the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2023) **113,** 1146–1159 doi: 10.1111/tpj.16081

Marsdenia tenacissima genome reveals calcium adaptation and tenacissoside biosynthesis

Yanli Zhou[1,†,‡], Wei Fan[2,3,†], Haoyue Zhang[1,2,†], Jingling Zhang[1,2], Guanghui Zhang[2,3], Ding Wang[2,3], Guisheng Xiang[2,3], Changhong Zhao[1,2], Lianhua Li[1,2], Simei He[2,3], Yingchun Lu[2,3], Jiuxia Zhao[1], Zhengui Meng[2,3], Xianmin Zhang[2,3], Hengling Meng[4], Xinhua Yin[5,*], Shengchao Yang[2,3,*] and Guangqiang Long[2,3,*]

[1]*Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Science, Kunming, Yunnan, 650201, China,*

[2]*Key Laboratory of Medicinal Plant Biology of Yunnan Province, Yunnan Agricultural University, Kunming, Yunnan, 650201, China,*

[3]*National & Local Joint Engineering Research Center on Germplasm Utilization & Innovation of Chinese Medicinal Materials in Southwestern China, Kunming, Yunnan, 650201, China,*

[4]*The Life Science and Technology College, Honghe University, Mengzi, Yunnan, 661199, China, and*

[5]*Department of Plant Sciences, University of Tennessee, Knoxville, Tennessee, 37996, USA*

Received 22 March 2022; revised 8 December 2022; accepted 21 December 2022; published online 27 December 2022.

*For correspondence (e-mail ynaulong2316@163.com; shengchaoyang@163.com; xyin2@utk.edu).

†These authors contributed equally to this work.

‡Present address: Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Science, Kunming, Kunming, 650201, China

SUMMARY

Marsdenia tenacissima **is a medicinal plant widely distributed in the calcium-rich karst regions of southwest China. However, the lack of a reference genome has hampered the implementation of molecular techniques in its breeding, pharmacology and domestication. We generated the chromosome-level genome assembly in Apocynaceae using combined SMRT sequencing and Hi-C. The genome length was 381.76 Mb, with 98.9% of it found on 11 chromosomes. The genome contained 222.63 Mb of repetitive sequences and 21 899 predicted gene models, with a contig N50 of 6.57 Mb. Phylogenetic analysis revealed that *M. tenacissima* diverged from *Calotropis gigantea* at least 13.43 million years ago. Comparative genomics showed that *M. tenacissima* underwent ancient shared whole-genome duplication. This event, together with tandem duplication, contributed to 70.71% of gene-family expansion. Both pseudogene analysis and selective pressure calculations suggested calcium-related adaptive evolution in the *M. tenacissima* genome. Calcium-induced differentially expressed genes (DEGs) were mainly enriched in cell-wall-related processes. Domains (e.g. Fasciclin and Amb_all) and *cis*-elements (e.g. MYB and MYC) frequently occurred in the coding and promoter regions of cell-wall DEGs, respectively, and the expression levels of these genes correlated significantly with those of calcium-signal-related transcription factors. Moreover, calcium addition increased tenacissoside I, G and H contents. The availability of this high-quality genome provides valuable genomic information for genetic breeding and molecular design, and lends insights into the calcium adaptation of *M. tenacissima* in karst areas.**

Keywords: genome assembly, Hi-C, Apocynaceae, Ca^{2+} adaptation, cell-wall remodeling, *Marsdenia tenacissima*.

INTRODUCTION

Calcium is one of the most abundant elements in karst regions and is an essential nutrient for plant growth (Gao et al., 2019; Tang et al., 2019). Relative to other elements, karst soils are enriched in calcium (Ca^{2+}), up to 37.68 g kg^{-1}. Comparatively, other nutrients like iron and zinc are found at concentrations that are 10 times lower (Hui et al., 2015; Maranhão et al., 2020). The calcium availability of these calcareous soils can reach 77.91%, with a mean availability of 50.9% (Wang et al., 2011). However, genomic data indicating how organisms adapt to such a high Ca^{2+} concentration in karst areas have only been reported for Langurs (i.e. animals; Liu et al., 2019); no such data have been reported for sessile plants. *Marsdenia tenacissima* is a perennial climbing plant widely distributed in the karst areas (Figure 1a) of tropical and

subtropical Asia (Li et al., 2016). According to the new Angiosperm Phylogeny Group (APG) IV classification, this plant belongs to the genus *Marsdenia* of the Apocynaceae family (Madani et al., 2017). Compared with other Apocynaceae species, the Marsdenieae clade seems widely endemic to limestone in karst landscapes (Rodda et al., 2020; Santo et al., 2018), but its adaptive mechanism to high Ca^{2+} levels is unknown. According to our investigation, *M. tenacissima* was capable of accumulating a large amount of calcium (≥3 g kg^{-1} in vined stems), presenting a calcium-adaptation model for the introduction of other plant species into karst areas. Moreover, other excellent traits, such as its drought and salt tolerance, make *M. tenacissima* ideal for the reforestation of harsh environments.

Marsdenia tenacissima has traditionally been used to treat inflammation due to asthma, tracheitis, pharyngitis and urethritis (Zheng et al., 2014), and the extracts of its dried stem have been used to develop a new anticancer medicine 'Xiao Ai Ping', which was demonstrated to be effective in clinical trials (Li et al., 2015). The major bioactive constituents in *M. tenacissima* are polyoxypregnane glycosides, which are derived from C21-steroids conjugated with various sugars (Panda et al., 2006). To date, six types of steroidal skeletons comprising more than 40 polyoxypregnane glycosides, the derivatives of tenacigenin B, have been successfully isolated from *M. tenacissima* (Deng et al., 2005; Li et al., 2014). Pharmacological studies suggest that tenacigenin B ester derivatives actively inhibit the proliferation of cancer cells and enhance the antitumor activity of other anticarcinogens *in vivo* (Xie et al., 2019).

With the ever-increasing demand for 'Xiao Ai Ping', natural sources of *M. tenacissima* are faced with possible overexploitation when supply falls short of demand. Wild *M. tenacissima* is widely cultivated in Honghe, Yunnan province, China. However, the need to obtain germplasms

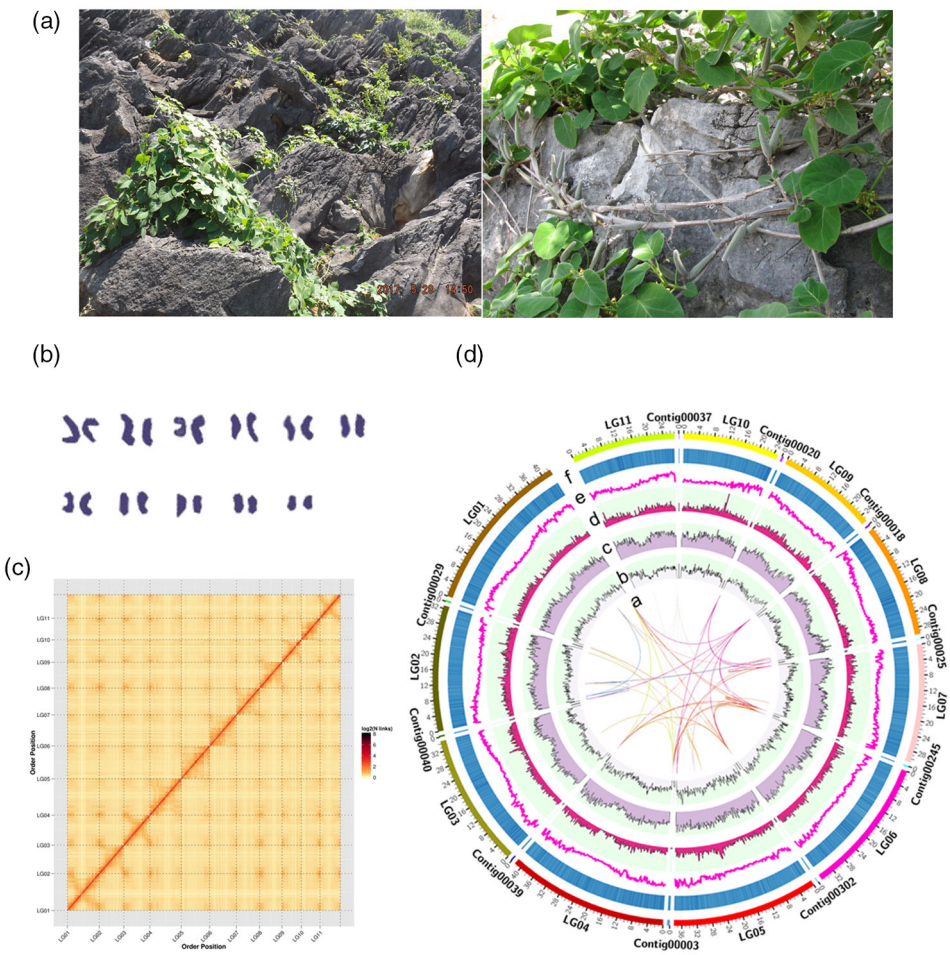

Figure 1. *Marsdenia tenacissima* natural habitat and chromosome-level genome overview. (a) *Marsdenia tenacissima* growing in karst areas.
(b) Karyotype analysis based on the chromosome squashing technique.
(c) Contact heatmap of high-throughput chromosome conformation capture (Hi-C).
(d) Circos exhibition of gene density. a, synteny of gene pairs in *M. tenacissima*; b, large retrotransposon derivative (LARD) distribution on chromosomes; c, SINE distribution; d, LINE distribution; e, gene density of chromosomes; f, GC content.

with a high quality and yield prompted us to investigate the genetics of this plant from multiple perspectives. As most studies have investigated the use of its end-products, little is known about the molecular breeding of *M. tenacissima* owing to the lack of genetic information. This knowledge-gap prompted us to investigate its genomic traits that developed during long-term adaptation in karst areas.

