University of Agricultural Sciences, Dharwad

“Association mapping for seed cotton yield, yield components and fiber quality traits in upland cotton *(Gossypium hirsutum* L.)”

Dr.I.S.Katageri
Professor and Head
Dept. Biotechnology(IABT)
College of Agriculture, Dharwad
katageriis@uasd.in

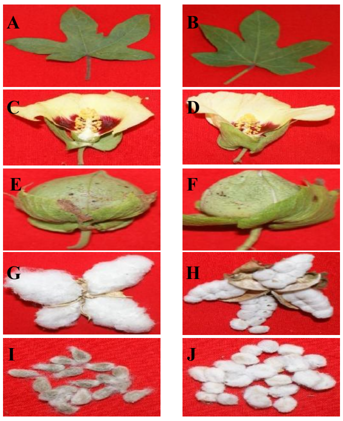
Mapping

- **Gene- Trait association**
- **Biparental mating**
- Gemplasm-**Diversity-Creation/ Recombination**
- **Increasing Productivity with desirable fibre Traits**

Morphological differences between *Fl* line compared to its normal line (*FL*)

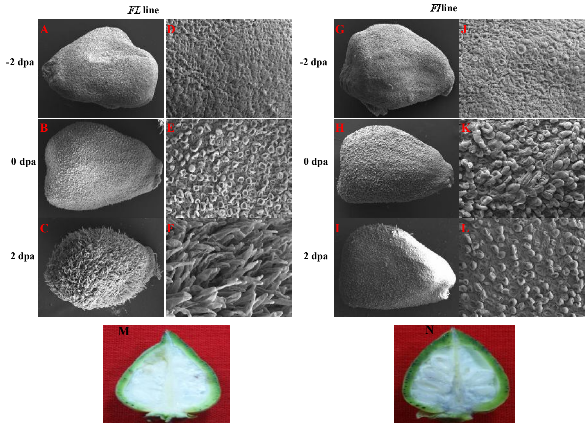
A. ***G. arboreum FL*** **line (Fuzzy linted).** **B.** ***G. arboreum*** **fuzzed-lintless (*Fl*) line plant**

Phenotypic differences of *FL* line and its *FL* line cotton.



A. Leaf of *FL* line ; B. Leaf of *FL* line plant; C. FLower boll of *FL* line ; D. FLower boll of *FL* line ; E. Closed boll of *FL* line plant; F. Closed boll of *FL* line ; G. Opened boll of *FL* line plant; H. Opened boll of *FL* line ; I. Opened boll of *FL* line ; J. Delinted seeds of WT and J. Delinted seeds of *FL* line . **(Atul etal 2013)**

Scanning electron microscopy (SEM) data



Images of SEM pictures of complete ovules (A, B, C, G, H, I) and epidermal layer of ovules showing development of fibre initials (D, E, F, J, K, L) in *G. arbor*eum (*FL* line) (A-F) and its near isogenic *FL* line (G-L) and mature seeds with elongated fibres in *FL* line (M) and in *Fl* line (N) are presented. **(Atul etal 2013)**

Why Association mapping?

- ✓ LD based association study is a
  - ✓ Non-random association between two markers or two genes/QTLs or between a gene/ QTL and a marker locus exists.
  - ✓ High-resolution,
  - ✓ **Broader allele coverage**
  - ✓ **Cost effective gene tagging** approach in plant germplasm resources.

- ✓ Provides an opportunity to widely **dissect and exploit existing natural variations** for crop improvement.

What we Report

- Genetic diversity: 200 germplasm, multi-environment
- Genotyping by single nucleotide polymorphic markers.
- Estimation of linkage disequilibrium and association mapping of markers with yield and fiber quality parameters.

