



A comparative genomics examination of desiccation tolerance and sensitivity in two sister grass species

Ricardo A. Chávez Montes^{a,b}, Anna Haber^{c,d}, Jeremy Pardo^{c,d,e}, Robyn F. Powell^f, Upendra K. Divisetty^g, Anderson T. Silva^h, Tania Hernández-Hernández^{a,i}, Vanildo Silveira^j, Haibao Tang^k, Eric Lyons^k, Luis Rafael Herrera Estrella^{a,b,1}, Robert VanBuren^{c,d,1}, and Melvin J. Oliver^{h,l,1}

^aNational Laboratory of Genomics for Biodiversity, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional 36824 Irapuato, Guanajuato, Mexico; ^bInstitute of Genomics for Crop Abiotic Stress Tolerance, Department of Plant and Soil Science, Texas Tech University, Lubbock, TX 79409; ^cDepartment of Horticulture, Michigan State University, East Lansing, MI 48824; ^dPlant Resilience Institute, Michigan State University, East Lansing, MI 48824; ^eDepartment of Plant Biology, Michigan State University, East Lansing, MI 48824; ^fTrait Diversity and Function, Royal Botanic Gardens, Richmond Surrey TW9 3AB, United Kingdom; ^gCyVerse Collaborative, School of Plant Sciences, University of Arizona, Tucson, AZ 85721; ^hDivision of Plant Sciences, University of Missouri, Columbia, MO 65211; ⁱDepartment of Research, Conservation and Collections, Desert Botanical Garden, Phoenix, AZ; ^jLaboratório de Biotecnologia, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, RJ 28013-602, Brazil; ^kSchool of Plant Sciences, BIO5 Institute, University of Arizona, Tucson, AZ 85721; and ^lInterdisciplinary Plant Group, University of Missouri, Columbia, MO 65211

Edited by Elizabeth Kellogg, Donald Danforth Plant Science Center, St. Louis, MO; received October 19, 2021; accepted December 14, 2021

Desiccation tolerance is an ancient and complex trait that spans all major lineages of life on earth. Although important in the evolution of land plants, the mechanisms that underlay this complex trait are poorly understood, especially for vegetative desiccation tolerance (VDT). The lack of suitable closely related plant models that offer a direct contrast between desiccation tolerance and sensitivity has hampered progress. We have assembled high-quality genomes for two closely related grasses, the desiccation-tolerant *Sporobolus stapfianus* and the desiccation-sensitive *Sporobolus pyramidalis*. Both species are complex polyploids; *S. stapfianus* is primarily tetraploid, and *S. pyramidalis* is primarily hexaploid. *S. pyramidalis* undergoes a major transcriptome remodeling event during initial exposure to dehydration, while *S. stapfianus* has a muted early response, with peak remodeling during the transition between 1.5 and 1.0 grams of water (gH₂O) g⁻¹ dry weight (dw). Functionally, the dehydration transcriptome of *S. stapfianus* is unrelated to that for *S. pyramidalis*. A comparative analysis of the transcriptomes of the hydrated controls for each species indicated that *S. stapfianus* is transcriptionally primed for desiccation. Cross-species comparative analyses indicated that VDT likely evolved from reprogramming of desiccation tolerance mechanisms that evolved in seeds and that the tolerance mechanism of *S. stapfianus* represents a recent evolution for VDT within the Chloridoideae. Orthogroup analyses of the significantly differentially abundant transcripts reconfirmed our present understanding of the response to dehydration, including the lack of an induction of senescence in resurrection angiosperms. The data also suggest that failure to maintain protein structure during dehydration is likely critical in rendering a plant desiccation sensitive.

desiccation tolerance | sister group contrast | comparative genomics | transcriptome remodeling | C4 grasses

Desiccation tolerance (DT) is a fundamental trait that is widespread and developed early in the evolution of the land plants (1, 2), and it is believed to have been critical in the colonization of the land by green algae (3). In tracheophytes, DT is generally limited to reproductive propagules, such as seeds and spores, while vegetative desiccation tolerance (VDT) occurs in only 0.086% of known vascular plant species (4). Our understanding of VDT (and its relationship to seed DT) has broadened with the recent expansion of whole-genome sequencing of resurrection plants, tracheophytes that can survive the desiccation of their vegetative tissues. Since the release of the *Boea hygrometrica* genome sequence (5), the genomes of four other resurrection angiosperms [*Xerophyta schlechteri* (6), *Oropetium thomaeum* (7, 8), *Lindernia brevidens* (9), and *Eragrostis nindensis* (10)], two lycophytes [*Selaginella tamariscina* (11) and *Selaginella lepidophylla* (12)], and the bryophyte

Syntrichia caninervis (13) have been published. Apart from the obvious benefits of obtaining genomic resources for individual resurrection species, the establishment of a collection of resurrection plant genomes offered the possibility of the reconstruction of an ancestral genome of a desiccation-tolerant progenitor that would reveal a genomic signature (blueprint) that defines a common mechanism for DT. However, a genomic blueprint for DT has not emerged (4), which may be related to the small number of genomes available and limited phylogenetic sampling, that all tracheophytes possess desiccation-tolerant propagules (seeds or spores), which would obfuscate the comparative analyses, or that the origin of DT lies deep in the land plant phylogeny and is thus cryptic in the recent plant lineages. It may also be a combination of these possibilities or that there is no genomic blueprint for this fundamental trait. Although a genomic blueprint for DT has not been revealed, comparative studies have demonstrated that certain gene families, such as those for early light-inducible proteins (ELIPs) and late embryogenesis-abundant proteins, have expanded in species that exhibit VDT (6, 14, 15).

Significance

This is a significant sister group contrast comparative study of the underpinning genomics and evolution of desiccation tolerance (DT), a critical trait in the evolution of land plants. Our results revealed that the DT grass *Sporobolus stapfianus* is transcriptionally primed to tolerate a dehydration/desiccation event and that the desiccation response in the DT *S. stapfianus* is distinct from the water stress response of the desiccation-sensitive *Sporobolus pyramidalis*. Our results also show that the desiccation response is largely unique, indicating a recent evolution of this trait within the angiosperms, and that inhibition of senescence during dehydration is likely critical in rendering a plant desiccation tolerant.

Author contributions: E.L., L.R.H.E., R.V., and M.J.O. designed research; J.P., R.F.P., T.H.-H., H.T., and M.J.O. performed research; R.A.C.M., A.H., J.P., R.F.P., U.K.D., A.T.S., T.H.-H., V.S., H.T., E.L., L.R.H.E., R.V., and M.J.O. analyzed data; and R.A.C.M., A.H., L.R.H.E., R.V., and M.J.O. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: luis.herrera-estrella@ttu.edu, bob.vanburen@gmail.com, or olivermj@missouri.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2118886119/-DCSupplemental>.

Published January 26, 2022.

Table 1. Estimation of the genome size (1C value) using flow cytometry

Species	Standard	Average: sample	Average: standard	GS (2C), pg	GS (1C), pg	SD sample
<i>S. stapfianus</i>	<i>P. crispum</i>	305.2	367.85	3.734	1.867	0.011533
<i>S. pyramidalis</i>	<i>P. crispum</i>	206.57	335.67	2.769	1.385	0.02623

Estimated genome size (GS), in bold, for the diploid (2C) and haploid (1C) states.

A corollary to the ancestral reconstruction approach to understanding the evolution of VDT and the genomic aspect of its phenotypic expression is the comparison of the genomes of closely related species that contrast the two extremes: sensitivity and tolerance. Such closely related contrasting species pairings are rare in resurrection plants, but this approach has been applied, albeit with species pairs that are not as close as would be ideal. The genomes and dehydration–rehydration transcriptomes of two resurrection eudicots within the Linderniaceae family (16), the desiccation-tolerant *L. brevidens* and the desiccation-sensitive (DS) *Lindernia subracemosa*, were sequenced and compared (9). The comparison revealed that at least in the *Lindernia* lineage, VDT evolved via a combination of gene duplications in gene families that are functionally associated with the desiccation response and a network-level rewiring of gene expression in vegetative tissue commonly associated with seed desiccation. More recently, a comparative analysis of two contrasting grass genomes along with their respective desiccation-related transcriptomes, the desiccation-tolerant *E. nindensis* and the related DS cereal *Eragrostis tef*, reinforced the potential role of gene duplications in the evolution of DT (10). Although there is still a significant phylogenetic distance between these two *Eragrostis* species (17), the comparative analysis and its extension to include other C4 grasses, including the desiccation-tolerant *O. thomaeum*, revealed chromatin restructuring and methylation patterns associated with down-regulated genes and specific seed-related orthologs whose expression is associated with VDT. The comparative transcriptome analyses indicated that genes having important roles in seed development and DT are broadly expressed under dehydration in both sensitive and tolerant species, with just a few genes uniquely expressed in the tolerant plants.

In this study, we have chosen two phylogenetically closely related C4 grasses, the homoiochlorophyllous desiccation-tolerant *Sporobolus stapfianus* and the DS *Sporobolus pyramidalis*, to develop detailed comparative genomic and transcriptomic analyses to further explore genomic inferences into the evolution of VDT. *S. stapfianus* and *S. pyramidalis* are members of the same clade, clade A, in the *Sporobolus* family of the Sporobolinae subtribe of the Chloridoideae grasses (18). *S. stapfianus* has been the subject of many mechanistic studies of its DT phenotype (19, 20) and along with *S. pyramidalis*, the subject of a detailed comparative leaf metabolomics study that highlighted differences in the metabolic responses of the two species to dehydration (21). We constructed Hi-C–derived assemblies of the sequenced genomes for both species and conducted transcript profiling analyses for parallel reductions in water contents for both species as well as a full desiccation drying series for *S. stapfianus*. We performed a detailed comparative genomic analysis for the two species and extended the analysis to include other grass species, both desiccation tolerant and DS. Our results offer insights into the mechanism and evolution of VDT in the Chloridoideae grasses.

