Theor Appl Genet (1998) 97: 756–761 © Springer-Verlag 1998

Zachary W. Shappley · Johnie N. Jenkins
William R. Meredith · Jack C. McCarty, Jr.

An RFLP linkage map of Upland cotton, *Gossypium hirsutum* L.

Received: 31 March 1998/ Accepted: 29 April 1998

Abstract Ninety-six $F_2.F_3$ bulked sampled plots of Upland cotton, *Gossypium hirsutum* L., from the cross of HS46 × MARCABUCAG8US-1-88, were analyzed with 129 probe/enzyme combinations resulting in 138 RFLP loci. Of the 84 loci that segregated as co-dominant, 76 of these fit a normal 1 : 2 : 1 ratio (non-significant chi square at $P = 0.05$). Of the 54 loci that segregated as dominant genotypes, 50 of these fit a normal 3: 1 ratio (non-significant chi square at $P = 0.05$). These 138 loci were analyzed with the MAPMAKER\EXP program to determine linkage relationships among them. There were 120 loci arranged into 31 linkage groups. These covered 865 cM, or an estimated 18.6% of the cotton genome. The linkage groups ranged from two to ten loci each and ranged in size from 0.5 to 107 cM. Eighteen loci were not linked.

Key words *Gossypium hirsutum* L. · Restriction fragment length polymorphism · Linkage map · Upland cotton · Linkage

Communicated by A. L. Kahler

Z. W. Shappley
Department of Plant and Soil Sciences, Mississippi State University, Mississippi State, MS 39762, USA

J. N. Jenkins (✉)
USDA, ARS, Crop Science Res. Laboratory, Mississippi State, MS 39762, USA
Tel.: 601 323-2230
Fax: 601 323-0915

W. R. Meredith
USDA, ARS, Cotton Physiology and Genetics Res. Unit, Stoneville, MS 38776, USA

J. C. McCarty, Jr.
USDA, ARS, Crop Science Res Lab., Mississippi State, MS 39762, USA

Introduction

Molecular markers provide increased potential for gathering information vital to the improvement of crop species. Paterson et al. (1991) noted that the primary uses for molecular markers include loci mapping, linkage studies, and employment in breeding programs. Restriction fragment length polymorphisms (RFLPs) are one type of molecular marker frequently used today.

In several crops RFLPs have been employed to further the knowledge of parental relationships, to map genes for insect and disease resistance, and for the identification of quantitative trait loci (Paterson et al. 1991; Dudley 1992; Giese et al. 1993; Schon et al. 1993; Lee et al. 1996). Molecular markers have two major advantages, they are very little influenced by the environment and heterozygous genotypes within specific plants in a given generation can be detected. These make molecular markers particularly valuable to plant scientists.

Meredith (1992), in a study of heterosis and varietal origins, reported the first RFLP evaluation in Upland cotton, *Gossypium hirsutum* L. To determine the amount of polymorphism for RFLP markers in Upland cotton, 68 individual lines from diverse genetic backgrounds were analyzed with 75 probe/enzyme combinations resulting in 179 polymorphic fragments. This large number of fragments provided sufficient evidence that genetic analyses of this type were feasible for Upland cotton (Shappley et al. 1993; Shappley 1994). Five linkage groups with 12 loci provided the first genetic linkage maps constructed with molecular markers in a cross of Upland cotton (Shappley 1994; Shappley et al. 1996). The objective of the present study was to establish a more-detailed linkage map of RFLP markers in Upland cotton.

Materials and methods

Choice of parents and methodology for successive generations

The breeding line MARCABUCAG8US-1-88 and the cultivar HS46 were chosen as parents for the present study. A consideration of the pedigrees of these two lines indicates that they are not closely related (Calhoun et al. 1997). The agronomic and fiber properties of these two lines are also very diverse (Shappley et al. 1996).

A cross of these lines was made in the summer of 1991 in the field at Mississippi State, Miss. In 1992 F_1 plants of this cross were grown in the field at Mississippi State University and individual plants were self-pollinated and their seed harvested. Nine F_1 plants were analyzed to determine if restriction fragment patterns were different among the F_1s (Shappley 1994). Some variability was observed among the F_1 plants. Therefore F_2 seed from one individual F_1 plant was used for this study.

