Mol Breeding (2010) 25:299–311
DOI 10.1007/s11032-009-9333-4

Development, characterization and utilization of GenBank microsatellite markers in *Phyllostachys pubescens* and related species

Ding-Qin Tang · Jiang-Jie Lu · Wei Fang · Shan Zhang · Ming-Bing Zhou

Received: 16 January 2009 / Accepted: 1 September 2009 / Published online: 2 October 2009

Abstract Public sequence databases provide a rapid, simple and cost-effective source of microsatellite markers. We analyzed 1,532 bamboo (*Phyllostachys pubescens*) sequences available in public domain DNA databases, and found 3,241 simple sequence repeat (SSR) loci comprising repeats of two or more nucleotides in 920 genomic survey sequences (GSSs) and 68 cDNA sequences. This corresponded to one SSR per 336 bp of GSS DNA and one SSR per 363 bp of cDNA. The SSRs consisted of 76.6 and 74.5% dinucleotide repeats, 20.0 and 22.3% trinucleotide repeats, and 3.4 and 3.2% higher-number repeats in the GSS DNA and cDNA sequences, respectively. The repeat motif AG/CT (or GA/TC) was the most abundant. Nineteen microsatellite markers were developed from Class I and Class II SSRs, showing that the limited polymorphism in *Ph. pubescens* cultivars and provenances could be attributed to clonal propagation of the bamboo plant. The transferability of the microsatellites reached 75.3%, and the polymorphism of loci successfully transferred was 66.7% for six additional *Phyllostachys* species. Microsatellite PBM014 transferred successfully to all six species, showed rich polymorphism, and could serve as species-specific alleles for the identification of *Phyllostachys* interspecies hybrids.

Keywords Microsatellites (SSRs) · *Ph. pubescens* · Polymorphism · Cross-species transferability · Hybrid identification

D.-Q. Tang (✉) · J.-J. Lu · W. Fang · M.-B. Zhou
Key Lab for Modern Silvicultural Technology of Zhejiang Province, Zhejiang Forestry University, 311300 Lin'an, Zhejiang Province, People's Republic of China
e-mail: tang@zjfc.edu.cn

S. Zhang
Eastwin Life Science Inc, 100085 Beijing, People's Republic of China

Introduction

Bamboo is a monocotyledonous plant in the family Poaceae and subfamily Bambusoideae. China has the largest reserve of bamboo in the world. *Phyllostachys pubescens* (synonyms *Ph. edulis*) is the most important bamboo species in China and the third most important plant species for timber production next to *Pinus massoniana* and *Cunninghamia lanceolata.* There are 4.2 million ha of bamboo forest in China, *Ph. pubescens* representing 3 million ha and corresponding to approximately 2% of China's total forest area, having doubled over the last 30 years (Fu 2001). It is the predominant source of bamboo shoots and plays an important ecological role.

Phyllostachys species flower at intervals of 60–120 years, often flowering simultaneously over an extensive area before dying (Janzen 1976; Watanabe et al. 1982). Populations recover mainly through the

development of seedling cohorts, which suffer high mortality rates during early development, the survivors extending rhizomes and producing culms. Therefore, it is difficult to achieve genetic improvements in *Ph. pubescens* through plus-tree selection and sexual hybridization. Molecular markers that facilitate the analysis of genetic traits are important for plant improvement (Gupta et al. 1999) and several types of marker have been applied in bamboo research including RAPDs (SCARs), AFLPs, ISSRs and unidentified EST-derived SSRs (reviewed by Zhang and Tang 2007). Research focuses on species identification (Das et al. 2005; Lin et al. 2009), genetic diversity (Friar and Kochert 1991; Lai and Hsiao 1997; Barkley et al. 2005; Ruan et al. 2008), the clonal structure of populations (Suyama et al. 2000; Isagi et al. 2004), phylogenetic relationships and species evolution (Friar and Kochert 1994; Loh et al. 2000; Li et al. 2002; Barkley et al. 2005).

Microsatellite markers, also known as simple-sequence repeats (SSRs), are DNA sequences 1–6 bp in length that are tandemly repeated a variable number of times (Tautz 1989). They are particularly valuable in plant-breeding programs because they are polymorphic, co-dominant, relatively abundant, widely dispersed across the genome, and easy to score using automated methods (Powell et al. 1996). Microsatellites have been used in cultivar identification and parentage assessment (Buteler et al. 2002; Malysheva et al. 2003; Rajora and Rahman 2003), genetic diversity analysis (Goldstein and Clark 1995; Cho et al. 2000), evolutionary and phylogenetic studies (Pupko and Graur 1999; Zhu et al. 2000), the construction of molecular maps (Bell and Ecker 1994; Temnykh et al. 2000), and the support of patents and property rights for plant varieties (Powell et al. 1996; Gupta et al. 1999). Recently, six microsatellite markers were developed from a genomic library of *Bambusa arundinacea*, a bamboo species in India with a caespitose rhizome system (Nayak and Rout 2005). However, developing microsatellite markers for a new plant species through the use of genomic DNA or enriched-SSR libraries is laborious, expensive and inefficient (Powell et al. 1996). Fortunately, searching through published sequence databases offers an alternative. In this study, we used sequence databases of *Ph. pubescens* with a scattered rhizome system to (1) evaluate the frequency and distribution of different classes and types of SSRs in the genome, (2) establish and validate microsatellite markers for the detection of polymorphisms in a reference set of *Ph. pubescens* cultivars and provenance populations, and (3) assess cross-species transferability and identify *Phyllostachys* interspecies hybrids.

Materials and methods

Database search and primer design for microsatellite markers

DNA sequences in GenBank (http://www.ncbi.nlm.nih.gov/) were searched for the phrase "*Phyllostachys pubescens*" and the search results were downloaded as FASTA-formatted sequence files. Web software RepeatMasker (Smit et al. 1996–2004: http://www.repeatmasker.org/) and Microsatellite Repeats Finder (Benson 1999: http://biophp.org/minitools/microsatellite_repeats_finder/) were used to detect tandem repeats of 2–6 nucleotides (Table 1). PCR primers were designed to anneal in the flanking regions of identified repeat sequences using the computer program Primer Primer 5 (PRIMIER Biosoft International, CA, USA). When two distinct microsatellite sequences were present at distant sites in one DNA sequence [for example, $(TA)_9$ and $(TTTTC)_4$ in ED018039], primer pairs were designed for each microsatellite. When two microsatellites in one DNA sequence were in close proximity [for example, $(TA)_{24}$ and $(GA)_{14}$ in ED018001], the primer pairs were designed outside these microsatellites (Table 2). Primers that met these requirements generated PCR products in the range 100–250 bp.

