ADVANCES IN SORGHUM GENETIC MAPPING WITH IMPLICATIONS FOR SORGHUM IMPROVEMENT

M. LEE
Department of Agronomy,
Iowa State University,
Ames, Iowa, United States of America

Abstract

DNA markers and genetic maps have become important tools for direct investigations of several facets of crop improvement and will provide vital links between plant breeding and basic plant biology. The markers and maps will become more important for increased crop production because plant genetics will be required to extend or replace extant management practices such as chemical fertilizers, pesticides, and irrigation [1]. Despite the importance of the sorghum crop, comprehensive genetic characterization has been limited. Therefore, the primary goal of this research program was to develop basic genetic tools to facilitate research in the genetics and breeding of sorghum. The first phase of this project consisted of constructing a genetic map based on restriction fragment length polymorphisms (RFLPs). The ISU sorghum map was created through linkage analysis of 78 F2 plants of an intraspecific cross between inbred CK60 and accession PI229828 [2]. Subsequent mapping efforts in several labs have enriched the sorghum map to the point where it now contains over 1,500 loci defined by RFLPs and many others defined by mutant phenotypes and QTLs. The ISU map consists of 201 loci distributed among 10 linkage groups covering 1299 cM. Comparison of sorghum and maize RFLP maps on the basis of common sets of DNA probes revealed a high degree of conservation as reflected by homology, copy number, and colinearity. Examples of conserved and rearranged locus orders were observed.

The same sorghum population was used to map genetic factors (mutants and QTLs) for several traits including vegetative and reproductive morphology, maturity, insect, and disease resistance. Four QTLs for plant height, an important character for sorghum adaptation in temperate latitudes for grain production, were identified in a sample of 152 F2 plants [3] whereas 6 QTLs were detected among their F3 progeny. These observations and assessments of other traits at 4 QTLs common to F2 plants and their F3 progeny indicate some of these regions correspond to loci (*dw*) previously identified on the basis of alleles with highly qualitative effects. Four of the six sorghum plant height QTLs seem to be orthologous to plant height QTLs in maize. Other possible instances of orthologous QTLs included regions for maturity and tillering. These observations suggest that the conservation of the maize and sorghum genomes encompasses sequence homology, colinearity, and function.

The genetic information and technology developed on the basis of DNA markers could be used in several facets of breeding, genetics, and other basic biological investigations. In addition, DNA markers have been used to survey large collections of elite sorghum germ plasm to determine the degree of genetic relationships and genetic diversity [4]. RFLP data seem to portray genetic relationships more accurately than the methods based exclusively on the coancestry coefficient. This information provides the basis for more accurate perceptions of genetic relationships and diversity. Recently, PCR-based markers (microsatellites; simple sequence repeats, SSRs) have become available for sorghum. Thirty-two SSR loci have been mapped throughout the sorghum genome. More SSR loci will be developed so that a rapid, reliable, and non-radioactive genetic marker system will be available for sorghum research in the near future.

1. INTRODUCTION

Knowledge of a crop's genetic architecture will become more important for increased crop production because plant genetics will be required to extend or replace extant management practices such as chemical fertilizers, pesticides, and irrigation. Such knowledge will include more detailed descriptions of genome organization, the crop's gene pools, and genes and pathways controlling important phenotypes. In many instances, DNA markers and genetic maps will be important tools for direct investigations of these areas and will provide vital links between plant breeding and basic plant biology [1].

Many of the limitations of plant breeding methods have been rooted in the status of the technical infrastructure for conducting genetic analyses. Breeders and geneticists of all crops have lacked an informative and integrated genetic context to aid interpretation and

conciliation of perspectives provided by seemingly different approaches to genetic improvement. The result has been a situation resembling the Tower of Babel with breeders, geneticists, cytogeneticists, taxonomists, molecular biologists, plant pathologists, and other factions contributing to the confusion. A key component of the infrastructure and context of future plant breeding programs will be genetic maps. The maps, when fully integrated, will have several roles: 1) provide a focal point and hub for data derived from the perspectives of myriad disciplines for each crop; 2) constitute a vital two-way avenue connecting plant breeding and basic plant biology; 3) contribute essential information for positional cloning of genes; 4) facilitate a considerable and directed expansion of a crop's gene pool through comparative mapping of related and unrelated taxa; 5) accelerate identification and incorporation of useful genes into cultivars; and 6) contribute important clues toward understanding the biological basis of complex traits and phenomena important to crop improvement. The significance of these and other roles and their implementation will vary with the repertoire of genetic technologies available to the crop, breeding methods and goals, and the nature of the crop's nuclear genome. However, the foundation provided by the maps have had, and will have a positive impact on the genetic improvement of crop species in many instances.

