Apomixis-related genes identified from a coexpression network in *Paspalum notatum*, a Neotropical grass

Fernanda Ancelmo de Oliveira[1], Bianca Baccili Zanotto Vigna[2], Carla Cristina da Silva[1], Alessandra Pereira Fávero[2], Frederico de Pina Matta[2], Ana Luisa Sousa Azevedo[3], Anete Pereira de Souza[1,4,§]

[1]Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brazil

[2]Embrapa Pecuária Sudeste, São Carlos, SP, Brazil

[3]Embrapa Gado de Leite, Juiz de Fora, MG, Brazil

[4]Departmento de Biologia Vegetal, Instituto de Biologia, UNICAMP, Campinas, SP, Brazil

[§]Corresponding author

E-mail: anete@unicamp.br

Telephone number: +55 19 3521-1132

Total Word Count:

6,440

Word Count for Introduction:

1,028

Word Count for Materials and Methods

2,123

Word Count for Results

1,695

Word Count for Discussion

1,495

Word Count for Acknowledgements

99

Figures Count

5

Tables Count

3

Summary

- Apomixis is a highly desirable trait in modern agriculture; however, incorporating it into breeding programs requires a deeper comprehension of apomictic regulatory mechanisms. *Paspalum notatum* is considered a good model for such studies because it exhibits both sexual and apomictic cytotypes. A transcriptomic approach to identifying differentially expressed genes between these cytotypes would constitute an important tool for the identification of genes in the apomictic pathway.
- We generated leaf and inflorescence transcriptomes of apomictic tetraploids and sexual diploids/tetraploids of *P. notatum* accessions. We *de novo* assembled and annotated 114,306 unigenes. Coexpression network and ontological annotation were applied to recover the interactions and biological processes of exclusive and differentially expressed genes from the apomictic libraries. Additionally, genes potentially linked to the apomixis-controlling region and several genes reported in the literature were identified.
- The reference transcriptome obtained in this study represents a robust set of expression data for *P. notatum*. Our approach to sequencing florets and leaves from different cytotypes enabled us to isolate a set of candidate genes. Using a gene expression network, we recovered transcriptionally coordinated genes related to mechanisms regulating apomixis.
- These findings provide an important contribution to our knowledge of apomixis and are valuable for *Paspalum* breeding programs.

Keywords

Apospory, Bahiagrass, Forage grass, Plant breeding programs, Plant reproduction, RNA-seq

Introduction

Apomixis is a mode of asexual reproduction through seeds in which plants produce offspring that are genetically identical to the female parent (Asker & Jerling, 1992). In apomixis, a non-reduced cell undergoes developmental pathways different from those of sexual cells. The multiple trajectories are classified as sporophytic or gametophytic according to the developmental origin of the cell from which the embryo is derived (Nogler, 1984; Koltunow, 1993; Conner *et al*., 2008). In sporophytes, the embryo develops from a somatic cell of the ovule through numerous mitotic divisions. In gametophytic apomixis, a chromosomally unreduced embryo sac can be formed either from the megaspore mother cell (diplospory) or from a nearby nucellar cell (apospory) without fertilization (parthenogenesis) through a process termed apomeiosis. Endosperm formation is required to produce a viable seed (Conner *et al.*, 2013, 2015). Apomixis is widely distributed among angiosperms, among which Poaceae represents the group with the largest number of apomictic genera, with the 47 apomictic species of *Paspalum* standing out in particular (Carman, 1997; Ortiz *et al*., 2013; Hojsgaard *et al.*, 2014).

The polyploid nature of all apomicts provides a challenge for genetic and genomic analysis (Ozias-Akins, 2006). Polyploidization is known to cause immediate and extensive genomic changes, including sequence rearrangements and/or elimination, changes in DNA methylation, recombination between homologous chromosomes, and the loss of balance in gene expression (Adams & Wendel, 2005; Cervigni *et al.*, 2008). Apomixis is frequently correlated with polyploidy and might have arisen through deregulation of the sexual developmental pathway due to the increase in number of genomic complements by a mechanism that could contain both genetic and epigenetic components (Spillane *et al.*, 2004; Polegri *et al.*, 2010; Calderini *et al.*, 2011; Selva *et al.*, 2012).

The genus *Paspalum* exhibits numerous characteristics that make it both unique and a very interesting system for the study of apomixis (Ortiz *et al.*, 2013), especially *Paspalum notatum* Flüggé, also known as bahiagrass. This species is considered an agamic complex that includes different ploidy levels and reproductive modes in which diploid cytotypes (2n=2×=20) are sexual and self-incompatible, while polyploids (3×, 4×, 5×, and 6×) are self-compatible pseudogamous apomicts (Burton, 1948). No 4× sexual cytotypes have been found in nature, but they have been obtained artificially (Martínez *et al*., 2001; Quarin *et al*., 2003). The inheritance of apomixis in *Paspalum* is controlled by a single complex dominant locus that confers apospory, i.e., epigenetically controlled parthenogenesis (Podio *et al.*, 2014a), with the capacity to form endosperm with a maternal excess genome contribution ratio of 4

(maternal):1 (paternal) (Ortiz *et al.*, 2013). The apomixis-controlling region (ACR) is small compared with other apomictic systems (Ortiz *et al.*, 2013), does not demonstrate recombination and conserves a relatively narrow region that is linked to apomixis among the different species (Pupilli *et al.*, 2004; Hojsgaard *et al.*, 2011). Comparative mapping of the ACR has shown synteny with a portion of the long arm of chromosome 12 of rice in *P. simplex*, *P. malacophyllum* and *P. procurrens* (Pupilli *et al.*, 2001, 2004; Hojsgaard *et al.*, 2011), while this region in *P. notatum* shows synteny with regions of rice chromosomes 2 and 12, suggesting chromosomal rearrangements (Martínez *et al.*, 2003; Pupilli *et al.*, 2004; Stein *et al.*, 2007; Podio *et al.*, 2012).

Apomixis presents potential significance for agriculture, allowing the maintenance of heterosis in hybrid progeny; thus, this mode of reproduction has become the subject of exhaustive cytoembryological, cytogenetic and molecular analyses (Ozias-Akins, 2006; Stein *et al.*, 2007). The discovery of key genes associated with the components of apomixis and the isolation of some sequence candidates has led to the understanding of some of the functions and processes involved in apomixis (Pessino *et al.*, 2001; Rodrigues *et al.*, 2003; Albertini *et al.*, 2004; Calderini *et al.*, 2006; Cervigni *et al.*, 2008; Laspina *et al.*, 2008; Sharbel *et al.*, 2009; Yamada-Akiyama *et al.*, 2009; Polegri *et al.*, 2010; Sharbel *et al.*, 2010; Felitti *et al.*, 2011; Okada *et al.*, 2013; Siena *et al.*, 2014; Podio *et al.*, 2014b; Conner *et al.*, 2015; Siena *et al.*, 2016; Ortiz *et al.*, 2017; Bocchini *et al.*, 2018). However, insights into the complete genetic mechanisms underlying asexual reproduction in natural apomicts are needed to verify whether candidate genes isolated from several species can be applied to develop a stable and universal apomixis system (Siena *et al*., 2016). In turn, given the complexity of this trait, an understanding of the genomic structure of the apomictic locus is likely to be an essential prerequisite for manipulation of the sexual pathway in model plants or economically important crops (Calderini *et al*., 2011). However, the suppression of recombination events around the ACR limits direct genetic strategies to isolate the mechanism triggering apospory (Laspina *et al*., 2008). In this context, reverse genetics approaches have garnered extreme interest in the study of apomixis both to identify and validate genes that are differentially expressed in apomicts and to conduct more meticulous investigations aiming to detect their positions and functions (Laspina *et al.*, 2008; Ortiz *et al*., 2017).

RNA sequencing (RNA-seq) is the most effective method of predicting new transcripts and differentially expressed genes, providing enhanced sensitivity for identifying gene expression patterns in different tissues or at different developmental stages (Huang *et al.*, 2017; Liang *et al.*, 2017). In contrast, considering the large amount of data generated from

RNA-seq, new approaches that efficiently extract meaningful associations from highly multivariate datasets are needed (Miao *et al.*, 2016). The construction of coexpression networks from gene expression data using pairwise correlation metrics provides us with valuable information on alterations in biological systems in response to differential gene expression patterns (de la Fuente, 2010; Gov & Arga, 2016; Miao *et al.*, 2016).

Therefore, the objective of this study was to use RNA-seq technology to comprehensively analyze the gene expression profiles of the leaves and florets of apomictic and sexual genotypes with different ploidy levels in *P. notatum*. Our study revealed several differentially expressed genes among the analyzed genotypes, and coexpression network analysis and Gene Ontology (GO) annotation allowed us to investigate the main biological processes of candidate genes potentially linked to the ACR, which may be used to further explore and clarify the mechanisms regulating apomixis in this important forage species.

Materials and methods

Plant materials and RNA extraction

The six accessions of *P. notatum* used in this study belong to the Germplasm Bank of *Paspalum*, maintained by Embrapa Pecuária Sudeste, São Carlos, SP, Brazil (22° 01'S and 47°54' W; 856 m above sea level). The accessions were chosen by considering their origins (different ecotypes), genetic dissimilarity (Cidade *et al.*, 2013) and ploidy levels and reproduction modes (Table 1). Tetraploid *P. notatum* is apomictic, while the diploid is sexual. The Q3664 (BGP 216) accession is a hybrid from a cross between a sexual colchicine-induced tetraploid (PT-2) and a white-stigma apomictic tetraploid of *P. notatum* (Forbes & Burton, 1961). The Q3664 is characterized as a facultative apomictic accession with a high level of sexuality (> 70%) (Quarin *et al.*, 1984).

For this study, five clones of each accession were planted in pots of eight liters each with a 1:1 soil/vermiculite mix and maintained in a greenhouse under the same conditions. The climate is humid subtropical (Koppen), with annual average low and high temperatures of 15.3 and 27.0°C, respectively, and total rainfall of 1,422.8 mm, occurring mainly during the spring and summer seasons (CEPAGRI, 2010).

In the summer season in Brazil (December), leaf samples from three biological replicates of each accession were collected. One clone each of the accessions BGP34 (apomictic tetraploid) and BGP22 (sexual diploid) presented just-emerged inflorescences, which were also collected. All leaf and floret samples were immediately placed in liquid

nitrogen and subsequently stored at -80°C until RNA extraction. Total RNA was isolated according to Oliveira *et al.* (2015). RNA integrity was assessed in a 1% agarose denaturing gel and quantified using a NanoVue Plus spectrophotometer (GE Healthcare Life Sciences, Little Chalfont, UK).

