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

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

Fernanda A. de Oliveira[1], Bianca B. Z. Vigna[2], Carla C. da Silva[1], Alessandra P. Fávero[2], Frederico de P. Matta[2], Ana L. S. Azevedo[3], Anete P. 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

E-mail addresses:

f.ancelmo.o@gmail.com, bianca.vigna@embrapa.br, silvacbio@yahoo.com.br, alessandra.favero@embrapa.br, frederico.matta@embrapa.br, ana.azevedo@embrapa.br, anete@unicamp.br

[*]**Corresponding author**

E-mail: anete@unicamp.br (AP)

Telephone number: +55 19 3521-1132

Abstract

Apomixis is a highly desirable trait in modern agriculture, due to the maintenance of characteristics of the mother plant in the progeny. However, incorporating it into breeding programs requires a deeper knowledge of its regulatory mechanisms. *Paspalum notatum* is considered a good model for such studies because it exhibits both sexual and apomictic cytotypes, facilitating the performance of comparative approaches. Therefore, we used comparative transcriptomics between contrasting *P. notatum* cytotypes to identify novel candidate genes involved in the regulation of the expression of this phenotype. We assembled and characterized a transcriptome from leaf and inflorescence from apomictic tetraploids and sexual diploids/tetraploids of *P. notatum* accessions, and then assembled a coexpression network based on pairwise correlation between transcripts expression profiles. We identified genes exclusively expressed in each cytotype and differentially expressed genes between pairs of cytotypes. Gene ontology enrichment analyses were performed for the interpretation of data. We *de novo* assembled 114,306 of reference transcripts. 536 novel candidate genes for the control of apomixis were detected through statistical analyses of expression data, contains in this set, the interactions among genes potentially linked to the apomixis-controlling region, differentially expressed, several genes also already reported in the literature and their neighbors transcriptionally related in the coexpression network. The reference transcriptome obtained in this study represents a robust set of expression data for *P. notatum*. Additionally, novel candidate genes identified in this work represent a valuable resource for future grass breeding programs.

Author Summary

Clonal mode of reproduction by seeds is termed apomixis, which results from the failure of gamete formation (meiosis) and fertilization in the sexual female reproductive pathway. The manipulation of seeds production genetically identical to the mother plant bears great promise for agricultural applications, however clarification regarding gene interactions involved in reproductive process is needed. *Paspalum* is considered a model genus for the analysis of apomixis mechanisms. Here, we describe an overall analysis of the expression profiles of *Paspalum notatum* transcripts in response to changes in reproductive mode (sexual to apomictic), which allowed us to identify several candidate apomixis genes. Among these, we found genes potentially associated with the apomixis control region, in addition to genes already described in the literature for *Paspalum*, which highlights the representativeness of assembled transcriptome. For the first time in the literature, we explored the main biological processes involved in controlling the expression of apomictic reproduction based on co-regulatory networks of candidate apomixis genes.

Introduction

Apomixis is a mode of asexual reproduction through seeds in which the plants produced offspring is genetically identical to the female parent [1]. 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 [2-4]. 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 [5,6]. Apomixis is widely distributed among angiosperms, among which Poaceae represents the family with the largest number of apomictic genera, with the 47 apomictic species of *Paspalum* standing out in particular [7-9].

The polyploid nature of all apomicts provides a challenge for genetic and genomic analysis [10]. Polyploidization is known to cause immediate and extensive genomic changes, including sequence rearrangements and/or elimination, changes in DNA methylation, and the loss of balance in gene expression [11,12]. 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, through a mechanism regulated by both genetic and epigenetic components [13-16].

84 *Paspalum notatum* Flüggé, also known as bahiagrass, belongs to the genus *Paspalum*,
85 which exhibits numerous characteristics that make it a very interesting system for the study of
86 apomixis [8]. This species is considered an agamic complex that includes different ploidy levels
87 and reproductive modes in which diploid cytotypes (2n=2×=20) are sexual and self-
88 incompatible, whereas polyploids (3×, 4×, 5×, and 6×) are self-compatible pseudogamous
89 apomicts [17]. No 4× sexual cytotypes have been found in nature, but they have been obtained
90 artificially [18,19]. The inheritance of apomixis in *Paspalum* is controlled by a single complex
91 dominant locus that confers apospory, i.e., epigenetically controlled parthenogenesis [20], with
92 the capacity to form endosperm with a maternal excess genome contribution ratio of 4:1
93 (maternal:paternal) [8]. The apomixis-controlling region (ACR) is small compared to other
94 apomictic systems [8], does not demonstrate recombination and conserves a relatively narrow
95 region that is linked to apomixis among the distinct species [21,22]. Comparative mapping of the
96 ACR has shown synteny with a portion of the long arm of chromosome 12 of rice in *P. simplex*,
97 *P. malacophyllum* and *P. procurrens* [21-23]. In *P. notatum* the ACR shows synteny with
98 regions of rice chromosomes 2 and 12, suggesting the presence of chromosomal rearrangements
99 [21,24-26].

100 Apomixis presents potential significance for agriculture, allowing the maintenance of
101 heterosis in hybrid progeny. Therefore, this mode of reproduction has become the subject of
102 exhaustive cytoembryological, cytogenetic and molecular analyses [10,25]. Some key genes
103 associated with the components of apomixis and the isolation of some sequence candidates have

already been discovered [6,12,14,27-41]. However, insights into the genetic mechanisms underlying asexual reproduction in natural apomicts are still needed to develop a stable and universal apomixis system to be applied in breeding programs [39]. 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 [15]. However, the suppression of recombination events around the ACR limits forward genetic strategies to isolate the mechanism triggering apomixis by map-based [31]. In this context, identify and validate genes that are differentially expressed in apomicts and to conduct more meticulous investigations aiming to detect their effects in phenotype, have garnered extreme interest in the study of apomixis [31,40].

RNA sequencing (RNA-seq) is the most effective method for simultaneously predicting new transcripts and identifying differentially expressed genes in distinct tissues, genotypes, abiotic conditions or developmental stages [42,43]. Conversely, considering the large amount of data generated from RNA-seq, new approaches that efficiently extract meaningful associations from highly multivariate datasets are needed [44]. 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 [44-46].

The objective of this study was to identify candidate genes possibly involved in the regulation of the expression of apomixis in *P. notatum*. To reach this goal, we used RNA-seq

124 technology to obtain a reference transcriptome and comprehensively analyze gene expression
125 profiles in leaves and florets from apomictic and sexual genotypes with different ploidy levels in
126 *P. notatum*. Our study revealed several differentially expressed genes among the analyzed
127 genotypes, and a coexpression network analysis and Gene Ontology (GO) annotation of
128 transcripts allowed us to investigate the main biological processes of candidate genes potentially
129 linked to the ACR. These candidate genes may be used to further explore and clarify the
130 mechanisms regulating apomixis in forage grasses.

131

132 # Results

133

134 ## Verification of the DNA content and mode of reproduction of *P.*
135 ## *notatum* accessions

136 The cytoembryological analysis confirmed the expected mode of reproduction of the
137 accessions based on the literature (Tables 1 and 2). The 2C DNA content of all plants from
138 accessions BGP_22 and BGP_306 was half the 2C DNA content from all plants from BGP_30,
139 BGP_34, BGP_115, and BGP_216. Therefore, all evaluated plants had DNA contents
140 compatible with the described ploidy levels of their respective accessions in the literature (Tables
141 1 and 2).

142 **Table 1. *Paspalum notatum* Accessions Evaluated in this Study.**

BGP	FORMER	REPRODUCTION	ORIGIN	COLLECTOR*	CHROMOSOME

CODE	CODE	MODE			NUMBER
306	BRA-024236	Sexual	Candói (PR), Brazil	VLiGu 14829	2n=2×=20 [47]
22	BRA-006173	Sexual	Bagé (RS), Brazil	VGnMaBd 9607	2n=2×=20 [48,49]
216	BRA-019470	Sexual**	Corrientes, Argentina	Q 3664	2n=4×=40 [50]
30	BRA-006467	Apomictic	Alegrete (RS), Brazil	VMrFrLw 9747	2n=4×=40 [51]
34	BRA-006513	Apomictic	Uruguaiana (RS), Brazil	VMrFrLw9782	2n=4×=40 [51,52]
115	BRA-010006	Apomictic	Laguna (SC), Brazil	VDBdSv 10137	2n=4×=40 [51]

* 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. [53].

Table 2. DNA Contents and Modes of Reproduction of the Bahiagrass Accessions Evaluated in this Study.

Accession	**DNA content**	**Mode of**

(BGP code)	Plants	2C (pg)	Δ2C among plants	Accession mean 2C (pg)	reproduction
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			
	BGP306_1[1,3]	1.5			
	BGP306_2	1.5			
306	BGP306_3	1.4	0.1	1.44	sexual
	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.

RNA-seq analysis and *de novo* transcriptome assembly

After removing barcodes sequences and low-quality and contaminant reads, 313,198,097 high-quality 72-bp paired-ends reads from floret and leaf tissues were used to assemble a *P. notatum* reference transcriptome. A total of 203,808 transcripts were assembled, out of which a set of 114,306 non-redundant transcripts were filtered (56.08% of all transcripts) (Table 3). Final transcripts in the reference transcriptome 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 non-redundant transcripts, we obtained 21,585 (18.88%) transcripts that were longer than 1 kb, a size range that commonly confers a high annotation rate. More than half of the total annotated transcripts were > 500 bp in length (Fig S1).

Table 3. Statistics for the Assembled Transcriptome of *Paspalum notatum*.

	All assembled transcripts	Non-redundant transcripts
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

The Bowtie aligner mapped 94.39% of sequenced reads onto the assembled transcripts. 73.78% mapped to the non-redundant transcripts set. Out of these, 82.97% and 83.48% of the sequencing reads belonged to sexual and apomictic samples, respectively, showing similar representation within the final transcriptome. 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%) fragmented sequences. One hundred and twenty-three (12%) BUSCO plant orthologs were not identified in the reference *P. notatum* transcriptome.

