The resurrection genome of *Boea hygrometrica*: A blueprint for survival of dehydration

Lihong Xiao[a], Ge Yang[a], Liechi Zhang[a], Xinhua Yang[b], Shuang Zhao[b], Zhongzhong Ji[a], Qing Zhou[b], Min Hu[b], Yu Wang[a], Ming Chen[b], Yu Xu[a], Haijing Jin[a], Xuan Xiao[a], Guipeng Hu[a], Fang Bao[a], Yong Hu[a], Ping Wan[a], Legong Li[a], Xin Deng[c], Tingyun Kuang[d], Chengbin Xiang[e], Jian-Kang Zhu[f,g,1], Melvin J. Oliver[h,1], and Yikun He[a,1]

[a]School of Life Sciences, Capital Normal University, Beijing 100048, China; [b]Beijing Genomics Institute-Shenzhen, Shenzhen 518083, China; [c]Key Laboratory of Plant Resources and [d]Key Laboratory of Photobiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China; [e]School of Life Sciences, University of Science and Technology of China, Hefei 230022, China; [f]Shanghai Center for Plant Stress Biology, Chinese Academy of Sciences, Shanghai 200032, China; [g]Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47907; and [h]Plant Genetics Research Unit, Midwest Area, Agricultural Research Service, United State Department of Agriculture, University of Missouri, Columbia, MO 65211

Contributed by Jian-Kang Zhu, March 26, 2015 (sent for review February 10, 2015; reviewed by Sagadevan G. Mundree and Andrew J. Wood)

"Drying without dying" is an essential trait in land plant evolution. Unraveling how a unique group of angiosperms, the Resurrection Plants, survive desiccation of their leaves and roots has been hampered by the lack of a foundational genome perspective. Here we report the ∼1,691-Mb sequenced genome of *Boea hygrometrica*, an important resurrection plant model. The sequence revealed evidence for two historical genome-wide duplication events, a compliment of 49,374 protein-coding genes, 29.15% of which are unique (orphan) to *Boea* and 20% of which (9,888) significantly respond to desiccation at the transcript level. Expansion of early light-inducible protein (ELIP) and 5S *rRNA* genes highlights the importance of the protection of the photosynthetic apparatus during drying and the rapid resumption of protein synthesis in the resurrection capability of *Boea*. Transcriptome analysis reveals extensive alternative splicing of transcripts and a focus on cellular protection strategies. The lack of desiccation tolerance-specific genome organizational features suggests the resurrection phenotype evolved mainly by an alteration in the control of dehydration response genes.

vegetative desiccation tolerance | resurrection plant | *Boea hygrometrica* | drought tolerance enhancement | genome

Resurrection plants constitute a unique cadre within the angiosperms: they alone have the remarkable capability to survive the complete dehydration of their leaves and roots. How the dry and visually "dead" plants come alive when water becomes available has long fascinated plant biologists and the lay public alike. The majority of plants, including all our crops, can rarely survive tissue water potentials of less than −4 Mpa. Resurrection plants can, in contrast, survive tissue water potentials of −100 MPa (equilibration to air of 50% relative humidity) and below. The ability to desiccate and resurrect vegetative tissues is considered a primal strategy for surviving extensive periods of drought (1). Desiccation tolerance (DT) has played a major role in plant evolution (1): Postulated as critical for the colonization of terrestrial habitats. DT, as it relates to seed survival and storage, is also arguably the primary plant trait that governs global agriculture and food security. Vegetative DT was lost early in the evolution of tracheophytes (1) and is rare in the angiosperms, but has since reappeared within several lineages, at least 13 of which belong to the angiosperms (2).

Vegetative DT is a complex multigenic and multifactorial phenotype (3–5), but understanding how DT plants respond to and survive dehydration has great significance for plant biology and, more directly, for agriculture. Resurrection plants offer a potential source of genes for improvement of crop drought tolerance (5, 6) as the demand for fresh water grows (7).

In recent decades, efforts have been focused on exploring the structural, physiologic, and molecular aspects of DT in a number of plant species (4). Although a functional genomic approach has been fruitful in revealing the intricacies of DT in resurrection plants (5, 8), and a system approach is contemplated (4), efforts are hampered by the lack of a sequenced genome for any of the resurrection plants. To fill this critical gap, we sequenced the genome of one of the important DT models (9), *Boea hygrometrica*.

B. hygrometrica is a homiochlorophyllous dicot in Gesneriaceae that grows in rocky areas throughout most of China (10). Not only is the whole plant DT (Fig. 1*A*), but a detached leaf or leaf segment retains the DT phenotype and can regenerate a new "seedling" even after several dehydration and rehydration cycles (Fig. 1*B* and *SI Appendix*, Fig. S1 *A* and *B*) (11). Drying leaf tissues exhibit classical dehydration-associated structural changes (12), including a folded cell wall and condensed cytoplasm (*SI Appendix*, Fig. S1 *C–E*).

