

## Research article

## Open Access

### The evolution of nuclear auxin signalling

Ivan A Paponov<sup>\*1</sup>, William Teale<sup>1</sup>, Daniel Lang<sup>2</sup>, Martina Paponov<sup>1</sup>, Ralf Reski<sup>2,3</sup>, Stefan A Rensing<sup>3</sup> and Klaus Palme<sup>1,3</sup>

Address: <sup>1</sup>Botany, Faculty of Biology, University of Freiburg, Schänzlestrasse 1, 79104 Freiburg, Germany, <sup>2</sup>Plant Biotechnology, Faculty of Biology, University of Freiburg, Schänzlestrasse 1, 79104 Freiburg, Germany and <sup>3</sup>FRISYS, Faculty of Biology, University of Freiburg, Schänzlestrasse 1, 79104 Freiburg, Germany

Email: Ivan A Paponov<sup>\*</sup> - [ivan.paponov@biologie.uni-freiburg.de](mailto:ivan.paponov@biologie.uni-freiburg.de); William Teale - [william.teale@biologie.uni-freiburg.de](mailto:william.teale@biologie.uni-freiburg.de); Daniel Lang - [Daniel.Lang@biologie.uni-freiburg.de](mailto:Daniel.Lang@biologie.uni-freiburg.de); Martina Paponov - [martina.trebar@biologie.uni-freiburg.de](mailto:martina.trebar@biologie.uni-freiburg.de); Ralf Reski - [ralf.reski@biologie.uni-freiburg.de](mailto:ralf.reski@biologie.uni-freiburg.de); Stefan A Rensing - [stefan.rensing@biologie.uni-freiburg.de](mailto:stefan.rensing@biologie.uni-freiburg.de); Klaus Palme - [klaus.palme@biologie.uni-freiburg.de](mailto:klaus.palme@biologie.uni-freiburg.de)

<sup>\*</sup> Corresponding author

Published: 3 June 2009

Received: 15 September 2008

BMC Evolutionary Biology 2009, 9:126 doi:10.1186/1471-2148-9-126

Accepted: 3 June 2009

This article is available from: <http://www.biomedcentral.com/1471-2148/9/126>

© 2009 Paponov et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** The plant hormone auxin directs many aspects of plant growth and development. To understand the evolution of auxin signalling, we compared the genes encoding two families of crucial transcriptional regulators, *AUXIN RESPONSE FACTOR* (ARF) and *AUXIN/INDOLE-3-ACETIC ACID* (Aux/IAA), among flowering plants and two non-seed plants, *Physcomitrella patens* and *Selaginella moellendorffii*.

**Results:** Comparative analysis of the *P. patens*, *S. moellendorffii* and *Arabidopsis thaliana* genomes suggests that the well-established rapid transcriptional response to auxin of flowering plants, evolved in vascular plants after their divergence from the last common ancestor shared with mosses. An N-terminally truncated ARF transcriptional activator is encoded by the genomes of *P. patens* and *S. moellendorffii*, and suggests a supplementary mechanism of nuclear auxin signalling, absent in flowering plants. Site-specific analyses of positive Darwinian selection revealed relatively high rates of synonymous substitution in the *A. thaliana* ARFs of classes IIa (and their closest orthologous genes in poplar) and Ib, suggesting that neofunctionalization in important functional regions has driven the evolution of auxin signalling in flowering plants. Primary auxin responsive gene families (GH3, SAUR, LBD) show different phylogenetic profiles in *P. patens*, *S. moellendorffii* and flowering plants, highlighting genes for further study.

**Conclusion:** The genome of *P. patens* encodes all of the basic components necessary for a rapid auxin response. The spatial separation of the Q-rich activator domain and DNA-binding domain suggests an alternative mechanism of transcriptional control in *P. patens* distinct from the mechanism seen in flowering plants. Significantly, the genome of *S. moellendorffii* is predicted to encode proteins suitable for both methods of regulation.

## Background

The evolution of signal transduction pathways since the divergence of plants and animals has been influenced by

very different selection pressures. Hormone signalling, though analogous in both kingdoms, differs in the signalling molecules employed as well as in their perception

and mode of action. Plants are adapted to a sessile life-style, being able continuously to form new organs during their postembryonic development. This process, in addition to embryonic development, is closely associated with specific growth regulators, effective at low concentrations. The signalling pathways of these growth regulators (also known as phytohormones) are relatively well understood, but their evolution, as well as their relationship to the evolution of embryonic and post-embryonic development in the plant kingdom, is less clear [1].

Auxin, one such phytohormone, is a principal regulator of growth and development in flowering plants [2], quickly triggering the transcription of auxin-responsive genes [3,4]. Proteins of two related families, AUXIN RESPONSIVE FACTOR (ARF) and AUXIN/INDOLE-3-ACETIC ACID (Aux/IAA), act together to regulate this transcription [5,6]. In flowering plants, ARF proteins possess a conserved DNA-binding domain which recognizes auxin responsive elements (AuxREs): short motifs which are found in the promoter sequences of many auxin-responsive genes [7,8]. Most ARFs, and all Aux/IAAs also contain a conserved dimerization domain which mediates protein-protein interactions within and between both protein families [9,10]. The middle region which joins ARF DNA-binding and dimerization domains is highly divergent and may be glutamine (Q) rich [11]. Those ARFs which contain such Q-rich regions are thought to be activators of gene transcription [11,12]. Conversely, those ARFs which repress gene transcription lack glutamine (or in one case methionine) -rich regions.

The N-terminal region of Aux/IAA proteins contains two other domains: domain I and II. Domain I contains a short amphiphilic repression motif, which binds to the co-repressor TOPLESS, enabling Aux/IAAs to repress ARF function [13,14]. Domain II contains a degron: a motif sufficient to signal Aux/IAAs for proteasome-mediated degradation [6,15,16]. Specific point mutations in domain II confer strong, auxin insensitive phenotypes [17].

At low cellular auxin concentrations, Aux/IAA proteins dimerize with ARF transcriptional activators, repressing their activity [18]. Auxin itself can bind at the interface of Aux/IAA proteins and TIR1-family F-box proteins, components of specific SCF E3 ubiquitin ligases, directly promoting their interaction. Accordingly, at high cellular auxin concentrations, Aux/IAAs are ubiquitinated and subsequently degraded [19-21]. Degradation of Aux/IAA proteins then allows ARF-mediated, auxin-dependent gene transcription.

*Physcomitrella patens* (a moss), *Selaginella moellendorffii* (a vascular non-seed plant) and angiosperms diverged from

each other at between 700 and 450 million years ago [22]. The genomes of both *P. patens* and *S. moellendorffii* encode all the proteins necessary for this primary auxin response [23,24]. Furthermore, *P. patens* has been shown to both synthesize auxin, and respond to exogenously applied auxin [25,26]. Here we use the complete genomic sequences of *P. patens* and *S. moellendorffii* to address how a relatively simple signalling mechanism has evolved into, in flowering plants, a central regulator of many essential and diverse developmental processes. A driving force of this evolution has been positive Darwinian selection. Such positive selection is a measure of the adaptation of amino acid sequences following a gene duplication event. The unambiguous indicator of positive selection, a high ratio of non-synonymous (dN) to synonymous (dS) nucleotide substitutions, was detected in the flowering plant ARFs.

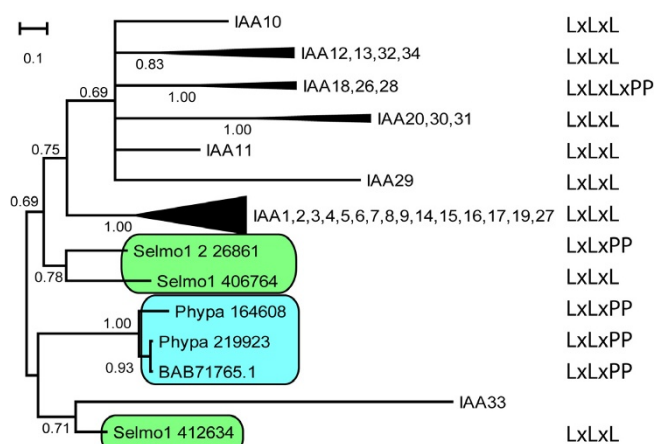
Based on a comparative analysis of the fully-sequenced genomes of *P. patens*, *S. moellendorffii*, and selected flowering plants, we are able to draw conclusions about ancestral auxin target genes and signalling mechanisms, and about the pressures which have driven the radiation of auxin-signalling genes in flowering plants.

## Results and discussion

Endogenous auxin is a widely used signalling molecule in vascular plants, but is also found in bryophytes, algae and prokaryotes [25]. In the present study, we identify similarities and differences between the auxin signalling components in moss and flowering plants by comparing the fully sequenced genomes of *P. patens* with those of model flowering plant species. Additional support, where appropriate, is drawn from the genome of *S. moellendorffii*, a vascular non-seed plant. Here we present an analysis of two gene families central to auxin signalling: the ARFs, encoding transcription factors, and Aux/IAAs, encoding their repressors. We also analyze three families of primary auxin responsive genes, which are among the first targets of auxin-induced transcription in flowering plants.

### Aux/IAA

The *P. patens* genome encodes three Aux/IAA proteins (PpAux/IAA) (Figure 1). These proteins, at between 484 and 503 amino acids, are significantly longer than all 29 *A. thaliana* Aux/IAA (AtAux/IAA) family members (the longest of which, IAA9, consists of 338 amino acids). The three Aux/IAAs of *S. moellendorffii* vary in length between 170 and 421 amino acids. Aux/IAA proteins typically comprise four domains: domain I confers the proteins' transcriptional repressor function, domain II is a degradation motif, and domains III and IV form a protein dimerization domain, evolutionarily related to the C-terminal dimerization (CTD) domain of ARF proteins.



**Figure 1**  
**Phylogenetic relationship of *A. thaliana*, *S. moellendorffii* and *P. patens* Aux/IAA proteins (Bayesian inference).** The presence of the domain I motif is marked.

#### Domain I

In all but one *Arabidopsis* Aux/IAA protein, domain I contains a sequence of amino acids reminiscent of an ERF-associated amphiphilic repression (EAR) motif. This LxLxL domain I motif interacts directly with TOPLESS (TPL), a transcriptional co-repressor. This interaction leads to a repression of the ARF-dependent transcription of a reporter gene driven by the DR5 promoter, a synthetic auxin-sensitive marker containing repeated TGTCTC auxin response elements (AuxREs) [13,14].

PpAux/IAAs do not contain an LxLxL motif in domain I. Instead they all contain a similar LxLxPP motif (Figure 1, Additional file 1). A corresponding and overlapping LxLxLxPP motif was found in three AtAux/IAAs (IAA18, 26 and 28), forming a cluster with good bootstrap support (Figure 1, Additional file 1). The genome of *S. moellendorffii* encodes three Aux/IAA proteins, one of which contains an LxLxPP motif in domain I. The other two contain the LxLxL motif typical of flowering plants (Figure 1, Additional file 1). The genes containing LxLxL motifs found in domain I of *S. moellendorffii* and flowering plants do not form monophyletic groups. Therefore the motif is likely to have become established at least twice in each lineage.

