JIPB Journal of Integrative Plant Biology

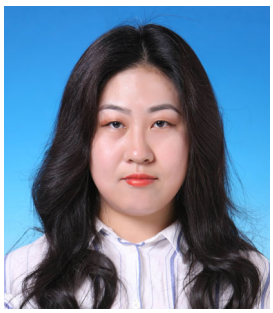
Research Article
https://doi.org/10.1111/jipb.13664

Establishment of genome-editing system and assembly of a near-complete genome in broomcorn millet[oo]

Yang Liu[1†], Zixiang Cheng[2†], Weiyao Chen[1,3], Chuanyin Wu[2], Jinfeng Chen[1*] and Yi Sui[2*]

1. State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, the Chinese Academy of Sciences, Beijing 100101, China
2. State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
3. University of Chinese Academy of Sciences, Beijing 100049, China
†These authors contributed equally to this work.
*Correspondence: Jinfeng Chen (chenjinfeng@ioz.ac.cn); Yi Sui (suiyi@caas.cn, Dr. Sui is fully responsible for the distributions of all materials associated with this article)

Yang Liu



Yi Sui

ABSTRACT

The ancient crop broomcorn millet (*Panicum miliaceum* L.) is an indispensable orphan crop in semi-arid regions due to its short life cycle and excellent abiotic stress tolerance. These advantages make it an important alternative crop to increase food security and achieve the goal of zero hunger, particularly in light of the uncertainty of global climate change. However, functional genomic and biotechnological research in broomcorn millet has been hampered due to a lack of genetic tools such as transformation and genome-editing techniques. Here, we successfully performed genome editing of broomcorn millet. We identified an elite variety, Hongmi, that produces embryogenic callus and has high shoot regeneration ability in *in vitro* culture. We established an *Agrobacterium tumefaciens*-mediated genetic transformation protocol and a clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-mediated genome-editing system for Hongmi. Using these techniques, we produced herbicide-resistant transgenic plants and edited *phytoene desaturase* (*PmPDS*), which is involved in chlorophyll biosynthesis. To facilitate the rapid adoption of Hongmi as a model line for broomcorn millet research, we assembled a near-complete genome sequence of Hongmi and comprehensively annotated its genome. Together, our results open the door to improving broomcorn millet using biotechnology.

Keywords: broomcorn millet, CRISPR/Cas9 editing, genome assembly, herbicide resistance, transformation system

Liu, Y., Cheng, Z., Chen, W., Wu, C., Chen, J., and Sui, Y. (2024). Establishment of genome-editing system and assembly of a near-complete genome in broomcorn millet. J. Integr. Plant Biol. **00:** 1–15.

INTRODUCTION

Orphan crops have been underutilized in modern agriculture and are cultivated in a limited number of regions (Ye and Fan, 2021). Compared with staple crops, orphan crops possess many beneficial traits, such as being more nutritious and well adapted to unfavorable environments, and therefore have great potential for sustainable agriculture (Yaqoob et al., 2023). Broomcorn millet (*Panicum miliaceum* L.), which originated and was domesticated in Northern China approximately 10,000 years ago (Lu et al., 2009; Hunt et al., 2018), is a valuable orphan crop that contributes to

food security in semi-arid areas. This fast-growing C4 plant has strong resilience to various abiotic stresses such as drought, thus performing better than staple crops and requiring lower inputs in agricultural management (Ghannoum et al., 2002; Seghatoleslami et al., 2008; Goron and Raizada, 2015; Liu et al., 2015; Washburn et al., 2015). Broomcorn millet is considered to be one of the Future Smart Foods (FSF), with the potential to combat hunger and malnutrition due to its high nutrient density and economic potential (Li and Siddique, 2020). Despite possessing these properties, genetics and genomics research on broomcorn millet is behind that of other crops. As a result, progress in functional genomics studies and variety improvement of broomcorn millet through biotechnology has been slow and needs to be accelerated.

Although conventional breeding has greatly contributed to crop improvement, it is a tedious process, often involving crossing, backcrossing, and selfing for many generations, accompanied by selection at each generation (Marone et al., 2023). Marker-assisted breeding increases selection efficiency, but the entire process is still time-consuming, and it usually takes several years to develop a commercial variety (Marone et al., 2023). Genetic engineering techniques have been evolving in recent decades, accelerating the improvement of crop varieties (Kumar et al., 2020). For example, transgenic approaches are now widely used to validate gene function in model species and to improve traits (such as herbicide tolerance and insect resistance) in crops (Kumar et al., 2020). Introducing foreign genes into the genomes of crop plants using transgenic approaches avoids many disadvantages of conventional breeding. However, this technology also faces challenges, as introducing novel genetic material may have unforeseen effects on the ecosystem. Therefore, in many countries, these approaches are subject to strict biosafety regulations for genetically modified organisms (GMOs) (Buchholzer and Frommer, 2023; Salt, 2023). By contrast, genome-editing technologies precisely and specifically introduce genetic variations into the target genes or promoter sequences *in vivo* (Chen and Gao, 2014; Kamburova et al., 2017; Puchta, 2017; Cable et al., 2021). Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated nuclease (Cas)-mediated genome editing targets a specific sequence in the genome to produce chromosomal breaks. The imperfect repair of the breaks generates sequence mutations at the target site (Zhang et al., 2020). Although this tool transiently introduces transgenes into the plant, they can be removed by backcrossing to the parent, resulting in transgene-free progeny (Buchholzer and Frommer, 2023). Many countries have recently introduced looser guidelines for the commercialization of genome-edited plants whose transgenes have been removed (Buchholzer and Frommer, 2023), thus clearing the way for crop improvement via genome editing. Therefore, a genome-editing system for broomcorn millet is urgently needed.

Although transgenic and genome-editing technologies have great potential to accelerate functional genomics and trait improvement in orphan crops, the low efficiency and genotype dependence of genetic transformation, as well as the lack of high-quality reference genomes and annotation information, have hindered their practical application (Anjanappa and Gruissem, 2021; Chen et al., 2022; Lee and Wang, 2023). These problems are especially pronounced for broomcorn millet. No efficient genetic transformation system is currently available for broomcorn millet and genome-editing technologies, such as CRISPR/Cas9 and its derivatives, have not yet been established for this orphan crop. Moreover, the genomes of several broomcorn millet accessions have been sequenced, including Longmi4 (longmi), Pm_0390, and Jinshu7. However, these genome assemblies have many gaps because broomcorn millet is an allotetraploid plant with numerous repetitive sequences, making sequence assembly quite challenging (Shi et al., 2019; Zou et al., 2019; Sun et al., 2023; Wang et al., 2023). A high-quality reference genome sequence could provide comprehensive, precise information (e.g., genomic variations) about the domestication and agricultural traits of a crop (Yaqoob et al., 2023). Therefore, developing an effective genetic transformation system and improving the continuity and quality of the reference genome of broomcorn millet are essential ways for facilitating genetic and functional studies and for breeding, which could increase the diversity of agriculture and contribute to global food security (Aganezov et al., 2022; Wang et al., 2022).

Here, we screened 13 broomcorn millet varieties for their ability in embryogenic callus induction and shoot regeneration, and identified Hongmi as a promising candidate for genetic transformation. We established an optimized *Agrobacterium tumefaciens* (*Agrobacterium*)-mediated genetic transformation protocol and a CRISPR/Cas9 genome-editing system for Hongmi. Moreover, we *de novo* assembled a near-complete genome of Hongmi, leading to improved performance in genomic analyses. These endeavors hold the potential to expedite functional genomics research and genetic enhancement efforts in broomcorn millet.

RESULTS

Identification of a genotype suitable for genetic transformation

Examining the genotype-dependent responses of plants to tissue culture is crucial for determining their suitability for genetic transformation. We therefore evaluated mature seeds from 13 broomcorn millet accessions, provided by the National GeneBank of the Institute of Crop Science, Chinese Academy of Agricultural Sciences (Table S1), as explants for callus induction. We incubated surface-sterilized mature seeds on a callus-inducing medium (CIM) for 20 d to initiate callus formation (Figure 1). All accessions were able to produce primary callus, although with variable frequencies of induction. Accessions Hongmi, Qingyanghongying, and Jinshu7 showed the highest induction rates (Figure S1). To

