Massive tandem proliferation of ELIPs supports convergent evolution of desiccation tolerance across land plants

Robert VanBuren[1,2*], Jeremy Pardo[2,3], Ching Man Wai[1,2], Sterling Evans[1], 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 author: bobvanburen@gmail.com

Short title: Desiccation tolerance arose via ELIP duplication

One-sentence summary: A recurrent gene duplication event is found in all surveyed resurrection plant genomes supporting convergent evolution of this trait.

Author Contributions: R.V. designed and conceived research; R.V., J.P., C.M.W., S.E., and D.B. analyzed data; R.V. wrote the paper. All authors read and approved the final manuscript.

Abstract

Desiccation tolerance was a critical adaptation for the colonization of land by early non-vascular plants. Resurrection plants have maintained or rewired these ancestral protective mechanisms and desiccation-tolerant species are dispersed across the land plant phylogeny. Though common physiological, biochemical, and molecular signatures are observed across resurrection plant lineages, features underlying the recurrent evolution of desiccation tolerance are unknown. Here we used a comparative approach to identify patterns of genome evolution and gene duplication associated with desiccation tolerance. We identified a single gene family with dramatic expansion in all sequenced resurrection plant genomes and no expansion in desiccation-sensitive species. This gene family of early light-induced proteins (ELIPs) expanded in resurrection plants convergent through repeated tandem gene duplication. ELIPS are universally highly expressed during desiccation in all surveyed resurrection plants and may play a role in protecting against

photooxidative damage of the photosynthetic apparatus during prolonged dehydration. Photosynthesis is particularly sensitive to dehydration and the increased abundance of ELIPs may help facilitate the rapid recovery observed for most resurrection plants. Together, these observations support convergent evolution of desiccation tolerance in land plants through tandem gene duplication.

Key words: convergent evolution, desiccation tolerance, land plants, tandem duplication, ELIP

Introduction

The ability to survive near complete anhydrobiosis is ubiquitous and desiccation tolerance is observed in prokaryotes, protists, fungi, plants, and even animals. The origin of desiccation tolerance in plants is unknown, but its prevalence throughout land plants and green algae suggest it is ancestral. Desiccation tolerance was essential during the transition from water to land, and these ancestral protective mechanisms are retained in some bryophytes, lycophytes, and ferns (Proctor, 1990; Oliver et al., 2005; Porembski, 2011). Early vegetative desiccation tolerance mechanisms likely formed the basis of pollen and seed desiccation pathways in angiosperms, as core abscisic acid-regulated pathways have conserved functions in seed and non-seed plants (Eklund et al., 2018). Plants with vegetative desiccation tolerance are broadly termed 'resurrection plants' because of their revival from typically lethal, prolonged dehydration. Vegetative desiccation tolerance in angiosperms arose more recently, and mounting evidence suggests it evolved through rewiring pathways leading to desiccation tolerance in seeds (Oliver et al., 2000; Illing et al., 2005; Farrant and Moore, 2011; Costa et al., 2017; VanBuren et al., 2017). Desiccation tolerance evolved independently in at least 13 lineages of angiosperms and over 300 diverse resurrection plant species have been identified to date across all major plant taxa outside of gymnosperms (Bewley and Krochko, 1982; Oliver et al., 2000; Oliver et al., 2005). This is likely an underestimation, as many desiccation tolerant fern and fern allies are uncharacterized (Porembski, 2011).

The successful induction of desiccation tolerance requires the coordinated expression of complex pathways with thousands of genes, and the deployment of physiological and biochemical safeguards to protect cellular macromolecules and machinery. The photosynthetic machinery is particularly sensitive to dehydration and must be maintained intact or repaired quickly after rehydration (Challabathula et al., 2016). Desiccation-associated responses are well characterized and a core set are conserved across all resurrection plants. These include accumulation of osmoprotectants, reactive oxygen species (ROS) scavengers, and late embryogenesis abundant (LEA) proteins as well as changes in cell structure and other physical properties (Illing et al., 2005; Zhang and Bartels, 2018). These conserved molecular signatures suggest desiccation tolerance evolved convergently through rewiring similar, preexisting pathways across independent lineages (Bewley, 1979; Gaff, 1997). This hypothesis is now

testable given the wealth of genomic resources for desiccation-tolerant species. The genomes of several model resurrection plants have been sequenced including *Boea hygrometrica* (Xiao et al., 2015), *Oropetium thomaeum* (VanBuren et al., 2015), *Xerophyta viscosa* (Costa et al., 2017), *Lindernia brevidens* (VanBuren et al. 2018b), *Selaginella lepidophylla* (VanBuren et al., 2018c), and *Selaginella tamariscina* (Xu et al., 2018). Genomes are also available for the abscisic acid (ABA)-inducible desiccation tolerant bryophytes *Physcomitrella patens* (Rensing et al., 2008) and *Marchantia polymorpha* (Bowman et al., 2017). Despite abundant genomic resources, large-scale comparisons of desiccation-related genes and pathways between these species are limited.

