nature plants

ARTICLES

https://doi.org/10.1038/s41477-020-0747-7

Check for updates

A mini foxtail millet with an *Arabidopsis*-like life cycle as a C_4 model system

Zhirong Yang[1,2,8], Haoshan Zhang[3,8], Xukai Li[1,4,8], Huimin Shen[4], Jianhua Gao[1,4], Siyu Hou[1,5], Bin Zhang[1,5], Sean Mayes[6], Malcolm Bennett[6], Jianxin Ma[7], Chuanyin Wu[3], Yi Sui[3]✉, Yuanhuai Han[1,5]✉ and Xingchun Wang[1,4]✉

Foxtail millet (*Setaria italica*) is an important crop species and an emerging model plant for C_4 grasses. However, functional genomics research on foxtail millet is challenging because of its long generation time, relatively large stature and recalcitrance to genetic transformation. Here we report the development of *xiaomi*, a rapid-cycling mini foxtail millet mutant as a C_4 model system. Five to six generations of *xiaomi* can be grown in a year in growth chambers due to its short life cycle and small plant size, similar to *Arabidopsis*. A point mutation in the *Phytochrome C* (*PHYC*) gene was found to be causal for these characteristics. *PHYC* encodes a light receptor essential for photoperiodic flowering. A reference-grade *xiaomi* genome comprising 429.94 Mb of sequence was assembled and a gene-expression atlas from 11 different tissues was developed. These resources, together with an established highly efficient transformation system and a multi-omics database, make *xiaomi* an ideal model system for functional studies of C_4 plants.

Over the past few decades, several plant species, including *Arabidopsis thaliana*, *Brachypodium distachyon* and rice (*Oryza sativa*), have been adopted as model plants for various aspects of research. These species, especially *Arabidopsis*, have had vital roles in making fundamental discoveries and technological advances[1]. However, all these model plants use C_3 photosynthesis, and discoveries made in these species are not always transferable to, or representative of, C_4 plants such as maize (*Zea mays*), sorghum (*Sorghum bicolor*) and millets, which are efficient fixers of atmospheric CO_2 into biomass. Thus, it is critical to develop a new model system for studies in these and many other C_4 plants[2].

Foxtail millet (*S. italica*) is a cereal crop that was domesticated from its wild ancestor, green foxtail (*Setaria viridis*). These two species are evolutionarily close to several bioenergy crops, including switchgrass (*Panicum virgatum*), napiergrass (*Pennisetum purpureum*) and pearl millet (*Pennisetum glaucum*), and major cereals such as sorghum, maize and rice[3]. In addition, extensive genetic diversity exists in *Setaria*, with approximately 30,000 accessions preserved in China, India, Japan and the United States[3] as valuable resources for gene-function dissection and elite-allele mining[4]. In recent years, the whole-genome sequences of foxtail millet and green foxtail have been made available[5–9], and both species have been proposed as C_4 model plant systems[3,6]. Between these two species, foxtail millet is more suitable as a model plant due to the seed shattering and dormancy in green foxtail. Nevertheless, the relatively long life cycle (usually 4–5 months per generation) and large plant size (1–2 m in height) limit the use of foxtail millet as a model plant[3,10–12]. To overcome such limitations, we have recently developed a large foxtail millet ethyl methane sulfonate (EMS)-mutagenized population using Jingu21, a high-yield, high-grain-quality elite variety widely grown in north China in the past few decades. From the mutant population, we identified a miniature mutant (dubbed *xiaomi*) with a life cycle similar to that of *Arabidopsis*. Subsequently, we developed genomics and transcriptomics resources and a protocol for efficient transformation of *xiaomi*, as essential parts of the toolbox for the research community.

Results

Creation and phenotypic characterization of *xiaomi*. The *xiaomi* mutant was identified from an EMS-mutagenized M2 population comprising approximately 20,000 mutant lines derived from Jingu21 (wild type (WT)). Under conditions in the field (37° 25′ 13″ N, 112° 35′ 26″ E), *xiaomi* exhibited an extremely early flowering phenotype with a heading date of about 39 d after sowing (DAS) (Fig. 1a and Supplementary Table 1). By contrast, WT plants showed an average heading date of around 82 DAS (Supplementary Table 1). *xiaomi* completed its life cycle in about 70 d, whereas WT plants matured at about 130 DAS (Fig. 1b,c and Supplementary Table 1). *xiaomi* was much smaller in height than the WT (Fig. 1b and Supplementary Table 1), but interestingly, the seed setting rate of *xiaomi* was 12.83% higher than that of the WT (Supplementary Table 1). Nevertheless, no significant difference was observed between *xiaomi* and WT in seed size, as represented by the 1,000-grain weight (Fig. 1d and Supplementary Table 1).

The botanical features of *xiaomi* in growth chambers under different photoperiod conditions were characterized. *xiaomi* headed about one month later under short-day (10 h light:14 h dark) conditions than under long-day (16 h light:8 h dark) conditions, indicating that early heading of *xiaomi* was dependent on the long-day conditions (Fig. 1e). By extending the day length and optimizing other conditions (see Methods for details), we were able to reduce the life cycle of *xiaomi* to 65 d with a plant height of about 29 cm

[1]Institute of Agricultural Bioengineering, Shanxi Agricultural University, Taigu, China. [2]College of Arts and Sciences, Shanxi Agricultural University, Taigu, China. [3]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China. [4]College of Life Sciences, Shanxi Agricultural University, Taigu, China. [5]College of Agriculture, Shanxi Agricultural University, Taigu, China. [6]Division of Plant and Crop Sciences, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, UK. [7]Department of Agronomy, Purdue University, West Lafayette, IN, USA. [8]These authors contributed equally: Zhirong Yang, Haoshan Zhang, Xukai Li. ✉e-mail: suiyi@caas.cn; hanyuanhuai@sxau.edu.cn; wxingchun@sxau.edu.cn

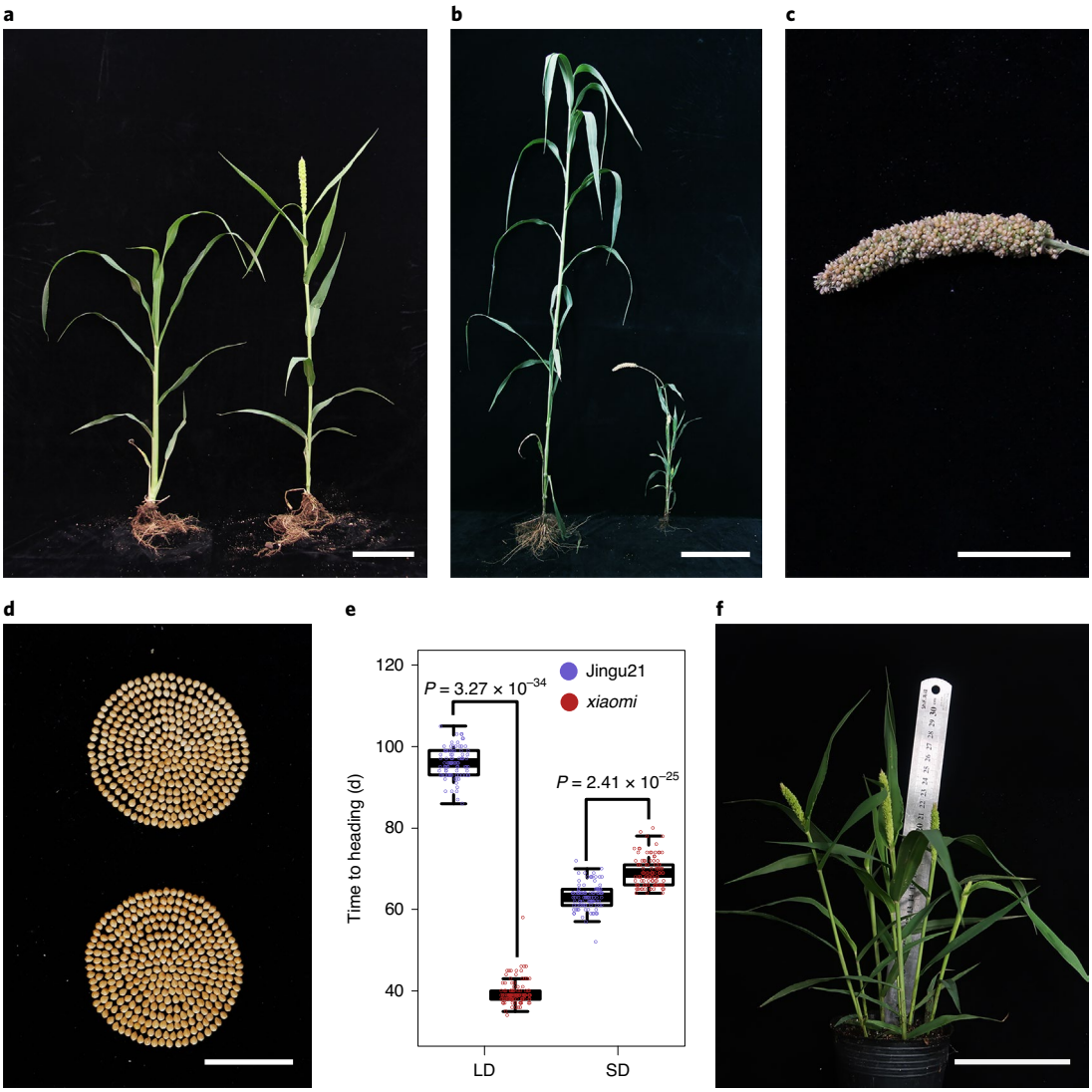


Fig. 1 | Phenotypic characterization of *xiaomi*. **a**, Field-grown Jingu21 (left) and *xiaomi* (right) plants at 40 DAS. **b**, Size comparison of an adult *xiaomi* plant (right) and a Jingu21 plant (left) at 68 DAS. **c**, An enlarged view of the panicle from the *xiaomi* plant shown in **b**. **d**, Seeds collected from the field-grown Jingu21 (top; 310 seeds) and *xiaomi* (bottom; 310 seeds) plants. **e**, Heading dates of *xiaomi* and Jingu21 under long-day (LD) or short-day (SD) conditions. The heading dates of at least 25 plants were measured for each replicate ($n = 3$ biologically independent replicates, ≥89 plants in total). In box plots, the middle line is the median, bottom and top edges represent the first and third quartiles, respectively, and the whiskers represent the maximum and minimum values. Statistical analysis was performed using two-tailed Wilcoxon rank-sum test. **f**, Image of 7 individual *xiaomi* plants in a pot at 39 DAS (10 cm diameter × 10 cm high) grown under the optimized conditions. Scale bars: 10 cm (**a**,**f**), 20 cm (**b**), 5 cm (**c**) and 2 cm (**d**).

(Fig. 1f and Supplementary Table 1). Thus, five to six generations of *xiaomi* can be completed in growth chambers in one year. Because of the small size of *xiaomi*, a set of 1,296 plants can be grown on two planting racks in a three-dimensional space of 1.2 m × 0.55 m × 2.0 m, similar to the space needed to grow an equal number of *Arabidopsis* plants.

A mutation in the *PHYC* gene is causal for the characteristics of *xiaomi*. To identify the causative mutation(s) responsible for early heading of *xiaomi*, we crossed it with G1—a landrace with a heading date of ~75 DAS under the long-day conditions. All 9 F_1 plants exhibited the G1-like late-heading phenotype, and the resulting F_2 populations comprising 268 individual plants showed a segregation ratio of 3:1 ($\chi^2 = 0.318 < \chi^2_{0.05(1)} = 3.841$) for the G1-like late heading date to the *xiaomi*-like early-heading date, suggesting that the early-heading date of *xiaomi* was caused by recessive mutation at a single locus. Using 106 early flowering F_2 individuals, this locus was mapped to a 212-kilobase (kb) region on chromosome 9, which harbours 27 genes, according to the annotation of the *xiaomi* reference genome (Fig. 2a). Comparison between the *xiaomi* genome sequence and the Jingu21 genome sequence, which was generated by genome resequencing, revealed the presence of only a single mutation in the mapped region—a transversion from G to T in the coding region of the gene *Si9g09200* in *xiaomi*, which encodes a putative PHYC protein (Fig. 2a and Supplementary Table 2). Of the 77 early-heading F_2 individuals examined, all were the T/T homozygous mutants. By contrast, of the 256 late-heading individuals examined, 92 were the G/G WT homozygotes and the other 164 were G/T heterozygotes, reflecting a perfect association between the genotypes and phenotypes. This mutation resulted in the creation of a stop codon, causing a truncated protein accounting for approximately 71% of transcripts (transcript 1) of the gene (Fig. 2a–d and Extended Data Fig. 1). Interestingly, this mutation also led to alternative splicing responsible for a frameshift deletion

