Speeding *Cis-Trans* Regulation Discovery by Phylogenomic Analyses Coupled with Screenings of an Arrayed Library of Arabidopsis Transcription Factors

Gabriel Castrillo¹, Franziska Turk², Magalie Leveugle³, Alain Lecharny³, Pilar Carbonero⁴, George Coupland², Javier Paz-Ares¹, Luis Oñate-Sánchez⁴*

¹ Department of Plant Molecular Genetics, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, Cantoblanco, Madrid, Spain, ² Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany, ³ Unité de Recherche en Génomique Végétale, Institut Scientifique de Recherche Agronomique and Centre National de la Recherche Scientifique, Evry, France, ⁴ Departamento de Biotecnología y Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid, Pozuelo de Alarón, Madrid, Spain

Abstract

Transcriptional regulation is an important mechanism underlying gene expression and has played a crucial role in evolution. The number, position and interactions between *cis*-elements and transcription factors (TFs) determine the expression pattern of a gene. To identify functionally relevant *cis*-elements in gene promoters, a phylogenetic shadowing approach with a lipase gene (*LIP1*) was used. As a proof of concept, in silico analyses of several *Brassicaceae* *LIP1* promoters identified a highly conserved sequence (*LIP1* element) that is sufficient to drive strong expression of a reporter gene in planta. A collection of ca. 1,200 *Arabidopsis thaliana* TF open reading frames (ORFs) was arrayed in a 96-well format (RR library) and a convenient mating based yeast one hybrid (Y1H) screening procedure was established. We constructed an episomal plasmid (pTUY1H) to clone the *LIP1* element and used it as bait for Y1H screenings. A novel interaction with an HD-ZIP (AtML1) TF was identified and abolished by a 2 bp mutation in the *LIP1* element. A role of this interaction in transcriptional regulation was confirmed in planta. In addition, we validated our strategy by reproducing the previously reported interaction between a MYB-CC (PHR1) TF, a central regulator of phosphate starvation responses, with a conserved promoter fragment (*IPS1* element) containing its cognate binding sequence. Finally, we established that the *LIP1* and *IPS1* elements were differentially bound by HD-ZIP and MYB-CC family members in agreement with their genetic redundancy in planta. In conclusion, combining in silico analyses of orthologous gene promoters with Y1H screening of the RR library represents a powerful approach to decipher *cis*- and *trans*-regulatory codes.


Editor: Miguel A. Blazquez, Instituto de Biología Molecular y Celular de Plantas, Spain

Received May 16, 2011; Accepted May 31, 2011; Published June 27, 2011

Copyright: © 2011 Castrillo 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.

Funding: This research was supported by two Knowledge Based Bio-Economy (KBBE) projects (REGULATORS and TRANSNET; French-German-Spanish trilateral program on plant Genomics), grants from the Spanish Ministry of Science and Innovation (MICINN; CONSOLIDER CSD2007-00057, BIO2008-04715, BFU2009-11809; BIO2010-17334) and a Marie Curie IRG grant to LO-S (036524). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: luis.onate@upm.es

† These authors contributed equally to this work.

Introduction

The control of gene expression is crucial for proper development in any living organism. Transcriptional regulation is an important mechanism underlying gene expression that has been a powerful driving force in the evolution of function and form [1] and is considered to have enormous biotechnological potential in the manipulation of agronomic traits [2 and references therein]. Transcriptional control is mediated by short DNA sequences (*cis*-elements) located in gene promoters that are bound by transcription factors (TFs). Combinatorial control driven by different *cis*-elements and their corresponding TF proteins at a given promoter is an important but not well understood area of plant gene regulation [3]. To reveal the complexity of gene transcriptional regulation it is necessary to identify all functionally relevant regulatory elements (*cis* regulatory code) as well as the TFs that interact with them (regulators in *trans*).

Non-coding sequences of orthologous genes diverge rapidly during evolution, except for those that are functionally important. This divergence in promoter sequences can be exploited to identify conserved sequences important for the regulation of gene expression, which reduces the need for time-consuming promoter analyses involving random deletions to generate promoter variants. Comparing the sequences of orthologous promoters from several related species increases the evolutionary divergence available and enables reliable detection of conserved non-coding elements whilst still allowing easy alignment of the sequences, an approach that has been called “phylogenetic shadowing” [4–6]. Phylogenetic shadowing was shown to be valuable in the identification of known as well as novel conserved motifs that are functionally important in various *A. thaliana* promoters [7–12]. The sequences identified by phylogenetic shadowing contain a combination of several *cis*-elements, providing valuable information on the conservation of the core and flanking sequences and
pinpointing the different elements to consider when studying the regulation of a given promoter.