One hundred F_2 seed from the selected F_1 individual were planted in the greenhouse in the winter of 1992, and 96 plants grew and were allowed to self-pollinate. This planting was the beginning of successive generations of F_2-derived families. These $F_2.F_3$ seed were planted in single-row plots, 12 m in length, at Mississippi State, Miss., in 1994. We did not have a complete planting of F_3 families in 1993 because we did not have enough seed from some F_2 plants; thus the complete planting of 96 F_3 families was made in 1994. Each plot of plants was a family, made up of seed from an individual F_2 plant. Bulk leaf-tissue samples were taken from these F_3 rows in 1994 and analyzed with RFLP probes. The F_3 plants were then self-pollinated in Mississippi in 1994 and the F_4 seed were sent to the winter nursery in Mexico in the winter of 1994–95 for self-pollination. In the spring of 1995, two-row plots of F_5 seed were planted and agronomic and fiber data were collected for a QTL study.

Probe construction

Biogenetic Services Incorporated constructed the initial cotton cDNA library using leaf material collected from six diverse Upland cotton cultivars. Total cellular RNA was isolated from leaf tissue of all six cultivars and prepped according to the methods of MacDonald et al. (1987). Poly(A)-RNA was purified and isolated according to the methods of Aviv and Leder (1972). The preparations were pooled and oligo-(dT) chromatography performed to isolate polyadenylated RNA. The poly(A)-RNA was then used as a template for double-stranded cDNA (ds-cDNA) synthesis (Gubler and Hoffman 1983). A dT-tailed *Not*I primer/adaptor oligonucleotide was used to prime a reverse transcriptase enzyme reaction. *Eco*RI adaptors were then joined to the 5′ ends of the ds-cDNA. Ligation and transformation techniques were followed according to Hanahan (1983). Subsequent to *Not*I digestion and size-fractionaction steps, the cDNa was inserted into the GEM-1zf(-) vector (Promega). Multiple copies of the probe were generated via PCR (Saki et al. 1988). The selection of transformants was based on Amp^R and Lac^-. Radiolabelling was done by random priming (Feinberg and Vogelstein 1983). The given references of methodology were slightly modified or adapted for use by Biogenetic Services Inc. (unpublished).

RFLP analyses

RFLP analyses were conducted at Biogenetic Services Inc., Brookings, South Dakota, via a purchase contract for the services. These analyses were conducted using bulk samples of leaf tissue from individual F_3 family single-row plots, collected at Mississippi State in 1994.

DNA was isolated and purified by sedimentation equilibrium in CsCl density gradients. Techniques for DNA isolation and other RFLP methods are found in "Current protocols in molecular biology" (Ausbel 1987) with slight modifications by Biogenetic Services, Inc. (unpublished). Either *Eco*RI or *Eco*RV restriction enzymes were used in the digest of the sample DNA, depending upon the probe employed. Gel electrophoresis was conducted with a TAE buffer. Southern blots were made on nitrocellulose membranes and the resulting complex of fragments was subject to hybridization techniques (Budowle and Baechtel 1990). After washing the membranes to remove the unhybridized probe solution, the membranes were placed on Kodak AR X-ray film and incubated for up to 14 days depending on the optimal exposure time. Restriction fragments were digitized using the DIGIGEL software package with phage lambda DNA as a standard. The autoradiographs and digitized data were then visually scored for the presence or absence of all fragments reported.

Fragments were entered as present or absent with their corresponding size in kilobases (kb). Allelic patterns were then coded as co-dominant or $+/-$ genotypes. If both alleles were visualized by the particular probe/enzyme combination, the locus was considered a co-dominant genotype. If only one allele was identified, the locus was considered a $+/-$ genotype. Eighty four-co-dominant and 54 $+/-$ loci were cataloged. Chi-square goodness-of-fit tests were conducted to determine the fit to the expected genetic ratios of the genotypes.

Genetic linkage maps were constructed using the MAPMAKER\EXP 3.0 (Lander et al. 1987) program. The "group" command was used to establish a subset of linked markers. The "compare" and "three point" commands were employed to order the markers within linkage groups. The parameters used for the detection of linkage groups during the analyses were a LOD score of 3.00 or greater and a genetic distance of 50 cM.

