

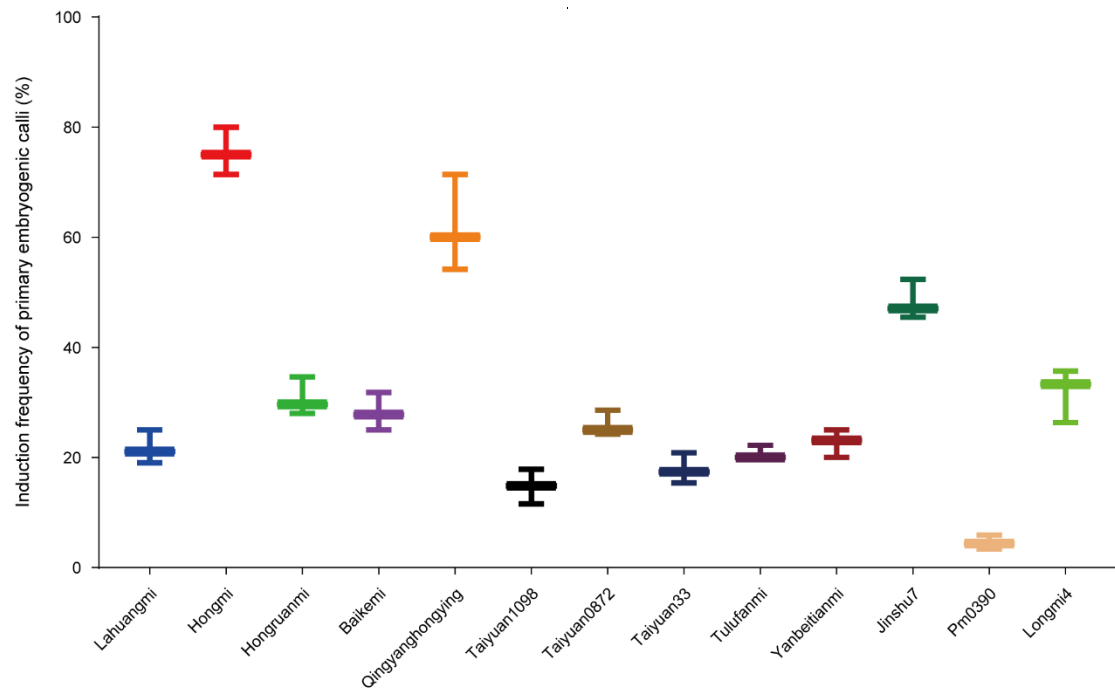


Figure S1. Frequency of primary calli induction across 13 broomcorn millet accessions.

The frequency was calculated primary calli numbers divided by germinated seeds numbers. Data present was three independent experiments with at least 100 starting seeds each.

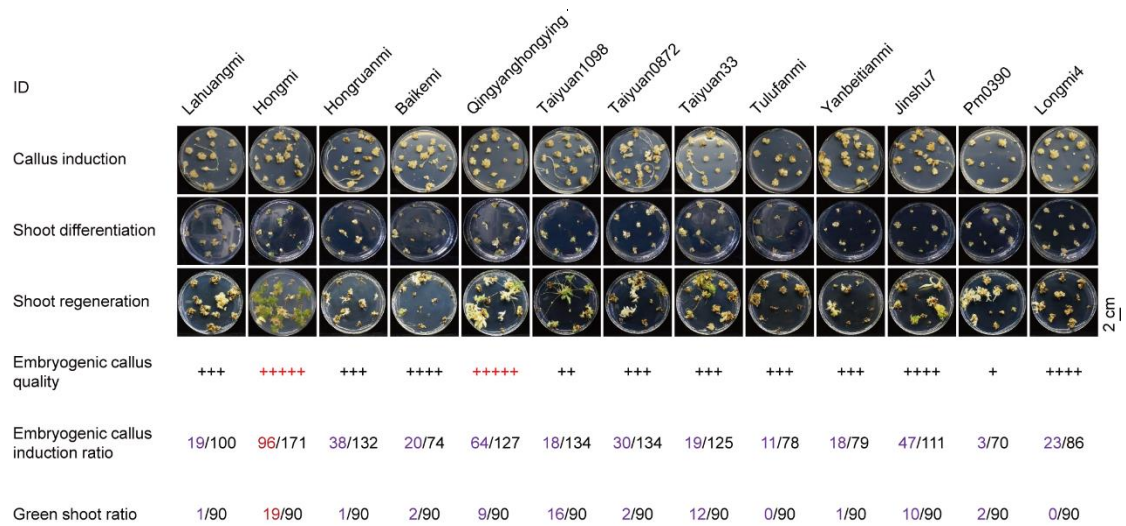


Figure S2. Embryogenic callus induction and shoot regeneration in Hongmi.

Representative images illustrate the embryogenic callus induction and shoot regeneration across 13 broomcorn millet accessions. Displayed are stages of callus formation after roughly 20 days in the induction medium, initial shoot differentiation within 5 days, and shoot elongation occurring after approximately 20 days in the regeneration medium. The quality of the embryogenic calli is evaluated and designated by the number of “+” symbols. The induction ratio of embryogenic calli induction is quantified as the number of embryogenic calli over the number of germinated seeds in the induction medium. The proportion of green to total shoots is determined based on the observation of 90 calli after approximately 20 days in the regeneration medium. Scale bar, 2 cm.

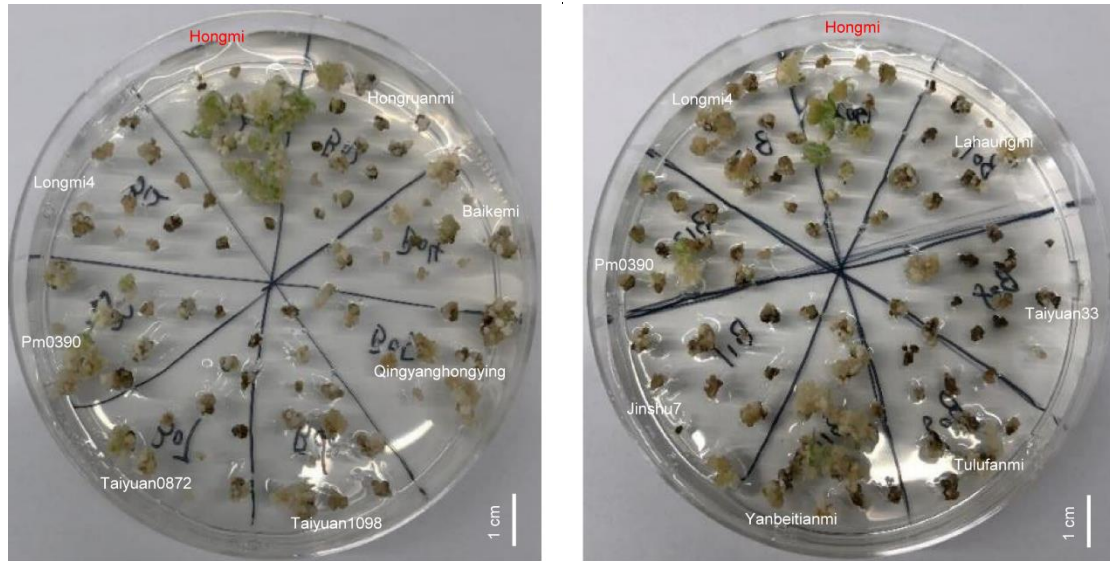


Figure S3. Regeneration capacity comparison among 13 broomcorn millet accessions.

For each accession, approximately 10 calli were estimated for the time and number of green shoots produced. Around 60 days, selected calli were transferred to the regeneration medium. Data was derived from three independent experiments. The diameter of all culture dishes is 90 mm. Scale bar, 1 cm.

