

# A footprint of desiccation tolerance in the genome of *Xerophyta viscosa*

Maria-Cecília D. Costa<sup>1,2†</sup>, Mariana A. S. Artur<sup>1†</sup>, Julio Maia<sup>1</sup>, Eef Jonkheer<sup>1,3</sup>, Martijn F. L. Derks<sup>1,3</sup>, Harm Nijveen<sup>1,3</sup>, Brett Williams<sup>4</sup>, Sagadevan G. Mundree<sup>4</sup>, José M. Jiménez-Gómez<sup>5</sup>, Tamara Hesselink<sup>6</sup>, Elio G. W. M. Schijlen<sup>6</sup>, Wilco Ligterink<sup>1</sup>, Melvin J. Oliver<sup>7</sup>, Jill M. Farrant<sup>2</sup> and Henk W. M. Hilhorst<sup>1\*</sup>

**Desiccation tolerance is common in seeds and various other organisms, but only a few angiosperm species possess vegetative desiccation tolerance. These ‘resurrection species’ may serve as ideal models for the ultimate design of crops with enhanced drought tolerance. To understand the molecular and genetic mechanisms enabling vegetative desiccation tolerance, we produced a high-quality whole-genome sequence for the resurrection plant *Xerophyta viscosa* and assessed transcriptome changes during its dehydration. Data revealed induction of transcripts typically associated with desiccation tolerance in seeds and involvement of orthologues of ABI3 and ABI5, both key regulators of seed maturation. Dehydration resulted in both increased, but predominantly reduced, transcript abundance of genomic ‘clusters of desiccation-associated genes’ (CoDAGs), reflecting the cessation of growth that allows for the expression of desiccation tolerance. Vegetative desiccation tolerance in *X. viscosa* was found to be uncoupled from drought-induced senescence. We provide strong support for the hypothesis that vegetative desiccation tolerance arose by redirection of genetic information from desiccation-tolerant seeds.**

The grass family (Poaceae) is arguably the most important contributor to global food security. However, poaceous staple crops, such as maize, corn, rice and wheat, do not survive the extreme water loss that is inevitably brought about by extended periods of drought. In contrast, their seeds are desiccation tolerant, and they can be dried to water contents as low as 1–5% on a fresh weight basis without losing viability<sup>1</sup>. There are some 135 angiosperm species, termed ‘resurrection plants’, which produce desiccation-tolerant seeds and possess desiccation-tolerant vegetative tissues<sup>2–4</sup>. Vegetative desiccation tolerance (DT) first arose with the transition from aquatic to terrestrial life forms, when both the probability of experiencing adverse conditions and the survival cost of such conditions were high<sup>5</sup>. As plants expanded into terrestrial habitats and developed tracheids to move water from the substrate to their aerial parts and more complex ecosystems were established, the slow growth characteristic of desiccation-tolerant plants was limiting their competitive ability. This favoured the loss of DT in vegetative tissues, the development of mechanisms to prevent water loss (physiological and morphological) and the confinement of DT to seeds, spores and pollen grains where it was required for dispersal and preservation of genetic resources<sup>6,7</sup>. It has been proposed that vegetative DT reappeared in the angiosperms, presumably in response to colonization of environmentally demanding habitats, in at least 13 separate lineages to evolve the present day resurrection plants<sup>4,7,8</sup>.

The myriad genetic changes experienced in the evolution of resurrection plants that enabled vegetative DT are not completely

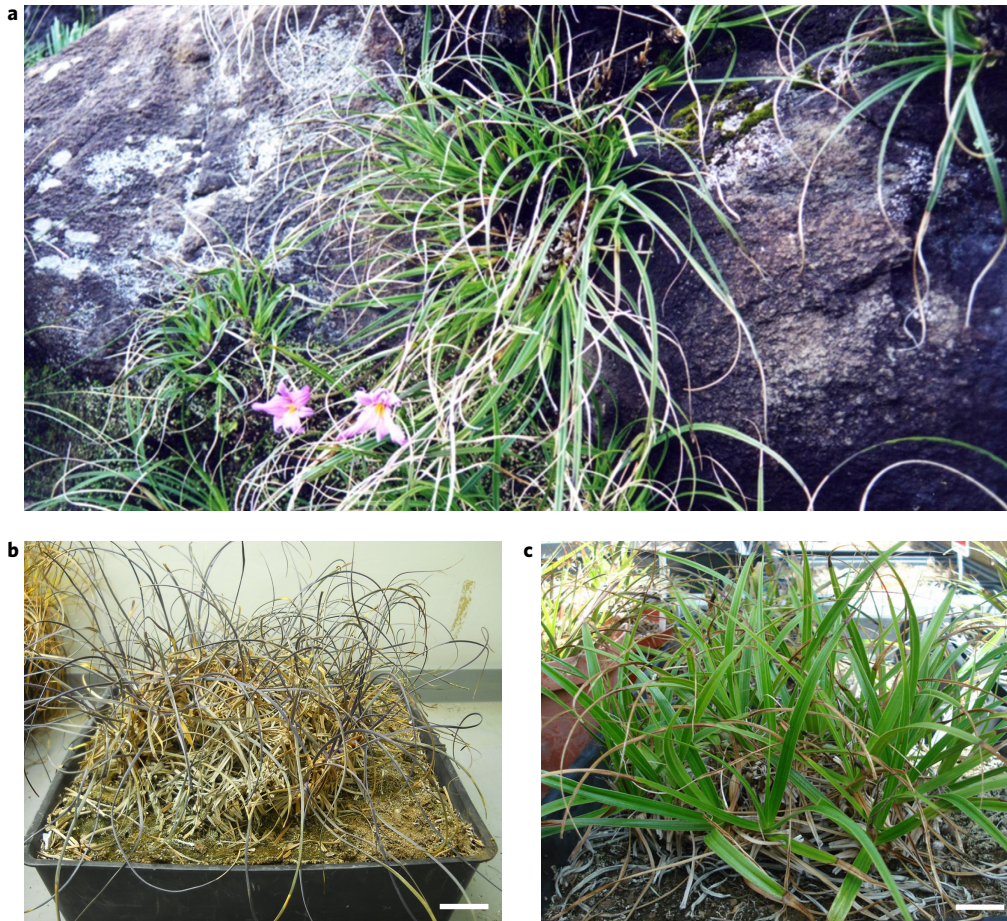
understood. An improved understanding of these changes will aid the development of crop improvement strategies for tolerance of water loss and survival of extreme drought conditions<sup>9</sup>. The pressure to develop such crop varieties has been intensified by predictions of a near future with increased drought and declining water resources in the world’s main agricultural areas<sup>10</sup>. This pressure is strongest in developing countries, which struggle to maintain robust breeding capabilities and need support to develop stress-tolerant crops critical for food security<sup>9</sup>. In this context, the study of resurrection plants, such as *Xerophyta viscosa* (Velloziaceae), will bring valuable information to bear on improving stress tolerance in crops. *X. viscosa* (Fig. 1) is a monocotyledonous plant species and, thus, phylogenetically closely related to staple cereal crops. It is thus an ideal model for understanding plant requisites to tolerate extreme dehydration.

We have produced a high-quality whole-genome sequence and assembly for *X. viscosa* from a mixed dataset of Illumina and Pac-Bio reads along with a full assessment of the transcriptomal changes that occur in young seedlings in response to exogenous abscisic acid (ABA) and in adult plants during desiccation and rehydration.

## A high-quality octoploid genome assembly

The source of genomic DNA consisted of individuals (five for Illumina and one for PacBio sequencing) grown from seeds harvested from a population of *X. viscosa* collected from the Buffelskloof Nature Reserve (Mpumalanga Province, South Africa). *X. viscosa* has a high level of heterozygosity due to

<sup>1</sup>Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands. <sup>2</sup>Department of Molecular and Cell Biology, University of Cape Town, Private Bag, 7701 Cape Town, South Africa. <sup>3</sup>Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands. <sup>4</sup>Centre for Tropical Crops and Biocommodities, Queensland University of Technology, PO Box 2434, Queensland 4001, Brisbane, Australia. <sup>5</sup>Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Cologne, Germany. <sup>6</sup>Bioscience, Wageningen Plant Research International, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands. <sup>7</sup>USDA-ARS-MWA-PGRU, 205 Curtis Hall, University of Missouri, Columbia, Missouri 65211, USA. <sup>†</sup>These authors contributed equally to this work. \*e-mail: [henk.hilhorst@wur.nl](mailto:henk.hilhorst@wur.nl)



**Figure 1 | *X. viscosa* phenotypes. a**, The chasmophytic nature of growth in the natural environment. **b**, A typical fully dehydrated plant (withholding water from the whole plant over a period of 25 days until <5% RWC). **c**, A rehydrated plant (at full turgor after 5 days of watering). Scale bars, 10 cm.

self-incompatibility. High levels of heterozygosity could cause several haplotypes to be assembled in different contigs causing duplicates in the final haploid assembly. However, we used Falcon<sup>11</sup> which is a tool designed to deal with high levels of heterozygosity and, thus, no measures were taken to reduce genome complexity. The genome was sequenced and assembled<sup>12–15</sup> using a whole-genome shotgun approach which combined 23 Gb of raw paired-end Illumina reads (~77× coverage) and 17.5 Gb of PacBio long reads (~58× coverage). This high-quality assembly covers 99.8% of the genome and consists of 1,811 contigs and 896 scaffolds (Table 1).

