Progress Toward the Positional Cloning of the Sorghum Grain Shattering Gene

Mark G. Wise[1], Stefan R. Schulze[1], Yang-Rong Lin[2], John E. Bowers[1], Hisato Okuizumi[3], Keith F. Schertz[4], Andrew H. Paterson[1]

[1]Applied Genetic Technology Center, Univ. of Georgia, Athens, GA 30602

[2]Dept. of Biology, Fu-Jen Catholic University, Hsing-Chung, Taipei 242, Taiwan, R.O.C.

[3]Dept. of Molecular Genetics, Nat. Inst. of Agrobiological Resources, Tsukuba, Ibaraki, 305-8602, Japan

[4]Dept. of Crop and Soil Sci., Texas A&M Univ., College Station, TX 77843

Abstract

Seed dispersal via disarticulation of the inflorescence, or "shattering", is an important agronomic trait contributing to significant yield loss in many common cereal crops. The shattering gene of sorghum, *Sh1*, has previously been mapped to a single genetic locus on sorghum linkage group C (Paterson, et al., Science 269:1714-1718). Fine resolution mapping of 370 F_2 individuals, including progeny testing, showed that one RFLP marker co-segregated with the shattering phenotype. The locus is flanked by two recombination events on one side and one on the other side that delineates it from adjacent markers. A 6-genome-equivalent *S. propinquum* BAC library was fingerprinted and used to construct a physical map of the shattering region. Two of the closely linked markers were present on a single 220kb BAC. This BAC has been completely sequenced, revealing some candidate genes that are under investigation.

Introduction

"Shattering", or seed dispersal via disarticulation of the mature inflorescence, in wild grass species (family: Poaceae) is an important mechanism for plant propagation. However, in crops shattering causes grain loss during either manual or automated harvesting. Repression of the shattering habit was likely a key event in the domestication of the world's major cereal crops on several continents. QTL studies have suggested that the control of the shattering trait range from simple to complex in related grasses. *Sorghum* appears to a favorable candidate for investigating the genetic basis for shattering, since one locus (*Sh1*) explains virtually all of the phenotypic variance in a cross between wild and domesticated sorghum (Paterson, et al., Science 269:1714-1718). In contrast at least ten and six QTLs affect shattering in maize and rice, respectively (Paterson, et al., Science 269:1714-1718; Kinoshita & Takahashi, J. Fac. Agric. Hokkaido Univ. 65:1-61). This work focused initially on the fine mapping of shattering with RFLP probes, and chromosome walking to the *Sh1* locus in *Sorghum propinquum*. Based on the combined physical/genetic map, there was a high likelihood that the locus was present on a single large BAC. This BAC was sequenced and a list of genes predicted to be present on the BAC is presented here.

Materials and methods

High resolution genetic mapping. For fine mapping of the chromosome segment associated with *Sh1*, substitution mapping was used (Paterson et al., Genetics 124: 735-742). DNA markers that mapped directly, or inferred by comparative data to locate close to *Sh1* were applied to a panel of recombinants in the region, and they defined most recombination events in the vicinity of shattering in a mapping population of 370 F_2 individuals (*S. bicolor* x *S. propinquum*). DNA extraction, restriction enzyme digestion, blotting, probing, hybridization and autoradiography were all done by standard methodology.

Physical mapping. A *S. propinquum* BAC library with an average insert size of 126 kb, which provided 6.6x coverage of the sorghum genome, was screened with DNA markers closely linked to *Sh1*. BACs were gridded using a Q-BOT robot (Genetix) and hybridized following standard techniques. BACs that hybridized markers in the region were fingerprinted via *Hin*DIII digestion, and used to construct a physical contig around *Sh1* with the FPC program (C. Soderlund, Cambridge, UK).

BAC sequencing. BAC 39E21 DNA was isolated using the QIAGEN Large Construct Kit, sheared with sonication, end-repaired and subcloned (avg. size 1.5 kb) into pPCR-Script (Stratagene) all following standard techniques. Subclones were sequenced using a 96-capillary ABI 3700 DNA sequencer. Sequence analysis and assembly was performed using PHRED/PHRAP/CONSED software (P. Green, Univ. of Washington). Gene prediction was by the GeneMark (M. Borodovsky, Georgia Tech) and Genscan (C. Burge, MIT) web servers.

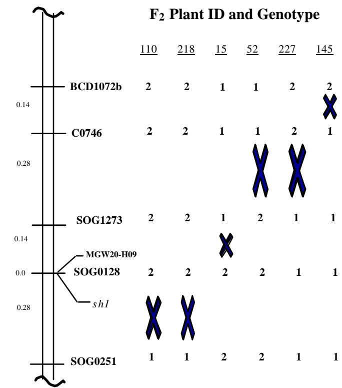

FIG. 1. Fine-scale genetic mapping of *Sh1*. Markers in the *Sh1* region were applied to a panel of recombinants in the target area and defined the critical recombination events surrounding *Sh1* in a mapping population of 370 F_2 individuals (740 informative gametes). The six columns of data are marker genotypes for the six individual plants known *a priori* to be recombinant in the region. Genotypes are 1=*S. bicolor* homozygote, 2=heterozygote. No recombinants in the region were *S. propinquum* homozygotes (genotype=3).