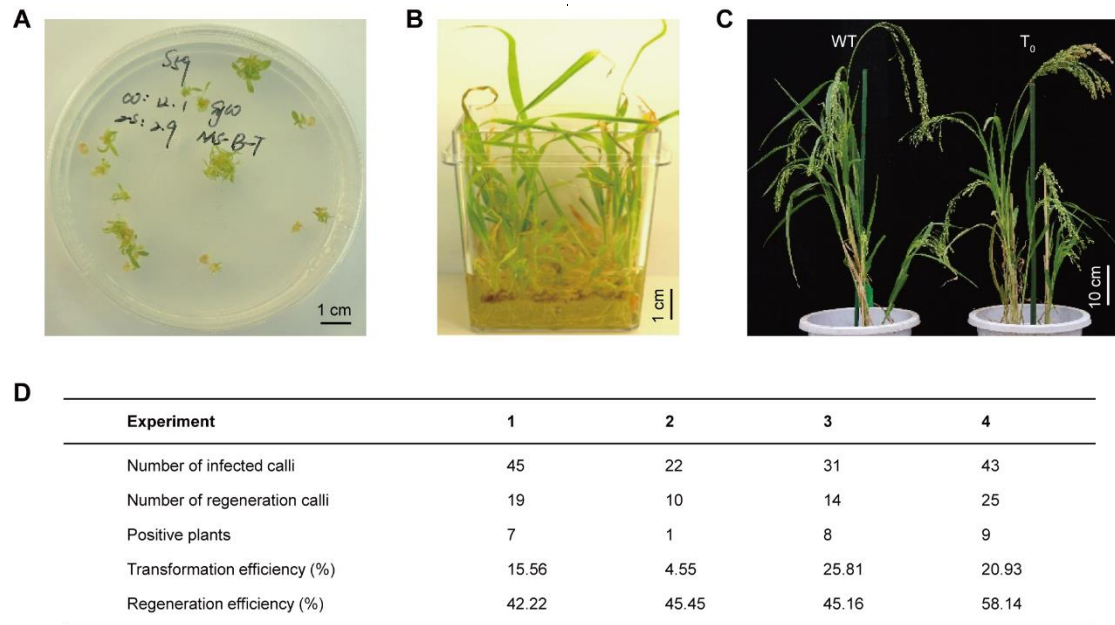


Figure S4. Genetic transformation assays in Hongmi.

(A) Regeneration of hygromycin-resistant calli expressing GFP.

(B) Rooting and hardening GFP-expressing plants.

(C) Comparison of the phenotype during heading between a T₀ GFP-expressing plant and its wild-type control.

(D) Efficiency of *Agrobacterium*-mediated genetic transformation and regeneration efficiency of GFP gene across four independent experiments. Transformation efficiency (%) = GFP positive plants / total infected calli * 100%; Regeneration efficiency (%) = the number of regeneration calli / total infected calli * 100%.

Scale bar, 1 cm in **A**, **B** and 10 cm in **C**.

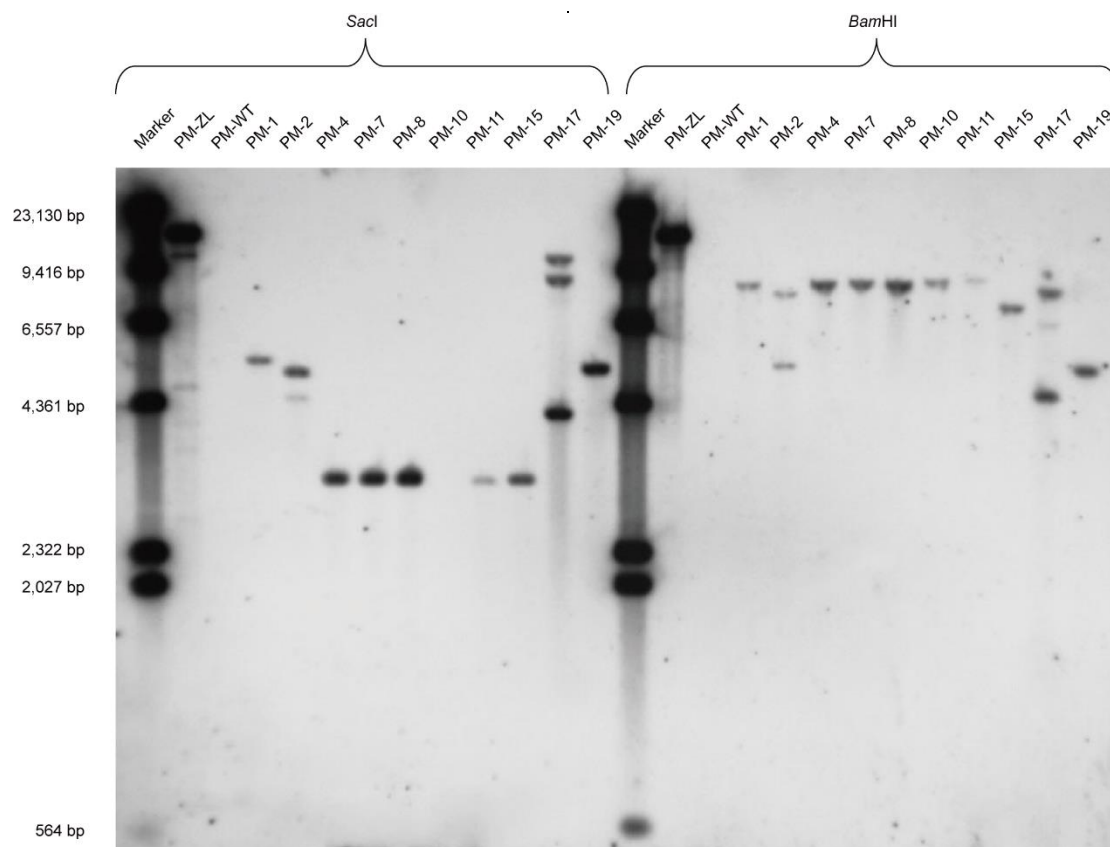


Figure S5. Identification of positive transgenic plants and T-DNA copy numbers analysis of T₀ plants.

Estimation of T-DNA copy number in 10 confirmed T₀ plants using Southern blotting analysis. Random selection of 10 plants for analysis. Genomic DNA (10 µg per sample) was digested with *SacI* and *BamHI*, respectively. The HPT probe was used for hybridization. Plasmid DNA (PM-ZL) and DNA from wild-type plants (PM-WT) were used as positive and negative controls, respectively.

