

Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton)

(quantitative variation/cotton fiber quality/crop evolution/intergenomic interactions/comparative quantitative trait locus analysis)

CHUN-XIAO JIANG*, ROBERT J. WRIGHT*†, KAMAL M. EL-ZIK†, AND ANDREW H. PATERSON*‡

*Plant Genome Mapping Laboratory, and †Multi-Adversity Resistance Cotton Breeding Program, Department of Soil and Crop Science, Texas A & M University, College Station, TX 77843-2474

Communicated by Jack Harlan, University of Illinois at Urbana-Champaign, Urbana, IL, October 29, 1997 (received for review September 18, 1997)

ABSTRACT A detailed restriction fragment length polymorphism map was used to determine the chromosomal locations and subgenomic distributions of quantitative trait loci (QTLs) segregating in a cross between cultivars of allotetraploid (AADD) *Gossypium hirsutum* (“Upland” cotton) and *Gossypium barbadense* (“Sea Island,” “Pima,” or “Egyptian” cotton) that differ markedly in the quality and quantity of seed epidermal fibers. Most QTLs influencing fiber quality and yield are located on the “D” subgenome, derived from an ancestor that does not produce spinnable fibers. D subgenome QTLs may partly account for the fact that domestication and breeding of tetraploid cottons has resulted in fiber yield and quality levels superior to those achieved by parallel improvement of “A” genome diploid cottons. The merger of two genomes with different evolutionary histories in a common nucleus appears to offer unique avenues for phenotypic response to selection. This may partly compensate for reduction in quantitative variation associated with polyploid formation and be one basis for the prominence of polyploids among extant angiosperms. These findings impel molecular dissection of the roles of divergent subgenomes in quantitative inheritance in many other polyploids and further exploration of both “synthetic” polyploids and exotic diploid genotypes for agriculturally useful variation.

Most angiosperm (flowering plant) genomes are thought to have incurred one or more polyploidization events (1, 2). Geneticists have long debated whether this simply reflects promiscuity of plants or whether a selective advantage is conferred by polyploid formation (3, 4). Among the best-studied polyploids are many of the world’s leading crops, including cotton, wheat, oat, soybean, peanut, canola, tobacco, coffee, and banana, each of which evolved by the joining of divergent genomes in a common nucleus (5).

The evolution of the genus *Gossypium* (cotton) has included a very successful experiment in polyploid formation. World cotton commerce of about \$20 billion annually is dominated by improved forms of two (among five extant) “AD” tetraploid ($2n = 4x = 52$) species, *Gossypium hirsutum* L. and *Gossypium barbadense* L. Tetraploid cottons are thought to have formed about 1–2 million years ago, in the New World, by hybridization between a maternal Old World “A” genome taxon resembling *Gossypium herbaceum* ($2n = 2x = 26$) and paternal New World “D” genome taxon resembling *Gossypium raimondii* (6) or *Gossypium gossypoides* (7) (both $2n = 2x = 26$). The antiquity of this New World event precludes human involvement in polyploid formation.

Wild A genome diploid and AD tetraploid *Gossypium* taxa produce spinnable fibers that were a likely impetus for do-

mestication (8, 9). Domesticated tetraploid cottons existed in the New World by 3500–2300 B.C. (10) and have been widely distributed by humans throughout the world’s warmer latitudes. Domesticated A genome diploids existed in the Old World by 2700 B.C. (11), and one (of only two extant) species, *Gossypium arboreum*, remains intensively bred and cultivated in Asia. Its close relative and possible progenitor, the other extant A genome diploid species *G. herbaceum* also produces spinnable fiber.

Although the seeds of D genome diploids are pubescent, none produce spinnable fibers (12). There is no evidence that domestication of D genome *Gossypium* taxa has ever been attempted, although their geographic distribution overlaps that of several wild tetraploids. No taxa from the other recognized diploid *Gossypium* genomes (B, C, E, F, and G) have been domesticated.

