LARGE-SCALE BIOLOGY ARTICLE

Desiccation Tolerance Evolved through Gene Duplication and Network Rewiring in *Lindernia*

Robert VanBuren[1,2*], Ching Man Wai[1,2], Jeremy Pardo[2,3], Valentino Giarola[4], Stefano Ambrosini[4], Xiaomin Song[4], Dorothea Bartels[4*]

[1]Department of Horticulture, Michigan State University, East Lansing, Michigan 48824, USA
[2]Plant Resilience Institute, Michigan State University, East Lansing, MI 48824, USA
[3]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA
[4]IMBIO, University of Bonn, Kirschallee 1, D-53115 Bonn, Germany
*Corresponding authors: bobvanburen@gmail.com, unb137@uni-bonn.de

Short title: Genomics of desiccation tolerance in *Lindernia*

One-sentence summary: Desiccation tolerance in *Lindernia* arose via a combination of gene duplication and network level rewiring of existing seed pathways.

The author(s) responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is (are): Robert VanBuren (bobvanburen@gmail.com)

ABSTRACT

Though several resurrection plant genomes have been sequenced, the lack of suitable dehydration-sensitive outgroups has limited genomic insights into the origin of desiccation tolerance. Here, we utilized a comparative system of closely related desiccation-tolerant (*Lindernia brevidens*) and -sensitive (*L. subracemosa*) species to identify gene and pathway level changes associated with the evolution of desiccation tolerance. The two high-quality *Lindernia* genomes we assembled are largely collinear and over 90% of genes are conserved. *L. brevidens* and *L. subracemosa* have evidence of an ancient, shared whole-genome duplication event, and retained genes have neofunctionalized, with desiccation-specific expression in *L. brevidens*. Tandem gene duplicates are also enriched in desiccation-associated functions including a dramatic expansion of early light induced proteins (ELIPs) from 4 to 26 copies in *L. brevidens*. A comparative differential gene co-expression analysis between *L. brevidens* and *L. subracemosa* supports extensive network rewiring across early dehydration, desiccation, and rehydration timecourses. Many *LEA* genes show significantly higher expression in *L. brevidens* compared to their orthologs in *L. subracemosa*. Co-expression modules uniquely upregulated during desiccation in *L. brevidens* are enriched with seed-specific and ABA-associated *cis* regulatory elements. These modules contain a wide array of seed-associated genes that have no expression in the desiccation-sensitive *L. subracemosa*. Together these findings suggest that desiccation tolerance evolved through a combination of gene duplications and network level rewiring of existing seed desiccation pathways.

INTRODUCTION

Comparative systems are a powerful tool for dissecting the molecular basis of complex biological traits. The origins of desiccation tolerance in resurrection plants are largely unknown, but the underlying genetic signatures could be traced using pairs of closely related desiccation-sensitive and -tolerant species. Such an approach has been applied to *Eragrostis* (Vander Willigen et al., 2001), *Selaginella* (Yobi et al., 2013), and *Sporobolus* (Oliver et al., 2011) at the morphological and biochemical level to identify signatures that distinguish drought and desiccation responses. Detailed pairwise comparisons have identified changes in cell wall composition (Plancot et al., 2014), metabolite and osmoprotectant accumulation (Oliver et al., 2011; Yobi et al., 2013), and physical properties unique to desiccation-tolerant species. Though genomes are available for several resurrection plants (VanBuren et al., 2015; Xiao et al., 2015; Costa et al., 2017; VanBuren et al., 2018), genomic resources in these comparative lineages are limited, and no genomes of closely related desiccation-sensitive species have been sequenced. High-quality reference genomes are available for the desiccation-sensitive *Selaginella moellendorffii* (Banks et al., 2011) and -tolerant *S. lepidophylla* (VanBuren et al., 2018), but their estimated divergence 248 MYA prevents detailed genomic comparisons (Baniaga et al., 2016).

Resurrection plants endure extreme and prolonged drought events through vegetative desiccation, entering a preserved and protected quiescent state that functionally mirrors seed dormancy in angiosperms. Desiccation tolerance was a critical adaptation during early land plant evolution, and many early-diverging fern, moss, and lycophyte lineages have retained or convergently evolved these ancestral resilience mechanisms (Proctor, 1990; Oliver et al., 2000; Lüttge et al., 2011). Vegetative desiccation tolerance is comparatively less common in angiosperms, and recent genomic and metabolic studies suggest it evolved through rewiring seed

desiccation pathways (Costa et al., 2017; VanBuren et al., 2017). Resurrection plants have a conserved set of molecular signatures associated with desiccation tolerance (Illing et al., 2005; Zhang et al., 2018), but the underlying genomic basis of this trait is largely unknown. The co-option of seed and desiccation-associated pathways in resurrection plants can occur through a broad range of mechanisms at the gene, pathway or network level. Changes in gene regulation at network hubs could drive pathway-level rewiring to upregulate a cascade of desiccation-related mechanisms. Changes in *cis*-regulation at isolated nodes in a network could shift the stoichiometry or abundance of endpoint metabolites and proteins related to desiccation. Gene and genome duplication can also drive adaptive evolution by providing additional copies for pathway- and gene-level sub- and neofunctionalization.

Desiccation tolerance is prominent in Linderniaceae (order Lamiales) within the clade spanning *Craterostigma* and *Lindernia* (Rahmanzadeh et al., 2005). *Craterostigma plantagineum* is a model resurrection plant (Bartels and Salamini, 2001) native to rocky outcrops of sub-Saharan Africa. All *Craterostigma* and some desert-adapted *Lindernia* species are desiccation tolerant, but most *Lindernia* species are desiccation sensitive. *Lindernia brevidens* is unusual as it displays desiccation tolerance despite an endemic habitat in the montane rainforests of eastern Africa that never experience seasonal drying (Phillips et al., 2008). Desiccation tolerance is likely an ancestral trait in this group (Fischer et al., 2013) and was retained in *L. brevidens* before its radiation to the tropical rainforest. *Lindernia* is paraphyletic (Fischer et al., 2013) and species outside of the clade containing *Craterostigma* and *L. brevidens* such as *L. subracemosa* are desiccation sensitive. This diversity in desiccation tolerance makes *Lindernia* an excellent comparative system to test the contribution of gene duplication, *cis*-elements, and pathway rewiring in the evolution of desiccation tolerance. Here, we assembled high-quality reference

genomes for the desiccation-tolerant *L. brevidens* and -sensitive *L. subracemosa.* Detailed comparative genomics and differential co-expression network analysis allowed us to survey the genetic basis of desiccation tolerance in *Lindernia*.

RESULTS

Comparative grade reference genomes for *Lindernia*

Craterostigma plantagineum is a well-studied model for the evolution of desiccation tolerance in eudicots, but its highly complex, octoploid genome has hindered genome-scale analyses. *L. brevidens* and *L. subracemosa* are diploid with relatively small genomes (270 and 250 Mb respectively), providing an excellent alternative system (Figure 1).

We generated high-quality reference genomes for both *Lindernia* species using a PacBio-based, single-molecule real-time (SMRT) sequencing approach. In total, we generated 21.7 Gb and 17.9 Gb of filtered PacBio data, collectively representing 80.3x and 71.6x coverage for *L. brevidens* and *L. subracemosa,* respectively (Supplemental Figure 1). Raw PacBio reads were error corrected and assembled using the long-read assembler Canu (Koren et al., 2017), which is optimized to avoid collapsing highly repetitive and tandemly duplicated regions. Contigs were polished using high-coverage Illumina data with Pilon (Walker et al., 2014) to remove residual errors. The *L. brevidens* assembly spanned 265 Mb across 267 contigs with a contig N50 of 3.6 Mb. The *L. subracemosa* assembly was slightly smaller, at 246 Mb with 328 contigs and an N50 of 1.9 Mb (Table 1). The total assembly sizes were consistent with the estimated genome sizes of 270 and 250 Mb based on flow cytometry. *Lindernia* species are primarily self-pollinated with low residual within-genome heterozygosity, which contributed to the high contiguity and relatively simple graph-based assembly structures (Supplemental Figures 2 and 3).

We used high-throughput chromatin conformation capture (Hi-C) to generate a chromosome-scale assembly of *L. brevidens*. The Hi-C based Illumina reads were mapped to the draft assembly using bwa (Li, 2013) followed by filtering and proximity-based clustering using the Juicer pipeline (Durand et al., 2016) (Supplemental Table 1). This approach yielded 14 high confidence clusters corresponding to the haploid chromosome number in *L. brevidens* (2n=2x=28; Figure 2). In total, 121 contigs were ordered and oriented into 14 scaffolds collectively representing 94.7% of the assembly (249 out of 263 Mb; Supplemental Table 2). This included anchoring 98.8% of the predicted gene models. The repetitive element density was inversely correlated with gene density, and most chromosomes contained large tracts of retrotransposons, which likely correlate with centromere position (Figure 3).

