DEVELOPMENT AND CHARACTERIZATION OF SSR MARKERS IN *Calotropis procera*

Project report submitted to Central University of Punjab

For the award of

Master of Science

in

Life Sciences (Specialization in Plant Sciences)

By

Vandana

Supervisor

Dr. Pankaj Bhardwaj

Department of Plant Sciences

School of Basic and Applied Sciences

Central University of Punjab, Bathinda

May, 2018

CERTIFICATE

I declare that dissertation entitled "DEVELOPMENT AND CHARACTERIZATION OF SSR MARKERS IN *Calotropis procera*" has been prepared by me under the guidance of Dr. Pankaj Bhardwaj, Assistant Professor, Department of Plant Sciences, School of Basic and Applied Sciences, Central University of Punjab, Bathinda. No part of this Project Report has formed the basis for the award of any degree or fellowship previously.

Vandana

Department of Plant Sciences,
School of Basic and Applied Sciences,
Central University of Punjab,
Bathinda-151001

Date:

CERTIFICATE

I certify that Vandana has prepared her Project Report entitled "DEVELOPMENT AND CHARACTERIZATION OF SSR MARKERS IN *Calotropis procera"* for the award of Master in Sciences (Specialization in Plant Sciences) from Central University of Punjab, under my guidance. She has carried out this work at the Department of Plant Sciences, School of Basic and Applied Sciences, Central University of Punjab.

Dr. Pankaj Bhardwaj

Department of Plant Sciences,
School of Basic and Applied Sciences,
Central University of Punjab,
Bathinda- 151001

Date:

ABSTRACT

Development and Characterization of SSR markers in *Calotropis procera*

Name of the student: Vandana

Registration number: 16mslsps01

Degree for which submitted: Master of Sciences (Specialization in Plant Sciences)

Name of supervisor: Dr. Pankaj Bhardwaj

Name of the Department: Department of Plant Sciences

Name of School: School of Basic and Applied Sciences

Keywords: Population structure, SSR markers, Calotropis procera, gene flow, cotton

Calotropis procera is a weed growing around the arid and semi-arid areas of the tropical and subtropical regions. The species has the potential to replace cotton and resolve the high rates of cotton-based products and at the same time reduce the use of synthetic non-biodegradable fibres. It has been used in various medicinal formulations from a long time. There is a need for developing molecular markers to facilitate breeding and genetic improvement of its varieties. Out of 12.6 million raw reads, 84636 reads were obtained after quality check and filtering. 93.7% of alignment rate and 76.31% of core genes were detected in Bowtie and BUSCO assessment, respectively. An average of 5.5 SSRs/Mb was obtained in 72.349 Mb transcriptome. 25 SSR markers were selected for characterization on 30 individuals representing 3 populations. 13 SSR loci were found polymorphic. In the population structure analysis a moderate level of genetic diversity (Na=3.9231, He=0.633) was predicted. The genetic variation among the population was found to be only 7% whereas within the population, it was 93%. The mean Shannon's information index obtained was 1.1, showing significant diversity richness in the populations. The Dendrogram showed that the population from Bathinda and Ambala are closely related to each other than those from Barnala. The gene flow between Bathinda and Ambala population was found significantly high (6.439). The present set of loci had a high PIC value of 0.506 that makes them very informative to be further used for large scale studies.

Signature of Student Signature of Supervisor

ACKNOWLEDGEMENTS

Foremost, I thank Almighty for the blessings, strength and fortitude.

I express my sincere gratitude to my supervisor, **Dr. Pankaj Bhardwaj**, Assistant Professor, Department of Plant Sciences, Central University of Punjab for his guidance, continuous support, motivation, supervision and constructive criticism throughout the course of study and preparation of the manuscript.

I extent my appreciations to **Dr. Sanjeev Thakur**, Head of Department (Department of Plant Sciences) and **Prof R. K. Kohli**, Vice Chancellor, Central University of Punjab for providing infrastructure and other technical facilities. I am grateful for the cooperation of all the faculty members of Department of Plant Sciences.

I sincerely thank my seniors specially Assim Majeed, Shruti Choudhary, Sapna Thakur and Amandeep singh for their continuous support and guidance. I deeply thank Assim Majeed for his kind endless help, generous advice and encouragement during the study. I thank my lab mate Diksha Bisht for her cooperation. I thank my friend Shweta Selpair for her emotional support and motivation throughout the study. I am thankful to technical staff for their help and support.

My warmest regards to my family for their love and blessings.

Vandana

TABLE OF CONTANTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

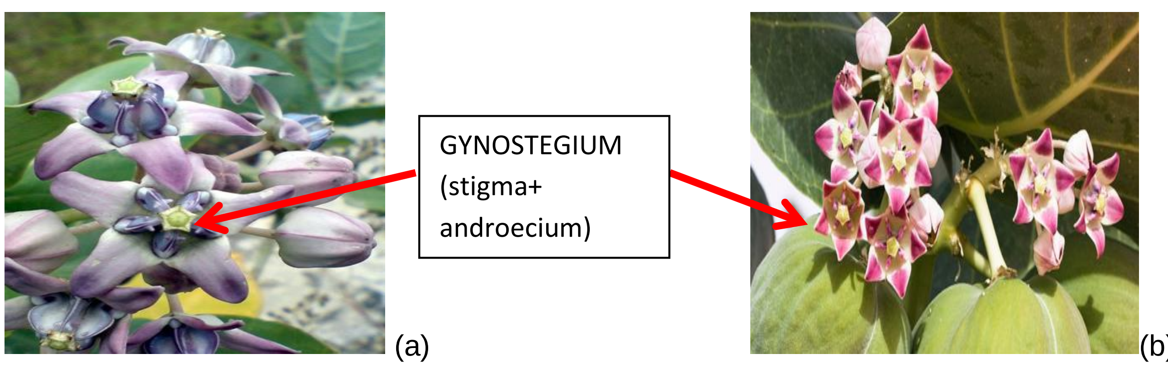
LIST OF ABBREVIATIONS

S.No.	Full form	Abbreviation
1	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide	MTT
2	Adenine	A
3	Amplified Fragment Length Polymorphism	AFLP
4	Bacterial Artificial Chromosome	BAC
5	Benchmarking Universal Single-Copy Orthologs	BUSCO
6	Centimetre	cm
7	Cetyl Trimethyl Ammonium Bromide	CTAB
8	Coding Sequences	CDS
9	Complementary Deoxyribo Nucleic Acid	cDNA
10	Cytosine	C
11	Degree Celsius	°C
12	Deoxyribo Nucleic Acid	DNA
13	Expressed Sequence Tag	EST
14	Glutathione	GSH
15	Gram	g
16	Guanine	G
17	Hardy- Weinberg	HW
18	Kilometre	Kg
19	Kyoto Encyclopedia of Gene and Genome	KEGG
20	Marker assistant selection	MAS
21	Microgram	µg
22	Milligram	mg
23	Millimetre	mm
24	Minute	Min
25	National Center for Biotechnology Information	NCBI
26	Nucleotide	nt
27	Polyacrylamide Gel Electrophoresis	PAGE
28	Polymerase Chain Reaction	PCR
29	Polymorphic Information Content	PIC
30	Quantitative Trait Locus	QTL
31	Random Amplification of Polymorphic DNA	RAPD
32	Restriction Fragment Length Polymorphism	RFLP
33	Sequence Read Archive	SRA
34	Simple sequence repeat	SSR
35	Single Nucleotide Polymorphism	SNP
36	Thymine	T
37	Untranstated Region	UTR
38	Unweight Pair Group Method with Arithmetic Mean	UPGMA

Chapter 1
Introduction

Calotropis procera ((Aiton) W. T. Aiton), Milk weed is a perennial shrub of Asclepiadaceae family. It is also known as Sodom Apple. Vernacular names of milk weed are Madar (Hindi), crown flower (English), Akanda (Bengali), Adityapuspiker (Sanskrit) Vellerukku (Tamil) and Aak (Punjabi) (Sharma et al., 2000). *Calotropis procera* grows in mainly semi-arid areas of sub-tropical and tropical regions. It grows mostly on sandy soils and has drought and salt stress tolerance. It is a weed growing in infertile overgrazed soils. It is highly spreading shrub reaching 2 to 6 meters in height. It has simple, opposite, obovate, sessile leaves. Its leaves are greyish green in colour and the fruit is green, fleshy and inflated up to 11 cm in diameter. The stem is hollow filled with latex which has similar constituent like foxgloves i.e. cardiac aglycones. The latex is toxic for both humans and other animals (Yao et al., 2015).

