SUMMARY

The *Mucuna pruriens* can meet three burgeoning needs of tropical agriculture: low-cost protein source, high-value medicinal plant, and an excellent green manure cover crop (GMCC). But like other underutilized crops, it lacks modern breeding resources. Identifying marker-trait associations (MTAs) could prompt marker-assisted breeding paving a way for the early development of the improved varieties in this plant. Recent studies have demonstrated the feasibility of identifying MTAs even using a small number of accessions (<100) triggering the hope for resource-poor crops. Taking a cue from there, we characterized a panel of 61 *M. pruriens* accessions across two consecutive years and performed association mapping using 66 genic-microsatellite markers. In addition, a detailed analysis of genetic diversity and seed trait variability among the 60 *M. pruriens* germplasm accessions from Northeast India has been carried out to supplement the existing association mapping (AM) panel in future genome-wide association studies (GWAS).

1. Variability for seed traits in *M. pruriens* natural population of Northeast India

M. pruriens is grown by several indigenous communities of Northeast India for edible pods and seeds. But, the germplasm is poorly explored for its offerings. In this study, 60 accessions collected from different locations of Northeast India revealed a significant variability for the six seed-based traits ($P < 0.05$). The two varieties- var. *pruriens* and var. *utilis* showed a wide difference in their seed characters and formed distinct clusters in principal component analysis (PCA) scatter plot. Almost all the seed-based traits were positively correlated ($P < 0.05$) suggesting their utility in the breeding programs. The presence of good phenotypic and genetic diversity in the *M. pruriens* natural population of Northeast India signifies their utility as an important resource base for future GWAS in this plant.

2. Genetic diversity among the Northeast Indian accessions

We observed a moderate to high genetic diversity in *M. pruriens* population from Northeast India (Shannon's information index, I = 0.496). This provides credence to the hypothesis of its origin in Eastern India. A moderate polymorphism information content (PIC) value was recorded for most of the microsatellite markers ($0.25 < PIC < 0.50$) suggesting their efficiency for the genetic diversity analysis. The distance-based neighbor-joining (NJ) tree, principal coordinate analysis (PCoA) plot, and Bayesian-based STRUCTURE revealed the absence of a structured population. We also observed that high gene flow has triggered low genetic differentiation among the population group. The overall high genetic diversity among the *M. pruriens* population from the Northeast region of India signifies a possible storehouse of several novel genes. Besides, high genetic diversity and weak population structure observed imply their use in future GWAS or genomic selection programs.

3. Feasibility analysis of association mapping panel

For association analysis, we used 61 accessions of *M. pruriens* procured from Sir Mokshagundam Visvesvaraya Institute of Technology (MVIT), Bangalore as an association mapping panel. The distance-based NJ tree, PCoA plot, and Bayesian-based STRUCTURE analysis revealed an analogous trend suggesting the minimal influence of the geographical origin and/or the varietal affiliation on the grouping of the accessions in our AM panel. Low pairwise kinship (*Fij*) suggested weak relatedness between the accessions. A low level of linkage disequilibrium (average $R^2 = 0.032$) for all the possible combinations of microsatellite marker pairs and its sharp decay implied the need for more markers for identifying reliable MTAs. Based on the weak population structure and low relatedness between accessions, the panel of accessions was determined to be suitable for the association mapping.

4. Association analysis

A significant phenotypic variability and high broad-sense heritability were observed for many of the traits measured in AM panel accessions. Fifteen MTAs were identified in association analysis for agronomically important traits with phenotypic variance explained (PVE) >10% from mixed linear model (MLM). Their reliability tested through annotation against the *Arabidopsis* genome database lends further credence to this. This is the first report on association mapping study in *M. pruriens* and the results are expected to offer significant groundwork for future marker-assisted breeding and mining candidate genes for important agronomic traits in this promising underutilized legume for the tropics.