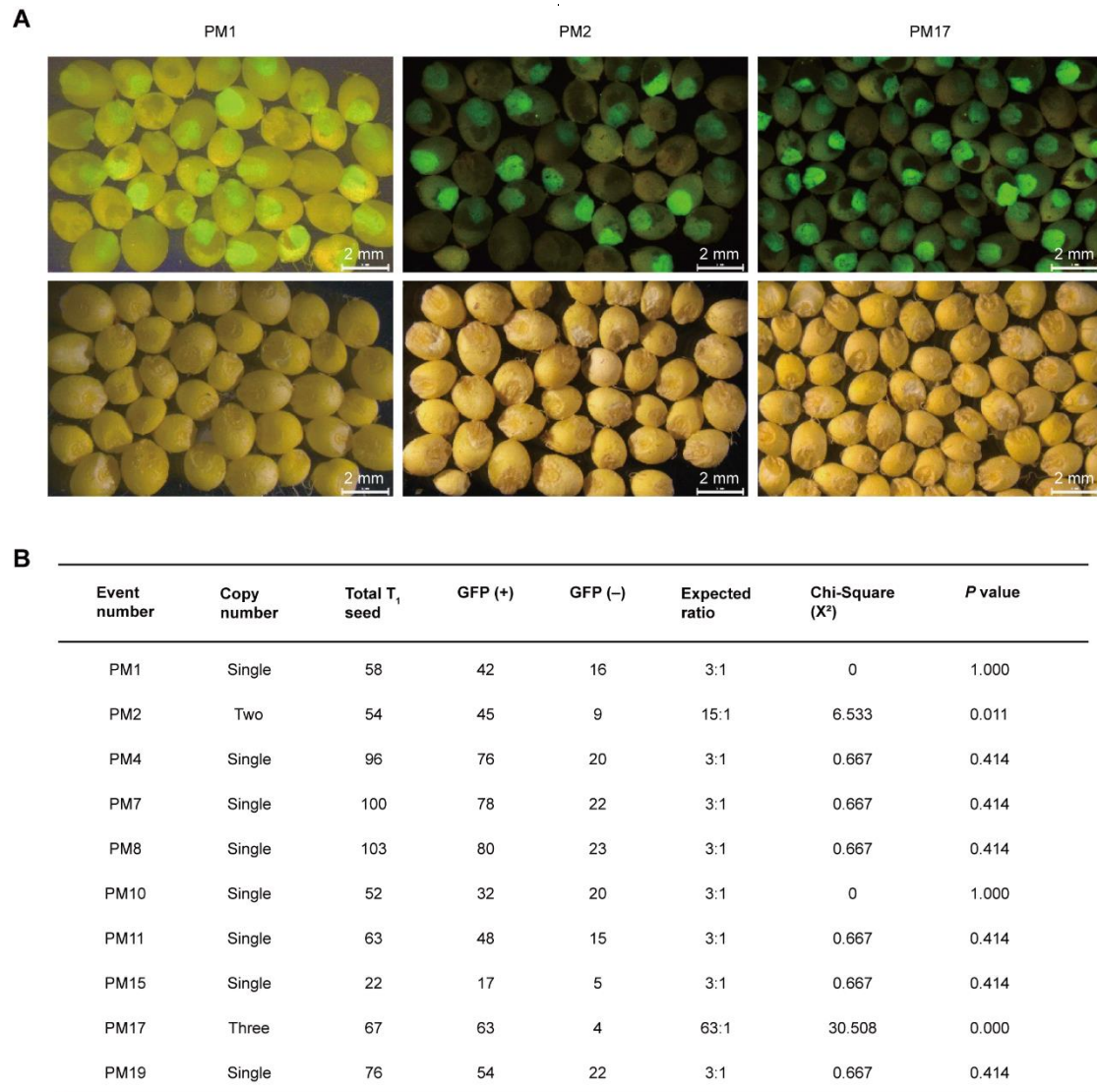


Figure S6. Segregation of the GFP transgene in T₁ seeds of Hongmi.

(A) GFP accumulation in dry dehusked seeds from three T₀ plants carrying one in PM1 (left), two in PM2 (middle), or three copies in PM17 (right) as visualized under UV (upper) and bright (lower). Scale bar, 2 mm. **(B)** Segregation analysis of the transgene in T₁ seedlings from southern verification T₀ plants carrying one, two, or three copies. The presence (+) or absence (-) of the GFP transgene was determined by observing GFP expression in seedlings under UV light seven days post-germination.

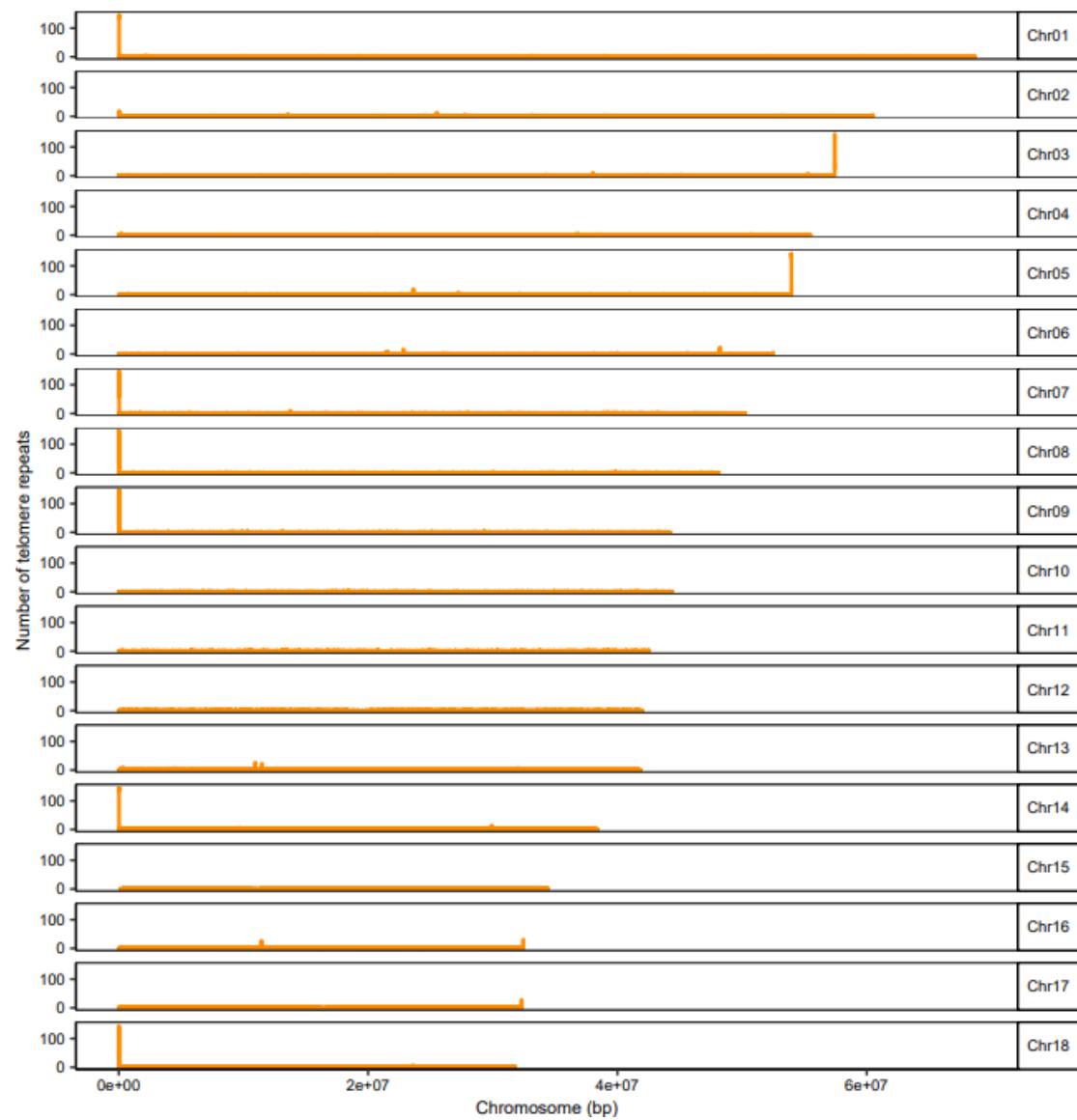


Figure S7. Distribution of telomere repeats along the genome of Hongmi.

Telomeres were identified using seven-base telomere repeats (5'-TTTAGGG-3') as sequence queries. Window size is 1 kb. Chromosomes 1-18 are shown.

