the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2020) doi: 10.1111/tpj.15116

To dry perchance to live: Insights from the genome of the desiccation-tolerant biocrust moss *Syntrichia caninervis*

Anderson T. Silva[1], Bei Gao[2], Kirsten M. Fisher[3], Brent D. Mishler[4], Jenna T. B. Ekwealor[4], Lloyd R. Stark[5], Xiaoshuang Li[2], Daoyuan Zhang[2], Matthew A. Bowker[6], John C. Brinda[7], Kirsten K. Coe[8] and Melvin J. Oliver[1,9*]

[1]*Division of Plant Sciences and Interdisciplinary Plant Group, University of Missouri, Columbia, Missouri 65211, USA,*
[2]*State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Science, Urumqi 830011, China,*
[3]*Department of Biological Sciences, California State University, Los Angeles, California 90032, USA,*
[4]*Department of Integrative Biology, University and Jepson Herbaria, University of California, Berkeley, California 94720-2465, USA,*
[5]*School of Life Sciences, University of Nevada, Las Vegas, Nevada 89154-4004, USA,*
[6]*School of Forestry, Northern Arizona University, Flagstaff, Arizona 86011, USA,*
[7]*Missouri Botanical Garden, St. Louis, Missouri 63110-0299, USA,*
[8]*Department of Biology, Middlebury College, Middlebury, Vermont 40506-0225, USA, and*
[9]*USDA-ARS-MWA, Plant Genetics Research Unit, Columbia, Missouri 65211, USA*

Received 8 July 2020; accepted 30 November 2020.
*For correspondence (e-mail olivermj@missouri.edu).

SUMMARY

With global climate change, water scarcity threatens whole agro/ecosystems. The desert moss *Syntrichia caninervis*, an extremophile, offers novel insights into surviving desiccation and heat. The sequenced *S. caninervis* genome consists of 13 chromosomes containing 16 545 protein-coding genes and 2666 unplaced scaffolds. Syntenic relationships within the *S. caninervis* and *Physcomitrella patens* genomes indicate the *S. caninervis* genome has undergone a single whole genome duplication event (compared to two for *P. patens*) and evidence suggests chromosomal or segmental losses in the evolutionary history of *S. caninervis*. The genome contains a large sex chromosome composed primarily of repetitive sequences with a large number of *Copia* and *Gypsy* elements. Orthogroup analyses revealed an expansion of *ELIP* genes encoding proteins important in photoprotection. The transcriptomic response to desiccation identified four structural clusters of novel genes. The genomic resources established for this extremophile offer new perspectives for understanding the evolution of desiccation tolerance in plants.

Keywords: *Syntrichia*, bryophyte, genome, desiccation tolerance, abiotic stress, comparative genomics, transcriptomics.

INTRODUCTION

Increasing water scarcity and climate change loom over efforts to improve global food security (Gosling and Arnell, 2016) and the need for more resilient crops has led to a surge in interest in abiotic stress tolerance mechanisms and extremophiles (Dhanker and Foyer, 2018). The search for crop improvement strategies has drawn attention to how plants prevent or repair cellular damage resulting from dehydration. Vegetative desiccation tolerance (DT), the ability to survive drying to −100 MPa and lower (Alpert and Oliver, 2002), is rare in the vascular plants (Procter and Pence, 2002), but less so in bryophytes (Wood, 2007).

To date, genomes have been published for six DT plants, often termed resurrection plants: five angiosperms, *Boea hygrometrica* (Xiao *et al.*, 2015), *Oropetium thomaeum* (VanBuren *et al.*, 2018a), *Xerophyta schlechteri* (Costa *et al.*, 2017a), *Lindernia brevidens* (VanBuren *et al.*, 2018b), and *Eragrostis nindensis* (Pardo *et al.*, 2020); and two lycophytes, *Selaginella tamariscina* (Xu *et al.*, 2018) and *Selaginella lepidophylla* (VanBuren *et al.*, 2018c). With such few resurrection plant genomes available, it is difficult to identify a 'genomic blueprint' that characterizes a DT plant, though the expansion of certain gene families, such as those for early light-inducible proteins (ELIPs) and late embryogenesis abundant (LEA) proteins, has been

linked to the DT phenotype (Costa *et al.*, 2016; VanBuren *et al.*, 2019).

It is thought that vegetative DT in the angiosperms evolved via reprogramming of the genetic networks controlling DT in seeds (Costa *et al.*, 2017b), and thus the capacity for DT may be present in the genomes of all angiosperms. Because DT is an ancient trait (Oliver *et al.*, 2005), its associated features are difficult to identify in genomes that are distant from the ancestral condition; thus, there is a need to better characterize the genomic context of DT in bryophytes. DT in bryophytes exhibits a high degree of plasticity (Stark, 2017) and is best represented as a 'norm of reaction' controlled by external and internal environmental parameters (Sultan, 2000). A case in point is the moss *Physcomitrella patens*; under standard equilibrium drying it can only withstand approximately −13 MPa (Koster *et al.*, 2010) but it can survive desiccation (to −100 MPa and lower) if dried extremely slowly (Xiao *et al.*, 2018). The physiologically inducible DT of *P. patens* approaches the desiccation-sensitive end of a gradient of inducibility called the 'continuum hypothesis of ecological desiccation tolerance' by Stark and Brinda, 2015. At the opposite end of this gradient of inducibility are species with a constitutive physiological (ecological) expression of DT such as the desert moss *Syntrichia caninervis* (Stark and Brinda, 2015), an important component of the biological soil crusts of both North American and Asian dryland ecosystems.

Here we report a chromosomal level assembly of the genome of *S. caninervis* derived from a clonally propagated female gametophyte. We developed this genomic resource to better understand the ability of *S. caninervis* to survive desiccation. One objective was to determine if there is an underlying genetic structure to DT that can be used to infer the evolution of this trait in land plants. To address this aim, we performed a comparative analysis of the *S. caninervis* and *P. patens* genomes, as well as a comprehensive analysis of gene families across the genomes of several plant species. In the context of the genome, we also characterized the transcriptomic response of *S. caninervis* to abiotic stress events, with a focus on DT. Mapping dominant patterns (DPs) of transcriptomic responses to the genome revealed structural features and novel transcripts that may have an important role in the ability of *S. caninervis* to survive desiccation.

RESULTS

Chromosome level genome assembly

The estimated size of the *S. caninervis* genome assembly was 331.83 Mb. It consisted of 2679 scaffolds with an N50/L50 of 20.459 Mb and seven scaffolds with an N90/L90 of 6.845 Mb, with the longest scaffold at 37.3 Mb. The bulk of the assembled genome consisted of 13 scaffolds, ranging in size from 14.9 to 37.3 Mb in length (Table S1). As shown in Figure S1, there was insufficient proximity ligation signal to join the remaining 2666 scaffolds (ranging in size from 1 to 6.8 Mb in length) to the larger 13 scaffolds. The 2666 unplaced scaffolds were characterized as more related to each other than to the 13 larger scaffolds and exhibited a significantly higher read density (Figure S1). The 13 larger scaffolds generated a 292.20 Mb smaller genome assembly (88% of the 331.83 Mb) that is considered near chromosome scale, with the 13 scaffolds, hereafter referred to as chromosomes, numbered according to size (Table 1).

Genome annotation revealed a total of 25 030 gene models in the 331.83 Mb genome that includes the 13 assembled chromosomes and the unplaced scaffolds; 16 545 gene models were annotated in the 13 chromosomes and 8486 gene models were annotated in the unplaced scaffolds. In an attempt to establish the identity of the unplaced scaffolds, the 8486 gene sequences in the unplaced scaffolds were subjected to a BLASTn (E-value 1^{-25}) search of the NCBI nt database and 6723 retrieved significant similarity hits. The vast majority of the similarity hits, 6324 (98.3%), were associated with bacterial genes and only 105 (1.56%) exhibited similarity to plant-related genes (Figure S2). The 105 plant-related genes were scattered across 70 unplaced scaffolds with 81 of the 105 genes (on 43 scaffolds) exhibiting significant similarity to *Syntrichia* chloroplast and mitochondrial genes deposited in the NCBI database (Table S2). BLASTn searches using the entire scaffold DNA sequences directly revealed only

Table 1 Statistics of *Syntrichia caninervis* genome assembly and annotation

Feature	Value
Genome assembly	
Genome assembly size (without the 2666 scaffolds considered to be non-*S. caninervis* DNA)	292.20 Mb
Number of chromosome level scaffolds	13
Longest chromosome	37.31 Mb
GC content	41.3%
Genome annotation	
Repetitive regions in genome	53.6%
Number of predicted gene models (with isoforms)	17 202
Number of protein-coding genes	16 545
Mean transcript length (mRNA)	1513 bp
Mean coding sequence length	1251.2 bp
GC content of coding sequences	57.86%
Average number of exons per gene	4.86
Mean exon length	317 bp
Mean intron length	280 bp
Gene functional annotation	
InterProScan	17 202
Gene Ontology	10 458
Transcription factors	542
With *P. patens* blast Hit	15 793

partial sequence similarity for 445 scaffolds and only 226 generated partial sequence similarity to plant DNA sequences. Transcriptomic analysis (RNA-Seq) revealed that after trimming reads for adaptor and sequence quality, 90% (average of 30 million) mapped to the whole *S. caninervis* genome (13 scaffolds and 2666 unplaced scaffolds) giving an average genome coverage of approximately 8×. While 99% of the total reads (MAPQ > 20) mapped to the 13 chromosomes, only 1% mapped to the unplaced scaffolds (Table S3).

The 16 545 gene models located on the 13 chromosomes were identified as plant nuclear encoded protein-coding genes that had an average coding sequence length of 1513 bp. Only 25 (0.15%) of the predicted genes lacked transcriptomic evidence, supporting the gene model, and 14 557 of the gene models (88%) had Annotation Edit Distance (AED) (Eilbeck *et al.*, 2009) values of ≤0.3. Using Benchmarking Universal Single-Copy Orthologs (BUSCO) (Waterhouse *et al.*, 2018), the genome annotation recovered 403 orthologs of the 430 highly conserved core proteins in the Viridiplantae, of which 379 (94%) were single-copy genes and 24 (6%) were duplicated. We identified 56.05% of the assembly as repeat regions (Figure 1; Table S4) composed of 25 identifiable repeat families. We also identified 122 intact long terminal repeat (LTR) retrotransposons within the assembly.

Of the 16 545 nuclear protein-coding genes predicted in *S. caninervis*, 15 793 (95.4%) demonstrated at least partial predicted protein sequence similarities (E-value < 1e−5) to *P. patens* proteins and all were annotated with InterProScan domain/family information (Table S5). Gene Ontology (GO) descriptions were also assigned to 10 458 of the *S. caninervis* genes. The protein length distribution in *S. caninervis* is comparable to that of *P. patens*, with slightly fewer proteins longer than 500 amino acids identified in the *S. caninervis* protein set. In addition, 290 tRNA, 150 snRNA, 20 microRNA, and 23 signal recognition particle (SRP) genes were detected (Table S6).

Of the 16 545 coding genes, 2452 are located within tandem repeats (Table S7) ranging from 2 to 15 genes. Of the six longest tandem repeats (9–15 genes), three occur on chromosome 4 and contain genes encoding protein kinases involved in signaling; the remaining three of the six longest tandem repeats occur on chromosomes 2, 6, and 9. Several of the tandem repeats include tandem duplications, for example, the 15-gene LRR receptor-like serine/threonine-protein kinase repeat contains three genes in succession separated from a second array of 10 genes by an unrelated two-gene spacer sequence.

The *S. caninervis* genome contains 35 *ELIP* genes, all of which annotate to the PTHR14154:SF5 ELIP ID and of which 23 have the conserved chloral b-binding protein domain (IPR022796). Of the 35 *ELIP* genes, 29 (83%) are in tandem repeats (Table S8).