Here we present a high-quality draft genome of *B. hygrometrica*, along with a full assessment of the changes in the leaf transcriptomes that occur during desiccation and that relate to the resurrection phenotype.

Results

Whole-Genome Features. The whole-genome shotgun sequenced draft genome of *B. hygrometrica* delivers a ∼1,548-Mb assembly,

Significance

The genome analysis presented here represents a major step forward in the field of desiccation tolerance and a much-anticipated resource that will have a far-reaching effect in many areas of plant biology and agriculture. We present the ∼1.69-Gb draft genome of *Boea hygrometrica*, an important plant model for understanding responses to dehydration. To our knowledge, this is the first genome sequence of a desiccation-tolerant extremophile, offering insight into the evolution of this important trait and a first look, to our knowledge, into the genome organization of desiccation tolerance. The underpinning genome architecture and response in relation to the hydration state of the plant and its role in the preservation of cellular integrity has important implications for developing drought tolerance improvement strategies for our crops.

Author contributions: L.X., T.K., and Y. He designed research; L.X., G.Y., L.Z., and Z.J. performed research; Y.W., Y.X., H.J., X.X., G.H., F.B., Y. Hu, L.L., X.D., and C.X. contributed new reagents/analytic tools; L.X., X.Y., S.Z., Q.Z., M.H., M.C., P.W., J.-K.Z., M.J.O., and Y. He analyzed data; and L.X., J.-K.Z., M.J.O., and Y. He wrote the paper.

Reviewers: S.G.M., Queensland University of Technology; and A.J.W., Southern Illinois University.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE48671), and the BioSample database, www.ncbi.nlm.nih.gov/biosample (accession no. SAMN02215335).

[1]To whom correspondence may be addressed. Email: jkzhu@purdue.edu, Mel.Oliver@ars.usda.gov, or yhe@cnu.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1505811112/-/DCSupplemental.

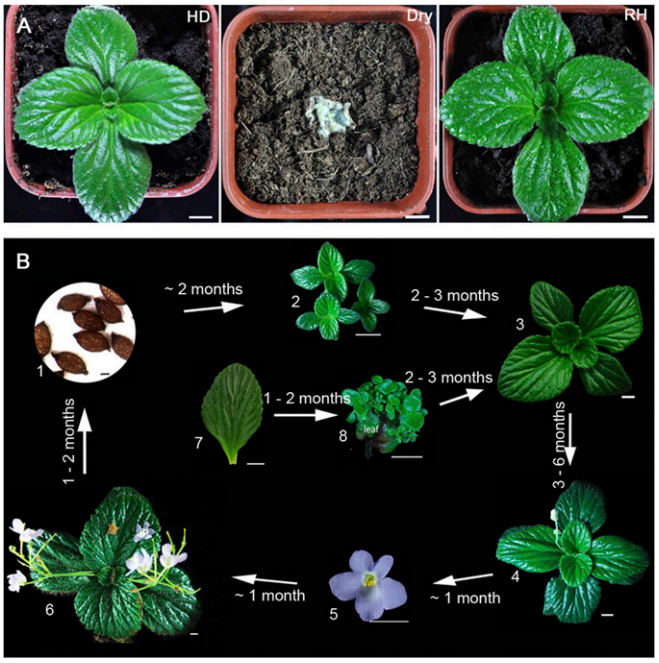

Fig. 1. Phenotypes during the dry–rehydration cycle and life cycle of *B. hygrometrica*. (*A*) Vegetative phenotypes of hydrated (HD), dry (2 weeks withholding water), and rehydrated for 48 h (RH) *B. hygrometrica*. (*B*) Life cycle of *B. hygrometrica* from seed germination or leaf regeneration to mature plant. (Scale bar for seed morphology, 1 mm; scale bar for plants, 1 cm.)

generated from 4.74×10^{11} high-quality reads (*SI Appendix*, Table S1), and represents 91.52% of the ~1,691-Mb estimated genome size (*SI Appendix*, Table S2) predicted from 17-nucleotide depth distribution (*SI Appendix*, Fig. S2 and Table S3). The assembly was generated by an iterative hybrid approach (Table 1 and *SI Appendix*, Fig. S3). Approximately 85.86% of the assembly is nongapped sequence. The quality of the assembly was assessed by alignment to Sanger-derived fosmid sequences, allowing only a limited potential for misassemblies (*SI Appendix*, Table S4 and Dataset S1). The extent of sequence coverage was confirmed by the mapping of 2,360 sequenced expressed sequence tags (*SI Appendix*, Table S5 and Dataset S2).

