Title

Transcriptomic analysis suggests a key role for
SQUAMOSA PROMOTER BINDING PROTEIN LIKE
,
NAC
and
YUCCA
genes in the heteroblastic development of the temperate rainfor...

Authors

Ostria-Gallardo, Enrique
Ranjan, Aashish
Chitwood, Daniel H
et al.

Publication Date

2015-12-01

DOI

10.1111/nph.13776

Peer reviewed

New Phytologist

Research

Transcriptomic analysis suggests a key role for *SQUAMOSA PROMOTER BINDING PROTEIN LIKE*, *NAC* and *YUCCA* genes in the heteroblastic development of the temperate rainforest tree *Gevuina avellana* (Proteaceae)

Enrique Ostria-Gallardo[1]*, Aashish Ranjan[2,3]*, Kristina Zumstein[2], Daniel H. Chitwood[4], Ravi Kumar[5], Brad T. Townsley[2], Yasunori Ichihashi[6], Luis J. Corcuera[1] and Neelima R. Sinha[2]

[1]Departamento de Botánica, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, 4030000 Biobío, Chile; [2]Department of Plant Biology, University of California, Davis, CA 95616, USA; [3]National Institute of Plant Genome Research, New Delhi 110067, India; [4]Donald Danforth Plant Science Center, St Louis, MO 63132, USA; [5]Novozymes, Davis, CA 95618, USA; [6]RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 230-0045, Japan

Author for correspondence:
Neelima R. Sinha
Tel: +1 530 754 8441
Email: nrsinha@ucdavis.edu

Received: *27 July 2015*
Accepted: *19 October 2015*

New Phytologist (2015)
doi: 10.1111/nph.13776

Key words: *Gevuina avellana*, heteroblasty, light availability, RNA-seq, temperate rainforest, transcriptional regulation.

Summary

• Heteroblasty, the temporal development of the meristem, can produce diverse leaf shapes within a plant. *Gevuina avellana*, a tree from the South American temperate rainforest shows strong heteroblasty affecting leaf shape, transitioning from juvenile simple leaves to highly pinnate adult leaves. Light availability within the forest canopy also modulates its leaf size and complexity. Here we studied how the interaction between the light environment and the heteroblastic progression of leaves is coordinated in this species.

• We used RNA-seq on the Illumina platform to compare the range of transcriptional responses in leaf primordia of *G. avellana* at different heteroblastic stages and growing under different light environments.

• We found a steady up-regulation of *SQUAMOSA PROMOTER BINDING PROTEIN LIKE* (*SPL*), *NAC*, *YUCCA* and *AGAMOUS-LIKE* genes associated with increases in age, leaf complexity, and light availability. In contrast, expression of *TCP*, *TPR* and *KNOTTED1* homeobox genes showed a sustained down-regulation. Additionally, genes involved in auxin synthesis/transport and jasmonate activity were differentially expressed, indicating an active regulation of processes controlled by these hormones.

• Our large-scale transcriptional analysis of the leaf primordia of *G. avellana* sheds light on the integration of internal and external cues during heteroblastic development in this species.

Introduction

Leaf shape is a conspicuous trait among angiosperms, exhibiting tremendous diversity and variability at all taxonomical levels (Nicotra *et al.*, 2011). One of the most fascinating mechanisms contributing to the diversification of leaf form, both within a single plant and between species, is heteroblasty. Heteroblasty refers to the temporal development of the shoot apical meristem that potentially affects all attributes of lateral organs, including their morphology. Since Goebel (1900), plant scientists have continued to describe gradual – or sometimes abrupt – changes in leaf size and shape at successive nodes that result from heteroblastic progression (Jones, 1999; Zotz *et al.*, 2011). In addition, recent studies show that the phenotypic expression of heteroblastic trajectories is highly contextual and responds in a plastic manner to prevailing environmental conditions (Jones, 1999; Burns, 2005; Gamage, 2011; Chitwood *et al.*, 2014). Because of their sessile life, plants anticipate forthcoming environmental conditions in their growing organs through a suite of mechanisms and initiate appropriate developmental responses (Diggle, 1997; Casal *et al.*, 2004). In this sense, it has been proposed that heteroblasty provides adaptive advantages to predictable changes in the environment during the lifetime of an individual plant (Day, 1998; Winn, 1999; Gamage, 2010). For example, in *Acacia* trees, the heteroblastic progression transitions from compound leaves to simple phyllodes. Interaction between the morphological/physiological stage in the trajectory from compound to simple leaves and prevailing environmental stresses (shade and then drought) faced throughout the development of an *Acacia* tree constitutes an adaptive mechanism used by plants in natural, heterogeneous habitats (Brodribb & Hill, 1993; Pasquet-Kok *et al.*, 2010; Wang *et al.*, 2011).

A central core of the genetic regulatory circuit underlying age-dependent changes in plants is the accumulation of

*These authors contributed equally to this work.

SQUAMOSA PROMOTER BINDING PROTEIN LIKE transcription factors (SPLs), which act as a timing cue for phase change transitions and for the developmental trajectory of the plant (Chen *et al.*, 2010). Recent findings suggest that microRNAs (miRNAs), specifically the opposing expression patterns of miR156 and miR172, regulate a conserved framework of phase changes in many, if not all, angiosperms (Wang *et al.*, 2011). miR156 represses the expression of *SPL* transcription factor genes and shows high correlation with juvenile-like vegetative leaf traits. As the expression of miR156 decreases, the expression of *SPLs* increases in parallel with adult-like vegetative traits (Poethig, 2013). For example, the heteroblastic increases in the serration of simple leaves in *Arabidopsis thaliana* and the number of leaflets in compound leaves of *Cardamine hirsuta* are regulated by the destabilization of TCP-CUC transcription factors mediated by *SPLs* (Chitwood & Sinha, 2014; Rubio-Somoza *et al.*, 2014). The *CUP SHAPED COTYLEDON* (*CUC*) genes, belonging to the NAC family of transcription factors, operate as a conserved boundary-specification program that modulates the sites of leaf and leaflet initiation in eudicots (Blein *et al.*, 2008). CUC proteins form homo- and heterodimers that are biologically active and promote an increase in leaf complexity (Rubio-Somoza *et al.*, 2014). In early leaves of *A. thaliana* and *C. hirsuta*, TCP transcription factors interfere with the formation of functional CUC complexes by generating TCP-CUC heterodimers. As leaves at successive nodes progress through the heteroblastic series, accumulation of SPL proteins acts as a heterochronic cue that destabilizes TCP-CUC complexes. Specifically, SPLs compete with CUC proteins for access to TCP, thus, activating functional CUC complexes that promote increases in leaf complexity in the newly formed organs of both species (Rubio-Somoza *et al.*, 2014). The antagonism between TCP and CUC protein complexes is reflected at the transcriptomic level in tomato (Chitwood *et al.*, 2015).

Despite increasing knowledge of the molecular control of phase change, the gene regulatory networks driving the phenotypic expression of heteroblasty within the context of the natural variation are largely unknown. Although recent studies identify quantitative trait loci regulating leaf shape in both a heteroblastic and ontogenetic context in tomato and wild relatives (Chitwood *et al.*, 2014), and specifically implicate the flowering time regulator *FLC* in mediating the phase change in *C. hirsuta* (Cartolano *et al.*, 2015), the extent to which these results can be extrapolated to nonmodel species remains in question. In order to understand how developmentally programmed processes (e.g. heteroblasty) and environmental signals are combined to generate diverse phenotypes across species, studies between model and nonmodel species must be compared, as well as results from controlled environmental conditions integrated with those found in natural environments (Rowan *et al.*, 2011). Fortunately, recent technical advances in functional genomics (through next-generation sequencing) have revolutionized the study of gene regulatory networks and their role in biological processes and molecular functions in nonmodel species, even those with uncharacterized genomes (Grabherr *et al.*, 2011). The enormous potential of these new sequencing techniques in extending the range of studied species provides the opportunity to reach a comprehensive, comparative overview of the molecular and developmental underpinnings of leaf shape in plants.

Because of their long life span as well as their perenniality, forest trees are excellent experimental systems for comparative studies of the relationship between genotypic and phenotypic diversity in natural environments (Neale & Ingvarsson, 2008). Understanding of gene function and genetic variation controlling complex traits in forest trees can provide new clues about how climate and microclimatic variation shape geographic and population genetic structure, as well as the evolution of unique arboreal adaptations (e.g. heteroblasty) to the seasonal cycles that accompany the life-history trajectories of long-lived plants (Sork *et al.*, 2013). In the temperate rainforest of Chile and Argentina, *Gevuina avellana* (Proteaceae) is a native tree, and the only species among coexisting trees that shows strong heteroblastic changes in leaf shape (Ostria-Gallardo *et al.*, 2015). Additionally, as a member of the Proteaceae, *G. avellana* is evolutionarily well positioned, providing a counterpoint as a basal eudicot to all too common studies of Rosid and Asterid model species. It grows in a wide range of light environments along the forest canopy, ranging from 5% to *c.* 50% of canopy openness (Lusk, 2002). In addition, when growing under high-light conditions, the species produces larger and more complex leaves (Fig. 1; Ostria-Gallardo *et al.*, 2015).

The molecular regulation of simple vs compound leaf morphology is increasingly well understood (Bharathan & Sinha, 2001; Champagne *et al.*, 2007; Blein *et al.*, 2010; Townsley & Sinha, 2012; Ichihashi *et al.*, 2014; Tsukaya, 2014). However, the mechanisms underlying life-history transitions from simple to highly compound leaves, especially those induced by natural environmental triggers, is less well known, probably because overt morphological transitions across the leaf series are not a common developmental strategy. Heterophylly – dramatic, sometimes discrete, transitions in leaf morphology – can be developmentally programmed but coincident with profound changes in the environment (light or predators) a plant encounters during its life history, as reported in some vines (Zotz *et al.*, 2011) and tree species, such as lancewood (*Pseudopanax crassifolius*) endemic to New Zealand (Gould, 1993) and *Acacia* (Brodribb & Hill, 1993; Pasquet-Kok *et al.*, 2010; Wang *et al.*, 2011). Heterophylly can also be strictly environmentally induced. The semiaquatic heterophyllic species *Rorippa aquatica* (Brassicaceae) produces both simple and compound leaf types responding to changes in the surrounding environment (e.g. above or below water, ambient temperature, and light intensity). The different leaf shapes and degrees of complexity are achieved by the effect of multiple environmental cues on the regulation of GA and cytokinin concentrations via gene networks regulated by *KNOX1* and *CUC* genes (Nakayama *et al.*, 2014). Given the contingent nature in which the strong, environmentally informed heteroblastic development of *G. avellana* is deployed, this species is an excellent model to study the molecular basis of life-history changes in leaf morphology, from simple to highly compound leaves in a natural setting.