There are 51 *LEA* genes in the *S. caninervis* genome, though one, Sc_g00795 on chromosome 12, has nine isoforms. Based on the presence of conserved LEA protein domains, InterPro classified 25 of the *LEA* genes into six families: two *DHN* (IPR000167), eleven *LEA_2* (IPR004864), one *LEA_3* (IPR004926), one *LEA_4* (IPR004238), nine *LEA_5* (IPR038956), and one *SMP* (IPR007011) (Table S9). Of the remaining 26 *LEA* genes, three were annotated as *LEA-1* (D-29, D-34), nine were annotated as *LEA-3* genes, three in the *LEA-5* gene family, two as *LEA-14* genes, and nine had no family designation. Overall, only 12 *LEA* genes appear in tandem repeat configurations (Table S7), the largest consisting of five *LEA-5* genes arranged in a direct repeat on chromosome 11.

Chromosome 13 – the putative sex chromosome

Chromosome 13 is the largest of the chromosomes with 37.3 Mb. It has the smallest number of protein-coding genes (386) and presents the largest percentage of LTRs of all the chromosomes, including 483 *Gypsy* and *Copia* elements. Sequence analysis revealed the existence of an approximately 700-bp sex-linked locus (Baughman *et al.*, 2017) annotated as part of Sc_g00229. The sequence, which covers both introns and exons of Sc_g00229, is identical to the female allele of the locus and is 18% and 4% divergent from the male allele in the intron and exon sequences, respectively (Figure S3). The complete identity match between the female allele of the sex-linked locus and Sc_g00229 coupled with the absence of any other locus in the *S. caninervis* genome with significant sequence similarity to the female or the male alleles suggests that chromosome 13 represents the non-recombining female U chromosome.

Comparative genomics and gene family analyses

The intra- and interspecies collinearity comparison analyses of the *S. caninervis* genome, using protein-coding genes as anchors, revealed evidence for an ancestral polyploidization of the genome. The intragenomic synteny analysis (Figure 1; Table S1) determined strong syntenic relationships between pairs of chromosomes with the exceptions of chromosomes 1, 8, and 13. The strongest syntenic pairings are between chromosomes 2 and 6, 3 and 5, 4 and 11, 7 and 9, and 10 and 12, indicating that at least one polyploidization event had occurred in the evolution of the genome.

Ks frequency analyses between *S. caninervis* and *P. patens* identified a predominant *Ks* peak at approximately 1.3 of duplicated syntelog pairs. The orthologous *Ks* peak between *S. caninervis* and *P. patens*, corresponding to their most recent common ancestor, was estimated at approximately 1.0 (Figure 2(a).1), suggesting that the whole genome duplication (WGD) event in mosses predated their divergence and is ancestrally shared by the two

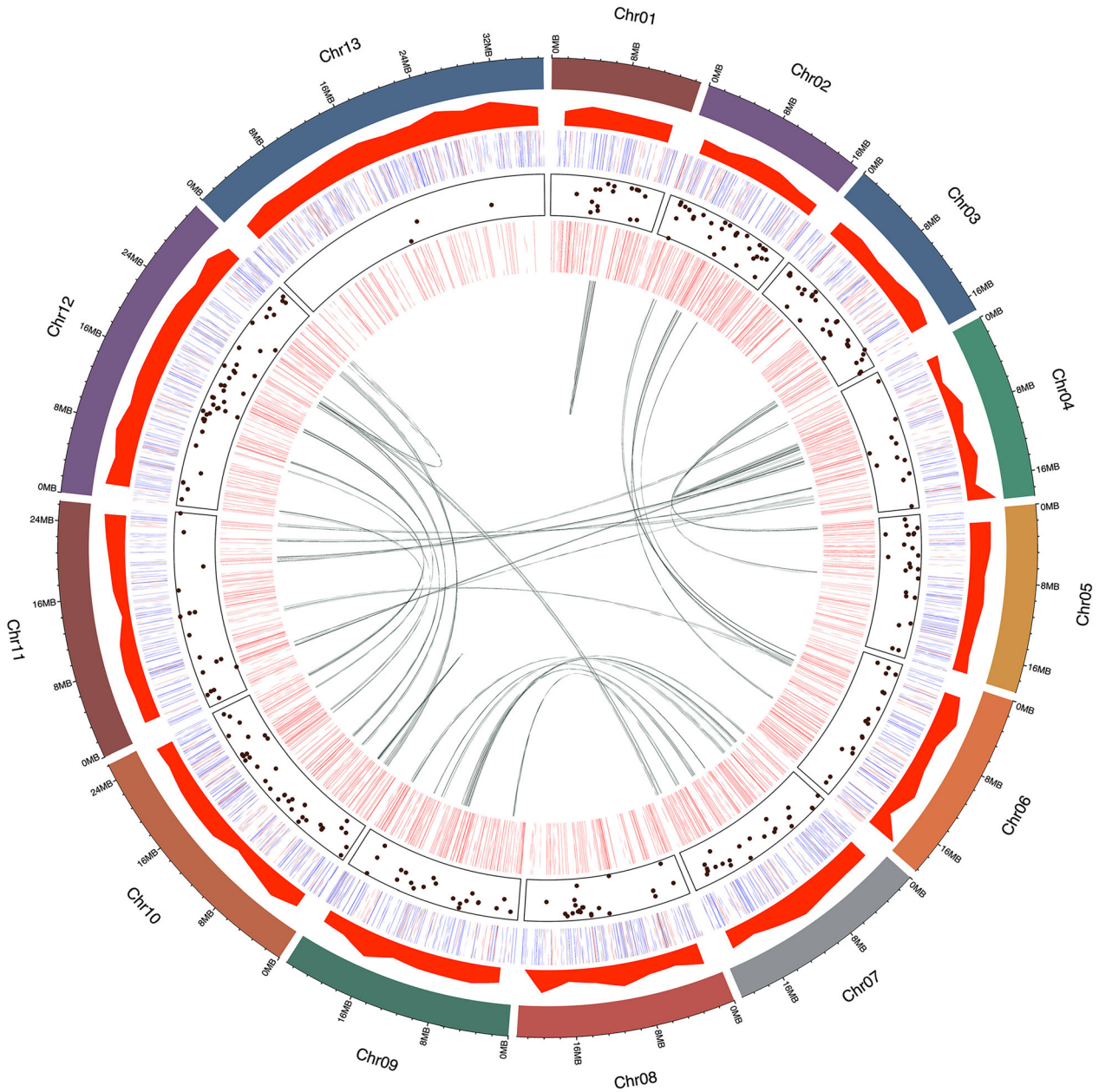

Figure 1. Characteristics of the *Syntrichia caninervis* genome. Circos plot depicting the structural characteristic of the *S. caninervis* genome. From outer to inner: The 13 scaffolds assembled into chromosomes; followed by GC content fluctuations in every 100-kb genomic region; LTR elements coverage as stacked histograms (red, *Copia*; blue, *Gypsy*); tRNA genes; orphan genes not clustered with other land plants; and interchromosome genomic collinearity depicted by connecting lines.

genomes. The fourfold degenerate transversion rate (4DTv) frequency distributions for the paralogous (4DTv peak approximately 0.43) and orthologous (4DTv peak approximately 0.3) divergence levels (Figure 2(a).2) also support this 'shared WGD' hypothesis.

The intergenomic synteny analysis between the genomes *S. caninervis* and *P. patens* detected a total of 1003 syntenic blocks with 5412 gene pairs (Figure 2(b); Table S10). There are clear syntenic relationships between each of the *S. caninervis* chromosomes and several of the chromosomes of *P. patens* (Figure 3). The syntenic depth analyses can be summarized as a 1:2 pattern (Figure 2(c)), which is consistent with the more recent *P. patens* species-specific WGD event (Pp-WGD2; Lang *et al.*, 2018). In addition, we detected 16 genes in *S. caninervis* that exhibited synteny with a maximum of four genes in the *P. patens* genome (Figures 2(c), 3; Table S10), corresponding to two successive WGDs in *P.*

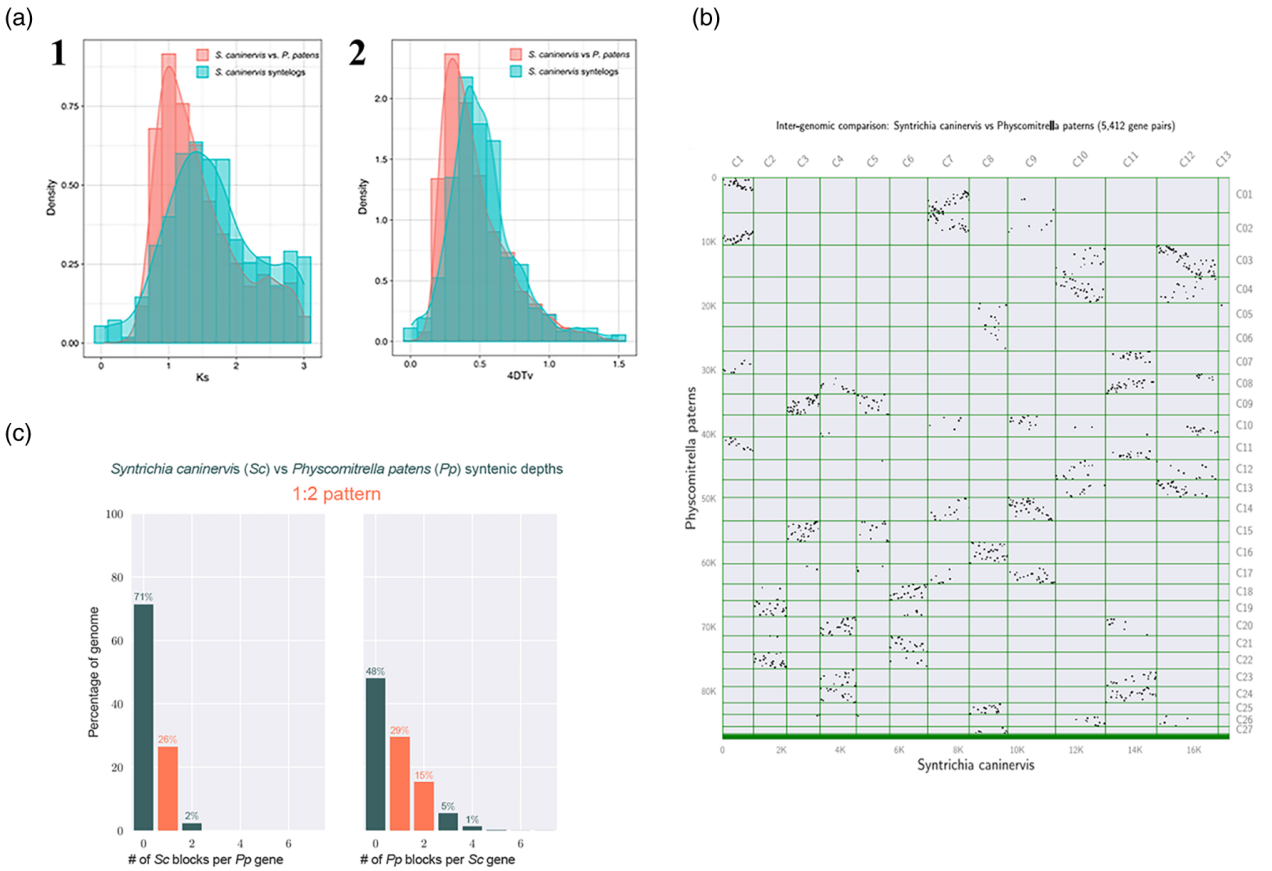

Figure 2. Comparative genomics between *S. caninervis* and *P. patens*. (a) Comparison of the genomes of *S. caninervis* and *P. patens* showing the frequency distributions of (1) syntenic paralogous and orthologous synonymous substitutions per synonymous site (*Ks*) and (2) fourfold degenerate transversion rate (4DTv). (b) Dot plot graph for genome synteny comparison revealed shared genome duplication in *Syntrichia caninervis* and *Physcomitrella patens*. A major syntenic relationship of 1:2 of *Syntricha caninervis* to *Physcomitrella patens* was observed, whereas two chromosomes (Chr1 and Chr8) revealed 1:4 syntenic relationships with *Physcomitrella patens*. (c) Synteny depth analyses between *Syntrichia caninervis* (*Sc*) and *Physcomitrella patens* (*Pp*) genomes. Refers to the number of times a genomic region is covered by synteny blocks against another genome. Syntenic depth of the *Syntrichia caninervis* genome for each *Physcomitrella patens* gene (left) and in the *Physcomitrella patens* genome for each *Syntrichia caninervis* gene (right).

patens and chromosomal or segmental losses in *S. caninervis*.

