Development of Novel Microsatellite Markers and Construction of Genetic Linkage Map of Pearl Millet [*Pennisetum glaucum* (L.) R. Br.]

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

by

MAHENDAR THUDI

Department of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad 500 046
Andhra Pradesh, India

Enrolment No. 04LPPH06

Dec 2009

Department of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad 500046
INDIA

CERTIFICATE

This is to certify that **Mr. Mahendar Thudi** has carried out the research work embodied in the present thesis entitled **"Development of novel microsatellite markers and construction of genetic linkage map of pearl millet [*Pennisetum glaucum* (L.) R. Br.]"** for the degree of Doctor of Philosophy under my supervision in the Department of Plant Sciences, School of Life Sciences, University of Hyderabad.

This work has not been submitted for the award of any degree or diploma of any other University or Institute.

Prof. Arjula Ramachandra Reddy _____
Supervisor

Head _____
Department of Plant Sciences

Dean _____
School of Life Sciences

DECLARATION

I hereby declare that the work presented in this thesis entitled "**Development of novel microsatellite markers and construction of genetic linkage map of pearl millet [*Pennisetum glaucum* (L.) R. Br.]"** has been carried out by me under the supervision of **Prof. Arjula R. Reddy** in the Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad- 500046, and that this work has not been submitted for any degree or diploma of any other University or Institute. All the assistance and help received during the course of the investigation have been duly acknowledged.

Mahendar Thudi

Prof.Arjula Ramachandra Reddy
Supervisor
Department of Plant Sciences

Acknowledgements

I express my sincere gratitude to my advisor, Professor Arjula Ramachandra Reddy, for his constant guidance, encouragement and moral support throughout this research and supervising this thesis with such great interest. I am indebted to him for my life time and promise to carry his sense of attitude in my future scientific endeavours. He has been an excellent mentor, introducing me as a scientist, encouraging and getting me funds to work in his lab. The scientific knowledge and training in advanced technological innovations I gained during my stay with Prof. Reddy has helped me in my scientific endeavours at its best.

I wish to take an opportunity to express my sincere and profound thanks to Dr. Rajeev K. Varshney, Principal Scientist, Applied Genomics, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). For giving me an opportunity to work in such a renowned and well equipped laboratory. Not only a great mentor and colleague, he has also been a cornerstone in my professional development. His amenable and stupendous guidance, friendliness and constructive criticism helped me to complete this piece of work and gain preliminary knowledge of research and management of time as well as resources.

I express my heart felt thanks to Dr. Dave Hoisington, Deputy Director General (Research), ICRISAT for permitting me to carry out my research in

Applied Genomic Laboratory, ICRISAT for giving me an opportunity to work in such a renowned and well equipped laboratory.

With respect regards and immense pleasure, I wish to acknowledge and express sincere thanks from my heart to Dr. C.T. Hash, Principal Scientist, sorghum breeding, ICRISAT, Patancheru, for his valuable counsel, assistance, and valuable advice rendered for successful completion of my project work at ICRISAT.

I am thankful to our collaborators Dr. S. Senthilvel (Center for excellence in Genomics, ICRISAT), Dr. R.K. Aggarwal (CCMB, Hyderabad), R.M.D Koebner for extending a collaborative research support in the form of valuable suggestions and materialistic support.

I am thankful to the former Deans, School of Life Sciences, Prof. T. Suryanarayana and the present Dean, Prof. A. S. Raghavendra, for extending facilities available in the School for my research work. I would also like to thank the former Head, Department of Plant Sciences, Prof. P.B. Kirti, Prof Appa Rao Podalie and the present Head Prof. A. R. Reddy for providing all the required facilities in the department.

I would like to thank Dr. G. Padmaja, Dept. of Plant Sciences for being helpful through out my present work.

My special thanks Dr. V. B. Reddy, for useful discussions during data analysis and interpreting the results.

I wish to thank my senior colleagues Dr. Markandeya Gorantla, Dr. Ravindra Babu Peram, Dr. Chandra Shekar Akila, Dr. Shivarama Prasad Lekkala, Dr. Madana Mohan Reddy Ambavaram, Dr. Linga Reddy Gutha, Dr. Kiran Kumar Velpula, and Maheedhar Reddy for their help and suggestions during my research work. I would like to thank the assistance received from Raju, Ravinder, Poorna, Satish and Narasimha in the lab whenever needed.

With immense pleasure, I express my cordial thanks to Mrs. Seetha Kannan, for her kind help and co-operative for successful completion of my project work. I shall be failing in my duty if I do not express my cordial thanks to all my lab mates, friends and technical staff and administration staff in AGL, ICRISAT, Patancheru and in the Department of Plant Sciences, University of Hyderabad for their kind help and co-operation.

I would like to thank KVSS Prasad for his assistance at ICRISAT

Assistance rendered by the members of Central Support Lab, who helped me completing my lab work efficiently and smoothly at ICRISAT. I am also very thankful for their help and support provided to me by library staff and Learning Systems Unit staff at ICRISAT.

I feel privileged to express heartfelt words of appreciation to the staff members of Department of Plant sciences, University of Hyderabad for their kind co-operation in successful completion of my research work.

I sincerely acknowledge the UGC-CSIR, Govt. of India for awarding me with Junior Research Fellowship and Senior Research Fellowship, University of Hyderabad.

I thank all my friends for their constant support and great time spent outside my laboratory with fond memories.

I thank my parents, wife and brother for their lots of love, affection patience and support. I owe a debt of gratitude for their love, supporting spiritually, emotionally and financially, and understanding me along the path of my longest academic pursuits.

Date: Thursday, September 03, 2015 MAHENDAR THUDI

CONTENTS PAGE NO

1	**INTRODUCTION**	1
2	**REVIEW OF LITERATURE**	5
2.1	Origin, Evolution and Taxonomy	5
2.2	Cultivation	6
2.3	Economic Importance	7
2.4	Production Constraints	7
2.5	Molecular Genetic Studies in Pearl Millet	8
2.5.1	Molecular marker repertoire for pearl millet genetics and genome analysis	9
2.5.2	Genetic diversity assessment in pearl millet	11
2.6	Importance of Molecular Genetic Maps and QTL studies	12
2.6.1	Molecular genetic maps and QTL studies for pearl millet	13
3	**MATERIALS AND METHODS**	15
3.1.	Comparative Polymorphism Assessment of SSR, SSCP-SNP and CISP Marker Systems	15
3.1.1	Isolation and quantification of DNA from inbred lines	15
3.1.2	EST-SSR, gSSR, PM-SSCP-SNP and CISP assay	15
3.1.3	Wheat SSCP–SNP assay	16
3.1.4	Separation and visualization PCR amplicons	16
3.2	Development of Novel Simple Sequence Repeat (SSR) Markers	18
3.2.1	Construction of microsatellite enriched genomic DNA library	18
3.2.2	Sequencing of microsatellite enriched clones	18
3.2.2a	Plasmid isolation	18
3.2.2b	Purification of plasmid DNA	19
3.2.2c	Quantification of plasmid DNA	19
3.2.2d	Sequencing the clones	20
3.2.3	Primer design and synthesis	21
3.2.4	Optimization and validation of SSR markers	22
3. 3	Development of Genetic Linkage Map of Pearl Millet	23
3.3.1	Genotyping of polymorphic markers on the mapping population	23
3.4	Data Analysis	23
3.4.1	Scoring of PCR amplicons resolved on PAGE gels	23
3.4.2	Scoring the PCR amplicons resolved using Capillary electrophoresis	23
3.4.3	Allele call	24
3.4.4	Heterozygosity	24
3.4.5	Polymorphism Information Content	24
3.4.6	Diversity analysis	24
3.4.6	Cluster analysis	24
3.4.7	Linkage map construction	25
4	**RESULTS**	26
4.1	Comparative Marker Polymorphism Assessment	26
4.1.1	Amplification status of marker systems	26

4.1.2	Marker informativeness	29
4.1.2.a	Number of alleles	29
4.1.2.b	Genetic dissimilarity detected by marker systems	29
4.1.2.c	Polymorphism Information Content (PIC)	29
4.1.2.d	Heterozygosity	30
4.1.2.e	Effective number of alleles	30
4.1.3	Marker polymorphism	32
4.1.4	Informative and common polymorphic markers	32
4.1.4a	Cluster analysis	42
4.1.5	Correspondence between marker systems	44
4.2	Development of Novel Microsatellite Markers	44
4.2.1	Construction of library, cloning and sequencing of microsatellites	44
4.2.2	Mining for Simple Sequence Repeats (SSRs)	44
4.2.3	SSR frequency and distribution	45
4.2.4	Primer designing	45
4.2.5	Marker validation	48
4.2.5.a	Optimization of PCR conditions	48
4.2.5.b	Efficiency of Marker Development	48
4.2.4c	Polymorphism assessment of the markers	52
4.3	Construction of Genetic Linkage Map	68
4.3.1	Genotyping of polymorphic markers	68
4.3.2	Segregation distortion	68
4.3.3	Linkage map	68
5	**DISCUSSION**	69
5.1	Comparative marker polymorphism	72
5.1.1	Consequences of differential marker informativeness	72
5.1.2	Marker polymorphism among the parents of the mapping populations	73
5.1.3	Implications of common polymorphic markers among populations	76
5.1.4	Correspondence between the marker systems	77
5.1.5	Genetic relationships among inbred lines	78
5.2	Novel Microsatellite Markers	80
5.2.1	Efficiency of SSR development from genomic libraries	81
5.2.2	Frequency and Distribution of SSRs in Pearl Millet Genome	82
5.2.4	Development of microsatellite markers	83
5.2.5	Validation of microsatellite markers for use in genetic studies	86
5.2.6	Microsatellite attritions	87
5.3	First Recombinant Inbred Line (RIL) based genetic linkage map	89
5.3.1	Mappability of novel SSR markers	90
5.3.2	Marker distance, order and distribution along the linkage groups	90
5.3.3	Segregation distortion	92
5.3.4	Excess heterozygosity and map expansion	94
6	**SUMMARY**	98

7 REFERENCES 100
8 APPENDIX A 127

LIST OF TABLES

Table No	Title	Page No
Table 3.1	Pearl millet inbred lines and their segregating traits	17
Table 4.1	Summary on amplification status and marker informativeness attributes	27
Table 4.2	Detailed polymorphism data for all the polymorphic markers assayed	33
Table 4.3	Marker polymorphism in parental combinations of eleven pearl millet mapping populations	40
Table 4.4	Informative and common polymorphic markers across eleven mapping populations	41
Table 4.5	Cophenetic correlation coefficients among marker systems	44
Table 4.6	Summary of *MISA* search	45
Table 4.7	Frequency of individual SSR motifs in 2274 microsatellites isolated from the microsatellite enriched library	47
Table 4.8	Summary of primer optimization	48
Table 4.9	Characteristics of *Pennisetum glaucum* SSRs and efficiency of marker development	49
Table 4.10	Information on primer sequence, amplification status, polymorphism information content of novel SSR markers	53
Table 4.11	Distribution of markers on seven linkage groups of pearl millet	70

LIST OF FIGURES

FIGURE NO	LEGEND	PAGE NO
Figure 4.1	Four CISP markers resolved on 6% PAGE	28
Figure 4.2	Amplification profiles all markers systems; 22 pearl millet inbred lines along with two control genotypes are resolved on 6% PAGE	28
Figure 4.3	Distribution of polymorphism information content (PIC) data for different pearl millet crosses.	31
Figure 4.4	Neighbour joining dendrogram constructed from profile data generated using all marker systems (eSSRs, gSSRs, ISHs and ESHs)	43
Figure 4.5	Distribution of different SSRs in the microsatellite enriched library	46
Figure 4.6	Distribution of different SSR repeat classes in the microsatellite enriched library	46
Figure 4.7	PCR amplicons of novel SSRs resolved on 1.2% agarose gels.	50
Figure 4.8	Gel image showing polymorphic, monomorphic unamplified and smeary PCR products resolved on 6% PAGE gels.	51
Figure 4.9	Polymorphic markers resolved on 6% PAGE gels	52
Figure 4.10	A representative gel image showing segregation of three markers (Xpsms74, Xpsms83 and Xpsms84) on RIL population ICMB 841- P3 and 863B-P2	68
Figure 4.11	A linkage map of *Pennisetum glaucum* based on 86 loci (28 gSSRs, 25 CISPs, 13 SSCP-SNPs and 20 novel SSRs).	71

LIST OF ABBREVIATIONS

%	:	Percent
ABI	:	Applied Biosystems
AFLP	:	Amplified Fragment Length Polymorphism
APS	:	Ammonium Persulphate
BAC	:	Bacterial Artificial Chromosome
BIBAC	:	Binary Bacterial Artificial Chromosome
BLAST	:	Basic Local Alignment Search Tool
bp	:	base pair
CAP3	:	Contig Assembly Program version 3
CTAB	:	Cetyl Trimethyl Ammonium Bromide
CISP	:	Conserved Intron Spanning Primers
DHL	:	Double Haploid Lines
DNA	:	Deoxyribonucleic Acid
dNTP	:	deoxy Nucleotide Tri Phosphate
EDTA	:	Ethylenediamine Tetra Acetic Acid
EST	:	Expressed Sequence Tags
FAO	:	Food and Agriculture Organization
g	:	Gram
ha	:	Hectare
ISSR	:	Inter Simple Sequence Repeat
kg	:	Kilogram
M	:	Molar
MAS	:	Marker Assisted Selection
Mbp	:	Mega basepair
mg	:	Milligram
MDE	:	Mutation Detection Enhancement
Mha	:	Million Hectare
miRNA	:	micro RNA
µl	:	Microlitre
mM	:	Millimolar
mRNA	:	messenger RNA
Mt	:	Million tonns
ng	:	Nanogram
^{o}C	:	Degree Celsius
PAGE	:	Poly Acrylamide Gel Electrophoresis
PCoA	:	Principal Coordinates Analysis
PCR	:	Polymerase Chain Reaction
PIC	:	Polymorphism Information Content
pmole	:	Picomoles
QTL	:	Quantitative Trait Loci
RAPD	:	Randomly Amplified Polymorphic DNA
RFLP	:	Restriction Fragment Length Polymorphism
RNA	:	Ribonucleic Acid
s	:	Seconds
SNP	:	Single Nucleotide Polymorphisms

SSCP	:	Single Strand Confirmation Polymorphism
SSR	:	Simple Sequence Repeat
Taq	:	*Thermus aquaticus*
TE	:	Tris- EDTA
UPGMA	:	Unweighted Pair Group Method Based on Arithmetic average
UTR	:	Untranslated Regions
UV	:	Ultraviolet
V	:	Volt
µg	:	Microgram

Dedicated to

Dr. Rajeev K. Varshney

Principal Scientist (Applied Genomics)
International Crops Research Institute for the Semi-Arid Tropics
And
Leader, Sub Programme 2: Genomics towards Gene Discovery
Generation Challenge Programme (GCP, www.generationcp.org)

1. INTRODUCTION

Pearl millet [*Pennisetum glaucum* (L.) R. Br] is sixth most important C4 cereal crop, belonging to family Poaceae, cultivated in arid and semi-arid regions of the world. It has relatively small diploid genome (2*n*=2*x*=14; 2450 Mb), with a DNA content 1C=2.36 pg (Martel et al. 1997). It is the only cereal that reliably provides grain and fodder under dryland conditions, shallow or sandy soils with low fertility and low water holding capacity. It is cultivated on 26 m ha mostly in India and sub Saharan Africa and as forage in the United States of America (FAO and ICRISAT 2006). Pearl millet is excellent forage and because of its low hydrocyanic acid content, is the best annual grazing crop in the southern USA (Burton 1995) and an important summer forage crop in Australia and South America as well (Hanna 1996). From a nutritional standpoint, pearl millet is an attractive feed grain. Several studies indicate that metabolizable energy of pearl millet for non-ruminant animals is approximately equal to that of maize (Abate and Gomez 1984; Fancher et al. 1987; Amato and Forrester 1995). When compared to maize on a weight basis, pearl millet is 8% - 60% higher in crude protein, 40% richer in lysine and methionine, and 30% richer in threonine (Burton et al. 1972).

The greater part of the world's food production is restricted to a handful of plant species like rice, wheat, maize and barley and these species are not well adapted to the water shortages, so the changing climatic conditions may adversely affect the world's food security in the near future. Drought is the most important abiotic constraint challenging the world's food production. Though drought tolerance improvement is major concern in most breeding programmes, breeding for enhanced tolerance has been difficult due to genetic complexity of the trait, high genotype by environment interactions, lack of precise phenotypic evaluation strategies at the field level and moreover location specificity, variation in frequency, intensity and duration of drought. In the past decades, many consorted efforts have been made towards

dissecting physiological, biochemical, and genetic and molecular mechanism of drought tolerance among cereals. Declining rainfall in the past 25 years in combination with fast human population growth has led to severe production deficits, reinforcing the need of drought tolerant varieties in pearl millet as well as other crop species. But the pace and extent of efforts must dramatically be increased to be on par with parallel demands. Apart from drought major factors that restrict the grain and stover production potential of pearl millet are the high mean temperatures, low and erratic rainfall, and infertile sandy soils with low water-holding capacity that are typical of pearl millet production environments (Bidinger and Hash 2005). These abiotic stresses are accompanied by biotic production constraints diseases such as pearl millet downy mildew [caused by *Sclerospora graminicola* (Sacc.) J. Schroet.] (Breese et al. 2002), pearl millet rust [*Puccinia substriata* var *penicillaiare* (Speg.) Ramachar and Cumm.] (de Carvallo et al. 2006), blast [*Pyricularia grisea* (Cooke) Sacc.; teleomorph = *Magnaporthe grisea*] (Zeigler et al. 1994), smut (*Moesziomyces penicillariae* (Bref.) Vánky] (Chahal et al. 1986), and ergot (*Claviceps fusiformis* Loveless; Loveless 1967); insect pests including aphids, stem borers [*Coniesta ignefusalis* (Hampson)] (Peterson et al. 1997), and the millet headminer [*Heliocheilus albipunctella* (de Joannis)] (Henzell et al. 1997); parasitic weeds such as *Striga hermonthica*; and grain-feeding birds such as quelea.

With respect to the recent advances in the plant biotechnology, as the sequences of many plant genomes become available, the power of genomics for applied breeding has to be one of the most exciting advances of recent years. Complete sequence information, maps, and a huge array of molecular markers exist for rice; with more sequence information for other crops, new techniques for assessing allelic diversity, and a better understanding of synteny (Devos et al. 1995, 2000a, b), these are now being adapted for the breeding of other crops. Yet, for orphan crops like cowpea, common bean, the millets, tef, and cassava, we still have

insufficient numbers of ESTs, bacterial artificial chromosome (BAC) libraries, molecular maps, and markers. Pearl millet is considered as an orphan crop (Feltus et al. 2006) for having few or very little genomic information is available. In case of pearl millet, molecular markers in the form of restriction fragment length polymorphisms (RFLPs) were reported as early as 1992 (Gale and Witcombe 1992).

Several DNA marker systems and associated techniques are available today for fingerprinting plant germplasm but information on their relative usefulness in particular crops is limited. It is necessary to establish a common basis for assessing the effectiveness of the various marker systems currently available for DNA fingerprinting and several published reports on a variety of plant species have addressed this matter (Liu and Furnier 1993; Powell et al. 1996; Vogel et al. 1996; Milbourne et al. 1997; Russell et al. 1997). However, to date no such study has been reported in case of pearl millet. To date, the diversity of pearl millet has been studied using iso-enzyme loci (Tostain et al. 1987; Tostain and Marchais 1989; Tostain 1994), AFLP markers (vom Brocke et al. 2003), and RFLP markers (Bhattacharjee et al. 2002). New markers such as SSCP-SNP (Bertin et al. 2005) and microsatellite loci (Allouis et al. 2001; Qi et al. 2001, 2004; Budak et al. 2003) have recently been developed. However, they have not yet been used to assess the genetic diversity of inbreed lines of pearl millet. At present, about 500 homologous RFLP markers are available for pearl millet that provide a robust and reliable, but slow system for using molecular markers in applied breeding programs. Further, to date, about 140 SSR markers are available for pearl millet, although only 82 of these have been mapped. Interestingly, the majority of the pearl millet molecular markers mapped to date are clustered around the seven pearl millet chromosome centromeres and only a few marker loci are mapped to more distal regions (Qi et al. 2004). Such an uneven distribution of SSR and RFLP markers on the genetic map may reflect an uneven distribution

of recombination in pearl millet genome, or is simply the result of the low number of markers that have been mapped so far. Hence development of additional markers will be of great use. The first molecular marker-based genetic linkage map of pearl millet comprised largely of RFLP loci supplemented by a few isozyme loci (Liu et al. 1994a, b). In subsequent years, the linkage map has been expanded (Qi et al. 2004) and its complex relationships with the foxtail millet, rice, and sorghum genomes have been established (Devos et al. 1995, 2000a, b; Gupta and Varshney 2000; Bowers et al. 2003). The linkage maps available were based on F2 populations and no map to date is based on the recombinant inbred lines is available of pearl millet.

Hence, in view of above the current research has been initiated with the following objectives:

1. Polymorphism assessment using gSSRs, EST-SSRs, SSCP-SNP and CISP marker systems
    a. Determine the marker informative attributes
    b. Diversity analysis of 22 inbred lines of pearl millet
2. Development of novel SSR markers for pearl millet
    a. Construction of microsatellite enriched library
    b. Sequencing of SSR enriched clones
    c. SSR mining
    d. Development and validation of novel SSR markers
3. Construction of RIL based genetic linkage map for pearl millet

2. REVIEW OF LITERATURE

Pearl millet is the sixth most important cereal grain crop grown extensively as a food grain and is the staple source of nutrition for millions of people in arid and semi-arid tropics of Africa and India (Khairwal 2007). These areas are characterized by low or erratic rainfall, high temperature and low soil fertility however; pearl millet gives stable grain yields in such harsh environments (Hash et al. 2003). It is a crop that is able to produce nourishment from the poorest soils in the driest regions in the hottest climates, where no other cereals can grow. It is grown as a cover crop, or for forage or grain in the USA, Canada, Brazil, Australia, and Europe. This grain crop contributes significantly to food and nutritional security of the rural and urban poor in drier areas where it is valued equally for both its grain and fodder.

2.1 Origin, Evolution and Taxonomy

Pearl millet has originated in western Africa and was introduced to eastern Africa and then to Indian sub-continent some 2000 years ago (Brunken 1977). The genus *Pennisetum* includes pearl millet (*P. glaucum*), an important crop widely cultivated in Africa and Asia, and its two wild relatives *P. violaceum* and *P. mollissimum*. The wild forms of pearl millet are found only in Africa, where they have been involved in the domestication of the crop for several thousand years. These species are not reproductively isolated and are subunits of the same gene pool (Brunken 1977). Other species of this genus are prized for their fodder (*P. purpureum, P. ramosum*) or are used as ornamental plants (*P. uillosum, P.setaceum*). This genus is a heterogeneous assemblage of species with different basic chromosome numbers *(x* = 5, 7, 8, and 9), ploidy levels (diploid to octoploid), reproductive behaviour (sexual or apomictic), and life cycle (annual, biennial, or perennial). According to morphological characters, species of this genus are classified into five sections: *Penicillaria, Breuiualvula*,

Eu-pennisetum, *Heterostachya*, and *Gymnothrix* (Stapf and Hubbard 1934). The most important transformations in pearl millet, compared with its wild relatives, are the suppression of spikelet shedding, the size reduction of bristles and bracts leading to uncoated seed, the increase in seed size and spikelet pedicel length and the loss of dormancy. Regarding plant architecture and phenology, drastic changes are evidenced by the tillering habit (low number of tillers and hierarchy in the flowering of tillers) and spike length (gigantism), both resulting from an increase in apical dominance (e.g. in maize: Doebley et al. 1997). Taxonomically pearl millet belongs to family *Poaceae* sub family *Panicoideae* section *Penicillarium* and possesses $2n = 2x = 14$ chromosomes.

2.2 Cultivation

India is the largest producer of this crop, both in terms of area (9.1 m ha) and production (7.3 m t), with an average productivity of 780 kg ha^{-1} during the last five years (Khairwal 2007). Nearly 50% of the millet area is under hybrid cultivars. In Western and Central Africa (WCA), pearl millet open pollinated varieties are cultivated in 16 m ha with a production of 11.5 m t and productivity of 800 kg ha^{-1}. The other major pearl millet producing countries are Senegal, Mali, Burkina Faso, Niger, Nigeria, Chad, Sudan, and India. It is also grown in Oceania and the Americas, predominantly as a forage and/or mulch component of minimum tillage-based cropping systems. Pearl millet is cultivated in the hot arid regions of India that spreads across parts of Rajasthan, Haryana and Gujarat. Rajasthan constitutes about 49% area and 39% of production of pearl millet in the country, followed by Maharashtra with 17% area and 15% production and Gujarat with 10% area and 14% production. The dual purpose nature of pearl millet offers both food and fodder security in semi arid tropical regions of the country. Improvements in pearl millet yields in the developing countries, mainly in India, have occurred largely due to the development, release and widespread multiplication of

improved open-pollinated and hybrid cultivars. Nearly 50% of the millet area is under hybrid cultivars.

2.3 Economic Importance

Pearl millet grain contains 27% to 32% more protein, higher concentration of essential amino acids, twice the extract (fat) and higher gross energy than maize (Ejeta et al. 1987; Davis et al. 2003). The energy density of pearl millet grain is relatively high, arising from its higher oil content relative to maize, wheat, or sorghum (Hill and Hanna 1990). Collins et al. (1997) noted commercial layers given feed containing pearl millet grain had lower omega-6 to omega-3 fatty acid ratio, endowing the eggs with a fatty acid profile more favorable to human health. The amino acid profile of pearl millet grain is better than that of normal sorghum or normal maize and is comparable to those of the small grains wheat, barley, and rice (Ejeta et al. 1987) with a less disparate leucine/isoleucine ratio (Hoseney et al. 1987; Rooney and McDonough 1987). The lysine content of the protein reported in pearl millet grain ranges from 1.9 to 3.9 g per 100 g protein (Ejeta et al. 1987). Pearl millet grain appears to be generally free of any major anti-nutritional factors, such as the condensed tannins in sorghum grain having a pigmented testa, which reduces protein availability.

2.4 Production constraints

Pearl millet production in the developing world is hampred by a number of biotic and abiotic factors. Primary biotic constraints in India include downy mildew (*Sclerospora graminicola* (Sacc.) Schroet.), *Striga* (*Striga hermonthica* Benth.), rust (*Puccinia substriata* var. *indica*), pyricularia leaf blight (*Magnoporthe grisea*), and root knot nematode (*Meloidogyne arenaria*). Among the fungal diseases downy mildew caused by the fungus *Sclerospora gramincola*, is the most widespread and destructive disease of the crop. In India

during years of severe attack, up to 30 percent of the harvest can be lost. Other major diseases affecting pearl millet are smut (*Moeszimyces penicikkariae*), ergot (*Claviceps fusiformis* Loveless) and rust *(Puccinia substriata*). Drought is a major production constraint, reducing pearl millet yields in arid and semi-arid regions of the world. Drought stress is a regular occurrence in these regions, making stress tolerance an essential attribute of new pearl millet cultivars. Breeding for drought tolerance has been slow due to the complex physiological responses to drought, various environmental factors, and their interactions. Major factors that restrict the grain and stover production potential of pearl millet are the high mean temperatures, low and erratic rainfall, and infertile sandy soils with low water-holding capacity that are typical of pearl millet production environments (Bidinger and Hash 2005).

2.5 Molecular Genetic Studies in Pearl Millet

The discovery of restriction enzymes (Smith and Wilcox 1970) and the polymerase chain reaction (PCR; Mullis and Faloona 1987) have opened new vistas in genetics and plant breeding. The development of molecular marker techniques has lead to a great increase in our knowledge of cereal genomics and our understanding of structure and behaviour of the cereal genomes (Gupta and Varshney 2000). In recent years a number of molecular marker techniques have been developed and applied to a range of species including cereals (Varshney et al. 2004). Various kinds of molecular markers are available for plant genome analysis to date for instance, hybridization based markers like restriction fragment length polymorphism (RFLP), PCR based marker systems like random amplified polymorphic DNA (RAPD), inter-simple sequence repeats (ISSR), sequence tagged sites (STS), sequence tagged microsatellite sites (STMS), simple sequence repeats (SSR) and single nucleotide polymorphism (SNP). However, the application of the marker systems depends on the purpose of the study. Molecular markers have been proven to useful in characterization of the available germplasm

and estimation of genetic diversity with the aim of using this information for selection of parents for hybridization programmes (e.g. Bhat et al. 1999; Srivastava et al. 2001; Prashanth et al. 2002; Roy et al. 2002); for tagging genes/QTLs for qualitative and quantitative traits through marker-trait association for marker assisted selection (MAS) (e.g. Hittalmani et al. 1995; Nair et al. 1996; Naik et al. 1998; Prasad et al. 1999; Negi et al. 2000; Sharma et al. 2002; Yadav et al. 2002); and preparation of molecular maps (Gupta et al. 2002b) and their use for interval mapping of QTL for polygenic traits (Prasad et al. 2002). Among various molecular marker systems employed for genetic enhancement of various crops, SSRs are considered to be the markers of choice for being abundant and dispersed throughout all of the eukaryotic genomes (Morgante et al. 2002). The codominant nature and allelic polymorphism revealed by SSR markers has provided detailed information on genetic structure (Bonnin et al. 2001) and gene flow (Konuma et al. 2000) in natural plant populations.

2.5.1 Molecular marker repertoire for pearl millet genetics and genome analysis

Although pearl millet is staple food for over 90 million people round the world; this crops lags behind in the genomics revolution and is considered as genomics research orphan (Feltus et al. 2006). However, during recent past, with the advances in genomics research technologies, efforts are being made world wide for the development of genomic resources for genetic improvement of pearl millet. As a result genomic resources like bacterial artificial chromosome (BAC) libraries (Qi et al. 2001; Allouis et al. 2001), small insert libraries (Budak et al. 2003), cDNA libraries (Mishra et al. 2007) are being made available for pearl millet research community. Apart from this with the establishment of expressed sequence tag (EST) projects for gene discovery programs in several plant species, a wealth of DNA sequence information has been generated and deposited in on-line databases (Rudd et al. 2003). Nevertheless, under the Generation Challenge Program (GCP), subtractive cDNA libraries for

high throughput cloning of stress-responsive genes that are differentially expressed in response to salinity, cold, and dehydration stress in pearl millet are established (Mishra et al. 2007). The multitude of potential applications of DNA marker technologies to the improvement of pearl millet are comprehensively reviewed by Hash et al. (2002). In case of pearl millet molecular markers in the form of restriction fragment length polymorphism (RFLPs) are first employed as early as 1992. However, the usage of microsatellite is reported in 2001 (Qi et al. 2001). A bacterial artificial chromosome (BAC) library was constructed using nuclear DNA from pearl millet (*Pennisetum glaucum*), and used as a resource for the isolation of microsatellite sequences. The library corresponds to 5.8 haploid genome equivalents. The frequency of BAC clones carrying inserts of chloroplast DNA was estimated to be less than 1% by hybridisation with a rice chloroplast probe.

Earlier, development of SSRs was very costly as it involved high cost of library screening and clone sequencing. But now large public SSR datasets exist for several crop species including pearl millet. As on 25th feb 2009 24, 642 ESTs, 4900 genome survey sequences (GSS) and 757 nucleotide sequences are publicly avalible in the National Center for Biotechnology Information (NCBI) Genbank. A set of 25 new pearl millet SSR markers were developed using publicly available sequence information from 3520 expressed sequence tags (ESTs) by Senthilvel et al. (2004). A new marker system, single-strand conformational polymorphism (SSCP)-SNP, was developed using annotated rice genomic sequences to initially predict the intron-exon borders in millet ESTs and then to design primers that would amplify across the introns (Bertin et al. 2005). However, little research has been conducted on the transferability of SSR and other markers like conserved intron scanning primer (CISP) from major cereal crops to pearl millet. Feltus et al. (2006) designed 384 PCR primer pairs to

conserved exonic regions flanking introns, using *Sorghum/Pennisetum* expressed sequence tag alignments to the *Oryza* genome.

2.5.2 Genetic diversity assessment in pearl millet

Genetic diversity studies in pearl millet germplasm offer possibilities for their use in improving pearl millet varieties. The breeding behavior of pearl millet, and the structure of genetic diversity within this species, has strong implications for the use of molecular markers in its diversity assessment. Mostly in all crop species, phenotypic estimates of genetic diversity are biased by the environment(s) in which evaluation occurs. Pearl millet is a cross pollinated species, the genetic resource characterisation and evaluation have demonstrated high phenotypic diversity in cultivated material (Singh 1987; Tostain et al. 1987; Wilson et al. 1990; Ouendeba et al. 1995). Genetic diversity in pearl millet has been assessed using isozymes in landraces from Africa and India found an intra-population diversity of 70–90% of the total diversity, but this was dependant on their regions of origin (Tostain et al. 1989). Isozymes have failed in the differentiation and classification of landraces because of the small number of marker loci available, which provide poor coverage of the genome, and a low level of allelic polymorphism. The past limitations associated with pedigree data and morphological, physiological and cytological markers for assessing genetic diversity in cultivated and wild plant species have largely been circumvented by the development of DNA-based markers. Further, these DNA-based markers are environmentally neutral and can provide estimates of the degree and distribution of variation between and within accessions. RFLP analyses showed that genetic polymorphism in the pearl millet gene pool is very high, not only between species (Liu et al. 1992), but also within landraces of the cultigen (Pilat-Andre et al. 1992). This is because the crop is allogamous and, most importantly, subject to frequent genetic exchange between wild and cultivated genotypes (Brunken 1977). The potential of DNA markers such as

microsatellites, minisatellites and RAPDs was investigated in pearl millet with respect to their abundance and variability (Choudary et al. 1998). The influence of farmer management on pearl millet landrace diversity was estimated using variation at 163 amplified fragment length polymorphism marker (AFLP) loci (Busso et al. 2000). Genetic diversity among 70 maintainers and two pollinators of sub-Saharan and Indian origin was studied for simple sequence repeat (SSR) loci using 34 primer pairs. A significantly lower number of alleles and lower gene diversity was reported in cultivated pearl millet accessions than in wild accessions (Mariac et al. 2006). Both RAPD and SSR markers were employed by Chandra-Shekara and colleagues (2007) to ascertain the genetic relationships among elite inbred lines of pearl millet. The combined data sets revealed moderate genetic divergence among the elite pearl millet germplasm lines.

