University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Sibout, R;Proost, S;Hansen, BO;Vaid, N;Giorgi, FM;Ho-Yue-Kuang, S;Legée, F;Cézart, L;Bouchabké-Coussa, O;Soulhat, C;Provart, N;Pasha, A;Le Bris, P;Roujol, D;Hofte, H;Jamet, E;Lapierre, C;Persson, S;Mutwil, M

Title:
Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in Brachypodium distachyon

Date:
2017-08-01

Citation:
Sibout, R., Proost, S., Hansen, B. O., Vaid, N., Giorgi, F. M., Ho-Yue-Kuang, S., Legée, F., Cézart, L., Bouchabké-Coussa, O., Soulhat, C., Provart, N., Pasha, A., Le Bris, P., Roujol, D., Hofte, H., Jamet, E., Lapierre, C., Persson, S. & Mutwil, M. (2017). Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in Brachypodium distachyon. New Phytologist, 215 (3), pp.1009-1025. https://doi.org/10.1111/nph.14635.

Persistent Link:
https://hdl.handle.net/11343/293035

Article type : MS - Regular Manuscript

Author Manuscript

Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in *Brachypodium distachyon*

Richard Sibout[1], Sebastian Proost[2], Bjoern Oest Hansen[2], Neha Vaid[2], Federico M. Giorgi[3], Severine Ho-Yue-Kuang[1], Frédéric Legée[1], Laurent Cézart[1], Oumaya Bouchabké-Coussa[1], Camille Soulhat[1], Nicholas Provart[4], Asher Pasha[4], Philippe Lebris[1], David Roujol[5], Herman Hofte[1], Elisabeth Jamet[5], Catherine Lapierre[1], Staffan Persson[6] and Marek Mutwil[2]

[1]Institut Jean-Pierre Bourgin, UMR 1318, INRA, AgroParisTech, CNRS, Université Paris-Saclay, RD10, Versailles, Cedex 78026, France; [2]Max Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, 14476 Potsdam, Germany; [3]Cancer Research UK, Cambridge Institute, Robinson Way, Cambridge, CB20RE, UK; [4]Department of Cell and Systems Biology, Centre for the Analysis of Genome Evolution and Function, University of Toronto, 25 Willcocks St, Toronto, ON, M5S 3B2, Canada; [5]Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRs, UPS, Castanet-Tolosan, France; [6]School of Biosciences, University of Melbourne, Parkville, Victoria 3010, Australia

Authors for correspondence:

Marek Mutwil

Tel: +49 173 5411 619

Email: mutwil@mpimp-golm.mpg.de

Richard Sibout

Tel: +33 0 1 30 83 30 00

Email: richard.sibout@inra.fr

Received: *28 February 2017*

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/nph.14635

This article is protected by copyright. All rights reserved

Accepted: *26 April 2017*

Summary

- While *Brachypodium distachyon* is an emerging model for grasses, no expression atlas and gene coexpression network is available. Such tools are of high importance to provide insights into the function of Brachypodium genes.
- We present a detailed Brachypodium expression atlas, capturing gene expression in its major organs at different developmental stages. The data was integrated into large-scale coexpression database (www.gene2function.de), enabling identification of duplicated pathways and conserved processes across ten plant species, thus allowing genome-wide inference of gene function.
- We highlight the importance of the atlas and the platform through the identification of duplicated cell wall modules, and show that a lignin biosynthesis module is conserved across angiosperms. We identified and functionally characterized a putative ferulate 5-hydroxylase gene through overexpression of it in Brachypodium, which resulted in an increase in lignin syringyl units and reduced lignin content of mature stems, and led to improved saccharification of the stem biomass.
- Our Brachypodium expression atlas thus provides a powerful resource to reveal functionally related genes, which may advance our understanding of important biological processes in grasses.

Key words: *Brachypodium distachyon*, coexpression, comparative coexpression, expression atlas, functional modules, gene function, lignin.

Introduction

Grasses provide the bulk of human nutrition and serve as important feedstock for animals. In addition, grass crops constitute an emerging source for lignocellulosic biofuels (Sanderson, 2011). However, the major grass species used as crops are poor experimental models, as they tend to have large and complex genomes, long life cycles, demanding growth conditions and/or intellectual property restrictions (International Rice Genome Sequencing Project *et al.*, 2005; Mayer *et al.*, 2012; The International Wheat Genome Sequencing Consortium, 2014). On the other hand, Brachypodium (*Brachypodium distachyon*), a wild annual member of the *Pooideae* subfamily, has one of the smallest known grass genomes (270 Mbp) (Vogel *et al.*, 2010), similar number of gene families as rice and sorghum, small stature, short life cycle,

This article is protected by copyright. All rights reserved

Author Manuscript

self-fertilization and undemanding growth conditions. Thus, Brachypodium has all the desirable features to serve as a model system for grasses (Draper *et al.*, 2001). In addition, physiological traits, such as root and grain development, biotic and abiotic stress tolerance, cell wall composition and yield, can be easily studied in Brachypodium and outcomes may subsequently be transferred to improve other grasses. These advantages have intensified the development of efficient transformation and crossing protocols, and the establishment of mutant (T-DNA and chemically-mutagenized lines) and natural accession collections, a microarray platform, DNA clone libraries (cDNA, BAC, and transformation plasmids) and a large number of accessions (Brkljacic *et al.*, 2011).

Despite these resources, the majority of genes in monocots are annotated only by sequence similarity and functional domain detection methods, such as BLAST2GO and Interproscan (Quevillon *et al.*, 2005; Conesa & Götz, 2008). Identification of orthologous genes across multiple species is made difficult by evolutionary events such as whole-genome or segmental duplications and consequently, improved annotations via other means are clearly needed to uncover the functions of the many grass-specific gene families. Combining sequence and expression profile similarity can be devised by determining equivalent tissues between species, with the best expression profile match being termed the 'expressolog' (Patel *et al.*, 2012). Analysis of transcriptomes of *Oryza sativa*, *B. distachyon* and *Sorghum bicolor* revealed that genes with properties such as high expression in many tissues and conserved synteny tend to evolve more slowly than genes not showing these properties (Davidson *et al.*, 2012). Since genes involved in related biological processes tend to have similar expression patterns across organs, developmental stages and biotic and abiotic perturbations, identification of transcriptionally co-regulated (coexpressed) genes has been used to identify functionally related genes (Stuart *et al.*, 2003; Yu *et al.*, 2003; Persson *et al.*, 2005; Itkin *et al.*, 2013; Proost & Mutwil, 2016). In plants, coexpression analyses have been used to find genes relevant for plant viability (Mutwil *et al.*, 2010), seed germination (Bassel *et al.*, 2011), shade avoidance (Jiménez-Gómez *et al.*, 2010), cyclic electron flow (Takabayashi *et al.*, 2009), cell division (Takahashi *et al.*, 2008) and others. In coexpression networks, nodes (or vertices) correspond to genes and edges (or links) connect genes that display similar expression profiles (Lee *et al.*, 2004). To facilitate user-friendly access to these networks, several web-based tools have been generated (Lee *et al.*, 2010; Mutwil *et al.*, 2010, 2011; Ficklin & Feltus, 2011; Movahedi *et al.*, 2011; Obayashi *et al.*, 2011; Tzfadia *et al.*, 2012; De Bodt *et al.*, 2012; Proost & Mutwil, 2017).

This article is protected by copyright. All rights reserved

Author Manuscript

Apart from predicting gene function, coexpression networks can be used to identify gene modules, that is, groups of coexpressed genes, that may be conserved across species and even across different kingdoms of life, as exemplified by cell cycle, ribosome biogenesis, and proteasome gene modules (Stuart *et al.*, 2003; Zarrineh *et al.*, 2014; Gerstein *et al.*, 2014). Such conserved gene modules are useful to transfer knowledge about pathways or co-functional processes from better characterized model species to crop plants, as functionally equivalent genes are found by combining DNA sequence and coexpression contexts (Ruprecht *et al.*, 2011; Mutwil *et al.*, 2011; Park *et al.*, 2013; Tzfadia *et al.*, 2016). Several tools, including PlaNet, ATTED-II, PlantGenIE, CoExpNetViz are available to detect conserved modules (Mutwil *et al.*, 2011; Aoki *et al.*, 2015; Sundell *et al.*, 2015; Serin *et al.*, 2016; Tzfadia *et al.*, 2016; Proost & Mutwil, 2017). It is important to note that analysis of coexpression networks faces several limitations. First, coexpression relationships are dependent on the used expression data. For example, if pollen expression data is absent from the dataset, pollen-specific modules will not be detected. Second, the conserved modules could obtain new expression patterns and functions during plant evolution. Third, *c*. 40% of protein coding genes are not represented on many microarrays, which could result in certain modules not being detected (Mutwil *et al*., 2011).

Recent studies revealed that gene modules can also duplicate within one organism (Ruprecht *et al.*, 2016, 2017b). This implies that if the gene module is associated with a certain function, the duplicated gene modules may perform related functions in different parts of a plant. To study conserved and/or duplicated gene modules, we recently updated the PlaNet platform with tools that allow their identification and analysis (Ruprecht *et al.*, 2016). Furthermore, by combining comparative genomic analyses with coexpression networks, we showed that it is possible to deduce when certain gene modules appeared in plant evolution (Ruprecht *et al.*, 2017a,b).

Broad patterns of gene expression and the identity of gene modules in Brachypodium are largely unknown. To address this knowledge gap, we generated a comprehensive expression atlas for Brachypodium that captures gene expression in major organ types and developmental stages, and incorporated it into the coexpression database PlaNet (www.gene2function.de). We used these tools together with transgenic approaches to identify and characterize a putative ferulate 5-hydroxylase gene and show that overexpression of this gene affected lignin architecture in Brachypodium, and led to improved saccharification of the corresponding biomass.

This article is protected by copyright. All rights reserved

Author Manuscript

Materials and Methods

Experimental growth conditions and organ sampling

The collected samples were grown in three locations (Table 1). Toulouse, France (T) growth chamber: seeds of *Brachypodium distachyon* (L.) P. Beauv line Bd21 were maintained in the dark in a wet chamber at 4°C during 4 d to favor a homogeneous germination. Seeds were then sowed in a Tref H4 compost (Jiffy France SARL, Trévoux, France). They were put in a growth chamber with the following day : night conditions: 16 h 25°C : 8 h 22°C and 80% humidity. Pots were covered with Saran® wrap until the first leaves reached a size of *c*. 3 cm. Two month-old plants were used for the experiments. Two types of organs at two different developmental stages were sampled: young leaves (from 3 to 6 cm in length) and fully-expanded mature leaves; apical internodes (first internodes at the top of the culm) and basal internodes (first lignified internodes at the basis of the culm). The plants were not flowering at the time of sampling and contained five to seven internodes per culm. Versailles, France (P) growth chambers: seeds were maintained at 4°C for 3 d, germinated in soil and grown in a Phytotron with the following day : night conditions: 20 h 23°C : 4 h 21°C and 65% humidity. Versailles glasshouse (V) material was grown in soil, with the following day : night conditions: 20 h 23°C : 4 h 21°C. Samples were collected, immediately placed in liquid nitrogen and then stored at -80°C until RNA extraction. The plants contained two to three internodes at 10 d after germination (DAG), three to four internodes at 17 DAG, five to seven internodes at 27 DAG and five to nine internodes at 30–100 DAG. The plants flowered 25 DAG.

RNA preparation, labelled cDNA synthesis, and microarray hybridization

Samples from Versailles: the plant material was ground in liquid nitrogen and extracted with RNeasy QIAGEN kit with a DNase step as recommended by instructions. Samples from Toulouse: preparation of total RNAs was performed using the SV Total RNA isolation system (Z3100, Promega, Charbonnières-les-Bains, France). The material was treated with RQ1 RNase-Free DNase according to manufacturer's instructions (Promega) before being precipitated with ethanol (2 vol) and 3 M sodium acetate pH 5.2 (0.1 vol) for storage at -80°C. To assess RNA quantity and purity, NanoDrop™ 2000C Spectrophotometer was used to measure 260–280 nm ratio, where the ratio close to 2.0 was considered suitable for microarray analysis. Quality and quantity of each RNA sample was further checked by Agilent 2100 Bioanalyzer and the RIN (RNA integrity number) values were above 6.8. cDNA synthesis,

This article is protected by copyright. All rights reserved

Author Manuscript

array hybridisation, scanning were done by AtlasBiolabs (http://www.atlas-biolabs.com) according to Affymetrix' GeneChip™ technology.

Microarray data analysis

The microarray platform used in this study was the BrachypodiumV1 chip from Affymetrix (Priest *et al.*, 2014). The chip contains 2142558 probes, whose nucleotide sequences we mapped over the Brachypodium CDS (coding sequence) database v1.2 (Vogel *et al.*, 2010) using blastn-short with default parameters. Only perfect matches were kept (25 nucleotides. BLAST bit score 50.1). The analysis showed that 1561378 probes had at least one perfect match to a Brachypodium CDS. A total of 333072 probes with ambiguous matches to more than one CDS were discarded **[Author, please confirm inserted text 'A total of' is correct]**. The surviving probes covered a total of 23564 Brachypodium CDS models (out of a total of 31029 described in the Brachypodium v1.2 transcriptome). A list of probesets (one per CDS) was generated and stored into a new binary CDF annotation file that was generated using the makecdfenv R package (https://bioconductor.org/packages/release/bioc/html/makecdfenv.html) and is available as Supporting Information Notes S1. The microarrays were RMA (Robust Multi-Array Average) normalized (Irizarry *et al.*, 2003) with Affymetrix Power Tools (http://www.affymetrix.com/estore/partners_programs/programs/developer/tools/powertools.affx). To infer similarities between the samples, we have calculated the euclidean distance between the expression vectors (samples), performed hierarchical clustering with complete linkage method and visualized the sample clustering as a dendrogram.

Functional annotation of Brachypodium genes

Gene annotations, pfam domains and gene ontology (GO) terms for genome version 1.2 were obtained from Phytozome V11 (Goodstein *et al.*, 2012) (http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=PhytozomeV11). Gene families for Brachypodium were downloaded from PLAZA2.5 (http://bioinformatics.psb.ugent.be/plaza/versions/plaza_v2_5/, Proost *et al.*, 2009).