The advances in knowledge of genetic mechanisms underlying the architecture of simple and compound leaves across

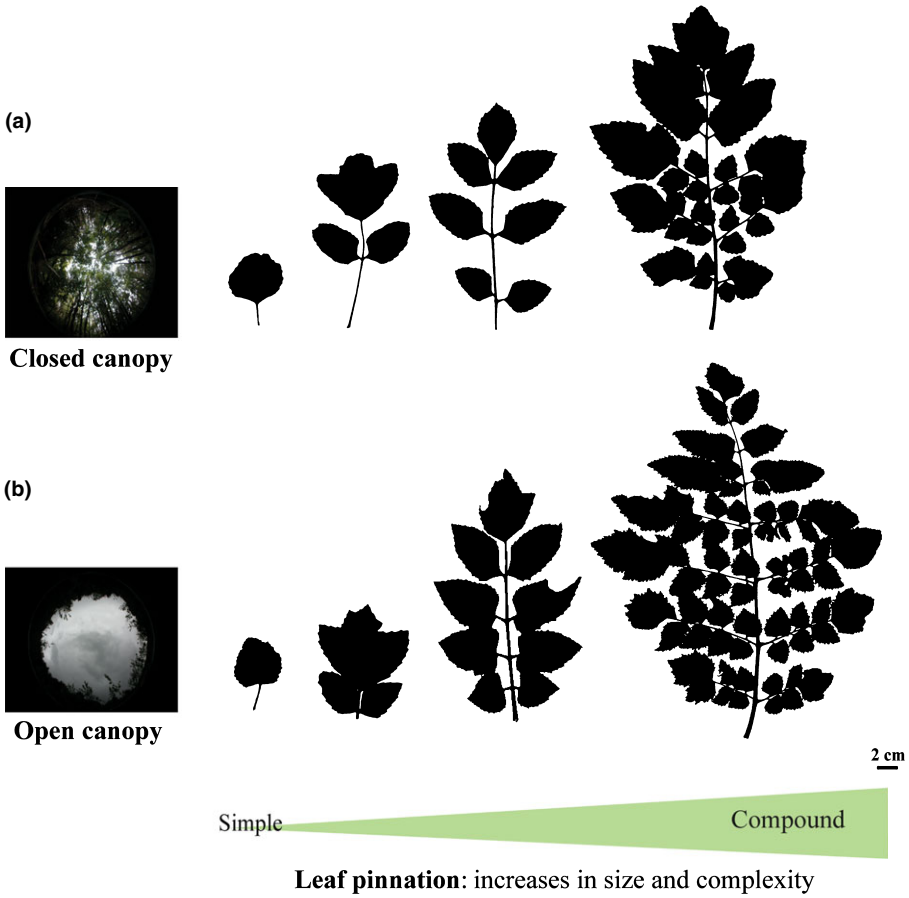

Fig. 1 Heteroblastic leaf development in *Gevuina avellana*. This species occupies a wide range of light environments, which has a small but significant effect on the phenotypic trajectory of leaves. Plants inhabiting closed canopy microsites have smaller and less complex leaves (a) than plants inhabiting more open canopy microsites (b).

evolutionary distances and the control of age-dependent changes in leaf shape/complexity, combined with the availability of new, high-throughput sequencing technologies to study genomic datasets in nonmodel species, make transcriptomic analyses of leaf development in nonmodel organisms feasible. We used RNA-seq on the Illumina platform to study the transcriptional responses of leaf primordia in *G. avellana* trees to the heteroblastic progression and natural light environments. Specifically, we asked if in *G. avellana* the expression of genes related to phase changes (e.g. *SPL*) is correlated to transcripts previously implicated in regulating leaf complexity and to what extent natural light environments regulate the heteroblastic leaf trajectory in *G. avellana* at a molecular level. Our analysis identified key genes correlated with plant development, leaf dissection, and light availability, and provides a broad view of the patterns of gene expression associated with the interaction between life history and light availability in the leaf primordia of a heteroblastic tree.

Materials and Methods

Study site, experimental design, and sampling

The study site corresponds to 30 hectares of a secondary temperate rainforest stand located in south central Chile (Katalapi Park, 41°31′07.5″S, 72°45′2.2″W). The forest structure corresponds to a coastal regenerating forest with patches of primary forest that have been protected from cattle grazing for 20 yr. The area produces *c.* four adult reproductive *Gevuina avellana* Mol. trees ha^{-1}, ensuring a homogeneous and sufficient supply of seeds. The canopy is almost exclusively composed of evergreen angiosperms, including *Nothofagus nitida* (Phill.) Krasser, *Laureliopsis philippiana* (Looser) Schodde, *Luma apiculata* (DC.) Burret, *Amomyrtus luma* (Molina) Legr. et Kaus, *Aextoxicum punctatum* R. et P., *Eucryphia cordifolia* Cav., and *Drimys winteri* J.R. et G. Forster (Saldaña & Lusk, 2003; Lusk & Corcuera, 2011). The climate is temperate with a humid oceanic influence (Di Castri & Hajek, 1976) although frost events occur from early autumn to spring (Reyes-Díaz *et al.*, 2005). Annual precipitation is *c.* 2000 mm, with maximum mean temperatures of 10°C in winter and 22°C in summer (for further climatic details see Coopman *et al.*, 2010). The growing season is concentrated between December and March, coinciding with a mild dry period where the lowest air relative humidity ranges between 45% and 55% (Escandón *et al.*, 2013).

Sampling was done from 09:00 h to 12:00 h at the late stage of the growing season (late March). The average temperature recorded at sampling time was 11.9°C. The average photosynthetic photon flux density recorded outside the canopy was 603.8 $\mu mol\ m^{-2}\ s^{-1}$. Three areas, 100 m long and 4 m wide, randomly distributed across the park, starting at the edge of the forest and continuing through the closed forest, were chosen in order to include most of the light gradient (Fig. 2a). We

characterized the light availability along the areas by the analysis of hemispherical photographs taken under homogeneous overcast days using a Nikon Coolpix 4500 digital camera equipped with an FC-E8 fisheye lens (Nikon, Tokyo, Japan). The camera was hand-leveled and -oriented so that the top of the image faced north (Chazdon & Fields, 1987). Photographs were analyzed for the percentage of canopy openness (%CO) using the GAP LIGHT ANALYZER 2.0 software (Institute of Ecosystem Studies, Millbrook, NY, USA), which allows a quantitative description of the canopy structure and amount of transmitted light through the canopy (GLA, Frazer *et al.*, 1999). We then established four light environment classes by interquartile ranges of %CO (hereafter called deep-shade (DS), shade (SH), semi-shade (SS), and sun (S)). Within each light environment, the healthiest seed-grown *G. avellana* plants were carefully selected, ranging from 4 to 200 cm in height, ensuring different heteroblastic stages within the vegetative phase by quantifying the number of leaflets on the last, fully expanded leaf on the main shoot. Given height differences of plants within each light environment, we characterized the light availability above each individual. The young leaf primordia from the shoot apex were collected in liquid nitrogen and stored at −80°C. To study the patterns of differential gene expression of the young leaf primordia of plants grown in different light environments and developmental stages, we ranked the plant height, number of leaflets, and light availability into regular intervals by quartiles to establish four classes each (Supporting Information Table S1).

RNA-seq library preparation and sequencing

RNA-seq libraries were prepared from the collected leaf primordia using a custom high-throughput protocol for Illumina RNA-seq library preparation (Fig. 2b; Kumar *et al.*, 2012). Because of the high content of secondary metabolites (interfering substances) that coprecipitated during our first attempts to extract nucleic acids from the samples, we modified the step of direct mRNA purification by adding 120 µl of polyvinylpyrrolidone (PVP-40) 20% per 1 ml of lysis/binding buffer (LBB) containing 2-mercaptoethanol and Antifoam A (Sigma-Aldrich). After bead beating, samples went through incubation at 65°C, shaking for 20 min. The supernatants were transferred to a Qiashredder column and were spun at maximum speed for 30 min. Subsequent steps were as described in Kumar *et al.* (2012). We obtained a total of 24 libraries (each library was made from the shoot apex of one *G. avellana* plant and we ensured a minimum of three biological replicates for each class; see details for each sample ID in Table S1), which were pooled and sequenced at the UC-Berkeley Genomics Sequencing Laboratory on a single lane of the HiSeq 2000 platform (Illumina Inc., San Diego, CA, USA), obtaining 100 bp paired-end reads.

Preprocessing of Illumina reads

Reads were preprocessed as described in Ranjan *et al.* (2014) and detailed here. The preprocessing of reads involved a quality filter trimming to remove all the reads with average Phred quality scores < 20 and low-quality bases from the 3′ end of the reads (Fig. S1). We then removed adapter/primer contamination and duplicated reads using custom Perl scripts. After quality filter processing, the reads were sorted into individual samples based on barcodes using fastx_barcode_splitter.pl and then trimmed using the FASTQ/A Trimmer script fastx_trimmer from FASTX_-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/).

De novo transcriptome assembly and prediction of open reading frames (ORFs)

We used the TRINITY software package (v. r2013-02-25) for efficient and robust *de novo* assembly of a transcriptome without a reference genome from RNA-seq data (Grabherr *et al.*, 2011). The assembly was performed at The Lonestar Linux Cluster at Texas Advance Computing Center (TACC, University of Texas, TX, USA) as described in Ranjan *et al.* (2014). All the subsequent bioinformatics and statistical analyses were performed either on our local servers or in the iPlant atmosphere and Discovery computing environment (Goff *et al.*, 2011).

