ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2020) **18**(3); 159–167
doi:10.1017/S1479262120000167

Elucidation of genetic diversity base in *Calotropis procera* – a potentially emerging new fibre resource

Aasim Majeed†, Bhawana Goel†, Vandana Mishra†, Ravinder Kohli and Pankaj Bhardwaj*

Molecular Genetics Laboratory, Department of Botany, Central University of Punjab, Bathinda, India

Received 16 December 2019; Accepted 19 June 2020 – First published online 16 July 2020

Abstract

Calotropis procera is emerging as a new, yet undomesticated, resource of fibre comparable to cotton and kapok. Screening of efficient genotypes from its wild populations would be a useful pre-domestication process. The desired genotypes can then be improved through conventional breeding programmes to develop a domesticated variety. Molecular markers play a major role in modern breeding systems. Thus, an efficient marker resource for *C. procera* would prove useful in germplasm selection during breeding programmes. In this study, we undertook an initial step of Simple sequence repeats (SSR) marker development for *C. procera*, which could be applied for germplasm selection. Furthermore, using the developed markers, we assessed the genetic diversity base within its wild populations which could be useful to identify the hotspot areas of germplasm collection. Out of 94,636 *de novo* assembled transcripts, 9148 sequences were found to contain 12,884 SSRs at a density of 5.5 SSRs/Mb. Twelve SSRs were found as polymorphic with a mean polymorphic information content of 0.575. We observed a moderate level of genetic diversity ($N_a = 3.625$, $H_o = 0.58$) in the studied populations. Mantel's test showed significant correlation between the geographic distance and the genetic distance ($r = 0.147$, $P = 0.010$). Sirsa was found as a genetically most diverse population followed by Barnala while Gurdaspur was found with the least genetic diversity. These genetically diverse populations can serve as an important resource for effective germplasm collection for breeding programmes.

Keywords: genetic diversity, northern plains of India, RNAseq, SSR markers

Introduction

Calotropis procera is perennial evergreen milkweed, which belongs to the family Asclepiadaceae. It is adapted to grow under saline soil and dry conditions. The fibre of *C. procera* is mainly composed of cellulose, hemicellulose and lignin. It is used for making bowstrings, fishing nets and ropes. It has higher tensile strength and water retention capacity than cotton and wool. The fibre exhibits high-insulating features, and has higher crystallinity and breaking strength than kapok. However, its breaking resistance is not strong as cotton (Qi *et al.*, 2018). It is biodegradable and renewable and has the potential to serve as a new member of natural fibre sources in textile industries (Sakthivel *et al.*, 2005; Cheema *et al.*, 2010). *Calotropis gigantea* is a closely related species to *C. procera* and produces a similar kind of fibre. Its blend with cotton is patented by Shanghai Palando Textile Technology Development Co., Ltd. (Patent No. CN105887296A, https://patents.google.com/patent/CN105887296A/en). The fibres from its bark and seeds were found to have enough potential to replace or supplement other fibrous raw materials as reinforcing agent (Ashori and Bahreini, 2009) and can also be used in generating natural fibre-reinforced composites (Ga *et al.*, 2014). Thus, the fibres of *C. procera* and *C. gigantea* have great potential in the industrial sector and can serve as an alternative source of fibre. Besides, their latex is widely known for anticancer, antibacterial, antifungal and anti-

*Corresponding author. E-mail: pankajihbt@gmail.com, pankajbhardwaj@cup.edu.in
†These authors contributed equally.

inflammatory properties (Choedon *et al.*, 2006; Silva *et al.*, 2010; Sobrinho *et al.*, 2013).

The idea of *C. procera* as a source of high-quality fibre is relatively naive. The economic and medical importance of *C. procera* allures towards its domestication, which requires identification, and breeding of desired genotypes to develop a variety of appropriate traits. An effective SSR marker resource needs to be developed for this species which could play an important role in germplasm selection during breeding programmes. Besides, these markers could prove useful for elucidation of its genetic diversity base over a landscape. Here, we attempted to design an effective SSR marker resource through an RNAseq approach followed by characterization to assess their polymorphic potential. Furthermore, the polymorphic SSR markers were used to elucidate the genetic diversity of *C. procera* populations in Indian northwestern plains to recognize the hub areas for germplasm collection. The cross-species transferability of *C. procera* with *C. gigantea* was also assessed for the effectiveness of the developed markers in cross-species studies.

Materials and methods

Sample collection and DNA isolation

A total of 79 leaf samples of *C. procera* were collected from the regions of Indian states of Delhi, Haryana and Punjab, constituting six populations. Eleven samples of *C. gigantea* were collected from Gorakhpur (Uttar Pradesh), India and used for cross-transferability analysis. These samples constitute the seventh population. The details of sample locations are given in online Supplementary information S1. The genomic DNA was isolated by the Cetyl Trimethyl Ammonium Bromide (CTAB) method (Doyle, 1987). Both the quality and quantity of the isolated DNA were assessed through agarose gel electrophoresis and Nanodrop spectrophotometer (Nanodrop 2000 Thermofisher, USA), respectively.