TF proteins have a modular structure with a DNA-binding domain (DBD) and a regulatory domain. They are classified into families according to sequence similarities of their DBDs, and some families with plant specific roles are not found in other eukaryotic organisms [13–14]. Plants devote a large proportion of their genome capacity to transcriptional control and around 1,500 genes in A. thaliana are estimated to encode TFs, representing 6–7% of the genome [15-19]. Although many of the characterized TFs have been isolated based on mutant phenotypes, these approaches have limitations because many TFs belong to large families, which often leads to functional redundancy. DNA-protein interactions can be detected in eukaryotic cells by using the yeast one hybrid Y1H system. It derives from the original yeast two hybrid (Y2H) method [20] and detection is based on the interaction of a prey TF with a bait DNA-sequence cloned upstream of a reporter gene. When cDNA expression libraries are used as preys, a limitation is that low abundant messengers, such as those derived from many TF encoding genes, tend to be underrepresented. In recent years, the availability of the A. thaliana genome sequence and cloning systems based on recombination, have greatly facilitated the generation of several TF open reading frame (ORF) collections that can be used to overcome this limitation [21,26–28].

In this work, we have identified a short promoter fragment from a lipase gene (LIPI element) by phylogenomic analyses which was shown to enhance the expression of a minimal promoter-reporter construct in planta. Then, we generated an arrayed yeast library of ca. 1,200 A. thaliana TF ORFs by extending that generated previously under the REGIA project [2]. The TFs in this library (REGIA + REGULATORS; RR library) are fused to the GAL4 activation domain (GAL4AD) into a Gateway compatible plasmid and were introduced into yeast and individually arrayed in 96-well plates. An episomal plasmid to clone bait sequences has been constructed (pTUY1H) and a mating liquid assay compatible with the arrayed library has been implemented to perform screenings with less than ten hours of labour spread throughout one week. As a proof of principle, we have uncovered novel as well as known DNA-TF interactions by screening the RR library using two promoter fragments selected by phylogenetic shadowing as baits. First, we identified a homebox TF (HD-ZIP) that specifically binds to a LIPI element present in the LIPI1 element but not to a mutated version carrying 2 bp changes. This interaction induces the expression of a reporter gene construct in planta. Second, we reproduced the previously published interaction of the PHR1 TF, a central regulator of phosphate starvation responses in A. thaliana, with a conserved promoter fragment from the IPS1 gene containing its cognate cis-element (P1BS) [12,29]. Finally, we demonstrated that binding of other members of the HD-ZIP and PHR1 TF families that are present in the RR library reflect their relatedness and functional redundancy in the plant. Thus, our results suggest that phylogenomics coupled to Y1H screenings is a powerful approach to unravel cis-trans regulation that may also be used to filter genetic redundancy based on DNA-binding specificity of Arabidopsis TFs.

**Results**

**Identification of functionally relevant promoter sequences to be used for yeast one hybrid screening**

We are interested in identifying cis-elements involved in the expression control of hydrolyase genes putatively involved in the mobilization of seed reserves required for seedling establishment [30]. GDSL-lipases belong to a large protein family of lipolytic enzymes with broad substrate specificity [31] but no information is available regarding their regulatory code. We focused on the promoter of a GDSL-lipase gene (At5g45670; hereafter LIPI) that is highly induced upon germination [32–34]. LIPI1 orthologous gene promoters from several plant species belonging to the Brassicaceae family (A. thaliana, S. oleracea, C. rubella, D. sophia, B. oleracea) were amplified by using a PCR-based approach with degenerated oligonucleotides (see materials and methods). In *sensu* analyses of these promoters identified several conserved sequences with over 70% identity and between 25–50 bp in length (Figure 1A and S1). A 50 bp fragment (LIPI element) with the highest identity score (83%) was chosen for functional analysis in * planta*. A plasmid containing a minimal upstream of a luciferase reporter gene (−58F8−pPYO; control) was used to clone four copies of the LIPI1 element (4xLIPI1−58F8−pPYO; 4xLIPI1). A. thaliana transgenic plants were generated with either the control or the 4xLIPI1 construct (Figure 1B) and 10 transgenic lines for each construct (20 seeds each) were used to quantify luciferase expression in vivo 24 h after seed imbibition. The 4xLIPI1 construct was able to increase luciferase expression 13-fold over the control construct suggesting that the LIPI1 element is being bound by TFs that activate gene expression in germinating seeds (Figure 1B).