Results

Generating Genomic Resources for *Sporobolus* Grasses. One-step flow cytometric assays generated size estimates for each of the *Sporobolus* genomes. The haploid genome of *S. stapfianus* had an average of 1,385 pg of DNA per nucleus, which is approximately equal to a complete genome sequence of 1.354 Gbp, and the haploid genome of *S. pyramidalis* had an average of 1,867 pg of DNA per nucleus, which is ~1.826 Gbp (Table 1).

Draft genome assemblies were generated for each grass using Illumina whole-genome shotgun sequencing combined with Chicago and Hi-C proximity ligation (*Materials and Methods*). The final assemblies consisted of 11,574 scaffolds with an N50 of 19.4 Mb for *S. stapfianus* and 2,518 scaffolds with an N50 of 21.6 Mb for *S. pyramidalis*, with the longest scaffolds for both species greater than 60 Mbp. Despite their high contiguity, the assembled genomes are smaller than the estimated genome size, at 1.080 and 1.055 Gbp for *S. stapfianus* and *S. pyramidalis*, respectively. These differences between the estimated and assembled genome sizes are likely caused by collapsed homologous regions in these complex polyploid species as described in detail below. Both genomes have similar levels of repetitive elements, 39.7 and 41.3% for *S. stapfianus* and *S. pyramidalis*, respectively (Table 2), with almost identical distributions of known repeat families (*SI Appendix, Table S1*). Gypsy and Copia retrotransposons are the most predominant families of the known repeats at 36 and 10 to 12%, respectively, for the two genomes.

The *Sporobolus* genomes were annotated using MAKER with a combination of RNASeq and PacBio Iso-Seq full-length transcripts as expressed sequence tag (EST) evidence and protein homology from other high-quality plant genomes. After filtering, the final annotations contained 52,208 and 51,207 gene models for *S. stapfianus* and *S. pyramidalis*, respectively (Table 2). Annotation completeness was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) with the poales_odb10.2019–11–20 database of 4,896 conserved genes. The genome annotations recovered 93.5 and 92.4% of complete BUSCOs for *S. stapfianus* and *S. pyramidalis*, respectively, indicating that both genomes were well annotated and contained the vast majority of the coding portion of these two genomes (Table 3). Gene models were functionally annotated using a simplified maize–GAMER pipeline; 96% of genes were annotated with InterProScan domain/family information, and 66% were annotated with Gene Ontology (GO) descriptions for both genomes.

Comparative Genomics of Chloridoideae Grasses. *Sporobolus* belongs to the Chloridoideae subfamily of grasses, a large and diverse group of predominantly C4 species with remarkable drought, heat, and salinity tolerance. The orphan grain crops finger millet and teff are found within Chloridoideae, as are several model desiccation-tolerant plants in the genera *Oropetium*, *Eragrostis*, *Tripogon*, and *Sporobolus* among others. Most of the surveyed Chloridoideae species (~90%) are polyploid, including species from many of the aforementioned taxa. The availability of several high-quality chloridoideae genomes facilitates detailed comparative genomic comparisons within these grasses. Macro-synteny between *S. stapfianus* and *S. pyramidalis* shows a clear 2:3 pattern, consistent with the tetraploid and hexaploid nature of these grasses, respectively (Fig. 1 and *SI Appendix, Fig. S1*). Comparisons with the closely related diploid chloridoideae grass *O. thomaeum* also revealed 1:2 and 1:3 patterns of synteny for *S. stapfianus* and *S. pyramidalis*, respectively, supporting their polyploidy (Fig. 1 and *SI Appendix, Fig. S2*). Although neither *Sporobolus* genome is scaffolded into complete chromosomes, large 20-Mb+-sized scaffolds are highly collinear with the *Oropetium* genome with few structural large-scale rearrangements (*SI Appendix, Fig. S2*), which is consistent with the unusually high conservation of karyotype and collinearity observed among other chloridoideae grass genomes (22).

Table 2. *S. pyramidalis* and *S. stapfianus* genome assemblies

	<i>S. stapfianus</i>	<i>S. pyramidalis</i>
Assembly size, bp	1,079,709,429	1,055,099,634
No. of contigs	11,574	2,518
Longest contig, bp	74,381,802	60,313,943
N50	19,421,551	21,593,735
N90	6,548,862	7,502,351
Repetitive elements, bp (%)	428,320,271 (39.7)	435,795,325 (41.3)
Genome GC content, %	44.0	43.1
No. of genes	52,808	51,207
No. of transcripts	61,021	62,960
No. of CDS	61,021	62,960
Median mRNA length, bases	1,239	1,694
Median CDS length, bases	1,071	1,221
CDS GC content, %	52.6	51.9
Median exons per transcript	4	5
Median exon length, bases	148	150
Median intron length, bases	157	154
No. of transcription factors	2,722	2,897
No. of genes with InterProScan entry (%)	52,601 (99.6)	51,023 (99.6)
No. of genes with GO biological process entry (%)	34,874 (66.0)	34,039 (66.5)

Genome assembly statistics for both *S. stapfianus* and *S. pyramidalis*. N50 and N90 represent the length of the shortest contig for which longer and equal length contigs cover at least 50 % and 90% of the assembly respectively. GC represents the guanine-cytosine contents and CDS represents the number of protein coding sequences for each genome.

Macrosyntentic analysis between the *Sporobolus* species and *O. thomaeum* exposed an overall more complex polyploid structure than the more straightforward tetraploid and hexaploid compositions (*SI Appendix, Fig. S2*). Roughly half the hexaploid *S. pyramidalis* genome has the expected 3:1 pattern of syntenic blocks compared with *O. thomaeum*, while 37% is only 2:1. The pattern is similar for tetraploid *S. stapfianus*, where ~44% of syntenic blocks are 2:1 to *O. thomaeum* as expected and ~42% of blocks are 1:1 (*SI Appendix, Fig. S2*). Similar assembly issues were observed in the tetraploid chloridoid grass *E. nindensis*, where one to four regions were assembled for each syntenic region in *O. thomaeum* (10). These discrepancies, combined with differences between the estimated and assembled genome sizes, suggest the *Sporobolus* genomes were partially collapsed during assembly in homologous regions. *S. pyramidalis* and *S. stapfianus* may be segmental allopolyploids with varying degrees of homology between chromosomes from separate subgenomes. Partial collapse during assembly would result in divergent homologous regions assembling separately and highly similar regions collapsing, which is supported by the observation that the ratio of assembled syntenic blocks is maintained across large syntenic blocks and whole chromosomes in *O. thomaeum*. For instance, two homologous regions are assembled in *S. pyramidalis* for chromosomes 3 and 4 from *O. thomaeum*, while three regions in *S. pyramidalis* were identified for most of chromosome 2 in *O. thomaeum*. Similar patterns were observed between *S. stapfianus* and *O. thomaeum*. To account for these issues related to polyploidy, syntenic gene pairs and orthogroups were used for

downstream comparative genomics and transcriptomics analyses between the *Sporobolus* genomes and other chloridoid grasses.

Transcriptional Landscape of Desiccation in *S. pyramidalis* and *S. stapfianus*.

We generated RNASeq data from RNA isolated from leaf tissues at different stages of dehydration for both species (*SI Appendix, Fig. S3*). Differentially expressed genes were identified using edgeR (23), and the resulting gene lists were assigned to GO biological process categories enrichment using the Cytoscape (23) plugin Bingo (24). These analyses indicate that *S. pyramidalis* and *S. stapfianus* transcriptomes respond differently to dehydration and share few biological process adaptations during the drying process. When water content decreases from 3 to 2 grams of water (gH₂O) g⁻¹ dry weight (dw), *S. pyramidalis* exhibits a strong response with 11,978 statistically differentially abundant transcripts (SDATs), in contrast to the more moderate response of 1,776 SDATs in *S. stapfianus* (Fig. 2 A and B). A GO enrichment analysis of SDAT lists further demonstrates that during the 3 to 2 gH₂O g⁻¹ dw water content transition, few biological processes are shared between the two species (Fig. 2 C and D and *SI Appendix, Fig. S4*). Some biological process categories, including response to heat and response to reactive oxygen species, are common to both species (*SI Appendix, Fig. S4*). Moreover, while *S. pyramidalis* responds to the change in water content from 3 to 2 gH₂O g⁻¹ dw by modulating processes involving the ribosome and the cell wall, *S. stapfianus* initiates alterations in the abundance of transcripts that relate to the response to oxidative stress, response to water deficit, and protein refolding (*SI Appendix, Fig. S4*).

As dehydration advances from 2 to 1.5 gH₂O g⁻¹ dw in *S. pyramidalis*, the functional categories of SDATs remain relatively unchanged from that activated at the initial loss of water, and as it is undergoing senescence during the 1.5 to 1 gH₂O g⁻¹ dw transition, further acclimation appears unlikely. By contrast, *S. stapfianus* exhibits an increase to 3,730 SDATs during the 2 to 1.5 gH₂O g⁻¹ dw transition, but starting at the 1.5 to 1 gH₂O g⁻¹ dw transition, it initiates a major remodeling of its transcriptome (*SI Appendix, Fig. S3*), as indicated by a significant increase to 14,557 and 16,047 SDATs during these two transitions in water content, respectively (Fig. 2D). Global transcriptional remodeling continues during the 0.75 to 0.5 gH₂O g⁻¹

Table 3. Genome assemblies BUSCO v4 statistics vs. the grass (poales_odb10) dataset

	<i>S. pyramidalis</i>	<i>S. stapfianus</i>
Complete BUSCOs	4,575 (93.5)	4,521 (92.4)
Complete and single-copy BUSCOs	2,162 (44.2)	3,421 (69.9)
Complete and duplicated BUSCOs	2,413 (49.3)	1,100 (22.5)
Fragmented BUSCOs	65 (1.3)	76 (1.6)
Missing BUSCOs	256 (5.2)	299 (6.0)

Data are presented as the no. of BUSCOs (%).

