1 # 1. Extended Data

Figure #	Figure title One sentence only	Filename This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: *Smith_ED_Fig1.jpg*	Figure Legend If you are citing a reference for the first time in these legends, please include all new references in the Online Methods References section, and carry on the numbering from the main References section of the paper.
Extended Data Fig. 1	Alternative splicing site of the *PHYC* gene in *xiaomi*	Yang_ED_Fig1.jpg	**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	Yang_ED_Fig2.jpg	**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 Jingu21and *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). **c**, 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	Yang_ED_Fig3.jpg	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	Yang_ED_Fig4.jpg	The intensity of the dark color is proportional to the strength of the correlation.
Extended Data Fig. 5	Transgene segregation	Yang_ED_Fig5.jpg	Dry mature seeds from transgenic lines representing single (**a**), two (**b**)

	in T_1 transgenic seeds as visualized for GFP expression		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	Yang_ED_Fig6.jpg	**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 F/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.

2. Supplementary Information:

A. Flat Files

Item	Present?	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. The extension must be .pdf	A brief, numerical description of file contents. i.e.: *Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.*
Supplementary Information	Yes	*yang_Supplementary_information.pdf*	*Supplementary Figures 1-7*
Reporting Summary	Yes	Reporting_summary.pdf	

B. Additional Supplementary Files

Type	Number If there are multiple files of the same type this should be the numerical indicator. i.e. “1” for Video 1, “2” for Video 2, etc.	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_Supplementary_Video_1.mov*	Legend or Descriptive Caption Describe the contents of the file
Supplementary Table	1	*Yang_Supplementary_Tables.xlsx*	Supplementary Table 1-25 **Supplementary Table 1** Comparison of agronomic traits between Jingu21 and *xiaomi* **Supplementary Table 2** SNP and InDel between Jingu21 and *xiaomi* **Supplementary Table 3** Primers used in this study

			Supplementary Table 4 Differential expressed photoperiodic pathway genes in *xiaomi* **Supplementary Table 5** Statistic summary of PacBio sequencing data **Supplementary Table 6** The length distribution of *xiaomi* PacBio subreads **Supplementary Table 7** Details of *xiaomi* draft genome assembled with Canu, Falcon and Quickmerge **Supplementary Table 8** Statistic summary of the *xiaomi* Hi-C-based PE reads **Supplementary Table 9** Evaluation of completeness of the *xiaomi* genome assembly using BUSCOs **Supplementary Table 10** Repetitive elements in *xiaomi* assembly

			Supplementary Table 11 Summary of the predicted protein-coding genes **Supplementary Table 12** Statistical summary of the *xiaomi* genes distributed on chromosomes **Supplementary Table 13** Functional annotation of predicted genes **Supplementary Table 14** Comparison of the *xiaomi* genome with the previously published foxtail millet genomes **Supplementary Table 15** Present variations in x*iaomi* **Supplementary Table 16** Absent variations in *xiaomi* **Supplementary Table 17** *xiaomi* genes absent in Yugu1 gene models **Supplementary Table 18** *xiaomi* genes absent in Zhanggu gene

			models **Supplementary Table 19** Putative *xiaomi*-specific genes **Supplementary Table 20** TPM expression value across 11 different tissues in xiaomi **Supplementary Table 21** Constitutive expressed genes in *xiaomi* **Supplementary Table 22** Specific expressed genes in *xiaomi* **Supplementary Table 23** Preferred expressed genes in *xiaomi* **Supplementary Table 24** Transformation frequency of *xiaomi* with HPT and NPTII marker genes **Supplementary Table 25** Putative T-DNA insertion sites of 14 independent transgenic lines
Supplementary Data	1	*Yang_Supplementary_Data_1.zip*	Supplementary Data 1 SNPs and InDels between *xiaomi* and

			Yugu1

3. Source Data

Figure	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_SourceData_Fig1.xls,* or *Smith_Unmodified_Gels_Fig1.pdf*	Data description i.e.: Unprocessed Western Blots and/or gels, Statistical Source Data, etc.
Source Data Fig. 1	Yang_SourceData_Fig1.xlsx	Statistical Source Data
Source Data Fig. 2	Yang_SourceData_Fig2.pdf	Unprocessed gel
Source Data Fig. 5	Yang_SourceData_Fig5.pdf	Unprocessed gels
Source Data Extended Data Fig. 2	Yang_SourceData_ED_Fig2.xlsx	Statistical Source Data

Source Data Extended Data Fig. 6	Yang_SourceData_ED_Fig6.pdf	Unprocessed gels

Establishment of a mini-foxtail millet with an Arabidopsis-like life cycle as a C_4 model system

Running title: Development of a model system for C_4 plants

Zhirong Yang[1,2†], Haoshan Zhang[3†], Xukai Li[1,4†], 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*], Xingchun Wang[1,4*]

1. Institute of Agricultural Bioengineering, Shanxi Agricultural University, Taigu, Shanxi, China
2. College of Arts and Sciences, Shanxi Agricultural University, Taigu, Shanxi, China
3. Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China
4. College of Life Sciences, Shanxi Agricultural University, Taigu, Shanxi, China
5. College of Agriculture, Shanxi Agricultural University, Taigu, Shanxi, China
6. Division of Plant and Crop Sciences, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK
7. Department of Agronomy, Purdue University, West Lafayette, Indiana, USA

† These authors contribute equally to this work.

*Corresponding authors:

Xingchun Wang, e-mail: wxingchun@sxau.edu.cn, wxingchun@163.com

Yuanhuai Han, e-mail: hanyuanhuai@sxau.edu.cn, swgctd@163.com

Yi Sui, e-mail: suiyi@caas.cn

Abstract

Foxtail millet (*Setaria italica*) is an important crop and an emerging model plant for C_4 grasses. However, functional genomics research on foxtail millet is challenged by its long generation time, relatively large stature and recalcitrance to genetic transformation. Here, we report the development of a rapid cycling mini-foxtail millet mutant, named *xiaomi*, as a C_4 model system. *xiaomi* can be grown for 5–6 generations a year in growth chambers due to its short life cycle and small plant size similar to Arabidopsis. A point mutation in *Phytochrome C* (*PHYC*) gene was found to be causal, which encodes a light receptor essential for photoperiodic flowering. A reference-grade *xiaomi* genome comprising 429.45 Megabases (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 (http://sky.sxau.edu.cn/MDSi.htm), make *xiaomi* an ideal model system for functional studies of C_4 plants.

Key words: *Setaria italica*; model plant; reference genome; gene expression atlas; genetic transformation system; *PHYC* gene

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 played 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 that can efficiently fix 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 (*Setaria italica*) is a cereal crop 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 USA[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 of them were proposed as C_4 model plant systems[3, 6]. In terms of 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) mutant 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 an extremely mini-mutant (dubbed *xiaomi*) with a life cycle similar to 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 M_2 population comprised of ~20,000 mutant lines derived from Jingu21 (wild type, WT). Under field conditions (37°25'13" N, 112°35'26" E), *xiaomi* exhibited an extremely early flowering phenotype with a heading date of ~39 days after sowing (DAS) (Fig. 1a and Supplementary Table 1). By contrast, WT plants showed an average heading date of ~82 DAS (Supplementary Table 1). *xiaomi* completed its life cycle in 70 days, whereas WT plants matured ~130 DAS (Fig. 1b,c and Supplementary Table 1). *xiaomi* was much shorter than WT (Fig. 1b and Supplementary Table 1), but, interestingly, the seed setting rate of *xiaomi* was 12.83% higher than that of 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

(SD, 10 h light/ 14 h dark) conditions than under long day (LD, 16 h light/ 8 h dark) conditions, indicating early heading of *xiaomi* was dependent on the LD conditions (Fig. 1e). Through extending the day-length and optimizing other conditions (See Materials and Methods for details), we were able to reduce the life cycle of *xiaomi* to 62 days with plant height of ~29 cm (Fig. 1f and Supplementary Table 1). Thus, five to six generations of *xiaomi* per year can be completed in growth chambers. With such a small size, a set of 1,296 *xiaomi* plants can be grown on two planting racks within a three-dimensional space of 1.2 m × 0.55 m × 2.0 m, similar to that for growing 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 LD conditions. All 9 F_1 plants exhibited the G1-like late-heading phenotype, and the resulting F_2 populations comprised of 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-kb region on chromosome 9, which harbors 27 genes according to the annotation of the *xiaomi* reference genome (Fig. 2a). Comparison between the *xiaomi* genome sequence and the Jingu21 genome sequence, with the latter generated by genome re-sequencing, revealed the presence of only a single mutation in the mapped region – a transversion from ‘G’ to ‘T’ in the coding

115 region of the gene *Si9g09200* in *xiaomi*, which encodes a putative PHYC protein (Fig.
116 2a, Supplementary Table 2). Of the 77 early heading F_2 individuals examined, all
117 were the T/T homozygous mutants. By contrast, of the 256 late heading individuals
118 examined, 92 were the G/G WT homozygotes and the other 164 were G/T
119 heterozygotes, reflecting a perfect association between the genotypes and phenotypes.
120 This mutation resulted in the formation of a stop codon to form a truncated protein
121 accounting for ~71% transcripts (transcript 1) of the gene (Fig. 2a-d and Extended
122 Data Fig. 1). Interestingly, this mutation also led to alternative splicing responsible for
123 a frame shift deletion of 113 bp that also formed a truncated protein accounting for
124 ~29% transcripts (transcript 2) of the gene (Fig. 2a-d and Extended Data Fig. 1).
125 Based on the prediction, the truncated proteins lack about two-third (peptide 1) or
126 one-third (peptide 2) of the second PAS domain, and the entire HK and HD domains[13]
127 (Fig. 2d).

128 We sequenced another mutant, named *xiaomi-2*, which also derived from Jingu21.
129 The *xiaomi-2* mutant showed an early heading phenotype similar to *xiaomi* (Extended
130 Data Fig. 2a-d). Sequence comparison of the *PHYC* locus revealed a single point
131 mutation (T674A) in the first exon of *PHYC* in *xiaomi-2* resulting in a change from a
132 conserved leucine to histone (Extended Data Fig. 2e,f and Extended Data Fig. 3). This
133 SNP is perfectly associated with phenotypic segregation of 82 early heading (A/A
134 genotype) and 84 late heading (49 T/A genotype and 35 T/T genotype) M_3 plants
135 derived from an M_2 heterozygous mutant. Together, these observations confirm that
136 the early-heading phenotype was resulted from the mutation at the *PHYC* locus.

To understand how the mutation at the *PHYC* locus affects flowering time under the LD conditions, we performed RNA-Seq analysis using the second leaves from 30-DAS plants (~10 days before *xiaomi* began heading), with the WT leaves collected at 30-DAS as a control. We found that the expression of several genes orthologous to the Arabidopsis oscillator genes, *PSEUDO-RESPONSE REGULATORs* (*PRRs*), *PHYTOCLOCK 1* (*PCL1*) and *GIGANTEA* (*GI*), respectively, which are critical for photoperiodic flowering, was significantly affected in *xiaomi* (Supplementary Fig. 1 and Supplementary Table 4). As expected, a putative downstream photoperiod gene orthologous to *Ghd7* showed ~95-fold decrease in the level of expression, whereas the putative flowering genes orthologous to *EHD1*, *Hd3a*/*FT*, *APETALA1* (AP1)/*FRUITFULL* (*FUL*) and *MADS*, respectively, exhibited significant increases in the level of expression in *xiaomi* (Supplementary Fig. 1 and Supplementary Table 4). Overall, these observations suggested that the early heading phenotype of *xiaomi* was caused by disruption of the photoperiodic pathways.

Assembly and annotation of *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 Table 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, Supplementary Table 8). After removing scaffolds less than 1 kb in length, our final assembly, dubbed *xiaomi* genome v1.0, is 429.94 Mb,

159 which contains 366 scaffolds with an N50 length of 42.41 Mb and 48 gaps (Table 1).
160 *K*-mer analysis suggests that the draft assembly covers approximately 98.10% of the
161 entire genome (Supplementary Fig. 2). The error rate of the assembly is about 0.001%
162 (one error per 100 kb) as estimated by Illumina DNA short reads. Single-copy
163 ortholog analysis showed that 97.78% of the 1,440 benchmarking universal
164 single-copy orthologs (BUSCO) genes were completely covered by the *xiaomi*
165 genome, with only 0.90% incomplete and 1.32% not assembled or annotated
166 (Supplementary Table 9). Collectively, these results indicate that the *xiaomi* genome
167 v1.0 can be used as a gold standard reference by the research community.