Intense directional selection by humans has consistently produced AD tetraploid cottons that have superior yield and/or quality characteristics compared to the A genome diploid cultivars. Selective breeding of *G. hirsutum* (AADD) has emphasized maximum yield, whereas *G. barbadense* (AADD) is prized for its fibers of superior length, strength, and fineness. Side-by-side trials of 13 elite *G. hirsutum* genotypes and 21 *G. arboreum* diploids (AA) adapted to a common production region (India) show average seed cotton yield of $1,135 \pm 90$ kg/ha for the tetraploids, a 30% advantage over the 903 ± 78 kg/ha of the diploids, at similar quality levels (13). Such an equitable comparison cannot be made for *G. barbadense* and *G. arboreum*, because they are bred for adaptation to different production regions. However, the fiber of “extra-long-staple” *G. barbadense* tetraploids, representing ~5% of the world’s cotton, commands a premium price due to ~40% higher fiber length (≈ 35 mm), strength (≈ 30 g per tex or more), and fineness over leading A genome cultivars (13), at similar yield levels. Obsolete *G. barbadense* cultivars reportedly had up to 100% longer fibers (50.8 mm; ref. 14) than modern *G. arboreum* (25.5 ± 1.6 mm; ref. 13).

To further investigate cotton fiber evolution, a detailed restriction fragment length polymorphism (RFLP) map (15) was used to determine the chromosomal locations and subgenomic (A versus D) distributions of quantitative trait loci (QTLs) segregating in a cross between a high-fiber-quality *G. barbadense* cultivar and a high-yielding *G. hirsutum* cultivar (both AADD). The D subgenome, from the non-fiber-producing ancestor, accounts for much more genetic variation in fiber traits of *G. barbadense* and *G. hirsutum* than does the A subgenome, from the fiber-producing ancestor. Lack of correspondence between QTLs in the A and D subgenomes,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/954419-6\$2.00/0
PNAS is available online at <http://www.pnas.org>.

Abbreviations: PV, phenotypic variance; PVE, phenotypic variance explained. QTL, quantitative trait locus; RFLP, restriction fragment length polymorphism; Ln, length by number; Lw, length by weight; SFCw, short fiber content by weight; CV, coefficient of variation; logSDCT, mass of cotton seed; LB, ratio of log(locule number) to log(boll number); lod, logarithm of odds; cM, centimorgan(s).

‡To whom reprint requests should be addressed.

diverged by about 10 million years, is in striking contrast to the extensive correspondence among QTLs in other genomes diverged by as much as 65 million years (16). The large influence of the D genome on fiber properties of tetraploid cotton suggests that polyploidy permits unique avenues for response to selection (3, 4, 17). These findings impel molecular dissection of the roles of divergent subgenomes in quantitative inheritance in many other polyploids and further exploration of both "synthetic" polyploids and exotic diploid genotypes for agriculturally useful variation.

MATERIALS AND METHODS

Mapping Population. A total of 271 F₂ progeny from a cross of *Gossypium hirsutum* cv. "CAMD-E" (hereafter GH) × *G. barbadense* cv. "Sea Island Seaberry" (GB) were grown under typical production conditions at the Texas A & M Experimental Farm near College Station, TX. Mature fiber was harvested from each plant by hand at three dates (September 3, 12, and 19). GH and GB were homozygous at all marker loci examined. GB was generously provided by Ed Percival (U.S. Department of Agriculture–Agricultural Research Service, College Station, TX).

Phenotypes. Fiber quality traits were measured by the International Center for Textile Research and Development (Texas Tech University, Lubbock, TX). Fiber length, thickness (diameter in micrometers), and elongation were measured by using the Advanced Fiber Information System instrument, whereas strength (g per tex) and micronaire were measured by standard techniques. Six traits reflecting fiber length were measured, including length by number (Ln) and weight (Lw), short fiber content by number (SFCn) and weight (SFCw), upper quartile length by number (UQLn) and weight (UQLw). SFCw was closely correlated to SFCn ($r = 0.93$), Ln ($r = -0.93$), Lw ($r = -0.75$), UQLw ($r = -0.56$), and UQLn ($r = -0.82$) and revealed all QTLs detected by the other length parameters, so that only data for SFCw are shown. Only 21 of the 271 fiber samples were large enough for traditional micronaire analysis; micronaire was closely correlated with fiber thickness ($r = 0.70$) so that only data for fiber thickness are shown. The coefficient of variation (CV) of fiber length by number (LnCV) and weight (LwCV) were closely correlated ($r = 0.85$), so that only LnCV data are shown. Four yield components were measured: number of bolls (fruits) per plant (BLs), number of locules per boll (LCs), mass of seed cotton (logSDCT; to improve normality the log₁₀ transformation was used for QTL analysis), and ratio of log(locule number) to log(boll number) (LB). The LogSDCT was highly correlated to both BLs ($r = 0.89$) and LCs ($r = 0.92$), so that only LogSDCT and LB are shown. Early maturity was evaluated on the percent of the seed cotton mass harvested by September 12, the second of the three weekly harvests.