Sporobolus pyramidalis

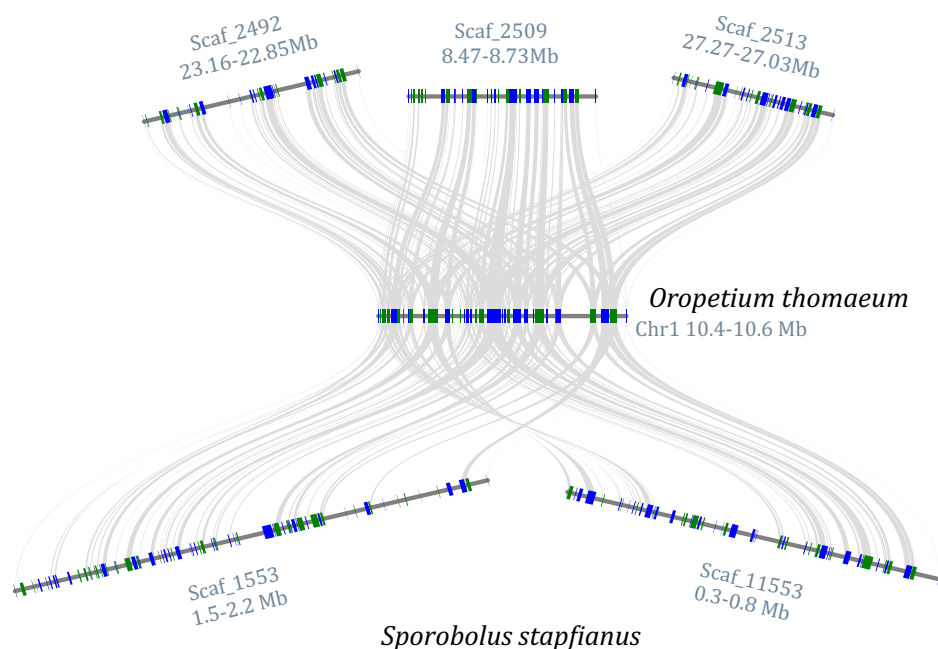


Fig. 1. Microsynteny within Chloridoideae grasses. A collinear region between *O. thomaeum*, *S. stapfianus*, and *S. pyramidalis* is highlighted, reflecting the ploidy of each species (diploid, tetraploid, and hexaploidy, respectively). Genes are shown in blue and green, and syntenic gene pairs are connected by gray lines.

dw transition, albeit at a lower degree, with 8,146 SDATs (Fig. 2D). When desiccated *S. stapfianus* plants are rehydrated, another strong transcriptome reprogramming, with 27,280 SDATs 12 h after rehydration, is evident and shifts to a transcriptome functional expression profile more similar to that of the fully hydrated control (SI Appendix, Fig. S3). Although *S. stapfianus* appeared morphologically fully recovered after 24 h of rehydration, the transcriptional profile is not equivalent to that observed in leaves of plants with a water content of 3 gH₂O g⁻¹ dw (SI Appendix, Fig. S3), with 24,659 SDATs between the two conditions (Fig. 2B). Leaves from plants 24 h after rehydration have up-regulated SDATs classified in ribosome biogenesis GO categories and down-regulated SDATs in photosynthesis categories, as well as remnants of stress-responsive adaptations, including the response to water categories, and altered metabolism, suggested by the presence of glucose 6-phosphate, fructose 1,6-bisphosphate, and several other metabolism-related categories (SI Appendix, Fig. S4B).

Syntenic Ortholog Transcripts Comparison of the Hydrated State in *S. pyramidalis* and *S. stapfianus*. To directly compare the transcriptomes for *S. stapfianus* and *S. pyramidalis* and identify differentially regulated transcripts that relate to the differences between the two species in the hydrated state prior to dehydration, we created a custom list of syntenic ortholog “genes” (Materials and Methods). Differential expression was accomplished using a contrast *S. stapfianus* vs. *S. pyramidalis* in edgeR (23), and the resultant syntenic ortholog gene lists were probed with GO enrichment as described previously for the intraspecies dehydration transcriptome analyses. The analyses demonstrate that *S. stapfianus* and *S. pyramidalis* have very different transcriptional landscapes under hydrated conditions that reflect functionally different priorities for each species. The *S. stapfianus* transcriptome significantly favors nitrogen, starch, and photosynthetic metabolic processes, whereas the *S.*

pyramidalis transcriptome significantly favors processes involved in growth, primarily the biogenesis of cell wall components (SI Appendix, Fig. S5A). These differences are also reflected at the cellular component and molecular levels (SI Appendix, Fig. S5B and C), with the majority of cellular functions related to the chloroplast and photosystems in *S. stapfianus* and the symplast, cytoskeleton, cell wall, and cell wall modification activities in *S. pyramidalis*.

Syntenic Ortholog Protein Comparisons of Desiccation Response in *S. pyramidalis* and *S. stapfianus*. To further compare the response of *S. pyramidalis* and *S. stapfianus* to dehydration, we performed a proteomic analysis using young leaves at 3 and 1.5 gH₂O g⁻¹ dw and focused on proteins encoded by syntenic genes in a comparison of enriched GO biological process categories of accumulating and decreasing proteins in both water content conditions (SI Appendix, Fig. S6). At 1.5 gH₂O g⁻¹ dw, *S. pyramidalis* had increased accumulation of proteins that are almost exclusively involved in stress responses; *S. stapfianus* had increased accumulation of stress response proteins but also, accumulated proteins involved in the response to misfolded proteins and protein catabolism (SI Appendix, Fig. S6A), and it decreased the abundance of proteins involved in energy production (SI Appendix, Fig. S6B). The protein data demonstrate that, as observed for the transcriptomic profiles, *S. pyramidalis* and *S. stapfianus* follow predominantly different approaches of protein accumulation in their response to dehydration.

Expression Dynamics of DS and Desiccation-Tolerant Grasses. To explore the evolution of VDT in the Chloridoideae subfamily of grasses, we made use of several high-quality genomes with similar dehydration expression datasets that were available for this group of grasses: the desiccation tolerant (*S. stapfianus*, *O. thomaeum*, and *E. nindensis*) and the DS (*E. tef* and *S. pyramidalis*). To facilitate comparisons between species with different ploidy, we

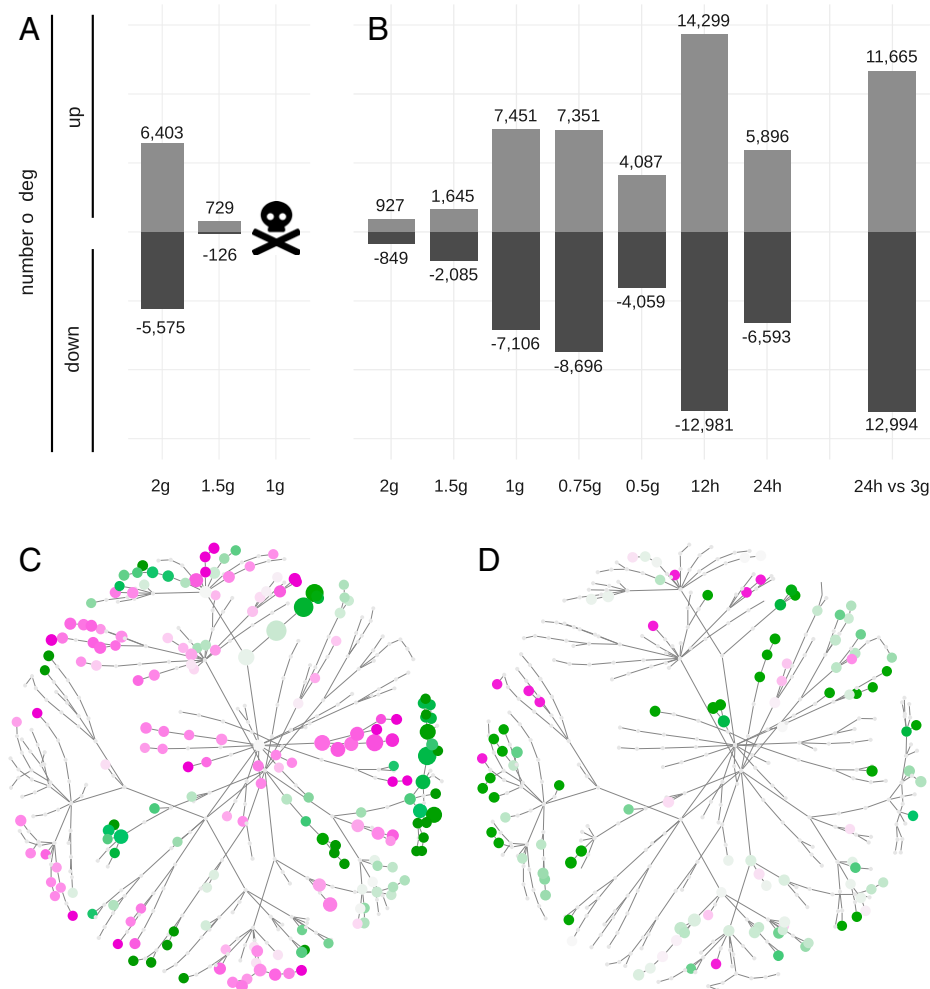


Fig. 2. *S. pyramidalis* and *S. stapfianus* transcriptional landscape during desiccation/rehydration. (A and B) Bar plots of the numbers of differentially expressed genes (FDR \leq 0.01) for *S. pyramidalis* (A) and *S. stapfianus* (B) from edgeR contrasts of sequential conditions; 2g corresponds to the contrast 2 vs. 3 gH₂O g⁻¹ dw, 1.5g corresponds to 1.5 vs. 2 gH₂O g⁻¹ dw, 1g corresponds to 1 vs. 1.5 gH₂O g⁻¹ dw, and so on. The last *S. stapfianus* contrast is 24 h after recovery irrigation vs. 3 gH₂O g⁻¹ dw. The numbers of up- and down-regulated genes are indicated at the top and bottom of each bar, respectively. The skull and bones icon indicates that *S. pyramidalis* is severely affected when at 1 gH₂O g⁻¹ dw and enters into senescence. (C and D) Graphs of enriched GO biological process categories in the contrast 2 vs. 3 gH₂O g⁻¹ dw for *S. pyramidalis* (C) and *S. stapfianus* (D). Nodes represent categories and edges represent the parent-child relationships in the ontology. Node identities and positions are identical in both graphs. Color is proportional to the ratio of increased abundance vs. decreased abundance transcripts in the category, with a green color indicating a ratio of more than one (a majority of increased abundance transcripts) and a magenta color indicating a ratio of less than one (a majority of decreased abundance transcripts). Category identifications and names are listed in *SI Appendix*, Fig. S4.

clustered genes into syntenic orthologs using MCScan (25) and orthologous groups (orthogroups) using OrthoFinder (26) and compared expression patterns between genes in the same orthogroups. We identified 49,418 orthogroups from OrthoFinder containing 806,075 genes across 23 diverse land plant genomes and focused the subsequent analyses on orthogroups, orthologs, or syntenic gene pairs present in the genome of all chloridoid grasses.