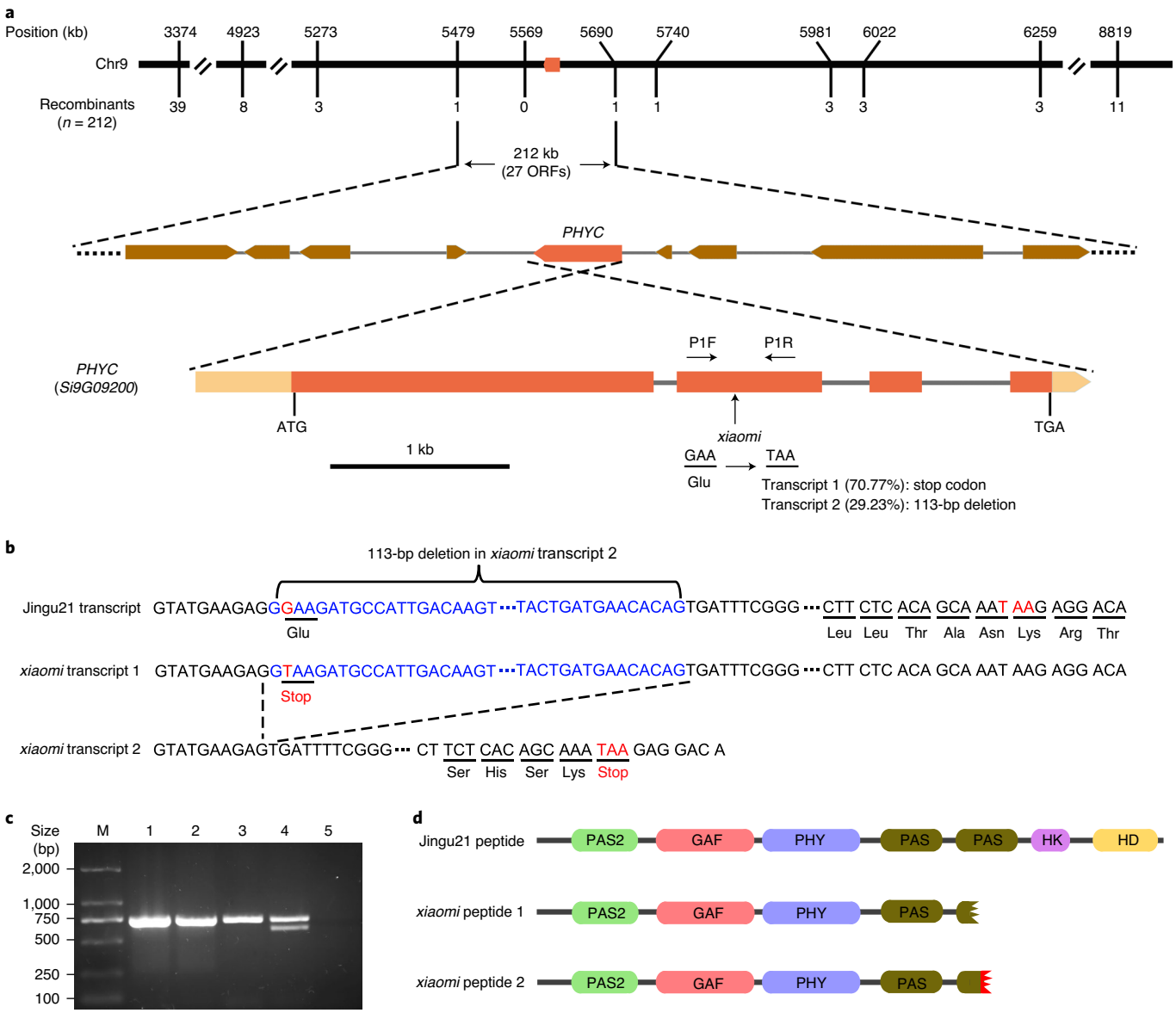


Fig. 2 | Molecular characterization of *xiaomi*. **a**, Genetic mapping of *xiaomi*. Top: schematic showing the positions of the mutation at the *PHYC* locus on chromosome 9 (chr9). The numbers (*n*) of recombinants used in mapping are given below the genetic maps. Middle: putative genes in the mapping region. Bottom: *PHYC* genomic structure, as deduced from its complementary DNA in Jingu21. Exons and introns are denoted by filled boxes and lines, respectively. P1F and P1R represent a pair of primers used to amplify the fragments harbouring the mutation site from the segregating F_2 individuals (primer sequences are listed in Supplementary Table 3). **b**, Normal (*xiaomi* transcript 1) and abnormal (*xiaomi* transcript 2) splicing sequences of *xiaomi* around the G/T mutation site. The red nucleotides represent either the mutation site (G/T) in *xiaomi* or stop codon (TAA) resulted by this mutation. **c**, Detection of the shortened transcripts in *xiaomi* by PCR. Lane 1, genomic DNA from Jingu21; lane 2, genomic DNA from *xiaomi*; lane 3, cDNA from Jingu21; lane 4, cDNA from *xiaomi*; lane 5, water control; M, a 2-kb DNA ladder. Note the presence of a 614-bp fragment in addition to the expected 727-bp fragment in *xiaomi*. The alternative splicing of mRNA in the second leaf at filling stage is presented here and similar results were observed in two other detected tissues including stem and seedling. **d**, Structure of PHYC protein and truncated versions deduced according to mutations in *xiaomi*. GAF, cGMP-specific phosphodiesterase, adenylyl cyclase and FhlA domain; PHY, phytochrome domain.

of 113 base pairs (bp) that also formed a truncated protein accounting for ~29% transcripts (transcript 2) of the gene (Fig. 2a–d and Extended Data Fig. 1). On the basis of the prediction, the truncated proteins lack about two-thirds (peptide 1) or one-third (peptide 2) of the second Per–Arn–Sim (PAS) domain, and the entire histidine kinase A (phospho-acceptor) (HK) domain and histidine kinase, DNA gyrase B and HSP90-like ATPase (HD) domain[13] (Fig. 2d).

We sequenced another mutant, named *xiaomi-2*, which also derived from Jingu21. The *xiaomi-2* mutant showed an early-heading phenotype similar to that of *xiaomi* (Extended Data Fig. 2a–d). Sequence comparison of the *PHYC* locus revealed a single point mutation (T674A) in the first exon of *PHYC* in *xiaomi-2* resulting in a change from a conserved leucine to histone (Extended Data Figs. 2e,f and 3). This single nucleotide polymorphism (SNP) is perfectly associated with phenotypic segregation of 82 early-heading (A/A genotype) and 84 late-heading (49T/A genotype and 35T/T genotype) M3 plants derived from an M2 heterozygous mutant. Together, these observations confirm that the early-heading phenotype resulted from the mutation at the *PHYC* locus.

Table 1 | Statistics for the *xiaomi* genome assembly and annotation

Genome assembly	Estimated genome size		438.26 Mb	
	Assembled genome size		429.94 Mb	
	GC content		45.96%	
	Number of contigs		414	
	Total contig length		429,934,041 bp	
	Longest contig		49,165,788 bp	
	Contig N50		18,838,472 bp	
	Number of scaffolds		366	
	Total scaffold length		429,936,786 bp	
	Longest scaffold		59,244,420 bp	
	Scaffold N50		42,406,388 bp	
Transposable elements	**Annotation**	**Number**	**Length (bp)**	**Percentage (%)**
	Retrotransposon	206,786	187,277,124	43.55
	DNA transposons	132,533	71,648,350	16.67
	Others	51,519	19,295,056	4
	Total without overlaps	390,838	235,013,481	54.66
Predicted genes	Protein-coding genes		34,436	
	Pseudogenes		2,631	
	rRNA		919	
	tRNA		3,516	
	miRNA		340	
	lncRNA		28,260	
	circRNA		1,318	

To understand how the mutation at the *PHYC* locus affects flowering time under the long-day conditions, we performed RNA-sequencing (RNA-seq) analysis using the second leaves from 30-DAS plants (approximately 10 d before *xiaomi* began heading), with WT leaves collected at 30 DAS as a control. We found that the expression levels of several genes orthologous to the *Arabidopsis* oscillator genes *PSEUDO-RESPONSE REGULATORs* (*PRRs*), *PHYTOCLOCK 1* (*PCL1*) and *GIGANTEA* (*GI*), which are critical for photoperiodic flowering, were significantly altered in *xiaomi* (Supplementary Fig. 1 and Supplementary Table 4). As expected, a putative downstream photoperiod gene orthologous to *Ghd7* showed about 95-fold decrease in expression level, whereas the putative flowering genes orthologous to *EARLY HEADING DATE 1* (*Ehd1*), *HEADING DATE 3a* (*Hd3a*) and *APETALA1* (*AP1*)/*FRUITFULL* (*FUL*)-like *MADS* box genes, respectively, exhibited significant increases expression level in *xiaomi* (Supplementary Fig. 1 and Supplementary Table 4). *Ehd1* is a key transcriptional regulator in photoperiodic flowering pathway in rice, which could active the transcription of the florigen *Hd3a* in leaves. The *Hd3a* protein produced in leaves is then transported to shoot apical meristem, where it induces the expression of *AP1*/*FUL*-like *MADS* box genes. Overall, these observations suggested that the early-heading phenotype of *xiaomi* was caused by disruption of the photoperiodic pathways.

Assembly and annotation of the *xiaomi* genome. To facilitate the use of *xiaomi* as a model plant, a total of 41.54 Gb (94.78× coverage) high-quality single-molecule real-time (SMRT) subread sequences were generated and assembled into 429.45 Mb of scaffold sequences, with a contig N50 of 19.85 Mb (Supplementary Tables 5–7). Of these, 399.40 Mb of scaffold sequences were anchored to nine super-scaffolds (chromosomes) with 137.33 million Hi-C-based paired-end reads (Extended Data Fig. 4 and Supplementary Table 8). After removing scaffolds of less than 1 kb in length, our final assembly, designated *xiaomi* genome v.1.0, contained 429.94 Mb of sequence, comprising 366 scaffolds with an N50 length of 42.41 Mb and 48 gaps (Table 1). *k*-Mer analysis suggests that the draft assembly covers approximately 98.10% of the entire genome (Supplementary Fig. 2). The error rate of the assembly is about 0.001% (one error per 100 kb), as estimated by Illumina DNA short reads. Single-copy orthologue analysis showed that 97.78% of the 1,440 benchmarking universal single-copy orthologs (BUSCO) genes were completely covered by the *xiaomi* genome, with only 0.90% incomplete and 1.32% not assembled or annotated (Supplementary Table 9). Collectively, these results indicate that the *xiaomi* genome v.1.0 can be used as a gold-standard reference by the research community.

Using a combination of de novo prediction and homology-based comparison, a total of 237.28 Mb (55.19%) of the *xiaomi* genome sequences were annotated as repetitive elements (Table 1 and Supplementary Table 10). We annotated 34,436 protein-coding genes using 671,853 full length non-chimeric (FLNC) isoform reads produced by PacBio RS II and approximately 1,054.5 million short RNA-seq reads produced by the HiSeq X Ten platform, and a combination of ab initio prediction and protein-homology-based searches (Table 1 and Supplementary Table 11), of which 32,743 (95.08%) were located in the nine pseudochromosomes (Supplementary Table 12). These genes were searched against Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Eukaryotic Orthologous Groups (KOG), TrEMBL and nr databases and compared with the annotation of *Arabidopsis* and rice to retrieve homologues with known functions and a total of 33,789 genes (98.12%) were annotated (Supplementary Table 13). In addition, we annotated 919 rRNA genes, 3,516 transfer RNA genes, 2,631 pseudogenes, 340 microRNA (miRNA) precursors, 28,260 long non-coding RNA (lncRNA) precursors, and 1,318 circular RNA (circRNA) precursors (Table 1).

