https://doi.org/10.1093/plphys/kiac558 PLANT PHYSIOLOGY 2023: 191: 844–848

Plant Physiology®

Letter

Sophora moorcroftiana genome analysis suggests association between sucrose metabolism and drought adaptation

Xin Yin [1,†] **Danni Yang** [1,2,†] **Yongming Liu** [1,3,†] **Shihai Yang** [1,4] **Rui Zhang** [3] **Xiaoling Sun** [5] **Hongxu Liu** [5] **Yuanwen Duan** [1] **Yunqiang Yang** [1,*] **and Yongping Yang** [1,*]

1 The Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China
2 University of Chinese Academy of Sciences, Beijing, 100049, China
3 Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing, 100081, China
4 Tibet Yunwang Industrial Corporation, Ltd, Tibet 857007, China
5 Shigatse Bureau of Science and Technology, Science and Technology Road, Sangzhuzi District, Shigatse City, Tibet 857007, China

*Author for correspondenceyangyunqiang@mail.kib.ac.cn (Y.Y), yangyp@mail.kib.ac.cn (Y.Y)
†These authors contributed equally to this work.

Conceptualization and Supervision, Y.P.Y. and Y.Q.Y.; Methodology and Software, D.N.Y., Y.Q.Y., and Y.M.L.; Validation, X.Y., S.H.Y., R.Z., X.L.S., and H.X.L.; Formal Analysis, X.Y., Y.W.D.; Writing–Original Draft, X.Y.; Writing–Review & Editing, X.Y., and Y.Q.Y.; Funding Acquisition, Y.P.Y., Y.Q.Y., and Y.W.D.
The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) are Yunqiang Yang (yangyunqiang@mail.kib.ac.cn) and Yongping Yang (yangyp@mail.kib.ac.cn).

Dear Editor,

Sophora moorcroftiana (Benth.) Baker is a highly drought-resistant and sand-resistant endemic *Sophora* shrub species in the Fabaceae family that can be found along the Yarlung Tsangpo River (elevation: 2,800–4,400 m) of the Tibetan Plateau (Liu et al., 2020). *Sophora moorcroftiana* has an extensive root system that enables its survival in arid regions with little rainfall (Figure 1A; Supplemental Figure 1). Previous work has examined the physiological and biochemical features underlying the drought tolerance of *S. moorcroftiana* (Guo et al., 2014). Additionally, only RNA-Seq data have been collected to assess its drought tolerance, and dehydration-responsive-element binding genes (*DREB*) are the only drought-associated genes that have been cloned from *S. moorcroftiana* (Li et al., 2015, 2017). Although *S. moorcroftiana* holds great potential for ecological restoration in the Tibetan Plateau, efforts to understand the molecular basis of its drought resistance have been hindered by a lack of functional genomics resources.

We herein report a reference genome of *S. moorcroftiana*, which we complemented with metabolomic and transcriptomic profiling to provide insights into the expansion of an α-amylase gene that is likely involved in starch degradation in source tissues and a β-fructofuranosidase gene that cleaves sucrose in sink tissues. Analysis of sucrose metabolism gene duplication suggests that gene duplications may be responsible for the long roots of *S. moorcroftiana*, which may, in turn, be responsible for its drought resistance.

The diploid ($2n = 18$) reference genome of *S. moorcroftiana* was assembled using a combination of Nanopore technology long reads and Illumina short reads (Supplemental Tables 1 and 2). All contigs were anchored and oriented on nine pseudomolecules using Lachesis (Burton et al., 2013), generating 737.35 Mb of pseudomolecule-level sequences, with a contig N50 value of 22.5 Mb (Figure 1B;

Received November 7, 2022. Accepted December 3, 2022. Advance access publication December 6, 2022