The genomes of *L. brevidens* and *L. subracemosa* were of similar size and the same karyotype, suggesting that they should have comparable repetitive element and gene composition. Long terminal repeat retrotransposons (LTR-RTs) were the most abundant repetitive elements in both genomes, and they collectively spanned 34% (92.0 Mb) and 31% (77.4 Mb) of the *L. brevidens* and *L. subracemosa* genomes, respectively (Table 1). Despite the similar overall LTR composition, *L. subracemosa* had significantly more intact LTRs compared to *L. brevidens* (1,972 vs 1,025; Wilcoxon Rank-Sum, $P < 0.05$). The distribution of LTR-RT insertion time was similar in both species, and most intact elements inserted within the last million years (Figure 4). These findings suggest that LTR-RTs are similarly active in both genomes but may fractionate more quickly in *L. brevidens*.

The overall gene composition was similar in both *Lindernia* species, though *L. subracemosa* had more annotated gene models. *Ab initio* gene prediction using the dehydration time course RNAseq data and protein similarities to other angiosperms identified 27,204 and

33,344 gene models in *L. brevidens* and *L. subracemosa,* respectively (Table 1). We assessed annotation quality using the Benchmarking Universal Single-Copy Orthologs (BUSCO) pipeline and found 91% and 90% (1,319 and 1,298) of the 1,440 genes in the Embryophyta dataset present in the *L. brevidens* and *L. subracemosa* assemblies. This proportion is comparable with results from other recent PacBio-based genomes.

Comparative genomics of *Lindernia*

The *L. brevidens* and *L. subracemosa* genomes were largely collinear based on whole-genome alignment, and 24,053 L*. brevidens* genes had syntenic orthologs in *L. subracemosa*. Roughly 70% of the genomes ere conserved in 2:2 syntenic blocks, supporting a shared, ancient whole-genome duplication event (WGD) in both species (Figure 5; Supplemental Figures 4 and 5). Six of the seven ancestral homeologous chromosome pairs from the WGD were intact in *L. brevidens*, including modern chromosome pairs: 1 and 13, 2 and 14, 3 and 5, 6 and 9, 7 and 10, 8 and 11 (Figure 5A). Two of the ancestral homeologous chromosomes were fused in modern chromosome 5, and chromosome 12 contained fragments from several ancestral chromosomes. Chromosomal rearrangements were difficult to identify in *L. subracemosa* given its contig-level assembly, but there were no obvious rearrangements based on macrosynteny with *L. brevidens* (Supplemental Figure 5). The ancestral subgenomes were heavily fractionated, and only 7,742 gene pairs were retained in duplicate in *L. brevidens* and 8,452 in *L. subracemosa* based on synteny. Gene-level fractionation was biased toward a dominant subgenome that contained significantly more genes (Figure 5B; Supplemental Figure 6).

We identified patterns of gene duplication and loss that may be related to the evolution of desiccation tolerance and other lineage-specific traits. Most gene pairs from the WGD were

either retained in duplicate or fractionated to single copies in both species, including 11,874 single-copy genes (1:1) and 7,568 duplicated (2:2) genes in both genomes (Table 2). We identified 3,200 lineage-specific genes in *L. brevidens* (1:0 or 2:0) and 7,067 lineage-specific genes in *L. subracemosa* (0:1 or 0:2) based on synteny. The higher number of lineage-specific genes in *L. subracemosa* was likely related to differences in total annotated gene number (27,204 vs 33,344). The lineage-specific genes in *L. brevidens* were enriched in gene ontology (GO) terms related to chlorophyll biosynthesis and metabolism, regulation of mitosis, and response to heat, which may suggest a role for these pathways in desiccation tolerance (Supplemental Table 3).

New genes can arise through tandem gene duplication (TDs), and TDs are associated with adaptive evolution (Cannon et al., 2004), including of desiccation tolerance in other resurrection plants (VanBuren et al., 2015; VanBuren et al., 2018). *L. brevidens* and *L. subracemosa* had a similar overall number of tandem genes but major differences in array size. *L. brevidens* had 2,673 tandem arrays containing 5,345 genes with array sizes ranging from 2 to 24 members. *L. subracemosa* had 3,404 tandem arrays across 6,809 genes with array sizes ranging from 2 to 31 (Figure 6). Through cross-referencing with syntenic gene pairs, we found that most tandem arrays were conserved between *L. brevidens* and *L. subracemosa*. Only 153 tandem arrays were specific to *L. brevidens* and 247 arrays were specific to *L. subracemosa*. Though tandem gene arrays were generally conserved, array sizes were highly variable and few contained the same number of genes between species (Figure 6B). Together these data suggest that most TD events are ancestral, but that each species has undergone unique array expansion and contraction.

182

183 **Global expression patterns and desiccation-related network rewiring**

184 To construct a comparative framework of genes related to desiccation, we conducted parallel
185 sampling of leaf tissue during desiccation and rehydration time courses in *L. brevidens* and *L.*
186 *subracemosa*. Parallel sampling between species allowed us to distinguish between genes
187 involved in typical dehydration responses from those specifically related to desiccation tolerance.
188 Sampling ranged from mild dehydration stress (relative water content (RWC) 53-56%; 3 days)
189 through severe dehydration (RWC 23-27%; 7 days) and desiccation (RWC 6-9%; 10 and 14
190 days), followed by 24 and 48 hours post rehydration (Figure 7). RWC was 53-56% at day three
191 and fell below 10% after ten days of drought in both species (Figure 7A). *L. subracemosa* plants
192 were largely dead upon rehydration and *L. brevidens* plants were mostly viable and
193 physiologically active at 48 hours post rehydration (RWC 44%) (Figure 7).

194 The greatest changes in gene expression occurred at two timepoints during the transition
195 from mild to severe dehydration stress and from desiccated to rehydrated (Supplemental Table
196 4). The number of differentially expressed (DE) genes between well-watered and mild
197 dehydration (F vs D3) were relatively similar in both species (5,322 vs 4,824 in *L. subracemosa*
198 and *L. brevidens*). Many syntenic gene pairs had similar expression levels, with 581 upregulated
199 and 133 downregulated in both species. Significantly more genes were DE between mild and
200 severe dehydration stress (D3 vs D7), with 4,329 and 9,227 DEGs respectively. A similar
201 proportion of syntenic gene pairs was upregulated in both species at D7 (581), but significantly
202 more gene pairs were similarly down-regulated (1,396) compared to mild dehydration stress.
203 This pattern suggests that there is conservation of downregulated pathways in desiccation-
204 sensitive and -tolerant species.

In L. *brevidens*, gene expression was relatively stable from severe dehydration to desiccation (D7, D10, and D14), whereas a high proportion of genes were DE in *L. subracemosa* during the transition to desiccation. This reflects the stability of desiccated *L. brevidens* and the imminent death of *L. subracemosa*. Few genes were similarly DE in both species under severe dehydration and desiccation (D7 vs D10: 50 and 62; D10 vs D14: 3 and 0, up and down-regulated, respectively; Supplemental Table 4). A substantial proportion of syntenic gene pairs (2,065) were similarly upregulated in both species during early rehydration (24 hours), supporting the conclusion that there is conserved activation of repair pathways. Expression changes in both species were minimal between 24 and 48 hours post rehydration. Though *L. subracemosa* had some transcriptional response post rehydration, this was not sufficient to repair the extensive desiccation-induced damage. Together, the divergent expression patterns suggest that there is extensive upregulation of distinct pathways with desiccation-specific roles.

We conducted GO enrichment analysis of the gene pairs that were uniquely upregulated in *L. brevidens*, with no change or a decrease in expression in *L. subracemosa*. We reasoned that such genes are likely to be specific to the induction of desiccation tolerance. Most of the GO terms enriched among genes upregulated in mild dehydration stress (D3) were related to responses to abiotic stress and secondary metabolite biosynthesis (Supplemental Table 5), suggesting early activation of protective mechanisms. There were only a few GO terms enriched among genes upregulated in severe dehydration (D7) and desiccation (D10 and D14), including terms related to transport, vacuole organization, ion homeostasis, and RNA modification (Supplemental Table 5). Most GO terms of genes uniquely downregulated in *L. brevidens* under mild and severe dehydration stress were related to photosynthesis processes, suggesting that the

227 photosynthetic apparatus is inactivated early under mild dehydration compared to *L.*
228 *subracemosa* (Supplemental Table 6).