The fourfold degenerate synonymous site of the third codon position (4DTv) values for coding regions for each of the duplicate gene pairs in the pairwise orthologous segments within *B. hygrometrica* genome revealed two whole-genome duplication events (4DTv ~0.5 and ~1.0; Fig. 2*A*). The species divergence event between *B. hygrometrica* and *Solanum tuberosum* or *Solanum lycopersicum* (4DTv ~0.54 or 0.49) that occurred around the most recent duplication event in the *B. hygrometrica* genome (4DTv ~0.5) likely reflects the divergence of the Lamiales from the Solanales (Fig. 2*A*). The ancient duplications, composed of several intermittent small duplication events (4DTv ~0.9 to ~1.3), may explain the large genome size, high level of repetitive sequences, and multicopy genes in the *B. hygrometrica* genome. The *B. hygrometrica* genome possessed a higher guanine-cytosine (GC) content (42.30%) than *S. tuberosum*, *S. lycopersicum*, or *Arabidopsis thaliana* (Table 1 and *SI Appendix*, Fig. S4), which is close to the upper limit for dicots (13). More than three fourths of the genome is composed of repeat sequences (75.75% of the assembled genome; Table 1 and *SI Appendix*, Fig. S5 and Table S6), which is similar to other dicots (14) but somewhat higher than *S. tuberosum* (62.2%) (15). Much of the unassembled genome is also composed of repetitive sequences, and the majority of the repetitive sequences could not be associated with known transposable element families. Plant transposable elements (TEs) are a significant source of small RNAs that function to epigenetically regulate TE and gene activity and are known to regulate DT in dicots (16). A recently discovered retroelement expressed in *B. hygrometrica*, osmotic and alkaline resistance 1, strengthens the possible role for LTRs in stress tolerance, and perhaps DT (17).

The draft genome also encodes 196 microRNA (miRNA), 538 tRNA, 1,512 *rRNA*, and 151 snRNA genes (*SI Appendix*, Table S7). In comparison with other dicot genomes (18), the *B. hygrometrica* genome encodes a large number of *rRNA* genes, especially 5S *rRNA* genes. Apart from their obvious structural role in ribosomes, large numbers of *rRNA* repeats (rDNA) have been linked with DNA stability, at least in yeast (19): a function that would be advantageous for surviving desiccation. There are 1,119 5S *rRNA* genes interspersed throughout the genome. This is 25–50 times the number contained in the only two other Asterid genomes that have been sequenced: *S. lycopersicum* (47 5S *rRNA* genes) and *S. tuberosum* (23 5S *rRNA* genes). The majority of the 5S *rRNA* genes are interspersed throughout the genome (Dataset S3); only 34 were clustered in four scaffolds (*SI Appendix*, Fig. S6).

Gene prediction protocols revealed 49,374 protein-coding genes, 40.68% of which are supported by RNA-Seq data and 23,250 (47.09%) of which had sufficient similarity to database entries to tentatively assign gene function (see *SI Appendix*, Table

Table 1. Overview of assembly and annotation for the *B. hygrometrica* draft genomes

Item	Features
Genome size (predicted and assembled)	1,691 and 1,548 Mb
Assembled in predicted genome	91.52%
No gap sequences in assembled genome	85.86%
Number of scaffolds (>100 bp)	520,969
Total length of scaffolds	1,547,684,042
N50 (scaffolds)	110,988
Longest scaffold	1,434,191
Number of contigs (>100 bp)	659,074
Total length of contigs	1,328,817,553
N50 (contigs)	11,187
Longest of contigs	691,061
GC content	42.30%
Number of predicted gene models	49,374
Mean transcript length (mRNA)	2,535.41
Mean coding sequence length	977.30
Mean number of exons per gene	3.58
Mean exon length	273.12
Mean intron length	604.33
Number of genes annotated	23,250
Number of genes unannotated	47.09%
Number of miRNA genes	196
Mean length of miRNA genes	112.4 bp
miRNA genes share in genome	0.00142%
Number of *rRNA* fragments	1512
Mean length of *rRNA* fragments	101.6 bp
rRNA fragments share in genome	0.00988%
Number of tRNA genes	538
Mean length of tRNA genes	76.2 bp
tRNA genes share in genome	0.00264%
Number of snRNA genes	151
Mean length of snRNA genes	117.0 bp
snRNA genes share in genome	0.00114%
Total size of repeat sequences	1,172,433,882
Repeat sequences share in genome	75.75%
Total size of transposable elements	1,163,296,466
TEs share in genome	75.16%
Total size of tandem repeats	62,678,253
Tandem repeats share in genome	4.05%

