications inside chromosome in *Cissus* and grape.

Table S13 GO enrichment of the specific genes on Chromosome 01 of *C. rotundifolia* .

Table S14 The GO enrichment of expanded gene families in *C. rotundifolia* .

Table S15 The selective amplifications of genes in *Cissus* and grape.

Table S16 The expression pattern of expanded HSP-20 in *C. rotundifolia.*

Table S17 The orthogroup between 18 species. 'O13'and 'S5'respresent significant orthogroup in 13 non-succulent and 5 succulent plants, respectively.

Table S18 Outlier annotation of orthogroup between 18 species (labled in Figure 3b).

Table S19 The significant orthogroup 5 succulent plants.

Table S20 The significant orthogroup in 13 non-succulent plants.

Table S21 The GO enrichment analysis of tandem repeat genes in the genome of *C. rotundifolia* .

Table S22 The gene number, GO and KEGG enrichment analysis of shared and unique tandem repeat genes in *C. rotundifolia* (Cre) and *V. vinifera* (Vvi).

Table S23 Functional bias in tandem duplicate genes retention between *Cissus* and grape.

Table S24 Gene number of expanded lineage-specific TD in succulent plants and other non-succulent plants from different GO categories.

Table S25 The List of putative CAM-related carbon fixation genes in *C. rotundifolia* and *V. vinifera* .

Table S26 The number of CAM gene family in *Cissus rotundifolia* ,*Vitis vinifera* , *Ananas comosus* , *Oryza sativa* ,*Zea mays* , *Kalanchoe fedtschenkoi* and *Phalaenopsis equestris* .

Table S27 Biological processes enriched in co-expression modules listed in Figure S10.

Table S28 Number of genes shared between co-expression modules (Figure S10) and genes clusters (Figure S11). The genes with significant diel expression patterns were used to constracte cluster analysis.

Table S29 The hub genes with the top 1% most connected nodes as well as putative CAM genes with at least 10 directed edges from each gene cluster shown in Table S28 and Figure S11.

Table S30 The occurrence frequency of motif in promoter region of CAM pathway genes.

Table S31 The circadian clock *cis* -elements annotated at promoter regions of orthologs in *C. rotundifolia, A. thaliana* and *V. vinifera* .

Table S32 Comparison of gene density between *C. rotundifolia* and *V. vinifera* .

List of Supplementary Figures

Figure S1 The chromosome number, genome size, heterozygosity and Hi-C contact map of *Cissus rotundifolia* .

Figure S2 Distribution of microRNA, tRNA, rRNA and snoRNA in each chromosome of *C. rotundifolia* .

Figure S3 Phylogenetic tree showing gene family expansion/contraction analysis compared with 13 representatives of the eudicot plants.

Figure S4 The density distribution of *Ks* of *V. vinifera* vs*V. vinifera* , *C. rotundifolia* vs *C. rotundifolia* vs and *V. vinifera* vs *C. rotundifolia* .

Figure S5 The highly conserved regions between the genomes of *C. rotundifolia* and *V. vinifera* cv. Pinot Noir (PN40024).

Figure S6 Scatter plot displayed the expanded orthogroups in *C. rotundifolia* and *V. vinifera* .

Figure S7 The relative expression level of four PEPC genes in *C. rotundifolia* and two orthologs in *Arabidopsis* .

Figure S8 The relative expression level of βCA1 and PPDK in *C. rotundifolia* and two orthologs in *Arabidopsis* .

Figure S9 The module and phase relationships of dial expressed genes in*C. rotundifolia.* Figure S10 The clusters of dial expressed genes using gens by maSigPro package in *C. rotundifolia.*

Figure S11 The dial expression pattern of a subset of genes that be responsible for the stomata open or close in night and day.

