Ecological Indicators 131 (2021) 108143

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Development of genic-SSR markers and their application in revealing genetic diversity and population structure in an Eastern and North-Eastern Indian collection of Jack (*Artocarpus heterophyllus* Lam.)

Devendra K. Singh [a,d], Avinash Pandey [a], Shashi Bhushan Choudhary [b], Sudhir Kumar [a], Kishor U. Tribhuvan [a], Dwijesh C. Mishra [c], Jyotika Bhati [c], Madan Kumar [a], J.B. Tomar [b], S. K. Bishnoi [b], M.A. Mallick [d], V.P. Bhadana [a], T.R. Sharma [a], A. Pattanayak [a], Binay K. Singh [a,*]

[a] ICAR – Indian Institute of Agricultural Biotechnology, Ranchi 834 003, Jharkhand, India
[b] ICAR – National Bureau of Plant Genetic Resources, Regional Station, Ranchi 834 003, Jharkhand, India
[c] ICAR – Indian Agricultural Statistics Research Institute, New Delhi 110 012, India
[d] Vinoba Bhave University, Hazaribagh 825 301, Jharkhand, India

ARTICLE INFO

Keywords:
Moraceae
Transcriptome
Illumina sequencing
Molecular markers
Genetic variation

ABSTRACT

Jack (*Artocarpus heterophyllus*) is a multi-purpose out-breeding tree species of the family Moraceae. We generated 42,928,887 high-quality expressed sequence reads, assembled them into 89,356 unigenes, and discovered 16,853 unigene-based perfect SSRs in *A. heterophyllus*. Thirty-eight polymorphic SSRs were used to analyze the genetic diversity and population structure of 224 germplasm accessions of *A. heterophyllus* constituting three populations from three agro-climatic zones, namely Eastern Plateau and Hills, Middle Gangetic Plain Region, and Eastern Himalayan Region, encompassing five Eastern and North-Eastern states of India. At the 38 SSR loci, we detected 142 alleles with a mean of 3.74 alleles per locus. The PIC values for the loci ranged from 0.25 to 0.69. The maximum genetic diversity was recorded in Eastern Plateau and Hills (I = 0.98, He = 0.52). The ANOVA analysis indicated significantly higher within-population variation (90%) than between populations (10%). The indirect estimation of gene flow (Nm) from PhiPT indicated significant gene flow among all three populations. The population structure analysis showed at least four distinct groups among the three populations with different introgression degrees. The NJ-based clustering grouped the 224 germplasm accessions into three main clusters, each with three sub-clusters. However, we did not observe distinct geographical structure among populations except some clustering among the germplasm accessions of the populations of geographically close locations. The transcriptome dataset and the SSR markers developed in the study would boost the species' molecular characterization, conservation, and specific need-based improvement.

1. Introduction

Artocarpus heterophyllus (2n = 4X = 56), commonly known as Jack, is the most important and widespread tree of the family Moraceae (Darlington and Wylie, 1956; Thomas, 1980). It is indigenous to the Western Ghats of India and the Malaysian Rain Forests (Brown, 1941; Chandler, 1958). *A. heterophyllus* is a wonder tree, as every part of the tree is used for different purposes (Burkill and Birtwistle, 1966). However, this multi-utility tree species has received little attention from the scientific community. To date, there are minimal reports available on the development of molecular markers in *A. heterophyllus* (Kavya and Shyamalamma, 2019). Consequently, the molecular-marker-based improvement for the species lags far behind the other economically important species of the genus Artocarpus (Gardner et al., 2016). Moreover, the lack of molecular markers has hindered the absolute measurement of genetic diversity and population parameters critical to interpreting ecological-indicators-based assessments and devising effective strategies for species' conservation and sustainability (Bagley et al., 2003). Thus, it is essential to develop a broad set of suitable markers for *A. heterophyllus* for its genetic improvement and sustainable conservation.

Next-generation sequencing (NGS) technologies, coupled with

* Corresponding author.
E-mail address: binay.singh@icar.gov.in (B.K. Singh).

https://doi.org/10.1016/j.ecolind.2021.108143
Received 22 May 2020; Received in revised form 4 August 2021; Accepted 21 August 2021

powerful computational methods, have revolutionized omics-based research in plants (Yadav et al., 2018). Plant scientists use these modern, cost-effective technologies to develop large-scale genomic and transcriptomic resources in different plant/tree species (Thumilan et al., 2016). The goal of the present study was to create a comprehensive transcriptome dataset, discover a large number of transcript-based SSR markers and demonstrate their utility in providing a deep insight into the genetic background of the species.

2. Material and methods

2.1. Object of study

A. heterophyllus is an outcrossing, underutilized fruit-tree species of the family Moraceae. It is monoecious and bears unisexual flowers. People eat the syncarpous ripe fruit fresh, while the unripe fruit is used as a vegetable and has a consistency similar to meat (Pushpakumara, 2006). *A. heterophyllus* has immense potential to cater to pharmaceutical/biomaterial, paint, agro-based, and several other product-specific sectors (Jagadeesh et al., 2007).

2.2. Plant material

The study included 224 germplasm accessions of *A. heterophyllus* collected by the ICAR – National Bureau of Plant Genetic Resources (NBPGR), Regional Station, Ranchi, Jharkhand – 834 003 from three agro-climatic zones comprising five different states of India (Fig. 1).

The zone and state-wise collection of germplasm accessions are as follows: Eastern Plateau and Hills (Jharkhand – 173; Odisha – 19), Middle Gangetic Plain Region (Bihar – 23), and Eastern Himalayan Region (Assam – 08; Meghalaya – 01) (Supplementary Table 1). ICAR-NBPGR, Regional Station, Ranchi, currently maintains these germplasm accessions under ex-situ in-field conditions.

2.3. RNA extraction and cDNA library preparation

Total RNA was isolated separately from one gram of developing seeds, leaves, inflorescence, and the roots of the *A. heterophyllus* (Acc. No. IC436479) using ZR Plant RNA MiniPrep Kit (ZYMO RESEARCH, CA, USA). The RNA was qualitatively analyzed on 1.0 % denaturing agarose gel and quantified on NanoDrop (Wilmington, Delaware, USA). One μg of total RNA extracted from each of the four tissue types was mixed to prepare an RNA pool that ensured the representation of the maximum number of expressed genes in the cDNA library (Dutta et al., 2011; Singh et al., 2016). The mRNA enriched RNA pool served as the substrate to prepare a paired-end cDNA sequencing library using Illumina TruSeq RNA Sample Preparation Kit (Illumina, San Diego, USA). An Agilent 4200 TapeStation System (Agilent Technologies, CA, USA) was used for the library's quality check.

2.4. Sequencing and sequence assembly

An Illumina NextSeq1000 platform was used to sequence the library. The quality of the raw data obtained by paired-end sequencing of the normalized cDNA library was assessed using FastQC software (Andrews, 2010). The adaptor contamination and low-quality reads were removed using Trimmomatic v.0.32 software (Bolger et al., 2014). The study included only those sequence reads whose lengths were >50 nucleotides. The sequences were cleaned thoroughly by removing ambiguous reads (unknown nucleotides 'N' > 5%), low-quality reads (QV < 20), and adaptor sequences. The clean reads were subjected to the de novo assembly using Trinity-v2.5.1 software (Grabherr et al., 2011) with the