Genotyping and Data Analysis. RFLP analysis used laboratory methods and DNA probes as described (15), supplemented by new probes (A.H.P., unpublished results). Phenotypic distributions for most traits were approximately normal—SDCT was skewed left (toward low yield). However, log₁₀ transformation (logSDCT) improved normality and was used in all analyses. Trait means and correlations were calculated by using SAS (18) and Microsoft EXCEL version 5.0. Linkage maps were made using MAPMAKER (19) and the Kosambi centimorgan (cM) function. QTL likelihood maps, gene action, and phenotypic variance (PV) explained by both individual QTL and multiple-QTL models were determined by interval mapping (20–22), using MAPMAKER/QTL. A stringent logarithm of odds (lod) threshold of ≥ 3.0 was used to keep the likelihood of even one false positive below 5% in the large genome of cotton (>4,000 cM; see ref. 21). For each lod peak, 1 and 2 lod support intervals were determined (Fig. 1). In cases where the maximum likelihood location of a QTL fell more

than 10 cM from the nearest marker, the lod score and phenotypic effects reported (Tables 1 and 2) were at the marker locus, to avoid any possible bias due to artifactual assignment by MAPMAKER/QTL of unlinked genotypic variance to the centers of large intervals between markers.

RESULTS

Genetic Map of the Interspecific Cross. The GH × GB F₂ linkage map included 261 RFLP markers in 27 linkage groups and spanned 3,767 cM with an average spacing of 14.4 cM between markers. Sixteen markers were not linked to the map, suggesting an overall map length that exceeded 4,000 cM. The linear orders of markers showed only small differences from the previously published map (15), usually associated with short distances between markers, or polymorphisms at different homoeologous (duplicated) loci detected by common probes.

QTLs Identified for Fiber-Related Traits. A total of 14 QTLs affecting fiber-related traits met the required significance threshold (lod 3.0). Allele effects for 12 of the 14 were consistent with the difference between parents; with GB alleles associated with long, strong, and fine fibers; and with GH alleles associated with higher yield and early maturity. The two exceptions were each fiber strength QTLs, where the GB alleles reduced fiber strength. Although the maturity of individual fibers was not assessed in this study, subsequent data (C.J., L. Decanini, C. W. Smith, and A.H.P., unpublished results) have suggested that immaturity of GB fibers can be associated with reduced strength. Individual QTLs detected were as follows:

Fiber strength. Three QTLs (Table 1) collectively explained 30.9% of PV (Table 2). The GB allele increased fiber strength on LGD02 and decreased it on chromosomes 20 and 22 (Fig. 1).

Fiber length (SFCw). One QTL was detected (Table 1), explaining 14.7% of PV in SFCw (Table 2). The GB allele on LGD03 increased fiber length (Fig. 1).

Fiber thickness. One QTL was detected (Table 1), explaining 12.6% of PV (Table 2). The GB allele on chromosome 10 decreased fiber thickness (Fig. 1). Fiber thickness was measured as the diameter in micrometers.

CV in mean length of fiber by number (LnCV). Two QTLs (Table 1) collectively explained 23.0% of PV (Table 2). The GB alleles on both chromosomes 10 and 15 increased the CV (Fig. 1).

Fiber elongation. Two QTLs (Table 1) collectively explained 21.1% of PVE (Table 2). The GB alleles on both LGD03 and LGA02 increased fiber elongation (Fig. 1).

Fiber yield components (logSDCT and LB). Four QTLs (Table 1) were detected, two explaining 17.0% of PV in logSDCT and two explaining 58.5% of PVE in LB (Table 2). The GB alleles on linkage groups A02 and D07 decreased seed cotton yield, and those on chromosomes 23 and 25 decreased locules per boll (Fig. 1).

Earliness (September 12). Only one QTL was detected (Table 1), on LGD04, explaining 8.1% of PV (Table 2). The GB allele delayed fiber harvest, increasing the percentage of fiber harvested after September 12, the second of three weekly harvest dates (Fig. 1).

Subgenomic Distribution of QTLs. The A versus D subgenomic origin of chromosomes and linkage groups was determined on the basis of concordance of tetraploid restriction fragments with those found in only one of the two diploid progenitor genomes (15). These determinations agreed with prior assignment of cytologically discernible chromosomes to subgenomes in all of the 14 cases for which chromosomal identity of "linkage groups" was known (15). When the term "linkage group" is used, it means that we do not yet know the identity of the corresponding chromosome (usually due to lack of availability of the diagnostic aneuploid genetic stock).

