PNAS
www.pnas.org

# Supplementary Information for

## Shared upregulation and contrasting downregulation of gene expression distinguish desiccation tolerant from intolerant green algae.

Elena L. Peredo[a*], Zoe G. Cardon[a]

[1]To whom correspondence may be addressed. E-mail: elperedo@mbl.edu.

Address: The Ecosystems Center, Marine Biological Laboratory, 7 MBL St., Woods Hole, Massachusetts, 02543, USA. Phone: +1 508496 289 7701

**This PDF file includes:**

Supplementary text
Figs. S1 to S9
References for SI

**Other supplementary materials for this manuscript include the following:**

Dataset S1. DEGs detected in pairwise comparisons among time points in *A. deserticola*, *F. rotunda* and *E. costatus.*
Dataset S2. Enriched GO terms associated with upregulated and downregulated DEGs identified during the time course in *A. deserticola*, *F. rotunda* and *E. costatus.*
Dataset S3. Enriched GO terms in "Late", "Dry" and "Rehydrated" stages of *A. deserticola.* For each stage, the full and REVIGO filtered list of GO terms enriched in total DEGs is provided.
Dataset S4. Enriched GO terms in "Late", "Dry" and "Rehydrated" stages of *F. rotunda.* For each stage, the full and REVIGO filtered list of GO terms enriched in total DEGs is provided.
Dataset S5. Enriched GO terms in "Late", "Dry" and "Rehydrated" stages of *E. costatus.* For each stage, the full and REVIGO filtered list of GO terms enriched in total DEGs is provided.
Dataset S6. Transcription factors and protein kinases differentially expressed during the time course in *A. deserticola*, *F. rotunda* and *E. costatus*.
Dataset S7. Expression patterns in *A. deserticola*, *F. rotunda* and *E. costatus* of genes traditionally associated with desiccation tolerance. (Source Data for Fig. 3).



www.pnas.org/cgi/doi/10.1073/pnas.1906904117

**Supplementary Information Text**

**Materials and Methods.**

**Algal isolates used in this study**

• The three green algal taxa within the Scenedesmaceae (Chlorophyta) in this study are highly related, diverging less than 2% in the sequence of the barcoding gene 18S. A phylogenetic tree is shown in Fig. S1A. Two are unicellular, independently evolved desert microalgae, extremely tolerant to multiple cycles of desiccation and rehydration. The third is a coenobial, aquatic, desiccation intolerant, closely related species (1) (Fig. S1B, C). The phylogenetic relationships among these taxa are described in detail in (2).

- *Acutodesmus deserticola* (L.A.Lewis & Flechtner ex E.Hegewald, C.Bock & Krienitz) E.Hegewald, C.Bock & Krienitz, 2013 [Basionym: *Scenedesmus deserticola* L.A.Lewis & V.R.Flechtner ex E.Hegewald, C.Bock & Krienitz], isolate BCP-SNI-2 from L. Lewis, GenBank accession: AY510462. Desert microalga collected at San Nicolas Island, CA, USA. 33.2˚N119.2˚W. Coll: July 1993, J. Belnap (2).
- *Flechtneria rotunda* Sciuto & L.A.Lewis, 2015 [=*Scenedesmus rotundus* L.A.Lewis & Flechtner, 2004], isolate BCP-SEV3-VF49 from L. Lewis. Genbank accession: AF513373. Desert microalga collected at the Sevilleta LTER, NM, USA. 34˚12’43.5”N 106˚45’28.5”W. Coll: October 1998, L. Lewis (2).
- *Enallax costatus* (Schmidle) Pascher, 1943 [≡*Scenedesmus costatus* Schmidle, 1985; Basionym: *Scenedesmus costatus* Schmidle], isolate CCAP276-31 from the Culture Collection of Algae and Protozoa, GenBank accession: HG514428. Freshwater microalga collected in Lake Tuomiojärvi, Jyväskylä, Finland. 62.2˚N, 25.7˚E. Coll: 1980, Hegewald.

**Culturing conditions**

Non-axenic unialgal stocks are maintained on 1.5% agar slants with a 1:1 mix of Bold’s Basal Medium with micronutrients and Woods Hole Medium (BBM+WH) (3). To analyze the differences in transcriptomic responses of each algal taxon during desiccation-rehydration, we used a common garden experimental design. Algal species were grown in liquid medium. For each species, two independent algal cultures were grown for six weeks in 250 mL Erlenmeyer flasks containing 150 mL of autoclaved BBM+WH medium (Fig. S2A). Each flask was inoculated with algal cells from stocks on slants. Ambient air was bubbled into the culture through a Pasteur pipette (with cotton plug at top) inserted through the middle of a foam stopper used to seal the flask mouth. To sustain vigorous growth, culture medium was replenished weekly by replacing half of the volume (75 mL) with fresh medium. All cultures were grown at 25ºC in a Conviron PGW36DE growth chamber (Conviron, Winnipeg, Canada) with a 12:12 h Light/Dark cycle and 40 µE from mixed metal halide and sodium lamps. All culturing procedures were performed under sterile conditions in a culturing hood. For each flask, algal identity was confirmed by sequencing the 18S barcoding region (primers SSU1/C18G; N18J/C18J) (2).

**Generation of manipulable units (algal dots)**

1 d after culture medium was replenished, 50 mL of six-week-old cultures were transferred to Falcon tubes 2 h after the onset of light. Algal cells were concentrated in a volume of 5 mL, after settling by gravity for one hour under low light (~5 µE) (Fig. S2B). Fresh medium had been added to the cultures 24 h before. All cultures were adjusted to a cell density of 3-5 $x10^7$ algal cells $mL^{-1}$ ($3.2 \pm 0.4$ $x10^7$ algal cells $mL^{-1}$ in *A. deserticola*, $4.9 \pm 1.1$ $x10^7$ algal cells $mL^{-1}$ in *F. rotunda*, and $3.4 \pm 1.2$ $x10^7$ algal cells $mL^{-1}$ *E. costatus)*. To generate manipulable experimental units, referred to from here on as algal dots, 40 µL of concentrated algal cultures were pipetted onto round microscopy covers lips (8 mm diameter German Glass, Electron Microscopy Sciences, Hatfield, PA, US) (Fig. S2C).

**Slow desiccation in dark conditions**

Cover slips were then placed on microscopy slides and introduced into the desiccation chamber of a custom-made drying apparatus (1) (Fig. S2D-F). Humidified air flowed through the desiccation chamber, slowly desiccating the algal dots. A low evaporation rate (<5 µL $h^{-1}$) was achieved by maintaining a high relative humidity in the desiccation chamber. Room air was bubbled through warmed distilled water and passed through a glass condenser held at 0.7°C below room temperature by an Isotemp 3016D water bath (Thermo Fisher Scientific). Humid air was pumped constantly into the desiccation chamber (Fig. S2F).

All desiccation experiments were carried out in near darkness (<1 µE). Desiccation was monitored by recording the diminishing size of the algal dots using an astronomy camera (Acton PI 1 kb Versarray Cooled Camera, Princeton Instruments, Trenton, NJ, with WinView/32 software). Images were captured with an exposure time of 1 min, at an imaging interval of 30 min for 16 h. Evaporation rates were quantified by volumetric estimation of the remnant size of algal dots after 2.5, 5, and 7.5 h in the drying chamber (Fig. S2G). Full evaporation of culturing medium (the "Dry" stage) was achieved after 12-13 h (Fig. S2E and G).

Dry algal dots were rehydrated in near darkness by adding 40 µL of sterile distilled water into the center of each dot and gently mixing by pipetting. To prevent water loss during rehydration, algal dots were transferred to a custom-made wet chamber composed of a holder inside a lidded petri dish containing enough room temperature distilled water to cover the flat dish bottom. Representative images of algal cells under hydrated conditions and after rehydration are shown in Fig. S1BC. Upon rehydration, no external damages were observed in cells of desert species (Fig. S1B) while rehydrated cells of the aquatic species *E. costatus* display extensive detachment of protoplast from cell wall (Fig. S1C, red arrow). In previous research we have observed this damage is is accompanied by high accumulation of reactive oxygen species (ROS) in the aquatic species (see 1). In that previous work, intracellular ROS was quantified in desert and aquatic taxa using $DCFH_2$-DA, a dye sensitive to intracellular ROS. In presence of ROS, $DCFH_2$-DA transforms into the fluorescent species DCF ($\lambda_{excitation}$=488 nm, $\lambda_{emission}$=502-589 nm). We quantified the presence of DCF signal in rehydrated cultures using a plate reader and at the individual cell level using quantitative confocal microscopy. See (1) for details. No resistant changes in morphology consistent with production of resistant autospores were observed in any of the species.

**Oxygen concentration and photophysiology of algal dots**

We monitored the oxygen saturation in fully hydrated algal dots over the experimental drydown period to check that algal respiration in the dark did not turn the culture medium anoxic over time (Fig. S3A). Using all three algal species, for one set of measurements, we replicated

the growth conditions, algal concentration protocol and the 40 µL dot-making procedure described above. However, rather than drying them slowly, we maintained them fully hydrated in a dark, humid environment for 24 h (nearly twice the original experimental period required to achieve a dry stage, Fig. S3B). For a second set of measurements, we ran the same experiment but used cultures that had not be replenished with fresh nutrient solution for 7 days. The algae were thus slightly nutrient stressed.