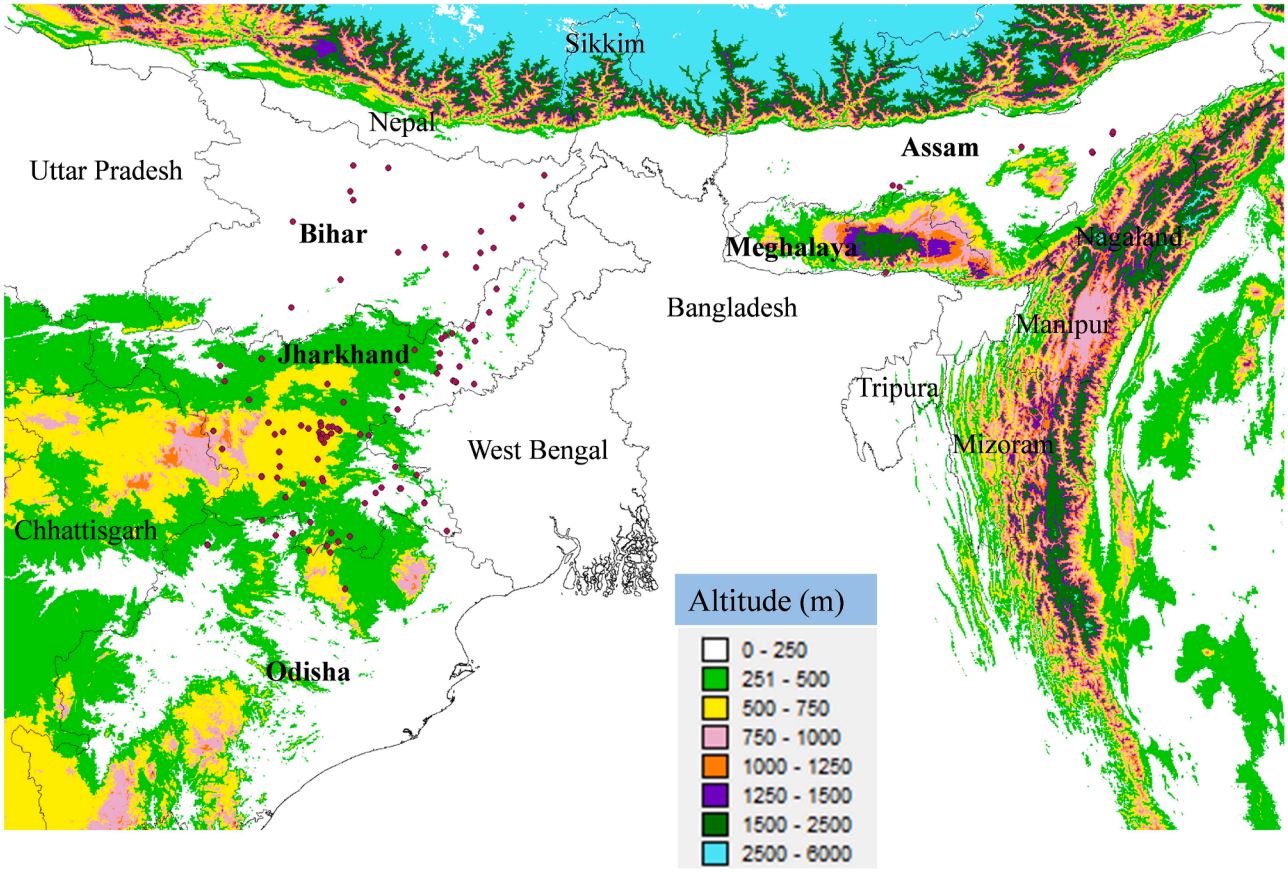

Fig. 1. Sampling distribution of germplasm accessions of *A. heterophyllus* in Eastern and North-Eastern India.

default k-mer, K = 25, and redundancies present in the assembled unigene sequences were removed using CD-HIT-EST -v4.6 software (Li and Godzik, 2006) with global sequence identity threshold cut-off set at of 90%. The assembly's reliability was validated by remapping clean reads to the assembled unigenes using BWA MEM -v0.7.12 software (Li and Durbin, 2009), and only those sequences were considered that paired adequately with a mapping quality of at least 20 (−q = 20). The assembly statistics were calculated using SAMtools (Li et al., 2009). We have deposited the sequence data at the National Centre for Biotechnology Information (NCBI) Short Read Archive (SRA) under the BioProject accession number SRR7250836.

2.5. Mining of genic-SSRs and designing of primers

MIcroSAtellite (MISA) tool (Thiel et al., 2003) was employed to detect SSRs in the assembled unigene sequences. Only the genic-SSR loci with 2–6 nucleotides simple sequences repeated at least four times were selected. Mononucleotide repeats and complex genic-SSRs were filtered out in the study. BatchPrimer3 v1.0 software (You et al., 2008) was used to design the primers for the mined genic-SSRs. The following criteria were applied for designing the primers: primer length = 18–27 bases (optimal of 22 bases), GC content = 40–60% (optimal 50%), annealing temperature = 50–60 °C (optimal 55 °C), and the product size = 100–200 bp.

2.6. DNA extraction and PCR amplification of genic-SSRs

Cetyl Trimethyl Ammonium Bromide (CTAB) method devised by Murray and Thompson (1980) was followed to extract the high molecular weight genomic DNA from young leaves of *A. heterophyllus*. DNA quality check was performed on 1.0% agarose gel. The PCR amplifications were performed in reaction mixtures (10-μl) containing purified genomic DNA (20 ng), PCR buffer (1×), $MgCl_2$ (1.5 μM), dNTPs (200 μM each), primer (250 nM each), and Taq DNA polymerase (0.25 U). All the PCR reagents were procured from Sigma-Aldrich, St. Louis, USA. The thermocycling conditions for PCR were as follows: 94 °C for 5 min followed by 35 cycles of 94 °C for 45 s, 60 °C for 30 s, 72 °C for 30 s, and a final extension at 72 °C for 7 min. MetaPhor (FMC BioProducts, Rockland, ME, USA) agarose gels (3.5%) containing ethidium bromide (0.5 ng/ml) (Sigma-Aldrich, St. Louis, USA) were used to resolve and visualize the PCR amplicons.

2.7. Genetic diversity analysis

PowerMarker v3.25 software (Liu and Muse, 2005) was used to calculate the heterozygosity, polymorphism information content (PIC) values, and pairwise genetic distance (Nei et al., 1983) between the germplasm accessions. GenAlEx 6.5 software (Peakall and Smouse, 2012) was used to calculate the population genetic parameters, the number of unique alleles, analysis of variance (ANOVA), and principal coordinates analysis (PCoA) based on Nei's genetic distance. Pairwise group PhiPT values and estimates of gene flow (Nm; Number of migrants per generation = 0.25[(1/PhiPT)–1]) were also calculated using the GenAlEx 6.5 software. The analyses involved three agro-climatic zone-based populations constituting the entire germplasm collection. The MEGA 6 software (Tamura et al., 2013) was used to generate the neighbour-joining dendrogram (cluster analysis).

2.8. Population structure analysis

The population structure of the entire collection of 224 germplasm accessions of *A. heterophyllus* was inferred by the Bayesian model-based clustering algorithm of STRUCTURE v.2.3.4 software (Pritchard et al., 2000). The K-value indicating the number of sub-populations was identified by performing Markov Chain Monte Carlo (MCMC) runs for each value of K from 1 to 10. Each run comprised of 10,000 burn-in periods and 100,000 MCMC replicates. The analyses were carried out using the admixture model option. We considered the germplasm accessions with membership fraction ≥0.8 as pure while the rest were considered as admixtures. The ΔK value was determined using Evanno's method (Evanno et al., 2005) executed in STRUCTURE HARVESTER software.

3. Results

3.1. Paired-end Illumina sequencing and de novo transcriptome assembly

The cDNA library's paired-end sequencing generated 42,928,887 high-quality reads representing a total of 6.46 Gb of expressed sequence. The de novo assembly of the sequence reads generated 89,356 unigenes with an average length of 1205 bp and the N50 value of 1703 bp. These unigenes covered 107,760,178 bp (~17.6%) of the estimated 1900 Mbp *A. heterophyllus* genome. The size of the majority of the unigenes (83.4%) ranged between 0.2 and 2.0 kbp.

3.2. Identification and validation of genic-SSRs

We identified a total of 16,853 perfect SSRs across 15,012 unigenes, accounting for 1.12 SSRs per unigene. Considering 107,760,178 bp as the unigene sequences' total size analyzed to discover SSRs, each 6.4 kbp of unigene sequence contained one SSR locus. Among the 16,853 genic-SSR loci analyzed, we identified a total of 192 distinct types of repeat motifs (Table 1). Trinucleotide SSRs were most abundant (80.57%), followed by di-, tetra-, penta- and hexanucleotide repeat with frequencies of 14.71, 3.04, 1.1, and 0.62%, respectively (Fig. 2a). Among the tri- and dinucleotide repeats, GAA/TTC and AG/CT were the most abundant motifs, with frequencies of 10.67 and 4.98%, respectively. The number of repetitions of the simple sequences ranged from 4 to 22. The frequency of reiterations ranged from 48.26% (n = 4) to <1% (n > 12) (Fig. 2b). Length-wise, 12 bp SSRs were most frequent (48.26%) followed by 15 bp (17.11 %), 18 bp (7.54%) and 20 bp (6.67%). The maximum length of the SSR was 56 bp (Fig. 2c).

Of the 16,853 genic-SSR loci identified in the study, we successfully designed PCR primers for 9469 (56.2%) of the genic-SSRs (Supplementary Table 2). From the 9469 genic-SSRs loci, we randomly selected 200 loci (n ≥ 18) to assess their utility in studying genetic diversity, and population structure of 224 germplasm accessions of *A. heterophyllus* collected from three agro-climatic zones of India, namely Eastern Plateau and Hills, Middle Gangetic Plain Region, and Eastern Himalayan Region (Supplementary Table 3). One hundred seventy of the 200 PCR primer pairs yielded a single band in *A. heterophyllus* (IC436479). Of these, 167 primer pairs yielded amplicons of the expected size, while three primer pairs amplified products that exceeded the expected size. We used all these PCR primer pairs for the preliminary polymorphism survey in a panel of eight germplasm accessions comprising the morphologically most diverse set. From the initial analysis, we identified 38 primer pairs with PIC > 0.5 and used them to analyze the genetic diversity and population structure of the entire set of 224 germplasm accessions of *A. heterophyllus*.