2.6 Importance of Molecular Genetic Maps and QTL studies

A genetic linkage map is a fundamental organizational tool for genomic research. The most important applications of genetic maps are towards:

(1) A basic knowledge of genome organization and evolution;

(2) The localization of monogenic and oligogenic traits; and

(3) Studies of genetic diversity.

Wide applications of DNA-based markers and resulting molecular genetic linkage maps allow the molecular dissection of genetic variation of complex phenotypes through the design, execution, and analysis of QTL mapping experiments. Nowadays the molecular markers, as a toolbox available to plant breeders, offers several new possibilities for increasing productivity, crop diversification and production, while developing a more sustainable agriculture. Substantial progress has been made in recent years in mapping, tagging and

isolating many agriculturally important genes using molecular markers due in large part to improvements in the techniques that have been developed to help find markers of interest.

2.6.1 Molecular genetic maps and QTL studies for pearl millet

Mapping population development was as described by Hash and Witcombe (1994), with RFLP skeleton mapping, trait phenotyping, and QTL mapping as described by Yadav et al. (1999a, b). In case of pearl millet nevertheless efforts were concentrated on the development of the first marker-based genetic maps using restriction RFLP markers (Liu et al. 1992, 1994a, 1994b). This began with a focus on the use of RFLP markers detected using homologous pearl millet probes, with ^{32}P-mediated autoradiography to visualize banding differences. A few heterologous probes from rice, wheat and barley along with several isozymes and known function probes were also included in this base map. The result was a map of over 180 loci that covered approximately 350 cM and contained the expected seven linkage groups, one extra couplet, and a floating point. This map probably corresponds to the centromeric regions of the seven pearl millet chromosome pairs. Furthermore, the development and mapping of simple sequence repeats (SSRs), was reported by Qi et al. (2001, 2004). Now more than 600 molecular markers have been created and mapped for pearl millet, a more detailed linkage map has been produced. More recently, subset of 21 polymorphic EST-SSRs and 6 recently developed genomic SSR markers were mapped using existing mapping populations (Senthilvel et al. 2008). Linkage map positions of these EST-SSRs were compared by homology search with mapped rice genomic sequences on the basis of pearl millet-rice synteny. Most new EST-SSR markers mapped to distal regions of linkage groups, often to previous gaps in these linkage maps. These new EST-SSRs are now are used by ICRISAT in pearl millet diversity assessment and marker-aided breeding programs.

Recent breeding research has mapped several quantitative trait loci (QTLs) for components of grain and stover yield *per se*, as well as yield maintenance, under terminal drought stress conditions. QTLs for seedling thermotolerance (Howarth et al. 1997), disease resistance (Jones et al. 1995, 2002), drought tolerance (Yadav et al. 2002, 2004), flowering time and grain and stover yield (Yadav et al. 2003), and ruminant nutritional quality of straw (Hash et al. 2003) have been mapped. Development of molecular maps for economically important traits such as downy mildew resistance (Jones et al. 1995) and drought tolerance (Yadav et al. 1999a) are available. Markers linked with resistance QTLs, and addition allow copy probes having high levels of polymorphism in initial variability studies, were converted to sequence tagged site (STS) markers (Money et al. 1994), but these have proven to be disappointing because much of the polymorphism of the RFLP markers on which they were based can no longer be detected without the use of multiple restriction enzymes. New resistance genes have been used to strengthen the crop's natural resistance to disease. One such hybrid, HHB 67, has proved exceptionally successful in India and has yet to succumb to the onslaught of downy mildew (Hash et al. 2003). Currently, genetic maps are available covering the seven pearl millet chromosomes (Liu et al. 1994b; Devos et al. 1995b). However, these maps are short, and the extent of genome coverage provided by the pearl millet maps is low.

3. MATERIALS AND METHODS

3.1. Comparative Polymorphism Assessment of SSR, SSCP-SNP and CISP Marker Systems

A set of 22 pearl millet inbred lines, which include parents of 11 mapping populations used in various breeding programs at ICRISAT (Table 1) were employed for comparative marker polymorphism assessment. For this a total of 627 primer pairs of which 100 pearl millet SSRs (Qi et al. 2004), 100 pearl millet Single Strand Confirmation Polymorphism (PM_SSCP-SNP) primers (Bertin et al. 2005), 57 wheat SSCP-SNP (W_SSCP-SNP; Botley et al. 2006), and 310 Conserved Intron Spanning Primers (CISP; Feltus et al. 2006), were screened on the 22 inbred lines.

3.1.1 Isolation and quantification of DNA from inbred lines

Genomic DNA of all the 22 inbred lines was isolated adopting high throughput DNA isolation protocol suggested by Mace et al. 2003. In order to determine the quality of genomic DNA and for the presence of the any contaminants, an aliquot of 1 µl of DNA from each sample along with 100 ng of molecular weight marker (λ DNA, Amarsham Biosciences,) were initially analyzed by electrophoresis on 0.8 % agarose gel containing ethidium bromide (0.5 µl/10 ml of gel) and run in 0.5X TBE (Tris Borate EDTA) buffer at a constant voltage (100 V) for one hour. The gel was viewed under UV illumination and recorded using UVi Tech gel documentation system (DOL-008.XD, England).

3.1.2 EST-SSR, gSSR, PM-SSCP-SNP and CISP assay

Polymerase chain reactions (PCR) were performed for EST-SSR, gSSR, PM-SSCP-SNP and CISP marker systems on 24 genotypes in a 5 µl reaction volume [0.5 µl of 10X PCR buffer

(Bioline, U.K.), 1.00 µl of 10 mM Mg^{++} (Bioline, U.K.), 0.25 µl of 2 mM dNTPs, 1.0 µl of 2 pM primer (MWG- Biotech AG), 0.1 U (0.2 µl of 0.5 U/µl) Taq polymerase (Bioline, U.K.) and 1.0 µl of template DNA (5 ng/µl)] in 96-well microtiter plate. A common touch down PCR amplification cycle was used for EST-SSR, gSSR, PM-SSCP-SNP and CISP markers as follows: The three min initial denaturation cycle (95˚C) step followed five cycles for 20 sec at 94˚C, 20 second at 60°C (-1˚C per cycle) and 30 sec at 72°C then by 30 cycles of 94˚C for 20 seconds, 20 sec at 56˚C and 30 sec at followed by 20 min final extension.

3.1.3 Wheat SSCP–SNP assay

Wheat SSCP–SNP primers were amplified in 10 µl reaction volume, the PCR components are same as mentioned above. The PCR were carried out with an initial denaturation cycle for 5 min at 94˚C followed by 94˚C for 30 sec and 34 cycles of 59˚C for 1 min, 72˚C for 1 min followed by a final extension for 5 min at 72˚C.

3.1.4 Separation and visualization PCR amplicons

Initially few random PCR amplicons of each gSSR and eSSR marker was initially tested for PCR amplification on 1.2% agarose gel and resolved on 6% polyacrylamide gel electrophoresis (PAGE) gels and visualized through silver staining (Tegelstrom et al. 1992). The PCR amplicons of SSCP-SNP, CISP markers were denatured at 94 °C for 5 min and separated on mutation detection enhancement (MDE) gels (Rockland, ME 04841 USA) and visualized through silver staining. Sequi-Gen® GT IPC Assembly, 38 × 30 cm BIORAD sequencing gel units were used for casting both PAGE and MDE gels for size separation and to detect single strand confirmation polymorphism respectively. Sixty fluorescently labelled gSSRs were screened on ABI 3100 genetic analyzer. Allele sizing was done using Genotyper version 3.7.

Table 3.1: Pearl millet inbred lines and their segregating traits

S. No	Parents	Traits segregating
1	H 77/833-2 × PRLT 2/89-33	Drought tolerance, seedling heat tolerance, salinity tolerance, downy mildew resistance, tillering, grain size, foliage pubescence
2	ICMB 841-P2 × 863-P3	Drought tolerance & downy mildew resistance, Stover quality, grain size, salinity tolerance, zinc and iron density of grain, pigmentation of anthers, nodes, glumes
3	Tift $23D_2B_1$-P5 × WSIL-P8	Downy mildew resistance, panicle length, and various morphological markers (sheath color, plant base pigmentation, pubescence, anther color, glossiness, yellow vs. green foliage color)
4	PT 732B-P2 × P1449-2-P1	Plant height (d2), downy mildew resistance, Stover quality
5	LGD 1-B-10 × ICMP 85410-P7	Flowering time, downy mildew resistance, grain size, panicle
6	81B-P6 × ICMP 451-P8	World Reference Mapping Population; Plant height (d2), panicle bristling, downy mildew resistance, rust resistance (forage project), foliage pubescence, CMS maintenance vs fertility restoration for A1 system
7	ICMP 451-P6 × H 77/833-2-P5 (OT)	Plant height, downy mildew resistance, rust resistance, seedling heat tolerance, tillering, grain size
8	W 504-1-P1 × P310-17	Downy mildew resistance, glume pigmentation, seedling vigor
9	IP 18293-P152 × Tift 238D1-P158	Plant height (d1 and d2), foliage color (purple vs green), downy mildew resistance
10	ICMB 89111-P6 × ICMB 90111-P6	Plant height (d2), downy mildew resistance
11	IPC 804 × 81B	Plant height (d2), panicle bristling, foliage pubescence, CMS maintenance vs. fertility restoration for A1 system

3.2 Development of Novel Simple Sequence Repeat (SSR) Markers

3.2.1 Construction of microsatellite enriched genomic DNA library

As a first step towards development of novel microsatellite or simple sequence repeat (SSR) markers a microsatellite enriched genomic DNA library was constructed as a collaborative effort between ICRISAT and Centre for Cellular and Molecular Biology (CCMB, Hyderabad) at Dr. Ramesh K. Aggarwal's laboratory. The pearl millet genotype "Tift $23D_2B_1$-P5" was chosen for the generation of a micosatellite enriched genomic DNA (gDNA) library. Tift $23D_2B_1$-P5 was derived by several generations of selfing and selection of typical plants in elite downy mildew susceptible seed parent maintainer line Tift $23D_2B_1$. The microsatellite enrichment procedure adopted was similar to the enrichment procedure described by Aggarwal et al. (2004). The microsatellite library was enriched for CA, GA, CAA and AGA repeats.

3.2.2 Sequencing of microsatellite enriched clones

3.2.2a *Plasmid isolation*

A set 960 putative microsatellite enriched positive clones were picked and plasmid DNA was isolated using alkaline lysis method described here under:

1. Positive clones were picked and inoculated into 5 ml of Luria Broth (LB) medium containing ampicilin (50 mg/ml)
2. Inoculated culture was incubated at 37 $^{\circ}$C for 16 h with a gentle shaking at 225 rpm
3. The overnight culture was centrifuged at 6000 rpm for 10 min and supernatant was discarded
4. The pellet was gently dispensed in 200 µl of Solution-I [0.9 gm of glucose, 5 ml of Tris-HCl (0.5M; pH-8) and 2 ml of EDTA (0.5M; pH-8)] and suspension was transferred into 1.5 ml centrifuge tubes

5. for each sample 300 µl of Solution-II (1% SDS, 0.2N NaOH) was added and mixed gently until the solution is uniform
6. then 300 µl of Solution III (7.5 M Ammonium acetate) was added,vortexed to mix well; kept in ice for 10 min and centrifuged at 10000 rpm for 10 min
7. Supernatant was collected in a fresh tube and 20 µl of RNase (10 mg/ml) was add mixed well and kept in incubator for 1 h.

3.2.2b *Purification of plasmid DNA*

1. To the RNase treated plasmid DNA samples 350 µl of chloroform : isoamyl alcohol (IAA; 24:1) was added and mixed gently and centrifuged at 5000 rpm for 5 min
2. The aqueous layer was collected into the fresh 1.5 ml microcentifuse tubes and equal amounts of chloroform isoamylalcohol (CIAA) was added, mixed well and centrifuged at 5000 rpm for 5 min
3. The aqueous layer was collected into the fresh microcentrifuge tube and equal volume of chilled isopropanol was added, mixed gently, kept at -20 $^{\circ}$C for 30 min and centrifuged at 12000 rpm for 10 min
4. Supernatant was poured off, pellet was ethanol washed by adding 500 µl of 70% ethanol and centrifuged at 8000 rpm for 5 min
5. Ethanol was poured off, and air dried for about 30- 40 min.
6. Plasmid DNA was then dissolved in 30 µl of $T_{10}E_1$

3.2.2c *Quantification of plasmid DNA*

The plasmid DNA thus isolated was checked for quality on 0.8% agarose gels and quantity using standard λ-DNA weight markers (Fermentas, USA).

3.2.2d *Sequencing the clones*

A set of 960 positive clones were sequenced from both (5' and 3') directions using M13F-pUC (-40) and M13R-pUC (-40) as sequencing primers, by adopting Sanger's dideoxy sequencing method and BigDye Terminator version 3.1 kit on ABI 3700 (Applied biosystems, USA).

For sequencing, microsatellite enriched clones, forward and reverse sequencing polymerase chain reaction (PCR) were performed separately in 10 µl reaction volume containing 2 µl of Big Dye Terminator Version 3.1 (BDT v 3.1) reaction mix (Applied Biosystems, Foster city, USA), 0.5 µl of 5X reaction buffer (Applied Biosystems, USA). 3.2 picomoles of forward and reverse primer for respective reactions, 1 µl of plasmid (100 ng/µl) and 6 µl of sterile water. The cycle sequencing PCR profile used involved 30 sec of initial denaturation followed by 40 cycles of 10 sec at 96 $^{\circ}$C (denaturation), 5 sec at 50 $^{\circ}$C (primer annealing) and 60 $^{\circ}$C for 4 min (primer extension) as per instruction manual of BDT v3.1 sequencing Kit (Applied Biosystems, USA).

In order to remove excessive polyA overhangs, unused dNTPs PCR products were treated with ExoSAP. Similarly to remove the dye blobs, PCR products were purified by ethanol treatment as follows:

1. For each sample 2.5 µl of 125 mM EDTA and 25 µl of 100 % ethanol were added, mixed well and incubated at room temperature for 15 min and centrifuged at 4000 rpm for 30 min at 4°C.
2. Ethanol and EDTA mix was poured off by inverting the plate on paper towels.
3. To remove any trace of ethanol and ethanol-EDTA mix the plate was inverted onto paper towel and centrifuged at 200 rpm for 1 min.

4. Followed by of addition of 60 µl of 70% ethanol, centrifuged at 4000 rpm for 20 min at 4°C.
5. After centrifugation, 70% ethanol was poured off by inverting plate on tissue onto a paper towel and centrifuged at 200 rpm for 1 min.
6. The samples in the plate were then air dried denatured (94°C for 5 min) after adding 10 µl of HiDi Formamide (Applied Biosystems, USA) and then the samples were sequenced on ABI 3700 Genetic Analyzer (Applied Biosystems, USA).

3.2.3 Primer design and synthesis

On sequencing 960 micosatellite enriched clones 1920 sequence reads were obtained from ABI 3700 Genetic Analyzer. These sequences were cured to remove the remnants of vector sequences from both 5' and 3' ends using VecScreen at NCBI (http://www.ncbi.nlm.nih.gov/vecscreen/vecscreen.html). Following vector trimming, a total of 1920 sequence reads thus obtained was allowed to form contigs (alignment of forward and reverse sequences) using DNABaser v2. The CAP3 programme was used to remove the sequence redundancy. As a result, the contigs and singletons thus obtained from CAP3 assembly (Huang and Madan, 1999) were used in the FASTA format in a single file for microsatellite search using MIcroSAtellite (Thiel et al. 2003), a tool for the identification and localization of (i) perfect microsatellites as well as, (ii) compound microsatellites (two individual microsatellites interrupted by up to 100 bases).

Primers were designed from the sequences containing SSR repeat motifs using Primer 3 (v. 0.3.0) (Rozen and Skaletsky 2000), a web based software that picks primers from DNA sequences. Ideal primers for PCR were designed using parameters viz., T_m in the range of 50–65°C, random nucleotide composition, a 40–60% GC-content, and be 18 - 24 bases long. To

avoid formation of hairpin structures, the intra-primer or inter-primer homology was kept as low as possible (<3 bp complementarity within primer) or primer dimers (<3 bp complementarity between primers), which will interfere with annealing of primer to the DNA template.

3.2.4 Optimization and validation of SSR markers

In order to check the amplification of 454 primer pairs, initially the polymerase chain reaction (PCR) was performed on two genotypes (ICMB 841-P3 and 863B-P2) in a 5 µl reaction volume [0.5 µl of 10X PCR buffer (Bioline, U.K.), 1.00 µl of 10 mM Mg^{++} (Bioline, U.K.), 0.25 µl of 2 mM dNTPs, 1.0 µl of 2 pmoles primer (MWG- Biotech AG), 0.1 U (0.2 µl of 0.5 U/µl) *Taq* polymerase (Bioline, U.K.) and 1.0 µl of template DNA (5 ng/µl)] in 96-well microtiter plate. A common touch down PCR profile was adopted for all pearl millet SSR primers: three minutes initial denaturation cycle (95˚C) followed by five cycles for 20 sec at 94˚C, 20 sec at 60˚C (-1˚C per cycle) and 30 sec at 72˚C then by 30 cycles of 94˚C for 20 sec, 20 sec at 56˚C and 30 sec at 72˚C followed by a 20 min final extension.

The primer pairs that failed to produce scorable amplification by using above mentioned PCR conditions were tested for amplification using modified PCR conditions. This time, the PCR reaction mix contained 0.5 µl 10X buffer, 1.0 µl of 10 mM Mg^{++}, 0.375 µl dNTPs, and 1.0 µl of 2 pM primer, 0.20 µl of *Taq* DNA polymerase and 0.5 µl of template DNA. The modified PCR amplification profile included 3 min of initial denaturation (95˚C) followed by five cycles for 20 sec at 94˚C, 20 sec at 60˚C (-1˚C per cycle) and 30 sec at 72˚C then followed by 30 cycles of 94˚C for 20 sec, 20 sec at 50˚C and 72˚C for 30 sec then final extension at 72˚C for a 20 min.

3. 3 Development of Genetic Linkage Map of Pearl Millet

3.3.1 Genotyping of polymorphic markers on the mapping population

In order to construct the genetic linkage map for pearl millet the polymorphic markers from the comparative marker polymorphism studies and 25 novel SSRs developed in the present study were employed. A set of 145 polymorphic markers comprising (50 CISPs, 27 gSSRs, 23_PM_SSCP-SNPs)were genotyped on mapping population ICMB 841- P3 × 863B-P2 with 101 recombinant inbred lines (RILs). The PCR components and the PCR amplification profile conditions were same as the conditions used for parental screening for polymorphism assessment.

3.4 Data Analysis

3.4.1 Scoring of PCR amplicons resolved on PAGE gels

For diversity studies The unlabelled markers (40 gSSRs, intron and exon sequence haplotypes) were scored as continuous variables, using “a” to indicate the first band, “b” the second band and so on.

3.4.2 Scoring the PCR amplicons resolved using Capillary electrophoresis

After completion of CE, files created by ABI machine were processed thorough GENESCAN 3.7 v software. GeneScan assigns the product sizes to all PCR amplicons based on their relative mobility with internal LIZ size standard. After this, assigned product sizes for each marker can be viewed in GENOTYPER 3.7 v software elecropherogram. Output files from ABI3130 machine were processed through GeneScan and Genotyper softwares. But data was scored as allele sizes (raw alleles) directly from Genotyper software. Raw data was processed though AlleloBin program.

3.4.3 Allele call

Raw allele calls derived directly from Genotyper software were processed through AlleloBin program. AlleloBin program uses repeat motif as a reference to call the perfect allele. All the markers screened against reference set were processed through AlleloBin to get perfect allele calls. Processed allele calls for each the markers were combined for further analysis.

3.4.4 Heterozygosity

Heterozygosity is simply the proportion of heterozygous individuals in the population. At a single locus it is estimated as:

$$\hat{H}_l = 1 - \sum_{i=1}^{k} P_i$$

where

P_i = i th allele frequency

3.4.5 Polymorphism Information Content

A closely related diversity measure is the polymorphism information content (PIC) (Botstein *et al.* 1980). It is estimated as:

$$PIC = 1 - \left[\sum_{i=1} Pi^2 \right] - \left[\sum_{i-1}^{n-1} \sum_{j=i+1}^{n} 2Pi^2 Pj^2 \right]$$

where P_i and P_j are the frequencies of i^{th} and j^{th} allele

3.4.6 Diversity analysis

3.4.6a Cluster analysis

Cluster analysis was carried out using DARwin 5.0 software (Perrier et al. 2003). Un-weighted neighbor joining (NJ) method was used to generate the dendogram using DARwin program

using the dissimilarity matrix to determine the aggregation of the accessions into clusters. Un-weighted neighbor joining method gives a same unitary weight given to all units. An identifier file was prepared consisting the all the genotypes, and other related information. This was used to identify the individuals in all the steps.

3.4.7 Linkage map construction

The segregation data for all polymorphic markers was scored as ‘A’, ‘B’, ‘H’ and ‘-’.

‘A’ – Homozygous allele resembles the allele present in recurrent parent (ICMB841-P3)

‘B’ – Homozygous allele resembles the allele present in donor parent (863B- P2)

‘H ’– Heterozygous or presence of both alleles (recurrent and donor parents)

‘-’ – Represent the missing (no amplification)

The genetic linkage map was constructed with MAPMAKER/EXP Version 3.0 (Lincoln et al. 1992). Loci were grouped using the criteria LOD = 3.0 and recombination fraction = 0.15. Linkage groups belonging to the same chromosomes were integrated using the criteria LOD = 3.0 and recombination fraction = 0.40. Multipoint analysis was used to construct the framework map. The order of “Framework” loci was determined using the “order” command with a LOD threshold of 3.0 and checked using the “ripple” command. Other markers were placed on the framework map using the “try” and “ripple” commands. The maps were drawn with MAPMAKER Version 2.0 (Lander et al. 1987), with the recombination distances calculated with the Kosambi mapping function (Kosambi 1944). Double crossovers between closely linked markers were checked to avoid misscoring. The segregation Wt of each locus to a 1:1 ratio was examined using the chisquare test.

4. RESULTS

4.1 Comparative Marker Polymorphism Assessment

In order to assess the relative informativeness and level of polymorphism of different PCR-based marker systems, a total of 627 primer pairs (comprising 100 pearl millet gSSRs; 60 pearl millet EST- SSRs, eSSRs; 100 pearl millet SSCP-SNPs; 57 wheat SSCP-SNPs, and 310 CISPs) were screened on a panel of 22 inbred lines which represent parents of eleven mapping populations available at ICRISAT (Table 3.1).

4.1.1 Amplification status of marker systems

The PCR amplicons resoved on 6% PAGE gels for CISP markers are shown in Fig 4.1 and amplification profiles of SSR, PM_SSCP-SNP, W_SSCP-SNP and CISP markers are shown in Fig 4.2. As SNPs are well-defined as a single base difference in sequence and the polymorphisms observed in the current study are both due to sequence variation and indels, based on origin of polymorphism, we have chosen to call CISPs and SSCP-SNPs as "intron sequence haplotypes" (ISH) and the wheat SSCP-SNPs as “exon sequence haplotypes” (ESH). Therefore data generated by CISP and SSCP-SNP primer pairs were analyzed together under ISH class. In total 455 (72.56%) primer pairs produced scorable amplicons (Table 4.1). Of 100 gSSRs tested, scorable amplification was observed in case of 82 (82%) primer pairs. Similarly the percent amplification in case of eSSRs, ISHs and ESHs was 75, 73.9 and 43.8 respectively. Among different PCR based marker systems tested highest success rate of amplification was observed in case of gSSRs.

Table 4.1: Summary on amplification status and marker informativeness attributes

Marker attributes	Marker systems			
	gSSRs[a]	eSSRs[b]	ISHs[c]	ESHs[d]
Total markers screened	100	60	410	57
Number of markers amplified	82	45	303	25
Polymorphic with one or the other populations	45	44	164	14
Percent marker polymorphism	45.0	73.3	73.9	24.6
Alleles range	2-14	2-8	2-4	2-4
Average number of alleles	6.06	3.09	2.60	3.10
Total number of alleles	273	136	113	44
Genetic similarity range	0.23-0.82	0.46-0.92	0.58-0.84	0.33-0.97
Average genetic similarity	0.31	0.64	0.72	0.62
PIC value range	0.08-0.91	0.08-0.84	0.09-0.50	0.17-0.49
Average PIC value	0.62	0.36	0.39	0.35
Observed heterozygosity range	0.08-0.91	0.09-0.83	0.08-0.83	0.11-0.70
Average heterozygosity	0.67	0.40	0.49	0.43
Range of effective number of alleles	1.10-12.02	1.20-11.55	1.21-13.02	1.42-9.03
Average effective number of alleles	2.11	4.11	2.86	3.08

[a]gSSRs, genomic SSRs; [b]eSSRs, EST-SSRs; [c]ISH, intron sequence haplotypes; [d]ESH, exon sequence haplotypes

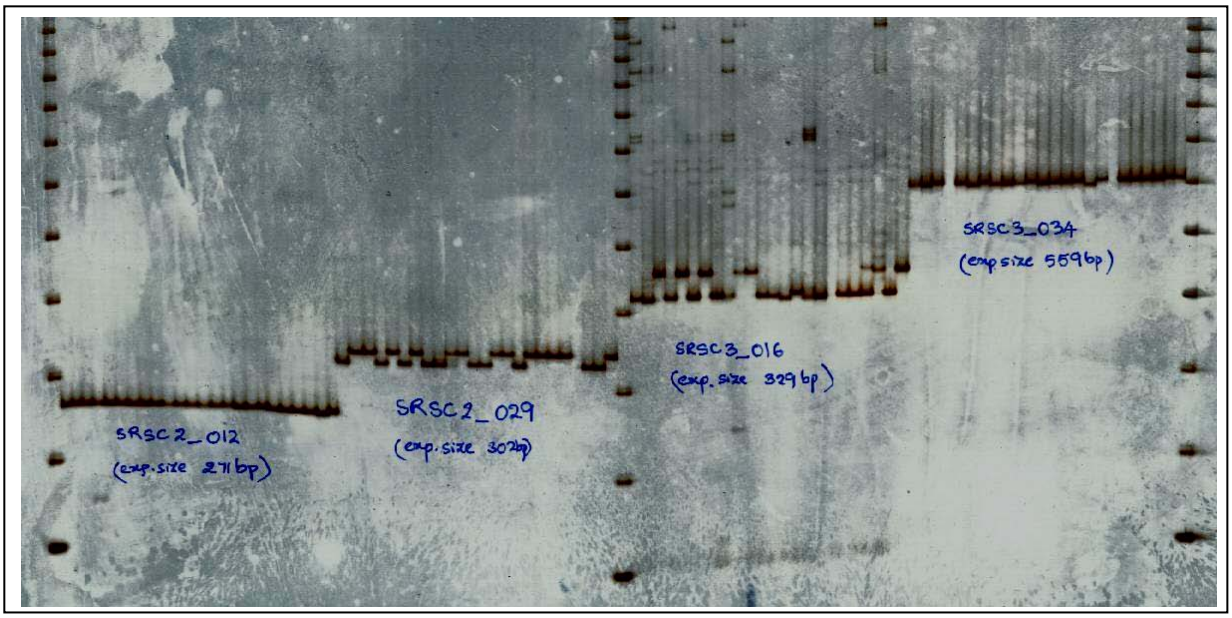


Fig 4.1: Four CISP markers resolved on 6% PAGE. The marker names and expected size of the PCR products are indicated on the gel

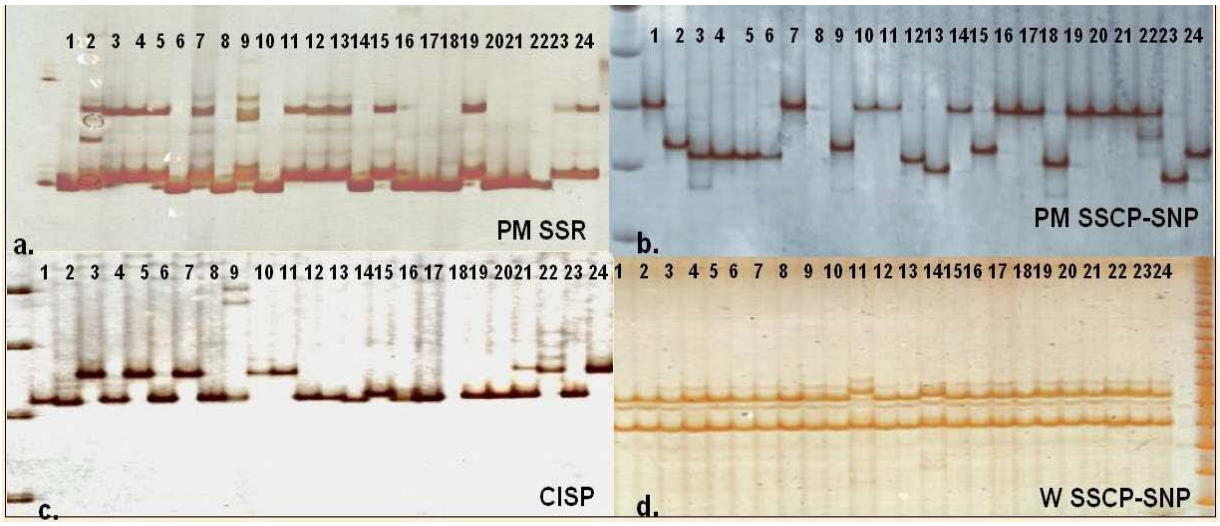


Figure 4.2: Amplification profiles of all markers systems; 22 pearl millet genotypes along with two control genotyopes are resolved on 6% PAGE gels

4.1.2 Marker informativeness

The informativeness of a marker system is determined by number of parameters for instance number of alleles, PIC values, heterozygosity, effective number of alleles etc. The observed, effective number of alleles, PCR product size range, observed heterozygosity, and PIC value of each marker investigated is presented in Table 4.1.

4.1.2.a *Number of alleles*

The number of alleles observed for each marker is presented in Table 4.1 and the amplicon size rage of each marker is presented in Table 4.2. The total number of alleles ranged from 44 (ESHs) to 436 (ISHs). The ranges of alleles detected by each marker system were 2-14, 2-8, 2-4, and 2-4, for the informative gSSRs, eSSRs, ISHs, and ESHs respectively (Table 4.1). The average number of alleles per polymorphic marker ranged from 2.6 (ISH) to 6.06 (gSSRs).

4.1.2.b *Genetic dissimilarity detected by marker systems*

A wide-range of genetic dissimilarities were revealed by marker systems among the inbred lines examined (Table 4.1). The range of genetic dissimilarity detected by gSSRs, eSSRs, ISHs and ESHs were 0.23-0.82, 0.46-0.92, 0.58-0.84 and 0.33-0.97 respectively. The average genetic dissimilarity observed among marker systems was highest among the ISH (0.72) followed by eSSRs (0.64), ESH (0.62) and gSSRs (0.31) (Table 4.1). All marker systems except CISP could reveal a broad range of genetic similarities among the inbred lines.

4.1.2.c *Polymorphism Information Content (PIC)*

The polymorphism information content (PIC) is another important measure of DNA polymorphism besides being a measure of genetic variation; it is also used in the context of

gene mapping. The values of PIC are lower than heterozygosity for the corresponding marker because in PIC, a quantity is subtracted from heterozygosity that corresponds to the probability of offspring being uninformative. The average and range of PIC values per each marker system are presented in Table 4.1. PIC value of each marker is given in Table 4.2. The gSSRs were associated with the highest average PIC value (0.62), followed by the ISH (0.39) eSSRs (0.36) and ESH (0.35) (Table 4.1). The PIC values ranged from 0.08 to 0.91 in case of gSSRs and the PIC values for eSSRs, ISH and were in the range of 0.08-0.84, 0.09-0.5 and 0.17-049 respectively. The differences in distribution of number of markers under different PIC classes can be visualized in Fig 4.3.

4.1.2.d Heterozygosity

The observed heterozygosity is based on the number of heterozygous individuals in the population under investigation. The observed heterozygosity at individual marker loci for all the markers is presented in Table 2. In the present study the overall mean for the observed heterozygosity was 0.40, 0.43, 0.49 and 0.67 for eSSRs, ESH, ISH and gSSRs respectively. However, the marker systems diplayed a broad range of observed heterozygosites for instance gSSRs recorded the broad range 0.08-0.91 followed ISH (0.09-0.83) eSSRs (0.08-0.83) and ESH (0.11-0.70).

