# BambooBase: A comprehensive database of bamboo omics and systematics

Dear Editor,

Bamboos (Bambusoideae, Poaceae) are of significant economic and ecological importance, supporting billions of people and contributing to international trade (Wu et al., 2023). They play a crucial role in carbon fixation and support a multitude of other species including the giant panda (Clark et al., 2015). Since the release of the first draft genome of moso bamboo (*Phyllostachys edulis*) (Peng et al., 2013), there has been a notable increase in the sequencing, assembly, and annotation of diverse bamboo genomes (Guo et al., 2019; Zheng et al., 2022; Ma et al., 2024). An existing resource for bamboo genomics, BambooGDB (Zhao et al., 2014), provides information based on the individual draft reference genome of moso bamboo. What is needed, however, is a resource reinforcing multiple genome assemblies in Bambusoideae to enable researchers to unravel the evolutionary history of chromosomes and genes through clade-specific comparative genomic investigations. Here, we present BambooBase (https://bamboo.genobank.org/), a newly developed web-based database designed to provide a seamless user experience for bamboo omics and systematics.

BambooBase provides a comprehensive collection of 18 genomes, 476 transcriptomes, 16 epigenomes, and taxonomic resources of all 135 currently recognized genera of bamboos (Figure 1A). Notably, it provides valuable information on subgenome-scale sequence similarity- and collinearity-based homologous groups (Figure 1B). A diverse range of versatile analytical tools have been incorporated to assist users in performing various comparative, evolutionary, and functional analyses (Figure 1C). In addition to the genomic resources, BambooBase also serves as a comprehensive knowledge hub for the Bamboo Phylogeny Group II (Clark, 2023), offering inclusive and integrative information on the 135 genera within the Bambusoideae (Soreng et al., 2022) (Figure 1D).
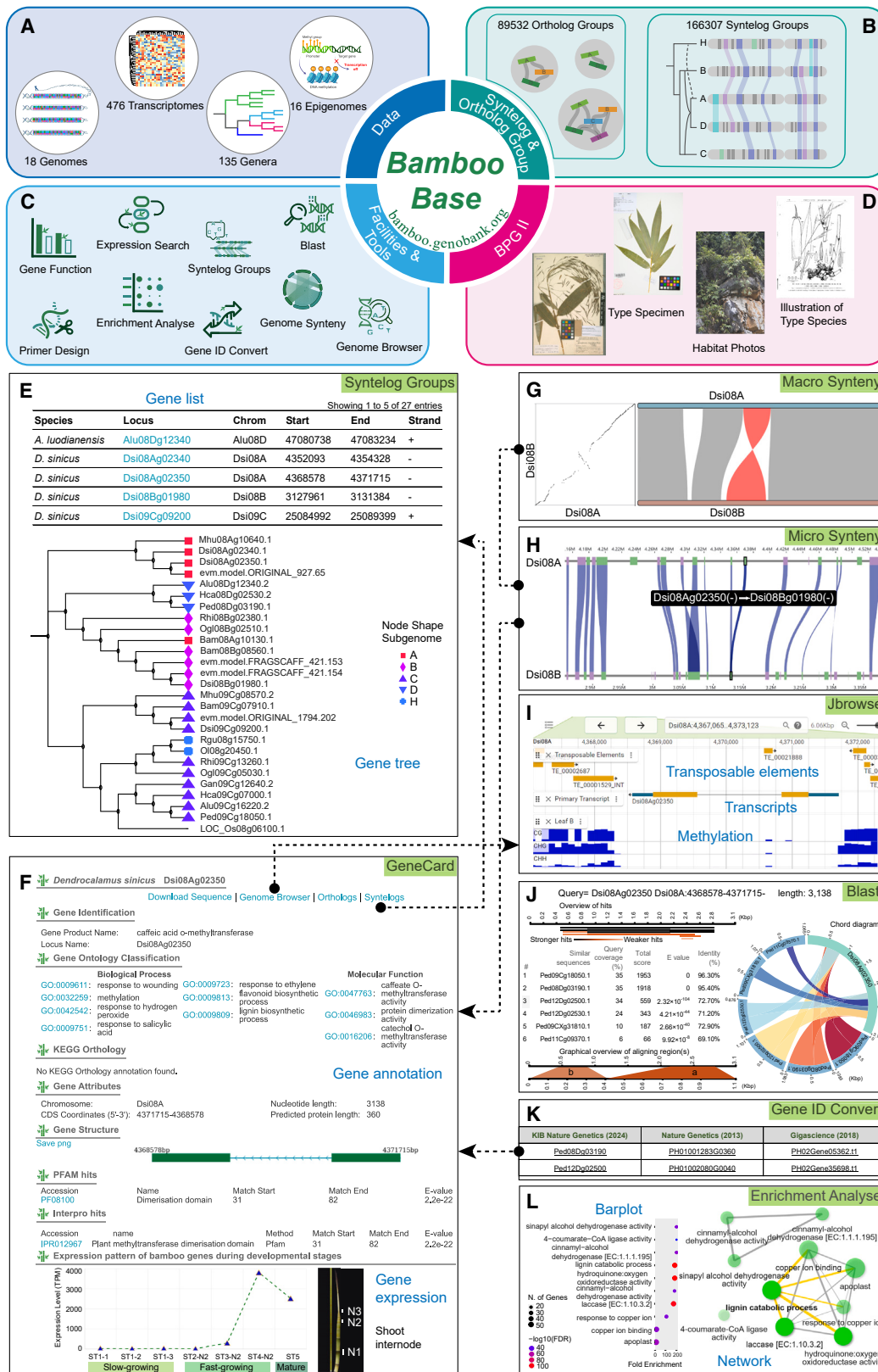
BambooBase now has 18 genome assemblies covering all 12 sequenced bamboo species to date (Supplemental Figure 1). Out of these, six are earlier draft genomes (Peng et al., 2013; Zhao et al., 2018; Guo et al., 2019), and the remaining 12 are reference-quality genomes (Zheng et al., 2022; Ma et al., 2024). The species with reference-quality genomes include two herbaceous bamboos with a diploid level (2n = 2x = 22), as well as 10 woody bamboos from three distinct clades of temperate, neotropical, and paleotropical bamboos representing different tetraploid (2n = 4x = 46–48) and hexaploid (2n = 6x = 70–72) species (Supplemental Figure 1).

