scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly and annotation of *Phyllostachys violascens* 'Prevernalis'

Haixia Wang[1,2]✉, Hui Ding[1], Ping Cheng[1], Xianjiang Jin[1], Yanru Fu[1] & Yizhen Peng[1]

Phyllostachys violascens **'Prevernalis' is one of the best species for shoots, and the Prevernalis shoots are widely favored by people for its high nutritional value, high medicinal value and delicious taste. However, high-quality reference genome for Prevernalis has not been published, primarily due to its polyploidy and large genome size. Here, we generated a high-quality genome for tetraploid Prevernalis using a combination of PacBio long reads and Hi-C sequencing techniques. The final genome length of Prevernalis is 2156.14 Mb, with a scaffold N50 size of 56.48 Mb, a BUSCO score of 97.51% and LAI score of 20.03. The assembled genome comprises 68.45% repetitive elements, and was anchored to twenty-four chromosomes. Moreover, we identified a total of 53,558 protein-coding genes, of which 94.13% were functionally annotated. Furthermore, comparing the genome of Prevernalis with seven other species showed that the gene numbers and structures of Prevernalis are similar with *Ph.edulis*. Overall, the high-quality genome assembly and gene annotation resources will provide crucial resources for functional genomic studies and molecular breeding in Prevernalis.**

Background & Summary

Phyllostachys violascens 'Prevernalis' (2n = 4x = 48), a member of the Bambusoideae subfamily in the grass family and the *Phyllostachys* genus is native to Lin 'an, Anji and Yuhang in Zhejiang province and widely distributed and cultivated in southern China[1,2]. Prevernalis is a variety of *Phyllostachys violascens* and commonly known as "Lei bamboo", because it grows bamboo shoots when thunder hits in early spring. The Prevernalis shoot is one of the most valuable bamboo shoots and has the characteristics of early shoot emergence, long shoot periods, high yield and delicious taste[3–5]. In addition, the Lei bamboo shoots have high nutritional value, containing essential amino acids, minerals, vitamins and dietary fiber, with low sugars and fats, and is called "the first bamboo shoot in Jiangnan". Moreover, Lei bamboo shoots are rich in phenols and phytosterols and other bioactive compounds, which contributes to the prevention of obesity, diabetes, hypertension, cancer and other diseases[6,7].

Bamboo is among the fastest growing plants on Earth[8], and comprise one diploid herbaceous and three polyploid (tetraploid or hexaploid) woody lineages[9,10]. In the last decade, there has been great progress in the genomics of bamboo. Four draft genome assemblies of the three different ploidy levels (diploid, tetraploid, and hexaploid) were generated, providing insight into the evolutionary characteristics and unique flowering and woody trait of bamboo[11]. The tetraploid woody bamboo Moso bamboo (*Phyllostachys edulis*) was the first to construct chromosome level genome assembly, and then a genome variation map containing 5.45 million single nucleotide polymorphisms (SNPs) was constructed by resequencing 427 individuals from 15 representative regions[12–14]. Recently, chromosomal-level genome assembly has been carried out for another 13 bamboo species, including two allohexaploid woody bamboo, Ma bamboo (*Dendrocalamus latiflorus* Munro) and *Dendrocalamus brandisii* (Munro) Kurz by single-molecule real-time (SMRT) sequencing technology from the Pacific Biosciences (PacBio) Sequel platform, the Oxford Nanopore Technology (ONT) and chromosome conformation capture sequencing (Hi-C)[6,15–17]. Moreover, both genetic transformation and gene editing techniques mediated by (clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein 9 (Cas9) have been established for Ma bamboo[18–20]. These efforts have greatly facilitated the study of molecular regulatory

[1]Jiangxi Academy of Forestry, Jiangxi, 330013, China. [2]Jiangxi Provincial Key Laboratory of Improved Variety Breeding and Efficient Utilization of Native Tree Species, Jiangxi, 330013, China. ✉e-mail: 44916972@qq.com

Fig. 1 Morphological characteristics of Prevernalis sequenced in this study.

mechanisms of agriculturally important traits in bamboo. For example, the mechanisms underlying rapid internodes growth (114.5 cm/day) in Moso bamboo have been unveiled through a combination of morphological investigations, molecular biology, phylostratigraphy and evolutionary transcriptomic assays[8]. MicroRNAs also play important roles in regulating the rapid growth of the *Ph.edulis* culm internode[21,22]. However, although plant hormones, such as auxin and cytokinin were reported to regulate bamboo shoot development[23,24], the genetic basis and molecular regulatory mechanisms of nutritional and flavor quality traits in bamboo shoots remain obscure. Prevernalis is a widely cultivated tetraploid woody bamboo, and its shoots are sweet, fresh, delicious and rich in nutrients[1]. Therefore, Prevernalis may be an important species to study the nutritional and flavor quality of bamboo shoots, and there is an urgent need to obtain a high-quality reference genome for Prevernalis.

