The draft genome of the fast growing non-timber forest species Moso Bamboo (*Phyllostachys heterocycla*)

Zhenhua Peng[1,4], Ying Lu[2,4], Lubin Li[1,4], Qiang Zhao[2,4], Qi Feng[2,4], Zhimin Gao[3,4], Hengyun Lu[2], Tao Hu[3], Na Yao[1], Kunyan Liu[2], Yan Li[2], Danlin Fan[2], Yunli Guo[2], Wenjun Li[2], Yiqi Lu[2], Qijun Weng[2], Congcong Zhou[2], Lei Zhang[2], Tao Huang[2], Yan Zhao[2], Chuanrang Zhu[2], Xinge Liu[3], Xuewen Yang[3], Tao Wang[1], Kun Miao[1], Caiyun Zhuang[1], Xiaolu Cao[1], Wenli Tang[3], Guanshui Liu[3], Yingli Liu[3], Jie Chen[1], Zhenjing Liu[1], Licai Yuan[3], Zhenhua Liu[1], Xuehui Huang[2], Tingting Lu[2], Benhua Fei[3], Zemin Ning[2], Bin Han[2*] & Zehui Jiang[1,3*]

[1]Research Institute of Forestry, Chinese Academy of Forestry, Key Laboratory of Tree Breeding and Cultivation, State Forestry Administration, Beijing 100091, China.

[2]National Center for Gene Research，Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 500 Caobao Road, Shanghai 200233, China.

[3]International Center for Bamboo and Rattan, 8 Fu Tong Dong Da Jie, Chaoyang District, Beijing 100102, China.

[4]These authors contributed equally to this work.

[*]Correspondence should be addressed to B.H. (bhan@ncgr.ac.cn); Z.J. (jiangzehui@icbr.ac.cn).

Nature Genetics: doi:10.1038/ng.2569

Supplementary Notes

Bamboo and the individual subjected to genome sequencing.

Bamboo is the general name of plants which belong to a Bambusoideae in Gramineae family, about 1,250 species and 90 genera widely distributed in the tropical and subtropical areas of Asia, Africa, America, and Pacific islands, and a few in the temperate and frigid zone[1].

According to the statistics, there are about 500 species of bamboo and totally 53,800 km^2 of bamboo forest in China, the largest cultivation area and the maximum gross output value of bamboo industry in the world. The export volume of bamboo products was US$ 14 billion in 2009 and the gross output value of bamboo industry was about US$ 130 billion in 2010. Bamboo is one of the most important non-timber forest resources and forest products in terms of their ecological conservation, feasible economic profit and social benefit, increasingly playing an important role in increasing farmer income and promoting local economic development. Fast growing, high productivity, strongly regeneration capability and various benefits of bamboo attracted attention in countries, where it was a sustainable material to supply works of the planting and industry to billions of people.

Moso bamboo is a large woody bamboo with the highest ecological, economic, and cultural values of all bamboos in Asia and accounting for ~70% of total area of bamboo growth and 5 billion US dollars of annual forest production in China[1]. The moso bamboo is one of the plants with an incredible growth speed in the world. The height growth of its shoot is rapid and steady in suitable condition in spring depended on its strong rhizome-dependent system. At the growth peak, the shoot of the moso Bamboo can grow as long as 100-cm within 24 hours, and reach to its maximum height about 20 meters in shortly 45 to 60 days. It grows predominantly by vegetative propagation due to its unique rhizome-dependent system. Bamboo has a very term flowering interval as long as 60 to 120 years. All of these mysteries

Nature Genetics: doi:10.1038/ng.2569

attracted our interesting in its biological research.

The moso bamboo genome contains 24 pairs of chromosomes, as well as basic number of chromosomes characters in other bamboos. Prior to the whole genome sequencing, we checked different moso bamboo communities growing in main planting areas, crossing Eastern, Southeastern, and Southern China. We finally chose the individual growing in the Tianmu-Mountain National Nature Reserve in Zhejiang Province of Eastern China (119º26'55.0”E, 30º19'13.4”N; 480-meter in elevation) where its growth had not been interrupted by human activities for a long time.

The tissues for the transcriptome sequencing.

Five vegetative tissues (young leaves, rhizome, root, tip of the 20cm-high shoot, and tip of the 50cm-high shoot) were collected in the Tianmu-Mountain National Nature Reserve in Zhejiang Province of China in spring, which was from the same individual used in genome sequencing. To perform the transcriptome sequencing of floral tissues, we spent over two years to look for the floral tissues of moso bamboo in 8 provinces of China because its flowering was too rare. Finally, in early summer of 2010, two reproductive tissues (panicles at early stage and panicles at flowering stage) were obtained in suburban Guilin (110º31'20.2”E, 25º10'42.7”N; 216 meter in elevation), Guangxi Province of Southern China, more than 1800 kilometers (1,100 miles) from our institute. The collected panicles from plant with no flowering or post-flowering spikelet were considered as panicle at early stage, while those from the plant growing at least 50% of flowering or post-flowering spikelet were considered as that at flowering stage.

Cytogenetic analysis of moso bamboo chromosomes.

Fluorescence In situ hybridization.

Meiotic pachytene chromosomes were prepared from the root tip of freshly germinated seedling. Individual pachytene chromosomes were identified by fluorescence *in situ* hybridization using rice 45s rDNA probe[2], according the method

Nature Genetics: doi:10.1038/ng.2569

described in Jiang *et al.*[3]. Digital images were recorded from an Olympus BX60 fluorescence microscope.

Estimation of genome size of the moso bamboo by the flow cytometry.

We put 2-month-old leaves from the sequenced bamboo individual into a flow cytometric analysis to estimate genome size as mentioned by Galbraith[4]. Finally, over 10,000 nuclei were analyzed per sample with a FACSAria flow cytometer (Becton, Dickinson and Company), equipped with 488 nm argon laser. 24 samples were analyzed using rice as the standard species. The software BDFACSDiva was used for data analysis with the coefficient variation controlled in 5%. The peak values of the fluorescence intensity of 24 bamboo and rice samples, the genome size of moso bamboo was estimated to be about 2,075.025 ± 13.08 Mb, or 2C DNA about 4.24 pg (1 pg DNA = 0.978×10^9 bp [5]).

Estimation of genome size of the moso bamboo by the frequency of k-mer occurency.

Values of K-mers were plotted against the frequency at their occurency (**Supplementary Fig.3**). At a K-mer size of 51, the peak occurrency is at 36. Study of the panda genome used K-mer frequency to estimate the genome size[6]. To avoid potential over-estimation of genome sized introduced by base errors, we turned to the modified method as described in the paper of the Tasmanian devil genome[7] to estimate the genome size of the moso bamboo. As the definition of genome size, the total number of effective K-mer words divided by the K-mer depth or the K-mer occurrence number at the peak kmer frequency D_p, $G_s = (K_n - K_s)/D_p$. Here K_n is the total number of K-mer words and K_s is the number of single or unique K-mer words. So we estimated the genome size to be (80477036861 - 9537946584) / 36 = 1.97 Gb.

BAC and BAC-end sequencing.

The moso bamboo BAC library was constructed by Amplicon Express, USA,

Nature Genetics: doi:10.1038/ng.2569

composed of 165,888 clones harvested from a Hind-III BAC library with epicenter pCC1BAC Cloning-Ready Vector. The nuclear DNA was isolated from the same individual as used in genome sequencing. Average insert size of bamboo BAC is about 135 Kb. Eight randomly selected BACs were sequenced by using subcloning and standard Sanger sequencing methods. A total of 10,327 BACs were isolated with MACHEREY-NAGEL plasmid and large-construct DNA purification kit (NucleoSpin®96 Flash, Cat. No.740618.24). Both ends of all these BACs were sequenced by the dideoxy chain termination method using BigDye Terminator Cycle sequencing kit V3.1 (Applied Biosystems, Life Technologies). BAC end sequencing was carried out on ABI3730xl DNA analyzer. The assembling of the BACs and re-basecalling of raw BAC-end sequences were performed by the PHRED and PHRAP programs[8]. Manual editing was utilized to validate the accuracy of the re-basecalled reads.

Prediction of protein-coding genes.

We build a 7-step pipeline to construct the gene model set (**Supplementary Fig. 7**).

1) The prediction software program, FgeneSH++ with gene model parameters trained from monocots, was used in *ab initio* gene prediction to build the preliminary gene models.

2) Coding sequences of each predicted gene model were aligned to both the Repbase TE library and the moso bamboo TE library created by RepeatModeler, using the Blastn at E-value of 1e-5.

3) The Illumina RNA-seq sequences from five vegetative and two reproductive tissues were mapped onto the coding sequences of FgeneSH gene models by the aligner SMALT with parameters set to minimum Smith-Waterman (-m) at 80 (for 2*120 bp reads) or 60 (for 2*100 bp reads), maximum insert size (-i) at 1,500, and minimum insert size (-j) at 20. Only uniquely matched reads were selected in assistance with gene prediction. Information between unique matches and corresponding gene models were collected by using 2 thresholds to screen SMALT cigar outputs: A) cigar:S and cigar:A items with score at 50 or more were selected. B)

Nature Genetics: doi:10.1038/ng.2569

for the selected cigar:A items, the available paired-end insert size should be at least 200 bp.

4) A total of 8,253 moso bamboo cDNAs, carrying entire coding sequences but not TE-derived, were selected from the 10,608 putative full-length cDNAs and were then mapped to the scaffolds by an mRNA/EST genome mapping program of GMAP[9] with the parameters set to "-n 1 -f 2 -B 1 -A -t 4".

5) The gene models were screened by integrating the information from outputs of the step 2), 3), and 4), using the following 4 thresholds.

A, the gene models coding TE-elements or overlapping TE-elements greater than 10% of gene coding region were firstly discarded.

B, the models aligned to the full-length cDNAs were preferentially collected. The splicing sites were manually adjusted according to the alignment.

C, the models detected not by the FgeneSH++ but by the cDNAs were created and their coding information was added into gene model set.

D, candidate gene models without the evidence of full-length cDNAs should be supported by 2 different uniquely matched RNA-seq sequences. And at least 20% of their coding region was covered by RNA-seq reads. A pair of PE reads was treated as a single RNA-seq sequence when counting number of the mapped transcriptome reads for each model.

6) Information of cDNA-supporting UTR ends was attached to the gene model set.

7) The single-exon genes were manually checked by experts and the genes with no hits to homologs of grass genes were also discarded.

8) For the gene with different transcripts, the longest one was selected.

Comparison of parameters of gene models among plant genomes.

We did comparative analysis between the bamboo genes and the genes identified from *Arabidopsis*, Brachypodium, rice, sorghum, and maize. The bamboo gene models exhibited very high similarity to other grass species in all of these parameters, such as the distribution of gene length, coding sequences (CDS), exon length, intron length, GC content in coding region, and exon number per gene

Nature Genetics: doi:10.1038/ng.2569

(**Supplementary Fig. 8**). Of compared species, only dicot *Arabidopsis* is obviously different from any other species in gene length, CDS GC content, and intron length.

Comparison of the assembled scaffolds to the available moso bamboo sequences in database.

The assemblies were compared with available sequences in the public database to assess the genomic coverage and assembling accuracy. In the GenBank till October of 2011, there have been 1,086 genome survey sequences, and 18 gene sequences. Alignment of the known genomic sequences with the length over 2 Kb to our assemblies showed that over 98% of sequence region were covered by the assembled scaffolds and 91% covered by a single best match (**Supplementary Table 2**). Similar coverage, 96% of all matches and 94% of the single best match, was observed in alignment of the rest 996 genome survey sequences less than 2 Kb. Most of the sequences with low coverage have putative sequencing errors because of more biases distribution within them. The known 18 gene coding sequences with total length of 28,741 bp were parallelly mapped onto the assemblies. Almost each one had a perfect match located on a single scaffold except unmatched bases at the end of the genes (**Supplementary Table 3**). The average sequence identities in aligned region were over 98%. Our manual check revealed that most unmatched bases at the end should be the low quality bases introduced by sequencing of amplified DNA fragments. Prior to this study, over ten thousand putative full-length cDNAs were cloned and sequenced for the moso bamboo. Of them, 8,253 cDNA sequences were picked up when those TE-derived, false ORF-coding, and non-moso-bamboo items were removed. We mapped the cDNAs onto the assemblies by means of GMAP. A total of 8118 (98.4% of 8,253) cDNAs were uniquely aligned to the assembled scaffolds with very high identities (averagely at 99.1%, **Supplementary Fig. 4**). All of the sequence comparison consisted with the estimated 98% coverage of genome assembly.

To evaluate the quality of whole genome shotgun assembly, the assembled scaffolds were aligned to 8 finished bacterial artificial chromosomes (BACs)

Nature Genetics: doi:10.1038/ng.2569

sequences with average length 133 Kb by Sanger sequencing technology. Seven BACs were well aligned to a single scaffold and 1 BACs were aligned to 2 scaffolds each (**Supplementary Fig. 5**). The coverage of the scaffolds and initial contigs on the BACs were up to 88.8% and 98.8% (**Supplementary Table 4**), which supported our estimation of whole genome coverage of the assemblies. The frequency of single-base difference and insertion/deletion were approximately 0.19 and 0.09 per Kb, without regard to heterozygous single nucleotide polymorphisms (SNPs) and short indels detected by the annotation. The average PE read depth in aligned regions was at 100- to 132-fold coverage. The incompatible bases were inclined to located near the unclosed gaps, indicative of assembling quality were lower at the end of the initial contigs. Prior to sequence alignment, we have removed most of the sequence errors in assembly of the Sanger BACs. The detected SNPs or short indels were probably derived from potential heterozygosity or low rate assembling errors.

Repeat annotation.

The *de novo* repeat annotation revealed that the moso bamboo genome comprised approximately 59% transposable elements (TEs). Detection of the TEs in the Sanger-BACs showed 53% of TE content, similar with that in whole genome. With comparison to other grass species, the moso bamboo genome had similar TE content to that of the sorghum (62%) (main text ref. 36), and more TE content than rice (40%) (main text ref.18) and Brachypodium (28%)[10], but lower than maize (84%) (main text ref.36, 26). Of the observed TEs, retrotransposons were the dominating repetitive sequences (39%), as well as 9.5% of DNA transposons. Like the rice, sorghum and maize genomes, the most abundant repeats in bamboo were long-terminal repeat elements (LTRs), 24.6% of *Gypsy*-type LTRs and 12.3% of *Copia*-type LTRs.

Bamboo genome has the highest copy numbers of TEs, *Gypsy/Copia*-type LTR retrotransposons and En/Spm transposons. Rice and sorghum have the highest copy numbers of MITE transposons (Tourist & Stowaway), and Harbinger transposons

Nature Genetics: doi:10.1038/ng.2569

and *Gypsy/Copia*-type LTR retrotransposons, respectively (**Supplementary Table 10b and 10c**). It was inferred that insertion of the LTR retrotransposons played the most important role in expansion of the higher plant genome size, though some DNA transposons also had very high copies. The TEs covering approximately 11% of the bamboo genome were not classified. However, their average unit length was within the range of the LTRs, implying that there should still be some unknown TEs active in the bamboo. A total of 9,412 intact LTR retrotransposons were predicted in our genome assemblies. The average length was approximately 10.3 Kb, 3 times of the average gene length.

We performed *de novo* prediction for LTR retrotransposons with LTRharvest and LTR_FINDER on the large-sized scaffolds (>10 Kb), using default parameter but -maxdistltr set at 30,000. The quality criteria were the existence of one or more typical retrotransposon protein domains and the simple-repeat/tandem-repeat content less than 35%. A total of 9,142 remaining candidates were considered as the full-length LTRs, including 3,103 *Gypsy*-type (PR-RT-INT), 2,677 *Copia*-type (PR-INT-RT), and 3,632 unclassified.

Identification of gene synteny and whole-genome duplication.

Detection of syntenic genes between bamboo and rice and between bamboo and sorghum.

To generate a pair-wise alignment of gene models between bamboo and rice (MSU RGAP 6.1) and between bamboo and sorghum (v1.4), 30,379 bamboo genes located on the larger scaffolds (>50 Kb) were aligned to the reference gene models by Blastp with E-value < 1e-20. The evidence-based gene prediction approach scarcely concerned gene colinearity between bamboo and other grasses and also expected to miss some genes. So two criteria were used to call syntenic gene blocks in bamboo scaffolds: i) number of the genes in one syntenic block >= 5; ii) number of non-syntenic bamboo genes between two adjacent syntenic genes <5. A perl script and following manual check was applied to determine the syntenic blocks and breakpoints between the blocks.

Nature Genetics: doi:10.1038/ng.2569

Identification of the WGD by investigating the collinear orthologous genes between bamboo and rice.

According the location of rice genes in chromosome, the collinear gene blocks of bamboo were mapped to the rice chromosomes. The overlapped gene blocks were manually checked to remove redundancy. As shown in **Supplemental Fig. 10a**, moso bamboo seemed to carry as two duplicates as that of rice gene model sets, though lots of bamboo gene lost within these regions during the duplication. It was suggested that the large-scale genome duplication in bamboo resulted from whole genome duplication not from segmental duplication, which supported a tetraploid origin of bamboo. Interestingly, rice chromosomes are characteristic of diploid (2*n*=24) while moso bamboo chromosomes are 2*n*=48. There might be unknown connection between them. We collected the orthologous pairs of bamboo, which share unique rice ortholog in collinear blocks, to estimate the divergence time with universal substitution rate of 6.5×10^{-9} mutations per site per year. Thus, estimated 7 to 15 mya was the potential time when two bamboo porgenitors diverged **(Supplemental Fig.10b**), and the WGD should occur more recent. However, the moso bamboo chromosomes now exhibit to be a diploid according to the FISH analysis. Consequently, we speculated that there is a long progress from tetraploidization to diploidization in moso bamboo since approximately 7-12 mya. But for the bamboo species with different chromosome number or ploidy number, the process should be variable.

Annotation of gene function and comparison of fundamental pathways.

Prediction of gene function motifs and domains were performed by Interpro[11] against available databases, including ProDom, PRINTS, Pfam, Panther, Profile, PIR, Smart, and Pattern. The gene functional ontology was retrieved from the outputs of InterPro using Gene Ontology[12]

The bamboo gene models were aligned to entries of sorghum, rice, and maize from the KEGG database (release till April 2011) by Blastp under E-value 1e-10 to find the best hit for each gene. The similarity of each pathway is the ratio of number of shared enzymatic steps and sum of referenced enzymatic steps. For instance, similarity (bamboo vs. rice) = number of the enzymatic steps shared by rice and

Nature Genetics: doi:10.1038/ng.2569

bamboo / sum of the rice-gene-involved enzymatic steps.

Annotation of conserved non-coding RNA (ncRNA) genes.

Identification of transfer RNAs (tRNA).

The tRNAScan-SE[13] algorithms with default parameters were applied to prediction of tRNA genes in the Arabidopsis, sorghum, maize, rice, Brachypodium, and bamboo genomes. Bamboo had 1,167 tRNA genes in the assemblies, nearly 0.5 times as many as that of maize because most of bamboo pseudogenes were not detected in the current assemblies (containing 10% unclosed gaps). The same analysis of the Arabidopsis[14] and Brachypodium[10] found 699 and 615 tRNA loci, respectively, very closed to 711 and 614 identified by their Genome Initiatives, suggesting that most of bamboo tRNAs had been found. Of all conserved tRNA genes, selenocysteine and suppressor tRNAs were involved in a special coding way by the stop codens. It was interesting that 6 Selenocysteine and 1 suppressor tRNAs were detected in the bamboo genome, which were only found in maize and sorghum of Panicoideae but not in its sister groups, Brachypodium of Pooideae and rice of Ehrhartoideae.

Identification of rRNA genes.

The rRNA fragments were identified by aligning the rRNA template sequences (Rfam database[15], release 10.0) of *Arabidopsis thaliana, Oryza sativa, Sorghum bicolor*, and *Zea mays* using Blastn with E-value at 1e-10 and identity cutoff at 95% or more.

Identification of other non-coding RNA genes.

The miRNA and snRNA genes were predicted by INFERNAL[16] software against the Rfam database (release 9.1, 1,412 families). To accelerate the speed, a rough filtering prior to INFERNAL was performed by Blastn against the Rfam sequence database under E-value at 1. For the miRNA prediction, the assemblies were aligned to the precursor sequences of *Arabidopsis thaliana, Brachypodium distachyon, Oryza sativa, Sorghum bicolor, Saccharum officinarum, Triticum aestivum, Hordeum vulgar*e, and *Zea mays*, derived from the Rfam sequence database. The extended

Nature Genetics: doi:10.1038/ng.2569

sequences covering detected loci region and 50 bp franking sequences from both ends of the region were put into the INFERNAL prediction with cutoff score at 30 or more. The predicted mature sequences of the bamboo miRNA were aligned to gene model set to detected miRNA target genes with Blastn under E-value at 1e-10. In the snRNA predictions, the assemblies were firstly aligned to the snRNA sequences of *Arabidopsis thaliana*, *Oryza sativa, Sorghum bicolor, Triticum aestivum* and *Zea mays* from Rfam database. The extended sequences, similar as that in miRNA prediction, were put into the INFERNAL prediction with cutoff score at 50 or more.

The C/D snoRNA were predicted using snoScan[17] with the yeast rRNA 16 methylation sites and yeast rRNA sequences provided by the snoScan distribution. The minimum cutoff score was based on the settings which yield a false positive rate of 30 bits. Similarly, H/ACA snoRNAs were detected by snoGPS using the yeast score tables and target pseudouridines[18].

Quantity of all predicted non-coding RNA genes was listed in **Supplementary Table 9**.

Reconstruction of phylogeny among 6 fully sequenced grass genomes.

The OrthoMCL clustered a total of 968 single-copy gene families among 6 fully grass and 2 dicot genomes, which was used to reconstruct the phylogeny. The coding sequences of the genes were concatenated to a supergene for each species. When the best substitution model (GTR+gamma+I) were determined by Modeltest[19], the supergene sequences were subjected to phylogenetic analyses by Mrbayes (main text ref.54) with the parameter set to 1,000,000 (1 sample / 100 generations) and the first 250 sample were burned in. Two independent runs reached the same result using Arabidopsis as an outgroup. Branch-specific *dN* and *dS* were estimated with codeml of PAML[20]. The output of OrthoMCL and phylogenic tree structure were subjected to a computational analysis of changes in gene family size with the software CAFE (Online methods ref. 55).

Estimation of divergence time of paralogous pairs.

Nature Genetics: doi:10.1038/ng.2569

To estimate the divergence time of the paralogs, we selected the gene families consisting of exactly 2 members to calculate the Ks of the pairs because the 2-member gene families had a single divergence relationship between two members. Thus, 3,786 bamboo 2-member gene clusters were put into the calculation of Ks, together with 2,552 rice, 2,161 sorghum, 1,874 Brachypodium, 3,285 maize, and 2,665 foxtail millet clusters. The Ks was calculated by a model that averages parameters across 14 candidate models (*P*-value < 0.001)[30] and then converted to the divergence time using a substitution rate of 6.5×10^{-9} mutations per site per year.

Estimation of divergence time between orthologous genes.

To estimate the divergence time, the Ks values were calculated between orthologs of bamboo and other genome from the 968 single-copy gene clusters determined by the OrthoMCL calculation. The Ks distribution of the one-to-one orthologous pairs of bamboo-Brachypodium, bamboo-rice, bamboo-foxtail-millet, bamboo-sorghum, bamboo-maize, and bamboo-wheat suggested the different divergence time between bamboo and other grass genome, which was consisted with the phylogenic relationship generated by Mrbayes analysis. The mean Ks was used to estimated the divergence time between different genomes. The internal duplication during the WGD was estimated by calculating the Ks of the paralogs in 2-member gene families of bamboo and maize, which was then converted to the divergence time to indicate the WGD time.

Calling of heterozygous SNPs and small indels.

To detect the heterozygous sequence polymorphism, all of the used PE reads (around 120× coverage) were firstly mapped to the assembled scaffolds by aligner SMALT. The SNPs were then called by SSAHA_Pileup (version 0,8) and 6 thresholds were used to post-filter unreliable SNPs: 1) SSAHA_Pileup SNP score >= 20; 2) ratio of two alleles between 3:17 to 17:3; 3) the highest sequencing depth of SNP position <= 240; 4) the lowest sequencing depth for each allele >= 5; 5) the

Nature Genetics: doi:10.1038/ng.2569

minimum distance for adjacent SNPs >= 10 bp; 6) only one polymorphism detected at each SNP position. The small indel (length <= 6 bp) were called by the Pindel[21] and 4 thresholds were used to remove unreliable small indels: 1) length of indels <=6 bp; 2) the highest sequencing depth of indel position <= 240; 3) the lowest sequencing depth for each allele >= 5; 4) ratio of gapped and ungapped reads at the indel position between 3:17 to 17:3.

Nature Genetics: doi:10.1038/ng.2569

References

1. Jiang, Z.H. World Bamboo and Rattan (in Chinese). *Liaoning Science and Technology Publishing House, Shenyang, China* (2002).
2. Fukui, K., Ohmido, N. & KhushG, S. Variability in rDNA loci in the genus *Oryza* detected through fluorescence *in situ* hybridization. *Theor. Appl. Genet.* **87**, 893-899 (1994).
3. Jiang, J. & Gill, B.S. Sequential chromosome banding and *in situ* hybridization analysis. *Genome* **36**, 792-795 (1993).
4. Galbraith, D.W. *et al.* Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* **220**, 1049-1051 (1983).
5. Dolezel, J., Greilhuber, J. & Suda, J. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* **2**, 2233-2244 (2007).
6. Li, R. *et al.* The sequence and *de novo* assembly of the giant panda genome. *Nature* **463**, 311-317 (2009).
7. Murchison, E.P. *et al.* Genome Sequencing and Analysis of the *Tasmanian Devil* and Its Transmissible Cancer. *Cell* **148**, 780-791 (2012).
8. Ewing, B. & Green, P. Base-calling of automated sequencer traces using PHRED. II. Error probabilities. *Genome Res.* **8**, 186-194 (1998).
9. Wu, T.D. & Watanabe, C.K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875 (2005).
10. Vogel, J.P. *et al.* Genome sequencing and analysis of the model grass *brachypodium distachyon*. *Nature* **463**, 763-768 (2010).
11. Mulder, N.J. *et al.* New developments in the InterPro database. *Nucleic Acids Res.* **35**, D224-228 (2007).
12. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25-29 (2000).
13. 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).
14. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815 (2000).
15. Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33** (Database issue), D121-124 (2005).
16. Nawrocki, E.P., Kolbe, D.L. & Eddy, S.R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335-1337 (2009).

Nature Genetics: doi:10.1038/ng.2569

17. Lowe, T.M. & Eddy, S.E. A computational screen for methylation guide snoRNAs in yeast. *Science* **283**, 1168-1171 (1999).
18. Schattner, P. *et al.* Genome-wide Searching for Pseudouridylation Guide snoRNAs: Analysis of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res.* **32**, 4281-4296 (2004).
19. Posada, D. & Crandall, K.A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818 (1998).
20. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**, 1586-1591 (2007).
21. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865-2871 (2009).
22. Thurston, M.I. & Field, D, Msatfinder: detection and characterization of microsatellites. [http://www.genomics.ceh.ac.uk/msatfinder/] website Oxford UK (2005).
23. Tadege, M. *et al.* Reciprocal control of flowering time by *OsSOC1* in transgenic Arabidopsis and by *FLC* in transgenic rice. *Plant Biotechnol. J.* **1**, 361-369 (2003).
24. Endo-Higashi, N. & Izawa, T. Flowering time genes *Heading date 1* and *Early heading date 1* together control panicle development in rice. *Plant Cell Physiol.* **52**, 1083-1094 (2011).
25. Li, D. *et al.* Functional characterization of rice *OsDof12*. *Planta* **229**, 1159-1169 (2009).
26. Li, D., Yang, C., Li, X., Ji, G. & Zhu, L. Sense and antisense *OsDof12* transcripts in rice. *BMC Mol. Biol.* **17**, 80 (2008).
27. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
28. Tanaka, T. *et al.* The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Res.* **36** (Database issue), D1028-1033 (2008). Rap-db IRGSP5 at http://rapdb.dna.affrc.go.jp/
29. Zhang, X.M. *et al. De Novo* Sequencing and Characterization of the Floral Transcriptome of *Dendrocalamus latiflorus* (Poaceae: Bambusoideae). *PLoS One* **7**:e42082 (2012).
30. Zhang, Z. *et al.* KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* **4**, 259-263 (2006).

