Received: 31 July 2019 | Accepted: 13 January 2020

DOI: 10.1111/1755-0998.13137

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

Comparing control options for time-series RNA sequencing experiments in nonmodel organisms: An example from grasses

Fan Qiu | Seton Bachle | Jesse B. Nippert | Mark C. Ungerer

Division of Biology, Kansas State University, Manhattan, KS, USA

Correspondence
Fan Qiu, Division of Biology, Kansas State University, Manhattan, KS 66506, USA.
Email: fqiu@ksu.edu

Funding information
National Science Fundation, Grant/Award Number: NSF DEB 1342787

Abstract

RNA sequencing (RNA-seq) is a widely used approach to investigate gene expression and increasingly is used in time-course studies to characterize transcriptomic changes over time. Two primary options are available as controls in time-course experiments: samples collected at the first sampling time are used as controls (temporal control, TC) and samples collected in parallel at each individual sampling time are used as controls (biological control, BC). While both approaches are used in experimental studies, we know of no analyses performed to date that directly compare effects of control type choices on identifying differentially expressed genes (DEGs) and subsequent functional analysis. In the current study, we compare experimental results using these different control types for time-course RNA-seq drought stress experiments in two wild grass species in the genus *Paspalum*. Our results showed BC assemblies gave a higher number of loci in both species. The number of DEGs increased with increasing stress and then decreased dramatically at the recovery time point using both control types. Expression levels of the same DEGs were highly correlated between control types in both species, ranging from $r = .653$ to $r = .852$. We also observed similar rank orders of shared enriched Gene Ontology term lists using the two different control types. Collectively, our findings suggest similar results in differential gene expression and functional annotation between control types. The ultimate choice of control type will rely on the experimental length and organism type, with labour time and sequencing costs as additional factors to be considered.

KEYWORDS
differentially expressed gene, functional annotation, RNA-seq, time-course experiment, RNA-seq control

1 | INTRODUCTION

The rapid progress of next generation sequencing (NGS) technologies over the past decade has led to higher throughput data generation at a lower cost for RNA sequencing (RNA-seq) studies (Goodwin, McPherson, & McCombie, 2016). RNA-seq has become the gold standard for global gene expression analyses because no prior knowledge of the genomic content is required (Hrdlickova, Toloue, & Tian, 2017) and RNA-seq allows for detection of the whole transcriptome, including known and novel transcripts (Wang, Gerstein, & Snyder, 2009). These advantages make RNA-seq an especially powerful tool for the study of nonmodel organisms whose genomes are not yet sequenced. In such cases, a de novo transcriptome assembly strategy is utilized to assemble a reference transcriptome before downstream differential gene expression analysis (Conesa et al., 2016).

When studying processes such as responses to biotic or abiotic stress, it is often desirable to collect samples in a time-course manner (Bar-Joseph, Gitter, & Simon, 2012) to characterize dynamic responses,

 © 2020 John Wiley & Sons Ltd

such as revealing gene expression changes over time (Tai & Speed, 2005, 2006). Quantifying mRNA levels temporally can illustrate how the transcriptional machinery is affected by the stress and provide insights into the molecular mechanisms that have evolved in response to that stress. Analyses of differential gene expression rely on comparing experimental samples to control samples, and thus there are two possible experimental designs to select controls in a time-course experiment: (a) control samples can consist of experimental individuals sampled before imposing a stress treatment (we refer this as temporal control, TC); and (b) control samples can consist of parallel sets of individuals collected at each separate time point that are not subjected to the stress treatment (we refer to this as biological control, BC). Figure 1 provides a representative schematic of these control options. Both approaches are utilized in experimental studies, for example, TC: (Gao et al., 2015; Min et al., 2016; Rabara et al., 2015); BC: (Fracasso, Trindade, & Amaducci, 2016; Garg et al., 2016; Meyer et al., 2014). Use of TCs requires fewer samples in total, which can save on sequencing costs and/or enable increased sampling efforts over the time-course experiment. For some organisms, such as plants, where repeated tissue sampling from the same individual is possible, this approach also allows for control of genotypic differences among individuals. A downside of the TC approach is that it may not properly control for extraneous environmental variation that may be present across the sampling time points that is not associated with the specific stress treatment under investigation. While use of BCs can eliminate this concern and better control for natural development/change in the experimental system over time, a drawback of the BC approach is higher potential cost and/or less intensive sampling over the time-course.

A sound experimental design is essential for success in RNA-seq studies (Conesa et al., 2016; Robles et al., 2012; Williams, Thomas, Wyman, & Holloway, 2014). Previous technical studies have examined issues such as library type, sequencing depth, and the number of biological replicates and their effects on RNA-seq experiments (Corley, MacKenzie, Beverdam, Roddam, & Wilkins, 2017; Liu, Zhou, & White, 2014). Previous studies have not, however, empirically addressed the specific question raised here: how does selection of control type (i.e., TC vs. BC) affect corresponding results in gene expression changes during time-course experiments? In this paper, we utilize time course RNA-seq experiments to characterize gene expression changes using different control types in two nonmodel grass species in the genus *Paspalum* during simulated drought stress. Drought is among the most widely experienced abiotic stressors for plants and impacts multiple aspects of growth, development and physiology (Jogaiah, Govind, & Tran, 2013; Valliyodan & Nguyen, 2006). Plants have evolved various dynamic responses at the physiological, biochemical and molecular levels to handle drought stress (Huber & Bauerle, 2016).

We evaluate how the use of specific control types (i.e., TC vs. BC) influences overall de novo transcriptome assembly, differential gene expression analyses and functional enrichment tests. Our results indicate similar patterns in transcriptome assembly, detection and fold-change levels of differentially expressed genes (DEGs), and functional categories related to drought stress in the time-course experiment. Some interesting differences also were found, however, and the pros and cons of the two different approaches are discussed. These experiments additionally provide insights into transcriptional stress responses of these two wild grass species.

2 | MATERIALS AND METHODS

2.1 | Plant materials, growing conditions and dry-down

Seeds of *Paspalum notatum* (Accession no. 241878, tetraploid population) (Fachinetto, Dall'Agnol, Schifino-Wittmann, Simioni, & Avila, 2018) and *P. juergensii* (Accession no. 508779, diploid) (Williams, Williamson, & Real, 2011) were acquired from the United States Department of Agriculture (USDA) National Plant Germplasm System (https://www.ars-grin.gov/npgs/). Seeds were germinated on moist filter paper in Petri dishes and then transferred to 10-cm plastic pots containing a mixture (2:1) of Metro-mix 360 and all-purpose sand with soil collected from a nearby tallgrass prairie added at 10% by volume to provide a source of natural soil microbial diversity. This prairie soil was untreated before addition. Plants were grown in the Kansas State University greenhouse facility under a daily light cycle of 16 hr light/8 hr dark, with supplemental lighting consisting of six lighting fixtures using 400-W high-pressure sodium bulbs. Watering was conducted daily or as needed. Plants were grown for 21 days before initiation of the dry-down experiment and thus were well established at the time of the experiment with multiple leaves per plant. Fourteen individuals (seven drought-treated and seven BC) per species were well watered on day 0 of the experimental dry-down, and water was withheld for seven drought-treated plants until photosynthetic rates reached 0 μmol m^{-2} s^{-1}, after which rewatering was performed and recovery was monitored for 2 days. Photosynthesis measures were made daily (dry-down: days 1–5; recovery: days 6–7 for *P. notatum* and dry-down: days 1–4; recovery: days 5–6 for *P. juergensii*) between 11:00 and 14:00 CDT with a LI-6400 system (LiCOR) equipped with an LED light source (light intensity was maintained at 2,000 μmol m^{-2} s^{-1}, CO_2 concentration at 400 ppm, and relative humidity at ambient levels, ~40%–60%). Measurements from the LI-6400 were logged after gas exchange measurements were stable for 2 min. For each species, six individual plants (three drought-stressed and three control) were selected for transcriptomic analyses. Leaf samples were harvested from these individuals on days 1, 3, 4, 5 and 7 for *P. notatum* and on days 1, 3, 4 and 6 for *P. juergensii*. A total of 54 RNA samples (*P. notatum*: 2 treatments × 3 replicates × 5 time points = 30; *P. juergensii*: 2 treatments × 3 replicates × 4 time points = 24) were thus generated for this study.