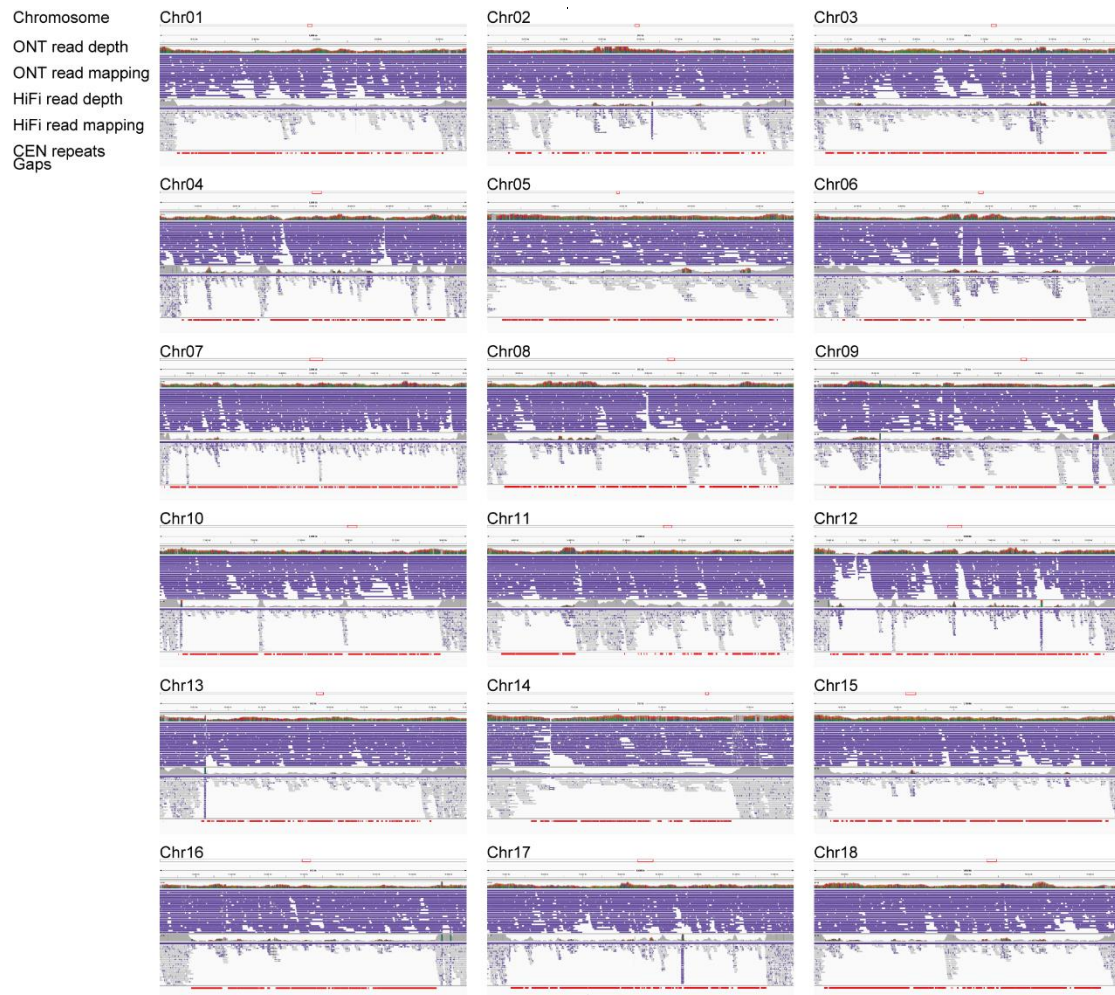


Figure S8. ONT and HiFi read alignment to the centromeric regions in Hongmi. Centromeric regions in all 18 chromosomes are shown. Tracks from top to bottom are Chromosome, ONT read coverage, ONT read mapping, HiFi read coverage, HiFi reads mapping, centromeric repeats, and gaps, respectively.

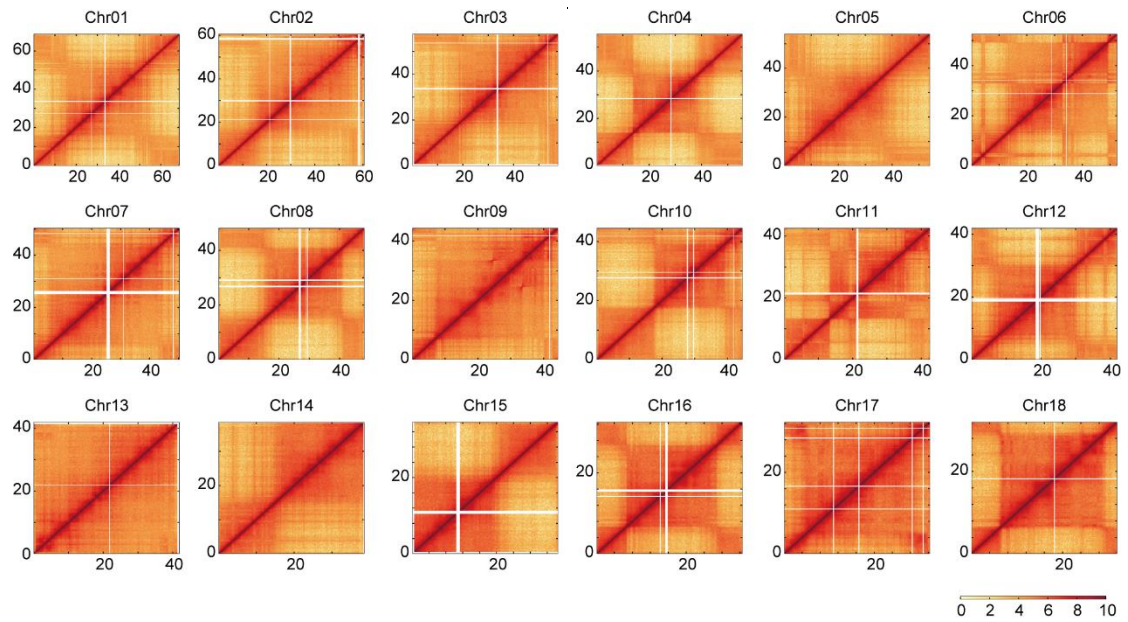


Figure S9. Detailed Hi-C chromatin interaction map of Hongmi.

Chromosomes 1-18 of Hongmi genome are shown. For each image, the axes represent chromosome position in Mb. The Hi-C interaction matrices are shown by 100 kb resolution. The color bar below indicates the intensity of interaction.

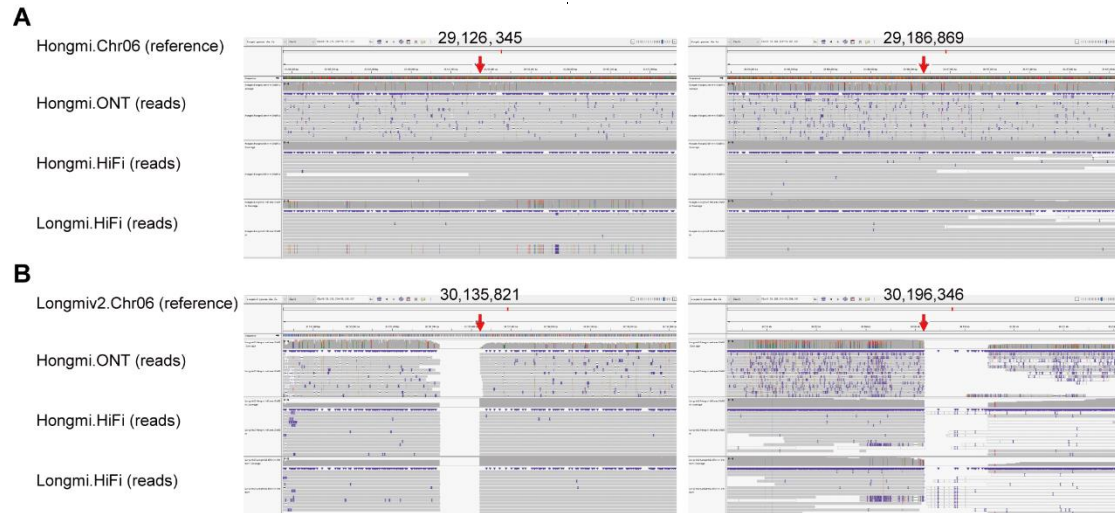


Figure S10. An example of manually inspection the SV breakpoint between Hongmi and Longmi_v2 genomes.