Gene duplications drive innovation and facilitate rapid adaptation to abiotic stresses (Hanada et al., 2008). Tandem gene duplicates in *O. thomaeum* (VanBuren et al., 2017), *S. lepidophylla* (VanBuren et al., 2018c), and *X. viscosa* (Costa et al., 2017) are enriched in functions related to dehydration stress and desiccation. Though similar enrichment patterns are observed across these taxa, overlap of duplicated orthologous genes has not been analyzed between independent resurrection plant lineages. Here, we identified patterns of gene duplication and gene family expansion associated with the evolution of desiccation tolerance. We identified a single gene family uniquely expanded in all sequenced resurrection plants. This gene family expansion is driven primarily by tandem duplication and genes in this family are universally highly expressed in all surveyed resurrection plants. Similar duplication patterns despite ~475 million years of separation (Smith et al., 2010) supports convergent evolution of desiccation tolerance.

Results

Patterns of gene family dynamics in resurrection plant genomes

The evolution of desiccation tolerance may occur through recurrent duplication of shared genes and pathways across independent lineages of resurrection plants. Duplication of these common features could increase the abundance of important osmoprotectants and end-point metabolites, or facilitate high-level pathway rewiring through regulatory neo or subfunctionalization. To test if conserved gene duplication events are contributing to desiccation tolerance, we surveyed changes in gene family composition across diverse desiccation-tolerant and sensitive land-plant genomes. Orthogroups were identified using OrthoFinder (Emms and Kelly, 2015) for nine monocot, six eudicot, two bryophyte and two lycophyte genomes (Supplemental Table S1). Three desiccation-tolerant angiosperms (*Lindernia brevidens*, *Oropetium thomaeum*, and *Xerophyta viscosa*) and two bryophytes with ABA-induced desiccation tolerance (*P. patens* and *M. polymorpha*) were included for analysis. In total, 26,406 orthogroups were identified across the 19 species and 456,776 (78.8%) of the input genes were assigned to orthogroups. Of these, 4,625 orthogroups had representatives from all 19 species. This subset of conserved orthogroups was used to identify gene families that were expanded in all resurrection plant genomes but showed no expansion in desiccation-sensitive plants. Three orthogroups have expanded in all resurrection plants compared to desiccation-sensitive species: OG0212, OG1697, and OG2512. The Arabidopsis (*Arabidopsis thaliana*) orthologs in OG2512 include the glycine-rich domain

protein AtGRDP1, which has a role in seed germination and ABA signaling (Rodríguez-Hernández et al., 2014). The *AtGRDP1* null mutant has higher sensitivity to salt and osmotic stress, and overexpression increases stress tolerance (Rodríguez-Hernández et al., 2014). OG1697 contains orthologs related to the rice (Oryza sativa) domestication gene *SH4*, which is involved in in seed shattering and development of the seed abscission layer (Li et al., 2006). OG0212 contains the early light-induced proteins (ELIPs), which have a well-defined role in protection against photooxidative damage under high light conditions (Hutin et al., 2003). Together, this supports gene duplication-mediated rewiring of seed-related processes in resurrection plants, consistent with previous hypotheses (Bewley, 1979; Illing et al., 2005; Costa et al., 2017; VanBuren et al., 2017). *P. patens* and *M. polymorpha* require exogenous ABA to induce desiccation tolerance and are not classified as true resurrection plants. Excluding these two lineages from enrichment tests identified a single expanded orthogroup (OG0212; ELIPs) common to all surveyed resurrection plants. The number of ELIPs in this orthogroup ranged from 10-26 for resurrection plants and 1-8 for desiccation-sensitive lineages. Gene family dynamics, evolutionary history, and expression patterns of ELIPs were further characterized across land plants and green algae.

ELIP gene family dynamics across land plants and green algae

We expanded the analysis of ELIP composition to seventy-five sequenced land plant and Chlorophyta genomes to test if the dramatic expansion we observed is universally unique to desiccation-tolerant species. ELIPs were identified using a homology-based approach with BLAST and classified as singletons or tandem duplicates based on their physical proximity in the genomes (see methods). All desiccation-sensitive land plants have less than ten ELIPs with an average of 3.1 per genome, compared to an average of 20.7 in desiccation-tolerant species (Figure 1). Desiccation-sensitive (DS) monocots tend to have more ELIPs than DS eudicots (avg. 3.7 vs 3.0) but the difference is not significant (Wilcoxon rank-sum, $P = 0.09$). Desiccation-sensitive grasses (family Poaceae) have an enrichment of ELIPs compared to all angiosperms, with an average of 5.0 per genome (Wilcoxon rank-sum, $P < 0.05$). The only Charophyte genome used in this study (*Chara braunii*) has one ELIP. Chlorophyta have considerably more ELIPs than DS land plants (avg. 7.5) but fewer than resurrection plants. The high copy number of ELIPs may help combat rapid changes in light intensity and the increased ultraviolet radiation-mediated photobleaching that green algae experience in dynamic aquatic or exposed environments.