Sequence homology search

To identify Brachypodium homologs of ferulate 5-hydroxylase (AtF5H, *At4g36220*) and caffeoyl coenzyme A O-methyltransferase 1 (CCoAOMT1, *At4g34050*), we blasted their

This article is protected by copyright. All rights reserved

Author Manuscript

protein sequences against Brachypodium proteome at Phytozome https://phytozome.jgi.doe.gov/pz/portal.html#!search?show=BLAST.

Construction and clustering of highest reciprocal rank coexpression networks

The normalized expression matrix (Table S1) was used to generate highest reciprocal rank (HRR) coexpression network with standard edge cut-off of 30 (Table S2; Mutwil *et al.*, 2010) The edges in the HRR network indicate how strongly coexpressed the connected genes are, with value of 0 showing higher coexpression, while value of 30 indicating lower coexpression. To obtain the clusters, we used heuristic cluster chiseling algorithm (HCCA), which is optimized for HRR-based networks, with standard parameters (step-size = 3, Mutwil *et al.*, 2010). The coexpression networks and clusters were uploaded into PlaNet (Mutwil *et al.*, 2011).

Data availability

The raw microarray data is available from EMBL ArrayExpress as accession E-MTAB-5491. The RMA-normalized data can be found in Table S2.

Tissue specificity measure

For each gene's expression profile, we calculated the specificity measure (SPM) value based on the formula shown later, where x is the list of a gene's expression values in all samples, and x_i indicates the expression value in sample i (Xiao *et al.*, 2010). The value can range from 0 to 1, where 0 indicates that a gene is not expressed in a given sample, while value of 1 indicates that the gene is expressed specifically in a given sample.

$$SPM = \frac{x_i^2}{\sum_{i=1}^{n} x_i^2}.$$

To set the cutoff required to ensure a gene is expressed specifically in a given sample, we generated 100000 random expression profiles with the same mean and SD as a randomly selected existing profile. Then, for each of the 100000 profiles, the maximum SPM value was determined. The resulting distribution of the maximum SPM values was then used to determine the cutoff above which <5% of the randomly generated profiles were found. Genes with SPM > 0.41 in a given sample were deemed to be preferentially expressed in the sample.

This article is protected by copyright. All rights reserved

Author Manuscript

The preferentially expressed genes were used to calculate MapMan enrichment analysis of the samples.

MapMan and GO enrichment analyses

MapMan terms for Brachypodium were obtained from http://mapman.gabipd.org/web/guest/mapmanstore, while GO annotations for Brachypodium v1.2 genome were downloaded from Phytozome (https://phytozome.jgi.doe.gov) (Goodstein *et al.*, 2012). MapMan and GO enrichment analysis of a module/neighbourhood/cluster/sample was performed by using hypergeometric distribution, which estimates the probability of finding a number of genes with a given ontology term, in a collection of a given size (e.g. all genes in a cluster), when considering the number of genes that are assigned to the given ontology term. To account for multiple hypothesis testing, we have applied false discovery rate (FDR) correction to the *P*-values (Benjamini & Hochberg, 1995).

Detection of conserved, duplicated and overlapping modules in PlaNet

To detect similar modules, we applied the FamNet pipeline, which identifies coexpression neighbourhoods that contain same gene families and Pfam domains (Ruprecht *et al.*, 2016). Briefly, when two neighbourhoods of sizes A and B are compared, the number of pfam domains and gene families (labels) in common between the two neighborhoods is counted. To estimate if the neighbourhoods are significantly similar, an empirical *P*-value is obtained by comparing the observed number of labels with the number of labels obtained when comparing two random neighbourhoods of size A and B. The number of permutations was set to 10000, and the FamNet pipeline rejected conserved/duplicated modules with *P*-value>0.05. Overlapping modules were found by identifying neighbourhoods that have at least one gene in common. PlaNet indicates if similar neighbourhoods can be found in other species (conserved modules) or within same species (duplicated modules) and visualizes these relationships as a module network (Ruprecht *et al.*, 2016). The nodes in the module network correspond to modules (i.e. groups of coexpressed genes), while the edge colours (blue, green and orange) indicate conserved, duplicated and overlapping modules, respectively.

Brachypodium genetic transformation

Brachypodium plants were genetically transformed using a method adapted from (Vogel & Hill, 2008) which is outlined in Methods S1.

This article is protected by copyright. All rights reserved

Author Manuscript

Quantitative RT-PCR analysis

Quantitative PCR were performed as described in (Bouvier D'Yvoire *et al.*, 2013). Total RNA were extracted from Brachypodium leaves using Qiagen RNeasy mini kit (http://www.qiagen.com). One microgram of RNA was retro-transcribed using a SuperScript II reverse transcriptase kit (Invitrogen, http://fr-fr.invitrogen.com). Quantitative RT-PCR **[Author, please insert expansion for 'RT-PCR']** was performed on the Eppendorf Realplex2 Mastercycler using the SYBR Green kit (Bio-Rad) and the following program: one activation cycle for 10 min at 95°C, 40 amplification cycles for 10 s at 95°C, 15 s at 60°C and 20 s at 72°C, and one melting curve cycle from 65 to 95°C. Quantitative RT-PCR data for *Bradi3g30590* were normalized with the housekeeping gene SAMDC encoding a S-adenosylmethionine decarboxylase (Hong *et al.*, 2008). Primers used are BdSamDCfw: CGGCAAGCTTGCTAATCTGCTGGAAT, BdSamDCrev: CAGAGCAACAATAGCCTGGCTGGC and BdF5Hfw: TCGACATGGGCGACGTCTTCGG and bdF5Hrev: GCGTCACCGTCGGCGTTCTGCAGGG.

Determination of lignin monomer levels by thioacidolysis

The material consisted of green plants before flowering (20 DAG) and mature dried plants (100 DAG). The analysis was performed as outlined in (Lapierre *et al.*, 1985) and Methods S1.

Determination of ferulic acid and *p*-coumaric acid levels

The analysis was performed on extractive-free samples as previously described (Sibout *et al.*, 2016) and outlined in Methods S1.

Klason lignin content determination

Klason lignin content was measured from extractive-free samples by the Klason method (Dence, 1992; Méchin *et al.*, 2014) which is outlined in Methods S1.

Saccharification assay

Saccharification assays were performed as described by (Berthet *et al.*, 2011) and outlined in Methods S1.

Microscopy and histochemistry

Staining of the cross-sections are outlined in the Methods S1.

This article is protected by copyright. All rights reserved

Author Manuscript

Statistical analysis of thioacidolysis results, Klason lignin content and saccharification efficiency

Statistics used for chemistry data analysis were made with PAST3 (Hammer *et al*. (2001) **[Author, please confirm amended citation 'Hammer *et al*. (2001)' is correct]** http://palaeo-electronica.org/2001_1/past/issue1_01.htm). One-way ANOVA was used to detect significant differences. Homogeneity of variance was tested using Levene's test. Pairwise comparisons were tested with Tukey's HSD test.

Results and Discussion

Generation of a Brachypodium expression atlas

We used an Affymetrix Brachypodium Genome Array (BradiAR1b520742) microarray platform to construct a gene expression atlas for Brachypodium (Priest *et al.*, 2014). We analysed 117 microarrays covering nine organ types during different developmental stages (given as DAG, days after fertilization (DAF), days after heading (DAH), or years) of Brachypodium grown in growth chambers in Versailles (P samples), in glasshouse in Versailles (V), in *in vitro* culture (I) and in glasshouse in Toulouse (T, Table 1; Fig. 1). To provide easy access to the expression data, we created a Brachypodium eFP browser instance (http://bar.utoronto.ca/efp_brachypodium/cgi-bin/efpWeb.cgi), which visualizes gene expression by colour coding organs in an image corresponding to the average gene expression level in these organs (Winter *et al.*, 2007).

The normalized expression data (Table S1) was used to generate a sample similarity dendrogram (Fig. S1), which revealed an expected clustering of replicates. To visualize transcriptomic profile similarities among the sampled organs, we generated another dendrogram based on averaged replicates (Fig. S2). The dendrogram revealed that organ identity and growth conditions had strong influence on clustering of the samples. The influence of organ identity was observed by clustering of leaf developmental series (V: leaf 10-60 DAG) and the seed developmental series (V: whole grain and endosperm). Growth conditions influenced clustering as well, as leaves grown in the growth chamber in Toulouse (T: young leaf, T: mature leaf), were more similar to other samples grown in the same conditions (T: first internodes), than to leaves grown in growth chamber in Versailles (P: Leaf 10-60 DAG, Fig. S2). While growth condition-dependent clustering of samples is surprising, a study of grapevine transcriptome revealed that samples clustered according to organ maturity, rather than organ identity (Fasoli *et al.*, 2012).To identify genes that are preferentially

This article is protected by copyright. All rights reserved

Author Manuscript

expressed in the sampled organs and developmental stages, we used the specificity measure (SPM), which can range from 0 (gene not expressed in a given sample) to 1 (gene expressed exclusively in a given sample, Xiao *et al.*, 2010). By identifying the sample with the highest SPM value for a given gene, we could infer which genes were preferentially expressed in a given organ and developmental stage (Table S3). To determine a threshold above which a gene is specifically expressed in a given sample, we have calculated maximum SPM values for randomized gene expression profiles with the same mean and standard deviation as 100000 randomly selected genes. The resulting distribution of SPM values revealed that 95% the randomized profiles had the maximum SPM value below 0.41 (Fig. S3), which we have used as a threshold to determine the specificity of gene expression. We observed the highest amount of specifically expressed transcripts in seeds (3966 preferential transcripts), leaves (1302 transcripts) and internodes (1176, Table 1). Different developmental stages of an organ also showed specific expression, as roots at 35 DAG contained more specifically expressed transcripts (838) than 10 DAG roots (303, Table 1).

To uncover which biological processes are preferentially expressed at the different developmental stages, we undertook MapMan ontology enrichment analysis, which revealed specific expression of numerous biological processes in the captured organs and developmental stages (Table S4). Among others, we observed that the transcript level of photosynthesis-related genes was low in etiolated shoots, but was strongly increased upon shift to light, while the wax biosynthesis pathway was expressed in young spikelets (FDR adjusted *P*-value < 0.05, Table S4). To conclude, the gene expression atlas can reveal spatio-temporal gene expression and activity of biological processes in Brachypodium.

Integration of Brachypodium transcriptome into PlaNet database

Biological networks, including those based on expression data, typically show a scale-free topology (also called power law behaviour), where many nodes have few connections, and few nodes have many connections (Barabási & Bonabeau, 2003). Scale-free biological networks also show a small-world effect, where any two nodes are connected with a path of few edges only (Barabási & Oltvai, 2004). To investigate if the generated expression data produces the expected scale-free networks, the data was used to construct a coexpression network based on the Pearson correlation coefficient (PCC), with the PCC cut-off set to 0.7 (Usadel *et al.*, 2009). Indeed, the Brachypodium coexpression network followed a scale-free topology (Fig. 2a), which is represented by points forming a line with a negative slope. Since scale-freeness is a good common topology of biological networks (Langfelder *et al.*, 2008), and the plot

This article is protected by copyright. All rights reserved

Author Manuscript

resembles those observed for other species (Zhang & Horvath, 2005; Mutwil *et al.*, 2010), we concluded that the Brachypodium expression data reflects an appropriate network topology. The expression data was uploaded into the PlaNet database (Mutwil *et al.*, 2011), which offers a variety of tools to mine co-expression networks of four monocot species (*B. distachyon*, *Hordeum vulgare*, *O. sativa*, *Triticum aestivum*), five dicot species (*Arabidopsis thaliana*, *Medicago truncatula*, *Populus trichocarpa*, *Glycine max*, *Nicotiana tabacum*) and one bryophyte species (*Physcomitrella patens*). The gene expression, coexpression relationships and GO enrichment analyses for each gene's coexpression neighbourhood can be accessed by a search box via the front page of PlaNet.

We exemplify these functionalities with *Bradi4g40780*, a photosystem I reaction center subunit N (*PsaN*, http://aranet.mpimp-golm.mpg.de/responder.py?name=gene!bdi!1039). As expected for a gene involved in photosynthesis, *PsaN* was lowly expressed in roots, but highly expressed in de-etiolated shoots (eFP browser output: Fig. 1, PlaNet output: Fig. 2b). Genes coexpressed with *PsaN* are shown as an interactive coexpression network, containing all genes within two steps away from the query gene (i.e. second neighbourhood). Nodes in this network represent genes, while edges connect coexpressed genes (Fig. 2c). PlaNet indicates which genes belong to the same gene family and contain same domains by a node colour/shape combination, for example, all genes represented by a beige circle belong to the family HOM086919. A more detailed description of the genes can be obtained by moving the mouse pointer over a node. The first neighbourhood of *PsaN* (i.e. genes directly connected to *PsaN*) identified other genes involved in photosynthesis (*Psi_PsaK*, *PsbQ* and *PsaD* pfam domains; Fig. 2c; Allen *et al.*, 2011), while GO enrichment analysis indicated that *PsaN* neighbourhood was significantly (FDR adjusted *P*-value < 0.05) enriched for genes involved in photosynthesis. As the neighbourhood of *PsaN* is involved in photosynthesis, genes belonging to the uncharacterized gene family HOM086919 are also likely to be involved in the same biological process and are thus good candidates for further functional analysis in the context of photosynthesis.

Functional analysis of network clusters

Coexpression networks contain groups of densely connected genes (clusters), that can represent distinct biological processes (Stuart *et al.*, 2003). Coexpression relationships of a gene can be viewed by investigating the neighbourhood (module) of the gene (e.g. *PsaN* in Fig. 2b), or the cluster the gene belongs to. While a neighbourhood is centred on a single gene, clusters are identified from global topological properties of the networks, and can associate

This article is protected by copyright. All rights reserved

Author Manuscript

genes that are not directly coexpressed. PlaNet uses the HCCA which was optimized to cluster HRR-based co-expression networks used in PlaNet (Mutwil *et al.*, 2010).

