Deciphering the molecular mechanisms of apomixis in polyploid *Paspalum* species with a SNP-based approach

Fernanda Ancelmo de Oliveira [1]; Alexandre Hild Aono [2]; Bianca Baccili Zanotto Vigna [3]; Alessandra Pereira Fávero [3]; Frederico de Pina Matta [3]; Anete Pereira de Souza [1,2]
[1] Plant Biology Department, Biology Institute, University of Campinas (UNICAMP), Campinas, SP, Brazil.; [2] Center for Molecular Biology and Genetic Engineering, UNICAMP, Campinas, SP, Brazil; [3] Brazilian Agricultural Research Corporation (Embrapa) Southeast Livestock, São Carlos, SP, Brazil

ABSTRACT

Paspalum is an important genus of tropical forage grass, predominantly composed of polyploid and apomictic (asexual reproduction mode via seeds) species with a great genetic variability. This genus represents a valuable resource to be explored not only in the agricultural sector, but also as a biological model for scientific research. Despite its unquestionable importance, studies at the molecular level are still limited and hindered by the mixed-polyploidy species, the absence of an appropriate reference genome, and the very recent breeding program, which makes the development of SNP markers for *Paspalum* a great scientific challenge. This work aims to study the allelic variation of SNPs by the estimation of allele dosages for several polyploid species of *Paspalum* with the development of a diverse panel of molecular markers associated with apomixis. For such a purpose, we analyzed 28 samples from sexual and apomictic plants. Two independent genotyping-by-sequencing (GBS) libraries were built, with two distinct restriction enzyme combinations i) *Pst*I + *Msp*I and ii) *Nsi*I + *Mse*I. DNA fragments were ligated to the common and barcode adapters, and each library was independently sequenced as 150-bp single-end reads using the Mid Output Kit v2.5 for the NGS NextSeq 500 platform. Raw sequencing data were evaluated and filtered using FASTQC and FASTX-Toolkit respectively. Using the BWA-MEM software, GBS reads were aligned against different pseudo-references obtained in Phytozome v.13 (i) *Panicum hallii*; (ii) *Panicum virgatum*; (iii) *Setaria italica*; and (iv) *Setaria viridis*; and with the assembly of two RNA-Seq experiments, the transcriptome of *Urochloa humidicola* and *Paspalum notatum*. SNP calling was performed with (I) Tassel-GBS pipeline for polyploids; (II) SAMtools and (III) FreeBayes, being retained the biallelic SNPs with a maximum of 25% missing data. The final set of SNPs was established through intersection of the results from all the tools and minimum locus depth of 20 reads for sample. The two sequencing experiments were combined into a final set of 208,774,226 reads (80.5%) of excellent quality. Alignments with the *Paspalum* transcriptome resulted in a larger set of aligned sequences (30.08%), followed by *P. virgatum* (22.81%). The alignment of the reads with the references used was low, which was expected because of the different grass genera. However, we could establish a final set with 30,134 SNPs. The potential of these SNPs to be true markers is high, as they were independently detected by different tools. Based on a preliminary analysis with apomixis related sequences described in the literature, we found a set of 615 sequences with several SNPs markers, which can be potentially related to different modes of reproduction. The results generated and the molecular markers developed have a great potential to benefit the use of this valuable germplasm and the genetic breeding programs of *Paspalum*.

KEYWORDS: GBS; forage grass; apomixis