Cytoembryological analysis

The mode of reproduction of the plants was confirmed using the embryo sac clarification method proposed by Young *et al.* (1979), with minor modifications. Inflorescences at anthesis were collected and fixed in FAA (95% ethanol, glacial acetic acid, distilled water, formalin 40%, 40:13:3:3 v/v) for 24 h. Afterward, the FAA was replaced with 70% ethanol and the samples were stored at 4°C. The embryo sacs were extracted by dissection under a stereoscopic microscope and stored in 70% ethanol. The embryo sacs were clarified by replacing 70% alcohol with the following series of solutions: 85% ethanol; absolute ethanol; ethanol:methyl salicylate (1:1); ethanol:methyl salicylate (1:3); 100% methyl salicylate (two times). The samples remained in each solution for 24 h. Observations were carried out with an Axiophot microscope (Carl Zeiss, Jena, Germany) using the differential interference contrast (DIC) microscopy technique. A total of 100 embryo sacs per accession were evaluated.

Flow cytometry

Flow cytometry analyses were performed to confirm the exact ploidy level of each experimental plant and compare it to the literature data on chromosome counts. Approximately 5 mg of young leaf tissue from each plant was used (Table 1). The same amount of leaf tissue was used for *Pisum sativum* cv. Ctirad (2C=9.09 pg), which served as an internal standard (Dolezel *et al.*, 1992; Greilhuber *et al.*, 2007). The samples were triturated in a petri dish containing 800 µL of LB01 buffer (0.45425 g TRIS, 0.186125 g NaEDTA, 0.0435 g_spermine, 0.29225 g NaCl, 1.491 g KCl, and 250 µL of Triton in 250 mL of distilled water, 7.5 pH, 0.11% v/v of β-mercaptoethanol) to obtain a nuclear suspension (Dolezel, 1997). The nuclear suspension was filtered through a mesh of 40 µm and incubated at room temperature for 5 min, after which 25 µL of propidium iodide and 25 µL of RNase were added. For each sample, at least 10,000 nuclei were analyzed. The material was analyzed with a FACSCalibur flow cytometer (Becton Dickinson, New Jersey, USA), and histograms were obtained in Cell Quest software and analyzed in Flowing Software (available at http://www.flowingsoftware.com). Generally, only histograms with both peaks (sample and

standard) of approximately the same height were considered. The nuclear DNA index (pg) of the plants was estimated based on the value of the G1 peak as an internal reference, and then the mean value of C was calculated by biological replicate and by accession. In addition, the difference (delta) between the highest and the lowest values found in the replicates of an accession was calculated.

RNA-seq library construction, Illumina sequencing, and data filtering

The cDNA libraries were constructed separately for each sample (18 leaf and two floret libraries) using a TruSeq RNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instructions. Library quality was confirmed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and quantified via quantitative real-time PCR (qRT-PCR) (Illumina protocol SY-930-10-10). Clustering was conducted using a TruSeq PE Cluster Kit on a cBot (Illumina Inc., San Diego, CA, USA). Paired-end sequencing was performed on an Illumina Genome Analyzer IIx with TruSeq SBS 36-Cycle kits (Illumina, San Diego, CA, USA) according to the manufacturer's specifications for 72-bp PE reads. RNA-seq was performed with the floret libraries of the accessions BGP22 (diploid/sexual) and BGP34 (tetraploid/apomictic) in two separate lanes, without biological replicates. The 18 leaf libraries from six different accessions, each with three independent biological replicates, were distributed randomly on the flow cell, with three libraries per lane.

The raw sequence data acquired from each of the libraries were stored in FastQ files containing 72-bp reads. Filtering for high-quality (HQ) reads was performed using the NGS QC Toolkit v2.3.3 (Patel & Jain, 2012). Initially, low-quality reads (Phred quality score < 20 in 75% of all reads) and reads with fewer than 60 bases were removed. Subsequently, the remaining reads were trimmed at the 3' end, and bases with Phred scores < 20 were removed. All reads were deposited in the NCBI Short Read Archive (SRA) under accession number SRP150615.

Transcriptome assembly

A *de novo* assembly of the filtered sequences was carried out using the Trinity program v2.0.2 (Grabherr *et al.*, 2011) using default settings to obtain a representative dataset for the *P. notatum* transcriptome from florets and leaves. After assembly, the unigenes were filtered using criteria to retain the first Butterfly transcript generated per Chrysalis component, which is considered representative. To assess the integrity of the assembled transcriptomes, the HQ short reads were mapped back into the transcript dataset using the Bowtie2 sequence

aligner v2.2.5 (Langmead *et al.*, 2009). Benchmarking Universal Single-Copy Orthologs (BUSCO), a recognized benchmark approach to single-copy orthologs providing an assessment of orthologs conserved among species (Simão *et al.*, 2015), was used to estimate the completeness of our transcriptome assembly based on the presence, absence, duplication, and/or fragmentation of conserved plant genes.

Gene annotation and analysis

All unigenes were aligned against the NCBI non-redundant protein (Nr) database UniProtKB/Swiss-Prot and the grass protein database from Phytozome v9.0, using the BLASTX algorithm with an e-value cutoff of 1e-06. GO (Ashburner *et al.*, 2000) terms were retrieved for the sequences that were functionally annotated in the NCBI Nr database using Blast2GO software (Conesa *et al.*, 2005). The GO terms were mapped to each annotated transcript based on the hits with highest sequence similarities, and the genes were distributed among distinct functional categories. We used REVIGO (Supek *et al.*, 2011) to summarize and visualize the GO terms. Pathways were assigned according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa & Goto, 2000).

Expression level estimates

The expression levels of the unigenes were calculated via the FPKM method (expected number of fragments per kilobase of transcript sequence per million base pairs sequenced) using Bowtie v2-2.1.0 by RSEM software (RNA-seq by expectation maximization) (Li & Dewey, 2011). Using the assembled unigenes as a reference transcriptome, the clean raw reads from each library were separately realigned to the assembled transcriptome. A Venn diagram of the genes was created with the online platform available at http://bioinformatics.psb.ugent.be/webtools/Venn/. Principal component analysis (PCA) was used to identify the expression patterns among the genotypes in RNA-seq to highlight their similarities and differences.

Differential expression analysis

EBSeq (Leng *et al.*, 2013) was used to identify differentially expressed genes (DEGs) in the counts matrix, and a false discovery rate (FDR) ≤ 0.05 was applied to extract all genes that were significantly expressed. Transcripts with a log2 fold change (FC) ≥ 1.5 in transcript abundance were regarded as overexpressed. GO enrichment analysis was carried out using R software (version 3.3.1) and the package 'goseq' in version 1.24.0 of Bioconductor (Young *et*

al., 2010). The p-values were subjected to Bonferroni correction, considering a corrected p-value ≤ 0.05 as enriched. The results were visualized with REVIGO (Supek *et al.*, 2011).

The accessions were grouped according to their biological conditions as sexual diploids (2×SEX), sexual tetraploids (4×SEX) and apomictic tetraploids (4×APO). The DEGs were identified in two stages. In the first step, the objective was to select a list of equally expressed genes among the different genotypes that shared the same biological condition. Thus, EBSeq was used to estimate the probabilities of gene expression between the pairwise genotypes, sexual diploids (accessions: BGP22 and BGP306), and apomictic tetraploids (accessions: BGP30, BGP34 and BGP115). The genes that presented PPEE ≥ 0.05 (posterior probability gene/transcript is equally expressed) were maintained for the second stage of the analysis, which consisted of DEG analysis between two biological groups, such as a) 2×SEX vs 4×APO; b) 2×SEX vs 4×SEX; and c) 4×APO vs 4×SEX. For these comparisons, only first-stage genes that had detectable expression in the three biological groups were used.

Quantitative real-time PCR (qRT-PCR) validation

To verify the reliability and accuracy of our transcriptome data, 18 DEGs were randomly selected. *P. notatum* has no reference genes described in the literature; thus, eight unigenes that displayed similar expression patterns based on FPKM were selected for evaluation as reference genes. Primer sets were designed using Primer3Plus software (Untergasser *et al.*, 2007). All the primer pairs were initially evaluated by PCR using genomic DNA as a template from the same genotypes used in RNA-seq. Only those that amplified the DNA of all genotypes and showed an amplification efficiency of 90-110% (R^2> 0.99) in qRT-PCR were used for expression analysis.

Total RNA (500 ng) was used for cDNA synthesis using the QuantiTect Reverse Transcription Kit (Qiagen Inc., Chatsworth, CA, USA). qRT-PCR was performed using a CFX384 Real-Time PCR Detection System with iTaq Universal SYBR Green Supermix (Bio-Rad Laboratories Inc., Hercules, CA, USA) according to the manufacturer's instructions. The reaction conditions were 95°C for 10 min and then 40 cycles of 95°C for 30 s and 60°C for 1 min. All experiments were performed in triplicate, and no-template controls were included. The presence of single amplicons in the PCR products was confirmed by a melting curve analysis. The baseline and Cq values were automatically determined, and the expression analysis (ΔΔCt method) was performed using CFX Manager 2.1 software (Bio-Rad Laboratories, Inc., USA). Reference genes were selected according to the gene expression stability values (M < 0.5) and coefficients of variance (CV < 0.25) among the samples. A

Mann-Whitney U-test was performed to estimate statistical significance.

Search for potential apomixis genes

To perform a detailed comparative analysis of *P. notatum* unigenes from this study with their syntenic counterparts on rice chromosomes that are conservatively linked to apomixis in *Paspalum* species (Ortiz *et al*., 2013), we aligned our unigenes against the rice transcripts with the BLASTN tool and isolated those that presented homology onto genes in the highlighted area of rice chromosomes 2 and 12. Then, we selected only the *P. notatum* unigenes contained in the ACR, which were delimited by a set of molecular markers that were completely linked to the apospory locus in this species. Thus, the region encompassed the C1069 and C996 markers for rice chromosome 12 and between C560 and C932 markers for chromosome 2 (Labombarda *et al.*, 2002; Martínez *et al.*, 2003; Pupilli *et al.*, 2004; Stein *et al.*, 2007; Podio *et al.*, 2012; Ortiz *et al.*, 2013). We also compared our transcriptome with sequences reported for the genus *Paspalum*, potentially associated with apomixis. The sequences were used as queries for nucleotide BLAST search against the 114,306 unigenes, with an e-value cutoff of 1e-05.