We first surveyed the global expression profiles of the five Chloridoid grasses under well-watered, drought/desiccation, and rehydration conditions using transformed expression data of 19,267 shared syntenic orthologs across all species. We applied a dimensionality reduction on the resulting expression matrix through principal component analysis. The first two principal components collectively explain 62% of the variance and separate the expression datasets by species and stress (Fig. 3). Well-watered RNASeq samples are found in a single tight cluster of all five species, while desiccation and rehydration samples are found in dispersed but distinct clusters.

Samples from dehydration and rehydration time courses in the DT species fall into two clusters, with *E. nindensis* and *O. thomaeum* samples intertwined in one cluster and *S. stapfianus* in the second. The dehydration samples from the two DS species (*E. tef* and *S. pyramidalis*) clustered together in a third distinct cluster. Samples of *E. nindensis* and *O. thomaeum* are separated by relative water content in principal component (PC)1 and by dehydration vs. rehydration in PC2, but interestingly, they are not delineated by species. Together, these results indicate that expression patterns are broadly conserved in leaf samples of all species but that dehydration and rehydration samples are distinct between the three lineages of DT species and their DS relatives.

The same leaf RNASeq data were analyzed in a pairwise fashion to identify genes with significantly increased transcript abundance under dehydrating conditions in all five species. These SDATs were clustered based on orthogroup using OrthoFinder (as described above) and compared between species. Orthogroups were used in this set of analyses as they contained more genes than the synteny-based analyses, and orthogroups

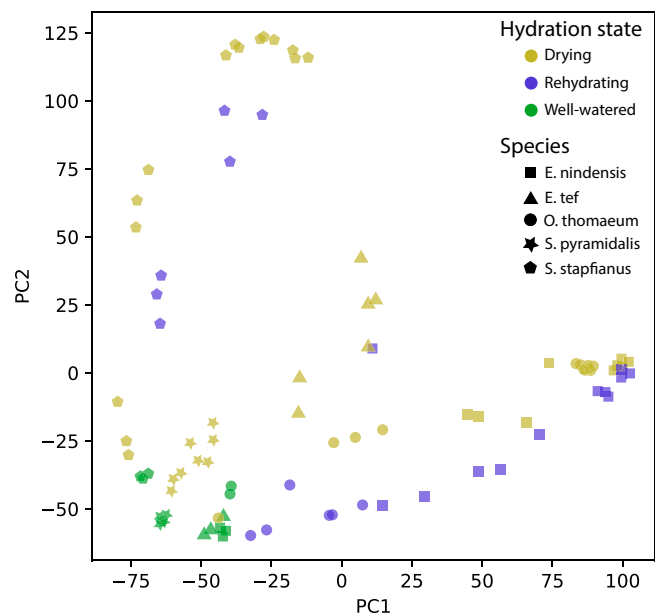


Fig. 3. Dimensional reduction of drought expression profiles across DS and DT Chloridoid grasses. Raw expression values for syntenic orthogroups were transformed by z score prior to principal component analysis. The first two principal components are plotted for the two DS Chloridoid grasses (*E. tef* and *S. pyramidalis*) and three tolerant grasses (*E. nindensis*, *O. thomaeum*, and *S. stapfianus*) with comparative expression datasets. Points are colored by species or hydration state as indicated in the key.

have better resolution of recently duplicated genes. Across the five sequenced chloridoid grasses, the largest number of up-regulated orthogroups under dehydrating conditions was observed between the two *Sporobolus* species (Fig. 4), as expected since they are sister taxa. The second largest number of up-regulated orthogroups was shared between the two

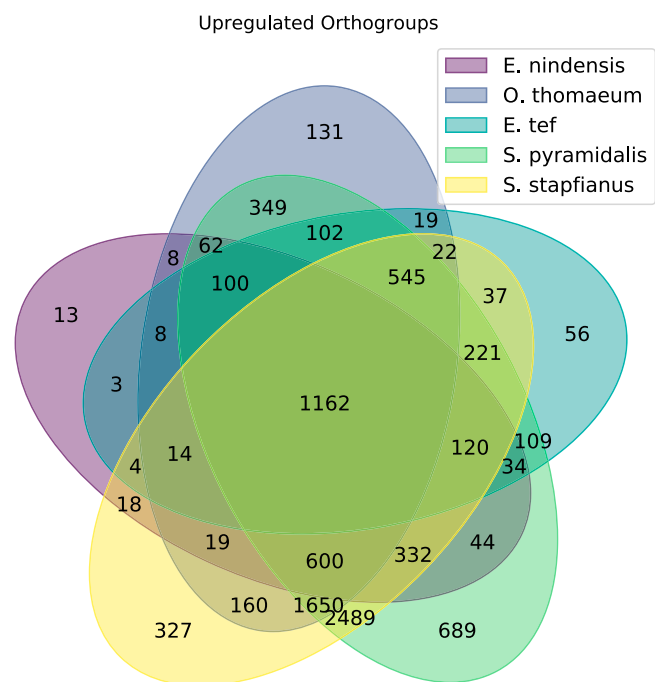


Fig. 4. Venn diagram of up-regulated orthogroups across the five surveyed chloridoid grasses. The number of overlapping orthogroups with up-regulated expression under drought is shown for each comparison.

Sporobolus species and *O. thomaeum* (Fig. 4), which is consistent with their phylogenetic placement within the Chloridoideae. Many other orthogroups are up-regulated similarly in all five species (Fig. 4). The orthogroups uniquely up-regulated in all VDT species are enriched in 214 biological process GO terms (*SI Appendix, Fig. S7*). Highly enriched GO terms include ultraviolet UV light response, chlorophyll catabolism, reactive oxygen species (ROS) metabolism, seed dormancy maintenance by abscisic acid (ABA), and gene expression in response to heat stress, among others (*SI Appendix, Fig. S7A*). These GO terms are consistent with well-characterized processes related to DT. Other GO terms with a lower magnitude of enrichment include those related to lipids, osmoprotectant biosynthesis, high light response, energy metabolism, protein degradation, and ABA signaling (*SI Appendix, Fig. S7 B and C*). Seventy-one biological process GO terms were uniquely up-regulated in only the DS species (*SI Appendix, Fig. S8*). These included several terms related to salicylic acid as well as ethylene and ABA signaling, arabinose biosynthesis, cell wall biogenesis, and notably, leaf senescence, among others (*SI Appendix, Fig. S4*). We then asked whether any of the GO terms uniquely up-regulated in DT species would overlap with those uniquely down-regulated in DS species and vice versa (*SI Appendix, Table S3*). The GO term “protein folding” was uniquely up-regulated in DT and down-regulated in DS species. Across these five species, most seed-related orthogroups are up-regulated similarly (*SI Appendix, Fig. S9*). There are no seed orthogroups that are up-regulated in all three DT species without also being up-regulated in one or more DS species.

Expansion of ELIPs. ELIPs have a conserved role in photoprotection during desiccation, and they have undergone massive tandem gene duplication in all sequenced resurrection plant genomes surveyed to date (14). We observed a similar duplication of ELIPs in the *Sporobolus* genomes (Fig. 5A). The *S. stapfianus* genome has 65 ELIPs in three tandem arrays, and the *S. pyramidalis* genome has 30 ELIPs in two tandem arrays (Fig. 5B). The largest array in *S. stapfianus* has 49 ELIPs compared with 17 in its corresponding homologous region, suggesting the duplications occurred after the divergence of the two *S. stapfianus* subgenomes. Both *O. thomaeum* and *S. stapfianus* have large tandem arrays of ELIPs, but the duplication events originated from different syntenic orthologs. The total number of ELIPs in *S. pyramidalis* is higher than some other desiccation-tolerant species, but when gene counts are normalized for ploidy, the ELIPs are within the range of other sensitive grasses.

ELIPs have little to no detectable expression in well-watered tissue, but they are highly induced in desiccating *S. stapfianus* leaf tissue after they reach $1.0 \text{ gH}_2\text{O g}^{-1} \text{ dw}$, and their expression continues 12 and 24 h postrehydration (Fig. 5). ELIPs are also up-regulated under drought in *S. pyramidalis*, and this occurs quickly in the dehydration process at 2.0 and $1.5 \text{ gH}_2\text{O g}^{-1} \text{ dw}$. However, their combined expression is less than *S. stapfianus* (Fig. 5C), similar to what has been observed in other DS grasses (14).