A comprehensive annotation of protein-coding genes was performed on the 12 bamboo genome assemblies, resulting in the annotation of a total of 566 676 genes (Supplemental Table 1). This annotation process utilized seven public protein and domain databases. A total of 507 720 bamboo genes were found to have homologous sequences in the NCBI NR databases through Blastp (Supplemental Table 1). The putative domains, Gene Ontology (GO) terms, and KEGG categories of bamboo genes were annotated using InterProScan with default parameters. This annotation process assigned 211 234 genes with Pfam domains, 391 712 genes with GO terms, and 203 086 genes with KEGG categories (Supplemental Table 1). Additionally, gene expression profiles of bamboo species in various tissues, spanning shoot to spikelet developmental stages (Supplemental Table 2), are also accessible in BambooBase. 43 333 transcription factor genes have been identified and classified into 68 families using iTAK, with an average of 3611 transcription factor genes per bamboo genome (Supplemental Table 3).

Bambusoideae is known for its multiple allopolyploidization events and reticulate evolution, resulting in complex relationships among the clades (Figure 1B and Supplemental Figure 1) (Ma et al., 2024). The management of a homology database in Bambusoideae thus poses greater challenges compared to many other grasses. To set up a solid foundation for this database, we have utilized the chromosome-level reference genomes (Supplemental Table 1) of 12 bamboo species that represent all the major clades of Bambusoideae. To identify inter-diploid syntenic orthologs (syntelogs), pairwise alignments were conducted for each pair of diploid genomes or subgenomes of allopolyploids, using rice as a diploid outgroup. Subsequently, a synteny-based gene family cluster called SynPan was utilized to categorize approximately 545 082 genes from the resulting 27 (sub)genomes into 166 307 syntelog groups (SGs) (Supplemental Figure 2; Supplemental Table 4). The size of SGs is larger than that of ortholog groups obtained using the Markov clustering approach OrthoFinder. Moreover, SynPan demonstrated superior performance compared to the Markov clustering method in distinguishing paralogs from orthologs, particularly in the core (798 vs. 701) and soft-core (2996 vs. 2684) groups (Supplemental Table 4). SGs also provide more flexible and comprehensive clade-specific and subgenome-based orthologous gene data than the approach of "perfect-copy" (as defined by Guo et al., 2019) syntenic genes. Within BambooBase, there is a "homolog" section presenting a link table that showcases all the corresponding homologs belonging to the same SG or ortholog group (Figure 1E). This section also incorporates a phylogenetic tree (Figure 1E) to visually depict the evolutionary relationships among these homologous genes.