IGV screenshots show the Oxford Nanopore and PacBio HiFi read alignments, which aligned to the Hongmi and Longmi_v2 assemblies. The red arrows represent the breakpoints of an inversion between Hongmi and Longmi_v2 genomes (Hongmi.Chr06:29126345-29186869 vs. Longmi_v2.Chr06:30135821-30196346). The results showed that no reads supporting the assembly of Longmi4_v2, whereas all reads supporting the assembly of Hongmi.

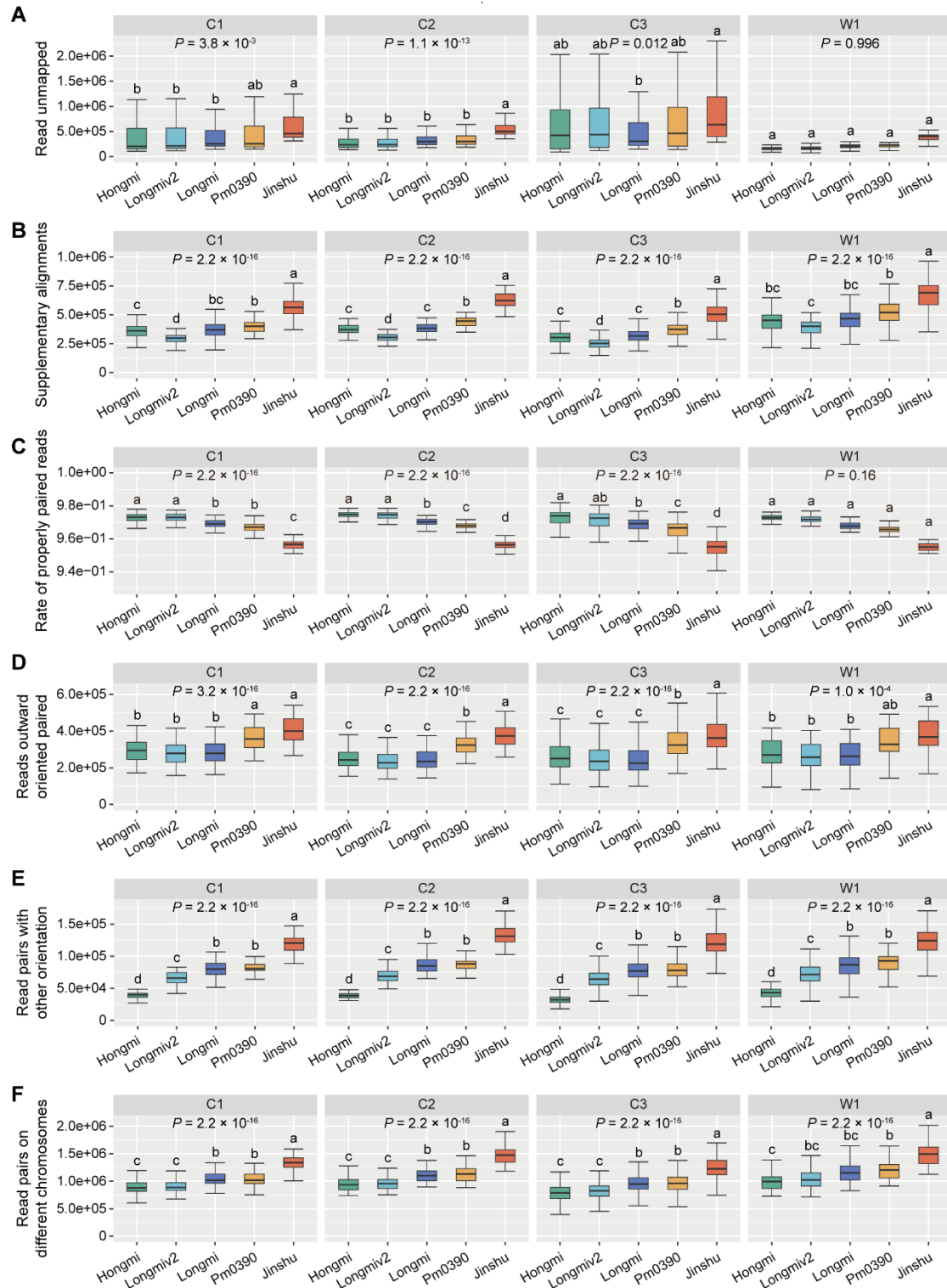


Figure S11. Short-read mapping statistics with Hongmi and other genomes as alignment target references.

(A) Number of unmapped reads aligning genome of Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. (B) Number of supplementary alignments aligning genome of Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. (C) Number of properly paired

reads aligning genome of Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. **(D)** Number of outward orientation paired reads aligning genome of Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. **(E)** Number of read pairs with other orientation aligning genome of Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. **(F)** Number of read pairs on different chromosomes aligning genome of Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. For each box plot, the edges and centerlines represent the interquartile range (IQR) and medians, with the whiskers extending to the most extreme points ($1.5 \times \text{IQR}$). Significance was tested by one-way ANOVA; multiple comparison was conducted by the least significant difference (LSD) method with Bonferroni correction. The lowercase letters above the box plots represent significant differences ($P \leq 0.05$).

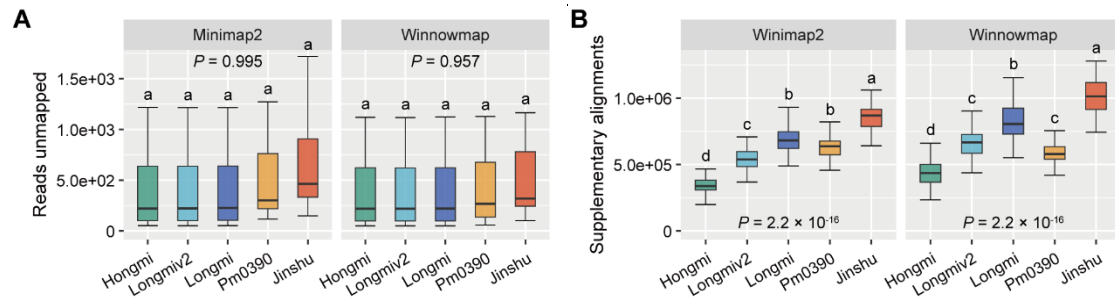


Figure S12. Long-read mapping statistics comparison using Hongmi and other available broomcorn millet genomes as reference genomes.

(A) Number of unmapped reads aligning genome of Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. **(B)** Number of supplementary alignments aligning genome of Hongmi, Longmi_v2, Longmi, Pm0390, and Jinshu7. For each box plot, the edges and centerlines represent the IQR and medians, with the whiskers extending to the $1.5 \times$ IQR. Significance was tested by one-way ANOVA; multiple comparison was conducted by LSD method with Bonferroni correction. The lowercase letters above the box plots represent significant differences ($P \leq 0.05$).