Plant materials and DNA extraction

Two groups of cultivars or forms and provenances were used to evaluate microsatellite marker polymorphism (Fig. 1). The first group included a 1-year-old seedling and nine cultivars or forms showing morphological differences in stem shape and color, and leaf color (Lin et al. 2009). The second group consisted of 17 provenances collected from bamboo stands in eight provinces (Jiangsu, Fujian, Hunan, Hubei, Jiangxi, Guangdong, Zhejiang and Anhui), representing almost all *Ph. pubescens* habitats. The provenances were genetically divergent in growth

Table 1 Size distribution of microsatellite motifs observed in *Ph. pubescens* sequences in the GenBank database

SSR motif	Number of repeat units in GSS/cDNA								
	3	4	5	6	7	8	9	≥10	Total
Di-nucleotide	2043/121	217/14	39/1	15/0	3/0	7/1	4/0	14/0	2342/137
AG/GA/CT/TC	750/48	79/1	7/1	9/0	1/0	5/0	2/0	5/0	858/50
AC/CA/GT/TG	642/28	66/4	10/0	2/0	2/0		1/0	1/0	724/32
AT/TA	475/32	52/7	18/0	3/0		2/1	1/0	8/0	559/40
CG/GC	176/13	20/2	4/0	1/0					201/15
Tri-nucleotide	554/36	50/4	5/1	1/0	1/0	1/0			612/41
AAC/ACA/CAA/GTT/TGT/TTG	60/4	1/0							61/4
AAG/AGA/GAA/CTT/TCT/TTC	96/9	10/0							106/9
AAT/ATA/TAA/ATT/TAT/TTA	48/0	2/0							50/0
ACC/CCA/CAC/GGT/TGG/GTG	46/3	2/0							48/3
ACG/CGA/GAC/CGT/TCG/GTC	27/1	2/1	1/0						30/2
ACT/CTA/TAC/AGT/TAG/GTA	27/0	4/0	0/1						31/1
AGC/GCA/CAG/GCT/TGC/CTG	36/5	1/0							37/5
AGG/GGA/GAG/CCT/TCC/CTC	90/6	7/3	1/0		1/0	1/0			100/9
ATC/TCA/CAT/GAT/TGA/ATG	44/3	8/0							52/3
CCG/CGC/GCC/CGG/GCG/GGC	80/5	13/0	3/0	1/0					97/5
Tetra-nucleotide	61/4	10/0	2/0	1/0					74/4
Penta-nucleotide	19/2	3/0	1/0						23/2
Hexa-nucleotide	6/0								6/0

characteristics (Chen et al. 2001) and physical and mechanical properties of timber (Liu et al. 2008). *Phyllostachys pubescens* and six additional *Phyllostachys* species (Table 3) collected from Anji Bamboo Germplasm Garden, Anji, Zhejiang Province, were used to test the amplification, sequencing and identification of SSRs (Fig. 2). Four clones of *Phyllostachys* interspecies hybrids sampled from Jiangxi province were used to test the new microsatellite loci (personal communication with Professor Liao of Jiangxi Forestry Research Institute). Genomic DNA was extracted from young leaves with the hexadecyltrimethylammonium bromide (CTAB) method (Doyle and Doyle 1987), with some modifications.

PCR and sequencing of microsatellite loci

Newly synthesized primer pairs (Shanghai Sangon Biological Engineering Technology & Services Co., Ltd) were tested for PCR amplification using DNA from *Ph. pubescens*. PCR amplification was performed in a thermal cycler (PE 9700, ABI) in 20-μl reactions comprising 50–100 ng of template DNA, 0.2 μM of each primer, 200 μM of each dNTP and 1 unit of *Taq* DNA polymerase with 1× PCR universal buffer (10 μM Tris–HCl, pH 8.3 at 25°C; 50 μM KCl), and 1.5 μM $MgCl_2$ (Shanghai Sangon Biological Engineering Technology & Services Co., Ltd). The reaction consisted of heating to 95°C for 5 min, followed by 30 cycles of 1 min denaturation at 95°C, 1 min annealing at 46–59°C depending on the primer pair (Table 2) and 2 min extension at 72°C, and a final step at 72°C for 5 min. Amplified microsatellite loci were further tested in six *Phyllostachys* cross-species (Table 3) and interspecies hybrids (Fig. 3). The PCR primer annealing temperature was lowered by 2–5°C according to the evolutionary distance between *Ph. pubescens* and related species (Rossetto 2001). PCR products were separated on 6% polyacrylamide denaturing gels, and marker bands were revealed by silver staining as described by Panaud et al. (1996). Desired bands were excised from the gel, purified using the EZ-10 Spin Column DNA Gel Extraction Kit (Biobasic Inc.) and ligated into the pUC18 vector (TaKaRa, Japan). Three positive

Table 2 Candidate sequences and microsatellite marker information

Marker name	GenBank accession no.	Motif	Position	Length of PCR fragment	5′ Primer sequence	3′ Primer sequence	Annealing temperature
PBM001	ED018039	$(CA)_{24}(TA)_{10}$	71				
PBM002	ED018039	$(AT)_{10}$	374	298	ATGGAGCATTCATTTGAG	TTGACAAGTTTGGAGGGT	48
PBM003	ED018039	(TTTTC)4	644				
PBM004	ED018001	(AT)24	175	291	CGATGAAGTTGCCTGTGG	TCAAGTCGCATTGGAGTA	48
PBM005	ED018001	(AG)14	224				
PBM006	ED018473	(TA)10	75				
PBM007	ED018473	(AT)17	224	189	CTCGAGGTGAAGTTACATTA	TCGTAGAGCACTATCCATTT	53
PBM008	ED018674	(TA)17	1				
PBM009	ED018674	(TA)12	37				
PBM010	ED018674	(GGTTG)3	190				
PBM011	ED018475	(TA)8	295	169	TAAAGAAGTATAATGTATTCG	AGCTGTTTTAGCATCTATGTT	50
PBM012	ED018770	(AT)15	1				
PBM013	ED018016	(TC)11	191	174	TTTTACTAACGTGCCCGTATT	GGAGGATCGCTGGATTGG	53
PBM014	ED018306	(CT)11	245	294	GCCGTCCAAACGCTCCTT	CACCCATCCATCTTATGCTAT	52
PBM015	ED018426	(CT)20	245				
PBM016	ED018730	(AG)10	56	88	CCCACAGAAGGTCGCACTA	CCGCCGAGAACGTCTACAT	56
PBM017	ED018105	(CCT)8	432	237	TATGCCTCCAATAATCCG	AAGGCGTAGCCACCGAAT	55
PBM018	ED018452	(CAG)3	418	285	GACGACGAGTGTAACGCTGAG	AGGATGATGCCGGTGAGG	52
PBM019	ED018615	(CTC)8	645	202	TCCTAGCCCTACCCTGTCC	CACGGTGCTCGCTTAAATAG	59
PBM020	ED018616	(TCT)4	572	117	TCGCATCTAAACAACATC	GAGAAGAAGAAAGGAGGA	46
PBM021	ED018721	(GCC)4	827	260	AGGGTGTTATTTGCTATTGT	AAATCCGACGCTGGAGGC	55
PBM022	ED018528	(TTTG)5	255	146	AAAGATGTTGCCTTGTTC	GTGCATGTCTATTCCCTC	47
PBM023	ED017945	(TTCCG)3	199	208	TTTCATGGAGGGCGACAG	GCCGACCAGGAAACAAACGAA	58
PBM024	ED018034	(ATGG)4-6	379				
PBM025	ED018083	(ATTTC)5	359	281	GTTGTGCCCAGCACTTTA	TGACGCCACTCATTTCTT	47
PBM026	ED018344	(GCTCG)4	395	144	CTATCATCAAATGTGCCTCAA	CCGAGCCTCACTTTCTGC	50
PBM027	ED018387	(TTCTCC)3	725	287	AAAGCAACCACGCCATTA	TCGGAGGAGAAGGAGAAG	49
PBM028	ED018541	(CGAGC)5	684	157	CAGGTAGTCGTCTTATCGAGT	ATAAGTCAATTTCGAATGAA	53
PBM029	ED018589	(CTCC)5	795				
PBM030	ED018773	(ACAT)4	161	224	GTTGCCCGTGGTTACTGT	CGTGAGCCATCTGAGTTT	50