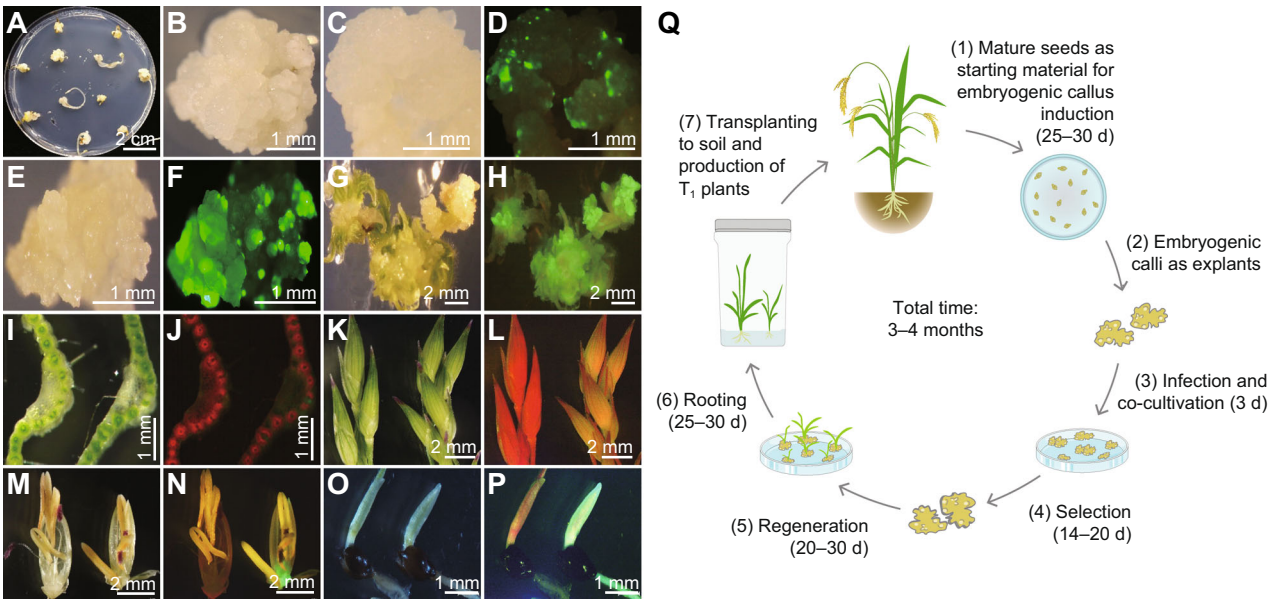


Figure 1. *Agrobacterium*-mediated transformation of broomcorn millet

(A) Primary calli induced from mature seeds of Hongmi, 20 d after inoculation. **(B)** Embryogenic callus (typical cream color and smooth surface) suitable for transformation at 30 d after seed inoculation. **(C)** Callus under bright light, photographed after 3 d of co-cultivation with *Agrobacterium*. **(D)** Highly efficient transient expression of the green fluorescent protein (GFP), as visualized under UV light, following *Agrobacterium* infection. **(E)** Appearance of callus under bright light after 10 d of the first selection step. **(F)** The same callus as in (E) but visualized under UV light, showing strong GFP signals in the protruding sectors. **(G, H)** Shoot regeneration from transgenic calli visualized under bright (G) or UV light (H). **(I, J)** Leaf cross-sections showing GFP signals in a transgenic T_0 plant (right). The leaf sections were photographed under bright (I) or UV light (J); a wild-type (WT) leaf section (left) was used as a control for comparison. **(K, L)** GFP-expressing spikelets imaged under bright (K) or UV (L) light, compared to no GFP expression in the WT (left). **(M, N)** An open floret with GFP expression (right) compared to a WT sample (left), as observed under bright (M) or UV light (N). **(O, P)** Germinating T_1 seeds were observed under bright (O) or UV (P) light. Positive or negative GFP expression among seedlings is an indication of transgene segregation. **(Q)** Schematic workflow of *Agrobacterium*-mediated transformation of Hongmi. Bars, 2 cm in A, 1 mm in B–F, I, J, O, and P, 2 mm in G, H, and K–N.

track callus proliferation and embryogenic callus growth, we subcultured primary calli on the same CIM for an additional 10–20 d. Most of the induced calli continued to swell and grow to become friable calli (Type I), whereas a few calli gave rise to compact, granular, cream-yellow embryogenic calli with smooth surfaces (Type II, Figure 1B). We quantified the proportions of embryogenic calli originating from primary calli and rated the quality of embryogenic calli into five categories based on texture, color, and degree of looseness. Hongmi and Qingyanghongying produced superior embryogenic calli, with embryogenic callus-induction rates of 56.14% (96/171) and 50.39% (64/127), respectively (Figure S2). We selected typical Type II calli from all accessions and transferred them onto the shoot regeneration medium (SRM) (Figure S3). While most accessions showed very low rates of shoot regeneration, Hongmi (21.11%, 19/90), Taiyuan1098 (17.78%, 16/90), Taiyuan33 (13.33%, 12/90), and Jinshu7 (11.11%, 10/90) had relatively high green shoot ratios (Figure S2). Therefore, we selected Hongmi to develop an *Agrobacterium*-mediated genetic transformation protocol based on its callus quality and shoot regeneration capacity.

Development of an *Agrobacterium*-mediated genetic transformation protocol for Hongmi

We aimed to develop an efficient, reliable, reproducible *Agrobacterium*-mediated genetic transformation protocol for broomcorn millet using mature seeds as the initial material for callus induction. The selection of high-quality starting callus is critical for successful transformation. Therefore, we selected cream-colored calli with a smooth surface and high density among calli that had been subcultured for 20–30 d (Figure 1A) and subcultured them for an additional 7–14 d (Figure 1B) before *Agrobacterium* infection. We performed infection experiments using *Agrobacterium* strain EHA105 harboring the pCAMBIA1305-GFP vector, which carries the green fluorescent protein (GFP) reporter gene (Yang et al., 2020). After 3 d of co-cultivation, multiple GFP spots were observed (Figure 1C, D) in typical Type II calli (Figure 1B), indicating that the infection process was efficient. These fluorescent spots became larger due to continued GFP expression in a selection medium. The subsequent survival of the GFP-expressing sectors following the selection on medium containing hygromycin (30 mg/L) confirmed the stable integration of the *GFP* transgene into the genome (Figure 1E, F). We transferred GFP-expressing calli to SRM and cultured them for 7–14 d, leading to successful regeneration (Figures 1G, H, S4A). Then, the GFP-positive transgenic plants were successfully rooted on rooting medium (RM) acclimatized to soil, and grown to maturity (Figure S4B, C).

We obtained 25 T_0 plants in four experiments, with average transformation and regeneration efficiencies of 16.7% and 47.7%, respectively (Figure S4D). Of the 24 surviving plants,

23 were verified to be transgenic by PCR. For further confirmation, we visualized GFP signals in leaf cross-sections (Figure 1I, J), panicles (Figure 1K, L), and florets (Figure 1M, N). The entire *Agrobacterium*-mediated transformation process took 3–4 months (Figure 1Q). To examine the copy number of the *hygromycin phosphotransferase* (*HPT*) transgene, we randomly selected ten T_0 plants and performed Southern blotting with a probe for the *HPT* selectable marker using two restriction enzymes (*Sac*I and *Bam*HI). Seven T_0 plants contained a single copy of *HPT* (PM1, PM11, PM15, PM19, PM4, PM7, and PM8), while PM2 and PM17 carried two and three copies, respectively (Figure S5). Among the T_0 plants with single-copy genes, PM4, PM7, and PM8 might have been derived from the same event (Figure S5). We also observed GFP expression in dry and germinating T_1 seeds (Figures 1O, P, S6A). The segregation ratio of the GFP expression in the T_1 progeny also confirmed the copy number of the transgene in the corresponding T_0 plants (Figure S6B).

To test the reliability of our *Agrobacterium*-mediated genetic transformation protocol, we introduced an herbicide-resistance gene into Hongmi to create herbicide-resistant broomcorn millet. The *bialaphos resistance* (*BAR*) or *phosphinothricin acetyltransferase* (*PAT*) genes are widely used to develop glufosinate-resistant crops using transgenic approaches (Cable et al., 2021). We generated a construct harboring *BAR* driven by the CaMV35S promoter and introduced it into Hongmi via *Agrobacterium*-mediated transformation, followed by selection on medium containing 30 μM bialaphos to screen for glufosinate resistance. We identified ten putative T_0 transformants using a BAR test strip and transplanted them into pots to produce T_1 seeds (Figure 2A–C). The T_1 transgenic plants (the event S352-1) exhibited normal growth, with an approximately 80% survival rate, on ½MS medium containing 30 μM bialaphos, whereas wild-type (WT) seeds failed to germinate (Figure 2D, E). These results provide compelling evidence that we successfully developed herbicide-resistant broomcorn millet, highlighting the utility of our transformation system.

Development of a CRISPR/Cas9-mediated genome-editing system in broomcorn millet

An efficient and stable genetic transformation technique is required to establish a genome-editing system. We therefore developed a CRISPR/Cas9-mediated genome-editing system based on our transformation protocol optimized for Hongmi. The *phytoene desaturase* (*PDS*) gene (Banakar et al., 2020), which is essential for chlorophyll biosynthesis, is often targeted in genome editing studies due to the visual albino phenotype conferred by its mutation. We utilized a CRISPR/Cas9 vector containing *Cas9* under the control of the maize *Ubiquitin* promoter and a single-guide RNA (sgRNA) targeting the fifth exon of the *PmPDS* gene driven by the rice *U3* promoters (*OsU3*) or the maize *U6* promoter (*ZmU6*) (Figure 3A, top). Considering the tetraploid nature of broomcorn millet, we designed the sgRNA to target a conserved sequence in *PmPDS* in both subgenomes A (*PmPDS-A*) and B (*PmPDS-B*). *PmPDS* in the two subgenomes can be distinguished by the presence of a specific single nucleotide polymorphism (SNPs; ATC for *PmPDS-A* and GTA for *PmPDS-B*), allowing allele-specific mutagenesis to be detected by Sanger sequencing (Figure 3A, bottom).

We used our *Agrobacterium*-mediated transformation protocol to transfer the CRISPR/Cas9 vector into Hongmi calli and obtained 23 transgene-positive T_0 plants using the *OsU3* promoter for sgRNA expression and 25 T_0 plants using the *ZmU6* promoter. Among the 23 T_0 plants carrying the *OsU3* promoter construct, no mutation was observed. However, sequencing of the 25 T_0 plants carrying the *ZmU6* promoter construct revealed that 11 plants exhibited genome editing at the *PmPDS* locus, resulting in an editing efficiency of 44.0%. Of these 11 edited plants, nine were complete-knockout mutants, while two plants contained one homeolog that was not edited or heterozygous, resulting in a knockout efficiency of 36.0% (Figure 3A, B; Table S2). The *pmpds* knockout plants had an albino phenotype, providing visible confirmation of their editing (Figure 3C). In summary, we successfully developed an efficient CRISPR/Cas9 genome-editing system for Hongmi. This breakthrough lays a solid foundation for functional genomics studies and precision breeding of this crop.

Genome assembly and annotation of the Hongmi genome

To construct a high-quality genome assembly for Hongmi, we generated ~38.0 Gb Oxford Nanopore Technology (ONT) ultra-long reads and ~22.6 Gb PacBio high-fidelity (HiFi) long reads, covering ~45× and ~27× of the Hongmi genome, respectively (Table S3). The N50 of the ONT and HiFi reads were 59.4 and 15.8 kb, respectively, covering the length of most repetitive elements (Table S3). We assembled the ONT ultra-long reads using NextDenovo, resulting in 33 contigs with an N50 of 42.5 Mb (Table 1). We then assembled the PacBio HiFi reads using hifiasm (Cheng et al., 2021), resulting in 1,422 contigs with an N50 of 23.9 Mb (Table 1). The contigs assembled from ONT and HiFi reads were separately anchored, oriented, and manually curated into 18 pseudochromosomes using Hi-C data with ALLHiC (Zhang et al., 2019), resulting in two assemblies: the ONT assembly and the HiFi assembly. The anchoring rate of ONT and HiFi contigs was 99.9% and 88.9%, respectively, with 13 gaps in the ONT assembly and 53 gaps in the HiFi assembly.

