

Mol Breeding (2010) 26:409–418
DOI 10.1007/s11032-009-9365-9

In silico mapping of important genes and markers available in the public domain for efficient sorghum breeding

P. Ramu · S. P. Deshpande · S. Senthilvel ·
B. Jayashree · C. Billot · M. Deu ·
L. Ananda Reddy · C. T. Hash

Received: 2 July 2009 / Accepted: 1 December 2009 / Published online: 29 June 2010
© Springer Science+Business Media B.V. 2010

Abstract Crop genome sequencing projects generate massive amounts of genomic sequence information, and the utilization of this information in applied crop improvement programs has been augmented by the availability of sophisticated bioinformatics tools. Here, we present the possible direct utilization of sequence data from a sorghum genome sequencing project in applied crop breeding programs. Based on sequence homology, we aligned all publicly available simple sequence repeat markers on a sequence-based physical map for sorghum. Linking this

physical map with already existing linkage map(s) provides better options for applied molecular breeding programs. When a new set of markers is made available, the new markers can be first aligned on a sequence-based physical map, and those located near the quantitative trait locus (QTL) can be identified from this map, thereby reducing the number of markers to be tested in order to identify polymorphic flanking markers for the QTL for any given donor × recurrent parent combination. Polymorphic markers that are expected (on the basis of their position on the sequence-based physical map) to be closely linked to the target can be used for foreground selection in marker-assisted breeding. This map facilitates the identification of a set of markers representing the entire genome, which would provide better resolution in diversity analyses and further linkage disequilibrium mapping. Filling the gaps in existing linkage maps and fine mapping can be achieved more efficiently by targeting the specific genomic regions of interest. It also opens up new exciting opportunities for comparative mapping and for the development of new genomic resources in related crops, both of which are lagging behind in the current genomic revolution. This paper also presents a number of examples of potential applications of sequence-based physical map for sorghum.

P. Ramu and S. P. Deshpande have contributed equally to this study.

Electronic supplementary material The online version of this article (doi:[10.1007/s11032-009-9365-9](https://doi.org/10.1007/s11032-009-9365-9)) contains supplementary material, which is available to authorized users.

P. Ramu · S. P. Deshpande · S. Senthilvel ·
B. Jayashree · C. T. Hash (✉)
M.S. Swaminathan Applied Genomics Lab (AGL),
International Crops Research Institute for the Semi-Arid
Tropics (ICRISAT), Patancheru PO, Hyderabad, Andhra
Pradesh 502324, India
e-mail: c.hash@cgiar.org

P. Ramu · L. Ananda Reddy
Department of Genetics, Osmania University,
Hyderabad 500007, Andhra Pradesh, India

C. Billot · M. Deu
CIRAD, UMR DAP, TA, A-96/03, Av Agropolis, 34398
Montpellier Cedex 5, France

Keywords Arabidopsis · Genes ·
Genome sequence · Markers · Physical map ·
Rice · Sorghum

Introduction

Recent advances in sequencing technologies and the advent of bioinformatics tools have led to the rapid availability of crop genome sequences and their annotation. The genome sequences of many important plant species are currently available in the public domain, including those of both monocotyledons [rice (*Oryza sativa* L.), sorghum (*Sorghum bicolor* (L.) Moench), maize (*Zea mays* L.)] and dicotyledons [arabidopsis (*Arabidopsis thaliana*), barrel medic (*Medicago truncatula*), black cottonwood (*Populus trichocarpa*), among others]. Efforts are underway to sequence other important species, such as tomato (*Solanum lycopersicum*), potato (*Solanum tuberosum*), lotus (*Lotus japonicus*), soybean (*Glycine max*), cassava (*Manihot esculenta*), foxtail millet (*Setaria italica*), eucalyptus (*Eucalyptus grandis*), and *Brachypodium distachyon*, and draft genome sequences of a number of these species are already available for public use.

Knowledge of crop genome sequences improve our understanding of the evolutionary processes and genome organization, not only in the sequenced species but also in related species. The availability of whole-genome sequences triggers the development of new molecular resources and tools, such as molecular markers for precise genetic mapping and comprehensive molecular analysis of genome structure and function. The availability of a high-quality genome sequence for a crop enables further exploration of the diversity among different varieties. In turn, the presence of diversity within the genomic component is a very important component in the search for new alleles for utilization in plant breeding programs aimed at developing novel varieties. Information on the genome sequence has direct application in crop improvement through the targeted development of molecular markers and proper design of marker-assisted selection (MAS) in breeding programs. In this paper, we present the possible direct utilization of data from a genome sequencing project in applied breeding programs using sorghum as an example.