Construction of coexpression networks

To create the Pearson-based coexpression networks, we used the expression values in FPKM of assembled unigenes from floret and leaf tissues. Genes showing null values for most of the replicates were excluded to reduce noise and eliminate residuals in the analysis. We then calculated an all-versus-all coexpression network matrix using a Pearson correlation coefficient cutoff of ≥ 0.8. The highest reciprocal rank (HRR) method proposed by Mutwil *et al.* (2010) was used to select only the strongest edges, considering an HRR limit of 30. The Heuristic Cluster Chiseling Algorithm (HCCA) was applied to partition the networks into manageable clusters, with three-step node vicinity networks (NVN) (Mutwil *et al*., 2010). The interactive coexpression network was visualized using Cytoscape 3.4.0 software (Shannon, 2003).

Results

DNA content and mode of reproduction

The cytoembryological analysis confirmed the expected mode of reproduction of the accessions based on the literature (Tables 1 and 2). The 2C DNA content of all plants from

accessions BGP_22 and BGP_306 was half the 2C DNA content from all plants from BGP_30, BGP_34, BGP_115, and BGP_216. Therefore, all evaluated plants had DNA contents compatible with the described ploidy levels of their respective accessions in the literature (Tables 1 and 2).

RNA-seq analysis and de novo assembly

After removing the adaptors and the low-quality and contaminated reads, 313,198,097 HQ clean paired-ends reads from floret and leaf tissues were obtained and used to construct a *P. notatum* reference transcriptome. A total of 203,808 transcripts were assembled, of which a set of 114,306 non-redundant sequences were filtered as unigenes (56.08% of all transcripts) (Table 3). The unigenes had an average length of 750.51 bp with an N50 of 906 bp and GC percentage of 47.60%. According to the length distribution of unigenes, we obtained 21,585 (18.88%) unigenes that were longer than 1 kb, a size range that commonly confers a high annotation rate. Finally, more than half of the total annotated unigenes were > 500 bp in length (Fig. S1).

The Bowtie aligner mapped 94.39% of the reads onto the assembled transcripts. When compared within the selected unigenes, 73.78% of the reads were remapped. Of these, 82.97% and 83.48% of the sequencing reads from sexual and apomictic samples, respectively, were similarly represented in the final assembly. The BUSCO analysis included 956 conserved single-copy plant orthologs; the *P. notatum* transcriptome showed high assembly completeness, with 664 (69%) complete sequences (532 as a single copy and 132 as multiple copies) and 169 (17%) gene fragments. One hundred and twenty-three (12%) BUSCO plant genes were not identified among the unigenes of *P. notatum*.

Gene function analysis

Unigene annotation was performed using the BLAST algorithm against several databases. Among the 114,306 unigenes, 51,939 (45.44%) were categorized in the NCBI Nr database, 34,750 (30.40%) were categorized in UniProtKB/Swiss-Prot, and 54,568 (47.74%) were categorized in Phytozome.

The GO database was then utilized to retrieve the GO identifiers for each unigene annotated in NCBI Nr. Overall, 21,537 (41.47%) unigenes were assigned to 4,614 GO terms in three main categories: 2,510 biological process, 1,486 molecular function, and 618 different cellular components. The results of the KEGG analysis showed that the 9,221 annotated unigenes participated in 129 pathways. Among these, Metabolic Pathway was the

most represented, with an emphasis on Purine Metabolism Pathway (1,079 members), followed by Metabolism of Thiamine (907 members) and Starch and Sucrose Metabolism (336 members). Overall, 33,873 unigenes remained without hits after searching against all protein databases. We then performed a search for the presence of open reading frames (ORFs), which were found in 15,107 unigenes, 2,347 of which showed complete ORFs. These sequences require further investigation since they may represent genes that have not yet been described and possibly new genes that may be unique to *P. notatum.*

Expression levels of genes and identification of DEGs

The HQ clean reads were realigned against the 114,306 unigenes of the reference transcriptome to estimate the expression level of the unigenes in each genotype. Most identified genes had expression levels of FPKM ≤ 100 (110,621 unigenes), whereas only 49 unigenes had high expression levels (FPKM ≥ 400). The remaining unigenes were not mapped and thus, their expression levels could not be determined. A comparison of the 110,656 unigenes that showed different levels of expression in the two tissues of the studied samples revealed genes that were expressed in the genotypes of only one biological group and that were not expressed in the others (Fig. 1), which were considered "unique" genes. Thus, 19,304 unique unigenes were found for the 4×APO group, 2,173 for the 2×SEX group, and 217 for the 4×SEX group. All the unigenes expressed in the floret tissues were also contained in the transcriptome profile of leaf; no unique genes were found in the floret samples for either diploid sexual or tetraploid apomictic samples (Fig. 1).

To compare the differences among all three phenotypic groups, we chose to use only the unigenes with FPKM values from leaf samples because they contained all the unigenes representative of the transcriptome assembly and had three biological replicates for each genotype, maximizing the accuracy and reliability of the differential gene expression analysis. However, within the same biological group, different expression patterns were observed among genotypes, mainly in group 4×APO (Fig. S2). Thus, to reduce the effects of the differences in expression caused by genotype, we performed a two-step analysis. In the first stage, we identified 72,318 unigenes that were equally expressed among all genotypes of the apomictic tetraploid group and 47,069 equally expressed unigenes in the sexual diploid group. This initial analysis was not necessary for the sexual tetraploid because only one genotype was analyzed. From these, we selected 28,969 unigenes that were common to all three biological groups.

Pairwise comparisons of 28,969 unigenes allowed the identification of 2,072 DEGs

between 2×SEX and 4×APO, which included 772 and 1,173 overexpressed unigenes, respectively. There were 1,173 DEGs between 2×SEX and 4×SEX, including 308 and 798 overexpressed unigenes, respectively, and 1,317 genes were identified as DEGs between 4×APO and 4×SEX, 618 and 661 of which were overexpressed, respectively. Next, we inferred the enriched functions for the identified DEGs that were over- or underexpressed (Fig. S5, S6 and S7). Among the identified DEGs, we isolated a set of 510 unigenes that were consistently differentially expressed between the apomictic and sexual groups independent of ploidy level (diploid or tetraploid) (Fig. 2). These unigenes correspond to potential genes related to the reproductive modes, sexual and apomictic, found in this species. We identified the enriched GO terms of these unigenes, which belonged to 26 biological processes (BP), 13 molecular functions (MF), and three cellular components (CC).

Among the BP enriched in this set, we can emphasize "positive regulation of miRNA metabolic process (GO:2000630)" and "regulation of primary miRNA processing (GO:2000636)", both of which either activate or increase the frequency, rate, or extent of miRNA production. miRNA, in turn, is directly related to "DNA metabolic process (GO:0006259)", which decreases the rate of gene expression (negative regulation of gene expression - GO:0010629) through an epigenetic RNA-based gene silencing process (RNA-directed DNA methylation - GO:0080188). Five unigenes are mainly involved in these processes and overexpressed in apomicts in relation to their sexual counterparts; these unigenes correspond to a pseudoARR-B transcription factor; pre-mRNA splicing factor, putative (*Ricinus communis*); hypothetical protein SORBIDRAFT_05g016770; uncharacterized protein, LOC100501330 (*Zea mays*); and an uncharacterized protein LOC103641690.

qRT-PCR validation

We confirmed a subset of the DEGs between the apomictic and the sexual groups using qRT-PCR. Of the genes selected for testing as a reference, five were successfully amplified and matched the parameters (Table S1). We used one of these primer pairs as a reference for qRT-PCR for the groups 2×SEX x 4×APO (01RefGen-*Pnot*) and 4×APO x 4×SEX (04RefGen-*Pnot*).

Among the 18 primer pairs, two were not evaluated for qRT-PCR but showed interesting results in terms of DNA amplification. The first was designed from a unigene that was unique to the apomicts according to the RNA-seq results (i.e., showing a zero expression value in the sexual (diploid/tetraploid) genotypes and amplifying apomictic DNA). For the

second, although DNA amplification occurred in all groups, when cDNA gel amplification was observed, only the apomictic genotypes presented amplicons. The expression level determined through RNA-seq analysis of the same genotypes was very low in 2×SEX. These results suggest that these unigenes, which are expressed only in certain cytotypes, may be exclusive to the genomes of these groups and otherwise poorly expressed or silenced. Eight primer pairs were then used for qRT-PCR (Table S2) for validation (Fig. S3 and S4). The qRT-PCR analyses revealed consistent expression trends with those identified through RNA-seq.

Detailed search for potential apomixis genes

Overall, 1,612 unigenes that were potentially integrated into the ACR were revealed. Of these, 1,356 unigenes corresponded to rice chromosome 2 and 256 corresponded to the chromosome 12 region. Curiously, 40 of these unigenes were part of the set identified as unique to apomicts, 9 were unique to sexual diploids, and 6 were unique to sexual tetraploids. Moreover, considering the analysis of differential expression, 16 unigenes were differentially expressed in 2×SEX vs 4×APO, 20 in 4×APO vs 4×SEX, and 17 in 2×SEX vs 4×SEX. All these findings, especially this set of 108 unique or differentially expressed unigenes, represent a valuable set of genes that can be carefully investigated to determine whether they are effectively genetically linked to the ACR in the species.

The BLAST results in Table S10 show the similarity of the unigenes to *Paspalum* sequences from the literature, revealing that our transcriptome contains genes already found to be involved in asexual reproductive development in *Paspalum.*

Network gene expression

We obtained 879,481 coexpressed connections among 53,262 unigenes (Fig. 3). The genes were grouped into 642 clusters according to coexpression pattern. In the network, the unique unigenes from each biological group tended to form coexpression modules (Fig. 3b-d). Despite this division, the relevant biological relationships of these genes with the other genes can be recovered to integrate the network.

We identified a direct correlation of the 108 DE and/or unique putative ACR unigenes with their nearest neighbors in the gene expression network. Thus, we retrieved a small network composed of 536 strongly correlated unigenes (Fig. 4), allowing the identification of GO enriched terms for these genes. We observed the participation of these unigenes in 43 BP, 31 MF, and 17 CC. The enriched BP terms (Fig. 5) demonstrate that the overall gene network

is transcriptionally directed toward reproductive BP, such as “male gamete generation (GO: 0048232)”; “sister chromatid segregation (GO: 0000819)”; “meiosis I (GO: 0007127)”; “resolution of meiotic recombination intermediates (GO: 0000712)”; “regulation of cytokinesis (GO: 0032465)”; “mitotic cytokinesis (GO: 0000281)”; “spindle assembly (GO: 0051225)”; “mitotic spindle organization (GO: 0007052)”; and “mitotic DNA integrity checkpoint (GO: 0044774)”. At the same time, the network links enriched epigenetic processes such as “macromolecule methylation (GO: 0043414)”; “histone-serine phosphorylation (GO: 0035404)”; “regulation of protein phosphatase type 2A activity (GO: 0034047)”; and “negative regulation of MAP kinase activity (GO: 0043407)”.