Open Access

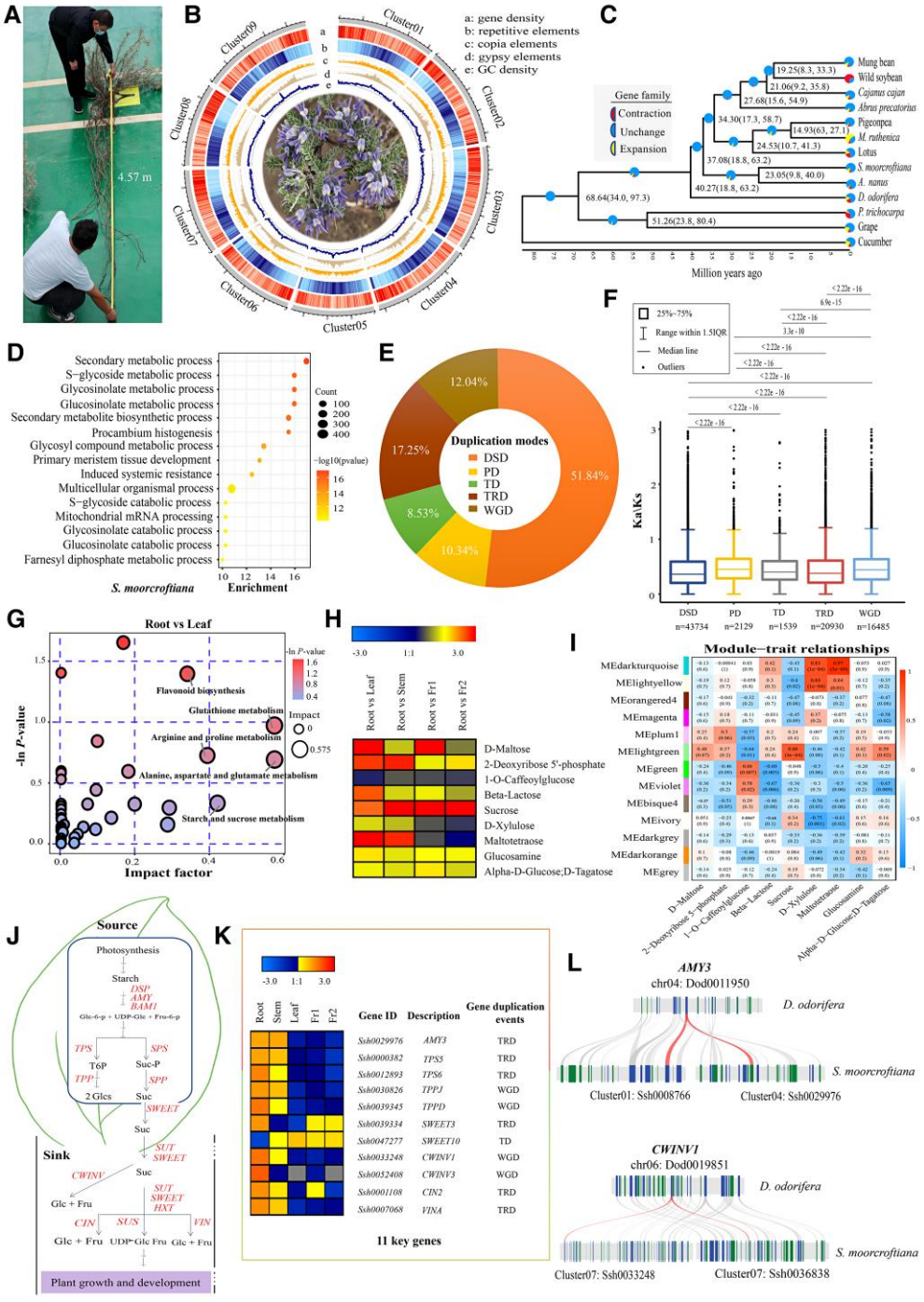

Figure 1 A, Data on root depth, length, spread, collard diameter, and soil penetration were measured ($n = 5$) for *S. moorcroftiana* (Supplemental Figure 1), with the most representative root length shown in Figure 1A. B, The genomic landscape across nine pseudomolecules. C, Phylogenetic analysis of 13 plant species, including 10 Fabaceae species, with cucumber (Cucurbitaceae), grape (Vitaceae), and *Populus trichocarpa* (Salicaceae) as outgroups. The estimated divergence times (million years ago) are indicated at each node, and the numbers in parentheses represent 95% credibility intervals of the estimated dates. The pie diagram on each branch of the tree represents the proportion of gene families undergoing contraction (red) or expansion (yellow) events. Unchanged families are shown in blue. D, GO-based annotation of the biological processes of all expanded gene families of *S. moorcroftiana*. The enrichment factor indicates −Log10 (*P*-value). The 15 most significantly enriched pathways are shown. E, Categories and proportions of different gene duplication events (whole-genome duplication, WGD; dispersed duplication, DSD; transposed duplication, TRD; proximally duplicated, PD; tandemly duplicated, TD) in *S. moorcroftiana*. F, Boxplots showing selective pressures (Ka/Ks) on genes originating from different gene duplication modes. The line in each box represents the median Ka/Ks. The $P < 0.001$ indicates a significant difference between two pairs found by using Student's *t* tests. G, Metabolic pathway analyses of roots compared to leaves are presented

(continued)

Supplemental Figure 2; Supplemental Table 3). The *S. moorcroftiana* genome harbors 496.98 Mb of repetitive sequences, with long terminal repeats (LTRs) accounting for 51.64% of the whole genome. Gypsy and Copia elements were found to be the predominant LTRs (Supplemental Figure 3; Supplemental Table 4). Furthermore, the LTR Assembly Index (LAI) score and the Benchmarking Universal Single-Copy Orthologs (BUSCO) were 18.2% and 98.3%, respectively, and approximately 96.01%–96.64% of the Illumina paired-end reads were mapped to pseudomolecules (Supplemental Tables 5-8). A total of 53,640 protein-coding genes were annotated for putative functions (Supplemental Tables 2 and 9).