Of the 35 Aux/IAAs encoded by the *Populus trichocarpa* genome, 27 are predicted to contain an LxLxL motif and six an LxLxPP motif [27]. In rice, it is predicted that 27 out of 33 Aux/IAAs contain an LxLxL motif; of these, two contain an LxLxLxPP motif. In rice, one Aux/IAA contains LxLxPP (Table 1). At present, there is no evidence that this LxLxPP motif, in any species, represents a functional repression domain. There are also no experimental data available which test the role of flowering plant Aux/IAAs

which contain no apparent functional domain I motif. These proteins may function as competitive regulators of the auxin response. Gaining empirical functional evidence on these proteins will allow hypotheses on their evolution to be tested.

There is one homologous position for the EAR-like motif of domain I. Based on the alignment of *A. thaliana*, *S. moellendorffii* and *P. patens* Aux/IAAs (Additional file 1) this domain I motif can be expanded to LXL [A, G] [L, P] [P, G, S, T]. This allows the detection of domain I in all sequences tested of these three species. If expanded further, an [L, I]X [L, I] [A, G] [L, P] [P, G, S, T] motif can, according to our present knowledge, be used to detect domain I in all land plant Aux/IAAs. This analysis does not preclude the possibility that other non-homologous domains serving a similar function are also present.

Mutations in the leucine positions of the domain I motif of *Arabidopsis* have been shown to result in significantly weaker repression of ARF-mediated transcription to Aux/IAA proteins [14]. Nevertheless, the widespread conservation of the LxLxPP sequence suggests it is a functional motif. The predicted presence in *P. patens* of two TPL-like transcriptional co-repressors (Additional file 2) also suggests the LxLxPP motif is able to inhibit (at least to some extent) ARF-mediated transcription. In flowering plants, however, the LxLxPP sequence appears to have been superseded by the LxLxL domain (Additional file 1). Notably, the genome of *S. moellendorffii* encodes proteins predicted to contain both motifs. Although the relative efficiency of domain I-dependent transcriptional repression in non-seed plants (via the LxLxPP motif) and flowering plants (via the LxLxL motif) is not possible to assess with the data that are currently available, it is highly significant, as they would be expected profoundly to influence the role of auxin-dependent transcriptional activation.

#### Domain II

The alignment of domain II from several Aux/IAA proteins indicates that not all 13 amino acids of the consensus sequence, which in flowering plants mediate the specific proteasomal degradation of Aux/IAAs in response to auxin, are faithfully conserved (Additional file 3). Nevertheless, a central core of five residues, representing amino acids 4–8 (GWPPV), is required for targeted protein degradation [16]. Though not sufficient to confer protein instability to a luciferase reporter fusion on its own, the functionally essential central motif (which can be represented by VGWPP [L, V, I]) is conserved in all Aux/IAAs, including those from *P. patens* and *S. moellendorffii* (Additional file 3).

**Table 1: Number of ARFs and Aux/IAA with different domains and motifs.**

Proteins	P.p.	S.m.	A.t.	P.t. *	O.s.	S.b.	V.v.	G.m.
ARF with DBD domain								
Total number	12	7	23	38 (39)	25	25	20	59
No III & IV domain	1	1	4	12 (6)	6	6	5	12
Q-rich	0	3	4**	8 (10)	8	8	3	18
ARF without DBD domain								
Only III&IV domain	1	4	0	11 (0)	0	0	0	8
Q-rich, no DBD domain	2	2	0	0 (0)	0	0	0	0
Aux/IAA								
Total number	3	3	29	33 (35)	33	26	27	62
LxLxL	0	2	28	23 (31)	27	21	17	55
LxLxPP	3	1	3	0 (6)	3	3	2	4
No domain II	0	0	5	1 (2)	6	2	8	21
No KR domain	0	0	10	9 (10)	11	7	9	22

(P.p. – *P. patens*, S.m. – *S. moellendorffii*, A.t. – *A. thaliana*, P.t. – *P. trichocarpa*, O.s. – *O. sativa*, S.b. – *S. bicolor*, V.v. – *Vitis vinifera* G.m – *G. max*).

\* The number of ARFs and Aux/IAAs from Kalluri et al. [27] is shown in brackets.

\*\* In addition, one *A. thaliana* ARF, ARF5, has a methionine-rich middle region.

Aux/IAAs are degraded after domain II binds to the TIR1 family of F-box proteins [19-21]. The presence of the core motif of domain II and four paralogs of the Aux/IAA-specific TIR1 family of F-box proteins (Additional file 4) in *P. patens*, suggests that PpAux/IAAs are degraded in an auxin-dependent manner. Homology modelling has shown that the auxin binding pocket of PpTIR1 is intact [26]. Together, these data suggest that auxin-mediated targeted protein degradation is relevant in *P. patens*, and that the relatively slow response of *P. patens* to auxin [28,29] is not due to an impaired ability to degrade Aux/IAAs in response to auxin.

#### Diversification of Aux/IAA

The dramatic radiation of Aux/IAA genes in land plants (from three in *P. patens* to 29 in *A. thaliana*, 35 in *P. trichocarpa* and 33 *O. sativa*) (Figure 1, Table 1) underpins a corresponding increase in the complexity of auxin signalling. After the separation of lycophytes and seed plants, the Aux/IAA family in the *A. thaliana* lineage was expanded by 25 additional duplication events (Additional file 5). To test whether this radiation has been driven by neofunctionalization at the amino acid level (for example in response to specific changes in ARF protein structures), rates of positive selection were measured [30]. Specifically, we applied a likelihood ratio test (LRT) to selected Aux/IAA sub-families of *A. thaliana* and *P. trichocarpa*, and compared data fits to two models: M1 vs. M2 and M7 vs. M8 (Table 2; Additional file 6, A). A comparison of these models measures the likelihood that differences in non-synonymous/synonymous substitution ratios happened by chance. For Aux/IAA proteins, no significant differences between test and null hypotheses were found in any of the data sets tested.

Aux/IAA genes have been retained in the *A. thaliana* genome at a high rate. A two-way analysis of variance (ANOVA) test of microarray data has previously shown that the gene expression patterns of Aux/IAA sister pairs of *A. thaliana* are significantly different [31]. We extended this analysis by widening the conditions tested. Two-way ANOVA results for ten pairs of Aux/IAA genes are reported as graphs of expression levels at 63 conditions in Additional file 7, A-J (after [32]). All ten sister pairs of Aux/IAA showed significant gene (G), sample (S), and gene by sample (GxS) effects (Additional file 7, A-J).

Aux/IAA genes have radiated through segmental duplication events [33]. In *P. trichocarpa* and *O. sativa*, both ARF and Aux/IAA gene families have been expanded, also largely due to segmental duplication [27,34,35]. After such events, the gradual appearance of deleterious mutations generally leads to the loss of one of the duplicated genes [36]. If both gene copies are retained, there is a higher probability that mutations leading to a split in the expression pattern of the ancestral gene between duplicated genes, rather than mutations that lead to a new function in one copy, have occurred [37]. Such a split can occur through changes in transcription-factor binding sites within promoter regions that result in differential expression of the two gene copies. We therefore conclude that changes in expression pattern have driven Aux/IAA radiation. Indeed, when compared to amino acid substitution rates, changes in expression pattern contribute more to Aux/IAA function [38,39]. Studies in *P. trichocarpa* [27] also showed that genes of the expanded PtIAA3 subgroup, which is represented by six members, are differentially transcribed. These data lend further support to the hypothesis that the diversification of Aux/IAA

family members in flowering plants has been sustained by changes in their expression patterns.

### ARF

In *A. thaliana*, all ARFs contain a DNA binding domain, but some lack a C-terminal dimerization domain (CTD). The genomes of *S. moellendorffii* and *P. patens* also encode ARFs with C-terminal truncations, as well as those with N-terminal truncations. All of these variants are discussed below.

#### Full-length and C-terminally truncated ARFs

To examine evolutionary relationships among *P. patens* (PpARF), *S. moellendorffii* (SmARF) and *A. thaliana* (AtARF) ARF proteins, a rooted phylogenetic tree was constructed from the alignment of the predicted protein sequences of the 12 PpARFs, the 7 SmARFs and the 23 AtARFs predicted to contain a DNA binding domain (DBD). All 42 ARF genes analysed could be grouped into five major classes (Figure 2). In addition to the five previously described classes of ARFs [35], we detected an additional cluster of four *P. patens* genes and one cluster of two *S. moellendorffii* genes, each with good bootstrap support. Six PpARFs are similar to subclass IIa (AtARF5-8 and 19) and two are similar to class III (AtARF10, 16 and 17). PpARFs, therefore, fell into one of three classes (Figure 2). As in *P. patens*, *S. moellendorffii* has representatives of subclass IIa (three genes) and class III (two genes). The *S. moe-*

*llendorffii*- and *P. patens*-specific subclasses are not monophyletic.

In addition, *P. patens* and *S. moellendorffii* each encode one C-terminally truncated ARF with no CTD (Figure 2, Figure 3, Table 1). Flowering plants encode more CTD-truncated ARFs. This trend is seen in *A. thaliana* (4 out of 23 ARFs), *O. sativa* (6 out of 25 ARFs) and *P. trichocarpa* (6 out of 39 ARFs) (Table 1). Diversification of CTD-truncated ARFs in flowering plants suggests a role for auxin-independent regulation of auxin responsive genes.