Discussion

We constructed a *P. notatum* reference transcriptome using next-generation sequencing approaches to investigate the gene expression changes associated with apomictic and sexual reproduction. RNA-seq analysis identified 114,306 unigenes based on two tissues (florets and leaves) from six genotypes with different ploidy levels and reproductive modes. Robust metrics indicated that the transcriptome was a quality assembly with a high degree of integrity, which further expands the bahiagrass transcriptome database.

GO analysis showed that the unigenes could be grouped into 4,614 known terms, which were used herein as the main biological GO classifications and ensured the integrity of the functional analyses of the apomixis candidate genes. Various unique GO terms were enriched at different time points. Based on 508 DEGs between apomictic and sexual plants, the analysis of enriched GO terms suggests that DEGs between plants with distinct reproductive modes are related to the regulation of gene expression by epigenetic silencing in apomictic bahiagrass. This result is consistent with the growing body of evidence that suggests that apomixis arises from deregulation of the sexual pathway, where epigenetic mechanisms play a significant role in at least some elements of apomictic development (Calderini *et al.*, 2011; Ortiz *et al.*, 2013; Brukhin, 2017). The differential expression analysis strategy carried out herein, aiming both to remove genes with expression that may be related to the effect of genotype and to isolate a set of genes that are commonly expressed by all genotypes presenting the same biological condition, allowed us to identify DEGs that are representative of the mode of reproduction and free of much of the influence of ploidy and genotype. These DEGs represent a strong set of candidate genes that, along with the 19,304 unique unigenes for apomixis, certainly merit further investigation.

In apospory, gene expression occurs at specific stages of apomictic development such

as apomeiosis, parthenogenesis, and autonomous endosperm development; increased or decreased expression of some genes due to these specific stages may hinder the analysis of differential expression between sexual and apomictic genotypes (Ortiz *et al.*, 2013, 2017). However, we believe that the identification of DEGs as performed herein, using only genes with expression among all genotypes of all biological groups and including our other modifications, may have minimized this influence. In addition, this method may explain the large number of identified "unique" genes. Gene expression observed exclusively in apomicts, for example, may be the result of the occurrence of heterochromatic expression or expression at particular stages of development or in specific cells.

In addition, since functionally related genes tend to be transcriptionally coordinated (Stuart *et al.*, 2003; Persson *et al.*, 2005), the construction of the gene expression network performed here provides a powerful resource, allowing us to identify genes that are coexpressed with the unique and DE genes and thus recover BP and closely related pathways. The 536-unigene network presented in this work is an example of how we can recover the relevant biological relationships of genes of interest using the whole transcriptome. By starting on a smaller scale with the DEGs and/or unique genes and integrating the information by adding their nearest neighbors, we can obtain a broader view of the processes related to the candidate genes. The genes in the smallest network are transcriptionally directed toward reproductive processes and the regulation of epigenetic changes by modulating histones. In particular, MAP kinase activity is negatively regulated. The mitogen-activated protein 3-kinase (MAP3K) gene family has been identified as differentially expressed in the apomictic and sexual flowers of *P. notatum* (Laspina *et al*., 2008) and in the flowers of *P. simplex* (Polegri *et al*., 2010), where it might have a role in parthenogenetic development of the embryo in both species (Ortiz *et al*., 2013).

The DEGs and unique unigenes used to form the smallest network are those that are possibly integrated into the ACR. The size of the ACR in *Paspalum* is relatively small compared with other apomictic systems (Ortiz *et al*., 2013), but the lack of recombination makes it difficult to study this region by map-based cloning (Laspina *et al*., 2008). Thus, the discovery of DEGs that are potentially located in this region is extremely valuable to our attempts to understand the complex functional network of gene–gene communication. These DEGs could also be used for future manipulation of the apomixis trait, which has outstanding importance in agricultural biotechnology. Increases in cattle production in tropical, subtropical, and temperate areas have been achieved due to the use of apomictic forage grass cultivars (Ortiz *et al*., 2013).

Apomixis is simultaneously intriguing and complex to elucidate, providing an inspiring area of research. In the genus *Paspalum*, the first approach to understanding the molecular nature of apomixis was applied by Pessino *et al*. (2001). This approach led to the identification of a few sequences that are highly expressed during early megagametophyte development in apomictic plants through differential display experiments in the inflorescences of sexual and aposporous *P. notatum*. Here, we recovered all three of those reported sequences as a single representative unigene (Table S10) annotated as kinesin motor expressed, involved in the biological process of microtubule-based movement. A kinesin motor protein has also been reported by other authors as differentially expressed between apomictic and sexual plants (Albertini *et al*., 2004, 2005; Laspina *et al*., 2008). Interestingly, our unigene was also identified as differentially expressed in the transcriptome data.

Since then, several other candidate genes have been reported using *P. notatum* and *P. simplex* species (Pessino *et al.*, 2001; Calderini *et al.*, 2006; Laspina *et al*., 2008; Polegri *et al*., 2010) and more recently, the first approach using RNA-seq was published for bahiagrass based on expression profiling of apomictic and sexual plant flowers (Ortiz *et al*., 2017). Some of these candidate genes have been investigated in more detail (Felitti *et al.*, 2011; Siena *et al.*, 2014, 2016; Podio *et al.*, 2014b). However, many questions remain, and further research is needed to define the relationships between the structure, the position, and the function of the known apomixis-linked genes (Ortiz *et al*., 2013). Furthermore, we recovered some of the previously published sequences in our *P. notatum* transcriptome.

One of these candidate genes is n20gap-1, a lorelei-like *P. notatum* gene (Laspina *et al*., 2008; Felitti *et al*., 2011) encoding a GPI-anchored protein that supposedly plays a role in the final stages of the apomixis developmental cascade. Laspina *et al*. (2008) previously reported this sequence as linked to the chromosomal locus governing apospory at a genetic distance of 22 cM. Here, one of these unigenes was in the identified set of unigenes potentially integrated into the ACR, corresponding to rice chromosome 2. Additionally, we found some unigenes, including some that were differentially expressed, that aligned with the *P. notatum* sequences characterized by Podio *et al*. (2014b), corresponding to *Somatic Embryogenesis Receptor-Like Kinase* (*SERK*), in addition to other numerous unigenes with the same annotation. The candidate SERK-like genes play crucial roles in somatic embryogenesis in angiosperms and have been reported as related to apomixis. Albertini *et al*. (2005) analyzed two members, *PpSERK1* and *PpSERK2*, and found that *PpSERK1* expression was high during premeiosis and decreased during the meiosis and post-meiotic stages, whereas *PpSERK2* expression was high from premeiosis to anthesis in the nucellar cells of

aposporous genotypes. The authors proposed that the expression pattern of *PpSERK* in *Poa pratensis* was compatible with its role in the specification of aposporous initials. In *P. notatum*, the expression of two different members of the *SERK* family (*PnSERK1* and *PnSERK2*) was observed, and *PnSERK2* displayed a spatial expression pattern similar to that reported for the *PpSERKs*, which are expressed in nucellar cells at meiosis in the apomictic genotype (Podio *et al*., 2014b). We also identified unigenes that showed similarity to other interesting candidate genes related to apospory, such as the PnTgs1-like gene that encodes a trimethylguanosine synthase-like protein, which plays a fundamental role in nucellar cell fate, as its diminished expression is correlated with initiation of the apomictic pathway in plants (Siena *et al*., 2014). Additionally, two unigenes were similar to the sequence of *Ps*ORC3, which seems to play an active role in mechanisms repressing sexuality in apomictic *P. simplex*, acting in the development of apomictic seeds, which deviate from the canonical 2 (maternal):1 (paternal) genome ratio (Siena *et al*., 2016). The recovery of these candidate gene sequences demonstrates the representative and informative nature of our transcriptome. This outcome reaffirms that high-throughput sequencing technology enhances our understanding of global RNA expression.

In this study, we applied RNA-seq technology and bioinformatics methods to identify the transcriptomic differences between apomictic and sexual reproduction. Our results reveal many DEGs, i.e., genes exclusively expressed in apomicts or sexuals, many of which are in potential association with the genomic region, including the apo locus. This set of genes may aid in the identification or validation of candidate genes, which will enable valuable insights through understanding apospory. Moreover, these genes may be considered to screen for molecular markers linked to apomixis in *P. notatum*, which will be crucial to boost breeding programs for apomictic forage grasses.

Acknowledgements

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2008/52197-4) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Computational Biology Program). FAO received a PhD fellowship from FAPESP (2013/20447-0), the CAPES-Computational Biology Program, and the CAPES-EMBRAPA Program; CCS received a PhD fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and FAPESP (2015/24346-9); APS is the recipient of a research fellowship from CNPq. We are grateful to Maria Augusta C. Horta and Natalia F. Murad for their help in generating the coexpression

networks.

Author Contributions

A.P.S., B.B.Z.V. and F.A.O. conceived and designed the experiments; F.A.O. performed the experiments, A.P.S, B.B.Z.V and F.A.O analyzed the data and wrote the manuscript; F.A.O and C.C.S. performed the qPCR analyses; A.P.F. prepared and provided the plant material; F.P.M. performed the cytoembryological analyses; A.L.S.A. obtained and analyzed the data for the flow cytometry analyses; and A.P.S. and B.B.Z.V coordinated the work. All authors critically revised and approved the manuscript.

References

Adamowski EDV, Pagliarini MS, Bonato ABM, Batista LAR, Valls JFM. 2005. Chromosome numbers and meiotic behavior of some *Paspalum* accessions. *Genetics and Molecular Biology* **28**: 773-780.

Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* **8**: 135-141.

Albertini E, Marconi G, Barcaccia G, Raggi L, Falcinelli M. 2004. Isolation of candidate genes for apomixis in *Poa pratensis* L. *Plant Molecular Biology* **56**: 879-894.

Albertini E, Marconi G, Reale L, Barcaccia G, Porceddu A, Ferranti F, Falcinelli M. 2005. *SERK* and *APOSTART*. Candidate genes for apomixis in *Poa pratensis*. *Plant Physiology* **138**: 2185-2199.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. 2000. Gene ontology: tool for the unification of biology. *Nature Genetics* **25**: 25-29.

