Biochemical Genetics (2019) 57:652–672
https://doi.org/10.1007/s10528-019-09914-4

ORIGINAL ARTICLE

Discovery and Utilization of EST-SSR Marker Resource for Genetic Diversity and Population Structure Analyses of a Subtropical Bamboo, *Dendrocalamus hamiltonii*

Abhishek Bhandawat, et al. ***[full author details at the end of the article]***

Received: 20 August 2018 / Accepted: 26 March 2019 / Published online: 5 April 2019

Abstract

Dendrocalamus hamiltonii is a giant bamboo species native to Indian subcontinent with high economic importance. Nevertheless, highly outcross nature and flowering once in decades impose severe limitation in its propagation. Identification and mixed cultivation of genetically diverse genotypes may assist successful breeding and natural recombination of desirable traits. Characterization of existing genetic diversity and population structure are indispensable for efficient implementation of such strategies, which is facing a major challenge due to non-availability of sequence-based markers for the species. In this study, 8121 EST-SSR markers were mined from *D. hamiltonii* transcriptome data. Among all, tri-repeats were most represented (52%), with the abundance of CCG/CGG repeat motif. A set of 114 polymorphic markers encompassing epigenetic regulators, transcription factors, cell cycle regulators, signaling, and cell wall biogenesis, detected polymorphism and interaction (in silico) with important genes, that might have role in bamboo growth and development. Genetic diversity and population structure of the three *D. hamiltonii* populations (72 individuals) revealed moderate to high-level genetic diversity (mean alleles per locus: 5.8; mean PIC: 0.44) using neutral EST-SSR markers. AMOVA analysis suggests maximum diversity (59%) exists within population. High genetic differentiation ($G_{st} = 0.338$) and low gene flow ($N_m = 0.49$) were evident among populations. Further, PCoA, dendrogram, and Bayesian STRUCTURE analysis clustered three populations into two major groups based on geographical separations. In future, SSR marker resources created can be used for systematic breeding and implementation of conservation plans for sustainable utilization of bamboo complex.

Keywords Bamboo · genetic diversity · microsatellite · population structure · SSR · transcriptome

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10528-019-09914-4) contains supplementary material, which is available to authorized users.

Introduction

Bamboo, also known as "poor man's timber" is the fastest growing, ligno-cellulose rich, perennial plant. It possesses tremendous ability to regenerate and produce multiple shoots annually, making it one of the most indispensable plants (Lessard and Chouinard 1980). Bamboo dwells in diverse environmental regimes and edaphic conditions of tropics, subtropics, and temperate regions. With its high carbon-sequestration efficiencies compared to Chinese fir and pines (Zhang et al. 2014), bamboo is a promising bioresource for managing climate change. Extraordinary versatility, high biomass, and easy machinability makes bamboo useful for multiple applications, including paper making, scaffolding, handicraft, and construction (Shukla and Das 1981; Tewari 1992; Das and Rout 1994). India, with over 136 species (23 genera) has nearly 17% bamboo forest cover (Tewari 1992). Among the 10 commercially valuable species, *Dendrocalamus hamiltonii* (hexaploid: $2n = 6x = 72$) is a giant bamboo species popular in India, having a life cycle of about 30 years (Ghorai and Sharma 1980; Bhatt et al. 2003). It occurs throughout Northeast and Northwest India providing necessary raw material for paper industry, buildings, furniture, support, and firewood. Due to the high demand of woods, the species is harvested at a faster pace, reducing the natural diversity of the species. Selection of elite genotypes in existing genetic diversity is constrained due to unavailability of reliable sequence-based marker resource, hindering efficient management, genetic improvement, and sustainable utilization of the species.

Microsatellites or Simple Sequence Repeats (SSRs) are short iterations of 1–6 bp repeats, often associated with replication slippage and DNA repair, widely distributed in the genome in all organisms (Tautz and Renz 1984; Epplen et al. 1993). SSRs play a crucial role in regulating gene expression, recombination, cell cycle, and chromatin organization (Li et al. 2002). Considering their obvious advantages such as hypervariability, abundance, specificity, high prevalence, co-dominance, and reproducibility, SSRs have emerged as a powerful marker for various genotyping applications like genetic diversity, population genetics, marker-assisted breeding, QTL mapping, and evolutionary studies (Varshney et al. 2005). Next-generation sequencing has revolutionized plant research by generating high-throughput genomic information at a much lower cost and labor, providing tremendous thrust for the genomic exploration in non-model plants (Abdelkrim et al. 2009; Unamba et al. 2015). Transcriptome sequencing generates immense sequence information representing functional genome, leaving behind untranscribed and repetitive regions. EST-SSRs derived from transcribed genome can be quickly developed through electronic sorting, are highly cross-transferable across diverse taxa, and often their putative function may be predicted (Zalapa et al. 2012; Bhandawat et al. 2016). Earlier studies on SSRs identified from transcribed genome suggest their utility in facilitating marker association with phenotype (Li et al. 2002). Further, such markers are useful for genetic characterization, molecular breeding, marker-assisted selections, and genetic mapping studies (Varshney et al. 2005). Previously, we successfully utilized public expressed

sequence data for creation of SSRs in *Dendrocalamus latiflorus* (Bhandawat et al. 2013, 2016). In this study, tissue-specific transcriptome data of *D. hamiltonii* were utilized to develop large-scale EST-SSR marker resource, establish their functional relevance, followed by polymorphism survey. Informative markers thus identified were successfully utilized for characterization of unexplored diversity and population stratification of the species in India for their efficient management, molecular breeding, and sustainable utilization in future.

Materials and Methods

De Novo Assembly, SSR Mining, and Primer Designing

Multi-tissue (dormant rhizome, growing rhizome, growing shoot, and mature shoot) transcriptome data (deposited at NCBI's sequence read archive under accessions: SRR3822239 to SRR3822242) (Bhandawat et al. 2017) generated to understand the molecular mechanism underlying growth of *D. hamiltonii* were utilized for SSR marker prediction. Low-quality sequences were filtered out using NGS QC toolkit (Patel and Jain 2012). De novo sequence assembly was performed using CLC Genomics Workbench (CLC Bio, Denmark) with sequence length cut-off = 150bp. Microsatellites (2-6bp repeats) were mined from assembled sequences using MISA (Thiel et al. 2003) with minimum repeat length for di-repeat = 6, tri-repeat = 4, tetra-repeat = 3, penta-repeat = 3, and hexa-repeat = 3 as described earlier (Bhandawat et al. 2016). SSR characteristics of *D. hamiltonii* were compared with two other species of bamboo (*D. latiflorus* and *Phyllostachys heterocycla*) (Liu et al. 2012; He et al. 2013). BatchPrimer3 was subsequently used to design primers (designated as DHTMS) flanking SSR repeats with default parameters (You et al. 2008).

Functional Annotation

To ascertain the functional relevance of SSR-containing transcripts, sequences were subjected to homology search against known proteins in NCBI's Nr, TAIR11, Swiss-Prot, plant transcription factor database (https://planttfdb.cbi.pku.edu.cn), and KEGG (https://www.genome.jp/kaas-bin/kaas_main) using the blastx algorithm with the cut-off *E*-value < $1.0e^{-5}$. KO number and transcripts encoding enzymes were determined from KEGG annotations. Conserved orthologous sequences were identified and classified using Eukaryotic Cluster of Orthologous Groups (KOG) database (Tatusov et al. 2003). For functionally relevant transcripts, AgriGO software (https://bioinfo.cau.edu.cn/agriGO) was used to determine GO terms and classified into three categories: biological process, molecular function, and cellular component using WEGO software (Ye et al. 2006). Overrepresented GO categories were identified using singular enrichment analysis (SEA) of AgriGO performed with *Oryza* orthologs against rice gene model as the reference. Fisher's statistical test with Hochberg (FDR) adjustment < 0.05 was performed for optimal enrichment of genes.

