



10° CBMP

CONGRESSO BRASILEIRO DE
MELHORAMENTO DE PLANTAS
ÁGUAS DE LINDÓIA/SP | 2019

28 A 31
JULHO
2019

TEMA:
**PESQUISA E
INOVAÇÃO**
PARA O
**DESENVOLVIMENTO
DA SOCIEDADE**

REFERENCE GRADE GENOME SEQUENCE ASSEMBLY FOR A BRAZILIAN SOYBEAN GERMPLASM

Orzenil B Silva-Junior^{1*}; Maria E.L Sá^{1,2}; Marco Pessoa-Filho³; Carlos L.P. de Melo⁴; Carlos A.A. Arias⁴; Dario Grattapaglia¹; Elibio L. Rech¹

¹Embrapa Recursos Genéticos e Biotecnologia, Brasília, Brasil. ²Empresa de Pesquisa Agropecuária de Minas Gerais, Belo Horizonte, Brasil. ³Embrapa Cerrados, Brasília, Brasil. ⁴Embrapa Soja, Londrina, Brasil.

*orzenil.silva@embrapa.br

Palavras-chave: genomics, DNA sequencing, soybeans.

In selective breeding, remarkable morphological, physiological and genetic changes occur in plant populations in a relatively short time. The extent of variation in these systems offers a formidable opportunity to detect allele frequency shifts correlated to directional processes favoring specific phenotypes. However, many of these phenotypes are complex, influenced by multiple genes and environmental factors and do not follow readily detectable patterns of inheritance. The complex interplay of sampling, genetic drift and natural selection imposed by the environment at play during selective breeding, result, together, in domestication, which ultimately encompasses modifications in the gene pool. As a matter of course, the fluctuations in population size during the domestication can bring dramatic consequences to the genetic diversity available for long-term breeding progress. Data analytical tools and multiple experimental techniques for the investigation of complex trait inheritance have contributed to the emergence of a new understanding on how genome-wide effects change dramatically over relatively fast time scales during domestication. The generation of a genome sequence assembly is a key driver to develop such powerful studies. On the basis of a single linear reference genome sequence for a species, methodologies for retrieving genome-wide information are applied, such as the construction of dense genotyping panels based on a priori-established polymorphic markers (SNP genotyping panels). While this approach allows genotyping large number of individuals at low costs, it usually suffers from ascertainment bias and ignores rare variants that might be underlying important traits. Whole-genome resequencing of individual plants has become an economically viable alternative to recover most of the genetic variation in a germplasm base. However, the use of a single reference genome sequence to analyze the sequence variation across individual samples can impede the accurate alignments of reads, which, in turn, can affect the downstream analyses such as the identification of large and small DNA variants. Here, we report a reference grade genome sequence assembly for the soybean (*Glycine max*, $2n = 40$) germplasm accession EMBRAPA “BR14-3465”, generated by a combination of SMRT, Hi-C and synteny-based approaches. The genome assembly totals 1.091 Gb in size, with a contig N50 of 1.64 Mb and a scaffold N50 of 56.31 Mb, mostly arranged in 20 large sequences and with gaps accounting only to 2.57 Mb. This first Brazilian soybean genome sequence assembly adds to the recent availability of similar genome sequences for modern soybean germplasm accessions worldwide (Lee, USA; ZhongHuang13, China; Enrei, Japan) and wild soybean (W05, China). More specifically, this quality assembly for soybean can provide information on genomic variation not captured in the single reference (Williams82, USA) and support ongoing efforts for linking genotypic to phenotypic data during domestication for (sub)tropical regions of low latitudes.

Agradecimentos: INCT BioSyn, CNPq; Embrapa-Monsanto Fund.