To reconstruct high-confidence phylogenetic trees, 169 single-copy ortholog sets derived from 13 species were used for protein sequence alignments (Supplemental Figure 4). Separation between *S. moorcroftiana* and *Dalbergia odorifera* was found to have occurred approximately 40.27 Mya (Figure 1C; Supplemental Figure 5), which is in agreement with previous findings (Hong et al., 2020). The expanded gene families found in medium-/high-altitude plants, including *S. moorcroftiana*, *Ammopiptanthus nanus*, and *D. odorifera*, were significantly enriched in genes associated with "metabolic processes" according to Gene Ontology (GO) annotation. However, expanded gene families in wild soybean (*Glycine soja*) and *Medicago ruthenica* were enriched in "mitochondrial mRNA processing" and "induced systemic resistance", which are related to nitrogen fixation (Stokstad, 2016; Yu et al., 2022) (Figure 1D; Supplemental Figure 6; Supplemental Datasets 1 and 2). These differences suggested that the gene expansions of *S. moorcroftiana* are more likely to enhance metabolite-related functions in addition to nitrogen fixation (Supplemental Figure 6).

Similar to *D. odorifera* (Hong et al., 2020), the *S. moorcroftiana* genome did not undergo an extra whole-genome duplication (WGD) event (Supplemental Figure 7). Next, we classified the gene duplication events in *S. moorcroftiana* (Figure 1E; Supplemental Dataset 3). Of the 25,097 duplicate genes, nearly 51.84% were dispersed duplication (DSD)-, 12.04% were WGD-, and 17.25% were transposed duplication (TRD)-derived gene pairs. Selection strength analysis indicated that proximally duplicated (PD) and WGD gene pairs had higher Ka/Ks ratios and smaller Ks values, implying that these duplications were subjected to rapid sequence divergence and stronger positive selection than genes derived from other duplication modes (Figure 1F; Supplemental Figure 8). We further found that DSD, WGD, and tandemly duplicated (TD) modes all contributed to duplications of genes related to metabolic functions (Supplemental Figure 9; Supplemental Dataset 4). Combined with the GO annotation of expanded gene families from medium and high-altitude legume species, these results imply that duplications played a substantial role in metabolic processes associated with high-altitude adaptation in *S. moorcroftiana*.