Although the rapid development of long-read sequencing has allowed the exploration of the genetic code associated with the domestication and evolution of many crops (Zeng et al., 2019), very few studies have focused on the Apocynaceae family. Moreover, no reference genomes are available for *M. tenacissima* and its related species in the Asclepiadoideae subfamily.

The aim of this study was to provide genetic data and theoretical support for understanding tenacissoside biosynthesis and calcium adaptation in fragile karst areas. We used single-molecule real-time (SMRT) sequencing and high-throughput chromosome conformation capture (Hi-C) to reveal the full genome of *M. tenacissima*. Furthermore, we investigated the transcriptional regulation of Ca^{2+} adaptation to karst soils in this plant using the high-quality chromosome-scale genome sequence that we developed.

RESULTS

Sequencing and assembly of genome

The genome of *M. tenacissima* was predicted to be 355.89 Mb by a 17 *k*-mer survey, with an estimated heterozygosity rate of 0.50% and repetitive sequence rate of 55.94% (Figure S1). Clean data (100.60 Gb, ~263.51 ×) were obtained using the long-read SMRT platform, with a mean read length of 29.64 kb and maximal read length of 264.29 kb. After polishing, a total assembly size of 381.76 Mb was found, yielding a draft genome with an N50 contig value of 6.57 Mb and an LTR Assembly Index score of 10.59. To further assess the completeness of the assembled genome, we first mapped gene models to 1440 Benchmarking Universal Single-Copy Orthologs (BUSCO) of Embryophyta, allowing the annotation of 91.11% of the genes. The whole genome contained 96.07% of the 458 core eukaryotic genes (CEGs) according to Core Eukaryotic Genes Mapping Approach (CEGMA) analysis (Table S1). We then mapped 315 071 505 Illumina HiSeq clean reads back to the genome. We obtained a mapping ratio of ~97.20%, suggesting the high quality and accuracy of the assembled genome.

Using the Hi-C technique, we generated 110 Gb (~288.14 ×) clean reads with Q30 scores larger than 92.66% to prepare a fine genome of *M. tenacissima*. We could uniquely map 218 898 765 read pairs to the draft genome, 85.30% of them were verified to be valid. Based on the chromosome karyotypes (Figure 1b), and fragmentation, clustering, sorting and orientation, we assembled the draft genome into 11 clusters, with 377.56 Mb (~98.9%) successfully anchored to 11 chromosome-level super-scaffolds (pseudochromosomes). The remaining 400 contigs comprising 4.2 Mb remained unanchored. The high quality of the Hi-C assembly was revealed by the compact anti-diagonal pattern of intra-chromosomal interactions on the genetic map, which indicated that the chromosome groups were explicit (Figure 1b). The six largest pseudochromosomes contained 95% of the assembled genome and 96.44% of the confirmed strand direction. Based on the fine genome, we calculated a gene density of 57 per Mb and an average GC content of 31.55% (Figure 1d). In summary, we *de novo* assembled a chromosome-scale reference genome for *M. tenacissima*, with a total length of ~381.76 Mb located on 11 chromosomes.

Gene annotation and repetitive sequences

Through a combination of *de novo* and homology-based approaches, we found that 222.63 Mb of the *M. tenacissima* genome contained repetitive elements. Repetitive sequences comprised 58.3% of the assembled genome, among which transposable elements (TEs) accounted for 92.29%. Class I TEs, class II TEs and simple sequence repeats contributed 47.76%, 6.05% and 1.06% to the genome size, respectively. For class I TE, retrotransposons, gypsy, large retrotransposon derivatives (LARD) and copia were the most abundant superfamilies, representing 21.38%, 12.77% and 10.70% of the genome, respectively (Table S2). After processing in RepeatMasker, the repetitive sequences of the whole genome were masked to allow annotation. A set of 21 899 gene models was predicted in the genome of *M. tenacissima*, whereof 19 740 (90.14%) and 17 857 (81.54%) were evidenced by homologous proteins and RNA-seq data, respectively. A total of 21 015 gene translated proteins (95.96%) could be functionally annotated to five public databases: NCBI non-redundant (NR), Clusters of Orthologous Groups of proteins (COG), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and TrEMBL (Table S3). Further COG classification found that sequences related to post-translational modification, protein turnover, chaperones and signal transduction mechanisms were most prominent among genes with a known function (Figure S2). In addition, we identified non-coding genes, including 107 rRNA, 510 tRNA, 69 microRNA and 1180 pseudogenes. These results further deepened our understanding of the *M. tenacissima* genome sequence.

Gene family and phylogeny

The gene families of the *M. tenacissima* genome, and protein sequences from *Calotropis gigantea*, *Apocynum venetum*, *Gentiana dahurica*, *Cucumis sativus*, *Catharanthus roseus*, *Coffea canephora*, *Arabidopsis thaliana*, *Vitis vinifera* and *Amborella trichopoda* were clustered in

OrthoFinder. Overall, 15 118 orthogroups containing 228 507 genes were clustered and covered 79.91% of all genes; all species were characterized by the predominance of one copy number for each gene (Figure S3). All 10 species shared 7992 orthogroups of these conserved genes, representing an ancestral core gene set (Figure 2a). However, 412 orthogroups containing 5170 genes were species-specific. There were 233 genes belonging to 115 orthogroups specific to *M. tenacissima*. KEGG enrichment analyses of the *Arabidopsis* homologs of these genes demonstrated that they were related to amino/nucleotide sugar metabolism and fatty acid degradation. Enrichment analysis based on their GO background revealed that the genes were enriched in NADH dehydrogenase, oxidoreductase activity and ion binding pathways. However, only 51.33% of *M. tenacissima* species-specific genes had expression signals. A phylogenetic tree with 10 species was constructed by concatenating 1170 aligned single-copy genes into a supermatrix to reveal the systematic evolution of these species. Figure 2(c) shows that *M. tenacissima* is closely related to *C. gigantea*. The current systematic position is in line with the literature that they belong to the same Asclepiadaceae subfamily (Hassan et al., 2015), validating our constructed tree. To estimate the divergence time, calibration points of *C. gigantea* and *A. trichopoda* were retrieved from the TimeTree database and used as normal priors to constrain the ages of all nodes. *Marsdenia tenacissima* and *C. gigantea* were estimated to diverge ~13.43 million years ago (Figure 3a), shortly after divergence from their common ancestor *A. venetum*.

As gene-family expansion or contraction has a profound role in driving adaptive evolution in flowering plants (Shang et al., 2020), we further investigated the gene-family dynamics of *M. tenacissima* by comparing the phylogeny-based cluster size between each species and their ancestor using the CAFÉ program. Compared with those of its common closest ancestor, 492 gene families were significantly expanded, while 331 gene families showed significant contraction in *M. tenacissima* (Figure 2c). Of the former, 2492 genes were annotated using the GO database. These expanded gene families were mainly annotated to secondary metabolism, photosynthesis, phenylpropanoid biosynthesis and oxidative phosphorylation. In contrast, the contracted genes primarily participated in plant hormone signal transduction, tyrosine metabolism, fatty acid degradation and glycerolipid metabolism (Figure 2d). It is worth noting that 72.5% of the expressed, expanded gene family members were responsive to Ca^{2+} treatment, with a fragments per kilobase of transcript per million mapped reads (FPKM) fold-change of at least 1.5 (Figure S4). However, genes related to binding some other ions, such as zinc and iron, were pseudogenized (Figure 2d; Figure S4).

Genome duplication analysis

We identified 3092 syntenic blocks containing 9675 genes within *M. tenacissima*, which represented the whole-genome duplication (WGD) event for this species and accounted for 44.18% of 21 899 annotated gene models. The percentages of proximal duplication, tandem duplication, dispersed duplication and singleton genes were 4.76% (1043), 11.52% (2523), 25.12% (5501) and 14.42% (3157), respectively. Because WGD occupied the largest proportion of gene pairs, the whole *M. tenacissima* genome may have been duplicated.

Age distributions of synonymous substitutions per synonymous site (K_s) for all paralogous genes represent the natural occurrence of background nucleotides related to evolutionary lineage (Hasegawa et al., 1985; Li et al., 2009). We observed that *M. tenacissima* underwent a common WGD at a K_s value of ~1.04, as observed in other species (Figure 3a). The well-maintained collinearity of paralogous gene pairs within the *M. tenacissima* genome (Figure S5) and the widespread synteny between the *M. tenacissima* and *C. gigantea* genomes also provided evidence of a shared WGD occurring in Apocynaceae (Figure 3b,c; Figure S5). We subsequently investigated the role of WGD in generating patterns. Interestingly, the expanded paralogous genes were mainly derived from WGD (37.61%) and tandem duplication (37.43%), while the contracted genes mainly resulted from WGD (49.35%). *Marsdenia tenacissima*-specific gene families were principally associated with WGD and dispersed duplication (32%), implying that idioadaptive evolution may have occurred during its speciation.

To explore the selective pressure on *M. tenacissima*, the coding sequences of 1170 single-copy genes were aligned and submitted for positive selection analysis. We observed 116 genes that likely contained positive selection sites. These genes were functionally enriched in thiamine metabolism, mismatch repair, N-glycan biosynthesis, inositol phosphate metabolism, glycerophospholipid metabolism and other KEGG pathways, suggesting that these cell components may have functionally evolved during adaptation. As phosphorus levels are correlated with calcium levels in karst soils (Gao et al., 2019; Wei et al., 2018), we evaluated the non-synonymous substitutions per non-synonymous site (K_a) and the K_s of phosphorus-related genes. The K_a/K_s ratio indicated that most genes involved in phosphorus homeostasis and acquisition were under a purifying selection that functions as a functional constraint (Figure S6), which possibly contributed to the environmental adaptation of *M. tenacissima*. However, ion binding genes related to zinc and iron were pseudogenes (Figure 2d).