3.3. Genic-SSR polymorphism and frequency spectrum

We detected 142 alleles at the 38 genic-SSR loci in 224 germplasm accessions of *A. heterophyllus*. The number of alleles per locus ranged from 2 to 5, with a mean value of 3.74. The PIC values for the genic-SSR loci ranged from 0.25 to 0.69, with an average of 0.51 (Table 2). The frequency distribution analysis of the alleles indicated that 38 alleles were rare (frequency < 0.05), 74 alleles were common (frequency 0.05–0.5), and 30 alleles were frequent (frequency > 0.5). Fig. 3 depicts a representative display gel showing the profile generated by the primer JFSSR-4 in different *A. heterophyllus* germplasm accessions.

The frequency of major alleles per genic-SSR locus ranged from 0.30

Table 1
Frequency distribution of the ten most abundant SSR repeat motifs in *A. heterophyllus*.

S. No.	Repeat motif	Number of reiterations of the motif																		Total
		4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	22	
1	GAA/TTC	959	445	172	118	61	16	19	3	4	2	–	–	–	–	–	–	–	–	1799
2	AAG/CTT	773	275	131	103	50	24	15	6	–	4	–	4	–	–	–	–	–	–	1385
3	AGA/TCT	772	299	146	56	36	7	18	4	2	2	1	–	–	–	–	–	–	–	1343
4	AAT/ATT	569	222	122	73	54	8	11	4	2	–	–	–	–	–	–	–	–	–	1065
5	AG/CT	–	–	–	–	–	–	230	213	185	40	37	60	30	22	9	7	4	2	839
6	AT/AT	–	–	–	–	–	–	184	141	121	49	40	20	24	24	1	1	–	–	605
7	GA/TC	–	–	–	–	–	–	176	138	111	40	32	22	24	13	6	3	1	1	567
8	TCA/TGA	417	89	40	11	7	1	–	–	–	–	–	–	–	–	–	–	–	–	565
9	CAA/TTG	331	137	64	23	5	2	1	–	–	–	–	–	–	–	–	–	–	–	563
10	TAA/TTA	268	82	88	34	34	6	9	–	7	–	–	–	–	–	–	–	–	–	528
	Other motifs*	4045	1904	686	259	150	30	212	137	82	30	25	17	12	5	–	–	–	–	7594
	Total	8134	3453	1449	677	397	94	875	646	514	167	135	123	90	64	16	11	5	3	**16,853**

* Additional 182 types of repeat motifs in *A. heterophyllus* transcriptome.

to 0.75, with a mean value of 0.56. All the alleles except those amplified by the primers JFSSR-12 and JFSSR-172 showed a significant deviation from the Hardy–Weinberg equilibrium ($P < 0.05$). Analysis of the distribution pattern of the 142 alleles among the 224 germplasm accessions of *A. heterophyllus* constituting three populations from three distinct agro-ecological zones of India, namely Eastern Plateau and Hills, Middle Gangetic Plain Region, and Eastern Himalayan Region indicated that (1) 124 alleles were present in all the three populations, (2) seven alleles were unique to Eastern Plateau and Hills population, (3) seven alleles were present only in Eastern Plateau and Hills, and Middle Gangetic Plain Region populations, and (4) four alleles were present only in Eastern Plateau and Hills, and Eastern Himalayan Region populations. The seven alleles unique to the Eastern Plateau and Hills population were those amplified by the primers JFSSR-4, JFSSR-18, JFSSR-20, JFSSR-101, JFSSR-134, JFSSR-135, and JFSSR-172. Out of these seven unique alleles of the Eastern Plateau and Hills population, five alleles were specific to a small pocket near the Jharkhand-Odisha border area.

3.4. Genetic diversity, differentiation, gene flow, and population structure analysis

The different diversity indices calculated for each of the three populations of *A. heterophyllus* are presented in Table 3. The observed number of alleles (Na) and the effective number of alleles (Ne) per population ranged from 3.36 to 3.61 and 2.26 to 2.65. The expected heterozygosity (He) varied from 0.45 to 0.52. We recorded the maximum and the minimum genetic diversity in Eastern Plateau and Hills ($I = 0.98$, $He = 0.52$) and Eastern Himalayan Region ($I = 0.81$, $He = 0.45$) populations, respectively. The polymorphism percentage of the genic-SSRs across the three populations varied from the maximum 100% in Eastern Plateau and Hills, 89.47% in the Middle Gangetic Plain Region, and the minimum of 84.21% in the Eastern Himalayan Region populations.

The analysis of variance (ANOVA) revealed little genetic differentiation among the three populations. The level of variation among the populations was only 10 percent. The genetic variation between and within the populations was significant ($P < 0.001$). The results indicated greater within-population variation (90%) than between populations (10%), and the genetic variation within populations was the primary source of the total variation (Table 4). We observed a higher level of gene flow between the populations from the Eastern Plateau and Hills and Eastern Himalayan Region (3.12) than between populations from the Middle Gangetic Plain Region and Eastern Himalayan Region (1.48) (Supplementary Fig. 1). We observed all pairwise Nm values among the populations > 1, indicating gene flow among all the populations.

The genotypic data analysis using STRUCTURE v.2.3.4 software revealed the maximal ΔK at $K = 4$, indicating that at least four distinct groups existed among the three populations (Fig. 4a). In the bar-plot with one bar for each accession, we have indicated the subpopulations 1, 2, 3, and 4 with green, blue, red, and yellow colours, respectively. The subpopulation-wise distribution of germplasm accessions was as follows: subpopulation-I = 41, subpopulation-II = 81, subpopulation-III = 33, subpopulation-IV = 69. The number of admixtures in the subpopulations 1–4 was 14, 25, 12, and 24, respectively. The remaining germplasm accessions in the populations were pure (Fig. 4b).

3.5. Genetic relationships

The neighbour-joining (NJ) based clustering grouped the 224 germplasm accessions of *A. heterophyllus* into three main clusters designated as clusters 1, 2, and 3, with three sub-clusters each (Fig. 5). We did not, however, observe a definitive population-wise clustering in the study. All three populations showed the affiliation to all the three main clusters. The sub-cluster-wise distribution of the populations, however, indicated some correlation with their geographical origin. The germplasm accessions of the Eastern Plateau and Hills showed a general affiliation to all the sub-clusters. However, the germplasm accessions of the Middle Gangetic Plain Region and the Eastern Himalayan Region showed the selective affiliation to different sub-clusters. Since the Eastern Plateau and Hill population constituted the maximum number of germplasm accessions, we split the population into two groups: Eastern Plateau and Hills (Jharkhand) and Eastern Plateau and Hills (Odisha) for a better resolution in the sub-cluster-wise distribution of the populations. Most of the Middle Gangetic Plain Region germplasm accessions clustered with the germplasm accessions of the Eastern Plateau and Hills (Jharkhand). However, a few germplasm accessions of the Middle Gangetic Plain Region also clustered with the Eastern Plateau and Hills (Odisha) and the Eastern Himalayan Region germplasm accessions.

Similarly, most Eastern Plateau and Hills (Odisha) germplasm accessions shared the sub-clusters with the Eastern Plateau and Hills (Jharkhand) and the Eastern Himalayan Region germplasm accessions. The PCoA (Fig. 6) showed almost similar results, as observed in NJ clustering. In the PCoA, the first three coordinates explained a 63.98% cumulative variation between the germplasm accessions. The vertical and the horizontal axes individually explained 26.56 and 21.32% of the variation, respectively. The graphical dispersion of the PCoA generated a noticeable cloud for the Middle Gangetic Plain Region germplasm accessions. These germplasm accessions exhibited a considerable genetic similarity with Eastern Plateau and Hills (Jharkhand) germplasm accessions but not Eastern Plateau and Hills (Odisha) and the Eastern Himalayan Region germplasm accessions. The germplasm accessions of the Eastern Plateau and Hills (Odisha) and the Eastern Himalayan Region showed a high level of genetic similarity with each other and also with the germplasm accessions collected from the Eastern Plateau and Hills (Jharkhand).