2.2 | RNA extraction, library construction and sequencing

Harvested leaves were flash-frozen in liquid nitrogen and stored at − 70°C before RNA extraction. Total RNA was extracted using TRIzol reagent (Invitrogen) following the manufacturer instructions.

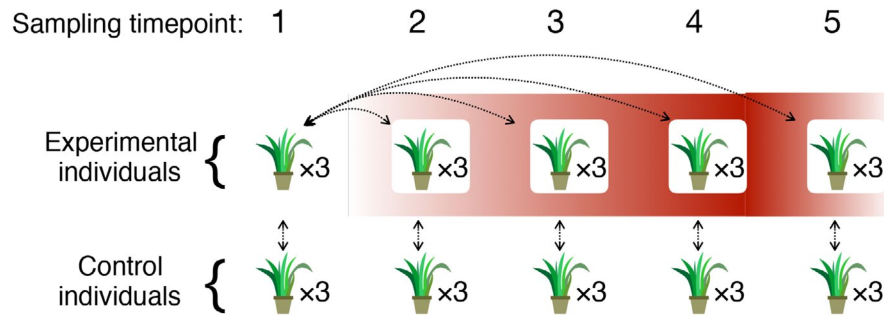

FIGURE 1 Representative schematic of the control options evaluated in the current study. The illustration depicts sampling at five time points where experimental individuals (top row) experience increasing stress (red shading) during time points 2–4 and an early stage of recovery from stress at time point 5. The temporal control (TC) approach uses experimental individuals from time point 1 as controls for tests of differential gene expression in experimental individuals sampled at time points 2–5 (curved horizontal double sided arrows). The biological control (BC) approach uses parallel sets of individuals sampled at the same time points but not subjected to stress (vertical double sided arrows). Diagram depicts three replicates sampled for all experimental and control individuals per time point [Colour figure can be viewed at wileyonlinelibrary.com]

Total RNA was purified to avoid genomic DNA contamination using a RNeasy Mini Kit and an on-column DNAse I digestion (Qiagen). The quality and quantity of the total RNA were examined using Agilent Tape Station (Agilent Technologies) and a Qubit fluorometer (Thermo Fisher Scientific). The average RNA integrity number (RIN) of all samples was 6.9. One microgram of total RNA per sample was utilized for library preparation and sequencing on an Illumina HiSeq2500 platform, generating 2 × 100-bp paired-end reads. Library preparation was performed following the Illumina TruSeq Stranded mRNA Library Prep Kit (Illumina). Library construction and sequencing were performed at the University of Kansas Genome Sequencing Core Facility (http://gsc.drupal.ku.edu/). The average number of reads obtained per sample was 29,953,855 and 31,356,449, respectively, for *P. notatum* and *P. juergensii*.

2.3 | Sequence assembly and differential gene expression

Raw sequence reads were trimmed and filtered using TRIMMOMATIC version 0.35 (Bolger, Lohse, & Usadel, 2014) according to the following criteria: (a) adapters and barcodes removed, (b) reads < 40 bases removed, (c) bases trimmed from read ends if quality < 30, and (d) read ends trimmed while mean quality < 30 in a 5-bp sliding window. The trimmed reads from samples from all time points were combined and used for de novo assembly, which was performed with TRINITY version 2.2.0 (Grabherr et al., 2011) with default normalization and other parameters except to keep contigs with length ≥ 300 bp. Two assemblies were produced to mimic experiments utilizing two different control types: (a) an assembly only based on drought-treated samples (we refer this as TC assembly); and (b) an assembly based on both drought-treated and BC samples (we refer this as BC assembly). In total, 765,233,012 reads versus 395,284,388 reads, respectively, were used in BC and TC assemblies in *P. notatum*, while 642,881,124 reads versus 314,066,636 reads, respectively, were used in *P. juergensii*. Following assembly, CD-HIT version 4.6.8 (Fu, Niu, Zhu, Wu, & Li, 2012) was used to obtain distinct sequences. The following parameters were used in CD-HIT analysis: (a) sequence identity threshold: 0.95, and (b) alignment coverage for the shorter sequence: 0.9. These nonredundant sequences were used for the downstream analysis. We also assessed the completeness of BC and TC assemblies using Benchmarking of Universal Single-Copy Orthologs (BUSCO) (Simao, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) with the Embryophyta database from ORTHODB version 9 (Zdobnov et al., 2017).

The nonredundant sequences from de novo assemblies were used as references for read mapping. Trimmed reads were aligned to nonredundant TC and BC assemblies using BOWTIE version 1.1.2 (Langmead, Trapnell, Pop, & Salzberg, 2009) and transcript and gene abundance was estimated with RSEM version 1.2.28 (Li & Dewey, 2011) and PERL script align_and_estimate_abundance.pl from TRINITY (Haas et al., 2013). EDGER version 3.3 (Robinson, McCarthy, & Smyth, 2010) was used to determine DEGs with a false discovery rate (FDR) <0.05 and $\log_2$ fold-change of 2. The weighted Trimmed Mean of M-values (TMM) approach (Robinson & Oshlack, 2010) was used to calculate library size normalization factors. Three biological replicates were treated as independent samples in EDGER analysis. To compare gene expression levels for the same DEGs detected using the two different control types, we first performed all-to-all blast between the BC and TC assemblies. We assumed that two genes across assemblies were the same if they were reciprocal best hits by BLAST with an e-value cutoff of 10^{-5}. Pearson's r was calculated based on the absolute values of the $\log_2$-transformed fold-change for DEGs detected by the two different control types at each sampling time point.

2.4 | Functional annotation and enrichment test

The final nonredundant contigs (i.e., those of ≥300 bp) from TC and BC assemblies in each species were compared to the entire UNIPROT and NCBI nonredundant (nr) databases with the BLASTX tool of BLAST version 2.2.31 (Altschul, Gish, Miller, Myers, & Lipman, 1990) and an e-value cutoff of 10^{-5}. Contigs with significant BLAST hits were annotated with the Gene Ontology (GO) terms of their top matches using

BLAST2GO version 3.0.6 (Conesa et al., 2005). GO enrichment analyses were then performed for DEGs at each time point with Fisher's exact test of BLAST2GO. The significance level for these tests was set to $\alpha = 0.05$ after correcting for the FDR due to multiple testing with the Benjamini and Hochberg method (Benjamini & Hochberg, 1995).