Asker S, Jerling L. 1992. *Apomixis in plants*. Boca Raton, FL: CRC Press.

Bocchini M, Galla G, Pupilli F, Bellucci M, Barcaccia G, Ortiz JPA, Pessino SC, Albertini E. 2018. The vesicle trafficking regulator PN_SCD1 is demethylated and overexpressed in florets of apomictic *Paspalum notatum* genotypes. *Scientific Reports* **8**: 3030.

Brukhin V. 2017. Molecular and genetic regulation of apomixis. *Russian Journal of Genetics* **53**: 943-964.

Burton GW. 1948. The method of reproduction in common bahia grass, *Paspalum notatum*. *Journal of the American Society of Agronomy* **40**: 443-452.

Calderini O, Chang SB, Jong H, Busti A, Paolocci F, Arcioni S, de **Vries SC, Abma-Henkens MHC, Lankhorst RMK, Donnison IS et al. 2006.** Molecular cytogenetics and DNA sequence analysis of an apomixis-linked BAC in *Paspalum simplex* reveal a non pericentromere location and partial microcolinearity with rice. *Theoretical and Applied Genetics* **112**: 1179-91.

Calderini O, Donnison I, Polegri L, Panara F, Thomas A, Arcioni S, Pupilli F. 2011. Partial isolation of the genomic region linked with apomixis in *Paspalum simplex*. *Molecular Breeding* **28**: 265-276.

Carman JG. 1997. Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. *Biological Journal of the Linnean*

Society **61**: 51-94.

CEPAGRI. 2010. *Centro de pesquisas meteorológicas e climáticas aplicadas à agricultura*. http://www.cpa.unicamp.br/o-cepagri.html. [accessed 17 August 2016].

Cervigni GDL, Paniego N, Pessino S, Selva JP, Díaz M, Spangenberg G, Echenique V. 2008. Gene expression in diplosporous and sexual *Eragrostis curvula* genotypes with differing ploidy levels. *Plant Molecular Biology* **67**: 11-23.

Cidade FW, Vigna BBZ, de Souza FHD, Valls JFM, Dall'Agnol M, Zucchi MI, de Souza-Chies TT, Souza AP. 2013. Genetic variation in polyploid forage grass: assessing the molecular genetic variability in the *Paspalum* genus. *BMC Genetics* **14**: 50.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674-3676.

Conner JA, Goel S, Gunawan G, Cordonnier-Pratt MM, Johnson VE, Liang C, Wang H, Pratt LH, Mullet JE, DeBarry J et al. 2008. Sequence analysis of bacterial artificial chromosome clones from the apospory-specific genomic region of *Pennisetum* and *Cenchrus*. *Plant Physiology* **147**: 1396-1411.

Conner JA, Gunawan G, Ozias-Akins P. 2013. Recombination within the apospory specific genomic region leads to the uncoupling of apomixis components in *Cenchrus ciliaris*. *Planta* **238**: 51-63.

Conner JA, Mookkan M, Huo H, Chae K, Ozias-Akins P. 2015. A parthenogenesis gene of apomict origin elicits embryo formation from unfertilized eggs in a sexual plant. *Proceedings of the National Academy of Sciences U.S.A.* **112**: 11205-11210.

Dahmer N, Schifino-Wittmann MT, Dall'Agnol M, Castro BD. 2008. Cytogenetic data for *Paspalum notatum* Flügge accessions. *Scientia Agricola* **65**: 381-388.

de la Fuente A. 2010. From 'differential expression' to 'differential networking' – identification of dysfunctional regulatory networks in diseases. *Trends in Genetics* **26**: 326-333.

Dolezel J 1997. Flow cytometry, its application and potential for plant breeding. In: Lelley T ed. *Current topics in plant cytogenetics related to plant improvement*. Vienna: WUV-Universitätsverlag, 80-90.

Dolezel J, Sgorbati S, Lucretti S. 1992. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiologia Plantarum* **85**: 625-631.

Felitti SA, Seijo JG, González AM, Podio M, Laspina NV, Siena L, Ortiz JPA, Pessino SC. 2011. Expression of lorelei-like genes in aposporous and sexual *Paspalum notatum* plants. *Plant Molecular Biology* **77**: 337-354.

Forbes I, Burton GW. 1961. Cytology of diploids, natural and induced tetraploids, and intra-species hybrids of bahiagrass, *Paspalum notatum* Flugge. *Crop Science* **1**: 402-406.

Gov E, Arga KY. 2016. Interactive cooperation and hierarchical operation of microRNA and transcription factor crosstalk in human transcriptional regulatory network. *IET Systems Biology* **10**: 219-228.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**: 644-652.

Greilhuber J, Temsch EM, Loureiro JCM 2007. Nuclear DNA content measurements. In: Doležel J, Greilhuber J, Suda J eds. *Flow cytometry with plant cells: analysis of genes, chromosomes and genomes*. Weinheim, Germany: Wiley-VCH Verlag GmbH and Co., 67-101.

Hojsgaard D, Klatt S, Baier R, Carman JG, Hörandl E. 2014. Taxonomy and biogeography of apomixis in *Angiosperms* and associated biodiversity characteristics. *Critical Reviews in Plant Sciences* **33**: 414-427.

Hojsgaard DH, Martínez EJ, Acuña CA, Quarin CL, Pupilli F. 2011. A molecular map of the apomixis-control locus in *Paspalum procurrens* and its comparative analysis with other species of *Paspalum*. *Theoretical and Applied Genetics* **123**: 959-971.

Huang W, Guo Y, Du W, Zhang X, Li A, Miao X. 2017. Global transcriptome analysis identifies differentially expressed genes related to lipid metabolism in Wagyu and Holstein cattle. *Scientific Reports* **7**: 5278.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**: 27-30.

Koltunow AM. 1993. Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules. *The Plant Cell* **5**: 1425-1437.

Labombarda P, Busti A, Caceres ME, Pupilli F, Arcioni S. 2002. An AFLP marker tightly linked to apomixis reveals hemizygosity in a portion of the apomixis-controlling locus in *Paspalum simplex*. *Genome* **45**: 513-519.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**: R25.

Laspina NV, Vega T, Seijo JG, González AM, Martelotto LG, Stein J, Podio M, Ortiz

JPA, Echenique VC, Quarin CL et al. 2008. Gene expression analysis at the onset of aposporous apomixis in *Paspalum notatum*. *Plant Molecular Biology* **67**: 615-628.

Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, Haag JD, Gould MN, Stewart RM, Kendziorski C. 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* **29**: 1035-1043.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

Liang R, Han B, Li Q, Yuan Y, Li J, Sun D. 2017. Using RNA sequencing to identify putative competing endogenous RNAs (ceRNAs) potentially regulating fat metabolism in bovine liver. *Scientific Reports* **7**: 6396.

Martínez EJ, Hopp HE, Stein J, Ortiz JPA, Quarin CL. 2003. Genetic characterization of apospory in tetraploid *Paspalum notatum* based on the identification of linked molecular markers. *Molecular Breeding* **12**: 319-327.

Martínez EJ, Urbani MH, Quarin CL, Ortiz JPA. 2001. Inheritance of apospory in bahiagrass, *Paspalum notatum*. *Hereditas* **135**: 19-25.

Miao X, Luo Q, Zhao H, Qin X. 2016. Co-expression analysis and identification of fecundity-related long non-coding RNAs in sheep ovaries. *Scientific Reports* **6**: 39398.

Mutwil M, Usadel B, Schutte M, Loraine A, Ebenhoh O, Persson S. 2010. Assembly of an interactive correlation network for the *Arabidopsis* genome using a *Novel Heuristic Clustering Algo*rithm. *Plant Physiology* **152**: 29-43.

Nogler GA 1984. Gametophytic apomixis. In: Johri BM ed. *Embryology of angiosperms*. Berlin: Springer, 475-518.

Okada T, Hu Y, Tucker MR, Taylor JM, Johnson SD, Spriggs A, Tsuchiya T, Oelkers K, Rodrigues JCM, Koltunow AMG. 2013. Enlarging cells initiating apomixis in *Hieracium praealtum* transition to an embryo sac program prior to entering mitosis. *Plant Physiology* **163**: 216-231.

Oliveira RR, Viana AJC, Reátegui ACE, Vincentz MGA. 2015. Short communication an efficient method for simultaneous extraction of high-quality RNA and DNA from various plant tissues. *Genetics and Molecular Research* **14**: 18828-18838.

Ortiz JPA, Quarin CL, Pessino SC, Acuña C, Martínez EJ, Espinoza F, Hojsgaard DH, Sartor ME, Cáceres ME, Pupilli F. 2013. Harnessing apomictic reproduction in grasses: what we have learned from *Paspalum*. *Annals of Botany* **112**: 767-787.

Ortiz JPA, Revale S, Siena LA, Podio M, Delgado L, Stein J, Leblanc O, Pessino SC. 2017. A reference floral transcriptome of sexual and apomictic *Paspalum notatum*.

BMC Genomics **18**: 318.

Ozias-Akins P. 2006. Apomixis: developmental characteristics and genetics. *Critical Reviews in Plant Sciences* **25**: 199-214.

Pagliarini MS, Carraro LR, Freitas Pm, Adamowski EV, Batista LAR, Valls JFM. 2001. Cytogenetic characterization of Brazilian *Paspalum* accessions. *Hereditas* **135**: 27-34.

Pagliarini MS, Takayama SY, de Freitas PM, Carraro LR, Adamowski EV, Silva N, Batista LAR. 1999. Failure of cytokinesis and 2n gamete formation in Brazilian accessions of *Paspalum. Euphytica* **108**: 129-135.

Patel RK, Jain M. 2012. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* **7**: e30619.

Persson S, Wei H, Milne J, Page GP, Somerville CR. 2005. Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proceedings of the National Academy of Sciences U.S.A.* **102**: 8633-8638.

Pessino SC, Espinoza F, Martínez EJ, Ortiz JPA, Valle EM, Quarín CL. 2001. Isolation of cDNA clones differentially expressed in flowers of apomictic and sexual *Paspalum notatum*. *Hereditas* **134**: 35-42.

Podio M, Cáceres ME, Samoluk SS, Seijo JG, Pessino SC, Ortiz JPA, Pupilli F. 2014a. A methylation status analysis of the apomixis-specific region in *Paspalum* spp. suggests an epigenetic control of parthenogenesis. *Journal of Experimental Botany* **65**: 6411-6424.