To prevent evaporation, we placed the algal dots (40 µL) in a sealed wet chamber. Typically, up to 39 µL of culture were recovered at the end of 25 h time course. Continuous measurement of $O_2$ concentration in the dots was performed using a $O_2$ microsensor Unisense Ox (Unisense, Aarhus, Denmark) connected to a Unisense picoammeter PA2000 (Unisense, Aarhus, Denmark). Data ware recorded every 5 minutes. For each algal dot, we also monitored the quantum yield of photosynthesis (Φ PSII) using a Walz Junior PAM chlorophyll fluorometer (Heinz Walz GmbH, Effeltrich, Germany). Chlorophyll measurements were captured by pointing the fiber optic of the Walz PAM fluorometer at the underside of the dot, through the clear bottom of the cover slips. In order to avoid perturbing the cells, we applied the necessary 600 msec saturating flash of light only once per hour to measure Φ PSII. TWee used a custom script to turn on the measuring light 3 s before each saturating flash of light. After each flash, the measuring light was turned off. Φ PSII was calculated using the PC software WinControl V3.29 for Windows 10 (Heinz Walz GmbH, Effeltrich, Germany). An autoclaved medium-only control dot was also prepared with 40 µL of autoclaved culture medium. The $O_2$ concentration of that control dot (lacking algae) remained stable at ~8.7 mg $L^{-1}$, matching that in the room air (dotted line Fig. S3A).All data were graphed using R v.3.1.

Fig. S3 shows data for all species, all measurements. Legends distinguish those dots treated exactly as the transcriptomic dots were treated (labeled “1d”) and those dots made from cultures that were allowed to become slightly nutrient stressed over one week (labeled “7d”). Fully hydrated algal dots held in the dark for 24 hours remained oxic, though the nutrient stressed cultures pulled $O_2$ concentrations down slightly more than the cultures that matched the transcriptomic experiment. Oxygen diffusion from room air into the dots was clearly sufficient to maintain oxic conditions and no evidence of accumulating stress was observed in the measurements of Φ PSII. Φ PSII remained stable throughout the 25 hours of the experimental period (Fig. S3A). Furthermore, after 24 h in darkness, a brief exposure to low light (20 µE) activated photosynthesis in all three species and $O_2$ concentrations began to increase rapidly in the liquid (Fig. S3A).

**Time course sampling**

For each species and replicate, five time points of the time course were used in RNA-seq experiments (Fig. S2H). Those corresponding to the “Hydrated” stage were collected immediately after transferring 40 µL of algal cultures to a round cover glass. “Early” and “Late” desiccation stages corresponded to a volume-loss of 25% and 60% respectively and were collected 2.5 h and 7.5 h into the desiccation time course. The “Dry” stage corresponded to fully flattened dots (Fig. S2E). Dry algal dots remained in the drying chamber 11 additional hours until collection (circa 24 h after the beginning of the experiment). Finally, the “Rehydrated” stage corresponded to algal dots collected one hour after rehydration. These dots retained at least 90% of the initial rehydration volume (Fig. S2G). For each stage, replicate and species, three algal dots were collected and transferred using clean forceps under very low light conditions (<1 µE) to 1.7 mL sterile Eppendorf tubes. Samples were immediately placed at -80°C until RNA extraction.

**RNA extraction**

Eppendorf tubes were flash frozen in liquid Nitrogen. Algal dots were ground using sterile micropestles, homogenized until the algal material and cover glass had a fine powder consistency. Total RNA was extracted from ground algal dots using ZR Plant RNA MiniPrep™ (Zymo, Irvine, USA) followed by RNA Clean & Concentrator™-5 (Zymo, Irvine, USA). Quantity, purity and integrity of each sample were determined using Qubit™ RNA BR Assay Kit for Qubit 2 (ThermoFisher Scientific, Waltham, MA, USA), NanoDrop microvolume spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA) and Agilent RNA 6000 Pico Kit in an Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, United States. RNA samples had $A_{260}/A_{280}$ ratios ~2 and RNA integrity number (RIN) in excess of 8.

**Library construction**

Strand-specific libraries were prepared from total RNA using the Ovation Arabidopsis RNA-seq Systems (NuGEN, San Carlos, CA, USA) following manufacturer instructions. This approach enriches for messenger RNA in the sample without using polyA selection that could affect the recovery of organellar mRNA (4). Nuclear, chloroplast, mitochondrial and bacterial ribosomic RNA (rRNA) was removed during the InDA-C adaptor cleavage step using a combination of commercially available oligos targeting prokaryote and green plant rRNA genes (*Arabidopsis* kit; NuGEN S02070, S02076, R01758, F01278) supplemented with 105 custom-made oligos designed from rRNA sequence data from these algae including Genbank sequences AY510465.1, KJ680140.1, KC145438.1, HQ246446.1.

Library concentrations were determined using a Qubit™ DNA BR Assay Kit for Qubit 2 (ThermoFisher Scientific, Waltham, MA, USA) and Agilent DNA 7500 Kit in an Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, United States). Six libraries per lane were pooled in equimolar concentrations for paired-end multiplexed sequencing (2 x 150 nt) using Illumina NextSeq 500 (Illumina, San Diego, CA, USA) at the W. M. Keck Ecological and Evolutionary Genetics Facility (Marine Biological Laboratory, Woods Hole, MA, USA). A total of 223 million paired-end reads were generated for *F. rotunda*, 301 million for *E. costatus*, and 172 million for *A. deserticola* (Fig. S4A). Prior to *de novo* reference transcriptome assembly and data analysis, raw reads were QC trimmed using Trimmomatic 0.35 (5) (phred33 leading:3 trailing:3 slidingwindow:4:15 minlen:36). Only perfect read pairs were retained (> 95% of each library, see Fig. S4A). Sequencing QC reports were generated using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw data are available at the NCBI's Sequence Read Archive (SRA) database, *de novo* transcriptome assemblies at the Transcriptome Shotgun Assembly (TSA) database, and expression data at Gene Expression Omnibus (GEO) database under the SuperSeries record GSE133354. *A. deserticola*, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA529464 . SRA accessions SRR8794168–SRR8794177, GEO (GSE133353), TSA (GHRQ00000000). *F. rotunda*, PRJNA529457 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA529457 . SRASRR8793708–SRR8793717, GEO (GSE133352), TSA (GHRR00000000). *E. costatus*, PRJNA529437 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA529437 . SRA accessions SRR8793529–SRR8793538, GEO (GSE133350), TSA (GHUV00000000).

***De novo* transcriptome assembly**

For each species, pooled filtered reads were used for *de novo* assembly of reference transcriptomes using Trinity v2.1 (6) (Jaccard clip option, strand-specific, k-mer length of 25).

Each assembly was subjected to a three-step refinement by removal of contigs (a) with <10 accumulated aligned reads across all time points and replicates, (b) corresponding to rRNA and (c) encoding bacterial transcripts. Non-algal transcripts were identified using blastx as implemented in BLAST+ against the nr NCBI database with a cut-off e-value of $10^{-5}$. Taxonomical information for filtering bacterial transcripts was obtained using the flag -outfmt "6 qseqid sseqid pident length mismatch gapopen qstart qend sstart send evalue bitscore staxids sskingdoms sscinames sblastnames stitle". Quality of assemblies was evaluated with standard statistical descriptors (Fig. S4B). Representation of full-length reconstructed CDS was calculated using BLAST+ locally against Uniprot90 SwissProt and nr NCBI databases. Completeness of each reference transcriptome was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0 (7) and the Chlorophyte odb10 pre-release dataset (downloaded June 2019). Additional estimation of completeness, tailored to the Scenesdesmaceae dataset, was performed through the Agalma pipeline (8) by estimating gene occupancy (Fig. S4B and C). In brief, we identified a total of 2,157 orthologous genes in between reference transcriptomes assembled and annotated here (*A. deserticola*, *F. rotunda* and *E. costatus*) and algal and seed plant flagship species with highly curated genomes (*Volvox carteri*, *Chlamydomonas reinhardtii*, *Coccomyxa subellipsoidea*, *Physcomytrella patens* and *Arabiodopsis thaliana*).

**Annotation of reference transcriptomes**

Reference transcriptomes were annotated using BLAST homologies captured against Uniprot (9) and nr NCBI (10) databases (e-value $<10^{-5}$) (Fig. S4B). The BLASTx top-hit species distribution of gene annotation followed that expected from phylogenetic relationships (Fig. S4D). Organelle genes were highly homologous to those in the chloroplast and mitochondria of the closely related species *T. obliquus* (11, 12). Protein sequences were predicted with TransDecoder v. 2.0.1 (http://transdecoder.github.io/ ). PFAM domains were identified with HMMER (13), signal peptides with SignalP v. 4.1 (14), transmembrane regions with TMHMM v. 2.0 (15). All annotations were integrated following Trinotate v 3.0.1 pipeline (https://github.com/Trinotate/Trinotate.github.io/wiki). Functional annotations included KEGG (16), GO (17), eggnog (18), and InterPro (19) (Fig. S4B).