Transcriptome annotation

Among the 114,306 unigenes, 51,939 (45.44%) were similar to proteins from the NCBI Nr database, 34,750 (30.40%) to proteins from the UniProtKB/Swiss-Prot, and 54,568 (47.74%) to proteins from Phytozome.

177 The GO database was then used to retrieve the GO identifiers for all possible transcripts.
178 Overall, 21,537 (41.47%) transcripts were assigned to 4,614 GO terms in three main categories:
179 2,510 biological process, 1,486 molecular function, and 618 cellular components. The KEGG
180 annotation was possible for 9,221 annotated transcripts belonging to 129 pathways. Among
181 these, the most represented were Purine Metabolism Pathway (1,079 members), followed by
182 Metabolism of Thiamine (907 members) and Starch and Sucrose Metabolism (336 members).
183 Overall, 33,873 transcripts remained without hits after searching against all protein databases.
184 We then performed a search for the presence of open reading frames (ORFs), which were found
185 in 15,107 transcripts, 2,347 of which were complete. These sequences that showed ORFs without
186 annotation require further investigation since they may represent genes that have not yet been
187 described and possibly new genes that may be unique to *P. notatum*.
188
189 ## Expression levels of genes and identification of DEGs
190 The sequenced reads were realigned against the 114,306 transcripts of the reference
191 transcriptome to estimate the expression level of the transcripts in each genotype. Most identified
192 transcripts had expression levels of FPKM $\leq$ 100 (110,621 transcripts), whereas only 49
193 transcripts had high expression levels (FPKM $\geq$ 400), in these, most of them involved in
194 photosynthesis process, house-keeping genes and response to stress. The remaining transcripts
195 were not mapped and thus, their expression levels could not be determined. A comparison of the
196 110,656 transcripts that showed different levels of expression in the two tissues of the sequenced

197 samples revealed genes that were expressed in the genotypes of only one phenotypic class and
198 that were not expressed in the others (Fig 1), which were considered "unique" transcripts. Thus,
199 19,304 unique transcripts were found exclusively in the 4×APO group, 2,173 in the 2×SEX
200 group, and 217 in the 4×SEX group. All transcripts expressed in florets were also expressed in
201 leaves; there were no unique transcripts expressed in the florets from either diploid sexual or
202 tetraploid apomictic samples (Fig 1).

203 **Fig 1. Venn diagram showing the distribution of *P. notatum* transcripts.** Expression level of
204 transcripts from two tissues (floret and leaf) and each phenotypic class (sexual diploid, sexual
205 tetraploid, and apomictic tetraploid).

206 For pairwise differential expression analyses of all three phenotypic classes, we
207 considered only leaf samples because they expressed all representative transcripts from the
208 transcriptome assembly and had three clones for each genotype, maximizing the accuracy and
209 reliability of the differential gene expression analysis. However, within the same phenotypic
210 class, different expression patterns were observed among genotypes, mainly in 4×APO samples
211 (Fig S2). Thus, to reduce the effects of the differences in expression caused by genotype, we
212 performed a two-step pipeline. In the first step, we identified 72,318 transcripts that were equally
213 expressed among all 4×APO genotypes and 47,069 equally expressed transcripts in the 2×SEX
214 genotypes. This initial analysis was not necessary for the 4×SEX because only one genotype was
215 analyzed. From these, we selected 28,969 transcripts that were expressed in all three phenotypic
216 classes. In the second step of the differential expression analysis, pairwise comparisons of 28,969

transcripts allowed the identification of 2,072 DEGs between 2×SEX and 4×APO, which included 772 and 1,173 overexpressed transcripts, respectively. There were 1,173 DEGs between 2×SEX and 4×SEX, including 308 and 798 overexpressed transcripts, respectively, and 1,317 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 (Figs S5-S7).

Among the identified DEGs, we isolated a set of 510 transcripts that were consistently differentially expressed between apomictic and sexual samples independently of the ploidy level (diploid or tetraploid). We identified GO terms enriched in this set of transcripts, which belonged to 26 biological processes (BP), 13 molecular functions (MF), and three cellular components (CC) (Fig 2). We considered these transcripts as potentially involved in the determination of the reproductive mode.

Fig 2. Enriched ontology terms of 510 transcripts differentially expressed in the *P. notatum* transcriptome. Transcripts consistently differentially expressed between apomictic and sexual samples.

qRT-PCR validation of RNA-seq data

We selected a subset of 18 DEGs detected in-silico between apomictic and sexual samples to confirm the expression profiles through an independent technique, qRT-PCR. Five equally expressed transcripts detected in-silico showed desirable expression stability and

coefficient of variance in qRT-PCR reactions (Table S1), out of which one was selected to be an internal control for qRT-PCR verification of 2×SEX vs. 4×APO analyses (01RefGen-*Pnot*) and one for the verification of 4×APO vs. 4×SEX analyses (04RefGen-*Pnot*).

Out of 18 primer pairs designed for target DEGs, eight (Table S2) were used for qRT-PCR validation of RNA-seq data. The qRT-PCR analyses revealed consistent results compared to those detected through RNA-seq (Figs S3 and S4), emphasizing the power of the latter technique. Two primers pairs were not evaluated in qRT-PCR but showed interesting results in terms of DNA amplification. The first was designed from a transcript that was exclusively expressed in apomict samples in RNA-seq analyses (i.e., showing a zero expression value in the sexual samples). This transcript did not amplify using genomic DNA from sexual samples as PCR templates. For the second primer pair, although genomic DNA amplification occurred in all samples, when complementary DNA (cDNA) was used as PCR template, only apomictic samples presented amplicons. The expression level of this transcript estimated from RNA-seq of the same samples was very low in 2×SEX samples. These results suggest that the first transcript, expressed only in apomictic cytotypes, is exclusive to the genomes of these samples, whereas the second transcript was poorly expressed or silenced in sexual samples.

Detailed search for potential apomixis genes

We detected 1,612 transcripts that showed high similarity to the ACR in the *P. notatum* reference transcriptome. Out of these, 1,356 transcripts aligned to corresponded rice

chromosome 2, and 256 aligned to rice chromosome 12. Curiously, 40 of these transcripts were part of the set identified as unique to apomictic samples, nine were unique to sexual diploids, and six were unique to sexual tetraploid samples expression. Moreover, considering the analyses of differential expression, 16 putative ACR transcripts were differentially expressed in 2×SEX vs. 4×APO, 20 DEGs in 4×APO vs. 4×SEX, and 17 DEGs in 2×SEX vs. 4×SEX. All these findings, especially this set of 108 unique or differentially expressed transcripts, represent a valuable set of genes that deserve to be carefully investigated to determine their roles in apomixis and determine whether they are effectively genetically within the ACR in *P. notatum*.

The BLAST results in Table S3 show the similarity scores of the *P. notatum* transcripts to previously *Paspalum* sequences associated to the apomictic mode of reproduction, revealing that our transcriptome contains genes already found to be involved in asexual reproductive development in *Paspalum*.

Transcripts coexpression network

We identified 879,481 connections among 53,262 transcripts from RNA-seq data (Fig 3). Transcripts were grouped into 642 clusters according to their correlated pattern of expression level. In the network, the unique transcripts from each phenotypic class tended to form coexpression modules (Fig 3B-D). Despite this subdivision, the relevant biological relationships of these transcripts with all remaining transcripts can be recovered in a fully integrated network.

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

We identified a direct correlation among the 108 differentially expressed and/or unique putative ACR genes and their first neighbors in the gene expression network. Thus, we retrieved a sub-network composed of 536 strongly correlated transcripts (Fig 4). We used GO enrichment analysis to summarize the main putative functions of this set of transcripts, which included 43 biological processes (BP), 31 molecular functions (MF) and 17 cellular components (CC) (Fig. 5).

Fig 4. Gene expression sub-network for 536 *Paspalum notatum* transcripts. The 108 differentially expressed and/or unique putative ACR transcripts are presented in this sub-network, along with their first 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 transcripts in the gene expression sub-network.

Discussion

We constructed a *P. notatum* reference transcriptome using the next-generation sequencing approaches, RNA-seq, to investigate the gene expression changes associated with apomictic and sexual reproduction. RNA-seq enabled the assembly of a non-redundant transcriptome, containing 114,306 transcripts from florets and leaves from six distinct 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.

Even though we did not sequence inflorescences samples at different developmental stages (premeiosis, late premeiosis/meiosis, postmeiosis and anthesis) to search for candidate apomixis-linked genes, differential expression analyses between leaf samples from sexual and apomictic cytotypes allowed us to identify DEGs that may be representative of the mode of reproduction and not dependent of the level of ploidy. We isolated expression patterns of phenotypic classes, aimed both at removing transcripts whose expression could be related to the effect of genotype and at identifying transcripts expressed in all genotypes within the same phenotypic class.

GO classification of the assembled transcripts into 4,614 known terms, was used to perform functional enrichment analyses of candidate apomixis genes. Based on a set of 510 DEGs detected between apomictic and sexual, among the enriched GO terms, we emphasize “positive regulation of miRNA metabolic process (GO:2000630)” and “regulation of primary

316 miRNA processing (GO:2000636)", both of which either activate or increase the frequency, rate,
317 or extent of miRNA production. miRNA, in turn, is directly related to "DNA metabolic process
318 (GO:0006259)", which decreases the rate of gene expression (negative regulation of gene
319 expression - GO:0010629) through an epigenetic RNA-based gene silencing process (RNA-
320 directed DNA methylation - GO:0080188). Five transcripts were mainly involved in these
321 processes and showed higher expression levels in apomict samples in comparison to sexual
322 samples; these transcripts correspond to a pseudoARR-B transcription factor; pre-mRNA
323 splicing factor, putative (*Ricinus communis*); hypothetical protein SORBIDRAFT_05g016770;
324 uncharacterized protein, LOC100501330 (*Zea mays*); and an uncharacterized protein
325 LOC103641690. The GO analysis reveals the significant enrichment of terms associated with the
326 regulation of gene expression by epigenetic silencing among transcripts showing higher
327 expression levels among apomictic samples in comparison to sexual samples of the bahiagrass.
328 This result is consistent with the growing body of evidence that suggests that apomixis arises
329 from deregulation of the sexual pathway, where epigenetic mechanisms play a significant role in
330 at least some elements of apomictic development [8,15,54].