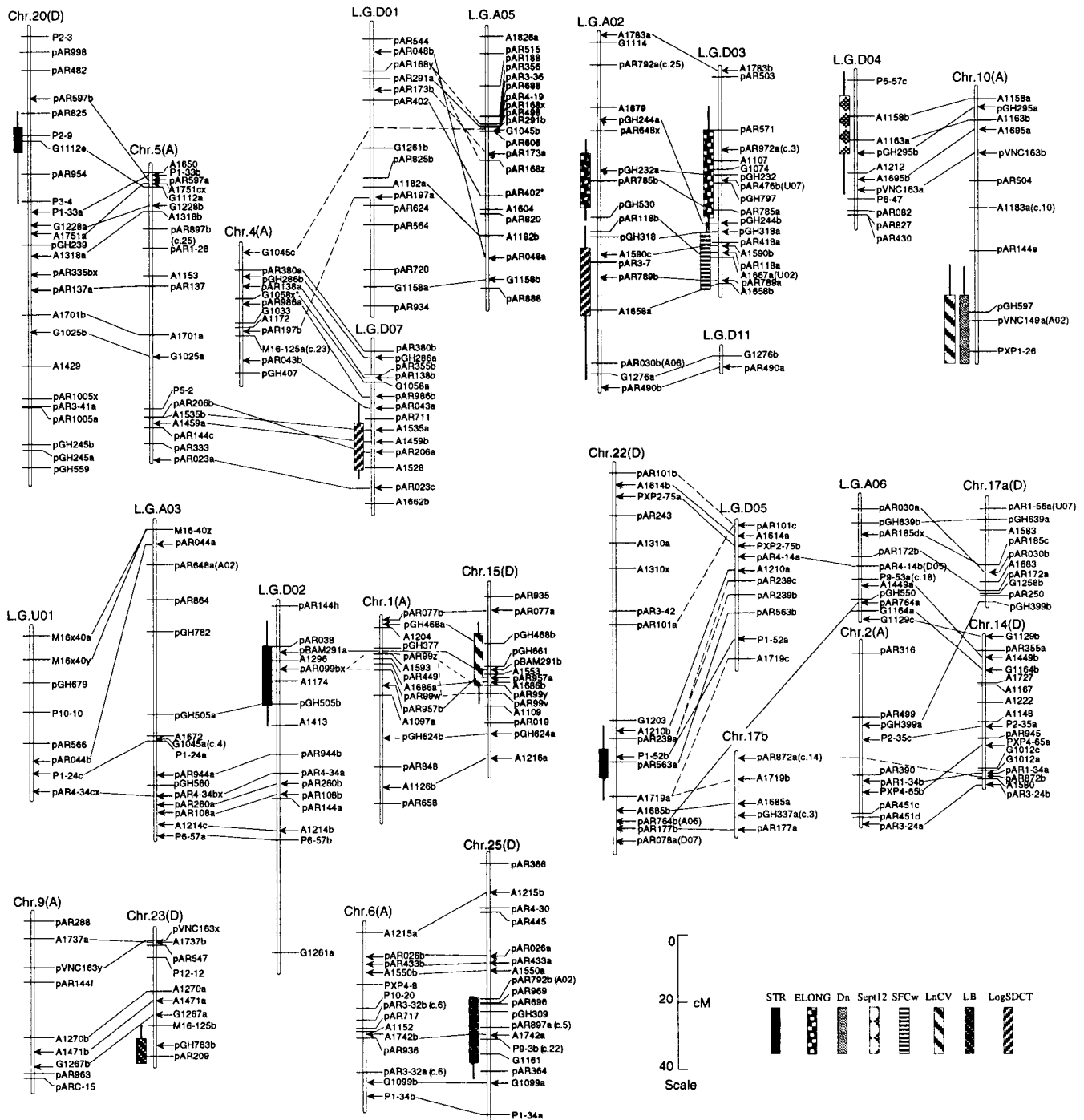


FIG. 1. Likelihood intervals for QTLs mapped. Bars and whiskers indicate 1 lod (10-fold) and 2 lod (100-fold) likelihood intervals. STR, fiber strength; ELONG, fiber elongation; Dn, fiber thickness; Sept12, earliness; SFCw, fiber length; LnCV, CV in mean length of fiber by number; LB, ratio of log(locule number) to log(boll number); LogSDCT, mass of seed cotton. Solid lines connecting probes on different linkage groups indicate homoeologous chromosomal segments supported by three or more pairs of adjacent duplicated loci (15). Dotted lines connecting LGD01, chr4, and LGA05; chr1, chr15, and LGD02; and chr17b, LGD05, and chr22 indicate cases where orthology (homoeology) cannot be distinguished from paralogy (multiple duplication events; see ref. 15). Arrows indicate the inferred locations of markers used to align the homoeologous linkage groups, based on a published map (15). An asterisk denotes loci not on the published map (15). For any duplicated loci that are not consistent with inferred relationships among chromosomes (linkage groups), the homoeologous (or paralogous) loci are indicated in parentheses. Chromosomes/linkage groups that neither contain QTLs nor are homoeologous to regions containing QTLs are not shown.

