JIPB Journal of Integrative Plant Biology

Journal of Integrative Plant Biology 2010, **52** (11): 1008–1015

Research Article

Insights into the Bamboo Genome: Syntenic Relationships to Rice and Sorghum

Yi-Jie Gui[1†], Yan Zhou[2,4†], Yu Wang[1†], Sheng Wang[3], Sheng-Yue Wang[4], Yan Hu[5], Shi-Ping Bo[1], Huan Chen[1], Chang-Ping Zhou[6], Nai-Xun Ma[7], Tian-Zhen Zhang[5] and Long-Jiang Fan[1*]

[1]Institute of Crop Science & Institute of Bioinformatics, Zhejiang University, Hangzhou 310029, China
[2]State Key Laboratory of Genetic Engineering, Department of Microbiology, School of Life Sciences, Fudan University, Shanghai 200032, China
[3]Hangzhou Dianzi University, Xiasha, Hangzhou 310018, China
[4]MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, China
[5]National Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agriculture University, Nanjing 210095, China
[6]Anji Bamboo Exposition Garden, Anji 313300, China
[7]The Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Fuyang 311400, China
[†]These authors contributed equally to this work.
[*]Corresponding author
Tel: +86 571 8697 1730; Fax: +86 571 8697 1117; E-mail: fanlj@zju.edu.cn
Available online on 14 July 2010 at www.jipb.net and www.wileyonlinelibrary.com/journal/jipb
doi: 10.1111/j.1744-7909.2010.00965.x

Abstract

Bamboo occupies an important phylogenetic node in the grass family and plays a significant role in the forest industry. We produced 1.2 Mb of tetraploid moso bamboo (*Phyllostachys pubescens* E. Mazel ex H. de Leh.) sequences from 13 bacterial artificial chromosome (BAC) clones, and these are the largest genomic sequences available so far from the subfamily Bambusoideae. The content of repetitive elements (36.2%) in bamboo is similar to that in rice. Both rice and sorghum exhibit high genomic synteny with bamboo, which suggests that rice and sorghum may be useful as models for decoding Bambusoideae genomes.

Gui YJ, Zhou Y, Wang Y, Wang S, Wang SY, Hu Y, Bo SP, Chen H, Zhou CP, Ma NX, Zhang TZ, Fan LJ (2010) Insights into the bamboo genome: syntenic relationships to rice and sorghum. *J. Integr. Plant Biol*. **52**(11), 1008–1015.

Introduction

Bamboos are a group of woody and herbaceous plants that have many food, garden and industrial applications. It is estimated that 18 million hectares of bamboo are distributed worldwide, with approximately 3 and 8 million hectares in China and India, respectively (Gielis et al. 1996). Bamboos belong to the subfamily Bambusoideae and are woody members of the grass family Poaceae, which also includes rice, sorghum and other cereals. However, just like cereals but not woody plants or trees, bamboos only flower once. Therefore bamboos present a unique group of the grass family. Bamboos serve essential functions in foods, timber, papermaking, biofuel, etc.

Bamboos are thought to have originated approximately 30–40 million years ago and evolved in forests from an ancestral grass that closely resembled *Streptochaeta* Schrad (Clark 1996; Klinkenborg 2001). Bamboos probably have the same basic chromosome number as rice ($x = 12$), and tropical and temperate bamboos are hexaploids ($2n = 72$) and tetraploids ($2n = 48$), respectively, while rice is diploid ($2n = 24$). The genome sizes of bamboos are estimated to be 2.45–5.30 pg (Gielis et al. 1996; Gui et al. 2007), which is generally within the mid-range of size in the grass family (Gaut 2002). The Bambusoideae include over 70 genera and 1 200 species (Jiang 2002) and economically important species are found in several genera or sections. For example, the genus *Phyllostachys*, which includes temperate moso and timber bamboos, contains

the ecologically and economically most important member of the Bambusoideae.

Considering the high economic importance of bamboo in rural economies and industrial applications, genetic and genomic analyses of bamboo need to be significantly advanced. While genome sequences are available for several cereals, there has been no significant progress in the genomic investigation of bamboo except for a recent genomic survey (Gui et al. 2007). In that study, almost 1 000 genomic survey sequences (GSS) were generated from moso bamboo (*P. pubescens*), the most widely distributed and economically important species in China. In GenBank, only 17 789 nucleotide sequences (less than 0.1% of the total sequences from Poaceae) have currently been determined from Bambusoideae (as of 11 November 2009, including expressed sequence tags (EST) and GSS).