#### Full-length ARF transcriptional activators

In *A. thaliana*, the first transcriptional response to exogenously applied auxin is a rapid up-regulation of auxin-responsive genes [4]. The so-called middle regions (MRs) of five AtARFs of sub-class IIa (AtARFs 5, 6, 7, 8 and 19) mediate this transcription [11]. All five of these MRs (as defined by the region between the CTD and DBD) are significantly longer than those of all other ARFs, with the exception of AtARF2 [40]. PpARFs and SmARFs of class IIa also contain an extended MR (Additional file 8 and 9). A second feature of the MRs of those AtARFs which function as transcriptional activators is a relatively high proportion of glutamine residues (except for AtARF5) (Additional file 9 and 10). The MRs of canonical PpARFs of this group contain fewer glutamine residues than their vascular plant counterparts at between 7.8 and 10% of all amino acid residues, compared to between 17.1 and 22.3 for the Q-

**Table 2: Sites under PDS in the *A. thaliana*/*P. trichocarpa* Aux/IAA gene family: „Site-specific analysis“.**

Site-specific analyses	n	dN/dS ( $\omega$ ) under M0	2 $\Delta\ell$ M2 vs. M1 (df 2)	2 $\Delta\ell$ M8 vs. M7 (df 2)	Parameter estimates under M8	Positively selected sites under M2 (BEB)	Positively selected sites under M8 (BEB)
<b>Aux/IAA</b>							
Node IAA1	9	0.112	0.00	0.00	$\rho_0 = 1.000$ ( $\rho_1 = 0.000$ ) ( $\rho = 0.285$ ) $q = 1.705$ $\omega = 2.40$	None	22
Node IAA5	5	0.132	0.00	1.44	$\rho_0 = 0.944$ ( $\rho_1 = 0.056$ ) ( $\rho = 0.315$ ) $q = 1.634$ $\omega = 2.11$	None	21 23 39 40 <u>68</u>
Node IAA7	10	0.090	0.00	4.62	$\rho_0 = 0.948$ ( $\rho_1 = 0.052$ ) ( $\rho = 0.318$ ) $q = 3.848$ $\omega = 1.00$	None	16 <u>19</u> 40 42 49 75 80 81
Node IAA8	7	0.125	0.00	0.00	$\rho_0 = 1.000$ ( $\rho_1 = 0.000$ ) ( $\rho = 0.537$ ) $q = 2.898$ $\omega = 2.302$	None	10 11 171
Node IAA10	6	0.076	0.00	0.10	$\rho_0 = 0.977$ ( $\rho_1 = 0.023$ ) ( $\rho = 0.522$ ) $q = 3.630$ $\omega = 1.000$	None	124
Node IAA18	6	0.112	0.00	0.00	$\rho_0 = 1.000$ ( $\rho_1 = 0.000$ ) ( $\rho = 0.505$ ) $q = 0.301$ $\omega = 5.133$	None	20 115 116
Node IAA20	5	0.155	0.74	4.54	$\rho_0 = 0.898$ ( $\rho_1 = 0.102$ ) ( $\rho = 0.617$ ) $q = 4.304$ $\omega = 1.307$	43	13 <u>43</u> 49

Each comparison has n sequences, dN/dS is average ratio over sites under a codon model with one  $\omega$ . Bold underline,  $PP \geq 0.99$  of being under positive selection; bold,  $0.99 > PP \geq 0.95$ ; italic,  $0.95 > PP \geq 0.90$ ; underline,  $0.90 > PP \geq 0.70$ ; normal,  $0.70 > PP > 0.50$

rich ARFs of *A. thaliana*. PpARFs are unidentifiable as Q-rich both by the normalized amino acid frequency used for Additional file 10 and by a PROSITE domain search. Nevertheless, these MRs all contain a higher proportion of glutamine residues than all but two of the repressor AtARFs (Additional file 9). Given the character states of the MR length (Additional file 8) and glutamine content (Additional file 10) in the phylogenetic tree, a single gain of the domain (basal to the cluster starting with AtARF7 and 19) seems to have occurred. The MR seems to have been secondarily reduced in one SmARF (Selmo1\_2\_438333) and secondarily expanded in AtARF2. The subsequent enrichment of the MR with glutamine residues apparently evolved several times independently within the genes containing the prolonged MR. *S. moellendorffii* contains three class IIa canonical ARF transcriptional activators. These proteins all contain an extended, Q-rich MR. The simultaneous appearance of an LxLxL motif in *S. moellendorffii* Aux/IAAs allows the possibility that this motif co-evolved with the appearance of canonical Q-rich ARFs.

The exogenous application of auxin to *P. patens* has been shown to have only a weak effect on the expression of transgenic flowering plant auxin-responsive markers [28,29,41]. In contrast, auxin-responsive transcription in *A. thaliana* is observed rapidly, and at relatively low auxin concentrations [4,42]. The slower response in *P. patens* could be due to a number of reasons relating either to an inability of the moss to recognize auxin-responsive flowering plant promoter elements, or to a slower auxin response in *P. patens per se*. Direct experimental evidence is needed if we are to state firmly that there is indeed a slower auxin response in *P. patens*, and that this is due to a relatively weak activation of gene transcription by ARFs. However, the observations that i) the LxLxPP motif of Aux/IAA domain I has been gradually replaced by an LxLxL motif in most flowering plants Aux/IAAs, ii) Q-rich ARFs and the LxLxL EAR-like domain appear together in *S. moellendorffii*, iii) mutations in the canonical LxLxL motif confer weaker transcriptional repression in *A. thaliana* [14], and iv) there is a relatively slow transcriptional response of *P. patens* to auxin together lead us to hypothesize that the Q-enriched subclass IIa ARFs of *P. patens* are moderate rather than strong transcriptional activators.

#### *N-terminally truncated ARFs are candidate trans-acting ARF regulators*

Two proteins encoded by the *P. patens* genome, Phypa\_171888 and Phypa\_170581, contain both a CTD and an extended Q-rich MR (at 14.0 and 14.5% Q) (Figure 3, 4). However, neither protein contains a recognizable DNA-binding domain (DBD); both gene models were manually checked for accuracy. This phylogenetic analysis placed both truncated *P. patens* proteins in class IIa with

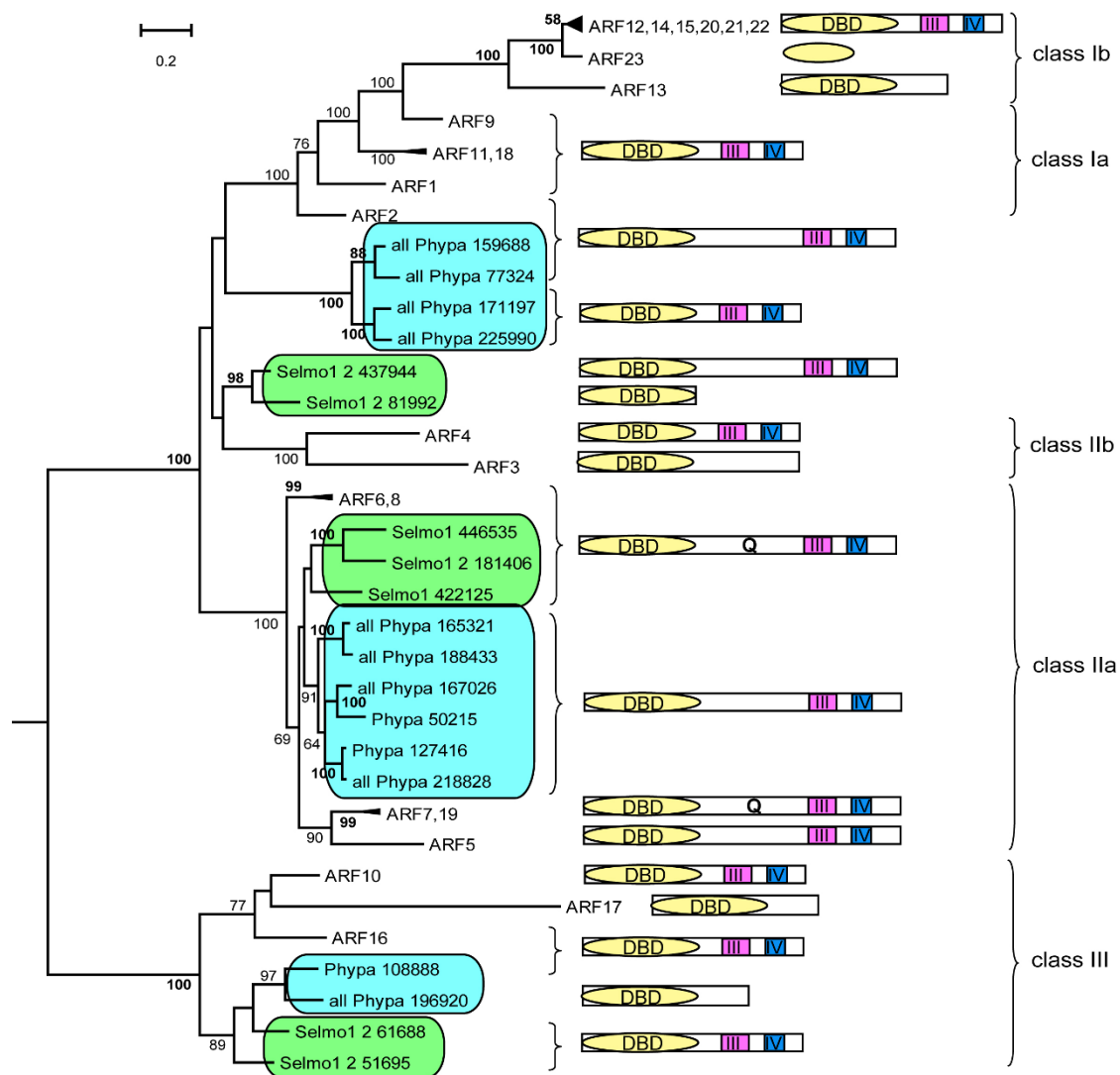
AtARF transcriptional activators 6, 7, 19, 5 and 8. The genome of *S. moellendorffii* also encodes two DBD-truncated ARFs in a monophyletic group with N-terminally truncated *P. patens* ARFs. Artificially truncated AtARF5, 6, 7 and 8 proteins (with their DNA-binding domains removed) have previously been shown to activate strongly (15- to 20-fold) transcription of an auxin-responsive reporter gene by dimerizing with canonical ARFs [11]. Therefore, it is possible that the activation of an auxin response in *P. patens* could be relayed by a CTD-dependent heterodimerization between a DBD-truncated ARF and a canonical ARF (Figure 5[1]).

The presence of DBD-truncated Q-rich ARFs allows an alternative transcriptional control, alongside the evolution of a functional motif in domain I of Aux/IAAs. Such an N-terminal truncation enables the spatial separation of transcription-activating MRs and DBDs by the competitive inhibition of ARF CTDs by Aux/IAAs (Figure 5[1]). A functional domain I-motif would not be necessary for such inhibition. Since such an inhibitory mechanism is not able to separate the DNA-binding and activation domains present in a single ARF, we hypothesize that a strong selection pressure on domain I of Aux/IAAs for the efficient recruitment of transcriptional co-repressors could have been a feature of Aux/IAA evolution after the appearance of canonical Q-rich ARFs. This hypothesis would predict that at least two mechanisms have evolved through which the evolution of a strong ARF activation domain has been accommodated: firstly, the appearance of a strong Aux/IAA repressor domain, as seen in flowering plants, and secondly, the spatial separation of the ARF activation domain from the DNA-binding domain, as seen in *P. patens*. Notably, *S. moellendorffii* is predicted to employ both.

A second group of proteins with an N-terminal truncation is encoded by the genomes of *P. patens* and *S. moellendorffii* (Table 1, Figure 3, Figure 5[2]). Here the truncation is larger, and the encoded proteins are predicted to have neither a DBD-domain, nor a middle region. We propose that proteins of this group act as auxin-independent competitive inhibitors of ARF dimerization, inhibiting both potentiation (via ARF-ARF dimerization) and repression (via ARF-Aux/IAA dimerization) of the auxin response.