2. DEVELOPMENT OF A GENETIC MAP FOR SORGHUM AND COMPARATIVE MAPPING WITH MAIZE

In terms of cultivated area, sorghum (*Sorghum bicolor* L. Moench) is the world's fifth leading cereal [5]. Despite the importance of the sorghum crop, comprehensive genetic characterization has been limited. Therefore, the primary goal of this research program was to develop basic genetic tools to facilitate research in the genetics and breeding of sorghum. The specific objectives were to 1) develop a complete genetic linkage map for sorghum based on DNA markers, 2) identify the genetic locations of genes of interest to sorghum breeding programs, 3) integrate the sorghum genetic map(s) with maps of allied grass species, and 4) assess the genetic diversity and relationships of sorghum germ plasm.

In sorghum, several maps have been developed for various objectives by different research groups. These maps are being compared and integrated on the basis of probes exchanged among various laboratories (G. Hart, 1996 personal communication). The RFLP linkage map produced at ISU (Iowa State University) was created through linkage analysis of 78 F2 plants of an intraspecific cross between inbred CK60 and accession PI229828 [2]. These parents were selected to produce the mapping population for several reasons; 1) adequate DNA polymorphism, 2) parental difference for resistance to an important insect pest, the aphid, *Schizapus graminum* (greenbug, race E), 3) parental difference for resistance to the fungal pathogen, *Peronsclerospora sorghi*, the causal agent of downy mildew, 4) parental divergence for numerous morphological and developmental traits (e.g. plant height, flowering, and tillering), 5) no known chromosomal polymorphisms and 6) presumably, rates and patterns of recombination more representative of other intraspecific crosses. Selection of this population and use of maize DNA probes to detect RFLPs permitted rather efficient collection of extensive genetic information for sorghum and the means of relating the information to that of maize and other grasses.

The map consists of 201 loci distributed among 10 linkage groups covering 1299 cM. Presumably the number of linkage groups correspond to the basic chromosome complement of sorghum (n=10). The RFLP loci were detected through hybridizations with probes of maize genomic (52), maize cDNA (124), and sorghum genomic (10) clones. Most probes detected a single RFLP locus (172) but there was evidence of genomic duplication as many probes

(76) detected more than one band with a strong signal. However, in this population the additional bands were usually monomorphic. Segregation data at 95% of the loci fit expected ratios for an F2 generation of a cross between two homozygous parents. Loci with deviant ratios were located predominantly to one region of linkage group B. All features of this initial map have been verified in a second and larger sample from this F2 population used for mapping quantitative trait loci (QTL [3]).

Recently, a collaboration among my lab, the University of Milan, Italy (E. Pe and G. Taramino) and the USDA/ARS (S. Kresovich, Griffin, GA, USA) has identified genetic loci using simple sequence repeats (SSRs or microsatellites). Genomic DNA clones or their sequences containing di-, tri- or tetra-nucleotide repeats were selected from either Genbank sequences or specially constructed libraries. Flanking sequences were then used to construct primers of 25-30 base pairs. PCR amplification of sorghum genomic DNA with these primers typically amplifies one predominant DNA fragment that is easily observed on agarose (4% Metaphor) gels stained with ethidium bromide. The first report of this collaboration identified seven SSR loci in the CK60/PI229828 mapping population [6]. Mapping with SSR primers developed by the USDA/ARS (S. Kresovich) will bring the total number of loci to 32 [7]. Additional efforts to develop primers for detecting SSRs in sorghum have been recently initiated (G. Hart, 1996, personal communication). Based on rather limited assessments and efforts, the utility of the primers has been limited to sorghum. In sorghum, however, SSRs do detect more DNA polymorphisms than RFLPs (M. Lee, unpublished). Therefore, this DNA marker system offers several advantages for breeding and genetics programs.

The maize DNA clones used to construct maize RFLP maps, and subsequent mapping of new maize cDNA clones, permitted comparative linkage analysis. Maize and sorghum are both diploid (2n=2x=20) but the maize nuclear genome is 3-4 times larger than the sorghum genome. Comparison of sorghum and maize RFLP maps revealed a high degree of conservation as reflected by homology, copy number, and colinearity. Over 60% of the maize clones, genomic and cDNA, produce strong hybridizations signals with sorghum. Many of the loci linked in maize (45 of 55 tested) were also linked in this sorghum population. Examples of conserved and rearranged locus orders were observed.

