

High-quality genome of guar (*Cyamopsis tetragonoloba*) provides the new insights into the biosynthesis and origination of galactomannans in Fabaceae family

Yuanhao Ding

Hainan University

Xinglong Zhao

Hainan University

Hui Yao

Hainan University

Shicai Huang

Hainan University

Xuyao Wang

Hainan University

Yayu Chen

Hainan University

Ping Li

Hainan University

Chengzhi Xie

Hainan University

Jiaquan Huang

Hainan University

Zhuang Yang

Hainan University

Haiyan Hu

huhaiyan@hainanu.edu.cn

Hainan University

Short Report

Keywords:

Posted Date: March 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9132094/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Plant Cell Reports on May 14th, 2026. See the published version at <https://doi.org/10.1007/s00299-026-03832-7>.

Abstract

Guar (*Cyamopsis tetragonoloba*) is a Fabaceous crop widely cultivated in tropical and subtropical regions. Guar gum, a galactomannan polysaccharide extracted from its seeds, is extensively utilized across food, pharmaceutical, cosmetic, and petroleum industries, underscoring its considerable economic importance. In this study, we present a 516 Mb chromosome-level genome assembly of guar and explore its evolutionary history within the Fabaceae family. Genome synteny and synonymous nucleotide substitution rates (Ks) analyses indicate that guar, along with other Faboideae species, shared one whole-genome duplication (WGD) events, occurring ~58 million years ago (MYA). Transcriptome profiling revealed that key genes of ManS and GMGT1 involved in galactomannan biosynthesis are highly expressed in the endosperm, while other Fabaceae species that do not synthesize galactomannan show significantly lower expression of these genes, suggesting that transcriptional regulation of ManS and GMGT1 play a central role in galactomannan production. This study offers a high-quality genomic resource for guar, clarifies its phylogenetic placement within the Fabaceae, and provides novel molecular insights into the biosynthesis and evolutionary origin of galactomannan.

Full Text

Guar (*Cyamopsis tetragonoloba*, $2n = 2x = 14$) is an annual herbaceous plant belonging to the Faboideae subfamily of Fabaceae (Fig. 1A), believed to originate in either Africa or Asia. As an important source of food and industrial raw materials, guar exhibits strong drought resistance. Mature seeds of guar contain a white to light yellow endosperm, commonly known as guar gum, which consists primarily of galactomannan polymers. However, the molecular mechanisms underlying the biosynthesis, regulation and origination of galactomannans in Fabaceae species remain poorly understood. Although two guar genomes have been published recently (Gaikwad et al., 2023; Li et al., 2024), a high-quality reference genome can deepen our understanding into the formation and evolution of galactomannans across the Fabaceae family.

To obtain a high-quality genome assembly of guar, we employed an integrated approach combining third-generation long-read sequencing with second-generation short-read error correction (Zan et al., 2025). The genome size was initially estimated to be ~ 510 Mb using flow cytometry and ~ 539 Mb via k-mer analysis (Supplemental Fig. 1 and Table S1-S2), and fluorescence in situ hybridization (FISH) confirmed that guar is a diploid containing 14 chromosomes ($2n = 2x = 14$) (Fig. 1B). A total of 152.41G (~ 295.4× coverage) clean PacBio Sequel long-reads were generated for de novo contig assembly, generating a 523 Mb draft genome with contig N50 of 3.98 Mb after polished with 86.67 Gb (~ 168.9×coverage) of Illumina short-read data (Table S3). Subsequently, 119.42 Gb (~ 231.41×coverage) Hi-C data were used to scaffold and anchor 661 contigs onto 7 chromosomes, resulting in a chromosome-level genome assembly of 516 Mb, with individual chromosomes ranging from 42 Mb to 120 Mb in size (Fig. 1C and Table S3-S5).

LTR Assembly Index (LAI) analysis indicated this assembled genome is a reference level genome (LAI = 12.98) (Table S4). Illumina and Hi-C data were realigned to the genome, achieving alignment rates of 99.70% and 99.5%, respectively (Table S6). Additionally, RNA-seq data from 15 samples showed an average alignment rate of 97.98% (Table S6). A total of 47,440 protein-coding genes were predicted, and assembly completeness was supported by the identification of 1,906 (94.1%) of the 2,026 highly conserved core eukaryotic genes in the BUSCO dataset (Table S4). Repetitive sequences accounted for 53.96% of the genome, and the overall GC content was 32.81% (Fig. 1C, Table S4).

