# NAM population – a novel genetic resource for soybean improvement: development and characterization for yield and attributing traits

Shivakumar Maranna[1]* ⓘ, Giriraj Kumawat[1], Vennampally Nataraj[1], C. Gireesh[2] ⓘ,
Sanjay Gupta[1], Gyanesh K. Satpute[1], Milind B. Ratnaparkhe[1] and Devendra P. Yadav[1]

[1]*ICAR-Indian Institute of Soybean Research, Indore, Madhya Pradesh, India* and [2]*ICAR-Indian Institute of Rice Research, Hyderabad, India*

## Abstract

Nested association mapping (NAM) captures the best features of both linkage and association mapping and enables the high power and high resolution of quantitative trait locus mapping through joint linkage-association analysis. In the current study, NAM population was developed by hybridizing JS 335, a popular variety of central India with 20 diverse soybean genotypes. The parents used in the study have various traits of economic importance such as drought and water-logging tolerance, bacterial pustule and yellow mosaic virus resistance, wider adaptability, resistance to mechanical damage and higher yield potential. High variability in the $F_2$ populations of 20 crosses for grain yield and days to maturity indicated scope for development of high-yielding varieties. Genetic variability studies, correlation, regression, principal component analysis (PCA) and genetic diversity analyses were carried out in 900 NAM-recombinant inbred lines (RILs) derived from 11 crosses. Correlation and regression analysis indicated a significant positive effect of biomass, pods/plant, harvest index, branches/plant, nodes/plant and plant height on grain yield. Genetic diversity analysis grouped 900 NAM-RILs into 10 clusters. PCA revealed first two principal components to explain 63.78% of total variation mostly contributed by grain yield, biomass and number of pods. The inbred lines developed in this study will serve as an elite soybean genetic resource in understanding the genetic architecture underlying different traits of economic significance.

## Introduction

Soybean [*Glycine max* (L.) Merr.] is a leading oilseed crop accounting for 358.77 million metric tonnes in 2018–19 (USDA, 2019). Soybean in India, grown annually as rain fed crop with a production of 11.50 million metric tonnes in 2018–19 (USDA, 2019). Suitable mapping population is one of the major prerequisite for identification of genomic locations associated with quantitative traits. Biparental mapping populations have been extensively utilized to understand the genetic architecture of complex traits such as yield (Li *et al*., 2007), early maturity (Kong *et al*., 2014), seed weight (Karikari *et al*., 2019), etc. However, biparental mapping populations such as $F_{2:3}$ $BC_1$, $BC_2$ not suitable for replicated/multi-location evaluations which undermines precise phenotyping. Further, mapping resolution of biparental approach is poor due to limited recombinations during development of population and lower population size (usually <400). To overcome limitations of biparental mapping populations, association mapping (AM) is utilized in soybean and other crops for mapping of complex traits with higher resolution. AM depends on linkage disequilibrium in natural populations and historical recombinant

---

*Corresponding author. E-mail: shivaiari9683@gmail.com

events occurred through the evolutionary process. AM has been employed in soybean to genetically dissect quantitative traits such as seed weight (Wang *et al.*, 2016), seed protein, oil contents (Li *et al.*, 2019), etc. However, resolution of quantitative trait locus (QTL) mapping in AM is influenced by selection, migration, genetic drift and linkage disequilibrium (LD) decay. Therefore, to overcome the shortfalls of biparental mapping and AM, multiparent mapping population such as multi-parent advanced generation intercross (MAGIC) and nested association mapping (NAM) have be conceptualized in crop breeding. MAGIC population is developed by three levels (two way, four way and eight way) of inter-crosses among the founder parents and further intermating of siblings (Shivakumar *et al.*, 2018). NAM captures the best features of both linkage and AM. It enables the high power and high resolution of QTL mapping through joint linkage-association analysis (McMullen *et al.*, 2009). NAM population was conceptualized and developed by Buckler *et al.* (2009) in maize wherein 25 NAM families were developed crossing 25 diverse maize genotypes with a common reference genotype, B75. Since then, many workers have employed it in a wide range of crops including sorghum (Jordan *et al.*, 2011), wheat (Bajgain *et al.*, 2016) and rice (Fragoso *et al.*, 2017).

Soybean NAM population comprises 5600 recombinant inbred lines (RIL) developed by crossing 40 diverse soybean lines with common parent (IA3023, a high-yielding MG III cultivar) through the single seed descent method (Song *et al.*, 2017) and utilized to study the flowering date in soybean (Li *et al.*, 2017), yield and agronomic traits (Diers *et al.*, 2018) and drought tolerance (Buezoa *et al.*, 2019). Considering the significance of NAM approach in soybean improvement, a soybean NAM population was developed using JS 335 (a widely adapted cultivar in India) as common founder parent and 20 diverse founder parents. The common founder parent JS 335 was crossed with 20 diverse founder parents and developed >2000 ($F_{2.5}$ generation) RIL population through the single seed descent method. For genetic analysis of yield and attributing characters 900 RILs derived from eleven crosses were used and analysed for correlation, regression, principal component analysis (PCA) and cluster analysis.

