Environmental gradients reveal stress hubs predating plant terrestrialization

Armin Dadras[1,*], Janine M. R. Fürst-Jansen[1,*], Tatyana Darienko[1], Denis Krone[1], Patricia Scholz[2], Tim P. Rieseberg[1], Iker Irisarri[1,3,4], Rasmus Steinkamp[1], Maike Hansen[5], Henrik Buschmann[6], Oliver Valerius[7], Gerhard H. Braus[7], Ute Hoecker[5], Marek Mutwil[8], Till Ischebeck[9], Sophie de Vries[1], Maike Lorenz[10], Jan de Vries[1,2,11#]

1 — University of Goettingen, Institute of Microbiology and Genetics, Department of Applied Bioinformatics, Goldschmidtstr. 1, 37077 Goettingen, Germany
2 — University of Goettingen, Albrecht Haller Institute of Plant Science, Department of Plant Biochemistry, Justus-von-Liebig-Weg, 37077 Goettingen, Germany
3 — University of Goettingen, Campus Institute Data Science (CIDAS), Goldschmidtstr. 1, 37077 Goettingen, Germany
4 — Leibniz Institute for the Analysis of Biodiversity Change (LIB), Centre for Molecular Biodiversity Research, Museum of Nature, Hamburg, Martin-Luther-King Platz 3, 20146 Hamburg, Germany
5 — University of Cologne, Institute for Plant Sciences and Cluster of Excellence on Plant Sciences (CEPLAS), Biocenter, Zülpicher Str. 47b, Cologne 50674, Germany
6 — University of Applied Sciences Mittweida, Faculty of Applied Computer Sciences and Biosciences, Section Biotechnology and Chemistry, Molecular Biotechnology, Technikumplatz 17, 09648 Mittweida, Germany
7 — University of Goettingen, Institute of Microbiology and Genetics and Göttingen Center for Molecular Biosciences (GZMB) and Service Unit LCMS Protein Analytics, Department of Molecular Microbiology and Genetics, University of Goettingen, Grisebachstr. 8, 37077 Goettingen, Germany
8 — School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore, 637551, Singapore
9 — University of Münster, Green Biotechnology, Institute of Plant Biology and Biotechnology (IBBP), 48143 Münster, Germany
10 — University of Goettingen, Albrecht-von-Haller Institute for Plant Science, Department of Experimental Phycology and SAG Culture Collection of Algae, Nikolausberger Weg 18, 37073 Goettingen, Germany
11 — University of Goettingen, Goettingen Center for Molecular Biosciences (GZMB), Department of Applied Bioinformatics, Goldschmidtstr. 1, 37077 Goettingen, Germany
*contributed equally
#author for correspondence: devries.jan@uni-goettingen.de

ORCID: Armin Dadras 0000-0001-7649-2388, Janine MR Fürst-Jansen 0000-0002-5269-8725, Tatyana Darienko 0000-0002-1957-0076, Patricia Scholz, 0000-0003-0761-9175, Tim P Rieseberg 0000-0003-3548-8475, Henrik Buschmann 0000-0003-3022-6150, Oliver Valerius 0000-0003-4430-819X, Gerhard H. Braus 0000-0002-3117-5626, Ute Hoecker 0000-0002-5636-9777, Marek Mutwil 0000-0002-7848-0126, Iker Irisarri 0000-0002-3628-1137, Till Ischebeck 0000-0003-0737-3822, Sophie de Vries 0000-0002-5267-8935, Maike Lorenz, 0000-0002-2277-3077, Jan de Vries, 0000-0003-3507-5195

ABSTRACT

Plant terrestrialization brought forth the land plants (embryophytes). Embryophytes account for most of the biomass on land and evolved from streptophyte algae in a singular event. Recent advances have unraveled the first full genomes of the closest algal relatives of land plants; among the first such species was *Mesotaenium endlicherianum*. Here, we used fine-combed RNAseq in tandem with photophysiological assessment on *Mesotaenium* exposed to a continuous range of temperature and light cues. Our data establish a grid of 42 different conditions, resulting in 128 transcriptomes and ~1.5 Tbp (~9.9 billion reads) of data to study combinatory effects of stress response using clustering along gradients. We describe major hubs in genetic networks underpinning stress response and acclimation in the molecular physiology of *Mesotaenium*. Our data suggest that lipid droplet formation, plastid and cell wall-derived signals denominate molecular programs since more than 600 million years of streptophyte evolution—before plants made their first steps on land.

KEYWORDS

plant evolution; functional genomics; evolutionary genomics; stress physiology; streptophyte algae; co-expression networks

MAIN

Plant terrestrialization changed the face of our planet. It gave rise to and plants (Embryophyta), the major constituents of Earth's biomass (Bar-On et al. 2018) and founders of the current levels of atmospheric oxygen (Lenton et al. 2016). Land plants belong to the Streptophyta, a monophyletic group that includes the paraphyletic freshwater and terrestrial streptophyte algae and the monophyletic land plants. Meticulous phylogenomic efforts have established the relationships of land plants to their algal relatives (Wickett et al. 2014; Leebens-Mack et al., 2019). These data brought a surprise: the filamentous and unicellular Zygnematophyceae—and not other morphologically more elaborate algae—are the closest algal relatives of land plants. Now the first three genomes of major

orders of Zygnematophyceae (see Hess et al., 2022) are at hand: *Mesotaenium endlicherianum*, *Spirogloea muscicola* (Cheng et al., 2019), and *Penium margaritaceum* (Jiao et al., 2020). Using these, we are beginning to redefine the molecular chassis shared by land plants and their closest algal relatives. Included in this shared chassis will be those genes that facilitated plant terrestrialization. We here focus on one critical aspect: the molecular toolkit for the response to environmental challenges. For this, we use the unicellular freshwater/subaerial alga *Mesotaenium endlicherianum*.

Land plants use a multilayered system for the adequate response to environmental cues. This involves sensing, signaling, and response mainly by the production of, e.g., protective compounds. Some of the most versatile patterns in land plant genome evolution concerns genes for environmental adaptation (Golicz et al., 2016; Gordon et al., 2017; Bayer et al., 2020). That said, there is a shared core of key regulatory and response factors that are at the heart of plant physiology. These include phytohormones such as abscisic acid (ABA) found in non-vascular and vascular plants (for an overview, see Umezawa et al., 2010; Bowman et al., 2019), protective compounds resting on specialized metabolic routes such phenylpropanoid-derived compounds as well as proteins such as LATE EMBRYOGENESIS ABUNDANT (LEA; Hundertmark and Hincha 2008; Carella et al., 2019). Many of the genes integrated into these stress-relevant metabolic routes have homologs in streptophyte algae (Rieseberg et al., 2022). Taking angiosperms as reference, such stress-relevant pathways are often patchy. Whether these are also used under the relevant conditions is currently unknown. For example, while Zygnematopyhceae have a homolog to the ABA-receptor PYL (de Vries et al. 2018, Cheng et al., 2019), this homolog works in a different, ABA-independent fashion (Sun et al., 2019). Thus, it is important to put the genetic chassis that could act under environmental shifts to the test.

Here, we used a fine grid of a bifactorial gradient for two key terrestrial stressors, variation in irradiance and temperature, to probe the genetic network that the closest algal relatives of land plants possess for its responsiveness to abiotic environment. Correlating growth, physiology, and global differential gene expression patterns from 128 transcriptomes (9,892,511,114 of reads, 1.5 Tbp of data) across 126 distinct samples covering a temperature range of >20°C and light range of >500 µmol photons m^{-2} s^{-1}, we pinpoint hubs in the circuits that have been shared along more than 600 million years of streptophyte evolution.

RESULTS

A physiological grid: co-dependency of eurythermy and euryphoty in *Mesotaenium*

We studied the genome-sequenced strain SAG 12.97 of the freshwater alga *Mesotaenium endlicherianum*, a member of the Zygnematophyceae, the closest algal relatives of land plants (Cheng et al., 2019; Figure 1a and 1b). We cultivated *Mesotaenium* in a large-scale setup in 1.5 liters of C medium up to a cell density of 0.33 AU at 680 nm. The culture was distributed across 504 wells (42 twelve well plates; 2.5 mL of culture per well). The well plates were placed on a table with a temperature gradient from 8.6±0.5 °C to 29.2±0.5 °C on the x-axis. On top of the table, white LED lamps created an irradiance gradient from 21.0±2.0 to 527.9±14.0 µmol photons m^{-2} s^{-1} across the y-axis, thus creating a 2D gradient table (Figure 1b, Suppl. Table 1). The 504 cultures were exposed to this gradient setup for 65 hours. The physiological status of the algae was assessed by determining the maximum quantum yield (F_v/F_m) using pulse amplitude modulation fluorometry (PAM; IMAGING PAM, Walz, Germany); growth was assessed using a microplate reader with absorption at 480 nm, 680 nm, and 750 nm (Figure 1c); the entire procedure was repeated in three successive biological replicates (i.e. three runs of the table, 504 F_v/F_m and 4,536 absorption measurements per replicate).

The algae showed significant differences ($p \leq 0.001$) in growth and gross physiology: F_v/F_m values as well as absorption values decrease (for F_v/F_m values at 20.5±1.0 °C: from 0.66±0.02 for I=21.14 µmol photons m-2 s-1 to 0.042±0.04 for I=534.7 µmol photons m-2 s-1) with rising intensities of irradiance (Figure 1d, Suppl. Fig. 1, Suppl. Table 2). The lowest F_v/F_m values (down to zero) were recorded at conditions of highest irradiance and lowest temperature. Here, low temperature had a stronger negative impact on growth and physiology than light (for Fv/Fm values: at 8.6±0.5 °C, 0.011±0.02 at 133±27 µmol photons m-2 s-1 compared to 0.463±0.02 at 29.2±0.5 °C at 118±25 µmol photons m-2 s-1). Values on growth and physiology clustered by light were less broadly distributed than if clustered by temperature (Figure 1e, 1f). Even the highest light intensity (527.9±14.0 µmol photons m-2 s-1) was stressful, but tolerable for the physiology of *Mesotaenium* at temperatures

between 20.5±0.1°C (Fv/Fm=0.042±0.04) to 25.3±0.1°C (Fv/Fm= 0.045±0.04); more extreme temperatures resulted in undetectable Fv/Fm values. Thus, eurythermy might establish the foundation for euryphoty in *Mesotaenium endlicherianum*.

Fine-combed global differential gene expression profiles and a gene models for *Mesotaenium*

To shed light on the molecular mechanisms that underpin the switch from tolerable steady-state conditions to adverse environmental cues in *Mesotaenium*, we applied global gene expression analyses using RNAseq. We pooled all twelve wells per plate and extracted RNA from a total of 126 samples (42 plates, three biological replicates). 114 samples yielded usable RNA that was used to build 128 libraries (with a minimum of three biological replicates and additional technical replicates) for sequencing on the Illumina NovoSeq6000 platform. We generated a total of 1.5 Tbp of 150 bp paired read data at an average depth of 37.7 million reads per sample (~9.9 billion reads in total). Building on this wealth of data, we updated the *Mesotaenium* gene models. The number of protein-coding mRNAs increased from 11080 in the original annotation (V1; Cheng et al. 2019) to 40326 protein-coding mRNA (26009 high confidence, 14317 low confidence; including splice variants) in 19233 genes; an additional 4408 mRNA (in 4312 genes) labeled as "predicted gene" in our gene models (Suppl. Table 3). The new gene models of annotation V2 brings the number of genes in *Mesotaenium* closer to other Zygnematophyceae with similar genome sizes; V2 has 43 more BUSCO genes (+10%; 21 less fragmented, 22 less missing; viridiplantae_odb10) than V1 (Suppl. Fig. 2). Besides, we calculated Annotation Edit Distance metrics (AED) to assess the congruence (0 to 1, with 0 being the best) between biological evidence and V1 and V2. In the cumulative fraction of annotation against AED score, V2 has more mRNAs with AED < 0.5. For example, 70% of mRNAs in V1 (7756 mRNAs) have an AED score < 0.5 compared to 60% in V2 (26840 mRNAs). This is sensible since V2 was built based on the same set of evidence used to calculate AED and it shows higher congruence with them (Suppl. Fig. 3). Thus, we pseudoaligned our data onto the new *Mesotaenium* transcriptome V2 (average alignment rate was 87.31%; Suppl. Table 4 Table).

To understand the gross profile of the gene expression data, we performed principal component analysis (PCA; Figure 2a). Independent biological replicates from the same condition clustered in close proximity. The most variation in data was explained by temperature (PC1; describes 35% of variance), followed by irradiance (PC2; describes 18.1% of variance). We evaluated the distance (Figure 2b) and Spearman correlation (Figure 2c) using all genes to look for trends among different growth conditions. The data can be grouped into at least three categories: (1) samples with high light and/or high temperature, (2) a collection of low-temperature (8, 13, 17 °C) samples, and (3) samples at stead-state. Large clusters included steady-state, high light + heat, and high light. Most distinct was the cluster formed by samples from the high temperature + high light (Small multiples; Figure 2d and 2e).

Plastid-related genes stand out in differential gene expression profiles

For dissecting the differential gene expression responses, we divided the table into nine sectors and, additionally, a cohort of stressed algae based on Fv/Fm < 0.5 (Figure 3). 36 comparisons were performed, among which we focused on nine, which additionally included the Fv/Fm-based comparison. Genes were considered to be differentially expressed between groups at an absolute fold change ≥ 2 and a Benjamini-Hochberg corrected $p \leq 0.01$ (Figure 3a and b). Gross gene expression profiles were titratable by the intensity of environmental cues, i.e., with increasing disparity between conditions compared, and overall following the pattern in the PCA (*cf*. Figure 3b and Figure 2a). The most differentially regulated genes (6,578) were pinpointed by comparing low light and low temperature (LLI_LT) versus high light and high temperature (HLI_HT). Enriched GO terms among regulated genes most frequently included plastid biology-associated genes (Figure 3c). To scrutinize these data for specific genes that show a robust and universal response to alterations to the environment, we intersected all 8,157 significantly regulated genes pinpointed by the different comparisons. 3, 30, and 124 genes overlapped among all 9, 8, and 7 comparisons, respectively. These concertedly pinpointed genes were mostly light harvesting genes, corroborating the importance of plastids in the overall cell biology of *Mesotaenium* (Figure 3d). Indeed, the 30 genes found in all comparisons included for example reactive oxygen species (ROS)-relevant genes such as ELIP and

fatty acid metabolic genes. To understand whether these genes integrate into the context of molecular programs, we next looked at gene co-expression.