229 The large-scale expression changes unique to desiccation in *L. brevidens* may be driven
230 by changes in *cis* regulation. Genes with unique desiccation-related expression in *L. brevidens*
231 were enriched with *cis*-regulatory elements associated with dehydration and ABA-mediated
232 responses as well as seed development pathways (Supplemental Table 7). Enriched *cis*-elements
233 associated with typical ABA-mediated dehydration responses included ABF1 and ABF2
234 (Yoshida et al., 2015) among others. Enriched seed maturation-associated *cis*-elements included
235 bZIP53 (Alonso et al., 2009), ABA-responsive element binding protein 3 (AREB3) (Nakashima
236 et al., 2009), and ABI5 (Lopez-Molina et al., 2001). *Cis*-elements at the interface of heat and
237 dehydration-mediated ABA responses such as Heat Stress Factor A6b (HSFA6B) (Huang et al.,
238 2016) and HSF7, were also enriched in desiccation-related genes. Enriched *cis*-elements in genes
239 downregulated during desiccation in *L. brevidens* had wide roles in plant growth and
240 development, hormone responses, and photosynthesis (Supplemental Table 8).

241 To compare network-level gene expression in *L. brevidens* and *L. subracemosa*, we
242 utilized a weighted correlation network analysis (WGCNA) approach (Langfelder and Horvath,
243 2008) across the dehydration and rehydration RNAseq time courses. This comparative co-
244 expression network approach allowed us to parse conserved dehydration-related patterns from
245 desiccation specific pathway rewiring. After filtering genes with low expression (see methods),
246 we constructed two co-expression networks, with 14,246 genes in ten modules for *L. brevidens*
247 and 14,075 genes in nine modules for *L. subracemosa* (Figure 8, Supplemental Figures 7 and 8).
248 Based on their temporal dynamics co-expression modules could be broadly classified into three
249 groups: 1) high expression in well-watered tissue but downregulation in dehydration

/desiccation, 2) expression during early dehydration, 3) sustained high expression throughout dehydration and desiccation (Figure 8). Modules 1, 2, 3, and 5 in the *L. brevidens* network and modules 2, 4, and 5 in *L. subracemosa* had high expression in well-watered and rehydrating conditions with decreasing expression throughout dehydration/desiccation timepoints (Figure 8). Modules 7 and 8 in *L. brevidens* and 3, 7, and 8 in *L. subracemosa* were involved in early dehydration responses, with a peak expression at 3 or 7 days of dehydration. Modules 4, 6, and 10 in *L. brevidens* and modules 1 and 3, in *L. subracemosa* had sustained dehydration and desiccation-induced expression.

We compared module overlap between the networks to identify patterns of conservation and species-specific divergence. Modules down-regulated during severe dehydration and desiccation were largely conserved between the two species, with 74% of gene pairs falling in the same group of modules (Supplemental Figure 9). Modules upregulated during dehydration and desiccation had comparatively little overlap between species, and only 43% of gene pairs were in overlapping modules (Supplemental Figure 9). This indicates that there was significant pathway rewiring during desiccation. Network wide *cis* regulatory element enrichment patterns mirrored the observations comparing pairwise differentially expressed genes. Desiccation-associated modules from the *L. brevidens* network were enriched in dehydration-associated ABA-responsive *cis* elements and seed maturation-associated *cis* elements including bZIP53, AREB3, and ABI5 among others (Supplemental Table 9).

Unique desiccation-related pathways in *L. brevidens*

The similarities between seed and vegetative desiccation suggest overlapping pathways, which is supported by expression data from several resurrection plant lineages (Costa et al., 2017; VanBuren et al., 2017). We identified a wide range of seed-specific genes and pathways that were expressed only under dehydration in *L. brevidens* compared to syntenic orthologs in *L. subracemosa* (Supplemental Table 10). Seed storage proteins serve as a reserve of nitrogen, carbon and sulfur for germinating seeds, and they likely play a role in seed longevity (Nguyen et al., 2015). Orthologs to genes encoding 2S and 12S seed storage proteins were generally upregulated in *L. brevidens* under desiccation, and syntenic orthologs in *L. subracemosa* were not expressed or were expressed highly in well-watered conditions (Supplemental Table 10). Delay of germination 1 (DOG1) is an essential component of seed dormancy regulation, and its expression affects hundreds of seed-related genes (Dekkers et al., 2016). *DOG1* was highly expressed in well-watered *L. brevidens* tissues but was down-regulated during desiccation. The *L. subracemosa DOG1* transcript had a low basal level expression in all time points.

Oil bodies are lipid organelles filled with triacylglycerols that function as high-density energy reserves during seed germination. Oil bodies accumulate in desiccated leaf tissue of *Oropetium*, and likely play a role in desiccation tolerance. Oil bodies are enveloped with oleosin structural proteins that prevent membrane coalescence and protect membrane integrity during freeze-thaw cycles (Shimada et al., 2008). Oil body membranes are also studded with calcium binding caleosin proteins that are associated with oil body degradation (Poxleitner et al., 2006) and general stress response pathways (Shen et al., 2014). *L. brevidens* and *L. subracemosa* had a similar number of genes for oleosin (8 vs 7) and caleosin (4 vs 3) proteins, though *L. brevidens* had more retained whole genome and tandem duplicates (Figure 9). Most oleosin and caleosin genes had low or undetectable expression in well-watered tissue, but several were induced during

295 progressive dehydration and desiccation. Six oleosin genes in *L. brevidens* and three oleosin
296 genes in *L. subracemosa* were upregulated in desiccating tissue, with most having a peak
297 expression of less than 30 transcripts per million (TPM). *L. brevidens* had a pair of retained
298 duplicated oleosin genes where one ortholog peaked at > 500 TPMs in desiccating tissue and the
299 single syntenic ortholog in *L. subracemosa* had a relatively low expression. No caleosin genes
300 were upregulated in *L. subracemosa* upon desiccation, but a pair of syntenic orthologs in *L.*
301 *brevidens* were abundantly expressed (Figure 9).

302 Early light induced proteins (ELIPs) are predicted to bind chlorophyll and function in
303 photoprotection under high light and other abiotic stresses. The *L. subracemosa* genome had four
304 genes encoding ELIP proteins, including a pair of syntelogs retained from the WGD event
305 (Figure 10). Expression of two *ELIP* genes was hardly detectable during the surveyed
306 timepoints, and two others were highly expressed during dehydration. The number of ELIPs in *L.*
307 *subracemosa* was similar to that in other desiccation-sensitive angiosperms, and their
308 dehydration-induced expression was consistent with the hypothesized protective mechanisms
309 (Hayami et al., 2015). By contrast, the *L. brevidens* genome had undergone a dramatic expansion
310 of ELIP genes with 26 in total, including a large tandem array of 19 duplicates. This large
311 tandem array was collinear to a pair of retained syntenic orthologs in *L. subracemosa* and a
312 single retained gene copy in *L. brevidens*. Nearly all the *ELIP* genes in this array, and dispersed
313 copies throughout the genome, were highly expressed during severe dehydration, desiccation,
314 and rehydration, but they were hardly expressed in well-watered and mildly dehydrated tissue
315 (Figure 10B). The tandem array was syntenic with the highly expressed ortholog in *L.*
316 *subracemosa*, and the single copy syntelog in *L. brevidens* and its syntenic ortholog in *L.*
317 *subracemosa* were not expressed in dehydrated tissue. This suggests an ancestral

318 subfunctionalization of this duplicated pair where only one gene copy was involved in
319 dehydration-related responses. After the divergence of *L. brevidens* and *L. subracemosa*, the
320 dehydration-specific syntelog likely underwent massive tandem proliferation in *L. brevidens*.
321 In Arabidopsis, STAY-GREEN (SGR) proteins are key regulators of chlorophyll
322 degradation and they are typically upregulated under abiotic stresses (Sakuraba et al., 2014b).
323 Syntenic orthologs of SGR were highly expressed in both *Lindernia* species during
324 dehydration/desiccation. STAY-GREEN LIKE (SGRL) proteins are negative regulators of
325 chlorophyll degradation, and overexpression of SGRL2 leads to a stay-green phenotype
326 (Sakuraba et al., 2014a). The syntenic ortholog of SGRL was highly induced during desiccation
327 in *L. brevidens*, but the *L. subracemosa* ortholog had no detectable expression (Supplemental
328 Table 10).
329 Carbohydrate metabolism is heavily shifted during desiccation, and sucrose, trehalose,
330 and short chain oligosaccharides function as osmoprotectants to stabilize cellular
331 macromolecules. Sucrose is the most abundant carbohydrate in most resurrection plants, and
332 accumulation of sucrose distinguishes desiccation-sensitive and -tolerant *Eragrostis* species
333 (Illing et al., 2005). *Craterostigma* and *L. brevidens* accumulate the unusual C8 sugar 2-octulose
334 in photosynthetic tissues, which serves as a reservoir of sucrose accumulation during desiccation
335 (Bianchi et al., 1991; Phillips et al., 2008). Transketolase 7 and 10 catalyze the formation of
336 octulose-8-phosphate in *Craterostigma* (Zhang et al., 2016), and the orthologous transketolase
337 genes were highly expressed in *L. brevidens* leaf tissue (Supplemental Table 10). This included
338 two pairs of syntenic 1:1 orthologs and a trio of retained 2:1 duplicates with upregulation in well-
339 watered and rehydrating tissue in *L. brevidens* and no or little expression in *L. subracemosa*.