Results

In preliminary work the parents and F_1 plants were evaluated with 73 probe enzyme combinations. The 9 F_1 plants were scored for 53 fragments that were polymorphic in the analysis of DNA from bulked leaf samples of each of the parents. Of the 53 polymorphic fragments in the parents, only five fragments were not present in all nine F_1 plants. We scored a total of 281 total fragments among the parents and the nine F_1 plants. Only seven fragments were not present in all nine F_1 plants.

In the present study 96 $F_2.F_3$ bulk-sampled families were analyzed with 129 probe/enzyme combinations resulting in 138 loci. Thus, some probe/enzyme combinations revealed more than a single locus. Single locus chi-square values showed that the majority of progeny arrays fit the expected 3 : 1 (dominant) or 1 : 2 : 1 (co-dominant) genotypic ratios. There were 84 co-dominant loci with 76 of these segregating normally with a 1 : 2 : 1 ratio (non-significant chi-square at $P = (0.05)$ and 54 dominant loci with 50 segregating normally in a 3 : 1 ratio (non-significant chi-square at $P = (0.05)$.

Thirtyone linkage groups with 120 of the 138 loci were established with the use of MAPMAKER\EXP. The number of loci associated with a particular linkage group ranged from two to ten while 18 of the loci were not linked. Map distances between loci ranged from 0.0 to 45.7 cM as determined by two-point analyses and

measured in probability by LOD scores. The average distance between loci was 7.0 cM.

Linkage groups were determined using multi-point analyses and the orders are shown in Fig. 1. In most instances gene order was well established. In a few cases, with very closely linked loci, the gene order may not be precise. The total distance covered by the linkage groups ranged from 107.5 cM (10 loci) to 0.5 cM (2 loci). The overall map distance covered by all linkage groups was 865 cM. Thus, the 120 loci covered 18.6% of the minimum map distance which has been estimated to be approximately 4660 cM for the cotton genome (Reinisch et al. 1994).

Four of the linkage groups were previously identified. Linkage group 3 was previously identified as linkage group B, and 10 as group A (Shappley 1994; Shappley et al. 1996). Linkage group 18 is what Shappley (1994) called group C and Shappley et al. (1996) called group D. Linkage group 21 is what Shappley (1994) called group D and Shappley et al. (1996) called group C. The genetic distances in the present work are sometimes different from earlier reports and this should be expected as markers are added to the linkage groups. Shappley (1994) could not distinguish the gene order of markers C34F5RV and C26D5RV in his linkage group A with only four markers, so he presented two alternative orders, A1 and A2, with respect to the four markers in this linkage group. The present work, with nine markers, established the order for these two markers with respect to other markers in linkage group 10. The order is that shown in linkage group A2 of Shappley et al. (1996). The agreement of these four linkage groups in the present work with those of previous groups adds additional markers to two linkage groups and provide confirming evidence for Shappley's (1994) work.

Twelve of one hundred and thirty eight loci showed distorted segregation ratios (Table 1). Two linkage, 10 and 31, had only one locus each with abnormal segregation; thus, these linkage groups were well established. In linkage group 5 both loci showed abnormal segregation, so that this may not be a well-established linkage group. In linkage group 11, two of five loci showed abnormal segregation ratios. In linkage group 19, four