Although massive genomic sequences have been determined in the grass family, many gaps still remain in our understanding of the variation across different species in this family, in which bamboos represent an important branch (Paterson et al. 2005). In the present study, we produced 1.2 Mb of *Phyllostachys* sequences from 13 bacterial artificial chromosome (BAC) clones, annotated them and compared them with orthologous regions in the rice and sorghum genomes.

Results and Discussion

Cloning and sequencing of bamboo BACs

Three hundred and eighty four BAC clones were constructed according to the procedure described by Zhang (2000) using shoots of moso bamboo collected from Anji Bamboo Exposition Garden, China. Thirteen BAC clones with insertion sizes of over 100 Kb were selected to be further sequenced. Two clones were finished by a traditional sequencing method using ABI 3730 sequencers while the other 11 clones plus one of the finished BACs were pooled together and sequenced using a massively parallel pyrosequencing technology (454 GS FLX). The two finished BACs (GQ252886 and GQ252887) were 113.2 and 139.3 Kb in length, respectively (**Table 1**). A total of 142 775 reads by the 454 technology, with an average size of 220 bp and a total length of 31 470 115 bp, provided about 22× coverage over the target BACs (120.9 Kb per clone estimated based on Pulsed Field Gel Electrophoresis). These reads were finally assembled into 90 contigs (>2.0 Kb) spanning 877.27 Kb. The average size of the contigs was 9.75 Kb, and 32 and three contigs were over 10 Kb and 30 Kb in length, respectively. All contig sequences from the present study were deposited in GenBank under accession numbers GQ252796-GQ252887.

Table 1. Summary of bacterial artificial chromosome (BAC) sequencing

BAC number	Stage	Contig numbers	Average size (kb)	Total size (kb)	Method
2	Finished	2	126.3	252.5	Sanger
11	Draft	90	9.8	877.3	Pyrosequencing

Gene annotation

In silico gene prediction was carried out using FGENESH (http://www.softberry.com) with monocot parameters. In addition to transposable elements (TEs), a total of 112 non-TE-related protein-coding genes were predicted and 46% of them showed evidence of transcriptional sequences or were highly similar (BLASTP E-value less than 1e-10) to genes from other plants. Among these 112 genes, we found some bona fide genes, such as four resistance gene homologs (*RGH* or *RGA*), one putative alcohol dehydrogenase (*Adh*) gene that was highly similar to rice *Adh3*, and two pollen-specific kinase genes, which provide clues for further genetic analysis in bamboos. Some of the predicted bamboo genes currently have no hits in GenBank. We selected five genes (Genes 3, 4, 6, 8 at GQ252886 and Gene 11 at GQ252805; **Figure 1**) to validate the predictions. Five primer pairs were designed based on the coding regions of the five genes, respectively, and amplified in the cDNA pool extracted from bamboo leaves or roots. Single expected bands were observed in the five primers, respectively (data not shown), which suggests that these genes are active and imply that they may be bamboo-specific.

The features of the 112 predicted bamboo genes in comparison to rice and sorghum are shown in **Table 2**. The average gene size of bamboo seems to be shorter than those of the other two cereals. This result needs to be further confirmed based on more bamboo genomic sequences or full-length cDNA sequences. Longer genes may be incomplete or unannotated in our bamboo sequences due to their incompleteness, causing underestimation of the mean gene length. In the annotation of the sorghum genome, 5 197 gene models, usually short (often < 150 amino acids) and with no EST support, were believed to be of low confidence and therefore were not included in the gene dataset (Paterson et al. 2009). If we exclude predicted genes with less than 150 amino acids from our dataset, the average size of bamboo genes increased to 2 370 bp. At the genomic level, the average G+C content of all bamboo contigs was 44.3%, which has no significant difference with those of rice (44.2%, Feng et al. 2002) and sorghum (44.4%, Paterson et al. 2009).