In this study, we generated 66.65 Gb of PacBio High fidelity (HiFi) long-reads for genome assembly, and the N50 length of the HiFi reads was 18.26 kb. An additional 246.85 Gb of Hi-C sequencing data was generated to validate the genome assembly, and we finally generated a high-quality chromosome-level reference genome of Prevernalis. The assembled genome has a total length of 2156.14 Mb, with a scaffold N50 length of 56.48 Mb, and 89.09% (1.92 Gb) of the genome sequences are successfully anchored to 24 pseudo-chromosomes. Genome annotation predicted 53,558 protein-coding genes, and 50416 (94.13%) of them were functionally annotated. The genome evolution analysis showed that the Prevernalis genome was most similar to *Ph.edulis*, including gene number and gene structure. In addition, comparative genomic analysis showed that a total of 487 significantly expanded gene families were identified, including 2629 genes, and 117 significantly contracted gene families, including 378 genes in Prevernalis. GO and KEGG enrichment analysis showed that these genes were mainly involved in the regulation of metabolites such as amino acids, sugars and nucleic acids, indicating that Prevernalis may have evolved more metabolic regulatory genes, further optimizing metabolic regulatory pathways, increasing the abundance and variety of metabolites, and thus increasing nutritional and flavor quality of bamboo shoots. Taken together, the high-quality chromosome-level genome in this study provides an important basis for studying bamboo shoot growth, development, and quality improvement in Prevernalis.

Methods

Sample collection and DNA extraction. Young fresh leaves of Prevernalis were collected from one sample individual grown in Jiangxi National Bamboo Germplasm Resource Bank in Nanchang City, Jiangxi Province, China (Fig. 1), and immediately frozen in liquid nitrogen for genomic DNA extraction. Total genomic DNA was extracted from from Prevernalis young fresh leaves according to the cetyltrimethylammonium bromide (CTAB) method, and the quality and quantity of the genomic DNA samples were assessed through 1% agarose gel electrophoresis, Nanodrop and Qubit.

PacBio library construction, sequencing and assembly. To perform *de novo* genome assembly, single-molecule real-time (SMRT) PacBio library was prepared using the SMRTbell Express Template Preparation Kit 2.0 (Pacifc Biosciences, USA) following the manufacturer's instructions, and sequenced on the PacBio Sequel II platform in Circular Consensus Sequence (CCS) mode to produce PacBio HiFi long reads. The total CCS reads

Mode	Total length (bp)	Total number	Total number (>=2 kb)	max length (bp)	N50 (bp)	N90 (bp)	GC content (%)
hifiasm	2,834,556,489	166	166	131,458,989	44,244,632	13,579,970	44.34
hifiasm + purge_haplotigs	2,156,138,571	60	60	131,458,989	56,476,697	20,599,053	44.33

Table 1. Statistics of the Prevernalis draft genome assembly.

length was 66.65 Gb (~30.9 × coverage), with an average length of 17.75 kb, and the CCS reads N50 size is 18.26 kb (Table S1). Then, HiFi reads generated by PacBio sequencing were assembled to produce the primary contigs, using the default settings of the Hifiasm v 0.16.1 algorithm[25]. Subsequently, the Purge Haplotigs program15[26] was employed to eliminate redundant contigs. The final draft genome assembly of Prevernalis was 2156.14 Mb with a scaffold N50 of 56.48 Mb, and the maximum contig size was 131.46 Mb (Table 1).

Hi-C library preparation, sequencing and chromosome anchoring. To achieve the chromosome-level genome assembly of Prevernalis, Hi-C data was generated to anchor and orient the draft genome contigs. Young fresh leaves (~1 g) of Prevernalis were fixed with 1% formaldehyde to coagulate the conformation of DNA. After cell lysis, the genomic DNA was then enzymatically digested with MboI (a 4-cutter restriction enzyme), generating fragments with sticky ends. The restriction fragment overhangs were repaired and biotin was introduced to label the end of oligonucleotide. Subsequently, the DNA fragments were ligated together to form chimeric circles using DNA ligase. The ligated DNAs were then uncrosslinked, purifed, and sheared into 300–500 bp in size, and pulled down with streptavidin beads. Finally, the Hi-C sequencing library was sequenced on the Illumina NovaSeq 6000 platform, and the total Hi-C data was 251.72 Gb (~116.7 × coverage) (Table S2). After filtering reads with average quality scores less than 20 and removing adapters using fastp v0.23.2[27], a total of 246.85 Gb clean data were obtained (Table S2). We also utilized the HiCUP pipeline[28] to remove the erroneous mappings and duplicated contigs to yield the interaction matrix. This matrix served as the foundation for anchoring the contigs into chromosomes via the 3D-DNA pipeline[29], and the chromosome segmentation boundaries and assembly errors were manually checked and adjusted utilizing Juicebox Assembly Tools[30]. Using the Hi-C data, the assembled draft genome sequences were further anchored to 24 chromosomes, with a cumulative length of 1.92 Gb, covering ~89.09% of the scaffold-level genome (Figs. 2 and 3).

Repetitive sequence annotation. Repeat elements in the Prevernalis genome were identified by the combination of homologous prediction method based on the RepBase library (http://www.girinst.org/repbase)[31] and *de novo* predictions method based on own sequence alignment and repeated sequence features. The homology search of repeat elements were executed using RepeatMasker and RepeatProteinMask[32] and the *de novo* repetitive elements were predicted by RepeatModeler[33], Piler[34] and Trf[35]. The combination of Repbase and our *de novo* TE library revealed that 66.94% of the assembled Prevernalis genome was annotated as repetitive elements, of which DNAs, LINEs, SINEs and LTRs accounted for 10.86%, 1.58%, 0.02% and 52.82% of the whole genome, respectively (Fig. 2B and Table 2).