#### **Evolution of ARF activators**

To test whether positive selection, and therefore possible neofunctionalization, has driven evolution within the extended ARF MR of class IIa, we compared the relative rates of synonymous and non-synonymous substitutions in full-length coding sequences from all ARFs of two fully-sequenced dicotyledonous species: *A. thaliana* and *P. trichocarpa*. A likelihood ratio test (LRT) was applied to selected ARF sequences from *A. thaliana* and *P. trichocarpa*

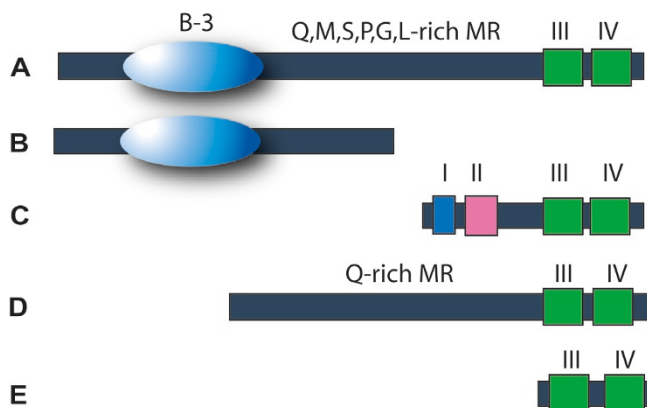
**Figure 2**

**Phylogenetic relationship of *A. thaliana*, *S. moellendorffii* and *P. patens* ARF proteins, aligned using DBDs (Maximum Likelihood (ML) method).** Bootstrap values greater than 49 are recorded. These numbers are given in bold if the NJ value is equal to or higher than the ML value. This applies only to those branches common to both topologies.

(Additional file 6, B). The maximum likelihood estimates (MLEs) of parameters under model M2a and M8 are listed in Table 3, together with the sites inferred to be under positive selection by the Bayes empirical Bayes (BEB) approach. For node ARF12, both M2a and M8 have significantly higher likelihood values than their corresponding null models M1a and M7, suggesting the presence of sites under positive selection. For node ARF7, M8 (but not M2a) had significantly higher likelihood values than its corresponding null model M7 (Table 3). Consequently, this model identified 24 sites under positive selection (21 sites are presented in Additional file 11), all within the extended MR (Table 3), supporting the hypothesis that

positive selection within the MR plays a role in neofunctionalization, possibly directly influencing the acquisition of a transcriptional activation function. It is however possible that positive selection influences an unrelated function of the MR. For example, the MR may influence protein stability, as is the case for the MR of ARF1 [43]. Yet, if the enrichment of glutamine is important for transcriptional activation, the fact that 14 of the 24 sites under positive selection in the MR encode a glutamine in at least one protein clearly argues for the involvement of positive selection in the acquisition of that particular function.





**Figure 3**  
**Domain structures of the Aux/IAA and ARF families found in *A. thaliana*, *S. moellendorffii* and *P. patens*.**  
 Canonical, full length ARFs (A); C-terminally truncated ARFs (B); Aux/IAAs (C); N-terminally truncated ARFs (D); *S. moellendorffii*-specific ARF with a large N-terminal truncation (E).

ARF7 and ARF19 dimerize with Aux/IAAs to regulate the expression of partially overlapping sets of auxin-responsive genes in the control of lateral root development and gravitropism [44]. However, ARFs do not only dimerize with Aux/IAAs. In Arabidopsis, a member of a second class of transcription factors, MYB77, interacts with the CTD ARF7 to control auxin-responsive gene expression and lateral root number [45]. Therefore a third interaction, besides DNA or Aux/IAA interaction, influences ARF evolution. As the interaction between MYB77 and ARFs occurs with the ARF CTD, it cannot explain positive selection within the proteins' MR. It does, however, represent a precedent for Aux/IAA independent protein-protein interactions (and possible subsequent post-translational modification) influencing protein function within the ARF7 node, and the presence of, as yet unconsidered, evolutionary pressures influencing ARF function.

#### Evolution of ARFs which lack a Q-rich middle region

In contrast to the relatively constant numbers of class IIa ARF transcriptional activators encoded by the genomes of *P. patens*, *S. moellendorffii* and *A. thaliana* (six, three and five respectively), the number of ARF repressors has increased from five and four in *P. patens* and *S. moellendorffii* to fourteen in *A. thaliana*. There is only one *P. patens*-specific and one *S. moellendorffii*-specific ARF class, whilst there are three flowering plant-specific sub-classes (class Ia, class Ib and class IIb) indicating that evolution within flowering plants has favoured strongly the diversification of auxin-regulated repressor ARFs (Figure 2).

Two polyphyletic groups of ARF lacking a Q-rich region can be differentiated: those with a CTD and those without. This suggests at least two distinct mechanisms of tran-

scriptional regulation. ARFs which lack a CTD are responsible for the auxin-independent (or basal) regulation of auxin-responsive genes (Figure 5[4]). These ARFs cannot interact with Aux/IAAs and therefore their transcriptional activity is independent of cellular auxin concentration. However, identity within their DNA-binding domain suggests they are able to bind to auxin responsive promoter elements. The second type of ARF has a CTD and is, at least according to the accepted paradigm, able to dimerize with Aux/IAAs [9,18,46] (Figure 3, 5[3]). Phosphorylation of ARF2 (a full length ARF) by BIN2, a kinase involved in brassinosteroid-dependent transcription decreases its ability to bind DNA [47]. This path for cross-talk between two hormone signalling pathways (auxin and brassinosteroid) represents a precedent for ARF repressors to perform in other signalling functions.

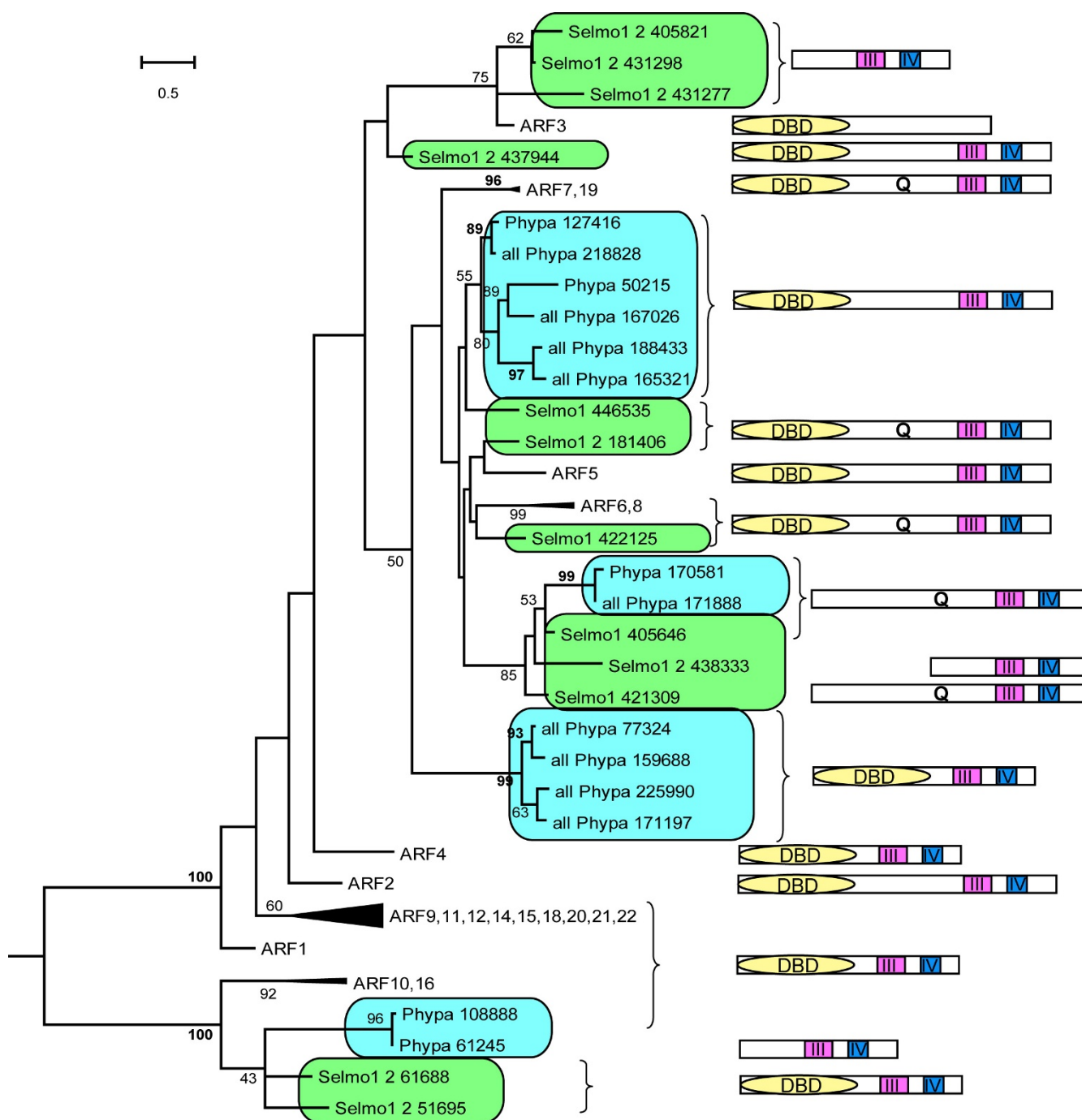
After analysis of all genes encoding *A. thaliana* ARF transcriptional repressors, positive selection was only observed in the ARF12 node (Class Ib), where little is known about protein function (Additional file 6, B). In this class, single knockouts do not show obvious aberrant phenotypes [44], and the generation of double knockout lines has been hampered by the genes' close proximity on chromosome 1.

Class Ib ARFs are absent from the *P. trichocarpa* and *O. sativa* genome, raising the possibility of a specific role within the order Brassicales [27,35]. Positive selection does not prove the acquisition of novel and specific function in the class Ib ARFs of *A. thaliana*. However, together with the subgroup's rapid and significant diversification, followed by the retention of duplicated genes, it suggests neofunctionalization. Any putative new function is also likely to be related to the amino acid residues under positive selection in the middle region of the protein, possibly facilitating new protein-protein interactions, protein stability, or post-translational modifications.