Network Analysis of SSR Transcripts

Protein–protein interaction (PPI) network of functionally relevant SSR transcripts was built using STRING PPI database of *Arabidopsis thaliana* (https://string-db.org/) to predict their interaction and create a regulatory network. The network was visualized using Cytoscape v3.4 (Smoot et al. 2011). Based on the significant conserved correlation edge with its respective TAIR orthologs, the correlation between the SSR transcripts was determined. Subsequently, a sub-network of polymorphic SSRs was built from overall SSR network to assess the interactions among themselves.

Plant Material and SSR Marker Evaluation

Northeast and Northwest Himalaya are the major biodiversity hotspot for *Dendrocalamus hamiltonii*. To evaluate the amplification potential of EST-SSRs, leaves of 72 accessions (samples) from three geographically distinct natural populations of *D. hamiltonii* (family: Poaceae) were collected and preserved in silica gel. Of these, 23 accessions belonged to Kangra, 25 to Mandi (Northwestern Himalaya), and 24 to Mizoram (Northeastern Himalaya) populations of India (Table S1). Leaves of individual samples were processed for DNA extraction using CTAB method (Doyle and Doyle 1990). The quantity and quality of DNA were assessed using Nanodrop (Thermo Scientific, USA) and 0.8% agarose gel with respect to λ uncut (Fermentas, USA). Ten random accessions from the three populations of *D. hamiltonii* were used for initial validation and polymorphism screening on denaturing polyacrylamide gel using 192 di- or tri-repeat entailing microsatellite markers. PCR reaction, polyacrylamide gel (PAGE) separation and amplicon visualization were performed as previously described (Bhandawat et al. 2013). A subset of polymorphic markers was used for diversity assessment of 72 accessions (including those used for screening). *D. hamiltonii* being hexaploid, all the allelic variants of given SSR locus identified using PAGE were scored as 1 (present) or 0 (absent) for binary data matrix according to dominant (Pfeiffer et al. 2011) scoring pattern. Numbers of alleles per marker locus were used to calculate polymorphic information content (PIC) of each marker (Anderson et al. 1993).

Outlier Test

To check if candidate loci are influenced by environmental gradient, and whether these SSR loci are suitable for genetic analysis, neutrality testing was performed across the 72 *D. hamiltonii* individuals across three different populations. In "Outlier Test," the SSR loci that demonstrate F_{st} significantly larger or smaller show the evidence of selection. To test for neutrality, Bayescan outlier detection program which implements multi-test corrections to estimate posterior probability was used (Foll and Gaggiotti 2008; Manel et al. 2009; Parisod and Joost 2010). We explored the locus under selection based on the “priority odds (PO) of neutrality” using threshold

value of 10 (for every 10 neutral loci in the data set, odds are that 1 locus is under selection). For the Markov Chain Monte Carlo algorithm implemented in BayeScan 2.1, 20 pilot runs each with 5000 iterations were used to adjust the proposal distribution to have acceptance rates between 0.25 and 0.45 for the runs. Afterwards, a burn-in of 50,000 was used for estimation using a thinning interval of 10.

Data Analyses for Diversity and Population Structure

Nei's genetic distance, gene diversity (h), percent polymorphic loci ($\%P$), Shannon's Information index (I), pairwise population differentiation (ϕ_{pt}), principal coordinate analysis, and Analysis of Molecular Variance (AMOVA) to study distribution of diversity across populations were estimated using GenAlEx v6.502 (Peakall and Smouse 2006). Genetic differentiation coefficient (G_{st}) was measured using POPGENE v1.32. Gene flow was measured as $N_m = (1/G_{st} - 1)/4$ (Slatkin and Barton 1989). Cluster analysis (with 1000 bootstraps) was performed based on Jaccard dissimilarity coefficient for determining genetic relationships among accessions using Darwin software v5.0.158 (Perrier and Jacquemoud-Collet 2006). To estimate the population stratification, Bayesian clustering algorithm of STRUCTURE version: 2.3.4 (Pritchard et al. 2000; Falush et al. 2007) was employed. Five independent runs (K value ranging from 1 to 10 with five iterations for each value of K) were performed. The length of burn-in period and number of Markov Chain Monte Carlo (MCMC) repeats were set to 500,000 each. The number of clusters were determined using structure harvester based on the most suitable value for K (Evanno et al. 2005; Earl 2012), while also considering the supporting results of above cluster analysis and recently observed biases towards $K = 2$ (Janes et al. 2017). Accessions were assigned to different clusters using Clumpp v1.1.2 (Jakobsson and Rosenberg 2007). The output of Clumpp was directly used as input for Distruct v1.1 (Rosenberg 2004) to obtain results in graphical format.

SSR Locus Variant Resequencing

Three randomly chosen EST-SSR markers, namely DHTMS-45878, DHTMS-1233, and DHTMS-22356 amplifying specific tri-repeat motifs: GAG, TCA, and CTT, respectively, were further validated for allelic variation by Sanger sequencing. Allelic variants of respective marker were separated on 6% polyacrylamide gel. Polymorphic amplicons were excised and suspended in 10 µL nuclease-free water. A second amplification was performed in 50 µL PCR reaction volume (as performed earlier) and subsequently purified using QIAquick PCR purification kit (Qiagen, USA). The concentration of purified product was measured on Nanodrop (Thermo Scientific, USA). 250ng template was used for direct PCR sequencing in 10 µL reaction containing 1 µL DMSO, 4 µL ready reaction mix using BigDyeTerminator v3.0Cycle Sequencing kit (Applied Biosystem, USA), and 15 ng of forward or reverse primer. Base calling and sequence analysis were done with ABI sequence analysis software (Applied Biosystem v5.2, USA). Multiple sequences were aligned using ClustalX v2.0.11.

In Silico Comparative Mapping

As map-based reference genome was unavailable for the species, all the EST-SSR transcripts were mapped to its closest reference genome, *Oryza sativa* using the blastn tool (*E*-value cut-off ≤ 1E−05). SSRs showing unique and reliable mapping were considered for comparative mapping. Transcripts showing multi-chromosomal hits (*E*-value difference < tenfold) were discarded from the analysis.