To evaluate the similarity of the shared enriched GO terms related to drought stress recovered at each time point based on the two different controls, we first ranked the GO terms based on FDR scores. We then evaluated similarity of rank order of the GO term lists using a permutation approach implemented in the program ORDEREDLIST (Lottaz, Yang, Scheid, & Spang, 2006).

3 | RESULTS

3.1 | Physiological results

The dry-down manipulation negatively impacted leaf-level physiological processes of both species. For *Paspalum notatum*, photosynthetic rates remained stable and indistinguishable from control plants for 3 days, after which it decreased by 37.8% (Day 4) and 64.8% (Day 5) relative to the mean Day 1 value for experimental plants and was reduced by 44.5% (Day 4) and 72.3% (Day 5) compared to control plants measured on the same days (Figure 2a). Rewatering of *P. notatum* plants was performed following physiological measurements and tissue sampling on Day 5, with photosynthetic rate increasing sharply and reaching 83.4% of the mean Day 1 value for experimental plants by Day 7, but with values still lower than corresponding control plants (Figure 2a). For *P. juergensii*, a slight decline in photosynthetic rate was evident by Day 3, with a sharp decrease of 98.0% relative to the mean Day 1 value for experimental plants by Day 4. The difference between experimental and control plants observed on Day 2 appears attributable to a mean increase in photosynthetic rate of control plants. Rewatering of *P. juergensii* plants was performed following physiological measurements and tissue sampling on Day 4, with photosynthetic rate increasing sharply and reaching 77.7% of the mean Day 1 value for experimental plants by Day 6. As observed for *P. notatum*, photosynthetic rates for *P. juergensii* experimental plants remained lower than corresponding control plants over the 2-day recovery period (Figure 2b).

3.2 | RNA-seq data sets and de novo assemblies

Transcriptome sequencing generated 23.7–38.6 million raw reads per sample (Table S1). After trimming, the number of reads was reduced by ~15% on average across samples and read lengths were reduced from their original size of 101 bases to a mean of 94.7–96.2 bases (Table S1). These trimmed, high-quality reads were used for de novo assembly in TRINITY. Two separate assemblies were performed to mimic experiments with and without biological controls for each sampling time point. Assemblies mimicking experiments using BC were conducted using sequence data from all experimental and control plants whereas assemblies mimicking experiments using TC were performed using sequence data from experimental plants only. The former, referred to here as "BC assemblies," were thus based on approximately twice the amount of sequence data used for the "TC assemblies": for *P. notatum*, 765,233,012 reads versus 395,284,388 reads, respectively; and for *P. juergensii*, 642,881,124 versus 314,066,636 reads, respectively. The completeness of transcriptome assemblies was assessed using BUSCO (https://busco.ezlab.org/). The numbers of complete and missing BUSCOs identified in our study (Table S2) were similar to those found in five other grass species (Schubert, Gronvold, Sandve, Hvidsten, & Fjellheim, 2019), indicating that the transcriptome assemblies were well represented and of high quality. The BC assemblies resulted in a higher number of loci detected in both species and corresponding higher numbers of loci with BLAST hits and annotation (Table 1). In sets consisting of the top 20, 50, 100, 200 and 300 annotated GO terms based on the number of genes assigned, the concordance between the two assemblies in each species was highly similar, ranging between 90% and 96% (Table 2).

3.3 | Differential expression analysis based on different control types

Trimmed reads were mapped to the respective sets of nonredundant loci to quantify gene expression changes using different controls (BC and TC) during dry-down on Days 3, 4 and 5 and at recovery (Day 7) in *P. notatum* and during dry-down on Days 3 and 4 and at recovery (Day 6) in *P. juergensii*. Changes in gene expression were additionally determined on Day 1 for the BC dataset. On days omitted from DEG analyses, photosynthetic rates remained largely unchanged (Figure 2). For both species, the number of DEGs increased with increasing stress and then decreased sharply at the recovery time point (Figure 2c,d). For *P. notatum*, more DEGs were detected using TC on Days 3, 4 and 5, whereas on Day 7, more DEGs were detected using BC (Figure 2c). The difference in the number of DEGs detected between BC and TC ranged from 319 on Day 7 to 880 on Day 4. For *P. juergensii*, more DEGs were detected using TC on Days 4 and 6, whereas on Day 3, more DEGs were detected using BC (Figure 2d). The difference in the number of DEGs detected between BC and TC ranged from 107 on Day 3 to 730 on Day 4. Analyses using BC revealed 114 and 122 DEGs at the earliest time point (Day 1) for *P. notatum* and *P. juergensii*, respectively (Figure 2c,d). To explore these data further, we conducted cross-comparison analyses by aligning the TC and BC data to the BC and TC assemblies, respectively. Similar numbers of DEGs were detected at each time point in these cross-comparisons versus where TC and BC approaches were conducted independently (see Figure S1 and Figure 2).

To compare variation in expression levels of DEGs detected using the two different controls, we used conservative criteria to identify the same loci across assemblies that were differentially expressed using both controls (see Methods) and we then compared relative fold-change values. For *P. notatum*, we identified between 47 (Day 3) and 5,542 (Day 5) such loci and for *P. juergensii* between 134 (Day 6) and 5,711 (Day 4) such loci (Figure 3). Fold-changes for these sets

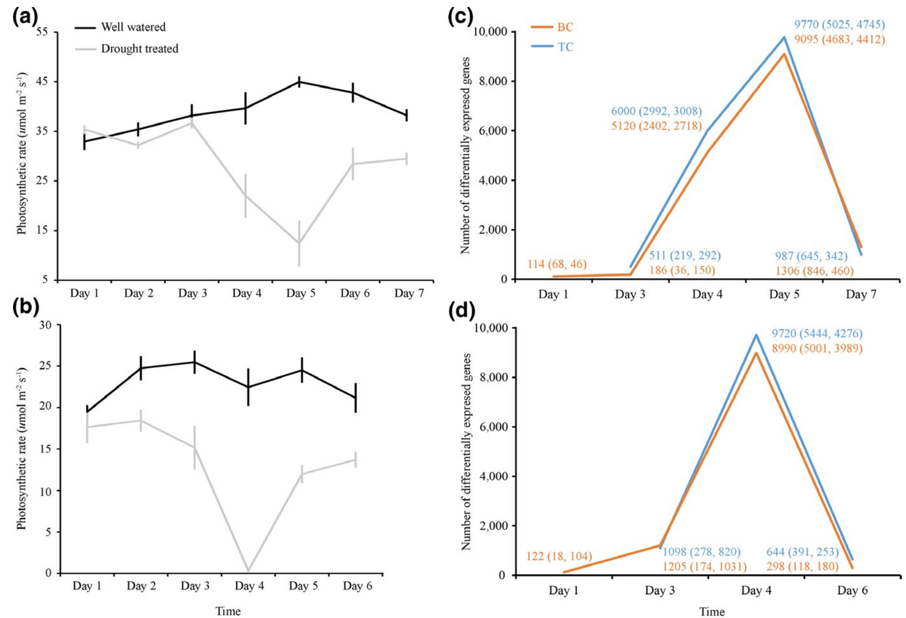

FIGURE 2 Physiological response (photosynthesis) in *Paspalum notatum* (a) and *Paspalum juergensii* (b) and changes of differential expressed genes (DEGs) during drought stress and water recovery period in *P. notatum* (c) and *P. juergensii* (d). In (a) and (b), bar represents standard error of the mean. In (c) and (d), among the three numbers at each time point, the first number represents the total number of DEGs, the second the number of down-regulated genes, the third the number of up-regulated genes [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Assembly and annotation statistics

	Paspalum notatum		*Paspalum juergensii*	
	BC assembly	TC assembly	BC assembly	TC assembly
Number of loci	96,749	85,587	67,641	63,739
Number of loci with BLASTX hit	46,186 (47.7%)	41,366 (48.3%)	35,033 (51.8%)	33,697 (52.9%)
Number of loci with annotation	37,356 (38.6%)	33,311 (38.9%)	28,700 (42.4%)	27,522 (43.2%)

Abbreviations: BC, biological control; TC, temporal control.