In order to filter out transcriptional artifacts, such as chimeric or poorly supported contigs, original reads were mapped to assembled transcripts using BOWTIE2 with the following parameters: -a –rdg 6,5 –rfg 6,5 –score-min L,-6,-.4. SAMTOOLS was subsequently used to generate a bam alignment file (Li *et al.*, 2009; Langmead & Salzberg, 2012). EXPRESS software was then used to calculate abundance estimates for each transcript in terms of fragments per kilobase per transcript per million mapped reads (FPKM) and transcripts with ≥1 FPKM were retained for downstream analysis (Roberts and Pachter, 2013). Highly similar/redundant contigs were clustered using the CD-HIT clustering algorithm based on a similarity threshold of 95%, and a representative contig from each cluster was retained (Fu *et al.*, 2012). Subsequently ORFs were predicted from the filtered and clustered contigs using TRANSDECODER (https://transdecoder.github.io/), and these predicted and clustered ORFs were used for annotation and downstream analysis.

Functional annotation of the transcriptome

We compared predicted ORFs from the final transcriptome with the National Center for Biotechnology Information (NCBI) nonredundant (nr) database and with the Arabidopsis protein database (TAIR10 database) using BLASTX with an e-value threshold of 1e–3 in both cases (Altschul *et al.*, 1997). The BLASTX output file was used for BLAST2GO analysis to annotate the ORFs with gene ontology (GO) terms (e-value filter 1e–6) describing biological processes, molecular functions, and cellular components (Götz *et al.*, 2008). ANNEX and GO slim files were then used to enrich the annotation into functional categories (Ranjan *et al.*, 2014). EC numbers were also generated from the Kyoto Encyclopedia of Genes and Genomes pathway (Kanehisa & Goto, 2000). Finally, we obtained a sequence description file with arbitrary nomenclature based on degrees of similarity identified in both nr and TAIR10 databases according to e-value and identity with genes present in BLAST references.

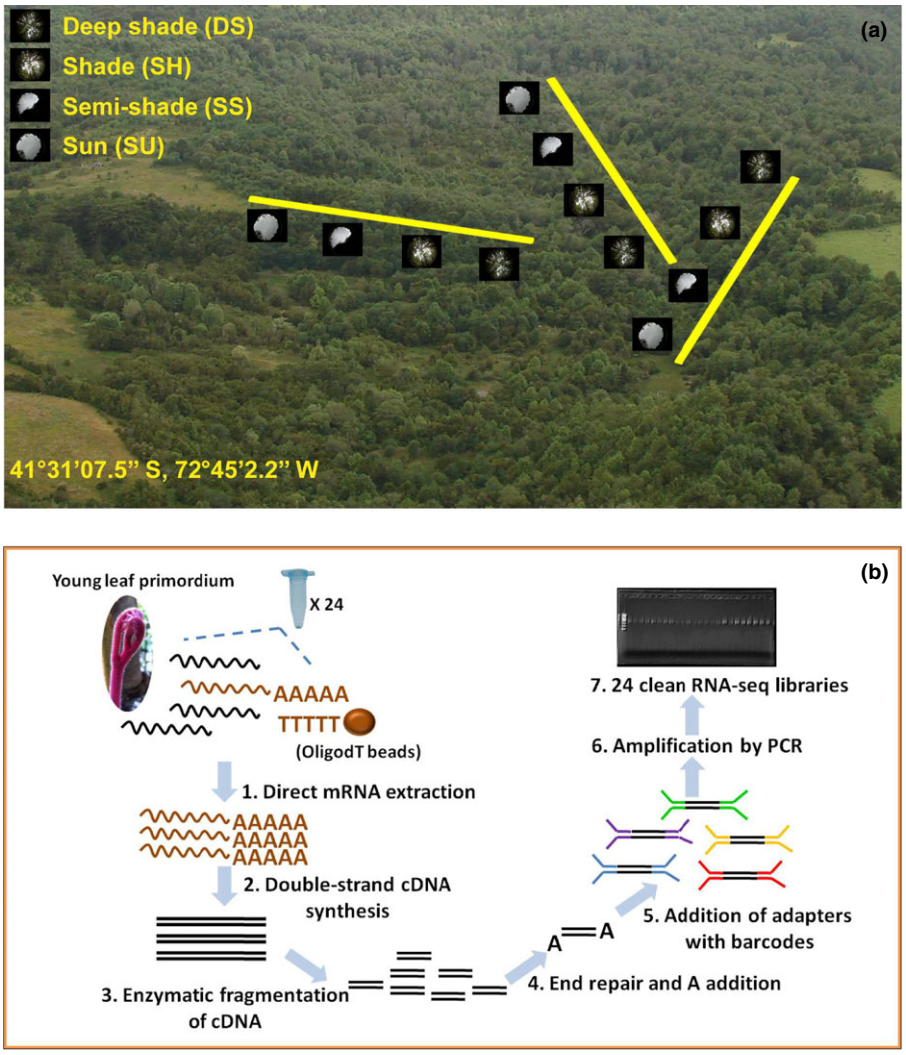

Fig. 2 (a) Study site with the three areas for collection of leaf primordia from *Gevuina avellana* trees of different sizes, under different canopy openness. (b) High-throughput RNA-seq libraries preparation (modified from Kumar *et al.*, 2012).

Differential expression analysis and GO enrichment analysis

The reads from each sample were mapped to the predicted ORFs using default RSEM parameters as described by Ranjan *et al.* (2014), and the abundances of the ORFs were extracted (Li & Dewey, 2011). To ensure a reliable differential gene expression analysis, we removed transcripts with very low estimated counts and then normalized the RSEM-estimated abundance values. Next, we used the run_DE_analysis.pl script based on the EdgeR Bioconductor package in the R statistical programming language to identify differentially expressed transcripts (Robinson & Oshlack, 2010; R Development Core Team 2012; Haas *et al.*, 2013). Differentially expressed transcripts were identified for all pairwise comparisons between each class of height, light, and degree of pinnation, with a *P*-value < 0.01. The RSEM and Perl scripts used are bundled in the TRINITY software package.

Gene ontology enrichment analysis of differentially expressed genes was conducted with the GOSeq Bioconductor package and GO terms and GOslim terms generated by BLAST2GO (Young *et al.*, 2010).

Principal component analysis (PCA) and self-organizing map (SOM) clustering

Normalized RSEM-estimated counts were used for clustering assembled ORFs based on expression patterns (Chitwood *et al.*, 2013). In order to detect the effects of light availability and ontogeny on gene expression, we selected genes from the upper 75% quartile of coefficient of variation for expression across plant development and light environments. The scaled expression values within samples were used to cluster these genes for a multidimensional 1×4 hexagonal SOM for plant height, number of leaflets, and light environment using the KOHONEN package on R (Wehrens & Buydens, 2007). In total, 100 training interactions were used during clustering with a decrease in the alpha learning rate from *c.* 0.0060 to 0.0040 (Fig. S2). SOM outcome was visualized in PCA space where PC values were calculated based on the gene expression of samples across height, number of leaflets, and light environments. In addition, we clustered genes for height, number of leaflets, and light environment in a single 1×3 hexagonal SOM to look at the distribution of clusters (nodes) and interaction among factors, resulting in groups of

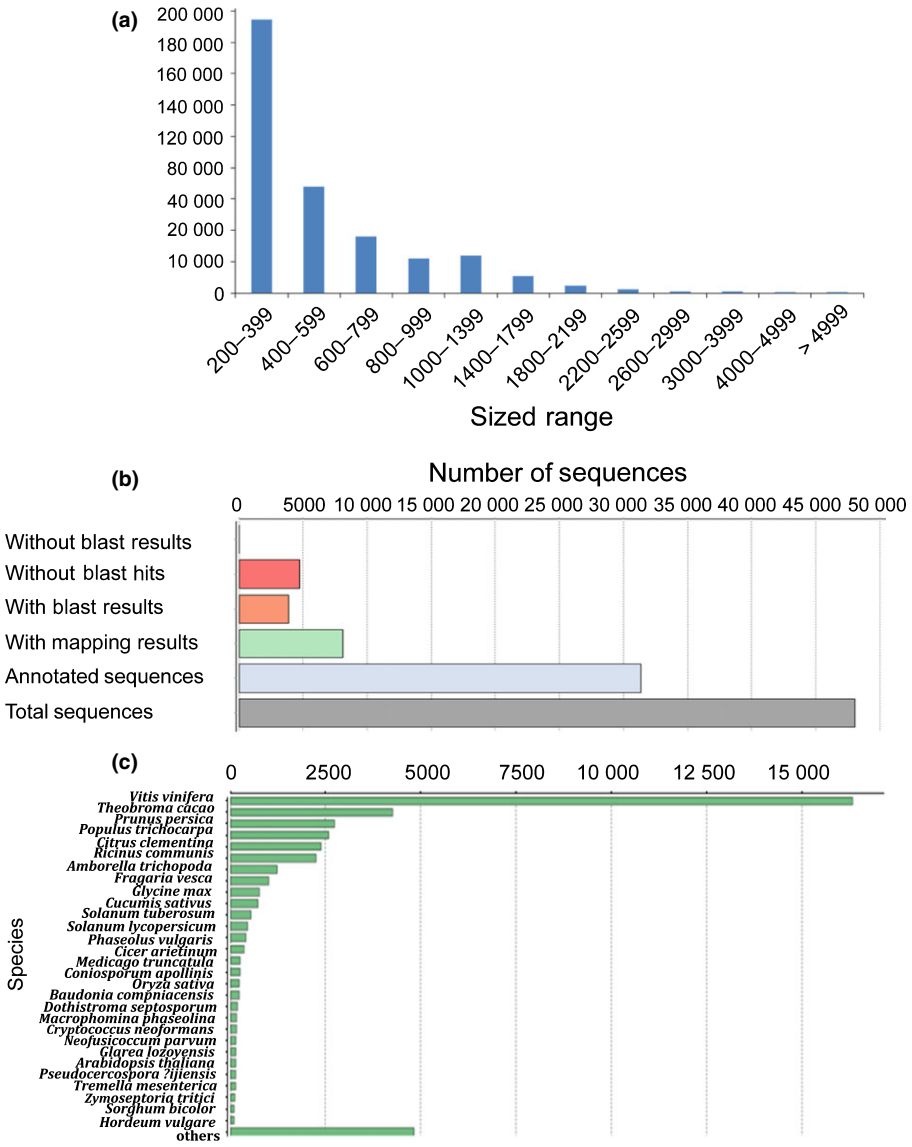

Fig. 3 (a) Transcript size distribution showing high proportion of small transcripts in the final transcriptome assembly of *Gevuina avellana*. (b) Sequence distribution after BLAST2GO analysis showing 70% of total sequences with annotation for predicted open reading frames (ORFs). (c) Top-hit species distribution of *G. avellana*'s final transcriptome showing abundance of top hits to the sequences of *Vitis vinifera* and tree species.

genes with common properties between factors (e.g. node 1 for height and light environment; Fig. S3).