Results

Transcriptome Assembly and Microsatellite Marker Development

A total of 56.9 million clean transcriptomic reads were assembled into 116,488 consensus transcripts, representing a cumulative sequence of 38.7 Mb, an average length of 332 bases and N50 of 352 bp. A total of 9857 transcripts encompassing 11,624 microsatellites (repeat length > 12) with 8439 transcripts containing single SSR and 1418 containing more than one SSR were identified (Table 1). Among all repeats, tri-repeats were most abundant ones (52%), while hexa- repeats were least (3%) (Fig. S1). Within tri-repeats $(CCG/CGG)_n$ and $(AGG/CCT)_n$, within tetra-repeats $(AAAG/CTTT)_n$ and $(AGCT)_n$, and within di-repeats $(AG/CT)_n$ were most abundant repeat motifs. From the frequency of various repeat motifs, it could be inferred that GC-rich repeats were highly predominant in bamboo. The frequency of microsatellites was found to be one per ten kb of *D. hamiltonii* transcripts. Transcriptome assembly of a temperate (*P. heterocycla*) and a tropical/subtropical (*D. latiflorus*) bamboo was used for SSR mining and comparison with *D. hamiltonii* SSRs as shown in Fig. 1. Tri-repeats were most predominant in each species

Table 1 Summary of data mining and SSR identification

Total raw reads	66.5 million
Total number of high-quality reads mined	56.9 million
Number of assembled transcripts	116,488
size of examined transcripts (bp)	38,670,049
Maximum transcript length (bp)	5004
Average transcript length (bp)	332
N50 length (bp)	352
GC content	47.7%
Identified SSRs	11,624
SSR-containing transcripts	9857
Transcripts with > 1 SSRs	1418
SSRs involved in compound formation	1156
SSR primers designed	8121
Number of SSRs tested	192
Number of Working SSR markers	155
Number of Polymorphic SSR markers	114

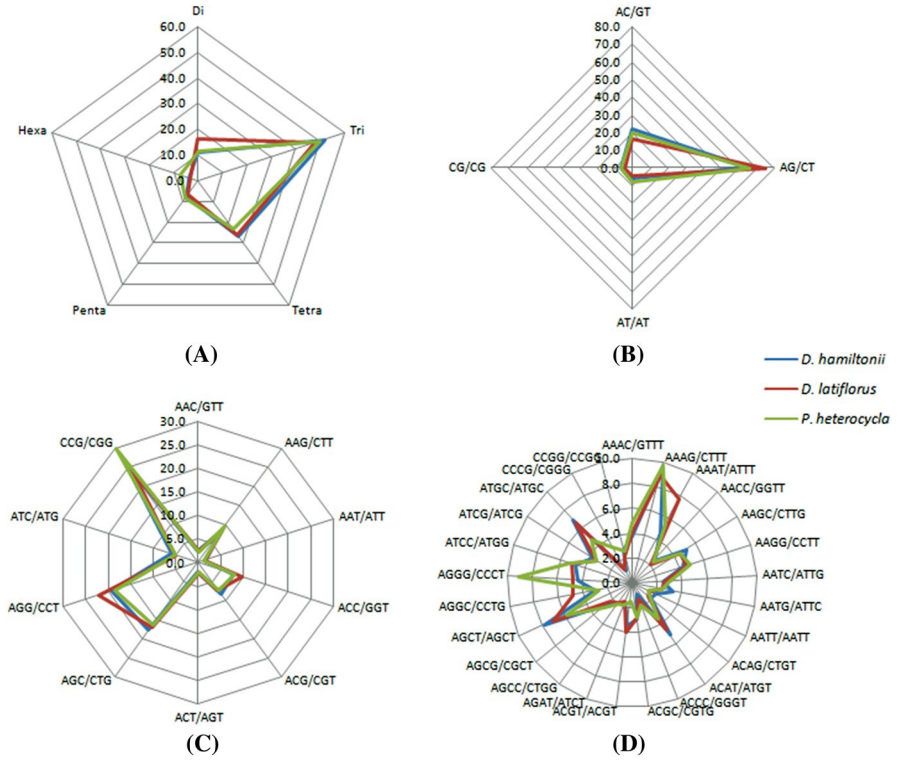


Fig. 1 Comparison of microsatellites in coding regions of three bamboo species: **a** repeat type and **b–d** most predominant repeat motifs

followed by tetra- and di-repeats. Among di-repeats, AG/CT motif was found most abundant, while CCG/CGG and AGG/CCT repeats were most abundant tri-repeat motifs in each case. A similar pattern of di- and tri-repeats was observed in all the three species; however, at the level of tetra SSR repeats, *D. hamiltonii* and *D. latiflorus* were showing higher similarities with each other. In contrast to *D. latiflorus* and *D. hamiltonii,* AGGG/CCCT repeat motifs were most represented in *P. heterocycla*. Similarly, ACGT/ACGT were most abundant in both the *Dendrocalamus* species but not in *Phyllostachys*.

Through SSR mining of *D. hamiltonii* transcripts, a sum of 8121 unique primer pairs [designated as DHTMS (*Dendrocalamus Hamiltonii* Transcriptome-derived MicroSatellite)] flanking specific SSR motifs were identified and submitted to NCBI probe database under accessions Pr032816837 to Pr032824957 (Table S2).

Functional Annotation and Enrichment of Microsatellites

Putative functions of SSR-containing transcripts were predicted based on sequence similarity search against various protein databases (Nr, Swiss-Prot,

TAIR, and Plant TF). 4729 (48.0%) SSR transcripts showed significant matches with the known proteins. Maximum annotations were found with Nr (4596; 46.6%) followed by TAIR (3133; 31.8%) and Swiss-Prot (2659; 27.0%). Based on blast search against plant transcription factor database, 1336 (13.6%) SSRs encoding transcription factors were predicted. Details of functional annotation are given in Table S2.

KEGG knowledgebase provides information about genes and pathways in an organism establishing links and relationships between them. A total of 916 transcripts with KO (KEGG Orthology) numbers were successfully mapped to 168 pathways and broadly grouped under four categories (Fig. S2). Maximum transcripts belonged to metabolism (476; 50.9%), which includes carbohydrate metabolism (110; 23.1%) and amino acid metabolism (91; 19.1%) as major subcategories. Other important subcategories include energy metabolism (63; 13.2%) and glycan metabolism (22; 4.6%). Genetic information processing was the second major category (228; 24.4%), which includes translation (92; 40.4%) and folding, sorting and degradation (68; 29.8%). Under environment information processing, signal transduction forms the major portion (119; 98.3%). Cell growth and death (50; 45.0%), and transport and catabolism (49; 44.1%) formed major fraction under cellular processes. Six-hundred and four SSR transcripts were encoding enzymes based on KEGG classification.

A total of 1920 SSR transcripts were classified into 25 functional classes based on KOG classification (Fig. S3). General function prediction was the most abundant class with 320 SSRs (16.6%). Other important classes include signal transduction mechanisms (140; 7.3%), carbohydrate transport and metabolism (74; 3.9%), cell cycle control, cell division, chromosome partitioning (29; 1.5%), and cell wall/membrane/envelope biogenesis (12; 0.6%).

Gene ontology (GO) is a globally standardized functional classification system of genes across all organisms. A total of 5679 GO terms (1–20) were identified in 1661 SSR transcripts and classified under three broad categories: biological processes (21 subcategories), cellular component (10 subcategories), and molecular function (12 subcategories) (Fig. S4). Under the biological process category, metabolic process (863 transcripts) and cellular process (763) were the most abundant subcategories. Other important categories include biological regulation (242), response to stimulus (63), and anatomical structure formation (23). Under molecular function, binding (1034) and catalytic activity (729) were the first and second largest subcategories, respectively. Regarding cellular components, cell (628) and cell part (628) were most predominant subcategories.