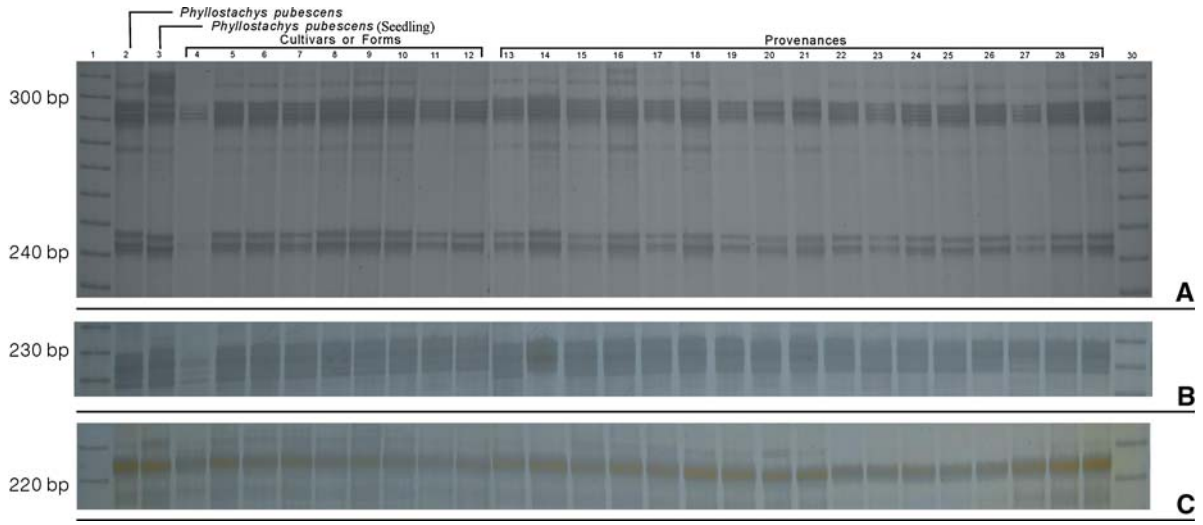

Fig. 1 **a–c** Polyacrylamide gel electrophoresis patterns of microsatellites derived from GSS sequences on a panel of 11 varieties and 17 provenances: Lane 1, 30: size marker; Lane 2: *Ph. pubescens* as a reference; Lane 3: seedling; Lanes 4–12: cultivars or forms of *Ph. pubescens* cv. Ventricosa, *Ph. pubescens* cv. Tao Kiang, *Ph. pubescens* cv. Viridisulcata, *Ph. pubescens* cv. Luteosulcata, *Ph. pubescens* cv. Gracilis, *Ph. pubescens* cv. Obliquinoda, *Ph. pubescens* cv. Tubaeformis, *Ph. pubescens* cv. Heterocycla, *Ph. pubescens* cv. Pachyloen; Lanes 13–29: provenances in Jurong of Jiangsu, Yixing of Jiangsu, Huoshan of Anhui, Wuhan of Hubei, Anji and Zhuzhou of Zhejiang, Jiujiang of Jiangxi, Shangrao of Jiangxi, Lechang and Conghua of Guangdong, Wuyi, Songxi, Jian'ou, Shaxian, Hua'an and Longhai of Fujiang province. **a**: PBM014; **b**: BPBM017; **c**: PBM019

clones for each species were selected for sequencing using BigDye terminator V3.1 in a cycle sequencing protocol according to the manufacturer's specifications (PE Applied Biosystems, ABI PRISM 3100-Avant Automatic DNA Sequencer). Sequences were deposited in the NCBI GenBank database (accession numbers FJ588714–FJ588848).

Analysis of sequence data

Vector sequences were removed, then edited using Vector NTI software (version 10.0, Invitrogen Corporation, USA). The DNA sequences were then aligned using the CLUSTAL method included in the software. Multiple gaps were closed manually to group the repeat sequences.

Results

Screening of GenBank and characteristics of *Ph. pubescens* SSRs

A total of 1,532 *Ph. pubescens* DNA sequences (including 200 cDNA sequences) was downloaded from GenBank using "*Phyllostachys pubescens*" as a search keyword (accessed before November 3rd 2008). After screening with RepeatMasker and Microsatellite Repeats Finder and excluding single-nucleotide repeats, we found 3,057 SSRs in 920 out of 996 GSSs (Gui et al. 2007) covering ~722 kb including 770 GSSs containing more than one SSR. In addition, 58 out of 68 cDNA sequences representing 66.8 kb contained 184 SSRs (Table 1). This corresponds to average distances of approximately 336 bp between SSRs in GGSs and 363 bp between SSRs in cDNAs. The size of the repeat unit was not evenly distributed among the SSR loci: 76.6 and 74.5% of the SSRs were dinucleotide repeats, 20.0 and 22.3% trinucleotide repeats, and 3.4 and 3.2% were higher-number repeats in GSS and cDNA sequences, respectively. This indicates that dinucleotide SSRs are the most frequent in both GSS and cDNA sequences. Among the dimeric SSRs, AG/CT (or GA/TC) repeat was the most common in both GSSs (36.7%) and cDNAs (39.7%), whereas GC or CG repeat was comparatively rare in GSSs (8.6%) and cDNAs (10.7%). SSRs ≥20 nucleotides in length (Class I microsatellites) accounted for only 0.75% of GSS SSRs (a total of 23 sequences), whereas SSRs