BambooBase further offers "GeneCard" pages that display comprehensive information about individual genes, along with convenient links to other modules and external databases, all on a single page (Figure 1F). Users can access the

---

**A** 18 Genomes | 476 Transcriptomes | 16 Epigenomes | 135 Genera

**B** 89532 Ortholog Groups | 166307 Syntelog Groups

Bamboo Base
bamboo.genobank.org

Data | Syntelog & Ortholog Group | BPG II | Facilities & Tools

**C** Gene Function | Expression Search | Blast | Syntelog Groups | Primer Design | Enrichment Analyse | Gene ID Convert | Genome Synteny | Genome Browser

**D** Type Specimen | Habitat Photos | Illustration of Type Species

**E** Syntelog Groups

Gene list

Showing 1 to 5 of 27 entries

| Species | Locus | Chrom | Start | End | Strand |
|---|---|---|---|---|---|
| *A. luodianensis* | Alu08Dg12340 | Alu08D | 47080738 | 47083234 | + |
| *D. sinicus* | Dsi08Ag02340 | Dsi08A | 4352093 | 4354328 | - |
| *D. sinicus* | Dsi08Ag02350 | Dsi08A | 4368578 | 4371715 | - |
| *D. sinicus* | Dsi08Bg01980 | Dsi08B | 3127961 | 3131384 | - |
| *D. sinicus* | Dsi09Cg09200 | Dsi09C | 25084992 | 25089399 | + |

Gene tree

Mhu08Ag10640.1
Dsi08Ag02340.1
Dsi08Ag02350.1
evm.model.ORIGINAL_927.65
Alu08Dg12340.2
Hca08Dg12530.2
Ped08Dg03190.1
Rhi08Bg02380.1
Ogl08Bg02510.1
Bam08Ag10130.1
Bam08Bg08560.1
evm.model.FRAGSCAFF_421.153
evm.model.FRAGSCAFF_421.154
Dsi08Bg01980.1
Mhu09Cg08570.2
Bam09Cg07910.1
evm.model.ORIGINAL_1794.202
Dsi09Cg09200.1
Rgu08g15750.1
Ol08g20450.1
Rhi09Cg13260.1
Ogl09Cg05030.1
Gan09Cg12640.2
Hca09Cg07000.1
Alu09Cg16220.2
Ped09Cg18050.1
LOC_Os08g06100.1

Node Shape Subgenome
A
B
C
D
H

**F** GeneCard

*Dendrocalamus sinicus* Dsi08Ag02350

Download Sequence | Genome Browser | Orthologs | Syntelogs

Gene Identification
Gene Product Name: caffeic acid o-methyltransferase
Locus Name: Dsi08Ag02350

Gene Ontology Classification

Biological Process
GO:0009611: response to wounding
GO:0032259: methylation
GO:0042542: response to hydrogen peroxide
GO:0009751: response to salicylic acid
GO:0009723: response to ethylene
GO:0009813: flavonoid biosynthetic process
GO:0009809: lignin biosynthetic process

Molecular Function
GO:0047763: caffeate O-methyltransferase activity
GO:0046983: protein dimerization activity
GO:0016206: catechol O-methyltransferase activity

KEGG Orthology
No KEGG Orthology annotation found.

Gene annotation

Gene Attributes
Chromosome: Dsi08A
CDS Coordinates (5'-3'): 4371715-4368578
Nucleotide length: 3138
Predicted protein length: 360

Gene Structure
Save png
4368578bp ———————————— 4371715bp

PFAM hits
| Accession | Name | Match Start | Match End | E-value |
|---|---|---|---|---|
| PF08100 | Dimerisation domain | 31 | 82 | 2.2e-22 |

Interpro hits
| Accession | name | Method | Match Start | Match End | E-value |
|---|---|---|---|---|---|
| IPR012967 | Plant methyltransferase dimerisation domain | Pfam | 31 | 82 | 2.2e-22 |

