Supplementary materials

Article title: Molecular mechanisms underlying tree host-pathogen interactions under drought stress and subsequent rewatering in *Eucalyptus grandis*

Authors: Demissew Tesfaye Teshome, Godfrey Elijah Zharare, Raphael Ployet and Sanushka Naidoo

The following Supplementary materials are available for this article:

Fig. S1 Soil moisture content for the glasshouse experiment on combined drought stress/rewatering and pathogen infection in *Eucalyptus grandis*.

Fig. S2 Principal component analysis (PCA) of RNA-seq data for the sample sets used in the study before and after batch correction.

Fig. S3 Co-expression network analyses and identification of key modules

Fig. S4. Representative photos showing stem lesion development in *Eucalyptus grandis* inoculated with *Chrysoporthe austroafricana* under well-watered, water stressed, and rewatered conditions at different measurement timepoints.

Fig. S5 Heatmap of expression for selected co-expression modules

Table S1 Gene ontology biological processes enrichment analysis for unpreserved co-expression modules of the network on 'unstressed' samples

Table S2 Significantly enriched transcription factors in co-expression modules selected using module preservation and module-trait correlation analyses

Table S3 Gene ontology biological processes enrichment analysis for selected co-expression modules of the network on the 'stressed' samples

Table S4 Potential hub genes identified from co-expression modules selected using module preservation and module-trait correlation analyses

Table S5 Gene ontology biological process enrichment for DEGs in the different Venn-intersection categories of pathogen infection at 17 days post inoculation, recovery from drought stress, and recovery from combined stress in Eucalyptus grandis

Table S6 Selected differentially expressed genes of Eucalyptus grandis during recovery from combined drought stress and pathogen infection

Table S7 Differentially expressed enriched transcription factors of Eucalyptus grandis during recovery from combined drought stress and pathogen infection

Table S8 Differentially expressed gene of Chrysoporthe austroafricana during drought stress and rewatering in Eucalyptus grandis

Methods S1 RNA-seq data and bioinformatic analysis pipeline used in the study

Methods S2 Co-expression network construction, module detection, and identification of key modules and hub genes