Calcium adaptation of *Marsdenia tenacissima* in karst soil

As *M. tenacissima* displays genomic adaptation to high soil Ca^{2+} levels, we explored the transcriptional regulation

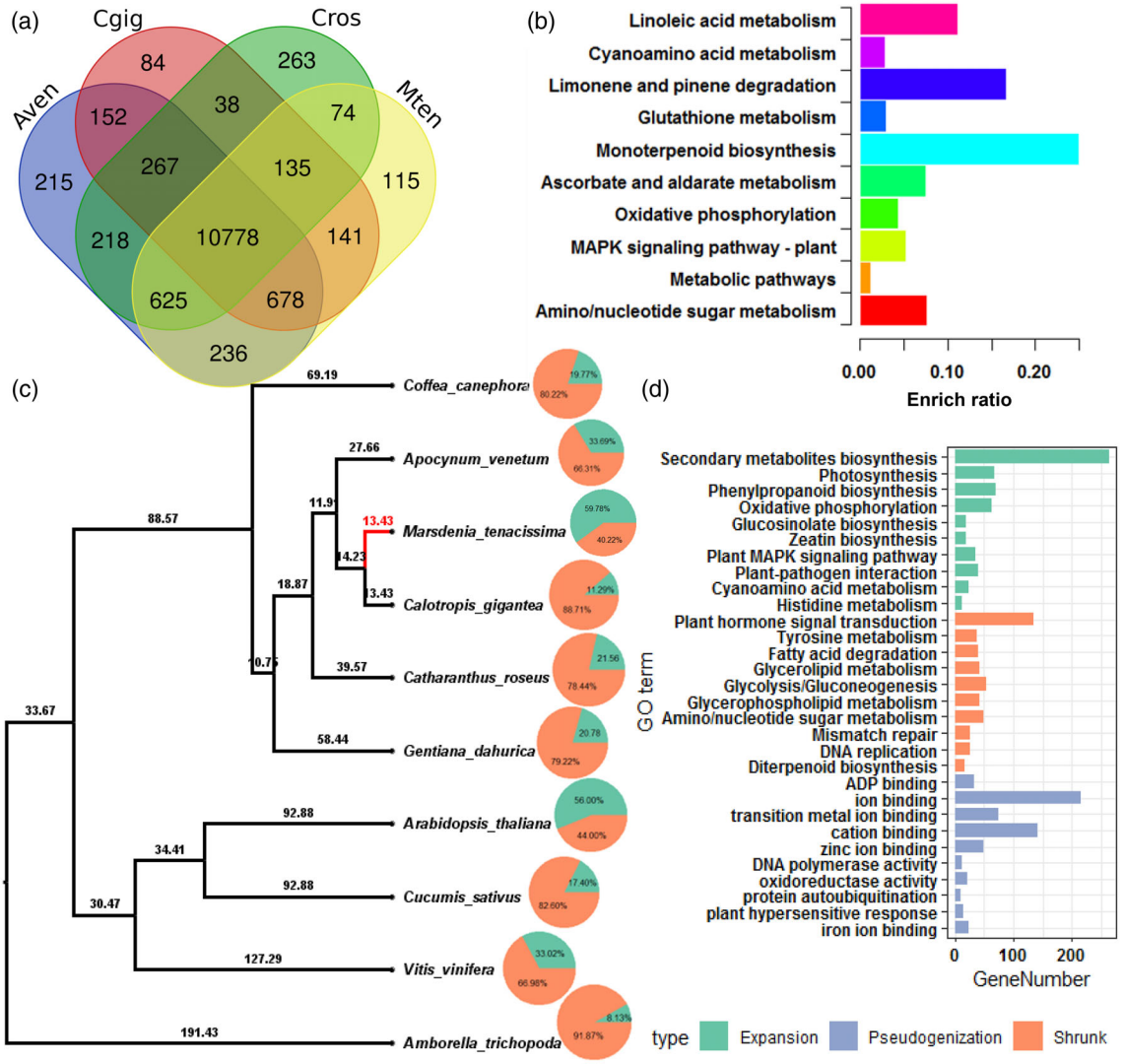

Figure 2. Comparative genomic analysis of *Marsdenia tenacissima*.
(a) Venn diagram of orthogroups in *Coffea arabica* (Cara), *Calotropis procera* (Cpro), *Catharanthus roseus* (Cros) and *M. tenacissima* (Mten).
(b) Gene Ontology (GO) enrichment of *M. tenacissima*-specific orthogroup genes.
(c) Phylogenetic tree of *M. tenacissima* among 10 sequenced plants. Tree was constructed using a supermatrix of 1170 single-copy genes. Orange sector indicates percentage of contracted gene family, while blue sector for expanded gene family.
(d) GO enrichments of expanded, contracted and pseudogenized gene families.

of this process. We investigated the role of calcium on seedling growth by conducting hydroponics culture tests with three different Ca^{2+} concentrations, namely 0 mmol L^{-1} (Ca^{2+}-free), 15 mmol L^{-1} (normal Ca^{2+}) and 50 mmol L^{-1} (Ca^{2+} addition). Our experiment confirmed that calcium is indispensable for normal seedling growth because ~90.9% of the seedlings died after cultivation in Ca^{2+}-free medium for 9 days, but all seedlings grew well in the normal hydroponic medium (Figure 4a).

The expression profiles of *M. tenacissima* seedlings showed 3027 differentially expressed genes (DEGs) in seedlings supplemented with additional Ca^{2+} relative to those grown in normal Ca^{2+} conditions (Figure 4b). The expanded genes, from genome duplication, showed significantly ($P < 0.05$) higher expression levels than those from *M. tenacissima*-specific gene models (Figure S4). The DEGs were mainly enriched for cell-wall organization or biogenesis (GO:0071554) and cell-wall polysaccharide metabolic processes (GO:0010383; Figure S7), and were mainly involved in the 'transporters', 'signaling processes' and 'biosynthesis of phenylpropanoid' and other 'secondary metabolites' pathways (Table S4). In the Ca^{2+} pathway, some Ca^{2+}-receptor genes that encode cyclic nucleotide channels, plasma membrane ATPase, glutamate-like receptor family, signal sensors (e.g. calmodulin-like genes) or responder of calmodulin, calcineurin B-like/calcineurin B-

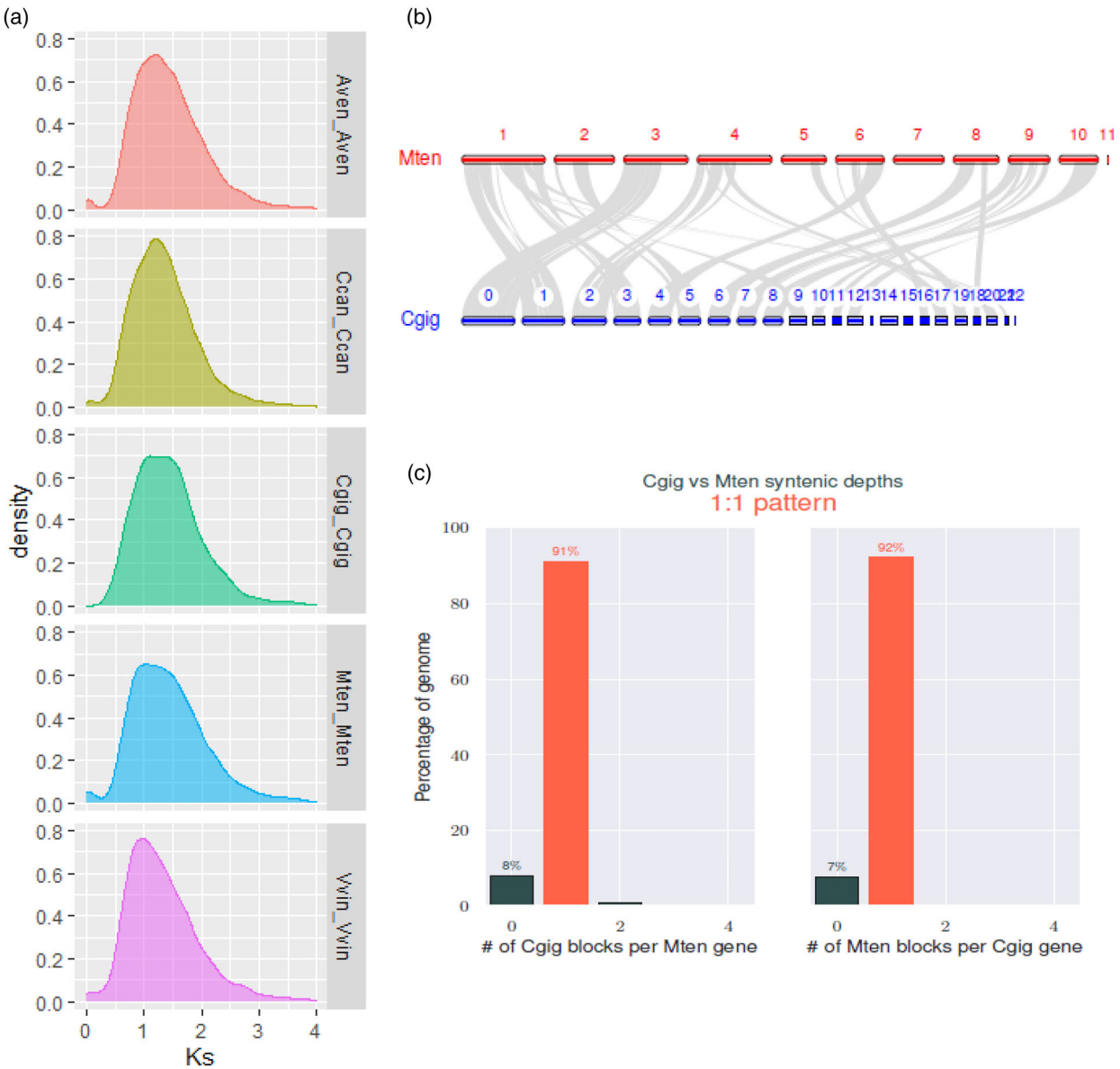

Figure 3. Comparative genomics between *Marsdenia tenacissima* and other species.
(a) Synonymous substitutions per synonymous site (K_s) distribution of closely related species.
(b) Collinearity analysis between *M. tenacissima* and *Calotropis gigantea*.
(c) Genome syntenic depths between *M. tenacissima* and *C. gigantea*. Abbreviations: Aven, *Apocynum_venetum*; Ccan, *Coffea arabica*; Cgig, *C. gigantea*; Mten, *M. tenacissima*; Vvin, *Vitis vinifera*.

like-interacting protein kinases and calcium-dependent protein kinases were upregulated by Ca^{2+} (Figure 5). All these genes were reported to directly or indirectly interact with transcription factors such as Myb-like protein (MYB) and bZIP (Kawamoto et al., 2015; Wang et al., 2017; Yoo et al., 2005).