Expansion of ELIPs was previously reported in the genomes of *Boea hydrometrica* (Xiao et al., 2015), *Selaginella lepidophylla* (VanBuren et al., 2018c), and *Lindernia brevidens* (VanBuren et al. 2018b). No such enrichment of ELIPs was reported in the resurrection plants *Oropetium thomaeum* (VanBuren et al., 2015), *Xerophyta viscosa* (Costa et al., 2017), or *Selaginella tamariscina* (Xu et al., 2018), suggesting a role for ELIPs in desiccation tolerance is not universal. The *O. thomaeum* genome was recently updated and reassembled, and the *O. thomaeum* V2 assembly has 22 newly annotated ELIPs (VanBuren et al., 2018a). Only one ELIP was reported in *Selaginella tamariscina* (Xu et al., 2018), but BLAST against the genome

identified 74 complete or partial unannotated ELIPs (Supplemental Table S2). Because these ELIPs were not annotated in the *S. tamariscina* genome, this species was excluded from the expression analysis. Numerous 'clusters of desiccation-associated genes' (CoDAGs) were described in the *X. viscosa* genome (Costa et al., 2017), but the ten annotated ELIPs were not classified as CoDAGs. This is surprising given the tandemly duplicated nature of the ELIPs, and their high expression during desiccation in *X. viscosa*.

Gene duplications can arise in mass during whole-genome duplication events, or at the single-gene level through tandem or retrotransposon-mediated duplication. We further characterized the origin of ELIPs to identify the mechanism driving the large-scale duplication observed in resurrection plants. Tandemly duplicated ELIPs were identified using a syntenic approach based on their physical proximity within the genome. Most ELIPS in desiccation-sensitive species are dispersed randomly across the genome as singletons (Figure 1). All ELIPs from plants within the Brassicales are singletons, and species from other angiosperm orders have a mix of singleton and tandem gene copies. In contrast, most ELIPs in resurrection plants (74%; 111 out of 149) are found in large tandem arrays. All ELIPs in *O. thomaeum* and *X. viscosa* are tandemly duplicated and other resurrection plants have a mix of singletons and tandem duplicates. *B. hydrometrica* has the lowest number of tandemly duplicated ELIPs among resurrection plants, but this may be an artifact of the fragmented nature of the genome assembly (Xiao et al., 2015). Desiccation-induced accumulation of ELIP transcripts was also observed in the model moss *Syntrichia* (Zeng et al., 2002; Oliver et al., 2004), but the number of ELIPs in the genome is unknown.

O. thomaeum has the largest tandem array of ELIPs, with 22 copies on chromosome 8 (Figure 2). Despite a high degree of gene-level collinearity between *O. thomaeum* and the C4 panicoid grasses sorghum (*Sorghum bicolor*) and Setaria (*Setaria italica*), ELIPs from the *O. thomaeum* tandem array have no syntenic orthologs (Figure 2a). The ELIP array in *O. thomaeum* is divided into two segments of eleven copies split by a 60 kb region spanning nine genes (Figure 2a and 2b). *O. thomaeum* genes that flank and bisect the ELIP tandem array are collinear with Sorghum and Setaria. Several *O. thomaeum* ELIPs contain fragments of, or are flanked by intact retrotransposons (Figure 2b). These closely associated Ty3-gypsy retrotransposons may have facilitated repeated gene duplication.

Expression of *ELIP genes during desiccation and rehydration*

The repeated duplication of ELIPs across resurrection plants suggests a conserved role in desiccation tolerance. To test for desiccation-related expression, we re-analyzed available RNAseq data from *O. thomaeum* (VanBuren et al., 2017), *B. hydrometrica* (Xiao et al., 2015), *S. lepidophylla* (VanBuren et al., 2018c), *L. brevidens* (VanBuren et al. 2018b), *X. viscosa* (Costa et al., 2017), and the moss *B. argenteum* (Gao et al., 2015). Expression data was processed using the same pipeline for each species, to more accurately compare patterns across experiments. A genome for *B. argenteum* is not available, so a set of representative transcripts was assembled

using the RNAseq data with Trinity (Haas et al., 2013). Eight ELIP transcripts were identified in *B. argenteum*, but this may not represent the total number of ELIPs in the genome.

ELIPs are consistently among the highest expressed genes in resurrection plants during varying dehydration and rehydration timecourses (Figure 3). Each of the surveyed dehydration and rehydration timecourses were sampled at different relative water contents, different length and severity of water deficit, and different times post rehydration. Because of this, we broadly classified timepoints into three groups to facilitate cross-species comparisons: well-watered (green), dehydrating/desiccating (yellow), and rehydration (blue). ELIPs have low or undetectable expression under well-watered conditions in all surveyed tolerant and sensitive species. ELIPs have the highest expression in desiccated tissues, and expression increases throughout the progression of dehydration stress. The high expression of ELIPs is generally maintained during early rehydration (>24 hours), with expression decreasing as plants return to normal relative water content. ELIP downregulation during rehydration is reflected by the rate of recovery. *B. argenteum* is able to recover in > 1 hour post rehydration, and ELIP expression returns to basal levels in less than 24 hours. Recovery in *X. viscosa* and *L. brevidens* occurs more slowly, and ELIPs are highly expressed during all of the sampled rehydration timepoints.

