DEVELOPMENT AND APPLICATIONS OF EST-SSR MARKERS IN SORGHUM

THESIS SUBMITTED TO

OSMANIA UNIVERSITY FOR THE AWARD OF

DOCTOR OF PHILOSOPHY

IN GENETICS

PUNNA RAMU

DEPARTMENT OF GENETICS

OSMANIA UNIVERSITY, HYDERABAD

2009

ABSTRACT

Microsatellites or simple sequence repeat (SSR) markers are playing a significant role in molecular breeding programs. But, conventional methods of SSR markers development are laborious and expensive. The alternative way is to explore the existing databases for the development of new markers. Complete genome sequencing projects depositing numerous amounts of genomic and genic (expressed sequence tag, EST) sequence in the public databases. This facilitates to explore for the development of new marker systems including SSRs, SNPs, etc.

In the present study, 600 EST-SSR markers were developed in sorghum by mining the J. Craig Research Institute [formerly, The Institute for Genome Research (TIGR)] database. Non-redundant EST sequences which are having the homology with rice genome were selected for the designing these 600 markers (by exploiting rice-sorghum synteny). PCR conditions for these primer pairs were optimized on four parental pairs of sorghum genotypes constituting two mapping populations. About 23% of markers were found polymorphic in parents of shootfly resistance mapping population (BTx623 × IS18551). These polymorphic markers were screened against the shootfly RIL mapping population. A total of 82 EST-SSR markers were mapped onto the previously existing skeleton map of the sorghum using shootfly resistance mapping population. After adding these 82 new EST-SSR markers, the map distance was extended to 2966.4 cM. This map distance was ~250 cM larger than the previously reported longest linkage map in sorghum. These EST-SSR markers, however, filled gaps or marker-rare regions in the existing linkage map of sorghum targeted based on rice-sorghum sequence similarity. This study demonstrates the utility of comparative genomic information in targeted development of markers to fill gaps in linkage maps of related crop species for which sufficient genomic tools are not available.

During the course of this study, complete sorghum genome sequence is available. Aligned sorghum genome sequence was downloaded (Sorbi v1.0) (http://genome.jgi-psf.org/Sorbi1/Sorbi1.info.html) and made as a database. SSR markers (available and unavailable for public use) were used to develop *in-silico* sequence-based physical maps for sorghum for all linkage groups. The order of EST-SSR markers from the *in-silico* mapping agrees with the linkage map positions for all ten sorghum linkage groups. The markers which had conflicting linkage and physical mapping positions can be attributed to the presence of duplicate loci or the presence of best hit for forward and reverse primer sequences on the different chromosomes. Upon integration of sorghum physical and linkage maps, it was found that 1 cM was equal to 223 kb (in other words, 1Mb = 4.47 cM). This indicates that if recombination rates are similar across the genome (which is not likely), then if 4–5 markers are present in each bin (1 Mb) the map can be considered as saturated, which is essential for map-based cloning. Integration of physical maps and genetic maps provides the bridge to isolate the genes through positional or map-based cloning. This physical map facilitated the researchers to select more number of markers in a given genomic region for effective transfer of QTLs from donor parent to locally improved varieties through markers-assisted selection (MAS).

A reference set of sorghum consisting 384 accessions representing all botanical races available in sorghum was selected for diversity analysis in the present study. This

reference set (representing the maximum diversity) was assembled from global composite germplasm collection of sorghum. A set of 45 EST-SSR markers having reasonable genome coverage on sorghum (based on their linkage map positions) were selected for the diversity analysis. In total, 40 markers were used in the analysis after deleting poor quality markers. These accessions were clustered primarily by race within the geographic origin in harmony with previous studies. In previous studies, bicolor race was identified as the most heterogeneous. But in the present study, ESR-SSR markers clearly identified two major distinct groups within bicolor. The margaritiferum group with in the guinea race (Gma) formed as separate cluster from the guinea race accessions and closely associated with wild genotypes suggesting independent domestication of this Gma group. The markers used in this study for diversity have discriminating power similar to that of SSR markers previously used to constitute the sorghum reference set. A few landraces from bicolor race and several intermediate race accessions were clustered with wild accessions suggesting gene flow between landraces and wild sorghums because of their co-cultivation in Africa.

This reference set represents well the diversity present in the sorghum global composite germplasm collection and can serve as an international public good not only as an entry point to global collections for breeders seeking variation in a particular trait of interest, but also for allele mining and association mapping. Proper phenotyping of the reference set along with genotyping data using these functional markers is expected to facilitate the association mapping in sorghum provided that the phonological variation (plant height and flowering time) present is not so great that it interferes with phenotyping for target traits of interest.

CERTIFICATE

This is to certify that the thesis entitled **"Development and applications of EST-SSR markers in sorghum"**, submitted for award of the degree of Doctor of Philosophy in Genetics, Osmania University, is a record of the bonafide research carried out by **Mr. Punna Ramu,** under my supervision, and no part of the thesis has been submitted for any other degree or diploma.

The assistance and help taken during the course of this investigation and the sources of literature referenced have been fully acknowledged.

Date:

Place: Hyderabad

(L. ANANDA REDDY)

Supervisor,
Associate Professor,
Dept. of Genetics,
Osmania University,
Hyderabad – 500 007**.**

(C. T. Hash)

Co-supervisor,
Principal Scientist,
Molecular Breeding,
AGL, ICRISAT,
Patancheru – 502 324.

DECLARATION

I hereby declare that the research work presented in this thesis entitled "**Development and applications of EST-SSR markers in sorghum**", has been carried out by me at the Department of Genetics, Osmania University, Hyderabad under the supervision of Dr. L. Ananda Reddy and at ICRISAT, Patancheru, Andhra Pradesh, India under the co-supervision of Dr. C.T. Hash.

This work is original and no part of the thesis has been submitted earlier for the award of any degree or diploma of any university.

Date: **(Punna Ramu)**

Place: Hyderabad

ACKNOWLEDGMENTS

*I am sincerely obliged and indebted to **Dr. L. Ananda Reddy**, Associate Professor, Department of Genetics, Osmania University, Hyderabad, and supervisor, for his sagacious advice and able direction throughout my research project. I am very thankful to for extending all possible help, constitutive suggestions and guidance and support during my project work.*

*Fervently and modesty, I extol the genuine co-operation and inspiration offered to me by **Dr. C. Tom Hash**, Principal Scientist, Molecular Breeding, ICRISAT, Patancheru, for providing research facilities in Applied Genomics, Laboratory (AGL) and for his principles of emphasizing clear writing and his uncompromising standards of accuracy and kind back-up during the process of achieving the final form of dissertation. It would be only a minuscule gesture on part to express my gratitude for his encouragement, support, punctual and judicious guidance and sustenance. It was indeed a rare privilege for me to work under his emending inspiration and indomitable spirit.*

*With respect regards and immense pleasure, I wish to acknowledge and express sincere thanks from my heart to **Dr. S. Senthilvel**, Scientist, Centre of Excellence in Genomics, ICRISAT, Patancheru, for his valuable counsel, assistance, and valuable advice rendered for successful completion of my project work at ICRISAT.*

*It gives me gratification in expressing my heartfelt gratitude to **Dr. Rajeev K. Varshney**, Principal Scientist, Applied Genomics Laboratory, ICRISAT, for his kind co-operation and help rendered during my research project at ICRISAT.*

*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 Genetics, Osmania University, Hyderabad for their kind help and co-operation.*

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 Genetics, Osmania University, Hyderabad for their kind co-operation in successful completion of my research work.

*It was the utmost blessings and good wishes of my parents, **Satyanarayana** and **Ramanarsamma**, affection of my sisters, **Ramadevi** and **Ramana**, and my loving wife, **Prashanthi,** that I could attain academic heights to accomplish my doctoral degree. And I express my deepest adoration to them for teaching me the etiquette of life.*

The Council of Scientific and Industrial Research (CSIR), New Delhi, India is greatly acknowledged for providing the financial assistance to carryout my research project.

I convey my whole hearted thanks to many of my well wishers and other friends requesting their forgiveness for not mentioning them here by name.

Place: Hyderabad **(Punna Ramu)**

CONTENTS

ABBREVIATIONS

˚C	:	degree Celsius
µl	:	microliter
AFLP	:	Amplified Fragment Length Polymorphism
B	:	race Bicolor
BLAST	:	Basic Local Alignment Search Tool
bp	:	base pair
C	:	race Caudatum
cDNA	:	complementary DNA
CISP	:	Conserved Intron Scanning Primers
cM	:	centiMorgan
COS	:	Conserved Orthologous Set
D	:	race Durra
DArT	:	Diversity Array Technology
DNA	:	Deoxyribonucleic Acid
EMBL	:	European Molecular Biology Laboratory
EST	:	Expressed Sequence Tag
G	:	race Guinea
GCP	:	Generation Challenge Program
Gma	:	sub-race Guinea margaritiferum
GRD	:	Genetic Resource Division
HPC	:	High Performance Computer
ICRISAT	:	International Crops research Institute for the Semi-Arid Tropics
ISEP	:	ICRISAT Sorghum EST Primer
JGI	:	Joint Genome Institute
kb	:	kilo bases
LD	:	Linkage Disequilibrium
LG	:	Linkage Group
LOD	:	Logarithm of odds (base 10)
MAB	:	Marker-Assisted Breeding
MAS	:	Marker-Assisted Selection

Mb	:	Million bases
MDS	:	Multi Dimensional Scaling
MISA	:	MIcroSAtellite
mM	:	milliMolar
NCBI	:	National Center for Biotechnology Information
ng	:	nanograms
NJ	:	Neighbor-Joining
PAGE	:	Polyacrylamide Gel
PCoA	:	Principal Coordinate Analysis
PCR	:	Polymerase Chain Reaction
PIC	:	Polymorphism Information Content
pm	:	picomole
QTL	:	Quantitative Trait Loci
RAPD	:	Random Amplified Polymorphic DNA
RFLP	:	Restricted Fragment Length Polymorphism
RIL	:	Recombinant Inbred Line
SNP	:	Single-Nucleotide Polymorphism
SSR	:	Simple Sequence Repeats
SSRIT	:	Simple Sequence Repeat Identification Tool
TIGR	:	The Institute for Genome Research
TRF	:	Tandem Repeat Finder

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Sorghum is the fifth most important grain crop in the world after wheat, maize, rice, and barley and the second most important cereal crop after maize in sub-Saharan Africa. It is thought to have originated from Northeast Africa (the Ethiopia, Eritrea area), the center of diversity for the crop. Sorghum is cultivated mostly in the developing world (especially in Africa and Asia), although it has become an important industrial crop in many developed countries. Its value in arid climates is due to its ability to withstand dry conditions. Sorghum is an annual and predominantly selfing cereal (Dje et al. 2004). Harlan and de Wet (1972) recognized five basic botanical races [bicolor (B), caudatum (C), durra (D), kafir (K) and guinea (G)] and ten intermediate races of cultivated sorghum, defined on the basis of spikelet, seed, and panicle morphology. The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), headquartered in India, serves as a world center for sorghum germplasm conservation and breeding.

Sorghum (*Sorghum bicolor* (L.) Moench) belongs to the Andropogoneae family and has a complex genome ($2n = 2x = 20$) compared to the model grass species, rice. Sorghum is an important target for plant genomics due to its adaptations to harsh environments, diverse germplasm collection and relatively small genome size (Menz et al. 2002). The genome size of sorghum is approximately 740 Mb (Paterson et al. 2009). Maize and sorghum have a common ancestor and diverged approximately 15-20 million years ago (Paterson et al. 1995; Doebly 1990). Although sorghum and maize have differences in genome size, they retain similar chromosomal organization (Paterson 1995). Sorghum and sugarcane may have shared a common ancestor as recently as 5 million years ago (Sobral et al. 1994) and retain very similar gene order (Ming et al. 1998).

At present, sorghum is the emerging crop model for many grass species that employ C_4 photosynthesis. Even though sorghum has a relatively large genome size compared to rice, it can act as a model grass species for tropical grasses. Over the last two decades, crops with large genome sizes have been studied effectively by using the model crop

genomes as templates through comparative genome mapping (Gale and Devos 1998). Sorghum is the second cereal after rice for which complete genome sequencing is available. Third place is occupied by maize (a draft genome was released in February 2008) which is a close relative of sorghum and belongs to the same Andropogoneae family. Another important commercial crop from the Andropogoneae family is sugarcane. The sequence information from sorghum and maize is expected to greatly help to understand genome organization in biofuel grasses with more complex genomes like that of sugarcane.

Crop improvement programs mainly rely on the variation present in the genotypes selected as parents. Variation at the molecular level is gaining importance in the breeder's perspective. In this context, molecular markers are playing the major role in characterization of crop germplasm. Molecular markers are considered constant landmarks in the genome. During the initial stages of molecular markers in sorghum, restricted fragment length polymorphism (RFLP) played a significant role. Later polymerase chain reaction (PCR)-based marker systems like amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), and simple sequence repeats (SSRs or microsatellites) came into existence. Every marker system has its own merits and demerits.

SSR markers have clear advantages over RFLP and AFLP markers in terms of technical simplicity, amenability to high-throughput automation, reproducibility, multi-allelic nature, co-dominant inheritance, relative abundance and good genome coverage (Varshney et al. 2005). They are the preferred marker system for many breeding applications. SSR markers are widely used in fingerprinting, genetic diversity analysis, mapping and marker-assisted selection. Hence, enriching the existing sorghum linkage map with more SSR markers is a valuable objective for the sorghum breeding community globally. SSR markers have been recently developed and exploited for different purposes in sorghum (Brown et al. 1996; Taramino et al. 1997; Bhattramakki et al. 2000; Klein et al. 2000; Kong et al. 2000; Tao et al. 2000; Subudhi and Nguyen 2000; Subudhi et al. 2000; Menz et al. 2002; Schloss et al. 2002; Bowers et al. 2003; Agrama and Tuinstra 2003; Uptmoor et al. 2003; Menz et al. 2004; Casa et al. 2005; Folkertsma et al. 2005; Wu and Huang 2006; Perumal et al. 2007; Barnaud et al. 2007; Caniato et al. 2007; Ali et

al. 2008; Due et al. 2008). However conventional SSR marker development is an expensive and time intensive process. Alternatively, SSR can be developed at cheaper costs through the mining of databases. Due to the availability of genomic or expressed sequence tag (EST)/genic sequences in public databases and the recent advent of bioinformatics tools, SSR marker development has become easier through mining the databases and more cost-effective (Jayashree et al. 2006).

On-going and future sequencing projects will contribute many more genomic and genic sequences to publicly available databases, facilitating development of different types of markers at lower cost. In the past, SSR markers have been successfully developed by mining EST databases in several crops (reviewed in Varshney et al. 2005). EST-SSR markers have the following advantages over genomic SSR markers: 1). Ability to detect the variation at expressed regions of genome helps in 'perfect marker-trait association' 2). No direct cost is involved in development of markers, as they were derived from EST databases as 'by-products' 3). Because of their greater transferability in comparison with genomic SSRs, they can be used in related crop species facilitating functional comparative genome mapping. As economically important grasses exhibit more conserved regions across taxa than found in other crop lineages, EST-SSRs should become an extremely powerful tool for better understanding functional relationships between the grass species (functional genetic diversity) and for genetic mapping.

EST-SSR markers have been developed and exploited effectively for different purposes in different crops (rice: Cho et al. 2000; wheat: Gupta et al. 2003, Zhang et al. 2006, Xue et al. 2008; tall fescue grass: Saha et al. 2004; rye: Studer et al. 2008). Construction of high-density linkage maps is a crucial step for identifying gene locations and their isolation. SSR markers, including both genic and genomic markers, will allow comparative analysis of sorghum genes with orthologous genes in rice, and also in closely related species like maize, sugarcane and Johnson grass. Because, EST-SSR markers are derived from transcribed regions of the genome, they have higher rates of transferability than genomic SSRs (Saha et al. 2004; Wang et al. 2005). Thus, EST-SSR markers can be exploited in relative crop species for which sufficient genomic tools are not available and greatly facilitate in developing comparative genome maps in related crop species.

Genetic improvement of sorghum can help farmers in semi-arid areas where sorghum is a key food crop. In the past, studies have been devoted to assessing patterns of sorghum genetic variation based on morphology or pedigree. However, this approach has its limitations. Complex and quantitatively inherited traits are difficult to trace based solely on morphology. For this reason, DNA-based methods have been employed in studies of sorghum genetic diversity and in genetic improvement of this crop. To date relatively few marker-based genetic diversity studies have been carried out in sorghum (Deu et al. 1994, 1995, 2006, 2008; Menkir et al. 1997; Dje et al. 2000; Grenier et al. 2000a, 2000b; Smith et al. 2000; Ghebru et al. 2002; Agrama and Tuinstra 2003; Uptmoor et al. 2003; Menz et al. 2004; Casa et al. 2005; Folkertsma et al. 2005; Perumal et al. 2007; Barnaud et al. 2008; Caniato et al. 2008; Ali et al. 2008; Mace et al. 2008). Genetic diversity in cultivated sorghum in Asia, Africa and Latin America is threatened by loss of landraces due to introduction of improved open-pollinated varieties and single-cross hybrids. Evaluating germplasm diversity can help to identify landraces with the greatest novelty and thus are most suitable for incorporation into crop improvement programmes. Genetic distance estimates determined by molecular markers help to identify suitable germplasm for incorporation into plant breeding programs. EST-SSR markers may help in understanding these genetic relationships at a functional level. In this context, this study was carried out with the following objectives:

OBJECTIVES

1. To develop EST-SSR markers in sorghum by exploiting rice-sorghum synteny
2. To assay the polymorphism between the parents of two sorghum mapping populations using EST-SSR markers
3. To develop genetic linkage map of sorghum with polymorphic EST-SSR markers using a sorghum recombinant inbred line (RIL) population segregating for shoot fly resistance
4. To develop sequence-based physical map for all linkage groups of sorghum using complete sorghum genome sequence
5. To assess allelic diversity using mapped EST-SSR markers in a reference set of sorghum germplasm (384 accessions representing diversity from a global composite collection of 3367 accessions)

2. REVIEW OF LITERATURE

2.1 Sorghum

Sorghum is the 5th most important cereal crop globally, after maize, wheat, rice and barley (www.fao.org) and an important food grain crop in the semi-arid tropics. Sorghum is defined as an important 'failsafe' crop in the agro-ecosystem (Paterson 2008) because of its dual nature (food and feed). Its wide adaptation to harsh environments, tolerance to stress conditions, diverse germplasm collections and its small genome size made sorghum as an important botanical model crop for many tropical grasses with complex genomes, which employ C_4 photosynthesis (Kresovich *et al*. 2005). Sorghum as biofuel crop occupies 2nd position (after maize) in grain-based ethanol production in Unites States (US). Sorghum offers novel learning opportunities relevant to weed biology as well as to improvement of a wide range of other forage, turf and biomass crops.

Cultivated sorghum is divided into five botanical races, based on their grain shape and glume morphology, namely, bicolor (B), kafir (K), caudatum (C), durra (D) and guinea (G), and ten intermediate races obtained by combining any two races (Harlan and de Wet 1972).

Sorghum is more closely related than rice to other major tropical cereal grasses with complex genomes and high level genome duplication. *Sorghum* and *Zea* (maize) diverged from a common ancestor ~15-20 million years ago. *Zea* has undergone at least one complete genome duplication since its divergence from sorghum, whereas *Saccharum* has undergone at least two such duplication events (Paterson et al. 1995, 2008).

Among the abiotic stresses, drought is the major constraint in the sorghum production in the arid, semi-arid tropics and subtropics. The stay-green component of terminal drought tolerance helps sorghum to stay stable even in high temperatures at grain filling stages. Stay-green sources are available to transfer this trait into locally-adapted improved verities. Shoot fly is the major biotic constraint in sorghum production in South Asia. Shoot fly infestation is characterized by formation of dead-hearts resulting in complete death of seedlings that ultimately affects the yield. Shoot fly resistant sorghum lines were also identified. Growing of shootfly resistance lines or transferring the resistance

genes/quantitative trait loci (QTLs) into locally-adapted improved varieties is the best option to avoid losses from shoot fly infestation.

2.2 Sorghum Genomics

2.2.1 Sorghum Genome Sequencing

Sorghum acts as a model for grasses employing C_4 photosynthetic pathway (Kresovich *et al*. 2005). Genome sequencing of sorghum was initiated at the end of 2005 through the 'Community sequencing program' of the US Department of Energy-Joint Genome Institute (JGI) (Paterson *et al*. 2009), and complete genome sequence (8X scaffold) information with ~10.5 million reads was made available for public use (Sorbi v1.0) (http://genome.jgi-psf.org/Sorbi1/Sorbi1.info.html). This sorghum genome sequence was generated from a leading US sorghum inbred, BTx623. Preliminary analysis of sorghum genome sequence assembly indicated that more than 97% of sorghum protein-coding genes (ESTs) were captured in 229 longest scaffolds. This recently completed genome sequencing provides a foundation for invigorating progress towards relating sorghum genes to their function.

2.2.2 Molecular Markers in Sorghum

Molecular markers play a vital role in the present scenario in many aspects of crop breeding ranging from identification of diverse lines to mapping of genomic regions controlling desirable traits and their use in marker-assisted breeding (MAB) programs. Nowadays availability of markers is not the limiting factor in crops like sorghum hence it is timely to explore the extent to which marker technology has to date – and is likely in future to – affect the practices of sorghum breeding. Use of molecular markers in sorghum was started in the 1990's. The RFLP marker system was greatly exploited in sorghum (Hulbert et al. 1990; Binelli et al. 1992; Whitkus et al. 1992; Ragab et al. 1994; Chittenden et al. 1994; Xu et al. 1994; Pereira and Lee 1995; Dufour et al. 1997; Tao et al. 1998a). Later, other marker systems came in to existence which include AFLP (Boivin et al. 1999; Klein et al. 2000; Haussmann et al. 2002a; Menz et al. 2002), RAPD (Xu et al., 2001; Haussmann et al., 2002a), SSRs (Taramino et al. 1997; Bhattramakki et al. 2000; Klein et al. 2000; Kong et al. 2000; Subudhi et al. 2000; Menz et al. 2002; Schloss et al. 2002; Bowers et al. 2003; Wu and Huang 2006) and most recently DArT markers

(Mace et al. 2008). All these marker systems were used in sorghum for different purposes like fingerprinting, assessing genetic diversity and QTL mapping.

Among the existing marker systems, SSR markers have clear advantages over RFLP and AFLP in terms of technical simplicity, throughput level and potential for automation (Varshney et al. 2005). SSRs are the preferred marker system for many breeding applications. SSR markers are widely used in fingerprinting, genetic diversity analysis, mapping and marker-assisted selection. Hence, enriching the existing sorghum linkage map with more SSR markers is a valuable objective for the sorghum breeding community globally.

2.2.2.1 EST-SSR markers

Conventional SSR marker development is a costly and time-consuming process. Thanks to the availability of genomic or EST/genic sequences in public databases and the recent advent of bioinformatics tools, SSR marker development has become easier and more cost-effective (Jayashree et al. 2006). In the past, SSR markers have been successfully developed by mining EST databases in several crops including monocots (reviewed in Varshney et al. 2005; Jayashree et al. 2006; Ellis and Burke 2007) and dicots (Kumpatla and Mukhopadhyay 2005, Jayashree et al. 2006). EST-SSRs were reported in many cereals/grass species, including rice (Cho et al. 2000), sugarcane (Cordeiro et al. 2001), durum wheat (Eujayl et al. 2002; Gupta et al. 2003; Yu et al. 2004; Zhang et al. 2006), rye (Hackauf and Wehling 2002; Studer et al. 2008), barley (Thiel et al. 2003), tall fescue grass (Saha et al. 2004), finger millet (Dida et al. 2007) and pearl millet (Senthilvel et al. 2008).