4.1.2.e Effective number of alleles

The effective number of alleles (n_e) is an estimate of the number of alleles with equal frequencies corresponding to a particular PIC value. The number of equally frequent alleles that would create the same heterozygosity as observed in the population. The number of alleles at individual marker loci for all the markers is presented in Table 2. A lower average effective number of alleles per locus (2.11) were observed in case of gSSRs (Table 4.1) and a high

average effective number of alleles were observed for eSSRs (4.11). Across all gSSR loci studied the effective number of alleles ranged from 1.10-12.02. Similarly the range of effective number of alleles in case of eSSRs, ISHs and ESHs were 1.20-11.55, 1.21-13.02, and 1.42-9.03. Although ISH recorded highest effective number of alleles the average effective number alleles were high in case of eSSRs.

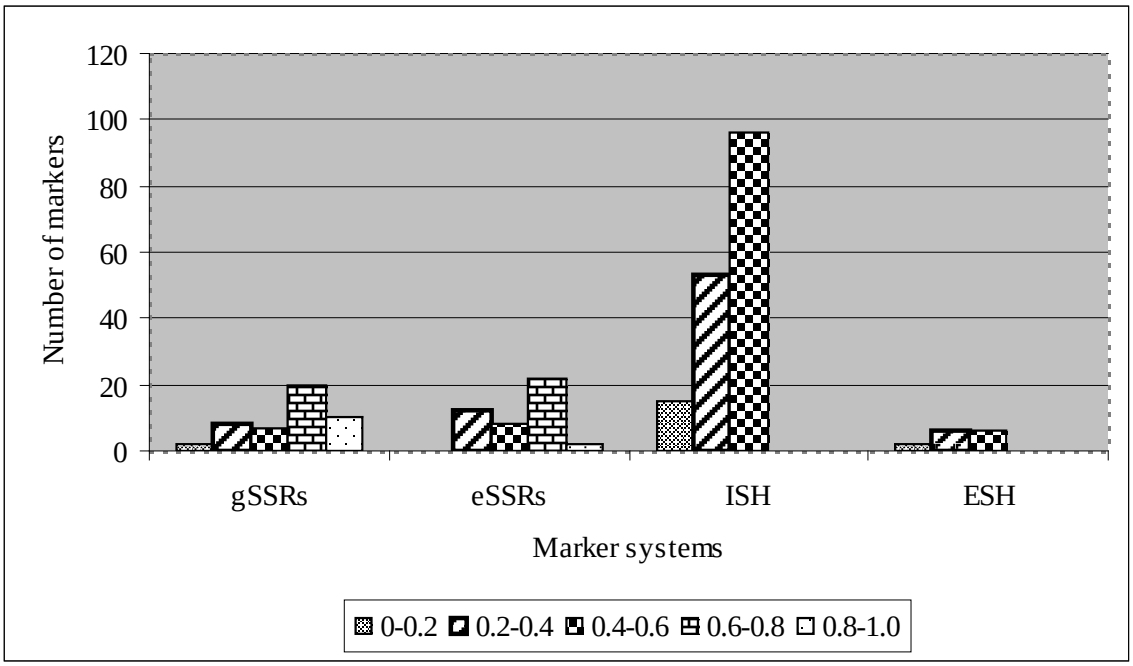


Figure 4.3: Distribution of polymorphism information content (PIC) data for different pearl millet crosses. The data was obtained using genomic simple sequence repeats (SSR), EST-SSRs (eSSRs), intron sequence haplotypes (ISH) and exon sequence haplotypes (ESH) markers.

4.1.3 Marker polymorphism

Of 627 primer pairs tested, a total of 268 (59%) markers showed polymorphism with at least one parental genotype combination (Table 4.3). On an average 116 (25.49%) markers were polymorphic with 11 mapping population parental genotype combinations that were tested. The number of polymorphic markers ranged from 103 (Tift23D_2B_1-P5 × WSIL-P8) to 122 (ICMB 841-P3 × 863B-P2) (26.81%) among the tested mapping population parents. The world reference mapping population 81B-P6 × ICMP451-P8 recorded a maximum of 118 (25.93%) polymorphic markers. Among eleven mapping population parental pairs, ICMB 841-P3 × 863-P2 recorded maximum number of polymorphic markers. The percent polymorphism among 11 mapping populations ranged 22.63-26.81% (Table 3).

4.1.4 Informative and common polymorphic markers

The number of informative and common polymorphic markers between populations varied with the marker system and between populations (Table 4. 4). High number of common and informative polymorphic markers was observed in case of ISH followed by gSSRs between populations. The pair wise common informative markers were also computed. A set of 8-11 common pair wise polymorphic markers were observed in case of genopmic SSRs and 3-6 eSSRs among the mapping populations studied. However, ISH markers exhibited highest pairwise common polymorphic markers (20-56) and ESH were found to show least common polymorphic markers.

Table 4.2: Detailed polymorphism data for all the polymorphic markers assayed

Marker[e]	Size range (bp)	Alleles	H_o	n_e	PIC
gSSRs					
Xpsmp2001	200-215	6	0.53	1.89	0.56
Xpsmp2017	194-220	11	0.84	3.50	0.85
Xpsmp2018	203-210	2	0.79	1.26	0.35
Xpsmp2027	193-220	9	0.82	1.22	0.8
Xpsmp2033	238-245	2	0.08	12.02	0.09
Xpsmp2038	365-380	7	0.70	2.50	0.79
Xpsmp2040	200-215	6	0.78	1.29	0.71
Xpsmp2043	157-170	4	0.68	1.47	0.69
Xpsmp2048	178-190	9	0.86	1.16	0.86
Xpsmp2063	112-125	6	0.79	1.27	0.59
Xpsmp2066	195-200	3	0.90	1.11	0.24
Xpsmp2069	315-330	10	0.89	1.12	0.90
Xpsmp2070	207-220	9	0.83	1.21	0.80
Xpsmp2072	209-225	11	0.84	1.19	0.80
Xpsmp2074	146-155	6	0.67	1.48	0.60
Xpsmp2076	146-160	7	0.70	1.42	0.70
Xpsmp2077	135-145	5	0.42	2.40	0.40
Xpsmp2078	135-150	14	0.91	1.10	0.90
Xpsmp2079	137-140	3	0.47	2.12	0.40
Xpsmp2080	193-205	7	0.78	1.29	0.70
Xpsmp2081	173-180	9	0.84	1.19	0.80
Xpsmp2084	190-200	13	0.84	1.19	0.80
Xpsmp2085	206-210	4	0.69	1.45	0.60
Xpsmp2087	170-175	4	0.63	1.58	0.60
Xpsmp2088	120-134	6	0.77	1.30	0.70
Xpsmp2089	135-150	8	0.85	1.18	0.80
Xpsmp2201	180-195	6	0.70	1.44	0.70
Xpsmp2203	184-198	8	0.79	1.26	0.80
Xpsmp2204	130-150	4	0.68	1.48	0.60
Xpsmp2206	352-360	3	0.36	2.76	0.30
Xpsmp2213	256-260	2	0.15	6.55	0.10
Xpsmp2214	194-200	4	0.64	1.56	0.60
Xpsmp2219	194-206	2	0.49	2.06	0.40
Xpsmp2220	127-130	5	0.74	1.36	0.70
Xpsmp2221	205-210	4	0.72	1.39	0.70
Xpsmp2222	360-370	2	0.50	2.00	0.40
Xpsmp2225	243-250	3	0.66	1.52	0.60
Xpsmp2227	216-220	3	0.34	2.91	0.30
Xpsmp2231	193-200	6	0.74	1.35	0.70

Marker[e]	Size range (bp)	Alleles	H_o	n_e	PIC
Xpsmp2233	243-250	7	0.79	1.27	0.80
Xpsmp2237	255-260	3	0.61	1.64	0.50
Xpsmp2255	229-232	3	0.51	1.96	0.40
Xpsmp2263	142-145	12	0.84	1.19	0.80
Xpsmp2273	217-220	7	0.82	1.23	0.80
Xpsmp2275	251-270	8	0.81	1.24	0.80
Expressed sequence tag -SSRs					
Xicmp3001	195-200	2	0.35	2.85	0.35
Xicmp3002	200-230	2	0.49	2.04	0.49
Xicmp3004	200-210	3	0.63	1.58	0.57
Xicmp3005	185-190	2	0.24	4.25	0.24
Xicmp3013	210-250	2	0.45	2.24	0.48
Xicmp3014	220-230	2	0.09	11.52	0.09
Xicmp3017	180-185	3	0.58	1.72	0.24
Xicmp3024	190-200	4	0.68	1.46	0.72
Xicmp3025	200-205	2	0.09	11.52	0.09
Xicmp3027	210-220	4	0.71	1.41	0.71
Xicmp3028	160-165	2	0.09	11.52	0.09
Xicmp3029	220-230	2	0.36	2.76	0.36
Xicmp3032	190-200	3	0.58	1.72	0.58
Xicmp3037	100-102	2	0.09	11.55	0.09
Xicmp3038	95-97	2	0.09	11.00	0.09
Xicmp3039	790-800	3	0.63	1.58	0.63
Xicmp3042	580-585	2	0.10	9.53	0.10
Xicmp3043	200-210	2	0.39	2.54	0.24
Xicmp3045	310-320	2	0.17	6.05	0.17
Xicmp3048	250-260	2	0.10	10.53	0.10
Xicmp3049	180-182	2	0.09	11.52	0.09
Xicmp3050	220-250	4	0.59	1.71	0.59
Xicmp3056	160-170	2	0.17	6.05	0.17
Xicmp3057	750-755	2	0.42	2.38	0.42
Xicmp3058	165-175	2	0.17	5.80	0.17
Xicmp3063	170-180	3	0.63	1.58	0.38
Xicmp3066	140-200	5	0.61	1.65	0.61
Xicmp3077	330-370	3	0.32	3.11	0.32
Xicmp3078	240-250	2	0.42	2.38	0.38
Xicmp3079.1	790-800	4	0.36	2.78	0.36
Xicmp3079.2	205-210	2	0.40	2.52	0.40
Xicmp3080	500-540	5	0.68	1.47	0.70
Xicmp3081	180-210	4	0.64	1.55	0.64
Xicmp3085	190-200	3	0.54	1.85	0.54

Marker[e]	Size range (bp)	Alleles	H_o	n_e	PIC
Xicmp3086	140-150	3	0.53	1.90	0.53
Xicmp3088	130-160	8	0.83	1.20	0.83
Xicmp3091.1	530-550	2	0.17	5.80	0.17
Xicmp3091.2	180-185	3	0.38	2.63	0.38
Xicmp3092	225-230	4	0.70	1.43	0.64
Xicmp3093	175-190	4	0.62	1.63	0.62
Xicmp3096	210-220	3	0.29	3.45	0.29
Xicmp4010.1	375-380	4	0.38	2.60	0.38
Xicmp4010.2	290-300	4	0.49	2.03	0.49
Xicmp4014	210-220	4	0.43	2.33	0.58
Intron sequence haplotypes					
Xpsms103	470-475	4	0.29	3.48	0.32
Xpsms13	657-660	3	0.5	2.01	0.48
Xpsms14	880-890	4	0.54	1.84	0.50
Xpsms15	875-890	2	0.71	1.41	0.30
Xpsms16	312-320	2	0.60	1.68	0.09
Xpsms17	302-320	2	0.80	1.24	0.43
Xpsms2	780-785	2	0.50	2.00	0.09
Xpsms21	375-380	2	0.32	3.13	0.43
Xpsms22	380-385	2	0.46	2.18	0.24
Xpsms23	755-760	3	0.34	2.94	0.49
Xpsms24	255-260	2	0.50	2.00	0.30
Xpsms26	420-430	2	0.54	1.85	0.48
Xpsms34	330-340	2	0.30	3.36	0.30
Xpsms35	315-320	2	0.38	2.67	0.46
Xpsms4	1075-1080	4	0.50	2.00	0.44
Xpsms40	590-600	2	0.55	1.82	0.30
Xpsms41	490-495	2	0.53	1.88	0.40
Xpsms44	1600-1620	2	0.69	1.44	0.24
Xpsms47	495-500	2	0.30	3.36	0.35
Xpsms48	395-400	2	0.72	1.40	0.40
Xpsms49	250-255	3	0.23	4.41	0.43
Xpsms52	910-915	4	0.51	1.95	0.46
Xpsms57	415-420	2	0.39	2.59	0.09
Xpsms60	550-560	3	0.60	1.67	0.42
Xpsms61	575-585	2	0.65	1.53	0.17
Xpsms64	220-225	3	0.08	13.02	0.48
Xpsms65	265-270	4	0.56	1.79	0.50
Xpsms66	1800-1810	3	0.15	6.55	0.24
Xpsms68	375-380	3	0.66	1.51	0.54
Xpsms69	506-515	2	0.69	1.46	0.46

Marker[e]	Size range (bp)	Alleles	H_o	n_e	PIC
Xpsms70	580-585	3	0.16	6.15	0.48
Xpsms73	335-350	2	0.49	2.04	0.50
Xpsms74	1150-1160	2	0.62	1.61	0.30
Xpsms75	420-440	2	0.76	1.32	0.40
Xpsms76	410-415	4	0.49	2.03	0.44
Xpsms78	1000-1050	2	0.56	1.80	0.35
Xpsms80	970-980	4	0.48	2.07	0.45
Xpsms82	1400-1420	2	0.76	1.31	0.30
Xpsms83	815-820	4	0.56	1.80	0.49
Xpsms84	510-520	3	0.75	1.34	0.43
Xpsms86	900-910	2	0.51	1.95	0.50
Xpsms88	900-910	2	046	1.59	0.30
Xpsms89	980-1000	2	0.77	1.29	0.30
Xpsms90	650-655	2	0.50	2.02	0.40
PRSC1_003	650-655	3	0.08	12.52	0.31
PRSC1_004	290-300	2	0.08	12.52	0.09
PRSC1_005	265-257	2	0.08	12.52	0.30
PRSC1_006	235-240	2	0.08	12.52	0.17
PRSC1_010	185-190	2	0.08	12.52	0.30
PRSC1_014	1000-1030	4	0.08	12.52	0.48
PRSC1_023	1600-1650	2	0.08	12.52	0.24
PRSC1_024	900-910	3	0.10	10.03	0.48
PRSC1_031	1300-1350	4	0.15	6.55	0.48
PRSC1_035	900-940	3	0.15	6.55	0.48
PRSC1_039	1100-1150	3	0.15	6.55	0.37
PRSC1_041	950-1000	4	0.15	6.55	0.49
PRSC1_043	1350-1400	2	0.15	6.55	0.43
PRSC1_044	587-595	2	0.16	6.30	0.40
PRSC1_045	1000-1010	2	0.17	6.05	0.46
PRSC1_047	390-400	3	0.22	4.60	0.46
PRSC1_049	405-425	4	0.23	4.41	0.38
PRSC10_070	1300-1310	2	0.23	4.27	0.30
PRSC4_050	650-655	2	0.24	4.08	0.35
PRSC4_053	720-900	4	0.28	3.60	0.50
PRSC4_060	800-805	3	0.28	3.60	0.48
PRSC4_063	400-420	2	0.28	3.60	0.30
SRSC1_003	190-210	2	0.28	3.51	0.43
SRSC1_004	283-290	2	0.29	3.47	0.50
SRSC1_005	650-660	3	0.30	3.36	0.67
SRSC1_007	515-520	2	0.32	3.10	0.17
SRSC1_008	1000-1010	2	0.33	3.03	0.35

Marker[e]	Size range (bp)	Alleles	H_o	n_e	PIC
SRSC10_001	700-710	3	0.33	3.03	0.24
SRSC10_018	510-520	2	0.34	2.94	0.09
SRSC11_005	895-900	2	0.35	2.85	0.48
SRSC11_007	530-535	2	0.36	2.82	0.09
SRSC11_008	540-710	2	0.36	2.81	0.50
SRSC11_009	720-725	3	0.36	2.75	0.48
SRSC11_011	397-405	3	0.38	2.67	0.46
SRSC11_014	600-605	2	0.38	2.67	0.30
SRSC11_016	225-240	2	0.39	2.55	0.46
SRSC11_017	700-705	2	0.42	2.36	0.24
SRSC11_018	1600-1605	2	0.42	2.36	0.35
SRSC11_023	350-355	3	0.42	2.36	0.46
SRSC11_024	1100-1120	2	0.42	2.36	0.40
SRSC12_001	598-605	2	0.43	2.30	0.50
SRSC12_002	990-1000	3	0.44	2.28	0.48
SRSC12_003	720-725	3	0.44	2.25	0.48
SRSC12_019	398-405	2	0.44	2.25	0.09
SRSC2_002	820-900	2	0.45	2.23	0.46
SRSC2_004	300-310	4	0.46	2.16	0.50
SRSC2_005	505-520	3	0.46	2.16	0.49
SRSC2_007	1080-1100	2	0.47	2.13	0.17
SRSC2_010	495-503	3	0.48	2.08	0.48
SRSC2_012	275-300	2	0.48	2.08	0.30
SRSC2_015	410-415	4	0.48	2.07	0.48
SRSC2_017	590-620	3	0.49	2.06	0.46
SRSC2_018	400-405	4	0.49	2.06	0.48
SRSC2_020	610-700	4	0.49	2.06	0.48
SRSC2_023	520-525	2	0.49	2.04	0.50
SRSC2_026	455-460	3	0.49	2.04	0.37
SRSC2_028	300-305	2	0.49	2.03	0.50
SRSC2_029	320-325	3	0.49	2.02	0.49
SRSC2_030	400-415	2	0.49	2.02	0.50
SRSC2_032	450-455	3	0.50	2.02	0.38
SRSC2_037	300-305	3	0.50	2.01	0.48
SRSC2_038	600-610	4	0.50	2.01	0.49
SRSC2_039	320-325	2	0.50	2.01	0.35
SRSC2_040	1500-1510	4	0.50	2.01	0.49
SRSC3_001	1000-1010	2	0.50	2.00	0.43
SRSC3_002	1100-1110	3	0.50	2.00	0.48
SRSC3_003	240-250	2	0.50	2.00	0.30
SRSC3_004	580-585	3	0.50	2.00	0.37

Marker[e]	Size range (bp)	Alleles	H_o	n_e	PIC
SRSC3_005	390-395	4	0.5	2.00	0.49
SRSC3_008	1300-1320	2	0.5	1.98	0.40
SRSC3_011	870-880	3	0.54	1.86	0.48
SRSC3_015	1600-1700	2	0.54	1.86	0.50
SRSC3_016	420-470	3	0.54	1.85	0.46
SRSC3_019	360-365	2	0.56	1.80	0.43
SRSC3_023	990–890	2	0.56	1.79	0.24
SRSC3_032	715-720	3	0.57	1.77	0.31
SRSC3_034	605-610	3	0.59	1.68	0.49
SRSC3_035	280-290	2	0.60	1.67	0.50
SRSC3_036	720-725	4	0.60	1.66	0.48
SRSC3_038	600-610	4	0.61	1.65	0.44
SRSC4_003	320-350	3	0.61	1.64	0.46
SRSC4_006	300-305	3	0.62	1.62	0.48
SRSC4_008	650-670	2	0.62	1.62	0.09
SRSC4_011	490-495	2	0.63	1.60	0.50
SRSC4_012	412-420	3	0.63	1.59	0.31
SRSC4_013	280-300	2	0.64	1.57	0.17
SRSC4_014	700-710	3	0.64	1.57	0.48
SRSC4_018	395-400	3	0.64	1.55	0.31
SRSC4_019	495-510	3	0.65	1.54	0.24
SRSC4_020	240-250	3	0.66	1.52	0.48
SRSC4_021	535-545	2	0.66	1.53	0.17
SRSC4_023	1100-1110	3	0.66	1.52	0.46
SRSC4_028	520-525	3	0.66	1.52	0.50
SRSC4_031	630-640	3	0.66	1.51	0.42
SRSC5_001	310-340	2	0.66	1.51	0.24
SRSC5_002	1230-1245	2	0.67	1.49	0.46
SRSC5_005	498-520	3	0.68	1.48	0.46
SRSC5_007	450-460	3	0.68	1.48	0.49
SRSC5_010	295-300	2	0.68	1.46	0.30
SRSC5_016	1090-1100	3	0.68	1.46	0.24
SRSC5_017	580-595	3	0.69	1.45	0.24
SRSC6_001	1920-1940	2	0.69	1.45	0.43
SRSC6_004	800-900	4	0.69	1.45	0.48
SRSC6_008	307-320	2	0.70	1.44	0.48
SRSC6_010	905-910	3	0.70	1.42	0.46
SRSC6_011	330-340	2	0.70	1.42	0.48
SRSC6_020	585-600	2	0.71	1.41	0.24
SRSC6_022	570-585	3	0.72	1.39	0.48
SRSC7_001	480-490	3	0.72	1.39	0.43

Marker[e]	Size range (bp)	Alleles	H_o	n_e	PIC
SRSC7_002	545-560	3	0.72	1.38	0.48
SRSC7_003	800-810	2	0.72	1.38	0.30
SRSC7_008	1400-1425	3	0.73	1.37	0.49
SRSC7_009	1070-1120	2	0.74	1.35	0.17
SRSC7_012	905-910	2	0.75	1.33	0.48
SRSC7_020	370-400	3	0.77	1.30	0.37
SRSC8_002	620-690	4	0.78	1.29	0.49
SRSC8_008	500-510	3	0.80	1.25	0.49
SRSC8_009	695-615	3	0.81	1.23	0.49
SRSC8_014	695-710	3	0.82	1.21	0.46
SRSC9_010	1200-1250	4	0.83	1.21	0.48
Exon Sequence Haplotypes					
BE406148	950-960	3	0.54	1.86	0.37
BE406808	1120-1125	3	0.16	6.40	0.17
BE426364	1240-1250	2	0.49	2.06	0.48
BE445506	880-900	2	0.41	2.42	0.40
BE445693	1300-1320	3	0.24	4.08	0.43
BE495400	1090-1100	4	0.70	1.42	0.48
BE496834	1300-1320	3	0.52	1.93	0.43
BE499478	450-455	2	0.28	3.60	0.30
BE604737	390-400	4	0.59	1.70	0.48
BF145484	700-710	3	0.35	2.85	0.43
BF201129	1250-1260	3	0.44	2.25	0.49
BF202681	1400-1410	4	0.56	1.78	0.44
BF473379	970-980	4	0.57	1.75	0.32
BM138439	1000-1025	4	0.11	9.03	0.49

[e]Xicmp = pearl millet eSSRs; Xpsmp = pearl millet gSSRs; Xpsms = rice-pearl millet SSCP-SNPs; SRSC = rice-sorghum CISPs; PRSC = rice-*Pennisetum* CISPs; and BE, BF and BM = wheat SSCP-SNPs; H_o=Observed hterozygosity; n_e=effective number of alleles, PIC=Polymorphism information content

Table 4.3: Marker polymorphism in parental combinations of eleven pearl millet mapping populations

	gSSRs[a]	eSSRs[b]	ISHs[c]	ESHs[d]	Total polymorphic markers	% polymorphism
H 77/833-2 × PRLT 2/89-33	30	20	59	3	112	24.61
ICMB 841-P3 × 863B-P2	27	18	72	5	120	26.81
Tift $23D_2B_1$-P5 × WSIL-P8	31	16	54	2	103	22.63
PT 732B-P2 × P1449-2-P1	30	18	65	3	116	25.49
LGD 1-B-10 × ICMP 85410-P7	39	14	62	3	118	25.93
81B-P6 × ICMP 451-P8	35	15	63	5	118	25.93
ICMP 451-P6 × H 77/833-2-P5(NT)	33	15	67	3	118	25.93
W 504-1-P1 × P 310-17-Bk	39	11	67	2	109	23.95
IP 18293-P152 × Tift $238D_1$-P158	37	19	58	2	116	25.49
ICMB 89111-P6 × ICMB 90111-P6	36	23	55	2	116	25.49
IPC 804 × 81B	40	17	57	4	118	25.93

gSSRs, genomic SSRs; [b]eSSRs, EST-SSRs; [c]ISHs, intron sequence haplotypes; [d]ESHs, exon sequence haplotypes

Table 4.4: Number of informative and common polymorphic markers in the present study across eleven mapping populations

	MP1[f]	2MP2	MP3	MP4	MP5	MP6	MP7	MP8	MP9	MP10	MP11
[a]gSSRs	30	20	20	17	15	12	9	7	5	3	1
[b]eSSRs	21	8	5	3	1	1	1	1	1	1	1
[c]ISHs	58	31	25	20	12	12	9	7	4	2	2
[d]ESHs	3	1	1	1	1	1	1	1	0	0	0

[a]gSSRs, genomic SSRs; [b]eSSRs, EST-SSRs; [c]ISHs, intron sequence haplotypes; [d]ESHs, exon sequence haplotypes; [f]MP= mapping populations

[f] the parental genotypes of the mapping populations are present in Table 4.3. The common polymorphic Markers are calculated between population 1 and 2 are designated as MP2, similarly MP3 indicates common polymorphic markers between population 1, 2 and 3.

4.1.4 Cluster analysis

Neighbour joining (NJ) trees for individual marker systems were constructed and presented in Appendix 1. The grouping pattern of the inbred lines using different marker systems differed. As the marker data on all the inbred lines examined were found to show high degree of correlation a NJ based phylogeny was constructed based on the combined genotypic data set (Fig 4.2). The 22 inbred lines clustered into three groups, I, II and III with seven, nine and six inbred lines respectively. Of seven inbreds grouped in cluster I, four inbreds were derived from land races ex. PRLT 2/89-33, 863B-P2, LGD 1-B-10 (all three inbreds were derived from the *Iniadi* landrace from Togo and Ghana) and H 77/833-2 (derived from a Rajasthani landrace population) (Fig 4.2). However, the implied genetic relationship between H 77/833-2 and the three *Iniadi*-derived inbreds was unexpected and remains unexplained. Among other three inbreds grouped in cluster I, ICMB 841-P3 is associated with stable resistance to downy mildew and both PT 732B- P2, Tift23D_2B_1-P5, carry the d_2 dwarfing gene. Among nine members of cluster II, three were downy mildew resistant inbred lines (ICMB 90111-P6, 81B-P8 and P310-17-Bk) two were derived from LCSN 72-1-2-1-1 (ICMP 451-P6, ICMP 451-P8), two were susceptible to downy mildew (H 77/833-2-P5(NT) and W504-1-P1) IPC 804 (triple-restorer of the A_1, A_4 and A_{egp} CMS systems, with profuse pollen producing capacity) and ICMB 89111-P6 (maintainer of the A_1 cytoplasmic-genetic male-sterility system), Cluster III comprised of three downy mildew resistant inbred lines (81B-P6, IP 18293 –P152 and P 1449-2-P1), and three inbreds that possess d_1/d_2 dwarfing gene (ICMP 85410 –P7, WSIL- P8, Tift 238D_1 –P158).

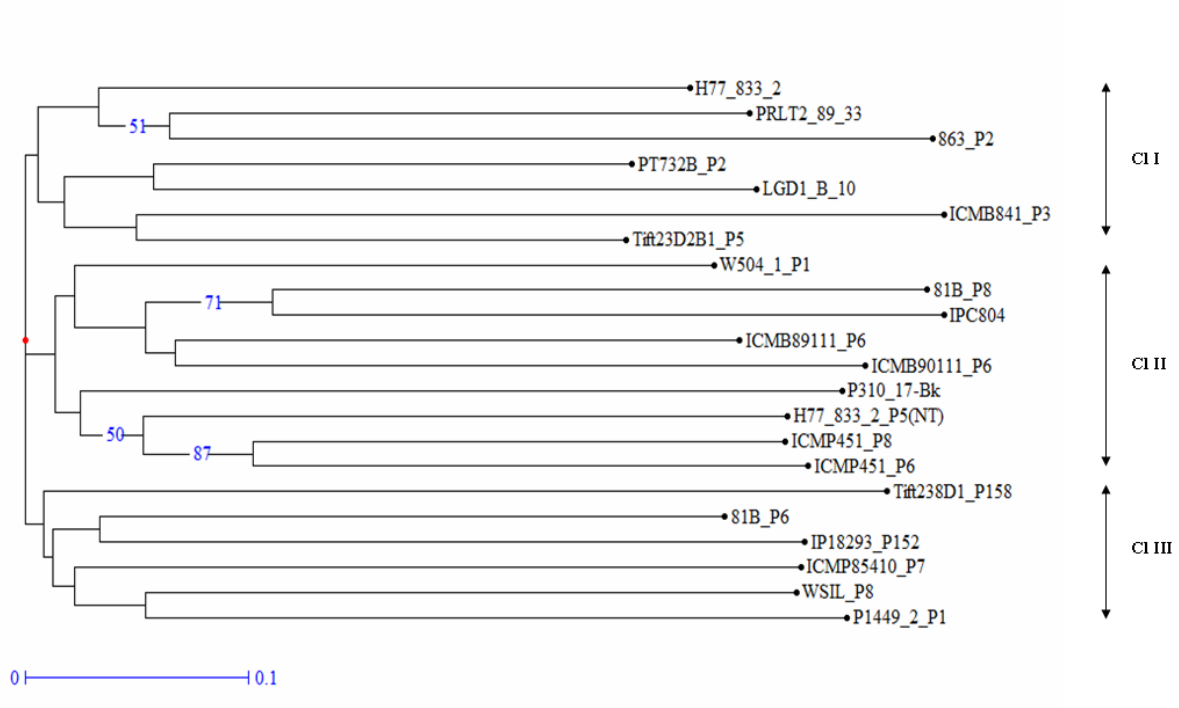


Figure 4.4: Neighbour joining dendrogram constructed from profile data generated using all marker systems (eSSRs, gSSRs, ISHs and ESHs)

4.1.5 Correspondence between marker systems

The correlations are given in Table 5. Mantel's tests indicated the best correspondence between ISH and eSSR (0.98), followed by eSSR and ESH, ISH and ESH (0.97).

Table 4.5: Cophenetic correlation coefficients among marker systems

	Marker systems			
Marker systems	gSSRs*	eSSRs†	ISHs	ESHs
gSSRs	1			
eSSRs	0.92	1		
ISH	0.89	0.98	1	
ESH	0.90	0.97	0.97	1

4.2 Development of Novel Microsatellite Markers

4.2.1 Construction of library, cloning and sequencing of microsatellites

Microsatellite enriched gDNA library was constructed from the pearl millet genotype "Tift $23D_2B_1$-P5". In the first instance, gDNA library was picked in ten 96-well plates. In total a set of 1152 putative positive clones were picked from the microsatellite library enriched for CA, GA, CAA and AGA repeats. However plasmid DNA was successfully isolated from 960 clones. On sequencing 960 clones from both (5' and 3') directions using M13F-pUC (-40) and M13R-pUC (-40), a total of 1920 sequence reads obtained were allowed to form contigs (alignment of forward and reverse sequences) using DNABaser v2 (DNA Baser Sequence Assembler v2.x 2009). As a result 1010 forward reverse sequence contigs were obtained. These sequences were cured to remove the remnants of vector sequences from both 5' and 3' ends using VecScreen at NCBI (http://www.ncbi.nlm.nih.gov/vecscreen/vecscreen.html). The CAP3 programme was used to remove the sequence redundancy (Huang and Madan 1999).

4.2.2 Mining for Simple Sequence Repeats (SSRs)

On mining 1010 sequences that represent 770.1 Kb, SSRs were found in only 636 sequences. In total 2274 SSR motifs were identified in 636 sequences with a frequency of one SSR

per 3.4 Kb (Table 4.6). Among 636 SSR containing sequences, 321 (50.5%) sequences contained one SSR, while the remaining 315 (49.5%) sequences harboured more than one SSR. Furthermore, among 2274 SSRs, 767 SSRs were compound type.

Table 4.6: Summary of *MISA* search

Total number of sequences examined	1010
Total size of examined sequences	77.01 kb
Total number of identified SSRs	2274
Number of SSR containing sequences	636
Number of sequences containing more than 1 SSR	315
Number of SSRs present in compound formation	767

4.2.3 SSR frequency and distribution

Among different classes of SSRs, di-nucleotide repeat motifs were found to be the most abundant (830; 36.5%) followed by compound (737; 32.4%), mono-(370; 16.3%) and tri-nucleotide SSRs (305; 13.4%) (Table 4.5). However, tetra- and penta-nucleotide SSRs were represented by only one number. Further, among dinucleotide AG/CT repeats and similarly ACC/GTT among trinucleotide repeats were found to be the most abundant repeat motif in the microsatellites enriched sequences examined (Fig 4.3). The distribution of different SSR motifs and SSR types are represented in Fig 4.3 and Fig 4.4.

4.2.4 Primer designing

A total of 454 primer pairs could be designed for 333 SSR containing sequences using Primer 3 (Rozen and Skaletsky 2000). However no primer could be designed for 303 SSR containing sequences. Of 454 primer pairs, 56% (254) primer pairs were designed to amplify the simple SSRs and 44% (204) primer pairs were designed for compound SSRs. More than one primer pair was designed for 92 sequences for instance four sets of primer pairs were designed for three sequences (ICMM02B04, ICMM02G14 and ICMM02O04). Among the 250 primer pairs that were designed for simple SSRs, 4.4 % (11), 48.8% and 46.8% were designed for mon-, di- and trinucleotide SSRs

motifs respectively. However no primers could be designed for tetra- and pentanucleotide repeats in the current study.

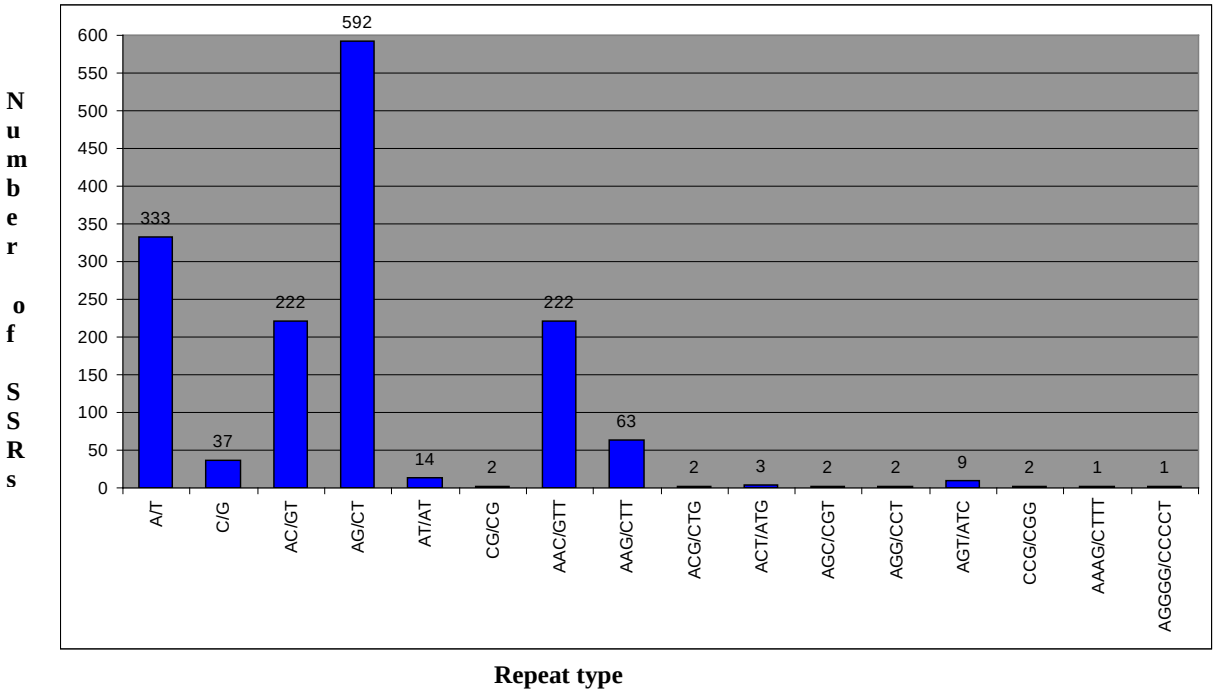


Figure 4.5: Distribution of different SSRs in the microsatellite enriched library

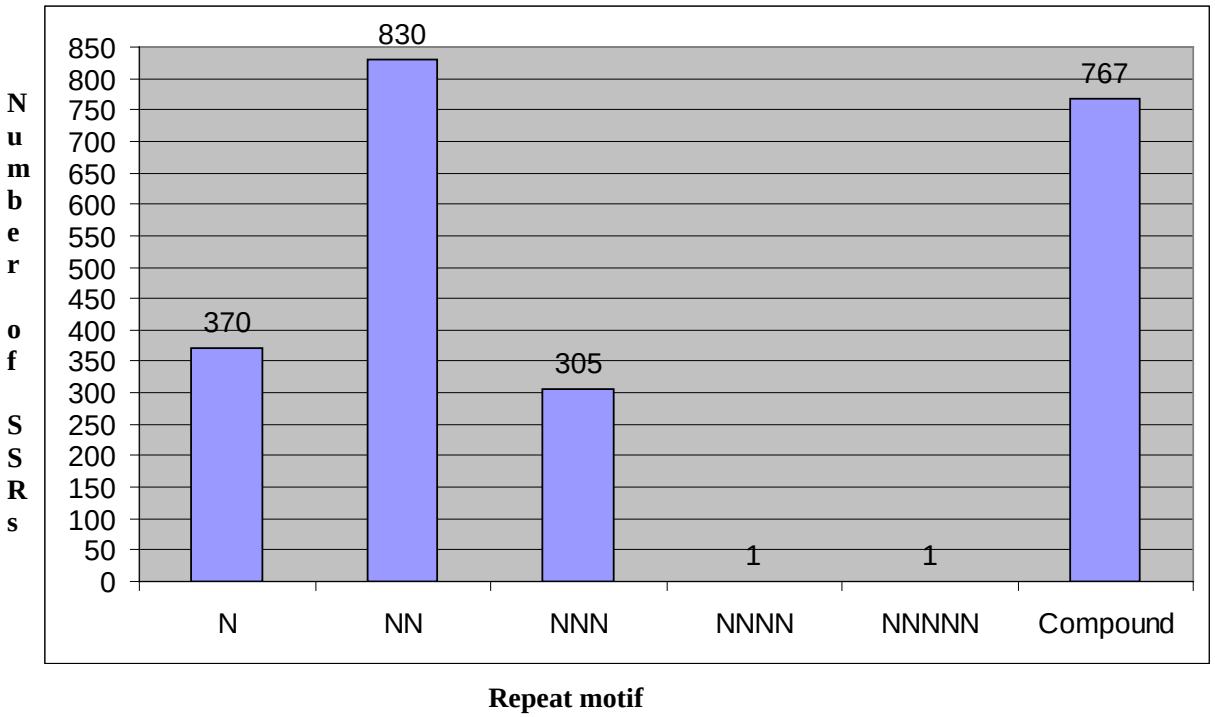


Figure 4.6: Distribution of different SSR repeat classes in the microsatellite enriched library

Table 4.7: Frequency of individual SSR motifs in 2274 microsatellites isolated from the microsatellite enriched library

SSR type	Repeats	No. of repeats												Total	%
		5	6	7	8	9	10	11-20	21-30	31-40	41-50	51-60	61-87		
Mononucleotides	A	-	-	-	-	-	33	187	34	19	11	2	3	289	19.18
	C	-	-	-	-	-	15	15	0	0	0	0	0	30	1.99
	G	-	-	-	-	-	5	2	0	0	0	0	0	7	0.46
	T	-	-	-	-	-	16	28	0	0	0	0	0	44	2.92
Dinucleotides	AC	13	9	5	11	4	3	11	4	3	1	4	1	69	4.58
	AG	54	15	4	5	3	4	33	28	8	0	0	0	154	10.22
	AT	2	2	1				0	0	0	0	0	0	5	0.33
	CA	23	14	12	4	5	3	3	4	5	3	1	0	77	5.11
	CG		1					0	0	0	0	0	0	1	0.07
	CT	18	11	10	3	6	3	17	12	0	0	0	0	80	5.31
	GA	7	8	9	3	12	2	55	14	4	3	2	0	119	7.90
	GC	1						0	0	0	0	0	0	1	0.07
	GT	5	3	3	1	1		5	2	6	1	0	0	27	1.79
	TA	6	1		1		1	0	0	0	0	0	0	9	0.60
	TC	77	15	9	4	8	3	53	53	15	2	0	0	239	15.86
	TG	9	6	3	2	5	2	12	5	0	3	2	0	49	3.25
Trinucleotides	AAC	1	2	2				0	1	0	0	0	0	6	0.40
	AAG	2						0	0	0	0	0	0	2	0.13
	ACA	1	3	1		1	1	3	2	0	0	0	0	12	0.80
	ACT		1					0	0	0	0	0	0	1	0.07
	AGA	1			1			2	2	1	0	0	0	7	0.46
	AGC	1						0	0	0	0	0	0	1	0.07
	AGT	1	2					3	0	0	0	0	0	6	0.40
	CAA	36	21	14	7	1	2	5	1	0	0	0	0	87	5.77
	CAG	1						0	0	0	0	0	0	1	0.07
	CTA							1	0	0	0	0	0	1	0.07
	CTC		1					0	0	0	0	0	0	1	0.07
	CTG	2						0	0	0	0	0	0	2	0.13
	CTT	1	3	1	2			1	3	2	0	0	0	13	0.86
	GAA	2	1		1			3	2	1	0	0	0	10	0.66
	GGC	1	1					0	0	0	0	0	0	2	0.13
	GTA	1						0	0	0	0	0	0	1	0.07
	GTT	18	6	3	8	8	3	3	5	1	0	0	0	55	3.65
	TCA			1	1			0	0	0	0	0	0	2	0.13
	TCC	1						0	0	0	0	0	0	1	0.07
	TCT	2	2					3	2	0	0	0	0	9	0.60
	TGA			1				0	0	0	0	0	0	1	0.07
	TGT	1	1	1		1		4	2	0	0	0	0	49	3.25

		5	4	3									
	TTC	1	2	3	2	5	4	5	0	0	0	22	1.46
	TTG	6	5			0	2	0	0	0	0	13	0.86
Tetranuclotides	TTCT	1				0	0	0	0	0	0	1	0.07
Pentanucleotides	CCTCC	1				0	0	0	0	0	0	1	0.07
Compound												767	
Total												2274	100.00

4.2.5 Marker validation

4.2.5.a *Optimization of PCR conditions*

Initially the PCR conditions for all the primer pairs (454) were optimized using three different PCR profiles differing in annealing temperature on two genotypes ICMB 841-P3 and 863B- P2. As a result 192, 28 and 6 primer pairs produced scorable amplification with PCR profile I (55-45), PCR profile II and PCR profile III respectively. The PCR amplicons were resolved on 1.2 % agarose gels (Fig 4.7). In order to identify the scorable amplicons for each primer pair, the PCR amplicons were further resovlved on 6% PAGE gels (Fig 4.8).Thus, in total 226 (49.7%) of the designed primer pairs produced successful amplification. However, 228 (50.3%) primer pairs did not give any PCR product of scorable amplification. Among 228 primer pairs, 78 produced smears, 30 primer pairs produced non specific/ non-scorable amplicons and remaining 120 (52.6%) primer pairs did not produce any amplification.

Table 4.8: Summary of primer optimization

	Simple*	**Compound**	**Total**
No. of primer pairs designed	250 (56)	204(44)	454
Amplified	131 (52)	95(46.5)	226
No amplification	59(23.6)	61(29.9)	120
Smears	46(18.4)	32(15.6)	78
Non-specific amplification	15(6)	15(7.3)	30

*percentages are mentioned in the parenthesis

4.2.5.b Efficiency of Marker Development

The efficiency of marker development was examined for each repeat motif. The success rate of PCR amplification and the level of polymorphism of the novel SSR markers for each SSR motif are listed in Table 4.8. The average success rate of PCR amplification was 49.77%.

Table 4.9: Characteristics of *Pennisetum glaucum* SSRs and efficiency of marker development

SSR motif	Primer pairs designed	Amplified primer[a]	Polymorphic primers[b]
Mononucleotide	**11**	**7(63.63)**	**1(14.28)**
A	3	1(14.28)	-
C	5	4(57.14)	1(14.28)
G	2	1(14.28)	-
T	1	1(14.28)	-
Dinucleotide	**103**	**54(52.42)**	**20(37.03)**
AC	16	13(24.07)	3(23.07)
AG	14	6(11.11)	3(50)
AT	2	1(1.85)	-
CA	11	4(7.40)	1(25)
CG	2	1(1.85)	-
CT	19	8(14.81)	4(50)
GA	13	3(5.55)	3(100)
GC	2	-	-
GT	4	2(3.70)	-
TA	2	-	-
TC	2	10(18.5)	3(30)
TG	16	6(11.11)	3(50)
Trinucleotide	**116**	**65(56.03)**	**21(32.30)**
AAC	7	6(9.23)	3(50)
AAG	2	2(3.07)	1(50)
ACA	7	2(3.07)	1(50)
AGA	1	1(1.538)	1(100)
AGC	2	-	-
CAA	38	24(36.92)	3(12.5)
CAG	1	-	-
CTT	3	-	-
GAA	1	1(1.53)	1(100)
GGT	1	1(1.53)	-
GTG	2	2(3.07)	2(100)
GTT	18	9(13.84)	1(11.11)
TCT	4	2(3.076)	1(50)
TGT	16	8(12.30)	4(50)
TTC	5	2(3.076)	2(100)
TTG	6	4(6.15)	1(25)
Tetranucleotide	**2**	**1(50)**	**-**
Compund	**203**	**95(46.79)**	**21(22.10)**
Total	**454**	**226(49.77)**	**63 (27.87)**

a Percentage of successfully amplified SSRs per designed primer pair
b Percentage of polymorphic markers per amplified primer pair

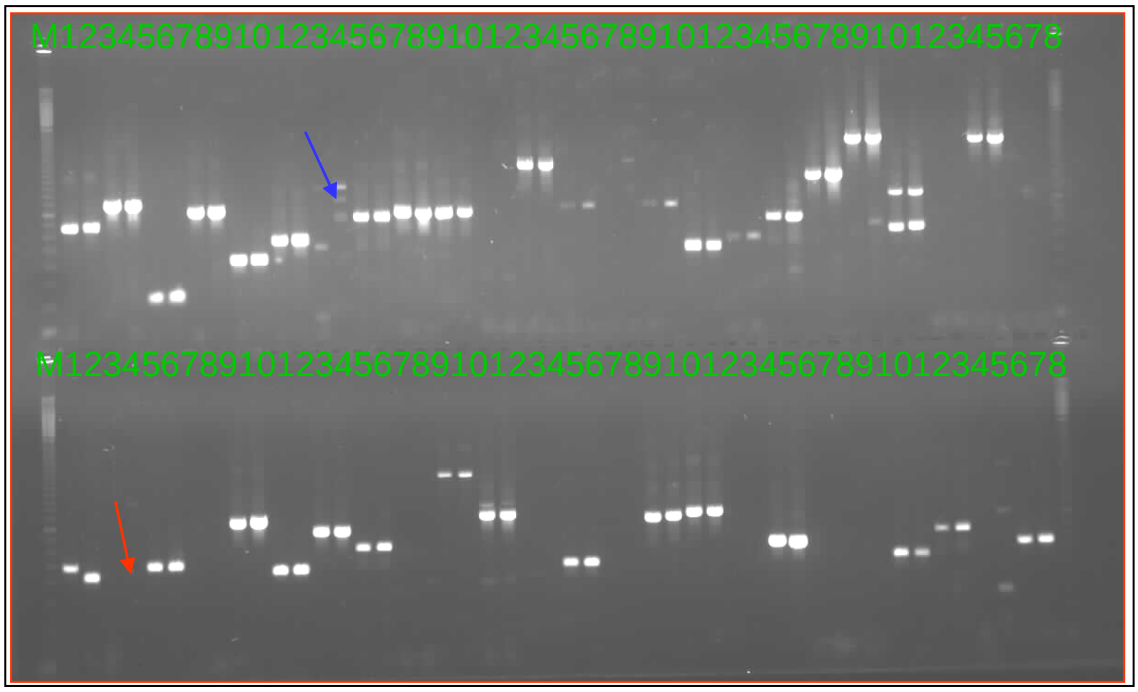


Figure 4.7: PCR amplicons of novel SSRs resolved on 1.2% agarose gels. The unamplified are sshown using red arrow and the non specific amplicfication is indicated using blue arrow.

Fig 4.8: Gel image showing polymorphic, monomorphic unamplified and smeary PCR products resolved on 6% PAGE gels.The primer ICMM03022 produced smears. Polymorphic (ICMM03O04 and ICMM03O08 etc.,) and monomorphic (ICMM02J13) are identified.

4.2.4c Polymorphism assessment of the markers

The details of the new markers viz., locus designation, repeat motifs, primer sequences, allele attributes are summarized in Table 4.7. All 226 primer pairs were further genotyped on 22 pearl millet inbred lines for parental polymorphism assessment. As a result 63 markers were found polymorphic with one or the other mapping population oparental combinations. The gel image of the 3 polymorphic markers (ICMM01C11a, ICMM01E08_X and ICMM01F01_X) on 22 parental lines is hsown in the Fig 4.9.

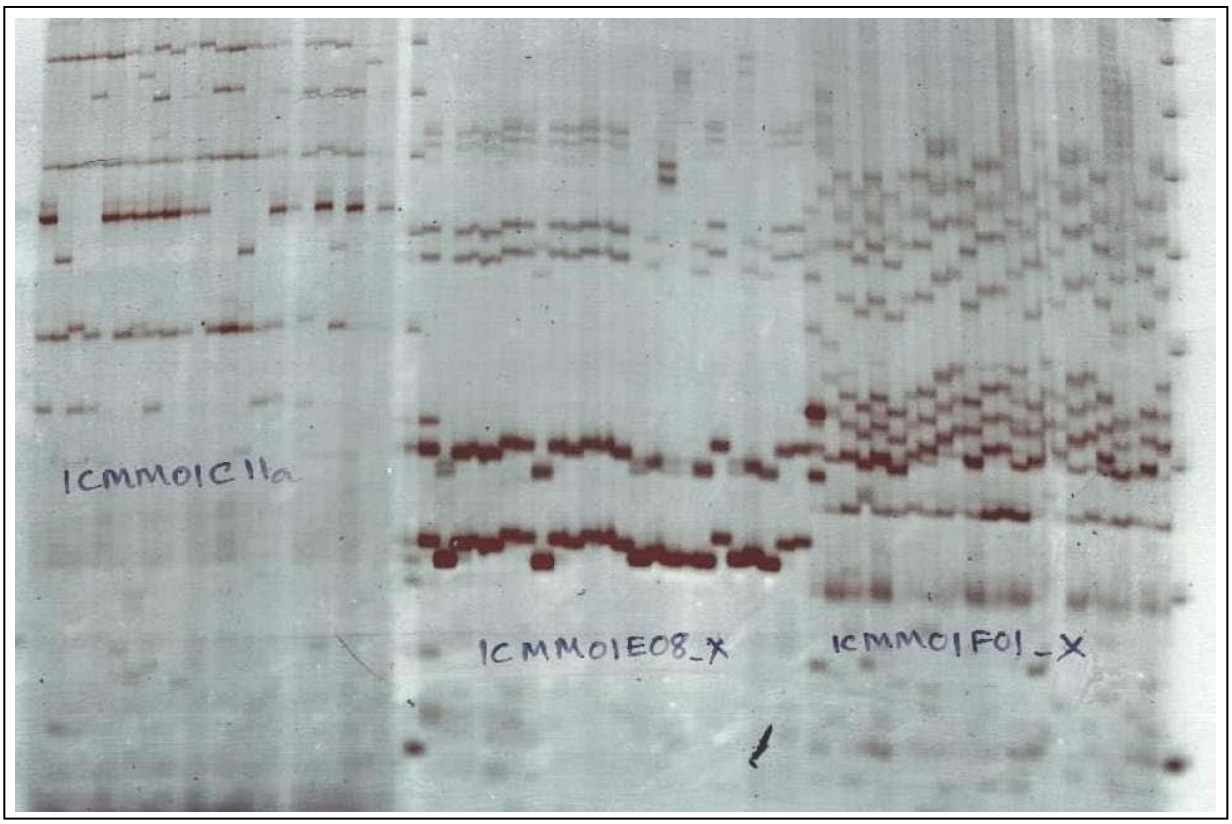


Fig 4.9: Polymorphic markers resolved on 6% PAGE gels

Table 4.10: Information on primer sequence, amplification status, polymorphism information content of novel SSR markers

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size range	No. of alleles	PIC value
ICMM01A01a	(GA)14(AG)34n(AG)5	F: TCCAGGAAAGTAGGATGCGT R: CGTCTGGTTCTCTCGCTTTC	55-45	180	1	0
ICMM01A01b	(CT)6	F: GCGGAGTATGAGTCTGGCAT R: CTCAGGCACACAATCAAGGA	55-45	130-150	2	0
ICMM01A02_X	(CAA)31n (AAC)8n(AGC)5n(GA)8	F: CGTGGACTACCCAACAACAA R: TGTCTTTATGCAACATGTGTTTTC	55-45	110-115	2	0.47
ICMM01A02_Y	(GA)10n (AG)6n(GA)6	F: ACCGGAGTATTGGTGACGAG R: TCTTGCTTACGCGTGGACTA	55-45	320	1	0
ICMM01A06_X	(A)11n(GA)9n(AG)6n(AG)10	F: TCTTGCTTACGCGTGGACTA R: GCTTACGCGTGGACTAAAAA	55-45	400	1	0
ICMM01A07a	(TC)5n(TC)7n(TC)9	F: TTTTGTGAACCTGTCGTTTCC R: TGTCCCCTTCTCCCTTTCTT	55-45	250-500	4	0.71
ICMM01A07b	(GA)13n(AG)20n(AT)6	F: TCCAGGAAAGTAGGATGCGT R: GAGGAAACGGGGTATTAGGG	55-45	300	1	0
ICMM01A09	(TC)5n(TC)7n(TC)9	F: TTTTGTGAACCTGTCGTTTCC R: TGTCCCCTTCTCCCTTTCTT	55-45	310	1	0
ICMM01A09a	(CT)25	F: TCTTGCTTACGCGTGGACTA R: TCCAGGAAAGTAGGATGCGT	55-45	180	1	0
ICMM01A09b	(TC)13n(GA)5	F: TCTTGCTTACGCGTGGACTA R: TCCCTTCTCCTTCCCTTGTT	55-45	270	4	0.69`
ICMM01A12	(CAA)5n(CAA)6	F: CAGCAGCAACAACAACAACC R: GGAATTGGTCGATTTCGTCT	55-45	240	1	0
ICMM01A24	(TGT)4n(TGT)7n(TGT)6	F: TCAATAGACATACCATCGTGCTG R: CCCACTGAGGAGGAAATGAA	65-60	140	1	0
ICMM01B09_X	(CAA)5	F: GCGACACTAGCCAGGTCAAT R: ACGTGGTTGGCTCCAGATAC	55-45	170	1	0
ICMM01B01	(GGC)5n(CAA)8	F: GGATGAAGAAAACAAGGCCA R: GCCACCAGAGTTCTTTTTCG	65-60	150	1	0

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM01C02	(AG)5n(AG)5	F: GGAAAGTAGGATGCGTCTGC R: GTGGCGTGAGAGAAAGGAAG	55-45	210	1	0
ICMM01C04_c	(AG)10	F: CCTCGTCAAAGGAGAGTTGC R: GGAGGTAGACTCCGAAAGCC	55-45	120-160	6	0.8
ICMM01C11a	(AG)14	F: ATCGAACACAACAAAACCCC R: CCACCTCTTTTTCCTCTCCC	55-45	320	8	0.8
ICMM01C11b	(TC)6	F: TTACGCGTGGACTACCTGTG R: TTGAAGGCGGATAGAGAGGA	55-45	225	1	0
ICMM01C22b	(TC)9n(CT)5n(A)15n(A)14n(A)10	F: TTTCTTTCTTGTGCATATCTCTCA R: TTCTTTGTGTTTTAGTTATTTGTTTGA	55-45	200	1	0
ICMM01D01	(GTT)8	F: CCTTCAGCAGGTCCTTGAAC R: GATGAAGAGGACAAGGCCAA	65-60	160	1	0
ICMM01D04	(CT)18n(CA)37	F: TCACATGAACCAATCTCCCA R: TCTTGCTTACGCGTGGACTA	55-45	600	2	0
ICMM01E02a	(GTT)5	F: TTGTGCTTCTTGAGCCTCCT R: CATGGTGTGTTTCAACTGCC	65-60	160	1	0
ICMM01E02b	(GTT)4	F: CAAAGTTGTTGTTGCTGCGT R: AACCAGAACACCATGAAGCC	65-60	180	1	0
ICMM01E02c	(CAA)5	F: AACCAGAACACCATGAAGCC R: CAAAGTTGTTGTTGCTGCGT	55-45	315	1	0
ICMM01E04_Yc	(TG)5	F: TGTTTTGGGTGGTGTTGTGT R: CCCCAACCCAAAAACAAA	55-45	200-700	6	0.8
ICMM01F01	(TC)29	F: GTTCGACCATCCTTTCCTCA R: GCCCTTAGGACATCCTCCTC	65-60	200	4	0.67
ICMM01F01_X	(GA)29	F: ACATCCTCCTCCTTGGTCCT R: GTTCGACCATCCTTTCCTCA	55-45	300-360	6	0.93
ICMM01E05	(AG)8n(AG)8n(AG)6	F: TCCTGGAAAGTAGGATGCGT R: TTAATCAGTGTCGTGGTCCG	55-45	140-145	2	0.44
ICMM01E08_X	(CA)13	F: GCTCGTCACGTTCTACCTCC R: ACTCCGGAATTCCAAACCTC	55-45	235-240	8	0.88

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM01F02	(TA)4n(GA)24	F: TTGCCTTTTCACTTCAGCTTC R: TGTAAATATTGTTCTTCTTTAGCTGC	55-45	200	7	0.81
ICMM01F03_Y	(CAA)5n(CAA)6	F: TACGCGTGGACTACCCTTCT R: TGAGGTAAAATCCTCTGCCAA	55-45	320	1	0
ICMM01F07	(AC)6	F: GGTTCGCCACCTCTAACAGA R: GTGGCTTGCGAACAAGGTAT	55-45	190	1	0
ICMM01G02	(AG)15	F: GCGGTTGTTGTCCATTTTCT R: CCACCTCTTCTTCCTCTCCC	65-60	200	2	0.5
ICMM01G02_X	(ACA)9	F: ACATCACCGACAAGGAGGTC R: CCTGAGTATTGTTGCTGCCA	55-45	190-205	2	0.5
ICMM01G08_Y	(CAA)5	F: TCATCGACCTCTTCCAGGAC R: CGACAGTGTTGTCTGGCTTG	55-45	440	1	0
ICMM01H01	(CAA)24(CTA)12	F: CCTGGGCAGATGAAGAAGAC R: CTTGCGCTTACGGTTAGGAC	65-60	200	1	0
ICMM01H08_X	(CAA)5	F: TCTTGCTTACGCGTGGACTA R: TTGGCCTGTTCACTGAGTTG	55-45	350	1	0
ICMM01H19a	(TC)5(TC)29n(TC)5(AT)7	F: TTTCCTTTTTCTTTTCTTTTTCTTTC R: TTTGAAAAAGCTCGTCATTTTG	55-45	110-120	2	0
ICMM01I02	(TGT)9	F: CCTGAGTATTGTTGCTGCCA R: ACATCACCGACAAGGAGGTC	65-60	-	2	0.41
ICMM01I06_X	(TTC)6n(TTC)15n(TCT)18	F: CATGGCTAGGTTGACGGATT R: AGACATGGGAAGAAGATGCG	55-45	900	4	0.83
ICMM01I08	(GA)5	F: TCCAGGAAAGTAGGATGCGT R: CCAACACATTTCCCTCTTCA	55-45	700	6	0.72
ICMM01J01a	(CT)13	F: TCTTGCTTACGCGTGGACTA R: CCTGGAGCAAAAGAACAAGC	60-55	250-250	3	0.64
ICMM01I02_X	(AC)5	F: CGCAAGGCTAAGACCTTCAC R: CCGTTTGTGATGAGTGATCG	55-45	220	2	0.5
ICMM01I02_Y	(AT)5	F: GTTCTCCTTTGTTGCGCTTC R: TGGCTCATAGTGTGTATGCGA	55-45	200	1	0

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM01J04a	(CTT)25n(TCT)17	F: GAGGCCGTTGAGATACTTGC R: GGATCATGGAAATGGGAAGA	55-45	155	1	0
ICMM01J04b	(CA)5	F: CAACGAGTTGACCCATTTCA R: TGTGTAAATGGATTGCGTTGA	55-45	150	1	0
ICMM01J06_X	(GTT)7	F: ACGGTCACACACTCCTCCTC R: TGCTTACGCGTGGACTACAG	55-45	170-190	12	0.88
ICMM01J08_X	(TTG)5	F: CCGAGTAGTTGTTGTTGCGA R: GTCGACGAAGAAGACAAGGC	55-45	120	1	0
ICMM01K02	(TC)5n(TC)7	F: TCTTGCTTACGCGTGGACTA R: GAATTCGAGCTCGGTACCTC	55-45	180-185	2	0
ICMM01K09	(GTT)6	F: CACGGGCACTTCTCCTTTAG R: GATGAAGAGGACAAGGCCAA	55-45	240	2	0
ICMM01L03	(AAG)5	F: GCTCAAAAAGGCTGTGAAGG R: AGATGAGGACGACGATGAGG	55-45	210-240	3	0.61
ICMM01M01_X	(CAA)5n(CAA)6	F: CAGCAGCAACAACAACAACC R: GGAATTGGTCGATTTCGTCT	55-45	160	1	0
ICMM01M09_X	(CAA)5	F: TACAGAACCCAAACCCAAGC R: TTTAGCATAGGGGCAGTTGG	55-45	160	1	0
ICMM01N05_Y	(CAA)12	F: GACGAAGAGGACAAAGCCAA R: TTCAGCAGGTCTTTGAACCC	55-45	210	1	0
ICMM01N19b	(CAA)5	F: GTGGACTACCAACGCCAAGT R: TGCTCTGTAGTGTGGTTGCC	55-45	300	1	0
ICMM01O01a	(GTT)5	F: GACTTGCTTGGGACAAAAGG R: TGCACTACATGTTGGGCTGT	65-60	200	1	0
ICMM01O01b	(TC)5n(TC)29n(TC)15	F: TCTTGCTTACGCGTGGACTA R: GGAGGGGAGAGAAAGAGAAAA	55-45	270	1	0
ICMM01O08_X	(TCT)10	F: GGGATCCGATTCTCTTGCTT R: GTACCTCGCGAATGCATCTA	55-45	180	1	0
ICMM01O08_Y	(TTG)5	F: TTTAGCATAGGGGCAGTTGG R: GAGAAACTTTGGACACCCCA	55-45	230-550	4	0.72

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM01O21	(CTT)7	F: TCATGCAAGGGAAAAAGACC R: TCGTGCCTGTTTCAACTGAG	55-45	350	1	0
ICMM01P01	(A)12n(TC)5	F: ACAAATAAAATATACCCACCATACAAT R: GGATATTTTAGATTTTTGGAAAGGTG	55-45	120-125	2	0.44
ICMM01P06_X	(CAA)5	F: AACCAGAACACCATGAAGCC R: TGCTGTTACTGTTGCCGAAG	55-45	800	1	0
ICMM01P06_Y	(AGT)6n(TGT)19	F: AACCCGCCAGAGTTCTTCTT R: GATGAAGAGGACAAGGCCAA	55-45	120	1	0
ICMM01P08b	(TC)5n(TC)30n(AT)6	F: TCTTGCTTACGCGTGGACTA R: TTTGTAAATATCAAAGAGGATGTTTTT	55-45	180	1	0
ICMM01P13	(GTT)5	F: TCTCCCATAGCCAGTTGTCC R: TGATGGCAGTACAGACCCAG	55-45	370	1	0
ICMM01P19	(C)10	F: TCGCACTTACAACCATGAGC R: CGTGTATACCGCCGGATAAT	55-45	430	1	0
ICMM023L23	(TC)15	F: TCTTGCTTACGCGTGGACTA R: CTTCGCCGTTCGAGTTTTAG	55-45	170	1	0
ICMM02A02	(AG)16n(AG)32n(AG)5n(A)10	F: TCCAGGAAAGTAGGATGCGT R: TCTTGCTTACGCGTGGACTA	55-45	225	1	0
ICMM02A03	(AAG)4	F: GCGGCAAGATTGGACACTAT R: ACGTGTCCTCGGAATCAGAC	65-60	200	1	0
ICMM02A04	(CAA)8	F: GATGAAGAGGACAAGGCCAA R: GCACTTCTCCTTCAGCAGGT	65-60	315	1	0
ICMM02A05a	(TC)5n(TC)45n(TA)4	F: TTTCCTTTTTCTTTTCTTTTTCTTTC R: GGCATAATTGACAAGGACCA	65-60	440	1	0
ICMM02A06	(TA)4n(GA)51n(AG)5	F: CAAGGGCCATCTTTAGTGGT R: TTTCCTTTTTCTTTTCTTTTTCTTTC	65-60	160	1	0
ICMM02A10	(AG)9n(AG)20	F: TTGAGGAAATGTAGACCTTCCA R: CCCCTTTGAAGCTATTGTCC	65-60	300	4	0
ICMM02A12	(AC)14	F: TGATGCCCGTCATATTGAAA R: CCCGTGTCCGTACATGTTTT	65-60	120	1	0

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM02A13	(TC)5n(TC)31n(TC)15n(TA)4	F: CACCTGTCGTTTCCTTTTTCTT R: CAAGGACCATCCTTAGTGGTTC	65-60	180	1	0
ICMM02A14	(GGT)4	F: ATCTGCTGGTTGCTGGTGTT R: CCAACTGCCCCTATGCTAAA	65-60	370	1	0
ICMM02A15	(CT)4n(TG)15	F: ATTCAAGCGCCCAATATCAC R: AGGACTGGAGCTGTGAATGC	65-60	430	1	0
ICMM02A16	(GA)9n(AG)6n(GA)7	F: ACCGGAGTATTGGTGACGAG R: TCTTGCTTACGCGTGGACTA	55-45	160	1	0
ICMM02A17	(TGT)5	F: GGTTCGTTTACCGAGTTGGA R: TACAGAACCCAAACCCAAGC	65-60	240	1	0
ICMM02A18	(GA)9n(AG)5n(GA)7	F: ACCGGAGTATTGGTGACGAG R: TCTTGCTTACGCGTGGACTA	55-45	410	1	0
ICMM02A19	(GT)34n(GT)6	F: TGACGAGGAGCAAAAAGGAT R: GGGTAGTATGTAAGGCGGCA	65-60	240	1	0
ICMM02A22	(TG)5n(TG)9	F: ATCCGTCAGCAGAAATGGAG R: GCACAGAGATAATCACACAAGCA	65-60	240	1	0
ICMM02A24	(TG)4	F: AGGGGGAGCTTGTGAGTTTT R: CAAGAAATGCCCAAAGGAAA	65-60	250	3	0.51
ICMM02B01	(AAC)4	F: CGATCAGAATGCTGCAGGTA R: CGCATCACCTTGCTGTTCT	65-60	280-380	3	0.66
ICMM02B02	(G)13n(AAC)24	F: CATCATTGTGCCCAGCATAC R: TCCATGTGTTTCCTTCCTCA	60-55	290	7	0.81
ICMM02B03	(CAA)5n(CAA)6	F: CAGCAGCAACAACAACAACC R: GGAATTGGTCGATTTCGTCT	55-45	175	1	0
ICMM02B04a	(TGT)4n(TGT)7n(TGT)5	F: TCAATAGACATACCATCGTGCTG R: CCCACTGAGGAGGAAATGAA	65-60	290	1	0
ICMM02B04c	(CAA)5n(CAA)7	F: CCCACTGAGGAGGAAATGAA R: ATACCATCGTGCTGTTGCTG	55-45	230	1	0
ICMM02B04d	(TC)5n(TC)30n(TC) 16n(TA)4	F: TTCCTTTTTCTTTTCTTTTTCTTTC R: GGCATAATTGACAAGGACCA	60-55	290	1	0