Drought-induced expression of ELIPs is not unique to resurrection plants and ELIPs are upregulated under water deficit in desiccation-sensitive species (Figure 4). Tandem duplication in resurrection plants may increase the absolute transcript abundance of ELIPs and improve photoprotective capacity. We surveyed the combined expression of ELIPs under the most desiccated timepoint for each species and compared this to drought samples from Arabidopsis (9 days drought, RWC 60%) (Crisp et al., 2017), maize (*Zea mays*, soil water content 40%), and *L. subracemosa*. ELIPs have, on average, 622-fold higher expression abundance in resurrection plants than desiccation-sensitive species under water deficit (Figure 4; Wilcoxon ran sum p=0.023). *O. thomaeum* and *L. brevidens* have the highest absolute expression of ELIPs and *X. viscosa* has the lowest expression among resurrection plants.

Discussion

Resurrection plants span all major taxa outside of gymnosperms and independent desiccation-tolerant lineages diverged up to 475 MYA (Smith et al., 2010). Despite this ancient divergence, all resurrection plants seem to utilize a set of conserved molecular mechanisms to combat the damages associated with desiccation. This phenotypic convergence could arise through independent co-option of selection-constrained pathways or through rewiring of non-overlapping (independent) pathways. Based on genome-scale comparisons of desiccation-tolerant and sensitive species, we identified a single gene family that has expanded in all sequenced resurrection plant genomes. Given the wide taxonomic sampling, we can infer expansion of the ELIP gene family occurred independently in each desiccation-tolerant lineage. The rapid and dramatic expansion arose through repeated tandem gene duplication, and ELIPs are found in large arrays ranging in size from 10 to 22 copies. Desiccation-sensitive species typically have 1-5 ELIPs with high collinearity across plant genomes. The ELIP tandem array in *O. thomaeum*

has recently been translocated to a different chromosome, and is non-syntenic with orthologous ELIPs in other grass genomes. Several ELIPs in *O. thomaeum* are nested around intact or fragmented retrotransposons, providing a source for their duplication and translocation. Taken together, repeated ELIP duplication supports convergent evolution of desiccation tolerance across land plants.

Desiccation induces tremendous stress on the macromolecules and molecular machinery of the cell. Photosynthesis is inhibited in desiccated tissues, and homoiochlorophyllous (chlorophyll retaining) resurrection plants are susceptible to increased photoinhibition and oxidative damage from unbound chlorophyll or unstable light harvesting complexes (LHC). ELIPs are induced under high light stress and protect against photooxidative damage via chlorophyll binding and stabilization of photosynthetic complexes (Hutin et al., 2003). ELIPs accumulate in photosynthetic tissue under other abiotic stresses including cold, drought, heat, and low nutrients (Bartels et al., 1992; Beator et al., 1992; Levy et al., 1993; Adamska and Kloppstech, 1994; Montané et al., 1997), though their role in some of these stresses is unknown. ELIPs are also induced under other conditions where chlorophyll is degraded such as chromoplast biogenesis in tomato (*Solanum lycopersicum*), suggesting a broader role of photoprotection during thylakoid restructuring (Bruno and Wetzel, 2004). Overexpression of a *Medicago truncatula* ELIP in*Nicotiana benthamiana* increased resistance to freezing, chilling, osmotic stress, and high light (Araújo et al. 2013).

ELIPs were among the earliest desiccation-related proteins identified, and ELIPs have a conserved role during desiccation in *Craterostigma plantagineum* (Bartels et al., 1992; Alamillo and Bartels, 2001) and *Syntrichia ruralis* (Zeng et al., 2002). ELIPs are also highly expressed during desiccation in *Haberlea rhodensis* (Gechev et al., 2013) and *Sporobolus stapfianus* (Yobi et al., 2017), but these lineages lack reference genomes, so they were not included in this study. The two ELIPs in Arabidopsis are functionally redundant, and overexpression of either is sufficient to rescue the photosensitivity in the pleiotropic *chaos* mutant (Hutin et al., 2003). ELIPs in resurrection plants likely maintain their ancestral photoprotective role, but the repeated tandem duplication increases their relative abundance. ELIPs are among the most highly expressed genes in desiccating leaf tissues and the proteins accumulate in the thylakoid of desiccating *C. plantagineum* (Bartels et al., 1992; Alamillo and Bartels, 2001). ELIP accumulation in *C. plantagineum* was hypothesized to bind free chlorophyll and stabilize the thylakoid to reduce photooxidative damage during prolonged desiccation. Outside of this single study in *C. plantagineum*, ELIP abundance and localization has not been surveyed in resurrection plants, and the link between duplication, transcript accumulation, and protein production remains to be tested. Overexpression of a *C. plantagineum* ELIP in *Medicago truncatula* increased drought tolerance and recovery, supporting a potential role in desiccation tolerance (Araújo et al., 2013).