Sorghum acts as a model for grasses having the C₄ photosynthetic pathway (Kresovich et al. 2005). Genome sequencing of sorghum was initiated at the end of 2005 through the ‘Community sequencing program’ of the U.S. Department of Energy–Joint Genome Institute (JGI) (Paterson et al. 2006, 2009),

and the complete genome sequence (8×) information for this species was made available for public use (Sorbi v1.0; <http://genome.jgi-psf.org/Sorbi1/Sorbi1.info.html>) (Paterson et al. 2009). This information has provided an opportunity for gaining a better understanding of genome organization in cultivated sorghum and its wild relatives, including Johnson grass (*Sorghum halepense*), as well as in maize and sugarcane. Here we present a case study in which published information on simple sequence repeat (SSR) markers and genes from cereals, especially sorghum, were utilized to align these markers and genes on the sorghum genome sequence for a better understanding of linkage-based mapping and in silico physical mapping. We collated information from published papers on approximately 7000 SSR markers and 100 genes and aligned them on a sequence-based physical map. We also compared SSR marker positions between a physical and consensus linkage map. The distances on the physical map and their corresponding linkage map distances are two different entities, and it is therefore an interesting objective to compare them. While the physical map distances are the actual nucleotide sequence-based distances, the linkage distances are governed by the recombination events occurring within the given chromosome segment. For marker-assisted backcrossing (MABC) programs, the co-segregation of the trait and flanking markers is more important. However, for fast recovery of the recurrent parent genome, a knowledge of the size of chromosomal fragment giving near-perfect target trait expression is important. This information would further facilitate the integration of genetic information obtained from quantitative trait loci (QTL) mapping studies and cloning experiments of important genes and also help in assigning proper annotations for newly mapped genes. SSR marker information along with this genetic information will suitably enhance the efficiency of sorghum breeding efforts for important traits.

Materials and methods

Aligned 8× sorghum genome sequence information was downloaded and formatted into a database on the high-performance computer available at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). This aligned sorghum genome

sequence contains 659 Mb (89% of 739 Mb) arranged in ten chromosome pairs, while the remaining 79 Mb is in super clusters and has yet to be assigned to individual chromosomes.

The sequences of all publicly available sorghum SSR primer pairs [Electronic Supplementary Material (ESM) Table 1] were collected and collated in FASTA format. These sequences included genomic SSR markers [*Xtxp* series (Kong et al. 2000; Bhatramakki et al. 2000, <http://sorgblast3.tamu.edu/search/marker.htm>), *Xsb* series (Taramino et al. 1997), *Xgap* series (Brown et al. 1996)], cDNA-based SSR markers [*Xcup* series (Schloss et al. 2002)], SSRs derived from the complete genome shotgun (WGS) sequence of sorghum [*SB* series (Yonemaru et al. 2009)] and expressed sequence tag (EST)–SSR markers [*Xisep* series (Ramu et al. 2009), *Xiabt* series (Arun 2006, Reddy et al. 2008)], other SSR markers of an unknown type [*gpsb* and *mSbCIR* series (developed at CIRAD, France and partially published in Mace et al. 2009)], and other marker systems, such as conserved intron scanning primers (CISPs, Feltus et al. 2006). Primer sequences were then BLAST (basic local alignment search tool) searched against the sorghum genome sequence information. Searches were performed separately for forward and reverse primer sequences with a default value of 11 words as the filter for sequence homology. BLAST searches are accelerated by the heuristic nature of the algorithm, but results in a loss of sensitivity in terms of the identification of sequence similarity. We decided to utilize criteria for primer pair hits for individual SSR marker (as discussed in following section) to improve the sequence similarity and putative SSR alignment with the genome sequence.

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). The BLAST hits of each individual's sequence (forward and reverse) of each primer pair were grouped together. The BLAST results were then analyzed using the following criteria:

1. Presence of both the forward and reverse primer pair BLAST hit on the same sorghum chromosome; this avoids mismatches in terms of SSR alignment on the genome sequence.
2. Occurrence of BLAST hits of forward and reverse primer sequences (for the same primer pair) on the opposite strands (positive and negative) of the sequence assembly of the sorghum genome; this indicates a perfect physical assignment of the SSR marker to the genome sequence.
3. The estimated difference between BLAST hits of the forward and reverse primer sequence hit (for the same primer pair) in terms of start and end nucleotide numbers on the genome sequence, (expected amplicon size) should be between 50 and 500 bp. This would increase the opportunity for utilizing these SSRs with the most commonly used SSR fragment analysis techniques (polyacrylamide gel electrophoresis and DNA sequencers).

The primers satisfying the above criteria were placed on the sequence-based physical map of the sorghum genome. This allocation was based on matching BLAST hit positions of primers to the progressive units along the sequence assembly of each sorghum chromosome using Mapchart (Voorrips 2002). Likewise, all available markers were placed onto the sequence-based physical map of sorghum. Any primer pair that did not meet any one of the above-mentioned criteria was excluded from the physical map. For comparing the arrangement of these SSRs on the sorghum genome sequence based in silico physical map with the linkage map, we utilized the consensus linkage map for sorghum developed by Mace et al. (2009). In this map, we averaged the linkage distances for non-bridge markers (common set of markers across the six populations used to build the framework map) within the recombination window where they mapped in different mapping populations.