Unsupervised gene expression clusters recover genetic programs shaped by physiology

The environmental gradients triggered changes in the expression of gene cohorts. We wanted to understand their concerted action independent of any prioritization guided by homology to any land plant genes—solely from the molecular programs that operated in the algae. To do so, we applied weighted gene co-expression network analysis (WGCNA) for unsupervised clustering (Figure 4). To then understand the driving forces behind these changes, we turned to the highly connected genes (nodes) in the network—the hubs (Figure 5).

The 17,905 genes expressed in our samples (and that passed the minimum expressison threshold) were clustered into 26 modules, which we refer to with colors (Figure 4a). Orange is the smallest module (39 genes), the largest modules are Turquoise, Blue, and Brown with 3568, 3101, and 1746 genes, respectively. The samples were taken under a range of distinct physiological conditions. Resulting data are a combined expression of the different environmental cues and the modulation of the algal physiology. To investigate the biological role of each module, we used their eigengenes as representatives for the modules' gene expression profiles and correlated their behavior with the two environmental cues light and temperature as well as the algal parameters absorption (culture density and pigmentation) and Fv/Fm (overall physiological status). One of the foremost general patterns in cellular response to stress are ROS. These act as signals as well as culprits that, if not quenched, damage biomolecules; this was represented in GO terms of module Green that positively correlated with light intensity (r = 0.88, p = $6x10^{-43}$) and negatively with Fv/Fm (r = -0.79, p = $6x10^{-29}$) (Figure 4d and Suppl. Fig. 4 to 7 and Suppl. Table 5 and 6).

The clusters also recovered the genetic signatures of thriving algae. Module Purple negatively correlates with increasing light (r = -0.94, p = $3x10^{-60}$) and positively with absorption and Fv/Fm (r = 0.67, p = $2x10^{-18}$ and r = 0.67, p = $2x10^{-18}$). These dense and physiologically healthy cell populations (experiencing no light stress) ramped up cell division (see Figure 4D and Suppl. Table 6), signified by homologs of cyclin and TPX2 appearing as hub genes. The 9th most connected hub gene was a kinesin homologous to important proteins such as PHRAGMOPLAST ORIENTING KINESIN 2 (Figure 5; Suppl. Table 7), which thus is a likely conserved cell division hub of all Phragmoplastophyta—going back to a common ancestor that lived in the late Cryogenium.

Conserved hubs for plastid-derived signals

Chloroplasts act as environmental sensors in land plant cells (Kleine et al., 2021). In concert with this, many of the clusters we identified were associated with plastid biology and/or physiology (Figure 4d, Suppl. Fig. 4 to 7, Suppl. Table 6). The brown cluster showed many plastid-related terms and negatively correlates with temperature (r = -0.95, p = $7x10^{-65}$) (Suppl. Fig. 5) and showed enrichment in GO-terms related to plastids, general transcription and translation. Among the top 20 hub genes in cluster brown, 12 were associated with translation and ribosomes. The light cyan cluster positively correlates with increasing light (r = 0.93, p = 10^{-56}) (Suppl. Fig. 6) and negatively with Fv/Fm (r = -0.67, p = $5x10^{-18}$) (Suppl. Fig. 4) whereas the blue cluster negatively correlates with increasing light (r = -0.76, p = 10^{-25}) and positively with Fv/Fm (r = 0.67, p = $2x10^{-18}$). Concomitantly, the blue module had a high number of enriched GO-terms (Suppl. Fig. 5 and Suppl. Table 6), many of which were plastid-related terms, cellular signaling, and terms that tie the two together; that is, signaling processes emanating from the plastid. This was also prominent in the light cyan module, where several terms related to terpenoid and apocarotenoid metabolism were enriched.

The hubs of many clusters, including those blue, light cyan, and yellow mentioned before, reflect an association with plastid-related processes. To highlight a few, the second most connected gene in module Blue was a homolog of GLK1 (Suppl. Fig. 8), a transcriptional factor (TF) that regulates chloroplast development and the activity of nuclear genes involved in photosynthetic light reaction and chlorophyll biosynthesis (Rossini et al., 2001; Yasumura et al., 2005; Waters et al., 2009). Blue also featured hydroxypyruvate reductase, important in photorespiration (Timm et al., 2008), as the fourth most connected gene. A CYP450 gene homologous to LUTEIN DEFICIENT 5 (LUT5), was the 7th most connected, suggesting the involvement of pigment-related signaling. Moreover, a homolog of ABA responsive elements-binding factor 2 (ABF2) was part of cluster blue,

bolstering previous discussions that parts of the ABA signaling module consist of ancient wires whose relevance in stress response predate plant terrestrialization and ABA dependency (de Vries et al., 2018; Sun et al., 2019; Fürst-Jansen et al., 2020).

Next to GLK—the most connected TF—other highly connected TFs appeared in Blue. These included the photomorphogenesis-regulating CONSTANS-like 3 (COL3; 4th most connected TF). Noteworthily, also a homolog of CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1) was present in module Blue; CO/COL and GLKs are both degradation targets of COP1 (Liu et al., 2008; Sarid Krebs et al., 2015; Ordoñez-Herrera et al., 2018). Further, the circadian regulator BROTHER OF LUX ARRHYTHMO (2nd most connected TF). Further, homologs of ETHYLENE-INSENSITIVE3-like 1 (6th most connected TF) and several ERFs were among the most connected TFs. A link to ethylene is noteworthy, because investigations of the Zygnematophyceae *Spirogyra pratensis* (*Sp*) have shown that *SpEIN3* can rescue Arabidopsis *ein3-1* mutant plants (Ju et al., 2015). Furthermore, exogenous application of ethylene on *Spirogyra* triggers stress-, plastid- and photosynthesis-associated gene expression responses similar to land plants (Van de Poel et al., 2016). This speaks to a conserved regulatory framework that involves ethylene-associated factors, and maybe ethylene itself, in environmental signaling cascades in the common ancestor of and plants and their closest algal relatives.

Light cyan featured not only hubs related to ROS homeostasis from the thioredoxin superfamily and other light-induced proteins, but also pigment and apocarotenoid metabolism; these are the source of important signals from the chloroplast that likely have deep evolutionary roots (Rieseberg et al., 2022) and are also formed by light dependent oxidative reactions (recently reviewed by Moreno et al., 2021). Module Yellow correlated positively with light intensity (r = 0.62, p = 10^{-14}) and negatively with absorption and Fv/Fm (r = -0.79, p = 10^{-28} and r = -0.81, p = $3x10^{-31}$; Figure 3B); GO terms associated with plastids and proteolytic enzymes (FtsH, ClpP; Kato et al., 2012), recapitulating well-known ties of protein homeostasis and plastid maintenance. Indeed, cluster yellow featured five hubs that are homologous to CLP proteases, critical for chloroplast protein homeostasis (Sjögren et al., 2006; Nishimura et al., 2016), and hubs homologous to genes that orchestrate the coordination of transcriptional activity between chloroplasts and the nucleus; the latter included homologs of (i) pTAC6, which is essential for plastid gene expression and thus chloroplast development in Arabidopsis (Pfalz et al., 2006), and (ii) a homolog of GENOMES UNCOUPLED 2, one of the foremost genes in the classical plastid–nucleus communication pathway (Susek et al., 1993). Among the TFs in cluster yellow was a homolog of the bZIP light signaling master regulator ELONGATED HYPOCOTYL 5 (HY5; reviewed in Jiao et al., 2007).

Of ancient signaling cascades and cell wall perturbance

Mitogen-activated protein kinases (MAPK) constitute environmental response pathways in all eukaryotes (Chen and Thorner, 2007). In land plants, several abiotic and biotic cues have been described to trigger MAPK-mediated signaling (Nakagami et al., 2005; Rodriguez et al., 2010; Meng and Zhang, 2013; Chen et al., 2021); MAPK and phototropin kinases appeared as hubs in cluster ‘blue’. Moreover, plant MAPK-based signaling is interwoven with wound response and brassinosteroid signaling (Nishiyama et al., 2005). Stress often coincides with a perturbance of plant cell wall homeostasis. Cluster pink includes hubs for such wounding and cell-wall derived signals. This was paired with the GO term brassinosteroid signaling, which balances growth, cell wall homeostasis, and stress in *Arabidopsis* (Sun et al., 2010; Planas-Riverola et al., 2019). Among the hubs in cluster pink were homologs for (i) diverse receptor kinases known from *Arabidopsis* to sense alterations in cell wall integrity (Hématy et al., 2007), and (ii) EXORDIUM (of which *Mesotaenium* has 12 homologs), which integrates growth with environmental signaling (Schröder et al., 2009). This was paired with the COBRA family proteins being the most and third most connected hubs in the module. These proteins are known to be involved in cell expansion and balancing pathogen response with growth (Schindelmann et al., 2001; Roudier et al., 2002; Ko et al., 2006). It appears that *Mesotaenium* bears parts of a loop that senses physico-chemcial perturbance of cell wall homeostasis; in land plants, these loops include includes brassinosteroid signaling (Wolf et al., 2014).

Lipid droplet formation constitutes a stress response predating plant terrestrialization

In land plants lipid droplet (LD) formation and TAG accumulation is common to many stress responses, including heat, cold and drought (Higashi et al., 2015; Mueller et al., 2015; Gidda et al., 2016; Doner et al., 2021; Krawczyk et al., 2022). We observed that cells of *Mesotaenium* accumulated inclusions resembling LDs (Figure 6a) upon prolonged exposure to stress. Consistently, these globular structures were stained by BODIPY™ 493/503 (EM/EX), a common dye for lipid and oil-rich compartments (Listenberger and Brown, 2007; Kretzschmar et al., 2020). Under different conditions of temperature and light conditions, counts of LDs per cell showed significant differences (Figure 6b, Suppl. Table 8). We observed that the CGI-58 homolog was the 10th most connected hub in cluster green (Figure 5b). CGI-58 is key to lipid homeostasis, causing, if perturbed, the Chanarin-Dorfman syndrome in humans and LD overaccumulation in *Arabidopsis* (Lass et al., 2006; James et al., 2010; Figure 5c). Further, differential gene expression profiles pinpointed elevation of transcripts for characteristic LD protein homologs such as HSD1 and OLE7 under high temperature and moderate light conditions (29 °C, 21 – 130 µmol photons m-2 s-1) and LDAP and PUX10 under high temperature and light conditions (21-29 °C, 130 – 528 µmol photons m-2 s-1) (Figure 6c).

To scrutinize whether these structures are comparable to LDs of land plants, we performed sub-cellular fractionizations, obtained lipid-rich phases, and subjected them to proteomics using liquid chromatography-mass spectrometry (LC-MS). We identified 739 proteins in the putative LD fraction and 1574 proteins in the total extract (Suppl. Table 9). Of these, 14 were significantly enriched in the putative LD fraction (Figure 6 d, volcano plot) including hallmark LD proteins such as oleosin (OLE), caleosin (CLO), steroleosin (HSD), and LD-associated protein (LDAP) (Figure 6 d, bar plots). Overall, *Mesotaenium* responds to stress conditions by formation of LDs containing signature proteins for embryophytic LDs.

DISCUSSION

Owing to their plain morphology, Zygnematophyceae emerged as unexpected closest algal relatives of land plants (Wickett et al., 2014; OneKP, 2019; Hess et al., 2022). That said, the molecular programs of Zygnematophyceae speak of their closes relationships to land plants. These point to a conserved chassis that likely operated in the last common ancestor of land plants and algae, featuring the proposed action of various hallmark genes (e.g., PYL homologs, GRAS family TFs and more) that were once considered land plant innovations. Building on the genomic resources for *Mesotaenium*, we have here delved into the molecular physiology and genetic programs of this alga, revealing which programs bear out when challenged with environmental cues.

Recent studies have proposed homology for the chassis of plastid–nucleus communication upon adverse environmental conditions between land plants and phragmoplastophytic streptophyte algae (Nishiyama et al., 2018; de Vries et al., 2018; Zhao et al., 2019). The GUN pathway likely has a conserved role in chloroplast transcription and streptophyte algal GUN1 homologs can rescue chloroplast retrograde signaling of Arabidopsis *Atgun1* mutants (Honkanen and Small, 2022); the degree of evolutionary conservation in the retrograde signaling pathway across streptophytes remains obscure (Honkanen and Small, 2022). Signals from damaged chloroplasts inhibit GLK1 expression in Arabidopsis (Martin et al., 2016). The negative correlation of module Blue (featuring *Me*GLK) with high light (leading to damaged chloroplasts) supports a role of *Me*GLK in operational retrograde signaling. Our data underscore that the wires between these components in plastid–nucleus communication are likely shared across more than 600 million years of streptophyte evolution and correlate with dealing with light regimes and adjustment of photosynthetic performance in the chloroplast also in the closest relatives of land plants.

One of the special features of plant and algal cells is their cell wall, forming their main interface with the environment. It is therefore not surprising that the cell wall is woven into a signaling network for environmental cues. In land plants, brassinosteroid-mediated signaling is part of a feedback loop for cell wall homeostasis and integrity (Wolf et al., 2014). While the involvement of brassinosteroids in streptophyte algae is doubtful—as is the case for many other phytohormones—our data suggest that there is a homologous chassis for a feedback loop for cell wall damage-based signaling that predates plant terrestrialization.

In land plants, the formation of LDs is known to occur under a variety of adverse environmental conditions (Gasulla et al., 2013; Mueller et al., 2015; Gidda et al., 2016). Stress-dependent formation of LDs likely evolved before land plants came to be (Li-Beisson et al., 2019; de

Vries et al., 2020; de Vries and Ischebeck, 2020), but their molecular underpinnings outside of land plants remain unclear. Here, we confirmed the identity of these *Mesotaenium* LDs using confocal microscopy, LD-specific staining and proteomics. Our comprehensive transcriptomic data illuminate co-expressed modules that might constitute a homologous program for stress-dependent LDs that acted before plants conquered land.