## Materials and method

### Plant material

A NAM population was developed by hybridizing JS 335, a popular variety of central India with 20 diverse soybean genotypes for mapping and improving towards drought and water-logging tolerance, yellow mosaic virus (YMV) and rust resistance, early maturity and higher yield (Table 1). The hybridization was carried out pairwise between JS 335 and 20 parents by pollination without emasculation procedure (Talukdar and Shivakumar, 2012; Shivakumar *et al.*, 2016) to produce 20 different $F_1$s. Due to small quantity of $F_1$ seed produced for two combinations (NRC 37 × JS 335 and Valder × JS 335) only 18 $F_1$ combinations along with 21 parents were grown in completely randomized block design with three replications in research field at ICAR-Indian Institute of Soybean Research (ICAR-IISR), Indore, India. The hybridity of the $F_1$ plants was confirmed based on morphological markers such as stem pigmentation, flower colour, pubescence on stem and pod. The $F_2$ seeds from each of true $F_1$ plants of respective crosses were grown under field conditions. The seeds from each of $F_2$ plants were harvested separately for all cross combination and $F_{2-3}$ families were again planted and then single plant was randomly selected from each family for generation advancement. The same procedure was followed up to $F_5$ generation. The details of 11 crosses used for the genetic variability and multivariate analysis are presented in online Supplementary Table S3. Standard cultivation practices recommended for soybean crop has been followed to raise good crop (ICAR, 2009).

## Plant characters and data observations

To estimate heterosis, we recorded days to maturity and grain yield traits on five randomly selected plants of 18 different hybrids and respective parents of each replication. Similarly, $F_2$ population was grown under field conditions and observations on maturity and grain yield per plant was recorded on 20 $F_2$ populations. A total of 900 NAM-RILs derived from 11 crosses were sown in augmented design in 3 m row length along with six checks JS 335, JS 20-29, JS 20-34, NRC 86, JS 97-52 and JS 71-05 replicated in 12 blocks. The data on seven quantitative traits on these 900 NAM-RILs were phenotyped for plant height, number of branches, number of nodes, number of pods, biomass (g), grain yield (g) and harvest index. For recording of biomass trait, five randomly selected plants were harvested at physiological maturity and then oven dried plants' sample weights were recorded and averaged.

## Statistical analysis

Heterosis over standard check (standard heterosis) was calculated as per Hayes *et al.* (1955). Significance of heterosis estimates was tested by employing *t* test as per the formula given by Arunachalam (2017). Adjusted means for all the traits under considerations for all the 900 NAM-RILs were estimated by using the R package 'augmentedRCBD' (Aravind *et al.*, 2018). Genetic variability parameters, frequency distribution and descriptive statistics were also estimated by using the R package 'augmentedRCBD'.

**Table 1.** Details of NAM families used in the current study

| Crosses | No. of RILs | Generation | Targeted traits |
|---|---|---|---|
| DOKO × JS 335 | 75 | F5 | Resistance to mechanical seed damage |
| JS 335 × AGS 191 | 75 | F5 | Higher yield potential |
| JS 97-52 × JS 335 | 72 | F5 | YMV resistance and water-logging tolerance |
| Bragg × JS 335 | 78 | F5 | Resistance to bacterial pustule |
| JS 20-38 × JS 335 | 100 | F5 | Water-logging tolerance |
| PK 472 × JS 335 | 95 | F5 | YMV resistance |
| EC546882 × JS 335 | 105 | F5 | Higher yield potential |
| Jackson × JS 335 | 50 | F5 | Drought tolerance |
| Santamaria × JS 335 | 95 | F5 | Long juvenility |
| Kalitur × JS 335 | 55 | F5 | Anthracnose resistance |
| RVS 2007-6 × JS 335 | 100 | F5 | Higher yield potential |
| JS 335 × IC 501198 | 76 | F5 | Higher yield potential |
| EC 4811347 × JS 335 | 87 | F5 | Higher yield potential |
| JS 335 × Gaurav 2 | 80 | F5 | Multiple stress resistance |
| JS 335 × EC 656641 | 82 | F5 | Higher yield potential |
| JS 335 × IC 15759A | 85 | F5 | Higher yield potential |
| Hardee × JS 335 | 92 | F5 | Drought tolerance and bacterial pustule resistance |
| G-11 × JS 335 | 88 | F5 | Higher yield potential |
| NRC 37 × JS 335 | 90 | F5 | Bacterial pustule and bud blight resistance |
| Valder × JS 335 | 85 | F5 | Drought tolerance |
| Total | 1665 | – | – |

Phenotypic variance, genotypic variance, phenotypic coefficient of variation (PCV), genotypic coefficient of variation (GCV) (Aravind *et al*., 2018), heritability ($H^2$) (Lush, 1940) and genetic advance (GA) (Johnson *et al*., 1955) were calculated by using the following formulae:

Phenotypic Variance ($\sigma_p^2$)

$$= \text{Sum of squares of test treatments (genotypes)}$$

Error Variance $(\sigma_e^2)$ = Sum of squares of residuals (error)

$$\text{Genotypic Variance } (\sigma_g^2) = \sigma_p^2 - \sigma_e^2$$

Phenotypic Coefficient of Variation (PCV)

$$= \frac{\text{Phenotypic Variance}}{\sqrt{\text{mean}}} \times 100$$

Genotypic Coefficient of Variation (GCV)

$$= \frac{\text{Genotypic Variance}}{\sqrt{\text{mean}}} \times 100$$

$$\text{Broad} - \text{sense heritability } (H^2) = \sigma_g^2/\sigma_p^2$$

$$\text{Genetic Advance (GA)} = k \times \sigma_g \times H^2/100$$

where, $k$ is the selection intensity and $\sigma_g$ is the genotypic standard deviation.