**Quantification of gene expression**

Libraries were individually aligned to their respective reference transcriptomes using Bowtie 2 v.2.2 (20). Transcript abundance was estimated using RSEM v.1.2.28 (21). Expression matrices including read counts matrix, normalized 'transcripts per million transcripts' (TPM), and standardized 'trimmed mean of M-values' (TMM) were generated using Trinity script abundance_estimates_to_matrix.pl. All downstream analysis but WGCNA were performed at the 'gene' level, that is by grouping expression values of different gene isoforms.

**Cross-validation of biological replicates and relationship among samples using Trinity pipeline**

For each species, a $Log_2$ transformed counts-per-million (CPM) matrix was generated from the read count abundance matrix. A data point was considered divergent if expression (at gene level) varied more than 2-fold change (FC) between biological replicates. We used Trinity PtR.pl to calculate Pearson's product-moment correlation and p-values. Following Trinity pipeline, divergence at the replicate level was evaluated with pairwise comparisons of expression values across replicates (Fig. S5). Across stages, similarity among all samples was visualized using Principal Components Analysis (Fig. S5A) and heatmaps (Fig. S5B). Similarity analysis placed

samples from “Dry” and “Rehydrated” stages together and highly divergent from those collected during fully and partially hydrated stages (“Hydrated”, “Early”, “Late”). Biological replicates showed high levels of correlation during the desiccation time course with a Pearson’s r value of >0.91 in *A. deserticola*, >0.95 in *F. rotunda* and >0.86 in *E. costatus*. For all taxa, the highest discrepancy was detected at the “Rehydrated” stage and it was especially noticeable in the desiccation-intolerant species, unable to fully recover upon rehydration (0.62). When DEGs were used to test for effects of replicate; only for *E. costatus* at the “Rehydrated” stage did replicates vary from one another (Fig. 3C).

**Cross-validation of biological replicates and identification of expression modules using Weighted Gene Correlation Network Analysis**

WGCNA, an unsupervised analysis method for detection of co-expressed gene modules based on their overall expression profiles (not only DEGs) was used to quantify the association of transcripts’ expression to external variables. Specifically, we tested the association of transcripts’ expression to variables (1) “replicate” to identify possible batch effects and (2) the different stages of desiccation rehydration during the time course. Analyses were conducted with R package WGCNA (22) on normalized expression data at the isotranscripts level from RNA-seq data. After discarding the presence of outliers, we calculated signed, weighted correlation networks for each species using an appropriate soft-thresholding power. Isotranscripts with high positive co-expression relationships were grouped together by average linkage hierarchical clustering on the topological overlap matrices (TOM) (criteria: manual stepwise module, minimum cluster size= 30). Highly correlated network modules were merged (pairwise correlation >0.75). For each module we calculated the module eigengene (E), summarizing the expression value of all the isotranscripts in a given module. Module membership (eigengene-based connectivity, kME), was calculated by determining the positive or negative Pearson’s correlation ($p < 0.001$) of each gene to the module eigengene. Gene Significance (GS) indicated the correlation between each isotranscript and a given external “trait” (here stage and replicate). For each external trait, highly correlated modules of interest were identified by calculating the correlation coefficients (and p-values) of the module eigengenes and each external trait (Fig. S6A-C). Replicate was rejected as a trait driving the changes in expression in all stages and taxa except for a subset of transcripts in rehydrated samples of *E. costatus*. (Fig. S6C). Isotranscripts in modules significantly correlated to the “Hydrated” stage still displayed correlation with “Early” desiccation and were uncorrelated with “Dry”. In desert species, modules highly correlated with “Dry” stage also correlated with “Rehydrated”. For all species, WGCNA analysis confirmed stage, and not replicate, as the factor affecting the categorization of isotranscripts in distinct modules. (Fig. S6D-E).

**Differential expression analysis**

Differential expression analysis was performed on count data generated with RSEM (21) using DESeq2 (23). A gene was classified as a DEG if the change in its expression exceeded 4-fold ($Log_2$ 2-FC) in pairwise comparisons with a significance level <0.001 (adjusted p-value after Benjamini-Hochberg (BH) correction for multiple testing (23)). In all cases, each stage during the desiccation-rehydration time course was compared against the stage “Hydrated”. Results of pairwise comparisons were represented in volcano plots ($Log_2$ FC vs. $-1*Log_{10}$ adj_p-value, Fig. S7). In total, during the time course, over 1,500 DEGs were identified for each algal species (Dataset S1). For all taxa, the “Dry” stage accounted for the highest number of DEGs (Fig. S7A). One hour after rehydration, “Rehydrated” samples of desert taxa still presented very

different gene expression patterns from “Hydrated” samples (Fig. S7B) and higher similarity with “Dry” samples (Fig. S7C). In the case of *F. rotunda*, no DEGs were detected in pairwise comparison between “Dry” and “Rehydrated” samples. In *E. costatus* most cells are dead or dying, and transcript abundances drop (Fig. 1B).

Expression values of DEGs were visualized in heatmap plots (Fig. 1A). Samples and DEGs were organized by similarity from Euclidean distance matrices and grouped with hclust (complete-linkage method). An automatic partitioning of similarity of dendrograms (60% maximum height) was used to identify clusters of co-expressed DEGs along the desiccation/rehydration time course. DEGs largely fell into three main clusters (DWN, UpM, and UpH). Upregulated DEGs were grouped in the clusters UpM (blue, moderate upregulation) and UpH (orange, highly upregulated), and downregulated DEGs in the cluster DWN (green) (Fig. 1A). For each cluster, we plotted the expression data of all individual genes within the cluster as centered to the mean $Log_2$ (fpkm +1) (fragments per kilobase million). In each plot, dark blue and dark green lines represent the average expression values of DEGs within the cluster for each biological replicate (Figure 1B). The complete list of the DEGs in each group is provided in Dataset S1. Expression data of genes of interest are presented in box plots (Fig. 3) generated with SigmaPlot10 (Systat Software, San Jose, CA).

**Gene Ontology (GO) and Transcription Factors**

GO term assignments were extracted from the annotated reference transcriptome for each species. GO term enrichment tests were performed using GOseq (24). A GO term was considered enriched with a significance level <0.05 of the adjusted p-value after Benjamini-Hochberg (BH) correction for multiple testing (24).

We conducted two complementary GO term enrichment analyses. First, we explored the response to desiccation by identifying those functions enriched in DEGs with overall upregulation during the time course (clusters UpM and UpH in Fig. 1A), and downregulation (Cluster DWN). The results of this analysis are fully detailed in Dataset S2. Figure 1C summarizes the results for the subontology Biological Processes (BP). To reduce possible redundancy in the analysis, significantly enriched GO terms were filtered using a semantic clustering approach as implemented in REVIGO (25) (Dataset S2) using the semantic similarity measure SimRef and following criteria of large similarity (0.9). The database of *Chlamydomonas reinhardtii* was selected as size reference. Results are summarized in scatterplots generated in REVIGO. In the plots (Fig. 1C), GO terms are grouped in space (x,y) according to their overall similarity. The size of the bubble corresponds to the $-1*|Log_{10}adj\_p\ value|$. To ease comparisons across taxa, all terms related to a given process are indicated in the same color. Related GO terms were defined as child and parent GO terms of a given term identified in the same pathway (Fig. S8B,C). These hierarchical relationships among GO terms were visualized in directed acyclic graph generated GOView (26). Overlap among lists of enriched GO terms were identified using Venn diagrams (Fig. 1C, Fig. 2A, Fig. S9) generated with Venny 2.0 http://bioinfogp.cnb.csic.es/tools/venny_old/venny.php.

Secondly, we compared the specific response functions that were identified in common across all three taxa as affected by desiccation. For each species, we conducted GO term enrichment tests using GOseq (24) on DEGs (independent of the direction of their response) identified in “Late” and “Dry” stage. We used REVIGO (25) to remove GO term redundancy (Datasets S3-S5). Venn diagrams indicated 15 GO terms were enriched in all three taxa (Fig. 2A and B). We used the GOplot v.1.0.1 (27) package run in R v.3.1 to summarize and visualize the results from GOseq across all taxa. GOplot quantifies the overall direction of response of each

GO term using a metric known as its Z-score (27). This metric is not related to the standard score but modeled after the "activation Z-score" implemented in the software package Ingenuity Pathway Analysis (IPA) from QIAGEN Bioinformatics (https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis). Z-scores were calculated as $z = \frac{(upregulated - downregulated)}{\sqrt{total\ DEGs}}$ (27). A negative value indicates a downregulated function and a positive, an upregulated function (27). Results are presented in polar graphs generated using circ command in GOplot (27). For each species, we show the expression changes (as $Log_2FC$) of each DEG within a given term (Fig. 2C, outer circle) and the overall behavior of the term (Fig. 2C, inner circle). To avoid redundancy, if a term was enriched in "Late" and "Dry" stages, only data from "Dry" were plotted.

Transcription factors, transcription regulators and protein kinases were identified using the software package iTAK v.1.7 (28) (database 17.09, 167 genomes including multiple Chlorophyte algae) (Dataset S6).

