Theor Appl Genet (2017) 130:2219–2230
DOI 10.1007/s00122-017-2952-y

ORIGINAL ARTICLE

Segregation distortion and genome-wide digenic interactions affect transmission of introgressed chromatin from wild cotton species

Rahul Chandnani[1] · Baohua Wang[1,2,5] · Xavier Draye[3] · Lisa K. Rainville[1] · Susan Auckland[1] · Zhimin Zhuang[2,5] · Edward L. Lubbers[2] · O. Lloyd May[2,4] · Peng W. Chee[2] · Andrew H. Paterson[1]

Received: 25 January 2017 / Accepted: 26 July 2017 / Published online: 11 August 2017

Abstract

Key message **This study reports transmission genetics of chromosomal segments into *Gossypium hirsutum* from its most distant euploid relative, *Gossypium mustelinum*. Mutilocus interactions and structural rearrangements affect introgression and segregation of donor chromatin.**

Abstract Wild allotetraploid relatives of cotton are a rich source of genetic diversity that can be used in genetic improvement, but linkage drag and non-Mendelian transmission genetics are prevalent in interspecific crosses. These problems necessitate knowledge of transmission patterns of chromatin from wild donor species in cultivated recipient species. From an interspecific cross, *Gossypium hirsutum* × *Gossypium mustelinum*, we studied *G. mustelinum* (the most distant tetraploid relative of Upland cotton) allele retention in 35 BC_3F_1 plants and segregation patterns in BC_3F_2 populations totaling 3202 individuals, using 216 DNA marker loci. The average retention of donor alleles across BC_3F_1 plants was higher than expected and the average frequency of *G. mustelinum* alleles in BC_3F_2 segregating families was less than expected. Despite surprisingly high retention of *G. mustelinum* alleles in BC_3F_1, 46 genomic regions showed no introgression. Regions on chromosomes 3 and 15 lacking introgression were closely associated with possible small inversions previously reported. Nonlinear two-locus interactions are abundant among loci with single-locus segregation distortion, and among loci originating from one of the two subgenomes. Comparison of the present results with those of prior studies indicates different permeability of Upland cotton for donor chromatin from different allotetraploid relatives. Different contributions of subgenomes to two-locus interactions suggest different fates of subgenomes in the evolution of allotetraploid cottons. Transmission genetics of *G. hirsutum* × *G. mustelinum* crosses reveals allelic interactions, constraints on fixation and selection of donor alleles, and challenges with retention of introgressed chromatin for crop improvement.

Introduction

Interspecific hybridization has been a major source of selectable variation for important traits (Anderson 1949). Interspecific hybrids and introgression populations are an attractive natural means for introducing novel variation into crop improvement (Benbouza et al. 2010; Levi et al. 2009; Tanksley and Nelson 1996). Strong directional selection has resulted in narrow genetic diversity in many self-pollinated crops, and has become a major barrier in crop improvement (Chee et al. 2004). Genetic diversity studies have revealed genetic bottlenecks affecting elite cultivars of several crops such as rice, wheat, beans and cotton (Galvan et al. 2003; Mercado et al. 1996; Wendel et al. 1992; Yu and Nguyen 1994). Wild relatives have shown great potential for broadening the genetic resources that facilitate crop improvement

Communicated by Joshua Udall.

✉ Andrew H. Paterson
paterson@uga.edu

[1] Plant Genome Mapping Laboratory, University of Georgia, 111 Riverbend Road, Athens, GA 30605, USA

[2] NESPAL Molecular Cotton Breeding Laboratory, University of Georgia, Tifton, GA 31793, USA

[3] Unité d'écophysiologie et amélioration végétale, Université Catholique de Louvain, Croix du Sud 1-10, 1348 Louvain-la-Neuve, Belgium

[4] Monsanto Cotton Breeding, Tifton, GA 31793, USA

[5] School of Life Sciences, Nantong University, Nantong 226019, Jiangsu, China

(Hodgkin et al. 2007). Exotic genetic resources have been under-utilized and have huge potential to introduce novel genetic variation into crop gene pools (Zamir 2001). The development of interspecific introgression (IL) lines in tomato has included increased knowledge of QTL for economically important traits, cloning of genes underlying QTLs and new insight into trait evolution (Gur et al. 2011). Development of high yielding and disease resistant synthetic hexaploid wheat by crossing a cultivated tetraploid (*Triticum durum*) and wild relative *Aegilops tauschii*, is another example of the potential of wild relatives for contributing to crop gene pools (Dreisigacker et al. 2005).

Realizing the benefits of introgression from wild relatives may necessitate knowledge of interspecific transmission genetics. While Mendelian principles should predict the inheritance pattern of an allele at a genetic locus, introgressed chromosome segments often include genes or structural rearrangements that result in selection. Widespread deviations from Mendelian segregation ratios have been studied for nearly a century (Mangelsdorf and Jones 1926) and remain of much interest (Li et al. 2011). Segregation distortion can reduce power to determine marker orders on genetic maps and increase the rate of false detection of QTL (Zhang et al. 2010). In backcross populations that resemble introgressions likely occur in nature, loci with segregation distortion can represent incompatibilities between new interspecific allelic combinations, and may be selected against as populations advance (Jiang et al. 2000). The rate of retention of introgressed chromosomal segments has a practical impact on population size needed to find particular gene combinations. Genomic rearrangements such as inversions and/or translocations may occur during the evolution of polyploids and may explain why some regions lacked introgression in advanced backcross cotton populations (Jiang et al. 2000; Waghmare et al. 2016). Availability of genomic resources such as molecular markers, genetic maps and genome sequence has made identification of type of genes involved in creation of gene flow barriers much more feasible (Yang et al. 2012). For example, genetic analysis helped to identify genomic regions that underwent possible selection and whether the nature of selection was zygotic or gametic (Li et al. 2011).

Tetraploid cotton is an attractive platform to study interspecific introgression, with potential practical benefits for crop improvement. The cotton genus *Gossypium* L., contains more than 40 diploid ($2n = 26$) and 7 allotetraploid species ($2n = 4x = 52$). The diploid *Gossypium* species are composed of eight different genome types (A through G plus K) and allotetraploids evolved by joining of A and D genome progenitors (Endrizzi et al. 1984; Wendel et al. 1995). Genetic mapping coupled with use of aneuploid or hyperploid genetic stocks has revealed in cotton the identity and subgenomic origins of most linkage groups (Rong et al. 2004). In tetraploid cotton, the A genome-derived chromosomes (1–13) are designated the At subgenome, and the D genome-derived chromosomes (14–26) are designated the Dt subgenome (Brown 1980). Several lines of evidence have suggested that different subgenomes of allopolyploid species have different roles in the evolution of phenotypes and may show strong bias in homeolog expression (Jiang et al. 1998; Zhu et al. 2011).