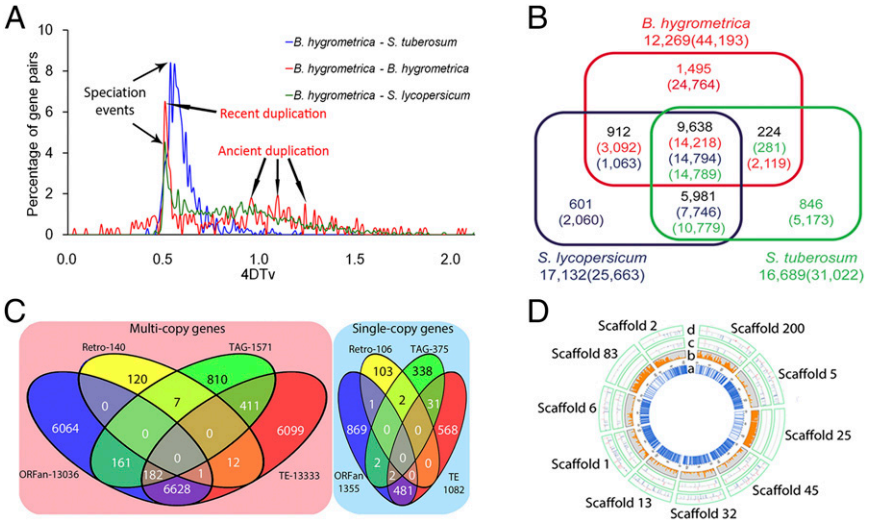

Fig. 2. *B. hygrometrica* genome features. (*A*) Genome duplication in genomes of *B. hygrometrica*, *S. tuberosum*, and *S. lycopersicum*, as revealed through 4DTv analyses. (*B*) A Venn diagram illustrating shared and specific gene families and genes (within brackets) in *B. hygrometrica*, *S. tuberosum*, and *S. lycopersicum*. The gene family and its related number of genes are listed in each of the components. (*C*) A Venn diagram of gene set. (*D*) Profiles integrating genome structures with DEGs of the longest 10 scaffolds. (*a*–*d*) Scaffolds indicating the distribution of ORFs (*a*, in blue), repetitive sequences with DNA II and RNA transposon (*b*, in yellow and orange), and DEG distribution on scaffolds in HD vs. 70% RWC and HD vs. 10% RWC (*c*, pink, accumulating DEGs; *d*, green, declining DEGs).

S8 and *SI Appendix*, *Results* for details). The structural features of the protein-coding gene complements for *B. hygrometrica* were closely comparable to those reported for *S. tuberosum* and *S. lycopersicum* but differed substantially from those reported for *Arabidopsis* (*SI Appendix*, Fig. S7 and Table S8). Of the predicted 12,269 potential gene families, 9,638 (~78.56%), involving 14,218 genes, are shared with *S. tuberosum* and *S. lycopersicum* genomes, reflecting the common origin between Lamiales and Solanales in asterids (Fig. 2*B*).

Predicted genes were functionally annotated by a consensus approach, using InterPro (20), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) (21), Swissprot, and Translated EMBL Nucleotide Sequence Data Library (TrEMBL) (22). The largest number of genes exhibited homology with proteins in the TrEMBL (46.12%) and InterPro (37.71%) databases (*SI Appendix*, Table S9). In total, 23,250 genes (47.09%) had sufficient similarity to database entries to tentatively assign gene function. Of the annotated protein-coding genes, multicopy genes outnumber single-copy genes by a factor of two (Fig. 2*C* and Dataset S4). Both categories contain an almost equal number of genes contained in TEs and genes classified as orphans [genes that are not a member of a gene family and have no significant sequence similarity to any entry in protein databases outside the taxon of interest (23)]. Up to 97% of the orphan genes originated from duplication events (*SI Appendix*, Table S10).

Of the genes that are historically associated with DT, in the *Boea* genome, only the early light-inducible protein (ELIP) gene family exhibits evidence of expansion. *B. hygrometrica* has seventeen ELIP genes (15 *ELIP1* and two *ELIP2*). One of the Asterid sequenced genomes, the *S. tuberosum* genome, reports a single ELIP gene (15), similar to the pea and tobacco genome (24), and *S. lycopersicum* has two ELIP genes (ELIP1 and ELIP2) (25), similar to *Arabidopsis* and barley.

The Genome and Desiccation Tolerance. To examine the response of the genome to desiccation, and to understand the architecture of its tolerance mechanisms within the genome, we profiled the dehydration-induced alteration of gene expression (Dataset S5). We constructed a genome-wide dehydration response profile by integrating the scaffold protein-coding and repetitive sequence mapping analysis with 9,888 differentially expressed genes (DEGs; identified as greater than twofold change in transcript abundances from that for hydrated controls, at a P value of < 0.05) during drying (Fig. 2*D* and Dataset S5). There was no obvious clustering of DEGs, the majority of which are located, as expected, predominantly in scaffolds that contain few repetitive sequences and that are gene-rich (Dataset S6). The lack of clustering of any significant number of DEGs with their scattered location among a large number of contigs suggests DT was not acquired in a recent evolutionary or restructuring event (sufficient time for dispersal of genes throughout the genome) but, rather, as a retooling of existing genetic elements to deliver the DT phenotype in vegetative tissues.