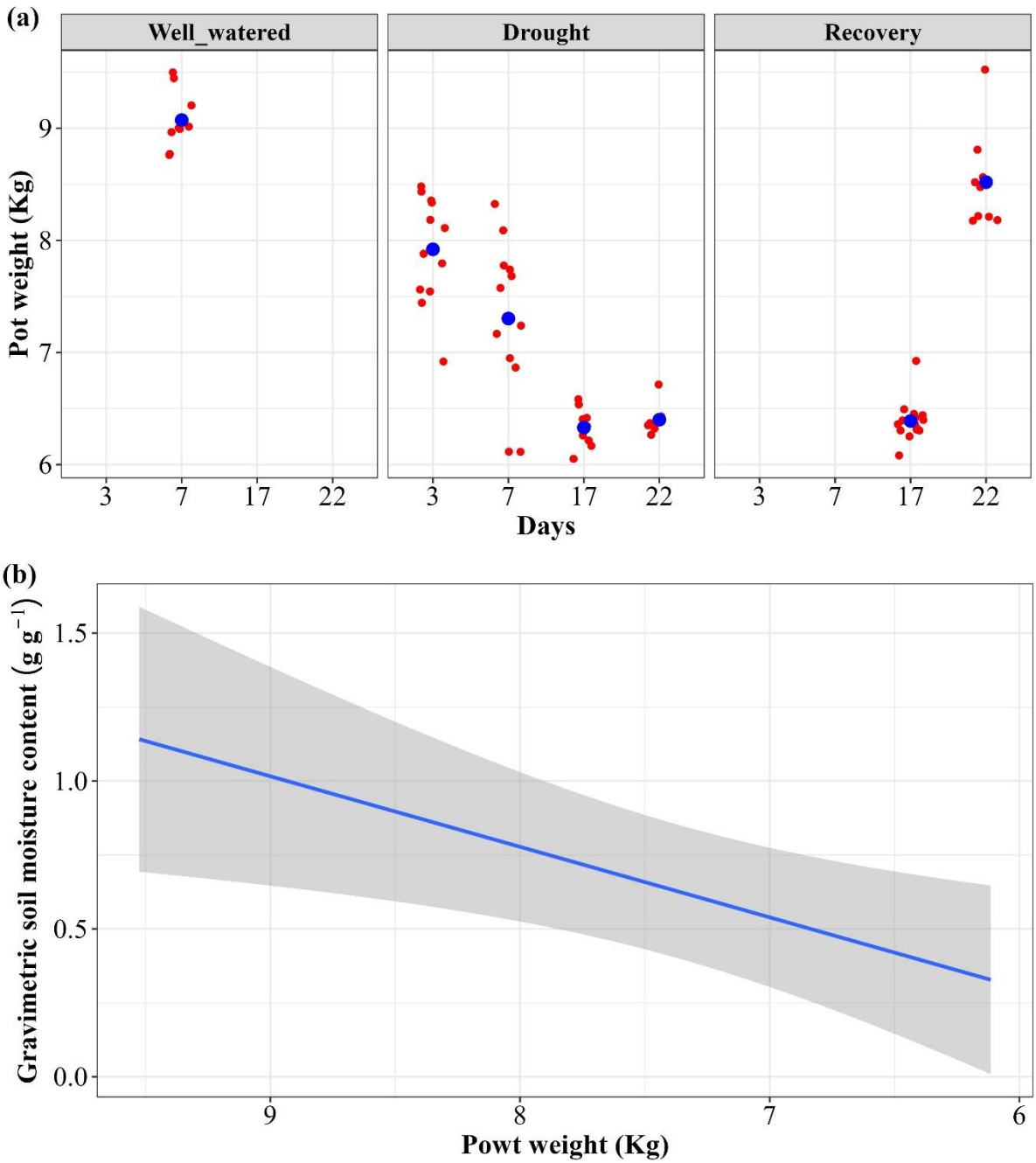

Fig. S1 Soil moisture content for the glasshouse experiment on combined drought stress/rewatering and pathogen infection in *Eucalyptus grandis*. a. Pot weight at different timepoints. Changes in moisture content were monitored by randomly selecting pots from well-watered, water-stressed and re-watered treatments at different timepoints. Days 3, 7, 17, and 22 represent measurement and stem sampling timepoints (Fig. 1). The small red and large blue dots respectively indicate the mean and individual pot weight measurements. **b.** Relationship between gravimetric soil moisture content and pot weight. One pot was randomly selected in each measurement time point, and gravimetric soil moisture content was determined. The line represents a mean of soil moisture content determined for soil from the top, middle and bottom of the pot.