LATE EMBRYOGENESIS ABUNDANT (LEA) proteins are predicted to have protective functions that are essential for desiccation tolerance (Hoekstra et al., 2001; Goyal et al., 2005; Hundertmark and Hincha, 2008). We identified 77 and 82 LEA protein-encoding genes in *L. brevidens* and *L. subracemosa*, respectively (Supplemental Table 11). Orthologs were assigned for 70 of these genes. About half of the identified LEAs were classified to the LEA_2 group whereas the second largest LEA group was the LEA_4 group (14 and 12 genes in *L. brevidens* and *L. subracemosa*, respectively). Nine LEA genes in *L. brevidens* (one Dhn, five LEA_2, two LEA_4 and one LEA_5) were derived from gene duplication events, which suggests that these genes may have functions related to desiccation tolerance in *L. brevidens*. For example, *LEA5-2* occurs in one copy in *L. subracemosa* (*LsLEA5-2*) and has two orthologs in *L. brevidens*, i.e., *LbLEA5-2* and *LbLEA5-3*. The *LsLEA5-2* gene showed negligible expression in *L. subracemosa* under control and dehydration conditions, whereas *LbLEA5-2* was among the highest expressed *LEA* genes upon dehydration in *L. brevidens* (Supplemental Data Set 1).

Almost one third of *L. brevidens LEA* genes showed expression levels 30 times higher or more than *L. subracemosa* orthologs during late dehydration (10d and 14d) (Supplemental Data Set 1). Most *L. subracemosa* orthologs had reduced or no expression in all surveyed timepoints, suggesting that there was a massive rewiring of expression networks (Supplemental Data Set 1). For example, *LEA1-3*, *LEA1-4*, *LEA2-14*, *LEA2-19*, *LEA4-1*, *LEA4-6*, *LEA4-7*, and *LEA5-2* showed very high expression in fully hydrated *L. brevidens* but no expression in fully hydrated *L. subracemosa*.

DISCUSSION

Genomic resources are abundant for resurrection plants, but the lack of suitable dehydration-sensitive outgroups has limited genomic insights into the origin and pathways controlling

desiccation tolerance. Here, we leveraged a unique comparative system of closely related desiccation-tolerant and -sensitive species to identify gene and pathway level changes associated with the evolution of desiccation tolerance. This approach allowed us to distinguish dehydration pathways conserved in all plants from desiccation-specific processes observed only in resurrection plants.

L. brevidens and *L. subracemosa* have a similar overall genome size and gene number, and most genes were likewise retained as singletons (1:1) or duplicates (2:2) after their shared WGD event. The genomes have no significant differences in architecture, rRNAs, repetitive element composition, or clustering of desiccation-related genes. These features were previously proposed to contribute to desiccation tolerance in other resurrection plant lineages (Xiao et al., 2015; Costa et al., 2017). Instead, our data indicate that desiccation tolerance in *L. brevidens* is driven by a complex cascade of *cis*-regulatory element-mediated pathway rewiring, tandem duplication, and preferential gene retention and neofunctionalization.

Gene expression patterns are dramatically divergent in dehydration and rehydration time-course data between *L. brevidens* and *L. subracemosa*. Only a few of the syntenic orthologs are similarly expressed in both species, and co-expression network modules are largely rewired. Early dehydration responses have surprisingly little overlap, suggesting that the gene expression program and signals leading to tolerance are already apparent upon mild dehydration. Gene expression is most dynamic between the day 3 and day 7 timepoints, when the plants shift from moderate to severe dehydration stress. This likely reflects major shifts in leaf water potential, photosynthesis, oxidative stress, and cellular damage. Gene expression is stabilized after moderate drought and desiccation in *L. brevidens*, which reflects the successful deployment of

protective mechanisms. By contrast, the dynamic and chaotic expression patterns in desiccating *L. subracemosa* may reflect last-ditch efforts to avoid imminent senescence.

Drought and seed development are linked by the common stress of water deficit. Vegetative and seed desiccation processes are strikingly similar, and overlapping pathways have been identified in resurrection plants (Costa et al., 2017; VanBuren et al., 2017). This includes accumulation of osmoprotectants, expression of LEA proteins, and free radical scavenging systems, as well as down-regulation of photosynthesis and dismantling the photosynthetic apparatus. Drought responses and seed development are similarly regulated by ABA-related signaling, and both elicit comparable downstream responses (Nakashima and Yamaguchi-Shinozaki, 2013). Several important transcription factors involved in dehydration and seed-related processes are preferentially retained in *L. brevidens* compared to *L. subracemosa*, which may allow high-level pathway rewiring. Desiccation-related genes such as *ELIP*s or *LEA*s have increased in copy number in *L. brevidens* via tandem gene duplication. For genes encoding proteins with structural, enzymatic or chaperone functions, tandem duplications may serve to increase their absolute abundance to surpass a threshold required for desiccation tolerance. Expansion of *ELIP*s has been observed in several resurrection plants including *C. plantagineum* (Bartels et al., 1992), *Selaginella lepidophylla* (VanBuren et al., 2018) and *Boea hygrometrica* (Xiao et al., 2015). ELIPs likely play an important role in protecting the photosynthetic apparatus and bind excess chlorophyll during prolonged desiccation (Ref Alamillo and Bartels 2001). The repeated duplication of *ELIP*s may be a hallmark of convergent evolution of desiccation tolerance across land plants. Expression patterns can also be shifted by changes in *cis*-regulatory elements, as was previously observed in *LEA* genes from *C. plantagineum* and *L. brevidens* (van den Dries et al., 2011; Giarola et al., 2018). The enrichment of seed related *cis*-regulatory

409 elements in modules uniquely upregulated in *L. brevidens* is likely the result of novel *cis*-element
410 acquisition in desiccation-related genes and activation of seed-related transcription factors.
411 Desiccation tolerance likely evolved from a complex, additive series of gene duplications
412 and pathway rewiring rather than a simple master regulatory switch. Naturally drought-tolerant
413 species could undergo favorable duplication of *ELIP*s or rewiring of LEA proteins to promote a
414 quasi-desiccation tolerant state. These responses could be further refined through accumulation
415 of additive features to surpass the threshold required for surviving anhydrobiosis. This step-wise
416 hypothesis is supported by the continuum of desiccation tolerance, where the magnitude and
417 duration of tolerance varies across species. *Craterostigma* can tolerate more rapid desiccation,
418 and recover more completely than *L. brevidens*, and older leaf tissue in *L. brevidens* is often
419 desiccation sensitive.
420 The recovery rate in *L. brevidens* is related to environmental factors including developmental
421 stage, rate of drying, and dehydration priming. This comparatively weak desiccation tolerance
422 may reflect relaxed selection in the drought-free rainforest habitat of *L. brevidens*. Desiccation
423 tolerance is ancestral in the clade spanning *L. brevidens* and *Craterostigma* (Fischer et al., 2013),
424 and some protective mechanisms were likely present in the shared ancestor with *L. subracemosa*.
425 This may also explain the partial induction of seed and vegetative desiccation-associated
426 pathways in *L. subracemosa* compared to the typical dehydration responses in other species. The
427 trajectory from sensitive to desiccation tolerant is a complex, multistep process, and future work
428 in intermediate or weakly desiccation-tolerant species will help uncover the origins of this trait.
429
430 **METHODS**
431 **Growth conditions and sampling**

Lindernia brevidens and *L. subracemosa* were grown as previously described (Phillips et al., 2008). Voucher specimens have been deposited: *Lindernia brevidens* Kenya, Taita Hills, E. Fischer 8022 (KOBL)(= Herbarium Koblenz) and *Lindernia subracemosa* Rwanda, Uwinka, E. Fischer 1350 (BG-Bonn 19990-2, KOBL). Plants were propagated via cuttings and maintained under day/night temperatures of 22 and 18°C respectively, under fluorescent lighting with an intensity of 80 μE m^{-2} sec^{-1} and 16/8 h photoperiod. *L. brevidens* and *L. subracemosa* were grown in the same chamber to minimize environmental variance. For the desiccation and rehydration time courses plants were allowed to gradually dry for a period of 30 days with sampling in triplicate with three independent plants at D3, D7, D14, D21, and D30. Plants were rehydrated and sampled at 24 and 48 h post rehydration. Samples were always taken at the same time of the day, 6 h after the onset of light, to minimize effects associated with circadian oscillation. Leaf tissue for RNAseq was flash frozen in liquid nitrogen and stored at -80°C. Relative water content (RWC) measurements were calculated using the equation: RWC = [(FW-DW)/ (SW-DW)], where FW, DW and SW indicate fresh weight of the leaf tissue, dry weight, and saturated weight. Dry weight was obtained after drying tissue at 80°C for 48 h and saturated plant weights were obtained after submerging leaf tissue in water for 24 h. Three replicates of RWC measurements were collected for each time point.