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM02B05	(GTGTGG)4	F: AAAGCATCCATGATGGTGAAG R: GAAATGCTCGCAAACAACAA	60-55	300	1	0
ICMM02B07a	(CAA)8	F: GAAGAGGACAAGGACAACGC R: CGCCTGAGTTCTTCTTCGAG	55-45	150-170	2	0
ICMM02B07b	(TA)4n(GA)16n(AG)27n(AG)5	F: GGCATAATTGACAAGGACCA R: TGTCTTTTCCTTTTTCTTTTCTTTTT	65-60	300	1	0
ICMM02B08	(TA)4n(GA) 51n(AG)5	F: CAAGGGCCATCTTTAGTGGT R: TTTCCTTTTTCTTTTCTTTTTCTTTC	60-55	310	1	0
ICMM02B09b	(TCA)7n(CAA)19	F: TGTCCAATGATGATGCCTCT R: GGGAGCTGGAAGAGGAAAAG	60-55	450-550	5	0.71
ICMM02B23	(TC)5n(TC)24n(TC)16n(TA)4	F: CACCTGTCTTTTCCTTTTTCTTTT R: GGCATAATTGACAAGGACCA	55-45	270	1	0
ICMM02B24b	(CAA)4	F: AGCAAATAGAACACGGGTGG R: TTGAGTGCGTGCCTAAGTTG	55-45	220, 270	2	0
ICMM02C03b	(GTT)9	F: CGGAGGTCTTCTTCGAGTTG R: AGAAGACAAGGCTAACGCCA	55-45	800-810	2	0
ICMM02C10b	(TA)n(GA)42n(AG)5	F: GGCATAATTGACAAGGACCA R: CACCTGTCTTTTCCTTTTTCTTTT	55-45	400-500	4	0.52
ICMM02C11b	(TGT)n(TGT)4n(TGT)6	F: TCAATAGACATACCATCGTGCTG R: CCCACTGAGGAGGAAATGAA	55-45	290	1	0
ICMM02C12a	(CTT)19	F: CCTTCCTCATGCTCCTCTTG R: ACGCGTGGACTAACTTCGTT	55-45	110	1	0
ICMM02C12b	(GTG)4	F: AAGCAGATGCGGAAGCTAAG R: TATCCCAACACCAGCTCTCC	55-45	200-210	4	0.72
ICMM02C14a	(GTG)4	F: GGAGAGCTGGTGTTGGGATA R: TCGAAGCAGTAGGTCTCCGT	61-51	400-500	8	0.88
ICMM02C15b	(CAA)5n(CAA)6	F: CAGCAGCAACAACAACAACC R: GGAATTGGTCGATTTCGTCT	55-45	385	1	0
ICMM02C21	(AG)6	F: GTCAAAAAGGCCATGGAAAA R: TCTTTTCCAACCATGTGCAA	55-45	240	1	0

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM02C24a	(AG)6	F: GTCAAAAAGGCCATGGAAAA R: TCTTTTCCAACCATGTGCAA	55-45	600	1	0
ICMM02D15a	(AGA)5	F: AACGGAAGGGTAAGGCAGTT R: CCAGTGAGGGTGGTGTCTCT	61-51	280-300	2	0.35
ICMM02D18b	(TC)26	F: TCTTGCTTACGCGTGGACTA R: TGGACCTTCATGTGTCCAAC	55-45	195	1	0
ICMM02D22a	(CAA)5	F: AAGAACAAGCAAAACACGGG R: AGTTGGTAGCACTGCTCGGT	55-45	250	1	0
ICMM02D23	(TG)10n(TG)27	F: TCTTGCTTACGCGTGGACTA R: GAGCATAGACAAGACCCCCA	55-45	295	1	0
ICMM02E03a	(AC)6	F: TTACGCGTGGACTAACTCCC R: AGACGCCCATCACATTTTTC	55-45	290	1	0
ICMM02E03b	(TA)5	F: AGGGCGTAAGTCCATCCTCT R: TGCATGAGCTGTGTGGTACA	55-45	290	1	0
ICMM02E09	(TC)5n(TC)19	F: TCTTGCTTACGCGTGGACTA R: CCTCGAAACTGAAAGCTCAGA	55-45	210	1	0
ICMM02E11	(AC)5	F: TCTTGCTTACGCGTGGACTA R: TAAAAACCTGGGGCGTAACA	55-45	320	1	0
ICMM02E21	(GA)17n(A)11n(GA)5	F: TTGGAAAGTTTTAGCAGTTGGA R: TTCTTCTTCTTTTCTTCCCTCTTTC	55-45	170	1	0
ICMM02F10	(TG)15	F: ACGTAACTGGCTCACCTGCT R: TCTTGCTTACGCGTGGACTA	55-45	600-620	4	0.73
ICMM02F13a	(CT)6	F: TCTTGCTTACGCGTGGACTA R: ATCAAGGAAGAGCCGGAAAT	55-45	500-510	3	0.64
ICMM02F13b	(TC)35	F: TTTTGTGACACCTGTCTTTTCC R: GTGGGGAGGTTGATTGTGTT	55-45	270	1	0
ICMM02F13c	(CAA)5	F: TGCGCAGCAACAATAACTTC R: TCCTGTGCATTGATGTGGTT	55-45	230-240	2	0.44
ICMM02G10	(CT)21	F: ATTCTCTTGCTTACGCGTGG R: AGGGGTGGTCTGCATGTAAG	55-45	270	1	0

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM02G14c	(TC)35n(TC)6	F: TGCTTACGCGTGGACTAAAA R: GGGAAAGGGAAAAAGAGAGAAA	55-45	420-510	3	0.64
ICMM02G14d	(TTC)12	F: TGGACTACCCGTCACACAAA R: GTACCTCGCGAATGCATCTA	55-45	300-410	5	0.78
ICMM02G18	(GA)9	F: GTGGTCGTTCTTCGCTTGAT R: TACGCGTGGACTAACCACAA	55-45	800	1	0
ICMM02G21	(A)10	F: TCTTGCTTACGCGTGGACTA R: ATCCTGTCCTTGTCTCGTGG	55-45	180	1	0
ICMM02G22	(AC)8	F: AACACCTCCCGCATATTCAC R: TCTTGCTTACGCGTGGACTA	55-45	460	1	0
ICMM02G24	(AG)5	F: TATCTACTGTGCGCGGGAAT R: AGTTGCGCGTGGACTAAAAT	55-45	290	1	0
ICMM02H07	(TC)9	F: TCTTGCTTACGCGTGGACTA R: AAAAAGAACGGCGAAGGAAT	55-45	300	1	0
ICMM02H11	(GTT)6	F: CCTTCTTCTGGAAGCCCTCT R: TCTTGCTTACGCGTGGACTA	55-45	200	1	0
ICMM02H15	(CAA)7	F: AACCAGAACACCATGAAGCC R: CCGAGTTGGAGTAGGTGGAA	55-45	230	1	0
ICMM02H18a	(AAC)6	F: GCGTGGACTAACTTGCATCA R: CGTTCCTGAGCCTTGACTTC	55-45	290	1	0
ICMM02H23	(TTG)5	F: TGCTCATCACCTTGCTGTTC R: ATGACAACAGTGCTGTGGGA	55-45	320	1	0
ICMM02I03	(TGT)5	F: TCTTCGAGTTCCCTTTTGGA R: GATGAAGAGGACAAGGCCAA	55-45	280	1	0
ICMM02I17	(AC)5	F: TGTAATTCGCCCTCTCCTTG R: CAAGTCCCTGGTGGAGTGAT	55-45	225-230	4	0.63
ICMM02I19	(TG)7n(TG)5n(TG)5	F: AGATCGACGACCTGCGAG R: AGATTGCTTACGCGTGGACT	55-45	150-160	4	0.61
ICMM02I21	(CAA)5	F: AAGAACAAGCAAAACACGGG R: AGTTGGTAGCACTGCTCGGT	55-45	270-280	2	0

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM02I22	(TG)8	F: TCTTGCTTACGCGTGGACTA R: AACACCTCCCGCATATTCAC	55-45	205	1	0
ICMM02J09b	(GAA)5	F: AGGCTGTGAAGGAAAGCTCA R: AGGACGACGATGAGGATGAC	55-45	150	3	0.66
ICMM02J13	(ACA)19	F: ATGCAGGCCTCTGCAGTC R: TCTTGCTTACGCGTGGACTA	55-45	210	1	0
ICMM02J15	(GTT)5	F: CTTCGAGTTCTCCTTCGTGG R: GAAGAAGACAAGGCCAATGC	55-45	215	1	0
ICMM02J16	(TGT)5	F: TTTAGCATAGGGGCAGTTGG R: ACGCGTGGACTAACAGAACC	55-45	200	1	0
ICMM02K16	(GA)16n(AG)28n(AG)5	F: AGGAAAGTAGGATGCGTTTGT R: TGACACCTGTCTTTTCCTTTTTC	55-45	170-220	3	0.66
ICMM02K18a	(TC)5n(TC)18n(TC)6	F: TCTTGCTTACGCGTGGACTA R: GATAGGAAGGCGGGTGAAAG	55-45	220	1	0
ICMM02K18b	(CAA)5	F: TACAGAACCCAAACCCAAGC R: CGGCTGGCTTTTTAGCATAG	55-45	410	3	0.49
ICMM02K19	(TG)9	F: GTGGACTACATGCCCTGGAT R: GCATGTGACAACCAAGTTGC	55-45	220-410	3	0
ICMM02K24	(TTG)5n(GTT)6	F: AACCTGTCCAGAACCACCAG R: CGTGTTCGAGGGATTTCTTC	55-45	200-220	2	0.44
ICMM02M02_X	(CTT)32	F: TCTTGCTTACGCGTGGACTA R: AAGACATGGGCTCGAGAGAA	55-45	600	1	0
ICMM02M16	(TGT)16n(TGT)6	F: CATACCATCGTGCTGTTGCT R: CACCAACAAACACACCCAAA	55-45	225	1	0
ICMM02N06a	(TG)46n(GT)12n(GT)6	F: TTAGGCCATACCCACCAAAC R: TGGGTGGGGATCTTATCAAA	55-45	410	1	0
ICMM02N22b	(TC)5n(TC)17n(CT)5	F: TCTTGCTTACGCGTGGACTA R: TTGTGCTCTCGTTGGTTCTG	55-45	330	1	0
ICMM02N22c	(GT)5	F: GCGTGTGTGCATGTGTGTAG R: TCTTGCTTACGCGTGGACTA	55-45	475	1	0
ICMM02N24	(GA)7	F: TCCAGGAAAGTAGGATGCGT	55-45	320	1	0

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
		R: TGGGGATTGGGATGTAGGTA				
ICMM02O02a	(TC)5n(TC)25	F: TCTTGCTTACGCGTGGACTA	55-45	475	1	0
		R: TGGTGATGATTGCGAGTAGC				
ICMM02O04d	(AGT)12n(TGT)24	F: CTTGCGCTTACGGTTAGGAC	55-45	600-700	4	0.73
		R: CCTGGGCAGATGAAGAAGAC				
ICMM02O11	(TGT)5	F: CTTTGAAACCGCCAGAGTTC	55-45	500-570	4	0.7
		R: GATGAAGAGGACAAGGCCAA				
ICMM02P02_Xb	(TC)5	F: TCTTGCTTACGCGTGGACTA	55-45	280-310	4	0.5
		R: GAATTCGAGCTCGGTACCTC				
ICMM02P09	(CA)41	F: GGGTAGTATGTAAGGCGGCA	55-45	500	1	0
		R: ATTCTCTTGCTTACGCGTGG				
ICMM02P12	(TTC)6	F: CTGCTGTGTCTGAGCAGAGG	55-45	300-600	6	0.77
		R: AAGCAGATGCGGAAGCTAAG				
ICMM02P13	(CT)10	F: CTCTTGCTTACGCGAGGACT	55-45	210-290	2	0
		R: TGCGCAGCAATAACAACTTC				
ICMM02P15b	(AC)8	F: AACACCTCCCGCATATTCAC	55-45	1000	2	0.5
		R: TCTTGCTTACGCGTGGACTA				
ICMM02P17	(TC)10	F: TGGTTCTTGGTCTTAGTGACCC	55-45	310-330	4	0.5
		R: AGGGAGATTTCCATCCCAAC				
ICMM02P19	(GA)21	F: CGACAGTGTTGTCTGGCTTG	55-45	300-350	3	0.54
		R: TCATCGACCTCTTCCAGGAC				
ICMM03A09b	(TC)5n(TC)28	F: TCTTGCTTACGCGTGGACTA	55-45	180-190	4	0.7
		R: GTATGAGGGGGTTTGTGGTG				
ICMM03A11	(TC)12n(CT)9n(CT)7	F: ACGCGATCTTGCAACTCTTT	55-45	290	1	0
		R: GCATGACTGCTGAAGGAACA				
ICMM03A14	(CAA)5	F: TGCGCAGCAACAATAACTTC	55-45	250-270	4	0.74
		R: TCCTGTGCATTGATGTGGTT				
ICMM03A15_X	(AAC)7	F: TGATGGCAGTACAGACCCAG	55-45	105	1	0
		R: GAGTTGTCGAAAGTAGGCGG				
ICMM03B04	(CT)8n(TC)18	F: TCTTGCTTACGCGTGGACTA	55-45	250	1	0
		R: CACGCCAAAGATAGGGAAAA				

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM03B20	(CAA)6	F: ACACTATGAAGCCCCCACAG R: GGCTTCTTGGCATATGGGTA	55-45	670	1	0
ICMM03C08	(CAA)19	F: AGAAGACAAGGCTAACGCCA R: CCACGACCATGTTGTTTGTC	55-45	150	1	0
ICMM03C11	(TG)5	F: CTAAAGGGTCCCCTGTCGAT R: GGCATGACCATCGAAGAAGT	55-45	245	1	0
ICMM03C12	(GA)17n(AG)19	F: TCCAGGAAAGTAGGATGCGT R: TTTCCCTTTTTCTTTCCCTCTT	55-45	290	1	0
ICMM03C14a	(TC)5n(TC)29n(CT)6	F: TGCTTACGCGTGGACTAAAA R: TGGGAATTAGGAAGAGTTCACG	55-45	320	1	0
ICMM03C14b	(CG)5	F: CGTTGGCCTCCCTCCTATAA R: GCTTCGTGGGAGAACTCAAG	55-45	270	1	0
ICMM03C16_X	(AC)56	F: GGGATCCGATTCTCTTGCTT R: TTAGGCCATACCCACCAAAC	55-45	300-310	2	0
ICMM03C18	(GT)7n(TC)20	F: AGGGAACAAGGTTTTGTCCA R: GGAGGAAGAAAAACCGCATT	55-45	260	1	0
ICMM03C22a	(TC)5n(TC)13	F: TCTTGCTTACGCGTGGACTA R: GCCTCTCCATTTCATCAATCA	55-45	270-290	6	0.64
ICMM03C22b	(TG)8	F: TCTTGCTTACGCGTGGACTA R: AACACCTCCCGCATATTCAC	55-45	300	1	0
ICMM03D12	(G)10	F: CCCCCGCCATTATACCTTAT R: CCCTGTTCTGTCAGTCGGTT	55-45	140	1	0
ICMM03D16	(CAA)6	F: AACCAGAACACCATGAAGCC R: CCATAAGACTGTTGGGGCTG	55-45	255	1	0
ICMM03D22	(TTC)15n(TCT)14	F: TCGGAATTTATGTTCGAGGC R: GATGAGGAAGATGATGGGGA	55-45	600	1	0
ICMM03E05_Y	(TCT)6	F: ATACCATGCACGTAGGAGCC R: GATCAACCAAGGAAAAGGCA	55-45	200-300	8	0.8
ICMM03E17	(TGT)5	F: CAGTTGAAACACACTCGGGA R: AAGAAGGAAAACAAGCGCAA	55-45	190-300	6	0.78

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM03F12	(GA)6n(GA)21	F: TGGATCACGTTGTGTGTCCT R: ACTGCATTCCACACCTCTCC	55-45	400	1	0
ICMM03G06	(CAA)5	F: AAAAATCAGGCAAGGGTCAG R: TCTTGCTTACGCGTGGACTA	55-45	260	1	0
ICMM03G11a	(TC)30	F: TCTTGCTTACGCGTGGACTA R: TCCTCAATAATTCGATGTCTTTTG	55-45	600	1	0
ICMM03G11b	(AAC)7	F: TGTAGTCGCCACCAGAAGTG R: GACTTGCTTGGGACAAAAGG	55-45	470	1	0
ICMM03H04_X	(CT)15	F: GCTCCATTTTGAAGCCCATA R: GATGAACCTTGACGTCACCC	55-45	300-310	3	0.66
ICMM03H09b	(TG)46n(GT)16	F: TCTTGCTTACGCGTGGACTA R: GCAGAGGCACACACAGAGAA	55-45	440	1	0
ICMM03H12a	(AC)17	F: TACGCGTGGACTAACCAACA R: ATCGCTCAGGAAACACCATC	55-45	100	1	0.73
ICMM03H12b	(AAC)7	F: TGATGGCAGTACAGACCCAG R: GAGTTGTCGAAAGTAGGCGG	55-45	300	3	0.65
ICMM03H16	(GA)15n(AG)16	F: TCCAGGAAAGTAGGATGCGT R: GGTAATCCACCTTGATGGCA	55-45	150-500	6	0.81
ICMM03H18a	(TC)5n(TC)30	F: TCTTGCTTACGCGTGGACTA R: TCTGTTGCTATTGTGATATTTGGAA	55-45	300	1	0
ICMM03H18b	(CA)5n(AC)5	F: CCCCTTCACGTGGCTAAATA R: GTGGTGTGCTCAAGGCACTA	55-45	420	1	0
ICMM03I14_X	(TTG)5	F: CTTCAAGTTCCCTTTCGCAG R: CAGCAATACAAACCAAGGCA	55-45	380	1	0
ICMM03I21	(AAC)5	F: CGGCAAGAGTAAGCACAACA R: CGTGGGTCTGAGGTATCGTT	55-45	550-560	4	0.7
ICMM03J04_X	(GT)8	F: GATTGCCCCTTCGTCAGTAA R: TGGAGAGAGCTGGGAAAGAA	55-45	160	1	0
ICMM03J04_Ya	(T)10	F: TTTCCACCATCCTTGGTCTC R: TCCATCCATCCATCCATTCT	55-45	170	1	0

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
ICMM03J04_Yb	(TC)18	F: TTACGCGTGGACTAAAAGCA R: TGGACTACGCGTGGACTAAA	55-45	180	1	0
ICMM03J08	(GA)9n(AG)6n(AG)9	F: ACCGGAGTATTGGTGACGAG R: TCTTGCTTACGCGTGGACTA	55-45	270	1	0
ICMM03J09	(AC)9n(AC)12	F: TACGCGTGGACTAACCAACA R: TCTTGCTTACGCGTGGACTA	55-45	320	1	0
ICMM03J11	(C)10	F: CTTAATCCCTCTTCCCACCC R: TAGCATTCAGATGGCGTCAG	55-45	400-450	4	0.7
ICMM03J12	(GT)24n(GA)13	F: GATCCGATTCTCTTGCTTCG R: TCCGAGGACACCCTGATTAC	55-45	300	1	0
ICMM03J13	(CAA)7	F: AACCAGAACACCATGAAGCC R: CCGAGTTGGAGTAGGTGGAA	55-45	705	1	0
ICMM03K07	(GA)15n(AG)28n(AG)5	F: TCCAGGAAAGTAGGATGCGT R: TCTTGCTTACGCGTGGACTA	55-45	320	1	0
ICMM03L02_X	(TTCT)5	F: TGCACTTTAGAGAATCGGGG R: GTGATCTACTCGGTCCTCGC	55-45	170	1	0
ICMM03L06_Y	(CT)8n(TC)10	F: GGGATCCGATTCTCTTGCTT R: GTACCTCGCGAATGCATCTA	55-45	260	1	0
ICMM03L06_Z	(CAA)7	F: GATGAAGAGGACAAGGCCAA R: AGAAAATCCCGAAAAATCCG	55-45	170	1	0
ICMM03L07_Y	(TC)5n(TC)21	F: TCCAGGAAAGTAGGATGCGT R: GTCCGGGATCGGATTATTTT	55-45	200-400	5	0.83
ICMM03L14	(GA)17n(AG)5	F: TCCAGGAAAGTAGGATGCGT R: TCTTGCTTACGCGTGGACTA	55-45	105	1	0
ICMM03L20b	(C)14n(AC)8	F: GCATAAGACACAGTCATGCCA R: TGGCGGATAATCAAAATTGC	55-45	410	1	0
ICMM03L23	(CAA)5n(CAA)6	F: GCAACAGCAACAACCTCATC R: CGATTTCGCCTCTCTGTTGT	55-45	280-310	8	0.68
ICMM03M18	(GA)17n(AG)28n(AG)5	F: TCCAGGAAAGTAGGATGCGT R: TCTTGCTTACGCGTGGACTA	55-45	220	1	0
ICMM03M20_X	(AC)8	F: AACACCTCCCGCATATTCAC	55-45	195-200	2	0

Primer ID	SSR motif	Primer sequence(5'-3')	PCR profile	Amplicon size	No. of alleles	PIC value
		R: TCTTGCTTACGCGTGGACTA				
ICMM03N15a	(TC)17n(TC)6n(GT)7n(TG)9	F: TGCCTACTCACGATGCTTTG	55-45	350-360	4	0.74
		R: ATAGGGACAAGGTGCAGGTG				
ICMM03O04	(CT)6	F: ACGACCCGTACGTCTCCTTA	55-45	140-190	6	0.8
		R: CCGCACTCTGCATTGAGATA				
ICMM03O05a	(CA)5	F: TGTACTCGTCGTAAGGTGCG	55-45	210-220	2	0
		R: CGTATTCGATCTGCGAGTCA				
ICMM03O05b	(AC)7	F: GACCCGTAGGTCCTCCTTTC	55-45	310	1	0
		R: CAAAGCAGTTAGAGGCAAAAA				
ICMM03O08_X	(TGT)7	F: ATACCATCGTGCTGTTGCTG	55-45	150-200	4	0.8
		R: CCCACTGAGGAGGAAATGAA				
ICMM03O15_X	(AC)39n(CA)10	F: TGCTTATGGCAGTCATCAAAA	55-45	290	1	0
		R: TCTTGCTTACGCGTGGACTA				
ICMM03P03	(TGT)20	F: TGCCTGAGTTGGTAGCACTG	55-45	750	1	0
		R: GCAATACAAACCAAGGTGGC				
ICMM03P06_X	(C)11	F: TGCATCCACACCTCCTAACA	55-45	250-350	2	0
		R: TTAGGTCGGAAAATTGTGGG				
ICMM03P09b	(CA)5n(AC)9	F: GATGTGCGCTTTTATTGGGT	55-45	290	1	0
		R: GAATTGAGCTCGGTACTCGC				
ICMM03P13	(C)10	F: CCCCTCTACTGCCAACACTC	55-45	275	1	0
		R: GGGATAGTGGGGTTGTTGTG				
ICMM03P15a	(GA)16n(AG)14	F: TCCAGGAAAGTAGGATGCGT	55-45	280	1	0
		R: GTGGGGTTTACGGTAGGGTT				
ICMM03P15b	(TC)5n(TC)27	F: TCTTGCTTACGCGTGGACTA	55-45	120	1	0
		R: AGTGTTGGCAGTAGAGGGGA				
ICMM03P15c	(CAA)5	F: TACAGAACCCAAACCCAAGC	55-45	280	1	0
		R: CTTGGCCCATTCACAGAGTT				
ICMM03P20	(TG)54n(GT)13	F: TTAGGCCATACCCACCAAAC	55-45	350	1	0
		R: TCCACACATATCTCCCCTCC				
ICMM03P21b	(AC)21at(CA)8ctcact(CA)10	F: GGAGGGGAGATATGTGTGGAA	55-45	244	1	0
		R: CCATCAATCACACCCTCACA				

4.3 Construction of Genetic Linkage Map

4.3.1 Genotyping of polymorphic markers

A total of 144 markers, including 48 CISPs, 31 SSCP-SNPs, 26 SSRs and 39 novel SSRs developed in this study were genotyped on the mapping population ICMB 841-P3 × 863B- P2 consisting of 149 RILs. As a result 147 loci were produced by 144 markers. Three SSR markers (Xpsmp2229, Xpsmp2270 and Xpsmp2231) amplified more than one locus. The genotyping data of each marker was used for computation of the map distances. A representative gel image showing the segregation of three markers Xpsms84, Xpsms86 and Xpsms74 can be visualized on Fig 4.10.

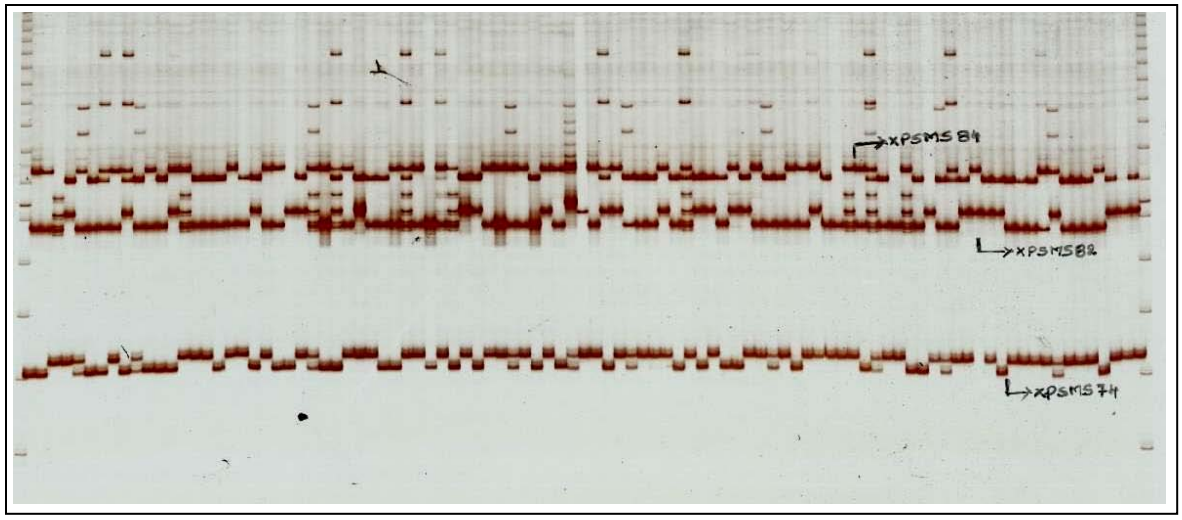


Figure 4.10. A representative gel image showing segregation of three markers (Xpsms74, Xpsms83 and Xpsms84); P1 and P2 indicate the parental genotypes (ICMB841-P3 and 863B-P2) M= 100bp marker

4.3.2 Segregation distortion

Forty two percent of the mapped markers segregated in 1:1 ratio. Forty one markers (58%) showed distorted segregation ($P \leq 0.05$, chi-square test). Results of linkage analysis revealed that markers with distorted segregation were distributed throughout the

genome. The number of markers showing segregation distortion varied from 1 to 10 per LG. The most extreme examples of segregation distortions in the RIL was found with marker ICMM02M02_X on LG 2 where only 22 of 152 plants were “BB” homozygotes, and Xpsms 21 on the LG1 where only 21 of 80 plants scored were also “AA” homozygotes. A set of ten mapped loci in LG 2 were significantly skewed towards ICMB 841- P3 parent and clustered at two different positions. However, in case of LG 7 two mapped loci were skewed towards the parent 863B- P2.

4.3.3 Linkage map

A total of 144 markers, including 48 CISP, 31 SSCP-SNP, 26 SSRs and 39 novel SSRs developed in this study, were used for the construction of a linkage map. Of which 69 of these markers could be assigned to 7 linkage groups (LG1 – LG7). The linkage map of the F_6 population spans a total genetic distance of 3593 cM (Haldane cM), with 61 markers remaining unlinked. Table 4.11 provides a summary of marker distribution on different linkage groups showing the size, number of markers and the average marker interval of each LG. The distance between the markers on the map also varies greatly across the different linkage groups. The length of the linkage groups ranged from 87.1 cM (LG 6) to 1328.4 cM (LG2). The average marker distance was 41.7 cM, with intervals between loci ranging from 0 to 345.4 cM (Fig. 4.11). Maximum number of markers (33) was mapped on to LG 2 followed by LG 3 (14 markers), LG 7 (12 markers), LG 5 (10 markers). However, LG1, LG4 and LG 6 contained only 8, 5 and 4 mapped loci respectively.

Table 4.11: Distribution of mapped markers on seven pearl millet chromosomes

Linkage group	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Total**
Map length (cM)	493.1	1028.4	677.2	94.9	516.9	87.1	395.4	**3293**
gSSR (Qi et al. 2004)	1	7	5	3	2	2	7	**27**
gSSRs (novel)	1	12	3	-	3	-	2	**21**
SSCP-SNP	3	4	2	-	3	-	2	**14**
CISP	3	10	4	2	2	2	1	**24**
Total Marker loci	**8**	**33**	**14**	**5**	**10**	**4**	**12**	**86**
Average map density	**61.63**	**31.16**	**48.37**	**18.98**	**51.69**	**21.77**	**32.95**	**38.29**

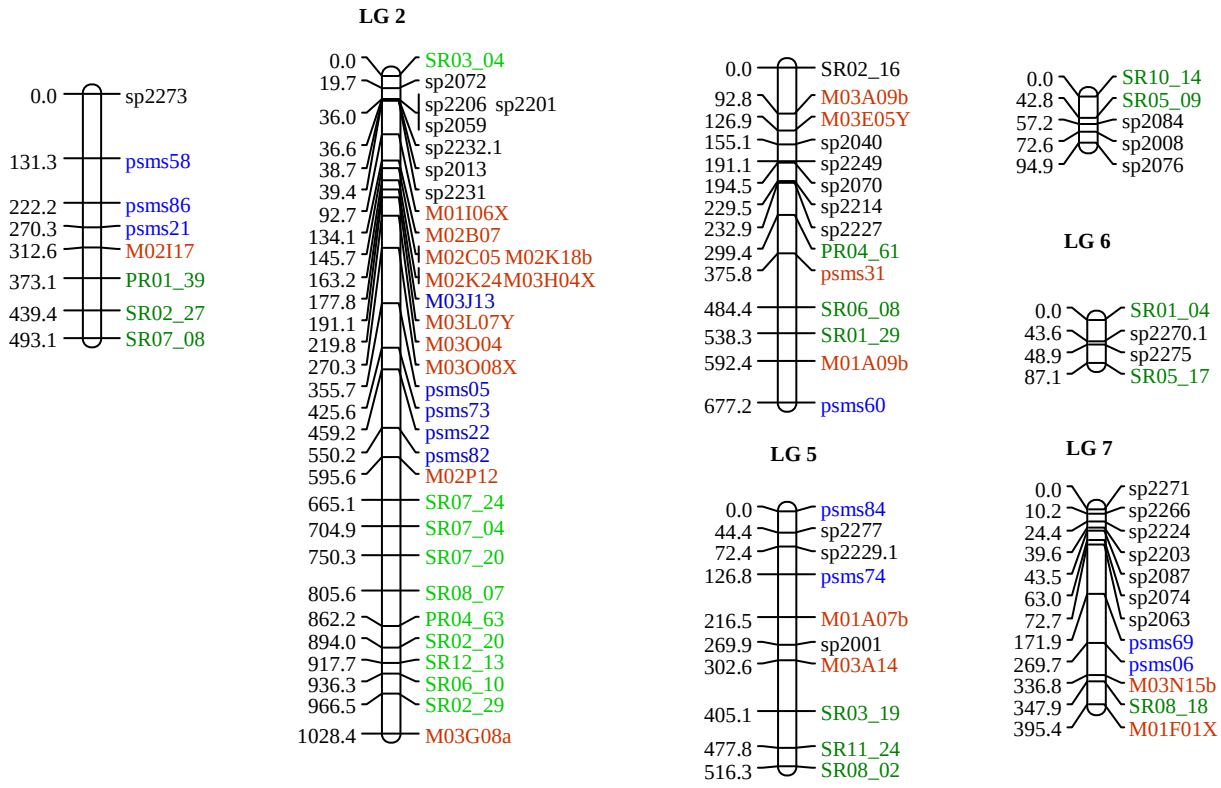


Fig 4.11: A linkage map of *Pennisetum glaucum* based on 86 loci (28 gSSRs, 25 CISPs, 13 SSCP-SNPs and 20 novel SSRs). All locus names are shown on the right side of linkage groups while inter marker distances are shown on the left side of linkage groups. The loci names in blue are markers developed from the microsatellite enriched library, loci in reddish brown are anchor markers (Qi et al. 2004) the loci in black are the intron sequence haplotypes.