168 By a combination of *de novo* prediction and homology-based comparison, a total
169 of 237.28 Mb (55.19%) of the *xiaomi* genome sequences were annotated as repetitive
170 elements (Table 1 and Supplementary Table 10). We annotated 34,436 protein-coding
171 genes using 671,853 full length non-chimeric (FLNC) Iso-reads produced by PacBio
172 RS II and ~1,054.5 M short RNA-Seq reads produced by the HiSeq X-ten platform,
173 and a combination of *ab initio* prediction and protein-homology-based searches (Table
174 1 and Supplementary Table 11), of which 32,743 (95.08%) were located in the nine
175 pseudochromosomes (Supplementary Table 12). These genes were searched against
176 GO, KEGG, KOG, TrEMBL, nr database and compared with the annotation of
177 Arabidopsis and rice to retrieve homologs with known functions and a total of 33,789
178 genes (98.12%) were annotated (Supplementary Table 13). In addition, we annotated
179 919 rRNA genes, 3,516 tRNA genes, 2,631 pseudogenes, 340 microRNA (miRNA)
180 precursors, 28,260 long non-coding RNA (lncRNA) precursors, and 1,318 circular

RNA (circRNA) precursors (Table 1).

All the genomic and transcriptomic data are now 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, CDS or peptide sequences.

Comparison of genome sequences from *xiaomi* and other foxtail millet varieties.

Compared with the three previously released genome sequences from varieties Yugu1[5], Zhanggu[8] and TT8[9], 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 Insertions/Deletions (InDels, <100 bp), 2,804 (total 15.32 Mb) presence variations (>1,000 bp) and 2,722 (total 17.38 Mb) absence variations (>1,000 bp) between *xiaomi* and Yugu1 genomes (Fig.3, Supplementary Data 1 and Supplementary Table 15,16).

Of the annotated 34,436 protein-coding genes in the *xiaomi* genome, 32,112 genes (93.25%) are shared by the Yugu1 genome (v2.2), 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, and considered to be *xiaomi*-specific (Supplementary Table 19). A remarkable phenotypic difference between *xiaomi* and Yugu1 is their susceptibility/resistance to downy mildew (Supplementary Fig. 3), but it is unclear whether such a 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/investigation of gene function, we measured transcript levels in eleven diverse tissues representing the major organs over various developmental stages of *xiaomi* (See materials and methods for details). A total of 1,054.51 M raw reads (~30 M reads per sample) were produced and analyzed. A total of 31,226 (90.68%) genes were expressed in at least one of the eleven *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 top-second 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 the eleven *xiaomi* tissues. Of these genes, 85 (0.25%), including one transcription initiation factor (*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 prior to comparative gene expression analysis. Moreover, we identified 1,218 organ/tissue-specific genes and 1,226 organ/tissue preferentially expressed genes (Supplementary Fig. 4b,c, Supplementary Fig. 5 and Supplementary Table 22,23).

To make these expression data more user-friendly, we developed a *xiaomi*

Electronic Fluorescent Pictograph (xEFP) browser (http://sky.sxau.edu.cn/MDSi.htm). In 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 realized 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 (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 medium during subculture (Fig. 5f). Roots of transgenic plants expressing *GFP* could be easily induced on the rooting medium and rooted plants survived well after transplanting to soil (Fig. 5g-i). We compared two commonly used selectable markers *NPTII* (*Neomycin Phosphotransferase II*) and *HPT* (*Hygromycin Phosphotransferase*), 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

248 average of 8.72% for *HPT* (Supplementary Table 24). We then confirmed the presence
249 of transgenes in primary putative transgenic plants (T_0) by PCR using primers
250 amplifying the *GFP* gene, *UBI* promoter, *HPT* or *NPTII* selectable marker gene,
251 respectively (Fig. 5j and Supplementary Fig. 6). We also observed GFP expression in
252 both dry and germinating seeds, indicating transmission of transgene to progeny (Fig.
253 5k,l and Extended Data Fig. 5). Insertions of transgenes into the *xiaomi* genome were
254 verified by genome sequencing of 13 independent transgenic lines produced with the
255 *HPT* or *NPTII* marker, and insertion sites of T-DNA were further confirmed by PCR
256 in three lines examined (Extended Data Fig. 6 and Supplementary Table 25). We grew
257 T_1 plants representing eight transgenic events in pots and observed no obvious
258 phenotypic differences from non-transformed *xiaomi* plants (Supplementary Fig. 7).
259 Collectively, we demonstrated the transgenic nature of the plants generated by
260 *Agrobacterium*-mediated transformation. Thus, we have established an efficient
261 protocol that allows production of transgenic plants ready to transplant to soil in 2-3
262 months counting from *Agrobacterium* infection or 4-5 months counting from callus
263 initiation from mature seeds (Fig. 6).
264
265 **Discussion**
266 Foxtail millet is an emerging C_4 model plant suitable for investigation of various
267 biological phenomena absent in other model plants such as Arabidopsis and rice.
268 Through this study, we identified a mini-plant *xiaomi* with short generation time,
269 created a reference genome and a gene expression atlas from various tissues, and
270 developed a highly efficient transformation protocol. The results demonstrate the

suitability of *xiaomi* as an ideal 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 to C_3 species, C_4 plants usually show higher rates of photosynthesis as well as higher nitrogen and water use 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 primary biomass production on the planet despite comprising only 3% of all land plant species[15]. Due to such high productivity, introducing the 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 underlining 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] proposed green foxtail as a C_4 model plant considering its relatively short stature, simple growth requirements, and rapid life cycle. Since then, great progresses have been made in genome assembly, transformation technology improvement, germplasm generation as well as mutant isolation and characterization in green foxtail[5, 17-20]. Compared to its wild progenitor green foxtail, foxtail millet is more suitable as a model plant. Firstly,

293 the seeds of foxtail millet are generally non-shattering and non-dormant, easier to
294 collect and germinate; secondly, foxtail millet has been widely cultivated for both
295 human food and fodder in the arid and semi-arid regions of the world, particularly in
296 China and India, and would be easier to deploy for grain production. These facts,
297 together with the short life cycle and small plant size, makes *xiaomi* an ideal model
298 plant to accelerate research in millet and many other C_4 plants.

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

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

314 At present, ~10% of the *xiaomi* genes are not captured in the tissues used for

315 construction of the gene expression atlas; nevertheless, the majority of these
316 'unexpressed' genes have homologs/orthologs in Arabidopsis and rice, suggesting that
317 they would be expressed in other tissues, at other developmental stages, or under
318 specific growth conditions, and additional RNA-seq should enable to the construction
319 of a more comprehensive gene expression atlas in the future.

320 Transformability is an essential prerequisite for a plant to be a model. Arabidopsis
321 can be efficiently transformed by floral dipping, which has enabled its rapid adoption
322 for basic research in plant biology worldwide. Recently, a similar approach (spike dip
323 transformation) has been explored in green foxtail with reported success[19, 25].
324 However, we were unable to recover any transgenic plant from *xiaomi* despite various
325 efforts made using this method. Transgenic plants were successfully produced using
326 calli induced from immature embryos/inflorescences in both foxtail millet and green
327 foxtail, although at low efficiency[26, 27]. The disadvantage of using fresh tissues is that
328 plants must be grown periodically to ensure a constant supply all year around, which
329 is obviously time-consuming and costly. Thus, we tried mature seeds as an explant
330 source for callus induction. After repeated trials, we realized that the primary calli
331 induced from mature embryos were not suitable for direct use in transformation most
332 likely due to their watery and soft nature. Then we focused our efforts on the
333 development of embryogenic calli by subculture and optimization of the infection and
334 selection steps by monitoring expression of the reporter gene *GFP*. We also compared
335 selectable markers and for the first time recognized *NPTII* as an efficient marker in
336 foxtail millet transformation. The transformation method established in this study has

a 3.5-fold higher efficiency than that previously reported[26], and upon further improvement should encourage broad adoption of *xiaomi* as a model for basic and applied research, especially in C_4 plants.

Materials and Methods

Plant materials and growth conditions.

xiaomi was identified from an ethyl methanesulfonate (EMS) mutagenized M_2 population of Jingu21, an elite 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 ($37^{o}25'13''$ N, $112^{o}35'26''$ E). For indoor research, plants were grown in the auto-controlled growth chamber/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 $\mu mol \cdot m^{-2} \cdot s^{-1}$ light intensity unless otherwise specified. To shorten the life cycle and reduce plant stature, we optimized growth conditions for *xiaomi*. Briefly, *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 above ground 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 plant of *xiaomi* or WT were harvested for genome re-sequencing.

For the expression atlas sequencing, eleven diverse tissues representing the major organ systems were collected, with three biological replications. These tissues were 3-day 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 days 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 days *xiaomi* seedlings (10 days before heading). Samples of stem, leaf 3 (flag leaf), leaf 4 (the top forth leaf) and panicle 3 were all harvested 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 re-sequencing 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 (CTAB) 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% CTAB 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 NaAc to precipitate the DNA. The precipitate was dissolved in 100 μL ddH_2O containing 10 ng/μL 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, Carlsbad, USA) and separated in 0.8% agarose gels.

Total RNA was isolated with RNAprep Pure Plant Kit (Tiangen Biotech Co., Ltd., Beijing, China) or Plant RNA kit (OMEGA, USA) according to the manufacturer's instructions. The integrity and quantity of extracted RNA were analyzed on the Agilent 2100 bioanalyzer and 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 Co., LTD (Beijing, China). Briefly, genomic DNA from a single *xiaomi* plant was randomly sheared into an average size of 20 kb, using a g-Tube (Covaris Inc., Woburn, MA, USA). 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, Beverly, USA) and sequences were generated on a PacBio Sequel (9 Cells) and PacBio RSII (1 cell) platform by Biomarker Technologies Co., LTD (Beijing, China).

Illumina HiSeq DNA libraries were made following standard protocols provided by Illumina. About 5 micrograms of 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 to the adapter sequences, and used to construct 270 bp paired-end library. The library was sequenced on an Illumina HiSeq X Ten sequencing platform by Biomarker Technologies Co., LTD (Beijing, China).

PacBio assembly, correction and validation.

The single-molecule sequencing data were assembled following a hierarchical approach, correction, assembly and polishing[29]. Briefly, a subset of longer reads was selected as seed data and corrected through Canu (v1.5)[30] and Falcon (v0.3.0)[31]. The error-corrected reads were assembled using Falcon and Canu. Since the Canu and the Falcon assemblies both contained some regions that were missing from the other one, the two initial assemblies were merged using Quickmerge (v0.2, 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 *Hin*dIII 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 by proteinase K and purified by a phenol extraction, followed by a phenol/chloroform/isoamylalcohol extraction. Then, the purified DNA was sheared to a size of 300–700 bp with a Covaris S220 instrument (Covaris, Woburn, MA, USA). 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 then the quality was assessed using HiC-Pro (v2.8.0, 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 (v1.05)[34], MITE-Hunter (v1.0.0)[35], RepeatScout (v1.0.5)[36] and PILER-DF (v2.4)[37] with the default parameters. Secondly, the primary repeat library was classified with PASTEClassifier (v1.0)[38] and then combined with Repbase[39] to build the final repeat library of *xiaomi*. Finally, repeats throughout the *xiaomi* genome were identified by RepeatMasker (v4.0.6) 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 Iso-Seq were used to conduct a

comprehensive search for consensus gene sets. For *ab initio* based gene prediction: five gene□finding program, Genscan (v3.1) [40], Augustus (v2.4)[41], GlimmerHMM (v1.2)[42], GeneID (v1.4)[43] and SNAP (v2006-07-28)[44] were used to detect genes in the repeat masked *xiaomi* genome with the default parameters. For homology-based prediction, proteins previously annotated in *Arabidopsis thaliana*, *Setaria italica*, *Oryza sativa*, *Sorghum bicolor* and *Zea mays* were downloaded and mapped to the *xiaomi* genome using BLAST and homologous genes were identified using GeMoMa (v1.3.1)[45]. Newly generated *xiaomi* PacBio Iso-Seq and RNA-Seq data were directly mapped to the *xiaomi* genome and assembled by PASA (v2.0.2)[46]. Finally, the results obtained from the above approaches were integrated into a consensus gene set of *xiaomi* using EVM (v1.1.1) with default parameters. These protein-coding genes were named using the following gene model nomenclature: Si (for *Setaria 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 (v2.2.31, E-value ≤1.0e-05). The Gene Ontology (GO)[47] terms for each gene were obtained using Blast2GO program[48] based on the above nr annotation.