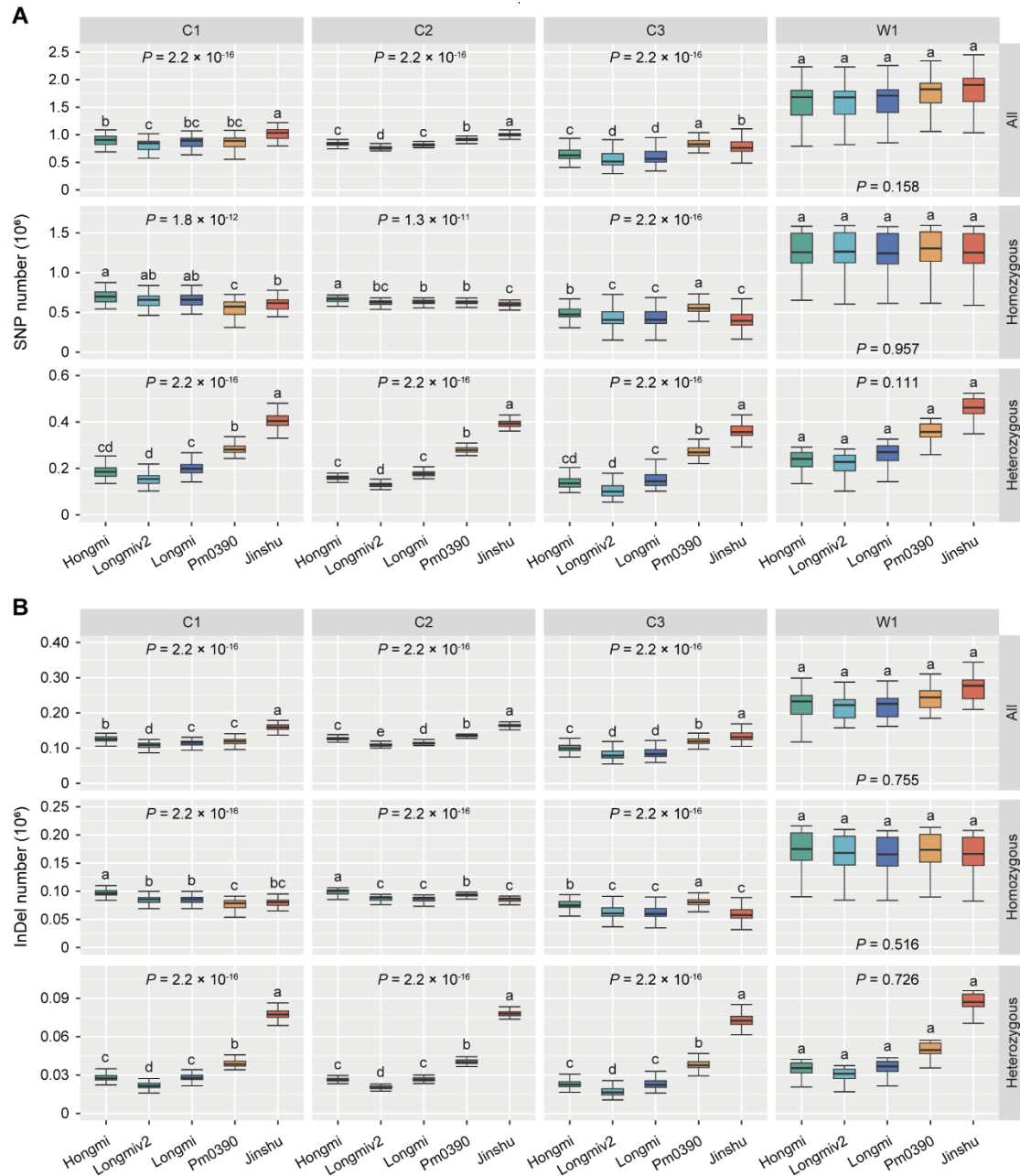


Figure S13. Improvements for SNP and InDel callings.

(A, B) Number of overall, homozygous, and heterozygous SNPs (A) and InDels (B) found in each sample across populations. For each box plot, the edges and centerlines represent the IQR and medians, with the whiskers extending to the $1.5 \times$ IQR. Significance was tested by one-way ANOVA; multiple comparison was conducted by LSD method with Bonferroni correction. The lowercase letters above the box plots represent significant differences ($P \leq 0.05$).

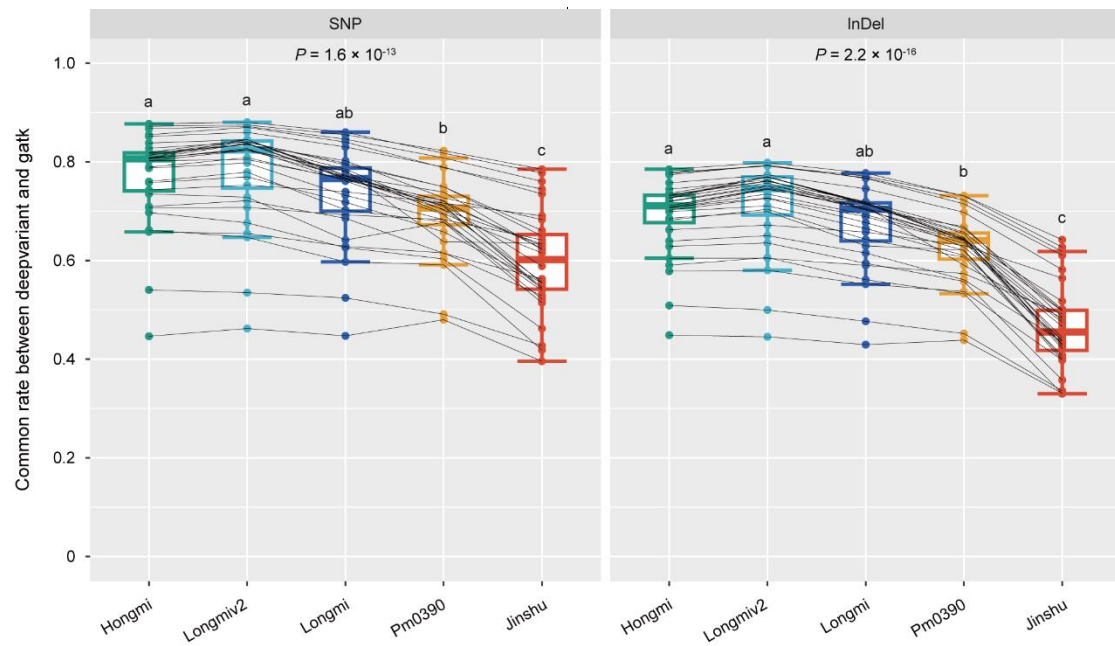


Figure S14. Common rate of SNPs and InDels between DeepVariant using long-reads and GATK using short-reads.

(A, B) Common rate of SNPs (A) and InDels (B) found in 32 samples between long- and short-read callings. For each box plot, the edges and centerlines represent the IQR and medians, with the whiskers extending to the $1.5 \times$ IQR. Significance was tested by one-way ANOVA; multiple comparison was conducted by LSD method with Bonferroni correction. The different lowercase letters above the box plots represent significant differences ($P \leq 0.05$).

