I J P S INTERNATIONAL JOURNAL OF PLANT SCIENCES
ISSN-0973-1547 Volume **15** | Issue 2 | July, 2020 | 121-134
DOI: 10.15740/HAS/IJPS/15.2/121-134
Visit us - *www.researchjournal.co.in*

RESEARCH ARTICLE

Identification and tagging of QTLs for arjunolic acid in *Terminalia arjuna* among Indian sub populations by association mapping and linkage disequilibrium

■ **Sonu Bharti**

SUMMARY

The content of cardiotonic arjunolic acid in *Terminalia arjuna* vary among the population. We studied the population structure and the association between the molecular markers and its active ingredient among 140 plants collected from various agroclimatic zones in India. Large variation was detected for the arjunolic acid in this study showing suitableness of the genotypes. The maximum arjunolic acid content was approximately 238 per cent higher than the lowest value for the genotypes and was found to be considerably correlated to bark thickness, bark fresh weight and bark dry weight. The population structure studies described the existence of nine subpopulations. As the distance increased between the associated markers, Linkage disequilibrium (LD) reduction and a considerable reduction in LD decay was ascertained. Eleven QTL regions associated with arjunolic acid were identified from a genome-wide marker-trait association study. Fine-scale resolution detected significant LD among 3.4 per cent RAPD paired loci and 8.7 per cent ISSR paired loci and 6.7 per cent RAPD paired loci and 13.3 per cent ISSR paired loci. Importantly LD decay found to start at a distance of >20bp from the loci on the genome of *T. arjuna* accessions. Finally, association mapping (AM) in arjun tightly linked to OPT09 which can be a possible substitute to QTL mapping methodology.

Key Words : Arjunolic acid, Association mapping, Linkage disequilibrium, QTL, Marker-trait association

How to cite this article : Bharti, Sonu (2020). Identification and tagging of QTLs for arjunolic acid in *Terminalia arjuna* among Indian sub populations by association mapping and linkage disequilibrium. *Internat. J. Plant Sci.*, **15** (2): 121-134, **DOI: 10.15740/HAS/IJPS/15.2/121-134,** Copyright@ 2020: Hind Agri-Horticultural Society.

Article chronicle : Received : 15.03.2020; **Revised :** 25.05.2020; **Accepted :** 16.06.2020

Arjun [*Terminalia arjuna* (Roxb. Ex DC) Wight and Arnot] is the significant medicinal plant for its chemical constituents such as triterpenoids, glycosides, β-Sitosterol, Flavonoids, tannins and trace elements/minerals. India accounts for 20 per cent of global exports in generics. India's pharmaceutical exports stood at US$ 17.27 billion in 2017-18 and are expected to reach US$ 20 billion by 2020. In 2018-19 these exports are expected to cross US$ 19 billion (According to India Brand Equality Foundation (IBEF). Because of the

AUTHOR FOR CORRESPONDENCE
Sonu Bharti, Department of Life Science, National Institute of Technology, **Rourkela (Odisha) India**
Email : sonugpb@rediffmail.com

increasing demand for Arjun at the global level, enhanced production of the genetically improved species attainment is primacy countrywide. Conventional breeding methods have contributed considerably towards the improvement of genetically and morphologically superior cultivars with triterpenoid saponins in Arjun. The identification of superior germplasm with desirable traits could be facilitated by marker-assisted selection (MAS).

From last two decades, mapping of genes that control the requisite phenotypes has been employed in many plant species for tagging the gene (s) Holland (2007), which however, is of low resolution and lengthy (Holland, 2007). This study demands a large size mapping population across different environments to acquired resilient composition data. In distinction, this will offer to take advantage of genetic variation in natural populations with the high-resolution mapping of complicated traits. Zhu *et al*. (2008) described the use of association mapping as a suitable approach that provides information about the high-resolution mapping based genetic variation for the complex traits in the natural population. It depends on the linkage disequilibrium (LD). If LD occurs between a marker and locus related to an attribute, then specific marker alleles or haplotypes are related to phenotypes at an elevation of applied statistical consequence Cardon and Bell (2001). Population structure and genetic association among the traits could result in spurious associations.

In the last ten years, several qualitative traits such as phenology of leaf and bud, form, growth, resistance against disease, time of flowering and quantitative traits such as wood quality and properties, fibre quality and contents have been used to map by the association mapping. Association mapping was done for several characters in *Eucalyptus* species (Freeman *et al*., 2013); *Pinus* sp. (Sewell *et al.*,1999); *Gossypium barbadense* (Abdurakhmonov *et al.,* 2008); *Hevea* sp. (Lespinasse *et al.,* 2000); *Fagus sylvatica* (Scalfi *et al.,* 2003); *Populus* sp. (Cervera *et al.,* 2004). In the present study, we screened the association of multiple markers to the arjunolic acid among 140 accessions of arjun collected from various parts of India. Association analysis detected eleven RAPD and six ISSR out of forty-two (30 RAPD and 12 ISSR) that attribute associations with arjunolic acid content. The identified markers could be highly used to screen for arjunolic content among various accessions of arjun.

MATERIAL AND METHODS

Plant materials:

Mapping population:

Mapping populations were obtained from all India germplasm of arjun (*Terminalia arjuna*) established at Central Tasar Research and Training Institute (CTRTI), Nagri, Ranchi that maintains superior arjun accessions that representing the nine states and five agro-climatic zones of India (Fig. A). A total of 140 accessions were selected and obtained for bark patch and leaf for extraction of arjunolic acid and genomic DNA for further studies (Table A).

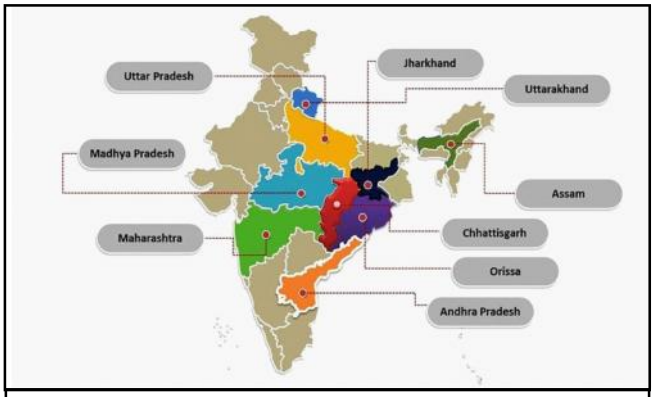


Fig. A: Map showing state-wise location of the sampled mapping population

Genome-wide screening:

Genome-wide screening of the association mapping was performed with 30 RAPD and 12 ISSR markers covering across the whole Arjun genome (Table B). Genomic DNA from the leaf samples of one hundred and forty plants was isolated using CTAB method with some modifications for obtaining quality genomic DNA (ref). Briefly, leaf tissue was macerated in lysis buffer supplemented with PVP, sorbitol and sodium chloride. The macerated material was washed in sorbitol buffer for easy extraction and getting rid of mucilaginous substances present in the sample. Subsequently, the inclusion of 4 per cent of β-mercaptoethanol, 5 mM NaCl, 1 per cent Triton-X and 4 per cent CTAB helped to remove polyphenols and polysaccharides. The quantity of the genomic DNA was determined by Nanodrop.