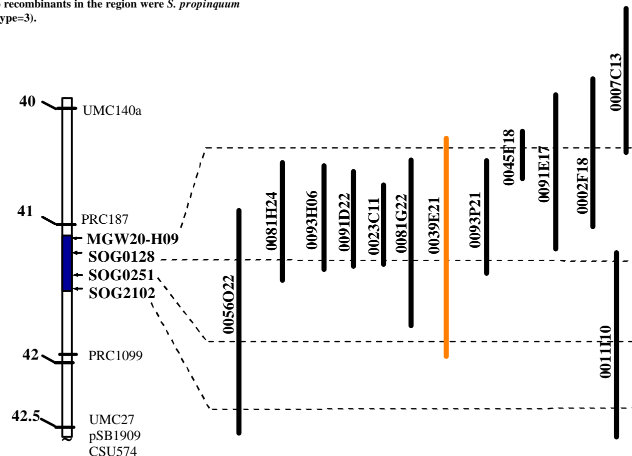

FIG. 2. Integrated genetic and physical map in Lg C-shattering (*Sh1*) gene region of *Sorghum*. BACs in the region were aligned on the basis of fingerprint and hybridization data. Orange color indicates sequenced BAC.

Results

Fine Genetic Mapping. Interval mapping showed that *Sh1* co-segregated with one marker SOG0128, and located between marker SOG0251 and SOG1273, in a genetic interval of 0.42 cM. Two recombinants occurred between SOG0251 and *Sh1*, giving an estimated genetic distance of 0.28 cM, and one recombinant was observed between *Sh1* and SOG1273, yielding a genetic distance of 0.14 cM (Fig. 1).

Physical Mapping. Thirteen BACs that hybridized markers in the region formed one contig (Fig. 2). One BAC, 39E21, spanned a large part of the contig with SOG0251 at one end, and the *Sh1* co-segregating marker SOG0128 near the middle.

Sequencing of 39E21. Sequencing revealed the BAC to be 220 kb in size. Table 1 shows the BLASTP scores of some of the translated genes predicted to be present on the BAC. Many of the highest hits are to genes found on retrotransposons and similar repetitive elements, as well as hypothetical genes found in the rice and *Arabidopsis* genomes. However, some candidate genes appear to code for proteins homologous to known protein families.

Table 1. List of genes predicted to be present on BAC 39E21. BLASTP scores of the predicted amino acid sequence are listed in descending order. *Present in two or more copies on the BAC

ID	Best Hit	Organism	BLASTP score (E-value)
Genscan31_2	DnaJ protein homolog ZMDJ1	Maize	0
Genemark31_4	Beta-glucosidase	Arabidopsis	0
Genscan28_1	Putative polyprotein*	Rice	0
Genscan31_5	Putative reverse transcriptase*	Rice	0
Genemark28_21	GAG-POL precursor*	Rice	0
Genemark28_15	Polyprotein*	Sorghum	1.00e-179
Genemark28_16	Sulfate transporter ST1	Maize	1.00e-137
Genscan33_1	RNA polymerase II Beta subunit	Tomato	1.00e-132
Genscan33_2	T27G7.10 (probable phosphoprotein phosphatase)	Arabidopsis	1.00e-113
Genemark30_2	Unknown protein	Arabidopsis	1.00e-112
Genemark28_23	22 kDa kafirin cluster; Ty3-Gypsy retroelement	Rice	1.00e-109
Genemark30_6	Argonaute-like protein	Arabidopsis	1.00e-100
Genemark28_17	Retrotransposon TNP2	Rice	8.00e-95
Genemark28_10	Hypothetical protein F4F15.170	Arabidopsis	2.00e-71
Genscan28_6	Hypothetical protein F26G5.100	Arabidopsis	8.00e-63
Genemark33_3	Hypothetical protein	Rice	9.00e-53
Genemark28_4	Hypothetical protein	Rice	2.00e-51
-many other hypothetical proteins-		Arabidopsis/Rice	2.00e-51 and lower
Genemark30_4	Methyl-binding domain protein	Maize	8.00e-43
Genemark28_6	Transcription factor	Vitis	5.00e-18
Genscan31_3	Uclacyanin I	Arabidopsis	5.00e-16
Genscan31_1	Similar to SF16 protein	Arabidopsis	7.00e-14
Genemark31_3	Putative zinc finger protein	Rice	6.00e-10
Genscan31_6	Similar to RING-H2 finger	Arabidobsis	7.00e-09

Discussion and Future Aims

On the basis of genetic/physical mapping, we determined that there was a high likelihood that the *Sh1* gene resided on a single large BAC, 39E21 (Fig. 2). Sequencing of the BAC revealed several putative genes (Table 1), including some homologous to members of families involved in developmental regulation, like the Argonaute/Zwille family (Lynn, et al. Development 126: 469-481). However, at this point we cannot be 100% certain that the *Sh1* locus resides on this BAC. While the "left" end of the contig is delineated by SOG0251, the "right-most" BAC end (subclone MGW20-H09) co-segregates with shattering (Figs. 1 & 2). We were unable to extend the BAC contig at satisfactory stringency to include BACs hybridizing marker SOG1273. Therefore, we are currently focusing on closing the contig by searching for polymorphic markers on the "right-most" BAC in the contig, 07C13 (Fig. 2). Once the contig is closed and fully sequenced, we intend to transform mutant *S. bicolor* (recessive non-shattering genotype) with the most promising candidate genes in an attempt to complement the non-shattering phenotype.