De novo *assembly*

Paired-end raw reads were retrieved from the NCBI SRA database (accession: ERX2103344) (Mutwakil *et al.*, 2017). Quality checks and trimming of the raw reads were performed through FastQC (Andrews, 2010) and Trim galore (https://github.com/FelixKrueger/TrimGalore), respectively. *De novo* assembly was generated using Trinity (Grabherr *et al.*, 2011). The removal of redundant sequences and generation of unigenes was performed through CD-HIT-EST (Li and Godzik, 2006) at a sequence similarity threshold of 95%. The quality of the assembly was assessed through read-representation wherein raw reads were mapped back to the assembly using Bowtie2 (Langmead, 2010). The quality was also assessed through Blastx against the SwissProt database for counting the number of full-length transcripts. The completeness assessment of the assembly was performed using BUSCO v2 (Simão *et al.*, 2015).

SSR screening and characterization

Screening of the transcripts for SSR regions was performed using MISA (MicroSAtellite identification tool) (http://pgrc.ipk-gatersleben.de/misa/). The coding sequences (CDS) were identified through ORF predictor v2.3 (Min *et al.*, 2005). The positions of the identified SSRs and the ORFs were correlated to find the location of the SSRs within the transcripts using an in-house Python script. The SSRs present at the 5′-UTR region were selected for primer designing. The primers were designed for the selected SSRs using Batch Primer3 (You *et al.*, 2008). Thirty-six primer pairs were synthesized for analysing the polymorphism among the selected accessions. The characterization was performed through a non-denaturing PAGE followed by silver staining (Huang *et al.*, 2018). The resulting polymorphic primer pairs were subsequently utilized for genotyping of 79 individuals of *C. procera* to elucidate its genetic diversity base in northwestern plains of India. The polymorphic markers were also utilized for cross-transferability analysis with *C. gigantea*.

Data analysis

The polymorphic information content (PIC), Hardy-Weinberg equilibrium (HWE), total number of alleles (N_a), the effective number of alleles (N_e), Shannon's information index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), fixation index (F_{is}) and gene flow (N_m) were determined by using Cervus 3.0.7 (Kalinowski *et al.*, 2007) and GenAlEx v6.5 (Peakall and Smouse, 2006, 2012). The Mantel's test was used to find a possible link between geographic distance and genetic distance. The genetic structure among the populations was elucidated by a Bayesian approach using STRUCTURE v 2.3.4 (Pritchard *et al.*, 2000). A burn-in period of 10,000 and Markov chain Monte Carlo repeats of 10,000 after the burn-in period were used for 100 iterations per K ($K = 1$–6) using the admixture model with allele frequency correlated. Structure harvester (Earl, 2012) was used to determine the optimal number of genetic pools based on the highest value of delta K. To determine the similarity/dissimilarity, the populations were clustered through principle coordinate analysis (PCoA) in MS excel and Hierarchical Clustering with hclust function in R.

Results

Assembly

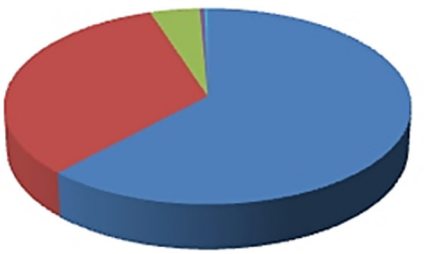
The raw reads were *de novo* assembled into 94,636 transcripts. After filtering off the sequences <500 nt, 39,743

https://doi.org/10.1017/S1479262120000167 Published online by Cambridge University Press

(a)

Total number of sequences examined	**39743**
Total size of examined sequences (bp):	**58238373**
Total number of identified SSRs:	**12884**
Number of SSR containing sequences:	**9148**
Number of sequences containing more than 1 SSR:	**2526**
Number of SSRs present in compound formation:	**1528**

(b)



(c)

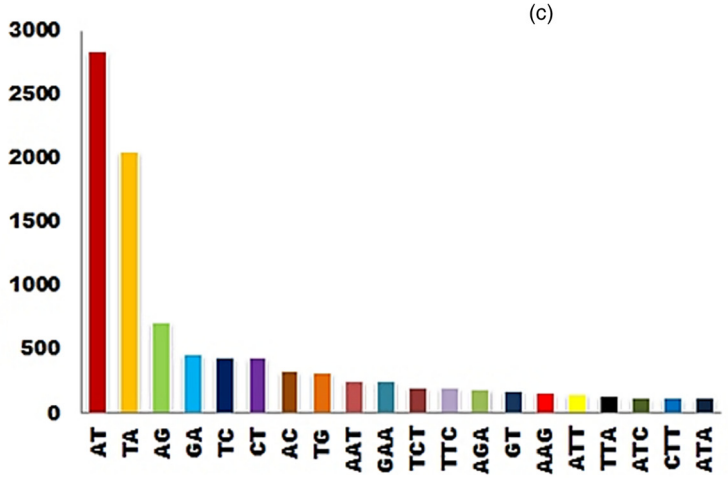


Fig. 1. (a) Statistics of SSR screening. (b) Distribution of different SSR types. (c) Top SSR repeats observed.

sequences were retained for further analysis. Read representation through Bowtie2 showed an overall alignment rate of 93.7%. Out of 1440 core genes queried, 1068 genes were detected during completeness assessment, thereby showing 74.17% assembly completeness. The mean sequence length of 1465 was observed. N_{50}, L_{50} and GC content were found to be 1820, 10,139 and 40.09% respectively. A total of 4053 sequences representing the full-length transcripts and having an alignment coverage of ⩾80% were obtained during blast against the SwissProt database. A total of 35,917 unigenes were retained for downstream analysis after the removal of sequence redundancy and clustering through CD-HIT EST.