For PCR amplification, the initial temperature was set at 94°C for 5 min for denaturation followed by 40 cycles of thermal cycler programme: denaturation at 94°C and primer annealing at 38º C for 1 min each and

Table A : List of Arjun genotypes used for association mapping with their source and population in which they are grouped during population structure study

Sr. No.	Accessions	Source	Population	Sr. No.	Accessions	Source	Population
1.	APAB01	Andhra Pradesh	1	71.	JHRNSL02	Chhattisgarh	4
2.	APAB02	Andhra Pradesh	1	72.	JHRNSL03	Chhattisgarh	4
3.	APAB03	Andhra Pradesh	1	73.	JHRNSL04	Chhattisgarh	4
4.	APAB04	Andhra Pradesh	1	74.	JHRNSL05	Chhattisgarh	4
5.	APAB05	Andhra Pradesh	1	75.	JHWSAD01	Chhattisgarh	4
6.	APABCN	Andhra Pradesh	1	76.	JHWSAD02	Chhattisgarh	4
7.	APHBMM	Andhra Pradesh	1	77.	JHWSAT	Chhattisgarh	4
8.	APKN01	Andhra Pradesh	1	78.	JHWSBG	Chhattisgarh	4
9.	APKN02	Andhra Pradesh	1	79.	JHWSBT01	Chhattisgarh	4
10.	APMD01	Andhra Pradesh	1	80.	JHWSBT02	Chhattisgarh	4
11.	APMD02	Andhra Pradesh	1	81.	JHWSGS01	Chhattisgarh	4
12.	APMD03	Andhra Pradesh	1	82.	JHWSGS02	Chhattisgarh	4
13.	APMD04	Andhra Pradesh	1	83.	JHWSGS03	Chhattisgarh	4
14.	APRR01	Andhra Pradesh	1	84.	JHWSHJ	Chhattisgarh	4
15.	APRR02	Andhra Pradesh	1	85.	JHWSKS01	Chhattisgarh	4
16.	APRRVB01	Andhra Pradesh	1	86.	JHWSKS02	Chhattisgarh	4
17.	APRRVB02	Andhra Pradesh	1	87.	JHWSLD	Chhattisgarh	4
18.	APWGNP	Andhra Pradesh	1	88.	MHBDAL	Maharashtra	5
19.	APWGWG	Andhra Pradesh	1	89.	MHBDBD01	Maharashtra	5
20.	ASKRKR	Assam	2	90.	MHBDBD02	Maharashtra	5
21.	CGBT01	Chhattisgarh	3	91.	MHBDBD03	Maharashtra	5
22.	CGBT02	Chhattisgarh	3	92.	MHBDBD04	Maharashtra	5
23.	CGBTAP	Chhattisgarh	3	93.	MHBDBD05	Maharashtra	5
24.	CGDT01	Chhattisgarh	3	94.	MHBDBD06	Maharashtra	5
25.	CGDT02	Chhattisgarh	3	95.	MHBDBD07	Maharashtra	5
26.	CGDT03	Chhattisgarh	3	96.	MHBDBD08	Maharashtra	5
27.	CGDT04	Chhattisgarh	3	97.	MHBDGV	Maharashtra	5
28.	CGDT05	Chhattisgarh	3	98.	MHBDNJ	Maharashtra	5
29.	CGDTSG	Chhattisgarh	3	99.	MHBDPH	Maharashtra	5
30.	CGSGBN	Chhattisgarh	3	100.	MHBDPU01	Maharashtra	5
31.	CGSGDM	Chhattisgarh	3	101.	MHBDPU02	Maharashtra	5
32.	CGSGKH	Chhattisgarh	3	102.	MHCPCP01	Maharashtra	5
33.	CGSGMP	Chhattisgarh	3	103.	MHCPCP02	Maharashtra	5
34.	CGSGSG	Chhattisgarh	3	104.	MHCPCP03	Maharashtra	5
35.	CGSGWN01	Chhattisgarh	3	105.	MHCPCP05	Maharashtra	5
36.	CGSGWN02	Chhattisgarh	3	106.	MHCPCP06	Maharashtra	5
37.	CGSGWN03	Chhattisgarh	3	107.	MHCPCP07	Maharashtra	5
38.	CGSJ04	Chhattisgarh	3	108.	MHCPCP08	Maharashtra	5
39.	CGSJOD	Chhattisgarh	3	109.	MHCPCP09	Maharashtra	5
40.	JHDGGK	Chhattisgarh	3	110.	MHCPGG	Maharashtra	5
41.	JHESBT	Chhattisgarh	3	111.	MHCPMD01	Maharashtra	5
42.	JHGWBG	Chhattisgarh	3	112.	MHCPMD02	Maharashtra	5
43.	JHGWBW	Chhattisgarh	3	113.	MHCPMD03	Maharashtra	5

Table A: Contd……..

Table A: Contd……

44.	JHGWGW	Chhattisgarh	3	114.	MHGCGC01	Maharashtra	5
45.	JHGWKJ	Chhattisgarh	3	115.	MHGCGC02	Maharashtra	5
46.	JHGWRK01	Chhattisgarh	3	116.	MHGDGD01	Maharashtra	5
47.	JHLTLT	Chhattisgarh	3	117.	MHGDGD02	Maharashtra	5
48.	JHRNBD01	Chhattisgarh	3	118.	MHGDGD03	Maharashtra	5
49.	JHRNBD02	Chhattisgarh	3	119.	MHGDGD04	Maharashtra	5
50.	JHRNBD03	Chhattisgarh	3	120.	MHGDNV	Maharashtra	5
51.	JHRNBD04	Chhattisgarh	3	121.	MPTGOC01	Madhya Pradesh	6
52.	JHRNBD05	Chhattisgarh	3	122.	MPTGOC02	Madhya Pradesh	6
53.	JHRNBD06	Chhattisgarh	3	123.	ODMBBP	Odisha	7
54.	JHRNBD07	Chhattisgarh	3	124.	ODSGBK	Odisha	7
55.	JHRNBD08	Chhattisgarh	3	125.	ODSGSG01	Odisha	7
56.	JHRNRN01	Chhattisgarh	3	126.	ODSGSG02	Odisha	7
57.	JHRNRN02	Chhattisgarh	3	127.	UKCMKP	Uttrakhand	8
58.	JHRNRN03	Chhattisgarh	3	128.	UKDDDD01	Uttrakhand	8
59.	JHRNRN04	Chhattisgarh	3	129.	UKDDDD02	Uttrakhand	8
60.	JHRNRN05	Chhattisgarh	3	130.	UKDDLP	Uttrakhand	8
61.	JHRNRN06	Chhattisgarh	3	131.	UKHDHD	Uttrakhand	8
62.	JHRNRN07	Chhattisgarh	3	132.	UKNTHD	Uttrakhand	8
63.	JHRNRN08	Chhattisgarh	3	133.	UKPGPG	Uttrakhand	8
64.	JHRNSH01	Chhattisgarh	3	134.	UKUSTD	Uttrakhand	8
65.	JHRNSH02	Chhattisgarh	3	135.	UPBDBD01	Uttar Pradesh	9
66.	JHRNSH03	Chhattisgarh	3	136.	UPBDBD02	Uttar Pradesh	9
67.	JHRNSH04	Chhattisgarh	3	137.	UPBDBD03	Uttar Pradesh	9
68.	JHRNSH05	Chhattisgarh	3	138.	UPJNJN01	Uttar Pradesh	9
69.	JHRNSH06	Chhattisgarh	3	139.	UPJNJN02	Uttar Pradesh	9
70.	JHRNSL01	Chhattisgarh	3	140.	UPSBRG	Uttar Pradesh	9

extension at 72°C for 3 min, respectively. The PCR reaction -mix of 20 µl contained genomic DNA (50 ng) (primers (20 pM), dNTPs (5 mM), Taq polymerase (1 U) and PCR buffer (1x) with 1.5 mM $MgCl_2$. It was run in T100 thermal cycler. The amplified products were analyzed through agarose gel electrophoresis in 1.2 per cent agarose gel. The 1kb DNA ladder (5 µl, MBTB051, HIMEDIA, Mumbai) containing ten double-stranded fragments ranging 200bp to 10kb was loaded onto both extreme sides of wells. The electrophoresis was performed at 80-100 volts for 3h until the bromophenol blue dye front reached the end of the plate. The gel was viewed and recorded in the gel documentation system. The scoring of bands of RAPD and ISSR markers were recorded as 1 for the presence and 0 for the absence of the concerned pattern of the amplified product which was manually done.

The size (in base pairs) of each locus was also recorded aligning with the DNA ladder. Polymorphic information content (PIC) or average heterozygosity was calculated by the formula given by Roldan- Ruiz *et al*. (2000) which is given as PIC = 2fi (1-fi), where, fi= frequency of the amplified allele and 1 - fi = the frequency of the null allele.

Study of population structure:

For analyzing the population structure and kinship, a subset of 42 (30 RAPD and 12 ISSR) dominant markers were selected (Table B). To assume the genetic structure in the dataset Model-based cluster analysis was performed with the help of Statistical software STRUCTURE V. 2.2 (Pritchard *et al*., 2000). The number of recognized subpopulations (K) was set from two to twenty and this practice was repeated 3 to 5 times. For each run, burn-in and MCMC iterations were set to 50,000 and 100,000, respectively and a model without

admixture and correlated allele frequencies was used. ΔK method (Evanno *et al*., 2005) was used for the determination of the number of the subpopulation. Based on affiliation probabilities, genotypes assignment was done for the individual subpopulation. A genotype was given to a specific subpopulation, with which it has more than 80 per cent association chance and treated as ''admixture''. The effect of population structure on phenotypic traits was investigated employing multiple regression analysis.

Linkage disequilibrium analysis and marker-traits associations:

The pair-wise recombination co-efficient (R^2 values) was calculated by the software Tassel 2.1 Bradbury *et al.* (2007) to estimate the level of LD between markers on sampled genotypes. Based on significant LD ($p < 0.05$), the percentage of the marker combinations was calculated for the total number of marker combinations. Marker combinations with significant LD at a different level of R^2 values ranging 0-1 were calculated thereby tabulating them in a range.

The marker-trait association for different traits, the best fit model was found using the lowest mean square difference (MSD) value among the four available models such as *viz.,* Naive, Q, K and Q+K. The expected P values used for MSD calculation were obtained by the equation given by Stich *et al.* (2008), in which the rank of an observed P-value divided by the total number of markers. The best models were selected for this study, considering the lowest MSD between observed and expected P values of all marker loci and percentage of observations below nominal level ($\alpha = 0.05$) in a P (expected)-P (observed) plot.

The statistical software TASSEL V. 3.0 was used to estimate the associations between molecular markers and trait, using the following models: GLM (general linear model-naive model), the general linear model including Q-matrix derived from STRUCTURE (Q-model) the mixed linear model based on the kinship matrix (K-model) and the mixed linear model based on both the Q-matrix and the kinship matrix (Q+K-model).

Table B : List of RAPD and ISSR markers used for whole genome-wide screening, population structure and kinship analysis

Sr. No.	Code	Nucleotide sequence (5'-3')	Sr. No.	Code	Nucleotide sequence (5'-3')
RAPD primers					
1.	OPT06	CAAGGGCAGA	16.	OPP-8	ACATCGCCTA
2.	OPT09	CACCCCTGAG	17.	OPP-10	TCCCGCCTAC
3.	OPT11	TTCCCCGCGA	18.	OPP-12	AAGGGCGAGT
4.	OPT12	GGGTGTGTAG	19.	OPP-15	GGAAGCCAAC
5.	OPT15	GGATGCCACT	20.	OPW-1	CTCAGTGTCC
6.	OPT18	GATGCCAGAC	21.	OPW-2	ACCCCGCCAA
7.	OPT20	GACCAATGC	22.	OPW-5	GGCGGATAAG
8.	OPJ19	GGACACCACT	23.	OPW-6	AGGCCCGATG
9.	OPJ20	AAGCGGCCTC	24.	OPW-8	GACTGCCTCT
10.	OPA18	AGGTGACCGT	25.	OPW-12	TGGGCAGAAG
11.	OPP-1	GTAGCACTCC	26.	OPW-14	CTGCTGAGCA
12.	OPP-2	TCGGCACGAC	27.	OPW-15	ACACCGGAAC
13.	OPP-3	GTGATACGCC	28.	OPW-16	CAGCCTACCA
14.	OPP-4	GTGTCTCAGG	29.	OPW-17	GTCCTGGGTT
15.	OPP-5	CCCCGGTAAG	30.	OPW-18	TTCAGGGCAC
ISSR primers					
1.	ISSR 808	AGAGAGAGAGAGAGAGC	7.	UBC-5	TCTCTCTCTCTCTCTCA
2.	ISSR 814	CTCTCTCTCTCTCTCTA	8.	UBC-6	ACACACACACACACACG
3.	ISSR 815	CTCTCTCTCTCTCTCTG	9.	UBC-7	TCTCTCTCTCTCTCTCG
4.	ISSR 816	CACACACACACACACAA	10.	UBC-8	AGAGAGAGAGAGAGAGC
5.	UBC-2	GAGAGAGAGAGAGAGAA	11.	UBC-9	GAGAGAGAGAGAGAGAYT
6.	UBC-4	CACACACACACACACAA	12.	UBC-13	CACACACACACACACARG

RESULTS AND DISCUSSION

The results obtained from the present investigation as well as relevant discussion have been summarized under following heads :

Phenotypic variation:

GBH exhibited a very high significant ($p < 0.01$) positive correlation with bark thickness, fresh weight and dry weight. Correlation co-efficient (r) value was 0.61 for bark thickness, 0.62 for bark fresh weight and 0.64 for bark dry weight.The assembled accessions comprising 140 trees/genotypes were characterized for girth at breast height (GBH) and bark thickness, fresh and dry weight (Table 1). The average value for the assembled

Table 1: Summary of the descriptive statistics for the four phenotypic traits examined during the study

Population	Trait	Minimum	Maximum	Mean	SD
AP	GBH (cm)	21	60	35.75	9.34
	Bark thickness (cm)	0.4	1.4	0.78	0.23
	Bark fresh weight (g)	10.72	34.73	20.54	6.10
	Bark dry weight (g)	3.78	15.05	77.12	3.03
AS	GBH (cm)	25	25	25	0
	Bark thickness (cm)	0.5	0.5	0.5	0
	Bark fresh weight (g)	15.67	15.67	15.67	0
	Bark dry weight (g)	5.46	5.46	5.46	0
CG	GBH (cm)	19.5	52	32.62	10.11
	Bark thickness (cm)	0.4	1.1	0.72	0.20
	Bark fresh weight (g)	9.88	29.43	18.30	5.11
	Bark dry weight (g)	3.15	10.94	6.88	2.35
JH	GBH (cm)	23	57	36.66	7.36
	Bark thickness (cm)	0.5	1.5	0.77	0.21
	Bark fresh weight (g)	12.11	41.21	19.82	5.42
	Bark dry weight (g)	4.17	17.84	7.17	2.29
MH	GBH (cm)	20	69	34.14	10.65
	Bark thickness (cm)	0.4	1.1	0.71	0.18
	Bark fresh weight (g)	11.43	30.21	18.38	4.16
	Bark dry weight (g)	3.95	12.25	6.51	1.74
MP	GBH (cm)	24.5	27	25.75	1.76
	Bark thickness (cm)	0.6	0.8	0.7	0.14
	Bark fresh weight (g)	14.89	26.31	20.6	8.07
	Bark dry weight (g)	5.44	9.94	7.69	3.18
OD	GBH (cm)	33	46	40.5	5.5
	Bark thickness (cm)	0.8	1.2	0.95	0.19
	Bark fresh weight (g)	17.84	31.66	23.84	5.73
	Bark dry weight (g)	6.99	10.39	8.98	1.43
UK	GBH (cm)	20.5	39	29.42	7.46
	Bark thickness (cm)	0.4	1.1	0.77	0.26
	Bark fresh weight (g)	8.75	26.85	18.65	7.10
	Bark dry weight (g)	3.30	9.96	6.88	2.70
UP	GBH (cm)	26	47.5	35.08	7.90
	Bark thickness (cm)	0.5	1.5	0.91	0.34
	Bark fresh weight (g)	15.88	38.30	23.73	8.09
	Bark dry weight (g)	5.22	14.36	8.44	3.20

accessions was 34.9±9.0 cm for girth, 0.8±0.2 cm for bark thickness, 19, 581±5,578 mg fresh weight/100cm^2 bark and 7,163±2402 mg dry weight/100cm^2 bark (Fig. 1).

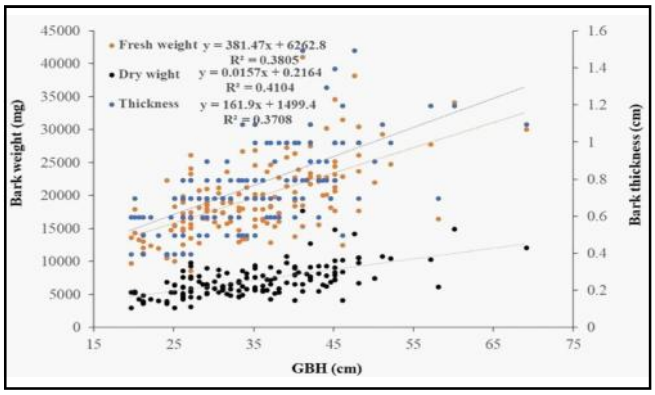


Fig. 1: Correlation and linear regression equation for GBH with bark thickness bark fresh weight and bark dry weight in *T. arjuna* accessions. Correlation co-effi cients are significant at p<0.01(n-2=140)

Analysis of population structure:

Structure analysis employing four models, *viz*., admix and no-admix with independent or correlated allele frequencies was performed. The consistent, reproducible and robust population structure results were obtained with admix with independent allele frequencies (Fig. 2 A, C) in congruence with correlated allele frequencies (Fig. 2 B, D) supporting two sub-population clusters (K=2) to which all sampled accessions belonged. The proportion of both sub-population clusters in the accession genomes across nine states and five agro-climatic zones are depicted in bar plot (Fig. 3).

Genome-wide LD and association study:

In the course of whole genome screening of the genotypes, a total of 17 primers (11 RAPD and 6 ISSR markers) out of 42 seemed to be polymorphic. A total of 118 loci were detected ranging from four to ten with an average of 7.09 loci per primer for RAPD and 6.66 for ISSR were identified. The Polymorphic content information values of the two dominant markers used in this study ranged from 0.02 to 0.37 with an average value of 0.23 for RAPD and 0.027 to 0.35 with an average value of 0.31 for ISSR. The LD plot for r^2 and *D* values between the markers showed the presence of significant LD between the linked other than the unlinked markers.

In the present study on *T. arjuna*, 700 pairs of

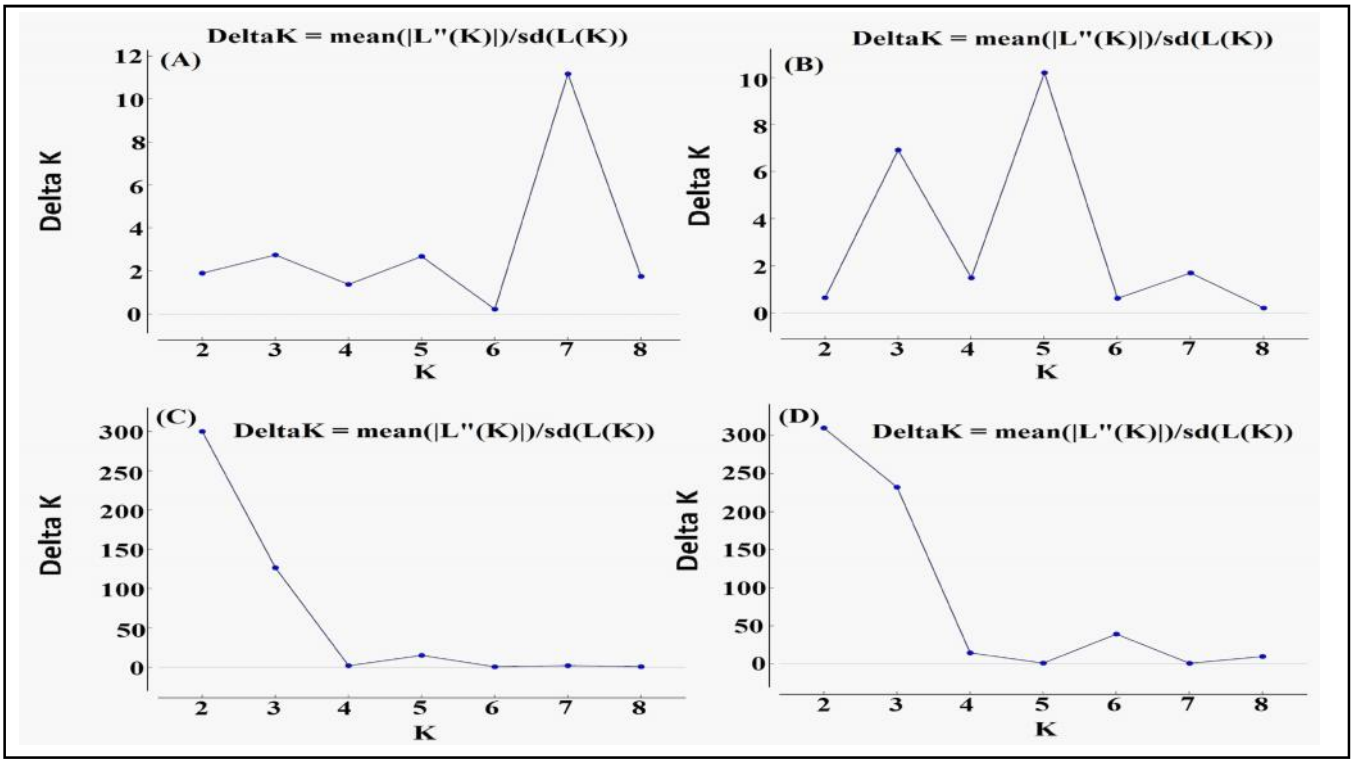


Fig. 2: Depiction of clusters (Delta K values) obtained structure analysis employing no-admix (A, B) and admix (C, D) with allele frequencies independent (A, C) and allele frequencies correlated (B, D). The combined RAPD and ISSR marker data set used for the analysis

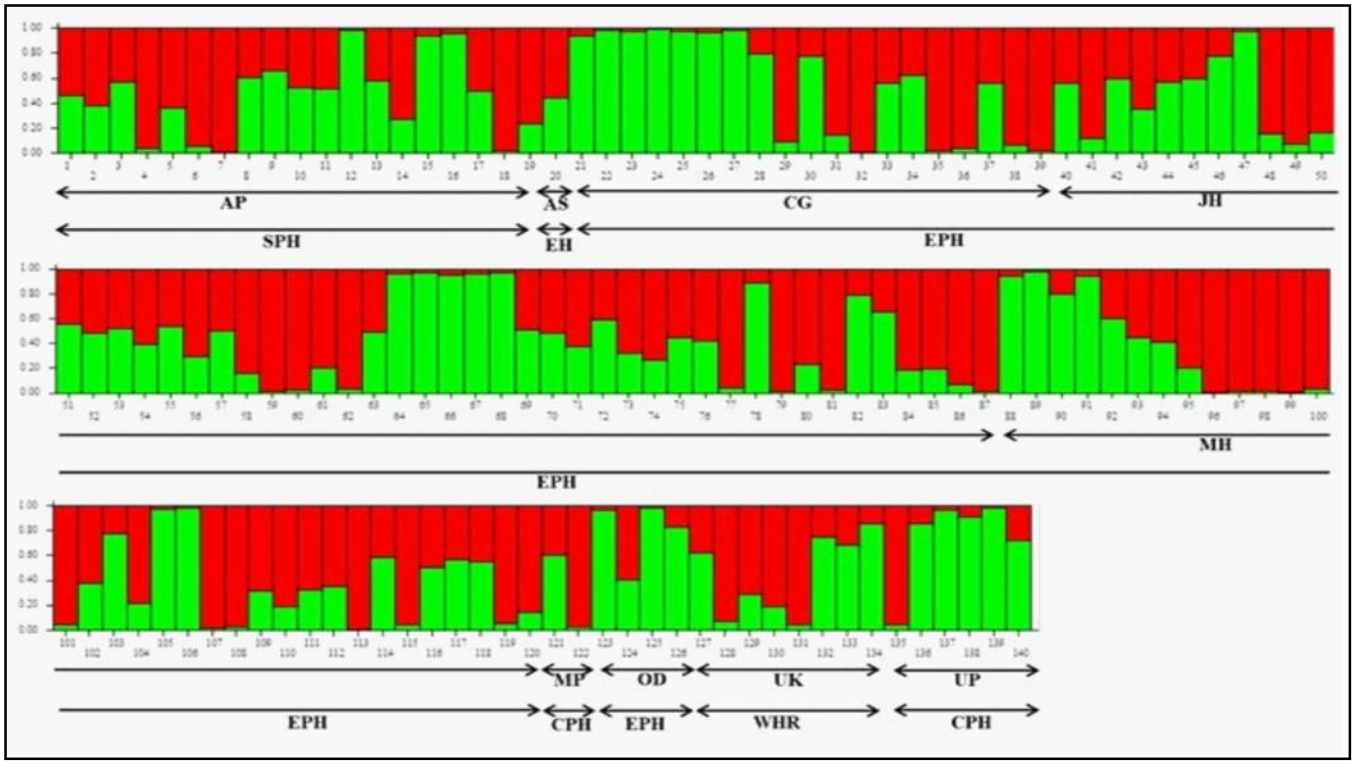


Fig. 3: Bar plot depicting proportion of detected two clusters in sampled accessions of *T. arjuna*

RAPD and ISSR loci out of 4625, *i.e.* 15 per cent exhibit significant LD at $p<0.01$ or R^2 value around 0.4 (Table 3). The R^2 value between 0.2 – 0.25 is usually considered as the minimum level of LD expected between two markers combination or between two loci that are significantly associated (Fig. 4). Results of LD obtained by RAPD and ISSR markers are comparable and within range of, those reported in the earlier limited publication available with dominant markers, e.g. 3.64 per cent (Remington *et al.*, 2001) and 17.28 per cent (Campoy *et al.*, 2016). It may be asserted that with weak population structure and existence of adequate LD in the sampled accessions under investigation, *T. arjuna* germplasm

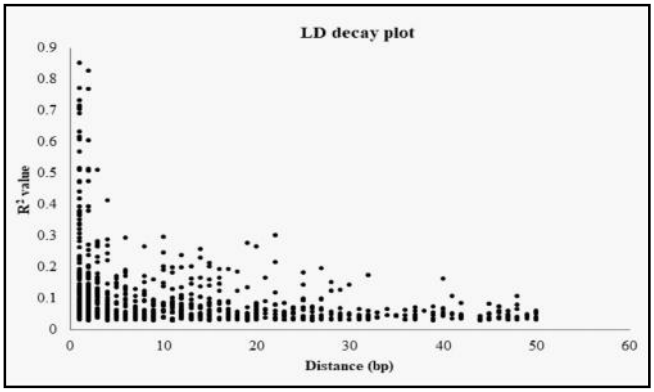
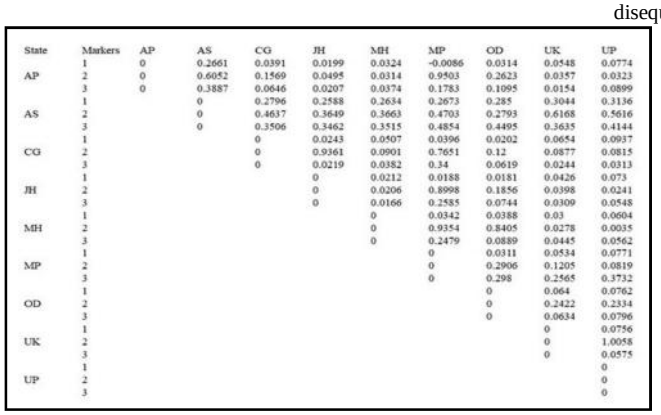


Fig. 4: LD decay lot for detected RAPD and ISSR marker pairs on genome of *T. arjuna*

Table 2: Analysis of molecular variance (AMOVA) for genetic variability partitioning within and among populations for sampled *T. arjuna* accessions under RAPD, ISSR and RAPD + ISSR markers sytems

Source of variation	SS	VC	V%	F_{ST}
RAPD				
Accessions among populations	281.335	0.405	3.264	0.0326
Accessions within population	3149.079	12.019	96.735	
ISSR				
Accessions among populations	160.208	0.146	1.790	0.0179
Accessions within population	2098.568	8.009	98.209	
RAPD + ISSR				
Accessions among populations	438.275	1.28506	6.174	0.0617
Accessions within population	5292.025	19.527	93.825	



State	Markers	AP	AS	CG	JH	MH	MP	OD	UK	UP
	1	0	0.2661	0.0391	0.0199	0.0324	-0.0086	0.0314	0.0548	0.0774
AP	2	0	0.6052	0.1569	0.0495	0.0314	0.9503	0.2623	0.0357	0.0323
	3	0	0.3887	0.0646	0.0207	0.0374	0.1783	0.1095	0.0154	0.0899
	1		0	0.2796	0.2588	0.2634	0.2673	0.285	0.3044	0.3136
AS	2		0	0.4637	0.3649	0.3663	0.4703	0.2793	0.6168	0.5616
	3		0	0.3506	0.3462	0.3515	0.4854	0.4495	0.3635	0.4144
	1			0	0.0243	0.0507	0.0396	0.0202	0.0654	0.0937
CG	2			0	0.9361	0.0901	0.7651	0.12	0.0877	0.0815
	3			0	0.0219	0.0382	0.34	0.0619	0.0244	0.0313
	1				0	0.0212	0.0188	0.0181	0.0426	0.073
JH	2				0	0.0206	0.8998	0.1856	0.0398	0.0241
	3				0	0.0166	0.2585	0.0744	0.0309	0.0548
	1					0	0.0342	0.0388	0.03	0.0604
MH	2					0	0.9354	0.8405	0.0278	0.0035
	3					0	0.2479	0.0889	0.0445	0.0562
	1						0	0.0311	0.0534	0.0771
MP	2						0	0.2906	0.1205	0.0819
	3						0	0.298	0.2565	0.3732
	1							0	0.064	0.0762
OD	2							0	0.2422	0.2334
	3							0	0.0634	0.0796
	1								0	0.0756
UK	2								0	1.0058
	3								0	0.0575
	1									0
UP	2									0
	3									0

Fig. 5: Pair-wise Nei's (1983) genetic distance matrices for accessions of nine states obtained by (1) RAPD, (2) ISSR and (3) RAPD+ ISSR markers

bank is suitable for the marker-trait association. In all situations, *T. arjuna* accessions belonging to all locations, except those of Assam, Madhya Pradesh, Uttar Pradesh cluster together. Alternatively the accessions from Assam, Madhya Pradesh and Uttar Pradesh remained as outliers and are discrete. The clustering has been well supported by high bootstrap values (Fig. 7 A, B, C). As for individual accessions, the detected clusters exhibit excessive mixing with all locations, indicating adequate gene flow across *T. arjuna* meta-population in India.

For population hierarchies, F_{ST} values between populations by RAPD, ISSR and RAPD + ISSR markers have been found very low, indicating high gene flow and

Table 3: Detection of significant LD in paired loci obtained for RAPD and ISSR marker systems in *T. arjuna* accessions

Marker system	Total paired loci (No.)	Significant LD among paired loci (No.)		
		p<0.001	p<0.01	Total
RAPD	2,675	90	180	270
ISSR	1,950	170	260	430
Total	4,626	260	440	700

Table 4: Marker loci exhibiting significant association with arjunolic acid content under GLM and MLM analyses

Marker loci	GLM		MLM	
	F-value	p-value	F-value	p-value
OPP15_3	4.044436	0.046265	4.928739	0.028045
*UBC002_5	4.291211	0.040172	-	-
OPP15_5	4.382934	0.038130	3.96817	0.048342
OPP03_2	4.644277	0.032893	-	-
OPP15_6	4.823684	0.029741	5.229991	0.023722
OPP15_1	5.010378	0.026797	4.114627	0.044438
OPT18_4	6.070491	0.014975	4.347685	0.038902
*UBC004_1	6.832830	0.009943	-	-
*UBC002_4	3.978430	0.048056	-	-
OPP08_8	4.100011	0.044812	-	-
OPP08_3	4.248959	0.041152	-	-
OPT18_5	4.409249	0.037565	-	-
*UBC009_2	4.707229	0.031748	-	-
OPT09_6	4.820986	0.029786	4.132668	0.043981
OPP08_7	5.089619	0.025642	-	-
OPP08_1	5.452986	0.020977	4.438784	0.036941
OPT09_4	5.541235	0.019984	-	-
*UBC007_1	5.630811	0.019027	-	-
OPP03_8	6.246957	0.013613	5.106977	0.025396
OPP10_8	6.380187	0.012670	4.924269	0.028115
OPP03_3	6.812944	0.010049	-	-
OPT18_6	7.694787	0.006305	5.772263	0.017611
OPT09_8	-	-	3.933228	0.049326

*ISSR loci

low differentiation. The F_{ST} value in *T. arjuna* accessions is comparable to those in *T. bellerica* populations (Dangi *et al.*, 2014). A similar range of F_{ST} has also been recorded in *Torreya jackii* (Li *et al.*, 2007). Wright (1978) identifies four population genetic differentiation groups based F_{ST} values. A range of 0.05-0.15 F_{ST} to which *T. arjuna* accessions belong indicates moderate genetic differentiation of population. Nei's G_{ST} value, an analog F_{ST}, is high, *i.e.* 0.28 in *T. arjuna* accessions (Fig.5). Nei's G_{ST} value is supposed to depreciate recurrence of inbreeding (selfing) in the population. RAPD and ISSR markers being dominant ones are unable to detect and deduct inbreeding within the population and thereby provide an over estimation of Nei's G_{ST}. Therefore, F_{ST} values obtained from AMOVA are more reliable than Nei's G_{ST} value when dominant markers namely RAPD and ISSR are employed due to their inability to distinguish heterozygous loci from homozygous dominant loci in the meta-population (Table 2). It is inferred that the sampled *T. arjuna* accessions depict moderate population genetic differentiation.

As mentioned above, Low F_{ST} also indicates the ease of gene flow. As a result, *T. arjuna* accessions appear to have freely traded gene exchange that points out the existence of ineffective geophysical barriers in locations of meta-population of *T. arjuna*. AMOVA analysis also allocates maximum genetic variability to accessions within populations (locations) and a minuscule fraction (< 7%) of genetic variability to among populations due to the free exchange of gene flow (Table 3).

In the present study, it detected the variations for arjunolic acid content across the 140 accessions under study. Arjun is a species that have the narrow genetic base and less variability but the moderate variability was detected, due to the presence of genetically diverse germplasm of arjun collected from the different site of origin including indigenous and exotic. Due to the occurrence of genetic variability, the population seemed to be suitable for population structure and association mapping study.

Model-based analysis of population structure was shown that the occurrence of nine subpopulations in the

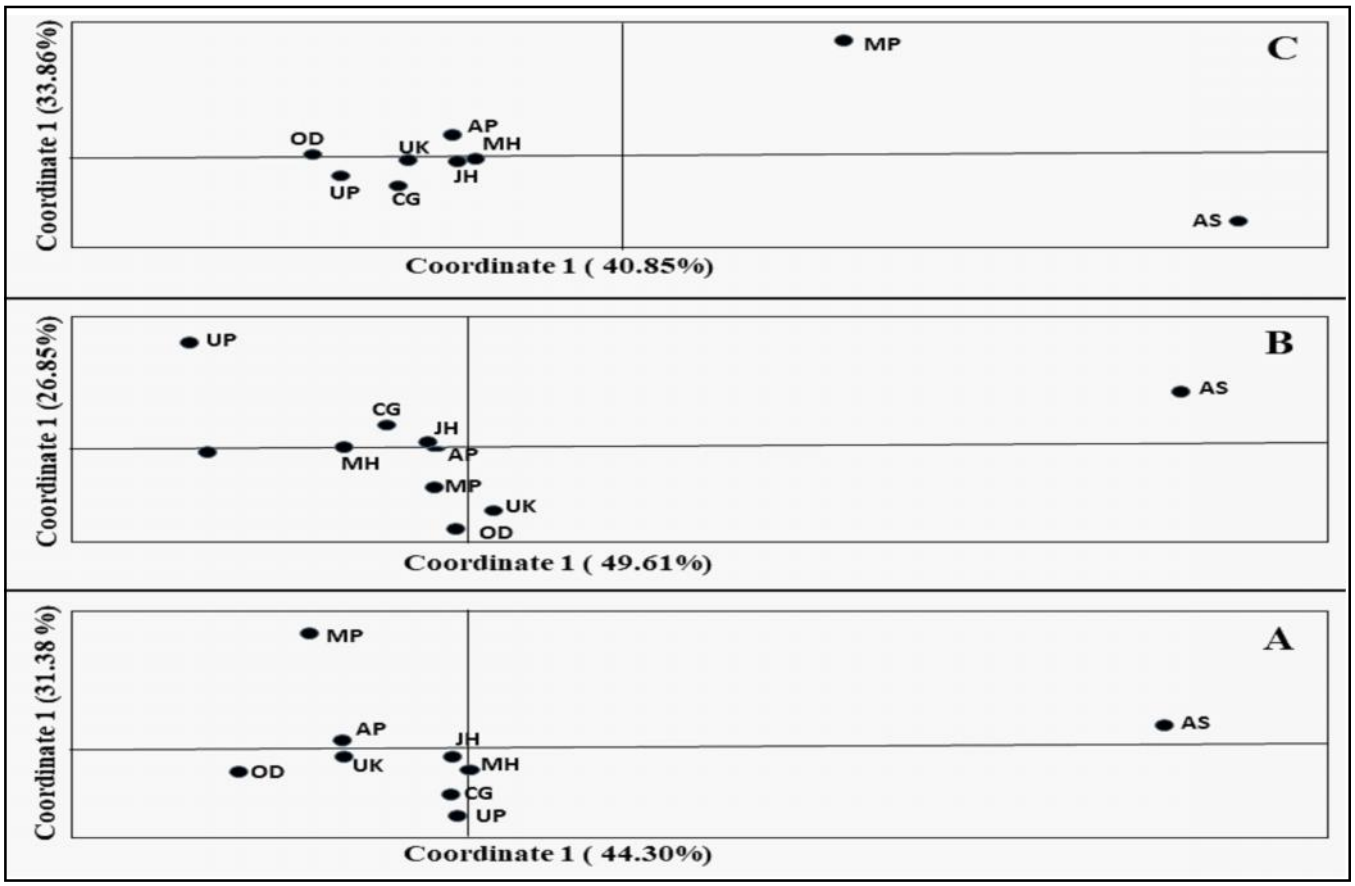


Fig. 6: Clustering of accessions belonging to nine states obtained for RAPD markers (A), ISSR markers (B) and RAPD + ISSR markers (C) on two coordinates, employing principal co-ordinate analysis (PCoA)

genotypes studied. These subpopulations resembled the major geographic regions of their origin or collection. Nei's genetic distance matrices obtained by RAPD, ISSR and RAPD + ISSR markers were used for clustering by PCoA (Fig. 6). The RAPD markers based genetic distance matrix made four clusters, three clusters comprising accessions from a single state, *i.e.* AS, MP, and OD and the fourth largest cluster with accessions from the remaining six states distributed among two coordinates cumulatively contributing to 75.68 per cent separation (Fig. 6A). The ISSR markers based genetic distance matrix (Fig.6B) also produced similar four clusters as those of RAPD markers based genetic distance matrix. However, the fourth largest cluster was not very compact. The clusters were distributed among two coordinates cumulatively accounting for 76.19 per cent separation. RAPD + ISSR markers based genetic distance matrix consolidated only three clusters, two clusters comprising accessions from a single state (AS and MP) and the third largest-cluster with accessions from seven states among both coordinates cumulatively accounting for 74.71 per cent separation (Fig. 6C).

RAPD and ISSR markers combined generated UPGMA dendrogram that separated accessions of Assam (AS) state from those of all other states supported by very high bootstrap value, *i.e.* 800 or 80 per cent. The large cluster comprising accessions from eight (AP, MH, CG, UK, JH, UP, OD, MP) states was further split into two asymmetrical sub-clusters supported by a high bootstrap value of 700 or 70 per cent. One sub-cluster had accessions from one state, *i.e.* MP, while the large sub-clusters comprised accession from seven states (AP, MH, CG, UK, JH, UP, OD). The large sub-cluster further divided into two asymmetrical groups supported by a high bootstrap value, *i.e.* 600 or 60 per cent. The small group contained accessions from a single state, *i.e.* OD. The large group comprised accessions from six states (AP, MH, CG, UK, JH, UP) that again split into two asymmetrical sub-groups supported with an adequate bootstrap value of 500 or 50 per cent. The small sub-group had accessions from a single state, *i.e.* UP. The large sub-group contained accessions from five states (AP, MH, CG, UK, JH), whose further resolution was not supported by adequate bootstrap value (Fig. 7 A, B, C). RAPD+ISSR marker data was applied for construction of UPGMA dendrogram for individual accessions that revealed their admixing across different states and agro-climatic zones, indicating very little population differentiation (Fig.8).

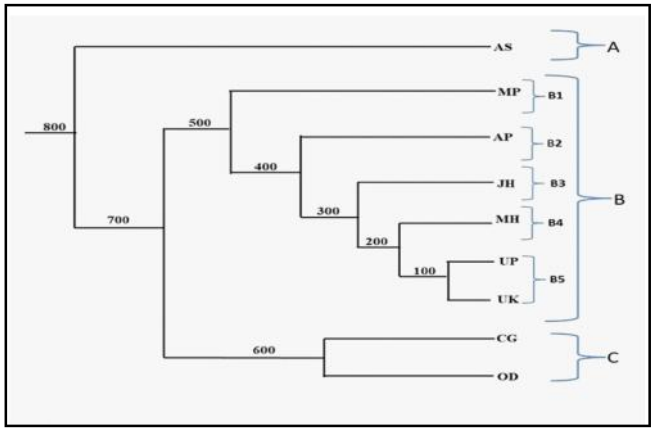


Fig. 7.B: UPGMA dendrogram based on Nei's genetic distance of ISSR markers for clustering of *T. arjuna* accessions belonging to nine states. Bootstrap values are shown on nodes

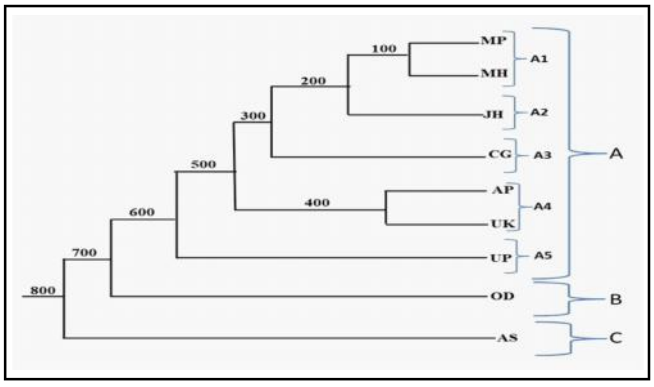


Fig. 7.A: UPGMA dendrogram based on Nei's genetic distance of RAPD markers for clustering of *T. arjuna* accessions belonging to nine states. Bootstrap values are shown on nodes

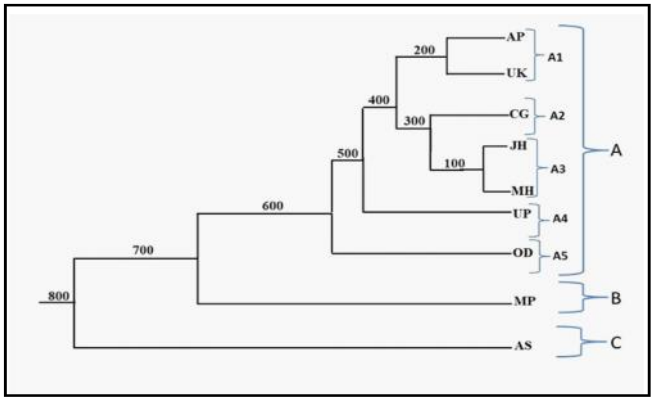


Fig. 7.C: UPGMA dendrogram based on Nei's genetic distance of RAPD + ISSR markers for clustering of *T. arjuna* accessions belonging to nine states. Bootstrap values are shown on nodes

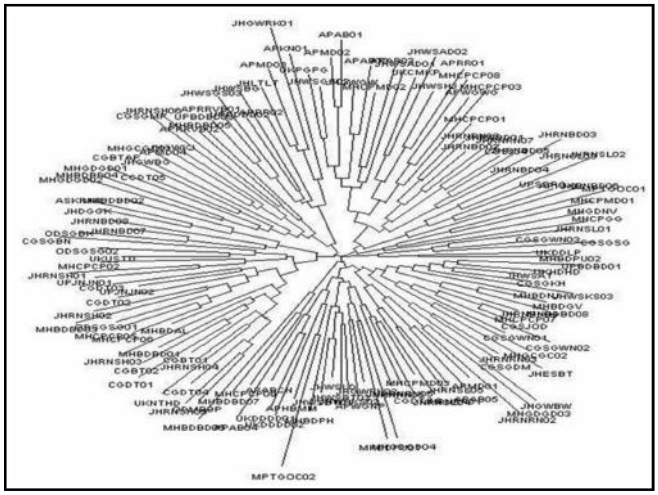


Fig. 8: UPGMA dendrogram showing the genetic relatedness among 140 *T. arjuna* accessions based on RAPD + ISSR markers

We studied that, LD between linked markers decreased with increased distances, and a significant drop was noticed beyond > 20bp distance on the genome of *T. arjuna* accessions. Long-range LD observed in this study demonstrated the potential for genome-wide association mapping with fewer markers in Arjun; however, such steps are certain to compromise the resolution. A genetic distance of ~10 cM, the high LD was found. In different studies done by the researchers different values of LD were observed. Tenaillon *et al.* (2001) studied, in *Zea mays*, genome wide LD decays was noticed within 200-1,500 bp and by Maccaferri *et al.* (2005), in durum wheat, the value of LD decay estimated was 10-20 cM. The cross-pollinated (Arjun) species keep LD to a shorter longer genetic distance as compared to self-pollinated crop species.

According to Zhu (2008), In plant genetics, association mapping is the best approach for the detection of genetic variants that cause the complex traits in the natural population. Nevertheless, the spurious associations in AM are the main concern and it influenced by the size of the population and no uniformity among the individuals. The General linear model (GLM) and mixed linear model (MLM) are the two main two major methodologies that used to see the association of a marker with the trait of interest. Neumann *et al*. (2011) described in his studies that the association between a marker with the trait is always higher in the GLM approach as compared to the MLM approach. In the population structure study only Q matrix used by the GLM approach to find out the marker-trait associations. Though, MLM at the same time describe both population structure as well as kinship (genetic relatedness among individuals), and hence, more consistent. We found that OPP03, OPP10, OPP15, OPT09 and OPT18 significantly associated with the arjunolic acid (both in GLM and MLM) (Table 4). The marker OPP15 was found to be significantly associated with the arjunolic acid in GLM studies. Similar results were also reported by Neumann *et al*. (2011) and Zhu *et al*. (2008) in wheat. The phenotypic variation defined by GLM was higher than the MLM. MLM method was confirmed to be suitable in controlling spurious associations than GLM. We studied that the loci controlled the complex traits can be mapped by the association mapping, the existence of weak population structure and adequate LD among paired loci *vis-à-vis* significant positive correlations among morphometric traits and the observed variability in arjunolic acid content in *T. arjuna* accessions indicate their suitability for marker-trait association investigations. GWA confers the advantage of detecting unknown loci associated with the trait over candidate gene sequencing approach.

Conclusion:

Finally, we conclude that there is plenty of genetic diversity to be found in the nine subpopulations of arjun genotypes collected from different agroclimatic regions. We studied wide-range of LD that shows the probability for association mapping in arjun. RAPD and ISSR markers exhibit a significant association with morphometric traits and arjunolic content in the sampled *T. arjuna* accessions. However, AM with morphometric traits and arjunolic acid content maintains a preponderance of RAPD markers over ISSR markers. Interestingly, both general and mixed linear models detect association of ISSR markers with characteristics of bark whereas ISSR markers have a significant association with bark arjunolic acid content in the GLM model only. Loci detected by OPJ20, OPT17 (RAPD) markers and UBC004, UBC815 (ISSR) markers display significant association with morphometric traits and those by OPP08, OPP10, OPP15, OPT17 (RAPD) markers with GBH, bark thickness/fresh weight/ dry weight and arjunolic acid content. In contrast, loci detected by OPT09 (RAPD) markers are significantly associated with arjunolic acid content only. The highest number of marker loci in both GLM and MLM models are associated with arjunolic acid content followed by bark thickness > bark dry weight > GBH and bark fresh weight. MLM model eliminates

association of ISSR markers as false discovery, except UBC15_3 for bark thickness, bark fresh weight and bark dry weight and also discovers additional association of UBC004_2 with bark fresh weight and OPT09_8 with arjunolic acid from the debris of spurious association. Consequently, MLM model discards false association to the tune of 14 per cent in GBH, 28 per cent in bark fresh weight and 50 per cent in each of bark thickness, bark dry weight and arjunolic acid content in the bark. By using this information one can choose superior accessions that can be used in further research for the mass multiplication of Arjun with the highest content of arjunolic acid.

Acknowledgment:

I acknowledge with thanks, Director of Tropical Forest Research Institute, Jabalpur, to provide me necessary facilities during the present work. I am thankful to Dr. S. A. Ansari, my Ph.D. supervisor for his unconditional support and encouragement during this work.

REFERENCES

Abdurakhmonov, I.Y., Kohel, J. R., Yu, J.Z., Pepper, A.E., Abdullaev, A. A., Kushanov, F.N., Salakhutdinov, I.B., Buriev, Z.T., Saha, Sz., Scheffler, B.E. and Jenkins, J.N. (2008). Molecular diversity and association mapping of fibre quality traits in exotic *G. hirsutum L. germplasm*. *Genomics*, **92** : 478–487.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**: 2633–2635.

Campoy, J.A., Lerigoleur-Balsemin, E., Christmann, H., Beauvieux, R., Girollet, N. and Quero-García, J. (2016). Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. *BMC Plant Bio.*, **16** (49): 1-15.

Cardon, R.L. and Bell, J.I. (2001). Association study designs for complex disease. *Nat. Rev. Genet.*, **2** : 91- 99.

Cervera, M.T., Sewell, M.M., Faivre-Rampant, P., Storme, V., Van Montagu, M. and Boerjan, W. (2004). Genome mapping in populus. In: Kumar S, Fladung M (eds) *Molecular genetics and breeding of forest trees*. Haworth's Food Products Press, New York, pp. 387-410.

Dangi, B., Varsha, K.K., Kothari, S.L. and Kachhwaha, S. (2014). Micropropagation of *Terminalia bellerica* from nodal explants of mature tree and assessment of genetic fidelity using ISSR and RAPD markers. *Physiol. Mol. Bio. Plants.*, **20** : (4): 509–516.

Evanno, G., Regnaut, S. and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.*, **14**: 2611-2620.

Freeman, J.S., Potts, B.M., Downes, G.M., Pilbeam, D., Thavamanikumar, S. and Vaillancourt, R. (2013). Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. *New Phytol.*, **198** : 1121-1134.

Holland, J.B. (2007). Genetic architecture of complex traits in plants. *Curr. Opin. Plant. Biol.*, **10**: 156–161.

Lespinasse, D., Rodier-Goud, M., Grivet, L., Leconte, A., Legnate, H. and Seguin, M. (2000). A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme markers. *Theor. Appl. Genet.*, **100** : 127–138.

Li, J. and Jin, Z. (2007). Genetic variation and differentiation in *Torreya jackii* Chun, an endangered plant endemic to China. *Plant Sci.*, **172**: 1048-1053.

Maccaferri, M., Sanguineti, M.C., Noli, E. and Tuberosa, R. (2005). Population structure and long-range linkage disequilibrium in a drum wheat elite collection. *Mol. Breed.*, **15**: 271–289.

Mather, K. A., Caicedo, A.L., Polato, N.R., Olsen, K.M., Mc Couch, S. and Purugganan, M. D. (2007). The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics*, **177** (4): 2223-2232.

Neumann, K., Kobiljski, B., Dencic, S., Varshney, R.K. and Borner, A. (2011). Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Mol. Breed.*, **27**: 37–58.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000). Inference of population structure using multi-locus genotype data. *Genetics.*, **155**: 945–959.

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M. M. and Buckler, E.S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *PNAS*, **98** : 11479-11484.

Roldán-Ruiz, I., Dendauw, J., Van Bockstaele, E., Depicker, A. and De Loose, M. (2000). AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). *Mol. Breed.*, **6** : 125–134.

Scalfi, M., Troggio, M., Piovani, P., Leonardi, S., Magnaschi, G., Vendramin, G. G. and Menozzi, P. (2003). A RAPD, AFLP and SSR linkage map, and QTL analysis in European beech (*Fagus sylvatica* L.). *Theo. Appl. Genet.*, **108** (3): 433–441.

Sewell, M.M., Sherman, B.K. and Neale, D.B. (1999). A consensus map for loblolly pine (*Pinus taeda* L.). I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. *Genetics*, **151**: 321-330.

Stich, B., Mohring, J., Piepho, H. P., Heckenberger, M., Buckler, E. S. and Melchinger, A.E. (2008). Comparison of mixed-model approaches for association mapping. *Genetics*, **178**: 1745–1754.

Tenaillon, M.I., Sawkins, M.C., Long, A. D., Gaut, R. L., Doebley, J. F. and Gaut, B. S. (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* L.). *Proc. Natl. Acad. Sci. USA.*, **98**: 9161–9166.

Wright (1978). *Evolution and the genetics of populations: variabilitywithin and among natural populations*. Univ Chicago Press, Chicago.IL.

Zhu, C., Gore, M.A., Buckler, E.S. and Yu, J. (2008). Status and prospects of association mapping in plants. *Plant Genome*, **1**: 5-20.