**Generation of an arrayed expression library of A. thaliana TFs in yeast and construction of the pTUY1H plasmid**

The strategy to generate a normalized library of A. thaliana TF ORFs in yeast is schematically represented in Figure S2. An ORF collection containing 288 TFs (Table S1) was generated (hereafter REGULATORS). This collection does not overlap with that previously generated under the REGIA project [2] but incorporates 71 ORFs found to be mutated or/and truncated in the REGIA collection (Table S1). We have generated a database integrating information from both collections (http://urgv.evry.inra.fr/projects/flagdb/) that also incorporates other useful features such as an up-to-date list of TF genes in A. thaliana grouped by families and, through FLAGdb+ (http://urgv.evry.inra.fr/projects/FLAGdb+/HTML/index.shtml) [33], links to TF transcriptomes from CATdb [36] and GENEVESTIGATOR [37] databases, as well as links to full-length cDNAs, T-DNA insertion mutants and miRNAs. Nested PCR with appropriate oligonucleotides (Table S2) and a high fidelity polymerase (Pfx; Invitrogen) was used to amplify TF ORFs flanked by attB sites that were recombined into a donor plasmid containing attP sites (pDONR221; Invitrogen). Sequencing results (Table S1) confirmed the suitability of our strategy since we found that most ORFs matched their annotated cDNAs or predicted sequences and were free of PCR derived errors. After amplifying 288 ORFs with an average size of 917 bp for 45 cycles, only 21 nucleotide changes, presumably derived from the PCR, were found. Five of these did not produce any amino acid change, six produced a conservative change and ten a non-conservative one (Table S1). We also found five discrepancies with previously annotated ORF sequences likely to represent a bona fide new gene model (Table 1 and S3). Together, both collections (REGIA + REGULATORS; RR library) contain 1,172 unique TF ORFs representing 43 different TF families (Table 2). The TF ORFs were recombinated into a yeast compatible plasmid conferring auxotrophy to tryptophan (W) as C-terminal translational fusions to the GAL4AD (pDEST152; Invitrogen). After verifying the quality of the resulting plasmid collection, clones were individually introduced into yeast (Saccharomyces cerevisiae YM4271; A mating type) and the resulting yeast library was stored as glycerol stocks in a 96-well plate format (Figure S2).
independent lines for each construct (20 seeds/line) are shown. Luciferase activity was quantified in vivo constructs were used to produce A. thaliana used to clone four copies of the 50 bp sequence (minimal promoter fused to the luciferase reporter gene (control) was present in the A. thaliana promoter of the GDSL-lipase gene (At5g45670) and their orthologous promoters in other Brassicaceae species were subjected to in silico analysis to identify conserved sequences (shaded boxes). Although over 1 Kb of each promoter was analyzed, significantly conserved sequences were found only along the first 500 bp upstream of the translation start site (TSS). The sequence with the highest degree of conservation (83% identity) is represented as a dark blue box and spans 50 bp (LIP1 element). A binary plasmid (pPYRO) containing a minimal promoter fused to the luciferase reporter gene (control) was used to clone four copies of the 50 bp sequence (4xLIP1) and both constructs were used to produce A. thaliana transgenic plants. Luciferase activity was quantified in vivo from transgenic seeds 24 h after imbibition. Average values and standard errors from 10 independent lines for each construct (20 seeds/line) are shown. doi:10.1371/journal.pone.0021524.g001

Figure 1. Identification of functionally relevant promoter cis-elements in a GDSL-lipase gene from A. thaliana. (A) The promoter of the A. thaliana GDSL-lipase gene (At5g45670) and their orthologous promoters in other Brassicaceae species were subjected to in silico analysis to identify conserved sequences (shaded boxes). Although over 1 Kb of each promoter was analyzed, significantly conserved sequences were found only along the first 500 bp upstream of the translation start site (TSS). The sequence with the highest degree of conservation (83% identity) is represented as a dark blue box and spans 50 bp (LIP1 element). (B) A binary plasmid (pPYRO) containing a minimal promoter fused to the luciferase reporter gene (control) was used to clone four copies of the 50 bp sequence (4xLIP1) and both constructs were used to produce A. thaliana transgenic plants. Luciferase activity was quantified in vivo from transgenic seeds 24 h after imbibition. Average values and standard errors from 10 independent lines for each construct (20 seeds/line) are shown. 