Marker characterization and genetic diversity

SSR screening revealed 12,884 SSRs distributed among 9148 sequences. Dinucleotide repeat motifs were most abundant followed by trinucleotide repeat motifs. Figure 1 shows the distribution of different repeat types. Within the identified SSRs, AT-motifs were most abundant followed by TA- and AG-motifs. Positional distribution revealed that out of 12,884 SSRs, 4305 were present in the 3′-UTR region, 3448 in 5′-UTR region and 3238 within the CDS region. Out of 36 SSRs selected for characterization, 21 exhibited successful amplification among which twelve showed polymorphism and nine revealed monomorphism. Population genetic analysis was carried out utilizing the twelve polymorphic SSRs. The N_a ranged from 2.0 to 5.3 with a mean value of 3.6, N_e ranged from 1.34 to 3.81 with a mean value of 2.57. The PIC value ranged from 0.21 to 0.78 with a mean value of 0.57. Six SSRs showed a significant deviation from the HWE. Mean Ho and He was found to be 0.58 and 0.54, respectively (Table 1).

Population characterization

The observed heterozygosity of the individual populations of *C. procera* ranged from 0.48 (Gurdaspur population) to 0.65 (Sirsa population) with a mean value of 0.58 (Table 2). The observed heterozygosity of *C. gigantea* was found to be 0.63. Assessment of inbreeding showed that the fixation

https://doi.org/10.1017/S1479262120000167 Published online by Cambridge University Press

Table 1. Characteristics of the polymorphic markers

Sl. no.	ID	Sequence	Motif	Obs. size range	T_a	PIC	N_a	N_e	I	H_o	H_e	F	HW
1	C1	F: CCATTCTTCACTCCCTCCAA R: CACTGAGCACACACTGTTGTTG	(AG)10	148–156	48	0.698	4.333	3.352	1.266	0.705	0.663	−0.067	NS
2	C2	F: GGACAGCGGAGACAAGAGTC R: CGTGTCCTTCGGCTTTCAT	(AG)10	108–113	55	0.520	3.000	2.188	0.853	0.665	0.502	−0.319	NS
3	C4	F: CCAGAATTCCGTAAGCTGGA R: GTAGCCATGGCATGAATCTG	(GA)12	152–160	48	0.712	4.000	3.181	1.222	0.775	0.676	−0.161	NS
4	C8	F: GGCTCGCTATTCGTTCCTTT R: TTGGATCGTTGTCGTAGTCG	(AG)13	151–158	49	0.573	3.667	2.597	1.066	0.907	0.601	−0.509	***
5	C15	F: GGACCCGAACCATTACAA R: AAGACCTTTGCGTCAACAGC	(AG)7	144–150	46	0.591	3.333	2.388	0.964	0.596	0.552	−0.072	NS
6	C16	F: GGACCCGAACCATTACAA R: AAGACCTTTGCGTCAACAGC	(AG)7	144–150	46	0.634	4.000	2.810	1.127	0.433	0.613	0.327	***
7	C21	F: CAAGCACACGGTTGGAAACT R: ATGTTGCAGCTCTGGTGCTA	(GA)6	140–146	47	0.500	3.000	2.049	0.779	0.270	0.461	0.414	***
8	C22	F: CAAGCACACGGTTGGAAACT R: ATGTTGCAGCTCTGGTGCTA	(GA)6	142–148	47	0.488	3.000	1.854	0.704	0.260	0.389	0.242	*
9	C26	F: GGGTGGTGTGATTGATAC R: GGGAATTCACCGTCAACAGA	(AG)6	150–156	46	0.216	2.000	1.348	0.408	0.238	0.249	0.034	NS
10	C28	F: ATCGGGGTTCTCTCCCTCT R: CTCTGCCTCCCAAGAACAAG	(AG)15	172–182	50	0.785	5.167	3.817	1.446	0.869	0.729	−0.202	**
11	C30	F: CTCCCTTCATTCTTGAGG R: GCGCCATTTCTCTCTGTCTC	(AG)9	148–154	47	0.692	5.333	3.288	1.328	0.664	0.655	0.004	NS
12	C36	F: CCCTGCGTCTCATAATTCCA R: GGTTCACTCACTCGGCAACT	(GA)9	170–174	50	0.491	2.667	2.065	0.794	0.659	0.497	−0.331	***
				Average		0.575	3.625	2.578	0.996	0.587	0.549	−0.053	

T_a, annealing temperature; N_a, number of alleles; N_e, number of effective alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; PIC, polymorphic information content, F, fixation index; HW, Hardy–Weinberg; NS, not significant.
***Significant at 0.1% level, ** at 1% level, * at 5% level.

