Chloroplast genome assemblies for four *Vanilla* species provide targets for the development of molecular markers for species differentiation

Pappas, Marília (1); Nunes, Tayara C. (1); Alves, Rosa B. N. (1); Rocha, Fernando S. (2); Vieira, Roberto F. (1); Bianchetti, Luciano B. (1); Pappas Júnior, Georgios J. (3); Pessoa-Filho, Marco (1)

(1) Embrapa Recursos Genéticos e Biotecnologia; (2) Embrapa Cerrados; (3) Universidade de Brasília; marilia.pappas@embrapa.br

Vanilla is a valuable spice crop, and continuous initiatives in Brazil have expanded germplasm collections with local species possessing economic potential. These efforts focus on morphological identification, chemical characterization, propagation protocol development, and capacity building. Understanding the genetic diversity and taxonomy of natural populations is crucial for selecting *Vanilla* species with aromatic potential. Chloroplast genome assemblies for six *Vanilla* species are publicly available and could provide useful structural and sequence variations for phylogenetic analyses. In this work, we present plastid assemblies for three species (*V. phaeantha* Rchb. f., *V. chamissonis* Klotzsch, and *V. columbiana* Rolfe) for which no genomic information was available, and a new assembly for *V. pompona* Schiede. Accessions belonging to the *Vanilla* germplasm collection at Embrapa Recursos Genéticos e Biotecnologia were selected for DNA extraction from previously desiccated and ethanol-treated leaf tissues. Genomic libraries were pooled and paired-end sequenced (2x150bp). Raw reads were trimmed and cleaned of contaminants with BBDuk. Clean reads were input to GetOrganelle v1.7.7.1, with default parameters for plant plastids. For *V. phaeantha*, word size was set to 102, and a previous incomplete chloroplast assembly obtained with NOVOPlasty was used as seed to generate a complete plastid genome assembly. Assembled genomes were uploaded to the GeSeq web server for semi-automated annotation of CDS, rRNA, organellar tRNAs, and Inverted Repeats (IRs). Graphical maps of chloroplast genome annotations were generated with OGDRAW. Chloroplast genome assembly sizes ranged between 147,372 bp for *V. phaeantha* and 150,748 for *V. calyculata*, all of which had a typical structure of one Large Single Copy (LSC) region, two inverted repeats (IRs) and one Short Single Copy (SSC) region. The SSC is exceptionally short in Vanilloideae, and this was observed for the four assemblies, with sizes ranging from 1,962 bp in *V. phaeantha* to 2,154 bp in *V. chamissonis*. Manual inspection of annotated regions and a phylogenomic comparison of chloroplast genomes of 11 Vanilla species will be performed, including the new assemblies described in this work and those already available in public databases. Future efforts will also include the identification of insertion/deletion (InDel) regions between species for the design and test of potential PCR-based markers for species differentiation.

Apoio: FAPDF, Embrapa

Palavras-chave: Phylogenomics. Plant Genetic Resources.