



# Genome-Wide Characterization and Expression Profiling of the *AUXIN RESPONSE FACTOR (ARF)* Gene Family in *Eucalyptus grandis*

Hong Yu<sup>1</sup>, Marçal Soler<sup>1</sup>, Isabelle Mila<sup>2</sup>, H el ene San Clemente<sup>1</sup>, Bruno Savelli<sup>1</sup>, Christophe Dunand<sup>1</sup>, Jorge A. P. Paiva<sup>3,4</sup>, Alexander A. Myburg<sup>5</sup>, Mondher Bouzayen<sup>2</sup>, Jacqueline Grima-Pettenati<sup>1\*</sup>, Hua Cassan-Wang<sup>1</sup>

**1** LRSV Laboratoire de Recherche en Sciences V eg etales, UMR5546, Universit e Toulouse III/CNRS, Castanet Tolosan, France, **2** Universit e de Toulouse, Institut National Polytechnique-Ecole Nationale Sup erieure Agronomique de Toulouse, Laboratoire de G enomique et Biotechnologie des Fruits, Castanet-Tolosan, France, **3** Instituto de Investigac o Cient fica e Tropical (IICT/MCTES), Lisboa, Portugal, **4** Instituto de Biologia Experimental e Tecnol gica (IBET), Oeiras, Portugal, **5** Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute (GRI), University of Pretoria, Pretoria, South Africa

## Abstract

Auxin is a central hormone involved in a wide range of developmental processes including the specification of vascular stem cells. Auxin Response Factors (ARF) are important actors of the auxin signalling pathway, regulating the transcription of auxin-responsive genes through direct binding to their promoters. The recent availability of the *Eucalyptus grandis* genome sequence allowed us to examine the characteristics and evolutionary history of this gene family in a woody plant of high economic importance. With 17 members, the *E. grandis* ARF gene family is slightly contracted, as compared to those of most angiosperms studied hitherto, lacking traces of duplication events. *In silico* analysis of alternative transcripts and gene truncation suggested that these two mechanisms were preeminent in shaping the functional diversity of the ARF family in *Eucalyptus*. Comparative phylogenetic analyses with genomes of other taxonomic lineages revealed the presence of a new ARF clade found preferentially in woody and/or perennial plants. High-throughput expression profiling among different organs and tissues and in response to environmental cues highlighted genes expressed in vascular cambium and/or developing xylem, responding dynamically to various environmental stimuli. Finally, this study allowed identification of three ARF candidates potentially involved in the auxin-regulated transcriptional program underlying wood formation.

**Citation:** Yu H, Soler M, Mila I, San Clemente H, Savelli B, et al. (2014) Genome-Wide Characterization and Expression Profiling of the *AUXIN RESPONSE FACTOR (ARF)* Gene Family in *Eucalyptus grandis*. PLoS ONE 9(9): e108906. doi:10.1371/journal.pone.0108906

**Editor:** Marcel Quint, Leibniz Institute of Plant Biochemistry, Germany

**Received:** July 2, 2014; **Accepted:** August 27, 2014; **Published:** September 30, 2014

**Copyright:**   2014 Yu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. All *Eucalyptus* ARF sequences used in this paper are available in phytozome <http://www.phytozome.net/>. *E. globulus* RNA Seq Illumina reads are provided in File S1.

**Funding:** This work was supported by the Centre National pour la Recherche Scientifique (CNRS), the University Paul Sabatier Toulouse III (UPS), the French Laboratory of Excellence project "TULIP" (ANR-10-LABX-41; ANR-11-IDEX-0002-02) and the Plant KBBE TreeForJoules project (ANR-2010-KBBE-007-01 (FR) and P-KBBE/AGR\_GPL/0001/2010 (FCT, PT)). HY was supported by PhD grants from the China Scholarship Council. MS was supported by a Postdoc fellowship "Beatriu de Pin s" thanks to the Departament d'Universitats, Recerca i Societat de la Informac o de la Generalitat de Catalunya. JAPP acknowledges Fundac o para a Ci ncia e a Tecnologia (FCT, Portugal) and his research contract funded by Programa Ci ncia 2008 and POPH (QREN). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors declare that no competing interests exist.

\* Email: [yu@lrsv.ups-tlse.fr](mailto:yu@lrsv.ups-tlse.fr)

## Introduction

The plant hormone auxin plays a prominent role in the regulation of plant growth in response to diverse developmental and environmental cues such as organogenesis, tropic movement, root growth, fruit development, tissue and organ patterning and vascular development [1]. Auxin plays a crucial role in the specification of vascular stem cells (procambium) and in cambial activity [2]. Analysis of auxin distribution across the cambial region in hybrid aspen trees showed a radial auxin gradient reaching a peak level in the cambial zone or at the border between the cambial zone and the expansion zone towards developing wood cells [3,4]. The auxin gradient was indeed shown to overlap with the sequential and numerous auxin-regulated genes responding dynamically to the change in auxin levels in wood forming cells [5].

As trees are long living organisms with sessile lifestyle, they have to adapt to changing environmental conditions throughout their lifetimes which may span decades and centuries in some cases. In particular, vascular stem cell activity shows plasticity in response to mechanical stress which affects wood formation and quality. In angiosperm woody species, a local increase in cambial cell division induces the formation of tension wood in the upper side of the leaning tree stems [6,7]. Auxin has been proposed to be implicated in the tension response, and application of either exogenous auxin or auxin transport inhibitors was shown to induce the gelatinous G-fibres characteristics of tension wood [8]. Although measurements of endogenous auxin failed to reveal significant changes in auxin balance in the cambial region tissues, a rather large set of auxin-related genes were found to be differentially expressed in developing poplar tension wood [9]. A recent study indicated that the auxin signalling pathway is significantly disrupted during

cambial dormancy in hybrid aspen [10]. Despite the fact that auxin has long been proposed as primary regulator of cambial activity and wood formation [4,11], the auxin-regulated transcriptional programs underlying wood formation remain largely under investigated.

Auxin exerts its function through modulating the expression of numerous genes among which is a set of transcriptional regulators. Auxin Response Factors (ARFs) and Aux/IAAs are two well-known mediators which regulate auxin responsive gene expression [12,13]. Most ARF proteins contain a highly conserved N-terminal B3-like DNA binding domain that recognizes an auxin-response element (AuxRE: TGTCTC) present in the promoters of auxin-responsive genes. The C-terminal domain contains two motifs, called III and IV, also found in Aux/IAA proteins and shown to enable the formation of homo- and heterodimers among ARFs and Aux/IAAs [14,15]. The middle region whose sequence is less conserved confers transcription activation or repression depending on its amino acid composition [13]. Biochemical and genetic studies in *Arabidopsis* and other plants have led to a working model of the mediation of auxin response by ARF proteins [14,16]. In the absence of auxin, Aux/IAAs bind to ARFs and recruit co-repressors of the TOPLESS (TPL) family, preventing the ARFs from regulating target genes [17]. The presence of auxin induces Aux/IAA protein degradation via the 26S proteasome through SCF-TIR1 ubiquitin ligase complex; thus liberating the trapped ARF proteins, allowing them to modulate the transcription of target auxin-responsive genes (for review, see Guilfoyle and Hagen) [12]. This model based on limited ARF-Aux/IAA interaction studies which provides a framework for understanding how members of these families may function. More recently, a large-scale analysis of the Aux/IAA-ARF interactions in the shoot apex of *Arabidopsis* showed that the vast majority of Aux/IAAs interact with all ARF activators, suggesting that most Aux/IAAs may repress the transcriptional activity of ARF activators [18]. In contrast, Aux/IAAs have limited interactions with ARF repressors suggesting that the role of the latter is essentially auxin-independent and that they might simply compete with the ARF activators for binding to the promoter of auxin-inducible genes [18]. This finding is particularly important taking into account that auxin predominantly activates transcription [19–21] and that a large complement of the ARF family acts as transcriptional repressors [12]. Whereas the above proposed scenario applies to the shoot apical meristem, it is likely that specific interactions between Aux/IAAs and ARFs might also affect the dynamics of the ARF-Aux/IAA signalling pathway in other developmental processes such as cambial development.

The ARF gene family has been most extensively studied in *Arabidopsis* where phenotyping of mutants revealed involvement of specific ARF genes in various plant growth and development processes [20,22–30]. For instance, ARF5/MONOPTEROS (MP) is a transcriptional activator known to play a critical role in the specification of vascular stem cells [27,31].

The ARF family has also been characterized in several annual herbaceous plants including monocots (rice, maize) [32,33] and dicots (*Arabidopsis*, tomato, soybean, *Brassica rapa*) [24,34–37] and in only two woody perennial genera, *Populus* [38] and *Vitis* [39]. However, so far, no ARF candidate has been identified as specifically involved in vascular cambium activity and xylem differentiation.

The recent availability of *Eucalyptus grandis* genome [40], the second hardwood forest tree genome fully sequenced, offers new opportunities to get insights into the regulation of secondary growth and cambial activity by ARFs, especially because *Eucalyptus* belongs to evergreen trees that do not present

dormancy in their cambial activity in sharp contrast with deciduous trees like *Populus*. *Eucalyptus* is also the most planted hardwood in the world, mainly for pulp and paper production but is also foreseen as a dedicated energy crop for lignocellulosic biofuel production. Thus, understanding the mechanisms that underlying auxin regulation in *Eucalyptus* wood formation is of interest both in the context of plant development and as a path to improve lignocellulosic biomass production and quality.

In the present paper, we report a genome-wide identification and characterization of the ARF family in *Eucalyptus grandis*. We analyzed gene structure, protein motif architecture, and chromosomal location of the members of the *E. grandis* ARF family. We also performed comparative phylogenetic relationships and large scale transcript profiling with a special focus on vascular tissues to get insights in their evolution, expression characteristics and possible functions.