A k-mer analysis of PacBio data indicated a (haploid) genome size of 295.5 Mb. Based on allele frequencies<sup>16</sup>, we determined the *X. viscosa* genome to be an octoploid (Supplementary Fig. 1). Polyploidy has been associated with vigour and size in crops, has been proposed to facilitate better environmental adaptability and is regarded as one of the major drivers of speciation, and thus evolution. It follows that polyploid species may be more competitive and invasive in new environments<sup>17</sup>. The African Velloziaceae species are all polyploid and assumed to have originated on the South-American continent, where family members are largely diploid<sup>17</sup>. Thus, polyploidization may have contributed to successful invasion of the Velloziaceae in southern Africa. Our data indicates that members of the Velloziaceae may be octoploids and not hexaploids, which is the current consensus of opinion.

*X. viscosa* has 48 small chromosomes (estimated length of 1.5–2.0 microns, Supplementary Fig. 1). We assembled a 295.5 Mb haploid genome and 90% of the assembly is contained in 238 scaffolds. The scaffold N50 (the scaffold size above which 50% of the total length of the sequence assembly can be found) is 1.67 Mb, which is very high for an octoploid genome. It approaches the N50 of the diploid

resurrection species *Oropetium thomaeum* (N50 2.4 Mb, genome size 245 Mb)<sup>18</sup> and is much greater than the N50 of the diploid resurrection species *Boea hygrometrica* (N50 110 kb, ~1,691-Mb sequenced genome)<sup>19</sup> or the tetraploid *Chenopodium quinoa* (N50 87 kb, genome size 1,448 Mb)<sup>20</sup>. In addition to the 295 Mb final assembly, we assembled 125 Mb of additional haplotypes in 6,103 contigs.

The GC content is 36.5% across the genome (Table 1). Greater GC contents have been associated with species that grow in seasonally cold and/or dry climates, perhaps suggesting an advantage of GC-rich DNA during cell freezing and desiccation<sup>21</sup>. Whereas *B. hygrometrica* has a relatively high GC content of 42.30%, it is only 34.86% in *O. thomaeum*, thus not supporting a general positive correlation between DT and GC content.

A total of 25,425 protein-coding genes were annotated; 97% of the encoded proteins exhibit high sequence similarity to proteins in the TrEMBL database<sup>22</sup>, 85% in Swiss-Prot<sup>22</sup> and 84% in InterPro<sup>23</sup>. The percentage of orphan genes (5.4%, Table 1), or genes that do not share any similarity with genes in other species, is considerably lower than expected for eukaryotic genomes (10–20%)<sup>24</sup>. This low percentage may suggest that the acquisition of vegetative DT by *X. viscosa* relied more on the redirection of genetic information than on the genesis of novel genes.

Transposable elements (TEs) account for 18% of the genome, which is a surprisingly low number (Table 1) considering that TEs in the genomes of *O. thomaeum* and *B. hygrometrica* account for 75% and 43%, respectively. Our high-quality data rules out collapsed or incomplete sequence assemblies. Plant TEs have been proposed to regulate DT in the dicot resurrection species *Craterostigma plantagineum*<sup>25</sup>. Taken together, the proportion of TEs in a plant genome does not appear to be related to DT in general.

**Table 1 | Properties of the *X. viscosa* genome.**

| Assembly                      | Number             | N50 (Mb)                    | L90                | Total length (Mb)     | Alignment rate (%) |
|-------------------------------|--------------------|-----------------------------|--------------------|-----------------------|--------------------|
| Scaffolds                     | 896                | 1.67                        | 238                | 295.5                 | 96.7               |
| Contigs                       | 1,811              | 1.11                        | 448                | 293.6                 | 96.3               |
| Additional haplotype contigs  | 6,103              | 0.03                        |                    | 125                   |                    |
| Annotation                    | Number             | Mean length (bp)            | Density            | Genome percentage (%) |                    |
| BUSCO (missing/total)         | 57/956             |                             |                    |                       |                    |
| BUSCO* (missing/fragmented)   | 51/31              |                             |                    |                       |                    |
| GC content                    |                    |                             |                    | 36.51                 |                    |
| Protein-coding genes          | 25,425             | 4,444.5                     | -                  | 38.2                  |                    |
| Exons                         | 149,027            | 265.9                       | 5.9 exons per gene | 13.4                  |                    |
| Introns                       | 120,004            | 610.7                       | 4.7 exons per gene | 24.8                  |                    |
| rRNA                          | 165                | 429.5                       | -                  | <0.01                 |                    |
| snRNA                         | 140                | 86.69                       | -                  | <0.01                 |                    |
| tRNA                          | 289                | 74.91                       | -                  | <0.01                 |                    |
| TEs                           | 116,932            | -                           | -                  | 18.3                  |                    |
| miRNA                         | 165                | 126.6                       | -                  | <0.01                 |                    |
| Orphan genes <sup>†</sup>     | 1,372              | 143.4                       | -                  | 0.066 (5.4% of genes) |                    |
| Polymorphisms                 | Number             | Density (kb <sup>-1</sup> ) |                    |                       |                    |
| SNPs                          | 1,384,518          | 4.7                         |                    |                       |                    |
| INDELS                        | 375,931            | 1.3                         |                    |                       |                    |
| Multi-allelic sites           | 65,241             | 0.2                         |                    |                       |                    |
| Repeat class                  | Number of elements | Genome percentage (%)       |                    |                       |                    |
| Retrotransposon               | 63,029             | 0.12                        |                    |                       |                    |
| LTR                           | 45,072             | 0.10                        |                    |                       |                    |
| Gypsy (RLG)                   | 30,293             | 0.08                        |                    |                       |                    |
| Copia (RLC)                   | 12,725             | 0.02                        |                    |                       |                    |
| Penelope (RPX)                | 119                | <0.01                       |                    |                       |                    |
| Unknown LTR (RLX)             | 288                | <0.01                       |                    |                       |                    |
| LINE (RIL)                    | 15,055             | 0.01                        |                    |                       |                    |
| SINE (RSX)                    | 2,902              | <0.01                       |                    |                       |                    |
| DIRS (RYD)                    | 127                | <0.01                       |                    |                       |                    |
| Unknown retrotransposon (RXX) | 292                | <0.01                       |                    |                       |                    |
| DNA transposon                | 56,603             | 0.07                        |                    |                       |                    |
| Maverick (DMX)                | 584                | <0.01                       |                    |                       |                    |
| Unknown DNA transposon (DXX)  | 1,776              | <0.01                       |                    |                       |                    |
| No category                   | 154,855            | 0.18                        |                    |                       |                    |

N50: scaffold size above which 50% of the total length of the sequence assembly can be found. L90: number of contigs whose summed length contains at least 90% of the sum of the total length of the sequence assembly. BUSCO, Benchmarking Universal Single-Copy Orthologs; rRNA, ribosomal RNA; snRNA, small nuclear RNA; SNP, single nucleotide polymorphism; INDEL, insertion or deletion of bases in the DNA; LTR, long terminal repeat. \*After MAKER annotation. <sup>†</sup>No hits in Swissprot, TrEMBL with cut off  $1 \times 10^{-10}$  and PFAM cut off  $1 \times 10^{-6}$ .

Orthologous clustering of the *X. viscosa* proteome with 15 plant genomes identified 15,450 orthologous groups (OGs) in common, with 20 OGs present specifically in *X. viscosa* (Supplementary Table 1). The OGs were used to map signatures of expansions and contractions of gene families (Supplementary Fig. 2). Overall, *X. viscosa* did not undergo extensive expansions or contractions of gene families. Although many gene families were shared between *X. viscosa* and other monocots, such as the desiccation-sensitive but relatively drought-tolerant *Eragrostis tef*<sup>26</sup> (*X. viscosa* and *E. tef* share 70% of OGs) and the resurrection plant *O. thomaeum* (*X. viscosa* and *O. thomaeum* share 72% of OGs), clearly there are several expansions and contractions unique to *X. viscosa* (Supplementary Fig. 2). The *X. viscosa* exclusive expansions include several gene families associated with cellular stress and metabolic regulation, such as putative late embryogenesis abundant (LEA) proteins, heat shock domain-containing proteins, putative WRKY transcription-factors and MADS-box domain-containing proteins. The contractions unique to *X. viscosa* include gene families related to control of gene expression, such as plastid transcriptionally active chromosome proteins and DNA helicases. The low overlap in expanded or contracted gene families between *X. viscosa* and *O. thomaeum* may indicate different genetic architectures underlying the resurrection phenotype

observed in both species and hint at independent evolution to reacquire vegetative DT<sup>27</sup>.