Genus Calotropis has only two species i.e. *Calotropis procera* and *Calotropis gigantea*. These species are morphologically very similar except for their flower in which the structures of their gynostegiums are different. It is the only identifying feature to differentiate them morphologically. Figure 1 shows the different structures of gynostegiums.



https://en.wikipedia.org/wiki/Calotropis_gigantea

https://www.smartseedstore.com/products/calotropis-procera-crown-flower

Figure 1: Morphological difference between (a) *C. gigantea* and (b) *C. procera*

Calotropis is used in a number of ayurvedic formulations to cure various ailments like leprosy, fever; diarrhoea, menorrhagia, malaria, rheumatism, snake bite etc. (Parrotta, 2001). In addition to this, the seed coat of *Calotropis* produces silk like shiny fibre also known as the crystal cotton. The tensile strength and the water retention capacity is more than that of the cotton. Because of the hollow structure of these fibres, *Calotropis* has a higher insulating property that wool (Bajwa et al., 2013).

C. procera is a potential candidate to replace cotton (Tuntawiroon et al., 1984) and is undomesticated. In view of its economic and medicinal potential, the domestication of this plant is eye-catching. For domestication, identification and breeding between desired genotypes is crucial to develop a variety of appropriate traits. Marker assisted selection has been fruitful to identify the desired genotypes (Collard et al., 2008). A reliable marker resource is required for MAS. So, the development of a reliable marker resource will prove useful to decipher the genetic make-up of *C. procera*. The developed SSR markers can be used in breeding programme of this species. Further, elucidation of its population genetic parameters will throw light on its evolution. So the study is undertaken to develop and characterize the SSR markers for this species. The significance of the study lies in providing polymorphic SSRs to elucidate the genetic makeup of this species.

Chapter 2

Review of literature

1. Calotropis procera

Calotropis procera is an important perennial shrub found in the tropical and subtropical regions. It is well adapted to high temperature, low fertility and drought stress. Apart from that it has a number of medicinal properties, the fibre from the stem is used to make ropes and bags and the latex is used as coagulating agent in West Africa for cheese making (Iqbal et al., 2005).

1.1 Distribution, Habit and Habitat

The origin of Calotropis is the Afro-Asian monsoon regions. It is believed that Calotropis spread from the western Africa through middle-east and Arabian Peninsula to the Indian subcontinent. It was also introduced in the subtropical regions of America, drier parts of Australia and South-East Asia. Calotropis is found in semi-arid condition at around 1300m above the sea level with an annual rainfall of 150 to 1000mm. It establishes itself even in excessively drained soils with little competition along roadsides, edges of lagoons, overgrazed pastures, rangelands etc. and withstands high salinity stress (Orwa et al., 2009). It also develops suckers in response to any damage to the roots (Parsons et al., 2001).

1.2 Phytochemistry

Leaves of *C. procera* mainly contain large amounts of calotropagenin and calotropin. In addition to these, leaves also have calotoxin, calactin, α-amyrins, β-amyrins, taraxasteryl acetate, β-sitosterol, organic carbonates and stigmasterol (Orea et al., 2002). The latex consists of 80-93% of water and contains cardinolides like uscharin, uscharidin, calotoxin, calotropin, voruscharin etc. (Seiber et al., 1982). The chief constituents of roots are proceranol and procerursenyl acetate. It also accumulates glyceryl mono-oleolyl-2-phosphate, glceryl-1, 2-dieapriate-3-phosphate, n-Dotriacont-6-ene, methyl myrisate and methyl behenate (Alam and Ali, 2009).

1.3 Uses

C. procera is used in homemade medicinal preparations from ancient times. The various scientific studies pertaining to its uses are enlisted in Table 1.

Table 1: List of various scientific studies done on *Calotropis procera* proving its usefulness.

S.No.	Activity analysed	Plant part used	Result	References
1	Analgesic activity	Dry latex	Oral dose of 170 to 180 mg/kg produced analgesic effect against acetic acid induced writhing. Oral dose of 830mg/kg didn't show any toxic effect on mice.	(Kumar et al., 1994)
2	Hepatoprotective activity .	Flower extract	Treated the altered levels of biochemical markers such as bilirubin, cholesterol, HDL, GSH etc. against Paracetamol-induced hepatitis in rats.	(Shethy et al., 2007)
3	Anti-diarrhoeal activity	Dry latex	500mg/kg doses of dry latex of C. procera decreased the rate of defecation in 80% of the rats treated with castor oil to increase fluid accumulation in intestine and also increase the electrolyte concentration in the intestinal fluid.	(Kumar et al., 2001)
4	Anti-diabetic activity	Dry latex	Daily oral intake of 400mg/kg of the dry latex lowered the blood glucose level and increased the hepatic glycogen content. Anti-diabetic activity was comparable to the standard anti-diabetic drug, glibenclamide.	(Roy et al., 2005)
5	Anti-cancerous activity	Leaf extract	Cytotoxicity was examined using MTT essay on Hep2 cell lines by treating with methanol extract at	(Prabha and Vasanta, 2011)

			500µg/ml that showed 100% results.	
6	Antimicrobial activity	Leaf and latex	Ethanolic and chloroform extract of leaf and latex was used to determine antimicrobial activity using agar well diffusion and paper disc methods. The ethanolic extract showed the best inhibition followed by the chloroform extract. The aqueous extract didn't show any inhibition.	(Kareem et al., 2008)
7	Antifertility activity	Root extract	At 250mg/kg treatment with root extract of *C. procera* 100% anti-implantation and uterotropic activity was observed in albino rats.	(Kamath and Rana, 2002)
8	Antioxidant activity	Leaf extract	Using 1,1-Diphenyl-2-picryl hydrazyl radicals the methanoic extract of the leaf was tested for antioxidant activity. It was found that the methanoic extract has a greater capacity to scavenge DPPH radicals proving its bioactive constituents can be used for the production of new antioxidants.	(Murti et al., 2011)
9	Bast fibre analysis	Stem	Study reveals that the bast fibres blended with the cotton fibres in 1:1 ratio are inferior to the cotton but can be improved by improving the evenness and fitness.	(Varshney and Bhoi, 1988)

2. Molecular markers

The changing environmental conditions, reduction in the natural resources and increase in the biotic and abiotic stresses pose a challenge to the plant growth. It becomes a matter of concern to improve the crop varieties so that they can adapt to the diverse agro-ecologies. With the advancement of agricultural technologies, plant breeding has shown remarkable progress. Emergence of DNA marker technology led to the development of different molecular markers that are now utilized to track down loci and genomic regions linked to various agronomic and disease resistant traits in plant species (Varshney et al., 2007). Molecular markers can be classified into two types: (a) Markers based on hybridization such as RFLP (restriction fragment length polymorphism), (b) PCR based Markers: RAPD (Random Amplification of Polymorphic DNA), AFLP (Amplified Fragment Length Polymorphism) and SSR (Simple Sequence Repeat or Microsatellite, and (c) Sequence-based markers: SNP (single nucleotide polymorphism). The RFLP markers are known as the first generation markers as they were used in 1980s and 1990s for the first time for plant genetic studies. Due to the changes/ mutation in nucleotide sequences present in the restriction sites, different sized fragments are generated that helps detecting polymorphism. The main properties of RFLP markers are high reproducibility, co-dominance, high locus specificity and there is no need of prior sequence information (Lateef et al., 2015). In several crops such as rice maize and wheat, RFLP markers are used to establish genetic maps (Nagaoka et al., 1997; Cho et al., 1998). RAPD markers are based on PCR amplification of DNA segments randomly using primers with random nucleotide sequences. These markers are easy to use and inexpensive. RAPD markers have been extensively used for assessment of genetic variations, study of phylogenetic relationships among species and fingerprinting (Gupta et al., 1999). AFLP markers are PCR based and DNA restriction fragments are selectively amplified through PCR. It is a multi-loci technique that involves fragment restriction digestion as well as selective PCR amplification which further can be used for DNA of any origin (Lateef et al., 2015).

SSR (Simple Sequence repeats) are short tandemly repeated sequences (1-6 nucleotides) that are found randomly throughout the genome in most taxa. It may have 10-40 bp repeat motifs; higher repeat motifs are also possible. SSRs tend to have very high mutation rate (estimated 10^{-2} to 10^{-4} per generation) due to what they tend to have high genetic diversity. They are among most polymorphic molecular markers. The variations in SSRs arise due to the replication slippage and unequal crossing-over during meiosis. The SSRs with longer repeat motifs are more polymorphic than the shorter ones (Park et al., 2009; White et al., 2007).