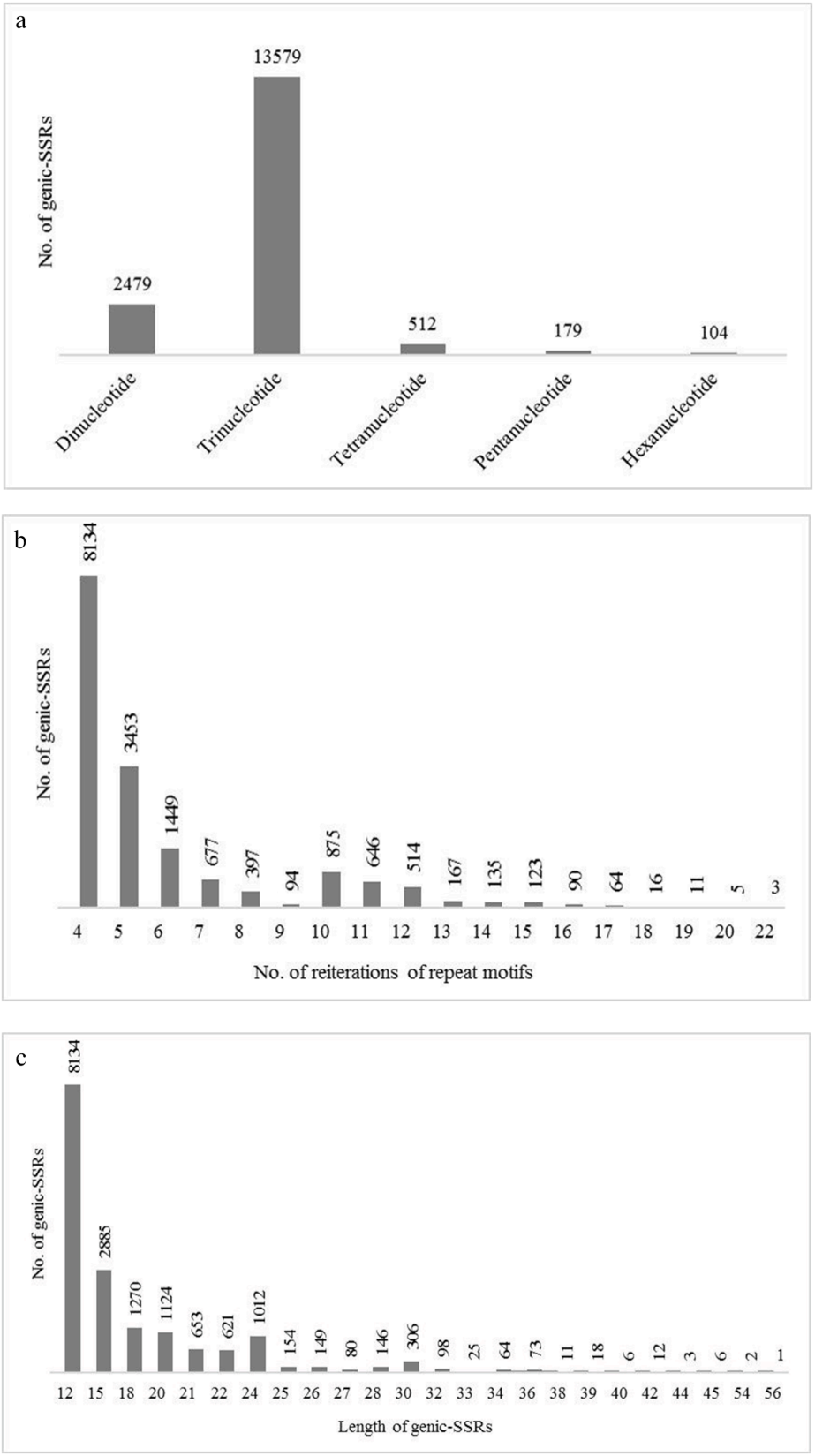

Fig. 2. Occurrence and distribution of genic-SSRs of *A. heterophyllus* (a) Length of repeat motifs, (b) Number of reiterations of repeat motifs, (c) Length of SSRs.

Table 2
Total number of alleles, frequency of the major alleles, heterozygosity, and polymorphism information content (PIC) generated by 38 primer pairs from 224 germplasm accessions of *A. heterophyllus*.

S. No.	Primer ID	Forward Primer	Reverse Primer	Expected Amplicon Size (bp)	Observed Amplicon Size (bp)	Allele No	Heterozygosity	PIC	Major Allele Frequency
1	JFSSR-2	GCAAGTATAACAACCCAAGTG	CAAAGACTCCTGAGGGAAA	374	500	2	0.21	0.25	0.72
2	JFSSR-4	ACAAGAGACTCAAGCCTCAA	TCTTTGACTGCAATGAAGC	354	354	5	0.36	0.66	0.36
3	JFSSR-8	CAACCCAAGAAGTAGTATCCA	GAGAAGGACGGAGATTCTTT	354	354	3	0.02	0.38	0.74
4	JFSSR-12	TGGATCTTGGTGTCGGTCT	TACCACTCCGTACAGTGGAAA	396	396	3	0.07	0.35	0.75
5	JFSSR-13	ACTCAGTTATTGAGCCTTCCT	CAAAGAGACGTGTGATGTTG	386	386	5	0.42	0.70	0.40
6	JFSSR-15	CTCCCTGGGAAAGGAATA	CTTAACGATTTGCATTAGGG	353	353	4	0.28	0.47	0.65
7	JFSSR-18	CACTAGTGAAACCTTCTGTGG	CTGTTCAGCAGGCAATAACT	351	351	5	0.83	0.63	0.41
8	JFSSR-19	GAGTTTATTTCTTGCCCTTG	CTCATCATTTCGCTTTAACC	352	352	5	0.54	0.63	0.50
9	JFSSR-20	AAATACCCAGCTCGCAGT	CACCAACAAACTCTTCCTTTC	370	370	5	0.61	0.49	0.63
10	JFSSR-25	GTTATCAGAATATCGACAGCAG	GTGTTATTGCAGCTCACAAG	364	364	4	0.62	0.53	0.57
11	JFSSR-28	GCCACTAGAAACTCCAAACTT	GGGAATGGGTATGAGTATGAG	363	363	4	0.28	0.48	0.63
12	JFSSR-32	CAATCTCCCACTTGTTTGT	CTGTTATATCTTCGGGAAGTG	353	353	3	0.30	0.50	0.62
13	JFSSR-35	AAGCGATCAAAGGTGACAT	CTCTCCTCCTTCTTCTTCATC	351	351	4	0.68	0.61	0.50
14	JFSSR-39	AATGGCAACTTAGCAAGGTA	ATACGATGTCCCAGAAGTCAT	389	389	3	0.41	0.51	0.58
15	JFSSR-44	CAAACCAACACTTATCACTCC	CCATCTTCAAAGTCCGTTC	368	368	4	0.52	0.52	0.62
16	JFSSR-45	TTCTCTCACAAACACCCAAA	GATGATGACAAAGGAAACTCC	353	353	4	0.35	0.69	0.30
17	JFSSR-47	GAGAAGCAACCCAATTAAGA	GTGTTCGTGGAACGTAGTG	363	502	5	0.60	0.57	0.59
18	JFSSR-48	ACCAAGTCACACATGATTACC	ACAGTGTAGGCACTGTTTCTG	374	374	4	0.69	0.50	0.61
19	JFSSR-52	GATCTCGTCGTACTTCTTGG	ACTCTCAACAATGGCAGTTT	379	379	3	0.32	0.38	0.68
20	JFSSR-64	CAGAATACGATGACAAGATGG	CTTCTTCATCGGTAAAGTGC	370	370	2	0.47	0.29	0.71
21	JFSSR-76	TCGAAGCAGACAATCAGAAT	AGAGGAGAAGGGACTGAATTT	377	377	3	0.52	0.36	0.67
22	JFSSR-88	AGAACTACTATCAATGGCATCC	CGATTCAGAAGTTTGGAGAG	370	370	3	0.58	0.41	0.62
23	JFSSR-97	AGAATCCAGGAGGTACCCTAT	TTGTGAGGAAGAAGAGGAGA	368	368	3	0.54	0.46	0.65
24	JFSSR-101	CCTAGCCGAGGAACTCTTA	GAGATGAAATGGAATGGATG	377	520	4	0.12	0.58	0.39
25	JFSSR-104	CTTCCCTGTTCATTACACAA	ACCTTGTTTGTGTTCCAGTT	354	354	4	0.40	0.49	0.65
26	JFSSR-114	AATTCAGCCAACTCAGAATC	AGCAACTATATTCGGATGGA	372	372	3	0.34	0.49	0.61
27	JFSSR-116	GAGGAAGCCATTGAACCT	CTCCGACTCCAATCTTCA	356	356	4	0.33	0.53	0.56
28	JFSSR-122	AAATCAGATGCGTTTAGGG	AAACAGAACAGACATCACTCG	381	381	3	0.10	0.45	0.58
29	JFSSR-126	TTTCCAAGTCACATGAATACC	TCGGTAAGCTTTGACTAAACA	150	150	4	0.29	0.68	0.36
30	JFSSR-132	AGCAAACAAAAGGAAAACAA	TTTGTTGAGCAACAGAAGC	153	153	4	0.21	0.64	0.38
31	JFSSR-134	AGAAACGAGAGAAGTGGAGAT	ACAATTCAACCAAACAACAAC	150	150	4	0.14	0.58	0.50
32	JFSSR-135	CAAAGTCCCTACCACCTTATC	TGCTCATCCTCATAATATTCG	147	147	4	0.01	0.51	0.60
33	JFSSR-141	CCTAGAATTACCAAGTGAGCA	CTGTCTCTGATTCAAAAATCG	155	155	3	0.34	0.50	0.61
34	JFSSR-170	CGTTTCTCCTCTGTAGTAGCA	CATTGTCTCTCTCTTCACCAG	141	141	4	0.15	0.53	0.50
35	JFSSR-172	TGGGTTAAGAGGGTTATTGAC	AAGTGCAACACTCTGAAGAAC	152	152	4	0.28	0.48	0.63
36	JFSSR-180	TACATTTGCTTCCAACTCATT	ACAGCAGAGCACTGACAATAC	159	159	4	0.38	0.67	0.34
37	JFSSR-181	CACTCAACCGTCTTCTCTTTA	CGCATAATAGCACTTTTGAGT	149	149	3	0.31	0.49	0.61