TABLE 2 Concordance for GO terms found between BC and TC assemblies

	Paspalum notatum			*Paspalum juergensii*		
GO term category	BP	MF	CC	BP	MF	CC
Top 20 GO terms	0.950	0.900	0.900	0.950	0.900	0.950
Top 50 GO terms	0.940	0.920	0.960	0.960	0.940	0.940
Top 100 GO terms	0.930	0.930	0.950	0.930	0.940	0.960
Top 200 GO terms	0.915	0.925	0.945	0.945	0.930	0.935
Top 300 GO terms	0.923	0.923	0.943	0.947	0.927	0.930

Abbreviations: BC, biological control; BP, biological process; CC, cellular component; GO, Gene Ontology; MF, molecular function; TC, temporal control.

of DEGs were highly correlated, ranging from $r = .653$ to $r = .852$ for *P. notatum* (Figure 3a) and from $r = .774$ to $r = .803$ for *P. juergensii* (Figure 3b). A small number of DEGs (~0.1% in *P. notatum* and ~ 0.3% in *P. juergensii*) were found to be up-regulated based on one control and down-regulated based on the other. These are depicted in Figure 3 in the lower right and upper left quadrants of panels.

3.4 | GO enrichment tests and concordance analysis of enriched GO terms

GO enrichment tests were used to identify enriched functional categories of up-regulated and down-regulated DEGs at different sampling time points detected using the different controls. GO terms were detected at most time points based on both controls, typically with similar numbers of terms detected (Table S3) and considerable overlap of those terms, with the number of shared terms typically greater than the number of unique terms found using either of the two controls (Figure 4). Examining shared GO terms only, similarities in rank order based on FDR were tested using the ORDEREDLIST BIOCONDUCTOR package (Lottaz et al., 2006). With one exception (i.e., for up-regulated genes on Day 7 in *P. notatum*), rank orders of GO term lists using the two different control types were highly similar (Figure 5).

While GO enrichment results were largely similar across control types, exceptions to this general pattern include lack of GO terms detected for BC on Day 3 (*P. notatum* and *P. juergensii*, down-regulated

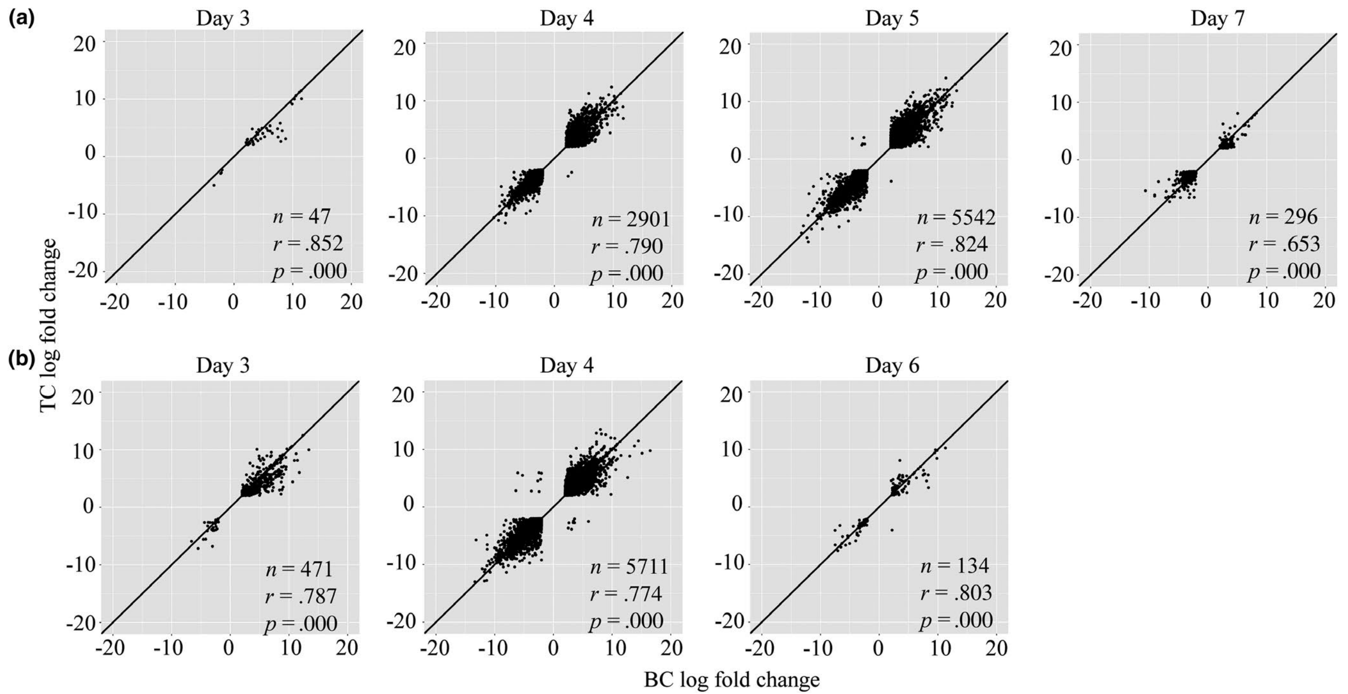

FIGURE 3 Fold-change comparison of DEGs using TC versus BC for four sampling time points for *Paspalum notatom* (a) and three sampling time points for *Paspalum juergensii* (b). *n* represents the number of DEGs analysed (see Methods for selection criteria) and *r* represents Pearson's correlation coefficient based on the absolute fold-change value (|fold-change|). Black diagonal lines indicate one-to-one ratios

genes) and BC on Day 6 (*P. juergensii*, down-regulated genes), and asymmetry in the number of detected GO terms on Day 7 for *P. notatum*, with considerably more GO terms detected with BC versus TC for both up- and down-regulated genes (Figure 4; Table S3). A complete lack of detected GO terms is probably associated with small numbers of DEGs detected for those specific time points/classifications (i.e., 36, 174 and 118 DEGs).

Among the enriched GO terms, common drought-related terms typically recovered in drought stress experiments (Fracasso et al., 2016; Wu et al., 2014; Zhang, Lei, Lai, Zhao, & Song, 2018) were recovered here using both control types. For up-regulated genes, "response to water deprivation (GO:0009414)" and "response to water (GO:0009415)" were over-represented on all days before recovery in both species. "Response to stress (GO:0006950)," "response to abiotic stress (GO:0009628)," "response to abscisic acid (GO:0009737)," "response to hormone (GO:0009725)," "response to oxidative stress (GO:0006979)" and "response to jasmonic acid (GO:0009753)" were over-represented on Day 4 and Day 5 in *P. notatum* and Day 3 and Day 4 in *P. juergensii*. After rewatering, "thylakoid (GO:0009579)," "chloroplast thylakoid (GO:0009534)," and "photosynthetic membrane (GO:0034357)" were over-represented in both species. Moreover, "photosystem (GO:0009521)," "photosynthesis (GO:0019684)" and "photosynthesis, light reaction (GO:0019684)" were over-represented in *P. juergensii*.