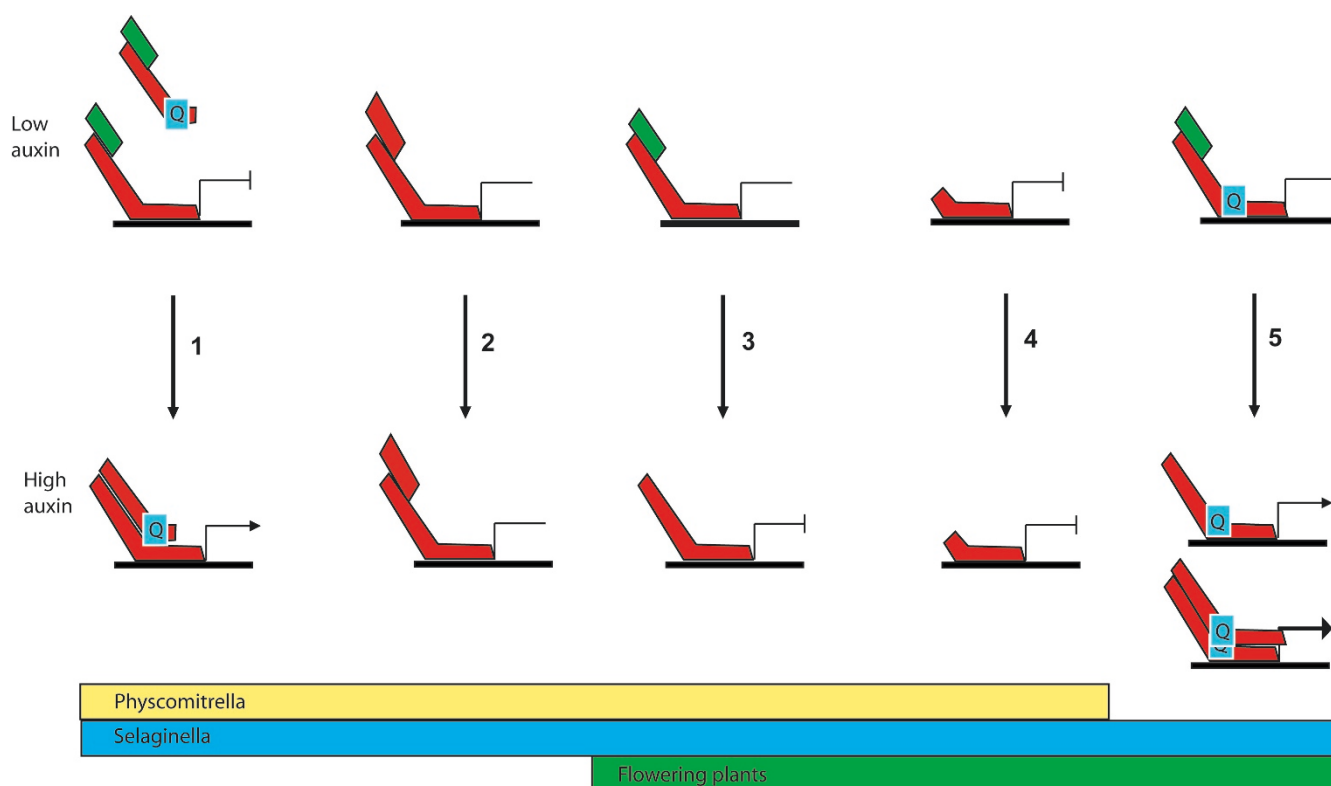
#### Auxin-independent regulation of ARF activity

ARFs without a CTD cannot dimerize with Aux/IAAs and are therefore not expected to be regulated directly by auxin. But is auxin-independent ARF signalling relevant to flowering plants, or is ARF function dependent on a functional CTD? CTD-deficient *arf* mutants do indeed have aberrant phenotypes. Four ARF proteins which lack a CTD are predicted to be encoded by the *A. thaliana* genome. Of these, *ARF3* mutants show pleiotropic effects in flower development [48]. Plants expressing a miRNA-resistant version of a second CTD-deficient ARF, *ARF17*, have increased *ARF17* mRNA levels and display dramatic developmental defects. These include embryo and emerging leaf symmetry anomalies, leaf shape defects, premature inflorescence development, altered phyllotaxy, reduced petal size, abnormal stamens, sterility, and root growth





**Figure 4**  
**Phylogenetic relationship of *A. thaliana*, *S. moellendorffii* and *P. patens* ARF proteins, aligned using CTDs (Maximum Likelihood (ML) method).** Bootstrap values greater than 49 are recorded. These numbers are given in bold if the NJ value is equal to or higher than the ML value. This applies only to those branches common to both topologies.

**Figure 5**

**Model of auxin signalling in *A. thaliana*, *S. moellendorffii* and *P. patens*.** (1) In *P. patens* and *S. moellendorffii*, N-terminally truncated Q-rich MRs are predicted to dimerise with full length ARFs to mediate transcription in the presence of auxin. In the absence of auxin, Aux/IAAs inhibit this interaction. Dimirization between other canonical ARFs is also possible. (2) *A. thaliana*, *S. moellendorffii* and *P. patens* full-length ARFs have a MR that is not Q-rich, which can dimerize with either truncated regulatory ARFs in an auxin-independent manner (in *P. patens* and *S. moellendorffii*), (3) or with Aux/IAAs in an auxin dependent manner (*A. thaliana*, *S. moellendorffii* and *P. patens*). (4) *A. thaliana*, *S. moellendorffii* and *P. patens* C-terminally truncated ARFs mediate transcription in an auxin-independent manner. (5) *A. thaliana* and *S. moellendorffii* ARFs with Q-enrichment in their MR. These ARFs directly activate expression of auxin responsive genes. ARF activators also dimerize and potentiate the activation of auxin responsive genes. These ARFs are regulated by Aux/IAAs in an auxin-dependent manner.

defects [49]. The search for alternative mechanisms of ARF regulation has centred on small RNAs. Two *P. patens* ARF transcripts (*Phypa\_159688* and *Phypa\_171197*), both encoding full-length ARFs (Figure 4), have been identified as targets of small RNAs [50]. Regulation of ARFs by miRNA in *A. thaliana* can be considered as auxin-independent because auxin treatment does not alter appreciably *miR160*, *miR164*, and *miR167* accumulation, at least in seedlings [49]. In *A. thaliana*, mRNAs encoding two out of the four ARFs which have no CTD have also been identified as targets of small RNAs: *ARF3* is the target of *AtTAS3a-c* and *ARF17* is the target of *miR160* [50]. Small RNAs do not only target transcripts of ARFs without a CTD, but also Aux/IAA-binding ARFs. The regulation of ARF activity is therefore complex and involves the integration of auxin-dependent and auxin-independent mechanisms (Figure 5). miRNAs are also potentially important regulators of cross-talk between auxin and other signalling

pathways, for example between auxin and abscisic acid [51].

Auxin-independent, cell-dependent regulation of auxin signalling activity has previously been identified as an important factor in plant development [52]. Indeed, endogenous small regulatory RNAs seem to play a relatively important role in the regulation of ARF gene expression [53]. For example, in *A. thaliana* the expression pattern of both *ARF6* and *ARF8* (involved in female and male reproductive organ development) is controlled by *miR167*, with *miRNA160* also involved in the control of ARF expression in *P. patens* and *A. thaliana* as well as in *S. moellendorffii*, suggesting a conserved mechanism of ARF post-transcriptional regulation [52,54]. The *P. patens* genome encodes a surprisingly diverse population of miRNAs. However, in contrast to ARF and Aux/IAA genes, the

**Table 3: Sites under PDS in the *A. thaliana*/*P. trichocarpa* ARF gene family: „Site-specific analysis“.**

Site-specific analyses	n	dN/dS ( $\omega$ ) under M0	$2\Delta\ell$ M2 vs. M1 (df 2)	$2\Delta\ell$ M8 vs. M7 (df 2)	Parameter estimates under M8	Positively selected sites under M2 (BEB)	Positively selected sites under M8 (BEB)
<b>ARF</b>							
Node ARF2	5	0.183	0.00	0.56	$\rho_0 = 0.957$ ( $\rho_1 = 0.043$ ) ( $\rho = 0.431$ ) $q = 1.635$ $\omega = 1.09$	None	478 540 563 631 <u>758</u>
Node ARF6	7	0.122	0.00	1.04	$\rho_0 = 0.997$ ( $\rho_1 = 0.003$ ) ( $\rho = 0.355$ ) $q = 1.994$ $\omega = 4.49$	None	<u>372464476</u> 479 482 <u>485</u> 539
Node ARF7	6	0.154	0.00	13.78**	$\rho_0 = 0.973$ ( $\rho_1 = 0.027$ ) ( $\rho = 0.409$ ) $q = 1.814$ $\omega = 3.086$	None	<u>460</u> 526 <u>532534</u> 539 <u>542544</u> 546 550 556 <u>557</u> 558 559 561 562 <u>566568</u> 569 570 <u>574599</u> 608 677 785
Node ARF10	8	0.137	0.00	0.64	$\rho_0 = 0.998$ ( $\rho_1 = 0.016$ ) ( $\rho = 0.413$ ) $q = 1.988$ $\omega = 2.69$	None	96 449 <u>486570</u>
Node ARF11	5	0.151	0.00	1.74	$\rho_0 = 0.932$ ( $\rho_1 = 0.068$ ) ( $\rho = 0.528$ ) $q = 2.885$ $\omega = 1.00$	None	6 334 370 372 376 383 399 412 413 478 529
Node ARF12	6	0.566	19.42**	19.4**	$\rho_0 = 0.982$ ( $\rho_1 = 0.018$ ) ( $\rho = 0.056$ ) $q = 0.054$ $\omega = 10.50$	<u>175335389</u> 433 <u>444447</u> 527 <u>567</u> <u>569</u> <u>570</u> 572	<u>175335</u> 351 359 <u>389</u> 433 444 447 527 <u>567</u> <u>569</u> <u>570</u> <u>572</u>

Each comparison has n sequences, dN/dS is average ratio over sites under a codon model with one  $\omega$ . Bold underline,  $PP \geq 0.99$  of being under positive selection; bold,  $0.99 > PP \geq 0.95$ ; italic,  $0.95 > PP \geq 0.90$ ; underline,  $0.90 > PP \geq 0.70$ ; normal,  $0.70 > PP > 0.50$

number of miRNAs conserved between *P. patens* and *A. thaliana* is relatively large [55].

#### Primary auxin response genes

Primary auxin response genes (those genes whose expression is directly regulated by ARFs) can be grouped into three major families: *Aux/IAAs*, *GH3s* and *SAURs*. Recently, transcription of certain *LOB domain (LBD)* genes has also been shown to be rapidly and specifically up-regulated by auxin [4,56]. All four of these major gene families are represented in the genomes of *P. patens* and *S. moellendorffii*. However, a detailed analysis of their response to auxin application is precluded by the lack of global transcriptional data from these species.

Microarray analysis has showed that, in *A. thaliana*, only the transcription of group II *GH3* genes (which encode auxin conjugating enzymes) is regulated by auxin [4]. Similarly, in *O. sativa*, the transcription of *GH3* genes which were most strongly up-regulated in response to auxin treatment also belong to group II [57]. The *P. patens* genome contains two genes that are homologous to the *GH3* family of flowering plants. Both conjugate IAA to amino acids, with PpGH3-2 showing a far broader range of substrate specificity than PpGH3-1 [58]. Surprisingly, the moss *GH3* genes form a common clade with the group I genes of *A. thaliana*, and not with those encoding the auxin conjugating enzymes of group II. Furthermore, the clades are separated by a relatively high genetic distance, suggesting that they diverged a relatively long time ago (Additional file 12). Auxin application increases tran-

scription of specific flowering plant *GH3* genes of group II. This increase has never been demonstrated in *P. patens* [59]. The genome of *S. moellendorffii* is predicted to encode one group II *GH3* enzyme, and one protein belonging to group I (Additional file 12). The remaining 19 *SmGH3* genes cannot be clearly assigned based on phylogeny. The transcriptional response to auxin of these genes has never been tested.

*P. patens* *GH3* enzymes are nevertheless able to conjugate auxin. Direct measurements of auxin conjugates in moss plants have give valuable insights into the developmental role of auxin conjugation by *GH3s*. *P. patens* plants lacking both *GH3* enzymes, when grown on IAA, still conjugate auxin. These results suggest other classes of enzymes may also conjugate auxin in *P. patens* [58].