SR=SRSC marker loci designed based on sorghum-rice syntenny; M=ICMM ICRISAT millet microsatellites developded in this study; PR=PRSC marker loci designed based on pearl millet-rice syntenny; SP=Xpsms, pearl millet SSR markers mapped by Qi et al. 2004; PSMS=Xpsms pearl millet SSCP-SNP markers

5. DISCUSSION

Pearl millet is one of the most drought tolerant crop species among cereals, grown in marginal and semi-arid regions of Sub-Saharan Africa and India. The past limitations associated with pedigree, morphological, physiological and cytological markers data for assessing genetic diversity in cultivated and wild plant species have largely been circumvented by the development of DNA markers. Although several DNA based marker systems and associated techniques are available their relative usefulness in particular crops is limited. To date, majority of the reports where correlation between molecular marker techniques has been found to exist are on autogamous and no reports on the cross-pollinated crop plants like pearl millet. Microsatellite or SSR markers amongst different classes of markers have been preferred for a variety of studies including genome mapping, genetic diversity studies, trait mapping in several plant species (Gupta and Varshney 2000). Although about 140 SSR markers are available in pearl millet, at present more markers are required for a successful deployment of markers in pearl millet breeding through marker-assisted selection (MAS). The available pearl millet linkage maps to date are F2 based and the consensus map developed by Qi et al. (2004) comprises 353 RFLP loci and only 65 SSR loci with large gaps towards the telomeres. Therefore, the present study was initiated for comparative assessment of various PCR-based marker systems, development, and characterization and mapping novel SSR markers for enhancing the marker density in pearl millet linkage map.

5.1 Comparative marker polymorphism

Recent advances in molecular marker technologies have enabled high-throughput, low cost markers to routinely be used to characterize germplasm and to select for favourable alleles in plant breeding programs. To obtain unbiased estimates of the genetic diversity, attention has to be paid to the choice of the marker system utilized as well as of the statistical methods. However, an ideal marker system or marker system of choice depends on the purpose of the study. The marker systems can be considered ideal provided, it is highly polymorphic, codominant, accurate, reproducible, high-throughput and low cost (both in terms of capital investment and cost per assay). Efficiency of a molecular marker technique depends upon the amount of polymorphism it can detect among the set of accessions under investigation. The genetic materials used in this study encompassed a broad spectrum of the inbred lines segregating for different biotic and a biotic stress.

5.1.1Consequences of differential marker informativeness

Although the four profiling techniques (SSR, eSSRs, ISH and ESH) are all based on DNA amplification by PCR, it was expected from the nature of the primers and reactions that the techniques would differ in the specific sequences targeted and in the number of fragments amplified. Interest then revolves around the comparison of techniques with respect to the number of unique groups of cultivars identified in each case, and in the formulation of some kind of common index reflecting the degree or level of DNA polymorphism generated in each case. Marker informativeness is an important

element when comparing different assay systems, but other factors such as cost per assay, level of skills required, reliability and reproducibility of assays should also be considered (Karp and Edwards 1997). Fluorescently based detection methods are normally more expensive than other methods, but the added advantages that these systems offer, make them very attractive.

The ability of a marker system to detect alleles at varying frequencies has applications in studying populations of individuals for germplasm analysis and maintenance, pre-screening populations for polymorphism. Marker informativeness was assessed using a number of different criteria. For instance, the number of alleles is the most basic criterion, where markers with a larger number of alleles are more likely to be polymorphic for any given germplasm set. Further, the informativeness of a genetic marker is also measured by allele frequencies, heterozygosity and polymorphic information content. All examined marker systems were able to uniquely fingerprint 22 inbred lines.

The number of observed alleles, their sizes and frequencies indicated wide variation in distribution of alleles among the inbred lines studied. The total number of alleles ranged from 44 (ESH) to 436 (ISH). The ranges of alleles detected by each marker system were 2-14, 2-8, 2-4, and 2-4, for the informative gSSRs, eSSRs, ISH, and ESH respectively (Table 4.1). The average number of alleles per polymorphic marker ranged from 2.6 (ISH) to 6.06 (gSSRs). The range in profile number (2-14) and the average number of alleles (6.06) detected by the gSSRs were higher than those reported both by

Budak et al. (2003; 2-9 alleles across 15 inbred lines) and Chandra-Shekara et al. (2007; 2.62 alleles/locus; 21 genotypes). In contrast, Mariac et al. (2006) reported 2-18 alleles (6.8 alleles per locus) in a collection of 46 wild accessions along with 421 of highly heterozygous, cross-pollinated African landraces. The reason for these differences, if statistically significant, presumably relates to the quantity and/or type of germplasm sampled. The comparison of effective number of alleles and the observed alleles at a locus explains the predominance of certain alleles. The heterozygosity levels from medium to high indicate that there is high genetic variability among the inbred lines studied (Table 4.1). Medium to high heterozygosity levels observed within the genotypes studied demonstrate that they have not lost their heterozygosity, even while living in captivity. The PIC value of a SSR marker provides an estimate of the discriminatory power of that marker locus by taking into the account not only the number of alleles that are detected but also the shared frequency of those alleles. The gSSRs were associated with the highest average PIC value (0.62), followed by the ISH (0.39) eSSRs (0.36) and ESH (0.35) (Table 4.1). The PIC value calculated for the gSSRs and intron sequence haplotypes was in line with what others have shown (Budak et al. 2003; Mariac et al. 2006; Chandra-Shekara et al. 2007).

More alleles (mean 7.7 alleles per marker), as well as a higher PIC value (0.67 per marker), were achieved using the g-SSR markers compared to the levels generated by the e-SSR markers (4.1 alleles per marker, mean PIC value 0.47). A similar result has also been reported elsewhere (Russell et al. 2004), and is attributed to the fact that e-SSRs are derived from coding sequence, which is likely to be more highly conserved than non-

coding sequence (Varshney et al. 2005a, b). However, since e-SSRs are present within genes, they tend to produce higher-quality, more robust markers, and furthermore, in contrast to g-SSRs, since they are associated with a putative function, they represent a better platform for the assessment of functional diversity (Kota et al. 2001; Eujayl et al. 2002; Thiel et al. 2003).

5.1.2 Marker polymorphism among the parents of the mapping populations

In general, the level of polymorphism depends on the degree of out-crossing behaviour of a species and genetic divergence with in a species and primers/marker system used. Of the 627 loci screened, 455 (73%) yielded a scorable amplicon. The highest success rate was achieved with gSSRs (82%) and the poorest rate was from the ESH (44%). The most informative marker type were the eSSR markers (73% of which showed at least one polymorphism across the 22 accessions) while ESH were least informative (25%) (Table 4.1). However, the eSSR loci were not associated with a high number of alleles, so their PIC value was low. As eSSR markers are developed from coding sequence, which is relatively well conserved, polymorphism is typically limited (Varshney et al. 2005a). The ESH primers also target coding sequence, and so are similarly expected to be rather non-informative. In all, 45% of the gSSR and 40% of the intron sequence haplotypes were informative, and this high proportion reflects the multi-allelic nature of the former, and the targeting of variable intronic DNA by the latter.

About 59% (268 out of 455) of the markers were informative in at least one of the populations (Table 4.2). On a per population basis, the number of polymorphic markers

ranged from 103 to 120 (mean 115). The ICMB 841-P3 × 863B-P2 population was associated with the highest number of informative markers, while Tift23D_2B_1-P5 × WSIL-P8 was associated with the least. The international reference pearl millet mapping population 81B-P6 × ICMP 451-P8 has been used to generate a 198 point genetic map (Qi et al. 2004), and some of the 118 polymorphic markers identified in the present study can be readily incorporated to enhance the density of this genetic map.

5.1.3 Implications of common polymorphic markers among populations

For many crop species, multiple maps, often constructed independently by different research groups using different sets of marker systems and mapping populations for instance, recombinanat inbred lines (RILs), near isogenic lines (NILs), double haploid (DH) lines, F_2 and Back-Cross populations that segregate for agronomically important traits, are available. This is especially important when attempting to map specific genes of interest (e.g., morphological markers or candidate genes for economically important traits) that are unlikely to segregate within a single mapping population (Diab et al. 2006). However, mapping with multiple populations provide several advantages over mapping based on a single population. It is possible to compare and integrate these different maps; as long as common subsets of markers (that serve as anchor markers) have been used among the different mapping studies. Further, integration of these maps provides a higher density of markers and greater genome coverage than is possible using a single study. Markers and mapped genes or quantitative trait loci (QTLs) that could not be mapped in one study may be placed on the basis of their relative positions in another study. Henceforth, the marker polymorphism information on different mapping

populations revealed in the current study can be utilized for development of consensus map and mapping trait QTLs for pearl millet. The present study provides 60 common polymorphic markers across two parental genotype combinations (H 77/833-2 × PRLT 2/89-33 and ICMB 841-P3 × 863-P2), 39 among three parental combinations (H 77/833-2 × PRLT 2/89-33, ICMB 841-P3 × 863-P2 and LGD 1-B-10 × ICMP 85410-P7) and 25 among four parental combinations (H 77/833-2 × PRLT 2/89-33, ICMB 841-P3 × 863-P2, LGD 1-B-10 × ICMP 85410-P7 and 81B-P6 × ICMP 451-P8). The gSSRs provided the highest number of such common markers (Table 4.4).

5.1.4 Correspondence between the marker systems

A remarkable degree of genome conservation has been established in comparative genetic mapping experiments for the Poaceae family, although genome sizes vary as much as 40-fold between some of the species, and despite the fact that they diverged as long as 60 million years ago (Gale and Devos 1998). The Mantel test was applied to compare the diversity patterns exposed by the various marker systems (Table 4.3). The marker systems were all highly inter-correlated, indicating that these genotypic data can legitimately be pooled for further analysis. In other species, EST-derived markers have been shown to be highly correlated ex., barley (Kota et al. 2001; Varshney 2005b) and rye (Varshney et al. 2007a; Khlestkina et al. 2006). Researchers have examined the existence of correlation between different molecular marker techniques in various species. Powell et al. (1996) reported similar genetic relationships in wild and cultivated soybean using four marker systems. Russel et al. (1997) found that RFLP and AFLP, but not SSR, were correlated in barley. Pejic et al. (1998) reported similar patterns of

relationships by three out of four marker techniques in maize inbred lines. RAPD and AFLP revealed similar relationships among clones of various willow species (Barker et al. 1999), *Vigna angularis* accessions (Yee et al. 1999), pepper inbreds (Lefebvre et al. 2001), apple cultivars (Goulao et al. 2001), and rice genotypes (Virk et al. 2000). The majority of the reports where correlation between molecular marker techniques has been found to exist have either been in autogamous crops or in inbred lines, where genotypes tend to be homozygous. Pearl millet is a cross-pollinated C4 cereal, is expected to have high levels of heterozygosity. If the genetic backgrounds of the accessions are too diverse, characteristic correspondence among different molecular techniques may not show up.

5.1.5 Genetic relationships among inbred lines

Detection and utilization of the genetic variation in crop plant genomes has been one of the most important tasks for plant geneticists and breeders for understanding the genome architecture and also to devise strategies for crop improvement.The average genetic similarity detected between pairs of inbred lines was consistently around 0.7 for the gene-associated markers (ISH 0.72, eSSRs 0.64, ESH 0.62), but only 0.31 for the gSSRs.

A Neighbour Joining based phylogeny was constructed based on the combined genotypic data set (Fig 1). Although the phylogenies derived from the individual marker data sets varied slightly from one another, this was thought to reflect sampling error resulting from small (marker) sample size, particularly for the exon sequence haplotypes.

The 22 inbred lines clustered into three groups, I, II and III with seven, nine and six inbred lines respectively. Of seven inbreds grouped in cluster I, four inbreds were derived from land races ex. PRLT 2/89-33, 863B-P2, LGD 1-B-10 (all three inbreds were derived from the *Iniadi* landrace from Togo and Ghana) and H 77/833-2 (derived from a Rajasthani landrace population) (Fig 4.4). However, the implied genetic relationship between H 77/833-2 and the three *Iniadi*-derived inbreds was unexpected and remains unexplained. Among other three inbreds grouped in cluster I, ICMB 841-P3 is associated with stable resistance to downy mildew and both PT 732B- P2, Tift23D_2B_1- P5, carry the d_2 dwarfing gene. Among nine members of cluster II, three were downy mildew resistant inbred lines (ICMB 90111-P6, 81B-P8 and P310-17-Bk) two were derived from LCSN 72-1-2-1-1 (ICMP 451-P6, ICMP 451-P8), two were susceptible to downy mildew (H 77/833-2-P5(NT) and W504-1-P1) IPC 804 (triple-restorer of the A_1, A_4 and A_{egp} CMS systems, with profuse pollen producing capacity) and ICMB 89111-P6 (maintainer of the A_1 cytoplasmic-genetic male-sterility system), Cluster III comprised of three downy mildew resistant inbred lines (81B-P6, IP 18293 –P152 and P 1449-2-P1), and three inbreds that possess d_1/d_2 dwarfing gene (ICMP 85410 –P7, WSIL- P8, Tift 238D_1 –P158). The clustering of both downy mildew resistant with susceptible lines, and of tall with dwarf lines, indicates that while members of a cluster may be genetically more similar to one another than to those in another cluster, the genetic differences underlying economically important traits have not contributed to the marker-based phylogeny. The genetic basis for this discordance reflects the fact that these traits are likely controlled by few genes, and thus their influence is swamped by that of the rest of the genome.

We have assessed here the relative informativeness of five marker systems, and identified sets of polymorphic markers from each. These are suitable for the *de novo* construction, or the supplementation of pearl millet linkage maps. The most informative marker type was the gSSR. We have identified sets of >100 informative markers for each of 11 mapping populations. The genetic relationships identified among the panel of inbred lines may be useful in designing strategies to improve the use of available genetic variation in the context of pearl millet breeding.

5.2 Novel Microsatellite Markers

Microsatellites are most abundant, ubiquitous, hypervariable, co-dominant and have been used for a variety of studies including genome mapping, assessing genetic relationships in several animal, insect, and plant species (Gupta and Varshney, 2000; Varshney et al. 2005a; Torada et al. 2006). The extraordinary level of informative polymorphism at a given SSR locus stems from the apparent propensity for expansion or contraction of repeat number through slippage replication or unequal crossing over events during meiosis (Goldstein and Schlötterer 1999). The identification and development of SSR markers are typically time consuming and labour intensive. Traditionally, microsatellite loci have been isolated from partial genomic DNA libraries (selected for small insert size) of the species of interest, followed by screening several thousands of clones through colony hybridization with repeat containing probes. As mentioned above, this approach can turn out to be extremely tedious and inefficient for species with low microsatellite frequencies. Therefore, several alternative microsatellite enrichment strategies have been devised for instance, streptavidin-coated magnetic beads (e.g. Kijas

et al. 1994; Fischer and Bachmann 1998; Hamilton et al. 1999), microsatellite probes attached to small nylon membranes (e.g. Karagyzov et al. 1993; Edwards et al. 1996). In order to reduce the time invested in microsatellite isolation, microsatellite enrichment procedures have been adopted in a number of plant species such as maize (Sharopova et al. 2002), wheat (Khlestkina et al. 2002), sunflower (Tang et al. 2002), and sugarbeet (Rae et al. 2002) and proven successful. In the current study a microsatellite enriched library was developed from pearl millet inbred line Tift $23D_2B_1$-P5. The library was enriched for CA, GA, CAA and AGA repeats adopting enrichment procedure suggested by Aggarwal et al. 2004.

5.2.1 Efficiency of SSR development from genomic libraries

Sequence analysis of 960 genomic DNA clones from SSR enriched library showed occurrence of 2274 SSRs in 636 sequences. The presence of SSRs in 63 % of genomic DNA clones investigated indicates an excellent SSR enrichment procedure for SSRs in the current study. In non-enriched libraries for instance, the frequency of fragments containing SSRs may be as low as 0.7%, as was reported by Chiba et al. (2003) for melon. However, in the recent past, different enrichment technologies significantly increased the efficiency of microsatellite marker development (Kijas et al. 1994; Edwards et al. 1996; Zane et al. 2002). Indeed, in several other studies, different enrichment procedures provided the SSRs at 20 % in sugarcane (Cordeiro et al. 2001), 80 % in cotton (Kumpatala et al. 2004) and 60 % in sorghum (Bhattramakki et al. 2000). Even in case of pearl millet, only 53 % of the clones from SSR enriched library

harboured SSRs (Budak et al 2003). The higher frequency of SSRs in the present study clearly indicates highly efficient procedure of capturing the SSRs.

5.2.2 Frequency and Distribution of SSRs in Pearl Millet Genome

Although AT/TA is the most common dinucleotide motif in plant genomic sequences, this motif is not usually used in SSR-enrichment procedures, owing to its self-complementary nature; therefore, we did not use AT/TA in our study. Of 1507 perfect SSRs, the motifs occurring at the highest rates were GA/TC (57.3%), CA/GT (14.3%), AAG/ TTC (16.9%), and AAT/TAA (10.3%). These results are close to those reported by Cardle et al. (2000) and Cai et al. (2003). As pointed out by Cardle et al. (2000), the most common dinucleotide motif found in plant genomic sequences is AT/TA, followed by GA/CT and CA/GT, and the most common trinucleotide motifs are AAT/ TAA and ATC/TAG. The repeat motifs AG/CT and AAC/GTT were found most abundant among the NN and NNN SSRs, respectively (Lagercrantz et al. 1993; Li et al. 2002; Morgante et al. 2002; Saha et al. 2006). Similar trend in relative frequencies of these SSR motifs were reported in earlier studies (Gupta and Varshney 2000). In the recent past many studies reported the relative abundance of microsatellite repeat motifs in cereal species for instance, barley (Thiel et al. 2003), wheat (Kantety et al. 2002), maize (Sharopova et al. 2002), sorghum (Punna et al. 2006), rice (Temnykh et al. 2001). Among the di-nucleotide motifs, GA/TC was found to be most abundant. It was interesting to note that the tri nucleotide motifs, AAC/GTT that were found to be most abundant in the present study were different from the earlier reported. In cereal species, however, the most common triplet is CCG in all the species, ranging from 32% in wheat to 49% in sorghum

(Varshney et al. 2002; Thiel et al. 2003), to 39.3% in sugarcane (Cordeiro et al. 2001). The abundance of CCG repeats is a specific feature of monocot genomes, and it may be due to their increased GC content. In plants, the most frequent triplet motif is AAG subclass (28.3%–42.1%) in *A. thaliana*, grape, and endophytes (Morgante et al. 2002). The AAT motifs were the least common (<1%) in monocot species (Cordeiro et al. 2001; Varshney et al. 2002; Thiel et al. 2003) and in other species. In transcribed regions, according to available large-scale observation in *Arabidopsis* databases, UTRs harbour more SSRs than the coding regions (Wren et al. 2000; Morgante, et al. 2002). Among different repeat unit classes, di- nucleotide (NN) SSRs were shown to have maximum relative abundance (36.5 %) followed, by compound, mono-, tri- nucleotide (NNN) SSRs. While only one SSR (0.55 %) each in case of tetra- (NNNN) and penta- nucleotide (NNNNN) categories was observed, while no SSR of hexa- nucleotide (NNNNNN) repeat units was observed. Although there are many reports on frequency of different repeat units in the SSRs isolated from genomic DNA libraries and expressed sequence tags (ESTs) in several plant species e.g. cereals like wheat, barley, rice and maize only one report on the relative abundance of repeat types is available in case of pearl millet (Senthilvel et al. 2004). As compared to work of Senthilvel et al. (2004), the relative frequencies of NN SSRs are higher in the present study. However, the source of SSRs in these two reports was different as Senthilvel et al. (2004) isolated SSRs from ESTs while in the present study SSRs were generated from the SSR enriched library. This study reaffirms the fact that genomic region contain more SSRs than the expressed regions (Wang et al. 2007).

5.2.4 Development of microsatellite markers

Out of 636 sequences containing SSRs, the primer pairs could be designed for only 69 SSRs (84.1 %). The higher success rate of primer synthesis compared to other reports indicates: (i) enough rooms are present in the flanking sequences of SSRs, (ii) quality of the sequence data was good, and (iii) good GC contents were present in the generated sequence data. Primer pairs could not be synthesized for about 48% of the SSR-containing sequences, as the microsatellite was located close to one of the ends of the sequence or flanking sequences were inappropriate for designing high quality primer pairs. Thus, primer pairs were synthesized from 331 unigene sequences only. These primer pairs were used to amplify genomic DNA of 11 pairs of pearl millet mapping population parental lines. Out of 454 primer pairs developed, 128 did not work, 17 produced multiple or non-specific fragments and 226 were functional giving simple PCR products. Most of the primer pairs that failed and the primer pairs that amplified non-scorable products had been designed for clones with compound SSRs (Table 4.8).

Thirty-five percent of the primer pairs designed for pearl millet SSRs did not amplify a product or give simple PCR products. Similarly, a larger percentage of genomic SSR primer pairs have reportedly failed to amplify in pearl millet (55%; Qi et al. 2001). The lack of amplification by some of the SSR markers has been routinely reported in several previous studies in other crops (Theil et al. 2003; Gupta et al. 2003; Yu et al. 2004; Saha et al. 2004; Varshney et al. 2006b), which were attributed to primer mismatches, the extension of primers across a splice site or the presence of large introns in the genomic DNA template. The quality of genomic sequence is therefore crucial if

one is to obtain functional primers with good amplification profiles. Nevertheless, the success rate is higher with EST-SSRs than with genomic SSRs (Senthilvel et al. 2008). In maize, it has been noted that a higher percentage of primer pairs from enriched genomic libraries failed to give consistent amplification products than from ESTsequence- derived primer pairs in maize (Sharapova et al. 2002). The authors commented that this could be due to SSR candidates being located in repeated or complex sequences in the genomic clones as opposed to lower-copy sequence origins for the EST-derived candidates. In addition, redundancy of clones within the enriched genomic libraries limited the efficiency of identifying novel SSRs. Non-specific amplifications were a problem for 6.6% of the primer pairs that were analysed. These primer pairs produced more bands than expected. A similar trend has also been observed in tall fescue and this might be due to amplifications of loci from the duplicated genomic regions (Saha et al. 2004).

Analysis of 226 markers on 22 inbred lines showed 63 polymorphic markers with at least one parental combination of eleven mapping populations. The number of polymorphic SSR markers recorded among different parental combinations in the present investigation varied from 23-45. The observed polymorphism (27.87%) is lower than anticipated polymorphism though the pearl millet is out crossing species. Nevertheless, the lower level of polymorphism indicates the lower genetic base among the elite inbred lines investigated. Another probable reason can be attributed to more recently evolved microsatellites in the pearl millet genome, sampled and analyzed in the present study. The mapping population ICMB 841-P3 × 863-P2 recorded the highest number (45) of polymorphic markers.

Sixty three polymorphic markers provided a total of 273 alleles with an average of 4.33 alleles per marker in 22 inbred lines. The PIC values for the polymorphic SSR marker were recorded in the range of 0.35 to 0.81 with an average value of 0.66 which is in congruence with the report of Budak et al. (2003). The average PIC value for 63 polymorphic markers in the present study (0.66) is higher than that of reported by Budak et al. (2003). Although a higher number of genotypes (53) were used by Budak et al. (2003) as compared to the present study (24), the higher PIC value reported in the present study may be attributed to; (i) the SSR markers of the present study are highly polymorphic, (ii) the genotypes examined here are more diverse as compared to Budak et al. (2003), and (iii) the number of markers analyzed here are less.

5.2.5 Validation of microsatellite markers for use in genetic studies

Functional primer pairs are those which amplified a fragment of the size predicted by its sequence. By corollary, the non-functional primer pairs amplify either a large number of fragments (resulting in a smear on a gel), a fragment of the wrong size, or nothing. In this study, 226 primer pairs were functional and 232 were non-functional. This high proportion of functional primers (~50%) is much higher than the ~30% functional primers previously reported (Senthilvel et al. 2004, 2008; Budak et al 2003; Qi et al. 2001; Röder et al. 1995, 1998; Bryan et al. 1997) in studies in which primer pairs were designed from clones selected from a random library. In another study, when 232 primer pairs from a microsatellite enriched library were examined, 72% were found to be functional primers (Prasad et al. 1999; Roy et al. 1999). This high frequency may be

attributed to the use of enriched libraries. In the present study, primers were designed to the sequences containing all mono-, di- and the tri-nucleotide motifs, and compound. The rate of polymorphic markers was 14%, 37%, 33%, and 22% for mono-, di-, tri-, and compound SSRs respectively. A previous investigation indicated that the [ATT/TAA]n motif was superior to all other trinucleotide repeats for the successful development of polymorphic microsatellite markers (Song et al. 2002). Although dinucleotide repeats are the most commonly used class of microsatellite markers in plants, dinucleotide-based loci often prove difficult to genotype due to a high frequency of strand-slippage artifacts (Levinson and Gutman 1987). In contrast, tri and tetra-nucleotide repeat-based markers generally produce a higher proportion of discrete PCR products as a result of a reduced level of strand-slippage artifacts (Gastier et al. 1995).

In this study totally 226 SSR markers were developed and they may be used in analysis of genetic diversity. To verify the usefulness of these markers, 226 SSR markers were used to detect genetic variation of 22 pearl millet genotypes. The average number of alleles per locus was 6.16, which was comparable to other studies, e.g. 3.9 in lychee (Viruel and Hormaza 2004), 5.7 in barley (Russell et al. 1997) and 7.38 in wheat (Prasad et al. 2000). The average PIC value in our study was 0.697, higher than that reported in soybean (0.60, Powel et al. 1996), but lower than that reported in maize (0.72, Pejic et al. 1998). For other DNA markers such as RFLP and RAPD, the PIC value ranged from 0.3 to 0.4 (Powel et al. 1996). Therefore, the SSR markers developed in the present study were informative and suitable for diversity study of pearl millet germplasm resources.

5.2.6 Microsatellite attritions

From the 454 microsatellite primer pairs developed in this work, 226 functional loci were obtained. This is an unexpectedly high attrition rate for microsatellite marker development.

In the process of SSR development, Squirrell et al. (2003) defined the successive loss of sequenced fragments and designed primers, until arriving at a final number of ‘‘working SSRs’’ producing discrete bands of expected size, as ‘‘attrition rate’’. This is an erosion of the originally high number of sequenced clones to a small final number of working SSRs. The reasons for erosion are (1) the absence of a useful SSR sequence in the clone, (2) duplications, (3) chimera formation (4) SSR sequences too close to one end of the fragment, and (5) primer pairs that produce either no fragment, or a complex banding pattern, or a smear.

However, with the accurate reporting of attrition rates at each step, the SSR development process can be further refined and improved to give greater efficiency of marker production. In our case this attrition rate was 80.2%, that is, approximately every fifth sequenced clone contained a working SSR. In different *Brassica* species, using enzyme digestion followed by colony hybridization for enrichment, Lowe et al. (2004) achieved similar enrichment efficiencies. After digesting genomic DNA with restriction enzymes in tall fescue (*Festuca arundinacea* Schreb.), Saha et al. (2006) reported 70% of clones containing SSR motifs. However, only every tenth of the sequenced clones produced a working SSR. One major reason for the large difference in efficiency to produce working SSRs in *Brassica* and *Cucurbita*, in comparison to tall fescue, may be due to genome

complexity. The size of the *Cucurbita* genome is 1C ~ 0.5 pg DNA, and that of diploid *Brassica* is slightly higher, 1C between 0.6 and 0.7 pg DNA. In contrast, the genome of tall fescue has 6.1 pg DNA and contains a large amount of repetitive sequences. Examination of the attrition rates we experienced at the various stages of SSR development is also informative. Due to the relatively high enrichment of libraries (an average of 85.8% across the libraries), attrition due to the lack of an SSR repeat was relatively low (14.2%). Implementing a screening stage to determine which clones contain SSR inserts could reduce redundancy, but such a process is time consuming. The library enrichment process could also be improved, although few enrichment techniques are able to realise the proportions achieved in this study (Zane et al. 2002), and it is likely that this process has already reached an optimum. The highest rate of attrition was due to an inability to sequence beyond the SSR repeat within a clone insert (17.2%). Improvements in PCR, sequencing and dye chemistry should help with this problem and it may be alleviated by future sequencing developments. An alternative is to sequence the reverse strand of a refractory insert, but this process requires initial screening and/or additional sequencing. A bi-directional sequencing tactic may become standard as sequencing costs reduce in the future. Another significant cause of attrition was location of the SSR repeat tract too close to the clone insertion site, thus offering little or no flanking region from which to design primers (12.8%).

5.3 First Recombinant Inbred Line (RIL) based genetic linkage map

Genetic linkage maps offer the possibility of developing genetic studies on various agronomic traits through the localization of major genes and quantitative trait loci

(QTLs), as well as helping breeding programs with marker-assisted selection (MAS). Homogenous distribution of markers and good genome coverage are two central desirable features of markers in mapping. To date, the available gentic maps for pearl millet are based on F2 populations. Nevertheless, recombinant inbred lines (RILs) can serve as powerful tools for genetic mapping as one need to genotype each line only once; one can phenotype multiple individuals from of the each line to reduce individual, environmental, measurement variability; multiple invasive phenotypes be obtained on the same set of genomes; and, as the breakpoints in RILs are more dense than those that occur in any one meiosis, greater mapping resolution can be achieved. Presence of large gaps in distal regions of the chromosomes is the characteristic feature of pearl millet linkage maps. Most of the RFLP and SSR markers that were developed to date have been mapped to centromeric regions. Past attempts to develop more markers targeting more distal regions of the pearl millet chromosomes have resulted in identification of few additional markers [Qi et al. 2001]. The present study is a first report of linkage map constructed based on RIL population of pearl millet. The map comprises 27 gSSRs, 38 (ISH; ie., 24 CISPs, 14SSCP-SNPs) and 21 novel SSRs.

5.3.1 Mappability of novel SSR markers

The new SSR markers were tested for their mappability on ICMB 841-P3 × 863-P2 linkage map. In total, 21 of the 44 new markers (47.7%) were found to be polymorphic for the notably, seven of the markers were mapped on independent LGs, which indicated the new markers to be randomly distributed on the pearl millet genome (Fig 4.6, Table 4.11).

5.3.2 Marker distance, order and distribution along the linkage groups

The distance between the markers on the map varied greatly across the different linkage groups. The average marker distance was 41.7 cM, with intervals between loci ranging from 0 to 345.4 cM (Fig. 1). The size of the LG does not necessarily reflect the number of linked markers. For instance, LG 7, with a total linkage distance of 395.4 cM had 12 mapped loci, whereas in LG 5, with a distance of 516.9 cM was covered by only 10 markers. The average map distance between new markers mapping to the ends of the LGs was comparatively high in this study. This supports the hypothesis of high recombination rates in distal regions of pearl millet chromosome postulated by previous studies (Liu et al. 1994b; Devo et al. 2000a, b; Qi et al. 2001).

Comparison of different linkage maps constructed from different populations with a different genetic background using different marker sets indicated that most markers showed the consensus order, although some intervals or regions always displayed some discrepancy in the marker order or positions. This phenomenon may be due to inversion, insertion, deletion, or transition of genomic regions as well as meiotic drive and gametic or zygotic selection. Also, possible errors in genotyping scoring may distort marker orders and segregation ratios (Kassem et al. 2006). Hayashi et al. (2001) reported that differences in locus order on a linkage map represent chromosomal rearrangements in *Lotus japonicus.* These results suggest the possibility of a chromosomal rearrangement in this region. However, the overall conservation of locus order indicates that chromosomal rearrangements have not occurred frequently in red clover.

Some SSR loci were clustered in certain genome regions (telomeric end of LG 7), which has been observed in sorghum (Bhattramakki et al. 2000) and rice (McCouch et al. 2002), possibly due to nonuniform distribution of recombination events in mapping populations (Castiglioni et al. 1999). In the present study, the novel microsatellites were not also randomly or evenly distributed on the three linkage groups of pearl millet. For instance, LG 2 and LG7 harboured 85% of the 18 SSR markers. The probable reasons included: (1) much higher polymorphism of the SSRs on these two chromosomes existed between the two mapping parents; and (2) more SSRs on these two chromosomes were isolated than those on other chromosomes. It should be pointed out that only a few SSR markers were located on some chromosomes of pearl millet. For example, only one SSR was located on LG 1 and LG6 in this study. Some researchers proposed that the use of BAC libraries for SSRs isolation rather than enriched libraries could map more SSRs in low-density regions (Cregan et al. 1999; Bhattramakki et al. 2000).

Marker distribution along the linkage groups (LG) was not uniform, as evident by the mixture of tightly linked loci and regions with low density as observed in the constructed map. This suggests that either recombination events or mapped loci were not evenly distributed throughout the genome. The low density of markers in some of the linkage groups might also correspond to regions highly homozygous and subject to higher recombination frequencies events (Castiglioni et al. 1999). Clustering or uneven distribution are usually found in linkage maps of plant species, for example, tomato (Areshchenkova and Ganal 1999), Italian ryegrass (Hirata et al. 2006), rice (McCouch et

al. 2002), barley (Ramsay et al. 1999), conifer (Elsik and Williams 2001), and sunflower (Tang et al. 2002). There are three possible explanations for the uneven distribution of markers:

1. Cardle et al. (2000) and Morgante et al. (2002) reported that microsatellites are preferentially associated with the nonrepetitive fraction of plant genomes. In *Arabidopsis thaliana*, AG/CT SSRs occur significantly more frequently in transcribed regions and 5′UTRs than in introns and 3′UTRs (Morgante et al. 2002). Moreover, sequence analysis of DNA contigs in wheat and barley has shown the presence of clusters of closely linked genes forming gene islands that are separated by large stretches of repetitive DNA (Wicker et al. 2001, 2005; Varshney et al. 2006a).
2. Homozygous regions between mapping parents may lead to gaps and insufficient coverage in the map.
3. The development of SSR markers from genomic DNA affects their distribution on the linkage maps. Ramsay et al. (1999) reported that SSRs isolated from genomic DNA libraries tend to concentrate in retrotransposons and dispersed repetitive element DNA.

5.3.3 Segregation distortion

Segregation can defined as a deviation of the observed genotypic frequencies from their expected values, violates the law of segregation and renders conventional genetic theory and analysis to be invalid (Lu et al. 2002). A variety of genetic factors could result in segregation distortion, including pollen tube competition, pollen lethal,

preferential fertilization, and selective elimination; the first three types were defined as gametic selection. In many plant species, the most commonly reported genetic factors associated with the distorted segregation ratio were at the gamete level. At present, segregation distortion has been reported in maize (Lu et al. 2002; Sibov et al. 2003), rice (Matsushita et al. 2003), alfalfa (Kalo´ et al. 2000), coffee (Ky et al. 2000). In the present study, the distortion ratios of the tested markers, as well as for mapped loci, were different. Segregation distortion was observed across seven linkage groups of pearl millet. And moreover, all marker types employed, in the current study, for map construction exhibited segregation distortion indicating the distortion is not due to the marker type but because of the underlying genetic factors. In a separate study, Sentilvel et al. (2008) observed skewed distribution of eSSRs (11 out of 17) towards either of the parents. Previously such kind of segregation distortion has often been noted in case of pearl millet. For instance on LG 3 and LG 6 of the ICMB 841-P3 × 863B-P2 cross (Yadav et al. 2004; Bidinger et al. 2007) with an excess of ICMB 841 alleles. The level of segregation distortion observed in this study (58%) is within the range reported in plant molecular studies (Schon et al. 1993; Lin et al. 1996; Wang et al. 1998). High segregation distortion observed in the present study may be due to out crossing nature of pearl millet. Similar results have been reported outcrossing species (Kubisiak et al. 1995; Hanley et al. 2002; Dettori et al. 2001; Liebhard et al. 2002). In rice and maize, the segregation distortion ratio of SSR markers in mapping populations ranged from 10 to 23% (Xu et al. 1997). This may associate with mapping population type (the DH population and RIL population usually show higher distortion ratios), selection against closely linked lethal or sub-lethal genes, and linkage with incompatibility alleles, and

man-made errors etc. In addition, some studies reported that during hybridization between genetically distant parents (e.g. cultivated and wild relatives), the alleles of wild species could be frequently lost, which led to severe distorted segregation (Xu et al. 1997). The proportion of loci showing significant segregation varies greatly with the species, population type and specific cross.

A part from above mentioned, the previous research reports indicated that the segregation distortion ratio in interspecific population was higher than in intraspecific population. For instance in case of sorghum, 40% and 27% of all markers showed significant segregation distortion in interspecific and intraspecific backcross population respectively (Gebhardt et al. 1991). Similarly, high rates of segregation distortion in interspecific populations have been reported in tomato (Xu et al. 1997), tetraploid cotton (Lin et al. 2005), *G. hirsutum* (Shappley et al. 1998) than in intaspecific populations. In this study, 77.42% markers showed segregation distortion, which was higher than that previously reported. Liu et al. (2000) and Zhang et al. (1997) used the same parents, cultivated variety Changnong4 and semi-wild variety Xinmin2, to construct F2 and RIL populations respectively, and found that the main reason of segregation was gametic selection. Yan et al. (2003) and Zhao et al. (1997) drew the same conclusion in maize and rice respectively. In this study, the allele frequency of most codominant markers skewed toward the female genotype, and the reason for segregation distortion may have been influenced by some gametophytic genes.

5.3.4 Excess heterozygosity and map expansion

The theory of the process by which excess heterozygosity leads to map expansion is based on the relationship between Mendelian segregation ratios, recombination frequency, and crossover number. This relationship is the fundamental principle on which linkage mapping is based (Mendel 1866; Morgan 1911; Sturtevant 1913; Haldane and Waddington 1931). The persistence of heterozygotes predicted by Haldane and Waddington (1931), for different crosses, has been taken into account in genetic mapping in animal and plant systems (Taylor 1978; Burr *et al*. 1988; Burr and Burr 1991). In a given meiosis four possible gametes can be derived from one chiasma exchange between four chromatids. Two of these will be parental and two will be recombinant for markers either side of the chiasma. Thus one chiasma corresponds to 0.5 recombinants or 50 cM. For each marker in an F2 population it is expected that the segregation will be 1:2:1, but if the ratio is distorted to 1: h: 1 (where h>2) then the excess number of chiasma in F2 is [2h/ (h+2)]-1. So for example if the segregation ratio is 1:3:1 then the fraction of heterozygotes is 0.6 and the proportion of recombinants will be increased by a factor of 1.2 ie., an excess of 0.2 chiasma. Therefore 10 such segments are equal to 2 extra chiasma or 100 cm more added to the map length. Where the excess heterozygote ratio persists during the development of an RI population the effect on genetic map length will be even greater. If the recombination fraction in a small interval is r in the F2 then in the RI the proportion of observed recombinants will be $r(h+2)/2$. This suggests that map expansion, resulting from the persistence of heterozygotes, is misrepresented by the map expansion function of Haldane and Waddington (1931) when F2 segregation ratios are distorted.

In a nutshell segregation distortion is a characteristic of most markers and mapping systems and has a major effect on map length and marker order. From this current study it is evident that excess heterozygosity can lead to genetic map expansion especially in RI mapping populations. Excess map length is discomforting, but the excess is from an expectation that assumes that markers are immune from segregation distortion, unexpected heterozygosity and nongenetic effects such as DNA methylation. Although the first RIL based linkage map was constructed in this study, there is still some work which is needed to be done in the future. Particularly, more markers should be added to the linkage map which can be useful in QTL locating since only 69 markers were mapped here, leading to the average map distance of 51.1 cM and some big gaps between the markers.

6. SUMMARY

Pearl millet [*Pennisetum glaucum* (L.) R. Br] is a major crop in the semi-arid, low input dryland regions of Africa and Southeast Asia. It grows rapidly and is adapted for growing in most areas, including many harsh arid climates with poor acidic, high aluminum soils unsuitable for wheat, rice, maize, sorghum and barley. It is most widely grown type of millet accounting for about 50 % of total world's production.

To supplement the conventional breeding activities aimed at crop improvement, genomics tools provides the opportunities for assaying and utilizing the genetic variation through molecular breeding approaches. In this regard, at present, microsatellite or SSR markers are the preferred class of molecular markers for plant breeding applications. Though, about 140 SSR markers are currently available in pearl millet, there is strong need to develop more SSR markers. The present study was carried out for comparative marker polymorphism assessment and for mining and characterization of SSRs from microsatellite enriched library, validation of SSR markers, polymorphism assessment and mapping of the novel SSRs along with the intron sequence haplotypes and genomic SSRs

The salient features of this study are as follows:

1. A set of 22 pearl millet inbred lines including the parents of eleven mapping populations, was screened with 627 markers including 100 pearl millet genomic SSRs (gSSRs), 60 pearl millet EST-SSRs (eSSRs), 410 intron sequence haplotypes (ISH), and 57 exon sequence haplotypes (ESH). In all, 267 (59%) of

the markers were informative for at least one of the 11 mapping populations. An average of 116 polymorphic markers was identified per mapping population.

2. The average PIC values and numbers of profiles (*P*) per polymorphic marker were: gSSRs (PIC = 0.62, *P* = 6.1), ISH (PIC = 0.39, *P* = 2.6), eSSRs (PIC = 0.36, *P* = 3.1) and ESH (PIC = 0.35, *P* = 3.1). A high correlation ($r > 0.89$, $p < 0.005$) was observed between the patterns of diversity exposed by the different marker systems.
3. The most informative marker type was the gSSR analyzed on PAGE. The polymorphic markers identified are suitable for trait mapping and genetic relationships studied among the inbred lines may be useful for harnessing genetic variation in pearl millet breeding.
4. The polymorphic markers identified are suitable for trait mapping and genetic relationships studied among the inbred lines may be useful for harnessing genetic variation in pearl millet breeding
5. Microsatellite enriched gDNA library was constructed from the pearl millet genotype "Tift $23D_2B_1$-P5". In the first instance, gDNA library was picked in ten 96-well plates. In total a set of 1152 putative positive clones were picked from the microsatellite library enriched for CA, GA, CAA and AGA repeats. On sequencing 960 putative positive clones 636 sequences were found to possess the SSRs
6. Among different classes of SSRs, di-nucleotide repeat motifs were found to be the most abundant (830; 36.5%) followed by compound (737; 32.4%), mono-(370; 16.3%) and tri-nucleotide SSRs (305; 13.4%)

7. Among dinucleotide AG/CT repeats and similarly ACC/GTT among trinucleotide repeats were found to be the most abundant repeat motif in the microsatellites enriched sequences examined
8. A total of 454 primer pairs were designed of which 226 produced scorable amplicons. However, 69 markers were found polymorphic across of the panel of 22 pearl millet inbre lines studied. The PIC values for the polymorphic SSR marker were recorded in the range of 0.35 to 0.81 with an average value of 0.66 which is in congruence with the report of Budak et al. (2003).
9. The present study is a first report of linkage map constructed based on RIL population of pearl millet. The map comprises 27 gSSRs, 38 (ISH; ie., 24 CISPs, 14 SSCP-SNPs) and 21 novel SSRs.
10. In total, 21 of the 44 new markers (47.7%) were found to be polymorphic for the notably, seven of the markers were mapped on independent LGs, which indicated the new markers to be randomly distributed on the pearl millet genome
11. The length of the linkage groups ranged from 87.1 cM (LG 6) to 1328.4 cM (LG2).
12. The size of the LG does not necessarily reflect the number of linked markers. For instance, LG 7, with a total linkage distance of 395.4 cM had 12 mapped loci, whereas in LG 5, with a distance of 516.9 cM was covered by only 10 markers.

7. REFERENCES

Abate AN, Gomez M (1984) Substitution of finger millet (*Eleusine coracana*) and bulrush millet (*Pennisetum typhoides*) for maize in broiler feeds. Anim Feed Sci Tech 10:291–299

Aggarwal RK, Velavan TP, Udaykumar D, Hendre S, Shanker K, Choudhury B C and Singh L (2004) Development and characterization of novel microsatellite markers from the olive ridley sea turtle (*Lepidochelys olivacea*). Mol Eco Notes 4:77–79

Allouis S, Qi X, Lindup S, Gale MD, Devos KM (2001) Construction of a BAC library of pearl millet, *Pennisetum glaucum*. Theor Appl Genet 102:1200–1205

Amato SV, Forrester RR (1995) Evaluation of pearl millet as a feed ingredient for broiler rations. In: Teare ID (ed), Proc. 1st Nat. Grain Pearl Millet Symp., Univ. Georgia, Tifton. pp 125–128

Areshchenkova T, Ganal MW (1999) Long tomato microsatellites are predominantly associated with centromeric regions. Genome 42:536–544

Barker JH, Matthes M, Arnold GM, Edwards KJ, Ahman I, Larsson S, Karp A (1999) Characterization of genetic diversity in potential biomass willows (*Salix* spp.) by RAPD and AFLP analyses. Genome 42:173-83

Bertin I, Zhu JH, Gale MD (2005) SSCP–SNP in pearl millet– a new marker system for comparative genetics. Theor Appl Genet 110:1467–1472

Bhat KV, Babrekar PP, Lakhanpaul S (1999) Study of genetic diversity in Indian and exotic sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. Euphytica 110:21–33

Bhattacharjee R, Bramel PJ, Hash CT, Kolesnikova–Allen MA, Khairwal IS (2002) Assessment of genetic diversity within and between pearl millet landraces. Theor Appl Genet 105:666–673

Bhattramakki D, Dong JM, Chhabra AK, Hart GE (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 43:988–1002

Bidinger FR, Hash CT (2005) Pearl millet. In: Nguyen HT and Blum A (eds) Physiology and biotechnology integration for plant breeding. Marcel Dekker, New York. pp. 225–270

Bidinger FR, Nepolean T, Hash CT, Yadav RS, Howarth CJ (2007) Quantitative trait loci for grain yield in pearl millet under variable post flowering moisture conditions. Crop Sci 47:969–980

Bonnin I, Ronfort J, Wozniak F, Olivieri I (2001) Spatial effects and rare outcrossing events in *Medicago truncatula* (Fabaceae). Mol Ecol 10:1371–1383

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bottley A, Xia GM, Koebner RMD (2006). Homoeologous gene silencing in hexaploid wheat. The Plant J 47:897–906

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z, Lin Y-R, Liu S-C, Luo L, Marler BS, Ming R, Mitchell SE, Qiang D, Reischmann K, Schulze SR, Skinner DN, Wang Y-W, Kresovich S, Schertz KF, Paterson AH (2003) A high–density genetic recombination map of

sequence–tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367–386

Breese WA, Hash CT, Devos KM, Howarth CJ (2002) Pearl millet genomics – an overview with respect to breeding for resistance to downy mildew. In: Leslie JF (ed). Sorghum and Millets Pathology, Iowa State Press: Ames, Iowa, USA, pp 243–246

Brunken JN (1977) A systematic study of *Pennisetum* sect. *Pennisetum* (Gramineae). Am J Bot 64:161–176

Bryan GJ, Collins AJ, Stephenson P, Orry A, Smith JB, Gale MD (1997) Isolation and characterisation of microsatellites from hexaploid bread wheat. Theor Appl Genet 94:557–563

Budak H, Pedraza F, Cregan PB, Baenzinger PS, Dweikat I (2003) Development and utilization of SSRs to estimate the degree of genetic relationships in a collection of pearl millet germplasm. Crop Sci 43:2284–2290

Burr B, Burr FA (1991) Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. Trends in Genet 7:55-60

Burr B, Burr FA, Thompson KH, Albertson MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519-526

Burton G (1995) History of hybrid development in pearl millet. In: Proc. First Grain Pearl Millet Symp. Univ. of Georgia, Tifton. 5–8.Tifton, GA. pp17–18

Burton GW, Wallace AT, Rachie KO (1972) Chemical composition and nutritive value of pearl millet. Crop Sci 12:187–188

Busso CS, Devos KM, Ross G, Mortimore M, Adams WM, Ambrose MJ, Alldrick S, Gale MD (2000) Genetic diversity within and among landraces of pearl millet (*Pennisetum glaucum*) under farmer management in West Africa. Genetic Res Crop Evol 47: 561–568

Cai HW, Yuyama N, Tamaki H, Yoshizawa A (2003) Development and characterization of simple sequence repeat (SSR) markers in hexaploid forage grass timothy (*Phleum pratense* L). Theor Appl Genet 107:1337–1349

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 156:847–854

Castiglioni P, Ajmone–Marsan P, van Wijk R, Motto M (1999) AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. Theor Appl Genet 99:425–431

Chahal SS, Kumar K, Gill GS (1986) Identity and variation in two isolates of *Moesziomyces penicillariae*, the causal organism of pearl millet smut. Plant Disease Res 1:59–64

Chandra–Shekara AC, Prasanna BM, Bhat SR, Singh BB (2007) Genetic diversity analysis of elite pearl millet inbred lines using RAPD and SSR markers. J Plant Biochem Biotech 16:23–28

Chiba N, Suwabe K, Nunome T, Hirai M (2003) Development of microsatellite markers in melon (*Cucumis melo* L.) and their application to major Cucurbit crops. Breeding Science 53:21–27

Chowdari KV, Davierwala AP, Gupta VS, Ranjekar PK, Govila OP (1998) Genotype identification and assessment of genetic relationships in pearl millet [*Pennisetum glaucum* (L.) R. Br.] using microsatellites and RAPDs. Theor Appl Genet 97:154–162

Collins VP, Cantor AH, Pescatore AJ, Straw ML, Ford MJ (1997) Pearl millet in layer diets enhances egg yolk "n–32" fatty acids. Poult Sci 76:326–330

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp) ESTs across transferable to *Erianthus* and *Sorghum*. Plant Sci 160:1115-1123

Cregan PB, Mudge J, Fickus EW, Marek LF, Danesh D, Denny R, Mathews BF, Jarvik T, Young ND (1999) Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. Theor Appl Genet 98:919–928

Davis AJ, Dale NM, Ferreira FJ (2003) Pearl millet as an alternative feed ingredient in broiler diets. J Appl Poult Res 12: 137–144

Davis AJ, Dale NM, Ferreira FJ (2003) Pearl millet as an alternative feed ingredient in broiler diets. J Appl Poult Res 12:137–144

de Carvalho A, de O. Soares DJ, do Carmo M.G.F, da Costa ACT, Pimentel C (2006) Description of the life-cycle of the pearl millet rust fungus–*Puccinia substriata* var. *penicillariae* with a proposal of reducing var. *indica* to a synonym. Mycopathologia 161:331-336

Dettori MT, Quarta R, Verde I (2001) A peach linkage map integrating RFLPs, SSRs, RAPDs, and morphological markers. Genome 44:783–790

Devos KM, Gale MD (2000a) Genome Relationships: The Grass model in current research. The Plant Cell 12:637–646

Devos KM, Pittaway TS, Busso CS, Gale MD, Witcombe JR, Hash CT (1995) Molecular tools for the pearl millet nuclear genome. Int Sorghum and Millets Newsl 36:64–66

Devos KM, Pittaway TS, Reynolds A, Gale MD (2000b) Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. Theor Appl Genet 100:190–198

Diab AA (2006) Construction of barley consensus map showing chromosomal regions associated with economically important traits. African J Biotech 5: 235–248

Dida M, Srinivasachary, Ramakrishnan S, Bennetzen JL, Gale MD, Devos KM (2007) The genetic map of finger millet, *Eleusine coracana.* Theor Appl Genet 114:321–332

Dinakar Bhattramakki, Jianmin Dong, Ashok K, Chhabra, Hart GE (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 43:988–1002

DNA Baser Sequence Assembler v2.x (2009) HeracleSoftware, http://www.DnaBaser.com

Doebley J, Stec A (1993) Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. Genetics 134: 559–570

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nature 386:485-488

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:758

Ejeta G, Hassen MM, Mertz ET (1987) *In vitro* digestibility and amino acid composition of pearl millet (*Pennisetum typhoides*) and other cereals. Proc Natl Acad SciUSA 84:6016–6019

Elsik CG, Williams CG (2001) Families of clustered microsatellites in a conifer genome. Mol Genet Genomics 265:535–542

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Fancher BI, Jensen LS, Smith RL, and Hanna WW (1987.) Metabolizable energy content of pearl millet. Poultry Sci. 66:1693–1696

FAO and ICRISAT (2006) The World Sorghum and Millet Economies: Facts, Trends and Outlook. Food and Agriculture Organization of the United Nations (FAO): Rome, Italy, and International Crops Research Institute for the Semi–Arid Tropics (ICRISAT), Patancheru, India. pp 72

Feltus FA, Singh HP, Lohithaswa HC, Schulze SR, Silva TD, Paterson AH (2006) A comparative genomics strategy for targeted discovery of single–nucleotide polymorphisms and conserved–noncoding sequences in orphan crops. Plant Physiol. 140:1183–1191

Fischer D, Bachmann K (1998) Microsatellite enrichment in organisms with large genomes (*Allium cepa L.*). BioTechniques 24:796-802

Gale M, Devos KM (1998) Plant comparative genetics after 10 years. Sci 282:656–659

Gale MD, Witcombe JR (1992) DNA markers and marker–mediated applications in plant breeding, with particular reference to pearl millet breeding. In: Biotechnology and Crop crop Improvement improvement in Asia Moss JP (Ed.). ICRISAT, Patancheru, Hyderabad, India, pp 323–332

Gastier MJ, Pulido JC, Suden S, Brody T, Kenneth HB, Murray JC, Weber JL, Hudson TJ, Sheffield VC, Duyk GM (1995) Survey of trinucleotide repeats in the human genome: assessment of their utility as genetic markers. Hum Mol Genet 4:1829–1836

Gebhardt C, Ritter E, Barone A. (1991) RFLP maps of potato and their alignment with the homoeologous tomato. Genome 83: 49–57

Goldstein DB, Schlötterer C (1999) Microsatellites: Evolution and Applications. Oxford University Press; Oxford, New York

Goulao L, Cabrita L, Oliveira CM, Leitao JM (2001) Comparing RAPD and AFLP analysis in discrimination and estimation of genetic similarities among apple (*Malus domestic* Borkh.) cultivars. Euphytica 119:259–270

Gupta PK, Balyan HS, Edwards KJ, Isaac P, Korzun V, Roder M, Joudrier P, Schlatter AR, Dubcovsky J, de la Pena RC, Khairallah M, Hayden M, Keller B, Wang R, Hardouin JP, Jack P, Leroy P (2002b) Genetic mapping of sixty six new SSR loci in bread wheat (*Triticum aestivum*). Theor Appl Genet 105:412–422

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST–SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 270:315–323

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Gupta PK, Varshney RK, Prasad M (2002a) Molecular markers: principles and Methodology. In: Jain SM, Brar DS, Ahloowalia BS (eds.), Molecular Techniques in Crop Improvement, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 9–54

Haldane JBS, Waddington CH (1931) Inbreeding and linkage. Genetics 16:357-374

Hamilton MB, Pincus EL, Di-Fiore A, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. Biotechniques 27:500-507

Hanley S, Barrer JHA, Van Ooijen JW, Aldam C, S.L. Harris, I. Ahman, S. Larsson & A. Kart, 2002. A genetic linkage map of willow (*Salix viminalis*) based on two *Lycopersicon esculentum* × *L. pennellii* F_2 populations. Theor Appl Genet 99:254–271

Hanna WW (1996) Improvement of millets—Emerging trends. 139– 146. In: Proc. 2nd Int. Crop Sci. Congress, Delhi, India. 17–23 Oxford and IBM Publ., New Delhi, India.

Hash CT, Bhasker Raj AG, Lindup S, Sharma A, Beniwal CR, Folkertsma RT, Mahalakshmi V, Zerbini E, Blümmel M (2003) Opportunities for marker–assisted selection (MAS) to improve the feed quality of crop residues in pearl millet and sorghum. Field Crops Res 84:79–88

Hash CT, Bramel–Cox P (2000) Marker applications in pearl millet *In* Application of molecular markers in plant breeding. B.I.G. Haussmann, H.H. Geiger, D.E. Hess, C.T. Hash and P. Bramel–Cox (Eds). Training manual for a seminar held at IITA, Ibadan, Nigeria, from 16–17 August 1999. International Crops Research Institute for the Semi–Arid Tropics (ICRISAT), Patancheru 502324, Andhra Pradesh, India.

Hash CT, Witcombe JR (1994) Pearl millet mapping populations at ICRISAT. p. 69–75 *in* J.R. Witcombe and R.R. Duncan (eds.): Use of molecular markers in sorghum and pearl millet breeding for developing countries. Overseas Development Administration: London, U.K.

Hayashi M, Miyahara A, Sato S, Kato T, Yoshikawa M, Taketa M, Hayashi M, Pedrosa A, Onda R, Imaizumi–Anraku H, Bachmair A, Sandal N, Stougaard J, Murooka Y, Tabata S, Kawasaki S, Kawaguchi M, Harada K (2001) Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F2 population. DNA Res 8:301–310

Henzell RG, Peterson GC, Teetes GL, Franzmann BA, Sharma HC, Youm O, Ratnadass A, Toure A, Raab J, Ajayi O (1997) Breeding for resistance to panicle pests of sorghum and pearl millet. Pages 255–280 *in* Proceedings of the International Conference on Genetic Improvement of Sorghum and Pearl Millet, Sep 22–27, 1996, Holiday Inn Plaza, Lubbock, Texas. INTSORMIL: Lincoln, NE, USA

Hill GM, Hanna WW (1990) Nutritive characteristics of pearl millet grain in beef cattle diets. J Anim Sci 68:2061–2066

Hirata M, Cai H, Inoue M, Yuyama N, Miura Y, Komatsu T, Takamizo T, Fujimori M (2006) Development of simple sequence repeat (SSR) markers and construction of an SSR–based linkage map in Italian ryegrass (*Lolium multiflorum* Lam.). Theor Appl Genet 113:270–279

Hittalmani S, Foolad MR,Mew T, Rodriguez RL, Huang N (1995) Development of a PCR–based marker to identify rice blast resistance gene, *Pi–2(t)*, in a segregating population. Theor Appl Genet 91:9–14

Hoseney RC, Andrews DJ, Clark H (1987) Sorghum and pearl millet. In: Nutritional quality of cereal grains: Genetic and agronomic improvement. ASA, Monograph. 28:397–456

Howarth CJ, Skot KP, Weltzien E, Bidinger FR (1997) Development and assessment of methods for evaluating seedling thermotolerance in pearl millet. In: Proceedings of the International Conference on Genetic Improvement of Sorghum and Pearl Millet. 22-27 September 1996. Lubbock, TX. INTSORMIL University of Nebraska, Publ. 97-5 pp 651-652

Huang X, Madan A (1999) CAP3: A DNA Sequence Assembly Program. Genome Res 9:868–877

Jones ES, Breese WA, Liu CJ, Singh SD, Shaw DS, Witcombe JR (2002) Mapping quantitative trait loci for resistance to downy mildew in pearl millet: Filed and greenhouse screens detect the same QTL. Crop Sci 42:1316–1323

Jones ES, Liu CJ, Gale MD, Hash CT, Witcombe JR (1995) Mapping quantitative trait loci for downy mildew resistance in pearl millet. Theor Appl Genet 91:448–456

Kalo´ P, Endre G, Zima´nyi L and Csana´di G (2000) Construction of an improved linkage map of diploid alfalfa (*Medicago sativa*). Theor Appl Genet 100:641–657

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum, and wheat. Plant Mol Biol 48:501–510

Karagyozov L, Kalcheva ID, Chapman VM (1993) Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. Nucleic Acids Res 21:3911- 3912

Karp A, Edwards K (1998) DNA markers: a global overview. In: Caetano-Annolés G. Gresshoff PM (Eds.), DNA Markers: Protocols, Applications and Overviews, pp. 1–14. Wiley-VCH, New York.

Kassem MA, Shultz J, Meksem K, Cho Y, Wood A J, Iqbal M J, Lightfoot DA (2006) An updated 'Essex' by 'Forrest' linkage map and first composite interval map of QTL underlying six soybean traits. Theor Appl Genet 113:1015–1026

Khairwal IS, Hash CT (2007) "HHB 67–improved" –the first product of marker–assisted crop breeding in India. Asia–Pacific Consortium on Agricultural Biotechnology. http://www.apcoab.org/special_news.html.

Khlestkina EK, Pestsova EG, Salina E, Roder MS, Arbuzova VS, Koval SF, Borner A (2002) Genetic mapping and tagging of wheat genes using RAPD, STS and SSR markers. Cell Mol Biol Lett 7:795–802

Khlestkina EK, Than MHM, Pestsova EG, Röder MS, Malyshev SV, Khorzum V, Börner A (2004) Mapping of 99 new microsatellite–derived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. Theor Appl Genet 109:725–732

Khlestkina EK, Varshney RK, Röder MS, Graner A, Börner A (2006) Comparative assessment of genetic diversity in cultivated barley collected at different periods of the last century in Austria, Albania and India by using genomic and genic SSR markers. Plant Genet Res 4:125–133

Kijas JMH, Fowler JCS, Garbett CA, Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. BioTechniques 16: 657-662

Konuma A, Tsumura Y, Lee CT, Lee SL, Okuda T (2000) Estimation of gene flow in the tropical–rainforest tree *Neobalanocarpus heimii* (Dipterocarpaceae), inferred from paternity analysis. Mol Ecol 9:1843–1852

Kosambi DD (1944). The estimation of map distances from recombination values. Ann Eugen 12:172-175

Kota R, Varshney RK, Thiel T, Dehmer KJ, Graner A (2001) Generation and comparison of EST derived SSR and SNP markers in barley (*Hordeum vulgare* L.). Hereditas 135:141–151

Kubisiak TL, Nelson CD, Nance WL, Stine M (1995) RAPD linkage mapping in a longleaf pine × slash pine F1 family. Theor appl Genet 90:1119–1127

Kumpatla SP, Manley MK, Horne EC, Gupta M, Thompson SA (2004) An improved enrichment procedure to develop multiple repeat classes of cotton microsatellite markers. Plant Mol Biol Rep 22:85a–85i

Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Louarn J, Charrier A, Hamon S, Noirot M (2000) Interspecific genetic linkage map, segregation distortion and genetice conversion in coffee (*Coffea* sp.). Theor Appl Genet 101:669–676

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21:1111–1115

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) Mapmaker an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lefebvre V, Goffinet B, Chauvet JC, Caromel B, Signoret P, Brand R, Palloix A (2001) Evaluation of genetic distances between pepper inbred lines for cultivar protection purposes: comparison of AFLP, RAPD and phenotypic data. Theor Appl Genet 102:741–750

Levinson GL, Gutman GA (1987) High frequencies of short frame shifts in poly–CA/TG tandem repeats borne by bacteriophageM13 in *Escherichia coli* K12. Nucleic Acids Res 15:5323–5338

Li Y, Du J, Wang T, Shi Y, Song Y, Jia JJ (2002) Genetic diversity and relationships among Chinese maize inbred lines revealed by SSR markers. Maydica 47: 93–101

Liebhard R, Gianfrancesschi L, Koller B, Ryder CD, Tarchini R, Van de Weg E, Gessler C (2002) Development and characterization of 140 new microsatellites in apple (*Malus domestica* Borkh). Mol Breed 10:217–241

Lin ZX. (2005) Linkage maps construction in cotton and QTL mapping for yield and fiber–related traits [Dissertation]. Huazhong Agricultural University, (in Chinese).

Lincoln S, Daly M, Lander ES (1992) Constructing genetic maps with MAPMAKER/EXP 3.0. Whitehead Institute Technical Report. 3rd ed. Whitehouse Technical Institute, Cambridge, Mass.

Liu CJ, Witcombe JR, Hash CT, Busso CS, Pittaway TS, Nash M, Gale MD (1994a) In: Use of molecular markers in sorghum and pearl millet breeding for developing countries. Witcombe JR and Duncan RR (Eds): Overseas Development Administration: London, U.K. pp 57–64

Liu CJ, Witcombe JR, Pittaway TS, Nash M, Hash CT, Busso CS, Gale MD (1994b) An RFLP–based genetic map of pearl millet (*Pennisetum glaucum*). Theor Appl Genet 89:481–487

Liu CJ, Witcombe JR, Pittaway TS, Nash M, Hash CT, Gale MD (1992) In: Restriction fragment length polymorphism in pearl millet, *Pennisetum glaucum*. Mounolou JC, Sarr A (Eds). Complexes d'espèces, flux de gènes et resources génétiques des plantes. Colloque international, 8–10 janvier, Publications du Bureau des Ressources Génétiques, Paris. pp. 233–241

Liu F, Wu XL, Chen SY (2000) Segregation distortion of molecular markers in recombinant inbred populations in soybean (*G. max*). Acta Genet Sin 27:883−887

Loveless AR (1967) *Claviceps fusiformis* sp. nov., the causal agent of an aglactia of sows. Transactions of the British Mycological Society 50:15-18

Lowe AJ, Moule CL, Trick M, Edwards KJ (2004) Efficient large–scale development of microsatellites for marker and mapping applications in *Brassica* crop species. Theor Appl Genet 108:1103–1112

Lu H, Romero–Severson J and Bernardo R (2002) Chromosomal regions associated with segregation distortion in maize. Theor Appl Genet 105:622–628

Lui Z, Furnier GR (1993) Comparison of allozyme, RFLP and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. Theor Appl Genet87:97–105

Mace ES, Buhariwalla HK, Crouch JH (2003) A high throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol Biol Rep 21:459a–459h

Mahender T, Hash CT, Senthilvel S, Bottley A, Koebner RMD, Paterson AH, Reddy AR, Hoisington DA, and Varshney RK (2007) Polymorphism Assessment Of SSR, SNP And CISP Marker Systems In Pearl Millet (*Pennisetum glaucum* (L.) R. Br.). Plant & Animal Genomes XV Conference (http://www.intlpag.org/15/abstracts/PAG15_P05d_373.html)

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Can Res 27:209–220

Mariac C, Luong V, Kapran I, Mamadou A, Sagnard F, Deu M, Chantereau J, Gerard B, Ndjeunga J, Bezancon G, Jean–Louis P, Vigouroux Y (2006) Diversity of wild and cultivated pearl millet accessions (*Pennisetum glaucum* [L.] R. Br.) in Niger assessed by microsatellite markers. Theor Appl Genet 114:49–58

Martel E, De Nay D, Siljak–Yakovlev S, Brown S, Sarr A (1997) Genome size variation and basic chromosome number in pearl millet and fourteen related *Pennisetum* species. J Hered 88: 139–143

Matsushita S, Iseki T, Fukuta Y, Araki E, Kobayashi S, Osaki M, Yamagishi M (2003) Characterization of segregation distortion on chromosome 3 induced in wide

hybridization between *indica* and *japonica* type rice varieties. Euphytica 134:27–32

McCouch SR, Teytelman L, Xu YB, Lobos KB, Clare K, Walton M, Fu BY, Maghirang R, Li ZK, Xing YZ, et al. (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

Mendel G (1866) Versuche über Pflanzen-Hybriden. Verhandlungen des Naturforschenden Vereins. Brünn 4:3-47

Milbourne D, Meyer R, Bradshaw JE, Baird E, Bonar N, Provan J, Powell W, and Waugh R, (1997) Comparison of PCR based marker systems for the analysis of genetic relationships in cultivated potato. Mol Breed 3:127–136

Mishra RN, Palakolanu Sudhakar Reddy, Suresh Nair, Gorantla M, Reddy AR, Sopory S K, Reddy MK (2007) Isolation and characterization of expressed sequence tags (ESTs) from subtracted cDNA libraries of *Pennisetum glaucum* seedlings. Plant Mol Biol 64:713–732

Money TA, Liu CJ, Gale MD (1994) Conversion of RFLP markers for downy mildew resistance in pearl millet to sequence–tagged sites. In: Witcombe JR and Duncan RR (eds.): Use of molecular markers in sorghum and pearl millet breeding for developing countries. Overseas Development Administration: London, U.K. pp 65–68

Morgan TH, (1911) An attempt to analyze the constitution of the chromosomes on the basis of sex-limited inheritance in *Drosophila.* J Exp Zool 11:365-413

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with non repetitive DNA in plant genomes. Nat Genet 30:194–200

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Mullis KB and Faloona FA (1987) Specific synthesis of DNA in vitro via the polymerase catalysed reaction. Meth Enzymol 255:335–350

Naik S, Gill KS, Rao VS, Gupta VS, Tamhankar SA, Pujar S, Gill BS & Ranjekar PK (1998) Identification of a STS marker linked to the *Aegilops speltoides*–derived leaf rust resistance gene *Lr28* in wheat. Theor Appl Genet 97:535–547

Nair S, Kumar A, Srivastava MN, Mohan M (1996) PCR–based DNA markers for Gm4t gene that confers resistance against gall midge and their potential in marker–aided selection in rice. Theor Appl Genet 92:660–665

Negi MS, Devic M, Delseny M, Lakshmikumaran M (2000) Identification of AFLP fragments linked to seed coat colour in *B. juncea* and conversion to SCAR marker for rapid selection. Theor Appl Genet 101:146–152

Ouendeba B, Ejeta G, Hanna WW, Kumar AK (1995) Diversity among African pearl millet landrace populations. Crop Sci 35:919–924

Pejic I, Ajmone Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. Theor Appl Genet 97:1248–1255

Peng Y, Liang YS, Wang SQ, Wu FQ, Li SC (2006) Analysis of segregation distortion of SSR markers in the RI population of rice. Mol Plant Breed 6:786−790

Perrier X, Flori A, Bonnot F (2003) Data analysis methods, In: Genetic diversity of cultivated tropical plants Hamon M et al (Eds), CIRAD/Science, Montpellier, France. pp 43–76

Peterson GC, Reddy BVS, Youm O, Teetes GL, Lambright L (1997) Breeding for resistance to foliar– and stem–feeding insects of sorghum and pearl millet. In: Proceedings of the International Conference on Genetic Improvement of Sorghum and Pearl Millet, Sep 22–27, 1996, Holiday Inn Plaza, Lubbock, Texas. INTSORMIL: Lincoln, NE, USA, pp. 281–302

Pilat–Andre, JS, Lamy F, Sarr A, (1992) Etude de la région Adh du mil (*Pennisetum typhoides*) par utilisation des techniques RFLP. In: Mounolou, JC, Sarr A (Eds), Complexes d'espèces, flux de gènes et ressources génétiques des plantes. Colloque international, 8–10 janvier, 1992, Publications du Bureau des Ressources Génétiques, Paris. pp. 58–71

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–248

Prasad M, Kumar N, Kulwal PL, Roder M, Balyan HS, Dhaliwal HS, Gupta PK (2002) QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. Theor Appl Genet 106:659–667

Prasad M, Kumar N, Kulwal PL, Roder M, Balyan HS, Dhaliwal HS, Gupta PK (2002) QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. Theor Appl Genet 106:659-67

Prasad M, Varshney RK, Kumar A, Balyan HS, Sharma PC, Edwards KJ, Singh H, Dhaliwal HS, Roy JK, Gupta PK (1999) A microsatellite marker associated with a QTL for grain protein content on chromosomal arm 2DL of bread wheat. Theor Appl Genet 99:341–345

Prasad M, Varshney RK, Roy JK (2000) The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. Theor Appl Genet 100:584–592

Prashanth SR, Parani M, Mohanty BP, Talame V, Tuberosa R, Parida A (2002) Genetic diversity in cultivars and landraces of *Oryza sativa* subsp. *indica* as revealed by AFLP markers. Genome 45:451–459

Punna R, Bantte K, Ashok Kumar Ch, Jayashree B, Folkertsma RT, Senthilvel S, Mahalakshmi V, Ananda Reddy L, Fakrudin B, Hash CT (2006) Development and mapping of EST-SSR markers in sorghum for comparative mapping with rice. Plant and Animals Genome Conference XIV, January 14-18, 2006 San Diego, CA
http://www.intl-pag.org/14/abstracts/PAG14_P200.html

Qi X, Lindup S, Pittaway TS, Allouis S, Gale MD and Devos KM (2001) Development of simple sequence repeat markers from bacterial artificial chromosomes without subcloning. BioTechniques 31:355–361

Qi X, Pittaway TS, Lindup S, Liu H, Waterman E, Padi FK, Hash CT, Zhu J, Gale MD, Devos KM (2004) An integrated genetic map of pearl millet, *Pennisetum glaucum*. Theor Appl Genet 109:1485–1493

Rae SJ, Aldam C, Dominguez I, Hoebrechts M, Barnes SR, Edwards KJ (2002) Development and incorporation of microsatellite markers into the linkage map of sugar beet (*Beta vulgaris* spp.). Theor Appl Genet 100:1240–1248

Ramsay L, Macaulay M, Cardle L, Ivanissevich SD, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J 17:415–425

Röder MS, Plaschke J, Konig SU, Borner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellite wheat. Mol Genet Genomics 246:327–333

Roder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Rooney LW, McDonough CM (1987) Food quality and consumer acceptance in pearl millet. Proc. Int. Pearl Millet Workshop. ICRISAT, Patancheru, India. pp 43–61

Roy JK,Prasad M, Varshney RK, Balyan HS, Blake TK, Dhaliwal HS, Singh H Edwards KJ and Gupta PK (2002) Identification of a microsatellite on chromosome 5B and a STS on 6D of bread wheat showing association with preharvest sprouting tolerance. Theor Appl Genet 88:225-239

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (Eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ pp 365–386

Rudd S, Mewes HW, Mayer K F X (2003) Sputnik: a database platform for comparative plant genomics Nucleic Acids Res 31:128-132

Russel JR, Fuller JD, Macaulay M, Hatz BGB, Jahoor A, Powell W, Waugh R (1997) Direct comparisons of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714–722

Russell J, Booth A, Fuller J, Harrower B, Hedley P *et al*. (2004). A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley. Genome 47:389-398

Saha MC, Cooper JD, Rouf Mian MA, Chekhovskiy K, May GD (2006) Tall fescue genomic SSR markers: development and transferability across multiple grass species. Theor Appl Genet 113:1449–1458

Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang L, May D (2004) Tall fescue EST–SSR markers with transferability cross several grass species. Theo Appl Genet 109:783–791

Senthilvel S, Jayashree B, Mahalakshmi V, Sathish Kumar P, Nakka S, Nepolean T, Hash CT (2008) Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags. BMC Plant Biology 8:119

Senthilvel S, Mahalakshmi V, Satish Kumar P, Reddy AR, Markandeya G, Reddy MK, Misra R, Hash CT (2004) New SSR markers for pearl millet from data mining of expressed sequence tags. In: Proc 4th International Crop Sci Cong, Brisbane, Australia, 26th Sept.–1st Oct. 2004, pp 232

Shappley ZW, Jenkins JN, Meredith R, McCarty JC (1998) An RFLP linkage map of Upland cotton (*Gossypium hirsutum* L.). Theor Appl Genet 97: 756−761

Sharapova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia-Hancock S, Musket T, Duru N, Polacco M, Edwards K, Ruff T, Register JC, Brouwer C, Thompson R, Velasco R, Chin E, Lee M, Woodman-Clikeman W, Long MJ, Liscum E, Cone K, Davis G, Coe EH (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48:463–481

Sharma R, Aggarwal RAK, Kumar R, Mohapatra T, Sharma RP (2002) Construction of an RAPD linkage map and localization of QTLs for oleic acid level using recombinant inbreds in mustard (*Brassica juncea*). Genome 45:467–472

Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez- Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia- Hancock S, Musket T, Duru N, Polacco M, Edwards K, Ruff T, Register JC, Brouwer C, Thompson R, Velasco R, Chin E, Lee M, Woodman-Clikeman W, Long MJ, Liscum E, Cone K, Davis G, Coe EH (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48:463–481

Sibov ST, de Souza CL Jr, Garcia AAF, Silva AR, Mangolin CA, Benchimol LL, de Souza AP (2003) Molecular mapping in tropical maize (Zea mays L.) using microsatellite markers. 1. Map construction and localization of loci showing distorted segregation. Hereditas 139:96–106.

Singh SP (1987) Genetic divergence over environments in pearl millet (*Pennisetum typhoides* Stapf & Hubb.). Genetica Agraria 42:35–41.

Smith HO and Wilcox KW (1970) A restriction enzyme from *Hemophilus influenzae*. Purification and general properties. J Mol Biol 51:379–392

Song QJ, Fickus EW, Cregan PB (2002) Characteristics of trinucleotide markers in wheat. Theor Appl Genet 104:286–293

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol 12:1339–1348

Srivastava A, Gupta V, Pental D, Pradhan AK (2001) AFLP–based genetic diversity assessment amongst agronomically important natural and some newly synthesized lines of *Brassica juncea*. Theor Appl Genet 102:193–199

Stapf and Hubbard CE (1934) *Pennisetum.* In: Flora of tropical Africa (Praln D, ed). Kent, England: L Reeve & Co. pp 954–1070

Sturtevant AH (1913) The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. J Exp Zool 14:43-59

Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105:1124–1136

Taylor BA (1978) Recombinant inbred strains: use in gene mapping, In: *Origins of Inbred Mice*, H. C. Morse (ed). Academic Press, New York. pp 423–438

Tegelstrom H (1992) Detection of mitochondrial DNA fragments In: Hoelzel AR, (eds) Molecular Genetic Analysis of Populations: A Practical Approach, IRL Press, Oxford, 1992, pp 89–114

Temnykh S, DeClerck G, Lukashova A et al. (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

Thiel T, Michalek W, Varshney RK, 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

Torada A, Kike M, Mochida K, Ogihara Y (2006) SSR–based linkage map with new markers using an intra specific population of common wheat. Theor Appl Genet 112:1042–1051

Tostain S (1994) Isozymic classification of pearl millet (*Pennisetum glaucum*, Poaceae) landraces from Niger (West Africa). Plant Syst Evol 193:81–93

Tostain S, Marchais L (1989) Enzyme diversity in pearl millet (*Pennisetum glaucum*) 2. Africa and India. Theor Appl Genet 77:634–640

Tostain S, Riandey MF, Marchais L (1987) Enzyme diversity in pearl millet (*Pennisetum glaucum*) 1. West Africa. Theor Appl Genet 74:188-193

van Deynze AE, Sorrells ME, Park WD, Ayres NM, Fu H, Cartinhour SW, Paul E, McCouch SR (1998) Anchor probes for comparative mapping of grass genera. Theor Appl Genet 97:356–369

Varshney RK, Beier U, Khlestkina E, Korzun V, Graner A, Börner A (2007a) Single nucleotide polymorphisms in rye (*Secale cereale* L.): discovery, frequency and applications for genome mapping and diversity studies. Theor Appl Genet 114:1105–1116

Varshney RK, Beier U, Khlestkina E, Kota R, Korzun V, Graner A, Börner A (2007) Single nucleotide polymorphisms in rye (*Secale cereal* L.): discovery, frequency and application for genome mapping and diversity studies. Theor Appl Genet 114:1105–1116

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

Varshney RK, Grosse I, Hahnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmied L, Graner A (2006a) Genetic mapping and BAC assignment of EST–derived SSR markers shows non–uniform distribution of genes in the barley genome. Theor Appl Genet 113:239–250

Varshney RK, Hoisington DA, Tyagi AK (2006b) Advances in cereal genomics and applications in crop breeding. Trends Biotech 24:490–499

Varshney RK, Prasad M, Graner A (2004) Molecular marker maps of barley, a resource for intra– and interspecific genomics. In: Molecular markers in improvement of agriculture and forestry'. GWenzel, Horst L (Eds) Springer Verlag, Germany, pp 229–243

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A (2005b) Interspecific transferability and comparative mapping of barley EST–SSR markers in wheat, rye and rice. Plant Sci 168:195–202

Varshney RK, Thiel T, Sretenovic–Rajicic T, Baum M, Valkoun J, Guo P, Grando S, Ceccarelli S, Graner A (2008) Identification and validation of a core set of informative genic SSR and SNP markers for assaying functional diversity in barley. Mol Breed 22: 1–13

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546

Virk PS, Zhu J, Bryan GJ, Jackson MT, Ford–Lloyd BV (2000) Effectiveness of different classes of markers for classifying and revealing variation in rice (*Oryza sativa*) germplasm. Euphytica 112: 275–284

Viruel MA, Hormaza JI (2004) Development, characterization and variability analysis of microsatellites in lychee (*Litchi chinensis* Sonn., Sapindaceae). Theor Appl Genet 108:896–902

Vogel JM, Rafalski A, Powell W, Morgante M, Andre C, Hanafey M, Tingey SV (1996) Application of genetic diagnostics to plant genome analysis and plant breeding. Hort Sci 31:1107–1108

vom Brocke K, Christinck A, Presterl T, Weltzien RE, Geiger HH (2003) Effects of farmers' seed management on performance and adaptation of pearl millet in Rajasthan, India. Euphytica 130:267–280

Wang Dan, Liao X, Cheng L, Yu X, Tong J (2007) Development of novel EST-SSR markers in common carp by data mining from public EST sequences.Acquaculture 271:558-574Wang ZM, Devos KM, Liu CJ et al (1998) Construction of RFLP–based maps of foxtail millet, *Setaria italica* (L.) P. Beauv. Theor Appl Genet 96:31–36

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contigous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26:307–316

Wicker T, Zimmermann W, Perovic D, Paterson AH, Ganal M, Graner A, Stein N (2005) A detailed look at 7 million years of genome evolution in a 439 kb contiguous

sequence at the barley Hv-eif4e locus: recombination, re-arrangements, and repeats. Plant J 41:184–194

Wilson JP, Burton GW, Zongo JD, Dicko IO (1990) Diversity among pearl millet landraces collected in central Burkina Faso. Crop Sci 30:40–43

Wren JD, Forgacs E, Fondon JW, Pertsemlidis A, Cheng SY, Gallardo T, Williams RS, Shohet R V, Minna JD, Garner HR (2000) Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. Am J Hum Genet 67:345-356

Xiao–Ping J, Yun–Su S, Yan–Cun S, Guo–Ying Wang, Tian–Yu Wang, Yu Li (2007) Development of EST–SSR in foxtail millet (*Setaria italica*). Genet Resour Crop Evol 54:233–236

Xu Y, Zhu L, Xiao J, Huang N, McCouch SR (1997) Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled–haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). Mol Gen Genet 53: 535−545

Yadav OP, Mitchell SE, Zamora A, Fulton TM, Kresovich S (2007) Development of new simple sequence repeat markers for pearl millet. An Open Access J published by ICRISAT. http://www.icrisat.org/Journal/mpii/v3i1/Sorgu_Millet%20_other_Cereals/ISMN%2022_Development%20of%20new...%20(OPYadav%20et%20al.).pdf

Yadav RS, Bidinger FR, Hash CT, Yadav YP, Yadav OP, Bhatnagar SK, Howarth CJ (2003) Mapping and characterization of QTL × E interactions for traits determining grain and stover yield in pearl millet. Theor Appl Genet 106:512–520

Yadav RS, Hash CT, Bidinger FR, Devos KM, Howarth CJ (2004) Genomic regions associated with grain yield and aspects of post–flowering drought tolerance in pearl millet across stress environments and tester background. Euphytica 136:265–277

Yadav RS, Hash CT, Bidinger FR, Dhanoa MS, Howarth CT (1999a) Identification and utilisation of quantitative trait loci (QTLs) to improve drought tolerance in pearl millet (*Pennisetum glaucum* (L.) R. Br.). Forth coming from CIMMYT: El Batan, Mexico.

Yadav RS, Hash CT, Bidinger FR, Howarth CJ (1999b) QTL analysis and marke r assisted breeding of traits associated with drought tolerance in pearl millet. Forth coming from IRRI: Los Banos, Philippines.

Yadav RS, Hash CT, Cavan GP, Bidinger FR, Howarth CJ (2002) Quantitative trait loci associated with traits determining grain and stover yield in pearl millet under terminal drought stress conditions. Theor Appl Genet 104:67–83

Yan JB, Tang H, Huang YQ, Zheng YL, Li JS (2003) Genetic analysis of segregation distortion of molecular markers in maize F2 Population. Acta Genet Sin 30:913−918

Yee E, Kidwell KK, Sills GR, Lumpkin TA (1999) Diversity among selected *Vigna angularis* (Azuki) accessions on the basis of RAPD and AFLP markers. Crop Sci 39:268–275

Yu J–K, La Rota M, Kantety RV, Sorrells ME (2004) EST derived SSR markers for comparative mapping in wheat and rice. Mol Genet Genomics 271:742–751

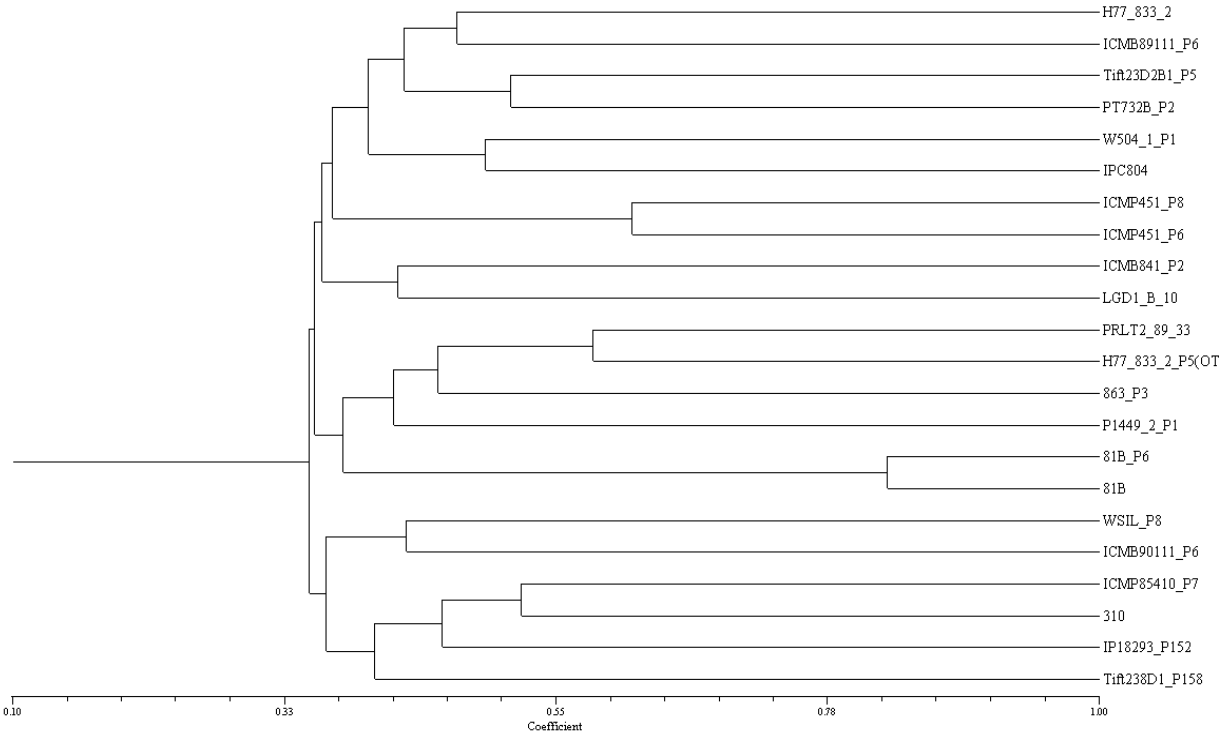
Zane L, Bargelloni L, Patarnello T (2002) Statregies for microsatellite isolation: a review. Mol Ecol 11:1–16

Zeigler RS, Leong SA, Teng PS (1994) Rice Blast Disease. CAB International, Oxon, UK

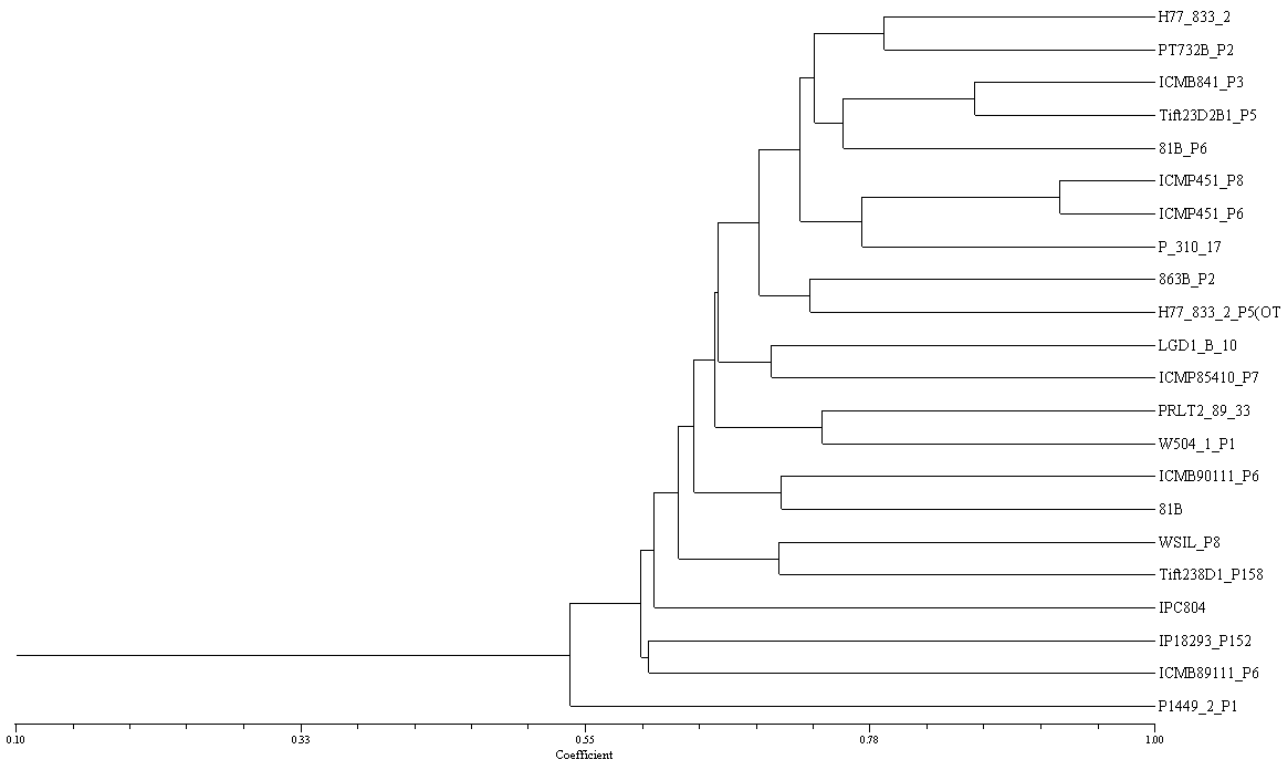
Zhang DS, Cheng SY, Hui DW, Zhuang PC (1997) Segregation distortion of RFLP markers in F2 population of cultivated /semi–wild soybean and the case analysis. Acta Genet Sin 24:62−67

Zhao XP, Kochert G (1993) Phylogenetic distribution and genetic mapping of a (GGC) microsatellite from rice (*Oryza sativa* L.). Plant Mol Biol 21:607–614

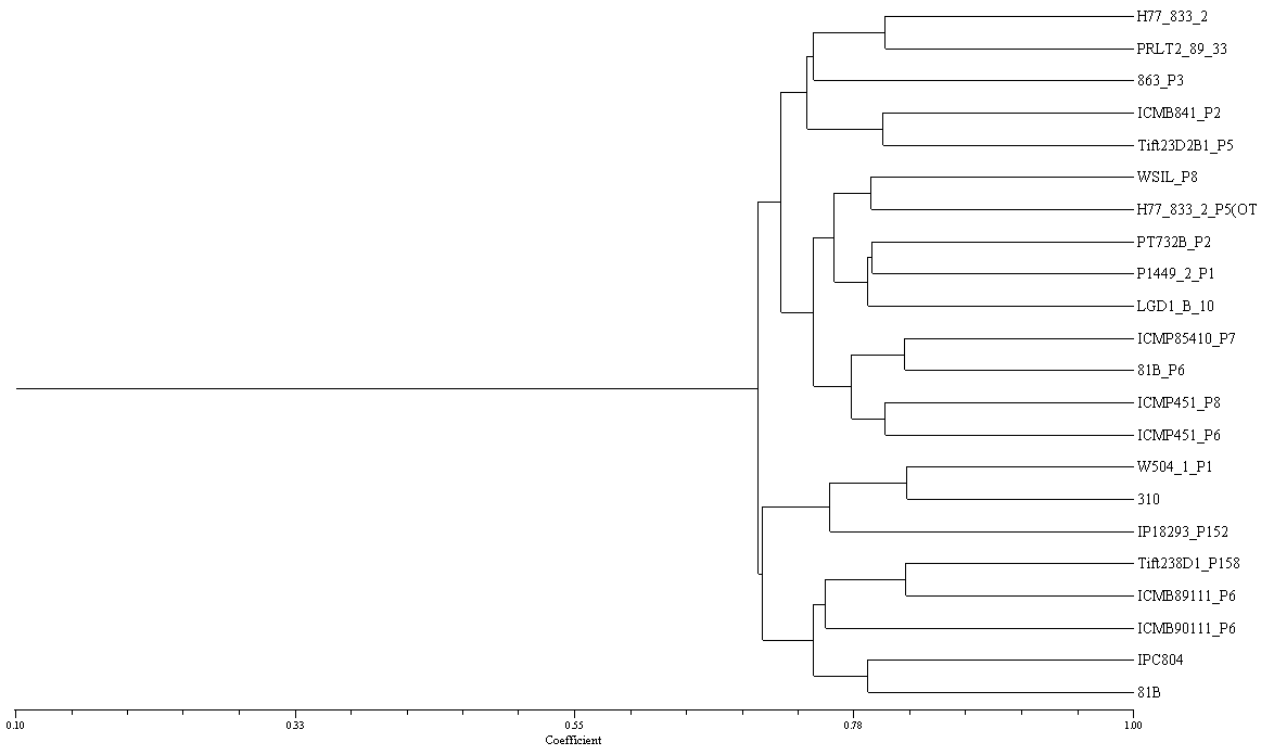
APPENDIX A



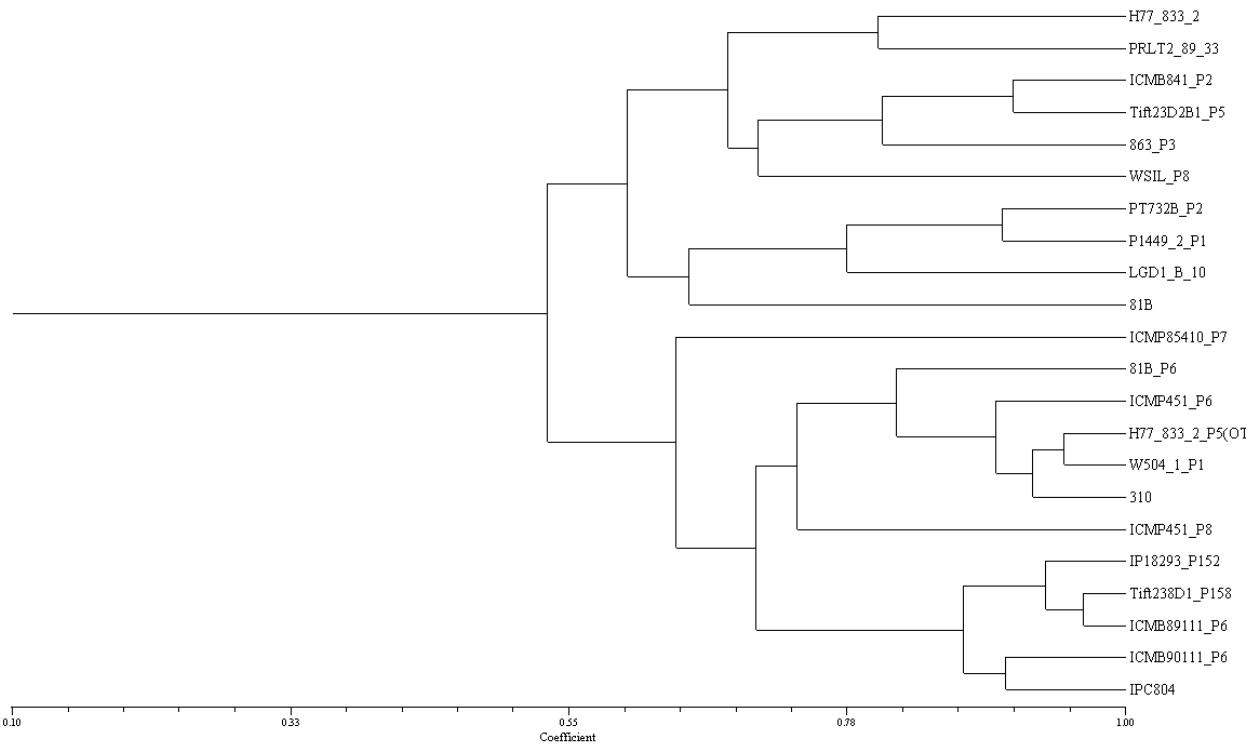
Genetic relationships among 22 pearl millet inbred lines using gSSR marker data



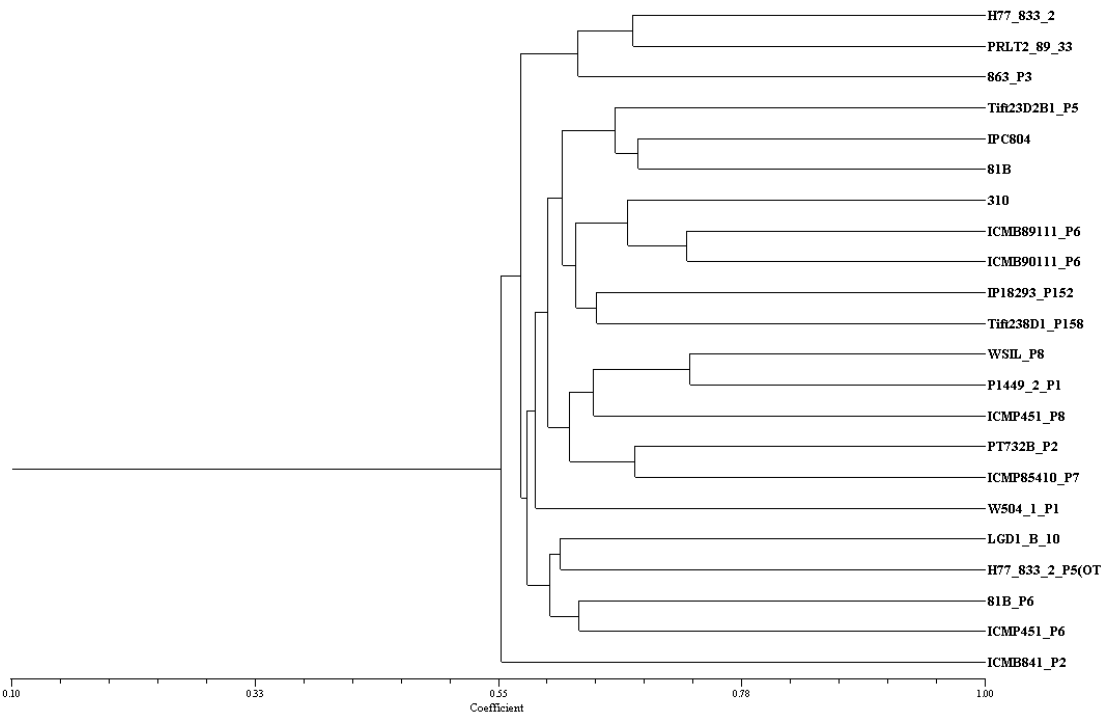
Genetic relationships among 22 pearl millet inbred lines using eSSR marker data



Genetic relationships among 22 pearl millet inbred lines using PM_SSCP-SNP marker data



Genetic relationships among 22 pearl millet inbred lines using W_SSCP-SNP marker data



Genetic relationships among 22 pearl millet inbred lines using CISP marker loci data