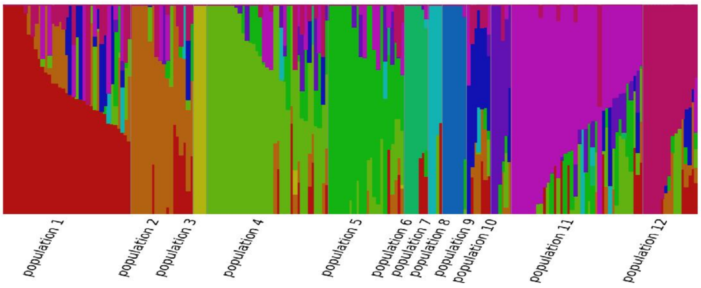
Mean and range

	Season	Plant height (cm)	Mon	Sym	Bolls	Boll weight(g)	Ginning Outturn (%)	Seed index (g)	Lint index (g)	SCY (Kg/ha)
Mean	Pooled	95.8	1.67	16.1	13.7	4.23	36.36	8.98	5.14	1491
Min.	Pooled	71.5	0.74	10.3	8.1	3.12	30.0	6.60	3.64	770
Max.	Pooled	121.4	2.89	20.5	21.2	5.23	40.23	12.15	6.95	2263
CD at 95%	Pooled	20.51	0.53	4.70	3.93	0.61	4.20	1.28	0.79	341.5

Mean, Range

	2.5% Span length (mm)	Uniformity ratio (%)	Micronaire (g/in)	Tenacity (g/tex)
Mean	28.0	47.8	3.98	21.83
Min	23.2	44.0	2.85	18.45
Max	33.2	52.5	4.85	24.85
CD @ 5%	3.87	3.50	0.921	3.26

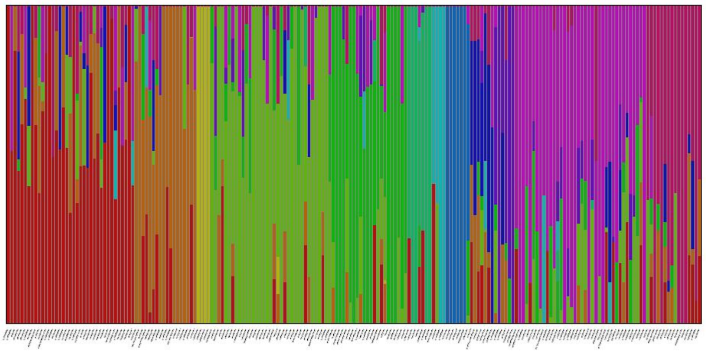
Population structure



Unique Genotypes belong to group not sharing

Q1 (5)	Q2 (6)	Q 3 (4)	Q4 (2)	Q5 (5)	Q6 (4)	Q7 (5)	Q8 (1)	Q9 (4)	Q10 (6)	Q11 (7)	Q12 (10)
W247-85-T14-26	BN	EC560430	EC137592	543395A03N98	IC356543	CNH-120	EC560379	IC357200	EC137593	54335302A012	IC357196
FQT-38	RAH-221	Bed-10m	EC560387	HBS-148	IC357226	EC560323		JBWR-23	IC358790	499887-MONBC	IC359059
128333-Acala-44	HS-289	RAH-3		PVK-Rajat	IC359088	SVPR-2		EC560395	499886-MONAB	EC560376	Cat4186ALAPO-40
RDT-13	BN-SGNR	Sharada		CPD-424	IC359047	ACP-71		EC560399	AK-32	JK-119	LRK-516
Sahana	CSHH-243F-Sirsa			TSH-2005		RAS-303			Vikram	CNH-012	EC559031
	CSH-3088								CNH-50	MCU-5	EC560418
										DS-28	PSHEC-15
											CCHL-76
											L-761
											AH-107

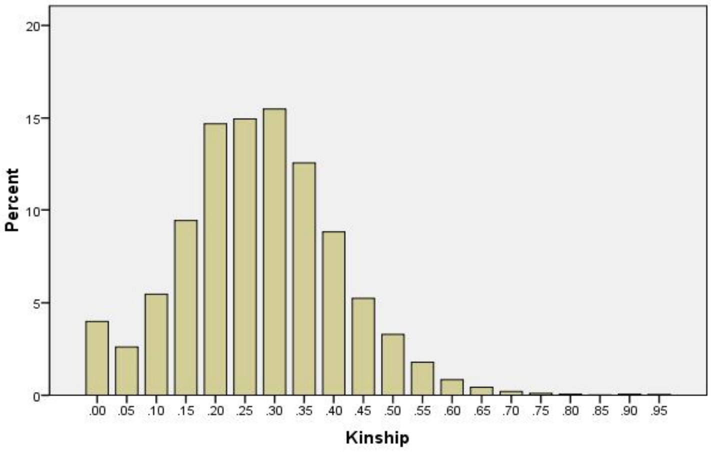
Population structure (fastSTRUCTURE)