Results and discussion

We were able to collate information on a total of 7013 SSR markers, 125 CISP markers, and 100 gene/gene products available in public domain (ESM Table 2) published up to April 30, 2009. Of these, 7013 SSR markers, 1063 SSR (15.2% of total SSR markers), and two CISP markers were dropped by following the criteria discussed in the [Materials and](#)

methods section: (1) 599 SSR markers (56.4% of discarded SSR markers) had no hits; (2) 144 SSR markers (13.5% of discarded SSR markers) had a fragment size >500 bp and were not assigned to a specific chromosome; (3) the remaining markers either had mis-match hits (on different chromosome), one or both primer pairs (forward or reverse) of the SSR did not align perfectly, or the primer pairs were assigned to super clusters. Among the perfectly assigned SSR markers on sorghum genome sequence, 118 were duplicate, 708 produced putatively overlapping fragments, and 132 annealed to regions which contained more than one SSR fragment (ESM 2). The small proportion of non-aligned, duplicate, and overlapping SSR markers indicate that the criteria used for identifying SSRs published in all of the studies were stringent. Physical maps for all ten linkage groups are given in Fig. 1. The linkage groups are named as per Kim et al. (2005). Saturation of markers was mainly noted at the telomeric regions of the chromosomes. No marker was located near and around the centromeric regions of chromosomes SBI-03, SBI-06, SBI-09, and SBI-10 (Fig. 1), and only a very few markers were located near the centromeric regions in the remaining chromosomes. The heat maps from the sorghum genome sequence project revealed that the gene-rich regions (and nearly all SSR markers) are mainly located at the ends of the sorghum chromosomes (Paterson et al. 2009).

Validation of physical map positions

For validation, markers already mapped on one or more sorghum genetic linkage maps (e.g., Bhatramakki et al. 2000; Tao et al. 2000; Klein et al. 2001; Haussmann et al. 2002, 2004; Wu and Huang 2006) were compared to their linkage map positions on the sequence-based physical map. Among the 369 SSR loci for which linkage map positions were available in the public domain, 290 markers (78.5%) mapped to similar positions in both the linkage and physical maps, nine markers (2.4%) mapped to different chromosomes on the physical map than expected from their reported linkage map positions (this may be due to putative duplicate loci), while 70 markers (18.9%) with linkage map positions were not mapped onto the sorghum physical map. These are the markers not fulfilling the criteria as explained earlier. The comparison of data related to the positions of

these differentially mapped markers on the physical map and linkage map revealed that either the primer pair(s) for these markers (20 out of 70) had BLAST hits on super clusters, indicating gaps in the genome sequence assembly, or they were genetically mapping in the regions with poor marker densities and/or near the centromeric regions (19/70).

Similarly, we compared the sequence-based physical map with the sorghum consensus linkage map (Mace et al. 2009). For this comparison, only the SSR marker position was used as physical map positions because other markers, including DArT clones, were not available. For all of the linkage groups (LGs), the SSR marker coverage was restricted and genome coverage was partial. Due to a lack of space, we illustrate only comparisons of LG 5 and LG 6 (Fig. 2) with sorghum chromosomes SBI-05 and SBI-06, respectively. There were 26 and 23 common SSR markers mapped on LG 5 and LG 6, respectively. Of these SSR markers, ten and six SSR markers from LG 5 and LG 6, respectively, were not mapped on the physical map. One marker each on LG 5 and LG 6 was found to map on different chromosome in the physical map. In most cases, the marker order in the linkage map and physical map was similar. As these results are based on a very limited number of SSRs, they require further confirmation by comparison with high-density linkage maps involving not only SSRs, but also other type of markers.

Agronomically important genes and their locations in the sorghum genome

Similar to primer sequences, all previously published gene sequences from sorghum and other cereals were BLAST searched against the sorghum genome sequence database and then assigned to the respective chromosome where they had BLAST hits. For all of the genes, the physical map positions on the sorghum genome, along with their GenBank IDs, are listed in ESM Table 3.

Sorghum is a C_4 grass and a logical complement to rice, the first cereal for which the complete set of genome sequence information was available. The phosphoenolpyruvate carboxylase (PEPC) gene plays a key role in C_4 photosynthesis: it acts as a CO_2 pump in mesophyll cells to concentrate CO_2 at the site of ribulose 1,5-biphosphate carboxylase/oxygenase in the bundle sheet cells (Westhoff and Gowik 2004).