331 These DEGs represent a set of candidate genes that, together with the 19,304 transcripts
332 exclusively expressed in apomictic samples, deserve further investigation. In apospory, gene
333 expression occurs at specific stages of apomictic development such as apomeiosis,
334 parthenogenesis, and endosperm development. Increased or decreased expression of some genes
335 during these specific stages may hinder the analysis of differential expression between sexual

and apomictic genotypes [40,55]. Nonetheless, we believe that the identification of DEGs as performed herein, using selected transcripts with detectable expression pattern among all genotypes of all phenotypic classes, may have minimized this influence. At the same time, our approach increased the potential for discovery of candidate genes involved in apomictic process and not only in a single step. Thus, future reverse genetics experiments based on qRT-PCR and in situ hybridization could be useful to identify the role of specific genes in the whole apomixis process.

Functionally related genes tend to be transcriptionally coordinated [56,57]. Therefore, the construction of transcripts coexpression network provided a powerful resource, for the identification of transcripts that are coexpressed with unique and differentially expressed genes, despite showing undetectable differences in expression between samples. The sub-network containing 536 candidate coexpressed transcripts associated with the expression of apomixis is an example of how we can recover the relevant biological relationships of genes of interest using the transcriptome sequencing approaches. By starting on a smaller scale of transcripts that are DEGs and/or unique genes and possibly integrated into the ACR and combining the information by adding their nearest neighbors, we can obtain a broader view of the processes involved in the regulation of candidate genes. Significantly enriched BP were mainly associated with plant reproduction, for instance, “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)". In addition, the sub-network was enriched in 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)". Transcripts from this sub-network were associated with reproductive processes and the regulation of epigenetic changes by modulating histones. In particular, MAP kinase activity was negatively regulated. The mitogen-activated protein 3-kinase (MAP3K) gene family has been identified as differentially expressed in apomictic and sexual flowers of *P. notatum* [31] and in flowers of *P. simplex* [14], and it might play a role in parthenogenetic development of the embryo in both species [8]. Recently, an essential gene to the formation of unreduced embryo sacs in *P. notatum* was identified from the characterization of a MAP3K retrieved in previous transcriptomic surveys [58].

The size of the ACR in *Paspalum* is relatively small compared with other apomictic systems [8], but the lack of recombination makes it difficult to study this region by map-based cloning [31]. Thus, the discovery of DEGs that are potentially located within this region is extremely valuable for understanding the complex regulatory network of gene–gene interaction. These DEGs could also be used for future manipulation of the apomixis trait, which has outstanding importance in agricultural biotechnology.

375 In the genus *Paspalum,* the first approach for understanding the molecular basis of
376 apomixis was applied by Pessino et al. [27]. This approach led to the identification of three small
377 sequences that are highly expressed during early megagametophyte development in apomictic
378 plants through differential display experiments in the inflorescences of sexual and aposporous *P.*
379 *notatum*. Here, we recovered the reported sequences as a single representative transcript (Table
380 S3) similar to a kinesin motor protein, involved in the biological process of microtubule-based
381 movement. The kinesin motor protein has also been reported by other authors as differentially
382 expressed between apomictic and sexual plants [29,31,59]. Interestingly, the transcript assembled
383 in this work also showed differential expression levels between apomictic and sexual samples.
384 Since then, several other candidate genes have been reported using *P. notatum* and *P. simplex*
385 species [14,27,30,31] and more recently, the first approach using RNA-seq was published for
386 bahiagrass based on expression profiling of apomictic and sexual flowers [40]. Some of these
387 candidate genes had been investigated in more detail [35,37-39]. However, many questions
388 remain unanswered, and further research is needed to define the relationships between the
389 structure, the position, and the function of the known apomixis-linked genes [8]. Furthermore,
390 based on the set of genes selected through differential expression patterns, we recovered
391 previously published sequences in our *P. notatum* transcriptome.

392 One of the candidate genes previously reported in the literature is n20gap-1, a lorelei-like
393 *P. notatum* gene [31,35] encoding a GPI-anchored protein that supposedly plays a role in the
394 final stages of the apomixis developmental cascade. Laspina et al. [31] previously reported this

395 sequence as linked to the chromosomal locus governing apospory at a genetic distance of 22 cM.
396 Here, one transcript similar to this gene present in the set of transcripts potentially integrated
397 within the ACR, corresponding to rice chromosome 2. Additionally, we found transcripts,
398 including some that were differentially expressed, that aligned with *P. notatum* sequences
399 characterized by Podio et al. [38], corresponding to *Somatic Embryogenesis Receptor-Like*
400 *Kinase* (*SERK*), in addition to other numerous transcripts with the same annotation. The
401 candidate SERK-like genes play crucial roles in somatic embryogenesis in angiosperms and have
402 been reported as related to apomixis. Albertini et al. [59] analyzed two members of this protein
403 family, namely *PpSERK1* and *PpSERK2*, and found that *PpSERK1* expression levels was high
404 during premeiosis and decreased during the meiosis and post-meiotic stages, whereas *PpSERK2*
405 expression was high from premeiosis to anthesis in the nucellar cells of aposporous genotypes.
406 The authors proposed that the expression pattern of *PpSERK* in *Poa pratensis* was compatible
407 with its role in the specification of aposporous initials. In *P. notatum*, the expression of two
408 different members of the *SERK* family (*PnSERK1* and *PnSERK2*) was observed, and *PnSERK2*
409 displayed a spatial expression pattern similar to that reported for the *PpSERKs*, which are
410 expressed in nucellar cells at meiosis in the apomictic genotype [38]. We also identified
411 transcripts that showed similarity to other interesting candidate genes related to apospory, such
412 as the PnTgs1-like gene that encodes a trimethylguanosine synthase-like protein, which plays a
413 fundamental role in nucellar cell fate, as its diminished expression is correlated with initiation of
414 the apomictic pathway in plants [37]. Additionally, two transcripts 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 [39]. The recovery of these previously published sequences demonstrates the representative and informative nature of assembled transcriptome. Indeed, 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 assemble a useful transcriptome and identify differences between apomictic and sexual reproduction. Our results reveal DEGs, genes exclusively expressed in apomicts or sexuals, of which are in potential association with the ACR genomic region, including the apo locus. The validation of genes from this set of candidates may enable valuable insights to the understanding of 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.

Materials and methods

Plant materials collections 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) (Table 1). The accessions were chosen based on their origins

(different ecotypes), genetic dissimilarity [60] and ploidy levels and reproduction modes. Tetraploid *P. notatum* is apomictic, whereas 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* [50]. The Q3664 is characterized as a facultative apomictic accession with a high level of sexuality (> 70%) [53].

For this study, five clones (biological replicates) of each accession were planted in 8 L pots, filled with 1:1 soil/vermiculite and grown in a greenhouse under the same environmental conditions. The climate is humid subtropical (according to the Köppen-Geiger classification system), 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 [61].

Young leaf samples from three clones of each accession were collected during summer (December). One clone each of the accessions BGP34 (apomictic tetraploid) and BGP22 (sexual diploid) presented 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. [62]. RNA integrity was assessed in a 1% agarose denaturing gel and quantified using a NanoVue Plus spectrophotometer (GE Healthcare Life Sciences, Little Chalfont, UK).

Verification of the mode of reproduction of *P. notatum* accessions through cytoembryological analysis

455 The mode of reproduction of the plants was confirmed using the embryo sac clarification
456 method proposed by Young et al. [63], with minor modifications. Inflorescences at anthesis
457 (when the embryo sac is fully developed) were collected and fixed in FAA (95% ethanol,
458 distilled water, glacial acetic acid, formalin 40%, 40:13:3:3 v/v) for 24 h at room temperature.
459 Afterward, the FAA was replaced with 70% ethanol and the samples were stored at 4°C. Ovaries
460 were extracted by dissection under a stereoscopic microscope and stored in 70% ethanol. The
461 embryo sacs were clarified by replacing 70% alcohol with the following series of solutions: 85%
462 ethanol; absolute ethanol; ethanol:methyl salicylate (1:1); ethanol:methyl salicylate (1:3); 100%
463 methyl salicylate (two times). The samples remained in each solution for 24 h. Observations
464 were carried out with an Axiophot microscope (Carl Zeiss, Jena, Germany) using the differential
465 interference contrast (DIC) microscopy technique. A total of 100 embryo sacs per accession were
466 evaluated.