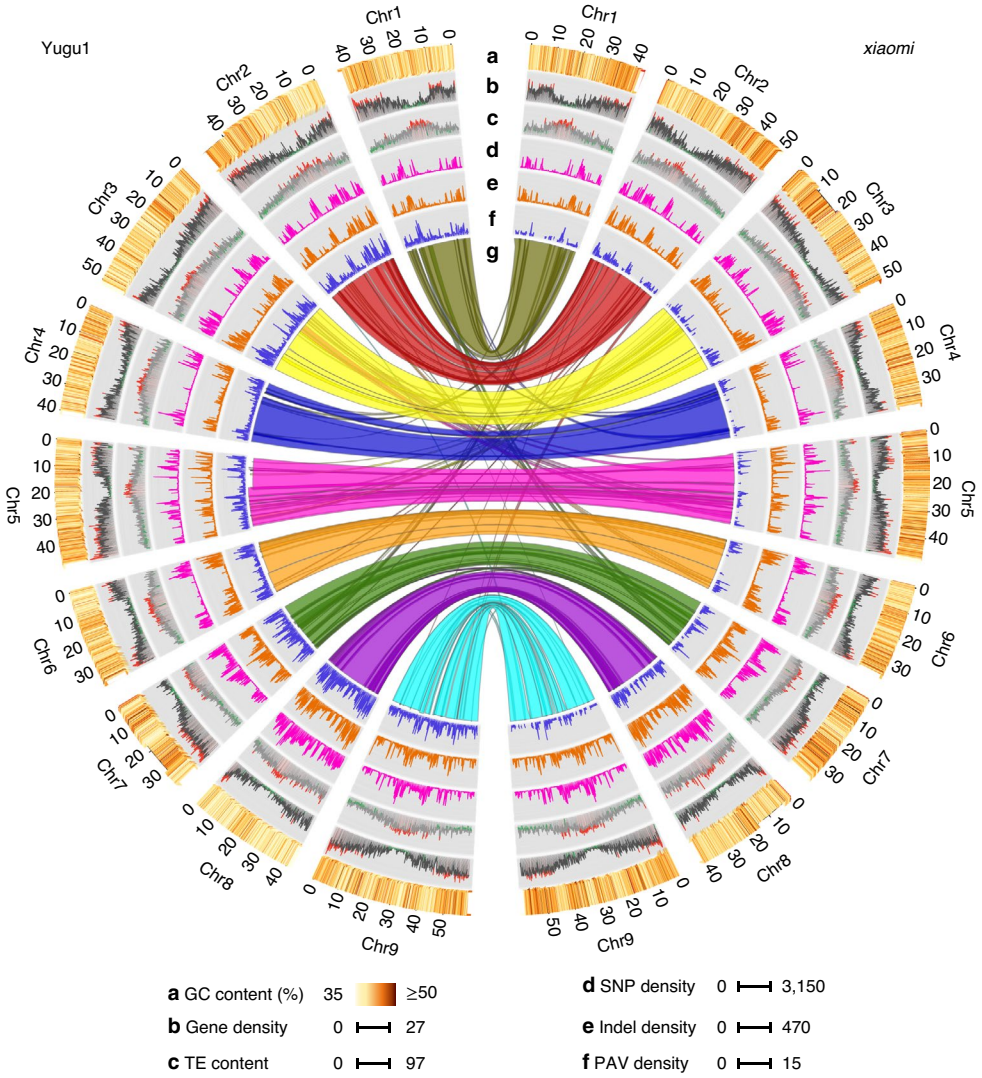


Fig. 3 | Circular plot of the *xiaomi* genome sequence compared with the Yugu1 genome. a–**f**, GC content (**a**), gene density (**b**), transposon element (TE) content (**c**), SNP density (**d**), indel density (**e**) and presence/absence variation (PAV) distributions (**f**) in sliding windows of 100 kb. **g**, Links display homologous genes in *xiaomi* and Yugu1. Bottom: scale bars for **a**–**f**.

All the genomic and transcriptomic data are publicly accessible through our user-friendly database (http://sky.sxau.edu.cn/MDSi.htm). In this database, researchers can navigate the genome by chromosome coordinates, gene or transcript symbols, or by BLAST search against the *xiaomi* genome, coding sequences or peptide sequences.

Comparison of genome sequences from *xiaomi* and other foxtail millet varieties. Compared with the three previously released genome sequences from Yugu1[5], Zhanggu[8] and TT8[9] varieties, the *xiaomi* genome showed the highest quality in terms of the genome coverage, contig N50 values, and contig and gap numbers (Supplementary Table 14). Intergenomic comparison revealed that 414.58 Mb (96.44%) of the *xiaomi* sequences correspond to 383.52 Mb (95.67%) of the Yugu1 sequences, with 1,577,935 SNPs (Supplementary Data 1). The size difference in the corresponding regions was mainly caused by 259,731 small indels of less than 100 bp, 2,804 (total 15.32 Mb) presence variations (over 1,000 bp) and 2,722 (total 17.38 Mb) absence variations (over 1,000 bp) between *xiaomi* and Yugu1 genomes (Fig. 3, Supplementary Data 1 and Supplementary Tables 15 and 16).

Of the annotated 34,436 protein-coding genes in the *xiaomi* genome, 32,112 genes (93.25%) are shared by the Yugu1 genome (v.2.2), with 2,324 predicted genes in *xiaomi*, of which 1,215 genes are supported by RNA-seq data (the maximum transcripts per million (TPM) > 0), absent in Yugu1 (Supplementary Table 17). A total of 2,280 predicted genes in *xiaomi* were not found in the Zhanggu genome (Supplementary Table 18). Only 1,030 genes in *xiaomi* were absent in both Yugu1 and Zhanggu; these were considered to be *xiaomi*-specific (Supplementary Table 19). A marked phenotypic difference between *xiaomi* and Yugu1 is their susceptibility and resistance, respectively, to downy mildew (Supplementary Fig. 3), but it is unclear whether this difference is associated with any of the detected variety-specific genes.

Construction of a dynamic gene-expression atlas for *xiaomi*. To develop a reference gene-expression atlas for functional interpretation and investigation of gene function, we measured transcript levels in 11 diverse tissues representing the major organs over various developmental stages of *xiaomi* (see Methods for details). A total of 1,054.51 M raw reads (~30 M reads per sample) were produced and analysed. A total of 31,226 (90.68%) genes were expressed in

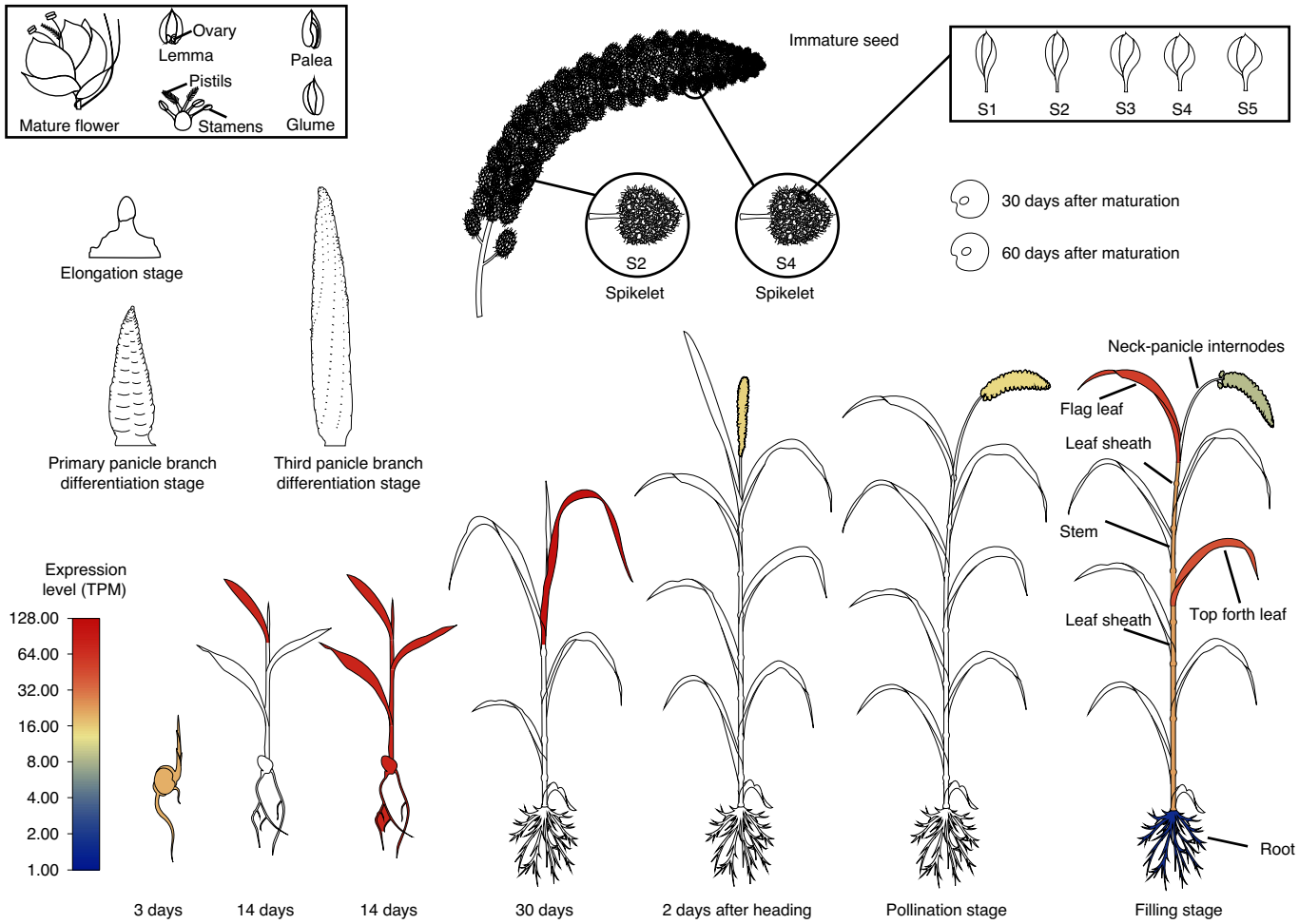


Fig. 4 | Expression pattern of the *Si9g04830* gene. The *Si9g04830* gene encodes a magnesium chelatase D subunit that catalyses the insertion of magnesium into protoporphyrin IX in chlorophyll biosynthesis. Expression data in the coloured tissues are described in this paper. The highest expression of *Si9g04830* is seen in the leaves, consistent with its known biological role and with published data[60]. A scale bar is shown on the left. The expression data of the tissues shown in greyscale are being analysed or are about to be analysed, and will be presented in the MDSi database in the near future.

at least one of the 11 *xiaomi* tissues (Supplementary Fig. 4a and Supplementary Table 20). The proportions of genes with expression detected in individual tissues ranged from 74.26% in the second top leaf (leaf 2) to 82.95% in the panicle at the pollination stage (panicle 2). A total of 22,202 genes were expressed in all 11 *xiaomi* tissues. Of these genes, 85 (0.25%), including one transcription initiation factor gene (*Si3G07600*) and two ubiquitin-conjugating enzyme coding genes (*Si1G37980* and *Si2G05250*), were constitutively expressed in all assayed tissues (Supplementary Table 21). These genes would be useful for transcript normalization before comparative gene-expression analysis. Moreover, we identified 1,218 organ- or tissue-specific genes and 1,226 organ- or tissue-preferentially expressed genes (Supplementary Figs. 4b,c and 5 and Supplementary Tables 22 and 23).

To make these expression data more user friendly, we developed a *xiaomi* Electronic Fluorescent Pictograph (xEFP) browser (http://sky.sxau.edu.cn/MDSi.htm). Using the xEFP browser, gene-expression data can be displayed with idealized images (Fig. 4).

Establishment of an efficient *Agrobacterium*-mediated genetic transformation system. To pave the way for functional genomics studies, we tested various factors to develop an *Agrobacterium*-mediated transformation protocol for *xiaomi*. We challenged mature seeds as a starting material for callus induction to avoid the costly need for growing plants if fresh tissues such as the young inflorescence or immature embryos are used for callus induction[14]. After a series of trials, we observed that primary calli were not suitable for use in transformation, mainly due to the soft texture. Following three rounds of subculture on an improved callus-induction medium (CIM), however, compact embryogenic calli were obtained (Fig. 5a). We used the green fluorescent protein (GFP) reporter gene to monitor *Agrobacterium* infection efficiency, as indicated by multiple green spots in the calli (Fig. 5b,c), and effectiveness for selecting out the transgenic callus (Fig. 5d,e). The regeneration ability of *xiaomi* was well maintained on the CIM during subculture (Fig. 5f). Roots of transgenic plants expressing GFP could be induced easily on the rooting medium and rooted plants survived well after transplanting to soil (Fig. 5g–i). We compared two commonly used selectable markers neomycin phosphotransferase II (NPT-II) and hygromycin phosphotransferase (HPT), and obtained transformation efficiency ranging from 8.05% to 38.75%, with an average of 23.28% for *NPTII*, and from 3.08% to 16.67%, with an average of 8.72% for *HPT* (Supplementary Table 24). We then confirmed the presence of transgenes in primary putative transgenic plants (T0) by PCR using primers to amplify the GFP gene, *UBI* promoter, and *HPT* or *NPTII* selectable marker genes (Fig. 5j and Supplementary Fig. 6). We also observed GFP expression in both