Expression pattern of bamboo genes during developmental stages

Gene expression

Shoot internode
N3 N2 N1
ST1-1 ST1-2 ST1-3 ST2-N2 ST3-N2 ST4-N2 ST5
Slow-growing | Fast-growing | Mature

**G** Macro Synteny
Dsi08A
Dsi08B
Dsi08A Dsi08B

**H** Micro Synteny
Dsi08A
Dsi08Ag02350(-) → Dsi08Bg01980(-)
Dsi08B

**I** Jbrowse
Dsi08A:4,367,065..4,373,123 6.06Kbp
Transposable Elements
TE_00002687 TE_00001888 TE_00003
TE_00001529_INT
Primary Transcript
Dsi08Ag02350
Transposable elements
Transcripts
Leaf B
CG CHG CHH
Methylation

**J** Blast
Query= Dsi08Ag02350 Dsi08A:4368578-4371715- length: 3,138
Overview of hits
Stronger hits | Weaker hits
Chord diagram

| # | Similar sequences | Query coverage (%) | Total score | E value | Identity (%) |
|---|---|---|---|---|---|
| 1 | Ped09Cg18050.1 | 35 | 1953 | 0 | 96.30% |
| 2 | Ped08Dg03190.1 | 35 | 1918 | 0 | 95.40% |
| 3 | Ped12Dg02500.1 | 34 | 559 | 2.32×10⁻¹⁰⁴ | 72.70% |
| 4 | Ped12Dg02530.1 | 24 | 343 | 4.21×10⁻⁴⁴ | 71.20% |
| 5 | Ped09Cg31810.1 | 10 | 187 | 2.66×10⁻⁴⁰ | 72.90% |
| 6 | Ped11Cg09370.1 | 6 | 66 | 9.92×10⁻⁸ | 69.10% |

Graphical overview of aligning region(s)
b a

**K** Gene ID Convert

| KIB Nature Genetics (2024) | Nature Genetics (2013) | Gigascience (2018) |
|---|---|---|
| Ped08Dg03190 | PH01001283G0360 | PH02Gene05362.t1 |
| Ped12Dg02500 | PH01002080G0040 | PH02Gene35698.t1 |

**L** Enrichment Analyse

Barplot
sinapyl alcohol dehydrogenase activity
4-coumarate-CoA ligase activity
cinnamyl-alcohol dehydrogenase [EC:1.1.1.195]
lignin catabolic process
hydroquinone:oxygen oxidoreductase activity
cinnamyl-alcohol dehydrogenase activity
laccase [EC:1.10.3.2]
response to copper ion
copper ion binding
apoplast

N. of Genes
20 30 40 50
-log10(FDR)
40 60 80 100

Fold Enrichment

Network
cinnamyl-alcohol dehydrogenase activity
cinnamyl-alcohol dehydrogenase activity [EC:1.1.1.195]
copper ion binding
apoplast
lignin catabolic process
response to copper ion
4-coumarate-CoA ligase activity
laccase [EC:1.10.3.2]
hydroquinone:oxygen oxidoreductase activity

*(legend on next page)*

GeneCard page by entering a bamboo gene ID in the search box on the homepage and "locus search" page. The GeneCard is organized into two parts. The first part provides essential details about genes, including their unique gene IDs, sequences, structural information, and functional annotations. The second part presents their expression profiles across diverse tissues and developmental stages (Figure 1F). The results of the co-expression network analysis, derived from 476 transcriptomes utilizing WGCNA, can be accessed via the "co-expression" module (Supplemental Figure 3).

BambooBase allows users to conveniently select any two bamboo species and visualize their genomic collinear blocks ("macro synteny") and collinear genes ("micro synteny") in the database (Figures 1G and 1H). This, in combination with the "genome browser," provides valuable assistance to users for exploring and visualizing genomic characteristics and methylation levels associated with individual bamboo genes (Figure 1I). In terms of sequence-based searches, the widely used "Blast" tool enables users to query amino acid or nucleotide sequences against all the protein sequences (Figure 1J). Figures 1E–1J depict the utilization of the caffeic acid O-methyltransferase gene from *Dendrocalamus sinicus* to demonstrate its annotation, expression, collinearity, and homologs across genomes and subgenomes. Additionally, the "primer designer" module can be used to design PCR primers (Supplemental Figure 4), thereby facilitating subsequent experimentation. The "gene ID convert tool" enables the linking of gene models across different assemblies of the same species (Figure 1K and Supplemental Figure 1). An "enrichment analysis" tool was developed for efficient functional determination of a list of genes using ShinyGO. Users can perform functional gene enrichment analysis using GO term and KEGG pathway catalogs (Figure 1L). The results highlight significantly enriched functional categories, which are cross-linked to specific public databases.