3. MAPPING MUTANTS AND QTL'S IN SORGHUM AND MAIZE

The same sorghum population was used to map genetic factors (mutants and QTLs) for several traits including vegetative and reproductive morphology, maturity, insect, and disease resistance. This presentation will emphasize analysis of genetic factors affecting plant height, an important character for sorghum adaptation in temperate latitudes for grain production. Evaluations of the traits were conducted with 152 F2 plants [3] and their F3 progeny (Ahnert, D.A., Pereira, M.G. and Lee, M. unpublished). Analysis of the F2 plants detected 4 unlinked QTLs for plant height accounting for 63% of the variation. The QTLs were located to linkage groups A, B, E, and H. Positive, additive genetic effects were estimated at 15-32 cm and alleles for increased stature were derived from the tall parent, PI229828. Tallness was dominant or overdominant at 3 QTLs whereas short stature was dominant at the fourth on linkage group H. Epistasis was evident for one pair of QTLs on linkage groups A and E. Analysis of F3 progeny verified all features of the QTLs detected in F2 plants and detected 2 additional QTLs for plant height in two other linkage groups, D and F.

Analysis of other traits identified several QTLs linked and unlinked to the plant height QTLs. Two QTLs for maturity (number of heat units to flowering) were identified as being

linked to two plant height QTLs. Regions initially identified as plant height QTLs were associated with several other traits (from 1 to 10). Only the QTL on linkage group E was specific for plant height. Collectively, these data and assessments of other traits at the 4 QTLs common to the F2 plants and F3 progeny indicate some of these regions correspond to loci (*dw*) previously identified on the basis of alleles with highly qualitative effects. These observations and linkage relationships will be assessed with sets of near isogenic lines for each to the 4 *dw* loci.

4. COMPARATIVE QTL MAPPING

On the basis of integrated RFLP maps, the positions and effects of sorghum and maize plant height QTLs were compared. Four of the six sorghum plant height QTLs seem to be orthologous to plant height QTLs in maize. The putative orthologous regions are (sorghum linkage group and maize chromosome) as follows: A & long arm of chromosome 1, D & chromosome 5, E & long arm of chromosome 6, and H and chromosome 9. The regions of the maize plant height QTLs also contain genetic loci defined by mutants with qualitative effects on stature such as *br1* and *an1* on chromosome 1, *na1* and *td1* on chromosome 5, *py1* on chromosome 6, and *d3* on chromosome 9. The effects of some of these maize mutants strongly resemble those of the sorghum plant height QTLs and *dw* loci. At least 3 of those maize loci, *an1, br1,* and *d3* have been tagged with transposons or cloned by various laboratories. These sequences could be used to isolate the related gene from sorghum and further assess the degree and nature of conservation between these two genomes.

Comparative QTL analysis identified evidence for several other orthologous regions. For example, a region of linkage group A (*isu033-isu123*) was strongly associated with tillering and production of lateral branches. The LOD values were 2.8 and 8.7 in F2 plants and F3 families. As indicated by comparative mapping with RFLP loci and QTLs for plant height, this region of the sorghum genome is most closely related to the long arm of chromosome 1 of maize. This region of the maize genome is the site of a genetic locus, *tb1*. The mutant phenotype at that locus is characterized by the production of many tillers and lateral branches in a manner strongly resembling the tillering QTL in sorghum. Other possible instances of orthologous QTLs included regions for maturity. These observations suggest that the conservation of the maize and sorghum genomes encompasses sequence homology, colinearity, and function despite their divergence millions of years ago and subsequent evolution in different hemispheres with contrasting ecogeographical conditions.

5. SORGHUM BREEDING AND GENETICS WITH DNA MARKERS

The genetic information and technology developed on the basis of DNA markers could be used in several facets of breeding, genetics, and other basic biological investigations of sorghum. For example, sorghum breeders in temperate regions routinely transfer short stature and photoperiod insensitivity from adapted temperate inbred lines into exotic (tropical) germ plasm through several generations of backcrossing. The goal of the backcross breeding is to maximize the recovery of the tropical germplasm with a growth habit adapted to temperate latitudes and mechanized harvesting. Once the traits necessary for temperate adaptation have been adequately recovered, the converted tropical germ plasm may be evaluated in temperate regions for traits. Similar programs are conducted for many crops (e.g. rice, maize, and wheat), traits (e.g. resistance to diseases and insects, grain quality, male sterility), and for diversifying and enhancing gene pools (e.g. adapting temperate germplasm to tropical environments and vice versa).