Identification of a TF binding to the LIP1 element

To screen the RR library, one copy of the LIP1 element was inserted into the pTUY1H plasmid (LIP1-pTUY1H) and introduced into S.cerevisiae Y187a. Leaky expression of the HIS3 reporter gene was titrated by using diploid cells obtained after mating the strain containing the LIP1-pTUY1H construct (Y187a) with the RR library strain (YM4271) containing a GFP-pDEST34 construct (AD-GFP). Growth of diploid cells was suppressed by using 1 mM 3-AT, a competitive inhibitor of the product of the HIS3 gene, and therefore this concentration was used in the RR library screen. Only one positive clone was identified in well H3 from library plate 3 (hereafter 3-H3). Cells derived from all wells from plate 3 were able to form diploids and grow at similar densities on diploid plates (Figure 2). However, diploid cells from 3-H3 were able to grow in media that selected for a positive DNA-protein interaction (screening plates + 1 mM 3-AT; Figure 2). To confirm the interaction, diploid cells from 3-H3 and 15-G9 (randomly selected negative control) were grown in liquid diploid media and similar number of cells were used to inoculate diploid plates and screening plates with increasing amounts of 3-AT. Diploid cells from 3-H3 and 15-G9 (both containing the LIP1-pTUY1H construct) produced colonies on diploid plates, but only those containing the 1xLIP1-pTUY1H and the 3-H3 ORF constructs were able to grow on screening plates even at 60 mM 3-AT, while growth of the negative control was completely blocked at 1 mM 3-AT (Figure 3). According to its position in the library plate, the ORF construct responsible for the activation of the HIS3 reporter gene contained a class IV homeodomain-leucine zipper gene (AtHDL, At4g21750). This was also confirmed by sequencing the ORF-pDEST34 construct present in the diploid cells from 3-H3. The same result observed for diploid cells was obtained when the experiment was repeated using haploid cells of both strains transformed with the appropriate plasmids: 1) The Y187a strain carrying the 1xLIP1-pTUY1H construct transformed either with the 3-H3 ORF or the 15-G9 ORF constructs. 2) The YM4271 strain carrying the 3-H3 ORF or the 15-G9 ORF constructs transformed with the 1xLIP1-pTUY1H construct. These results ruled out possible effects of the yeast genotype or ploidy on the interaction.

Binding of AtML1 to a L1-box present in the LIP1 element is abolished by a 2bp mutation and this interaction is relevant in planta

The AtML1 protein specifically binds to a motif with a conserved 6 bp core sequence (L1-box: 5-TAAATG-3’) and a
The L1-box abolishes both binding of AtML1 in vitro and reporter gene expression in transgenic plants [38]. We found that the L1-box contains a L1-box, a finding compatible with the positive interaction observed with AtML1 (Figure 4A). To demonstrate that AtML1 binds to the L1-box present in the L1-box element and in order to test the specificity of the Y1H system, we prepared a mutated version of this element containing two nucleotide changes in the core of the L1-box (L1-box, Figure 4A). The mutated and wild type constructs were introduced into a yeast strain containing the L1-box (AD-AtML1) or the AD-GFP (negative control) constructs and growth of the resulting transformants were scored on plates containing increasing concentrations of 3-AT. When using AD-GFP, yeast cells carrying the L1-box construct required 15 mM 3-AT to suppress the basal activity of the reporter gene instead of the 1mM required when they contained the L1-box construct (Figure 4B). However, when the AD-AtML1 construct was used in combination with the L1-box construct, growth was observed even at 100 mM 3-AT while growth of cells carrying the L1-box construct ceased at 15 mM 3-AT as observed for the control (Figure 4B). These results indicate that AtML1 binds specifically to the L1-box sequence present in the L1-box element.

To further validate the system, we carried out a screen to reproduce a well characterized DNA-protein interaction reported for the Phosphate Starvation Response 1 protein (PHR1; R1MYB TF) with the P1BS cis-element, a motif enriched in promoters of phosphate (Pi) starvation induced genes [12,29,39]. For this purpose, we used a conserved 50 bp promoter fragment (IPS1 element) from a Pi starvation induced gene (IPS1 containing the P1BS motif) that was identified by phylogenomic shadowing [12]. The promoter fragment was cloned into the pTU15H plasmid, introduced into S.cerevisiae Y187a cells and the resulting strain mated with the strain containing the AD-GFP construct. After titration of the basal expression of the HIS3 reporter gene, a screening was performed using appropriate plates without 3-AT and one strong positive was identified in well H7 from library plate 5 (hereafter 5-H7). The ORF construct responsible for the activation of the HIS3 reporter gene was confirmed by sequencing and, as expected, it was found to contain the PHR1 coding sequence (R1MYB; At4g28610). To confirm this interaction, similar numbers of diploid cells from wells 5-H7 (PHR1) and 5-H6 (a R1MYB gene not related to PHR1) were plated on plates without or with increasing amounts of 3-AT (Figure 4D). Diploid cells from wells 5-H7 and 5-H6 produced colonies on plates containing 3-AT while those containing the IPS1-pTU15H from the Y1H experiments. The negative control was blocked as soon as the histidine was removed from the media (Figure 4D).

The L1-box and IPS1 fragments are differentially bound by HD-ZIP and MYB-CC family members, respectively

AtML1 was the only positive obtained in our screening with the L1-box element. However, AtML1 belongs to the class IV HD-ZIP TF family that contains members known to bind to the L1-box sequence in vitro and in vivo [38,41-42]. In our screening, 1mM of 3-AT was used and weaker interactions of the L1-box element with other class IV HD-ZIP TFs may have been missed. We used a wider range of 3-AT concentrations to re-examine the ability to bind to the L1-box element of GL2 and HDG10, two HD-ZIP proteins present in our library that have different phylogenetic relationships with AtML1 (Figure 5A) as well as different loss-of-function phenotypes and/or expression patterns [41]. As can be seen in Figure 5B, only AtML1 was able to interact with the L1-box element even at 100mM 3-AT while GL2 and HDG10 did not activate reporter gene expression even in the absence of 3-AT.