METHODS

Algal culturing and gradient table setup

We used the axenic and genome-sequenced *Mesotaenium endlicherianum* SAG 12.97 (https://sagdb.uni-goettingen.de/detailedList.php?str_number=12.97) from the Algal Culture Collection, Göttingen, Germany (SAG, Friedl and Lorenz 2012, SAG). *Mesotaenium* was cultivated in C-Medium (Ichimura, 1971) for an average of 12 days in an aerated culture glass flasks (SCHOTT, Germany) at 80 µmol photons m^{-2} s^{-1} 16:8 L/D cycle. Prior to the experiment, cell density was analyzed using a LUNA™ Automated Cell Counter (Logos Biosystems, Annandale, VA, USA) and set to $2.03*10^7$ cells/ml (diluting with C-Medium if needed; settings for cell counting: Cell roundness: 60%, minimum size: 3 µm, maximum size: 60 µm), corresponding to Abs680nm = 0.33 (Epoch Microplatereader, BioTek Instruments, USA). For the gradient table setup algal suspension was distributed across 504 wells (42 twelve-well plates [tissue culture testplates 12 No. 92412, TPP, Switzerland]; 2.5 mL of culture per well). Plates were sealed with Surgical tape, Micropore™ tape (3M, Germany) to minimize evaporation. The 42 twelve-well plates were then placed on a table that generates a cross-gradient of temperature (8.6±0.5 °C to 29.2±0.5 °C on the x-axis) and irradiance (21.0±2.0 to 527.9±14.0 µmol photons m-2 s-1 on the y-axis) (see Suppl. Table 1). The temperature gradient was generated using a custom-made table (Labio, Czech Republic) equipped with true-daylight LEDs (sTube 2W 120 ver 11:11, Snaggi, Czech Republic) set to a 16:8 L/D cycle (Light from 6 am to 22 pm, Central European winter time). *Mesotaenium* samples exposed to the 504 different conditions 65 hours (for sampling for RNAseq and physiological measurements) and 89 hours (for detailed light microscopy) on the gradient table. Condensed water at the top of the 12-well plates lids was removed three times in the 65 hours timespan by lightly tapping the lids twice.

Plate reader

Abs480nm, Abs680nm, and Abs750nm of all 42 plates was measured after 65h exposition (4-6 hours after light on) with an absorbance microplate reader Epoch (BioTek Instruments, VT, USA). 9 data points per well were analyzed and averaged using software Gen5 2.0 (Biotek, USA), resulting in 108 measurements per 12-well plate per wavelength. For downstream analyses these values were averaged resulting in one value per 12-well plate per wavelength (Suppl. Fig. 1). After 89 h exposition 16 plates were chosen from the prominent gradients (the four most extreme conditions in the corners and a cross of vibrant growth along the two gradients) for analyzing a full absorption spectrum (300-900nm) using the same setup. (Suppl. Fig. 9, Suppl. Table 10).

Photophysiological measurements

For maximum-quantum yield measurements (Fv/Fm) the maxi version of the IMAGING PAM (ImagMAX/L, M-series, Walz, Germany) with an IMAG-K5 CCD camera, controlled with the ImagingWinGigE (V2.32) software, was used. The *Mesotaenium* cultures in the 12-well plates were dark adapted for 10-30 min before measurement. Before measurements, the lid was removed. For the Fv/Fm measurement a short saturation pulse (Intensity 3) was applied. The measurement settings on the IMAGING PAM were the following: measuring light 1, gain 3, damping 2, mean over AOI (area of interest) was turned off. No special SP-routine was applied to modify the signal to noise ratio of the chlorophyll fluorescence measurement.

Statistical analysis of absorption and Fv/Fm values and temperature/light cluster analysis

Statistical analysis of the absorption and the Fv/Fm values was done using Kruskal-Wallis test with post hoc test Fisher's least significant difference (Conover, 1999) using R (version 4.1.3). P-values were Bonferroni corrected and grouped into significant groups using R packages ‘agricolae’ version 1.3-5 and ‘dplyr’ version 1.0.9. For heatmap generation of physiological values plotted against temperature/light R package ‘pheatmap’ version 1.0.12 was used. For cluster analysis the R package

‘factoextra’ version 1.0.7 was used. Clusters were generated using the eclust function with clustering function ‘kmeans’, with number of clusters set to six and for hierarchical clustering ‘euclidean’ was used as distance measure. Clusters were visualized with principal component analysis (PCA) in R.

RNA extraction and sequencing

After absorption measurements, the twelve-well plates were put back on the table to let cells adjust to the table conditions again for a minimum of 5 minutes before harvesting them. For RNA extraction 0.4 mL were taken from every well of the 42 twelve-well plates on the table after pipetting the cells up and down twice to homogenize them. In total 4.8 mL liquid culture was taken per condition on the table (i.e., pooling 0.4 mL of each 12 wells per each of the 42 conditions). Samples were then centrifuged for 5 min at 20 °C and 4000 rpm. The supernatant was removed and the pellet was frozen at -80 °C. To extract RNA the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich Chemie GmbH, Germany) was used according to the manufacturer’s instructions. For cell disruption samples in lysis buffer were ultrasonicated for 1 min and vortexed. RNA samples were treated with DNAse I (Thermo Fisher, Waltham, MA, USA) and shipped on dry ice to Novogene (Cambridge, UK) where they were quality checked with a Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA, USA). Libraries were built based on total RNA using poly-T oligo-attached magnetic beads. Following fragmentation, synthesis of the first strand cDNA was carried out using random hexamer primers and second strand cDNA using dUTP, instead of dTTP. A directional size-selected library was built that included PCR-based amplification. Sequencing adapters were 5' Adapter: 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3' and 3' Adapter: 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGACTATCTCGTATGCCGTCTTCTGCTTG-3'. The library was sequenced on an illumine NovaSeq6000 platform.

Quality control of reads

We checked the quality of our raw reads via FastQC (Andrews, Simon 2010) (v0.11.9) and summarized the results via MultiQC (Ewels et al. 2016) (v1.11). Based on these and the used adapter sequence, we filtered and trimmed reads via Trimmomatic (Bolger, Lohse, and Usadel 2014) (v 0.36) with these parameters: ("ILLUMINACLIP: novogene_adapter_sequences_Trimmomatic.fa:2:30:10:2:True LEADING:26 TRAILING:26 SLIDINGWINDOW:4:20 MINLEN:36"). We checked the quality of the trimmed reads with FastQC and MultiQC again.

Genome annotation

The original annotation of *M. endlicherianum* (Cheng et al., 2019) had a lower number of genes compared to other Zygnematophyceae algae. We took advantage of our newly generated RNAseq dataset to improve genome annotation. Trimmed reads were mapped via HISAT2 (Pertea et al. 2016, 2) and assembled via StringTie (Pertea et al. 2016, 2). StringTie results showed many novel isoforms as well as novel transcripts. We also used BUSCO v5 (Manni et al. 2021) to measure the completeness of the gene models in annotation V1 independent of StringTie. Although the gene prediction method which used by BUSCO at the genome level is very efficient, it is not unexpected if it misses some proteins that were annotated in a genome via experimental, based on bioinformatic methods and NGS data, or ab-initio based gene prediction methods. Therefore, we expect that the BUSCO score based on the proteins of a gene model should be equal to or greater than the BUSCO score of the genome. When we compared the BUSCO score between the genome and protein sequences for *M. endlicherianum* with “viridiplantae.odb.10-2020-09-10”, we noticed that they show similar numbers (Suppl. Fig. 2). Therefore, we decided to re-annotate the genome of *M. endlicherianum* with our comprehensive RNA-Seq datasets as well as public protein and genome sequences published for its close relatives.

We annotated the *M. endlicherianum* genome using REAT (v0.6.1). Various gene models were predicted based on different types of evidence and methods. The final gene models and annotation V2 were based on agreement with the experimental evidence. At the end, we tried to quantify “completeness” and quality of the new annotation V2 and the old V1.

First, we used RNAseq evidence with REAT's "Transcriptome Workflow" with HISAT2 (v2.2.1), Scallop (Shao and Kingsford 2017) (v0.10.5) and StringTie (v2.1.5). We also used Portcullis (Mapleson et al. 2018) (v1.2.4) to identify genuine junctions based on short reads alignments. This workflow uses Mikado (Venturini et al. 2018) (v2.3.4) to identify the "best" set of transcripts from multiple transcript assemblies.

Then, we used gene homology information from representative streptophytes in REAT's "Homology Workflow". PALN (Gotoh 2008a; 2008b) (v2.4.7) was used to align representative protein sequences onto the *M. endlicherianum* genome. The representative dataset consisted on genome, gene models, and protein sequences of *Anthoceros agrestis* (Oxford strain) (Li et al. 2020), *Arabidopsis thaliana* (C.-Y. Cheng et al. 2017), *Azolla filiculoides* (Li et al. 2018), *Chara braunii* (Nishiyama et al. 2018), *Chlorokybus melkonianii* (Wang et al. 2020), *Chlamydomonas reinhardtii* (Merchant et al. 2007) (v5.6), *Klebsormidium nitens* (Hori et al. 2014), *Mesostigma viride* (Liang et al. 2019), *Marchantia polymorpha* (Montgomery et al. 2020) (v6.1r1), *Penium margaritaceum* (Jiao et al. 2020), *Physcomitrium patens* (Lang et al. 2018) (v3.3), *Selaginella moellendorffii* (Banks et al. 2011), and *Spirogloea muscicola* (S. Cheng et al. 2019). We also used the junction file produced by Portcullis. Since there were no close relatives of *M. endlicherianum* on the SPALN species-specific parameter set, we used three different closest possibilities (Angiosp, Chlospec, and MossWorts) and built three models. These alignments are filtered using a predefined set of criteria (*cf.* code on GitHub) including exon length, intron length, internal stop codon, among others. The final gene models of V2 were prepared by Mikado.

Afterwards, we used REAT's "Prediction Workflow" to predict gene models *ab initio* and based on RNAseq evidence. This uses Augustus (Stanke et al. 2006; Stanke, Tzvetkova, and Morgenstern 2006; Hoff and Stanke 2019) (v 3.4.0), SNAP (Korf 2004) (version 2006-07-28), Glimmer (Kelley et al. 2012) (v0.3.2), and CodingQuarry (Testa et al. 2015) (v2.0), which generate different gene models as the raw material for EvidenceModeler (Haas et al. 2008) (v1.1.1) that chooses the best set of exons and combine them in a gene model using weights (see GitHub) that could be adjusted for each sort of prediction and evidence. To include UTRs where possible, the EVM output is then processed by Mikado using UTR-containing gene models from the transcriptome and homology workflows as inputs, as well gene models classified by REAT as gold, silver, and bronze based on their agreement with the set of protein sequences from other streptophyte genomes (streptophyte algae and land plants), transcriptome alignment, homology alignment, and junctions. To train *ab initio* predictors, a user-defined number of models are randomly chosen in a user-defined ratio between (10%) mono-exonic and (90%) multi-exonic. These models were chosen from best classified models (gold and silver). For Augustus, we performed meta parameter optimization and train a model with kfold=8. Beside *ab initio* predictions, we used Augustus to predict gene models with three different weights for each evidence type as suggested by Augustus authors.

At last, we used Minos ("Minos - a Gene Model Consolidation Pipeline for Genome Annotation Projects" [2019] 2022) which is gene model consolidation pipeline and produces external metrics based on DIAMOND "BLASTp/BLASTx" (Buchfink, Xie, and Huson 2015), Kallisto (Bray et al. 2016) (v0.46.2) expression quantification, coding potential calculator (CPC2 v0.1) (Kang et al. 2017, 2) and BUSCO assessments. These metrices pass through Mikado in combination with various gene models produced with different methods (as mentioned above), Minos determines the best gene model for each region based on user defined criteria (for details, see GitHub) and external metrics. Minos also put a tag on each gene model to categorize them based on a user defined threshold (we used default values) for sequence similarity coverage of homologs, BUSCO score, CPC score, TPM expression, and transcript score into "high confidence", "low confidence", and "predicted genes".

Genome annotation assessment

We used two methods to compare the quality of the new gene model with the published one. We compared the BUSCO scores of the annotated protein sequences as well as genome sequence using the reference "viridiplantae.odb.10-2020-09-10" dataset. We also used maker (Campbell et al. 2014) (v3.01.04) to calculate the AED (Eilbeck et al. 2009) to evaluate the agreement of the gene models with external evidences. Maker-P was used to build the *M. endlicherianum* gene model v1.

Further, we used the maker package to perform functional annotation via InterProScan and BLAST using agat (Dainat 2020) package (v0.9.2). Additionally, we performed a BLAST search

against A. thaliana protein sequences (Araport11) and reported the best hit for each sequence in (Suppl. Table 11) and used eggNOGmapper (Huerta-Cepas et al. 2017; 2019) (v2.1.8) to perform functional annotation. We used DIAMOND (Buchfink, Xie, and Huson 2015) (v2.0.15) with ultra-sensitive mode, *e* value cutoff of $1e^{-7}$ and in an iterative manner. We used the protein sequences as our inputs and Viridiplantae (33090) as our taxonomy scope.

RNA-Seq analysis: Pseudoalignment

In order to quantify gene expression, we used Kallisto (Bray et al. 2016) (v0.45.0). We indexed the transcriptome file with --kmer-size=31 parameter and used --bootstrap-samples 100 and --rf-stranded to quantify gene expression based on pseudoaligned reads. We used MultiQC to obtain an overview of alignment for each condition.

Filtering, normalization, modeling mean-variance relationship, and data exploration

Kallisto quantification files were imported into R (v4.2.0) with tximport (Soneson, Love, and Robinson 2016) (v1.24.0) to calculate the counts from abundance via "lengthScaledTPM" based on our study design file (Suppl. Table 12). We used edgeR (Robinson, McCarthy, and Smyth 2010) (v3.38.1) for filtering and TMM-normalization (Robinson and Oshlack 2010) of the reads (genes with >1 count per million (CPM) at log2 scale in a least 3 samples—the number of replicates—were kept). Then, we used the voom function from limma (Ritchie et al. 2015; Phipson et al. 2016; Law et al. 2014; Liu et al. 2015) (v3.52.2) to model mean-variance relationship. The normalized expression table on the log2 scale is available in (Suppl. Table 13). We performed principal component analysis based on the expression table output of voom and visualized the result with ggplot2 (Ggplot2 n.d., 2) (v3.3.6). We visualized the heatmap of distance and Spearman correlation between all samples considering all genes via pheatmap (v1.0.12) and calculated clusters via the Euclidian method.