Using the transcription factor (TF) family classification schemes established in the PlantTFDB database (Jin *et al.*, 2017), we identified a total of 542 genes in 50 TF families (Table S11) for the *S. caninervis* genome, in contrast to 1156 TF genes that have been reported for the *P. patens* genome. A detailed comparison of the *S. caninervis* to *P. patens* TF genes revealed a family-size ratio of 1:2 for the majority of the TF families (Figure 4).

OrthoFinder2 (Emms and Kelly, 2019) assigned 321 783 (79%) of the input genes to 15 765 OGs across the 15 species (Table S12). Of these, 4276 OGs had representative genes from all 15 species. Gene family comparisons revealed that significantly more *S. caninervis* OGs were shared with the mosses *P. patens* (14 004 OGs) and *Sphagnum fallax* (12 573 OGs) than with other species. *Syntrichia caninervis* had the smallest number of species-specific OGs (4) and the species-specific OGs contained fewer genes on average (19) than similar families in other species (Table S12). The *S. caninervis* genome contained 1526 highly divergent or species-specific 'orphan genes' that could not be clustered into any OG, 686 (45%) of which could not be assigned a GO ID (Figure 1 and Table S13). Of the 17 most frequently occurring GO biological process IDs in the orphan gene set, nine were found to significantly exceed ($P < 0.001$) their expected mean occurrence based on bootstrapped subsamples from the whole genome gene set (Table 2). Included in the orphan genes that have been ascribed a function are three of the 35 *ELIP* genes (Sc_g00945, Sc_g04799, and Sc_g15046) as well as a genes encoding antioxidant enzymes, P450 proteins, kinases (including three LRR receptor-like serine/threonine-protein kinases, one of which is located in the 15-gene tandem repeat), heat shock proteins, and several ethylene-responsive TFs.

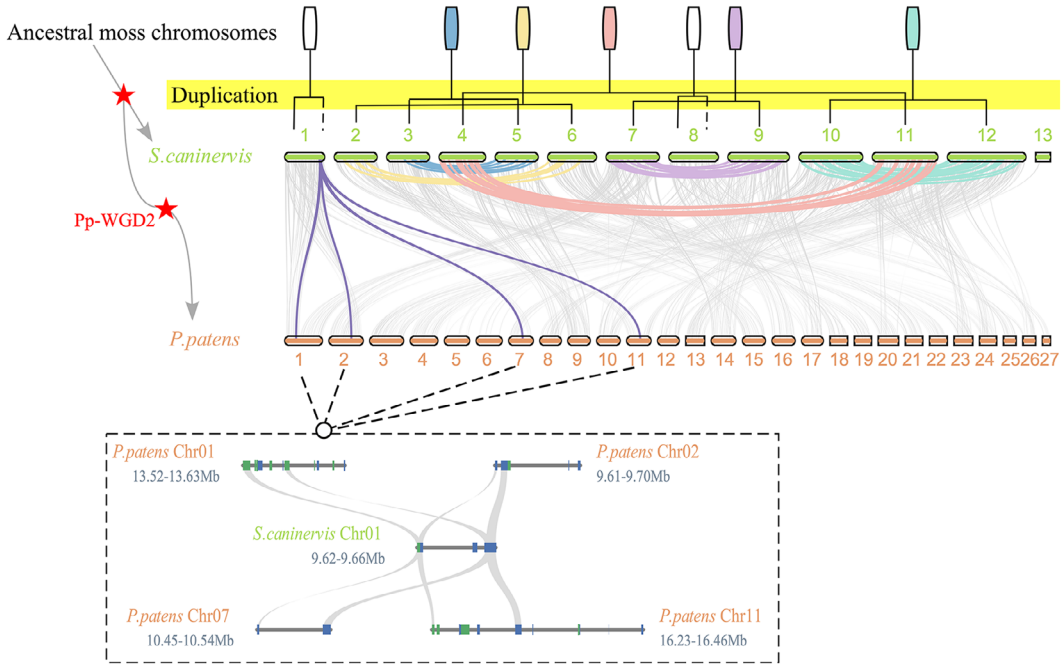

Figure 3. Intra- and interspecies genomic synteny comparisons with *Physcomitrella patens*. The syntenic comparison depicts the shared Bryopsida-wide ψ duplication (highlighted in yellow) of the postulated seven ancestral chromosomes, followed by two chromosome losses in *S. caninervis* and a more recent WGD event in *P. patens* (denoted as Pp_WGD2). Colored lines track intragenomic synteny within the *Syntrichia caninervis* genome; gray lines track the inter-species synteny between the *Syntrichia caninervis* and *Physcomitrella patens* genomes. Insert depicts the areas where the Pp_WGD2 event resulted in syntenic depth relationships of 1:4 between a number of regions between the two species.

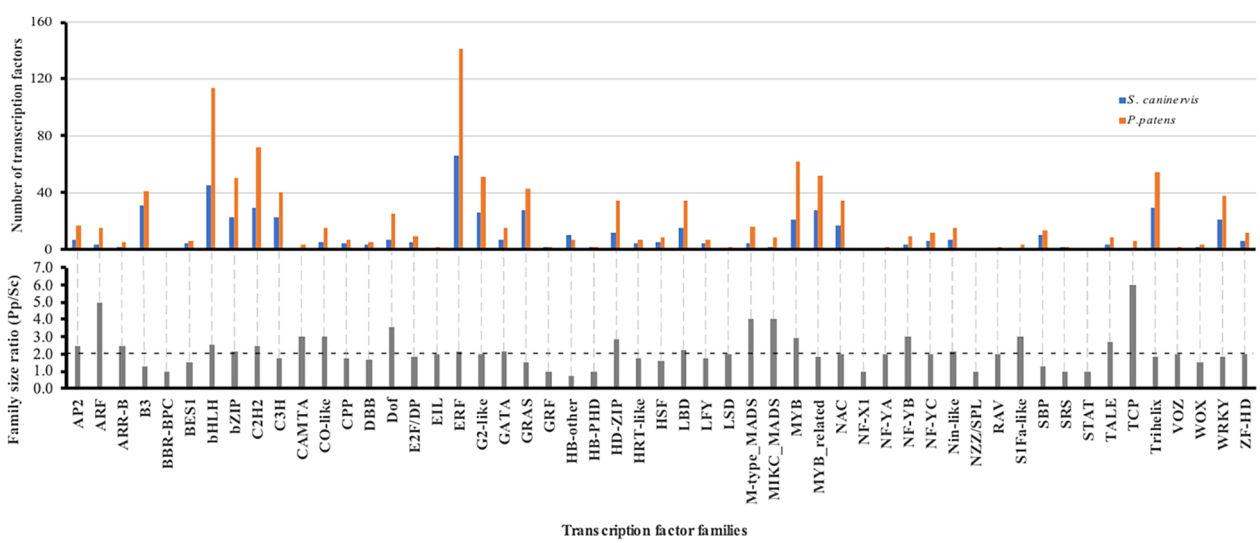

Figure 4. Comparison of transcription factor families in the *Syntrichia caninervis* and *Physcomitrella patens* genomes. The upper panel of the graph details the number of transcription factors (*y*-axis) in each transcription factor gene family (*x*-axis); blue columns indicate *Syntrichia caninervis* transcription factor genes, orange columns indicate *Physcomitrella patens* transcription factor genes. The lower panel details the ratio of transcription factor genes in *Physcomitrella patens* versus transcription factor genes for *Syntrichia caninervis* in each transcription factor family. Most transcription factor families are represented by fewer genes in *S. caninervis* than *P. patens*. And family size ratios were around two, in parallel with the recent *P. patens* species-specific WGD event.

Gain, loss, expansion, and contraction of *ELIP* and *LEA* gene family members were assessed in OGs (Table S14). Of the 35 *ELIP* genes in the *S. caninervis* genome, five were classified as orphan genes and the remaining 30 could be assigned to two different OGs, 29 to OG0000141, and three to OG0009798 (Tables S9, S14). Within the 51

Table 2 Gene Ontology biological functions significantly overrepresented in the set of *Syntrichia caninervis* genes lacking clear orthologs in the other plant genomes investigated (*t*-test; $P < 0.001$)

Gene Ontology biological function ID	Description	Bootstrap occurrence mean	95% confidence interval	Occurrences in orphan set
P:GO:0006468	Protein phosphorylation	81.423	80.87238–81.97362	102
P:GO:0006508	Proteolysis	30.856	30.51288–31.19912	36
P:GO:0006457	Protein folding	12.587	12.36977–12.80423	22
P:GO:0009651	Response to salt stress	10.304	10.10903–10.49897	21
P:GO:0042542	Response to hydrogen peroxide	5.815	5.665726–5.964274	17
P:GO:0009408	Response to heat shock	7.492	7.327433–7.656567	17
P:GO:0051259	Protein complex oligomerization	4.488	4.357918–4.618082	16
P:GO:0006511	Ubiquitin-dependent protein catabolic process	8.272	8.091566–8.452434	12
P:GO:0006486	Protein glycosylation	8.036	7.865262–8.206738	12

LEA genes, two were classified as orphan genes and the remaining 49 *LEA* genes were assigned to 25 different OGs (Tables S10, S14). The genes in the main *ELIP* OG, OG0000141, are present in the genomes of 13 of the 15 species utilized in the OG analysis (Table S14). Within the 25 OGs that *LEA* genes of *S. caninervis* populate, only two contain significant numbers of genes from *S. caninervis*: OG0000714 (nine) and OG0000931 (eight) (Table S14). Nine of the 15 genomes analyzed contained *LEA* genes in OG0000714, and only the species that can acquire DT under certain conditions have significant numbers of genes within this OG. Of those species that exhibit vegetative DT, only *S. lepidophylla* (a DT lycophyte) has *LEA* genes that fall within OG0000714. The genomes from all 15 species have genes that are contained within OG0000931 but there is no relationship between the number of genes in this OG and the DT phenotype of the plants (Table S14).

Transcriptomic analysis identifies stress response signatures

A total of 11 113 transcripts (representing 67% of the protein-coding genes) exhibited significantly altered abundance levels ($P < 0.05$) in the treated (stressed) samples in comparison to the abundance levels in the hydrated control samples grown at 20°C. Of the 11 113 transcripts, 6994 increased in abundance, 7011 decreased in abundance, and 2842 increased in one stress condition but declined in another (Table S14 and Figure S4). Individual environmental stresses varied in their impact on transcript abundance. Mapping the responsive transcripts for each individual stress treatment did not reveal any recognizable clustering pattern of stress-responsive genes within the genome.

Of the transcripts representing the 35 *ELIP* genes in the *S. caninervis* genome, 16 accumulated significantly in response to slow drying and two declined in abundance in comparison to the hydrated controls (Tables S9, S15). Only one *ELIP* transcript (Sc_g02119 on Chr12) accumulated in response to rehydration and one transcript for Sc_g10331 on Chr6 accumulated in response to the cold treatment. Of those ELIP transcripts that responded to either elevated temperature or heat shock, all exhibited a significant decline compared to hydrated controls.