Previous studies have demonstrated that Fabaceae plants evolved from a common tetraploid ancestor approximately 58 million years ago (MYA) (Young et al., 2011; Zhao et al., 2021). In this study, genome synteny comparisons between guar, soybean (*Glycine max*), and grape (*Vitis vinifera*) revealed that soybean underwent two whole-genome duplication (WGD) events, whereas guar experienced only one (Fig. 1D). To further investigate WGD events and the evolutionary history of guar and other Fabaceous plants, we performed a corrected synonymous nucleotide substitution rates (*Ks*) analysis to estimate divergence times across six species: *C. tetragonoloba*, *G. max*, *V. vinifera*, *Robinia pseudoacacia*, *Medicago truncatula* and *Lotus japonicus* (Fig. 1E). The results indicate that Fabaceous species share two ancient polyploidization events: an early whole-genome triplication (WGT) shared with *V. vinifera* ~ 130 MYA, a subsequent WGD event specific to the Fabaceae family ~ 58 MYA, which facilitated the divergence of major lineages within the family (Fig. 1E). In addition, *G. max* underwent a more recent, lineage-specific WGD event ~ 13 MYA, consistent with previous reports (Feng et al., 2024; Pan et al., 2023). Intergenomic collinearity analyses between Fabaceous species and *V. vinifera* provided clear evidence for the recent WGD events, whereas signals of the ancient WGT event were largely undetectable through synteny-based methods (Fig. 1F). Besides, among Fabaceae family species, the ancestor of guar first diverged from *R. pseudoacacia*, and then rapidly branched from *L. japonicus*, *G. max* and *M. truncatula* (Fig. 1E).

To further unveil the phylogenetic position of guar within the Fabaceae family, a species tree was constructed using two species from Caesalpinioideae, 11 species from Faboideae, and *V. vinifera* as the outgroup (Fig. 1G). The resulting phylogenetic topology is consistent with previously reported relationships (Feng et al., 2024; Li et al., 2024). Divergence times estimation indicated that the subfamilies of Faboideae and Caesalpinioideae diverged from a common ancestor ~ 76.44 MYA, while guar split from other Faboideae species ~ 46.89 MYA, suggesting that guar represents a relatively ancient lineage within this subfamily. Furthermore, we observed that species producing bean gums (highlighted in green in Fig. 1G) are distributed across both Faboideae and Caesalpinioideae subfamilies. This dispersed phylogenetic pattern implies that the genetic basis for bean gum biosynthesis likely originated in an early common ancestor of the Fabaceae family.

As the principal component of guar gum, galactomannan is widely distributed among Fabaceous species with different galactose: mannose (G/M) ratios, such as G/M = 2:1 in *C. tetragonoloba*, G/M = 3:1 in *Gleditsia sinensis*, G/M = 5:1 in *Senna tora* and G/M = 2:1 in *Sesbania cannabina*. Here, we reconstructed the biosynthetic pathway of galactomannan in guar, and visualized tissue-specific

expression patterns of related genes via a heatmap based on transcriptomic data from 8 tissues (Fig. 1H). Two key enzymes, *Mannan synthase (ManS)* and *galactomannan galactosyltransferase (GMGT1)*, which are critical for galactomannan synthesis (Dhugga et al., 2004), exhibited high expression in the endosperm (Fig. 1H), consistent with the endosperm being the primary site of guar gum biosynthesis. Additionally, other genes involved in galactomannan biosynthesis, such as *BGAL*, *GALM*, *GALK*, *PMM*, and *GDPMP*, also showed preferential expression in the endosperm. In contrast, the expression of genes related to GDP-D-Mannose degradation were suppressed (including *GDPM*, α -*MS* and *MAN*), thereby ensuring the efficiency for galactomannan biosynthesis (Fig. 1H). Correlation analysis further revealed a strong positive relationship ($r > 0.9$) between *ManS* and *GMGT1*, while they also showed high correlations ($r = 0.8 \sim 0.95$) with *GDPMP*, *BGAL* and *GALK*, which are involved in channeling metabolic flux toward galactomannan synthesis (Supplemental Fig. 2).