RNA-Seq analysis: Weighted gene co-expression network analysis

We used WGCNA (Langfelder and Horvath 2008; 2012) package (v1.71) with the expression table produced by limma . We checked for and filtered out outliers as suggested by WGCNA authors (Suppl. Fig. 10). Then, we visualized the scale free topology model fit (R^2) against the soft thresholds (ßs) to pick a ß for our network construction (Suppl. Fig. 11). We used signed network type and bicor as our correlation function for WGCNA. Based on these results, we picked 16 as our soft threshold 'ß'. We experimentally chose a merging threshold if 0.25 after exploring different values from 0.2 to 0.4 and investigating the relationship between eigengenes and temperature, light intensity, FvFm, and absorption (Suppl. Fig. 12). We built the gene co-expression network using a merging threshold of 0.25 for modules, maximum portion of outliers as 0.05 and minimum module size of 30. Then, we visualized the correlation between each module eigengene and temperature, light intensity, FvFm, and absorption to identify which modules are more related to each treatment (Suppl. Fig. 13). We provided a table for all genes, their module assignment, inter- and intramodular connectivity, gene significance for temperature and light intensity, correlation with temperature and light intensity, and their module membership (aka. Signed eigengene-based connectivity) in (Suppl. Table 5). We also visualized the graphical representation of the topological overlap matrix of our samples (Suppl. Fig. 14). In order to have a visual representation of gene expression in each module, we drew heatmaps for each module via pheatmap (using Euclidean method for calculating the distance and complete method clustering) (Suppl. Fig. 15). GO enrichment analysis was performed via clusterProfiler package (Yu et al. 2012; Wu et al. 2021) (v4.4.4) using the output of eggNOGmapper and adjusted p-value cut-off 0.05 and q-value cut-off of 0.05, considering only genes that are present in our GO term-to-gene table which was expressed as our background gene universe (Suppl. Table 6). Determining the proper background gene list has significant importance in enrichment analysis (Wijesooriya et al. 2022).

To see how *A. thaliana*'s well-known genes in stress-response mechanisms (downloaded from TAIR database via keyword search) were distributed across different modules we performed BLASTp searches against the new *M. endlicherianum* annotated proteins. We visualized the distribution of these IDs for different stress-related keywords in (Suppl. Fig. 16) and the expression of these genes across different samples via pheatmap (Suppl. Fig. 17). We defined as module hubs the top 20 genes (nodes) with the highest connectivity within each module (Suppl. Table 5 and 14).

Differential gene expression analysis

We performed differential gene expression analysis using the limma package. We divided samples into multiple groups as follows: low light intensity (21 and 39 µmol photons m^{-2} s^{-1}), medium light intensity (72 and 129 µmol photons m^{-2} s^{-1}), high light intensity (329 and 527 µmol photons m^{-2} s^{-1}), low temperature (8 °C and 12 °C), medium temperature (17 °C, 20 °C, and 23 °C), high temperature (26 °C and 29 °C; see grid/colored table layout in Figure 3). We performed all-against-all comparisons and an additional comparison of those samples from an Fv/Fm < 0.5 versus low light intensity + medium temperature. We used duplicateCorrelation as suggested by Smyth et al. (2005) to consider technical replicates. We used clusterprofiler for GO-enrichment analysis (Wu et al. 2021) with adjusted p-value and q-value cutoff of 0.01 and only genes that were expressed and passed filtering as our background universe. The heatmap of gene expression profiles, dot plot and cnetplot of enriched GO-terms for each comparison is available in ((Suppl. Table 15 and Suppl. Fig. 18 to 26).

Differential interference contrast and confocal laser scanning microscopy

Differential interference contrast (DIC) imaging was done for all replicates from the table with a Olympus BX-60 microscope (Olympus, Japan) with a ProgRes C14plus camera and the ProgRes® CapturePro Software (version 2.9.01) (JENOPTIK AG, Jena, Germany). The morphology of chosen conditions (see Supplemental Figure 1) of *Mesotaenium* cells that were 89 h on the table was analyzed.

For algae that were used for quantifying the abundance of lipid droplet per cell, a ZEISS Axioscope 7 microscope (Carl Zeiss, Germany) was used including the ZEN software (Carl Zeiss, Germany). Lipid droplet count was carried out in FIJI (Schindelin et al., 2012). For statistical analysis of the lipid droplet count data, we first used a Shapiro-Wilk test (Shapiro and Wilk, 1965) to assess normality and used Mann-Whitney U tests (Mann and Whitney, 1947) with R (version 3.6.1) accordingly.

Confocal laser scanning microscope was done on a Zeiss LSM780 (Carl Zeiss) set as in Müller et al. (2017). For the staining of the LD structures, we used the neutral lipid specific stain BODIPY™ 493/503 (EM/EX) (Merck). *Mesotaenium* cells were grown for 22 days on WHM-medium at 70-80 µmol photons m-2 s-1 and 22°C. These cells were ultrasonicated for 1 min with 1:500 BODIPY and incubated on a shaker for 5 min before visualization.

Lipid droplet isolation and proteomics

For lipid droplet isolation 23 days old *Mesotaenium* cells grown on WHM-Medium at 70-80 µmol photons m^{-2} s^{-1} and 22 °C were homogenized using a Tenbroeck or potter homogenizer in lipid droplet isolation buffer (10 mM sodium phosphate buffer pH 7.5, 200 µM PMFS, 0.5 mM DSP, 10 mM N-Ethylmaleimide). The resulting centrifuged supernatant of a 100 x g spin for 1 min was considered as total extract (TE). After two further high speed centrifugations (SW40 Ti for 1h, 4°C at 100000 x *g*, TLA120 for 1h at 100000 x *g* and 4°C) the floating fat pad was precipitated at -20 °C using 100% ethanol overnight. The precipitated pellet was washed with 80% ethanol twice, dried and then suspended in 6M urea. Protein concentration was determined using BCA. An in-gel SDS gel digestion was done with trypsin adapted from Shevchenko et al. (1996). C18 Stage tip purification was done according (Rappsilber et al., 2003; 2007). Protein samples were analyses using LC-MS. For this, peptide samples were reconstituted in 20 µl LC-MS sample buffer (2% acetonitrile, 0.1% formic acid). 2 µl of each sample were subjected to reverse phase liquid chromatography for peptide separation using an RSLCnano Ultimate 3000 system (Thermo Fisher Scientific). Therefore, peptides were loaded on an Acclaim PepMap 100 pre-column (100 µm x 2 cm, C18, 5 µm, 100 Å; Thermo Fisher Scientific) with 0.07% trifluoroacetic acid at a flow rate of 20 µL/min for 3 min. Analytical separation of peptides was done on an Acclaim PepMap RSLC column (75 µm x 50 cm, C18, 2 µm, 100 Å; Thermo Fisher Scientific) at a flow rate of 300 nL/min. The solvent composition was gradually changed within 94 min from 96 % solvent A (0.1 % formic acid) and 4 % solvent B (80 % acetonitrile, 0.1 % formic acid) to 10 % solvent B within 2 minutes, to 30 % solvent B within the next 58 min, to 45% solvent B within the following 22 min, and to 90 % solvent B within the last 12 min of the gradient. All solvents and acids had Optima grade for LC-MS (Thermo Fisher Scientific). Eluting peptides were on-line ionized by nano-electrospray (nESI) using the Nanospray Flex Ion Source (Thermo Fisher Scientific) at 1.5 kV (liquid junction) and transferred into a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Full scans in a mass range of 300 to 1650 m/z were

recorded at a resolution of 30,000 followed by data-dependent top 10 HCD fragmentation at a resolution of 15,000 (dynamic exclusion enabled). LC-MS method programming and data acquisition was performed with the XCalibur 4.0 software (Thermo Fisher Scientific). Afterwards the raw proteome data were analyzed using Max Quant software version 1.6.2.10 (Cox and Mann, 2008). The database for this analysis was our new V2 gene model data. The data were then further processed by the Perseus (1.6.2.2) software (Cox et al., 2008; Tyanova et al., 2016).

Data availability

All RNAseq reads have been uploaded to NCBI SRA and can be accessed under Bioproject PRJNA832564 and SRA accessions SRR18936040 to SRR18936170. Codes and Data used for genome re-annotaiton, WGCNA and differential gene expression analysis are available on our GitHub page https://github.com/deVries-lab/Response_to_a_gradient_of_environmental_cues_in_mesotaenium_endlicherianum. Proteomic data have been uploaded to PRIDE. Furthermore, data can be interactively explored at https://mesotaenium.uni-goettingen.de

ACKNOWLEDGEMENT

We thank René Heise for excellent technical support. J.d.V. thanks the European Research Council for funding under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 852725; ERC-StG "TerreStriAL"). J.d.V., U.H., and H.B. are grateful for support through the German Research Foundation (DFG) within the framework of the Priority Programme "MAdLand – Molecular Adaptation to Land: Plant Evolution to Change" (SPP 2237; VR 132/4-1; BU 2301/6-1), in which T.R. is a PhD student and A.D., J.M.R.F.-J, and I.I. partake as associate members. A.D. is grateful for being supported through the International Max Planck Research School (IMPRS) for Genome Science. J.M.R.F.-J. and T.R. gratefully acknowledge support by the Ph.D. program "Microbiology and Biochemistry" within the framework of the "Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences" (GGNB) at the University of Goettingen. M.M. is supported by Singaporean Ministry of Education grant T2EP30122-0001. P.S. is grateful for support from the Studienstiftung des Deutschen Volkes.

CONTRIBUTIONS

J.d.V. and M.L. conceived the project. J.d.V. coordinated the project with M.M. M.L. provided plant materials. J.M.R.F.-J., T.D., and T.R. performed experimental work. A.D. carried out computational analysis. O.V., J.M.R.F.-J., P.S., T.I., D.K. and G.H.B. performed proteomics. H.B. investigated cell division patterns. M.H. and U.H. investigated photomorphogenesis patterns. A.D. and R.S. built web resources. J.d.V., A.D., and J.M.R.F.-J. contributed to writing the manuscript. J.d.V. organized the manuscript. All authors commented, discussed, and provided input on the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL FIGURES

Supplemental Figure 1. Fv/Fm and absorption values of all replicates of gradient tables; representative micrographs of the most extreme corners and of vividly growing algae along the two gradients.

Supplemental Figure 2. BUSCO comparison between genome, protein sequences v1, protein sequences v2

Supplemental Figure 3. Cumulative fraction of annotation vs AED plot for gene model v1 and v2

Supplemental Figure 4. Module membership versus Gene Significance for genes in different modules with respect to FvFm

Supplemental Figure 5. Module membership versus Gene Significance for genes in different modules with respect to Temperature

Supplemental Figure 6. Module membership versus Gene Significance for genes in different modules with respect to light intensity

Supplemental Figure 7. Heatmap of the correlation between module eigengenes and light intensity, temperature, absorption, replicate, and Fv/Fm as well as student test p-value

Supplemental Figure 8. The GLK alignment
Supplemental Figure 9. Absorption spectra of all replicates at chosen conditions.
Supplemental Figure 10. Sample dendrogram and trait heatmap to identify outliers for WGCNA
Supplemental Figure 11. Picking a soft threshold for WGCNA based on scale independence and Mean connectivity
Supplemental Figure 12. Clustering of different modules and traits based for identifying a merging threshold
Supplemental Figure 13. Module-trait relationship between eigengenes and light intensity, temperature, FvFm, and absorption
Supplemental Figure 14. The graphical representation of the topological overlap matrix
Supplemental Figure 15. Heatmap of gene expression Z-score values for each module
Supplemental Figure 16. Distribution of best blast hit of *A. thaliana* stress response genes among WGCNA modules
Supplemental Figure 17. Heatmap of best blast hit of *A. thaliana* stress response genes in *M. endlicherianum* across different growth conditions
Supplemental Figure 18. Dotplot, cnetplot and heatmaps of DEGs comparing FvFm control vs stress
Supplemental Figure 19. Dotplot, cnetplot and heatmaps of DEGs comparing HLI_HT vs LLI_MT
Supplemental Figure 20. Dotplot, cnetplot and heatmaps of DEGs comparing MLI_HT vs LLI_MT
Supplemental Figure 21. Dotplot, cnetplot and heatmaps of DEGs comparing LLI_HT vs LLI_MT
Supplemental Figure 22. Dotplot, cnetplot and heatmaps of DEGs comparing HLI_MT vs LLI_MT
Supplemental Figure 23. Dotplot, cnetplot and heatmaps of DEGs comparing MLI_MT vs LLI_MT
Supplemental Figure 24. Dotplot, cnetplot and heatmaps of DEGs comparing HLI_LT vs LLI_MT
Supplemental Figure 25. Dotplot, cnetplot and heatmaps of DEGs comparing MLI_LT vs LLI_MT
Supplemental Figure 26. Dotplot, cnetplot and heatmaps of DEGs comparing MLI_LT vs LLI_MT
Supplemental Figure 27. Fully-labeled phylogenies of hub genes.
Supplemental Figure 28. Lipid droplet count setup 2.
Supplemental Figure 29. LDAP phylogeny.

SUPPLEMENTAL TABLES

Supplemental Table 1. Temperature and light intensity measurements of all 504 coordinates on the gradient table.
Supplemental Table 2. All 504 Fv/Fm and absorption measurements of all replicates.
Supplemental Table 3. Number of genes and transcripts in gene model V2
Supplemental Table 4. The general stats of raw reads, trimmed reads, and pseudoalignment
Supplemental Table 5. Summary of WGCNA Results
Supplemental Table 6. The results of GO-enrichment analysis for all modules of WGCNA
Supplemental Table 7. The list of top20 hubs for each module
Supplemental Table 8. Counts of lipid droplets in micrographs.
Supplemental Table 9. Full proteomic results, showing Mesotaenium gene model V2 identifiers, Arabidopsis gene identifers, and IBAQ values.
Supplemental Table 10. All data on absorption spectra of all replicates at chosen conditions.
Supplemental Table 11. Best blast hit of *M. endlicherianum* gene model against *A. thaliana* (Araport11)
Supplemental Table 12. Study design file used for RNASeq analysis
Supplemental Table 13. The CPM normalized expression table on the log2 scale
Supplemental Table 14. The hubs (top 20 mostly connected genes) of WGCNA modules
Supplemental Table 15. The GO-enrichment results of 9 pairwise comparisons

FIGURE LEGENDS

Figure 1: A fine-combed setup for assessing environmental cues in Mesotaenium. (a) Cladogram of Streptophyta, highlighting that *Meostaenium enlicherianum* SAG 12.97 is a representative of the closest algal relatives of land plants. (b) *Meostaenium enlicherianum* grown in C-medium in 42 twelve-well plates on a gradient table that produces a temperature range of 8.6±0.5 °C to 29.2±0.5 °C

on the x-axis and an irradiance gradient of µmol photons m-2 s-1 from 21.0±2.0 to 527.9±14.0 on the y-axis. (c) Overview of the measured physiological values Fv/Fm (blue) and Absorption (Abs.) at 480 (orange) and 680 nm (green). Three biological triplicates for the respective measurements are shown on the right. (d) Statistical analysis of the physiological values (Fv/Fm, Abs. 480 nm, Abs. 680 nm). Numbers correspond to environmental conditions on the table. Biological triplicates were grouped in significant groups (a-u) with R (version 4.1.3) using a Kruskal-Wallis test coupled with Fisher's least significance, *p* values were Bonferroni corrected. Significant differences at $p \leq 0.001$ are shown as letters. (e) Heatmaps displaying either temperature or light sorted averaged physiological values of the 42 conditions. A cut-off was set based on the distribution of the highest values, which were then summed up to determine a positive correlation with temperature or light conditions. (f) Two principal component analyses (PCA) showing the correlation of light conditions (left) or temperature conditions (right) to physiological values (Fv/Fm, Abs. 480, 680 nm). Clusters are shown in different colors, which are also visualized in an overview scheme of the gradient table at the top right of the plots.