Gene prediction and annotation. Based on the repeat-masked genome, the prediction of protein-coding genes within the Prevernalis genome assembly was approached through three distinct strategies: homologous, *de novo* prediction and RNA-sequencing methods. The homology-based prediction involved the retrieval of protein sequences from eleven grass family species downloaded from the NCBI database. The *de novo* prediction was performed using Augustus[36] and GlimmerHMM[37] with default settings. Transcriptome data assembled by Tophat and Cufflinks was also used for the genome annotation by using HISAT2[38]. Finally, the measured gene sets are integrated into a non-redundant, more complete gene set using MAKER[39] and HiFAP (Wuhan OneMore Tech Co., Ltd., https://www.onemore-tech.com/) with default settings, and a total of 53,558 protein-coding genes were predicted for the Prevernalis genome (Fig. 3 and Table S3). The average length of the predicted protein-coding gene was 5000 bp, and each gene has an average of 4.87 exons. The average lengths of the exon and intron were 281.91 bp and 936.59 bp, respectively (Table S3). Gene functions were assigned by aligning the protein sequences to the NCBI nonredundant protein (NR), Swiss-Prot, TrEMBL, GO, InterPro, TF, Pfam, eukaryotic orthologous groups of proteins (KOG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, respectively, and approximately 94.13% (50,416 genes) of the total predicted genes were successfully annotated by at least one of the databases (Table 3).

Non-coding RNA prediction and annotation. The tRNAscan-SE[40] software is used to predict the transfer RNAs (tRNA) in the Prevernalis genome according to the structural characteristics of tRNA, while BLASTN program was used for ribosomal RNAs (rRNA) prediction, and other ncRNAs, including micro-RNAs (miRNAs) and small nuclear RNAs (snRNAs) were identified by searching in the miRBase[41] and Rfam[42] database using the INFERNAL[43] software with default parameters. Finally, a total of 7,745 non-coding RNAs were predicted, including 593 miRNAs, 852 tRNAs, 4,844 rRNAs, and 1,456 snRNAs (Table 4).

Data Records

The raw data of this study have been deposited in the NCBI Sequence Read Archive (SRA) database with Bioproject ID: PRJNA1081620 and BioSample ID: SAMN40964207[44]. The final genome assembly data have been deposited in the NCBI with the accession ID JBINAA000000000[45]. DNA sequencing data from the PacBio long-reads were deposited in the SRA at SRR31132316[46]. DNA sequencing data from the Hi-C sequencing data were deposited in the SRA at SRR31132317[47]. The transcriptome data are available under accession number SRR29138491[48]. The gff or gft files for gene and TE annotations are available with the Figshare link, https://doi.org/10.6084/m9.figshare.27314118.v1[49].