The alteration in the abundance of LEA protein transcripts exhibited a more complex pattern (Tables S9, S15). Transcripts for 26 of the 51 *LEA* genes significantly increased in abundance during slow drying and one (Sc_g13301 on Chr4) declined. Only one of the two genes annotated as dehydrins, dehydrin 5 (Sc_g01426 on Chr12), responded to dehydration. Transcripts for 18 *LEA* genes increased in abundance during rehydration, and all but one (Sc_g04910 on Chr10) also increased in abundance during drying. Only two LEA protein transcripts, Sc_g00802 on Chr12 and Sc_g10451 on Chr6, increased in abundance in response to the cold treatment; none decreased in abundance during this treatment. Eight LEA protein transcripts increased and 18 decreased in abundance in response to the elevated temperature treatment and 14 increased and 10 decreased in abundance in response to the heat shock treatment. Of the *LEA* genes that were located in tandem repeats, transcripts for only two genes, Sc_g03602 and Sc_g03607, within the largest *LEA* gene (encoding em-like protein GEA1) repeat exhibited significant changes in abundance in response to the stress treatments (Table S16).

The abundance of transcripts derived from the genes that are situated within the six longest tandem repeats exhibited significant changes in response to the abiotic stress treatments but not all transcripts for each gene within the repeat respond in the same fashion (Table S16). As an example, in the largest *ELIP* gene tandem repeat on chromosome 12, only transcripts from three of the eight genes exhibited significant changes in abundance levels in response to the stress treatments (Sc_g20118, Sc_g20119, and Sc_g20120). All three genes exhibited an increase in abundance during slow drying and decreased in abundance in response to the elevated temperature treatment.

Transcripts for Sc_g20119 increased in abundance during rehydration and two declined in response to heat shock (Sc_g20118 and Sc_g20119).

DP analysis generated 11 DPs (Figure 5) from the subset of 3972 transcripts that were changed in abundance in at least one treatment, relative to the hydrated control (Table S17). Five of the 11 DPs exhibited co-response patterns in transcript abundance levels associated with a single stress treatment. Pattern DP2 (244 transcripts) exhibited increased transcript abundance predominantly associated with desiccation; DP10 (93 transcripts) exhibited increased transcript levels predominantly associated with rehydration; DP11 exhibited reduced transcript abundance predominantly associated with desiccation, and two patterns, DP5 (497 transcripts) and DP7 (452 transcripts), exhibited increased transcript levels predominantly associated with elevated temperatures, both 30°C and 35°C. The remaining seven DPs exhibited transcript sets that accumulated or declined (e.g., DP6) across several stresses. None of the patterns generated in this analysis indicated a specificity of a transcript abundance response associated with the 4°C treatment.

The 244 transcripts that constitute the desiccation-associated pattern DP2 include 16 of the 26 *LEA* gene transcripts, but only two of the 16 *ELIP* gene transcripts that accumulate in response to drying (Tables S8 and S18). A transcript encoding a LRR receptor-like serine/threonine-protein kinase, although not a member of the 15-gene tandem repeat, is also represented in DP2. Transcripts encoding two dehydration-responsive element binding proteins, four E3 ubiquitin-protein ligases, and three TFs are also members of this group. The collection of transcripts that constitute the rehydration-associated pattern DP10 is characterized by transcripts from three aldehyde oxidase genes, three cytochrome P450 protein transcripts, and three ethylene-responsive TFs (Table S19). This group also includes two of the LRR receptor-like serine/threonine-protein kinase transcripts, one of which is present in the large tandem repeat. The DP11 pattern, a collection of transcripts that exhibit a significant decrease in abundance in response to dehydration, is noteworthy for the inclusion of three TF transcripts (bHLH49-like, TFIIIA, and HEC2-like) (Table S20). Both DP5 and DP7 exhibited increased transcript abundance levels associated with elevated temperatures and included transcripts from several heat shock genes, primarily for low-molecular weight heat shock proteins (17 to 18.1 kDa range), and other chaperone proteins, as well as transcripts encoding protein components of the proteasome (Tables S21 and S22).

The transcripts associated with desiccation, DP2 and DP10, did not exhibit a recognizable genomic architecture but rather are dispersed across all 12 autosomes of the genome (Figure 6(a,b)). However, the 'island' analysis revealed four regions for DP2 where desiccation-responsive transcripts map within close proximity to each other (as indicated on the Circos plot for DP2 in Figure 6(a)). The four physical clusters, each containing three identical genes, are located on chromosomes 2, 9, 10, and 11 (Figure 6(a)). The chromosome 2 cluster was annotated as *SALP1*; the chromosome 9 cluster encodes a predicted rubredoxin; the chromosome 10 cluster encodes a predicted Golgi membrane protein; and the chromosome 11 cluster encodes a predicted phosphokinase of the threonine family (Table S23).

DISCUSSION

The 331.83-Mb genome of the haploid gametophyte of the DT moss *S. caninervis* is likely to be overestimated in size as the unplaced scaffolds (39.63 Mb) in the assembly likely contain non-moss DNA. This is based on higher read density; the lack of any linkage of these scaffolds with the 13 large scaffolds that constitute the rest of the genome; the lack of sequence similarity of the gene models mapped to the unplaced scaffolds with known plant genes; the low percentage of RNA-Seq reads that map to the unplaced scaffolds; and the below 1× coverage of the RNA-Seq reads that do map to the unplaced scaffolds. These attributes and the fact that the scaffolds were more closely linked to each other indicated that they were not associated with the *S. caninervis* genome and were likely contaminants derived from other organisms associated with the cultured plants, possibly symbionts as there were no obvious contaminants in the cultures. It is clear, however, that the 13 large assembled scaffolds represent the minimum size of the *S. caninervis* genome and are of a size and quality that suggests they represent 13 chromosomes. Although the genome size for *S. caninervis* has not been estimated using flow cytometry, 292.2 Mb is within the range of genome sizes for mosses (Voglmayr, 2000). We generated a high-quality annotation of the 16 545 protein-coding genes contained within the *S. caninervis* genome, which is almost half of the number, 35 307, in the genome of *P. patens* (Lang *et al.*, 2018), presumably because the *P. patens* genome has undergone a more recent WGD.

Of the *S. caninervis* genome, 56% is composed of repeat elements, nearly identical to the 57% repeat element estimate for the genome of *P. patens* (Lang *et al.*, 2018). Unlike most plant genomes where LTR-*Gypsy* retrotransposons are the most abundant type, unknown repeat elements were the most abundant (65.8%). The LTR-*Gypsy* and *Copia* repeat families comprise 7% and 3.4% of the *S. caninervis* genome, respectively, in comparison to 47.9% and 3.5% in the *P. patens* genome, indicating a significant expansion of *Gypsy* family elements in *P. patens* (Lang *et al.*, 2018). These differences in repeat structures between the two genomes might be related to the species' differences in habitat and stress tolerance, perhaps indicating repeats act as a source of regulatory elements (Hollister

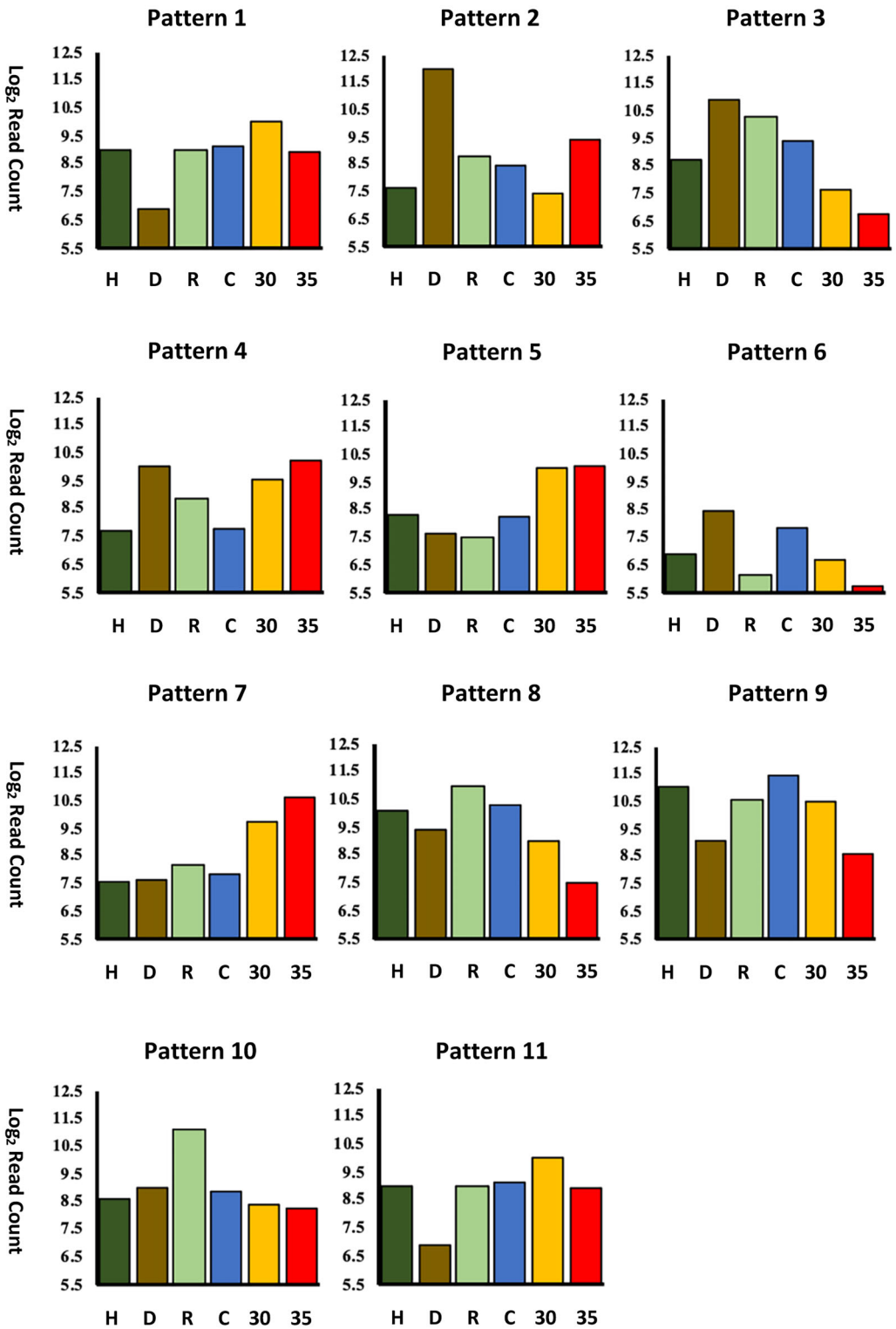

Figure 5. Overview of the dominant patterns (DPs) of transcript abundance in each stress treatment and the hydrated control. General tendency and average log_2 transformed abundance of transcripts at each stress. This analysis shows the maximum accumulation of transcripts at different stresses, indicating the stress specificity of the transcript abundance changes. H, hydrated control (dark green); D, dry (brown); R, rehydrated for 30 min (light green); C, 4°C for 90 min (blue); 30, 30°C for 90 min (yellow); 35, 35°C heat shock for 90 min (red).