Figure 2: Global profiles of environment-governed gene expression response. (a) Principal component analysis (PCA), visualizing PC1 and PC2. Backgrounds were drawn to highlight our interpretation of the observed trends; samples are coded by color (temperature) and symbols (irradiance in µmol photons m^{-2} s^{-1}). (b) Visualization of Euclidean distance between samples via heatmap. (c) Heatmap of Spearman correlation between samples. The clusters were calculated via the Euclidean distance. (d) PC1 and PC2 small multiples of light intensity and (e) temperature. In (d) shades of gray corresponds to different light intensities. In (e) different colors represent different temperatures and were mapped with the same colors as (a).

Figure 3: Stress-titratable global differential gene expression profiles. Linear model were fitted for each gene and empirical Bayes statistics computed for differentially expressed genes (DEGs) by limma package. DEGs defined as genes with absolute fold change greater than 2 and adjusted p-value lesser than 0.01. (a) Volcano plots of DEGs for 9 different comparisons. (b) Heatmaps of Z-score values of DEGs for each comparison; grey bars label the first component (treatment) for calculating the contrasts (treatment vs. control). (c) Biological theme comparison, GO-term enrichment analysis with adjusted p-value ≤ 0.01 of DEGs against all genes that were expressed and passed the filtering in our analyses as background. The size of each circle is proportional to the count of each term. Only the top 30 enriched terms are shown. (d) Wordle of the 124 genes that showed significant regulation across multiple comparisons shown in Figure 3a.

Figure 4: Unsupervised gene expression clusters recover genetic programs separated by environmental cues. Gene expression clustering into 26 colored modules was performed using WGCNA; grey is the module of unclustered genes. (a) Hierarchical cluster tree of 17,095 genes. The heat map below the dendogram shows the gene significance measure (from red, positive correlation, to white, no correlation, to blue, negative correlation) for the four different conditions / physiological parameters. (b) Heat map of the module–trait correlation based on eigengenes (from red, positive correlation, to white, no correlation, to blue, negative correlation); see supplemental figure S_D (c) Bar plots of the mean gene significance across modules (given in the corresponding module color) towards the parameters light intensity, temperature, and Fv/Fm. (d) Enriched GO-terms for eight of the 26 modules; each inset shows the gene expression profiles of all genes in a given module. (e) . Abbreviations: proc. = process; reg. = regulation; biogen. = biogenesis; develop. = development; pos. = positive; neg. = negative; init. = initiation; GEP = Gene expression profile; med. = mediated; dep. = dependent; modif. = modification; conjug. = conjugation; anneal. = annealing; compl. = complex; synth. = synthesis; resp. = response; transf. = transferring.

Figure 5: Molecular programs for environmental responses around recurrent plant hubs. Visualization of the co-expression networks clustered by WGCNA into the modules blue (3101 genes), yellow (1427 genes), green (1220 genes), purple (506 genes), and pink (718 genes). Nodes, circles representing genes, are connected by edges whose weight is based on a weighted topological overlap matrix (TOM); weight is shown in a color gradient from light to dark indicating low to high topological overlap values in the TOM. Brightly colored nodes represent the 20 most connected genes

(hubs) and are annotated; all other nodes are depicted in the corresponding paler color. Hubs are annotated based on homology. Around the clusters, different protein-coding hub genes are highlighted, giving information such as predicted domain structures or phylogenetic relationships; fully-labelled phylogenies are deposited in Supp. Fig. 27. Circles in phylogenies represent ultrasfast bootstrap support, with larger circles represent high/full support; diamond symbols indicate high (>90) support for branches separating highlighted clades. An alignment of GLK homologs can be found in Suppl. Fig. 8.

Figure 6: Lipid droplets accumulate in *Mesotaenium* upon changing environments. (a) Differential interference contrast (DIC) and confocal laser scanning micrographs of *Mesotaenium endlicherianum* SAG 12.97 cells accumulating lipid droplets (LDs; arrows). Cells were either subjected to different temperature/light conditions (see abbreviations below) of the gradient table for 89 h or 216 h. For confocal microscopy, algae were cultured independent of table conditions at 75 µmol photons m-2 s-1 and 22°C for 22 days. LDs are visible as distinct globular structures in NIC and were stained by the lipid stain BODIPY^{TM} (false-colored green; 493 nm excitation; 503 nm emission) and chlorophyll autofluorescence (false-colored purple). (b) Violin plots of LD quantification after 9 days of exposure to different environmental conditions including statistical analysis using Mann-Whitney *U* statistics (significance grouping based on p value < 0.05; see also Suppl. Fig. 28) (c) Heat map of row-scaled z-scores of the expression of homologs for LD biogenesis and function (see also Suppl. Fig. 29). Temperature/light conditions are displayed at the bottom as symbols in different colors, *Arabidopsis thaliana* (AT) identifiers based on BLASTp search are shown on the right (d) Proteomic investigation into a lipid enriched phase extracted from *Mesotaenium endlicherianum* SAG 12.97 cells showing enrichment in hallmark proteins of LDs. Volcano plot showing significantly (FDR<0.05) enriched *Mesotaenium* proteins in the lipid enriched phase (right side) compared to proteins of the total extract (left side). Hallmark *A. thaliana* LD protein identifiers are annotated based on BLAST. Top bar on the left plot shows the IBAQ (per mille) values for ten LD signature protein detected in *Mesotaenium*. Bottom bar plot shows the log2 enrichment of proteins characteristic for sub-cellular compartments. (LL = Low light, LT = low temperature, ML = moderate light, MT = moderate temperature, HL = high light, HT = high temperature).

REFERENCES

Andrews, Simon. 2010. "FastQC: A Quality Control Tool for High Throughput Sequence Data." Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Banks, Jo Ann, Tomoaki Nishiyama, Mitsuyasu Hasebe, John L. Bowman, Michael Gribskov, Claude dePamphilis, Victor A. Albert, et al. 2011. "The Selaginella Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants." Science 332 (6032): 960–63. https://doi.org/10.1126/science.1203810.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." Bioinformatics 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. "Near-Optimal Probabilistic RNA-Seq Quantification." Nature Biotechnology 34 (5): 525–27. https://doi.org/10.1038/nbt.3519.

Buchfink, Benjamin, Chao Xie, and Daniel H. Huson. 2015. "Fast and Sensitive Protein Alignment Using DIAMOND." Nature Methods 12 (1): 59–60. https://doi.org/10.1038/nmeth.3176.

Campbell, Michael S., Carson Holt, Barry Moore, and Mark Yandell. 2014. "Genome Annotation and Curation Using MAKER and MAKER-P." Current Protocols in Bioinformatics 48 (1): 4.11.1-4.11.39. https://doi.org/10.1002/0471250953.bi0411s48.

Cheng, Chia-Yi, Vivek Krishnakumar, Agnes P. Chan, Françoise Thibaud-Nissen, Seth Schobel, and Christopher D. Town. 2017. "Araport11: A Complete Reannotation of the Arabidopsis Thaliana Reference Genome." The Plant Journal 89 (4): 789–804. https://doi.org/10.1111/tpj.13415.

Cheng, Shifeng, Wenfei Xian, Yuan Fu, Birger Marin, Jean Keller, Tian Wu, Wenjing Sun, et al. 2019. "Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution." Cell 179 (5): 1057-1067.e14. https://doi.org/10.1016/j.cell.2019.10.019.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26, 1367-1372.

Dainat, J. 2020. "AGAT: Another Gff Analysis Toolkit to Handle Annotations in Any GTF/GFF Format." Version V0 4: 10–5281.

Doner, N. M. *et al.* Arabidopsis thaliana EARLY RESPONSIVE TO DEHYDRATION 7 Localizes to Lipid Droplets via Its Senescence Domain. *Front. Plant Sci.* **12**, 658961 (2021).

Eilbeck, Karen, Barry Moore, Carson Holt, and Mark Yandell. 2009. "Quantitative Measures for the Management and Comparison of Annotated Genomes." BMC Bioinformatics 10 (1): 67. https://doi.org/10.1186/1471-2105-10-67.

Ewels, Philip, Måns Magnusson, Sverker Lundin, and Max Käller. 2016. "MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report." Bioinformatics 32 (19): 3047–48. https://doi.org/10.1093/bioinformatics/btw354.

Friedl, T., Lorenz, M., The Culture Collection of Algae at Göttingen University (SAG): a biological resource for biotechnological and biodiversity research. (2012) Procedia Environ Sci 15:110–117.

Ggplot2. n.d. Accessed July 20, 2022. https://link.springer.com/book/10.1007/978-3-319-24277-4.

Gidda, S. K. *et al.* Lipid Droplet-Associated Proteins (LDAPs) Are Required for the Dynamic Regulation of Neutral Lipid Compartmentation in Plant Cells. *Plant Physiol.* **170**, 2052–2071 (2016).

Gotoh O. 2008a. "A Space-Efficient and Accurate Method for Mapping and Aligning CDNA Sequences onto Genomic Sequence." Nucleic Acids Research 36 (8): 2630–38. https://doi.org/10.1093/nar/gkn105.

Gotoh O. 2008b. "Direct Mapping and Alignment of Protein Sequences onto Genomic Sequence." Bioinformatics 24 (21): 2438–44. https://doi.org/10.1093/bioinformatics/btn460.

Haas, Brian J., Steven L. Salzberg, Wei Zhu, Mihaela Pertea, Jonathan E. Allen, Joshua Orvis, Owen White, C. Robin Buell, and Jennifer R. Wortman. 2008. "Automated Eukaryotic Gene Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced Alignments." Genome Biology 9 (1): R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Higashi, Y., Okazaki, Y., Myouga, F., Shinozaki, K. & Saito, K. Landscape of the lipidome and transcriptome under heat stress in Arabidopsis thaliana. *Sci Rep* **5**, 10533 (2015).

Hoff, Katharina J., and Mario Stanke. 2019. "Predicting Genes in Single Genomes with AUGUSTUS." Current Protocols in Bioinformatics 65 (1): e57. https://doi.org/10.1002/cpbi.57.

Hori, Koichi, Fumito Maruyama, Takatomo Fujisawa, Tomoaki Togashi, Nozomi Yamamoto, Mitsunori Seo, Syusei Sato, et al. 2014. "Klebsormidium Flaccidum Genome Reveals Primary Factors for Plant Terrestrial Adaptation." Nature Communications 5 (1): 3978. https://doi.org/10.1038/ncomms4978.

Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K Forslund, Helen Cook, Daniel R Mende, et al. 2019. "EggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses." Nucleic Acids Research 47 (D1): D309–14. https://doi.org/10.1093/nar/gky1085.

Huerta-Cepas, Jaime, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. 2017. "Fast Genome-Wide Functional Annotation through Orthology Assignment by EggNOG-Mapper." Molecular Biology and Evolution 34 (8): 2115–22. https://doi.org/10.1093/molbev/msx148.

Ichimura, T. 1971 Sexual cell division and conjugation-papilla formation in sexual reproduction of Closterium strigosum. In Proceedings of the Seventh International Seaweed Symposium, University of Tokyo Press, Tokyo, p. 208-214

Jiao, Chen, Iben Sørensen, Xuepeng Sun, Honghe Sun, Hila Behar, Saleh Alseekh, Glenn Philippe, et al. 2020. "The Penium Margaritaceum Genome: Hallmarks of the Origins of Land Plants." Cell 181 (5): 1097-1111.e12. https://doi.org/10.1016/j.cell.2020.04.019.

Kang, Yu-Jian, De-Chang Yang, Lei Kong, Mei Hou, Yu-Qi Meng, Liping Wei, and Ge Gao. 2017. "CPC2: A Fast and Accurate Coding Potential Calculator Based on Sequence Intrinsic Features." Nucleic Acids Research 45 (W1): W12–16. https://doi.org/10.1093/nar/gkx428.

Kelley, David R., Bo Liu, Arthur L. Delcher, Mihai Pop, and Steven L. Salzberg. 2012. "Gene Prediction with Glimmer for Metagenomic Sequences Augmented by Classification and Clustering." Nucleic Acids Research 40 (1): e9. https://doi.org/10.1093/nar/gkr1067.

Kleine, T. et al. Acclimation in plants – the Green Hub consortium. Plant J 106, 23–40 (2021).

Korf, Ian. 2004. "Gene Finding in Novel Genomes." BMC Bioinformatics 5 (1): 59. https://doi.org/10.1186/1471-2105-5-59.

Krawczyk, H. E. *et al.* Heat stress leads to rapid lipid remodeling and transcriptional adaptations in *Nicotiana tabacum* pollen tubes. *Plant Physiology* kiac127 (2022) doi:10.1093/plphys/kiac127.

Kretzschmar, F. K., Doner, N. M., Krawczyk, H. E., Scholz, P., Schmitt, K., Valerius, O., Braus, G. H., Mullen, R. T., and Ischebeck, T. (2020) Identification of Low-Abundance Lipid Droplet Proteins in Seeds and Seedlings. Plant Physiol. 182, 1326–1345

Lang, Daniel, Kristian K. Ullrich, Florent Murat, Jörg Fuchs, Jerry Jenkins, Fabian B. Haas, Mathieu Piednoel, et al. 2018. "The Physcomitrella Patens Chromosome-Scale Assembly Reveals Moss Genome Structure and Evolution." The Plant Journal 93 (3): 515–33. https://doi.org/10.1111/tpj.13801.

Langfelder, Peter, and Steve Horvath. 2008. "WGCNA: An R Package for Weighted Correlation Network Analysis." BMC Bioinformatics 9 (1): 559. https://doi.org/10.1186/1471-2105-9-559.