492 The tRNA genes in the assembly were identified by tRNA scan-SE (v1.3.1)[49] with
493 eukaryote parameters. rRNA and miRNA were identified by searching the Rfam
494 (v12.1)[50] with an E-value threshold of 1.0e-05.
495 Pseudogene GeneWise (v2.4.1) was used to predict the candidate gene structure
496 based on the homogenous alignments. We filtered GeneWise's results to retain only
497 those with at least 95% coverage of the protein. The gene structures with frame shift
498 mutations were considered to be candidate pseudogenes.
499 **Genome quality evaluation.**
500 The quality of the *xiaomi* assembly was assessed by examining the alignment ratio of
501 HiSeq short reads and the presence of well conserved core eukaryotic genes. The
502 short reads generated by Illumina HiSeq platform were aligned to the *xiaomi*
503 assembly using BWA (v0.7.10-r789). To further evaluate the completeness of the
504 *xiaomi* gene models, BUSCO (v2.0)[51] analysis was undertaken with genome mode
505 and embryophyta_odb9 dataset
506 (http://busco.ezlab.org/datasets/embryophyta_odb9.tar.gz) as a reference. The
507 embryophyte_db9 dataset contains 1,440 protein sequences and orthologous group
508 annotations for major clades. The proportion of complete and partial core eukaryotic
509 genes was assessed as a measure of the completeness of the *xiaomi* assembly.
510 **RNA-sequencing library preparation, Isoform- and HiSeq-sequencing.**
511 For Iso-sequencing, eight tissues, including seed, seedling, root, stem, young leaf (leaf
512 1), mature leaf (leaf 3), pollinated panicles (panicle 1) and panicles at the filling stage
513 (panicle 3), were harvested for RNA isolation. Equal amounts of total RNA from each

514 tissue were pooled together to identify as many isoforms as possible. SMRTbell
515 libraries were prepared according to the Iso-Seq protocol (Isoform Sequencing
516 (Iso-Seq™) using the Clontech SMARTer PCR cDNA Synthesis Kit and the
517 BluePippin™ Size Selection System). The first cDNA strand was synthesized using
518 SMARTer™ PCR cDNA Synthesis Kit (Takara Biotechnology, Dalian, China). After
519 cycle optimization, large-scale PCR was performed to generate double-strand cDNA
520 for size selection on the BluePippin System (Sage Science, Inc., Beverly, MA, USA).
521 Then, another large-scale PCR was performed using the eluted DNA to generate more
522 double-stranded cDNA. Re-amplified cDNA was purified, repaired and ligated with
523 hairpin adapters. To minimize the bias that favors sequencing of shorter transcripts,
524 multiple size-fractionated libraries (0–1, 0.5–1, 1–2, 1–3, 2–3 and 2–8 kb) were
525 constructed according to the manufacture's instruction. Finally, a total of 15 SMRT
526 cells were sequenced on a PacBio RS II platform.

527 For transcriptome atlas sequencing with the Illumina HiSeq X-ten platform,
528 RNA-Seq libraries were constructed using NEBNext Ultra RNA Library Prep Kit for
529 Illumina (E7770, New England BioLabs, USA) according to the manufacture's
530 instruction. Briefly, mRNA was purified from total RNA using NEBNext Poly (A)
531 mRNA Magnetic Isolation Module (E7490, New England BioLabs, USA) and
532 fragmented into approximately 200 nt RNA short fragments. The fragmented mRNAs
533 were then used as templates to synthesize the first-strand cDNA and the second-strand
534 cDNA. After end repair and adaptor ligation, the products were selected by Agencourt
535 AMPure XP beads (Beckman Coulter, Inc., CA, USA) 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 Co., LTD (Beijing, China).

RNA-seq read processing, clustering analyses, Z-score and coefficient of variation expression analysis.

Illumina RNA-seq reads of *xiaomi* were cleaned using Trimmomatic (v0.38)[52] 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 (v2.04)[53] with default parameters. Genes expression analysis and quantile normalization were conducted using R with the transcripts per million (TPM)[54]. The genes with TPM >0 in a given organ were considered to be expressed in this organ. Tissue-preferential and -specific genes were identified according to their TPM. Fold changes >10 between the tissues showing the highest and the second highest expression levels were considered to be tissue specific genes, while 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. CV values 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 no more than a two-fold difference between the highest and lowest levels of a gene transcript in any organs was considered to be constitutively expressed.

Hierarchical clustering analysis (HCA) was performed using pheatmap package (v1.0.12) of R software. 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 Co., Ltd, Beijing, China) 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 NEB Multiplex Small RNA Library Prep Kit for Illumina kit (New England Biolabs, USA) following the manufacturer's recommendations. Briefly, the small RNAs were ligated with 3′ and 5′ SR adapters, reverse-transcribed and amplified. Amplified cDNA constructs between 140–160 bp were selected and sequenced using 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, CA, USA) and NEBNext Ultra Directional RNA Library Prep Kit (E7420, New England BioLabs, USA) following the manufacturer's recommendations. Briefly, 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 adaptors 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 (v0.38)[52]. 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, lncRNA was identified using the following parameters: 1) FPKM≥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 (v3.23)[57] (http://mummer.sourceforge.net/) between the *xiaomi* and Yugu1 genomes. Briefly, the *xiaomi* pseudochromosome sequence was mapped to its corresponding Yugu1 pseudochromosomes with MUMmer, and then SNPs and InDels were identified using

602 Show-SNPs. PAVs were extracted by scanPAV[58] with default parameters. The
603 resulting PAVs ≤ 1000 bp were filtered out as noise.

604 **Syntenic analysis and identification of *xiaomi*-specific genes.**

605 All-versus-all BLASTP analysis of protein sequences was performed between *xiaomi*
606 and Yugu1 using an E-value cutoff of 1e-10 and syntenic blocks were then identified
607 using MCscan (http://chibba.pgml.uga.edu/mcscan2/) based on the all-to-all BLASTP
608 results with the following parameters: MATCH_SCORE >50, MATCH_SIZE=10,
609 GAP_PENALTY=-1, OVERLAP_WINDOW=5, MAX GAPS=25. *xiaomi*-specific
610 genes were determined by BLASTP analysis of protein sequences using an E-value
611 cutoff of 1e-5.

612 ***Agrobacterium*-mediated genetic transformation, GFP fluorescence observation**
613 **and molecular analysis.**

614 This method was developed following mature seed-based transformation protocol
615 in rice[14], with improvement to make it suitable for foxtail millet. For callus induction,
616 palea and lemma of *xiaomi* mature seeds were mechanically removed to reduce
617 potential contamination. The dehusked seeds were surface-sterilized in 70% (v/v)
618 ethanol for 2 min, and then in 10% bleach containing 0.1% tween 20 for 20 minutes,
619 and finally rinsed five times with autoclaved water. The sterilized seeds were
620 transferred onto sterile paper towels to remove excess water. The seeds were placed
621 on the callus induction medium (CIM, 4 g/L CHU (N6) basal salt with vitamins, 30
622 g/L sucrose, 2 mg/L 2,4-D, 0.3 g/L casein acid hydrolysate, 2.8 g/L proline, 0.1 g/L
623 myo-inositol, 0.1 mg/L 6-benzylaminopurine, 8 g/L agar, pH 5.7). The seeds were

624 incubated at 28 °C in the dark. After 8-10 weeks induction, the callus could be seen.
625 To obtain high-quality, regenerable calli, the initially formed callus was divided into
626 2–3 mm pieces and transferred onto fresh CIM. After 3 rounds of subculture, the calli
627 became yellowish and were ready for transformation.

628 The vectors pCAMBIA1305-GFP and p8-GFP, both harboring the *GFP* gene as a
629 reporter, were used for protocol development and method optimization. The *HPT*
630 gene in pCAMBIA1305-GFP and *NPT II* gene in p8-GFP were tested for their
631 effectiveness in selection of transformants. Both vectors were introduced into the
632 *Agrobacterium* strain EHA105 by electroporation. The EHA105 cells were cultured in
633 the YEB medium (5 g/L beef extract, 5 g/L peptone, 1 g/L yeast extract, 5 g/L sucrose,
634 10 mM magnesium sulphate, pH 7.0) overnight until an OD_{600nm} = 1.0. For infection
635 and co-cultivation, the actively proliferating calli were infected with *Agrobacterium*
636 cells (OD_{600nm} = 0.5) in the infection medium (IM, 0.44 g/L MS salts, 1×B5 vitamins,
637 68 g/L sucrose, 36 g/L glucose, 1 g/L asparagine, 1 g/L casamino acids, 0.2 g/L
638 cysteine, 2 mg/L 2, 4-D, 200 μM acetosyringone, pH 5.2) for 5 min. The calli were
639 then blotted on sterile filter paper and transferred onto IM solidified with 8 g/L
640 agarose for 3-day co-cultivation at 22 °C in the dark.

641 After co-cultivation, the calli were sub-cultured on the CIM resting medium
642 containing 250 mg/L carbenicillin at 28 °C in the dark for 3 days and then transferred
643 to the CIM selection medium containing 100 mg/L paromomycin or 50 mg/L
644 hygromycin B and 250 mg/L carbenicillin for another two weeks. Yellowish calli
645 were sub-cultured on the same medium every two weeks until fast-growing resistant

calli were formed. The resistant calli were then transferred to the shoot induction medium (SIM, 4.43 g/L MS basal salt with vitamins, 30 g/L sucrose, 1 g/L proline, 1 g/L aspartic acid, 0.5 g/L casein acid hydrolysate, 0.25 mg/L copper sulfate, 2 mg/L 6-benzylaminopurine, 0.2 mg/L NAA, 250 mg/L carbenicillin, 100 mg/L paromomycin or 50 mg/L hygromycin B, 8 g/L agar, pH 5.7) and cultured at 28 °C under 16 h light/8 h dark condition for 4–5 weeks. Regenerated shoots of 1–2 cm in length were transferred to the root induction medium (RIM, half-strength MS basal salt with vitamins, 30 g/L sucrose, 0.1 g/L myo-inositol, 2.6 g/L Gelzan, pH 5.6) for root formation. Healthy roots were developed in 2–3 weeks. 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* and *HPT*/*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 seven pCAMBIA1305-GFP and six p8-GFP transgenic lines, by genome resequencing. Approximate 50 T_1 transgenic young seedlings of each line were used for DNA extraction and sequencing. Approximate 12 Gb 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.

Data availability

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. 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. 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.

References

1. Provart, N. J. et al. 50 years of Arabidopsis research: highlights and future

691 directions. *New Phytol.* **209**, 921-944 (2016).
692 2. Brutnell, T. P., Bennetzen, J. L. & Vogel, J. P. *Brachypodium distachyon* and
693 *Setaria viridis*: model genetic systems for the grasses. *Annu. Rev. Plant Biol.* **66**,
694 465-485 (2015).
695 3. Doust, A. N., Kellogg, E. A., Devos, K. M. & Bennetzen, J. L. Foxtail millet: a
696 sequence-driven grass model system. *Plant Physiol.* **149**, 137-141 (2009).
697 4. Jia, G. et al. A haplotype map of genomic variations and genome-wide association
698 studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat. Genet.* **45**,
699 957-961 (2013).
700 5. Bennetzen, J. L. et al. Reference genome sequence of the model plant Setaria. *Nat.*
701 *Biotechnol.* **30**, 555 (2012).
702 6. Brutnell, T. P. et al. *Setaria viridis*: a model for C_4 photosynthesis. *Plant Cell* **22**,
703 2537-2544 (2010).
704 7. Acharya, B. R. et al. Optimization of phenotyping assays for the model monocot
705 *Setaria viridis*. *Front. Plant Sci.* **8**, 2172 (2017).
706 8. Zhang, G. et al. Genome sequence of foxtail millet (*Setaria italica*) provides
707 insights into grass evolution and biofuel potential. *Nat. Biotechnol.* **30**, 549
708 (2012).
709 9. Tsai, K. J. et al. Assembling the *Setaria italica* L. Beauv. genome into nine
710 chromosomes and insights into regions affecting growth and drought tolerance.
711 *Sci. Rep.* **6**, 35076 (2016).
712 10. Diao, X., Schnable, J., Bennetzen, J. L. & Li, J. Initiation of Setaria as a model

713 plant. *Front. Agr. Sci. Eng.* **1**, 16-20 (2014).
714 11. Lata, C., Gupta, S. & Prasad, M. Foxtail millet: a model crop for genetic and
715 genomic studies in bioenergy grasses. *Crit. Rev. Biotechnol.* **33**, 328-343 (2013).
716 12. Li, P. & Brutnell, T. P. *Setaria viridis* and *Setaria italica*, model genetic systems
717 for the Panicoid grasses. *J. Exp. Bot.* **62**, 3031-3037 (2011).
718 13. Rockwell, N. C., Su, Y. S. & Lagarias, J. C. Phytochrome structure and signaling
719 mechanisms. *Annu. Rev. Plant Biol.* **57**, 837-858 (2006).
720 14. Hiei, Y. & Komari, T. *Agrobacterium*-mediated transformation of rice using
721 immature embryos or calli induced from mature seed. *Nat. Protoc.* **3**, 824-834
722 (2008).
723 15. Sage, R. F. The evolution of C_4 photosynthesis. *New Phytol.* **161**, 341-370 (2004).
724 16. Ermakova, M., Danila, F. R., Furbank, R. T. & von Caemmerer, S. On the road to
725 C_4 rice: advances and perspectives. *Plant J.* **101**, 940-950 (2020).
726 17. Yang, J. et al. Brassinosteroids modulate meristem fate and differentiation of
727 unique inflorescence morphology in *Setaria viridis*. *Plant Cell* **30**, 48-66 (2018).
728 18. Huang, P. et al. *Sparse panicle1* is required for inflorescence development in
729 *Setaria viridis* and maize. *Nature Plants* **3**, 17054 (2017).
730 19. Saha, P. & Blumwald, E. Spike-dip transformation of *Setaria viridis*. *Plant J.* **86**,
731 89-101 (2016).
732 20. Huang, P. et al. Population genetics of *Setaria viridis*, a new model system. *Mol.*
733 *Ecol.* **23**, 4912-4925 (2014).
734 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 (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 (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**, W265-268 (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**, i351-358 (2005).