Based on phylogeny *P. patens* *GH3s* are more closely related to *GH3*-11 of *A. thaliana*, which catalyses the synthesis of jasmonic acid conjugates [58,60]. *P. patens* plants lacking the *GH3*-2 gene show an increased sensitivity to high jasmonic acid concentrations, suggesting a potential role for jasmonic acid conjugation as well for this enzyme [58]. A broad substrate specificity of *GH3*-2 in *P. patens* could suggest that the enzyme has retained this characteristic from the common ancestor of all land plants.

In flowering plants, *SAUR* genes are a diverse family of unknown function with differing responsiveness to auxin [4]. The *P. patens* genome contains 18 *SAUR* genes (*A. thaliana* approximately 70), which cluster in two groups

with low bootstrap support (Additional file 13). All *AtSAUR* genes of group A are auxin-responsive [4,62]. This group shows relatively high similarity to nine *PpSAUR* genes (albeit with low bootstrap support) (Additional file 13) and therefore could participate in the auxin response in *P. patens*. The LOB domain family of transcription factors also contains important auxin-responsive signalling proteins. In *P. patens*, the LBD gene family has 17 members, forming five clades (Additional file 14). One clade, encoding four LBDs (Phypa\_18666, 7278, 25219 and 48669), is monophyletic with important auxin-responsive regulators of lateral root formation in *A. thaliana*, LBD16 and 29 [56], and therefore represents candidates for *P. patens* auxin primary response, an attractive target for future research.

## Conclusion

It is clear that auxin signalling is responsible for many aspects of vascular plant growth and development. In this manuscript, we demonstrate that the genome of *P. patens* encodes all of the basic components necessary for an auxin response. We also suggest that the evolution of an alternative, competitive mechanism of transcriptional control in *P. patens*, involving the truncation of ARF transcriptional activators, substitutes for a mechanism which, in *S. moellendorffii* and flowering plants, confers a rapid auxin response.

However, without a systematic analysis of the auxin transcriptional response in *P. patens* and *S. moellendorffii* it remains difficult to assess (i) whether these plants are capable of rapidly synthesizing specific mRNAs in response to auxin in the same manner as flowering plants, and (ii) the role any such response plays in auxin homeostasis and plant development.

It is, however, clear that an expansion of the *Aux/IAA* gene family accounts for much of the diversification of auxin signalling proteins in flowering plants. Furthermore, the smaller size of many gene families relevant to auxin signalling in *P. patens* is probably correlated to the lower structural complexity of this plant. This correlation is especially pronounced in *Aux/IAA* gene families.

Auxin and its polar transport are crucial factors in flowering plant development, and have come to direct many processes which are not relevant to mosses such as apical dominance, formation and maintenance of shoot and root apical meristems and vascular differentiation. Mosses nevertheless require auxin for cell differentiation and division. Understanding the differences in the underlying mechanisms of auxin signalling, which drive these different physiological processes, and of their evolutionary relationship, will be a fascinating challenge for the future.

## Methods

### Candidate gene family member selection and curation

To define and extract the *ARF*, *Aux/IAA* and *TIR1* gene families we screened the published genomes of *A. thaliana* (TAIR7; [ftp://ftp.arabidopsis.org/Sequences/blast\\_datasets/TAIR7\\_blastsets/TAIR7\\_pep\\_20070425/](http://ftp.arabidopsis.org/Sequences/blast_datasets/TAIR7_blastsets/TAIR7_pep_20070425/)), *O. sativa* (Osa1 version 5.0; [ftp://ftp.tigr.org/pubdat/Eukaryotic\\_Projects/o\\_sativa/annotation\\_dbs/pseudomolecules/version\\_5.0/all.chrs/all.pep/](http://ftp.tigr.org/pubdat/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_5.0/all.chrs/all.pep/)), *P. trichocarpa* (Poptr1\_1; [ftp://ftp.jgi-psf.org/puJGI\\_datPoplar/annotation/v1.proteins.Poptr1\\_1.lamboree\\_Models.fasta.gz](http://ftp.jgi-psf.org/puJGI_datPoplar/annotation/v1.proteins.Poptr1_1.lamboree_Models.fasta.gz)), *S. bicolor* (Sbi1.4; [ftp://ftp.jgi-psf.org/pub/JGI\\_data/Sorghum\\_bicolor/v1.0/Sbi/annotation/Sbi1.4/Sbi1.4.pep.fasta.gz](http://ftp.jgi-psf.org/pub/JGI_data/Sorghum_bicolor/v1.0/Sbi/annotation/Sbi1.4/Sbi1.4.pep.fasta.gz)), *V. vinifera* (Vitis\_vinifera\_v1; [http://www.genoscope.cns.fr/externe/Download/Projets/Projet\\_ML/data/annotation/Vitis\\_vinifera\\_peptide\\_v1.fasta](http://www.genoscope.cns.fr/externe/Download/Projets/Projet_ML/data/annotation/Vitis_vinifera_peptide_v1.fasta)), *G. max* (Glyma0.1b.pep.fasta.gz; [ftp://ftp.jgi-psf.org/pub/JGI\\_data/Glycine\\_max/Glyma0/annotation/](http://ftp.jgi-psf.org/pub/JGI_data/Glycine_max/Glyma0/annotation/)) and *P. patens* (Phypa1\_1; [ftp://ftp.jgipsf.org/puJGI\\_datPhyscomitrella\\_patenv1.protein\\_syp\\_1.Filtered\\_Models.fasta.gz](http://ftp.jgipsf.org/puJGI_datPhyscomitrella_patenv1.protein_syp_1.Filtered_Models.fasta.gz)) by BLASTP against a database containing all (predicted) proteins of the respective organisms. As queries, the known members of the *A. thaliana* gene families were used. For *ARF* queries, At1g35240, At1g77850 and At5g60450 were selected; for *Aux/IAA*, At1g04550, At2g01200 and At4g14560; and for *TIR1*, At3g62980 and At5g49980. Based on the protein domain architecture of the *A. thaliana* proteins, BLAST results were inspected manually to determine query specific filtering criteria. For *ARF* sequences, we required that 30% of amino acids be identical and 50% of aligned amino acid sites be shared; for *Aux/IAA* sequences, we used an E-value threshold of 1E-40, and 1E-62 for *TIR1*. The *S. moellendorffii* candidate gene family members were detected using BLASTP against the filtered models 2 predicted proteins using the filtering criteria mentioned above. All candidate loci were manually inspected using the JGI genome browser <http://genome.jgi-psf.org/Selmo1/> and curated to select the "optimal" gene model. Additionally, the genomic contexts (~40 kbp) of highly conserved gene model pairs were compared to exclude redundancies due to gene models representing loci from the two sequenced haplotypes. Furthermore, the *P. patens* genome v1.1 was screened for additional, as yet undetected, gene family members using Exonerate [63]. All detected *P. patens* candidate loci were inspected manually using the cos-moss.org genome browser <http://www.cos-moss.org/rowsys/come>. Under consideration of all available cDNA, EST and protein evidences the "optimal" predicted gene model was derived for each locus. To reduce complexity and maintain readability of the resulting phylogenetic trees, further analysis only included the candidate proteins from *P. patens*, *S. moellendorffii* and *A. thaliana*. *ARF*

and Aux/IAA *P. patens* protein IDs: 108888, 127416, 170581, 50215, 61245, 164608, 219923, 159688, 165321, 167026, 171197, 171888, 188433, 196920, 218828, 225990, 77324 and BAB71765. ARF and Aux/IAA *S. moellendorffii* protein IDs: 431277, 431298, 405821, 438333, 181406, 61688, 51695, 437944, 81992, 406764, 412634, 26861, 405646, 421309, 446535, and 422125.

#### Domain annotation and multiple sequence alignments

Protein domain architectures of the ARF and Aux/IAA candidate hits were annotated using the Pfam [64] Hidden Markov Profiles (HMMs) PF02362.12 (B3, representing the DBD), PF06507.4 (Auxin\_resp), PF02309.7 (Aux\_IAA, representing the CTD) and the PROSITE [65] profile PS50962 (IAA\_ARF) using the hmmpfam and the ps\_scan tools and applying each domain profile's "trusted cutoff" as filtering criteria. To extract CTD domain region from both, ARFs and Aux/IAs (CTD+), the FASTA output option of ps\_scan was used. CTD+ domain sequences were aligned with MAFFT L-INSI [66], ProbCons, Muscle and T-coffee and subsequently combined into an optimal alignment using the combiner function of T-coffee [67]. Full-length multiple sequence alignments (MSAs) were calculated using Dialign [68]. Full-length MSAs including the protein domain annotation were visualized and manually inspected and curated using the Jalview [69] alignment editor. In order to generate data for the domain-based phylogenies, the full-length MSAs were clipped to either the N-terminal DNA-binding (DBD; extending the B3 + Aux\_resp domain matches) or the C-terminal interaction domain (CTD; extending the Aux\_IAA domain matches) regions, according to the domain annotation and alignment quality. Proteins missing both individual domains were discarded and the clipped MSAs were realigned using the MAFFT [66] L-INSI algorithm.

#### Phylogenetic analyses

Bayesian inference was performed using MrBayes for the clipped Aux/IAA and the CTD+ MSA with 2 runs with a mixed model prior, a proportion of invariable sites and gamma distribution for a maximum of 2,000,000 using a temperature of 0.2 and a sampling rate of 5. Maximum Likelihood (ML) and Neighbor Joining (NJ) phylogenies were calculated for the full-length Aux/IAA MSA and the clipped MSAs for the DBD, the ARF-specific CTD and the CTD including the Aux/IAs (CTD+). Bootstrapped (100×) NJ trees were calculated using a modified version of the quicktree software [70], with the Scoredist [71] matrix. ProtTest [72] was used to select the most appropriate evolutionary model for ML inference (DBD:JTT+G; CTD:JTT+G; Aux/IAA:JTT+G+F). Bootstrapped (100×) best-known likelihood topologies were calculated using the parallelized version of RAXML [73]. Generally, phylogenetic trees were rooted by midpoint-rooting. The CTD, as the common feature of both families, was used to root

the ARF and AUX/IAA trees. To infer the history of duplications and losses, the CTD+ phylogeny was reconciled with Notung [74], as used in [75-77] applying the species tree (Phyph, (Selmo, Arath)).

#### Character state analyses

The MR was defined as the region between DBD and CTD, or in case of a lack of the DBD as the region from the N-terminus to the start of the CTD. The length of the MR was transformed into a continuous character matrix comprising eight characters. Q-rich regions were represented by the amino acid frequency normalized to the length of the MR. The resulting character matrix was analyzed using the Mesquite [78] analysis tool "Trace Character History" on the basis of the Notung reconciled CTD+ MrBayes phylogeny. Nucleotide alignments of coding sequences were performed on the basis of protein alignments. The protein sequences were aligned with MAFFT [69]. DAMBE 4.5.55 [79] was used to translate protein alignments to nucleotide alignments.