Repetitive elements

Previous studies have suggested that long terminal repeats (LTR) retrotransposons may be predominant among repetitive

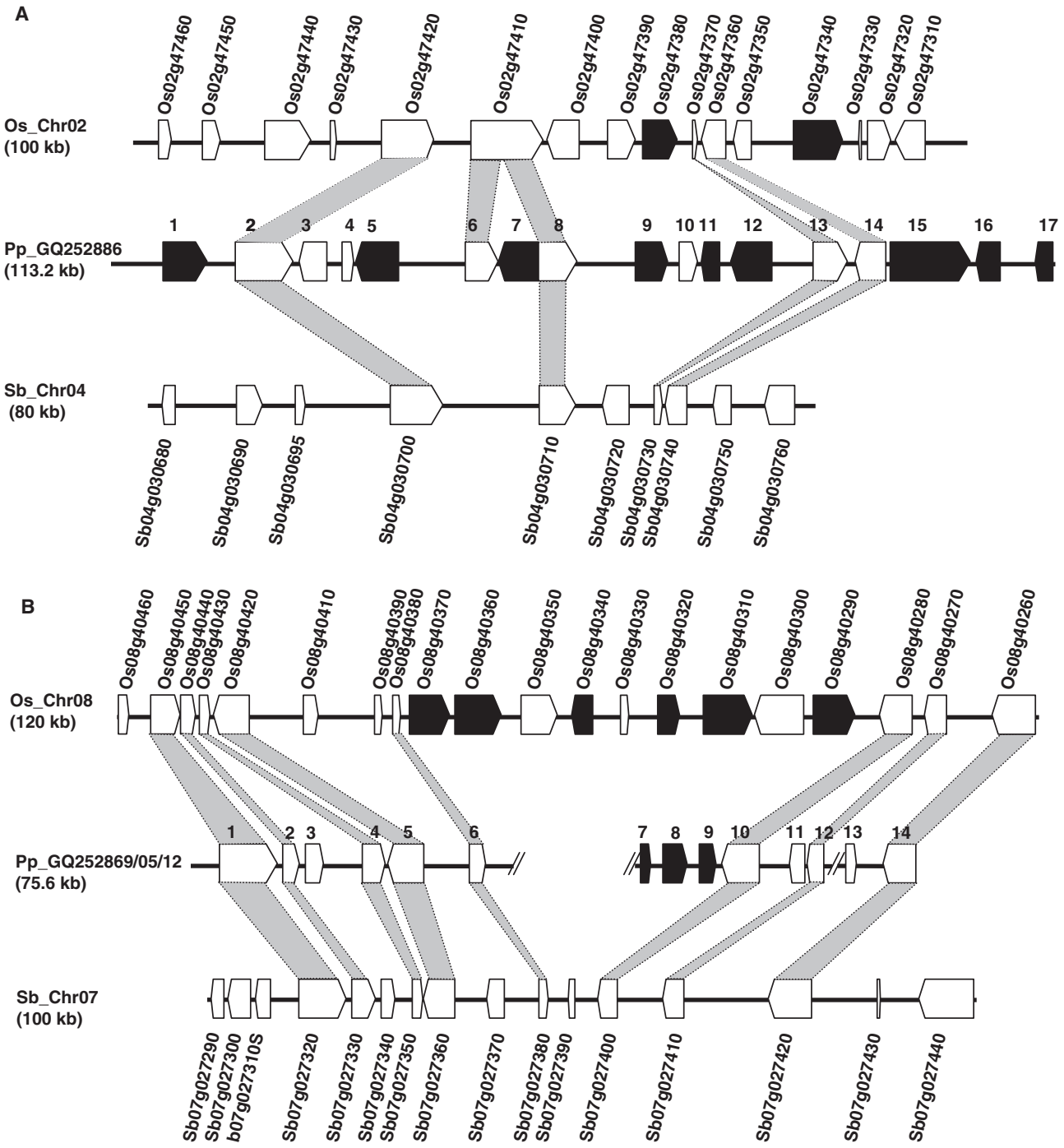

Figure 1. Two bamboo-rice-sorghum syntenic regions.

(A) A finished bacterial artificial chromosome (BAC) sequence (GQ252886).

(B) A putative scaffold including three bamboo contigs (from right to left: GQ252869, GQ252805 and GQ252812).