Excess light is a pervasive challenge in desiccated leaf tissue. Poikilochlorophyllous resurrection plants dismantle their photosynthetic apparatus during desiccation to mitigate light-associated damage. In contrast, homoiochlorophyllous resurrection plants retain thylakoids and chlorophyll during desiccation, but recover more quickly upon rehydration. Photosystem II has

negligible activity in desiccated tissues of *C. pumilum*, but near normal maximum quantum yield of PSII (Fv/Fm) is observed only 18 hours post rehydration (Charuvi et al., 2015). Recovery in mosses is even faster and Fv/Fm reaches two-thirds the well-watered ratio in less than 40 minutes post rehydration (Proctor and Smirnoff, 2000). Homiochlorophyllous resurrection plants have considerably more ELIPs (avg. 21) than poikilochlorophyllous species and this is likely reflective of the increased photooxidative damage resulting from maintaining the photosynthetic apparatus during desiccation. The increased relative abundance of ELIPs in the thylakoid of homiochlorophyllous resurrection plants may protect and stabilize PSII complexes and facilitate more rapid neutralization and degradation of unbound chlorophyll.

X. viscosa has the lowest number of ELIPs (10) among resurrection plants, and it is the only poikilochlorophyllous species with a sequenced genome. The comparatively low number of ELIPs may reflect the alternative strategy of dismantling photosynthetic machinery then degrading chlorophyll to reduce photooxidative damage. High expression of ELIPs in *X. viscosa* is maintained throughout rehydration. This species takes several days to fully recover and to resynthesize chlorophyll, therefore ELIPs may have a protective role during chlorophyll synthesis, similar to greening seedlings. Sequences of additional resurrection plant genomes are needed to further test this link between ELIP copy number and strategies for mitigating damage from excess light.

ELIP duplication supports convergent evolution. ELIPs, together with other abundantly expressed protective proteins such as LEA proteins, are likely central components of the desiccation response. Desiccation pathways typically associated with seeds are induced during vegetative desiccation, suggesting this trait evolved through pathway rewiring (Illing et al., 2005; Farrant and Moore, 2011; Costa et al., 2017; VanBuren et al., 2017). Transcription factors are under strict dosage constraints, and no conserved duplications of genes encoding seed-related transcription factors such as ABI3 and ABI5 (Lopez-Molina et al., 2001) were identified in this study. This suggests either independent lineages duplicated and neofunctionalized different genes to induce seed-associated desiccation pathways, or existing genes were rewired via *cis*-regulatory elements as previously shown (Giarola et al., 2018). This remains to be tested but will be possible with additional genomes and genome-scale analysis of *cis*-elements and transcription factor binding sites. Taken together, we hypothesize rewired-desiccation pathways enabled plants to withstand prolonged drying and the duplication of ELIPs allowed plants to withstand excessive light in the absence of water.

Methods

Orthogroup identification and enrichment patterns

Orthologous genes across a subset of 19 land plant species were analyzed to identify gene families that are expanded in desiccation tolerant lineages. The gene family analysis included

nine monocot (*Ananas comosus, Brachypodium distachyon, Oryza sativa, Oropetium thomaeum, Panicum virgatum, Sorghum bicolor, Setaria italic, Xerophyta viscosa, Zostera marina*), six eudicot (*Lindernia brevidens, L. subracemosa, Arabidopsis thaliana, Medicago truncatula, Solanum lycopersicum, Vitis vinifera*), two bryophyte (*Marchantia polymorpha* and *Physcomitrella patens*) and two lycophyte genomes (*Selaginella lepidophylla* and *S. moellendorffii*). This includes three desiccation-tolerant angiosperms and two bryophytes with ABA-induced desiccation tolerance. The predicted protein sequences for each species were first clustered into orthologous groups using Orthofinder (v2.2.6) (Emms and Kelly, 2015). Diamond (v0.9.24) (Buchfink, et al. 2014). was used to conduct pairwise protein alignments and all other parameters were set to defaults. The resulting 26,406 orthogroups were filtered to include only the 4,625 groups with at least one ortholog in all 19 species. For each orthogroup, the following values were calculated: the proportion of genes in that orthogroup among the 7 desiccation tolerant plants, the proportion of genes in that orthogroup among the 14 desiccation sensitive plants, and the combined proportion using the following formulas:

$$p_{dt} = \frac{\sum_{dt\ spesies\ i}^{dt\ species\ 1} \#\ genes\ in\ orthogroup}{\sum_{dt\ species\ i}^{dt\ species\ 1} total\ \#\ of\ genes\ \ in\ any\ orthogroup}$$

$$p_{ds} = \frac{\sum_{ds\ species\ i}^{ds\ species\ 1} \#\ genes\ in\ orthogroup}{\sum_{ds\ species\ i}^{ds\ species\ 1} total\ \#\ of\ genes\ \ in\ any\ orthogroup}$$

$$p = \frac{\sum_{species\ i}^{species\ 1} \#\ genes\ in\ orthogroup}{\sum_{species\ i}^{species\ 1} total\ \#\ of\ genes\ in\ any\ orthogroup}$$

The two proportions were then compared using a z-score calculated as follows:

$$Z = \frac{(p_{dt} - p_{ds})}{\sqrt{p(1-p)\left(\frac{1}{n_{dt}} + \frac{1}{n_{ds}}\right)}}$$

A 1-sided hypothesis test was calculated by comparing the Z-score to a normal distribution. The resulting p-value was then adjusted using the Benjamini & Hochberg procedure (Benjamini and Hochberg, 1995) to obtain q-values. Orthogroups with a q-value of less than 0.05 were considered to be significantly enriched.