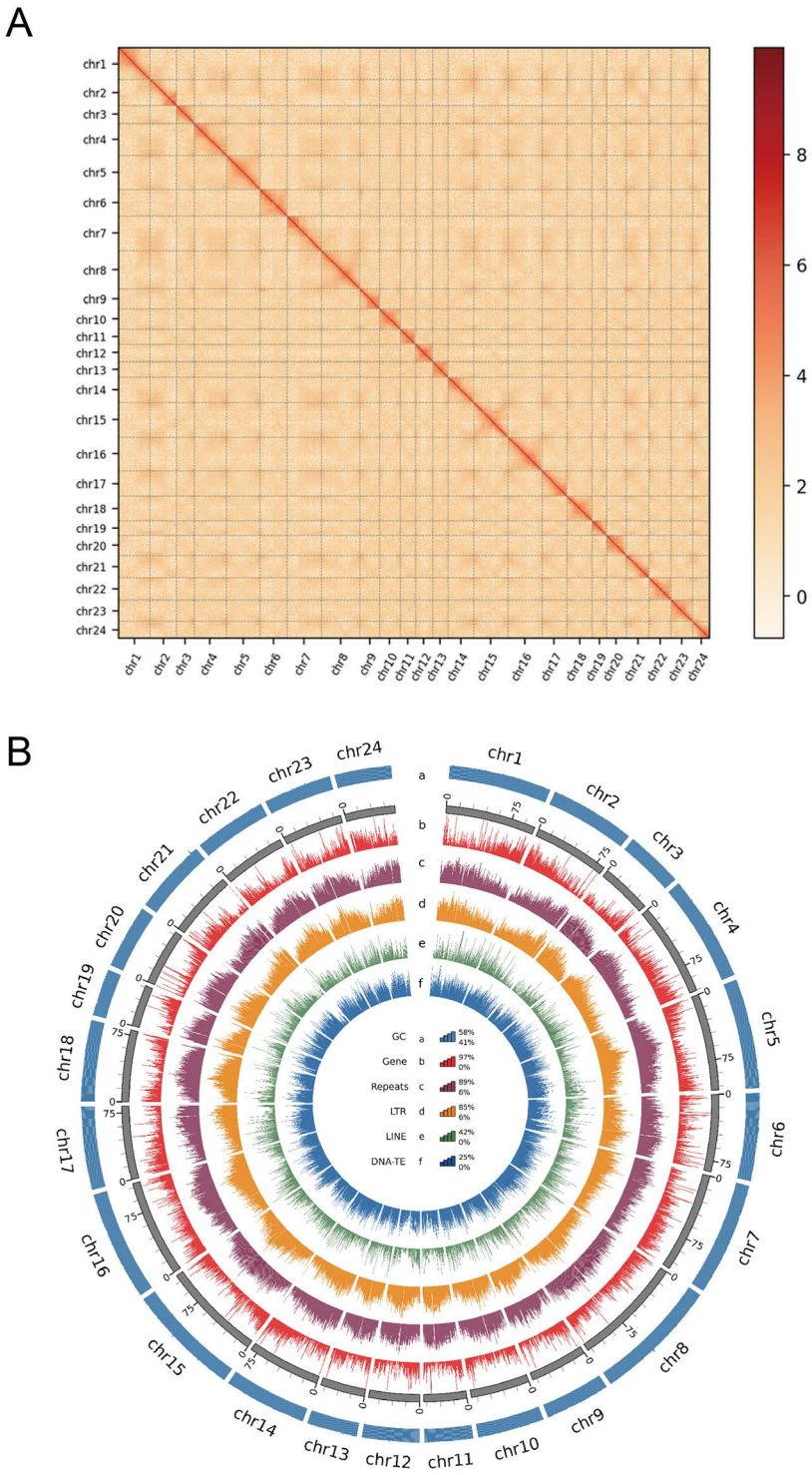

Fig. 2 The chromosome-level genome assembly results of Prevernalis. (**A**) Genome-wide chromatin interaction heatmap of Prevernalis based on Hi-C data. (**B**) Circos plot of distribution of the genomic elements in Prevernalis. From the outer circle to the inner circle: (a) GC content of the genome; (b) gene distribution; (c) Total repeat density distribution (d) long terminal repeats (LTR); (e) long interspersed nuclear elements (LINE); (f) DNA transposable elements.

Technical Validation

To evaluate the quality and completeness of the Prevernalis genome assembly, we took three approaches. First, we assessed the integrity of the genome using BUSCO[50] (Benchmarking Universal Single-Copy Orthologs), which is a quantitative evaluation software for the integrity of genome assembly using a orthologous database,

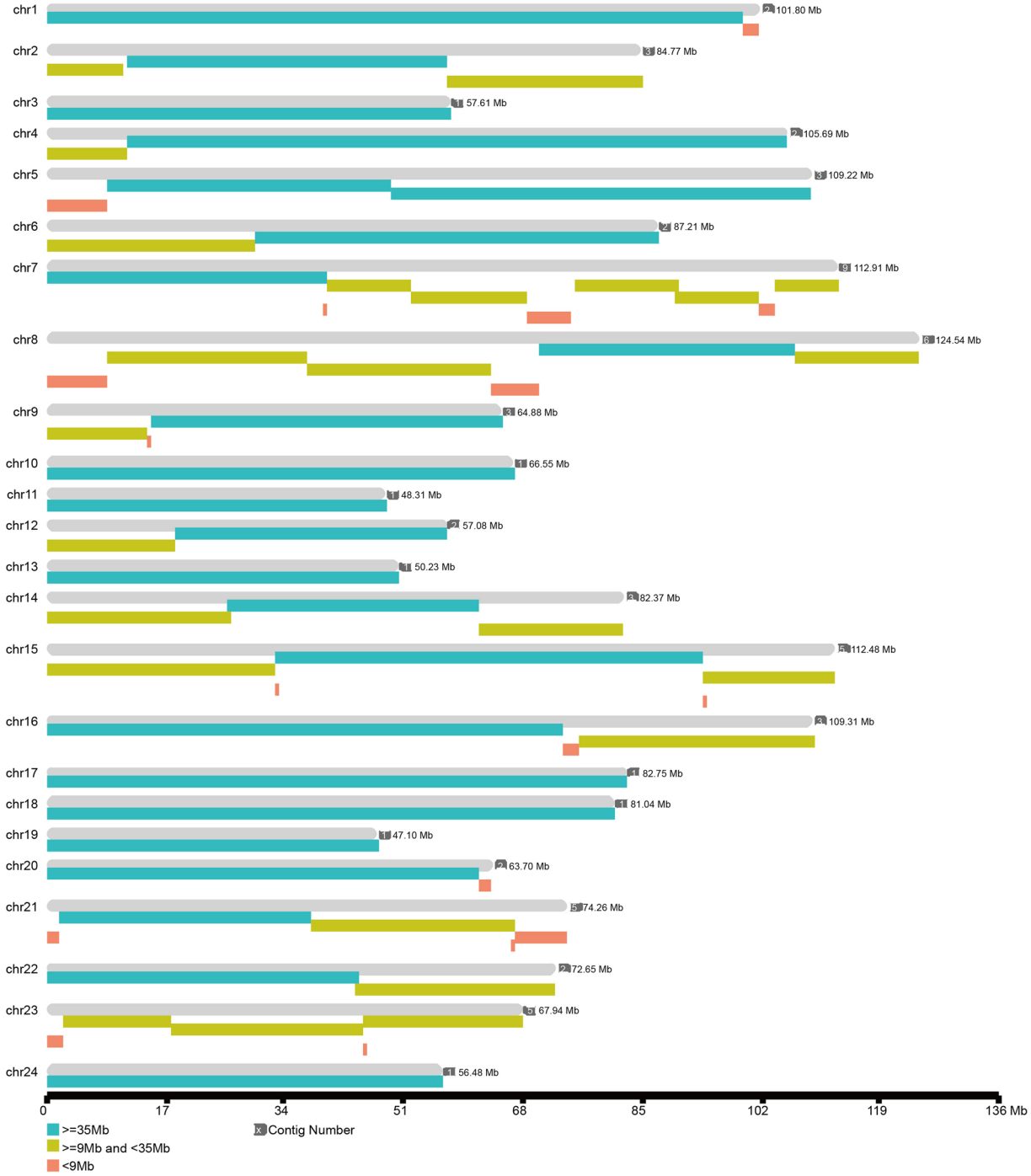

Fig. 3 Contigs distribution map for chromosomes of Prevernalis. The bars in grey represent entire lengths of chromosomes, and other colors represent contigs with different length.