The prediction of functional domains revealed that pectin-related enzymes such as pectinesterase (PF01095), pectate lyase (PF00544) and pectin methylesterase inhibitor (PMEI; PF04043) were most abundant in the gene sequences of cell-wall-related DEGs (Figure 4c). Interestingly, in the 76 cell-wall-related DEGs, MYB and myc proto-oncogene (MYC) regulatory elements were found in 66 (86.84%) and 65 (85.53%) promoters, respectively. A correlation analysis revealed that 71.74% of pairs showed significant ($R > 0.7$) correlation between 26 transcription factors and 76 pectin-related enzymes (Figure S8).

As we previously found *cis*-elements such as MYB and MYC in the cloned promoter of a polyoxypregnane glycosides pathway gene (Long et al., 2020), we

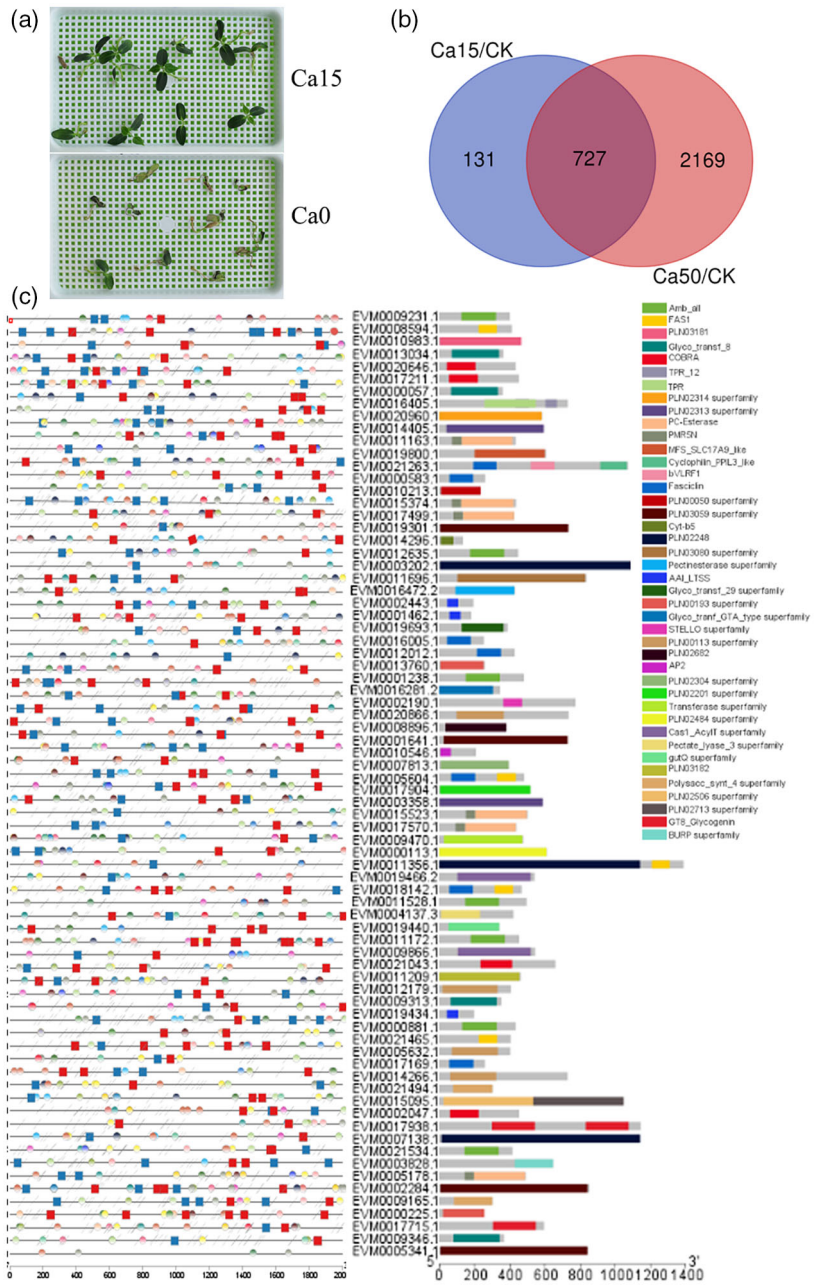

Figure 4. Effect of Ca^{2+} on the plant growth and gene expression of *Marsdenia tenacissima*.
(a) Hydroponic culture of *M. tenacissima* seedlings in normal (left) and Ca^{2+}-free (right) nutrient solution.
(b) Venn diagram of differentially expressed genes (DEGs) under hydroponic culture with 15 mmol L^{-1} Ca^{2+} (Ca15) and 50 mmol L^{-1} Ca^{2+} (Ca50).
(c) Functional domains of cell-wall-related DEGs. Left, promoter *cis*-elements; right, gene functional motifs.

investigated the effects of Ca^{2+} on tenacissoside secondary metabolites. High-performance liquid chromatograph (HPLC) analysis revealed that Ca^{2+} treatment increased the contents of tenacissoside I, G and H (Figure 6), suggesting that these active metabolites result from genomic adaptation to karst calcium levels.

Biosynthetic pathway prediction of tenacissosides in *Marsdenia tenacissima*

To better understand the biosynthesis of tenacissosides, we determined the putative pathways by searching against the whole genome (Figure 7). Based on previous studies (Fujita et al., 2006; Lindemann & Luckner, 1997; Morikawa et al., 2009; Rieck et al., 2019), cytochrome P450 (CYP) was postulated to catalyze cholesterol transformation to pregnenolone, which was then converted stepwise to 3β-hydroxy-5α-pregnan-20-one, the intermediate 5α-8,14-oxidopregnan-3,12β,11α-trol-20-one, and finally to tenacissosides under the action of enzymes such as UDP-glucuronosyltransferase (UGT) and ACT.

DISCUSSION

Only a few Asclepiadoideae genomes have been sequenced, using Illumina short reads (Hoopes

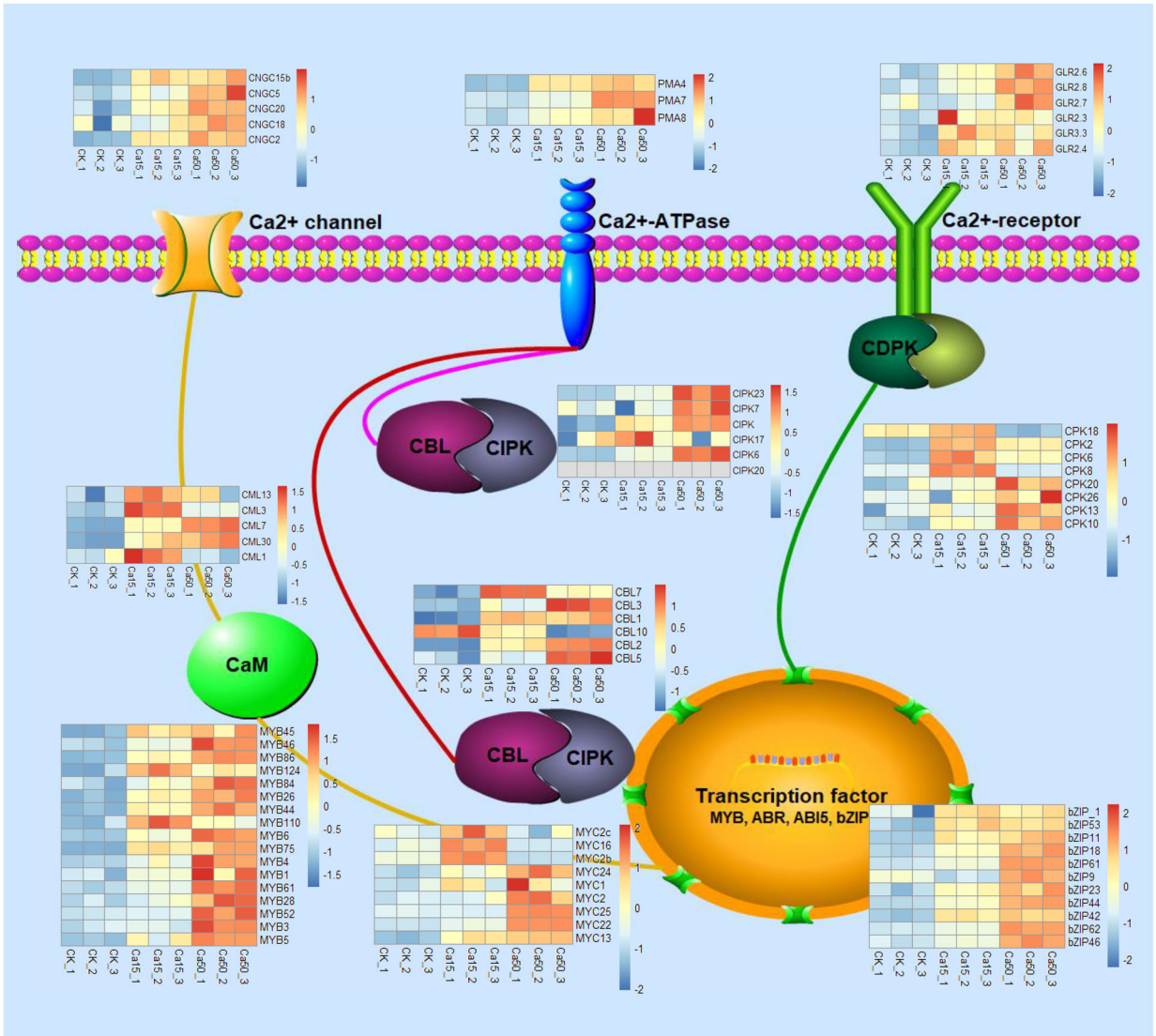

Figure 5. Expression of genes involved in Ca^{2+} sensor and signal-relay pathways in *Marsdenia tenacissima*. CML, calmodulin-like gene; CBL, calcineurin B-like protein; CIPK, CBL-interacting protein; GLR, glutamate-like receptors; CDPK, calcium-dependent protein kinase; MYB, Myb-like protein; MYC, myc proto-oncogene; bZIP, basic leucine zipper.

et al., 2018), but none is available for the *Marsdenia* genus, despite its medicinal potential. *Marsdenia tenacissima*, an Asclepiadoideae species with documented anti-inflammatory, asthma-resistance, furunculosis-treating and anticancer effects, is widespread in karst areas of the tropics and subtropics. This study generated the chromosome-level genome assembly for *M. tenacissima* by combining a long-read sequencing technique, short-read polishing and chromosome conformation capture. The *M. tenacissima* (21899) gene number is much lower than the 55 986 gene models of the MSU *Oryza sativa* genome (Kawahara et al., 2013), despite both species having a similar genome size. This may be ascribed to the scarcity of gene-based research in the Apocynaceae family and the high proportion of repetitive sequences in *M. tenacissima* (58.32% versus 39.50% in *O. sativa*). This high-quality genome of *M. tenacissima* provides a reference for comparative genomics and further functional studies.

