



Brief Communication

The ricebean genome provides insight into *Vigna* genome evolution and facilitates genetic enhancement

Aleena Francis¹, Nagendra Pratap Singh¹, Mohar Singh², Paras Sharma^{3,4}, Gayacharan², Durgesh Kumar⁵, Udit Basu¹, Deepak Bajaj¹, Nidhi Varshney¹, Dinesh Chandra Joshi⁶, Dinesh Prasad Semwal², Vandana Tyagi², Dhammaprakash Wankhede², Rakesh Bharadwaj², Amit Kumar Singh², Swarup Kumar Parida^{1,*}  and Debasis Chattopadhyay^{1,*} 

¹National Institute of Plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi, India

²ICAR-National Bureau of Plant Genetic Resources (NBPGR), Pusa Campus, New Delhi, India

³ICMR-National Institute of Nutrition (NIN), Hyderabad, Telangana, India

⁴Department of Food Technology, Mizoram University, Aizawl, Mizoram, India

⁵Division of Genetics, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

⁶ICAR-Vivekananda Institute of Hill Agriculture, Almora, Uttarakhand, India

Received 6 October 2022;

revised 29 April 2023;

accepted 8 May 2023.

*Correspondence (Debasis Chattopadhyay, Tel +91-11-26735189; fax +91-11-26741658; email debasis@nipgr.ac.in); (Swarup Kumar Parida, Tel +91-11-26735228; fax +91-11-26741658; email swarup@nipgr.ac.in)

Keywords: agronomic trait, evolution, genome assembly, ricebean, *Vigna umbellata*.

Ricebean [*Vigna umbellata* (Thunb.) Ohwi and Ohashi] ($2n = 2x = 22$) is a warm-season dietary pulse legume crop and was originated in the Indo-China region. It is known to provide food security to the small and marginal farmers of South and South-East Asia. Ricebean is well known for its high nutritional quality and resistance to bacterial leaf spot, Mungbean yellow mosaic virus and bruchid, which are devastating for the other *Vigna* family crops (Dhaliwal *et al.*, 2022). We report a reference grade *de novo* genome assembly, which is anchored to the genetic linkage groups and covered almost the whole estimated genome length of ricebean and so far, the largest among the sequenced *Vigna* species.

A pure high-yielding Indian ricebean variety VRB3 (Him Shakti) was used for single molecule real time (SMRT) long-read whole genome sequencing (126.53 Gb; Figure S1a,b) to assemble 1287 contigs with an N50 of 1.71 Mb and a total length of 605.22 Mb. Optical mapping (589.03 Gb) and Hi-C library sequence reads (145.40 Gb) were used for correcting and scaffolding these contigs to produce an assembly of 889 scaffolds and contigs with a total length 626.88 Mb with N₅₀ and the longest scaffold length being 31.81 and 74.76 Mb, respectively (Tables S1 and S1). Illumina short reads (186.45 Gb) were employed for the post-processing by polishing and gap filling. The assembled scaffolds were anchored and oriented following an ultra-high-density consensus genetic linkage map comprising of 25 633 SNPs and constructed by using two mapping populations [VRB3 (IC595428) × PRR1 (IC360590) and VRB3 (IC595428) × SKMRB1] (Table S3, Figure S3). The said genetic linkage map was used to construct 11 chromosome pseudomolecules with a total chromosomal length of 619.01 Mb (Table S4). Optical mapping, k-mer analysis (Figure S1c) and flow-cytometry (Datta and Gupta, 2009; Figure S2) suggested the estimated

genome length of ricebean between ~578 and 633 Mb, respectively. With 630.50 Mb as the average size based on k-mer and optical mapping, this assembly covered nearly 99.5% of the estimated genome length, the largest chromosome-scale assembly so far of a crop belonging to *Vigna* genus (<https://phytozome-next.jgi.doe.gov>, Guan *et al.*, 2022). A reference normalised transcriptome was generated by IsoSeq method to produce 33 004 transcripts of total 94.43 Mb in length (Table S5). The repeat-masked genome sequence was used to annotate a non-redundant and high-confidence consensus gene model set of total 37 489 protein-coding genes (Tables S6 and S7, Text S1). Orthologs of about 92% (36883) of the annotated ricebean proteins were found in the NCBI nr database. (Table S8; Text S2). BUSCO analysis showed a presence of 97.2% of 2326 near-universal single copy homologue genes from Eudicot_s_odb10 suggesting a near-complete genome assembly of ricebean. The authenticity, robustness and continuity of our genome assembly (VRB3) derived by integrating data generated from various technical platforms/strategies (PacBio, Optical mapping, Hi-C, Illumina WGS, high-density genetic linkage map) and its collinearity with the assemblies of the related species is clearly evident from our results (Text S3; Figures S4–S7).