OrthoDB. BUSCO analysis showed that among 4896 single copy homologues in the poales_odb10 database, 97.51% were complete genes (47.20% single copy genes, 50.31% repeat genes), 0.59% were fragmented and 1.9% were missing (Table 5), indicating that the assembly of the Prevernalis genome is remarkably complete. Secondly, the whole-genome high long terminal repeat (LTR) assembly index (LAI)[51] is an important indicator of genome assembly quality and completeness, and the LAI score for Prevernalis genome assembly was 20.03 using LTR_Finder[52] and LTR_retriever[53], indicating that the assembly quality of Prevernalis reached the reference genome level (Table S4). Moreover, BWA[54] and Minimap2[55] were used to aligned illumina paired-end clean reads and PacBio long reads to the final assembly Prevernalis genome, resulting in mapping rates of 99.53% and 100%, respectively. When aligning Illumina paired-end clean reads to the final assembly, the homozygous SNP rate and homozygous InDel rate were 0.001% and 0.001%, respectively, indicating the exceptional comprehensiveness of the complete Prevernalis genome.

Statistics of repetitive elements								
Type	**Repeat Size**	**% of genome**						
Trf	93808574	4.88						
Repeatmasker	585843336	30.5						
Proteinmask	304313851	15.84						
De novo	1013050177	52.74						
Total	1314944272	68.45						
Classifcation of repetitive elements								
Type	**Repbase TEs**		**TE protiens**		***De novo***		**Combined TEs**	
	Length (bp)	**% in genome**	**Length (bp)**	**% in genome**	**Length (bp)**	**% in genome**	**Length (bp)**	**% in genome**
DNA	117294853	6.11	39533057	2.06	150643017	7.84	208564069	10.86
LINE	24306287	1.27	12995635	0.68	19249313	1	30384437	1.58
SINE	251561	0.01	0	0	313415	0.02	422353	0.02
LTR	445406001	23.19	251797678	13.11	794547276	41.36	1014578269	52.82
Satellite	1369004	0.07	0	0	315999	0.02	1535040	0.08
Simple_repeat	0	0	0	0	446867	0.02	446867	0.02
Other	2020	0	0	0	0	0	2020	0
Unknown	308977	0.02	0	0	62277483	3.24	62581489	3.26
Total	585843336	30.5	304313851	15.84	1013050177	52.74	1285895312	66.94

Table 2. Statistics and classifcation of repetitive elements in the Prevernalis genome.

			Number	**Percent (%)**
Total			53558	100
	Annotated		50416	94.13
		NR	49946	93.26
		SwissProt	33079	61.76
		TrEMBL	50173	93.68
		KOG	38518	71.92
		TF	3652	6.82
		InterPro	43118	80.51
		GO	31236	58.32
		KEGG_ALL	49259	91.97
		KEGG_KO	17883	33.39
		Pfam	38550	71.98
	Unannotated		3142	5.87

Table 3. Prediction and annotation of protein-coding genes in the Prevernalis genome.

Type		**Copy**	**Average length (bp)**	**Total length (bp)**	**% of genome**
miRNA		593	111	65565	0.003413
tRNA		852	75	63902	0.003327
rRNA	rRNA	4844	158	764335	0.03979
	18S	104	1730	179947	0.009368
	28S	201	204	40922	0.00213
	5.8S	97	156	15129	0.000788
	5S	4442	119	528337	0.027505
snRNA	snRNA	1456	121	175858	0.009155
	CD-box	1106	114	126263	0.006573
	HACA-box	133	129	17114	0.000891
	splicing	217	150	32481	0.001691
	scaRNA	0	0	0	0

Table 4. Statistics and classification of non-coding RNAs identified in the Prevernalis genome.

Type	Proteins	Percentage (%)
Complete BUSCOs (C)	4,774	97.51
Complete and single-copy BUSCOs (S)	2,311	47.2
Complete and duplicated BUSCOs (D)	2,463	50.31
Fragmented BUSCOs (F)	29	0.59
Missing BUSCOs (M)	93	1.9
Total BUSCO groups searched	4,896	100

Table 5. Evaluate the quality and completeness of the Prevernalis genome assembly by BUSCO analysis.