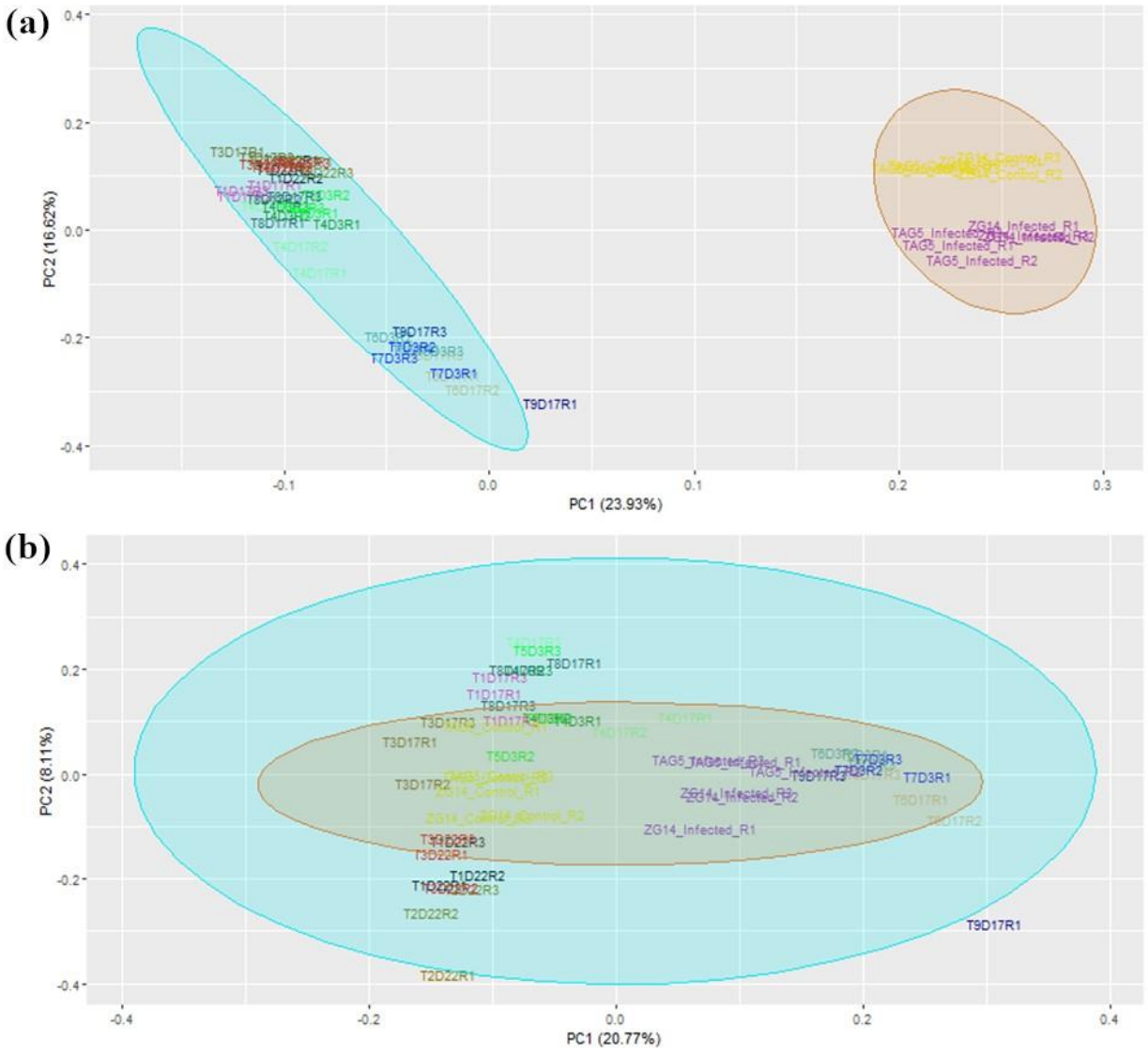

Fig. S2 Principal component analysis (PCA) of RNA-seq data for the sample sets used in the study before and after batch correction. a. PCA before batch correction **b.** PCA after batch correction. The ellipses indicate samples in the different batches.

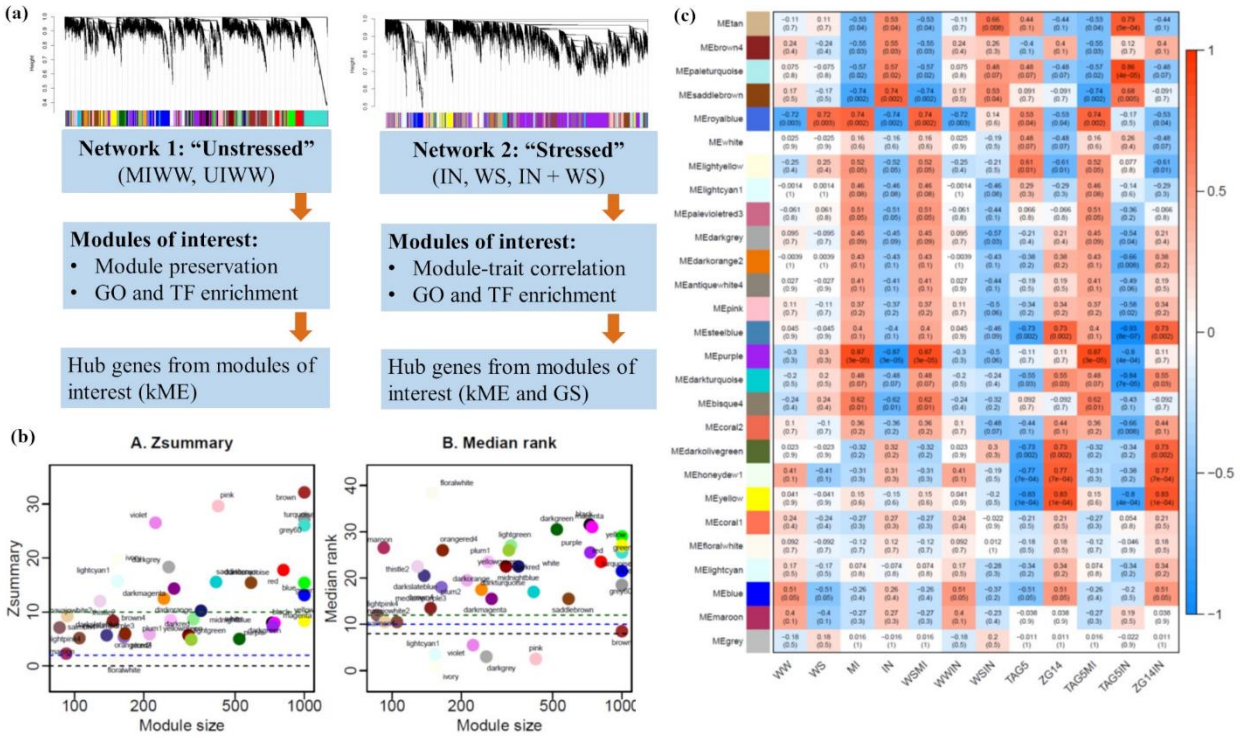

Fig. S3 Co-expression network analyses and identification of key modules. a. Pipeline of co-expression network analyses **b.** Module preservation analysis **c.** Module-trait correlation analysis. MIWW = Mock-inoculated/well-watered, UIWW = uninoculated/well-watered, INWW = inoculated/well-watered, MIWS = Mock-inoculated/water-stressed, INWS = inoculated/water-stressed, kME = module membership, GS = gene significance


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)

Fig. S4. Representative photos showing stem lesion development in *Eucalyptus grandis* inoculated with *Chrysoporthe austroafricana* under well-watered, water stressed, and rewatered conditions at different measurement timepoints. a. Well-watered at 3 days post inoculation (dpi) **b.** Water stressed at 3 dpi **c.** Well-watered at 7 dpi **d.** Water stressed at 7 dpi **e.** Well-watered at 18 dpi **f.** Rewatered at 18 dpi **g.** Well-watered at 22 dpi **h.** Water stressed at 22 dpi **i.** Rewatered at 22 dpi. The experimental design and measurement timepoints are depicted in Figure 1. The timepoints 3, 7, 18, and 22 dpi respectively correspond to 10-, 14-, 25-, and 29-days after withholding water. For the rewatered treatments, the timepoints 18 and 22 dpi respectively correspond to 1 and 5 days following rewatering.