For down-regulated genes, "chloroplast (GO:0009507)," "thylakoid (GO:0009579)," "photosynthetic membrane (GO:0034357)," "photosynthesis (GO:0015979)" and "photosynthesis, dark reaction (GO:0019685)" were over-represented on Day 4 and Day 5 in *P. notatum* and Day 4 in *P. juergensii*. "Photosynthesis, light reaction (GO:0019684)" was over-represented on Day 5 in *P. notatum* and Day 4 in *P. juergensii*. After recovery, "response to stress (GO:0006950)," "response to abiotic stress (GO:0009628)," "response to abscisic acid (GO:0009737)," "response to water deprivation (GO:0009414)" and "response to water (GO:0009415)" were over-represented in *P. notatum*.

4 | DISCUSSION

With rapid developments in sequencing technology in the past decade, RNA-seq has become the standard method for transcriptome analysis and identification of DEGs across samples and/or treatments (Hrdlickova et al., 2017). Time course gene expression experiments characterize gene expression changes across multiple time points. For example, during organismal development, different sets of genes are activated or suppressed at different stages. Time course gene expression experiments can provide a better understanding of the genetic control of various developmental transitions (Rauwerda et al., 2017). They also provide a powerful tool for studying perturbation-response dynamics, where temporal gene expression changes can reveal common and unique features of organismal responses to environmental perturbations (Gasch et al., 2000).

Two main options are available for the selection of control samples for identifying DEGs in such experiment: (a) experimental samples at the first sampling time point can be used as controls (temporal

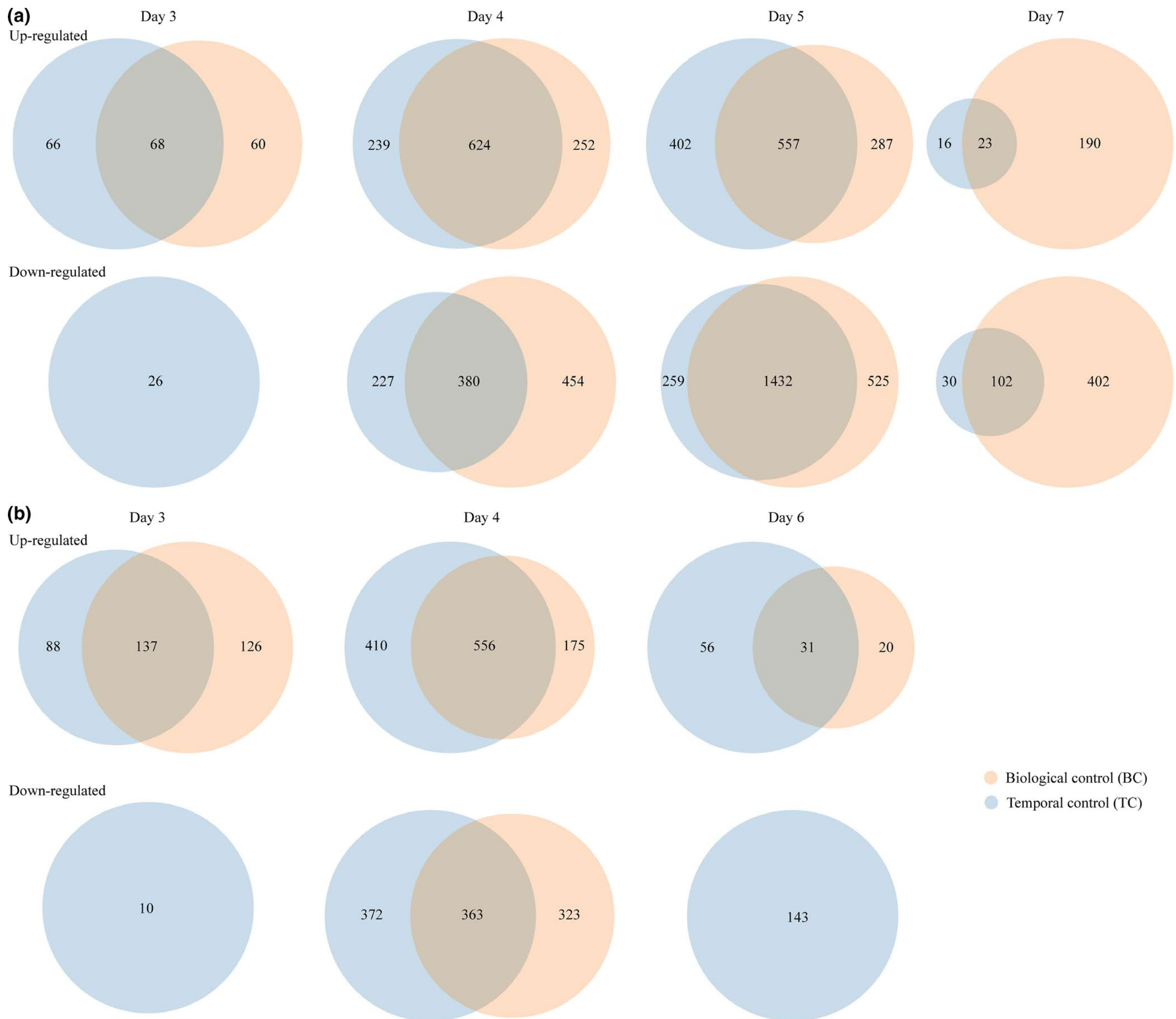

FIGURE 4 Venn diagram of the enriched GO terms for up- and down-regulated genes in *Paspalum notatum* (a) and *Paspalum juergensii* (b) [Colour figure can be viewed at wileyonlinelibrary.com]

control, TC), with all other time points compared to the TC, and (b) parallel sets of unmanipulated samples can be used as controls (biological control, BC), with all experimental samples compared to corresponding unmanipulated controls collected at the same time point. Use of BC may better control for gene expression changes attributable to environmental fluctuations over the experimental time-course, but can lead to significant extra cost and labour in experiments requiring sampling at multiple time points. While TC cannot account for influences of extraneous environmental fluctuations over the experimental time-course, this approach may be preferable and more cost-effective under circumstances when more frequent sampling is desired or when higher numbers of replicates are necessary to increase statistical power for detecting DEGs (Corley et al., 2017; Liu et al., 2014). Moreover, for experimental systems such as plants where tissue can be sampled repeatedly from the same individuals, TC may better control for DEGs attributable to genotypic differences between experimental and control individuals. Indeed, a previous study showed significant differences in drought tolerance for individual genotypes within accessions (Cui, Wang, Wang, & Jiang, 2015). The use of TC also may be more appropriate when analyses of transcriptomic changes are coupled with labour-intensive measures of physiological responses (e.g., gas exchange measurements in plants), where physiological responses to stress are evaluated in reference to starting values (Osborne, Wythe, Ibrahim, Gilbert, & Ripley, 2008; Phillips, Oren, Licata, & Linder, 2004; Prasad, Pisipati, Ristic, Bukovnik, & Fritz, 2008; Seelig, Wolter, & Schroder, 2015).