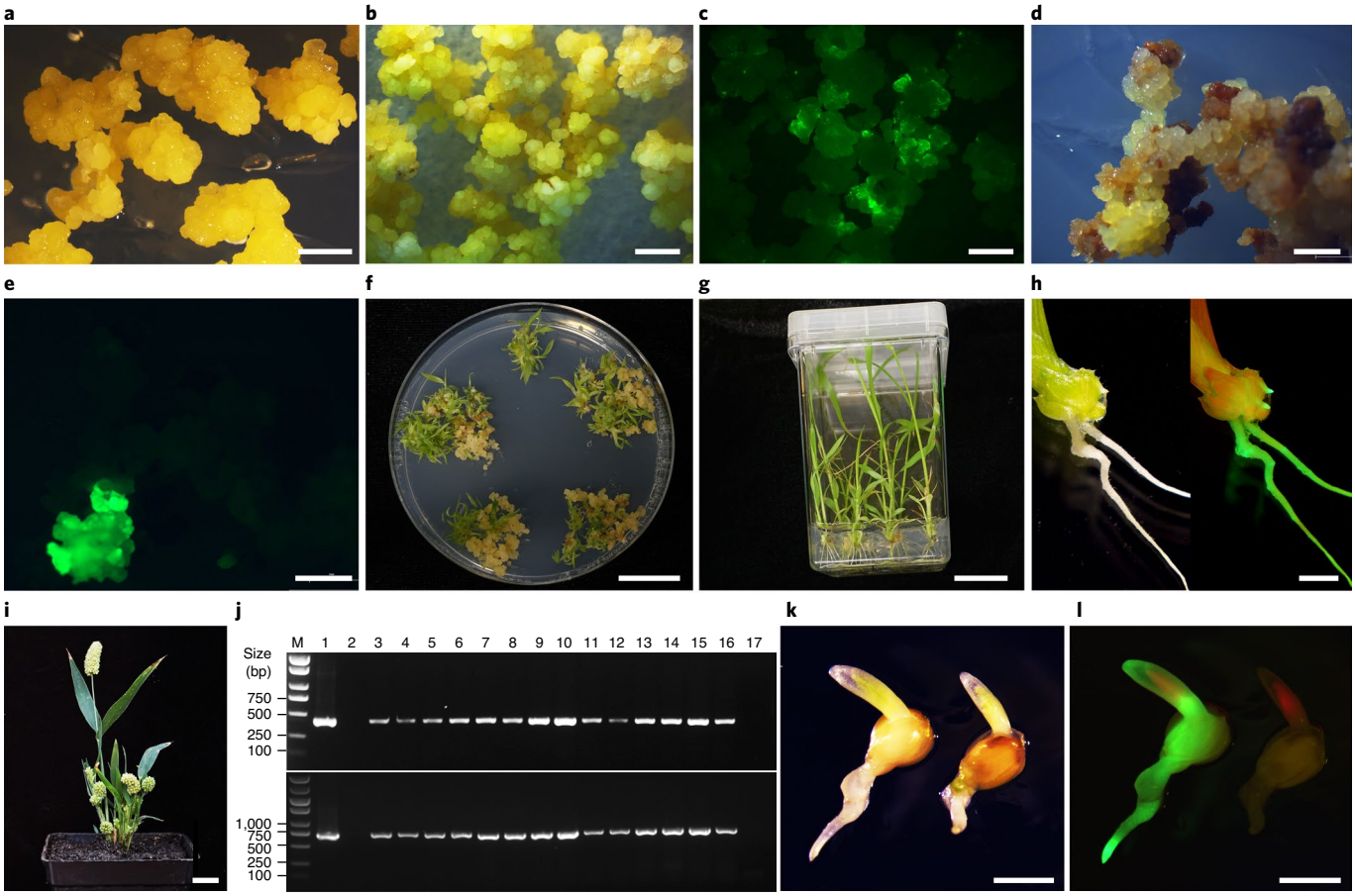


Fig. 5 | *Agrobacterium*-mediated transformation of *xiaomi*. **a**, Embryogenic calli suitable for transformation, two months after seed inoculation. **b**, Calli co-cultivated with *Agrobacterium* for 3 d under bright light. **c**, UV visualization of infected calli in **b**, showing transient expression of the GFP reporter. **d**, A transformed callus sector (pale yellow) proliferating on selection medium at the end of the second round of selection. **e**, The same callus as in **d** visualized under UV light. **f**, Shoot regeneration from transgenic calli. **g**, Root formation on root-induction medium. **h**, A healthy GFP-expressing plant imaged under white (left) and UV (right) light. **i**, An adult primary transgenic plant. **j**, PCR confirmation of the transgenic plantlets generated with the *HPT* selectable marker using specific markers for the *GFP* gene (top) or *UBI* promoter (bottom). M, molecular marker; lane 1, plasmid DNA; lane 2, non-transformed *xiaomi* plant; lanes 3–16, independent T0 transformants; lane 17, water control. **k**,**l**, Segregation of *GFP* transgene in germinated seeds under white (**k**) or UV (**l**) light, confirming transmission of transgene to progeny. All experiments were performed with eight independent biological repeats and at least six samples were tested for each biological repeat except **j**, which was performed with three repeats. Scale bars, 2 mm (**a**–**e**,**h**,**k**,**l**) and 2 cm (**f**,**g**,**i**).

dry and germinating seeds, indicating transmission of the transgene to progeny (Fig. 5k,l and Extended Data Fig. 5). Insertions of transgenes into the *xiaomi* genome were verified by genome sequencing of 13 independent transgenic lines produced with the *HPT* or *NPTII* marker, and insertion sites of transfer DNA (T-DNA) were further confirmed by PCR in three lines examined (Extended Data Fig. 6 and Supplementary Table 25). We grew T1 plants representing eight transgenic events in pots and observed no obvious phenotypic differences from non-transformed *xiaomi* plants (Supplementary Fig. 7). Collectively, we demonstrated the transgenic nature of the plants generated by *Agrobacterium*-mediated transformation. Thus, we have established an efficient protocol that allows production of transgenic plants ready to transplant to soil in 2–3 months from *Agrobacterium* infection or 4–5 months from callus initiation from mature seeds (Fig. 6).

Discussion

Foxtail millet is an emerging C_4 model plant suitable for investigation of various biological phenomena that are absent in other model plants such as *Arabidopsis* and rice. In this study, we identified *xiaomi*, a mutant with reduced stature and short generation time, created a reference genome and a gene-expression atlas from various tissues, and developed a highly efficient transformation protocol. The results demonstrate the suitability of *xiaomi* as a model system to investigate C_4 grass biology and other important molecular mechanisms including, but not limited to, higher nitrogen use efficiency, abiotic and biotic stress responses, downy mildew resistance, domestication and evolution.

Compared with C_3 species, C_4 plants usually show higher rates of photosynthesis as well as higher nitrogen and water efficiencies. The most productive crop species, such as maize, sorghum and sugarcane, are C_4 plants, and C_4 plants contribute to about a quarter of global primary biomass production despite comprising only 3% of all land plant species[15]. Due to such high productivity, introducing C_4 pathway genes into major C_3 crops such as rice seems to be a promising strategy to meet the growing demand for food production[16]. Towards implementation of such a strategy, it is essential to elucidate genetic and molecular mechanisms underlying the differentiation of C_3 and C_4 anatomical, physiological and biochemical features. Among C_4 plants, maize and sorghum are the major contributors to world food production, whereas sugarcane and switchgrass are major bioenergy plants. However, all these plants possess relatively large statures, large genomes, long life cycles, and are difficult to transform. About ten years ago, Brutnell et al.[6]

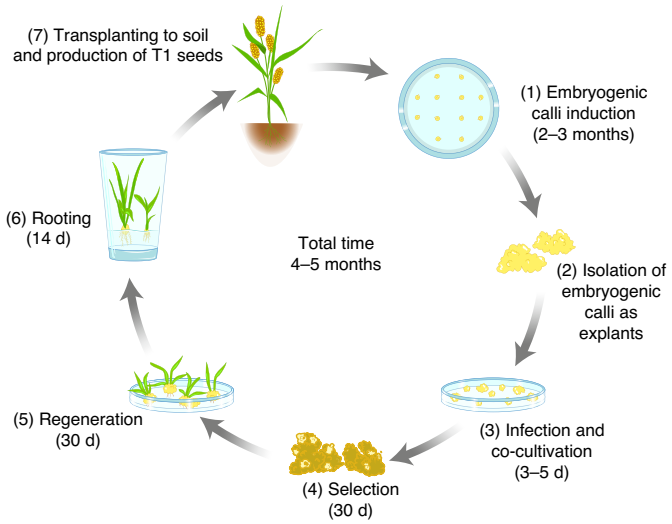


Fig. 6 | Schematic of *Agrobacterium*-mediated transformation of *xiaomi*. The described procedure for *xiaomi* transformation.

proposed green foxtail as a C_4 model plant, considering the advantages of its relatively short stature, simple growth requirements and rapid life cycle. Since then, great progress has been made in genome assembly, transformation technology and germplasm generation as well as mutant isolation and characterization in green foxtail[5,17–20]. Compared with its wild progenitor green foxtail, foxtail millet is more suitable as a model plant. First, the seeds of foxtail millet are generally non-shattering and non-dormant, and easier to collect and germinate; second, foxtail millet has been widely cultivated for both human food and animal feed in the arid and semi-arid regions of the world, particularly in China and India, and would be easier to deploy for grain production. These characteristics, together with the short life cycle and small plant size, makes *xiaomi* an ideal model plant to accelerate research in millet and other C_4 plants.

We acknowledge that there remain limitations to the direct use of *xiaomi* for certain studies. For example, it may not be suitable for directly evaluating grain yield of foxtail millet cultivars in the field condition, although such traits may be dissected into individual yield-related components such as grain size, 1,000-grain weight and seed number per panicle. Nevertheless, some of the limitations may be at least partially overcome by growing *xiaomi* under short-day conditions or crossing *xiaomi* plants with WT plants to produce progeny for use in subsequent investigations of the traits of interest.

Other early flowering mutants, such as Xiaowei[21] in rice and Micro-Tom[22] in tomato, have been used for conducting large-scale indoor research. Similar to those of *xiaomi*, the phenotypic changes of Xiaowei were caused by the deficiency of a haem oxygenase involved in the biosynthesis of a phytochrome chromophore[21]. Indeed, accelerated flowering under noninductive photoperiods was also observed in the *phyC* mutants in *Arabidopsis*[23] and rice[24]. Thus, it is highly probable that *xiaomi*-like mutants can be created using any millet variety by editing the *PHYC* gene.

At present, around 10% of *xiaomi* genes are not captured in the tissues used for construction of the gene-expression atlas; nevertheless, the majority of these 'unexpressed' genes have homologues or orthologues in *Arabidopsis* and rice, suggesting that they would be expressed in other tissues, at other developmental stages, or under specific growth conditions, and additional RNA-seq analysis should enable construction of a more comprehensive gene-expression atlas in the future.

Transformability is an essential prerequisite for a model plant. *Arabidopsis* can be efficiently transformed by floral dipping, which has enabled its rapid adoption for basic research in plant biology worldwide. Recently, a similar approach (spike dip transformation) has been explored in green foxtail with reported success[19,25]. However, we were unable to recover any transgenic plant from *xiaomi* using this method. Transgenic plants were successfully produced using calli induced from immature embryos or inflorescences in both foxtail millet and green foxtail, although at low efficiency[26,27]. The disadvantage of using fresh tissues is that plants must be grown periodically to ensure a constant supply all year round, which is time consuming and costly. Thus, we attempted to use mature seeds as an explant source for callus induction. In repeated trials, we observed that the primary calli induced from mature embryos were not suitable for direct use in transformation, most probably due to their watery and soft nature. We then focused our efforts on the development of embryogenic calli by subculture and optimization of the infection and selection steps by monitoring expression of the reporter GFP. We also compared selectable markers and identified *NPTII* as an efficient marker for foxtail millet transformation. The transformation method established in this study has a 3.5-fold higher efficiency than the previously reported method[26], and with further improvement should encourage broad adoption of *xiaomi* as a model for basic and applied research, especially in C_4 plants.

Methods

Plant materials and growth conditions. *xiaomi* was identified from an EMS-mutagenized M2 population of Jingu21, a variety of foxtail millet widely cultivated in North China for its good grain quality and high yield. The *xiaomi* mutant was maintained by self-pollination in the laboratory for ten generations, leading to a very low level of heterozygosity. Foxtail millets were grown in the experimental field in Taigu, Shanxi, China (37° 25′ 13″ N, 112° 35′ 26″ E). For indoor research, plants were grown in an auto-controlled growth chamber or culture room equipped with full spectrum (420–730 nm) LED light sources, under 28 °C:22 °C day:night cycle with a 14 h photoperiod and 350–700 μmol m^{-2} s^{-1} light intensity unless otherwise specified. To shorten the life cycle and reduce plant stature, we optimized growth conditions for *xiaomi*. In brief, *xiaomi* seeds were soaked in water overnight at room temperature and sown in a soil mix of nutrient soil, sandy soil and vermiculite (3:2:1, v/v/v) watered with B5 solution (water content approximately 25%, w/w). Plants were grown under 16 h photoperiod and watered to maintain 10%–15% water content.

For genome sequencing, the aboveground tissues, including leaves, stem and young panicle were collected from a single healthy *xiaomi* plant at the pollination stage for PacBio SMRT DNA sequencing. Young leaves from a single healthy *xiaomi* or WT plant were collected for genome resequencing.

For the expression atlas sequencing, 11 diverse tissues representing the major organ systems were collected, with 3 or 5 biological replications. These tissues were 3 d imbibed seeds (seed), 2-week-old whole seedling (seedling), root, stem, the top first fully extended leaf of 2-week-old seedling (leaf 1), the top second leaf of 30-day-old plants (leaf 2), flag leaf (leaf 3), the fourth leaf (leaf 4), immature panicle (panicle 1), panicle at pollination stage (panicle 2) and panicle at grain-filling stage (panicle 3). For seed germination, the surface-sterilized seeds were placed on Whatman no. 1 filter paper soaked with distilled water and cultured for 3 d, allowing them to germinate. For the 2 week seedling stage, the seeds were sown in soil and the whole seedlings (seedling) and the first immature leaves (leaf 1) were sampled at 2 weeks after germination. Leaf 2 is the top second leaf of 30 d *xiaomi* seedlings (10 d before heading). Samples of stem, leaf 3 (flag leaf), leaf 4 (the top fourth leaf) and panicle 3 were all collected at the grain-filling stage. Each biological replicate included at least five healthy *xiaomi* plants randomly selected from the field or auto-controlled growth chamber. All samples were immediately frozen in liquid nitrogen and stored until use.