Nucleic acid extraction, library construction, and sequencing

High molecular weight (HMW) genomic DNA for PacBio and Illumina sequencing was isolated from young leaf tissue of growth chamber-grown *L. brevidens* and *L. subracemosa*. DNA was isolated using a modified nucleus preparation (Zhang et al., 1995) followed by phenol chloroform purification to remove residual contaminants. PacBio libraries were constructed and

size selected for 25 kb fragments on the BluePippen system (Sage Science) followed by purification using AMPure XP beads (Beckman Coulter). Libraries were sequenced on a Sequel platform with V4 software and V2 chemistry. In total, 2,054,566 filtered subreads spanning 21.7 Gb were sequenced for *L. brevidens* and 1,615,065 reads spanning 17.9 Gb were sequenced for *L. subracemosa*. This represents 80.3x and 71.6x coverage for *L. brevidens* and *L. subracemosa,* respectively. Illumina DNAseq libraries for error correction were constructed using the KAPA HyperPrep Kit (Kapa Biosystems) following the manufacturer`s instructions. Libraries were sequenced on an Illumina HiSeq4000 under paired-end 150 bp mode.

For RNAseq analysis, total RNA was extracted from 200 mg of ground *L. brevidens* and *L. subracemosa* leaf tissues using an Omega-biotek E.Z.N.A. Plant RNA Kit (Omega-biotek), according to the manufacturer`s instructions. RNA quality was validated using gel electrophoresis and Qubit RNA IQ Assay (ThermoFisher). Two micrograms of total RNA was used for constructing Illumina TruSeq-stranded mRNA libraries following the manufacturer's protocol (Illumina). Libraries were pooled and sequenced on the Illumina HiSeq4000 under paired-end 150 bp mode. Three replicates were sequenced for each timepoint in each species.

Genome assembly

Genome sizes for *L. brevidens* and *L. subracemosa* were estimated using flow cytometry as previously described (Arumuganathan and Earle, 1991). The flow cytometry-based estimates of 270 Mb and 250 Mb for *L. brevidens* and *L. subracemosa*, respectively, were consistent with K-mer based analysis using Illumina WGS data. *Linderia* are mostly selfing, and unimodal k-mer distribution suggests low within-genome heterozygosity for both species. Raw PacBio reads were error-corrected and assembled using Canu (V1.4) (Koren et al., 2017). Based on previous

experience (Edger et al., 2017; VanBuren et al., 2018), Canu produced the most contiguous and accurate assembly for homozygous, diploid species compared to other leading long-read assemblers. The following Canu parameters were modified and all others were left as default: minReadLength=2500, GenomeSize=270Mb (or 250Mb), minOverlapLength=1000. Assembly graphs were visualized in Bandage (Wick et al., 2015). Draft contigs were polished with Pilon (V1.22) (Walker et al., 2014) using 79x and 58x coverage of Illumina paired-end 150 bp data for *L. brevidens* and *L. subracemosa,* respectively. Illumina reads were quality-trimmed using Trimmomatic (v0.33) (Bolger et al., 2014) and aligned to the draft contigs using bowtie2 (V2.3.0) (Langmead and Salzberg, 2012) with default parameters. Alignment rates in the first round of corrections were 96% and 97% respectively, suggesting that both the *L. brevidens* and *L. subracemosa* assemblies were largely complete. Parameters for Pilon were modified as follows: --flank 7, --K 49, and --mindepth 20. Pilon was run recursively three times with minimal corrections in the third round, supporting accurate indel correction.

HiC library construction and analysis

The *L. brevidens* draft genome was anchored into a chromosome-scale assembly using a Hi-C proximity-based assembly approach. The Hi-C library was constructed using 0.2 g of young leaf tissue from well-watered *L. brevidens* plants with the Proximo™ Hi-C Plant kit (Phase Genomics) following the manufacturer's protocol. The final library was size-selected for 300-600 bp and sequenced on the Illumina NexSeq 500 under paired-end 75 bp mode. In total, 178 million reads were used as input for the Juicer and 3d-DNA Hi-C analysis and scaffolding pipelines (Durand et al., 2016; Dudchenko et al., 2017) (Supplemental Table 1). Quality-filtered reads were aligned to the PacBio contigs using bwa (V0.7.16) (Li, 2013) with strict parameters (-

n 0) to prevent mismatches and non-specific alignments and the resulting sam files were used as input into the Juicer pipeline. Read pairs were merged and duplicates or near duplicates were removed prior to constructing the distance matrix. Contigs were ordered and oriented and assembly errors were identified using the 3d-DNA pipeline with default parameters (Dudchenko et al., 2017). The resulting hic contact matrix was visualized using Juicebox, and misassemblies and misjoins were manually corrected based on neighboring interactions. This approach identified 14 high-confidence clusters representing the haploid chromosome number in *L. brevidens*. The manually validated assembly was used to build pseudomolecules using the finalize-output.sh script from 3d-DNA, and chromosomes were renamed and ordered by size.

Genome annotation

Prior to genome annotation, long terminal repeat (LTR) retrotransposons were predicted using LTR harvest (genome tools V1.5.8) (Ellinghaus et al., 2008) and LTR Finder (v1.07) (Xu and Wang, 2007), and the LTR library was refined using LTR retriever (v1.8.0) (Ou and Jiang, 2018). Retrotransposons were classified as intact if they were flanked by full-length LTRs. The insertion time for each intact element was calculated using LTR_retriever with the formula of $T=K/2\mu$, where K is the divergence rate approximated by percent identity and μ is the neutral mutation rate estimated as $\mu=1 \times 10\text{-}8$ mutations per bp per year.

The filtered, non-redundant LTR library from LTR retriever was used as input for whole-genome annotation of LTR retrotransposons using RepeatMasker (http://www.repeatmasker.org/)(Chen, 2004). The *L. brevidens* and *L. subracemosa* genomes were annotated using the MAKER-P pipeline (Campbell et al., 2014). Transcript-based evidence for gene predictions was produced using the desiccation/rehydration time course RNAseq data. RNAseq reads were aligned to the

L. brevidens and *L. subracemosa* genomes using the splice aware aligner STAR (v2.6) (Dobin et al., 2013). Transcripts were identified using StringTie (v1.3.4) (Pertea et al., 2015) with default parameters and the –merge flag was used to combine the output from individual libraries. The sets of non-redundant transcripts were used as expressed sequence tag evidence and protein sequences from *Arabidopsis* (Lamesch et al., 2011) and UniprotKB plant databases(Boutet et al., 2007) were used as protein evidence. The custom LTR retrotransposon library produced by LTR retriever and Repbase libraries were used for repeat masking. *Ab initio* gene prediction was done using SNAP (Korf, 2004) and Augustus (3.0.2) (Stanke and Waack, 2003) with two rounds of iterative training for each species. The raw gene models were filtered to identify any residual repetitive elements using BLAST with a non-redundant transposase library. After filtering, a final set of 27,204 and 33,344 gene models was produced for *L. brevidens* and *L. subracemosa* respectively. Annotation quality was assessed using the benchmarking universal single-copy orthologs (BUSCO; v.2) (Simão et al., 2015) with the plant-specific dataset (embryophyta_odb9).