Sequence submission

The quality-filtered, barcode-sorted, and trimmed short read data set, which was used for transcriptome assembly and gene expression analysis, was deposited in the NCBI Sequence Read Archive (SRA) under accession numbers SRR2787260, SRR2787261, SRR2787320, SRR2787501, SRR2787578, SRR2787722, and SRR2787810–SRR2787827.

The assembled transcripts (Gavel_transcriptome) have been deposited at DDBJ/EMBL/GenBank under accession number GEAC00000000. The version described in this article is the first version, GEAC01000000. Sequences of all predicted ORFs from Gavel_transcriptome can be downloaded as a FASTA file at http://de.iplantcollaborative.org/dl/d/9C816F17-3164-4B5F-B35B-DE01EB7FECC1/Gavel_predicted_ORF_CDS.fa.

Results

De novo assembly and transcriptome annotation

A total of 99 642 474 high-quality 100 bp paired-end reads were obtained after sequencing the RNAseq libraries. *De novo* transcriptome assembly using preprocessed reads and subsequent clustering of the assembled contigs at the threshold of 95% sequence identity yielded 185 301 *G. avellana* contigs (> 200 bp). The N50, which is the largest contig length, such that using equal or longer contigs produces half the bases of the transcriptome, was 686 bp long, while the average contig length was 538 bp (Fig. 3a; Notes S1). The prediction of coding sequences from the 185 301 transcripts resulted in 48 074 predicted ORFs (Fig. 3b), which were used for downstream analysis of differential expression. BLAST searches of predicted ORFs against the nr database and the TAIR10 Arabidopsis protein database resulted in the annotation of 43 341 and 38 975 ORFs, respectively. An insight

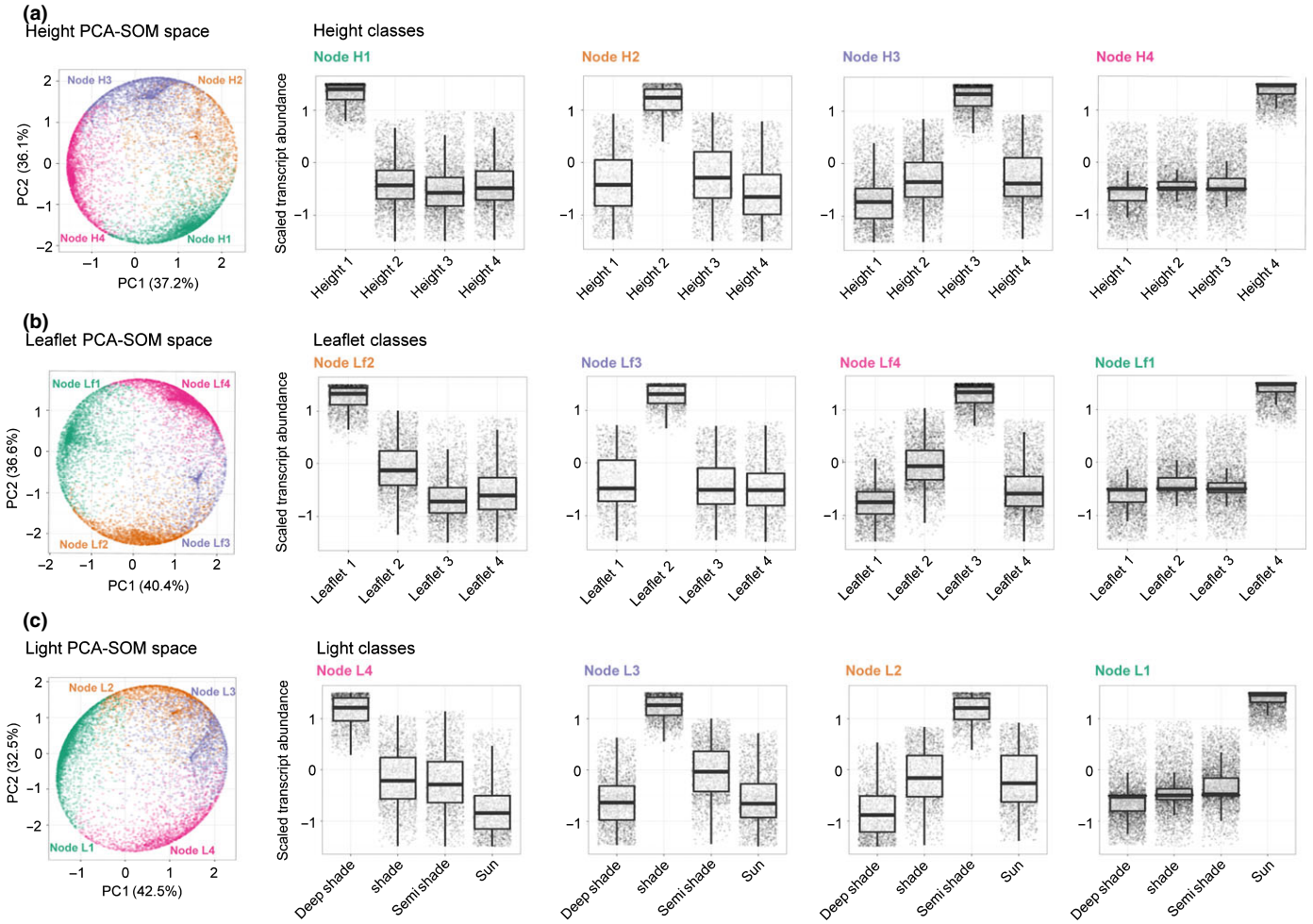

Fig. 4 (a–c) Principal component analysis (PCA) with self-organizing map (SOM) clustering of gene expression for the different plant height (a), number of leaflets (b), and light environment (c) classes. Each PCA-SOM space represents the expression profile of transcripts, indicating node membership by different colors, and node number plus suffix for each evaluated factor. A total of four clusters were defined for each factor showing the expression pattern specific to each class of *Gevuina avellana*'s size, pinnation, and light availability. Horizontal lines and bars in each boxplot represent the median and the maximum and minimum values of the scaled transcript abundance, respectively.

into the taxonomic distribution of top blast hits of ORFs against nr database revealed 73% of top hits to model woody species like *Vitis* and other trees (Fig. 3c).

Among predicted ORFs, the functional annotation resulted in 97 258 (52.48%), 42 858 (21.48%), and 39 815 (23.13%) annotated counts for biological process (BP), molecular function (MF), and cellular component (CC) categories, respectively (Fig. S4). Additionally, 12 963 ORFs were annotated as enzymes, with transferases being the most abundant class, followed by hydrolases and oxidoreductases (Fig. S5).

Transcript expression patterns across a heteroblastic series and light availability

A major focus of this study was to investigate the heteroblastic expression of genes correlated with height, leaf pinnation, and light availability. SOM clustering was used to describe transcripts with similar patterns of accumulation related to these factors, each partitioned into four clusters (nodes). We first studied the pattern of accumulation of transcripts related to plant height, which is the proxy for plant age (for details see Ostria-Gallardo *et al.*, 2015). For plant height, the nodes (hereafter node H*n*) were organized along height classes 1 to 4, showing accumulation patterns of transcripts through the progression of plant height from node H1 to node H4, as visualized using a combination of SOM and PCA (Fig. 4a). Each node was highly enriched for specific GO term, with an important contribution of BP categories. Node H1 transcripts were enriched in functional categories corresponding to response to red and far-red light, meristem activity, regulation of shape, shade avoidance, monooxygenase activity and auxin transport/signaling GOs. Transcripts in nodes H2 and H4 were enriched for ribosome and translation, apoptotic process, and growth rate GOs, whereas node H3 transcripts are enriched in the functional categories of anatomical structure, morphogenesis, regulation of flower development and floral organ formation, cell plate formation and cell proliferation, and epigenetic GOs (Notes S2).

Next, we studied the accumulation pattern of transcripts for the number of leaflets. As in plant height, the nodes (hereafter node Lf*n*) were organized by increases in leaf pinnation (Fig. 4b). Transcripts in nodes Lf1 and Lf3 were highly enriched for ribosome and translation GOs. Transcripts in node Lf2 were enriched for the functional categories of meristem initiation, monooxygenase activity, and electron transport GOs, whereas transcripts in node Lf4 were highly enriched for cell proliferation, regulation of meristem growth, ontogeny, leaf shape and leaf morphogenesis, phyllome, flower development, and epigenetic GOs (Notes S3). Finally, the accumulation pattern of transcripts related to light availability behaved similarly to the pattern in plant height and number of leaflets, explaining prominent densities of transcripts along increasing light availability (Fig. 4c). For example, node L4 exhibits high transcript accumulation in the DS light environment class and progresses along light environments to node L1 transcripts, which accumulate at high levels in the SU class. Transcripts in node L1 were enriched for ribosome and translation GOs. Transcripts in nodes L2 and L3 were enriched for the functional categories of cell proliferation, signaling and development, flower development, and epigenetic GOs. Transcripts in node L4 associated with DS were highly enriched in the functional categories of circadian rhythm, photoperiodism, and flowering GOs (Notes S4).

Differential transcript expression and GO enrichment analysis

Three independent differential transcript expression analyses were conducted using annotated ORFs for plant height, number of leaflets, and light environment to identify candidate genes that would be involved in heteroblastic development of *G. avellana* (Notes S5–S19). First, we studied gene expression changes (false discovery rate < 0.05) related to heteroblastic development by comparing height classes. As expected, increases in the number of differentially expressed genes were detected with increases in plant height (Notes S5–S10). Among all plant height comparisons, H1 vs H4 showed the largest number of differentially expressed genes, with 669 and 1514 down-regulated and up-regulated transcripts, respectively. GOslim categories enriched in up-regulated transcripts included transcription factor complex and sequence-specific DNA-binding transcription factor activity, oxidation-reduction process, catalytic activity, transporter activity, transmembrane transport and cell wall GOs. Transcripts of special interest under those GO terms were: transcription factors, monooxygenase activity and hormone synthesis and signaling, kinases, and cell wall formation/expansion (Table 1).