(*continued on next page*)

Table 2 (*continued*)

S. No.	Primer ID	Forward Primer	Reverse Primer	Expected Amplicon Size (bp)	Observed Amplicon Size (bp)	Allele No	Heterozygosity	PIC	Major Allele Frequency
38	JFSSR-182	CGTGTTTATGTATTGTTTCCTG	TCAATTGAGGTTTCTTGTCAT	153	153	4	0.15	0.54	0.60
Mean						**3.74**	**0.36**	**0.51**	**0.56**

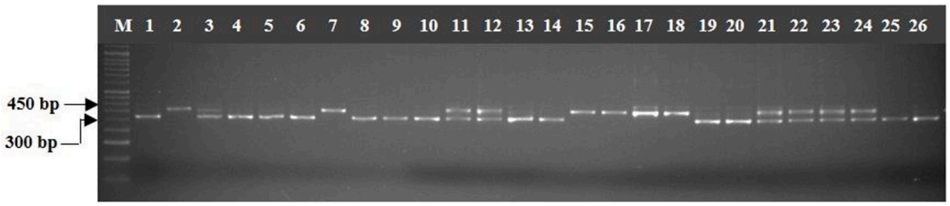

Fig. 3. Polymorphism analysis of JFSSR -4 in different *A. heterophyllus* germplasm accessions. (1) IC21617, (2) IC21619, (3) IC21622, (4) IC21625, (5) IC21627, (6) IC21628, (7) IC21629, (8) IC21630, (9) IC21631, (10) IC21632, (11) IC21633, (12) IC21635, (13) IC21636, (14) IC21637, (15) IC21638, (16) IC21640, (17) IC21642, (18) IC21643, (19) IC21644, (20) IC21645, (21) IC21646, (22) IC21647, (23) IC21649, (24) IC24323, (25) IC24324, (26) IC24325. M = 50 bp DNA marker.

Table 3
Summary statistics of genetic diversity parameters in the populations of *A. heterophyllus* calculated based on 38 polymorphic genic-SSR markers.

S. No.	Population	Na	Ne	I	He	P	P%	N	U
1.	Eastern Plateau and Hills	3.61 ± 0.13	2.65 ± 0.14	0.98 ± 0.05	0.52 ± 0.03	38	100	142	7
2.	Middle Gangetic Plain Region	3.45 ± 0.13	2.45 ± 0.12	0.89 ± 0.07	0.48 ± 0.04	34	89.47	131	0
3.	Eastern Himalayan Region	3.36 ± 0.15	2.26 ± 0.11	0.81 ± 0.06	0.45 ± 0.04	32	84.21	128	0

Na = Observed number of alleles; Ne = Effective number of alleles ; I = Shannon's nformation index; He = Expected heterozygosity; P = Number of polymorphic loci; P % = Polymorphism percentage; N = Total number of alleles; U = Number of unique alleles.

Table 4
Summary of analysis of variance (ANOVA).

Source	df	SS	MS	Estimated variance	% variance	PhiPT	p-value
Among populations	2	176.77	88.39	2.39	10	0.104	<0.001
Within populations	221	4538.84	20.54	20.54	90	–	–
Total	223	4715.62	–	22.93	100	–	–

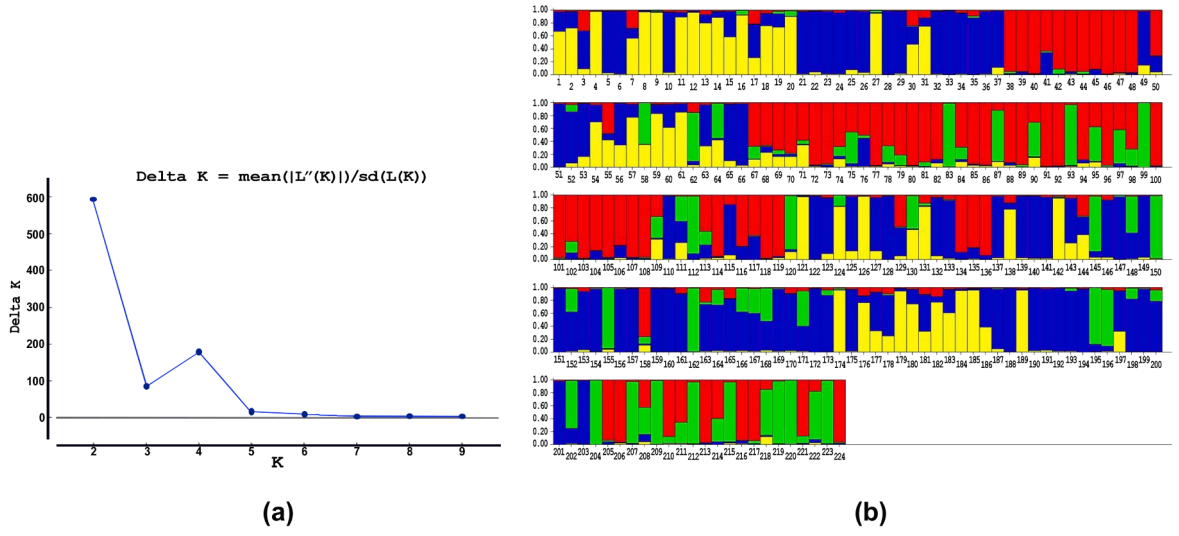

Fig. 4. Population genetic structure based on 38 genic-SSR markers in 224 germplasm accessions of *A. heterophyllus*. (a) ΔK graph showing peak value at K = 4, (b) population structure at ΔK = 4.

4. Discussion

Genetic diversity is a fundamental component of biodiversity (Hughes et al., 2008). It encompasses the extent of diversity present between a species' individuals, critical to its sustainability. A standing level of genetic diversity within a population enables it to adapt to

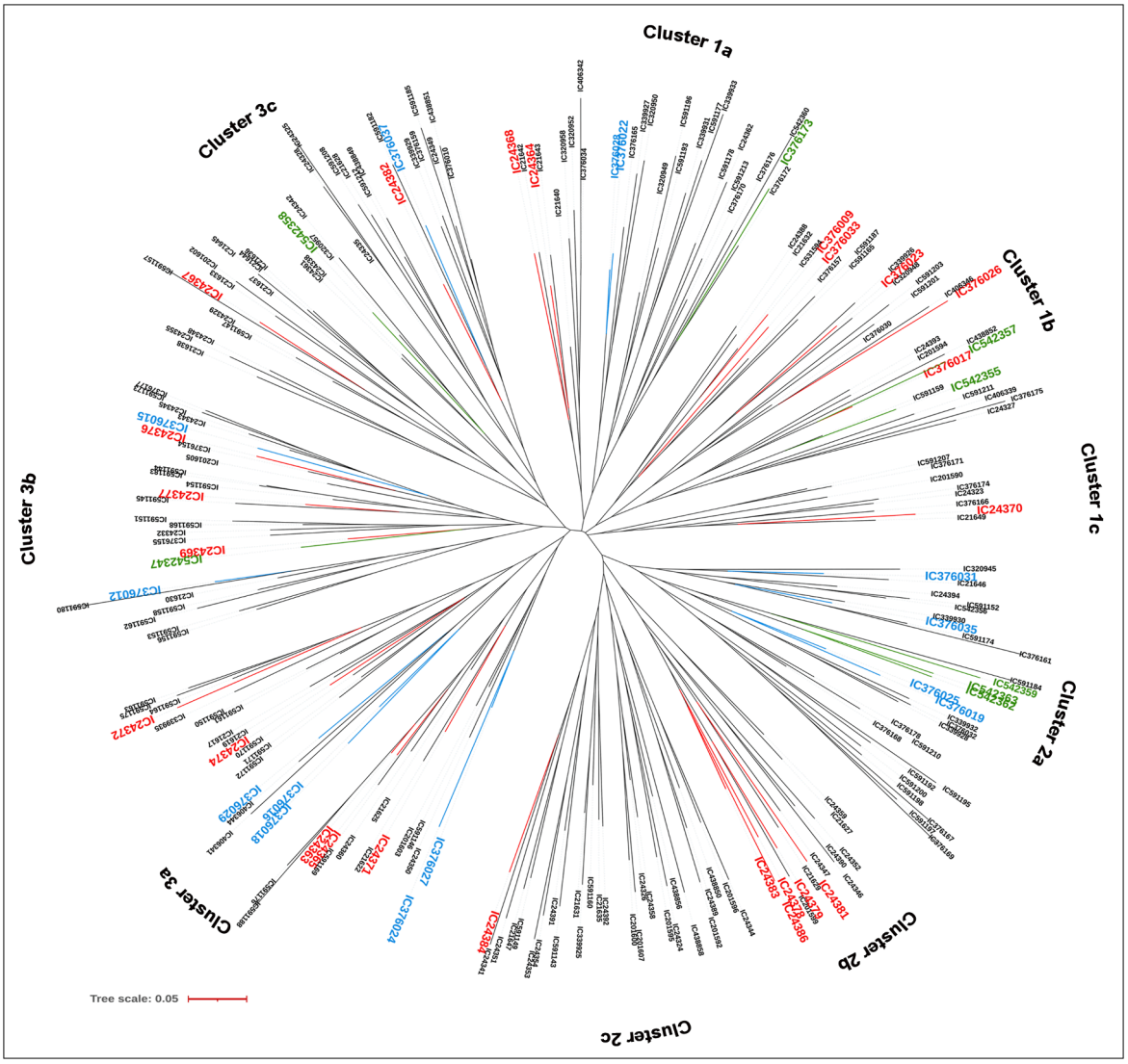

Fig. 5. NJ tree showing the genetic relationships among the 224 germplasm accessions of *A. heterophyllus*.