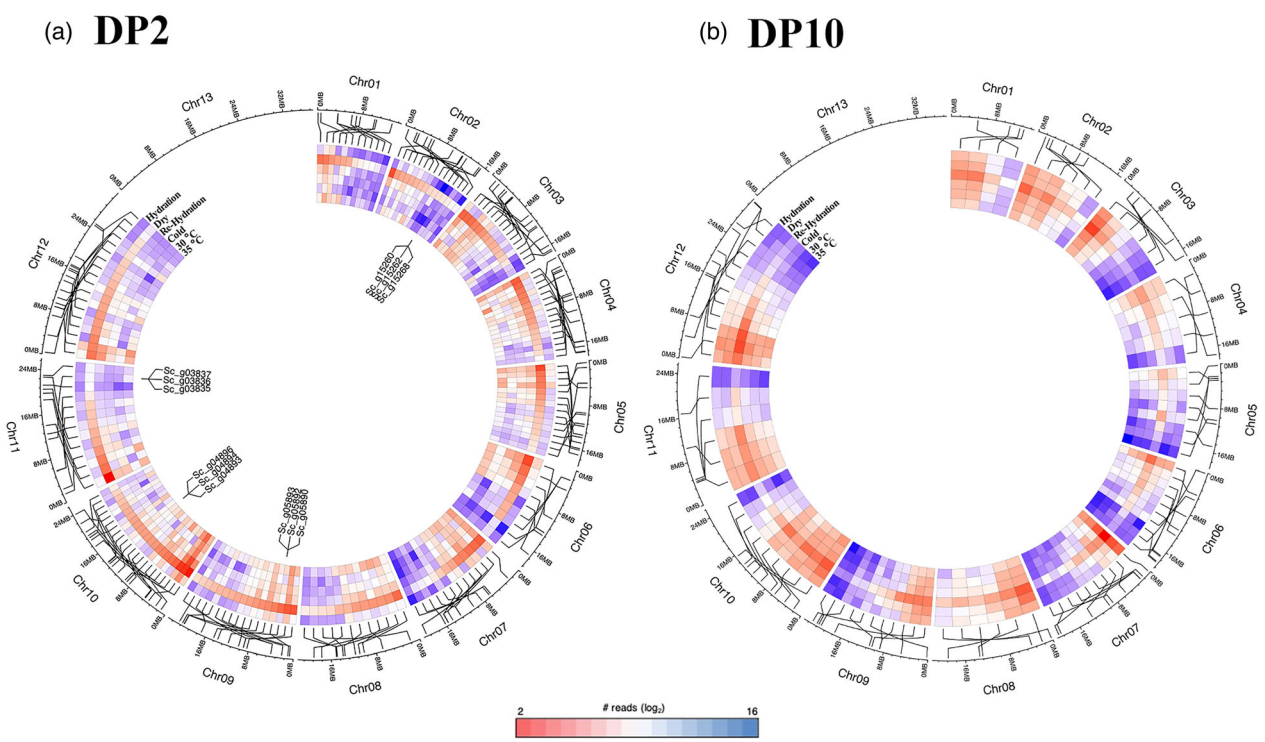

Figure 6. Genome mapping of genes represented by transcripts in dominant patterns DP2 (a) and DP10 (b). Circos plot depicting the chromosomal locations for the genes represented by transcripts that are grouped in dominant patterns DP2 (a) and DP10 (b). (a) Transcript abundance profiles for genes that cluster in the desiccation-associated pattern DP2 derived from the 3972 genes that are differentially expressed in at least one treatment. The location of the physical clusters (islands) identified in DP2 are depicted in the center of the Circos plot. (b) Transcript abundance profiles for genes that cluster in the rehydration-associated pattern DP10 derived from the 3972 stress-related differentially accumulating transcripts. For both (a) and (b) circus plots, from outer to inner layers: chromosome, connecting lines for each gene to its chromosomal position, transcript abundance for the hydrated control treatment at 20°C, transcript abundance after desiccation to equilibrium at 67% relative humidity, transcript abundance after rehydration from the desiccated state for 30 min at 20°C, transcript abundance after cold treatment for 90 min at 4°C, transcript abundance after elevated temperature for 90 min at 30°C, and transcript abundance after heat shock for 90 min at 35°C. Key: heat map of $\log_2$ transformed transcript abundance levels from blue (2) to red (16) in the indicated treatments.

et al., 2011), but such nuances were not a part of our analyses.

The size of the 13 largest scaffolds coupled with the use of HiC lead us to conclude that each scaffold represents a chromosome in the *S. caninervis* genome. A chromosome count of 13 for *S. caninervis* is consistent with chromosome counts for *Syntrichia* reported in the Chromosome Counts Database (CCDB http://ccdb.tau.ac.il/). We conclude that chromosome 13, the largest chromosome, is the female sex chromosome based on the presence of the 700-bp sex-linked locus previously reported for this moss (Baughman *et al.*, 2017) and its singular topology amongst the rest of the chromosomes (low gene number and high LTR content).

The *S. caninervis* genome has retained in large part the characteristics of the hypothesized seven ancestral moss chromosomes (Rensing *et al.*, 2012; Lang *et al.*, 2018), as evidenced by the genomic synteny analyses. The syntenic blocks mapped within the *S. caninervis* genome exhibit clear evidence of the WGD event that occurred early in the evolution of mosses, termed WGD1 (Lang *et al.*, 2018) or ψ duplication event (Gao *et al.*, 2020). Accepting the hypothesis that the ancestral moss had seven chromosomes (Rensing *et al.*, 2012), it appears that subsequent to the WGD, large segments of chromosomes 1 and 8 were transposed or lost from the lineage to generate the 12 autosomes. The origin of the sex chromosome, chromosome 13, may have occurred after the WGD event as there is no evidence for its derivation from either chromosome 1 or 8 (no syntenic blocks). However, there is one small area of synteny between chromosome 13 and chromosome 5 of *P. patens* which may indicate an earlier derivation. The loss of one ancestral chromosome for *P. patens* and the loss of two ancestral chromosomes for *S. caninervis* during the WGD1 (or ψ) event would explain the two syntenic depth ratios of 1:2 and 1:4. Although some chromosome rearrangements could be discerned, the *S. caninervis* genome in large part retains the characteristics of the ancestral moss genome.

Earlier studies suggest gene retention following WGD is biased for different gene functional categories (Van der Peer *et al.*, 2017). Unlike most genes that tend to return to single-copy status following WGD, transcription regulatory

genes are biased towards being retained and gaining novel functions (Li *et al.*, 2016). We predicted 542 TF genes in 40 gene families (Table S9), which contrasts with 1156 TF genes in the *P. patens* genome (1:2 ratio). In the *S. caninervis* genome, five genes represent an intact HSF DNA binding domain TF family, whereas in *P. patens* 16 TFs represent the HSF DNA binding domain family. This indicates that for these genes there is a syntenic relationship of 1:4 and as such one would expect 20 of the HSF TFs in *P. patens*. We can thus hypothesize that one member of the functional HSF TF family was lost in *P. patens*. In addition, we determined that 29 TF genes were located in tandem repeat clusters in the *S. caninervis* genome and 128 in the *P. patens* genome, indicating that tandem duplications also contribute to the expansion of TF gene families.

Our analysis revealed 1526 orphan genes unique to *S. caninervis*, at least within the limits of available genomes, which are likely to reflect species-specific genetic requirements for unique functions or control elements as proposed by Khalturin *et al.* (2009). Orphan genes can be derived in several ways, including gene duplication or horizontal gene transfer followed by sequence divergence (Van Oss and Carnuvis, 2019), and perhaps reflect unique and strong evolutionary pressures imposed upon a species as it adapts to new environments. The observation that many of the orphan genes of *S. caninervis* encode proteins that have documented roles in abiotic stress tolerance mechanisms lends some credence to the assertion that many orphan genes arise from environmentally driven selective factors, as suggested by Arendsee *et al.* (2014).

ELIP proteins are thought to act as photoprotectants by binding to chlorophylls and carotenoids (Adamska *et al.*, 1999), protecting them from the damage associated with high light and desiccation. The elevation of *ELIP* transcript abundance, along with an expansion of *ELIP* genes within the genome, is currently considered a genomic signature of DT plants (VanBuren *et al.*, 2019). The *S. caninervis* genome has 35 *ELIP* genes, well above the 10-gene threshold that delineates the lower limit of expansion in DT plants (VanBuren *et al.*, 2019). Of the 35 *ELIP* genes in *S. caninervis*, 29 are captured in the ELIP OG, OG0000141, which is characterized by *ELIP* gene expansion in all but one of the DT plant genomes used in the analysis. The one exception is for the DT grass *O. thomaeum*, which is likely due to its distant evolutionary relationship to the other species in the OG analysis. Notably, 16 of the 35 *S. caninervis ELIP* genes are characterized by an increase in transcript abundance during desiccation and two of the *ELIP* genes are also represented in DP2, which associated with the desiccation response of the moss. The majority of the *ELIP* genes (29) are situated within eight tandem repeats located on four chromosomes. However, only some members of each repeat were characterized by an increase in transcript abundance in response to desiccation and so it is unlikely that this structural aspect of *ELIP* genes within the genome is critical to the DT response.

The LEA proteins are diverse in both form and function, but all are hydrophilic proteins thought to protect various cellular components and are associated with DT in both seeds and vegetative tissues (Oliver *et al.*, 2020). Although *S. caninervis* has 51 *LEA* genes, this is neither an expansion (Arabidopsis has 81 *LEA* genes) nor an indication of a vegetative DT phenotype (Oliver *et al.*, 2020). However, of the 51 *LEA* genes in the *S. caninervis* genome, 26 are characterized by an increase in transcript abundance during desiccation and 16 of the 26 are represented by transcripts in DP2. Only 12 of the *LEA* genes are situated within tandem repeats and only some members of each repeat were characterized by an increase in transcript abundance in response to desiccation. So again, it is unlikely that this structural aspect of the *LEA* genes in the genome can be ascribed to have an influence on the DT response of the moss.

The magnitude of the *S. caninervis* transcriptomic response to an abiotic stress event is similar to that of most plants (Hruz *et al.*, 2008; Dash *et al.*, 2012), with perhaps the exception of the response to exposure to 4°C. Clearly 4°C is not a severe stress for this desert moss. As seen in many studies of transcriptomic responses to abiotic stress, there is crosstalk between the responses to individual stresses (Tahmasebi *et al.*, 2019). Almost 36% of the differentially accumulated transcripts were present in at least two treatments (3972 of 11 113 transcripts). Each of the stress treatments, with the exception of the 4°C treatment, resulted in complex changes in the transcriptome with anywhere from 4880 (slow drying) to 8118 (heat shock) transcripts significantly departing from abundance levels seen in the control hydrated plants at 20°C. Changes in transcript abundance, at least with regard to the response to desiccation, may be determined by changes in mRNA stability through sequestration in messenger ribonucleoprotein particles (mRNPs) as previously demonstrated for *Syntrichia ruralis* (Wood and Oliver, 1999). Mapping of the differentially accumulating transcripts for each treatment to the genome of *S. caninervis* did not indicate any structural signature that would indicate an evolutionary blueprint for the stress responses that were investigated.

However, the clustering of the transcript data to reveal DPs of alterations in transcript abundance during the stress treatments reduced the complexity of the transcriptome data and helped to frame them in the context of the genome. The DP5 and DP7 heat and heat shock patterns, where transcript abundance likely reflects transcriptional activity, highlight the increase in abundance of heat shock proteins, chaperones, and proteins involved in protein homeostasis, all of which are commonly associated with the response to elevated temperatures and heat shock (Richter *et al.*, 2010; Xu and Xue, 2019). Similarly, the desiccation pattern, DP2, highlights transcripts that are

commonly associated with plant desiccation responses, including *ELIPs*, *LEAs*, and ABA signal transduction pathways (Oliver *et al.*, 2020). Transcripts encoding proteins critical for protein homeostasis (E3 ubiquitin-protein ligases) are also highlighted in DP2, indicating the probable denaturation of proteins during desiccation; thus, it is likely that these transcripts are required during the initial stages of rehydration. The rehydration pattern DP10 features transcripts encoding aldehyde oxidase, which is known to detoxify toxic aldehydes that are produced during stress (Srivastava *et al.*, 2017), indicative of metabolic damage incurred during desiccation (38). The stressful aspect of rehydration from the dried state could also be reflected in the increased abundance of ethylene-responsive TF transcripts, which are often involved in abiotic and biotic stress responses as ethylene is generated from cellular damage (Singh *et al.*, 2002). The rehydration pattern also includes transcripts of P450 94A2 proteins, which are implicated in fatty acid metabolism (Pinot and Beisson, 2011) and are an important aspect of membrane repair during rehydration.

