

Association mapping of important agronomic traits in *Mucuna pruriens* (L.) DC.

Patrush Lepcha

Department of Botany, Sikkim University, P.O. Tadong, Gangtok, Sikkim 737102

Mahesh S

Department of Biotechnology, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka 562157

Leelambika M

Department of Biotechnology, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka 562157

Mahammad Jaheer

Department of Biotechnology, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka 562157

Ratan Chopra

Department of Plant and Microbial Biology, University of Minnesota, Saint Paul MN 55108

Vikas Belamkar

Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583

N Sathyanarayana (✉ sathyan@cuk.ac.in)

Department of Botany, Sikkim University, P.O. Tadong, Gangtok, Sikkim 737102

Research Article

Keywords: Genetic diversity, marker-trait association, population structure, seed traits

Posted Date: June 14th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3038281/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The *Mucuna pruriens* (L.) DC. can meet three burgeoning needs of tropical agriculture: low-cost protein source, high-value medicinal plant, and an excellent green manure cover crop. But like other underutilized crops, it lacks modern breeding resources. Identifying marker-trait associations (MTAs) can prompt marker-assisted breeding paving a way for development of the improved varieties. Recent studies have demonstrated feasibility of identifying MTAs using a small number of accessions (<100) triggering the hope for resource poor crops. Taking cues from there, we characterized a panel of 70 *M. pruriens* accessions across two consecutive years and performed association analysis for 16 phenotypic traits using 66 genic-microsatellite markers. The results showed significant phenotypic ($P < 0.05$) and genetic diversity (Shannon's information index, $I = 0.62$) in our germplasm collection. Many of the tested traits were highly heritable (42.86% to 99.93%). A total of 15 MTAs were detected at an adjusted significance level of $P < 5.55 \times 10^{-3}$ for nine economic traits such as – seed length, seed thickness, seed width, hundred seed weight, seed yield per plant, inflorescence length, flower buds per inflorescence, flower length, and petiole length with phenotypic variance explained (PVE) ranging from 14.72 to 31.12%. Their annotation using the *Arabidopsis* genome database affirmed reliability of our MTAs. This is the first report on association mapping in *M. pruriens* and results are expected to galvanize marker-assisted breeding as well as mining of candidate genes in this promising legume.

Introduction

Tapping the potential of neglected and underutilized species (NUS) is vital to reduce over reliance on staples and improve global food security. The prospects of *Mucuna pruriens* (L.) DC. for the tropical agriculture is well recognized (Pugalenthi et al. 2005). But little modern genetic tools are available for the improvement of its key agronomic and medicinal traits including the edible protein and anti-Parkinson's drug L-DOPA (l-3,4-dihydroxyphenylalanine). Identification of marker-trait association can accelerate marker-assisted breeding programs and propel its use for commercial cultivation.

The *M. pruriens* is classified under the phaseloid clade of Leguminosae/Fabaceae. It is a self-pollinated annual species with a diploid genome ($2n = 2x = 22$) of approximately 1281 to 1361 Mbp (Sastrapradja et al. 1974; Sathyanarayana et al. 2017), and occurs both in wild (var. *pruriens* and var. *hirsuta*) and cultivated (var. *utilis*) forms (Wilmot-Dear 1987; Sasidharan 2004). With a seed to seed duration of 90–120 days, it flourishes well under the acidic soil ($\text{pH} < 5.8$), elevation below 1600 m, warm (19–27 °C) and moist tropical climatic conditions (Pugalenthi et al. 2005; Kumar and Saha 2013). The plant is native to Eastern India and Southern China (Duke 1981; Wilmot-Dear 1987); but now distributed across the Asia, Americas, West Indies, Africa and the Pacific Islands (Fung et al. 2011). It is adapted to the less fertile and dry soils (Siddhuraju et al. 2000), produces a seed yield of about 1.3–2.4 t/ha (Kumwenda and Gilbert 1998; Gurumoorthi et al. 2003), and possesses disease-resistant (Eilitta et al. 2002), nematocidal (Carsky and Ndikawa 1998) as well as allelopathic properties (Fujii et al. 1991). *M. pruriens* (var. *utilis*) contain high seed protein (20–30%) akin to other edible legumes such as soybean, pigeon pea, chickpea (Kumar et al. 1991; Hira and Chopra 1995; Mang et al. 2016) and thus offers a cheap source of edible protein.

Genetic diversity is a key determinant in the success of any crop improvement programs. In *M. pruriens*, earlier works found wide-ranging phenotypic and trait-specific variability in different germplasm collections (Mahesh and Sathyanarayana 2011, 2015; Sathyanarayana et al. 2012). DNA markers such as: randomly amplified polymorphic DNA (RAPD) (Padmesh et al. 2006; Leelambika et al. 2010; Patil et al. 2016), inter simple sequence repeat (ISSR) (Patil et al. 2016; Chinapolaiah et al. 2018), amplified fragment length polymorphism (AFLP) (Sathyanarayana et al. 2011; Mahesh and Sathyanarayana 2015; Tripathi et al. 2018) and simple sequence repeat (SSR) (Shetty et al. 2015; Sathyanarayana et al. 2017; Kumar et al. 2019) have reinforced these findings. In the realm of molecular breeding, studies aiming at the development of the linkage maps and tagging of agronomic traits using AFLP markers were initiated (Capo-chichi et al. 2004; Mahesh et al. 2016). However, these efforts didn't advance beyond preliminary findings, perhaps due to resource constraints often faced by such crops. It is well known that conventional linkage analysis is both time and resource intensive due to the requirement of bi-parental populations. It also suffers from the efficacy issues due to low-resolution and limited allelic variation (Flint-Garcia et al. 2005; Yu and Buckler 2006; Oraguzie and Wilcox 2007; Abdurakhmonov and Abdurakhmov 2008).

Alternatively, association mapping offers a powerful method for the identification of marker-trait association as it determines relationship between phenotypic variations and genetic polymorphisms taking into account naturally occurring or historical recombination (Ambreen et al. 2018). It has been successfully used for identifying marker-trait associations (MTAs) in different legume species (Lui et al. 2017; Singh et al. 2017; Zhao et al. 2017; Vaijayanthi et al. 2018). However, to date, not even a single study has explored this opportunity in *M. pruriens*. Among the molecular markers, EST-SSRs have been proved to be more reliable for the association analysis as they are directly linked with the genes expressing a trait (Varshney et al. 2005). Several species-specific EST-SSR markers have been developed using RNA-seq analysis recently in *M. pruriens* (Sathyanarayana et al. 2017). The SSR markers are also the marker of choice for plant breeders due to co-dominant nature, multiallelic, high polymorphism, and greater genome abundance (Kalia et al. 2011).

Conventionally, core collection, which is available only for the selected/mainstream crop species has done well as association mapping panels (Lui et al. 2017; Zhao et al. 2017; Ambreem et al. 2018). Alternatively, for crops that lack these resources like NUS, a diverse panel of accessions from the germplasm collection can be used, provided they demonstrate the presence of high genetic variance, low kinship association among its individuals, and lack of strong population structure (Hu et al. 2014; Mahajan et al. 2017). Association analysis is usually carried out on a large number of populations (> 300). But in cases of NUS, where the numbers are limited, it is still possible to explore MTAs (Soumya et al. 2021). Recently, several genome-wide association studies (GWAS) have been successfully carried out in population sizes between 60 and 150 (Lehnert et al. 2018; Ma et al. 2018; Rohila et al. 2020). Taking cues from these studies, in the present study we evaluated a panel of seventy diverse *M. pruriens* accessions for suitability as an association mapping panel and performed association analysis on sixteen agronomically important traits using microsatellite markers.

Materials and methods

Plant material and phenotypic evaluation

We used a subset of 70 *M. pruriens* accessions from our germplasm collection as an association mapping panel. This comprised 11 accessions of *M. pruriens* var. *utilis* (cultivated variety), 28 of var. *pruriens* (wild variety), and 31 of var. *hirsuta* (wild variety). They represented three large geographic regions in India - Eastern, Southern, and West-Central India (Table S1). Sixteen traits related to seed (seed length, seed width, seed thickness, seed yield per plant, hundred seed weight); pod (pod length, pod width, number of pods per cluster, number of pods per plant); inflorescence (inflorescence length, flower buds per inflorescence, flower length, pedicel length), and biochemical attributes (L-DOPA, total protein, total carbohydrate) were evaluated for two consecutive years (2014 and 2015) as per the descriptor (Sathyanarayana et al. 2012). For data scoring, we used manual measurements using Vernier caliper, weighing balance, and counting as relevant for the trait. The L-DOPA content was estimated as per the method of Daxenbichler et al. (1972). The total protein and total carbohydrate contents were estimated using Lowry's (Lowry et al. 1951) and Anthrone (Hedge and Hofreiter 1962) methods, respectively. All statistical analyses such as one-way analysis of variance (ANOVA), correlation, and principal component analysis (PCA), were performed using the R program (R Core Team 2014).

Estimation of variance components and broad-sense heritability

To estimate the variance, genotypic and phenotypic coefficients of variation were evaluated using Syukur et al. (2012) as follows:

$$\sigma^2G = (MSG - MSE) / r$$

$$\sigma^2P = \sigma^2G + \sigma^2E/r$$

Where, σ^2G = genotypic variance, σ^2P = phenotypic variance, σ^2E = environmental variance (error mean square from the analysis of variance); MSG = mean square of genotypes; MSE = error mean square; r = number of replications.

The broad-sense heritability for each trait was estimated according to Allard (1960) as follows:

$$H^2 = (\sigma^2G / \sigma^2P) \times 100$$

Where, H^2 = heritability in broad-sense; σ^2G = genotypic variance; σ^2P = phenotypic variance.

DNA isolation and SSR genotyping

Leaf tissues from 15-days old seedlings were used for the DNA isolation as described previously (Doyle and Doyle 1990). Genotyping was carried out using a sub-set of 90 species-specific genic SSR primer pairs (Table S2) chosen from Sathyanarayana et al. (2017). Polymerase chain reaction (PCR) mixture

included: template DNA (50ng/μl), primers (1μM each), dNTPs (2.5mM), Taq polymerase (1U), PCR buffer (1X), and MgCl₂ (1.5mM), maintaining the final volume of 25μl. DNA amplification condition comprised of: initial denaturation at 94°C for 3 min followed by 35 cycles of 94°C for 30s, appropriate annealing temperature (*T_m*) for 30 s, and an extension at 72°C for 20 s with a final extension of 7 min at 72°C. Subsequently, one (1) μl each of PCR amplicons generated with different dye-labeled primers were mixed with 2.95μl distilled water, 7μl of formamide and 0.05μl of the GeneScan™ 500 LIZ® Size Standard. Denatured DNA was size fractionated with capillary electrophoresis. As each SSR primer-pair produced multiple amplicons, primer-pairs are referred to as SSR marker and, amplicons as bands or alleles here onwards. We applied stringent filtering criteria (minor allele frequency > 0.05 and missing percentage less than 20%) for the marker bands produced. Those SSRs fulfilling these criteria were retained and used for the further analysis.

Marker attributes and genetic diversity

Marker attributes such as polymorphism information content (PIC) and major allele frequency (MAF) were calculated by PowerMarker v 3.25 (Liu and Muse 2005). The total number of alleles (*N_a*), effective number of alleles (*N_e*), observed heterozygosity (*H_o*), expected heterozygosity (*H_e*), percentage of polymorphic loci (%P), gene flow (*N_m*), and Hardy-Weinberg equilibrium (HWE) were assessed by GenAlEx v 6.5 (Peakall and Smouse 2012). HP-Rare v 1.0 (Kalinowski 2005) was used to determine private/rare allelic richness per locus (*R_p*) based on rarefaction approach. The genetic diversity indices such as - Shannon's information index (*I*), Nei's gene diversity (*h*), and total genetic diversity (*H_T*) were determined by using POPGENE v 1.32 (Yeh et al. 1999).