Gene Expression and Desiccation. The majority of genes expressed in the leaves of *B. hygrometrica* belong to gene families. The large number of orphan genes, ~29% of all annotated genes and 8.51–10.48% of expressed annotated genes, was within the expected range for orphan gene content of eukaryotic genomes (*SI Appendix*, Table S11) (23), of which only a small number (a maximum of 128) were significantly responsive to dehydration (*SI Appendix*, Table S11). Of the 9,888 DEGs, 58.18% responded to moderate dehydration [70% relative water content (RWC)] and 87.47% responded to dehydration to 10% RWC (Fig. 3*A* and Dataset S5). There were 1,239 DEGs that only responded to moderate dehydration (769 increase and 470 decline), and 4,135 specifically responded during desiccation (2,188 increase and 1,947 decline).

The assignment of GO terms for 7,716 DEGs (Dataset S5) focuses on membrane components and organelle structure, biopolymer molecular processes and intermediary metabolism, and metal binding, hydrolytic, and oxidoreductase activities (Fig. 3*B* and *SI Appendix*, Table S12). Enrichment analysis of the 7,758 DEGs with KEGG annotation (Fig. 3*C* and Datasets S5 and S7) revealed that glycerophospholipid metabolism and soluble *N*-ethylmaleimide sensitive fusion attachment protein receptor interactions in vesicular trafficking (both processes involved in membrane maintenance) are favored during dehydration. Dehydration also favored transcripts involved in the pathogen defense system, a common observation for abiotic stress responses, and one often brokered by plant hormones [e.g., abscisic acid (ABA) (26)]. As tissues approach desiccation, transcripts that populate the mRNA surveillance pathway appear and accumulate, indicating a need to remove damaged transcripts from the drying cells. Dehydration also resulted in depletion of transcripts that represent a wide range of metabolic processes (Fig. 3*C*), primarily for pathways involved in growth (photosynthesis and nitrogen metabolism). A more focused clustering of 734 high-level DEGs revealed three major clusters ($\log_2$ base mean value in one sample is more than fourfold higher than that in any other sample; Fig. 3 *D* and *E*, *SI Appendix*, *Results*, and Dataset S8), offering a broad assessment of the response to desiccation and a broad comparison with similar transcriptomes of other resurrection dicots (5).

This and other studies of vegetative dehydration/desiccation transcriptomes (27) point toward a central core of genes and gene products associated with the ability to survive drying: ABA metabolism and signaling, phospholipid signaling, late