Pair wise kinship coefficients

Kinship values	Percent	Cumulative Perce
0.00	4.0	4.0
0.05	2.6	6.6
0.10	5.5	12.1
0.15	9.4	21.5
0.20	14.7	36.2
0.25	15.0	51.1
0.05 to 0.25	**47.1**	
0.30	15.5	66.6
0.35	12.5	79.1
0.40	8.8	88.0
0.45	5.2	93.2
0.50	3.3	96.5
0.30 to 0.50	**45.1**	
0.55	1.8	98.3
0.60	0.8	99.1
0.65	0.4	99.5
0.70	0.2	99.7
0.75	0.1	99.8
0.80	0.1	99.9
0.85	0.0	99.9
0.90	0.1	100
0.95	0.0	100
0.55 to 0.95	**3.5**	
Total	**100**	

Frequency distribution of pair wise kinship coefficient values



Frequency distribution of LD (*R^2*) of marker pairs

R^2value	A genome (%)		D genome (%)		Whole genome (%)	Unmapped (chr 27,28, 29)
	mean	Range	mean	Range		
0 - 0.1	89.2	85.36-97.94	81.88	73.56-91.75	78.31	88.34
0.1 - 0.2	1.37	0.20-2.92	4.26	2.68-5.15	3.00	0.86
0.2 - 0.3	0.60	0.18-1.29	2.02	1.42-3.02	1.41	0.13
0 to 0.3	**91.17**	**-**	**88.16**	**-**	**82.72**	**89.33**
0.3 - 0.4	0.34	0.11-0.71	1.34	0.58-2.15	0.91	0.10
0.4 - 0.5	0.26	0.03-0.55	0.95	0.37-1.29	0.65	0.03
0.5 - 0.6	0.17	0.00-0.51	0.72	0.25-1.45	0.49	0.01
0.6 - 0.7	0.12	0.01-0.23	0.89	0.25-3.65	0.57	0.70
0.7 - 0.8	0.22	0.04-0.75	0.69	0.13-3.33	0.49	1.27
0.8 - 0.9	0.32	0.06-1.50	0.77	0.17-12.86	0.57	4.37
0.9 – <1.0	2.22	0.89-7.80	6.31	1.36-16.01	4.51	4.19
1.0	-	-	-	-	3.13	-
0.3 to 1.0	**3.85**	**-**	**11.7**	**-**	**8.19**	**10.67**

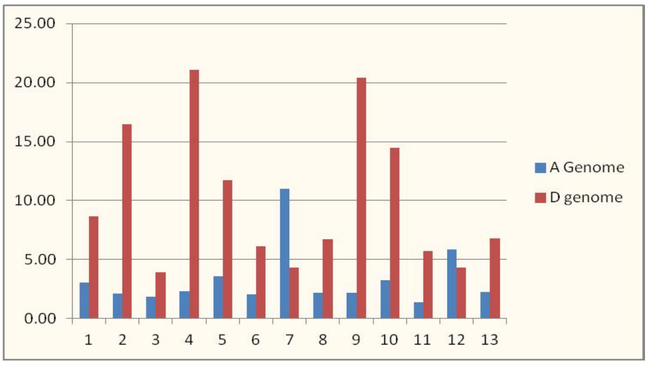
Number of markers covered on each chromosome and average LD value (R^2)

Chromosome number	Number of SNP	Mean R^2	Chromosome number	Number of SNP	Mean R^2
A1	715	0.0865	D14	1148	0.1244
A2	594	0.0664	D15	1203	0.2425
A3	511	0.0574	D16	649	0.0745
A4	548	0.0673	D17	**1459**	**0.2840**
A5	883	0.0945	D18	1227	0.1678
A6	465	0.0644	D19	896	0.0861
A7	**912**	**0.2644**	D20	820	0.0736
A8	410	0.0582	D21	1023	0.0971
A9	594	0.0767	D22	**1779**	**0.2532**
A10	484	0.0747	D23	1178	0.2062
A11	396	0.0563	D24	935	0.0878
A12	565	0.1639	D25	513	0.0688
A13	258	0.0637	D26	966	0.0986
Mean (A)	564	0.0919	Mean (D)	1061	0.1434

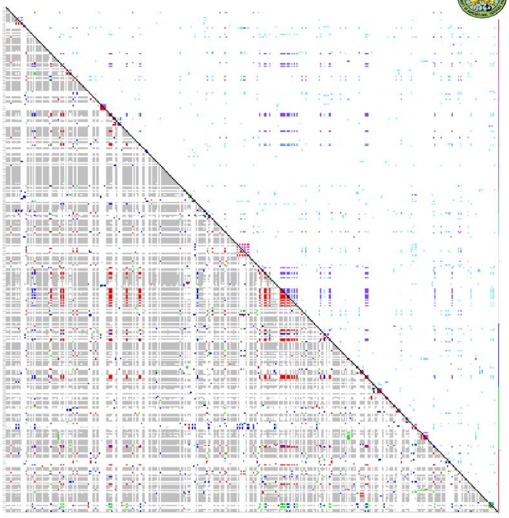
Chromosome wise percentage of marker pairs in LD (>0.3)

Chromosome No.	A Genome	D genome
1	3.06	8.68
2	2.14	16.49
3	1.84	3.95
4	2.30	21.04
5	3.60	11.72
6	2.09	6.11
7	11.01	4.32
8	2.17	6.72
9	2.17	20.41
10	3.26	14.46
11	1.36	5.70
12	5.85	4.35
13	2.23	6.82
Mean	**3.85**	**11.70**
Whole genome mean	**8.19**	

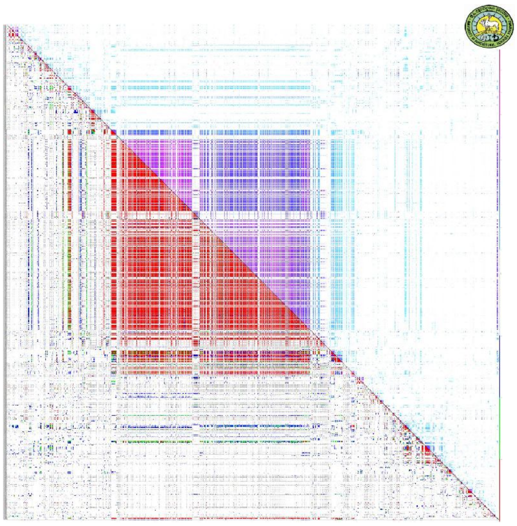
Linkage disequilibrium on different chromosomes of A and D genome



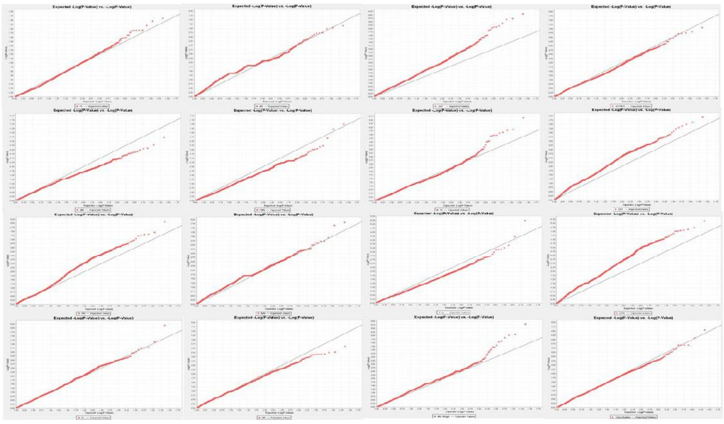
- Linkage disequilibrium plot for chromosome 11



- Linkage disequilibrium plot for chromosome 17

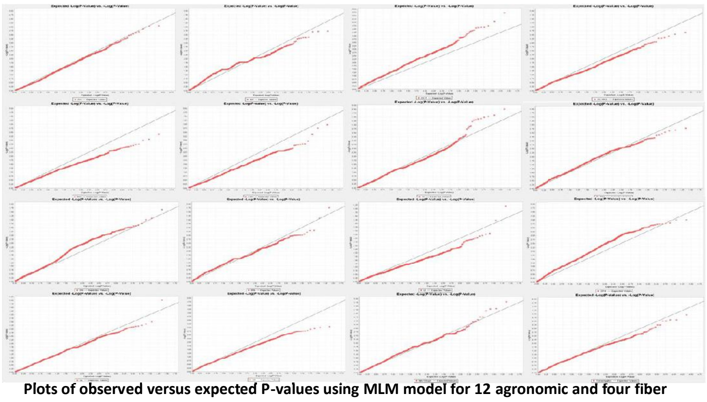


Marker trait associations



Plots of observed versus expected P-values using GLM model for 12 agronomic and four fiber quality traits. The line represents expected P-values, and the red symbol represents the observed P-values

---



Plots of observed versus expected P-values using MLM model for 12 agronomic and four fiber quality traits. The line represents expected P-values, and the red symbol represents the observed P-values

---

Number of markers Associated with Traits

	GLM	MLM	Common	Only in GLM	Only in MLM	R^2>0.1 in
Plant height (cm)	23	19	19	4	0	4
Number of monopodia	4	2	2	2	0	0
Number of sympodia	89	58	56	33	2	12
Number of bolls	23	11	11	12	0	0
Number of fruiting points	8	3	3	5	0	2
Boll weight (g)	26	12	12	14	0	2
Ginning outturn (%)	110	74	66	44	8	20
Seed index (g)	30	23	23	7	0	9
Lint index(g)	9	8	7	2	1	2
Seed cotton yield (kg)/ha	16	9	9	7	0	1
Days to 50% flowering	88	16	15	73	1	3
Days to 50% boll opening	135	74	73	62	1	5
Total yield component traits	**561**	**309**	**296**	**271**	**13**	**60**
2.5% Span length (mm)	23	12	12	11	0	1
Uniformity ratio (%)	10	2	2	8	0	0
Micronaire (g/in)	33	16	16	17	0	6
Tenacity (g/tex)	15	10	10	5	0	1
Total fiber quality traits	**81**	**40**	**40**	**41**	**0**	**8**
Total	642	349	336	312	13	68

Summary of marker-trait associations (more than 10% of phenotypic variation)

Trait	Marker	Chro*	Site/position (bp)	P value	R^2 >0.1
Plant height (cm) Total – 19 (4)	i29306Gh	D24	5199420	8.64E-05	0.1145
	i42807Gh	A7	12295607	1.34E-04	0.1094
	i00181Gh	D22	11756857	2.05E-04	0.1043
	i40368Gh	A12	88775369	2.76E-04	0.1009
Number of sympodia Total – 58 (12)	i49462Gh	A5	89287313	2.26E-05	0.1322
	i29532Gh	A12	49255387	8.15E-05	0.1167
	i42093Gh	A8	16719162	1.01E-04	0.1142
	i05737Gh	D18	63421907	1.73E-04	0.1077
	i21201Gh	A12	95855844	2.28E-04	0.1044
	i61826Gt	D16	475375	2.49E-04	0.1034
	i52085Gb	D16	299932	2.55E-04	0.1031
	i15586Gh	D18	63422332	3.00E-04	0.1012
	i17767Gh	D25	5985273	3.06E-04	0.1009
	i50535Gb	29	48658399	3.08E-04	0.1008
	i53982Gb	D16	270819	3.25E-04	0.1002
	i52441Gb	A8	94591281	3.29E-04	0.1001
Number of fruiting points Total – 3 (2)	i15272Gh	D18	368790	1.07E-04	0.1099
	i04833Gh	D18	371148	1.79E-04	0.1040
Boll weight (g) Total – 12 (2)	i36680Gh	27	16444	$1.74E\text{-}^{5}$	0.1301
	i33806Gh	D26	38783253	8.08E-05	0.1122

Trait	Marker	Chro*	Site/position (bp)	P value	$R^2 > 0.1$
Ginning outturn (%) Total – 74 (20)	i12232Gh	D24	59297518	1.93E-06	0.1525
	i09321Gh	D22	10175360	8.54E-06	0.1354
	i40736Gh	A9	22537047	1.81E-05	0.1268
	i47922Gh	A3	79988301	2.31E-05	0.1240
	i04579Gh	D17	56268264	2.34E-05	0.1238
	i28655Gh	D24	54160132	2.47E-05	0.1232
	i28277Gh	A8	69703782	2.62E-05	0.1225
	i20324Gh	D26	40241229	3.99E-05	0.1178
	i04549Gh	D17	55643092	4.38E-05	0.1167
	i04550Gh	D17	55643296	4.74E-05	0.1158
	i43328Gh	D14	8727117	5.27E-05	0.1146
	i14055Gh	D19	4148125	6.48E-05	0.1123
	i52173Gb	A8	69758391	7.44E-05	0.1107
	i18881Gh	D18	11394888	1.14E-04	0.1059
	i06056Gh	D19	34200312	1.24E-04	0.1050
	i10251Gh	D22	46999061	1.35E-04	0.1040
	i44908Gh	D21	42747114	1.38E-04	0.1038
	i40301Gh	D22	45113205	1.44E-04	0.1033
	i18801Gh	D17	55637445	1.45E-04	0.1032
	i10547Gh	D23	731628	1.49E-04	0.1030

Trait	Marker	Chro*	Site/position (bp)	P value	$R^2 > 0.1$
Seed index (g) Total – 23 (9)	i21971Gh	D18	60318472	9.80E-06	0.1240
	i56626Gb	A12	23453858	2.41E-05	0.1145
	i60970Gt	D20	23430506	2.96E-05	0.1123
	i57170Gb	A11	5029544	3.35E-05	0.1110
	i60609Gb	A5	19818126	4.02E-05	0.1091
	i57394Gb	A11	111215717	4.29E-05	0.1085
	i52725Gb	D23	26396844	4.79E-05	0.1073
	i46956Gh	A11	68228049	5.07E-05	0.1067
	i53857Gb	D23	47606579	6.10E-05	0.1048
Lint index (g) Total – 8 (2)	i09321Gh	D22	10175360	6.70E-06	0.1425
	i03415Gh	D16	36948537	6.18E-05	0.1164
Seed Cotton Yield (Kg) / ha Total – 9 (1)	i52027Gb	D14	6738182	2.32E-04	0.1015
2.5% span length (mm) Total – 12 (1)	i53586Gb	D14	25796117	9.69E-06	0.1182
Micronaire (10^{-6}g/in) Total – 16 (6)	i29870Gh	D25	13732650	1.64E-05	0.1289
	i42819Gh	A4	53107031	4.75E-05	0.1167
	i40086Gh	29	18800490	4.98E-05	0.1162
	i23655Gh	A4	53076924	6.47E-05	0.1132
	i12683Gh	D25	13927103	1.74E-04	0.1020
	i34957Gh	D19	33570928	1.75E-04	0.1019
Tenacity (g/tex) - (10)	i23131Gh	A5	12351293	1.15E-04	0.1050

Number of markers reported in MLM on A and D genomes for yield component and fiber traits

Traits	All significant markers				With $R^2>0.1$			
	A genome	D genome	Unmapped	Total	A genome	D genome	Unmapped	Total
Plant height	3	15	1	19	2	2		4
Number of monopodia		2		2				
Number of sympodia	28	24	6	58	5	6	1	12
Number of bolls	5	5	1	11				
Number of fruiting points	1	2		3		2		2
Boll weight	1	10	1	12		1	1	2
Ginning outturn	18	54	2	74	4	16		20
Seed index	7	14	2	23	5	4		9
Lint index	1	7		8		2		2
SCY (kg)/ha	**5**	**4**		9		1		1
Days to 50% flowering	8	6	2	16	2		1	3
Days to 50% boll opening	14	53	7	74	2	3		5

Traits	All significant markers				With $R^2>0.1$			
	A genome	D genome	Unmapped	Total	A genome	D genome	Unmapped	Total
2.5% Span length (mm)	3	8	1	12		1		1
Uniformity ratio (%)		2		2				
Micronaire (g/in)	4	11	1	16	2	3	1	6
Tenacity (g/tex)	**6**	**3**	1	10	1			1
Total	104	220	25	349	19	37	4	68

Acknowledgements

- My university for funding
- Different Cotton Research Stations and Central Institute for Cotton Research, Nagpur
- Students
- Dr.Rafiq Choudhary
- Dr.Mohamed Negm and all others

From Excellence in Plant Sciences to Value for Society

Thanks