Nine chromosomes in the ONT assembly (chromosomes 1, 3, 5, 7, 9, 11, 14, 15, and 16) contained no gaps, highlighting the high continuity of the ONT assembly. We filled four sequence gaps in the ONT assembly using long reads spanning the gaps and three gaps using locally assembled contigs (Kolmogorov et al., 2019). To ensure the highest base-pair accuracy, we replaced the low-accuracy ONT contigs with the HiFi assembly. The size of the final assembled genome is 831 Mb (Figure 4A; Table 1), which is slightly smaller than that of Longmi_v2 (837 Mb) and larger than those of Pm0390 (822 Mb) and Jinshu7 (804 Mb).

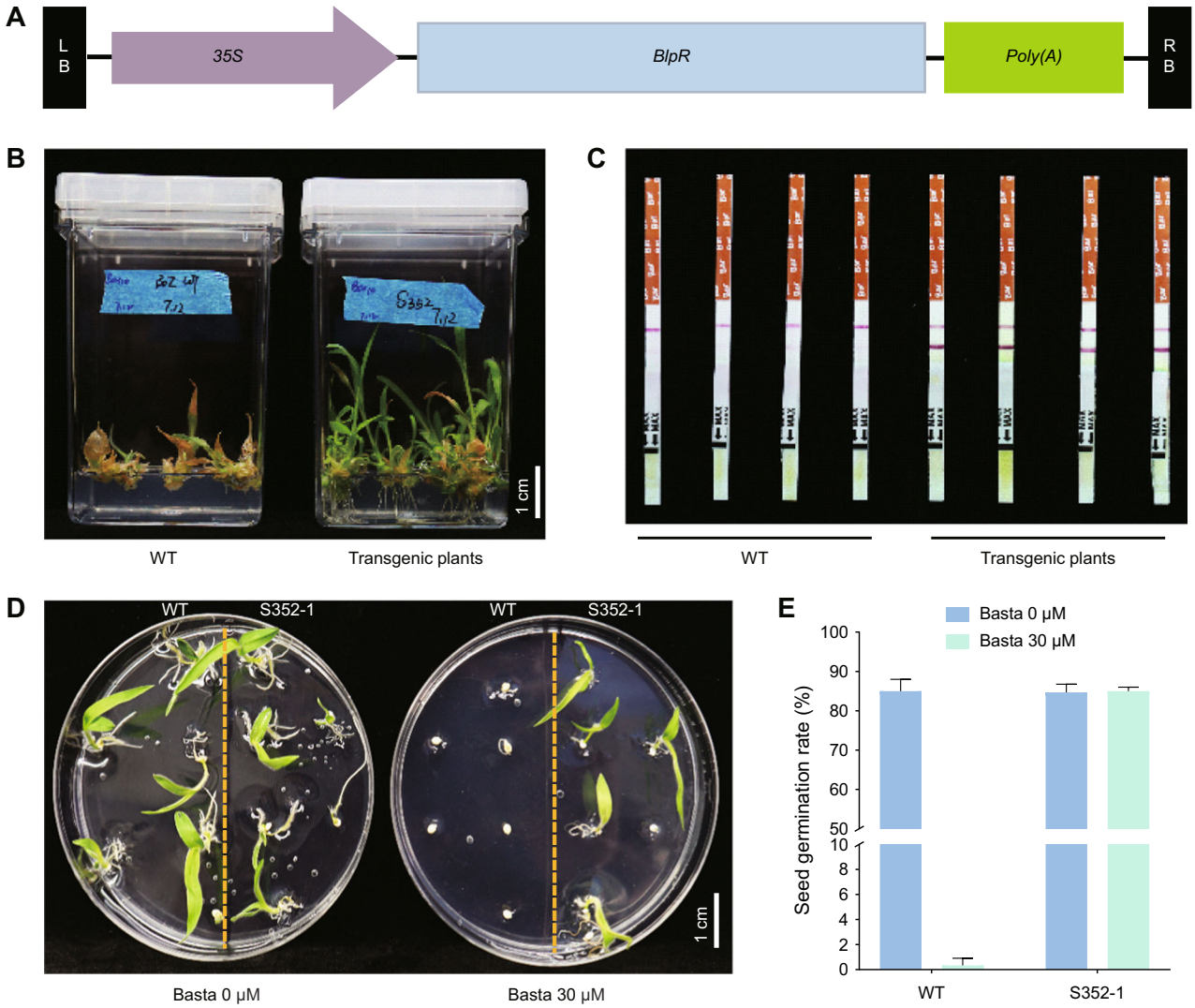


Figure 2. Development of herbicide resistance in broomcorn millet by genetic transformation

(A) Schematic illustration of the vector containing the *BAR* (*BlpR*) gene for herbicide resistance. **(B)** Transgenic T_0 plants on rooting medium containing 30 μM bialaphos. Note the sensitivity of non-transformed wild-type (WT) shoots and roots to the herbicide on the same medium. **(C)** Rapid confirmation of transgenic plants by a PAT/BAR strip test. The presence of the lower band indicates transgene-positive plants. **(D)** Transgenic T_1 seedlings are healthy when grown on ½MS medium containing 30 μM bialaphos, in contrast with the failed germination of WT seeds under the same conditions. **(E)** Seed germination rates of WT and transgenic S352-1 plants on medium with 0 or 30 μM bialaphos.

Furthermore, we identified ten telomeres using seven-base telomere repeats (5′-TTTAGGG-3′) as sequence queries (Figures 4B, S7; Table S4). In addition, we identified candidate centromeric tandem repeats using TRF (Benson, 1999) and HmmerSearch. The centromeric regions of the chromosomes are complete except for chromosomes 6 and 7, each with one gap in the centromeric region (Figure S8).

To assess the quality of the assembly, we evaluated the completeness of the assembly with 1,614 Benchmarking Universal Single Copy Orthologs (BUSCO) genes from Embryophyta (odb10), and found that 98.9% of intact genes were included in the assembly (Table S5). We also estimated the genome quality using the LTR assembly index (LAI) (Ou and Chen, and Jiang, 2018), which yielded a value of 17.94, indicating that the assembly has high contiguity and quality. We mapped PacBio HiFi reads, Illumina reads, and RNA-seq reads to the assembly, resulting in mean mapping rates of 99.15%, 96.69%, and 97.13%, respectively. Further evaluation using Merqury (Rhie et al., 2020) showed that the mean base-pair quality value (QV) of the assembly was 42.0, indicating high assembly accuracy (Table S6). Additionally, the Hi-C interaction matrices showed that the assembly had high consistency (Figures 4C, S9). Finally, we manually checked 35 large structural variations (SVs) between the Hongmi and Longmi_v2 assemblies and found that over half (18/35) of these SVs were false assemblies in Longmi_v2, which were corrected in the Hongmi assembly (Figure S10; Table S7). Overall, the Hongmi assembly represents the most

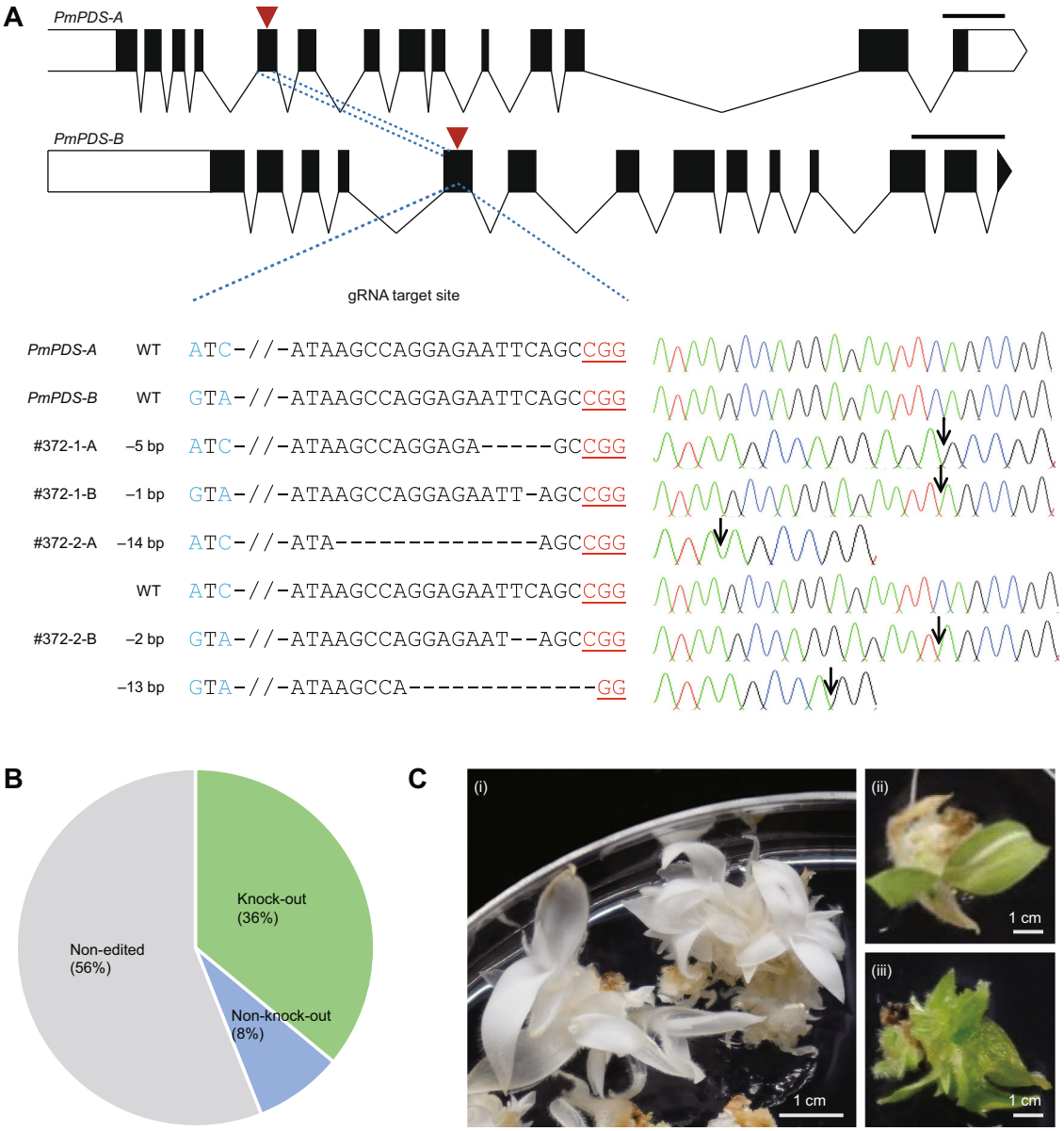


Figure 3. Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-mediated genome editing in broomcorn millet
(A) Genomic structure of *PmPDS* in subgenomes A and B and targeted mutagenesis. The target sites in both subgenomes are marked by red arrowheads, and the protospacer adjacent motif (PAM) sites are highlighted in red. Deletions within the targets are represented by dashes, and nucleotide polymorphisms (SNPs) differentiating the two subgenomes are shown in blue. The right panel shows partial Sanger sequencing chromatograms of edited *pmpds* alleles. The arrows indicate the edited positions. Scale bar corresponds to 500 bp. **(B)** Editing efficiency of *PmPDS* in T_0 plants. **(C)** Phenotypes of *PmPDS* knockout (i), chimeric (ii), and WT (iii) T_0 plants. Bars, 1 cm.

contiguous and complete assembly of the broomcorn millet genome to date.