SSR markers are highly reproducible as it can produce same profile regardless of the state of template DNA. These are hyper-variable and produce high allelic variation even in closely related varieties that can vary from 1-37. Co-dominance of SSR makers is appropriate for genetic analysis in segregating F2 populations and parentage analysis in hybrids (Scott et al., 2000). SSR markers are very abundant and well distributed among genomes of various species (Wang et al., 1994).Microsatellite markers are present in both the coding and non-coding regions of the genome (Miah et al., 2013). The SSR markers corresponding to the transcribed region are known as the genic markers that have significant application in gene function characterization, QTL analysis and association analysis for gene tagging (Breseghello and Sorrells, 2006).

The development of molecular markers like SSRs can overcome the limitations of morphological markers. SSR markers are ideal for the mapping experiments to track loci and genomic regions linked to desirable traits. (Miah et al., 2013). The main factors that affect MAS are analysed sample numbers, the number of lines that can be improved in provided time limit and requirement of QTL identification every time additional germplasm is used. Since SSRs target highly variable region of the genome, they can detect high level of polymorphism, independent of the species considered. Genomic manipulation has become easier because of the microsatellite markers since their use allows earlier sampling as they require small amount of tissue, faster preparation of DNA (small amount of template is required), and efficient handling of large samples (Ribaut and Hoisington, 1998). Integrating the conventional breeding strategies and MAS can develop an overall strategy to develop a better variety.

Strategies for developing SSR markers

A new set of polymorphic microsatellite markers can be generated by isolation or identification of microsatellite repeats with sufficient flanking sequence data for primer designing. The polymorphic potential of the SSR markers can be estimated in optimized PCR conditions and screening in a set of related and non-related individuals. Development of SSR markers can be achieved by various methods viz. construction of enriched or non-enriched genomic libraries, using the product of other molecular marker, by sequencing PCR products of universal or consensus chloroplast primers or by using EST sequences from public databases, by checking cross-species transferability of SSR markers of some other species (Senan et al., 2014).

SSR markers from SSR-enriched genome libraries can be developed by two ways i.e. by selective hybridization or by primer extension methods. In selective hybridization repeat specific probes are used to isolate SSR containing portions of DNA. After enrichment, the DNA fragments are amplified, can be cloned or sequenced and can also be directly sequenced to search SSR motifs. In primer extension method repeat specific primers are used to selectively amplify the SSR containing DNA. This method depends on primary genomic library constructed in phagemid vector to get a single stranded DNA library. Streptavidin coated magnetic beads are used to selectively extract the primer extended products. A second round of primer extension is done to convert the ssDNA into dsDNA. SSR markers can also be developed by using EST sequences in various public databases. With the help of SSR mining tools like MISA, TROLL, QDD, Msatcommander etc. searching microsatellite repeats in EST database has become fast(Senan et al., 2014).

Studies exploiting Microsatellite markers

SSR markers have evolved to become versatile and popular molecular marker because of the abundance in the plant community. Because of the high mutation rate microsatellite markers are very informative and because of what they are being exploited in various scientific studies. Table 2 is enlisting various studies involving SSR marker development and characterization.

Table 2: List of various scientific studies done using microsatellite markers.

s. no.	Species	Purpose	Approach	Result	Reference
1	*Hordeum vulgare* L. (Barley)	Development and characterization of SSR markers	EST derived SSR markers	2,019 SSR markers were identified in 1,856 ESTs out of which primers were designed for 1453 sequences. These SSR markers were less polymorphic than those derived from genomic regions.	(Thiel et al., 2003)
2	*Ipomoea batatas* (Sweet potato)	Denovo assembly and characterization of root transcriptome.	GAII platform of Illumina sequencing was performed for the development of SSRs	59.2 M raw reads were generated out of which 72.4% were more than 100bp in length. 4,114 potential cSSRs were identified and a total of 92 working primers were screened. 31.14% of unigenes were mapped in 124 KEGG pathways.	(Wang et al., 2010)
3	*Oryza sativa* L. (Rice)	Development and Mapping of SSR markers	SSR containing sequences retrieved from GenBank and IRGSP	2408 primers were designed from the Monsento Rce Genome database and 199 SSRs were developed fron IRGSP. 1825 newly	(McCouch et al., 2002)

				designed markers were mapped along the rice chromosomes.	
4	*Oryza sativa* L. (Rice)	Experimental and computational analysis if SSRs	SSRs were developed using BAC-end sequences and partially sequences cDNAs	200 class-I SSRs were developed and added to the existing SSR map of rice that makes a total of 500 where the map position and genetic diversity was evaluated.	(Temnykh et al., 2001)
5	*Sorghum bicolor*	Characterization and mapping of SSRs	Enriched genomic DNA library and hybridization screening for SSR	13 SSR markers were developed and amplified with 9 sorghum inbreed lines. 70% of the primers showed the presence of at least five alleles. Out of the 13 SSRs, 9 were mapped with the help of RFLP map.	(Taramino et al., 1997)
6	*Medicago truncatula*	Development characterization and transferability among legume and non-legumes	EST derived microsatellite markers	14637 SSRs were identified in 11750 sequences containing SSRs. Of the total SSRs 82% were mononucleotide repeats. High level of transferability was shown ranging 53-71% in legumes and 33% to 45% in non-legumes.	(Gupta and Prasad, 2009)
7	*Arachis hypogaea*	Isolation, characterization and diversity assessment of SSRs	Enriched microsatellite construction through primer hybridisation and sequence over ABI 3700 sequencer.	490 SSRs generated from 720 SSR positive clones. Primers were designed for 170 markers and 46 markers showed polymorphism providing 112 alleles.	(Cuc et al., 2008))

Chapter 3

Materials and Method

SSR Marker development

Paired end raw reads were retrieved from NCBI SRA Dtabase (accession: ERX2103344) (Mutwakil et al., 2017). Quality check and trimming was done using FastQC (Andrews, 2010) and Trim galore (https://github.com/FelixKrueger/TrimGalore). Denovo assembly was done using Trinity (Grabherr et al., 2011). The quality of the assembly was assessed by analysing read representation wherein raw reads were mapped back to the assembly by using Bowtie 2 (Langmead, 2010). Further, completeness of the assembly was done by using BUSCO v2 (Simão et al., 2015). Screening of the transcripts for SSR regions was done by using MISA (MicroSAtellite identification tool) (http://pgrc.ipk-gatersleben.de/misa/). Positional distribution of the identified SSRs was done by ORF predictor version 2.3 (Min et al., 2005). The identified SSRs and the ORFs were corelated to find the position of the SSRs relative to the coding regions using in-house Python script. The SSRs present at the 5'UTR region were selected for primer designing. Using Batch Primer 3 (You et al., 2008) primers were designed for the selected SSRs.

Primer characterization

10 Samples each from 3 populations (Ambala, Barnala and Bathinda) were collected. Herbariums were made for *C. procera.* DNA was isolated using CTAB method (Doyle, 1987) with some modifications. Quantification and quality check was done on nanodrop spectrophotometer and agarose gel electrophoresis(0.8% agarose gel) respectively. The primers were then charecterized by carrying out the PCR reactions on 30 selected genotypes. The PCR product was analyzed through UREA-PAGE followed by silver staining. The bands were scored and fed to excel.

Table 3: Geographical locations of *C. procera* populations.

Samples	Pop 1, Ambala		Pop 2, Barnala		Pop 3, Bathinda	
	Latitude	Longitude	Latitude	Longitude	Latitude	Longitude
1	30°18'48.72''	76°49'34.92"	30°22'39.8"	75°32'12.29"	30°10.513'	74°57.968'
2	30°18'59.80''	76°49'32.21"	30°22'41.30"	75°32'8.60"	30°10.511'	74°57.945'
3	30°18'52.70''	76°49'27.14"	30°22'42.08"	75°32'9.27"	30°10.493'	74°57.939'
4	30°18'53.27''	76°49'25.66"	30°22'40.05"	75°32'0.25"	30°10.454'	74°57.933'
5	30°18'53.99''	76°49'24.45"	30°22'40.55"	75°31'58.99"	30°10.469'	74°57.977'
6	30°18'52.32''	76°49'25.31"	30°22'40.12"	75°31'57.55"	30°10.696'	74°58.018'
7	30°18'50.51''	76°49'24.64"	30°22'39.8"	75°31'52.85"	30°10.689'	74°58.017'
8	30°18'51.06''	76°49'21.74"	30°22'41.61"	75°31'51.56"	30°10.671'	74°58.017'
9	30°18'48.96''	76°49'21.71"	30°22'42.83"	75°31'49.06"	30°10.641'	74°58.021'
10	30°18'46.50''	76°49'26.23"	30°22'43.4"	75°31'51.29"	30°10.683'	74°58.018'

Data analysis

Using the primer profiles, parameters of genetic diversity were estimated using POPGENE version 1.31 (Yeh, 1999) i.e. number of alleles (Na), effective number of alleles (Ne), expected and observed Heterozygosity (He, Ho), Shannons information index (I) and Nei’s genetic distance (Nei, 1978) and UPGMA dendrogram was contructed. F-statistic analysis, and genetic distance between inividuals was done using GenAlEx 6 (Peakall and Smouse, 2006). PIC values for each locus was calculated using Cervus 3.0 software (Kalinowski et al., 2007).