Table 2. Population wise diversity parameters

Population	N_a	N_e	H_o	H_e	F_{is}
Bathinda	4.333	2.780	0.585	0.573	−0.02
Barnala	3.417	2.629	0.618	0.552	−0.12
Gurdaspur	3.083	2.160	0.483	0.479	0.03
Delhi	3.750	2.618	0.595	0.571	−0.03
Ambala	3.667	2.741	0.583	0.570	0.005
Sirsa	3.500	2.542	0.657	0.548	−0.20
Gorakhpur	3.917	2.858	0.636	0.608	−0.01

index of *C. procera* populations ranged from −0.20 (Sirsa population) to 0.03 (Gurdaspur population) with a mean value of −0.05 per population. The fixation index of *C. gigantea* was found to be −0.01. The analysis of molecular variance (AMOVA) showed that 84% of the variation lies within the populations and 16% among the populations (Fig. 2(a)). The clustering of populations by PCoA and hierarchical clustering showed similar results, separating the *C. procera* and *C. gigantea* into distinct clusters (Fig. 3(a) and (b)). The Bayesian analysis displayed a genetic structure among the sampled populations by clustering them into two distinct genetic pools (Fig. 3(c) and (d)). The mean F_{st} of all the loci across all the populations was observed to be 0.147. The pairwise genetic differentiation assessment (online Supplementary file S2) revealed that the pairwise F_{st} ranged from 0.045 (between Barnala and Delhi populations) to 0.174 (between populations Bathinda and Gorakhpur). The *C. gigantea* population showed relatively higher F_{st} values with all the population of *C. procera* (online Supplementary file S2). We observed a significant positive correlation of genetic distance with geographic distance through the Mantel's test with $r = 0.147$, $P = 0.010$ (Fig. 2(b)).

Discussion

The properties of *C. procera* fibres are appropriate to use them as a new resource of biodegradable and renewable natural fibres (Tuntawiroon *et al.*, 1984; Sakthivel *et al.*, 2005; Ashori and Bahreini, 2009; Cheema *et al.*, 2010; Ga *et al.*, 2014; Qi *et al.*, 2018). Despite its importance, it is still undomesticated. Its emerging role as a new source of natural fibre may turn the attention towards its domestication and selection. The information on genetic variation in natural populations of *C. procera* is scarce, especially from the northwestern plains of India. An effective polymorphic marker resource was developed to study its genetic diversity in the northern plains of India. The polymorphic SSR marker resource could prove useful to assist in the breeding of *C. procera*. Furthermore, the assessment of genetic diversity could be beneficial in selecting the area of broad

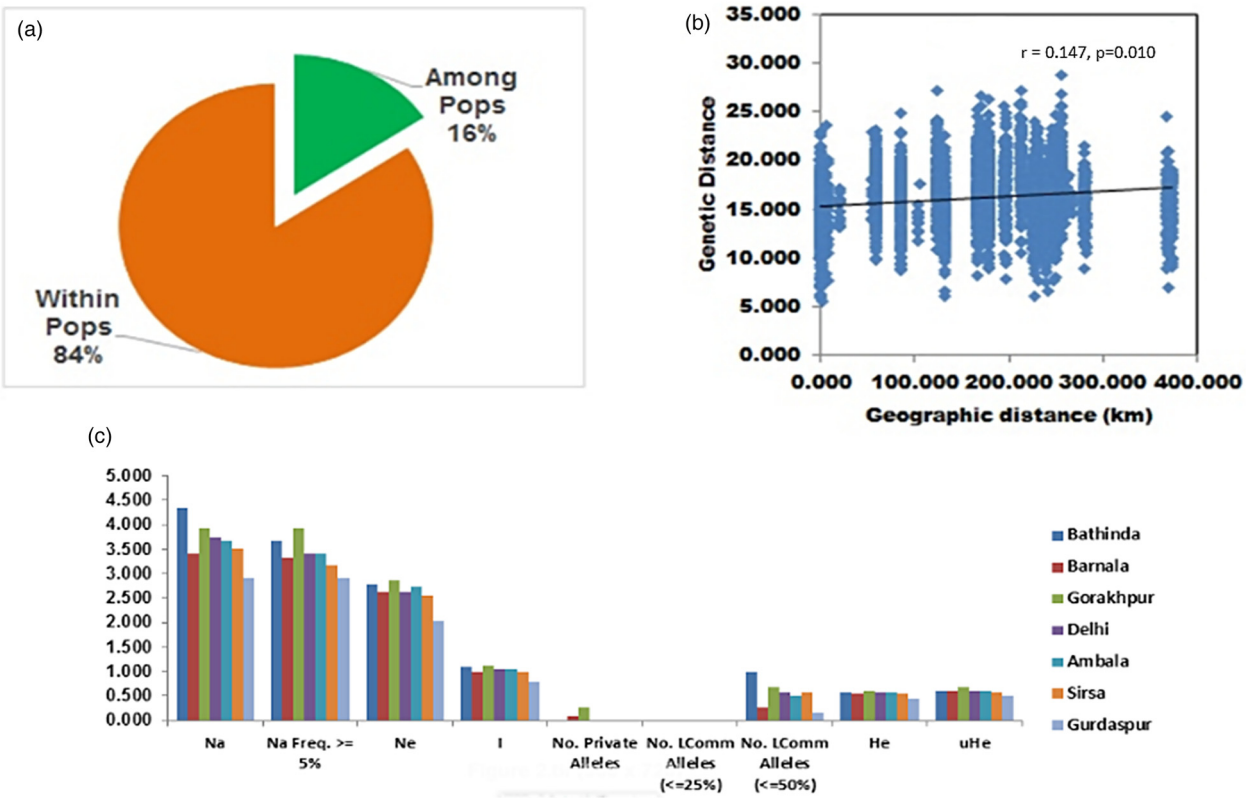


Fig. 2. (a) Representation of variation within and among populations. (b) Correlation between geographic and genetic distance. (c) Allelic pattern across the populations.