Table 3 Amplicon size, microsatellite motifs and polymorphism in species of genus *Phyllostachys*

Locus	Species							Transferability/ polymorphism
	Ph. pubescens	*Ph. nidularia*	*Ph. atrovaginata*	*Ph. heteroclada*	*Ph. praecox*	*Ph. kwangsiensis*	*Ph. bambusoides*	
PBM014	294/244	316/298/244	274/244	292/244	274/244	294/244	283/244	100%/83.3%
	$(CT)_{11}$-$(CAT)_3$/$(CT)_4$	$(CT)_{21}$-$(CAT)_4(CT)_{12}$-$(CAT)_5$/$(CT)_4$	$(CT)_7$-$(CAT)_4$/$(CT)_4$	$(CT)_{10}$-$(CAT)_4$/$(CT)_4$	$(CT)_8$-$(CAT)_3$/$(CT)_4$	$(CT)_{11}$-$(CAT)_3$/$(CT)_4$	$(CT)_7$-$(CAT)_3$/$(CT)_4$	
PBM016	88	86	88	104	126	88	94	100%/83.3%
	$(AG)_{10}$	$(AG)_9$	$(AG)_9$	$(AG)_{18}$	$(AG)_4$	$(AG)_{10}$	$(AG)_{13}$	
PBM018	285	285	285	285	285	285	285	100%/0%
	$(CAG)_3$	$(CAG)_3$	$(CAG)_3$	$(CAG)_3$	$(CAG)_3$	$(CAG)_3$	$(CAG)_3$	
PBM020	117	117	117	117	117	117	117	100%/0%
	$(TCT)_4$	$(TCT)_4$	$(TCT)_4$	$(TCT)_4$	$(TCT)_4$	$(TCT)_4$	$(TCT)_4$	
PBM022	146	146	146	146	138	146	146	100%/83.3%
	$(TTTG)_5$	$(TTTG)_4$	$(TTTG)_4$	$(TTTG)_4$	$(TTTG)_3$	$(TTTG)_3$	$(TTTG)_5$	
PBM025	281	281	281	281	281	281	281	100%/0%
	$(ATTTC)_5$	$(ATTTC)_5$	$(ATTTC)_5$	$(ATTTC)_5$	$(ATTTC)_5$	$(ATTTC)_5$	$(ATTTC)_5$	
PBM023	208	208	208	208	208	208	208	100%/0%
	$(TTCCG)_3$	$(TTCCG)_3$	$(TTCCG)_3$	$(TTCCG)_3$	$(TTCCG)_3$	$(TTCCG)_3$	$(TTCCG)_3$	
PBM027	287	271	281	269	285	258	281	100%/100%
	$(TTCTCC)_5$	$(TTCTCC)_2$	$(TTCTCC)_4$	$(TTCTCC)_2$	$(TCC)_3(TTCTCC)_3$	$(TC)_3$-$(TTCTCC)_2$	$(TTCTCC)_3$	
PBM028	157	157	157	157	157	157	157	100%/100%
	$(CGAGC)_5$	$(CGAGC)_4$	$(CGAGC)_4$	$(CGAGC)_4$	$(GAGCC)_3$	$(CGAGC)_3$	$(GAGCC)_3$	
PBM004	291/241/233	291/209	313/306/225	291/278/231/211	277/219	209	216	100%/100%
(PBM005)	$(AT)_{24}(AG)_{14}$/$(AT)_{15}$/$(AT)_{11}$	$(AT)_6(GTATAT)_2(GT)_7$ $ATGT(AT)_4GT(AT)_{20}$/$(AT)_5$	$(AC)_{32}(AT)_{23}$/$(AC)_{32}$ $(AT)_{19}$/$(AT)_4AC(AT)_8$	$(AT)_{32}$/$(AC)_{13}GC(AC)_{24}$/$(AT)_9$/$(AT)_5$	$(AT)_{18}(AG)_{13}$/$(AG)_5$	$(AT)_5$	$(AT)_5$	
PBM017	237	231	242	242/230	240	NA	240	66.7%/100%
	$(CCT)_8$	$(CCT)_7$	$(CCT)_9$	$(CCT)_9$/$(CCT)_7$	No		$(CCT)_9$	
PBM021	260	244	264	301	331	NA	279	66.7%/50%
	$(GCC)_4$	No	(GCC)4	(GCC)3	$(GCC)_3$		$(GCC)_4$	
PBM011	169	229/151	NA	NA	176	NA	178	50%/100%
	$(TA)_8$	$(TA)_{16}$/$(TA)_7$			$(TA)_{20}$		$(TA)_{27}$	

Table 3 continued

Locus	Species							Transferability/ polymorphism
	Ph. pubescens	*Ph. nidularia*	*Ph. atrovaginata*	*Ph. heteroclada*	*Ph. praecox*	*Ph. kwangsiensis*	*Ph. bambusoides*	
PBM019	202	193	191	188	NA	189	202	83.3%/100%
	$(CTC)_8$	$(CTC)_4$	$(CTC)_5$	$(CTC)_5$		$(CTC)_4$	$(CTC)_4$	
PBM013	175	163	161	162	NA	NA	171	66.7%/100%
	$(TC)_{11}$	$(TC)_5$	$(TC)_4$	$(TC)_4$			$(TC)_9$	
PBM030	224	383	299	374	383	220	255	66.7%/100%
	$(ACAT)_4$	No	$(GA)_3$- $(AT)_3$	$(AT)_3$- $(AT)_3$- $(AT)_3$	No	$(AT)_5$	$(GA)_3$- $(AT)_5$	
PBM002	298	NA	NA	NA	NA	NA	NA	0%/-
	$(AT)_{10}$							
PBM007	189	NA	NA	NA	NA	179	NA	15.6%/100%
	$(AT)_{17}$					$(AT)_{12}$		
PBM026	144	158	159	137	159	144	159	15.6%/0%
	$(GCTCG)_4$	No	No	No	No	$(GCTCG)_4$	No	

The size (in bp) of the amplicon is listed above the sequence of the microsatellite motif in the amplicon. The slash between size estimates of microsatellite motifs separates data for multiple amplicons at that locus

NA, No indicate no amplified product and no microsatellite sequence detected in the amplicon, respectively. Polymorphism is calculated only from the successfully transferred loci reference to *Ph. pubescens*