### Resurrection physiology

*X. viscosa* seeds and adult plants are desiccation tolerant. However, individuals are not continuously desiccation tolerant from seed to adult plant. They lose DT briefly upon germination, which recovers gradually during seedling development, first in shoots and later in roots. This recovery of DT can be induced earlier in shoots, but not in roots, by application of ABA (Supplementary Fig. 3). In shoots, exogenous ABA induced the accumulation of transcripts of genes involved in chlorophyll degradation, translational control of gene expression and transport, whereas it decreased transcript abundance of genes related to photosynthesis, energy metabolism and ABA biosynthesis (Supplementary Table 2). In roots, transcript abundance of genes involved in chlorophyll degradation and responsiveness to brassinosteroids increased, whereas transcripts of genes related to energy metabolism and plant responses to external signals (mainly by modifying cell walls) were depleted (Supplementary Table 2).

In adult plants, stomata closure and carbon gain from photosynthesis ceases when water content drops below 55% relative water content (RWC)<sup>9,28</sup>. In desiccation-sensitive species, cessation of

carbon gain coupled with continued water loss results in metabolic stress, whereas resurrection plants redirect their metabolism towards subcellular protection and ultimate quiescence<sup>29,30</sup>. In *X. viscosa*, the molecular signature of this metabolic redirection is reflected in the higher number of genes that exhibit differential expression at 40% RWC (1.0 gH<sub>2</sub>O g<sup>-1</sup> dwt) compared to other hydrated states (Supplementary Fig. 4). At 40% RWC, energy metabolism is strongly repressed and transcripts of genes related to chlorophyll degradation significantly accumulated. Significant accumulation of transcripts related to chlorophyll synthesis and chloroplast and thylakoid formation were also observed during dehydration at 40% RWC (Supplementary Table 3), reflecting the strategy of poikilochlorophyll adopted by *X. viscosa*. Poikilochlorophyllous resurrection plants dismantle their photosynthetic apparatus (thylakoids and chlorophyll) in a controlled manner during dehydration, and reconstitute this upon rehydration utilizing transcripts stably stored in desiccated tissues<sup>9,31–33</sup>.

This metabolic redirection is also evident from the grouping of the differentially expressed genes in four distinct self-organizing maps (SOMs, Fig. 2). From these, it is evident that two major changes in gene expression occur: between 60% RWC (1.5 gH<sub>2</sub>O g<sup>-1</sup> dwt) and 40% RWC (1.0 gH<sub>2</sub>O g<sup>-1</sup> dwt, SOMs 1 and 3) and between 40% RWC and 20% RWC (0.5 gH<sub>2</sub>O g<sup>-1</sup> dwt, SOMs 2 and 4). Genes related to protein folding, protection and translation control are enriched exclusively in SOM 1, whereas genes related to nuclear import and control of gene expression are enriched exclusively in SOM 2 (Supplementary Table 2). Likewise, a number of transcription factor-encoding genes and co-regulators were highly induced in desiccated leaves of the resurrection species *Haberlea rhodopensis*<sup>34</sup>, reflecting the massive transcriptional reprogramming behind the metabolic redirection. Genes related to lipid metabolism, nucleotide metabolism and protection against oxidative stress are enriched exclusively in SOM 3, whereas genes related to energy metabolism, photosynthesis, water transport and genetic information processing are enriched exclusively in SOM 4. Decline of transcripts related to energy metabolism and photosynthesis was also observed in the resurrection species *H. rhodopensis* and *C. plantagineum* as decreasing metabolic activity is a general primary target during dehydration<sup>34–36</sup>.

A gene co-expression network approach revealed a large overlap in the patterns of gene activation and repression in ABA-treated seedlings and drying mature leaves (Fig. 2). Young seedlings of *X. viscosa* are desiccation sensitive and obtain DT only later in development, but can be made tolerant by application of ABA (Supplementary Fig. 3). Thus, this overlap points to similarities between seed- and vegetative-DT<sup>37</sup>. In both systems, metabolism is shut down in an orderly manner and protection mechanisms activated<sup>30,38,39</sup>. Moreover, upon dehydration, both in seeds and in poikilochlorophyllous resurrection species, chloroplasts are disassembled and chlorophyll degraded. All these (ABA-controlled) events have so far been considered seed-specific but here we show that vegetative-DT in Angiosperms is also regulated by ABA, employing the same genes as in seeds. This supports a hypothesis that DT was reactivated in vegetative tissues, presumably utilizing genes associated with DT in the seeds of the species or family as a 'blueprint', and then later modified to result in species- or family-specific DT mechanisms<sup>27</sup>.

In the context of the gene co-expression network, the genes in SOMs 3 and 4 are more tightly co-regulated than the genes in SOMs 1 and 2, highlighting the importance of a coordinated shut-down of metabolic and cellular processes during dehydration. Accordingly, most of the network's hubs (nodes with high number of connections to other nodes, or high degree) represent genes related to energy metabolism (Supplementary Fig. 4 and Supplementary Data Table).

The network aided the identification of ABA-independent gene expression (Supplementary Fig. 5). Although ABA-independent

signalling pathways mediated by DREB2 and DREB1/CBF have been described<sup>40</sup>, other regulators not described yet may also be acting. Around 20% of the nodes in the network represent genes differentially expressed in response to desiccation but not in response to exogenous ABA. The gene ontology (GO) categories enriched in this group of genes are related to signalling ('consequence of signal transduction' and 'generation of a signal involved in cell-cell signalling') and transcriptional regulation ('positive regulation of transcription regulator activity', 'regulation of sequence-specific DNA binding transcription factor activity' and 'regulation of transcription regulator activity'). These categories suggest the presence of ABA-independent regulators not described yet.

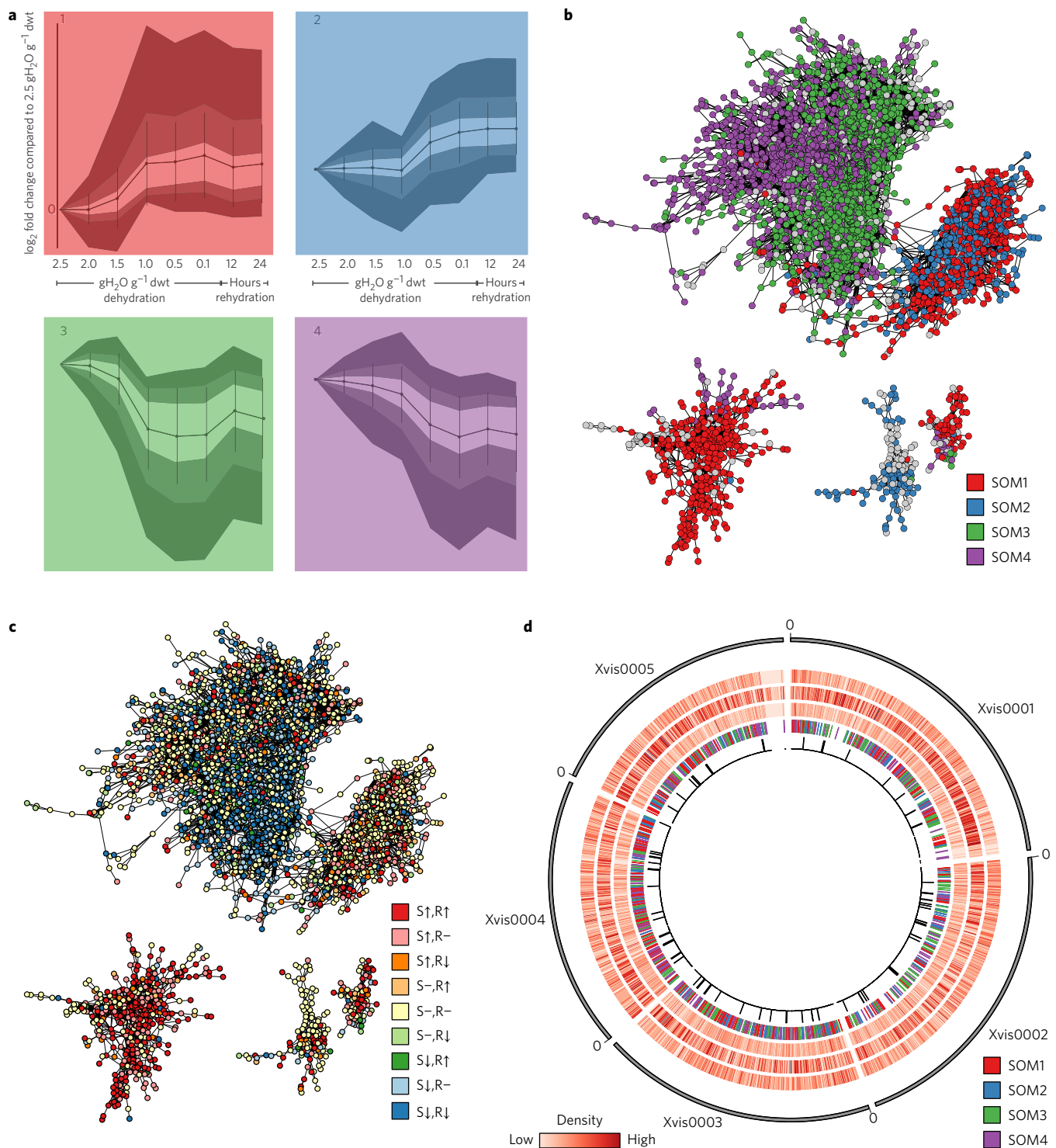
### Clusters of desiccation-associated genes