Map-based cloning. We crossed *xiaomi* with the cultivar G1 to generate a F_2 mapping population. Using 45 recessive F_2 plants with the typical *xiaomi*-like early-heading phenotype, we firstly mapped the *XIAOMI* locus to a 5.45 Mb interval between the two indel markers, M3374 and M8819 on chromosome 9. We further developed 9 new markers within this interval and finally narrowed down the locus to a 212-kb region between two SNP markers, M5479 and M5690. A candidate gene was then identified by genome resequencing of and comparison of this region between Jingu21 and *xiaomi*. Sequences of all primers used in map-based cloning are listed in Supplementary Table 3.

DNA and RNA isolation. For PacBio single-molecule sequencing, DNA was extracted from a single healthy *xiaomi* plant as described in the 'Preparing *Arabidopsis* Genomic DNA for Size-Selected ~20 kb SMRTbell Libraries' protocol (http://www.pacb.com/wp-content/uploads/2015/09/

Shared-Protocol-Preparing-Arabidopsis-DNA-for-20-kb-SMRTbell-Libraries.pdf). For Illumina HiSeq sequencing, DNA was isolated from leaf tissues using cetyltrimethylammonium bromide methods[28] with modifications. About 100 mg young leaf was ground to a fine powder in liquid nitrogen. The powder was then placed in 2 ml microtubes containing 1 ml preheated 2% cetyltrimethylammonium bromide extraction buffer (adding 0.5% β-mercaptoethanol just before use) and incubated at 65 °C for 30 min. The samples were then centrifuged and the resultant supernatant was extracted with 800 µl chloroform: isoamyl alcohol (24:1, v/v). The supernatant DNA was transferred to a new microtube containing 800 µl cold isopropanol and 80 µl 3 mol l^{-1} sodium acetate to precipitate the DNA. The precipitate was dissolved in 100 µl ddH_2O containing 10 ng $µl^{-1}$ RNase and incubated at 37 °C for 30 min. Finally, the DNA was isolated using magnetic beads. The quality and integrity of extracted DNA was assessed with a Qubit Fluorometer (Life Technologies) and separated in 0.8% agarose gels.

Total RNA was isolated with RNAprep Pure Plant Kit (Tiangen Biotech) or Plant RNA kit (OMEGA) according to the manufacturer's instructions. The integrity and quantity of extracted RNA were analysed on an Agilent 2100 bioanalyzer and by agarose gel electrophoresis.

Genome-sequencing library construction, PacBio SMRT and HiSeq sequencing. The DNA libraries for PacBio SMRT sequencing were prepared following the PacBio standard protocols and sequenced on a Sequel platform by Biomarker Technologies. In brief, genomic DNA from a single *xiaomi* plant was randomly sheared to an average size of 20 kb, using a g-Tube (Covaris). The sheared gDNA was end-repaired using polishing enzymes. After purification, a 20-kb insert SMRTbell library was constructed according to the PacBio standard protocol with the BluePippin size-selection system (Sage Science) and sequences were generated on a PacBio Sequel (9 cells) and PacBio RS II (1 cell) platform by Biomarker Technologies.

Illumina HiSeq DNA libraries were made following standard protocols provided by Illumina. About 5 µg extracted DNA was fragmented randomly and DNA fragments of the desired length were gel purified. These DNA samples were end-repaired and ligated to the adapter and were then pooled, purified and amplified with primers compatible with the adapter sequences, and used to construct a 270-bp paired-end library. The library was sequenced on an Illumina HiSeq X Ten sequencing platform by Biomarker Technologies.

PacBio assembly, correction and validation. The single-molecule sequencing data were assembled following a hierarchical approach, with correction, assembly and polishing[29]. In brief, a subset of longer reads was selected as seed data and corrected through Canu[30] (v.1.5) and Falcon[31] (v.0.3.0). The error-corrected reads were assembled using Falcon and Canu. Since the Canu and Falcon assemblies each contained some regions that were missing from the other, the two initial assemblies were merged using Quickmerge v.0.2 (https://github.com/mahulchak/quickmerge) to produce a more contiguous assembly. Finally, the draft assembly was polished to obtain the final assembly. The first-round polishing adopted the quiver/arrow algorithm using SMS data with the 40 threads. The second polishing adopted the pilon algorithm (v1.22, https://github.com/broadinstitute/pilon) using Illumina HiSeq-sequencing data.

Hi-C library preparation, sequencing, and raw read processing. The Hi-C library was prepared as described previously[32] with minor modifications. Nuclear DNA was cross-linked in situ with formaldehyde, extracted and then digested with HindIII at 37 °C overnight. After digestion, the sticky ends were filled in, biotinylated and then ligated to each other randomly to form chimeric circles. Biotinylated DNA fragments were reverse cross-linked with proteinase K and purified by phenol extraction, followed by a phenol/chloroform/isoamyl alcohol extraction. Then, the purified DNA was sheared to a size of 300–700 bp with a Covaris S220 instrument (Covaris). The sheared DNA was end-repaired with T4 DNA polymerase. The biotin-tagged ligation products were isolated with MyOne Streptavidin C1 Dynabeads (Life Technologies). Bead-bound Hi-C DNA was amplified and purified for preparing the sequencing library. Finally, the Hi-C library was paired-end sequenced on an Illumina HiSeq X Ten platform.

The Hi-C reads were aligned to the draft assembly using the BWA aln algorithm[33] with default parameters, and the quality was then assessed using HiC-Pro v.2.8.0 (http://github.com/nservant/HiC-Pro). The invalid interaction pairs, including self-circle ligation, dangling ends, PCR duplicates and other potential assay-specific artefacts were discarded. The unique valid interaction pairs (non-redundant, true ligation products) were uniquely mapped onto the draft assembly contigs, which were grouped into 9 chromosome clusters, and scaffolded by Lachesis[32] using the following parameters: cluster min re sites = 52, cluster max link density = 2; cluster noninformative ratio = 2; order min n res in trun = 46; order min n res in shreds = 42.

Repeat annotation, gene prediction and functional annotation. For the repeat annotation of the *xiaomi* genome, both structural predictions and de novo approaches were adopted. Specifically, the primary repeat library of *xiaomi* was built from the de novo approach using LTR_Finder[34] (v.1.05), MITE-Hunter[35] (v.1.0.0), RepeatScout[36] (v.1.0.5) and PILER-DF[37] (v.2.4) with the default parameters. Secondly, the primary repeat library was classified with PASTEClassifier[38] (v.1.0) and then combined with Repbase[39] to build the final repeat library of *xiaomi*. Finally, repeats throughout the *xiaomi* genome were identified by RepeatMasker (v.4.0.6) with the parameters '-nolow -no_is -norna -engine wublast -qq -frag 20000'.

For predicting genes, a combination of ab initio-based approaches, homology-based methods and supporting PacBio isoform sequencing (Iso-seq) were used to conduct a comprehensive search for consensus gene sets. For ab initio-based gene prediction, five gene-finding programs, Genscan[40] (v.3.1), Augustus[41] (v.2.4), GlimmerHMM[42] (v.1.2), GeneID[43] (v.1.4) and SNAP[44] (v.2006-07-28) were used to detect genes in the repeat masked *xiaomi* genome with the default parameters. For homology-based prediction, proteins previously annotated in *A. thaliana*, *S. italica*, *O. sativa*, *S. bicolor* and *Z. mays* were downloaded and mapped to the *xiaomi* genome using BLAST and homologous genes were identified using GeMoMa[45] (v.1.3.1). Newly generated *xiaomi* PacBio Iso-seq and RNA-seq data were directly mapped to the *xiaomi* genome and assembled by PASA[46] (v.2.0.2). Finally, the results obtained from the above approaches were integrated into a consensus gene set of *xiaomi* using EVM (v.1.1.1) with default parameters. These protein-coding genes were named using the following gene model nomenclature: Si (for *S. italica*) followed by the chromosome number and gene number on the chromosome, going from top to bottom in steps of 10. For example, the first and the second genes on chromosome 1 were named *Si1G00010* and *Si1G00020*, respectively.

For the functional annotation of gene models of *xiaomi*, the final protein-coding regions were aligned to sequences in public databases including nr (https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/), KOG (ftp://ftp.ncbi.nih.gov/pub/COG/KOG/), KEGG (https://www.kegg.jp/) and TrEMBL (https://www.ebi.ac.uk/uniprot) using BLAST (v.2.2.31, *E*-value $\leq 1.0 \times 10^{-5}$). The GO[47] terms for each gene were obtained using the Blast2GO program[48] based on the above nr annotation.

The tRNA genes in the assembly were identified by tRNA scan-SE[49] (v.1.3.1) with eukaryote parameters. rRNA and miRNA were identified by searching Rfam[50] (v.12.1) with an *E*-value threshold of 1.0×10^{-5}.

Pseudogene GeneWise (v.2.4.1) was used to predict the candidate gene structure on the basis of the homogenous alignments. We filtered the GeneWise results to retain only those with at least 95% coverage of the protein. The gene structures with frameshift mutations were considered to be candidate pseudogenes.

Genome quality evaluation. The quality of the *xiaomi* assembly was assessed by examining the alignment ratio of HiSeq short reads and the presence of well-conserved core eukaryotic genes. The short reads generated by the Illumina HiSeq platform were aligned to the *xiaomi* assembly using BWA (v.0.7.10-r789). To further evaluate the completeness of the *xiaomi* gene models, BUSCO[51] (v.2.0) analysis was undertaken with genome mode and embryophyta_odb9 dataset (http://busco.ezlab.org/datasets/embryophyta_odb9.tar.gz) as a reference. The embryophyte_db9 dataset contains 1,440 protein sequences and orthologous group annotations for major clades. The proportion of complete and partial core eukaryotic genes was assessed as a measure of the completeness of the *xiaomi* assembly.

RNA-sequencing library preparation, Iso-seq and HiSeq sequencing. For Iso-seq, eight tissues, including seed, seedling, root, stem, young leaf (leaf 1), mature leaf (leaf 3), pollinated panicles (panicle 1) and panicles at the filling stage (panicle 3), were collected for RNA isolation. Equal amounts of total RNA from each tissue were pooled together to identify as many isoforms as possible. SMRTbell libraries were prepared according to the Iso-seq protocol using the Clontech SMARTer PCR cDNA Synthesis Kit and the BluePippin size-selection system (Sage Science). The first cDNA strand was synthesized using SMARTer PCR cDNA Synthesis Kit (Takara Biotechnology). After cycle optimization, large-scale PCR was performed to generate double-stranded cDNA for size selection on the BluePippin system (Sage Science). Then, another large-scale PCR was performed using the eluted DNA to generate more double-stranded cDNA. Re-amplified cDNA was purified, repaired and ligated with hairpin adapters. To minimize the bias that favours sequencing of shorter transcripts, multiple size-fractionated libraries (0–1, 0.5–1, 1–2, 1–3, 2–3 and 2–8 kb) were constructed according to the manufacture's instruction. Finally, a total of 15 SMRT cells were sequenced on a PacBio RS II platform.

For transcriptome atlas sequencing with the Illumina HiSeq X Ten platform, RNA-seq libraries were constructed using the NEBNext Ultra RNA Library Prep Kit for Illumina (no. E7770, New England BioLabs) according to the manufacturer's instructions. In brief, mRNA was purified from total RNA using NEBNext Poly (A) mRNA Magnetic Isolation Module (no. E7490, New England BioLabs) and fragmented into approximately 200-nt RNA short fragments. The fragmented mRNAs were then used as templates to synthesize the first-strand and the second-strand cDNAs. After end repair and adapter ligation, the products were selected by Agencourt AMPure XP beads (Beckman Coulter) and amplified to create a cDNA library by PCR. In total, 35 RNA-seq libraries were made from 11 different tissues with five biological replicates for leaf 2 and three biological replicates for others. All libraries were sequenced using an Illumina HiSeq X Ten platform by Biomarker Technologies.

RNA-seq read processing, clustering analyses, *Z*-score and coefficient of variation expression analysis. Illumina RNA-seq reads of *xiaomi* were cleaned using Trimmomatic[52] (v.0.38) with parameters 'ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:30 HEADCROP:10'. The clean reads were mapped to the *xiaomi* genome using HISAT2[53] (v.2.04) with default parameters. Gene-expression analysis and quantile normalization were conducted using R with the TPM (ref. [54]). Genes with TPM > 0 in a given organ were considered to be expressed in that organ. Tissue-preferential and -specific genes were identified according to their TPM. Fold changes greater than ten between tissues showing the highest and the second highest expression levels were considered to be tissue-specific genes, whereas fold changes of at least 5 but no more than 10 were considered as tissue-preferentially expressed genes. Constitutively expressed genes were identified by coefficient of variation (CV) analysis. CVs ranged from 10.30% to 331.66%, representing the most stably expressed genes to the most differentially expressed genes. A gene with CV < 20% and a difference of no more than twofold between the highest and lowest levels of a gene transcript in any organ was considered to be constitutively expressed.