Discussion

The genomic resources we developed for the sister species *S. stapfianus* and *S. pyramidalis* offer a robust contrast that facilitates a strong comparison between a VDT and a DS grass species. The addition of the genomic resources from other resurrection grasses, *O. thomaeum* (8) and *E. nindensis* (10), broadens the comparison further into the Chloridoideae subfamily of grasses. The two genome assemblies revealed the complex mixed ploidy of these two grasses, with *S. stapfianus* primarily tetraploid and *S. pyramidalis* primarily hexaploid. The structural complexity of the two genomes likely contributed to

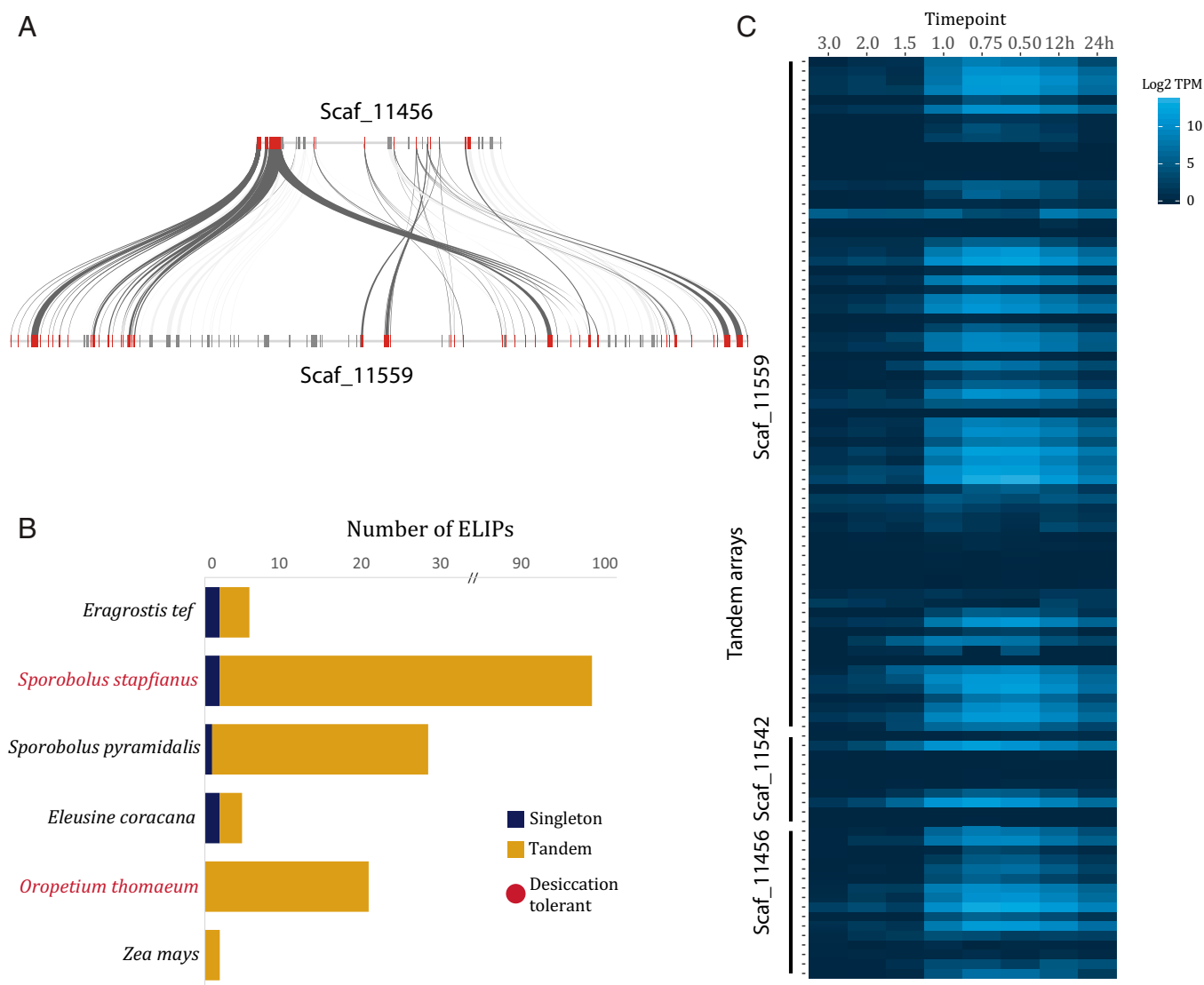


Fig. 5. ELIPs tandem duplication in *S. stapfianus* and ELIP gene abundance in leaf tissues. (A) Microsynteny of two ELIP tandem arrays is shown in *S. stapfianus*. ELIPs are shown in red, other genes are shown in gray, and syntenic homeologs between the scaffolds are denoted by gray connections. (B) The number of ELIPs in sequenced Chloridoideae grasses (*E. tef*, *S. stapfianus*, *S. pyramidalis*, *E. coracana*, *O. thomaeum*, and *Z. mays*) is plotted. The two desiccation-tolerant grasses are denoted in red. (C) Log₂-transformed gene abundance (TPM) of the 30 ELIPs in *S. pyramidalis* and 65 ELIPs in *S. stapfianus* across each replicate of the leaf desiccation time courses.

the inability to assemble the genomes into chromosome-level contigs or to record sequenced genome sizes equivalent to those determined cytologically. The increase in ploidy between the two species probably occurred immediately after the divergence of the *S. pyramidalis* clade from the common ancestor of the two species (18). The assemblies did not reveal any genomic structural characteristics, with the exception perhaps of tandem arrays of ELIP genes (14), that could be attributed to the difference in VDT between the two species, which is consistent with the general observation that there is not a genomic “blueprint” for VDT in resurrection species (4). However, the assemblies did allow for a thorough comparative analysis, both structural and functional, of the gene space for each genome, and coupled with the in-depth transcriptome data, we were able to explore a detailed genomic assessment of the dehydration/desiccation responses within the *Sporobolus* sister species contrast.

The generation of transcriptomic and proteomic data for dehydrating young leaf tissue at specific water contents during a dry-down experiment such that the dehydration levels are

survivable for both grasses provides a broad assessment of the stress response for each species. DS *S. pyramidalis* mounted a messenger RNA (mRNA)-level response to an initial drop in hydration as has been observed for the majority of dehydration-sensitive plants (27, 28). However, as dehydration to 1.5 gH₂O g⁻¹ dw was reached, the transcript abundance response declined dramatically (Fig. 24), perhaps as the leaf water content reached a critical level for *S. pyramidalis*. The leaves of *S. pyramidalis* are wilted at 1.5 gH₂O g⁻¹ dw (21) but otherwise, appear undamaged, so it is tempting to speculate that the decline in the transcript abundance response may be related to wilting and perhaps, loss of turgor during wilting in *S. pyramidalis*. Although *S. pyramidalis* responds quickly to a loss of water, the early increased transcript abundance response appears to be focused on protein translational processes and transcripts common to heat and cold stress (SI Appendix, Fig. S4), and only later, as dehydration deepens, do transcripts associated with proline metabolism (osmoregulation) and redox proteins, common to water-deficit responses (27), accumulate. The early decline in transcripts involved in photosynthesis and

cell wall homeostasis is also common to the dehydration response in most angiosperms (4). The later decline in transcripts that are associated with general biosynthetic processes is consistent with the general lack of a metabolic response to dehydration seen in metabolite profiling studies of *S. pyramidalis* at similar levels of water loss (21). Desiccation-tolerant *S. stapfianus*, in contrast, exhibited a significantly different qualitative transcriptional response to dehydration with a low-magnitude response in the early phase of dehydration. With the comparatively muted response and although there are some common transcript abundance responses between the two species, *S. stapfianus* clearly targets remodeling a completely different functional aspect of the transcriptome than does *S. pyramidalis* at similar water contents. Indeed, it appears that *S. stapfianus* targets the accumulation of transcripts that function more in stress-related activities unlike *S. pyramidalis*, which does not. The differences between the two transcriptional responses for the two species were unexpected as other studies have indicated that there was extensive overlap in functionality of the transcriptomes of both sensitive and tolerant grasses exposed to dehydration (10). Although there are a few common transcript abundance functional categories in the early response to dehydration in both species, it is clear that the overall transcriptome remodeling during dehydration is very different between them, as exemplified by the different dehydration thresholds for the accumulation ELIP transcripts.

For *S. stapfianus*, the primary remodeling of the transcriptome during dehydration appears to occur as the plants reach the 1.0- to 0.75-gH₂O g⁻¹ dw part of the drying curve, which appears to be a critical period in the desiccation response of all resurrection angiosperms studied so far (29) and concurs with early microarray data (30). In *S. stapfianus*, the transition from 1.0 to 0.75 gH₂O g⁻¹ dw occurs during leaf curling (19) and is likely at water contents just prior to and during a change in membrane fluidity that occurs as leaf water potentials approach -12 MPa (4). The functional aspects of the transcriptome remodeling during desiccation of *S. stapfianus* leaves have been documented previously and are in accord with the observation that transcript abundance is concordant with changes in metabolism associated with cellular protection aspects of DT (30). There was a dramatic alteration of the transcriptome upon rehydration of *S. stapfianus* leaves, which likely reflects the complex nature of the dehydration event. The magnitude of the change in the transcriptome, reflecting a change in abundance of at least half of the known transcripts, and the functional processes they represent indicate not only the stress incurred from the inrush of water and mechanical aspects of cellular expansion but also, the need to repair damage (from both desiccation and rehydration), reactivate energy metabolism, and reinstate the physiological integrity of the cells and tissues (4). The observation that transcripts encoding proteins involved in ribosome biogenesis are accumulated and those encoding proteins involved in photosynthesis have not recovered control levels at 24 h following rehydration highlights the extent of the impact that desiccation and rehydration have on plant cells and tissues even in DT plants. *S. stapfianus* requires between 48 and 72 h to regain the structural and physiological integrity seen in well-watered plants (19, 31).