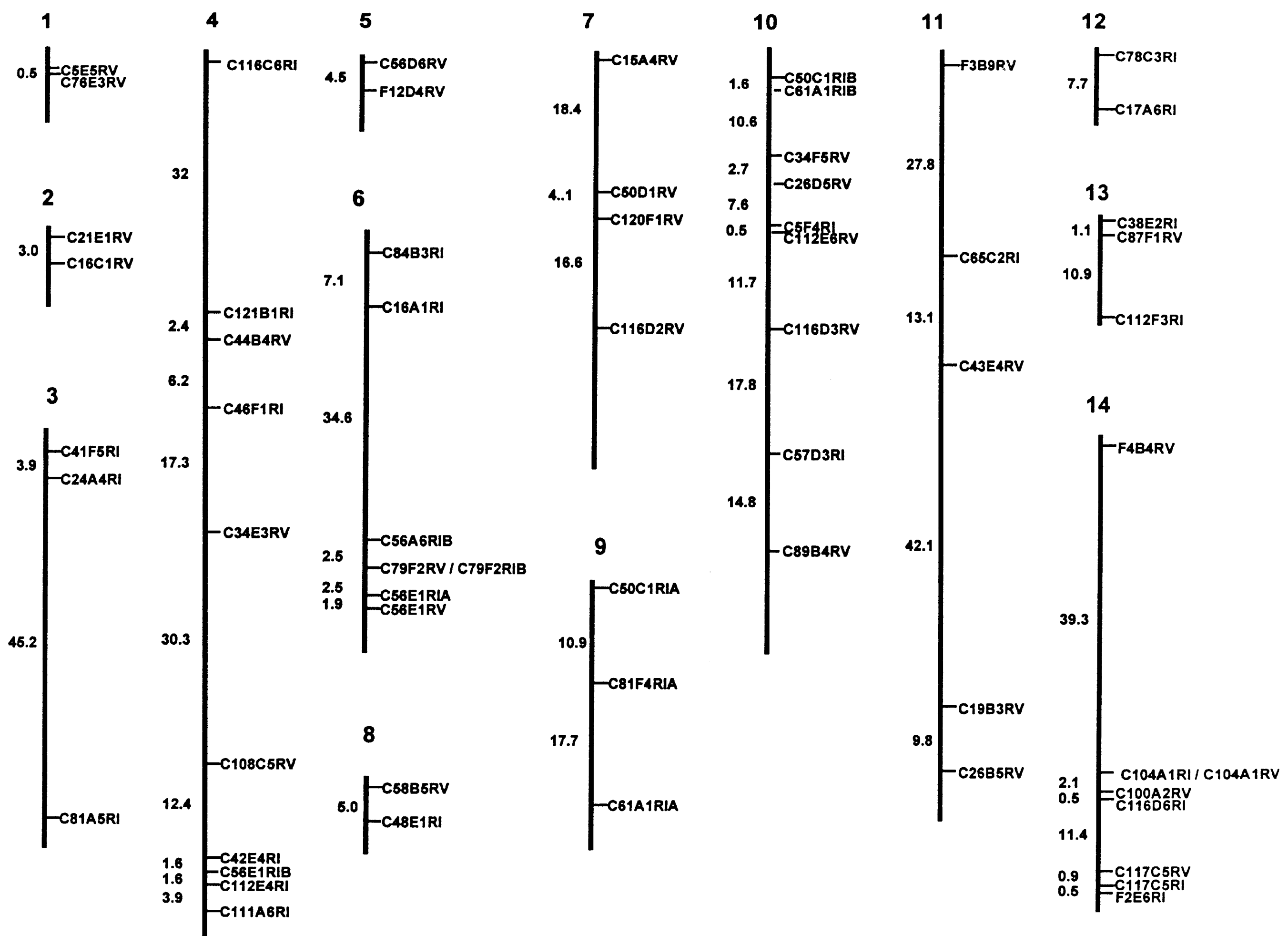

Fig. 1 See page 759 for legend

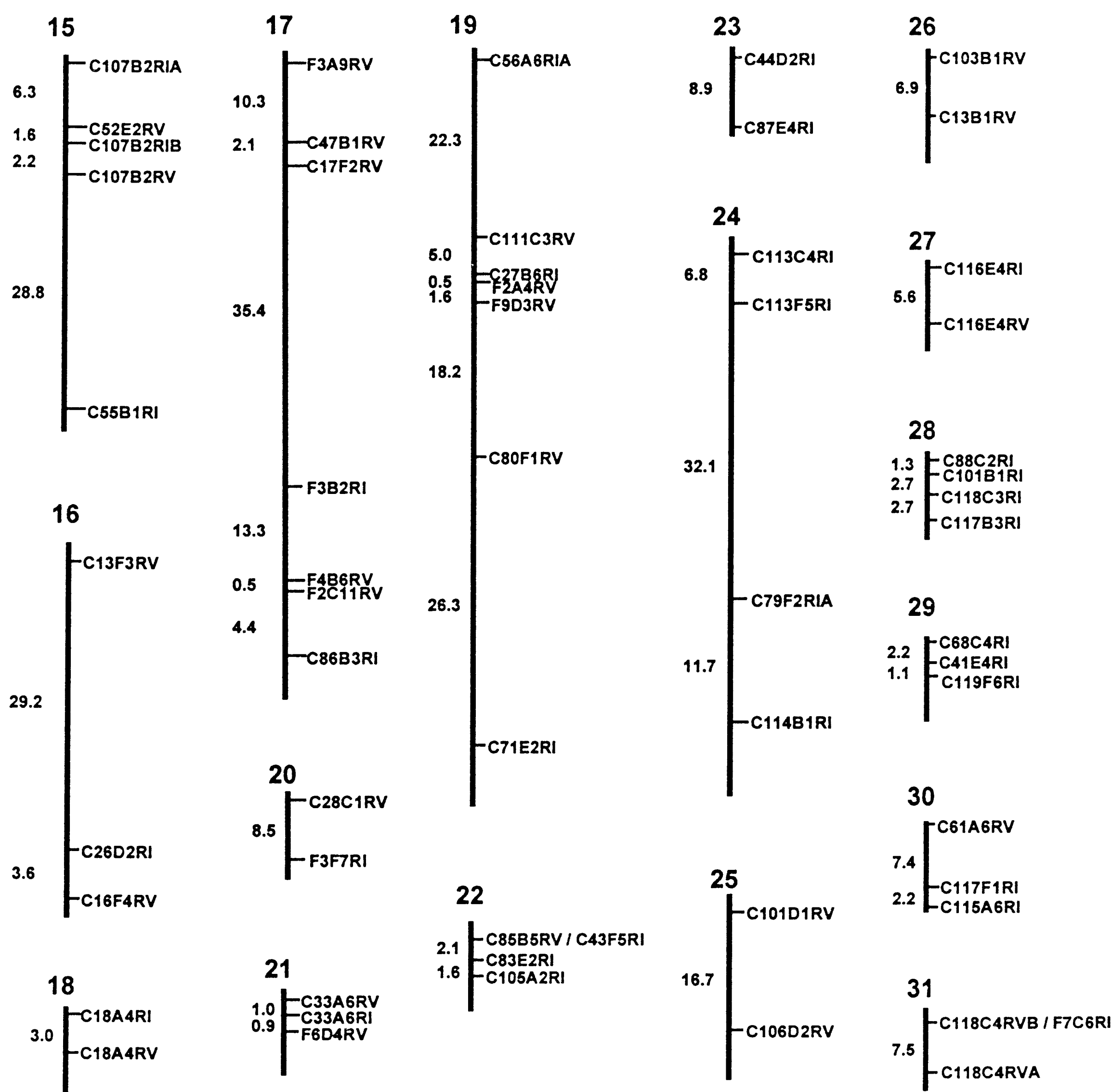

Fig. 1 Thirtyone linkage groups of Upland cotton (*G. hirsutum* L.) containing 120 loci, based on RFLP analyses of a 96-family population of HS46 × MARCABUCAG8US-1-88. Map distances between adjacent markers are in centiMorgans. This linkage map was constructed with the aid of MAPMAKER/EXP

of seven loci showed abnormal segregation; thus this linkage group may not be well established. Conversely, this abnormal segregation in linkage group 19 may have a cytological or biological cause. Thus, there were 3 of 31 linkage groups that showed abnormal segregation for more than one locus. The 12 of the 138 loci with distorted segregation ratios were perhaps due to such things as gametophyte selection, genetic drift, or cytological attributes. Recent research in our laboratory with cytogenetic deficiency lines has shown abnormal segregation for three of the 12 markers which we found to be showing abnormal segregation in the present study. These are markers C56D6RV in linkage group 5, and F2A4RV and C80F1RV in linkage group 19. Thus, there may be some cytological reason for the abnormal segregation.