Comparative Genomics

Syntenic gene pairs within and between *L. brevidens* and *L. subracemosa* were identified using the MCSCAN toolkit (V1.1)(Wang et al., 2012) implemented in python (https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)). Gene models were aligned using LAST and hits were filtered to find syntenic blocks. Tandem gene duplicates were identified using all vs. all BLAST with a minimum e-value of 1e-5 and maximum gene distance of 10 genes. Macro and microsynteny plots, and syntenic block depths were plotted using the python version of MCScan. Genes were classified as lineage specific, if they had no syntenic

orthologs between the two species or hits from LASTAL with >70% nucleotide identity. The whole-genome duplication event (WGD) within *Lindernia* was identified using a combination of synteny and synonymous substitution rate (Ks) estimation between duplicated gene pairs. Comparison of *L. brevidens* and *L. subracemosa* in MCSCAN identified a 2:2 syntenic pattern with 7,742 and 8,452 duplicated gene pairs retained respectively. Duplicated regions span 70% of the *L. brevidens* genome and 72% of the *L. subracemosa* genome. Most duplicated regions were retained in large blocks, allowing chromosome pairs to be identified. Ks was estimated using KaKs_calculator with the NG model (Zhang et al., 2006) and a peak of 0.65 in *L. brevidens* and 0.69 for *L. subracemosa* was identified, which indicates that the event is shared between both species.

RNAseq analysis

Paired end Illumina RNAseq reads were trimmed by quality score and by adapter contamination using Trimmomatic (v0.33)(Bolger et al., 2014) with default parameters. The expression level of each gene was quantified using the pseudoaligner Kallisto (Bray et al., 2016) against the final gene models for *L. brevidens* and *L. subracemosa.* Parameters were left as default with 100 bootstraps per sample. Expression was quantified in Transcript Per Million (TPM), and a mean across the three replicates was used for single gene analysis and for constructing log2 transformed expression based heatmaps. Pairwise differentially expressed genes were identified using sleuth (Pimentel et al., 2017) implemented in R.

Co-expression network construction

The time course RNA-seq data were clustered into gene co-expression networks using the R package WGCNA (Langfelder and Horvath, 2008). Genes with less than an average TPM of 5 across all seven timepoints were filtered prior to network construction. A signed co-expression network was constructed for each species using a soft-thresholding power of 8 and a tree cut height of 0.15. All remaining parameters were left as default. In total, 14,246 genes were clustered into 10 modules for *L. brevidens* and 14,075 genes were clustered into 9 modules for *L. subracemosa.*

Cis-element identification

Cis-regulatory elements were identified using the Hypergeometric Optimization of Motif EnRichment program (HOMER; v4.10) (Heinz et al., 2010) using *cis*-elements from 529 plant transcription factors (O'Malley et al., 2016). *Cis* elements were identified in the 1 kb region upstream of the transcriptional start site (TSS) when known, or directly upstream of the start codon of each gene model. Promoters of gene models with detectable expression (TPM > 1) were used as background. Enrichment tests were performed using syntenic gene pairs with differential expression specific to *L. brevidens*, or gene models unique to modules upregulated or downregulated during desiccation in *L. brevidens*. A *P* value of < 0.00001 was used as cutoff for identifying enriched motifs in any comparison.

Identification of *LEA* genes

LATE EMBRYOGENESIS ABUNDANT (*LEA*) genes were retrieved from *L. brevidens* and *L. subracemosa* transcriptomes by BLAST and HMMER (http://hmmer.org/). *A. thaliana* and *C. plantagineum* LEA protein sequences were used for BLAST searches. HMM profiles for the

eight LEA families (DHN - PF00257, LEA_1 - PF03760, LEA_2 - PF03168, LEA_3 - PF03242, LEA_4 - PF02987, LEA_5 - PF00477, LEA_6 - PF10714 and SMP - PF04927) obtained from the Pfam database (http://pfam.xfam.org) (Finn et al., 2015) were used with the program hmmscan to search for LEA domain-containing proteins. Proteins identified with hmmscan were queried against the nr databank to confirm their classification as LEAs. Orthologs pairs were additionally confirmed by pairwise sequence alignments of the predicted protein sequences using EMBOSS Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/).

Accession Numbers

The genome assemblies, raw PacBio data, Illumina DNAseq, and RNAseq data are available from the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA). The RNAseq reads were deposited to NCBI SRA under BioProject PRJNA488068. The genome assemblies for *L. brevidens* and *L. subracemosa* have been deposited under BioProjects PRJNA489464 and PRJNA489465 respectively.

Supplemental Data

Supplemental Figure 1. Histogram of filtered PacBio subreads for (a) *Lindernia brevidens* and (b) *L. subracemosa*.

Supplemental Figure 2. Graph-based assembly of the *L. brevidens* genome.

Supplemental Figure 3. Graph based assembly of the *L. subracemosa* genome.

Supplemental Figure 4. Summary of genome wide syntenic blocks in *Lindernia*.

Supplemental Figure 5. Macrosyntenic dotplot between the *L. brevidens* and *L. subracemosa* genomes.

Supplemental Figure 6. Microsynteny showing a region with biased fractionation between *L. brevidens* and *L. subracemosa*.

Supplemental Figure 7. Weighted gene co-expression network in *L. brevidens*.

Supplemental Figure 8. Weighted gene co-expression network in *L. brevidens*.

Supplemental Figure 9. Overlap between *L. brevidens* and *L. subracemosa* co-expression networks.

Supplemental Table 1. Statistics of read mapping, filtering, and interactions for the HiC data.

Supplemental Table 2. Summary of HiC-based scaffolding.

Supplemental Table 3. Summary of GO terms enriched in *L. brevidens*-specific genes.

Supplemental Table 4: Summary of differential expressed genes during dehydration and rehydration in the two *Lindernia* species

Supplemental Table 5. Enriched GO terms in syntenic orthologs uniquely upregulated in *L. brevidens* with no change in expression in *L. subracemosa*

Supplemental Table 6. Enriched GO terms in syntenic orthologs uniquely downregulated in *L. brevidens* with no change in expression in *L. subracemosa*

Supplemental Table 7. Enriched *cis*-regulatory elements (CREs) in genes uniquely upregulated under desiccation in *L. brevidens.*

Supplemental Table 8. Enriched *cis*-regulatory elements (CREs) in genes uniquely downregulated under desiccation in *L. brevidens.*

Supplemental Table 9. Enriched CREs in desiccation associated co-expression modules in *L. brevidens*.

Supplemental Table 10. Expression of desiccation-related genes.

Supplemental Table 11. Number of LEA genes in *Lindernia brevidens* and *Lindernia subracemosa*

Supplemental Data Set 1. Expression of *LEA* genes.

Author Contributions

R.V. and D.B. designed and conceived research; R.V., C.M.W., and J.P. annotated genome features; X.S., collected desiccation and rehydration data. C.M.W constructed RNAseq, DNAseq, and HiC libraries; R.V., C.M.W., J.P., X.S., S.A., V.G., and D.B. analyzed data; R.V. wrote the paper. All authors read and approved the final manuscript.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgements

We thank Eberhard Fischer (University Koblenz Germany) for making *Lindernia* plants available originally.

FIGURE LEGENDS

Figure 1. Comparative desiccation tolerance system within Linderniaceae. Inferred phylogeny from Fischer et al. 2013 showing the two model desiccation-tolerant species (*Crateristigma plantagineum* and *Lindernia brevidens*) and the desiccation-sensitive outgroup (*Lindernia subracemosa*). Ploidy, karyotype, and genome size are shown on branches.

Figure 2. Hi-C clustering for pseudomolecule construction in *L. brevidens*. Post-clustering heat map of HiC-based intrachromosomal interactions in *Lindernia brevidens*. Pseudomolecules corresponding to the 14 haploid chromosomes are delineated by gray boxes.

Figure 3. Landscape of the *L. brevidens* genome. LTR-RT and CDS density are plotted in sliding windows of 50 kb with 25 kb step size for the 14 *L. brevidens* scaffolds (chromosomes). Red indicates high density and blue indicates low density of CDS and LTR-RTs in the heat map below each landscape.

Figure 4. Insertion time of intact LTR-RTs. The average insertion time calculated from the divergence of LTR pairs is plotted for the 1,025 intact LTR-RTs in *L. brevidens* and 1,972 intact LTR-RTs in *L. subracemosa.*

Figure 5. Comparative genomics of *L. brevidens* and *L. subracemosa*. (A) Syntenic dotplot of *L. brevidens* showing retained gene pairs from the recent WGD event. Each black dot represents a pair of retained genes. (B) Microsynteny between syntenic blocks of the *L. brevidens* (top) and *L. subracemosa* (bottom) genomes. Syntenic gene pairs between *L. brevidens* vs. *L. subracemosa* are shown by brown connections and retained WGD gene pairs within each genome are shown by gray connections. Genes are colored by orientation in *L. brevidens* (light blue are forward, dark blue are reverse) and *L. subracemosa* (light red are forward, dark red are reverse).