## Materials and Methods

### Identification of ARF gene family in *Eucalyptus grandis* and chromosomal location

The identification procedure is illustrated in Fig. S1. Firstly we used *Arabidopsis* ARF proteins as queries in BLASTP searches for predicted protein in *Eucalyptus* genome (JGI assembly v1.0, annotation v1.1, <http://www.phytozome.net/eucalyptus>). A total of 64 *Eucalyptus* proteins identified in this initial search were examined by manual curation of protein motif scan using Pfam domain IDs (<http://pfam.wustl.edu>) and NCBI conserved domain database (<http://www.ncbi.nlm.nih.gov/cdd>). Redundant and invalid gene models were eliminated based on gene structure, intactness of conserved motifs and EST support. Three incomplete gene models were identified and completed by FGeneSH (<http://linux1.softberry.com>). To complete partial sequence of Eucgr.K02197.1, we cloned the corresponding genomic fragment using forward primer: 5'-AATTGACCGCGGTTGGATA-3' and reverse primer 5'-GAGCAGGCCAACATCCTCA-3', which located up-stream and down-stream respectively of the non-determined sequence (N). According to sequencing result we complete the missing part (1156 bp), corresponding to a part of promoter region and a part of 5' end CDS of the Eucgr.K02197.1 (submitted to GenBank data library under the accession number KC480258). All these manual curations enabled us to obtain 17 complete *Eucalyptus* ARF proteins sequences. We then used them as query in two subsequent additional searches: 1) BLASTP against *Eucalyptus* proteome for exhaustive identification of divergent *Eucalyptus* gene family members, and 2) tBLASTn searches against *Eucalyptus* genome for seeking any possible non-predicted genes. For validation, we also used poplar ARF proteins as queries to do the search procedure described above, and we obtained exactly the same result.

In the course of the above identification process we completed and expertly re-annotated three partial sequences (accession numbers Eucgr.F02090.1, Eucgr.F04380.1, and Eucgr.K03433.1 in the Phytozome database) initially annotated in the *Eucalyptus* genome-sequencing project (Table 1). In addition, we found one gene (accession number Eucgr.K02197.1) that corresponded to a partial sequence for which the 5' end was not determined (1240 N as sequencing results). Information on chromosomal location was retrieved from the *Eucalyptus* genome browser (<http://www.phytozome.net/eucalyptus>). *EgrARF* genes were mapped to their loci using MapChart 2.2 [41].

**Table 1.** ARF gene family in *Eucalyptus*.

| Accession no. <sup>a</sup> | Arabidopsis best hit (blastp) <sup>b</sup> | Amino acid identity % <sup>c</sup> | Short name <sup>d</sup> | Number of predicted alternative transcript <sup>e</sup> | Chromosome <sup>f</sup> | Genome location <sup>g</sup> | ORF (bp) <sup>h</sup> | Deduced polypeptide <sup>i</sup> Length (aa) | MW (kDa) | PI   | Exon No. |
|----------------------------|--|------------------------------------|-------------------------|---|-------------------------|------------------------------|-----------------------|--|----------|------|----------|
|                            |  |                                    |                         |   |                         |                              |                       |  |          |      |          |
| <i>Eucgr.G00076.1</i>      | <i>AtARF1</i>                              | 62.2                               | <i>EgrARF1</i>          | 4   | 7                       | 848,248..860,087             | 2028                  | 675  | 75.3     | 6.06 | 14       |
| <i>Eucgr.K02197.1</i>      | <i>AtARF2</i>                              | 59.3                               | <i>EgrARF2A</i> **      | 5   | 11                      | 29,232,770..29,239,103       | 2508                  | 835  | 93.13    | 6.55 | 14       |
| <i>Eucgr.803551.1</i>      | <i>AtARF2</i>                              | 49.5                               | <i>EgrARF2B</i>         | 2   | 2                       | 60,136,697..60,143,120       | 2364                  | 787  | 88.09    | 6.62 | 14       |
| <i>Eucgr.D00588.1</i>      | <i>AtARF3</i>                              | 46.2                               | <i>EgrARF3</i>          | 2   | 4                       | 10,835,264..10,841,977       | 2172                  | 723  | 79.19    | 6.36 | 10       |
| <i>Eucgr.802480.1</i>      | <i>AtARF4</i>                              | 57.8                               | <i>EgrARF4</i>          | 3   | 2                       | 46,950,723..46,956,747       | 2394                  | 797  | 88.53    | 6.31 | 12       |
| <i>Eucgr.F02090.1</i>      | <i>AtARF5</i>                              | 59.3                               | <i>EgrARF5*</i>         | 1   | 6                       | 28,226,789..28,233,884       | 2835                  | 944  | 103.53   | 5.43 | 14       |
| <i>Eucgr.D00264.1</i>      | <i>AtARF6</i>                              | 69.1                               | <i>EgrARF6A</i>         | 2   | 4                       | 4,224,849..4,232,295         | 2694                  | 897  | 99.04    | 6.04 | 14       |
| <i>Eucgr.A02065.1</i>      | <i>AtARF6</i>                              | 65.6                               | <i>EgrARF6B</i>         | 3   | 1                       | 31,463,716..31,473,722       | 2613                  | 870  | 96.64    | 6.06 | 14       |
| <i>Eucgr.D01764.1</i>      | <i>AtARF9</i>                              | 58.6                               | <i>EgrARF9A</i>         | 1   | 4                       | 31,589,333..31,593,948       | 1971                  | 656  | 73.53    | 6.15 | 14       |
| <i>Eucgr.E00888.1</i>      | <i>AtARF9</i>                              | 58.4                               | <i>EgrARF9B</i>         | 2   | 5                       | 9,303,573..9,308,209         | 2064                  | 687  | 76.46    | 6.09 | 14       |
| <i>Eucgr.J00923.1</i>      | <i>AtARF10</i>                             | 60.1                               | <i>EgrARF10</i>         | 1   | 10                      | 10,070,677..10,074,840       | 2139                  | 712  | 77.68    | 8.32 | 4        |
| <i>Eucgr.G02838.1</i>      | <i>AtARF16</i>                             | 58                                 | <i>EgrARF16A</i>        | 1   | 7                       | 46,569,603..46,573,031       | 2079                  | 692  | 76.33    | 6.80 | 3        |
| <i>Eucgr.K01240.1</i>      | <i>AtARF16</i>                             | 52.5                               | <i>EgrARF16B</i>        | 1   | 11                      | 15,584,770..15,588,716       | 2124                  | 707  | 77.91    | 6.91 | 3        |
| <i>Eucgr.F04380.1</i>      | <i>AtARF17</i>                             | 42.1                               | <i>EgrARF17*</i>        | 1   | 6                       | 52,730,940..52,733,856       | 1881                  | 626  | 67.75    | 7.54 | 3        |
| <i>Eucgr.C03293.1</i>      | <i>AtARF19</i>                             | 60.4                               | <i>EgrARF19A</i>        | 2   | 3                       | 62,460,408..62,469,132       | 3360                  | 1119   | 124.93   | 6.30 | 14       |
| <i>Eucgr.C02178.1</i>      | <i>AtARF19</i>                             | 44.8                               | <i>EgrARF19B</i>        | 2   | 3                       | 39,505,080..39,513,066       | 3360                  | 1119   | 123.24   | 5.98 | 14       |
| <i>Eucgr.K03433.1</i>      | -  | -                                  | <i>EgrARF24*</i>        | 1   | 11                      | 43,352,151..43,357,873       | 1836                  | 611  | 68.44    | 7.24 | 14       |

\*Using FGeneSH to complete the complete sequence.

\*\*using specific primers to amplify the genomic DNA to complete the sequence.

<sup>a</sup>Gene model of *Eucalyptus* (version 1.1) in phytozome V8.0.<sup>b</sup>The best hit of *EgrARF* in Arabidopsis by using blastp.<sup>c</sup>The amino acid identity percentage between *EgrARF* and corresponding *AtARF*.<sup>d</sup>Designation related to Arabidopsis best hit.<sup>e</sup>The number of predicted alternative transcripts of *EgrARF* in phytozome.<sup>f</sup>Location of the *EgrARF* genes in the Chromosome.<sup>g</sup>Length of open reading frame in base pairs.<sup>h</sup>The number of amino acids, molecular weight (kilodaltons), and isoelectric point (pI) of the deduced polypeptides.<sup>i</sup>doi:10.1371/journal.pone.0108906.t001

## Sequence, phylogenetic, gene structure analysis

Conserved protein motifs were determined by Pfam [42]. Multiple protein sequences alignment was performed using Clustal X program (Version 2.0.11). Using full length sequences of all predicted protein, phylogenetic trees were constructed with MEGA5 program by neighbor-joining method with 1000 bootstrap replicates. Their exon-intron structures were extracted from Phytozome (<http://www.phytozome.net/eucalyptus>) and visualized in Fancy Gene V1.4 (<http://bio.ieu.eu/fancygene/>). The prediction of small RNA target sites in *EgrARF* genes was performed through the web application psRNATarget (<http://plantgrn.noble.org/psRNATarget/>). The stem-loop structures were predicted using RNAfold web server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) and visualized by RNAstructure 5.3 program.

## Plant material

The plant materials provenance and preparation are described in Cassan-Wang et al. [43]. Hormone treatments were performed in an *in vitro* culture system. 10  $\mu$ M NAA (1-Naphthaleneacetic acid, for auxin), or 20  $\mu$ M gibberellic acid or 100  $\mu$ M ACC (1-aminocyclopropane-1-carboxylic-acid, for ethylene) were added to the medium of 65-d-old young trees, and trees were sampled 14 days after treatments.

## Total RNA extraction, cDNA synthesis, quality controls and high throughput quantitative RT-PCR

All the procedures used for the qRT-PCR, from the RNA extraction to the calculation of transcript abundance are described in Cassan-Wang et al. [43]. Only samples with a RNA integrity number >7 (assessed by Agilent 2100 Bioanalyzer) were retained for reverse transcription. cDNA quality was assessed as described by Udvardi et al. [44] using housekeeping genes *IDH* and *PP2A3* (primers see Table S1). Primer pairs were designed using the software QuantPrime (<http://www.quantprime.de>) [45], showing in Table S1. qRT-PCR was performed by the Genotoul service in Toulouse (<http://genotoul.genotoul.fr/>) using the BioMark 96:96 Dynamic Array integrated fluidic circuits (Fluidigm Corporation, San Francisco, USA) described in Cassan-Wang et al. [43]. The specificity of the PCR products was confirmed by analysing melting curves. Only primers that produced a linear amplification and qPCR products with a single-peak melting curves were used for further analysis. The efficiency of each pair of primers was determined from the data of amplification Ct value plot with a serial dilution of mixture cDNA and the equation  $E = 10^{(-1/\text{slope})} - 1$ .  $E^{-\Delta\Delta C_t}$  method was used to calculate relative mRNA fold change compared to control sample using formula  $(E_{\text{target}})^{\Delta C_{t_{\text{target}}(\text{control-sample})}} / (E_{\text{reference}})^{\Delta C_{t_{\text{reference}}(\text{control-sample})}}$  [46] and five reference genes (*IDH*, *PP2A1*, *PP2A3*, *EF-1a* and *SAND*, Table S1) were used for data normalization. We chose *in vitro* plantlets as control sample, because it contains the main organs and tissues of our studies such as stem, leaves, shoot tips, xylem, phloem and cambium, and it is a relative stable and less variable sample as being grown under the same *in vitro* culture condition from one experiment to another.