In the current study, we examined how use of different control type (i.e., TC vs. BC) affected the overall results of differential gene expression and subsequent gene function analysis during time-course experiments. We utilized data from a time-course RNA-seq experiment to characterize gene expression changes in two nonmodel grass species, *Paspalum notatum* and *P. juergensii*, during drought

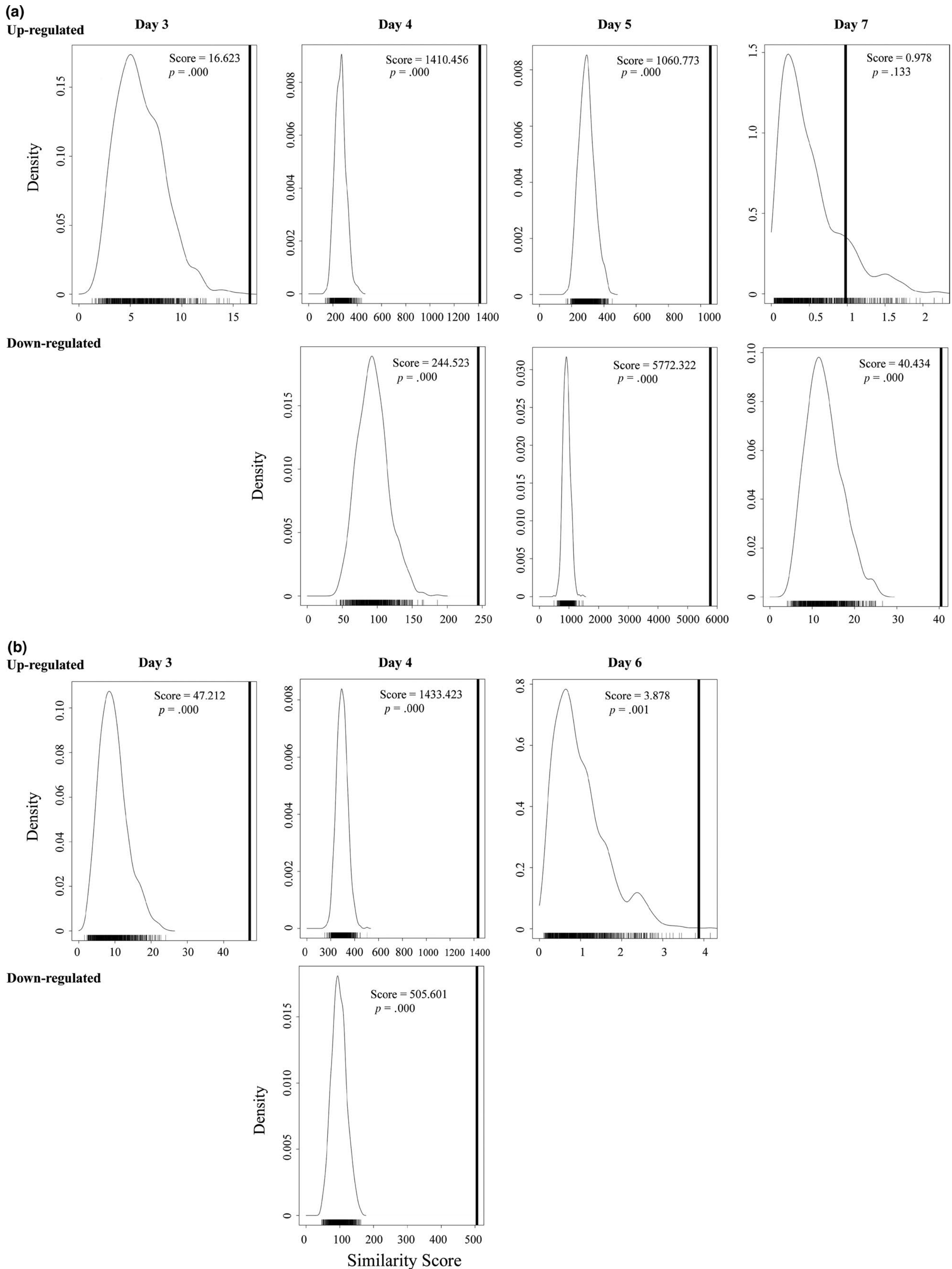

FIGURE 5 Similarity score between the two ranked enriched GO term lists from up- and down-regulated DEGs in *Paspalum notatum* (a) and *Paspalum juergensii* (b). Black vertical line represents the empirical similarity score. Significance was determined by comparing the empirical score to a distribution derived from 1,000 permutations where rank order was randomized and similarity remeasured

stress and recovery. Two assemblies were produced corresponding to these alternative experimental designs: the first was based only on drought-treated samples (TC assembly) and the second based on both drought-treated and nondrought-treated (control) samples (BC assembly). We compared properties of the two de novo transcriptome assemblies and examined temporal aspects of differential gene expression and functional enrichment based on the two control options.

BC and TC assemblies produced different numbers of loci, with the BC assemblies producing ~6% (*P. juergensii*) to ~13% (*P. notatum*) more unigenes versus the TC assemblies. Similar observations (i.e., increased assembled loci with higher sequencing depth) have been found in other nonmodel organisms lacking a reference genome (Qiu et al., 2015; Riesgo et al., 2012; Zhao et al., 2011). One possibility is that lower sequencing depth resulted in incomplete transcriptome sampling due to an inability to detect rare and lowly expressed transcripts. Indeed, studies have shown that up to 500 million reads may be required to cover the full sequence diversity, including rare and lowly expressed transcripts (Blencowe, Ahmad, & Lee, 2009; Toung, Morley, Li, & Cheung, 2011). Functional annotation analysis based on total assembled genes revealed a high concordance (90%–96%) between TC and BC assemblies for both species (Table 2), even though the GO terms in each set were not necessarily ranked in the same order. These results indicate overall similarities in gene functions were recovered in both assemblies despite different numbers of identified unigenes.

Patterns of temporal gene expression changes were similar between BC and TC control options: the number of DEGs increased in a similar fashion with increasing stress and decreased sharply following rewatering. While trends were generally similar, a higher number of DEGs was detected in five out of seven total time points from both species when using TC control (Figure 2). This may be because using TC control does not account for expression changes in response to extraneous environmental fluctuations. Use of the BC control revealed 114 (*P. notatum*) and 122 (*P. juergensii*) DEGs on Day 1 of the experiment (Figure 2). Indeed, GO terms related to drought stress, such as “oxidoreductase activity (GO:0016491)” and “response to abiotic stimulus (GO:0009628),” were enriched on Day 1 in *P. juergensii*, but not in *P. notatum*, indicating early drought responses in *P. juergensii*. These terms were recovered in both species using either controls when stress was increased.