467

468 Verification of the ploidy level of *P. notatum* accessions through flow 469 cytometry

470 Flow cytometry analyses were performed to confirm the exact ploidy level of each plant
471 used in the experiment and results were compared to the literature data. Approximately 5 mg of
472 young leaf tissue from each plant was used (Table 1). *Pisum sativum* cv. Ctirad samples
473 (2C=9.09 pg), were used as an internal control [64,65]. Samples were triturated in a petri dish
474 containing 800 μL of LB01 buffer (0.45425 g TRIS, 0.186125 g NaEDTA, 0.0435 g-spermine,

475 0.29225 g NaCl, 1.491 g KCl, and 250 µL of Triton X-100 in 250 mL of distilled water, 7.5 pH,
476 0.11% v/v of β-mercaptoethanol) to obtain a nuclear suspension [66]. The nuclear suspension
477 was filtered through a mesh of 40 µm and incubated at room temperature for 5 min, followed by
478 the addition of 25 µL of propidium iodide and 25 µL of RNase. For each sample, at least 10,000
479 nuclei were analyzed. Samples were analyzed with a FACSCalibur flow cytometer (Becton
480 Dickinson, New Jersey, USA). Histograms were obtained in Cell Quest software and analyzed in
481 Flowing Software (available at http://www.flowingsoftware.com). Only histograms with both
482 peaks (sample and standard) of approximately the same height were considered. The nuclear
483 DNA index (pg) of the plants was estimated based on the value of the G1 peak as an internal
484 reference. We calculated the mean value of C per biological replicate and accession and the
485 difference between the highest and lowest values (Δ) observed in replicates of each accession.
486

487 RNA-seq library construction, Illumina sequencing, and data
488 quality control

489 The cDNA libraries were constructed from each RNA sample (18 leaf and two floret
490 libraries) using specific barcodes and the TruSeq RNA Sample Preparation Kit (Illumina Inc.,
491 San Diego, CA, USA) following manufacturer's instructions. Library quality was confirmed
492 using the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and quantified
493 via quantitative real-time PCR (qPCR) (Illumina protocol SY-930-10-10). Clustering was
494 conducted using a TruSeq Paired-End Cluster Kit on a cBot (Illumina Inc., San Diego, CA,

495 USA). Paired-end sequencing was performed on the Illumina Genome Analyzer IIx Platform
496 with TruSeq SBS 36-Cycle kits (Illumina, San Diego, CA, USA) following manufacturer's
497 specifications. RNA-seq was performed with floret libraries of the accessions BGP22
498 (diploid/sexual) and BGP34 (tetraploid/apomictic) in two separate lanes, without biological
499 replicates. The remaining 18 leaf libraries from six different accessions (Table 1), each with
500 three independent biological replicates (clones), were distributed randomly in the flow cell, with
501 three libraries per lane.

502 Raw data was converted to FastQ files containing 72-bp reads. Quality control was
503 performed using the NGS QC Toolkit v2.3.3 [67]. Initially, high-quality reads (Phred quality
504 score $\geq$ 20 in at least 75% of bases) and reads with more than 60 bases were selected.
505 Subsequently, reads were trimmed at the 3' end for the removal of barcodes. All reads were
506 deposited in the NCBI Short Read Archive (SRA) under accession number SRP150615.

507

508 ## Transcriptome assembly and completeness assessment

509 High-quality reads from leaves and florets were *de novo* assembled into a reference
510 transcriptome for *P. notatum*, performed with Trinity program v2.0.2 [68] using default settings.
511 Contigs redundancy was minimized through the selection of the first Butterfly sequence from
512 each Chrysalis component, which is considered the most representative contig. To assess the
513 sequenced reads-support of the assembled contigs, the high-quality short reads were mapped
514 back to the transcriptome using the Bowtie2 sequence aligner v2.2.5 [69]. The Benchmarking

Universal Single-Copy Orthologs (BUSCO) tool, an approach for the assessment of conserved orthologs among plant species in a set of sequences [70], was used to estimate the completeness of the transcriptome assembly.

Transcriptome annotation

All transcripts were aligned to proteins from the the NCBI non-redundant (Nr) database, from the UniProtKB/Swiss-Prot and from the grass protein database from Phytozome v9.0, using the BLASTX algorithm with an e-value cutoff of 1e-06. Gene Ontology (GO) [71] terms were retrieved for transcripts showing similarity to sequences from NCBI Nr or UniProtKB databases using Blast2GO software [72]. We used REVIGO [73] to summarize and visualize GO terms sets. Pathways were assigned to metabolic pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [74].

Estimation of transcript expression levels

The expression levels of the transcripts were estimated using the FPKM method (expected number of fragments per kilobase of transcript sequence per million base pairs sequenced). Read counts for each transcript from each sample (Table 1) was obtained using Bowtie v2-2.1.0 and normalized using the RSEM software (RNA-seq by expectation maximization) [75]. A Venn diagram for the visualization of the amount of exclusive and shared transcripts among samples was created using the online platform available at

535 http://bioinformatics.psb.ugent.be/webtools/Venn/. Principal component analysis (PCA) was
536 used to assess expression patterns of sequenced genotypes.

537

538 ## Differential expression analysis

539 EBSeq [76] was used to identify differentially expressed genes (DEGs) using FPKM
540 values, at a false discovery rate (FDR) ≤ 0.05. Transcripts with a log2 fold change (FC) ≥ 1.5 in
541 transcript abundance were regarded as overexpressed. GO enrichment analysis was carried out
542 using R software (version 3.3.1) and the 'goseq' 1.24.0 Bioconductor package [77]. P-values
543 were subjected to Bonferroni correction, and we considered adjusted p-values ≤ 0.05 as enriched.
544 Resulting enriched GO terms were summarized using REVIGO [73].

545 Genotypes were grouped according to their phenotypic classes as sexual diploids
546 (2×SEX), sexual tetraploids (4×SEX) and apomictic tetraploids (4×APO) to be compared in
547 pairwise differential expression analyses. A two-step pipeline was used for the identification of
548 the DEGs (Fig 6). In the first step, the objective was to select a list of equally expressed
549 transcripts among genotypes belonging to the same phenotypic class. This procedure allowed the
550 detection of gene expression patterns related with characteristics that were shared among
551 genotypes, independently of physiological or developmental particularities. Thus, EBSeq was
552 used to estimate the pairwise posterior probabilities of a transcript being equally express between
553 genotypes: sexual diploids (accessions: BGP22 and BGP306), and apomictic tetraploids
554 (accessions: BGP30, BGP34 and BGP115). Transcripts that presented PPEE ≥ 0.95 were kept for

555 the second step of the analysis. The second step consisted of another round of pairwise
556 differential expression analyses between phenotypic classes: (i) 2×SEX vs. 4×APO; (ii) 2×SEX
557 vs. 4×SEX; (iii) 4×APO vs. 4×SEX.

558 **Fig 6. Workflow of differential expression analyses used in this study.** We realized pairwise
559 differential expression analyses of all three phenotypic classes (apomictic tetraploid "4×APO",
560 sexual diploid "2×SEX" and sexual tetraploid "4×SEX"). Each phenotypic class had different
561 genotype with three clones. In the first step, the objective was to select a list of equally expressed
562 transcripts among genotypes belonging to the same phenotypic class. EBSeq was used to
563 estimate the pairwise posterior probabilities of a transcript being equally express (PPEE)
564 between genotypes of the same phenotypic class: apomictic tetraploid (A×B; A×C; B×C) and
565 sexual diploid (D×E). Transcripts that presented PPEE ≥ 0.95 were kept for the second step of
566 the analysis. The second step consisted of another round of pairwise differential expression
567 analyses between phenotypic classes: (i) 4×APO vs. 2×SEX; (ii) 4×APO vs. 4×SEX; (iii) 2×SEX
568 vs. 4×SEX. The objective was to select a list of transcripts that presented pairwise posterior
569 probabilities of being differentially expressed (PPDE ≥ 0.95). Additionally, we identified a list of
570 transcripts that were consistently differentially expressed between sexual and apomictic samples,
571 independently of the ploidy level (diploid or tetraploid).

572

573 Quantitative reverse transcription PCR (qRT-PCR) validation of
574 differential expression results

575 To verify the reliability and accuracy of transcriptome data and differential expression
576 analyses pipeline, 18 DEGs were randomly selected for quantification through reverse
577 transcription PCR (qRT-PCR). Eight transcripts showing similar expression patterns based on
578 FPKM values were used as internal reaction controls. Primer sets for controls and targets were
579 designed using Primer3Plus software [78]. All primer pairs were initially tested in regular PCR
580 reactions using genomic DNA as a template from the same genotypes and clones used in RNA-
581 seq. Only primers that amplified the genomic DNA of all genotypes were used in the following
582 qRT-PCR amplification efficiency tests. Primer pairs with an amplification efficiency of 90-
583 110% and R^2> 0.99 were used for relative expression analyses.

584 Total RNA (500 ng) was used for cDNA synthesis using the QuantiTect Reverse
585 Transcription Kit (Qiagen Inc., Chatsworth, CA, USA). qRT-PCR were performed using a
586 CFX384 Real-Time PCR Detection System with iTaq Universal SYBR Green Supermix (Bio-
587 Rad Laboratories Inc., Hercules, CA, USA), according to the manufacturer's instructions. The
588 reaction conditions were 95°C (10 min) and then 40 cycles of 95°C (30 s) and 60°C (1 min). All
589 experiments were performed using technical triplicate, and no-template controls were included.
590 The specificity of amplicons was confirmed through the analysis of the melting curve. The
591 baseline and Cq (quantification cycle) values were automatically determined, and the expression
592 values were estimated using the ΔΔCt method implemented by the CFX Manager 2.1 software
593 (Bio-Rad Laboratories, Inc., USA). Reference genes were selected according to their expression
594 stability (M < 0.5) and coefficient of variance (CV < 0.25) among samples. Mann-Whitney U-

595 tests were performed to estimate statistical significance of distribution differences between
596 samples.

597

598 ## Search for putative apomixis genes

599 To perform a detailed comparative analysis of *P. notatum* transcripts obtained in this
600 study with their syntenic counterparts on rice chromosomes that are conservatively linked to
601 apomixis in *Paspalum* species [8], we aligned all *P. notatum* transcripts to rice transcripts using
602 the BLASTN tool and selected those that presented putative homology to genes in the
603 highlighted area of rice chromosomes 2 and 12. Then, we selected only *P. notatum* transcripts
604 present in the ACR, which were delimited by a set of molecular markers completely linked to the
605 apospory locus in this species. Thus, the region encompassed the C1069 and C996 markers for
606 rice chromosome 12 and between C560 and C932 markers for chromosome 2 [8,21,24-26,79].
607 We also compared the *P. notatum* reference transcriptome with sequences reported for the genus
608 *Paspalum*, potentially associated with apomixis. The sequences were used as queries for
609 BLASTN search in the 114,306 transcripts database, with an e-value cutoff of 1e-05.