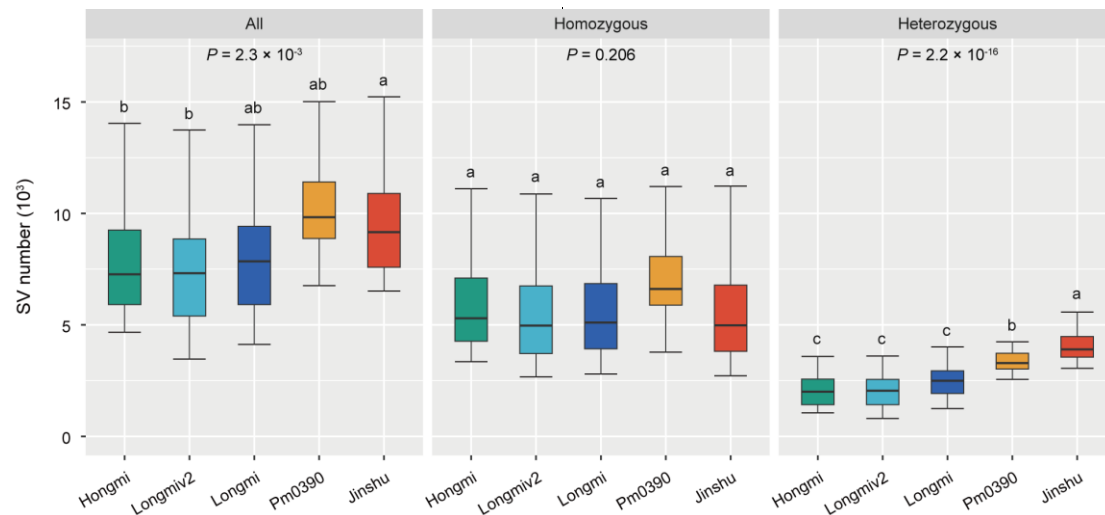


Figure S15. Improvements for structural variant calling using PacBio HiFi reads.

Number of overall, homozygous, and heterozygous structural variants found in 32 samples. For each box plot, the edges and centerlines represent the IQR and medians, with the whiskers extending to the most extreme points ($1.5 \times \text{IQR}$). Significance was tested by one-way ANOVA; multiple comparison was conducted by LSD method with Bonferroni correction. The different lowercase letters above the box plots represent significant differences ($P \leq 0.05$).

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

Accession ID	Experiment ID	Name	Original producing area	Province
00007417	B01	Lahuangmi	China	Qinghai
00007449	B02	Hongmi	China	Qinghai
00004677	B03	Hongruanmi	China	-
00004704	B04	Baikemi	China	-
00005218	B05	Qingyanghongying	China	-
00006541	B06	Taiyuan1098	China	Shanxi
00006550	B07	Taiyuan0872	China	Shanxi
00006559	B08	Taiyuan33	China	Shanxi
00007344	B09	Tulufanmi	China	Xinjiang
00007361	B10	Yanbeitianmi	China	Shanxi
NA	B11	Jinshu7	China	Shanxi
NA	B12	Pm_0390	China	-
NA	B13	Longmi4	China	Gansu

Table S2. CRISPR/Cas9-mediated *PmPDS* knocked out in Hongmi

ID	Edited type in <i>PmPDS-A</i>	Edited sequence in <i>PmPDS-A</i>	Edited type in <i>PmPDS-B</i>	Edited sequence in <i>PmPDS-B</i>
S372-1	-5 bp/-5 bp	-AATTC/AATTC	-1 bp/-1 bp	-C/-C
S372-2	-14 bp/WT	-AGCCAGGAGAATT C/WT	-2 bp/-13 bp	-TC/-GGAGAATT CAGCC
S372-3	+2 bp/+2 bp	+TT/+TT	-4 bp/-4 bp	-ATTC/-ATTC
S372-4	+2 bp/+2 bp	+TT/+TT	-4 bp/-5 bp	-ATTC/-AATTC
S372-5	-44 bp/-44 bp	-GAATAAGCCAGGA GAATTCAGCCGGTT TGATTTCCCAGAAA CTT/-GAATAAGCCA GGAGAATTCAGCCG GTTTGATTTCCCAGA AACTT	-2 bp/-4 bp	-TC/-TCAG
S372-6	-5 bp/+1 bp	-TCTTA/+1A	-4 bp/-4 bp	-TCAG/-TCAG
S372-7	-7 bp/-7 bp	-AGAATTC/- AGAATTC	-1 bp/-1 bp	-C/-C
S372-8	-5 bp/-5 bp	-AATTC/-AATTC	-18 bp/-18 bp	- AATAAGCCAGG AGAATTC/- AATAAGCCAGG AGAATTC
S372-9	-3 bp/-3 bp	-CAG/-CAG	-4 bp/-4 bp	-TCAG/-TCAG
S372-10	-5 bp/-5 bp	-AATTC/-AATTC	-4 bp/-18 bp	-ATTC/- AATAAGCCAGG AGAATTC
S372-11	WT/WT	WT/WT	+2 bp/+2 bp	+TT/+TT

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

	PacBio HiFi	Nanopore
N50 (bp)	15,790	59,395
N90 (bp)	12,022	27,386
N95 (bp)	11,350	19,821
Read number	1,461,405	899,125
Total length (bp)	22,550,212,300	38,020,024,488
Mean length (bp)	15,431	42,286
Max length (bp)	44,874	556,355
Min length (bp)	116	14
Depth (×)	27	45

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

Chromosome	Hongmi		Longmi_v2	
	chrLeft	chrRight	chrLeft	chrRight
Chr01	4135	0	0	0
Chr02	0	0	558	0
Chr03	0	933	2,193	0
Chr04	0	0	2,435	2,235
Chr05	0	847	30	1,990
Chr06	0	0	0	2,067
Chr07	5,064	0	3,464	1,529
Chr08	4,088	0	254	0
Chr09	6,282	0	183	0
Chr10	0	0	101	0
Chr11	0	0	491	71
Chr12	0	0	4,116	2,657
Chr13	0	0	0	0
Chr14	3,173	0	249	0
Chr15	0	0	0	1,501
Chr16	0	36	352	1,747
Chr17	0	44	0	1,845
Chr18	445	0	62	0

Table S5. Statistics of Hongmi genome assembly evaluation by BUSCO

	Number	Percent (%)
Complete BUSCOs (C)	1,596	98.9
Complete and single-copy BUSCOs (S)	370	22.9
Complete and duplicated BUSCOs (D)	1,226	76.0
Fragmented BUSCOs (F)	6	0.4
Missing BUSCOs (M)	12	0.7
Total BUSCO groups searched	1,614	

Table S6. Statistics of Hongmi genome assembly evaluation by Merqury

Chromosome	QV	Error rate
Chr01	43.7227	4.24E-05
Chr02	42.7101	5.36E-05
Chr03	44.8592	3.27E-05
Chr04	40.3513	9.22E-05
Chr05	44.8317	3.29E-05
Chr06	42.9964	5.02E-05
Chr07	43.4479	4.52E-05
Chr08	40.973	7.99E-05
Chr09	43.194	4.79E-05
Chr10	41.2402	7.52E-05
Chr11	41.4394	7.18E-05
Chr12	38.3751	1.45E-04
Chr13	42.3688	5.80E-05
Chr14	44.8888	3.24E-05
Chr15	40.9832	7.97E-05
Chr16	40.5297	8.85E-05
Chr17	40.2197	9.51E-05
Chr18	41.5391	7.02E-05