We further examined fold-change values to determine if the same DEGs identified by both controls have similar detected expression levels. DEGs were considered as the same gene if they were reciprocal best hits across assemblies. Notably, we confirmed highly significantly positive correlations in fold-change of DEGs detected by the two different control types (Figure 3). While the DEGs used in this comparison represent only subsets of the total numbers of DEGs detected (selected based on reciprocal best hit criteria), they are probably an unbiased sample, and the high correlation coefficients in these analyses indicate similar fold-change estimates between the two control types. Interestingly, a small number of DEGs examined (<0.3%) exhibited significant up-regulation based on one control type and significant down-regulation based on the other. No significant enriched GO terms were found for these genes. Three of these genes, however, were classified as “response to stimulus (GO:0050896)”: with one involved in jasmonic acid metabolic process, one involved in oxidoreductase activity and one involved in response to salt stress.

Enriched GO terms were found for both up- and down-regulated genes for all time points and for both control types, with the exception that no terms were detected for down-regulated genes at three time points based on the BC: Day 3 for *P. notatum* and Days 3 and 6 for *P. juergensii* (Figure 4). Lack of detected GO terms may be associated with small numbers of DEGs detected for those specific time points/categories (i.e., 36, 174 and 118 DEGs, see Figure 2c,d). For most time points, however, there was considerable overlap in enriched GO terms based on the different control types, and the rank order of these shared GO terms based on FDR was highly similar (Figure 5). For up-regulated genes, for example, shared terms included essential drought-related terms such as “response to water deprivation (GO:0009414),” “response to abiotic stress (GO:0009628)” and “response to abscisic acid (GO:0009737)” (Fracasso et al., 2016; Wu et al., 2014; Zhang et al., 2018). For nonshared GO terms, when using the TC control type, “secondary metabolic process (GO: 0019748)” was over-represented on Day 3, Day 4, and Day 5 in *P. notatum* and “response to external stimulus (GO:0009605)” was over-represented on Day 3 and Day 4 in *P. juergensii*. This may be due to the fact that the TC control does not account for environmental perturbations through time. When using the BC control type, GO terms related to biosynthetic process were enriched. For example, “lipid biosynthetic process (GO:0008610)” and “alcohol biosynthetic process (GO:0046165)” were over-represented on Day 4 and Day 5 in *P. notatum*, while “cellular carbohydrate biosynthetic process (GO:0034637)” was over-represented on Day 3 and Day 4 in *P. juergensii*. Together, these results indicated high concordance in detecting expression levels of DEGs and drought-stress-related GO terms across control types.

5 | CONCLUSIONS

Comparison of transcriptome assemblies, patterns and magnitudes of differential gene expression and downstream GO enrichment indicate generally similar results between the two different control types for the experiments conducted here. Use of the TC resulted in detection of slightly more DEGs across more time points while use of BC resulted in a failure to detect enriched GO terms for down-regulated genes for some time points where the total numbers of DEGs were low. While similar results across independent experiments suggest that either control type could be used effectively, a clear caveat is the relatively short duration of these time-course studies (four or five sampling time points across 6–7 days). For longer time-course experiments where extraneous environmental influences could be more pronounced, TC and BC could yield more dissimilar results. Multiple factors ultimately should be considered when making a decision on which control type to choose, including time-course length, frequency of sampling and desired replication, the magnitude of potential extraneous environmental perturbations, and of course the tradeoffs of these factors with labour and sequencing costs. A

TC approach may be especially advantageous when attempting to control for genotypic differences among individuals, especially in systems such as plants where the same individuals can be sampled repeatedly over time.

ACKNOWLEDGEMENTS

The project was funded by NSF DEB 1342787.

AUTHOR CONTRIBUTIONS

F.Q., J.B.N. and M.C.U. planned and designed the research. F.Q., S.B., J.B.N. and M.C.U. performed experiments, F.Q. and S.B. analysed the data, F.Q. and M.C.U. wrote the manuscript. S.B. and J.B.N. provided comments on drafts of the manuscript. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

Raw sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) database under BioProject PRJNA557657 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA557657).

ORCID

Fan Qiu https://orcid.org/0000-0003-3618-7631

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bar-Joseph, Z., Gitter, A., & Simon, I. (2012). Studying and modelling dynamic biological processes using time-series gene expression data. *Nature Reviews Genetics*, *13*(8), 552–564. https://doi.org/10.1038/nrg3244

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blencowe, B. J., Ahmad, S., & Lee, L. J. (2009). Current-generation high-throughput sequencing: Deepening insights into mammalian transcriptomes. *Genes & Development*, *23*(12), 1379–1386. https://doi.org/10.1101/gad.1788009

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., … Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*, 13. https://doi.org/10.1186/s13059-016-0881-8

Corley, S. M., MacKenzie, K. L., Beverdam, A., Roddam, L. F., & Wilkins, M. R. (2017). Differentially expressed genes from RNA-Seq and functional enrichment results are affected by the choice of single-end versus paired-end reads and stranded versus non-stranded protocols. *BMC Genomics*, *18*(1), 399. https://doi.org/10.1186/s12864-017-3797-0

Cui, Y., Wang, J. S., Wang, X. C., & Jiang, Y. W. (2015). Phenotypic and genotypic diversity for drought tolerance among and within perennial ryegrass accessions. *HortScience*, *50*(8), 1148–1154. https://doi.org/10.21273/HORTSCI.50.8.1148

Fachinetto, J. M., Dall'Agnol, M., Schifino-Wittmann, M. T., Simioni, C., & Avila, M. R. (2018). New wild diploids in *Paspalum notatum* Flugge (Poaceae): Potential accessions for use in breeding. *Crop Breeding and Applied Biotechnology*, *18*(4), 432–436. https://doi.org/10.1590/1984-70332018v18n4n63

Fracasso, A., Trindade, L. M., & Amaducci, S. (2016). Drought stress tolerance strategies revealed by RNA-Seq in two sorghum genotypes with contrasting WUE. *BMC Plant Biology*, *16*(1), 115. https://doi.org/10.1186/s12870-016-0800-x

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Gao, F., Wang, J., Wei, S., Li, Z., Wang, N., Li, H., … Zhang, F. (2015). Transcriptomic analysis of drought stress responses in *Ammopiptanthus mongolicus* leaves using the RNA-seq technique. *PLoS ONE*, *10*(4), e0124382. https://doi.org/10.1371/journal.pone.0124382

Garg, R., Shankar, R., Thakkar, B., Kudapa, H., Krishnamurthy, L., Mantri, N., … Jain, M. (2016). Transcriptome analyses reveal genotype- and developmental stage-specific molecular responses to drought and salinity stresses in chickpea. *Scientific Reports*, *6*, 19228. https://doi.org/10.1038/srep19228

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., … Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*, *11*(12), 4241–4257. https://doi.org/10.1091/mbc.11.12.4241

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–U130. https://doi.org/10.1038/nbt.1883

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hrdlickova, R., Toloue, M., & Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Reviews: RNA*, *8*(1), e1364. https://doi.org/10.1002/wrna.1364

Huber, A. E., & Bauerle, T. L. (2016). Long-distance plant signaling pathways in response to multiple stressors: The gap in knowledge. *Journal of Experimental Botany*, *67*(7), 2063–2079. https://doi.org/10.1093/jxb/erw099

Jogaiah, S., Govind, S. R., & Tran, L. S. (2013). Systems biology-based approaches toward understanding drought tolerance in food crops. *Critical Reviews in Biotechnology*, *33*(1), 23–39. https://doi.org/10.3109/07388551.2012.659174

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323. https://doi.org/10.1186/1471-2105-12-323