To elucidate the metabolic features involved in the adaption of *S. moorcroftiana* to the Tibetan Plateau, we performed UHPLC-MS-based metabolomics of the root, stem, leaf, and fruit (unripe-Fr1 and ripe-Fr2) (Supplemental Figure 10; Supplemental Datasets 5 and 6). Principal component analysis (PCA) showed that the root could be easily separated from other tissues (Supplemental Figure 11). The differentially expressed metabolites in roots compared to those in other tissues were mainly related to "arginine and proline metabolism" and "starch and sucrose metabolism" (Figure 1G; Supplemental Figures 12 and 13; Supplemental Dataset 7), which have been associated with drought stress (Furlan et al., 2020; Chen et al., 2022). A total of nine common differential carbohydrates were detected, with sucrose substantially elevated in roots (Figure 1H).

A high concentration of sucrose, which is the primary photoassimilate transported from source to sink, improves root growth and drought resistance (Ruan, 2014; Chen et al., 2022). In our analysis, a total of 11 different types of duplicated genes were found to be involved in sucrose metabolism from leaves (source tissue) to roots (sink tissue), showing strong correlation coefficient values ($r > 0.8$) with sucrose metabolites (Figure 1, I–K), as analyzed by weighted

Figure 1 (Continued)
as bubble plots. Each bubble in the bubble diagram represents a metabolic pathway. The horizontal coordinate of the bubble and the size of the bubble indicate the influence factor of the pathway in the topology analysis. Larger sizes indicate larger influence factors, and the vertical coordinate of the bubble and the color of the bubble indicate the *P*-value (negative natural logarithm) of the enrichment analysis, with darker colors indicating smaller *P*-values with more significant enrichment. H, Heatmap diagram showing the ratio of the relative content of carbohydrate metabolites in roots compared to other tissues (Root vs. Leaf, Root vs. Stem, Root vs. Fr1, Root vs. Fr2, respectively). I, Module-trait associations. Each row corresponds to a module characteristic gene (eigengene), and each column corresponds to a trait. Each cell contains a corresponding correlation coefficient and Student's asymptotic *P*-value (parentheses). The scale bar on the right indicates the range of correlations from positive (red, 1) to negative (blue, −1). J, Model showing sucrose synthesis and export from mesophyll cells to the apoplasm of the phloem parenchyma for loading, followed by unloading, transport, and metabolism in sink tissues mediated by sugar transporters (see the Supplemental methods for more details). Multiple arrows and solid arrows represent multi-step and direct response relationships, respectively. The genes are highlighted in red with capital abbreviations. Metabolites are capitalized with initials and in regular black font. K, Heatmap diagram showing the expression levels of 11 key genes involved in sucrose metabolism and their duplicated patterns. L, Syntenic blocks between *D. odorifera* and *S. moorcroftiana* containing *AMY3* and *CWINV1*. Both the blue and green boxes represent genes with synteny in this interval between *D. odorifera* and *S. moorcroftiana*. A gray line connects the covariance between them, where *AMY3* and *CWINV1* expand in *S. moorcroftiana*, from one in *D. odorifera* to two in *S. moorcroftiana*, which we have highlighted in red.

gene co-expression network analysis (WGCNA) based on transcriptome (RNA-Seq) and metabolome (Supplemental Table 10 and 11; Supplemental Figure 14-24; Supplemental Dataset 8–11) Notably, *AMY3* (Ssh0029976, α-amylase) and *CWINV1* (Ssh0033248, β-fructofuranosidase), which were derived from TRD and WGD, respectively, underwent gene expansion (Figure 1L; Supplemental Figure 24). Given the role that these genes play in sugar metabolism, these events may be associated with the increased root length and higher drought resistance of *S. moorcroftiana*.

Overall, the reference-quality genome of *S. moorcroftiana* provides insights into the expansion of genes derived from gene duplication events, many of which are involved in sucrose metabolism and could therefore represent candidates for genetic engineering of plateau plants with improved drought tolerance.