Population structure, relative kinship, and genetic relationship

Population structure was analyzed using the Bayesian clustering method implemented in STRUCTURE v 2.3.4 (Pritchard et al. 2000). For each accession, the proportion of ancestral contribution was estimated using the admixture model and correlated allele frequencies. K-values ranging from 1 – 10 were tested with 10 independent replications, 100,000 lengths of the burn-in period, and 200,000 Markov Chain Monte Carlo (MCMC) repetitions for each K. The optimal K-value was obtained by using STRUCTURE analysis results with STRUCTURE HARVESTER (Earl 2012). Accessions were assigned to a subgroup depending on the Q-value (membership proportion), and if the Q-value is < 80%, they were termed as admixtures. Analysis of molecular variance (AMOVA) and pairwise *F_{ST}* of geographic population and subpopulations were carried out using GenAlEx v 6.5 (Peakall and Smouse 2012) with 1,000 permutations. To estimate the relatedness between the individuals, the kinship coefficients (*F_{ij}*; individual level) was estimated following Loiselle et al. (1995). TASSEL v 5.0 (Bradbury et al. 2007) was used to generate the kinship coefficient matrix among all pairs of accessions. The kinship heat map was obtained using GAPIT, R package (Wang et al. 2014). Genetic relationships based on the distance was arrived through the construction of unrooted neighbor-joining (NJ) dendrogram and principal coordinate analysis (PCoA)

using Darwin v 6.0 (Perrier and Jacquemoud-Collet 2006). The reliability of the NJ dendrogram was tested with the bootstrap value of 1000 replicates.

Association mapping and annotation of the MTAs

TASSEL v 5.0 (Bradbury et al. 2007) was used for the association analysis. Mixed linear model (MLM) was employed to determine the MTAs. The MLM is considered superior over the general linear model (GLM), as GLM incorporates only population structure (Q matrix) which often results in false-positives; instead, MLM incorporates both Q and K matrix which overcomes this limitation (Yu et al. 2006). Normally in such studies, Bonferroni multiple test correction ($P = 0.05/n$; n is the number of markers used in the study) is applied to obtain the threshold P -value as it provides a stringent cut-off to avoid false-positives. However, it assumes that the loci are independent, which is not always true, given that certain loci may be in linkage disequilibrium. Therefore, to avoid loss of beneficial MTAs, in addition to Bonferroni correction, we also used a less stringent criterion, $P = 1/n$ as additional cut-off. This approach has been beneficially used in some earlier studies (Li et al. 2012; Yang et al. 2014; Xu et al. 2018). The Manhattan plots and quantile-quantile (Q-Q) plots MTAs were generated using CMplot, R package (Yin et al. 2021). Further, to test the reliability of our MTAs, the putative function of each trait-associated marker was determined using the *Arabidopsis thaliana* (L.) Heynh. genome database (<https://www.arabidopsis.org>). For this, the transcripts of SSR markers were used as a query sequence and matched against the reference genome sequences of *A. thaliana* with the BLAST tool, and then the top hit was selected as the putative gene/function of the respective marker. An $e\text{-value} \leq 1e-05$ was used as a threshold.

Results

Phenotypic variability

The one-way ANOVA revealed significant variability ($P < 0.05$) for majority of the phenotypic traits evaluated in our association mapping panel, except for pedicel length, and L-DOPA content. Inflorescence length and flower length showed the highest (CV = 84.97%) and the lowest variations (CV = 6.90%) (Table 1). The values of correlation coefficients indicated significantly high positive correlation ($R > 0.70$) between some important seed, pod and inflorescence traits (Fig. 1). For instance, hundred seed weight was strongly correlated ($R > 0.70$) with seed thickness, seed width, seed length, and pod length, and moderately correlated ($0.30 < R < 0.70$) with pod width and flower length. Similar trend was observed for seed yield as well.

Table 1
Variability and broad-sense heritability of sixteen seed-based traits in *M. pruriens* association mapping panel

Traits (unit of measurement)	Range	Average $\pm$ SE	CV(%)	ANOVA	H ² (%)
Inflorescence length (cm)	1.60–93.00	18.85 $\pm$ 2.05	84.97	**	96.93
No of flower buds per inflorescence	4.25–83.75	25.54 $\pm$ 2.40	73.37	**	62.23
Flower length (cm)	4.05–5.18	4.61 $\pm$ 0.04	6.90	ns	52.94
Pedicle length (cm)	0.30–4.50	0.60 $\pm$ 0.07	80.45	ns	42.86
Pod length (cm)	6.43–14.03	8.80 $\pm$ 0.21	18.71	***	99.93
Pod width (cm)	1.43–2.55	1.82 $\pm$ 0.03	14.07	***	98.14
No of pods per cluster	2.50–28.00	10.83 $\pm$ 0.77	55.71	*	98.83
No of pods per plant	27.25–405.75	147.43 $\pm$ 11.66	61.77	**	96.77
Seed length (mm)	7.73–16.90	10.88 $\pm$ 0.28	20.19	***	96.62
Seed width (mm)	6.15–12.77	8.18 $\pm$ 0.20	19.17	***	93.92
Seed thickness (mm)	3.01–9.61	5.33 $\pm$ 0.18	26.08	***	97.27
Seed yield per plant (gm)	41.13–1136.60	303.61 $\pm$ 30.94	79.60	***	93.34
Hundred seed weight (gm)	12.83–146.36	41.71 $\pm$ 3.75	70.31	***	95.93
L-DOPA (%)	0.95–3.24	2.07 $\pm$ 0.06	22.81	ns	62.50
Total protein (%)	15.07–29.19	22.94 $\pm$ 0.42	14.14	*	98.58
Total carbohydrate (%)	6.54–28.25	14.60 $\pm$ 0.68	36.44	**	95.71
SE: Standard error; CV: Coefficient of variation; ANOVA: One-way analysis of variance; H ² : Broad-sense heritability					
ns: Statistically not significant at $P < 0.05$					
***: Statistically significant at $P < 0.001$					
**: Statistically significant at $P < 0.01$					
*: Statistically significant at $P < 0.05$					

In principal component analysis (PCA), the first five principal components (PCs) explained 81.39% of the total phenotypic variance of which PC1 accounted for 34.15% and the PC2 accounted for 24.68% (Fig. 2) (Table S3). The resultant scatter plot distinctly separated the cultivated and wild varieties. However, we couldn't find clear-cut separation for the two wild varieties - var. *pruriens* and var. *hirsuta* (Fig. 2). We observed moderate to high estimates (42.86–99.93%) of broad-sense heritability (H²) for the 16

phenotypic traits (Table 1). Traits such as: number of pods per cluster, number of pods per plant, hundred seed weight, seed yield per plant etc., showed a very high heritability values ranging from 93.34 to 99.93%. Among the biochemical traits, total protein (98.58%) and total carbohydrate (95.71%) recorded high heritability, and L-DOPA content revealed a moderately high heritability (62.50%).

Genetic diversity

Initially, 90 markers tested on 70 accessions of *M. pruriens* (Table S4) produced 298 alleles with an average of 3.39 alleles per locus. At this stage, we applied a stringent filtering criterion to retain only 180 SSR alleles contributed by 66 markers based on minor allele frequency > 0.05, along with 61 accessions which showed maximum missing site < 20% for further analysis. The major allele frequency (MAF, revealed by filtered SSR data) varied from 0.50 to 0.94, with an average of 0.73 indicating high genetic diversity and polymorphism at the observed loci. The heterozygosity (Ho) values ranged from 0.00 to 0.59 with a mean value of 0.12 and the gene diversity (He) varied from 0.03 (MPU_122) to 0.73 (MPU_42) with a mean value of 0.36. The summary of the marker attributes at different loci, as well as individual markers retained after stringent filtration are given in Table 2 and Table S5, respectively.

Table 2
Marker attributes of sixty-six SSR markers used in the study

Major allele frequency (MAF)	
Average	0.73
SSR marker with lowest MAF	0.5 (MPU_52)
SSR marker with highest MAF	0.95 (MPU_20, MPU_122)
Observed heterozygosity (Ho)	
Average	0.12
SSR marker with lowest Ho	0.00 (MPU_39, MPU_47, MPU_80)
SSR marker with highest Ho	0.59 (MPU_90)
Expected heterozygosity (He)	
Average	0.36
SSR marker with lowest He	0.03 (MPU_122)
SSR marker with highest He	0.73 (MPU_42)
Polymorphism Information Content (PIC)	
Average	0.29
SSR marker with the lowest PIC	0.09 (MPU_20)
SSR marker with the highest PIC	0.38 (MPU_52)

The polymorphism information content (PIC), which provides relative informativeness of each marker, ranged from 0.09 (MPU_22) to 0.38 (MPU_52) with an average of 0.29 (Table 2). The major proportion of SSR markers (43) in our study was moderately polymorphic (PIC-value > 0.25) with PIC-values ranging from 0.26 to 0.38 (Table S5). The mean and private allelic richness per marker were 2.82 and 0.06, respectively. The majority of the SSR markers deviated from the Hardy-Weinberg equilibrium (HWE) at a significance level of $P < 0.001$ and very few markers deviated at $P < 0.01$ (MPU_46 and MPU_115) and $P < 0.05$ (MPU_92) (Table S5). The SSR markers MPU_122 and MPU_130 showed no significant deviation.

The Shannon's information index ($I = 0.62$) and Nei's gene diversity ($h = 0.32$) suggests high genetic diversity within our association mapping panel. Total gene diversity value ($H_T = 0.34$) (Table S5) substantiated this finding. For more perceptive analyses, we subdivided the association panel accessions into two groups: first one based on three regional gene pools (Eastern, Central-West, and Southern India) and the second one on three botanical varieties (*var. utilis*, *var. pruriens*, and *var. hirsuta*). We estimated genetic diversity statistics independently for each one of these gene pools (Table 3). In the case of regional gene pools, I varied from 0.56 to 0.67 and h varied from 0.28 to 0.34. Accessions from West-Central India ($I = 0.67$, $h = 0.34$) and Southern India ($I = 0.64$, $h = 0.34$) were more diverse as compared to ones from Eastern India ($I = 0.56$, $h = 0.31$). The percentage of polymorphic SSR loci was 96.97% for the accessions of South India followed by 89.39% and 87.88% for the accessions West-Central and Eastern India, respectively. Among the three varietal gene pools, I varied from 0.44 to 0.72 and h varied from 0.28 to 0.36 (Table 3). Wild varieties, *var. pruriens* ($I = 0.72$, $h = 0.36$) and *var. hirsuta* ($I = 0.62$, $h = 0.34$) were genetically more diverse than the cultivated *var. utilis* ($I = 0.44$, $h = 0.28$). The percentage of polymorphic SSR loci was 98.48%, 96.97%, and 72.73% in *var. pruriens*, *var. hirsuta*, and *var. utilis*.

Table 3
Genetic diversity estimates for the regional gene pools and botanical varieties within *M. pruriens* association mapping panel

Regional gene pool	N	Na	Ne	I	h	Ho	He	%P
Eastern India	15	2.48	1.65	0.56	0.31	0.09	0.33	87.88
West-Central India	11	2.56	1.95	0.67	0.34	0.13	0.40	89.39
Southern India	35	3.15	1.75	0.64	0.34	0.13	0.35	96.97
Varietal gene pool	N	Na	Ne	I	h	Ho	He	%P
<i>M. pruriens</i> var. <i>utilis</i>	8	1.98	1.53	0.44	0.28	0.06	0.27	72.73
<i>M. pruriens</i> var. <i>pruriens</i>	28	3.11	1.92	0.72	0.36	0.12	0.41	98.48
<i>M. pruriens</i> var. <i>hirsuta</i>	25	3.00	1.72	0.62	0.34	0.14	0.35	96.97
N: No. of accessions; Na: Total no. of alleles; Ne: Effective no. of alleles; I: Shannon's information index; h: Nei's gene diversity; Ho: observed heterozygosity; He: Expected heterozygosity; %P: Percentage of polymorphic loci								

Population structure

STRUCTURE analysis (ΔK vs. K plot) revealed a sharp peak at $K = 2$, indicating the presence of two genetic groups in our association mapping panel (Fig. 3a). Forty-eight accessions showed >80% membership within respective subgroups, and 13 were admixtures with <80% of shared ancestry (Fig. 3b). Majority of the 17 accessions in the subgroup-1 (designated as MpSTR-I), and 44 accessions in the subgroup-2 (MpSTR-II) aligned with Eastern (14) and Southern India (34) respectively. Accessions from the West-Central India did not form an independent subgroup, but merged with one of the two major clusters.

To ascertain the partitioning of variation, we performed AMOVA on three population groups assumed based on: (a) geographic distribution (b) varietal affiliation and (c) subgroups identified in STRUCTURE analysis (Table 4). All three revealed higher within gene pool variance vis-à-vis among gene pools. The ratios for within to between gene pools variance were 93:7 in case of geographical distribution, 97:3 for varietal affiliations, and 89:11 for structure derived sub-populations. We estimated population differentiation (F_{ST}) within each group separately. The results revealed low to moderate differentiation in groups based on the varietal affiliation ($F_{ST} = 0.032$) and geographic origin ($F_{ST} = 0.070$) and high genetic differentiation ($F_{ST} = 0.105$) in groups based on STRUCTURE analysis (Table 4).

Table 4
Population differentiation and AMOVA based on geographic distribution, varietal affiliation, and STRUCTURE analysis within *M. pruriens* association mapping panel

Geographic distribution					
Sources of variation	Estimated Variance	Percentage of Variance	P-value*	F _{ST}	
Among populations	1.048	7%	0.001	0.070	
Within populations	13.841	93%	0.001		
Total	14.889	100%	0.001		
Varietal affiliation					
Sources of variation	Estimated Variance	Percentage of Variance	P-value*	F _{ST}	
Among populations	0.463	3%	0.001	0.032	
Within populations	14.172	97%	0.001		
Total	14.635	100%	0.001		
STRUCTURE					
Sources of variation	Estimated Variance	Percentage of Variance	P-value*	F _{ST}	
Among populations	1.619	11%	0.001	0.105	
Within populations	13.794	89%	0.001		
Total	15.413		0.001		
*: With 999 data permutations					

To determine the genetic relationship, we constructed the NJ dendrogram based on a simple matching coefficient and performed PCoA. Two main clusters were revealed by NJ algorithm, which were named MpNJ-I and MpNJ-II (Fig. 4). Similar to STRUCTRE results, majority of the 31 accessions in MpNJ-I were derived from Southern India (25). But considerable mixing was observed among the 33 accessions in MpNJ-II, as 15 of them belonged to Eastern India, eight were from West-Central India, and seven were from Southern India. The accessions in the NJ tree and PCoA plots (PC1 vs. PC2) are color-coded from the information of two subgroups identified using the STRUCTURE analysis (MpSTR-I and MpSTR-II). We didn't find any indication of grouping based on geographic origin or varietal affiliation as accessions representing these groups were pretty much mixed up among both the clusters.

In principal coordinate analysis (PCoA), the principal axes 1 and 2 explained 28 and 26% of the total variance (Fig. 5). The dispersion was relatively homogeneous in all the four quadrants signifying the diverse nature of the accessions. The clustering pattern in PCoA largely corresponded with that of NJ-dendrogram. We also estimated the relative kinship between these accessions to measure their

relatedness. About 55.29% of kinship coefficient (F_{ij}) values between any two accessions were within 0 to 0.05 and there was subsequent reduction in frequency with increase in kinship value (Fig. 6). The heat map (Fig. 7) revealed substantial differences among the accessions.

Association mapping

For association analysis, we employed the mixed linear model (MLM) which integrates both population structure (Q matrix) and kinship (K matrix) to avoid false-positive MTAs. Only two MTAs involving two SSR markers (MPU_83 and MPU_122) passed the stringent Bonferroni adjusted threshold $P < 2.77 \times 10^{-4}$ ($P = 0.05/180$, where 180 is the number of markers used in the analysis). These two associations were for the seed yield per plant (phenotypic variance explained, PVE = 31.12%) and the pedicel length (PVE = 25.07%) (Fig. 8a, b).

We also used a less stringent criterion of $P < 5.55 \times 10^{-3}$ ($1/180$) as cut-off for the reasons explained in the method section. By applying, a total of 15 additional MTAs related to inflorescence length, flower buds per inflorescence, flower length, pedicel length, seed length, seed width, seed thickness, hundred seed weight, and seed yield per plant were identified (Table 5). The PVE for these MTAs ranged from 14.72 to 31.12%. These MTAs were contributed mainly by 10 SSR markers: MPU_19, MPU_42, MPU_54, MPU_57, MPU_58, MPU_83, MPU_89, MPU_108, MPU_111, and MPU_122. We found four MTAs contributing for hundred seed weight, two each for seed length, seed width, and pedicel length, and at least one each for inflorescence length, flower buds per inflorescence, flower length, seed thickness, and seed yield per plant. The highest PVE (31.12%) was observed in MTA between SSR marker MPU_83 and important agronomic trait, seed yield per plant. This crucial MTA was also spotted at Bonferroni adjusted threshold $P < 2.77 \times 10^{-4}$ (Fig. 8a). The Manhattan and Q-Q plots of MTAs at $P < 5.55 \times 10^{-3}$ are given in Fig. S1 and S2, and the MTAs significant at threshold $P < 0.05$ and PVE > 10% are provided in Table S6.

Table 5
Marker-trait associations identified using mixed linear model (MLM) at $P < 5.55 \times 10^{-3}$ with phenotypic variance explained (PVE) > 10%

Trait	Marker	P-value	PVE (%)
Inflorescence length	MPU_58	4.50×10^{-3}	14.72
Flower buds per inflorescence	MPU_19	2.33×10^{-3}	19.43
Flower length	MPU_19	4.09×10^{-3}	15.48
Pedicel length	MPU_57	3.03×10^{-3}	15.81
	MPU_122	2.60×10^{-4}	25.07
Seed length	MPU_42	4.35×10^{-3}	15.43
	MPU_83	4.99×10^{-3}	14.94
Seed width	MPU_42	5.20×10^{-3}	15.61
	MPU_83	2.55×10^{-3}	18.48
Seed thickness	MPU_54	5.04×10^{-3}	15.15
Seed yield per plant	MPU_83	1.00×10^{-4}	31.12
Hundred seed weight	MPU_54	2.88×10^{-4}	26.7
	MPU_89	1.10×10^{-3}	20.27
	MPU_108	7.68×10^{-4}	22.08
	MPU_111	2.78×10^{-3}	16.85
PVE: Phenotypic variance explained			

The association analysis also revealed four SSR markers which associated with multiple agronomic traits ($P < 5.55 \times 10^{-3}$; PVE > 10%). Marker MPU_83 was not only associated with the seed length and seed width, but also with the seed yield per plant. SSR loci MPU_54 was associated with seed thickness and hundred seed weight. MPU_42 was associated with seed length and seed width, and MPU_19 was associated with flower length and flower buds per inflorescence. These MTAs were justified by high the significant correlations ($P < 0.05$) values between these traits of agronomic importance (Fig. 1).

Annotation of marker-trait associations

To test the reliability of our MTAs, we determined putative gene/function for each trait-associated marker ($P < 5.55 \times 10^{-3}$; PVE > 10%) by annotating the related SSR primers using the *Arabidopsis thaliana* (L.)

Heynh. genome database. Out of the 10 trait-associated markers, the BLAST analysis annotated seven markers *viz.*, MPU_19, MPU_42, MPU_54, MPU_89, MPU_108, MPU_111, and MPU_122 (Table 6). Importantly, the SSR marker MPU_19, associated with flower length and flower buds per inflorescence, was annotated as 4-coumarate-CoA ligase 2 and MPU_42 associated with seed length and seed width was annotated for Zinc transporter 7 precursor. Likewise, the SSR marker MPU_54 associated with traits such as seed thickness and hundred seed weight was annotated for sucrose non-fermenting 1- related protein kinase 2 (SnRK2) gene, and MPU_89 associated with hundred seed weight was annotated as 3-dehydroquinate synthase. The SSR MPU_108, associated with hundred seed weight matched with NAC domain-containing protein which is a plant-specific transcription factor (TF) associated with multiple aspects of stress and development. Further, MPU_111 (hundred seed weight) and MPU_122 (pedicel length) were annotated with Ubiquitin-specific protease and LEUNIG homolog (LUG and LUH), respectively.

Table 6

Putative Gene ID/description of the trait-associated markers obtained from *Arabidopsis* annotated genes using blast results

Markers	P-value	Associated Traits	Gene ID/Description	Functions attributed
MPU_19	0.00671	FBpl, FL	4-coumarate-CoA ligase 2 (<i>4CL</i> ; <i>EC 6.2.1.12</i>)	Catalyzes the activation of 4-coumarate and a few related substrates to the respective CoA esters and thus channels the common, phenylalanine-derived building block into the otherwise widely distinct branches of general phenylpropanoid metabolism (Hamburger and Hahlbrock 2004).
MPU_42	0.00326	SL, SW	Zinc transporter 7 precursor (<i>AT2G04032</i>)	Mediates zinc uptake from the rhizosphere (Grotz et al. 1998; Milner et al. 2013).
MPU_54	6.95E-04	ST, HSW	Sucrose nonfermenting 1-related protein kinase 2 (<i>SnRK2</i>)	Involved in the abscisic acid (ABA) signaling and plays a central role in plant stress signal transduction (Feng et al. 2018).
MPU_89	0.00294	HSW	3-dehydroquinate synthase	Catalyzes the transformation of the seven-carbon sugar 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAH7P) into the carbocycledehydroquinate (DHQ) (Negron et al. 2011).
MPU_108	0.00133	HSW	NAC domain-containing protein	Plant-specific transcription factors (TFs) are associated with multiple aspects of the stress and development region (Mathew et al. 2016).
MPU_111	0.00311	HSW	Ubiquitin-specific protease 23 (<i>AT5G57990</i>)	The organ size in plants is reported to be regulated by two reversible processes called ubiquitination and deubiquitinating (Shi et al. 2019).
MPU_122	1.53E-04	PedL	LEUNIG homolog (LUG and LUH)	The two proteins may act cooperatively to coordinate inflorescence architecture through their influences on auxin biosynthesis, transport, and perception (Douglas et al. 2017).
FBpl: Flower buds per inflorescence; FL: Flower length; SL: Seed length; Seed width; ST: Seed thickness; HSW: Hundred seed weight; PL: Pedicel length				

Discussion

Phenotypic diversity and heritability

M. pruriens is an underutilized legume species; hence, its germplasm is maintained by few international and national research institutes/organization such as International Institute of Tropical Agriculture (IITA), Nigeria; United States Department of Agriculture (USDA), the USA; National Bureau of Plant Genetic Resources (NBPGR), India etc., (Jorge et al. 2007). We used our own germplasm collection containing 70 accessions as an association mapping panel. We found significant variability for majority of the flower, pod, seed, and biochemical characteristics in our association mapping panel. This is consistent with earlier reports in different *M. pruriens* germplasm (Gurumoorthi et al. 2003; Kalidas and Mohan 2011; Sathyanarayana et al. 2012). Majority of the seed and pod-based traits revealed significant positive correlation possibly due to the key role played by the pod in the seed development in terms of protection and nutrient source (Bennet et al. 2011). Few accessions (IC-369144, 500113-MH, 500126-KA, 500135-KA, 500178-MH, 500184-MH, and 500190-MH) showed high seed protein content akin to soybean [*Glycin max* (L.) Merr.]. These stocks will be useful in breeding for protein content. But, unlike previous studies (Tripathi et al. 2018; Kumar et al. 2019), we found less variability for the L-DOPA content in our germplasm. The estimation method used (Daxenbichler et al. 1972) by us in this study is old and laborious; thus, it is plausible that poor recovery and associated short comings might have skewed this experiment. Moderately-high to high heritability ($H^2 > 60\%$) were observed for different traits in our study. Two earlier studies found analogous observation in *M. pruriens* for traits such as pod length, pod width, pod weight, hundred seed weight, seed yield per plant, and inflorescence length (Hadapad et al. 2018; Chinapolaiah et al. 2019). Our results reinforce them. High heritability means less influence of the environmental factors rendering phenotypic selection reliable, besides contributing to high additive effect in the breeding programs (Tiwari et al. 2011; Rosmania et al. 2016).

Genetic diversity, population structure and kinship analysis

Initially, we used 90 SSR markers on 70 *M. pruriens* accessions, but after stringent filtration, only 66 SSR markers and 61 accessions were retained for further analysis which generated a moderate PIC-value ($0.25 < \text{PIC} < 0.50$). This corroborates with the earlier report in *M. pruriens* (Kumar et al. 2019) albeit slightly higher allelic range and allele per locus. The result endorses the choice of SSR markers for our study. The Shannon's information index suggests high genetic diversity ($I = 0.62$) in our germplasm collection. This is higher than the values reported using AFLP (0.34) and SSR (0.47) markers in earlier studies (Tripathi et al. 2018; Kumar et al. 2019). In a predominantly self-pollinating species like *M. pruriens*, we expect lower levels of genetic diversity. But higher values recorded here and in some earlier studies (summarized in Sathyanarayana et al. 2016) categorically points to presence of out-crossing in this species as observed initially by Padmesh et al. (2006). This is partly supported by a high average gene flow among the population groups ($N_m = 3.03$). Thus, further studies on pollination mechanism in *M. pruriens* can throw more light on the drivers of genetic diversity in this species. Among the three geographical regions, accessions from West-Central and Southern India revealed higher diversity than that of Eastern India which is consistent with Kumar et al. (2019). Of the three botanical varieties, wild varieties (var. *pruriens* and var. *hirsuta*) were more diverse than the cultivated variety (var. *utilis*) – the finding often documented in many earlier studies (Leelambika and Sathyanarayana 2011;

Sathyanarayana et al. 2012; Tripathi et al. 2018; Kumar et al. 2019). Nonetheless, wide phenotypic and genetic diversity in our germplasm collection signifies its utility for the association analysis.

The low to moderate F_{ST} values (0.032, 0.070) in *M. pruriens* varietal and geographical populations indicated less divergence. The population structure analyses from the distance-based NJ-tree, PCoA, and Bayesian-based STRUCTURE largely conformed to each other in suggesting minimal influence of the geographical origin and/or the varietal affiliation on the grouping of the accessions. High gene flow estimates ($N_m = 3.03$), presence of genetic admixtures and higher within population variance in AMOVA further concur with this. These observations are also chronicled in different germplasm collections in earlier studies (summarized in Sathyanarayana et al. 2016). Put together, these imageries suggest one or more of the-extensive pollen flow across long geographic distance, lack of inter-varietal barriers for hybridization, long seed dispersal, and high seed germination rates are contributing to low divergence of *M. pruriens* gene pools. The resultant returns appear to have shaped it highly adaptable species as evident from its wide and diverse distribution range. Alternatively, this might also suggest incomplete lineage sorting during its diversification as observed in safflower (*Carthamus tinctorius* L.) (Ambreen et al. 2018). However, we warrant a caution here as our experiments are based on the limited sample size. The future studies must examine more samples from a broad geographical range to confirm these hypotheses. Further, we noticed low pairwise kinship estimates between the majorities of the *M. pruriens* accessions in our association mapping panel suggesting weak relationship. Low population divergence and kinship estimates further support the suitability of our panel for an association analysis.

Association mapping and significant marker-trait associations

Among the recent approaches used for fine-scale mapping of the desirable traits, association mapping has produced fast and reliable results in several legume taxa. This includes markers for seed-related traits in peanut (*Arachis hypogaea* L.) (Zhao et al. 2017), iron and zinc concentration in lentil (*Lens culinaris* Medikus) (Singh et al. 2017), frost tolerance in pea (*Pisum sativum* L.) (Lui et al. 2017), flowering, pod yield per plant and fresh pod per pant in lablab [*Lablab purpureus*(L.) Sweet] (Vaijayanthi et al. 2018), crude protein and mineral concentration in alfalfa (*Medicago sativa* L.) (Jia et al. 2017), etc. However, in *M. pruriens* it has never been put to use (Lepcha et al. 2019), perhaps due to lack of resources common to NUS. Thus, in the present study, we attempted association mapping of key agronomic traits in *M. pruriens*. Usually large number of accessions (> 300) are used in the association studies. In comparison, our number is small (61). But several recent studies have demonstrated feasibility of finding right marker-trait associations even using a small number of accessions (Lehnert et al. 2018; Ma et al. 2018; Rohilla et al. 2020; Soumya et al. 2021).

We identified 15 significant MTAs ($P < 5.55 \times 10^{-3}$) using MLM with high PVE ranging from 14.72 to 31.12%. These associations are also benefitted by the moderate to high heritability observed for all the trait-associated markers (H^2 ranging from 42.86 to 97.27%). Four SSR markers were associated with multiple traits. They mostly comprised of seed or inflorescence traits which are highly correlated. Such an

account of a single marker associated with multiple traits can be attributed to closely linked QTLs affecting different traits (Rakshit et al. 2010). Alternatively, it may also be due to the pleiotropic effect of the linked QTLs on different traits (Miller and Rawlings 1967).

Annotation of SSR markers associated with the traits

We ascertained the putative functions of the trait-associated markers by annotating against corresponding genes/loci in the *Arabidopsis thaliana* (L.) Heynh. database (<https://www.arabidopsis.org>). The results produced some useful insights. For instance, the SSR marker MPU_111, associated with hundred seed weight corresponded with the ubiquitin-specific protease 23 gene in *A. thaliana*. Another gene belonging to the same family in rice (*Oryza sativa* L.) viz., ubiquitin-specific protease 23 is known to exert positive regulatory influence on grain width and size (Shi et al. 2019). SSR marker MPU_54 associated with the traits such as seed thickness and hundred seed weight was annotated for sucrose non-fermenting 1-related protein kinase 2 (SnRK2) gene. This gene is involved in the abscisic acid (ABA) signaling during seed germination, dormancy, and seedling growth and has a central role in plant stress signal transduction (Feng et al. 2018). The SSR MPU_108, associated with hundred seed weight was traced to NAC domain-containing protein which is a plant-specific transcription factor (TF) associated with the multiple aspects of the stress and development. It is reported that three NAC TF encoding genes (*ONAC020*, *ONAC026*, and *ONAC023*) are expressed at very high levels during the seed development in rice (*Oryza sativa* L.) and exhibits a strong association with the seed size or weight with the sequence variations located in the upstream regulatory region (Mathew et al. 2016). SSR marker MPU_122 associated with pedicel length was annotated for LEUNIG homolog (LUG and LUH). Likely, these two proteins act cooperatively to coordinate inflorescence architecture through their influence on auxin biosynthesis, transport, and perception (Douglas et al. 2017). The marker MPU_89 associated with hundred seed weight was annotated to 3-dehydroquinate synthase gene which catalyzes the transformation of the seven-carbon sugar 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAH7P) into the carbocycledehydroquinate (DHQ) (Negron et al. 2011). This pathway controls the production of precursors of aromatic amino acids in several prokaryotes, fungi, and plants (Ganem 1978). Scrutinizing its relationship with L-DOPA biosynthesis in *M. pruriens* might throw open fascinating insights on the regulation of aromatic amino acids, particularly Tyrosine production during L-DOPA production. Thus, MTAs identified in the present study will be useful for improving the seed yield and related economic traits in *M. pruriens*. To the best of our understanding, this is the first report on marker-trait association based on association mapping in any *Mucuna* species from anywhere in the world.

But our study has following limitations: (a) Firstly, we have used less number of accessions and SSR marker. This is because *M. pruriens* is an underutilized legume species, thus, their germplasm collection and genomic resources are limited. (b) Secondly, we have used single environment data for our analysis. So, there is a need to analyze the stability and utility of these marker-trait associations with larger number of accessions and markers under additional environments through multi-location trials. The future studies must also focus on developing MTAs for other important traits relevant to this crop, particularly indeterminate growth habit - a major detriment in commercial cultivation of this crop.

Conclusion

In summary, a high phenotypic/genotypic diversity, low population divergence, moderate to high heritability of the evaluated traits, and low kinship estimates in our association mapping panel confirm its utility for the association mapping. Fifteen MTAs identified for important traits with PVE > 10% from MLM, offers a strong resource base for improvement of this crop. Their reliability tested through annotation against the *Arabidopsis* genome database lends further credence to this. Thus, the results from this study offer significant ground work for future marker-assisted breeding and mining candidate genes for important agronomic traits in *M. pruriens*-a promising underutilized legume for the tropics.

Declarations

Acknowledgement

The authors acknowledge financial supports from Science and Engineering Research Board (SERB); Department of Science and Technology (DST), Government of India; Department of Biotechnology (DBT), Government of India; and Ministry of Tribal Affairs (MoTA), Government of India.

Conflict of Interest

The authors declare that they have no conflict of interest in the publication.

Author Contributions

NS conceptualized the study, design experiment, acquired funding, and supervised the study; PL performed data analysis and wrote the manuscript; MS, LM, and MJ performed field investigation and compiled data; RC investigated the gene annotation; VB reviewed and edited the manuscript. All authors read and approved the final version of the manuscript.

Data availability

The datasets generated during and/or analyzed during the current study are available in the supplementary information section.

References

1. Abdurakhmonov IY, Abdukirimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. *Int J Plant Genom* 2008:574927. <https://doi.org/10.1155/2008/574927>
2. Allard RW (1960) Principles of plant breeding. John Wiley and Sonc Inc., New York.
3. Ambreen H, Kumar S, Kumar A, Agarwal M, Jagannath A, Goel S (2018) Association mapping for important agronomic traits in safflower (*Carthamus tinctorius* L.) core collection using microsatellite markers. *Front Plant Sci* 9:402. <https://doi.org/10.3389/fpls.2018.00402>

4. Bennett EJ, Roberts JA, Wagstaff C (2011) The Role of the pod in seed development: Strategies for manipulating yield. *New Phytol* 190:838-853. <https://doi.org/10.1111/j.1469-8137.2011.03714.x>
5. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633-2635. <https://doi.org/10.1093/bioinformatics/btm308>
6. Capo-chichi LJA, Morton CM, Weaver DB (2004) An intraspecific genetic map of velvet bean (*Mucuna* sp.) based on AFLP markers. *Theor Appl Genet* 108:814-821. <https://doi.org/10.1007/s00122-003-1493-8>
7. Carsky RJ, Ndikawa R (1998) Identification of cover crops for the semi-arid savannah zone of West Africa. In: Buckles D, Eteka A, Osiname M, Galiba M, Galiano G(ed) *Cover crops in West Africa - contributing to sustainable agriculture*. International Development Research Centre (IDRC), Ottawa, pp. 179-187.
8. Chinapolaiah A, Bindu HH, Khadke GN, Manjesh GN, RaoNH, Kumar SS, et al. (2018) Genetic diversity analysis in underutilized medicinal climber *Mucuna pruriens* (L.) DC. germplasm revealed by inter simple sequence repeats markers. *Legume Res* 43:32-37. <https://doi.org/10.18805/LR-3954>
9. Chinapolaiah A, Bindu KH, Manjesh GN, Thondaiman V, Shivakumara KT (2019) Genetic variability, correlation and path analysis for yield and biochemical traits in velvet bean [*Mucuna pruriens* (L.)]. **J Pharmacogn Phytochem** 8:2698-2704.
10. Daxenbichler ME, Van Etten CH, Earle FR, Tallent WH(1972) L-DOPA recovery from *Mucuna* seed. *J Agric Food Chem* 20:1046-1048.
11. Douglas SJ, Li B, Kliebenstein DJ, Nambara E, Riggs, C.D.(2017) A novel filamentous flower mutant suppresses brevipedicellus developmental defects and modulates glucosinolate and auxin levels. *PLoS One* 12:e0177045. <https://doi.org/10.1371/journal.pone.0177045>
12. Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13-15.
13. Duke (1981) *Handbook of legumes of world economic importance*. Plenum Press, New York, pp. 170-184.
14. Earl DA (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conser Genet Resour* 4:359-361. <https://doi.org/10.1007/s12686-011-9548-7>
15. Eilitta M, Bressani R, Carew LB, Carsky RJ, Flores M, Gilbert R, Huyck L, St. Laurent L, Szabo NJ (2002) *Mucuna* as a food and feed crop: An overview. In: Flores BM, Eilitta M, Myhrman R, Carew LB, Carsky RJ (ed) *Food and feed from Mucuna: Current Uses and the Way Forward*, Workshop. CIDICCO, CIEPCA and World Hunger Research Center, Tegucigalpa, pp. 18-47.
16. Feng J, Wang L, Wu Y, Luo Q, Zhang Y, Qiu D, et al. (2018) *TaSnRK2.9*, a sucrose non-fermenting 1-related protein kinase gene, positively regulates plant response to drought and salt stress in transgenic tobacco. *Front Plant Sci* 9:2003. <https://doi.org/10.3389/fpls.2018.02003>
17. Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, et al. (2005) Maize association population: A high- resolution platform for quantitative trait locus dissection. *Plant J* 44:1054-1064.

18. Fujii Y, Shibuya T, Usami Y (1991) Allelopathic effect of *M. pruriens* on the appearance of weeds. J Res Weed Sci 36: 43-49.
19. Fung SY, Tan NH, Sim SM, Marinello E, Guerranti Aguiyi JC (2011) *M. pruriens* Linn. seed extract pretreatment protects against cardiorespiratory and neuromuscular depressant effects of *Naja sputatrix* (Javan spitting cobra) venom in rats. Indian J Exp Biol 49:254-259.
20. Ganem B (1978) From glucose to aromatics: Recent developments in natural products of the shikimic acid pathway. Tetrahedron 34:3353-3383.
21. Grotz N, Fox T, Connolly E, Park W, Guerinot ML, Eide D (1998) Identification of a family of zinc transporter genes from *Arabidopsis* that respond to zinc deficiency. Proc Natl Acad Sci 95:7220-7224. <https://doi.org/10.1073/pnas.95.12.7220>
22. Gurumoorthi P, Kumar SS, Vadivel V, Janardhanan K (2003) Studies on agrobotanical characters of different accessions of velvet bean collected from Western Ghats, South India. Trop Subtrop Agroecosyst 2:105-115.
23. Hadapad B, Ravi CS, Shivaprasad M, BinduH, Nadukeri S, Devaraju (2018) Genetic variability and correlation studies for quantitative and qualitative traits in velvet bean (*Mucuna pruriens* L.) genotypes in rubber plantation under hill zone of Karnataka. **J Pharmacogn Phytochem SP3:83-90.**
24. Hamberger B, Hahlbrock K (2004) The 4-Coumarate: CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proc Natl Acad Sci 101:2209-2214. <https://doi.org/10.1073/pnas.0307307101>
25. Hedge JE, Hofreiter BT (1962) Determination of reducing sugars and carbohydrates. In: Whistler RL, BeMillerJN (ed) Methods in carbohydrate chemistry. Academic Press, New York, pp 380-394.
26. Hira CK, Chopra N (1995) Effects of roasting on protein quality of chickpea (*Cicer arietinum*) and peanut (*Arachis hypogaea*). J Food Sci Technol 32:501-503.
27. Hu Z, Zhang D, Zhang G, Kan G, Hong D, Yu D (2014) Association mapping of yield-traits and SSR markers in wild soybean (*Glycine soja* Sieb. and Zucc.). Breed Sci 63:441-449. <https://doi.org/10.1270/jsbbs.63.441>
28. Jia C, Wu X, Chen M, Wang Y, Liu X, Gong P, et al. (2017) Identification of genetic loci associated with crude protein and mineral concentrations in alfalfa (*Medicago sativa*) using association mapping. BMC Plant Biol 17:97. <https://doi.org/10.1186/s12870-017-1047-x>
29. Jorge M, Eilitta M, Proud, F (2007) *Mucuna* species: Recent advances in application of biotechnology. Fruit Veg Cereal Sci Biotech 1:80-94.
30. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica. <https://doi.org/10.1007/s10681-010-0286-9>
31. Kalidass C, Mohan VR (2011) Nutritional and anti-nutritional composition of itching bean [*Mucuna pruriens* (L.) DC. var. *pruriens*]: An underutilized tribal pulse in Western Ghats, Tamil Nadu. Trop Subtrop Agroecosyst 14:279-293.

32. Kalinowski ST (2005) HP-Rare 1.0: A Computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* 5:187-189. <https://doi.org/10.1111/j.1471-8286.2004.00845.x>
33. Kumar P, Saha S (2013) An updated review on taxonomy, phytochemistry, pharmacology and toxicology of *Macuna pruriens*. *J Pharmacogn Phytochem* 2: 306-314.
34. Kumar S, Singh GK, Kumar R, Bhatia NK, Awasthi CP (1991). Variation in quality traits of pigeon pea (*Cajanus cajan* L. Millsp.) varieties. *J Food Sci Technol* 28: 173-174.
35. Kumar PR, Sundeep S, Sathyanarayana N (2019) Microsatellite analysis reveals low inter-population differentiation in velvet bean (*Mucuna pruriens* var. *utilis*) of India. *Nucleus* 63:35-45. <https://doi.org/10.1007/s13237-019-00276-1>
36. Kumwenda JDT, Gilbert RA (1998) Biomass production by legume green manures on exhausted soils in Malawi: A soil fertility network trial. In: Waddington SR, Murwira HK, Kumwenda JDT, Hikwa D and Tagwira F (ed) Soil fertility research for maize-based farming systems in Malawi and Zimbabwe. Proceedings of the soil fertility network results and planning workshop, Mutare, pp. 85-86.
37. Leelambika M, Mahesh S, Jaheer M, Sathyanarayana N (2010) Comparative evaluation of genetic diversity among Indian *Mucuna* species using morphometric, biochemical and molecular approaches. *World J Agric Sci* 6:568-578.
38. Leelambika M, Sathyanarayana N (2011) Genetic characterization of Indian *Mucuna* (Leguminosae) species using morphometric and random amplification of polymorphic DNA (RAPD) approaches. *Plant Biosyst* 145:786-797. <https://doi.org/10.1080/11263504.2011.610924>
39. Lehnert H, Serfling A, Friedt W, Ordon F (2018) Genome-wide association studies reveal genomic region associated with response of wheat (wheat (*Triticum aestivum* L.) to mycorrhizae under drought stress conditions. *Front Plant Sci* 9:1728. <https://doi.org/10.3389/fpls.2018.01728>
40. Lepcha P, Kumar PR, Sathyanarayana N (2019) Exploring genomics research in the context of some underutilized legumes-a review. In: Banerjee R, Kumar GV, Kumar SPJ (ed) OMICS-based approaches in plant biotechnology. Scrivener Publishing, Beverly, pp. 1-8. <https://doi.org/10.1002/9781119509967.ch1>
41. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, et al. (2012) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet* 45:43-50. <https://doi.org/10.1038/ng.2484>
42. Liu K, Muse SV (2005) PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128-2129. <https://doi.org/10.1093/bioinformatics/bti282>
43. Liu R, Fang L, Yang T, Zhang X, Hu J, Zhang H, et al. (2017) Marker-trait association analysis of frost tolerance of 672 worldwide pea (*Pisum sativum* L.) collections. *Sci Rep* 7:5919. <https://doi.org/10.1038/s41598-017-06222-y>
44. Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* 1:1420-1425. <https://doi.org/10.2307/2445869>
45. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with folin phenol reagent. *J Biol Chem* 193:265-275.

46. Ma L, Liu M, Yan Y, Qing C, Zhang X, Zhang Y, et al. (2018) Genetic dissection of maize embryonic callus regenerative capacity using multi-locus genome-wide association studies. *Front Plant Sci* 9:561. <https://doi.org/10.3389/fpls.2018.00561>
47. Mahajan R, Zargar SM, Salgotra RK, Singh R, Wani AA, Nazir M, et al. (2017) Linkage disequilibrium based association mapping of micronutrients in common bean (*Phaseolus vulgaris* L.): A collection of Jammu & Kashmir, India. *3 Biotech* 7:295. <https://doi.org/10.1007/s13205-017-0928-x>
48. Mahesh S, Leelambika M, Jaheer M, Anithakumari AM, Sathyanarayana N (2016) Genetic mapping and QTL analysis of agronomic traits in Indian *Mucuna pruriens* using an intraspecific F2 population. *J Genet* 95:35-44. <https://doi.org/10.1007/s12041-015-0602-1>
49. Mahesh S, Sathyanarayana N (2011) The genotype X environment interaction and stability analysis for L-DOPA trait in *M. pruriens* seeds. *Indian J Genet PI Br* 71:279-282.
50. Mahesh S, Sathyanarayana N (2015) Intra-Specific variability for salinity tolerance in Indian *Mucuna pruriens* (L.) DC. germplasm. *J Crop Sci Biotechnol* 18:181-194. <https://doi.org/10.1007/s12892-015-0019-7>
51. Mang YD, Njintang YN, Abdou BA, Scher J, Bernard C, Mbofung MC (2016) Dehulling reduces toxicity and improves in vivo biological value of proteins in vegetal milk derived from two mucuna (*Mucuna pruriens* L.) seeds varieties. *J Food Sci Technol* 53: 2548-2557.
52. Mathew IE, Das S, Mahto A, Agarwal P (2016) Three rice NAC transcription factors heteromerize and are associated with seed size. *Front Plant Sci* 7:1638. <https://doi.org/10.3389/fpls.2016.01638>
53. Miller PA, Rawlings JO (1967) Breakup of initial linkage blocks through intermating in a cotton breeding population. *Crop Sci* 7:199-204.
54. Milner MJ, Seamon J, Craft E, Kochian LVJ (2013) Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. *Exp Bot* 64:369-381. <https://doi.org/10.1093/jxb/ers315>
55. Negron L, Patchett ML, Parker EJ (2011) Expression, purification, and characterization of dehydroquinatesynthase from *Pyrococcus furiosus*. *Enzyme Res* 2011:134893. <https://doi.org/10.4061/2011/134893>
56. Oraguzie NC, Wilcox PL (2007) An overview of association mapping. In: Oraguzie NC, Rikkerink EHA, Gardiner SE (ed) *Association Mapping in Plants*. Springer, New York, pp 1–9. https://doi.org/10.1007/978-0-387-36011-9_1
57. Padmesh P, Reji JV, Dhar JM, Seeni D (2006) Estimation of genetic diversity in varieties of *Mucuna pruriens* using RAPD. *Biol Plant* 50:367-372. <https://doi.org/10.1007/s10535-006-0051-z>
58. Patil RR, Pawar KD, Rane MR, Yadav SR, Bapat VA, Jadhav JP (2016) Assessment of genetic diversity in *Mucuna* species of India using randomly amplified polymorphic DNA and inter simple sequence repeat markers. *Physiol Mol Biol Plants* 22:207-217. <https://doi.org/10.1007/s12298-016-0361-3>
59. Peakall R, Smouse PE (2012) GenAEx 6.5: Genetic analysis in excel. Population genetic software for teaching and research - an update. *Bioinformatics* 28:2537-2539.

<https://doi.org/10.1093/bioinformatics/bts460>

60. Perrier X, Jacquemoud-collet JP (2006) DARwin software, version 5.0.158. Montpellier: Department Systems Biologiques (BIOS), CIRAD.
61. Pugalenth M, Vadivel V, Siddhuraju P (2005) Alternative food/feed perspectives of and underutilized legume *M. pruriens* var. *utilis* - a review. Plant Foods Hum Nutr 60:201-218.
<https://doi.org/10.1007/s11130-005-8620-4>
62. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959. <https://doi.org/10.1093/genetics/155.2.945>
63. R Core Team (2014) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. <http://www.R-project.org/>
64. Rakshit A, Rakshit S, Singh J, Chopra SK, Balyan HS, Gupta PK, et al. (2010) Association of AFLP and SSR markers with agronomic and fiber quality traits in *Gossypium hirsutum* L. J Genet 89:155-162.
<https://doi.org/10.1007/s12041-010-0055-5>
65. Rohilla M, Singh N, Mazumder A, Sen P, Roy P, Chowdhury D, et al. (2018) Genome-wide association studies using 50 K rice genic SNP chip unveil genetic architecture for anaerobic germination of deep-water rice population of Assam, India. Mol Genet Genomics 295:1211-1226.
<https://doi.org/10.1007/s00438-020-01690-w>
66. Rosmaina, Syafrudin, Hasrol, Yanti, Juliyanti and Zulfahmi (2016) Estimation of variability, heritability and genetic advance among local chilli pepper genotypes cultivated in peat lands. Bulg J Agric Sci 22:431-436.
67. Sasidharan N (2004). Biodiversity documentation for Kerala Part 6: Flowering plants. Kerala Forest Research Institute (KFRI), Peechi, India.
68. Sastrapradja S, Sastrapradja D, Aminah SH, Lubis I and Idris S (1974) Morphological and cytological investigation on some species of *Mucuna* (Papilionaceae). Ann Bogor 5: 173-178.
69. Sathyanarayana N, Leelambika M, Mahesh S, Jaheer M (2011) AFLP assessment of genetic diversity among Indian *Mucuna* accessions. Physiol Mol Biol Plants 17:171-180.
<https://doi.org/10.1007/s12298-011-0058-6>
70. Sathyanarayana N, Mahesh S, Jaheer M, Leelambika M (2012) Genetic diversity of wild and cultivated *Mucuna pruriens* (L.) DC. accessions analyzed using thirty morpho-agronomical characters. Trop Subtrop Agroecosyst 15:249-259.
71. Sathyanarayana N, Mahesh S, Leelambika M, Jaheer M, Chopra R, Rashmi KV (2016) Role of genetic resources and molecular markers in *M. pruriens* (L.) DC. improvement. Plant Genet Resour 14:270-282. <https://doi.org/10.1017/S1479262116000071>
72. Sathyanarayana N, Pittala RK, Tripathi PK, Chopra R, Singh HR, Belamkar V, et al. (2017) Transcriptomic resources for the medicinal legume *Mucuna pruriens*: *de novo* transcriptome assembly, annotation, identification and validation of EST-SSR markers. BMC Genom 18:409.
<https://doi.org/10.1186/s12864-017-3780-9>

73. Shetty P, Sharma S, Sathyanarayana N (2015) Exploiting legume EST data bases for the development of gene-derived SSR-markers in medicinal legume *Mucuna pruriens* L. (DC.). Electron J Plant Breed 6:1041-1051.
74. Shi C, Ren Y, Liu L, Wang F, Zhang H, Tian P, et al. (2019) Ubiquitin specific protease 15 has an important role in regulating grain width and size in rice. Plant Physiol 180:381-391. <https://doi.org/10.1104/pp.19.00065>
75. Siddhuraju P, Becker K, Makkar HP (2000) Studies on the nutritional composition and antinutritional factors of three different germplasm seed materials of an under-utilized tropical legume, *Mucuna pruriens* var. *utilis*. J Agri Food Chem 48: 6048-6060.
76. Singh A, Sharma V, Dikshit K, Aski M, Kumar H, Thirunavukkarasu N et al. (2017) Association mapping unveils favorable alleles for grain iron and zinc concentrations in lentil (*Lensculinaris* subsp. *culinaris*). PLoS One12:e0188296. <https://doi.org/10.1371/journal.pone.0188296>
77. Soumy PR, Burrige AJ, Singh N, Batra R, Pandey R, Kalia S, et al. (2021) Population structure and genome-wide association studies in bread wheat for phosphorus efficiency traits using 35 K wheat breeder's affymetrix array. Sci Rep 11:7601. <https://doi.org/10.1038/s41598-021-87182-2>
78. Syukur M, Sujiprihati S, Yuniarti R (2012) Teknik Pemuliaan Tanaman. PenebarSwadaya, Jakarta.
79. Tiwari DK, Pandey P, Tripathi S, Giri SP, Dwivedi JL (2011) Studies on genetic variability for yield components in rice (*Oryza sativa* L.). AAB Bioflux 3:76-81.
80. Tripathi PK, Jena SN, Rana TS, Sathyanarayana N (2018) High levels of gene flow constraints population structure in *Mucuna pruriens* L. (DC.) of Northeast India. Plant Gene 15:6-14. <https://doi.org/10.1016/j.plgene.2018.05.005>
81. Vijayanthi PV, Ramesh S, Gowda MB, Rao AM, Keerthi CM (2018) Genome-wide marker-trait association analysis in a core set of dolichos bean germplasm. Plant Genet Resour 3:1-11. <https://doi.org/10.1017/S1479262118000163>
82. Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621-630. <https://doi.org/10.1016/j.tplants.2005.10.004>
83. Wang Q, Tian F, Pan Y, Buckler ES, Zhang ZA (2014) A SUPER powerful method for genome wide association study. PLoS One 9:e107684. <https://doi.org/10.1371/journal.pone.0107684>
84. Wilmot-Dear CM (1987) A revision of *Mucuna* (Leguminosae Phaseoleae) in the Indian sub-continent and Burma. Kew Bulletin 42: 23-46.
85. Xu Y, Yang T, Zhou Y, Yin S, Li P, Liu J, et al. (2018) Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. Front Plant Sci 9:1311. <https://doi.org/10.3389/fpls.2018.01311>
86. Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, et al. (2014) Genome-wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. PLoS Genetics 10:e1004573. <https://doi.org/10.1371/journal.pgen.1004573>
87. Yeh FC, Yang RC, Boyle T, Ye ZH, Mao JX (1999) POPGENE, version 1.32: the user-friendly software for population genetic analysis. Edmonton, AB: Molecular Biology and Biotechnology Centre,

University of Alberta.

88. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, et al. (2021) rMVP: A memory-efficient, visualization enhanced, and parallel-accelerated tool for genome-wide association study. *Genom Proteom Bioinf*. <https://doi.org/10.1016/j.gpb.2020.10.007>
89. Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17:155-160. <https://doi.org/10.1016/j.copbio.2006.02.003>
90. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203-208. <https://doi.org/10.1038/ng1702>
91. Zhao J, Huang L, Ren X, Pandey MK, Wu B, Chen Y, et al. (2017). Genetic variation and association mapping of seed-related traits in cultivated peanut (*Arachis hypogaea* L.) using single-locus simple sequence repeat markers. *Front Plant Sci* 8:2105. <https://doi.org/10.3389/fpls.2017.02105>

Figures

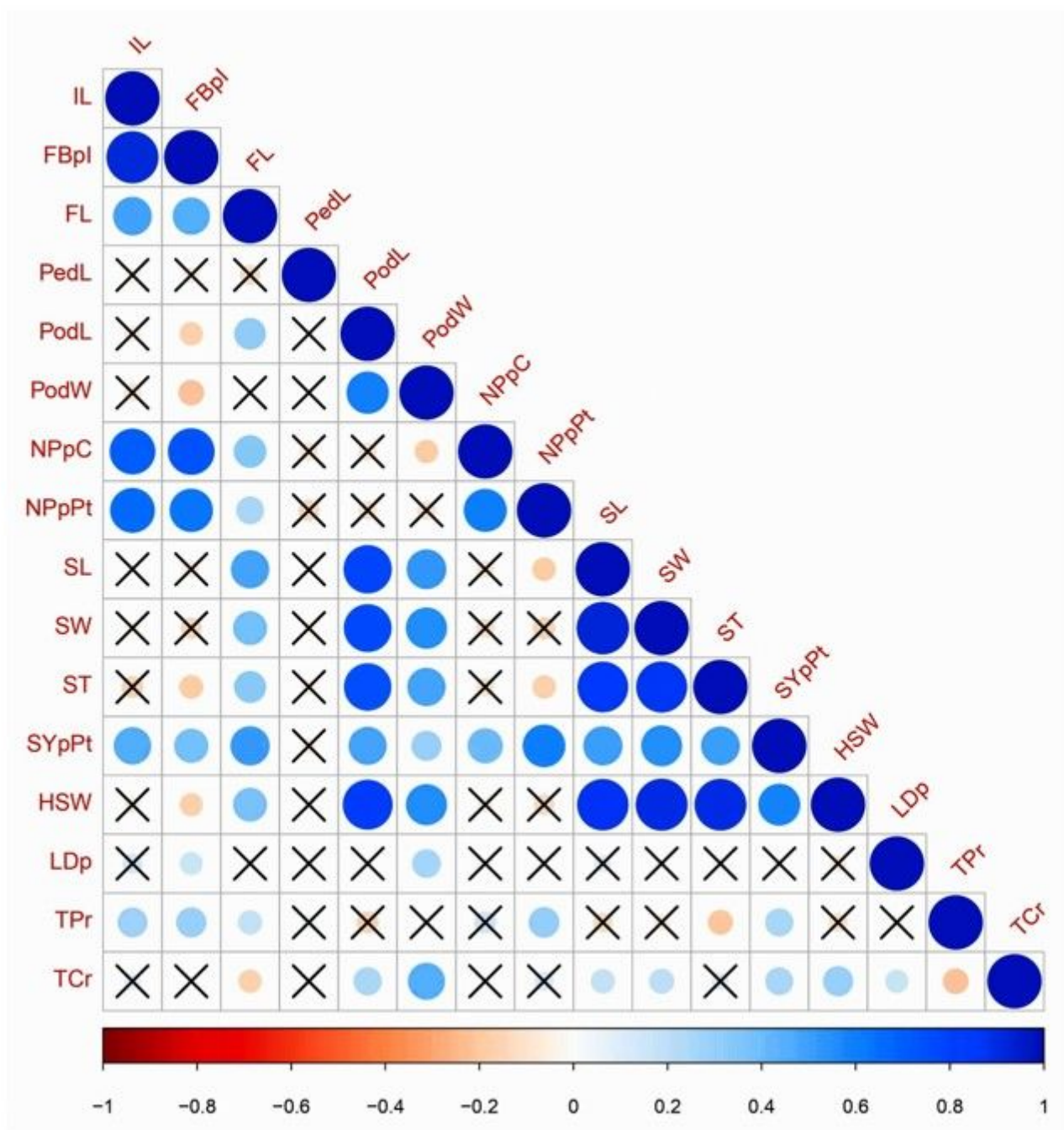


Figure 1

Correlation matrix of sixteen phenotypic traits in *M. pruriens* association mapping panel. *Note: The intensity of color ranges from blue (positive) to brown (negative) and the size of the circles show strength of significant correlation ($P < 0.05$)*

IL: Inflorescence length; FBpl: Flower buds per inflorescence; FL: Flower length; PedL: Pedicel length; PodL: Pod length; PodW: Pod width; NPpC: Number of pods per cluster; NPpPt: Number of pods per plant; SL: Seed length; SW: Seed width; ST: Seed thickness; SYpPt: Seed yield per plant; HSW: Hundred seed

weight; LDp: Total L-DOPA; TPr: Total protein; TCr: Total carbohydrate; ✕: Statistically non-significant correlation values

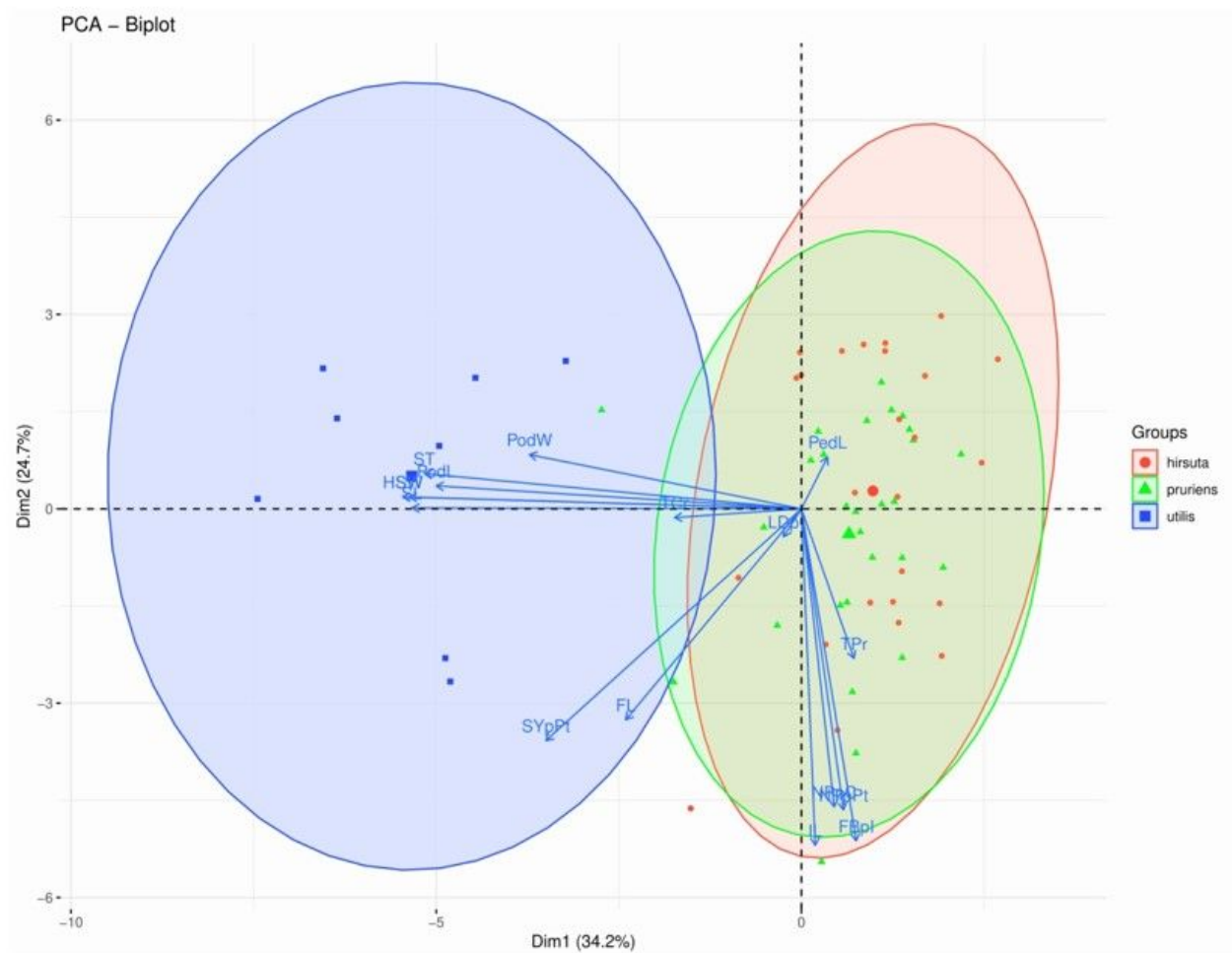


Figure 2

Scatter plot of *M. pruriens* accessions based on sixteen phenotypic traits from the principal component analysis (PCA). Note: Red circles, green triangles and blue squares represent *M. pruriens* var. *hirsuta*, var. *pruriens*, and var. *utilis*, respectively

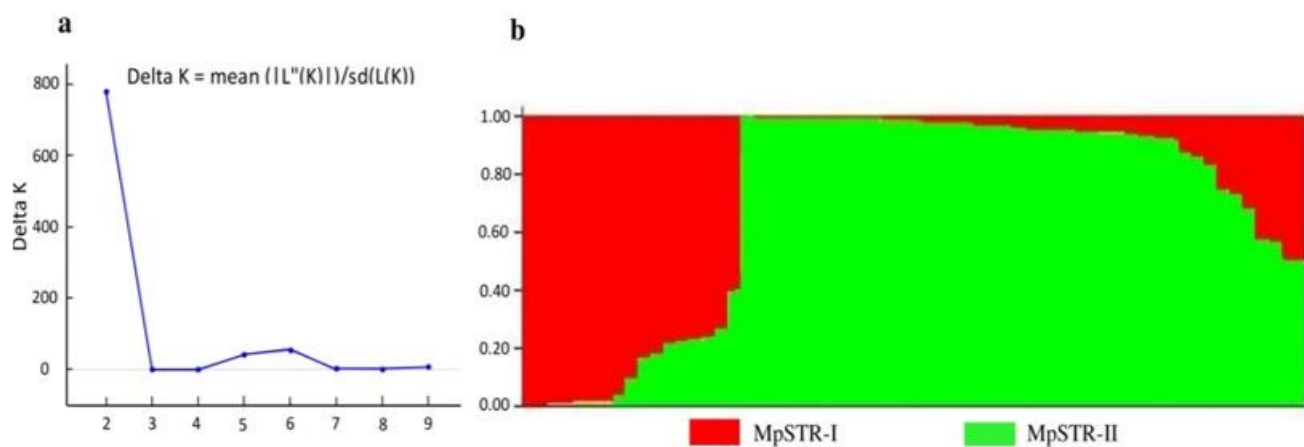


Figure 3

Population structure of *M. pruriens* association mapping panel inferred using STRUCTURE software **(a)** Hypothetical sub-population estimation using ΔK -values ($K = 2$) indicating two subpopulations **(b)** Population structure at $K = 2$ based on inferred ancestry (Q-matrix) in which two sub-populations are indicated as MpSTR-I (red color) and MpSTR-II (green color)

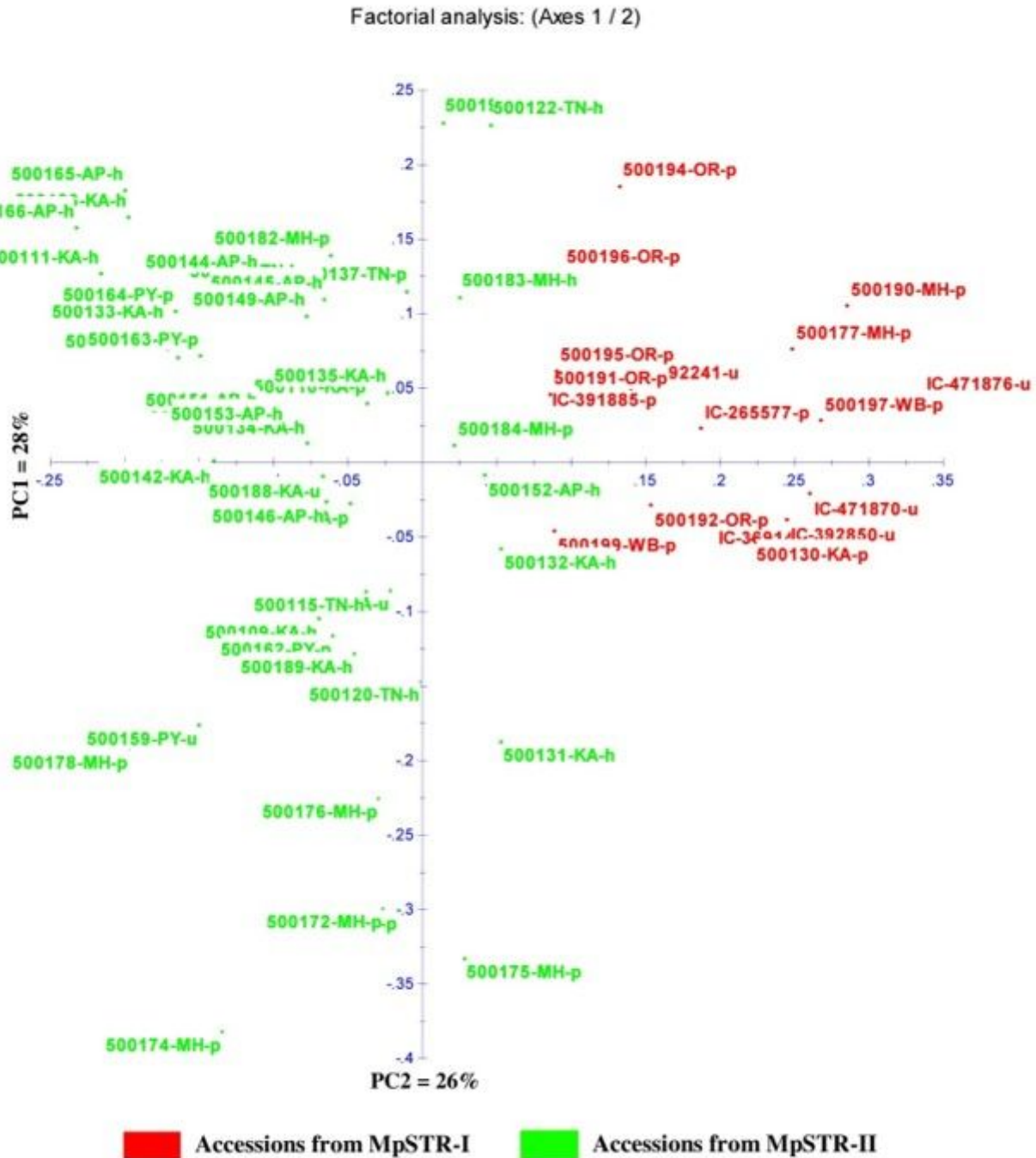


Figure 5

Scatter plot depicting dispersion of *M. pruriens* accessions based on principal coordinate analysis. *Note:* Red and green color indicates subgroup-1 (MpSTR-I) and subgroup-2 (MpSTR-II) from STRUCTURE analysis

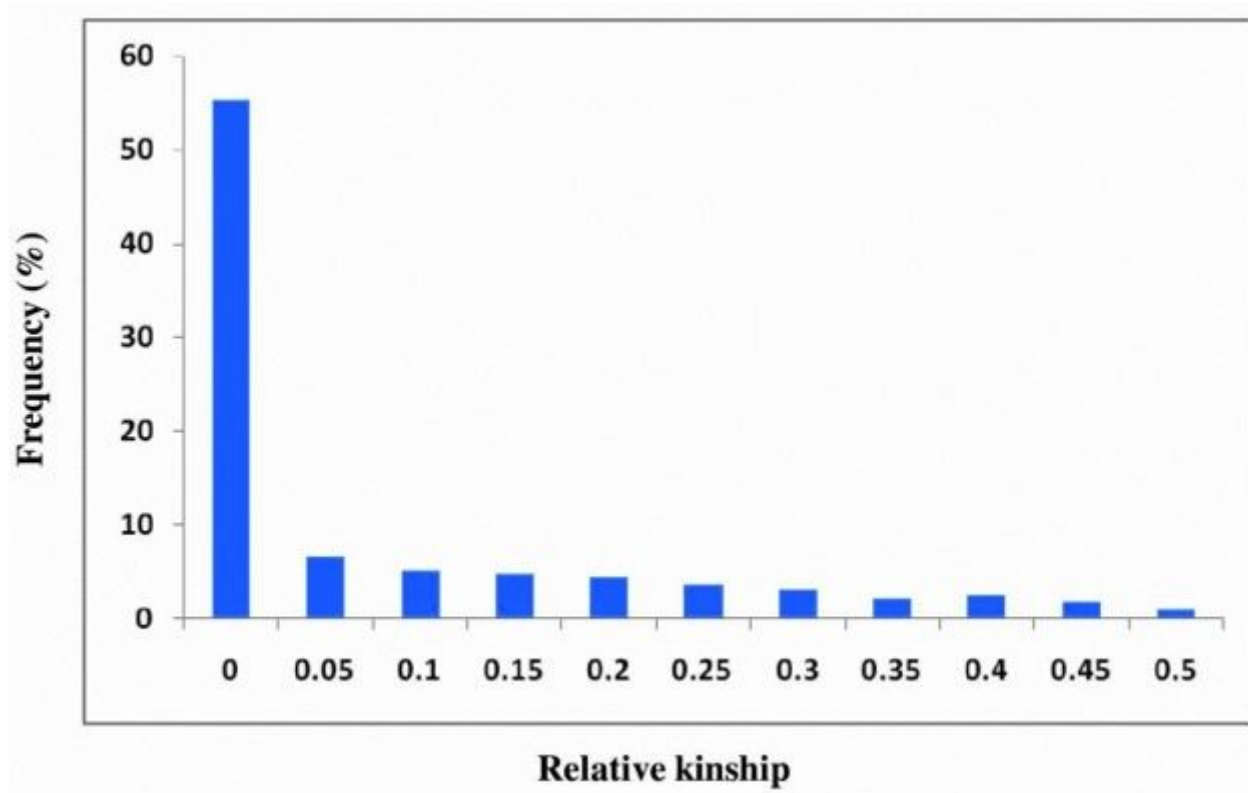


Figure 6

Distribution of global pairwise kinship coefficients (F_{ij}) of *M. pruriens* association panel

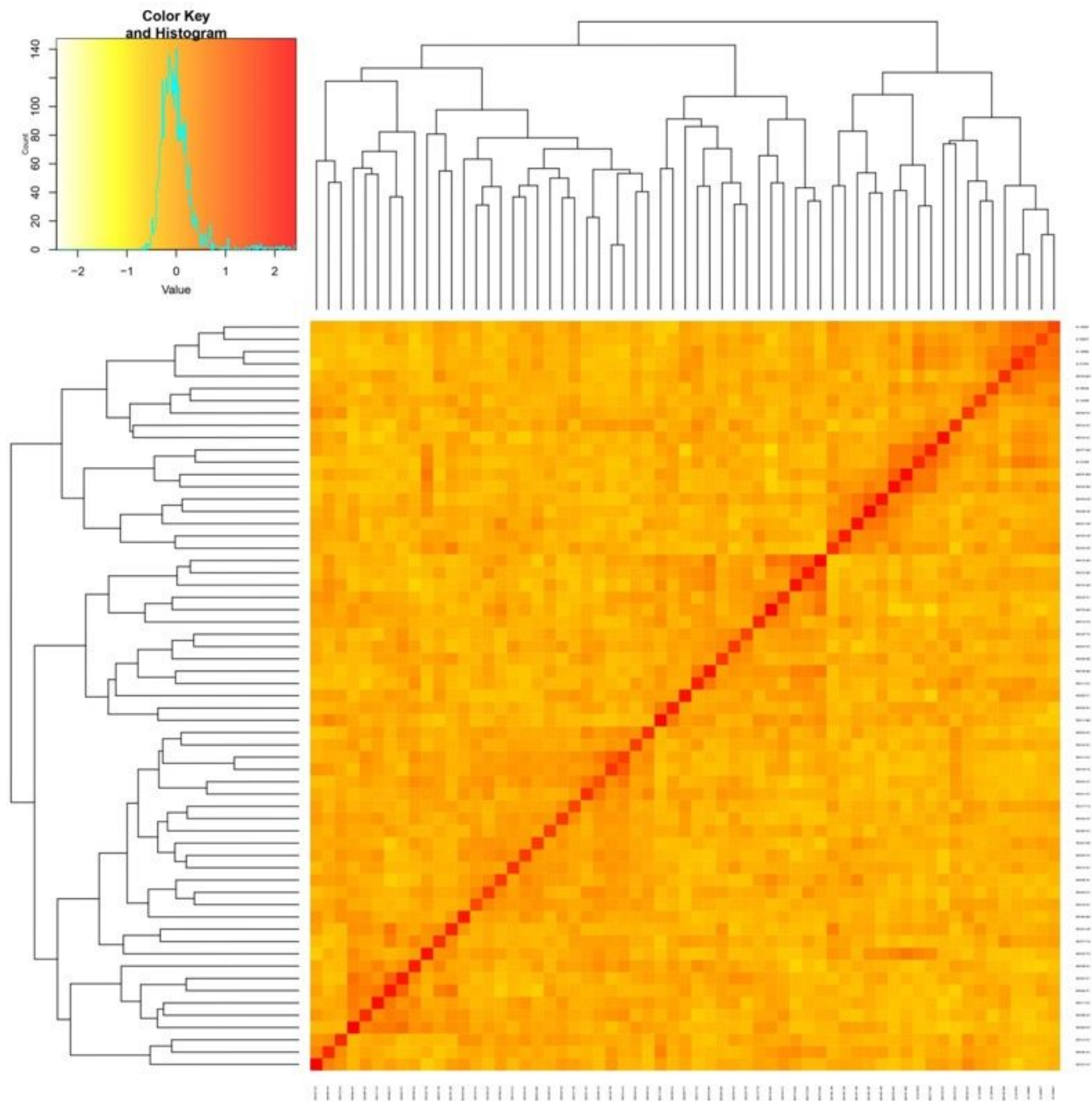


Figure 7

Heat map of kinship matrix generated for *M. pruriens* association mapping panel based on 180 filtered SSR markers data. Dendrogram are shown on the top and left side of the figure

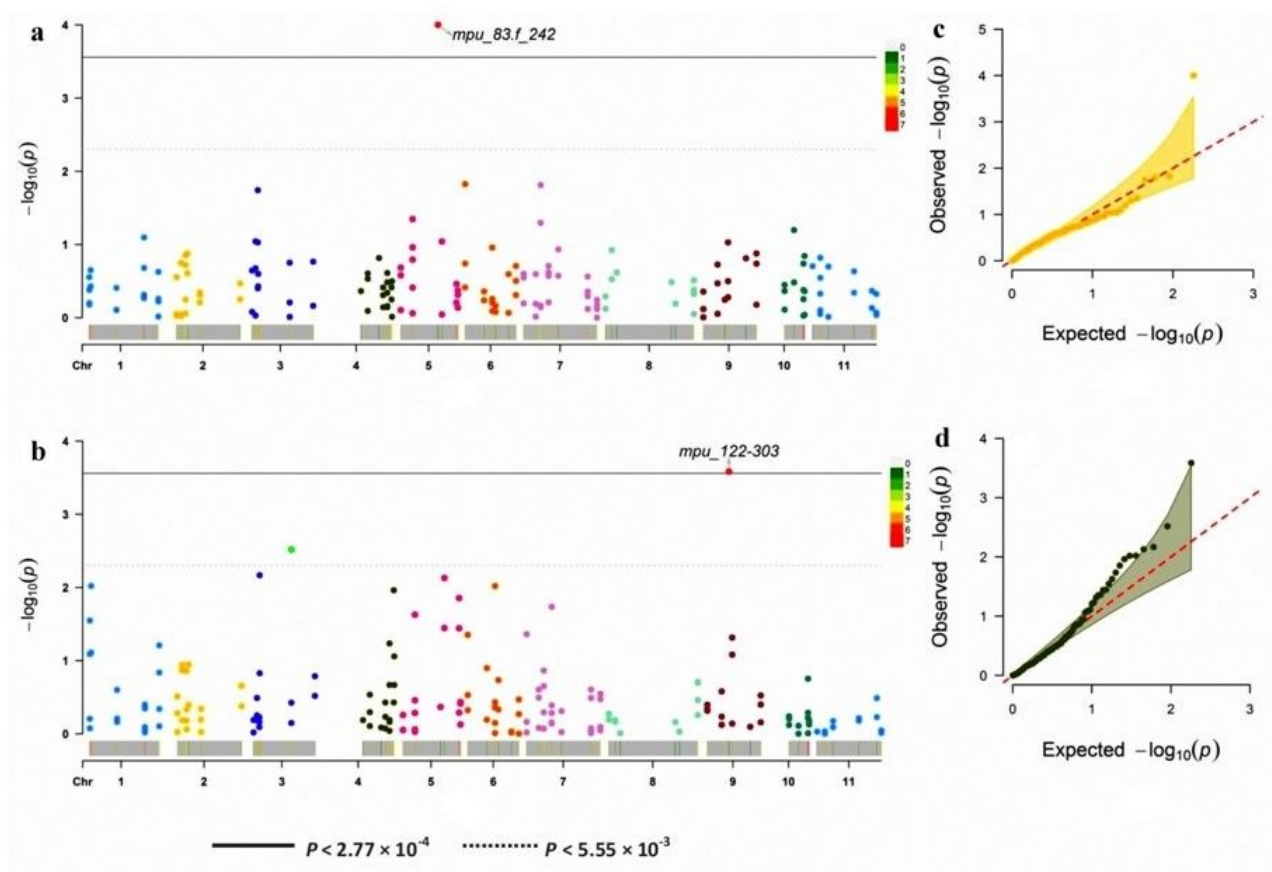


Figure 8

Manhattan plot (P -value) for MLM depicting significant marker-trait associations at adjusted Bonferroni threshold $P < 2.77 \times 10^{-4}$ (a) seed yield per plant (b) Pedicel length; Quantile-quantile (Q-Q) plot of MLM (c) Seed yield per plant (d) Pedicel length

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterials.docx](#)
- [SupplementaryTables.xlsx](#)