With the availability of the reference genome and resequencing data, we had the opportunity to compare *M. tenacissima* to species from other genetic backgrounds. Both K_s peak and collinearity analyses revealed one shared WGD event during *M. tenacissima* speciation, which may have driven the adaptation for its colonization of karst regions. The WGD not only contributed to both gene family expansion and species-specific gene origination but also to gene family contraction, which may have been reshaped under intensive selective pressure. We traced the

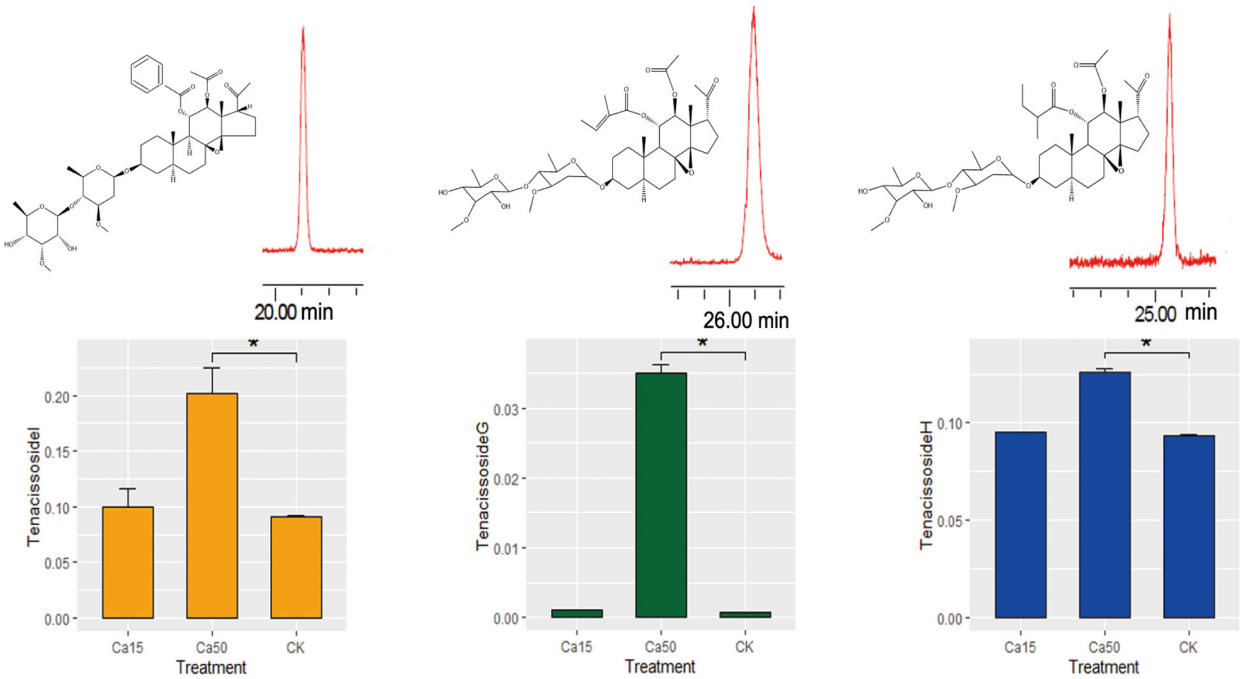

Figure 6. Effect of Ca^{2+} on the tenacissoside content of *Marsdenia tenacissima* under hydroponic culture with 15 mmol L^{-1} Ca^{2+} (Ca15) and 50 mmol L^{-1} Ca^{2+} (Ca50).

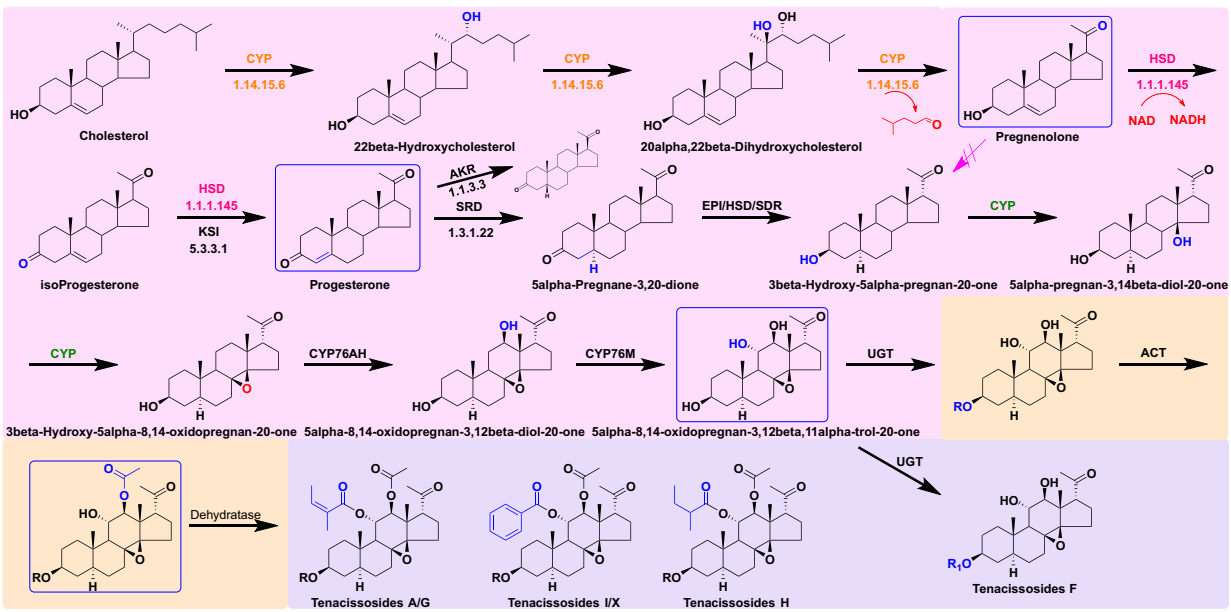

Figure 7. Putative pathway of tenacissosides biosynthesis.

Ca^{2+} response of expanded genes along the phylogeny, which peaked at 75.53% at a common ancestor in the Apocynaceae family (*A. venetum*), suggesting that this event may be shared at least before Asclepiadoideae subfamily differentiation. Santo et al. (2019) also found the karst distribution of several other Marsdenieae species. We also found two inorganic phosphate transporters (EVM0003134.1 and EVM0003701.1) that newly originated out of the nine other species examined. The K_a/K_s ratios between *M. tenacissima* and *G. dahurica* suggest that genes encoding phosphorus-acquisition transporters are under more functional constraints to be utilized through

natural selection (Figure S6), inducing phosphorus precipitation with calcium and its presence in karst-type soils (Wei et al., 2018). For example, calcium-bound phosphorus is the dominant fraction in karst farmland-wetland ecosystem (Gao et al., 2019). In contrast, we also found that cations such as zinc and iron were constrained by related-gene pseudogenization because highly calcareous soils lead to deficiencies in these ions (Hemmaty et al., 2012). This study, therefore, supports the hypothesis that calcium drove the adaptive evolution of *M. tenacissima*. Because new genes usually drive adaptive evolution (Chen et al., 2013), the genes specific to this species may also participate in calcium adaptation. However, in the investigation of calcium-related DEGs, the expanded genes showed much higher expression levels than *M. tenacissima*-specific gene models (Figure S4). This is consistent with the low-expression trait of new genes, and indicates that expanded genes outweigh species-specific genes in terms of adaptation.

The high Ca^{2+} content in karst soils prompts plants to absorb this abundant element. In plant cells, cytoplasmic Ca^{2+} is rigorously regulated by passive Ca^{2+}-channel fluxes and Ca^{2+}-related transporters across the cell wall, plasma membrane and endomembranes (Lecourieux et al., 2006). It is, therefore, understandable that genes of the cell wall are largely modulated by Ca^{2+}, as the cell wall is the first barrier against external factors (Loix et al., 2017; Miedes et al., 2014). Excess calcium in cells is detrimental to plant survival (Case et al., 2007; White & Broadley, 2003). Therefore, plants often take preventative actions against excess calcium. On one hand, they can sequester excess ions in organelles by compartmentalization (Orrenius et al., 2003; Petersen, 2002). On the other hand, plants can alleviate ion accumulation by controlled absorption or ion excretion (Gilliham et al., 2011; Jin et al., 2018; Li et al., 2007). However, the inhibition of absorption or efflux by the plant itself is extremely limited (Jin et al., 2018). In this study, *M. tenacissima* seemed to adopt another strategy to overcome this difficulty, which was to utilize Ca^{2+} locally for cell-wall biogenesis. This could explain the high Ca^{2+} concentration that we observed in the cellulase lysed cell-wall fraction (Table 1).