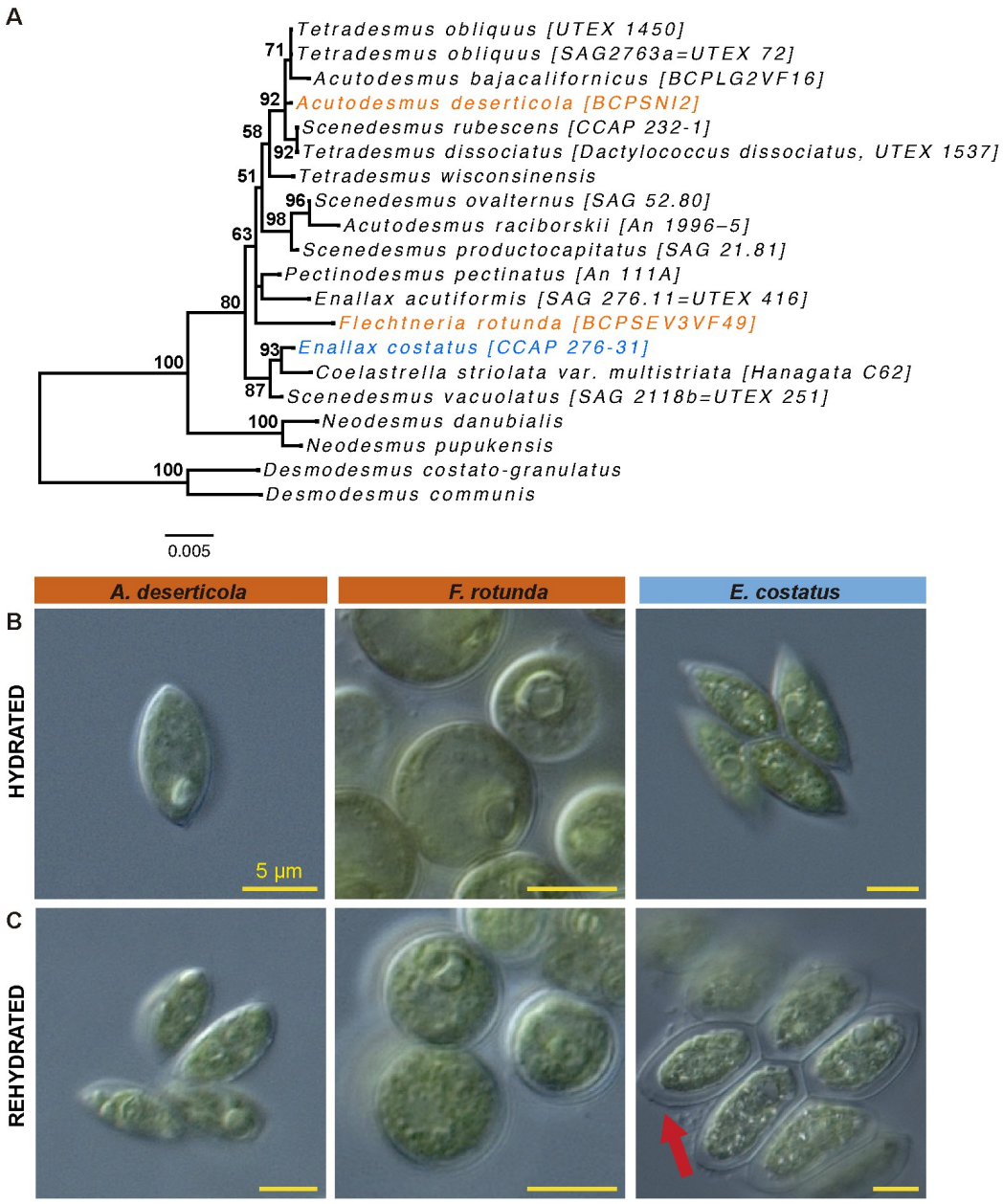


**Fig. S1.**
(A) Maximum likehood (ML) tree depicting the phylogenetic relationships among the green algal species included this study (indicated in color). ML tree was built from aligned 18S gene sequences obtained from GenBank in raxML GUI 1.5 (29) under the ML + rapid bootstrap option (1000 reps) and with GTR GAMMA model of rate heterogeneity. Bootstrap values are indicated for each node. ML Optimization Likelihood: -3554.525780. Sequences included: AF513372, AY510462, AF513373, AB037084, AB012846, AB037089, AB037090, X56103, AJ249515, X81966, AB037092, X91266, AB037094, X74002, X56104, AB037097, X73994, X91265, X91267, and AB037086. (B) Morphology of algal cells under hydrated conditions and (C) upon rehydration. From left to right, images correspond to desert taxa *A. deserticola* and *F. rotunda*, and aquatic *E. costatus*. Red arrow (C, right panel) indicate cell wall detachment. In all images, yellow bars indicate 5 µm. All images were captured using with a 60X objective using Zeiss Axioskop2 optical microscope.

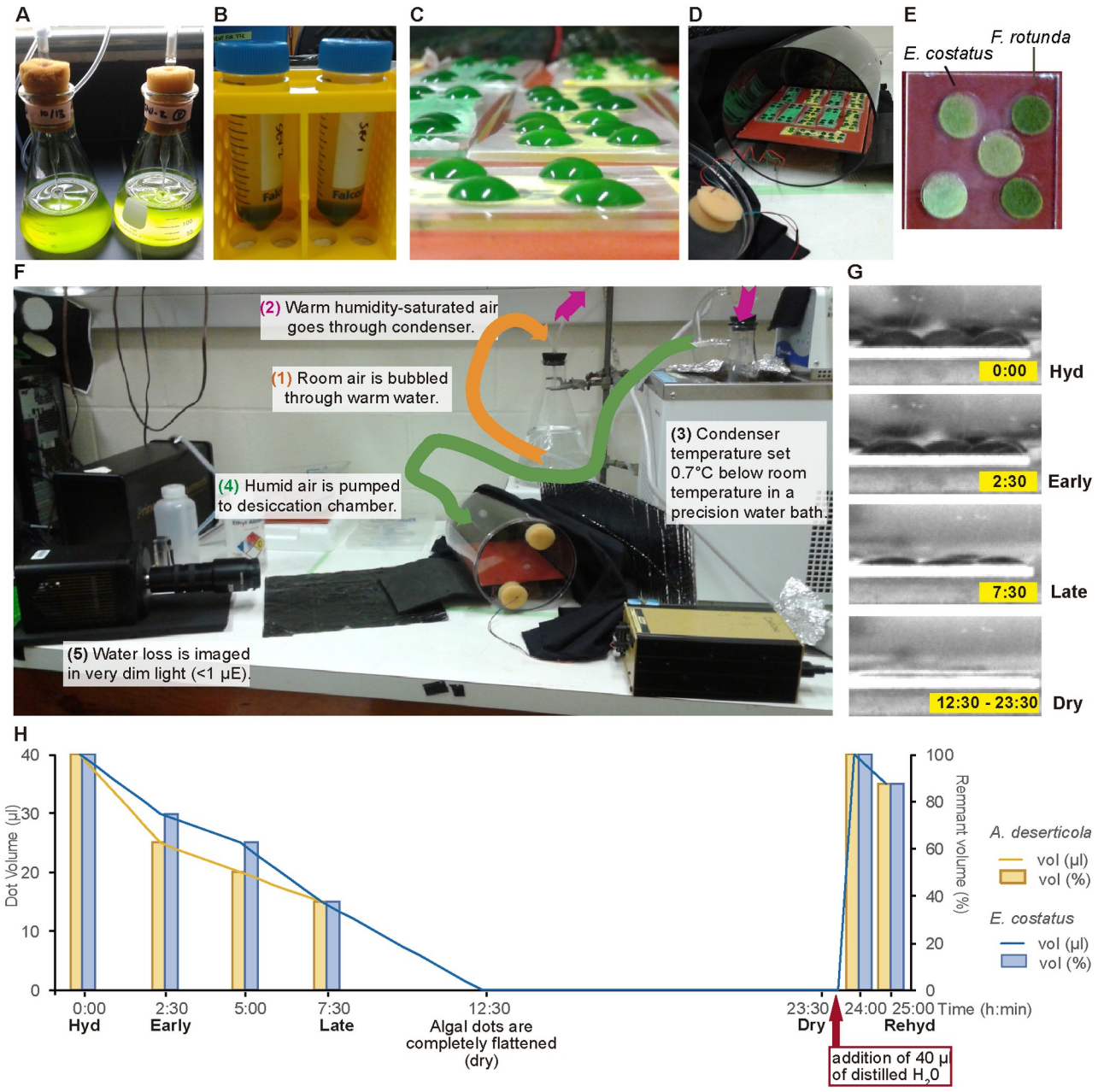


**Fig. S2.**
Overview of the experimental set-up used for desiccation. (A) Algal cultures growing 250 mL flasks with 150 mL of BBM+WH medium. (B) Cultures concentrated by gravity and adjusted to Abs600= 1.8. (C) Algal dots. 40 µL of concentrated algal culture were placed on microscopy cover slips (prior to desiccation). (D) Algal dots in a custom-made apparatus optimized for slow desiccation. (E) Dry algal dots. (Different colors correspond to different species.) (F) Components of the desiccation apparatus. (G) Images of algal dots in the desiccation chamber at time points t = 0, t= 2.5 h, t= 7.5, t = 12.5 h. Images were captured with 1 minute exposure. (H) Volumetric changes in algal dots during desiccation and rehydration. Aquatic *E. costatus* is indicated in blue, desert-evolved *A. deserticola* in orange. The left vertical axis is estimated volume (µL). The right vertical axis is estimated percentage of initial volume.