Podio M, Felitti SA, Siena LA, Delgado L, Mancini M, Seijo JG, González AM, Pessino SC, Ortiz JPA. 2014b. Characterization and expression analysis of *SOMATIC EMBRYOGENESIS RECEPTOR KINASE (SERK)* genes in sexual and apomictic *Paspalum notatum. Plant Molecular Biology* **84**: 479-495.

Podio M, Rodríguez MP, Felitti S, Stein J, Martínez EJ, Siena LA, Quarin CL, Pessino SC, Ortiz JPA. 2012. Sequence characterization, in silico mapping and cytosine methylation analysis of markers linked to apospory in *Paspalum notatum. Genetics and Molecular Biology* **35**: 827-837.

Polegri L, Calderini O, Arcioni S, Pupilli F. 2010. Specific expression of apomixis-linked alleles revealed by comparative transcriptomic analysis of sexual and apomictic *Paspalum simplex* Morong flowers. *Journal of Experimental Botany* **61**: 1869-1883.

Pozzobon MT, Valls JFM. 1997. Chromosome number in germplasm accessions of *Paspalum notatum* (Gramineae). *Brazilian Journal of Genetics* **20**: 29-34.

Pupilli F, Labombarda P, Caceres ME, Quarín CL, Arcioni S. 2001. The chromosome

segment related to apomixis in *Paspalum simplex* is homoeologous to the telomeric region of the long arm of rice chromosome 12. *Molecular Breeding* **8**: 53-61.

Pupilli F, Martinez EJ, Busti A, Calderini O, Quarin CL, Arcioni S. 2004. Comparative mapping reveals partial conservation of synteny at the apomixis locus in *Paspalum spp*. *Molecular Genetics and Genomics* **270**: 539-548.

Quarin CL, Burson BL, Burton GW. 1984. Cytology of intra- and interspecific hybrids between two cytotypes of *Paspalum notatum* and P. cromyorrhizon. *Botanical Gazette* **145**: 420-426.

Quarin CL, Urbani MH, Blount AR, Martinez EJ, Hack CM, Burton GW, Quesenberry KH. 2003. Registration of Q4188 and Q4205, sexual tetraploid germplasm lines of bahiagrass. *Crop Science* **43**: 745-746.

Rodrigues JCM, Cabral GB, Dusi DMA, de Mello LV, Rigden DJ, Carneiro VTC. 2003. Identification of differentially expressed cDNA sequences in ovaries of sexual and apomictic plants of *Brachiaria brizantha*. *Plant Molecular Biology* **53**: 745-757.

Selva JP, Pessino SC, Meier MS, Echenique VC. 2012. Identification of candidate genes related to polyploidy and/or apomixis in *Eragrostis curvula*. *American Journal of Plant Sciences* **3**: 403-416.

Shannon P. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**: 2498-2504.

Sharbel TF, Voigt ML, Corral JM, Galla G, Kumlehn J, Klukas C, Schreiber F, Vogel H, Rotter B. 2010. Apomictic and sexual ovules of *Boechera* display heterochronic global gene expression patterns. *The Plant Cell* **22**: 655-671.

Sharbel TF, Voigt ML, Corral JM, Thiel T, Varshney A, Kumlehn J, Vogel H, Rotter B. 2009. Molecular signatures of apomictic and sexual ovules in the *Boechera holboellii* complex. *Plant J* **58**: 870-882.

Siena LA, Ortiz JPA, Calderini O, Paolocci F, Cáceres ME, Kaushal P, Grisan S, Pessino SC, Pupilli F. 2016. An apomixis-linked ORC3-like pseudogene is associated with silencing of its functional homolog in apomictic *Paspalum simplex*. *Journal of Experimental Botany* **67**: 1965-1978.

Siena LA, Ortiz JPA, Leblanc O, Pessino S. 2014. PNTGS1-like expression during reproductive development supports a role for RNA methyltransferases in the aposporous pathway. *BMC Plant Biology* **14**: 297.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.

Bioinformatics **31**: 3210-3212.

Spillane C, Curtis MD, Grossniklaus U. 2004. Apomixis technology development—virgin births in farmers' fields? *Nature Biotechnology* **22**: 687-691.

Stein J, Pessino SC, Martínez EJ, Rodriguez MP, Siena LA, Quarin CL, Ortiz JPA. 2007. A genetic map of tetraploid *Paspalum notatum* Flügge (bahiagrass) based on single-dose molecular markers. *Molecular Breeding* **20**: 153-166.

Stuart JM, Segal E, Koller D, Kim SK. 2003. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**: 249-255.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**: e21800.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. 2007. Primer3plus, an enhanced web interface to Primer3. *Nucleic Acids Research* **35**: W71-W74.

Yamada-Akiyama H, Akiyama Y, Ebina M, Xu Q, Tsuruta SI, Yazaki J, Kishimoto N, Kikuchi S, Takahara M, Takamizo T et al. 2009. Analysis of expressed sequence tags in apomictic guineagrass (*Panicum* maximum). *Journal of Plant Physiology* **166**: 750-761.

Young BA, Sherwood RT, Bashaw EC. 1979. Cleared-pistil and thick-sectioning techniques for detecting aposporous apomixis in grasses. *Canadian Journal of Botany* **57**: 1668-1672.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**: R14.

Figure Legends

Fig. 1 Venn diagram showing the distribution of unigenes of *P. notatum* from two tissues (floret and leaf) and three biological groups (sexual diploid, sexual tetraploid, and apomictic tetraploid).

Fig. 2 Enriched ontology terms of 510 unigenes differentially expressed between apomictic and sexual groups in the *Paspalum notatum* transcriptomes.

Fig. 3 Coexpression network for *Paspalum notatum*. (a) Coexpression network of all analyzed genes; the more centralized genes in light gray are those common to all genotypes analyzed with non-differential expression. Node color denotes the differentially expressed genes between the biological groups: purple (2×SEX vs 4×APO), yellow (2×SEX vs 4×SEX), and dark green (4×APO vs 4×SEX). The networks of unique genes for each group are highlighted: (b) unigenes unique to sexual diploids (blue); (c) unigenes unique to sexual tetraploids (green); and (d) unigenes unique to apomictic tetraploids (red). Edges denote interaction strength. Circular nodes represent genes.

Fig. 4 Gene expression network for 536 *Paspalum notatum* unigenes. The 108 differentially expressed and/or unique putative ACR unigenes are presented in this network, along with their nearest neighbors. The color patterns are the same as those used in the complete coexpression network in Fig. 3.

Fig. 5 Enriched biological processes in the 536 unigenes in the gene expression network.

Tables

Table 1. *Paspalum notatum* accessions evaluated in this study.

BGP CODE	FORMER CODE	REPRODUCTION MODE	ORIGIN	COLLECTOR*	CHROMOSOME NUMBER
306	BRA-024236	Sexual	Candói (PR), Brazil	VLiGu 14829	2n=2×=20 (Dahmer *et al.*, 2008)
22	BRA-006173	Sexual	Bagé (RS), Brazil	VGnMaBd 9607	2n=2×=20 (Pagliarini *et al.*, 1999; Pagliarini *et al.*, 2001)
216	BRA-019470	Sexual**	Corrientes, Argentina	Q 3664	2n=4×=40 (Forbes & Burton, 1961)
30	BRA-006467	Apomictic	Alegrete (RS), Brazil	VMrFrLw 9747	2n=4×=40 (Pozzobon & Valls, 1997)
34	BRA-006513	Apomictic	Uruguaiana (RS), Brazil	VMrFrLw9782	2n=4×=40 (Pozzobon & Valls, 1997; Adamowski *et al.*, 2005)
115	BRA-010006	Apomictic	Laguna (SC), Brazil	VDBdSv 10137	2n=4×=40 (Pozzobon & Valls, 1997)

* Collectors: Bd, I. I. Boldrini; D, M. Dall'Agnol; Fr, J. M. O. Freitas; Gn, J. O. N. Gonçalves; Gu, A. Guglieri; Li, L. Essi; Lw, H. M. Longhi-Wagner; Ma, M. C. Assis; Mr, C. O. C. Moraes; Q, C. L. Quarín; Sv, G. P. da Silva; V, J. F. M. Valls; LAT, latitude (decimal degrees); LONG, longitude (decimal degrees). ** Facultative apomictic with a high level of sexuality (> 70%), according to Quarín *et al*. (1984).

Table 2. DNA contents and modes of reproduction of the bahiagrass accessions evaluated in this study.

Accession (BGP code)	DNA content				Mode of reproduction
	Plants	2C (pg)	Δ2C among plants	Accession mean 2C (pg)	
30	BGP30_1[1]	3.2	0.4	3.20	facultative apomixis
	BGP30_2[1]	3.1			
	BGP30_3	2.9			
	BGP30_4	2.8			
	BGP30_5[1,3]	3.0			
34	BGP34_1[1,2,3]	2.9	0.5	2.80	facultative apomixis
	BGP34_2	2.9			
	BGP34_3[1]	2.5			
	BGP34_4	2.7			
	BGP34_5[1]	3.0			
115	BGP115_1	2.7	0.3	2.68	obligate apomixis
	BGP115_2[1,3]	2.9			
	BGP115_3[1]	2.6			
	BGP115_4	2.6			
	BGP115_5[1]	2.7			
216	BGP216_1[1,3]	2.6	0.6	2.82	sexual
	BGP216_2[1]	2.7			
	BGP216_3	2.9			
	BGP216_4	3.2			
	BGP216_5[1]	2.7			
22	BGP22_1	1.5	0.3	1.52	sexual
	BGP22_2	1.7			
	BGP22_3[1,2,3]	1.4			
	BGP22_4[1]	1.4			
	BGP22_5[1]	1.5			
306	BGP306_1[1,3]	1.5	0.1	1.44	sexual
	BGP306_2	1.5			
	BGP306_3	1.4			
	BGP306_4[1]	1.4			
	BGP306_5[1]	1.4			

[1]Indicates plants with foliar RNA evaluated; [2]Indicates plants with inflorescence RNA evaluated; [3]Indicates plants with DNA evaluated.

Table 3. Statistics for the assembled transcriptome of *Paspalum notatum*.

	Transcripts	**Unigenes**
Total	203,808	114,306
Total assembled bases	219,683,637	85,787,466
Average length (bp)	1,077.90	750.51
N50 length (bp)	1,599	906
Percent GC (%)	46.76	47.60

Supporting Information

Fig. S1 Comparison of the number of unigenes and the number of unigenes annotated in the NCBI non-redundant protein database by size range.

Fig. S2 Principal component analysis (PCA) according to the FPKM values of *P. notatum* leaf and floret transcriptomes for all genotypes and biological replicates used in RNA-seq.