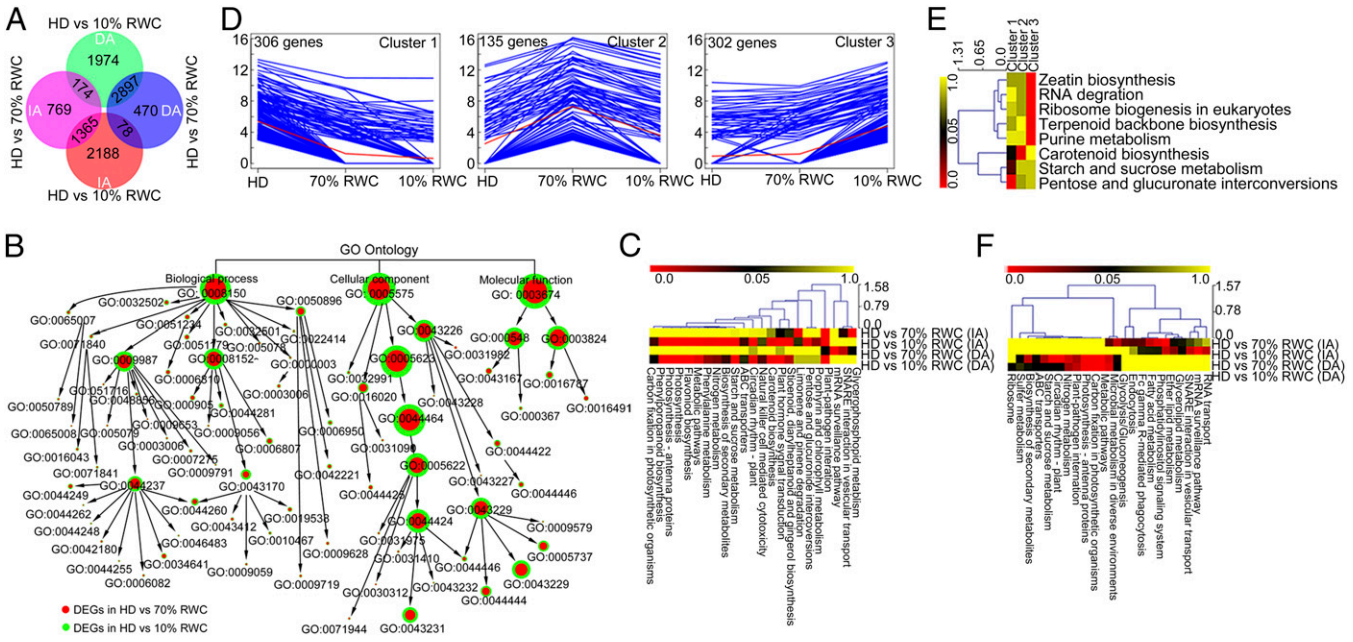

Fig. 3. Transcriptional responses during dehydration. (*A*) Venn diagrams show the number of differentially expressed genes during dehydration and rehydration: hydrated (HD), dehydration (70% RWC), and desiccation (10% RWC). IA, increased abundance; DA, decline in abundance. (*B*) GO classifications of the DEGs respond to dehydration and desiccation. Only GO terms with a gene number larger than 150 are shown. (*C*) Heat maps of significantly enriched pathways in DEGs during dehydration. The yellow and red colors indicate the Q-value for significantly enriched pathways. (*D*) Clusters of high-level ($\log_2$ fold change > 4) DEGs during dehydration. The *y* axis gives the normalized expression level by DESeq software (on a log scale) of DEGs. Each blue line represents a different gene, and the red line indicates the gene expression trend of DEGs in each cluster. (*E*) The heat map describes the significantly enriched pathways in each cluster. The yellow and red color delineates the Q-value for significantly enriched pathways. (*F*) Significantly enriched pathways for those DEGs for which alternative splicing occurred during dehydration. The yellow and red color shows the Q-value for significantly enriched pathways.

embryogenesis abundant proteins (LEAs) (protective proteins), components of reactive oxygen species (ROS) protection and detoxification pathways, and ELIPs (Dataset S9).

Of the 21 DEGs associated with ABA metabolism, eight positive DEGs encode enzymes directly involved in ABA biosynthesis and catabolism, indicating tight control of ABA levels during dehydration (*SI Appendix*, *Results*). A single phospholipase D gene, *PLD-1α*, controlled, in part, desiccation response of the resurrection dicot *Craterostigma plantagineum* (28). This may also be the case for *B. hygrometrica*, as evidenced by the increased abundance of transcripts from one of the two PLD-1α genes during dehydration (Dataset S9). Other *PLDs* (three *PLD-γs*, a *PLD-β*, and a *PLD-P1/Z1*) also responded positively to dehydration, indicating that phospholipid signaling may be more complex in *B. hygrometrica*.

The *B. hygrometrica* genome contains a plethora of LEA protein genes [65 with 51 expressed and 47 DEGs (Dataset S9)], which is a much greater number than reported for the transcriptomes of *C. plantagineum* (27) or *Haberlea rhodopensis* (29). The greater number of expressed LEA genes may reflect the length and severity of the seasonal dehydration periods experienced by *Boea* compared with the other resurrection species (*SI Appendix*, *Results*). The proteins derived from two *LEA1s*, *Bhs4_093* and *Bhs4_094*, have been demonstrated to stabilize the photosynthetic proteins (such as LHCs) in transgenic tobacco seedlings during dehydration and rehydration (30).

The response to dehydration for genes involved in ROS protection and mitigation of oxidative damage is a complex one. Early studies revealed the importance of glutathione metabolism in the dehydration response of *Boea* species (11). Specific members of the GST gene family responded to dehydration stress, along with several peroxidases (Dataset S9), indicative of a need for detoxification and repair of oxidative damage (*SI Appendix*, *Results*).

The increase in abundance of ELIP transcripts is a common feature of the response of DT plants to dehydration (4), as observed in *H. rhodopensis* and *C. plantagineum* (27, 29). Thirteen of the 17 ELIP orthologs in the *B. hygrometrica* genome were ranked as positive DEGs (*SI Appendix*, *Results* and Dataset S9). It thus appears that the protection of photosystem II is a major aspect of the DT mechanism for *B. hygrometrica*.

Relating the RNA-Seq data for dehydrating to the draft genome revealed that 7,127 of the genes represent two or more alternative splicing (AS) products, delivering more functional variation than specified by the annotated gene complement alone (Dataset S10). Of the DEGs, 4,491 (45.42%) exhibited AS during dehydration (*SI Appendix*, Table S13 and Dataset S11). Alternative 5′ splice sites dominated the four major AS patterns. Pathway enrichment of AS-DEGs favored an increase in abundance of transcripts related to endocytosis and Fc gamma R-mediated phagocytosis, fatty acid metabolism, and peroxisomal functions, suggestive of needs for membrane component and protein removal or recycling as cells lose water, as well as an ongoing repair of membranes and removal of ROS. AS was also involved in transcript selections for the processes that were revealed in the overall analysis of DEGs mentioned previously (Fig. 3*F*) (31).

Discussion

Vegetative DT most likely evolved in certain angiosperm lineages from selection pressures exerted by an environment that delivered lengthy periods of little or no soil water. The lack of DT-specific genome organizational features in *B. hygrometrica*, such as clustering of DEGs, supports the contention that vegetative DT evolved primarily from an alteration in the regulation of preexisting genetic modules. This most likely involved those genetic components that deliver developmentally controlled DT to seeds and pollen (32). A portion of that alteration in the regulation of gene expression in *B. hygrometrica* clearly involves AS of transcripts and the plant hormone ABA.

