Expanding genetic diversity in common bean through EMS mutagenesis: a novel platform for variant discovery and candidate genes validation

Angelo Gaiti[1] ,Chiara Russo[2], Eleonora Cominelli[2], Giulio Testone[3], Paolo Cozzi[2], Dario Paolo[2], Massimo Galbiati[2], Alessia Losa[4], Tea Sala[4], Elena Avite[5], Paolo Andreatta[6], Carlo Pozzi[7], Francesca Sparvoli[2]

[1] *Department of Food, Environmental and Nutritional Sciences, University of Milan, Milan, Italy*
[2] *Institute of Agricultural Biology and Biotechnology, CNR, Milan Italy*
[3] *Institute for Biological Systems, CNR, Rome, Italy*
[4] *Research centre for Genomics and Bioinformatics, CREA, Montanaso Lombardo (LO), Italy*
[5] *Verisem Italia, Piacenza, Italy*
[6] *SmeUp, Erbusco (BS), Italy*
[7]*Department of Agricultural and Environmental Sciences - Production, Landscape, Agroenergy, University of Milan, Milan, Italy*

KEYWORDS

Common bean; mutagenesis; WGS.

TEXT

Common bean (*Phaseolus vulgaris* L.) is the primary grain legume consumed directly by humans worldwide and is increasingly valued for its high nutritional quality, making it a prime target for genomics research. Despite its importance, genomic resources for this species remain limited.

Within the PhasTILL and AGRITECH projects, an EMS-mutagenized population of 2,345 M1 lines was recently developed from an Italian cultivated variety (Meccearly dwarf Borlotto type, Verisem Italia). This population represents a unique resource for isolating mutants with desirable traits. To date, M3 seeds obtained from approximately 2,300 M2 plants derived from 590 M1 lines are available. Phenotypic screening in M2 and M3 generations has identified major alterations in germination and plant morphology. DNA has been collected from all M2 plants, and whole-genome sequencing (WGS) is ongoing to characterize induced variants.

A pilot WGS study was conducted to optimize variant calling pipelines for EMS-induced mutations. The dataset included two wild-type controls and six M2 pools of four individuals each (4X pooling) sequenced at 40X coverage, and four individual plants from two M2 pools sequenced at 10X coverage. Variant calling was performed using five strategies varying in caller, pooling, and merging approaches, including Freebayes (single-run, batch-wise, octaploid pooling) and HaplotypeCaller with a per-sample gVCF workflow. After filtering, an average of 26,713 variants per sample was identified. HaplotypeCaller detected approximately five times more canonical variants than Freebayes (85,657 vs 11,110–23,363), but with lower confirmation rate in 1X samples (53–62% vs 92–99%). Importantly, 53–94% of canonical variants specific to two M2 pools and confirmed in 1X samples were also recovered by HaplotypeCaller.

These results demonstrate that EMS-induced variants can be reliably identified using short-read WGS of both pools and individual mutants. The observed differences in variant calling highlight the importance of balancing sensitivity and precision. In conclusion, this population broadens genetic diversity in common bean and represents a valuable platform to validate candidate genes, identify potential targets for susceptibility to diseases and accelerate functional genomics studies aimed at improving performance, productivity, and nutritional quality.