Figure 6. Comparison of tandem gene arrays in *Lindernia*. (A) Histogram of tandem array sizes. (B) Heatmap of tandem array size in syntenic orthologs between *L. brevidens* and *L. subracemosa*. Values are plotted as the proportion of tandem genes in each category against all the genes in that array size.

Figure 7. Overview of desiccation and rehydration processes in *Lindernia*. (A) Relative water content (RWC) of fresh leaf tissues (F), 3, 7. 10, and 14 days drought (D), and 24 and 48 hours post rehydration (R) in *L. brevidens* and *L. subracemosa*. Error bars represent the standard error of the mean with three replicates for each RWC measurement. (B) Representative *L. brevidens* (top) and *L. subracemosa* at various RWC.

Figure 8. Comparative co-expression networks during desiccation and rehydration in *Lindernia*. The mean expression of genes from modules in the co-expression network in *L. brevidens* (left) and *L. subracemosa* (right) are plotted for the seven timepoints. Conserved modules with upregulation during desiccation are highlighted in yellow in both networks, and modules with downregulation during desiccation are highlighted in gray.

Figure 9. Subfunctionalization of oleosin and caleosin genes in *L. brevidens*. Heatmaps of Log2 transformed expression of genes encoding oleosins (top) and caleosins (bottom) in *L. brevidens* and *L. subracemosa*. Syntenic orthologs are connected by brown lines including 1:2 and 2:2 orthologs between species.

691 **Figure 10.** Tandem proliferation of desiccation associated *ELIP* genes in *L. brevidens.* (A)
692 Microsynteny of a large tandem gene array in *L. brevidens* compared to the single gene syntenic
693 ortholog in *L. brevidens* and the two syntenic regions in *L. subracemosa*. (B) Log2 transformed
694 expression patterns of *ELIP*s in *L. subracemosa* (top) and *L. brevidens* (bottom). The large
695 tandem array and syntenic whole-genome duplicates are labeled.

696

References

Alonso, R., Oñate-Sánchez, L., Weltmeier, F., Ehlert, A., Diaz, I., Dietrich, K., Vicente-Carbajosa, J., and Dröge-Laser, W. (2009). A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation. The Plant Cell **21,** 1747-1761.

Arumuganathan, K., and Earle, E. (1991). Estimation of nuclear DNA content of plants by flow cytometry. Plant molecular biology reporter **9,** 229-241.

Baniaga, A.E., Arrigo, N., and Barker, M.S. (2016). The small nuclear genomes of Selaginella are associated with a low rate of genome size evolution. Genome biology and evolution **8,** 1516-1525.

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A., and Ashton, N.W. (2011). The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. science **332,** 960-963.

Bartels, D., and Salamini, F. (2001). Desiccation tolerance in the resurrection Plantcraterostigma plantagineum. A contribution to the study of drought tolerance at the molecular level. Plant Physiology **127,** 1346-1353.

Bartels, D., Hanke, C., Schneider, K., Michel, D., and Salamini, F. (1992). A desiccation-related Elip-like gene from the resurrection plant Craterostigma plantagineum is regulated by light and ABA. The EMBO journal **11,** 2771-2778.

Bianchi, G., Gamba, A., Murelli, C., Salamini, F., and Bartels, D. (1991). Novel carbohydrate metabolism in the resurrection plant Craterostigma plantagineum. The Plant Journal **1,** 355-359.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, btu170.

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., and Bairoch, A. (2007). Uniprotkb/swiss-prot. In Plant bioinformatics (Springer), pp. 89-112.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nature biotechnology **34,** 525.

Campbell, M.S., Law, M., Holt, C., Stein, J.C., Moghe, G.D., Hufnagel, D.E., Lei, J., Achawanantakun, R., Jiao, D., and Lawrence, C.J. (2014). MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. Plant physiology **164,** 513-524.

Cannon, S.B., Mitra, A., Baumgarten, A., Young, N.D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. BMC plant biology **4,** 10.

Chen, N. (2004). Using Repeat Masker to identify repetitive elements in genomic sequences. Current protocols in bioinformatics **5,** 4.10. 11-14.10. 14.

Costa, M., Artur, M., Maia, J., Jonkheer, E., Derks, M., Nijveen, H., Williams, B., Mundree, S.G., Jiménez-Gómez, J.M., and Hesselink, T. (2017). A footprint of desiccation tolerance in the genome of Xerophyta viscosa. Nature plants **3,** 17038.

Dekkers, B.J., He, H., Hanson, J., Willems, L.A., Jamar, D.C., Cueff, G., Rajjou, L., Hilhorst, H.W., and Bentsink, L. (2016). The Arabidopsis DELAY OF GERMINATION 1 gene affects ABSCISIC ACID INSENSITIVE 5 (ABI5) expression and genetically interacts with ABI3 during Arabidopsis seed development. The Plant Journal **85,** 451-465.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29,** 15-21.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., and Aiden, A.P. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science **356,** 92-95.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell systems **3,** 95-98.

Edger, P.P., VanBuren, R., Colle, M., Poorten, T.J., Wai, C.M., Niederhuth, C.E., Alger, E.I., Ou, S., Acharya, C.B., and Wang, J. (2017). Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (Fragaria vesca) with chromosome-scale contiguity. GigaScience.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC bioinformatics **9,** 18.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., and Sangrador-Vegas, A. (2015). The Pfam protein families database: towards a more sustainable future. Nucleic acids research **44,** D279-D285.

Fischer, E., Schäferhoff, B., and Müller, K. (2013). The phylogeny of Linderniaceae—The new genus Linderniella, and new combinations within Bonnaya, Craterostigma, Lindernia, Micranthemum, Torenia and Vandellia. Willdenowia **43,** 209-238.

Giarola, V., Udo Jung, N., Singh, A., Satpathy, P., and Bartels, D. (2018). Analysis of pcC13-62 promoters predicts a link between cis-element variations and desiccation tolerance in Linderniaceae. Journal of experimental botany**,** ery173.

Goyal, K., Walton, L.J., and Tunnacliffe, A. (2005). LEA proteins prevent protein aggregation due to water stress. Biochemical Journal **388,** 151-157.

Hayami, N., Sakai, Y., Kimura, M., Saito, T., Tokizawa, M., Iuchi, S., Kurihara, Y., Matsui, M., Nomoto, M., and Tada, Y. (2015). The responses of Arabidopsis early light-induced protein2 to ultraviolet B, high light, and cold stress are regulated by a transcriptional regulatory unit composed of two elements. Plant physiology **169,** 840-855.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell **38,** 576-589.

Hoekstra, F.A., Golovina, E.A., and Buitink, J. (2001). Mechanisms of plant desiccation tolerance. Trends in plant science **6,** 431-438.

Huang, Y.-C., Niu, C.-Y., Yang, C.-R., and Jinn, T.-L. (2016). The heat stress factor HSFA6b connects ABA signaling and ABA-mediated heat responses. Plant physiology **172,** 1182-1199.

Hundertmark, M., and Hincha, D.K. (2008). LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC genomics **9,** 118.

Illing, N., Denby, K.J., Collett, H., Shen, A., and Farrant, J.M. (2005). The signature of seeds in resurrection plants: a molecular and physiological comparison of desiccation tolerance in seeds and vegetative tissues. Integrative and Comparative Biology **45,** 771-787.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome research **27,** 722-736.

Korf, I. (2004). Gene finding in novel genomes. BMC bioinformatics **5,** 59.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., and Garcia-Hernandez, M. (2011). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic acids research **40,** D1202-D1210.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics **9,** 559.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods **9,** 357-359.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Lopez-Molina, L., Mongrand, S., and Chua, N.-H. (2001). A postgermination developmental arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in Arabidopsis. Proceedings of the National Academy of Sciences **98,** 4782-4787.

Lüttge, U., Beck, E., and Bartels, D. (2011). Plant desiccation tolerance. (Springer Science & Business Media).

Nakashima, K., and Yamaguchi-Shinozaki, K. (2013). ABA signaling in stress-response and seed development. Plant cell reports **32,** 959-970.

Nakashima, K., Fujita, Y., Kanamori, N., Katagiri, T., Umezawa, T., Kidokoro, S., Maruyama, K., Yoshida, T., Ishiyama, K., and Kobayashi, M. (2009). Three Arabidopsis SnRK2 protein kinases, SRK2D/SnRK2. 2, SRK2E/SnRK2. 6/OST1 and SRK2I/SnRK2. 3, involved in ABA signaling are essential for the control of seed development and dormancy. Plant and Cell Physiology **50,** 1345-1363.