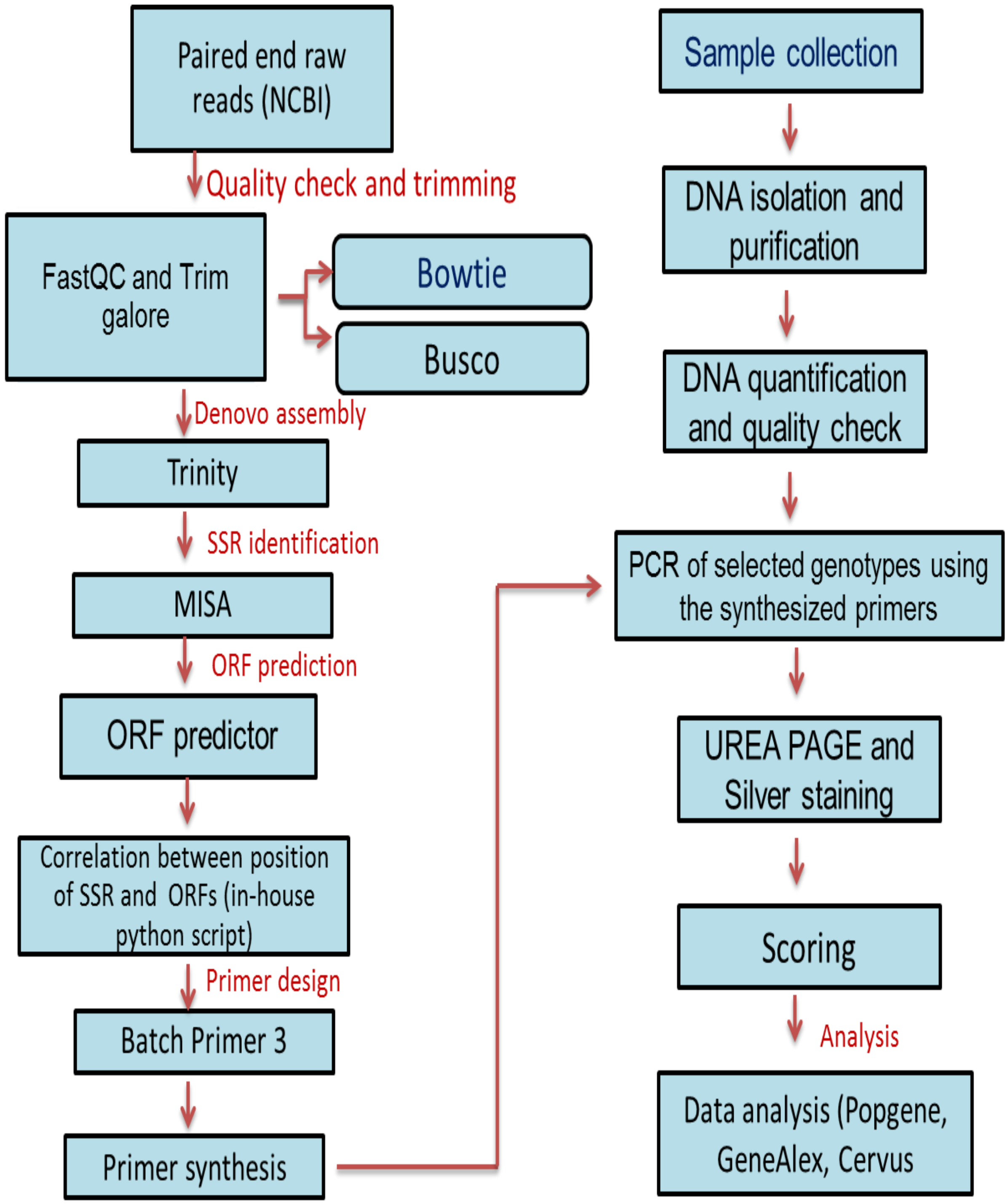


Figure 2: Schematic representation of SSR Marker development and Primer Characterization

Chapter 4

Results

Marker Development

From the total of 12.6 M raw reads, 94636 *denovo* transcripts were obtained. The sequences with less than 500bp were filtered out leaving 39743 sequences for further analysis. Read alignment through Bowtie2 showed an overall alignment rate of 93.7%. The completeness assessment (figure 3) showed that out of 1440 core genes queried, 1099 (76.31%) genes were detected with 62.57% of complete genes and 13.75% of partial genes. An average of 1.5 orthologs was found per core gene and 37.18% of gene had more than one orthologs.

Table 4: Length statistics and base composition of the assembled transcripts

Number of sequences	84636
Total length (nt)	72139129
Longest sequence (nt)	17643
Shortest sequence (nt)	201
Mean sequence length (nt)	852
Median sequence length (nt)	460
N50 sequence length (nt)	1467
L50 sequence count	14390
Number of sequences > 1K	22723 (26.8%)
Number of sequences > 10K	46 (0.1%)
Number of sequences > 100K	0 (0.0%)
Number of sequences > 1M	0 (0.0%)
Number of sequences > 10M	0 (0.0%)
GC-content (%)	39.32
Number of sequences containing non-ACGTN (nt)	0

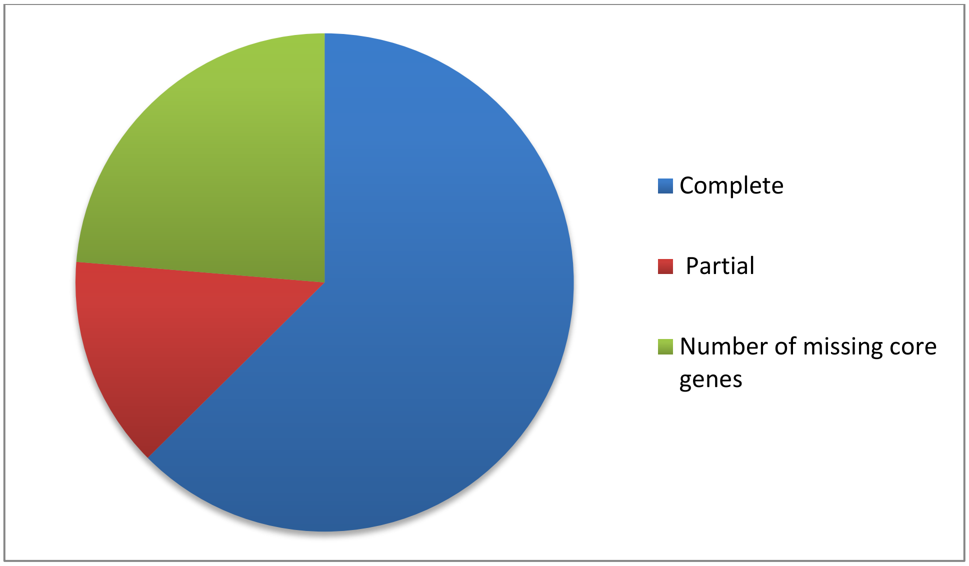


Figure 3: Completeness assessment result of assembled transcripts.

SSR screening identified 12884 SSRs distributed among 9148 sequences. Dinucleotide repeat motifs were most abundant followed by trinucleotide repeat motifs. Figure 4 shows the distribution of different repeat types. Among the identified SSRs, AT motifs are most abundant followed by TA and AG. Positional distributional analysis revealed that out of 12884 SSR markers, 4305 were present in the 3'UTR region, 3448 were present in 5'UTR region and 3238 were present in CDS region.