Anhydrobiosis-related gene islands (ARIDs) are defined as genomic regions hosting compact clusters of genes which are anhydrobiosis-related and accumulate transcripts upon desiccation<sup>39</sup>. In analogy with ARIDs, we define clusters of desiccation-associated genes (CoDAGs) as genomic regions hosting compact clusters of genes associated with desiccation and differentially expressed upon desiccation. These regions are of interest owing to their potential role in the evolution of DT. We performed a genome-wide search for CoDAGs in *X. viscosa* considering the same criteria as for ARIDs<sup>39</sup>: (1) they host a paralogous set of genes (containing at least a pair of genes); (2) their localization in the genome is not necessarily related to that of the potential ancestor of the expanded set of genes; and (3) all genes located within CoDAGs are differentially expressed in response to desiccation. CoDAGs are abundant in the genome of *X. viscosa* (277 CoDAGs and 600 genes involved) and are mainly composed of genes with declining transcript abundance upon dehydration (Fig. 2 and Supplementary Table 3). The expression pattern of the genes in the same CoDAG is not necessarily the same. Genes involved in various environmental interactions form CoDAGs, such as the ABA receptor PYL (Pyrabacin Resistance-Like), ABC transporters, ethylene-responsive transcription factors, intracellular ribonucleases, LEA proteins, pathogenesis-related proteins, polyphenol oxidases and uncharacterized proteins (Supplementary Table 4). That the majority of CoDAGs consist of transcripts that decline upon dehydration appears contradictory, but may be related to mechanisms that suppress growth and development or energy metabolism in vegetative tissues during dehydration, and may trigger or inhibit the induction of the mechanisms that lead to vegetative DT. A less parsimonious possibility is that the genes in these CoDAGs may suppress the expression of DT in vegetative tissues, allowing the induction of DT only at critical levels of water loss, and that these genes have to be suppressed in order to acquire DT.

### LEA proteins

LEA proteins have been implicated in plant embryo development and maturation, response to high salinity, freezing and DT<sup>6,41</sup>. They accumulate in seeds, pollen, fungal spores, yeast cells, nematodes, rotifers and embryos of some crustaceans<sup>6,41</sup>. In resurrection plants, desiccation-induced expression of LEA transcripts is a common response as shown in *H. rhodopensis* and *C. plantagineum*<sup>34–36</sup>.

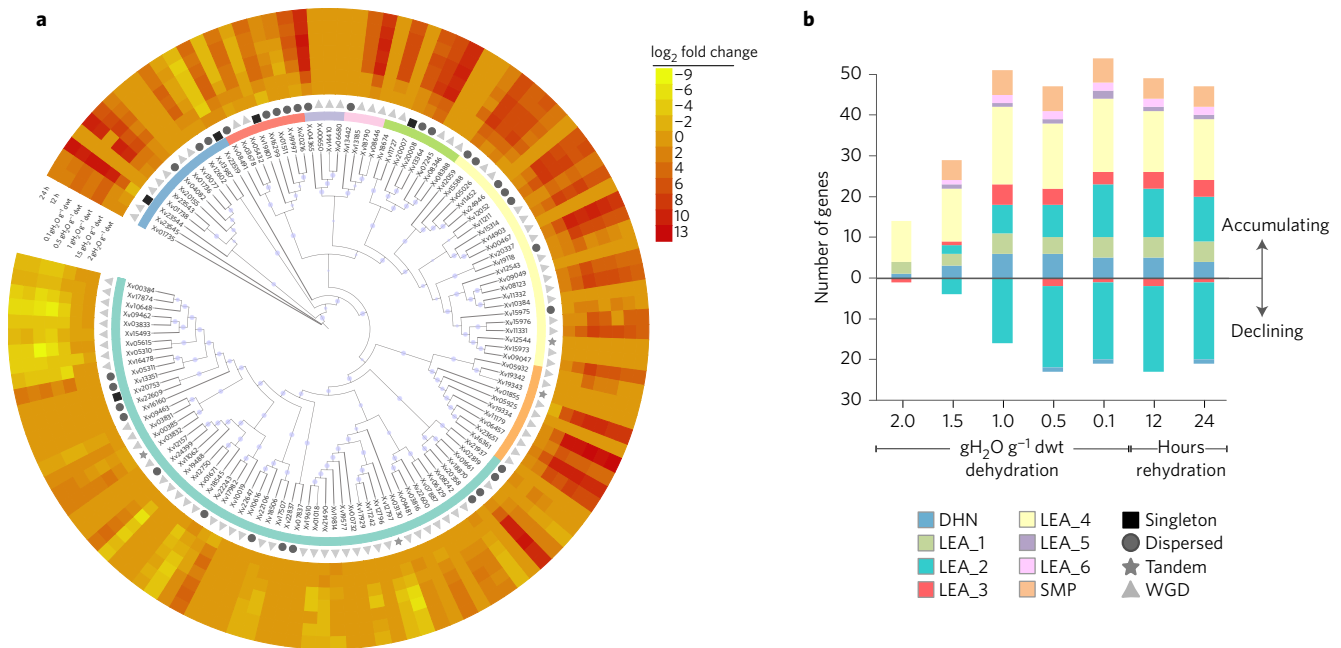
HMM (Hidden Markov Model) profiles provided by PFAM were used to search for LEAs in the genome of *X. viscosa*. We found 126 putative LEA motif-containing proteins divided into eight families (Fig. 3a), of which 90 were differentially expressed during drying and rehydration of adult plants (Fig. 3b). Most of the duplicated gene pairs have dispersed origins, which are paralogs that do not show conserved synteny or are not near each other on chromosomes<sup>42</sup> (Supplementary Table 5). By performing a BLASTP search to assess the similarities between each of the LEA proteins (Supplementary Table 5), we found that, at the protein level, there are no repeated sequences in the genome. Thus, each LEA gene



**Figure 2 | Genomic organization of *X. viscosa*.** **a**, SOMs of gene expression displaying a  $\log_2$  fold change of transcripts relative to full hydration. **b**, The gene co-expression network with nodes coloured according to SOMs. **c**, The co-expression network of differentially expressed genes after ABA application to the shoots (S) and roots (R) of the seedlings. **d**, A circos map with features of the five longest assembled scaffolds of the *X. viscosa* genome. The tracks from outside to inside display the gene density, the TE density, the domain density, the fold change of gene expression to the hydrated state (colour codes as in the SOMs) and the presence of putative CoDAGs.

that we identified is unique, and the high number of LEAs found is due to small (punctual) duplications, and not due to the octoploid nature of the genome. The number of LEAs is significantly higher than identified in the genomes of *B. hygrometrica* (67 LEA genes)<sup>19</sup> and *O. thomaeum* (94 LEA genes; [http://resources.ropetium.org/Oropetium\\_v01/](http://resources.ropetium.org/Oropetium_v01/))<sup>18</sup>, as well as in a phylogenetic cross-section of

25 plant species (Supplementary Table 5). Whole-genome duplications played critical roles in the expansion of LEA families in *X. viscosa* (Fig. 3a). In contrast, in *O. thomaeum* and *B. hygrometrica* dispersed duplications were more evident (Supplementary Fig. 6). A comparative genomic analysis with 25 plant species spanning the land plant phylogeny indicated that two of the eight LEA



**Figure 3 | LEAs transcript expression and accumulation patterns during dehydration and rehydration.** **a**, The maximum likelihood tree showing the phylogenetic relationship between the 126 LEA proteins along with the gene duplication mode, and a heat map showing the log<sub>2</sub> fold change of the transcripts in each point relative to the hydrated state (2.5 gH<sub>2</sub>O g<sup>-1</sup> dwt). The dots on the branches represent bootstrap support values. **b**, The number of LEA transcripts that accumulate and decline relative to the hydrated state. The colours on the chart represent each LEA family shown in the legend. WGD, whole-genome duplication.

families, namely LEA\_4 and LEA\_6, are significantly expanded in the genome of *X. viscosa* (Supplementary Fig. 6 and Supplementary Table 5). These differences might relate to the poikilochlorophyllous nature of *X. viscosa*, in contrast to the homoiochlorophyllous species *B. hygrometrica* and *O. thomaeum*. This, in turn, might relate to longevity in the dry state. Although longevity of *X. viscosa* has not been systematically assessed, two other poikilochlorophyllous species of *Xerophyta* have demonstrated longevity. Dried excised leaves of *Xerophyta scabrada* can survive up to 2 years under laboratory conditions<sup>43</sup>. In the only study reporting on longevity of whole plants in the dry state, it was demonstrated that *Xerophyta humilis* retains viability for more than 2 years, whereas the homoiochlorophyllous *Craterostigma wilmsii* lost viability within 3 months<sup>44</sup>. Furthermore, *X. viscosa* (unlike *X. humilis*) is a chasmophyte (Fig. 1a) and thus tolerates not only frequent periods of desiccation, but also extremes of temperature ranging from sub-zero to above +50 °C (ref. 9).