The Hongmi genome contains 478.0 Mb of repetitive sequences, accounting for 57.5% of the genome, including 45.6% LTR retrotransposons and 5.3% DNA transposons (Table S8), which is similar to that of the Longmi, Jinshu7, and Pm0390 genomes (Shi et al., 2019; Zou et al., 2019; Sun et al., 2023; Wang et al., 2023). We used the repeat-masked genome to predict gene structure by combining *ab initio* gene predictions, transcriptome data, and homologous protein evidence using MAKER2 (Holt and Yandell, 2011). A total of 60,163 protein-coding genes (75,101 transcripts) were annotated in the Hongmi genome, 99.7% of which were located on the 18 chromosomes. We evaluated the completeness of gene annotation using BUSCO and revealed that 95.6% (1,543/1,614) of intact BUSCO genes were included in the annotated gene set (Table S9). Of all the annotated genes, 68.9% have functional annotations in the InterPro database, and 66.0% contain Pfam domains (Table S10).

Table 1. Characteristics of Hongmi, Longmi, Pm0390, and Jinshu7 genome assemblies

	Hongmi	Longmi_v2	Longmi	Pm0390	Jinshu7
Sequencing					
Platform	Nanopore; PacBio HiFi	PacBio HiFi	PacBio	PacBio	PacBio
Read size (depth)	38.0 Gb (45×); 22.6 Gb (27×)	42.3 Gb (51×)	150.7 Gb (170×)	81.0 Gb (88×)	86.0 Gb (97×)
Read N50	59.4 kb; 15.8 kb	15.6 kb	12.6 kb	–	11.2 kb
Assembly					
Contig number	55; 1,422	1,260	1,308	5,541	2,414
Contig size	832.1 Mb; 931.9 Mb	846.0 Mb	838.9 Mb	839.0 Mb	840.3 Mb
Contig N50	42.6 Mb; 23.9 Mb	26.2 Mb	2.6 Mb	368.6 kb	1.1 Mb
Final assembly size (bp)	832,397,787	846,020,423	848,436,801	854,674,422	862,020,904
Chromosome size (bp)	831,478,740	837,228,188	838,873,930	822,124,240	804,026,145
Gap number	6	260	829	4,232	2,277
Complete BUSCO	98.9%	98.9%	98.8%	98.7%	98.7%
LAI index	17.9	20.8	15.9	10.0	11.0
Merqury					
Kmer completeness	96.1207	98.9817	99.0888	97.8385	97.4693
QV	41.9651	42.5554	47.1215	40.5183	41.0501
Accuracy	0.999936395	0.999944479	0.999980598	0.99991125	0.999921478
Error rate	6.36E-05	5.55E-05	1.94E-05	8.87E-05	7.85E-05
Annotation					
Gene number	60,163	60,096	63,671	55,930	61,782
Complete BUSCO	95.6%	96.9%	83.7%	94.7%	98.2%

The Hongmi genome facilitates analysis of genetic variation in broomcorn millet

A high-quality reference genome is crucial for identifying genetic variations when performing population genetic analyses or studying domestication and agricultural traits (Deng et al., 2022). To investigate the efficacy of the Hongmi genome and other available broomcorn millet genomes in variant identification, we used resequencing data from a population of 366 individuals (Chen et al., 2023), including 58 wild and 308 cultivated accessions, to assess the efficiency of the genomes for short-read mapping and variation calling across populations.

We mapped the Illumina reads of 366 accessions to the Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7 genomes using BWA-MEM (Li, 2013). More properly paired reads were mapped to the Hongmi genome than to the other genomes (97.11%, 97.02%, 96.71%, 96.41%, and 95.34% properly paired reads for Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7, respectively) (Figure S11; Table S11). Moreover, even though more reads were mapped to the Hongmi genome, the mismatch rate of mapped reads was the lowest in Hongmi (7.99×10^{-3}, 8.33×10^{-3}, 8.74×10^{-3}, 9.28×10^{-3}, and 1.01×10^{-2} in Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7, respectively), suggesting the Hongmi genome has the highest sequence accuracy (Figure 5A; Table S11). Additionally, when considering the diversity of samples in the populations, Hongmi yielded the lowest the number of read pairs that were misoriented (36,253, 67,608, 80,869, 82,922, and 123,396 in Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7, respectively) and those on different chromosomes (862,081, 897,664, 1,022,075, 1,039,106, and 1,336,311 in Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7, respectively) and the lowest standard deviation of read coverage (93.13, 125.43, 127.81, 129.80, and 130.18 in Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7, respectively) (Figures 5B, S11E, F; Table S11). Taken together, these results demonstrate the power of the assembled Hongmi genome as a reference for short-read-based population-level studies.

We also compared the performance of Hongmi and other broomcorn millet genomes as references for long-read mapping. We mapped PacBio HiFi long reads from 32 accessions (Chen et al., 2023) using two aligners, minimap2 (Li, 2018) and winnowmap (Jain et al., 2022). There were no significant differences in the number of mapped reads among these reference genomes (Figure S12A; Table S11). However, the mismatch rate (0.01650, 0.01961, 0.02012, 0.01956, and 0.02574 with the genome of Hongmi, Longmi_v2, Longmi, Pm0390 and Jinshu7, respectively) and the read number in the supplementary alignments (393,520, 600,976, 761,660, 617,865, and 950,746 in Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7, respectively) were lowest when we used the Hongmi genome as the reference (Figures 5C, S12B; Table S11). In addition, there was a significant reduction in the standard deviation of mapping coverage when reads were aligned to the Hongmi genome (133.6, 257.1, 258.7, 273.1, 266.1 with Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7, respectively) (Figure 5D; Table S11), indicating that the alignments were more evenly distributed across the genome. These results indicate that the Hongmi genome performed best when used as a reference in long-read mapping.

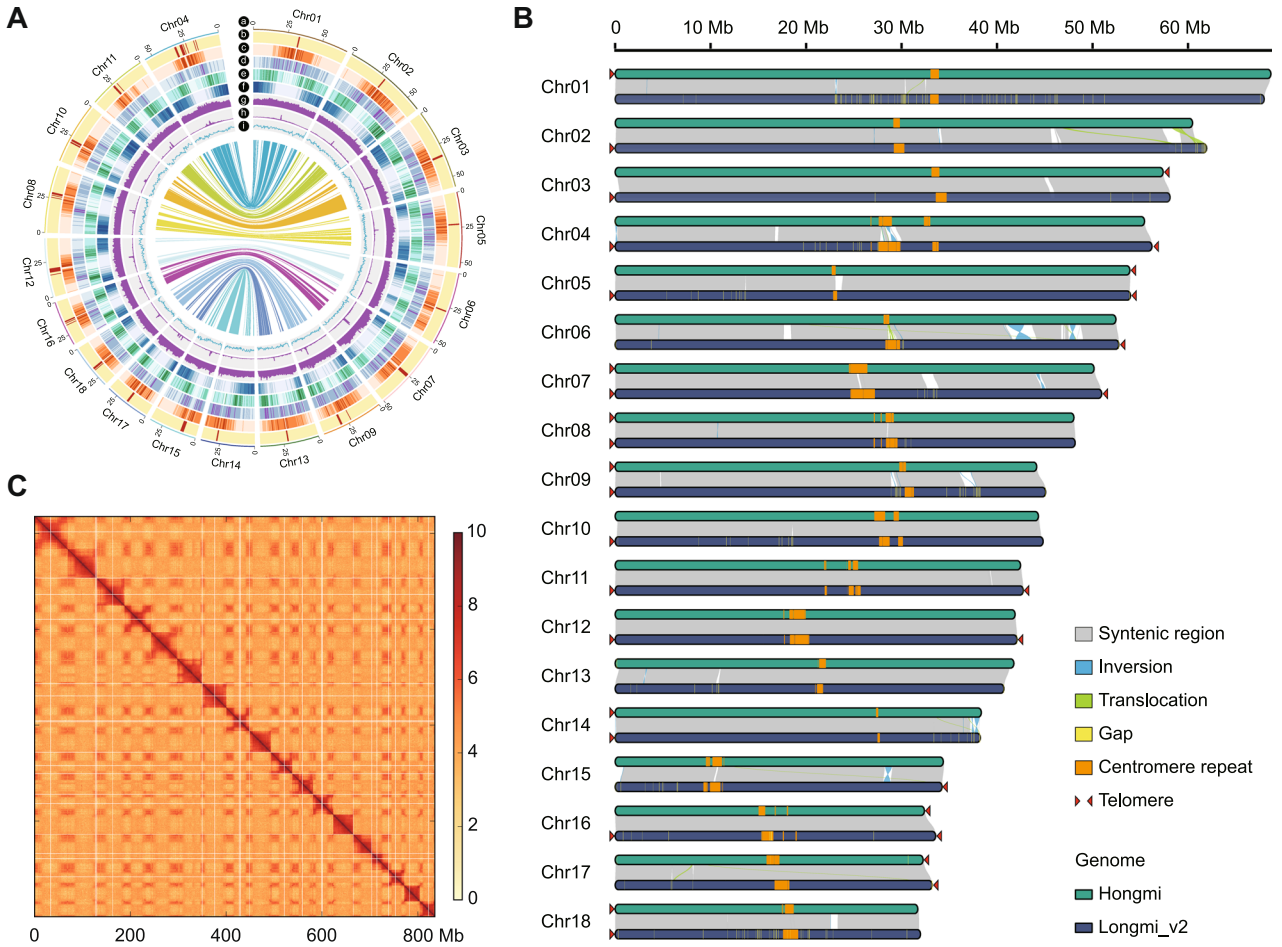


Figure 4. Genome assembly of the Hongmi genome

(A) Circos plot of features across the chromosomes of Hongmi. Tracks a–i represent chromosomes (length in Mb), centromere repeats, LTR/Gypsy, LTR/Copia, DNA TEs, genes, nucleotide polymorphisms (SNPs), SVs, and GC content, respectively. **(B)** Collinearity between Hongmi (top) and Longmi_v2 (bottom). The collinear regions between Hongmi and Longmi_v2 are linked by gray ribbons. Gap regions are shown as yellow blocks. Centromere repeat regions are as orange blocks. The triangles at the ends of chromosomes indicate the presence of telomere sequence repeats. **(C)** Hi-C chromatin interaction map of the 18 chromosomes of Hongmi. The interaction resolution is 100 kb, and the color bar to the right indicates the intensity of interaction.