## Transactivation analysis in single cell system

For testing the ability of ARF transcription factors to up or down regulate the expression of auxin responsive promoter DR5, the full-length cDNAs of the ARF transcription factors were cloned in pGreen vector under 35SCaMV promoter to create the effector constructs. The reporter constructs use a synthetic auxin-responsive promoter DR5 fused with the GFP reporter gene. Tobacco BY-2 protoplasts were co-transfected with the reporter

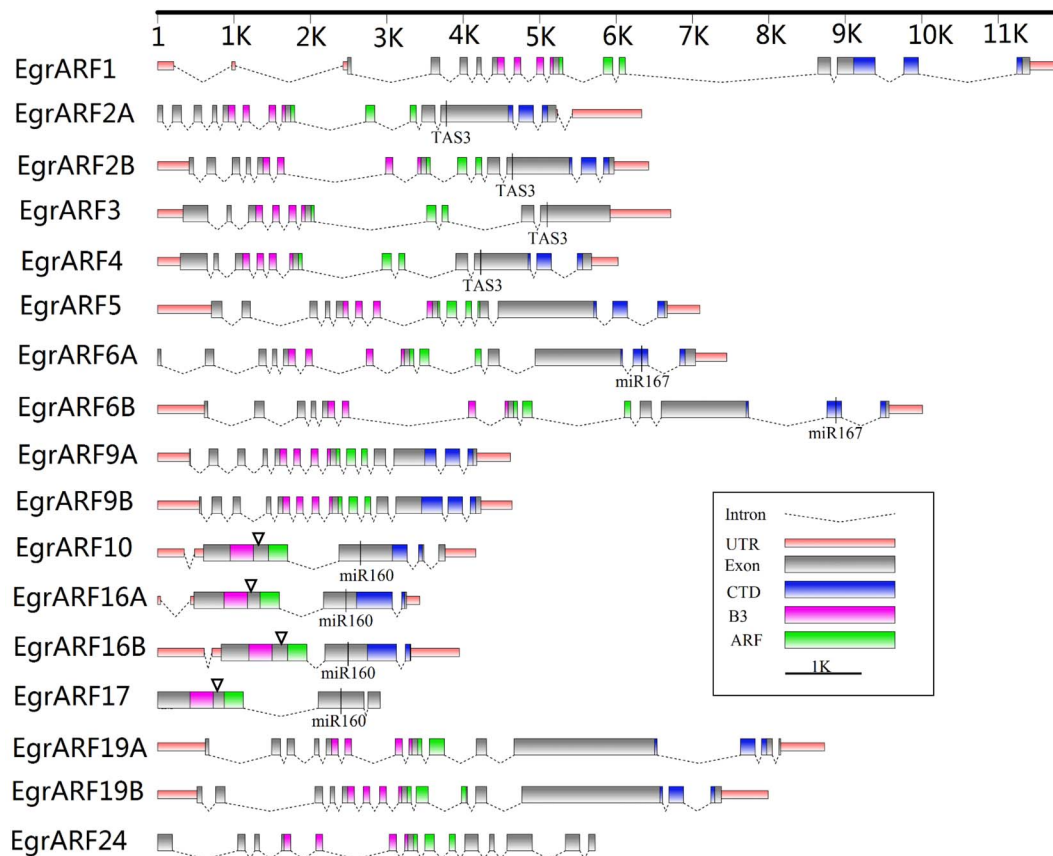
and effector constructs as described in Audran-Delalande et al. [47]. After 16 h incubation, GFP expression was quantified by flow cytometry (LSR Fortessa, BD Biosciences). Data were analysed using BD FACS Diva software. Transfection assays were performed in three independent replicates and 3000–4000 protoplasts were gated for each sample. GFP fluorescence corresponds to the average fluorescence intensity of the protoplasts population after subtraction of auto-fluorescence determined with non-transformed protoplasts. 50  $\mu$ M 2, 4-D was used for auxin treatment. We tested two independent protoplast preparations and for each of them, we performed in three independent transformation replicates. Similar results were obtained with the independent protoplast preparations and the data were represented by one of the preparations. For normalization, protoplasts were transformed with the reporter vector and the effector plasmid lacking the *ARF* gene.

## Results and Discussion

### Identification and chromosomal distribution of *Eucalyptus* ARF genes

The procedure to identify all members of the *ARF* family in the *E. grandis* genome (JGI assembly v1.0, annotation v1.1 (<http://www.phytozome.net/cgi-bin/gbrowse/eucalyptus/>), included expert manual curation as illustrated in Fig. S1. It allowed the identification of 17 genes encoding full length *Eucalyptus* ARF proteins (henceforth referred to as *EgrARF* genes). We named these genes according to their potential orthologs in *Arabidopsis* (Table 1). Where two *EgrARFs* matched the same potential *Arabidopsis* ortholog *AtARFx*, they were named as *EgrARFxA* and *xB*, with *xA* being the closest to the *Arabidopsis* ortholog; e.g. *EgrARF2A* and *EgrARF2B*. The percentage of identity between the *Arabidopsis* and the *Eucalyptus* predicted ARF protein sequences, and among the *Eucalyptus* ARFs themselves are given as Table S2 and S3, respectively. Eight *Arabidopsis* genes have no corresponding *Eucalyptus* orthologs (*AtARF12* to *15* & *20* to *23*), while only one *EgrARF* gene, *EgrARF24*, has no ortholog in *Arabidopsis* (Table 1). *In silico* chromosomal mapping of the gene loci revealed that the 17 *EgrARF* genes are scattered on nine of the eleven chromosomes, with one to three *EgrARF* genes per chromosome and with chromosomes 8 and 9 being devoid of *ARF* genes (Fig. S2).

The predicted proteins encoded by the *EgrARF* genes ranged from 593–1119 amino acid residues (Table 1), with PIs in the range of 5.43–8.32, suggesting that they can work in very different subcellular environments. Sequence analyses of the predicted proteins and Pfam protein motif analysis showed that most of them (14 of the 17 predicted proteins) harbour the typical ARF protein structure comprising a highly conserved DNA-binding domain (DBD) in the N-terminal region composed of a plant specific B3-type subdomain and an ARF subdomain, a variable middle region (MR) that functions as an activation or repression domain, and a carboxy-terminal dimerization (CTD) domain consisting of two highly conserved dimerization subdomains III and IV, similar to those found in Aux/IAAs (Fig. 1). We analysed and aligned the predicted amino acid sequences of the *EgrARFs* (Fig. 1 and Fig. S3). Four out of the 17 *EgrARFs* (*10*, *16A*, *16B* and *17*) exhibited an additional short segment of amino acids (between 15 to 43 amino-acids) in their DBD, between the B3 and ARF subdomains (Fig. 1 and Fig. S3). Such a feature has already been reported in *Arabidopsis* and soybean [36]. At the end of the DBD domain, all of the *EgrARFs* excepty *EgrARF6A*, *6B* and *19A* contain a conserved putative mono-partite nuclear localization signal (NLS) (Fig. S3) shown to direct the proteins into the nucleus [36,48].



**Figure 1. Gene structure of the *EgrARF* family.** The information on exon–intron structure was extracted from the Phytozome database and visualized by using the FancyGene software (<http://bio.ieu.eu/fancygene/>). The sizes of exons and introns are indicated by the scale at the top. The domains of *EgrARF* gene were predicted by Pfam (<http://pfam.xfam.org/>) and are indicated by different colours. The B3 together with ARF subdomains constitute the DNA binding domain (DBD). The CTD contains two sub-domains III and IV. The TAS3 and microRNA target sites are marked on the corresponding target genes. The triangles underline the insertion sites of additional short amino-acids segments between the B3 and ARF subdomains.

doi:10.1371/journal.pone.0108906.g001

Five *EgrARF*s (*EgrARF*5, 6A, 6B, 19A and 19B) harbour a glutamine (Q)-rich middle region (Fig. S3) implying that these proteins are likely transcriptional activators since glutamine enrichment seems to be a distinctive feature of ARF activators in all plant lineages [15,49]. The other 12 *EgrARF*s may function as repressors based on their middle regions enriched either in S (serine), SPL (Serine, Proline, Leucine) or SGL (Serine, Glycine, Leucine) [36] (Fig. S3).

The predicted protein structures of *EgrARF*3 and *EgrARF*17 are lacking dimerization domains III and IV like their potential orthologs in *Arabidopsis* (Fig. 1 and Fig. S3). *EgrARF*24, which has no ortholog in *Arabidopsis*, has a truncated CTD since only Aux/IAA subdomain III is present. The percentage of *EgrARF*s displaying a truncated CTD (17.6%) is similar to that in *Arabidopsis* (17.4%), but lower than in rapeseed (22.6%) or tomato (28.6%) [24,37,50]. These truncated *EgrARF*s are predicted to be unable to interact with Aux/IAA, a sequestration mechanism which may regulate their activity, and hence, they are likely to be insensitive to auxin. However, ARF repressors seem to display very limited interactions with Aux/IAA proteins [18], therefore the lack of domains III and/or IV could also have consequences for the interaction of ARFs with other transcriptional regulators [49].

Compared to *Arabidopsis*, the ARF family in *Eucalyptus* is slightly contracted with 17 versus 23 members. It is worth noting

that we found the exact same number of ARF genes in another *Eucalyptus* species, *E. camaldulensis* (<http://www.kazusa.or.jp/eucaly/>). Indeed when comparing to other species, in which the ARF family has been characterized (Table 2), *Eucalyptus* and grapevine appeared to have the smallest families with 17 and 19 members respectively, whereas poplar and soybean had the largest families with 39 and 51 members, respectively. We did not find evidence that any of the 17 *EgrARF* genes arose by tandem, segmental, or whole genome duplication, or even the more ancient hexaploidization in the *E. grandis* genome [40] and it appears that any such duplicates have been lost in *Eucalyptus* as is the case for 95% of whole-genome duplicates. This is sharply contrasting with the intensive tandem duplication events found for *Arabidopsis* ARF members [14,51], the segmental duplication found in *Populus* [38], and the whole-genome duplication events in soybean [36].