610

611 ## Construction of a transcripts coexpression networks

612 To generate coexpression network, we used expression values in FPKM of all assembled
613 transcripts from floret and leaf tissues. Transcripts showing null values for most of the replicates
614 were excluded to reduce noise and eliminate residuals in the analysis. We calculated an all-

615 versus-all coexpression network matrix using the Pearson correlation coefficient cutoff of ≥ 0.8.
616 The highest reciprocal rank (HRR) method proposed by Mutwil et al. [80] was used to select
617 only the strongest edges, considering an HRR limit of 30. The Heuristic Cluster Chiseling
618 Algorithm (HCCA) was applied to partition the networks into manageable clusters, with three-
619 step node vicinity networks (NVN) [80]. The interactive coexpression network was visualized
620 using the Cytoscape 3.4.0 software [81].
621

622 # Acknowledgments

623 We are grateful to Maria Augusta C. Horta and Natalia F. Murad for their help in
624 generating the coexpression networks and Mariana V. Cruz for critically reading the manuscript.
625
626
627

References

1. Asker S, Jerling L. Apomixis in plants. Boca Raton, FL: CRC Press; 1992.
2. Nogler GA. Gametophytic apomixis. In: Johri BM, editor. Embryology of angiosperms. Berlin: Springer; 1984. pp. 475-518.
3. Koltunow AM. Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules. Plant Cell. 1993;5: 1425-1437.
4. Conner JA, Goel S, Gunawan G, Cordonnier-Pratt MM, Johnson VE, Liang C, et al. Sequence analysis of bacterial artificial chromosome clones from the apospory-specific genomic region of *Pennisetum* and *Cenchrus*. Plant Physiol. 2008;147: 1396-1411.
5. Conner JA, Gunawan G, Ozias-Akins P. Recombination within the apospory specific genomic region leads to the uncoupling of apomixis components in *Cenchrus ciliaris*. Planta. 2013;238: 51-63.
6. Conner JA, Mookkan M, Huo H, Chae K, Ozias-Akins P. A parthenogenesis gene of apomict origin elicits embryo formation from unfertilized eggs in a sexual plant. Proc Natl Acad Sci U S A. 2015;112: 11205-11210.
7. Carman JG. Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biol J Linn Soc. 1997;61: 51-94.
8. Ortiz JPA, Quarin CL, Pessino SC, Acuña C, Martínez EJ, Espinoza F, et al. Harnessing apomictic reproduction in grasses: what we have learned from *Paspalum*. Ann Bot. 2013;112: 767-787.

9. Hojsgaard D, Klatt S, Baier R, Carman JG, Hörandl E. Taxonomy and biogeography of apomixis in *Angiosperms* and associated biodiversity characteristics. CRC Crit Rev Plant Sci. 2014;33: 414-427.
10. Ozias-Akins P. Apomixis: developmental characteristics and genetics. CRC Crit Rev Plant Sci. 2006;25: 199-214.
11. Adams KL, Wendel JF. Polyploidy and genome evolution in plants. Curr Opin Plant Biol. 2005;8: 135-141.
12. Cervigni GDL, Paniego N, Pessino S, Selva JP, Díaz M, Spangenberg G, et al. Gene expression in diplosporous and sexual *Eragrostis curvula* genotypes with differing ploidy levels. Plant Mol Biol. 2008;67: 11-23.
13. Spillane C, Curtis MD, Grossniklaus U. Apomixis technology development—virgin births in farmers' fields? Nat Biotechnol. 2004;22: 687-691.
14. Polegri L, Calderini O, Arcioni S, Pupilli F. Specific expression of apomixis-linked alleles revealed by comparative transcriptomic analysis of sexual and apomictic *Paspalum simplex* Morong flowers. J Exp Bot. 2010;61: 1869-1883.
15. Calderini O, Donnison I, Polegri L, Panara F, Thomas A, Arcioni S, et al. Partial isolation of the genomic region linked with apomixis in *Paspalum simplex*. Mol Breeding. 2011;28: 265-276.
16. Selva JP, Pessino SC, Meier MS, Echenique VC. Identification of candidate genes related to polyploidy and/or apomixis in *Eragrostis curvula*. Am J Plant Sci. 2012;03: 403-416.

17. Burton GW. The method of reproduction in common bahia grass, *Paspalum notatum1*. J Am Soc Agron. 1948;40: 443-452.

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

19. Quarin CL, Urbani MH, Blount AR, Martinez EJ, Hack CM, Burton GW, et al. Registration of Q4188 and Q4205, sexual tetraploid germplasm lines of bahiagrass. Crop Sci. 2003;43: 745-746.

20. Podio M, Cáceres ME, Samoluk SS, Seijo JG, Pessino SC, Ortiz JPA, et al. A methylation status analysis of the apomixis-specific region in *Paspalum* spp. suggests an epigenetic control of parthenogenesis. J Exp Bot. 2014;65: 6411-6424.

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

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

23. Pupilli F, Labombarda P, Caceres ME, Quarín CL, Arcioni S. The chromosome segment related to apomixis in *Paspalum simplex* is homoeologous to the telomeric region of the long arm of rice chromosome 12. Mol Breed. 2001;8: 53-61.

687 24. Martínez EJ, Hopp HE, Stein J, Ortiz JPA, Quarin CL. Genetic characterization of
688 apospory in tetraploid *Paspalum notatum* based on the identification of linked molecular
689 markers. Mol Breed. 2003;12: 319-327.
690 25. Stein J, Pessino SC, Martínez EJ, Rodriguez MP, Siena LA, Quarin CL, et al. A genetic
691 map of tetraploid *Paspalum notatum* Flügge (bahiagrass) based on single-dose molecular
692 markers. Mol Breed. 2007;20: 153-166.
693 26. Podio M, Rodríguez MP, Felitti S, Stein J, Martínez EJ, Siena LA, et al. Sequence
694 characterization, *in silico* mapping and cytosine methylation analysis of markers linked to
695 apospory in *Paspalum notatum*. Genet Mol Biol. 2012;35: 827-837.
696 27. Pessino SC, Espinoza F, Martínez EJ, Ortiz JPA, Valle EM, Quarín CL. Isolation of
697 cDNA clones differentially expressed in flowers of apomictic and sexual *Paspalum*
698 *notatum*. Hereditas. 2001;134: 35-42.
699 28. Rodrigues JCM, Cabral GB, Dusi DMA, de Mello LV, Rigden DJ, Carneiro VTC.
700 Identification of differentially expressed cDNA sequences in ovaries of sexual and
701 apomictic plants of *Brachiaria brizantha*. Plant Mol Biol. 2003;53: 745-757.
702 29. Albertini E, Marconi G, Barcaccia G, Raggi L, Falcinelli M. Isolation of candidate genes
703 for apomixis in *Poa pratensis* L. Plant Mol Biol. 2004;56: 879-894.
704 30. Calderini O, Chang SB, de Jong H, Busti A, Paolocci F, Arcioni S, et al. Molecular
705 cytogenetics and DNA sequence analysis of an apomixis-linked BAC in *Paspalum*

706 *simplex* reveal a non pericentromere location and partial microcolinearity with rice. Theor
707 Appl Genet. 2006;112: 1179-1191.
708 31. Laspina NV, Vega T, Seijo JG, González AM, Martelotto LG, Stein J, et al. Gene
709 expression analysis at the onset of aposporous apomixis in *Paspalum notatum*. Plant Mol
710 Biol. 2008;67: 615-628.
711 32. Sharbel TF, Voigt ML, Corral JM, Thiel T, Varshney A, Kumlehn J, et al. Molecular
712 signatures of apomictic and sexual ovules in the *Boechera holboellii* complex. Plant J.
713 2009;58: 870-882.
714 33. Yamada-Akiyama H, Akiyama Y, Ebina M, Xu Q, Tsuruta S, Yazaki J, et al. Analysis of
715 expressed sequence tags in apomictic guineagrass (*Panicum* maximum). J Plant Physiol.
716 2009;166: 750-761.
717 34. Sharbel TF, Voigt ML, Corral JM, Galla G, Kumlehn J, Klukas C, et al. Apomictic and
718 sexual ovules of *Boechera* display heterochronic global gene expression patterns. Plant
719 Cell. 2010;22: 655-671.
720 35. Felitti SA, Seijo JG, González AM, Podio M, Laspina NV, Siena L, et al. Expression of
721 lorelei-like genes in aposporous and sexual *Paspalum notatum* plants. Plant Mol Biol.
722 2011;77: 337-354.
723 36. Okada T, Hu Y, Tucker MR, Taylor JM, Johnson SD, Spriggs A, et al. Enlarging cells
724 initiating apomixis in *Hieracium praealtum* transition to an embryo sac program prior to
725 entering mitosis. Plant Physiol. 2013;163: 216-231.

726 37. Siena LA, Ortiz JPA, Leblanc O, Pessino S. PNTGS1-like expression during
727 reproductive development supports a role for RNA methyltransferases in the aposporous
728 pathway. BMC Plant Biol. 2014;14: 297.
729 38. Podio M, Felitti SA, Siena LA, Delgado L, Mancini M, Seijo JG, et al. Characterization
730 and expression analysis of SOMATIC EMBRYOGENESIS RECEPTOR KINASE
731 (SERK) genes in sexual and apomictic *Paspalum notatum*. Plant Mol Biol. 2014;84: 479-
732 495.
733 39. Siena LA, Ortiz JPA, Calderini O, Paolocci F, Cáceres ME, Kaushal P, et al. An
734 apomixis-linkedORC3-like pseudogene is associated with silencing of its functional
735 homolog in apomictic *Paspalum simplex*. J Exp Bot. 2016;67: 1965-1978.
736 40. Ortiz JPA, Revale S, Siena LA, Podio M, Delgado L, Stein J, et al. A reference floral
737 transcriptome of sexual and apomictic *Paspalum notatum*. BMC Genomics. 2017;18:
738 318.
739 41. Bocchini M, Galla G, Pupilli F, Bellucci M, Barcaccia G, Ortiz JPA, et al. The vesicle
740 trafficking regulator PN_SCD1 is demethylated and overexpressed in florets of apomictic
741 *Paspalum notatum* genotypes. Sci Rep. 2018;8: 3030.
742 42. Huang W, Guo Y, Du W, Zhang X, Li A, Miao X. Global transcriptome analysis
743 identifies differentially expressed genes related to lipid metabolism in Wagyu and
744 *Holstein cattle*. Sci Rep. 2017;7: 5278.