The nature and frequency of SSRs in EST collections have been comprehensively discussed in Kantety et al. (2002), Varshney et al. (2002), La Rota et al. (2005) and Jayashree et al. (2006) for monocots and Kumpatla and Mukhopadhyay (2005) and Jayashree et al. (2006) for dicots. These EST-SSR markers were used for diversity analysis in rice (Cho et al. 2000), wheat (Leigh et al. 2003, Gupta et al. 2003, Zhang et al. 2006), barley (Thiel et al, 2003; Varshney et al. 2007, 2008), and for mapping in wheat (Gupta et al. 2003, Yu et al. 2004; Balyan et al. 2005), barley (Varshney et al. 2006), finger millet (Dida et al. 2007) and pearl millet (Senthilvel et al. 2008).

Generally, the EST-derived SSR markers are found to be less polymorphic than genomic SSRs as these are derived from transcribed regions of the genome (Cho et al. 2000; Eujayl et al. 2002; Thiel et al. 2003; Varshney et al. 2002, 2005; Jayashree et al. 2006; Ellis and Burke 2007). Most transcribed regions are greatly conserved across species. In rice, the model organism for cereals, Cho et al. (2000) reported 54% of polymorphism using EST-SSR markers across seven mapping population parents. EST-SSR markers developed in durum wheat identified only 25% polymorphism (Eujayl et al. 2002), whereas Thiel et al. (2003) reported 8-54% polymorphism on three different mapping population parental line pairs in barley. Even though EST-SSR markers exhibit a lower percentage of polymorphism than their genomic SSR counterparts, this marker system has been greatly exploited in the transition from structural genomics to functional genomics.

EST-SSR markers are superior in terms of cross-species transferability, as they were derived from the most conserved regions of genome, and thus are well suited for application in phylogenetic analysis and comparative genome mapping (Zhang et al. 2006). Wang et al (2005) developed a small number of EST-SSR markers (30) in sorghum along with wheat, rice and maize. The transfer rate of EST-SSR markers from sorghum to paspalum (*Paspalum* spp.) and to maize was 68% and 61%, respectively. Saha et al. (2004) reported about 57% transferability across six grass species using tall fescue EST-SSRs. EST-SSR markers developed in wheat (Yu et al. 2004) found 62% transferability across four species including wheat, rice, maize and barley. Thiel et al. (2003) identified 40% transferability of barley EST-SSRs to rice.

Development of EST-SSR markers was spread even to other species in functional genomics era, including *Brassica* (Tonguc and Griffith 2004), cotton (Qureshi et al 2004; Wang et al. 2004), coffee (Aggarwal et al. 2007), citrus (Luro et al. 2008)and tree species, red raspberry (Graham et al. 2004), loblolly pine and spruce (Berube et al. 2007).

2.2.2.2 Diversity Array Technology (DArT) markers

DArT markers are one of the new generation markers. DArT provides high quality markers that can be used for diversity analyses and to construct medium-density genetic

linkage maps. The high number of DArT markers generated in a single assay not only provides a precise estimate of genetic relationships among genotypes, but also their even distribution over the genome offers real advantages for a range of molecular breeding and genomics applications. DArT was first developed in rice (Jaccoud et al. 2001). Subsequently, it was developed for different crops and used in linkage map construction and diversity analysis. The important plant species for which DArT has been developed include rice (Xie et al. 2006) barley (Wenzel et al. 2004, 2006), *Arabidopsis* (Witenberg et al. 2005), eucalyptus (Lezar et al. 2004), wheat (Semagn et al. 2006; Akbari et al. 2006), Cassava (Xia et al. 2005) and pigeon pea (Young et al. 2006). Recently, DArT was successfully developed for sorghum (Mace et al. 2008). Over 500 markers detected variation among 90 sorghum accessions used in a diversity analysis and 596 DArT markers were mapped onto integrated linkage map of sorghum.

2.2.3. Marker based genetic linkage maps in sorghum

Construction of a genetic linkage map is a fundamental step and prerequisite for a detailed genetic study and broad application of marker-assisted breeding approaches in any crop. Linkage maps developed in sorghum by different research groups using different marker systems were summarized in Table 1. Construction of genetic linkage map in sorghum was started in the 1990's. Hulbert et al. (1990) reported the first linkage map in the cross between Shangui Red (a kaoliang cultivar from North Central China) and M91051 (a zera zera ciltivar from East Africa) using RFLP probes from maize clones. In general, sorghum contains 10 linkage groups (LGs), but in their study, they reported only 8 LGs with a map length of 283 cM. Later, many groups reported genetic linkage maps for sorghum. Early stage linkage maps for sorghum were constructed mainly with RFLP probes (Hulbert et al. 1990; Binelli et al. 1992; Whitkus et al. 1992; Chittenden et al. 1994; Ragab et al. 1994; Xu et al. 1994; Pereira and Lee 1995; Dufour et al. 1997; Tao et al. 1998a). Sorghum-specific probes derived from gDNA and cDNA libraries were added to linkage maps along with exogenous probes developed from other related genomes. A map constructed by Ragab et al. (1994) using 38 sorghum gDNA and 33 maize gDNA probes contained 15 LGs, and covered 633 cM with an average distance of 8.9 cM between markers. A linkage map was developed from the inter-specific cross between BTx623 and *S. propinquum*, comprising 10 LGs corresponding to the 10 pairs of sorghum chromosomes using 56 F_2 plants (Chittenden et al. 1994). Up to 1997, only F_2

mapping populations were used for construction of sorghum genetic linkage maps. Later sorghum RIL populations were used as they are more stable as compared to F_2 mapping populations.

Later, linkage maps for sorghum were generated adding other marker systems like amplified fragment length polymorphisms (AFLPs) (Boivin et al. 1999; Klein et al. 2000; Haussmann et al. 2002a; Menz et al. 2002), simple sequence repeats (SSRs) (Tao et al. 1998a; Bhattramakki et al. 2000; Klein et al. 2000; Kong et al. 2000; Tao et al. 2000; Subudhi et al. 2000; Haussmann et al. 2002a; Menz et al. 2002; Bowers et al. 2003; Feltus et al. 2006a; Brown et al. 2006; Wu and Huang 2006, 2008) and recently diversity array technology (DArT) (Mace et al. 2008).

The linkage maps were saturated recently by adding more markers. Bhattramakki et al. (2000) developed a high density genetic linkage map for sorghum using RFLP probes and SSR markers with an average distance between markers of 3.1 cM. Maps were still further saturated by Menz et al. (2002) with average marker distance of 0.6 cM where as Bowers et al (2003) developed a map with 0.4 cM interval between markers. An integrated linkage map for sorghum was developed using 145 SSR and 323 RFLP markers (Bowers et al. 2003). Later, Wu and Huang (2006) developed an exclusive linkage map of sorghum with 118 publicly available SSR markers, which spanned 997.5 cM with an average distance of 8.5 cM. A recent linkage map developed by Mace et al (2008) consists of 358 DArT markers, 47 SSRs and 188 RFLP markers distributed across all 10 linkage groups and spanning 1431.6 cM. The average distance between adjacent markers is 2.39 cM.

2.2.4 QTL mapping

Genomic regions associated with important agronomic traits controlled by multiple genes are termed as quantitative trait loci (QTL). In sorghum QTL have been identified for agronomically important traits including a range of abiotic stress tolerances [primarily mid-season and terminal drought tolerance (Tuinstra et al. 1996, 1997; Crasta et al. 1999; Subudhi et al. 2000; Tao et al. 2000; Xu et al. 2000; Kebede et al. 2001; Haussmann et al. 2002b; Sanchez et al. 2002; Harris et al. 2007), early-season cold (Knoll et al. 2008) and aluminum tolerance (Magalhaes et al. 2004)], host plant resistance to hemi-parastic

weeds [*Striga* spp. (Haussmann et al. 2004)], insect pests [aphids (Agrama et al. 2002; Katsar et al. 2002; Wu and Huang 2008), midge (Tao et al. 2003), and shoot fly (Folkertsma et al. 2003)], and diseases [of both foliage (Oh et al. 1996; Tao et al. 1998b; Nagy et al. 2007) and panicles (Klein et al. 2001)], and a range of traits related to grain quality (Lijavetsky et al. 2000), crop phenology [height and maturity (Lin et al. 1995; Pereira and Lee 1995; Klein et al. 2001; Feltus et al. 2006a)], and yield components (Hart et al. 2002; Hicks et al. 2002).

Among the abiotic constraints, drought plays a key role in crop production in tropical countries. Drought occurring at grain filling stage is termed as "terminal drought". 'Stay-green' considered as one of the traits responsible for tolerance to terminal drought. Different research communities have studied the terminal drought (Tuinstra et al. 1996, Crasta et al. 1999, Xu et al. 2000; Sanchez et al. 2002, Haussmann et al. 2002b, Harris et al. 2007) and identified QTLs responsible for terminal drought tolerance. Six QTLs were identified in sorghum responsible for terminal drought tolerance (Tuinstra et al. 1996). Whereas Crasta et al. (1999) reported 7 QTLs for terminal drought tolerance and 2 QTLs for maturity. Among these seven, three major QTLs contributed about 42% of phenotypic variability and four minor QTLs together contributed 25% phenotypic variability in terms of stay-green ratings. Two groups, Sanchez et al. (2002) and Harris et al. (2007) reported only four major QTLs responsible for stay-green trait in sorghum. In both studies, B 35 genotype was the donor parent for stay-green trait. In another study, E 36-1 genotype was used as the source material for the identification QTLs responsible for stay-green sorghum (Haussmann et al. 2002b).

Among biotic constraints, shoot fly (*Atherigona soccata*) causes more damage to sorghum in terms of yield loss than other insect pests. Shoot fly infestation on sorghum seedlings result in formation of dead-hearts, and further leading to complete death of seedlings. Shoot fly resistance has quantitative inheritance. The genomic regions (QTLs) associated with various components and direct measures of shootfly resistance have been identified (Folkertsma et al. 2003).

Among the sucking pests of sorghum, QTLs have been identified for greenbug (*Schizaphis graminum* Rondani) resistance (Agrama et al. 2002; Katsar et al. 2002; Wu

and Huang 2008). Greenbug is a major insect pest of several cereals in the Great Plains of USA. Greenbug resistance was controlled by two QTLs present on sorghum chromosome 9 (SBI-09), one major QTL accounting 55-80% of phenotypic variation and one minor QTL accounting 1-6% of phenotypic variability. A diversity analysis was performed using greenbug resistant lines in sorghum (Wu et al. 2006) using AFLP markers. This type of analysis helps the breeding community to select the most diverge greenbug resistance parental lines to transfer the resistance into locally-adapted improved varieties.

2.3 Sequence Databases

De novo generation of microsatellite markers through laboratory-based screening of SSR-enriched genomic libraries is a time-intensive and expensive endeavor (Gupta et al. 2003, Varshney 2005, Jayashree et al. 2006). An alternative approach to develop microsatellite markers is to screen the public databases of related model species where abundant sequence data is already available. Availability of public databases for crop plants and powerful bioinformatics tools for data mining helped in development of new molecular tools to strengthen plant breeding (Mahalakshmi and Ortiz 2001; Jayashree et al. 2006). The availability different techniques and tools to analyze massive amounts of nucleotide sequence data has led to the development of innovative ways to examine these data as reflected in their functions.

Genome-related public databases have become an invaluable tool for the scientific community and are the public window for the high-throughput genomics projects. Genome sequencing of a plant species is not the end of itself, but it is the beginning of a new venture to unravel the genetic information of that species and to gain better insight into genetics of closely related plant species (Mahalakshmi and Ortiz 2001). Sequencing results suggest that improvement of other important food crops can benefit from information obtained in the model crop systems. Identification of a gene of interest in the present scenario starts with sequence information, including EST, genome sequence and protein sequence information. This information is publicly available for large-scale analysis from GenBank at the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov) and the European Molecular Biology Laboratory (EMBL) (www.ebi.ac.uk/embl.com). Separate databases are also maintained for each crop (Arabidopsis (www.arabidopsis.org), gramineae species (www.gramene.org), maize

(www.maizegdb.org), etc.). Public databases, most notably the model organism databases, have two major consumers: the focused scientific community actively studying a particular species, and the large scientific community interested in relating this specialized information to and from other systems (Mahalakshmi and Ortiz 2001).

In general, most of the databases maintain both genomic and genic (EST) sequences where as the J. Craig Venter Institute database [formerly, The Institute for Genome Research (TIGR) (http://www.tigr.org)] exclusively maintains an EST sequence database for plant and animal species. Expressed sequence tags (ESTs) are single-pass cDNA sequences derived from transcribed regions of the genome. EST projects in several crops have resulted in generation and deposition of many EST sequences in the public databases. These sequences are useful for gene discovery, new marker development and for comparative genomics between related crop species. Microsatellites in the ESTs are a major resource for the breeding community in developing these markers in larger numbers and at lower costs, which will increase the potential of the breeding community to use them for effective transfer of genomic regions through marker-assisted breeding (MAB) programs.

SSR markers developed by data mining are considered as the by-products of these databases (Varshney et al. 2002). Beyond the cost savings achieved by exploiting existent DNA sequence information, this approach also offers the possibility of identifying rare microsatellite motifs that would be uneconomical to identify through conventional laboratory protocols.

2.4 Marker-assisted selection (MAS)

Molecular markers allow the traditional breeding methods with the transfer of a greater variety of genetic information in a more precise and controlled manner. Marker-assisted selection for important but complex traits, which are often difficult to select in the routine breeding programs, will enhance the breeding programs by providing better-focused products and save time and resources. MAS could expedite the introgression of genes from landraces into elite breeding lines; however, very few studies have empirically demonstrated the efficacy of MAS for quantitatively inherited agronomic traits. A fully

integrated sorghum molecular genetic map is the basis for mapping, marker-assisted selection, candidate gene cloning and sequencing of the full sorghum genome.

Early season cold tolerance in sorghum is controlled by three major QTLs represented by 3 genomic SSRs (*Xtxp*43, *Xtxp*51 and *Xtxp*211). These three QTLs were transferred from Shan Qui Red (SQR), a Chinese cold tolerant line, into two locally-adapted improved cold susceptible varieties, Tx2794 and Wheatland by Knoll and Ejeta 2008. They evaluated the QTL effects in the field for cold tolerance in both genetic backgrounds.

ICRISAT has initiated a large-scale marker-assisted back crossing (MABC) program to transfer the stay-green component of terminal drought tolerance – a trait that is likely to be associated not only with more stable grain and stover yield, but also expected to contribute the maintenance of ruminant nutritional value of the stover produced under drought stress conditions (Hash et al. 2003). The stay-green trait is controlled by at least six QTLs. These QTLs are being introgressed from donor parent, B35, into locally-adapted varieties of sorghum, including R 16 and ISIAP Dorado, through MABC at ICRISAT.

2.4.1 MAS in other cereals

Even though sufficient microsatellite markers are available in barley, MAS is not widely used except for barley yellow mosaic virus (BaYMV) (Thomas 2003). In wheat, MAS targets related to genes for disease resistance, most of which were well characterized with efficient phenotypic screens. Malting quality and stripe rust resistance also gained the importance in later stages through MAS in wheat. Markers for two cereal cyst nematode resistance genes derived from a landrace and a wild relative of wheat have been employed successfully to pyramid two resistance genes (Eagles et al. 2001). MAS is greatly successful where the markers are the products of target gene themselves, e.g., the wheat cereal cyst nematode resistance genes (Eagles et al. 2001), the *Xa21* locus in rice (Chen et al. 2000), and grain storage protein subunits in wheat (Eagles et al. 2001).

Pyramiding of disease resistance was given the priority for MAS in rice, particularly for bacterial blight (*Xanthomonos oryzae*) and blast (*Pyricularia oryzae*) (Chen et al. 2000; Sanchez et al. 2000; Singh et al. 2001). To meet the needs of farmers in flood-prone

areas, Neeraja et al (2007) successfully transferred the submergence QTL (*Sub*1) present on rice chromosome 9 through MAS to locally-adapted improved variety of rice, Swarna, using flanking SSR markers.

2.6 Comparative genome mapping

Comparative genomics studies suggest that much of the plant genome has been conserved over the process of evolution (Gale and Devos 1998). Geneticists and evolutionary botanists have long held interest in the phylogeny of crops and in chromosomal evolution. Comparative genome mapping adds a powerful technique for assessing the mode and tempo of chromosomal evolution. The utility of comparative mapping outside the aspect of evolutionary biology is in the development of molecular markers for specific traits. Genetic content and co-linearity in large chromosomal segments have been conserved over millions of years, thereby leading to the construction of first grass genome maps that align several maps. Anchored common markers across the species are used as landmark references in comparative genome mapping.

Among the cereals, rice was selected first for genome sequencing because of its small genome size (430 Mb). It acts as a model for crop plant species, so it is positioned in the center of cereal genome circle (Figure 1), developed by using RFLP probes (Gale and Devos 1998).

During the initial days of DNA-based molecular marker research, restriction fragment length polymorphism (RFLP) played a vital role in identifying low copy number genes that could be mapped across different plant species. This led to elucidation of extensive genome comparisons and conservation of gene order with in homoeologous chromosomal segments. Complete co-linearity between even closely related species is rare, as co-linearity is often interrupted by insertions, deletions and inversions (Cook 1998; Adam 2000; Mahalakshmi and Ortiz 2001).

Sorghum, sugarcane and maize belong to Andropogoneae tribe and the genome of sorghum is smallest among these three. Comparative map between sorghum, maize and sugarcane revealed that maize is an ancient tetraploid while sugarcane is polyploid (Guimaraes et al. 1997). In these maps, sorghum stands between these two species.

Reports on comparative mapping of rice probes on sorghum indicate a high degree of co-linearity and synteny (Paterson et al. 1995; Harushima et al., 1998; Nagamura et al. 1998; Gale and Devos. 1998, Mahalakshmi and Ortiz 2001; McCouch 2001; Ventelon et al. 2001). Several efforts have been already made for comparative genome mapping in cereals such as sorghum and maize (Hulbert et al. 1990; Binelli et al. 1992; Whitkus et al. 1992), sugarcane, maize and sorghum (Dufour et al. 1997), rice and sorghum (Paterson et al. 1995; Nagamura et al. 1998; Ventelon et al. 2001), rice and maize (Ahn et al. 1993), and wheat, rye and barley (Devos and Gale 1997). Comparative genome maps were developed between rice and *Arabidopsis* (Liu et al. 2001; Mayer et al. 2001, Salse et al. 2002). Comparing the rice and *Arabidopsis* genomes, very high level of rearrangement was found characterizing 60 syntenic regions illustrating the ancient nature of speciation between them. However, micro-synteny was observed between these two model plant species in these three studies.

Comparative genome mapping is not limited to cereals but also extends to other family groups like Solanaceae [tomato and pepper (Tanksley et al. 1988), tomato and potato (Tanksley et al. 1992), tomato and egg plant (Doganlar et al. 2002)]. Comparative maps were also developed between *Arabidopsis* and soybean (Grant et al. 2000).

Practical utility of comparative genome mapping information would be in map-based cloning of specific genes of interest, the dissection of trait-based markers especially expressed sequence tags (ESTs), genome evolutionary studies and utility of markers developed in one species in the related crop species.

2.7 Physical mapping and its integration with linkage map in sorghum

Integration of genetic and physical maps is a valuable tool for gene isolation through the map-based cloning approach and comparative genome mapping. Most of physical maps (in humans: Vollrath and Jaramillo-Babb 1997, in *Arabidopsis*: Meinke et al. 2003, in *Arabidopsis* and tomato: Fransz et al. 1996, and in rice: Zhu et al. 1999, Cheng et al. 2005) were constructed using fingerprinting and PCR- or hybridization- based methods in complex genomes. Integration of genetic and physical maps in sorghum was done for the first time using hybridization and fingerprinting data on BAC libraries by Klein et al. (2000). This is a low cost and highly efficient method to construct integrated genetic and

physical maps. Klein et al. (2000) aimed at the development of 2500 links between sorghum genetic and physical maps, or an average of one link/marker for every 300 kb. This was a big task at that time because of insufficient numbers of markers available to achieve this target. Later, Draye et al. (2001) developed integrated maps by combining genetic, physical, diversity and cytomolecular maps for grasses by taking sorghum genome as a template using 156 RFLP probes specific to LG C of sorghum. But, in the present scenario, the complete sorghum genome sequence is available and many more molecular markers (especially SSRs and AFLPs) are available. This encouraged us to develop a physical map exclusively with SSR markers and the sorghum genome sequence and integrate this with linkage maps. Integrated genetic and physical maps open the way to more insight towards functional genomics from structural genomics by looking directly into the sequences underlying the genes or QTL positions.

2.8 Genetic diversity studies in sorghum

Cultivated sorghums present a wide phenotypic diversity. Marker-based genetic diversity assessment among the cultivated sorghums was first reported by Aldrich and Doebley (1992) using 38 RFLP probes, and followed by Tao et al. (1993) using 16 RFLP and 29 RAPD markers. Because of the small number of accessions (used less than 50), their studies didn't find groupings based on race or geographic origin. Racial differentiation was first observed by Deu et al. (1994), by using 33 RFLP probes in 94 accessions representing all races and geographic origins. Average number of alleles in this study was 2.8 per locus. Among the races, bicolor did not form any homogenous group, while the guineas formed three major groups based on their geographic origin. In another study (Cui et al. 1995), less racial differentiation was observed because kafir and guinea races from southern Africa and Asia formed a single group. Guinea margaritiferum (Gma) accessions showed clear differentiation from other guinea race accessions and were found to be more closely related to wild sorghum than other cultivated sorghum (Cui et al. 1995). Gma accessions were greatly devoid of common alleles and mainly characterized by unique alleles (de Oliveira et al. 1996). Kafir race was found to be least diverse in comparison to bicolor and guinea races in the study revealed by Menkir et al. (1997) using 82 RAPD primers on 190 accessions.

Development of simple sequence repeat (SSR) markers in sorghum was started in late 90's. Different research groups developed SSR markers independently in sorghum (Brown et al. 1996; Taramino et al 1997; Bhattramakki et al. 2000; Kong et al. 2000; Schloss et al. 2002). After the development of SSR markers, utility of these markers was greatly increased in genome fingerprinting, genetic diversity analysis and linkage mapping (Dean et al. 1999; Smith et al. 2000; Grenier et al. 2000b; Menz et al. 2004; Agrama and Tuinstra 2003; Folkertsma et al. 2005; Wang et al. 2006; Perumal et al. 2007; Barnaud et al. 2007; Caniati et al. 2007; Wu et al. 2008; Deu et al. 2008).

By the early 2000's, RFLP marker utility in sorghum was greatly reduced as other marker systems had been developed having greater simplicity and high through-put potential. These replacement marker systems included RAPD, AFLP and SSR markers. Smith et al. (2000) used 15 SSRs in combination with 104 RFLP probes to study genetic similarity among elite sorghum lines. Twenty-two trait-specific sorghum accessions were studied for diversity using 32 RAPD primers and 28 SSR markers (Agrama and Tuinstra 2003). Results indicated that SSR markers have the great potential to differentiate the accessions with an average of 4.5 alleles per locus. SSR markers have clearly differentiated the accessions according to race and geographic origin than RAPD markers. Menz et al (2004) also used SSRs in combination with RFLP markers to assess the genetic diversity among public inbred lines of sorghum in US breeding programs.

A panel consisting of 104 accessions (73 landraces and 31 wild sorghums) was used to identify the genetic diversity and simultaneously identified selection signals that might be associated with sorghum domestication using 98 SSR markers (Casa et al. 2005). Their study showed that about 86% of genetic diversity observed in wild sorghums was retained in landraces. SSR markers proved powerful to greatly differentiate the wild and landrace accessions.

Later, genetic studies were extended to specific races. Diversity within the guinea race (100 accessions from ICRISAT's guinea race core collection) was studied using 21 genomic SSR markers (Folkertsma et al. 2005). Within the guinea race, most accessions were grouped based on their geographic distribution, whereas guinea margaritiferum

accessions formed a distinct group. Average number of allele per locus in this study was 5.6 per locus and average Polymorphic Information Content (PIC) value was 0.44.