### Table 1. ORFs showing differences with their corresponding sequences present in databases.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Family</th>
<th>bp</th>
<th>a.a.</th>
<th>Database</th>
<th>Regulators</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>At3g24520</td>
<td>HSF</td>
<td>993</td>
<td>330</td>
<td>bp</td>
<td>990</td>
<td>329</td>
</tr>
<tr>
<td>At2g33550</td>
<td>Trihelix</td>
<td>945</td>
<td>314</td>
<td>bp</td>
<td>936</td>
<td>311</td>
</tr>
<tr>
<td>At4g31620</td>
<td>B3</td>
<td>1479</td>
<td>492</td>
<td></td>
<td>1485</td>
<td>494</td>
</tr>
<tr>
<td>At5g49230</td>
<td>ZZ</td>
<td>636</td>
<td>211</td>
<td></td>
<td>621</td>
<td>206</td>
</tr>
<tr>
<td>At1g67580</td>
<td>MYB</td>
<td>900</td>
<td>299</td>
<td></td>
<td>997</td>
<td>190</td>
</tr>
</tbody>
</table>

Two base pair mutation in the L1-box abolishes both binding of AtML1 in vitro and reporter gene expression in transgenic plants [38]. We found that the L1-box element contains a L1-box, a finding compatible with the positive interaction observed with AtML1 (Figure 4A). The mutated and wild type constructs were introduced into a yeast strain containing the AtML1-pDEST TM (AD-AtML1) or the AD-GFP (negative control) constructs and growth of the resulting transformants were scored on plates containing increasing concentrations of 3-AT. When using AD-GFP, yeast cells carrying the L1-box constructs required 15 mM 3-AT to suppress the basal activity of the reporter gene instead of the 1mM required when they contained the L1-box construct (Figure 4B). However, when the AD-AtML1 construct was used in combination with the L1-box construct, growth was observed even at 100 mM 3-AT while growth of cells carrying the L1-box construct ceased at 15 mM 3-AT as observed for the control (Figure 4B). These results indicate that AtML1 binds specifically to the L1-box sequence present in the L1-box element.

To further validate the system, we carried out a screen to reproduce a well characterized DNA-protein interaction reported for the Phosphate Starvation Response 1 protein (PHR1; R1MYB TF) with the P1BS cis-element, a motif enriched in promoters of phosphate (Pi) starvation induced genes [12,29,39]. For this purpose, we used a conserved 50 bp promoter fragment (IPS1 element) from a Pi starvation induced gene (IPS1 containing the P1BS motif) that was identified by phylogenomic shadowing [12]. The promoter fragment was cloned into the pTU15H plasmid, introduced into S.cerevisiae Y187a cells and the resulting strain mated with the strain containing the AD-GFP construct. After titration of the basal expression of the HIS3 reporter gene, a screening was performed using appropriate plates without 3-AT and one strong positive was identified in well H7 from library plate 5 (hereafter 5-H7). The ORF construct responsible for the activation of the HIS3 reporter gene was confirmed by sequencing and, as expected, it was found to contain the PHR1 coding sequence (R1MYB; At4g28610). To confirm this interaction, similar numbers of diploid cells from wells 5-H7 (PHR1) and 5-H6 (a R1MYB gene not related to PHR1) were plated on plates without or with increasing amounts of 3-AT (Figure 4D). Diploid cells from wells 5-H7 and 5-H6 produced colonies on plates containing 3-AT while those containing the IPS1-pTU15H from the Y1H experiments. The negative control was blocked as soon as the histidine was removed from the media (Figure 4D).

The L1-box and IPS1 fragments are differentially bound by HD-ZIP and MYB-CC family members, respectively

AtML1 was the only positive obtained in our screening with the L1-box element. However, AtML1 belongs to the class IV HD-ZIP TF family that contains members known to bind to the L1-box sequence in vitro and in vivo [38,41-42]. In our screening, 1mM of 3-AT was used and weaker interactions of the L1-box element with other class IV HD-ZIP TFs may have been missed. We used a wider range of 3-AT concentrations to re-examine the ability to bind to the L1-box element of GL2 and HDG10, two HD-ZIP proteins present in our library that have different phylogenetic relationships with AtML1 (Figure 5A) as well as different loss-of-function phenotypes and/or expression patterns [41]. As can be seen in Figure 5B, only AtML1 was able to interact with the L1-box element even at 100mM 3-AT while GL2 and HDG10 did not activate reporter gene expression even in the absence of 3-AT.
A Novel Approach to Unravel Cis-Trans Regulation