The remodeling of the transcriptome in response to dehydration starts from two very different resting-state (fully hydrated) transcriptomes. Our functional analysis of the gene-level expression of the syntenic orthologs of the sister grasses, although somewhat confounded by the structural complexity of the two genomes, revealed that for *S. stapfianus*, the biosynthesis of starch and nitrogen compounds was perhaps a priority for young leaves under normal conditions, while for young leaves of *S. pyramidalis*, the priority appeared more focused on the construction of cell walls. Although somewhat speculative, the

increase in nitrogen compounds, primarily amino acids from a combination of new synthesis and redistribution, was the focus of a recent study that demonstrated that these compounds are apparently used to fuel central metabolism or for other metabolic adjustments related to the acquisition of DT, such as osmoregulation (32). The differences in priorities are consistent with the changes in protein abundance from 3 to 1.5 gH₂O g⁻¹ dw. Although *S. pyramidalis* protein abundance changes did not reflect cell wall processes, perhaps due to the difficulty in extracting the majority of wall-related proteins (33), they show that *S. pyramidalis* was almost exclusively focused on the accumulation of stress response proteins. At the same desiccation stage, *S. stapfianus* had similarly accumulated stress response proteins but also, proteins involved in protein catabolism, and it had down-accumulated energy-related proteins, suggesting a scaling down, at the protein level, of the energy metabolism transcriptomic activity of the hydrated state and the continuation of N metabolism prioritization through protein salvage, possibly from misfolded proteins. Syntenic orthologs transcriptomic data are also consistent with information from the metabolomes of young leaves of these two grasses in that fully hydrated leaves of *S. stapfianus* were focused on the accumulation of a variety of amino acids and photosynthate derivatives, while for *S. pyramidalis*, the metabolome was focused on energy metabolism and growth (21). The conclusion from the metabolomics analyses was that leaves of *S. stapfianus* were prepared (primed) for a dehydration/desiccation event by accumulating osmolytes in times of water abundance and that *S. pyramidalis* needed to generate energy and components to support a faster growth rate, perhaps to deal with competition in its more mesic habitats. The hydrated transcriptome functional analysis fully supports this conclusion, and our transcriptomic and proteomic data, although somewhat speculative in nature, extend the hypothesis to include a focus on the maintenance of chloroplast function in *S. stapfianus* in the priming mechanism and cell wall biogenesis in *S. pyramidalis* as a target for the focus on energy metabolism and growth.

Although transcriptomic analyses were useful in comparing the functional aspects of the response to dehydration of the contrasting sister *Sporobolus* species and the desiccation and rehydration response of *S. stapfianus*, the availability of a high-quality genome for each of these two species allowed for a direct comparison of the genetic components (and their functions) of the response and allowed us to extend the comparison with other desiccation-tolerant and DS grass species. The broad comparison of the expression patterns of orthogroups and syntenic gene sets common in all five of the chloridooid grasses included in the analysis confirmed the disparate nature of the dehydration response between *S. stapfianus* and *S. pyramidalis*. It also revealed that the overall dehydration expression pattern for *S. stapfianus* was distinctly different from those observed for the other two desiccation-tolerant grasses, *E. nindensis* and *O. thomaeum*. The most recent phylogenetic analyses of the Chloridoideae indicate that the common ancestor of the Eragrostideae, which contains *E. nindensis* and *E. tef*, gave rise to the Zoysieae and the Cynodonteae, within which *O. thomaeum* resides; the Zoysieae then diversified into the Zoysiinae and the Sporobolinae, within which the *Sporobolus* clade containing both *S. stapfianus* and *S. pyramidalis* is located (18, 34). The phylogeny indicates that *O. thomaeum* and *S. stapfianus* are closer to one another than either are to *E. nindensis*, which is consistent with results of our analysis of orthogroups representing SDATs that increase in abundance. However, the overall expression response to dehydration for *O. thomaeum* appears to be more similar to the distantly (ancestrally) evolved response of *E. nindensis*. This might also explain why there is less overlap between the dehydration transcriptome of *S. stapfianus* and the transcriptomes of both sensitive and tolerant

grasses exposed to dehydration (10). Thus, although we have used only a three-way comparison, it does allow for the hypothesis that the desiccation response of *S. stapfianus* represents a more recent evolution of a mechanism for VDT within the Chloridoideae.

The orthogroup analysis of the SDATs that increase in abundance in all of the VDT species underscored the importance of most of the well-characterized processes that deliver cellular DT (4). The orthogroup analysis of the SDATs that increase in abundance in all of the DS species also reconfirmed what we understand of the response of most plants to a water deficit stress and highlighted the induction of senescence, which is thought to be blocked in resurrection angiosperms during desiccation (reviewed in ref. 4). However, the observation that transcripts classified as involved in protein folding accumulate in the VDT species and decline in abundance in the DS species indicates not only that maintaining protein structure is important in VDT, as has been well documented, but that the lack of the necessary components to do so might be critical in rendering a plant DS. The observation that all seed-related orthogroups are up-regulated in all VDT species and in one or more of the DS species reinforces the hypothesis that VDT likely evolved from a reprogramming of DT mechanisms that evolved in seeds (10).

Materials and Methods

Plant Material. *S. stapfianus* Gandoger (original provenance: Verena, Transvaal, South Africa) and *S. pyramidalis* Beauv. (also known as *Sporobolus indicus* var. *pyramidalis*) were grown and maintained as described in ref. 21. For genome sequencing, a single, healthy 3-mo-old fully hydrated plant from each species was selected, and young leaf tissue was collected, flash frozen in liquid N₂, and stored at -80 °C. For RNASeq experiments, seeds were collected from selfed clonal plants derived from the individuals used for the genome sequencing and germinated and plants grown to the 3-mo-old stage under greenhouse conditions (16-h light and day/night temperatures of 28 °C/19 °C).

Sampling and Drying Treatments. Plants were grown and maintained and seed stocks were increased (as described in ref. 35) in 1-gallon pots under greenhouse conditions. Three-month-old plants were subjected to a drying event by withholding water. *S. stapfianus* plants were dried until desiccated (after 3 wk), whereas *S. pyramidalis* plants were dried to a water content of ~1.5 gH₂O g⁻¹ dw before rewatering. Drying rates were as described by Oliver et al. (21) to simulate field drying rates that occur over a period of ~7 d to reach the 1.5-gH₂O g⁻¹ dw stage for both grasses and ~14 d for full desiccation of *S. stapfianus* (plants were left dry for a further 7 d). Young leaf tissue was collected at daily intervals, between 9 and 10 AM, from individual plants, flash frozen in liquid N₂, and stored at -80 °C. Dried plants were maintained dry for a week before rehydration. Duplicate samples were harvested for water content measurements at the time of sampling. The water content was calculated as fresh weight minus the dry weight (dried to equilibrium at 70 °C for 4 h). Triplicate samples were chosen for RNA extraction. Rehydration was achieved by placing the desiccated *S. stapfianus* plants under a continuous misting system in the greenhouse, and young leaves were sampled in triplicate at 12 and 24 h following the addition of water.

Estimation of the Genome Size (1C Value) Using Flow Cytometry. The genome size was estimated using the one-step flow cytometry procedure described in ref. 36. Approximately 1 cm² of leaf material from the *Sporobolus* species and leaf material of the calibration standard *Petroselinum crispum* (Mill.) Fuss (37) (haploid genome [1C] value = 2,201 Mbp) were diced in 1 mL of "general purpose buffer" (GPB) (38) supplemented with 3% polyvinylpyrrolidone of average molecular weight of 40,000. A further 1 mL of GPB was added, and the homogenate was filtered through a 30-μm nylon mesh (Celltrics 30-μm mesh; Sysmex); 100 μL propidium iodide (1 mg/mL) was added and incubated on ice for 10 min. The relative fluorescence of 5,000 particles was recorded using a Partec Cyflow SL3 flow cytometer (Partec GmbH) fitted with a 100-mW green solid-state laser (532 nm; Cobolt Samba). Three replicates of species were processed, and output histograms were analyzed using FlowMax software v.2.4 (Partec GmbH).

DNA Isolation, Library Preparations, and Sequencing. High-molecular weight DNA was isolated from 5 g of flash-frozen young leaf tissue using the PacBio

SampleNet-Shared Protocol (<https://www.pacb.com/support/documentation/>) as described. Random shotgun genomic libraries with various insert sizes, both paired end and mated pair libraries, were constructed for the Illumina HiSeq 2000 sequencing system (Illumina) according to the manufacturer's protocols. Sequencing of was conducted using an Illumina HiSeq 2500 ultrahigh-throughput DNA sequencing platform (Illumina) at the DNACore facility at the University of Missouri, Columbia, MO (<https://dnacore.missouri.edu/ngs.html>).

For Chicago sequencing, genomic DNA isolation, library preparation, sequencing, and assembly were conducted by Dovetail Genomics and are detailed in *SI Appendix, Methods*. Chicago genomic DNA libraries were prepared as described in ref. 39. Dovetail Hi-C libraries were prepared as described in ref. 40 after fixation of chromatin in place in the nucleus by incubation of leaf tissue for each species in 1% formaldehyde for 15 min under vacuum.

Assembly of the *S. stapfianus* and *S. pyramidalis* Genomes. A de novo assembly was constructed using a combination of paired end (mean insert size ~350 bp) libraries and mated pair libraries with inserts ranging from 7 to 12 kbp. De novo assembly was performed using Meraculous v2.2.2.5 (diploid mode 1) (41) with a k-mer size of 109. Reads were trimmed for quality, sequencing adapters, and mate pair adapters using Trimmomatic (42). The de novo assembly, shotgun reads, Chicago library reads, and Dovetail Hi-C library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (39) and detailed in *SI Appendix, Methods*.

RNA Extraction and RNA Sequencing. RNA was extracted from young leaf samples using the RNeasy (Qiagen) kit with RLC buffer following the manufacturer's protocol. The RNA isolates were treated with deoxyribonuclease 1 and cleaned using the DNA-free RNA Kit (Zymo Technologies). RNA quality was assessed by use of a fragment analyzer (Advanced Analytical Technologies), and concentration was determined with a Nanodrop Spectrophotometer (ThermoFisher). RNA libraries were individually bar-coded from 2.7 μg of template total RNA utilizing the TruSeq RNA Sample Prep Kit (Illumina) as described in the manufacturer's protocol. Libraries were pooled in groups of 12 and sequenced (12 samples per lane) on an Illumina HiSeq 2500 ultrahigh-throughput DNA sequencing platform (Illumina) at the DNACore facility at the University of Missouri.