Among the 14 QTLs affecting fiber-related traits that met the stringent lod 3.0 threshold, 10 (71%) fell on D subgenome linkage groups. In each subgenome, only 2 QTLs mapped to overlapping intervals (LGs A02 and D03).

The D subgenome bias of QTLs is not explained by the subgenomic composition of the overall genetic map, which is virtually identical in the two subgenomes (D = 1,918 cM or

48.7%; A = 1,958 cM or 49.7%; and 64.9 cM or 1.6% is uncertain), as reported for a second mapping population (15). The DNA content of the A subgenome is about 50% larger than the D subgenome (15).

The D subgenome bias of QTLs also is not explained by a difference in the levels of genetic variation in the two subgenomes—prior RFLP mapping revealed that among 705 poly-

Table 1. Biometrical parameters of individual QTLs affecting fiber-related traits of cotton

QTL	lod	% variation	a	d/a	Mode	lod at homeologous site
STR-D02	4.05	13.3	2.36	0.52	DA	0.68
STR- <i>chr22</i>	3.96	12.0	-2.12	-0.25	RA	0.45
STR- <i>chr20</i>	3.34	9.7	-1.31	1.42	D	0.02
SFCw-D03	3.23	14.7	-1.13	0.60	DA	0.22
Dn- <i>chr10</i>	3.03	12.6	-0.49	0.03	DA	NA
LnCV- <i>chr15</i>	3.44	14.9	2.09	1.24	D	0.95
LnCV- <i>chr10</i>	3.54	13.2	2.02	0.74	DA	NA
ELONG-D03	4.42	12.3	0.42	-0.49	RA	3.40
ELONG-A02	3.40	14.0	0.47	0.05	DA	4.42
LogSDCT-D07	4.12	12.3	-0.23	0.43	DA	0.68
LogSDCT-A02	3.10	6.4	-0.18	-0.00	RA	0.00
LB- <i>chr23</i>	4.67	54.1	-0.23	1.31	D	1.54
LB- <i>chr25</i>	3.07	7.9	0.02	-7.10	R	1.03
Sept12-D04	3.01	8.1	0.13	0.38	DA	1.15

NA, no homoeologous chromosome found for this region. Locus name includes trait abbreviation (see Fig. 1) and the linkage group/chromosome (*chr*). Chromosomes 1–13 belong to the A subgenome, and chromosomes 14–26 belong to the D subgenome. For “linkage groups” that have not yet been assigned to chromosomes, A and D followed by numbers, such as A02 and D03, represent linkage groups previously assigned to the respective subgenomes (15). Calculation of additive effects (*a*), dominance deviations (*d*), *d/a* ratios, and mode of gene action was as described (22). Modes of gene action that could not be deemed unlikely are listed in order of decreasing likelihood.

morphic DNA marker loci, 327 (46%) mapped to the D subgenome and 295 (42%) mapped to the A subgenome, with the remainder uncertain (15).

Finally, during the scientific breeding of GH and GB, some introgression has occurred between the two taxa—however, introgressed chromatin is approximately equally distributed across the two subgenomes. About five chromosomal segments account for the majority of introgression between these taxa; two on the A subgenome, two on the D, and one uncertain (23).

Lack of Homoeology Among Cotton Fiber QTLs. The lod scores for chromosomal regions that are homoeologous to mapped QTLs (Table 1) reveal that few of the detected fiber QTLs correspond between the two subgenomes. Only one (25%) A subgenome QTL corresponded to one (10%) D subgenome QTL. The two QTLs affect fiber elongation and map to corresponding sites on LGs A02 and D03 (Table 1 and Fig. 1). In all other cases, lod scores at homoeologous loci fall 30-fold (1.5 lod) or more below the significance threshold. This level of correspondence between fiber QTLs in the cotton subgenomes could occur by chance in about 38.5% of cases (calculated as described in ref. 24).