Deu et al. (2006) assessed the global structure of diversity in sorghums worldwide using 74 RFLP probes scattered across all linkage groups. These markers were selected to study the genetic diversity in a mini-core collection (210 accessions) available in CIRAD, France. This study also clearly distinguished the guinea margaritiferum sub-race, as well as geographical and racial patterns of genetic diversity of other cultivated sorghums. Guinea sorghums were clustered into 3 major groups with respect to their geographic distribution.

Genetic diversity at the village level was assessed among the locally cultivated varieties in Cameroon using SSR markers (Barnaud et al. 2007). AFLP markers were used to distinguish the sweet sorghum accessions from grain sorghum (Ritter et al. 2007). Genetic similarity was assessed among resistant lines of sorghum for fungal diseases; rust and anthracnose (Wang et al. 2006) and for greenbug (Wu et al. 2008) using SSR markers. Niger-wide sorghum accessions were evaluated for diversity using SSR markers (Deu et al. 2008) where as Mace et al (2008) studied genetic similarities between 90 elite breeding and inbred sorghum lines using DArT markers.

3. MATERIALS AND METHODS

3.1 Development of EST-SSR markers

3.1.1 Identification of SSRs in EST sequences

Sorghum EST sequences were downloaded in FASTA format from the J. Craig Venter Institute [formerly, The Institute for Genome Research (TIGR)] database (http://www.tigr.org) as of May 2004 (Figure 2a). The JCVI database exclusively maintains EST sequences for all plant species.

Microsatellites/SSRs were identified using the Simple Sequence Repeat Identification Tool (SSRIT) program (www.gramene.org/db/searches/ssrtool) (Figure 2b). Simple scripts (written in Visual Basic) were then used to parse SSRIT output into the relational database. EST sequences less than 200 bp were removed. SSR-containing EST sequences were clustered using Cap3 program to identify the non-redundant EST sequences (Figure 2c). The Cap3 algorithm computes overlaps between sequences and then joins the reads in decreasing order of overlaps to form contigs.

3.1.2 Selection of candidate EST sequences

Non-redundant SSR containing EST sequences of sorghum were used for homology search against the aligned genome sequence of rice (*Oryza sativa*) in May 2004 using the Basic Local Alignment Search Tool (BLAST) available in the GRAMENE database (http://www.gramene.org/db/searches/blast) (Figure 2d). Among all the BLAST hits identified for a particular sorghum EST, one having the maximum score was selected. This was followed by identification of homologous genomic region on rice chromosome (Figure 2e). BLAST searches were performed to provide complete coverage across the rice genome sequence. For this purpose, each rice chromosome was divided arbitrarily into five equal parts and named as “top”, “middle”, “bottom”, “region between top and middle” and “region between middle and bottom” (Figure 3). The top scoring ten SSR-containing sorghum EST sequences for each region of each rice chromosome were picked. Thus a total of 600 SSR-containing non-redundant sorghum EST sequences were selected (10 each for each region of the 12 rice chromosomes) having homology with rice

sequences. These 600 sorghum SSR-containing EST sequences were expected to sparsely cover the entire nuclear genome of rice, and thereby similarity covers the corresponding sorghum genomic regions.

3.1.3 Primer designing

Primer pairs were designed for selected 600 non-redundant SSR-containing sorghum EST sequences using Primer3 program after masking the repeat units (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) (Figure 2f). The primer pairs were designed to meet the following conditions: primer length (min-18nt, opt-22nt, max-24nt), Tm (min-54˚C, opt-57˚C, max-60˚C) and GC content (min-45%, opt-50%, max-60%).

3.1.4 PCR optimization and polymorphism assessment

Polymerase chain reaction (PCR) conditions for all 600 EST-SSR primer pairs were optimized using template DNA of the genetically diverse parental pairs of two sorghum mapping populations available at ICRISAT-Patancheru, *viz*., N13 × E 36-1 and BTx623 × IS 18551. N13 is *Striga* resistant parent while E 36-1 is stay-green drought tolerant and *Striga* susceptible. BTx623 is the elite hybrid parental line chosen for genome sequencing (Paterson et al. 2009). It is susceptible to shoot fly and is being used as the recurrent parent for ICRISAT's exploratory shoot fly resistance marker-assisted backcrossing program. IS 18551 is a shoot fly resistant donor parent being used by ICRISAT for shoot fly resistance QTL mapping and marker-assisted backcrossing. DNA of the four parental lines was isolated using a high-throughput DNA extraction protocol as reported by Mace et al. (2003) and normalized to a working concentration of 2.5 ng/µl. PCR was performed in 5 µl reaction volumes with following protocol:

Component	Stock Concentration	Volume
DNA	**2.5 ng/µl**	**1.0 µl**
Primers	**2 pm/µl**	**1.0 µl**
$MgCl_2$	**10 mM**	**1.0 µl**
dNTPs	**2 mM**	**0.3 µl**
Buffer	**10X**	**0.5 µl**
Enzyme	**0.5 U/µl**	**0.2 µl**
(AmpliTaq Gold®, Applied Biosystems, USA)		
Water		**1.0 µl**
	Total	5.0 µl

PCR reactions were carried out in a GeneAmp® PCR System 9700 thermal cycler (Applied Biosystems, USA) with a Touchdown (61-51) program using the following cyclic conditions:

Step 1**:** **Denaturation at 94˚C for 15 min**

Step 2**:** **Denaturation at 94˚C for 15 sec**

Step 3**:** **Annealing at 61˚C for 20 sec**

(temperature reduced by 1˚C for each cycle)

Step 4**:** **Extension at 72˚C for 30 sec**

Step 5**:** **Go to Step 2 for 10 times**

Step 6**:** **Denaturation at 94˚C for 10 sec**

Step 7**:** **Annealing at 54˚C for 20 sec**

Step 8**:** **Extension at 72˚C for 30 sec**

Step 9**:** **Go to Step 6 for 40 times**

Step 10**:** **Extension of 20 min at 72˚C.**

Step 11**:** **Store at 4˚C**

Amplified PCR products were resolved on 6% native polyacrylamide gels (PAGE) coupled with silver staining as described by Tegelstrom (1992). SSR primer pairs detecting polymorphism between the four parental lines were selected based on mobility differences in their amplified products in 6% PAGE gels.

Forward primer of primer pairs detecting polymorphism between the four mapping population parents were synthesized again by adding a universal M13 sequence (5'CACGACGTTGTAAAACGAC3') at the 5' end of the primer. Addition of a fluorescently-labeled M13 primer into the PCR reaction, along with reverse and forward primers (with M13 sequence), adds the fluorescent label to the PCR amplicon. This facilitates high throughput genotyping and marginally reduces the cost of such genotyping.

3.2 Mapping of EST-SSR markers

PCR conditions for EST-SSR primer pairs detecting polymorphism between shoot fly resistance mapping population parents (BTx623 and IS 18551) were optimized again with M13 tails. For this purpose, forward primer with M13 sequence, reverse primer and fluorescently labeled M13 primer (FAM / HEX / NED) were mixed in a ratio of 1:10:10, respectively. With this concentration ratio, only some of the markers were worked. Remaining markers were tried with the increased primer concentrations of 1:5:5 ratio or 1:2:2 ratio. Finally, markers were optimized at one ratio (1:2:2 primer concentrations) and used for screening progenies from the shoot fly mapping RIL population.

3.2.1 Recombinant Inbred Population (RIL)

A sorghum recombinant inbred population (RIL) targeted for shoot fly resistance available in ICRISAT was used for mapping a portion of the newly developed EST-SSR markers. The mapping population was derived from the cross between inbred genotypes BTx623 and IS 18551. The RIL mapping population consists of 252 inbred lines. DNA samples for these lines were arranged in three 96-well plates. Each plate consists of 94 RIL individuals along with the two parents, BTx623 and IS 18551.

3.2.2. Screening of RIL population with polymorphic EST-SSRs

All optimized polymorphic markers with M-13 tails were screened against 252 inbred lines of shoot fly mapping population. The same PCR protocol was used with some modifications using M13-tails as follows:

Component	Working Conc.	Final Conc.	Volume
DNA	**2.5 ng/µl**	**2.5 ng**	**1.0 µl**
Primer			
M13-tailed forward	**2.0 pm/µl**	**0.08 pm/µl**	**0.2 µl**
M13-Label	**2.0 pm/µl**	**0.16 pm/µl**	**0.4 µl**
Reverse primer	**2.0 pm/µl**	**0.16 pm/µl**	**0.4 µl**
$MgCl_2$	**10 mM**	**2 mM**	**1.0 µl**
dNTPs	**2 mM**	**0.12 mM**	**0.3 µl**
Buffer	**10 X**	**1 X**	**0.5 µl**
Enzyme	**0.5 U/µl**	**0.1 U**	**0.2 µl**
(AmpliTaq Gold®, Applied Biosystems, USA)			
Water			**1.0 µl**
		Total	5.0 µl

The same PCR program, Touchdown 61-51, was used for amplification of all polymorphic markers. After amplification, PCR products were pooled based on their amplicon size. Pooling was done using the following protocol:

Component	**Volume**
Hi-Di Formamide	7.0 µl
ROX size standard	0.1 µl
PCR products	
FAM labeled	1.0 µl
HEX labeled	1.0 µl

NED labeled	1.0 µl
Water	2.9 µl
Total	**10.0 µl**

Pooled PCR products of different markers labeled with different fluorescent dyes were denatured at 95˚C for 5 minutes. Capillary electrophoresis of denatured pooled products was performed using ABI3130*xl* DNA genetic analyzer. After completion of electrophoresis run, the raw data files created by the ABI machine were processed through GENESCAN 3.7 v software (Applied Biosystems) for sizing the PCR amplified fragments based on their relative mobility compared to the internal ROX size standards. Allele calling was done by GENOTYPER 3.7 v software (Applied Biosystems).

3.2.3 Data scoring

Based on the amplicon sizes, data were scored for all optimized primers run against template DNA from the RIL individuals. Data points were scored as 'A', 'B', 'H' or '-'.

'A' – Allele of female parent (BTx623)
'B' – Allele of male parent (IS 18551)
'H '– Heterozygous (presence of both parental alleles)
'-' – Missing data (failed amplification)

3.2.4 Linkage map construction

The segregation data was used to place the new markers onto the existing skeleton linkage map with 109 previously mapped SSR markers (Folkertsma et al. 2003) at LOD scores of 3.0 using MAPMAKER v 3.0 (Lander et al. 1987). Recombination frequencies were converted into centiMorgan (cM) distances using the Haldane mapping function (Haldane 1919). Linkage groups were designated as per the nomenclature suggested by Kim et al. (2005b). Previously mapped markers were used as anchor markers to assign the new markers into linkage groups. Using the commands 'compare', 'order' and 'ripple', each LG was constructed using already mapped markers in the previous studies as anchor markers. 'Anchor' and 'framework' commands were used to designate a group of markers to a particular chromosome (SBI). Mapmaker output was used in the Mapchart program to draw the linkage map.

3.3 Comparative genome mapping

The mapping positions of the EST-SSR markers on sorghum linkage map were compared with comparative maps between rice and sorghum developed by the JCVI database (as on 12th Jan, 2007). These comparative maps in JCVI were developed by calculating the colinear order of sorghum markers on the sorghum genetic map and on rice chromosome pseudomolecules. High density genetic recombination maps for sorghum were downloaded from the University of Georgia. The corresponding marker sequences were retrieved from GenBank and used to search the rice pseudomolecules.

3.4 Development of sequence based *in-silico* map of sorghum

Aligned 8X sorghum genome sequence information, made available for public use (Sorbi v1.0) (http://genome.jgi-psf.org/Sorbi1/Sorbi1.info.html), was downloaded and formatted into a database on the High Performance Computer (HPC) available at ICRISAT. This aligned sorghum genome sequence contains 659 Mb (89% of 739 Mb) arranged in 10 chromosome pairs, while the remaining 79 Mb is in super clusters and are yet to be assigned to individual chromosomes.

The sequences of all publicly available sorghum SSR primer pairs were collected and collated in FASTA format. This included both genomic SSR markers [*Xtxp* series (Kong et al., 2000; Bhattramakki et al., 2000), *XSb* series (Taramino et al., 1997), *Xgap* series (Brown et al., 1996), *Xisp* series (unpublished)], cDNA-based SSR markers [*Xcup* series (Schloss et al., 2002)], EST-based SSR markers including the *Xisep* series from the present study and the *Xiabt* series developed at the University of Dharward, Karnataka, India, (unpublished)] and other SSR markers of unknown type [*gpsb* and *mSbCIR* series (unpublished sorghum SSR markers from CIRAD, France)] and other marker systems such as Conserved Intron Scanning Primers (CISPs, Feltus et al., 2006b). Available gene sequences in the public domain and primer sequences (published and unpublished) were then BLAST searched against the sorghum genome sequence information. Searches were performed separately for forward and reverse primer sequences.

For both forward and reverse sequences, BLAST results were converted into a relational database using simple scripts (written in Visual Basic) to parse BLAST output (Altschul

et al. 1990). When the best BLAST hits of an individual marker's primer sequences (forward and reverse) were grouped together, then the BLAST results were analyzed using the following criteria:

1. Presence of forward and reverse primer pair BLAST hits on the same sorghum chromosome.
2. Occurrence of BLAST hits of forward and reverse primer sequences (for the same primer pair) on opposite strands (positive and negative) of the sequence assembly of the sorghum genome.
3. Estimated difference between BLAST hits of forward and reverse primer sequence hits (for the same primer pair) in terms of start and end nucleotide numbers on the genome sequence (expected amplicon size), being less than 1.5Kb.

For aligning the markers on the genome sequence assembly, each chromosome was compartmented into cells of 1 Mb each (e.g. Chromosome 1 being 73.8 Mb long, it was divided into 74 cells). Then primers satisfying the above criteria were placed on this sequence-based physical map of the sorghum genome, by allocating each primer pair to a cell of 1Mb along the complete chromosome sequence. This allocation was based on matching BLAST hit positions of primer pair to progressive units of 1Mb along the sequence assembly of each sorghum chromosome. Likewise, all the markers meeting the above requirements were placed on to the sequence-based physical map of sorghum. Any primer pair that did not meet one or more of the above mentioned criteria was excluded from the physical map. Genes were mapped based on their hit location on aligned sorghum genome sequence. All the linkage groups were named following Kim et al. (2005b).

3.5 Diversity analysis

3.5.1 Plant material

Reference set comprising of 384 sorghum accessions (Appendix 1), assembled from sorghum global composite germplasm collection under Generation Challenge Program (GCP) project, was used for diversity analysis using EST-SSR markers mapped in this

present study. This reference set represents the diversity present in sorghum global composite germplasm collection which contains 3367 accessions collected from different parts of the world, that were screened with 41 genomic SSR markers. This reference set contains a mini-core collection (176 genotypes) from CIRAD (France), landraces, important parental genotypes involved in molecular breeding programs at ICRISAT and some wild genotypes.

DNA for the reference set was isolated from 10-days old seedlings using the DNA extraction protocol described by Mace et al. (2003). DNA quantity and quality were assessed using 0.8 per cent agarose gels. Based on the relative intensity when compared with uncut lambda (λ) DNA standards, DNA concentration of each sample was estimated and then normalized to 2.5 ng/µl.

3.5.2 Selection of EST-SSR markers for diversity analysis

After mapping EST-SSR markers onto previously existing skeleton map of the sorghum shoot fly resistance mapping population, markers were selected based on their mapping position on each linkage group. At least three to four markers per linkage group were selected based on their position to provide reasonably complete genome coverage. In total, 45 markers were selected based on their positions. Multiplex sets were created using all selected markers based on expected amplicon size of BTx623. Each multiplex set consists of 3 markers, labeled with different dyes.

3.5.3 Screening of reference set with EST-SSR markers

M13-tailed forward primers were used for genotyping reference set of sorghum. DNA samples of reference set consisting 384 accessions were arranged in five 96-well plates. Each plate consisted of 92 genotypes and 4 standard entries. Standards included were BTx623 and the pools of 3 to 4 standard genotypes: Control A (pool of 3 accessions), Control B (pool of 3 accessions) and Control C (pool of 4 accessions). The PCR conditions and PCR program used were the same as used for the mapping experiment. After PCR was completed, PCR products were pooled. After denaturing the pooled PCR products along with Hi-Di Formamide and the internal ROX size standard, capillary electrophoresis was performed using ABI3130*xl* DNA analyzer.

3.5.4 Allele calling

Raw allele calls obtained directly from Genotyper software were processed through AlleloBin program (developed at ICRISAT, unpublished). AlleloBin uses repeat motif unit size as a reference to assign the base pair (bp) value into discrete bins. All the markers screened against the reference set were processed through AlleloBin to get corrected allele calls. Processed allele calls for each of the markers were combined for further analysis.

3.5.5 Statistical parameters

3.5.5.1 Summary statistics

Summary statistics for all the markers was derived using PowerMarker v 3.25 software (Liu and Muse, 2005). This software uses the following formulas to calculate different parameters:

Major allele frequency

$$\text{Major allele frequency} = \frac{\text{Number of genotypes having major allele}}{\text{Total number of genotypes}} \times 100$$

Gene diversity

Gene diversity, often referred to as expected heterozygosity, is defined as the probability that two randomly chosen alleles from the population are different. An unbiased estimator of gene diversity at the l^{th} locus is

$$H_e = (1 - \sum_{i=1}^{n} P_{il}^{\ 2}) / (1 - \frac{1+f}{n})$$

where

$P_i = i^{th}$ allele frequency

f = inbreeding coefficient

n = number of individuals

Heterozygosity

Heterozygosity is 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

Polymorphism information content

As per Botstein et al. (1980), polymorphism information content (PIC) was 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} alleles

3.5.5.2 Dissimilarity matrix

Processed data from AlleloBin was directly used for calculating the dissimilarity matrix using DARwin 5.0 software (Perrier et al. 2003). Dissimilarity was calculated by pair-wise simple matching using the following formula:

$$d_{ij} = 1 - \frac{1}{L} \sum_{i=1}^{L} \frac{mi}{\Pi}$$

where

d_{ij} : dissimilarity between units i and j

L : number of loci

π : ploidy

m_i : number of matching alleles for locus i

3.5.5.3 Factorial Analysis

Principal coordinate analysis (PCoA) is a member of the factorial analysis family working on distance matrices and is related to multidimensional scaling methods (MDS). It considers the high dimensional space defined by the distances between units two by two. The output is a list of coordinates of each unit on each axis that are sufficient to exhibit the main structure of the data. Factorial methods aim mainly to give an overall representation of diversity and are not really affected by individual effects. The simple-matching dissimilarity matrix was used to perform the factorial analysis using DARwin 5.0 software.

3.5.5.4 Dendogram/Tree construction

The un-weighted neighbor joining (NJ) method as implemented in DARwin 5.0 software was used to generate dendogram using the simple-matching dissimilarity matrix to determine the aggregation of the accessions into clusters. Un-weighted neighbor joining gives a same unitary weight to all units.

4. RESULTS

4.1 EST-SSR marker development

4.1.1 EST sequences for sorghum and SSRs identification

All 187,282 EST sequences available for sorghum in the TIGR database were downloaded in FASTA format. EST sequences of more than 200 bp only were considered for identification of SSR using SSRIT. A total of 39,106 sorghum EST sequences were found to contain microsatellites. Among SSR containing EST sequences, only 7.9% belonged to Class I (motif length ≥ 20 nucleotides) and remaining were belonged to class II type (motif length < 20 nucleotides). Frequency of SSRs in sorghum EST sequences was 1 SSR/1.79 kb or 0.56 SSRs/kb.

Clustering of SSR-containing EST sequences using Cap3 program eliminated redundancy in the dataset: 39,106 sorghum EST sequences (redundant) were reduced to 10.044 non-redundant EST sequences or unigene sequences that contain microsatellites motifs. In total, only about 5.3% of sorghum EST sequences were found with microsatellite motifs or 25.68% of non-redundant sorghum EST sequences are containing microsatellite motifs The AG class of di-nucleotide repeats which includes CT, GA and TC repeats and the CCG class of tri-nucleotide repeats which includes CCG, GGC, CGC, GCC, GCG and CGG were the most abundant in sorghum. The frequencies of tri-nucleotide repeats belonging to ATT (Isoleucine) and AAT (Asparagine) classes were very low (0.57% and 0.27%, respectively).

4.1.2 Identification of candidate SSR-containing non-redundant EST sequences having homology with rice

Rice-sorghum synteny was exploited for selection of a sub-set of non-redundant SSR-containing sorghum EST sequences. Sorghum EST sequences were BLAST searched against the rice genome in the GRAMENE database to get top ten hits on each of the five arbitrary regions of all rice chromosomes (Figure 4 and 5). Rice has 12 linkage groups and each was divided into 5 divisions. Each division was target to identify 10 candidate

(50 per linkage group) sorghum EST sequences containing SSRs, hence targeted to select 600 candidate sequences across the complete rice genome sequence.

To identify these 600 candidate sequences from 10,044 non-redundant sorghum ESTs containing SSR sequences, 1483 EST sequences (14.76%) were BLAST searched against the rice genome to obtain uniform and complete coverage on rice with the expectation that this would similarly provide uniform and complete coverage across the sorghum genome. Table 2 shows the BLAST results for these 1483 non-redundant sorghum EST sequences having homology with the rice genome.

The list of 600 putative sorghum EST containing SSR sequences selected based on BLAST results were given in Appendix 2. This contains the information on sorghum EST identity, BLAST score, on which chromosome of rice having the hit, nucleotide range spanned by the BLAST hit, and location on rice linkage group. Only the BLAST scores and their locations are listed in Table 3 (part of the sheet).

Among the selected 600 candidate sequences, 470 sequences (78%) had putative annotations (Appendix 2). The majority of these are classified as transcription factors or DNA binding proteins. Among the 600 selected sorghum ESTs containing SSRs, 63 were di-nucleotide repeats (10.5%); 425 were tri-nucleotide repeats (70.8%); 78 were tetra-nucleotide repeats (13%) and 34 were penta-nucleotide repeats (5.6%). The most abundant repeats among the selected EST sequences were AG and CCG classes as found across the full set of 10,044 non-redundant sorghum EST sequences containing SSRs. A total of 79 of these EST-SSRs (13%) were Class I and 496 EST-SSRs were Class II, while the remaining 25 EST-SSRs were only 10 nucleotides (five di-nucleotide units) in length. Distribution of microsatellites with in the selected 600 candidate sequences was given in Figure 6.

4.1.3 Primer Designing

For the selected 600 candidate SSR-containing non-redundant sorghum EST sequences, primer pairs were designed using Primer3 program after masking the repeat motif. These primers were named as **ICRISAT Sorghum EST Primer (ISEP)** pairs with four digit serial number. First two digits represent the linkage group of rice on which that particular

sorghum EST has the best hit and remaining two digits represent the serial number from top to bottom on that linkage group of rice.

4.2 Characterization of EST-SSRs in sorghum

Four sorghum inbred lines involved in mapping population development (BTx623, IS 18551, N 13 and E 36-1) were screened with all designed primer pairs. Out of 600 primer pairs, 457 (76.1%) amplified the template DNA and 386 (84.5%) of them produced simple and easy to score amplification products, whereas the remaining 15.5% produced multiple fragments that were difficult to score (Figure 7). Most of the primer pairs that produced no amplification or gave non-specific amplification targeted tri-nucleotide repeats. Very few primer pairs gave the amplicon sizes more than 500 bp while most of the primer pairs gave the amplicons sizes in the range of 200-250 bp (Appendix 2).

Polymorphism between the parents involved was scored simultaneously with PCR optimization. Of the 386 primer pairs that produced good amplification profiles, 133 primer pairs (22%) detected polymorphism between N13 and E 36-1 and 140 primer pairs (23%) detected polymorphism between BTx623 and IS 18551 (Appendix 2). In both crosses, di-nucleotide repeats were found most polymorphic (30-38%) followed by tetra-nucleotide repeats in the N13 × E 36-1 cross and penta-nucleotide repeats in the BTx623 × IS 18551 cross. The polymorphic markers and their expected position on each rice linkage group are listed in Table 4.