To investigate the role of SSR in functionally important transcripts, GO enrichment analysis was performed (Table S3). Under biological process (Fig. S5), we found two subcategories, namely gene expression and transcription to be highly enriched, while other subcategories like regulation of cellular process and transport were moderately enriched. Low level of enrichment was seen in response to abiotic stimulus, catabolic processes, lipid metabolic processes, and translation. Under molecular function category (Fig. S6), hydrolase activity (nucleoside-triphosphatase activity) was moderately enriched. Transcription factor activity, structural activity, and transporter activity were showing low enrichment. Most of the subcategories of

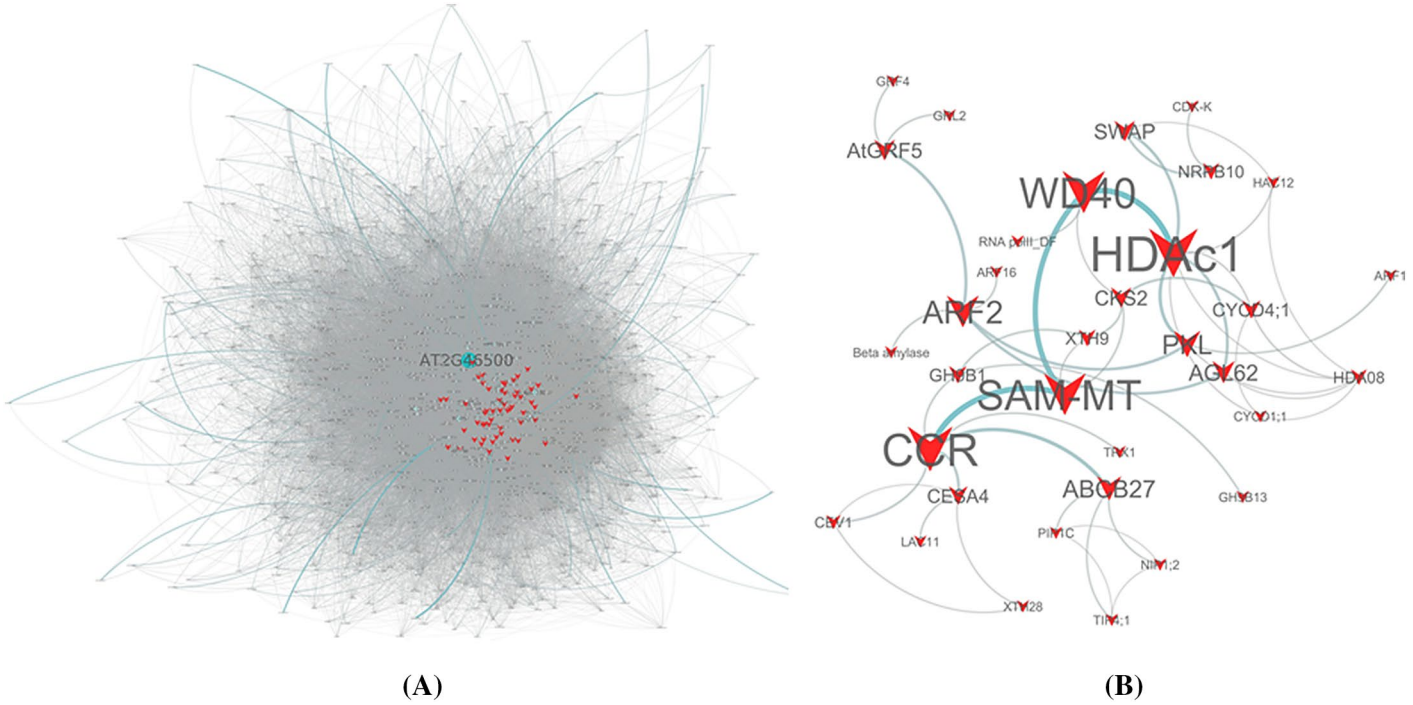
cellular component (Fig. S7), namely ribosome, nucleus, nucleoplasm, endoplasmic reticulum, were lowly enriched.

Protein–Protein Interaction Network

Since SSR sequences are liable to show higher replication slippages, variation in SSRs encompassing genes may severely influence gene function. Further, it is essential to determine the interacting proteins which might be indirectly influenced by such variations. Considering enriched information about Arabidopsis protein–protein interactions (PPI), SSR-containing transcripts were mapped to predetermined PPI network of Arabidopsis for identification of putative interacting proteins. Out of 2664 functionally relevant SSR transcripts, 1912 unique TAIR-IDs were successfully mapped to Arabidopsis PPI network. Among these, 55 polymorphic SSR markers recorded direct interaction with 1140 nodes/genes (46,154 edges) with clustering coefficient 0.355 and 77 average number of neighbors (Fig. 2a; Table S4). Thirty-four polymorphic SSR markers were directly interacting among themselves with 50 edges. These 34 markers include various epigenetic regulators, transcription factors, meristem maintenance, cell cycle regulators, water transporters, and cell wall biogenesis-related genes (Fig. 2b). S-adenosyl-L-methionine-dependent methyltransferases (SAM-MT), histone deacetylase (HDAc1), and RNA polymerase II degrading factor showed direct interaction with transducin/WD40 repeat-containing transcription factor (TF), thus WD40 might be playing major role in regulating expression of epigenetic regulators. Further, HDAc1 was directly interacting with histone deacetylase 8, histone acetylase, cyclin D, agamous-like, protein kinase, and SWAP (expressed in growing tissues) proteins. Aside from this, SAM-MT was found to show direct interaction with xyloglucan endotransglucosylase, Auxin response factor (ARF), agamous-like, and cinnamoyl-CoA reductase (CCR). CCR was directly interacting with ABC transporters (directly interacting with PIP, NIP and TIP transporters), cellulose synthase (directly interacting with laccase, xyloglucan endotransglucosylase, other cellulose synthase family members), and glycosyl hydrolase (GH9B1: cell wall modification). ARF showed direct interaction with the growth regulating factors, Beta-amylase, protein kinase-like, and agamous-like proteins thus might influence bamboo growth.

Detection of Polymorphism

A set of 192 SSR primers flanking di- or tri-repeats, including transcription factors (31), epigenetic modulators (21), cell wall biogenesis (23), environment and signaling (6), phytohormones (6), cell cycle regulators (6), water transport (8), sugar metabolism and transport (5), and disease resistance (1) genes implicated to have role in growth were selected for amplification validation in 10 random accessions of *D. hamiltonii* collected from three geographically distinct populations of India (Table S1). A total of 155 SSR markers (80.7%) showed successful amplification in at least one accession. Among these, 114 markers (73.5%) were capable of detecting polymorphism and were tagged as "polymorphic marker"; rest 41



(A) **(B)**

Fig. 2 Protein–protein interaction (PPI) network showing interactions among functionally relevant SSR transcripts and their first neighbors. **a** PPI network of all SSR transcripts, **b** PPI network of polymorphic SSRs. Size of symbol signifies their interactions with multiple proteins

markers amounting 26.5% could not differentiate between accessions and were tagged "monomorphic marker." In all, 114 polymorphic SSRs produced a total of 666 alleles ranging from 2–15 (mean alleles 5.8 per SSR locus) in *D. hamiltonii* accessions. The PIC varied from 0.255 to 0.5 and the average was 0.44 (Table S5). Interestingly, 56 SSRs encompassing various growth-related genes, including two markers for meristem maintenance, epigenetic regulators (12 markers), cell signaling (2), cell cycle regulators (7), cell wall biogenesis (13), cell expansion (8), phytohormones (2), transcription factors (10), also detected variable level of polymorphism.