Data availability

All raw sequencing data of the genome and RNA-Seq have been deposited in the Genome Sequence Archive in BIG Data Center under accession numbers CRX306560 and CRA006939, respectively.

Supplemental data

The following materials are available in the online version of this article.

Supplemental Figure S1. *Sophora moorcroftiana* morphology.

Supplemental Figure S2. The Hi–C assisted assembly of *S. moorcroftiana* pseudomolecules.

Supplemental Figure S3. Insertion time distribution of LTR retrotransposons.

Supplemental Figure S4. Comparison of the number of gene families in the *S. moorcroftiana* genome and in the genomes of other Fabaceae species and *P. trichocarpa*, grape, and cucumber.

Supplemental Figure S5. Genomic collinearity analysis.

Supplemental Figure S6. Gene ontology (GO)-based annotation analysis under biological processes of expanded gene families of nine Fabaceae species, respectively.

Supplemental Figure S7. Comparison of Ks distributions of inter- and intra-species homologous gene pairs.

Supplemental Figure S8. Analysis of the expression pattern of five types of duplicated genes in different tissues in the stem, leaf, root, and fruit (unripe-Fr1 and ripe-Fr2).

Supplemental Figure S9. GO-based annotation analysis of duplicate genes under biological processes of Fabaceae species, respectively.

Supplemental Figure S10. Classification of detected metabolites.

Supplemental Figure S11. PCA of metabolic data.

Supplemental Figure S12. The number of differential metabolites between leaf, stem, fruit (unripe-Fr1 and ripe-Fr2), and root (Root vs. Leaf, Root vs. Stem, Root vs. Fr1, Root vs. Fr2), respectively.

Supplemental Figure S13. Pathway analysis among the comparison groups.

Supplemental Figure S14. TPM distribution of gene expression values.

Supplemental Figure S15. PCA of RNA-Seq data.

Supplemental Figure S16. Number of differentially expressed genes.

Supplemental Figure S17. Soft threshold (power) versus scale-free topology model fit significance.

Supplemental Figure S18. Dendrogram showing modules identified by the WGCNA and clustering dendrogram of expressed genes.

Supplemental Figure S19. Network heatmap. One thousand and five hundred genes were selected and displayed randomly.

Supplemental Figure S20. Hierarchical cluster tree showing correlation modules identified by WGCNA.

Supplemental Figure S21. A scatterplot of gene significance for sucrose versus module membership in the light green module.

Supplemental Figure S22. Phylogenetic relationship of sucrose metabolism pathway proteins in *S. moorcroftiana*, *D. odorifera*, and cucumber.

Supplemental Figure S23. Phylogeny of AMY3 and CWINV1 shows the expansion during evolution.

Supplemental Figure S24. Heat map of expression profiles (in log2-based TPM) for sucrose metabolism pathway genes of *S. moorcroftiana* in different tissues.

Supplemental Table S1. Nanopore, Illumina, Hi-C, sequencing data for the *S. moorcroftiana*.

Supplemental Table S2. Summary of genome assembly and annotation for the *S. moorcroftiana*.

Supplemental Table S3. Pseudomolecules length data statistics.

Supplemental Table S4. Summary of transposable elements in the *S. moorcroftiana* genomes.

Supplemental Table S5. The Illumina paired-end reads mapped to the *S. moorcroftiana* pseudomolecules.

Supplemental Table S6. The correspondence between contigs from draft assembled genome and pseudomolecules of the *S. moorcroftiana*.

Supplemental Table S7. Genome BUSCO results for the *S. moorcroftiana* final pseudomolecules assemblies.

Supplemental Table S8. Protein BUSCO results for the *S. moorcroftiana* final pseudomolecules assemblies.

Supplemental Table S9. The functional annotations of genes in *S. moorcroftiana*.

Supplemental Table S10. Summary statistics of RNA-Seq.

Supplemental Table S11. The clean reads of RNA-Seq mapped to the *S. moorcroftiana* pseudomolecules.