ManS and *GMGT1* play decisive roles in determining whether galactomannan is successfully synthesized. To elucidate the origin and evolutionary trajectory of galactomannan biosynthesis within the Fabaceae family, we examined the microsynteny of *ManS* and *GMGT1* across both galactomannan-producing and non-producing species. Homologs of *ManS* and *GMGT1* were first identified (Supplemental Fig. 3–4), and JCVI software was employed to visualize the locations of *ManS* and *GMGT1* gene syntenic regions between genomes (Fig. 1I, Supplemental Fig. 5). The results showed that syntenic regions containing *ManS* and *GMGT1* are widely present across Fabaceae species (Fig. 1I and Supplemental Fig. 5). Interestingly, *ManS* was duplicated in *G. max* and *S. cannabina*, while *GMGT1* were tripled in *G. max* and quadrupled in *S. cannabina*, while *G. max* and *S. cannabina* both suffered two WGD event. This suggests that *ManS* and *GMGT1* originated in a common ancestor of the Fabaceae prior to the divergence of the Faboideae and Caesalpinioideae subfamilies. We infer that one copy of *ManS* was lost in the Faboideae lineage after the shared WGD event but before genus-level differentiation. In contrast, *GMGT1* syntenic regions were largely retained across the family, with the exception of a recent loss in *G. max* following its lineage-specific WGD.

In summary, we have assembled a high-quality chromosome-level genome of guar and refined the understanding of its phylogenetic placement within the Fabaceae family. Furthermore, through comprehensive analysis of the origin and evolutionary pathways of *ManS* and *GMGT1*, we propose that the coordinated transcriptional regulation of these two genes is critical for galactomannan biosynthesis. This study provides valuable genomic resources and novel insights that will support future molecular breeding efforts aimed at improving galactomannan yield and quality in guar.

Declarations

Research funding

This study was supported by the Hainan Province Science and Technology Special Fund (ZDYF2023XDNY055), Hainan Natural Science Foundation of High Level Talents Project (325RC670), the

Hainan Provincial Natural Science Foundation of China (318QN192), and the Hainan Province Science and Technology Special Fund (ZDYF2021XDNY273).

Author contributions

HYH, YHD designed the study. ZY, XLZ, and HY analyzed the data. ZY, XLZ, and YHD wrote the manuscript. SCH and XYW prepared the samples. YYC, PL, and CZX helped data proceeding. JQH provided the materials. YHD and HYH revised the manuscript. HYH provide funding support.

Data availability statement

The genome sequence data resulting from this birch genome project have been deposited in the China National GeneBank (CNCB) with the accession number PRJCA042194. The original sequencing data, including Hi-C, PacBio, and RNA-seq data, are available in the Genome Sequence Archive (GSA) at the National Genomics Data Center (NGDC) under accession number CRA027220.

Conflict of interest statement

The authors declare no competing interests.

References

1. Dhugga, K.S., Barreiro, R., Whitten, B., Stecca, K., Hazebroek, J., Randhawa, G.S., Dolan, M., Kinney, A.J., Tomes, D., Nichols, S. and Anderson, P. (2004) Guar seed beta-mannan synthase is a member of the cellulose synthase super gene family. *Science* 303, 363–366.
2. Feng, L., Teng, F., Li, N., Zhang, J.C., Zhang, B.J., Tsai, S.N., Yue, X.L., Gu, L.F., Meng, G.H., Deng, T.Q., Tong, S.W., Wang, C.M., Li, Y., Shi, W., Zeng, Y.L., Jiang, Y.M., Yu, W., Ngai, S.M., An, L.Z., Lam, H.M. and He, J.X. (2024) A reference-grade genome of the xerophyte *Ammopiptanthus mongolicus* sheds light on its evolution history in legumes and drought-tolerance mechanisms. *Plant Communications* 5, 100891.
3. Gaikwad, K., Ramakrishna, G., Srivastava, H., Saxena, S., Kaila, T., Tyagi, A., Sharma, P., Sharma, S., Sharma, R., Mahla, H.R., Kumar, K., Sv, A.M., Solanke, A.U., Kalia, P., Rao, A.R., Rai, A., Sharma, T.R. and Singh, N.K. (2023) The chromosome-scale genome assembly of cluster bean provides molecular insight into edible gum (galactomannan) biosynthesis family genes. *Scientific Reports* 13, 9941.
4. Li, J.H., Li, M.J., Li, W.L., Li, X.Y., Ma, Y.B., Tan, X., Wang, Y., Li, C.X. and Ma, X.R. (2024) Leguminous industrial crop guar (*Cyamopsis tetragonoloba*): The chromosome-level reference genome de novo assembly. *Industrial Crops and Products* 216.
5. Pan, L., Liu, M., Kang, Y., Mei, X., Hu, G., Bao, C., Zheng, Y., Zhao, H., Chen, C., and Wang, N. 2023. Comprehensive genomic analyses of *Vigna unguiculata* provide insights into population differentiation and the genetic basis of key agricultural traits. *Plant Biotechnology Journal* 21:1426–1439.