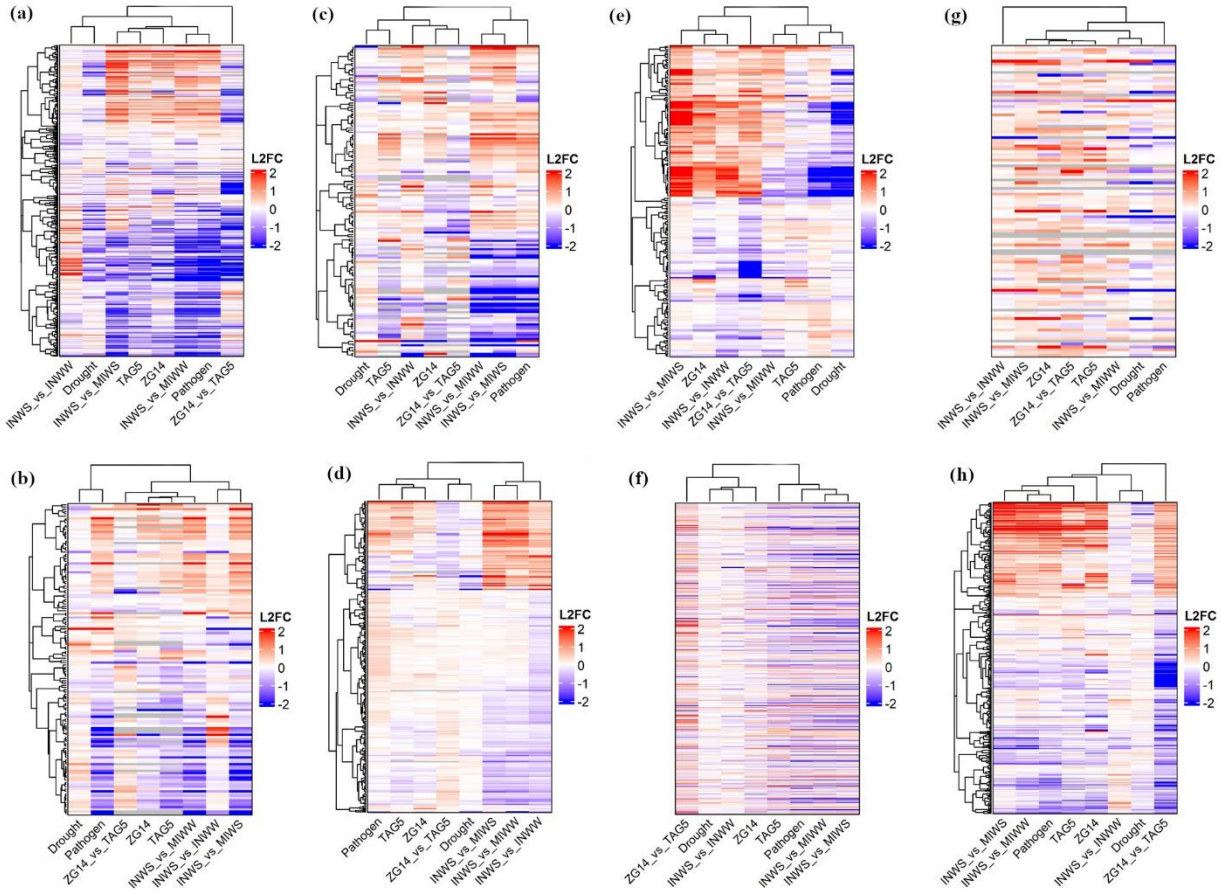

Fig. S5 Heatmap of expression for selected co-expression modules. a. Darkgreen **b.** Darkslateblue **c.** Floralwhite **d.** Darkgrey **e.** Darkolivegreen **f.** Darkgrey **g.** Honeydew1 **h.** Lightyellow **i.** Royalblue **j.** Saddllebrown **k.** Steelblue **l.** Tan **m.** Yellow The heatmaps represent the expression pattern of genes in the key modules selected using module-trait correlation analysis (highly correlated modules, a-j) and those selected using module preservation analysis (unpreserved modules, k-m). The $\log_2$ foldchange (L2FC) between different comparisons was used to plot the heatmap using the R package ComplexHeatmap. MIWW = Mock-inoculated-well-watered, MIWS = Mock-inoculated-water-stressed, INWW = Inoculated-well-watered, INWS

= Inoculated-water-stressed, TAG5 and ZG14 refer to comparisons between the respective inoculated and mock-inoculated treatments using the data from Mangwanda et al., (2015), drought and pathogen respectively refer to comparisons of MIWS and INWW with MIWW based on the data from the present experiment.