Outlier Test

As SSRs were scored dominantly, all the 231 loci (40 SSR) across 72 individuals of *D. hamiltonii* were tested to detect loci under selection. BayeScan program revealed none of the 231 loci were under selection (as maximum log10 (PO) of -0.12509 which corresponds to a posterior probability of. 0.42849) (Fig. S8). As all the loci were neutral, they can be utilized for phylogeographic and demographic studies.

Genetic Diversity Analyses

A subset of 40 highly polymorphic neutral markers were randomly chosen to study genetic diversity and population structure of *D. hamiltonii* across 72 accessions representing three Indian populations. Forty markers produced a sum of 231 alleles with an average of 5.8 alleles per SSR locus (Table S5). Mean PIC of tested markers across all individuals was 0.456, suggesting the markers were highly informative. Maximum polymorphic bands (78.36%) were observed in Mizoram population, while 69.40% and 68.66% in Mandi and Kangra populations, respectively. Among the three populations, Mizoram exhibited maximum diversity as inferred from the number of effective alleles (1.39), Shannon's information index (0.358), and unbiassed diversity (0.243) (Table 2).

Genetic Differentiation and Partitioning

Genetic differentiation was analyzed among three populations. Nei's genetic distance ranged from 0.126 (Kangra–Mandi) to 0.494 (Mandi–Mizoram) (Table 3).

Table 2 Estimates of genetic diversity in *D. hamiltonii* populations

Population	*N*	*% P*	N_a	N_e	*I*	*h*	uh
Kangra	23	68.66	1.470 ± 0.072	1.224 ± 0.024	0.248 ± 0.019	0.150 ± 0.013	0.157 ± 0.014
Mandi	25	69.40	1.493 ± 0.070	1.158 ± 0.019	0.200 ± 0.016	0.114 ± 0.011	0.119 ± 0.011
Mizoram	24	78.36	1.664 ± 0.059	1.391 ± 0.032	0.358 ± 0.022	0.233 ± 0.016	0.243 ± 0.017
Total	72	72.14	1.542 ± 0.039	1.258 ± 0.015	0.269 ± 0.012	0.166 ± 0.008	0.173 ± 0.009

N number of accessions, *%P* percent polymorphic loci, N_a number of different alleles, N_e number of effective alleles, *I* Shannon's information index, *h* diversity, *uh* unbiased diversity

Table 3 Pairwise comparison of Nei's genetic distance (matrix below diagonal) and genetic differentiation (ϕ_{PT}, matrix above diagonal) in *D. hamiltonii*

Kangra	Mandi	Mizoram	
0.000	0.030	0.238	Kangra
0.126	0.000	0.248	Mandi
0.448	0.494	0.000	Mizoram

Nei's genetic differentiation (G_{st}) was found to be 0.338 suggesting high genetic differentiation among populations. Pairwise population differentiation showed high differentiation between Mandi and Mizoram populations ($\phi_{PT}=0.248$), while minimum differentiation was observed between Kangra and Mandi populations ($\phi_{PT}=0.030$) (Table 3). Analysis of molecular variance (AMOVA) revealed maximum variance was dwelling within population (59%), while 41% variance was recorded among populations ($P<0.0001$) (Table 4). Gene flow (N_m) (calculated based on G_{st}) among all the populations was found to be 0.49. High gene flow was recorded between Kangra and Mandi populations ($N_m=1.73$), while Mandi and Mizoram recorded low gene flow ($N_m=0.26$) estimated using pairwise ϕ_{PT} values.

Clustering and Population Structure

Three complementary approaches (combined NJ tree, principal coordinate analysis, and Bayesian model) were used to evaluate the genetic relationships among the three populations (72 accessions). Out of 72 *D. hamiltonii* accessions, 68 were grouped into two major clusters. Cluster 1 was further represented as sub-cluster 1a and 1b, mostly comprising of Kangra (15 accessions) and Mandi populations (22 accessions), respectively, with intermixing of few accessions (Fig. 3). Interestingly, all the accessions (except one) of Mizoram were grouped under Cluster 2. Four accessions irrespective of the tested populations did not fall under these clusters and remained as outgroup. The 72 accessions of *D. hamiltonii* were further investigated for population stratification using STRUCTURE. Bayesian model of population structure grouped all the 72 accessions into two major clusters ($K=2$) based on the maximum value of ΔK (Fig. 4). Inferred ancestry revealed that 68 accessions were strongly associated with assigned clusters as pure genetic populations ($ql \geq 0.80$). Four accessions, one from Mizoram (dh-acc/57), two from Kangra (dh-acc/67 and dh-acc/71), and one from Mandi (dh-acc/25–18), showed admixed ancestry. Results of genetic structure analysis suggested accessions were broadly grouped according

Table 4 Analysis of molecular variance of *D. hamiltonii populations*

Source	df	SS	MS	Est. Var.	% variation	*P*
Among pops	2	408.010	204.005	8.023	41%	< 0.001
Within pops	69	797.352	11.556	11.556	59%	< 0.001
Total	71	1205.361		19.579	100%	

Df degree of freedom, *P* probability, *SS* sum of squares, mean square

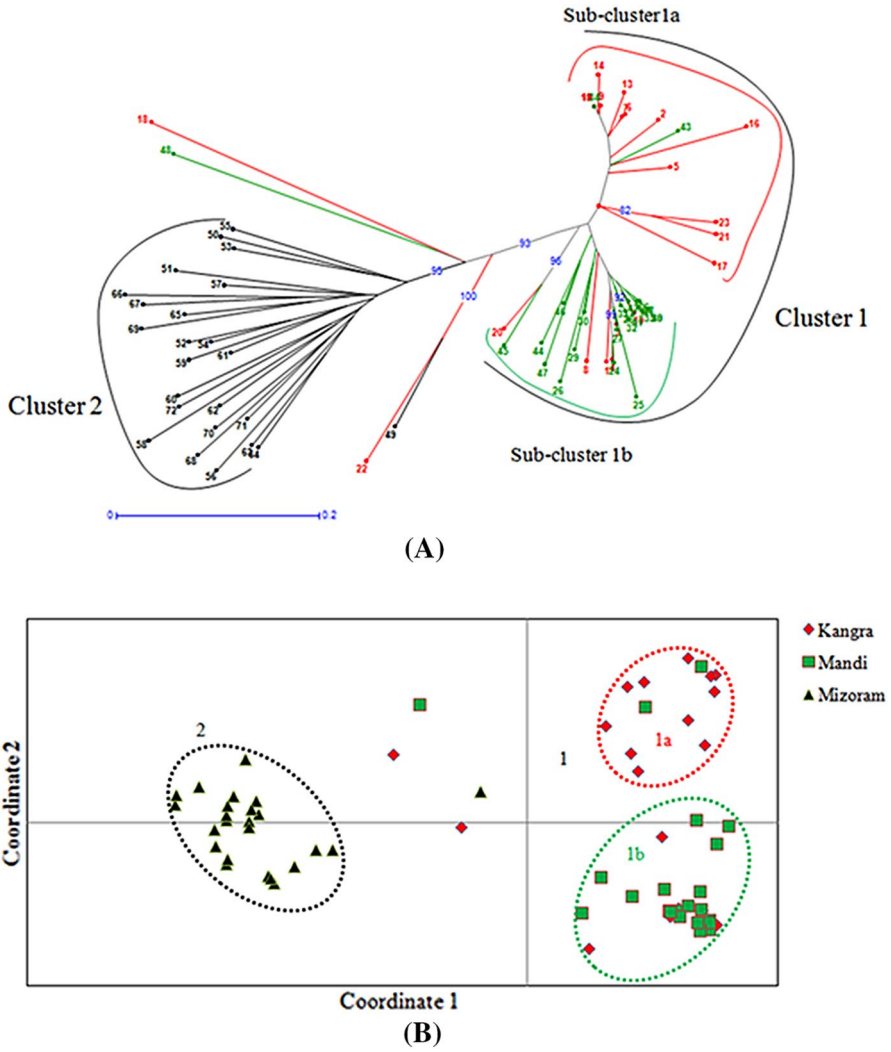


Fig. 3 Clustering pattern of 72 *D. hamiltonii* genotypes: **a** Neighbor Joining dendrogram based on Jaccard dissimilarity coefficient along with bootstrap values and **b** Principal coordinates analysis representing two populations (Northeast and Northwest) as two distinct clusters

to geographical location. Population structure also indicated clear distinction of Mizoram (Northeast) from Kangra and Mandi (Northwest) populations, thus complementing other analyses.