Motivated by practical needs and empowered with genomic tools, the introgression tolerance, retention and transmission genetics of both subgenomes have been investigated in several cotton species. Repetitive use of a small group of elite tetraploid cultivars in cotton breeding has resulted in narrow genetic variation in present cultivars and limited the potential for further improvement (May et al. 1995; Tyagi et al. 2014). In contrast, wild species show potential for contributions to cotton improvement (Wang et al. 2016a; Zhang et al. 2011). Interactions between unlinked loci contribute to skewed transmission of *Gossypium barbadense* chromatin in crosses with *G. hirsutum* (Jiang et al. 2000). While a preponderance of inter-subgenomic epistatic interactions contribute to novel features that differentiate tetraploid cotton from its diploid ancestors (Jiang et al. 2000), a preponderance of intra-subgenomic interactions and chromosomal rearrangements restrict introgression in cultivated species from wild relatives (Waghmare et al. 2016).

Here, we explore the transmission genetics of *G. mustelinum*, the sole member identified to date of the tetraploid cotton clade that is most distant from Upland cotton, *G. hirsutum* (Wendel and Cronn 2003). Among the seven tetraploid cotton species, *G. hirsutum* and *G. barbadense* are widely cultivated whereas *Gossypium tomentosum* (Nuttall ex Seemann), *G. mustelinum* (Miers ex Watt), *Gossypium darwinii* Watt, *Gossypium ekmanianum* (Wittm.) and *Gossypium* Sp. Nov. are wild relatives. Together with prior studies of *G. barbadense* and *G. tomentosum*, this will complete our survey of the transmission genetics of crosses among the three clades of tetraploid cottons, and provide for useful comparisons among them.

Materials and methods

Gossypium hirsutum inbred line "PD98042" was used as a maternal parent to cross with *G. mustelinum* accession "AD4-8" to yield 35 F_1 plants. Three backcrosses involved repeated back crossing of one hybrid individual from each generation to PD98042. F_1 hybrids and backcrosses were carried out in the greenhouse, using plants grown in commercial potting mix. Greenhouse temperature was set to 78 °F (day) and 68 °F (night). A total of 12–13 h of light was provided and plants were fertilized twice a month.

Among 35 BC_3F_1 plants, 21 that produced large progeny numbers were selected for this study, with BC_3F_2 families totaling 3202 individuals (an average of 152.4 per family). BC_3F_1 and BC_3F_2 populations were planted in the field (Tifton, GA), following population development and agronomic practices described by Wang et al. (2016a). DNA extractions from BC_3F_1 and BC_3F_2 plants were performed as described by Paterson et al. (1993). The genome composition of the BC_3F_1 plants and the segregation patterns in 21 BC_3F_2 populations were analyzed by genotyping 216 SSR marker loci drawn from genetic map of the F_2 generation (Wang et al. 2016b). SSR markers were amplified as described by Wang et al. (2016b). Distortion from Mendelian segregation, i.e., expected genotypic and allelic segregation ratios were calculated using SAS. SAS procedure FREQ and TABLE statement were used to calculate deviations (SAS and Software 2011). Chi-squared contingency tests (CHISQ option) were performed for association between alleles at pairs of unlinked loci (digenic interactions) while Microsoft Excel spreadsheet software was used to perform tests of homogeneity among two-locus genotypes in different backcross families as described in Snedecor and Cochran (1980).

Results

Genomic composition of BC_3F_1 plants

Gossypium mustelinum alleles were introgressed at much higher than the expected percentage of loci in BC_3F_1 individuals, ranging from 21.75 to 40.27% with an average of 27.86% in contrast to the expected 12.5%. Surprisingly, not a single BC_3F_1 plant fell below the expected rate of retention. Only 9.2% (20 of 216) of loci contain *G. mustelinum* alleles in fewer than the expected numbers of families (an average of 2.6 families, or 12.5% of 21). At two loci, NAU3695 and NAU3074, *G. mustelinum* alleles were retained in 76.19% (16) of BC_3F_1 families. Only 3 of 26 chromosomes (chr. 2, 4, 10) were introgressed throughout their lengths. Regions lacking introgression were required to have six or more linked markers spanning 15 cM or more than showed no introgression—46 such regions were found on 23 chromosomes, spanning 15.06–132.4 cM and totaling 2040.9 cM, or 36.47% of the recombinational length (5595 cM) of the F_2 map. The numbers and lengths of regions lacking introgression for individual chromosomes are shown in Table 1. The average frequency of *G. mustelinum* alleles across all loci and BC_3F_1 individuals was 27.87%, with no significant difference between the At subgenome (26.98%) and Dt subgenomes (28.79%).

Table 1 Summary of correlation between average frequency of donor allele in BC_3F_1 and BC_3F_2 families, and number and length of introgression lacking region for all the chromosomes

Chrom	Correlation	Total length of introgression lacking region	Number of regions
1	0.86	27.7	1
2	0.43	0	0
3	−0.53	87.8	2
4	−0.31	0	0
5	−0.62	151.3	3
6	0.41	196	3
7	−0.20	90.9	2
8	−0.66	91	3
9	0.55	132.4	1
10	0.14	0	0
11	0.41	39	2
12	0.00	140.2	2
13	−1.00	61.2	1
14	−0.65	85.3	3
15	0.14	101	2
16	−0.23	38.7	1
17	0.44	24.1	1
18	−0.16	44.5	2
19	0.11	44.4	1
20	0.20	120	2
21	0.68	208.8	3
22	0.67	40.4	1
23	0.35	99.3	2
24	−0.30	56.5	3
25	−0.05	72	2
26	−0.44	88.4	3
Total		2040.9	46

Correlation is in the average frequency of *G. mustelinum* allele between BC_3F_1 and BC_3F_2 population

Segregation and retention of donor chromatin in BC_3F_2 families

A total of 21 BC_3F_2 progeny lines comprising a total of 3202 individuals (an average of 152.4 per family) derived by selfing of BC_3F_1 individuals were subjected to study of segregation ratios. The average frequency of *G. mustelinum* alleles across all segregating families was lower (38.17%) than expected (50% according to Mendelian expectations). The average *G. mustelinum* allele frequencies of individual loci varied from 18.94 to 60.84%. Both genotypic and allelic segregation studies were conducted and there was widespread segregation distortion. Regarding genotypic segregation ratios (1:2:1), 98.6% of markers distorted significantly ($p < 0.01$) across the whole population and 89.4% of markers were significantly distorted in at least one family for allelic segregation (1:1). There

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_X
_1	39									39
_2	38	6								44
_3	33	11	2	1						47
_4	26	7	2							35
_5	9	4	2			1				16
_6	2	4				1				7
_7	1	1	1				1			4
_9	1	1								2
_Y	149	34	7	1		2	1			194

MUSB0846a

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_X
_1	44	5								49
_2	43	20	1							64
_3	14	15	1							30
_4	7	13	3							23
_5	2	6								8
_Y	110	59	5							174