The functional aspects of the DPs of transcript abundance did not reveal co-localization of any of the co-regulated genes. However, using the method of Pecrix *et al.* (2018), we uncovered four interesting sets of genes that appear to be important in the desiccation response (in DP2) and that perhaps accumulate transcripts during dehydration in a location-dependent manner or reflect a strong evolutionary signal that is important to the DT phenotype in this moss. SALP1 is a plasma membrane protein that may be involved in osmoregulation (Yuan *et al.*, 2016) and indicates an important role for membrane transport in the response to desiccation. Rubredoxin is a small redox-responsive chloroplast protein involved in the assembly of Photosystem II (Kiss *et al.*, 2019) and may serve as an electron carrier in ROS scavenging mechanisms (Sewelam *et al.*, 2016). The Golgi apparatus membrane protein TVP15 is involved in membrane trafficking (Inadome *et al.*, 2007) that might impact membrane stability during drying and rehydration. Finally, one of the phosphokinases annotates as an LRR receptor-like serine/threonine-protein kinase belonging to a group of receptor-like kinases, which have been implicated in adaptation to environmental stresses (Liang and Zhou, 2018). The expression of these four genes increases during dehydration in a location-dependent manner and this may reflect a strong evolutionary signal that is important to the DT phenotype in this moss.

METHODS

Syntrichia caninervis cultures

Shoots of *S. caninervis* Mitt., free of visible algae and bacteria, were grown in the Stark lab at the University of Nevada, Las Vegas. The shoots originated from a single specimen collected in the Mojave Desert and vouchered in the UNLV herbarium (*Stark NV-107* [UNLV-B-0003271], USA, Nevada, Clark County, Newberry Mts, Christmas Tree Pass). This specimen is female and was selected for its healthy appearance and sustained growth under culture conditions. The shoots were grown on sterile fine sand collected locally near Las Vegas (pH neutral, sieved at 500 μm), and watered on alternating weeks with sterile distilled water and with a 30% inorganic nutrient solution (Hoagland and Arnon, 1950). Cultures were placed in a growth chamber set to a 12-h photoperiod (20°C light, 8°C dark), at approximately 90 μmol m^{-2} s^{-1} photosynthetically active radiation (PAR). The single clonal line used for genome sequencing had been sub-cultured and grown to maturity through at least three asexual generations.

Several gametophytes from the original clonal line were used to expand and establish subcultures for generating sufficient material for isolation of RNA both for the generation of a full-length transcriptome (described later) to aid in the annotation of the genome and for the construction of transcriptome libraries for transcript abundance studies (RNA-Seq). These cultures were grown under identical conditions as described for the original clone; however, the sterile sand used was collected from a dune near Moab, Utah and had a different composition from that collected in Nevada; 93.9% sand, 5.5% silt, 0.6% clay with a pH of 8.4 and the light was at 100 μmol m^{-2} s^{-1} PAR. Expansion of the cultures was achieved by isolating individual shoots following branching and by fragment regeneration.

Genomic DNA isolation and Chicago library preparation and sequencing

Genomic DNA isolation, library preparation, sequencing, and assembly were conducted by Dovetail Genomics (Scotts Valley, CA). To isolate genomic DNA, 1–2 grams of flash frozen gametophytes was ground to a fine powder in a mortar and pestle chilled with liquid nitrogen. The ground material was incubated in 20 ml of CTAB buffer (2% CTAB, 1.4 M NaCl, 100 mM Tris pH 9.5, 20 mM EDTA, 1% PVP-40, 1% PEG 8000, 0.5% BME) at 68°C for 15 min. The extraction was clarified by centrifugation at 10 000 × ***g***, and the supernatant was extracted with equal volumes of phenol/chloroform (1:1) followed by chloroform:iso-amyl alcohol 24:1 (v/v). High-molecular weight genomic DNA was precipitated with isopropanol and resuspended in Qiagen Buffer G2 with RNAse A, and after incubation at 50°C for 30 min, the DNA was purified using Qiagen Genomic-tips as described by the manufacturer (Qiagen Waltham, MA). The DNA was precipitated with isopropanol and resuspended in TE.

A Chicago genomic DNA library was prepared as described previously (Putnam *et al.*, 2016). Approximately 500 ng of HMW gDNA (mean fragment length = 100 kb) was isolated from hydrated *S. caninervis* gametophytes

utilizing the Qiagen HMW DNA Isolation Kit (Qiagen, Waltham, MA). The isolated DNA was reconstituted into chromatin overnight at 27°C using the Active Motif *in vitro* chromatin assembly kit (Active Motif, Carlsbad, CA) and fixed with formaldehyde. Fixed chromatin was digested with *Dpn*II, the 5′ overhangs were filled in with biotinylated nucleotides, and the free blunt ends were ligated. After ligation, crosslinks were reversed and the DNA was purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to approximately 350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters (New England Biolabs, Ipswich, MA). Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeq platform to produce 225 million 2 × 150-bp paired end reads, which provided 144.9× physical coverage of the genome (1–100 kb pairs).

HiC library preparation and sequencing

A Dovetail HiC library was prepared as described previously (Lieberman-Aiden *et al.*, 2009). For each library, chromatin was fixed in place in the nucleus by incubation of the gametophytes in 1% formaldehyde for 15 min under vacuum. The fixed chromatin was extracted from the treated gametophytes using the Dovetail™ Hi-C Kit (Dovetail Genomics, Santa Cruz, CA) and digested with *Dpn*II, the 5′ overhangs were filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed, and the DNA was purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to approximately 350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeq platform to produce 247 million 2 × 150-bp paired end reads, which provided 39 587.8× physical coverage of the genome (10–10 000 kb pairs).

***De novo* assembly of the *S. caninervis* genome**

A *de novo* assembly was constructed using a combination of paired end (mean insert size approximately 350 bp) libraries. *De novo* assembly was performed using Meraculous v2.2.2.5 (diploid mode 1) (Chapman *et al.*, 2011) with a kmer size of 109. The input data consisted of 430.8 million read pairs sequenced from paired end libraries (totaling 129.3 Gbp). Reads were trimmed for quality, sequencing adapters, and mate pair adapters using Trimmomatic (Bolger *et al.*, 2014).

Scaffolding the assembly with HiRise

The input *de novo* assembly, shotgun reads, Chicago library reads, and Dovetail HiC library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam *et al.*, 2016). An iterative analysis was conducted. First, Shotgun and Chicago library sequences were aligned to the draft input assembly using a modified SNAP read mapper (http://snap.cs.berkeley.edu). The separations of Chicago read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and to make joins above a threshold. After aligning and scaffolding Chicago data, Dovetail HiC library sequences were aligned and scaffolded following the same method. After scaffolding, shotgun sequences were used to close gaps between contigs.

Abiotic stress treatments

Dehydration (slow drying) was achieved by placing the gametophytes in small wire baskets over saturated ammonium nitrate (67% relative humidity) in a closed glass desiccator placed in the same incubator as the moss cultures at 20°C. Under these conditions the gametophytes reach a stable weight (equilibrium) at 6 h within the light period (100 μmol m^{-2} s^{-1}) of the day/night cycle. Rehydration was achieved by placing the desiccated gametophytes in a petri dish in the incubator at 20°C in the light and adding sufficient distilled water to ensure full hydration. Heat treatments were achieved by placing hydrated gametophytes in small glass petri dishes, with only sufficient water to maintain hydration, in a glass fronted incubator in ambient light at the appropriate temperature, 30°C or 35°C, for 90 min. Cold treatments, 4°C, were achieved by placing hydrated gametophytes in small glass petri dishes, with only sufficient water to maintain hydration, covered in wet ice in ambient light for 90 min. After completion of all treatments, including hydrated controls, the samples were blotted to quickly remove surface water (when appropriate) and flash frozen in liquid nitrogen for storage at −80°C prior to processing.

RNA extraction and RNA sequencing

RNA was extracted using the RNeasy (Qiagen, Hilden, Germany) kit with RLC buffer following the manufacturer's recommended protocol. The RNA isolates were treated with DNase1 and cleaned using the DNA-free RNA Kit (Zymo Technologies, Irvine, CA). RNA quality was assessed by use of a fragment analyzer (Advanced Analytical Technologies, Ankeny, IA) and the concentration was determined by use of a Nanodrop Spectrophotometer (Thermo Fisher, Waltham, Massachusetts). RNA libraries were created and

individually bar-coded from 2.7 μg of template total RNA utilizing the TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA) as described in the manufacturer's recommended protocol. Libraries were pooled in groups of 12 and sequenced (12 samples per lane) on an Illumina HiSeq 2500 ultra high-throughput DNA sequencing platform (Illumina, San Diego, CA) at the DNA Core facility at the University of Missouri, Columbia, MO, USA (https://dnacore.missouri.edu/HiSeq.html).

Full-length sequencing (IsoSeq) and *de novo* transcriptome assembly

To identify as many transcripts as possible in *S. caninervis*, high-quality RNA was extracted from *S. caninervis* submitted to seven different conditions: hydrated, 4°C for 90 min, 30°C for 90 min, 35°C for 90 min, slow dried at 67% relative humidity, rapid dried over activated silica gel (equilibrated after 30 min), and slow-dried rehydrated for 30 min. The RNAs were pooled for subsequent amplification. Barcoded SMRT libraries were prepared and sequenced on the PacBio platform with X SMRT cells by Novogene Corporation Inc. (Sacramento CA), yielding 328 566 consensus reads. Sequence reads were processed using IsoSeq3 (https://github.com/PacificBiosciences/IsoSeq3). Each transcript length was selected ranging from 300 to 15 000 bp. A large portion (255 292; 78%) of reads were classified as full-length based on the presence of bar-coded primers and polyA tails. IsoSeq3 cluster process yielded 17 431 high-quality isoforms. These high-quality isoforms were used to improve the *de novo S. caninervis* transcriptomics on Trinity (Grabherr *et al.*, 2011) and the genome annotation training. The sets from the high-quality single end Illumina RNA-Seq library described above were aligned to the unmasked *S. caninervis* genome using HISAT2 (Kim *et al.*, 2015) and the bam file was used as input for Trinity using the --genome_guide_bam flag. Default parameters were used for Trinity for a genome guide run adding the IsoSeq full-length transcripts to the --long_reads flag to produce a high-quality *de novo* transcriptome assembly.

Repeat annotation

Repeat families found in the *S. caninervis* genome assembly were first *de novo* identified and classified using the RepeatModeler (v1.0.11) (http://www.repeatmasker.org/RepeatModeler) and LTR_retriever (Ou and Jang, 2018) (which was employed to identify intact LTRs) pipelines. We removed *de novo* repeats from the LTR library that matched known *P. patens* genes (BLASTn E-value < 1e−5). The unclassified sequences were further curated based on BLASTn comparison with the previously characterized plant protein database (alluniRefprexp082813). We combined our *de novo* library with the RepBase plant repeat database (Bao *et al.*, 2015) and used it as input into RepeatMasker (v4.0.7) to discover and identify repeats.

Genome annotation

The *S. caninervis* genome was annotated using MAKER (2.31.8) (Cantarel *et al.*, 2008) and the *ab initio* annotation was performed using SNAP (Zaharia *et al.*, 2011), Augustus (Stanke and Moregenstern, 2005), and BRAKER (Hoff *et al.*, 2019) with two and one round of reiterative training, respectively. In MAKER, the *de novo* transcriptome assembly was treated as expressed sequence tag evidence and protein sequences from *Arabidopsis thaliana*, *Oryza sativa*, *P. patens*, and UniprotKB plant databases were used as protein homology evidence. The custom LTR library, as described above, was input as custom repeat database along with the default protein repeat libraries in MAKER. The initial iteration of the MAKER gene set was filtered for gene models with an AED score equal to 0.25 and retained only the gene models with more than 50 amino acids. Genes were functionally classified using Blast2Go software (Conesa and Götz, 2008) to generate BLASTx high-quality sequence matches and InterPro IDs.