A total of 669 duplicated syntenic blocks containing 12 912 genes (34.44% of all genes) were detected in the assembly (Table S9; Figure 1a). Clustering of orthologous gene families among the five sequenced *Vigna* species along with *P. vulgaris* and *G. max* all belonging to the tribe Phaseoleae showed these group shares 16 696 ortho-groups comprising of 195 860 genes (48.5%; Figures S8–S12). The distribution analysis of synonymous substitution rates (K_s) between the paralogous gene-pairs (Figure 1b) suggested a whole genome duplication (WGD) about 51.23 million years ago (mya), which is consistent with the WGD event of Papilionoideae (Lavin *et al.*, 2005). The K_s between the single-copy orthologous genes revealed a divergence period between ricebean and adzuki bean, and between mungbean and blackgram at 2.54 mya (million years ago) while the *Vigna* crown clade diverged from the ancestor of common bean about 8.5 mya, consistent with the crown age calculated based on the fossil (Lavin *et al.*, 2005; Figure 1c). This divergence caused expansion and contraction of 72 and 173 gene families, respectively, in the clade of ricebean and adzuki bean, while the divergence from adzuki bean resulted in expansion and contraction of 200 and 988 gene families, respectively, in ricebean (Figure 1c) with an

enrichment in the squalene epoxidase/monooxygenase genes, the rate limiting enzymes for highly abundant triterpenoid saponin biosynthesis in ricebean (Zhao *et al.*, 2010; Tables S10 and S11, Text S4).

Construction of *Vigna* ancestral karyotype by taking syntenic blocks of the five sequenced *Vigna* crops after genome-to-genome alignment revealed maintenance of large syntenic blocks between their genomes. The outgroup common bean shares 9 out of 11 of the *Vigna* ancestral proto-chromosomes (Figure 1d). Chromosome 6 of common bean did not possess any of the syntenic blocks of the *Vigna* ancestral proto-chromosomes while, proto-chromosome 5 of the *Vigna* ancestral karyotype did not show any presence in common bean chromosomes. Cowpea (*V. unguiculata*) possesses the highest (10 out of 11 chromosomes) one-to-one chromosomal alignment with the ancestral *Vigna* karyotype suggesting its closest proximity to the *Vigna* ancestral proto-chromosomes. The recombination between Chr 1 and 9 of the ancestral *Vigna* species was found to be a signature of all the sequenced extant *Vigna* species and is absent in the outlier common bean genome (Text S5).

A total of 1983 Gb sequence data with an average of 8.31x genome coverage per accession was generated by sequencing whole genomes of 353 ricebean accessions representing diverse eco-geographical regions of South-East Asia and India to identify a total of 10.56 million SNPs (Table S12) and was used to calculate pair-wise nucleotide diversity. Despite a high nucleotide diversity across the 11 chromosomes, a genomic region of ~15 Mb in the chromosome 2 showed unusually low nucleotide

diversity (Figure 1e; Table S13). This region codes for 30S ribosomal protein S5, WER-like transcription factor, ABC transporter and Ec-AMP-D2-like Defensin which has antifungal activity (Odintsova *et al.*, 2020). Nine genomic regions of total 1.03 Mb showed strong purifying selection (Tajima's D < -2). Genes which code for disease resistance and protein Exordium-like, a potential mediator of brassinosteroid (BR)-promoted growth were found in this region (Coll-Garcia *et al.*, 2004; Table S14, Text S6).

Phylogenetic analysis grouped all 353 accessions into seven clusters (Figure 1f). The neighbour-joining tree clustered the most of the North India-Nepal accessions with the North-East Indian accessions and both the populations showed similar nucleotide diversity (7.2×10^{-2} and 7.3×10^{-2} , respectively). Principal component analysis and distance matrix (Figure S13; Table S15) suggested that accessions from North India-Nepal and North-East India underwent a limited and wider diffusion, respectively, and are clustered with the accessions collected from other parts suggesting that this region was the major centre of diffusion of ricebean. However, these two populations have the same origin as inferred from the similar nucleotide diversity and genetic distance (F_{st}) (7.2×10^{-2} and 7.3×10^{-2} ; Table S16). Our data suggest that one path diffused to Africa (6.4×10^{-2}) and South India (6.8×10^{-2}) and another path to South-East Asia (1.5×10^{-1}) and Sri Lanka (1.1×10^{-1}). Admixture analysis predicted seven ancestral populations with the least cross-validation error (Figure S14; Text S7).

We constructed a ricebean pangenome of 679.32 Mb in size using 353 ricebean accessions (Table S17; Text S8). GWAS

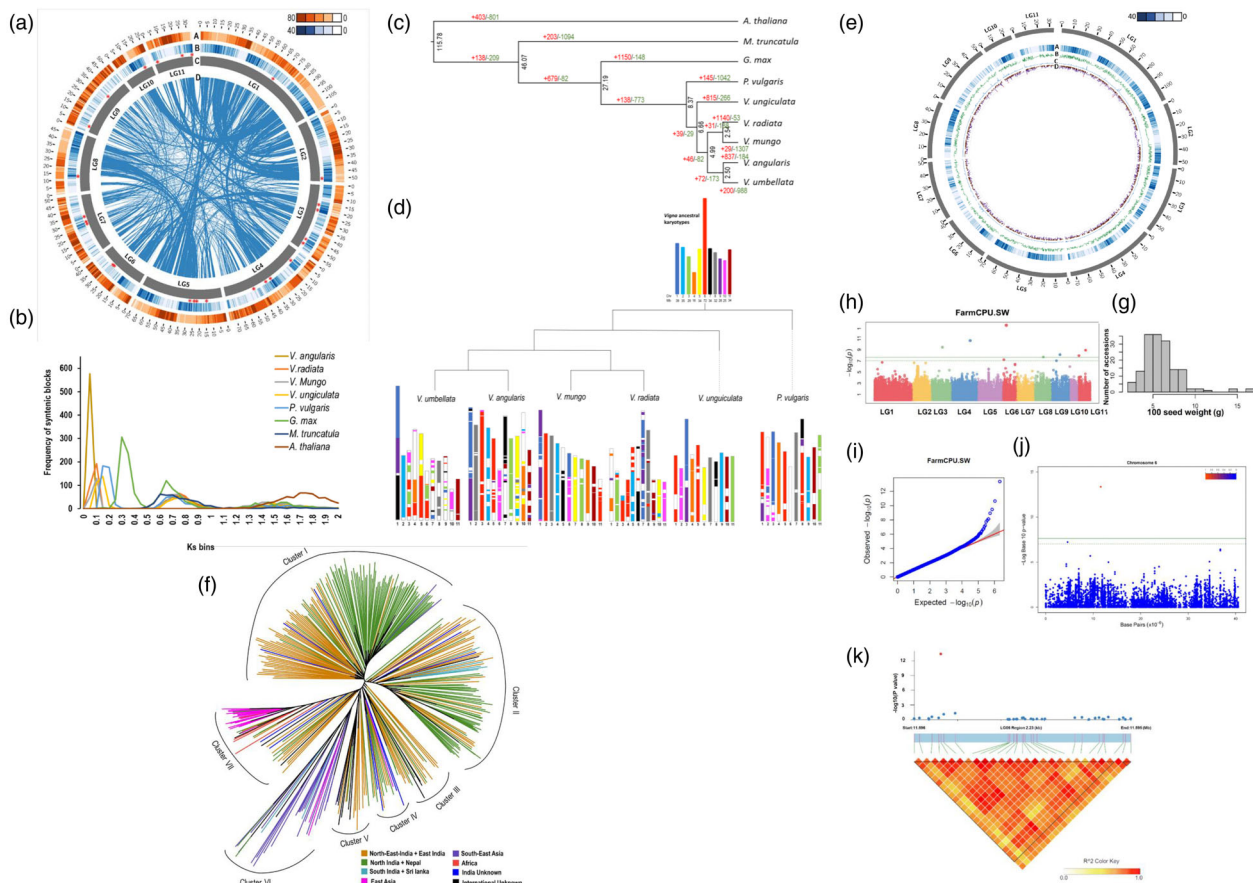


Figure 1 Various genomic features of assembled *V. umbellata* var. VRB3 genome (Detail in Supplementary Text).

analysis using the genotyping information of 2 145 937 SNPs and phenotyping data evaluated (Table S18; Figure S15) in a diversity panel of 353 ricebean accessions identified 241 and 64 genomic loci showing significant association with the 11 agro-morphological and 9 nutritional quality traits, respectively, at a $P \leq 10^{-6}$ (with FDR cut-off ≤ 0.05) (Figure S16–S18; Figure 1g–k; Table S19; Text S9). All the genomic resources created in this study has been curated in the ricebean portal database (www.ricebeanportal.com; Figure S19).

Author Contributions

AF and NPS assembled, annotated, analysed the genome and pangenome, and analysed the resequencing data of the ricebean population. MS developed the mapping populations and maintained ricebean lines. PS and RB evaluated the nutritional parameters, and Gayacharan and DCJ evaluated the agromorphological parameters. DK estimated the genome size by flow cytometry. UB, DB and NV constructed the linkage map and prepared the samples for sequencing. DPS and VD were responsible for germplasm collection and curation. DW and AKS guided the transcriptome analyses. SKP and DC conceived, designed and coordinated the research project and wrote the manuscript.

Acknowledgements

Authors acknowledge Department of Biotechnology, Ministry of Science and Technology, Government of India (DBT), Indian Council of Agricultural Research (ICAR) and National Institute of Plant Genome Research (NIPGR), India. Assistance of DBT-eLibrary Consortium (DeLCON) for providing access to e-Resources is acknowledged. AF acknowledges Council of Scientific and Industrial Research (CSIR, Govt. of India), and DC acknowledges Science and Engineering Board, Department of Science and Technology for fellowships (JCB/2020/000014).

Conflict of interest

Authors declare no conflict of interest.

Funding

This project was funded by the Department of Biotechnology (DBT) by a grant (BT/Ag/Network/Pulses-I/2017-18). The funder

had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability statement

Submitted to NCBI (<https://www.ncbi.nlm.nih.gov/>) with BioProject ID: *PRJNA821999: Vigna umbellata cultivar: VRB3 Genome sequencing and assembly PRJNA822062: Vigna umbellata cultivar: VRB3 Raw sequence reads and Transcriptome assembly* NCBI Release Genebank ID (<https://www.ncbi.nlm.nih.gov/nucleotide/JALIRJ000000000>). Submitted to Figshare (https://figshare.com/articles/dataset/Ricebean_Genome/22704943).

References

- Coll-Garcia, D., Mazuch, J., Altmann, T. and Müssig, C. (2004) EXORDIUM regulates brassinosteroid-responsive genes. *FEBS Lett.* **563**, 82–86.
- Datta, S. and Gupta, S. (2009) Use of genomic resources in improvement of *Vigna* species. *J. Food Legumes* **22**, 1–10.
- Dhaliwal, S.K., Gill, R.K., Sharma, A., Kaur, A., Bhatia, D. and Kaur, S. (2022) A large-effect QTL introgressed from ricebean imparts resistance to Mungbean yellow mosaic India virus in blackgram (*Vigna mungo* (L.) Hepper). *Theor. Appl. Genet.* **135**, 4495–4506.
- Guan, J., Zhang, J., Gong, D., Zhang, Z., Yu, Y., Luo, G., Somta, P. et al. (2022) Genomic analyses of rice bean landraces reveal adaptation and yield related loci to accelerate breeding. *Nat. Commun.* **13**, 5707.
- Lavin, M., Herendeen, P.S. and Wojciechowski, M.F. (2005) Evolutionary rates analysis of leguminosae implicates a rapid diversification of lineages during the tertiary. *Systematic Biol.* **54**, 575–594.
- Odintsova, T.I., Slezina, M.P. and Istomina, E.A. (2020) Defensins of grasses: A systematic review. *Biomolecules* **10**, 1029.
- Zhao, C.L., Cui, X.M., Chen, Y.P. and Liang, Q. (2010) Key enzymes of triterpenoid saponin biosynthesis and the induction of their activities and gene expressions in plants. *Nat. Prod. Commun.* **5**, 1147–1158.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1 Supplementary Materials and Methods.

Figure S1–S19 Supplementary Figures.

Table S1–S19 Supplementary Tables.

Text S1–S9 Supplementary Texts.