As duplication and alternative splicing are the two main mechanisms involved in diversification of function within gene families, sometimes viewed as opposite trends in gene family evolution, we performed an *in silico* survey of the alternative transcripts predicted in the *E. grandis* genome JGI assembly v1.0, annotation v1.1 (<http://www.phytozome.net/eucalyptus>), and compared them to those in *Arabidopsis* (Table 1 and Fig. S4). More than half of the *Eucalyptus* ARF family members (10 out of 17) have evidence of alternative splicing (Fig. S4). Taking into



**Table 2.** Summary of ARF gene content in angiosperm species.

| Species                     | ARFs content | Reference                          |
|-----------------------------|--------------|------------------------------------|
| <i>Eucalyptus grandis</i>   | 17           | This study                         |
| <i>Vitis vinifera</i>       | 19           | [39] Wan <i>et al.</i> (2014)      |
| <i>Solanum lycopersicon</i> | 22           | [35] Zouine <i>et al.</i> (2014)   |
| <i>Arabidopsis thaliana</i> | 23           | [24] Okushima <i>et al.</i> (2005) |
| <i>Oryza sativa</i>         | 25           | [33] Wang <i>et al.</i> (2007)     |
| <i>Zea mays</i>             | 31           | [32] Xing <i>et al.</i> (2011)     |
| <i>Brassica rapa</i>        | 31           | [37] Mun <i>et al.</i> (2012)      |
| <i>Populus trichocarpa</i>  | 39           | [38] Kalluri <i>et al.</i> (2007)  |
| <i>Glycine max</i>          | 51           | [36] Ha <i>et al.</i> (2013)       |

doi:10.1371/journal.pone.0108906.t002

account the number of possible alternative transcripts in *Eucalyptus* (17) and in *Arabidopsis* (15), the total number of possible transcripts in both species becomes very similar, 34 and 38, respectively. Some of the transcripts resulted in truncated versions of the genes like *EgrARF1.4*, *4.3* and *9B.2* lacking the Aux/IAA interaction domain and *EgrARF2B.2* lacking the B3/DBD domain. We further compared the *in silico* predicted ARF alternative transcripts from *E. grandis* to those expressed in a dataset of in-house RNA-Seq data from *E. globulus* (Table S4, Fig. S5, File S1). Remarkably, the vast majority of the alternative transcripts predicted in *E. grandis* were found expressed in *E. globulus* providing strong experimental support to their occurrence and conservation in the two *Eucalyptus* species. The importance of alternative splicing in the ARF family, has been highlighted recently by Finet *et al.* [49], who have shown that two *Arabidopsis* alternative transcripts of *AtARF4* have very different functions in flower development, and by Zouine *et al.* [35] who have shown that in tomato, one third of the ARF members displays alternative splicing as a mode of regulation during the flower to fruit transition. In *Arabidopsis* and in many other species, not only domain rearrangement through alternative splicing but also extensive gene duplication played a significant role in ARF functional diversification [49], whereas in *Eucalyptus* the first mechanism appeared to be preeminent.

### Comparative Phylogenetic analysis of the ARF family

To assess the relationship of *Eucalyptus* ARF family members to their potential orthologs in other landmark genomes, we constructed a comparative phylogenetic tree using all predicted ARF protein sequences from genomes of relevant taxonomic lineages. The core rosids were represented by *Arabidopsis* and *Populus* (Malvids) while the Myrtales, the Vitales and the Asterides were represented by *Eucalyptus grandis*, *Vitis vinifera* and *Solanum lycopersicum*, respectively. The monocots were represented by the *Oryza sativa* genome (Fig. S6). A simplified tree with only *Arabidopsis*, *Populus* and *Eucalyptus* (Fig. 2A) showed that ARFs are distributed into four major groups I, II, III, and IV. *Eucalyptus* (and also grapevine) which harbour the smallest number of ARF genes as compared to all other species (Table 2), have the fewest number of ARF proteins in each of the four groups. The positions and phases of the introns were well conserved within each group (Fig. 1 and Fig. 2), whereas their sizes were poorly conserved even within the same group. All five predicted *Eucalyptus* ARF transcriptional activators fell within group II as their potential orthologs from *Arabidopsis* and other

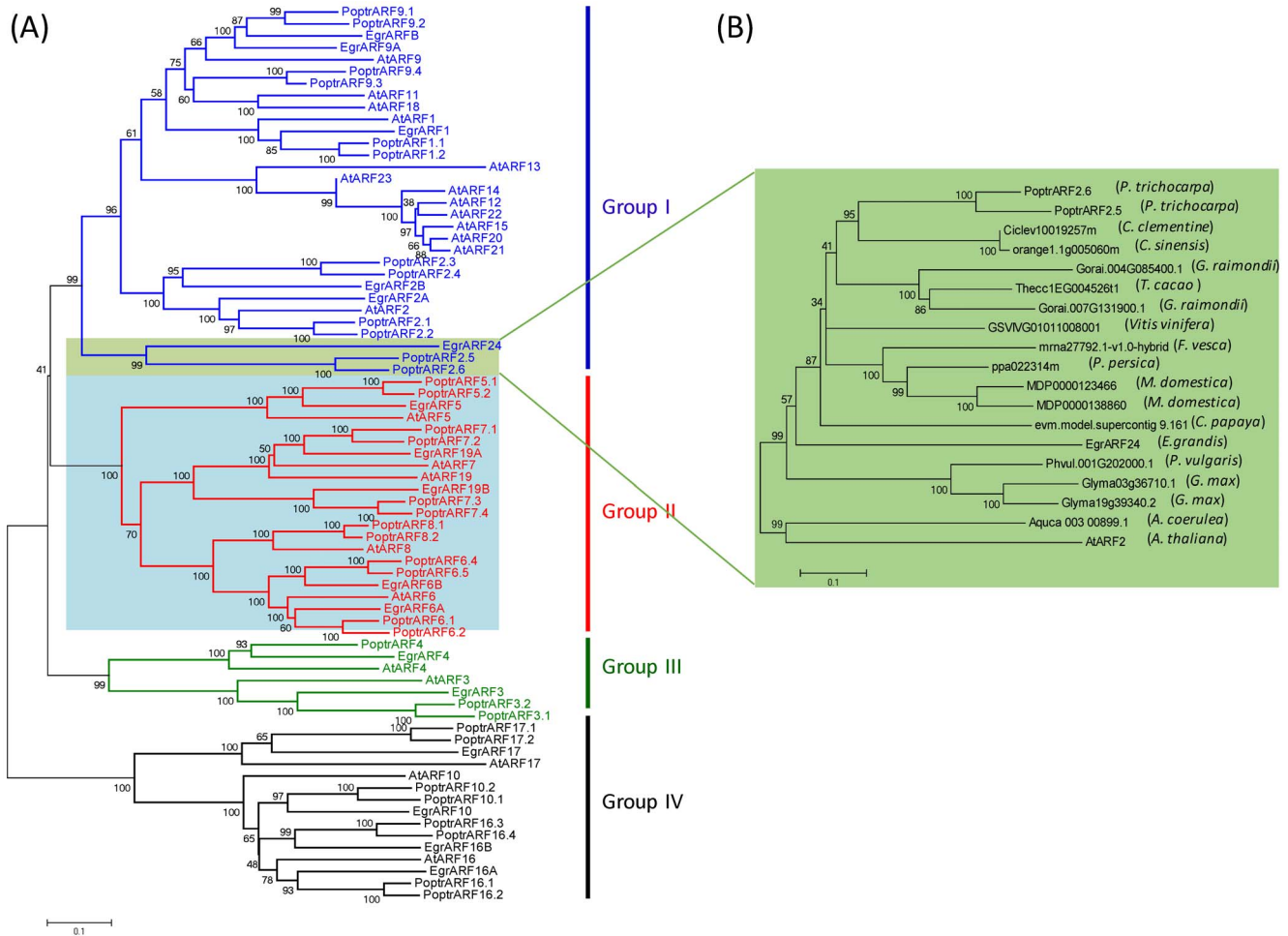
species; the remaining *Egr*ARFs were distributed among the three other groups.

Some lineage-specific clades were found in the *Solanaceae* ARF family [35] as well as in *Arabidopsis* ARF family [24]. In *Arabidopsis*, group I was substantially expanded with a subgroup containing seven tandem duplicated genes (encoding proteins AtARF12 to 15 and AtARF20, to 22), and the sister pair of AtARF11–AtARF18, for which orthologs were found only in Brassicaceae.

In group I, an isolated clade (highlighted in green in Fig. 2) contained *Egr*ARF24 clustering with *Ptr*ARF2.5 and *Ptr*ARF2.6 and did not contain any obvious *Arabidopsis* ortholog. This clade was absent from the herbaceous annual plants (*Arabidopsis*, tomato and rice), but present in woody perennial plants (*Eucalyptus*, *Populus* and *Vitis*; Fig. S6). To verify if this clade could be more specific to woody perennial plants, we performed a BLAST similarity search in 33 plant genomes available in Phytozome and found potential orthologs of *Egr*ARF24 in 13 plant species out of 33 (Table S5) which are presented in a phylogenetic tree (Fig. 2B). Among these 13 plant species, 11 are trees such as *M. domestica*, *C. sinensis*, *C. clementina*, *P. persica*, or tree-like plants and shrubs such as *C. papaya*, *T. cacao*, *G. raimondii*, although the latter is often grown as an annual plant. *A. coerulea* and *F. vesca* are perennial herbaceous plants. The two notable exceptions are two members of the Fabaceae family (*G. max*, and *P. vulgaris*) which are annual herbaceous plants. We thus considered this clade as woody-preferential. Regarding Group III, there was no evidence of large expansion of *ARF3* and *ARF4* genes in any of the three species, with only *ARF3* duplicated in *Populus*. Group IV contained four members from *Eucalyptus*, i.e. one more than in *Arabidopsis*. All of the *Egr*ARFs belonging to this group have in common an additional fragment (between 15 to 43 amino-acids residues) within their DBD (Fig. 1 and Fig. 2) and, noteworthy, alternative splicing was not detected for any of these genes in *Eucalyptus* and *Arabidopsis* (Fig. S4).