In our present data, five of eight co-dominant loci with abnormal segregation showed an excess of heterozygotes and of homozygotes for alleles from the MAR parent. Three of eight co-dominant markers segregating abnormally showed a deficiency of heterozygotes with homozygotes for alleles from HS 46 higher than expected in two of these three markers (Table 1). In the four dominant markers for which we could only score one allele, the data indicated that this allele was probably from the HS 46 parent. We do not know if the abnormal segregation is related to gamete selection or to some other cause. Four of the 12 loci were in linkage group 19 and two each were in linkage groups 5 and 11. Perhaps cotton has several regions of chromosomes that are only distantly related. In our study the two

Table 1 Molecular markers with segregation ratios significantly different from the 1 : 2 : 1 or 3 : 1 expected ratios at the 0.05 level

Linkage group	Loci with distorted ratios	Class 1 2 3 4[a]	Chi-square value	No. loci in linkage group
5	C56D6RV	29 54 13 0	6.83 *	2
5	F12D4RV	23 60 13 0	8.08 *	
10	C116D3RV	39 36 21 0	12.75 *	9
11	C43E4RV	25 34 37 0	11.17 *	5
11	F3B9RV	37 0 0 59	9.39 *	
19	C27B6R1	27 56 13 0	6.75 *	7
19	F2A4RV	27 57 12 0	8.06 *	
19	F9D3RV	28 56 12 0	8.00 *	
19	C80F1RV	38 0 0 58	10.89 *	
31	C118C4RVA	29 30 37 0	14.83 *	3
Unlinked	C40E1RV	47 0 0 49	29.39 *	18
Unlinked	C102C5R1	37 0 0 59	9.39 *	

* Significant at the $P = 0.05$ level with the chi-square test

[a] 1 = homozygous for MAR allele, i.e. AA; 2 = heterozygous for MAR and HS 46 alleles, i.e. A/B; 3 = homozygous for HS 46 allele, i.e. BB; 4 = not homozygous for the MAR allele, i.e. the plant is either BB or A/B

parental lines are not closely related. HS 46 traces to a cross of AZ7209 × Acala 90 (Calhoun et al. 1997). The parent MARCABUCAG8US-1-88 is from the Multiple Adversity Resistance (MAR) program at Texas Agricultural Experiment Station.

Abnormal segregation is not unusual with molecular markers. Schon et al. (1993) reported that 18 of 87 markers in an F_2 population in corn showed significant deviation from the expected 1 : 2 : 1 ratios. They also reported an overall excess of heterozygotes. They found distorted segregation ratios for alleles on all ten corn chromosomes. Saha (1989) found an excess of heterozygotes in his analysis of isozyme alleles in cotton. In F_2-derived inbred lines in corn, Bernardo et al. (1997) reported that the frequency of alleles clustered around 0.05 but the frequency ranged from 0.248 to 0.801 indicating a wide variability in the proportion of the genome derived by these inbreds from one of the parents in a biparental cross. In BC_1-derived inbreds the allele frequency clustered around 0.75 with a range of 0.54 to 0.989 (Bernardo et al. 1997). Thus, our observations of 12 loci segregating abnormally should not be surprising.

Although 18.6% of the cotton genome is represented by the 120 mapped loci, further study with additional molecular markers is needed to confirm the precise locus positions within the linkage groups; however, our best estimate has been given for each position. The close proximity of some loci made it difficult to determine precise map orders within a few linkage groups. We had several markers which were closely linked (3 cM or less). We do not know why this occurred; however, it may relate to the diversity of the two parents or to the derivation of the probes from a leaf cDNA library involving several cultivars. Map orders will continue to evolve as additional markers are added to the overall map.

These studies increase the size of the previous RFLP linkage map of this population; however, further mapping work will be needed to resolve closely linked marker orders and for the assignment of linkage groups to specific cotton chromosomes. The addition of more polymorphic RFLP marker loci will also enhance the depth and resolution of the Upland cotton molecular map. Research is under way in our laboratory to assign the linkage groups to specific chromosomes using germplasm lines with known chromosome identities.

Acknowledgements We thank Dr. Alex L. Kahler, Dr. Eugene Butler III, and Dee French for conducting the RFLP assays at Biogenetic Services Incorporated, Brookings, SD 57006, USA.