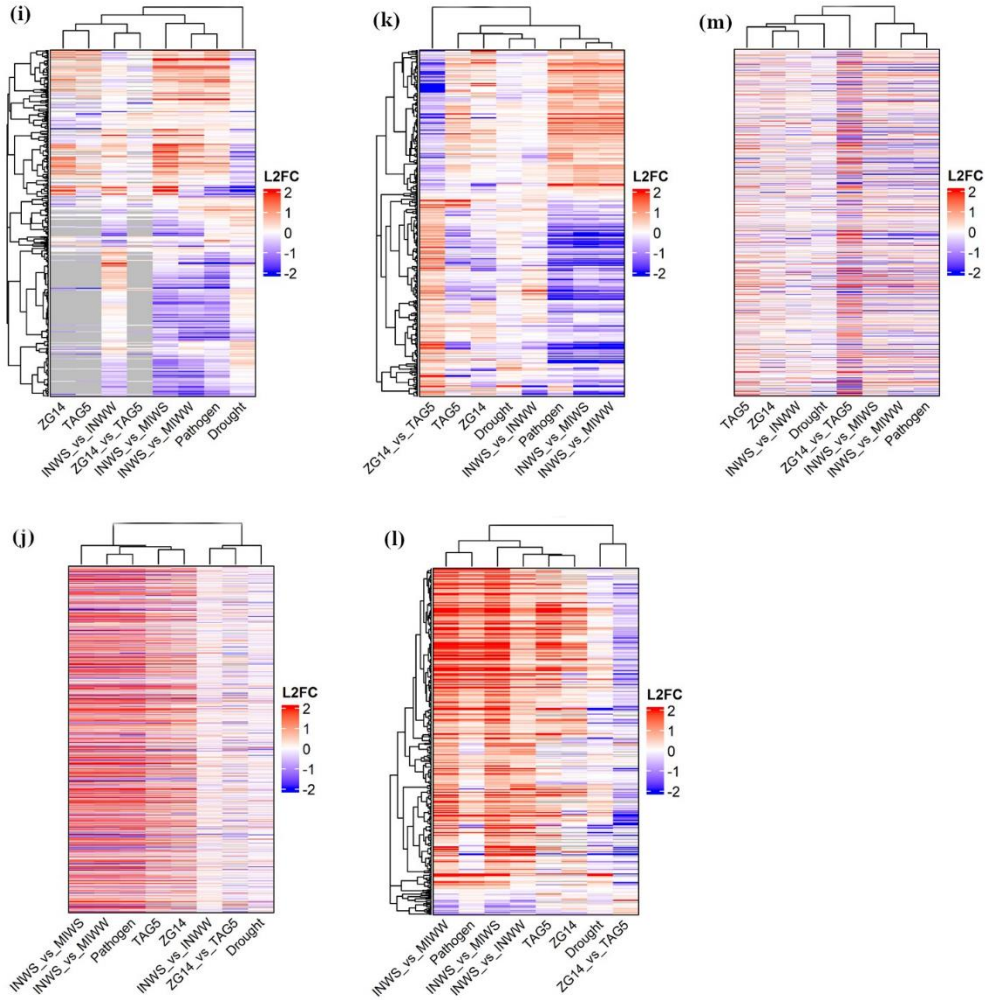

Fig. S5 Heatmap of expression for selected co-expression modules. a. Darkgreen **b.** Darkslateblue **c.** Floralwhite **d.** Darkgrey **e.** Darkolivegreen **f.** Darkgrey **g.** Honeydew1 **h.** Lightyellow **i.** Royalblue **j.** Saddllebrown **k.** Steelblue **l.** Tan **m.** Yellow**.** The heatmaps represent the expression pattern of genes in the key modules selected using module-trait correlation analysis (highly correlated modules, a-j) and those selected using module preservation analysis (unpreserved modules, k-m). The $\log_2$ foldchange (L2FC) between different comparisons was used to plot the heatmap using the R package ComplexHeatmap. MIWW = Mock-inoculated-well-watered, MIWS = Mock-inoculated-water-stressed, INWW = Inoculated-well-watered, INWS = Inoculated-water-stressed, TAG5 and ZG14 refer to comparisons between the respective inoculated and mock-inoculated treatments using the data from Mangwanda et al., (2015), drought and pathogen respectively refer to comparisons of MIWS and INWW with MIWW based on the data from the present experiment.