Searching the functional domains of these Ca^{2+}-induced cell-wall proteins revealed that the fasciclin and Amb_all (pectate lyase) domains occurred most frequently,

Table 1 Ca^{2+} concentration in supernatant and residual of cellulase-treated cell wall (mg L^{-1})

Component	CK	Ca15	Ca50
Supernatant	1.96 ± 0.02a	2.42 ± 0.02b	3.48 ± 0.02c
Residual	11.09 ± 0.04a	12.84 ± 0.04b	13.62 ± 0.06c

Different letters in the same row indicate significant difference between Ca^{2+} treatment level.

consistent with the fact that pectin and cellulose constitute the main components of cell walls (Broxterman & Schols, 2018; Jamet et al., 2008). Fasciclin-like arabinogalactan proteins are usually involved in fiber initiation and development (Huang et al., 2013). Pectate lyase catalyzes cell elongation and cell-wall degradation via depolymerization (Leng et al., 2017; Sun et al., 2020). Ca^{2+}-linked alkali residues could be involved in the scaffolding of pectin with other polysaccharides, such as cellulose, to remodel cell-wall integrity (Broxterman & Schols, 2018; Feng et al., 2018). Among the candidate genes, a pectate lyase (EVM0004001), SHN-type cutin and suberin gene (EVM0019724), and COBRA-like cellulose microfibril and hemicellulose interaction protein (EVM0021043) showed the highest expression levels in roots, which may have directly contributed to the high-calcium soil adaptation.

In summary, we successfully assembled the genome of *M. tenacissima* and reported its chromosome-level sequencing, assembly and annotation. Genome analyses revealed the inclusion of Ca^{2+}-pathway gene families for environment adaptation and putative tenacissoside biosynthesis. The transcriptional profile of seedlings responding to excess Ca^{2+} revealed that cell-wall genes were induced and related to MYB and MYC transcription factors. Consequently, we propose a model that mitigates Ca^{2+} stress in this karst-soil-adapted plant where excess calcium is utilized in cell-wall biogenesis and remodeling. In the near future, we can deepen our understanding on Ca^{2+} adaptation in more karst plants and perform more intensive studies on cell-wall layers.

EXPERIMENTAL PROCEDURES

Plant material

Marsdenia tenacissima seedlings were pre-cultured for 1 month before uniform seedlings were selected and cultured in Hoagland hydroponics solution. The nutrient solution was supplemented with Ca^{2+} at 0, 15 and 50 mmol L^{-1}, for Ca^{2+}-free, normal-Ca^{2+} and Ca^{2+}-addition treatments, respectively, and renewed every 3 days. There were four seedlings in each treatment group and all treatments had three replicates. After culturing in an artificial climate chamber (light/dark cycle of 12/12 h, light intensity of 8000 Lux, humidity of 65–70%) for 15 days, samples were collected for subsequent analysis and stored at −80°C. For tissue-specific RNA sequencing, samples were taken from 3-year-old plants grown at the Mengzi Base for Scientific Research and Demonstration (23°8′8″ N, 103°22′50″ E). For genome sequencing, fresh plant material was collected from tender leaves.

Transcriptome analysis

RNA was extracted with the Omega Plant RNA Kit (R6827, Omega Bio-Tek, Norcross, GA, USA) and validated via gel electrophoresis, micro-spectrophotometry (Nanodrop 2000) and RNA integrity number calculation (Agilent 2100). The mRNAs with polyA tails were enriched with oligo(dT) magnetic beads and fragmented by sonication. Thereafter, first-strand cDNA was synthesized in

reverse transcriptase M-MuLV reaction solution (NEB, USA) using random oligonucleotide primers. Second-strand cDNA was synthesized using dNTP and DNA polymerase I. The purified cDNA was end-repaired and ligated to a polyA adaptor. Finally, cDNAs with lengths of ~200 bp were selected via AMPure XP bead-purification and amplified for library construction. RNA sequencing was performed on an Illumina HiSeq 2000 instrument.

The transcriptome data were deposited in NCBI under accession number PRJNA715100. The raw reads were subjected to quality control and trimming. We mapped several sets of our RNA-seq data back to our *M. tenacissima* genome using HISAT2 (Kim et al., 2019) with default parameters, before quantification with StringTie (Pertea et al., 2015). DEGs were identified using DESeq2 (Love et al., 2014) with a threshold of $\log_2FC > 1$ and $P < 0.01$. We extracted read alignments for cell-wall-related genes and retrieved their FPKM in different tissues and under different treatments.

DNA extraction

Genomic DNA was extracted from fresh leaf material as described by Xu et al. (2021). DNA quality was assessed using a Thermo Fisher NanoDrop 8000 spectrophotometer and Invitrogen Qubit Fluorometer. The 260/280 absorbance ratio was 1.84, and the 260/230 absorbance ratio ranged from 1.84 to 2.35. The concentration of genomic DNA was 546 ng μL^{-1}, and the total yield was 30.24 μg. Subsequently, 100 ng was subjected to 0.3% agarose gel electrophoresis, which showed that the main DNA band was larger than 48.5 kb; this met the quality standard for sequencing library construction. The high-quality DNA was sheared into fragments of approximately 20 kb, which were enriched and purified using Agencourt AMPure XP beads.

Genomic size estimation and preliminary assembly

Short-read sequencing libraries were constructed following the manufacturer's protocol. The genome size of *M. tenacissima* was estimated via the *k*-mer method (Chor et al., 2009). Heterozygosity was determined from the *k*-mer distribution and using GenomeScope (Vurture et al., 2017).

For SMRT sequencing, 2 μg of high-molecular-weight genomic DNA was sheared in a SonicMan microplate sonicator (Brooks Automation) and repaired with the NEB Next FFPE DNA Repair Mix Kit (M6630, NEB, USA) before processing using the Oxford Nanopore Technologies (ONT) Template prep kit (SQK-LSK109, Oxford, UK), following the manufacturer's protocol. The candidate DNA library was purified and quantified on a Qubit Fluorometer (Thermo Fisher). The large segment libraries were premixed with loading beads, transferred into clean R9 flow cells, and sequenced with the ONT sequencing reagent kit (EXP-FLP001.PRO.6, UK) on the ONT PromethION platform with a corresponding R9 cell. Raw data from short- and long-read sequencing were submitted to NCBI under the accession number PRJNA715027.

Hi-C library construction and sequencing

The Hi-C library construction was performed as previously described (Rao et al., 2014). Fresh leaf tissue, as described above, was fixed under vacuum with 36% formaldehyde in fixation buffer for 90 min to crosslink the DNA. After digestion with HindIII, the sticky ends were filled in with biotin-labeled dNTPs and cyclized by ligation. Each pair was de-crosslinked, fragmented into 300–700-bp pieces, and captured by streptavidin-magnetic to construct the corresponding library of interacting DNA. The concentration and insert size of the Hi-C library were determined using Qubit 2.0 Fluorometer and Agilent 2100 Bioanalyzer, respectively. The Hi-C raw data were generated by the Illumina high-throughput PE150 sequencing platform.

Genome polishing and assembly

Raw reads of the single-molecule sequencing data were first quality-assessed by Q20 and Q30, with low-quality reads and adapters removed to obtain clean reads. A subset of longer reads was selected as seed data and submitted to error-correction by the remaining reads in Canu v1.5 (Koren et al., 2017). The initial error-corrected reads were assembled in Canu, and self-correction was conducted twice in Racon (Vaser et al., 2017), followed by three rounds of polishing in Pilon with more than 200 × Illumina data to obtain a draft genome.

Based on the chromosome karyotypes of *M. tenacissima* (Figure 1b), we used the Hi-C data to assemble the genome into 11 pseudochromosomes. Clean Hi-C data were mapped to the draft genome by Burrows–Wheeler Aligner (BWA) after discarding junctions in the raw data (Li & Durbin, 2009). The valid Hi-C interaction pairs were isolated in Hi-C-Pro with the limitation that they interact spatially but not linearly. The draft genome was fragmented into 50-kb segments, with scaffolds clustered, ordered and oriented. Finally, the fine genome was reassembled in LACHESIS (Burton et al., 2013) based on valid interaction pairs. The completeness of the draft genome was assessed by CEGMA prediction, BUSCO analysis and mapping of RNA-seq reads.

Genome annotation

Species-specific repetitive sequences of *M. tenacissima* were *de novo* and structurally retrieved by LTR_FINDER and RepeatScout, classified in PASTEClassifier, and merged with Repbase to construct an integrated background database. Finally, repeat annotation was accomplished using the WUBlast alignment algorithm in RepeatMasker (Tarailo-Graovac & Chen, 2009).

Gene prediction was based on an integrated *ab initio*, homology and RNA-seq evidence strategy. Gene models were initially *de novo* predicted using GENSCAN, AUGUSTUS (v2.4), GlimmerHMM, GeneID v1.4 and SNAP, and the homologs were then aligned to the genome information of *A. thaliana*, *O. sativa*, *C. canephora* and *C. roseus* in GeMoMa. Gene structures for the RNA-seq data were predicted in TransDecoder (v2.0) and GeneMarkS-T (v5.1) after reference assembly using HISAT (v2.0.4) and StringTie (v1.2.3), or predicted using PASA (v2.0.2) after non-reference assembly in Trinity. All related results were merged into a gff file by EVM (v1.1.1) integration. Pseudogene-prediction was carried out via GeneWise (Birney et al., 2004). The functional annotation of genes was conducted by BLAST (v2.2.3) searches against GO, KEGG, KOG, TrEMBL and NR databases, using default parameters except for an e-value of 1×10^{-5}. The Circos tool (http://www.circos.ca; Krzywinski et al., 2009) was used to show gene density, GC content, repeat distribution and gene synteny on individual pseudochromosomes.