Fig. S3. Histograms of gene expression obtained by qRT-PCR showing the relative expression patterns of 4 genes that are differentially expressed between tetraploid apomictic "4×APO" (red box) and diploid sexual "2×SEX" (blue box). *P < 0.05; **P < 0.01; ***P < 0.001: statistically significant differences in gene expression between biological groups compared using the Mann-Whitney U-test.

Fig. S4. Histograms of gene expression obtained by qRT-PCR showing the relative expression patterns of 4 genes that are differentially expressed between tetraploid apomictic "4×APO" (red box) and tetraploid sexual "4×SEX" (green box). *P < 0.05; **P < 0.01; ***P < 0.001: statistically significant differences in gene expression between the biological groups compared using the Mann-Whitney U-test.

Fig. S5. Functional classification of enriched overexpressed DEGs in the 2×SEX vs the 4×APO comparison. Gene Ontology biological process (blue boxes), Gene Ontology cellular component (yellow boxes), and Gene Ontology molecular function (orange boxes). a) Categories enriched in overexpressed genes in 2×SEX and b) categories enriched in overexpressed genes in 4×APO.

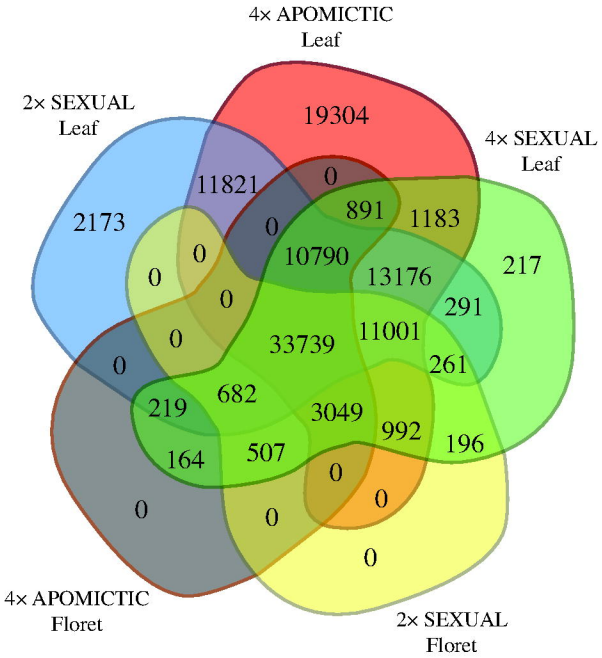
Fig. S6. Functional classification of enriched overexpressed DEGs in the 2×SEX vs 4×SEX comparison. Gene Ontology biological process (blue boxes), Gene Ontology cellular component (yellow boxes), and Gene Ontology molecular function (orange boxes). a) Categories enriched in overexpressed genes in 2×SEX and b) categories enriched in overexpressed genes in 4×SEX.

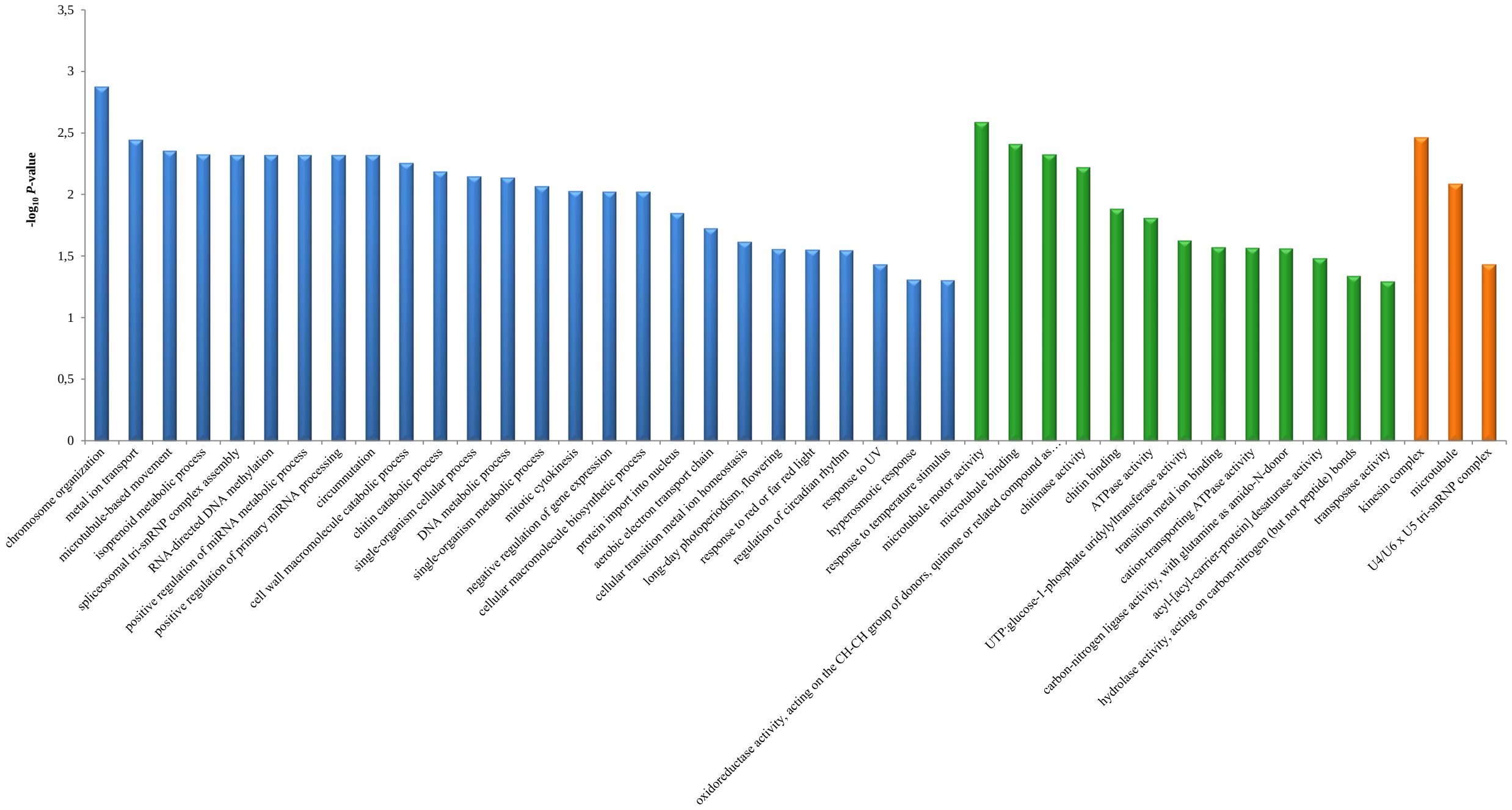
Fig. S7. Functional classification of enriched overexpressed DEGs in the 4×APO vs 4×SEX comparison. Gene Ontology biological process (blue boxes), Gene Ontology cellular component (yellow boxes), and Gene Ontology molecular function (orange boxes). a) categories enriched in overexpressed genes in 4×APO and b) categories enriched in overexpressed genes in 4×SEX.

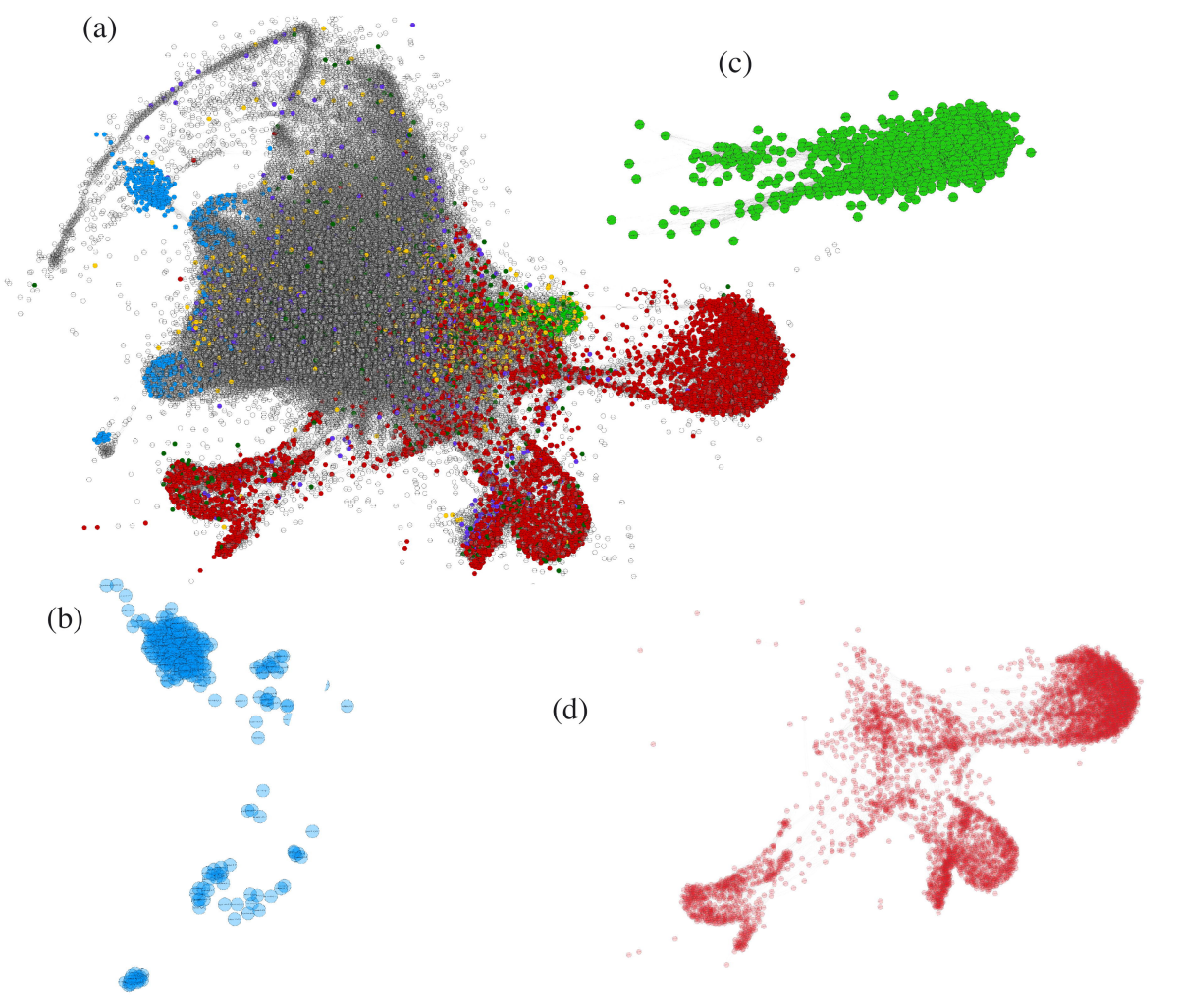
Table S1 Primer sequences and amplicons of the candidate reference genes evaluated in this study.