Nature Genetics: doi:10.1038/ng.2569

31. Cocuron, J.C. *et al.* A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. *Proc. Natl. Acad. Sci. USA.* **104**, 8550-8555(2007).
32. Liepman, A.H., Wilkerson, C.G. & Keegstra, K. Expression of cellulose synthase-like (*Csl*) genes in insect cells reveals that *CslA* family members encode mannan synthases. *Proc. Natl. Acad. Sci. USA.* **102**, 2221-2226 (2005).
33. Davis, J., Brandizzi, F., Liepman, A.H. & Keegstra, K. Arabidopsis mannan synthase CSLA9 and glucan synthase CSLC4 have opposite orientations in the Golgi membrane. *Plant J.* **64**, 1028-1037 (2010).
34. Wan, L. *et al.* Transcriptional Activation of *OsDERF1* in *OsERF3* and *OsAP2-39* Negatively Modulates Ethylene Synthesis and Drought Tolerance in Rice. *PLoS One* **6**, e25216 (2011).
35. Xiang, Y., Tang, N., Du, H., Ye, H. & Xiong L. Characterization of *OsbZIP23* as a Key Player of the Basic Leucine Zipper Transcription Factor Family for Conferring Abscisic Acid Sensitivity and Salinity and Drought Tolerance in Rice. *Plant Physiol.* **148**, 1938-1952 (2008).
36. Nijhawan, A., Jain, M., Tyagi, A.K. & Khurana, J.P. Genomic Survey and Gene Expression Analysis of the Basic Leucine Zipper Transcription Factor Family in Rice. *Plant Physiol.* **146**, 333-350 (2008).
37. Jeong, J.S. *et al.* Root-Specific Expression of *OsNAC10* Improves Drought Tolerance and Grain Yield in Rice under Field Drought Conditions. *Plant Physiol*. **153**, 185-197 (2010).
38. Hu, H. *et al.* Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proc. Natl. Acad. Sci. USA.* **103**, 12987-12992 (2006).
39. Huang, J. *et al.* SRWD: A novel WD40 protein subfamily regulated by salt stress in rice (*Oryza sativa* L.). *Gene* **424**, 71-79 (2008).
40. Toriba, T. *et al.* Molecular characterization the YABBY gene family in *Oryza sativa* and expression analysis of *OsYABBY1*. *Mol Genet Genomics* **277**, 457-468 (2007).
41. Sato, Y. & Yokoya, S. Enhanced tolerance to drought stress in transgenic rice plants overexpressing a small heat-shock protein, sHSP17.7. *Plant Cell Rep.* **27**, 329-334 (2008).
42. Kurusu, T. *et al.* Regulation of Microbe-Associated Molecular Pattern-Induced Hypersensitive Cell Death, Phytoalexin Production, and Defense Gene Expression by Calcineurin B-Like Protein-Interacting Protein Kinases,

Nature Genetics: doi:10.1038/ng.2569

OsCIPK14/15, in Rice Cultured Cells. *Plant Physiol.* **153**, 678-692 (2010).

43. Pi-Fang L.C. *et al.* Induction of a cDNA clone from rice encoding a class II small heat shock protein by heat stress, mechanical injury, and salicylic acid. *Plant Science,* **172**, 64-75 (2007).
44. Lee, B.H. *et al.* Expression of the chloroplast-localized small heat shock protein by oxidative stress in rice. *Gene* **245**, 283-290 (2000).
45. Zou, J. *et al.* Expression analysis of nine rice heat shock protein genes under abiotic stresses and ABA treatment. *J Plant Physiol*. **166**, 851-861 (2009).
46. Liu, J.G. *et al. OsHSF7* gene in rice, *Oryza sativa* L., encodes a transcription factor that functions as a high temperature receptive and responsive factor. *BMB Rep.* **42**, 16-21 (2009).
47. Singh, A. *et al. OsHsfA2c* and *OsHsfB4b* are involved in the transcriptional regulation of cytoplasmic *OsClpB* (Hsp100) gene in rice (*Oryza sativa* L.). *Cell Stress Chaperones* **17**, 243-254 (2012).
48. Welinder, K.G. *et al.* Structural diversity and transcription of class III peroxidases from *Arabidopsis thaliana*. *Eur. J. Biochem.* **269**, 6063-6081 (2002).
49. Rizhsky, L. *et al.* When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. *Plant Physiol.* **134**, 1683-1696 (2004).
50. Tehseen, M., Cairns, N., Sherson, S. & Cobbett, C.S. Metallochaperone-like genes in *Arabidopsis thaliana*. *Metallomics* **2**, 556-564 (2010).
51. Murphy, A., Zhou, J. Goldsbrough, P.B., Taizk,. L. Purification and immunological identification of metallothioneins 1 and 2 from *Arabidopsis thaliana*. *Plant Physiol.* **113**, 1293-1301 (1997).
52. Abe, H. *et al.* Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell* **9**, 1859-1868 (1997).
53. Weig, A., Deswarte, C. & Chrispeels, M.J. The major intrinsic protein family of Arabidopsis has 23 members that form three distinct groups with functional aquaporins in each group. *Plant Physiol.* **114**, 1347-1357 (1997).
54. Imaizumi, T., Schultz, T.F., Harmon, F.G., Ho, L.A. & Kay, S.A. FKF1 F-box protein mediates cyclic degradation of a repressor of *CONSTANS* in Arabidopsis. *Science* **309**, 293-297 (2005).
55. Takahashi, Y. & Shimamoto, K. *Heading date 1* (*Hd1*), an ortholog of Arabidopsis *CONSTANS*, is a possible target of human selection during domestication to diversify flowering times of cultivated rice. *Genes Genet Syst*.

Nature Genetics: doi:10.1038/ng.2569

86, 175-182 (2011).

56. Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M.& Araki, T. A pair of related genes with antagonistic roles in mediating flowering signals. *Science* **286**, 1960-2 (1999).
57. Nakagawa, M., Shimamoto, K. & Kyozuka, J. Overexpression of *RCN1* and *RCN2*, rice *TERMINAL FLOWER 1/CENTRORADIALIS* homologs, confers delay of phase transition and altered panicle morphology in rice. *Plant J.* **29**, 743-750 (2002).
58. Ryu, J.Y., Park, C.M.& Seo, P.J. The floral repressor *BROTHER OF FT AND TFL1* (*BFT*) modulates flowering initiation under high salinity in Arabidopsis. *Mol. Cells* **32**, 295-303 (2011).
59. Xi, W. & Yu, H. *MOTHER OF FT AND TFL1* regulates seed germination and fertility relevant to the brassinosteroid signaling pathway. *Plant Signal Behav.* **5**, 1315-1317 (2010).
60. Valdés, A.E. *et al. Arabidopsis thaliana TERMINAL FLOWER2* is involved in light-controlled signalling during seedling photomorphogenesis. *Plant Cell Environ.* **35**, 1013-1025 (2012).
61. Rao, N.N., Prasad, K., Kumar, P.R. & Vijayraghavan, U. Distinct regulatory role for *RFL*, the rice *LFY* homolog, in determining flowering time and plant architecture. *Proc Natl Acad Sci U S A.* **105**, 3646-3651 (2008).
62. Liu. C., Xi, W., Shen, L., Tan, C. & Yu, H. Regulation of floral patterning by flowering time genes. *Dev. Cell* **16**, 711-722 (2009).
63. Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M.& Araki, T. A pair of related genes with antagonistic roles in mediating flowering signals. *Science* **286**, 1960-1962 (1999).
64. Hanano, S. & Goto, K. Arabidopsis *TERMINAL FLOWER 1* is involved in the regulation of flowering time and inflorescence development through transcriptional repression. *Plant Cell* **23**, 172-3184 (2011).
65. Gocal, G.F. *et al.* GAMYB-like genes, flowering, and gibberellin signaling in Arabidopsis. *Plant Physiol.* **127**, 1682-1693 (2011).
66. Banas, A.K., Łabuz, J,. Sztatelman, O., Gabrys, H. & Fiedor, L. Expression of enzymes involved in chlorophyll catabolism in Arabidopsis is light controlled. *Plant Physiol*. **157**, 1497-1504 (2011).
67. El-Assal, S. E. D. *et al.* The role of Cryptochrome 2 in flowering in Arabidopsis. *Plant Physiol*. **133**, 1504-1516 (2003).
68. Piñeiro, M., Gómez-Mena, C., Schaffer, R., Martínez-Zapater, J.M. & Coupland,

Nature Genetics: doi:10.1038/ng.2569

G. *EARLY BOLTING IN SHORT DAYS* is related to chromatin remodeling factors and regulates flowering in Arabidopsis by repressing *FT*. *Plant Cell* **15**, 1552-1562 (2003).

69. Reeves, P. H., Murtas, G., Dash, S., & Coupland, G. early in short days 4, a mutation in Arabidopsis that causes early flowering and reduces the mRNA abundance of the floral repressor *FLC*. *Development* **129**, 5349-5361 (2002).
70. Sawa, M., Nusinow, D. A., Kay, S. A., & Imaizumi, T. *FKF1* and *GIGANTEA* complex formation is required for day-length measurement in Arabidopsis. *Science* **318**, 261-265 (2007).
71. Lim, M. H. *et al.* A new Arabidopsis gene, *FLK*, encodes an RNA binding protein with K homology motifs and regulates flowering time via *FLOWERING LOCUS C*. *Plant Cell* **16**, 731-740 (2004).
72. Kania, T., Russenberger, D., Peng, S., Apel, K., & Melzer, S. *FPF1* promotes flowering in Arabidopsis. *Plant Cell* **9**, 1327-1338 (1997).
73. Johanson, U. *et al.* Molecular analysis of *FRIGIDA*, a major determinant of natural variation in Arabidopsis flowering time. *Science* **290**, 344-347 (2000).
74. Kardailsky, I. *et al.* Activation tagging of the floral inducer *FT*. *Science* **286**, 1962-1965 (1999).
75. Yamaguchi, A., Kobayashi, Y., Goto, K., Abe, M., & Araki, T. *TWIN SISTER OF FT (TSF)* Acts As a Floral Pathway Integrator Redundantly with *FT*. *Plant Cell Physiol*. **46**, 1175-1189 (2005).
76. Ausin, I., Alonso-Blanco, C., Jarillo, J. A., Ruiz-Garcia, L., & Martinez-Zapater, J. M. Regulation of flowering time by FVE, a retinoblastoma-associated protein. *Nat. Genet.* **36**, 162-166 (2004).
77. Silverstone, A. L., Chang, C., Krol, E., & Sun, T. P. Developmental regulation of the gibberellin biosynthetic gene GA1 in *Arabidopsis thaliana*. *Plant J.* **12**, 9-19 (1997).
78. Peng, J. *et al.* The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses. *Gen. Dev.* **113**, 194-207 (1997).
79. Silverstone, A. L., Ciampaglio, C. N., & Sun, T. The Arabidopsis RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell* **10**, 155-169 (1998).
80. Tyler, L. *et al.* Della proteins and gibberellin-regulated seed germination and floral development in Arabidopsis. *Plant Physiol.* **135**, 1008-1019 (2004).
81. Voegele, A., Linkies, A., Müller, K. & Leubner-Metzger, G. Members of the

Nature Genetics: doi:10.1038/ng.2569

gibberellin receptor gene family *GID1* (*GIBBERELLIN INSENSITIVE DWARF 1*) play distinct roles during *Lepidium sativum* and *Arabidopsis thaliana* seed germination. *J. Exp. Bot.* **62**, 5131-5147 (2011).

82. Lee, H. *et al.* The Arabidopsis HOS1 gene negatively regulates cold signal transduction and encodes a RING finger protein that displays cold-regulated nucleo--cytoplasmic partitioning. *Genes Dev.* **15**, 912-924 (2001).
83. Lee, I., Michaels, S. D., Masshardt, A. S., & Amasino, R. M. The late-flowering phenotype of *FRIGIDA* and mutations in *LUMINIDEPENDENS* is suppressed in the *Landsberg erecta* strain of Arabidopsis. *Plant J.* **6**, 903-909 (1994).
84. Iñigo, S., Alvarez, M.J., Strasser, B., Califano, A. & Cerdán, P.D. *PFT1*, the MED25 subunit of the plant Mediator complex, promotes flowering through *CONSTANS* dependent and independent mechanisms in Arabidopsis. *Plant J.* **69**, 601-612 (2012).
85. Maloof, J. N. *et al.* Natural variation in light sensitivity of Arabidopsis. *Nat. Genet.* **29**, 441–446 (2001).
86. Noh, Y. S. & Amasino, R. M. *PIE1*, an *ISWI* family gene, is required for *FLC* activation and floral repression in Arabidopsis. *Plant Cell* **15**, 1671-1682 (2003).
87. McGinnis, K. M. *et al.* The Arabidopsis *SLEEPY1* gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. *Plant Cell* **15**, 1120-1130 (2003).
88. Lee, J., Oh, M., Park, H. & Lee, I. *SOC1* translocated to the nucleus by interaction with *AGL24* directly regulates leafy. *Plant J.* **55**, 832-843 (2008).
89. Jacobsen, S. E., Binkowski, K. A. & Olszewski, N. E. SPINDLY, a tetratricopeptide repeat protein involved in gibberelling signal transduction in Arabidopsis. *Proc. Natl. Acad. Sci. U S A.* **93**, 9292-9296 (1996).
90. Sung, S. B. & Amasino, R. M. Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. *Nature* **427**, 159-164 (2004).
91. Greb, T. *et al.* The PHD finger protein VRN5 functions in the epigenetic silencing of Arabidopsis FLC. *Curr. Biol.* **17**, 73-78 (2007).
92. Boonburapong, B. & Buaboocha, T. Genome-wide identification and analyses of the rice calmodulin and related potential calcium sensor proteins. *BMC Plant Biol.* **7**, 4 (2007).
93. Jan, A. *et al.* Gibberellin Regulates Mitochondrial Pyruvate Dehydrogenase Activity in Rice. *Plant Cell Physiol.* **47**, 244-253 (2006).
94. Ge, L.F. *et al.* Overexpression of the trehalose-6-phosphate phosphatase gene *OsTPP1* confers stress tolerance in rice and results in the activation of stress

Nature Genetics: doi:10.1038/ng.2569

responsive genes. *Planta* **228**, 191-201 (2008).

95. Huang, Y., Xiao, B. & Xiong, L. Characterization of a stress responsive proteinase inhibitor gene with positive effect in improving drought resistance in rice. *Planta* **226**, 73-85 (2007).
96. Ohnishi, T. *et al. OsNAC6*, a member of the NAC gene family, is induced by various stresses in rice. *Genes Genet Syst.* **80**, 135-139 (2005).
97. Wang, C., Zhang, Q. & Shou H.X. Identification and expression analysis of *OsHsfs* in rice. *J Zhejiang Univ. Sci. B.* **10**, 291-300 (2009).
98. Shan, G.Y. *et al.* Analysis of the rice *SHORT-ROOT5* gene revealed functional diversification of plant neutral/alkaline invertase family. *Plant Science*, **176**, 627-634 (2009).
99. Kawasaki, T. *et al.* Cinnamoyl-CoA reductase, a key enzyme in lignin biosynthesis, is an effector of small GTPase Rac in defense signaling in rice. *Proc Natl. Acad. Sci. U S A.* **103**, 230-235 (2006).
100. Garg, R., Jhanwar, S., Tyagi, A. & Jain, M. Genome-Wide Survey and Expression Analysis Suggest Diverse Roles of Glutaredoxin Gene Family Members During Development and Response to Various Stimuli in Rice. *DNA Res.* **17**, 353-367 (2010).
101. Siriporn, S. *et al.* Exogenous ABA induces salt tolerance in indica rice (*Oryza sativa* L.): The role of *OsP5CS1* and *OsP5CR* gene expression during salt stress. *Env Exp Bot* **86**, 94-105 (2011).
102. Ramamoorthy, R., Jiang, S.Y. & Ramachandran, S. *Oryza sativa* Cytochrome P450 Family Member *OsCYP96B4* Reduces Plant Height in a Transcript Dosage Dependent Manner. *PLoS One* **6**, e28069 (2011).
103. Jung, K.H. *et al.* Wax-deficient anther1 Is Involved in Cuticle and Wax Production in Rice Anther Walls and Is Required for Pollen Development. *Plant Cell* **18**, 3015-3032 (2006).
104. Tatsuro, H., Graham N.S. & Tomio, T. An expression analysis profile for the entire sucrose synthase gene family in rice. *Plant Science* **174**, 534-543 (2008).
105. Kong, Z., Li, M., Yang, W., Xu, W. & Xue, Y. A Novel Nuclear-Localized CCCH-Type Zinc Finger Protein, *OsDOS*, Is Involved in Delaying Leaf Senescence in Rice. *Plant Physiol.* **141**, 1376-1388 (2006).
106. Morsy, M.R. *et al.* The *OsLti6* genes encoding low-molecular-weight membrane proteins are differentially expressed in rice cultivars with contrasting sensitivity to low temperature. *Gene* **344**, 171-180 (2005).

Nature Genetics: doi:10.1038/ng.2569

107. Singh, A., Singh, U., Mittal, D. & Grover, A. Genome-wide analysis of rice *ClpB*/*HSP100*, *ClpC* and *ClpD* genes. *BMC Genomics* **11**, 95 (2010).

108. Wang, Y. *et al.* An ethylene response factor *OsWR1* responsive to drought stress transcriptionally activates wax synthesis related genes and increases wax production in rice. *Plant Mol Biol.* **78**, 275-288 (2012).

109. Ishida, S. *et al.* Allocation of Absorbed Light Energy in PSII to Thermal Dissipations in the Presence or Absence of *PsbS* Subunits of Rice. *Plant Cell Physiol.* **52**, 1822-1831 (2011).

110. Zhong, R. *et al.* Transcriptional Activation of Secondary Wall Biosynthesis by Rice and Maize NAC and MYB Transcription Factors. *Plant Cell Physiol.* **52**, 1856-1871 (2011).

111. Huang, J. *et al.* A novel rice C2H2-type zinc finger protein lacking DLN-box/EAR-motif plays a role in salt tolerance. *Biochim. Biophys. Acta.* **1769**, 220-227 (2007).

112. Zhang C.J. *et al.* An Apoplastic H-Type Thioredoxin Is Involved in the Stress Response through Regulation of the Apoplastic Reactive Oxygen Species in Rice. *Plant Physiol.* **157**, 1884-1899 (2011).

113. Esther, N. *et al.* The expression of the large rice FK506 binding proteins (FKBPs) demonstrate tissue specificity and heat stress responsiveness. *Plant Science* **170**, 695-704 (2006).

114. Akiyama, T., Pillai, M.A. & Sentoku, N. Cloning, characterization and expression of *OsGLN2*, a rice endo-1,3-β-glucanase gene regulated developmentally in flowers and hormonally in germinating seeds. *Planta* **220**, 129-139 (2004).

115. Yadav, S.R., Khanday, I., Majhi, B.B., Veluthambi, K. & Vijayraghavan, U. Auxin-Responsive *OsMGH3*, a Common Downstream Target of *OsMADS1* and *OsMADS6*, Controls Rice Floret Fertility. *Plant Cell Physiol.* **52**, 2123-2135 (2011).

116. Li, H. *et al.* Rice *MADS6* Interacts with the Floral Homeotic Genes *SUPERWOMAN1*, *MADS3*, *MADS58*, *MADS13*, and *DROOPING LEAF* in Specifying Floral Organ Identities and Meristem Fate. *Plant Cell* **23**, 2536-2552 (2011).

117. Seino, J. *et al.* Characterization of rice nucleotide sugar transporters capable of transporting UDP-galactose and UDP-glucose. *J Biochem.* **148**, 35-46 (2010).

118. Wang, R., Shen, W.B., Liu, L.L., Jiang, L. & Wan, J.M. Cloning and expression of

Nature Genetics: doi:10.1038/ng.2569

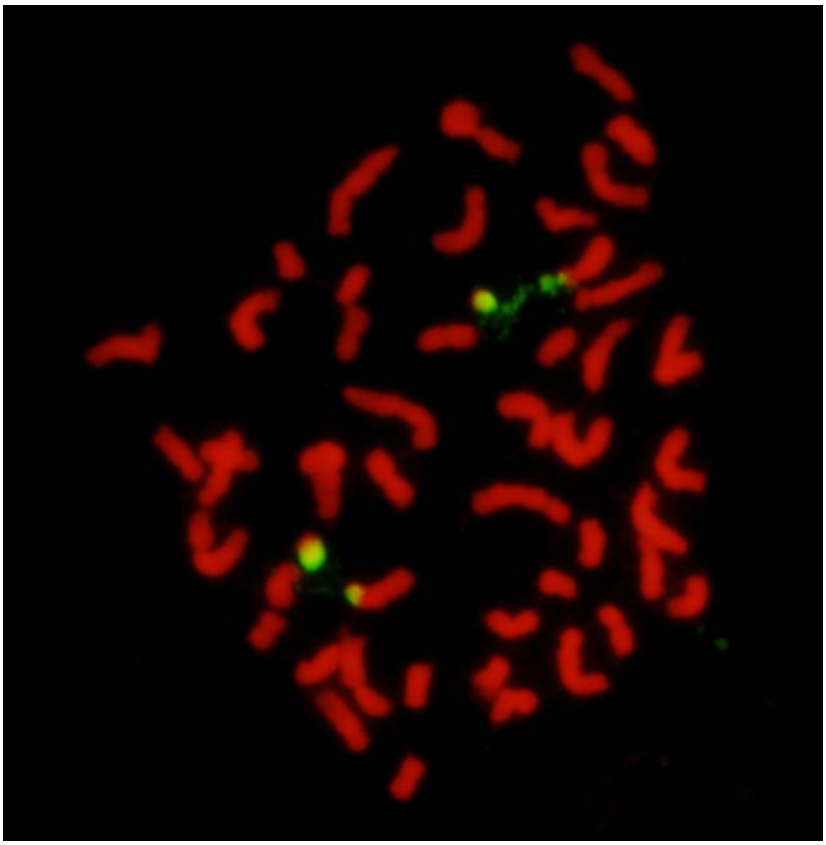
OsLOX1 gene encoding rice lipoxygenase. *Rice Genetics Newsletters* **24**, 37-40 (2008).

119. Islam, M.A., Du, H., Ning, J., Ye, H. & Xiong, L. Characterization of *Glossy1*-homologous genes in rice involved in leaf wax accumulation and drought resistance. *Plant Mol. Biol.* **70**, 443-456 (2009).

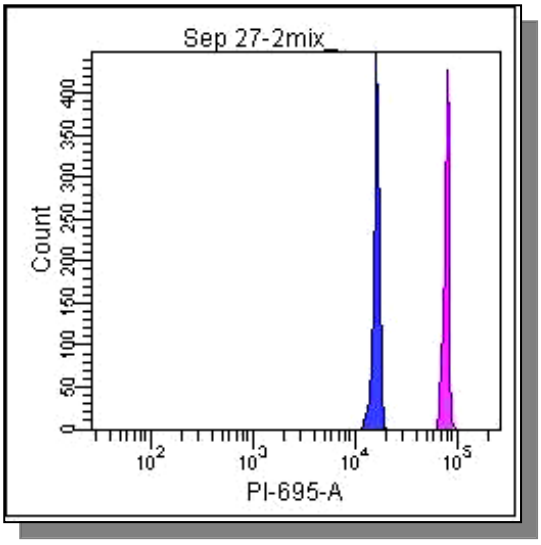
Nature Genetics: doi:10.1038/ng.2569

Supplementary Figures

a



b

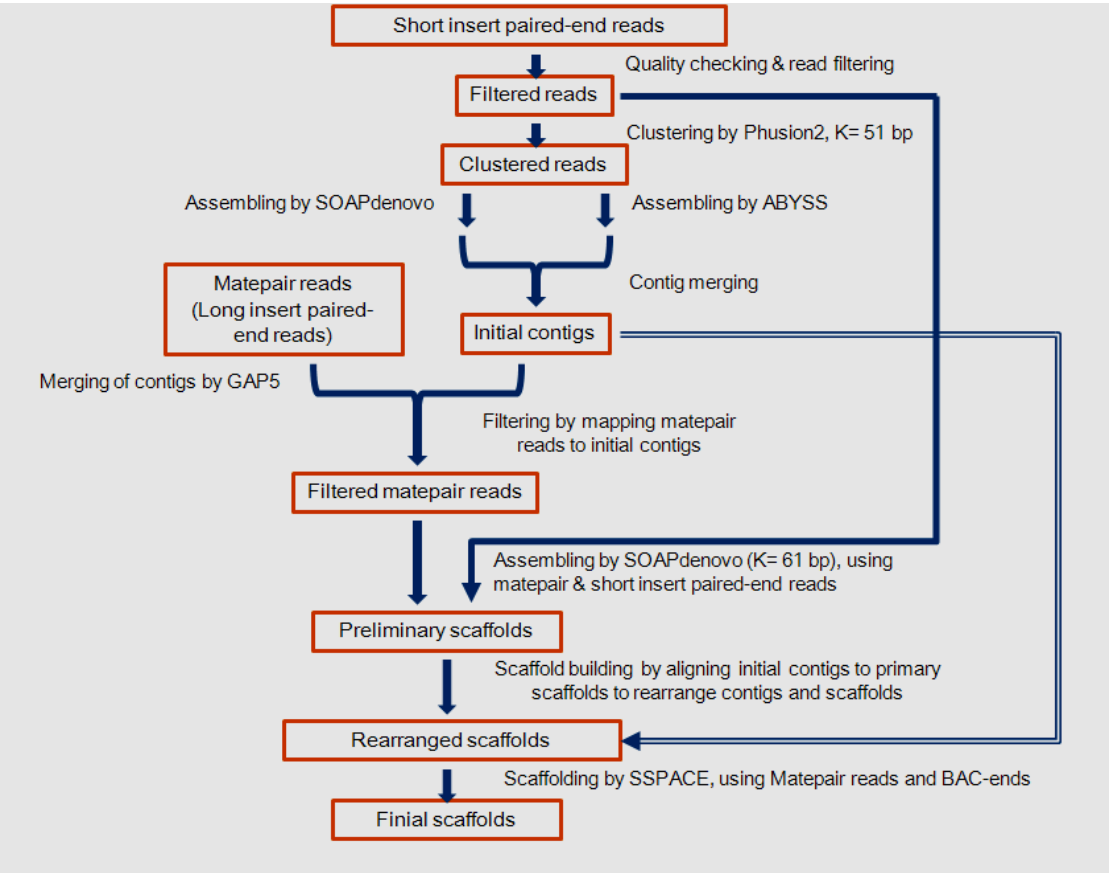


Supplementary Figure 1 Cytogenetic analysis of bamboo chromosomes. (**a**)

Nature Genetics: doi:10.1038/ng.2569

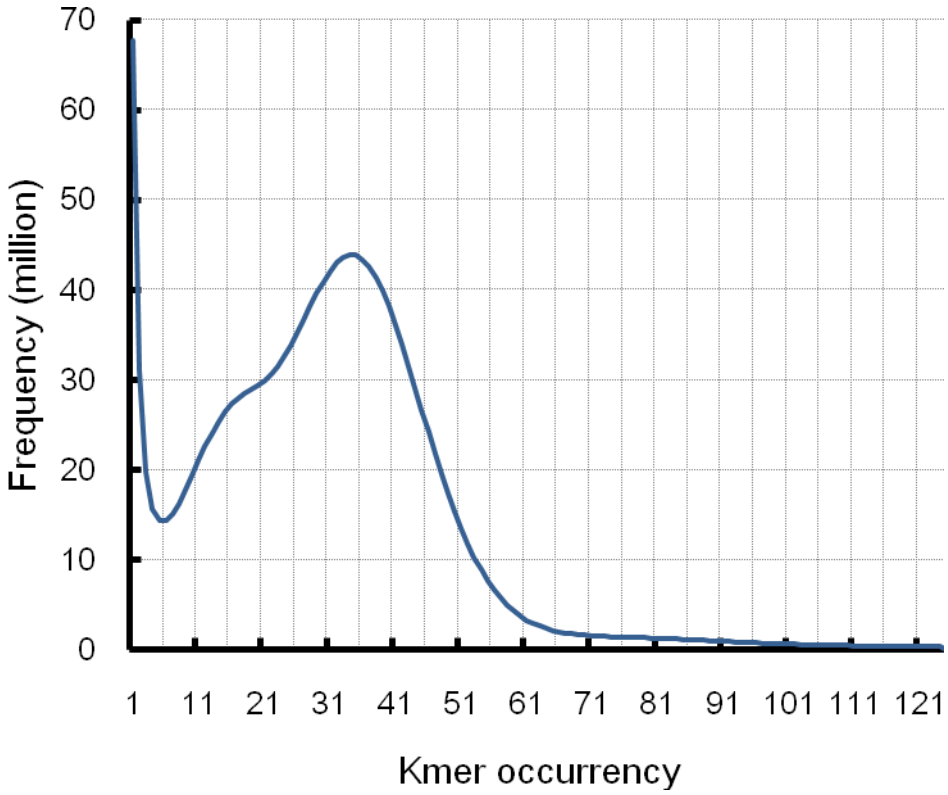
Fluorescence *in situ* hybridization of the moso bamboo at mitotic metaphase with the probe of rice 45S rDNA. The chromosomes were dyed by DAPI. The 45S rDNA probe was labeled with digoxin. Signals were detected by FITC with a magnification of microscope at 200 times. Two copies of chromosome sets were displayed. (**b**) Test result of *Oryza sativa* and *Phyllostachys heterocycla* mixed samples by using flow cytometry. The term C-value refers to the amount (picograms) of DNA contained within a haploid nucleus or one half the amount in a diploid somatic cell of a eukaryotic organism. Blue peak indicated 2 C DNA of *Oryza sativa* at 16,378. Pink peak indicated 2 C DNA of *Phyllostachys heterocycla* at 79,892. Compared with that of rice (430 Mb), genome size of the moso bamboo was estimated to be 2 075.025 ± 13.08 Mb.