Finally, we identified SNPs and insertions/deletions (InDels, <50 bp) in the broomcorn millet population based on the above alignments of short reads using GATK (McKenna et al., 2010). Using Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7 as reference genomes, we obtained 19,351,615, 19,244,325, 19,575,472, 19,650,629, and 20,079,907 high-quality SNPs and 3,668,899, 3,636,987, 3,671,730, 3,762,112, and 3,845,586 high-quality InDels, respectively. The number of variants was lowest when using Longmi_v2 as a reference genome compared with the others (Figure S13; Table S11). In humans and rice, the use of an updated genome usually leads to a lower number of variants identified in individuals due to improvements in the number of rare alleles, consensus errors, and structural errors in the updated genome (Aganezov et al., 2022; Shang et al., 2023). Although the number of variants was lower in Longmi_v2, the number of the homozygous variants was higher in Hongmi (Figure S13; Table S11).

We also used DeepVariant (Poplin et al., 2018) to call SNPs and InDels with HiFi reads from 32 broomcorn millet accessions. The number of shared SNPs or InDels from the short reads and the long reads was the highest when Hongmi and Longmi_v2 were used as the reference genomes (Figure S14; Table S11). For example, the proportion of shared SNPs between DeepVariant and GATK was 0.77 and 0.78 for Hongmi and Longmi_v2, respectively, whereas these numbers were much lower when using Longmi, Pm0390, and Jinshu7 as the reference genomes (0.74, 0.69, and 0.60). In addition, the numbers of large SVs (≥50 bp) called from long reads were similar to the numbers of small variations described above (Figure S15; Table S11). These results indicate that using the Hongmi genome as a reference can improve the performance of variant calling using either short reads or long reads.

DISCUSSION

Transgenic and genome-editing technologies are powerful approaches for introducing specific new traits into crops and modifying or recombining existing traits. These techniques

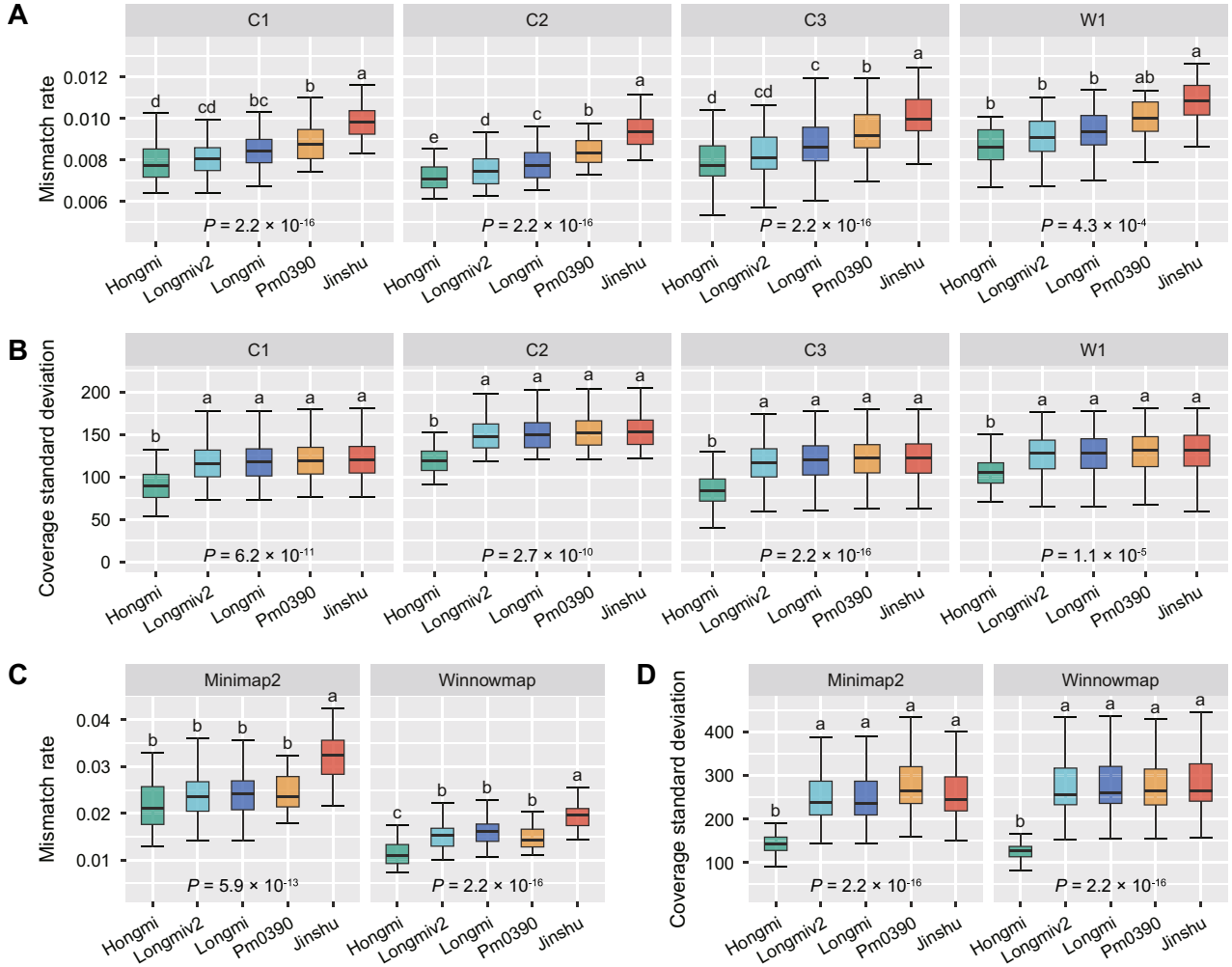


Figure 5. Improvements in short- and long-read mapping

(A) Mismatch rate of an alignment of short reads to the genomes of Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. **(B)** Standard deviation of the short-read mapping coverage for Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. The standard deviation was calculated for each 500-bp bin of the genome. **(C)** Mismatch rate of an alignment of long reads to the genomes of Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. Alignment was performed using minimap2 and winnowmap. **(D)** Standard deviation of the long-read mapping coverage for Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. The standard deviation was calculated for each 500-bp bin of the genome. In A–D, the edges and centerlines of the boxes represent the interquartile range (IQR) and medians, respectively, with the whiskers extending to the most extreme points (1.5 × IQR). Significance was tested by one-way ANOVA; multiple comparison was conducted by the least significant difference (LSD) method with Bonferroni correction. Different lowercase letters above the box plots represent significant differences ($P \leq 0.05$).

often rely on the availability of a suitable genetic transformation system (Anjanappa and Gruissem, 2021; Gao, 2021). Unfortunately, the dependence of the transformation process on specific target tissues and genotypes remains a major bottleneck for the widespread use of transgenic technologies in many crop species, including broomcorn millet (Anjanappa and Gruissem, 2021; Chen et al., 2022; Lee and Wang, 2023). Hence, to identify an optimal broomcorn millet genotype for genetic transformation, we assessed 13 widely used accessions for their tissue culture. The Hongmi accession exhibited the highest rates of embryogenic callus formation and shoot regeneration. Focusing on Hongmi, we developed a reliable genetic transformation protocol and a CRISPR/Cas9 genome-editing system for broomcorn millet.

This study presents a genetic transformation system for broomcorn millet, providing a technical foundation for its future studies of functional genes in this orphan crop, including genes that enhance plant stress resistance, thereby contributing to innovations in other crops. However, we only accessed 13 accessions for their ability for callus induction and shoot regeneration. We were unable to identify the genetic factors that control these features in broomcorn millet. Incorporating morphogenic genes, such as *Baby Boom* (*Bbm*) and *Wuschel2* (*Wus2*), can improve transformation efficiency and overcome genotype dependence (Aregawi et al., 2022; Cable et al., 2021). Therefore, further experiments including testing these morphogenic genes in broomcorn millet may help to improve the transformation

efficiency in other accessions. Moreover, optimizing the Cas9 protein based on the genome codon usage of broomcorn millet and cloning its *U6* or *U3* promoter to drive sgRNA expression may further improve genome-editing efficiency. These modifications could ensure better compatibility and expression of genome-editing components in broomcorn millet, which may lead to more efficient and precise genome-editing outcomes.

A high-quality reference genome sequence facilitates target gene discovery and transformation vector construction. This is especially important for orphan crops due to the lack of research and inadequate genetic and genomic resources (Anjanappa and Gruissem, 2021; Yaqoob et al., 2023). Broomcorn millet is an allotetraploid with a high degree of heterozygosity (Hamoud et al., 1994), which makes it difficult to assemble. Due to the development of various sequencing technologies, several genome assemblies for broomcorn millet have been reported (Shi et al., 2019; Zou et al., 2019; Sun et al., 2023; Wang et al., 2023), but they are of variable quality and contain many gaps as we described in Table 1.