Nguyen, T.-P., Cueff, G., Hegedus, D.D., Rajjou, L., and Bentsink, L. (2015). A role for seed storage proteins in Arabidopsis seed longevity. Journal of experimental botany **66,** 6399-6413.

O'Malley, R.C., Huang, S.-s.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. Cell **165,** 1280-1292.

Oliver, M.J., Tuba, Z., and Mishler, B.D. (2000). The evolution of vegetative desiccation tolerance in land plants. Plant Ecology **151,** 85-100.

Oliver, M.J., Guo, L., Alexander, D.C., Ryals, J.A., Wone, B.W., and Cushman, J.C. (2011). A sister group contrast using untargeted global metabolomic analysis delineates the biochemical regulation underlying desiccation tolerance in Sporobolus stapfianus. The Plant Cell Online **23,** 1231-1248.

Ou, S., and Jiang, N. (2018). LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. Plant physiology **176,** 1410-1422.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature biotechnology **33,** 290.

Phillips, J.R., Fischer, E., Baron, M., Van Den Dries, N., Facchinelli, F., Kutzer, M., Rahmanzadeh, R., Remus, D., and Bartels, D. (2008). Lindernia brevidens: a novel desiccation-tolerant vascular plant, endemic to ancient tropical rainforests. The Plant Journal **54,** 938-948.

Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. Nature methods **14,** 687.

Plancot, B., Vanier, G., Maire, F., Bardor, M., Lerouge, P., Farrant, J.M., Moore, J., Driouich, A., Vicré-Gibouin, M., and Afonso, C. (2014). Structural characterization of arabinoxylans from two African plant species Eragrostis nindensis and Eragrostis tef using various mass spectrometric methods. Rapid Communications in Mass Spectrometry **28,** 908-916.

Poxleitner, M., Rogers, S.W., Lacey Samuels, A., Browse, J., and Rogers, J.C. (2006). A role for caleosin in degradation of oil-body storage lipid during seed germination. The Plant Journal **47,** 917-933.

Proctor, M. (1990). The physiological basis of bryophyte production. Botanical Journal of the Linnean Society **104,** 61-77.

Rahmanzadeh, R., Müller, K., Fischer, E., Bartels, D., and Borsch, T. (2005). The Linderniaceae and Gratiolaceae are further lineages distinct from the Scrophulariaceae (Lamiales). Plant Biology **7,** 67-78.

Sakuraba, Y., Kim, D., Kim, Y.-S., Hörtensteiner, S., and Paek, N.-C. (2014a). Arabidopsis STAYGREEN-LIKE (SGRL) promotes abiotic stress-induced leaf yellowing during vegetative growth. FEBS letters **588,** 3830-3837.

Sakuraba, Y., Lee, S.-H., Kim, Y.-S., Park, O.K., Hörtensteiner, S., and Paek, N.-C. (2014b). Delayed degradation of chlorophylls and photosynthetic proteins in Arabidopsis autophagy mutants during stress-induced leaf yellowing. Journal of experimental botany **65,** 3915-3925.

Shen, Y., Xie, J., Liu, R.-d., Ni, X.-f., Wang, X.-h., Li, Z.-x., and Zhang, M. (2014). Genomic analysis and expression investigation of caleosin gene family in Arabidopsis. Biochemical and biophysical research communications **448,** 365-371.

Shimada, T.L., Shimada, T., Takahashi, H., Fukao, Y., and Hara-Nishimura, I. (2008). A novel role for oleosins in freezing tolerance of oilseeds in Arabidopsis thaliana. The Plant Journal **55,** 798-809.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31,** 3210-3212.

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics **19,** ii215-ii225.

van den Dries, N., Facchinelli, F., Giarola, V., Phillips, J.R., and Bartels, D. (2011). Comparative analysis of LEA-like 11-24 gene expression and regulation in related plant species within the Linderniaceae that differ in desiccation tolerance. New Phytologist **190,** 75-88.

VanBuren, R., Wai, C., Zhang, Q., Song, X., Edger, P., Bryant, D., Michael, T., Mockler, T., and Bartels, D. (2017). Seed desiccation mechanisms co-opted for vegetative desiccation in the resurrection grass Oropetium thomaeum. Plant, cell & environment **40,** 2292.

VanBuren, R., Wai, C.M., Ou, S., Pardo, J., Bryant, D., Jiang, N., Mockler, T.C., Edger, P., and Michael, T.P. (2018). Extreme haplotype variation in the desiccation-tolerant clubmoss Selaginella lepidophylla. Nature communications **9,** 13.

VanBuren, R., Bryant, D., Edger, P.P., Tang, H., Burgess, D., Challabathula, D., Spittle, K., Hall, R., Gu, J., and Lyons, E. (2015). Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum. Nature **527,** 508-511.

Vander Willigen, C., Pammenter, N., Mundree, S., and Farrant, J. (2001). Some physiological comparisons between the resurrection grass, Eragrostis nindensis, and the related desiccation-sensitive species, E. curvula. Plant Growth Regulation **35,** 121-129.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., and Young, S.K. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS one **9,** e112963.

Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.-h., Jin, H., Marler, B., and Guo, H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic acids research **40,** e49-e49.

Wick, R.R., Schultz, M.B., Zobel, J., and Holt, K.E. (2015). Bandage: interactive visualization of de novo genome assemblies. Bioinformatics **31,** 3350-3352.

Xiao, L., Yang, G., Zhang, L., Yang, X., Zhao, S., Ji, Z., Zhou, Q., Hu, M., Wang, Y., and Chen, M. (2015). The resurrection genome of Boea hygrometrica: A blueprint for survival of dehydration. Proceedings of the National Academy of Sciences **112,** 5833-5837.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic acids research **35,** W265-W268.

Yobi, A., Wone, B.W., Xu, W., Alexander, D.C., Guo, L., Ryals, J.A., Oliver, M.J., and Cushman, J.C. (2013). Metabolomic profiling in Selaginella lepidophylla at various hydration states provides new insights into the mechanistic basis of desiccation tolerance. Molecular plant **6,** 369-385.

Yoshida, T., Fujita, Y., Maruyama, K., Mogami, J., Todaka, D., Shinozaki, K., and YAMAGUCHI-SHINOZAKI, K. (2015). Four Arabidopsis AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signalling in response to osmotic stress. Plant, cell & environment **38,** 35-49.

Zhang, H.B., Zhao, X., Ding, X., Paterson, A.H., and Wing, R.A. (1995). Preparation of megabase-size DNA from plant nuclei. The Plant Journal **7,** 175-184.

Zhang, Q., Bartels, D., and Varshney, R. (2018). Molecular responses to dehydration and desiccation in desiccation-tolerant angiosperm plants. Journal of experimental botany.

Zhang, Q., Linnemann, T.V., Schreiber, L., and Bartels, D. (2016). The role of transketolase and octulose in the resurrection plant Craterostigma plantagineum. Journal of experimental botany **67,** 3551-3559.

Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G.K.-S., and Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics, proteomics & bioinformatics **4,** 259-263.

Tables:

Table 1: *Lindernia* genome assembly metrics

	L. subracemosa	*L. brevidens* V1 (draft)[#]	*L. brevidens* V2 (chromosome scale)
Number of contigs	328	327	144 (unplaced)[†]
Number of scaffolds	NA[*]	NA	14
Contig N50	1.9 Mb	3.7 Mb	18.7 Mb
Total length	246,514,821 bp	271,249,608 bp	263,221,403 bp
LTR composition	77.4 Mb (31%)[+]	92.0 Mb (34%)	92.0 Mb (34%)
Number of gene models	33,344	27,204	27,204

Notes:

[#]The *L. brevidens* V1 represents the contig level PacBio based assembly and V2 represents the chromosome scale assembly of *L. brevidens* anchored using Hi-C data.

[†] Unplaced scaffolds were not anchored into the chromosome scale assembly.

[*]The *L. subracemosa* and *L. brevidens* V1 assemblies are contig level and thus contain no scaffolds.

[+]Proportion (percentage) of the genome represented by LTRs.

Table 2: Comparison of biased fractionation following the shared WGD in *Lindernia*. Gene ratios are shown as *L. brevidens*: *L. subracemosa*.

Gene classification	Gene ratio	Number of Genes/pairs
Single copy (both species)	1:1	11,874
Duplicate retained (both species)	2:2	7,568
Duplicate retained (*L. brevidens*)	2:1	1,276
Duplicate retained (*L subracemosa*)	1:2	3,286
L. brevidens specific (single copy)	1:0	3,026
L. subracemosa specific (single copy)	0:1	6,183
L. brevidens specific (duplicated)	2:0	174
L. subracemosa specific (duplicated)	0:2	884