Figure 1

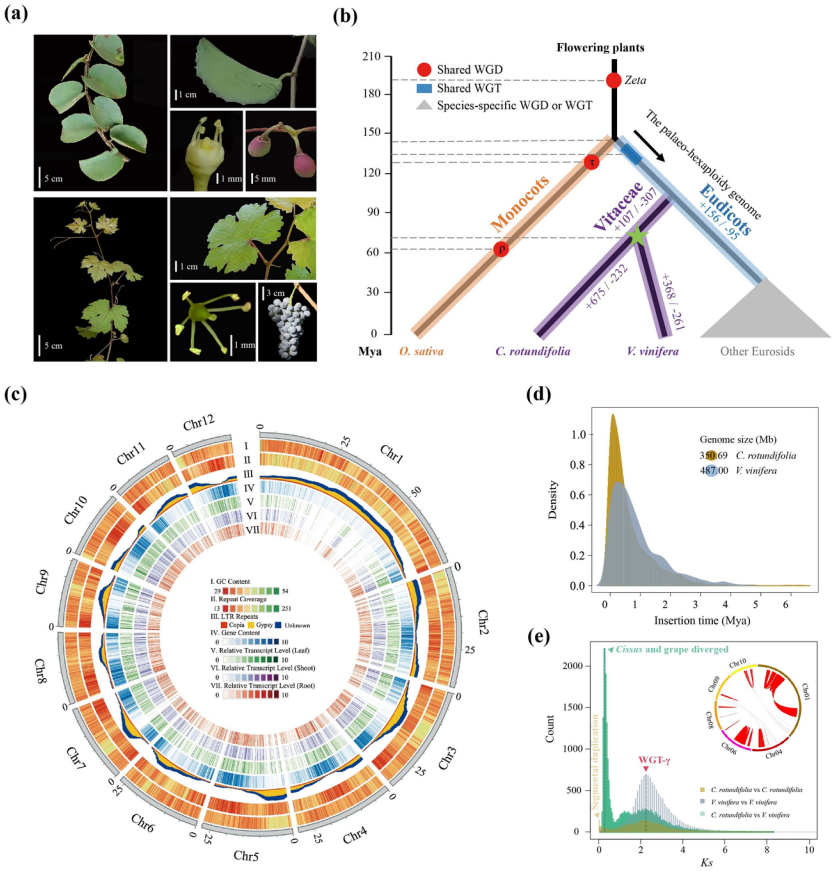


Figure 2

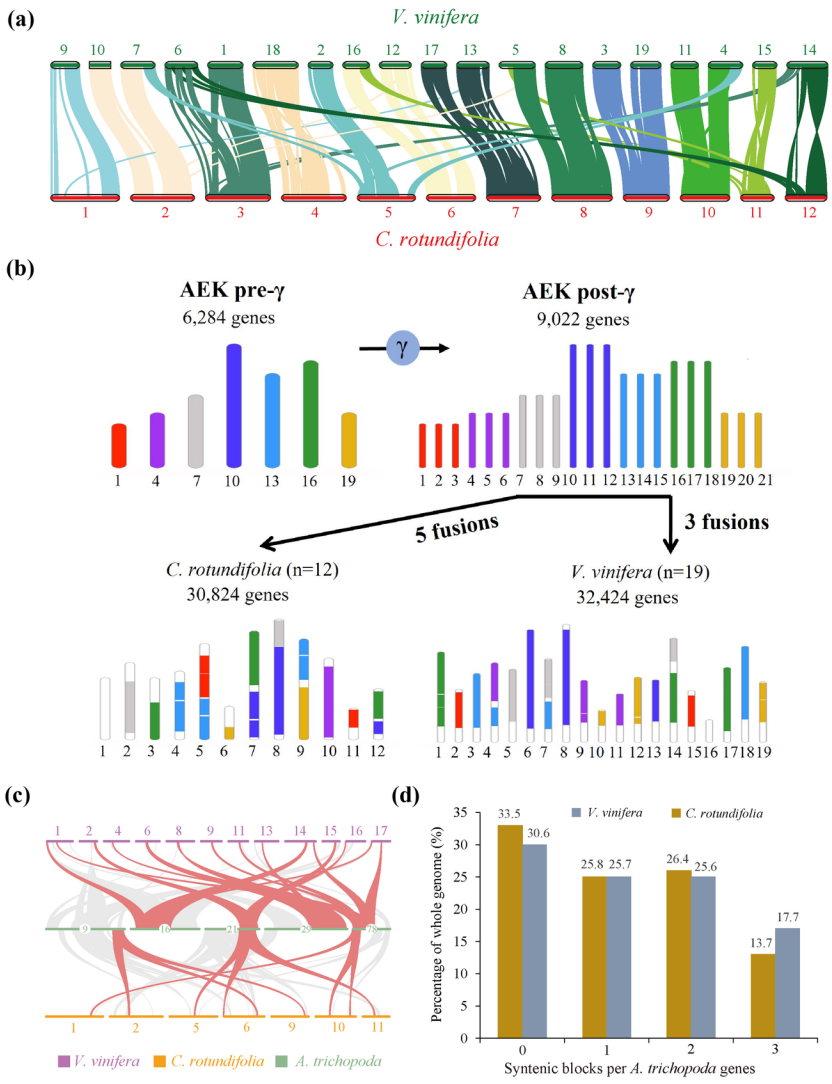


Figure 3

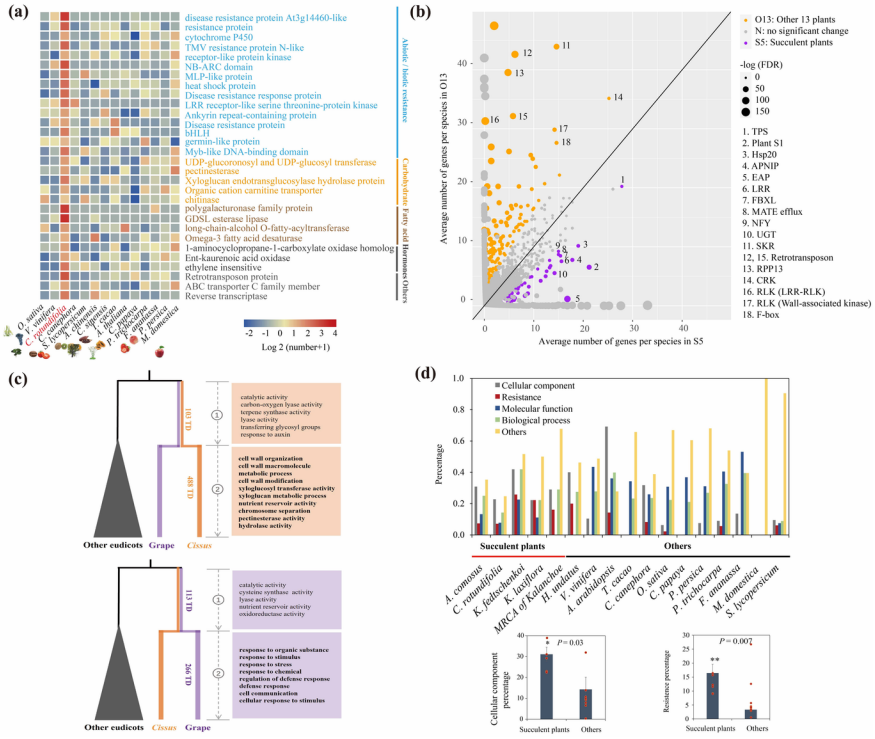
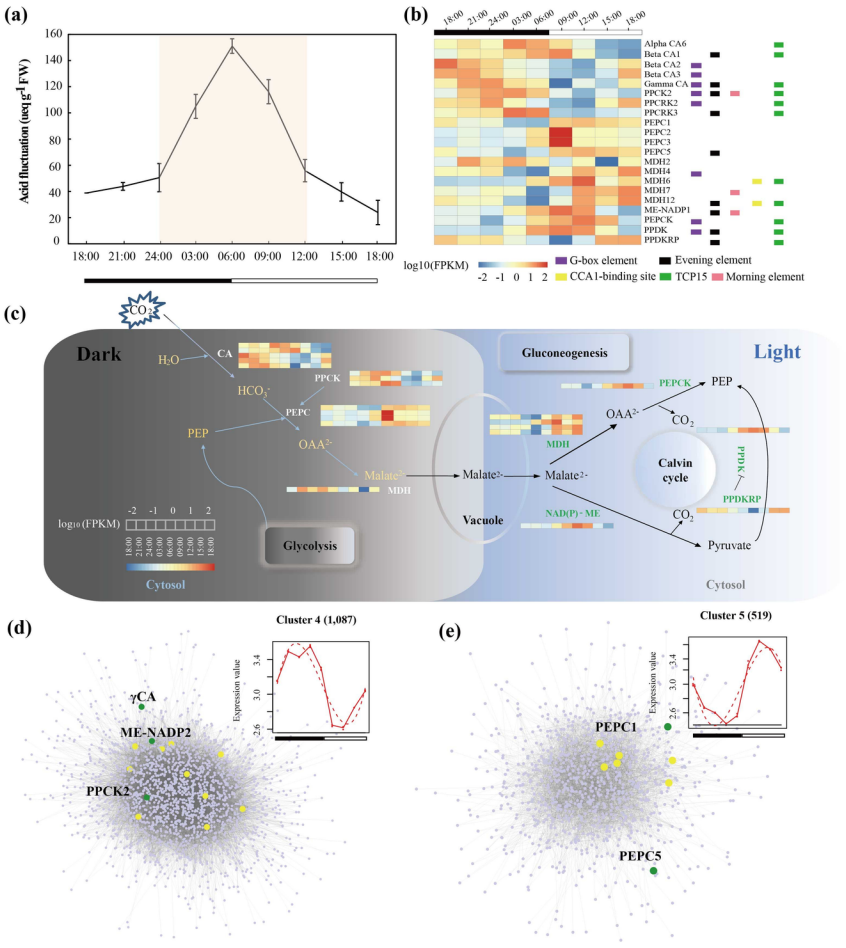


Figure 4



Hosted file

3. Supplementary Information-Cissus genome (final).docx available at https://authorea.com/users/496435/articles/577804-a-genome-for-cissus-illustrates-features-underlying-the-evolutionary-success-in-dry-savannas