6. Young, N.D., Debellé, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F., Gouzy, J., Schoof, H., et al. 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 480:520–524.
7. Zan, Y., Chen, S., Ren, M., Liu, G., Liu, Y., Han, Y., Dong, Y., Zhang, Y., Si, H., Liu, Z., et al. 2025. The genome and GeneBank genomics of allotetraploid *Nicotiana tabacum* provide insights into genome evolution and complex trait regulation. *Nat Genet* 57:986–996.
8. Zhao, Y., Zhang, R., Jiang, K.W., Qi, J., Hu, Y., Guo, J., Zhu, R., Zhang, T., Egan, A.N., Yi, T.S., Huang, C.H. and Ma, H. (2021) Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. *Molecular Plant* 14, 748–773.

Figures

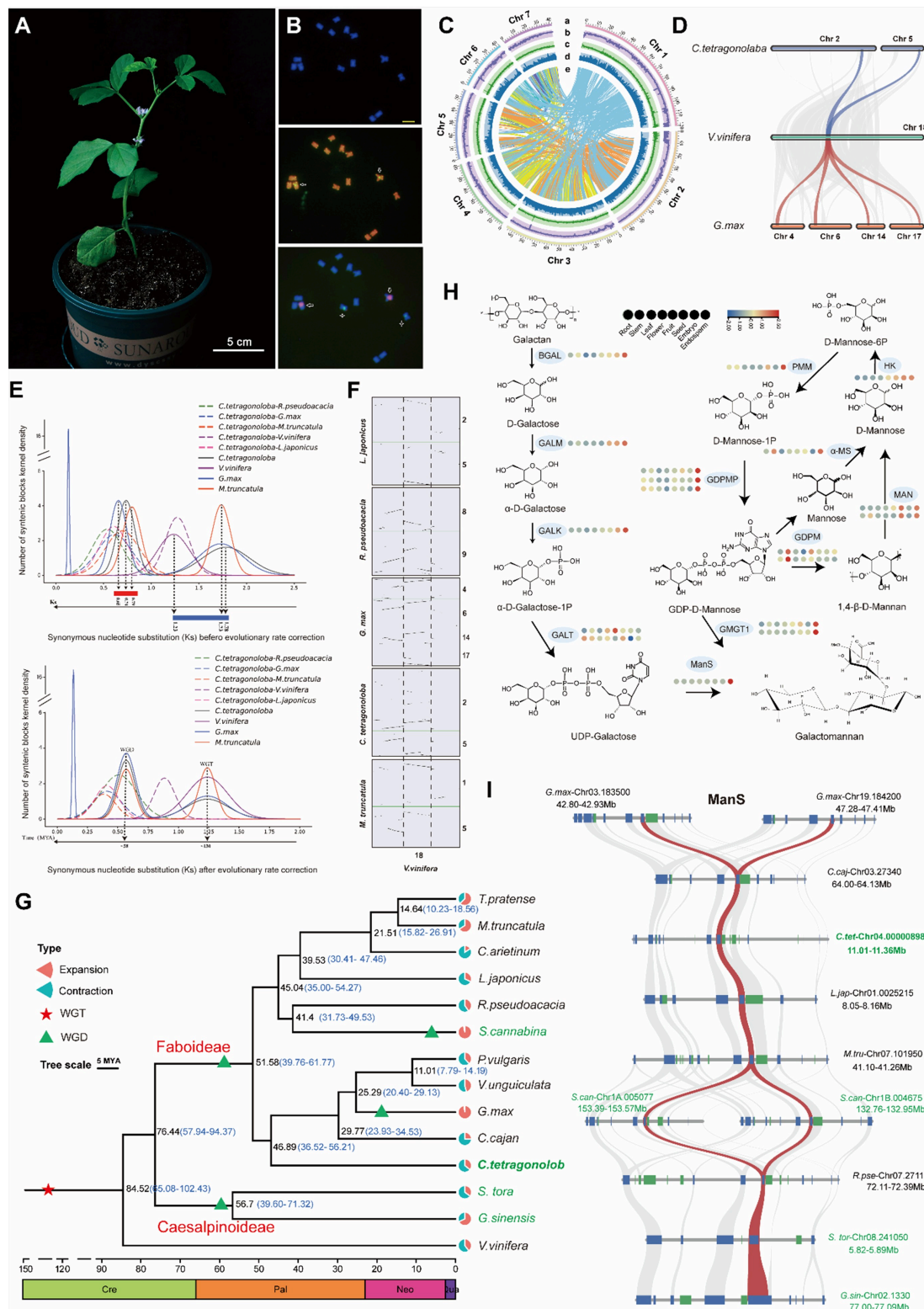


Figure 1

Genome assembly, comparative genomics, and galactomannan biosynthesis and evolution in *C. tetragonoloba*.

(A) Morphology of *C. tetragonoloba*. Scale bar = 5 cm.

- (B) Karyotype of *C. tetragonoloba* determined by Fluorescence *in situ* Hybridization (FISH). Red signals: 5S rDNA, green signals: 18S rDNA. Scale bars = 5 μ m.
- (C) Circos plot of the *C. tetragonoloba* genome. From outer to inner rings: (a): chromosomes; (b): transposable element (TE) density; (c): GC content; (d): gene density; (e): intragenomic synteny.
- (D) Syntenic blocks among *C. tetragonoloba*, *V. vinifera* and *G. max*. Highlighted are connections between chromosome 18 of *V. vinifera* and chromosomes 2 and 5 of *C. tetragonoloba*, and chromosomes 4, 6, 14, and 17 of *G. max*. Blue ribbons: synteny between *V. vinifera* and *C. tetragonoloba*, red ribbons: denote synteny between *V. vinifera* and *G. max*.
- (E) Distribution of synonymous substitution rates (K_s) for paralogous gene pairs in selected Fabaceae species, based on raw and corrected K_s values. Abbreviated: *R. pseudoacacia* (*Robinia pseudoacacia*), *M. truncatula* (*Medicago truncatula*) and *L. japonicus* (*Lotus japonicus*).
- (F) Local synteny between chromosome 18 of *V. vinifera* and homologous regions in *C. tetragonoloba*, *G. max*, *R. pseudoacacia*, *M. truncatula* and *L. japonicus*.