Predicted genes and their orientation are shown as boxed areas. Putative retrotransposons are shown in black. Conserved genes between bamboo and rice/sorghum regions are connected by shaded areas. Annotations of rice and sorghum are based on the MSU Rice Genome Annotation Project at http://rice.plantbiology.msu.edu/ and the DOE JGI sorghum genome project at http://www.jgi.doe.gov. Os, rice; Pp, bamboo; Sb, sorghum.

Table 2. Features of bamboo genes in comparison with those of rice and sorghum

	Bamboo[a]	Rice[b]	Sorghum[c]
Average gene size (bp)	1 862	2 844	2 794
Average exon length (bp)	258	313	279
Average intron length (bp)	388	414	456
Average exon/gene	3.5	4.9	4.4
Average intron/gene	2.5	3.9	3.4
Average exon GC content (%)	51.0	51.5	NA
Average intron GC content (%)	40.0	37.6	NA

[a]Based on the 112 annotated bamboo genes in this study; [b]From the MSU Rice Genome Annotation Project at http://rice.plantbiology.msu.edu; [c]From the DOE JGI at http://genome.jgi-psf.org/Sorbi1/. NA, not available.

elements of tetraploid moso bamboo (Gui et al. 2007). The large genomic segments generated in the present study make it possible for us to re-evaluate this point. The retroelements were annotated using gene and LTR prediction programs, which combined repeated scanning against custom repeat databases (see Materials and Methods). In the two finished BAC clones, high densities of retrotransposons were found (**Figures 1**, **2**). In GQ252887, for example, retrotransposons comprised most of the repetitive elements (**Figure 2**; detailed list in Table S1). For these LTR retrotransposons, their insertion times were estimated based on solo-LTR sequences (SanMiguel et al. 1998). Most of them seem to have been inserted recently (<2 million years ago).

We further estimated the amount and composition of repetitive elements in the bamboo genome using all large contigs (>10 Kb). Following the sorghum repeat annotation pipeline with minor changes, we annotated the bamboo contigs and compared our results to those of three cereals annotated by Paterson et al. (2009) (**Table 3**). Consistent with an estimation in a previous study (Gui et al. 2007), the repetitive content (36.2%) of bamboo is not significant different ($P = 0.32$) to that of rice (39.5%). However, it should be noted that several factors would lead to underestimation of proportionate content in the bamboo genome. First, our dataset is not big, and additional repeats with low copy numbers (such as less than 1 000 in the genome) will be found as more sequence data are available. Therefore more genomic sequences are necessary to confirm this estimation. Second, our contigs are not big (9.75 Kb average size) and an underestimation of repetitive elements might be caused by contigs generated by the 454 approach, which gives short reads that are generally difficult to assemble for repetitive regions. Finally, the repetitive contents of the three cereals are much better known because their genomes have been sequenced extensively and annotated carefully for TEs.

There is a slight difference in the composition of repetitive elements between bamboo and rice. The content of DNA transposons in the bamboo genome (6.14%) was significantly ($P = 0.02$) lower than that in rice (13.7%). In rice, the contents of CACTA superfamily (3.43%) and miniature inverted-repeat transposable elements (MITEs) (5.24%) of DNA transposons seem to have increased (but $P = 0.16$ and 0.21) during evolution of the rice lineage relative to that in bamboo (1.59 and 2.24%). A similar trend was also observed in the other two cereals (except for the CACTA family in sorghum) (Paterson et al. 2009).