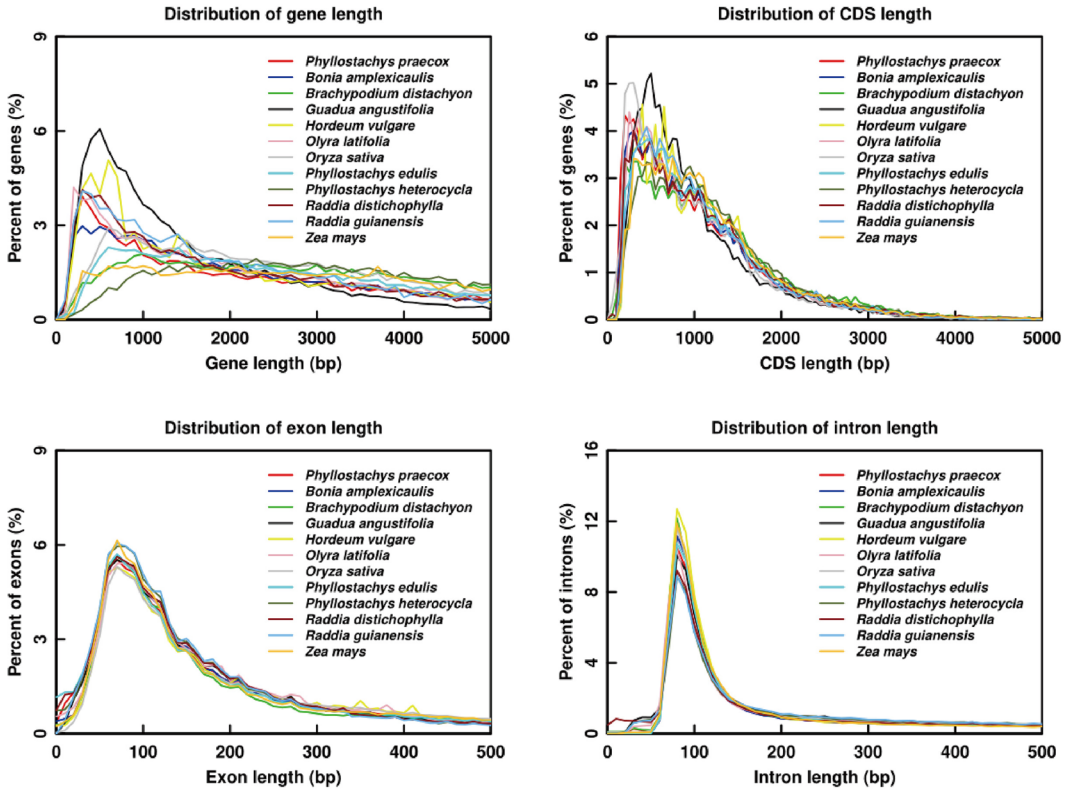

Fig. 4 Gene component of the genomes of Prevernalis and eleven other grass family species, namely *Raddia_distichophylla*, *Raddia_guianensis*, *Bonia_amplexicaulis*, *Phyllostachys_edulis*, *Olyra_latifolia*, *Phyllostachys_heterocycla*, *Oryza_sativa*, *Zea_mays*, *Brachypodium_distachyon*, *Guadua_angustifolia* and *Hordeum_vulgare*.

To validate gene annotation, we firstly studied the structure and number of genes in Prevernalis and eleven other grass family species, including *Raddia_distichophylla*, *Raddia_guianensis*, *Bonia_amplexicaulis*, *Phyllostachys_edulis*, *Olyra_latifolia*, *Phyllostachys_heterocycla*, *Oryza_sativa*, *Brachypodium_distachyon*, *Zea_mays, Guadua_angustifolia and Hordeum_vulgare*. The average lengths of transcripts, CDS, exons, and introns in Prevernalis and eleven other grass family species were found to be almost identical, and the average number of exons per gene was also nearly the same across all species (Table S3, Fig. 4). In addition, we also performed comparative genomic analysis and evolutionary analysis in Prevernalis and other seven species, including *Phyllostachys edulis* (*Ph.edulis*), *Broachypodium distachyon* (*B.distachyon*), *Oryza sativa* (*O.sativa*), *Sorghum bicolor* (*S.bicolor*), *Zea mays* (*Z.mays*), *Arabidopsis thaliana* (*A.thaliana*) and *Amborella trichopoda* (*A. trichopoda*). Through MAFFT[56], Gblocks[57] and RAxML[58] software, a phylogenetic tree was constructed using single-copy gene families, and mcmctree program in PAML[59] software package was used to estimate the differentiation time. Finally, it was found that Prevernalis and *Ph.edulis* had the closest relatives (Fig. 5). In addition, we used CAFE[60] to analyze the contraction and expansion of gene families of Prevernalis and seven other species, and 487 significantly expanded and 117 significantly contracted gene families were identified in Prevernalis. (Fig. 5). The significantly expanded gene family contained 2629 genes, and the the significantly contracted gene family contained 378 genes. Genes that had significantly expanded or contracted in Prevernalis were used for GO and KEGG annotation, respectively, and then, subjected to enrichment analysis using clusterProfiler in R (Table S5, S6, S7, S8, S9). Furthermore, BUSCO analysis showed that among 4896 single copy homologues in the poales_odb10 database, 98.2% were complete genes (45.4% single copy genes, 52.8% repeat genes), 0.7%

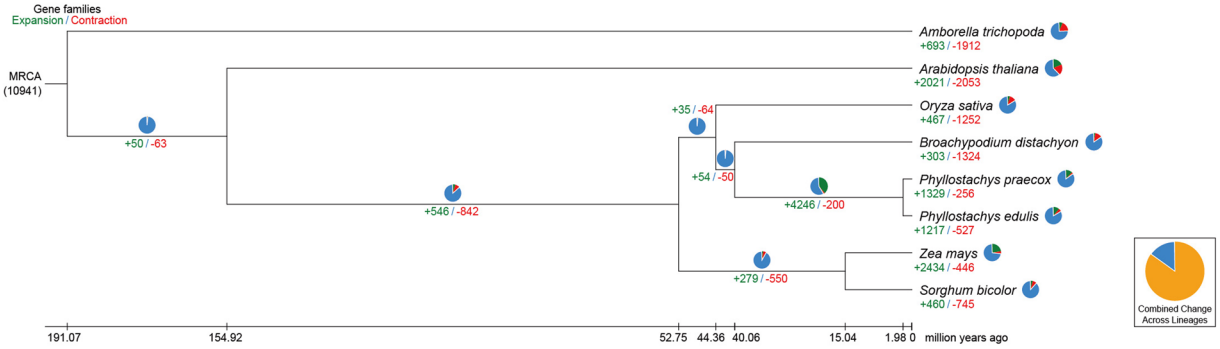

Fig. 5 Phylogenetic tree and gene families expansion and contraction of Prevernalis and seven other species. The scale at the bottom of the figure represents the divergence time.

were fragmented and 1.1% were missing (Table S10), indicating that the annotaion of the Prevernalis genome is remarkably accurate and complete.Taken together, we constructed a high-quality reference genome of Prevernalis and obtained accurate gene annotation.