Nature Genetics: doi:10.1038/ng.2569



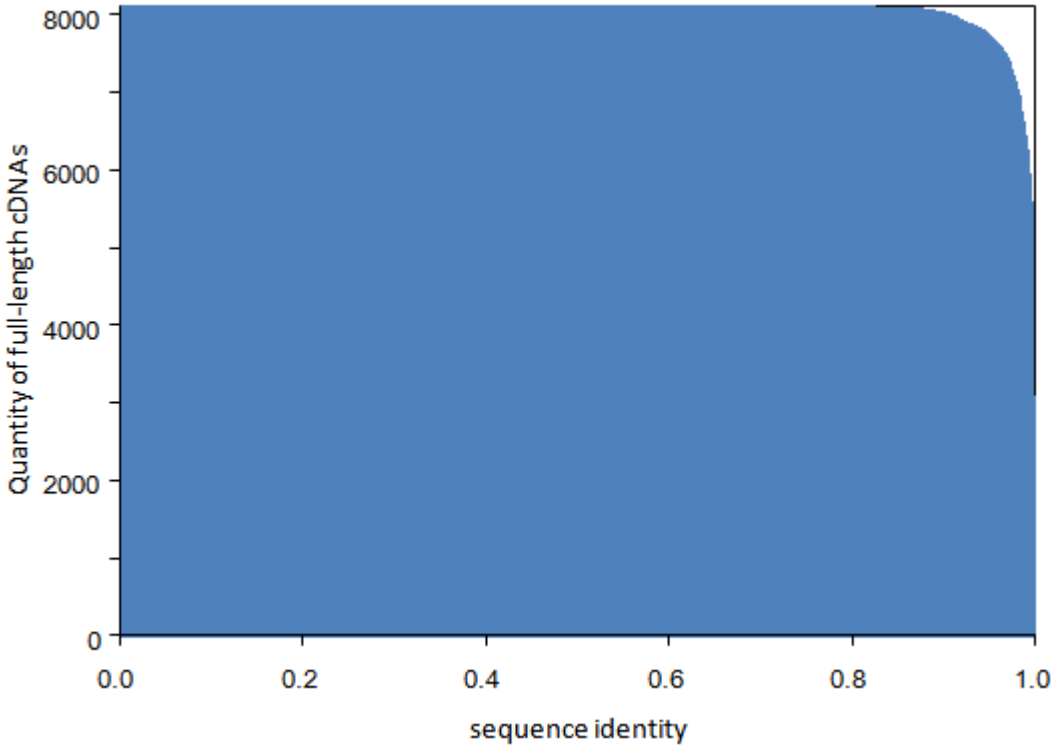
Supplementary Figure 2 Phusion-meta pipeline of short read assembly. To complete assembly of this highly heterozygous genome, we developed an integrated *de novo* assembly pipeline for large-sized genomes using short read sequencing data. Although this pipeline used existing algorithms and assembler, there were three critical variables introduced into the assembling strategy: i) filtering of the paired-end (PE) reads by K-mer occurrency to lower sequence error; ii) clustering of the reads before assembling to reduce assembling errors derived by direct assembling of the short read; iii) multiple use of reads and contigs in different cycles to make up for deficiencies of different algorithms. Thus, we generated comparatively high-quality assemblies by using nearly entire short reads. This pipeline therefore can be efficiently used in *de novo* assembly of complex and large-sized genomes.

Nature Genetics: doi:10.1038/ng.2569



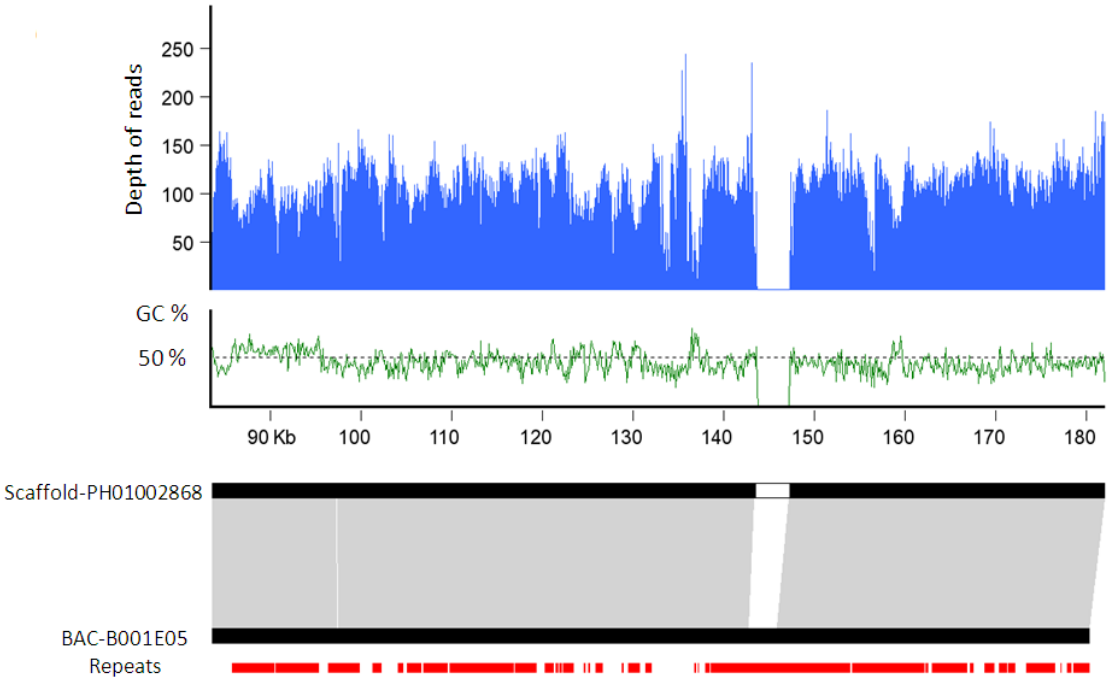
Supplementary Figure 3 Distribution of Kmer frequency. Distribution of 51-mer frequency in the reads of short insert size libraries (350 - 400 bp). Values of K-mers were plotted against the frequency (*y*-axis) at their occurrency (*x*-axis). The leftmost truncated peak at low occurrence (1-2) was mainly due to random base errors in the raw sequencing reads. The frequency exhibited a bi-modality caused by heterozygosity.

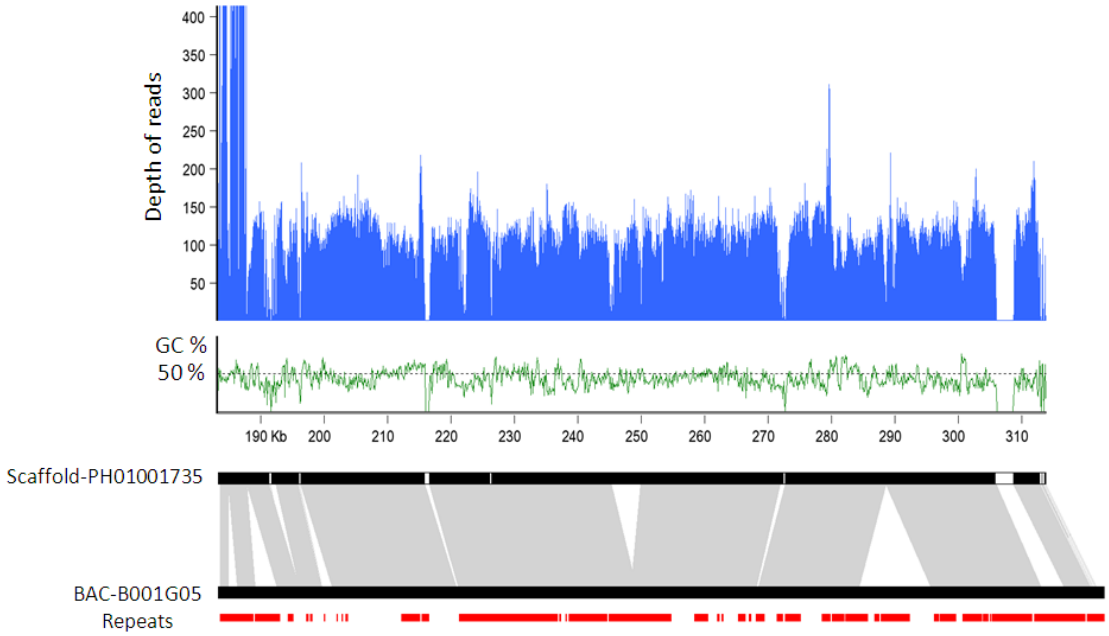
Nature Genetics: doi:10.1038/ng.2569



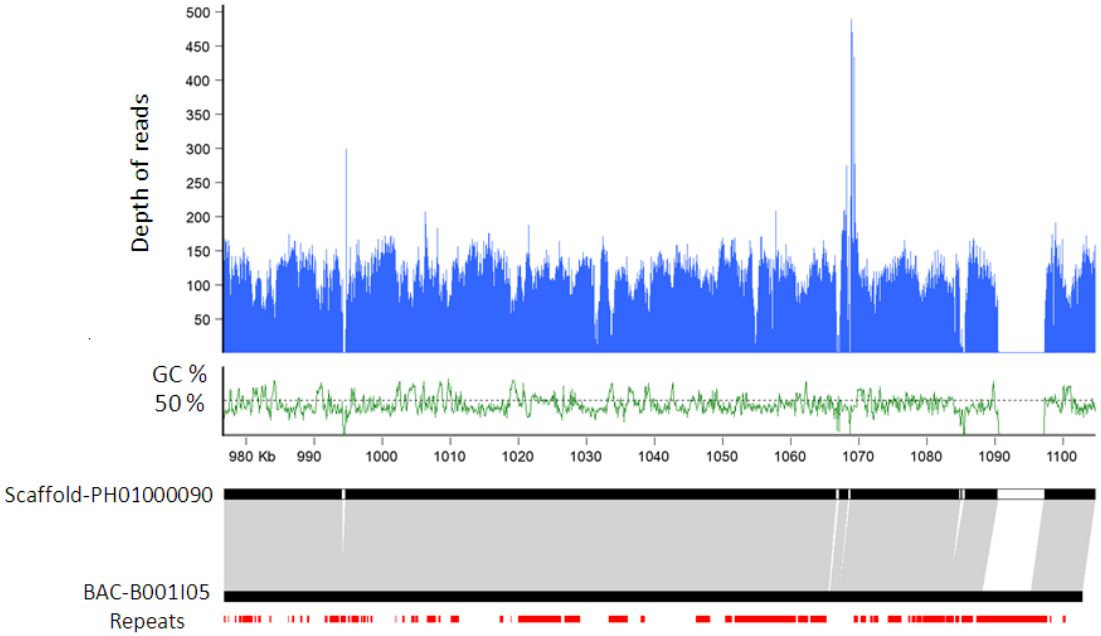
Supplementary Figure 4 Mapping of the full-length cDNA sequences to the moso bamboo genome. The cDNA sequences were aligned to the assembled scaffolds. Totally, 8,118 of 8,253 (98.4%) cDNAs were uniquely mapped onto the assemblies with high identities at 99.1% using the GMAP. The *Y* axis showed accumulative frequency of the aligned cDNAs. The *X* axis exhibited the identity of sequence alignment.

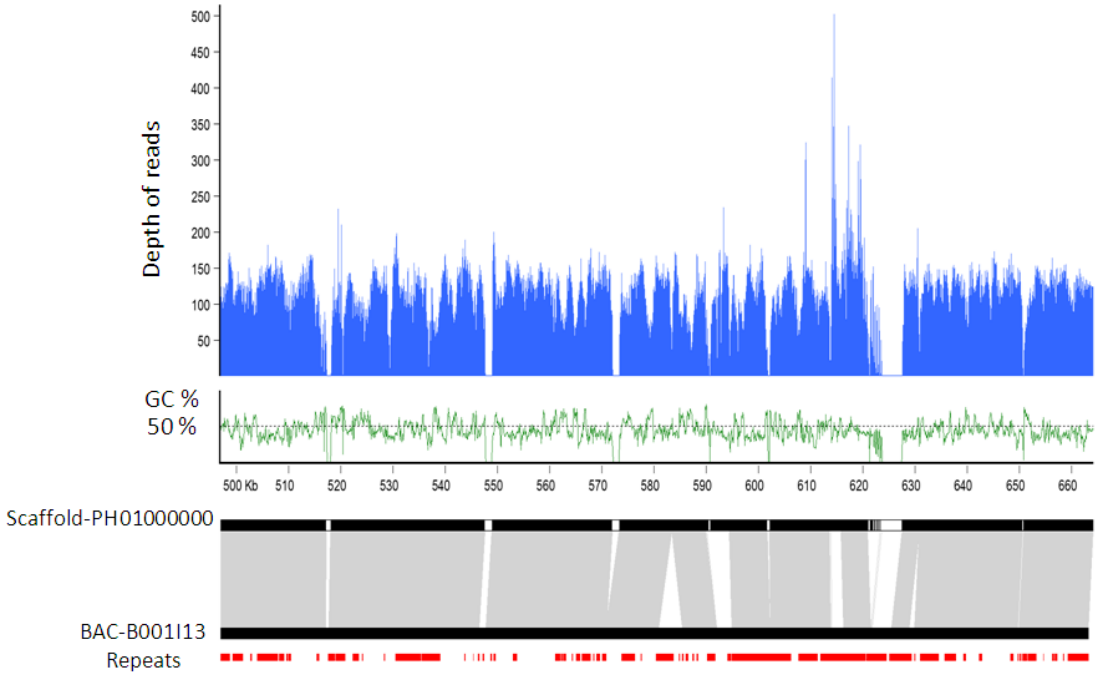
Nature Genetics: doi:10.1038/ng.2569



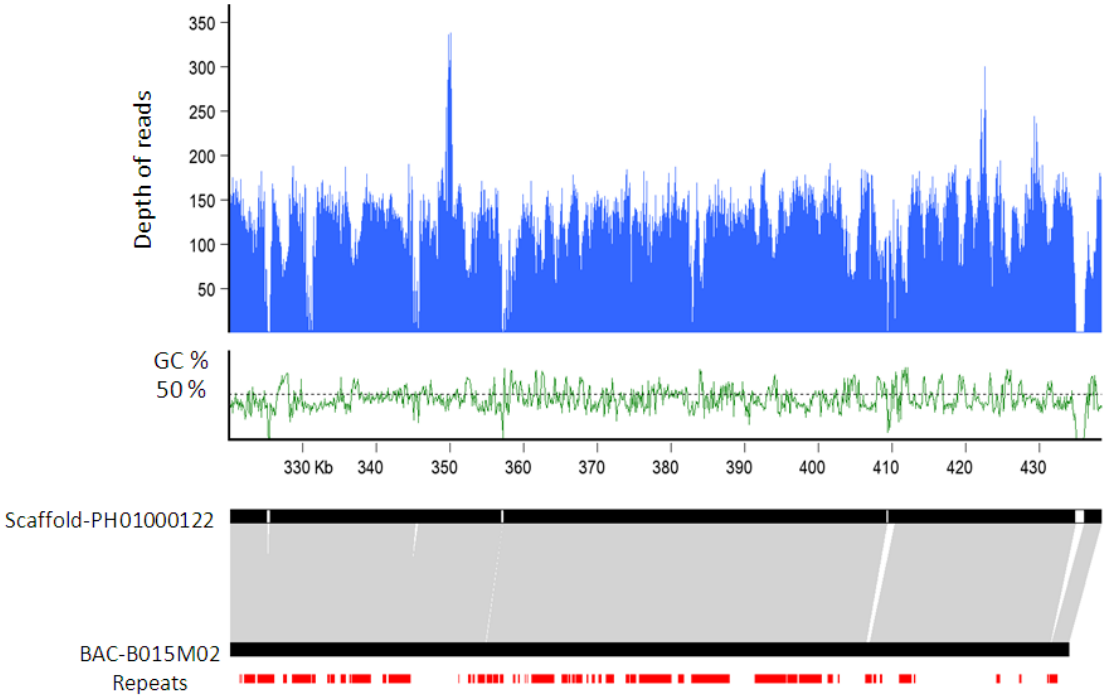


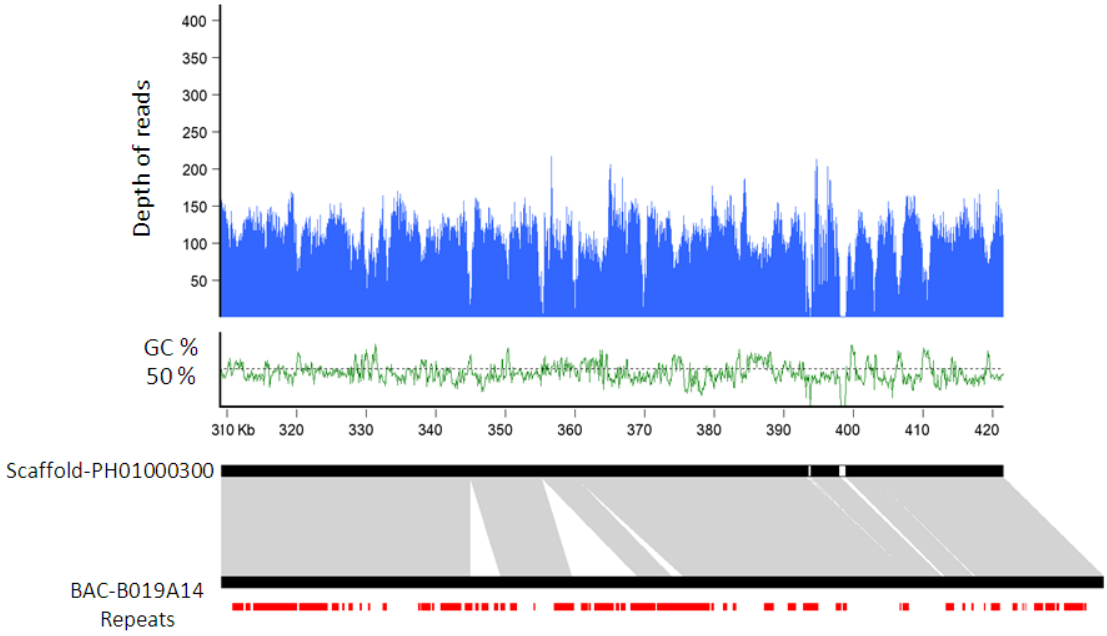
Nature Genetics: doi:10.1038/ng.2569



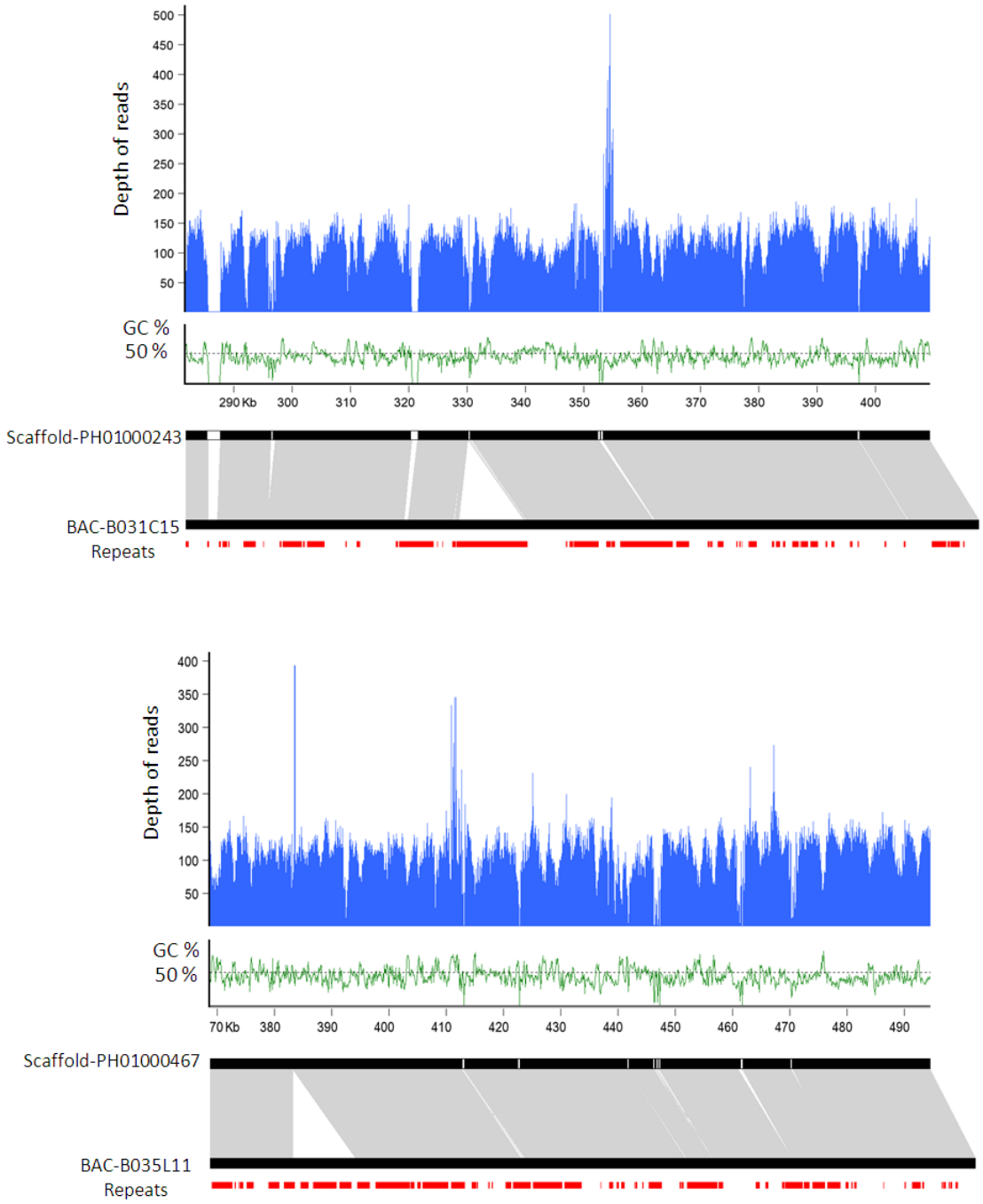


Nature Genetics: doi:10.1038/ng.2569



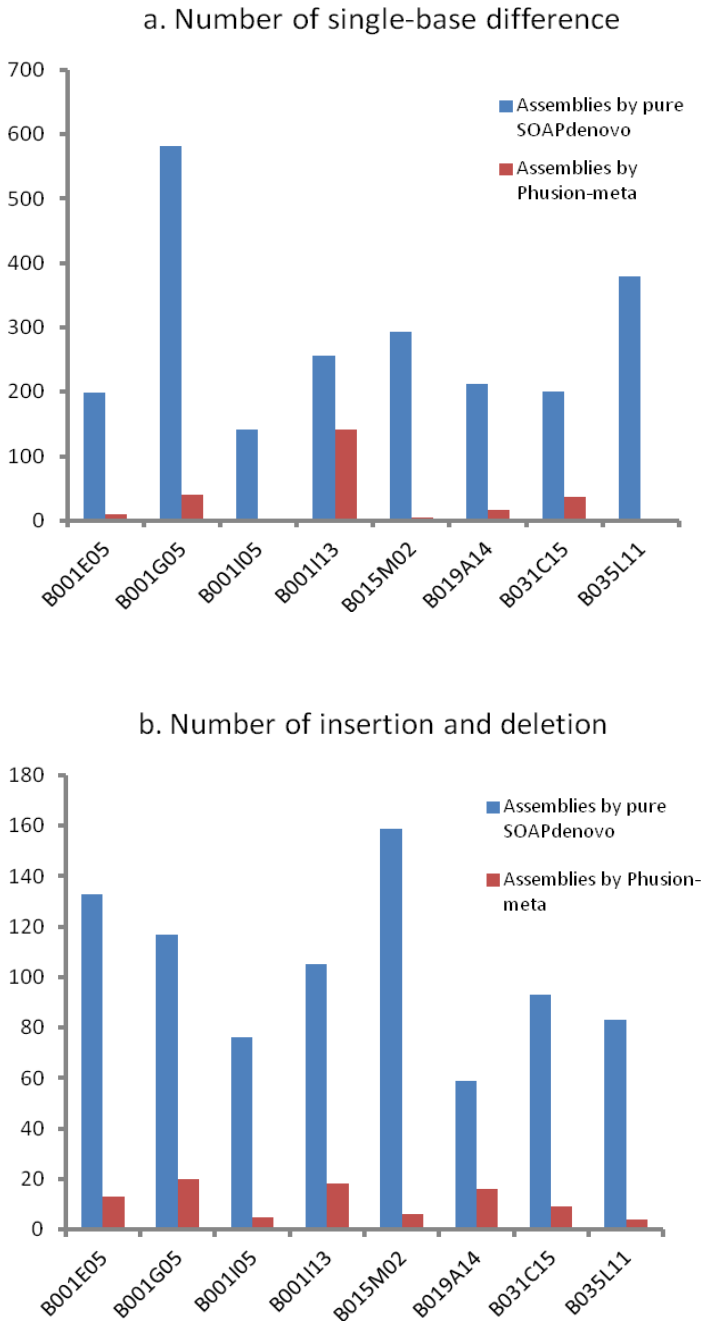


Nature Genetics: doi:10.1038/ng.2569



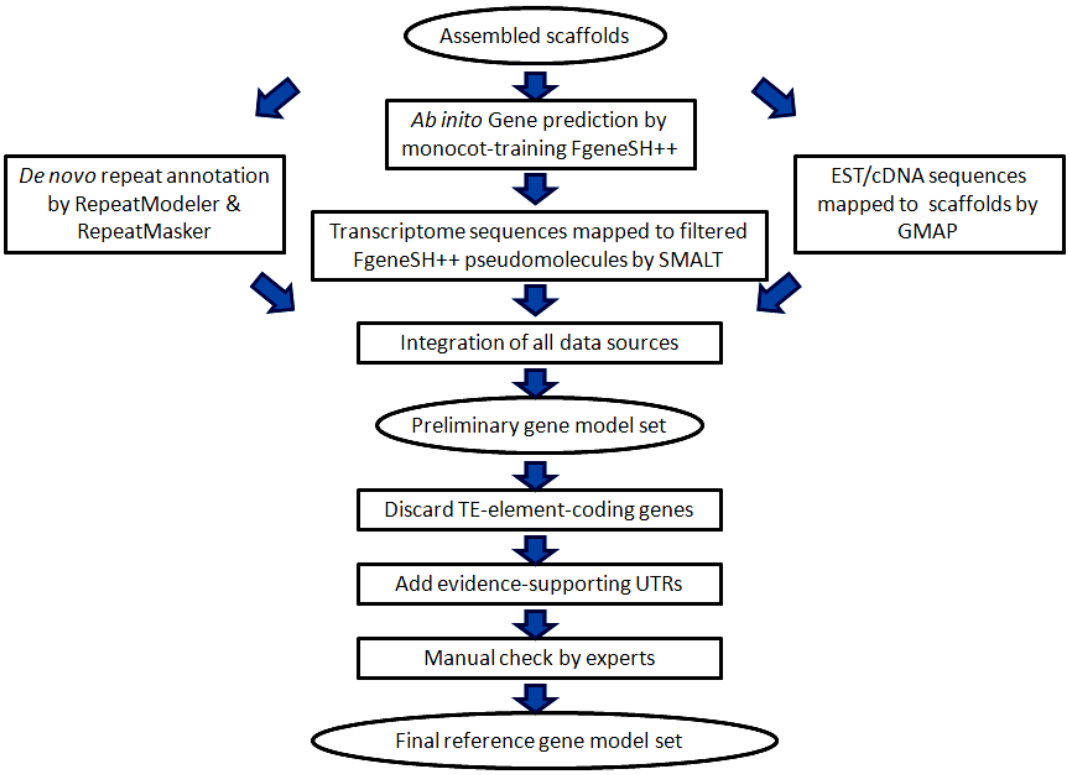
Supplementary Figure 5 Alignment of assembled scaffolds to the BACs sequenced by Sanger method**.** Depth of reads in blue was calculated by mapping PE reads onto the BAC sequences. Repeats in red showed the RepeatMasker-annotated TEs on the BAC sequences. The blocks in white indicated the unfilled gaps on the scaffolds. The grey blocks showed aligned region between Sanger BACs and scaffolds. GC contents were shown in green curves.

Nature Genetics: doi:10.1038/ng.2569



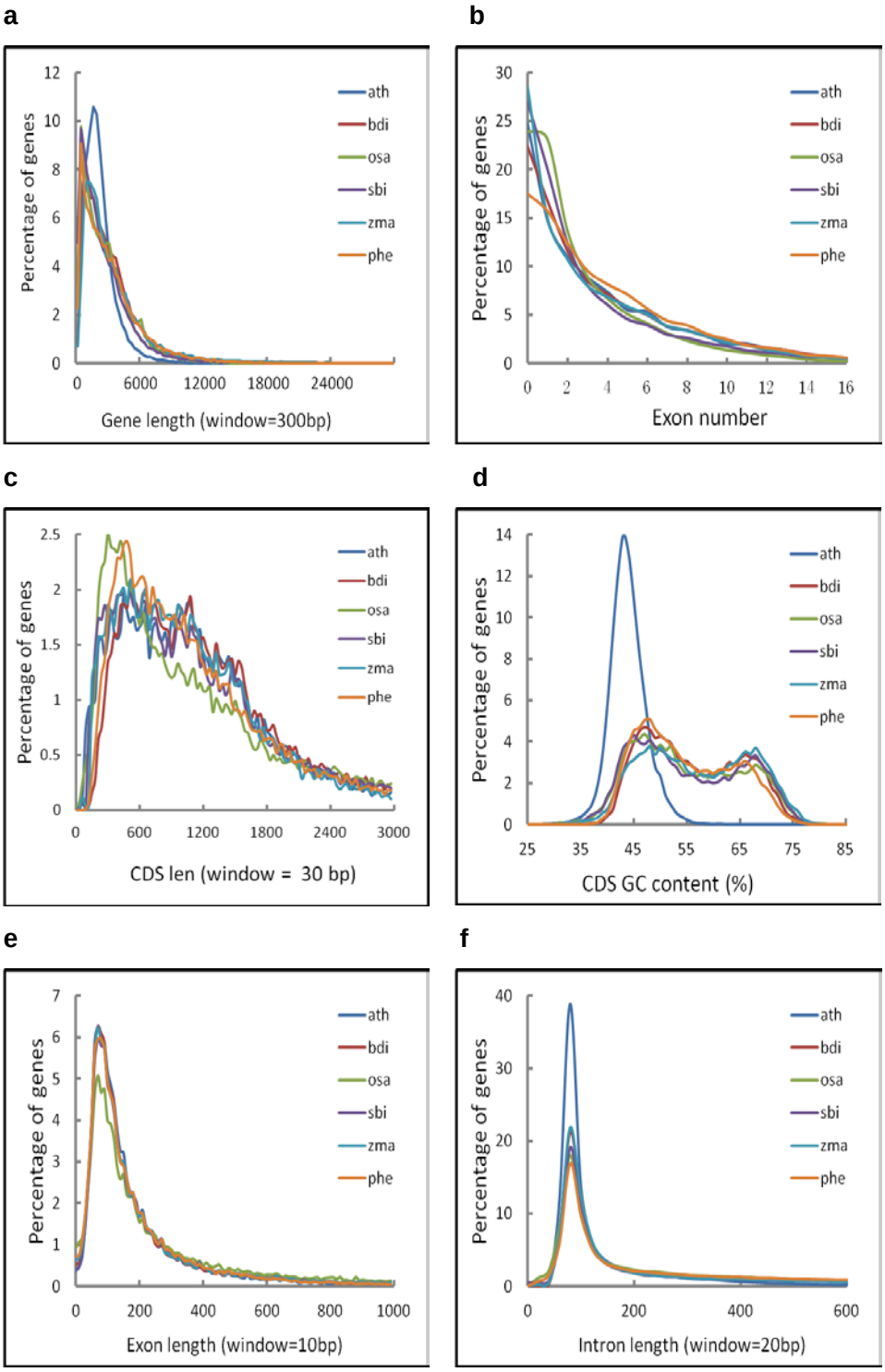
Supplementary Figure 6 Comparison of detected assembling errors by different assembling methods.

Nature Genetics: doi:10.1038/ng.2569



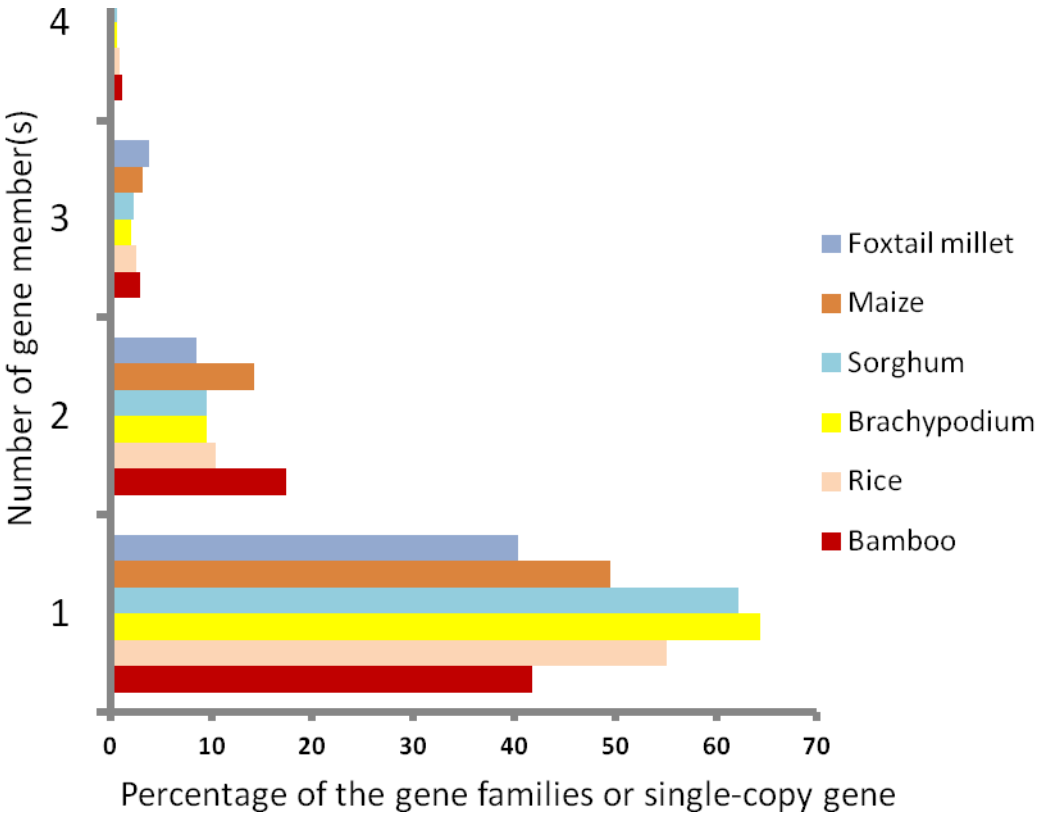
Supplementary Figure 7 Simplified pipeline of gene prediction with the combination of *ab initio* gene prediction, mapped RNA-seq reads, and cDNA sequences. Using both RNA-seq data and 8,253 cDNA sequences, we initially detected 35,378 transcribed loci in the genome. By applying the stringent criteria to gene prediction, a total of 31,987 high-confidenced genes were finally identified in the annotation, which were in the same range as those of other grass families.

Nature Genetics: doi:10.1038/ng.2569



Supplementary Figure 8 Comparison of gene parameters among fully sequenced genomes. Gene structure of the *Phyllostachys heterocycla* (phe) showed highly consistent with that of other grass speices, including *Arabidopsis thaliana* (ath), *Brachypodium distachyon* (bdi), *Oryza sativa ssp. Japonica* (osa), *Sorghum bicolor* (sbi), and *Zea mays* (zma), in distribution of gene length, exon number per gene, coding sequence length, GC content in coding region, exon and intron length.

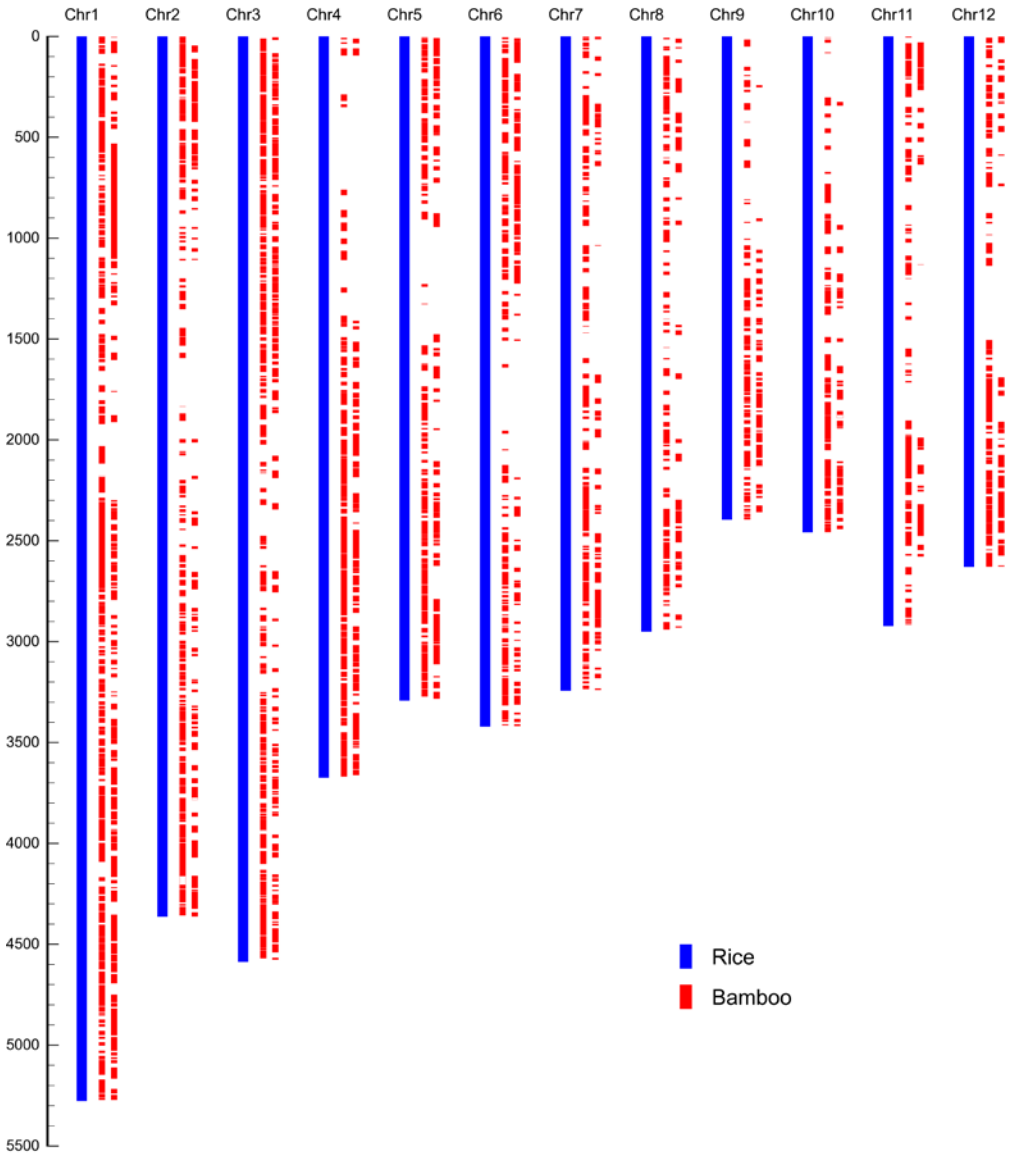
Nature Genetics: doi:10.1038/ng.2569



Supplementary Figure 9 Quantitative comparison of single-copy genes and gene families consisting of 2 to 4 members among foxtail millet, maize, sorghum, Brachypodium, rice, and bamboo. The gene families and single-copy genes were categorized by OrthoMCL analysis. For the *y* axis, number of gene member at 1 indicated the single-copy gene and 2 to 4 meant that the gene families consisted of 2 to 4 member(s). The *x* axis indicated their proportion (%).

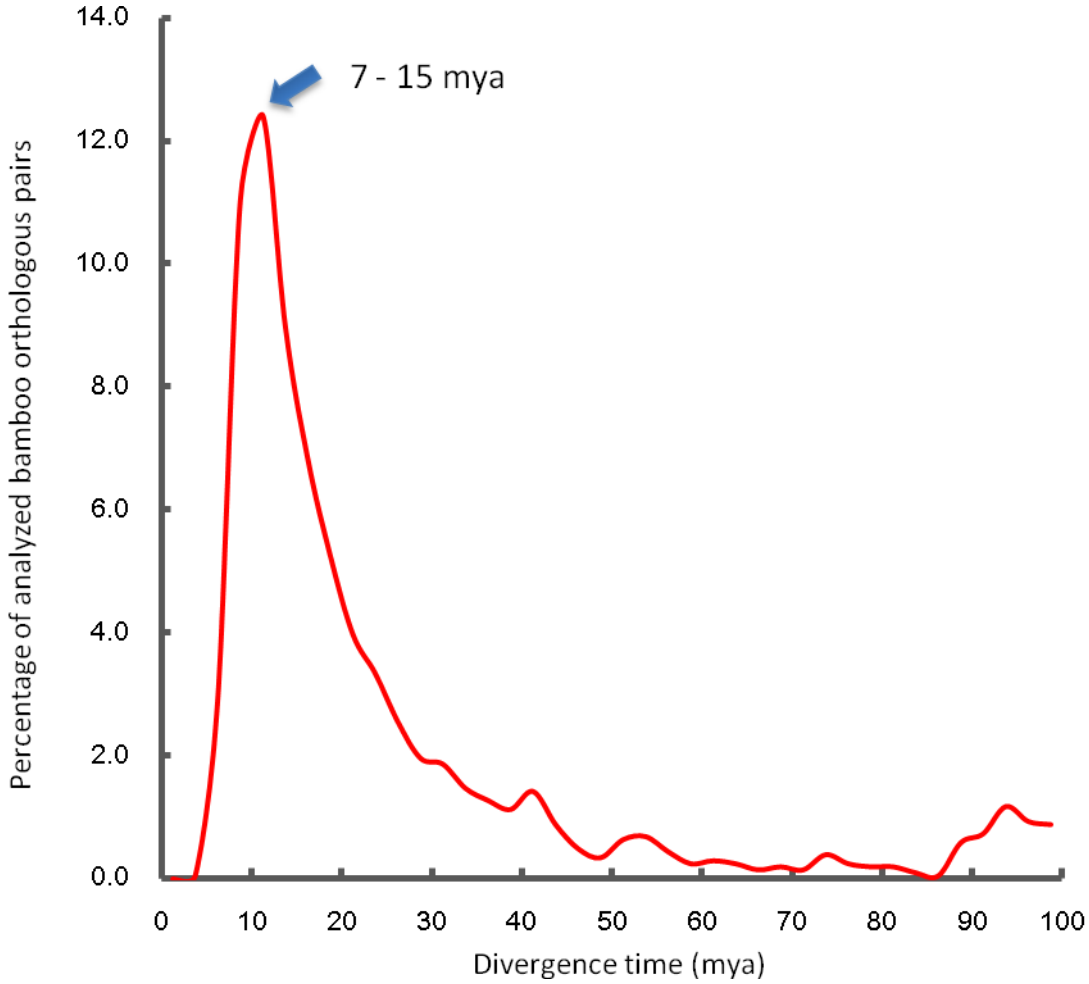
Nature Genetics: doi:10.1038/ng.2569

a



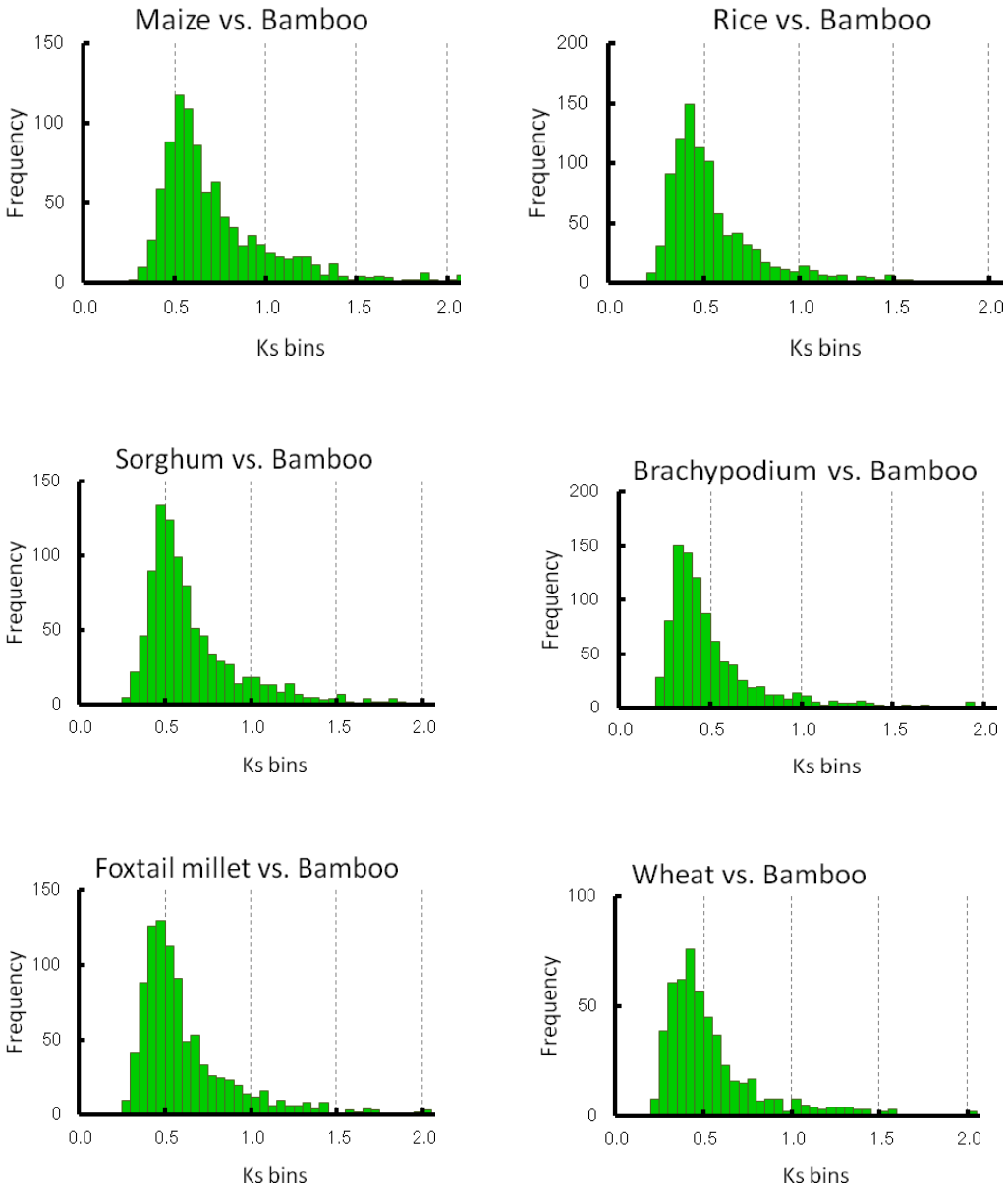
Nature Genetics: doi:10.1038/ng.2569

b



Supplementary Figure 10 The bamboo WGD identified by analysis of gene collinearity between bamboo and rice orthologs. (**a**) Collinear gene blocks between bamboo and rice genome. The rice genes are arranged according to their gene order. Rice gene sets on different chromosomes were exhibited in blue rough lines. The ordinal number of the genes were measured by the bar in the left. The collinear gene blocks of bamboo were shown in red blocks, which implicated that moso bamboo carried nearly two duplicates of rice genome. (**b**) Estimated divergence time of bamboo orthologous pairs. Only the orthologous pairs of bamboo sharing unique rice orthologs in collinear blocks were used to estimate the divergence time. The 7 to 15 mya was the potential time when two bamboo porgenitors diverged, and the WGD should occur more recent than it.

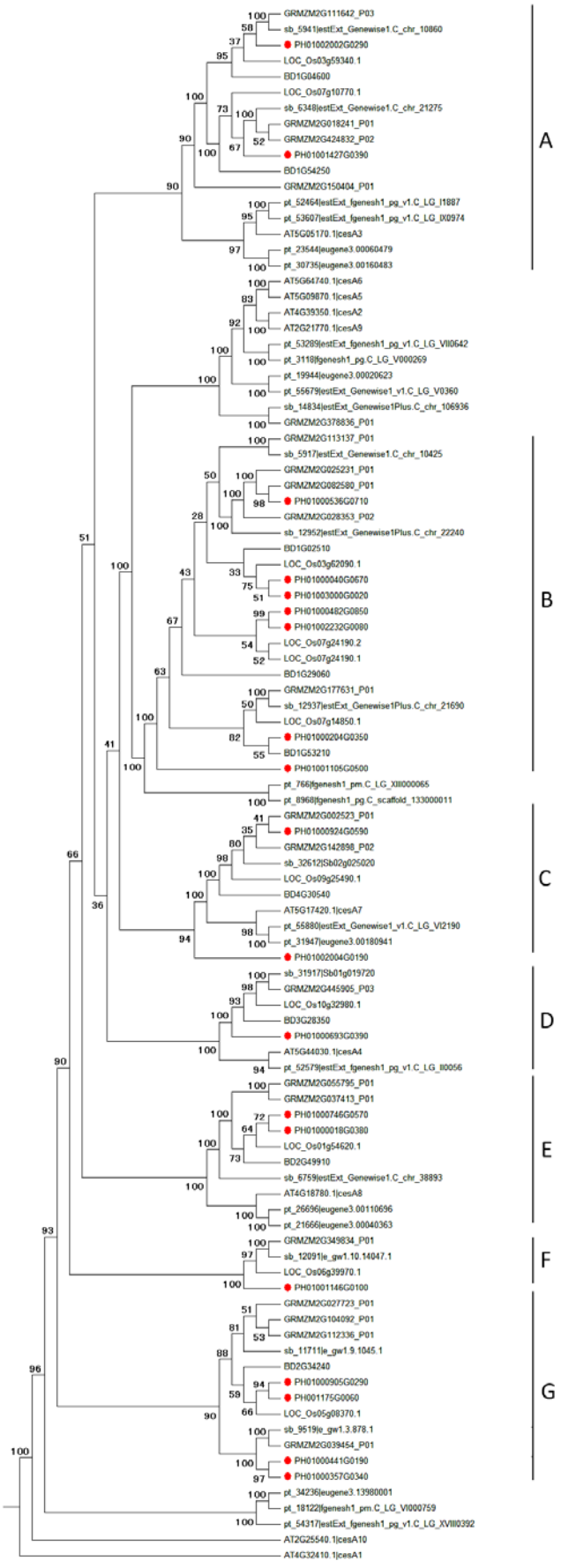
Nature Genetics: doi:10.1038/ng.2569



Supplementary Figure 11 Ks distribution of orthologous genes between bamboo and grass species. The bin size of Ks value was 0.05. Frequency was quantity of the one-to-one gene clusters.

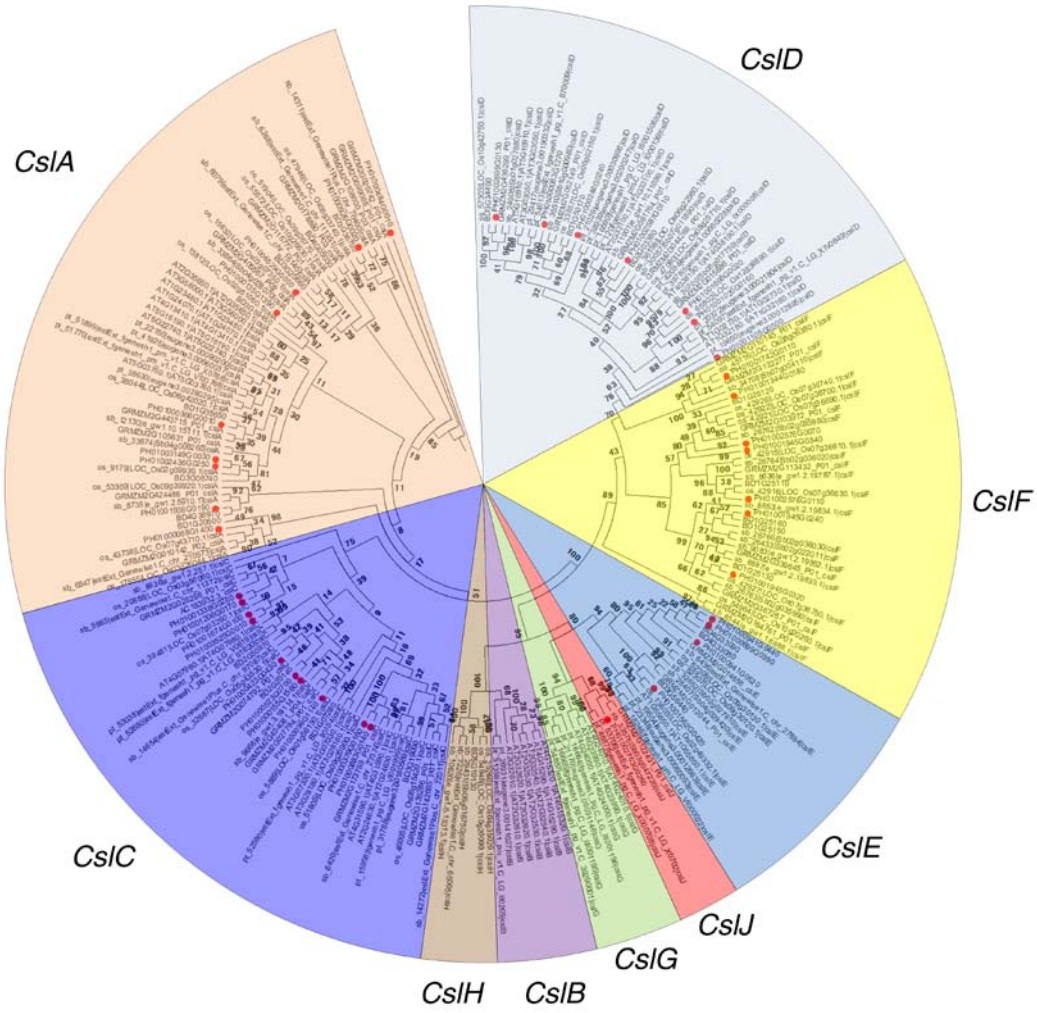
Nature Genetics: doi:10.1038/ng.2569

a



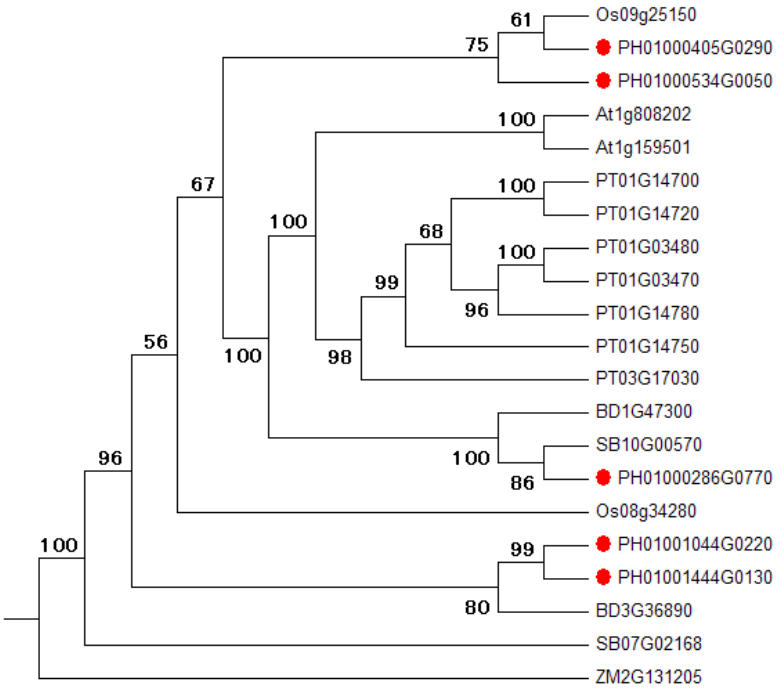
Nature Genetics: doi:10.1038/ng.2569

b

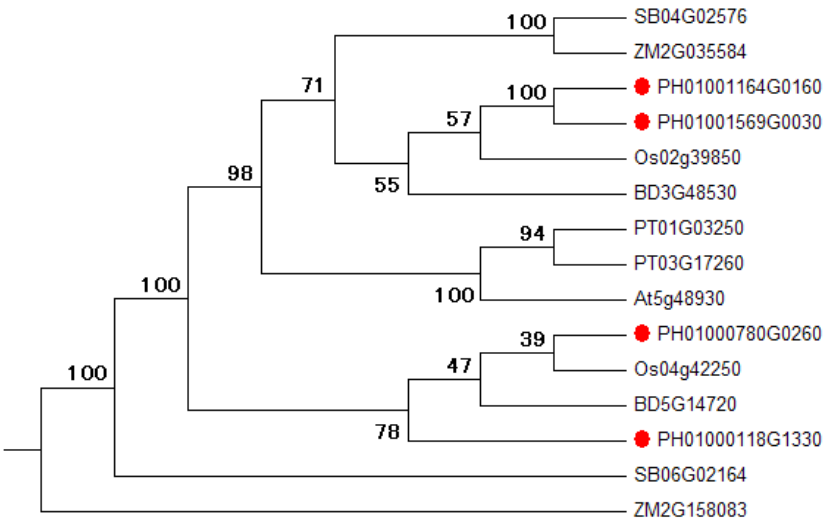


Nature Genetics: doi:10.1038/ng.2569

c

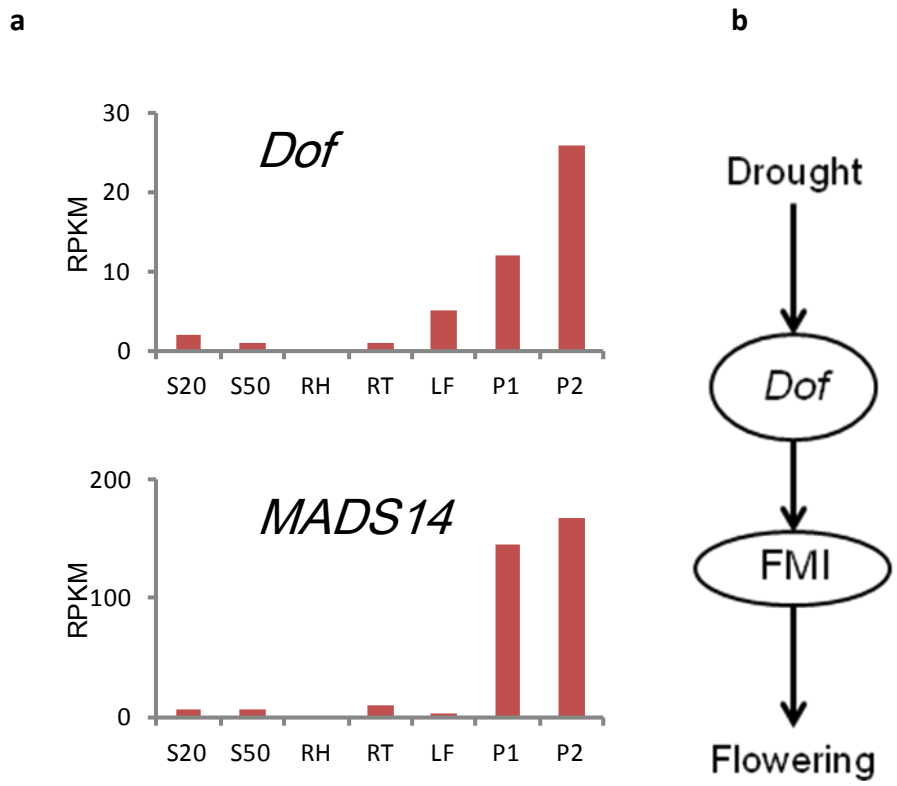


d



Supplementary Figure 12 Phylogenic tree of *CesA* and *Csl* gene families among Arabidopsis, poplar, rice, maize, sorghum, Brachypodium, and bamboo. (**a**) NJ tree of *CesAs*. A, B, C, D, E, F, and G indicated 7 clades where the bamboo genes were located. (**b**) NJ tree of *Csls*. Different subfamilies were shown in different colors. (**c**) NJ tree of CCR genes. (**d**) NJ tree of HCT genes. The bamboo genes were labeled by red point. The numbers beside the branches were bootstrap percentage.

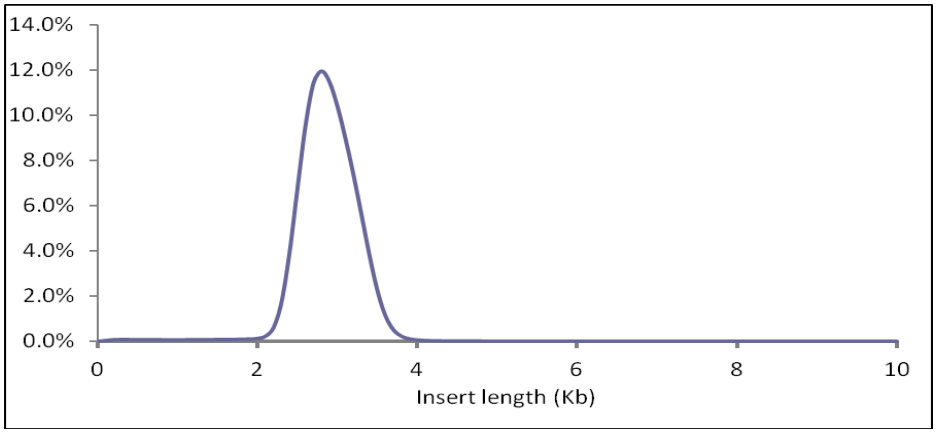
Nature Genetics: doi:10.1038/ng.2569



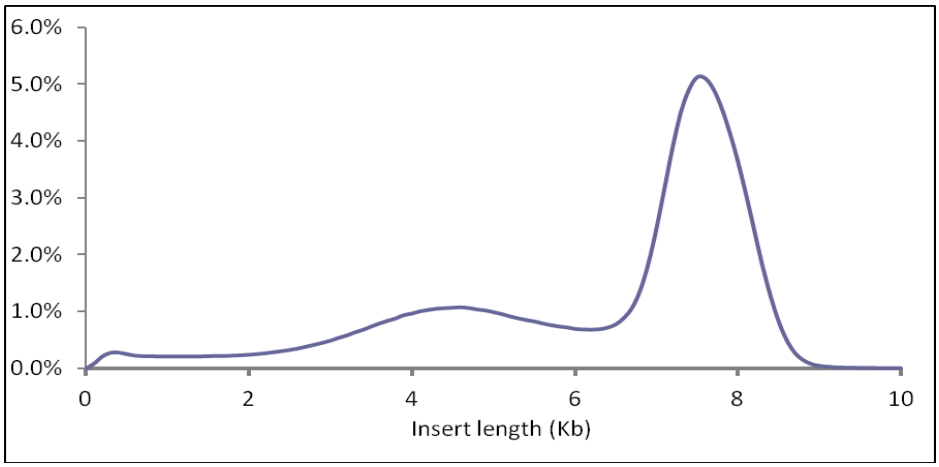
Supplementary Figure 13 A hypothesized pathway in activation of flowering. (**a**) Quantified expression levels of bamboo *Dof* (the homolog of the *OsDof12*) and *MADS14* (one of homologs of the *OsMADS14*, involved in the FMI) in different tissues. The expression is indicated as the normalized quantified transcript levels (RPKM). (**b**) A predicted pathway in controlling of flower-time in bamboo. Of the identified floral genes, the homologs of *OsMADS14* (main text ref.30) (bamboo *MADS14s*, Identity >70%, FMI genes) were highly expressed in panicles (16- to 84-fold over vegetative tissue, Q-value < 0.001). However, the expressions of the homologs of its upstream regulatory genes, such as *OsSOC1*[23], and *Ehd1*[24], were not detected (**Supplementary Table 17**) except for a homolog of *OsDof12*[25] (bamboo *Dof,* Identity >80%, **Supplementary Table 16**) with significantly higher expression in panicles (5- to 26-fold, Q-value < 0.001), implying that the bamboo *Dof* might also be functional in regulating *MADS14*s at the flowering of bamboo. As the previous studies in rice, the *OsDof12* can be induced under long-day conditions[25] or drought stress[26]. The flowering bamboo were grown in the typical short-day growing area in Southern China where a severe drought had just occurred, suggesting that the bamboo *Dof* was likely induced by drought stress. Taken together, a pathway of drought-*Dof*-*MADS14*-flowering might be active during the flowering.

Nature Genetics: doi:10.1038/ng.2569

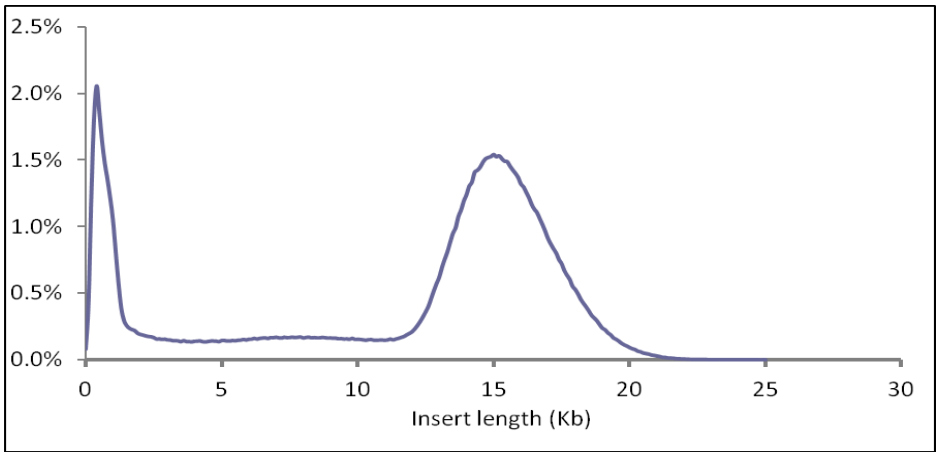
a



b

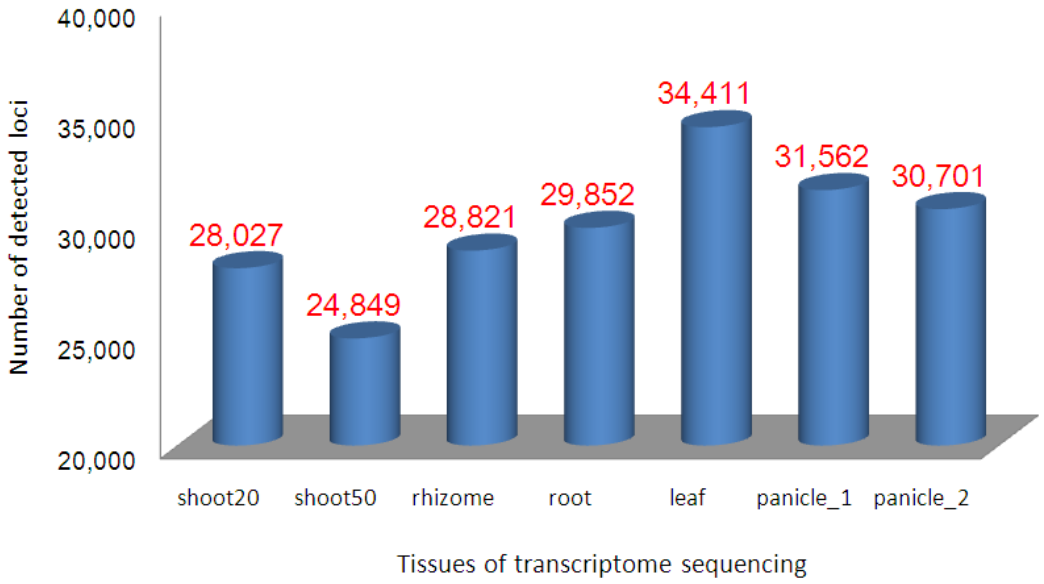


c



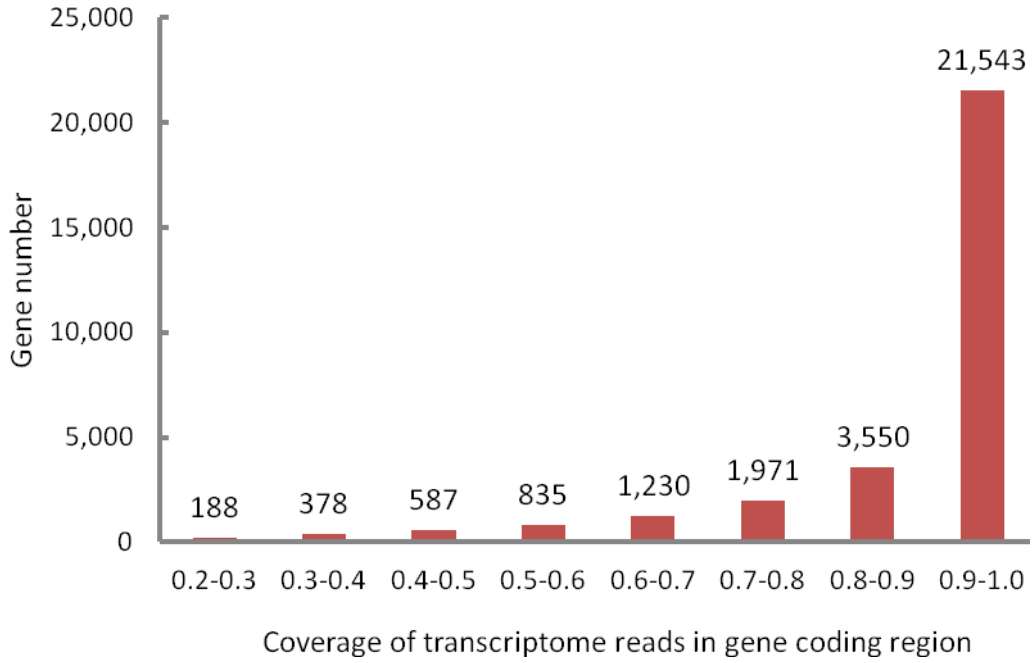
Supplementary Figure 14 Distribution of insert sizes in paired-end sequencing libraries with inserts at around 3 to 18 Kb. (**a**) Distribution of the library with insert sizes around 3Kb. (**b**) Distribution of the library with insert sizes around 7-8 Kb. (**c**) Distribution of 16-18 Kb insert library. The insert sizes and their distribution were estimated by counting the number of read pairs located on existed initial contigs.

Nature Genetics: doi:10.1038/ng.2569



Supplementary Figure 15 Number of detected non-TE-derived loci by transcriptome sequences from 7 different tissues, tip of 20cm-high shoot (shoot20), tip of 50cm-high shoot (shoot50), rhizome, root, leaf, panicle at early flowering stage (panicle_1), and panicle at late flowering stage (panicle_2). The number indicated quantity of the detected loci in corresponding tissues. Some of the detected loci were discarded in filtering of gene models, which resulted in the detected number in some tissues was more than final gene models.

Nature Genetics: doi:10.1038/ng.2569



Supplementary Figure 16 Transcriptome evidence for gene prediction in moso bamboo. A potential gene model should be supported by at least 20% coverage of transcriptome reads in gene coding region. Over 27,000 (87% of 31,987) genes' coding regions were strongly supported by transcriptome sequences.

Nature Genetics: doi:10.1038/ng.2569

Supplementary Tables

Supplementary Table 1 Summary of sequence assembly.

Supplementary Table 1a Summary of the BAC-end sequences used in scaffolding

Quantity of BAC-ends	20,654 (10,327 pairs)
Total length of BAC-ends	18.4 Mb
Genome coverage	0.66 ×
Average read length	890 bp
Average distribution density	every 103 Kb
Average BAC insert length	140 Kb
Quantity of aligned BAC-ends to the scaffolds	19,069
Total length of aligned BAC-ends to the scaffolds	17.3 MB (94.2%)
Shared identites of BAC-ends in the scaffolds	96.8%

Supplementary Table 1b Summary of sequencing and assembly of the moso bamboo genome.

	Added paired-end insert size (bp)	Sequence coverage (×)	N50 (bp)	N80 (bp)	Length of max. contig (bp)	Total length (bp)
Initial contig	350 - 400	120	11,622	2,338	186,163	1,862,588,005
Scaffold 1			13,599	2,876	220,151	1,915,326,230
Scaffold 2[†]	2,400 - 3,600 6,800 - 8,700 13,500 - 17,500 120,000 (10,327 pairs of BAC-ends)	13 12 2 0.66	328,698	62,052	4,869,017	2,051,719,643

[†] Final scaffolds with less than 500 bp were excluded.

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 1c Comparison of length of contigs and scaffolds assembled by different methods.

			Assemblies by pure SOAPdenovo	Assemblies by Phusion-meta
Initial contigs	N50	Length (bp)	5,076	11,882
		n	89,573	42,412
	N80	Length (bp)	472	3,582
		n	412,943	114,186
	Average length (bp)		598	2,088
	Maximum length (bp)		99,128	152,833
	Total length (bp)		1,837,124,999	1,871,331,085
Scaffolds	N50	Length (bp)	81,579	328,698
		n	6,285	1,626
	N60	Length (bp)	55,746	234,025
		n	9,120	2,362
	N70	Length (bp)	31,874	149,820
		n	13,550	3,450
	N80	Length (bp)	13,996	62,052
		n	22,289	5,499
	N90	Length (bp)	5,368	1,733
		n	43,146	44,100
	N100	Length (bp)	500	500
		n	172,645	277,278
	Average length (bp)		11,215	7,400
	Maximum length (bp)		767,478	4,869,017
	Total length (bp)		1,936,298,937	2,015,719,643
	Genome coverage (%)		92.2	97.7

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 2 Comparison of assembled scaffolds and genomic sequences in Genbank. The known genome sequences with the length at 2 – 40 Kb were downloaded from GenBank (Accession NO. GQ252841 - 252869). Of the downloaded sequences, gi|284434746|gb|GQ252834.1| was probably from chloroplastic DNA. Blast with E-value at less than 1e-05 was used in alignment.

GeneBank Accession No.	Length (bp)	Coverage of all matches	Coverage of single best match	GeneBank Accession No.	Length (bp)	Coverage of all matches	Coverage of single best match
gi|284434480|gb|GQ252796.1|	6,939	0.96	0.85	gi|284434609|gb|GQ252853.1|	3,345	0.99	0.99
gi|284434482|gb|GQ252797.1|	31,231	0.99	0.99	gi|284434611|gb|GQ252854.1|	8,050	0.99	0.99
gi|284434487|gb|GQ252799.1|	6,445	0.98	0.94	gi|284434613|gb|GQ252855.1|	4,129	0.90	0.90
gi|284434489|gb|GQ252800.1|	7,811	0.99	0.99	gi|284434615|gb|GQ252856.1|	2,858	0.99	0.99
gi|284434492|gb|GQ252801.1|	6,410	0.99	0.99	gi|284434617|gb|GQ252857.1|	5,290	0.99	0.99
gi|284434494|gb|GQ252802.1|	4,037	0.99	0.99	gi|284434619|gb|GQ252858.1|	26,484	0.99	0.99
gi|284434496|gb|GQ252803.1|	23,256	0.99	0.98	gi|284434622|gb|GQ252859.1|	22,895	0.98	0.95
gi|284434500|gb|GQ252805.1|	24,578	0.99	0.98	gi|284434629|gb|GQ252860.1|	7,028	0.99	0.99
gi|284434507|gb|GQ252806.1|	3,794	0.99	0.99	gi|284434631|gb|GQ252862.1|	15,053	0.99	0.98
gi|284434509|gb|GQ252807.1|	13,227	0.99	0.99	gi|284434635|gb|GQ252863.1|	8,067	0.98	0.98
gi|284434511|gb|GQ252809.1|	6,059	0.99	0.99	gi|284434637|gb|GQ252864.1|	11,009	0.99	0.96
gi|284434513|gb|GQ252810.1|	6,189	0.98	0.88	gi|284434642|gb|GQ252865.1|	8,200	0.96	0.83
gi|284434516|gb|GQ252812.1|	9,531	0.99	0.99	gi|284434644|gb|GQ252866.1|	5,680	0.99	0.99
gi|284434519|gb|GQ252813.1|	5,709	0.99	0.99	gi|284434647|gb|GQ252867.1|	5,306	0.99	0.99
gi|284434522|gb|GQ252814.1|	9,927	0.99	0.99	gi|284434649|gb|GQ252868.1|	25,304	0.96	0.72
gi|284434524|gb|GQ252815.1|	8,244	0.99	0.99	gi|284434653|gb|GQ252869.1|	42,561	0.97	0.50
gi|284434527|gb|GQ252816.1|	18,487	0.99	0.99	gi|284434660|gb|GQ252870.1|	11,864	0.99	0.90
gi|284434533|gb|GQ252818.1|	16,728	0.99	0.95	gi|284434663|gb|GQ252871.1|	9,881	0.96	0.55
gi|284434536|gb|GQ252819.1|	13,505	0.99	0.99	gi|284434666|gb|GQ252872.1|	11,757	0.99	0.99
gi|284434539|gb|GQ252821.1|	3,102	0.94	0.79	gi|284434668|gb|GQ252873.1|	8,127	0.98	0.97
gi|284434541|gb|GQ252822.1|	10,294	0.99	0.99	gi|284434671|gb|GQ252874.1|	20,132	0.96	0.68
gi|284434543|gb|GQ252823.1|	14,582	0.99	0.97	gi|284434676|gb|GQ252875.1|	18,539	0.97	0.53
gi|284434547|gb|GQ252824.1|	7,262	0.98	0.93	gi|284434678|gb|GQ252876.1|	10,060	0.99	0.99
gi|284434549|gb|GQ252825.1|	11,647	0.99	0.99	gi|284434681|gb|GQ252877.1|	7,687	0.99	0.99
gi|284434551|gb|GQ252826.1|	21,041	0.99	0.99	gi|284434684|gb|GQ252879.1|	13,424	0.99	0.86
gi|284434557|gb|GQ252827.1|	21,656	0.99	0.84	gi|284434687|gb|GQ252881.1|	31,459	0.99	0.99
gi|284434561|gb|GQ252828.1|	3,905	0.99	0.99	gi|284434694|gb|GQ252882.1|	5,403	0.99	0.98
gi|284434563|gb|GQ252829.1|	5,314	0.93	0.84	gi|284434697|gb|GQ252884.1|	5,338	0.99	0.97
gi|284434565|gb|GQ252830.1|	6,685	0.99	0.99	gi|284434699|gb|GQ252885.1|	3,379	0.89	0.82
gi|284434568|gb|GQ252831.1|	5,688	0.99	0.99	gi|284434740|gb|GQ252798.1|	3,880	0.98	0.90
gi|284434570|gb|GQ252832.1|	10,633	0.99	0.99	gi|284434741|gb|GQ252804.1|	12,042	0.99	0.99
gi|284434572|gb|GQ252833.1|	10,037	0.99	0.94	gi|284434742|gb|GQ252808.1|	4,347	0.98	0.95
gi|284434574|gb|GQ252836.1|	17,988	0.99	0.98	gi|284434743|gb|GQ252811.1|	4,920	0.99	0.99

Nature Genetics: doi:10.1038/ng.2569

gi|284434580|gb|GQ252837.1|	3,610	0.99	0.99	gi|284434744|gb|GQ252817.1|	4,026	0.99	0.99
gi|284434582|gb|GQ252838.1|	3,803	0.99	0.97	gi|284434745|gb|GQ252820.1|	3,125	0.99	0.80
gi|284434584|gb|GQ252839.1|	6,741	0.99	0.99	gi|284434746|gb|GQ252834.1|	3,233	0.33	0.23
gi|284434586|gb|GQ252840.1|	9,734	0.94	0.92	gi|284434747|gb|GQ252835.1|	3,415	0.99	0.99
gi|284434589|gb|GQ252841.1|	2,050	0.99	0.99	gi|284434748|gb|GQ252844.1|	4,285	0.99	0.97
gi|284434591|gb|GQ252842.1|	7,689	0.99	0.99	gi|284434749|gb|GQ252846.1|	3,882	0.99	0.98
gi|284434594|gb|GQ252843.1|	7,973	0.99	0.99	gi|284434750|gb|GQ252849.1|	2,512	0.99	0.99
gi|284434596|gb|GQ252845.1|	2,750	0.99	0.86	gi|284434751|gb|GQ252851.1|	6,882	0.99	0.99
gi|284434598|gb|GQ252847.1|	19,143	0.98	0.98	gi|284434752|gb|GQ252861.1|	5,083	0.99	0.99
gi|284434602|gb|GQ252848.1|	3,363	0.99	0.86	gi|284434753|gb|GQ252878.1|	3,917	0.98	0.97
gi|284434604|gb|GQ252850.1|	14,653	0.99	0.99	gi|284434754|gb|GQ252880.1|	2,894	0.99	0.99
gi|284434607|gb|GQ252852.1|	6,356	0.99	0.99	gi|284434755|gb|GQ252883.1|	3,085	0.99	0.99
Sum	**429,837**	**0.98**	**0.91**				

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 3 Comparison of assembled scaffolds and 18 available genes' mRNA/coding sequences in the public database. The mRNA and coding sequences were downloaded from GenBank. Poly(A) sequences at the end of mRNA were removed anterior to the alignment. Alignment indicated that most of the unmatched bases were located at the end of the GenBank genes, which were probably introduced by DNA amplification using homologous genes' primers of different species.

GenBank Accession No.	Length (bp)	Coverage of all matches	Coverage of single best match	Identities in aligned region	Description
gi|145845825|gb|EF549577.1|	1,035	0.960	0.884	0.960	cinnamyl alcohol dehydrogenase
gi|145845829|gb|EF549579.1|	801	0.983	0.983	0.997	caffeoyl-CoA O-methyltransferase
gi|162568699|gb|EU295482.1|	1,130	0.983	0.947	0.998	DRE-binding protein DREB2 (DREB2)
gi|169743367|gb|EU366146.1|	1,060	0.983	0.983	0.993	chloroplast chlorophyll a/b binding protein
gi|169743369|gb|EU366147.1|	1,141	0.976	0.976	0.990	chloroplast chlorophyll a/b binding protein
gi|175050407|gb|EF549578.2|	2,302	0.984	0.984	0.894	phenylalanine ammonia-lyase
gi|190694830|gb|EU780143.1|	554	0.978	0.978	0.994	chloroplast chlorophyll a/b binding protein
gi|195546525|gb|EU860441.1|	1,224	0.984	0.984	0.998	DRE-binding protein DREB1 (DREB1)
gi|222154090|gb|FJ594467.1|	2,171	0.985	0.985	0.949	phenylalanine ammonia-lyase (PAL1)
gi|222154092|gb|FJ594468.1|	2,294	0.985	0.985	0.942	phenylalanine ammonia-lyase (PAL1)
gi|237506882|gb|FJ495287.1|	3,293	0.985	0.985	0.998	cellulose synthase (cesA1)
gi|251766020|gb|FJ475350.1|	3,262	0.985	0.985	1.000	cellulose synthase (CesA2)
gi|251766022|gb|FJ475351.1|	3,306	0.978	0.978	0.986	cellulose synthase (CesA4)
gi|255764546|gb|FJ600727.1|	824	0.983	0.983	0.999	PsbS protein
gi|294818264|gb|GU944762.1|	590	0.981	0.981	0.997	putative pathogenesis protein (WRKY10)
gi|301071262|gb|GU434145.1|	1,245	0.979	0.979	0.997	Actin
gi|312232178|gb|HM747940.1|	1,770	0.985	0.985	0.980	MYB protein
gi|312232180|gb|HM747941.1|	739	0.943	0.943	0.978	VAH protein
Sum	28,741	0.981	0.976	0.981	

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 4 Comparison of the assembled scaffolds 8 independently sequenced Sanger BACs. The assemblies were aligned to the BACs using an aligner MUMmer[27] and Blastn with 98% or more identity. The difference of single-base and insertion/deletion in each aligned block was counted by manual checked. Content of TE-element were estimate by running RepeatMasker against constructed bamboo repetitive sequence library.

BAC ID	TE %	Length of BAC (bp)	Coverage by initial contigs (%)	Coverage by scaffolds (%)	# of single-base difference	# of insertion and deletion	Average read depth on scaffold	Rate of single-base difference (per Kb)
B001E05	69.30	96,839	95.72	100.00	9	13	106	0.09
B001G05	74.11	167,736	75.35	92.43	41	20	132	0.29
B001I05	46.26	126,170	93.64	99.99	2	5	113	0.02
B001I13	49.65	166,343	88.97	100.00	142	18	114	0.85
B015M02	41.36	113,962	99.20	100.00	5	6	127	0.04
B019A14	45.04	126,859	85.49	100.00	16	16	112	0.13
B031C15	43.67	136,024	88.13	100.00	37	9	117	0.27
B035L11	49.83	133,794	90.71	100.00	3	4	105	0.02
Average	52.70	133,466	88.78	98.81	31.9	11.4	116	0.19

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 5 Comparison of detected assembling errors by different methods.

	Assemblies by pure SOAPdenovo	Assemblies by Phusion-meta
Rate of single-base difference (# per Kb)	2.27	0.19
Rate of insertion and deletion (# per Kb)	0.83	0.09
Coverage by initial contigs	0.80	0.89
Coverage by supercontigs	0.93	0.99

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 6 Statistics of heterozygous polymorphisms. The potential sites of SNPs and short indels were detected by unique read coverage in genic and intergenic regions.

Source	Quantity of heterozygous loci			Size of analyzed sequences (bp)	heterozygous rate ($\times 10^{-3}$)
	Short indel	SNP	Indel + SNP		
gene region	6,818	92,068	98,886	115,641,313	0.85
Intergenic region	44,405	1,917,419	1,961,824	1,936,078,330	1.01
Total	51,223	2,009,487	2,060,710	2,051,719,643	1.00

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 7 Overview of gene prediction in some fully sequenced higher plants.

Organism	Sorghum	Maize	Rice	Brachypodium	Poplar	Arabidopsis	Foxtail millet	Bamboo
Assembly size (Mb)	730	2,300	382	272	385	125	401	**2,051**
Transposable element	0.62	0.842	0.4	0.281	0.374	0.14	0.4	**0.60**
Gene number	27,640	32,540	34,792	25,532	45,555	25,498	35,472	**31,987**
Average gene length (bp)	2,873 (without UTR)	3,757	3,039	2,956 (without UTR)	2,300	2,011	987 (CDS)	**3,350 (without UTR) 1,213 (CDS)**
Gene density (kb/gene)	24	13.1 (non-repeat region)	11.0	10.7	8.5	4.5	11.3	**64.1**
Average exon # per gene	4.7	5.3	3.7	5.5	4.3	5.2	4.5	**5.3**
Average exon len (bp)	268	304	256	268	254	250	163 (median)	**227**
Average intron length (bp)	436	516	409	391	379	168	135 (median)	**492**
Reference	Main text Ref.36	Main text Ref.26	28	10	Main text Ref.11	14	29	This study

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 8 Pathway similarity between bamboo and selected grass genomes.

Entry ID	Class of pathway	P.he vs. O.sa			P.he vs. S.bi			P.he vs. Z.ma		
		O. sa	P.he	Similarity	S. bi	P.he	Similarity	Z. ma	P.he	Similarity
KO00010	Glycolysis / Gluconeogenesis	91	123	0.929	66	62	0.818	85	100	0.840
KO00020	Citrate cycle (TCA cycle)	48	59	0.895	30	32	0.867	45	34	0.778
KO00030	Pentose phosphate pathway	40	59	0.929	30	27	0.857	42	46	0.929
KO00051	Fructose and mannose metabolism	42	50	0.765	26	16	0.600	38	35	0.714
KO00052	Galactose metabolism	31	31	0.833	24	13	0.667	24	18	0.857
KO00061	Fatty acid biosynthesis	18	20	1.000	24	10	0.400	24	26	0.800
KO00071	Fatty acid metabolism	31	46	0.700	14	13	0.571	21	25	0.667
KO00100	Steroid biosynthesis	18	24	0.769	36	10	0.467	22	21	0.769
KO00130	Ubiquinone and other terpenoid-quinone biosynthesis	25	15	0.357	19	0	0.000	16	3	0.200
KO00190	Oxidative phosphorylation	122	109	0.605	89	65	0.413	99	74	0.525
KO00195	Photosynthesis	71	21	0.259	59	15	0.217	66	18	0.216
KO00230	Purine metabolism	96	102	0.768	96	42	0.406	79	51	0.537
KO00240	Pyrimidine metabolism	83	66	0.656	76	29	0.316	70	33	0.447
KO00250	Alanine, aspartate and glutamate metabolism	40	36	0.652	25	22	0.450	37	27	0.611
KO00260	Glycine, serine and threonine metabolism	36	35	0.714	28	17	0.500	26	23	0.765
KO00270	Cysteine and methionine metabolism	60	72	0.917	33	21	0.500	38	40	0.778
KO00280	Valine, leucine and isoleucine degradation	32	39	0.765	20	14	0.500	27	29	0.563
KO00290	Valine, leucine and isoleucine biosynthesis	29	27	0.600	27	11	0.429	29	16	0.615
KO00300	Lysine biosynthesis	14	10	0.875	12	2	0.250	12	6	0.625
KO00330	Arginine and proline metabolism	44	64	0.750	29	26	0.478	39	48	0.652
KO00340	Histidine metabolism	19	26	0.500	11	0	0.000	15	14	0.222
KO00350	Tyrosine metabolism	25	31	0.583	17	8	0.222	25	18	0.385
KO00360	Phenylalanine metabolism	47	57	0.667	30	11	0.182	28	22	0.417
KO00380	Tryptophan metabolism	29	40	0.688	15	5	0.222	22	17	0.364
KO00400	Phenylalanine, tyrosine and tryptophan biosynthesis	33	34	0.611	24	7	0.188	23	20	0.571
KO00410	beta-Alanine metabolism	20	30	0.800	17	8	0.300	20	22	0.667

Nature Genetics: doi:10.1038/ng.2569

KO00450	Selenoamino acid metabolism	24	25	0.643	15	6	0.250	17	10	0.556
KO00460	Cyanoamino acid metabolism	16	16	0.875	12	5	0.143	10	6	0.500
KO00480	Glutathione metabolism	45	53	0.813	25	16	0.571	31	27	0.667
KO00500	Starch and sucrose metabolism	83	76	0.808	55	23	0.429	54	36	0.684
KO00510	N-Glycan biosynthesis	34	37	0.667	33	15	0.296	32	26	0.560
KO00520	Amino sugar and nucleotide sugar metabolism	78	73	0.833	47	20	0.476	49	39	0.722
KO00561	Glycerolipid metabolism	32	30	0.471	29	5	0.357	28	18	0.375
KO00562	Inositol phosphate metabolism	29	31	0.769	18	10	0.500	18	12	0.750
KO00564	Glycerophospholipid metabolism	46	38	0.542	34	9	0.389	33	12	0.474
KO00620	Pyruvate metabolism	56	79	0.762	44	24	0.529	49	46	0.650
KO00630	Glyoxylate and dicarboxylate metabolism	33	28	0.818	18	15	0.636	19	15	0.700
KO00640	Propanoate metabolism	17	31	0.889	18	9	0.364	22	25	0.667
KO00650	Butanoate metabolism	26	22	0.667	18	12	0.556	28	14	0.545
KO00670	One carbon pool by folate	14	15	0.778	18	12	0.571	12	8	0.429
KO00710	Carbon fixation in photosynthetic organisms	71	80	0.773	18	32	0.750	56	58	0.895
KO00770	Pantothenate and CoA biosynthesis	17	14	0.583	18	7	0.400	15	7	0.444
KO00860	Porphyrin and chlorophyll metabolism	35	31	0.714	18	8	0.240	34	21	0.500
KO00900	Terpenoid backbone biosynthesis	38	36	0.750	18	9	0.222	33	20	0.684
KO00906	Carotenoid biosynthesis	17	15	0.615	18	3	0.182	18	5	0.250
KO00910	Nitrogen metabolism	29	42	0.789	18	23	0.500	22	15	0.700
KO00920	Sulfur metabolism	18	14	0.444	18	2	0.400	14	5	0.833
KO00940	Phenylpropanoid biosynthesis	45	52	0.786	18	14	0.375	20	16	0.625
KO00970	Aminoacyl-tRNA biosynthesis	51	47	0.778	18	12	0.375	34	16	0.409
KO01040	Biosynthesis of unsaturated fatty acids	24	18	0.778	18	14	0.778	26	19	0.889
KO03010	Ribosome	212	236	0.707	18	224	0.684	218	230	0.712
KO03018	RNA degradation	50	63	0.643	18	12	0.237	39	20	0.419
KO03020	RNA polymerase	35	22	0.720	18	12	0.400	27	14	0.632
KO03022	Basal transcription factors	26	19	0.550	18	6	0.188	16	9	0.462
KO03030	DNA replication	37	34	0.690	18	4	0.125	23	11	0.368
KO03040	Spliceosome	99	117	0.656	18	37	0.292	93	43	0.443
KO03050	Proteasome	61	70	0.829	18	58	0.800	56	54	0.758
KO03060	Protein export	42	37	0.680	18	23	0.524	34	22	0.500
KO03410	Base excision repair	35	21	0.480	18	1	0.059	20	6	0.200
KO03420	Nucleotide excision repair	45	41	0.600	18	7	0.185	26	16	0.421
KO03430	Mismatch repair	26	26	0.650	18	3	0.118	15	7	0.364
KO03440	Homologous recombination	24	21	0.611	18	2	0.083	14	5	0.222

Nature Genetics: doi:10.1038/ng.2569

KO04070	Phosphatidylinositol signaling system	26	52	0.727	18	17	0.556	11	18	0.500
KO04120	Ubiquitin mediated proteolysis	81	87	0.653	18	42	0.319	60	53	0.475
KO04130	SNARE interactions in vesicular transport	26	30	0.824	18	12	0.600	20	17	0.889
KO04141	Protein processing in endoplasmic reticulum	106	112	0.632	18	61	0.446	110	78	0.537
KO04144	Endocytosis	54	78	0.793	18	24	0.385	35	37	0.556
KO04145	Phagosome	59	96	0.963	18	66	0.870	43	66	0.952
KO04146	Peroxisome	52	46	0.621	18	15	0.250	43	24	0.458
KO04626	Plant-pathogen interaction	59	42	0.250	18	27	0.231	52	41	0.450
KO04650	Natural killer cell mediated cytotoxicity	16	26	0.750	18	15	0.750	16	22	1.000

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 9 Predicted non-coding RNA genes.

Supplementary Table 9a Summary of tRNA genes identified in maize (Z. ma), rice (O. sa), sorghum (S. bi), Brachypodium (B. di), bamboo (P. he), and Arabidopsis (A. th).

	Z.ma	O.sa	S.bi	B.di	P.he	A.th
tRNAs decoding Standard 20 AA	1,413	720	535	593	1,076	685
Selenocysteine tRNAs (TCA)	**4**	0	**1**	0	**6**	0
Possible suppressor tRNAs (CTA,TTA)	**7**	0	**1**	0	**1**	0
tRNAs with undetermined/unknown isotypes	14	0	8	7	2	1
Predicted pseudogenes	768	26	61	15	82	13
Total tRNAs	2,206	746	606	615	1,167	699

Supplementary Table 9b Conserved non-coding RNA genes in the moso bamboo genome.

ncRNA Type	Loci #	Average length (bp)	Total length (bp)	% of genome
tRNA	1,167	75	87,363	0.0043
rRNA	279	714	199,180	0.0097
SnoRNA	321	118	37,985	0.0019
C/D box	277	123	34,025	0.0017
H/ACA	44	90	3,960	0.0002
snRNA	173	140	24,248	0.0019
miRNA	225	114	25,748	0.0013

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 9c Prediction of microRNA target genes. The INFERNAL-predicted miRNAs were aligned to the bamboo gene models by Blastn with e-value at 1e-10. The microRNAs with were clustered into the different families according to outputs of the INFERNAL prediction against Rfam database. The functional domain of each gene was searched by InterproScan against pfam database.

MicroRNA family	Bamboo microRNA ID	Target gene	Pfam
mir156	PH01000043m01 PH01000118m01 PH01000586m01 PH01000780m01 PH01000906m01 PH01000906m02 PH01000906m03 PH01000906m03 PH01001039m01 PH01001164m01 PH01001488m01 PH01001488m02 PH01001488m03 PH01002501m01	PH01000002G1660	PF03110 SBP
		PH01000050G0170	PF03110 SBP
		PH01000095G1560	PF03110 SBP
		PH01000117G1390	PF03110 SBP
		PH01000145G0070	Unknown
		PH01000150G0460	PF03110 SBP
		PH01000164G0630	PF03110 SBP
		PH01000176G0670	PF03110 SBP
		PH01000300G1120	PF04499 SAPS
		PH01000327G0270	PF02518 HATPase_c; PF00183 HSP90
		PH01000450G0480	PF03110 SBP
		PH01000457G0590	PF00076 RRM_1; PF00398 RrnaAD
		PH01000548G0580	Unknown
		PH01000594G0440	Unknown
		PH01000770G0270	PF03110 SBP
		PH01000969G0180	PF03110 SBP
		PH01001337G0400	PF00450 Peptidase_S10
		PH01002673G0070	PF03110 SBP
		PH01002789G0180	PF03110 SBP
		PH01003178G0220	PF03110 SBP
		PH01003773G0220	PF03110 SBP
		PH01007654G0010	PF03110 SBP
MIR159	PH01000077m01 PH01000130m01 PH01000257m01 PH01001252m01	PH01000345G0790	PF00072 Response_reg; PF06203 CCT
		PH01000564G0790	PF00085 Thioredoxin
		PH01000630G0430	PF00085 Thioredoxin
		PH01001806G0140	PF06424 PRP1_N
		PH01003596G0230	Unknown
		PH01003618G0130	PF00072 Response_reg; PF06203 CCT
		PH01005144G0100	PF07690 MFS_1
mir160	PH01000093m01 PH01000264m01 PH01000290m01 PH01000323m01	PH01000044G0540	PF02362 B3; PF06507 Auxin_resp
		PH01000069G1210	PF06507 Auxin_resp
		PH01000138G0430	PF04539 Sigma70_r3; PF04542 Sigma70_r2; PF04545 Sigma70_r4

Nature Genetics: doi:10.1038/ng.2569

	PH01000331m01 PH01000483m01	PH01000305G0690	PF02309 AUX_IAA; PF02362 B3; PF06507 Auxin_resp
	PH01000702m01 PH01001046m03	PH01000556G0480	PF00856 SET; PF02182 YDG_SRA; PF05033 Pre-SET
	PH01002185m01	PH01000845G0410	PF00168 C2; PF01412 ArfGap; PF02362 B3; PF06507 Auxin_resp
		PH01001026G0300	PF02309 AUX_IAA; PF02362 B3; PF06507 Auxin_resp
		PH01001175G0060	PF03552 Cellulose_synt
		PH01001285G0430	PF02362 B3; PF06507 Auxin_resp
		PH01002498G0280	PF02309 AUX_IAA; PF02362 B3; PF06507 Auxin_resp
		PH01002685G0120	PF02309 AUX_IAA; PF02362 B3; PF06507 Auxin_resp
		PH01004610G0130	PF00892 EamA; PF03151 TPT; PF03188 Cytochrom_B561
MIR164	PH01000210m01 PH01000543m01	PH01000015G2120	Unknown
		PH01000041G2170	PF02365 NAM
		PH01000057G1650	PF00642 zf-CCCH; PF00013 KH_1
		PH01000084G0850	Unknown
		PH01000093G0340	PF02365 NAM
		PH01000110G0680	PF02365 NAM
		PH01000183G1320	PF02365 NAM
		PH01000409G0160	Unknown
		PH01000483G1000	PF02365 NAM
		PH01000501G0450	PF02365 NAM
		PH01001122G0430	PF00070 Pyr_redox; PF02852 Pyr_redox_dim
		PH01001309G0120	PF02365 NAM
		PH01001318G0370	Unknown
		PH01001320G0360	Unknown
		PH01002276G0270	PF02365 NAM
		PH01002494G0010	Unknown
		PH010004131G0150	Unknown
MIR167_1	PH01000015m01 PH01010914m01 PH01254050m01	PH01000070G2000	PF05071 NDUFA12
		PH01000152G1310	PF00327 Ribosomal_L30; PF08079 Ribosomal_L30_N
		PH01000224G0820	PF00327 Ribosomal_L30; PF08079 Ribosomal_L30_N
		PH01001042G0370	PF00664 ABC_membrane; PF00005 ABC_tran
		PH01001940G0210	PF00337 Gal-bind_lectin; PF01762 Galactosyl_T
		PH01002304G0040	PF00294 PfkB
MIR168	PH01000280m01 PH01000585m01	PH01000003G4150	PF04844 DUF623
		PH01000004G2090	PF10369 ALS_ss_C

Nature Genetics: doi:10.1038/ng.2569

PH01000008G2140	PF01426 BAH; PF00403 HMA; PF05641 Agenet
PH01000009G2100	PF03106 WRKY
PH01000013G1620	Unknown
PH01000016G0550	Unknown
PH01000017G1630	PF00225 Kinesin; PF11721 Malectin
PH01000019G2340	Unknown
PH01000024G0010	Unknown
PH01000026G0210	PF08417 PaO; PF00355 Rieske
PH01000028G2550	PF07690 MFS_1
PH01000036G1230	PF00069 Pkinase
PH01000042G1600	Unknown
PH01000045G0170	Unknown
PH01000069G1200	PF02630 SCO1-SenC
PH01000102G0710	Unknown
PH01000164G0270	PF01554 MatE
PH01000188G0960	Unknown
PH01000197G0530	PF00185 OTCace; PF02729 OTCace_N
PH01000223G0540	PF05003 DUF668; PF11961 DUF3475
PH01000224G0730	PF00125 Histone
PH01000245G0160	Unknown
PH01000349G0440	Unknown
PH01000352G0040	PF00155 Aminotran_1_2
PH01000356G0540	PF10447 EXOSC1
PH01000358G0390	PF00249 Myb_DNA-binding
PH01000358G0650	PF00612 IQ
PH01000361G0710	PF02576 DUF150
PH01000367G0560	PF01985 CRS1_YhbY
PH01000367G0850	PF04755 PAP_fibrillin
PH01000433G0520	PF00400 WD40
PH01000437G0470	Unknown
PH01000439G0060	PF00278 Orn_DAP_Arg_deC; PF02784 Orn_Arg_deC_N
PH01000448G0580	PF00010 HLH
PH01000538G0570	PF02134 UBACT; PF00899 ThiF; PF10585 UBA_e1_thiolCys
PH01000542G0140	PF01501 Glyco_transf_8; PF00403 HMA
PH01000590G0720	Unknown
PH01000591G0360	Unknown
PH01000597G0190	Unknown
PH01000603G0510	PF00201 UDPGT
PH01000623G0580	PF03081 Exo70
PH01000666G0100	PF00271 Helicase_C; PF09369 DUF1998
PH01000695G0190	PF03083 MtN3_slv

Nature Genetics: doi:10.1038/ng.2569

PH01000698G0590	PF01842 ACT
PH01000735G0110	PF03106 WRKY
PH01000748G0450	Unknown
PH01000753G0540	Unknown
PH01000780G0490	PF05703 DUF828; PF08458 PH_2
PH01000795G0520	PF01565 FAD_binding_4; PF08031 BBE
PH01000842G0630	PF01535 PPR
PH01000866G0400	Unknown
PH01000875G0370	PF02836 Glyco_hydro_2_C; PF05282 AAR2
PH01000890G0140	PF00481 PP2C
PH01000895G0460	PF00004 AAA
PH01001004G0250	PF10998 DUF2838
PH01001065G0050	PF10288 DUF2392
PH01001117G0310	PF02701 zf-Dof
PH01001135G0080	PF05770 Ins134_P3_kin
PH01001175G0340	PF00097 zf-C3HC4
PH01001188G0490	PF01486 K-box
PH01001194G0040	Unknown
PH01001262G0230	PF01501 Glyco_transf_8
PH01001528G0340	PF00226 DnaJ
PH01001597G0080	Unknown
PH01001599G0350	Unknown
PH01001740G0210	Unknown
PH01001760G0230	PF03690 UPF0160
PH01001874G0080	PF00076 RRM_1; PF00098 zf-CCHC
PH01001896G0320	PF00076 RRM_1
PH01001979G0270	PF01106 NifU
PH01001998G0030	Unknown
PH01002124G0020	Unknown
PH01002232G0310	PF02540 NAD_synthase
PH01002235G0180	PF01163 RIO1
PH01002316G0200	PF00650 CRAL_TRIO; PF03765 CRAL_TRIO_N
PH01002375G0200	PF00069 Pkinase
PH01002439G0150	PF01535 PPR
PH01002529G0110	PF04570 DUF581
PH01002705G0160	PF01762 Galactosyl_T
PH01002825G0210	Unknown
PH01003170G0110	PF03106 WRKY
PH01003342G0140	Unknown
PH01003422G0190	PF01926 MMR_HSR1; PF06071 YchF-GTPase_C
PH01003917G0020	PF01535 PPR
PH01004682G0110	PF02686 Glu-tRNAGln
PH01004717G0040	Unknown

Nature Genetics: doi:10.1038/ng.2569

		PH01004719G0070	PF04597 Ribophorin_I
		PH01005322G0010	PF02309 AUX_IAA; PF02362 B3; PF06507 Auxin_resp
		PH01005724G0010	Unknown
		PH01007546G0020	PF07719 TPR_2
		PH01040671G0010	Unknown
		PH01262640G0010	PF01171 ATP_bind_3
MIR169_2	PH01000450m01	PH01000148G1100	PF00091 Tubulin; PF12327 FtsZ_C
MIR169_5	PH01000117m01	PH01000006G0960	Unknown
	PH01000117m02	PH01000235G0700	Unknown
	PH01001476m01	PH01000463G1100	Unknown
	PH01002131m01	PH01000752G0210	Unknown
MIR171_1		PH01000027G2350	PF00847 AP2
		PH01000095G1440	PF00171 Aldedh
		PH01000123G0110	PF00171 Aldedh
		PH01000170G1390	PF03514 GRAS
	PH01000169m01	PH01000229G1370	PF02681 DUF212
	PH01000366m01	PH01000498G0680	PF00310 GATase_2; PF01380 SIS
	PH01000390m01	PH01000666G0830	PF08389 Xpo1
	PH01001439m01	PH01000770G0710	Unknown
	PH01003493m01	PH01000850G0320	Unknown
	PH01003786m01	PH01001577G0130	PF03514 GRAS
	PH01004538m01	PH01001692G0030	PF03514 GRAS
		PH01001780G0280	PF00931 NB-ARC
		PH01002325G0040	PF03514 GRAS
		PH01002597G0150	PF12214 TPX2_importin
		PH01099851G0010	Unknown
MIR171_2		PH01000369G0580	Unknown
		PH01000464G0470	PF08030 NAD_binding_6; PF01794 Ferric_reduct; PF08414 NADPH_Ox
	PH01004540m01	PH01000596G0230	PF08022 FAD_binding_8; PF08030 NAD_binding_6; PF01794 Ferric_reduct; PF08414 NADPH_Ox
	PH01000211m01		
	PH01000200m01	PH01001043G0440	PF00069 Pkinase
		PH01001716G0120	Unknown
		PH01002279G0010	PF03070 TENA_THI-4
		PH01002838G0190	Unknown
		PH01003904G0030	PF01535 PPR
mir172		PH01000037G1490	Unknown
	PH01000021m01	PH01000052G2050	Unknown
	PH01000466m01	PH01000097G0900	PF00646 F-box
	PH01004738m01	PH01000127G0210	Unknown
		PH01000213G0510	PF00646 F-box

Nature Genetics: doi:10.1038/ng.2569

		PH01000365G0970	Unknown
		PH01001028G0270	Unknown
		PH01002138G0280	PF00067 p450
		PH01002503G0090	Unknown
		PH01002747G0060	PF05450 Nicastrin
		PH01002963G0090	Unknown
		PH01003375G0030	PF04859 DUF641
		PH01005644G0020	PF00646 F-box
mir395	PH01000021m01 PH01000466m01 PH01004738m01	PH01000314G0270	PF00078 RVT_1; PF00789 UBX; PF08284 RVP_2
		PH01000607G0360	PF00026 Asp; PF05184 SapB_1; PF10551 MULE
		PH01000669G0280	PF03463 eRF1_1; PF03464 eRF1_2; PF03465 eRF1_3
		PH01000853G0290	PF03637 Mob1_phocein
		PH01001115G0420	Unknown
		PH01001142G0220	PF00560 LRR_1; PF08263 LRRNT_2; PF00069 Pkinase
		PH01001208G0380	PF00026 Asp; PF05184 SapB_1; PF03489 SapB_2
		PH01001821G0330	PF05703 DUF828
		PH01002024G0310	Unknown
		PH01003316G0070	PF03637 Mob1_phocein
mir399	PH01004613m01 PH01000814m01 PH01000429m01 PH01000000m01	PH01000010G1850	PF03016 Exostosin
		PH01000157G1010	PF03016 Exostosin
		PH01000271G0200	PF09258 Glyco_transf_64
		PH01000436G0030	Unknown
		PH01000502G0800	PF09258 Glyco_transf_64
		PH01001655G0500	PF00566 TBC
		PH01002153G0330	Unknown
		PH01002839G0140	PF04116 FA_hydroxylase; PF12076 Wax2_C
		PH01003152G0320	Unknown
		PH01003440G0030	Unknown
		PH01003658G0100	Unknown
		PH01004857G0110	PF00566 TBC
		PH01006818G0010	PF12076 Wax2_C
MIR535	PH01001265m01 PH01001265m02 PH01001768m01 PH01004182m01	PH01000095G1560	PF03110 SBP
		PH01000145G0070	Unknown
		PH01000164G0630	PF03110 SBP
		PH01000176G0670	PF03110 SBP
		PH01000474G0490	PF03005 DUF231
		PH01000548G0580	Unknown
		PH01000552G0240	Unknown
		PH01000740G0630	PF01255 Prenyltransf
		PH01001050G0540	PF04577 DUF563

Nature Genetics: doi:10.1038/ng.2569

		PH01001376G0370	PF01490 Aa_trans
		PH01001876G0100	PF00190 Cupin_1; PF08700 Vps51
		PH01001940G0250	PF03171 2OG-FeII_Oxy
		PH01001963G0380	PF00400 WD40
		PH01002134G0090	PF00400 WD40
		PH01002190G0080	PF00067 p450
		PH01003819G0090	PF00139 Lectin_legB; PF00069 Pkinase
		PH01004078G0110	PF00069 Pkinase
		PH01007654G0010	PF03110 SBP
MIR821	PH01000753m01	PH01000272G1270	PF00505 HMG_box
	PH01000535m01	PH01001058G0400	PF00657 Lipase_GDSL
	PH01000378m01	PH01001316G0020	PF03110 SBP
	PH01003487m01		

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 10 Statistics of Repetitive sequences.

Supplementary Table 10a Repetitive sequences in the moso bamboo genome.