The Bamboo Phylogeny Group II section provides cross-referenced data on the currently accepted three tribes, 19 subtribes, and 135 genera of the Bambusoideae (Supplemental Table 5). This includes the scientific name of each accepted genus, its protologue, synonym(s), type species, morphological description, distribution, and phylogenetic remarks. In addition, related research articles, photos, species covered within each genus, and other information are also provided when available (Supplemental Figure 5).

BambooBase was constructed on an Apache web server, which includes an indexed database powered by MySQL. A graphical user interface was developed using the spring-boot framework in Java. The browser and database communicate and manipulate data through a series of jQuery function calls, executed using JavaScript, Perl, and R script (v.4.3.0). The third-party databases and software employed by the platform are detailed in Supplemental Table 6.

BambooBase represents a significant advancement in bamboo genomics research by providing an integrated platform that combines high-quality genome assemblies and diverse transcriptomic datasets with comprehensive knowledge on bamboo systematics. This enables the interrogation, utilization, and exploitation of the 12 available chromosomal-level genomes of bamboos. Like the Rice Gene Index (https://riceome.hzau.edu.cn), the Triticeae-GeneTribe (http://wheat.cau.edu.cn/TGT/), and the Tea Plant Information Archive 2 (http://tpia.teaplants.cn), BambooBase provides a homology inference platform for genomes at different ploidy levels by incorporating gene collinearity, demonstrating superior accuracy and scalability compared to the general sequence-similarity-based methods. Moreover, it offers enhanced search and visualization capabilities and will serve as a valuable scalable resource for the global bamboo research community. BambooBase will be regularly updated when new reference genome assemblies and other omics datasets of bamboos become available.

### SUPPLEMENTAL INFORMATION
Supplemental information is available at *Molecular Plant Online*.

### AUTHOR CONTRIBUTIONS
D.-Z.L., Y.-L.L., and P.-F.M. conceived and designed the project. Y.-L.L., G.J., S.-Y.G., C.G., M.-Y.Z., Y.-Z.Y., and L.-Z.N. collected data and performed analysis. S.-Y.G., and Q.G. built the database and managed the computing platforms. Y.-L.L., D.-Z.L., and P.-F.M. wrote and edited the

---

**Figure 1. Schematic design and online display examples of BambooBase.**
The caffeic acid O-methyltransferase (*COMT*) gene exemplifies the potential of utilizing diverse modules in data mining **(E–J)**.
**(A)** The data layer summarizing the source data provided in BambooBase.
**(B)** Schematic illustration of ortholog groups and syntenic ortholog (syntelog) groups deduced through the Markov clustering method by OrthoFinder and the collinearity approach by SynPan.
**(C)** Integrated access and analytical facilities and tools in BambooBase for comparative, evolutionary, and functional analyses.
**(D)** Bamboo Phylogeny Group II (BPG II). A comprehensive knowledge repository for the 135 recognized genera within the Bambusoideae.
**(E)** Syntelog groups section displaying *COMT*-corresponding syntelogs and phylogenetic gene tree illustrating evolutionary relationships.
**(F)** GeneCard pages displaying comprehensive gene information and expression profiles: an example of a *COMT-A* gene (Dsi08Ag02350) in *Dendrocalamus sinicus*.
**(G)** Macro synteny module depicting collinear blocks and highlighted inversions in chromosomes 8A and 8B of *D. sinicus*.
**(H)** Micro synteny module illustrating local-scale homologous relationships surrounding *COMT* in *D. sinicus*.
**(I)** "JBrowse" facilitating exploration and visualization of genomic characteristics and methylation levels in *COMT-A* genes and adjacent regions.
**(J)** Blast tool enabling sequence-based searches of amino acid or nucleotide sequences against all protein sequences.
**(K)** Gene ID convert tool for linking gene models across different assemblies of the same species.
**(L)** Enrichment analysis tool for efficient functional determination of gene lists using GO term and KEGG pathway catalogs.

paper with input from Z.-H.G., E.X., S.-Y.G., and Q.G. All authors read and approved the final manuscript.