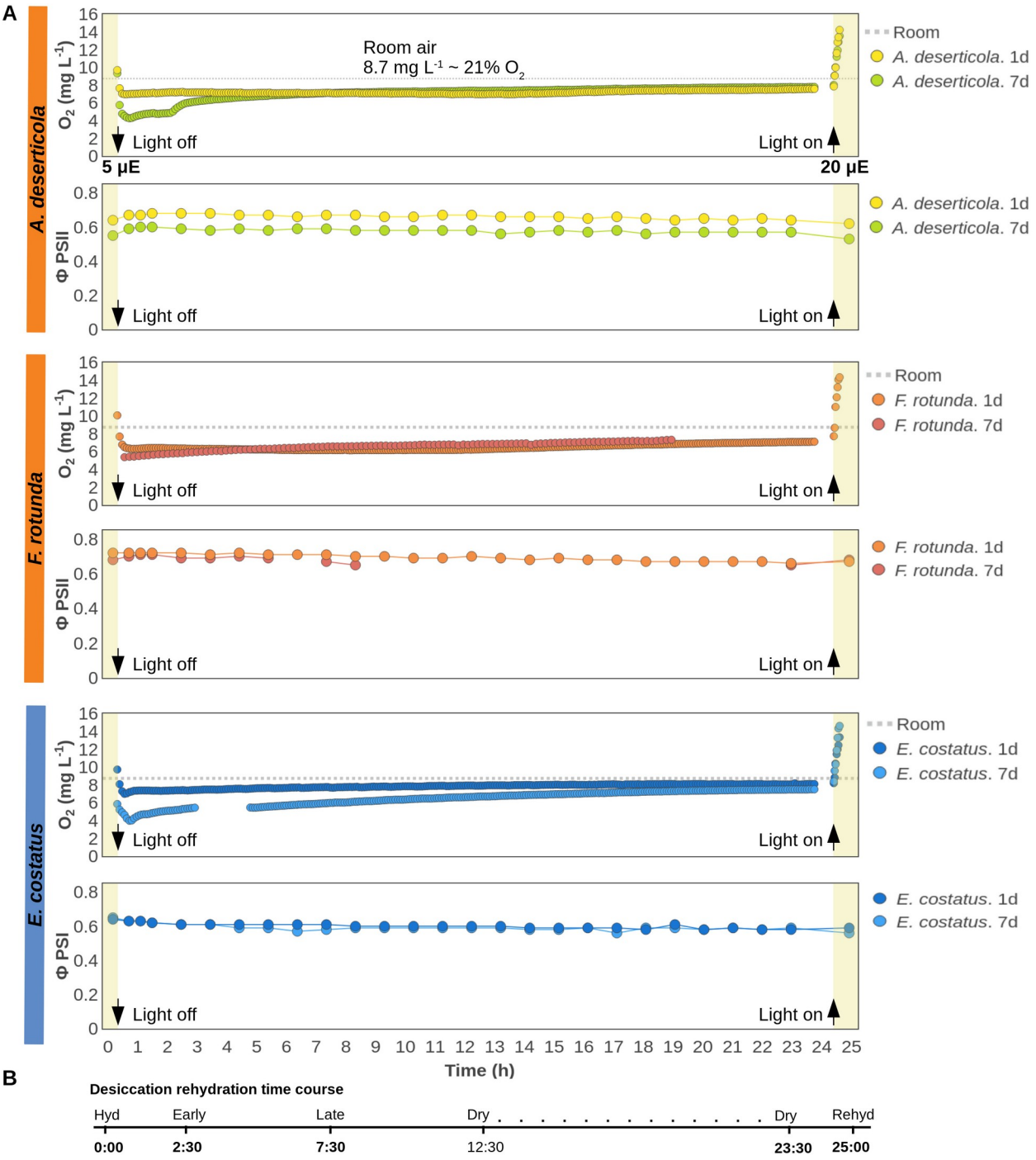


**Fig. S3.**

Oxygen concentration ($O_2$) and the quantum yield of photosynthesis (Φ PSII) in algal dots maintained hydrated for 25 h. (A) For each species, graphs present $O_2$, concentration (top) and Φ PSII (bottom) over the experimental period. Light periods are indicated with dark yellow shadowing and the light intensity levels are provided in the top most graph. Arrows signal the beginning (↓) and end (↑) of the 24 h of darkness. "1d" and "7d" identify algal dots prepared from cultures replenished with fresh medium 1 day or 7 days before experiments. (B) Schema of the desiccation-rehydration time course. The stages sampled for the RNA-seq experiment are indicated in bold.

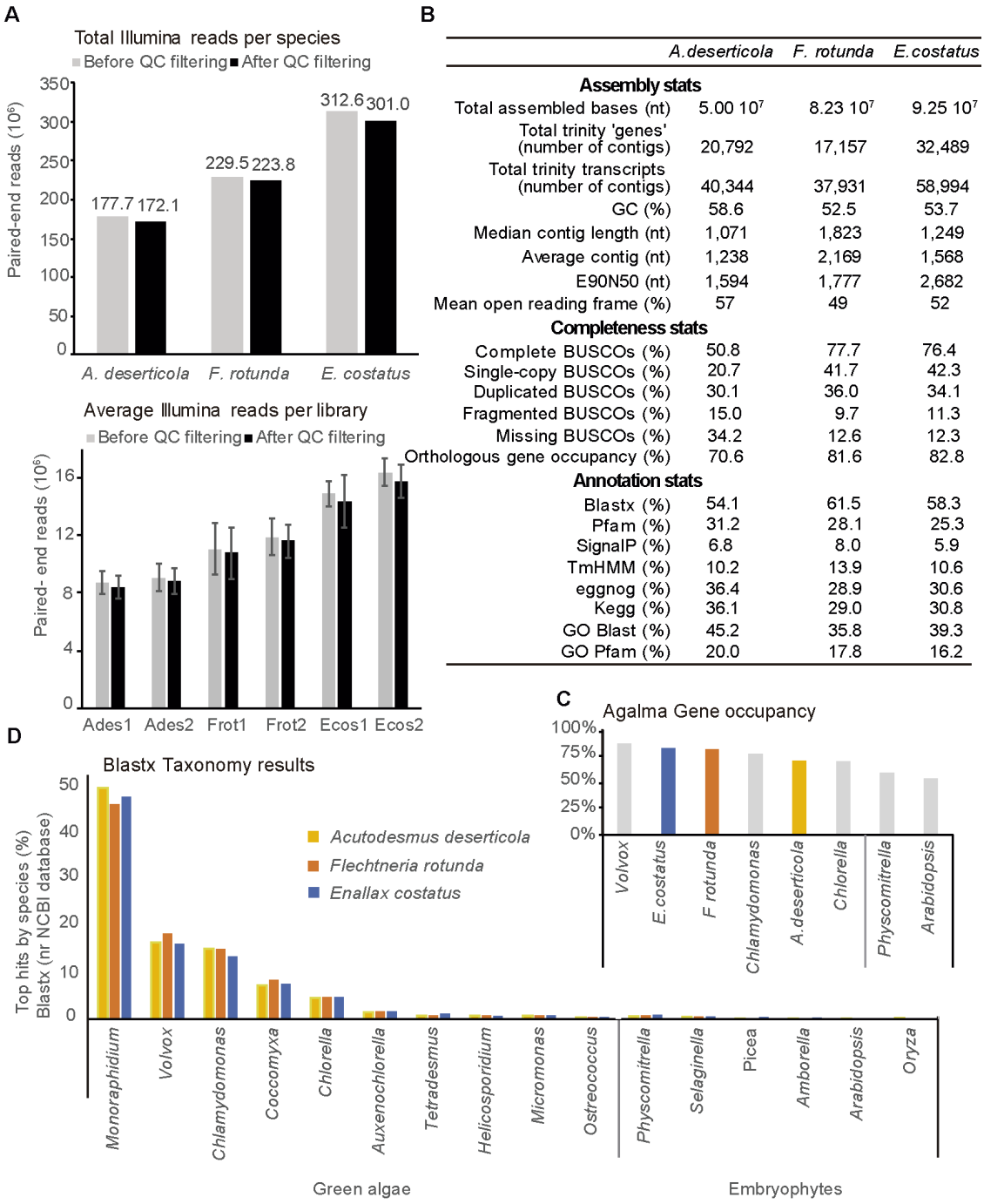


**Fig. S4**

Overview of sequencing, reference transcriptome assembly, and annotation. (A) Total paired-end reads per species and average (± s.d.) paired-end reads per library for each species and replicate before and after quality trimming. (B) Summary of quality parameters of assembly and annotation for each reference transcriptome. (C) Estimation of completeness of each reference transcriptome as a percentage of orthologous genes identified in each assembly using the Agalma pipeline. (D) Blastx top hit species distribution. Contigs assembled using Trinity were blasted against non-redundant NCBI database for annotation. *Monoraphidium* is the closest species with a sequenced and annotated genome to the taxa used in this study. Available annotation data from *Tetradesmus* (algal species within Scenedesmaceae) are restricted to organellar genes.