37. Edgar, R. C. & Myers, E. W. PILER: identification and classification of genomic repeats. *Bioinformatics* **21 Suppl 1**, i152-158 (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

779 elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
780 40. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic
781 DNA. *J. Mol. Biol.* **268**, 78-94 (1997).
782 41. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new
783 intron submodel. *Bioinformatics* **19 Suppl 2**, ii215-225 (2003).
784 42. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two
785 open source *ab initio* eukaryotic gene-finders. *Bioinformatics* **20**, 2878-2879
786 (2004).
787 43. Blanco, E., Parra, G. & Guigó, R. Using geneid to identify genes. *Curr. Protoc.*
788 *Bioinformatics* **18**, 4.3.1-4.3.28 (2007).
789 44. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
790 45. Keilwagen, J. et al. Using intron position conservation for homology-based gene
791 prediction. *Nucleic Acids Res.* **44**, e89 (2016).
792 46. Haas, B. J. et al. Improving the Arabidopsis genome annotation using maximal
793 transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654-5666 (2003).
794 47. Dimmer, E. C. et al. The UniProt-GO annotation database in 2011. *Nucleic Acids*
795 *Res.* **40**, D565-570 (2012).
796 48. Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and
797 analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676 (2005).
798 49. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of
799 transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955-964 (1997).
800 50. Griffiths-Jones, S. et al. Rfam: annotating non-coding RNAs in complete genomes.

Nucleic Acids Res. **33**, D121-124 (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).

Acknowledgement

We thank Professor Don Grierson (University of Nottingham), Dr. Zhixi Tian (Institute of Genetics and Developmental Biology, Chinese Academy of Sciences), Dr. Rupert Fray (University of Nottingham) and Dr. Yiwei Jiang (Purdue University) for their critical reading of the manuscript. We are grateful to Professor Rui Xia at South China Agricultural University for help in developing 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 analyzed the sequence data. H. Z., Y. S. and C. W. established the *Agrobacterium*-mediate 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.

Figure Legends and Tables

Fig. 1 | Phenotypic characterization of *xiaomi*. **a**, Forty DAS field-grown Jingu21 (left) and *xiaomi* (right) plants. **b**, An adult small-sized *xiaomi* plant (right) compared to Jingu21 (left), at the 68th day in field. **c**, An enlarged view of the panicle from the *xiaomi* plant shown in **b**. **d**, Seeds harvested from the field-grown Jingu21 (top, 310 seeds) and *xiaomi* (bottom, 310 seeds). **e**, Heading dates of *xiaomi* and Jingu21 under the LD or the SD conditions. The heading dates of ≥ 25 plants were measured for each replicate (n = 3 biologically independent replicates, ≥ 89 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. **f,** Image of seven 39-DAS individual *xiaomi* plants in a pot (dimeter 10 cm, height 10 cm) grown under the optimized conditions. Scale bars, 10 cm (**a** and **f**), 20 cm (**b**), 5 cm (**c**) and 2 cm (**d**), respectively.

Fig. 2 | Molecular characterization of *xiaomi*. **a**, Genetic mapping of *xiaomi*. Top: a schematic diagram showing positions of the mutation at the *PHYC* locus. 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 cDNA 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 harboring 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. **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 ladders. 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 top second leaf at filling stage was presented here and similar results were observed in two other detected tissues including stem and seedling. **d**, Structure of PHYC and its truncated versions deduced according to mutations in *xiaomi*. PAS, Per (period circadian protein) Arn (Ah receptor nuclear translocator protein) and Sim (single-minded protein) domain; GAF, cGMP-specific phosphodiesterase, adenylyl cyclase and FhlA domain; PHY, phytochrome domain; HK, His kinase A (phospho-acceptor) domain; HD, histidine kinase, DNA gyrase B and HSP90-like ATPase domain.

Fig. 3 | Circular plot of the *xiaomi* genome sequence compared with the Yugu1 genome. GC content (**a**), gene density (**b**), transposon element content (**c**), SNP density (**d**), InDel density (**e**) and PAV distributions (**f**) in sliding windows of 100 kb. **g**, Links display homologous genes in *xiaomi* and Yugu1.

Fig. 4 | Expression pattern of the *Si9g04830* gene. The *Si9g04830* gene encodes a magnesium chelatase D subunit which catalyzes the insertion of magnesium into protoporphyrin IX in chlorophyll biosynthesis. The strongest expression of *Si9g04830* is seen in the leaves, consistent with its known biological role and with published data[60]. The scale bar (in TPM) is displayed on the left. Note: The eleven tissues described here were presented in color. The expression data of the gray tissues/organs is being analyzed or is about to be analyzed, and will be presented in the MDSi database in the near future.

Fig. 5 | *Agrobacterium*-mediated transformation of *xiaomi*. a, Embryogenic calli suitable for transformation, 2 months after seed inoculation. **b**, Calli co-cultivated with *Agrobacterium* for 3 days under bright light. **c**, UV visualization of infected calli in **b**, showing transient expression of the *GFP* reporter. **d**, A transformed callus sector with light yellow color proliferating on selection medium at the end of the second-round selection. **e**, The same callus as in **d** but 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, respectively. **i**, An adult primary transgenic plant. **j**, PCR confirmation of the transgenic plantlets generated with the *HPT* selectable marker using *GFP* gene (top) or *UBI* promoter (bottom) specific primers. M, molecular marker; lane 1, plasmid DNA; lane 2, non-transformed *xiaomi* plant; lanes 3–16, independent T_0 transformants; lane 17, water control. **k** and **l**, Segregation of *GFP* transgene in

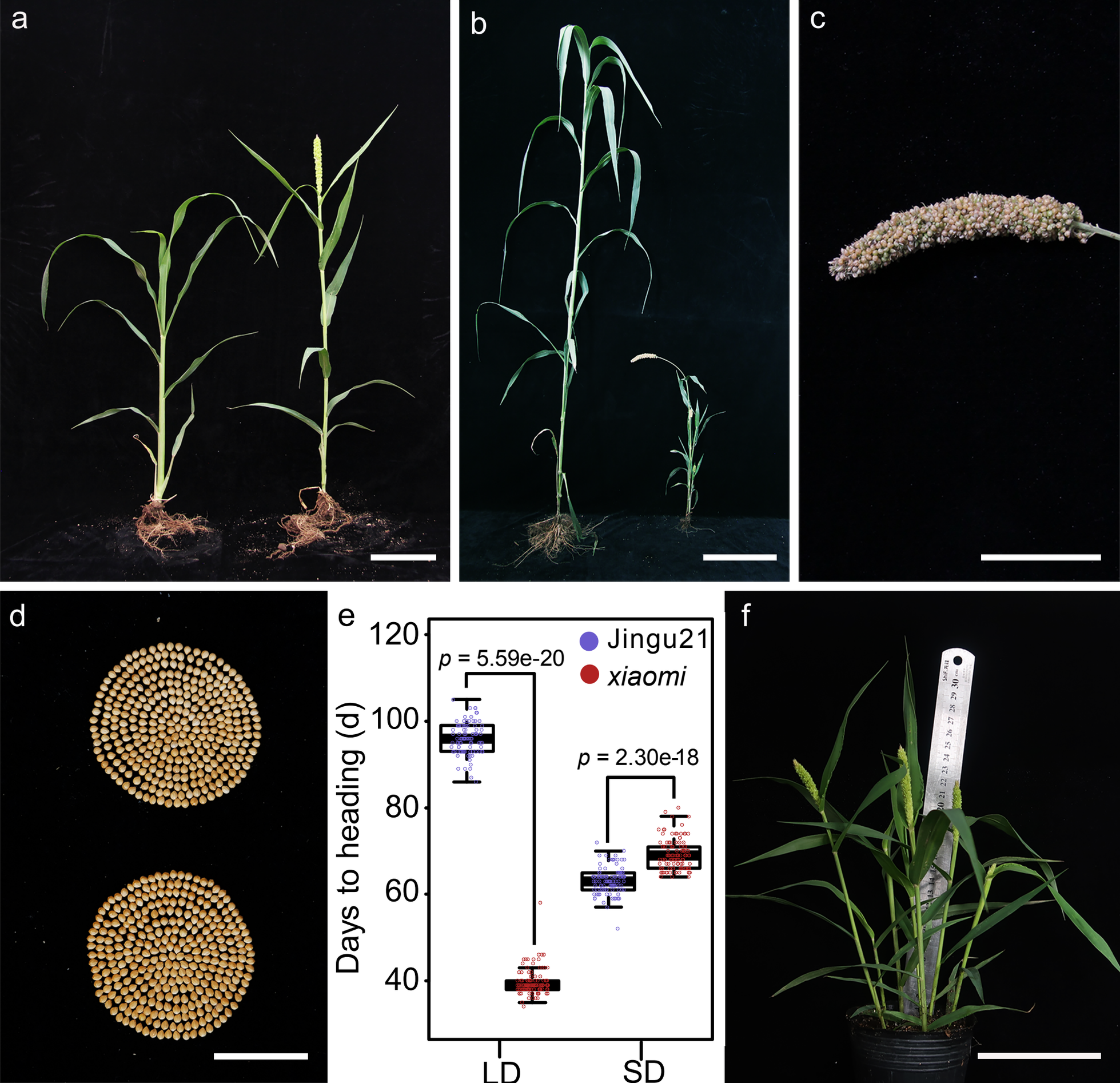
germinated seeds under white (**l**) or UV light (**m**), confirming transmission of transgene to progeny. All experiments were performed for eight independent biological repeats and at least six samples were tested for each biological repeat except for Panel **j** that were performed for three repeats. Scale bars, 2 mm (**a**-**e**, **h**, **k** and **l**), 2 cm (**f**, **g** and **i)**.