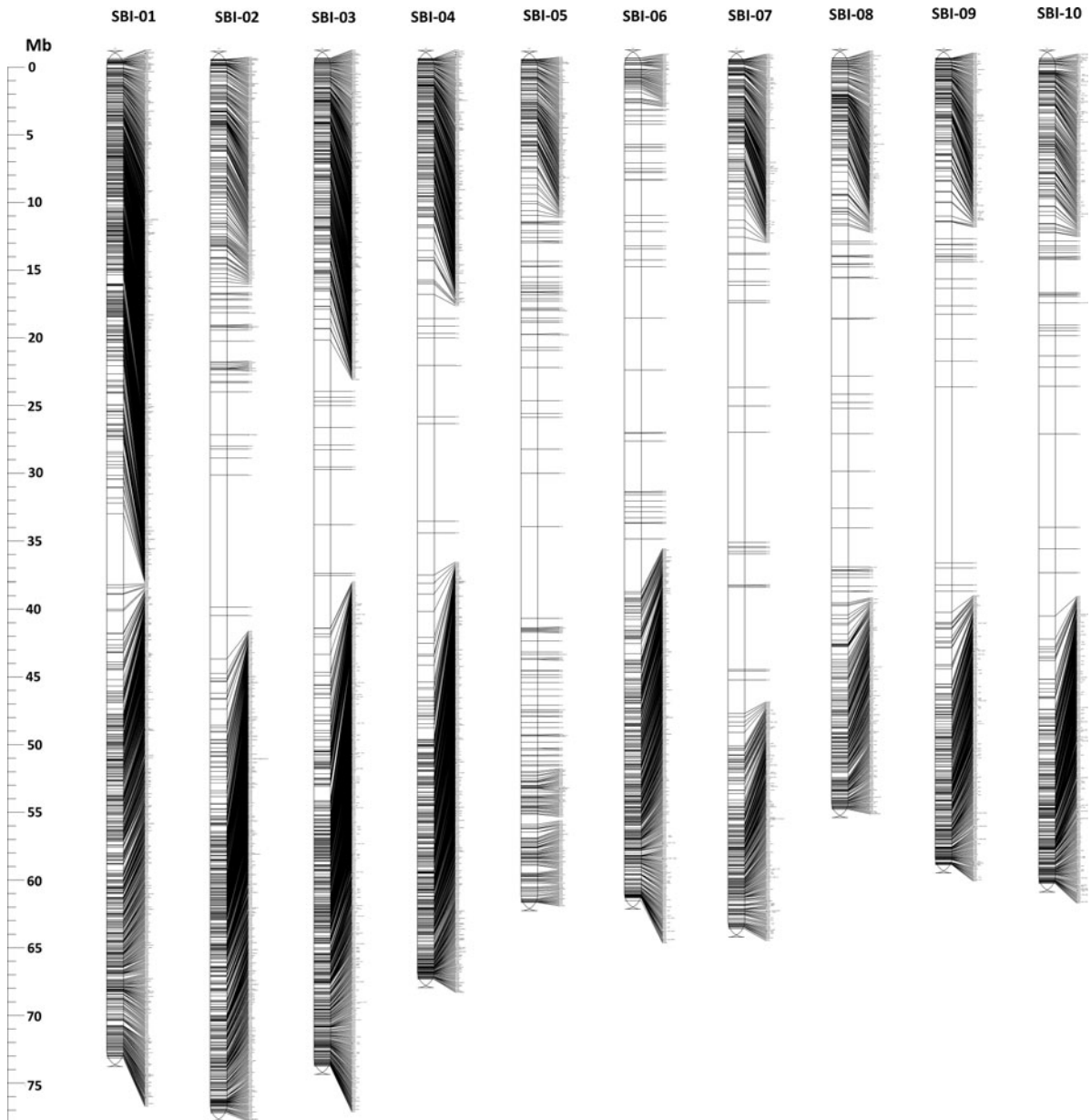


Fig. 1 Positions of important genes and markers on the physical maps for sorghum chromosomes SBI-01 to SBI-10 generated using the sorghum genome sequence. Mapped genes are prefixed with ‘>’ symbol

Gene sequences for PEPC are available for sorghum from two different species, *S. verticilliflorum* (AJ318575) and *S. vulgare* (X17379). These two gene sequences mapped to the distal end of the long arm of sorghum chromosome 10. Another PEPC sequence from *S. verticilliflorum* (AJ293347) is located near by on the same chromosome. The two sequences from *S. verticilliflorum* (AJ293347 and

AJ318575) are separated by about 1.2 kb. These different PEPC sequences may belong to two exons coding for PEPC with one or more intervening introns spanning 1.2 kb.

The nicotianamine aminotransferase (NAAT1) gene is responsible for synthesis and secretion of mugineic acid (MA) from the roots of graminaceous plants. MA family phytosiderophores act as metal