The response to desiccation appears LEA-family specific. Members of dehydrin (DHN), LEA\_1 and LEA\_4 families accumulate transcripts during drying and rehydration. It is likely that members of these LEA families are involved in a rapid primary response to dehydration and as transcripts continue to accumulate these LEA families may guarantee homeostasis in the desiccating cells. Moreover, as LEA proteins decline shortly after full rehydration, they might serve as a readily available nitrogen source to support resumption of photosynthesis and metabolism upon rehydration. Most members of LEA\_2 family and a few members of DHN and LEA\_3 families significantly decrease transcript abundance during drying and rehydration, suggesting that these LEAs are not involved in the protection mechanism activated by desiccation.

The LEA\_2 family is the most diverse in the genome of *X. viscosa*, with 57 members. A promoter enrichment analysis showed that the LEA\_2 family is activated by a more diverse range of transcription factors than other LEA families (Supplementary Fig. 6). The ABI5 motif is prominent in the promoters of the LEA\_4 family-members. The ABI5 transcription factor is mostly commonly associated with the regulation of seed maturation, and regulates

transcription of LEA genes and degradation of chlorophyll essential to the establishment of seed longevity (measured by a moderate artificial aging method) in *Medicago truncatula* and *Pisum sativum*<sup>45,46</sup>.

Predicted subcellular localization shows that LEAs can be found in seven subcellular compartments (Supplementary Fig. 6), supporting the diversity of LEA expression in adult leaves. Most LEA families are overrepresented in the nucleus. Seed maturation protein (SMP) family-members are overrepresented in the cytosol. LEA\_2 and LEA\_3 families are overrepresented in the plastids and decrease transcript abundance during drying, supporting their poikilochlorophyllous origin.

During desiccation, the most profound change in LEA expression occurs between 1.5 gH<sub>2</sub>O g<sup>-1</sup> dwt and 1.0 gH<sub>2</sub>O g<sup>-1</sup> dwt, with a 43% increase in LEA expression which was maintained during further desiccation and rehydration (Fig. 3). Thus, these LEAs are largely represented by the early responses of SOMs 1 and 3 (Fig. 2). These observations indicate an activation of regulatory pathways essential for the desiccation response around 1.0 gH<sub>2</sub>O g<sup>-1</sup> dwt (40% RWC). This activation is likely caused by a regulatory 'switch' between the dehydration response (to initial water loss) and cellular preparation response (protection) for the desiccated state.

### The ABI3 regulon

ABI3 is a transcription factor that mediates DT in seeds through a highly conserved gene regulatory network<sup>37,47–49</sup>. *X. viscosa* has two structural orthologues of ABI3. Due to different gene contexts (different neighbouring genes), these paralogs are not considered alleles of the same gene. They are expressed, but their expression did not change significantly in response to exogenous ABA or drying and rehydration of adult plants. We identified in *X. viscosa* 139 structural orthologues of the 98 target genes of the ABI3 regulon from *Arabidopsis thaliana*<sup>47</sup>. The majority displayed an increase in transcript abundance in response to exogenous ABA (62% in shoots of seedlings) and to drying (45% in SOM 1). Fifty of these structural orthologues were located in our co-expression network. Although the nodes representing these genes are not directly connected, they share first neighbours, forming tightly connected

sub-networks (Supplementary Fig. 5). The GO categories enriched in these genes are primarily related to energy metabolism ('cellular carbohydrate metabolic process', 'generation of precursor metabolites and energy') and photosynthesis ('plastid organization', 'regulation of photosynthesis', 'stomatal complex morphogenesis' and 'cofactor metabolic process'). Structural orthologues of seed-specific members of the ABI3 regulon were also found in our analysis, such as embryonic proteins DC-8, oleosins, seed maturation proteins and 1-Cys peroxiredoxin. Overall, these analyses shed light on the conserved mechanisms of the gene regulatory network orchestrated by ABI3 in seeds and vegetative tissues of resurrection plants.

### Autophagy and senescence

Autophagy is a catabolic cellular process that attempts to restore homeostasis during severe stress<sup>38</sup>. It involves the recycling of nutrients and removal of damaged and potentially harmful cellular material<sup>38</sup>. In resurrection plants, effective regulation of autophagy aids survival under extreme stress conditions<sup>38</sup>. Although drought-induced leaf senescence is thought to be an efficient strategy for reducing transpiration and allowing remobilization of water and nutrients, it does not occur in resurrection plants in the tissues undergoing induction of DT<sup>38,50</sup>. It is intriguing to investigate how resurrection plants regulate autophagy and avoid senescence during dehydration. On the basis of sequence similarity to proteins predicted to be involved in autophagy and senescence in the resurrection grass *Tripogon loliiformis*<sup>38</sup>, we found a number of similar proteins in *X. viscosa* (Supplementary Table 6). *X. viscosa* homologues of *A. thaliana* AUTOPHAGY (e.g. *ATG2*, *ATG8*, *ATG9*, *ATG12* and *ATG18*), known to delay senescence in *A. thaliana*<sup>51</sup>, cluster in SOMs 1 and 2, whereas *X. viscosa* homologues of *A. thaliana* SENESCENCE-ASSOCIATED GENE and SENESCENCE-ASSOCIATED RECEPTOR-LIKE KINASE, known to promote senescence in *A. thaliana*<sup>51</sup>, cluster in SOM4. Overall, both in *X. viscosa* and in *T. loliiformis*<sup>38</sup> transcripts associated with senescence and pro-apoptosis exhibit a reduction in accumulation, whereas transcripts associated with the delay of senescence and anti-apoptosis accumulated during drying. Despite different strategies concerning chlorophyll retention during drying (*X. viscosa* is poikilochlorophyllous and *T. loliiformis* is homoiochlorophyllous), processes related to autophagy and senescence appear to operate in similar ways in *X. viscosa* and *T. loliiformis*.

Stress signalling in the endoplasmic reticulum (ER) is transduced through the unfolded protein response (UPR) pathway<sup>52</sup>. The UPR activates the ER-located molecular chaperone binding protein BiP, that assists in folding newly synthesized proteins, acts in ER stress signalling and prevents senescence<sup>50,52</sup>. We identified ten putative BiP genes in *X. viscosa*. Most of them clustered in SOM 2, indicating activation when the stress becomes more severe. Thus, apparently, *X. viscosa* engages a mechanism to prevent activation of the UPR cell death response during severe dehydration.

### Conclusions

Combining genome sequence information with genome-wide gene expression data is a powerful approach to gain insight into the 'footprint' of DT in a resurrection species. Among the currently known resurrection plant species, *X. viscosa* is probably the most resilient<sup>9</sup>. Vegetative DT was reactivated in at least 13 families of Angiosperm plant species<sup>27</sup> as adaptation to environmental demands, and we propose that the basal DT mechanism was further modified according to specific environmental and habitat (niche) requirements. Reactivation of vegetative DT must have been based on the presence of genes associated with DT in reproductive structures, such as seeds, and hence the genomic information for DT was redirected towards vegetative tissues. Further genome modifications to deliver DT may have included polyploidization, as in the African Velloziaceae, during the establishment of this family on the

African continent. Poikilochlorophyllous resurrection species have a clear seed imprint in their phenotype. For example, the degradation of chlorophyll during the maturation stage is common among orthodox seeds, presumably to avoid the generation of reactive oxygen species during subsequent long periods of storage in the dry state. During dismantling of photosynthetic machinery, *X. viscosa* engages anti-senescence mechanisms, giving evidence that in this species vegetative DT is uncoupled from common drought-induced senescence, again similar to seeds. In conjunction, the bZIP transcription factor ABI5 is strongly associated with the acquisition of longevity in orthodox seeds, assessed by artificial aging, as well as expression of several LEAs<sup>45</sup>. Here we show that ABI5 may be a regulator of expression of the LEA\_4 family which, consequently, may be an important factor in the longevity of *X. viscosa* in the dry state. Furthermore, we identified two structural orthologues of ABI3, a major regulator of seed maturation and DT along with the majority of the ABI3 regulon expressed in leaves. Again, this exemplifies the strong 'seed character' of vegetative DT in this species.

The CoDAGs that are downregulated upon dehydration may represent genes which are related to mechanisms that suppress growth and development or energy metabolism in vegetative tissues and may trigger or inhibit the induction of the mechanisms that lead to vegetative DT. The genes in these CoDAGs may also suppress the expression of DT in vegetative tissues, allowing the induction of DT only at critical levels of water loss. Their down regulation may thus be part of the reactivated mechanism that evolved to deliver DT in vegetative tissues. Further study of possibly higher-order regulation of this set of genes may identify targets for modification of crops towards DT.