Methods S1 RNA-seq data and bioinformatic analysis pipeline used in the study

We used RNA-seq data from the combined pathogen infection and drought stress/recovery treatments of the present study as well as previous single stress studies. Paired end (PE) RNA-seq data from 100 and 150 base pairs (bp) libraries were obtained for three biological replicates each of the well-watered/mock-inoculated (WWMI), well-watered/inoculated (WWIN), water-stressed/mock-inoculated (WSMI), and water-stressed/inoculated (WSIN) combined stress treatments on day 3 as well as the WWMI, WWIN, rewatered/mock-inoculated (RWMI), and rewatered/inoculated (RWIN) recovery treatments on day 17 (Fig. 1). The RNA-seq data from previous studies included 100 and 150 bp PE RNA-seq data for unwounded-well-watered TAG5 stem samples deposited on the repository of the national centre for biotechnology information (NCBI, accession number PRJNA896601) (Teshome *et al.*, 2023), and 50 bp PE RNA-seq data for samples inoculated with *Chrysoporthe austroafricana* with the respective mock-inoculated controls at 3 days post inoculation for both TAG5 and ZG14 with three biological replicates each deposited on the NCBI repository (accession number PRJNA280236) (Mangwanda *et al.*, 2015). Additionally, 50 bp PE RNA-seq data from a previous *in vitro* experiment involving culturing of *C. austroafricana* on minimal and complete media with three biological replicates each were used for the pathogen (Mangwanda *et al.*, 2016).

The RNA-seq data were pre-processed using the same tools and pipeline described in Teshome *et al.* (2023). We used FASTQC version 0.11.7 for quality assessment both before and after quality trimming using Trimmomatic version 0.36 (Bolger *et al.*, 2014). We used Spliced Transcripts Alignment to a Reference (STAR) (Dobin *et al.*, 2013) to align the RNA-seq data to the concatenated references of the *E. grandis* genome v2.0 (Myburg *et al.*, 2014) and *C. austroafricana* genome (Wingfield *et al.*, 2015) followed by transcript quantification using StringTie version 1.3.4d (Pertea *et al.*, 2015). Subsequently, we used tximport to import the transcript quantification files into the R package DESeq2 version 1.32.0 (Love *et al.*, 2014) and summarize them into a transcripts per million (TPM) matrix for the combined drought stress-pathogen infection treatments and differential expression analyses for the recovery treatments and pathogen transcripts.

To identify genes that were differentially expressed (DE) in one or more of the experimental conditions for the rewatering treatments in the host, we conducted Venn intersection analysis using GOVenn in the R package GOplot (Walter *et al.*, 2015). We then

used the R package GOseq (Young *et al.*, 2010) to analyse the overrepresentation of gene ontology biological processes (GO) terms in each Venn intersection. GO enrichment analysis was conducted for the sets of up- and downregulated DEGs in each Venn intersection separately. DEGs with contrasting pattern of expression between the different time points formed separate data sets for GO enrichment analysis. We also identified transcription factors (TFs), whose target genes were significantly ($P < 0.05$) overrepresented in each set of unique and shared DEGs identified by Venn intersection using the TF enrichment analysis online tool provided by the Plant Transcriptional Regulatory Map database (Tian *et al.*, 2020). The expression patterns of enriched TFs, which were significantly DE in one or more of the experimental conditions were visualized in heatmaps generated using the R package ComplexHeatmap (Gu *et al.*, 2016).

We took a selection of specific genes based on the co-expression network analysis for the combined stress treatments as well as previous studies. Then, we identified the genes which were DE in one or more of the experimental conditions during recovery with a specific consideration of the major GO terms overrepresented in the different sets of DEGs. We selected genes related to response to *C. austroafricana* infection in *E. grandis* (Naidoo *et al.*, 2013; Mangwanda *et al.*, 2015; Visser *et al.*, 2015), growth and response to drought stress in different *Eucalyptus* spp. (Thumma *et al.*, 2012; Spokevicius *et al.*, 2017; Laubscher *et al.*, 2018; Favreau *et al.*, 2019; Wierzbicki *et al.*, 2019; Takawira *et al.*, 2023), response to and recovery from drought stress in *E. grandis* (Teshome *et al.*, 2023), and the combined effect of drought stress and pathogen infection in *E. grandis* based on the co-expression network analysis in this manuscript. In addition, we used some well-known genes from *Arabidopsis* and other annual crops to select DEGs related to response to drought stress (Benny *et al.*, 2019; Balti *et al.*, 2020) and pathogen infection during recovery from drought stress (Gupta & Senthil-Kumar, 2017). We visualized the expression pattern of these genes across the different experimental conditions in heatmaps generated using the R package ComplexHeatmap (Gu *et al.*, 2016). For convenience with visualization, we manually categorized the genes according to the major signalling pathway or biological process they are involved in based on Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000), MAPMAN (Thimm *et al.*, 2004), and Plant Metabolite Network 15 (PMN 15) (Hawkins *et al.*, 2021) annotations as well as the information from the respective previous studies that investigated the individual genes.

For the pathogen transcripts, we selected the top 100 highly expressed genes for drought

stress and rewatering treatments as well as the respective controls and compared the expression patterns with *in vitro* and *in planta* data from previous studies (Mangwanda *et al.*, 2015; 2016). We also computed the differential expression of genes and conducted BLASTp against the pathogen host interactions database (PHI-base) and the carbohydrate-active enzyme annotation database (dbCAN3) to identify DEGs specifically related to the pathogenicity of *C. austroafricana*. We visualized the expression pattern of the selected genes in heatmaps generated using the R package ComplexHeatmap (Gu *et al.*, 2016).

Methods S2 Co-expression network construction, module detection, and identification of key modules and hub genes

To check for a possible batch effect among the different RNA-seq data sets as suggested in Chowdhury *et al.* (2019), we conducted principal component analysis (PCA) which showed variations between the different data sets (Fig. S1A). After conducting batch correction on the TPM matrix using the ComBat function (Johnson *et al.*, 2007) of the R package SVA (Leek *et al.*, 2012), a repeat of PCA (Fig. S1B) confirmed that the batch correction was successful.

We used the batch corrected RNA-seq data to construct two co-expression networks which consisted of 'unstressed' and 'stressed' sets of samples using similar methods of data pre-processing and co-expression network construction (Fig. S2A). The 'unstressed' set consisted of either unwounded or mock-inoculated well-watered samples while the 'stressed' set included samples exposed to *C. austroafricana* infection, drought stress, or a combination of both stresses. The mock-inoculated samples were included in the 'unstressed' set because wounding was common to all the samples in the 'stressed' group, and thus, will allow a better comparison of the individual and combined stresses. We confirmed that there were no outlier samples using cluster plots and selected genes which were expressed in at least 80% of the samples. Then, we selected 17887 genes with TPM values of at least one and standardized the expression of these genes across samples. We used Pearson correlation coefficient to measure co-expression between genes and the R package Weighted Gene Co-expression Network Analysis (WGCNA) version 1.70-3 (Langfelder & Horvath, 2008) to detect co-expression modules. Soft thresholding powers of 9 and 5, which gave a scale free topology fit R^2 of at least 0.8, were selected to create unsigned adjacency matrices for the 'unstressed' and 'stressed' sample networks, respectively. These matrices were then converted to topological overlap matrices. Module detection using hierarchical clustering and dynamic tree cutting methods with at least 50 genes in a module was followed by calculation of module eigengenes and merging close modules at a height of 0.2.

To confirm the biological relevance of the detected co-expression modules, gene ontology biological processes (GO) and transcription factor (TF) enrichment analyses were conducted using the R package GO-seq and the TF Enrichment web tool of plant transcriptional regulatory map (PlantRegMap) (Tian *et al.*, 2020), respectively. We employed two different but complementary methods for the identification of key modules with respect to the experimental conditions of interest from the two networks. For the network of the 'unstressed' samples, modules of interest were identified using module preservation analysis where the 'unstressed' set of samples was used as a reference to conduct module preservation analysis (MPA) using the WGCNA R package. Two composite module preservation statistics and cut-off points, $Z_{summary} < 10$ and $medianRank > 8$, suggested in (Langfelder *et al.*, 2011) were used to identify unpreserved modules. For the network of the 'stressed' set, the biweight midcorrelation coefficients between module eigengenes and experimental conditions (module-trait correlation analysis, MTCA) was used to identify modules with strong correlations with the target experimental conditions. GO and TF enrichment analyses were used to further select key modules and hub genes were identified from the key modules selected from both networks. For the key modules selected using MPA, 88 hub genes were identified using the absolute value of module membership (kME) > 0.9 as a selection criterion. For the key modules selected using MTCA, 305 hub genes were selected using |kME| > 0.8 and the absolute value of gene significance (GS) > 0.8 as selection criteria.

References for methods S1 and S2

Balti I, Benny J, Perrone A, Caruso T, Abdallah D, Salhi-Hannachi A, Martinelli F. 2020. Identification of conserved genes linked to responses to abiotic stresses in leaves among different plant species. *Functional Plant Biology* **48**(1): 54-71.

Benny J, Pisciotta A, Caruso T, Martinelli F. 2019. Identification of key genes and its chromosome regions linked to drought responses in leaves across different crops through meta-analysis of RNA-Seq data. *BMC Plant Biology* **19**(1): 1-18.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15): 2114-2120.

Chowdhury HA, Bhattacharyya DK, Kalita JK. 2019. (Differential) co-expression analysis of gene expression: a survey of best practices. *IEEE/ACM transactions on computational biology and bioinformatics* **17**(4): 1154-1173.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**(1): 15-21.

Favreau B, Denis M, Ployet R, Mounet F, Peireira da Silva H, Franceschini L, Laclau J-P, Labate C, Carrer H. 2019. Distinct leaf transcriptomic response of water deficient *Eucalyptus grandis* submitted to potassium and sodium fertilization. *PloS one* **14**(6): e0218528.

Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**(18): 2847-2849.

Gupta A, Senthil-Kumar M. 2017. Transcriptome changes in *Arabidopsis thaliana* infected with *Pseudomonas syringae* during drought recovery. *Scientific Reports* **7**(1): 9124.

Hawkins C, Ginzburg D, Zhao K, Dwyer W, Xue B, Xu A, Rice S, Cole B, Paley S, Karp P. 2021. Plant Metabolic Network 15: A resource of genome-wide metabolism databases for 126 plants and algae. *Journal of Integrative Plant Biology* **63**(11): 1888-1905.

Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**(1): 118-127.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**(1): 27-30.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**(1): 1-13.

Langfelder P, Luo R, Oldham MC, Horvath S. 2011. Is my network module preserved and reproducible? *PLoS Computational Biology* **7**(1): e1001057.

Laubscher M, Brown K, Tonfack L, Myburg AA, Mizrachi E, Hussey SG. 2018. Temporal analysis of *Arabidopsis* genes activated by *Eucalyptus grandis* NAC transcription factors associated with xylem fibre and vessel development. *Scientific Reports* **8**(1): 10983.

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**(6): 882-883.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**(12): 1-21.

Mangwanda R, Myburg AA, Naidoo S. 2015. Transcriptome and hormone profiling reveals *Eucalyptus grandis* defence responses against *Chrysoporthe austroafricana*. *BMC Genomics* **16**(1): 1-13.

Mangwanda R, Zwart L, van der Merwe NA, Moleleki L, Berger D, Myburg A, Naidoo S. 2016. Localization and transcriptional responses of *Chrysoporthe austroafricana* in *Eucalyptus grandis* identify putative pathogenicity factors. *Frontiers in microbiology* **7**: 1953.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D. 2014. The genome of *Eucalyptus grandis*. *Nature* **510**(7505): 356-362.

Naidoo R, Ferreira L, Berger DK, Myburg AA, Naidoo S. 2013. The identification and differential expression of *Eucalyptus grandis* pathogenesis-related genes in response to salicylic acid and methyl jasmonate. *Frontiers in plant science* **4**: 42592.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**(3): 290-295.

Spokevicius AV, Tibbits J, Rigault P, Nolin M-A, Müller C, Merchant A. 2017. Medium term water deficit elicits distinct transcriptome responses in *Eucalyptus* species of contrasting environmental origin. *BMC Genomics* **18**(1): 1-17.

Takawira LT, Bachir IH, Ployet R, Tulloch J, San Clemente H, Christie N, Ladouce N, Dupas A, Grima-Pettenati J, Myburg AA. 2023. Functional investigation of five R2R3-MYB transcription factors associated with wood development in *Eucalyptus* using DAP-seq-ML. *Plant Molecular Biology* **113**(1): 33-57.

Teshome DT, Zharare GE, Ployet R, Naidoo S. 2023. Transcriptional reprogramming during recovery from drought stress in *Eucalyptus grandis*. *Tree Physiology* **43(6)**: 979-994.

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal* **37**(6): 914-939.

Thumma BR, Sharma N, Southerton SG. 2012. Transcriptome sequencing of *Eucalyptus camaldulensis* seedlings subjected to water stress reveals functional single nucleotide polymorphisms and genes under selection. *BMC Genomics* **13**(1): 1-21.

Tian F, Yang D-C, Meng Y-Q, Jin J, Gao G. 2020. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Research* **48**(D1): D1104-D1113.

Visser EA, Mangwanda R, Becker JV, Külheim C, Foley W, Myburg AA, Naidoo S. 2015. Foliar terpenoid levels and corresponding gene expression are systemically and differentially induced in *Eucalyptus grandis* clonal genotypes in response to *Chrysoporthe austroafricana* challenge. *Plant Pathology* **64**(6): 1320-1325.

Walter W, Sánchez-Cabo F, Ricote M. 2015. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics* **31**(17): 2912-2914.

Wierzbicki MP, Christie N, Pinard D, Mansfield SD, Mizrachi E, Myburg AA. 2019. A systems genetics analysis in *Eucalyptus* reveals coordination of metabolic pathways associated with xylan modification in wood-forming tissues. *New Phytologist* **223**(4): 1952-1972.

Wingfield BD, Ades PK, Al-Naemi FA, Beirn LA, Bihon W, Crouch JA, De Beer ZW, De Vos L, Duong TA, Fields CJ. 2015. Draft genome sequences of *Chrysoporthe austroafricana*, *Diplodia scrobiculata*, *Fusarium nygamai*, *Leptographium lundbergii*, *Limonomyces culmigenus*, *Stagonosporopsis tanaceti*, and *Thielaviopsis punctulata*. *IMA fungus* **6**(1): 233-248.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**(2): 1-12.