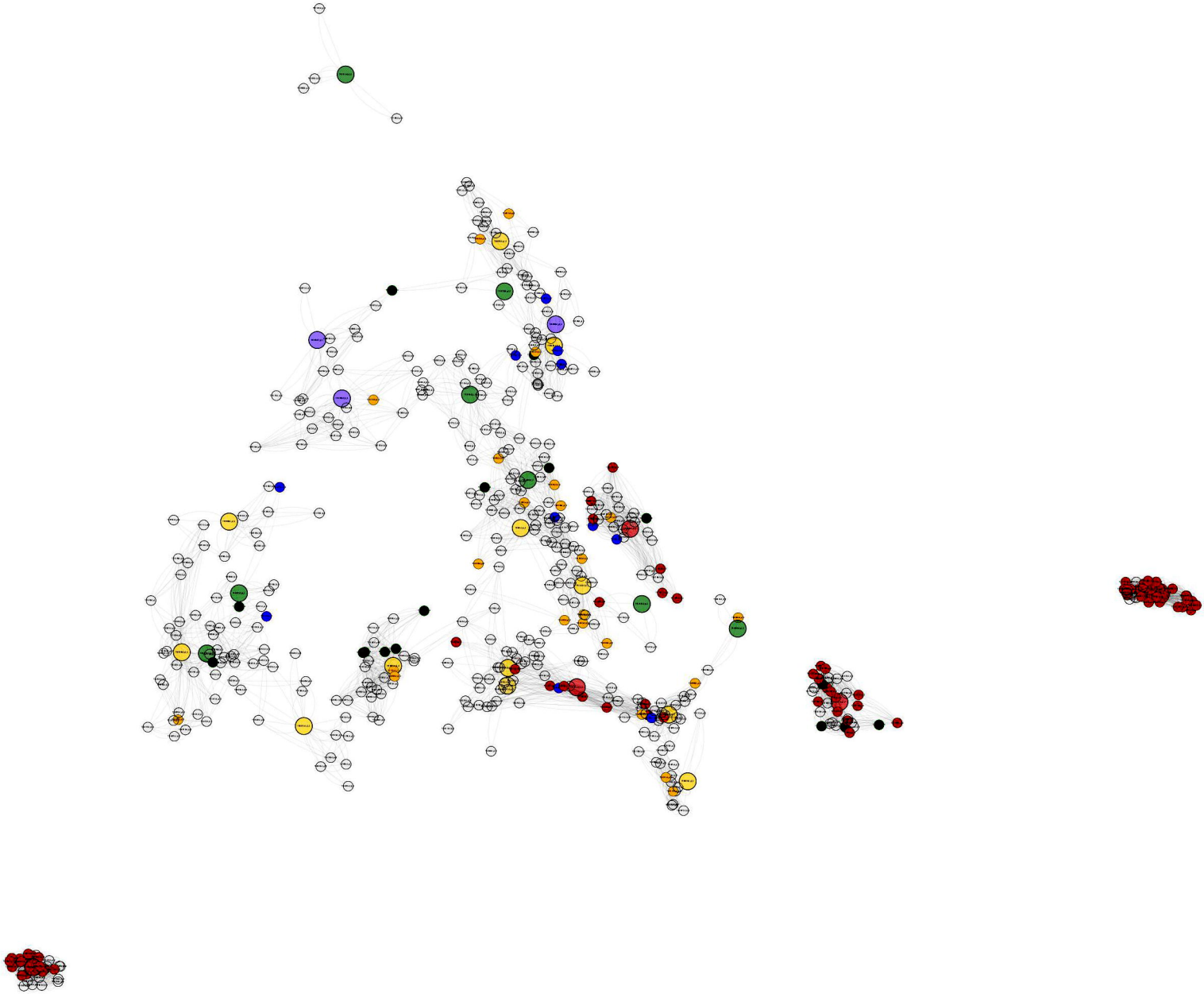
Table S2 Quantitative RT-PCR primer sequences.

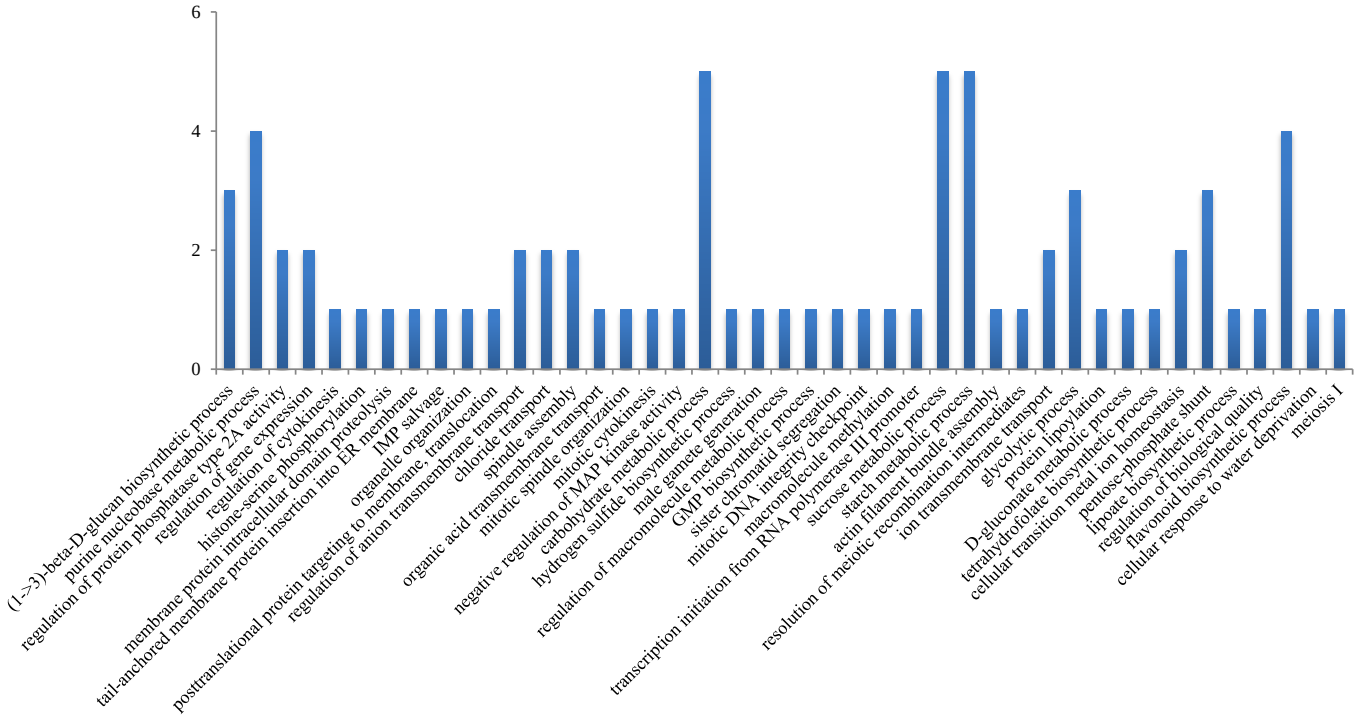
Table S3 BLAST search results for sequences of *Paspalum notatum* from the literature against the assembled unigenes.


4× APOMICTIC
Leaf
2× SEXUAL
Leaf
4× SEXUAL
Leaf
4× APOMICTIC
Floret
2× SEXUAL
Floret
19304
11821
0
2173
0
891
1183
0
10790
217
0
13176
0
291
0
33739
11001
0
261
682
219
3049
992
196
164
507
0
0
0
0
0


-log10 P-value
3,5
3
2,5
2
1,5
1
0,5
0
chromosome organization
metal ion transport
microtubule-based movement
isoprenoid metabolic process
spliceosomal tri-snRNP complex assembly
RNA-directed DNA methylation
positive regulation of miRNA metabolic process
positive regulation of primary miRNA processing
circumnutation
cell wall macromolecule catabolic process
chitin catabolic process
single-organism cellular process
DNA metabolic process
single-organism metabolic process
mitotic cytokinesis
negative regulation of gene expression
cellular macromolecule biosynthetic process
protein import into nucleus
aerobic electron transport chain
cellular transition metal ion homeostasis
long-day photoperiodism, flowering
response to red or far red light
regulation of circadian rhythm
response to UV
hyperosmotic response
response to temperature stimulus
microtubule motor activity
microtubule binding
oxidoreductase activity, acting on the CH-CH group of donors, quinone or related compound as…
chitinase activity
chitin binding
ATPase activity
UTP:glucose-1-phosphate uridylyltransferase activity
transition metal ion binding
cation-transporting ATPase activity
carbon-nitrogen ligase activity, with glutamine as amido-N-donor
acyl-[acyl-carrier-protein] desaturase activity
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds
transposase activity
kinesin complex
microtubule
U4/U6 x U5 tri-snRNP complex


(a)
(c)
(b)
(d)




6
4
2
0
(1->3)-beta-D-glucan biosynthetic process
purine nucleobase metabolic process
regulation of protein phosphatase type 2A activity
regulation of gene expression
regulation of cytokinesis
histone-serine phosphorylation
membrane protein intracellular domain proteolysis
tail-anchored membrane protein insertion into ER membrane
IMP salvage
organelle organization
posttranslational protein targeting to membrane, translocation
regulation of anion transmembrane transport
chloride transport
spindle assembly
organic acid transmembrane transport
mitotic spindle organization
negative regulation of MAP kinase activity
mitotic cytokinesis
carbohydrate metabolic process
hydrogen sulfide biosynthetic process
regulation of macromolecule metabolic process
male gamete generation
GMP biosynthetic process
sister chromatid segregation
mitotic DNA integrity checkpoint
transcription initiation from RNA polymerase III promoter
macromolecule methylation
sucrose metabolic process
resolution of meiotic recombination intermediates
starch metabolic process
actin filament bundle assembly
ion transmembrane transport
glycolytic process
protein lipoylation
D-gluconate metabolic process
tetrahydrofolate biosynthetic process
cellular transition metal ion homeostasis
pentose-phosphate shunt
lipoate biosynthetic process
regulation of biological quality
flavonoid biosynthetic process
cellular response to water deprivation
meiosis I