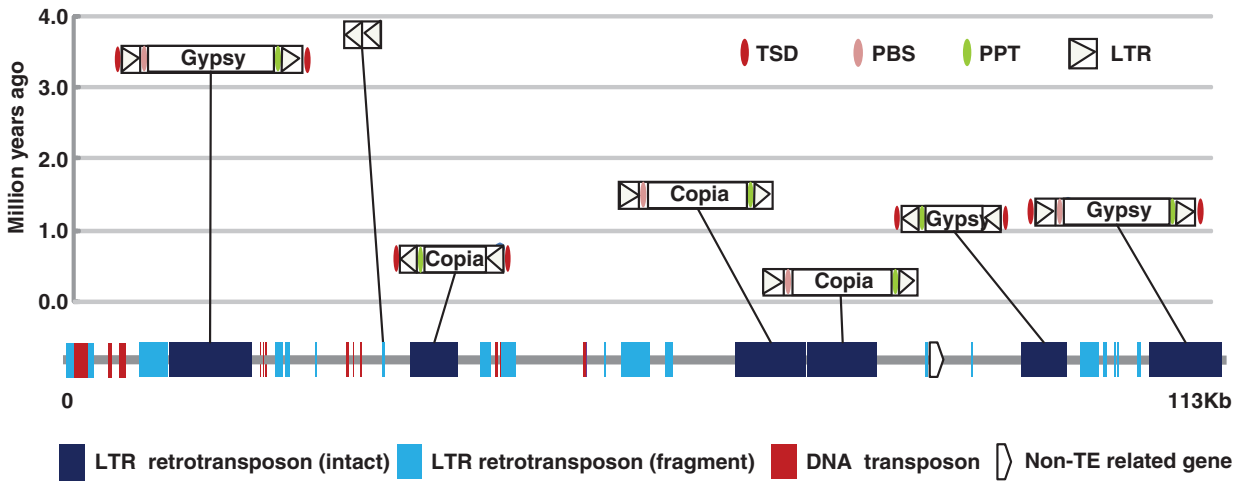

Figure 2. Repetitive elements in a bamboo bacterial artificial chromosome (BAC) sequence (GQ252887).

The potential insertion times and types of intact LTR retrotransposon are shown. LTR, long terminal repeats; PBS, primer binding site; PPT, polypurine tract; TE, transposable element; TSD, target site duplication.

Table 3. Repeat composition and major components of the bamboo genome in comparison to rice, sorghum and maize

Repetitive elements	Pp	Os	Sb	Zm
Transposon DNA	36.23	39.50	62.00	82.10
Class I: Retroelement	30.09	25.78	54.52	79.44
LTR Retrotransposon	29.99	23.47	54.43	75.08
Ty1/copia	13.90	2.47	5.18	21.75
Ty3/gypsy	13.07	12.03	19.00	37.73
Unclassified LTR	3.03	8.98	30.25	15.59
Non-LTR Retrotransposon	0.10	1.24	0.04	0.35
LINE	0.10	0.80	0.04	0.34
SINE	0.00	0.45	0.00	0.01
Unclassified retroelement	NA	1.07	0.05	4.02
Class II: DNA Transposon	6.14	13.67	7.46	2.68
DNA Transposon Superfamily	2.83	7.04	4.79	0.92
CACTA superfamily	1.58	3.43	4.69	0.47
hAT superfamily	0.61	0.52	0.02	0.10
Mutator superfamily	0.59	1.81	0.06	0.15
Tc1/Mariner superfamily	0.00	0.02	0.00	0.00
PIF/Harbinger	0.00	0.00	0.02	0.08
Unclassified	0.05	1.26	0.00	0.12
MITE	2.24	5.24	1.74	0.32
Stowaway	1.94	1.74	0.19	0.03
Tourist	0.30	1.50	0.94	0.08
Unclassified MITE	NA	2.00	0.61	0.21
Helitron	0.98	0.33	0.81	1.31
Unclassified DNA transposon	0.08	1.06	0.12	0.12

The estimations (percentage of genome bp) were based on 10 Kb genomic contigs generated in this study for the bamboo (Pp) genome and Paterson et al. (2009) for three cereals (Os, rice; Sb, sorghum; Zm, maize). LINE, long interspersed nuclear element; LTR, long terminal repeat; MITE, miniature inverted-repeat transposable element; PIF, P instability factor; SINE, short interspersed nuclear element.