Langfelder, Peter, and Steve Horvath. 2012. "Fast R Functions for Robust Correlations and Hierarchical Clustering." Journal of Statistical Software 46 (March): 1–17. https://doi.org/10.18637/jss.v046.i11.

Lass A, et al. (2006) Adipose triglyceride lipase-mediated lipolysis of cellular fat stores is activated by CGI-58 and defective in Chanarin-Dorfman Syndrome. Cell Metab 3: 309–319.

Law, Charity W., Yunshun Chen, Wei Shi, and Gordon K. Smyth. 2014. "Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts." Genome Biology 15 (2): R29. https://doi.org/10.1186/gb-2014-15-2-r29.

Li, Fay-Wei, Paul Brouwer, Lorenzo Carretero-Paulet, Shifeng Cheng, Jan de Vries, Pierre-Marc Delaux, Ariana Eily, et al. 2018. "Fern Genomes Elucidate Land Plant Evolution and Cyanobacterial Symbioses." Nature Plants 4 (7): 460–72. https://doi.org/10.1038/s41477-018-0188-8.

Li, Fay-Wei, Tomoaki Nishiyama, Manuel Waller, Eftychios Frangedakis, Jean Keller, Zheng Li, Noe Fernandez-Pozo, et al. 2020. "Anthoceros Genomes Illuminate the Origin of Land Plants and the Unique Biology of Hornworts." Nature Plants 6 (3): 259–72. https://doi.org/10.1038/s41477-020-0618-2.

Liang, Zhe, Yuke Geng, Changmian Ji, Hai Du, Chui Eng Wong, Qian Zhang, Ye Zhang, et al. 2019. “Mesostigma Viride Genome and Transcriptome Provide Insights into the Origin and Evolution of Streptophyta.” Advanced Science 7 (1): 1901850. https://doi.org/10.1002/advs.201901850.

Listenberger, L. L., and Brown, D. A. (2007) Fluorescent Detection of Lipid Droplets and Associated Proteins. Curr Protoc Cell Biol. 35, 24.2.1-24.2.11

Liu, L.-J. et al. COP1-Mediated Ubiquitination of CONSTANS Is Implicated in Cryptochrome Regulation of Flowering in Arabidopsis. The Plant Cell 20, 292–306 (2008).

Liu, Ruijie, Aliaksei Z. Holik, Shian Su, Natasha Jansz, Kelan Chen, Huei San Leong, Marnie E. Blewitt, Marie-Liesse Asselin-Labat, Gordon K. Smyth, and Matthew E. Ritchie. 2015. “Why Weight? Modelling Sample and Observational Level Variability Improves Power in RNA-Seq Analyses.” Nucleic Acids Research 43 (15): e97. https://doi.org/10.1093/nar/gkv412.

Mann, H.B. and Whitney, D.R. (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statist. 1, 50–60.

Manni, Mosè, Matthew R Berkeley, Mathieu Seppey, Felipe A Simão, and Evgeny M Zdobnov. 2021. “BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.” Molecular Biology and Evolution 38 (10): 4647–54. https://doi.org/10.1093/molbev/msab199.

Mapleson, Daniel, Luca Venturini, Gemy Kaithakottil, and David Swarbreck. 2018. “Efficient and Accurate Detection of Splice Junctions from RNA-Seq with Portcullis.” GigaScience 7 (12): giy131. https://doi.org/10.1093/gigascience/giy131.

Merchant, Sabeeha S., Simon E. Prochnik, Olivier Vallon, Elizabeth H. Harris, Steven J. Karpowicz, George B. Witman, Astrid Terry, et al. 2007. “The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions.” Science 318 (5848): 245–50. https://doi.org/10.1126/science.1143609.

Minos. “Minos - a Gene Model Consolidation Pipeline for Genome Annotation Projects.” (2019) 2022. Python. EI-CoreBioinformatics. https://github.com/EI-CoreBioinformatics/minos.

Montgomery, Sean A., Yasuhiro Tanizawa, Bence Galik, Nan Wang, Tasuku Ito, Takako Mochizuki, Svetlana Akimcheva, et al. 2020. “Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin.” Current Biology 30 (4): 573-588.e7. https://doi.org/10.1016/j.cub.2019.12.015.

Mueller, S. P., Krause, D. M., Mueller, M. J. & Fekete, A. Accumulation of extra-chloroplastic triacylglycerols in *Arabidopsis* seedlings during heat acclimation. *Journal of Experimental Botany* **66**, 4517–4526 (2015).

Nishiyama, T., H. Sakayama, J. de Vries, H. Buschmann, D. Saint-Marcoux, K.K. Ullrich, F.B. Haas, et al. 2018. “The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization.” Cell 174 (2): 448-464.e24. https://doi.org/10.1016/j.cell.2018.06.033.

Ordoñez-Herrera, N. et al. The Transcription Factor COL12 Is a Substrate of the COP1/SPA E3 Ligase and Regulates Flowering Time and Plant Architecture. Plant Physiol. 176, 1327–1340 (2018).

Pertea, Mihaela, Daehwan Kim, Geo M. Pertea, Jeffrey T. Leek, and Steven L. Salzberg. 2016. “Transcript-Level Expression Analysis of RNA-Seq Experiments with HISAT, StringTie and Ballgown.” Nature Protocols 11 (9): 1650–67. https://doi.org/10.1038/nprot.2016.095.

Phipson, Belinda, Stanley Lee, Ian J. Majewski, Warren S. Alexander, and Gordon K. Smyth. 2016. “Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression.” The Annals of Applied Statistics 10 (2): 946–63. https://doi.org/10.1214/16-AOAS920.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem 75, 663-670.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2, 1896-1906.

REAT. “REAT - Robust and Extendable Eukaryotic Annotation Toolkit.” n.d. Accessed July 19, 2022. https://github.com/EI-CoreBioinformatics/reat.

Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. “Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies.” Nucleic Acids Research 43 (7): e47–e47. https://doi.org/10.1093/nar/gkv007.

Robinson, Mark D., and Alicia Oshlack. 2010. “A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data.” Genome Biology 11 (3): R25. https://doi.org/10.1186/gb-2010-11-3-r25.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2010. “EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data.” Bioinformatics 26 (1): 139–40. https://doi.org/10.1093/bioinformatics/btp616.

Rossini, L., Cribb, L., Martin, D. J. & Langdale, J. A. The Maize Golden2 Gene Defines a Novel Class of Transcriptional Regulators in Plants. The Plant Cell 13, 1231-1244 (2001).

Sarid□Krebs, L. et al. Phosphorylation of CONSTANS and its COP 1□dependent degradation during photoperiodic flowering of Arabidopsis. Plant J 84, 451–463 (2015).

Schindelin, J., Arganda-Carreras, I., Frise, E. *et al.* Fiji: an open-source platform for biological-image analysis. Nat. Methods, **9**, 676–682 (2012)

Shao, Mingfu, and Carl Kingsford. 2017. “Accurate Assembly of Transcripts through Phase-Preserving Graph Decomposition.” Nature Biotechnology 35 (12): 1167–69. https://doi.org/10.1038/nbt.4020.

Shapiro, S.S. and Wilk, M.B. (1965) An analysis of variance test for normality (complete samples). Biometrika, 52, 591–611.

Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal. Chem. 68, 850–858 (1996).

Smyth, Gordon K., Joëlle Michaud, and Hamish S. Scott. 2005. “Use of Within-Array Replicate Spots for Assessing Differential Expression in Microarray Experiments.” Bioinformatics (Oxford, England) 21 (9): 2067–75. https://doi.org/10.1093/bioinformatics/bti270.

Soneson, Charlotte, Michael I. Love, and Mark D. Robinson. 2016. “Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences.” F1000Research. https://doi.org/10.12688/f1000research.7563.2.

Stanke, Mario, Ana Tzvetkova, and Burkhard Morgenstern. 2006. “AUGUSTUS at EGASP: Using EST, Protein and Genomic Alignments for Improved Gene Prediction in the Human Genome.” Genome Biology 7 (1): S11. https://doi.org/10.1186/gb-2006-7-s1-s11.

Stanke, Mario, Oliver Keller, Irfan Gunduz, Alec Hayes, Stephan Waack, and Burkhard Morgenstern. 2006. “AUGUSTUS: Ab Initio Prediction of Alternative Transcripts.” Nucleic Acids Research 34 (suppl_2): W435–39. https://doi.org/10.1093/nar/gkl200.

Testa, Alison C., James K. Hane, Simon R. Ellwood, and Richard P. Oliver. 2015. “CodingQuarry: Highly Accurate Hidden Markov Model Gene Prediction in Fungal Genomes Using RNA-Seq Transcripts.” BMC Genomics 16 (1): 170. https://doi.org/10.1186/s12864-015-1344-4.

Timm, S. et al. A Cytosolic Pathway for the Conversion of Hydroxypyruvate to Glycerate during Photorespiration in Arabidopsis. The Plant Cell 20, 2848–2859 (2008).

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 13, 731-740.

Venturini, Luca, Shabhonam Caim, Gemy George Kaithakottil, Daniel Lee Mapleson, and David Swarbreck. 2018. "Leveraging Multiple Transcriptome Assembly Methods for Improved Gene Structure Annotation." GigaScience 7 (8): giy093. https://doi.org/10.1093/gigascience/giy093.

W.J. Conover, Practical Nonparametrics Statistics. (1999)

Wang, Sibo, Linzhou Li, Haoyuan Li, Sunil Kumar Sahu, Hongli Wang, Yan Xu, Wenfei Xian, et al. 2020. "Genomes of Early-Diverging Streptophyte Algae Shed Light on Plant Terrestrialization." Nature Plants 6 (2): 95–106. https://doi.org/10.1038/s41477-019-0560-3.

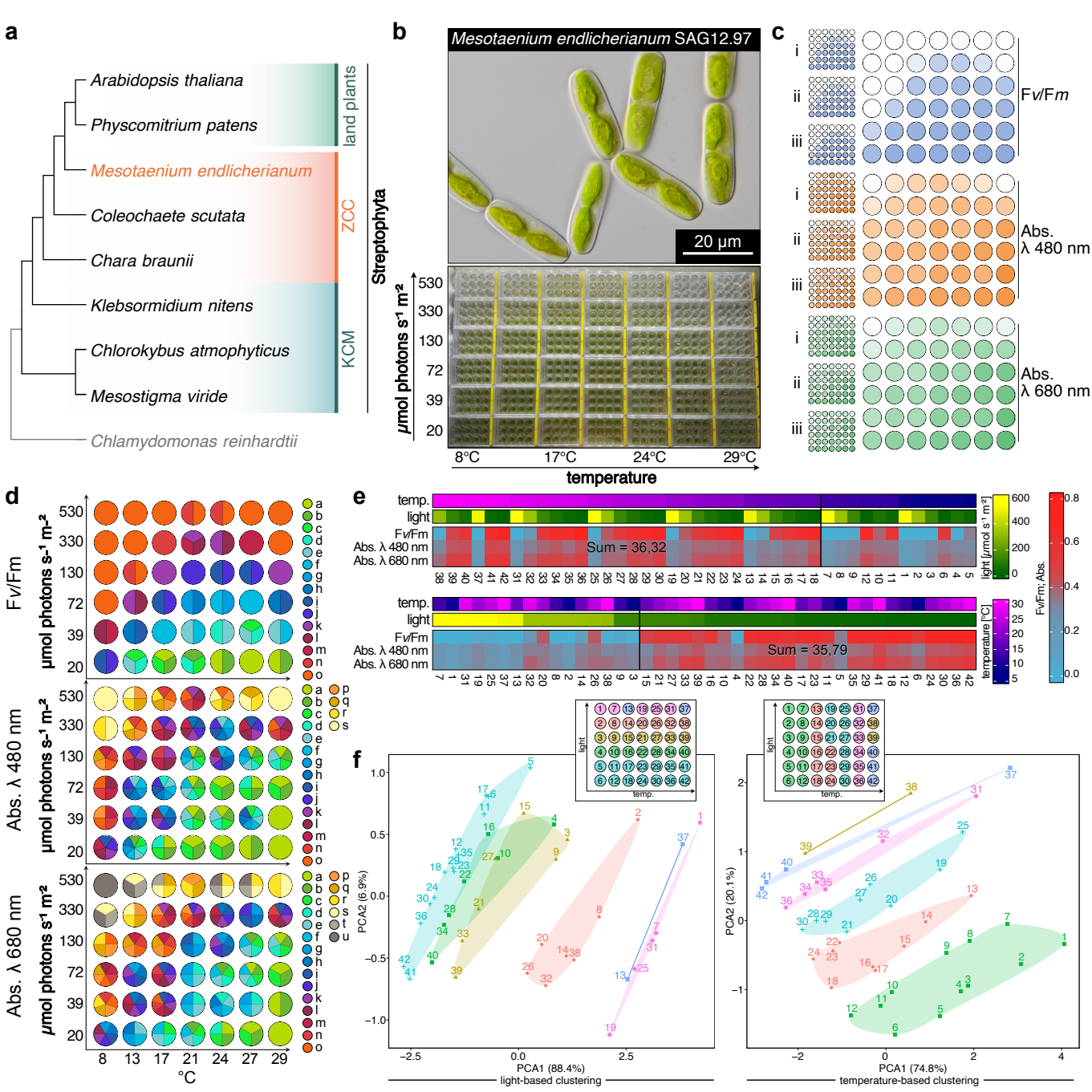
Waters, M. T. et al. GLK Transcription Factors Coordinate Expression of the Photosynthetic Apparatus in Arabidopsis. The Plant Cell 21, 1109–1128 (2009).