Table 5: General statistics of SSR screening by MIcroSAtellite identification tool

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**

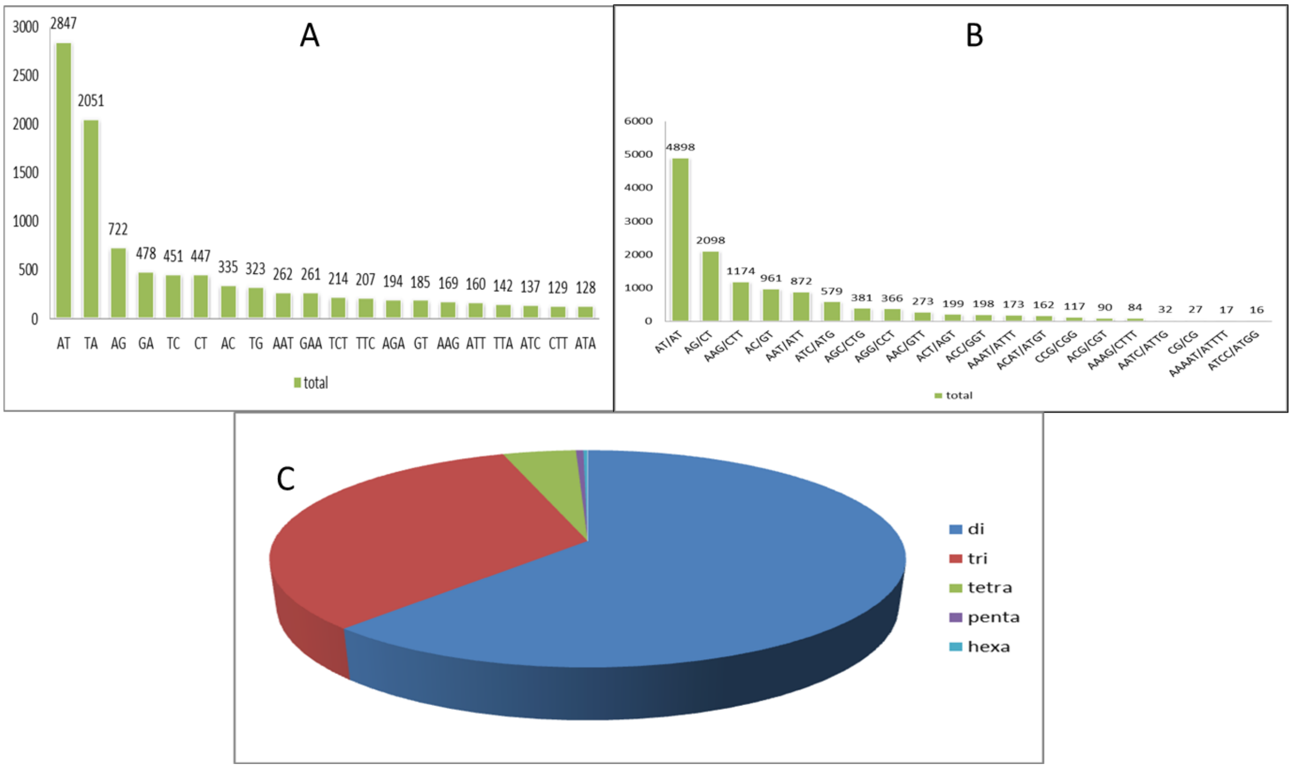


Figure 4: (A) Frequency of identified SSR motifs (B) Frequency of classified repeat types (considering sequence complementary) (C) Distribution of different repeat motifs.

Among the 25 selected markers for characterizing 30 genotypes, 13 markers showed successful amplification. 10 loci showed significant deviations from HW equilibrium. Observed heterozygosity was found to be 0.7897, effective number alleles ranged from 1.665 to 4.6632 with a mean value of 2.897 and mean Shannon's information index was 1.1. Certain alleles were found specific to particular populations. Figure 5 represents the graphical way of allelic pattern across populations. The polymorphic Information Content (PIC) value of the markers ranged from 0.356 to 0.751 with an average value of 0.506. Mean Fst value was found to be 0.0335. Shannon's diversity partition statistics showed that 93% of diversity lies within populations and 7% among populations.

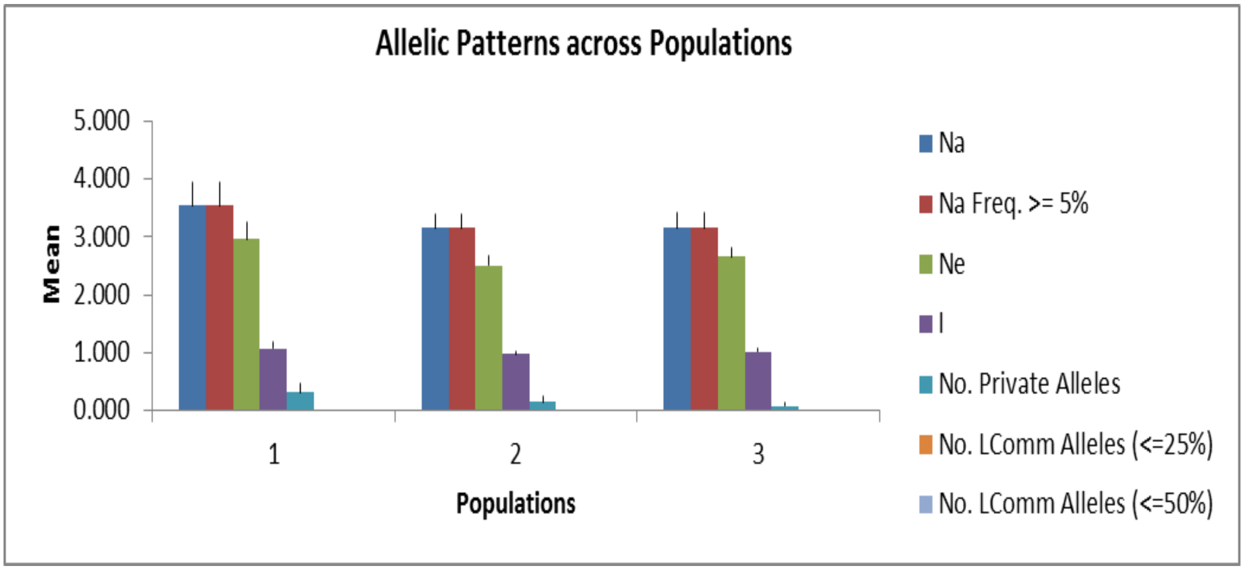


Figure 5: Allelic patterns across the three populations.

Table 6: Summary of F-Statistics

Locus	Fis	Fit	Fst	Nm
CP2	-0.3333	-0.2517	0.0612	3.8352
CP3	-1.0000	-1.0000	0.0000	*****
CP7	-1.0000	-1.0000	0.0000	*****
CP8	0.3237	0.3514	0.0409	5.8585
CP14	-1.0000	-1.0000	0.0000	*****
CP21	0.0741	0.1513	0.0835	2.7458
CP22	-0.1823	-0.1120	0.0595	3.9545
CP24	-0.1825	-0.1765	0.0051	48.6250
CP26	-0.1558	-0.0858	0.0606	3.8766
CP30	-0.1671	-0.1196	0.0407	5.8906
CP31	-0.1712	-0.1712	0.0360	6.6875
CP32	-0.5385	-0.5385	0.0000	*****
CP33	-0.2500	-0.2500	0.0156	15.7500
Mean	-0.3112	-0.2672	0.0335	7.2055