Code availability

No dedicated code was developed for this study. The data analyses adhered to the manuals and protocols offered by the creators of the respective bioinformatics tools, as outlined in the methods section.

Received: 26 July 2024; Accepted: 29 January 2025;
Published online: 30 May 2025

References

1. Hou, D. *et al*. Integrative transcriptomic and metabolomic data provide insights into gene networks associated with lignification in postharvest Lei bamboo shoots under low temperature. *Food Chem.* **368**, 130822, https://doi.org/10.1016/j.foodchem.2021.130822 (2022).
2. Cai, K. *et al*. Development and characterization of EST-SSR markers from RNA-Seq data in *Phyllostachys violascens*. *Front. Plant Sci.* **10**, 50, https://doi.org/10.3389/fpls.2019.00050 (2019).
3. Nongdam, P. & Tikendra, L. The nutritional facts of bamboo shoots and their usage as important traditional foods of northeast India. *Int. Sch. Res. Notices* **2014**, 679073, https://doi.org/10.1155/2014/679073 (2014).
4. Jiao, Y. *et al*. Comparative transcriptomic analysis of the flower induction and development of the Lei bamboo (*Phyllostachys violascens*). *BMC Bioinformatics* **20**, 687, https://doi.org/10.1186/s12859-019-3261-z (2019).
5. Mbukwa, D., Gui, R. & Deng, S. Effects of aeration treatments on root and rhizome growth parameters of *Phyllostachys violascens* (Lei bamboo) under intensive cultivation: A field study. *Sci. Total Environ.* **900**, 165738, https://doi.org/10.1016/j.scitotenv.2023.165738 (2023).
6. Jiang, J. *et al*. Chromosomal-level genome and metabolome analyses of highly heterozygous allohexaploid *Dendrocalamus brandisii* elucidate shoot quality and developmental characteristics. *J. Integr. Plant Biol.* https://doi.org/10.1111/jipb.13592 (2023).
7. Sun, J. *et al*. Major chemical constituents of bamboo shoots (*Phyllostachys pubescens*): qualitative and quantitative research. *J. Agric. Food Chem.* **64**, 2498–2505, https://doi.org/10.1021/acs.jafc.5b05167 (2016).
8. Chen, M. *et al*. Rapid growth of Moso bamboo (Phyllostachys edulis): Cellular roadmaps, transcriptome dynamics, and environmental factors. *Plant Cell.* **34**, 3577–3610, https://doi.org/10.1093/plcell/koac193 (2022).
9. Yang, J. B. *et al*. Genetic structure and differentiation in *Dendrocalamus sinicus* (Poaceae: Bambusoideae) populations provide insight into evolutionary history and speciation of woody bamboos. *Sci. Rep.* **8**, 16933, https://doi.org/10.1038/s41598-018-35269-8 (2018).
10. Jin, G. *et al*. New genes interacted with recent whole-genome duplicates in the fast stem growth of bamboos. *Mol. Biol. Evol.* **38**, 5752–5768, https://doi.org/10.1093/molbev/msab288 (2021).
11. Guo, Z. H. *et al*. 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 (2019).
12. Peng, Z. *et al*. The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat. Genet.* **45**, 456–461, https://doi.org/10.1038/ng.2569 (2013).
13. Zhao, H. *et al*. Analysis of 427 genomes reveals moso bamboo population structure and genetic basis of property traits. *Nat. Commun.* **12**, 5466, https://doi.org/10.1038/s41467-021-25795-x (2021).
14. Zhao, H. *et al*. Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*). *GigaScience* **7**, https://doi.org/10.1093/gigascience/giy115 (2018).
15. Zheng, Y. *et al*. 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 (2022).
16. Ma, P. F. *et al*. Genome assemblies of 11 bamboo species highlight diversification induced by dynamic subgenome dominance. *Nat. Genet.*, https://doi.org/10.1038/s41588-024-01683-0 (2024).
17. Liu, Y. L. *et al*. BambooBase: A comprehensive database of bamboo omics and systematics. *Mol. Plant*, https://doi.org/10.1016/j.molp.2024.02.017 (2024).
18. Ye, S. *et al*. An efficient plant regeneration and transformation system of Ma Bamboo (*Dendrocalamus latiflorus Munro*) started from young shoot as explant. *Front. Plant Sci.* **8**, 1298, https://doi.org/10.3389/fpls.2017.01298 (2017).
19. Ye, S. *et al*. Robust CRISPR/Cas9 mediated genome editing and its application in manipulating plant height in the first generation of hexaploid Ma bamboo (*Dendrocalamus latiflorus Munro*). *Plant Biotechnol. J.* **18**, 1501–1503, https://doi.org/10.1111/pbi.13320 (2020).
20. Min, T., Hwarari, D., Li, D., Movahedi, A. & Yang, L. CRISPR-based genome editing and its applications in woody plants. *Int. J. Mol. Sci.* **23**, https://doi.org/10.3390/ijms231710175 (2022).
21. Wang, K. L. *et al*. MicroRNAs play important roles in regulating the rapid growth of the *Phyllostachys edulis* culm internode. *New Phytol.* **231**, 2215–2230, https://doi.org/10.1111/nph.17542 (2021).