Correlation analysis was performed by using the R function cor(). Correlation plots were obtained by using the R packages 'corrplot' (Wei and Simko, 2017.) and 'PerformanceAnalytics' (Peterson and Carl, 2018). Linear regression analysis was performed using the R function lm(), and PCA was performed by using the R packages 'corrplot' (Wei and Simko, 2017.), 'factoextra' (Kassambara and Mundt, 2017) and 'FactoMineR' (Le *et al*., 2008). The *K*-means cluster analysis was carried out using SYSTAT software version 13.2.

## Results

### *Estimation of F₁ heterosis*

Analysis of variance revealed significant differences ($P < 0.05$) for grain yield per plant (F ratio: 49.09***) and days to maturity (F ratio: 12.47***) across all the genotypes including parents and 18 crosses. For grain yield per plant, highest standard heterosis was recorded for the cross JS 97-52 × JS 335 (77.35%) followed by Bragg × JS 335 (52.04%), JS 20-38 × JS 335 (35.58%), RVS 2007-6 × JS 335 (33.66%), EC 546882 × JS 335 (26.25%), EC 656641 × JS 335 (13.7%) and IC 15759A × JS 335 (12.6%) (online Supplementary Table S2). For days to maturity, standard heterosis was equally highest for PK 472 × JS 335 and IC 501198 × JS 335 (4.81%) followed by Hardee × JS 335 (4.54%), RVS 2007-6 × JS 335 (4.34%), EC 656641 × JS 335 (3%) and Gaurav 2 × JS 335 (2.73%) (online Supplementary Table S2).

## Mean and range in $F_2$ generation

NAM population consisting of 20 cross combination ($F_2$ generation) were evaluated for yield and attributing traits. A total of 2385 progenies from these 20 crosses were evaluated for seed yield and yield-attributing traits. Enormous variability was observed in the plants for all the traits under consideration. The crosses viz., G-11 × JS 335 (27.78 g) and JS 97-52 × JS 335 (22.50 g) had highest mean seed yield amongst all. Plant with highest seed yield (82.7 g) recovered from the cross JS 97-52 × JS 335. Plants from the cross JS 335 × EC 656641 were the earliest to mature (92 days) as compared to rest of the crosses (online Supplementary Table S4).

## Genetic properties of NAM population

### Mean, range, variance, coefficient of variance, heritability and genetic advance

The statistical analysis of data on seven quantitative traits indicated wide range of variability among the accessions (online Supplementary Tables S1 and S5). Number of pods, grain yield and biomass were found to have higher range when compared to other traits studied. Grain yield (46.15/39.88), nodes/ plant (34.51/30.52), biomass (36.03/22.21), number of pods (35.63/22.21), plant height (26.22/23.14) found to have high PCV and GCV (online Supplementary Table S5). The heritability was highest for number of nodes (87.18%), plant height (77.86%) and grain yield (74.68%), and lowest heritability was recorded for the number of branches (3.17%). GA as percent mean was recorded high for the traits viz.; grain yield (71.09%), number of nodes (55.66%), plant height (42.13%) and it was lowest for branches per plant (1.74%) (online Supplementary Table S5).

### Frequency distribution of quantitative traits

Frequency distribution showed that 22% of the lines produced grain yield of 30.1–40 g per plant and 48% of the lines showed harvest index of 40.1–50% (data not shown). The number of pods also showed large variation and 25% of the lines found to have pods per plant in the range of 34–44. The trait biomass was also recorded good amount of variability and 30% of the lines produced 90–120 g per plant biomass. The frequency distribution of seven quantitative traits is presented in online Supplementary Fig. S1.

### Correlation and regression

Correlation matrix (online Supplementary Table S7) showed significant positive correlation of grain yield with biomass (0.86), harvest index (0.60), pods per plant (0.51), branches per plant (0.30), plant height (0.21) and nodes per plant (0.11) and no trait was negatively associated grain yield however, harvest index was negatively correlated with plant height (−0.11), number of nodes (−0.06) and branches per plant (−0.07). Similarly, nodes and branches (−0.09) are also negatively correlated (Fig. 1 and online Supplementary Fig. S2). Simple linear regression analysis performed for grain yield versus other traits showed that biomass, had highest direct effect on seed weight followed by harvest index, pods per plant, branches per plant, nodes per plant and plant height (data not presented).