Day 1 (Friday)

Frozen library plates

Replicator

Square plate (-W)

Frozen bait

Inoculate & grow 3 days at 28°C

Day 2 (Monday)

Replicator

96-well plate

Inoculate and incubate with shaking 24 hrs at 28°C

Day 3 (Tuesday)

Replicator

100μl YPAD

Multichannel pipette

Add 100μl of bait to each well & Incubate 48hrs at 28°C

Day 4 (Monday)

200ml YPAD

Replicator

MATING

Day 5 (Tuesday)

Use mated cultures to inoculate new 96-well plates containing 200 μl of -L-W media. Shake & grow for 24 hrs at 28°C

Day 6 (Wednesday)

Replicator

DIPLOID SELECTION

Day 7 (Thursday)

Day 8 (Friday)

Inoculate selection plates & incubate for up to 5 days

Day 9-10

Replicator

1 h/2 persons

Day 10-13

Diploids (-L-W)

Positives (-L-W-H+3-AT)
The screening with the IPS1 element rendered several positives, being PHR1 the strongest one. PHR1 is a R1MYB TF that belongs to the MYB-CC family known to contain 15 members (Figure 5C). According to the functional redundancy observed for members of the PHR1 subfamily [12], other TFs belonging to this subfamily were also able to bind to the IPS1 element and produced positive interactions in our screening. Re-examination of these interactions showed that only PHR1-like TFs belonging to the MYB-CC group 1 subfamily but not those from group 2 subfamily, were able to bind the P1BS element (Figure 5D).

These results suggest that our system is able to discriminate DNA binding specificities among different members of these TF families according to their functional redundancy in the plant.

Discussion

In this study we describe an effective approach to decipher DNA-protein interactions underlying transcriptional control in A. thaliana. An arrayed library (RR) of ca. 1,200 A. thaliana TFs was prepared in yeast and a matrix interaction screening procedure established. We demonstrate that functionally relevant promoter sequences identified by phylogenetic shadowing can be used to screen the RR library to isolate specific DNA binding proteins.

A conserved and functional promoter fragment from a GDSL-lipase gene (LIP1) highly induced upon germination was identified by phylogenetic approaches and in silico analyses with open access tools (see materials and methods). We used a collection of Brassicaceae species with different degrees of phylogenetic closeness with A. thaliana to cover a wide range of evolutionary differences and we have applied a new method for isolation of promoter regions based in gene order conservation (synteny). Because of this, we were able to amplify orthologous promoters only in the cases where synteny exists and, high conservation of the ATG and the coding sequence fragments adjacent to the promoter regions isolated, were taken as indicators of orthology [43–44]. In case that a wrong orthology had been assigned, this would not invalidate the conclusion on the likely relevance of the conserved boxes identified. Rather this would have potentially resulted in the non-identification of some relevant boxes, i.e., conserved among orthologous genes, but not conserved among closely related genes. Although there are some examples of exceptions in which considerable shuffling and alteration in number of binding sites in enhancers may occur among related species [7,45–49], it is clear that conserved motifs are more likely to be functionally relevant [7–12]. With the recent developments in DNA sequencing technology, an increasing number of plant genomes are being sequenced and annotated, thus avoiding the need for experimental promoter isolation and speeding up the discovery of conserved cis-regulatory elements.

During the generation of the REGULATORS collection, ca. 2% of the clones showed differences with predicted transcripts or annotated cDNA clones available in TAIR9 (http://www.arabidopsis.org). This was comparable to the discrepancies observed during a similar project conducted by Gong et al (2004) and indicates that experimental data for transcripts may still contribute to improve genome annotation. The REGULATORS collection, together with that generated under the REGIA project, used the Gateway-recombination method and contains ca. 1,200 TFs. It contains 469, 519 or 258 new ORFs when compared to the collections generated by Gong et al (2004), Mitsuda et al (2010) and Ou et al (2011), respectively, indicating that they are additive resources.

Y1H and Y2H systems can be used for high throughput studies of DNA-protein and protein-protein interactions. Yeast cells can be used as convenient eukaryotic test systems that require little specific optimization for each interaction compared to other approaches, and are more likely to provide an appropriate environment for interactions that depend on post-transcriptional modifications. Using arrayed TF libraries instead of pooled TF collections, reduce labour time since this eliminates the effort required to characterize several positives produced by the same clone. For instance, Mitsuda et al (2010) detected 72 positive interactions with a promoter fragment, being the same TF responsible for 39 of them. In our system, mating is carried out