A recent pan-genomic study reported the genomes of 32 broomcorn millet accessions; although this study used PacBio HiFi reads, the resulting genomes still contained dozens to hundreds of gaps (Chen et al., 2023). Compared with the above studies, we utilized not only PacBio HiFi reads but also Nanopore ultra-long reads, which generated a near-complete genome containing only six gaps. Nanopore ultra-long reads have a great advantage in assembling complex genomes because these long reads can span repetitive sequences such as centromeric regions. Our analysis revealed that the efficiency and accuracy of read mappings and genomic variant callings can be greatly improved when using Hongmi as a reference, indicating that a high-quality genome could facilitate genomic and genetic research, as demonstrated in humans and rice (Aganezov et al., 2022; Shang et al., 2023). In addition, future exploring genomics, epigenomics, proteomics, and other omics studies could further annotate functional genes or regulatory elements in the Hongmi genome and facilitate its application in breeding (Varshney et al., 2021). For example, pan-genome and associated genomic and phenotypic variations in broomcorn millet were recently described (Chen et al., 2023), greatly advancing our understanding of domestication and functional genes in this crop. Together, these tools and resources will help unlock the full potential of broomcorn millet and accelerate its breeding.

In conclusion, these tools and resources developed in this study will open up new avenues for investigating stress-resistance genes and enhancing the resilience and productivity of broomcorn millet in the face of impending global climate change. The implications of our findings extend beyond this particular crop, as the techniques and resources we developed can be applied to other orphan crops. This study contributes to the long-term sustainability of global food production by advancing research aimed at creating a climate-resilient and diverse agricultural system.

MATERIALS AND METHODS

Agrobacterium-mediated genetic transformation, GFP observation, and generation of herbicide-resistant broomcorn millet

This method with mature seeds as explants was developed based on the rice and foxtail millet transformation protocol (Wu and Sui, 2019; Yang et al., 2020) to adapt to the broomcorn millet. There were a few modifications, including the following key differences in the transformation steps: 1, Instead of using an N6-based medium, an MS-based medium was used to induce callus formation, which had a higher propensity to generate Type II calli. 2, The infection medium was the same as that used for foxtail millet (Yang et al., 2020), which is different from that of rice. Briefly, big-plump seeds were selected and carefully removed seed coats to avoid damaging the embryo using abrasive paper. Here, 200–300 de-husked seeds were placed in a 50-mL sterilized centrifuge tube and were surface sterilized in 70% (v/v) ethanol for 2 min, then in 12% bleach for 10–15 min, and finally rinsed five times with autoclaved water. The sterilized seeds were rid of the extra water using sterile paper towels and then put in CIM (4.43 g/L MS basal salt with vitamins, 30 g/L sucrose, 2 mg/L 2,4-dichlorophenoxyacetic acid, 0.3 g/L casein acid hydrolysate, 1 g/L proline, 0.1 g/L myo-inositol, 0.1 mg/L 6-benzylaminopurine, 8 g/L agar, pH 5.7). In total, 12–15 seeds of each dish (60 × 90 cm) were incubated at 28°C in the dark for 20–30 d. The primary callus was subcultured on the same medium for another 10 d before being infected with *Agrobacterium*. For estimation regeneration ability, the Type II calli were selected and transferred on SRM (4.3 g/L MS basal salt mixture, 1.0 g/L casamino acids, 1.0 mg/L nicotinic acid, 1.0 mg/L pyridoxine·HCl, 5.0 mg/L thiamine·HCl, 100 mg/L myo-inositol, 1.25 mg/L cupric sulfate, 30 g/L sucrose 8.0 g/L agar, 2 mg/L 6-benzylaminopurine, pH 5.6). Shooting differentiation was cultured at 26°C under light (16 h light: 8 h dark) for 14–30 d.

The vector pCAMBIA1305-GFP harboring the GFP reporter gene was used to optimize and establish the transformation system. The *HPT* gene in pCAMBIA1305-GFP was used to select the marker gene for testing the optimal selection concentration of hygromycin. The *BAR* gene in the pCAMBIA1305-BAR vector was used as the select marker gene for testing the optimal selection concentration of bialaphos. The above vectors were introduced into the *Agrobacterium* strain EHA105 and a single clone EHA105 cell was cultured in YEB medium (5 g/L beef extract, 5 g/L peptone, 1 g/L yeast extract, 5 g/L sucrose, 10 mM magnesium sulfate, pH 7.0) with a stable rotation speed 200 r/min overnight until optical density up to 600 nm (OD_{600} nm) = 1.0.

For *Agrobacterium* transformation, subcultured embryogenic calli were infected with *Agrobacterium* cells (OD_{600} nm = 0.5) in the infection medium (IM: 0.44 g/L Murashige–Skoog salts, 1 × B_5 of vitamins, 68 g/L sucrose, 36 g/L glucose, 1 g/L asparagine, 1 g/L casamino acids, 0.2 g/L cysteine,2 mg/L 2,4-dichlorophenoxyacetic acid, 200 μM acetosyringone, pH 5.2)

for 30 min. Removing extra infection liquid using sterile paper, the calli were transferred onto the co-cultivation medium (infection medium solidified with 8 g/L agarose) for 3 d at 22°C in the dark. After co-cultivation, the calli were transferred to the CIM resting medium containing 250 mg/L carbenicillin for 5–7 d at 28°C in the dark, and then continued transferred onto the selection medium (SM: 4 g/L CHU (N6) basal salt with vitamins, 30 g/L sucrose, 2 mg/L 2,4-dichlorophenoxyacetic acid, 0.3 g/L casein acid hydrolysate, 2.8 g/L proline, 0.1 g/L myo-inositol, 0.1 mg/L 6-benzylaminopurine, 8 g/L agar, 50 mg/L hygromycin B or 3 mg/L bialaphos, and 250 mg/L carbenicillin pH 5.7) for every 2 weeks until the resistant calli emerging. After two round selections, the resistant calli were put on SRM medium with 200 mg/L carbenicillin for 4–8 weeks, under 16 h light and 8 h dark conditions. The SRM medium was replaced every 2 weeks until shoots of 1–2 cm were regenerated. The shoots were placed in the Magenta box with RM (half-strength Murashige–Skoog basal salt with vitamins, 30 g/L sucrose, 0.1 g/L myo-inositol, 2.6 g/L Gelzan, 30 mg/L hygromycin B or 3 mg/L bialaphos and 250 mg/L carbenicillin pH 5.6) for rooting. After 3–4 weeks, the healthy roots grew under light conditions, and then the transgenic plants were moved to pots for growing to maturity. GFP was monitored by a Leica M305FCA fluorescence stereomicroscope with a DMC6200 camera. Transgenic plants were confirmed by PCR genotyping with the GFP-specific primers or PAT/BAR strip test.

Southern blotting and molecular analysis

Genomic DNA was isolated from T_0 and Wild Hongmi leaves by the cetyltrimethylammonium bromide (CTAB) extraction method (Springer, 2010). Genomic DNA (10 μg) was digested with *Bam*HI and *Sac*I (New England Biolabs Inc., Ipswich, MA, USA) restriction enzyme for 16 hs at 37°C. The 0.8% agarose gel was prepared in TAE, and the enzyme product was electrophoretic for 16 h at 30 V voltage. The electrophoretic products were imprinted on the nylon membrane using a capillary method. Membranes were hybridized with Digoxigenin-labeled probes (Catalog#: 11585614910; Roche, Mannheim, Germany). The 556-bp probe contained the *HPT* gene sequence. *HPT* gene primer pairs for PCR included forward-*HPT* (5′-ACACTACATGGCGTGATTTCAT-3′); Reverse-*HPT* (5′-TCCACTATCGGCGAGTACTTCT-3′).

Segregation analysis of T_1 seeds containing *HPT* gene was germinated on hygromycin (30 mg/L) amended half-strength MS (½MS) medium and 7 d seedlings were scored as hygromycin resistance (HygR) or hygromycin sensitive (HygS). Meanwhile, the GFP visible signal of T_1 seeds was scored as positive and negative. Primers used were in Table S12.

CRISPR/Cas9 editing and Sanger sequencing

In the CRISPR/Cas9 editing system, Cas9 was driven under the maize ubiquitin promoter and sgRNA under the rice *U3* (*OsU3*) or maize *U6-6* promotor (*ZmU6*) (Li et al., 2017). sgRNA sequences (5′-ATAAGCCAGGAGAATTCAGCCGG-3′, the underlined "CGG" represents the proto-adjacent-motif), a designed sgRNA targeting the fifth exon of the *PmPDS* gene. The amplicons were designed to surround the intended target site, and PCR assay was conducted using high-fidelity 2× Phanta Flash Master Mix polymerase (Catalog#: P520-01; Vazyme, Nanjing, China) and primers spanning the target sites. The primers for PCR were as follows: 5′-TGCCATA GTTGAACCATACAGC-3′ and 5′-TGAGATGGCAGAATGTC TTGA-3′. Mutations were identified in each sample of T_0 transformation events positive sequences from clones carrying the PCR amplicons coming from pEASY®-Blunt Zero Cloning Kit (Catalog#: CB501; TransGen Biotech., Beijing, China).

Plant materials and genome sequence

Genomic DNA was extracted and purified from broomcorn millet (Hongmi) seedlings using an optimized CTAB method. We constructed 250-bp paired-end sequencing libraries and performed sequencing on the Illumina HiSeq platform (Illumina Inc., San Diego, CA, USA). Additionally, PacBio HiFi sequencing libraries were prepared from the Hongmi seedlings according to Pacific Biosciences Co (Menlo Park, CA, USA) standard protocols, with the resulting long reads obtained from the PacBio Sequel II platform. Nanopore ultra-long reads were produced using the PromethION platform (Oxford Nanopore Technologies, Oxford, UK). For Hi-C libraries construction, fresh Hongmi seedlings were subjected to vacuum infiltration with 2% formaldehyde for crosslinking, which was then stopped with glycine. The nuclei were isolated, digested with 100 units of *Dpn*II and tagged with biotin-14-dATP. Post ligation, the DNA was sheared into 300–700-bp fragments, end-repaired, A-tailed, and purified. Finally, Hi-C libraries were quantified and sequenced on the Illumina HiSeq platform. We collected five tissues (calli, kernels, mature leaves, seedlings, and young ears) from Hongmi with three biological replicates each. Total mRNA extraction was performed using TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA), and RNA-seq libraries were prepared and sequenced with paired-end reads on the Illumina HiSeq platform following Illumina's protocols. The sequencing services were provided by Annoroad Gene Technology Co., Ltd (Beijing, China).