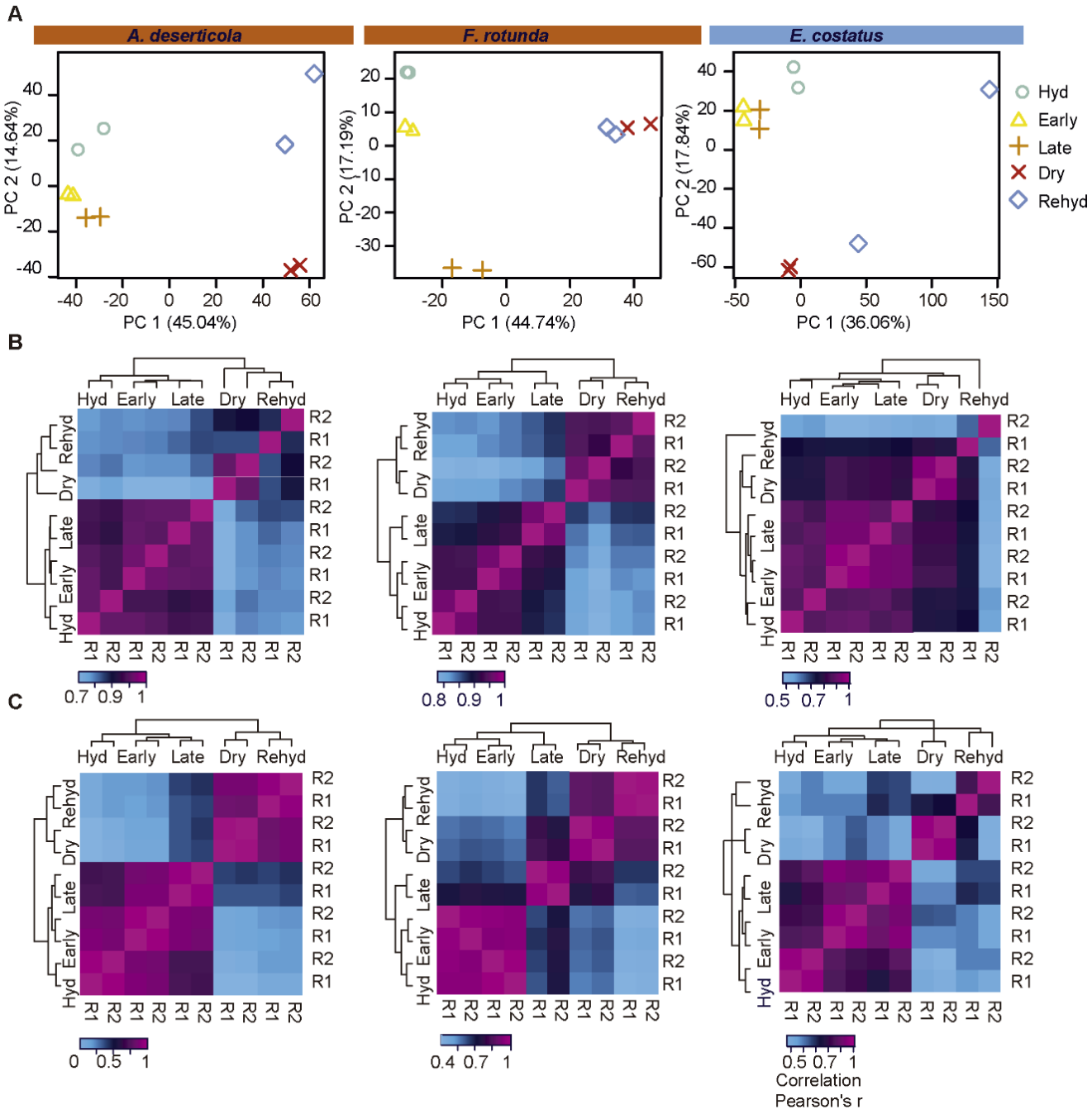


**Fig. S5.**
Comparative analysis of similarity across samples and replicates. (A) Principal Component Analysis (PCA) showing the overall similarity among samples. (B) Correlation matrix heatmaps displaying the Pearson correlation values calculated from pairwise comparisons among samples generated from all genes in the reference transcriptome. (C) Clustered heatmaps based only on differentially expressed genes (DEGs).

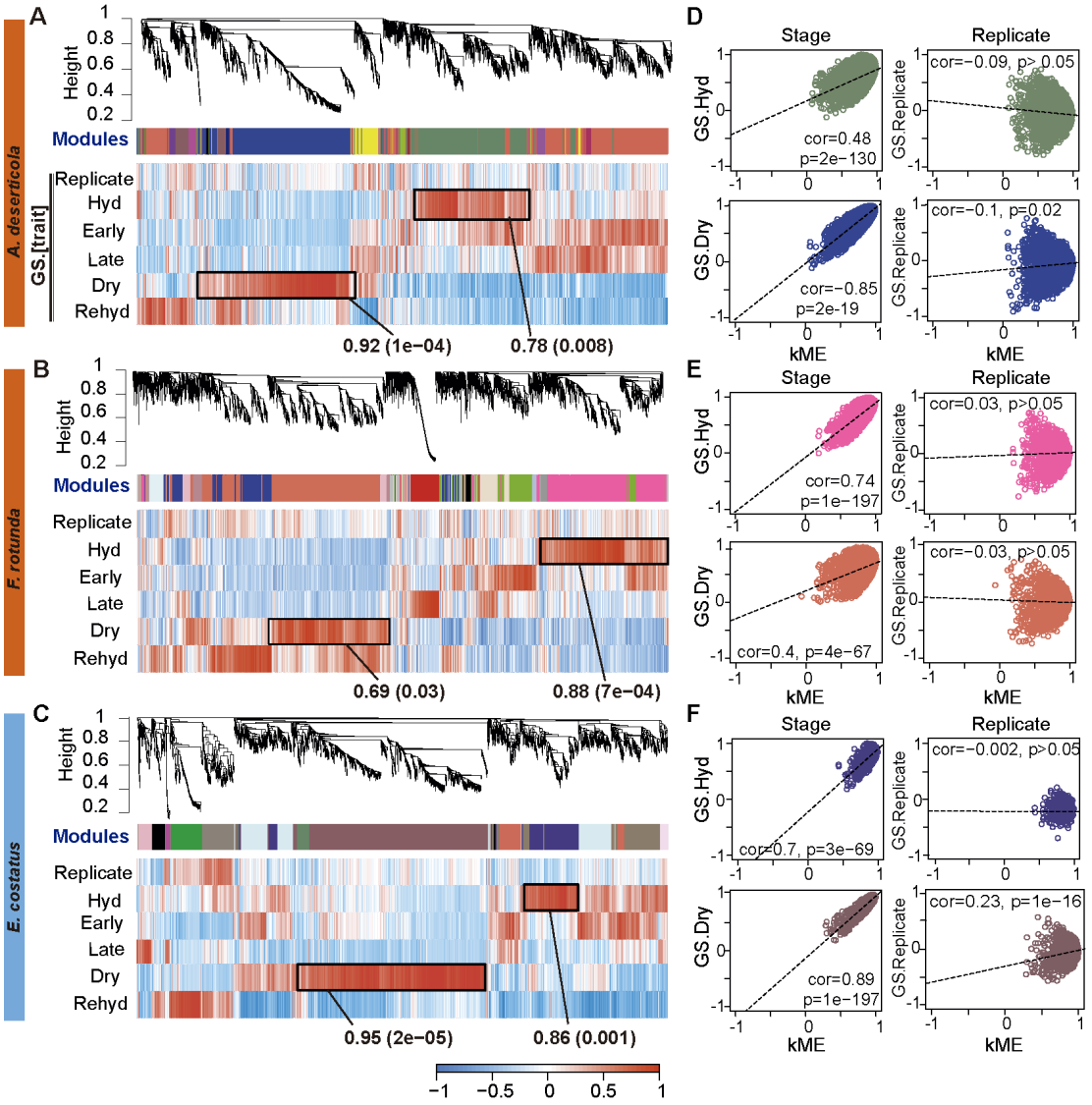


**Fig. S6.**
Results of co-expressed weighted correlation network analysis (WGCNA) of transcription data during desiccation and rehydration time course (top, *A. deserticola*; middle, *F. rotunda*; bottom, *E. costatus*). (A-C) Dendrogram (top) used to identified co-expressing transcripts during desiccation and rehydration. Highly correlated clusters of genes were grouped in 'network modules', shown in the first graph underneath each tree. Each module of genes is identified by a different color. Subsequent color bar graphs represent the correlation between each gene (as Gene Significance, GS) and a given external trait indicated on the left: Replicate, Hydrated, Early, Late, Dry and Rehydrated. Genes in which expression is highly correlated with a trait are shown in red, those anticorrelated in blue. For each taxon, the modules most correlated with the traits "Hydrated" and "Dry" are identified with a black box on the bar plot; the module eigengenes (overall correlation coefficient) and statistical significance (adjusted p-values) are indicated for each box. (D-F) Plots showing the correlation of each gene within a module of interest and an external driver ("Hydrated" and "Dry"). The correlation of each module of interest with the external trait "Replicate" is provided on the right. Scatterplots show GS versus gene module membership (kME). For each module, "cor" is the Pearson correlation coefficient; "p" is the Bonferroni adjusted p-value.