Synteny among bamboo, rice and sorghum

Extensive genomic colinearity has been shown within grasses. The availability of the genome sequences of rice (International Rice Genome Sequencing Project 2005; Yu et al. 2005), sorghum (Paterson et al. 2009) and segments of the bamboo genome provide an opportunity to estimate the synteny between bamboo and cereals. High syntenies were observed among bamboo, rice and sorghum (two examples are shown in **Figure 1**). The finished BAC sequence GQ252886 showed extensive synteny with two orthologous regions at rice chromosome 2 and sorghum chromosome 4, respectively, even to the point where many retrotransposons are inserted in GQ252886 (**Figure 1A**). Interestingly, a retrotransposon inserted into intron 11 of a putative protein kinase gene, which was split into two putative genes (Genes 6 and 8 at **Figure 1A**) in bamboo. The transcription sequences of Genes 6 and 8 have been confirmed by us (GQ273506 and GQ273507), and the results suggest that these genes are transcribed. In sorghum, only the 3′ region (Sb04g030710) of the putative kinase gene was retained and its 5′ region seems to be pseudo due to indels/mutations, although it still shows high similarity to the bamboo Gene 6 (data not shown). However, the putative kinase gene was kept intact in rice (Os02g47410, 8 675 bp in length) and its transcriptal sequence can be found (see the MSU Rice Genome Annotation Project at http://rice.plantbiology.msu.edu).

Another example shows a putative scaffold containing three contigs (**Figure 1B**). The three contigs (GQ252869, GQ252805 and GQ252812) were confirmed from the same BAC clone through polymerase chain reaction (PCR) amplification. Just like the above protein kinase gene in bamboo, we also found a ~4 Kb potential insertion sequence in the bamboo syntenic region (Gene 1 in the contig of GQ252869) of genes from rice (Os08g40450) and sorghum (Sb07g027320) (**Figure 1B**), and thus two genes were predicted from this region by several gene-prediction programs. The 4 Kb non-repetitive sequence is located in a putative intronic region based on the two orthologs from rice and sorghum. To test whether the gene, as with the above protein kinase gene, has indeed been split into two genes, a primer pair located on two predicted exons that covered the 4 Kb sequences was designed. We successfully amplified and sequenced a transcript (GQ273508) from the bamboo cDNA pool. Sequence alignment indicated that the bamboo ortholog (Gene 1) had the same gene structure as that of Os08g40450 and Sb07g027320, and the 4 Kb sequence simply added the length of the intronic sequence of its gene.

The comparison of the two othologous regions also demonstrated that some bamboo genes seem to have been lost or moved to other genomic regions after the divergence of bamboo from other members. For example, the pectinacetylesterase domain containing gene (Os02g47400 and Sb04g030720) could not be found in the syntenic region of bamboo (**Figure 1A**). Meanwhile, some bamboo genes (such as Genes 3, 4 and 10 at GQ252886; Genes 3 at GQ252869 and Gene 11 at GQ252805; **Figure 1**) have no hits to orthologous or even other regions of rice and sorghum, suggesting they might be bamboo-specific genes. These results reflect a dynamic evolutionary process of the bamboo genome.

The high genomic synteny between bamboo and rice/sorghum suggests that rice and sorghum may be very useful as models for decoding Bambusoideae genomes. It is difficult to construct a bamboo genetic population and genetic map since we can not control the flowering of bamboos, and the comparative genomic approach is therefore more important for studies on the bamboo genome. Rice has one of the smallest genomes and was the first member of the grass family to be sequenced. Since rice is relatively close to bamboo

in phylogeny (Kellogg 1998; Gaut 2002; Also our results at Figure S1 based on 28 annotated bamboo genes), it should be the best available reference genome for the comparative genomic analysis with bamboo. Meanwhile, sorghum also shows high genomic synteny with bamboo although it is relatively distant from bamboo in phylogeny. The recent publication of high-quality genome sequences of sorghum (Paterson et al. 2009) thus provides another important reference genome for bamboo.

Materials and Methods

Plant samples

Fresh shoots and leaves of moso bamboo (*P. pubescens* E. Mazel ex H. de Leh.) were collected from Anji Bamboo Exposition Garden and Hangzhou Botany Garden, China for the construction of BAC clones and total RNA extraction, respectively.

Construction of BAC clones

Bamboo BAC clones were constructed according to a protocol developed by Zhang (Zhang 2000) with minor modifications: (i) to avoid organelle contamination in the nuclei preparations, nuclei were isolated from bamboo shoot; (ii) BAC clones were constructed in the *Hin*dIII site of the vector pIndigoBAC-5 (*Hin*dIII-cloning Ready, Epicentre Technologies, Madison, WI, USA); (iii) ligation was transformed into ElectroMAX DH10B competent cells (Invitrogen, Carlsbad, CA, USA).