Genome assembly, gap-filling, and evaluation

To assemble the Nanopore ultra-long reads, we used NextDenovo v2.5.0 (https://github.com/Nextomics/NextDenovo) with parameters "read_cutoff = 1k; seed_depth = 40; nextgraph_options = -a 1 -q 16". The assembled contig sequences were polished in one round using Racon v1.5.0 (Vaser et al., 2017) by aligning the PacBio HiFi reads, and two rounds using Pilon v1.24 (Walker et al., 2014) by aligning the Illumina reads. We used hifiasm v19.3 (Cheng et al., 2021) with parameters "-l3 -u1" to assemble the PacBio HiFi reads. We additionally used the Nanopore ultra-long reads that are over 100-kb long as option "-ul". Hi-C data were used to construct the chromosome sequence for both Nanopore and HiFi assemblies using ALLHiC v0.9.13 (Zhang et al., 2019).

Finally, we checked the interaction matrix of each chromosome using HiC-Pro v3.1.0 (Servant et al., 2015).

To further fill these gaps in the Nanopore assembly, we applied several approaches. First, we extracted the reads mapping around each gap, and applied local assembling using Flye v2.9 (Kolmogorov et al., 2019) with default parameters, and three gaps in chromosomes 2, 10, and 12 were filled. For the remaining gaps, we then directly filled four gaps in chromosomes 8, 12, and 18 according to the spans of long reads. Finally, we replaced the low-accuracy Nanopore assembly with the HiFi assembly to ensure the highest base-pair accuracy.

To evaluate the assembly quality, we conducted several analyses. First, we assessed the gene completeness with BUSCO (v4.0.6) using the embryophyta_odb10 database (Manni et al., 2021) and repeat completeness based on the LAI (Ou et al., 2018) using LTR_retriever (v2.9.0) (Ou and Jiang, 2018). Then, we measured Kmer completeness, QV score, and false duplication using Merqury (v1.3) (Rhie et al., 2020).

Annotation of the transposable elements and protein-coding genes

For repeat annotation, we used RepeatModeler v2.0.1 (Flynn et al., 2020) to predict and construct a broomcorn millet-specific repeat database *de novo*. RepeatMasker v4.0.9 (http://www.repeatmasker.org/) was then employed to identify repeat sequences throughout the genome using the broomcorn millet-specific repeat database with the parameters: "-e rmblast -div 40 -norna." Protein-coding genes were annotated using the MAKER2 pipeline (v2.31.11) (Holt and Yandell, 2011), incorporating *ab initio* prediction, transcriptome evidence, and homologous protein evidence. In this process, AUGUSTUS (v3.4.0) (Stanke et al., 2008) and SNAP (v2006-07-28) (Korf, 2004) were utilized for *ab initio* gene prediction, employing customed Hidden Markov Models (HMM). Following this, we assembled Hongmi transcripts using RNA-seq datasets. The RNA-seq reads were processed and filtered using Trimmomatic v0.39 (Bolger et al., 2014) to eliminate adapters and low-quality sequences. The cleaned reads were mapped to the genome with HISAT2 v2.1.0 (Kim et al., 2019) using parameters: "--min-intronlen 20 --max-intronlen 15000", and subsequently assembled the transcripts with StringTie2 (v2.1.7) (Kovaka et al., 2019) using default settings. Homologous protein evidence was obtained from the protein sequences of Jinshu7, Longmi_v2, *Panicum hallii*, foxtail millet (*Seteria italica*), sorghum (*Sorghum bicolor*), maize (*Zea mays*), rice (*Oryza sativa*), *Arabidopsis thaliana*, and Embryophyta proteins from UniProt. Functional annotation of the protein-coding genes was conducted using InterProScan (v5.52-86.0) (Jones et al., 2014).

Short-read alignment, coverage analysis, and variant calling

We aligned 366 samples that including C1 (59), C2 (63), C3 (186), and W1 (58) populations from Chen et al. (2023) to the genome of Hongmi, Longmi_v2, Longmi, Pm0390 and Jinshu7 (Shi et al., 2019; Zou et al., 2019; Sun et al., 2023; Wang et al., 2023) using BWA-MEM v0.7.17 (Li, 2013) with default parameters. To assess the quality of the short-read alignments, we used function 'stats' in SAMtools v1.14 (Danecek et al., 2021) to quantify the mapping statistics, including error rate, reads unmapped, reads properly paired, supplementary alignments, outward-oriented pairs, pairs with other orientation, and pairs on different chromosomes. To compare the standard deviation of mapping depth, we computed coverage statistics using mosdepth v0.3.1 (Pedersen and Quinlan, 2018) on 500-bp non-overlapping genome-wide windows on the individual sample.

We called the small variants using GATK v4.2.6.1 (McKenna et al., 2010). First, we marked the duplicates using GATK MarkDuplicates for each sample. Then, we performed raw variant calls using GATK HaplotypeCaller. We then performed joint genotyping using GATK GenotypeVCFs. Finally, we filtered SNPs and InDels using GATK VariantFiltration with parameters 'QD<2.0||QUAL<30.0||SOR>3.0||FS>60.0||MQ<40.0'.

Long-read alignment and coverage analysis

To assess the impact of using Hongmi as a reference on long-read alignment, we used PacBio HiFi sequencing data from 32 samples (Chen et al., 2023). We performed alignments using minimap2 v2.24 (Li, 2018) and winnowmap v2.03 (Jain et al., 2022), and mapping statistics were compared among the five references using the function "stats" in SAMtools v1.14 (Danecek et al., 2021). In addition, we computed coverage statistics using mosdepth v0.3.1 (Pedersen and Quinlan, 2018) on 500-bp non-overlapping genome-wide windows on the individual sample.

Data availability statement

The whole genome sequence (WGS) reads (SRR27197004), PacBio HiFi reads (SRR17406970), ONT ultra-long reads (SRR21203430 and SRR21203431), Hi-C sequences (SRR17406724), and RNA-seq data (Callus: SRR17406736, SRR17406737, and SRR17406725; Seedling: SRR17406729, SRR17406730, SRR17406731, SRR17406732, and SRR17406733; Young ear: SRR17406726, SRR17406727, and SRR17406728; Mature leaf: SRR17406734, SRR17406735, and SRR17406718; Panicle: SRR17406719, SRR17406720, and SRR17406721) of Hongmi (BioProject accession no. PRJNA792682) have been deposited in the National Center for Biotechnology Information SRA. The genome assembly has been deposited in the National Center for Biotechnology Information Genome (JBBDGV000000000). The genome assembly, gene annotation, and transposable element annotation files are also available at Zenodo (https://doi.org/10.5281/zenodo.10369102). All study data are included in the main article and Supplementary Materials. All broomcorn millet accessions are available at the National Crop Genebank of China. All codes or tools used in this study are described in the methods and available from the corresponding authors. Please contact chenjinfeng@ioz.ac.cn for code access.

ACKNOWLEDGEMENTS

We thank Dr. Minxuan Liu for providing the germplasm resource of broomcorn millet. This work was funded by grants from the National Key Research and Development Program of China (2022YFD1400800; 2021YFF1000101-4), the Innovation Program of the Chinese Academy of Agricultural Sciences, the National Natural Science Foundation of China (32172093), the China Postdoctoral Science Foundation funded project (2023M743843), and Initiative Scientific Research Program, Institute of Zoology, Chinese Academy of Sciences (2023IOZ0203).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Y.S., C.W., and J.C. conceived and designed the study. Y.L. and W.C. performed genome assembly, annotation, and bioinformatic analyses. Y.S. and Z.C. performed the experiments. Y.L., Z.C., C.W., J.C., and Y.S. wrote and revised the manuscript. All authors read and approved the contents of this paper.

Edited by: Huilong Du, Hebei University, China

Received Jan. 18, 2024; **Accepted** Mar. 29, 2024

OO: OnlineOpen

References

Aganezov, S., Yan, S.M., Soto, D.C., Kirsche, M., Zarate, S., Avdeyev, P., Taylor, D.J., Shafin, K., Shumate, A., Xiao, C., et al. (2022). A complete reference genome improves analysis of human genetic variation. Science **376:** eabl3533.

Anjanappa, R.B., and Gruissem, W. (2021). Current progress and challenges in crop genetic transformation. J. Plant Physiol. **261:** 153411.

Aregawi, K., Shen, J., Pierroz, G., Sharma, M.K., Dahlberg, J., Owiti, J., and Lemaux, P.G. (2022). Morphogene-assisted transformation of *Sorghum bicolor* allows more efficient genome editing. Plant Biotechnol. J. **20:** 748–760.

Banakar, R., Schubert, M., Collingwood, M., Vakulskas, C., Eggenberger, A.L., and Wang, K. (2020). Comparison of CRISPR-Cas9/Cas12a ribonucleoprotein complexes for genome editing efficiency in the rice phytoene desaturase (*OsPDS*) gene. Rice **13**: 4.

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. **27:** 573–580.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics **30:** 2114–2120.

Buchholzer, M., and Frommer, W.B. (2023). An increasing number of countries regulate genome editing in crops. New Phytol. **237:** 12–15.

Cable, J., Ronald, P.C., Voytas, D., Zhang, F., Levy, A.A., Takatsuka, A., Arimura, S.I., Jacobsen, S.E., Toki, S., Toda, E., et al. (2021). Plant genome engineering from lab to field-a Keystone Symposia report. Ann. N. Y. Acad. Sci. **1506:** 35–54.

Chen, Z., Debernardi, J.M., Dubcovsky, J., and Gallavotti, A. (2022). Recent advances in crop transformation technologies. Nat. Plants **8:** 1343–1351.

Chen, K., and Gao, C. (2014). Targeted genome modification technologies and their applications in crop improvements. Plant Cell Rep. **33:** 575–583.

Chen, J., Liu, Y., Liu, M., Guo, W., Wang, Y., He, Q., Chen, W., Liao, Y., Zhang, W., Gao, Y., et al. (2023). Pangenome analysis reveals genomic variations associated with domestication traits in broomcorn millet. Nat. Genet. **55:** 2243–2254.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. Nat. Methods **18:** 170–175.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. Gigascience **10:** giab008.

Deng, Y., Liu, S., Zhang, Y., Tan, J., Li, X., Chu, X., Xu, B., Tian, Y., Sun, Y., Li, B., et al. (2022). A telomere-to-telomere gap-free reference genome of watermelon and its mutation library provide important resources for gene discovery and breeding. Mol. Plant **15:** 1268–1284.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U. S. A. **117**: 9451–9457.

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. Cell **184:** 1621–1635.

Ghannoum, O., Caemmerer, S.V., and Conroy, J.P. (2002). The effect of drought on plant water use efficiency of nine NAD-ME and nine NADP-ME Australian C4 grasses. Funct. Plant Biol. **29:** 1337–1348.

Goron, T.L., and Raizada, M.N. (2015). Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution. Front. Plant Sci. **6:** 157.

Hamoud, M.A., Haroun, S.A., Macleod, R.D., and Richards, A.J. (1994). Cytological relationships of selected species of *Panicum* L. Biol. Plant. **36:** 37–45.

Holt, C., and Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinform. **12:** 491.

Hunt, H.V., Rudzinski, A., Jiang, H., Wang, R., Thomas, M.G., and Jones, M.K. (2018). Genetic evidence for a western Chinese origin of broomcorn millet (*Panicum miliaceum*). Holocene **28:** 1968–1978.

Jain, C., Rhie, A., Hansen, N.F., Koren, S., and Phillippy, A.M. (2022). Long-read mapping to repetitive reference sequences using Winnowmap2. Nat. Methods **19:** 705–710.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: Genome-scale protein function classification. Bioinformatics **30:** 1236–1240.

Kamburova, V.S., Nikitina, E.V., Shermatov, S.E., Buriev, Z.T., Kumpatla, S.P., Emani, C., and Abdurakhmonov, I.Y. (2017). Genome editing in plants: An overview of tools and applications. Int. J. Agron. **2017:** 7315351.

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

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. **37:** 540–546.

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinform. **5:** 59.

Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. **20:** 278.

Kumar, K., Gambhir, G., Dass, A., Tripathi, A.K., Singh, A., Jha, A. K., Yadava, P., Choudhary, M., and Rakshit, S. (2020). Genetically modified crops: Current status and future prospects. Planta **251:** 91.

Lee, K., and Wang, K. (2023). Strategies for genotype-flexible plant transformation. Curr. Opin. Biotechnol. **79:** 102848.

Li, C., Liu, C., Qi, X., Wu, Y., Fei, X., Mao, L., Cheng, B., Li, X., and Xie, C. (2017). RNA-guided Cas9 as an *in vivo* desired-target mutator in maize. Plant Biotechnol. J. **15:** 1566–1576.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2.

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics **34:** 3094–3100.

Li, X., and Siddique, K.H.M. (2020). Future Smart Food: Harnessing the potential of neglected and underutilized species for Zero Hunger. Matern. Child Nutr. **16:** e13008.

Liu, M.X., Qiao, Z.J., Zhang, S., Wang, Y.Y., and Lu, P. (2015). Response of broomcorn millet (*Panicum miliaceum* L.) genotypes from semiarid regions of China to salt stress. Crop J. **3:** 57–66.

Lu, H., Zhang, J., Liu, K.B., Wu, N., Li, Y., Zhou, K., Ye, M., Zhang, T., Zhang, H., Yang, X., et al. (2009). Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. Proc. Natl. Acad. Sci. U. S. A. **106**: 7367–7372.

Manni, M., Berkeley, M.R., Seppey, M., Simao, F.A., and Zdobnov, E.M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. **38:** 4647–4654.

Marone, D., Mastrangelo, A.M., and Borrelli, G.M. (2023). From transgenesis to genome editing in crop improvement: Applications, marketing, and legal issues. Int. J. Mol. Sci. **24:** 7122.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20:** 1297–1303.

Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res. **46:** e126.

Ou, S., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. **176:** 1410–1422.

Pedersen, B.S., and Quinlan, A.R. (2018). Mosdepth: Quick coverage calculation for genomes and exomes. Bioinformatics **34:** 867–868.

Puchta, H. (2017). Applying CRISPR/Cas for genome engineering in plants: The best is yet to come. Curr. Opin. Plant Biol. **36:** 1–8.

Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. **21:** 245.

Salt, D.E. (2023). GMO or non-GMO? That is the question. New Phytol. **237:** 7–8.

Seghatoleslami, M.J., Kafi, M., and Majidi, E. (2008). Effect of drought stress at different growth stages on yield and water use efficiency of five proso millet (*Panicum miliaceum* L.) genotypes. Pak. J. Bot. **40:** 1427–1432.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Dekker, J., Heard, E., and Barillot, E. (2015). HiC-Pro: An optimized and flexible pipeline for Hi-C processing. Genome Biol. **16:** 259.

Shang, L., He, W., Wang, T., Yang, Y., Xu, Q., Zhao, X., Yang, L., Zhang, H., Li, X., Lv, Y., et al. (2023). A complete assembly of the rice Nipponbare reference genome. Mol. Plant **16:** 1232–1236.

Shi, J., Ma, X., Zhang, J., Zhou, Y., Liu, M., Huang, L., Sun, S., Zhang, X., Gao, X., Zhan, W., et al. (2019). Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. Nat. Commun. **10:** 464.

Springer, N.M. (2010). Isolation of plant DNA for PCR and genotyping using organic extraction and CTAB. Cold Spring Harb. Protoc. **2010**: db.prot5515.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics **24:** 637–644.

Sun, Y., Liu, Y., Shi, J., Wang, L., Liang, C., Yang, J., Chen, J., and Chen, M. (2023). Biased mutations and gene losses underlying diploidization of the tetraploid broomcorn millet genome. Plant J. **113:** 787–801.

Varshney, R.K., Bohra, A., Roorkiwal, M., Barmukh, R., Cowling, W.A., Chitikineni, A., Lam, H.M., Hickey, L.T., Croser, J.S., Bayer, P.E., et al. (2021). Fast-forward breeding for a food-secure world. Trends Genet. **37:** 1124–1136.

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

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One **9:** e112963.

Wang, Z., Huang, S., Yang, Z., Lai, J., Gao, X., and Shi, J. (2023). A high-quality, phased genome assembly of broomcorn millet reveals the features of its subgenome evolution and 3D chromatin organization. Plant Commun. **4:** 100557.

Wang, S., Gao, S., Nie, J., Tan, X., Xie, J., Bi, X., Sun, Y., Luo, S., Zhu, Q., Geng, J., et al. (2022). Improved 93-11 genome and time-course transcriptome expand resources for rice genomics. Front. Plant Sci. **12:** 769700.

Washburn, J.D., Schnable, J.C., Davidse, G., and Pires, J.C. (2015). Phylogeny and photosynthesis of the grass tribe Paniceae. Am. J. Bot. **102:** 1493–1505.

Wu, C., and Sui, Y. (2019). Efficient and fast production of transgenic rice plants by *Agrobacterium*-mediated transformation. Methods Mol. Biol. **1864:** 95–103.

Yang, Z., Zhang, H., Li, X., Shen, H., Gao, J., Hou, S., Zhang, B., Mayes, S., Bennett, M., Ma, J., et al. (2020). A mini foxtail millet with an Arabidopsis-like life cycle as a C4 model system. Nat. Plants **6:** 1167–1178.

Yaqoob, H., Tariq, A., Bhat, B.A., Bhat, K.A., Nehvi, I.B., Raza, A., Djalovic, I., Prasad, P.V., and Mir, R.A. (2023). Integrating genomics and genome editing for orphan crop improvement: A bridge between orphan crops and modern agriculture system. GM Crops Food **14:** 1–20.

Ye, C.Y., and Fan, L. (2021). Orphan crops and their wild relatives in the genomic era. Mol. Plant **14:** 27–39.

Zhang, X., Zhang, S., Zhao, Q., Ming, R., and Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nat. Plants **5:** 833–845.

Zhang, D., Zhang, Z., Unver, T., and Zhang, B. (2020). CRISPR/Cas: A powerful tool for gene function study and crop improvement. J. Adv. Res. **29:** 207–221.

Zou, C., Li, L., Miki, D., Li, D., Tang, Q., Xiao, L., Rajput, S., Deng, P., Peng, L., Jia, W., et al. (2019). The genome of broomcorn millet. Nat. Commun. **10:** 436.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: http://onlinelibrary.wiley.com/doi/10.1111/jipb.13664/suppinfo

Figure S1. Frequency of primary calli induction across 13 broomcorn millet accessions

Figure S2. Embryogenic callus induction and shoot regeneration in Hongmi

Figure S3. Regeneration capacity comparison among 13 broomcorn millet accessions

Figure S4. Genetic transformation assays in Hongmi

Figure S5. Identification of positive transgenic plants and T-DNA copy numbers analysis of T_0 plants

Figure S6. Segregation of the *GFP* transgene in T_1 seeds of Hongmi

Figure S7. Distribution of telomere repeats along the genome of Hongmi

Figure S8. ONT and HiFi read alignment to the centromeric regions in Hongmi

Figure S9. Detailed Hi-C chromatin interaction map of Hongmi

Figure S10. An example of manually inspecting the SV breakpoint between Hongmi and Longmi_v2 genomes

Figure S11. Short-read mapping statistics with Hongmi and other genomes as alignment target references

Figure S12. Long-read mapping statistics comparison using Hongmi and other available broomcorn millet genomes as reference genomes

Figure S13. Improvements for nucleotide polymorphism (SNP) and InDel calling

Figure S14. Common rate of nucleotide polymorphisms (SNPs) and InDels between DeepVariant using long reads and GATK using short reads

Figure S15. Improvements for structural variant calling using PacBio HiFi reads

Table S1. Accession information of broomcorn millet used in this study

Table S2. Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-mediated *PmPDS* knocked out in Hongmi

Table S3. Summary of PacBio HiFi, Nanopore ultra-long data for assembling

Table S4. Number of telomere repeats in Hongmi and Longmi_v2 assemblies

Table S5. Statistics of Hongmi genome assembly evaluation by BUSCO

Table S6. Statistics of Hongmi genome assembly evaluation by Merqury

Table S7. Summary of manual inspection of 35 large SVs in Hongmi (Ref) and Longmi_v2 (Qry) assemblies based on long-read mapping

Table S8. Summary of the transposable elements in the Hongmi genome

Table S9. Statistics of Hongmi gene annotation evaluated by BUSCO

Table S10. Gene function annotation of Hongmi

Table S11. Mean statistic values of short- and long-read mapping and variant calling

Table S12. Primers used for PCR, plasmid construction, and Southern blotting