Hierarchical clustering analysis was performed using the pheatmap package (v.1.0.12) in R. Distance analysis was calculated using pairwise Pearson correlation.

Non-coding RNA isolation, library preparation, sequencing and sequence data analysis. The miRNAs were isolated using the EASYspin Plant microRNA Kit (RN4001, Aidlab Biotechnologies) according to the manufacturer's protocol. The miRNAs from the tissues for atlas analysis were mixed equally and used for library construction. The sequencing library was prepared using the NEB Multiplex Small RNA Library Prep Kit for Illumina kit (New England Biolabs) following the manufacturer's recommendations. In brief, the small RNAs were ligated with 3′ and 5′ SR adapters, reverse-transcribed and amplified. Amplified cDNA constructs between 140–160 bp in size were selected and sequenced using the Illumina HiSeq X Ten platform with single-end reads of 50 nucleotides. The raw reads were trimmed by removing adapter sequences and low-quality reads containing ploy-N, with adapter contaminants of less than 18 nt. Then, the high-quality clean reads of small RNA were mapped to the *xiaomi* reference sequence and miRNAs were identified using MiRDeep2[55].

Strand-specific RNA-seq libraries for lncRNA and circRNA identification were generated using Ribo-Zero Magnetic Kit (Illumina) and NEBNext Ultra Directional RNA Library Prep Kit (no. E7420, New England BioLabs) following the manufacturer's recommendations. In brief, total RNA was treated with the Ribo-Zero Magnetic Kit to remove ribosomal RNA. After fragmentation, the rRNA-depleted RNA was reverse-transcribed using random primers, followed by the second-strand synthesis. The resulting double-stranded DNA was ligated to adapters after purification, end repair and ligation of a poly A tail. Subsequently, the cDNA was digested with uracil-specific excision reagent (USER) enzyme to degrade the cDNA strands containing U instead of T. The first-strand cDNA that preserved the direction of the RNA was amplified and the products were purified. Finally, the strand-specific cDNA library was sequenced using an Illumina HiSeq X Ten platform with 150-bp pair-end reads. The resulting directional paired-end reads were filtered and trimmed using Trimmomatic[52] (v.0.38). Then, the clean reads were mapped to the *xiaomi* genome sequence using HISAT2. To construct transcripts, the mapped reads were assembled de novo using Cufflinks (http://cole-trapnell-lab.github.io/cufflinks/). The assembled transcripts were annotated using the *xiaomi* genome to identify protein-coding transcripts. After filtering of the protein-coding genes, lncRNAs were identified using the following parameters: (1) fragments per kilobase of transcript per million mapped reads ≥ 0.5; (2) the transcripts were longer than 200 bp. CircRNAs were identified essentially according to the method described by Memczak et al.[56].

Identification of SNPs, small indels and PAVs. We identified SNPs and small indels (length <100 bp) with MUMmer[57] (v3.23) (http://mummer.sourceforge.net/) between the *xiaomi* and Yugu1 genomes. In brief, the *xiaomi* pseudochromosome sequence was mapped to its corresponding Yugu1 pseudochromosomes with MUMmer, and then SNPs and indels were identified using Show-SNPs. PAVs were extracted by scanPAV[58] with default parameters. The resulting PAVs of up to 1,000 bp were filtered out as noise.

Syntenic analysis and identification of *xiaomi*-specific genes. All-versus-all BLASTP analysis of protein sequences was performed between *xiaomi* and Yugu1 using an *E*-value cut-off of 1×10^{-10} and syntenic blocks were then identified using MCscan (http://chibba.pgml.uga.edu/mcscan2/) based on the all-to-all BLASTP results with the following parameters: MATCH_SCORE > 50, MATCH_SIZE = 10, GAP_PENALTY = -1, OVERLAP_WINDOW = 5, MAX GAPS = 25. *xiaomi*-specific genes were determined by BLASTP analysis of protein sequences using an *E*-value cut-off of 1×10^{-5}.

***Agrobacterium*-mediated genetic transformation, GFP fluorescence observation and molecular analysis.** This method was developed following the protocol for mature seed-based transformation in rice[14], with improvements to make it suitable for foxtail millet. For callus induction, palea and lemma of *xiaomi* mature seeds were mechanically removed to reduce potential contamination. The dehusked seeds were surface-sterilized in 70% (v/v) ethanol for 2 min, and then in 10% bleach containing 0.1% tween 20 for 20 min, and finally rinsed 5 times with autoclaved water. The sterilized seeds were transferred onto sterile paper towels to remove excess water. The seeds were placed on CIM (4 g l^{-1} CHU (N6) basal salt with vitamins, 30 g l^{-1} sucrose, 2 mg l^{-1} 2,4-dichlorophenoxyacetic acid, 0.3 g l^{-1} casein acid hydrolysate, 2.8 g l^{-1} proline, 0.1 g l^{-1} myo-inositol, 0.1 mg l^{-1} 6-benzylaminopurine, 8 g l^{-1} agar, pH 5.7). The seeds were incubated at 28 °C in the dark. After 8–10 weeks induction, the callus could be seen. To obtain high-quality, regenerable calli, the initially formed callus was divided into 2–3 mm pieces and transferred onto fresh CIM. After three rounds of subculture, the calli became yellowish and were ready for transformation.

The vectors pCAMBIA1305-GFP and p8-GFP, both harbouring the *GFP* gene as a reporter, were used for protocol development and method optimization. The *HPT* gene in pCAMBIA1305-GFP and *NPTII* gene in p8-GFP were tested for their effectiveness in selection of transformants. Both vectors were introduced into the *Agrobacterium* strain EHA105 by electroporation. The EHA105 cells were cultured in YEB medium (5 g l^{-1} beef extract, 5 g l^{-1} peptone, 1 g l^{-1} yeast extract, 5 g l^{-1} sucrose, 10 mM magnesium sulfate, pH 7.0) overnight until optical density at 600 nm (OD_{600nm}) = 1.0. For infection and co-cultivation, the actively proliferating calli were infected with *Agrobacterium* cells (OD_{600nm} = 0.5) in the infection medium (0.44 g l^{-1} Murashige–Skoog salts, 1×B5 vitamins, 68 g l^{-1} sucrose, 36 g l^{-1} glucose, 1 g l^{-1} asparagine, 1 g l^{-1} casamino acids, 0.2 g l^{-1} cysteine, 2 mg l^{-1} 2,4-dichlorophenoxyacetic acid, 200 μM acetosyringone, pH 5.2) for 5 min. The calli were then blotted on sterile filter paper and transferred onto infection medium solidified with 8 g l^{-1} agarose for 3 d co-cultivation at 22 °C in the dark.

After co-cultivation, the calli were subcultured on CIM resting medium containing 250 mg l^{-1} carbenicillin at 28 °C in the dark for 3 d and then transferred to the CIM selection medium containing 100 mg l^{-1} paromomycin or 50 mg l^{-1} hygromycin B and 250 mg l^{-1} carbenicillin for another 2 wk. Yellowish calli were subcultured on the same medium every 2 wk until fast-growing resistant calli were formed. The resistant calli were then transferred to the shoot induction medium (SIM, 4.43 g l^{-1} Murashige–Skoog basal salt with vitamins, 30 g l^{-1} sucrose, 1 g l^{-1} proline, 1 g l^{-1} aspartic acid, 0.5 g l^{-1} casein acid hydrolysate, 0.25 mg l^{-1} copper sulfate, 2 mg l^{-1} 6-benzylaminopurine, 0.2 mg l^{-1} 1-naphthaleneacetic acid, 250 mg l^{-1} carbenicillin, 100 mg l^{-1} paromomycin or 50 mg l^{-1} hygromycin B, 8 g l^{-1} agar, pH 5.7) and cultured at 28 °C under 16 h light:8 h dark conditions for 4–5 wk. Regenerated shoots of 1–2 cm in length were transferred to the root-induction medium (RIM, half-strength Murashige–Skoog basal salt with vitamins, 30 g l^{-1} sucrose, 0.1 g l^{-1} myo-inositol, 2.6 g l^{-1} Gelzan (gellan gum), pH 5.6) for root formation. Healthy roots were developed in 2–3 wk. The rooted putative transgenic plants were moved directly to pots or field.

Expression of GFP was monitored with a Leica M305FCA fluorescence stereo microscope equipped with a DMC6200 camera during co-cultivation, selection and shoot regeneration. Transgenic *GFP* plants were confirmed by PCR genotyping using the *GFP*, *UBI*, *HPT* and *NPTII* specific binding primers listed in Supplementary Table 3.

T-DNA identification of the transgenic *xiaomi* lines. We identified the T-DNA insertion sites in 13 independently transgenic *xiaomi* plants, including 7 pCAMBIA1305-GFP and 6 p8-GFP transgenic lines, by genome resequencing. Approximately 50 T1 transgenic young seedlings of each line were used for DNA extraction and sequencing. Approximately 12 Gb of data (~28× coverage) was obtained for each line. T-DNA insertion site(s) were identified using TDNAScan[59]. Primers used for PCR confirmation of insertion sites are listed in Supplementary Table 3.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The genome assembly, annotation and expression data can be easily accessed at our Multi-omics Database for *S. italica* (MDSi) (http://sky.sxau.edu.cn/MDSi.htm). The genome assembly and annotation of *xiaomi* are also available at Genome Warehouse in the Beijing Institute of Genomics Data Center (https://bigd.big.ac.cn/) under accession number GWHAAZD00000000. The raw sequence data have been deposited in the Beijing Institute of Genomics Data Center with the following accession numbers: CRA001973 (Genome sequencing of *xiaomi* by PacBio), CRA001968 (Hi-C of *xiaomi*), CRA001972 (isoform sequencing of *xiaomi*), CRA001967 (Genome resequencing of *xiaomi* and Jingu21), CRA001953 (RNA-seq of 11 *xiaomi* tissues), CRA001954 (RNA-seq of the top second leaf of 30-day-old Jingu21), CRA001974 (non-coding RNAs), CRA002603 (genome resequencing of *xiaomi-2*) and CRA002604 (genome resequencing of 13 transgenic lines). The Yugu1 genome was downloaded from public database Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). The Zhanggu genome was downloaded from ftp://ftp.genomics.org.cn/pub/Foxtail_millet. Other data can be obtained from the public databases nr (https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/), KOG (ftp://ftp.ncbi.nih.gov/pub/COG/KOG/), KEGG (https://www.kegg.jp/), TrEMBL (https://www.ebi.ac.uk/uniprot), GO (http://geneontology.org/) and BUSCO embryophyta_odb9 dataset (http://busco.ezlab.org/datasets/

embryophyta_odb9.tar.gz). All data and materials are available from the corresponding author upon request. Source data are provided with this paper.

Received: 12 October 2019; Accepted: 20 July 2020;
Published online: 31 August 2020