	Length occupied (bp)	Percentage of sequences
Class I elements (Retroelements)	**790,027,115**	**0.385**
LTR Retrotransposons	**764,632,374**	**0.373**
LTR/*Copia*	251,526,442	0.123
LTR/*Gypsy*	505,181,075	0.246
unclassified LTR	7,924,857	0.004
non-LTR Retrotransposons	**24,526,904**	**0.012**
LINE	23,701,208	0.012
SINE	825,696	0.000
unclassified retrotransposons	**867,837**	**0.000**
Class II elements (DNA Transposons)	**194,238,269**	**0.095**
DNA Transposons	**176,524,830**	**0.086**
DNA/En-Spm	74,871,248	0.036
DNA/hAT	26,546,823	0.013
DNA/MuDR	73,460,366	0.036
DNA/Harbinger	1,646,393	0.001
MITEs	**7,356,902**	0.004
DNA/TcMar-Stowaway	3,209,627	0.002
DNA/Tourist	4,147,275	0.002
RC/Helitrons	**3,241,423**	**0.002**
unclassified transposons	**7,115,114**	**0.003**
Unknown repeats	**226,597,546**	**0.110**
Total transposable elements	**1,210,862,930**	**0.590**
Low_complexity	**1,130,281**	**0.001**

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 10b Comparison of TEs with highest copies among moso bamboo, rice, and sorghum.

Bamboo			Rice			Sorghum		
Repeat ID	**Class of TE-element**	**Copies**	**Repeat ID**	**Class of TE-element**	**Copies**	**Repeat ID**	**Class of TE-element**	**Copies**
PH01R6F001124	LTR/Gypsy	20,624	WANDERER_OS	DNA/Tourist	8,531	TSB1	Harbinger	20,984
PH01R1F000001	DNA/En-Spm	16,987	STOWAWAY41_OS	DNA/TcMar-Stowaway	5,821	Tourist1a_SB	Harbinger	7,731
PH01R3F000573	DNA/En-Spm	16,918	Gaijin	DNA/Tourist	4,986	Copia-141_SB-LTR	Copia	7,430
PH01R1F000000	DNA/En-Spm	15,691	STOWAWAY47_OS	DNA/TcMar-Stowaway	4,518	ATHILA-1_SBi-LTR	Gypsy	6,810
PH01R6F000783	Unknown	13,094	STOWAWAY1_OS	DNA/TcMar-Stowaway	4,445	Gypsy-133_SBi-LTR	Gypsy	5,863
PH01R5F002220	LTR/Gypsy	12,658	SEVERIN-2	RC/Helitron	4,260	Gypsy-122_SBi-LTR	Gypsy	5,802
PH01R1F000002	LTR/Gypsy	11,637	TREP215	DNA/TcMar-Stowaway	4,142	TSB2	Harbinger	5,720
PH01R5F002786	LTR/Copia	10,725	Explorer	DNA	3,854	Gypsy-125_SBi-LTR	Gypsy	5,482
PH01R5F002970	LTR/Gypsy	9,602	STOWAWAY2_OS	DNA/TcMar-Stowaway	3,671	Gypsy-127_SBi-LTR	Gypsy	5,337
PH01R6F002209	LTR/Gypsy	9,446	SINE03_OS	SINE	3,507	Gypsy-125_SBi-I	Gypsy	5,002

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 10c Comparison of TEs occupying most genome size among the moso bamboo, rice, and sorghum.

Bamboo				Rice				Sorghum			
Repeat ID	Class of TE-element	Copies	Occupied genome size (bp)	Repeat ID	Class of TE-element	Copies	Occupied genome size (bp)	Repeat ID	Class of TE-element	Copies	Occupied genome size (bp)
PH01R6F001124	LTR/Gypsy	20624	21071416	SPMLIKE	DNA/En-Spm	2150	6292696	Gypsy-122_SBi-I	Gypsy	4366	27222661
PH01R5F003093	LTR/Gypsy	6366	16535123	RETRO2_I	LTR/Gypsy	628	3846962	Gypsy-133_SBi-I	Gypsy	4979	17779891
PH01R6F002209	LTR/Gypsy	9446	15516557	RIRE2_I	LTR/Gypsy	758	3498738	Gypsy-121_SBi-I	Gypsy	3137	14031073
PH01R1F000002	LTR/Gypsy	11637	12906984	RIRE3_LTR	LTR/Gypsy	1956	3338048	ATHILA-1_SBi-I	Gypsy	4976	11493996
PH01R6F000836	LTR/Copia	7388	12860851	ATLANTYS-I_OS	LTR/Gypsy	1086	3215554	Gypsy-136_SBi-I	Gypsy	3917	10004929
PH01R5F002786	LTR/Copia	10725	12543585	TRUNCATOR	LTR/Gypsy	1755	2952995	ATHILA-1_SBi-LTR	Gypsy	6810	8641228
PH01R6F004818	LTR/Gypsy	4337	12059267	RIRE3A_LTR	LTR/Gypsy	2301	2796270	Gypsy-125_SBi-I	Gypsy	5002	8122387
PH01R6F002298	LTR/Gypsy	6746	11580121	SZ-7_int	LTR/Gypsy	693	2394637	Gypsy-128_SBi-LTR	Gypsy	2820	7778326
PH01R5F002220	LTR/Gypsy	12658	10774197	TRUNCATOR2_OS	LTR/Gypsy	2411	2356738	Gypsy-122B_SBi-I	Gypsy	2366	7355980
PH01R1F000015	LTR/Copia	6969	9936341	RIREX_I	LTR/Gypsy	967	2179712	ATHILA-3_SBi-I	Gypsy	3267	7080252

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 11 Mean Ks and divergence time for the bamboo versus grass species. Mean Ks and divergence time between the bamboo and fully sequenced grass species were calculated from the Ks distribution of obtainned 968 single-copy gene families. Calculation of Ks were performed by MA model that averages parameters across 14 candidate models[30]. Divergence time were calculted using a substitution rate of 6.5×10^{-9} mutations per site per year. The Ks of internal dupliation was estimated by calculating the Ks of the paralogous pair from the 2-member gene clusters, of which the derived divergence time was indicative of the WGD time.

Species	Mean Ks	Divergence time (mya)
Brachypodium vs.Bamboo	0.610	46.9
Wheat[1] vs. Bamboo	0.621	47.8
Rice vs.Bamboo	0.632	48.6
Foxtail millet vs. Bamboo	0.701	53.9
Sorghum vs. Bamboo	0.761	58.5
Maize vs. Bamboo	0.840	64.6
Bamboo **internal duplication**	0.10 - 0.15	7.7 - 11.5
Maize **internal duplication**	0.15 - 0.20	11.5 - 15.4

[1] The wheat gene models were downloaded at ftp://ftp.ncbi.nih.gov/repository/UniGene/Triticum_aestivum/.

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 12 Change of gene family size in bamboo, with comparison to different plant genomes. Fields highlighted in green represented the families with expanding gene number. Fields highlighted in red represented the families with contracting gene number. The gene families were generated by the OrthoMCL analysis. The estimation of gene familiy size change was perfromed by a CAFE calculation with *P*-value < 0.01.

Family No.	Description of conserved function domains	Phe	Bdi	Osa	Sbi	Zma	Sit	Ath	P-value
# 4	PF03552 Cellulose_synt,PF01652 IF4E	31	15	23	25	31	27	16	0.005
# 18	PF00931 NB-ARC,PF00560 LRR_1,PF01657 DUF26,PF01419 Jacalin	27	2	39	25	2	30	0	0.000
# 24	PF02364 Glucan_synthase,PF04652 DUF605	18	11	12	12	8	23	12	0.001
# 702	PF00931 NB-ARC	14	0	6	1	0	0	0	0.000
# 66	PF00931 NB-ARC,PF00400 WD40,PF02671 PAH,PF03446 NAD_binding_2	13	0	15	14	8	19	0	0.000
# 43	PF00069 Pkinase	12	9	9	7	11	20	8	0.004
# 451	PF00403 HMA	12	4	3	2	2	1	1	0.009
# 82	PF07645 EGF_CA PF02797 Chal_sti_synt_C PF08392 FAE1_CUT1_RppA PF00069 Pkinase	11	7	12	12	5	15	0	0.000
# 116	PF00931 NB-ARC	11	0	9	13	2	15	0	0.000
# 89	PF00862 Sucrose_synth,PF00534 Glycos_transf_1	9	6	7	3	7	13	6	0.005
# 457	PF00931 NB-ARC	9	1	8	3	0	5	0	0.000
# 98	PF03822 NAF,PF00069 Pkinase	8	5	6	5	6	17	4	0.001
# 125	PF00400 WD40	8	3	3	2	4	12	5	0.001
# 175	PF00072 Response_reg,PF02518 HATPase_c,PF00512 HisKA,PF03924 CHASE	8	3	4	2	7	8	3	0.005
# 285	PF04578 DUF594	8	5	6	5	1	7	0	0.003
# 348	PF00931 NB-ARC	8	2	8	2	1	8	0	0.000
# 109	PF03030 H_PPase	7	4	6	6	8	16	1	0.002
# 140	PF04810 zf-Sec23_Sec24,PF04811 Sec23_trunk,PF04815 Sec23_helical,PF00626 Gelsolin,PF08033 Sec23_BS	7	4	4	4	4	13	5	0.008
# 331	PF00560 LRR_1,PF00931 NB-ARC	7	5	2	9	2	5	0	0.001
# 265	PF02309 AUX_IAA,PF02362 B3,PF06507 Auxin_resp	6	3	3	2	5	9	1	0.009
# 307	PF00560 LRR_1 PF00931 NB-ARC	6	4	8	3	1	9	0	0.001
# 454	PF00931 NB-ARC	6	2	6	8	0	4	0	0.000
# 10664	PF00560 LRR_1	6	0	0	1	0	0	0	0.001
# 215	PF01909 NTP_transf_2,PF04926 PAP_RNA-bind,PF04928 PAP_central	5	3	3	3	3	14	3	0.001
# 243	PF00026 Asp,PF05184 SapB_1,PF03489 SapB_2	5	2	3	1	3	14	3	0.000
# 323	PF00067 p450,PF02298 Cu_bind_like	5	1	3	9	5	7	0	0.004
# 385	PF00560 LRR_1 PF00931 NB-ARC PF00069 Pkinase	5	2	5	3	1	12	0	0.000
# 622	PF01909 NTP_transf_2	4	1	1	1	2	8	1	0.002

Nature Genetics: doi:10.1038/ng.2569

# 700	PF01428 zf-AN1,PF01754 zf-A20	4	1	2	0	3	6	1	0.003
# 246	PF00067 p450	3	1	2	1	1	10	6	0.002
# 618	PF02896 PEP-utilizers_C PF01326 PPDK_N PF00391 PEP-utilizers	3	1	2	1	2	11	1	0.001
# 1	PF00560 LRR_1,PF08263 LRRNT_2,PF00069 Pkinase	30	24	68	44	19	51	5	0.000
# 5	PF00139 Lectin_legB,PF00106 adh_short,PF00069 Pkinase	22	28	37	30	23	30	13	0.006
# 2	PF00954 S_locus_glycop,PF07714 Pkinase_Tyr,PF01453 B_lectin,PF08276 PAN_2,PF11883	19	18	27	28	18	39	26	0.000
# 14	PF00560 LRR_1,PF08263 LRRNT_2,PF00069 Pkinase,PF01280 Ribosomal_L19e	15	17	30	21	13	32	5	0.000
# 28	PF00954 S_locus_glycop,PF01453 B_lectin,PF08276 PAN_2,PF00069 Pkinase	15	16	34	15	13	5	1	0.000
# 11	PF07714 Pkinase_Tyr,PF00202 Aminotran_3,PF11721 Malectin ,PF00560 LRR_1	11	6	25	6	8	25	13	0.000
# 35	PF00954 S_locus_glycop,PF01453 B_lectin,PF08276 PAN_2,PF00069 Pkinase	9	11	29	14	5	12	0	0.000
# 70	PF00560 LRR_1,PF01464 SLT,PF08263 LRRNT_2	9	9	12	10	4	19	1	0.000
# 74	PF00125 Histone	8	13	10	11	2	1	8	0.000
# 21	PF00271 Helicase_C,PF01535 PPR,PF02365 NAM,PF02466 Tim17	7	7	10	18	7	12	23	0.001
# 126	PF00560 LRR_1,PF00931 NB-ARC	7	16	14	2	0	10	0	0.000
# 96	PF00560 LRR_1,PF00931 NB-ARC	6	14	4	13	3	17	0	0.000
# 122	PF00117 GATase,PF06418 CTP_synth_N	5	5	6	3	5	15	5	0.003
# 108	PF00931 NB-ARC,PF05725 FNIP,PF00560 LRR_1	4	7	12	15	1	14	0	0.000
# 159	PF00128 Alpha-amylase,PF07821 Alpha-amyl_C2	4	3	6	6	4	16	1	0.000
# 197	PF00931 NB-ARC	4	3	12	7	3	9	0	0.000
# 26	PF00560 LRR_1,PF00069 Pkinase	3	12	12	7	3	14	42	0.000
# 71	PF00069 Pkinase	3	4	17	2	8	4	0	0.000
# 131	PF02519 Auxin_inducible	3	1	8	5	3	9	9	0.005
# 138	PF00009 GTP_EFTU,PF03143 GTP_EFTU_D3,PF03144 GTP_EFTU_D2	3	4	4	1	7	16	4	0.000
# 287	PF00931 NB-ARC	3	9	7	4	2	7	0	0.004
# 294	PF07714 Pkinase_Tyr,PF07645 EGF_CA	3	8	7	7	2	5	0	0.002
# 169	PF00931 NB-ARC	2	3	7	6	8	15	0	0.000
# 170	PF07714 Pkinase_Tyr	2	4	12	3	3	3	0	0.010
# 191	PF00232 Glyco_hydro_1	2	2	3	8	4	12	3	0.003
# 205	PF03171 2OG-FeII_Oxy,PF11744 ALMT	2	3	2	3	2	13	5	0.001
# 206	PF04398 DUF538	2	2	9	5	7	13	0	0.000
# 262	PF00854 PTR2	2	3	5	7	3	9	1	0.010
# 313	PF00931 NB-ARC	2	3	5	6	1	12	0	0.000
# 332	PF00931 NB-ARC	2	2	7	5	9	5	0	0.005
# 461	PF02892 zf-BED	2	2	3	6	0	5	1	0.002
# 1316	PF00931 NB-ARC	2	2	7	3	0	3	0	0.004
# 55	PF00560 LRR_1,PF08263 LRRNT_2	1	5	25	6	4	8	18	0.000
# 68	PF01107 MP,PF00098 zf-CCHC,PF02160 Peptidase_A3,PF00078 RVT_1,PF00077 RVP	1	0	12	21	0	3	0	0.000

Nature Genetics: doi:10.1038/ng.2569

# 106	PF00067 p450	1	10	10	5	3	10	0	0.000
# 111	PF00394 Cu-oxidase,PF07731 Cu-oxidase_2,PF07732 Cu-oxidase_3	1	10	4	3	4	30	0	0.000
# 221	PF00232 Glyco_hydro_1	1	3	5	9	8	2	6	0.007
# 223	PF00891 Methyltransf_2,PF08100 Dimerisation	1	0	7	2	7	3	0	0.000
# 252	PF03330 DPBB_1	1	8	2	5	3	9	0	0.000
# 475	PF00931 NB-ARC	1	1	5	8	2	9	0	0.000
# 556	PF00067 p450	1	1	11	4	2	5	0	0.000
# 678	PF03478 DUF295,PF04493 Endonuclease_5	1	1	2	3	10	3	1	0.001
# 754	PF04578 DUF594	1	2	2	1	0	15	0	0.000
# 755	PF03087 DUF241	1	3	6	4	0	3	0	0.001
# 845	PF02362 B3	1	1	2	12	0	4	0	0.000
# 849	PF04578 DUF594	1	2	3	4	1	9	0	0.001
# 1138	PF00646 F-box	1	2	10	2	0	2	0	0.001
# 1139	PF07714 Pkinase_Tyr,PF07645 EGF_CA,PF08488 WAK	1	3	7	1	1	5	0	0.004
# 1140	PF00646 F-box	1	2	13	1	0	1	0	0.000
# 1536	PF00067 p450	1	3	2	9	1	0	0	0.000
# 2111	PF03088 Str_synth	1	1	3	5	0	4	0	0.001
# 88	PF01593 Amino_oxidase,PF02721 DUF223	0	2	1	2	54	0	0	0.000
# 171	PF05699 hATC	0	0	8	23	0	1	7	0.000
# 222	PF01357 Pollen_allerg_1	0	0	15	6	9	7	0	0.000
# 276	PF03004 Transposase_24	0	2	5	2	19	5	0	0.000
# 295	PF00069 Pkinase	0	8	1	6	8	9	0	0.000
# 333	PF00067 p450	0	2	1	5	1	10	7	0.000
# 335	PF00321 Thionin	0	4	12	1	0	9	4	0.000
# 472	PF00177 Ribosomal_S7	0	5	11	2	3	1	2	0.000
# 473	PF00646 F-box,PF03169 OPT	0	8	5	4	7	2	0	0.000

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 13 Gene synteny.

Supplementary Table 13a Statistics of syntenic bamboo loci on the aligned rice genome

Rice	No. of syntenic gene blocks	No. of syntenic genes	Average gene number per block
Chromosome-01	217	2,563	11.8
Chromosome-02	153	1,756	11.5
Chromosome-03	216	2,467	11.4
Chromosome-04	135	1,747	12.9
Chromosome-05	166	1,506	9.1
Chromosome-06	147	1,389	9.4
Chromosome-07	126	1,165	9.2
Chromosome-08	91	812	8.9
Chromosome-09	110	962	8.7
Chromosome-10	96	850	8.9
Chromosome-11	60	932	15.5
Chromosome-12	90	1,586	17.6
sum	1,607	17,735	11.0

Supplementary Table 13b Statistics of syntenic bamboo loci on the aligned sorghum genome

Sorghum	No. of syntenic gene blocks	No. of syntenic genes	Average gene number per block
Chromosome-01	296	3,319	11.2
Chromosome-02	212	1,893	8.9
Chromosome-03	199	2,364	11.9
Chromosome-04	159	1,667	10.5
Chromosome-05	48	352	7.3
Chromosome-06	134	1,688	12.6
Chromosome-07	102	909	8.9
Chromosome-08	72	708	9.8
Chromosome-09	167	1,507	9.0
Chromosome-10	150	1,339	8.9
Sum	1,539	15,746	10.2

Note: A total of 30,379 (94.8% of 31,987) Bamboo loci located on the scaffolds with length over 50 KB were aligned to the rice and sorghum gene models, respectively. At least 5 genes are required to call synteny. Within a syntenic gene block, the maximum number of non-syntenic genes between two adjacent syntenic genes should be less than 5.

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 14 Quantity of cell wall biosynthesis genes in plant genomes.

Supplementary Table 14a Comparison of copy numbers of cellulose synthase (*CesA*) and cellulose synthase-like (*Csl*) genes among grasses, Arabidopsis, and poplar. The subfamilies *CslA, C, D, E, F*, G, H, and *J* were classified by the referenced genes located on the same clade.

		P.he	A.th	Z.ma	B.di	S.bi	O.sa	P.tr
CesA		19	10	20	9	12	11	18
Csl	Total	38	29	33	24	37	34	37
	CslA	9	8	10	8	8	10	5
	CslB	0	6	0	0	0	0	2
	CslC	10	5	8	4	6	6	5
	CslD	7	6	5	3	5	5	11
	CslE	4	1	2	3	3	3	3
	CslF	7	0	7	5	11	8	0
	CslG	0	3	0	0	0	0	4
	CslH	0	0	0	1	3	2	0
	CslJ	1	0	1	0	1	0	2

Supplementary Table 14b Copy number of genes involved in phenylpropanoid and lignin biosynthetic pathways.

	PAL	C4H	C3H	4CL	HCT	CCR	CCoAOMT	CAD	F5H	COMT
P.he	**8**	**4**	**3**	**6**	**4**	**3 ***	**2**	**1**	**3**	**1**
A.th	4	1	3	4	1	2	1	2	2	1
Z.ma	10	4	2	3	2	1	2	1	2	1
B.di	8	3	1	5	2	2	1	1	2	1
S.bi	9	3	2	5	2	2	1	1	2	1
O.sa	9	4	2	5	2	2	1	1	3	1
P.tr	5	2	3	5	2	7	2	1	4	2

* Two interrupted bamboo CCR genes were not included.

Abbreviation of the encoded proteins: Phenylalanine amonnia lyase (PAL), Cinnamate-4-hydroxylase (C4H), p-Coumaroyl shikimate 3'-hydroxylase/Coumaroyl 3-hydroxylase (C3H), 4-Coumarate:CoA Ligas (4CL), Hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyltransferase (HCT), Cinnamoyl-CoA reductase (CCR), Trans-caffeoyl-CoA 3-O-methyltransferase (CCoAOMT), Cinnamyl alcohol dehydrogenase (CAD), Ferulate 5-hydroxylase (F5H), Caffeic acid 3-O-methyltransferase (COMT).

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 15 Expression of cell wall genes in bamboo.

Supplementary Table 15a Gene expression of the *CesA* subfamilies. The expression level was shown as the quantified transcript level (RPKM) in the 7 sequenced tissues.

Annotated bamboo loci of *CesA* subfamiy	Quantified transcript level (RPKM)						
	shoot20	shoot50	rhizome	root	leaf	panicle1	panicle2
PH01000018G0380	2	2	16	52	96	46	32
PH01000040G0670	15	13	4	5	2	2	1
PH01000204G0350	39	27	58	115	35	38	15
PH01000357G0340	23	22	9	5	8	6	5
PH01000441G0190	38	33	10	6	4	3	3
PH01000482G0850	27	25	0	3	1	1	0
PH01000536G0710	1	1	2	2	1	2	3
PH01000693G0390	3	1	24	104	125	80	55
PH01000746G0570	3	1	10	43	45	36	19
PH01000905G0290	32	19	17	57	9	10	4
PH01000924G0590	2	0	12	69	64	51	31
PH01001105G0500	23	16	10	22	11	6	2
PH01001146G0100	0	0	0	0	0	1	1
PH01001175G0060	38	28	16	37	8	13	9
PH01001427G0390	47	30	6	8	6	3	2
PH01002002G0290	46	48	46	20	27	26	14
PH01002004G0190	1	1	5	20	28	14	12
PH01002232G0080	8	7	1	3	1	2	0
PH01003000G0020	22	18	16	16	11	9	5

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 15b Gene expression of the *Csl* subfamilies. The expression level was shown as the quantified transcript level (RPKM) in the 7 sequenced tissues.

Csl subfamily	Annotated bamboo loci	Quantified transcript level (RPKM)						
		shoot20	shoot50	rhizome	root	leaf	panicle1	panicle2
CslA	PH01000068G1400	18	20	10	5	5	3	2
	PH01000135G1390	15	10	15	37	6	5	9
	PH01000366G0010	3	2	24	15	4	4	4
	PH01000484G0010	9	10	15	6	19	24	16
	PH01000926G0020	7	6	2	2	0	0	0
	PH01001938G0190	0	0	0	0	0	0	0
	PH01002436G0250	124	117	26	17	17	12	5
	PH01003149G0030	17	27	8	2	2	4	3
	PH01006529G0050	6	6	12	20	13	11	7
CslC	PH01000099G1140	5	3	3	4	1	1	0
	PH01000155G1560	8	3	0	1	0	0	0
	PH01000379G0160	16	13	6	5	10	4	4
	PH01000423G0430	7	3	4	9	2	2	2
	PH01000523G0700	4	3	4	7	3	2	1
	PH01000883G0140	7	6	6	1	1	1	1
	PH01001206G0170	14	13	10	4	21	8	5
	PH01001338G0260	22	18	10	1	18	4	1
	PH01001674G0100	13	13	6	0	2	0	0
	PH01003525G0010	2	3	2	0	1	1	2
CslD	PH01000083G1270	24	25	22	14	10	6	5
	PH01000246G0240	6	9	5	5	5	4	3
	PH01001020G0160	22	24	4	1	1	1	0
	PH01001105G0500	23	16	10	22	11	6	2
	PH01001695G0020	11	11	3	1	1	1	0
	PH01002587G0110	0	0	0	0	0	0	1
	PH01002699G0130	0	0	0	0	0	0	0
CslE	PH01000699G0390	0	0	1	9	4	4	4
	PH01000941G0440	2	2	3	2	13	11	11
	PH01000941G0520	0	0	0	1	87	8	4
	PH01001562G0420	2	2	3	7	2	4	3
CslF	PH01001344G0180	11	12	16	26	25	13	6
	PH01001742G0110	26	16	41	43	31	23	11
	PH01001945G0240	0	0	0	0	9	0	0
	PH01001945G0320	4	1	49	66	0	0	0
	PH01001945G0340	1	0	0	0	0	0	0
	PH01002576G0070	0	0	0	0	3	1	1
	PH01002576G0110	1	1	3	5	23	7	3
CslJ	PH01002763G0260	0	0	0	1	4	9	16

Nature Genetics: doi:10.1038/ng.2569

Note: The plant cell walls represented the most predominant determinant of overall form, grow, and development. The younger moso bamboo shoot has fastest growth speed in the world, which can grow around 4 feet within 24 hours. Our phylogenic analysis and RNA-seq data revealed gene expansion in *CslCs*, *CslAs*, and *CesAs* and many duplications involved genes with higher expression in shoot. The *CslC* has been identified to encode a xyloglucan glucan synthase to produce β-1-4-glucans backbone of xyloglucans in primary wall formation[31]. The CSLA proteins synthesized β-(1-4)-linked mannan in the cell wall[32]. The recent study found that some member of both *CslA* and *CslC* subfamily were involved in hemicellulose predominantly in Golgi membranes[33]. Correspondingly, the *CslA* and *CslC* subfamilies were observed to have the highest copies among the *Csls*. So potentially, evolution of *CslC* and *CesA* families were important to formation of bamboo-specific characters in shoot development.

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 16 Gene expression of selected loci with significantly increased transcription level (>2- fold increase, Q-value < 0.001) in the floral tissues. The involved pathways referenced the function information of the Arabidopsis/rice homologs (TAIR10 and MSU RGAP 6.1) and conserved domains identified by Interpro. The 7 sequenced tissues were shown as S20 (tip of 20-cm shoot), S50 (tip of 50-cm shoot), RH (rhizome), RT (root), LF (leaf), P1 (panicle at the early stage), and P2 (panicle at the flowering stage). Only the loci with higher expression in two or more copies were listed in the table. Table listed the genes carrying clustered conserved function domains (Interpro domains).