### Methods

**DNA sequencing, *de novo* assembly and validation.** The genome of *X. viscosa* was sequenced using raw paired-end Illumina (~77× coverage) and PacBio (~58× coverage) technologies. Reads originating for contaminants (insects, bacterial, fungal and human), as well as chloroplast and mitochondrial genomes, were removed from all sequence data prior to assembly. Illumina reads were error corrected using Lighter<sup>12</sup> and assembled using SparseAssembler<sup>13</sup>. PacBio reads were assembled using the diploid aware FALCON assembler<sup>11</sup>. A hybrid assembly was produced with DBG2OLC<sup>14</sup> and the contigs were reordered and connected into scaffolds using SSPACE-LongRead<sup>15</sup>. The assembly was polished using Sparc<sup>13</sup> and Pilon<sup>53</sup>. PBjelly2<sup>54</sup> was used for gap closure and genome improvement. Alignments due to gene duplication and repeats were filtered out using the delta-filter utility of the MUMmer package<sup>55</sup>. The assembly was validated by mapping the available RNA and DNA libraries to the genome with Bowtie2<sup>56</sup> and Blasr. Assembly statistics were calculated using QUAST<sup>57</sup>. Gene space completeness was measured by using BUSCO (Benchmarking Universal Single-Copy Orthologs). The genome assembly and validation pipeline is summarized in Supplementary Data Figure.

**Annotation.** The *ab initio* predictors AUGUSTUS<sup>58</sup> and SNAP<sup>59</sup> were trained on the transcriptome data using BRAKER1<sup>60</sup>. The MAKER2 annotation pipeline was applied for gene prediction and repeat annotation. Predicted genes were functionally annotated by a consensus approach using InterProScan<sup>23</sup>, Gene Ontology<sup>61</sup>, Kyoto Encyclopedia of Genes and Genomes (KEGG), Swiss-Prot<sup>22</sup>, Translated EMBL Nucleotide Sequence Data Library (TrEMBL<sup>22</sup>) and BLAST2GO<sup>62</sup>.

Repeatmodeller<sup>63</sup> was applied to build a *de novo* repeat library and identify ribosomal and small nuclear RNAs. Repetitive sequences in the assembly were soft masked using RepeatMasker<sup>63</sup>. Transfer RNAs (tRNAs) were annotated using tRNA-scan-SE<sup>64</sup>, microRNAs were predicted using BLAST and INFERNAL<sup>65</sup> against the RFAM database<sup>66</sup>.

Single nucleotide polymorphisms and insertions and deletions (INDELs) were called relative to the genome using FreeBayes<sup>67</sup>. Detected polymorphisms from short-read alignments were discarded when the quality was below 20.

**OG inference.** Orthology between 15 plant species (*Amborella tricopoda*, *A. thaliana*, *Elaeis guineensis*, *E. tef*, *Hordeum vulgare*, *Musa acuminata*, *O. thomaeum*, *Oryza sativa*, *Phalaenopsis equestris*, *Phoenix dactylifera*, *Physcomitrella patens*, *Spirodela polyrrhiza*, *X. viscosa*, *Zea mays* and *Zostera marina*) was determined with OrthoFinder<sup>68</sup>. The proteomes were downloaded from Phytozome V11.0 (<https://phytozome.jgi.doe.gov/pz/portal.html>).

Gene family expansions and contractions in comparison with other sequenced genomes were calculated for all OGs (excluding singletons). The z-score was calculated for each OG; those with z-score  $\geq 2$  represent significantly expanded gene

families and those with  $z$ -score  $\leq -2$  represent significantly contracted gene families. The number of genes per species for each OG was transformed into a matrix of  $z$ -scores. The  $z$ -score profile was hierarchically clustered using Pearson correlation as a distance measure. The functional annotation of each OG was predicted based on sequence similarity to entries in the Swiss-Prot database<sup>22</sup> and InterPro protein families database<sup>23</sup> where more than 50% of proteins in the family share the same protein annotation.

The ancestral gene content at key nodes in the phylogeny of the 15 plant species was reconstructed by Wagner parsimony with a 1.2 gain penalty using COUNT<sup>69</sup>.

**Cytogenetics.** Chromosome preparations were performed as described previously<sup>70</sup>. The chromosome preparations were screened under a phase-contrast microscope and late-pachytene cells with little or no cytoplasm, good chromosome spreading and well-differentiated chromatin morphology were selected (Supplementary Fig. 2).

The assessment of DNA content of *X. viscosa* was performed by Iribov (Enkhuizen, The Netherlands). Genome size was measured by flow cytometry on a CA-II cell analyser (Partec). Tetraploid *Brassica oleracea* was used as a reference for which the DNA content was calculated as 2.8 pg.

**DT in *X. viscosa* seedlings.** Seeds were harvested from *X. viscosa* plants collected in the Buffelskloof Nature Reserve and cultivated under glasshouse conditions at the University of Cape Town (South Africa). Seeds were sown on two layers of blue filter paper (Blue Blotter Paper, Anchor Paper Company) and 50 ml of distilled water and incubated in germination cabinets with constant white light at 25 °C for ~10 days.

To assess DT in *X. viscosa* seedlings, we used seedlings with the first leaf between 1 and 2 mm length. Three replicates of 20 seedlings were incubated for 3 days in 6-cm Petri dishes containing 1.3 ml of solutions of distilled water (control) or 50  $\mu$ M ABA on two sheets of white filter paper (grade 3hw, Biolab Products, Sartorius Stedim Biotec) in the dark at 20 °C. After incubation, seedlings were rinsed in distilled water, transferred to new Petri dishes with one sheet of white filter paper, dried and rehydrated. Drying was achieved by drying under an atmosphere of 32% relative humidity for 3 days at 20 °C, resulting in final water contents as low as 0.126 gH<sub>2</sub>O g<sup>-1</sup> dwt. Water contents were assessed gravimetrically by determination of the fresh weight and subsequently dry weight (dwt) after 17 h at 105 °C. After dehydration, seedlings were pre-humidified in humid air (100% relative humidity) for 24 h at 22 °C and rehydrated in H<sub>2</sub>O at 22 °C on a Copenhagen Table under a 12/12 h dark/light regime. DT was quantified as percentage of seedlings that exhibited growth resumption with both green leaves and development of a root system (Supplementary Fig. 3).

For RNA extraction, seedlings treated with ABA and non-treated were dissected in three parts: seed coat, shoots and roots. The seed coat was discarded and shoots and roots were used for RNA extraction. Total RNA was extracted following a modified hot borate protocol<sup>71</sup>.

The samples were sequenced (150 nt, single-end) with Illumina HighSeq 2,000 and 55.2 Gb were obtained after trimming (12 libraries). A *de novo* transcriptome assembly was constructed using Trinity<sup>72</sup>. The raw cDNA sequencing (RNA-seq) data have been deposited in NCBI SRA database under accession SRS1071017.

**DT in adult plants.** A population of 80 clonally propagated 5-month-old *X. viscosa* plants were grown in Premier Pro-Mix BX (Premier Tech Horticulture Pennsylvania, USA) under the following greenhouse conditions: 16-hour days at 27 °C during the day and 18 °C at night. Prior to the initiation of the drying period, well-watered plants were covered with a plastic bag overnight to maximize the hydrated water contents prior to sampling. Each of the 80 plants were treated as individual biological reps for sampling. Dehydration was achieved by withholding water from the pots. Young green leaf tissue was harvested from 16 individual plants which were sampled daily to monitor their total water content (TWC). A portion of the tissue from each of the 16 plants was flash-frozen in liquid nitrogen and stored at -80 °C. Turgid total water fraction (TWC), expressed as grams of water per gram of DWT, was determined gravimetrically using the remainder of the sample and calculated as  $TWC = (FWT - DWT)/DWT$ . DWT of each sample was determined gravimetrically after exposure to 70 °C in a convection oven until constant weight, which took 4 h<sup>73</sup>. Plants were rehydrated by fully saturating the soil and aerially by misting with water.

Triple samples were selected with TWCs of 2.5 (fully hydrated TWC), 2, 1.5, 1, 0.5 and 0.1 gH<sub>2</sub>O g<sup>-1</sup> dwt and rehydrated tissues at 12 and 24 h post-re-watering were selected for RNA extraction. RNA was extracted using the RNeasy (Qiagen, Hilden, Germany) kit with the RLC buffer following the manufacturer's recommended protocol. The RNA isolates were treated with DNaseI and cleaned using the DNA-free RNA Kit (Zymo Technologies, Irvine, CA). RNA quality was assessed by use of a fragment analyser (Advanced Analytical Technologies, Ankeny, IA) and concentration determined by use of a Nanodrop Spectrophotometer (ThermoFisher, Waltham, MA).

RNA libraries were created and individually bar-coded from 2.7  $\mu$ g of template total RNA utilizing the TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA) as described in the manufacturer's recommended protocol. Libraries were pooled in groups of six and sequenced (six samples per lane) on an Illumina HiSeq 2,500 ultra-high-throughput DNA sequencing platform (Illumina, San Diego, CA) at the DNA

Core facility at the University of Missouri, Columbia, MI, USA (<http://dnacore.missouri.edu/HiSeq.html>).

**Gene expression analyses.** Gene expression was calculated using TopHat and Cufflinks from the Tuxedo suite<sup>74</sup>. Differential expression was then computed using Cuffdiff, also from the Tuxedo suite<sup>74</sup>. The expression of 12 genes was analysed by quantitative PCR (qPCR) in shoots in order to verify the accuracy of the data. Candidate reference genes were chosen based on the RNA-Seq data applying a cut-off of  $\geq 0.9$  for  $q$ -value. These genes were further checked using qBase+ (Biogazelle) applying a cut-off of  $\leq 0.5$  for M-value (gene stability value) and  $\leq 0.15$  for CV-value (coefficient of variation), generating a shortlist of seven reference genes. The three reference genes with most stable expression in the qPCR were used for expression normalization of the target genes. Both RNA-Seq and qPCR showed comparable trends.