745 43. Liang R, Han B, Li Q, Yuan Y, Li J, Sun D. Using RNA sequencing to identify putative
746 competing endogenous RNAs (ceRNAs) potentially regulating fat metabolism in bovine
747 liver. Sci Rep. 2017;7: 6396.
748 44. Miao X, Luo Q, Zhao H, Qin X. Co-expression analysis and identification of fecundity-
749 related long non-coding RNAs in sheep ovaries. Sci Rep. 2016;6: 39398.
750 45. de la Fuente A. From 'differential expression' to 'differential networking' – identification
751 of dysfunctional regulatory networks in diseases. Trends Genet. 2010;26: 326-333.
752 46. Gov E, Arga KY. Interactive cooperation and hierarchical operation of microRNA and
753 transcription factor crosstalk in human transcriptional regulatory network. IET Syst Biol.
754 2016;10: 219-228.
755 47. Dahmer N, Schifino-Wittmann MT, Dall'Agnol M, Castro BD. Cytogenetic data for
756 *Paspalum notatum* Flügge accessions. Sci Agric. 2008;65: 381-388.
757 48. Pagliarini MS, Takayama SY, de Freitas PM, Carraro LR, Adamowski EV, Silva N, et al.
758 Failure of cytokinesis and 2n gamete formation in Brazilian accessions of *Paspalum*.
759 Euphytica. 1999;108: 129-135.
760 49. Pagliarini MS, Carraro LR, Freitas PM, Adamowski EV, Batista LAR, Valls JFM.
761 Cytogenetic characterization of Brazilian *Paspalum* accessions. Hereditas. 2001;135: 27-
762 34.
763 50. Forbes I, Burton GW. Cytology of Diploids, natural and induced tetraploids, and intra-
764 species hybrids of bahiagrass, *Paspalum notatum* Flugge. Crop Sci. 1961;1: 402-406.

765 51. Pozzobon MT, Valls JFM. Chromosome number in germplasm accessions of *Paspalum*
766 *notatum* (Gramineae). Braz J Genet. 1997;20: 29-34.
767 52. Adamowski EDV, Pagliarini MS, Bonato ABM, Batista LAR, Valls JFM. Chromosome
768 numbers and meiotic behavior of some *Paspalum* accessions. Genet Mol Biol. 2005;28:
769 773-780.
770 53. Quarin CL, Burson BL, Burton GW. Cytology of intra- and interspecific hybrids between
771 two cytotypes of *Paspalum notatum* and *P. cromyorrhizon*. Bot Gaz. 1984;145: 420-426.
772 54. Brukhin V. Molecular and genetic regulation of apomixis. Russ J Genet. 2017;53: 943-
773 964.
774 55. Schmidt A, Schmid MW, Klostermeier UC, Qi W, Gutho□rl D, Sailer C, et al.
775 Apomictic and sexual germline development differ with respect to cell cycle,
776 transcriptional, hormonal and epigenetic regulation. PLoS Genet. 2014;10: e1004476.
777 56. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global
778 discovery of conserved genetic modules. Science. 2003;302: 249-255.
779 57. Persson S, Wei H, Milne J, Page GP, Somerville CR. Identification of genes required for
780 cellulose synthesis by regression analysis of public microarray data sets. Proc Natl Acad
781 Sci U S A. 2005;102: 8633-8638.
782 58. Mancini M, Permingeat HR, Colono C, Siena LA, Pupilli F, Azzaro C, et al. The
783 MAP3K-coding QUI-GON JINN (QGJ) gene is essential to the formation of unreduced
784 embryo sacs in *Paspalum*. Front. Plant Sci. 2018. doi: 10.3389/fpls.2018.01547.

785 59. Albertini E, Marconi G, Reale L, Barcaccia G, Porceddu A, Ferranti F, et al. *SERK* and
786 *APOSTART*. Candidate genes for apomixis in *Poa pratensis*. Plant Physiol. 2005;138:
787 2185-2199.
788 60. Cidade FW, Vigna BBZ, de Souza FHD, Valls JFM, Dall'Agnol M, Zucchi MI, et al.
789 Genetic variation in polyploid forage grass: assessing the molecular genetic variability in
790 the *Paspalum* genus. BMC Genet. 2013;14: 50.
791 61. CEPAGRI. (2010) Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à
792 Agricultura. Available from: http://www.cpa.unicamp.br/o-cepagri.html Cited 17 August
793 2016.
794 62. Oliveira RR, Viana AJC, Reátegui ACE, Vincentz MGA. Short communication an
795 efficient method for simultaneous extraction of high-quality RNA and DNA from various
796 plant tissues. Genet Mol Res. 2015;14: 18828-18838.
797 63. Young BA, Sherwood RT, Bashaw EC. Cleared-pistil and thick-sectioning techniques for
798 detecting aposporous apomixis in grasses. Can J Bot. 1979;57: 1668-1672.
799 64. Dolezel J, Sgorbati S, Lucretti S. Comparison of three DNA fluorochromes for flow
800 cytometric estimation of nuclear DNA content in plants. Physiol Plant. 1992;85: 625-631.
801 65. Greilhuber J, Temsch EM, Loureiro JCM. Nuclear DNA content measurements. In:
802 Doležel J, Greilhuber J, Suda. J, editors. Flow cytometry with plant cells: analysis of
803 genes, chromosomes and genomes. Weinheim, Germany: Wiley-VCH Verlag GmbH and
804 Co.; 2007. pp. 67-101.

805 66. Dolezel J. Flow cytometry, its application and potential for plant breeding. In: Lelley T,
806 editor. Current topics in plant cytogenetics related to plant improvement. Vienna: WUV-
807 Universitätsverlag; 1997. pp. 80-90.
808 67. Patel RK, Jain M. NGS QC toolkit: a toolkit for quality control of next generation
809 sequencing data. PLoS One. 2012;7: e30619.
810 68. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length
811 transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol.
812 2011;29: 644-652.
813 69. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment
814 of short DNA sequences to the human genome. Genome Biol. 2009;10: R25.
815 70. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO:
816 assessing genome assembly and annotation completeness with single-copy orthologs.
817 Bioinformatics. 2015;31: 3210-3212.
818 71. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology:
819 tool for the unification of biology. Nat Genet. 2000;25: 25-29.
820 72. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a
821 universal tool for annotation, visualization and analysis in functional genomics research.
822 Bioinformatics. 2005;21: 3674-3676.
823 73. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists
824 of gene ontology terms. PLoS One. 2011;6: e21800.

825 74. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids
826 Res. 2000;28: 27-30.
827 75. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or
828 without a reference genome. BMC Bioinformatics. 2011;12: 323.
829 76. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, et al. EBSeq: an
830 empirical Bayes hierarchical model for inference in RNA-seq experiments.
831 Bioinformatics. 2013;29: 1035-1043.
832 77. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq:
833 accounting for selection bias. Genome Biol. 2010;11: R14.
834 78. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3plus,
835 an enhanced web interface to Primer3. Nucleic Acids Res. 2007;35: W71-W74.
836 79. Labombarda P, Busti A, Caceres ME, Pupilli F, Arcioni S. An AFLP marker tightly
837 linked to apomixis reveals hemizygosity in a portion of the apomixis-controlling locus in
838 *Paspalum simplex*. Genome. 2002;45: 513-519.
839 80. Mutwil M, Usadel B, Schutte M, Loraine A, Ebenhoh O, Persson S. Assembly of an
840 interactive correlation network for the *Arabidopsis* genome using a Novel Heuristic
841 Clustering algorithm. Plant Physiol. 2010;152: 29-43.
842 81. Shannon P. Cytoscape: a software environment for integrated models of biomolecular
843 interaction networks. Genome Res. 2003;13: 2498-2504.
844

Supporting Information

S1 Fig. Length distribution of the number of non-redundant transcripts successful annotated. Comparison of the all non-redundant transcripts and the number of transcripts annotated in the NCBI non-redundant protein database by size range.

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

S3 Fig. Histograms of gene expression obtained by qRT-PCR. qRT-PCR validation 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 phenotypic classes compared using the Mann-Whitney U-test.

S4 Fig. Histograms of gene expression obtained by qRT-PCR. qRT-PCR validation 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 phenotypic classes compared using the Mann-Whitney U-test.

S5 Fig. 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

865 enriched in overexpressed transcripts in 2×SEX and b) categories enriched in overexpressed
866 transcripts in 4×APO.

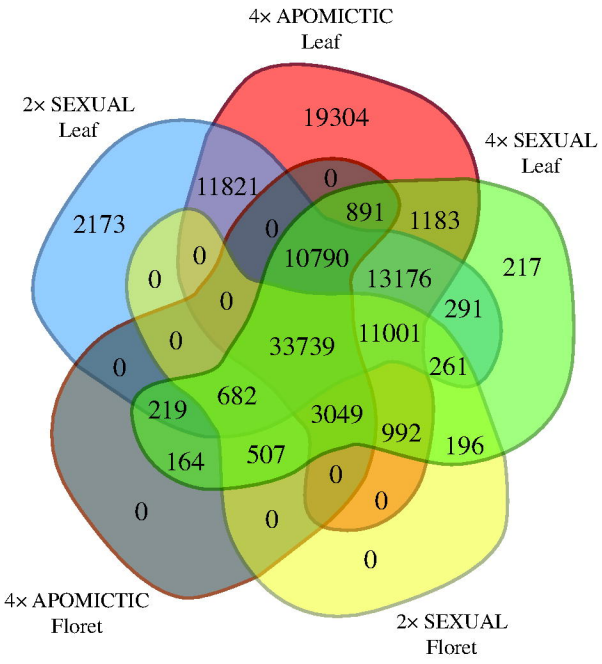
867 **S6 Fig. Functional classification of enriched overexpressed DEGs in the 2×SEX vs. 4×SEX**
868 **comparison.** Gene Ontology biological process (blue boxes), Gene Ontology cellular component
869 (yellow boxes), and Gene Ontology molecular function (orange boxes). a) Categories enriched in
870 overexpressed transcripts in 2×SEX and b) categories enriched in overexpressed transcripts in
871 4×SEX.