References

Ausbel F (ed) (1978) Current protocols in molecular biology. Green Publishing Company, New York

Aviv H, Leder P (1972) Purification of biologically active globin messenger RNA by chromotography on oligothymidylic acid-cellulose. Proc Natl Acad Sci USA 69 : 1408–1412

Bernardo R, Murigneux A, Maisonneuve JP, Johnsson C, Karaman Z (1997) RFLP based estimates of parental contribution to F_2- and BC_1-derived maize inbreds. Theor Appl Genet 94 : 652–656

Budowle B, Baechtel FS (1990) Modifications to improve the effectiveness of restriction fragment polymorphism typing. Appl Theor Electrophoresis 1 : 181–187

Calhoun DS, Bowman DT, May OL (1997) Pedigrees of Upland and Pima cotton cultivars released between 1970 and 1995. Bull 1069, Miss. Agric and For Exp Sta

Dudley JW, Saghai-Maroof A, Rufener GK (1992) Molecular-marker information and selection of parents in corn breeding programs. Crop Sci 32 : 301–304

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to highly specific activity. Anal Biochem 137 : 266 –267

Giese H, Holm-Jensen AG, Jensen J (1993) Localization of the *laevigatum* powdery mildew resistance gene to barley chromosome 2 by the use of RFLP markers. Theor Appl Genet 85 : 897–900

Gubler U, Hoffman BJ (1938) A simple and very effective method for generating cDNA libraries. Gene 25 : 263–269

Hanahan D (1983) Studies on the transformation of *Escherichia coli* with plasmids. J Mol Biol 166 : 557–580

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

Lee SH, Bailey MA, Mian MAR, Carter Jr. TE, Ashley DA, Hussey DA, Parrott WA, Boerma HR (1996) Molecular markers associated with soybean plant height, lodging, and maturity across locations. Crop Sci 36 : 728–735

Meredith WR (1992) RFLP association with varietal origin and heterosis. In : Herber D (ed) Proc Beltwide Cotton Prod Res Conf Nashville, TN, 6–10 Jan. National Cotton Council of America, Memphis, Tennessee, p 607

MacDonald, RJ, Swift GH, Przybyla AE, Chirgwin JM (1987) Isolation of RNA using guanidinium salts. Methods Enzymol 152 : 219–223

Paterson AH, Tanksley SD, Sorrells ME (1991) DNA markers in plant improvement. Adv Agron 46 : 39–90

Reinisch MJ, Doug J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense* : chromosome organization and evolution in a disomic polyploid genome. Genetics 138 : 829–847

Saki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239 : 47–491

Saha Sukumar (1989) Isozyme polymporhism in *Gossypium*. PhD dissertation submitted to Department of Plant and Soil Science, Texas A and M University, College Station, Texas

Schon CC, Lee M, Melchinger AE, Guthrie WD, Woodman WL (1993) Mapping and characterization of quantitative trait loci affecting resistance against second generation European corn borer in maize with the aid of RFLPs. Heredity 70 : 648–659

Shappley ZW (1994) Restriction fragment length polymorphisms in cotton (*Gossypium hirsutum L.*) : feasability of use, diversity among plants within a line, and establishment of molecular markers and linkage groups among two F_2 populations. MS thesis. Mississippi State University, Mississippi State, Mississippi

Shappley ZW (1996) Construction of RFLP linkage groups and mapping of quantitative trait loci in Upland cotton (*Gossypium hirsutum L.*). PhD dissertation, Department of Plant and Soil Sciences, Mississippi State University, Mississippi State, Mississippi

Shappley ZW, Jenkins JN, McCarty Jr JC (1993) Potential for RFLP mapping in cotton. In : Herber D (ed) Proc Beltwide Cotton Prod Res Conf, New Orleans, LA, 10–14 Jan. National Cotton Council of America, Memphis, Tennessee, p 610

Shappley ZW, Watson Jr CE, Jenkins JN, Robinson MR, Tang B (1995) Use of RFLPs in prediction agronomic performance. In : Herber D (ed) Proc Beltwide Cotton Prod Res Conf, San Antonio, TX, 4–7 Jan. National Cotton Council of America, Memphis, Tennessee, p 531

Shappley ZW, Jenkins JN, Watson Jr CE, Kahler AL, Meredith Jr WR (1996) Establishment of molecular markers and linkage groups in two F_2 populations of Upland cotton. Theor Appl Genet 92 : 915–919