The annotated genome was used for an over-representation analysis to recover over-represented biological processes (using Benjamini & Hochberg False Discovery Rate correction,  $P$  value  $\leq 0.05$ ) based on gene ontologies using the plugin BiNGO for Cytoscape. InterProScan and protein BLAST (using *A. thaliana*, *O. sativa* and *Z. mays* as a reference) were used to assign GO-terms to the total set of annotated genes. The whole annotation was used as reference set. The term's semantic distance with respect to other semantically close terms ('Dispensability') was calculated using the online tool ReviGO (<http://revigo.irb.hr/>) and used to remove redundant terms applying a cut-off of  $\leq 0.05$ .

SOMs were calculated using GeneMaths software (version 2.1, Applied Maths BVBA, Sint-Martens-Latem, Belgium) by importing gene expression data and mapping them into four groups (2  $\times$  2 node format) that provided optimal representation of gene expression patterns in a small number of independent bins.

Orthologues were defined as hits with lowest Expect value (E-value), with a threshold of  $\leq 10^{-10}$ . Multiple hits were considered orthologues when the difference between their E-values and the lowest hit's E-value was smaller than  $10^{-10}$ .

**Identification of LEA proteins.** To identify members of LEA protein families, all HMMs from the PFAM database<sup>75</sup> (<http://pfam.xfam.org>) were uploaded and used to generate HMM profile matrix using the program hmmbuild of the HMMER3.0 package<sup>76</sup>. The HMM profiles were used to identify members of 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) in the genome of *X. viscosa* using the program hmscan. All proteins with significant hits (E-value  $< 0.01$ ) were collected.

**Expansion of *X. viscosa* LEA families.** We collected protein sequences from 14 species (*Amborella trichopoda*, *A. thaliana*, *B. hygrometrica*, *Brachypodium distachyon*, *Brachypodium stacei*, *Chlamydomonas reinhardtii*, *E. guineensis*, *E. tef*, *H. vulgare*, *M. acuminata*, *O. thomaeum*, *O. sativa*, *Panicum hallii*, *P. equestris*, *P. dactylifera*, *P. patens*, *Selaginella moellendorffii*, *Setaria italica*, *Setaria viridis*, *Sorghum bicolor*, *Sphagnum fallax*, *S. polyrrhiza*, *Z. mays* and *Z. marina*). Protein sequences were retrieved in the same way as described above for *X. viscosa*, using HMM from PFAM and searching with HMMER3.0 program.

LEA family expansions were calculated as described above. The families that presented  $z$ -score  $\geq 2$  were considered significantly expanded.

**Phylogenetic and gene duplication analysis.** Multiple sequence alignment of the 126 putative LEAs was generated using MAFFT version 7 (<http://mafft.cbrc.jp/alignment/server/>) with default parameters. The test for the best amino acid substitution model and maximum likelihood tree were performed by IQ-Tree with default parameters, and the model WAG+I+G4 was selected according to Bayesian information criterion and 1,000 ultra-rapid bootstraps. The phylogenetic tree was edited and displayed using the online application iTOL v3.0 (Interactive Tree of Life, <http://itol.embl.de/>).

For the comparative analysis with 24 species, multiple sequence alignments of the significantly expanded LEA\_4 and LEA\_6 families were performed using MAFFT with default parameters and trimmed with the online tool TrimAl v1.3 (<http://trimal.cgenomics.org/>) implemented in Phylemon v2.0 with settings: conserving at least 50% of the positions and gap threshold 0.7. Maximum likelihood analysis and amino acid substitution model tests were performed with IQ-Tree with default parameters and 1,000 ultra-rapid bootstraps. The model JTT+F+G4 was selected for LEA\_4 family and JTTDCMut+G4 for LEA\_6 according to Bayesian information criterion. iTOL v3.0 was used for editing and displaying the phylogenetic trees. We used the 'duplicate\_gene\_classifier' command in MCScanX package to determine the origins of duplicate genes for the LEA families in *X. viscosa* genome.

**Promoter motif analysis.** The MEME program available within the web tool MEME-Lab (<http://meme-suite.org/tools/meme>) was used for discovery of motifs at the 1 kb region upstream of each LEA family. The parameters used were: occurrence per sequence 1, motif count 10, motif width between 6 wide and 15 wide (inclusive). Subsequently, the motifs found were compared with annotated motifs in the Jasp Core 2016 Plants database by Tomtom, which is part of the MEME suite platform, with default parameters.



**Protein structure and subcellular localization.** We used PSI program (Plant Subcellular-localization Integrative predictor, <http://bis.zju.edu.cn/psi/>) to predict the subcellular localization of LEAs of *X. viscosa* proteins. PSI was chosen because it integrates the main predictors such as Cello, mPloc, Predotar, mitoProt, MultiLoc, TargetP, Wolf PSORT, subcellPredict, iPSort, Yloc, PTS1. The grand average of hydropathy (GRAVY) value for protein sequences calculation was performed with GRAVY calculator (<http://www.gravy-calculator.de/>).

**Gene co-expression network.** Genes differentially expressed in at least one comparison between a time-point in the dehydration/rehydration curve and the hydrated state were used. Pearson correlation coefficients between all pairs of differentially expressed genes were calculated. A table with correlation coefficient values was exported to Cytoscape v.2.8.2 and correlation coefficients above a threshold of 0.96 (determined according to ref. 77) were used to filter the connections and determine the edges between nodes in the network. The resulting network was displayed with a yGraph Organic layout. The Cytoscape built-in app NetworkAnalyzer was used to compute network parameters.

**Data availability.** Sequence and transcriptome data are available from the National Center for Biotechnology Information (NCBI) under BioSample ID [PRJNA291133](https://www.ncbi.nlm.nih.gov/biosample/PRJNA291133) and BioProject ID [SAMN03940242](https://www.ncbi.nlm.nih.gov/bioproject/SAMN03940242).