Liu, Y., Zhou, J., & White, K. P. (2014). RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics*, *30*(3), 301–304. https://doi.org/10.1093/bioinformatics/btt688

Lottaz, C., Yang, X., Scheid, S., & Spang, R. (2006). OrderedList–a bioconductor package for detecting similarity in ordered gene lists. *Bioinformatics*, *22*(18), 2315–2316. https://doi.org/10.1093/bioinformatics/btl385

Meyer, E., Aspinwall, M. J., Lowry, D. B., Palacio-Mejia, J. D., Logan, T. L., Fay, P. A., & Juenger, T. E. (2014). Integrating transcriptional, metabolomic, and physiological responses to drought stress and recovery in switchgrass (*Panicum virgatum* L.). *BMC Genomics*, *15*, 527. https://doi.org/10.1186/1471-2164-15-527

Min, H., Chen, C., Wei, S., Shang, X., Sun, M., Xia, R., ... Xie, Q. I. (2016). Identification of drought tolerant mechanisms in maize seedlings based on transcriptome analysis of recombination inbred lines. *Frontiers in Plant Science*, *7*, 1080. https://doi.org/10.3389/fpls.2016.01080

Osborne, C. P., Wythe, E. J., Ibrahim, D. G., Gilbert, M. E., & Ripley, B. S. (2008). Low temperature effects on leaf physiology and survivorship in the C3 and C4 subspecies of *Alloteropsis semialata*. *Journal of Experimental Botany*, *59*(7), 1743–1754. https://doi.org/10.1093/jxb/ern062

Phillips, N. G., Oren, R., Licata, J., & Linder, S. (2004). Time series diagnosis of tree hydraulic characteristics. *Tree Physiology*, *24*(8), 879–890. https://doi.org/10.1093/treephys/24.8.879

Prasad, P. V. V., Pisipati, S. R., Ristic, Z., Bukovnik, U., & Fritz, A. K. (2008). Impact of nighttime temperature on physiology and growth of spring wheat. *Crop Science*, *48*(6), 2372–2380. https://doi.org/10.2135/cropsci2007.12.0717

Qiu, F., Ding, S., Ou, H., Wang, D., Chen, J., & Miyamoto, M. M. (2015). Transcriptome Changes during the life cycle of the red sponge, *Mycale phyllophila* (Porifera, Demospongiae, Poecilosclerida). *Genes (Basel)*, *6*(4), 1023–1052. https://doi.org/10.3390/genes6041023

Rabara, R. C., Tripathi, P., Reese, R. N., Rushton, D. L., Alexander, D., Timko, M. P., ... Rushton, P. J. (2015). Tobacco drought stress responses reveal new targets for Solanaceae crop improvement. *BMC Genomics*, *16*, 484. https://doi.org/10.1186/s12864-015-1575-4

Rauwerda, H., Pagano, J. F. B., de Leeuw, W. C., Ensink, W., Nehrdich, U., de Jong, M., ... Breit, T. M. (2017). Transcriptome dynamics in early zebrafish embryogenesis determined by high-resolution time course analysis of 180 successive, individual zebrafish embryos. *BMC Genomics*, *18*(1), 287. https://doi.org/10.1186/s12864-017-3672-z

Riesgo, A., Andrade, S. C. S., Sharma, P. P., Novo, M., Pérez-Porro, A. R., Vahtera, V., ... Giribet, G. (2012). Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. *Frontiers in Zoology*, *9*(1), 33. https://doi.org/10.1186/1742-9994-9-33

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(3), R25. https://doi.org/10.1186/gb-2010-11-3-r25

Robles, J. A., Qureshi, S. E., Stephen, S. J., Wilson, S. R., Burden, C. J., & Taylor, J. M. (2012). Efficient experimental design and analysis strategies for the detection of differential expression using RNA-sequencing. *BMC Genomics*, *13*(1), 484. https://doi.org/10.1186/1471-2164-13-484

Schubert, M., Gronvold, L., Sandve, S. R., Hvidsten, T. R., & Fjellheim, S. (2019). Evolution of cold acclimation and its role in niche transition in the temperate grass subfamily Pooideae. *Plant Physiology*, *180*(1), 404–419. https://doi.org/10.1104/pp.18.01448

Seelig, H. D., Wolter, A., & Schroder, F. G. (2015). Leaf thickness and turgor pressure in bean during plant desiccation. *Scientia Horticulturae*, *184*, 55–62. https://doi.org/10.1016/j.scienta.2014.12.025

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Tai, Y. C., & Speed, T. P. (2005). Statistical analysis of microarray time course data. In U. Nuber (Ed.), *DNA Microarrays* (pp. 257–280). Oxford, UK: BIOS Scientific Publishers Limited, Taylor & Francis.

Tai, Y. C., & Speed, T. P. (2006). A multivariate empirical Bayes statistic for replicated microarray time course data. *Annals of Statistics*, *34*(5), 2387–2412. https://doi.org/10.1214/009053606000000759

Toung, J. M., Morley, M., Li, M., & Cheung, V. G. (2011). RNA-sequence analysis of human B-cells. *Genome Research*, *21*(6), 991–998. https://doi.org/10.1101/gr.116335.110

Valliyodan, B., & Nguyen, H. T. (2006). Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Current Opinion in Plant Biology*, *9*(2), 189–195. https://doi.org/10.1016/j.pbi.2006.01.019

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. https://doi.org/10.1038/nrg2484

Williams, A. G., Thomas, S., Wyman, S. K., & Holloway, A. K. (2014). RNA-seq data: challenges in and recommendations for experimental design and analysis. *Current Protocols in Human Genetics*, *83*, 11–13. https://doi.org/10.1002/0471142905.hg1113s83

Williams, W. M., Williamson, M. L., & Real, D. (2011). Paspalum. In C. Kole (Ed.), *Wild crop relatives: Genomic and breeding resources* (pp. 197–216). Berlin, Germany: Springer.

Wu, Y., Wei, W., Pang, X., Wang, X., Zhang, H., Dong, B. O., ... Wang, M. (2014). Comparative transcriptome profiling of a desert evergreen shrub, *Ammopiptanthus mongolicus*, in response to drought and cold stresses. *BMC Genomics*, *15*, 671. https://doi.org/10.1186/1471-2164-15-671

Zdobnov, E. M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R. M., Simão, F. A., Ioannidis, P., ... Kriventseva, E. V. (2017). OrthoDB v9. 1: Cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research*, *45*(D1), D744–D749. https://doi.org/10.1093/nar/gkw1119

Zhang, X., Lei, L., Lai, J., Zhao, H., & Song, W. (2018). Effects of drought stress and water recovery on physiological responses and gene expression in maize seedlings. *BMC Plant Biology*, *18*(1), 68. https://doi.org/10.1186/s12870-018-1281-x

Zhao, Q. Y., Wang, Y., Kong, Y. M., Luo, D., Li, X., & Hao, P. (2011). Optimizing de novo transcriptome assembly from short-read RNA-seq data: A comparative study. *BMC Bioinformatics*, *12*(Suppl 14), S2. https://doi.org/10.1186/1471-2105-12-S14-S2

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Qiu F, Bachle S, Nippert JB, Ungerer MC. Comparing control options for time-series RNA sequencing experiments in nonmodel organisms: An example from grasses. *Mol Ecol Resour*. 2020;20:681–691. https://doi.org/10.1111/1755-0998.13137