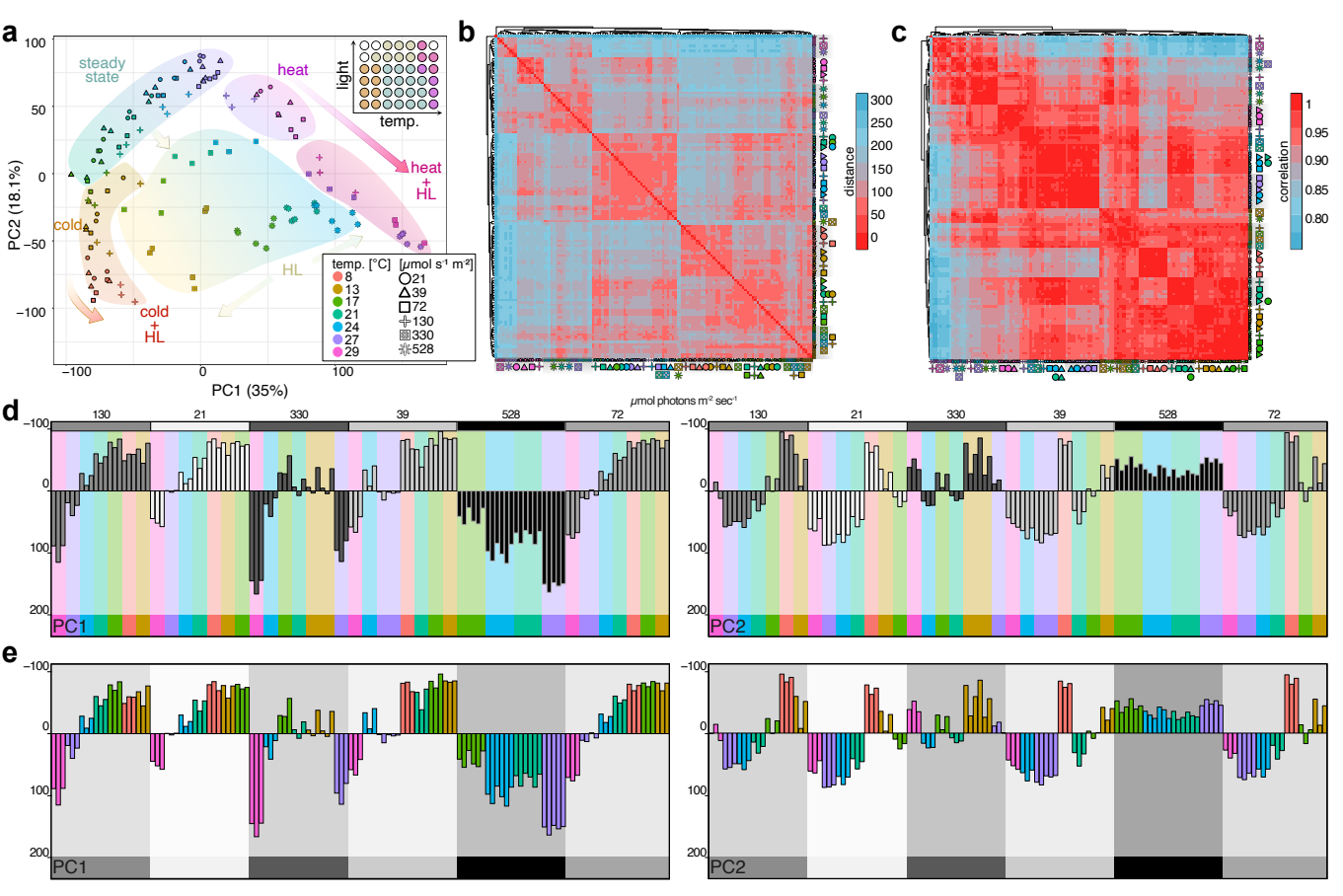
Wijesooriya, Kaumadi, Sameer A. Jadaan, Kaushalya L. Perera, Tanuveer Kaur, and Mark Ziemann. 2022. "Urgent Need for Consistent Standards in Functional Enrichment Analysis." PLOS Computational Biology 18 (3): e1009935. https://doi.org/10.1371/journal.pcbi.1009935.

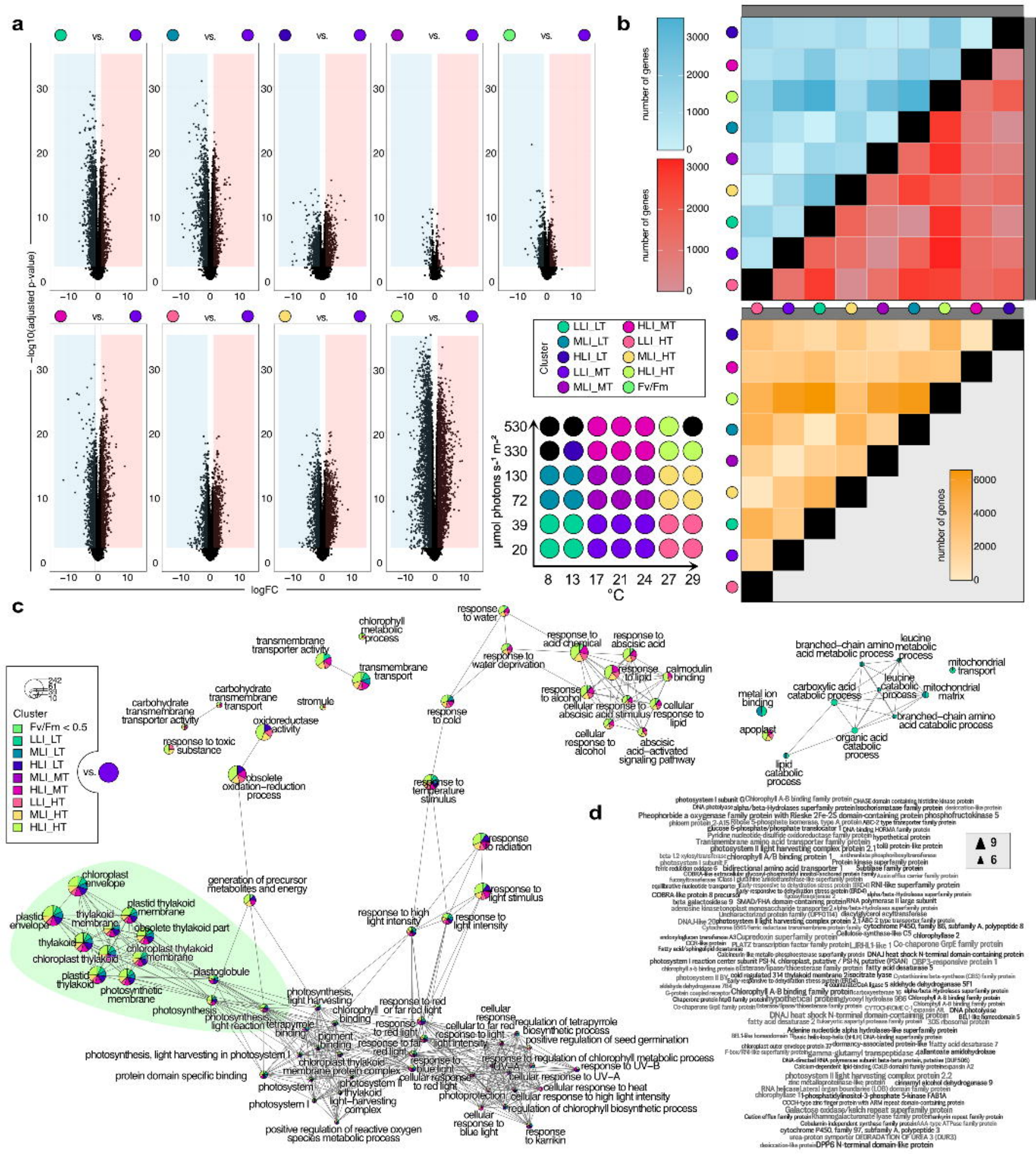
Wu, Tianzhi, Erqiang Hu, Shuangbin Xu, Meijun Chen, Pingfan Guo, Zehan Dai, Tingze Feng, et al. 2021. "ClusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data." The Innovation, July, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

Yasumura, Y., Moylan, E. C. & Langdale, J. A. A Conserved Transcription Factor Mediates Nuclear Control of Organelle Biogenesis in Anciently Diverged Land Plants. The Plant Cell 17, 1894–1907 (2005).

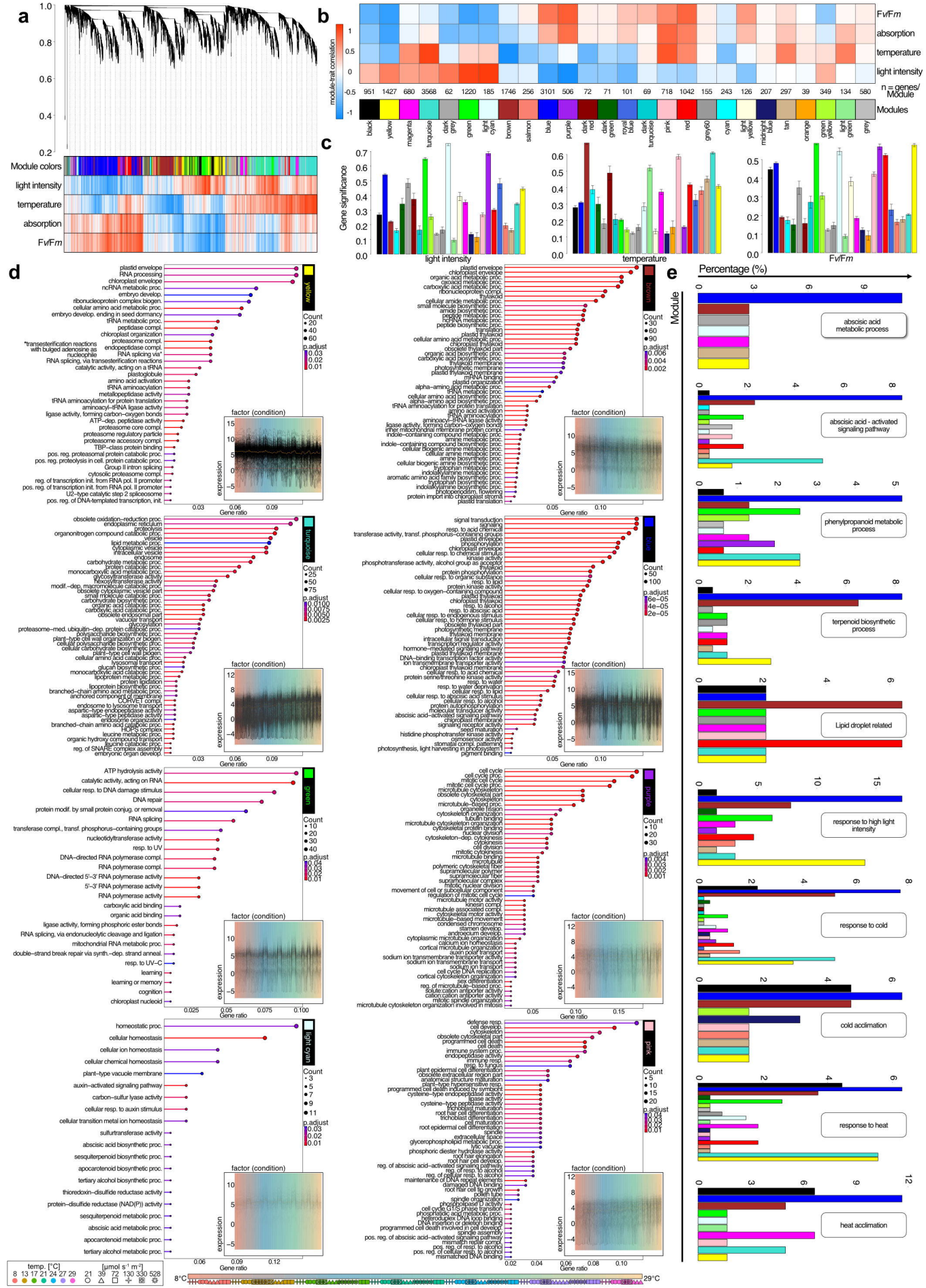
Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "ClusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters." OMICS: A Journal of Integrative Biology 16 (5): 284–87. https://doi.org/10.1089/omi.2011.0118.

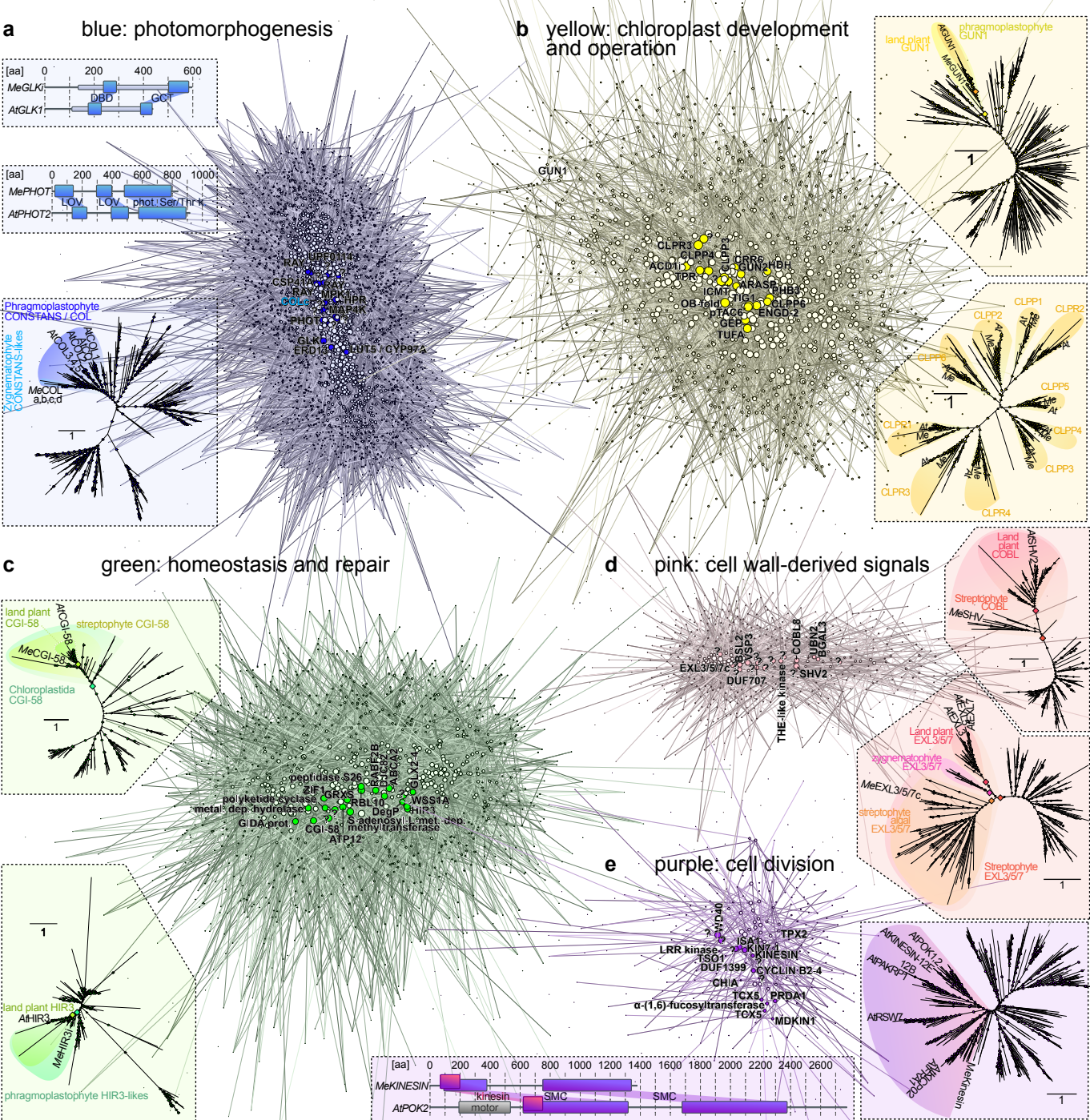

a
Arabidopsis thaliana
Physcomitrium patens
Mesotaenium endlicherianum
Coleochaete scutata
Chara braunii
Klebsormidium nitens
Chlorokybus atmophyticus
Mesostigma viride
Chlamydomonas reinhardtii
land plants
ZCC
KCM
Streptophyta
b
Mesotaenium endlicherianum SAG12.97
20 µm
µmol photons s-1 m-2
530
330
130
72
39
20
8°C
17°C
24°C
29°C
temperature
c
i
ii
iii
Fv/Fm
Abs. λ 480 nm
Abs. λ 680 nm
d
Fv/Fm
Abs. λ 480 nm
Abs. λ 680 nm
µmol photons s-1 m-2
8
13
17
21
24
27
29
°C
e
temp.
light
Fv/Fm
Abs. λ 480 nm
Abs. λ 680 nm
Sum = 36,32
Sum = 35,79
light [µmol s-1 m-2]
temperature [°C]
Fv/Fm; Abs.
f
PCA2 (6.9%)
PCA1 (88.4%)
light-based clustering
PCA2 (20.1%)
PCA1 (74.8%)
temperature-based clustering