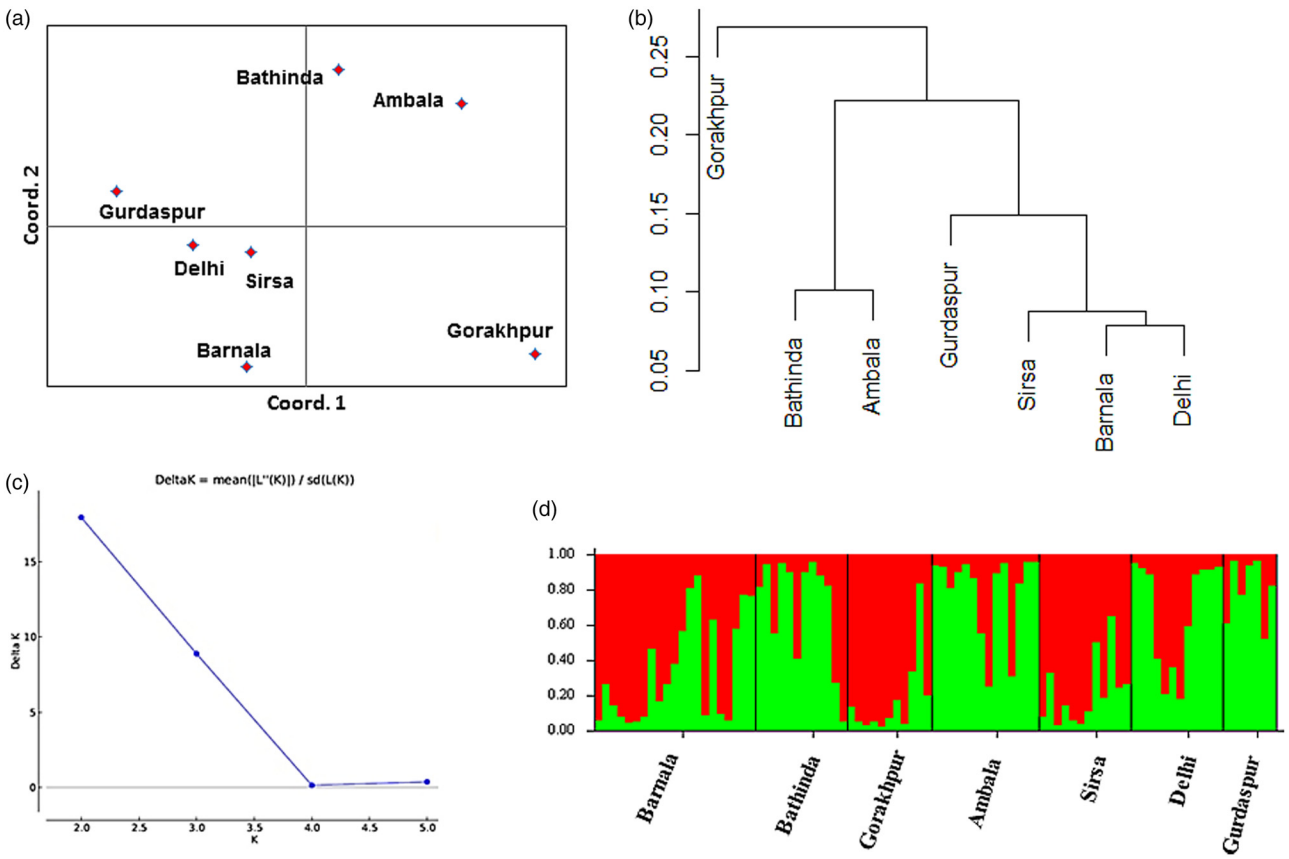


Fig. 3. (a) Clustering of populations through PCoA. (b) Dendrogram representing population clustering. (c) Depiction of highest delta *K* obtained through Structure Harvester. (d) Bar plot showing the number of overall genetic stocks.

genetic base for germplasm collection during breeding programmes.

We used the RNAseq approach to identify SSR regions. We obtained an alignment rate of 93.7% and detected 74.17% of the core genes during the assembly completeness assessment. A comparable quality of the assembly for SSR discovery was also achieved by Choudhary *et al.* (2018). This shows that the quality of our assembly was good for the downstream analysis. The N_{50} value of 1820 obtained in this study was better than that of Majeed *et al.* (2019), which also aimed at SSR screening in *Taxus contorta*. A total of 12,884 SSRs were identified from 72.139 Mb transcriptome with an average of 5.5 SSRs/Mb of the sequence. This shows a substantial density of SSRs in the *C. procera* genome. A total of 25% of the SSRs were located within the CDS region, in contrast to 75% within UTR regions. This higher UTR to CDS ratio is maintained by the selection pressure to avoid Open reading frame (ORF) change within the CDS region, which can cause frameshift mutations. A similar lower CDS to the UTR ratio was also observed by Majeed *et al.* (2019) and Qu and Liu (2013).