Full-Length Sequencing (Iso-Seq) and De Novo Transcriptome Assembly. High-quality RNA was extracted from whole-root tissues obtained from seedlings at the four-leaf stage when the first pair of leaves had matured, whole seedlings at the two-leaf stage, mature leaves, young leaves, floral inflorescences, and tissue samples identical to those used for the dehydration/desiccation/rehydration transcriptomes. The RNAs were pooled for each individual species for subsequent amplification. Bar-coded SMRT libraries were prepared and sequenced on the PacBio platform with X SMRT cells by Novogene Corporation Inc. Sequence reads were processed using Iso-Seq3 (<https://github.com/PacificBiosciences/IsoSeq>).

Genome Annotation. Genome assemblies were annotated using three rounds of MAKER-P. Briefly, round 1 used full-length nonchimeric sequences from PacBio transcriptome sequencing as EST evidence; a collection of *Arabidopsis thaliana* [Araport11 (43)], *Zea mays* [downloaded from Gramene's ftp server at <https://www.gramene.org/ftp-download>; AGPv4 release 59 (44, 45)], *Sorghum bicolor* [downloaded from Phytozome; <https://phytozome-next.jgi.doe.gov/pz/portal.html>, version 3.1.1 (46)], and *O. thomaeum* [downloaded from Phytozome, version 1.0 (7)] sequences as protein evidence; and a de novo repeats library obtained using LTR_Finder (47), LTRharvest (48), LTR retriever (49), and RepeatModeler (50) as inputs. Round 2 used the round 1 maker gff file and an SNAP (<http://korflab.ucdavis.edu/software.html>) hmm file obtained from the round 1 gff3 file. Round 3 used the round 2 maker gff3 file, the GeneMark-ES (51) HMM output file from a BRAKER (52) run from hisat (53) aligned RNASeq reads, and the corresponding Augustus (54) gene prediction models.

As a further filter, we decided to only keep genes that had expression evidence in our RNASeq Illumina or Pacbio data and/or whose corresponding protein is homologous to a known plant protein. Evidence of expression was at least one of the following two criteria: 1) an expression value of at least one transcripts per million (TPM) in all replicates of at least one sample in the RNASeq data after bowtie2 (55) alignment and Salmon (56) quantification or 2) at least one TPM in the gtf file obtained after a minimap2 (57) alignment and StringTie (58) quantification of IsoSeq3 polished long reads. *Sporobolus* proteins were considered as homologous if they satisfied at least one of three criteria: 1) a blastp match with an e value of 1e-6 or lower vs. either

Arabidopsis proteins [Araport11 annotation (43)]; 2) vs. a collection of *Glycine max*, *Oryza sativa* subsp. *japonica*, *Populus trichocarpa*, *Solanum lycopersicum*, *S. bicolor*, *Vitis vinifera*, *Brachypodium distachyon*, *Physcomitrella patens* subsp. *patens*, and *Chlamydomonas reinhardtii* UniProt Trembl proteins; or 3) proteins with a domain identified by InterProScan (59) with an *e* value of 1e-10 or lower.

Final gene identifiers are in the format Sp2s00000_00000 for *S. pyramidalis* and Ss2s00000_00000 for *S. stapfianus*. Sp stands for *S. pyramidalis*, Ss stands for *S. stapfianus*, 2 indicates the genome version, s00000 indicates the scaffold number, and the last five digits are an arbitrary gene number.

Genome Functional Annotation. GO annotation was done using a simplified version of the maize–GAMER pipeline (60). Transcript sequences were analyzed using BLAST vs. *Arabidopsis* Araport11 proteins and a collection of UniProt (61) TREMBL proteins from nine plant species (*G. max*, *O. sativa* subsp. *japonica*, *P. trichocarpa*, *S. lycopersicum*, *S. bicolor*, *V. vinifera*, *B. distachyon*, *P. patens* subsp. *patens*, *C. reinhardtii*), InterProScan with the -goterms option, and Pannzer2 (62). GO annotations of BLAST reciprocal best hits were retrieved from either the *A. thaliana* gaf file available at <http://geneontology.org> or the GOA file available at European Bioinformatics Institute. GO annotations from Blast, InterProScan, and Pannzer2 analyses were collated into a nonredundant gaf file and used for GO enrichment analyses.

Orthogroup Analysis. Comparative genomics analyses were completed using MCScan (25). The *O. thomaeum* genome was used as a common anchor as it is diploid and has a chromosome scale assembly. A minimum cutoff of five genes was used to identify syntenic gene blocks. A set of syntenic orthogroups was created containing genes present in all grass species analyzed.

We clustered proteins from 23 species into orthogroups using OrthoFinder (v2.3.8) (26). OrthoFinder using default parameters and the reciprocal DIAMOND search was used to identify similar proteins, which were clustered using the Markov Cluster Algorithm. The following species were included in OrthoFinder: *Ananas comosus*, *A. thaliana*, *B. distachyon*, *E. nindensis*, *E. tef*, *L. brevidens*, *L. subracemosa*, *Marchantia polymorpha*, *Medicago truncatula*, *O. sativa*, *O. thomaeum*, *P. patens*, *S. bicolor*, *Setaria italica*, *Selaginella lepidophylla*, *Selaginella moellendorffii*, *S. lycopersicum*, *S. pyramidalis*, *S. stapfianus*, *V. vinifera*, *Xerophyta viscosa*, *Zostera marina*, and *Z. mays*.

A set of orthogroups containing seed-related genes was previously identified based on seed and leaf expression datasets from *Z. mays*, *S. bicolor*, *O. sativa*, and *E. tef* (22). Syntenic orthologs of these seed-related genes were then identified in *O. thomaeum*, and these syntenic orthologs were used with OrthoFinder output to identify seed-related orthogroups.

Cross-Species Expression Analyses. Differential expression (DE) analyses were conducted using DESeq2 (63) (*E. nindensis*, *E. tef*, and *O. thomaeum*) or edgeR (23) (*S. stapfianus* and *S. pyramidalis*), and resulting outputs were processed using Pandas 0.25.0 in Python 3.6.8. Up-regulated and down-regulated genes were extracted for each species (SI Appendix, Table S2). OrthoFinder output was used to identify the orthogroup corresponding to each gene in the differential expression output. For seed orthogroups, the previously generated lists of seed-related orthogroups were used to extract differentially expressed seed orthogroups. The intersections and differences among the resulting sets of orthogroups were then extracted, and Venn diagrams were constructed using matplotlib_venn (version 3.1.1) (64) or Python package venn. Enrichment of GO terms was conducted using topGO (65) 2.38.1 in R 3.6.0 for various intersections and differences of DE orthogroups (SI Appendix, Table S3).

Differentially expressed genes in these orthogroups were extracted, and GO enrichment was conducted using Fisher's exact test via the weight01 algorithm. Following enrichment, unique biological process GO terms were extracted using the Python library Pandas. Unique GO terms for DS as compared with DT were also extracted for further study.

GO Categories Enrichment Analysis for the *S. stapfianus* vs. *S. pyramidalis* Comparison. A comparison of gene expression of *S. stapfianus* vs. *S. pyramidalis* leaves at 3 gH₂O g⁻¹ dw was achieved using tximport (66) and edgeR (23). We created a custom syntenic orthologs tx2gene file (<https://bioconductor.org/packages/release/bioc/vignettes/tximport/inst/doc/tximport.html>). GO annotation files for both species were merged, replacing each gene identifier with the custom gene identifier from our tx2gene file. In this way, each gene inherits the GO annotation of all its corresponding *S. stapfianus* and *S. pyramidalis* genes (SI Appendix, Methods). GO categories enrichment analysis was carried out for the list of up-regulated both_n genes and the list of down-regulated both_n genes using Bingo (24) in Cytoscape (67), with a false discovery rate (FDR)-adjusted *P* value cutoff of 0.05 and the list of genes in our tx2gene file as the universe.

Protein Extraction and Proteomics Analysis. Proteins were extracted from triplicate samples of 1 g of frozen leaf tissue, separated on 16-cm sodium dodecyl sulfate polyacrylamide gel electrophoresis gels, and cut into 10 equal slices; each slice was digested with trypsin, and liquid chromatograph mass spectrometer (LCMS) data were acquired on the LTQ Orbitrap at the Charles W. Gehrke Proteomics Center, University of Missouri using standard protocols (<http://proteomics.missouri.edu/protocols.php>). Raw data were analyzed with MaxQuant software v. 2.0.1.0 (68). Tandem mass spectrometer spectra were searched against the *S. pyramidalis* and *S. stapfianus* proteins, and potential contaminants by the built-in Andromeda search engine (69). Label-free quantification (LFQ) of the identified proteins was performed using normalized LFQ (LFQ intensity) using the MaxLFQ algorithms (70). The resulting identified proteins were filtered, keeping only proteins with an LFQ intensity greater than zero in all biological replicates or absent in all biological replicates. Proteins with significant Student's *t* test (two tailed; *P* < 0.05) results were considered up accumulated (log₂ fold change > 0.5) or down accumulated (log₂ fold change < 0.5). The lists of up- and down-accumulated protein identifiers were translated to their corresponding syntenic ortholog identifiers, and GO biological process categories enrichment was done using Bingo previously.

Data Availability. The two genomes, Illumina DNA-seq, Chicago-seq, and Hi-C data are available from the CyVerse CoGe genome evolution site (<https://genomeevolution.org/CoGe/index.pl>, <https://genomeevolution.org/coge/GenomeInfo.pl?gid=54805>, and <https://genomeevolution.org/coge/GenomeInfo.pl?gid=54807>). The *S. stapfianus* RNASeq data and the *S. pyramidalis* RNASeq data can be found at the National Center for Biotechnology Information Sequence Read Archive (BioProject accession nos. PRJNA719555 and PRJNA719804, respectively) (71, 72).