Annotation of non-coding RNAs

Non-coding RNAs were predicted using various programs and data sources. Transfer RNA genes were annotated using the scan-SE (v2.0) algorithm (Lowe and Chan, 2016) with default parameters (-E) to search for eukaryotic tRNAs. For the detection of rDNA fragments in the genome assembly, we used the Basic Rapid Ribosomal RNA Predictor pipeline (Barrnap v0.9, https://github.com/tseemann/barrnap [Quast *et al.*, 2013]), a pipeline that uses nhmmer from the hmmer package (v3.1b2) and the seed hmm profile constructed using corresponding seed rRNA gene sequences in the Rfam database (Release 14) with an E-value cutoff at 1e−6. The resultant rDNA sequences were then further extended by aligning (BLASTn E-value < 1e−5) against the genome. The snRNA, miRNA, and SRP genes were predicted using cmscan from the INFERNAL (http://infernal.janelia.org/) package (v1.1.2) to search against the Rfam database (Release 14) with the family-specific gather threshold (--cut_ga) and further filtered using the E-value cutoff at 1e−5.

Identification of tandem duplications and comparisons of gene families

To identify local tandemly arrayed genes (TAGs), putative paralogous genes were identified by all-by-all BLASTp with an E-value cutoff value of 1e−5. Tandem duplicated gene clusters were then identified by allowing no more than 10 spacer genes. To elucidate the overall distribution of gene families among different plants (from algae to angiosperms), we recruited 10 other sequenced plant genomes for protein family clustering using OrthoFinder2 (Emms and Kelly, 2019) with default parameters. We analyzed 15 genomes: (i) five genomes that exhibit vegetative DT under most environmental conditions (*S. caninervis*,

S. lepidophylla, *L. brevidens*, *O. thomaeum*, and *Xerophyta viscosa*), (ii) three genomes that can acquire vegetative DT under certain environmental conditions (*Marchantia polymorpha*, *P. patens*, and *S. fallax*), and (iii) seven genomes from plants that have desiccation-sensitive vegetative tissues but do have desiccation-tolerant propagules (seeds or spores) (*A. thaliana*, *O. sativa*, *Chenopodium quinoa*, *Azolla filiculoides*, *Salvinia cucullata*, *Selaginella moellendorffii*, and *Amborella trichopoda*). The interspecies comparisons of specific sets of genes were conducted based on the orthogroup clustering results.

Syntenic comparisons between genomes of *S. caninervis* and *P. patens*

Syntelog anchors in the *S. caninervis* genome were identified using MCScanX (Wang *et al.*, 2012). Protein sequences of syntelog pairs were aligned using muscle v3.8.31 (Edgar, 2004) and back-translated into coding sequence alignments using the pal2nal v14 script (Suyama *et al.*, 2006). *Ks* values were estimated for pairs of syntelogs using KaKs_Calculator v2.0 (Wang *et al.*, 2010) with the Goldman and Yang model (-m GY) (Goldman and Yang, 1994). Similarly, *Ks* values of one-to-one orthologous syntenic genes filtered by reciprocal best BLAST hits between the *S. caninervis* and *P. patens* genomes were also calculated. Then *Ks* frequency plots within the range of [0.05,3] were generated and *Ks* peaks were analyzed using the kernel density estimation (KDE) function implemented in R software. Additionally, frequency distributions for the paralogous and orthologous 4DTv values were also calculated and analyzed.

Transcriptome analysis

Trim Galore (V0.4.1) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to trim adapter sequences and low-quality bases from the single end RNA-Seq raw data. Trimmed sequences were mapped to the *S. caninervis* genome using HISAT2 (Pertea *et al.*, 2016). Transcript abundance levels (in read counts) were quantified using HTseq (Anders *et al.*, 2015) with 26 723 gene models from *S. caninervis* provided as input. An average of read count for each time-point was calculated from the mean read counts of the three biological replicates. The statistical analysis was performed using R (R Core Team, 2017, https://www.r-project.org), with Bioconductor packages including edgeR (Robinson *et al.*, 2010). Normalization and differential analysis were carried out according to the edgeR model and package.

DPs of transcript accumulation were identified as described by Silva *et al.* (Silva *et al.*, 2016). Briefly, differentially expressed transcripts that occur in at least two different treatments were selected if the read counts in a specific treatment were significantly ($P < 0.01$, $\log_2$ fold-change of 2) larger than the read counts for other treatments. The extracted data sets of 3972 transcripts were used to determine the DPs. DPs were generated by using the function FANNY in the CRAN package cluster in R, and a minimum Pearson correlation of 0.98 was used to evaluate the number of clusters (K) choices from 1 to 20 with a cutoff value for cluster membership of 0.4. The K choice that yielded the greatest number of transcriptional modules was 11. An R script was used to locate clusters or 'islands' of co-located and co-regulated genes (47).

ACKNOWLEDGMENTS

The authors acknowledge the excellent technical skills of Kate Guill in the preparation of libraries for the RNA-Seq and Shaune Hall and Mark Daly of Dovetail Genomics for their guidance and assistance in the sequencing and the interpretation of the genome assembly. This research used resources provided by the SCINet project of the USDA Agricultural Research Service, ARS project number 0500-00093-001-00-D. The research was supported by the National Science Foundation Dimensions of Biodiversity Program Awards 1638996 (to KMF), 1638956 (to BDM), 1638966 (to MAB), 1638955 (to KKC), 1638943 (to LRS), and 1638972 (to MJO) and by the NSFC-Xinjiang Key Project (grant number U1703233 to DZ).

AUTHOR CONTRIBUTIONS

All authors contributed to the research planning and experimental designs; all authors participated in editing the manuscript; MJO and AS directed the project; AS and MJO wrote the manuscript with assistance from BG; AS performed the genome annotation, orthogroup analyses, and RNA-Seq experiments with assistance from MJO; BG performed the comparative genomics with assistance from XL and DZ; KF performed the sex chromosome analysis. All authors have reviewed the final version of the manuscript and approved it and therefore are equally responsible for the integrity and accuracy of its content.

CONFLICT OF INTEREST

All authors have read and approved the manuscript. All authors declare there are no competing interests regarding the content of this manuscript.

DATA AVAILABILITY STATEMENT

The plant material is vouchered and available on request from Dr. Lloyd Stark at the University of Nevada, Las Vegas. The genome and transcriptome data are available from the National Center for Biotechnology Information (BioProject PRJNA630883 and BioSample SAMN14847761). Accession JADDRJ000000000. The version described in this paper is version JADDRJ010000000.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Link density histogram of the genome assembly for *Syntrichia caninervis*. The *x* and *y*-axes give the mapping positions of the first and second reads in the read pair, respectively, grouped into bins. The color of each square gives the number of

read pairs within that bin. White vertical and black horizontal lines have been added to show the borders between scaffolds. Scaffolds less than 1 Mb are excluded.

Figure S2. Taxonomic distribution of top BLASTn hits for gene models located on the unplaced scaffolds. The pie chart represents the percent taxonomic distribution derived from the gene ID information for the top hit for each of the 6324 positive hits from the full 8486 gene models associated with the unplaced scaffolds.

Figure S3. Maximum likelihood tree showing the relationships between male and female *Syntrichia caninervis* sequences from an approximately 700-bp gene region located on scaffold 13. Unigene 80515 from China (Gao *et al*., 2020) and SC_g00229_V2 are inferred to be female based on the tree topology; the remaining sequences were derived from ramets expressing male (n = 3) or female (n = 2) gametangia.

Figure S4. Comparison of transcripts that significantly change in abundance across treatments. (a) The comparisons of transcripts across treatments that increase in abundance. (b) The comparisons of transcripts across treatments that increase in abundance. Dark brown squares highlight the number of differentially accumulating transcripts for treatment. Brown squares depict the number of differentially accumulating transcripts in the range of 2000 to 3000; tan squares 500 to 2000; light brown squares 50 to 100; and beige squares between 0 and 50.

Table S1. Chromosome size and GC content for the *Syntrichia caninervis* genome.

Table S2. Plant-related gene list for unplaced scaffolds.

Table S3. Mapping statistics and scaffold coverage for RNA-Seq sequenced reads mapped to all scaffolds.

Table S4. DNA repeat families for the *Syntrichia caninervis* genome.

Table S5. Annotated gene list for the *Syntrichia caninervis* genome.

Table S6. Non-protein-coding gene list for the *Syntrichia caninervis* genome.

Table S7. Tandem repeat gene lists with descriptions.

Table S8. Early light-inducible protein (*ELIP*) genes of *Syntrichia caninervis*: annotation and transcript abundance profiles.

Table S9. Late embryogenesis abundant (*LEA*) genes of *Syntrichia caninervis*: annotation and transcript abundance profiles.

Table S10. Syntenic blocks associated with each chromosome of *Syntrichia caninervis* (intraspecies synteny) and between the chromosomes of *S. caninervis* and *Physcomitrella patens* (Interspecies synteny).

Table S11. Comparison of transcription factor genes and gene families of *Syntrichia caninervis* and *Physcomitrella patens* genomes.

Table S12. Orthogroup analysis for all 15 plant species including *Syntrichia caninervis*.

Table S13. Annotated list of 'orphan' genes for *Syntrichia caninervis*.

Table S14. Ortholog analysis for *ELIP* and *LEA* genes across 15 plant species.

Table S15. Differential transcript abundance profiles for slow-dried, rehydrated, cold-treated, elevated temperature-treated, and heat-shocked gametophytes of *Syntrichia caninervis*.

Table S16. Transcript abundance responses for genes included in the top six tandem gene repeats within the genome of *Syntrichia caninervis*.

Table S17. Log_2 transformed transcript abundance for transcripts that changed in abundance by a log_2 fold-change of 2 from the hydrated control level in at least one treatment: transcript subset for dominant pattern analysis.

Table S18. Transcripts included in Dominant Pattern 2.

Table S19. Transcripts included in Dominant Pattern 5.

Table S20. Transcripts included in Dominant Pattern 7.

Table S21. Transcripts included in Dominant Pattern 10.

Table S22. Transcripts included in Dominant Pattern 11.

Table S23. Physical clusters (islands) of co-located and co-regulated genes from the Dominant Pattern 2 transcripts.

REFERENCES

Adamska, I., Roobol-Bóza, M., Lindahl, M. and Andersson, B. (1999) Isolation of pigment-binding early light-inducible proteins from pea. *Eur. J. Biochem.* **260**, 453–460.

Alpert, P. and Oliver, M.J. (2002) Drying without drying. In *Desiccation and Survival in Plants: Drying without Dying* (Black, M. and Pritchard, H.W., eds.) Oxford: Oxford University Press, pp. 3–44.

Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

Arendsee, Z.W., Li, L. and Wurtele, E.S. (2014) Coming of age: orphan genes in plants. *Trends in Plant Sci.* **19**, 698–708.

Bao, W., Kojima, K.K. and Kohany, O. (2015) Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA*, **6**, 11. https://doi.org/10.1186/s13100-015-0041-9

Baughman, J.T., Payton, A.C., Paasch, A.E., Fisher, K.M. and McDaniel, S.F. (2017) Multiple factors influence population sex ratios in the Mojave Desert moss *Syntrichia caninervis*. *Am. J. Bot.* **104**, 733–742.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Cantarel, B.L.B.L., Korf, I., Robb, S.M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A.S. and Yandell, M. (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Gen. Res.* **18**, 188–196.

Chapman, J.A., Ho, I., Sunkara, S., Luo, S., Schroth, G.P. and Rokhsar, D.S. (2011) Meraculous: de novo genome assembly with short paired-end reads. *PLoS One*, **6**, e23501. https://doi.org/10.1371/journal.pone.0023501

Conesa, A. and Götz, S. (2008) Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Gen.* **2008**, 1–13.

Costa, M.C.D., Artur, M.A., Maia, J. *et al.* (2017) A footprint of desiccation tolerance in the genome of *Xerophyta viscosa*. *Nat. Plants*, **3**, 1–10.

Costa, M.C.D., Cooper, K., Hilhorst, H.M.W. and Farrant, J.M. (2017) Orthodox seeds and resurrection plants: two of a kind? *Plant Physiol.* **175**, 589–599.