### Prediction of small regulatory RNAs and their potential ARF targets

In *Arabidopsis*, several ARF genes are targets of microRNAs miR160 and miR167, or of a trans-acting short interfering RNA (tasiRNA) *TAS3* [29,52–54]. Since these small RNAs and their targets are very often conserved across plant species [32,55,56], we searched for their potential orthologs in the *Eucalyptus* genome. Their chromosomal locations, genomic sequences and the sequences of their mature forms are presented in Table S6. We identified three potential *Eucalyptus* *miR160* loci and three



**Figure 2. Phylogenetic relationships of ARF proteins between *Eucalyptus* and other species.** (A) Phylogenetic relationships between ARF proteins from *Arabidopsis*, *Populus* and *Eucalyptus*. Full-length protein sequences were aligned by using the Clustal\_X program. The phylogenetic tree was constructed by using the MEGA5 program and the neighbor-joining method with predicted ARF proteins. Bootstrap support is indicated at each node. The blue shade highlights the activators, and the green shade indicates the distinct likely woody preferential clade containing *EgrARF24*. (B) Phylogenetic relationships between the orthologs of *EgrARF24* in other species. *EgrARF24* proteins were used to blast 33 species genomes in Phytozome. An E-value of  $1.0E-50$  as used as a cut off to select the ARF potential orthologs from each species. A phylogenetic tree was constructed using the procedure as in (A) and using *AtARF2* was used as an outgroup. The species containing putative orthologs of *EgrARF24* were the followings: 1 *Aquilegia coerulea*, 2 *Glycine max*, 1 *Phaseolus vulgaris*, 1 *Carica papaya*, 2 *Malus domestica*, 1 *Prunus persica*, 1 *Fragaria vesca*, 1 *Vitis vinifera*, 2 *Populus trichocarpa*, 1 *Citrus sinensis*, 1 *Citrus clementine*, 2 *Gossypium raimondii*, 1 *Theobroma cacao*. doi:10.1371/journal.pone.0108906.g002

potential *miR167* loci, all predicted gene products formed typical microRNA stem-loop structures (Fig. S7). The three *EgrmiR160* genes encode a mature RNA identical to that in *Arabidopsis*. The three *miR167* genes produce two different mature RNA forms (Table S6) whereas in *Arabidopsis* three different mature *miR167* forms were detected. We also identified a potential *TAS3* locus in the *Eucalyptus* genome (Table S6).

We used these newly identified *Eucalyptus* small RNAs as probes to search *in silico* for their target sites in *EgrARF* genes. Ten of the 17 *EgrARF* genes were found to be potential targets of these three small RNAs (Table S7). We identified highly conserved target sites for *EgrmiR160* in *EgrARF10*, *16A*, *16B* and *17*, for *EgrmiR167* in *EgrARF6A* and *B*, and for *EgrTAS3* in *EgrARF2A*, *2B*, *3* and *4* (Table S7). The targeting of three different small RNA to their corresponding target genes was highly conserved between *Arabidopsis* and *Eucalyptus* suggesting common regulation of plant growth and development. For example, *miR160*, a highly conserved miRNA group across the plant

kingdom, is known to target *ARF10*, *ARF16* and *ARF17* to regulate various aspects of plant development [30,52,53]. In *Arabidopsis*, *miR167* regulates lateral root outgrowth [57], adventitious rooting [58], ovule and anther growth, flower maturation [20,29] and jasmonic acid homeostasis [59] by targeting both *AtARF6* and *AtARF8*. Very recently, it has also been shown that *miR167* regulates flower development and female sterility in tomato [60]. Because *Eucalyptus* is a woody perennial plant, one could expect that some small RNAs (for instance *miR160* and *Tasi 3*) could be involved in the regulation of wood formation through targeting of ARF genes preferentially expressed in cambial cells or developing xylem.

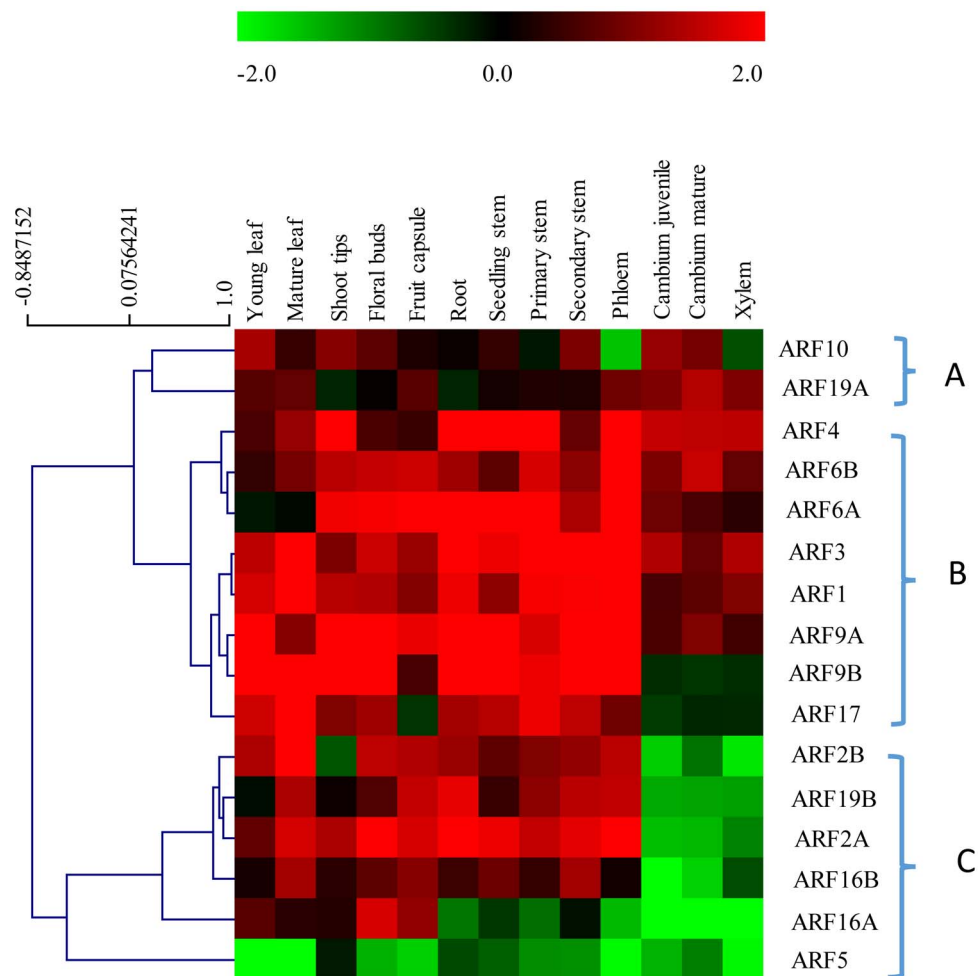
#### Expression of *EgrARFs* in different *Eucalyptus* organs and tissues and in response to environmental cues

To start investigating the functions of the *EgrARF* genes, we assessed their transcript expression levels in various *Eucalyptus* organs and tissues by qRT-PCR, with special attention to vascular

tissues (Fig. 3, Fig. S8 and Fig. S9). Transcript accumulation was detected for 16 *EgrARFs* in all 13 organs and tissues tested (Fig. 3), except for *EgrARF24*, which was detected only in shoot tips and young leaves (Fig. S8A). The very restricted expression profile of *EgrARF24* is surprising first because this gene belongs to a woody-preferential clade and second, because its poplar orthologs *PtrARF2.5* and *PtrARF2.6* could be detected in xylem based on microarray expression data [38], *PtrARF2.6* being highly expressed in developing wood (<http://popgenie.org/>). It should be noted however that this gene is truncated in *E. grandis*, it has lost domain III, whereas *PtrARF2.6* and their grapevine ortholog still have domain III and IV.

Heatmap representation (Fig. 3) indicated that *EgrARF* genes were expressed across various tissues and organs, but different members displayed preference to particular tissues and/or organs and could therefore be clustered into three main expression groups. Group A is the smallest with only two members *EgrARF10* (predicted repressor) and *EgrARF19A* (predicted activator) showing a relatively higher expression in vascular cambium as compared to other tissues and/or organs. *EgrARF10* was expressed at higher level in cambium (both mature and

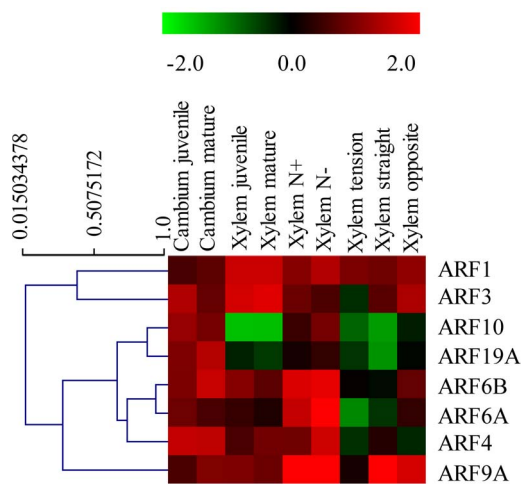
juvenile) than in differentiating xylem and/or phloem (Fig. 3 and Fig. S9). Its ortholog in *Populus*, *PtrARF10.1*, is highly expressed in developing xylem tissues [38], suggesting that *AtARF10* orthologs in trees might be involved in wood cell differentiation having a different/supplementary role as compared to that of the *Arabidopsis* sister pair *AtARF10* – *AtARF16* whose mutants exhibit root cap defects and abnormal root gravitropism [30]. *EgrARF19A* was expressed at similar levels in the three vascular tissues (Fig. 3 and Fig. S9). Group B is the largest with eight genes (*EgrARF4*, *6B*, *6A*, *3*, *1*, *9A*, *9B*, *17*) expressed in all tissues including vascular and non-vascular tissues (Fig. 3). The expression of *EgrARF3* and *EgrARF4* is highest in root, stem and phloem and differs from the specific expression of their *Arabidopsis* orthologs *AtARF3* and *AtARF4* associated with developing reproductive and vegetative tissues. This suggests that they might be involved in other processes than the control of the abaxial identity of the gynoecium, and lateral organs shown in *Arabidopsis* [26]. Group C includes six genes (*EgrARF2A*, *2B*, *5*, *16A*, *16B*, *19B*) preferentially expressed in leaves, floral buds and fruits and virtually absent from vascular tissues and particularly from cambium and xylem (Fig. 3 and Fig. S9). As its *Arabidopsis*



**Figure 3. Expression profiles of 16 *EgrARF* genes in various organs and tissues.** The heat map was constructed by using the relative expression values determined by qRT-PCR of 16 *EgrARF* genes (indicated on the right) in 13 tissues and organs (indicated at the top) normalized with a control sample (*in vitro* plantlets). In the heat map, red and green indicate relatively high and lower expression ( $\log_2$ ratios) than in the control, respectively. Each measurement is the mean of three independent samples. The heat map and the hierarchical clustering were generated by MultiExperiment Viewer (MEV).