a
steady state
heat
light
temp.
heat + HL
cold
HL
cold + HL
PC2 (18.1%)
PC1 (35%)
temp. [°C]
8
13
17
21
24
27
29
[μmol s⁻¹ m⁻²]
21
39
72
130
330
528
b
distance
300
250
200
150
100
50
0
c
correlation
1
0.95
0.90
0.85
0.80
d
μmol photons m⁻² sec⁻¹
130
21
330
39
528
72
PC1
PC2
e
PC1
PC2

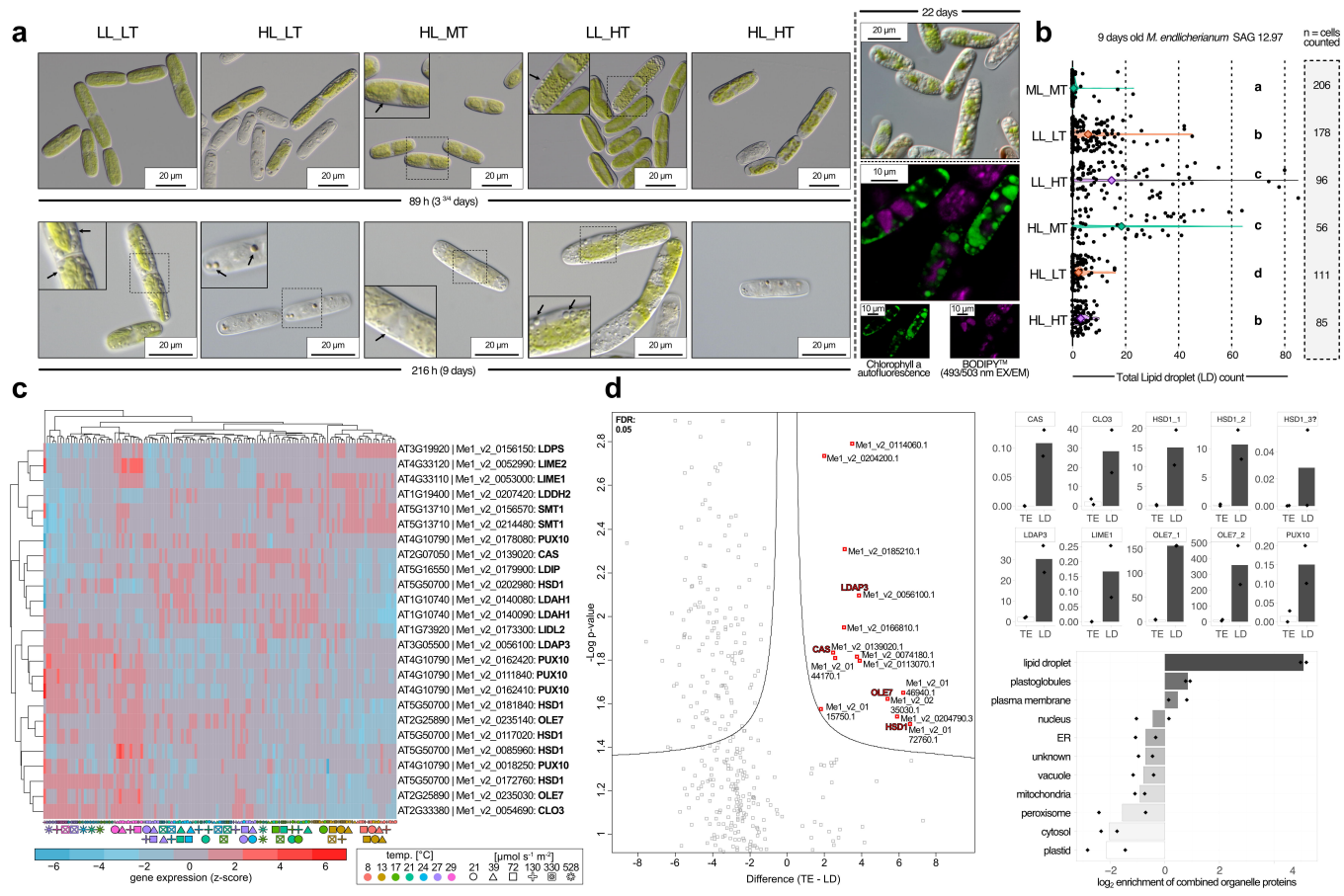


a
b
c
d
vs.
−log10(adjusted p-value)
logFC
Cluster
LLI_LT
MLI_LT
HLI_LT
LLI_MT
MLI_MT
HLI_MT
LLI_HT
MLI_HT
HLI_HT
Fv/Fm
μmol photons s-1 m-2
°C
number of genes
Fv/Fm < 0.5
response to water
response to water deprivation
response to acid chemical
response to abscisic acid
response to lipid
calmodulin binding
response to alcohol
cellular response to abscisic acid stimulus
cellular response to lipid
cellular response to alcohol
abscisic acid–activated signaling pathway
response to cold
response to temperature stimulus
chlorophyll metabolic process
transmembrane transporter activity
transmembrane transport
carbohydrate transmembrane transport
carbohydrate transmembrane transporter activity
stromule
oxidoreductase activity
response to toxic substance
obsolete oxidation–reduction process
branched–chain amino acid metabolic process
leucine metabolic process
mitochondrial transport
leucine catabolic process
mitochondrial matrix
carboxylic acid catabolic process
metal ion binding
apoplast
branched–chain amino acid catabolic process
organic acid catabolic process
lipid catabolic process
response to radiation
response to light stimulus
response to high light intensity
response to light intensity
generation of precursor metabolites and energy
chloroplast envelope
plastid envelope
thylakoid membrane
plastid thylakoid membrane
obsolete thylakoid part
thylakoid
chloroplast thylakoid
chloroplast thylakoid membrane
plastid thylakoid
plastoglobule
photosynthetic membrane
photosynthesis
photosynthesis, light harvesting
photosynthesis, light reaction
tetrapyrrole binding
chlorophyll binding
pigment binding
response to red or far red light
response to red light
response to far red light
cellular response to far red
cellular response to light intensity
cellular response to light intensity
regulation of tetrapyrrole biosynthetic process
positive regulation of seed germination
photosynthesis, light harvesting in photosystem I
chloroplast thylakoid membrane protein complex
response to blue light
response to regulation of chlorophyll metabolic process
response to UV–B
protein domain specific binding
photosystem
photosystem II
thylakoid light–harvesting complex
photosystem I
cellular response to red light
cellular response to UV–A
cellular response to heat
photoprotection
cellular response to high light intensity
regulation of chlorophyll biosynthetic process
cellular response to blue light
response to karrikin
positive regulation of reactive oxygen species metabolic process


a
Module colors
light intensity
temperature
absorption
Fv/Fm
b
module-trait correlation
Fv/Fm
absorption
temperature
light intensity
n = genes/ Module
Modules
c
Gene significance
light intensity
temperature
Fv/Fm
d
yellow
turquoise
green
light cyan
blue
brown
purple
pink
factor (condition)
expression
Gene ratio
Count
p.adjust
e
Percentage (%)
Module
abscisic acid metabolic process
abscisic acid - activated signaling pathway
phenylpropanoid metabolic process
terpenoid biosynthetic process
Lipid droplet related
response to high light intensity
response to cold
cold acclimation
response to heat
heat acclimation
temp. [°C]
[µmol s-1 m-2]
8°C
29°C



a
blue: photomorphogenesis
[aa] 0 200 400 600
MeGLKi
DBD
GCT
AtGLK1
[aa] 0 200 400 600 800 1000
MePHOT
LOV
LOV
phot. Ser/Thr k
AtPHOT2
Phragmoplastophyte CONSTANS / COL
Zygnematophyte CONSTANS-likes
MeCOL a,b,c,d
b
yellow: chloroplast development and operation
GUN1
CLPR3
CLPP4
CLPP3
CRR6
ACD1
GUN2
HDH
TPR1
ARASP
ICMT
PHB3
TIG1
OB-fold
CLPP6
pTAC6
ENGD-2
GEP
TUFA
land plant GUN1
phragmoplastophyte GUN1
CLPP1
CLPP2
CLPR2
CLPP5
CLPR1
CLPP4
CLPP3
CLPR3
CLPR4
c
green: homeostasis and repair
land plant CGI-58
streptophyte CGI-58
MeCGI-58
Chloroplastida CGI-58
peptidase S26
ZIF1
GRXS
polyketide cyclase
metal. dep. hydrolase
GIDA prot
CGI-58
ATP12
RABF2B
DJC82
ABCA2
GLX2-4
RBL10
WSS1A
DegP
HIR3
S-adenosyl-L-met.-dep. methyltransferase
land plant HIR3
AtHIR3
MeHIR3
phragmoplastophyte HIR3-likes
d
pink: cell wall-derived signals
EXL3/5/7c
BSL2
VSP3
DUF707
THE-like kinase
COBL8
UBN2
BGAL3
SHV2
Land plant COBL
Streptophyte COBL
MeSHV
Land plant EXL3/5/7
zygnematophyte EXL3/5/7
MeEXL3/5/7c
streptophyte algal EXL3/5/7
Streptophyte EXL3/5/7
e
purple: cell division
WD40
TPX2
LRR kinase
KIN7-1
KINESIN
TSO1
DUF1399
CYCLIN-B2-4
CHIA
TCX5
PRDA1
α-(1,6)-fucosyltransferase
TCX5
MDKIN1
AtPOK1,2
AtKINESIN-12E
AtPAKRP1
AtRSW7
AtPOK2
MeKinesin
[aa] 0 200 400 600 800 1000 1200 1400 1600 1800 2000 2200 2400 2600
MeKINESIN
kinesin
SMC
SMC
motor
AtPOK2


a
LL_LT
HL_LT
HL_MT
LL_HT
HL_HT
20 µm
89 h (3 3/4 days)
216 h (9 days)
22 days
10 µm
Chlorophyll a autofluorescence
BODIPY™ (493/503 nm EX/EM)
b
9 days old M. endlicherianum SAG 12.97
n = cells counted
ML_MT
LL_LT
LL_HT
HL_MT
HL_LT
HL_HT
a
b
c
c
d
b
206
178
96
56
111
85
0
20
40
60
80
Total Lipid droplet (LD) count
c
AT3G19920 | Me1_v2_0156150: LDPS
AT4G33120 | Me1_v2_0052990: LIME2
AT4G33110 | Me1_v2_0053000: LIME1
AT1G19400 | Me1_v2_0207420: LDDH2
AT5G13710 | Me1_v2_0156570: SMT1
AT5G13710 | Me1_v2_0214480: SMT1
AT4G10790 | Me1_v2_0178080: PUX10
AT2G07050 | Me1_v2_0139020: CAS
AT5G16550 | Me1_v2_0179900: LDIP
AT5G50700 | Me1_v2_0202980: HSD1
AT1G10740 | Me1_v2_0140080: LDAH1
AT1G10740 | Me1_v2_0140090: LDAH1
AT1G73920 | Me1_v2_0173300: LIDL2
AT3G05500 | Me1_v2_0056100: LDAP3
AT4G10790 | Me1_v2_0162420: PUX10
AT4G10790 | Me1_v2_0111840: PUX10
AT4G10790 | Me1_v2_0162410: PUX10
AT5G50700 | Me1_v2_0181840: HSD1
AT2G25890 | Me1_v2_0235140: OLE7
AT5G50700 | Me1_v2_0117020: HSD1
AT5G50700 | Me1_v2_0085960: HSD1
AT4G10790 | Me1_v2_0018250: PUX10
AT5G50700 | Me1_v2_0172760: HSD1
AT2G25890 | Me1_v2_0235030: OLE7
AT2G33380 | Me1_v2_0054690: CLO3
−6
−4
−2
0
2
4
6
gene expression (z-score)
temp. [°C]
8 13 17 21 24 27 29
[µmol s⁻¹ m⁻²]
21 39 72 130 330 528
d
FDR: 0.05
-Log p-value
Difference (TE - LD)
Me1_v2_0114060.1
Me1_v2_0204200.1
Me1_v2_0185210.1
LDAP3
Me1_v2_0056100.1
Me1_v2_0166810.1
CAS
Me1_v2_0139020.1
Me1_v2_0074180.1
Me1_v2_0113070.1
Me1_v2_0144170.1
OLE7
Me1_v2_0146940.1
Me1_v2_0135030.1
Me1_v2_0204790.3
HSD1
Me1_v2_0172760.1
Me1_v2_0115750.1
CAS
CLO3
HSD1_1
HSD1_2
HSD1_3?
LDAP3
LIME1
OLE7_1
OLE7_2
PUX10
TE
LD
lipid droplet
plastoglobules
plasma membrane
nucleus
ER
unknown
vacuole
mitochondria
peroxisome
cytosol
plastid
log2 enrichment of combined organelle proteins