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

Ref	refStart	refEnd	Qry	qryStart	qryEnd	Type	Length	Inspection
Chr01	23078648	23217692	Chr01	23051129	23191979	INV	140850	False LM2
Chr01	25653896	25956160	Chr01	26856324	27077840	TRL	221516	False LM2
Chr01	32485017	32557201	Chr01	30367827	30440011	TRL	72184	False LM2
Chr01	39594403	39662610	Chr01	39321027	39321027	DEL	68207	Complex region
Chr01	39962141	40057532	Chr01	39922018	40017408	TRL	95390	False LM2
Chr01	50486918	50569235	Chr01	30234656	30320029	TRL	85373	False LM2
Chr01	50565821	50638047	Chr01	30170935	30233964	TRL	63029	False LM2
Chr02	33742446	33873736	Chr02	33974095	34276283	INV	302188	True SV
Chr02	58043541	58122614	Chr02	61859126	61969531	TRL	110405	False LM2
Chr02	58079106	59047085	Chr02	60992407	61933967	TRL	941560	False LM2
Chr02	58934445	58987977	Chr02	61043052	61043052	DEL	53532	True SV
Chr02	59047086	59189206	Chr02	60827859	60969860	TRL	142001	False LM2
Chr04	22767	180086	Chr04	41554	209907	INV	168353	Complex region
Chr04	27846425	27855527	Chr04	28196419	28250772	INV	54353	False LM2
Chr04	28254414	28708129	Chr04	28982645	29565132	INV	582487	Complex region
Chr04	29031193	29082184	Chr04	29880043	29880043	DEL	50991	Complex region
Chr05	27314827	27371053	Chr05	27206579	27206579	DEL	56226	True SV
Chr06	3773540	3856759	Chr06	2	83219	TRL	83217	False LM2
Chr06	29126345	29186869	Chr06	30135821	30196346	INV	60525	False LM2
Chr06	34029681	34113817	Chr06	35065871	35065871	DEL	84136	Complex region
Chr06	40900461	42944762	Chr06	41817519	43821911	INV	2004392	True SV
Chr06	47476386	48270929	Chr06	47519303	48314398	INV	795095	True SV
Chr07	32015180	32134053	Chr07	33308346	33421717	TRL	113371	False LM2
Chr07	43985016	44395201	Chr07	44726000	45116225	INV	390225	Complex region
Chr08	10720805	10786677	Chr08	10674274	10739772	INV	65498	False LM2
Chr09	28820071	28886059	Chr09	29366972	29433013	INV	66041	False LM2
Chr10	41634066	41715110	Chr10	41949829	41949829	DEL	81044	Complex region
Chr13	3239393	3362085	Chr13	2891235	3010934	INV	119699	False LM2
Chr13	11014296	11086172	Chr13	10545655	10617575	INV	71920	False LM2
Chr14	37742442	38195797	Chr14	37538057	38018616	INV	480559	True SV
Chr15	28239345	29042272	Chr15	28055175	28859416	INV	804241	True SV
Chr17	8107242	8214818	Chr17	5882342	5989960	TRL	107618	False LM2
Chr17	14126830	14126830	Chr17	14523710	14583476	INS	59766	Complex region
Chr17	28325613	28389361	Chr17	29344447	29344447	DEL	63748	Complex region
Chr17	30700353	30811298	Chr17	31657912	31657912	DEL	110945	Complex region

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

Superfamily	Count	Length (bp)	Percentage (%)
Class I			
LTR_Copia	58,477	64,268,306	7.73
LTR_Gypsy	155,579	307,791,499	37.02
LTR_Unknown	19,386	6,660,793	0.8
LINE	22,793	11,550,272	1.38
SINE	7,861	1,314,687	0.16
Class II			
CMC-EnSpm	39,831	20,215,591	2.43
MULE-MuDR	25,440	7,748,652	0.93
hAT	19,235	5,728,828	0.69
PIF-Harbinger	29368	6586276	0.79
TcMar-Stowaway	20,466	3,075,328	0.37
DNA_Unknown	2,226	462,091	0.05
RC_Helitron	9,429	3,433,553	0.41
Unknown interspersed	73,987	16,655,326	2
Total interspersed	484,078	455,491,202	54.78
Low_complexity	14048	727,154	0.09
Satellite	1,539	15,477,967	1.86
Simple_repeat	126,101	6,193,537	0.74
Total repeat	625,766	477,889,860	57.47

Table S9. Statistics of Hongmi gene annotation evaluated by BUSCO

	Number	Percent (%)
Complete BUSCOs (C)	1,543	95.6
Complete and single-copy BUSCOs (S)	297	18.4
Complete and duplicated BUSCOs (D)	1,246	77.2
Fragmented BUSCOs (F)	17	1.1
Missing BUSCOs (M)	54	3.3
Total BUSCO groups searched	1,614	

Table S10. Gene function annotation of Hongmi

InterProScan Database	Hit number	Percent
CDD	15,169	25.2%
Gene3D	33,527	55.7%
Pfam	39,711	66.0%
PRINTS	5,934	9.9%
ProSiteProfiles	18,287	30.4%
SMART	14,097	23.4%
SUPERFAMILY	31,632	52.6%

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

Features	Hongmi	Longmi_v2	Longmi	Pm0390	Jinshu7
Illumina mapping					
Mismatch rate	0.007986012	0.008325179	0.00873652	0.009284047	0.01014568
Coverage standard deviation	93.1336	125.4334	127.8065	129.8025	130.1791
Reads unmapped	601,570	619,364	579,967	654,123	844,539
Supplementary alignments	348,560	290,394	353,807	410,763	563,309
Rate of properly paired reads	0.9711153	0.9701759	0.9670673	0.9641437	0.9533731
Reads outward oriented paired	280,987	265,435	263,136	353,612	393,899
Read pairs with other orientation	36,253	67,608	80,869	82,922	123,396
Read pairs on different chromosomes	862,081	897,664	1,022,075	1,039,106	1,336,311
PacBio mapping					
Mismatch rate	0.01650432	0.01961111	0.02011718	0.01955918	0.02574372
Coverage standard deviation	133.5535	257.0636	258.6814	273.1297	266.0863
Reads unmapped	697	697	699	773	867
Supplementary alignments	393,520	600,976	761,660	617,865	950,746

Be continued.

Continued.

Features	Hongmi	Longmi_v2	Longmi	Pm0390	Jinshu7
Variant calling					
SNP number (Homozygous)	674,520	627,314	628,246	683,945	608,044
SNP number (Heterozygous)	211,026	180,105	225,282	331,548	432,840
INDEL number (Homozygous)	99,005	86,871	85,930	96,638	84,165
INDEL number (Heterozygous)	36,832	30,980	37,012	50,823	86,716
SNP common rate between deepvariant and gatk	0.7719463	0.7832377	0.7378741	0.6938694	0.5982808
INDEL common rate between deepvariant and gatk	0.6918605	0.7126017	0.6736834	0.6245065	0.4674516
SV number (Homozygous)	6,240	5,960	6,049	7,411	6,053
SV number (Heterozygous)	2,049	2,076	2,505	3,389	4,060

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

Primer name	Primer (5'-3')	Note
Pm-PDS-F	TGCCATAGTTGAACCATACAGC	<i>PmPDS</i> target sequencing
Pm-PDS-R	TGAGATGGCAGAATGTCTTGA	
U6-6-1F	CGGCCAGTGCCAAGCTAATTGGCCCTTACAAAATAGC	<i>PmPDS</i> knockout
U6-6-1R	CTGCGACCACCGCTCCATAAGCCAGGAGAATTCAGC	
sgRNA1-F	ATAAGCCAGGAGAATTCAGCGTTTTAGAGCTAGAA	
sgRNA1-R	CACTGCACAAGCTAAAAAAGCACCGACTCGGTGC	
U3-sgRNA-F	AGATGATCCGTGGCAATAAGCCAGGAGAATTCAGCG	
U3-sgRNA-R	TTT TAGAGCTATGC GCATAGCTCTAAAACGCTGAATTCTCCTGGCTTATTG CCACGGATCATCT	Southern blotting
HPT-F	ACACTACATGGCGTGAT	
HPT-R	TCCACTATCGGCGAGTACTTCT	