Phylogeny construction and gene family comparison

The longest isoforms of proteins in eight species (*M. tenacissima*, *Calotropis procera*, *C. roseus*, *Coffea arabica*, *V. vinifera*, *C. sativus*, *Geum montanum* and *A. trichopoda*) were extracted in Orthofinder v2.2.6 to identify gene families (Emms & Kelly, 2015). Based on the blastp algorithm, all protein sequences were clustered into different orthogroups. Single-copy gene families were selected to

estimate the phylogeny of the eight species by concatenating their single-copy gene sequences into a supermatrix. The concatenated protein sequences of the single-copy genes were multiple-aligned in MAFFT (v7.429) and reverted to codons against the cds file using PAL2NAL (v14). The gene supermatrix was submitted to a model test in IQ-Tree, followed by phylogeny construction with the best model of GTRGAMMAIX in RAxML (v8.2.12). By adopting the MCMCTree module in paml4.9, the divergence time of each node was calibrated to the reference times of known species documented in TimeTree (http://www.timetree.org/; Kumar et al., 2017). Two calibration points were selected from the TimeTree website as priors to restrain the ages of the nodes. The tree was visualized in FigTree.

Based on the obtained phylogeny, we compared the cluster size differences between each species and its nearest ancestor to analyze the expansion and contraction of the gene families using CAFÉ software (v4.2). GO enrichment analyses of expanded or contracted gene families were conducted in KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/index.php; Bu et al., 2021).

Positive selection

Positive selection analysis was conducted on single-copy gene families. A tree of each gene was constructed based on phylogeny. Aligned codons of single-copy gene families were input for calculation in paml4.9 by codeML. *Marsdenia tenacissima* was set as the foreground branch in the gene tree. The positive site for each gene was determined by the branch-site model in the control file. GO functional enrichment of positive selected genes was performed in KOBAS.

Genome duplication and mutation analysis

The syntenic regions of genes were detected by MCScanX based on all-to-all BLASTP searches of the longest protein isoform of each species. The best alignment of each gene was identified by reciprocal best hits and then aligned by MAFFT, reverting to codons for K_s calculation. The peak of paralogous gene pairs (intra-species) represented WGD, while that of orthologous gene pairs (inter-species) denoted species divergence. The K_s values between different gene pairs were estimated by the yn00 method in paml (Yang, 2007). The visualization of K_s distribution was performed using R script.

Detection of calcium content

For detecting the calcium content in the cell wall, 0.1 g of sample was extracted overnight in cold extracting solution containing 3 mL of 250 mmol L^{-1} sucrose, 14 mL of 50 mmol L^{-1} Tris–HCl and 3 mL of 1 mmol L^{-1} disulfide erythritol solution, and then centrifuged at 7000 rpm for 20 min at 4°C to obtain the cell wall. The calcium content in the cell wall was determined by removing the supernatant and performing HNO_3:$HClO_4$ (4:1) digestion of the cell-wall residuals after digesting with cellulase or pectase. The content of calcium in the samples was determined by atomic absorption spectrophotometry.

Detection of tenacissoside content

Tenacissosides I, G and H were extracted from 1 g of ground powder by ultrasonic processing in 50 mL methanol for 45 min, followed by rotary evaporation, filtration, and reconstituted to 10 mL solution. The quantification of these three components in extracts was conducted by calibrating to commercial standards using reversed-phase HPLC with evaporative light-scattering detection. An XBridge C18 column was used to separate the components, and the samples were eluted in a gradient mobile phase containing acetonitrile (40–45%) and water (60–55%).

ACCESSION NUMBERS

The raw data from short- and long-read sequencing were submitted to NCBI under the accession number PRJNA715027. The transcriptome data were deposited in NCBI under the accession number PRJNA715100. The assembled genome and annotation file are available at https://data.cyverse.org/dav-anon/iplant/home/zhouyanli11/analyses/Marsdenia_tenacissima/.

AUTHOR CONTRIBUTIONS

LGQ, YSC, YXH, FW and ZYL designed the study. ZHY, ZJL, ZCH, HSM, LYC, ZJX, MZG and ZXM conducted the experiments. ZYL, ZHY, ZGH, WD, XGS, LLH and MHL performed data analysis. ZYL, FW and ZHY drafted the manuscript with help from LGQ, YXH and YSC. ZYL, FW and ZHY should be considered joint first author.

ACKNOWLEDGEMENTS

The authors thank Yupeng Cun, Wei Xu and Chengli Zhou for the help in data collection and analysis, and thank Zhengkun Liu and Pingli Wang for assistance in field sampling, and also thank Shixian Geng and Rui Xu for help in experiment, and thank Wei Hu, Shuangxia Jin and Guanze Liu for helpful discussions. The authors are grateful to Quanjun Hu for sharing the genome data of *Gentiana dahurica*, and to Fei Gao for providing the genome file of *Apocynum_venetum*. This work was financially supported by grant from National Natural Science Foundation of China (81660635), the Youth Innovation Promotion Association CAS (2021394), Yunnan Fundamental Research Projects (202001 AU070075), Project of Science and Technology in Yunnan (No. 2016RA009), CAS scholarship and the Project of Yunnan Youth Top Talent.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table S1. Genome assembly integrity assessment revealed by CEGMA.

Table S2. Statistical information of repeat sequences.

Table S3. Statistical information of 21 899 predicted gene models searched in five functional databases.

Table S4. KEGG enrichment of Ca^{2+}-induced differentially expressed genes (DEGs).

Figure S1. Kmer distribution at k = 17 depth analysis based on two 2 × 350-bp whole-genome shotgun data of *Marsdenia tenacissima*.

Figure S2. KOG classification of 21 899 annotated gene models.

Figure S3. Copy number variation of gene family in different species.

Figure S4. Expression heatmap of some expanded gene families and pseudogenes.

Figure S5. Circle synteny of *M. tenacissima* chromosomes.

Figure S6. Ka/Ks value of P transporter.

Figure S7. Go enrichment of Ca^{2+}-induced DEGs.

Figure S8. Correlation matrix between Ca^{2+}-signal transcription factors and cell-wall-related DEGs.

REFERENCES

Birney, E., Clamp, M. & Durbin, R. (2004) GeneWise and genomewise. *Genome Research*, **14**, 988–995.

Broxterman, S.E. & Schols, H.A. (2018) Interactions between pectin and cellulose in primary plant cell walls. *Carbohydrate Polymers*, **192**, 263–272.

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z. *et al.* (2021) KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Research*, **49**, W317–W325.

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O. & Shendure, J. (2013) Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology*, **31**, 1119–1125.

Case, R.M., Eisner, D., Gurney, A., Jones, O., Muallem, S. & Verkhratsky, A. (2007) Evolution of calcium homeostasis: from birth of the first cell to an omnipresent signalling system. *Cell Calcium*, **42**, 345–350.

Chen, S., Krinsky, B.H. & Long, M. (2013) New genes as drivers of phenotypic evolution. *Nature Reviews Genetics*, **14**, 645–660.

Chor, B., Horn, D., Goldman, N., Levy, Y. & Massingham, T. (2009) Genomic DNA *k*-mer spectra: models and modalities. *Genome Biology*, **10**, R108.

Deng, J., Liao, Z. & Chen, D. (2005) Marsdenosides A–H, polyoxypregnane glycosides from *Marsdenia tenacissima*. *Phytochemistry*, **66**, 1040–1051.

Emms, D.M. & Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, **16**, 157.

Feng, W., Kita, D., Peaucelle, A., Cartwright, H.N., Doan, V., Duan, Q. *et al.* (2018) The FERONIA receptor kinase maintains cell-wall integrity during salt stress through Ca^{2+} signaling. *Current Biology*, **28**, 666–675.e665.

Fujita, S., Ohnishi, T., Watanabe, B., Yokota, T., Takatsuto, S., Fujioka, S. *et al.* (2006) Arabidopsis CYP90B1 catalyses the early C-22 hydroxylation of C27, C28 and C29 sterols. *The Plant Journal*, **45**, 765–774.

Gao, P., Liu, Y., Wang, Y., Liu, X., Wang, Z. & Ma, L.Q. (2019) Spatial and temporal changes of P and Ca distribution and fractionation in soil and sediment in a karst farmland-wetland system. *Chemosphere*, **220**, 644–650.

Gilliham, M., Dayod, M., Hocking, B.J., Xu, B., Conn, S.J., Kaiser, B.N. *et al.* (2011) Calcium delivery and storage in plant leaves: exploring the link with water flow. *Journal of Experimental Botany*, **62**, 2233–2250.

Hasegawa, M., Kishino, H. & Yano, T.A. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Hassan, L.M., Galal, T.M., Farahat, E.A. & El-Midany, M.M. (2015) The biology of *Calotropis procera* (Aiton) WT. *Trees*, **29**, 311–320.

Hemmaty, S., Dilmaghani, M.R. & Naseri, L. (2012) Effects of sulfur application on soil pH and uptake of phosphorus, iron and zinc in apple trees. *Journal of Plant Physiology and Breeding*, **2**, 1–10.

Hoopes, G.M., Hamilton, J.P., Kim, J., Zhao, D., Wiegert-Rininger, K., Crisovan, E. *et al.* (2018) Genome assembly and annotation of the medicinal plant *Calotropis gigantea*, a producer of anticancer and antimalarial cardenolides. *G3 Genes, Genomes, Genetics*, **8**, 385–391.

Huang, G.-Q., Gong, S.-Y., Xu, W.-L., Li, W., Li, P., Zhang, C.-J. *et al.* (2013) A fasciclin-like arabinogalactan protein, GhFLA1, is involved in fiber initiation and elongation of cotton. *Plant Physiology*, **161**, 1278–1290.

Hui, Y., Jianhong, L., Jiarui, C. & Jianhua, C. (2015) Soil calcium speciation at different geomorphological positions in the Yaji karst experimental site in Guilin, China. *Journal of Resources and Ecology*, **6**(224–229), 226–229.

Jamet, E., Albenne, C., Boudart, G., Irshad, M., Canut, H. & Pont-Lezica, R. (2008) Recent advances in plant cell wall proteomics. *Proteomics*, **8**, 893–908.