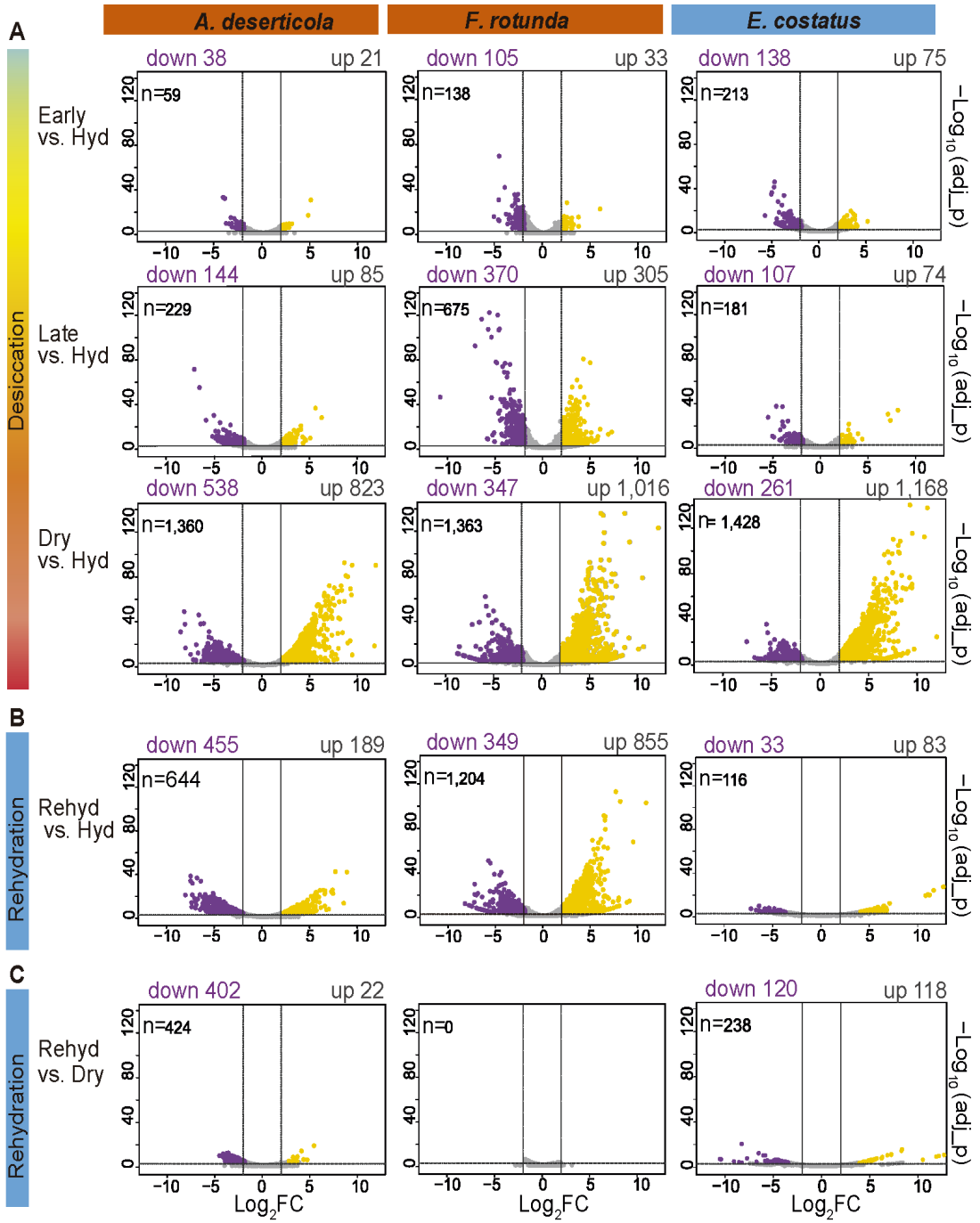


**Fig. S7.**
Volcano plots representing changes in expression during the desiccation and rehydration time course. From left to right, data corresponding to desert-evolved *A. deserticola* and *F. rotunda*, and the aquatic *E. costatus*. From top to bottom results from (A) each stage identified during desiccation when compared to "Hydrated" samples. (B) "Rehydrated" samples compared to "Hydrated" samples. (C) "Rehydrated" samples compared to "Dry" samples. In each plot, each dot corresponds to a gene. Y axis is statistical significance, -1*$Log_{10}$(adj_p-value). X axis is the change in expression as $Log_2$FC. Criteria for DEG (significance level <0.001 and FC>4) are indicated in the graph by dashed black lines. Upregulated DEGs are shown in gold, downregulated in purple. Total number of DEGs in each comparison is given in black.

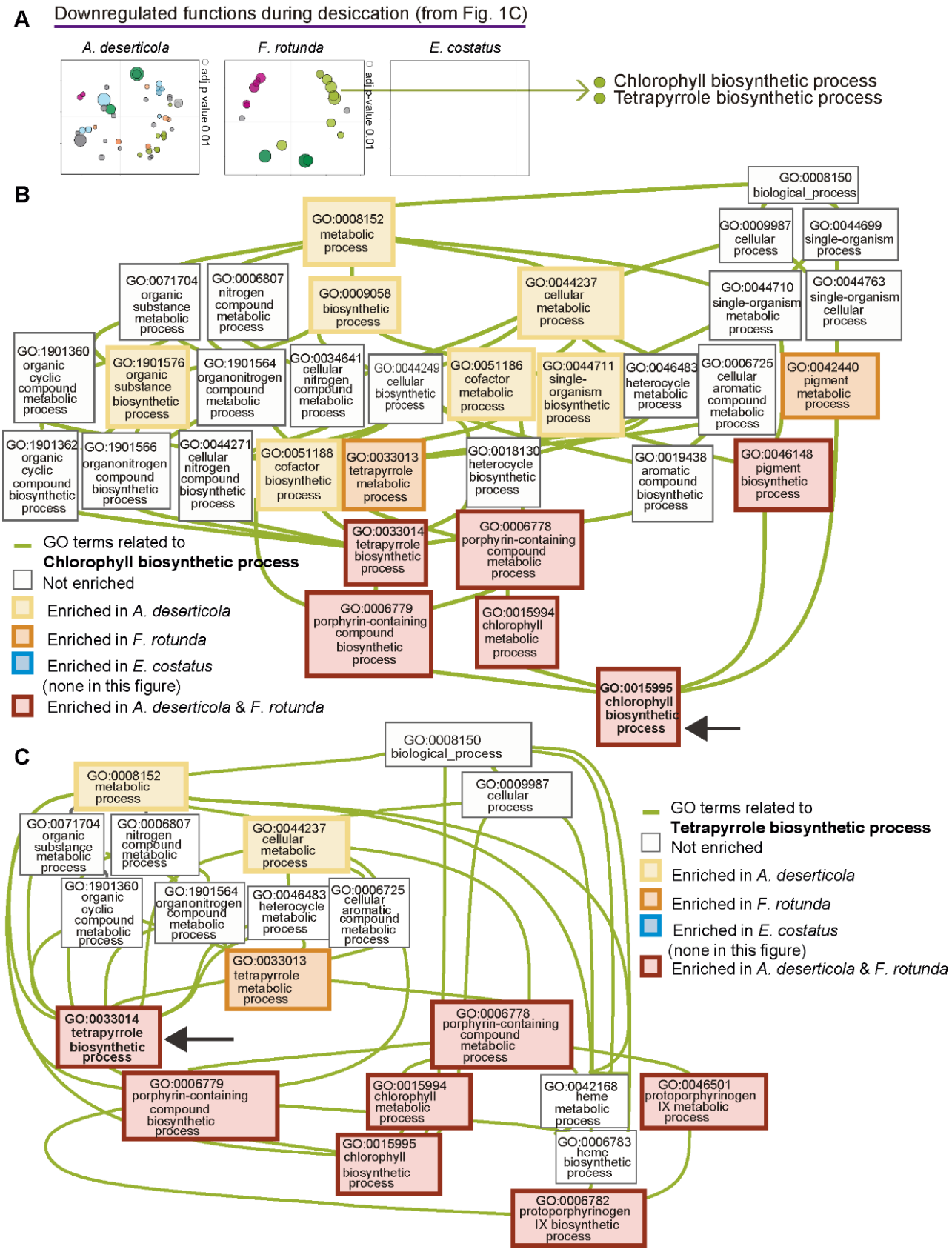
A
Downregulated functions during desiccation (from Fig. 1C)
A. deserticola
F. rotunda
E. costatus
adj p-value 0.01
adj p-value 0.01
Chlorophyll biosynthetic process
Tetrapyrrole biosynthetic process
B
GO:0008150 biological_process
GO:0008152 metabolic process
GO:0009987 cellular process
GO:0044699 single-organism process
GO:0071704 organic substance metabolic process
GO:0006807 nitrogen compound metabolic process
GO:0009058 biosynthetic process
GO:0044237 cellular metabolic process
GO:0044710 single-organism metabolic process
GO:0044763 single-organism cellular process
GO:1901360 organic cyclic compound metabolic process
GO:1901576 organic substance biosynthetic process
GO:1901564 organonitrogen compound metabolic process
GO:0034641 cellular nitrogen compound metabolic process
GO:0044249 cellular biosynthetic process
GO:0051186 cofactor metabolic process
GO:0044711 single-organism biosynthetic process
GO:0046483 heterocycle metabolic process
GO:0006725 cellular aromatic compound metabolic process
GO:0042440 pigment metabolic process
GO:1901362 organic cyclic compound biosynthetic process
GO:1901566 organonitrogen compound biosynthetic process
GO:0044271 cellular nitrogen compound biosynthetic process
GO:0051188 cofactor biosynthetic process
GO:0033013 tetrapyrrole metabolic process
GO:0018130 heterocycle biosynthetic process
GO:0019438 aromatic compound biosynthetic process
GO:0046148 pigment biosynthetic process
GO:0033014 tetrapyrrole biosynthetic process
GO:0006778 porphyrin-containing compound metabolic process
GO:0006779 porphyrin-containing compound biosynthetic process
GO:0015994 chlorophyll metabolic process
GO:0015995 chlorophyll biosynthetic process
GO terms related to
Chlorophyll biosynthetic process
Not enriched
Enriched in A. deserticola
Enriched in F. rotunda
Enriched in E. costatus
(none in this figure)
Enriched in A. deserticola & F. rotunda
C
GO:0008150 biological_process
GO:0008152 metabolic process
GO:0009987 cellular process
GO:0071704 organic substance metabolic process
GO:0006807 nitrogen compound metabolic process
GO:0044237 cellular metabolic process
GO:1901360 organic cyclic compound metabolic process
GO:1901564 organonitrogen compound metabolic process
GO:0046483 heterocycle metabolic process
GO:0006725 cellular aromatic compound metabolic process
GO:0033013 tetrapyrrole metabolic process
GO:0033014 tetrapyrrole biosynthetic process
GO:0006778 porphyrin-containing compound metabolic process
GO:0006779 porphyrin-containing compound biosynthetic process
GO:0015994 chlorophyll metabolic process
GO:0042168 heme metabolic process
GO:0046501 protoporphyrinogen IX metabolic process
GO:0015995 chlorophyll biosynthetic process
GO:0006783 heme biosynthetic process
GO:0006782 protoporphyrinogen IX biosynthetic process
GO terms related to
Tetrapyrrole biosynthetic process
Not enriched
Enriched in A. deserticola
Enriched in F. rotunda
Enriched in E. costatus
(none in this figure)
Enriched in A. deserticola & F. rotunda