22. Wang, N., Wang, W., Cheng, Y., Cai, C. & Zhu, Q. Uncovering the miRNA-mediated regulatory network involved in Ma bamboo (*Dendrocalamus latiflorus*) *de novo* shoot organogenesis. *Hortic. Res.* **10**, uhad223, https://doi.org/10.1093/hr/uhad223 (2023).
23. Bai, Y. *et al*. Phytohormone crosstalk of cytokinin biosynthesis and signaling family genes in Moso Bamboo (*Phyllostachys edulis*). *Int. J. Mol. Sci*. **24**, https://doi.org/10.3390/ijms241310863 (2023).
24. Li, L. *et al*. The association of hormone signalling genes, transcription and changes in shoot anatomy during moso bamboo growth. *Plant Biotechnol. J.* **16**, 72–85, https://doi.org/10.1111/pbi.12750 (2018).
25. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175, https://doi.org/10.1038/s41592-020-01056-5 (2021).
26. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460, https://doi.org/10.1186/s12859-018-2485-7 (2018).
27. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890, https://doi.org/10.1093/bioinformatics/bty560 (2018).
28. Wingett, S. *et al*. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* **4**, 1310, https://doi.org/10.12688/f1000research.7334.1 (2015).
29. Dudchenko, O. *et al*. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95, https://doi.org/10.1126/science.aal3327 (2017).
30. Durand, N. C. *et al*. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98, https://doi.org/10.1016/j.cels.2016.07.002 (2016).
31. Jurka, J. *et al*. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467, https://doi.org/10.1159/000084979 (2005).
32. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* Chapter 4, 4.10.11–14.10.14, https://doi.org/10.1002/0471250953.bi04.10s25 (2009).
33. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA*. **117**, 9451–9457, https://doi.org/10.1073/pnas.1921046117 (2020).
34. Edgar, R. C. & Myers, E. W. PILER: identification and classification of genomic repeats. *Bioinformatics* **21**(Suppl 1), i152–158, https://doi.org/10.1093/bioinformatics/bti1003 (2005).
35. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580, https://doi.org/10.1093/nar/27.2.573 (1999).
36. Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–439, https://doi.org/10.1093/nar/gkl200 (2006).
37. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879, https://doi.org/10.1093/bioinformatics/bth315 (2004).
38. Trapnell, C. *et al*. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515, https://doi.org/10.1038/nbt.1621 (2010).
39. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491, https://doi.org/10.1186/1471-2105-12-491 (2011).
40. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096, https://doi.org/10.1093/nar/gkab688 (2021).
41. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–d162, https://doi.org/10.1093/nar/gky1141 (2019).
42. Griffiths-Jones, S. *et al*. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–124,, https://doi.org/10.1093/nar/gki081 (2005).
43. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935, https://doi.org/10.1093/bioinformatics/btt509 (2013).
44. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.gca:GCA_044048585.1 (2024).
45. Wang, H. *et al*. Phyllostachys violascens cultivar Prevernalis, whole genome shotgun sequencing project. *GenBank* https://identifiers.org/ncbi/insdc:JBINAA010000000 (2024).
46. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR31132316 (2024).
47. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR31132317 (2024).
48. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR29138491 (2024).
49. Wang, H. *et al*. Phyllostachys violascens cultivar Prevernalis genome. *figshare* https://doi.org/10.6084/m9.figshare.27314118.v1 (2024).
50. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212, https://doi.org/10.1093/bioinformatics/btv351 (2015).
51. Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**, e126, https://doi.org/10.1093/nar/gky730 (2018).
52. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–268,, https://doi.org/10.1093/nar/gkm286 (2007).
53. Ou, S. & Jiang, N. LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422, https://doi.org/10.1104/pp.17.01310 (2018).
54. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).
55. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100, https://doi.org/10.1093/bioinformatics/bty191 (2018).
56. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066, https://doi.org/10.1093/nar/gkf436 (2002).
57. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577, https://doi.org/10.1080/10635150701472164 (2007).
58. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, https://doi.org/10.1093/bioinformatics/btu033 (2014).
59. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556, https://doi.org/10.1093/bioinformatics/13.5.555 (1997).
60. Mendes, F. K., Vanderpool, D., Fulton, B. & Hahn, M. W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518, https://doi.org/10.1093/bioinformatics/btaa1022 (2021).

Acknowledgements

This research was funded by Jiangxi Forestry S&T Innovation Special Project [2021]19, [2020]01, Jiangxi Academy of Forestry Basic Research and Talent Research Special Project (2022510902,2023510901), and the Project of Central Finance Forestry Development and Reform Special Project (2023,2024). We thank Wuhan Onemore-tech Co., Ltd. for their assistance with genome sequencing and analysis.

Author contributions

Haixia Wang conceived and designed the study. Hui Ding and Xianjiang Jin conducted the genome assembly and analysis. Haixia Wang, Hui Ding, Yanru Fu and Yizhen Peng wrote the paper. Ping Cheng revised the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-04556-1.

Correspondence and requests for materials should be addressed to H.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.