Jin, W., Long, Y., Fu, C., Zhang, L., Xiang, J., Wang, B. *et al.* (2018) Ca^{2+} imaging and gene expression profiling of *Lonicera Confusa* in response to calcium-rich environment. *Scientific Reports*, **8**, 7068.

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

Kawamoto, N., Sasabe, M., Endo, M., Machida, Y. & Araki, T. (2015) Calcium-dependent protein kinases responsible for the phosphorylation of a bZIP transcription factor FD crucial for the florigen complex formation. *Scientific Reports*, **5**, 8341.

Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**, 907–915.

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 adaptive *k*-mer weighting and repeat separation. *Genome Research*, **27**, 722–736.

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D. *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645.

Kumar, S., Stecher, G., Suleski, M. & Hedges, S.B. (2017) TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, **34**, 1812–1819.

Lecourieux, D., Ranjeva, R. & Pugin, A. (2006) Calcium in plant defence-signalling pathways. *The New Phytologist*, **171**, 249–269.

Leng, Y., Yang, Y., Ren, D., Huang, L., Dai, L., Wang, Y. *et al.* (2017) A Rice *PECTATE LYASE-LIKE* gene is required for plant growth and leaf senescence. *Plant Physiology*, **174**, 1151–1166.

Li, C., Guo, Q.-S., Yang, S.-C., Zheng, K.-Y., Li, W.-P., Meng, Z.-G. *et al.* (2015) Determination of multiple elements in samples of the medicinal plant *Marsdenia tenacissima* and estimation of geographic origin via pattern recognition techniques. *Journal of Natural Medicines*, **69**, 55–62.

Li, C., Yang, S., Guo, Q., Zheng, K., Wang, P., Xu, X. *et al.* (2014) FTIR fingerprint spectrograms of traditional Chinese medicine Marsdenia tenacissima (in Chinese). *China Journal of Chinese Materia Medica*, **39**, 3311–3315.

Li, C., Yang, S.-C., Guo, Q.-S., Zheng, K.-Y., Wang, P.-L. & Meng, Z.-G. (2016) Geographical traceability of *Marsdenia tenacissima* by Fourier transform infrared spectroscopy and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **152**, 391–396.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li, J., Zhang, Z., Vang, S., Yu, J., Wong, G.K.-S., Wang, J. *et al.* (2009) Correlation between Ka/Ks and Ks is related to substitution model and evolutionary lineage. *Journal of Molecular Evolution*, **68**, 414–423.

Li, Q., Yu, L.-J., Deng, Y., Li, W., Li, M.-T. & Cao, J.-H. (2007) Leaf epidermal characters of *Lonicera japonica* and *Lonicera confuse* and their ecology adaptation. *Journal of Forestry Research*, **18**, 103–108.

Lindemann, P. & Luckner, M. (1997) Biosynthesis of pregnane derivatives in somatic embryos of Digitalis lanata. *Phytochemistry*, **46**, 507–513.

Liu, Z., Zhang, L., Yan, Z., Ren, Z., Han, F., Tan, X. *et al.* (2019) Genomic mechanisms of physiological and morphological adaptations of limestone langurs to karst habitats. *Molecular Biology and Evolution*, **37**, 952–968.

Loix, C., Huybrechts, M., Vangronsveld, J., Gielen, M., Keunen, E. & Cuypers, A. (2017) Reciprocal interactions between cadmium-induced cell wall responses and oxidative stress in plants. *Frontiers in Plant Science*, **8**, 1–19.

Long, G., Zhao, C., Zhao, P., Zhou, C., Ntirenganya, E. & Zhou, Y. (2020) Transcriptomic response to cold of thermophilous medicinal plant *Marsdenia tenacissima*. *Gene*, **742**, 144602.

Love, M.I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.

Madani, I., Ali, L.I. & El Bushra, E.S.E.N. (2017) Evidences from morphological investigations supporting APGIII and APGIV classification of the family Apocynaceae Juss., nom. Cons. *European Academic Research*, **4**, 8259–8266.

Maranhão, D.D.C., Pereira, M.G., Collier, L.S., Anjos, L.H.C.d., Azevedo, A.C. & Cavassani, R.d.S. (2020) Pedogenesis in a karst environment in the Cerrado biome, northern Brazil. *Geoderma*, **365**, 114169.

Miedes, E., Vanholme, R., Boerjan, W. & Molina, A. (2014) The role of the secondary cell wall in plant resistance to pathogens. *Frontiers in Plant Science*, **5**, 1–13.

Morikawa, T., Saga, H., Hashizume, H. & Ohta, D. (2009) CYP710A genes encoding sterol C22-desaturase in Physcomitrella patens as molecular evidence for the evolutionary conservation of a sterol biosynthetic pathway in plants. *Planta*, **229**, 1311–1322.

Orrenius, S., Zhivotovsky, B. & Nicotera, P. (2003) Regulation of cell death: the calcium–apoptosis link. *Nature Reviews Molecular Cell Biology*, **4**, 552–565.

Panda, N., Banerjee, S., Mandal, N.B. & Sahu, N.P. (2006) Pregnane glycosides. *Natural Product Communications*, **1**, 665–695.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T. & Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, **33**, 290–295.

Petersen, O.H. (2002) Calcium signal compartmentalization. *Biological Research*, **35**, 177–182.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.

Rieck, C., Geiger, D., Munkert, J., Messerschmidt, K., Petersen, J., Strasser, J. *et al.* (2019) Biosynthetic approach to combine the first steps of cardenolide formation in Saccharomyces cerevisiae. *Microbiology Open*, **8**, e925.

Rodda, M., Simonsson, N., Ercole, E., Khew, G., Niissalo, M., Rahayu, S. *et al.* (2020) Phylogenetic studies in the *Hoya* group (*Apocynaceae*, *Marsdenieae*): the position of Anatropanthus and Oreosparte. *Willdenowia*, **50** (119–138), 120.

Santo, E., do, F.d.S., Rapini, A., Ribeiro, P.L., Liede-Schumann, S., Goyder, D.J. *et al.* (2019) Phylogeny of the tribe Marsdenieae (Apocynaceae), reinstatement of Ruehssia and the taxonomic treatment of the genus in Brazil. *Kew Bulletin*, **74**, 30.

Santo, F.D.S.D.E., Bitencourt, C., Ribeiro, P.L. & Rapini, A. (2018) Two new species of *Marsdenia* (Apocynaceae) from limestone outcrops in Brazil. *Willdenowia*, **48**(109–116), 108.

Shang, J., Tian, J., Cheng, H., Yan, Q., Li, L., Jamal, A. *et al.* (2020) The chromosome-level wintersweet (*Chimonanthus praecox*) genome provides insights into floral scent biosynthesis and flowering in winter. *Genome Biology*, **21**, 200.

Sun, H., Hao, P., Gu, L., Cheng, S., Wang, H., Wu, A. *et al.* (2020) Pectate lyase-like gene *GhPEL76* regulates organ elongation in *Arabidopsis* and fiber elongation in cotton. *Plant Science*, **293**, 110395.

Tang, J., Tang, X., Qin, Y., He, Q., Yi, Y. & Ji, Z. (2019) Karst rocky desertification progress: soil calcium as a possible driving force. *Science of the Total Environment*, **649**, 1250–1259.

Tarailo-Graovac, M. & Chen, N. (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, **25**, 4.10.11–14.10.14.

Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, **27**, 737–746.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J. *et al.* (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, **33**, 2202–2204.

Wang, C.-Y., Wang, S.-J., Rong, L. & Luo, X.-Q. (2011) Analyzing about characteristics of calcium content and mechanisms of high calcium adaptation of common pteridophyte in Maolan karst area of China (In Chinese). *Chinese Journal of Plant Ecology*, **35**, 1061–1069.

Wang, J., Cheng, G., Wang, C., He, Z., Lan, X., Zhang, S. *et al.* (2017) The bHLH transcription factor CgbHLH001 is a potential interaction partner of CDPK in halophyte *Chenopodium glaucum*. *Scientific Reports*, **7**, 8441.

Wei, X., Deng, X., Xiang, W., Lei, P., Ouyang, S., Wen, H. *et al.* (2018) Calcium content and high calcium adaptation of plants in karst areas of southwestern Hunan, China. *Biogeosciences*, **15**, 2991–3002.

White, P.J. & Broadley, M.R. (2003) Calcium in plants. *Annals of Botany*, **92**, 487–511.

Xie, B., Lu, Y.-Y., Luo, Z.-H., Qu, Z., Zheng, C.-G., Huang, X.-A. *et al.* (2019) Tenacigenin B ester derivatives from *Marsdenia tenacissima* actively inhibited CYP3A4 and enhanced in vivo antitumor activity of paclitaxel. *Journal of Ethnopharmacology*, **235**, 309–319.

Xu, B., Liao, M., Deng, H.-N., Yan, C.-C., Lv, Y.-Y., Gao, Y.-D. *et al.* (2021) Chromosome-level *de novo* genome assembly and whole-genome resequencing of the threatened species *Acanthochlamys bracteata* (Vellociaceae) provide insights into alpine plant divergence in a biodiversity hotspot. *Molecular Ecology Resources*, **22**, 1582–1595.

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

Yoo, J.H., Park, C.Y., Kim, J.C., Do Heo, W., Cheong, M.S., Park, H.C. *et al.* (2005) Direct interaction of a divergent CaM isoform and the transcription factor, MYB2, enhances salt tolerance in *Arabidopsis*. *Journal of Biological Chemistry*, **280**, 3697–3706.

Zeng, L., Tu, X.-L., Dai, H., Han, F.-M., Lu, B.-S., Wang, M.-S. *et al.* (2019) Whole genomes and transcriptomes reveal adaptation and domestication of pistachio. *Genome Biology*, **20**, 79.

Zheng, K., Zhang, G., Jiang, N., Yang, S., Li, C., Meng, Z. *et al.* (2014) Analysis of the transcriptome of *Marsdenia tenacissima* discovers putative polyoxypregnane glycoside biosynthetic genes and genetic markers. *Genomics*, **104**, 186–193.