BAC sequencing

The DNA of 12 BAC clones were equally pooled together and sequenced by using a massively parallel pyrosequencing technology (454 GS FLX) (Margulies et al. 2005). The gaps were closed by multiplex PCR (Tettelin et al. 1999). An ABi 3730xl capillary sequencer was used to determine the end sequences of each BAC clone. All contig sequences from this study were deposited in GenBank under accession numbers GQ252796-GQ252887.

Isolation of total RNA and synthesis of cDNA

RNA was extracted from moso bamboo shoot and leaves using a Classic Total RNA Isolation Kit (Sangon, Shanghai, China). cDNA was synthesized with a RevertAid First Strand cDNA Synthesis Kit (Fermentas, Vilnius, Lithuania) according to the manufacturer's instructions. The single-stranded cDNA was used as a template for reverse-transcription PCR amplification.

Sequence annotation and analysis

In silico gene prediction was performed using FGENESH (http://www.softberry.com) with monocot parameters. Predicted genes with less than 50 amino acids were excluded (http://rice.plantbiology.msu.edu/). The predicted genes were searched against protein sequence sets including the current the institute for genomic research (TIGR) rice annotation database and the SWISS-PROT database using BLASTP and BLASTX with a cutoff value of 1e-10.

The repeat annotation followed Paterson et al. (2009) except that (i) LTR FINDER (http://tlife.fudan.edu.cn/ltr_finder/) was used for the *de novo* prediction of long terminal repeat (LTR) retrotransposons with an *Oryza sativa* tRNA database and default parameters; (ii) MITEs were predicted with FINDMITE (Tu 2001); and (iii) a combined repeat database including mips_REdat_4.3 (http://mips.gsf.de/), the triticeae repeat sequence database (TREP) (http://wheat.pw.usda.gov/ITMI/Repeats/), RetrOryza (http://www.retroryza.org/) (Chaparro et al. 2007), TIGR plant repeats databases (http://www.tigr.org/tdb/e2k1/plant.repeats/) (Ouyang and Buell 2004) and RepBase (http://www.girinst.org/) (Jurka et al. 2005) was used. 100 random 800 Kb genomic sequences were selected from rice (http://rice.plantbiology.msu.edu/) and sorghum (http://www.jgi.doe.gov/) for statistical tests of repetitive element and GC contents. The insertion times of LTR retrotransposons were estimated according to SanMiguel et al. (SanMiguel et al. 1998). 5′ and 3′ solo LTRs were aligned using the water program (EMBOSS package, http://www.hgmp.mrc.ac.uk/Software/EMBOSS/) with default values and the nucleotide substitutions per site were calculated under the g-K2P model as implemented in MEGA v4.0 (Kumar et al. 2004). The substitution rate of $2 \times 6.5 \times 10^{-9}$ mutations per site per year (Gaut et al. 1996) was used to estimate the absolute time. Primer3 (Koressaar and Remm 2007) was used to design unique primer pairs.

Acknowledgements

We thank Hui Kang and Yongqiang Zhu (Shanghai) for their assistance in BAC sequencing. This work was supported by the Program for New Century Excellent Talents in University of China to Fan.

Received 15 Jan. 2010 Accepted 26 Apr. 2010

References

Chaparro C, Guyot R, Zuccolo A, Piegu B, Panaud O (2007) RetrOryza: a database of the rice LTR-retrotransposons. *Nucleic Acids Res.* **35**, D66–D70.

Clark LG (1996) Diversity, biogeography and evolution of Chusquea. In: Chapman GP, ed. *International Symposium on Bamboos*. Academic Press Ltd, London.

Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y, Zhu J, Liu Y, Hu X, Jia P, Zhao Q, Ying K, Yu S, Tang Y, Weng Q, Zhang L, Lu Y, Mu J, Zhang LS, Yu Z, Fan D, Liu X, Lu T, Li C, Wu Y, Sun T, Lei H, Li T, Hu H, Guan J, Wu M, Zhang R, Zhou B, Chen Z, Chen L, Jin Z, Wang R, Yin H, Cai Z, Ren S, Lv G, Gu W, Zhu G, Tu Y, Jia J, Chen J, Kang H, Chen X, Shao C, Sun Y, Hu Q, Zhang X, Zhang W, Wang L, Ding C, Sheng H, Gu J, Chen S, Ni L, Zhu F, Chen W, Lan L, Lai Y, Cheng Z, Gu M, Jiang J, Li J, Hong G, Xue Y, Han B (2002) Sequence and analysis of rice chromosome 4. *Nature* **420**, 316–320.