Thirty-six SSRs were selected for characterization among which 12 were found to be polymorphic. Around 25% of the SSRs showed monomorphism which is quite obvious in the case of genic markers. A similar higher monomorphic ratio was also observed by Olsson *et al.* (2018), Postolache *et al.* (2014) and Majeed *et al.* (2019). The mean PIC value of 0.57 revealed that the SSRs are efficient and substantially polymorphic. They showed 100% cross-species transferability with *C. gigantea*, indicating that markers have some regions linked with the conserved homologous genes (Varshney *et al.*, 2005). The success of transferability depends on the evolutionary distance, which inversely affect the transferability (Liewlaksaneeyanawin *et al.*, 2004). The 100% transferability between *C. procera* and *C. gigantea* indicates that they are evolutionary more related and share a higher degree of sequence conservation. A similar perfect transferability was also achieved between the rubber species, which decreased to only 73% in cross genera experiments (Feng *et al.*, 2009). A comparatively lesser percentage (75%) of cross transferability was observed for *T. contorta* and *Taxus baccata* (Majeed *et al.*, 2019).

Different marker systems were used to assess the genetic diversity of Calotropis. Priya *et al.* (2015) observed rich genetic diversity along with 84% genetic similarity between white and pink flowered *C. gigantea* using RAPD markers.

https://doi.org/10.1017/S1479262120000167 Published online by Cambridge University Press

A higher level of polymorphism was also observed in Egyptian accessions of *C. procera* through Random amplified polymorphic DNA (RAPD) (El-Bakry *et al.*, 2014). The isozyme analysis involving superoxide dismutase and peroxidase also showed a high diversity of three populations of *C. procera* (Pandeya *et al.*, 2007). Furthermore, SDS-PAGE of a protein and isozyme successfully discriminated the genotypes of *C. procera* populations (Hassan *et al.*, 2008). Moreover, a broad genetic base was observed in West African accessions of *C. procera* using Amplified fragment length polymorphism (AFLP) markers (Yao *et al.*, 2015). Our study showed an overall moderate diversity of *C. procera* in northern plains of India. The observed heterozygosity of the individual populations ranged from 0.48 to 0.65 with a mean value of 0.58. Thus, the populations of *C. procera* from the northern plains of India have greater genetic diversity than the African populations analysed by Muriira *et al.* (2018) as well as *C. gigantea* populations analysed from China and Bangladesh by Islam *et al.* (2019) using EST-SSR markers. In this study, the Sirsa population showed the highest diversity whereas the Gurdaspur population revealed the lowest diversity. The highest numbers of alleles were observed for the Bathinda population, whereas the lowest numbers of alleles were found in the Gurdaspur population. In contrast to our populations, the African populations of *C. procera*, as well as Chinese and Bangladeshi populations of *C. gigantea* were observed to have some degree of inbreeding (Muriira *et al.*, 2018; Islam *et al.*, 2019). Our study showed that the fixation index of the individual populations ranged from −0.20 (Sirsa population) to 0.03 (Gurdaspur population) with a mean value of −0.05 per population, indicating no inbreeding. This may explain the relatively lower diversity found by Muriira *et al.* (2018) and Islam *et al.* (2019). In our case, only two populations, Gurdaspur and Ambala were found to have a very low positive value of inbreeding coefficients, indicating negligible inbreeding. The population of *C. gigantea* showed higher genetic diversity ($H_o = 0.63$) than populations of *C. procera* and no signs of inbreeding ($F_{is} = -0.01$). The observed heterozygosity of *C. gigantea* from Gorakhpur was found to be higher than the populations analysed from Nepal, Dongchuan, Honghe and Hainan (Muriira *et al.*, 2018) and those of China and Bangladesh (Islam *et al.*, 2019). The allelic pattern across the populations showed that there is one allele (belonging to the locus C30) specific to the Barnala population of *C. procera* whereas three alleles (belonging to loci C2 and C26) were specific to Gorakhpur population of *C. gigantea*.

PCoA is often used to show the genetic similarity/dissimilarity among the populations. The results of PCoA revealed that the first three axes explained 75.87% of the variations (online Supplementary file S3). The populations of *C. procera* formed two sub-clusters with Bathinda and Ambala in one sub-cluster whereas Gurdaspur, Delhi, Sirsa, and Barnala in another sub-cluster. The *C. gigantea* population was distinct from these sub-clusters, which was also revealed in the dendrogram constructed by hierarchical clustering. This suggests that a distinct genetic structure exists between *C. gigantea* and *C. procera*. This finding is corroborated by Muriira *et al.* (2018) who also found a distinct genetic pattern between them.

The AMOVA results showed higher within-population genetic variations (84%) than among population diversity (16%). Studies on other Apocynaceae species such as *Aspidosperma polyneuron*, *Mandevilla velutina*, *Asclepias tuberosa*, *Asclepias viridis*, *Asclepias meadii* and *Asclepias incarnata* showed similar results. AMOVA showed the occurrence of genetic structure which was supported by the moderate level of genetic differentiation ($F_{st} = 0.147$). This genetic differentiation is lower than that of Islam *et al.* (2019) and Muriira *et al.* (2018) for *C. procera* and *C. gigantea*. In contrast to Islam *et al.* (2019) and Muriira *et al.* (2018), we observed relatively less impact of physical distance on genetic differentiation ($r = 0.147$, $P = 0.010$). The positive relation during Mantel's test provides useful information to select genetically diverse genotypes for future domestication and breeding programme.