Identification of ELIPs

Based on orthogroup enrichment, the analysis of ELIPs was extended to 72 land plant and green algae genomes downloaded from Phytozome (V12; https://phytozome.jgi.doe.gov/pz/portal.html). ELIPs were also annotated in the Charophyte *Chara braunii* (Nishiyama et al., 2018). ELIPs were identified using BLASTp with the Arabidopsis ELIP1 (AT3G22840.1) as a query and an e-value cutoff of 1e-15 for land plants and 1e-6 for green algae. ELIPs were further classified as singleton or tandemly duplicated based on their physical proximity in the genome (< 5 genes separating ELIPs).

Expression analysis

Illumina RNAseq data from *O. thomaeum* (VanBuren et al., 2017), *Boea hydrometrica*(Xiao et al., 2015), *Selaginella lepidophylla* (VanBuren et al., 2018c), *Lindernia brevidens* (VanBuren et al., 2018b), *Lindernia subracemosa* (VanBuren et al., 2018b), *Xerophyta viscosa* (Costa et al., 2017), Arabidopsis (Crisp et al., 2017), maize (PRJNA419326), and the moss *Bryum argenteum* (Gao et al., 2015) was downloaded from the NCBI sequence read archive (SRA) and reanalyzed. Raw Illumina reads were trimmed using TRIMMOMATIC (v0.33)(Bolger et al., 2014) with default parameters to remove adapters and low quality bases. A summary of the RNAseq data used in this study along with the relative water content and designation for each sample can be found in Supplemental Table S3. A set of representative transcripts for *B. argenteum* was assembled using Trinity (Haas et al., 2013) with the available RNAseq data (Gao et al., 2015). Expression levels were quantified using Kallisto (v0.44.0)(Bray et al., 2016) against the respective set of gene models for each species and the Trinity-based transcripts for *B. argenteum.* Parameters for Kallisto were left as default and 100 bootstraps were run per sample. Expression was quantified in Transcript Per Million (TPM), and a mean across all of the replicates was used to compare expression of ELIPs in each species. Log2 transformed TPM expression values of ELIPs were plotted. Samples were clustered into three groups (well-watered, desiccating, and rehydrating) to compare expression between species.

Grass comparative genomics

Syntenic gene pairs between the *Oropetium thomaeum* (VanBuren et al., 2015; VanBuren et al., 2018a)*, Setaria italica* (Zhang et al., 2012), and *Sorghum bicolor* (Paterson et al., 2009) genomes were identified using the python version of MCSCAN toolkit (V1.1) (Wang et al., 2012). Gene models from the three genomes were aligned using LAST and hits were filtered using default parameters to find the best syntenic blocks. A microsyntenic dotplot of the region containing the ELIP tandem array in Oropetium was constructed using MCScan.

Accession Numbers

Genomes and gene models were downloaded from Phytozome (V12; https://phytozome.jgi.doe.gov/pz/portal.html. Illumina RNAseq data was downloaded from the

NCBI SRA under the following BioProjects: *Oropetium thomaeum* (PRJNA286116) *Boea hydrometrica* (PRJNA277046), *Selaginella lepidophylla* (PRJNA420971), *Lindernia brevidens* (PRJNA488068), *Lindernia subracemosa* (PRJNA488068), *Xerophyta viscosa* (PRJNA295811), Arabidopsis (PRJNA391262), maize (PRJNA419326), and the moss *Bryum argenteum* (PRJNA272646),

Supplemental Data

Supplemental Table S1. Summary of orthologous gene family clustering

Supplemental Table S2. Annotation of ELIPs in the *Selaginella tamariscina* genome.

Supplemental Table S3. Summary of expression and physiology data used in this study.

Figure Legends

Figure 1. **ELIP composition in sequenced land plant and Chlorophyta genomes.** The number of ELIPs are plotted for 72 genomes. Tandemly duplicated ELIPs are plotted in red and single copy or interspersed ELIPs are plotted in blue. Desiccation-tolerant species are highlighted in orange.

Figure 2. **Unique tandem array of ELIPs in *O. thomaeum*.** (a) Microsynteny plot of the syntenic region flanking the ELIP array in *O. thomaeum*, Sorghum, and Setaria. Genes in the forward orientation are shown in yellow and blue genes are in reverse orientation. Syntenic orthologs between the three genomes are denoted by connecting gray lines. The region containing ELIPs is highlighted in brown. (b) Expanded view of the ELIP tandem array in *O. thomaeum*. ELIPs are shown in blue and other genes are shown in gray. Long terminal repeat (LTR) retrotransposons are shown in red.