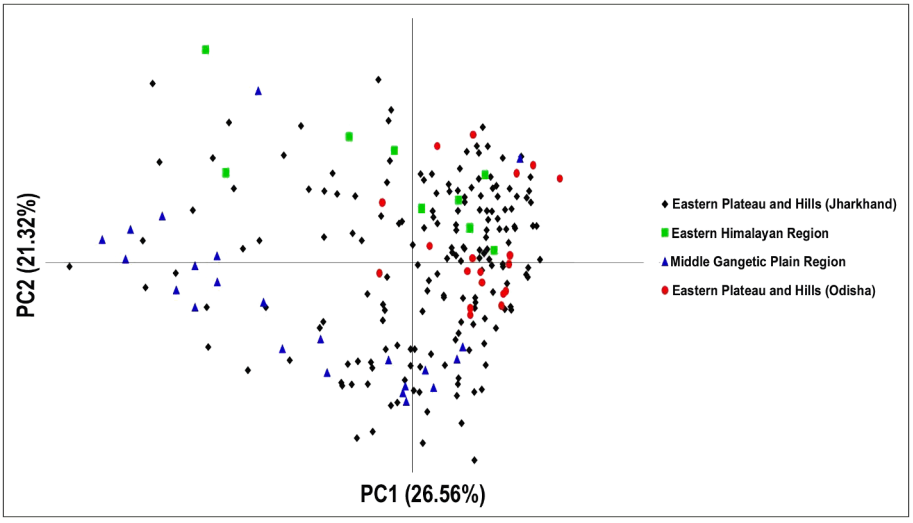

Fig. 6. PCoA based on pairwise Nei's genetic distance estimates for the populations.

changing environments. Small populations with suboptimal genetic diversity may experience fitness reductions and an increased risk of genetic erosion (Bagley et al., 2003).

A. heterophyllus is an underutilized fruit tree species. It has been introduced and is now both naturalized and cultivated in many tropical countries. ICAR-NBPGR, Regional Station, Ranchi, made a large-scale collection of *A. heterophyllus* from India's three different agro-climatic zones and maintaining them under ex-situ in-field conditions. However, any genetic improvement or diversity study could not be taken up due to the lack of suitable markers for the species.

In recent years, SSR markers are proven to be very useful for various molecular genetic and genomic studies (Varshney et al., 2005). To provide impetus to such studies in *A. heterophyllus*, we generated global transcriptome data and developed a comprehensive set of genic-SSRs for the species. We used the Illumina NextSeq1000 platform for the paired-end sequencing of the normalized cDNA library prepared from an equimolar mixture of total RNA extracted from developing seeds, leaves, inflorescence, and the roots of *A. heterophyllus*. A normalized cDNA sequencing library prepared from the pooled RNA generates maximally informative sequence resources (Hoang et al., 2019). The paired-end sequencing of a cDNA library often yields reads with significant overlaps critical for their assembly. It is beneficial for non-model tree species lacking reference genome sequence (Martin and Wang, 2011). We generated 42,928,887 high-quality reads and assembled them into 89,356 unigenes with an average length of 1205 bp and the N50 value of 1703 bp. The average and the N50 contig lengths obtained in our study were comparable or even better than many earlier reports (Singh et al., 2016; Singh et al., 2018). It implies that our sequence assembly was accurate and effective (Ge et al., 2019). From the assembled unigenes, we identified 16,853 perfect SSRs, accounting for one SSR per 6.4 kbp of unigene. The frequency of the genic-SSRs obtained in our study is higher than the earlier reports in wheat, tomato, poplar (Cardle et al., 2000; Kantety et al., 2002), comparable to those in cassava, barley, Arabidopsis, soybean (Cardle et al., 2000; Raji et al., 2009; Thiel et al., 2003) but significantly lower than in coffee, rubber, rice (Cardle et al., 2000; Aggarwal et al., 2007; Feng et al., 2009). The size of the genome, redundant unigenes, and the SSR detection tools and parameters are often accountable for these variations (Varshney et al., 2005). As to repeat motif frequencies and the length of the genic-SSRs, our results corroborate the earlier report in *Morus alba* (Thumilan et al., 2016), a closely related species of *A. heterophyllus*.

Of the 16,853 genic-SSR loci identified in the study, we could design PCR primers for only 9469 (56.2%) loci, possibly due to too short flanking sequences or the failure of the flanking sequences to meet the criteria for primer designing using BatchPrimer3 v 1.0 software (Dutta et al., 2011). We tested 200 randomly selected genic-SSRs, out of which 167 (83.5%) showed successful amplification. Earlier reports indicate that generally, 60–90% of genic-SSR loci show successful amplification in PCR assays (Liang et al., 2009; Saha et al., 2004; Wang et al., 2011). The quality of the sequence and the location of the primers within the SSR-containing genes largely determine the success of PCR assays (Varshney et al., 2005). The failure to amplify some of the loci in our study may be possible due to the presence of large-sized introns within the SSR containing unigenes (Singh et al., 2016).

The number of alleles present in a population (allelic richness) offers a significant measure of the population's long-term potential for adaptability to the changing environment (Greenbaum et al., 2014). The mean value of (3.74) genic-SSR alleles per locus recorded in our study was lower than that of 10.9 alleles per locus recorded in an earlier study on a Ugandan collection of *A. heterophyllus* (Nakintu et al., 2020). The moderate allelic richness and the prevalence of rare alleles (26.8%) obtained in the study underscore the evolutionary potential of the populations considered for analysis (Caballero et al., 2013; Hughes et al., 2008). Besides, it highlights the populations' exceptional ability to adapt to future environmental changes (Fisher, 1930; Wagner, 2008). A significant deviation ($P < 0.05$) of the majority of alleles from Hardy–Weinberg equilibrium highlights the complex interplay of evolutionary forces as well as humanmade selection pressure, referred to as founding events, on the populations (Dlugosch and Parker, 2008). The study revealed the presence of a few unique alleles in the populations. These alleles may serve as valuable markers in asserting germplasm accessions' genetic identity during germplasm characterization (Kalinowski, 2004). The prevalence of unique alleles of the Eastern Plateau and Hills population near the Jharkhand-Odhisa border area indicates a priority site for the in-situ conservation of *A. heterophyllus*. As the crop continues to gain global popularity as an industrial crop, failure to take timely conservation initiatives may lead to losing these locally adapted alleles due to intentional selection processes (Peñas et al., 2016; Witherup et al., 2019).

The assessment of genetic variation within a population forms the basis for selecting, improving, and conserving the species (Rao and Hodgkin, 2002). The study observed the maximum ($I = 0.98$, $He = 0.52$) and minimum ($I = 0.81$, $He = 0.45$) genetic diversity in Eastern Plateau and Hills and Eastern Himalayan Region populations. The higher genetic diversity explained by the newly developed genic-SSR markers among the populations substantiate their efficiency as effective markers for diversity studies.

The analysis of variance (ANOVA) revealed little genetic differentiation among the three populations. The results indicated more significant within-population variation than between populations, and the genetic variation within populations was the primary source of the total variation. In general, migration enhances the genetic variability within populations and homogenizes genetic differences among populations. It is particularly relevant for species with high or moderate dispersal abilities. Out-breeding plants often show higher variation levels within populations and less between populations (Hamrick and Godt 1990; Yang et al., 2019). *A. heterophyllus* is an out-breeding species (Moncur, 1985). An outcrossing mating system and long-distance seed dispersal in the genus Artocarpus promote gene flow among populations. Besides, these are most likely responsible for reducing population structuring and preserving a high genetic diversity level. We indirectly estimated the magnitude of gene flow (Nm) in the genus Artocarpus from the PhiPT values. The indirect estimation of gene flow involves several assumptions that are often not valid for natural populations (Whitlock and McCauley, 1999). Nevertheless, they can still provide helpful information about the approximate magnitude of gene flow (Neigel, 2002). All pairwise Nm values recorded in the study were >1, indicating gene flow among all the populations. The prevalence of gene flow among the geographically distant populations may be due to the dispersal of seeds by human migrations, which are very common in India. These migrations often occur due to poverty, lack of local options, and work availability elsewhere. The population structure analysis revealed different degrees of introgressions among the populations, an evident phenomenon for out-breeding populations.