doi:10.1371/journal.pone.0108906.g003



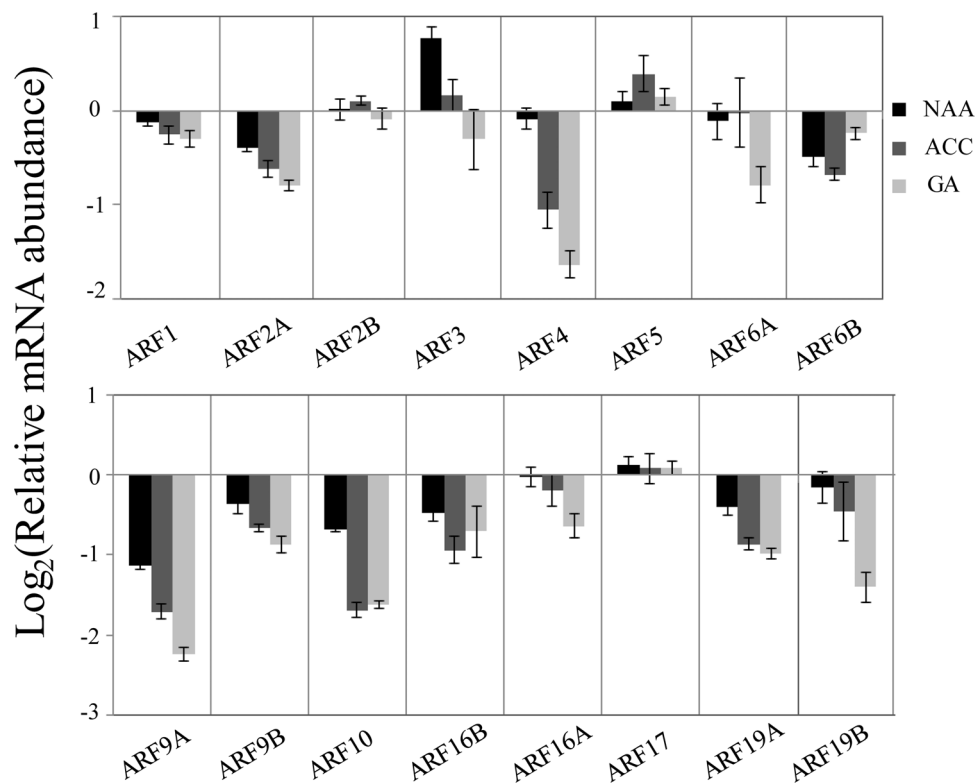


**Figure 4. Effect of environmental cues and developmental stages on *EgrARF* expression.** The heat map was constructed by using the relative expression values determined by qRT-PCR of *EgrARF* genes (indicated on the right) in various tissues and conditions (indicated at the top) normalized with a control sample (*in vitro* plantlets). In the heat map, red and green indicates relatively higher expression and lower expression ( $\log_2$ ratios) than in the control, respectively. The heat map and the hierarchical clustering were generated by MultiExperiment Viewer (MEV). doi:10.1371/journal.pone.0108906.g004

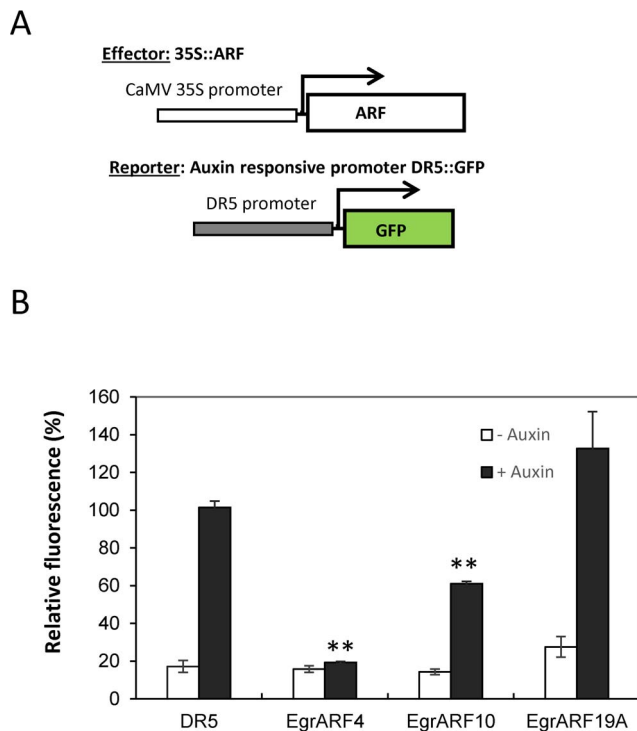
ortholog, *EgrARF19B* was highly expressed in root [24]. It should be noted that the activator *EgrARF5* is highly expressed in all samples analysed, with the highest expression in *in vitro* plantlets. Because *in vitro* plantlets were used to normalize the expression data in the heatmap, the expression of *EgrARF5* appeared in green in all other samples (Fig. 3). Its expression profile normalized using a different sample is given in Fig. S8.

Thirteen of the sixteen *EgrARF* genes examined (Fig. 3 and Fig. S9) exhibited higher expression in phloem than in xylem and/or cambium, suggesting that in *Eucalyptus* more *EgrARF* genes are involved in phloem than in xylem differentiation and/or function. *EgrARF5* was equally expressed in phloem and xylem. In *Arabidopsis*, *ARF5/MONOPTEROS (MP)* is known to play a critical role in the specification of vascular stem cells [27] but its role in secondary growth driven by vascular cambium activity has not been explored hitherto. *EgrARF10* and *EgrARF19A* were the only two genes more expressed in cambium and/or xylem than in other organs or tissues, supporting their possible involvement during the differentiation of meristematic cambium cells into xylem cells. No obvious difference in transcript levels were observed between juvenile and mature stages neither in cambium nor in differentiating xylem (Fig. 3 and Fig. 4).

We further examined the responsiveness to bending stress of the eight *EgrARF* genes which showed moderate to high expression in vascular tissues (Fig. 4). Half of *EgrARFs* were down-regulated in tension wood as compared to the control upright xylem, including three predicted repressors (*EgrARF3*, 4, and 9A) and one predicted activator (*EgrARF6A*). Conversely, in opposite xylem, four genes were up-regulated, including three predicted activators



**Figure 5. *EgrARF* genes expression levels in young tree stems after long term hormones treatments.** Hormone treatments are detailed in the Methods section. NAA, 1-naphthaleneacetic acid, a synthetic auxin usually used in *in vitro* culture. ACC, a precursor of ethylene biosynthesis. GA, gibberellic acid. Relative mRNA abundance was compared to expression in mock-treated young tree stems. Error bars indicated the SE of mean expression values from three independent experiments. doi:10.1371/journal.pone.0108906.g005



**Figure 6. *EgrARF* transcriptional activities in tobacco protoplasts.** (A) Schemes of the effector and reporter constructs used to analyse the function of *EgrARF*s in auxin-responsive gene expression. The effector constructs express the *EgrARF* of interest driven by the 35S promoter. The reporter construct consists of a reporter gene expressing GFP driven by the auxin-responsive promoter DR5 (DR5::GFP). (B) Effector and reporter constructs were co-expressed in tobacco protoplasts in the presence or absence of a synthetic auxin (50  $\mu$ M 2, 4-D). GFP fluorescence was quantified 16 h after transfection by flow cytometry. A mock effector construct (empty vector) was used as a control. In each experiment, protoplast transformations were performed in independent biological triplicates. Two independent experiments were performed and similar results were obtained; the figure indicates the data from one experiment. Error bars represent SE of mean fluorescence. Significant statistical differences (student T test,  $P < 0.001$ ) to control are marked with \*\*. doi:10.1371/journal.pone.0108906.g006

(*EgrARF6A*, *6B*, *19A*) and one repressor (*EgrARF10*). Only one gene (*EgrARF4*) was repressed. In general, *EgrARF* gene expression was repressed in tension xylem and induced in opposite xylem, except in the case of *EgrARF4*, which was down-regulated in both tension and opposite xylem (Fig. 4). These results are consistent with a study performed in *Populus* where seven *ARF* genes were detected in a poplar tension wood EST database, while the majority of genes were down-regulated in tension wood as compared to opposite wood [61].

Recent studies indicated that high nitrogen fertilization affects xylem development and alters fibre structure and composition in *Populus* [62,63] and induces some overlapping effects with tension wood on xylem cell walls. Interestingly, *EgrARF4* and *EgrARF6A* were down regulated in tension wood, but were down regulated when nitrogen was in excess (Fig. 4).

#### Effects of long-term hormone treatments on *EgrARF*s transcript levels

Several hormones are known to regulate cambium activity and xylem formation [[31], [64] and references therein]. For instance,

application of exogenous ethylene (ACC) on young poplar trees during 12 days was shown to stimulate cambial activity, while xylem cell size was decreased [65]. We performed similar long-term hormonal treatments (15 days) by growing young *Eucalyptus* trees on medium supplemented with either auxin, gibberellin or ethylene in order to evaluate the consequences on the transcripts levels of the *EgrARF* genes in stems (organs enriched in xylem). The phenotypes of the *Eucalyptus* trees after hormonal treatments were typical of each hormone: gibberellin stimulated plant growth resulting in longer stems, ethylene reduced plant growth and led to epinastic leaves, whereas auxin induced shortened and bolded roots (Fig. S10). All *EgrARF* transcripts except *EgrARF24* were detected in young tree stems and the expression levels of 13 were altered and mainly down-regulated by long-term hormonal treatments (Fig. 5). Although long-term hormonal treatments likely have both direct and indirect effects on *ARF*s expression, it is interesting to note distinct and differential behaviours: Five *ARF*s exhibited a kind of “hormonal preference” response since their transcripts levels were altered in stems treated only by one of the three hormones. For instance, *EgrARF3* was up-regulated only in auxin treated samples; *EgrARF5*, only in ethylene treated samples, whereas *EgrARF6A*, *EgrARF16A* and *EgrARF19B* were altered only in gibberellin treated samples. Most of the other *ARF*s were modulated at different degrees by the direct and/or indirect actions of each of three hormones with the notable exception of *EgrARF4* that was down-regulated in stems treated by ethylene and gibberellic acid but not affected in those treated by auxin.