NAU5443

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_X
_1	42	12								54
_2	31	17	3							51
_3	17	25	1	1						44
_4	10	9	6							25
_5	3	4								7
_6	1		1							2
_7		1								1
_Y	104	68	11	1						184

BNL226a

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_X
_1	54	6								60
_2	37	13	1							51
_3	21	9	3	2						35
_4	12	9	3							24
_5		3			1					4
_6							1			1
_7		1		1						2
_Y	124	41	7	3	1		1			177

BNL3779a

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_X
_1	17									17
_10			1							1
_14		1								1
_2	25	3	1							29
_3	27	11								38
_4	22	6	2							30
_5	12	10	3							25
_6	8	8	6							22
_7	9	10	4	1						24
_8	2	4	1							7
_9	1	5		1						7
_Y	123	58	18	2						201

NAU3695

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_X
_1	58	11								69
_2	24	21	3							48
_3	9	11	2							22
_4	3	2	3	1	1					10
_Y	94	45	8	1	1					149

DPL0325

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_X
_1	34	15								49
_2	24	21	4							49
_3	21	18	4	1						44
_4	2	10	7	1						20
_5		3	1	1						5
_6			2							2
_Y	81	67	18	3						169

DPL0085

was no significant difference in the average (48.67%) and expected (50%) frequency of heterozygotes but the average genotypic frequencies for *G. hirsutum* and *G. mustelinum* homozygotes were 38.58 and 14.44%, respectively, versus expected frequencies of 25% each. Correlations between the average frequencies of *G. mustelinum* alleles in BC_3F_1 and the average frequencies across all the BC_3F_2 families (presented in Table 1) were negative for 12 chromosomes, positive for 11, and not significant for one (chromosome 12).

◂**Fig. 1** Deviations from random assortment of two-locus genotypes in *G. mustelinum* × *G. hirsutum* BC_3F_2 populations. *Panels* show the number of significant deviations from random assortment ($p < 0.01$) between the indicated locus and unlinked loci that co-occur in different segregating families. *Rows 1–9 in the leftmost column of each panel* indicate the number of families in which the indicated marker co-occurs with other unlinked markers. The *Y row* (*bottom cells*) shows the sum of two-locus genotypes deviating significantly from random assortment in the indicated number of cases. *Columns 0–8 of each panel* indicate the number of families in which tests show significant deviation from random assortment of two-locus genotypes. The *X* (*rightmost*) *column* shows the total number of markers that can be tested for random assortment in the indicated number of families. The *bottom right cell of each panel* (*in bold*) shows the total number of two-locus genotypes that can be tested between the indicated marker and all other unlinked loci. For example, in **a**, among a total of 194 unlinked markers that occur with *MUSB0846a,* 39 co-occur in only one family. Of the 39 marker pairs tested, 0 cases show significant deviation from random assortment ($p < 0.01$). Also, 44 other unlinked loci co-occur with *MUSB0846a* in two families. Of those 44 tested interactions, 38 are not significant in both families, and six show significant deviation from random assortment in one of the two families

Two-locus and multilocus interactions and their assortment among BC_3F_2 families

Multilocus interactions can cause differences in genome composition of BC_3F_1 individuals, and deviations from expected segregation ratios in BC_3F_2 families (Jiang et al. 2000). We have studied genome-wide digenic interactions among all unlinked loci. A few loci that have striking differences in segregation ratios among different segregating families might be involved in such interactions. A total of 23 loci for which segregation ratios deviated by more than 0.2 standard deviation (SD) from the overall average were selected for studying interactions in detail. Deviations from random assortment of two-locus genotypes in different segregating families (Fig. 1) were 13 times more frequent than expected by chance (about 13% of total two-locus interactions were significant at $p < 0.01$). A total of 7079 interactions could be tested, among which 933 (13%) were highly significant ($p < 0.01$). Details of digenic interactions among these loci are presented in Table 2. Marker loci with higher standard deviations of segregation ratios among BC_3F_2 families were involved in a higher proportion of significant digenic interactions than other loci. Two loci, DPL0085 (0.23 SD) and BNL3779a (0.22 SD), were each involved in more than 18% of significant interactions. At DPL0085, the frequency of *G. hirsutum* homozygotes ranged from 82.24% in family B27 to 7.64% in family B17. Similarly, at BNL3779a, heterozygote frequency ranged from 11.26% in family B27 to 67.95% in family B05. Genotypic frequencies of eight loci (>0.25 SD) in BC_3F_2 segregating families are mentioned in Table 3. The two-locus genotype patterns for identical pairs of loci in different segregating families were also quite variable in some cases, for example two-locus interactions for DPL0085 × BNL3977 were similar in families B9 and B10 but quite different in family B27 (Fig. 2). Most digenic interactions were in favor of heterozygotes.

Subgenomic differentiation for donor allele retention and digenic interactions

Knowledge of the subgenomic origin of most marker loci enabled us to compare the introgression of *G. mustelinum* alleles in BC_3F_1 individuals, allele frequencies in segregating BC_3F_2 families, and the types of digenic interactions between and within the subgenomes. Among a total of 508 At subgenomic loci and 545 Dt subgenomic marker loci mapped in the F_2 generation, 105 (20.67%) and 111 (20.37%) were introgressed in advanced backcross populations (BC_3) from At and Dt subgenomes, respectively. The average frequencies of *G. mustelinum* alleles for At and Dt subgenomic loci, respectively, were similar in the BC_3F_1 population (26.98%, 28.79%) and BC_3F_2 families (38.72%, 37.64%). The correlation between the rate of retention of donor alleles in the BC_3F_1 plants and the segregation ratios at the same loci in their BC_3F_2 progenies was marginally significant for the At subgenome ($r = 0.21$, $p < 0.028$) but not for Dt subgenome ($r = 0.12$, $p < 0.206$).

Subgenomic origins of digenic interactions were compared in detail. Genome-wide digenic interactions between all possible combinations of 216 marker loci were calculated, finding 9.78% of total interactions significant at $p < 0.01$. From a total of 7079 digenic interactions calculated among the 23 loci studied in detail, 933 (13.18%) deviated significantly from random assortment. Among these 933 interactions, intergenomic interactions (A–D, 529) were significantly more frequent ($X^2 = 16.7$, $p < 0.01$) than intragenomic interactions (A–A or D–D, 404). The At subgenome (519) was involved in significantly more interactions ($X^2 = 11.8$, $P < 0.01$) than the Dt (414), with subgenomic origins of digenic interactions presented in Table 2. The proportion of intragenomic interactions was similar to the expected rate for the Dt subgenome (20.36 ~ 19.68%) but less than expected for the At subgenome (22.93 < 30.94%).

Selection against *Gossypium mustelinum* alleles

A total of 61 loci showed no *G. mustelinum* (MM) homozygotes in 107 cases across all the segregating families, which suggests strong negative selection against *G. mustelinum* alleles at or near these loci. Selection against *G. mustelinum* homozygotes was slightly (not significant, $X^2 = 1.2$) stronger against At (54.09%) than Dt subgenomic loci (45.90%). Segregation distortion from genotypic versus allelic frequency ratios indicates the type of selection. There was no significant difference in the number of loci deviating from genotypic and allelic frequencies, respectively. Among cases with

Table 2 Summary of two-locus genotype interactions of loci with higher standard deviation (>0.2 SD) in segregation distortion among *G. mustelinum* × *G. hirsutum* BC_3F_2 families

Locus	Chrom	Position	SD	Number of significant interactions (total tested)	Proportion of tested	Intra genomic (A–A/D–D)	Inter genomic (A-D)
DPL0085	Chr04	109.3	23.59	112 (396)	0.28	53	59
DPL0325	Chr11	58.9	21.63	68 (271)	0.25	27	41
BNL3779a	Chr09	55.1	22.81	74 (403)	0.18	32	42
MUSB0846a	Chr12	120.3	25.55	67 (576)	0.12	22	45
NAU3695	Chr11	199.4	24.53	100 (877)	0.11	46	54
NAU2640b	Chr12	197.6	20.64	57 (686)	0.08	21	36
NAU3791	Chr04	193.5	20.70	15 (238)	0.06	4	11
DPL0270	Chr11	48.6	32.00	6 (98)	0.06	2	4
BNL2835	Chr08	32	28.24	9 (155)	0.06	3	6
DPL0183a	Chr06	190.4	27.50	11 (302)	0.04	4	7
At sum				519 (4002)		214	305
BNL226a	Chr14	114.1	22.84	93 (442)	0.21	39	54
NAU5443	Chr17	100.8	28.20	69 (399)	0.17	31	38
NAU5138	Chr15	216.5	20.87	48 (281)	0.17	18	30
STV188	Chr23	0	23.83	57 (340)	0.17	21	36
STS511	Chr25	0	20.51	37 (239)	0.15	20	17
BNL1163	Chr25	175.1	32.27	14 (96)	0.15	6	8
BNL1059	Chr17	225.6	23.91	15 (123)	0.12	11	4
DPL0378	Chr23	109.5	21.97	19 (215)	0.09	6	13
BNL3443	Chr17	234.7	23.55	15 (179)	0.08	10	5
STS236	Chr14	94.5	30.85	8 (107)	0.07	5	3
NAU5465	Chr14	60.4	22.49	16 (237)	0.07	9	7
DPL0385a	Chr16	146.2	25.17	9 (151)	0.06	6	3
MUCS422	Chr15	222.3	22.40	14 (268)	0.05	8	6
Dt sum				414 (3077)		190	224
Total				933 (7079)		404	529

very strong deviations, ($p < 0.0001$) more have genotypic (189 loci) than allelic segregation distortion (156 loci).

Discussion

Transmission genetics of introgressed chromatin from *G. mustelinum*, the tetraploid cotton thought to be the most distant genetically from *G. hirsutum* contributes to knowledge of transmission patterns in allotetraploid *Gossypium* species (Jiang et al. 2000; Waghmare et al. 2016). Similarity in DNA markers and experimental procedures enabled us to compare transmission genetics among three species representing divergent allotetraploid cotton clades: *G. barbadense*, *G. tomentosum* and *G. mustelinum* (Jiang et al. 2000; Waghmare et al. 2016).

A common theme in the transmission genetics of crosses between each wild tetraploid and *G. hirsutum* is that some regions of the genome are recalcitrant to introgression. We found a total of 46 regions located on 23 *G. mustelinum* chromosomes completely lacking introgression, accounting for 36.47% of total recombinational length whereas in *G. tomentosum*, we found 21 regions representing 28.5% of total recombinational length. However, while underrepresentation of donor alleles was common in *G. barbadense* (7.3%) and *G. tomentosum* (4.63%) studies, the average BC_3F_1 individual retained *G. mustelinum* alleles at more than twice the expected (12.5%) percent of loci. This indicates that although each BC_3F_1 individual was introgressed with *G. mustelinum* alleles at a higher than expected fraction of loci, introgression was possible in limited regions only. Segregation distortion was generally evident from multiple linked DNA markers, thus was clearly not attributable to genotyping errors but was a result of biological factors. If the elimination of donor chromatin occurs randomly after a backcross then the probability of any one unlinked region lacking introgression in any one BC3F1 line would be $(1 - 0.125) = 0.875$. With the simplifying assumption that each of the 23 unlinked regions of segregation distortion behaves as single unit of inheritance and all segregate independently, the probability of all 23 unlinked regions lacking introgression in all 21 BC_3F_1 plants would be $[(0.875)^{23}]^{21} = 9.76$

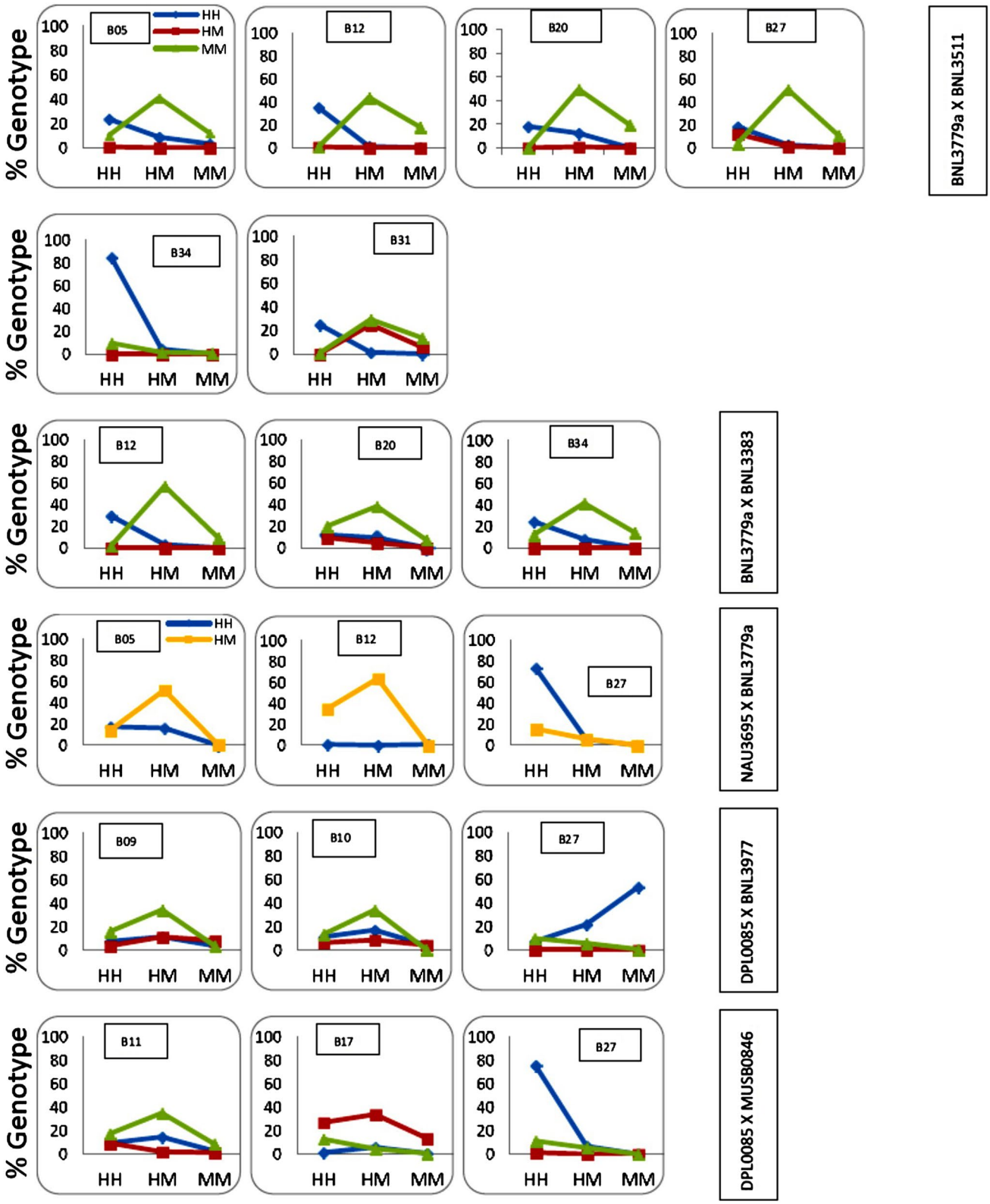
Table 3 Genotypic segregation pattern across all the BC_3F_2 segregating families for the loci with higher standard deviation (>0.25 SD) in genotypic frequencies among different families

BC3F2 family	DPL0085			BNL3779a			DPL0325			MUSB0846		
	HH	HM	MM	HH	HM	MM	HH	HM	MM	HH	HM	MM
B02	–	–	–	–	–	–	–	–	–	23.9	56.0	20.1
B03	–	–	–	–	–	–	–	–	–	19.4	80.6	0.0
B05	–	–	–	31.4	68.0	0.6	–	–	–	31.4	68.6	0.0
B08	78.7	21.3	0.0	–	–	–	–	–	–	–	–	–
B09	22.2	54.8	23.0	–	–	–	34.1	55.3	10.6	–	–	–
B10	32.6	48.2	19.2	–	–	–	–	–	–	–	–	–
B11	27.0	60.5	12.5	–	–	–	–	–	–	35.8	51.7	12.6
B12	–	–	–	36.0	63.3	0.7	–	–	–	29.1	53.0	17.9
B15	–	–	–	97.5	2.6	0.0	99.4	0.6	0.0	–	–	–
B16	–	–	–	–	–	–	–	–	–	20.8	78.5	0.8
B17	7.6	17.8	74.5	–	–	–	89.8	8.3	1.9	41.0	44.2	14.7
B20	–	–	–	21.0	65.6	13.4	–	–	–	96.1	3.9	0.0
B27	82.2	16.5	1.3	88.7	11.3	0.0	91.2	8.8	0.0	88.2	11.8	0.0
B31	–	–	–	26.6	42.4	31.0	–	–	–	–	–	–
B32	–	–	–	–	–	–	24.5	52.8	22.6	95.6	4.4	0.0
B34	–	–	–	32.2	67.8	0.0	–	–	–	–	–	–
B35	24.4	50.0	25.6	–	–	–	–	–	–	–	–	–
Percent genotype	39.3	38.4	22.3	47.6	45.9	6.5	67.8	25.2	7.0	48.1	45.3	6.6
BC3F2 family	BNL0226a			NAU5138			NAU5443			STV188		
	HH		M-	HH	HM	MM	HH	HM	MM	HH	HM	MM
B01	–	–	–	–	–	–	–	–	–	35.7	61.8	2.6
B03	25.6	–	74.4	–	–	–	–	–	–	–	–	–
B05	–	–	–	–	–	–	28.7	71.3	0.0	30.4	69.6	0.0
B06	–	–	–	–	–	–	16.3	81.3	2.5	–	–	–
B07	93.8	–	6.3	29.8	54.4	15.8	–	–	–	–	–	–
B10	24.1	–	75.9	97.2	2.8	0.0	27.7	72.3	0.0	–	–	–
B11	32.0	–	68.0	–	–	–	36.8	63.2	0.0	–	–	–
B12	–	–	–	–	–	–	–	–	–	37.8	62.3	0.0
B17	89.0	–	11.0	–	–	–	–	–	–	–	–	–
B20	31.8	–	68.2	–	–	–	–	–	–	39.5	60.5	0.0
B27	91.3	–	8.7	93.4	6.6	0.0	85.9	14.1	0.0	88.8	11.2	0.0
B31	–	–	–	27.5	51.9	20.6	–	–	–	–	–	–
B34	–	–	–	–	–	–	–	–	–	29.5	70.6	0.0
B35	–	–	–	35.6	54.4	10.0	32.3	67.1	0.6	–	–	–
Percent genotype	55.4		44.6	56.7	34.0	9.3	38.0	61.5	0.5	43.6	56.0	0.4

$\times 10^{-29}$. Thus, it is unlikely that all of these regions lack introgression in all BC_3F_1 plants (21) due to chance. Comparison of introgression from *G. mustelinum* and *G. tomentosum* to *G. hirsutum* revealed that 15 of 26 chromosomes lacking introgression were common in both studies, with four regions (Ch. 03, 08, 15 and 24) also lacking introgression in *G. hirsutum* from *G. barbadense* (Jiang et al. 2000; Waghmare et al. 2016).

Patterns of segregation in the *G. mustelinum* BC_3F_2 families were similar to those found in previous studies of *G. barbadense* and *G. tomentosum*, with lower than expected frequencies of donor alleles across most segregating families (Jiang et al. 2000; Waghmare et al. 2016). Segregation in BC_3F_2 families favors the recipient haplotype with a higher average frequency of *G. hirsutum* than *G. mustelinum* homozygotes, which adds to prior evidence of non-random maintenance of integrity of the recipient genome.

Several lines of evidence add further strength to the finding from our prior studies that multilocus interactions between unlinked loci are an important determinant of



single-locus segregation patterns. Segregation distortion was prevalent in the *G. mustelinum* BC_3F_2 families with respect to both genotypic and allelic segregation ratios (Table 3; Fig. 3). Most loci did not show independent assortment, and most introgressed alleles did not segregate according to single-locus models. For example, *G. mustelinum* alleles at

◂**Fig. 2** Examples of patterns of two-locus genotypic frequencies among different segregating families. DPL0085, BNL3779a, NAU3695 and MUSB0846a all have >0.2 SD among segregation families. Both locus X locus and family X locus interactions are exhibited here. For example, locus X locus interaction was shown by DPL0085, which was in significant interaction with two different loci MUSB0846 and BNL3977 in family B27. Family X locus interaction is exemplified by a pair of loci (BNL3779a × BNL3511) showing different segregation ratios in family B27 and B34. *Green lines with triangles*, *red lines with squares*, *blue lines with diamonds* and *yellow lines with squares* represent *G. mustelinum* homozygotes, heterozygotes, *G. hirsutum* homozygotes and genotypes with presence of *G. mustelinum* alleles for dominant markers

BNL1034 had homozygote frequencies ranging from zero (selected negatively) to 26.42% (near the Mendelian expectation) in different BC_3F_2 families, indicating that these frequencies were affected by alleles at other unlinked loci that differed among the families. Highly significant two-locus interactions (Chi-squared test, $p < 0.01$) were found in 13% of combinations with heterogeneity in the pattern of two-locus genotypes among different families (Fig. 2). Low correlation between retention of donor alleles in BC_3F_1 individuals and their segregation ratios in BC_3F_2 families suggests, as have our prior results, that two-locus interactions between unlinked loci are an important determinant of segregation patterns in individual families. Indeed, we found that the loci involved in higher numbers of significant interactions have the highest variation in segregation ratios among different families.

The respective tetraploid cotton subgenomes have contributed differently to multilocus interactions in the *G. mustelinum* BC_3F_2 families, another recurring theme in cotton introgression (Jiang et al. 2000; Waghmare et al. 2016). In our previous studies of *G. barbadense* and *G. tomentosum* transmission genetics, At and Dt subgenomes were involved in significant two-locus interactions but different subgenomes were favored for the respective donor genomes (Jiang et al. 2000; Waghmare et al. 2016). A higher number of significant interactions originated from At than Dt subgenomic loci in *G. mustelinum*, correlated with average frequency of donor alleles in each BC_3F_1 individual, for all three studies. In *G. tomentosum*, the At subgenome was involved in fewer interactions than the Dt subgenome, while in *G. barbadense* the two subgenomes were involved in similar number of interactions. Biased involvement of the At subgenome only in *G. mustelinum* coincides with a previous report of a closer relationship of only this tetraploid genome with the A genome than the D genome (Wu et al. 2013). A higher number of intergenomic than intragenomic interactions is similar in *G. mustelinum* and *G. barbadense*, whereas there was a preponderance of intragenomic interactions in *G. tomentosum*. Comparison of the three studies suggests evolution of different types and amount of interactions (intragenomic A–A or D–D; intergenomic A–D) between At and Dt subgenomes in divergent allotetraploid species after their divergence from a common ancestor.

Our data about selection against *G. mustelinum* alleles further support the notion that different subgenomes have different evolutionary fates. Selection against At subgenomic loci was slightly stronger than Dt subgenomic loci. Perhaps, this may be related to the observations that D genome has higher expression than A genome and most fiber quality QTLs have been mapped on Dt subgenomic loci in allotetraploid cotton (Flagel and Wendel 2010; Rong et al. 2004).

Different 'permeability' of various regions of the recipient genome may indicate differential levels of fitness for donor alleles, as noted by others (Rieseberg et al. 1999). Complete absence of *G. hirsutum* homozygotes at a few loci or fixation of *G. mustelinum* alleles at those loci show that only a single introgression event can be sufficient to fix the donor allele in a population. On the other hand, the absence of *G. mustelinum* homozygotes suggests that the donor allele dramatically reduces fitness at the relevant locus in the recipient genome. Fitness of *G. mustelinum* alleles at a locus seems to be related to *G. mustelinum* alleles at other loci that are present or absent in that family, and the multilocus interactions acting upon combinations of loci. Selection against *G. mustelinum* alleles was widespread—however, ironically, higher numbers of loci were able to tolerate homozygosity for alleles from *G. mustelinum* than from the more closely related *G. tomentosum* (Waghmare et al. 2016). Although *G. tomentosum* and *G. mustelinum* both are wild allotetraploid species, and in spite of higher average retention of donor alleles in the *G. mustelinum* BC_3F_1 (27.86%) than *G. tomentosum* (4.63%) population, the *G. tomentosum* population had fewer regions (17) than the *G. mustelinum* populations (46) in which introgression was completely absent. A number of possible reasons for these differences can be envisioned. For example, selection against individual *G. mustelinum* alleles may have been stronger, and more frequently met our criterion of affecting six consecutive loci. This may also have the consequence that fewer unfit alleles may have been introgressed in early generations and comparatively fewer segregating alleles remain under selection in advanced generations of *G. mustelinum*.

The nature of selection has been an important aspect (Li et al. 2011) of segregation distortion studies. Although not statistically significant, higher numbers of *G. mustelinum* loci are distorted from genotypic than allelic segregation, which suggests that zygotic selection may be somewhat more important than gametic selection in *G. mustelinum*.

Chromosomal rearrangements are one of the possible reasons for resistance of some genomic regions to introgression. Since many of the regions lacking introgression from *G. mustelinum* were in common with those lacking introgression from other species, we also looked for possible rearrangements in those regions. A recent study

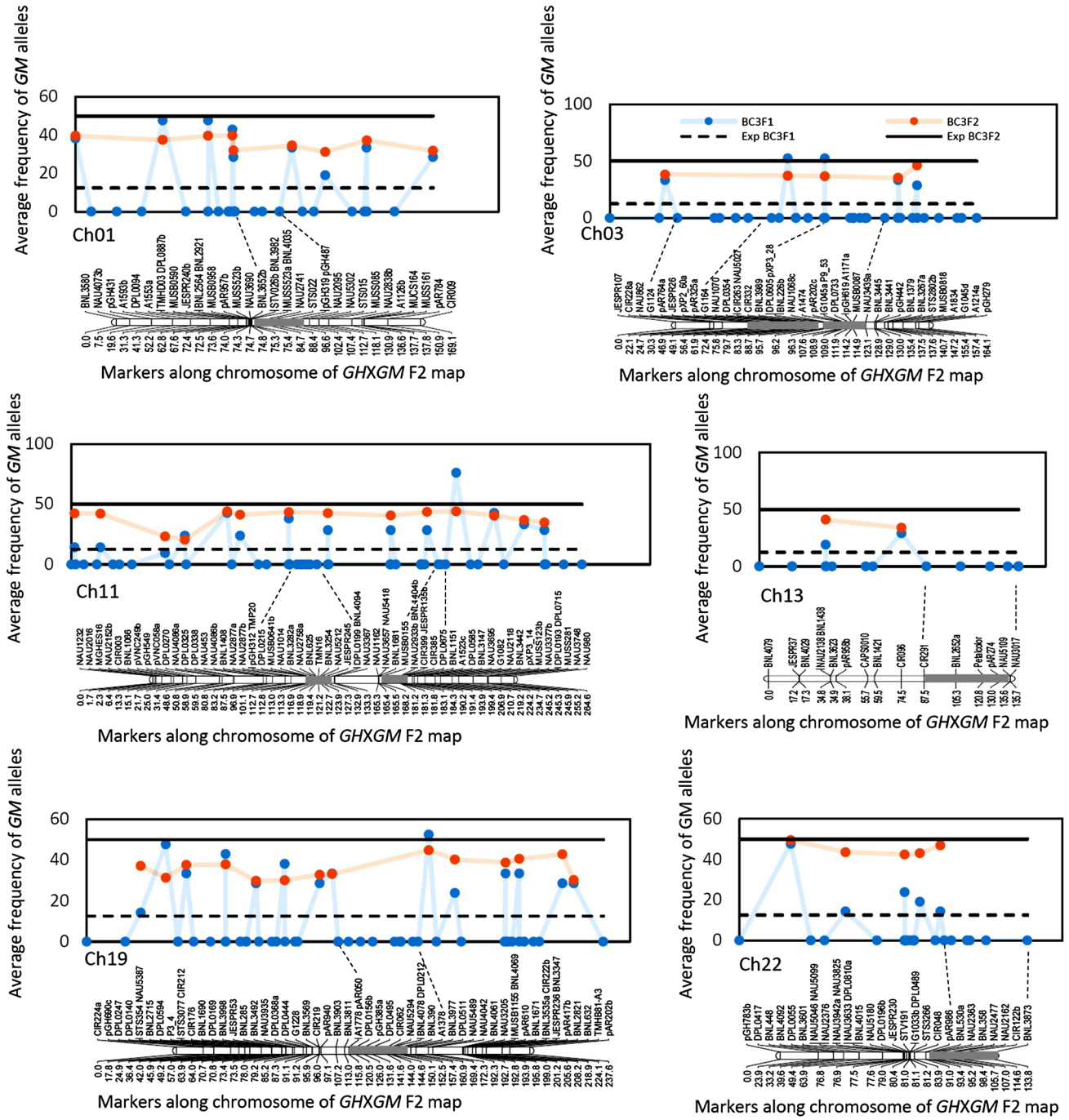


Fig. 3 Segregation pattern of *G. mustelinum* alleles in BC_3F_1 individuals and BC_3F_2 families. *Blue dots* represent observed average frequencies of *G. mustelinum* alleles in BC_3F_1 individuals and *black discontinuous lines* represent the expected frequency (12.5%) in BC_3F_1. *Orange dots* represent observed average frequency of *G. mustelinum* alleles in BC_3F_2 families and the *black solid line* represents the expected frequency (50%) of *G. mustelinum* alleles in BC_3F_2 families. Marker positions along the chromosomes are from an F2 map between *G. mustelinum* and *G hirsutum* (Wang et al. 2016b). The figure shows the segregation pattern of loci on chromosomes with the highest (Chr. 11, 19) and lowest number of introgressed loci (Chr. 3, 22) and the chromosomes with highest positive (Chr. 1, 0.86) and lowest negative (Chr. 3, −1.00) correlation between average frequencies of *G. mustelinum* alleles in BC_3F_1 and BC_3F_2 families

(Wang et al. 2016b) reported possible chromosome structural changes specific to *G. mustelinum*, particularly possible inversions on chromosomes 3, 15 and 17. We found regions lacking introgression from *G. mustelinum* on chromosomes 3 and 15 (Fig. 4) in the inverted regions, spanning 55 and 81.5 cM, respectively. On chromosome 17, we

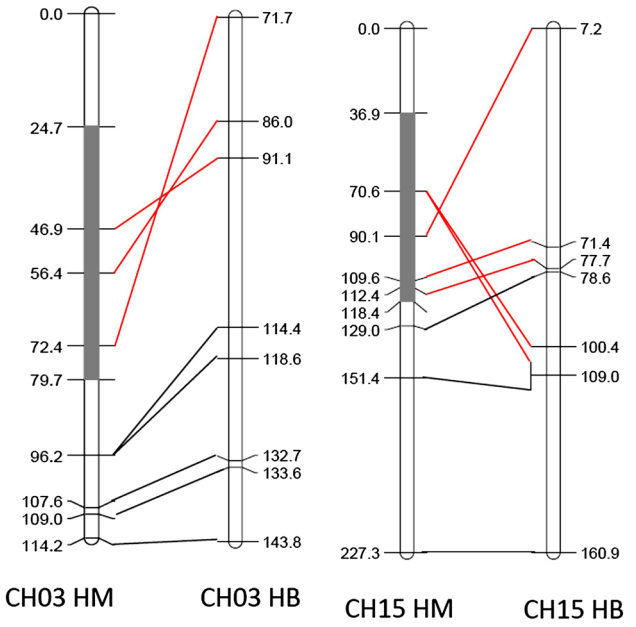


Fig. 4 Previously reported (Wang et al. 2016b) chromosomal regions with possible inversion in *G. hirsutum* × *G. mustelinum* that also harbor introgression lacking regions. In this figure, we have shown only marker loci that were common in *G. mustelinum* and reference maps. Introgression lacking regions are depicted as *shaded gray* regions on chromosomes

found more than six markers lacking introgression in the inverted regions but our criteria (6 consecutive markers lacking introgression) were not met because two groups of three non-introgressed loci were interrupted by one locus that showed introgression. Knowledge of these variations in the pattern of introgression from divergent species may be useful to breeders if novel alleles are sought on a specific chromosome. Regions that are impermeable to introgression may be important in the evolution of reproductive isolation between the tetraploid *Gossypium* species. Cytonuclear interactions may contribute to outbreeding depression which indirectly contributes to resistance to introgression in an interspecific population (Etterson et al. 2007).

In summary, transmission genetic study of *G. mustelinum* shows that the extent of introgression and the fate of introgressed chromatin depend on multilocus interactions, fitness of substituted alleles and allelic combinations, chromosomal rearrangements and cytonuclear interactions. Comparative knowledge of the extent and distribution of introgression and the transmission pattern among *G. barbadense*, *G. tomentosum* and now *G. mustelinum* enables us to develop hypotheses about evolutionary changes in genomic arrangements, and helps to choose specific populations for future breeding efforts to improve cotton for economically important traits.

Author contribution statement RC performed data analysis, interpretation and drafted manuscript. OLM and EL performed crossing and population development. BW, ZZ, LR and SA helped in genotyping and marker data collection. XD helped in data analysis. PC provided feedback on data interpretation and the manuscript. AP proposed the concept and experimental design, provided feedback on data interpretation and revised the manuscript before submission.

Acknowledgements Authors sincerely thank the members of Plant Genome Mapping Lab and Molecular Cotton Breeding Lab for help in lab and field work. We thank Curt Brubaker for initial crossing in population development. We also thank the National Science Foundation for financial support.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Anderson E (1949) Introgressive hybridization. Wiley/Chapman and Hall, New York/London

Benbouza H, Lacape J, Jacquemin J, Courtois B, Diouf F, Sarr D, Konan N, Baudoin J-P, Mergeai G (2010) Introgression of the low-gossypol seed & high-gossypol plant trait in upland cotton: analysis of [(*Gossypium hirsutum* × *G. raimondii*)^2 × *G. sturtianum*] trispecific hybrid and selected derivatives using mapped SSRs. Mol Breed 25:273–286

Brown MS (1980) Identification of the chromosomes of *Gossypium hirsutum* L. by means of translocations. J Hered 71:266–274

Chee PW, Rong J, Williams-Coplin D, Schulze SR, Paterson AH (2004) EST derived PCR-based markers for functional gene homologues in cotton. Genome 47:449–462

Cx Jiang, Wright RJ, El-Zik KM, Paterson AH (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). PNAS 95:4419–4424

Dreisigacker S, Zhang P, Warburton M, Skovmand B, Hoisington D, Melchinger A (2005) Genetic diversity among and within CIMMYT wheat landrace accessions investigated with SSRs and implications for plant genetic resources management. Crop Sci 45:653–661

Endrizzi J, Turcotte E, Kohel R (1984) Qualitative genetics, cytology, and cytogenetics. Cotton 24:81–129

Etterson JR, Keller SR, Galloway LF (2007) Epistatic and cytonuclear interactions govern outbreeding depression in the autotetraploid *Campanulastrum americanum*. Evolution 61:2671–2683

Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytol 186:184–193

Galvan M, Bornet B, Balatti P, Branchard M (2003) Inter simple sequence repeat (ISSR) markers as a tool for the assessment of both genetic diversity and gene pool origin in common bean (*Phaseolus vulgaris* L.). Euphytica 132:297–301

Gur A, Semel Y, Osorio S, Friedmann M, Seekh S, Ghareeb B, Mohammad A, Pleban T, Gera G, Fernie AR (2011) Yield quantitative trait loci from wild tomato are predominately expressed by the shoot. Theor Appl Genet 122:1–16

Hodgkin T, Hajjar R, Maxted N, Ford-Lloyd B, Kell S, Iriondo J, Dulloo M, Turok J (2007) Using crop wild relatives for crop improvement: trends and perspectives. In: Proceedings of the first international conference on crop wild relative conservation and use, Sicily, 14–17 Sep 2005. CABI, pp 535–548

Jiang CX, Chee PW, Draye X, Morrell PL, Smith CW, Paterson AH (2000) Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). Evolution 54:798–814

Levi A, Ovnat L, Paterson AH, Saranga Y (2009) Photosynthesis of cotton near-isogenic lines introgressed with QTLs for productivity and drought related traits. Plant Sci 177:88–96

Li X, Wang X, Wei Y, Brummer EC (2011) Prevalence of segregation distortion in diploid alfalfa and its implications for genetics and breeding applications. Theor Appl Genet 123:667–679

Mangelsdorf PC, Jones DF (1926) The expression of Mendelian factors in the gametophyte of maize. Genetics 11:423

May OL, Bowman DT, Calhoun DS (1995) Genetic diversity of US upland cotton cultivars released between 1980 and 1990. Crop Sci 35:1570–1574

Mercado L, Souza E, Kephart K (1996) Origin and diversity of North American hard spring wheats. Theor Appl Genet 93:593–599

Paterson AH, Brubaker CL, Wendel JF (1993) A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Mol Biol Rep 11:122–127

Rieseberg LH, Kim MJ, Seiler GJ (1999) Introgression between the cultivated sunflower and a sympatric wild relative, *Helianthus petiolaris* (*Asteraceae*). Int J Plant Sci 160:102–108

Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding X, Garza JJ, Marler BS (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389–417

Snedecor GW, Cochran WG (1980) Statistical methods. Iowa State University Press, Iowa

SAS S, Sofware S (2011) Version 9.3. NC SAS Institute Inc, Cary

Tanksley S, Nelson J (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthy V (2014) Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). Theor Appl Genet 127:283–295

Waghmare VN, Rong J, Rogers CJ, Bowers JE, Chee PW, Gannaway JR, Katageri I, Paterson AH (2016) Comparative transmission genetics of introgressed chromatin in *Gossypium* (cotton) polyploids. Am J Bot 103:719–729

Wang B, Draye X, Zhang Z, Zhuang Z, May OL, Paterson AH, Chee PW (2016a) Advanced backcross quantitative trait locus analysis of fiber elongation in a cross between *Gossypium hirsutum* and *G. mustelinum*. Crop Sci 56:1760–1768

Wang B, Liu L, Zhang D, Zhuang Z, Guo H, Qiao X, Wei L, Rong J, May OL, Paterson AH (2016b) A genetic map between *Gossypium hirsutum* and the Brazilian endemic *G. mustelinum* and its application to QTL mapping. G3 6:1673–1685

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78:139–186

Wendel JF, Brubaker CL, Percival AE (1992) Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. Am J Bot 79:1291–1310

Wendel JF, Schnabel A, Seelanan T (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). PNAS 92:280

Wu Q, Liu F, Li S, Song G, Wang C, Zhang X, Wang Y, Stelly D, Wang K (2013) Uniqueness of the *Gossypium mustelinum* genome revealed by GISH and 45S rDNA FISH. J Integr Plant Biol 55:654–662

Yang J, Zhao X, Cheng K, Du H, Ouyang Y, Chen J, Qiu S, Huang J, Jiang Y, Jiang L (2012) A killer-protector system regulates both hybrid sterility and segregation distortion in rice. Science 337:1336–1340

Yu L-X, Nguyen HT (1994) Genetic variation detected with RAPD markers among upland and lowland rice cultivars (*Oryza sativa* L.). Theor Appl Genet 87:668–672

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983–989

Zhang L, Wang S, Li H, Deng Q, Zheng A, Li S, Li P, Li Z, Wang J (2010) Effects of missing marker and segregation distortion on QTL mapping in F2 populations. Theor Appl Genet 121:1071–1082

Zhang Z, Rong J, Waghmare VN, Chee PW, May OL, Wright RJ, Gannaway JR, Paterson AH (2011) QTL alleles for improved fiber quality from a wild Hawaiian cotton, *Gossypium tomentosum*. Theor Appl Genet 123:1075–1088

Zhu H, Han X, Lv J, Zhao L, Xu X, Zhang T, Guo W (2011) Structure, expression differentiation and evolution of duplicated fiber developmental genes in *Gossypium barbadense* and *G. hirsutum*. BMC Plant Biol 11:1