- (G) Phylogenetic tree of Fabaceae species constructed from orthologous genes. Node values indicate divergence times (black) and confidence intervals (blue), calibrated using *G. max* and *P. vulgaris*. Pie charts show expansion (red) and contraction (blue) of gene families. Red stars: whole-genome triplication (WGT) events; green triangles: whole-genome duplication (WGD) events. Species names highlighted in green indicate those capable of galactomannan production. Green-highlighted species produce galactomannan; bold text highlights *C. tetragonoloba*. Additional species abbreviations: *T. pratense* (*Trifolium pratense*), *C. arietinum* (*Cicer arietinum*), *S. cannabina* (*Sesbania cannabina*), *V. unguiculata* (*Vigna unguiculata*), *C. cajan* (*Cajanus cajan*), *S. tora* (*Senna tora*), *G. sinensis* (*Gleditsia sinensis*).
- (H) Galactomannan biosynthetic pathway and tissue-specific expression of related genes in *C. tetragonoloba*. Genes are marked in blue ovals; adjacent dots represent expression levels across tissues. Gene abbreviations: BGAL (β -galactosidase), GALM (galactose mutarotase), GALK (galactokinase), GALT (galactose-1-phosphate uridylyltransferase), ManS (mannan synthase), HK (hexokinase), α -Ms (α -Mannosidase), PMM (phosphomannomutase), GDPM (1,4- β -D-mannosyltransferase), GDPMP (mannose-1-phosphate guanylyltransferase), MAN (mannan endo-1,4- β -mannosidase), GMGT1 (galactomannan galactosyltransferase 1).
- (I) Microsynteny analysis of *ManS* and its flanking regions across *G. max*, *C. cajan*, *C. tetragonoloba*, *L. japonicus*, *M. truncatula*, *S. cannabina*, *R. pseudoacacia*, *S. tora*, and *G. sinensis*. Blue and green boxes represent homologous genes; red bands indicate syntenic *ManS* genes. Green-highlighted species produce galactomannan; bold text highlights *C. tetragonoloba*.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementalinformation.docx](#)
- [SupplementalFigures.docx](#)
- [SupplementalTables.xlsx](#)