#### Transcriptional activities of *EgrARF4*, *EgrARF 10* and *EgrARF19A*

We decided to characterize the transcriptional activity of three *ARF* members: *EgARF10* and *19A* which were preferentially expressed in cambium/xylem, and *EgARF4* whose expression was modulated in xylem in response to mechanical stress and to nitrogen fertilization. For this purpose, tobacco protoplasts were co-transfected with an effector construct expressing the full-length coding sequence of the *ARF*s under the *35SCaMV* promoter and a reporter construct carrying the auxin-responsive *DR5* promoter fused to *GFP* coding sequence (Fig. 6A). *DR5* is a synthetic auxin-responsive promoter made of nine inverted repeats of the conserved Auxin-Responsive Element, (TGTCTC box), fused to a *35SCaMV* minimal promoter. This reporter construct has been widely used to assess auxin responsive transcriptional activation or repression *in vivo* and *in planta* [15,47]. The *DR5*-driven GFP showed low basal activity which was induced up to 4-fold by exogenous auxin treatment (Fig. 6B). Co-transfection with the effector genes *EgrARF4* and *EgrARF10* resulted in a very significant ( $p < 0.001$ ) repression of auxin-induced reporter gene. Expression of 80% and 38%, respectively hereby confirming their predicted repressors roles. On the other hand, the values obtained for *EgrARF19A* suggested that it could be an activator as predicted by its sequence analysis, but this tendency was not strongly supported by the student-T test.

#### Conclusions

The *ARF* family in *E. grandis* contains 17 members (5 activators and 12 repressors) and is slightly contracted as compared to most angiosperm *ARF* families studied hitherto. In contrast to these species, it is characterized by the absence of whole genome, segmental and/or tandem duplication events. Indeed, whole genome duplication in *Eucalyptus* occurred 109.9 Mya ago, considerably earlier than those detected in other rosids

and 95% of the paralogs were lost [40]. The absence of tandem duplication is remarkable especially because *E. grandis* has the largest number of genes in tandem repeats (34% of the total number of genes) reported among sequenced plant genomes. Indeed, tandem duplication shaped functional diversity in many gene families in *Eucalyptus*. The ARF family thus evolved in a very different way. Our data suggests that genomic truncation and alternative splicing were preeminent mechanisms leading to the diversity of domain architecture, shaping and increasing the functional diversity of the ARF family in *Eucalyptus*, thereby compensating for the lack of extensive gene duplication found in other species. Comparative phylogenetic studies pointed out the presence of a new clade, maintained preferentially in woody and perennial plants. Finally, large scale expression profiling allowed identifying candidates potentially involved in the auxin-regulated transcriptional programs underlying wood formation.

## Supporting Information

**Figure S1 Procedure used for identifying ARF genes in *Eucalyptus grandis*.** *Arabidopsis* ARF protein sequences were used to search their orthologs in the predicted *Eucalyptus* proteome by using in BLASTP. Sixty-four *Eucalyptus* proteins identified in this initial search were further examined by manual curation using protein motif scanning and the FgeneSH program to complete partial sequences. Redundant and invalid genes were eliminated based on gene structure, integrity of conserved motifs and EST support. Manual curation resulted in 17 complete *Eucalyptus* ARF protein sequences. These 17 protein sequences were used in two subsequent additional searches: first, a BLASTP search against the *Eucalyptus* proteome to identify exhaustively all divergent *Eucalyptus* ARF gene family members and, second, tBLASTn searches against the *Eucalyptus* genome for any possible unpredicted genes. To confirm our findings, we used poplar ARF proteins and repeated the complete search procedure described above and obtained identical results.

(TIFF)

**Figure S2 Locations of the 17 *EgrARF* genes on the 11 *Eucalyptus grandis* chromosomes.**

(TIFF)

**Figure S3 Multiple sequence alignment of predicted amino acid sequences of *EgrARF* and *AtARF* proteins.** The multiple sequence alignment was obtained with the MUSCLE software [66]. The highly conserved domains and nuclear localization signals (NLSs) proteins were noted on the bottom of the alignment with different colours.

(PDF)

**Figure S4 Comparative analysis of predicted ARF alternative variants between *Eucalyptus grandis* and *Arabidopsis thaliana*.** The alternative spliced protein sequences were extracted from Phytozome except for *AtARF4* (obtained from Finet *et al.* (2013), the motif structures were predicted by Pfam (<http://pfam.xfam.org/>).

(PDF)

**Figure S5 Structure of the ARF alternative transcripts in *E. globulus*.** The *E. globulus* alternative transcripts were obtained from a compendium of RNASeq data. The material and methods are described in Table S4. The illumina reads sequences are provided in File S1 in the FastQ format.

(PDF)

**Figure S6 Comparative Phylogenetic relationships between ARF proteins from poplar, *Eucalyptus*, grapevine,**

***Arabidopsis*, tomato and rice.** Full-length protein sequences were aligned using the Clustal\_X program. The phylogenetic tree was constructed by using the MEGA5 program and the neighbour-joining method with predicted full-length ARF proteins. Bootstrap supports are indicated at each node.

(PDF)

**Figure S7 Predicted stem-loop structures of three *EgrmiR160* and three *EgrmiR167*.** The part of the stem-loop from which the mature microRNA derives is highlighted in yellow.

(TIFF)

**Figure S8 Expression profiles of *EgrARF5* and *EgrARF24* in various organs and tissues.** Relative mRNA abundance of *EgrARF5* and *EgrARF24* was compared to expression in the control sample of mature leaves and *in vitro* plantlets, respectively. Error bars indicate the SE of mean expression values from three independent experiments.

(TIFF)

**Figure S9 Expression profiles of *EgrARF* genes in tissues involved in secondary growth.** Relative mRNA abundance was compared to expression in the control sample (*in vitro* plantlets).

(TIFF)

**Figure S10 Young *Eucalyptus grandis* trees phenotypes in response to various long-term hormonal treatments.** 10  $\mu$ M NAA, or 20  $\mu$ M gibberellic acid or 100  $\mu$ M ACC were added to the medium of 65-d-old young tree, and phenotypes were observed 14 days later.

(TIFF)

**Table S1 Primers for *EgrARF* genes and reference genes used in qRT-PCR experiments.**

(PDF)

**Table S2 Protein identity matrix between *EgrARF* and *AtARF*.**

(PDF)

**Table S3 Protein identity matrix among *EgrARF*.**

(PDF)

**Table S4 Comparison of the number of alternative transcripts predicted in phytozome for *E. grandis* to those found in a large compendium of transcriptomic data from in *E. globulus*.**

(PDF)

**Table S5 Comparison of the number of *EgrARF24* putative orthologs in other species.**

(PDF)

**Table S6 Potential small RNAs targeting *EgrARF* genes.**

(PDF)

**Table S7 Small RNAs target site prediction in *EgrARF* genes.**

(PDF)

**File S1 Sequences of the Illumina reads from RNA Seq used to predict the *E. globulus* alternative transcripts.** The origin of the material and the procedure are described in Table S4.

(ZIP)

## Acknowledgments

The authors are grateful to J. P. Combiar for his help and advices for miRNA and target prediction, to E. Camargo (Unicamp, Brazil), J.M.

Gion and E. Villar (CIRAD, FR), F. Melun and L. Harvengt (FCBA, France), C. Araujo and L. Neves (AltriFlorestal, Portugal) and C. Marques (RAIZ, Portugal) for kindly providing and/or allowing collection of *Eucalyptus* organ and tissue samples, C. Graça, and V. Carocha for help with sample collection and RNA extraction, N Ladouce for performing Fluidigm experiments. Thanks to the Bioinfo Genotoul Platform (<http://bioinfo.genotoul.fr>) for access to bioinformatic resources and to the Genome and transcriptome Platform (<http://get.genotoul.fr/>) for advice and technical assistance with high-throughput Biomark Fluidigm qRT-PCR amplifications. Thanks to Y. Martinez (FR3450) for assistance with

microscopy analysis to C. Pecher and A. Zakaroff-Girard for technical assistance and expertise in flow cytometry (Toulouse RIO imaging platform <http://tri.ups-tlse.fr>).

## Author Contributions

Conceived and designed the experiments: HY HCW MS JGP. Performed the experiments: HY HSC CD IM BS. Analyzed the data: HY HCW MB MS JGP. Contributed reagents/materials/analysis tools: JAPP AAM. Wrote the paper: JGP YH HCW.