The HCCA algorithm partitioned the Brachypodium coexpression network into 182 non-overlapping clusters (Table S5). To elucidate biological functions of these clusters, we calculated the MapMan ontology term enrichment of the genes found in the clusters. The analysis revealed that 24 out of 35 of the major biological categories defined by MapMan are significantly enriched in at least one cluster (FDR adjusted *P*-value < 0.05), with exception of C1 metabolism, S-assimilation and micro-RNA categories (Fig. 3a, black arrows; Table S5). Considering the importance for Brachypodium as model organism for second-generation biofuel crop species, we examined the content of cluster 48 (http://aranet.mpimp-golm.mpg.de/responder.py?name=gene!bdi!c48), which is significantly enriched for cell wall biosynthetic process (Fig. 3a; Table S5). This module contains known components of cellulose biosynthesis, such as cellulose synthases (brown squares, CESA, McFarlane *et al.*, 2014), COBRA (light-green square, Brown *et al.*, 2005), POM2/CSI1 (blue triangle, Bringmann *et al.*, 2012), CTL (yellow square, Sánchez-Rodríguez *et al.*, 2012), exostosins (cyan hexagon, EXO, Madson *et al.*, 2003), KORRIGAN (green triangle, KOR, Vain *et al.*, 2014) **[Author, please ensure that all references to coloured symbols are given with the associated figure number, if appropriate]**, and others (Fig. 3b; Ruprecht *et al.*, 2011; McFarlane *et al.*, 2014). Presence of MYB and NAM/NAC transcription factors suggested that they are also regulating secondary cell wall deposition in Brachypodium (Nakano *et al.*, 2015). Furthermore, the presence of several key lignin biosynthesis enzymes (PAL/PTAL, dark green circles; C3H, violet triangles; and 4CL, violet diamond; reviewed in Vanholme *et al.*, 2012 **[Author, please ensure that all references to coloured symbols are given with the associated figure number, if appropriate]**) indicated that synthesis of cellulose and lignin is coordinated in Brachypodium. This finding is in line with a previous analysis, which revealed a transcriptional association between cellulose and lignin biosynthesis in seven angiosperms (Ruprecht *et al.*, 2011). Similar to gene pages, the cluster pages also contain information about gene IDs, annotations, enriched GO terms, pfam domains and families of the genes found in the clusters. Thus, when combined with MapMan enrichment analysis, these clusters are a good starting point for discovery of genes relevant for many biological processes and pathways in Brachypodium (Table S5).

Duplication of cell wall modules in Brachypodium

This article is protected by copyright. All rights reserved

Author Manuscript

Previous studies showed that cell wall biosynthesis modules are duplicated in Arabidopsis (Ruprecht *et al.*, 2016). To investigate whether Brachypodium modules are duplicated, we analysed the second neighbourhood of several genes that are central to cell wall synthesis. We exemplify this approach by using *Bradi2g16560* belonging to a fasciclin arabinogalactan (FLA) gene family (http://aranet.mpimp-golm.mpg.de/responder.py?name=gene!bdi!18410), which is often transcriptionally associated with cell wall biosynthesis (Fig. 3; Ruprecht *et al.*, 2011; Tan *et al.*, 2012). GO enrichment analysis found on the gene page of *Bradi2g16560* revealed that the gene is part of a module involved in an aspect of cell wall biosynthesis (GO:0044036: cell wall macromolecule metabolic process). PlaNet detects whether there are other neighbourhoods (modules) within the genome-wide Brachypodium coexpression network that are similar to the *Bradi2g16560* neighbourhood (see http://aranet.mpimp-golm.mpg.de/publications.html for examples of such analyses). We refer to these similar neighbourhoods within one species as duplicated modules. Indeed, we observed other duplicated modules similar to *Bradi2g16560*, supporting that the cell wall biosynthesis module is duplicated in Brachypodium (Fig. 4a, duplicated modules connected by green edges). The genes *Bradi4g34420*, *Bradi1g76630*, *Bradi2g16570*, *Bradi2g46120* and *Bradi2g23270*, which represent centres of the modules, show similar expression profiles to each other (Fig. 4b), and these genes are connected in the coexpression network (Fig. S4). Consequently, the second neighbourhoods of the five genes are overlapping and constitute one large module (Fig. 4a, orange edges). Since *Bradi4g34420* module showed the highest similarity to *Bradi2g16560* (15 gene families and pfam domains in common), and GO enrichment analysis suggested its function in cell wall biosynthesis (GO:0016762: xyloglucan:xyloglucosyl transferase activity, http://aranet.mpimp-golm.mpg.de/responder.py?name=gene!bdi!19800), we selected it as a representative of this large module and used PlaNet's NetworkComparer tool to compare it with *Bradi2g16560* (Fig. 4a, black boxes).

While both modules are expressed in shoots, first and second internodes, *Bradi2g16560* is also expressed in peduncle and spikelet internode (Fig. 4b). The gene module contents revealed presence of cell wall biosynthesis-related families (Fig. 4c), such as cellulose synthases (white circles), glycosyl hydrolases 9 (red circles, glucanase, Sánchez-Rodríguez *et al.*, 2012), glycosyl hydrolases 1 (glucan endo-1,3-β-glucosidase, orange circle), DUF579 (violet square, involved in xylan synthesis, Brown *et al.*, 2011), exostosins (beige circle, xyloglucan synthesis, Madson *et al.*, 2003) and peroxidases (light green circle, modifying cell wall polymers, Francoz *et al.*, 2015) **[Author, please ensure that all**

This article is protected by copyright. All rights reserved

Author Manuscript

references to coloured symbols are given with the associated figure number, if appropriate].

To suggest a biological role of these modules, we checked the annotation of the cellulose synthases (*CESA*) genes by moving the mouse pointer over the nodes in the interactive coexpression network, and found that *Bradi2g16560* module contained primary (CESA1, 5, 6 and 9, Persson *et al.*, 2007) and secondary (CESA 4, 7 and 8, Brown *et al.*, 2005; Persson *et al.*, 2005) cell wall cellulose synthases, suggesting that this module is involved in biosynthesis of both cell wall types. Interestingly, both modules contained cellulose synthase-like F family protein (CSLF, a mixed-linkage glucan synthase, Burton *et al.*, 2006), indicating that the duplicated cell wall modules might be involved in biosynthesis of this polymer in different organs and developmental stages of Brachypodium. While the exact function of these modules in grasses is currently unknown, the genes found in these modules are likely involved in an aspect of cell wall biosynthesis.

Lignin biosynthesis module is conserved in Arabidopsis and Brachypodium

Lignins are recalcitrant cell wall biopolymers that hamper hydrolysis of cellulose into fermentable glucose. The knowledge of key molecular players involved in lignification of biofuel crops is a prerequisite to engineer the lignocellulosic biomass (Vanholme *et al.*, 2012).

To gain insights into lignin biosynthesis in Brachypodium and to demonstrate how PlaNet can be used to find conserved modules and to transfer knowledge between species, we queried the PlaNet database with a known phenylalanine ammonia-lyase (PAL) gene from Brachypodium (*BdPAL2*, *Bradi3g49260*, http://aranet.mpimp-golm.mpg.de/responder.py?name=gene!bdi!8486, Cass *et al.*, 2015). PAL is the first enzyme involved in the phenylpropanoid pathway that eventually leads to lignin formation. The gene page of *BdPAL2* contains a gene module network, which showed other modules similar to *BdPAL2*. As lignin can be found in all flowering plants, we found lignin gene modules in Arabidopsis, barley, poplar, rice, soybean and wheat (Fig. 5a, for brevity, only Arabidopsis, Brachypodium and rice are shown). Since the lignin biosynthesis pathway has been thoroughly studied in Arabidopsis (Vanholme *et al.*, 2012), we chose to compare *BdPAL2* with the Arabidopsis *At2g37040* (*AtPAL1*), *At3g53260* (*AtPAL2*) and At3g10340 (*AtPAL4*) modules. As *AtPAL1*, *2* and *4* modules overlap (Fig 5a, connected by orange lines **[Author, please confirm amended text 'lines' is correct]**), we can consider these genes part of the same large module as these genes are coexpressed.

This article is protected by copyright. All rights reserved

Author Manuscript

The analysis revealed that all of the genes involved in lignin biosynthesis can be found within the selected gene modules (Fig. 5b,c), with the exception of genes encoding *caffeoyl coenzyme A O-methyltransferase* 1 (*CCoAOMT1*) and *ferulate 5-hydroxylase* (*F5H*) in Brachypodium module. Interestingly, a gene encoding Tyrosine Ammonia-Lyase 1 (*BdTAL1)* that converts tyrosine into *p*-coumaric acid as an alternative to the PAL activity in Brachypodium (TAL, Fig. 5b; Barros *et al.*, 2016) was present in the module (*Bradi3g49250*, turquoise circle). In addition, we found a gene encoding a 2-deoxy-D-arabino-heptulosonate 7-phosphate synthase, which acts as the first step of the shikimate pathway (*At4g39980* and *Bradi1g60750*, yellow triangles **[Author, please ensure that all references to coloured symbols are given with the associated figure number, if appropriate]** , Entus *et al.*, 2002). This further highlights the metabolic link between phenylalanine/tyrosine and lignin biosynthesis.

We found three putative copies of CSE homologs (*Bradi3g13420*, *Bradi3g29970*, *Bradi3g31440*) in the PAL module (Fig. 5, beige triangle, *At1g52760*, hydrolase_4), despite the fact that crude protein extracts from stems of Brachypodium and corn (*Zea mays*) exhibited only weak caffeoyl shikimate esterase (CSE) activity with caffeoyl shikimate (Ha *et al.*, 2016). We also found genes encoding dirigent proteins in the Brachypodium module (green square, Pickel *et al.*, 2010 **[Author, please ensure that all references to coloured symbols are given with the associated figure number, if appropriate]**). These proteins are involved in biosynthesis of lignans and coexpression of dirigent proteins with lignin biosynthetic genes was previously observed in Arabidopsis (Wang *et al.*, 2013). Furthermore, a gene encoding *p-coumaroyl-CoA monolignol transferase* (*Bradi2g36910*, *PMT1*) involved in the acylation of sinapyl alcohol by *p*-coumaric acid (CA) in Brachypodium lignin was present in the module (Fig. 5; Petrik *et al.*, 2014).

Genes encoding transcription factors that are referred to as master regulators of lignification were also present. For instance, *AtMYB63* (*At1g79180*, blue circle), coding for a R2R3 MYB domain transcription factor, is a known regulator of lignin biosynthesis in Arabidopsis) (Zhou *et al.*, 2009), and the Brachypodium lignin module contained three uncharacterized *AtMYB63* homologues (*Bradi2g17982*, *Bradi5g20130*, *Bradi2g47590*). The modules also contained various transporters, among which an ABC transporter (dark blue triangle, *At3g16340*) is involved in p-coumaryl alcohol transport (Alejandro *et al.*, 2012). Other uncharacterized transporters were present, such as amino acid transporter (light blue triangle, *At3g56200*), amino acid exporter (gray triangle, *At2g40900*, Mueller *et al.*, 2015) and various major intrinsic proteins (purple square) **[Author, please ensure that all references to**

This article is protected by copyright. All rights reserved

Author Manuscript

coloured symbols are given with the associated figure number, if appropriate]. The uncharacterized genes present in these conserved modules can serve as candidates for further functional analysis of their possible involvement in lignin biosynthesis in angiosperms.

Identification of a putative F5H gene involved in lignin formation in Brachypodium

The absence of clear homologs of *F5H* and *CCoAOMT1* in the BdPAL2 module was surprising as they perform important steps in the biosynthesis of lignins. This might likely be due to the fact that the expression of the genes are slightly different compared with the other genes in the pathway. To see if we could find the *F5H* and *CCoAOMT1* genes associated with lignin synthesis in Brachypodium we first identified potential *F5H* genes by sequence homology searches. We found four proteins (*Bradi1g77740*, *Bradi1g36790*, *Bradi1g77597*, *Bradi3g30590*) that showed >50% sequence identity to the AtF5H (Table S6). We then used PlaNet's expression analysis to identify which of the corresponding four genes is expressed highly in Brachypodium stem, as the *PAL1* module genes are highly expressed in this tissue (Fig. S5). Querying PlaNet with the first three genes revealed expression profiles that were not stem-specific: *Bradi1g77740* (strongly expressed in the developing seed and the endosperm), *Bradi1g77597* (endosperm) and *Bradi1g77597* (35 DAG roots, de-etiolated shoots, endosperm, Fig. S4a). Conversely, by comparing the expression profile of *Bradi3g30590* and *BdPAL2* (*Bradi3g49260*), we observed that *Bradi3g30590* showed expression in specific developmental stages of the lignified tissues, such as mature root system (35 DAG), first and second internodes (35 DAG), while *BdPAL2* was more ubiquitously expressed (Fig. 6a, full expression profiles found in Fig. S5b). Based on the expression profile, we assumed that *Bradi3g30590* was not present in the BdPAL module (Fig. 5), because it was only expressed in a subset of developmental stages. However, we also observed that *Bradi3g30590* neighbourhood contained a *CAD* gene that is associated with the BdPAL module (Figs S6, 5, dark green circle; Bouvier D'Yvoire *et al.*, 2013), suggesting that *Bradi3g30590* and the BdPAL2 module are partially coexpressed. Taken together, based on the sequence similarity and high expression in stems, we hypothesized that *Bradi3g30590* was the most likely *F5H*-encoding gene to be involved in sinapyl alcohol synthesis. To identify gene candidates for *CCoAOMT* activity in lignin biosynthesis, we applied a similar homology search and arrived at nine putative *CCoAOMT* genes. However, we could not identify a clear candidate, as none of these genes showed specific expression in stems (data not shown).