4.3 Mapping of EST-SSR markers

Among the candidate polymorphic markers between shoot fly mapping population parents (BTx623 and IS 18551), 24 markers (17%) targeted di-nucleotide repeats, 87 markers (62%) targeted tri-nucleotide repeats, 22 markers (16%) targeted tetra-nucleotide repeats and 7 markers (5%) targeted penta-nucleotide repeats. All the new polymorphic marker candidates (140 markers) were re-optimized with M13-tailed primer pairs. After adding the M13-tail to the forward primer at 5' end, PCR conditions were optimized for only 83 primer pairs (with 1:2:2 ratio). These 83 markers were screened against the template DNA sample from all 252 RIL individuals. PCR products were checked for successful PCR amplification in 1.2 % agarose gel (Figure 8). Denatured PCR products

(in multiplexes) were run through capillary electrophoresis in the ABI3130*xl* DNA genetic analyzer.

The ABI3130*xl* output files were processed through GeneScan 3.7 and data was analyzed using Genotyper software (Figure 9). An electropherogram showing the marker profile for *Xisep*0948 is shown in Figure 10 as an example. Segregation data for all the markers across all RIL individuals were scored and combined with the previously existing marker data for this RIL population for further analysis. Table 5 shows a partial sheet of segregation data for some of the polymorphic markers.

EST-SSR markers were mapped on to all 10 sorghum linkage groups (Figure 11). Out of 83 markers tested, 82 markers were ordered on to the ten sorghum linkage groups, with an average of 8.2 additional markers per linkage group. Number of new loci mapped on to each linkage group ranged from 5 (SBI-08) to 14 (SBI-03). One marker, *Xisep*0511 remained unlinked. After adding the new EST-SSR markers on to the linkage map, its length was extended to 2966.4 cM. Linkage group SBI-03 having largest number of SSR markers (30) spanned 254.3 cM. SBI-08 was the shortest linkage group with 238.8 cM, while SBI-01 was the longest linkage group (485.6 cM). EST-SSR markers were distributed evenly across all ten sorghum linkage groups and the resolution of the map was greatly increased. Many of newly added markers mapped to gaps in the previously existing linkage maps of sorghum.

4.3.1 SBI-01 (LG A)

Linkage group SBI-01 consisted of 26 SSR markers (including 8 new EST-SSRs) and spanned 485.6 cM, with an average inter-marker distance of 18.7 cM. EST-SSR markers mapped on this chromosome were *Xisep*0327, *Xisep*0949, *Xisep*0728, *Xisep*0504, *Xisep*1039, *Xisep*1032, *Xisep*1046 and *Xisep*1035. *Xisep*0728 mapped exactly in the gap between *Xtxp*302 and *Xcup*73. *Xisep*0327 and *Xisep*0949 markers mapped at the distal end of SBI-01, increasing the length of this linkage group. Numerous markers were clustered in the middle of this linkage group.

4.3.2 SBI-02 (LG B)

Linkage group SBI-02 consisted of 25 SSR markers (including 6 new EST-SSRs) and spanned a map distance of 297.7 cM, with an average inter-marker distance of 11.9 cM. EST-SSR markers mapped on this chromosome were *Xisep*0747, *Xisep*1145, *Xisep*0938, *Xisep*0733, *Xisep*0841 and *Xisep*0522. *Xisep*0747 and *Xisep*0522 were mapped at the distal and proximal ends of SBI-02, respectively.

4.3.3 SBI-03 (LG C)

Linkage group SBI-03 comprised of 30 SSR markers (including 14 new EST-SSRs) spanning 254.3 cM, with an average inter-marker distance of 8.4 cM. EST-SSR markers mapped on to this linkage group were *Xisep*0107, *Xisep*1218, *Xisep*1042, *Xisep*0101, *Xisep*1248, *Xisep*1012, *Xisep*1031, *Xisep*0114, *Xisep*0132, *Xisep*0102, *Xisep*0117, *Xisep*0843, *Xisep*0612 and *Xisep*0824. Several of them were mapped to the distal end and filled the gaps on SBI-03.

4.3.4 SBI-04 (LG D)

Linkage group SBI-04 consisted of 19 SSR markers (including 9 new EST-SSRs) and spanning 323.7 cM, with an average inter-marker distance of 17.0 cM. EST-SSR markers mapped on to this linkage group were *Xisep*0948, *Xisep*0210, *Xisep*0209, *Xisep*0228, *Xisep*0203, *Xisep*0746, *Xisep*1103, *Xisep*0234 and *Xisep*0242. *Xisep*1103 and *Xisep*0746 filled the gap between *Xtxp*12 and *Xtxp*41 in the previously existing linkage map.

4.3.5 SBI-05 (LG J)

Linkage group SBI-05 consisted of 18 SSR markers (including 9 new EST-SSRs) spanning 315.7 cM, with an average inter-marker distance of 17.5 cM. EST-SSR markers mapped on this chromosome were *Xisep*1202, *Xisep*1208, *Xisep*1109, *Xisep*1029, *Xisep*1128, *Xisep*1133, *Xisep*0713, *Xisep*1140 and *Xisep*0120. Four markers (*Xisep*1109, *Xisep*1029, *Xisep*1128 and *Xisep*1133) mapped in the gap between *Xtxp*94 and *Xtxp*15 on the previously existing skeleton map, while *Xisep*1140 was mapped between *Xtxp*225 and *Xtxp*262.

4.3.6 SBI-06 (LG I)

Linkage group SBI-06 comprised of 15 markers (including 7 new EST-SSRs) spanning 251.8 cM, with an average inter-marker distance of 16.8 cM. EST-SSR markers mapped on this chromosome were *Xisep*0444, *Xisep*0432, *Xisep*0617, *Xisep*0443, *Xisep*0422, *Xisep*0427 and *Xisep*0449. All except two of these markers (*Xisep*0444 and *Xisep*0432) mapped to the proximal end of SBI-06 whereas *Xisep*0444 and *Xisep*0432 mapped at the distal end of this chromosome filling the gap between markers *Xtxp*6 and *Xtxp*265 in the previously existing skeleton linkage map. Markers

4.3.7 SBI-07 (LG E)

Linkage group SBI-07 consisted of 13 SSR markers (including 7 new EST-SSRs) spanning 250.7 cM, with an average inter-marker distance of 19.3 cM. EST-SSRs mapped on this chromosome were *Xisep*0131, *Xisep*0805, *Xisep*0806, *Xisep*1250, *Xisep*0828, *Xisep*0829 and *Xisep*0716. Many of these markers had filled the gaps in the previously existing linkage map for this RIL population.

4.3.8 SBI-08 (LG H)

Linkage group SBI-08 comprised of 12 SSR markers (including 5 new EST-SSRs) spanning 238.8 cM, with an average inter-marker distance of 19.9 cM. This linkage group had the least number of EST-SSR markers (*Xisep*1231, *Xisep*0632, *Xisep*1150, *Xisep*0108 and *Xisep*0809). However all these new markers mapped to gaps between the previously mapped markers. *Xisep*1231 and *Xisep*0632 mapped at the distal end while *Xisep*0809 mapped at the proximal end of linkage group SBI-08, leading to an increase in the map length of this linkage group.

4.3.9 SBI-09 (LG F)

Linkage group SBI-09 consisted of 14 SSR markers (including 8 new EST-SSRs) spanning 255.9 cM, with an average inter-marker distance of 18.3 cM. EST-SSR markers mapped on this linkage were *Xisep*1014, *Xisep*0506, *Xisep*0513, *Xisep*0550, *Xisep*1008, *Xisep*1241, *Xisep*0523 and *Xisep*0125. All the markers were mapped to the top and middle portions of linkage group except *Xisep*0125 which mapped to proximal end of SBI-09. *Xisep*0125 was mapped almost 111 cM away from *Xtxp*10 greatly increasing the

map length. This increase in map length and position of this marker was confirmed by physical mapping based on sorghum genome sequence information.

4.3.10 SBI-10 (LG G)

Linkage group SBI-10 comprised of 19 SSR markers (including 9 new EST-SSRs) spanning 292 cM, with an average inter-marker distance of 15.7 cM. EST-SSR markers mapped to this chromosome were *Xisep*0604, *Xisep*0608, *Xisep*0607, *Xisep*0621, *Xisep*0625, *Xisep*0639, *Xisep*0630, *Xisep*1011 and *Xisep*1038.

4.4 Comparative mapping between sorghum and rice

The new EST-SSR markers reported here were developed based on sorghum EST sequences having similarities with rice genomic sequences (exploiting rice-sorghum synteny), hence these markers have a great potential for use in comparative mapping of rice and sorghum. Thus, the genetic linkage map positions of these EST-SSR markers in the sorghum genome were then compared with expected genomic positions of rice using comparative maps available in the TIGR database (Figure 12).

4.4.1 SBI-01

SBI-01 (chromosome 1 of sorghum) shows synteny with rice chromosomes 3, 10, 4, 11, 7, 8 and 6 with the strongest relationship involving rice chromosome 3 and 10 (Figure 1). EST-SSR markers developed based on sequence similarity with rice chromosome 10 mapped in the middle of sorghum SBI-01. Comparative maps also suggest that middle portion of SBI-01 is syntenic with rice chromosome 10.

4.4.2 SBI-02

SBI-02 (chromosome 2 of sorghum) reportedly has synteny with rice chromosomes 7, 9, 3, 1, 6 and 8. Out of six newly mapped EST-SSR markers, four markers mapped in the expected regions of the sorghum based on the previously available comparative maps in the JCVI database.

4.4.3 SBI-03

SBI-03 (chromosome 3 of sorghum) is reported to have synteny with rice chromosomes 1, 5, 3 and 2. However, 6 of these markers targeted for rice chromosome 1 mapped to

SBI-03 of sorghum. Remaining 8 new sorghum EST-SSR markers mapped to this linkage group were derived from ESTs having the best hits on rice chromosomes 12, 10, 8 and 6.

4.4.4 SBI-04

SBI-04 (chromosome 4 of sorghum) reported to have synteny with rice chromosomes 2, 6, 3, 1, 5, 9, and 4. Six sorghum EST-SSR markers developed based on homology with rice chromosome 2 were mapped on sorghum chromosome 4, along with one each targeting rice chromosomes 7, 9 and 11.

4.4.5 SBI-05

SBI-05 (chromosome 5 of sorghum) has synteny with rice chromosomes 11, 1, 2 and 12. Most of the new markers mapped in the middle portion of this linkage group were targeted based the homology with rice chromosome 11. Two markers having the homology with rice chromosome 12 and one marker with homology to rice chromosome 1 having synteny with SBI-05 were also mapped along with one marker each having homology with rice chromosome 7 and 10 that were not expected to exhibit synteny with SBI-05.

4.4.6 SBI-06

SBI-06 (chromosome 6 of sorghum) has synteny with rice chromosomes 4, 2 and 6. All the new markers mapped on SBI-06 have the synteny with rice chromosome 4 except one marker exhibiting the synteny with rice chromosome 6.

4.4.7 SBI-07

SBI-07 (chromosome 7 of sorghum) has synteny only with rice chromosome 8. Four new markers mapped on SBI-07 having the synteny with rice chromosome 8. In addition, one new marker each having homology with portions of rice chromosome 1, 7 and 12 were also mapped on SBI-07.

4.4.8 SBI-08

SBI-08 (chromosome 8 of sorghum) has synteny with rice chromosomes 12, 11 and 3. Out of five newly mapped EST-SSR markers on this linkage group, only two, *Xisep*1231 and *Xisep*1150 are based on ESTs having synteny with rice chromosomes 12 and 11, respectively. The remaining three new EST-SSR markers mapped to SBI-08 based on

ESTs having the homology with portions of rice chromosome 1, 6 and 8, although linkage relationships for *Xisep*0108 and *Xisep*0809 were weak.

4.4.9 SBI-09

SBI-09 (chromosome 9 of sorghum) has synteny with rice chromosome 5, 3, 1 and 7. Four of new EST-SSR markers mapped on SBI-09 have syntenic relationship with rice chromosome 5, spanned from top to the middle of SBI-09. *Xisep*0125, which loosely mapped to the distal end of SBI-09, has synteny with rice chromosome 1. In addition, two new markers having synteny to rice chromosome 10 and one with synteny to rice chromosome 12, unexpectedly mapped to SBI-09.

4.4.10 SBI-10

SBI-10 (chromosome 10 of sorghum) has synteny with rice chromosome 6, 2, 1, 4, 10 and 5. However, only new EST-SSR markers targeting rice chromosomes 6 and 10 were mapped to SBI-10. Markers having synteny with rice chromosome 6 were mapped at the top and middle of chromosome whereas markers having synteny with rice chromosome 10 were mapped at the bottom of chromosome, SBI-10.

4.5 Physical map development for sorghum using genome sequencing

Aligned sorghum genome sequence (http://genome.jgi-psf.org/Sorbi1/Sorbi1.info.html; Paterson et al. 2009) was converted into a database. Aligned sequence contains 659 Mb (89% of the expected sorghum genome of 738.5 Mb) arranged in 10 chromosome pairs while the remaining 79 Mb of the sorghum nuclear genome is in super clusters that are yet to be assigned to individual chromosomes (Table 6).

For aligning the markers on the genome sequence assembly, each chromosome was compartmented into cells of 1 Mb each (e.g. SBI-01 is 73.8 Mb long, so was divided into 74 cells). Figure 13 shows the physical map for SBI-01 of sorghum developed based on the sequence information derived from BTx623 genotype, as an example. Physical maps for the remaining nine linkage groups are given in Figure 14.

All SSR markers meeting criteria for placement, when aligned on the sequence-based physical map, were found to map at the distal regions of all chromosome arms.

Especially, SBI-03, SBI-06 and SBI-10 are devoid of even a single marker located proximal to their centromeric regions. In the remaining SBIs also, only one or two markers are located to the centromeric regions. Only nine markers could be placed on short of the chromosome 6. The heat maps from the sorghum genome sequencing project revealed that the gene rich regions (and nearly all SSR markers) are mainly located at the ends of the sorghum chromosomes. There is good agreement with these observations on the sequence-based physical map locations of the EST-SST markers developed in the present study.

4.5.1 Validation of physical map positions

For validation, markers already mapped on one or more sorghum genetic linkage maps (Chittenden et al., 1994; Bhattramakki et al., 2000; Tao et al., 2000; Klein et al., 2000; Haussmann et al. (2002a); Menz et al. 2002; Bowers et al., 2003; Tao et al., 2003; Haussmann et al. 2004) were used to compare their linkage map positions on the sequence-based physical map. Among the 369 SSR loci for which linkage map position were available in the public domain, 290 markers (78.5%) mapped to similar positions in both linkage and physical maps. Nine markers (2.4%) mapped on different chromosomes on the physical map than expected from their reported linkage map positions (this may be of putative duplicate loci), while 70 markers (18.9%) that have linkage map positions were not mapped on to the sorghum physical map (Figure 13 and 14). These are the markers not following the criteria as explained earlier. The data comparison related to physical map and linkage map positions for these differentially mapped markers revealed that, either primer pair(s) for these markers (20 out of 70) had BLAST hits on super clusters indicating gap in the genome sequence assembly or they were genetically mapping in the regions with poor marker densities and/or near to centromeric regions.

4.5.2 Validation of EST-SSR markers

Mapped EST-SSR markers were BLAST searched against the sorghum genome sequence to align them on the physical map. EST-SSR markers had expected physical map locations (based on their genetic linkage map positions) for 74 of 79 markers that this was possible to assess. Five markers were mapped onto different linkage groups in physical map in comparison with linkage map and three markers which were mapped in linkage studies were not aligned on the physical map. Linkage map positions in

comparison with the physical map positions for all mapped markers are summarized in Table 7. In general, high degree of collinearity and synteny of the EST-SSR markers was observed between the linkage and physical maps of sorghum.

4.6 Integration of linkage and physical maps

In the physical map, each linkage group was divided into a number of bins based on their size; each bin is equivalent to 1 Mb. The linkage map positions for these newly added EST-SSR markers on sorghum linkage map are integrated with physical map positions (Figure 14). Order of newly mapped EST-SSR markers is in agree with the positions of sequence-based physical map of sorghum. The linkage map developed in this study was extended to 2966.4 cM. When the linkage map is integrated with the physical map, each bin (1 Mb) constituted an average of 4.47 cM, in other words, on average 1 cM = 223 kb on the physical map. However, average bin (1 Mb) size on different chromosomes ranged from 3.391 cM (SBI-03) to 6.562 cM (SBI-01). If we consider in terms of recombination frequencies, each cM ranged from 152 kb (SBI-01) to 295 kb (SBI-03) on the physical map (Table 8). Agronomically important genes and their location on different chromosomes of sorghum were listed in Table 9.

4.7 Diversity analysis

After extracting DNA from sorghum reference set, DNA samples were checked in 0.8% agarose gel to estimate the DNA concentration (Figure 15). Based on their relative intensity of the bands in comparison with uncut λ standards, DNA quantity was estimated and diluted further to 2.5 ng/µl (working stock concentrations).

In total, 45 EST-SSR markers were selected based on their linkage map position representing the entire genome, to screen the reference set of sorghum. After PCR, products were checked for amplification in 1.2% agarose gels (Figure 16). Based on their expected sizes (according to optimization on BTx623 genotype), these markers were defined into 15 multiplex sets (Table 10). Samples were run through the ABI3130*xl* for capillary electrophoresis according to their multiplexes. GENESCAN and GENOTYPER softwares were used for scoring the allele for each marker. Figure 17 showing the electropherograms for *Xisep*0101 (a) and *Xisep*0841 (b).

Alleles were scored for each individual entry across all the markers. Raw allele calls from Genotyper were first processed through the AlleloBin program to obtain adjusted calls using the marker repeat motif size. Raw allele calls and processed allele calls for *Xisep*0101 marker for some of individual entries included in the reference set are listed in Table 11. Called allelic data for each marker across entries was combined and analyzed for diversity.

Five markers (*Xisep*0209, *Xisep*0621, *Xisep*0728, *Xisep*1128 and *Xisep*1150) were removed from the data analysis because of poor quality and many missing data points. In order to assess the level of genetic diversity, basic statistics were computed for remaining 40 EST-SSR markers using PowerMarker v 3.25 software. Diversity parameters forest-SSR markers screened against the reference set of sorghum are listed in Table 12.

4.7.1 Availability of marker data

Availability of marker data ranged from 91 to 100%. There were no missing data points for the marker *Xisep*0101 whereas 8.8% of data points were missing for *Xisep*0108. On average, missing data across all the markers was 2.1%.

4.7.2 Alleles in reference set

All 40 EST-SSR markers screened against the sorghum reference set were polymorphic. A total of 362 alleles were observed with an average of 9 alleles/marker (Table 12). Out of 362 alleles observed, 318 (88%) were from cultivated sorghum accessions where as wild genotypes had contributed 257 alleles (71%). Number of alleles per locus ranged from 3 (*Xisep*0117) to 38 (*Xisep*1012).

4.7.3 Allele distribution according to the race

Maximum numbers of alleles were present in caudatum (207 alleles, 57.2%) followed by bicolor race (205 alleles, 56.6%). Durra and guinea race accessions individually contributed 180 alleles (50.0%) where as kafir race contribution was limited to only 30.0% of total alleles. Among the guinea race sub-groups, guinea margaritiferum (Gma) contributed 91 alleles (25.0%). Five markers (*Xisep*0114, *Xisep*0422, *Xisep*0829, *Xisep*1038 and *Xisep*1231) were found to be monomorphic across the kafir race accessions while one marker was found as monomorphic across the caudatum or guinea races. *Xisep*0120 and *Xisep*0422 markers were found monomorphic in caudatum and

guinea race, respectively. With in the guinea race, Gma race accessions alone had nine monomorphic markers, namely, *Xisep*0242, *Xisep*0422, *Xisep*0444, *Xisep*0607, *Xisep*0639, *Xisep*0824, *Xisep*0829, *Xisep*0948 and *Xisep*1035. No such monomorphic markers were found in races bicolor and durra.

4.7.4 Allele distribution according to geographic origin

Genotypes originating from the African continent contributed 324 alleles (89.5%) to total alleles where as 303 alleles (84%) were contributed by accessions originating from other parts of the world. All most equal percentage of alleles were added by the accessions derived from North America (229 alleles, 63%) and Asia (221 alleles, 61%). No monomorphic alleles were detected in any of these three geographic origins. Among the African-origin genotypes, Central African genotypes contributed 183 alleles (50.5%), East African genotypes added 247 alleles (68.0%), South African genotypes contributed 208 alleles (57.4%) and West African genotypes contributed 226 alleles (62.4%) to the total alleles.

Maximum number of alleles (38 alleles) was reported for *Xisep*1012. In these 35 alleles (92.0%) were contributed from genotypes originating from Africa where as 31 alleles (81.5%) were contributed by the genotypes originating from elsewhere across the world.

Reference collection samples from Asia were divided into three groups, namely, East Asia, Middle East and India. All the markers are polymorphic across the genotypes having Indian origins and these contributed 175 alleles (48.3%). Genotypes derived from East Asia added 127 alleles (35.0%) where as Middle East genotypes contributed 139 alleles (38.3%). In the latter two cases, three markers were found monomorphic across the respective accessions. *Xisep*0203, *Xisep*1038 and *Xisep*1231 were found monomorphic in accessions from East Asia accessions while *Xisep*0422, *Xisep*0504 and *Xisep*1231 were monomorphic in accessions from the Middle East. In both cases, *Xisep*1231 marker was found monomorphic.

4.7.5 Unique and rare alleles

The 361 cultivated sorghum genotypes reported 105 unique alleles while the 23 wild genotypes had only 44 unique alleles. Among the 44 unique alleles in wild genotypes, 13

alleles were contributed by only one accession, IS 18868 which belongs to *verticilliflorum* subspecies, and has passport data indicating its origin (unlikely) as USA. 213 alleles were detected as common between cultivated and wild genotypes.

A large number of alleles (139 alleles, 38.3% at frequency <1%) were found to be rare alleles. Only for three markers, *Xisep*0117, *Xisep*0523 and *Xisep*1029, rare alleles were not reported. Maximum number of rare alleles were reported for *Xisep*1012 (13 alleles) followed by *Xisep*0829 (11 alleles). *Xisep*0504 and *Xisep*0938 reported only one rare allele each. Rare alleles reported for both of these markers were present only in wild genotypes. 200 bp allele for *Xisep*0504 was reported in only one individual (IS 23166) whereas 205 bp allele for *Xisep*0938 was reported in three genotypes (IS 18874, IS 18800 and IS 18876). Two of them belonging to subspecies *verticilliflorum* and originated from African continent while IS 18874 belongs to subspecies *bicolor* and has passport data indicating its origin (unlikely) as North America.

4.7.6 Polymorphic Information Content (PIC)

PIC values for the markers used in this study ranged from 0.13 (*Xisep*0120) to 0.94 (*Xisep*1012). *Xisep*1012 was more informative as it was observed with more number of alleles (38) with similar frequencies hence a larger PIC value. Observed heterozygosity varied widely from 0.002 (*Xisep*1038) to 0.11 (*Xisep*1103) with a mean of 0.03. Gene diversity also varied greatly from 0.14 (*Xisep*0120) to 0.95 (*Xisep*1012) (Table 12).

4.7.7 Cluster analysis

Cluster analysis was carried out using DARwin software. Processed allele calls from AlleloBin for all markers were directly used to calculate a dissimilarity matrix. Dissimilarities between all pairs of individuals were estimated based on simple matching. This matrix was used for the factorial analysis which gave the preliminary groups among the accessions. Figures18 and 19 shows the results of factorial analysis (axis 2 and 3 only) based on the race and geographic origin of sorghum accessions, respectively.

Factorial analysis (FA) clearly identified five different clusters of genotype according to their races. Genotypes belonging to Guinea margaretiferum race were clustered in the top of quadrant II of the factorial diagram. Wild genotypes were clustered in the middle of

the intersection between quadrants I and II. Genotypes belonging to kafir race were clustered in quadrant III while genotypes belongings to caudatum were grouped in quadrants IV and I. Guinea race genotypes also clustered in between quadrant IV. Durra race grouped together away from other clusters (which was clear on axis 1 and 2, not shown). Bicolor race and intermediate races did not have any specific clusters in factorial analysis.