#### Statistical tests for positive selection

We applied the codon-based substitution model of Yang et al. [80] to identify amino acid sites under positive selection using PAML3.14 [81]. First, we ran a test for the existence of sites with a dN/dS ratio > 1 by using a likelihood ratio test (LRT) to compare null models M1a and M7(beta) (that do not allow for sites with dN/dS > 1) with alternative models M2a (PositiveSelection) and M8(beta&omega). If the LRT difference was statistically significant we identified the sites that were under positive selection. Naïve empirical Bayes (NEB) and Bayes empirical Bayes (BEB) approaches were used [82] to calculate the posterior probability that each site belongs to a particular site class. Sites with high posterior probabilities from the class with  $\omega > 1$  were inferred to be under positive selection.

#### Microarray experiments

The microarray gene expression data for paralogous pairs of Aux/IAA genes were analyzed in 63 diverse samples [32] (in our analysis, we included only data generated from wild type plants). gcRMA normalized data were used [32]. Three biological replications were used to generate the data sets. To identify which components contribute to expression pattern divergence within each duplicate pair, the two-way ANOVA used by Duarte et al. [31] to partition the gene (G), sample (S), and gene by sample interaction (G×S) effects was extended to all 63 microarray samples. Analysis was done using Statistica 5.0.

#### Abbreviations

ARF: AUXIN RESPONSIVE FACTOR; AuxREs: Auxin responsive elements; Aux/IAA: AUXIN/INDOLE-3-ACETIC ACID; BEB: Bayes empirical Bayes; CTD: C-terminal

dimerization; DBD: NDA binding domain; EAR: ERF-associated amphiphilic repression; ERF: ethylene response factor; HMMs: Hidden Markov Profiles; LBD: LOB domain; LRT: likelihood ratio test; MLEs: maximum likelihood estimates; MSAs: multiple sequence alignments; MR: middle region; Q: glutamine; NEB: Naïve empirical Bayes; NJ: Neighbor Joining; TPL: TOPLESS.

### Authors' contributions

IAP, WT, DL, RR, SAR, KP designed research. IAP and DL performed research. IAP, DL, SAR and MP analyzed data. WT and IAP wrote the paper.

### Additional material

#### Additional file 1

*Amino acid sequence alignment of Aux/IAA proteins of A. thaliana, S. moellendorffii and P. patens across conserved domain I. An LxLxL motif is encoded by Aux/IAAs of A. thaliana and S. moellendorffii, but not of P. patens.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S1.pdf]

#### Additional file 2

*Phylogenetic relationship of A. thaliana and P. patens TOPLESS-like transcriptional co-repressors (Neighbor Joining (NJ) method). The P. patens genome encodes two TOPLESS-like transcriptional co-repressors. Bootstrap values greater than 49 are recorded.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S2.pdf]

#### Additional file 3

*Amino acid sequence alignment of Aux/IAA proteins of A. thaliana, S. moellendorffii and P. patens domain II. The core motif of domain II of Aux/IAA proteins was present in all plant species tested. Bootstrap values greater than 49 are recorded.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S3.pdf]

#### Additional file 4

*Phylogenetic relationship of A. thaliana and P. patens TIR1-like F-box proteins (Neighbor Joining (NJ) method). Four paralogs of the TIR1-family of F-box proteins are present in P. patens. Bootstrap values greater than 49 are presented.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S4.pdf]

#### Additional file 5

*Phylogenetic relationship of A. thaliana, S. moellendorffii and P. patens ARF and Aux/IAA proteins (Bayesian inference). To infer the history of duplication and losses among the species tested, the CTD+ phylogeny was reconciled with Notung using the species tree (Phypha, (Selmo, Arath)).*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S5.pdf]

#### Additional file 6

*Phylogeny of A. thaliana and P. trichocarpa Aux/IAA (A) and ARF (B) proteins. Boxes identify nodes tested for positive selection.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S6.pdf]

#### Additional file 7

*Expression pattern of paralogous pairs of A. thaliana Aux/IAA genes (A-J). gcRMA normalized data were used. Three biological replications were used to generate the data set. The two-way ANOVA was used to partition the gene (G), sample (S) and GxS interaction effects.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S7.pdf]

#### Additional file 8

*Phylogenetic relationship of A. thaliana, S. moellendorffii and P. patens ARF proteins. Reconciled tree based on Bayesian inference. Length of middle region was normalized and transformed into a continuous character matrix.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S8.pdf]

#### Additional file 9

*Detailed comparison of A. thaliana, P. patens and S. moellendorffii ARFs. Here we present details of the middle region of ARFs, the presence of domain III and IV, amino acid frequency for Q, S, G, P, L, M, the total length of proteins, and the presence of amino acid-rich domains using ScanProsite.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S9.pdf]

#### Additional file 10

*Phylogenetic relationship of A. thaliana, S. moellendorffii and P. patens ARF proteins. Reconciled tree based on Bayesian inference. Q-rich regions are represented by the amino acid frequency normalized with the length of the MR.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S10.pdf]

#### Additional file 11

*ARF protein sequence alignment of the middle regions in the ARF7 node of A. thaliana and P. trichocarpa. Arrows indicate sites at which positive selection was detected. Boxed amino acids indicate putative phosphorylation motifs.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S11.pdf]

#### Additional file 12

*Phylogenetic relationship (neighbor-joining (NJ) method) of A. thaliana, S. moellendorffii and P. patens GH3 proteins. PpGH3s are indicated in light blue. SmGH3s are indicated in light green.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S12.pdf]

### Additional file 13

*Phylogenetic relationship (neighbor-joining (NJ) method) of A. thaliana and P. patens SAUR proteins. The P. patens SAURs are indicated in light blue. A. thaliana SAURs transcriptionally up-regulated by auxin are indicated in purple.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S13.pdf>]

### Additional file 14

*Phylogenetic relationship (neighbor-joining (NJ) method) of A. thaliana and P. patens LBD proteins. LBD proteins of P. patens are indicated in light green. A. thaliana LBDs transcriptionally up-regulated by auxin are indicated in purple.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-9-126-S14.pdf>]

## Acknowledgements

We are grateful to the *Selaginella* community <http://selaginella.genomics.purdue.edu/> and to the JGI <http://genome.jgi-psf.org/Selmo1/> for providing the *S. moellendorffii* genome sequence. Our work was supported by the Deutsche Forschungsgemeinschaft (SFB 592, grant Re 837/10-2), BMBF (grant 0313921, Freiburg Initiative in Systems Biology), ESA, EU, FCI, and the Landesstiftung Baden-Württemberg GmbH. D.L. is grateful for support by the GRK 1305 International Graduate School.

## References

- Davies PJ: **Plant Hormones: Biosynthesis, Signal Transduction, Action!** Dordrecht: Springer; 2004.
- Teale WD, Paponov IA, Palme K: **Auxin in action: signalling, transport and the control of plant growth and development.** *Nat Rev Mol Cell Biol* 2006, **7**:847-859.
- Theologis A, Ray PM: **Early auxin-regulated polyadenylated messenger-RNA sequences in pea stem tissue.** *Proc Natl Acad Sci USA* 1982, **79**:418-421.
- Paponov I, Paponov M, Teale WD, Menges M, Shakrabortee S, Murray JA, Palme K: **Comprehensive transcriptome analysis of auxin responses in Arabidopsis.** *Mol Plant* 2008, **1**:321-337.
- Guilfoyle T, Hagen G, Ulmasov T, Murfett J: **How does auxin turn on genes?** *Plant Physiol* 1998, **118**:341-347.
- Gray WM, Kepinski S, Rouse D, Leyser O, Estelle M: **Auxin regulates SCFTIR1-dependent degradation of AUX/IAA proteins.** *Nature* 2001, **414**:271-276.
- Ballas N, Wong LM, Theologis A: **Identification of the auxin-responsive element, AuxRE, in the primary indoleacetic acid-inducible gene, PS-IAA4/5, of pea (Pisum Sativum).** *J Mol Biol* 1993, **233**:580-596.
- Li Y, Liu ZB, Shi XY, Hagen G, Guilfoyle TJ: **An auxin-inducible element in soybean SAUR promoters.** *Plant Physiol* 1994, **106**:37-43.
- Ulmasov T, Hagen G, Guilfoyle TJ: **ARF1, a transcription factor that binds to auxin response elements.** *Science* 1997, **276**:1865-1868.
- Kim J, Harter K, Theologis A: **Protein-protein interactions among the Aux/IAA proteins.** *Proc Natl Acad Sci USA* 1997, **94**:11786-11791.
- Ulmasov T, Hagen G, Guilfoyle TJ: **Activation and repression of transcription by auxin-response factors.** *Proc Natl Acad Sci USA* 1999, **96**:5844-5849.
- Tiwari SB, Hagen G, Guilfoyle T: **The roles of auxin response factor domains in auxin-responsive transcription.** *Plant Cell* 2003, **15**:533-543.
- Szemenyei H, Hannon M, Long JA: **TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis.** *Science* 2008, **319**:1384-1386.
- Tiwari SB, Hagen G, Guilfoyle TJ: **Aux/IAA proteins contain a potent transcriptional repression domain.** *Plant Cell* 2004, **16**:533-543.
- Worley CK, Zenser N, Ramos J, Rouse D, Leyser O, Theologis A, Callis J: **Degradation of Aux/IAA proteins is essential for normal auxin signalling.** *Plant J* 2000, **21**:553-562.
- Ramos JA, Zenser N, Leyser O, Callis J: **Rapid degradation of auxin/indoleacetic acid proteins requires conserved amino acids of domain II and is proteasome dependent.** *Plant Cell* 2001, **13**:2349-2360.
- Ouellet F, Overvoorde PJ, Theologis A: **IAA17/AXR3: biochemical insight into an auxin mutant phenotype.** *Plant Cell* 2001, **13**:829-841.
- Ulmasov T, Murfett J, Hagen G, Guilfoyle TJ: **Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements.** *Plant Cell* 1997, **9**:1963-1971.
- Kepinski S, Leyser O: **The Arabidopsis F-box protein TIR1 is an auxin receptor.** *Nature* 2005, **435**:446-451.
- Dharmasiri N, Dharmasiri S, Estelle M: **The F-box protein TIR1 is an auxin receptor.** *Nature* 2005, **435**:441-445.
- Tan X, Calderon-Villalobos LIA, Sharon M, Zheng CX, Robinson CV, Estelle M, Zheng N: **Mechanism of auxin perception by the TIR1 ubiquitin ligase.** *Nature* 2007, **446**:640-645.
- Lang D, Zimmer AD, Rensing SA, Reski R: **Exploring plant biodiversity: the Physcomitrella genome and beyond.** *Trends in Plant Science* 2008, **13**:542-549.
- Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Peer Y Van de, Verrier PJ, Waters E, Wood A, Yang LX, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL: **The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants.** *Science* 2008, **319**:64-69.
- Floyd SK, Bowman JL: **The ancestral developmental tool kit of land plants.** *Int J Plant Sci* 2007, **168**:1-35.
- Cooke TJ, Poli D, Sztein AE, Cohen JD: **Evolutionary patterns in auxin action.** *Plant Mol Biol* 2002, **49**(3-4):319-38.
- Hayashi K, Tan X, Zheng N, Hatate T, Kimura Y, Kepinski S, Nozaki H: **Small-molecule agonists and antagonists of F-box protein-substrate interactions in auxin perception and signaling.** *Proc Natl Acad Sci USA* 2008, **105**:5632-5637.
- Kalluri UC, DiFazio SP, Brunner AM, Tuskan GA: **Genome-wide analysis of Aux/IAA and ARF gene families in Populus trichocarpa.** *BMC Plant Biology* 2007, **7**:59.
- Bierfreund NM, Reski R, Decker EL: **Use of an inducible reporter gene system for the analysis of auxin distribution in the moss Physcomitrella patens.** *Plant Cell Rep* 2003, **21**:1143-1152.
- Imaizumi T, Kadota A, Hasebe M, Wada M: **Cryptochrome light signals control development to suppress auxin sensitivity in the moss Physcomitrella patens.** *Plant Cell* 2002, **14**:373-386.
- Hill RE, Hastie ND: **Accelerated evolution in the reactive center regions of serine protease inhibitors.** *Nature* 1987, **326**:96-99.
- Duarte JM, Cui LY, Wall PK, Zhang Q, Zhang XH, Leebens-Mack J, Ma H, Altman N, dePamphilis CW: **Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis.** *Mol Biol Evol* 2006, **23**:469-478.
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU: **A gene expression map of Arabidopsis thaliana development.** *Nat Genet* 2005, **37**:501-506.
- Remington DL, Vision TJ, Guilfoyle TJ, Reed JW: **Contrasting modes of diversification in the Aux/IAA and ARF gene families.** *Plant Physiol* 2004, **135**:1738-1752.
- Jain M, Kaur N, Garg R, Thakur JK, Tyagi AK, Khurana JP: **Structure and expression analysis of early auxin-responsive Aux/IAA gene family in rice (Oryza sativa).** *Funct Integr Genomics* 2006, **6**:47-59.