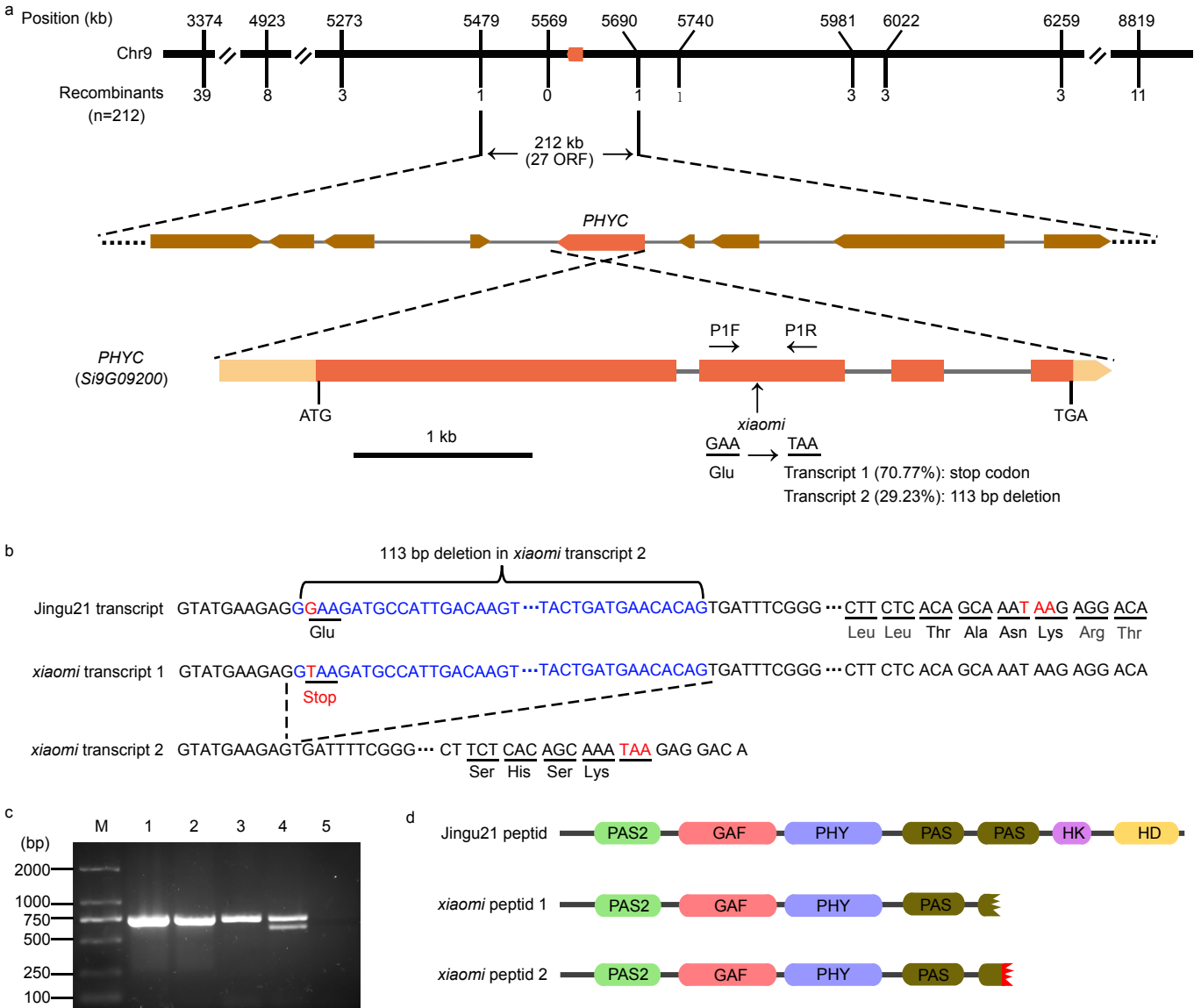
Fig. 6 | Flow chart for *Agrobacterium*-mediated transformation of *xiaomi*

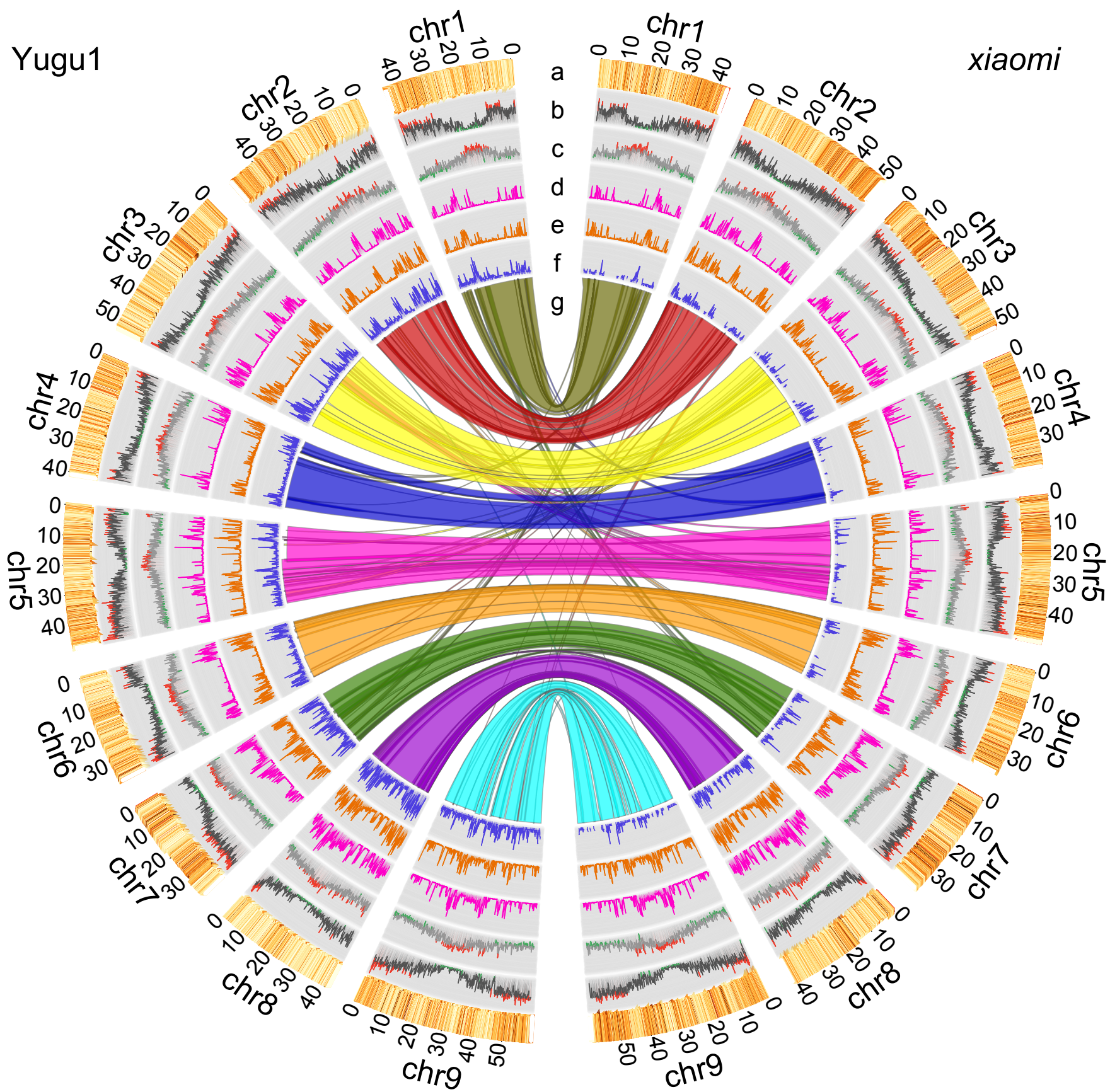
Table 1 | Statistics for the *xiaomi* genome assembly and annotation

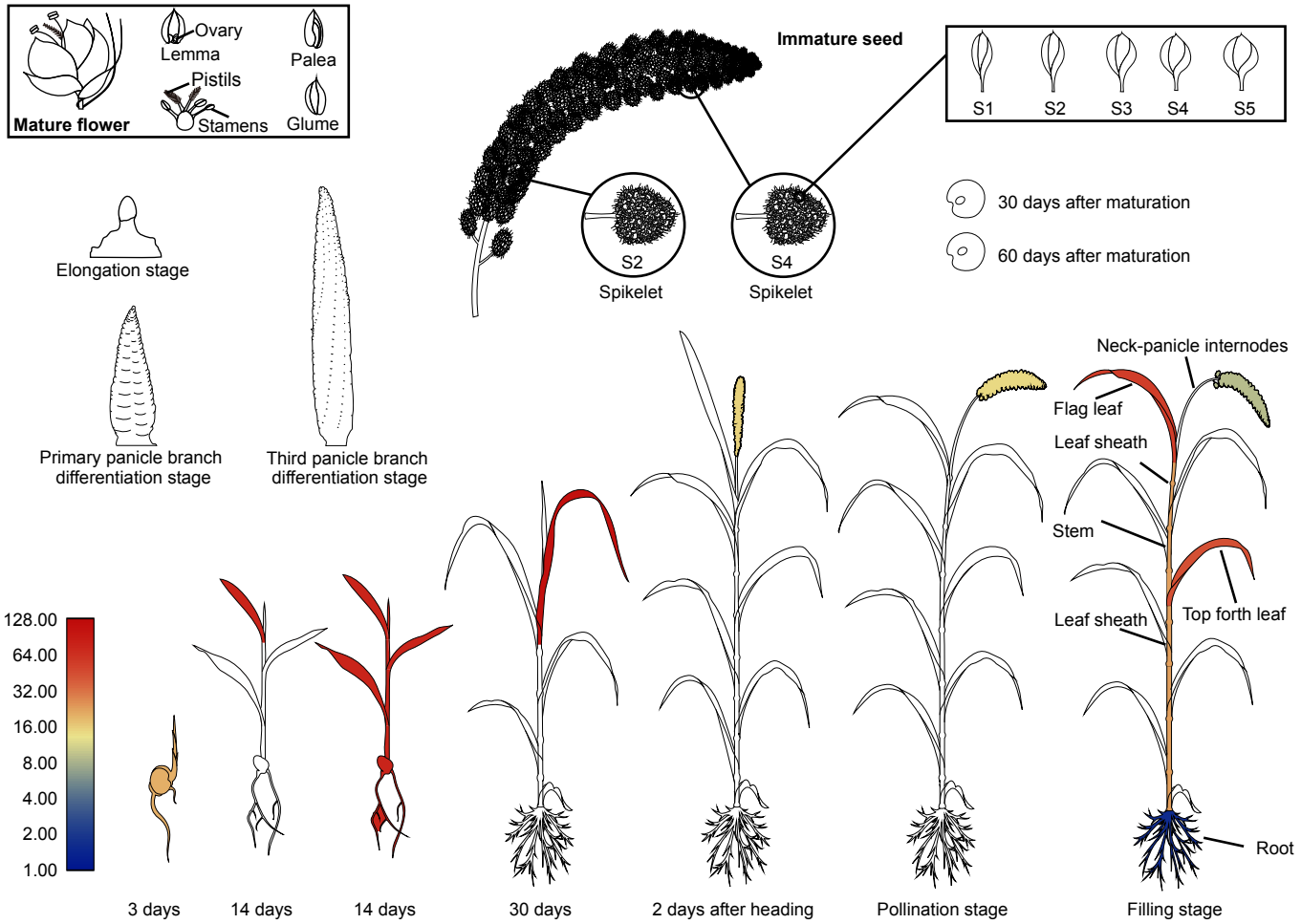
Genome assembly	Estimated genome size		438.26 Mb	
	Assembled genome size		429.94 Mb	
	GC content		45.96%	
	Number of contigs		414	
	Total contig length		429,934,041 bp	
	Longest contig		49,165,788 bp	
	Contig N50		18,838,472 bp	
	Number of scaffolds		366	
	Total scaffold length		429,936,786 bp	
	Longest scaffold		59,244,420 bp	
	Scaffold N50		42,406,388 bp	
	Annotation	**Number**	**Length (bp)**	**Percentage (%)**
Transposable elements	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	

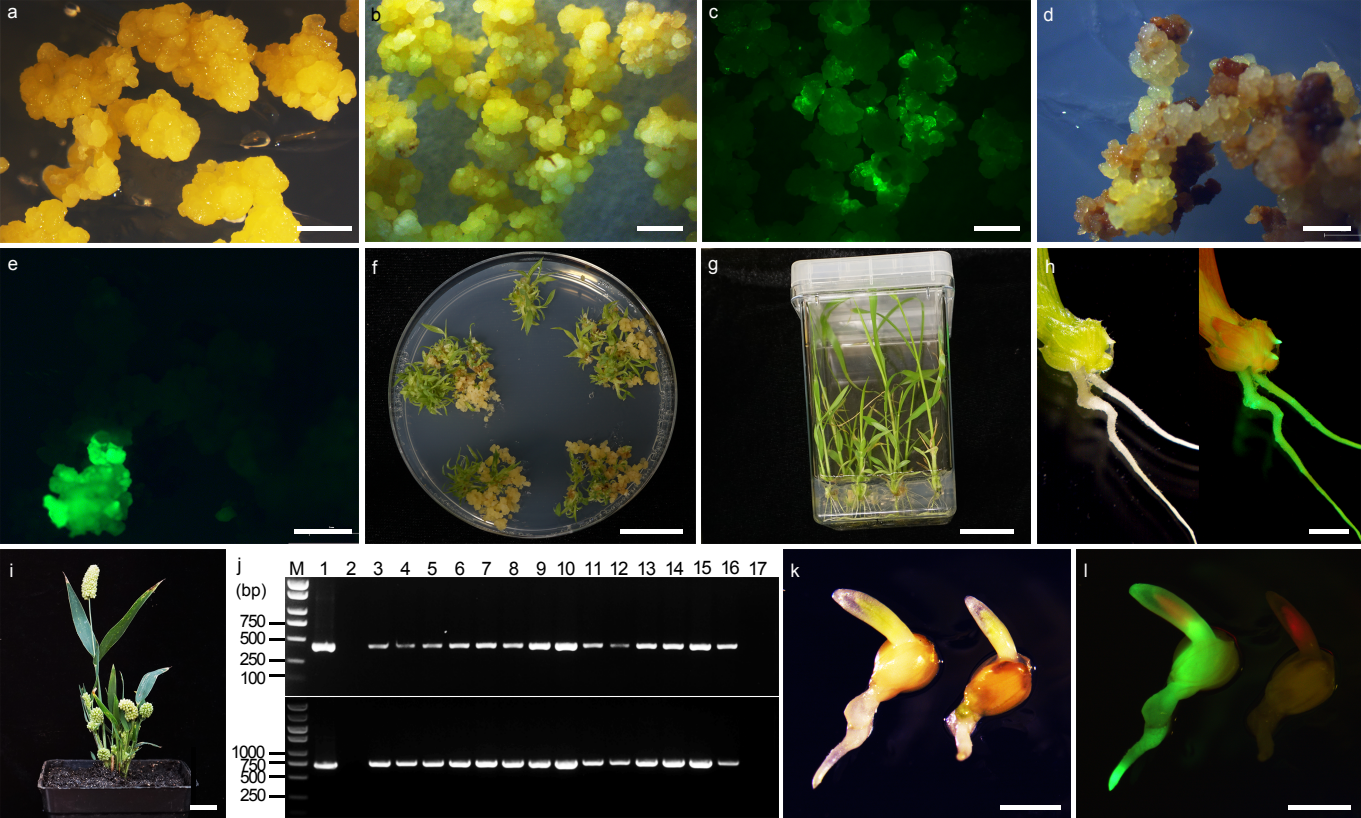
923

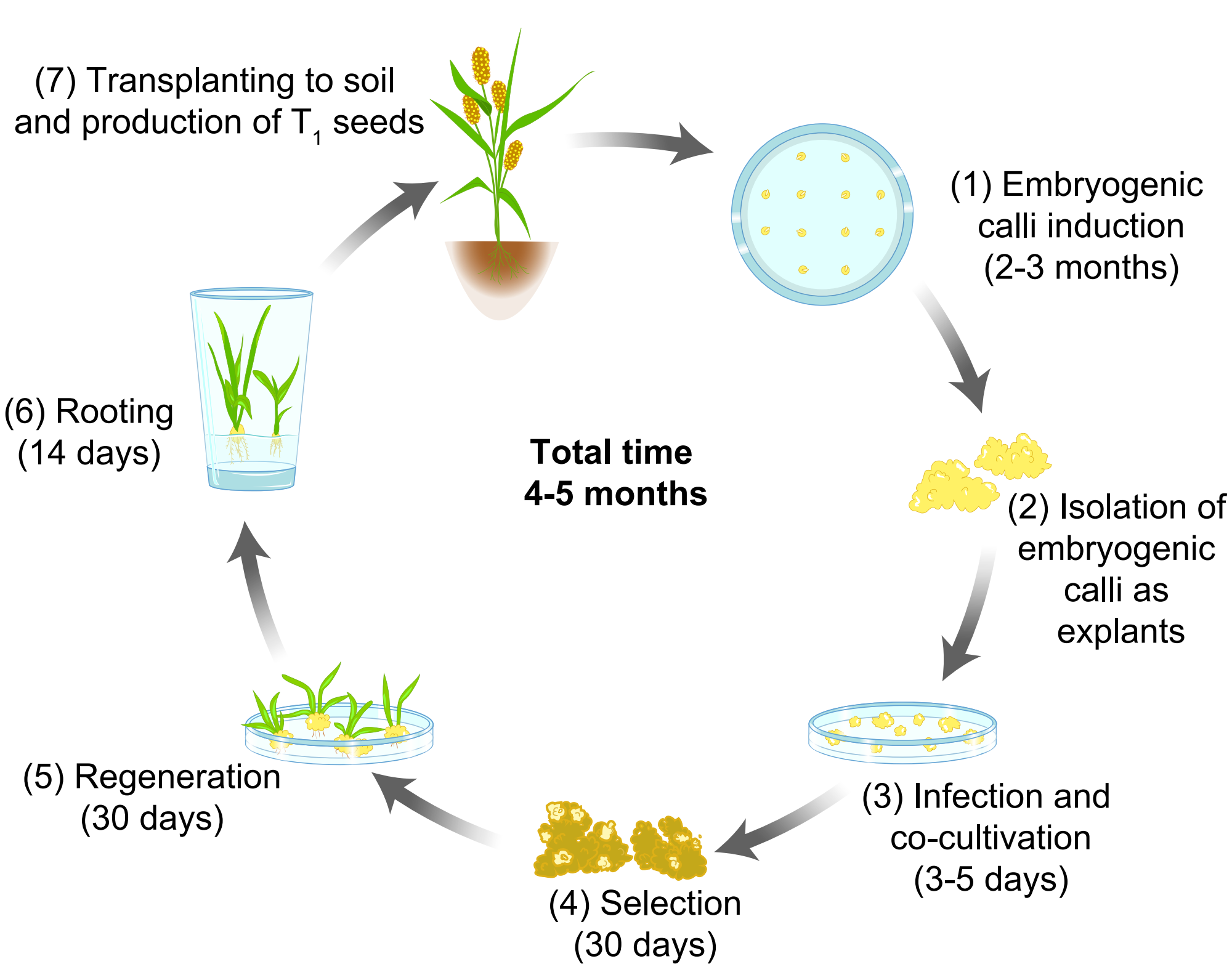

a
b
c
d
e
f
120
100
80
60
40
Days to heading (d)
p = 5.59e-20
p = 2.30e-18
Jingu21
xiaomi
LD
SD


a
Position (kb)
3374
4923
5273
5479
5569
5690
5740
5981
6022
6259
8819
Chr9
Recombinants
(n=212)
39
8
3
1
0
1
1
3
3
3
11
212 kb
(27 ORF)
PHYC
PHYC
(Si9G09200)
P1F
P1R
ATG
TGA
xiaomi
1 kb
GAA
TAA
Glu
Transcript 1 (70.77%): stop codon
Transcript 2 (29.23%): 113 bp deletion
b
113 bp deletion in xiaomi transcript 2
Jingu21 transcript
GTATGAAGAGGGAAGATGCCATTGACAAGT···TACTGATGAACACAGTGATTTCGGG···CTT CTC ACA GCA AAT AAG AGG ACA
Glu
Leu Leu Thr Ala Asn Lys Arg Thr
xiaomi transcript 1
GTATGAAGAGGTAAGATGCCATTGACAAGT···TACTGATGAACACAGTGATTTCGGG···CTT CTC ACA GCA AAT AAG AGG ACA
Stop
xiaomi transcript 2
GTATGAAGAGTGATTTTCGGG··· CT TCT CAC AGC AAA TAA GAG GAC A
Ser His Ser Lys
c
M
1
2
3
4
5
(bp)
2000
1000
750
500
250
100
d
Jingu21 peptid
PAS2
GAF
PHY
PAS
PAS
HK
HD
xiaomi peptid 1
PAS2
GAF
PHY
PAS
xiaomi peptid 2
PAS2
GAF
PHY
PAS


Yugu1
xiaomi
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
a
b
c
d
e
f
g
a GC content (%) 35 ≥50
b Gene density 0 27
c TE content 0 97
d SNP density 0 3,150
e InDel density 0 470
f PAV density 0 15

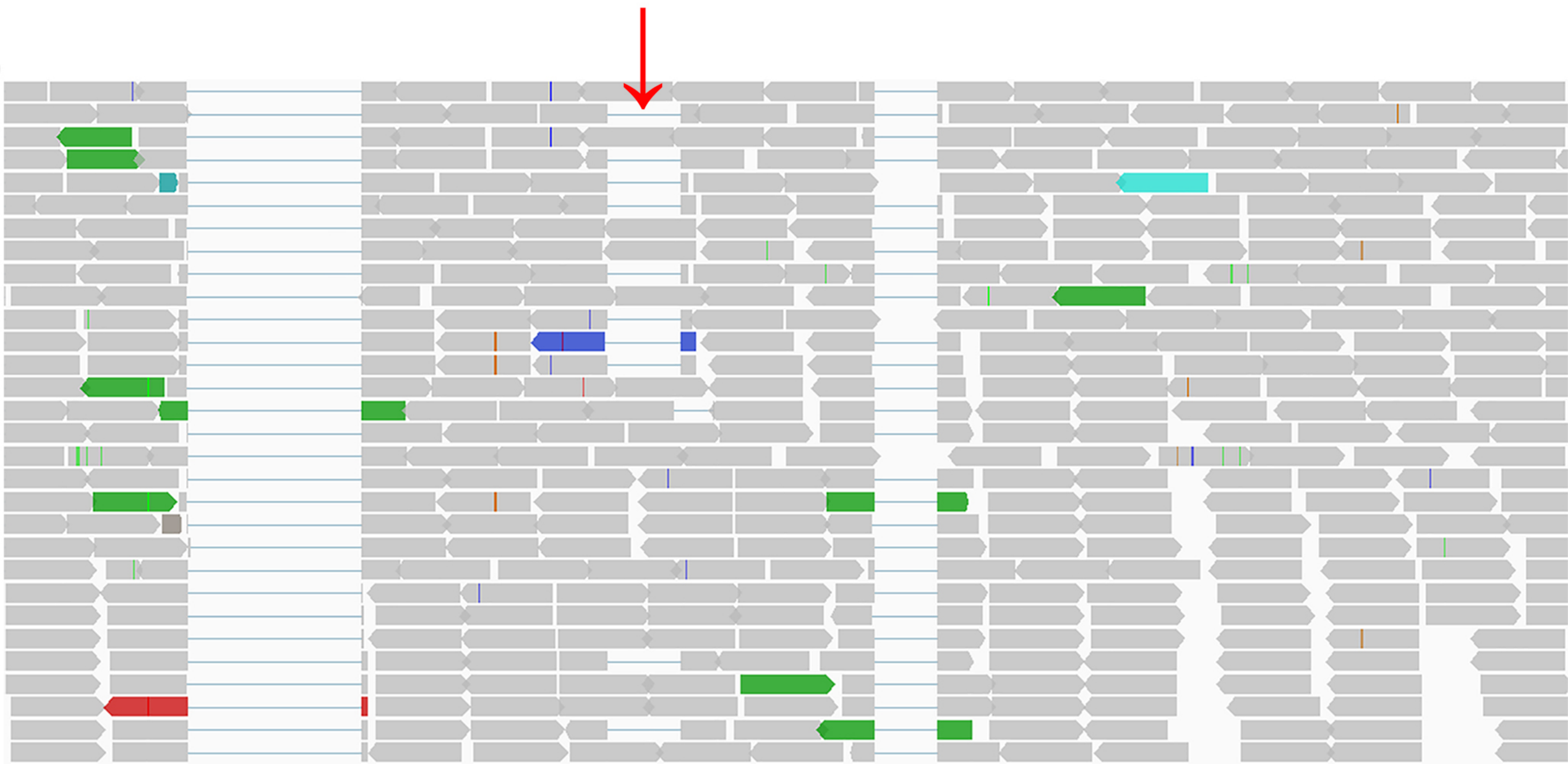

Mature flower
Ovary
Lemma
Pistils
Stamens
Palea
Glume
Immature seed
S1
S2
S3
S4
S5
S2
Spikelet
S4
Spikelet
30 days after maturation
60 days after maturation
Elongation stage
Primary panicle branch differentiation stage
Third panicle branch differentiation stage
Neck-panicle internodes
Flag leaf
Leaf sheath
Stem
Leaf sheath
Top forth leaf
Root
128.00
64.00
32.00
16.00
8.00
4.00
2.00
1.00
3 days
14 days
14 days
30 days
2 days after heading
Pollination stage
Filling stage