DISCUSSION

The joining in a common nucleus of A and D genomes, with very different evolutionary histories, appears to have created

Table 2. Subgenomic distribution of QTLs controlling fiber-related trait

Trait	Overall PVE (all QTLs), %	A subgenome		D subgenome	
		No. QTLs	PVE	No. QTLs	PVE
STR	30.9	0	—	3	30.9
SFCw	14.7	0	—	1	14.7
Dn	12.6	1	12.6	0	—
LnCV	23.0	1	13.2	1	14.9
ELONG	21.1	1	14.0	1	12.3
LB	58.5	0	—	2	58.5
LogSDCT	17.0	1	6.4	1	12.3
Sept12	8.1	0	—	1	8.1

Overall phenotypic variance explained (PVE) of multiple-QTL models is normally slightly less than the sum of the PVEs of the component single-QTL models, due to factors such as nominal confounding of variance explained by different QTLs or interactions between the QTLs.

unique avenues for response to selection in AD-tetraploid cottons. The D subgenome, from an ancestor that does not produce spinnable fiber, accounts for more genetic variation in fiber traits of modern *G. barbadense* and *G. hirsutum* than does the A subgenome, from a fiber-producing ancestor. Selection for variant alleles at D subgenome loci during domestication and scientific breeding is a likely basis for the observation that leading tetraploid cottons consistently exceed the yield and/or quality of comparable A genome diploid cultivars. The D subgenome bias of fiber QTLs is not explained by differences in recombinational or physical size or levels of genetic variation (as reflected by DNA marker alleles) in the two subgenomes. Although our data provide concrete evidence for the role of the D subgenome in cotton fiber traits, anecdotal information has long suggested that a wild D genome taxon, *Gossypium thurberi*, may have contributed alleles improving fiber quality of “PeeDee” cottons, via a synthetic polyploid (14).

The presence of fiber on wild A genome diploids suggests that when polyploid formation occurred, many A genome loci relevant to fiber development may already have contained “favorable” alleles as a result of natural selection. In about 10 million years since the divergence of the A and D genomes from a common ancestor (6), fiber-related genes evolved along very different paths in the two genomes. Seed-borne fibers can be under very strong natural selection pressure in taxa that use them as a dispersal agent (25–27). When human selection was “recently” imposed, there may have been little selective advantage to new mutations at “major” fiber loci in the A subgenome of tetraploid cotton—because favorable alleles had previously been selected at these loci. The several A subgenome QTLs that we found may represent fiber-related loci that only became limiting factors after favorable alleles were fixed at “major” loci.

By contrast, human selection for fiber attributes of tetraploid cotton may have conferred a unique fitness advantage to mutations at D subgenome loci. Ostensibly, the D subgenome had rarely if ever been under selection for seed-borne fiber, because its diploid progenitors show inadequate promise to warrant domestication. Mutations that enhanced fiber development may have become favorable only after the D genome was joined in the same nucleus with the fiber-producing A (sub)genome. This suggests that the locations of

D subgenome QTLs in tetraploid cottons may guide us to the corresponding (homoeologous) locations of A subgenome loci at which favorable alleles had already been fixed before domestication. Testing this hypothesis will require isolation of the underlying genes and comparative analysis of alleles in both A and D genome diploid and tetraploid cottons, as well as other wild diploid *Gossypium*.

Although extensive correspondence of QTLs has been found in other genomes diverged by up to 65 million years (16, 24), only one A subgenome fiber QTL corresponded to one D subgenome fiber QTL. This may be a further reflection of the idea that the A subgenome already contained favorable alleles for fiber development when polyploid formation occurred.

Cases that contrast with the cotton model, showing correspondence between QTLs on duplicated chromosomal segments, may reflect common or parallel evolutionary history of ancestral chromosomes regarding the trait(s) under selection. For example, many pairs of QTLs affecting domestication-related traits in crosses between wild and domestic forms of both maize (16) and sorghum (24) are at locations that appear to correspond, due to ancient chromosomal duplications. If chromosomal duplication precedes domestication, and ancestral chromosomes were subject to common or parallel selective forces, QTL correspondence is a logical outcome.

Although founder events such as polyploid formation clearly reduce genetic heterozygosity, nonlinear interactions among genes can limit or even reverse the loss of additive genetic variance in populations (28, 29). Our results support the view that polyploid formation provides evolutionary opportunity by permanently combining divergent diploid genotypes, rather than repeatedly sampling from a common pool of genes (30).