Supplemental Dataset S1. GO annotation of the expansion gene family in 10 Fabaceae species.

Supplemental Dataset S2. GO-term enrichment data under biological processes of expanded gene families in 10 Fabaceae species, including *P*-values.

Supplemental Dataset S3. Gene list from different gene duplication events (WGD, DSD, TRD, PD, and TD) in *S. moorcroftiana*.

Supplemental Dataset S4. GO-term enrichment data under biological processes of genes from different gene duplication events, including *P*-values.

Supplemental Dataset S5. Metabolomic data.

Supplemental Dataset S6. 1,141 metabolites were identified and can be divided into 16 classes.

Supplemental Dataset S7. Differential metabolites between root and other tissues, respectively.

Supplemental Dataset S8. Expression of differential genes in root versus other tissues, respectively.

Supplemental Dataset S9. The gene lists with expression data for modules.

Supplemental Dataset S10. Genes with module membership versus gene significance within modules were characterized.

Supplemental Dataset S11. Information of sucrose metabolism pathway protein sequences in *S. moorcroftiana*.

Acknowledgments

We thank the National Basic Science Data Center of China, and the numerical calculations in this paper have been done at the Hefei Advanced Computing Center.

Funding

This work was supported by the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (2019QZKK0502), the Regional Science and Technology Collaborative Innovation Project of Shigatse Bureau of Science and Technology (QYXTZX-RKZ2022–01 and QYXTZX-RKZ2021–07), the Strategic Priority Research Program of the Chinese Academy of Sciences, Pan-Third Pole Environment Study for a Green Silk Road (Pan-TPE) (Grant No. XDA2004010306), the National Natural Science Foundation of China (Grant Nos. 32070362 and 32100315), and "Cross-Cooperative Team" of the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences.

Conflict of interest statement. None declared.

References

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J (2013) Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol **31**(12): 1119–1125

Chen Q, Hu T, Li X, Song CP, Zhu JK, Chen L, Zhao Y (2022) Phosphorylation of SWEET sucrose transporters regulates plant root: shoot ratio under drought. Nat Plants **8**(1): 68–77

Furlan AL, Bianucci E, Giordano W, Castro S, Becker DF (2020) Proline metabolic dynamics and implications in drought tolerance of peanut plants. Plant Physiol Biochem **151**: 566–578

Guo Q, Zhang W, Li H (2014) Comparison of photosynthesis and antioxidative protection in *Sophora moorcroftiana* and *Caragana maximovicziana* under water stress. J Arid Land **6**(5): 637–645

Hong Z, Li J, Liu X, Lian J, Zhang N, Yang Z, Niu Y, Cui Z, Xu D (2020) The chromosome-level draft genome of *Dalbergia odorifera*. *GigaScience* **9**(8): giaa084

Li H, Yao W, Fu Y, Li S, Guo Q (2015) De novo assembly and discovery of genes that are involved in drought tolerance in Tibetan *Sophora moorcroftiana*. PLoS One **10**(1): e111054

Li H, Zhang Y, Guo Q, Yao W (2017) Molecular characterisation of a *DREB* gene from *Sophora moorcroftiana*, an endemic species of plateau. Protoplasma **254**(4): 1735–1741

Liu Y, Yi F, Yang G, Wang Y, Pubu C, He R, Xiao Y, Wang J, Lu N, Wang J, et al. (2020) Geographic population genetic structure and diversity of *Sophora moorcroftiana* based on genotyping-by-sequencing (GBS). PeerJ **8**: e9609

Ruan YL (2014) Sucrose metabolism: gateway to diverse carbon use and sugar signaling. Annu Rev Plant Biol **65**(1): 33–67

Stokstad E (2016) The nitrogen fix. Science **353**(6305): 1225–1227

Yu Y, Gui Y, Li Z, Jiang C, Guo J, Niu D (2022) Induced systemic resistance for improving plant immunity by beneficial microbes. Plants-Basel **11**(3): 386