## References

- Woodward AW, Bartel B (2005) Auxin: regulation, action, and interaction. *Ann Bot* 95: 707–735.
- Miyashima S, Sebastian J, Lee JY and Helariutta Y (2013) Stem cell function during plant vascular development. *EMBO J* 32: 178–193.
- Tuominen H, Puech L, Fink S, Sundberg B (1997) A radial concentration gradient of indole-3-acetic acid is related to secondary xylem development in hybrid aspen. *Plant Physiol* 115: 577–585.
- Uggla C, Moritz T, Sandberg G, Sundberg B (1996) Auxin as a positional signal in pattern formation in plants. *Proc Natl Acad Sci U S A* 93: 9282–9286.
- Nilsson J, Karlberg A, Antti H, Lopez-Vernaza M, Mellerowicz E, et al. (2008) Dissecting the molecular basis of the regulation of wood formation by auxin in hybrid aspen. *Plant Cell* 20: 843–855.
- Pilate G, Déjardin A, Laurans F, Leplé JC (2004) Tension wood as a model for functional genomics of wood formation. *New Phytol* 164: 63–72.
- Paux E, Carocha V, Marques C, Mendes de Sousa A, Borralho N, et al. (2005) Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. *New Phytol* 167: 89–100.
- Timell TE (1969) The chemical composition of tension wood. *Svensk Papperstidn* 72: 173–181.
- Andersson-Gunneras S, Mellerowicz EJ, Love J, Segerman B, Ohmiya Y, et al. (2006) Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *Plant J* 45: 144–165.
- Baba K, Karlberg A, Schmidt J, Schrader J, Hvidsten TR, et al. (2011) Activity-dormancy transition in the cambial meristem involves stage-specific modulation of auxin response in hybrid aspen. *Proc Natl Acad Sci U S A* 108: 3418–3423.
- Sundberg B, Uggla C, Tuominen H (2000) Cambial growth and auxin gradients. Oxford: BIOS Scientific Publishers. 169–188 p.
- Guilfoyle TJ, Hagen G (2007) Auxin response factors. *Curr Opin Plant Biol* 10: 453–460.
- Tiwari SB, Hagen G, Guilfoyle T (2003) The roles of auxin response factor domains in auxin-responsive transcription. *Plant Cell* 15: 533–543.
- Hagen G, Guilfoyle T (2002) Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Mol Biol* 49: 373–385.
- Ulmasov T, Hagen G, Guilfoyle TJ (1999) Activation and repression of transcription by auxin-response factors. *Proc Natl Acad Sci U S A* 96: 5844–5849.
- Gray WM, Kepinski S, Rouse D, Leyser O and Estelle M (2001) Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. *Nature* 414: 271–276.
- Szemenyei H, Hannon M and Long JA (2008) TOPLESS mediates auxin-dependent transcriptional repression during *Arabidopsis* embryogenesis. *Science* 319: 1384–1386.
- Vernoux T, Brunoud G, Farcot E, Morin V, Van den Daele H, et al. (2011) The auxin signalling network translates dynamic input into robust patterning at the shoot apex. *Mol Syst Biol* 7: 508.
- Paponov I, Paponov M, Teale W, Menges M, Chakrabortee S, et al. (2008) Comprehensive transcriptome analysis of auxin responses in *Arabidopsis*. *Mol Plant* 1: 321–337.
- Nagpal P, Ellis CM, Weber H, Ploense SE, Barkawi LS, et al. (2005) Auxin response factors *ARF6* and *ARF8* promote jasmonic acid production and flower maturation. *Development* 132: 4107–4118.
- Pufky J, Qiu Y, Rao MV, Hurbat P, Jones AM (2003) The auxin-induced transcriptome for etiolated *Arabidopsis* seedlings using a structure/function approach. *Funct Integr Genomics* 3: 135–143.
- Ellis CM, Nagpal P, Young JC, Hagen G, Guilfoyle TJ, et al. (2005) *AUXIN RESPONSE FACTOR 1* and *AUXIN RESPONSE FACTOR 2* regulate senescence and floral organ abscission in *Arabidopsis thaliana*. *Development* 132: 4563–4574.
- Okushima Y, Mitina I, Quach HL, Theologis A (2005) *AUXIN RESPONSE FACTOR 2 (ARF2)* a pleiotropic developmental regulator. *Plant J* 43: 29–46.
- Okushima Y, Overvoorde PJ, Arima K, Alonso JM, Chan A, et al. (2005) Functional genomic analysis of the *AUXIN RESPONSE FACTOR* gene family members in *Arabidopsis thaliana*: unique and overlapping functions of *ARF7* and *ARF19*. *Plant Cell* 17: 444–463.
- Sessions A, Nemhauser JL, McColl A, Roe JL, Feldmann KA, et al. (1997) *ETTIN* patterns the *Arabidopsis* floral meristem and reproductive organs. *Development* 124: 4481–4491.
- Pekker I, Alvarez JP, Eshed Y (2005) Auxin response factors mediate *Arabidopsis* organ asymmetry via modulation of KANADI activity. *Plant Cell* 17: 2899–2910.
- Hardtke CS, Berleth T (1998) The *Arabidopsis* gene *MONOPTEROS* encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J* 17: 1405–1411.
- Harper RM, Stowe-Evans EL, Luesse DR, Muto H, Tatematsu K, et al. (2000) The *NPH4* locus encodes the auxin response factor *ARF7*, a conditional regulator of differential growth in aerial *Arabidopsis* tissue. *Plant Cell* 12: 757–770.
- Wu MF, Tian Q, Reed JW (2006) *Arabidopsis microRNA167* controls patterns of *ARF6* and *ARF8* expression, and regulates both female and male reproduction. *Development* 133: 4211–4218.
- Wang JW, Wang LJ, Mao YB, Cai WJ, Xue HW, et al. (2005) Control of root cap formation by MicroRNA-targeted auxin response factors in *Arabidopsis*. *Plant Cell* 17: 2204–2216.
- Miyashima S, Sebastian J, Lee JY, Helariutta Y (2013) Stem cell function during plant vascular development. *EMBO J* 32: 178–193.
- Xing H, Pudake RN, Guo G, Xing G, Hu Z, et al. (2011) Genome-wide identification and expression profiling of auxin response factor (*ARF*) gene family in maize. *BMC Genomics* 12: 178.
- Wang D, Pei K, Fu Y, Sun Z, Li S, et al. (2007) Genome-wide analysis of the auxin response factors (*ARF*) gene family in rice (*Oryza sativa*). *Gene* 394: 13–24.
- Kumar R, Tyagi AK, Sharma AK (2011) Genome-wide analysis of auxin response factor (*ARF*) gene family from tomato and analysis of their role in flower and fruit development. *Mol Genet Genomics* 285: 245–260.
- Zouine M, Fu Y, Chateigner-Boutin AL, Mila I, Frasse P, et al. (2014) Characterization of the tomato *ARF* gene family uncovers a multi-levels post-transcriptional regulation including alternative splicing. *PLoS one* 9: e84203.
- Ha CV, Le DT, Nishiyama R, Watanabe Y, Sulaiman S, et al. (2013) The auxin response factor transcription factor family in soybean: genome-wide identification and expression analyses during development and water stress. *DNA Res* 20: 511–524.
- Mun JH, Yu HJ, Shin JY, Oh M, Hwang HJ, et al. (2012) Auxin response factor gene family in *Brassica rapa*: genomic organization, divergence, expression, and evolution. *Mol Genet Genomics* 287: 765–784.
- Kalluri UC, Difazio SP, Brunner AM, Tuskan GA (2007) Genome-wide analysis of *Aux/IAA* and *ARF* gene families in *Populus trichocarpa*. *BMC Plant Biol* 7: 59.
- Wan S, Li W, Zhu Y, Liu Z, Huang W, et al. (2014) Genome-wide identification, characterization and expression analysis of the auxin response factor gene family in *Vitis vinifera*. *Plant Cell Rep* 33: 1365–1375.
- Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, et al. (2014) The genome of *Eucalyptus grandis*. *Nature* 510: 356–362.
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 93: 77–78.
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, et al. (2014) Pfam: the protein families database. *Nucleic Acids Res* 42: D222–230.
- Cassan-Wang H, Soler M, Yu H, Camargo EL, Carocha V, et al. (2012) Reference genes for high-throughput quantitative RT-PCR analysis of gene expression in organs and tissues of *Eucalyptus* grown in various environmental conditions. *Plant Cell Physiol* 53: 2101–2116.
- Udvardi MK, Czechowski T, Scheible WR (2008) Eleven golden rules of quantitative RT-PCR. *Plant Cell* 20: 1736–1737.
- Arvidsson S, Kwansiewski M, Riano-Pachon DM, Mueller-Roeber B (2008) QuantPrime: a flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinformatics* 9: 465.
- Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29: e45.
- Audran-Delalande C, Bassa C, Mila I, Regad F, Zouine M, et al. (2012) Genome-wide identification, functional analysis and expression profiling of the *Aux/IAA* gene family in tomato. *Plant Cell Physiol* 53: 659–672.
- Shen C, Wang S, Bai Y, Wu Y, Zhang S, et al. (2010) Functional analysis of the structural domain of ARF proteins in rice (*Oryza sativa* L.). *J Exp Bot* 61: 3971–3981.
- Finet C, Berne-Dedieu A, Scutt CP and Marletaz F (2013) Evolution of the *ARF* gene family in land plants: old domains, new tricks. *Mol Biol Evol* 30: 45–56.

50. Wu J, Wang F, Cheng L, Kong F, Peng Z, et al. (2011) Identification, isolation and expression analysis of *auxin response factor (ARF)* genes in *Solanum lycopersicum*. *Plant Cell Rep* 30: 2059–2073.
51. Remington DL, Vision TJ, Guilfoyle TJ, Reed JW (2004) Contrasting modes of diversification in the *Aux/IAA* and *ARF* gene families. *Plant Physiol* 135: 1738–1752.
52. Liu PP, Montgomery TA, Fahlgren N, Kasschau KD, Nonogaki H, et al. (2007) Repression of *AUXIN RESPONSE FACTOR10* by microRNA160 is critical for seed germination and post-germination stages. *Plant J* 52: 133–146.
53. Mallory AC, Bartel DP, Bartel B (2005) MicroRNA-directed regulation of *Arabidopsis AUXIN RESPONSE FACTOR17* is essential for proper development and modulates expression of early auxin response genes. *Plant Cell* 17: 1360–1375.
54. Williams L, Carles CC, Osmont KS, Fletcher JC (2005) A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the *Arabidopsis ARF2*, *ARF3*, and *ARF4* genes. *Proc Natl Acad Sci U S A* 102: 9703–9708.
55. Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA (2006) Conservation and divergence of plant microRNA genes. *Plant J* 46: 243–259.
56. Lu S, Sun YH, Chiang VL (2008) Stress-responsive microRNAs in *Populus*. *Plant J* 55: 131–151.
57. Gifford ML, Dean A, Gutierrez RA, Coruzzi GM, Birnbaum KD (2008) Cell-specific nitrogen responses mediate developmental plasticity. *Proc Natl Acad Sci U S A* 105: 803–808.
58. Gutierrez L, Bussell JD, Pacurar DI, Schwambach J, Pacurar M, et al. (2009) Phenotypic plasticity of adventitious rooting in *Arabidopsis* is controlled by complex regulation of *AUXIN RESPONSE FACTOR* transcripts and microRNA abundance. *Plant Cell* 21: 3119–3132.
59. Gutierrez L, Mongelard G, Flokova K, Pacurar DI, Novak O, et al. (2012) Auxin controls *Arabidopsis* adventitious root initiation by regulating jasmonic acid homeostasis. *Plant Cell* 24: 2515–2527.
60. Liu N, Wu S, Van Houten J, Wang Y, Ding B, et al. (2014) Down-regulation of *AUXIN RESPONSE FACTORS 6* and *8* by microRNA 167 leads to floral development defects and female sterility in tomato. *J Exp Bot* 65: 2507–2520.
61. Jin H, Do J, Moon D, Noh EW, Kim W, et al. (2011) EST analysis of functional genes associated with cell wall biosynthesis and modification in the secondary xylem of the yellow poplar (*Liriodendron tulipifera*) stem during early stage of tension wood formation. *Planta* 234: 959–977.
62. Pitre FE, Lafarguette F, Boyle B, Pavy N, Caron S, et al. (2010) High nitrogen fertilization and stem leaning have overlapping effects on wood formation in poplar but invoke largely distinct molecular pathways. *Tree Physiol* 30: 1273–1289.
63. Pitre FE, Pollet B, Lafarguette F, Cooke JE, MacKay JJ, et al. (2007) Effects of increased nitrogen supply on the lignification of poplar wood. *J Agri Food Chem* 55: 10306–10314.
64. Sorce C, Giovannelli A, Sebastiani L, Anfodillo T (2013) Hormonal signals involved in the regulation of cambial activity, xylogenesis and vessel patterning in trees. *Plant Cell Rep* 32: 885–898.
65. Love J, Björklund S, Vahala J, Hertzberg M, Kangasjärvi J, et al. (2009) Ethylene is an endogenous stimulator of cell division in the cambial meristem of *Populus*. *Proc Natl Acad Sci U S A* 106: 5984–5989.
66. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res* 32: 1792–1797.