The F_{st} value 0.147 across all the loci shows that there is a moderate genetic differentiation of *C. procera*. The pairwise genetic differentiation assessment (online Supplementary file S2) revealed that the pairwise F_{st} ranged from 0.045 (between Barnala and Delhi populations) to 0.174 (between Bathinda and Gorakhpur populations). The *C. gigantea* population showed higher F_{st} values with all the population of *C. procera* (online Supplementary file S3). The genetic structure of a population is determined by the complex interplay of processes such as gene flow, mutation, selection and mating strategy (Schaal *et al.*, 1998). The PCoA, STRUCTURE and hierarchical clustering showed differentiation of *C. gigantea* with populations of *C. procera* but the differentiation is not as strong as observed by Muriira *et al.* (2018). The possible reason for sharp differentiation may be due to a very large physical distance between the *C. procera* and *C. gigantea* genotypes, as the former were mostly sampled from Africa whereas the later from Asia. The Bayesian analysis displayed a genetic structure by clustering the populations into two distinct genetic pools. However, each population was found to have a proportion form each gene pool. This shows that the gene flow is not severely restricted and no intense genetic barrier separate these populations in the northern plains of India. The mean gene flow of 2.9 indicates the significant level of allele sharing and is higher than the other Apocynaceae species like *Hancornia speciose* and *Vincetoxicum atratum* (1.463), and *Gymnema sylvestre* (Jimenez *et al.*, 2015; Rathore *et al.*, 2016; Yamashiro *et al.*, 2016).

https://doi.org/10.1017/S1479262120000167 Published online by Cambridge University Press

Conclusion

We developed a set of informative SSR markers for *C. procera* which showed perfect cross-transferability with *C. gigantea*. We revealed a moderate level of genetic diversity in *C. procera* from the northern plains of India along with its genetic structure. This information will be beneficial in the initial assessment of selecting the germplasm to ensure the maximum genetic diversity can be captured during the domestication process.

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S1479262120000167.

Acknowledgements

This study was funded by SERB-DST under the grant SR/52/JCB-11/2012.

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ashori A and Bahreini Z (2009) Evaluation of *Calotropis gigantea* as a promising raw material for fiber-reinforced composite. *Journal of Composite Materials* 43: 1297–1304.

Cheema HMN, Bashir A, Khatoon A, Iqbal N, Zafar Y and Malik KA (2010) Molecular characterization and transcriptome profiling of expansin genes isolated from Calotropis procera fibers. *Electronic Journal of Biotechnology* 13: 10–11.

Choedon T, Mathan G, Arya S, Kumar VL and Kumar V (2006) Anticancer and cytotoxic properties of the latex of *Calotropis procera* in a transgenic mouse model of hepatocellular carcinoma. *World Journal of Gastroenterology* 12: 2517–2522.

Choudhary S, Thakur S, Najar RA, Majeed A, Singh A and Bhardwaj P (2018) Transcriptome characterization and screening of molecular markers in ecologically important Himalayan species (*Rhododendron arboreum*). *Genome* 61: 417–428.

Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phyto chem Bulletin Botanical Society of America* 19: 11–15.

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

El-Bakry AA, Hammad IA and Rafat FA (2014) Polymorphism in *Calotropis procera*: preliminary genetic variation in plants from different phytogeographical regions in Egypt. *Rendiconti Lincei. Scienze Fisiche e Naturali* 25: 471–477.

Feng SP, Li WG, Huang HS, Wang JY and Wu YT (2009) Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). *Molecular Breeding* 23: 85–97.

Ga DB, Kb SB and Pc NK (2014) Tensile and wear behavior of *Calotropis gigentea* fruit fiber reinforced polyester composites. *Procedia Engineering* 97: 531–535.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I and Chen Z (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* 29: 644–652.

Hassan AM, El-Shawaf IIS, Bekhi, MMM, El-Saied FM and Masoud IM (2008) Genetic variation within Ushaar (*Calotropis procera* (ait) F.) genotypes using SDS-PAGE for protein and isozyme analysis. The fourth Conference of sustainable Agriculture Development, Faculty of Agriculture, Fayoum University, 20–22 Oct: 103–114.

Huang L, Deng X, Li R, Xia Y, Bai G, Siddique KH and Guo P (2018) A fast silver staining protocol enabling simple and efficient detection of SSR markers using a non-denaturing polyacrylamide gel. *Journal of Visualized Experiments* 134: e57192.

Islam MR, Li ZZ, Gichira AW, Alam MN, Fu PC, Hu GW and Chen LY (2019) Population genetics of *Calotropis gigantea*, a medicinal and fiber resource plant, as inferred from microsatellite marker variation in two native countries. *Biochemical Genetics* 57: 522–539.

Jimenez HJ, Martins LSS, Montarroyos AVV, Silva Junior JF, Alzate-Marin AL and Moraes Filho RM (2015) Genetic diversity of the Neotropical tree *Hancornia speciosa* Gomes in natural populations in Northeastern Brazil. *Genetics & Molecular Research* 14: 17749–17757.