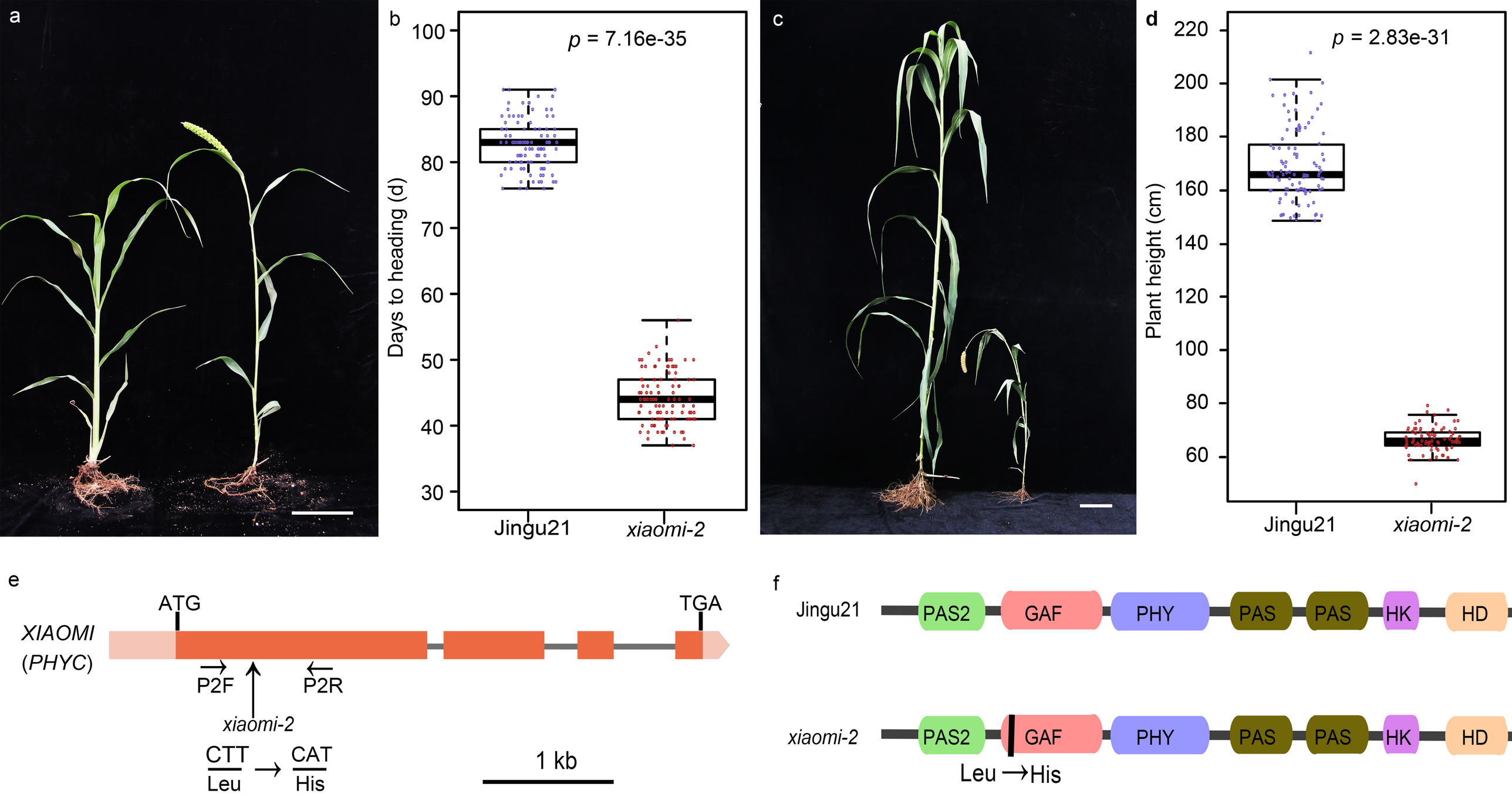

a
b
c
d
e
f
g
h
i
j
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
(bp)
750
500
250
100
1000
750
500
250
k
l

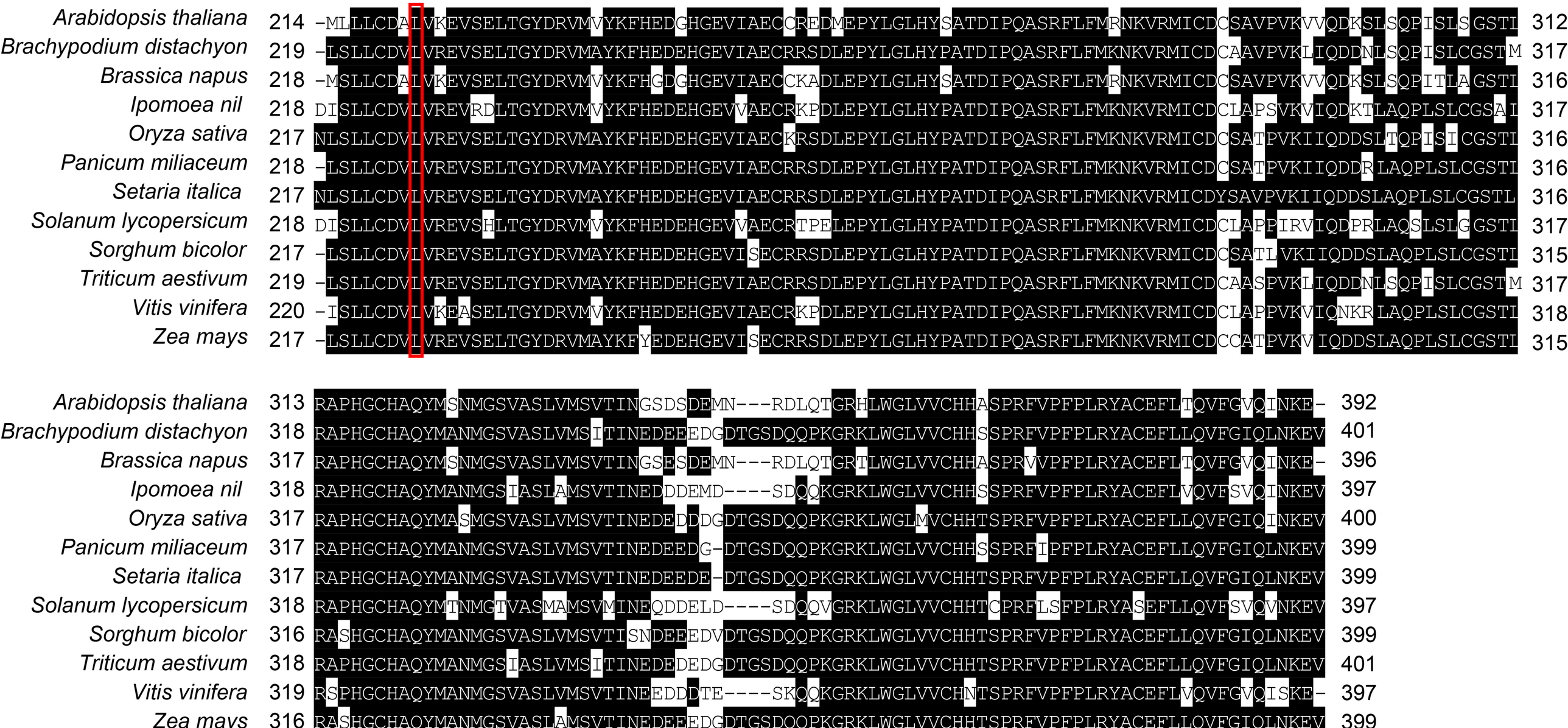

(7) Transplanting to soil and production of T_1 seeds
(1) Embryogenic calli induction (2-3 months)
(2) Isolation of embryogenic calli as explants
(3) Infection and co-cultivation (3-5 days)
(4) Selection (30 days)
(5) Regeneration (30 days)
(6) Rooting (14 days)
Total time 4-5 months

a

b




a
b
p = 7.16e-35
Days to heading (d)
100
90
80
70
60
50
40
30
Jingu21
xiaomi-2
c
d
p = 2.83e-31
Plant height (cm)
220
200
180
160
140
120
100
80
60
Jingu21
xiaomi-2
e
XIAOMI
(PHYC)
ATG
TGA
P2F
P2R
xiaomi-2
CTT
Leu
CAT
His
1 kb
f
Jingu21
PAS2
GAF
PHY
PAS
PAS
HK
HD
xiaomi-2
PAS2
GAF
PHY
PAS
PAS
HK
HD
Leu →His


Arabidopsis thaliana 214 -MLLLCDALVKEVSELTGYDRVMVYKFHEDGHGEVIAECCREDMEPYLGLHYSATDIPQASRFLFMRNKVRMICDCSAVPVKVVQDKSLSQPISLSGSTI 312
Brachypodium distachyon 219 -LSLLCDVLVREVSELTGYDRVMAYKFHEDEHGEVIAECRRSDLEPYLGLHYPATDIPQASRFLFMKNKVRMICDCAAVPVKLIQDDNLSQPISLCGSTM 317
Brassica napus 218 -MSLLCDALVKEVSELTGYDRVMVYKFHGDGHGEVIAECCKADLEPYLGLHYSATDIPQASRFLFMRNKVRMICDCSAVPVKVVQDKSLSQPITLAGSTI 316
Ipomoea nil 218 DISLLCDVLVREVRDLTGYDRVMVYKFHEDEHGEVVAECRKPDLEPYLGLHYPATDIPQASRFLFMKNKVRMICDCLAPSVKVIQDKTLAQPLSLCGSAI 317
Oryza sativa 217 NLSLLCDVLVREVSELTGYDRVMAYKFHEDEHGEVIAECKRSDLEPYLGLHYPATDIPQASRFLFMKNKVRMICDCSATPVKIIQDDSLTQPISICGSTI 316
Panicum miliaceum 218 -LSLLCDVLVREVSELTGYDRVMAYKFHEDEHGEVIAECRRSDLEPYLGLHYPATDIPQASRFLFMKNKVRMICDCSATPVKIIQDDRLAQPLSLCGSTI 316
Setaria italica 217 NLSLLCDVLVREVSELTGYDRVMAYKFHEDEHGEVIAECRRSDLEPYLGLHYPATDIPQASRFLFMKNKVRMICDYSAVPVKIIQDDSLAQPLSLCGSTL 316
Solanum lycopersicum 218 DISLLCDVLVREVSHLTGYDRVMVYKFHEDEHGEVVAECRTPELEPYLGLHYPATDIPQASRFLFMKNKVRMICDCLAPPIRVIQDPRLAQSLSLGGSTI 317
Sorghum bicolor 217 -LSLLCDVLVREVSELTGYDRVMAYKFHEDEHGEVISECRRSDLEPYLGLHYPATDIPQASRFLFMKNKVRMICDCSATLVKIIQDDSLAQPLSLCGSTI 315
Triticum aestivum 219 -LSLLCDVLVREVSELTGYDRVMAYKFHEDEHGEVIAECRRSDLEPYLGLHYPATDIPQASRFLFMKNKVRMICDCAASPVKLIQDDNLSQPISLCGSTM 317
Vitis vinifera 220 -ISLLCDVLVKEASELTGYDRVMVYKFHEDEHGEVIAECRKPDLEPYLGLHYPATDIPQASRFLFMKNKVRMICDCLAPPVKVIQNKRLAQPLSLCGSTI 318
Zea mays 217 -LSLLCDVLVREVSELTGYDRVMAYKFYEDEHGEVISECRRSDLEPYLGLHYPATDIPQASRFLFMKNKVRMICDCCATPVKVIQDDSLAQPLSLCGSTI 315
Arabidopsis thaliana 313 RAPHGCHAQYMSNMGSVASLVMSVTINGSDSDEMN---RDLQTGRHLWGLVVCHHASPRFVPFPLRYACEFLTQVFGVQINKE- 392
Brachypodium distachyon 318 RAPHGCHAQYMANMGSVASLVMSITINEDEEEDGDTGSDQQPKGRKLWGLVVCHHSSPRFVPFPLRYACEFLLQVFGIQLNKEV 401
Brassica napus 317 RAPHGCHAQYMSNMGSVASLVMSVTINGSESDEMN---RDLQTGRTLWGLVVCHHASPRVVPFPLRYACEFLTQVFGVQINKE- 396
Ipomoea nil 318 RAPHGCHAQYMANMGSIASLAMSVTINEDDDEMD----SDQQKGRKLWGLVVCHHSSPRFVPFPLRYACEFLVQVFSVQINKEV 397
Oryza sativa 317 RAPHGCHAQYMASMGSVASLVMSVTINEDEDDDGDTGSDQQPKGRKLWGLMVCHHTSPRFVPFPLRYACEFLLQVFGIQINKEV 400
Panicum miliaceum 317 RAPHGCHAQYMANMGSVASLVMSVTINEDEEDG-DTGSDQQPKGRKLWGLVVCHHSSPRFIPFPLRYACEFLLQVFGIQLNKEV 399
Setaria italica 317 RAPHGCHAQYMANMGSVASLVMSVTINEDEEDE-DTGSDQQPKGRKLWGLVVCHHTSPRFVPFPLRYACEFLLQVFGIQLNKEV 399
Solanum lycopersicum 318 RAPHGCHAQYMTNMGTVASMAMSVMINEQDDELD----SDQQVGRKLWGLVVCHHTCPRFLSFPLRYASEFLLQVFSVQVNKEV 397
Sorghum bicolor 316 RASHGCHAQYMANMGSVASLVMSVTISNDEEEDVDTGSDQQPKGRKLWGLVVCHHTSPRFVPFPLRYACEFLLQVFGIQLNKEV 399
Triticum aestivum 318 RAPHGCHAQYMANMGSIASLVMSITINEDEDEDGDTGSDQQPKGRKLWGLVVCHHTSPRFVPFPLRYACEFLLQVFGIQLNKEV 401
Vitis vinifera 319 RSPHGCHAQYMANMGSVASLVMSVTINEEDDDTE----SKQQKGRKLWGLVVCHNTSPRFVPFPLRYACEFLVQVFGVQISKE- 397
Zea mays 316 RASHGCHAQYMANMGSVASLAMSVTINEDEEEDGDTGSDQQPKGRKLWGLVVCHHTSPRFVPFPLRYACEFLLQVFGIQLNKEV 399