The advantages of tetraploid cottons over diploids may exemplify an evolutionary model that holds that a change in "genetic environment" for all genes at once may trigger "... genetic revolution, in which physiologically interacting genes re-adapt to one another in new genetic alignments" (31–33). Several recently described mechanisms may provide new mutations (and quantitative genetic variation) immediately after polyploid formation (7, 34–36), whereas the "buffering" effect of the alternative genome underpins the fitness of the organism. "Directional genomic change" of the paternally donated D subgenome (6) away from its progenitor, as suggested for *Brassica* (35), would be consistent with the phenotypic variation that we observe in cotton. Reciprocal chromosome exchange between the A and D subgenomes of modern tetraploid cottons is rare (37). However, it may have been more frequent in the past, especially if formation of polyploid cotton involved a chromosomally unbalanced intermediate as is thought to occur in many polyploidization events (38).

The discovery of genes from a noncultivated genome that improve agricultural productivity and quality lends fresh support to the importance of preservation and investigation of exotic germplasm. In cotton, the same polyploidization event is thought to have spawned not only *G. hirsutum* and *G. barbadense* but also the other nondomesticated AD polyploid species *Gossypium tomentosum*, *Gossypium mustelinum*, and *Gossypium darwinii* (6). Study of these taxa, and of feral or wild GH and GB, may reveal additional genes that can further improve cultivated cottons. Crosses between diploid crops and divergent wild relatives have recently shown similar evidence of such "cryptic valuable alleles" (39, 40).

Broadening of the cotton genetic base beyond what may have been present in the natural founder population, by combining divergent D genome taxa or even other *Gossypium* genomes with the A genome (cf. ref. 41), might expose additional valuable alleles or create the opportunity for such alleles to arise. Such "synthetic" polyploids developed by humans from wild plants have contributed to improvement of cotton (14, 41), wheat (42–44), peanut (45), and other crops. A fascinating possibility would be to apply recurrent selection

to synthetic polyploids (using genetic or chemical emasculation if necessary) to explore the potential of cotton and other polyploids to evolve extreme phenotypes such as those bred in diploid maize (46).

Genetic mapping in polyploid crops such as wheat, triticale, strawberry, banana, and others should shed further light on the role of intergenomic interactions in (crop) evolution. In cotton, molecular cloning of fiber QTLs is the next step in dissecting the evolution of agricultural productivity and quality. High-density mapping and progress toward alignment of the large genome of cotton with small genomes such as that of *Arabidopsis* (47) may facilitate this goal.

We thank Ed Percival for seed; Mark Burow, Peter Morrell, and Jonathan Wendel for helpful discussions; Guo-liang Wang, Peggy Thaxton, and Jian-min Dong for technical assistance; and the Texas Higher Education Coordinating Board, Texas Agricultural Experiment Station, and U. S. Department of Agriculture for funding.