Figure 3. **Desiccation related expression of ELIPs across divergent resurrection plants.** Log2 transformed expression of ELIPs is plotted using RNAseq data from six resurrection plants. Colors in violin plots correspond to hydration state: green are well-watered, yellow are dehydrating/desiccated, and blue are rehydrating. Stages are defined in Supplemental Table S3.

Figure 4. **Total expression of ELIPs in desiccation tolerant and sensitive species.** The Log2 transformed, summed TPM of ELIPs from each species are plotted for drought/desiccation timepoints (yellow) and comparable well-watered or rehydrated timepoints (green and blue respectively).

Acknowledgements: This work is supported in part by funding from the National Science Foundation (NSF-MCB 1817347to R.V.).

Competing Financial Interests: The authors declare no competing financial interests.

References:

Adamska, I., and Kloppstech, K. (1994). Low temperature increases the abundance of early light-inducible transcript under light stress conditions. Journal of Biological Chemistry **269,** 30221-30226.

Alamillo, J.M., and Bartels, D. (2001). Effects of desiccation on photosynthesis pigments and the ELIP-like dsp 22 protein complexes in the resurrection plant *Craterostigma plantagineum*. Plant Science **160,** 1161-1170.

Araújo, S., Duque, A., Silva, J., Santos, D., Silva, A., and Fevereiro, P. (2013). Water deficit and recovery response of *Medicago truncatula* plants expressing the ELIP-like DSP22. Biologia plantarum **57,** 159-163.

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.

Beator, J., Pötter, E., and Kloppstech, K. (1992). The effect of heat shock on morphogenesis in barley: Coordinated circadian regulation of mRNA levels for light-regulated genes and of the capacity for accumulation of chlorophyll protein complexes. Plant physiology **100,** 1780-1786.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society. Series B (Methodological)**,** **57**, 289-300.

Bewley, J., and Krochko, J. (1982). Desiccation-tolerance. In Physiological plant ecology II (Springer), pp. 325-378.

Bewley, J.D. (1979). Physiological aspects of desiccation tolerance. Annual Review of Plant Physiology **30,** 195-238.

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

Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., and Berger, F. et al. (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. Cell **171,** 287-304. e215.

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

Bruno, A.K., and Wetzel, C.M. (2004). The early light-inducible protein (ELIP) gene is expressed during the chloroplast-to-chromoplast transition in ripening tomato fruit. Journal of Experimental Botany **55,** 2541-2548.

Buchfink, B., Xie, C., and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. Nature Methods, **12**, 59-60.

Challabathula, D., Puthur, J.T., and Bartels, D. (2016). Surviving metabolic arrest: photosynthesis during desiccation and rehydration in resurrection plants. Annals of the New York Academy of Sciences **1365,** 89-99.

Charuvi, D., Nevo, R., Shimoni, E., Naveh, L., Zia, A., Adam, Z., Farrant, J.M., Kirchhoff, H., and Reich, Z. (2015). Photoprotection conferred by changes in photosynthetic protein levels and organization during dehydration of a homoiochlorophyllous resurrection plant. PlantPphysiology**,** pp. 114.255794.

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.

Crisp, P.A., Ganguly, D., Smith, A.B., Murray, K.D., Estavillo, G.M., Searle, I.R., Ford, E., Bogdanović, O., Lister, R., and Borevitz, J.O. et al. (2017). Rapid recovery gene downregulation during excess-light stress and recovery in Arabidopsis. The Plant Cell**, 29,** 1836-1863.

Eklund, D.M., Kanei, M., Flores-Sandoval, E., Ishizaki, K., Nishihama, R., Kohchi, T., Lagercrantz, U., Bhalerao, R.P., Sakata, Y., and Bowman, J.L. (2018). An Evolutionarily Conserved Abscisic Acid Signaling Pathway Regulates Dormancy in the Liverwort *Marchantia polymorpha*. Current Biology.

Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology **16,** 157.

Farrant, J.M., and Moore, J.P. (2011). Programming desiccation-tolerance: from plants to seeds to resurrection plants. Current opinion in plant biology **14,** 340-345.

Gaff, D. (1997). Mechanisms of desiccation tolerance in resurrection vascular plants. In Mechanisms of environmental stress resistance in plants**.** Harwood Academic Publishers (Amsterdam, Netherlands) 43-58.

Gao, B., Zhang, D., Li, X., Yang, H., Zhang, Y., and Wood, A.J. (2015). De novo transcriptome characterization and gene expression profiling of the desiccation tolerant moss *Bryum argenteum* following rehydration. BMC Genomics **16,** 416.

Gechev, T.S., Benina, M., Obata, T., Tohge, T., Sujeeth, N., Minkov, I., Hille, J., Temanni, M.-R., Marriott, A.S., and Bergström, E. et al. (2013). Molecular mechanisms of desiccation tolerance in the resurrection glacial relic *Haberlea rhodopensis*. Cellular and Molecular Life Sciences **70,** 689-709.