Received 18 October 2016; accepted 20 February 2017;  
published 27 March 2017

## References

- Kermode, A. R. Approaches to elucidate the basis of desiccation-tolerance in seeds. *Seed Sci. Res.* **7**, 75–95 (1997).
- Black, M. & Pritchard, H. W. (eds) in *Desiccation and Survival in Plants* 207–237 (CABI, 2002); <http://www.cabi.org/cabebooks/ebook/20023069464>
- Gaff, D. F. Desiccation-tolerant flowering plants in Southern Africa. *Science* **174**, 1033–1034 (1971).
- Porembski, S. in *Plant Desiccation Tolerance* Vol. 215 (eds Lüttge, U., Beck, E. & Bartels, D.) 139–156 (Springer, 2011).
- Jönsson, K. I. & Järemo, J. A model on the evolution of cryptobiosis. *Ann. Zool. Fennici.* **40**, 331–340 (2003).
- Alpert, P. Constraints of tolerance: why are desiccation-tolerant organisms so small or rare? *J. Exp. Biol.* **209**, 1575–1584 (2006).
- Oliver, M. J., Tuba, Z. & Mishler, B. D. The evolution of vegetative desiccation tolerance in land plants. *Plant Ecol.* **151**, 85–100 (2000).
- Oliver, M. J., Velten, J. & Mishler, B. D. Desiccation tolerance in bryophytes: a reflection of the primitive strategy for plant survival in dehydrating habitats? *Integr. Comp. Biol.* **45**, 788–799 (2005).
- Farrant, J. M. *et al.* A molecular physiological review of vegetative desiccation tolerance in the resurrection plant *Xerophyta viscosa* (Baker). *Planta* **242**, 407–426 (2015).
- Dai, A. Increasing drought under global warming in observations and models. *Nat. Clim. Change* **3**, 52–58 (2013).
- Chin, C.-S. *et al.* Phased diploid genome assembly with single molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
- Song, L., Florea, L. & Langmead, B. Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biol.* **15**, 509 (2014).
- Ye, C. *et al.* Exploiting sparseness in de novo genome assembly. *BMC Bioinformatics* **13**, S1 (2012).
- Ye, C., Hill, C., Wu, S., Ruan, J. & Ma, Z. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. *Sci. Rep.* **6**, 31900 (2016).
- Boetzer, M. *et al.* SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15**, 211 (2014).
- Yoshida, K. *et al.* The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife* **2**, e00731 (2013).
- de Melo, N. F. *et al.* Cytogenetics and cytotaxonomy of velloziaceae. *Plant Syst. Evol.* **204**, 257–273 (1997).
- VanBuren, R. *et al.* Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. *Nature* **527**, 508–511 (2015).
- Xiao, L. *et al.* The resurrection genome of *Boea hygrometrica*: a blueprint for survival of dehydration. *Proc. Natl Acad. Sci. USA* **112**, 5833–5837 (2015).
- Yasui, Y. *et al.* Draft genome sequence of an inbred line of *Chenopodium quinoa*, an allotetraploid crop with great environmental adaptability and outstanding nutritional properties. *DNA Res.* **23**, 535–546 (2016).
- Šmarda, P. *et al.* Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc. Natl Acad. Sci. USA* **111**, E4096–E4102 (2014).
- Boeckmann, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370 (2003).
- Mitchell, A. *et al.* The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43**, D213–D221 (2015).
- Wilson, G. A. *et al.* Orphans as taxonomically restricted and ecologically important genes. *Microbiology* **151**, 2499–2501 (2005).
- Hilbricht, T. *et al.* Retrotransposons and siRNA have a role in the evolution of desiccation tolerance leading to resurrection of the plant *Craterostigma plantagineum*. *New Phytol.* **179**, 877–887 (2008).
- Cannarozzi, G. *et al.* Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (*Eragrostis tef*). *BMC Genomics* **15**, 581 (2014).
- Gaff, D. F. & Oliver, M. J. The evolution of desiccation tolerance in angiosperm plants: a rare yet common phenomenon. *Funct. Plant Biol.* **40**, 315–328 (2013).
- Mundree, S. G. & Farrant, J. M. in *Plant Tolerance to Abiotic Stress in Agriculture: Role of Genetic Engineering* (eds Cherry, J. H., Locy, R. D. & Rychter, A.) 201–222 (Springer, 2000).
- Gaff, D. F. & Loveys, B. Abscisic acid levels in drying plants of a resurrection grass. *Trans. Malaysian Soc. Plant Physiol.* **3**, 286–287 (1993).
- Farrant, J. M., Cooper, K., Dace, H. J. W., Bentley, J. & Hilgart, A. in *Plant Stress Physiology* (ed Shabala, S.) 217–252 (CAB International, 2016).
- Bewley, J. D. Physiological aspects of desiccation tolerance. *Annu. Rev. Plant Physiol.* **30**, 195–238 (1979).
- Csintalan, Z., Tuba, Z., Lichtenthaler, H. K. & Grace, J. Reconstitution of photosynthesis upon rehydration in the desiccated leaves of the poikilochlorophyllous shrub *Xerophyta scabrida* at elevated CO<sub>2</sub>. *J. Plant Physiol.* **148**, 345–350 (1996).
- Tuba, Z., Protor, M. C. F. & Csintalan, Z. Ecophysiological responses of homoiochlorophyllous and poikilochlorophyllous desiccation tolerant plants: a comparison and an ecological perspective. *Plant Growth Regul.* **24**, 211–217 (1998).
- Gechev, T. S. *et al.* Molecular mechanisms of desiccation tolerance in the resurrection glacial relic *Haberlea rhodopensis*. *Cell. Mol. Life Sci.* **70**, 689–709 (2013).
- Dinakar, C. & Bartels, D. Desiccation tolerance in resurrection plants: new insights from transcriptome, proteome and metabolome analysis. *Front. Plant Sci.* **4**, 482 (2013).
- Rodriguez, M. C. S. *et al.* Transcriptomes of the desiccation-tolerant resurrection plant *Craterostigma plantagineum*. *Plant J.* **63**, 212–228 (2010).
- Costa, M. C. D. *et al.* Key genes involved in desiccation tolerance and dormancy across life forms. *Plant Sci.* **251**, 162–168 (2016).
- Williams, B. *et al.* Trehalose accumulation triggers autophagy during plant desiccation. *PLoS Genet.* **11**, 1–17 (2015).
- Challabathula, D., Puthur, J. T. & Bartels, D. Surviving metabolic arrest: photosynthesis during desiccation and rehydration in resurrection plants. *Ann. NY Acad. Sci.* **1365**, 89–99 (2015).
- Todaka, D., Shinozaki, K. & Yamaguchi-Shinozaki, K. Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. *Front. Plant Sci.* **6**, 84 (2015).
- Tunnacliffe, A. & Wise, M. J. The continuing conundrum of the LEA proteins. *Naturwissenschaften* **94**, 791–812 (2007).
- Wang, Y. *et al.* MCSscan: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
- Tuba, Z., Lichtenthaler, H. K., Maroti, I. & Csintalan, Z. Resynthesis of thylakoids and functional chloroplasts in the desiccated leaves of the poikilochlorophyllous plant *Xerophyta scabrida* upon rehydration. *J. Plant Physiol.* **142**, 742–748 (1993).
- Bajic, J. *Exploring the longevity of dry Craterostigma wilmsii (homoiochlorophyllous) and Xerophyta humilis (poikilochlorophyllous) under simulated field conditions.* PhD thesis, Univ. Cape Town (2006).
- Verdier, J. *et al.* A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. *Plant Physiol.* **163**, 757–774 (2013).
- Zinsmeister, J. *et al.* ABI5 is a regulator of seed maturation and longevity in legumes. *Plant Cell* **28**, 2735–2754 (2016).
- Mönke, G. *et al.* Toward the identification and regulation of the *Arabidopsis thaliana* ABI3 regulon. *Nucleic Acids Res.* **40**, 8240–8254 (2012).
- Delahaie, J. *et al.* LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-regulated LEA protein abundance linked to desiccation tolerance. *J. Exp. Bot.* **64**, 4559–4573 (2013).
- Khandelwal, A. *et al.* Role of ABA and ABI3 in desiccation tolerance. *Science* **327**, 546 (2010).
- Griffiths, C. A. *et al.* Drying without senescence in resurrection plants. *Front. Plant Sci.* **5** (2014).
- Li, Z., Peng, J., Wen, X. & Guo, H. Gene network analysis and functional studies of senescence-associated genes reveal novel regulators of *Arabidopsis* leaf senescence. *J. Integr. Plant Biol.* **54**, 526–539 (2012).
- Reis, P. A. A. *et al.* The binding protein BiP attenuates stress-induced cell death in soybean via modulation of the N-rich protein-mediated signaling pathway. *Plant Physiol.* **157**, 1853–1865 (2011).
- Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
- English, A. C. *et al.* Mind the gap: upgrading genomes with pacific biosciences RS long-read sequencing technology. *PLoS ONE* **7**, e47768 (2012).
- Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).

56. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
57. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUASt: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
58. Stanke, M. & Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, 465–467 (2005).
59. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
60. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767–769 (2016).
61. Harris, M. A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–D2261 (2004).
62. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
63. Smit, A. F. A., Hubley, R. & Green, P. *RepeatMasker Open-3.0* (RepeatMasker, 2008); <http://www.repeatmasker.org>
64. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
65. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
66. Nawrocki, E. P. *et al.* Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* **43**, D130–D137 (2015).
67. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at <https://arxiv.org/abs/1207.3907#> (2012).
68. Emms, D. M. & Kelly, S. Orthofinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
69. Csuo, M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**, 1910–1912 (2010).
70. Szinay, D. *et al.* High-resolution chromosome mapping of BACs using multi-colour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6. *Plant J.* **56**, 627–637 (2008).
71. Wan, C.-Y. & Wilkins, T. A. A modified hot borate method significantly enhances the yield of high quality RNA from cotton (*Gossypium hirsutum* L.). *Anal. Biochem.* **223**, 7–12 (1994).
72. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
73. Oliver, M. J. *et al.* A sister group contrast using untargeted global metabolomic analysis delineates the biochemical regulation underlying desiccation tolerance in *Sporobolus stapfianus*. *Plant Cell* **23**, 1231–1248 (2011).
74. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
75. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2015).
76. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
77. Freeman, T. C. *et al.* Construction, visualisation, and clustering of transcription networks from microarray expression data. *PLoS Comput. Biol.* **3**, 2032–2042 (2007).

### Acknowledgements

We thank E. Parker (owner) and J. Burrows (manager) of Buffelskloof Nature Reserve Mphumulanga for allowing collection of *Xerophyta viscosa* plants. We thank all members of the Wageningen Seed Lab for discussions. We thank K. Cooper for invaluable assistance in compiling Fig. 1. M.-C.D.C. received financial support from CNPq–National Council for Scientific and Technological Development (201007/2011-8). M.A.S.A. received financial support from CAPES–Brazilian Federal Agency for Support and Evaluation of Graduate Education (BEX0428/09-04, BEX0857/14-9). J.M.F. acknowledges use of funding supplied by the South African Research Chairs Initiative of the DST and NRF of SA (Grant No 98406).

### Author contributions

M.-C.D.C. and M.A.S.A. wrote the article; M.-C.D.C., H.N., E.J. and M.F.L.D. performed the bioinformatics; J.M. and W.L. contributed to the genome and transcriptome analysis; J.M.J.-G. and M.J.O. performed and analysed the transcriptomics; B.W. and S.G.M. provided the autophagy/anti-senescence dataset and performed blasting; T.H. and E.G.W. M.S. prepared the libraries and performed the PacBio sequencing and initial genome analysis; J.M.F. and H.W.M.H. initiated and coordinated the work and directed preparation of the article.

### Additional information

**Supplementary information** is available for this paper.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to H.W.M.H.

**How to cite this article:** Costa, M.-C. D. *et al.* A footprint of desiccation tolerance in the genome of *Xerophyta viscosa*. *Nat. Plants* **3**, 17038 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Competing interests

The authors declare no competing financial interests.