References

1. Provart, N. J. et al. 50 years of *Arabidopsis* research: highlights and future directions. *N. Phytol.* **209**, 921–944 (2016).
2. Brutnell, T. P., Bennetzen, J. L. & Vogel, J. P. *Brachypodium distachyon* and *Setaria viridis*: model genetic systems for the grasses. *Annu. Rev. Plant Biol.* **66**, 465–485 (2015).
3. Doust, A. N., Kellogg, E. A., Devos, K. M. & Bennetzen, J. L. Foxtail millet: a sequence-driven grass model system. *Plant Physiol.* **149**, 137–141 (2009).
4. Jia, G. et al. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat. Genet.* **45**, 957–961 (2013).
5. Bennetzen, J. L. et al. Reference genome sequence of the model plant *Setaria*. *Nat. Biotechnol.* **30**, 555–561 (2012).
6. Brutnell, T. P. et al. *Setaria viridis*: a model for C_4 photosynthesis. *Plant Cell* **22**, 2537–2544 (2010).
7. Acharya, B. R. et al. Optimization of phenotyping assays for the model monocot *Setaria viridis*. *Front. Plant Sci.* **8**, 2172 (2017).
8. Zhang, G. et al. Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat. Biotechnol.* **30**, 549 (2012).
9. Tsai, K. J. et al. Assembling the *Setaria italica* L. Beauv. genome into nine chromosomes and insights into regions affecting growth and drought tolerance. *Sci. Rep.* **6**, 35076 (2016).
10. Diao, X., Schnable, J., Bennetzen, J. L. & Li, J. Initiation of *Setaria* as a model plant. *Front. Agr. Sci. Eng.* **1**, 16–20 (2014).
11. Lata, C., Gupta, S. & Prasad, M. Foxtail millet: a model crop for genetic and genomic studies in bioenergy grasses. *Crit. Rev. Biotechnol.* **33**, 328–343 (2013).
12. Li, P. & Brutnell, T. P. *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid grasses. *J. Exp. Bot.* **62**, 3031–3037 (2011).
13. Rockwell, N. C., Su, Y. S. & Lagarias, J. C. Phytochrome structure and signaling mechanisms. *Annu. Rev. Plant Biol.* **57**, 837–858 (2006).
14. Hiei, Y. & Komari, T. *Agrobacterium*-mediated transformation of rice using immature embryos or calli induced from mature seed. *Nat. Protoc.* **3**, 824–834 (2008).
15. Sage, R. F. The evolution of C_4 photosynthesis. *N. Phytol.* **161**, 341–370 (2004).
16. Ermakova, M., Danila, F. R., Furbank, R. T. & von Caemmerer, S. On the road to C_4 rice: advances and perspectives. *Plant J.* **101**, 940–950 (2020).
17. Yang, J. et al. Brassinosteroids modulate meristem fate and differentiation of unique inflorescence morphology in *Setaria viridis*. *Plant Cell* **30**, 48–66 (2018).
18. Huang, P. et al. *Sparse panicle1* is required for inflorescence development in *Setaria viridis* and maize. *Nat. Plants* **3**, 17054 (2017).
19. Saha, P. & Blumwald, E. Spike-dip transformation of *Setaria viridis*. *Plant J.* **86**, 89–101 (2016).
20. Huang, P. et al. Population genetics of *Setaria viridis*, a new model system. *Mol. Ecol.* **23**, 4912–4925 (2014).
21. Hu, S. et al. Xiaowei, a new rice germplasm for large-scale indoor research. *Mol. Plant* **11**, 1418–1420 (2018).
22. Meissner, R. et al. A new model system for tomato genetics. *Plant J.* **12**, 1465–1472 (1997).
23. Monte, E. et al. Isolation and characterization of *phyC* mutants in *Arabidopsis* reveals complex crosstalk between phytochrome signaling pathways. *Plant Cell* **15**, 1962–1980 (2003).
24. Takano, M. et al. Distinct and cooperative functions of phytochromes A, B, and C in the control of deetiolation and flowering in rice. *Plant Cell* **17**, 3311–3325 (2005).
25. Martins, P. K. et al. *Setaria viridis* floral-dip: a simple and rapid *Agrobacterium*-mediated transformation method. *Biotechnol. Rep.* **6**, 61–63 (2015).
26. Liu, Y., Yu, J., Zhao, Q., Zhu, D. & Ao, G. Genetic transformation of millet (*Setaria italica*) by *Agrobacterium*-mediated. *J. Agric. Biotechnol.* **13**, 32–37 (2005).
27. Liu, Y., Yu, J., Ao, G. & Zhao, Q. Factors influencing *Agrobacterium*-mediated transformation of foxtail millet (*Setaria italica*). *Chin. J. Biochem. Mol. Biol.* **23**, 531–536 (2007).
28. Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S. & Thompson, W. F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **1**, 2320–2325 (2006).
29. Chakraborty, M., Baldwin-Brown, J. G., Long, A. D. & Emerson, J. J. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res.* **44**, e147 (2016).
30. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
31. Chin, C. S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
32. Burton, J. N. et al. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
34. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
35. Han, Y. & Wessler, S. R. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199 (2010).
36. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21** (Suppl. 1), i351–i358 (2005).
37. Edgar, R. C. & Myers, E. W. PILER: identification and classification of genomic repeats. *Bioinformatics* **21**(Suppl 1), i152–i158 (2005).
38. Wicker, T. et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
39. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
40. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
41. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**(Suppl 2), ii215–ii225 (2003).
42. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
43. Blanco, E., Parra, G. & Guigó, R. Using geneid to identify genes. *Curr. Protoc. Bioinforma.* **18**, 4.3.1–4.3.28 (2007).
44. Korf, I. Gene finding in novel genomes. *BMC Bioinf.* **5**, 59 (2004).
45. Keilwagen, J. et al. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89 (2016).
46. Haas, B. J. et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
47. Dimmer, E. C. et al. The UniProt-GO annotation database in 2011. *Nucleic Acids Res.* **40**, D565–D570 (2012).
48. Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
49. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
50. Griffiths-Jones, S. et al. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124 (2005).
51. Simao, 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 (2015).
52. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
53. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
54. Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. RNA-seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**, 493–500 (2010).
55. Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* **40**, 37–52 (2012).
56. Memczak, S. et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
57. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
58. Giordano, F., Stammnitz, M. R., Murchison, E. P. & Ning, Z. scanPAV: a pipeline for extracting presence–absence variations in genome pairs. *Bioinformatics* **34**, 3022–3024 (2018).
59. Sun, L. et al. TDNAscan: a software to identify complete and truncated T-DNA insertions. *Front. Genet.* **10**, 685 (2019).
60. Li, W. et al. Gene mapping and functional analysis of the novel leaf color gene *SiYGL1* in foxtail millet [*Setaria italica* (L.) P. Beauv]. *Physiol. Plant.* **157**, 24–37 (2016).

Acknowledgements

We thank D. Grierson, Z. Tian, R. Fray and Y. Jiang for their critical reading of the manuscript, and R. Xia for help in developing the xEGP browser. This work was supported by the National Key R&D Program of China (2018YFD1000700, 2018YFD1000704 and 2018YFD1000702), National Natural Science Foundation of China (31600289, 31471502 and 31371693) and Key R&D Projects of Shanxi Province (201703D211008).

Author contributions

X.W., Y.H., Z.Y. and Y.S. designed and coordinated the study. Y.H., J.G., S.H. and B.Z. constructed the Jingu21 EMS-mutagenized library and identified the *xiaomi* mutant. Z.Y., X.W. and H.S. characterized the *xiaomi* phenotype, cloned the *PHYC* gene and analysed the sequence data. H.Z., Y.S. and C.W. established the *Agrobacterium*-mediated genetic transformation system and wrote the relevant part of the manuscript. X.W., Y.H., X.L., Z.Y., J.M., S.M. and M.B. performed downstream analysis of the sequence data. H.S., J.G., S.H. and B.Z. collected the experimental materials. X.W. and J.M. wrote the manuscript. All authors edited and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41477-020-0747-7.

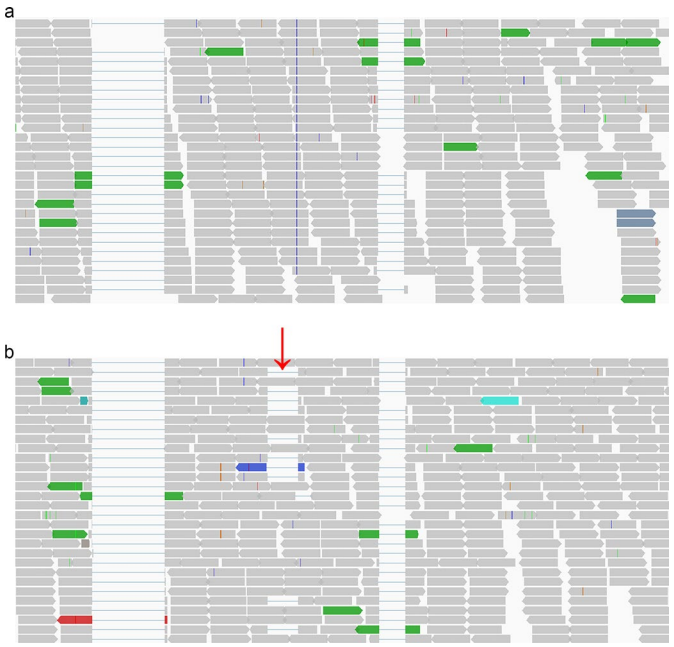
Supplementary information is available for this paper at https://doi.org/10.1038/s41477-020-0747-7.

Correspondence and requests for materials should be addressed to Y.S., Y.H. or X.W.

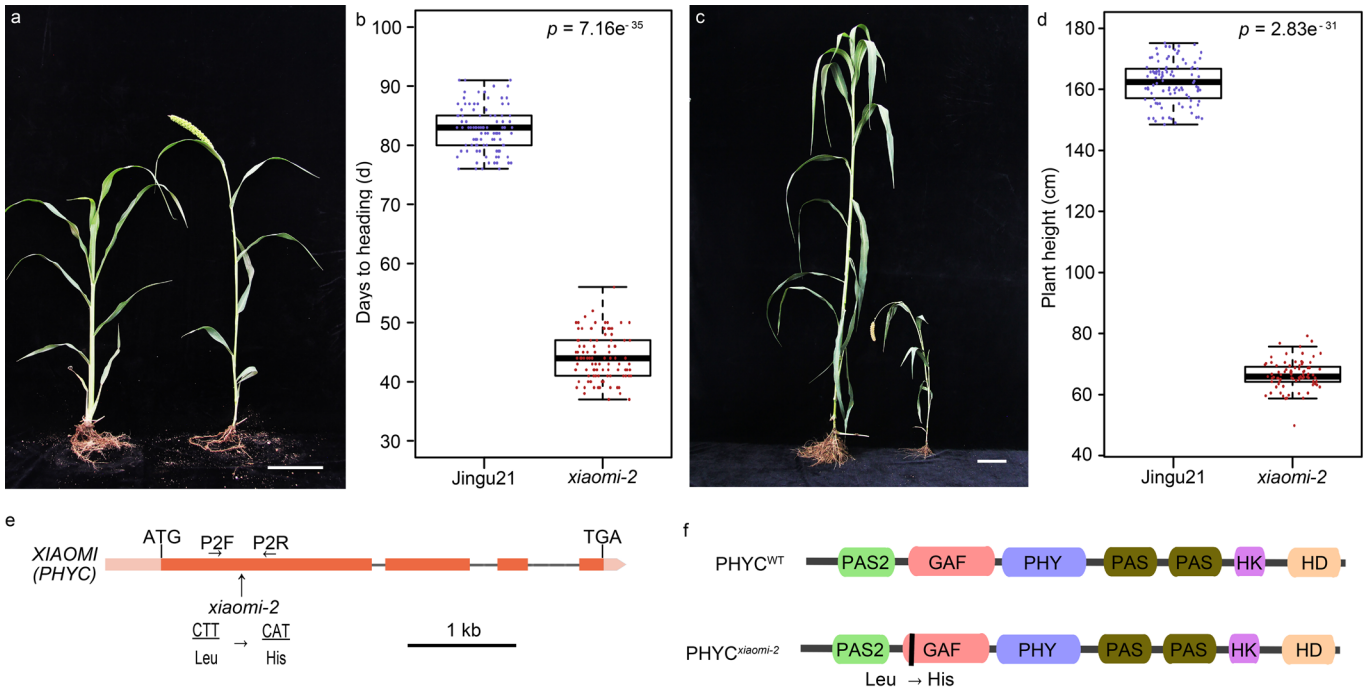
Peer review information *Nature Plants* thanks Andrew Doust, Manoj Prasad, Hong Yu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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.

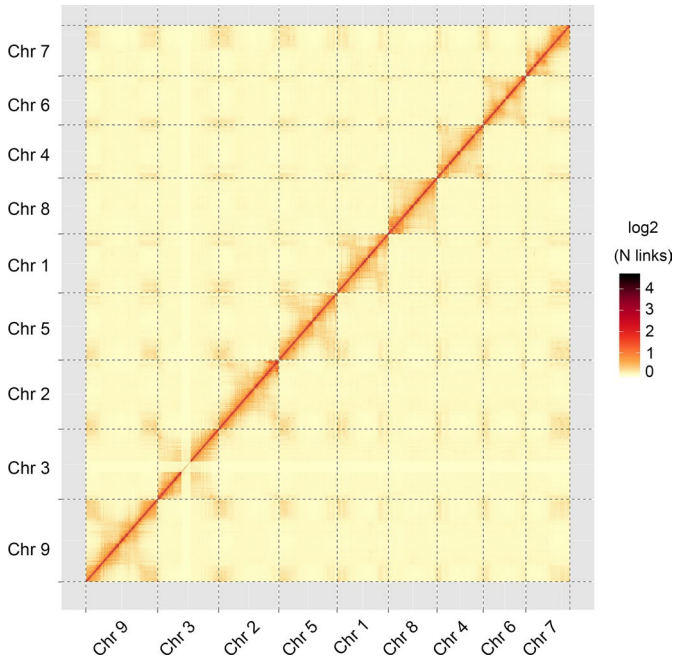


Extended Data Fig. 1 | Alternative splicing site of the *PHYC* gene in *xiaomi*. a, RNA-Seq reads of Jingu21. *xiaomi* genome sequences were used as reference genome. The blue vertical line shows the G-T mutation site. **b**, RNA-seq reads of *xiaomi*. The wrong splicing site was marked by a red arrow.