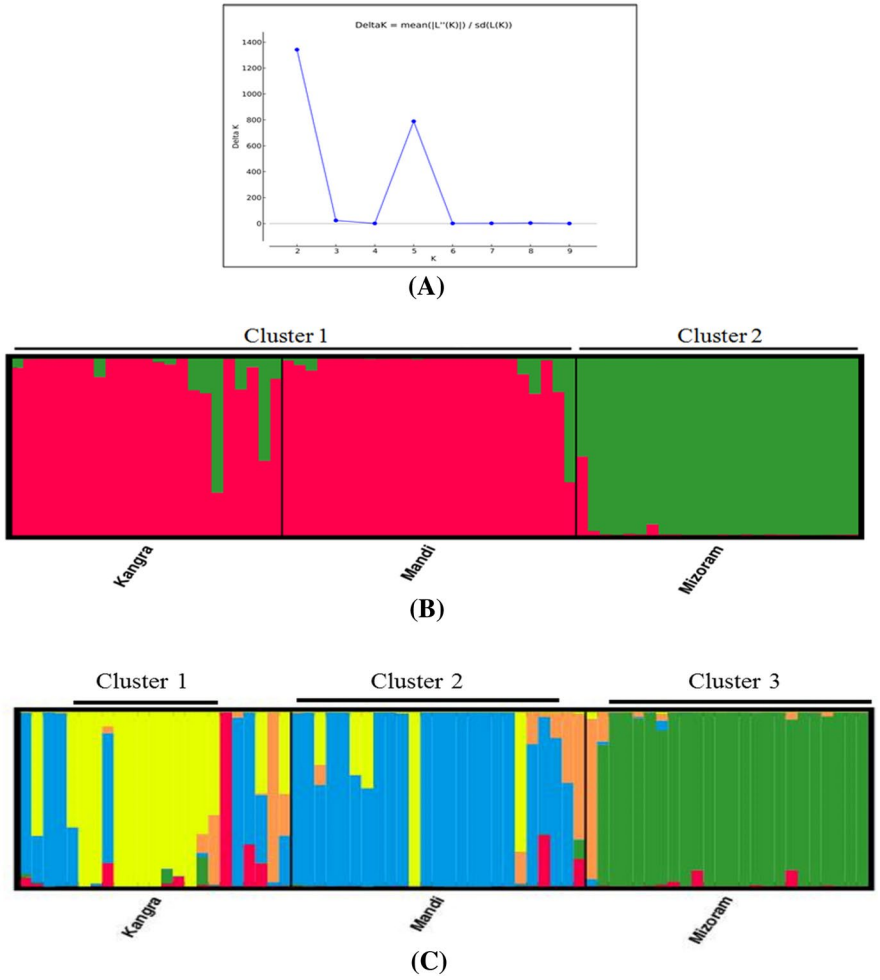


Fig. 4 Bayesian model of population genetic structure showing 2 major genepools of *D. hamiltonii.*. **a** Graph representing deltaK max, **b** Genotype wise population structure ($K=2$), **c** Genotype wise population structure ($K=5$)

SSR Locus Resequencing

To confirm the variation displayed by SSR loci, we sequenced polymorphic alleles of three SSR locus encompassing different tri-repeat motifs. In each case, variation in SSR locus was mainly due to addition or deletion of targeted repeats. However, in few cases, point mutations (transition/transversion) were observed along with few indels (Fig. S9).

In Silico Mapping of SSR Locus

A total of 3099 SSR markers were successfully mapped on twelve *Oryza sativa* orthologous chromosomes (Fig. S10). Maximum number of SSRs were mapped on chromosome 3 (490 SSRs) and 1 (453 SSRs). Least SSRs were mapped to chromosome 10. On an average, 258 SSRs were mapped to each rice chromosome. Distribution of SSRs based on comparative mapping broadly correlated with chromosomal length; also complementing genome-wide distribution of SSRs.

Discussion

Bamboos are sustainable alternatives to timber forests due to their fast growth and ability to produce multiple shoots annually. However, the natural habitat of many valuable bamboo species, including *D. hamiltonii,* is dwindling due to forest fire, extensive harvesting, and complex breeding behavior (infrequent flowering in 30–40 years), leading to rapid loss of natural bamboo stands which is not well documented (Bystriakova et al. 2003). Lack of genomic resources and molecular markers has severely obstructed characterization of many commercially important bamboos (including *D. hamiltonii*) for their efficient management and genetic improvement. In this study, spatio-temporal transcriptome sequence data were utilized for SSR mining and creation of a comprehensive EST-SSR marker resource in *D. hamiltonii*. SSR mining through electronic sorting of assembled transcriptomic reads is very quick and inexpensive approach. Additionally, NGS provides a global representation of genome compared to other low-throughput technologies, resulting in the accurate representation of SSR characteristics and frequencies (Ambreen et al. 2015). We identified 11,624 SSRs enriching the marker resource in *D. hamiltonii*. Frequency of SSRs in *D. hamiltonii* (1/10 kb) was found similar to that of sugarcane (1/10.9 kb), maize (1/9 kb), barley (1/8.9), and wheat (1/10.6 kb) (Parida et al. 2010), but lower as compared to *D. latiflorus* (1/4 kb) (Liu et al. 2012) and higher as compared to *P. heterocycla* (1/14.3 kb) (He et al. 2013), which in part also depends upon SSR mining parameters that vary from study to study. Prevalence of tri-repeats in expressed genome may be because they are less liable for frame-shift mutations (Metzgar et al. 2000). Broadly, microsatellite characteristics (frequency and distributions of SSRs and motif type) were generally conserved among the Poaceae species. Differences in frequencies of specific SSR repeat motifs may have an evolutionary role in plants which needs to be further investigated.