Next, we examined changes in gene expression associated with the degree of leaf pinnation. The highest number of differentially expressed transcripts was found between leaflet 1 and leaflet 3 classes. From a total of 929 differentially expressed transcripts, 486 and 443 showed down-regulation and up-regulation, respectively. GOslim categories in up-regulated genes included catalytic activity and metabolic process, oxygen binding, sequence-specific DNA-binding transcription factor activity, cell wall and translation regulator activity. Transcripts of interest under these GO terms for degree of pinnation were: members of transcription factor families, the argonaute RNA silencing family, hormone response genes, kinases, and cell wall expansion (Table 1).

In addition, considering that the progression of height and number of leaflets is part of the heteroblastic program, we found that three members of the *SPL* transcription factor family, *SQUAMOSA PROMOTER BINDING-LIKE* genes 4, 8 and 12 (*SPL4*, *SPL8*, *SPL12*), are up-regulated steadily as plants grow taller and leaves become more complex. Also, *NAC* genes from the *NAM/CUC* subgroup (of which several were in the CUC2/CUC3 clade – see Fig. S6), and *AGAMOUS*-like genes show, respectively, about twofold and sixfold up-regulation in sequential comparisons of plant height classes and between the extremes of leaflet numbers (L1 vs L3).

Next, we quantified the changes in expression of transcripts associated with light availability by comparing expressed transcripts in DS to those in SH, SS and S. The highest number of differentially expressed genes was found in DS vs S, with 574 and 2632 down-regulated and up-regulated transcripts, respectively. GOslim categories for up-regulated genes included catalytic activity, transcription factor activity, oxidation-reduction process, monooxygenases, and tryptophan metabolic process. Transcripts of particular interest were transcription factor families, kinases coding gene components of enzymatic reactions, phytochrome B-mediated light signal, hormone synthesis and signaling, and genes related with shade avoidance syndrome. Interestingly, we observed a sustained up-regulation of *NDPK1* genes from three- to eightfold along the light gradient. Also *NAC*, *AGAMOUS*-like and *JAZ* genes show steady up-regulation in sequential comparisons of light environment classes (Notes S14–S19).

Finally, we evaluated the patterns of overlap of up- and down-regulated genes between the three factors (Fig. S7) (Hulsen *et al.*, 2008). We found more overlap for up-regulated genes than for down-regulated genes, except between height and number of leaflets where the overlap was higher for down-regulated genes. Among factors, light showed the highest number of genes with no overlap in both up- and down-regulated groups of genes (3095 and 1667, respectively). For up-regulated genes, height and light availability showed the highest overlap, with 1248 shared genes, whereas height and number of leaflets showed the highest overlap in down-regulated genes, with 348. The total number of genes overlapped between the three factors was 130 for up-regulated genes and 65 for down-regulated genes. Genes that overlapped between each of the factors are shown in Table S2.

Discussion

In this study we examined gene expression in response to canopy openness in the natural habitat in the basal eudicot tree *G. avellana* and report on transcriptional dynamics of leaf primordia with respect to their morphology and plant age. This first *de novo* transcriptome assembly of a tree in the Proteaceae allowed us to identify key genes underlying heteroblasty, characteristic of other model organisms, and also to examine such well-established molecular pathways in the natural environment. Other studies have described 24 expressed sequence tag regions,

Table 1 Differential expression of transcripts (false discovery rate < 0.05) during heteroblastic development of *Gevuina avellana*

Type of transcript	Regulation	Height Genes (log FC)	Leaflets Genes (log FC)	Light Genes (log FC)
Transcription factors	Up	*AGL2* (4.7); *AGL6* (6.6); *AGL7* (3.8); *AGL9* (7.5); *AGL20* (1.7);*AGL22* (1.2); *ANAC002* (4.4); *ANAC039* (2.4); *ANAC055* (2.5); *ANAC058* (2.9); *ANAC100* (1.3); *ATHB7* (3.4);*ATHB17* (1.7); *ATHGD11* (2.5); *HDG2* (1.7); *GRF3* (6.5); *GRF5* (2.0); *MYB14* (1.8); *MYB62* (2.6); *MYB77* (6.1); *MYB90* (2.0); *MYB93* (3.0);*MYB102* (3.4); *NGA3* (3.4); *SPL4* (1.8); *SPL12* (4.6); *VRN1* (1.5); *YAB3* (2.9)	*AGL9* (4.9); *AGL22* (1.3); *ANAC002* (2.7); *ATHB7* (3.0); *ATHGD11* (2.0); *MYB102* (2.8); *MYB117* (2.3); *NAC007* (3.5); *NAC028* (2.5); *SCL8* (1.9); *SPL7* (1.5); *WOX1* (5.2); *YAB1* (3.1); *YAB2* (1.7); *YAB3* (2.9)	*AGL2* (5.4); *AGL6* (7.4); *AGL7* (5.1); *AGL9* (2.7); *AGL9* (4.9); *AGL20* (1.7); *ANAC002* (1.9); *ANAC036* (2.9); *ANAC039* (2.0); *ANAC042* (2.7); *ANAC043* (2.7); *ANAC055* (1.8); *ANAC058* (2.2); *ANAC083* (2.6); *ANAC100* (1.6); *AP3* (2.6); *ATHB7* (2.6); *GRF1* (4.6); *GRF2* (7.4); *GRF12* (4.3); *KNAT3* (1.1); *KNAT6* (1.1); *LBD1* (5.8); *LBD11* (3.7); *LBD15* (2.9); *LBD21* (4.9); *MYB4* (1.5); *MYB5* (1.1); *MYB6* (2.0); *MYB7* (3.4); *MYB31* (1.5); *MYB77* (7.0); *MYB82* (3.0); *MYB90* (2.3); *MYB102* (2.4); *MYB103* (6.2); *MYB107* (2.8); *MYB108* (6.5); *NAC073* (1.9); *NGA3* (2.8); *PHB* (2.1); *PHV* (1.7); *PI* (2.0); *SCL5* (1.9); *SCL8* (1.1); *SPL4* (1.5); *SPL7* (5.6); *SPL9* (2.6); *SPL13* (1.1); *WOX1* (3.1); *YAB1* (1.5); *YAB3* (2.2)
	Down	*ANAC043* (−6.9); *ATHB2* (−1.5); *KNAT1* (−2.7); *LBD414* (−4.0); *MYB42* (−6.3); *MYB46* (−4.7); *MYB66* (−1.4); *MYB85* (−3.8); *MYB103* (−6.2); *NAC007* (−3.4); *NAC028* (−5.9); *NAC071* (−4.2); *STM* (−1.9); *TCP20* (−1.3); *TPR2* (−1.2); *WOX4* (−2.1)	*ATHB2* (−1.3); *KNAT1* (−1.8); *MYB42* (−4.9); *MYB66* (−1.1); *MYB85* (−3.0); *MYB103* (−4.1); *STM* (−1.5); *TCP8* (−1.2); *TCP20* (−1.0)	*ANAC043* (−1.8); *KNAT1* (−1.7); *LBD21* (−4.2); *MYB6* (−7.1); *NGA3* (−3.4); *PHB* (−2.3); *PHV* (−1.6); *TCP7* (−6.2); *TCP8* (−1.8); *TPR2* (−1.3); *TPR3* (−1.6)
Monooxygenase activity	Up	*YUC7* (6.1); *YUC9* (6.8)		*YUC9* (7.9); *YUC11* (2.5)
Hormone synthesis, signaling, transport	Up	*IAGLU* (4.3); *LAX2* (7.3); *JAZ1* (2.6); *JMT* (4.3)	*JAZ10* (3.7); *JMT* (6.2)	*ARF4* (1.7); *ARF16* (1.1); *DLF2* (3.0); *GASA (3.3)*; *IAA9* (1.3); *IAA31* (5.6); *IAGLU* (2.3); *LAX2* (8.2); *JAZ1* (3.0); *JAZ2* (1.3); *JAZ10* (4.6); *JMT* (4.1); *PIN1* (2.4); *PIN5* (3.4); *PIN6* (1.7); *SAUR* (2.5)
	Down	*ARF9* (−7.7); *AUX1* (−2.4); *IAA9* (−1.2); *IAA13* (−1.8); *IAA16* (−2.0); *LAX1* (−1.8); *PIN1* (−1.5); *PIN3* (−1.7); *PIN5* (−3.7)	*IAA13* (−1.5); *IAA16* (−1.3); *LAX1* (−1.7)	*ARF4* (−7.4); *ARF9* (−7.4); *GASA* (−2.5); *IAA9* (−1.3); *IAA13* (−2.0); *IAA16* (−2.9); *IAGLU* (−1.5); *LAX2* (−2.7); *JAZ10* (−4.4); *JMT* (−4.2); *PIN1* (−1.4); *PIN5* (−4.5); *PIN6* (−2.2)
Kinases	Up	*NDPK1* (7.1); *CIPK11* (1.4); *CIPK19* (5.6); *PKS1* (3.3)	*PKS1* (3.4)	*NDPK1* (7.8); *CIPK11* (1.3); *CIPK12* (1.5); *CIPK20* (1.2)
	Down	*PKS4* (−1.7)		*PKS1* (−1.7)
Light responses	Up			*COP1* (1.9); *SRR1* (1.4)
	Down	*CIP8* (−1.6)	*CIP8* (−1.0)	*CIP8* (−1.3)
Argonaute	Up		*AGO9* (1.9)	*AGO2* (4.2); *AGO7* (2.0)
Cell wall formation/ expansion	Up	*COBL4* (3.0)	*COBL4* (3.4)	*COBL7* (1.9)
	Down	*COBL4* (−6.8)	*COBL4* (−4.0)	

The analyses were conducted for plants of different height, number of leaflets, and light environment. Gene identities were obtained with gene ontology (GO) enrichment by Blast2GO analysis. The level of regulation of each gene was estimated by value of the logarithm of fold-change (log FC) above and below 1.0 and −1.0, for up- and down-regulation, respectively.

representing the only data currently available at the NCBI database for this species (Riegel *et al.*, 2010), and thus our study could serve as a model for future works in the Proteacea family. The robustness of our transcriptome data allowed us to identify progressive enrichment in up-regulated genes related to phase change (Chen *et al.*, 2010), floral development (Kaufman *et al.*, 2009; Fernandez *et al.*, 2014), organ separation, and leaf complexity (Olsen *et al.*, 2005; Blein *et al.*, 2008) as plant development progressed and light availability increased. Surprisingly, genes reported as key components in the development of