Chr 7
Chr 6
Chr 4
Chr 8
Chr 1
Chr 5
Chr 2
Chr 3
Chr 9

log2
(N links)
4
3
2
1
0

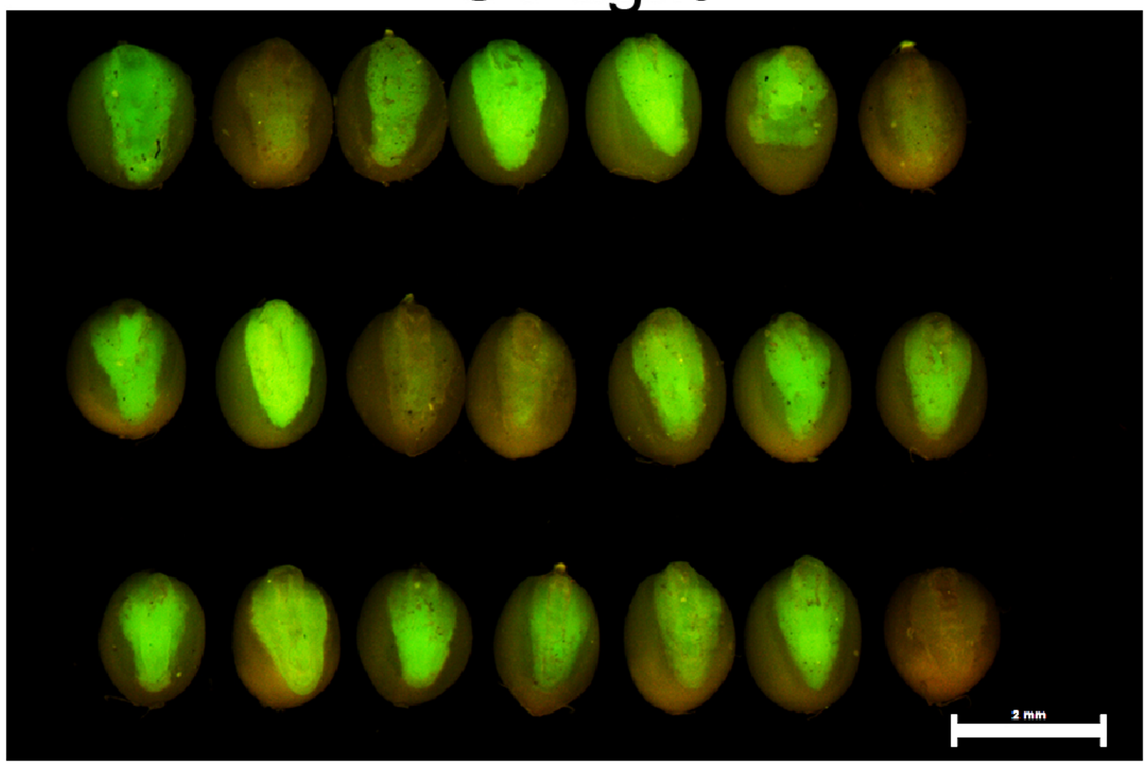
Chr 9
Chr 3
Chr 2
Chr 5
Chr 1
Chr 8
Chr 4
Chr 6
Chr 7

White light

UV light

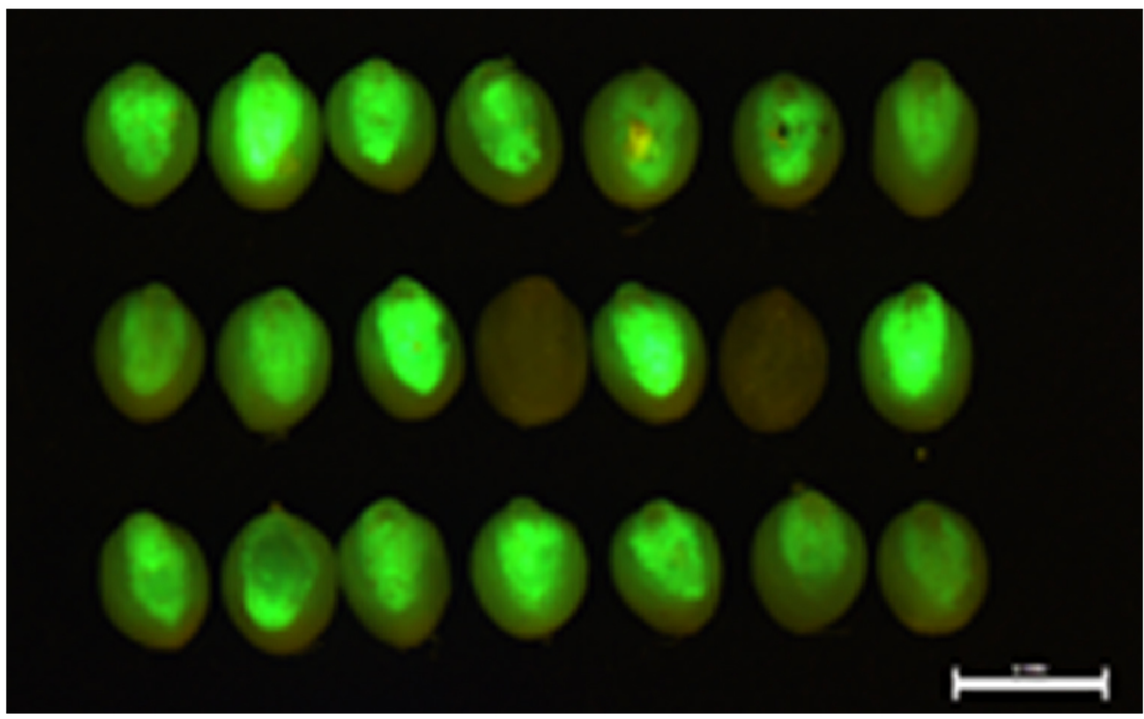
a

One



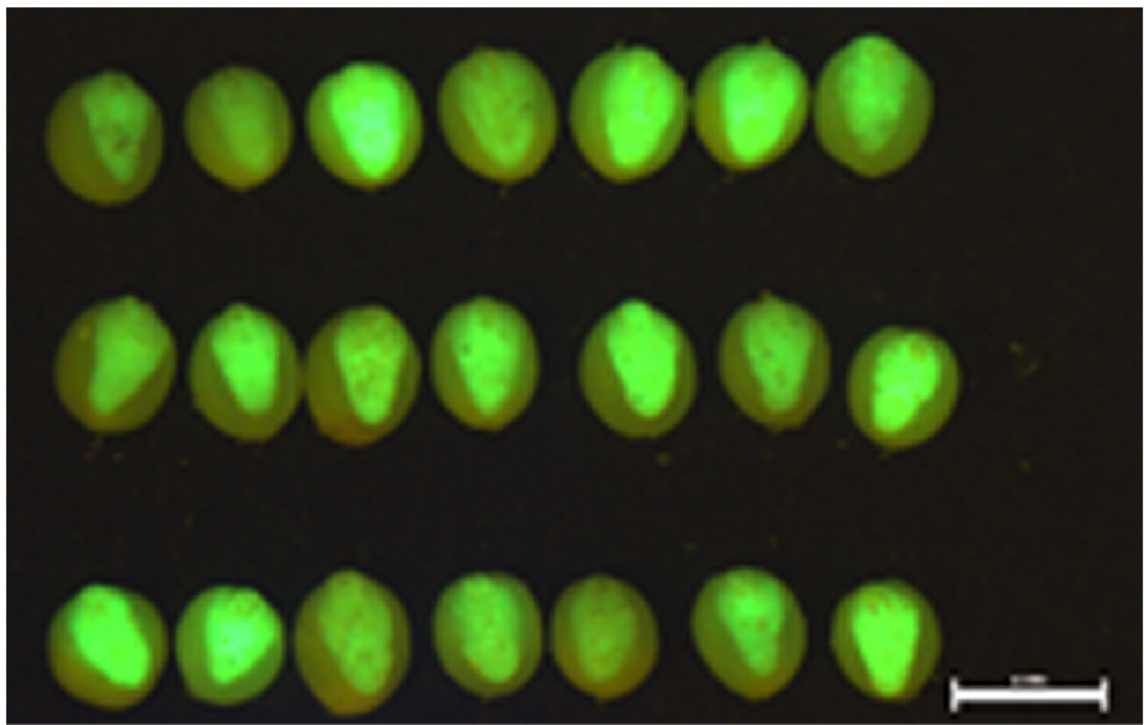
b

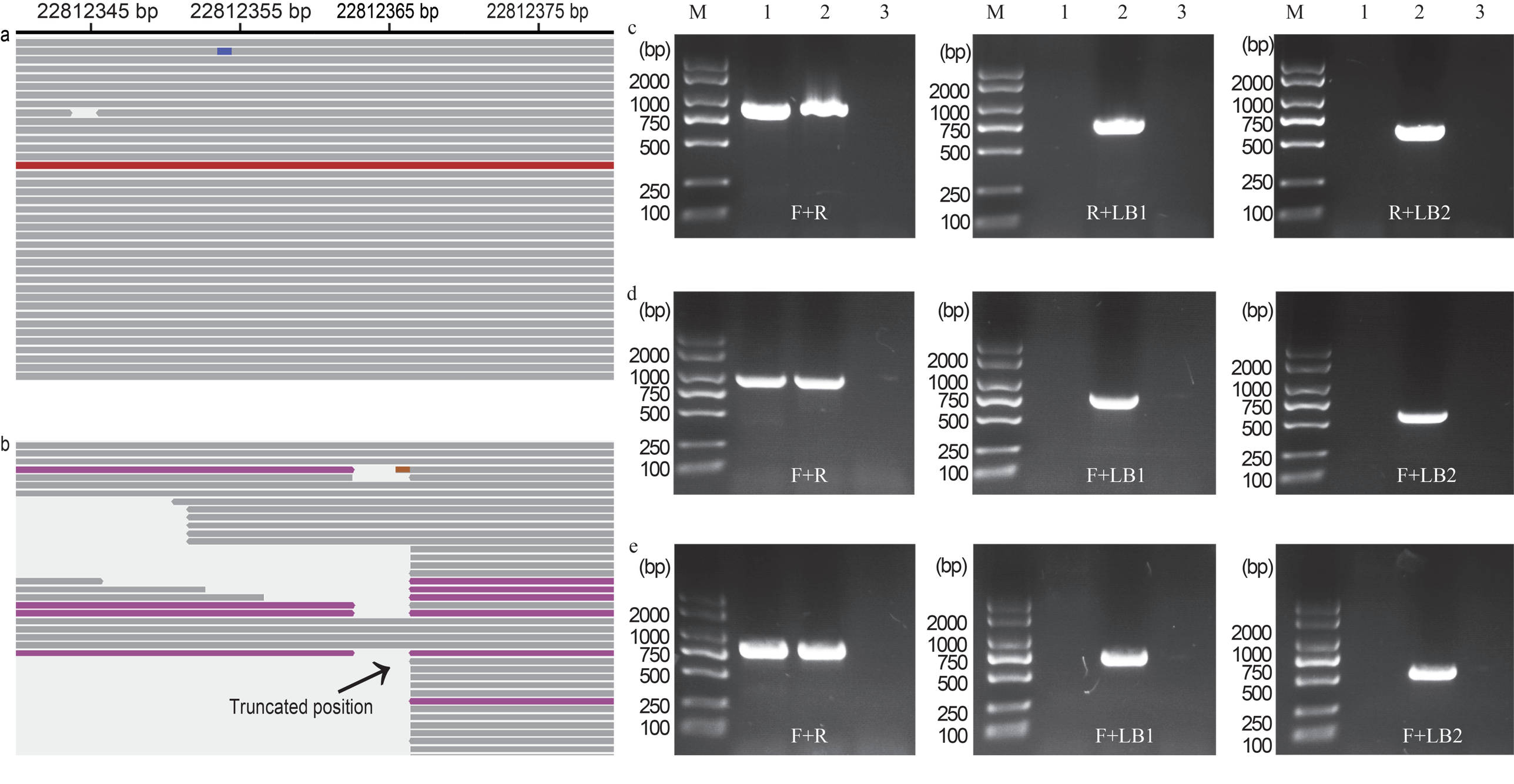
Two



c

Multiple




a
22812345 bp
22812355 bp
22812365 bp
22812375 bp
b
Truncated position
c
M
1
2
3
(bp)
2000
1000
750
500
250
100
F+R
R+LB1
R+LB2
d
F+R
F+LB1
F+LB2
e
F+R
F+LB1
F+LB2