SSRs developed from genomic libraries (SSR-enriched or genome sequencing projects) are largely regarded as evolutionarily neutral markers, while SSR markers derived from EST/transcriptome sequencing may play adaptive and evolutionary role through induced changes in amino acid sequence, protein length, 3D structure, or regulation of gene expression (Li et al. 2004; Hancock and Simon 2005). A large number of SSR encompassing important genes, transcription factors, epigenetic modulators, cell cycle controllers, signaling, and cell wall biogenesis regulating growth and development of bamboo were identified in our study (Bhandawat et al. 2017). Gene ontology helped us in the identification of important categories

involved in metabolisms, biological regulation, response to stimulus, and anatomical structure formation of bamboo, resulting in fast growth of bamboo. High level of enrichment of SSR transcripts detected in gene expression and transcription categories suggested an involvement of these loci in transcriptional regulations. Large number of SSR transcripts were mapped to various pathways, including carbohydrate metabolism, glycan metabolism, transcription, translation, replication and repair, signal transduction, transport, cell growth, and death, implying their role in various aspects of bamboo growth and development. TFs (Transcription Factors) regulate a lot of biological processes, including responses to environmental stimuli, and maintenance of metabolic and physiological pathways (Riechmann et al. 2000). In our study, important TF family members such as ARFs, GRFs, and WD40, having a role in bamboo growth, showed allelic variation among accessions and can therefore be potentially useful markers for dissection of the regulatory mechanism of growth in bamboo. Polymorphism was detected in SSRs encompassing epigenetic modulators (histone acetylase, histone deacetylases, and SAM methyltransferase) that affect gene expression through DNA methylation and/or histone tail modifications in response to intrinsic and extrinsic signals. Interestingly, phytohormones (ethylene biosynthesis, auxin transport), cell signaling genes (protein kinase, calmodulin), cell cycle regulators (cyclin D, CDKs, RNA polymerase II, SWAP), cell wall biogenesis genes (Cellulose synthase, endoglucanase, xyloglucan endotransglucosylase, cinnamoyl-CoA reductase, laccase, sugar transporters), and cell expansion genes (Expansin, aquaporin, peroxidase) also revealed substantial polymorphism, and thus may regulate bamboo growth and development.

Proteins control physiological processes either directly or via interaction with other proteins. Interactome analysis provides useful information about proteins and their interactions. Variations in the proteins present in hub have profound effect on multiple proteins, and thus affect plant growth and development. Potential growth-related markers occurring in the hub were predicted based on protein–protein interaction network of SSR transcripts. Direct interactions between transcription factors, epigenetic regulators, cell cycle regulators, cell wall expansion, and biogenesis-related genes harboring SSRs may have implications for selection of elite genotypes, marker-assisted selection, and association mapping studies in bamboo (Xiao et al. 2014).

A huge data set of functionally relevant SSRs enriched the available marker resource for *D. hamiltonii* and successfully utilized for genetic diversity and population genetic analysis of the species. The higher success rate of designed SSR primers was observed in *D. hamiltonii* (80.7%) as compared to previous reports in *D. latiflorus* (44%) (Bhandawat et al. 2016), suggesting the good quality of designed primers. Although majority of primers (86) produced reliable amplification at 55 °C, annealing temperature of 28 primers was optimized to maximize the number of working primers. Most markers produced amplified bands within targeted size range, however, few markers produced amplicons of the size larger than expected, may be due to the presence of large intronic regions as observed previously in *Dendrocalamus latiflorus* (Bhandawat et al. 2013). Few primers failed to produce amplification may be due to intronic region present in the target locus or presence of splice site across designed primers, or point mutations in the primer binding sites (in particular

towards the more binding-sensitive 3′-end) (Liu et al. 2013). In *D. hamiltonii*, polyploidy might play a crucial role in the evolution of its natural populations and providing plasticity for performing better in stressful conditions as well as emergence of new species (te Beest et al. 2011; Madlung 2013).Generally, genomic SSRs are more polymorphic than EST-SSRs, however, we found significant polymorphism in EST-SSRs too. Novel SSR markers showed an average number of alleles and PIC (corresponds to high discriminatory nature of SSRs) similar to that of *D. latiflorus* transcriptome-derived SSRs. SSR variation was further confirmed by resequencing of polymorphic allelic variants, conveying the reliability of designed markers.

Population genetic structure provides clues about diversity, reproductive strategies, and the extent of gene flow of the species across distant geographic locations/conditions. Success represents fitness in terms of adaptability, genetic stability, and variability (Loveless and Hamrick 1984). Thus, species with poor variability is liable to extinction due to abiotic and/or biotic stresses. This makes it crucial to study diversity and genetic structure of commercially important overexploited species for their efficient utilization and strategic management (Yang et al. 2012). *D. hamiltonii* is commercially important and one of the most utilized bamboo species in India, occurring most widely in Northeast and Northwest Himalayan regions. Existing diversity and population biology of the plant remain understudied due to non-availability of reliable marker resource. Distribution of across populations genetic diversity is determined by multiple factors such as geographical barriers, demography, life history, and reproductive behavior of the species (Sampson and Byrne 2012). Outlier test based on Bayesian analysis revealed geographic location has no effect on any of the forty tested SSR marker locus, i.e., were neutral to selection pressure, thus can be utilized for phylogeographic and demographic studies of bamboos. Population-wise genetic diversity estimates (%*P*, N_e, *h*, *I*, and uh) indicate that Mizoram population is highly diverse as compared to Kangra or Mandi populations. Northeast region (Mizoram) of the Indian subcontinent appears to be the chief diversity center of the species, which requires further investigation. Pairwise genetic estimates highlight high differentiation between Northeast and Northwest regions. AMOVA analysis suggests that most of the diversity exists within populations (59%) which corresponds to the degree of population differentiation. The extent of pollen and seed dispersal to distant locations determines the flow of genes from one population to another (Loveless and Hamrick 1984). High gene flow ($N_m = 1.73$) between Kangra and Mandi population, and low gene flow between Mizoram–Kangra ($N_m = 0.30$) and Mizoram–Mandi ($N_m = 0.26$) populations may be because Kangra–Mandi populations are geographically closer as compared to Mizoram–Kangra or Mizoram–Mandi populations. Further, the existing genetic structure of the species corresponds to low gene flow across populations isolated by large distances, long flowering intervals (30–40 years), and highly outcross nature of the species. A similar pattern of population structure was also observed earlier in the case of *Dendrocalamus membranaceus* (Yang et al. 2012).

Further, based on cluster analyses (NJ tree, PCoA, Bayesian model), Mizoram population was clearly distinguished from Kangra and Mandi populations (which showed ample of intermixing within themselves) due to large geographical distance acting as a major barrier to gene flow, and clustered separately. The accessions were

differentiated into two major clusters (sub-clustering of Kangra and Mandi population) suggesting two broad genetic populations existing with clear distinctness between Mizoram and Mandi/Kangra populations. This kind of population differentiation between Northeast and Northwest populations is due to landscape barriers in gene flow (via seed, pollen dispersal, and/or long flowering intervals) which is low between Kangra and Mandi populations.

Rising demands for wood are a serious threat to the forest diversity. *D. hamiltonii* is commercially one of the most important bamboo species of India which lacks proper management. Current research presents the first study to assess genetic diversity and differentiation of Indian *D. hamiltonii* populations using novel EST-SSR markers. Inferences on genetic diversity and population stratification would expedite effective conservation, breeding, and genetic improvement efforts in this species. Huge number of EST-SSR marker resources, which may be useful for establishing functional diversity and marker-trait association studies in future. Molecular estimates of genetic diversity provide indications that substantial genetic variation exists in *D. hamiltonii,* and that Northeast (Mizoram) and Northwest (Kangra and Mandi) populations form two genetic populations in which isolation by distance plays determining role. However, considering, extensive harvesting activities, long flowering intervals, poor seed set, and high outcross nature of the species, it is essential to select identified diverse germplasm using newly developed SSR marker for co-cultivation, breeding, and conservation programs to avoid the wood scarcity in future.

Acknowledgements We acknowledge funding from CSIR research Grant MLP071. This is IHBT communication No. 4126.