Nm = Gene flow estimated from $Nm = 0.25(1 - Fst)/Fst$

Table 7: Characteristics and evaluation details of 13 microsatellite markers

S.No.	Locus	Primer sequence	Repeat motif	Ta	Ne	Na	size range	Ho	He	I	PIC
1	CP2	F5’GGACAGCGGAGACAAGAGTC R5’CGTGTCCTTCGGCTTTCAT	(AG)10	48.48	1.6651	3	108-113	0.5938	0.4062	0.7073	0.356
2	CP3**	F5’GACGCCAAATTCTACGAGGA R5’AACTCTAACGTGGCGGAGAA	(AG)6	46.78	2.0000	2	146-149	1.0000	0.5085	0.6931	0.375
3	CP7**	F5’GGACATCGGGAAGAACAAAC R5’CCAGCAGCCATATCTCCTACA	(AG)6	48.07	2.0000	2	198-201	1.0000	0.5085	0.6931	0.375
4	CP8	F5’GGCTCGCTATTCGTTCCTTT R5’TTGGATCGTTGTCGTAGTCG	(AG)13	46.78	3.5644	5	151-158	0.4667	0.7316	1.6375	0.670
5	CP14**	F5’CTGGTGTCCTTTTCCTGCAT R5’ATGGTGGAAAGCTCGATGAC	(AG)6	46.78	2.0000	2	149-152	1.0000	0.5085	0.6931	0.375
6	CP21*	F5’CAAGCACACGGTTGGAAACT R5’ATGTTGCAGCTCTGGTGCTA	(GA)6	46.78	4.6632	5	140-146	0.6667	0.7989	1.5717	0.751
7	CP22*	F5’CAAGCACACGGTTGGAAACT R5’ATGTTGCAGCTCTGGTGCTA	(GA)6	46.78	3.5644	5	142-148	0.8000	0.7316	1.3719	0.672
8	CP24**	F5’TCAAAGACTTGTGCGACGAC R5’CCTCAAGGATGTGAGGGTTA	(GA)6	46.78	2.8708	3	130-132	0.7667	0.6627	1.0747	0.576
9	CP26*	F5’GGGTGGTGTGATTGATAC R5’GGGAATTCACCGTCAACAGA	(AG)6	44.91	3.4026	5	150-156	0.7667	0.7181	1.3204	0.651
10	CP30*	F5’CTCCCTTCATTCTTGAGG R5’GCGCCATTTCTCTCTGTCTC	(AG)9	45.93	2.8986	5	148-154	0.7333	0.6661	1.2222	0.592
11	CP31	F5’CTCCCTTCATTCTTGAGG R5’GCGCCATTTCTCTCTGTCTC	(AG)9	45.93	3.8462	8	144-156	0.8667	0.7525	1.5825	0.700
12	CP32***	F5’GGGAAGATGAAGTTGGGACA R5’AGGACATCCCGGAGGATATT	(GA)8	46.78	2.4096	3	118-120	0.9000	0.5949	0.9489	0.495
13	CP33***	F5’GGGAAGATGAAGTTGGGACA R5’AGGACATCCCGGAGGATATT	(GA)8	46.78	2.7778	3	118-120	0.8000	0.6508	1.0549	0.563
		Mean			2.8971			0.7897	0.6338	1.1001	0.506

(* P<0.05, ** P<0.01, *** P<0.001)

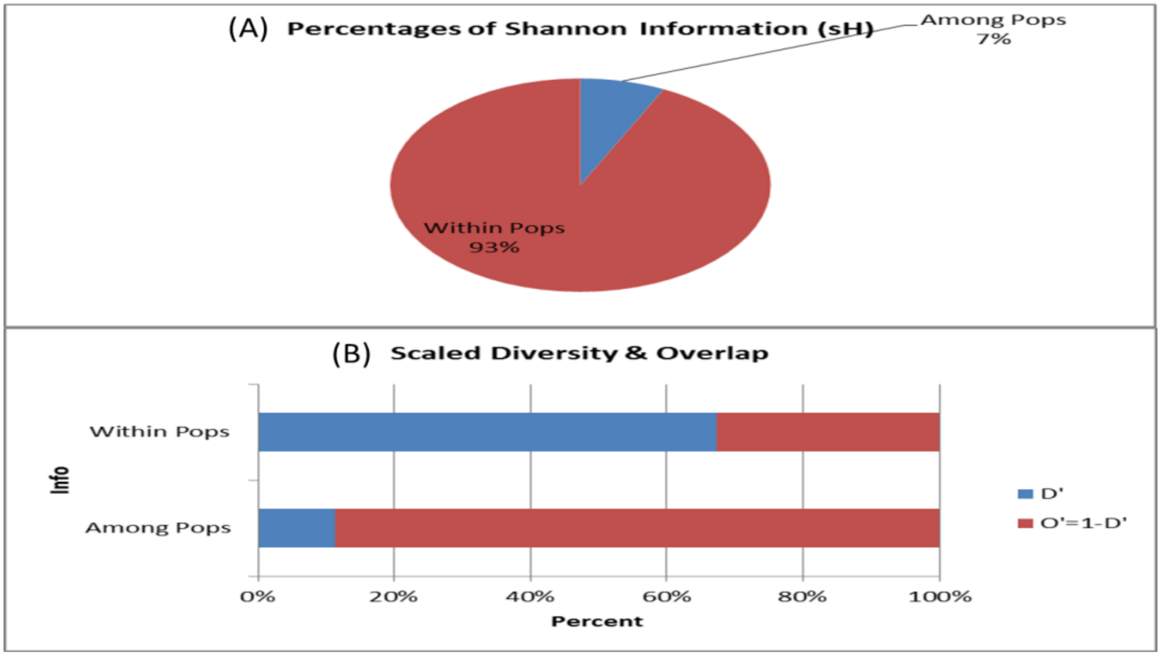


Figure 6: Representation of Shannon's diversity partition Statistics.

Pairwise Fst analysis revealed moderate population differentiation. It was found that the populations from Barnala and Bathinda are more differentiated with an Fst value of 0.028 and Nm of 2.0, while the populations from Ambala and Bathinda are least differentiated having an Fst value of 0.019. Both Nei's genetic distance (D) and identity (I) corroborate the Fst (a measure of population differentiation) and Nm (gene flow) based values. Clustering analysis also segregated Individuals from Barnala as separate cluster in contrast to population of Ambala and Bathinda which clustered together.

Table 8: Fst, Nm and Nei's genetic distance and identity between populations.

Population	Fst	D	I	Nm
Pop1-pop2	0.023	0.080	0.924	2.712
Pop1-pop3	0.019	0.061	0.941	6.439
Pop2-pop3	0.028	0.104	0.901	2.002

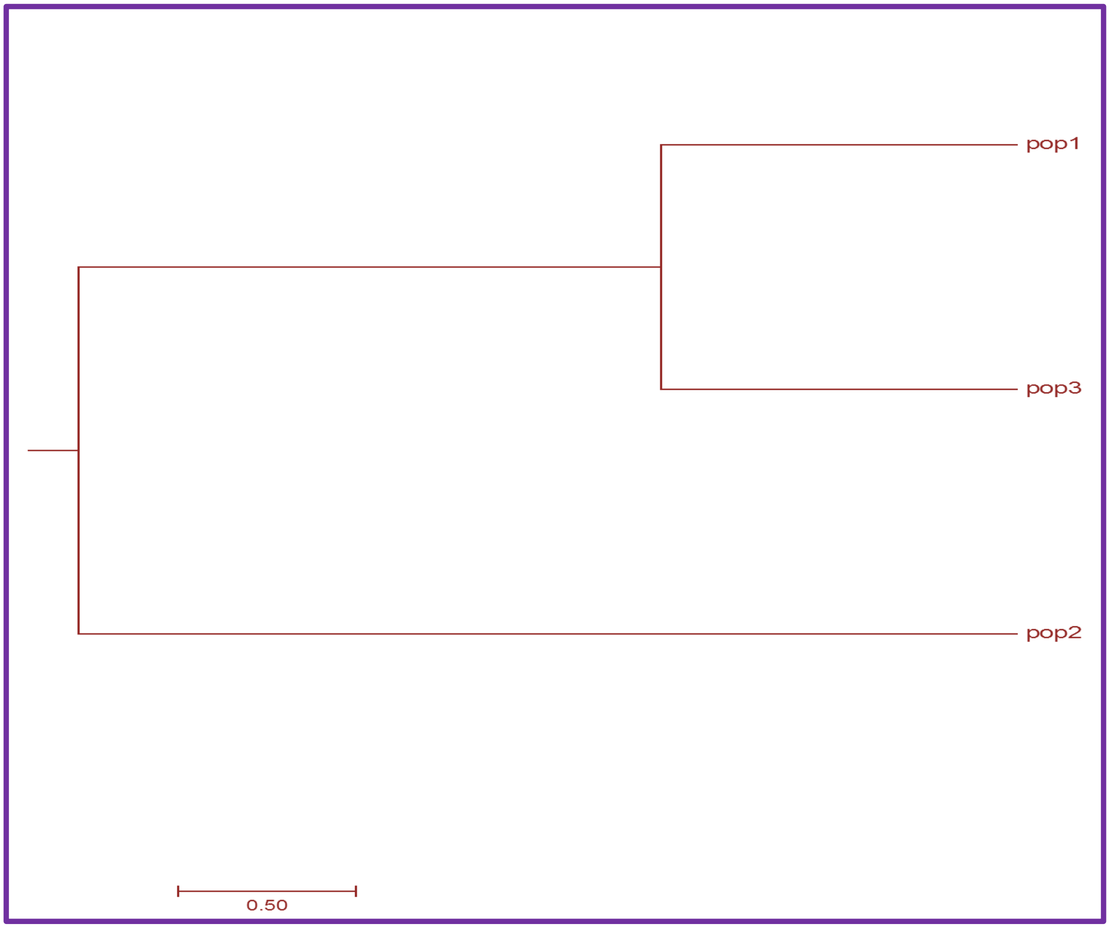


Figure 7: Dendrogram based on Nei’s genetic distance.