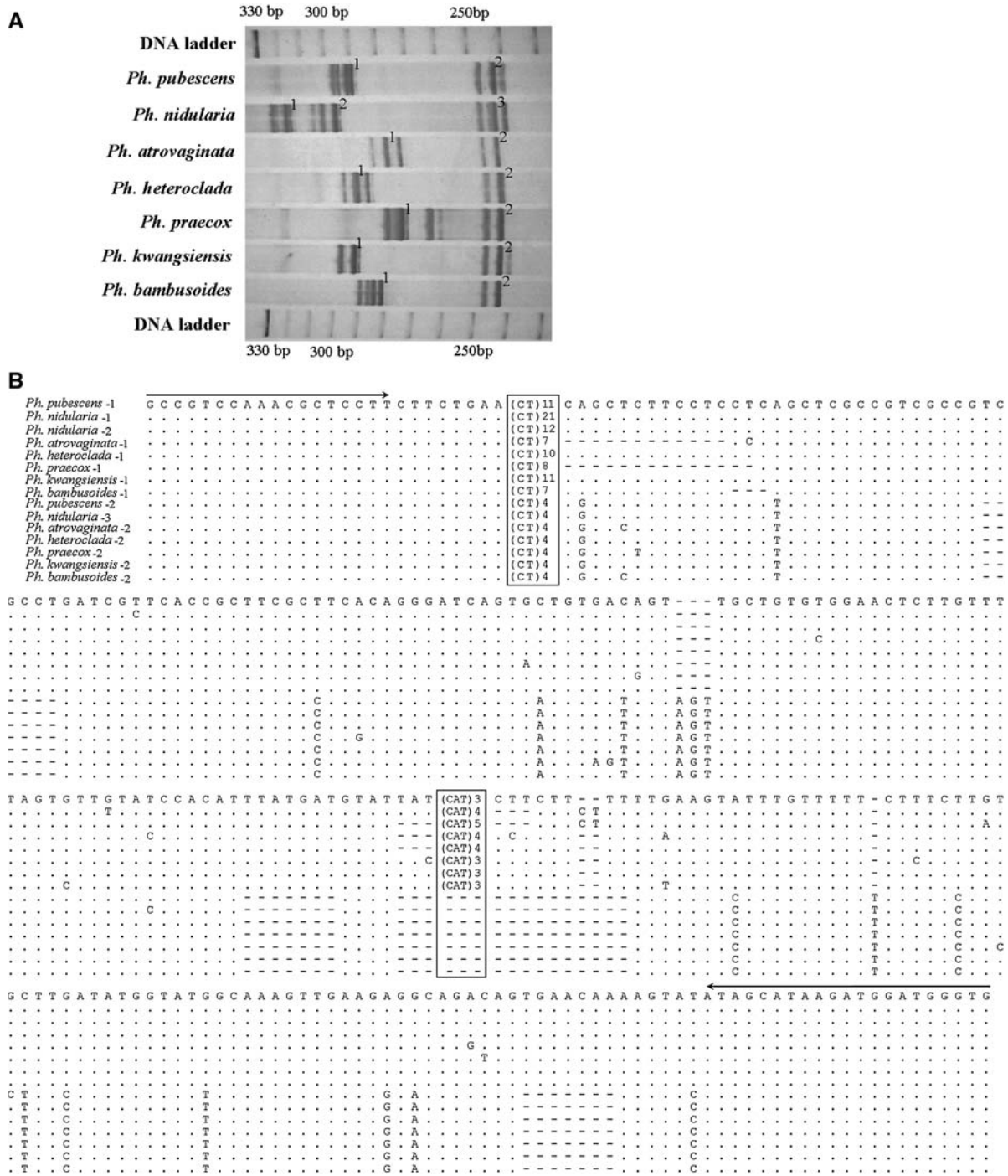

Fig. 2 **a** Polyacrylamide gel electrophoresis patterns of microsatellite alleles derived from locus PBM014. **b** Alignment of the nucleotide sequences of the microsatellite alleles at locus PBM014 amplified from seven representative bamboo species of genus *Phyllostachys*. Nucleotides conserved among these sequences (relative to *Ph. pubescens*) are shown by *dots*. The *dashes* indicate deletions. The *arrows* indicate the primer sequences used to amplify this microsatellite locus. The *box* highlights the microsatellite motif. The suffix numbers after bamboo species correspond to the DNA bands marked in **a**

12–19 nucleotides in length (Class II microsatellites) accounted for 2.4 and 2.6% of the total number of SSRs in GSS and cDNA sequences, respectively.

Assessing the polymorphism of *Ph. pubescens* microsatellite markers

We sought to develop microsatellite markers based on public genomic sequence information, incorporating empirically-derived data concerning the frequency, size polymorphism, and PCR-amplification properties of different types of *Ph. pubescens* SSR. Primers were designed for 30 of the longest SSR loci in 24 GSSs comprising 23Class I SSRs, six Class II SSRs and one SSR <12 nucleotides in length (Table 2). Primer pairs were successfully designed for 26 (86.7%) of the loci, the remainder either containing insufficient flanking sequences [$(TA)_{17}$ in ED018674 and $(TA)_{15}$ in ED018770] or were inappropriate for primer modeling [$(CA)_{24}(TA)_{10}$ in ED018039 and $(TA)_{12}$ in ED018674]. After PCR amplification with the appropriate primers, polyacrylamide gel electrophoresis and sequencing led to the further development of 19 microsatellite markers (summarized in Table 2).

To determine the polymorphism of the selected loci, one *Ph. pubescens* seedling, nine cultivars or forms and 17 provenances of were surveyed using the 19 corresponding primer sets listed in Table 2. We detected little size or sequence variation (data not shown) with the exception of locus PBM014 (Fig. 1), indicating that the 19 microsatellite loci display little polymorphism and the genetic diversity of *Ph. pubescens* is low.

Microsatellite analysis in *Phyllostachys* interspecies hybrids

To test the transferability of the 19 microsatellites to other bamboo species, six diverse *Phyllostachys* species were selected for cross-species amplification with the SSR primers identified in *Ph. pubescens* (Table 3). All but one of the markers (94.7%) transferred successfully to at least one other *Phyllostachys* species. Ten microsatellites (52.6%), corresponding to loci PBM014, PBM016, PBM018, PBM020, PBM022, PBM025, PBM023, PBM027, PBM028 and PBM004 (PBM005), successfully transferred to all six species. The remaining microsatellites failed to amplify correctly or efficiently (e.g. loci PBM002 and PBM007 produced weak amplification products in all six species, and PBM026 produced a product lacking the SSR in five of the six species). The aggregate transferability of the 19 *Ph. pubescens* SSRs was 75.3%.

We observed clear interspecies variation in most of the 18 microsatellite loci although the size polymorphism was generally subtle, the major exception being locus PBM030 (Table 3). No variation was observed in loci PBM018, PBM020, PBM025 or PBM023, but significant variation was observed in loci PBM014, PBM016, PBM028 and PBM004 (PBM005). Thirteen of the 18 loci showed polymorphism in at least one other species using *Ph. pubescens* as the comparator (72.2%), which gave an average polymorphism of 66.7% for the 18 successfully transferred loci. PBM014 showed the richest polymorphism among the seven *Phyllostachys* species, with three amplicons in *Ph. nidularia* and two in the other species. In the smallest amplicon, all seven *Phyllostachys* species shared a 244-bp sequence containing the common SSR motif $(CT)_4$ and minimal nucleotide difference. The largest amplicon contained both poly $(CT)_n$ with 7–21 copies and poly $(CAT)_n$ with 3–5 copies of compound repeat motifs. In addition, differences between the smallest and largest amplicons reflected the CT copy number, the presence or absence of $(CAT)_n$ and further insertions and deletions (indels) in the flanking region.

The different SSR copy numbers among these bamboo species provide a series of SSLP (simple sequence length polymorphism) markers that may be useful for the genetic analysis of interspecies hybrids (Fig. 3). Polymorphic microsatellites (SSLPs) were selected to characterize the unidentified interspecies hybrid samples. PBM014 was polymorphic for all of the bamboo species tested and selected for parental (*Ph. kwangsiensis*, *Ph. bambusoides* and *Ph. pubescens*) species-specific alleles at this locus (Fig. 2a). As indicated in Fig. 3a, the heterozygosity of clones I and II revealed *Ph. kwangsiensis* and *Ph. bambusoides* parents, whereas homozygous clone III was derived from the female parent *Ph. kwangsiensis*. Sequencing the corresponding bands demonstrated that all of the bands contained the $(CT)_n$ and $(CAT)_n$ SSR motifs. These data show that clones I and II are

interspecies hybrids of *Ph. kwangsiensis* and *Ph. bambusoides*, but clone III is not. Similarly, clone IV was confirmed as an interspecies hybrid of *Ph. bambusoides* and *Ph. pubescens*.

Discussion

This study sought to develop microsatellite markers for bamboo based on sequences containing SSRs deposited in public databases. In the past, this method has only been suitable for well-characterized plants such as *Arabidopsis thaliana* (Bell and Ecker 1994), rice (Cho et al. 2000; Temnykh et al. 2001), and other cereals (Cordeiro et al. 2001; Thiel et al. 2003; La Rota et al. 2005). However, with the advent of low-cost, large-scale DNA sequencing, the pool of DNA sequence information in public databases for bamboo species, especially *Ph. pubescens* and *Dendrocalamopsis oldhamii*, has increased rapidly (reviewed by Tang 2009). In this study, we detected more than 3,200 SSRs from 966 GSS and 200 cDNA sequences and determined several notable characteristics of SSRs in the *Ph. pubescens* genome. Most of the SSRs were dinucleotide repeats (>75%) and proportion of the repeated number less than four was 87.7%. These characteristics resulted in short SSRs, 99% of which were <20 bp in length, and 97% of which were <12 bp in length, corresponding to our previous characterization of SSR-enriched libraries (data not shown). Although rice shows a similar tendency (Temnykh et al. 2001), the bias in frequency and length variation appears much more pronounced in *Ph. pubescens*.

Microsatellites can be categorized as class I (SSRs ≥20 nucleotides) or class II (SSRs ≥12 but <20 nucleotides) (Temnykh et al. 2001). Class I microsatellites tend to be highly polymorphic, whereas class II microsatellites show less variability (Weber 1990; Cho et al. 2000; Temnykh et al. 2000). Microsatellites <12 bp tend to mutate at the same rate as unique sequences and therefore demonstrate stochastic variation (Pupko and Graur 1999). To develop polymorphic microsatellite markers for bamboo, we followed the procedure already adopted for rice (Temnykh et al. 2001). We identified 30 candidate microsatellite loci (predominantly class I) and selected 19 for further development (Table 2).

Polymorphism was limited among the nine cultivars or forms and 17 provenances of *Ph. pubescens* (Fig. 1), which nevertheless show genetic variation when typed with AFLP and ISSR markers (Ruan et al. 2008; Lin et al. 2009). At the genus level, however, polymorphism reached an average of 66.7% with *Ph. pubescens* as the comparator (Table 3), much lower than the 86.0% transferred polymorphism statistically calculated from 1,800 species/primers at the same genus level (Rossetto 2001). Rich polymorphism is the norm for microsatellites (Tautz 1989), e.g. in rice, the average polymorphism in microsatellites developed from the GenBank database is 54%, whereas that observed in genomic libraries is 83.8% at the intra-species level (Cho et al. 2000). Therefore, the limited polymorphism observed in this study appears to be another distinguishing feature for *Ph. pubescent* and other *Phyllostachys* species. Replication slippage (Schlotterer and Tautz 1992; Richards and Sutherland 1994) and recombination (Jakupiak and Wells 1999) during DNA replication are the principal methods for the diversification of microsatellite alleles. However, the long flowering interval (Janzen 1976; Watanabe et al. 1982) means that *Phyllostachys* species are propagated clonally more often than sexually, reducing the likelihood of allele diversification as described for *B. arundinacea* (Nayak and Rout 2005).

Microsatellite primers developed for one species can be used to detect polymorphism at homologous sites in related species. The transfer success of microsatellites was an average of 76.4% at the genus level and 35.2% at the family level (Rossetto 2001). In rice, the transfer success was >90% at the genus level (Wu and Tanksley 1993). In bamboo, the transfer success of microsatellites in *B. arundinacea* was 100% (6/6 loci) for species of *Bambusa* and 83.3% (5/6 loci) for other genera in the Bambusoideae subfamily (calculated from the data provided by Nayak and Rout, 2005). In this study, the transfer success of *Ph. pubescens* microsatellites was an average of 75.3% at the genus level, lower than that reported in *Bambusa*. However, when differences concerning amplicon size and the identification of repeat motifs are taken into account, our results are consistent with those of Nayak and Rout (2005).

The frequency of microsatellites in plants is higher in transcribed regions, especially in the untranslated region (Morgante et al. 2002). EST-derived and/or

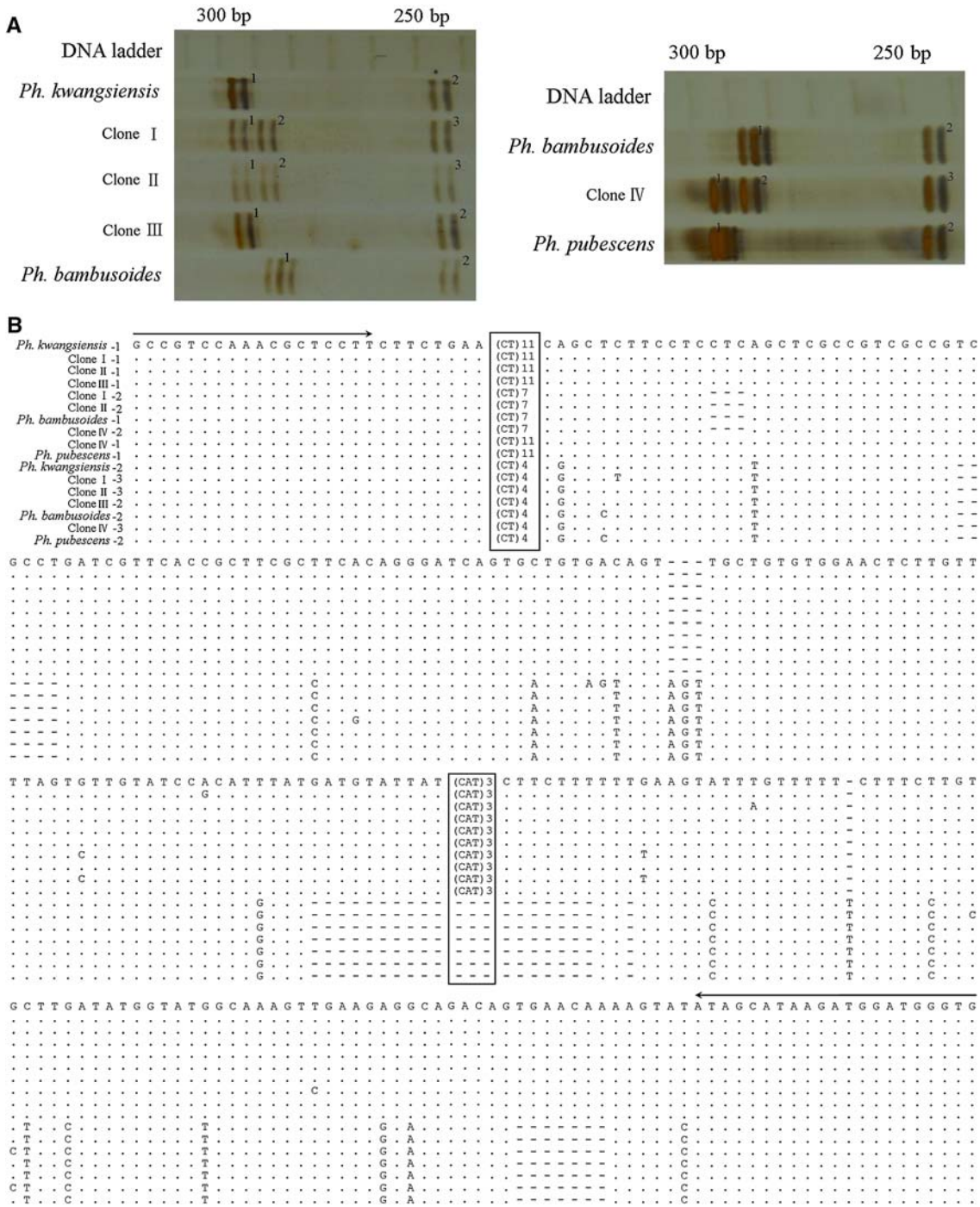

Fig. 3 **a** Microsatellite DNA fingerprints of *Ph. kwangsiensis* and *Ph. bambusoides* and their presumable hybrids at PBM014 locus. **b** Alignment of the nucleotide sequences of the microsatellite alleles at locus PBM014 amplified from *Ph. kwangsiensis* and *Ph. bambusoides* and their presumable hybrids. Nucleotides conserved among these sequences (relative to *Ph. kwangsiensis*) are shown by *dots*. The *dashes* indicate deletions. The *arrows* indicate the primer sequences used to amplify this microsatellite locus. The *box* highlights the microsatellite. The suffix numbers after bamboo species correspond to the DNA bands marked in **a**

unigene-derived microsatellites demonstrate a high-level of transferability to distantly related species, thereby providing additional markers for comparative genomics and evolutionary studies (Cordeiro et al. 2001; La Rota et al. 2005; Zhang et al. 2005; Parida et al. 2006). Barkley et al. (2005) assessed genetic diversity and phylogenetic relationships of temperate bamboo species using the transferred EST-derived SSRs from major cereals, but did not identify the transferability of SSRs between the cereals and bamboo species. In this study most microsatellite loci could be transferred to all the tested species, and might be present in the transcribed region. BLAST searches using these *Ph. pubescens* SSR loci showed no matches to any of the transcribed or EST regions except locus PBM014, which is closely related to the sequence encoding a hypothetical rice protein (1074–1412 bp of Os07g0569100). With polymorphism and transfer success among all six *Phyllostachys* species, the PBM014 locus could be used to identify *Phyllostachys* interspecies hybrids and the corresponding parental alleles. In another of our studies, the PBM014, PBM022, PBM025 and PBM028 loci transferred successfully to other bamboo species beyond *Phyllostachys*. PBM014 and PBM025 served as species-specific alleles for the identification of interspecies hybrids of *Sinobambusa tootsik* × *Pleioblastus distichus*, *Sasa tokugawana* × *S. borealis* and *Pleioblastus simoni* × *Phyllostachys praecos* (Lu et al. 2009). These results demonstrate that microsatellite markers (especially PBM014 etc.) are ideal markers for bamboo hybrid identification as reported in poplar (Rajora and Rahman 2003) and wheat-barley (Malysheva et al. 2003).

Compared with previous methods involving the construction of genomic DNA libraries for *Bambusa arundinacea* (Nayak and Rout 2005) and an SSR-enriched library for *Ph. pubescens* (data unpublished), developing microsatellite markers from public database searches is simple, rapid, cost-effective and highly suited to practical applications in plant breeding.

Acknowledgments We gratefully acknowledge C Chen of Fujian Agriculture and Forestry University and G Liao of the Jiangxi Forestry Research Institute for providing the materials on provenances and hybrids, and K Zheng of the China National Rice Research Institute for reviewing the manuscript and providing helpful suggestions. This work was financially supported by grants from the National Natural Science Foundation of China (grant no. 30371181 to DQ Tang) and the Talents Program of the Natural Science Foundation of Zhejiang Province (grant no. R303420 to DQ Tang).

References

Barkley NA, Newman M, Wang ML, Hotchkiss MW, Pederson GA (2005) Assessment of the genetic diversity and phylogenetic relationships of a temperate bamboo collection by using transferred EST-SSR markers. Genome 48:731–737

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of Arabidopsis. Genomics 19:137–144

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580

Buteler MI, Labonte DR, Jarret RL, Macchiavelli RE (2002) Microsatellite-based paternity analysis in polyploid sweet potato. J Am Soc Hort Sci 127:392–396

Chen C, Liang Y, Qiu E, Fan, Xie J (2001) A Study on synthetic selection for multi-shape and properties of *Phyllostachys pubescens*. Sci Silv Sin 37:18–23 (in Chinese with English abstract)

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci 160:1115–1123

Das M, Bhattacharya S, Pal A (2005) Generation and characterization of SCARs by cloning and sequencing of RAPD products: a strategy for species-specific marker development in bamboo. Ann Bot 95:835–841

Doyle JJ, Doyle JL (1987) A rapid isolation procedure for small quantities of fresh leaf materials. Phytochem Bull 19:11–15

Friar E, Kochert G (1991) Bamboo germplasm screening with nuclear restriction fragment length polymorphisms. Theor Appl Genet 82:697–703

Friar E, Kochert G (1994) A study of genetic variation and evolution of *Phyllostachys* (Bambusoideae: Poaceae) using nuclear restriction fragment length polymorphisms. Theor Appl Genet 89:265–270

Fu J (2001) Chinese moso bamboo: its importance. Bamboo 22(5):5–7

Goldstein DB, Clark AG (1995) Microsatellite variation in North American population of *Drosophila melanogastor*. Nucleic Acid Res 23:3882–3886

Gui Y, Sheng W, Quan LY, Zhou CP, Long SB, Zheng HJ, Jin L, Zhang XY, Ma NX, Fan LJ (2007) Genome size and sequence composition of moso bamboo: a comparative study. Sci China Ser C Life Sci 50(5):700–705

Gupta PK, Varshnet RK, Sharma PC, Ramesh B (1999) Molecular markers and their applications in wheat breeding. Plant Breed 118:369–390

Isagi Y, Shimada K, Kushima H, Tanaka N, Nagao A, Ishikawa T, Onodera H, Watanabe S (2004) Clonal structure and flowering traits of a bamboo [*Phyllostachys pubescens*

(Mazel) Ohwi] stand grown from a simultaneous flowering as revealed by AFLP analysis. Mol Ecol 13:2017–2021
Jakupiak JP, Wells RD (1999) Genetic instabilities in (CTG.CAG) repeats occur by recombination. J Biol Chem 274:23468–23479
Janzen DH (1976) Why bamboos wait so long to flower. Ann Rev Eco Syst 7:347–391
La Rota M, Kantety RV, Yu JK, Sorrells ME (2005) Non-random distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat and barley. BMC Genomics 6(23):1–12. doi:10.1186/1471-2164-6-23
Lai CC, Hsiao JY (1997) Genetic variation of *Phyllostachys pubescens* (Bambusoideae, Poaceae) in Taiwan based on DNA polymorphisms. Bot Bull Acad Sin 38:145–152
Li S, Yin T, Zou H, Ding Y, Huang M (2002) Preliminary study on molecular systematics of bamboo by SSR primers derived from rice. Sci Silv Sin 38:42–48 (in Chinese with English abstract)
Lin XC, Ruan XS, Lou YF, Guo XQ, Fang W (2009) Genetic similarity among cultivars of *Phyllostachys pubescens*. Plant Syst Evol 277:67–73
Liu Y, Gui R, Yu Y, Chen C, Fang W (2008) A preliminary study on the physical and mechanical properties of different provenances of moso bamboo. J Bamboo Res 27:50–54 (in Chinese with English abstract)
Loh JP, Kiew R, Set O, Gan LH, Gan YY (2000) A study of genetic variation and relationships within the bamboo subtribe Bambusinae using amplified fragment length polymorphism. Ann Bot 85:607–612
Lu JJ, Yoshinaga K, Fang W, Tang DQ (2009) Identification of the hybrid bamboo F1 by SSR markers. Sci Silv Sin 45(3):29–34 (in Chinese with English abstract)
Malysheva L, Sjakste T, Matzk F, Roder M, Ganal M (2003) Molecular cytogenetic analysis of wheat-barley hybrids using genomic in situ hybridization and barley microsatellite markers. Genome 46:314–322
Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with non-repetitive DNA in plant genomes. Nat Genet 30:194–200
Nayak S, Rout GR (2005) Isolation and characterization of microsatellites in *Bambusa arundinacea* and cross species amplification in other bamboos. Plant Breed 124:559–602
Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol Gen Genet 252:597–607
Parida SK, Raj Kumar KA, Dalal V, Singh NK, Mohapatra T (2006) Unigene derived microsatellite markers for the cereal genomes. Theor Appl Genet 112:808–817
Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222
Pupko T, Graur D (1999) Evolution of microsatellites in the yeast *Saccharomyces cerevisiae*: role of length and number of repeated units. J Mol Evol 48:313–316
Rajora P, Rahman H (2003) Microsatellite DNA and RAPD fingerprinting, identification and genetic relationships of hybrid poplar (*Populus* × *canadensis*) cultivars. Theor Appl Genet 106:470–477
Richards RI, Sutherland GR (1994) Simple repeat DNA is not replicated simply. Nat Genet 6:114–116
Rossetto M (2001) Sourcing of SSR markers from related plant species. In: Henry RJ (ed) Plant genotyping—the DNA fingerprinting of plant. CABI Publishing, pp 211–224
Ruan X, Lin X, Lou Y, Guo X, Fan W, Chen C (2008) Genetic diversity of *Phyllostachys pubescens* provenances by AFLP and ISSR. J Zhejiang For Sci Tech 28:29–33 (in Chinese with English abstract)
Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:2211–2215
Smit AFA, Hubley R, Green P (1996–2004) RepeatMasker Open-3.0. http://www.repeatmasker.org
Suyama Y, Obayashi K, Hayashi I (2000) Clonal structure in a dwarf bamboo (*Sasa senanensis*) population inferred from amplified fragment length polymorphism (AFLP) fingerprints. Mol Ecol 9:901–906
Tang DQ (2009) Genomic sequencing and its application for biological and evolutional research in bamboo. Bamboo J 26:12–17
Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471
Temnykh S, Park WD, Ayers N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712
Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch SR (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452
Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422
Watanabe M, Ueda K, Manabe I, Akai T (1982) Flowering, seeding, germination and flowering periodicity of *Phyllostachys pubescens*. J Japn For Soc 64:107–111
Weber JL (1990) Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. Genomics 7:524–530
Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225–235
Zhang S, Tang DQ (2007) A review on application of DNA marker in bamboos and its limits. J Bamboo Res 26(1):10–14 (in Chinese with English abstract)
Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P (2005) High transferability of bread wheat EST-derived SSRs to other cereals. Theor Appl Genet 111:677–687
Zhu Y, Queller DC, Strassman JE (2000) A phylogenetic perspective on sequence evolution in microsatellite loci. J Mol Evol 50:324–338