Both NJ and Bayesian model-based clustering studies failed to indicate any definitive clustering among the germplasm accessions. The non-clustering of germplasm accessions, even from the same agro-climatic zones, may be attributed primarily to the anthropogenic factors leading to the movement of seed/propagules across the agro-climatic zones (Asfaw et al., 2009). A significant difference in the number of germplasm accessions across the populations may also be responsible for a non-conclusive clustering pattern observed in the study. The origin of a considerable number of germplasm accessions from the areas near the geographical boundaries shared by agro-climatic zones, on the other hand, explains the cross-population sub-clustering of some of the germplasm accessions. The distortion in the populations' geographical structuring may be due to the complex histories of human migration in the region (Kundu and Niranjan, 2007) coupled with the cultural association of jackfruit since prehistoric times (Soepadmo, 1992). Our report agrees with an earlier report from China, wherein AFLP-based genetic diversity analysis did not find a significant correlation between genetic relationship and geographical origin within a

total of 50 germplasm accessions from three provinces of China (Yingzhi et al., 2010).

5. Conclusions

We developed the first dataset of expressed sequences and discovered a total of 16,853 genic-SSR markers in *A. heterophyllus*. Further, we demonstrated the genic-SSR markers' utility in providing a deep insight into the species' genetic background. We used a set of 38 genic-SSR markers to analyze the genetic diversity and population structure of 224 germplasm accessions of *A. heterophyllus* constituting three populations from different agro-climatic zones of India. The analysis revealed a moderate allelic richness and the prevalence of rare alleles in the populations, highlighting their exceptional ability to adapt to future environmental changes. Moreover, it indicated that the Eastern Plateau and Hills population harbour the maximum genetic diversity among the populations used in the study. The Eastern Plateau and Hills' germplasm accessions also contain a considerable number of private alleles based on which we identified a small pocket near the Jharkhand-Odisha border area as a priority site for the in-situ conservation of *A. heterophyllus*. We observed most of the genetic diversity among individuals within a population that underlines individual-level selection for genetic improvement of the species. Overall, our study provided an efficient tool for genetic analysis and opened up an avenue for more extensive future studies in the genus Artocarpus.

CRediT authorship contribution statement

Devendra K. Singh: Methodology, Investigation. **Avinash Pandey:** Formal analysis, Validation. **Shashi Bhushan Choudhary:** Resources. **Sudhir Kumar:** Formal analysis, Validation. **Kishor U. Tribhuvan:** Data curation, Software, Formal analysis. **Dwijesh C. Mishra:** Formal analysis, Methodology, Software. **Jyotika Bhati:** Methodology, Software. **Madan Kumar:** Formal analysis, Methodology. **J.B. Tomar:** Resources. **S.K. Bishnoi:** Resources. **M.A. Mallick:** Visualization, Supervision. **V.P. Bhadana:** Resources, Supervision. **T.R. Sharma:** Resources, Supervision, Visualization, Writing - review & editing. **A. Pattanayak:** Funding acquisition, Supervision, Writing - review & editing. **Binay K. Singh:** Conceptualization, Project administration, Validation, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We sincerely acknowledge the Director, ICAR – Indian Institute of Agricultural Biotechnology, Ranchi 834 010, Jharkhand, India, for providing financial support and the facilities to carry out this research work. We also gratefully acknowledge the Director, ICAR – National Bureau of Plant Genetic Resources, New Delhi 110 012, India, to provide plant materials of *A. heterophyllus* maintained at its Regional Centre located at Ranchi.

Funding

This study was conducted under the project "Development of transcriptome-based resources for indigenous agri-horticultural crops of eastern India," funded by ICAR - Indian Institute of Agricultural Biotechnology, Ranchi 834 010, Jharkhand, India.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2021.108143.

References

Aggarwal, R.K., Hendre, P.S., Varshney, R.K., Bhat, P.R., Krishnakumar, V., Singh, L., 2007. Identification, characterization, and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor. Appl. Genet. 114 (2), 359–372. https://doi.org/10.1007/s00122-006-0440-x.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.

Asfaw, A., Blair, M.W., Almekinders, C., 2009. Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. Theor. Appl. Genet. 120 (1), 1–12. https://doi.org/10.1007/s00122-009-1154-7.

Bagley, M.J., Franson, S.E., Christ, S.A., Waits, E.R., Toth, G.P., 2003. Genetic Diversity as an Indicator of Ecosystem Condition and Sustainability: utility for Regional Assessments of Stream Condition in the Eastern United States. US Environmental Protection Agency, Cincinnati, OH.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brown, W.H., 1941. Useful Plants of the Philippines, Dept. Agric. Natl. Res. Tech. Bull. 10. Bureau of Printing, Manila.

Burkill, I.H., Birtwistle, W., 1966. A Dictionary of the Economic Products of the Malaya Peninsula, second ed. Kuala Lumpur, Malaysia.

Caballero, A., García-Dorado, A., 2013. Allelic diversity and its implications for the rate of adaptation. Genetics 195 (4), 1373–1384. https://doi.org/10.1534/genetics.113.158410.

Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D., Waugh, R., 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156 (2), 847–854.

Chandler, W.H., 1958. Evergreen Orchards, second ed. Henry Kimpton, London.

Darlington, C.D., Wylie, A.P., 1956. Chromosome Atlas of Flowering Plants, second ed. The Macmillan Co., New York.

Dlugosch, K.M., Parker, I.M., 2008. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. Mol. Ecol. 17 (1), 431–449. https://doi.org/10.1111/j.1365-294X.2007.03538.x.

Dutta, S., Kumawat, G., Singh, B.P., Gupta, D.K., Singh, S., Dogra, V., Gaikwad, K., Sharma, T.R., Raje, R.S., Bandhopadhya, T.K., Datta, S., Singh, M.N., Bashasab, F., Kulwal, P., Wanjari, K.B., K Varshney, R., Cook, D.R., Singh, N.K., 2011. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. B.M.C Plant Biol. 11 (1), 17. https://doi.org/10.1186/1471-2229-11-17.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14 (8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Feng, S.P., Li, W.G., Huang, H.S., Wang, J.Y., Wu, T.Y., 2009. Development, characterization and cross-species/genera transferability of EST–SSR markers for rubber tree (*Hevea brasiliensis*). Mol. Breed. 23 (1), 85–97. https://doi.org/10.1007/s11032-008-9216-0.

Fisher, R., 1930. The Genetical Theory of Natural Selection. Oxford University Press, UK.

Gardner, E.M., Johnson, M.G., Ragone, D., Wickett, N.J., Zerega, N.J.C., 2016. Low-coverage, whole-genome sequencing of *Artocarpus camansi* (Moraceae) for phylogenetic marker development and gene discovery. Appl. Plant Sci. 4 (7), 1600017. https://doi.org/10.3732/apps.1600017.

Ge, H., Lin, K., Shen, M., Wu, S., Wang, Y., Zhang, Z., Wang, Z., Zhang, Y., Huang, Z., Zhou, C., Lin, Q., Wu, J., Liu, L., Hu, J., Huang, Z., Zheng, L., 2019. De novo assembly of a chromosome-level reference genome of red-spotted grouper (*Epinephelus akaara*) using nanopore sequencing and Hi-C. Mol. Ecol. Resour. 19 (6), 1461–1469. https://doi.org/10.1111/1755-0998.13064.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., de Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29 (7), 644–652. https://doi.org/10.1038/nbt.1883.

Greenbaum, G., Templeton, A.R., Zarmi, Y., Bar-David, S., 2014. Allelic richness following population founding events – a stochastic modelling framework incorporating gene flow and genetic drift. PLoS ONE 9 (12). https://doi.org/10.1371/journal.pone.0115203.

Hamrick, J.L., Godt, M.J.W., 1990. Allozyme diversity in plant species. In: Brown, A.D. H., Brown, M.T., Clegg, A.L., Weir, Kahler, B.S. (Eds.), Plant Population Genetics, Breeding and Germplasm Resource. Sinauer, Massachusetts, pp. 43–63.

Hoang, N.V., Furtado, A., Perlo, V., Botha, F.C., Henry, R.J., 2019. The impact of cDNA normalization on long-read sequencing of a complex transcriptome. Front. Genet. 10, 654. https://doi.org/10.3389/fgene.2019.00654.

Hughes, A.R., Inouye, B.D., Johnson, M.T., Underwood, N., Vellend, M., 2008. Ecological consequences of genetic diversity. Ecol. Lett. 11 (6), 609–623. https://doi.org/10.1111/j.1461-0248.2008.01179.x.

