

ADVANCES IN COMMON BEAN GENOME-WIDE ANALYSIS AND IMPACTS FOR THE BREEDING PROGRAMS

R. P. VIANELLO

Ph.D. in Plant Molecular Biology

Embrapa Arroz e Feijão, Santo Antônio de Goiás, Goiás.

rosana.vianello@embrapa.br

Key word: *Phaseolus vulgaris*; molecular breeding; association panel.

Conservation and knowledge of common bean genetic resources for sustainable management and proper utilization are of great importance to *increase genetic gain in breeding programs*. The Brazilian common bean core collection (CONFÉ) represents the genetic diversity of a large collection and presents a great potential to be widely explored to improve utilization of germplasm for association analysis of agronomic traits and genome selection. In the last years, SNP markers have been increasingly developed based on the analysis of the accessions from CONFÉ, and applied for genetic analysis, investigation of the genetic structure along the domestication sites and breeding programs, as well as the identification of genomic regions related to traits of agronomic interest through association analysis and genomic selection.

The first effort to provide useful SNPs for Brazilian germplasm bank was based on restriction site-associated DNA (RAD sequencing) from DNA pooled from 12 diverse common bean genotypes (Valdisser et al. 2016). A set of 1,032 RAD-SNPs were identified and a subset of 384 RAD-SNPs revealed a *successful amplification rate* (97%) and a great genetic diversity (H_e) for the Andean (0.161) and Mesoamerican group (0.156). This study integrates a group of 580 SNPs (289 RAD-SNPs and 291 BARC-SNPs) genotyped for the same set of accessions, of which 70% were in Linkage Disequilibrium (LD). In conclusion, RAD sequencing allowed the discovery of hundreds of useful SNPs that initially provided an excellent panel of molecular tools for whole genome analysis. Genotyping of these SNPs on a germplasm of the Agronomic institute of Campinas (IAC) allowed conducting a GWAS analysis for Anthracnose and Angular Leaf Spot, where new associations were detected with an increased mapping resolution (Perseguini et al. 2016). In parallel, using the same SNP set, other common bean accessions, evaluated for traits of agronomic interest, were genotyped and five genomic selection models, i.e., Bayes A (BA), B (BB), C (BC), LASSO (BL) and Ridge regression (BRR) were tested, of which the accuracy values were estimated, followed by the indication of the models that performed better (Barilli et al. submitted).

More recently, based on DArT technology using next generation sequencing platforms (DArTseq™), 6,286 SNPs (1 SNP/86.5 Kbp) were genotyped in genic (43.3%) and non-genic (56.7%) regions. Genetic subdivision associated to the gene pools ($K = 2$) and grain types ($K = 3$ and $K = 5$) were reported. In general, the genetic diversity analysis revealed an average H_e of 0.442 for the whole collection, 0.102 for Andean and 0.168 for Mesoamerican gene pools. The SNP effects prediction revealed a predominance of impact on non-coding regions (77.8%) and SNPs under selection were identified within gene pools comparing landrace and cultivar/line (Andean: 18 accessions; Mesoamerican: 69 accessions) and between the gene pools (59 SNPs). The LD extension estimate corrected for population structure and relatedness (r^2_{sv}) was ~ 88 kbp, while for the Andean gene pool was ~ 395 kbp, and for the Mesoamerican was ~ 130 kbp (Valdisser et al. 2017). This set of germplasm was also evaluated characterized to the traits plant architecture, lodging resistance and productivity across environments using DArT and DArTseq genotyping. Based on two analytical approaches, genome-wide association studies (GWAS) and regional heritability mapping (RHM), genetic variants affecting the complex traits were identified. RHM detected a larger number of QTLs and captured considerably more of the genomic heritability that accounted for larger fractions of the additive genetic variance regions than GWAS. For the trait lodging resistance, an important component for plant architecture, a genomic region with environmental stability and significant effect of allelic substitution was found, enabling the use for marker-assisted selection (Resende et al. submitted).

As part of our effort for drought tolerance breeding in Brazilian common bean, we conducted a transcriptome study, by RNA-seq, to discover the differentially expressed genes in response to drought stress. From 28,590 genes identified, 1,618 were first described for the common bean and 1,242 were differentially expressed genes in leaves and roots tissues of those contrasting genotypes, BAT477 and Pérola. These genes were extensively explored for their pattern of differential expression between genotypes and among treatments of dehydration, and categorized in GO terms. Validation of RNA-seq was performed using qPCR, and we demonstrated that transcriptome analysis by RNA-seq was reliable for gene identification. The raw data files were submitted to the NCBI Sequence Read Archive (SRA) database to be accessed for common bean researchers worldwide (Pereira et al. submitted). In addition, a set of 5,293 genes expressed under drought stress was targeted for resequencing in a sample composed of 175 common bean accessions to identify SNPs. Based on these SNPs and those previously obtained by DArTseq in a sample composed of 343 common bean accessions of Mesoamerican origin, a GWAS analysis for drought tolerance was conducted (*manuscript in elaboration*). The traits of 100-seed weight and yield were evaluated for three consecutive years in environments with and without drought stress. The association analyzes were performed using the mixed linear model implemented in the Tassel software and resulted in the identification of several SNPs associated with the evaluated traits. Most SNP effects predicted were of modifier type (80%). Additionally, in this study, an association panel for the Mesoamerican germplasm based on hundreds of thousands SNPs in LE was made available to be used for association analysis of any trait of interest in common bean.

These results account for progress towards the validation of a SNP array to be implemented into operational use in markers assisted selection (set of SNPs in LE) and integration of genomic selection strategies into breeding program of common bean. In addition, a common bean panel (343 Mesoamerican and 64 Andean genotyped with ~17.000 SNPs) was made available to be phenotyped for any traits of interest creating opportunities for diverse studies.

References:

BARILLI et al. Genome predictions for common bean through Bayesian models. *Scientia Agricola*, submitted.

PEREIRA W.J.; MELO A.T.O.; COELHO A.S.G.; RODRIGUES F.A.; MAMIDI S.; DE ALENCAR S.A.; LANNA A.C.; VALDISSER P.A.M.R.; BRONDANI C.; NASCIMENTO-JÚNIOR I.R.; BORBA T.C.O.; VIANELLO R.P. A drought transcriptional map reveals new genes for the Mesoamerican common bean germplasm. **BMC Genomics**, submitted.

RESENDE R.T.; DE RESENDE M.D.V.; AZEVEDO C.F.; SILVA F.F.; MELO L.C.; PEREIRA H.S.; SOUZA T.L.P.O.; VALDISSER P.A.M.R.; BRONDANI C.; VIANELLO R.P. Genome-wide association and regional heritability mapping of plant architecture, lodging and productivity in *Phaseolus vulgaris*. **Heredity**, submitted.

VALDISSER P.A.M.R.; PAPPAS G.J.; DE MENEZES I.P.P.; MÜLLER B.S.F.; PEREIRA W.J.; NARCISO M.G.; ET AL. SNP discovery in common bean by restriction-associated DNA (RAD) sequencing for genetic diversity and population structure analysis. **Mol Genet Genomics** 291: 1277–1291, 2016.

VALDISSER P.A.M.R.; PEREIRA W.J.; ALMEIDA FILHO J.E.; MÜLLER B.S.F.; COELHO G.R.C.; DE MENEZES I.P.P.; et al. In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. **BMC Genomics** 18: 423, 2017.

Acknowledgments: EMBRAPA Macroprograma; CNPq; CAPES.