Based on geographic origin genotypes were clustered into five different groups. All guinea margaretiferum accessions from West Africa were clustered at the top of quadrant II. Majority of genotypes from South Africa were clustered in quadrant III while those from East Africa were clustered in quadrants I and IV. Another group of accessions having West Africa origins were clustered in quadrant IV. Accessions having Indian origin were clustered away from all other clusters (which was clear on axis 1 and 5, not shown).

4.7.8 Dendogram/Tree

The dissimilarity matrix also was used to generate dendograms following the un-weighted Neighbor-Joining (NJ) analysis method implemented in DARwin. Dendograms were developed using un-weighted NJ analysis method are displayed in Figures 20 to 24.

NJ analysis revealed the same type of clustering as the factorial analysis. References set accessions were clustered according to geographic origin and their races. However, for better understanding, clusters were divided into eight groups. Accessions originating from West Africa (AfricaW) were grouped into 2 separate clusters. One cluster (VII) was exclusively contained the accessions belonging to guinea margaretiferum (Gma) which formed a group along with wild genotypes in cluster VIII. Another cluster of accessions were mostly belonged to guinea (G) and their intermediate races in cluster II. Cluster III comprised mainly of accessions belonging to the kafir (K) race most of which originated from South Africa (AfricaS). This cluster also included other accessions derived from South Africa belonging to race bicolor (B) and intermediate race durra-caudatum (DC) derived from East Africa.

Accessions originated from India mainly grouped into two clusters. One group of accessions of race guinea was in cluster IV along with some other guinea and

intermediate race materials from South Africa (AfricaS). A second larger group of accessions originated from India were in cluster V, which was a mostly durra race materials. This cluster included the accessions originating from both India and East Africa (AfricaE) belonging to durra race.

Accessions having East Africa (AfricaE) origin were mainly grouped mainly into two different clusters. Cluster I mainly comprised of accessions belong to caudatum race. Within cluster I, all caudatum breeding lines grouped together whereas landrace caudatums grouped separately within the same cluster (Figure 24). Another group of accessions having East Africa (AfricaE) origin clustered in VI, which comprised of accessions belonging to bicolor and intermediate races with caudatum. Cluster I also included a group of accessions originating from Central Africa (AfricaC). These accessions also formed two different groups in Cluster I. One group of accessions with caudatum race clustered together whereas another group of accessions having the intermediate races between caudatum and guinea were grouped.

Cluster VI included different accessions grouped according to their geographic origin. This group included accessions having the geographic origins from all parts of the world including Africa continent. With in the cluster IV, accessions of caudatum and their intermediate races originated from East Asia (AsiaE) grouped together. Intermediate race between durra and caudatum (DC) originating from the Middle East (MidE) were grouped. Similarly, bicolor race accessions having passport data indicating North American (AmericaN) origin grouped together in this same cluster.

Most wild genotypes were grouped in cluster VIII. There was no specific cluster differentiation according to their geographic origin. But there was clear demarcation in grouping pattern according to their subspecies. Nearly all *arundinaceum* and *verticilliflorum* subspecies were grouped separately in cluster VIII where as a*ethiopicuum* and *drummodii* subspecies genotypes were scattered. Surprisingly some of the landrace accessions were also grouped into the same cluster along with wild genotypes.

Breeding materials related to caudatum race were grouped together in cluster I while materials belonging to kafir race were grouped together in cluster III. The remaining breeding materials were scattered in clusters IV and VI according to their geographic

region. Landraces were grouped together into separate clusters according to their geographic distribution. Wild accessions were grouped in cluster VIII along with some landraces (Figure 24).

BTx623 is the standard sorghum genotype and recurrent parent in the shootfly resistance marker-assisted backcrossing programs. This genotype was grouped in cluster I. Source genotype for stay-green trait, E36-1, grouped together surprisingly with recurrent parent for stay-green trait, S35 in the cluster I. As both are relatively elite caudatum background breeding materials, this is probably not an unexpected result. Other important genotypes clustered together within cluster I are IRAT 204 (recurrent parent for stay-green trait) and PB 15520 (stem borer resistant and sorghum midge susceptible mapping population parent). ICSV 745 (stem borer susceptible and midge resistant mapping population parent) was grouped in a separate cluster and away from PB 15520. B 35 (converted durra race and stay-green donor parent in markers assisted-backcrossing programs) was clustered in V, which mainly comprises the accessions having durra race.

5. DISCUSSION

Characterization of genetic variation at the molecular or DNA level is a main target in breeder's perspective in the present scenario. DNA markers have gained considerable importance in molecular breeding. A vast number of molecular markers, namely, RFLP, AFLP, RAPD, SSRs, SNPs, etc. were developed in different crop species. Every marker system has its own merits and demerits. During the initial stages of DNA-based marker development, RFLP was used by the breeders to study genetic relationships, develop linkage map and comparative genome mapping (Gale and Devos 1998). Because, RFLP markers require the use of radioactivity and the process is time consuming, invention of markers systems which are quick and show immediate results was sought. Thereby, PCR-based markers were evolved. Among different types of PCR-based markers, simple sequence repeats (SSRs) proved to be the best option for molecular breeding programs because of their simplicity, reproducibility, high-throughput and co-dominance in nature. Hence, SSR markers were widely used in studies pertaining to assessment of genetic relationships and trait mapping. SSR marker data generation is relatively cheap, but its initial cost of development is very high.

During early 2000s, the genome sequencing projects were initiated in model crops like *Arabidopsis* (The Arabidopsis Genome Initiative, 2000) and rice (International Rice Genome Sequencing Project, 2005). Then, molecular biologists focus was changed from identifying the basis of genetic variation using markers to the study of complete genome level variation. As an outcome of plant genome sequencing projects along with EST projects, a large number of DNA sequences flooded into nucleotide databases which were made available for public use. During this time, full length gene sequences along with their characterization were generated and made available. These sequences permitted researchers to identify new molecular tools for application in molecular breeding which include identification of SSRs, SNPs and COS markers.

Generation of genome sequences for model crop species enabled the geneticists to shift their focus towards functional genomics. A simple study analyzing the types of genes being expressed and their level of expression is to read a single-pass from either end of

cDNA clones to generate ESTs. Analysis of ESTs is a simple strategy to understand transcribed regions of the genome. In this context, ESTs or cDNA sequences gained much more attention in the functional genomics era. The sudden increase in the volume of sequence data generated from EST projects in several plant species facilitated the identification of SSRs in large numbers. SSRs present within EST or cDNA sequences were termed as 'EST-SSRs' or 'genic microsatellites' (reviewed by Varshney et al. 2005). Identification of EST-SSRs does not require much input and is less expensive than *de novo* identification of genomic SSRs by sequencing SSR enriched genomic libraries. EST-SSRs are valuable source of functional markers and once EST-SSRs were mapped, they will be associated with the genes carrying them. Thus generation of EST-derived SSR markers has become a complement to existing SSR collections. Application of markers like EST-SSRs in sorghum genome analysis holds great promise for producing lasting insight into processes by which novel genotypes are generated.

To exploit the advantages of genic microsatellites, this study was aimed at the identification of EST-SSR markers in sorghum by data mining from public databases and their applications in molecular breeding.

5.1 EST-SSRs identification and their distribution

Advances in bioinformatics helped in handing large amounts of sequence information. Different software tools like SSRIT (www.gramene.org/db/searches/ssrtool), TRF (www.tandem.bu.edu/trf), MISA (http://pgrc.ipk-gatersleben.de/misa), etc. are helpful in identifying the SSRs from the existing databases. J. Craig Venter Institute [formerly, The Institute for Genome Research (TIGR)] exclusively maintain EST sequences for both plants and animals. At the onset of this study (as of May, 2004), there were about 187,282 sorghum EST sequences available, representing about 98.9 Mb (13.3%) of sorghum genome (740 Mb, Paterson et al. 2009). Only 21% of these sequences were found to contain SSRs. Studies on identification of SSRs, estimating their frequency and their distribution was started early in 1990s (Morgante and de Oliviers 1993). The frequency of SSRs reported in the earlier studies was 1 SSR/64.6 kb in cereals. Such low frequency was mainly due to the availability of less number of genic and genomic sequences in the databases at that point of time. Later, many EST projects contributed several hundred sequences to databases, hence the density of SSRs also greatly increased

from 1 SSR/6.0 kb (Varshney et al. 2002) to 1 SSR/1.83 kb (Jayashree et al. 2006) in cereals. In case of sorghum, SSR frequency was 1 SSR/5.5 kb (Varshney et al. 2002) when 41.6 Mb (5.6% of genome) was available in the form of EST sequences. This availability was increased more than two-fold by 2004 to 98.9 Mb (13% of genome). The frequency was increased to three-fold (1 SSR/1.79 kb) in sorghum.

Tri-nucleotide repeats are the predominant nucleotide repeat in sorghum followed by di-nucleotide repeats as identified earlier in other cereals such as barely (Kantety et al., 2002; Varshney et al., 2002; Thiel et al., 2003), bread wheat (Gupta et al., 2003). Among the tri-nucleotides, CCG class (includes GGC, CGC, GCC, GCG and CGG), is the most abundant class in sorghum in conformance with most other studies (Eujayl et al. 2002, Kantety et al. 2002, Varshney et al. 2002, Gupta et al. 2003, Saha et al. 2004). CCG class triplet codes for proline, which is the characteristic amino acid accumulated in drought tolerant varieties in crop plants.

Di-nucleotide repeat class AG (including CT, GA and TC repeats) is the most abundant in sorghum as identified in barley (Thiel et al. 2003), wheat (Gupta et al. 2003) and in tall fescue grass (Saha et al. 2004). However, Kantety et al. (2002), Varshney et al. (2002), and Jayashre et al. (2006) have reported AG-TC as predominant class of di-nucleotide repeat by mining the nucleotide databases. Yu et al. (2004) also found the same class of di-nucleotide repeats abundant in wheat and the majority of di-nucleotide repeats were located in non-transcribed regions of the genome in contrast to tri-nucleotide repeats (Jayashree et al 2006). This indicates that AG and CCG repeats are the most abundant class of repeats in cereals and is supported by the EST-SSR survey in major cereals by Jayashree et al. (2006).

The Cap3 program helped in clustering of the sequences and identification of redundant sequences with in SSR-containing EST sequences in sorghum. This analysis determined that 1/3 of the EST sequences with SSRs were redundant. At the end, only 10,044 sequences remained as non-redundant EST sequences containing SSRs for sorghum. This redundancy is an inherent characteristic of random or shotgun sequencing within cDNA libraries (Weber and Myers 1997; Bouck et al. 1998). Clustered datasets can help in

designing the primer pairs effectively because it provides longer lengths of contigs (Jayashree et al. 2006).

5.1.1 Selection of candidate EST sequences containing SSRs

Sequence-based alignment studies in one species can accelerate the filling of gaps on linkage maps of related species and can be effectively used for comparative genome mapping across species (Klein et al. 2003). Thereby, this study aimed at the development of EST-SSR markers in sorghum to enrich the existing genetic linkage maps on the basis of sequence similarity with rice. Identification of sequence similarities of non-redundant sorghum EST sequences with rice genome sequence (rice-sorghum synteny) played a major role in selecting candidate EST sequences which putatively target the desirable portion or gaps in sorghum. About, 1,483 non-redundant sorghum EST sequences were BLAST searched against the rice genome sequence to identify 600 candidate sequences which span the complete genome of rice, to develop EST-SSRs in sorghum. When searching the homology of 1,483 non-redundant sorghum SSR-containing EST sequences against the rice genome, the highest alignment scores was obtained for those aligning the long arms of rice linkage groups (Appendix 2).

Rice chromosome 3 had the highest number of hits followed by chromosomes 1 and 11. This indicates that the corresponding genomic regions of sorghum, SBI-01, SBI-03 and SBI-05, are more syntenic to rice than other regions of the sorghum nuclear genome. In the sequence similarity search, 150 of these sorghum EST sequences (10.1%) did not have any hit on the rice genome. Klein et al. (2003) also found that 10% of the sorghum ESTs studied had no homologs in the rice genome. Interestingly, this agrees well with the observation that 7% of the predicted genes from the aligned 8X shotgun sequence of sorghum inbred BTx623 have no apparent homologs in *Arabidopsis*, rice or poplar (Paterson et al. 2009). Sorghum is less closely related to rice than to sugarcane and maize. Rice and the maize/sorghum lineage may have diverged ~50 million years ago and show much more chromosomal rearrangements (Paterson et al. 1995, 2000).

Even in the selected 600 sorghum candidate sequences, tri-nucleotide repeats (present in 425 sequences, 71%) were predominant followed by tetra-nucleotide repeats (78 sequences, 13%) (Figure 6). The most abundant di- and tri-nucleotide repeats in selected

sequences were the same as present in the complete data set, *i.e*. AG and CCG repeats, respectively. Putative annotation was available for 70% of the sequences and most of them were classified as transcription factors and DNA binding proteins. Class I type repeats (>20 nt) were present in 13% of selected non-redundant sorghum EST sequences. Primer pairs were designed to all selected candidate sequence using the Primer3 program and named as ICRISAT Sorghum EST Primers (ISEP).

5.2 Characterization of EST-SSR markers in sorghum

Out of 600 markers, 457 (76%) amplified PCR products across sorghum four parent panel, but 386 markers (84.5%) gave good and easily scorable PCR products. Among these, 65% of amplified markers were tri-nucleotide repeats belonging to the CCG class (proline encoding codon). Most of the primer pairs that produced no amplification or gave non-specific amplification were tri-nucleotide repeats. The amplification failure (130 primers, 24%) may be because of i) presence of intron-exon borders at a primer annealing site especially at the 3' end, ii) presence of large introns between two consecutive exons regions of a gene for which primer pairs were designed, and iii) may be presence of null allele. Similar results were observed in rice (Cho et al. 2000), and wheat (Eujayl et al. 2002; Gupta et al. 2003; Yu et al. 2004). Very few markers produced amplicon sizes greater than the expected size (Appendix 2). This variation in allele sizes is not likely due to repeat length variability with in the SSRs, as a result of insertions and deletions (indels). Similar results were observed with EST-SSRs developed in wheat (Yu et al. 2004).

In this study, a total of 169 EST-SSR markers (28%) detected polymorphism in a survey of four mapping population parental lines (shootfly and *Striga* resistance mapping population parents). Among these, 104 EST-SSR markers were polymorphic in both the populations whereas 36 and 29 markers detected polymorphism specific to the shoot fly and the *Striga* resistance mapping populations, respectively. Individually, only 22% and 23% of the EST-SSR markers developed were polymorphic among the *Striga* and shootfly resistance mapping population parents. A similar percentage of polymorphism (25%) was reported for EST-SSRs in durum wheat (Eujayl et al. 2002), whereas Thiel et al. (2003) reported 8-54% polymorphism on three different mapping population parental line pairs in barley, and Yu et al. (2004) observed a very high level of polymorphism

(43%) in wheat which is almost comparable to genomic SSRs (53%) (Roder et al. 1998). Di-nucleotide repeats were more polymorphic, in relation to the total number of di-nucleotide markers developed, (Appendix 2) than other repeat classes as reported by Thiel et al. (2003) in barley and Yu et al. (2004) in wheat. Polymorphic markers were re-synthesized with a universal M13-forward primer sequence at the 5' end and further screened against the shootfly resistance RIL population for mapping.

5.3 Mapping

Genomic SSR markers which were already mapped served as anchor markers to assign the newly developed sorghum EST-SSR markers on to respective linkage groups. The new EST-SSR markers were scattered across all ten linkage groups of sorghum (Figure 11). Linkage groups were named according to Kim et al. (2005b) who developed a karyotypic map based on FISH (Fluorescent *In Situ* Hybridization) technique where short arm of the chromosome are placed at the top and long arm at the bottom of linkage groups and named as SBI-01, -02, -03, -04, -05 -06, -07, -08, -09 and -10 corresponding to LG A, B, C, D, J, I, E, H, F and G of Menz et al. (2002), respectively. Maximum number of markers (14 markers) was mapped on sorghum chromosome 3. Among 14 EST-SSR markers, six markers mapped onto sorghum chromosome 3 have synteny with rice chromosome 1. This indicates that chromosome 1 of rice and chromosome 3 of sorghum exhibit considerable synteny. Same conclusions were made individually while comparing the genetic maps between sorghum and rice (Ventelon et al. 2001; Klein et al. 2003; Buell et al. 2005).

After incorporation of newly developed EST-SSR markers to previously existing skeleton map of sorghum, map distance was extended to 2966.4 cM encompassing 191 SSR markers (109 genomic SSRs and 82 EST-SSR markers) with an average distance of 15.5cM. Length of the linkage groups ranged from 238.8 cM (SBI-08) to 485.6 cM (SBI-01) (Figure 11). Number of EST-SSR markers mapped on each linkage group ranged from 5 markers on SBI-08 to 14 markers on SBI-03 (Figure 11). Linkage group SBI-01 was the longest (485.6 cM) with 26 markers. Among these markers, 18 markers were genomic SSRs and 8 markers were EST-SSRs developed in the present study.

The present map is longer than sorghum linkage maps previously reported by several authors (Chittenden et al. 1994; Pereira and Lee 1995; Haussmann et al. 2002a; Bhatramakki et al. 2002; Menz et al, 2002, Bowers et al. 2003, Haussmann et al. 2004; Wu and Huang 2006; Mace et al. 2008). The greater map distance can be attributed to the increase in the recombination frequencies with increase in the population size included in the analysis, and also because of increase in marker density. Previous reports in sorghum (Boivin et al. 1999) and in rice (Maheswaran et al. 1997) also revealed an increase in the map length when marker density was increased by addition of new markers. Although, genetic distances can vary between maps, marker locus order should remain the same between the maps of a single species. Marker order in the present map was comparable with other existing maps (Bhattramakki et al. 2000; Kim et al. 2005a; Wu and Huang 2006), however, few rearrangements in certain regions where markers were tightly linked was observed, which is not unexpected.

So far, most existing maps in sorghum were developed using both RFLP and SSR markers (Bhattramakki et al. 2000; Haussmann et al. 2004). Recently, linkage map were developed exclusively with SSRs for mapping greenbug resistance in sorghum (Wu and Huang 2006, 2008) and shootfly resistance (Folkertsma et al. 2003). New markers, however, filled gaps or marker-rare regions in the existing linkage map of sorghum for shoot fly resistance.

In the case of the genetic linkage map developed using SSRs in the cross between Westland A line and PI 550610, 16 linkage groups were reported instead of expected 10 linkage groups (Wu and Huang 2006). Only four chromosomes out of ten were reported with single linkage groups (SBI-01, SBI-02, SBI-05 and SBI-06). And other chromosomes were reported with two linkage groups named as a and b. But in the present study with SSR, all chromosomes were having only one linkage group, representing one linkage group for one chromosome as reported by Menz et al. (2002) and Kim et al. (2005a). This is because of tight linkage between the markers on each chromosome. Even though the distance is large between the consecutive markers in parts of some of the linkage groups, the positions for all SSR markers (genomic and genic) were confirmed through the use of genome sequence (Figure 13 and 14).

As targeted (using the sequence similarity between rice and sorghum), EST-SSR markers were mapped in all gaps in the existing linkage groups of sorghum (Figure 11). The gap between *Xtxp*302 and *Xcup*73 on SBI-01 in the previous linkage map was filled with one marker, *Xisep*0728 which reduced gap between adjacent markers to 57.2 cM and 80.2 cM with respect to *Xtxp*302 and *Xcup*73, respectively. A cluster of four markers filling the gap between *Xtxp*228 and *Xtxp*009, and another cluster of five markers between *Xtxp*009 and *Xtxp*033 on SBI-03 were added. A huge gap (107.3 cM) between *Xtxp*012 and *Xtxp*041 was filled with markers *Xisep*0746 and *Xisep*1103 on SBI-04. Markers in the present study bridged a big gap (160 cM) between *Xtxp*006 and *Xtxp*265 on SBI-06 with two EST-SSR markers (*Xisep*0444 and *Xisep*0432). A couple of markers on SBI-07 and SBI-08 greatly reduced existing gaps on these linkage groups between *Xtxp*040 and *Xtxp*159, and *Xtxp*047 and *Xisp*198, respectively. Three markers were mapped between *Xcup*49 and *Xtxp*020 on linkage group SBI-10. These markers can be expected to greatly enhance the rate of transferring of shootfly resistance from donor parent to recipient parent. This has proved that the comparative genome mapping approach will help to bridge gaps in the existing linkage maps of crops.

This approach in sorghum not only filled the gaps in the linkage maps, but also extended the map length towards telomeric regions. Markers *Xisep*0327 and *Xisep*0949 on SBI-01 increased map length by 77.4 cM at the short arm of the chromosome. The short arm of SBI-02 was extended by 8.9 cM with the addition of *Xisep*0747 while the long arm was extended by 43.2 cM with the addition of *Xisep*0522. SBI-05 was extended by 17.1 cM at the top of the chromosome by adding markers *Xisep*1202 and *Xisep*1208. 11.1 cM (*Xisep*0131) and 19.4 cM (*Xisep*0716) were added to SBI-07 on the short arm and long arm, respectively. *Xisep*1231 and *Xisep*0632 were added to the short arm of SBI-08 extending the map distance of 39.2 cM while addition of *Xisep*0809 at the distal end of the long arm increases the map distance by 57.7 cM. 111.9 cM was added at the bottom of linkage group SBI-09 by inclusion of loosely linked marker *Xisep*0125. In total, 374 cM was added to the previously existing linkage map of this RIL population (Folkertsma et al. 2003), thus increasing the map length to 2966.4 cM. This is ~250 cM larger than the sorghum linkage map with longest distance reported so far (Tao et al. 2000). Increasing the map lengths by the addition of new markers at the telomeric ends of the chromosome arms was confirmed by the presence of these markers in these positions on the physical map of sorghum (Figure 13 and 14).

On SBI-09, *Xtxp*10 was mapped at the end of the chromosome and no markers were reported beyond this (Folkertsma et al. 2003). But however, an additional 69.5 cM was reported beyond marker *Xtxp*10 on LG F (SBI-09) by Bhattramakki et al. 2000. Only one SSR (*Xtxp*339) was reported and remaining were RFLP markers. In the present study, SBI-09 was extended by 111 cM beyond *Xtxp*10 but with only one loosely linked EST-SSR marker, *Xisep*0125. This loose association is because of using Haldane function and expected to have approximately 45% of recombination.

5.4 Comparative genome mapping between sorghum and rice

Grass species show extensive conservation of gene order even among species with dramatically different genome sizes (Devos and Gale 1997; Gale and Devos 1998). Rather than developing extensive genome mapping projects in orphan crops independently, the incorporation of markers from closely related species using comparative genome mapping approach can provide the scaffold upon which genome maps can be developed (Paterson 2000). An enriched sorghum genetic linkage map opened a new avenue into functional comparative genome mapping. Common anchor points in genome greatly facilitate translational and comparative genome mapping and opportunity to improve our understanding of the evolutionary process at functional level.

Comparative maps were prepared in relation to the existing comparative maps available between rice and sorghum (Figure 12). Previously it has been found that rice chromosome 1 shares a high degree of synteny with sorghum chromosome 3 (SBI-03) (Ventelon et al. 2001; Klein et al. 2003; Buell et al. 2005). Among the SSR markers developed from sorghum EST sequences having homology with rice chromosome 1, ten were found to be polymorphic in the shootfly mapping population parents. Out of these ten polymorphic markers, six (*Xisep*0101, *Xisep*0102, *Xisep*0107, *Xisep*0114, *Xisep*0117 and *Xisep*0132) were mapped to SBI-03 and the remaining four markers, *Xisep*0108, *Xisep*0120, *Xisep*0125 and *Xise*0131, were mapped to SBI-08, SBI-05, SBI-09 and SBI-07, respectively. *Xisep*0120 and *Xisep*0125 were mapped at the distal ends of the long arm of SBI-05 and SBI-09, respectively. *Xisep*0125 is mapped almost 112 cM distance away from the nearest marker, greatly extending the length of this linkage group. This marker position was confirmed by physical map (Figure 13 and 14) developed based on

sequence information derived from BTx623. The comparative genome map for SBI-09 with rice chromosomes also assumes that a small portion of chromosome 1 of rice contributed at the bottom of SBI-09 of sorghum (Figure 12). As these markers were developed based on sequence similarity with rice, they were expected to map in the corresponding targeted regions of sorghum. However, some of the markers did not map to the expected target regions of sorghum. This is may be because of 1) gene families with dispersed distribution, 2) the presence of paralogs instead of homologs 3) the movement due to transposon, and 4) the fact that only 10% of non-redundant EST sequences with SSRs were used instead of using all 10,044 sequences, leaving lesser chance of getting sequences with perfect homology with rice to target them on sorghum genome.