ACKNOWLEDGMENTS. We acknowledge the expert technical assistance of Jim Elder in the preparation and growth of the plant material. We also thank Dr. Brian Mooney and the Charles W Gehrke Proteomics Center for their expertise in the proteomics analysis. This work was partially supported by Governor University Research Initiative Program of the State of Texas Grant 05-2018 (to L.R.H.E.), NSF Grant MCB-1817347 (to R.V.), and Agricultural Research Services Project 5070-21000-038-00D (to M.J.O.).

1. M. J. Oliver, Z. Tuba, B. D. Mishler, The evolution of vegetative desiccation tolerance in land plants. *Plant Ecol.* **151**, 85–100 (2000).
2. M. J. Oliver, J. Velten, B. D. Mishler, Desiccation tolerance in bryophytes: A reflection of the primitive strategy for plant survival in dehydrating habitats? *Integr. Comp. Biol.* **45**, 788–799 (2005).
3. C. F. Delwiche, E. D. Cooper, The evolutionary origin of a terrestrial flora. *Curr. Biol.* **25**, R899–R910 (2015).
4. M. J. Oliver et al., Desiccation tolerance: Avoiding cellular damage during drying and rehydration. *Annu. Rev. Plant Biol.* **71**, 435–460 (2020).
5. L. Xiao et al., The resurrection genome of *Boea hygrometrica*: A blueprint for survival of dehydration. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5833–5837 (2015).
6. M. D. Costa et al., A footprint of desiccation tolerance in the genome of *Xerophyta viscosa*. *Nat. Plants* **3**, 17038 (2017).
7. R. VanBuren et al., Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. *Nature* **527**, 508–511 (2015).
8. R. VanBuren, C. M. Wai, J. Keilwagen, J. Pardo, A chromosome-scale assembly of the model desiccation tolerant grass *Oropetium thomaeum*. *Plant Direct* **2**, e00096 (2018).
9. R. VanBuren et al., Desiccation tolerance evolved through gene duplication and network rewiring in *Lindernia*. *Plant Cell* **30**, 2943–2958 (2018).
10. J. Pardo et al., Intertwined signatures of desiccation and drought tolerance in grasses. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10079–10088 (2020).
11. Z. Xu et al., Genome analysis of the ancient tracheophyte *Selaginella tamariscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. *Mol. Plant* **11**, 983–994 (2018).
12. R. VanBuren et al., Extreme haplotype variation in the desiccation-tolerant clubmoss *Selaginella lepidophylla*. *Nat. Commun.* **9**, 13 (2018).
13. A. T. Silva et al., To dry perchance to live: Insights from the genome of the desiccation-tolerant biocrust moss *Syntrichia caninervis*. *Plant J.* **105**, 1339–1356 (2021).
14. R. VanBuren, J. Pardo, C. Man Wai, S. Evans, D. Bartels, Massive tandem proliferation of ELIPs supports convergent evolution of desiccation tolerance across land plants. *Plant Physiol.* **179**, 1040–1049 (2019).
15. M. C. D. Costa et al., Key genes involved in desiccation tolerance and dormancy across life forms. *Plant Sci.* **251**, 162–168 (2016).
16. E. Fischer, B. Schäferhoff, K. Müller, The phylogeny of Linderniaceae—The new genus *Linderniella*, and new combinations within *Bonnaya*, *Craterostigma*, *Lindernia*, *Micranthemum*, *Torenia* and *Vandellia*. *Willdenowia* **43**, 209–238 (2013).

17. D. Girma, G. Cannarozzi, A. Weichert, Z. Tadele, Genotyping by sequencing reasserts the close relationship between Tef and its putative wild *Eragrostis* progenitors. *Diversity (Basel)* **10**, 17 (2018).
18. P. M. Peterson, K. Romaschenko, Y. H. Arrieta, J. M. Saarela, A molecular phylogeny and new subgeneric classification of *Sporobolus* (Poaceae: Chloridoideae: Sporobolinae). *Taxon* **63**, 1212–1243 (2014).
19. D. F. Gaff *et al.*, *Sporobolus stapfianus*, a model desiccation-tolerant grass. *Funct. Plant Biol.* **36**, 589–599 (2009).
20. D. F. Gaff, M. Oliver, The evolution of desiccation tolerance in angiosperm plants: A rare yet common phenomenon. *Funct. Plant Biol.* **40**, 315–328 (2013).
21. M. J. Oliver *et al.*, A sister group contrast using untargeted global metabolomic analysis delineates the biochemical regulation underlying desiccation tolerance in *Sporobolus stapfianus*. *Plant Cell* **23**, 1231–1248 (2011).
22. R. VanBuren *et al.*, Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nat. Commun.* **11**, 884 (2020).
23. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
24. S. Maere, K. Heymans, M. Kuiper, BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).
25. H. Tang *et al.*, Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
26. D. M. Emms, S. Kelly, OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
27. E. A. Bray, Plant responses to water deficit. *Trends Plant Sci.* **2**, 48–54 (1997).
28. K. Shinozaki, K. Yamaguchi-Shinozaki, Gene expression and signal transduction in water-stress response. *Plant Physiol.* **115**, 327–334 (1997).
29. Q. Zhang, D. Bartels, Molecular responses to dehydration and desiccation in desiccation-tolerant angiosperm plants. *J. Exp. Bot.* **69**, 3211–3222 (2018).
30. A. Yobi *et al.*, *Sporobolus stapfianus*: Insights into desiccation tolerance in the resurrection grasses from linking transcriptomics to metabolomics. *BMC Plant Biol.* **17**, 67 (2017).
31. F. Dalla Vecchia, T. El Asmar, R. Calamassi, N. Rascio, C. Vazzana, Morphological and ultrastructural aspects of dehydration and rehydration in leaves of *Sporobolus stapfianus*. *Plant Growth Regul.* **24**, 219–228 (1998).
32. A. Yobi, A. Batushansky, M. J. Oliver, R. Angelovici, Adaptive responses of amino acid metabolism to the combination of desiccation and low nitrogen availability in *Sporobolus stapfianus*. *Planta* **249**, 1535–1549 (2019).
33. E. Jamet *et al.*, Isolation of plant cell wall proteins. *Methods Mol. Biol.* **425**, 187–201 (2008).
34. P. Peterson, K. Romaschenko, Y. H. Arrieta, A molecular phylogeny and classification of the Cynodonteae (Poaceae: Chloridoideae) with four new genera: *Orthacanthus*, *Triplasiella*, *Tripogonella*, and *Zaqiqah*; three new subtribes: *Dactylocteniinae*, *Oriniinae*, and *Zaqiqahinae*; and a subgeneric classification of *Distichlis*. *Taxon* **65**, 1263–1287 (2016).
35. P. J. O'Mahony, M. J. Oliver, The involvement of ubiquitin in vegetative desiccation tolerance. *Plant Mol. Biol.* **41**, 657–667 (1999).
36. J. Dolezel, J. Greilhuber, J. Suda, Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* **2**, 2233–2244 (2007).
37. R. Obermayer, I. J. Leitch, L. Hanson, M. D. Bennett, Nuclear DNA C-values in 30 species double the familial representation in pteridophytes. *Ann. Bot.* **90**, 209–217 (2002).
38. J. Loureiro, E. Rodriguez, J. Dolezel, C. Santos, Two new nuclear isolation buffers for plant DNA flow cytometry: A test with 37 species. *Ann. Bot.* **100**, 875–888 (2007).
39. N. H. Putnam *et al.*, Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).
40. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
41. J. A. Chapman *et al.*, Meraculous: De novo genome assembly with short paired-end reads. *PLoS One* **6**, e23501 (2011).
42. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
43. C.-Y. Cheng *et al.*, Araport11: A complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **89**, 789–804 (2017).
44. P. S. Schnable *et al.*, The B73 maize genome: Complexity, diversity, and dynamics. *Science* **326**, 1112–1115 (2009).
45. M. Law *et al.*, Automated update, revision, and quality control of the maize genome annotations using MAKER-P improves the B73 RefGen_v3 gene models and identifies new genes. *Plant Physiol.* **167**, 25–39 (2015).
46. R. F. McCormick *et al.*, The Sorghum bicolor reference genome: Improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* **93**, 338–354 (2018).
47. Z. Xu, H. Wang, LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
48. D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).
49. S. Ou, N. Jiang, LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
50. J. M. Flynn *et al.*, RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457 (2020).
51. A. Lomsadze, V. Ter-Hovhannisyanyan, Y. O. Chernoff, M. Borodovsky, Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* **33**, 6494–6506 (2005).
52. K. J. Hoff, A. Lomsadze, M. Borodovsky, M. Stanke, Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65–95 (2019).
53. D. Kim, B. Langmead, S. L. Salzberg, HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
54. M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
55. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
56. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
57. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
58. M. Perteu *et al.*, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
59. P. Jones *et al.*, InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
60. K. Wimalanathan, I. Friedberg, C. M. Andorf, C. J. Lawrence-Dill, Maize GO annotation-methods, evaluation, and review (maize-GAMER). *Plant Direct* **2**, e00052 (2018).
61. UniProt Consortium, UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
62. P. Törönen, A. Medlar, L. Holm, PANNZER2: A rapid functional annotation web server. *Nucleic Acids Res.* **46**, W84–W88 (2018).
63. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
64. J. D. Hunter, Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
65. A. Alexa, J. Rahnenfuhrer, topGO: Enrichment Analysis for Gene Ontology (R Package Version 2, 2010). <http://www.bioconductor.org/packages/release/bioc/html/topGO.html>. Accessed 20 January 2022.
66. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000 Res.* **4**, 1521 (2015).
67. M. E. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang, T. Ideker, Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).
68. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
69. J. Cox *et al.*, Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
70. J. Cox *et al.*, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
71. M. J. Oliver, *Sporobolus stapfianus* genome sequencing, assembly, and dehydration/rehydration transcriptomes. NCBI. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA719555>. Deposited 10 January 2022.
72. M. J. Oliver, *Sporobolus pyramidalis* Genome Sequencing and Transcriptome project. NCBI. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA719804>. Deposited 10 January 2022.