Jagadeesh, S.L., Reddy, B.S., Swamy, G.S.K., Gorbal, K., Hegde, L., Raghavan, G.S.V., 2007. Chemical composition of jackfruit (*Artocarpus heterophyllus* Lam.) selections of

Western Ghats of India. Food Chem. 102 (1), 361–365. https://doi.org/10.1016/j.foodchem.2006.05.027.
Kalinowski, S.T., 2004. Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv. Genet. 5 (4), 539–543. https://doi.org/10.1023/B:COGE.0000041021.91777.1a.
Kantety, R.V., Rotal, M.L., Matthews, D.E., Sorrells, M.E., 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol. Biol. 48, 501–510. https://doi.org/10.1023/A:1014875206165.
Kavya, G.K., Shyamalamma, S., 2019. Characterization of selected jackfruit germplasm accessions for fruit shape through morphological and marker-based assay. Indian J. Agric. Res. https://doi.org/10.18805/IJARe.A-5427.
Kundu, A., Niranjan, S., 2007. Migration, employment status and poverty: an analysis across urban centres. Econ. Polit. Wkly. 42 (4), 299–306.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25 (14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25 (16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22 (13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.
Liang, X., Chen, X., Hong, Y., Liu, H., Zhou, G., Li, S., Guo, B., 2009. Utility of EST-derived S.S.R. in cultivated peanut (Arachis hypogaea L.) and Arachis wild species. B.M.C Plant Biol. 9, 35. https://doi.org/10.1186/1471-2229-9-35.
Liu, K., Muse, S.V., 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21 (9), 2128–2129. https://doi.org/10.1093/bioinformatics/bti282.
Martin, J.A., Wang, Z., 2011. Next-generation transcriptome assembly. Nat. Rev. Genet. 12 (10), 671–682. https://doi.org/10.1038/nrg3068.
Moncur, M.W., 1985. Floral ontogeny of the jackfruit, *Artocarpus heterophyllus* Lam. Aust. J. Bot. 33, 585–593.
Murray, M.G., Thompson, W.F., 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8 (19), 4321–4326. https://doi.org/10.1093/nar/8.19.4321.
Nakintu, J., Albrecht, C., Müller, C.M., Kagoro-Rugunda, G., Andama, M., Olet, E.A., Lejju, J.B., Gemeinholzer, B., 2020. Exploring the genetic diversity of jackfruit (*Artocarpus heterophyllus* Lam.) grown in Uganda based on SSR markers. Genet. Resour. Crop Evol. 67, 605–619. https://doi.org/10.1007/s10722-019-00830-5.
Nei, M., Tajima, F., Tateno, Y., 1983. Accuracy of estimated phylogenetic trees from molecular data. J. Mol. Evol. 19 (2), 153–170. https://doi.org/10.1007/BF02300753.
Neigel, J., 2002. Is Fst obsolete? Conserv. Genet. 3, 167–173. https://doi.org/10.1023/A:1015213626922.
Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics 28 (19), 2537–2539. https://doi.org/10.1093/bioinformatics/bts460.
Peñas, J., Barrios, S., Bobo-Pinilla, J., Lorite, J., Martínez-Ortega, M.M., 2016. Designing conservation strategies to preserve the genetic diversity of Astragalus edulis Bunge, an endangered species from western Mediterranean region. PeerJ 4. https://doi.org/10.7717/peerj.1474 e1474.
Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155 (2), 945–959.
Pushpakumara, D.K.N.G., 2006. Floral and fruit morphology and phenology of *Artocarpus heterophyllus* Lam. (Moraceae). Sri. Lankan J. Agric. Sci. 43, 82–106.
Raji, A.AJ., Anderson, J.V., Kolade, O.A., Ugwu, C.D., Dixon, A.GO., Ingelbrecht, I.L., 2009. Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, polymorphisms, and cross-taxa utility. B.M.C Plant Biol. 9 (1), 118. https://doi.org/10.1186/1471-2229-9-118.
Rao, R., Hodgkin, T., 2002. Genetic diversity and conservation and utilization of plant genetic resources. Plant Cell Tiss Org. 68, 1–19. https://doi.org/10.1023/A:1013359015812.
Saha, M.C., Mian, M.A., Eujayl, I., Zwonitzer, J.C., Wang, L., May, G.D., 2004. Tall fescue EST-SSR markers with transferability across several grass species. Theor. Appl. Genet. 109 (4), 783–791. https://doi.org/10.1007/s00122-004-1681-1.
Singh, B.K., Choudhary, S.B., Yadav, S., Malhotra, E.V., Rani, R., Ambawat, S., Priyamedha, Pandey, A., Kumar, R., Kumar, S., Sharma, H.K., Singh, D.K., Rai, P.K., 2018. Genetic structure identification and assessment of interrelationships between Brassica and allied genera using newly developed genic-SSRs of Indian Mustard (*Brassica juncea* L.). Ind. Crops Prod. 113, 111–120. https://doi.org/10.1016/j.indcrop.2018.01.023.
Singh, B.K., Mishra, D.C., Yadav, S., Ambawat, S., Vaidya, E., Tribhuvan, K.U., Kumar, A., Kumar, S., Kumar, S., Chaturvedi, K.K., Rani, R., Yadav, P., Rai, A., Rai, P. K., Singh, V.V., Singh, D., 2016. Identification, characterization, validation and cross-species amplification of genic-SSRs in Indian mustard (Brassica juncea). J. Plant Biochem. Biotechnol. 25 (4), 410–420. https://doi.org/10.1007/s13562-016-0353-y.
Soepadmo, E., 1992. *Artocarpus heterophyllus* Lam. In: Verheij, E.W.M., Coronel, R.E. (Eds.), Plant Resources of South-East Asia 2, Edible Fruits and Nuts. PROSEA, Bogor, Indonesia.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30 (12), 2725–2729. https://doi.org/10.1093/molbev/mst197.
Thiel, T., Michalek, W., Varshney, R.K., 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. https://doi.org/10.1007/s00122-002-1031-0.
Thomas, C.A., 1980. Jackfruit, *Artocarpus heterophyllus* (Moraceae), as source of food and income. Econ. Bot. 34 (2), 154–159. https://doi.org/10.1007/BF02858629.
Thumilan, M., Sajeevan, R.S., Biradar, J., Madhuri, T., Nataraja, K.N., Sreeman, S.M., Parida, S.K., 2016. Development and characterization of genic SSR markers from Indian Mulberry transcriptome and their transferability to related species of Moraceae. PLoS ONE. 11 (9), e0162909. https://doi.org/10.1371/journal.pone.0162909.
Varshney, R.K., Graner, A., Sorrells, M.E., 2005. Genic microsatellite markers in plants: features and applications. Trends Biotechnol. 23 (1), 48–55. https://doi.org/10.1016/j.tibtech.2004.11.005.
Wagner, A., 2008. Robustness and evolvability: a paradox resolved. Proc. Biol. Sci. 275 (1630), 91–100. https://doi.org/10.1098/rspb.2007.1137.
Wang, Z., Li, J., Luo, Z., Huang, L., Chen, X., Fang, B., Li, Y., Chen, J., Zhang, X., 2011. Characterization and development of EST-derived S.S.R. markers in cultivated sweet potato (*Ipomoea batatas*). B.M.C Plant Biol. 11 (1), 139. https://doi.org/10.1186/1471-2229-11-139.
Whitlock, M.C., McCauley, D.E., 1999. Indirect measures of gene flow and migration: FST not equal to 1/(4Nm+1). Heredity (Edinb). 82, 117–125. https://doi.org/10.1038/sj.hdy.6884960.
Witherup, C., Zuberi, M.I., Hossain, S., Zerega, N.J.C., 2019. Genetic diversity of Bangladeshi jackfruit (*Artocarpus heterophyllus*) over time and across seedling sources. Econ. Bot. 73 (2), 233–248. https://doi.org/10.1007/s12231-019-09452-5.
Yadav, P., Vaidya, E., Rani, R., Yadav, N.K., Singh, B.K., Rai, P.K., Singh, D., 2016. Recent perspective of next generation sequencing: applications in molecular plant biology and crop improvement. Proc. Natl. Acad. Sci., India - Sect. B: Biol. Sci. 88 (2), 435–449. https://doi.org/10.1007/s40011-016-0770-7.
Yang, S., Xue, S., Kang, W., Qian, Z., Yi, Z., Chiang, T.-Y., 2019. Genetic diversity and population structure of *Miscanthus lutarioriparius*, an endemic plant of China. e0211471 PLoS ONE 14 (2). https://doi.org/10.1371/journal.pone.0211471.
Ying-zhi, L.I., Qi, M.A.O., Feng, F.E.N.G., Chun-hai, Y.E., 2010. Genetic diversity within a jackfruit (*Artocarpus heterophyllus* Lam.) germplasm collection in China using AFLP markers. Agril. Sci. China. 9 (9), 1263–1270. https://doi.org/10.1016/S1671-2927(09)60215-7.
You, F.M., Huo, N., Gu, Y., Luo, M.-cheng., Ma, Y., Hane, D., Lazo, G.R., Dvorak, J., Anderson, O.D., 2008. BatchPrimer3: A high throughput web application for PCR and sequencing primer design. B.M.C. Bioinform. 9 (1), 253. https://doi.org/10.1186/1471-2105-9-253.