---

**Figure 2. Flowchart for the yeast screening procedure for the arrayed TF library in 96-well format.** TF library and bait clones are grown on plates with their corresponding auxotrophic media. These plates were used to inoculate either 96-well plates (TF library; preys) or Erlenmeyer flasks (cis-element in pTUY1H; bait) containing YPAD and incubated overnight. Bait and preys were then mixed and incubated for 48 h without shaking to allow mating. Mated cells were used to inoculate another set of 96-well plates containing diploid selection media (DOB-L-W). After incubation for 24 h, diploid cells were replicated onto diploid and screening (DOB-L-W-H ±3-AT) plates. Positives were visible after 2 to 5 days of growth. Hours of labour per person are indicated for each step of the protocol.
doi:10.1371/journal.pone.0021524.g002

**Figure 3. Yeast one hybrid screening with the LIP1-pTUY1H construct.** Growth of diploid cells at different concentrations of 3-AT from clones showing positive (3-H3) and negative (15-G9) interactions in the screening. Three serial dilutions of diploid cells from saturated cultures were plated.
doi:10.1371/journal.pone.0021524.g003
in liquid media so that diploid and screening plates are inoculated with similar numbers of cells and grown and scored in parallel, allowing eventual non-mating clones to be flagged as not screened. Diploid colony size can be taken into account to compare and normalize the strength of positive interactions. Moreover, the 96-well plates containing the diploid cells can be stored at 4°C and re-spotted at any time on different types of screening plates, for instance containing hormones or other chemicals, to re-evaluate positive and negative interactions from the initial screening. The simplicity of the procedure, offers the possibility to easily perform these screens with reduced labour and time. Also, more complex matrix interaction schemes involving several baits can be performed [27,50]. As an added value, this library constitutes a convenient tool for the plant community since it could also be used for Y2H and Y3H approaches. In fact, the RR library has been successfully used in two hybrid screenings, by using a pDEST™32-ORF clone in the S.cerevisiae P694α as bait, to reproduce previously published interactions [51-52].

We have shown that a single copy of a discrete promoter fragment can be used with an episomal plasmid in Y1H screenings without compromising specificity. It has been common to Y1H experiments to use large promoter fragments or generate tandemly repeated promoter sequences that need to be integrated into the yeast genome. Using large promoter fragments requires performing additional promoter deletions/mutations and experiments to pinpoint the exact sequence bound by the TF identified. Moreover, S.cerevisiae genome is more compact than that of A.thaliana and it is known that for UAS located over 300bp upstream of a reporter gene, transcription initiates proximally to the UAS and competes with that derived from the reporter gene located downstream [53]. For instance, this could explain why Brady et al (2011) carried out Y1H matrix assays between 167 TFs and 65
promoters (3 kb) mainly expressed in the stele and they only detected positive interactions for 16 promoters. Also, Mitsuda et al. (2010) used 500 bp promoter fragments for Y1H screenings with a pooled TF library and found that only one out of 2 positive interactions seemed to be biologically relevant when tested in planta. Their results suggest that such promoter fragments may be missing information contained in upstream parts of the promoters used and adding noise to the system by increasing the probability of having yeast TF derived positives.

Compared with the use of core cis-elements (typically 6–8 bp length), small promoter fragments such as those identified by phylogenomic shadowing, may allow discrimination between DNA binding specificities among different members of a TF family and the identification of several TFs binding to different target sequences, while focusing on a small part of the promoter likely to be involved in its regulation. We have uncovered a novel interaction between a lipase promoter and AtML1, a class IV HD-ZIP protein. AtML1 binds to a L1-box, a motif that is also bound in vitro by class IV HD-ZIP proteins [41]. However, the L1-box sequence present in the LIP1 element exclusively interacted with AtML1 but not with two other class IV HD-ZIP proteins expected to have different functions in the plant (Figure 5B). Also, several

Figure 5. HD-ZIP subfamily IV and MYB-CC proteins show differential binding capabilities. (A) Phylogenetic tree of HD-ZIP class IV TF proteins from A. thaliana constructed using the Phylogeny.fr platform. TFs used in part (B) are indicated by green and red colored circles. (B) Yeast strains containing either the LIP1 element (WT) or a 2 bp mutation in the L1-box (mut), were mated to strains containing the AD-AtML1, AD-GL2 or AD-HDG10 constructs. Diploid cells were grown on screening plates with increasing concentrations of 3-AT. Only the AtML1 protein is able to activate the reporter gene by binding to the WT element. (C) Phylogenetic tree of MYB-CC TF proteins from A. thaliana modified from a tree published elsewhere [12]. A dotted line separates Group 1 from Group 2 subfamily members characterized by having the MYB-CC domain at C or N-terminal position, respectively. TFs used in part (D) are indicated by green and red colored circles. (D) A yeast strain containing the IPS1 element was mated to strains containing the AD-PHR1 or AD-MYB-CC TFs. Diploid cells were grown on diploid and screening plates. Only PHR1 and PHR1-like proteins belonging to Group 1 were able to activate the reporter gene. Three serial dilutions of diploid cells were spotted.

doi:10.1371/journal.pone.0021524.g005
TFs belonging to the PHR1 subfamily were able to bind to the IPS1 element (Figure 5D) according to their functional redundancy and phylogenetic relatedness [12].