The U.S. Sorghum Conversion Program [8] has used a backcrossing scheme to introduce genes for short stature and photoperiod insensitivity from temperate, adapted parent to the tropical, unadapted parent. According to that scheme, adapted segregants are selected on the basis of stature and ability to flower at the temperate latitude. Self-pollinated seed from such segregants is selected and used to backcross to the recurrent, tropical parent in a tropical environment. That backcross generation is self-pollinated in the tropical location and selfed seed is sown in the temperate location to identify segregants of appropriate stature and flowering response. Typically, 4 or more cycles of backcrossing, selfing, and selection are used to convert a tropical parent.

In this instance, DNA markers could be used to eliminate generations of backcrossing and to ensure that the tropical germplasm was indeed transferred. For example, DNA fingerprints of adapted segregants could be determined and used as a selection criteria prior to backcrossing. This should be especially effective for expediting recovery of chromosome regions from the tropical parent unlinked to the genes needed for adaptation to the temperate latitudes. For example, Pereira and Lee [3] identified short-statured, adapted F2 plants with RFLP genotypes closely resembling (60-75% similarity over 111 loci) the tall or short parent. By selecting the short plants with the greatest degree of resemblance to the tall parent, the backcross conversion could be completed in fewer generations. In addition, DNA marker loci could be used to minimize linkage drag from the adapted, donor parent in the vicinity of the genes affecting adaptation. Through such a scheme, the conservation and conversion of the tropical germ plasm would be optimized. Also, breeders would have the greatest opportunity to identify new and valuable genes from exotic germ plasm.

6. ASSESSING GENETIC DIVERSITY IN SORGHUM BREEDING PROGRAMS

One consequence of modern agricultural practices, which generally emphasizes maximum productivity with acceptable quality and uniformity, has been a reduction of genetic diversity of the primary gene pool under cultivation with similar fates for the secondary and tertiary gene pools of most major crops. That trend may be exacerbated in crops such as sorghum in which F1 hybrid seed is produced using cytoplasmic male sterility. Even though the extent of the reduction may be largely unquantifiable, it is generally assumed that valuable and irreplaceable genes have been lost or ignored, plant genetic resources have been shrinking at accelerated rates, and crop-based agriculture has become more vulnerable to the vagaries of climate and associated biotic and abiotic stresses. Undoubtedly there is considerable merit, validity, and controversy associated with each point. Facts and anecdotes aside, the consequences of a narrow genetic base of major crops have been experienced sporadically throughout recent history often with significant human and economic costs. Therefore, an awareness of genetic diversity and management of crop genetic resources has been an important component of plant improvement programs.

The prospects of utilizing DNA marker technology for managing germ plasm collections has been the subject of a recent and comprehensive review [9]. How do plant breeders assess genetic diversity and relationships among elite germ plasm? Many of the methods used by germ-plasm managers have been used by plant breeders (e.g. morphology and ecogeographical data). In addition, plant breeders often have access to pedigree information, performance records (e.g. combining ability, progeny evaluation, selection and breeding history), and inferences gleaned from various mating designs. The strength of some of the methods is that they are often based on direct assessments of what the breeder needs to know about the germ plasm. Such methods will be extremely difficult to improve. However, some methods and concepts have relied on weak genetic foundations, if any. For

some plant breeding practices, that may constitute a weakness which reduces their efficiency. At least some of these deficiencies may be satisfied in part by DNA markers.

One of the most pervasive measures of genetic relationships in elite crop germ plasm has been Malecot's coefficient of coancestry (f), which provides an estimate of the degree of genetic similarity between two individuals [10]. This measure estimates the probability that two randomly drawn, homologous genes (alleles) from each of two individuals are identical by descent. The measure has been based on Mendelian inheritance and probability and has been calculated under several assumptions: 1) absence of selection, mutation, migration, and drift; 2) regular diploid meiosis; and 3) no relationship ($f = 0$) for individuals without verified common ancestors [10]. Several common features of plant breeding programs have represented departures from these assumptions: 1) intense selection, 2) drift due to small sample sizes, 3) irregular nondiploid meiosis for some crops, and 4) unknown or incorrect pedigree records. Nevertheless, this method of estimating the degree of similarity will create information each generation and has been used by crop breeding programs.

DNA markers have been used to assess large collections of elite sorghum germ plasm to determine the degree of genetic relationships and genetic diversity [4]. A set of 58 R lines, mostly from Kafir, and 47 B lines, mostly from Zera-zera or Feterita were surveyed with 104 DNA probes. The RFLP data clearly identify two different gene groups of inbred lines (pollen vs. seed parents of F1 hybrids; R and B lines, respectively) and document the high degree of genetic similarity among members of certain gene pools (e.g. B lines). On average, 3.6 RFLP-band patterns per probe were observed for R lines whereas only 3 were detected for B lines. Estimates of genetic similarity based on RFLP fingerprints were 0.67 and 0.76 for R and B lines. Cluster analysis of RFLP data further divided R lines into 2 distinct groups representing derivatives from Feterita and Zera-zera.