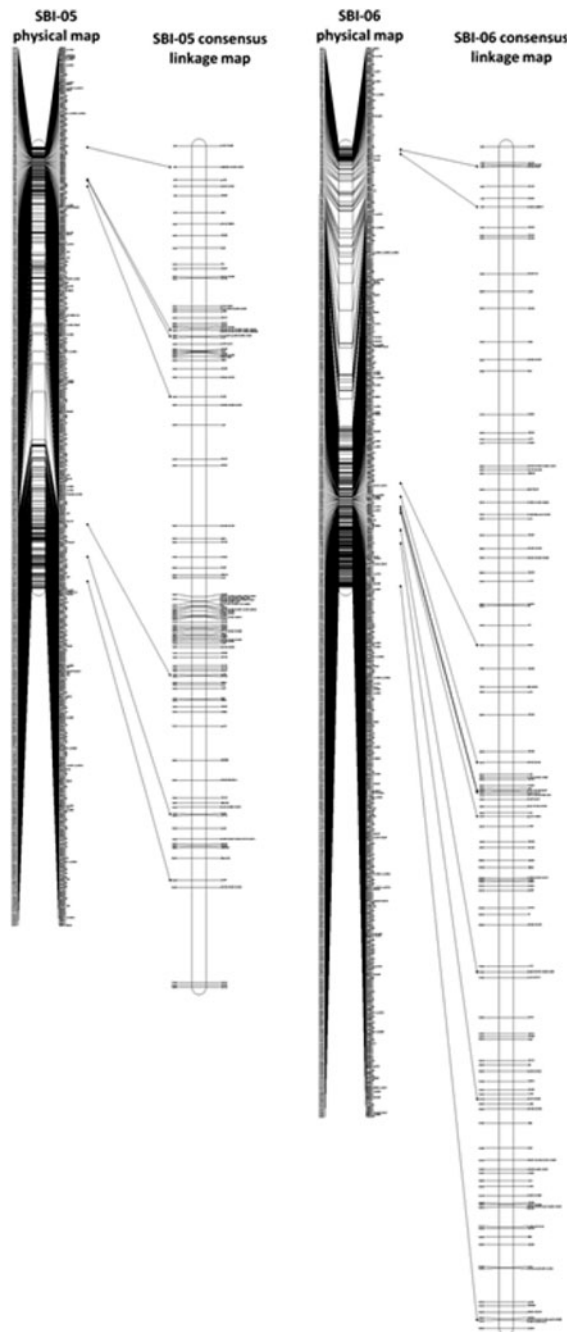


Fig. 2 Comparison of genome sequence based on the in silico physical map and consensus linkage map of sorghum (Mace et al. 2009) for SBI-05 and SBI-06. The scale for the physical map is 4 mm each for each 1 cM of the linkage map and 1 Mb of the genome sequence, respectively

chelators to chelate Fe(III) in the soil and help in uptake of Fe through roots (Cheng et al. 2007). NAAT1 gene homologs from barley (AB005788),

maize (AB375372), and rice (AB206814) have their sorghum counterpart present on the distal end of the short arm of chromosome 2. A major sorghum aluminum tolerance gene (EF611342), belonging to the multi-drug and toxic compound extrusion family (MATE), was found, as expected, on the long arm of chromosome 3 (Magalhaes et al. 2007). Similarly, a gene responsible for lignin biosynthesis (caffeic acid *O*-methyltransferase, the COMT gene) is present on sorghum chromosome 7. The other important gene involved in lignin biosynthesis is cinnamyl alcohol dehydrogenase (CAD), which has four variants characterized in maize (www.maizegdb.org). All of these variants were aligned to same position at the distal end of the short arm of sorghum chromosome 4.

Several genes contributing to the photoperiod responsiveness of flowering were located on sorghum chromosome 1, including phytochrome A (*phyA*) (U56729) and phytochrome C (*phyC*) (U56731) on the short arm of chromosome 1 and phytochrome B (*phyB*) (AF182394) on the long arm. A gene homologous to that responsible for vernalization in barley (*Hordeum vulgare*) (EF591648) and oats (*Avena sativa*) (EU283089), having a 99% similarity with the published genes, was located 53 kb away from the *phyC* gene on sorghum chromosome 1.

The gene responsible for the gamma-kafirin seed storage protein is present on sorghum chromosome 2. Two clones of this storage protein gene (X62480 and AY566299) are present at the same location on chromosome 2. The phytoene synthase genes (PSY3) from maize (DQ356430) and rice (DQ356431) exhibit complete homology with a sorghum gene sequence at the distal end of the long arm of chromosome 2. However, the PSY3 gene reported from sorghum (AY705390) is located 1.7 kb downstream of the homolog of the maize and rice PSY3 genes. PSY is the first enzyme involved in the plastid-localized carotenoid biosynthesis pathway, which controls the flux of carotenoids in seeds. Other important genes mapped on chromosome 2 are opaque-2, gene associated with the lysine content of seed storage proteins, a homolog of the DIMBOA biosynthesis gene (*Bx8*, AF331854) from maize that is involved in stem borer resistance (which maps on the short arm of chromosome 4 in maize), and a homolog of the wheat puroindoline gene (responsible for grain hardness).

The *Zea mays* centroradialis (ZCN) gene family from maize is well characterized. To date, about 24 ZCN genes have been reported in maize. The locations of the homologs of these genes were scattered throughout the sorghum genome (ESM Table 1). The *ZCN24* (EU241914) and *ZCN18* (EU241909) homologs are present at the distal end of sorghum chromosome 2, but at essentially the same position, as the *ZCN18* homolog is an integral part of that of *ZCN24*. However, it has been reported as two different ZCN proteins. Three previously reported gene pairs, namely, *ZCN1* (EU241892) and *ZCN3* (EU241894), *ZCN7* (EU241898) and *ZCN8* (EU241899), and *ZCN9* (EU241900) and *ZCN10* (EU241901), also belong to this category as the sorghum homolog of *ZCN1* is an integral part of the sorghum homolog of *ZCN3*, the *ZCN7* homolog is a part of *ZCN8*, and the *ZCN9* and *ZCN10* homologs have 100% similarity and are located on the short arm of sorghum chromosome 3. To date, only ZCN family genes have been mapped onto chromosome 6 of the sorghum physical map.

MADS box genes are responsible for control of flowering in plants. This family of genes is well characterized in *Arabidopsis*. The homologs of some genes belonging to the MADS box family have been mapped onto the sorghum genome, but they are scattered across the whole genome. The homolog of *Gigantia* (*Gi*), one of the regulators of flowering in long-day plants, is mapped to the short arm of sorghum chromosome 3. Likewise, the positions of sorghum homologs of *VRN* (on sorghum chromosome 1), *MADS2* and *MADS24* from rice (on sorghum chromosome 2), *AGL41* (on sorghum chromosome 3), and *VIN3* (on sorghum chromosome 9) and *VRN2* (on sorghum chromosome 8) were identified on different sorghum chromosomes. The nucleotide sequence of the sorghum sucrose synthase gene (Sivasudha and Kumar 2007) was BLAST searched against the sorghum genome sequence; as expected, there was a clear difference between positions of fragment I and II on sorghum chromosome 10.

Applications in molecular breeding programs

The Human Genome Project contributed enormously to biomedical research and development. However, such an approach is not likely to be the case in agriculture as the sequencing of one plant species can

not answer all of the questions in agriculture, which is the result of the collective interaction of human activities with the environments of different crops, livestock, and pest species genomes of varying size and complexity. However, it is expected that crop genome sequencing projects will be able to provide solutions to questions unanswered by conventional breeding programs and that these solutions in turn will be used by the next generation of agricultural researchers and in future breeding programs (Martienssen 2004).

SSR markers have remained a preferred marker system in molecular crop breeding programs. These markers are tremendously helpful for characterizing genetic diversity and germplasm population structure, for linkage map development and saturation, and for introgression of QTLs into locally adapted varieties and hybrid parental lines via backcrossing (BC) programs. To this end, we need QTL map positions and anchored markers. If we are to develop new markers, we first need to align the primer pairs to determine their position on the sequence-based physical map. By adopting this approach, instead of testing all new primer pairs for newly developed markers, we are able to select only those markers located near the QTL regions or genes of interest and then screen these for parental polymorphism. Those detecting scorable polymorphism can then be genetically mapped and used in BC programs to transfer the target QTLs. This approach will also allow the gaps on linkage maps to be filled and markers to be targeted directly to regions requiring further marker locus saturation. Thus, genome sequencing projects are not only helpful for identifying genes but they also facilitate the development of new markers (with little additional cost) in selected regions of chromosomes that are required to achieve the degree of linkage map marker saturation essential for effective QTL mapping and transfer.

A number of examples in which the direct utilization of sorghum genome sequence information has been applied in crop improvement programs are illustrated in the following sections.

Genome-wide scanning for diversity with the proper selection of SSRs

The availability of SSRs was not a constraint in the recent past in studies on crops such as sorghum. The

positioning of available SSRs, not yet mapped genetically, on a sequence-based physical map using sequence homology helps the researcher choose a set of markers with complete genome coverage for the effective scanning of genetic diversity. The mapping of all markers onto the sequence-based physical map of sorghum allowed us to select a set of about 100 markers (ESM Table 4) to represent the complete genome for effective genome-wide scanning of diversity. Furthermore, if any economically important gene is identified, the allele mining through association analysis will be made easy through the design of more markers to scan the target region.

Gap filling in the existing linkage maps of sorghum

In most mapping experiments, the proper coverage of the genome depends on the level of polymorphism between the parents, the type of marker system, and the availability of markers in the crop species under study. However, the use of polymorphic markers is still constrained by the unavailability of information on the exact genomic region to which they map. Here we cite an example of a recombinant inbred line (RIL) population segregating for *striga* resistance based on a cross N13 × E 36-1 (Ramu et al. 2009). The linkage map for the RIL population had considerable gaps, which has reduced the precision of identifying the regions harboring the *striga* resistance gene(s). After aligning SSR markers on physical map, we identified seven markers (at least one marker per QTL for five QTLs) to fill the gap in the QTL interval regions. The availability of physically mapped markers helped to achieve better genome coverage, which in turn improved QTL detection power. Importantly, we were able to add more markers in the genomic regions harboring the QTLs for *striga* resistance (Hausmann et al. 2004), which could improve the efficiency of BC program.