1. Stebbins, G. L. (1966) *Science* **152**, 1463–1469.
2. Masterson, J. (1994) *Science* **264**, 421–424.
3. Tal, M. (1980) in *Polyploidy: Biological Relevance*, ed. Lewis, W. H. (Plenum, New York), pp. 61–76.
4. Levin, D. A. (1983) *Am. Nat.* **122**, 1–25.
5. Hilu, K. W. (1993) *Am. J. Bot.* **80**, 1494–1499.
6. Wendel, J. F. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4132–4136.
7. Wendel, J. F., Schnabel, A. & Seelanan, T. (1995) *Mol. Phylogenet. Evol.* **4**, 298–313.
8. Stephens, S. G. (1967) *Cienc. Cult.* **19**, 118–134.
9. Fryxell, P. A. (1979) *The Natural History of the Cotton Tribe* (Texas A & M Univ. Press, College Station, TX).
10. Stephens, S. G. & Moseley, M. E. (1974) *Am. Antiquity* **39**, 109–122.
11. Chowdhury, K. A. & Burth, G. M. (1971) *Linn. Soc. London Biol. J.* **3**, 303–312.
12. Lee, J. (1984) in *Cotton*, eds. Kohel, R. J. & Lewis, C. F. (ASA/CSSA/SSSA Publishers, Madison, WI), pp. 6–24.
13. Anonymous (1997) *Zonal Coordinators Annual Report of All-India Coordinated Cotton Improvement Project*.
14. Niles, G. A. & Feaster, C. V. (1984) in *Cotton*, eds. Kohel, R. J. & Lewis, C. F. (ASA/CSSA/SSSA Publishers, Madison, WI), pp. 202–229.
15. Reinisch, A. R., Dong, J. M., Brubaker, C., Stelly, D., Wendel, J. & Paterson, A. H. (1994) *Genetics* **138**, 829–847.
16. Paterson, A. H., Lin, Y. R., Li, Z., Schertz, K. F., Doebley, J. F., Pinson, S. R. M., Liu, S. C., Stansel, J. W. & Irvine, J. E. (1995) *Science* **269**, 1714–1718.
17. Stebbins, G. L. (1971) *Chromosomal Evolution in Higher Plants* (Addison-Wesley, Reading, MA).
18. SAS Institute (1989) *SAS/STAT User's Guide* (SAS Institute, Cary, NC), Version 6, 4th Ed.
19. Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E. & Newburg, L. (1987) *Genomics* **1**, 174–181.
20. Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E. & Tanksley, S. D. (1988) *Nature (London)* **335**, 721–726.
21. Lander, E. S. & Botstein, D. (1989) *Genetics* **121**, 185–199.
22. Paterson, A. H., Damon, S., Hewitt, J. D., Zamir, D., Lincoln, S. E., Lander, E. S. & Tanksley, S. D. (1991) *Genetics* **127**, 181–197.
23. Wang, G., Dong, J., Paterson, A. H. (1995) *Theor. Appl. Genet.* **91**, 1153–1161.
24. Lin, Y. R., Schertz, K. F. & Paterson, A. H. (1995) *Genetics* **141**, 391–411.
25. Diamond, J. (1996) *Nature (London)* **380**, 103–104.
26. Cody, M. & Overton, J. (1996) *J. Ecol.* **84**, 53–62.
27. Carlquist, S. (1974) *Island Biology* (Columbia Univ. Press, New York).
28. Bryant, E. & Meffert, L. (1993) *Heredity* **70**, 122–129.
29. Cheverud, J. M. & Routman, E. J. (1996) *Evolution* **50**, 1042–1051.
30. Mendiburu, A. O. & Peloquin, S. J. (1977) *Euphytica* **26**, 573–583.
31. Coyne, J. A. (1996) *Science* **272**, 700–701.
32. Mayr, E. (1963) in *Animal Species and Evolution* (Belknap, Cambridge, MA).
33. Templeton, A. R. (1981) *Annu. Rev. Ecol. Syst.* **12**, 23–48.

34. Soltis, D. E. & Soltis, P. S. (1993) *Crit. Rev. Plant Sci.* **12**, 243–273.
35. Song, K., Lu, P., Tang, K. & Osborn, T. C. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7719–7723.
36. Wendel, J. F., Schnabel, A. & Seelanan, T. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 280–284.
37. Kimber, G. (1961) *Nature (London)* **191**, 98–99.
38. Harlan, J. R. & DeWet, J. M. J. (1975) *Bot. Rev.* **41**, 361–390.
39. Tanksley, S. D., Grandillo, S., Fulton, T. M., Zamir, D., Eshed, Y., Petiard, V., Lopez, J. & Beck-Bunn, T. (1996) *Theor. Appl. Genet.* **92**, 213–224.
40. Xiao, J., Li, J., Grandillo, S., Ahn, S. N., McCouch, S. R., Tanksley, S. D. & Yuan, L. (1996) *Nature (London)* **384**, 223–224.
41. Stewart, J. McD. (1979) in *Plant Tissue Culture*, ed. Barber, J. T. (Southern Section Am. Soc. Plant Physiol., Rockville, MD), pp. 44–56.
42. Feldman, M. (1976) in *Evolution of Crop Plants*, ed. Simmonds, N. W. (Longman, Essex, U.K.), pp. 120–128.
43. Villareal, R. L., Mujeeb-Kazi, A., Fuentes Davila, G. & Rajaram, S. (1996) *Crop Sci.* **36**, 218.
44. Watanabe, N., Kobayashi, S. & Furata, Y. (1997) *Euphytica* **94**, 303–309.
45. Burow, M. D., Starr, J. L., Simpson, C. E. & Paterson, A. H. (1996) *Mol. Breeding* **2**, 307–319.
46. Dudley, J. W. & Lambert, R. J. (1992) *Maydica* **37**, 1–7.
47. Paterson, A. H., Lan, T. H., Reischmann, K. P., Chang, C., Lin, Y. R., Liu, S. C., Burow, M. D., Kowalski, S. P., Katsar, C. S., DelMonte, T. A. *et al.* (1996) *Nat. Genet.* **14**, 380–382.