Yun-Long Liu[1,6,*], Shu-Yang Gao[1,2,6],
Guihua Jin[1,6], Meng-Yuan Zhou[1,6],
Qijuan Gao[3,4,6], Cen Guo[1], Yi-Zhou Yang[1,2],
Liang-Zhong Niu[1], Enhua Xia[3],
Zhen-Hua Guo[1], Peng-Fei Ma[1,*] and
De-Zhu Li[1,2,5,*]

[1]Germplasm Bank of Wild Species & Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China
[2]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan 650201, China
[3]State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, Anhui 230036, China
[4]School of Computer and Artificial Intelligence, Hefei Normal University, Hefei, Anhui 230061, China
[5]Key Laboratory for Plant Diversity and Biogeography in East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China
[6]These authors contributed equally to this article.
*Correspondence: Yun-Long Liu (liuyunlong@mail.kib.ac.cn), Peng-Fei Ma (mapengfei@mail.kib.ac.cn), De-Zhu Li (dzl@mail.kib.ac.cn)
https://doi.org/10.1016/j.molp.2024.02.017

## REFERENCES

Clark, L.G., Londoño, X., and Ruiz-Sanchez, E. (2015). Bamboo Taxonomy and Habitat. In Tropical Forestry (Springer International Publishing), pp. 1–30. https://doi.org/10.1007/978-3-319-14133-6_1.

Clark, L.G. (2023). Integrating genomic and morphological data into bamboo taxonomic and evolutionary studies. Plant Divers. **45**:123–124. https://doi.org/10.1016/j.pld.2023.02.002.

Guo, Z.H., Ma, P.F., Yang, G.Q., Hu, J.Y., Liu, Y.L., Xia, E.H., Zhong, M.C., Zhao, L., Sun, G.L., Xu, Y.X., et al. (2019). Genome sequences provide insights into the reticulate origin and unique traits of woody bamboos. Mol. Plant **12**:1353–1365. https://doi.org/10.1016/j.molp.2019.05.009.

Ma, P.F., Liu, Y.L., Guo, C., et al. (2024). Genome assemblies of 11 bamboo species highlight diversification induced by dynamic subgenome dominance. Nat. Genet. https://doi.org/10.1038/s41588-024-01683-0.

Peng, Z., Lu, Y., Li, L., Zhao, Q., Feng, Q., Gao, Z., Lu, H., Hu, T., Yao, N., Liu, K., et al. (2013). The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). Nat. Genet. **45**:456–461e4612. https://doi.org/10.1038/ng.2569.

Soreng, R.J., Peterson, P.M., Zuloaga, F.O., Romaschenko, K., Clark, L.G., Teisher, J.K., Gillespie, L.J., Barberá, P., Welker, C.A.D., Kellogg, E.A., et al. (2022). A worldwide phylogenetic classification of the Poaceae (Gramineae) III: An update. J. Systemat. Evol. **60**:476–521. https://doi.org/10.1111/jse.12847.

Wu, J., Che, J., and Smith, A.G. (2023). INBAR (2023) Trade Overview 2021: Bamboo and Rattan Commodities in China.

Zhao, H., Peng, Z., Fei, B., Li, L., Hu, T., Gao, Z., and Jiang, Z. (2014). BambooGDB: a bamboo genome database with functional annotation and an analysis platform. DATABASE-OXFORD **2014**:bau006. https://doi.org/10.1093/database/bau006.

Zhao, H., Gao, Z., Wang, L., Wang, J., Wang, S., Fei, B., Chen, C., Shi, C., Liu, X., Zhang, H., et al. (2018). Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*). GigaScience **7**:giy115. https://doi.org/10.1093/gigascience/giy115.

Zheng, Y., Yang, D., Rong, J., Chen, L., Zhu, Q., He, T., Chen, L., Ye, J., Fan, L., Gao, Y., et al. (2022). Allele-aware chromosome-scale assembly of the allopolyploid genome of hexaploid Ma bamboo (*Dendrocalamus latiflorus* Munro). J. Integr. Plant Biol. **64**:649–670. https://doi.org/10.1111/jipb.13217.