Giarola, V., Jung, N.U., 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**, 69,** 3773-3784.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., and Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols **8,** 1494.

Hanada, K., Zou, C., Lehti-Shiu, M.D., Shinozaki, K., and Shiu, S.-H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiology **148,** 993-1003.

Hutin, C., Nussaume, L., Moise, N., Moya, I., Kloppstech, K., and Havaux, M. (2003). Early light-induced proteins protect Arabidopsis from photooxidative stress. Proceedings of the National Academy of Sciences **100,** 4921-4926.

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.

Levy, H., Tal, T., Shaish, A., and Zamir, A. (1993). Cbr, an algal homolog of plant early light-induced proteins, is a putative zeaxanthin binding protein. Journal of Biological Chemistry **268,** 20892-20896.

Li, C., Zhou, A., and Sang, T. (2006). Rice domestication by reducing shattering. Science **311,** 1936-1939.

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.

Montané, M.-H., Dreyer, S., Triantaphylides, C., and Kloppstech, K. (1997). Early light-inducible proteins during long-term acclimation of barley to photooxidative stress caused by light and cold: high level of accumulation by posttranscriptional regulation. Planta **202,** 293-302.

Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K., Haas, F.B., Vanderstraeten, L., Becker, D., and Lang, D. (2018). The chara genome: Secondary complexity and implications for plant terrestrialization. Cell **174,** 448-464. e424.

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., Velten, J., and Mishler, B.D. (2005). Desiccation tolerance in bryophytes: a reflection of the primitive strategy for plant survival in dehydrating habitats? Integrative and Comparative Biology **45,** 788-799.

Oliver, M.J., Dowd, S.E., Zaragoza, J., Mauget, S.A., and Payton, P.R. (2004). The rehydration transcriptome of the desiccation-tolerant bryophyte *Tortula ruralis*: transcript classification and analysis. BMC Genomics **5,** 89.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., and Poliakov, A. (2009). The *Sorghum bicolor* genome and the diversification of grasses. Nature **457,** 551.

Porembski, S. (2011). Evolution, diversity, and habitats of poikilohydrous vascular plants. In Plant desiccation tolerance (Springer), pp. 139-156.

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

Proctor, M.C., and Smirnoff, N. (2000). Rapid recovery of photosystems on rewetting desiccation-tolerant mosses: chlorophyll fluorescence and inhibitor experiments. Journal of Experimental Botany **51,** 1695-1704.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.-F., Lindquist, E.A., and Kamisugi, Y. et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science **319,** 64-69.

Rodríguez-Hernández, A.A., Ortega-Amaro, M.A., Delgado-Sánchez, P., Salinas, J., and Jiménez-Bremont, J.F. (2014). AtGRDP1 Gene encoding a glycine-rich domain protein is involved in germination and responds to ABA signalling. Plant Molecular Biology Reporter **32,** 1187-1202.

Smith, S.A., Beaulieu, J.M., and Donoghue, M.J. (2010). An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. Proceedings of the National Academy of Sciences**,** **107**, 5897-5902.

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**, 96.

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. The Plant Cell. tpc.00517.2018

VanBuren, R., Wai, J., Zhang, Q., Song, X., Edger, P.P., Bryant, D., Michael, T.P., Mockler, T.C., and Bartels, D. (2017). Seed desiccation mechanisms co-opted for vegetative desiccation in the resurrection grass *Oropetium thomeaum*. Plant, Cell & Environment, **40,** 2292-2306.

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*. 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.

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

Xiao, L., Yang, G., Zhang, L., Yang, X., Zhao, S., Ji, Z., Zhou, Q., Hu, M., Wang, Y., and Chen, M. et al. (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., Xin, T., Bartels, D., Li, Y., Gu, W., Yao, H., Liu, S., Yu, H., Pu, X., and Zhou, J. et al. (2018). Genome analysis of the ancient tracheophyte *Selaginella tamariscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. Molecular Plant. **11,** 983-994.

Yobi, A., Schlauch, K.A., Tillett, R.L., Yim, W.C., Espinoza, C., Wone, B.W., Cushman, J.C., and Oliver, M.J. (2017). *Sporobolus stapfianus*: Insights into desiccation tolerance in the resurrection grasses from linking transcriptomics to metabolomics. BMC Plant Biology **17,** 67.

Zeng, Q., Chen, X., and Wood, A.J. (2002). Two early light-inducible protein (ELIP) cDNAs from the resurrection plant *Tortula ruralis* are differentially expressed in response to desiccation, rehydration, salinity, and high light. Journal of Experimental Botany **53,** 1197-1205.

Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., Xie, M., Zeng, P., Yue, Z., and Wang, W. (2012). Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nature Biotechnology **30,** 549.

Zhang, Q., and Bartels, D. (2018). Molecular responses to dehydration and desiccation in desiccation-tolerant angiosperm plants. Journal of Experimental Botany **69,** 3211-3222.