Kalinowski ST, Taper ML and Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099–1106.

Langmead B (2010) Current protocols in bioinformatics. In: Aligning Short Sequencing Reads with Bowtie. John Wiley & Sons, Inc., pp. 32:11.7.1-11.7.14.

Li W and Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)* 22: 1658–1659.

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA and Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. *Theoretical & Applied Genetics* 109: 361–369.

Majeed A, Singh A, Choudhary S and Bhardwaj P (2019) Transcriptome characterization and development of functional polymorphic SSR marker resource for Himalayan endangered species, *Taxus contorta* (Griff). *Industrial Crops & Products* 140: 111600.

Min XJ, Butler G, Storms R and Tsang A (2005) ORF predictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research* 33: 677–680.

Muriira NG, Muchugi A, Yu A, Xu J and Liu A (2018) Genetic diversity analysis reveals genetic differentiation and strong population structure in *Calotropis* plants. *Scientific Reports* 8: 7832.

Mutwakil MZ, Hajrah NH, Atef A, Edris S, Sabir MJ, Al-Ghamdi AK and El-Domyati FM (2017) Transcriptomic and metabolic responses of *Calotropis procera* to salt and drought stress. *BMC Plant Biology* 17: 231.

Olsson S, Pinosio S, González-Martínez SC, Abascal F, Mayol M, Grivet D and Vendramin GG (2018) De novo assembly of English yew (*Taxus baccata*) transcriptome and its applications for intra- and inter-specific analyses. *Plant Molecular Biology* 97: 337–345.

Pandeya SC, Chandra A and Pathak PS (2007) Genetic diversity in some perennial plant species with-in short distances. *Journal of Environmental Biology* 28: 83–86.

Peakall R and Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Resources* 6: 288–295.

Peakall R and Smouse PE (2012) Genalex 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics (Oxford, England)* 28: 2537–2539.

Postolache D, Leonarduzzi C, Piotti A, Spanu I, Roig A, Fady B and Vendramin GG (2014) Transcriptome versus genomic microsatellite markers: highly informative multiplexes for genotyping *Abies alba* Mill. and congeneric species. *Plant Molecular Biology Reporter* 32: 750–760.

Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Priya TA, Manimekalai V and Ravichandran P (2015) Intraspecific genetic diversity studies on *Calotropis gigantea* (L) R. Br. using RAPD markers. *European Journal of Biotechnology & Biosciences* 3: 7–9.

Qi Y, Xu F, Longdi C, Ruiyun Z, Lifang L, Wenhong F, Beina Z and Li J (2018) Evaluation on a promising natural cellulose fiber – *Calotropis gigantea* fiber. *Trends in Textile Engineering & Fashion Technology* 2: 205–211.

Qu J and Liu J (2013) A genome-wide analysis of simple sequence repeats in maize and the development of polymorphism markers from next-generation sequence data. *BMC Research Notes* 6: 403.

Rathore PK, Madihalli S, Hegde S, Hegde HV, Bhagwat RM, Gupta VS, Kholkute SD, Jha TB and Roy S (2016) Assessment of genetic diversity of *Gymnema sylvestre* (Retz.) R.Br. from Western Ghats and Eastern India. India. *Journal of Biodiversity & Environmental Sciences* 9: 82–92.

Sakthivel JC, Mukhopadhyay S and Palanisamy NK (2005) Some studies on Mudar fibers. *Journal of Industrial Textiles* 35: 63–76.

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT and Smith WA (1998) Phylogeographic studies in plants: problems and prospects. *Molecular Ecology* 7: 465–474.

Silva MCC, da Silva AB, Teixeira FM, de Sousa PCP, Rondon RMM, Júnior JERH and de Vasconcelos SMM (2010) Therapeutic and biological activities of *Calotropis procera* (Ait.) R. Br. *Asian Pacific Journal of Tropical Medicine* 3: 332–336.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV and Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England)* 31: 3210–3212.

Sobrinho MS, Tabatinga GM, Machado IC and Lopes AV (2013) Reproductive phenological pattern of *Calotropis procera* (Apocynaceae), an invasive species in Brazil: annual in native areas; continuous in invaded areas of caatinga. *Acta Botanica Brasilica* 27: 456–459.

Tuntawiroon N, Samootsakorn P and Theeraraj G (1984) The environmental implications of the use of *Calotropis gigantea* as a textile fabric. *Agriculture, Ecosystems & Environment* 11: 203–212.

Varshney RK, Graner A and Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* 23: 48–55.

Yamashiro T, Yamashiro A, Inoue M and Maki M (2016) Genetic diversity and divergence in populations of the threatened grassland perennial *Vincetoxicum atratum* (Apocynaceae-Asclepiadoideae) in Japan. *Journal of Heredity* 107: 455–462.

Yao DARA, Sprycha Y, Porembski S and Horn R (2015) AFLP assessment of the genetic diversity of *Calotropis procera* (Apocynaceae) in the West Africa region (Benin). *Genetic Resources & Crop Evolution* 62: 863–878.

You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hane D, Lazo GR, Dvorak J and Anderson OD (2008) Batchprimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 253: 1–13.