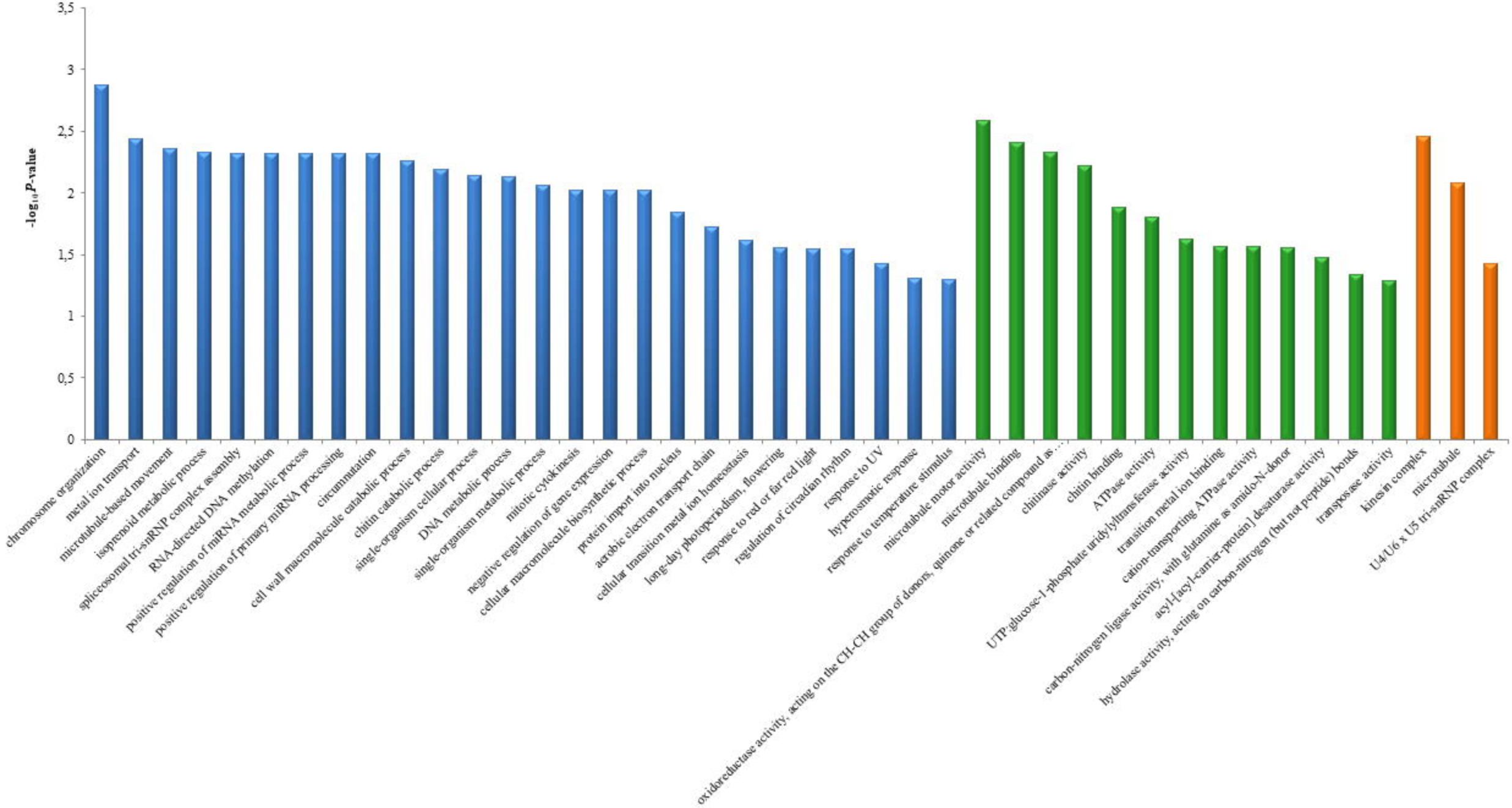
872 **S7 Fig. Functional classification of enriched overexpressed DEGs in the 4×APO vs. 4×SEX**
873 **comparison.** Gene Ontology biological process (blue boxes), Gene Ontology cellular component
874 (yellow boxes), and Gene Ontology molecular function (orange boxes). a) categories enriched in
875 overexpressed transcripts in 4×APO and b) categories enriched in overexpressed transcripts in
876 4×SEX.

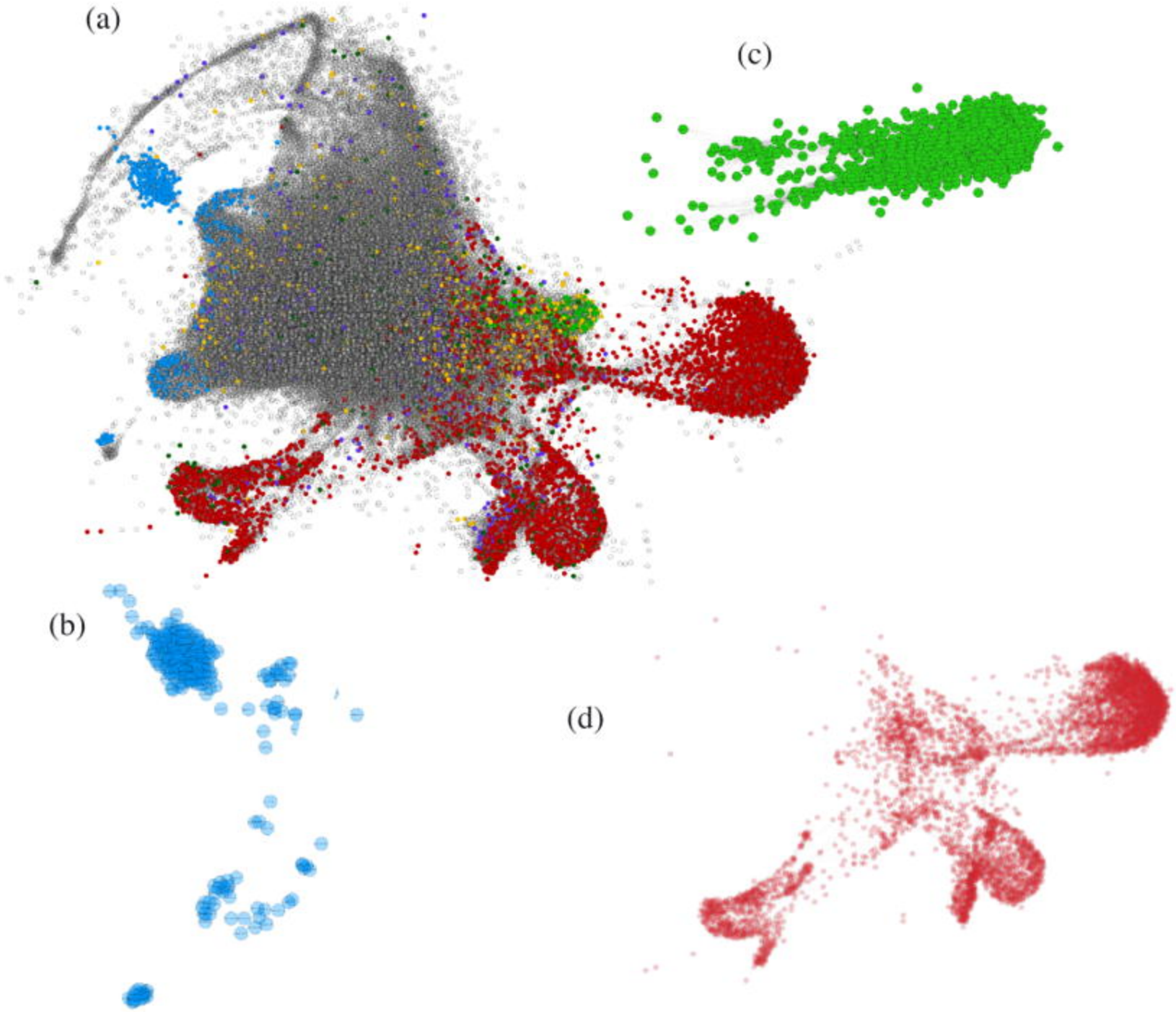
877 **S1 Table. Primer sequences and amplicons of the candidate reference genes evaluated in**
878 **this study.**