Improved resolution of QTLs and a better choice of markers for efficient MABC breeding projects

QTL mapping studies are usually hampered by a lack of proper genome coverage, which results in either the important QTLs either not being detected or being detected at a lower resolution. For a major QTL present on the longer arm of SBI-10 imparting resistance to shoot fly (*Atherigona socatta*) in

sorghum [mapped in a RIL population based on cross BT×623 (susceptible to shoot fly) × IS18551 (resistance to shoot fly)], there was a lack of polymorphic loci in the QTL interval and also among the flanking markers (Folkertsma et al. 2003). This QTL was associated with trichome density (on both leaf surface) and the expression of glossiness intensity and contributes to the non-preference mechanism. The support interval for this QTL in the original study was reported to be 30 cM with a SSR, *Xgap01*, falling in the QTL interval and another SSR, *Xtxp141*, being 30 cM further. Two additional EST–SSR markers identified from the sequence-based alignment of newly developed EST–SSRs (Ramu et al. 2009) were used for map saturation. The EST–SSRs *Xisep0630* and *Xisep0639* were mapped between the interval of *Xgap01* and *Xtxp141*. This has not only resulted in reducing the QTL interval to 18 cM, but also in *Xisep0639* falling in the support interval and *Xisep0630* being located at the rim of the support interval. Also, the supporting LOD scores increased from 8.59 to 23.58, and the genotypic variation contribution increased from 14.6 to 20.9.

Similar to many MABC programs, the MAS for this QTL was constrained by the low level of polymorphic markers flanking the QTL regions. Although the QTL had the best phenotypic variation to contribute to the target trait, the reduced number of polymorphic markers hinders its utilization in applied breeding programs. The sequence-based physical map of sorghum identified about 73 SSR loci spanning the flanking regions through the QTL interval. From these, 12 markers (ESM Table 5) were selected on the basis of information already available on successful amplification in sorghum. When tested for polymorphism, nine loci were found to be polymorphic, which has enormously improved the efficiency of MABC for transferring shoot fly resistance.

Sorghum and sugarcane belong to the same tribe within the Poaceae, namely the Andropogoneae. These two crops share considerable homology between their genomes even though sugarcane has a seven- to tenfold larger genome size than sorghum (Paterson et al. 2009). Sorghum markers can be effectively used in sugarcane QTL mapping and MAS programs, especially, those sorghum markers derived from transcribed genomic regions, such as EST–SSR markers (Yogesh Parmessur, personal communication).

The genome sequencing project for any crop species opens a new era of genomics and brings new challenges regarding the efficient use of huge amounts of sequence information. The direct application of this information is based on other genetic and genomic resources that are already available for the species. Genetic studies in the various fields of agronomy of the concerned crop species form the basis of utilization of this resource. Various genomic resources in addition to ESTs, such as completely characterized genes and their allele richness in concerned species, can act as carriers for successful application of the information in crop breeding programs. This physical mapping information can be integrated with information from various genetic studies in sorghum and then overlaid with information from QTL mapping, QTL cloning, and gene cloning studies. Such a total set of genetic information will form the basis of a meta-analysis, which will help to correlate and understand the nature of genetic interaction for different traits studied in different populations. Such an approach will enhance the scope of utilizing such genomic information for efficient crop improvement. Based on the results of our study, it can be easily be seen that the SSR markers published to date are quite unique and non-redundant and thus provide better choices of markers in applied breeding programs for capturing global genetic diversity and population structure. They consequently reveal polymorphism even between closely related germplasm accessions, provide an improved resolution of the genomic region harboring target trait(s) and fine mapping and, more importantly, allow allele mining and efficient marker-assisted introgression of these alleles.

Comparisons of genome organization, genes, and intergenic regions between cereal species will enable regions that are highly conserved or rapidly evolving to be identified. Such regions are expected to yield crucial insights into genome evolution, speciation, and domestication (International Rice Genome Sequencing Project 2005). The development of physical maps for crops for which genome sequences are available facilitates our understanding of the genomic relationships between related crop species and so aids the development of markers for related crops for which sufficient genomic tools are not yet available. Synteny and collinearity are the key factors in comparative genome mapping for studying

ancestral genome evolution (Tang et al. 2008). With the help of physical maps and synteny between genomes of crop species, the development of markers in the related crop species, in specific targeted regions, should now be a practical option. This approach will greatly speedup the development of genomic tools required for marker-assisted breeding in orphan crop species that would otherwise be unlikely to benefit from the current genomic revolution.

Acknowledgments The research fellowship provided to PR by the Council of Scientific and Industrial Research (CSIR), New Delhi, India is greatly acknowledged. We also thank to Dr. A.H. Paterson for his action that led to sorghum genome sequencing. We acknowledge the Generation Challenge Program (GCP) for providing the Paracel high-performance computing facility at ICRISAT, Patancheru. We also thank Dr. Fakrudin (University of Agricultural Sciences, Dharwad, Karnataka, India) for providing the EST–SSR primer pair sequences in sorghum and Mr. Yogesh for providing the information related to transferability of sorghum EST–SSRs in sugarcane.