Chapter 5

Discussion

12.6 M raw reads were retrieved, quality checked and filtered to get 84636 reads. The alignment quality and completeness was checked by Bowtie2 and BUSCO respectively. The alignment rate of 93.7% was obtained and 76.31% of core genes were detected through the completeness assessment. This shows that the quality of the assembly was good for the further analysis. 12884 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 *Calotropis procera* genome. 25 SSRs were selected for characterization among which 13 markers were found to be polymorphic.

Heterozygosity determines the genetic variability in a population. It was observed that the mean observed heterozygosity (0.7897) was greater than the expected value (0.6338). The value of heterozygosity varies from 0 to 1. Zero value of heterozygosity means a homozygous population i.e. there is no variation whereas a value near to 1 shows high variability. If the heterozygosity comes 1 that means the population has large number of alleles with same frequency in the population. The value of 'Ho' is near to 1, it can be concluded that the populations have a lot of variant alleles due to high rate of gene flow among them. Shannon's information index (I) gives the idea about the diversity in a population i.e. richness hence, greater the value of Shannon's information index more is the allele richness. A mean of 1.1 was obtained which signifies the richness of allele diversity in the populations. Polymorphic information content (PIC) shows the power of the marker in revealing the polymorphism among the genotypes. It gives an overall idea about the efficiency of a marker and varies from 0 to 1. A 0 value indicates no variation in the alleles hence monomorphic, and value 1.0 indicates the presence of different alleles hence absolute polymorphic. The mean PIC value was calculated to be 0.506 which means the markers are informative. Fst or fixation coefficient measures the overall divergence in a population. The mean Fst was found to be 0.0335. Fis, the inbreeding coefficient varies from -1 to +1. A -1 Fis value indicates that all the individuals in the population are heterozygous and the Fis value we obtained was -0.3112 which indicates the presence of heterozygosity in the populations overall. The high gene flow dictates the high observed heterozygosity and moderate population differentiation. This finding is corroborated by the Shannon's diversity partition statistics which showed that among the populations the diversity is 7% while 93% resides within the populations. The beta diversity partitioning showed that among the population an average 67% of differentiated alleles and 33% of overlapping or common alleles are present within the populations while among the population only 11% of the differentiated alleles were found and 89% of the alleles were overlapping. This result shows higher diversity within the population than among the populations which is attributed to high gene flow.

Based on the Nei's genetic distance and identity UPGMA dendrogram was constructed that shows the separate clustering of population 1 and population 3. The genetic distance between population1 and 2 was 0.080, between population 2 and 3 it was 0.104. The pairwise genetic identity between population 1 and 3 was highest (0.941) showing the highest gene flow between them. According to the dendrogram the gene flow between Bathinda and Barnala was least and highest between Bathinda and Ambala.

SUMMARY

C. procera growing in the tropics has been used as medicine from a long time and has the potential utility as biodegradable fibre. The high-throughput development of molecular markers based on the next-generation sequencing data can be extended to analyse the genetic structure and marker-based selection in the species. 84,636 high-quality reads were obtained after pre-processing of 12.6 M raw reads available in the SRA database of *C. procera* at NCBI. The parameters such as alignment quality and completeness assessment showed that the assembly was optimum for further analysis. Twenty five SSRs were selected for primer synthesis and 13 of them were found polymorphic. Further analysis revealed a significant PIC value for the present set of loci. The F-statistics showed that the populations were heterozygous as the value of observed heterozygosity was higher than that of the expected heterozygosity. Shannon's information index displayed richness in the allelic diversity within the populations. Based on the beta diversity partition, it was concluded that there is a higher variation in the alleles within the population than among populations. Among the three populations sampled from Bathinda, Ambala, and Barnala, the first two were grouped into a separate clade showing high rate gene flow among them than the latter. From the above study the SSR markers were found informative that can be used for further genetics studies.

References

References

(https://github.com/FelixKrueger/TrimGalore

Alam, P., & Ali, M. (2009). Phytochemical investigation of *Calotropis procera* Ait roots.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Bajwa, K. S., Shahid, A. A., Rao, A. Q., Kiani, M. S., Ashraf, M. A., Dahab, A. A., ... & Aftab, A. (2013). Expression of *Calotropis procera* expansin gene CpEXPA3 enhances cotton fibre strength. *Australian Journal of Crop Science*, ***7*(2)**; 206.

Breseghello, F., & Sorrells, M. E. (2006). Association analysis as a strategy for improvement of quantitative traits in plants. Crop Science, **46(3)**; 1323-1330.

Chen, Q., Zhao, T., Wang, M., & Wang, J. (2013). Studies of the fibre structure and dyeing properties of *Calotropis gigantea*, kapok and cotton fibres. Coloration Technology, **129(6)**; 448-453.

Cho, Y. G., McCouch, S. R., Kuiper, M., Kang, M. R., Pot, J., Groenen, J. T. M., & Eun, M. Y. (1998). Integrated map of AFLP, SSLP and RFLP markers using a recombinant inbred population of rice (Oryza sativa L.). Theoretical and Applied Genetics, **97(3)**; 370-380.

Collard, B. C., & Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences, **363(1491)**; 557-572.

El-Bakry, A. A., Hammad, I. A., & Rafat, F. A. (2014). Polymorphism in *Calotropis procera*: preliminary genetic variation in plants from different phytogeographical regions of Egypt. Rendiconti Lincei, **25(4)**; 471-477.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Chen, Z. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology, **29(7)**; 644.

Gupta, P. K., & Varshney, R. K. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica, **113(3)**; 163-185.

Gupta, P. K., Varshney, R. K., Sharma, P. C., & Ramesh, B. (1999). Molecular markers and their applications in wheat breeding. Plant breeding, **118(5)**; 369-390.

Gupta, S., & Prasad, M. (2009). Development and characterization of genic SSR markers in Medicago truncatula and their transferability in leguminous and non-leguminous species. Genome, **52(9)**; 761-771.

Hu, X., Wang, J., Lu, P., & Zhang, H. (2009). Assessment of genetic diversity in broomcorn millet (Panicum miliaceum L.) using SSR markers. Journal of Genetics and Genomics, **36(8)**; 491-500.

Iqbal, Z., Lateef, M., Jabbar, A., Muhammad, G., & Khan, M. N. (2005). Anthelmintic activity of *Calotropis procera* (Ait.) Ait. F. flowers in sheep. Journal of Ethnopharmacology, **102(2)**; 256-261.

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular ecology, **16(5)**; 1099-1106.

Kamath, J. V., & Rana, A. C. (2002). Preliminary study on antifertility activity of *Calotropis procera* roots in female rats. Fitoterapia, **73(2)**; 111-115.

Kareem, S. O., Akpan, I., & Ojo, O. P. (2008). Antimicrobial activities of *Calotropis procera* on selected pathogenic microorganisms. African journal of biomedical research, **11(1)**.

Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Current protocols in bioinformatics, 11-7.

Lateef, D. D. (2015). DNA marker technologies in plants and applications for crop improvements. Journal of Biosciences and Medicines, **3(05)**; 7.

Miah, G., Rafii, M. Y., Ismail, M. R., Puteh, A. B., Rahim, H. A., Islam, K. N., & Latif, M. A. (2013). A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. International journal of molecular sciences, **14(11)**; 22499-22528.

Min, X. J., Butler, G., Storms, R., Tsang, A. (2005). OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Research. **33(2)**; 677–680.

Muriira, N. G., Xu, W., Muchugi, A., Xu, J., & Liu, A. (2015). De novo sequencing and assembly analysis of transcriptome in the Sodom apple (*Calotropis gigantea*). BMC genomics, **16(1);** 723.

Mutwakil, M. Z., Hajrah, N. H., Atef, A., Edris, S., Sabir, M. J., Al-Ghamdi, A. K., ... & El-Domyati, F. M. (2017). Transcriptomic and metabolic responses of Calotropis procera to salt and drought stress. BMC plant biology, **17(1)**; 231.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics, **89(3);** 583-590.

Olea, R. G., Oliveira, A. V., Silveira, L. M. S., & Silveira, E. R. (2002). Organic carbonate from *Calotropis procera* leaves. Fitoterapia, **73(3)**; 263-265.