Costa, M.C.D., Farrant, J.M., Oliver, M.J., Ligterink, W., Buitink, J. and Hilhorst, H.W.M. (2016) Key genes involved in desiccation tolerance and dormancy across life forms. *Plant Sci.* **251**, 162–168.

Dash, S., Van Hemert, J., Hong, L., Wise, R.P. and Dickerson, J.A. (2012) PLEXdb: gene expression resources for plants and plant pathogens. *Nuc. Acids Res.* **40**, D1194–D1201.

Dhankher, O.P. and Foyer, C.H. (2018) Climate resilient crops for improving global food security and safety. *Plant Cell Environ.* **4**, 877–884.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nuc. Acids Res.* **32**, 1792–1797.

Eilbeck, K., Moore, B., Holt, C. and Yandell, M. (2009) Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinfor.* **10**, 67. https://doi.org/10.1186/1471-2105

Emms, D.M. and Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Gen. Biol.* **20**, 1–14.

Gao, B., Chen, M-X., Li, X-S., Liang, Y-Q., Zhang, D-Y., Wood, A.J., Oliver, M.J. and Zhang, J-H. (2020) Ancestral gene duplications in mosses characterized by integrated phylogenomic analyses. *Journal of Systematics and Evolution*. https://doi.org/10.1111/jse.12683

Goldman, N. and Yang, Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736.

Gosling, S.N. and Arnell, N.W. (2016) A global assessment of the impact of climate change on water scarcity. *Clim. Change*, **134**, 371–385.

Grabherr, M.G., Haas, B.J., Yassour, M. *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotech.* **29**, 644–652.

Hoagland, D.R. and Arnon, D.I. (1950) The water culture method for growing plants without soil. 2nd edition. *Circ. Univ. Calif. Agr. Stn.* **347**, 32.

Hoff, K.J., Lomsadze, A., Borodovsky, M. and Stanke, M. (2019) Whole-genome annotation with BRAKER. In *Gene Prediction. Methods in Molecular Biology* (Kollmar, M., ed.) New York, NY: Humana, vol 1962, pp. 65–95.

Hollister, D., Smith, L.M., Guo, Y.L., Ott, F., Weigel, D. and Gaut, B.S. (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc. Nat. Acad. Sci. USA*, **108**, 2322–2327.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W. and Zimmermann, P. (2008) Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinform.* **2008**, 420747. https://doi.org/10.1155/2008/420747

Inadome, H., Noda, Y., Kamimura, Y., Adachi, H. and Yoda, K. (2007) Tvp38, Tvp23, Tvp18 and Tvp15: novel membrane proteins in the Tlg2-containing Golgi/endosome compartments of *Saccharomyces cerevisiae*. *Exp. Cell Res.* **313**, 688–697.

Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J. and Gao, G. (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nuc. Acids Res.* **45**, D1040–D1045.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. and Bosch, T.C. (2009) More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* **25**, 404–413.

Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.

Kiss, É., Knoppová, J., Aznar, G.P., Pilný, J., Yu, J., Halada, P., Nixon, P.J., Sobotka, R. and Komenda, J. (2019) A photosynthesis-specific rubredoxin-like protein is required for efficient association of the D1 and D2 proteins during the initial steps of photosystem II assembly. *Plant Cell.* **31**, 2241–2258.

Koster, K.L., Balsamo, R.A., Espinoza, C. and Oliver, M.J. (2010) Desiccation sensitivity and tolerance in the moss *Physcomitrella patens*: assessing limits and damage. *Plant Growth Reg.* **62**, 293–302.

Lang, D., Ullrich, K.K., Murat, F. *et al.* (2018) The *Physcomitrella* patens chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515–553.

Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y. and De Smet, R. (2016) Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell*, **28**, 326–344.

Liang, X. and Zhou, J.M. (2018) Receptor-like cytoplasmic kinases: central players in plant receptor kinase–mediated signaling. *Ann. Rev. Plant Biol.* **69**, 267–299.

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289.

Lowe, T.M. and Chan, P.P. (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nuc. Acids Res.* **44**, W54–W57.

Oliver, M.J., Farrant, J.M., Hilhorst, H.W.M., Mundree, S., Williams, B. and Bewley, J.D. (2020) Desiccation tolerance: avoiding cellular damage during drying and rehydration. *Ann. Rev. Plant Biol.* **71**, 435–460.

Oliver, M.J., Velten, J. and Mishler, B.D. (2005) Desiccation tolerance in bryophytes: a reflection of the primitive strategy for plant survival in dehydrating habitats? *Integ. Comp. Biol.* **45**, 788–799.

Ou, S. and Jiang, N. (2018) LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422.

Pardo, J., Wai, C.M., Chay, H., Madden, C.F., Hilhorst, H.W.M., Farrant, J.M. and VanBuren, R. (2020) Intertwined signatures of desiccation and drought tolerance in grasses. *Proc. Nat. Acad. Sci. USA*, **117**, 10079–10088.

Pecrix, Y., Staton, S.E., Sallet, E. *et al.* (2018) Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nat. Plants*, **4**, 1017–1025.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T. and Salzberg, S.L. (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT. StringTie and Ballgown. *Nature Protoc.* **11**, 1650–1667.

Pinot, F. and Beisson, F. (2011) Cytochrome P450 metabolizing fatty acids in plants: characterization and physiological roles. *FEBS J.* **278**, 195–205.

Proctor, M.C.F. and Pence, V.C. (2002) Vegetative tissues: bryophytes, vascular resurrection plants and vegetative propagules. In *Desiccation and Survival in Plants: Drying without Dying*. (Black, M. and Pritchard, H.W., eds). Wallingford: CABI, pp. 207–237.

Putnam, N.H., O'Connell, B.L., Stites, J.C. *et al.* (2016) Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Gen. Res.* **26**, 342–350.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res.* **41**, D590–D596.

Rensing, S.A., Beike, A.K. and Lang, D. (2012) Evolutionary importance of generative polyploidy for genome evolution of haploid-dominant land plants. In *Plant Genome Diversity*. (Greilhuber, J., Wendel, J.F., Leitch, I.J. and Dolezel, J., eds). Vienna, New York: Springer, pp. 295–305.

Richter, K., Haslbeck, M. and Buchner, J. (2010) The heat shock response: life on the verge of death. *Mol. Cell.* **40**, 253–266.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Sewelam, N., Kazan, K. and Schenk, P.M. (2016) Global plant stress signaling: reactive oxygen species at the crossroad. *Front. Plant Sci.* **7**, 187. https://doi.org/10.3389/fpls.2016.00187

Silva, A.T., Ribone, P.A., Chan, R.L., Ligterink, W. and Hilhorst, H.W.M. (2016) A predictive coexpression network identifies novel genes controlling the seed-to-seedling phase transition in *Arabidopsis thaliana*. *Plant Physiol.* **170**, 2218–2231.

Singh, K.B., Foley, R.C. and Oñate-Sánchez, L. (2002) Transcription factors in plant defense and stress responses. *Curr. Opin. Plant Biol.* **5**, 430–436.

Srivastava, S., Brychkova, G., Yarmolinsky, D., Soltabayeva, A., Samani, T. and Sagi, M. (2017) Aldehyde oxidase 4 plays a critical role in delaying silique senescence by catalyzing aldehyde detoxification. *Plant Physiol.* **173**, 1977–1997.

Stanke, M. and Morgenstern, B. (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nuc. Acids Res.* **33**, W465–W467.

Stark, L.R. (2017) Ecology of desiccation tolerance in bryophytes: a conceptual framework and methodology. *Bryol.* **120**, 130–166.

Stark, L.R. and Brinda, J.C. (2015) Developing sporophytes transition from an inducible to a constitutive ecological strategy of desiccation tolerance in the moss *Aloina ambigua*: effects of desiccation on fitness. *Ann. Bot.* **115**, 593–603.

Sultan, S.E. (2000) Phenotypic plasticity for plant development, function and life history. *Trends Plant Sci.* **5**, 537–542.

Suyama, M., Torrents, D. and Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nuc. Acids Res.* **34**, W609–W612.

Tahmasebi, A., Ashrafi-Dehkordi, E., Shahriari, A.G., Mazloomi, S.M. and Ebrahimie, E. (2019) Integrative meta-analysis of transcriptomic responses to abiotic stress in cotton. *Prog. Biophys. Mol. Biol.* **146**, 112–122.

VanBuren, R., Pardo, J., Wai, C.M., Evans, S. and Bartels, D. (2019) Massive tandem proliferation of ELIPs supports convergent evolution of desiccation tolerance across land plants. *Plant Physiol.* **179**, 1040–1049.

VanBuren, R., Wai, C.M., Keilwagen, J. and Pardo, J. (2018a) A chromosome-scale assembly of the model desiccation tolerant grass *Oropetium thomaeum*. *Plant Direct.* **2**, 11. https://doi.org/10.1002/pld3.96/e00096

VanBuren, R., Wai, C.M., Pardo, J., Giarola, V., Ambrosini, S., Song, X. and Bartels, D. (2018b) Desiccation tolerance evolved through gene duplication and network rewiring in *Lindernia*. *Plant Cell.* **30**, 2943–2958.

VanBuren, R., Wai, C.M., Ou, S., Pardo, J., Bryant, D., Jiang, N., Mockler, T.C., Edger, P. and Michael, T.P. (2018c) Extreme haplotype variation in the desiccation-tolerant clubmoss *Selaginella lepidophylla*. *Nat. Commun.* **9**, 1–8.

Van de Peer, Y., Mizrachi, E. and Marchal, K. (2017) The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424.

Van Oss, S.B. and Carvunis, A.R. (2019) De novo gene birth. *PLoS Gen.* **15**, 5. https://doi.org/10.1371/journal.pgen.1008160

Voglmayr, H. (2000) Nuclear DNA amounts in mosses (Musci). *Ann Bot.* **85**, 531–546.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J. and Yu, J. (2010) KaKs Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Gen. Prot. Bioinf.* **8**, 77–80.

Wang, Y., Tang, H., Debarry, J.D. *et al.* (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nuc. Acids Res.* **40**, e49. https://doi.org/10.1093/nar/gkr1293

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V. and Zdobnov, E.M. (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548.

Wood, A.J. (2007) The nature and distribution of vegetative desiccation tolerance in hornworts, liverworts, and mosses. *Bryol.*, **110**, 163–167.

Wood, A.J. and Oliver, M.J. (1999) Translational control in plant stress: formation of messenger ribonucleoprotein complexes (mRNPs) in *Tortula ruralis* in response to desiccation. *Plant J.* **18**, 359–370.

Xiao, L., Yang, G., Zhang, L. *et al.* (2015) The resurrection genome of *Boea hygrometrica*: a blueprint for survival of dehydration. *Proc. Nat. Acad. Sci. USA*, **112**, 5833–5837.

Xiao, L., Yobi, A., Koster, K.L., He, Y. and Oliver, M.J. (2018) Desiccation tolerance in *Physcomitrella patens*: rate of dehydration and the involvement of endogenous ABA. *Plant Cell Environ.* **41**, 275–284.

Xu, F.Q. and Xue, H.W. (2019) The ubiquitin-proteasome system in plant responses to environments. *Plant Cell Environ.* **42**, 2931–2944.

Xu, Z., Xin, T., Bartels, D. *et al.* (2018) Genome analysis of the ancient tracheophyte *Selaginella tamariscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. *Mol. Plant.* **11**, 983–994.

Yuan, X., Sun, H., Tang, Z., Tang, H., Zhang, H. and Huang, J. (2016) A novel little membrane protein confers salt tolerance in rice (*Oryza sativa* L.). *Plant Mol. Biol. Rep.* **34**, 524–532.

Zaharia, M., Bolosky, W.J., Curtis, K., Fox, A., Patterson, D., Shenker, S., Stoica, I., Karp, R.M. and Sittler, T. (2011) Faster and more accurate sequence alignment with SNAP. arXiv:1111.5572.