References

- Altschul SF, Miller WG, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410
- Arun SS (2006) In silico EST data mining for elucidation of repeats biology and functional annotation in sorghum [*Sorghum bicolor* (L.) Moench.]. MSc thesis. University of Agricultural Sciences, Dharwad
- Bhatramakki D, Dong J, Chhabra AK, Hart G (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. *Genome* 43:988–1002
- 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
- Cheng L, Wang F, Shou H, Huang F, Zheng L, He F, Li J, Zhao F-J, Ueno D, Ma JF, Wu P (2007) Mutation in nicotianamine aminotransferase stimulated the Fe(II) acquisition system and led to iron accumulation in rice. *Plant Physiol* 145:1647–1657
- Feltus FA, Singh HP, Lohithaswa HC, Schilze SR, Silva TD, Paterson AH (2006) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved non-coding sequences in orphan crops. *Plant Physiol* 140:1183–1191
- 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: 11th Int Plant Animal

- Genome Conf. San Diego. Available at: http://www.intl-pag.org/11/abstracts/P5d_P462_XI.html
- Hausmann BIG, Hess DE, Seetharama N, Welz HG, Geiger HH (2002) 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
- Hausmann BIG, Hess DE, Omany GO, Folkertsma RT, Reddy BVS, Kayentao M, Weiz 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
- International Rice Genome Sequencing Project (2005) The map-based sequencing of the rice genome. *Nature* 436:793–800
- Kim J-S, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005) Chromosome identification and nomenclature of *Sorghum bicolor*. *Genetics* 169:1169–1173
- Klein RR, Klein PE, Chhabra AK, Dong J, Pammi S, Childs KL, Mullet JE, Rooney WR, Schertz KF (2001) Molecular mapping of the *rfl* gene for pollen fertility restoration in sorghum (*Sorghum bicolor* L.). *Theor Appl Genet* 102:1206–1212
- 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, Burke J, Clifton S, Cordonnier-Pratt M-M, Cox S, Dahlberg J et al (2005) Towards sequencing the sorghum genome. A U.S. national science foundation-sponsored workshop report. *Plant Physiol* 138:1898–1902
- Mace ES, Rami J-F, Bouchet S, Klein PE, Klein RR, Kilian A, Wenzl P, Xia L, Halloran K, Jordan DR (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput diversity array technology (DArT) markers. *BMC Plant Biol* 9:13
- Magalhaes JV, Liu J, Guimaraes CT, Lana UGP, Alves VMC, Wang Y-H, Schaffert RE, Hoekenga OA, Piner MA, Shaff JE, Klein PE, Carneiro NP, Coelho CM, Trick HN, Kochian LV (2007) A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nat Genet* 39:1156–1161
- Martienssen RA (2004) Crop plant genome sequence: what is it good for? *Crop Sci* 44:1898–1899
- Paterson AH, Bowers JE, Gingle A, Paterson DG, Kresovich SE, Messing J, Hash CT, Rokhsar DS (2006) Sequencing of the sorghum genome. In: 14th Int Plant Animal Genome Conf. San Diego. Available at: http://www.intl-pag.org/pag/14/abstracts/PAG14_W65.html
- 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, Ohtillar 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 WareD, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457:551–556
- Ramu P, Kassahun B, Senthilvel S, Kumar CA, Jayashree B, Folkertsma RT, Reddy LA, Kuruvinashetti MS, Hausmann BIG, Hash CT (2009) Exploiting rice-sorghum synteny for targeted development of EST-SSRs to enrich the sorghum genetic linkage map. *Theor Appl Genet* 119:1193–1204
- Reddy PS, Fakrudin B, Rajkumar Punnuri SM, Arun SS, Kuruvinashetti MS, Das IK, Seetharama N (2008) Molecular mapping of genomic regions harboring QTLs for stalk rot resistance in sorghum. *Euphytica* 159(1–2):191–198
- Schloss SJ, Mitchell SE, White GM, Kukutla 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
- Sivasudha T, Kumar PA (2007) Sequence analysis of cereal sucrose synthase genes and isolation of sorghum sucrose synthase gene fragment. *African J Biotech* 6(20):2386–2392
- Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. *Science* 320:486–488
- 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
- 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
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *T J Hered* 93:77–78
- Westhoff P, Gowik U (2004) Evolution of C4 phosphoenolpyruvate carboxylase genes and proteins: A case study with the genus *Flaveria*. *Ann Bot* 93:13–23
- 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
- Yonemaru JI, Ando T, Mizubayashi T, Kasuga S, Matsumoto T, Yano M (2009) Development of genome-wide simple sequence repeat markers using whole-genome shotgun sequences of sorghum (*Sorghum bicolor* (L.) Moench). *DNA Res* 16:187–193