Orwa, C., Mutua, A., Kindt, R., Jamnadass, R., & Simons, A. (2009). Agroforestree database: a tree species reference and selection guide version 4.0. World Agroforestry Centre ICRAF, Nairobi, KE.

Park, Y. J., Lee, J. K., & Kim, N. S. (2009). Simple sequence repeat polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops. Molecules, **14(11)**; 4546-4569.

Parrotta, J. (2001). Healing plants of Peninsular India. CAB International., Wallingford UK and New York, p. 944.

Parsons, W. T., Parsons, W. T., & Cuthbertson, E. G. (2001). Noxious weeds of Australia. CSIRO publishing.

Peakall, R. O. D., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Resources, **6(1)**; 288-295.

Ribaut, J. M., & Hoisington, D. (1998). Marker-assisted selection: new tools and strategies. Trends in Plant Science, **3(6)**; 236-239.

Roy, S., Sehgal, R., Padhy, B. M., & Kumar, V. L. (2005). Antioxidant and protective effect of latex of *Calotropis procera* against alloxan-induced diabetes in rats. Journal of Ethnopharmacology, **102(3)**; 470-473.

Scott, K. D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E. M., Lee, L. S., & Henry, R. J. (2000). Analysis of SSRs derived from grape ESTs. Theoretical and applied genetics, **100(5)**; 723-726.

Seiber, J. N., Nelson, C. J., & Lee, S. M. (1982). Cardenolides in the latex and leaves of seven Asclepias species and *Calotropis procera*. Phytochemistry, **21(9)**; 2343-2348.

SENAN, S., KIZHAKAYIL, D., SASIKUMAR, B., & SHEEJA, T. E. (2014). Methods for development of microsatellite markers: an overview. Notulae Scientia Biologicae, **6(1)**; 1.

Setty, S. R., Quereshi, A. A., Swamy, A. V., Patil, T., Prakash, T., Prabhu, K., & Gouda, A. V. (2007). Hepatoprotective activity of *Calotropis procera* flowers against paracetamol-induced hepatic injury in rats. Fitoterapia, **78(7-8)**; 451-454.

Sharma, P. C., Yelne, M. B., & Dennis, T. J. (2000). Database on medicinal plants used in Ayurveda central council for research in ayurveda and sidha. New Delhi, **1**; 114-117.

Sheng, Y., Zheng, W., Pei, K., & Ma, K. (2005). Genetic variation within and among populations of a dominant desert tree Haloxylon ammodendron (Amaranthaceae) in China. Annals of Botany, **96(2)**; 245-252.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, **31(19)**; 3210-3212.

Taramino, G., Tarchini, R., Ferrario, S., & Lee, M. (1997). Characterization and mapping of simple sequence repeats (SSRs) in Sorghum bicolor. Theoretical and Applied Genetics, **95(1-2)**; 66-72.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., & McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon associations, and genetic marker potential. Genome research, **11(8**); 1441-1452.

Thiel, T., Michalek, W., Varshney, R., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theoretical and applied genetics, **106(3)**; 411-422.

Tuntawiroon, N., Samootsakorn, P., & Theeraraj, G. (1984). The environmental implications of the use of Calotropis gigantea as a textile fabric. Agriculture, ecosystems & environment, **11(3)**; 203-212.

Varshney, A. C., & Bhoi, K. L. (1987). Some possible industrial properties of *Calotropis procera* (Aak) floss fibre. Biological Wastes, **22(2)**; 157-161.

Varshney, A. C., & Bhoi, K. L. (1988). Cloth from bast fibre of the *Calotropis procera* (Aak) plant. Biological wastes, **26(3)**; 229-232.

Varshney, R. K., Mahendar, T., Aggarwal, R. K., & Börner, A. (2007). Genic molecular markers in plants: development and applications. In Genomics-assisted crop improvement (pp. 13-29). Springer, Dordrecht.

Veeckman, E., Ruttink, T., & Vandepoele, K. (2016). Are we there yet? Reliably estimating the completeness of plant genome sequences. The Plant Cell, **28(8)**; 1759-1768.

Wang, Z., Weber, J. L., Zhong, G., & Tanksley, S. D. (1994). Survey of plant short tandem DNA repeats. Theoretical and applied genetics, **88(1)**; 1-6.

White, T. L., Adams, W. T., & Neale, D. B. (2007). Genetic markers-morphological, biochemical and molecular markers. Book: Forest Genetics, 53-74.

Yao, D. A. R. A., Sprycha, Y., Porembski, S., & Horn, R. (2015). AFLP assessment of the genetic diversity of *Calotropis procera* (Apocynaceae) in the West Africa region (Benin). Genetic resources and crop evolution, **62(6)**, 863-878.

Yeh, F. C. (1999). Microsoft window based freeware for population genetic analysis. Popgene Ver. 1. 31.

You, F. M., Huo, N., Gu, Y. Q., Luo, M. C., 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. BMC Bioinformatics. **253(5);** 1-13.

Appendices

APPENDIX – A

COMPOSITION OF CHEMICALS

I. DNA EXTRACTION

DNA EXTRACTION BUFFER

1M TrisCl (pH 8.0)	: 10.0 ml
0.5M Na.EDTA (pH 8.0)	: 4.0 ml
5M NaCl	: 35.0 ml
β- mercaptoethanol	: 0.2ml (add just before use)
4% CTAB	: 20.0 ml
10% PVP	: 15 ml
dH_2O	: 16 ml

II. DENATURING PAGE

6% Gel Pouring Solution

Urea	: 45g
Acrylamyde: Bis-acrylamide (19:1)	: 30ml
5X TBE	: 20ml
APS (prepare fresh)	: 0.75mg/ml
TEMED (v/v)	: 0.044% (add just before use)

Make final volume upto 100ml by adding distilled water.

III. 5X TBE

Tris base	: 54g
Boric acid	: 27.5g
EDTA (pH 8.0)	: 3.72g

Make final volume upto 1000ml by adding distilled water.

IV. **Silver stain** (prepare fresh)

Silver Nitrate	: 1g
Formaldihyde	: 1.5ml

Make final volume 1000ml by adding distilled water.

V. Developer solution (prepare fresh)

Sodium Carbonate : 30g

Formaldehyde : 1.5ml

Sodium Thiosulphate : 200µl (add just before use)

Make final volume upto 1000ml by adding distilled water.

VI. Fixer (prepare fresh)

Glacial acetic acid : 100ml

Distilled water : 900ml

VII. 10X sample loading buffer

Bromophenol blue : 25ml

Xylene cyanol : 25ml

0.5M EDTA (pH 8.0) : 200µl

99% formamide : 9.8ml

Mix PCR product and denaturing dye (1:1) and denature at 95°C for 5min, snap cool.

APPENDIX – B

DNA isolation

DNA was isolated by using the following protocol (Doyle, 1990):

1. 2g of plant material was homogenized to powder using liquid nitrogen in mortar pestle.
2. Homogenized plant material was transferred to 25ml centrifuge tube with pre-warmed DNA extraction buffer and incubated at 65°C for 1 hr with intermittent inversion.
3. The suspension was allowed to cool to room temperature and mixed with equal volume of chloroform: iso-amyl alcohol (24:1) with gentle mixing for 5min and centrifuged at 13,000 rpm for 35min.
4. The supernatant (aqueous phase) was transferred to a fresh centrifuge tube and 2/3rd volume of iso-propanol was added and mixed by gently. Centrifuged at 7500 rpm for 25min.
5. Supernatant was discarded and pallet was washed twice with 70% ethanol and absolute ethanol.
6. Washed DNA pallets were dried overnight at room temperature, later dissolved in TE buffer and at -20°C until used.

DNA purification

Purification of isolated DNA is required to remove contaminants such as RNA, proteins and polysaccharides. RNA contaminants are removed by the treatment of RNase A while proteins are removed by phenol: chloroform treatment. Following protocol was used to purify DNA:

1. To 500µL of DNA sample added 10µl RNase A (1µg/µL) and incubated at 37°C for 1hr.
2. To the above solution added equal volume of phenol: chloroform: iso- amyl alcohol (25:24:1) and centrifuged at 10,000 rpm for 5min. upper aqueous

phase was transferred to a new centrifuge tube and treated with equal volume of chloroform: iso-amyl alcohol (24:1).

3. The separated aqueous phase was transferred to new centrifuge tube and 1/10th volume of 3M sodium acetate and two volumes of absolute alcohol were added. DNA was pelleted by centrifuging at 13,000 rpm for 3min.
4. DNA pallet was washed twice with 70% ethanol and dried overnight. The dried DNA pallet was dissolved in TE buffer.