**Fig. S8.**
Identification of inter-related GO terms associated with a metabolic function through direct acyclic graphs. (A) Simplified version of a portion of Fig. 1C showing downregulated functions during desiccation in REVIGO scatterplots. To facilitate comparison across taxa, related GO terms to a Biological Process (BP) of interest are indicated with the same color in the scatterplots. Related GO terms were identified by mapping the lists of enriched GO terms in direct acyclic graphs and using the 'search path' function in GOView. Direct acyclic graphs representing the hierarchical structure of GO terms associated with (B) Chlorophyll biosynthetic process and (C) Tetrapyrrole biosynthetic process (indicated with black arrows). Each box represents a GO term and the green lines show the relationships among terms. A colored box indicates that a given term is significantly enriched in DEGs downregulated during desiccation. Those GO terms enriched in downregulated DEGs exclusively in *A. deserticola* are shown in yellow, in *F. rotunda* shown in orange. Terms enriched in downregulated DEGs in both desert taxa are highlighted in red and illustrate an important congruence in behavior. No GO terms related to chlorophyll synthesis were enriched among downregulated DEGs of *E. costatus*.

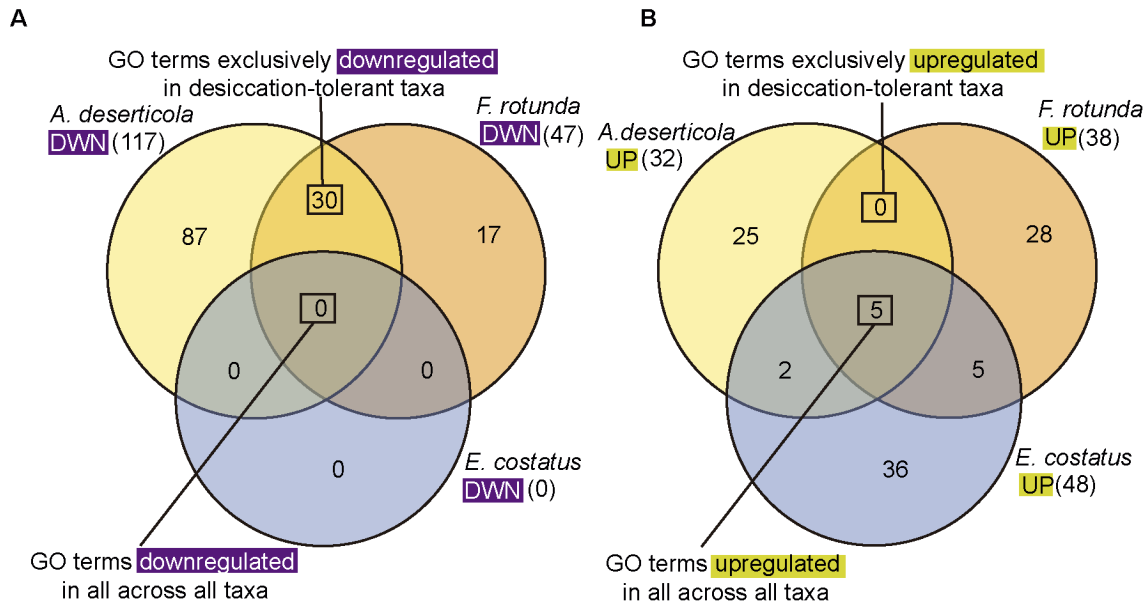


**Fig. S9.**

Venn diagrams showing the overlap of the GO terms significantly enriched (adjusted p-value <0.05) in DEGs during desiccation in desiccation tolerant and desiccation intolerant microalgae species. (A) GO terms enriched in downregulated DEGs during desiccation. (B) GO terms enriched in upregulated DEGs during desiccation. For each species and in each figure, the number between brackets is the total number of enriched GO terms. The full list of GO terms is presented in Dataset S2.

**References for SI Appendix**


1. Cardon ZG, Peredo EL, Dohnalkova AC, Gershone HL, Bezanilla M (2018) A model suite of green algae within the Scenedesmaceae for investigating contrasting desiccation tolerance and morphology. *J Cell Sci* 131:jcs212233.
2. Lewis LA, Flechtner VR (2004) Cryptic species of *Scenedesmus* (Chlorophyta) from desert soil communities of western North America. *J Phycol* 40(6):1127–1137.
3. Lunch CK, et al. (2013) The xanthophyll cycle and NPQ in diverse desert and aquatic green algae. *Photosynth Res* 115(2–3):139–51.
4. Stone JD, Storchova H (2015) The application of RNA-seq to the comprehensive analysis of plant mitochondrial transcriptomes. *Mol Genet Genomics* 290(1):1–9.
5. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
6. Haas BJ, et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8(8):1494–1512.
7. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
8. Dunn CW, Howison M, Zapata F (2013) Agalma: An automated phylogenomics workflow. *BMC Bioinformatics* 14:330.
9. The UniProt Consortium (2017) UniProt: The Universal Protein knowledgebase. *Nucleic Acids Res* 45:158–169.
10. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2016) GenBank. *Nucleic Acids Res* 44:67–72.
11. de Cambiaire J-C, Otis C, Lemieux C, Turmel M (2006) The complete chloroplast genome sequence of the chlorophycean green alga *Scenedesmus obliquus* reveals a compact gene organization and a biased distribution of genes on the two DNA strands. *BMC Evol Biol* 6:37.
12. Nedelcu A. M (2000) The complete mitochondrial dna sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res* 10(6):819–831.
13. Finn RD, Clements J, Eddy SR (2011) HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res* 39:29–37.
14. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–786.
15. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. *J Mol Biol* 305(3):567–580.
16. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(D1):109–114.
17. Ashburner M, et al. (2000) Gene Ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
18. Powell S, et al. (2012) eggNOG v3.0: Orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* 40(D1):284–289.

19. Hunter S, et al. (2009) InterPro: The integrative protein signature database. *Nucleic Acids Res* 37:211–215.
20. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
21. Li B, Dewey CN (2011) RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
22. Horvath S (2011) *Weighted Network Analysis: Applications in Genomics and Systems Biology* doi:10.1007/978-1-4419-8819-5.
23. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.
24. Young MD, Wakefield MJ, Smyth GK (2010) goseq: Gene Ontology testing for RNA-seq datasets reading data. *Genome biology 11.2: R14.*
25. Supek F, Bošnjak M, Škunca N, Šmuc T (2011) Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6(7).
26. Zhang B, Kirov S, Snoddy J (2005) WebGestalt: An integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 33:741–748.
27. Walter W, Sánchez-Cabo F, Ricote M (2015) GOplot: An R package for visually combining expression data with functional analysis. *Bioinformatics* 31(17):2912–4.
28. Zheng Y, et al. (2016) iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol Plant* 9(12):1667–1670.
29. Silvestro D, Michalak I (2012) - raxmlGUI: A graphical front-end for RAxML. *Org Divers Evol* 12: 335–337.