35. Wang DK, Pei KM, Fu YP, Sun ZX, Li SJ, Liu HQ, Tang K, Han B, Tao YZ: **Genome-wide analysis of the auxin response factors (ARF) gene family in rice (*Oryza sativa*)**. *Gene* 2007, **394**:13-24.
36. Walsh JB: **How often do duplicated genes evolve new functions?** *Genetics* 1995, **139**:421-428.
37. Ganko EW, Meyers BC, Vision TJ: **Divergence in expression between duplicated genes in *Arabidopsis***. *Mol Biol Evol* 2007, **24**:2298-2309.
38. Knox K, Grierson CS, Leyser O: **AXR3 and SHY2 interact to regulate root hair development**. *Development* 2003, **130**:5769-5777.
39. Weijers D, Benkova E, Jager KE, Schlereth A, Hamann T, Kientz M, Wilmoth JC, Reed JW, Jurgens G: **Developmental specificity of auxin response by pairs of ARF and Aux/IAA transcriptional regulators**. *EMBO J* 2005, **24**:1874-1885.
40. Guilfoyle TJ, Hagen G: **Auxin response factors**. *Curr Opin Plant Biol* 2007, **10**:453-460.
41. Fujita T, Sakaguchi H, Hiwatahi Y, Wagstaff SJ, Ito M, Deguchi H, Sato T, Hasebe M: **Convergent evolution of shoots in land plants: lack of auxin polar transport in moss shoots**. *Evol Dev* 2008, **10**:176-186.
42. Abel S, Theologis A: **Early genes and auxin action**. *Plant Physiol* 1996, **111**:9-17.
43. Salmon J, Ramos J, Callis J: **Degradation of the auxin response factor ARF1**. *Plant J* 2008, **54**:118-128.
44. Okushima Y, Overvoorde PJ, Arima K, Alonso JM, Chan A, Chang C, Ecker JR, Hughes B, Lui A, Nguyen D, Onodera C, Quach H, Smith A, Yu GX, Theologis A: **Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in *Arabidopsis thaliana*: Unique and overlapping functions of ARF7 and ARF19**. *Plant Cell* 2005, **17**:444-463.
45. Shin R, Burch AY, Huppert KA, Tiwari SB, Murphy AS, Guilfoyle TJ, Schachtman DP: **The *Arabidopsis* transcription factor MYB77 modulates auxin signal transduction**. *Plant Cell* 2007, **19**:2440-2453.
46. Tiwari SB, Wang XJ, Hagen G, Guilfoyle TJ: **AUX/IAA proteins are active repressors, and their stability and activity are modulated by auxin**. *Plant Cell* 2001, **13**:2809-2822.
47. Vert G, Walcher CL, Chory J, Nemhauser JL: **Integration of auxin and brassinosteroid pathways by Auxin Response Factor 2**. *Proc Natl Acad Sci USA* 2008, **105**:9829-9834.
48. Sessions A, Nemhauser JL, McColl A, Roe JL, Feldmann KA, Zambryski PC: **ETTIN patterns the *Arabidopsis* floral meristem and reproductive organs**. *Development* 1997, **124**:4481-4491.
49. Mallory AC, Bartel DP, Bartel B: **MicroRNA-directed regulation of *Arabidopsis* AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes**. *Plant Cell* 2005, **17**:1360-1375.
50. Axtell MJ, Snyder JA, Bartel DP: **Common functions for diverse small RNAs of land plants**. *Plant Cell* 2007, **19**:1750-1769.
51. Liu PP, Montgomery TA, Fahlgren N, Kasschau KD, Nonogaki H, Carrington JC: **Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages**. *Plant J* 2007, **52**:133-146.
52. Wu MF, Tian Q, Reed JW: ***Arabidopsis* microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction**. *Development* 2006, **133**:4211-4218.
53. Reinhardt BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP: **MicroRNAs in plants**. *Genes Dev* 2002, **16**:1616-1626.
54. Axtell MJ, Bartel DP: **Antiquity of microRNAs and their targets in land plants**. *Plant Cell* 2005, **17**:1658-1673.
55. Fattash I, Voss B, Reski R, Hess WR, Frank W: **Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution**. *BMC Plant Biology* 2007, **7**:13.
56. Okushima Y, Fukaki H, Onoda M, Theologis A, Tasaka M: **ARF7 and ARF19 regulate lateral root formation via direct activation of LBD/ASL genes in *Arabidopsis***. *Plant Cell* 2007, **19**:118-130.
57. Jain M, Kaur N, Tyagi AK, Khurana JP: **The auxin-responsive GH3 gene family in rice (*Oryza sativa*)**. *Funct Integr Genomics* 2006, **6**:36-46.
58. Ludwig-Müller J, Jülke S, Bierfreund NM, Decker EL, Reski R: **Moss (*Physcomitrella patens*) GH3 proteins act in auxin homeostasis**. *New Phytol* 2009, **181**(2):323-38.
59. Bierfreund NM, Tintelnot S, Reski R, Decker EL: **Loss of GH3 function does not affect phytochrome-mediated development in a moss, *Physcomitrella patens***. *J Plant Physiol* 2004, **161**:823-835.
60. Staswick PE, Tiryaki I: **The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in *Arabidopsis***. *Plant Cell* 2004, **16**:2117-2127.
61. Hagen G, Guilfoyle T: **Auxin-responsive gene expression: genes, promoters and regulatory factors**. *Plant Mol Biol* 2002, **49**(3-4):373-85.
62. Jain M, Tyagi AK, Khurana JP: **Genome-wide analysis, evolutionary expansion, and expression of early auxin-responsive SAUR gene family in rice (*Oryza sativa*)**. *Genomics* 2006, **88**:360-371.
63. Slater GS, Birney E: **Automated generation of heuristics for biological sequence comparison**. *BMC Bioinformatics* 2005, **6**:31.
64. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer ELL, Bateman A: **The Pfam protein families database**. *Nucleic Acids Res* 2008, **36**:D281-D288.
65. Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuhe BA, de Castro E, Lachaize C, Langendijk-Genevaux PS, Sigrist CJA: **The 20 years of PROSITE**. *Nucleic Acids Res* 2008, **36**:D245-D249.
66. Katoh K, Kuma K, Toh H, Miyata T: **MAFFT version 5: improvement in accuracy of multiple sequence alignment**. *Nucleic Acids Res* 2005, **33**:511-518.
67. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment**. *J Mol Biol* 2000, **302**:205-217.
68. Morgenstern B: **Alignment of genomic sequences using DIALIGN**. *Methods Mol Biol* 2007:195-203.
69. Clamp M, Cuff J, Searle SM, Barton GJ: **The Jalview Java alignment editor**. *Bioinformatics* 2004, **20**:426-427.
70. Hanekamp K, Bohnbeck U, Beszteri B, Valentin K: **PhyloGena – a user-friendly system for automated phylogenetic annotation of unknown sequences**. *Bioinformatics* 2007, **23**:793-801.
71. Sonnhammer ELL, Hollich V: **Scoredist: A simple and robust protein sequence distance estimator**. *BMC Bioinformatics* 2005, **6**:108.
72. Abascal F, Zardoya R, Posada D: **ProtTest: selection of best-fit models of protein evolution**. *Bioinformatics* 2005, **21**:2104-2105.
73. Stamatakis A: **RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models**. *Bioinformatics* 2006, **22**:2688-2690.
74. Durand D, Halldorsson BV, Vernet B: **A hybrid micro-macroevolutionary approach to gene tree reconstruction**. *J Comput Biol* 2006, **13**:320-335.
75. Wildman DE, Chen CY, Erez O, Grossman LI, Goodman M, Romero R: **Evolution of the mammalian placenta revealed by phylogenetic analysis**. *Proc Natl Acad Sci USA* 2006, **103**:3203-3208.
76. Als TD, Vila R, Kandul NP, Nash DR, Yen SH, Hsu YF, Mignault AA, Boomsma JJ, Pierce NE: **The evolution of alternative parasitic life histories in large blue butterflies**. *Nature* 2004, **432**:386-390.
77. Whittall JB, Hodges SA: **Pollinator shifts drive increasingly long nectar spurs in columbine flowers**. *Nature* 2007, **447**:706-712.
78. Maddison WP, Maddison DR: **Mesquite: A modular system for evolutionary analysis. Version 1.12**. 2006 [<http://mesquiteproject.org>].
79. Xia X, Xie Z: **DAMBE: Software package for data analysis in molecular biology and evolution**. *J Hered* 2001, **92**:371-373.
80. Yang ZH, Nielsen R, Goldman N, Pedersen AMK: **Codon-substitution models for heterogeneous selection pressure at amino acid sites**. *Genetics* 2000, **155**:431-449.
81. Yang ZH: **PAML: a program package for phylogenetic analysis by maximum likelihood**. *Comput Appl Biosci* 1997, **13**:555-556.
82. Yang ZH, Wong WSW, Nielsen R: **Bayes empirical Bayes inference of amino acid sites under positive selection**. *Mol Biol Evol* 2005, **22**:1107-1118.