The *B. hygrometrica* genome offers some important insights into the genetic strategies used for accomplishing vegetative DT and its evolution in this resurrection species. The large number of orphan genes housed within the genome, ∼10% of expressed

genes, reflects the somewhat unique nature of this resurrection species. Orphan genes are thought to represent lineage-specific adaptations and, in some plant species, to be linked to stress responses (e.g., rice) (33). This may also be true for the expressed orphan genes of *B. hygrometrica*, but only a small number (128) can, at this point, be associated with the resurrection phenotype and probably represent species-specific aspects of the DT mechanism.

The apparent expansion of 5S *rRNA* genes in the *Boea* lineage may reflect the need for a supply of active ribosomes during the rapid resumption of protein synthesis (and recovery) on rehydration. Because ribosomal 5S *rRNA* transcripts can only be amplified by transcription, it would seem reasonable to suggest the 5S *rRNA* gene expansion in *B. hygrometrica* evolved to meet the protein synthesis burden inherent in the resurrection phenotype. As this is the first resurrection genome, to our knowledge, to be sequenced, it remains to be seen whether this is a common genotypic feature of resurrection species.

The genome sequence and transcriptome also revealed an expansion of the ELIP gene family in *B. hygrometrica* concomitant with enhanced transcript abundance for 13 of the 17 gene family members. ELIP proteins are postulated to protect the photosynthesis machinery from photooxidative damage by preventing the accumulation of free chlorophyll by binding pigments and preserving the chlorophyll-protein complexes (34). ELIP proteins (and transcripts) have been reported to increase in abundance in a linear fashion with the amount of photoactivation and photodamage to the photosystem II reaction centers, D1 protein degradation, and changes in pigment level (24). Photooxidative damage is a primary stressor for resurrection species, as they spend a considerable amount of time in the dried state and under high-light conditions (35). Thus, it appears that *B. hygrometrica* has evolved a strategy of ELIP gene expansion to aid in its ability to protect its photosynthetic apparatus, particularly photosystem II, from oxidative damage: an essential and perhaps central aspect of its DT mechanism. The transcriptomic analysis provides a broader perspective on the nature of the cellular protection aspects of vegetative DT, highlighted by the increase in transcript abundance for LEA protein genes, GST gene family, and peroxidases.

The draft genome offers a unique opportunity to construct a systems approach to understanding the mechanistic aspects of DT and resurrection in plants. Such an approach can help influence our understanding of the evolution of the land plants and our attempts to design strategies for the improvement of the dehydration tolerance of our major crops as food security issues increase in importance globally.

Materials and Methods

The original accessions for *B. hygrometrica* were collected from a dry rock crack in Fragrant Hills in a Beijing suburb in China. The genome was sequenced using the whole-genome shotgun approach, using Illumina HiSeq and Roche 454 platforms. Whole-genome shotgun data were used to assemble the draft genome, using the hybrid assembly strategy by Newbler, SSPACE, and SOAP de novo algorithm. Genes were annotated using a combined approach on the repeat masked genome with ab initio gene predictions, protein similarity, and transcripts to build optimal gene models. Repeat sequences were identified by both de novo approach and sequence similarity at the nucleotide and protein levels. Detailed information of materials, methods, and any associated references are available in the *SI Appendix, Materials and Methods*.

ACKNOWLEDGMENTS. We thank the Beijing Genetics Institute staff members and Capital Normal University graduate students for their assistances on genome sequencing, assembling, and bioinformatic analyses. This study was supported by funds from the Chinese Ministry of Agriculture (2014ZX08009-23B, 2009ZX08009-058B), Chinese Ministry of Science and Technology (2007AA021405), and the Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality (Y. He). We are also thankful for the special financial support from the National Key Disciplines of China for this project.