During the process of evolution sorghum chromosome 3 (SBI-03) shared ancestral genomic regions that contributed to chromosomes 1, 2, 3, and 5 of rice (Figure 12). But when we compare the mapping positions, the markers having the synteny with rice chromosome 1 only mapped and the remaining markers did not show similarity as expected with rice chromosomes. The markers mapped other than those having the similarity with rice chromosome 1 suggest that other parts of the ancestral grass genome also contributed in the formation of sorghum chromosome 3 during the process of evolution.

In total 82 EST-SSR markers were mapped on ten linkage groups of sorghum. Of these, 57 markers (69.5%) exhibited synteny and co-linearity with their respective chromosomes of rice when we align the linkage group for comparison with the previously available maps (Figure 12). EST-SSR markers developed targeting rice chromosomes 4 and 11 were mapped to the corresponding syntenic linkage groups (SBI-06 and SBI-05) of sorghum, indicating high levels of synteny between the rice and sorghum nuclear genomes in these regions. Of 9 EST-SSR markers mapped on SBI-04, 7 markers shared perfect synteny with the respective chromosomes of rice. Out of 7 EST-SSR markers mapped on SBI-06 of sorghum, six (87.5%) were mapped on the same linkage group based on its syntenic relationship with rice chromosome 4. Sorghum genome sequence also confirms the synteny of SBI-06 with rice chromosome 4. Remaining EST-SSR marker mapped to SBI-06 was developed to target rice genome 6,

with which SBI-06 also shares the synteny. SBI-05 was mapped with 77.7% (7 of 9) markers having synteny to the expected regions. SBI-07, which is expected to share ancestry primarily with rice chromosome 8, exhibited the expected synteny at top and bottom of the linkage group (Figure 12). Markers have expected synteny with rice on the sorghum linkage map ranged from two (SBI-08) to nine (SBI-10). All nine markers mapped on SBI-10 of sorghum showed synteny with the expected rice chromosomes. SBI-10 shares a common ancestry with chromosome 1, 2, 4, 5, 6, and 10 of rice. Among these rice chromosomes, only markers developed targeting rice chromosome 6 and 10 were mapped on SBI-10.

Finger millet revealed an impressive amount of co-linearity and synteny with rice (Srinivasachary et al. 2007). In the present study, sorghum EST-SSR markers also have exhibited a reasonable amount of synteny and co-linearity with rice. Variation within the finger millet is very less (Dida et al. 2007). EST-SSR markers have reasonably good cross-species transferability (Varshney et al. 2005). Hence, the EST-SSR markers developed in this study can be evaluated as an alternate source of markers as well as anchor points for the development of functional comparative genome maps in between rice, sorghum and finger millet.

5.4.1 Comparative maps in Andropogoneae family: A future thrust

Sorghum has the smallest genome size among the Andropogoneae family member species. Maize and sugarcane have approximately 3-4 and 7-10 folds larger genomes, respectively, when compared to sorghum. Maize and sugarcane share considerable homology with sorghum (Paterson et al. 2009). Repetitive DNA is the major obstacle in chromosomal walking to isolate genes by positional cloning. This obstacle can be removd through parallel walks to the corresponding genes in closely related species (Kilian et al. 1995, Paterson et al. 2000). Because of its small genome size among the Andropogoneae family members, sorghum can occupy the central position in the crop circle of comparative genome maps for the Andropogoneae family. For map-based cloning of a gene though comparative genome mapping, fine-scale synteny mapping needs to be accelerated between sorghum, maize and sugarcane (Subudhi and Nguyen 2000). In the present map, the order of all the genomic SSR markers were collinear in comparison with previously published maps (Bhattramakki et al. 2002; Kim et al. 2005a; Wu and Huang

2006). This map was enriched with genic SSRs (EST-SSRs) derived from expressed regions of the genome. This opens new possibilities in comparative genome mapping at the functional level between sorghum and other grass species. EST-SSR markers have greater transferability than genomic SSR markers. These markers were successfully amplified in sugarcane (at Sugarcane Research Institute, Mauritius; data not shown). This will greatly enhance the development of comparative genome maps between the sorghum and sugarcane. Sorghum can act as a model to guide mapping and positional cloning in sugarcane, because of its close relationship and high degree of co-linearity with sugarcane (Paterson et al. 1995; Dillon et al. 2007).

5.5 Physical mapping of sorghum

A physical map for sorghum was constructed using sequence information derived from the elite inbred line, BTx623 (Figure 13 and 14; Table 6). Whole genome sequencing profiles provided this new opportunity for the broader application of reverse genetics, expression profiling and genetic mapping. Primer pair sequences were used to identify SSR marker positions on the physical map. Almost all, except a few genomic SSRs, were mapped onto the expected linkage group. The markers that could not be positioned on the physical map had forward and reverse primer hits on different chromosomes or in super clusters that have not yet been assigned to a chromosome. This indicating gap in the genome sequence assembly or they were genetically mapping in the regions with poor marker densities and/or near to centromeric regions. If super cluster can be defined clearly into different chromosomal groups, and then there is good chance of aligning these remaining markers onto the sequence-based sorghum physical map. Complementing the improvement of genetic maps, technological advancements are accelerating the incorporation of EST-hybridization landmarks into physical maps.

In the case of EST-SSRs on sorghum linkage map developed in the present study, 74 markers were mapped onto the sorghum physical map in the expected linkage groups (Table 9). Among the remaining eight markers, five markers, *Xisep*0746 on SBI-04, *Xisep*0713 on SBI-05, *Xisep*1250 on SBI-07, *Xisep*1008 on SBI-09 and *Xisep*1038 on SBI-10 were aligned on to SBI-02, SBI-02, SBI-05, SBI-01 and SBI-01 linkage groups, respectively, in the physical map (Figure 13 and 14). This discrepancy can be attributed to the presence of duplicate loci in the respective linkage groups. The remaining three

markers, *Xisep*1202 on SBI-05, *Xisep*0120 on SBI-05 and *Xisep*0716 on SBI-07 could not be aligned onto sorghum physical map (Figure 13 and 14). This is because of presence of forward and/or reverse primer sequence on different chromosome or present in super clusters that need further investigation to position them in to the sorghum chromosomes.

Most of the markers were clustered towards the telomeric regions of the chromosome arms and very few markers were located in the central portion or centromeric regions of the chromosomes. Gene rich regions are present at the telomeric ends of each sorghum chromosome, except SBI-06 (Paterson et al. 2009). Short arm of SBI-06 is completely devoid of genes/presence of very few genes as per sorghum genome sequence (Figure 25). In agreement with this, very few markers (genic and genomic) were located on physical map of SBI-06. SBI-03, SBI-06 and SBI-10 are devoid of even single markers located near the centromeric regions of these chromosomes.

5.5.1 Integration of physical and linkage maps of sorghum

In principle, genetic mapping serves to subdivide and order the genome by crossing over and recombination of its parts, whereas physical mapping allows ordering of genomic fragments (parts) based on overlapping pattern among them (Coe et al. 2002). Construction of highly saturated genetic linkage maps and their integration with physical, chromosomal and molecular cytogenomic maps have paved the way for the comparative genome mapping among cereals and map-based cloning for economically important genes (Zhi-Ben et al. 2008). High resolution maps will provide information to link sorghum genetic diversity and QTLs to the sorghum physical map. This will help in acceleration of gene discovery and analysis.

Technological advancement in plant genomics in association with sequencing efforts has allowed the research community to survey the genome-wide distribution and expression of genes. Identification of gene locations and gene sequences is a crucial step in crop improvement programs. Integration of physical and linkage maps in sorghum is a great land mark in the plant genomics era especially for the crops employing C_4 photosynthesis as it aids in identification of origins of gene sequences underlying in these important biochemical pathways. The integration of genetic, molecular and physical maps will ultimately permit more routine cloning while building scaffolds for sequencing projects

of the future. In fact, in sorghum genome sequence assembly, scaffolds were assembled (Paterson et al. 2009) by taking the advantage of high-density sorghum linkage maps developed using RFLP probes (Bowers et al. 2003).

The sorghum physical map developed in this study is based on the presence of SSR loci in genome sequence information derived from an elite inbred line, BTx623. These SSR loci include both genomic and genic (EST) SSRs. This localization of EST-SSR primers on physical map in combination with microarray-based gene expression analysis will provide efficient tools to study the sorghum genes and the basis of adaptation to adverse environmental conditions (Menz et al. 2002).

Construction of high-density maps is a crucial step in identifying gene locations and their isolation by map-based cloning. As genetic maps and the loci that constitute those maps become integrated into physical maps, recombination frequencies will be converted into physical map distances (cM to kb) at which point the resolution of the genetic map will no longer limit the progress of genomic applications. The linkage map positions for these newly added EST-SSR markers on sorghum linkage map are integrated with physical map positions (Figure 14). In the physical map each linkage group was divided into a number of bins based on their size, each bin is equivalent to 1 Mb. The linkage map developed in this study was 2966.4 cM. After integration of linkage map with physical map, 1 cM = 223 kb or 1 Mb = 4.467 cM. However, average bin (1 Mb) size on different chromosomes ranged from 3.391 cM (SBI-03) to 6.562 cM (SBI-01). If recombination frequency was considered, each cM ranged from 152 kb (SBI-01) to 295 kb (SBI-03) on the physical map (Table 7). The differences in average recombination length of these 1 Mb bins on SBI-01 and SBI-03 can be attributed to higher and lower recombination rates (and therefore map distances), respectively, despite having similar physical lengths.

Klein et al. (2000) considered that the presence of a link/marker/allele at every 3 Mb constitutes a high density map. In this 664 Mb physical map (excluding super clusters), primer pairs for a total of 1329 candidate SSR markers were positioned on the 10 sorghum linkage groups, with an average of 2 markers per Mb. In this map, 3 Mb equals approximately 13.4 cM. So, presence of a marker for every 13 cM distance interval (inter markers distance) can be considered as highly saturated map according to Klein et al.

(2000). But in this physical map, for every 223 kb or 4.467 cM, at least a single marker was located.

A total of 664 bins were present on physical map. But, only 358 bins (54%) were covered by the markers and these were more strongly saturated towards at the telomeric ends of each chromosome. The remaining 306 bins (46%) did not have even a single marker and these were almost exclusively located in the middle of the chromosomes. If we take into consideration of only bins covered, each bin is equal to 8.28 cM. So, we can consider this as a highly saturated integrated linkage and physical map of sorghum. Further, on the physical map, the covered bins are not only having a single marker, but ranged from one to fourteen markers per bin. So there is a great chance of further reducing the map distance between adjacent markers, if we consider only covered bins to calculate the distance between adjacent markers. Nearly all the markers for which information was available, including SSRs and EST-SSR markers, exhibited strong co-linearity between the physical and linkage maps, but there were some exceptions in their order at some genomic locations (for example, SBI-03). This may likely due to tight linkage between such markers with small number of recombination events between them in the relatively small populations upon which the linkage maps are based - resulting in imprecise positioning of very closely linked markers which are resolved in the physical map (assuming the genome sequence alignment is fully accurate).

Sorghum genome sequence consortium generated the heat maps for each chromosome of sorghum and suggesting that gene rich regions were present at the telomeric ends of the sorghum chromosomes (Paterson et al. 2009). Markers on the physical map are located mostly at the telomeric ends, especially EST-SSR markers. This greatly supports the heat maps generated in sorghum genome sequencing project which suggests that most of the genes are located towards the telomeric ends. Very few markers were present on the short arm of SBI-06. This shows that very few genes may be located at this end of SBI-06 (Figure 13 and 14) and this is in agreement with the heat map (Figure 25).

Genes which are well characterized in sorghum, based on their sequence, were aligned to the left of each linkage group in the physical map (Figure 13 and 14). If these genes have a significant role in plant breeding, agronomically important, superior alleles can be transferred to locally adapted improved varieties from source lines using the markers

present in the same or adjacent bins through MAS. PhyC and PhyA are present on SBI-01 at 7 and 9 Mb bins, respectively. These genes were flanked by SSR markers like *Xtxp*325, *Xtxp*350 and *Xtxp*208 as reported by Menz et al. (2002). However, the present physical map suggests that desirable alleles of these genes can be effectively moved to other backgrounds using a larger number of SSR markers through MAS. Agronomically important genes and their location on different chromosomes are listed in Table 8.

By virtue of the marker positions in the QTL regions in sorghum, these markers can help in identifying the genomic regions responsible for control of a particular trait. Thereby, genomic regions responsible for control of the particular trait can be isolated through map-based cloning. This helps in better understanding the complex traits in sorghum. For example, the Sucrose synthase gene (Sivasudha and Kumar, 2007) is located at the distal end of the short arm of chromosome SBI-10. Previously only *Xcup*49, *Xcup*50, *Xtxp*20 and *Xcup*67 were available to serve as flanking SSR markers for this gene. Among these, all *Xcup* series markers were developed from RFLP probes (Schloss et al. 2002) and these were also derived from expressed regions. The gap between flanking markers (*Xcup*49 and *Xtxp*20) in this region spans approximately 130 cM. In the present study, the gap between *Xcup*49 and *Xtxp*20 was filled with three additional EST-SSR markers, namely *Xisep*0604, *Xisep*0607 and *Xisep*0608. As these markers were derived from expressed regions, they could serve as good flanking markers to transfer favorable alleles for the sucrose synthase gene to locally adapted improved varieties. These marker positions were confirmed by physical mapping in sorghum. With the increased interest in ethanol production from sorghum, transfer of most favorable alleles for this gene has potential to make a significant contribution to improving production of bio-ethanol.

In summary, collinearity and synteny of the markers were widely observed between the linkage and physical maps of sorghum. Developing the physical map for sorghum helps the research community by providing opportunity for a prior test to have a preliminary idea on where new candidate markers have the greatest chance to map on sorghum genome, thus identifying those with the best potential instead of needing to test all of them empirically in wet lab conditions. The integration of the linkage map with physical map provides opportunity for the breeders to choose large numbers of polymorphic

markers between any combinations of parents for a particular targeted region in the linkage map where the genes or QTLs of interest reside.

5.6 Diversity among Reference Set of Sorghum

The important objective of crop improvement programs is the identification of variability among the available genotypes. Diversity analysis at the molecular level using PCR based markers is the cheapest and most rapid method of characterizing the relationships among different genotypes. User friendly nature of SSR markers have been successfully exploited in crop species for better understanding genetic diversity, the domestication process, and geographic divergence and distribution.

Initially, RFLP markers were used to study the genetic diversity (Aldrich and Doebley 1992, Tao et al. 1993). Because of very few numbers of genotypes were included in the earlier studies, underlying population structure could not be studied well using RFLP and RAPD markers. The population structure mainly depends on the type and number of markers used and representation of the samples analyzed. In the present study, samples (the reference set of 384 genotypes) were selected from an earlier global genetic diversity analysis of 3367 accessions using 41 SSR markers under the Generation Challenge Program (GCP). The reference set of sorghum (Appendix 1) selected for diversity analysis represent a significant portion of genetic variation with all 5 basic races, their intermediate races and wild genotypes. Markers selected for screening (Table 10) the reference set in this study provide complete genome coverage across all 10 linkage groups of sorghum.

All EST-SSR markers (45) used in this study were polymorphic (100%) in sorghum reference set. Gupta et al. (2003) identified only 55% of 20 EST-SSR markers used were polymorphic among 52 wheat accessions. Eujayl et al. (2002) reported still lower level of polymorphism when 42 EST-SSR markers were used to screen 64 durum wheat germplasm lines. This clearly indicates that the percentage of polymorphism depends on the number and nature of the material under analysis. In the present study, data analysis was carried out using only 40 EST-SSR markers. Five markers (*Xisep*0209, *Xisep*0621, *Xisep*0728, *Xisep*1128 and *Xisep*1150) were removed from the analysis because of many missing data. The AlleloBin program, used for converting raw allele calls to perfect allele

calls, uses a step-wise mutation model (SMM; Kimura and Ohta 1978) where the repeat unit is taken into consideration and assumes that alleles mutate back and forth with uniform repeat length (Table 11). In the present study, 33 EST-SSR markers followed the SSM model where as 7 markers (*Xisep*0203, *Xisep*0523, *Xisep*0617, *Xisep*0630, *Xisep*1038, *Xisep*1140 and *Xisep*1202) did not follow the SSM model. However, these 7 markers fit well in an alternative model, the infinite alleles model (IAM; Ohta and Kimura 1973), which assumes that each mutation (insertion and deletion) creates a new allele. Similar patterns of allele scoring have been observed with genomic SSR markers (Folkertsma et al. 2005).

In total, 40 EST-SSR markers screened against 384 accessions in the sorghum reference set produced 362 alleles with an average of 9 alleles per marker (Table 12). This is the maximum number of alleles per marker reported using EST-SSR markers in any cereal to date. In case of tall fescue grass, an average of 2.78 alleles/marker were reported (Saha et al. 2004), while 1.8 alleles/marker in bread wheat (Gupta et al. 2003) with 20 EST-SSRs, 4.5 alleles/markers in durum wheat with 42 EST-SSRs (Eujayl et al. 2002), and 3 alleles/markers in 54 barley accessions using 38 EST-SSR markers (Thiel et al. 2003). EST-SSR markers detected an average of 9 alleles per markers in this study (Table 12), is on par with 8.8 alleles per markers detected with genomic SSRs in sorghum (Barnaud et al. 2007) and also detected 8.7 alleles/marker in a set of aluminum tolerant sorghum accessions with 15 SSR markers (Caniato et al. 2007), but less than 10.43 alleles/marker detected in Niger-wide sorghum accessions using 28 SSRs (Deu et al. 2008). This illustrates that the EST-SSR markers included in this study have discriminating power similar to that of genomic SSRs.

The PIC values of markers provide an estimate of their discriminating power in a set of accessions by taking not only the number of alleles but also the relative frequencies of each allele (Smith et al. 2000). Average PIC value of EST-SSR markers (0.52) was a bit higher in this references set of sorghum (Table 12) when compared with previous studies using EST-SSR markers for genetic diversity analysis in other crops, for example 0.443 in bread wheat (Gupta et al. 2003) and 0.45 in barley (Thiel et al. 2003). However, the average PIC value was lower compared to the PIC values of genomic SSR markers in sorghum [0.62 in both studies by Agrama and Tuinsta (2003) and Caniato et al. (2007)].

However, this is higher than the PIC value reported by Folkertsma et al. (2005) using 100 guinea race accessions and 21 SSR markers. High PIC value and large number of alleles per marker can also be attributed to the nature of the materials studied. As the reference set was derived from global germplasm collection representing almost the complete diversity available in sorghum, it is expected to produce the large numbers of alleles. SSR markers containing di-nucleotide repeats produced more alleles and hence, greater PIC values. These results were in harmony with previous studies reported by Smith et al. (2000), Agrama and Tuinstra (2003), Casa et al. (2005) and Deu et al. (2008).

5.6.1 Allelic distribution

Maximum number of alleles were reported in caudatum race (57.2%) followed by bicolor race (56.6%). Even though the maximum numbers of alleles were contributed by caudatum race (79 accessions), one marker (*Xisep*0120) was found monomorphic where as five markers were found monomorphic across the kafir accessions (23 accessions). The lowest numbers of alleles (25%) were reported in Gma accessions (11 only) followed by kafir race (30%). This is the major reason for their separate clusters as compared to other races.

Genotypes from African continent reported the maximum number of alleles (89.5%) suggesting that maximum diversity in sorghum was found in that continent and also greatly supporting the idea that sorghum originated from Africa. In Africa, Eastern Africa contributed the maximum number of alleles (68%) supporting greatly the idea that sorghum originated from East Africa. No marker was found monomorphic in accessions originating from Africa where as three markers were found monomorphic among accessions from East Asia and Middle East. In both these cases, *Xisep*1231 was found monomorphic.

Wild species are the most diverged (capturing 71% of total alleles) in the present study as reported by Deu et al. (1994, 2006). Cultivated and wild genotypes shared 213 alleles. Among 44 unique alleles reported from wild accessions, 13 alleles were contributed only by one wild accession, IS 18868 which belongs to *S. bicolor* subspecies *verticilliflorum*. This accession has passport data indicating its origin as USA, but this likely means that it entered the global germplasm collection via USA rather than as a direct field collection from Africa.

5.6.2 Structure of genetic diversity

Based on floral and grain morphology, sorghum cultivars were grouped into five basic races and ten intermediate races. Reference set accessions were grouped primarily according to geographic origin and further characterized according to race (Figures 20 to 23) in agreement with previous studies using RFLP markers (Deu et al. 1994, 2006), SSRs and AFLPs (Menz et al. 2004), SSRs (Deu et al. 2008) and recently by using DArT markers (Mace et al. 2008). Cluster analysis of accessions based on EST-SSR allelic variation divided them in to eight groups (Figures 20 to 23). On a finer discrimination level, a racial pattern was found in the 8 clusters on the dendogram. Racial discrimination by markers was first observed by Deu et al. (1994) in sorghum.

Average gene diversity (H_e, also known as expected heterozygosity) among the reference set was 0.57 (Table 12). As expected with EST-SSR markers, this is lower as compared to previously published results with a small set of highly polymorphic SSRs in materials from Morocco (H_e = 0.84) by Dje et al. (1999), in Eritrea (H_e = 0.78) by Ghebru et al. (2002), in Niger accessions (H_e = 0.61) by Deu et al. (2008) and in South Africa (H_e = 0.60) by Uptmoor et al. (2003). Observed heterozygosity reported in the present study is very less (0.038) as most of the markers detect single loci, and sorghum being largely self-pollinated. The large differences between the expected and observed heterozygosity indicates that the materials are not in Hardy-Weinberg equilibrium. There is a considerable population structure and the materials studied are highly inbreds due to enforced selfing for several generations.

In the previous studies, accessions belong to bicolor race were found scattered across all clusters, and considered as the most heterogeneous and most ancient race with wider geographical distribution and diverse uses (forage, broom-corn and sweet stocks) (Doggett 1988; Dje et al. 2000; Deu et al. 2006; Mace et al. 2008). In the present study, EST-SSR markers differentiated the bicolor race into two major groups (Figures 20 to 23). Some of the African bicolor accessions were grouped in cluster III along with homogeneous kafir race derived from Southern Africa. Most of the remaining bicolors from different places, including Africa and America origin, were grouped in cluster VI. Cluster VI mostly consists of sorghum accessions derived from Asia (predominantly East

Asia and the Middle East) and North America along with Ethiopia-originated bicolors. A couple of bicolor accessions derived from India were grouped in cluster V which mainly consisted of Indian durras. Further, four accessions belong to bicolor race (IS 14206, IS 14449, IS 24503 and IS 27855) grouped with wild accessions confirming their resemblance to spontaneous weedy sorghum and also the most primitive grain sorghum (Casa et al. 2005). This suggests that EST-SSR markers have good discriminating power. Similarly, Mace et al. (2008) also found the grouping of one bicolor race with wild genotypes.

Caudatum is the race of greatest interest to breeders because of providing the genes responsible for higher grain yields with excellent seed quality (Mace et al. 2008) where as guinea race is of interest to breeders due to their great genetic diversity (Deu et al. 1994, 1995). Caudatum race was grouped mainly into three different clusters (Figures 20 to 23), which were largely geographic-origin-specific. Most of the African caudatums were grouped in cluster I whereas another group of accessions from East Africa (AfricaE) were grouped in cluster IV. East Asia and North America origin caudatums were grouped in cluster VI. One of the caudatum race accessions (IS 2730) from East Africa (AfricaE) was grouped with wild accessions in cluster VIII which warrants verification.

Durra race accessions were grouped primarily in cluster V (Figures 20 to 23), largely comprised of durra genotypes originating from India and Africa. Most of the durras are considered as drought tolerant genotypes based on their adaptation to very high temperatures and/or receding moisture conditions (Deu et al. 2006). Intermediate races with durra, durra-bicolor (DB) and durra-caudatum (DC) were distributed throughout the clusters except DC genotypes from the Middle East that were grouped in cluster VI closely with some durra race accessions (Figures 20 to 23). This is contrary to the conclusions drawn by Deu et al. (2008) where all intermediate races with durra were clustered in the same cluster with durra.

Kafir race is grouped in only cluster III. Interestingly an intermediate race with caudatum (KC) also grouped in the same cluster. This suggests that kafir race is the least diverged and expected to be the most homogeneous group among all the races. The same conclusions were drawn in previous studies (Deu et al. 1994, 1995, 2006; Cui et al. 1995;

Menkir et al. 1997; Dje et al. 2000). Kafir accessions were primarily derived from Southern Africa (AfricaS) (Figures 20 to 23). The homogenous nature of the kafir race was supported by the presence of lower number of alleles (30% of the total) and five monomorphic markers. These results are in agreement with the recent origin and restricted geographic distribution of the kafir race (Doggett et al. 1988, Deu et al. 2006).

Guinea race exhibited a moderate level of genetic diversity and grouped into three major clusters according to their origin and distribution (Figures 20 to 23). Similar grouping patterns were observed with isozymes, RFLP probes and SSRs (Deu et al. 1994, 2006, 2008; Cui et al. 1995; Folkertsma et al. 2005, Barnaud et al. 2007). Guinea accessions derived mainly from Western Africa (AfricaW) were grouped into two separate clusters, namely margaritiferums and non-margaritiferums. Guinea guineense (Ggu) and guinea gambicia (Gga) were grouped with some other guinea lines in cluster II, where as Gma accessions clustered in a small group (cluster VII) adjacent to most wild accessions in cluster VIII, which were mainly derived from Western Africa (AfricaW) (Figure 21 and 23). Hence, Gma may be considered as a recently evolved 'primitive form' of the guinea race. Guinea accessions originated from India were grouped separately in cluster IV and grouped closely with Southern African guineas (including guinea roxburghii, Gro) suggesting a recent introduction of Asian forms from Southern Africa. Deu et al. (2006) also found a separated cluster for guineas derived from the Southern Africa and the Asia.
The singularity of Gma race accessions is in harmony with previous studies (Deu et al. 1994, 1995, 2006, 2008; Cui et al. 1995, de Oliveira et al. 1996, Folkertsma et al. 2005). Singular nature of Gma was due to presence of only 25% of alleles found across the sorghum reference set and a maximum of nine monomorphic markers. All guinea margaritiferum accessions in cluter VII were derived from Western Africa (AfricaW). The distinct nature of Gma from other guinea race subgroups is remarkable, since they are infertile and cultivated in sympatry in the same season by the same farmers (Deu at al. 2006). The Gma sub-group was distinct from other guinea races and forming a close knit group with wild genotypes than other cultivated sorghums suggesting that the Gma group represents an independent domestication event. An accession of Southern Africa origin (IS 19455) was not grouped with Western African margaritiferums but it was grouped with other guineas derived from Southern Africa. Grouping of this accession is in agreement with previously reported by Deu et al. (1995, 2006). These results suggest

that Southern African margaritiferums shared a common ancestor with Western African margaritiferums and the change in its genetic background as compared to western African guinea margaritiferums is due to their isolation and selection pressure (Deu et al. 2006).

Surprisingly some of the landraces were grouped with wild accessions (Figures 20 to 24). This suggests a possible gene flow between landraces and wild accessions. Mace et al. (2008) also found the grouping of a bicolor race genotype with wild accessions. Ten accessions (nine landraces: four bicolor, one durra, two caudatum-bicolor, one durra-bicolor and one guinea-caudatum; and one breeding lines, caudatum race) were grouped with wild accessions in cluster VIII. Wild genotypes contributing 44 rare alleles and among these, IS 18868 (*S. bicolor* subspecies *verticilliflorum*), a wild accession contributed 13 alleles.

5.6.3 Gene-flow in sorghum

Exchange of genes is one of the major factors in evolution of domesticated plant species (Haraln 1992). Gene flow from cultivated to wild, weedy and feral relatives disturbs the size and dynamics of wild and weedy populations, resulting in the disturbance of natural gene pools and endangering wild relatives. This in turn leads to loss of natural genetic diversity (Akimoto et al. 1999; Snow et al. 2003). Molecular analysis clearly identified the existence of gene flow in crop plants (Mariac et al. 2006). In the present study, ten accessions (nine of the landraces and one breeding material, IS 2730, caudatum race) were grouped with wild genotypes and most of them were derived from African countries, except IS 14206 which was derived from Australia (Figure 21, 23 and 24). This can be attributed to gene flow between landraces or cultivars and wild genotypes. This may be because of wild accessions growing around the cultivated sorghum in Africa as observed by Tesso et al. (2008). Existing cultural practices in Africa is the major source of gene flow from cultivated lines to wild genotypes and vice-versa. Wild accession are found in Africa in crop margins, barren lands, hill bottom areas, and in fields where crops abandoned due to severe drought, pest and weeds infestation or extreme nutrient deficiency (Tesso et al. 2008). This suggests that wild accessions have desirable alleles to resist these factors. This could make the sorghum breeders to pay more attention to find the alleles from wild genotypes to transfer into locally adapted improved varieties in their crop improvement programs.

5.6.4 Association mapping – a future thrust

Association mapping or linkage disequilibrium (LD) mapping uses the nonrandom association of loci in naturally occurring diverse germplasm lines and hence LD mapping is considered as the powerful high-resolution mapping tool for complex quantitative traits as compared to conventional method of linkage mapping (reviewed in Abdurakhmonov and Abdukarimov 2008; Zhu et al. 2008). Association mapping requires phenotyping data generated over the multiple environments. In sorghum, association mapping was initiated recently for eight traits using 377 accessions and 47 markers (Casa et al. 2008). In the present study, highly diverged germplasm lines (reference set) representing the global sorghum composite collection, were screened with 40 functional markers. Thus, the available genotyping data with EST-SSR and genomic SSR markers combined with proper phenotyping of this reference set could permit association mapping in sorghum. Hence, the reference set germplasm used in this study is an useful genetic resource for sorghum breeding community for trait-specific allele mining and association mapping in sorghum provided that the phonological variation (plant height and flowering time) present is not so great that it interferes with phenotyping for target traits of interest.

6. SUMMARY

Sorghum [*Sorghum bicolor* (L.) Moench, $2n = 2x = 20$] is the fifth most important cereal crop. It is well adapted to harsh conditions like high temperatures and low rain fall. Now sorghum is the emerging model crop species for tropical grasses of the Andropogoneae family which employ C_4 photosynthesis. Thus, sorghum is considered as a logical complement to rice, the first monocot for which complete genome sequencing is available.

Initially, RFLP markers brought about revolution in use of markers in applied crop breeding programs, including diversity assessment, molecular mapping, MAS and also in comparative genome mapping. Later, other marker systems like RAPDs, AFLPs, SSRs, ISSRs, CAPs and SNPs came into existence. Among these, SSR markers, being co-dominant in nature, have their own significance and have proven themselves as a better markers system in applied genomics. The major limitation of SSRs is its high cost of development. On-going sequencing projects and other EST projects are generating large amounts of DNA sequence data that is available for public use. This includes both genomic and genic (EST) sequences that can be exploited to develop SSRs, SNPs or other types of markers. The existence of any marker type in the publicly available sequence is a low-cost product of the databases in which this sequence information is stored. If, SSRs are present in genic sequences, they are referred as 'EST-SSRs'.

EST-SSRs are a valuable source of functional markers and once EST-SSRs are mapped, they will be associated with the genes carrying them. Thus generation of EST derived SSR markers has become a cheaper alternative to conventional SSR development. Application of markers like EST-SSRs in sorghum genome analysis holds great promise for producing lasting insight into processes by which novel genotypes are generated. By taking advantage of this marker system, the present study was aimed at the development of EST-SSR markers in sorghum and their application in crop breeding programs.

Sorghum EST sequences from the J. Craig Research Institute [formerly, The Institute for Genome Research (TIGR)] were downloaded (187,282 sequences) and searched for the

presence of microsatellite markers. In total, 39,106 sorghum EST sequences were found with microsatellites. Over time, SSR density in sorghum EST databases has increased from 1 SSR/5.5 kb (in 2002) to 1 SSS/1.79 kb (in 2006). AG among di-nucleotide and CCG among tri-nucleotide repeats were found abundant in the sorghum as reported in previous studies. The Cap3 program was used to identify the redundant sequences among the ESTs containing SSRs. At the end, only 5.3% sequences (10,044 ESTs) were found non-redundant with microsatellites.

Candidate non-redundant sorghum EST sequences containing SSRs were selected for designing primer pairs based on their sequence similarity with rice using the BLAST tool available in the GRAMENE database. The sorghum EST sequences providing coverage across the complete genome in rice (and thereby corresponding to sorghum genome locations) were selected for primer design. A total of 600 primer pair were designed, and named as 'ICRISAT Sorghum EST Primers' (ISEP).

About 64.3% of the primer pairs produced good amplification profiles for a four-entry panel of sorghum genotypes representing two pairs of mapping population parents. About 23% of markers were found polymorphic in parents of shootfly resistance mapping population (BTx623 × IS18551). These polymorphic markers were screened against the shootfly RIL mapping population. A total of 82 EST-SSR markers were mapped onto the previously existing skeleton map of the sorghum using shootfly resistance mapping population. After adding these 82 new EST-SSR markers, the map distance was extended to 2966.4 cM. This map distance was ~250 cM larger than the previously reported longest linkage map in sorghum. This increase in map distance in sorghum was confirmed by comparing linkage map positions with locations on the sequence based sorghum physical map. The EST-SSR markers greatly helped in filling the gaps in marker-rare regions in the existing skeleton map of sorghum developed using sorghum shootfly resistance mapping population. Our initial attempt to develop EST-SSRs based on sorghum-rice synteny was successful for filling important gaps in the existing sorghum linkage map, for which we had been unable to detect mappable polymorphism with previously available genomic SSRs. This will help the breeder community to effectively map and transfer the QTL regions from shootfly resistant line IS 18551 to agronomically elite susceptible lines such as BTx623 and 296 B.

Grass species express extensive synteny and collinearity of markers and genes among species which differ in genome size. On successful completion of EST-SSR mapping in sorghum, their sorghum map locations were compared with their expected position on the rice genome sequence. Rice and sorghum comparative maps which were already available at the JCVI database were downloaded and compared to the maps developed in the present study. This comparison between the rice and sorghum may not give a perfect idea regarding the comparative maps between rice and sorghum. However, they suggest that this sequence similarity approach can help in targeted development of markers in closely related crop species using the comparative genome mapping approach employed here.

Upon completion of sorghum genome sequencing from inbred line BTx623, the complete genome sequence was made available for public use during early 2008. This enabled us to develop an *in-silico* sequence-based physical map for sorghum using the available SSR markers (published and unpublished). These markers were observed to be greatly concentrated towards the telomeric end of each chromosome arm. This confirms that all gene-rich regions of the genome were concentrated at the telomeric ends of the chromosomes as concluded by the sorghum genome sequencing community.

In-silico mapping of the ESTs from which these new SSRs were developed, and comparing marker order and position with results from conventional linkage analysis, confirms the quality of the sorghum genome sequence alignment. Many marker positions from the *in-silico* study exactly match results from conventional linkage analysis, both in terms of chromosome arm location and order, as one would predict from the rice genome given that these markers were developed based on sequence similarities between rice and sorghum. The order of these EST-SSRs from the *in-silico* mapping agrees with the linkage map for all ten sorghum linkage groups provided that the latter are oriented with the short chromosome arm at the top, as per the suggestion of Kim et al. (2005b). The markers which had conflicting linkage and physical mapping positions can be attributed to the presence of duplicate loci or the presence of best hit for forward and reverse primer sequences on the different chromosomes. Upon integration of sorghum physical and linkage maps, it was found that 1 cM was equal to 223 kb (in other words, 1Mb = 4.47

cM). This indicates that if recombination rates are similar across the genome (which is not likely), then if 4–5 markers are present in each bin (1 Mb) the map can be considered as saturated, which is essential for map-based cloning. Integration of physical maps and genetic maps provides the bridge to isolate the genes through positional or map-based cloning. Continuity, from phenotype to genotype, can be achieved better through the integration of physical and genetic maps.

Integrated sorghum genetic and physical maps can provide a valuable new tool for structural, functional, and comparative genome mapping investigations of many aspects of grass biology. Integration of physical and linkage maps in sorghum is a great land mark in plant genomics era especially for the crops employing the C_4 photosynthesis as it will aide in identifying the gene sequences underlying all important biochemical pathways. The integration of linkage and physical maps significantly enhances the breeders' ability to use large numbers of polymorphic markers for any combination of crossed parents in sorghum from a targeted region in the linkage map where the genes or QTLs reside.

A set of mapped markers were selected (45 EST-SSRs) based on their linkage map positions to screen the reference set of sorghum. Reference set of sorghum consists of 384 diverse accessions representing five basic races, their intermediate races and wild genotypes. These accessions were clustered primarily by race within the geographic origin in harmony with previous studies. Kafir race was the most homogenous group among all the races. In previous studies, bicolor race was identified as the most heterogeneous. But in the present study, ESR-SSR markers clearly identified two major distinct groups within bicolor. The margaritiferum group with in the guinea race (Gma) formed as separate cluster from the guinea race accessions and closely associated with wild genotypes suggesting independent domestication of this Gma group. The markers used in this study for diversity have discriminating power similar to that of SSR markers previously used to constitute the sorghum reference set. A few landraces from bicolor race and several intermediate race accessions were clustered with wild accessions suggesting gene flow between landraces and wild sorghums because of their co-cultivation in Africa. Proper phenotyping of the reference set along with genotyping data using these functional markers is expected to facilitate the association mapping in

sorghum provided that the phonological variation (plant height and flowering time) present is not so great that it interferes with phenotyping for target traits of interest. In the present scenario, association mapping is gaining more importance as compared to conventional linkage and QTL mapping because it consumes less time and is inexpensive; however, conventional QTL mapping and LD mapping approaches are complementary, and exploitation of both in applied plant breeding demand high density genetic maps well saturated with polymorphic markers.

Thus, the availability of genome sequences and other marker information can help in selecting or developing markers for targeted mapping not only in a given crop but also in related crops for which sufficient genomic tools are not yet available in the public domain. Species like sorghum and its larger-genome close relatives are characterized with huge amount of non-transcribed junk DNA of unknown function. Even, un-transcribed regions of genome can be covered by developing the genomic SSRs. Thereby, EST-SSRs in combinations with genomic SSRs provide opportunities for more complete genome scans and increase the power of transferring the desired portion of genome through marker-assisted breeding programs. Through this study, we have demonstrated the value of a comparative genomics approach for targeted development of PCR-compatible molecular markers for practical applications in crop genetics and breeding. EST-SSR markers show higher rates of cross-species transferability in comparison with genomic SSRs, hence these markers may also be used in other related grass species for which insufficient numbers of PCR-compatible markers are available. This will enable us to study the functional evolutionary process and to develop comparative genome maps.

7. LITERATURE CITED

Abdurakhmonov IY, Abdukarimov A (2008) Application of association mapping to understand the genetic diversity of plant germplasm resources. Int J Plant Genomics. doi:10.1155/2008/574927

Adam D (2000) Now for the hard ones. Nature 408:792-793

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analysis of coffee and related species. Theor Appl Genet 114:359-372

Agrama HA, Tuinstra MR (2003) Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. Afr J Biotechnol 10:334–340

Agrama HA, Widle GE, Reese JC, Campbell LR, Tuinstra MR (2002) Genetic mapping of QTLs associated with green bug resistance and tolerance in *Sorghum bicolor*. Theor Appl Genet 104:1373-1378

Ahn S, Anderson JA, Sorrells ME, Tanksley SD (1993) Homoeologous relationships of rice, wheat and maize chromosomes. Mol Gen Genet 241:483-490

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113:1409-1420

Akimoto M, Shimamoto Y, Morishima H (1999) The extinction of genetic resources of Asian wild rice, *Oryza rufi - pogon* Griff: A case study in Thailand. Genet Resour Crop Evol 46:419–425

Aldrich PR, Doebley J (1992) Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. Theor Appl Genet 85:293-302

Ali ML, Rajewski JF, Baenziger PS, Gill KS, Eskridge KM, Dweikat J (2008) Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers. Mol Breeding 21:497-509

Altschul SF, Miller WGW, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. 215:403-10

Balyan HS, Gupta PK, Rustgi S, Bandopadhyay R, Goyal A, Singh R, Kumar A, Kumar N, Sharma S (2005) Development and use of SSRs of bread wheat for genetic and physical mapping and transferability to the species of Triticum-Aegilops complex. Czech J Genet Plant Breed 41:141-145

Barnaud A, Deu M, Garine E, McKey D, Joly H (2007) Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces. Theor Appl Genet 114:237–248

Berube Y, Zhuang J, Rungis D, Ralph S, Bohlmann J, Ritland K (2007) Characterization of EST-SSRs in loblolly pine and spruce. Tree Genet Genomics 3:251-259

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

Binelli G, Ginafranceschi L, Pe ME, Taramino G, Busso C, Stenhouse J, Ottaviano E (1992) Similarity of maize and sorghum genome as revealed by maize RFLP probes. Theor Appl Genet 84:10–16

Boivin K, Deu M, Rami JF, Trouche G, Hamon PL (1999) Towards a saturated sorghum map using RFLP and AFLP markers. Theor Appl Genet 98:320-328

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

Bouck J, Miller W, Gorrell JH, Muzny D, Gibbs RA (1998) Analysis of the quality and utility of random shotgun sequencing at low redundancies. Genome Res 8:1074-1084.

Bowers JE, Abbey C, Anderson S, Chang C, Draye X et al. (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

Brown PJ, Klein PE, Bortiri E, Acharya CB, Rooney WL, Kresovich S (2006) Inheritance of inflorescence architecture in sorghum. Theor Appl Genet 113:931-942

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 93:190-198

Buell CR, Yuan Q, Ouyang S, Liu J, Zhu W et al. (2005) Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. Genome Res 15: 1284-1291

Caniato FF, Guimarães CT, Schaffert RE, Alves VM, Kochian LV, Borem A, Klein PE, Magalhaes JV (2007) Genetic diversity for aluminum tolerance in sorghum Theor Appl Genet 114:863-876

Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, Aquadro CF Kresovich S (2005) Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theor Appl Genet 111:23–30

Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S (2008) Community resources and strategies for association mapping in sorghum. Crop Science 48:30-40

Chen S, Lin XH, Xu CG, Zhang QF (2000) Improvement of bacterial blight resistance of 'Minghui 63', an elite restorer line of hybrid rice, by molecular marker-assisted selection. Crop Sci 40:239–244

Cheng CH, Chung MC, Liu SM, Chen SK, Kao FY, Lin SJ, Hsiao SH, Tseng IC, Hsing YI, Wu HP, Chen CS, Shaw JF, Wu J, Matsumoto T, Sasaki T, Chen HH, Chow TY (2005) A fine physical map of rice chromosome 5. Mol Genet Genomics 274:337-345

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) A detailed RFLP map of *Sorghum bicolor* × *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of *Sorghum* chromosomes or chromosomal segments. Theor Appl Genet 87:925-933

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.) Theor Appl Genet 100:713-722

Coe E, Cone K, McMullen M, Chen S-S, Davis G, Gariner J, Liscum E, Polacco M, Paterson AH, Sanchez-Villeda H, Soderlund C, Wing R (2002) Access to maize genome: An integrated physical and genetic map. Plant Physiolo 128:9-12

Cook RJ (1998) Towards a successful multinational crop plant genome initiative. Proc Natl Acad Sci 95:1993-1995

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

Crasta OR, Xu WW, Rosenow DT, Mullet J, Nguyen HT (1999) Mapping of post-flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. Mol Gen Genet 262:579-588

Cui YX, Xu GW, Magill CW, Schertz KF, Hart GE (1995) RFLP-based assay of *Sorghum bicolor* (L.) Moench genetic diversity. Theor Appl Genet 90:787-796

de Oliveira AC, Richter T, Bennetzen JL (1996) Regional and racial specificities in sorghum germplasm assessed with DNA markers. Genome 39:579-587

Dean RE, Dahlberg JA, Hopkins MS, Mitchell SE, Kresovich S (1999) Genetic redundancy and diversity among 'Orange' accessions in the US national sorghum collection as assessed with simple sequence repeat (SSR) markers. Crop Sci 39:1215–1221

Deu M, Gonzalez-de-Leon D, Glaszmann J-C, Degremont I, Chantereau J, Lanaud C, Hamon P (1994) RFLP diversity in cultivated sorghum in relation to racial differentiation. Theor Appl Genet 88:838–844

Deu M, Hamon P, Chantereau J, Dufour P, D'Hont A, Lanaud C (1995) Mitochondrial DNA diversity in wild and cultivated sorghum. Genome 38:635–645

Deu M, Rattunde F, Chantereau J (2006) A global view of genetic diversity in cultivated sorghums using a core collection. Genome 49:168–180

Deu M, Sagnard F, Chantereau J, Calatayud C, Hérault D, Mariac C, Pham J-L, Vigouroux Y, Kapran I, Traore PS, Mamadou A, Gerard B, Ndjeunga J, Bezancon G (2008) Niger-wide assessment of in situ sorghum genetic diversity with microsatellite markers. Theor Appl Genet 116:903-913

Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Mol Biol 35:3-15

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

Dillon SL, Shapter FM, Henry RJ, Cordeiro G, Izquierdo L, Lee S (2007) Domestication to crop improvement: Genetic resources for Sorghum and Saccharum (Andropogoneae). Annal Bot 100:975-989

Dje` Y, Forcioli D, Ater M, Lefebvre C, Vekemans X (1999) Assessing population genetic structure of sorghum landraces from North-western Morocco using allozyme and microsatellite markers. Theor Appl Genet 99:157-163

Dje` Y, Heuertz M, Ater M, Lefe`bvre C, Vekemans X (2004) *In situ* estimation of outcrossing rate in sorghum landraces using microsatellite markers. Euphytica 138:205–212

Dje` Y, Heuertz M, Lefe`bvre C, Vekemans X (2000) Assessment of genetic diversity within and among germplasm accessions in 408 cultivated sorghum using microsatellite markers. Theor Appl Genet 100:918–925

Doebley J (1990) Molecular evidence for gene fl ow among *Zea* species. BioScience 40:443–448

Doganlar S, Frary A, Daunay CM, Lester RN, Tanksley SD (2002) A comparative genetic map of eggplant (Solanum melongena) and its implication to genome evolution in the Solanaceae. Genetics 161:1697-1711

Doggett H (1988) Sorghum, 2nd edn. Longman Scientific & Technical, London

Draye X, Lin Y-R, Qian X-Y, Bowers JE, Burow GB, Morrell PL, Peterson DG, Presting GG, Ren S-x, Wing RA, Paterson AH (2001) Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the sorghum genome as a foundation. Plant Physiol 125:1325-1341

Dufour P, Deu M, Grivet L, D'Hont A, Paulet F, Bouet A, Lanaud C, Glaszmann, JC Hamon P (1997) Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theor Appl Genet 94:409-418

Eagles H, Bariana H, Ogbonnaya F, Rebetzke G, Hollamby G, Henry R, Henschke P,Carter M (2001) Implementation of markers in Australian wheat breeding. Aust J Agric Res 52:1349-1356

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analysis. Heredity 99:125-132

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

Feltus FA, Hart GE, Schertz KF, Casa AM, Kresovich S, Abraham S, Klein PE, Brown PJ, Paterson AH (2006a) Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. Theor Appl Genet 112:1295-1305

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

Folkertsma RT, Frederick H, Rattunde W, Chandra S, Raju GS, Hash CT (2005) The pattern of genetic diversity of Guinea-race Sorghum bicolor (L.) Moench landraces as revealed with SSR markers. Theor Appl Genet 111:399–409

Folkertsma RT, Sajjanar GM, Reddy BVS, Sharma HC, Hash CT (2003) Genetic mapping of QTL associated with sorghum shootfly (*Atherigona soccata*) resistance in sorghum (*Sorghum bicolor*). In: In: XI International Plant and Animal Genome conference, San Diego, CA, USA, W65 (http://www.intl-pag.org/11/abstracts/P5d_P462_XI.html)

Fransz PF, Alonso-Blanco C, Liharska TB, Peeters AJM, Zabel P, de Jong JH (1996) High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA probes. The Plant J 9:421-430

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci 95: 1971-1974

Ghebru B, Schmidt RJ, Bennetzen JL (2002) Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theor Appl Genet 105:229–236

Graham J, Smith K, MacKenzie K, Jorgenson L, Hackett C, Powell W (2004) The construction of genetic linkage map of red raspberry (*Ruus idaeus* subsp. *ideaus*) based on AFLPs, genomic-SSR and EST-SSR markers. Theor Appl Genet 109:740-749

Grant D, Cregan P, Shoemaker RC (2000) Genome organization in dicots: Genome duplication in Arabidopsis and synteny between soybean and Arabidopsis. Proc Natl Acad Sci 97:4168-4173

Grenier C, Bramel-Cox PJ, Noirot M, Prasadao Rao KE, Hamon P (2000a) Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs. non-random sampling procedures. A. Using morphoagronomical and passport data. Theor. Appl. Genet. 101: 190–196.