To corroborate that *Bradi3g30590* is associated with lignin formation, and thus illustrate how PlaNet can identify functionally relevant genes, we transformed wild-type

This article is protected by copyright. All rights reserved

Author Manuscript

Brachypodium (WT, accession Bd21-3) with a cDNA clone of *Bradi3g30590* under the control of the maize ubiquitin promoter. Despite several studies in dicots (Meyer *et al.*, 1998; Franke *et al.*, 2000; Stewart *et al.*, 2009; Weng *et al.*, 2010; García *et al.*, 2014), the impact of F5H overexpression has not been investigated in grasses. We selected two transgenic lines, which showed 30- and 68-fold higher expression of the transgene in the T_3 progenies, as compared with WT (Fig. 6b), and studied lignin content and structure before and after flowering (Table 2). Surprisingly, we observed a decrease of Klason lignin content by 18% in the *Bradi3g30590*-OE lines when compared with WT (Table 2). A decrease of lignin was similarly observed in Arabidopsis and poplar upon F5H overexpression (Franke *et al.*, 2000; Stewart *et al.*, 2009; Weng *et al.*, 2010; Vanholme *et al.*, 2012), but the biological cause behind these results (higher metabolic cost of sinapyl alcohol synthesis, shunt of the pathway) remains unknown.

Since overexpression (OE) of F5H can dramatically increase the amount of S units in poplar or Arabidopsis lignin (Stewart *et al.*, 2009), we studied lignin structure by thioacidolysis. This method provides H, G and S thioethylated monomers from H, G and S units that are exclusively involved in arylglycerol-β-ether structures (Bouvier D'Yvoire *et al.*, 2013). Not unexpectedly, the analysis revealed an average increase of 25% of the S thioacidolysis monomers in the overexpressors, leading to an increase of the S : G unit ratio on average from 2.3 (WT) to 8.1 (OE) (Table 2). Curiously, we also observed a slight but significant increase of H units in the *Bradi3g30590*-OE lines. To identify where the S units are deposited in the stem, we stained Brachypodium stem cross-sections with Maüle method (Bouvier D'Yvoire *et al.*, 2013), which colours S and G units with red and orange, respectively. At mature stage, the staining did not discriminate the transgenics from the WT (data not shown), but we observed an increased red staining in the mestome and intra-fascicular fibres of vascular bundles of the *Bradi3g30590*-OE lines at heading stage (Fig. 6c,d). Interestingly, the 25% increase of S-unit lignins was only accompanied by a very modest increase of 5-hydroxy-guaiacyl (5-OH-G, from 0.8% in WT to 1.1% in mature dried transgenics). This compound, which results from F5H activity, is often detected in F5H overexpressors or when caffeic acid O-methyltransferase (COMT) becomes limiting to produce sinapyl alcohol (Ralph *et al.*, 2001a,b; Weng *et al.*, 2010).

In contrast to most common dicots, grass cell walls are abundantly decorated with hydroxycinnamic acids which stem from the lignin pathway. Ferulic (FA) and *p*-coumaric acids (CA) can be linked both to lignins and to arabinoxylans (Hartley & Ford, 1989; Ralph, 2010; Withers *et al.*, 2012; Petrik *et al.*, 2014; Karlen *et al.*, 2016; Sibout *et al.*, 2016). We

This article is protected by copyright. All rights reserved

Author Manuscript

quantified hydroxycinnamic acid content in cell wall of WT and *Bradi3g30590*-OE lines by mild alkaline hydrolysis. While CA is mainly ester-linked to S lignin units in Brachypodium (Withers *et al.*, 2012; Petrik *et al.*, 2014), the increased frequency of S units in Bradi3g30590-OE did not induce a higher CA level (Table 2). This result strongly suggests that the sinapyl alcohol accumulation in our transgenics is uncoupled from some genetic elements necessary for the acylation of sinapyl alcohol by coumaric acid. Furthermore, we detected significantly higher FA content in the mature dried stems of *Bradi3g30590*-OE when compared with WT (Table 2). This result might be related to the decrease of lignin content in the *Bradi3g30590*-OE. Indeed, increase of FA content (released mild alkaline hydrolysis) was often associated with less lignin content in mutants of Brachypodium (Bouvier D'Yvoire *et al.*, 2013; Wang *et al.*, 2015), because FA ester-linked to arabinoxylans are thus less involved in ether bonds with lignins (Ralph, 2010).

As a probable consequence of the decrease in lignin content, the saccharification yield of the extractive free mature stems increased on average by 31.6% in the *Bradi3g30590-OE* (Table 3). To conclude, despite the absence of an enzymatic assay or knockout analyses of *Bradi3g30590*, the enrichment of S units in lignin of *Bradi3g30590*-OE strongly suggests that this gene is a functional ferulate 5-hydroxylase.

Conclusion

By quantifying gene expression in nine major organs at different developmental stages, the Brachypodium gene expression atlas can provide a foundation to assess gene activity and function in Brachypodium and grass-related crop species. The true potential of the atlas, however, can be further unlocked by identifying modules of transcriptionally co-regulated (coexpressed) genes. To this end, we have included the atlas in the PlaNet database (www.gene2function.de), which contains interactive, user-friendly tools to mine the expression profiles, coexpression network, and gene modules of Brachypodium. The comparative features of PlaNet provide an easy means to identify conserved and duplicated modules, enabling efficient transfer of functional information between species and discovery of duplicated pathways, respectively.

We underscore the importance of the expression atlas and coexpression based tools by identifying the F5H ortholog involved in lignin synthesis in Brachypodium. Furthermore, increasing S units, decreasing lignin content, and making cell wall more susceptible to enzymatic digestion by overexpression of F5H presents an achievable goal in grass crops.

This article is protected by copyright. All rights reserved

Author Manuscript

Acknowledgements

We would like to thank Debbie Maizels (http://www.scientific-art.com) for the Brachypodium drawing shown in Fig. 1. We are thankful to Sebastien Antelme for taking care of plant cultures.

Author contributions

R.S., H.H., S. Persson. planned the expression atlas. R.S. and S.H-Y-K. collected the RNA samples. F.L., L.C. and C.L. performed lignin analysis. C.S. and O.B-C. produced transgenic plants. P.L. constructed the overexpressor vectors. F.M.G. designed the CDF file for the microarrays. N.V., N.P. and A.P. provided eFP browser instance. E.J. and D.R. provided additional expression data. F.M.G. constructed the CDF file. M.M., S. Proost and B.O.H. performed the bioinformatical analyses. M.M. and R.S. wrote the article with help from all authors.

References

Alejandro S, Lee Y, Tohge T, Sudre D, Osorio S, Park J, Bovet L, Lee Y, Geldner N, Fernie AR *et al.* 2012. *AtABCG29* is a monolignol transporter involved in lignin biosynthesis. *Current Biology* **22**: 1207–1212.

Allen JF, de Paula WBM, Puthiyaveetil S, Nield J. **2011**. A structural phylogenetic map for chloroplast photosynthesis. *Trends in Plant Science* **16**: 645–655.

Aoki Y, Okamura Y, Tadaka S, Kinoshita K, Obayashi T. **2015**. ATTED-II in 2016: a plant coexpression database towards lineage-specific coexpression. *Plant & Cell Physiology* **2**: pcv165.

Barabási A-L, Bonabeau E. **2003**. Scale-free networks. *Scientific American* **288**: 60–69.

Barabási A-L, Oltvai ZN. **2004**. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* **5**: 101–113.

Barros J, Serrani-Yarce JC, Chen F, Baxter D, Venables BJ, Dixon RA. **2016**. Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nature Plants* **2**: 16050.

Bassel GW, Lan H, Glaab E, Gibbs DJ, Gerjets T, Krasnogor N, Bonner AJ, Holdsworth MJ, Provart NJ. **2011**. Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. *Proceedings of the National Academy of Sciences, USA* **108**: 9709–9714.

Benjamini Y, Hochberg Y. **1995**. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*

This article is protected by copyright. All rights reserved

Author Manuscript

(Methodological) **57**: 289–300.

De Bodt S, Hollunder J, Nelissen H, Meulemeester N, Inzé D. **2012**. CORNET 2.0: integrating plant coexpression, protein-protein interactions, regulatory interactions, gene associations and functional annotations. *New Phytologist* **195**: 707–720.

Bouvier D'Yvoire M, Bouchabke-Coussa O, Voorend W, Antelme S, Cézard L, Legée F, Lebris P, Legay S, Whitehead C, McQueen-Mason SJ *et al.* 2013. Disrupting the cinnamyl alcohol dehydrogenase 1 gene (*BdCAD1*) leads to altered lignification and improved saccharification in *Brachypodium distachyon*. *Plant Journal* **73**: 496–508.

Bringmann M, Li E, Sampathkumar A, Kocabek T, Hauser M-T, Persson S. **2012**. *POM-POM2*/*cellulose synthase interacting1* is essential for the functional association of cellulose synthase and microtubules in Arabidopsis. *Plant Cell* **24**: 163–177.

Brkljacic J, Grotewold E, Scholl R, Mockler T, Garvin DF, Vain P, Brutnell T, Sibout R, Bevan M, Budak H *et al.* 2011. Brachypodium as a model for the grasses: today and the future. *Plant Physiology* **157**: 3–13.

Brown D, Wightman R, Zhang Z, Gomez LD, Atanassov I, Bukowski JP, Tryfona T, McQueen-Mason SJ, Dupree P, Turner S. **2011**. Arabidopsis genes *IRREGULAR XYLEM* (*IRX15*) and *IRX15L* encode DUF579-containing proteins that are essential for normal xylan deposition in the secondary cell wall. *Plant Journal* **66**: 401–413.

Brown DM, Zeef LAH, Ellis J, Goodacre R, Turner SR. **2005**. Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell* **17**: 2281–2295.

Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, Stone BA, Newbigin EJ, Bacic A, Fincher GB. **2006**. Cellulose synthase-like *CslF* genes mediate the synthesis of cell wall (1,3;1,4)-ß-d-glucans. *Science* **311**: 1940–1942.

Cass CL, Peraldi A, Dowd PF, Mottiar Y, Santoro N, Karlen SD, Bukhman Y V., Foster CE, Thrower N, Bruno LC *et al.* 2015. Effects of *PHENYLALANINE AMMONIA LYASE* (*PAL*) knockdown on cell wall composition, biomass digestibility, and biotic and abiotic stress responses in *Brachypodium*. *Journal of Experimental Botany* **66**: 4317–4335.

Conesa A, Götz S. **2008**. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* **2008**: 619832.

Davidson RM, Gowda M, Moghe G, Lin H, Vaillancourt B, Shiu SH, Jiang N, Robin Buell C. **2012**. Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *Plant Journal* **71**: 492–502.

This article is protected by copyright. All rights reserved

Author Manuscript

Dence CW. **1992**. The determination of lignin. In: Lin SY, Dence CW, eds. *Methods in lignin chemistry*. Berlin, Heidelberg, Germany: Springer-Verlag, 33–58 **[Author, please confirm amended reference is correct]**.

Draper J, Mur LA-J, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge a P. **2001**. *Brachypodium distachyon*. A new model system for functional genomics in grasses. *Plant Physiology* **127**: 1539–1555.

Entus R, Poling M, Herrmann KM. **2002**. Redox regulation of Arabidopsis 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase. *Plant Physiology* **129**: 1866–1871.

Fasoli M, Dal Santo S, Zenoni S, Tornielli GB, Farina L, Zamboni A, Porceddu A, Venturini L, Bicego M, Murino V *et al.* 2012. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *Plant Cell* **24**: 3489–3505.

Ficklin SP, Feltus FA. **2011**. Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. *Plant Physiology* **156**: 1244–1256.

Francoz E, Ranocha P, Nguyen-Kim H, Jamet E, Burlat V, Dunand C. **2015**. Roles of cell wall peroxidases in plant development. *Phytochemistry* **112**: 15–21.

Franke R, McMichael CM, Meyer K, Shirley AM, Cusumano JC, Chappie C. **2000**. Modified lignin in tobacco and poplar plants over-expressing the Arabidopsis gene encoding ferulate 5-hydroxylase. *Plant Journal* **22**: 223–234.

García JR, Anderson N, Le-Feuvre R, Iturra C, Elissetche J, Chapple C, Valenzuela S. **2014**. Rescue of syringyl lignin and sinapate ester biosynthesis in *Arabidopsis thaliana* by a *coniferaldehyde 5-hydroxylase* from *Eucalyptus globulus*. *Plant Cell Reports* **33**: 1263–1274.

Gerstein MB, Rozowsky J, Yan K-K, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ *et al.* 2014. Comparative analysis of the transcriptome across distant species. *Nature* **512**: 445–448.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N *et al.* 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40:** D1178-D1186.

Ha CM, Escamilla-Trevino L, Yarce JCS, Kim H, Ralph J, Chen F, Dixon RA. **2016**. An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in *Medicago truncatula*. *Plant Journal* **86**: 363–375.

Hammer O, Harper DAT, Ryan PD. 2001. *PAST: Paleontological Statistics software*

This article is protected by copyright. All rights reserved

Author Manuscript

package for education. URL http://palaeo-electronica.org/2001_1/past/issue1_01.htm. **[Author, please confirm inserted reference Hammer *et al.* (2001) is correct.]**

Hartley JL. **2006**. Cloning technologies for protein expression and purification. *Current Opinion in Biotechnology* **17**: 359–366.

Hartley RD, Ford CW. **1989**. Phenolic constituents of plant cell walls and wall biodegradability. In: Lewis NG, Paice MG, eds. *Plant cell wall polymers*. Washington DC, USA: American Chemical Society, 137–145 **[Author, please confirm the amended reference is correct]**.

Himmelbach A, Zierold U, Hensel G, Riechen J, Douchkov D, Schweizer P, Kumlehn J. **2007**. A set of modular binary vectors for transformation of cereals. *Plant Physiology* **145**: 1192–1200.

Hong S-YY, Seo PJ, Yang M-SS, Xiang F, Park C-MM. **2008**. Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR. *BMC Plant Biol* **8**: 112.

International Rice Genome Sequencing Project, Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L *et al.* 2005. The B73 maize genome: complexity, diversity, and dynamics. *Nature* **326**: 1112–1115.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. **2003**. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**: 249–264.

Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S, Finkers R *et al.* 2013. Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* **341**: 175–179.

Jiménez-Gómez JM, Wallace AD, Maloof JN. **2010**. Network analysis identifies ELF3 as a QTL for the shade avoidance response in Arabidopsis. *PLoS Genetics* **6:** e1001100.

Karlen SD, Zhang C, Peck ML, Smith RA, Padmakshan D, Helmich KE, Free HCA, Lee S, Smith BG, Lu F *et al.* 2016. Monolignol ferulate conjugates are naturally incorporated into plant lignins. *Science Advances* **2:** e1600393.

Langfelder P, Horvath S, Fisher R, Zhou X, Kao M, Wong W, Steffen M, Petti A, Aach J, D'haeseleer P *et al.* 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559.

Lapierre C, Monties B, Rolando C. **1985**. Thiloacidolysis of lignin: comparison with acidolysis. *Journal of wood Chemistry and Technology* **5**: 277–292.

This article is protected by copyright. All rights reserved

Author Manuscript

Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P. **2004**. Coexpression analysis of human genes across many microarray data sets. *Genome research* **14**: 1085–1094.

Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY. **2010**. Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*. *Nature Biotechnology* **28**: 149–156.

Madson M, Dunand C, Li X, Verma R, Vanzin GF, Caplan J, Shoue D a, Carpita NC, Reiter W-D. **2003**. The *MUR3* gene of Arabidopsis encodes a xyloglucan galactosyltransferase that is evolutionarily related to animal exostosins. *Plant Cell* **15**: 1662–1670.

Mayer KFX, Waugh R, Langridge P, Close TJ, Wise RP, Graner A, Matsumoto T, Sato K, Schulman A, Muehlbauer GJ *et al.* 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**: 711–716.

McFarlane HE, Döring A, Persson S. **2014**. The cell biology of cellulose synthesis. *Annual Review of Plant Biology* **65**: 69–94.

Méchin V, Laluc A, Legée F, Cézard L, Denoue D, Barrière Y, Lapierre C. **2014**. Impact of the brown-midrib *bm5* mutation on maize lignins. *Journal of Agricultural and Food Chemistry* **62**: 5102–5107.

Meyer K, Shirley AM, Cusumano JC, Bell-Lelong DA, Chapple C. **1998**. Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in Arabidopsis. *Proceedings of the National Academy of Sciences, USA* **95**: 6619–6623.

Movahedi S, Van de Peer Y, Vandepoele K. **2011**. Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in Arabidopsis and rice. *Plant physiology* **156**: 1316–1330.

Muller B, Fastner A, Karmann J, Mansch V, Hoffmann T, Schwab W, Suter-Grotemeyer M, Rentsch D, Truernit E, Ladwig F *et al.* 2015. Amino acid export in developing Arabidopsis seeds depends on UmamiT facilitators. *Current Biology* **25**: 3126–3131.

Mutwil M, Klie S, Tohge T, Giorgi FM, Wilkins O, Campbell MM, Fernie AR, Usadel B, Nikoloski Z, Persson S. **2011**. PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. *Plant Cell* **23**: 895–910.

Mutwil M, Usadel B, Schütte M, Loraine A, Ebenhöh O, Persson S. **2010**. Assembly of

This article is protected by copyright. All rights reserved

Author Manuscript

an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. *Plant Physiology* **152**: 29–43.

Nakano Y, Yamaguchi M, Endo H, Rejab NA, Ohtani M. **2015**. NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. *Frontiers in Plant Science* **6**: 288.

Obayashi T, Nishida K, Kasahara K, Kinoshita K. **2011**. ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. *Plant & Cell Physiology* **52**: 213–219.

Park CY, Wong AK, Greene CS, Rowland J, Guan Y, Bongo LA, Burdine RD, Troyanskaya OG. **2013**. Functional knowledge transfer for high-accuracy prediction of under-studied biological processes. *PLoS Computational Biology* **9**: e1002957.

Patel RV, Nahal HK, Breit R, Provart NJ. **2012**. BAR expressolog identification: expression profile similarity ranking of homologous genes in plant species. *Plant Journal* **71**: 1038–1050.

Persson S, Paredez A, Carroll A, Palsdottir H, Doblin M, Poindexter P, Khitrov N, Auer M, Somerville CR. **2007**. Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. *Proceedings of the National Academy of Sciences, USA* **104**: 15566–15571.

Persson S, Wei H, Milne J, Page GP, Somerville CR. **2005**. Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proceedings of the National Academy of Sciences, USA* **102**: 8633–8638.

Petrik DL, Karlen SD, Cass CL, Padmakshan D, Lu F, Liu S, Le Bris P, Antelme S, Santoro N, Wilkerson CG *et al.* 2014. p-Coumaroyl-CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in *Brachypodium distachyon*. *Plant Journal* **77**: 713–726.

Pickel B, Constantin M-A, Pfannstiel J, Conrad J, Beifuss U, Schaller A. **2010**. An enantiocomplementary dirigent protein for the enantioselective laccase-catalyzed oxidative coupling of phenols. *Angewandte Chemie International Edition* **49**: 202–204.

Priest HD, Fox SE, Rowley ER, Murray JR, Michael TP, Mockler TC. **2014**. Analysis of global gene expression in *Brachypodium distachyon* reveals extensive network plasticity in response to abiotic stress. *PLoS ONE* **9:** e87499.

Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, Vandepoele K. **2009**. PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* **21**: 3718–3731.

This article is protected by copyright. All rights reserved

Author Manuscript

Proost S, Mutwil M. **2016**. Tools of the trade: studying molecular networks in plants. *Current Opinion in Plant Biology* **30**: 130–140.

Proost S, Mutwil M. **2017**. PlaNet: comparative co-expression network analyses for plants. In: van Dijk ADJ, ed. *Plant genomics databases: methods and protocols*. New York, NY, USA: Springer New York, 213–227.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. **2005**. InterProScan: protein domains identifier. *Nucleic Acids Research* **33**. W116-W120.

Ralph J. **2010**. Hydroxycinnamates in lignification. *Phytochemistry Reviews* **9**: 65–83.

Ralph J, Lapierre C, Lu F, Marita JM, Pilate G, Van Doorsselaere J, Boeŕjan W, Jouanin L. **2001a**. NMR evidence for benzodioxane structures resulting from incorporation of 5-hydroxyconiferyl alcohol into lignins of O-methyltransferase-deficient poplars. *Journal of Agricultural and Food Chemistry* **49**: 86–91.

Ralph J, Lapierre C, Marita JM, Kim H, Lu F, Hatfield RD, Ralph S, Chapple C, Franke R, Hemm MR *et al.* 2001b. Elucidation of new structures in lignins of CAD- and COMT-deficient plants by NMR. *Phytochemistry* **57**: 993–1003.

Ruprecht C, Mendrinna A, Tohge T, Sampathkumar A, Klie S, Fernie AR, Nikoloski Z, Persson S, Mutwil M. **2016**. FamNet: a framework to identify multiplied modules driving pathway expansion in plants. *Plant Physiology* **170**: 1878–1894.

Ruprecht C, Mutwil M, Saxe F, Eder M, Nikoloski Z, Persson S. **2011**. Large-scale co-expression approach to dissect secondary cell wall formation across plant species. *Frontiers in Plant Science* **2**: 1–13.

Ruprecht C, Proost S, Hernandez-Coronado M, Ortiz-Ramirez C, Lang D, Rensing SA, Becker JD, Vandepoele K, Mutwil M. **2017a**. Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. *Plant Journal* **90:** 447–465.

Ruprecht C, Vaid N, Proost S, Persson S, Mutwil M. **2017b**. Beyond genomics: studying evolution with gene coexpression networks. *Trends in Plant Science* **22:** 298–307.

Sánchez-Rodríguez C, Bauer S, Hématy K, Saxe F, Ibáñez AB, Vodermaier V, Konlechner C, Sampathkumar A, Rüggeberg M, Aichinger E *et al.* 2012. *CHITINASE-LIKE1*/*POM-POM1* and its homolog *CTL2* are glucan-interacting proteins important for cellulose biosynthesis in Arabidopsis. *Plant Cell* **24**: 589–607.

Sanderson K. **2011**. Lignocellulose: a chewy problem. *Nature* **474**: S12–S14.

Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W. **2016**. Learning from co-

This article is protected by copyright. All rights reserved

Author Manuscript

expression networks: possibilities and challenges. *Frontiers in Plant Science* **7**: 444.

Sibout R, Le Bris P, Legée F, Cézard L, Renault H, Lapierre C. **2016**. Structural redesigning arabidopsis lignins into alkali-soluble lignins through the expression of p-coumaroyl-CoA:monolignol transferase *PMT*. *Plant Physiology* **170**: 1358–1366.

Stewart JJ, Akiyama T, Chapple C, Ralph J, Mansfield SD. **2009**. The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar[1]. *Plant Physiology* **150**: 621–635.

Stuart JM, Segal E, Koller D, Kim SK. **2003**. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**: 249–255.

Sundell D, Mannapperuma C, Netotea S, Delhomme N, Lin Y-C, Sjödin A, Van de Peer Y, Jansson S, Hvidsten TR, Street NR. **2015**. The Plant Genome Integrative Explorer Resource: PlantGenIE.org. *The New phytologist* **208**: 1149–1156.

Takabayashi A, Ishikawa N, Obayashi T, Ishida S, Obokata J, Endo T, Sato F. **2009**. Three novel subunits of Arabidopsis chloroplastic NAD(P)H dehydrogenase identified by bioinformatic and reverse genetic approaches. *Plant Journal* **57**: 207–219.

Takahashi N, Lammens T, Boudolf V, Maes S, Yoshizumi T, De Jaeger G, Witters E, Inzé D, De Veylder L. **2008**. The DNA replication checkpoint aids survival of plants deficient in the novel replisome factor ETG1. *The EMBO Journal* **27**: 1840–1851.

Tan L, Showalter A, Egelund J, Hernandez-Sanchez A, Doblin M, Bacic A. **2012**. Arabinogalactan-proteins and the research challenges for these enigmatic plant cell surface proteoglycans. *Frontiers in Plant Science* **3**: 140.

The International Wheat Genome Sequencing Consortium (IWGSC). **2014**. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**: 1251788.

Tzfadia O, Amar D, Bradbury LMT, Wurtzel ET, Shamir R. **2012**. The MORPH algorithm: ranking candidate genes for membership in Arabidopsis and tomato pathways. *Plant Cell* **24**: 4389–4406.

Tzfadia O, Diels T, De Meyer S, Vandepoele K, Aharoni A, Van de Peer Y. **2016**. CoExpNetViz: comparative co-expression networks construction and visualization tool. *Frontiers in Plant Science* **6**: 1194.

Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ. **2009**. Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant, Cell & Environment* **32**: 1633–1651.

This article is protected by copyright. All rights reserved

Author Manuscript

Vain T, Crowell EF, Timpano H, Biot E, Desprez T, Mansoori N, Trindade LM, Pagant S, Robert S, Hofte H *et al.* 2014. The cellulase *KORRIGAN* is part of the cellulose synthase complex. *Plant Physiol* **165**: 1521–1532.

Vanholme R, Morreel K, Darrah C, Oyarce P, Grabber JH, Ralph J, Boerjan W. **2012**. Metabolic engineering of novel lignin in biomass crops. *New Phytologist* **196**: 978–1000.

Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smith M, Lail K *et al.* 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**: 763–768.

Wang Y, Bouchabké-Coussa O, Le Bris P, Antelme S, Soulhat C, Gineau E, Dalmais M, Bendahmane A, Morin H, Mouille G *et al.* 2015. *LACCASE 5* is required for lignification of the *Brachypodium distachyon* culm. *Plant Physiology* **168**: 192–204.

Wang Y, Chantreau M, Sibout R, Hawkins S. **2013**. Plant cell wall lignification and monolignol metabolism. *Frontiers in Plant Science* **4**: 220.

Weng J-K, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C. **2010**. Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte Selaginella and flowering plants. *Plant Cell* **22**: 1033–1045.

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. **2007**. An 'electronic fluorescent pictograph' Browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* **2:** e718**.**

Withers S, Lu F, Kim H, Zhu Y, Ralph J, Wilkerson CG. **2012**. Identification of grass-specific enzyme that acylates monolignols with p-coumarate. *The Journal of Biological Chemistry* **287**: 8347–8355.

Xiao SJ, Zhang C, Zou Q, Ji ZL. **2010**. TiSGeD: a database for tissue-specific genes. *Bioinformatics* **26**: 1273–1275.

Yu H, Luscombe NM, Qian J, Gerstein M. **2003**. Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends in Genetics* **19**: 422–427.

Zarrineh P, Sánchez-Rodríguez A, Hosseinkhan N, Narimani Z, Marchal K, Masoudi-Nejad A. **2014**. Genome-scale co-expression network comparison across *Escherichia coli* and *Salmonella enterica* serovar Typhimurium reveals significant conservation at the regulon level of local regulators despite their dissimilar lifestyles. *PLoS ONE* **9**: e102871.

Zhang B, Horvath S. **2005**. A General Framework For Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology* **4**:

This article is protected by copyright. All rights reserved

Author Manuscript

Article17.

Zhou J, Lee C, Zhong R, Ye Z-H. **2009**. *MYB58* and *MYB63* are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. *Plant Cell* **21**: 248–266.

Author Manuscript

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Sample relationship dendrogram of the 117 microarrays.

Fig. S2 Sample relationship between transcriptomic profiles of the sampled organs, at different developmental stages.

Fig. S3 Distribution of the maximum SPM values from the observed and randomized expression profiles.

Fig. S4 Coexpression network of *Bradi4g34420*, *Bradi1g76630*, *Bradi2g16570*, *Bradi2g23270* and *Bradi2g46120*.

Fig. S5 Expression of Brachypodium F5H candidate genes.

Fig. S6 Co-expression network of *Bradi3g30590* (large green square).

Table S1 Normalized expression data

Table S2 HRR-based coexpression network of Brachypodium

Table S3 Highest SPM values for genes

Table S4 MapMan enrichment analysis of preferentially expressed transcripts in the samples

Table S5 MapMan enrichment analysis of HCCA clusters

Table S6 Phytozome BLAST output of *AT4G36220* (*F5H*) compared with the Brachypodium proteome

Methods S1 The protocol used to perform Klason lignin analysis, determination of lignin monomer levels by thioacidolysis, and determination of ferulic acid and *p*-coumaric acid levels.

Notes S1 CDF file describing probesets on the Brachypodium microarray.

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

This article is protected by copyright. All rights reserved

Fig. 1 eFP browser of the sampled organs of *Brachypodium distachyon*. Expression values in the sampled organs are indicated by a colour gradient, where yellow indicates no detectable expression, while red signifies highest expression. The age of sampled organs is given as days after germination (DAG), days after fertilization (DAF), days after heading (DAH), or years. The legend describing the colour gradient and expression values is shown in the bottom right corner. Photosystem I reaction center subunit N (PsaN) gene *Bradi4g40780* is used as an example. P, growth chamber in Versailles; V, glasshouse in Versailles **[Author, please confirm inserted text 'P, growth chamber in Versailles; V,** glasshouse **in Versailles' is correct]**.

Fig. 2 PlaNet instance of Brachypodium. (a) Power-law behaviour of Brachypodium Pearson correlation coefficient (PCC) coexpression network. The network was generated by setting threshold of PCC > 0.7. The plot shows the relationship between the number of coexpression edges a given gene has (node degree, *x*-axis) vs the number of genes with the given node degree (node frequency, *y*-axis). Note that both axes have been $\log_{10}$-transformed. (b) Expression profile of *Bradi4g40780*, where *x*-axis and *y*-axis indicate organ samples and expression values, respectively. The green points indicate expression values from the individual microarrays, while the red points show average expression. For brevity, expression data from only five sampled organs are shown. P, growth chamber in Versailles; V, glasshouse in Versailles **[Author, please confirm inserted text 'P, growth chamber in Versailles; V,** glasshouse **in Versailles' is correct]**. (c) Coexpression neighbourhood of PsaN (*Bradi4g40780*, large light green circle). Nodes represent genes, while edges connect coexpressed genes. The edge colour shows how strongly two genes are coexpressed, with green, orange and red edges indicating strong (HRR < 10, highest reciprocal rank), medium (HRR < 20) and weak (HRR < 50) coexpression, respectively. Coloured shapes represent label co-occurrences, which are used to bin genes that contain same Pfam domains and gene families. For example, genes *Bradi1g24760* and *Bradi2g61500* (beige circles) belong to the same gene family (HOM086919). The annotation of label co-occurrences is shown to the right of the network. Functional annotation of PLAZA2.5 gene families can be found at http://bioinformatics.psb.ugent.be/plaza/versions/plaza_v2_5/, while the description of pfam domains can be found at http://pfam.xfam.org/. For simplicity, only one representative pfam domain (e.g. PsaD) or gene family (e.g. HOM001413) per label co-occurrence is shown.

This article is protected by copyright. All rights reserved

Author Manuscript

Fig. 3 MapMan enrichment analysis of coexpression clusters. (a) Rows represent MapMan terms, while columns represent heuristic cluster chiseling algorithm (HCCA) clusters. Clusters that are enriched (false discovery rate adjusted *P*-value < 0.05) for a given MapMan term are indicated by red boxes. For brevity, only 35 MapMan terms and 15 clusters are shown. Cluster 48 is highlighted by the blue rectangle. (b) Coexpression network of cell wall cluster 48. Similarly to Fig. 1(c), nodes represent genes, edges connect coexpressed genes, while coloured shapes indicate label co-occurrences. For brevity, identifiers of genes not involved in cellulose and lignin biosynthesis gene are hidden.

Fig. 4 Duplicated cell wall modules in Brachypodium. (a) Gene module network of *Bradi2g16560* (large node). Nodes represent gene modules, while line colours indicate duplicated modules (green lines), and overlapping modules (orange lines). Line styles depict the degree of similarity of the module pairs, as shown in the legend. **[Author, please confirm amended text 'lines' are correct.]** The similarity is represented by label co-occurrences, which indicates how many gene families and Pfam domains the modules have in common. Black boxes indicate gene modules selected for further analysis. (b) Representative expression profiles of the genes selected for the comparative analysis. White and blue colours indicate low and high expression, respectively. (c) Gene contents of the selected modules. Abbreviated description of the label co-occurrences is shown below.

Fig. 5 Lignin modules in Arabidopsis, Brachypodium and rice. (a) Gene module network of *Bradi3g49260* (large, central node). Blue lines connect conserved modules, orange lines indicate overlapping modules, while black boxes indicate gene modules selected for further analysis. **[Author, please confirm amended text 'lines' are correct.]** (b) Lignin biosynthesis pathway outlining the different steps of coniferyl and sinapyl alcohol biosynthesis. The coloured shapes are used to indicate the enzymes found in (c, d). (c, d) Gene contents of the selected (c) Arabidopsis and (d) Brachypodium gene modules. For brevity, only one representative pfam domain is shown per label co-occurrence.

Fig. 6 Analysis of *Bradi3g30590*-OE lines. (a) Heatmap showing expression profiles of *BdF5H* and *BdPAL1*, where dark blue colour indicates high expression. (b) Expression levels of *Bradi3g30590* cDNA in leaves of transgenic lines and wild-type (WT) plants. The data represent means and SD from four biological replicates. (c) Cross-section of WT stem. Orange

This article is protected by copyright. All rights reserved

Author Manuscript

and red colours indicate G and S units stained by Maüle method. Vascular bundles are indicated by arrows. (d) Cross-section of *BdF5H-OE* stem.

Author Manuscript

Table 1 Description of sampled organs and developmental stages of *Brachypodium distachyon*

Organ	No. of preferentially expressed genes	Developmental stage/part/treatment	Age, growth condition	No. of stage-specific expressed genes
Root	1141	Root	10 DAG (P); 35 DAG (V)	303; 838
Shoot	610	Etiolated	3 DAG (I)	326
		De-etiolated	3 DAG (I)	284
Node	787	First node	10 DAG (P)*; 17 DAG (P); 27 DAG (P)*; 60 DAG (P)*	123; 15; 7; 21
		First node + adventitious roots	35 DAG (P)	198
		Last node	35 DAG (V)	370
		Lower part of inclined node	42 DAG (V)*	41
		Upper part of inclined node	42 DAG (V)*	12
Internode	1176	First internode	10 DAG (P); 17 DAG (P)*; 27 DAG (P); 35 DAG (P)*; DAG 60 (T)*	33; 17; 8; 208; 295; 36
		Second internode	17 DAG (P); 27 DAG (P)*	19; 20
		Last internode	35 DAG (V)*; 60 DAG (P)*; 60 DAG (T)	237; 34; 269
Leaf	1302	Leaf	10 DAG (P)*; 17 DAG (P); 27 DAG (P); 60 DAG (P)	59; 13; 37; 37

This article is protected by copyright. All rights reserved

		Young leaf, <6 cm	60 DAG (T)	551
		Mature leaf, fully expanded	60 DAG (T)	605
Peduncle	268	First 2 cm	42 DAG (V)*	73
		Second 2 cm	42 DAG (V)*	73
		Third 2 cm	42 DAG (V)	94
		Last 2 cm	42 DAG (V)	18
		Spikelet pedicel	42 DAG (V)	10
Spikelet	381	Young spikelet	3 DAH (P)*	297
		First spikelet internode	42 DAG (V)*	51
		Last spikelet internode	42 DAG (V)	33
Seed	3966	Endosperm	11 DAF (V); 31 DAF (V)**	527; 840
		Whole grain	11 DAF (V); 31 DAF (V); 2 yr (V)	1561; 611; 427
Coleoptile	283	Coleoptile	10 DAG(P); 17 + 27 DAG(P)	22; 261

The second column indicates how many genes are preferentially expressed in a given organ. The third column describes the part of the organ that was sampled, while the fourth column gives the age of the organ, as days after germination (DAG), days after fertilization (DAF), days after heading (DAH) or years. Growth facilities are indicated: P, growth chamber in Versailles; V, glasshouse in Versailles; I, *in vitro* culture in Versailles; T, growth chamber in Toulouse. The fifth column indicates how many genes are preferentially expressed in a given sample. For example, 1036 and 1749 genes are preferentially transcribed in endosperm 11 DAF and 31 DAF, respectively. Samples marked with * and ** were taken in duplicates or as single sample, respectively. Otherwise, samples were taken in triplicates.

Table 2 Analyses of lignins, *p*-OH cinnamic acid (*p*-coumaric acid (CA) and ferulic acid (FA) ester-linked to the cell walls) and glucose yield in the mature stems of Brachypodium samples

Genotype	Lignin content	Relative frequency of lignin-derived thioacidolysis monomers	Ester-linked *p*-OH cinnamic acids

This article is protected by copyright. All rights reserved

Author Manuscript

	(%)	%H	%G	%S	%5-OHG	CA (mg g^{-1})	FA (mg g^{-1})
Culture 1: Mature dried plants (100 DAG)							
WT	19.57 ± 0.30	3.1 ± 0.2	29.3 ± 1.2	66.9 ± 1.1	0.8 ± 0.0	8.88 ± 0.41	5.25 ± 0.48
Bradi3g30590-OE line 1	16.48 ± 0.84[b]	4.1 ± 0.3[b]	10.4 ± 0.8[b]	84.4 ± 1.0[b]	1.1 ± 0.0[b]	8.18 ± 0.35	7.19 ± 0.30[b]
Bradi3g30590-OE line 2	16.10 ± 0.82[b]	4.2 ± 0.2[b]	11.2 ± 1.1[b]	83.6 ± 1.2[b]	1.1 ± 0.0[b]	8.60 ± 0.54	6.96 ± 0.48[b]
Culture 2: Green plants before flowering (20 DAG)							
WT	15.15 ± 0.23	4.4 ± 0.2	33.6 ± 0.9	61.4 ± 0.8	0.6 ± 0.0	12.04 ± 0.57	5.68 ± 0.18
Bradi3g30590-OE line 1	13.73 ± 0.19[b]	5.7 ± 0.2 [b]	11.8 ± 0.4 [b]	81.5 ± 0.5 [b]	0.9 ± 0.1[b]	12.49 ± 0.39	5.86 ± 0.04
Bradi3g30590-OE line 2	13.74 ± 0.47[b]	5.5 ± 0.2[b]	12.3 ± 0.4[b]	81.2 ± 0.4[b]	1.1 ± 0.0[b]	12.01 ± 0.23	5.47 ± 0.23

The lignin content and structure are evaluated by the Klason and thioacidolysis methods. The CA or FA levels ester-linked to the cell walls are measured by mild alkaline hydrolysis. Values are means ± SD from individually analyzed plants (n = 3 or 4). [a]All the plants were obtained from the same culture experiment in the glasshouse. [b]Significantly different from the corresponding wild type (WT) sample (one way ANOVA, Tukey's HSD) at $P < 0.05$. H, *p*-hydroxyphenyl; G, guaiacyl; S, syringyl; 5-OH-G, 5-hydroxy-guaiacyl; DAG, days after germination.

Table 3 Saccharification efficiency of extractive-free mature stems from *Brachypodium distachyon*

Genotype	Glucose yield (mg g^{-1})
WT	63.9 ± 2.9
BdF5H-OE line1	80.1 ± 4.2[a]
BdF5H-OE line2	83.9 ± 3.2[a]

Saccharification was evaluated by the glucose release induced by hydrolysis with a commercial cellulase preparation and without any pretreatment. Values are means ± SD from individually analysed plants (n = 3). All

This article is protected by copyright. All rights reserved

Author Manuscript

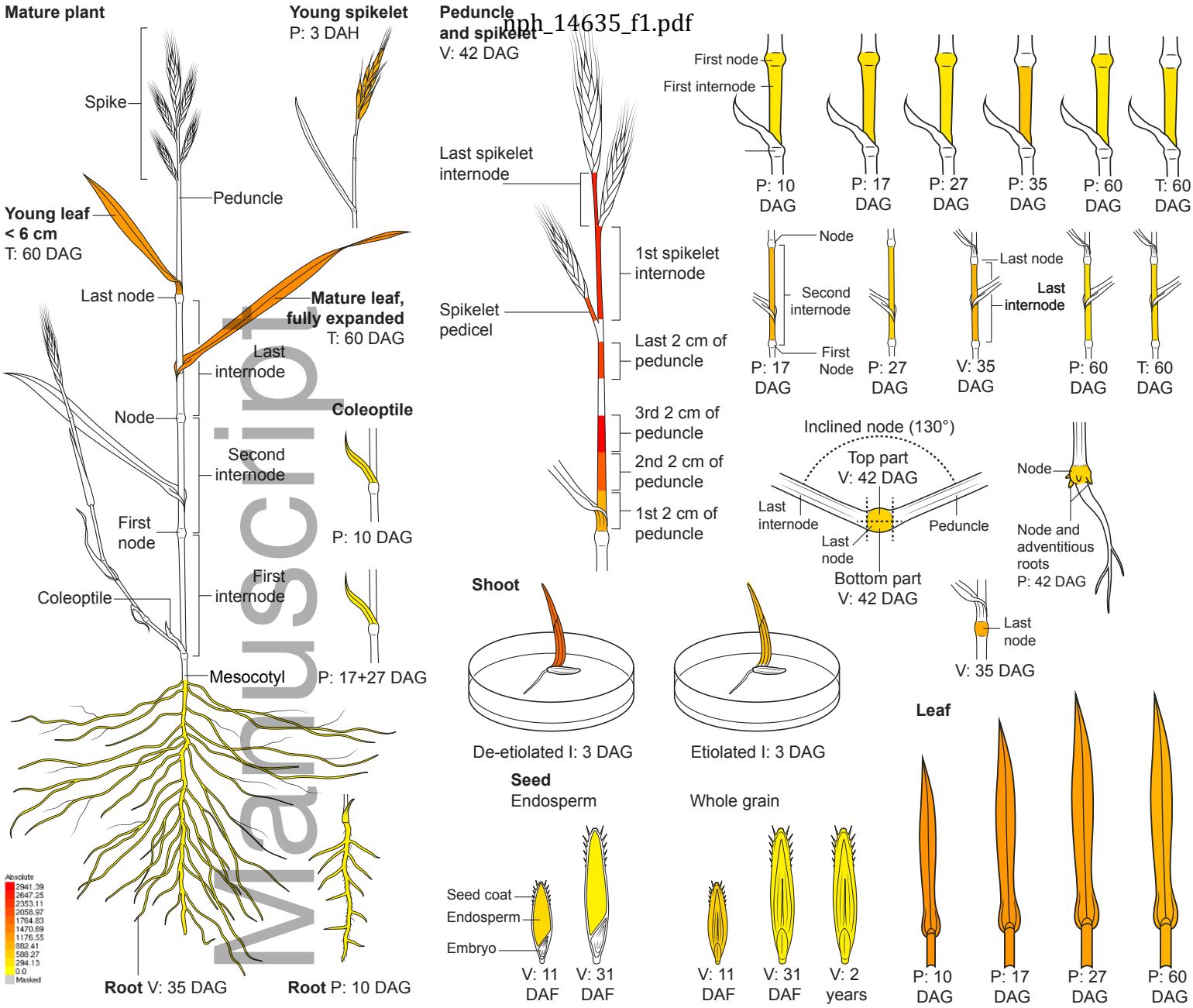
the plants were obtained from culture 1. [a]Significantly different from the corresponding wild-type (WT) sample (one way ANOVA, Tukey's HSD) at $P < 0.05$.

Author Manuscript

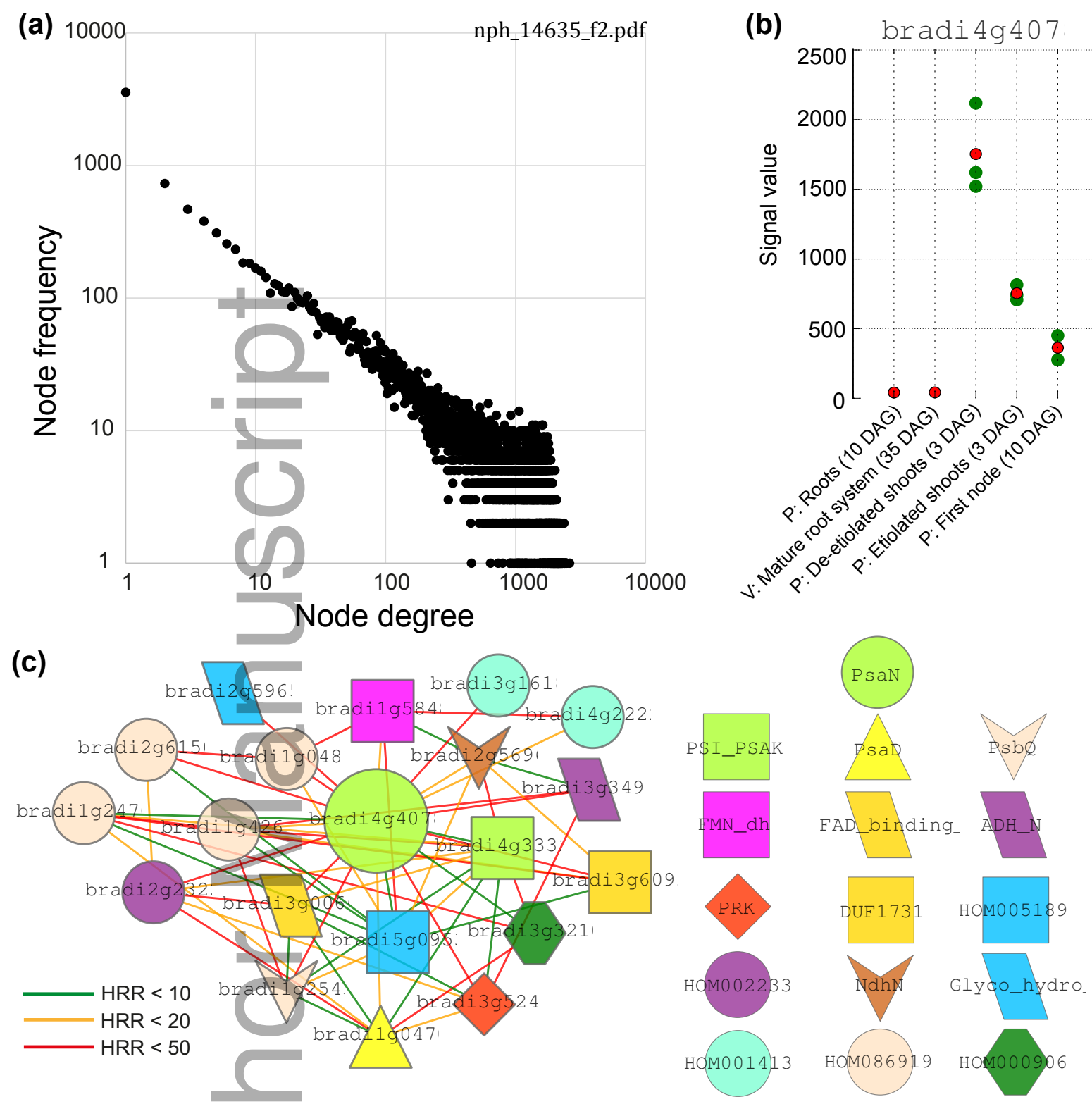
This article is protected by copyright. All rights reserved

Author Manuscript

nph_14635_f1.pdf

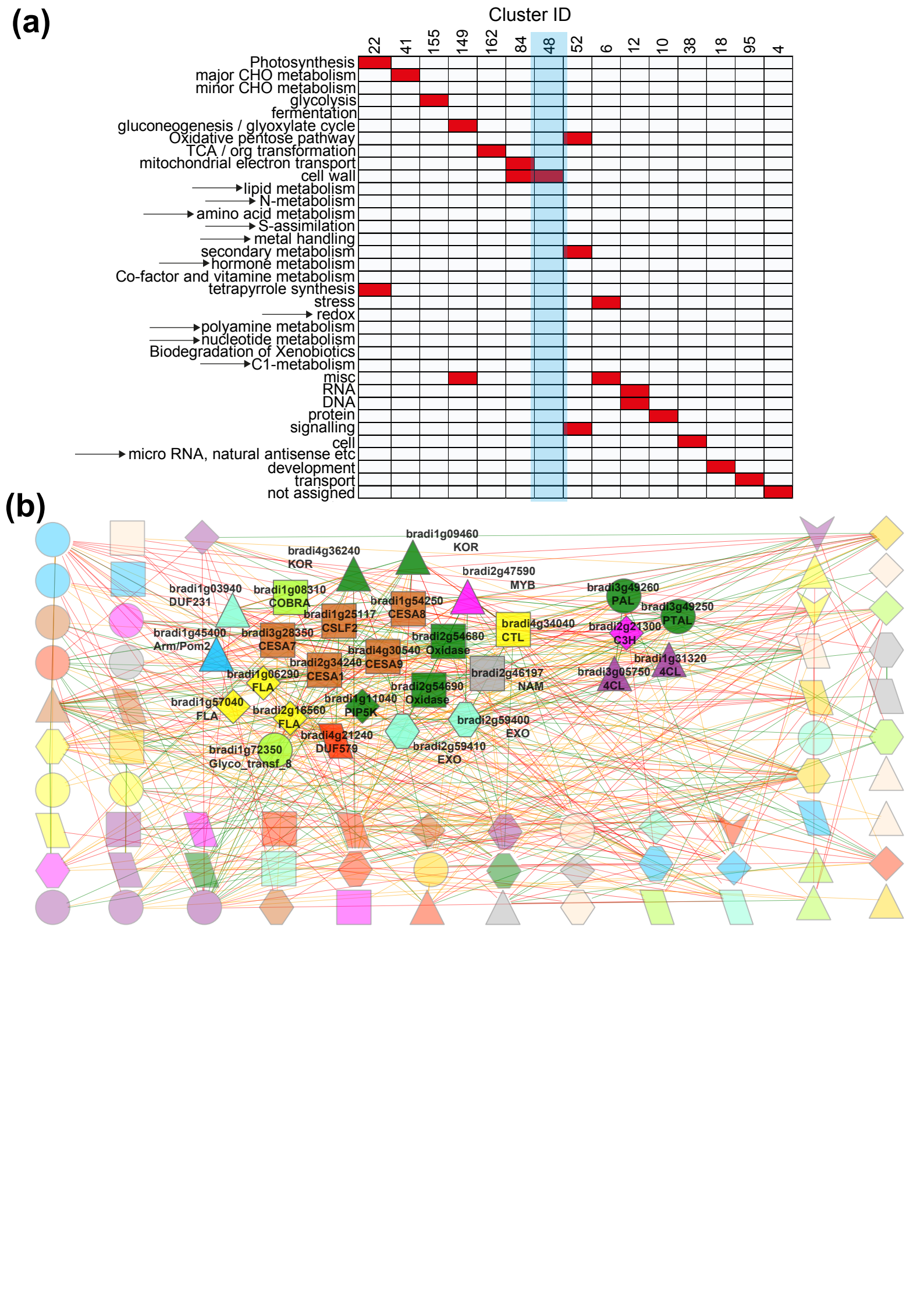


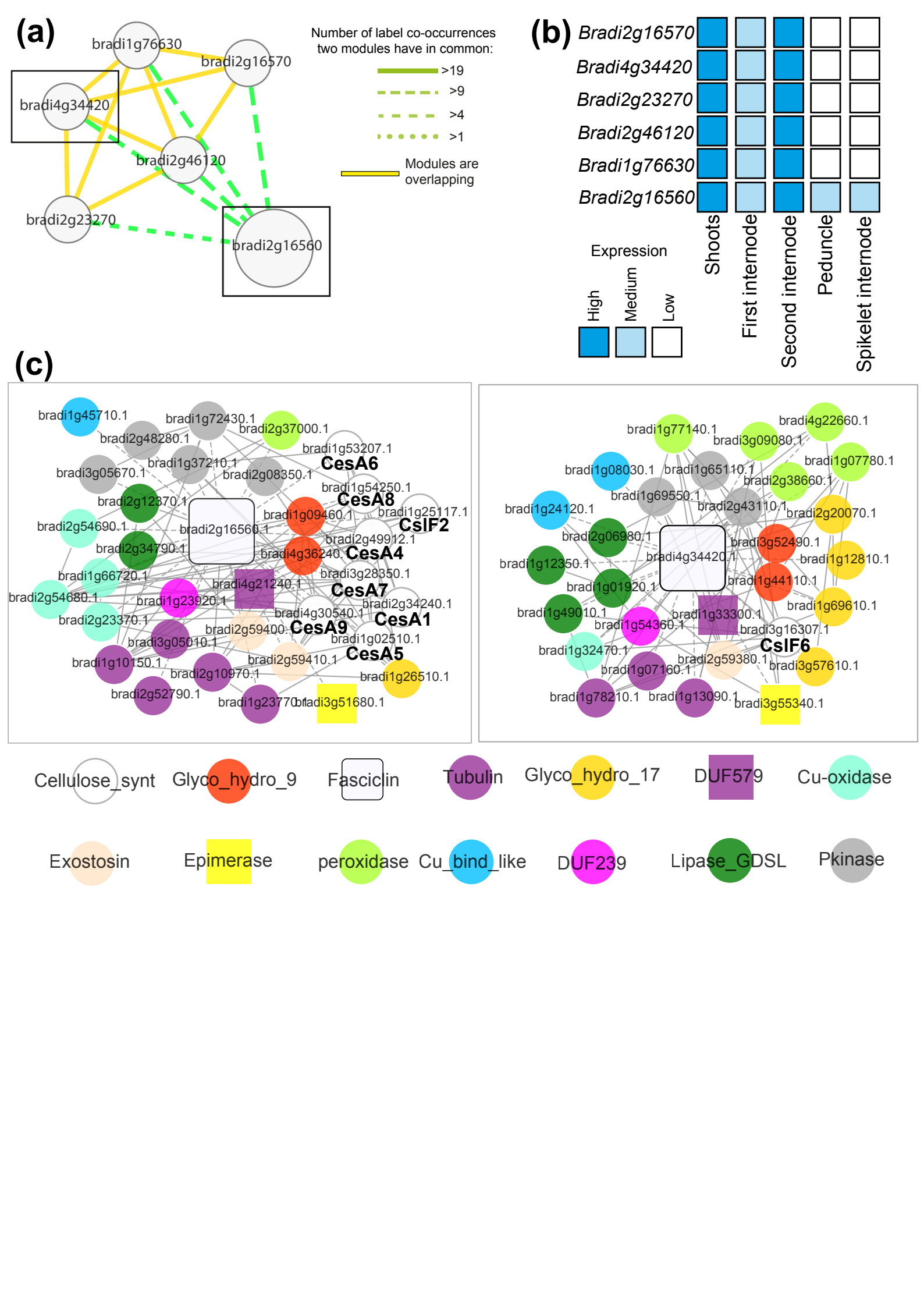
This article is protected by copyright. All rights reserved

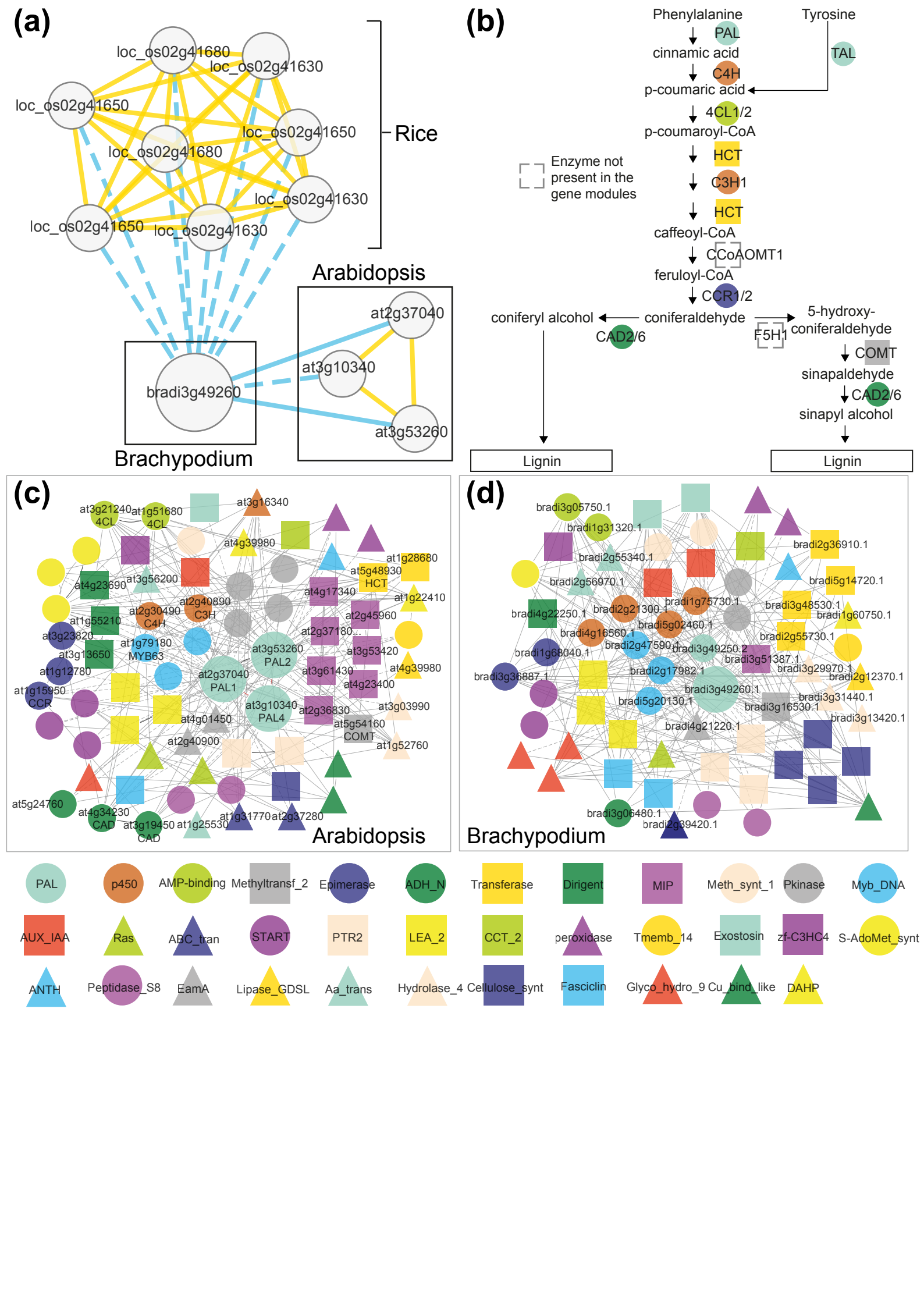


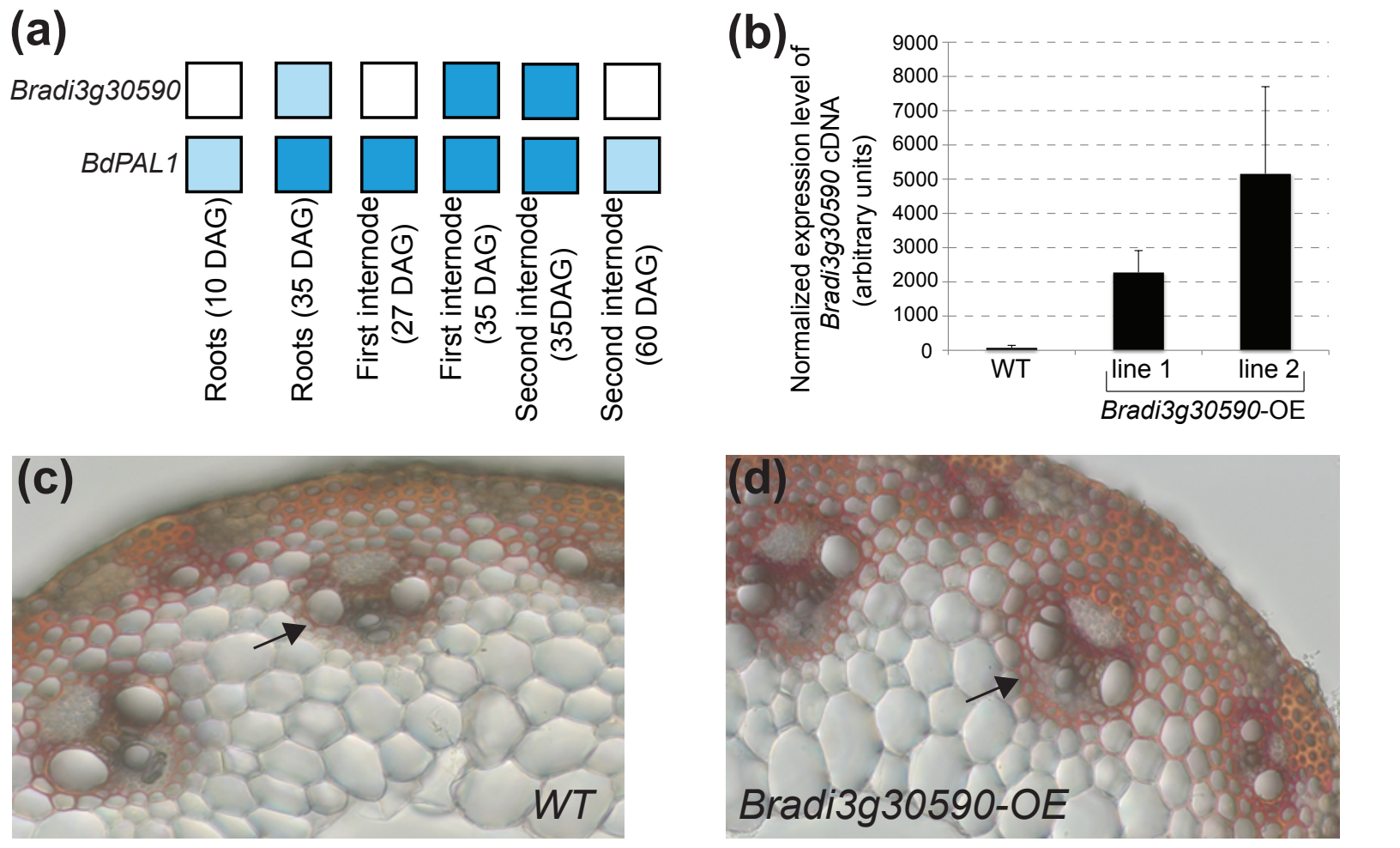
Author Manuscript

This article is protected by copyright. All rights reserved


(a)
Cluster ID
22
41
155
149
162
84
48
52
6
12
10
38
18
95
4
Photosynthesis
major CHO metabolism
minor CHO metabolism
glycolysis
fermentation
gluconeogenesis / glyoxylate cycle
Oxidative pentose pathway
TCA / org transformation
mitochondrial electron transport
cell wall
lipid metabolism
N-metabolism
amino acid metabolism
S-assimilation
metal handling
secondary metabolism
hormone metabolism
Co-factor and vitamine metabolism
tetrapyrrole synthesis
stress
redox
polyamine metabolism
nucleotide metabolism
Biodegradation of Xenobiotics
C1-metabolism
misc
RNA
DNA
protein
signalling
cell
micro RNA, natural antisense etc
development
transport
not assigned
(b)
bradi1g09460 KOR
bradi4g36240 KOR
bradi2g47590 MYB
bradi1g03940 DUF231
bradi1g08310 COBRA
bradi1g54250 CESA8
bradi1g25117 CSLF2
bradi3g49260 PAL
bradi3g49250 PTAL
bradi4g34040 CTL
bradi2g21300 C3H
bradi1g45400 Arm/Pom2
bradi3g28350 CESA7
bradi2g54680 Oxidase
bradi4g30540 CESA9
bradi1g31320 4CL
bradi2g34240 CESA1
bradi2g46197 NAM
bradi3g05750 4CL
bradi1g06290 FLA
bradi2g54690 Oxidase
bradi1g57040 FLA
bradi1g11040 PIP5K
bradi2g16560 FLA
bradi2g59400 EXO
bradi4g21240 DUF579
bradi1g72350 Glyco_transf_8
bradi2g59410 EXO


(a)
bradi1g76630
bradi2g16570
bradi4g34420
bradi2g46120
bradi2g23270
bradi2g16560
Number of label co-occurrences two modules have in common:
>19
>9
>4
>1
Modules are overlapping
(b)
Bradi2g16570
Bradi4g34420
Bradi2g23270
Bradi2g46120
Bradi1g76630
Bradi2g16560
Expression
High
Medium
Low
Shoots
First internode
Second internode
Peduncle
Spikelet internode
(c)
bradi1g45710.1
bradi1g72430.1
bradi2g37000.1
bradi2g48280.1
bradi1g53207.1
bradi1g37210.1
CesA6
bradi3g05670.1
bradi2g08350.1
bradi1g54250.1
bradi2g12370.1
CesA8
bradi1g09460.1
bradi1g25117.1
bradi2g54690.1
bradi2g16560.1
CslF2
bradi2g49912.1
bradi2g34790.1
bradi4g36240.1
CesA4
bradi1g66720.1
bradi3g28350.1
bradi2g54680.1
bradi4g21240.1
CesA7
bradi1g23920.1
bradi2g34240.1
bradi2g23370.1
bradi4g30540.1
CesA1
bradi2g59400.1
CesA9
bradi1g02510.1
bradi3g05010.1
CesA5
bradi1g10150.1
bradi2g59410.1
bradi2g10970.1
bradi1g26510.1
bradi2g52790.1
bradi1g23770.1
bradi3g51680.1
bradi1g77140.1
bradi4g22660.1
bradi3g09080.1
bradi1g07780.1
bradi1g08030.1
bradi1g65110.1
bradi2g38660.1
bradi1g69550.1
bradi1g24120.1
bradi2g43110.1
bradi2g20070.1
bradi2g06980.1
bradi4g34420.1
bradi3g52490.1
bradi1g12350.1
bradi1g12810.1
bradi1g01920.1
bradi1g44110.1
bradi1g49010.1
bradi1g33300.1
bradi1g69610.1
bradi1g54360.1
bradi3g16307.1
CslF6
bradi1g32470.1
bradi2g59380.1
bradi1g07160.1
bradi3g57610.1
bradi1g78210.1
bradi1g13090.1
bradi3g55340.1
Cellulose_synt
Glyco_hydro_9
Fasciclin
Tubulin
Glyco_hydro_17
DUF579
Cu-oxidase
Exostosin
Epimerase
peroxidase
Cu_bind_like
DUF239
Lipase_GDSL
Pkinase


(a)
loc_os02g41680
loc_os02g41630
loc_os02g41650
loc_os02g41650
loc_os02g41680
loc_os02g41630
loc_os02g41650
loc_os02g41630
Rice
Arabidopsis
at2g37040
at3g10340
at3g53260
bradi3g49260
Brachypodium
(b)
Phenylalanine
PAL
Tyrosine
TAL
cinnamic acid
C4H
p-coumaric acid
4CL1/2
p-coumaroyl-CoA
HCT
C3H1
HCT
caffeoyl-CoA
CCoAOMT1
feruloyl-CoA
CCR1/2
Enzyme not present in the gene modules
coniferyl alcohol
CAD2/6
coniferaldehyde
F5H1
5-hydroxy-coniferaldehyde
COMT
sinapaldehyde
CAD2/6
sinapyl alcohol
Lignin
Lignin
(c)
at3g21240 4CL
at1g51680 4CL
at3g16340
at4g39980
at1g28680
at5g48930 HCT
at4g23690
at3g56200
at4g17340
at1g22410
at2g40890 C3H
at2g30490 C4H
at2g45960
at1g55210
at2g37180...
at3g23820
at1g79180 MYB63
at3g53260 PAL2
at3g53420
at3g13650
at4g39980
at1g12780
at2g37040 PAL1
at3g61430
at4g23400
at1g15950 CCR
at3g10340 PAL4
at2g36830
at3g03990
at4g01450
at5g54160 COMT
at2g40900
at1g52760
at5g24760
at4g34230 CAD
at3g19450 CAD
at1g25530
at1g31770
at2g37280
Arabidopsis
(d)
bradi3g05750.1
bradi1g31320.1
bradi2g36910.1
bradi2g55340.1
bradi2g56970.1
bradi5g14720.1
bradi1g75730.1
bradi4g22250.1
bradi2g21300.1
bradi3g48530.1
bradi1g60750.1
bradi4g16560.1
bradi5g02460.1
bradi2g55730.1
bradi2g47590.1
bradi3g49250.2
bradi1g68040.1
bradi3g51387.1
bradi3g36887.1
bradi2g17982.1
bradi3g29970.1
bradi2g12370.1
bradi3g49260.1
bradi5g20130.1
bradi3g31440.1
bradi3g16530.1
bradi3g13420.1
bradi4g21220.1
bradi3g06480.1
bradi2g39420.1
Brachypodium
PAL
p450
AMP-binding
Methyltransf_2
Epimerase
ADH_N
Transferase
Dirigent
MIP
Meth_synt_1
Pkinase
Myb_DNA
AUX_IAA
Ras
ABC_tran
START
PTR2
LEA_2
CCT_2
peroxidase
Tmemb_14
Exostosin
zf-C3HC4
S-AdoMet_synt
ANTH
Peptidase_S8
EamA
Lipase_GDSL
Aa_trans
Hydrolase_4
Cellulose_synt
Fasciclin
Glyco_hydro_9
Cu_bind_like
DAHP


(a)
Bradi3g30590
BdPAL1
Roots (10 DAG)
Roots (35 DAG)
First internode (27 DAG)
First internode (35 DAG)
Second internode (35DAG)
Second internode (60 DAG)
(b)
Normalized expression level of Bradi3g30590 cDNA (arbitrary units)
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
WT
line 1
line 2
Bradi3g30590-OE

(c)
WT

(d)
Bradi3g30590-OE