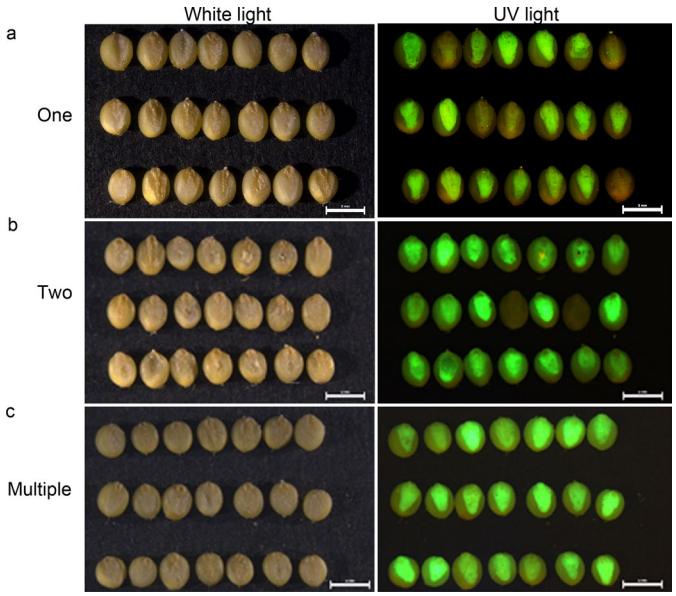


Extended Data Fig. 2 | Phenotypic and molecular characterization of the *xiaomi-2* mutant. a, Forty-day-old plants of Jingu21 (wild type, left) and *xiaomi-2* (right) plants grown under natural long-day conditions. **b**, Heading date of Jingu21 and *xiaomi-2* under natural field conditions. The heading date of ≥ 20 plants was measured for each replicate (n = 3 biologically independent replicates, ≥ 102 in total). The bottom and top of boxes represent the first and third quartile, respectively. The middle line is the median and the whiskers represent the maximum and minimum values. Statistical analysis was performed using two-tailed Wilcoxon rank-sum test. **c**, A mature small-sized *xiaomi-2* plant (right) compared to Jingu21 (left), at the 68th day in field. **d**, Plant height of Jingu21 and *xiaomi-2* under natural field conditions. The plant height of ≥ 23 plants was measured for each replicate (n = 3 biologically independent replicates, ≥ 83 in total). **e**, Molecular charicterization of *xiaomi-2*. Exons and introns are denoted by filled boxes and lines, respectively. P2F and P2R represent a pair of primers used to amplify the fragments harboring the mutation site from the segregating M_3 individuals (Primer sequences are listed in Supplementary Table 3). **f**, Structure of PHYC and its mutation version deduced according to mutations in *xiaomi-2*. Scale bars, 10 cm in **a** and **c**.

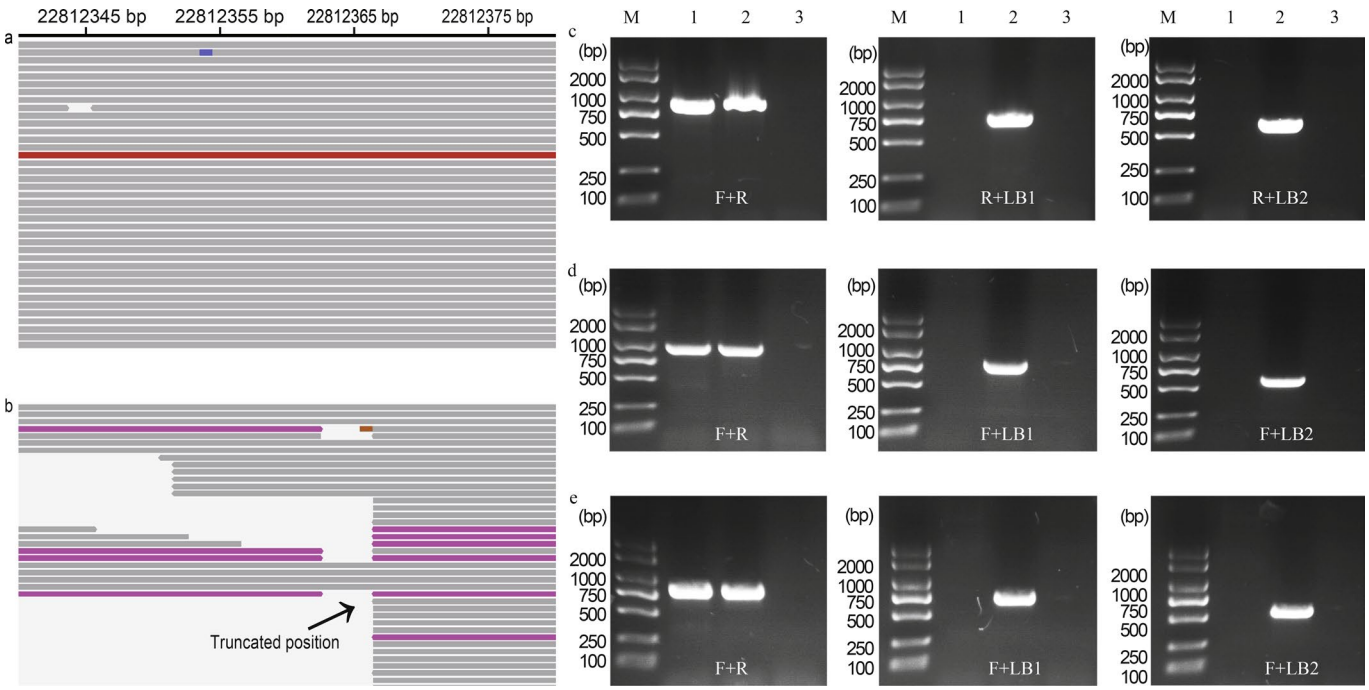
Extended Data Fig. 3 | Sequence alignment of the GAF domain of PHYC in foxtail millet and its homologs. Alignment was carried out using Clustal W method of the MegAlign software. Red box indicates the conserved residue Leu across all listed species that is substituted with His in *xiaomi-2*, demonstrating its functional importance for PHYC. Accession numbers for the aligned sequences: *Arabidopsis thaliana* NP_198433, *Brachypodium distachyon* XP_003559446, *Brassica napus* XP_013680236, *Ipomoea nil* XP_019162785, *Oryza sativa* AAF66603, *Panicum miliaceum*, RLN42126, *Solanum lycopersicum* NP_001307446, *Sorghum bicolor* XP_002466441, *Triticum aestivum* AAU06208, *Vitis vinifera* ACC6096 and *Zea mays* XP_008665426. PHYC protein in Jingu21 is presented as for *Setaria italica* (*Si9G09200*).



Extended Data Fig. 4 | Hi-C interaction matrices show the pairwise correlations between ordered scaffolds along the 9 pseudomolecules. The intensity of the dark color is proportional to the strength of the correlation.



Extended Data Fig. 5 | Transgene segregation in T_1 transgenic seeds as visualized for GFP expression. Dry mature seeds from transgenic lines representing single **a**, two **b**, or multiple **c**, T-DNA insertions were scanned with a dissection microscope equipped with UV light. All experiments were performed for eight independent biological repeats, and similar results were obtained. Scale bars, 2 mm.



Extended Data Fig. 6 | PCR confirmation of the site-specific T-DNA insertions identified by genome resequencing. a and **b**. An Integrative Genomics Viewer (IGV) display of genome sequencing reads from WT (**a**) or the transgenic line N2 (**b**) spanning the T-DNA insertion site 22812363 on chromosome 7. The break point caused by the insertion is marked by an arrow. **c**. PCR confirmation of the insertion site 33288299 on chromosome 6 in line H2. **d**. PCR confirmation of the insertion site 22812363 on chromosome 7 in line N2. **e**. PCR confirmation of the insertion site 39094661 on chromosome 5 in line N8. Note: The genomic DNA for sequencing and PCR was prepared from pooling approximate 50 T_1 transgenic seedlings, which explains the heterozygous nature of the T-DNA insertion seen in **b**-**e**. M, molecular marker; lane 1, no-transformed *xiaomi* plants; line 2, transgenic *xiaomi* plants; line 3, water control; F and R are primers for priming genomic regions flanking LB and RB ends of T-DNA, respectively; both the LB1 and LB2 primers are for T-DNA sequence close to the left border (LB). LB1 is 161 bp further apart from the border than LB2 for the vector pCAMBIA1305GFP, resulting in a band of bigger size in the R/LB1 pair in **c**. Similarly, LB1 and LB2 are distanced by 183 bp for the p8-GFP vector, thus resulting in different band size between F/LB1 and F/LB2 in **d** and **e**. All experiments were performed for three repeats, and similar results were obtained. Primers used are listed in Supplementary Table 3.

nature research

Corresponding author(s): Xingchun Wang

Last updated by author(s): Jul 8, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

1. PacBio SMRT sequencing data was acquired by PacBio Sequel and PacBio RSII.
2. Hi-C data was generated by Illumina HiSeq X Ten.
3. Iso-sequencing data was acquired by PacBio RSII.
4. RNA-Seq and non-coding RNA data were collected by Illumina HiSeq X Ten.
5. Re-sequencing data was acquired by Illumina HiSeq X Ten or sequencing platform.
6. qRT-PCR data was collected by Bio-Rad CFX96 real-time PCR system (Bio-Rad Laboratories Inc.).

Data analysis

We used lots of software for data analysis in this paper which was described in Methods section of manuscript.
1. Genome assembly and quality evaluation software
CANU (v1.5), https://canu.readthedocs.io/en/latest/
Falcon (v0.3.0), https://github.com/PacificBiosciences/FALCON/
Quickmerge (v0.2), https://github.com/mahulchak/quickmerge
pilon algorithm (v1.22), https://github.com/broadinstitute/pilon
HiC-Pro (v2.10.0), http://github.com/nservant/HiC-Pro
LACHESIS (https://github.com/shendurelab/LACHESIS)
BWA (v0.7.10-r789)
BUSCO (v2.0) analysis was undertaken with genome mode and embryophyta_odb9 dataset (http://busco.ezlab.org/datasets/embryophyta_odb9.tar.gz) as a reference.
2. Genome annotation software
Repeat sequences were annotated using the following softwares: LTR_Finder (v1.05), MITE-Hunter (v1.0.0, http://target.iplantcollaborative.org/mite_hunter.html), RepeatScout (v1.0.5), PILER-DF (v2.4), PASTEClassifier (v1.0) and RepeatMasker (v4.0.6).
Protein coding genes were annotated with Genscan (v3.1), Augustus (v2.4), GlimmerHMM (v1.2), GeneID (v1.4), SNAP (v2006-07-28) and EVM (v1.1.1).
tRNAs were identified by tRNA scan-SE (V1.3.1); rRNA and miRNA were identified by searching the Rfam (V12.1) with an E-value threshold of

1.0e-05.
Homologous genes were identified using GeMoMa (v1.3.1)
PacBio Iso-Seq and RNA-Seq data were directly mapped to the xiaomi genome and assembled by PASA (v2.0.2).
3. RNA-Seq analysis
Illumina RNA-seq reads were cleaned using Trimmomatic (v0.38) with parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:30 HEADCROP:10. The clean reads were mapped to the Xiaomi assembly using hisat2 (2.04) with default parameters. Genes expression analysis and quantile normalization was conducted using R with the transcripts per million (TPM).
Hierarchical clustering analysis was performed using pheatmap package (1.0.12) of R software (3.5.1).
4. Identification of SNPs, small InDels and PAVs
SNPs and small InDels were identified with MUMmer (v3.23, http://mummer.sourceforge.net/); PAVs were extracted by scanPAV (V2018-03-05, https://github.com/wtsi-hpag/scanPAV).
5. T-DNA identification in xiaomi transgenic lines
T-DNA insertion site (s) was identified using TDNAScan (V2018-09-13, https://github.com/noble-research-institute/TDNAscan).
6. Protein sequence alignment using MegAlign (v11.1.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

1. The genome assembly, annotation and expression data can be easily accessed at our Multi-omics Database for Setaria italica (http://sky.sxau.edu.cn/MDSi.htm). The genome assembly and annotation of xiaomi are also available at Genome Warehouse in the Beijing Institute of Genomics (BIG) Data Center (https://bigd.big.ac.cn/) under accession number GWHAAZD00000000.
2. The raw sequence data have been deposited in BIG Data Center with the following accession numbers: CRA001973 (Genome sequencing of xiaomi by PacBio), CRA001968 (Hi-C of xiaomi), CRA001972 (Iso-sequencing of xiaomi), CRA001967 (Genome re-sequencing of xiaomi and Jingu21), CRA001953 (RNA-seq of 11 xiaomi tissues), CRA001954 (RNA-Seq of the top second leaf of 30 day old Jingu21), CRA001974 (non-coding RNAs), CRA002603 (Genome re-sequencing of xiaomi-2) and CRA002604 (Genome re-sequencing of 13 transgenic lines). Yugu1 genome was downloaded from public database Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). Zhanggu genome was downloaded from ftp://ftp.genomics.org.cn/pub/Foxtail_millet.
3. Other data can be obtained from the public databases: NR (https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/), KOG (ftp://ftp.ncbi.nih.gov/pub/COG/KOG/), KEGG (https://www.kegg.jp/), TrEMBL(https://www.ebi.ac.uk/uniprot), GO (http://geneontology.org/) and BUSCO embryophyta_odb9 dataset (http://busco.ezlab.org/datasets/embryophyta_odb9.tar.gz),Repbase (https://www.girinst.org/repbase/).
4. All data supporting the findings of this study are available within the paper and its supplementary information, or from the corresponding author on request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes for plant growth experiments were chosen based on our past experience or the previous references performing similar analyses. The sample sizes were described in the corresponding figure legends or in the main text.
Data exclusions	No data were excluded from analysis in this study.
Replication	All experiments were conducted at least three times and produced similar results.
Randomization	Plants were grown in the field, greenhouses or growth chambers and randomly used for experiments.
Blinding	Experiments were not blinded. All data were collected according to the genotype of plants.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging