

Draft genome assembly of forage peanut (*Arachis pintoi*)

Oliveira, J.C (1), Assis, G.M.L (2), Campos, T. (3,4), Pessoa-Filho, M. (5)

1 Universidade Federal do Acre, Rio Branco, AC, Brazil

2 Embrapa Gado de Corte, Campo Grande, MS, Brazil

3 Embrapa Acre, Rio Branco, AC, Brazil

4 Embrapa Cerrados, Brasília, DF, Brazil

5 Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil

Forage peanut (*Arachis pintoi* Krapov. and W.C. Greg.), a forage legume native to Brazil, is widely used as a green cover crop in several countries. Its benefits include soil moisture retention, nutrient cycling, biological nitrogen fixation, recovery of degraded areas, and erosion prevention. Its usage in pastures intercropped with grasses has increased daily weight gain by up to 50% and reduced cattle slaughter time by up to 9 months. However, its adoption is limited by the high costs of seed production, low establishment, decreased yields in dry seasons, and susceptibility to viral diseases. The lack of molecular tools to support cultivar development has been listed as one of the bottlenecks of breeding programs. High-quality genome assemblies are fundamental for the development of genomic resources for crops. This work presents a draft genome assembly of *A. pintoi* based on third-generation HiFi reads. High molecular weight DNA was extracted from young leaves of a single diploid individual. Sequencing was performed on a Sequel II and a Revio system at Maryland Genomics, using two SMRT cells on CCS HiFi mode, with 30-hour and 24-hour runtimes, respectively. Approximately 68 Gbp were obtained in 6.6 million reads, with an average size of 10kbp, representing 61x genome coverage. Analysis of k-mer frequency spectra indicated a heterozygous and highly repetitive genome with a size of 1.1 Gbp. A partially phased draft assembly was obtained with Hifiasm in HiFi-only mode. The initial draft had 1,186 contigs with a total length of ~1.5 Gbp. Blobtools was used to filter contaminant contigs, and MITOhifi was used to assemble mitochondria and chloroplast genomes and filter organelle contigs. The final draft contains 428 contigs with a total size of ~1.5 Gbp, higher than the 1.1 Gbp genome size previously estimated for the species. Contig N50 was 31.6 Mbp, with 39 contigs longer than 10 Mbp (9.1% of the total), and the longest contig had 106.8 Mbp. The assembled gene space assessment

using BUSCO (embryophyta_odb10) showed 99% complete orthologs, with only 0.6% fragmented and 0.4% missing BUSCOs. RepeatModeler and RepeatMasker were used to annotate repetitive regions, and 78.25% of the assembly was masked as repeats. Hi-C will be used to group and orient contigs in chromosome-scale scaffolds. RNAseq data will be used to predict and annotate gene models in the final assembly. This is the first genome assembly in the genus for a species not belonging to the Arachis section, increasing the understanding of the genomic diversity of Arachis species. It will be an invaluable resource for forage peanut breeding and genetics, allowing the development of state-of-the-art genotyping technologies to support the development of new cultivars.

Keywords: forage legume, genomics, genomic resources, third-generation sequencing.

Funding: Embrapa 10.19.03.038.00.00, 20.20.01.001.00.00; Governo Federal/CNPq/Governo do Estado do Acre/FAPAC TO:003/2023.