Author contribution AB, RKS: Conceived and designed the experiments; AB, V.S., RS: Performed the experiments; AB, AN, PS, RS: Analyzed the data; AB, RKS: wrote the paper; JK: Helped in data interpretations; RKS: Editing and approval of final version of manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdelkrim J, Robertson BC, Stanton J-AL, Gemmell NJ (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. Biotechniques 46:185

Ambreen H, Kumar S, Variath MT, et al (2015) Development of genomic microsatellite markers in *Carthamus tinctorius* L.(safflower) using next generation sequencing and assessment of their cross-species transferability and utility for diversity analysis. PLoS ONE 10:e0135443

Anderson JA, Churchill GA, Autrique JE, et al (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Bhandawat A, Sharma V, Sharma H et al (2013) Development and crosstransferability of functionally relevant microsatellite markers in *Dendrocalamus latiflorus* and related bamboo species. TC 5:5

Bhandawat A, Singh G, Raina AS et al (2016) Development of genic SSR marker resource from RNA-Seq data in Dendrocalamus latiflorus. J Plant Biochem Biotechnol 25:179–190

Bhandawat A, Singh G, Seth R et al (2017) Genome-wide transcriptional profiling to elucidate key candidates involved in bud burst and rattling growth in a subtropical bamboo (Dendrocalamus hamiltonii). Front Plant Sci 7:2038

Bhatt BP, Singha LB, Singh K, Sachan MS (2003) Some commercial edible bamboo species of North East India: production, indigenous uses, cost-benefit and management strategies. Bamboo Sci Cult 17:4–20

Bystriakova N, Kapos V, Lysenko I, Stapleton CMA (2003) Distribution and conservation status of forest bamboo biodiversity in the Asia-Pacific Region. Biodivers Conserv 12:1833–1841

Das P, Rout GR (1994) Analysis of current methods and approaches on the micropropagation of bamboo. Proc Natl Acad Sci India Sect B 64:235

Doyle JJ, Doyle JL (1990) Isolation of DNA from small amounts of plant tissues. BRL Focus 12:13–15

Earl DA (2012) Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Epplen C, Melmer G, Siedlaczck I et al (1993) On the essence of "meaningless" simple repetitive DNA in eukaryote genomes. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds) DNA fingerprinting: state of the Science. Springer, Dordrecht, pp 29–45

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7:574–578

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–993

Ghorai A, Sharma A (1980) Bambuseae—a review. Feddes Repert 91:281–299

Hancock JM, Simon M (2005) Simple sequence repeats in proteins and their significance for network evolution. Gene 345:113–118

He C, Cui K, Zhang J et al (2013) Next-generation sequencing-based mRNA and microRNA expression profiling analysis revealed pathways involved in the rapid growth of developing culms in Moso bamboo. BMC Plant Biol 13:119. https://doi.org/10.1186/1471-2229-13-119

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806

Janes JK, Miller JM, Dupuis JR et al (2017) The K = 2 conundrum. Mol Ecol 26(14):3594–3602

Lessard G, Chouinard A (1980) Bamboo research in Asia: proceedings of a workshop held in Singapore, 28–30 May 1980. In: Bamboo research in Asia: proceedings of a workshop held in Singapore, 28–30 May 1980. International Development Research Centre.

Li Y, Korol AB, Fahima T et al (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11:2453–2465

Li Y-C, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991–1007

Liu M, Qiao G, Jiang J et al (2012) Transcriptome sequencing and de novo analysis for ma bamboo (Dendrocalamus latiflorus Munro) using the Illumina platform. PLoS ONE 7:e46766

Liu T, Zhu S, Fu L, et al (2013) Development and characterization of 1827 expressed sequence tag-derived simple sequence repeat markers for ramie (*Boehmeria nivea* L. Gaud). PLoS ONE 8:e60346

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Syst 15:65–95

Madlung A (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity (Edinb) 110:99–104

Manel S, Conord C, Després L (2009) Genome scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. BMC Evol Biol 9:288

Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10:72–80

Parida SK, Pandit A, Gaikwad K et al (2010) Functionally relevant microsatellites in sugarcane unigenes. BMC Plant Biol 10(1):2511

Parisod C, Joost S (2010) Divergent selection in trailing-versus leading-edge populations of Biscutella laevigata. Ann Bot 105:655–660

Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE 7:e30619

Peakall ROD, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Perrier X, Jacquemoud-Collet JP (2006) DARwin software: dissimilarity analysis and representation for windows. https://darwin.cirad.fr/õ [accessed 1 March 2013]

Pfeiffer T, Roschanski AM, Pannell JR et al (2011) Characterization of microsatellite loci and reliable genotyping in a polyploid plant, Mercurialis perennis (Euphorbiaceae). J Hered 102(4):479–488
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
Riechmann JL, Heard J, Martin G et al (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290:2105–2110
Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138
Sampson JF, Byrne M (2012) Genetic diversity and multiple origins of polyploid Atriplex nummularia Lindl. (Chenopodiaceae). Biol J Linn Soc 105:218–230
Shukla U, Das D (1981) On the occurrence of some medicinal grasses in Eastern India. Indian J For
Slatkin M, Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43(7):1349–1368
Smoot ME, Ono K, Ruscheinski J et al (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27:431–432
Tatusov RL, Fedorova ND, Jackson JD et al (2003) The COG database: an updated version includes eukaryotes. BMC Bioinform 4(1):41
Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12:4127–4138
te Beest M, Le Roux JJ, Richardson DM et al (2011) The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot 109(1):19–45
Tewari DN (1992) A monograph on teak (*Tectona grandis* Linn. f.). International book distributors
Thiel T, Michalek W, Varshney R, 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
Unamba CIN, Nag A, Sharma RK (2015) Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. Front Plant Sci 6:1074
Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55
Xiao Y, Zhou L, Xia W et al (2014) Exploiting transcriptome data for the development and characterization of gene-based SSR markers related to cold tolerance in oil palm (Elaeis guineensis). BMC Plant Biol 14(1):384
Yang H-Q, An M-Y, Gu Z-J, Tian B (2012) 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
Ye J, Fang L, Zheng H et al (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 34:W293–W297
You FM, Huo N, Gu YQ et al (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinform 9(1):253
Zalapa JE, Cuevas H, Zhu H et al (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot 99:193–208
Zhang H, Zhuang S, Sun B, et al (2014). Estimation of biomass and carbon storage of moso bamboo (Phyllostachys pubescens Mazel ex Houz.) in southern China using a diameter–age bivariate distribution model. Forestry 87(5):674–82.

Affiliations

Abhishek Bhandawat[1,2] · Vikas Sharma[1,3] · Pradeep Singh[1] · Romit Seth[1] · Akshay Nag[1] · Jagdeep Kaur[2] · Ram Kumar Sharma[1,4]

✉ Ram Kumar Sharma
rksharma.ihbt@gmail.com; ramsharma@ihbt.res.in

[1] Molecular Genetics & Genomics Lab, Department of Biotechnology, CSIR-Institute of Himalayan Bioresource Technology, Palampur, Himachal Pradesh, India

[2] Department of Biotechnology, Panjab University, Chandigarh, India

[3] Present Address: Sant Baba Bhag Singh University, Khiala, Jalandhar, India

[4] Academy of Scientific and Innovative Research (AcSIR), CSIR-IHBT, Palampur, Himachal Pradesh 176061, India