Grenier C, Deu M, Kresovich S, Bramel-Cox PJ, Hamon P (2000b) Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection

using random vs non-random sampling procedures. B. Using molecular markers. Theor Appl Genet 101:197–202

Guimaraes C, Sills GR, Sobral WS (1997) Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. Proc Natl Acad Sci 94:14261-14266

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 Gen Genomics 270:315-323

Hackauf B, Wehling P (2002) Identific ation of microsatellite polymorphisms in an expressed portion of rye genome. Plant Breed 121:17-25

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 8:299-309

Harlan JR (1992) Crops and man. ASA, Madison, WI

Harlan JR, de Wet JMJ (1972) A simplified classification of cultivated sorghum. Crop Sci 12:172–176

Harris K, Subudhi PK, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein P, Klein R, Mullet (2007) Sorhum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. J Exp Bot 58:327-338

Hart GE, Schertz KF, Peng Y, Syed NY (2002) Genetic mapping of *Sorghum bicolor* (L.) Moench QTLs that control variation in tillering and other morphological characters. Theor Appl Genet 103:1232-1242

Harushima Y, Yano M, Shomura A, Sata M, Shimano T, Kuboki Y, Tamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamato K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148:479-494

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 Crop Res 84:79-88

Haussmann BIG, Hess DE, Omanya GO, Folkertsma RT, Reddy BVS, Kayentao M, Welz HG, Geiger HH (2004) Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. Theor Appl Genet 109:1005-1016

Haussmann BIG, Hess DE, Seetharama N, Welz HG, Geiger HH (2002a) Construction of a combined sorghum linkage map from two recombinant inbred population using AFLP, SSR, RFLP, and RAPD markers, and comparison with other sorghum maps. Theor Appl Genet 105:629-637

Haussmann BIG, Mahalakshmi V, Reddy BVS, Seetharama N, Hash CT, Geiger HH (2002b) QTL mapping of stay-green in two sorghum recombinant inbred populations. Theor Appl Genet 106:133-142

Hicks C, Tuinstra MR, Pedersen JF, Dowell FE, Kofoid KD (2002) Genetic analysis of feed quality and seed weight of sorghum inbred lines and hybrids using analytical methods and NIRS. Euphytica 127:31-40

Hulbert SH, Richter TE, Axtell JD, Bennetzen JL (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proc Natl Acad Sci (USA) 87:4251-4255

International Rice Genome Sequencing Project (2005) The map-based sequencing of the rice genome. Nature 436: 793-800

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:25

Jayashree B, Punna R, Prasad P, Bantte K, Hash CT, Chandra S, Hoisington DA, Varshney RK (2006) A database of simple sequence repeats from cereal and legume expressed sequence tags mined *in silico*: survey and evaluation. In silico Biol 6:0054 (http://www.bioinfo.de/isb/2006/06/0054/)

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

Katsar C, Paterson AH, Teetes GL, Peterson GC (2002) Molecular analysis of sorghum resistance to the greenbug (Homoptera: Aphididae). J Econ Ent 95:448-457

Kebede H, Subudhi PK, Rosenow DT, Nguyen HT (2001) Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). Theor Appl Genet 103:266-276

Kilian A, Kudrna DA, Kleinhofs A, Yano M, Kurata N, Steffenson B, Sasaki T (1995) Rice–barley synteny and its application to saturation mapping of the barley *Rpg1* region. Nucleic Acids Res 23:2729–2733

Kim J-S, Islam-Faridi MN, Klein PE, Stelly DM, Price HJ, Klein RR, Mullet JE (2005a) Comprehensive molecular cytogenetic analysis of sorghum genome architecture: distribution of euchromatin, heterochromatin, genes and recombination in comparison to rice. Genetics 171:1963-1976

Kim J-S, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005b) Chromosome identification and nomenclature of *Sorghum bicolor*. Genetics 169:1169-1173

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci USA 75:2868-2872

Klein PE, Klein RR, Cartinhour SW, Ulanch PE, Dong J, Obert JA, Morshige DT, Schlueter SD, Childs KL, Ale M, Mullet JE (2000) A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map. Genome Res 10:789-807

Klein PE, Klein RR, Vrebalov J, Mullet JE (2003) Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. The Plant J 34:605-621

Klein RR, Rodriguez-Herrera R, Schlueter JA, Klein PE, Yu ZH, Rooney WL (2001) Identification of genomic regions that affect grain-mould incidence and other traits of agronomic importance in sorghum. Theor Appl Genet 102:307-319

Knoll J, Ejeta G (2008) Marker-assisted selection for early-season cold tolerance in sorghum: QTL validation across populations and environments. Theor Appl Genet 116:541-553

Knoll J, Gunaratna N, Ejeta G (2008) QTL analysis of early-season cold tolerance in sorghum. Theor Appl Genet 116:577-587

Kong L, Dong J, Hart GE (2000) Characteristics, linkage-map positions and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple sequence repeats (SSRs). Theor Appl Genet 101:438-448

Kresovich S, Barbazuk B, Bedell JA, Borrell A, Buell AR et al. (2005) Towards sequencing the sorghum genome. A U.S. National Science foundation-sponsored workshop report. Plant Physiol 138:1898-1902

Kumpatla SP, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48:985-998

La Rota M, Kantety RV, Yu J-K, Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics 6:23

Lander E, Green P, Abrahamson J, Barlow A, Daley M, Lincoln S, Newburg L (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181

Leigh F, Lea V, Law J, Wolters P, Powell W, Donini O (2003) Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. Euphytica 133:359–366

Lezar S, Myburg AA, Berger DK, Wingfield MJ, Wingfield BD (2004) Assessment of microarray-based DNA fingerprinting in Eucalyptus grandis. Theor Appl Genet 109:1329-1336

Lijavetsky D, Martinez MC, Carrari F, Hopp HE (2000) QTL analysis and mapping of pre-harvest sprouting resistance in Sorghum. Euphytica 112:125-135

Lin YR, Schertz. KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the Poaceae in reference to an interspecific sorghum population. Genetics 141:391-411

Liu H, Sachidanandam R, Stein L (2001) Comparative genomics between rice and *Arabidopsis* shows scant collinearity in gene order. Genome Res 12:2020-2026

Liu K, Muse SV (2005) PowerMarker: Integrated analysis environment for genetic markers data. Bioinformatics 21:2128-2129

Luro FL, Costantino G, Terol J, Argout X, Allario T, Wincker P, Talon M, Ollitrault P, Morillon R (2008) Transferability of the EST-SSRs developed on Nules Clementine (*Citrus clementina* Hort ex Tan) to other *Citrus* species and their effectiveness for genetic mapping. BMC Genomics 9:287

Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, Wenzl P, Kilian A (2008) DArT markers: diversity analyses and mapping in Sorghum bicolor. BMC Genomics 9:26

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

Magalhaes JV, Garvin DF, Wang Y, Sorrells ME, Klein PE, Schaffert RE, Li L, Kochian LV (2004) Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the Poaceae. Genetics 167:1905-1914

Mahalakshmi V, Ortiz R (2001) Plant genomics and agriculture: From model organisms to crops, the role of data mining for gene discovery.Electronic J Biotechnol 4:169178

Maheswaran K, Subudhi PK, Nandi S, Xu Jc, Parco A, Yang DC and Huang N (1997) Polymorphism, distribution and segregation of AFLP markers in a double haploid rice population. Theor Appl Genet 94: 39-45

Mariac C, Robert T, Allinne C, Remigereau MS, Luxereau A, Tidjani M, Seyni O, Bezancon G, Pham JL, Sarr A (2006) Genetic diversity and gene flow among pearl millet crop/weed complex: a case study. Theor Appl Genet 113:1003–1014

Mayer K, Murphy G, Tarchini R, Wambutt R, Volckaert G, Pohl T, Dusterrhoft A, Stiekema W, Entian KD, Terryn N, Lemcke K, Haase D, Hall CR, Van Dodeweerd AM, Tingey SV, Mewes HW, Bevan MW, Bancri[oft L (2001) Conservation of microstructure between a sequenced region of rice and multiple segments of the genome of *Arabidopsis thaliana*. Genome Res 11:1167-1174

McCouch, SR (2001) Genomics and synteny. Plant Physiol 125:152-155

Meinke DW, Meinke LK, Showalter TC, Schissel AM, Mueller LA, Tzafrir I (2003) A sequence-based map of arabidopsis genes with mutant phenotypes. Plant Physiol 131:409-418

Menkir A, Goldsbrough P, Ejeta G (1997) RAPD based assessment of genetic diversity in cultivated races of sorghum. Crop Sci 37:564-569

Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE (2002) A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP, RFLP and SSR markers. Plant Mol Bio 48:483-499

Menz MA, Klein RR, Unruh NC, Rooney WL, Klein PE, Mullet JE (2004) Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Sci 44:1236-1244

Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH et al. (1998) Detailed alignment of *Saccharum* and sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150: 1663–1682

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Nagamura Y, Tanaka T, Nozawa H, Kaidai H, Kasuga S, Sasaki T (1998) Syntenic regions between rice and sorghum genomes. National Grassland Res Insti Report (Japan) 9:97-103

Nagy ER, Lee TC, Ramakrishna W, Xu Z, Klein PE, SanMiguel P, Cheng CP, Li J, Devos KM, Schertz K, Dunkle L, Bennetzen JL (2007) Fine mapping of the *Pc* locus of *Sorghum bicolor*, a gene controlling the reaction to a fungal pathogen and its host-selective toxin. Theor Appl Genet 114:961-970

Neeraja CN, Maghirang-Rodriguez R, Pamplona A, Heuer S, Colard BCY, Septiningsih EM, Vergara G, Sanchez D, Xu K, Ismail AM, Machill DJ (2007) A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. Theor Appl Genet 115:767-776

Oh B-J, Frederiksen RA, Magill CW (1996) Identification of RFLP markers linked to a gene for downy mildew resistance (*Sdm*) in sorghum. Can J Bot 74:315-317

Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res 22:201–204

Paterson AH (2008) Genomics of sorghum. Intl J Plant Genomics. doi:10.1155/2008/362451

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551-556

Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang C-X, Katsar CS, Lan T-H, Lin Y-R, Ming R, Wright RJ (2000) Comparative genomics of plant chromosomes. The Plant Cell 12:1523-1539

Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714-1718

Peng Y, Schertz KF, Cartinhour S, Hart GE (1999) Comparative genome mapping of *Sorghum bicolor* (L.) Moench using a RFLP map constructed in a population of recombinant inbred lines. Plant Breed 118:225–235

Pereira MG, Lee M (1995) Identification of genomic regions affecting plant height in sorghum and maize. Theor Appl Genet 90:380-388

Perrier X, Flori A, Bonnot F (2003) Methods for data analysis. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic diversity of cultivated tropical plants. Science Publishers, Inc. and CIRAD, Montpellier, pp 31–63

Perumal R, Krishnaramanujam R, Menz MA, Katile S Dahlberg J, Magill CW, Rooney WL (2007) Genetic diversity among sorghum races and working groups based on AFLPs and SSRs. Crop Sci 47:1375-1383

Qureshi SN, Saha S, Kantety RV, Jenkins JN (2004) EST-SSR: A new class of genetic markers in cotton. The J cotton Sci 8:112-123

Ragab RA, Dronavalli S, Maroof MAS, Yu YGL (1994) Construction of a sorghum RFLP linkage map using sorghum and maize DNA probes. Genome 37:590-594

Ritter KB, McIntyre CL, Godwin ID, Jordan DR, Chapman SC (2007) An assessment of the genetic relationship between sweet and grain sorghums, within *Sorghum bicolor* ssp. *bicolor* (L.) Moench, using AFLP markers. Euphytica 157:161-176

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

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

Salse J, Piegu B, Cooke R, Delseny M (2002) Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. Nucleic Acids Res 30:2316-2328

Sanchez AC, Brar DS, Huang N, Li Z, Khush GS (2000) Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. Crop Sci 40:792-797

Sanchez AC, Subudhi PK, Rosenow DT, Nguyen HT (2002) Mapping QTLs associated with drought resistance in sorghum (*Sorghum bicolor* (L.) Moench). Plant Mol Biol 48:713-726

Schloss SJ, Mitchell SE, White GM, Kukatla R, Bowers JE, Paterson AH, Kresovich S (2002) Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L.) Moench. Theor Appl Genet 105:912-920

Semagn K, Bjornstad A, Skinnes H, Maroy AG, Tarkegne Y, William M (2006) Distribution of DArT, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. Genome 49:545-555

Senthilvel S, Jayashree B, Mahalakshmi V, Kumar PS, 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 Biol 8:119

Singh K, Vakil Y, Sidhu JS, Brar DS, Shaliwal HS (2001) Rice cultivar PR 106, pyramided with three bacterial blight resistance genes, using marker assisted selection, reaches plant-breeding trial. In: pp 68a-71, Proc 8th Natl Rice Biotechnol, Network meeting, Oct 21-25, Aurangabad, India

Sivasudha T and Kumar PA (2007) Sequence analysis of cereal sucrose synthase genes and isolation of sorghum sucrose synthase gene fragment. Afr J Biotechnol 6:2386-2392

Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K (2000) Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. Crop Sci 40:226–232

Snow AA, Pilson D, Rieseberg LH, Paulsen M, Pleskac N, Reagon MR, Wolf DE, Selbo SM (2003) A *Bt* transgene reduces herbivory and enhances fecundity in wild sunflowers. Ecol Appl 13:279–286

Sobral BWS, Braga DPV, LaHood ES, Keim P (1994) Phylogenetic analysis of chloroplast restriction enzyme site mutations in the *Saccharinae* Griseb. Subtribe of the *Andropogoneae* Dumort. Tribe. Theor Appl Genet 87, 843–853

Srinivasachary, Dida MM, Gale MD, Devos KM (2007) Comparative analyses reveal high levels of conserved collinearity between the finger millet and rice genomes. Theor Appl Genet 115:489-499

Studer B, Asp T, Frei U, Hentrup S, Meally H, Guillard A, Barth S, Muylle H, Roldan-Ruiz I, Barre P, Koning-Boucoiran C, Uenk-Stunnenberg G, Dolstra O, Skot KP, Turner LB, Humphreys MO, Kolliker R, Roulund N, Nielsen KK, Lubberstedt T (2008) Expressed sequence tag-derived microsatellite markers of perennial ryegrass (*Lolium perenne* L.). Mol Breeding 21:533-548

Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen HT (2000) Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench). Genome 43:461-469

Subudhi PK, Nguyen HT (2000) Linkage group alignment of sorghum RFLP maps using a RIL mapping population. Genome 43:240-249

Tanksley SD, Bernatzky R, Lapitan NJ, Prince JP (1988) Conservation of gene repertoire but not gene order in pepper and tomato. Proc Natl Acad Sci 85:6419-6423

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141-1160

Tao YZ, Hardy A, Drenth J, Henzell RG, Franzmann BA, Jordan DR, Butler DG, McIntyre CL (2003) Identification of two different mechanism s for sorghum midge resistance trough QTL mapping. Theor Appl Genet 107:116-122

Tao YZ, Henzell RG, Jordan DR, Butler DG, Kelly AM, McIntyre CL (2000) Identification of genomic regions associated with stay-green in sorghum by testing RILs in multiple environments. Theor Appl Genet 100:1225-1232

Tao YZ, Jordan DR, Henzell RG, McIntyre CL (1998a) Construction of a genetic map in sorghum RIL population using probes from different sources and its comparison with other sorghum maps. Aust J Agric Res 49:729-736

Tao YZ, Jordan DR, Henzell RG, McIntyre CL (1998b) Identification of genomic regions for rust resistance in sorghum. Euphytica 103:287-292

Tao YZ, Manners JM, Ludlow MM, Henzel RG (1993) DNA polymorphisms in grain sorghum, *Sorghum bicolor* (L.) Moench. Theor Appl Genet 86:679-688

Taramino G, Tarchini R, Ferrario S, Lee M, Pe´ ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor Appl Genet 95:66–72

Tegelstrom H (1992) Detection of mitochondrial DNA fragments. In: A.R. Hoelzel (eds), Molecular genetic analysis of populations: a practical approach. IRL Press, Oxford, pp 89-114

Tesso T, Kapran I, Grenier C, Snow A, Sweeney P, Pedersen J, Marx D, Bothma G, Ejeta G (2008) The potential for crop-to-wild gene flow in sorghum in Ethiopia and Niger: A geographic survey. Crop Sci 48:1425-1431

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796-815

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization f gene-derived SSR-markers in barley (*Hordeum vulgare* (L.) Theor Appl Genet 106:411-422

Thomas WTB (2003) Prospects for molecular breeding of barley. Ann Appl Biol 142:1-12

Tonguc M, Griffith PD (2004) Genetic relationships of Brassica vegetables determined using database derived simple sequence repeats. Euphytica 137:193-201

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1996) Identification of quantitative trait loci associated with pre-flowering drought tolerance in sorghum. Crop Sci 36:1337-1344

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1997) Genetic analysis of post-flowering drought tolerance and components of grain development in *Sorghum bicolor* (L.) Moench. Mol Breed 3:439-448

Uptmoor R, Wenzel W, Friedt W, Donaldson G, Ayisi K, Ordon F (2003) Comparative analysis on the genetic relatedness of Sorghum bicolor accessions from Southern Africa by RAPDs, AFLPs and SSRs. Theor Appl Genet 106:1316–1325

Varshney RK, Chabane K, Hendre PS, Aggarwal RK, Graner A (2007) Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Sci 173:638–649

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

Varshney RK, Grosse I, Hahnel U, Thiel T, Rudd S, Zhang H, Prasad M, Stein N, Langridge P, Graner A (2006) Genetic mapping and physical mapping (BAC-identification) of EST-derived microsatellite markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 113:239–250

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 of frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537-546

Ventelon M, Due M, Garsmeur O, Doligex A, Ghesquiere A, Lorieux M, Rami JF, Glaszmann JC, Grivet L (2001) A direct comparison between the genetic maps of sorghum and rice. Theor Appl Genet 102:379-386

Vollrath D, Jaramillo-Babb VL (1999) A sequence-ready BAC clone contig of a 2.2-Mb segment of human chromosome 1q24. Genome Res 9:150–157

Wang ML, Barkkley NA, Yu J-K, Dean RE, Newman ML, Sorrells ME, Pederson GA (2005) Transfer of simple sequence repeat (SSR) markers from major cereal crops to minor grass species for germplasm characterization and evaluation. Plant Genetic Resources 3:45-57

Wang ML, Dean R, Erpelding J, Pederson G (2006) Molecular genetic evaluation of sorghum germplasm differing in response to fungal diseases: Rust (*Puccinia purpurea*) and anthracnose (*Collectotrichum graminicola*). Euphytica 148:319-330

Wang PZ, Han ZG, Zhang T (2004) EST-SSR based genetic diversity assessment of xinjiang cultivars in upland cotton. J Genet Mol Biol 15:171-181

Weber JL, Myers EW (1997) Human whole-genome shotgun sequencing. Genome Res 7:401-409

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity arrays technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci 101:9915-9920

Wenzl P, Li H, carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Jaroslava O, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E, Kilian A (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7:206-228.

Whitkus R, Doebley J, Lee M (1992) Comparative genome mapping of sorghum and maize. Genetics 132:1119-1130

Wittenberg AHJ, van der Lee T, Cayla C, Kilian A, Visser RGF, Schouten HJ (2005) Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. Mol Gen Genomics 274:30-39.

Wu YQ, Huang Y (2006) An SSR genetic map of *Sorghum bicolor* (L.) Moench and its comparison to published genetic map. Genome 50:84-89

Wu YQ, Huang Y (2008) Molecular mapping of QTLs for resistance to the greenbug (*Schizaphis graminum* (Rondani) in *Sorghum bicolor* (Moench). Theor Appl Genet 117:117-124

Wu YQ, Huang Y, Tauer CG, Porter DR (2006) Genetic diversity of sorghum accessions resistant to greenbugs as assessed with AFLP markers. Genome 49:143-149

Xia L, Peng K, Yang S, Wenzl P, Carmen de Vicente M, Fregene M, Kilian A (2005) DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. Theor Appl Genet 110:1092–1098

Xie Y, McNally K, Li C-Y, Leung H, Zhu Y-Y (2006) A high-throughput genomic tool: Diversity array technology complementary for rice genotyping. J Integr Plant Biol 48:1069-1076

Xu GW, Magill CW, Schertz KF, Hart GE (1994) A RFLP linkage map of *Sorghum bicolor* (L) Moench. Theor Appl Genet 89:139-145

Xu JC, Weerasuriya YM, Bennetzen JL (2001) Construction of genetic map in sorghum and mapping of the germination stimulant production gene response to *Striga asiatica*. Acta Genetica Sinica 28:870-876

Xu W, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen NT (2000) Molecular mapping of QTLs conferring stay-green in grain sorghum (*Sorghum bicolor* L. Moench). Genome 43:461-469

Xue S, Zhang Z, Lin F, Kong Z, Cao Y, Li C, Yi H, Mei M, Zhu H, Wu J, Xu H, Zhao D, Tian D, Zhang C, Ma Z (2008) A high-density intervarietal map of the wheat genome enriched with markers derived from expressed sequence tags. Theor Appl Genet 117:181-189

Yang S, Pang W, Harper J, Carling J, Wenzl P, Huttner E, Zong X, Kilian A (2006) Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology (DArT). Theor Appl Genet 113:585–595

Yu J-K, Dake TM, Singh S, Benscher D, Li W, Gill B, Sorrells ME (2004) Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 47:805-818

Zhang LY, Ravel C, Bernard M, Balfourier F, Leroy P, Feuillet C, Sourdille P (2006) Transferable bread wheat EST-SSRs can be useful for phylogenetic studies among the Triticeae species. Theor Appl Genet 113:407-418

Zhi-Ben Y, Yi S, Xiao-Hong L, Wei-Jun Z, Min Y, Li-Xia C (2006) Advances in genetic mapping of the sorghum genome. Chinese J Agric Biotechnol 3:15-161

Zhu H, Blackmon BP, Sasinowski M, Dean RA (1999) Physical map and organization of chromosome 7 in the rice blast fungus, *Magnaporthe grisea*. Genome Res 9:739–750

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. T Plant Genome 1:5–20