### Genetic diversity and principal component analysis

Multivariate analysis was performed using genetic diversity and principal component analyses. All the 900 NAM-RILs were grouped into 10 clusters using *k*-means clustering. Each cluster was found to have varied number of accessions (online Supplementary Table S6). The number of accessions falling in each cluster was highest (183) in cluster 2 (C2) followed by C3 (139), C10 (133), C1 (123), C6 (81), C5 (79), C7 (71), C8 (47), C4 (37) and C9 (7). The mean value of accessions grouped into each cluster (online Supplementary Table S6) showed that accessions in C9 and C4 produced high grain yield per plant and biomass whereas cluster 4 found to has the highest harvest index. C9, C8, C7 and C4 produced highest number of pods per plant (online Supplementary Table S6). PCA used to eliminate the redundancy in dataset revealed that all the seven quantitatively measured traits have been loaded on first five components, however, major portion of variance (80.9%) in NAM-RILs is explained by the first three components with eigenvalue >1 (Table 2). The first component (PC1) accounted for 42.05% of variation largely through biomass, grain yield, number of pods and branches; PC2 accounted for 22% of variation contributed through harvest index, nodes, plant height and yield and PC3 contributed 17% through number of branches, harvest index and number of nodes. PC4, PC5 and PC6 contributed 8, 6 and 3% of total variation, respectively. Biploting of PC1 and PC2 indicated grain yield (20.31%) contributed highest to the first two principal components followed by biomass (17.74%) and number of pods (15.79%) (Fig. 2 and online Supplementary Fig. S3). Similarly, biplots of PC1 and PC2 involving only top hundred contributing RILs to the total variation identified number of superior genotypes forming clusters on the same axis (online Supplementary Fig. S4). Further, the relationship among 11 NAM families is depicted in online Supplementary Fig. S5.
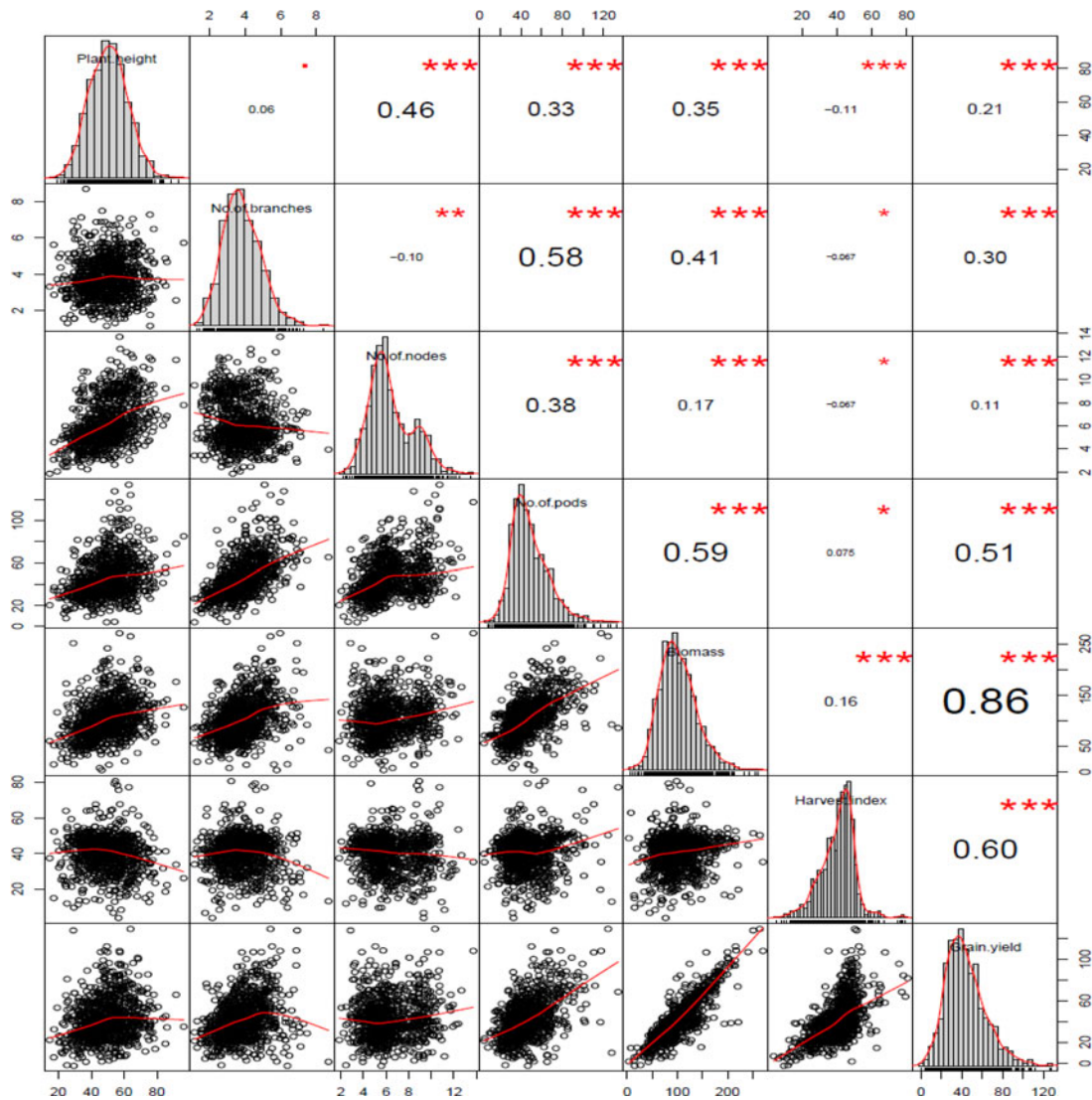
**Fig. 1.** Graphical depiction of correlation among the quantitative traits in NAM-RILs.

## Discussion

Soybean [*G. max* (L.) Merr.] is one of the most economical crops in the international market, being the fourth most produced and consumed crop worldwide (Gesteira *et al.*, 2018). Soybean [*G. max* (L.) Merr.] has been studied and enhanced for most of its economically important traits. The majority of studies on genetic variability and multivariate analysis were carried out on breeding population derived from two crosses commonly known as biparental populations. Even, the conventional plant breeding dealing with the development of new crop varieties is also based on the selection of superior progenies derived from biparental populations. Recently, researchers have started making multiparent-based populations such as MAGIC and

NAM and several successful examples of development and utilization of these populations are already demonstrated (Kover *et al.*, 2009; Cook *et al.*, 2012; Huang *et al.*, 2012; Bandillo *et al.*, 2013; Peiffer *et al.*, 2013; Mackay *et al.*, 2014; Bajgain *et al.*, 2016; Bouchet *et al.*, 2017; Fragoso *et al.*, 2017; Li *et al.*, 2017; Ren *et al.*, 2017; Song *et al.*, 2017; Shivakumar *et al.*, 2018). The most significant use of NAM population is gene mapping through the presence of two sources of recombination: shuffling of parental alleles over several generations through segregation and genetic recombination, and historical recombination of haplotypes present in the various diversity donors. The current study reports the creation of NAM population, and in the process of creation, we estimated number of genetic parameters such as estimation of $F_1$ heterosis, genetic

550 S. Maranna *et al.*

**Table 2.** Eigenvalues, variance contribution (%), variables coordinates and percent contribution of each variable on individual five principle components of 900 NAM-RILs

| | | Plant height | No. of branches | No. of nodes | No. of pods | Biomass (g) | Harvest Index | Grain yield (g) | Eigen values | Variance (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| PC1 | Coordinates | 0.47 | 0.54 | 0.36 | 0.81 | 0.88 | 0.31 | 0.85 | 2.94 | 42.06 |
| | Contribution (%) | 7.62 | 10.17 | 4.45 | 22.74 | 26.61 | 3.28 | 25.1 | | |
| PC2 | Coordinates | 0.62 | −0.07 | 0.66 | 0.18 | −0.09 | −0.68 | −0.41 | 1.52 | 21.72 |
| | Contribution (%) | 25.34 | 0.35 | 29.46 | 2.35 | 0.61 | 30.77 | 11.07 | | |
| PC3 | Coordinates | 0.25 | −0.74 | 0.42 | −0.24 | −0.02 | 0.53 | 0.22 | 1.19 | 17.1 |
| | Contribution (%) | 5.41 | 46.58 | 15.2 | 4.98 | 0.065 | 23.52 | 4.21 | | |
| PC4 | Coordinates | −0.4 | 0.12 | 0.41 | 0.33 | −0.3 | 0.23 | −0.11 | 0.62 | 8.97 |
| | Contribution (%) | 26.27 | 2.64 | 27.43 | 17.85 | 15.17 | 8.43 | 2.18 | | |
| PC5 | Coordinates | 0.39 | 0.23 | −0.15 | 0.034 | −0.32 | 0.31 | −0.11 | 0.45 | 6.49 |
| | Contribution (%) | 35.12 | 12.07 | 5.29 | 0.26 | 22.61 | 21.47 | 3.16 | | |

variability, correlations, regression, PCA and cluster analysis in NAM-RILs derived from 11 crosses. Soybean NAM developed from this study involves diverse donor founder parents for various complex traits such as yield attributing traits, biotic and abiotic stress tolerance which will facilitate the elucidation of genetic architecture of complex traits in soybean. Further, it helps in detection of QTLs and its novel alleles associated with complex traits, single-nucleotide polymorphisms and candidate genes associated with complex traits, and broadening the genetic base of soybean breeding population. Utilization of diverse parents in developing soybean NAM population enhanced the identification and introgression of allelic diversity in the breeding population. Recent advances in soybean genomics and next generation sequencing (NGS) technologies, the soybean NAM can be effectively utilized for fine mapping of complex traits and map-based cloning, genomic selection for genetic improvement of soybean.

Diers *et al.* (2018) developed soybean NAM population and evaluated 5600 inbred lines to broaden the narrow genetic base of crop. Heterosis does exist in soybean [*G. max* (L.) Merr.], soybean hybrid technology is possible, if appropriate parent combinations and an economical means of producing hybrid seed was identified (Perez *et al.*, 2009). The estimation of heterosis among 18 crosses revealed higher heterosis in some of the crosses; whereas negative heterosis is also found in few crosses studied for both grain yield and days to maturity. The differential expression of heterosis in soybean is reported by other researchers as well (Chaudhary and Singh, 1974; Gadag and Upadhyaya, 1995; Perez *et al.*, 2009). The variability observed in $F_2$ generation in 18 crosses of NAM population reflected the diversity among parents used for hybridization. The genetic variability parameters analysed in 900 NAM-RILs indicated high PCV and GCV for grain yield, biomass, number of pods, number of nodes and plant height. The similar trend of PCV and GCV for these traits reported in biparental populations (Shivakumar *et al.*, 2011; Besufikad, 2018). Our results found that high heritability coupled with high GA percent mean was recorded for plant height, number of nodes and grain yield and similar values for grain yield is reported (Shivakumar *et al.*, 2011; Besufikad, 2018; Shruti and Basavaraja, 2019) in biparental populations. Correlation studies revealed that traits such as biomass, harvest index and number of pods were having the highest significant positive effect on grain yield. The simple regression analysis also complimented these results indicating the potentiality of indirect selection of biomass, harvest index and number of pods to improve yield. The findings in biparental population and core collections in soybean also reported a similar type of association among these traits thereby indicating the traits to be kept in plant breeders eye during early generation selections (Shivakumar *et al.*, 2011;
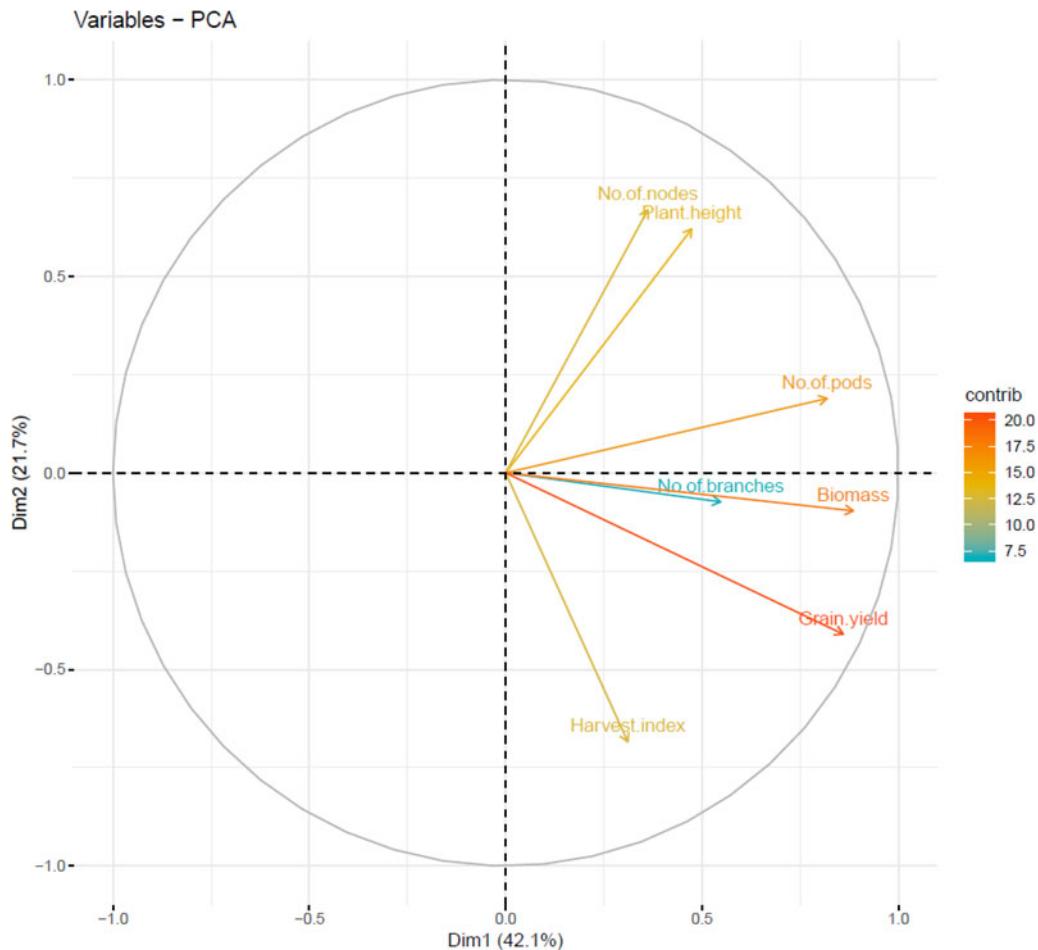


**Fig. 2.** Biploting of different variables loaded on PC1 and PC2 components.

Gireesh *et al*., 2015). Using PCA, we can able to assess the relative contribution of different components to the total divergence together with nature of forces operating at intra- and inter-cluster levels (Sharma *et al*., 2009). In the current study, first three components explained >80% of the total variation. In the first two PCs, maximum percent of variation is contributed by grain yield, biomass and number of pods. Therefore, these traits are important in explaining the genetic variation in the 900 RILs under study. The results are in agreement with various other studies that reported the maximum contribution of biomass and grain yield in soybean (Nidhi *et al*., 2018). Cluster analysis grouped the 900 RILs into 10 clusters with cluster C1, C2, C3 and C10 comprising maximum number of breeding lines indicating close relatedness of the genotypes. Similar way of grouping of 4274 accessions of common bean into 10 clusters was reported by Rana *et al*. (2015). Genotypes having highest grain yield per plant, biomass and harvest index are fallen under clusters C4 and C9. These genotypes can further be utilized in hybridization programmes for further improving yield potential.

## Conclusion

The success of development of variety is more if it involves an evaluation of large number of breeding lines derived from diverse crosses. In this direction, the current study has generated information on large number of recombinant inbred lines derived from multiple parents possessing number of useful traits such as water-logging tolerance, drought tolerance, resistance to mechanical damage, wider adaptability, long juvenility, bacterial pustule and YMV resistance. The 900 RILs derived from 11 crosses of NAM population will serve as a useful genetic resource for mapping of several economic traits segregating in these RILs and can be shared to soybean breeders to utilize in improving crop for various situations prevailing different agro-climatic zones. Thus NAM population developed in this study will

constitute useful genetic material for both applied and basic studies for genetic improvement of soybean crop.

## Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S1479262119000352.

## Acknowledgements

## References

Aravind J, Sankar MS, Dhammaprakash PW and Kaur V (2018) augmentedRCBD: Analysis of Augmented Randomised Complete Block Designs. R package version 0.1.0, https://aravind-j.github.io/augmentedRCBD/.

Arunachalam V (2017) *Quantitative genetics for quality experimentation*. Viva Books Private Limited, New Delhi. p. 154. ISBN: 978-81-309-1802-0.

Bajgain P, Rouse MN, Tsilo TJ, Macharia GK, Bhavani S, Jin Y and Anderson JA (2016) Nested association mapping of stem rust resistance in wheat using genotyping by sequencing. *PLoS ONE* 11: e0155760.

Bandillo N, Raghavan C, Pauline AM, Sevilla MAL, Irish TL, Christine JDE, Tung C, Susan M, Michael T, Ramil M, Singh RK, Gregorio G, Redona G and Leung H (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6: 1.

Besufikad EG (2018) Genetic variability, heritability and expected genetic advance in soybean [*Glycine max* (L.) Merrill] genotypes. *Agriculture, Forestry and Fisheries* 7: 108–112.

Bouchet S, Olatoye MO, Marla SR, Perumal R, Tesso T, Yu J, Tuinstra M and Morris GP (2017) Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. *Genetics* 206: 573–585.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Gracia A, Galubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Lasson S, Lepak LK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Villeda HS, Silva HS, Sun Q, Tian F, Upadyayula N, ware D, Yates H, Yu J, Zhang Z, Kresovich S and Mc Mullen MD (2009) The genetic architecture of maize flowering time. *Science* 325: 714–718.

Buezoa J, Sanz-Saezb A, Jose FM, David S, Iker A and Raquel E (2019) Drought tolerance response of high-yielding soybean varieties to mild drought: physiological and photochemical adjustments. *Physiologia Plantarum* 166: 88–104.

Chaudhary DN and Singh BB (1974) Heterosis in soybean. *Indian Journal of Genetics and Plant Breeding* 34: 69–74.

Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES and Flint-Gracia SA (2012) Genetic architecture of maize kernel composition in the nested

association mapping and inbred association panels. *Plant Physiology* 158: 824–834.

Diers BW, Specht J, Katy MR, Cregan P, Song Q, Ramasubramanian V, Graef G, Nelson R, Schapaugh W, Wang D, Shannon G, Mchale L, Kantartzi SK, Xavier A, Mian R, Stupar RM, Michno JM, An YQC, Gottel W, Ward R, Fox C, Lipka AE, Hyten D, Cary T and Beavis WD (2018) Genetic architecture of soybean yield and agronomic traits. *G3: Genes, Genomes, Genetics* 8: 3367–3375.

Fragoso CA, Moreno M, Wang Z, Heffelfinger C, Arbelaez LJ, Aguirre JA, Franco N, Romero LE, Labadie K, Zhao H, Dellaporta SL and Lorieux M (2017) Genetic architecture of a rice nested association mapping population. *G3: Genes, Genomes, Genetics* 7: 1913–1926.

Gadag RN and Upadhyaya HD (1995) Heterosis in soybean (*Glycine max* (l.) Merrill). *Indian Journal of Genetics and Plant Breeding* 55: 308–314.

Gesteira GDS, Bruzi AT, Zito RK, Vanoli F and Arantes NE (2018) Selection of early soybean inbred lines using multiple indices. *Crop Science* 58: 2494–2502.

Gireesh C, Husain SM, Shivakumar M, Satpute G, Kumawat G, Arya M, Agarwal DK and Bhatia VS (2015) Integrating principal component score strategy with power core method for development of core collection in Indian soybean germplasm. *Plant Genetic Resources: Characterization and Utilization* 15: 230–238.

Hayes HK, Immer FF and Smith DC (1955) Heterosis in hybrids In: Hayes HK and Immer FF (eds) *Methods of Plant Breeding II*. New York: McGraw-Hill, p. 551.

Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK and Cavanagh CR (2012) A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnology Journal* 10: 826–839.

ICAR (2009) *Handbook of Agriculture*. New Delhi: Indian Council of Agricultural Research, pp. 1143–1150.

Johnson HW, Robinson HF and Comstock RE (1955) Estimates of genetic and environmental variability in soybeans. *Agronomy Journal* 47: 314–318.

Jordan DE, Mace AC, Hunt C and Henzell R (2011) Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. *Crop Science* 51: 1444–1457.

Karikari B, Chen S, Xiao Y, Chang F, Zhou Y, Kong J, Akhter Bhat J and Zhao T (2019) Utilization of interspecific high-density genetic map of RIL population for the QTL detection and candidate gene mining for 100-seed weight in soybean. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2019.10:1001.

Kassambara A and Mundt F (2017) Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.5. https://CRAN.R-project.org/package=factoextra.

Kong F, Nan H, Cao D, Li Y, Wu F, Wang J *et al.* (2014) A new dominant gene *E*9 conditions early flowering and maturity in soybean. *Crop Science* 54: 2529–2535.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C and Mott R (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genetics* 5: e1000551.

Le S, Josse J and Husson F (2008) Factominer: an R package for multivariate analysis. *Journal of Statistical Software* 25: 1–18.

Li D, Pfeiffer TW and Cornelius PL (2007) Soybean QTL for yield and yield components associated with *Glycine soja* alleles. *Crop Science* 48: 571–581.

Li S, Cao Y, He J, Zhao T and Gai J (2017) Detecting the QTL-allele system conferring flowering date in a nested association mapping population of soybean using a novel procedure. *Theoretical and Applied Genetics* 130: 2297–2314.

Li D, Zhao X, Han Y, Li W and Xie F (2019) Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. *Genomics* 111: 90–95. https://doi.org/10.1016/j.ygeno.2018.01.004.

Lush JL (1940). Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. *Proceedings of the American Society of Animal Nutrition* 1940: 293–301.

Mackay IJ, Basler PB, Barber T, Alison RB, Cockram J, Gosman N, Andy JG, Horsnell R, Howells R, Donal MO, Gemma AR and Phil JH (2014) An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties and validation. *G3* 4: 1603–1610.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q *et al.* (2009) Genetic properties of the maize nested association mapping population. *Science* 325: 737–740.

Nidhi D, Avinashe HA and Shrivastava AN (2018) Principal component analysis in advanced genotypes of soybean [*Glycine max* (L.) Merrill] over seasons. *Plant Archives* 18: 501–506.

Peiffer JA, Flint-Garcia SA, De Leon N, McMullen MD, Kaeppler SM and Buckler ES (2013) The genetic architecture of maize stalk strength. *PLoS ONE* 8: E67066.

Perez PT, Silvia RC and Reid GP (2009) Evaluation of Soybean [*Glycine max* (L.) Merr.] $F_1$ Hybrids. *Journal of Crop Improvement* 23: 11–18.

Peterson BG and Carl P (2018) PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis. R package version 1.5.2. https://CRAN.R-project.org/package=PerformanceAnalytics.

Rana C, Sharma R, Tyagi RK, Chahota RK, Gautam NK, Singh M, Sharma PN and Ojha SN (2015) Characterisation of 4274 accessions of common bean (*Phaseolus vulgaris* L.) germplasm conserved in the Indian gene bank for phenological, morphological and agricultural traits. *Euphytica* 205: 441–457.

Ren Y, Hou W, Lan C, Basnet BR, Singh RP, Zhu W, Cheng X, Cui D and Chen F (2017) QTL analysis and nested association mapping for adult plant resistance to powdery mildew in two bread wheat populations. *Frontiers in Plant Science* 8: 1212.

Sharma MK, Mishra S and Rana NS (2009) Genetic divergence in French bean (*Phaseolus vulgaris* L.) pole type cultivars. *Legume Research* 32: 220–223.

Shivakumar M, Basavaraja GT, Salimath PM, Patil PV and Talukdar A (2011) Identification of rust resistant lines and their genetic variability and character association studies in soybean [*Glycine max* (L.) Merr.]. *Indian Journal of Genetics and Plant Breeding* 71: 235–240.

Shivakumar M, Gireesh C and Talukdar A (2016) Efficiency and utility of pollination without emasculation (PWE) method in intra and inter specific hybridization in soybean. *Indian Journal of Genetics and Plant Breeding* 76: 98–100.

Shivakumar M, Kumawat G, Gireesh C, Ramesh SV and Husain SM (2018) Soybean MAGIC population: a novel resource for genetics and plant breeding. *Current science* 114: 906–908.

Shruti K and Basavaraja GT (2019) Genetic variability studies on yield and yield component traits of soybean. *International Journal of Current Microbiology and Applied Sciences* 8: 1269–1274.

Song Q, Yan L, Quingley C, Jordan BD, Fickus E, Schroeder S, Song BH, Charles AYQ, Hyten D, Nelson R, Rainey K, Beavis WD, Specht J, Diers B and Cregan P (2017) Genetic characterization of the soybean nested association mapping population. *The Plant Genome* 10. doi:10.3835/plantgenome2016.10.0109.

Talukdar A and Shivakumar M (2012) Pollination without emasculation: an efficient method of hybridization in soybean (*Glycine max* Merr). *Current Science* 103: 628–630.

United States Department of Agriculture (USDA) (2019) Foreign Agriculture Service. *World Agriculture Production*, Table 11. Circular Series WAP 10-19 pp 24.

Wang J, Chu S, Zhang H, Zhu Y, Cheng H and Yu D (2016) Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Scientific Reports* 6: 20728.

Wei T and Simko V (2017) R package 'corrplot': Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot.