SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
PLANT MOLECULAR BIOLOGY
PLANT GENETICS

Received
3 October 2014

Accepted
29 December 2014

Published
26 January 2015

Correspondence and requests for materials should be addressed to Z.G. (gaozhimin@icbr.ac.cn)

* These authors contributed equally to this work.

Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus *Phyllostachys*

Hansheng Zhao*, Li Yang*, Zhenhua Peng, Huayu Sun, Xianghua Yue, Yongfeng Lou, Lili Dong, Lili Wang & Zhimin Gao

State Forestry Administration Key Open Laboratory on the Science and Technology of Bamboo and Rattan, International Center for Bamboo and Rattan, Beijing 100102, China.

Morphology-based taxonomy via exiguously reproductive organ has severely limitation on bamboo taxonomy, mainly owing to infrequent and unpredictable flowering events of bamboo. Here, we present the first genome-wide analysis and application of microsatellites based on the genome of moso bamboo (*Phyllostachys edulis*) to assist bamboo taxonomy. Of identified 127,593 microsatellite repeat-motifs, the primers of 1,451 microsatellites were designed and 1,098 markers were physically mapped on the genome of moso bamboo. A total of 917 markers were successfully validated in 9 accessions with ~39.8% polymorphic potential. Retrieved from validated microsatellite markers, 23 markers were selected for polymorphic analysis among 78 accessions and 64 alleles were detected with an average of 2.78 alleles per primers. The cluster result indicated the majority of the accessions were consistent with their current taxonomic classification, confirming the suitability and effectiveness of the developed microsatellite markers. The variations of microsatellite marker in different species were confirmed by sequencing and *in silico* comparative genome mapping were investigated. Lastly, a bamboo microsatellites database (http://www.bamboogdb.org/ssr) was implemented to browse and search large information of bamboo microsatellites. Consequently, our results of microsatellite marker development are valuable for assisting bamboo taxonomy and investigating genomic studies in bamboo and related grass species.

As one of the most important non-timber forest resources and fastest-growing plant in the world, bamboos belong to a group of flowering and evergreen plants, classified in the grass family Poaceae, subfamily Bambusoideae, tribe Bambusease. Because of having a strong carbon fixation capability and a high strength-to-weight ratio, like natural woody, bamboo had received much attention in the ecological and economic aspects. More than 1,250 species, belonging to 75 genera, are reported worldwide[1].

The vegetative features are quite essential for identification of the bamboo members. However, as one of important and special characteristics in bamboo, infrequent and unpredictable flowering events, such as a long flowering cycle of up to 120 years and flowering once before culm death, severely restricts the opportunity to study fresh reproductive materials. Even if their herbarium samples were available, enough morphological resolution was lack, causing confusion in the real field condition[2]. Therefore, traditional morphological taxonomy based on reproductive characteristics has to face great difficulties in bamboo. On the other hand, currently morphology-based taxonomy for bamboo, mainly depending on vegetative characteristics such as culm or culm-sheath, is hard to produce a confirmed taxonomic demarcation and approach for bamboo. Because some variations may be induced in culm, leaf or rhizome of bamboo during different development stages or under different environments. In particular, with the increasing number of identified bamboo species, a few of confused phenomena based on morphological taxonomy were emerged, such as homonym and synonym[3,4].

To address this issue properly and assist bamboo taxonomy, considerable progress had been made in the field of modern molecular techniques[5–10]. As one of the most powerful genetic markers in biology, simple sequence

Table 1 | Distribution of microsatellites involved genome size, microsatellite frequency and microsatellite type in six surveyed plant species

			Identified SSRs			
			Single SSRs (2–6 bp repeat motif)		Compound SSRs	
Plant species	Genome size (Mb)	SSR/Mb	Number	%	Number	%
Phyllostachys edulis	2051.7	62.2	106,582	83.53	21,011	16.47
Zea mays	2066.4	52.1	94,683	87.88	13,059	12.12
Oryza sativa	374.5	165.5	49,505	79.88	12,472	20.12
Sorghum bicolor	738.5	91.1	57,016	84.70	10,295	15.30
Brachypodium distachyon	271.9	76.5	17,865	85.93	2,924	14.07
Arabidopsis thaliana	119.7	135.5	12,259	75.59	3,958	24.41

repeats (SSRs), also known as microsatellites, refer to DNA sequences of 1–6 bp in length that are tandem repeated a variable number of times[11]. Generally, the repeat-motifs with more than mono-nucleotide were selected to develop molecular markers[12,13]. Compared with other molecular marker, the nature of microsatellites provides a number of advantages such as simple PCR screening, co-dominant, relatively abundant and inherent potential for variation. Therefore, microsatellites were considered to be better suited for construction of high-density linkage maps[14], QTL mapping[15], genetic diversity analysis[16], cultivar identification[17] and marker-assisted selection[18]. Meanwhile, microsatellites with a high degree of transferability among species have been widely utilized for comparative genetics and taxonomy research[19,20].

Nevertheless, traditional development of microsatellite markers is a time-consuming, money-cost process and limited utilization[21]. The previous results[22,23] demonstrated that a less number of bamboo microsatellites markers were identified with few polymorphism, due to lacking of a large number of sequence data. Therefore, a genome-wide analysis of bamboo microsatellite on the basis of the draft genome of moso bamboo (*Phyllostachys edulis*) was performed[24], to effectively expand our understanding and widely detect polymorphism of microsatellites in bamboo. Thus, we had identified the whole genome-based microsatellite loci and developed some microsatellite markers. Moreover, a selected set of validated and physically mapped markers were used in evaluating polymorphic and diversity potential in 78 accessions belonged to the genus *Phyllostachys*, which is one of the most important genus of bamboo including more than 50 species and accounting for 90% of commercial bamboos in the Bambusoideae. Genetic relationship was revealed among the accessions of *Phyllostachys* by cluster analysis to identify the tallies with morphological taxonomy. Furthermore, the genetic basis of sequence length variation by sequencing and *in silico* comparative genome mapping were investigated as well. Lastly, the module of bamboo microsatellite database was implemented in order to provide worldwide researchers with a diverse and powerful search and browse for bamboo microsatellite information.

Results

Microsatellites frequency analysis. Owing to mono-nucleotide repeat-motif having homo-polymorphism, it was rarely employed as microsatellite markers based on polymorphism analysis[12,13]. Thus, we identified a total of 127,593 microsatellites with more than mono-nucleotide repeat-motif in moso bamboo genome based on bioinformatics analysis (Table 1). Besides the genome of moso bamboo, a genome-wide investigation of microsatellite based on the 5 important plants (*Zea mays*, *Oryza sativa*, *Sorghum bicolor*, *Brachypodium distachyon* and *Arabidopsis thaliana*) were performed by utilizing their latest genomic sequences and annotations. The result demonstrated the highest microsatellites frequency was found in *O. sativa* (165.5/Mb) which was the minor size of genome, whereas the lower microsatellites frequency was in *Z. mays* (52.1/Mb) which owned the biggest size of genome. Moreover, among the six genomes analyzed, the genome size of moso bamboo (2,051.7 Mb) was ~17 times of that of *A. thaliana* (119.7 Mb), but the frequency of *A. thaliana* (135.5/Mb) was ~2-fold of that of moso bamboo (62.2/Mb). Compared with these of *Z. mays*, the genome size of moso bamboo was almost identical to that of *Z. mays* (2,066.4 Mb), but the number of microsatellites in moso bamboo was yet increased by 65.2% than that of *Z. mays*.

The microsatellite possessed more than one microsatellite locus contiguously with intervening non-repeat sequence less than 100 bp were identified as compound formation in this study. Compound microsatellite is a special aggregation of microsatellites in which two or more individual microsatellites are found directly adjacent to each other. All identified microsatellites have a minimum length of 15 bp. Whether two or more adjacent microsatellites account as a compound microsatellite depends on the distance separating these microsatellites[25]. Based on the definition of interrupted length was 1–100 bp in this study, the distribution of interrupted length was analyzed in the selected plants (see Supplementary Table S1 online). The result showed that the mainly interrupted length was concentrated on 3–16 bp. The most of interrupted length was present in 10 bp of moso bamboo, accounted for 6.4%. Moreover, the number of interrupted length was most in moso bamboo, followed by *Z. mays*, while that of *A. thaliana* was slight, revealing the lower complexity of genome than those of moso bamboo and *Z. mays*.

Frequency of motif repeat and length distribution. As a relative frequency based on different microsatellite repeat motifs and length repeat, the major statistical result of moso bamboo was shown in Supplementary Table S2. For the di- and tri- nucleotide repeats, the AG/CT and AT/AT were the dominant repeats compared with AC/GT and CG/CG, and the AAG/CTT was the most abundant, followed by CCG/CGG and AGG/CCT. Moreover, the frequency of ACT/AGT and ACC/CGT repeats were at lower level, accounted for only 2.98% and 3.65%, respectively.

As shown in Figure 1a, besides mono-nucleotide repeat-motif, the length of other microsatellite repeat-motifs was counted in this study and classified into 10 groups. According to the previous study[26], the length of microsatellites with more than 20 bp was defined as class I or hypervariable markers, while ones ranged from less than or equal to 20 bp was defined as class II or potentially variable markers. Comparative analysis of the 6 selected plants demonstrated class II occupied a large share, evenly accounted for up to 85%. The number of microsatellites correspondingly decreased with the increasing of microsatellites length.

Microsatellites distribution in different genomic regions. Since microsatellite distribution is the result of selection pressure during evolution[27], microsatellites in different genomic regions have different features and may perform varied functions[28–30]. Here, we analyzed the distribution of microsatellites in genomic regions based on latest genomic sequences and annotations in surveyed plants[3]. According to the location and annotation of genome, genomic

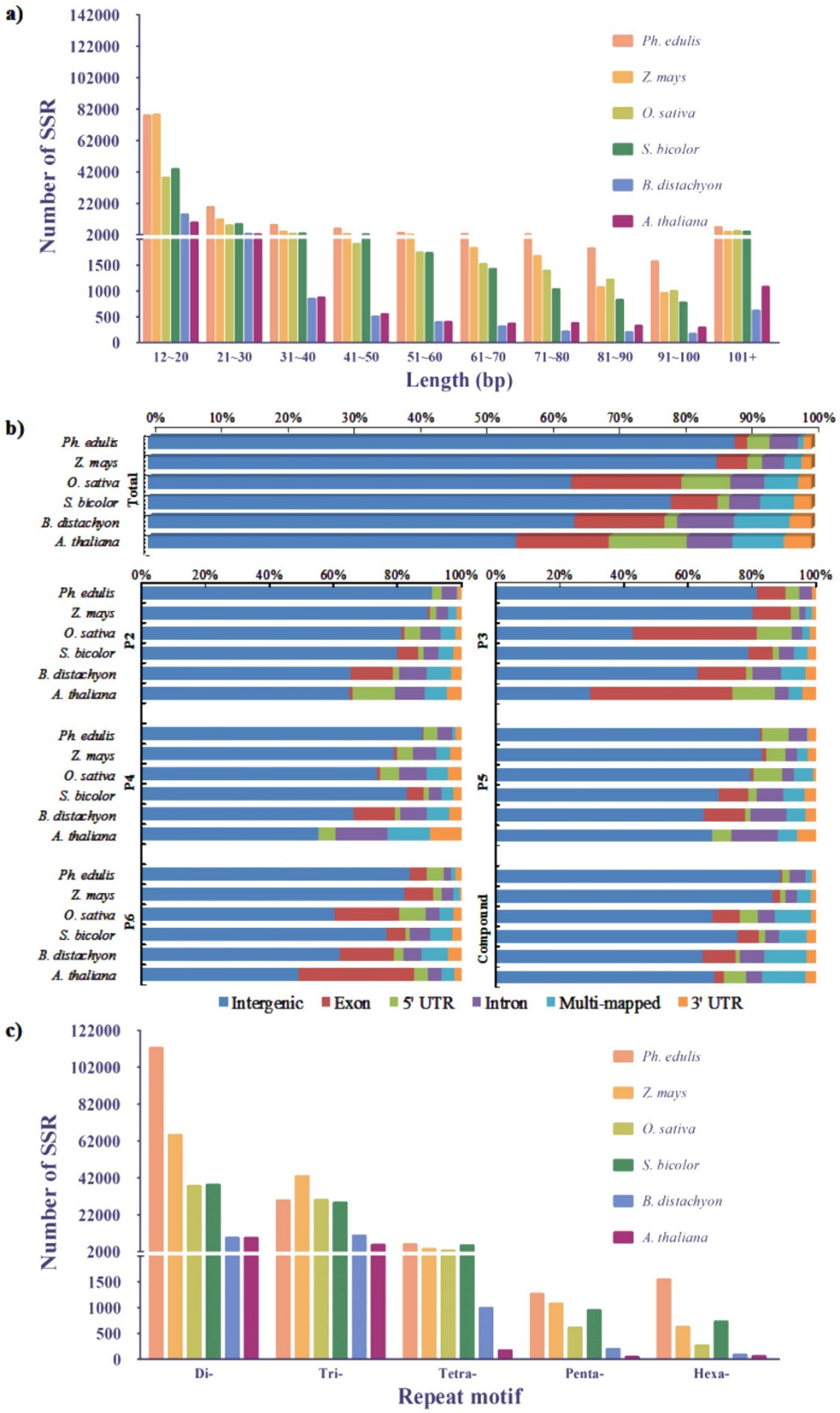

Figure 1 | Distribution of microsatellite repeat motif, length and different genomic regions. (a) Microsatellite length distribution. The x-axis represents the nucleotide length of microsatellites. The y-axis indicates the number of microsatellites with different length in the six selected plant species. (b) The x-axis indicates the proportion of microsatellites with various repeat motifs. The y-axis represents different repeat motif in the six selected plant species. P2: di-nucleotide repeats; P3: tri-nucleotide repeats; P4: tetra-nucleotide repeats; P5: penta-nucleotide repeats; P6: hexa-nucleotide repeats; compound: compound microsatellite. (c) Total number of each repeat motif. The x-axis indicates repeat motif with di- to hexa- nucleotides. The y-axis represents the number of microsatellites with various repeat motifs. The different plant species were marked with different color.

Table 2 | Universal and polymorphic potential of 1,098 microsatellite markers

Type of SSR	Number of SSR	Universality		Polymorphism	
		Number	Probability (%)	Number	Probability (%)
Di-nucleotide	581	474	81.6	193	40.7
Tri-nucleotide	442	377	85.3	130	34.5
Tetra-nucleotide	35	31	88.6	20	64.5
Penta-nucleotide	22	19	86.4	8	42.1
Hexa-nucleotide	9	7	77.8	5	71.4
Compound	9	9	100	9	100
Total	1,098	917	/	365	/

sequences in the 6 surveyed plant species were classified into 6 regions. Besides 5′ UTR (Untranslated Regions), intron, exon, 3′ UTR and intergenic region, the multi-mapped region was identified via bioinformatics analysis. For instance, one sequence of microsatellite was able to align two or more regions.

As shown in Figure 1b, according to the analysis of sequences, the result illustrated most sequences were commonly mapped onto intergenic regions. The remaining parts were aligned to other regions. For example, the second highest proportion belonged to intron region in moso bamboo, followed by 5′ UTR region. Whereas the second one appeared in exon region of *O. sativa*, followed by intron region.

To comprehensively unveil the detailed feature on different type of repeat, distribution based on each type of microsatellites was analyzed. In di-nucleotide repeat, the higher proportion of 5′ UTR and exon regions emerged in *A. thaliana* and *B. distachyon*, respectively. In tri-nucleotide repeats, exon region with tri-nucleotide repeat was rather abundant in overall analyzed plants. And the top mapping region was not intergenic regions, but rather exon regions in *A. thaliana*. Excluding tri-nucleotide repeat, exon regions accounted for larger proportion in hexa-nucleotides repeats. The highest proportion of multi-mapped region was found in compound type of motif, because compound microsatellites with longer nucleotides sequences made it possible to align more regions.

Universal and polymorphic potential of microsatellite markers. Polymorphism of microsatellite marker is the basis to develop molecular markers that are utilized in genetic mapping broadly[31]. Here, we have employed experimental method to detect polymorphism of microsatellite markers in bamboo. Firstly, of all identified microsatellites, 1,451 primer pairs in gene region were designed and 1,098 markers were physically mapped on the genome of moso bamboo. To evaluate universality and polymorphism of developed microsatellite markers, *Ph. edulis*, *Ph. sulphurea* var. *viridis*, *Ph. violascens*, *Ph. nigra*, *Ph. aureosulcata*, *Ph. propinqua*, *Ph. sulphurea*, *Ph. atrovaginata* and *Ph. heteroclada* were amplified using 1,098 primer pairs (see Supplementary Figure S1 online). The result showed 917 of the selected markers produced clear and stable bands with the expected size, among which 365 markers had polymorphism. The universality and polymorphism probabilities of different kind of microsatellites were predicted (Table 2). As effectively molecular marker, microsatellite marker was valuable for identifying relative species. For example, the amplification bands of locus PhEMS-653 were only produced in the *Ph. edulis* and its variant and cultivars (see Supplementary Figure S2a online), therefore the locus was specific for this species and its subordinates. Thus, retrieved from microsatellites validated, 23 primer pairs with high amplification effect were subsequently used for polymorphism analysis among 78 accessions. The 64 alleles ranged from 2 to 5 per loci were detected by 23 SSRs, with an average value of 2.78. The maximum number of alleles was 5 observed in locus PhEMS-561 (see Supplementary Figure S2b online), and then followed by 4 alleles detected in locus PhEMS-791 (see Supplementary Figure S2c online).

Phylogenetic analysis of *Phyllostachys* accessions. UPGMA (Unweighted Pair Group Method with Arithmetic Mean) analysis of microsatellite data showed that 78 accessions were clustered into 3 classes (Figure 2) based on the genetic distance (see Supplementary Figure S3 online). The accessions in Class I and Class II were belonged to Phyllostachys section, while those in Class III were belonged to Heterocladae section, in which the cultivars, variants or forma from *Ph. violascens*, *Ph. vivax*, *Ph. edulis*, *Ph. sulphurea*, *Ph. aurea*, *Ph. nigra*, and *Ph. aureosulcata* were together clustered, indicating the majority of the accessions in this clustered result were consistent with their current taxonomic classification[32]. Class I was the smallest cluster only including 4 accessions (*Ph. propinqua*, *Ph. propinqua* f. *lanugtginosa*, *Ph. nuda* and *Ph. sp.*). The genetic distance between *Ph. nuda* and *Ph. propinqua* was 0.2143 which supported that they were different species with having or not bristles on the back of sheath and coronatines on the internodes of young culm. Class II was the largest group and separated into five subclasses, among which subclass IIa was the group of *Ph. violascens*, including its variant (*prevernalis*) and cultivars ('Notata', 'Viridisulcata', 'Flavistriatus' and 'Anhuiensis'). *Ph. violascens* f. *prevernalis* was the forma of *Ph. violascens* noted in the Flora of China (Chinese), its middle of internode slightly slimed in morphology and shoot duration was earlier half a month than those of *Ph. violascens*, and the genetic distance between them was 0.111, which supported that they could not be considered as the same one. In addition, *Ph. tianmuensis*, *Ph. primotina*, *Ph. ncarnate*, *Ph. glabrata*, *Ph. dulcis* and *Ph. nigella* were also clustered together within the group of *Ph. violascens*, which their close relationship agreed with the previously morphological classification. The accessions clustered in subclass IIb, IIc and IId were agreed to their current taxonomic classification[32].

The genetic analysis based on sequence length variation. On the basis of differences in the copy of microsatellite repeats, allelic length variations were found[12]. The mutations, such as insertion, deletion, or translocations, in the regions flanking microsatellite may be considered as polymorphism in the sequence. Therefore, to investigate whether the PCR really amplified the targets, the PCR products were subcloned into T-easy vector and sequenced. For example, the results of sequences alignment showed the locus PhEMS-855 amplified variant alleles from 120 to 147 bp (Figure 3), which revealed mixed type of allelic distribution with the variable number of repeats in the microsatellite motifs (AT)n accompanied by several point mutations like insertion/deletion or substitution mutations.

***In silico* comparative genome mapping between moso bamboo and other surveyed species.** Based on the experimental results and the previous studies[12,13], of 1,098 validated microsatellite, 984 microsatellite markers on the genome of moso bamboo was successfully mapped on four sequenced Poaceae members (Table 3, Figure 4a1-d1). To display syntenic relationships of genome-widely

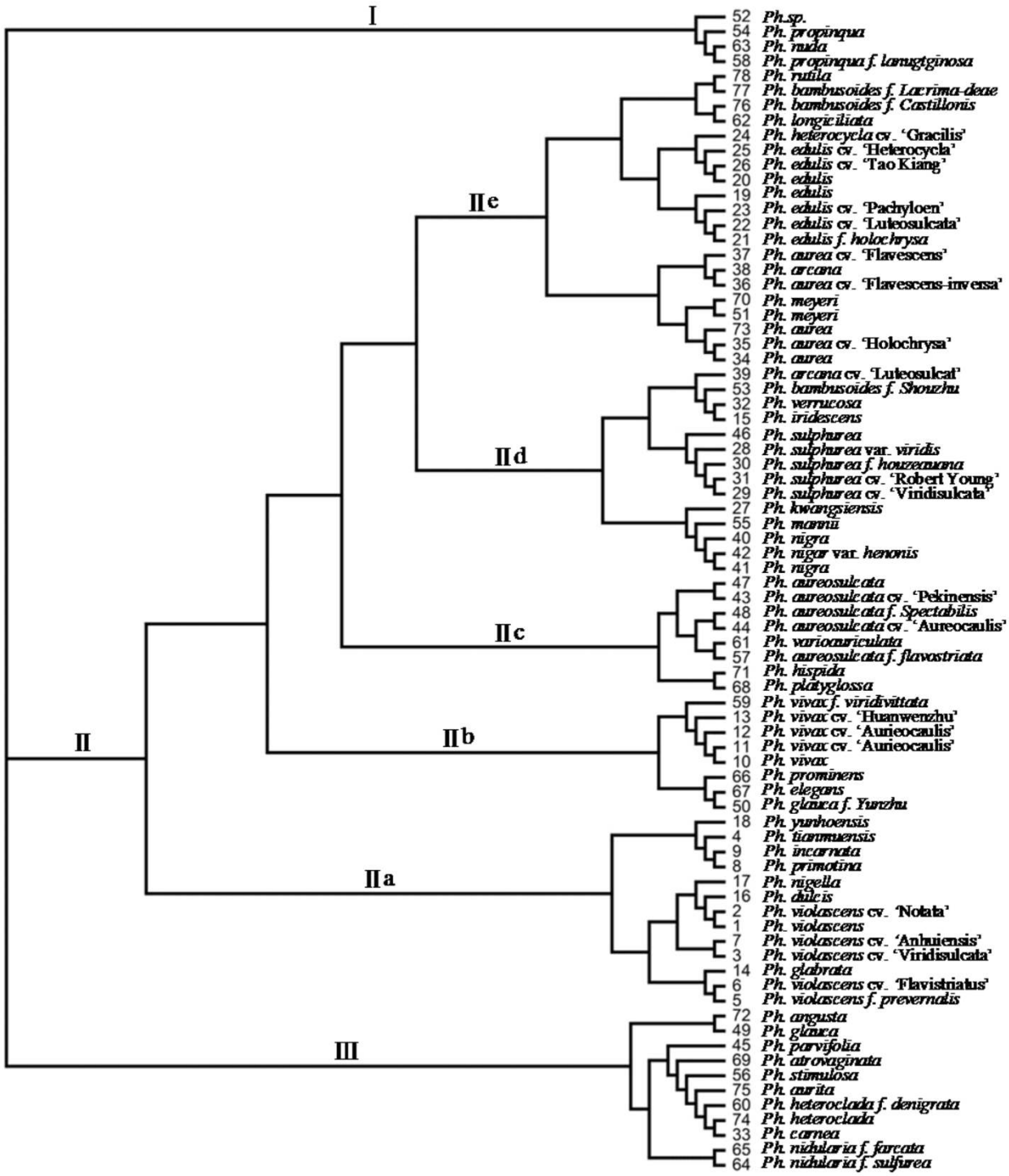

Figure 2 | Phylogenetic analysis of 78 bamboo accessions in genus *Phyllostachys* based on microsatellite data. No. 1–78 represent different bamboo accessions, detailed in Supplementary table S3.

microsatellites marker-based via *in silico* comparative analysis, the scaffolds of moso bamboo were clustering into 6 groups via aligning microsatellites to surveyed genomes, because the draft genome of moso bamboo is scattered and hard to assign scaffolds to chromosomes by now. The results demonstrated the significant proportion of orthology and syntenic relationship of microsatellite marker with *Z. mays* (116, 11.79%), *O. sativa* (185, 18.80%), *S. bicolor* (126, 12.80%), and *B. distachyon* (180, 18.29%). Meanwhile, based on the bioinformatics analysis, of 127,593 identified microsatellites, a total of 101,683 microsatellite markers with available for primers was mapped on four sequenced Poaceae members and their physical location on genome of other surveyed plants were compared (Table 3, Figure 4a2–d2). Detailed results on microsatellites marker-based comparative mapping were in Supplementary Table S4–S7, which demonstrated a number of moso bamboo microsatellite markers significantly matched with different chromosomes of *Z. mays*, *O. sativa*, *S. bicolor* and *B. distachyon*.

Constructed a database of microsatellites for moso bamboo. To facilitate access and utilization of microsatellite data in moso bamboo, the entire microsatellites had been organized into an online database that was freely accessible via the website (http://

Figure 3 | Multiple sequence alignment of PhEMS-855 demonstrating the presence of microsatellite repeat motif in *Phyllostachys edulis* and other species of genus *Phyllostachys*. The analysis indicates multiple point mutations and insertion/deletions happened in a large number of repeat motifs among different species.

www.bamboogdb.org/ssr), which provides diverse and powerful search tools for browsing and searching in the database (Figure 5). The search function with a Boolean search provided user with specific and combined query options by various characteristics, such as location information, repeat type, copy number and size of microsatellite (Figure 5a). As a valuable kind of molecular marker, initial microsatellites development, such as a large number of primer design, was costly, time consuming and labor intensive[21]. To facilitate primer design, five candidate primer pairs for each microsatellite were provided and stored in the database (Figure 5b).

Discussion

A larger number of microsatellites with single formation played essential roles in molecular markers for relationship, population and so on[33,34]. In this study, microsatellites with 2–6 bp (di-, tri-, tetra-, penta- and hexa- nucleotides) repeat motif were chiefly focused. In Figure 1c, the result of microsatellite repeat motif demonstrated nearly 80% of microsatellite repeat motifs were the di- and tri-nucleotide repeats. Of these, di-nucleotide repeats was the most abundant one. Besides, a small remaining share was contributed by tetra-, penta- and hexa-nucleotide repeats, in which penta- and hexa-nucleotide repeats were almost equal. In the six surveyed plant species, their distribution of microsatellite repeat motifs was similar. For instant, di-nucleotide repeats dominated over other type of repeats. The proportion of microsatellite repeat motifs (from di- to hexa-nucleotide repeats) indicated the proportion of microsatellites obviously decreased with increasing in repeat motif length, except in *B. distachyon* which the frequency of di-nucleotide repeats was lower than that of the tri-nucleotide repeats. Moreover, the total proportion of tetra-, penta- and hexa- nucleotide repeats was less than 10 percent of whole repeat motifs.

The occurrence of microsatellite frequencies, microsatellite distribution and length variation is the result of selection pressure during evolution[27]. However, the molecular mechanism for microsatellites is still a subject of controversy. Main theories, containing replication slippage and unequal crossing-over, suggested microsatellites in dif-

Table 3 | A summary of microsatellite comparative mapping showing syntenic relationships of the genome between moso bamboo and *Zea mays*, *Oryza sativa*, *Sorghum bicolor* and *Brachypodium distachyon*

Scaffold Group of Phyllostachys edulis	*Zea mays*		*Oryza sativa*		*Sorghum bicolor*		*Brachypodium distachyon*	
	Experiment	Prediction	Experiment	Prediction	Experiment	Prediction	Experiment	Prediction
PSG1	Chr.1 (25, 37.88%)	Chr.5 (244, 28.27%)	Chr.3 (42, 42.86%)	Chr.2 (440, 33.79%)	Chr.1 (42, 56.57%)	Chr.4 (314, 34.70%)	Chr.1 (47, 48.96%)	Chr.3 (644, 55.00%)
PSG2	Chr.3/8 (10, 32.29%)	Chr.1 (291, 34.04%)	Chr.1 (30, 62.50%)	Chr.3 (515, 38.40%)	Chr.3 (17, 56.67%)	Chr.1 (434, 46.82%)	Chr.2 (51, 96.23%)	Chr.1 (921, 80.51%)
PSG3	Chr.2 (5, 71.43%)	Chr.3 (232, 35.97%)	Chr.4 (9, 100.00%)	Chr.1 (547, 52.70%)	Chr.6 (5, 62.50%)	Chr.3 (399, 55.19%)	Chr.5 (8, 88.89%)	Chr.2 (771, 84.73%)
PSG4	Chr.4 (5, 45.45%)	Chr.2 (140, 54.90%)	Chr.11 (10, 71.43%)	Chr.4 (302, 79.47%)	Chr.5/6 (2, 25.00%)	Chr.6 (186, 71.81%)	Chr.4 (18, 100.00%)	Chr.5 (240, 81.36%)
PSG5	Chr.6 (1, 100.00%)	Chr.7 (76, 76.77%)	Chr.9 (6, 100.00%)	Chr.9 (47, 54.65%)	Chr.10 (6, 100.00%)	Chr.2 (122, 91.04%)	Chr.1 (2, 66.67%)	Chr.4 (39, 59.10%)
PSG6	N.A.	Chr.3 (3, 30.00%)	N.A.	Chr.12 (6, 60.00%)	N.A.	Chr.8 (80, 100.00%)	Chr.4 (1, 100.00%)	Chr.4 (4, 80.00%)

N.A. = No Available.

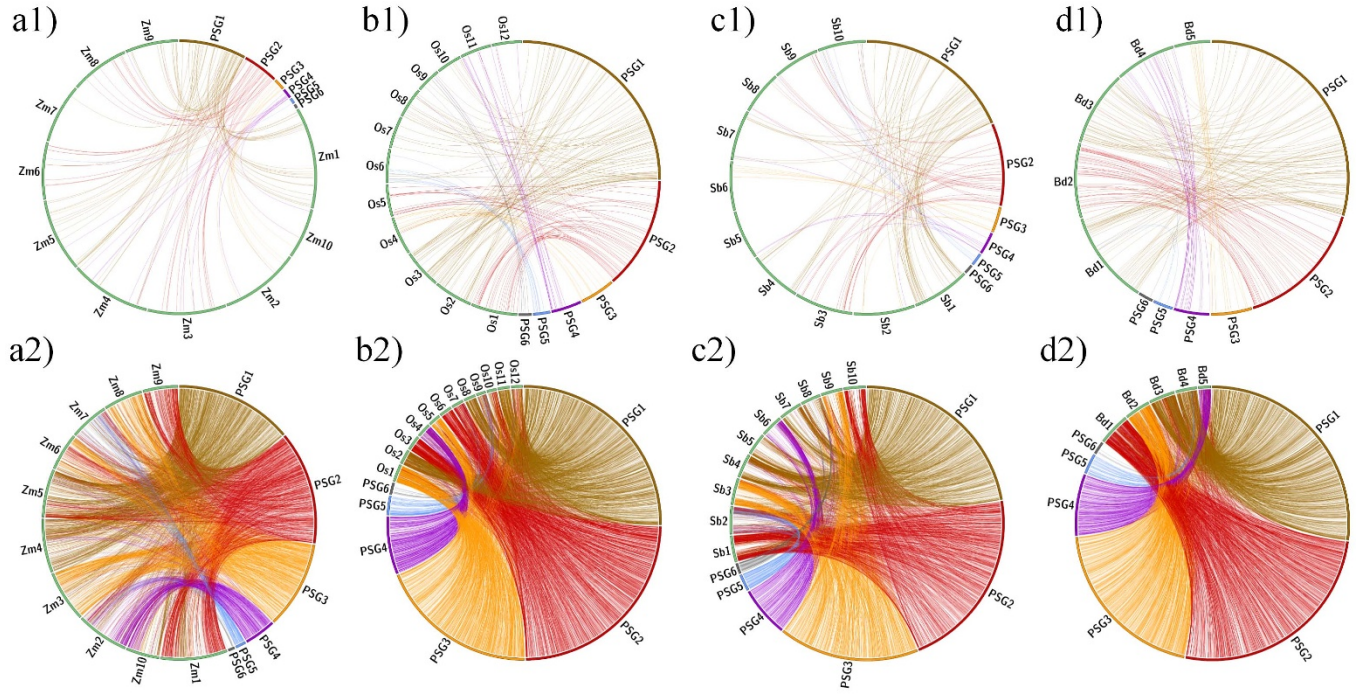

Figure 4 | Genome relationship of moso bamboo with other plant species. Based on the experimentally physical mapping, syntenic relationship of moso bamboo genome with (a1) *Zea mays*, (b1) *Oryza sativa*, (c1) *Sorghum bicolor* and (d1) *Brachypodium distachyon* chromosomes using 984 physically mapped moso bamboo microsatellite markers. Maximum syntenic relationships of the genome of moso bamboo with *Oryza sativa* chromosomes based on experimental microsatellite markers were apparent. Besides, based on the bioinformatics comparative mapping, syntenic relationship of moso bamboo genome with (a2) *Zea mays*, (b2) *Oryza sativa*, (c2) *Sorghum bicolor* and (d2) *Brachypodium distachyon* chromosomes using 101,683 mapped moso bamboo microsatellite markers. Similarly, maximum syntenic relationships of the genome of moso bamboo with *Oryza sativa* chromosomes based on predicted microsatellite markers were apparent as well. Zm1-10 used to be short for the chromosome 1–10 of *Zea mays*; Os1–12 used to be short for the chromosome 1–12 of *Oryza sativa*; Sb1–10 used to be short for the chromosome 1–10 of *Sorghum bicolor*; Bd1-1-5 used to be short for the chromosome 1–5 of *Brachypodium distachyon*; PSG1–6 used to be short for the group of scaffold 1–6 of *Phyllostachys edulis*.

ferent genomic regions have different features and may play various roles[28–30]. For example, microsatellite frequency in monocot coding regions is 2-fold of that of dicots[35,36]. Besides, another reason was that the large of scattered and discontinuous sequences may lead to low frequency of microsatellite in the relatively bigger size of genome with high complexity and incomplete assembling.

As a kind of molecular markers, di-nucleotide repeats were considered as the essential resource, because its mutant rate was the highest and the amount was abundant[37]. For di-nucleotide repeat category, AG/CT repeat was more frequent in *B. distachyon* and *Z. mays* with 50.72% and 47.74%, respectively, whereas AT/AT repeat were more abundant in *S. bicolor* and *A. thaliana* with 54.40% and 52.58%, respectively. Another noting that AG/CT and AT/AT repeat dominated over other di-nucleotide repeats in entire selected plant species. Moreover, the outcome of di-nucleotide indicated that GC-rich motifs were enrichment rather than AT-rich motifs.

Differentiated from di- nucleotide repeat motif, the remaining ones (from tri- to hexa- nucleotides) have a wide diversity in the distribution of repeat motif. For example, in tri-nucleotide repeats category, AAG/CTT repeat of moso bamboo was the abundant, accounted for 20.08%, followed by the CCG/CGG repeat (20.04%), whereas CCG/CGG repeat was plentiful in *O. sativa*, *B. distachyon* and *S. bicolor*. This result may be caused by the high GC content and codon bias in monocotyledons[38]. Subsequently, the frequency of tetra-, penta and hexa-nucleotide repeats was relatively at low level in all surveyed plants and the number of their motifs was slight.

As the development of microsatellite markers, the main resources derived from the cDNA and EST data owing to lacking of the genome data. cDNA-SSR and EST-SSR were on the basis of transcription region, high conversation and wide versatility. The comparative studies among cDNA-SSR, EST-SSR and genome-SSR depicted that the most of motif repeats and frequency were similar in moso bamboo[39,40]. However, there were some significantly differences between the size of examined sequences and the number of identified microsatellites (see Supplementary Table S8 online). For example, tri-nucleotide repeat was the dominant in EST-SSR and cDNA-SSR, while mono- and di-nucleotide repeats were abundant in genome-SSR. Moreover, the analysis indicated that the type and number of microsatellite repeats were more abundant in genome sequence. Taken together, it is necessary that developing microsatellite markers based on the whole genome of moso bamboo to comprehensively and effectively assisting bamboo taxonomy and systematically expand our understanding of microsatellite.

The taxonomy of bamboo has puzzled researchers because of bamboo reproductive characteristics. Especially in the genus *Phyllostachys*, some of same accessions may have different names, while some different accessions may have the same name. With the fast development of molecular biology, the molecular marker of bamboo was developed and used for bamboo genetic diversity and phylogenetic analysis which was helpful for the taxonomy of bamboo. ISSR and AFLP techniques were used to measure genetic diversity and similarity in cultivars of moso bamboo, three taxa including *Ph. kwangsiensis*, *Ph. nigra* var. *henonis* and moso bamboo were separated into three groups while the cultivars of moso bamboo were genetically similar[41]. Meanwhile, RAPD analysis was also capable in detecting bamboo species, although it was sometime difficult to distinguish similar species[6]. For example, the previous study on 23 bamboos of *Phyllostachys* based on RAPD method showed the cultivars of *Ph. vivax*, *Ph. violascens* and *Ph. aureosulcata* were not clustered into one clade[42]. There was a difference between this result of cluster and the traditional taxonomy based on classification of biological characteristics. To considerably investigate this problem, we

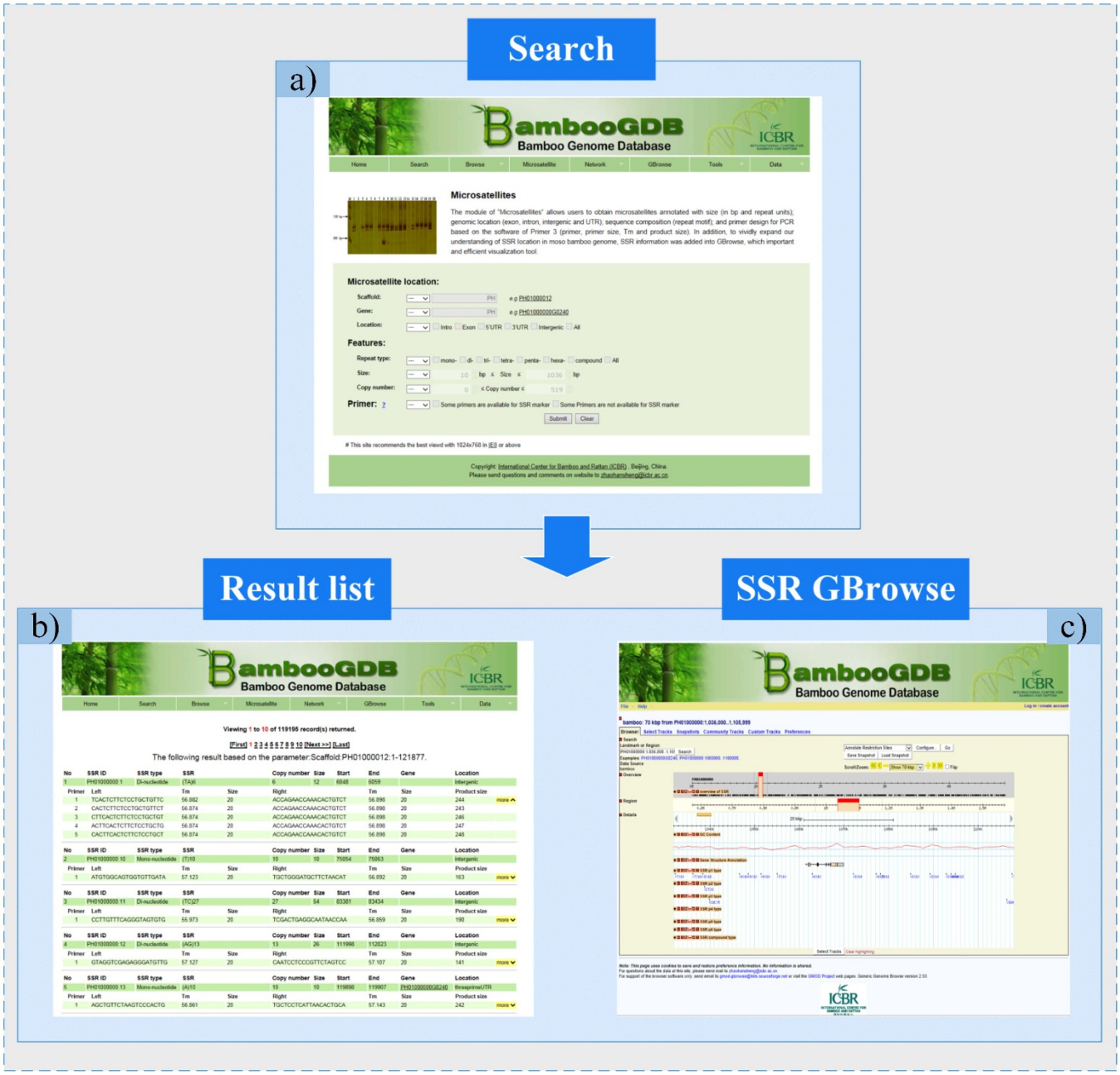

Figure 5 | Screenshot showing the pages of browse and results in bamboo microsatellite database. (a) The Boolean search was provided in the searching page of moso bamboo microsatellites. (b) The results page was included, such as microsatellite ID, microsatellite type, microsatellite position, microsatellite primers and so on. (c) The physical location of microsatellites in moso bamboo genome was displayed in GBrowse tool.

had adopted microsatellite technique based on genome sequence to detect the bamboos of the genus *Phyllostachys*.

Based on the development of microsatellite markers, the clustered result showed that the majority of the accessions were consistent with their current taxonomy. The accessions of synonyms were confirmed by the genetic distance of zero, such as accession 34 'Luo han zhu' and accession 73 'Ren mian zhu' belonged to the same one of *Ph. aurea*, accession 51 'Zhe jiang dan zhu' and accession 70 'Mao huan zhu' were the same one of *Ph. meyeri*. However, there were also few exceptions. For example, in subclass IIe, the cultivars of 'Holochrysa', 'Luteosulcata', 'Pachyloen', 'Gracilis', 'Heterocycla' and 'Tao Kiang' from *Ph. edulis* clustered together demonstrated close genetic relationship among them, the genetic distance was zero among 'Holochrysa', 'Luteosulcata' and 'Pachyloen', in addition to that among 'Gracilis', 'Heterocycla', 'Tao Kiang' and *Ph. edulis* was also zero, indicating that these cultivars had no genetic variation. The results also indicated that the 23 microsatellite loci selected were hardly used to distinguish the *Ph. edulis* from its cultivars. Another example, the accessions in Class III mainly contain the members of Heterocladae section except accession 49 'Dan zhu' and accession 74 'Shui zhu' which were classified into Phyllostachys section according to the reference[32]. Several of molecular techniques made similar taxonomy but including slight discrepancies. For example, *Ph. nigra* was classified in Phyllostachys section using microsatellite, which was consistent with the taxonomy defined by previous research[43]. However, *Ph. nigra* was grouped internal to Heterocladae section using the RFLP[44] and AFLP[45] methods.

These results may be caused by the following reasons, 1) Bias resulted from the cluster owing to the number of microsatellite molecular markers were not enough. 2) The resolution was relatively

lower for the products of amplification using non-denaturing PAGE electrophoresis, which resulted in allelic loci were not easily identified comparing with capillary electrophoresis. 3) Only the easily recognizable bands were recorded by artificial reading, the fuzzy ones were not adopted in data statistics. Therefore, the most abundant of allelic loci was only five in this study, while this number was nine in the population structure and phylogenetic relationships of *Ph. edulis* using capillary electrophoresis in diversity analysis[46]. These reasons for the difference with traditional taxonomy need to be further validated by expanding the number of microsatellite loci with the capillary electrophoresis method.

Due to complex genomic architecture, weak biology research on bamboo, infrequent sexual reproduction and the long periods of time between flowering intervals, the draft genome of moso bamboo is scattered with 277,278 scaffolds by now[24]. Therefore, it is difficult to assign scaffolds to chromosomes. Here, to vividly demonstrate physical location of microsatellites in genomic regions, information of microsatellite had been added into GBrowse, which is an efficient visualization tool[47]. In our GBrowse, various types of factors, including gene, coding sequence, messenger RNA, RNA-seq, were simultaneously in the context of genomic regions, expanding our understanding of microsatellites (Figure 5c). Users can access to the detailed feature page of corresponding entries from the browser.

Method

Plant materials, genomic DNA isolation and detection. Leaf samples of the accessions belonged to the genus *Phyllostachys* are collected from the base of Anhui Experimental Station of International Center for Bamboo and Rattan, Huangshan City in Anhui Province as well as Lin'an County, Hangzhou City in Zhejiang Province, respectively. The accessions were shown in Supplementary Table S3. These Latin name mainly referred to Flora of China[32]. Total genomic DNA was extracted with CTAB method described by Doyle and Doyle with minor modifications[48]. The quality and quantity of DNA were detected using Nanodrop 2000.

Identification of microsatellites. The MIcroSAtellite identification tool (MISA) (http://pgrc.ipk-gatersleben.de/misa/) was used to identify microsatellite occurrences in the genome sequence data (moso bamboo data from BambooGDB[49], others from Phytozome v10[3]). To identify the presence of microsatellites, the minimum repeat unit was defined as 6 for di, and 5 for tri-, tetra-, penta-, and hexa-nucleotides. The microsatellite position, number of different repeat types of microsatellite and length (motif bp × number of motif) was analyzed by a bespoke program written in MISA files[50].

Designing primers and validation of moso bamboo microsatellite markers. The primer pairs for flanking sequences of each unique microsatellite were designed automatically using the Primer3 (See Supplementary Table S9 online). Nine samples of *Ph. edulis*, *Ph. sulphurea* var. *viridis*, *Ph. violascens*, *Ph. nigra*, *Ph. aureosulcata*, *Ph. propinqua*, *Ph. sulphurea*, *Ph. atrovaginata* and *Ph. heteroclada* were selected to validate the suitability of microsatellite loci by PAGE. According to previous studies, PCR amplification was performed and PCR products were separated[28]. Seventy-eight accessions belonged to the genus *Phyllostachys* (see Supplementary Table S3 online) were used for microsatellite marker validation. UPGMA cluster analysis was performed based on the similarity matrix for 78 bamboo accessions using the PowerMarker program in the statistical analysis[51], then phylogeny was displayed in the TreeView software[52].

The variation analysis of microsatellite marker. Four primer pairs were selected to detect the microsatellite variation in different bamboo species. The genomic DNA samples of *Ph. edulis*, *Ph. violascens*, *Ph. aureosulcata*, and *Ph. heteroclada* were used as templates for PCR, and the PCR products were subcloned into T-easy vector and sequenced. The sequence alignment analysis was carried out by using DNAMAN version 5.0.

***In silico* comparative genome mapping between bamboo and other grass species.** Based on the previous studies[12,13], the flanking sequences of microsatellite marker loci were mapped the genome sequences of moso bamboo were BLAST searched with best hit against genome sequences of *Zea mays*, *Oryza sativa*, *Sorghum bicolor*, *Brachypodium distachyon* to analyze marker-based syntenic relationship between bamboo and other grass species. Before displayed with syntenic relationships in Circos software[53], the scattered scaffolds of moso bamboo were clustered into 6 groups via aligning microsatellites to surveyed genomes.

System design and implementation of microsatellite database. For moso bamboo, all data of microsatellite are stored and managed in a MySQL relational database. The website was developed in Centos Linux 6.4 environment using Java Platform Enterprise Edition, Apache Tomcat, Apache web server and so on. In addition, to vividly exhibit location of microsatellite in whole genome, microsatellite data was added in GBrowser after being converted into genome feature format (GFF). GBrowser uses MySQL as backend and was built following the configuration files provided by its developer (http://gmod.org).

1. Han, J. *et al.* Diversity of culturable bacteria isolated from root domains of moso bamboo (*Phyllostachys edulis*). *Microb. Ecol.* **58**, 363–373 (2009).
2. Singh, S. R. *et al.* Limitations, progress and prospects of application of biotechnological tools in improvement of bamboo-a plant with extraordinary qualities. *Physiol. Mol. Biol. Plants* **19**, 21–41 (2013).
3. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, 1178–1186 (2012).
4. Lai, G. A revision of some species of the genus *Phyllostachys* (Bambusoideae)(III)(*in Chinese*). *Guihaia* **22**, 390–393 (2002).
5. Shiobara, F., Ishi, T., Terachi, T. & Tsunewaki, K. Mitochondrial genome differentiation in the genus *Phyllostachys*. *JARQ (Japan)* **32**, 7–14 (1998).
6. Ramanayake, S., Meemaduma, V. & Weerawardene, T. Genetic diversity and relationships between nine species of bamboo in Sri Lanka, using random amplified polymorphic DNA. *Plant Syst. Evol.* **269**, 55–61 (2007).
7. Loh, J. P., Kiew, R., Set, O., Gan, L. H. & Gan, Y. Y. A study of genetic variation and relationships within the bamboo subtribe Bambusinae using amplified fragment length polymorphism. *Ann. Bot.* **85**, 607–612 (2000).
8. Marulanda, M. L., Márquez, P. & Londoño, X. AFLP analysis of *Guadua angustifolia* (Poaceae: Bambusoideae) in Columbia with emphasis on the coffee region. *Sci.Culture* **16**, 32–42 (2002).
9. Das, M., Bhattacharya, S. & Pal, A. 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 (2005).
10. Yang, H., An, M., Gu, Z. & Tian, B. Genetic diversity and differentiation of *Dendrocalamus membranaceus* (Poaceae: Bambusoideae), a declining bamboo species in Yunnan, China, as based on inter-simple sequence repeat (ISSR) analysis. *Int. J. Mol. Sci.* **13**, 4446–4457 (2012).
11. Jurka, J. & Pethiyagoda, C. Simple repetitive DNA sequences from primates: compilation and analysis. *J Mol Evol* **40**, 120–126 (1995).
12. Kumari, K. *et al.* Development of eSSR-markers in Setaria italica and their applicability in studying genetic diversity, cross-transferability and comparative mapping in millet and non-millet species. *PloS one* **8**, e67742 (2013).
13. Pandey, G. *et al.* Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [*Setaria italica* (L.)]. *DNA Res.* **20**, 197–207 (2013).
14. Sugita, T. *et al.* Development of simple sequence repeat markers and construction of a high-density linkage map of *Capsicum annuum*. *Mol. Breeding* **31**, 909–920 (2013).
15. Jeennor, S. & Volkaert, H. Mapping of quantitative trait loci (QTLs) for oil yield using SSRs and gene-based markers in African oil palm (*Elaeis guineensis* Jacq.). *Tree Genet. Genomes* **10**, 1–14 (2014).
16. Moretzsohn, M. C. *et al.* Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biol.* **4**, 11 (2004).
17. Ercisli, S., Ipek, A. & Barut, E. SSR marker-based DNA fingerprinting and cultivar identification of olives (*Olea europaea*). *Biochem. Genet.* **49**, 555–561 (2011).
18. Sui, C., Wei, J., Chen, S., Chen, H. & Yang, C. Development of genomic SSR and potential EST-SSR markers in Bupleurum chinense DC. *Afr. J. Biotechnol.* **8**, 8 (2009).
19. Marulanda, M. L., Lopez, A. M. & Claroz, J. L. Analyzing the genetic diversity of *Guadua* spp. in Colombia using rice and sugarcane microsatellites. *Crop Breed. Appl. Biot.* **7**, 43–51 (2007).
20. Chen, S. *et al.* Transferability of rice SSR markers to bamboo. *Euphytica* **175**, 23–33 (2010).
21. Iniguez-Luy, F. L., Voort, A. V. & Osborn, T. C. Development of a set of public SSR markers derived from genomic sequence of a rapid cycling *Brassica oleracea* L. genotype. *Theor. Appl. Genet.* **117**, 977–985 (2008).
22. Nayak, S. & Rout, G. R. Isolation and characterization of microsatellites in *Bambusa arundinacea* and cross species amplification in other bamboos. *Plant Breeding* **124**, 599–602 (2005).
23. Tang, D., Lu, J., Fang, W., Zhang, S. & Zhou, M. Development, characterization and utilization of GenBank microsatellite markers in *Phyllostachys pubescens* and related species. *Mol. Breeding* **25**, 299–311 (2010).
24. Peng, Z. *et al.* The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat. Genet.* **45**, 456–461, 461e1–2 (2013).
25. Wu, X., Zhou, L., Zhao, X. & Tan, Z. The analysis of microsatellites and compound microsatellites in 56 complete genomes of *Herpesvirales*. *Gene* **551**,103–109 (2014).
26. Temnykh, S. *et al.* 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 (2001).
27. Cho, Y. G. *et al.* Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **100**, 713–722 (2000).

28. Sonah, H. *et al*. Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in Brachypodium. *Plos one* **6**, e21298 (2011).
29. Levinson, G. & Gutman, G. A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203–221 (1987).
30. Schlötterer, C. & Tautz, D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20**, 211–215 (1992).
31. Agarwal, M., Shrivastava, N. & Padh, H. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep.* **27**, 617–631 (2008).
32. Flora of China Editorial Committee. *Flora of China* **Vol. 22** (Poaceae, Tribe Bambuseae). (Science Press, Beijing and Missouri Botanical Garden Press, St. Louis, 2006).
33. Joshi, S. P., Ranjekar, P. K. & Gupta, V. S. Molecular markers in plant genome analysis. *Curr. Sci. India* **77**, 230–240 (1999).
34. Schulman, A. H. Molecular markers to assess genetic diversity. *Euphytica* **158**, 313–321 (2007).
35. Han, X. Y. *et al*. Molecular characterization of tree peony germplasm using sequence-related amplified polymorphism markers. *Biochem. Genet.* **46**, 162–179 (2008).
36. Tóth, G., Gáspári, Z. & Jurka, J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* **10**, 967–981 (2000).
37. Chakraborty, R., Kimmel, M., Stivers, D. N., Davison, L. J. & Deka, R. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. USA* **94**, 1041–1046 (1997).
38. Morgante, M., Hanafey, M. & Powell, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* **30**, 194–200 (2002).
39. Peng, Z. *et al*. Genome-wide characterization of the biggest grass, bamboo, based on 10,608 putative full-length cDNA sequences. *BMC Plant Biol.* **10**, 116 (2010).
40. Zhang, Z., Guan, Y., Yang, L., Yu, L. & Luo, S. Analysis of SSRs information and development of SSR markers from Moso Bamboo (*Phyllostachys edulis*) ESTs. *Acta Horticulturae Sinica* **38**, 989–996 (2011).
41. Lin, X., Ruan, X., Lou, Y., Guo, X. & Fang, W. Genetic similarity among cultivars of *Phyllostachys pubescens*. *Plant Syst. Evol.* **277**, 67–73 (2009).
42. Suping, G., Yanjun, R. & Qibing, C. RAPD analysis of genetic relationships among the 23 ornamental bamboo Species in *Phyllostachys*. *Acta Horticulturae Sinica* **33**, 566 (2006).
43. Wang, J. *et al*. Genome-wide comparative analyses of microsatellites in papaya. *Tropical Plant Biol.* **1**, 15 (2008).
44. Friar, E. & Kochert, G. A study of genetic variation and evolution of *Phyllostachys* (Bambusoideae: Poaceae) using nuclear restriction fragment length polymorphisms. *Theor. Appl. Genet.* **89**, 265–270 (1994).
45. Hodkinson, T. R., Renvoize, S. A., Chonghaile, G. N., Stapleton, C. M. & Chase, M. W. A comparison of ITS nuclear rDNA sequence data and AFLP markers for phylogenetic studies in *Phyllostachys* (Bambusoideae, Poaceae). *J. Plant Res.* **113**, 259–269 (2000).
46. WX, J. Population structure and phylogenetic relationships of *Phyllostachys edulis* (Poaceae) in China. *Nanjing Forestry University doctoral thesis* (2013).
47. Donlin, M. J. Using the generic genome browser (GBrowse). *Curr. Protoc. Bioinform.*, **Chapter 9**, Unit 9 (2009).
48. Doyle, J. J. Isolation of plant DNA from fresh tissue. *Focus* **12**, 13–15 (1990).
49. Zhao, H. *et al*. BambooGDB: a bamboo genome database with functional annotation and an analysis platform. *Database* (Oxford) **2014**, bau006 (2014).
50. Thiel, T., Michalek, W., Varshney, R. K. & Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **106**, 411–422 (2003).
51. Liu, K. & Muse, S. V. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**, 2128–2129 (2005).
52. Page, R. D. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**, 357–358 (1996).
53. Krzywinski, M. *et al*. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

Acknowledgments

The work was supported by National Science and Technology Supporting Plan of the Twelfth Five-Year [2012BAD23B0504], the National Science Foundation of China [No. 31400557 and No. 31370588], and Fundamental Research Funds for International Center for Bamboo and Rattan [No. 1632013009].

Author contributions

H.Z. and Z.G. designed the experiments; H.Z. and L.Y. performed experiments; H.Z., L.Y., L.W., L.D. and H.S. analyzed data; Y.L., X.Y. and Z.P. contributed reagents/materials/analysis tools; H.Z. and Z.G. wrote the paper. All authors read and approved the final manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zhao, H. *et al*. Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus *Phyllostachys*. *Sci. Rep.* **5**, 8018; DOI:10.1038/srep08018 (2015).