Abbr.	Floral genes in bamboo	Known homologous genes	Referencing function	Interpro domains
ERF	PH01000032G1740	*OsERF3*[34]	Drought tolerance	IPR001471 Pathogenesis-related transcriptional factor/ERF, DNA-binding; IPR016177 DNA-binding, integrase-type
	PH01000046G1730			
	PH01000129G0360			
	PH01000573G0640	*OsAP2-39*[34]	Drought tolerance	
	PH01000573G0670			
	PH01001102G0050			
	PH01001360G0530			
	PH01001634G0140			
	PH01001704G0270			
	PH01002279G0250			
	PH01002393G0230			
	PH01002571G0300			
	PH01002648G0300			
	PH01003475G0200			
	PH01004791G0030			

Nature Genetics: doi:10.1038/ng.2569

	PH01000028G1600			
bZIP	PH01000173G1010	*OsbZIP23*[35]	Salinity and drought tolerance	IPR004827 Basic-leucine zipper (bZIP) transcription factor; IPR011616 bZIP transcription factor, bZIP-1
	PH01000437G0190	*OsbZIP23*[35]	Salinity and drought tolerance	
	PH01000682G0190			
	PH01000811G0380			
	PH01001986G0070			
	PH01000242G0910	*OsbZIP45*[36]		
CCT/B-box	PH01000831G0230			IPR010402 CCT domain
	PH01169704G0010			
F-box	PH01003868G0020			IPR001810 F-box domain, cyclin-like | IPR022364 F-box domain, Skp2-like
	PH01000198G0670			
	PH01000317G0460			
	PH01000610G0210			
HTH, Myb-type	PH01000383G0320			IPR001005 SANT, DNA-binding; IPR006447 Myb-like DNA-binding domain, SHAQKYF class; IPR017930 HTH transcriptional regulator, Myb-type, DNA-binding
	PH01002069G0200			
	PH01002344G0400			
	PH01002440G0410			
	PH01004165G0090			
homeobox	PH01000001G0190			IPR001356 Homeobox | IPR003106 Leucine zipper, homeobox-associated IPR017970 Homeobox, conserved site
	PH01000042G1220			
	PH01000496G0570			
	PH01000922G0150			
	PH01000630G0150			
MADS-box	PH01001303G0110			IPR002100 Transcription factor,

Nature Genetics: doi:10.1038/ng.2569

	PH01001952G0230			MADS-box
	PH01000306G0610	*OsMADS14*	FMI	
	PH01000317G0080	*OsMADS2*	FMI	
	PH01000606G0250	*OsMADS14*	FMI	
	PH01002127G0260	*OSMADS3*	FMI	
	PH01000222G1190	*OsMADS14*	FMI	
	PH01001188G0490			
	PH01001952G0190	*OsMADS1*	FMI	
NAC	PH01000053G1650	*OsNAC10*[37]	Salinity and drought tolerance	IPR003441 No apical meristem (NAM) protein
	PH01000122G1000	*SNAC1*[38]		
	PH01001843G0210	*SNAC1*[38]		
	PH01004006G0090	*OsNAC10*[37]		
WD-40	PH01000091G0440	*SRWD3*[39]	Salinity tolerance	IPR001680 WD40 repeat | IPR015943 WD40/YVTN repeat-like-containing domain | IPR019775 WD40 repeat, conserved site
	PH01000600G0380	*SRWD5*[39]	Salinity tolerance	
	PH01003526G0090			
YABBY	PH01000162G1010	*OsYABBY2*[40]	FMI	IPR006780 YABBY protein
zf-Dof	PH01000113G0300	*OsDof12*[23,24]	Regulation of FMI	IPR003851 Zinc finger, Dof-type
	PH01000290G0170			
	PH01000323G0330			
	PH01002061G0210			
HSP20	PH01000015G0220	*OsHSP17.7*[41]	Dought and heat tolerance	IPR002068 Heat shock protein Hsp20
	PH01000154G1240			
	PH01000268G0820			

Nature Genetics: doi:10.1038/ng.2569

	PH01000543G0160	*OsCIPK15*[42]	Defense response	
	PH01000906G0020	*Oshsp18.0-CII*[43]	response to heat stress, mechanical injury, and salicylic acid	
	PH01000943G0260			
	PH01000967G0270	*Oshsp26*[44]	Heat and oxidative tolerance	
	PH01001115G0640	*OsHSP17.7*[41]	Dought and heat tolerance	
	PH01001131G0040	*Oshsp18.0-CII*[43]	response to heat stress, mechanical injury, and salicylic acid	
	PH01003771G0070	*OsHSP17.0*[41]	Dought and heat tolerance	
	PH01004446G0090			
	PH01172955G0010	*Oshsp18.0-CII*[43]	response to heat stress, mechanical injury, and salicylic acid	
HSP70	PH01000101G0740		Abiotic stress tolerance	IPR001023 Heat shock protein Hsp70 | IPR018181 Heat shock protein 70, conserved site
	PH01000639G0270	*OsHSP71.1*[45]		
	PH01000974G0590	*OsHSP71.1*[45]		
	PH01001109G0400	*OsHSP71.1*[45]		
	PH01001215G0490			
	PH01001722G0380			
HSP DnaJ	PH01000430G0800			IPR001623 Heat shock protein DnaJ, N-terminal | IPR003095 Heat shock protein DnaJ | IPR018253 Heat shock protein DnaJ, conserved site
	PH01001102G0260			
	PH01003393G0080			
	PH01004637G0120			
	PH01035373G0010			
	PH01000298G0920			
	PH01000667G0420			

Nature Genetics: doi:10.1038/ng.2569

HSF	PH01000000G3800 PH01000081G0140 PH01000174G0590 PH01000194G0800 PH01000208G0690 PH01000314G0470 PH01000546G0840 PH01003169G0070 PH01002606G0040	*OsHSF7*[46] *OsHSF7*[46] *OsHsfA2c*[47] *OsHSF7*[46] *OsHsfB2b*[47] *OsHsfA2c*[47]	Heat and oxidative tolerance	IPR000232 Heat shock factor (HSF)-type, DNA-binding | IPR011991 Winged helix-turn-helix transcription repressor DNA-binding
peroxidase	PH01000207G0510 PH01004839G0040 PH01017692G0010	AT1G30870, AT1G44970[48]	Oxidative tolerance	IPR000823 Plant peroxidase | IPR019793 Peroxidases heam-ligand binding site | IPR019794 Peroxidase, active site
Dehydrin	PH01000279G0080 PH01001299G0140	AT3G50980[49]	Drought tolerance	IPR000167 Dehydrin
Thaumatin	PH01001035G0130 PH01001519G0090			IPR001938 Thaumatin, pathogenesis-related | IPR017949 Thaumatin, conserved site
HM	PH01000010G3140 PH01000689G0270 PH01002103G0360	AT4G16380[50]	Drought tolerance	IPR006121 Heavy metal transport/detoxification protein
MT	PH01001117G0170 PH01004749G0070 PH01001892G0210 PH01001967G0220	AT5G02380[51]	Defense response	IPR000347 Plant metallothionein, family 15

Nature Genetics: doi:10.1038/ng.2569

	PH01003107G0020			
BURP	PH01000333G0270 PH01000435G0330 PH01001626G0090 PH01001648G0010	AT5G25610[52]	Drought tolerance	IPR004873 BURP
MIP	PH01000021G0560 PH01001074G0010 PH01001117G0550	AT1G01620[53]	Drought tolerance	IPR000425 Major intrinsic protein

Note: description of the abbreviation was listed at the follows: ERF, ethylene-responsive transcriptional factor ; bZIP, Basic-leucine zipper (bZIP) transcription factor; CCT/B-box, CCT/B-box zinc finger protein; F-box, F-box domain containing protein; HTH myb-type, Helix-turn-helix transcriptional regulator, Myb-type; Homeobox, Homeobox domain containing protein; MADS-box, Transcription factor, MADS-box; NAC, NAC domain transcription factor; WD-40, WD-40 repeat family protein; YABBY, YABBY domain containing protein; zf-Dof, dof zinc finger domain containing protein. HSP20, Heat shock protein Hsp20; HSP70, Heat shock protein Hsp70; HSP DnaJ, Heat shock protein DnaJ; HSF, Heat shock factor (HSF)-type; Peroxidase, Plant peroxidase; Dehydrin, Dehydrin domain containing proteins; Thaumatin, Thaumatin domain containing proteins; HM, Heavy metal transport/detoxification protein); MT, Plant metallothionein, family 15; BURP, BURP domain containing proteins; MIP, Major intrinsic protein.

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 17 Insertion of TEs in homologs of *CO* and FPI genes.

Function	Known floral genes in rice and Arabidopsis†	Homologous loci in moso bamboo		Insertion of the repeats		Expression at flowering
		Scaffold	Position	TE type	Position‡	
Transcription factor *CONSTANS* (up-regulation of FPIs)	*CO* (At5g15840)[54] & *Hd-1* (LOC_Os06g16370)[55]	PH01000682	157487 - 159283	LINE/L1	3'-UTR, 390 bp from stop-codon	very low¶
		PH01005551	14180 - 15803			not detected
		PH01002508	56985 - 62300	LTR/Gypsy	intron, insertion >4,000 bp	not detected
FPIs (floral pathway integrators)	*FT* (At1g65480), *TSF* (At4g20370), TFL1 (At5g03840)[56], *RCN2* (LOC_Os02g32950)[57] & *Hd3a* (LOC_Os06g06320)	PH01000216	339337 - 340520	LINE/L1	promoter, 578 bp from start-codon	very low
		PH01002091	17501 - 18413	LINE/L1	3'-UTR, 456 bp from stop-codon	not detected
		PH01001461	145851 - 156173	LTR/Copia	coding region	not detected
		PH01000257	828385 - 839445	LTR/Gypsy & DNA/En-Spm	intron, insertion >8,000 bp	not detected
		PH01002149	213193 - 215000	DNA/TcMar-Stowaway	5'-UTR, 187 bp from start-codon	not detected
		PH01000020	1187666 - 1189438	Unclassified TE	promter, 661 bp form start-codon	very low
		PH01000089	982699 - 984616	Unclassified TE	coding region	not detected
		PH01000128	233584 - 235378	LTR/Gypsy	near 3'-UTR, 920 bp from	not detected

Nature Genetics: doi:10.1038/ng.2569

PH01000255	718847 - 720785	Unclassified TE	stop-codon promoter, <400 bp from start-codon	very low
PH01000191	326436 - 332802	DNA/hAT-Tag1 & unclassified TE	intron, insertion >5,000 bp	not detected
PH01000063	1107445 - 1113586	DNA/hAT-Tag1 & DNA/En-Spm	intron, insertion >5,000 bp	not detected
PH01000253	791493 - 793545	Unclassified TE	promoter, 240 bp from start-codon	very low
PH01002961	177515 - 178039	LTR/Copia	coding region	not detected
PH01001126	132864 - 133880	Unclassified TE	promter, 457 bp form start-codon	not detected
PH01002288	8168 - 9361	DNA/TcMar-Stowaway	coding region	not detected
PH01001134	281529 - 282416	DNA/MuDR	3'-UTR, 395 bp from stop-codon	not detected
PH01003363	146371 - 147247			very low
PH01002916	66198 - 67076	Unclassified TE	5'-UTR, 140 bp from start-codon	not detected
PH01004268	53808 - 55349	DNA/MuDR	promoter, 400 bp from start-codon	not detected
PH01003270	154451 - 153545			not detected
PH01001769	121285 - 125158	LTR/Gypsy & DNA/En-Spm	coding region	not detected
PH01001953	255266 - 256989	LINE/L1	promoter, 500 bp	very low

Nature Genetics: doi:10.1038/ng.2569

				from start-codon	
ATC (At2g27550)[56] & *BFT* (At5g62040)[58]	PH01000265	193381 - 195152	LINE/L1	promoter, 500 bp from start-codon	not detected
	PH01000019	26402 - 27560	LINE/L1	coding region	not detected
	PH01000354	535405 - 536549	DNA/MuDR	3'-UTR, 500 bp from stop-codon	very low
	PH01002570	19730 - 20888	Unclassified TE	promoter, 600 bp from start-codon	not detected
MFT (At1g18090)[59]	No homologs found				
TFL2 (At5g17690)[60]	No homologs found				
Ehd1 (LOC_Os10g32600)[22]	No homologs found				
OsSOC1 (LOC_Os03g03070)[21]	PH01000596	810562 - 811503	DNA/hAT-Ac	coding region, lost 2 exon	higher in leaf
	PH01000107	455178 - 459374	DNA/hAT-Ac	5'-UTR, 20 bp from start-codon	higher in shoot
	PH01000759	371677 - 372787	LTR/Copia	coding region, lost 2 exon	very low
	PH01002152	94108 - 96884	LTR/Gypsy	1st intron	higher in shoot and leaf
RFL (LOC_Os04g51000)[61]	PH01001425	66798 - 69307			very low
	PH01000386	696552 - 703389	DNA/hAT-Ac	coding region	not detected

† Abbreviations of gene description: *CONSTANS (CO), Heading date 1 (Hd-1), FLOWERING LOCUS T (FT), TWIN SISTER OF FT (TSF), BROTHER OF FT AND TFL2 (TFL1), Reduced Culm Number 2 (RCN2), Heading date-3a (Hd3a), Arabidopsis thaliana CENTRORADIALIS (ATC), BROTHER OF FT AND TFL1 (BFT), MOTHER OF FT AND TFL1 (MFT), BROTHER OF FT AND TFL2 (TFL2), Early heading date 1 (End1), rice LFY homolog (RFL),* and *SUPPRESSOR OF OVEREXPRESSION OF CO 1 (OsSOC1)*.

Nature Genetics: doi:10.1038/ng.2569

‡ For the position of repeat insertion, the 3'-UTR (3'-untranslated region), 5'-UTR, and promoter regions were estimated according to the analysis of the bamboo full-length cDNAs when the end of the gene model was not supported by the full-length cDNA. The estimated 3-UTR length was set at less than 500 bp from stop-codon, 5'-UTR less 250 bp from start-codon, and promoter region at over 250bp upstream of start-codon.

¶ Very low meant RPKM less than 5.

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 18 Quantified transcription levels of bamboo genes with homologs of Arabidopsis floral genes.

Identified genes in Arabidopsis	Abbreviation†	Function	Homologs in bamboo	RPKM						
				S20	S50	RH	RT	LF	P1	P2
AT4G24540, AT2G22540[62]	*AGL24, SVP*	Regulator of FPI/FMI	PH01000038G1550	9	9	4	21	2	3	2
			PH01000437G0930	72	71	52	36	26	11	10
AT2G27550[63], AT5G62040[58], AT5G03840[64]	*ATC, BFT, TFL1*	FPI	PH01001134G0390	0	0	0	0	0	0	0
			PH01002570G0010	4	2	0	0	0	0	0
			PH01003363G0220	0	1	0	0	0	0	0
AT5G06100[65]	*ATMYB33*	Regulator of FPI/FMI	PH01000029G1950	2	1	1	0	1	0	1
			PH01000009G0060	2	2	1	1	2	1	7
AT4G08920[66], AT1G04400[67]	*CRY1, CRY2*	Photoperiod pathway	PH01000263G1210	6	4	1	2	5	3	1
			PH01000968G0540	8	6	3	2	4	4	4
			PH01002304G0120	4	3	4	4	3	1	1
			PH01002373G0140	3	2	1	1	3	2	1
AT4G22140[68]	*EBS*	Regulator of FPI/FMI	PH01001266G0500	20	32	19	7	7	8	6
			PH01001406G0500	12	14	3	1	1	6	5
			PH01002328G0250	9	12	0	0	0	0	0
AT4G15880[69]	*ESD4*	Autonomous pathway	PH01000364G0790	2	3	2	1	2	1	1
			PH01000526G0230	8	12	9	7	7	5	7
			PH01001219G0240	6	7	10	5	7	4	6
AT1G68050[70]	*FKF1*	Photoperiod pathway	PH01002213G0250	9	9	2	2	3	3	5
			PH01002958G0010	11	9	6	16	7	10	22
			PH01007024G0030	15	15	6	6	9	6	7

Nature Genetics: doi:10.1038/ng.2569

			PH01000114G1110	11	12	6	4	6	5	5
			PH01000836G0340	9	7	3	9	6	12	16
AT3G04610[71]	*FLK*	Autonomous pathway	PH01000025G1210	11	10	9	9	3	7	8
			PH01000171G0620	158	175	157	133	76	76	75
			PH01000280G1350	3	1	5	2	4	3	6
			PH01001197G0230	3	4	4	3	1	1	1
AT5G24860[72]	*FPF1*	Gibberellin pathway	PH01000002G4150	2	1	6	2	9	15	18
			PH01001317G0310	0	0	0	2	5	1	2
			PH01002809G0210	13	10	10	16	0	18	47
AT4G00650[73]	*FRI*	Autonomous pathway	PH01000371G0250	2	1	3	1	2	2	4
AT1G65480[74], AT4G20370[75]	*FT,TSF*	FPI	PH01002288G0050	0	0	0	0	1	0	0
			PH01000020G1780	0	0	0	0	1	1	1
AT2G19520[76]	*FVE*	Autonomous pathway	PH01000048G0850	42	35	37	17	21	14	10
			PH01000241G0710	56	52	93	38	57	39	28
AT4G02780[77]	*GA1*	Gibberellin pathway	PH01000557G0660	4	3	2	2	3	6	4
AT1G14920, AT2G01570, AT1G66350, AT3G03450	*GAI,RGA,RGL*[78-80]	Gibberellin pathway	PH01004823G0070	3	1	2	1	1	1	2
			PH01143550G0010	0	0	0	0	0	0	0
			PH01190367G0010	25	33	9	10	3	4	9
			PH01000142G0910	27	40	92	60	19	33	65
			PH01000254G0530	0	0	0	0	0	0	0
AT3G05120, AT5G27320	*Gar*[81]	Gibberellin pathway	PH01001316G0350	4	3	3	2	6	8	10
			PH01002734G0310	3	3	1	0	3	2	4
			PH01000068G0090	8	6	65	6	2	2	4
			PH01000068G0120	0	0	3	1	0	0	1

Nature Genetics: doi:10.1038/ng.2569

AT2G39810[82]	*HOS1*	Gibberellin pathway	PH01000750G0240	22	20	12	11	9	7	6
AT4G02560[83]	*LD*	Autonomous pathway	PH01006816G0010	6	7	6	7	9	7	5
AT1G25540[84]	*PFT1*	Photoperiod pathway	PH01001819G0320	24	26	19	17	14	13	8
			PH01002482G0220	22	24	22	25	19	16	15
AT1G09570, AT2G18790, AT4G16250, AT4G18130	*PHY*[85]	Light-quality pathway	PH01000013G2230	11	11	4	1	3	1	2
			PH01000013G2240	8	11	9	3	2	3	2
			PH01000222G1330	11	13	4	3	3	3	3
			PH01000606G0390	17	20	2	1	1	2	1
AT3G12810[86]	*PIE1*	Autonomous pathway	PH01000672G0430	11	12	4	4	4	4	2
			PH01001540G0210	7	6	2	1	1	1	1
AT4G24210[87]	*SLY1*	Gibberellin pathway	PH01000146G1260	11	10	9	5	7	5	7
AT2G45660[88]	*SOC1*	FPI	PH01000616G0020	1	1	0	1	2	1	1
			PH01000759G0450	1	0	2	2	3	1	1
			PH01002152G0120	22	27	5	54	210	9	12
			PH01000059G1270	0	1	0	1	5	1	0
AT3G11540[89]	*SPY*	Gibberellin pathway	PH01000299G0650	24	23	25	12	14	18	15
			PH01003018G0160	11	10	9	3	8	7	6
AT5G57380[90], AT3G24440[91]	*VIN3,VIN3-L*	Ambient-temperature pathway	PH01000836G0140	20	18	17	16	22	23	24
			PH01001556G0190	4	5	1	0	1	1	0
			PH01000006G3670	4	4	1	1	1	1	1
			PH01000258G0590	5	5	6	3	6	4	3
			PH01000674G0720	7	7	1	1	0	2	1

Nature Genetics: doi:10.1038/ng.2569

† Abbreviations of the genes: *AGAMOUS-LIKE 24 (AGL24), SHORT VEGETATIVE PHASE (SVP), Arabidopsis thaliana CENTRORADIALIS (ATC), BROTHER OF FT AND TFL1 (BFT), TERMINAL FLOWER 1 (TFL1), MYB DOMAIN PROTEIN 33 (ATMYB33), CRYPTOCHROME 1 (CRY1), CRYPTOCHROME 2 (CRY2), EARLY BOLTING IN SHORT DAYS (EBS), EARLY IN SHORT DAYS 4 (ESD4), FLAVIN-BINDING KELCH DOMAIN F BOX PROTEIN 1 (FKF1), FLOWERING LOCUS KH DOMAIN (FLK), FLOWERING PROMOTING FACTOR 1 (FPF1), FRIGIDA (FRI), FLOWERING LOCUS T (FT), FVE (FVE), GA REQUIRING 1 (GA1), GA INSENSITIVE (GAI), REPRESSOR OF GA1-3 (RGA), RGA-LIKE (RGL), GAI AN REVERTANT (Gar), HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 1 (HOS1), LUMINIDEPENDENS (LD), PHYTOCHROME AND FLOWERING TIME 1 (PFT1), PHYTOCHROME (PHY), PHOTOPERIOD-INDEPENDENT EARLY FLOWERING (PIE), SLEEPY 1 (SLY1), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), SPINDLY (SPY), VERNALIZATION INSENSITIVE 3 (VIN3),* and *VERNALIZATION INSENSITIVE 3-LIKE 1 (VIN3-L)*.

Nature Genetics: doi:10.1038/ng.2569

Supplementary Table 19 Bamboo floral genes sharing high identities (> 50%) with known rice genes.

Bamboo gene ID	Homologous rice gene	Description of rice gene	Identities of amino acids	Function	Involved pathway†
PH01000015G0220	LOC_Os01g04380	*OsHSP17.0*[45]	0.90	Stress tolerance	ABA
PH01000032G1740	LOC_Os01g58420	*OsERF*[44]	0.66	Drought tolerance	ETH
PH01000053G1650	LOC_Os11g03300	*OsNAC10*[37]	0.64	Drought tolerance	ABA
PH01000058G1180	LOC_Os01g72530	*OsCML31*[92]	0.75	Signal transduction	Ca^{2+} sensor
PH01000068G0660	LOC_Os07g44330	*OsPDK1*[93]	0.92	Signal transduction	GA
PH01000074G0590	LOC_Os02g44235	*OsTPP1*[94]	0.51	Stress tolerance	ABA
PH01000081G0140	LOC_Os03g06630	*OsHSF7*[46]	0.82	Stress tolerance	
PH01000091G0440	LOC_Os03g26870	*SRWD3*[39]	0.80	Salinity tolerance	
PH01000099G1710	LOC_Os01g42860	*OCPI1*[95]	0.66	Drought tolerance	
PH01000111G0850	LOC_Os01g66120	*OsNAC6 ; SNAC2*[96]	0.82	Stress tolerance	ABA
PH01000113G0300	LOC_Os03g07360	*OsDof12* [23,24]	0.80	Regulator of FMI	Flowering
PH01000122G1000	LOC_Os03g60080	*SNAC1*[38]	0.70	Drought tolerance	
PH01000154G1240	LOC_Os03g16040	*OsHSP17.7*[41]	0.54	Drought and heat stress tolerance	
PH01000162G1010	LOC_Os03g44710	*OsYABBY2; OsYAB2*[40]	0.71	FMI	
PH01000173G1010	LOC_Os02g52780	*OsbZIP23*[35]	0.59	Drought and salinity tolerance	ABA
PH01000174G0590	LOC_Os07g08140	*OsHsfA2b*[97]	0.62	Heat stress tolerance	
PH01000192G1330	LOC_Os04g41540	*OsCML22*[92]	0.54	Signal transduction	Ca^{2+} sensor
PH01000194G0800	LOC_Os10g28340	*OsHsfA2c*[47]	0.80	Heat and oxidative stress tolerance	
PH01000208G0690	LOC_Os03g06630	*OsHSF7*[46]	0.78	Heat stress tolerance	
PH01000222G1190	LOC_Os03g54160	*OsMADS14*	0.69	FMI	Flowering
PH01000242G0910	LOC_Os05g49420	*OsbZIP45*[36]	0.78	Reproductive development and stress	ABA

Nature Genetics: doi:10.1038/ng.2569

				tolerance	
PH01000280G1220	LOC_Os02g34560	*SRT-5*[98]	0.61	Hydrolysis of sucrose	
PH01000284G0720	LOC_Os05g44340	*OsClpB-cyt; HSP100*[47]	0.87	Heat stress tolerance	
PH01000286G0770	LOC_Os02g56460	*OsCCR1*[99]	0.57	Defense and cell wall biosynthesis	Lignin biosynthesis
PH01000298G0570	LOC_Os05g48930	*OsGRX17*[100]	0.57	Glutaredoxin	
PH01000306G0610	LOC_Os03g54160	*OsMADS14*	0.57	FMI	Flowering
PH01000309G1110	LOC_Os10g36650	*OsActin*[101]	0.93	Salinity tolerance	ABA
PH01000314G0470	LOC_Os08g43334	*OsHsfB2b*[97]	0.54	Heat and oxidative stress tolerance	
PH01000317G0080	LOC_Os01g66030	*OsMADS2*	0.92	FMI	Flowering
PH01000344G0730	LOC_Os03g04680	*OsCYP96B4*[102]	0.74	Cytochrome P450	Lipid metabolism
PH01000356G0210	LOC_Os10g33250	*Wda1*[103]	0.53	Pollen Development	
PH01000416G0840	LOC_Os03g22120	*SUS4*[104]	0.95	Sucrose synthase	Sugar metabolism
PH01000437G0190	LOC_Os02g52780	*OsbZIP23*[35]	0.59	Salinity and drought tolerance	ABA
PH01000491G0570	LOC_Os01g09620	*OsDOS*[105]	0.72	Delaying Leaf Senescence	
PH01000534G0160	LOC_Os11g02240	*OsCIPK15*[42]	0.51	Signal transduction	Ca^{2+} sensor
PH01000546G0830	LOC_Os10g28340	*OsHsfA2c*[47]	0.83	Heat stress tolerance	
PH01000548G0440	LOC_Os11g35710	*OsOSC11; OsIAS1*	0.78	Isoarborinol synthase and defense	
PH01000573G0640	LOC_Os04g52090	*OsAP2-39*[44]	0.53	Drought tolerance	ABA/GA
PH01000595G0160	LOC_Os05g04700	*OsLti6b*[106]	0.54	Low temperature stress tolerance	ABA
PH01000606G0250	LOC_Os03g54160	*OsMADS14*	0.66	FMI	Flowering
PH01000639G0270	LOC_Os03g16860	*OsHSP71.1*[45]	0.57	Stress tolerance	ABA
PH01000654G0080	LOC_Os03g31300	*OsClpB-c*[107]	0.73	Stress tolerance	
PH01000664G0490	LOC_Os03g50885	*OsActin*[108]	0.88	Drought tolerance	ABA

Nature Genetics: doi:10.1038/ng.2569

PH01000794G0630	LOC_Os01g66120	*OsNAC6 ; SNAC2*[96]	0.73	Stress tolerance	ABA
PH01000845G0420	LOC_Os04g59440	*psbS2*[109]	0.67	Absorbed Light Energy	PHOTOSYSTEM
PH01000877G0160	LOC_Os08g01330	*OsSWN3*[110]	0.52	Secondary Wall Biosynthesis	
PH01000906G0020	LOC_Os01g08860	*Oshsp18.0-CII*[43]	0.67	Stress tolerance	
PH01000967G0270	LOC_Os03g14180	*Oshsp26*[44]	0.51	Heat and oxidative stress tolerance	
PH01000974G0590	LOC_Os03g16860	*OsHSP71.1*[45]	0.90	Stress tolerance	ABA
PH01001038G0420	LOC_Os03g60560	*ZFP182*[111]	0.65	Salinity and drought tolerance	
PH01001109G0400	LOC_Os03g16860	*OsHSP71.1*[45]	0.66	Stress tolerance	ABA
PH01001115G0640	LOC_Os03g16040	*OsHSP17.7*[41]	0.80	Drought and heat stress tolerance	
PH01001131G0040	LOC_Os01g08860	*Oshsp18.0-CII*[43]	0.71	Stress tolerance	
PH01001298G0250	LOC_Os07g08840	*OsTrx23,OsTRXh1*[112]	0.64	Oxidative stress tolerance	ABA
PH01001464G0430	LOC_Os02g28980	*rFKBP75*[113]	0.79	heat stress tolerance and seed development	
PH01001546G0030	LOC_Os07g08840	*OsTrx23,OsTRXh1*[112]	0.74	Oxidative stress tolerance	ABA
PH01001726G0230	LOC_Os05g44340	*OsClpB-cyt; HSP100*[47]	0.92	Heat stress tolerance	
PH01001843G0210	LOC_Os03g60080	*SNAC1*[38]	0.50	Drought tolerance	
PH01001888G0390	LOC_Os01g71670	*OsGLN2*[114]	0.53	Rice endo-1,3-β-glucanase	ABA/GA
PH01001952G0190	LOC_Os03g11614	*OsMADS1*[115]*; LHS1; AFO*	0.54	FMI	Flowering
PH01002127G0260	LOC_Os01g10504	*OSMADS3*[116]	0.79	FMI	Flowering
PH01002642G0160	LOC_Os02g28980	*rFKBP75*[113]	0.55	heat stress tolerance and seed development	
PH01003526G0090	LOC_Os03g26870	*SRWD5*[39]	0.73	Salinity tolerance	
PH01003638G0080	LOC_Os04g59550	*OsUGT1*[117]	0.83	Sugar transporters	Sugar metabolism

Nature Genetics: doi:10.1038/ng.2569

PH01003771G0070	LOC_Os01g04380	*OsHSP17.0*[41]	0.89	Drought and heat stress tolerance	ABA
PH01004006G0090	LOC_Os11g03300	*OsNAC10*[37]	0.53	Drought tolerance	ABA
PH01005579G0050	LOC_Os03g49380	*OsLOX1*[118]	0.55	Stress tolerance	JA
PH01005903G0030	LOC_Os02g08230	*OsGL1-2*[119]	0.82	Drought tolerance	
PH01006818G0010	LOC_Os10g33250	*Wda1*[103]	0.64	Pollen development	
PH01172955G0010	LOC_Os01g08860	*Oshsp18.0-CII*[43]	0.84	Stress tolerance	

† Abbreviation of the pathways: Floral pathways of FMI or FPI (Flowering), abscisic acid pathway (ABA), Gibberellin pathway (GA), ethylene-responsive pathway (ETH), jasmonic acid pathway (JA).

Nature Genetics: doi:10.1038/ng.2569