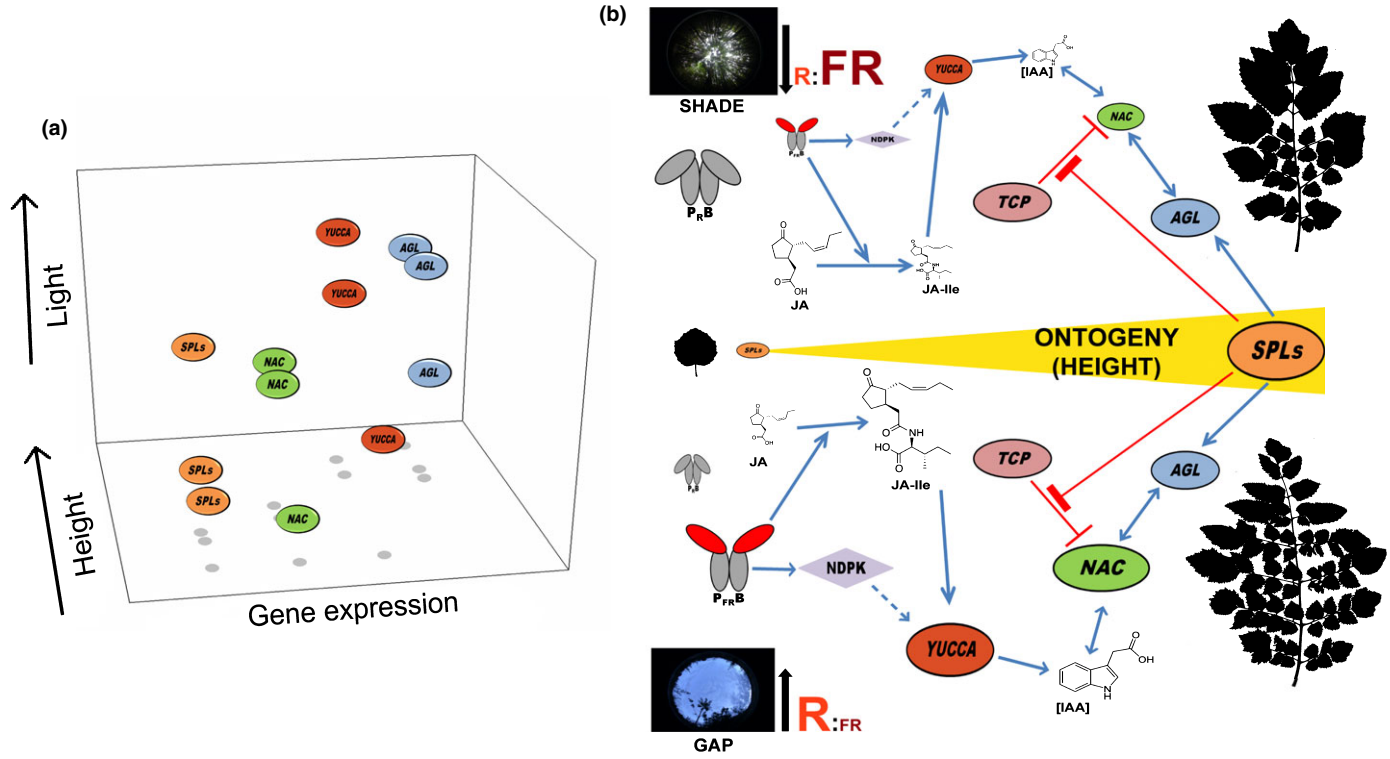

Fig. 5 (a) Up-regulation of *SPL*, *YUCCA*, and *AGL* genes in relation to increases in height and light availability. Relative values for gene expression derive from the maximum value of fold change (log FC) between quartile 1 and quartiles 2, 3 and 4, for both height and light. (b) Proposed model for the age-dependent changes in leaf shape and contribution of light, during the vegetative phase of *Gevuina avellana*. The change in leaf shape is regulated by the progressive expression of *SPL* genes during ontogeny, which achieve two pivotal roles: first, the up-regulation of *SPL* genes drives the destabilization of the TCP-NAC complex, and the active NAC proteins promote the increases in leaf complexity. Second, *SPL* genes act upstream of the expression of *AGAMOUS-like* genes (*AGL*) in a coordinated fashion, involving floral promoters and repressors. For the sequential changes in leaf complexity, *NAC* genes undergo an increased expression as a result of a positive feedback loop with *AGL* together with local auxin maxima along the rachis, controlled by the activity of nucleoside phosphate and the jasmonic acid-mediated expression of *YUCCA* genes. This is a process highly coordinated by phytochrome B in the context of light availability and the enrichment of red light sensed by the plants as they grow taller. Arrow, induction/activation; dashed arrow, signal cascade; perpendicular line, repression.

compound leaves, such as class I *KNOX* or *FLO/LFY* genes (Bharathan & Sinha, 2001; Champagne *et al.*, 2007; Ge *et al.*, 2014), turned out to be either down-regulated or absent in our *G. avellana* transcriptome data, respectively. One plausible explanation for these findings is that class I *KNOX* genes do not have a pivotal role in the heteroblastic increase of leaf complexity in *G. avellana* and that this role is assumed by other regulatory factors, such as *SPL*, *NAM/CUC* and *AGAMOUS*-like genes (*AGL*). Such an observation is consistent with transcriptomic results analyzing the shade avoidance response in tomato, where *KNOX* genes play a pivotal role in increasing leaf complexity in simulated shade conditions, but through a heteroblastic-independent mechanism (Chitwood *et al.*, 2015). In turn, the absence of *FLO/LFY* could be the result of assembly or sampling methods as the gene is known to be expressed at very low levels (Coen *et al.*, 1990). As discussed later, the current transcriptomic analyses suggest a direct effect of *SPL*, *NAM/CUC* and *AGL* genes on the regulation of heteroblastic development of *G. avellana*. This is possible, considering the presence of *AGL* genes and their progressive up-regulation in relation to height and light availability. If our assumptions are correct, then instead of the two common mechanisms regulating compound leaves described to date, these leaf types arose through divergent developmental paths and rewiring of gene regulatory networks (GRNs) that have converged to generate compound leaves (Hasson *et al.*, 2010; Townsley & Sinha, 2012).

Integration of internal-external cues for heteroblastic development

Based on differential gene expression analysis, transcript accumulation patterns, and enrichment in GO terms and functional categories for each class of the factors evaluated, we focused our attention on those genes that link the development of individual leaves under natural light conditions of *G. avellana* with the progression of heteroblasty in the species (Fig. 5). The transcriptome of *G. avellana* showed up-regulation of the *SPL* genes 4, 8, and 12 in relation to plant height, and *SPL* 4, 7, 9, 12 and 13 in relation to light availability. *SPL* 4, 8 and 12 are involved in GA signaling, floral induction/fruit ripening, and programmed cell death (Chen *et al.*, 2010). *SPL* 7 and 9 participate mainly in regulating shoot maturation and plant architecture. This suggests that the up-regulation of *SPL* genes has an important role in the heteroblastic development of *G. avellana* during the vegetative

phase. Probably *SPL* genes drive the timing and progress of heteroblasty but not leaf pinnation *per se*. The up-regulation of NAC transcription factors of the *CUC* type has been shown to facilitate the sequential change from simple to compound leaves. Specifically, the requirement of *NAM/CUC* activity is essential in promoting leaflet separation locally and leaflet formation at a distance by interacting with other regulators of compound leaf development (Ooka *et al.*, 2003; Blein *et al.*, 2008). This dual role is likely to be a conserved molecular framework for compound leaf development within eudicots (e.g. *C. hirsuta*, *Solanum lycopersicum*, and *Pisum sativum*; Blein *et al.*, 2008). In all the studied species with compound leaf development, the *NAM/CUC3* genes are required for proper expression of *KNOXI/LFY*-like genes and vice versa, indicating a coordinated regulation of leaflet formation through a positive feedback loop (Blein *et al.*, 2008). Interestingly, the comparison of gene expression in our *G. avellana* transcriptome shows steady up-regulation of members of the CUC clade (*ANAC039*, *ANAC058*, *ANAC100*; see Fig. S6) but an overall down-regulation of class I *KNOX* members *KNAT6*, *BP* and *STM* and the absence of *LFY*-like genes. Nonetheless, we found up-regulation of several *AGL* genes such as *AP1*, *SEP1* and *SEP3*, as well as the floral repressor *SHORT VEGETATIVE PHASE* (*SVP*) correlated with plant height. Fernandez *et al.* (2014) showed that the interaction of AGL transcription factors of both types, repressors and promoters of flowering, are key in controlling the transition from the vegetative to the reproductive phase. In addition, the expression of *AGL* floral promoting genes during the vegetative phase can produce changes in the leaf morphology of Arabidopsis. We observed the presence of *AGL* genes and their progressive up-regulation in relation to height and light availability. Thus, we assume that *AGL* genes have a direct effect on the regulation of heteroblastic development of *G. avellana*, and that during the vegetative phase of *G. avellana*, the expression of *AGL* floral promoter/repressor genes in leaf primordia would be a component of the mechanism regulating its leaf pinnation. For example, in Arabidopsis mutants, the expression of *AGL* genes in leaf tissues promotes changes in leaf morphology producing upward curling in rosette and cauline leaves, and also blocks the mechanism of floral induction during the vegetative phase (Fernandez *et al.*, 2014). We speculate that the mechanism that controls compound leaf development in *G. avellana* underwent GRN rewiring, selecting *AGL* genes for an analogous role to those of *KNOXI* genes seen in almost all compound-leaved species and the *LFY* gene in a derived clade in the Fabaceae. It is also possible that the up-regulation of *AGL* genes in sequentially older plants represents a more mature state that is verging towards reproductive development.

At the incipient stage, leaf development requires local auxin maxima in cells at the flanks of the shoot apical meristem (SAM) and the rachis, allowing expression of the genes controlling initiation and separation of both leaves (from the SAM) and leaflets (from each other; Blein *et al.*, 2010; Townsley & Sinha, 2012). Whether correlated with developmental stage and/or light availability, we found both up-regulated and down-regulated transcripts related to auxin transport and signaling, which suggests an active regulation of processes controlled by this hormone. However, *YUCCA* genes, particularly *YUC7* and *YUC9*, showed steady up-regulation with increasing light availability and plant development. Also, we found up-regulation of *JAZ*-domain protein coding genes, a key component of jasmonic acid (JA) signaling, with the increase in both age and light availability. *YUCCA* genes have a pivotal role in the tryptophan-dependent auxin biosynthesis pathway. In turn, activation of the *YUCCA* promoter is modulated by JA, which is also known to be involved in the distribution of the auxin-exporter *PIN-FORMED 2* (*PIN2*) (Hentrich *et al.*, 2013; Brumos *et al.*, 2014). Given that JA usually acts at long distances from the source to the target tissue (Lucas & Lee, 2004), an increase of *JAZ* genes serves as indirect evidence of JA activity. JAZ are ubiquitinated in the presence of JA, which triggers JA-responsive gene expression (Chini *et al.*, 2009). This process is balanced by a positive feedback loop between JA and JAZ, avoiding exacerbated responses (Kazan & Manners, 2008; Pauwels & Goossens, 2011). Thus, the up-regulation of JAZ genes in *G. avellana* might reflect a rapid turnover of this protein by JA and an increase in JA-mediated responses such as the activation of *YUCCA* promoters. Several JA-mediated responses (e.g. shade avoidance, defense mechanisms) are driven by light quality mediated by phytochromes (Ballaré, 2009). As the plant perceives an increment in the red to far-red ratio (R : FR), a signal cascade mediated by phytochrome B induces the conversion of inactive JA to the active JA-Ile (Moreno *et al.*, 2009; Radhika *et al.*, 2010). Given the contingent nature of development in *G. avellana*, as the plant grows there is a vertical increase in light availability concomitant with an increase in red light. It is known that phytochrome B (phyB) mediates several photomorphogenic responses under enriched red light through direct interaction with basic helix–loop–helix domain transcription factors, resulting in the activation of multiple genes (Alabadi *et al.*, 2001). Therefore, for this tree, we hypothesize that the JA-responsive expression of *YUCCA* genes progresses coordinately with the increase of plant development and light availability.

Together with our previous finding of the significant effects of light availability on leaf size and complexity (Ostria-Gallardo *et al.*, 2015), the morphological and physiological fate of leaf primordia is closely linked to the interaction between developmental progression and the prevailing light environment. The resulting heteroblastic development in *G. avellana* has most likely been shaped by gene interactions in GRNs that were selectively maintained as they were functionally advantageous in allowing the plant to cope with the predictable environmental changes throughout the life-history trajectory of the plant (e.g. increasing vapor pressure deficit and light availability). These interactions may have undergone stochastic rewiring during the course of evolution from the basal eudicots to the well-characterized model taxa. Thus, our analysis sheds light on the integration of internal and external clues in the progression of heteroblastic development in *G. avellana*. Further studies should be pursued using approaches that cover development within an environmental context to identify the signal perception and transduction mechanisms through which the environment informs and modulates

development, as pointed out by Sultan (2010). These studies might help to explain why *G. avellana* is the only strongly heteroblastic species in this temperate rainforest and how the resulting leaf phenotypes influence its natural distribution along the gradient of light availability.

Acknowledgements

E.O-G. thanks the Chilean National Commission for Scientific and Technological Research for a doctoral fellowship, and the internship grant supported by Universidad de Concepción, project Mecesup UCO0708. Also thanks are due to Dr León Bravo, Holly Forbes, and Alejandro Navarrete for fieldwork, and Katalapi Park for excellent research field facilities. We thank the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. Part of the work was supported by NSF PGRP grant IOS-1238243 (to Julia Bailey-Serres, N.R.S., Siobhan Brady and Roger Deal). This work used computational resources or cyberinfrastructure provided by the iPlant Collaborative (http://www.iplantcollaborative.org). The iPlant Collaborative is funded by National Science Foundation Grant DBI-0735191.

Author contributions

E.O-G., L.J.C. and N.R.S. planned and designed the research. E.O-G. and L.J.C. conducted fieldwork. E.O-G, A.R., K.Z., R.K. and Y.I. performed RNA-seq experiments and bioinformatics. A.R. performed *de novo* transcriptome assembly. E.O-G., A.R., D.H.C. and B.T.T. analyzed data. E.O-G. wrote the manuscript with contributions from other authors.

References

Alabadi D, Oyama T, Yanovsky MJ, Harmon FG, Más P, Kay SA. 2001. Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science* **293**: 880.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gepped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

Ballaré C. 2009. Illuminated behavior: phytochrome as a key regulator of light foraging and plant anti-herbivore defense. *Plant, Cell & Environment* **32**: 713–725.

Bharathan G, Sinha NR. 2001. The regulation of compound leaf development. *Plant Physiology* **127**: 1533–1538.

Blein T, Hasson A, Laufs P. 2010. Leaf development: what it needs to be complex. *Current Opinion in Plant Biology* **13**: 75–82.

Blein T, Pulido A, Vialette-Guiraud A, Nikovics K, Morin H, Hay A, Johansen IE, Tsiantis M, Laufs P. 2008. A conserved molecular framework for compound leaf development. *Science* **322**: 1835–1839.

Brodribb T, Hill RS. 1993. A physiological comparison of leaves and phylodes in *Acacia melanoxylon*. *Australian Journal of Botany* **41**: 293–305.

Brumos J, Alonso JM, Stepanova AN. 2014. Genetic aspects of auxin biosynthesis and its regulation. *Physiologia Plantarum* **151**: 3–12.

Burns KC. 2005. Plastic heteroblasty in beach groundsel (*Senecio lautus*). *New Zealand Journal of Botany* **43**: 665–672.

Cartolano M, Pieper B, Lempe J, Tattersall A, Huijser P, Tresch A, Darrah PR, Hay A, Tsiantis M. 2015. Heterochrony underpins natural variation in *Cardamine hirsuta* leaf form. *Proceedings of the National Academy of Sciences, USA* **112**: 10539–10544.

Casal JJ, Fankhauser C, Coupland G, Blázquez MA. 2004. Signalling for developmental plasticity. *Trends in Plant Science* **9**: 309–314.

Champagne C, Goliber T, Wojciechowski M, Mei R, Townsley B, Wang K, Paz M, Geeta R, Sinha NR. 2007. Compound leaf development and evolution in the legumes. *Plant Cell* **19**: 3369–3378.

Chazdon R, Field C. 1987. Photographic estimation of photosynthetically active radiation: evaluation of a computerized technique. *Oecologia* **73**: 525–532.

Chen X, Zhang Z, Liu D, Zhang K, Li A, Mao L. 2010. *Squamosa* promoter-binding protein-like transcription factors: star players for plant growth and development. *Journal of Integrative Plant Biology* **52**: 946–951.

Chini A, Boter M, Solano R. 2009. Plant oxylipins: *COI1*/*JAZs*/*MYC2* as the core jasmonic acid-signaling module. *FEBS Journal* **276**: 4682–4692.

Chitwood DH, Kumar R, Ranjan A, Pelletier JM, Townsley B, Ichihashi Y, Martinez CC, Zumstein K, Harada JJ, Maloof JN *et al.* 2015. Light-induced indeterminacy alters shade avoiding tomato leaf morphology. *Plant Physiology* **Pii**: 01229:2015.

Chitwood DH, Maloof JN, Sinha NR. 2013. Dynamic transcriptomic profiles between tomato and a wild relative reflect distinct developmental architectures. *Plant Physiology* **162**: 537–552.

Chitwood DH, Ranjan A, Kumar R, Ichihashi Y, Zumstein K, Headland LR, Ostria-Gallardo E, Aguilar-Martínez JA, Bush S, Carriedo L *et al.* 2014. Resolving distinct genetic regulators of tomato leaf shape within a heteroblastic and ontogenetic context. *Plant Cell* **26**: 3616–3629.

Chitwood DH, Sinha NR. 2014. Plant development: small RNAs and the metamorphosis of leaves. *Current Biology* **24**: R1087–R1089.

Coen ES, Romero JM, Doyle S, Elliott R, Murphy G, Carpenter R. 1990. *floricaula:* a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **6**: 1311–1322.

Coopman RE, Fuentes-Neira FP, Briceño VF, Cabrera HM, Corcuera LJ, Bravo LA. 2010. Light energy partitioning in photosystems I and II during development of *Nothofagus nitida* growing under different light environments in the Chilean evergreen rainforest. *Trees* **24**: 247–259.

Day JS. 1998. Light conditions and the evolution of heteroblasty (and the divaricate form) in New Zealand. *New Zealand Journal of Ecology* **22**: 43–54.

Di Castri F, Hajek E. 1976. *Bioclimatología de Chile*. Santiago, Chile: Ediciones de la Pontificia Universidad Católica de Chile.

Diggle PK. 1997. Ontogenetic contingency and floral morphology: the effects of architecture and resource limitation. *International Journal of Plant Sciences* **158**: 99–107.

Escandón AB, Paula S, Rojas R, Corcuera LJ, Coopman RE. 2013. Sprouting extends the regeneration niche in temperate rainforests: the case of the long-lived tree *Eucryphia cordifolia*. *Forest Ecology and Management* **310**: 321–326.

Fernandez DE, Wang CT, Zheng Y, Adamczyk BJ, Singhal R, Hall PK, Perry SE. 2014. The MADS-domain factors AGAMOUS-Like15 and AGAMOUS-Like18, along with SHORT VEGETATIVE PHASE and AGAMOUS-Like24, are necessary to block floral gene expression during the vegetative phase. *Plant Physiology* **165**: 1591–1603.

Frazer GW, Canham CD, Lertzman KP. 1999. *Gap Light Analyzer (GLA), version 2.0: imaging software to extract canopy structure and gap light transmission indices from true-colour fisheye photographs, user manual and program documentation*. Millbrook, NY, USA: Simon Frasr University, Burnaby, British Columbia, and the Institute of Ecosystem Studies.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-Hit: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150–3152.

Gamage HK. 2010. Leaf serration in seedlings of heteroblastic woody species enhance plasticity and performance in gaps but not in the understory. *International Journal of Ecology* **2010**: 683589.

Gamage HK. 2011. Phenotypic variation in heteroblastic woody species does not contribute to shade survival. *AoB Plants* **2011**: plr013.

Ge L, Peng J, Berbel A, Madueño F, Chen R. 2014. Regulation of compound leaf development by *PHANTASTICA* in *Medicago truncatula*. *Plant Physiology* **164**: 216–228.

Goebel K. 1900. *Organography of plants. I. general organography*. New York, NY, USA: Hafner Publishing Co.

Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, Gessler D, Matasci N, Wang L, Hanlon M, Lenards A *et al.* 2011. The iPlant Collaborative: cyberinfrastructure for plant biology. *Frontiers in Plant Science* **2**: 34.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**: 3420–3435.

Gould KS. 1993. Leaf heteroblasty in *Pseudopanax crassifolius*: functional significance of leaf morphology and anatomy. *Annals of Botany* **71**: 61–70.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**: 644–652.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M *et al.* 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**: 1494–1512.

Hasson A, Blein T, Laufs P. 2010. Leaving the meristem behind: the genetic and molecular control of leaf patterning and morphogenesis. *Comptes Rendus Biologies* **333**: 350–360.

Hentrich M, Böttcher C, Düchting P, Cheng Y, Zhao Y, Berkowitz O, Masle J, Medina J, Pollmann S. 2013. The jasmonic cid signaling pathway is linked to auxin homeostasis through the modulation of *YUCCA*8 and *YUCCA*9 gene expression. *Plant Journal* **74**: 626–637.

Hulsen T, de Vlieg J, Alkema W. 2008. BioVenn – a web application for comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**: 488–493.

Ichihashi Y, Aguilar-Martínez JA, Farhi M, Chitwood DH, Kumar R, Millon LV, Maloof JN, Sinha NR. 2014. Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proceedings of the National Academy of Sciences, USA* **111**: E2616–E2621.

Jones C. 1999. An essay on juvenility phase change and heteroblasty in seed plants. *International Journal of Plant Sciences* **160**: 105–111.

Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**: 27–30.

Kazan K, Manners JM. 2008. Jasmonate signaling: toward an integrated view. *Plant Physiology* **146**: 1459–1468.

Kaufmann K, Muiño JM, Jauregui R, Airoldi CA, Smaczniak C, Krajewski P, Angenent GC. 2009. Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS Biology* **7**: 854.

Kumar R, Ichihashi Y, Kimura S, Chitwood DH, Headland LR, Peng J, Maloof J, Sinha NR. 2012. A high-throughput method for Illumina RNA-Seq library preparation. *Frontiers in Plant Science* **3**: 202.

Langmead B, Slazberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

Li H, Handsker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 1754–1760.

Lucas WJ, Lee JY. 2004. Plasmodesmata as a supracellular control network in plants. *Nature Reviews Molecular Cell Biology* **5**: 712–726.

Lusk CH. 2002. Leaf area accumulation helps juvenile evergreen trees tolerate shade in a temperate rainforest. *Oecology* **132**: 188–196.

Lusk CH, Corcuera LJ. 2011. Effects of light availability and growth rate on leaf lifespan of four temperate rainforest Proteaceae. *Revista Chilena de Historia Natural* **84**: 269–277.

Moreno JE, Tao Y, Chory J, Ballaré CL. 2009. Ecological modulation of plant defense via phytochrome control of jasmonate sensitivity. *Proceedings of the National Academy of Sciences, USA* **24**: 4935–4940.

Nakayama H, Nakayama N, Seiki S, Kojima M, Sakakibara H, Sinha NR, Kimura S. 2014. Regulation of the *KNOX-GA* gene module induces heterophyllic alteration in North American lake cress. *Plant Cell* **26**: 4733–4748.

Neale DB, Ingvarsson PK. 2008. Population, quantitative and comparative genomics of adaptation in forest trees. *Current Opinion in Plant Biology* **11**: 149–155.

Nicotra A, Leigh A, Boyce CK, Jones CS, Niklas KJ, Royer DL, Tsukaya H. 2011. The evolution and functional significance of leaf shape in the angiosperms. *Functional Plant Biology* **38**: 535–552.

Olsen AN, Ernst HA, Leggio LL, Skriver K. 2005. *NAC* transcription factors: structurally disctinct, functionally diverse. *Trends in Plant Science* **2**: 79–87.

Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakami K, Matsubara K, Osato N, Kawai J, Carninci P *et al.* 2003. Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA Research* **10**: 239–247.

Ostria-Gallardo E, Paula S, Corcuera LJ, Coopman RE. 2015. Light environment has little effect on heteroblastic development of the temperate rainforest tree *Gevuina avellana* Mol. (Proteaceae). *International Journal of Plant Sciences* **176**: 285–293.

Pasquet-Kok J, Creese C, Sack L. 2010. Turning over a new 'leaf': multiple functional significances of leaves versus phyllodes in Hawaiian *Acacia koa*. *Plant, Cell & Environment* **33**: 2084–2100.

Pauwels L, Goossens A. 2011. The JAZ proteins: a crucial interface in the Jasmonate signaling cascade. *Plant Cell* **23**: 3089–3100.

Poethig RS. 2013. Vegetative phase change and shoot maturation in plants. *Current Topics in Developmental Biology* **105**: 125–152.

R Development Core Team. 2012. *R: a language and environment for statistical computing*. URL http://www.R-project.org/. Vienna, Austria: the R foundation for statistical computing.

Radhika V, Kost C, Mithöfer A, Boland W. 2010. Regulation of extrafloral nectar secretion by jasmonates in lima bean is light dependent. *Proceedings of the National Academy of Sciences, USA* **107**: 17228–17233.

Ranjan A, Ichihashi Y, Farhi M, Zumstein K, Townsley B, David-Schwartz R, Sinha NR. 2014. De Novo assembly and characterization of the transcriptome of the parasitic weed dodder identifies genes associated with plant parasitism. *Plant Physiology* **166**: 1186–1199.

Reyes-Díaz M, Alberdi M, Piper F, Bravo LA, Corcuera LJ. 2005. Low temperature responses of *Nothofagus dombeyi* and *Nothofagus nítida* two evergreen species from south central Chile. *Tree Physiology* **25**: 1389–1398.

Riegel R, Diaz L, Véliz D. 2010. *De novo* sequencing and development of EST-SSR for *Gevuina avellana* (Proteaceae), a nut tree from South America. *American Journal of Botany* **97**: e133–e135.

Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods* **10**: 71–73.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**: R25.

Rowan BA, Weigel D, Koenig D. 2011. Developmental genetics and new sequencing technologies: the rise of nonmodel organisms. *Developmental Cell* **21**: 65–76.

Rubio-Somoza I, Zhou CM, Confraria A, Martinho C, von Born P, Baena-Gonzalez E, Wang JW, Weigel D. 2014. Temporal control of leaf complexity by MiRNA-regulated licensing of protein complexes. *Current Biology* **24**: 2714–2719.

Saldaña A, Lusk CH. 2003. Influencia de las especies del dosel en la disponibilidad de recursos y regeneración avanzada en un bosque templado lluvioso del sur de Chile. *Revista Chilena de Historia Natural* **76**: 639–650.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB. 2013. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics and Genomes* **9**: 901–911.

Sultan SE. 2010. Plant developmental responses to the environment: eco-devo insights. *Current Opinion in Plant Biology* **13**: 96–101.

Townsley BT, Sinha NR. 2012. A new development: evolving concepts in leaf ontogeny. *Annual Review of Plant Biology* **63**: 535–562.

Tsukaya H. 2014. Comparative leaf development in angiosperms. *Current Opinion in Plant Biology* **17**: 103–109.

Wang JW, Park MY, Wang LJ, Koo Y, Chen XY, Weigel D, Poethig RS. 2011. MiRNA control of vegetative phase change in trees. *PLoS Genetics* **7**: e1002012.

Wehrens R, Buydens LM. 2007. Self- and Super-organizing maps in R: the kohonen package. *Journal of Statistical Software* **21**: 1–19.

Winn A. 1999. The functional significance and fitness consequences of heterophylly. *International Journal of Plant Sciences* **160**: 113–121.
Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**: R14.
Zotz G, Whilhem K, Becker A. 2011. Heteroblasty – a review. *Botanical Review* 77: 109–151.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Quality scores and accuracy of Illumina Hiseq 100 bp paired-end reads.

Fig. S2 Training progress of the average distances of genes for the self-organizing maps (SOMs) used for height, leaflets, and light environment factors.

Fig. S3 Accumulation pattern of transcripts among height, leaflets, and light.

Fig. S4 Bar charts for GO distribution of annotated transcripts in three categories: biological processes, cellular components, and molecular function.

Fig. S5 Enzyme code distribution for the final *G. avellana* transcriptome.

Fig. S6 Phylogenetic tree of NAC1 and CUC clade genes from *Gevuina avellana*, *Arabidopsis thaliana* and *Solanum lycopersicum*.

Fig. S7 Venn diagrams of genes up- and down-regulated between each of the factors evaluated.

Table S1 Experimental design for library preparation and identification of sequence membership of *G. avellana* leaf priomordia

Table S2 Up- and down-regulated genes of interest overlapped between each of the factor classes

Notes S1 *De novo* assembly statistics for *G. avellana* contigs, predicted ORFs, and BLAST hits of predicted ORFs.

Notes S2 Enriched GO categories, both all GO and GOslim, for all four clusters generated from the principal component analysis with self-organizing map for the height factor.

Notes S3 Enriched GO-categories, both all GO and GOslim, for all four clusters generated from the principal component analysis with self-organizing map for the leaflets factor.

Notes S4 Enriched GO-categories, both all GO and GOslim, for all four clusters generated from the principal component analysis with self-organizing map for the light availability factor.

Notes S5 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* height classes 1 and 2.

Notes S6 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* height classes 1 and 3.

Notes S7 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* height classes 1 and 4.

Notes S8 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* height classes 2 and 3.

Notes S9 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* height classes 2 and 4.

Notes S10 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* height classes 3 and 4.

Notes S11 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* leaflet classes 1 and 2.

Notes S12 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* leaflet classes 1 and 3.

Notes S13 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* leaflet classes 2 and 3.

Notes S14 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* light availability classes deep shade and shade.

Notes S15 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* light availability classes deep shade and semi-shade.

Notes S16 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* light availability classes deep shade and sun.

Notes S17 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* light availability classes semi shade and shade.

Notes S18 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* light availability classes semi-shade and sun.

Notes S19 Differentially expressed transcripts ($\log FC \geq 1$; $FDR \leq 0.05$) for pairwise comparison for *G. avellana* light availability classes shade and sun.

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.