Gaut BS (2002) Evolutionary dynamics of grass genomes. *New Phytol.* **154**, 15–28.

Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. *Proc. Natl. Acad. Sci. USA* **93**, 10274–10279.

Gielis J, Valente P, Bridts C, Verbelen JP (1996) Estimation of DNA content of bamboos using flow cytometry and confocal laser scanning microscopy. In: Chapman GP, ed. *International Symposium on Bamboos*. Academic Press Ltd, London.

Gui Y, Wang S, Quan L, Zhou C, Long S, Zheng H, Jin L, Zhang X, Ma N, Fan L (2007) Genome size and sequence composition of moso bamboo: a comparative study. *Sci. China Ser. C* **50**, 700–705.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* **436**, 793–800.

Jiang ZH (2002) *Bamboo and Rattan in the World*. Liaoning Science and Technology Publishing House, Shenyang.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467.

Klinkenborg V (2001) *Bamboo for Gardens*. Times Book Rev, New York.

Kellogg EA (1998) Relationships of cereal crops and other grasses. *Proc. Natl. Acad. Sci. USA* **95**, 2005–2010.

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291.

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* **5**, 150–163.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.

Ouyang S, Buell CR (2004) The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Res.* **32**, D360–D363.

Paterson AH, Freeling M, Sasaki T (2005) Grains of knowledge: genomics of model cereals. *Genome Res.* **15**, 1643–1650.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The sorghum bicolor genome and the diversification of grasses. *Nature* **457**, 551–556.

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20**, 43–45.

Tettelin H, Radune D, Kasif S, Khouri H, Salzberg SL (1999) Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics* **62**, 500–507.

Tu Z (2001) Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. *Proc. Natl. Acad. Sci. USA* **98**, 1699–1704.

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, Zhang J, Zhang Y, Li R, Xu Z, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Deng Y, Ran L, Shi X, Wang X, Wu Q, Li C, Ren X, Li D, Liu D, Zhang X, Ji Z, Zhao W, Sun Y, Zhang Z, Bao J, Han Y, Dong L, Ji J, Chen P, Wu S, Xiao Y, Bu D, Tan J, Yang L, Ye C, Xu J, Zhou Y, Yu Y, Zhang B, Zhuang S, Wei H, Liu B, Lei M, Yu H, Li Y, Xu H, Wei S, He X, Fang L, Huang X, Su Z, Tong W, Tong Z, Ye J, Wang L, Lei T, Chen C, Chen H, Huang H, Zhang F, Li N, Zhao C, Huang Y, Li L, Xi Y, Qi Q, Li W, Hu W, Tian X, Jiao Y, Liang X, Jin J, Gao L, Zheng W, Hao B, Liu S, Wang W, Yuan L, Cao M, McDermott J, Samudrala R, Wong GK, Yang H (2005) The genome of *Oryza sativa*: a history of duplications. *PLoS Biol.* **3**, e38.

Zhang H (2000). *Construction and Manipulation of Large-insert Bacterial Clone Libraries-Manual*. Texas A&M University, Texas.

(Co-Editor: Yin-Long Qiu)

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. A document file including repeat annotation of Contig 28 and a list of bamboo's orthologs in rice, sorghum and *Arabidopsis*.

Figure S1. Amino acid substitution rates of 28 annotated bamboo genes (Table S1) with their orthologs in rice, sorghum and *Arabidopsis*. The orthologs are the best hits (also BLASTP E-value less than 1e-20) for the 28 bamboo genes in the rice (http://rice.plantbiology.msu.edu/), sorghum (http://www.jgi.doe.gov/) and *Arabidopsis* genomes (http://www.arabidopsis.org/), respectively. The orthologous gene pairs were ranked based on the bamboo-rice substitution rates (from low to high). At, *Arabidopsis thaliana*; Os, rice; Pp, bamboo; Sb, sorghum.