Similar assessments may be made without DNA markers but the methods require extensive pedigree records and generations of breeding. This requires generations of careful record keeping and is based on several assumptions mentioned previously. However, there was only a modest positive correlation between estimates of genetic similarity based on RFLP data and those based on the coancestry coefficient ($r=0.46$ and 0.43 for related sets of R and B lines). In this instance, information derived from DNA markers is substituting for the considerable time and records needed for traditional analysis of pedigrees and lineages. Also, RFLP data seem to portray genetic relationships more accurately than the methods based exclusively on the coancestry coefficient.

7. SUMMARY AND CONCLUSIONS

The state of sorghum genetics has changed dramatically within the last five years. In 1990, the sorghum genetic map consisted of nine linkage groups with each group containing 2-9 loci. Most of the 260 identified genes had not been mapped. Less than six years later, there are several genetic maps of sorghum each with ten well defined linkage groups based on RFLPs and PCR-based maps are under development. Exchange of DNA clones and mapping information has facilitated map development to the point where the genetic map contains at least 1,500 RFLP loci. Mapping resources have also improved with the development of several populations of recombinant inbred lines (RILs) suitable for mapping factors controlling resistance to biotic and abiotic stress and other important attributes. The RILs are usually available to any research group in the world. Thus, a powerful and practical means of collecting and compiling genetic information for sorghum has been established. The RFLP, QTL and genetic (mutant) maps of sorghum are now being integrated with those of

other grasses such as maize and rice. These connections should create many opportunities to transfer information, technology and materials among research groups working with these species. In many regards, in the last five years, there has been more advancement in the basic infrastructure needed for sorghum improvement than there has been in the previous fifty years.

REFERENCES

[1] LEE, M., DNA Markers and plant breeding programs. Adv. Agron. 55 (1995) 265-343.

[2] PEREIRA, M.G., LEE, M., BRAMEL-COX, P., WOODMAN, W., DOEBLEY, J., WHITKUS, R., Construction of an RFLP map in sorghum and comparative mapping in maize. Genome 37 (1994) 236-243.

[3] PEREIRA, M.G., LEE, M., Identification of genomic regions affecting plant height in sorghum and maize, Theor. Appl. Genet. 90 (1995) 380-388.

[4] AHNERT, D., AUSTIN, D., LEE, M., LIVINI, C., OPENSHAW, S.J., SMITH, J.S. C., PORTER, K., DALTON, G., Genetic diversity among elite sorghum inbred lines assessed with DNA markers and pedigree information. Crop Sci. 36 (1996) 1385-1392.

[5] DOGGET, H., Sorghum, Longmans, Green & Co, London (1988).

[6] TARAMINO, G., TARCHINI, R., FERRARIO, S., LEE, M., PE, E., Characterization and mapping of simple sequence repeats (SSRs) in Sorghum bicolor. Theor. Appl. Genet. (1997).

[7] BROWN, S.M., HOPKINS, M.S., MITCHELL, S.E., SENIOR, M.L., WANG, T.Y., DUNCAN, D.D., GONZALEZ-CANDELAS, F., KRESOVICH, S., Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum (*Sorghum bicolor* L. Moench), Theor.Appl.Genet. 93 (1996) 190-198.

[8] DUNCAN, R., BRAMEL-COX, P., MILLER, F.R., Contributions of introduced sorghum germplasm to hybrid development in the USA. In: Use of Plant Introductions in Cultivar Development-Part 1. (H.L Shands, L.E. Weisner, Eds.), Special Publ. No. 17, Crop Sci. Soc. Am., Madison, WI, USA (1991) 69-101.

[9] BRETTING, P.K., WIDRLECHNER, M.P., Genetic markers and plant genetic resource management, Plant Breeding Revs. 13 (1995) 11-86.

[10] MELCHINGER, A.E., Use of RFLP markers for analysis of genetic relationships among breeding materials and prediction of hybrid performance, (Proc. Int. Crop Sci. Symp. Ames, IA, USA, 1992), (BUXTON, D.R., SHIBLES, R., FORSBERG, R.A., BLAD, B.L., ASAY, K.H., PAULSEN, G.M., WILSON, R.F., Eds), Crop Sci. Soc. Am., Madison, WI, USA (1993) 621-628.

NEXT PAGE(S) left BLANK