1. Oliver M-J, Tuba Z, Mishler B-D (2000) The evolution of vegetative desiccation tolerance in land plants. *Plant Ecol* 151(1):85–100.
2. Proctor M-C-F, Pence V-C (2002) Vegetative tissues: Bryophytes, vascular resurrection plants, and vegetative propogules. *Desiccation and survival in plants: Drying without dying*, eds Black M, Pritchard H-W (CABI, Wallingford, Oxon), pp 207–267.
3. Moore J-P, Le N-T, Brandt W-F, Driouich A, Farrant J-M (2009) Towards a systems-based understanding of plant desiccation tolerance. *Trends Plant Sci* 14(2):110–117.
4. Oliver MJ, Cushman JC, Koster KL (2010) Dehydration tolerance in plants. *Methods Mol Biol* 639:3–24.
5. Gechev T-S, Dinakar C, Benina M, Toneva V, Bartels D (2012) Molecular mechanisms of desiccation tolerance in resurrection plants. *Cell Mol Life Sci* 69(19):3175–3186.
6. Oliver M-J, et al. (2011) A sister group contrast using untargeted global metabolomic analysis delineates the biochemical regulation underlying desiccation tolerance in Sporobolus stapfianus. *Plant Cell* 23(4):1231–1248.
7. Gerten D, et al. (2011) Global water availability and requirements for future food production. *J Hydrometeorol* 12(5):885–900.
8. Cushman JC, Oliver MJ (2011) Understanding vegetative desiccation tolerance using integrated functional genomics approaches within a comparative evolutionary framework. *Plant desiccation tolerance in plants,* eds. Luttge U, Beck E, Bartels D (Springer: Heidelberg), pp 307–338.
9. Deng X, Hu Z, Wang H (1999) mRNA differential display visualized by silver staining tested on gene expression in resurrection plant *Boea hygrometrica*. *Plant Mol Biol Rep* 17(3):279–279.
10. Wilson C-L (1974) Floral anatomy in Gesneriaceae. I. Cyrtandroideae. Bot Gaz 135:247–268.
11. Jiang G, et al. (2007) Proteome analysis of leaves from the resurrection plant *Boea hygrometrica* in response to dehydration and rehydration. *Planta* 225(6):1405–1420.
12. Vicré M, Farrant J-M, Driouich A (2004) Insights into the cellular mechanisms of desiccation tolerance among angiosperm resurrection plant species. *Plant Cell Environ* 27(11):1329–1340.
13. Matassi G, Montero L-M, Salinas J, Bernardi G (1989) The isochore organization and the compositional distribution of homologous coding sequences in the nuclear genome of plants. *Nucleic Acids Res* 17(13):5273–5290.
14. Flavell R-B, Bennett M-D, Smith J-B, Smith D-B (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem Genet* 12(4):257–269.
15. Xu X, et al.; Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475(7355):189–195.
16. Hilbricht T, et al. (2008) Retrotransposons and siRNA have a role in the evolution of desiccation tolerance leading to resurrection of the plant *Craterostigma plantagineum*. *New Phytol* 179(3):877–887.
17. Zhao Y, et al. (2014) Identification of a retroelement from the resurrection plant *Boea hygrometrica* that confers osmotic and alkaline tolerance in *Arabidopsis thaliana*. *PLoS ONE* 9(5):e98098.
18. Gorman SW, Teasdale RD, Cullis CA (1992) Structure and organization of the 5S rRNA genes (5S DNA) in *Pinus radiata* (*Pinaceae*). *Plant Syst Evol* 183(3-4):223–234.
19. Kobayashi T (2011) Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. *Cell Mol Life Sci* 68(8):1395–1403.
20. Mulder N-J, et al. (2007) New developments in the InterPro database. *Nucleic Acids Res* 35(Database issue):D224–D228.
21. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
22. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1):27–30.
23. Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. *Nat Rev Genet* 12(10):692–702.
24. Yurina NP, Mokerova DV, Odintsova MS (2013) Light-inducible stress plastid proteins of phototrophs. *Russ J Plant Physiol* 60(5):577–588.
25. Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485(7400):635–641.
26. Lee S-C, Luan S (2012) ABA signal transduction at the crossroad of biotic and abiotic stress responses. *Plant Cell Environ* 35(1):53–60.
27. Dinakar C, Bartels D (2013) Desiccation tolerance in resurrection plants: New insights from transcriptiome, proteome, and metabolome analysis. *Front Plant Sci* 4(482):1–14.
28. Frank W, Munnik T, Kerkmann K, Salamini F, Bartels D (2000) Water deficit triggers phospholipase D activity in the resurrection plant *Craterostigma plantagineum*. *Plant Cell* 12(1):111–124.
29. Gechev T-S, et al. (2013) Molecular mechanisms of desiccation tolerance in the resurrection glacial relic *Haberlea rhodopensis*. *Cell Mol Life Sci* 70(4):689–709.
30. Liu X, et al. (2009) LEA 4 group genes from the resurrection plant *Boea hygrometrica* confer dehydration tolerance in transgenic tobacco. *Plant Sci* 176(1):90–98.
31. The Angiosperm Phylogeny Group (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot J Linn Soc* 141(4):399–436.
32. Galf DF, Oliver MJ (2013) The evolution of desiccation tolerance in angiosperm plants: A rare yet common phenomenon. *Funct Plant Biol* 40(4):315–328.
33. Guo W-J, Li P, Ling J, Ye S-P (2007) Significant comparative characteristics between orphan and nonorphan genes in the rice (*Oryza sativa L.*) genome. *Comp Funct Genomics* 2007:21676.
34. Alamillo J-M, Bartels D (2001) Effects of desiccation on photosynthesis pigments and the ELIP-like dsp 22 protein complexes in the resurrection plant Craterostigma plantagineum. *Plant Sci* 160(6):1161–1170.
35. Farrant J-M, Willigen V-C, Lofell D-A, Bartsch S, Whittaker A (2003) An investigation into the role of light during desiccation of three angiosperm resurrection plants. *Plant Cell Environ* 26(8):1275–1286.