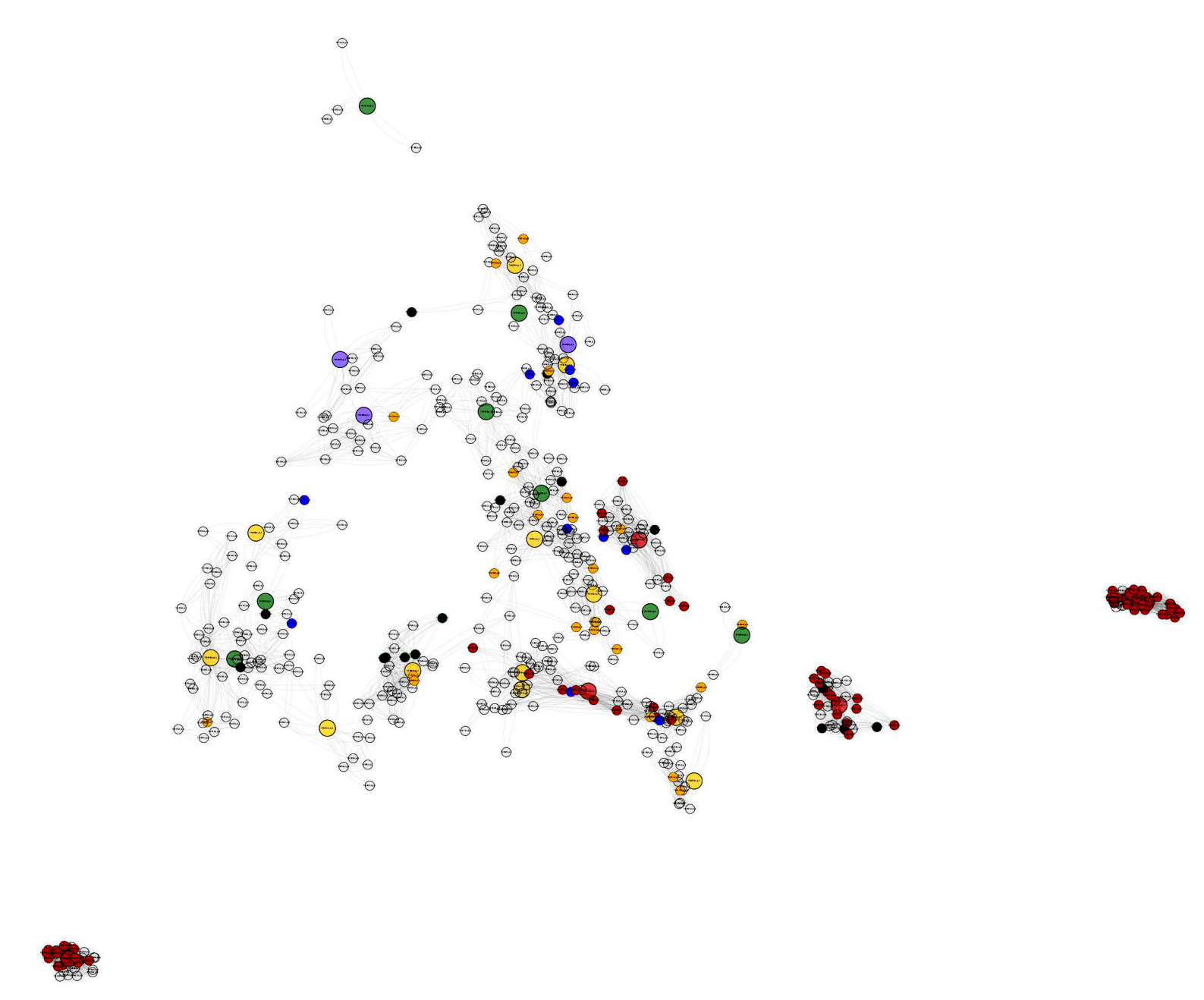
879 **S2 Table. Quantitative RT-PCR primer sequences.**

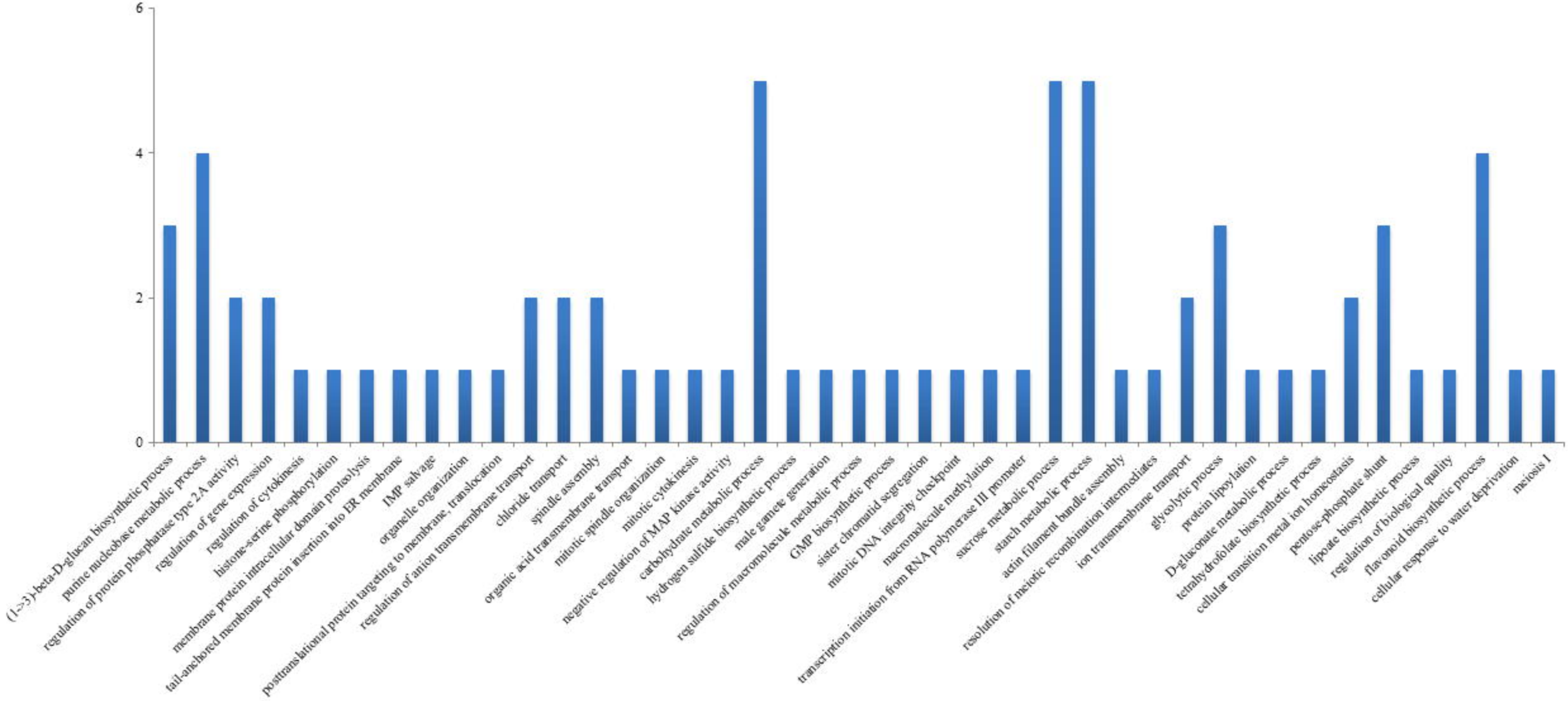
880 **S3 Table. BLAST search results for sequences of *P. notatum* from the literature against the**
881 **assembled transcriptome.**

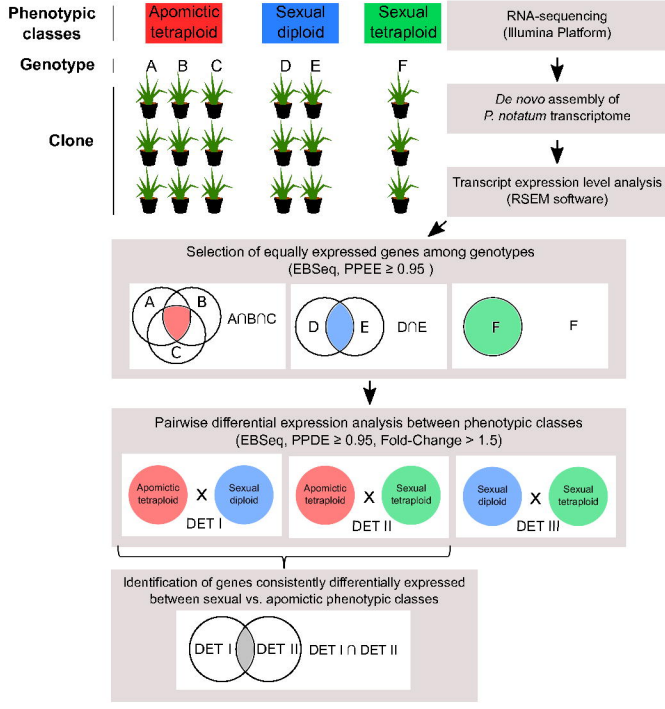



-log10 P-value
3,5
3
2,5
2
1,5
1
0,5
0
chromosome organization
metal ion transport
microtubule-based movement
isoprenoid metabolic process
spliceosomal tri-snRNP complex assembly
RNA-directed DNA methylation
positive regulation of miRNA metabolic process
positive regulation of primary miRNA processing
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
mitotic cytokinesis
negative regulation of MAP kinase activity
carbohydrate metabolic process
hydrogen sulfide biosynthetic process
male gamete generation
regulation of macromolecule metabolic process
GMP biosynthetic process
sister chromatid segregation
mitotic DNA integrity checkpoint
macromolecule methylation
transcription initiation from RNA polymerase III promoter
sucrose metabolic process
starch metabolic process
actin filament bundle assembly
resolution of meiotic recombination intermediates
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


Phenotypic classes
Apomictic tetraploid
Sexual diploid
Sexual tetraploid
RNA-sequencing (Illumina Platform)
Genotype
A B C
D E
F
Clone
De novo assembly of P. notatum transcriptome
Transcript expression level analysis (RSEM software)
Selection of equally expressed genes among genotypes (EBSeq, PPEE ≥ 0.95)
A∩B∩C
D∩E
F
Pairwise differential expression analysis between phenotypic classes (EBSeq, PPDE ≥ 0.95, Fold-Change > 1.5)
Apomictic tetraploid X Sexual diploid
DET I
Apomictic tetraploid X Sexual tetraploid
DET II
Sexual diploid X Sexual tetraploid
DET III
Identification of genes consistently differentially expressed between sexual vs. apomictic phenotypic classes
DET I
DET II
DET I ∩ DET II