Our results suggest that approaches such as the one presented here, which includes a phylogenetic shadowing-based motif identification step, may be more restrictive and specific in detecting relevant DNA-protein interactions. Identifying target DNA binding motifs and their partners is crucial to understand their function and the approach reported here will facilitate future studies on transcriptional regulation of gene expression.

Materials and Methods

Cloning orthologous LIP1 gene promoters from several Brassicaceae species and phylogenomic analyses

For the phylogenomic analysis of the LIP1 gene (At5g45670), a collection of 18 Brassicaceae species, that represents a wide evolutionary range inside this family, was used (Figure S4). Using this collection for promoter amplification of orthologous LIP1 genes, we first designed 32–33 mers oligonucleotide mixtures to be used as primers (LIP1forward and LIP1reverse; Table S2). These oligonucleotides mixtures were derived from the conserved protein coding regions of the LIP1 gene and the upstream gene flanking the LIP1 promoter in A. thaliana (At5g45660). For each primer, the oligonucleotide mixture reflected all sequence variants encoding the conserved regions as found in A. thaliana, Brassica oleracea, Brassica napus and Brassica napus. In addition, the 12 nts at the 3'-end of the primer mixture included all possible combinations, according to the genetic code, encoding the corresponding 4 amino acids of the conserved protein sequences. For all the promoter amplifications the Expand long template PCR system (Roche) was used. Cycling conditions were 5 min at 95°C, 10 cycles of 10 s at 95°C, 5 min 30 s at 68°C followed by 20 cycles of 15 s at 95°C, 5 min 30 s + 10 s/cycle at 68°C. Out of the Brassicaceae species in which clear PCR amplification products was observed, five of them displaying different degrees of phylogenetic closeness with A. thaliana were selected for sequencing. PCR products were separated by agarose electrophoresis, purified (QIA quick gel extraction kit, Qiagen) and cloned using the pCRII-TOPO TA Cloning Kit (Invitrogen).

Alignment of promoter sequences was performed by using DiAlign (http://www.genomatix.de/cgi-bin/dialign/dialign.pl) [54], a program with a high resolving power in the detection of conserved blocks. High conservation of the ATG and the coding sequence fragments adjacent to the promoter regions isolated were taken as indicators of orthology [43–44].

The phylogenetic tree of the HD-ZIP class IV proteins was performed on the Phylogeny.fr platform (www.phylogeny.fr) [55] as previously described [12]. The default substitution model was selected assuming an estimated proportion of invariant sites (0.076) and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma = 0.321). Nodes with bootstrap value <30 were collapsed. Only the conserved DNA binding domain was used for alignment and tree construction.

Generation of constructs for plant transformation and in vivo imaging of transgenic plants

A 969 bp promoter fragment from the lipase gene was amplified by PCR from A. thaliana (Columbia ecotype) genomic DNA using the LO843 and LO848 primers (Table S2) and cloned into the pGEM®-T Easy vector (Promega). This construct was used as a template to amplify a 50 promoter sequence (LIP1 element, -194-145 in Figure 1A) using the primers LO846 and LO847 that contained Bgl2 and BamH1/Hind3 restriction sites, respectively (Table S2). To introduce four tandemly repeated LIP1 elements into plants, we used a promoterless binary plasmid containing a luciferase (LUC+) encoding sequence (pPRO) [56]. First, we introduced a GSTF9 minimal promoter with primers -58F9-F (-58F9 and Hind3) and -58F9-R (SalI) containing engineered restriction sites. The fragment was digested with Bgl2 and SalI (filled in after digestion) and cloned into pPRO digested with BamH1 and Hind3 (filled in after digestion). Then, the engineered Bgl2 and Hind3 sites of the LIP1 element were used to clone it into the BamH1 and Hind3 sites of the 58F9-pPRO plasmid upstream of the minimal promoter and the LUC+ gene (1xLIP1)-58F9-pPRO. This process was repeated three more times to generate a 4xLIP1-58F9-pPRO plasmid. The -58F9-pPRO and 4xLIP1-58F9-pPRO constructs were introduced into A. thaliana (Col-0 ecotype) using the floral-dip method and the Agrobacterium tumefaciens strain GV3101 [56-59]. Seeds from these transgenic plants were sown on MS agar plates containing 50 μM of luciferin and stratified for 2 days (4°C in the dark). Luciferase activity was recorded in vivo after 24 h of imbibition at a constant temperature of 22°C and a photoperiod of 16 h light/8 h dark. In vivo imaging and quantification was performed using a CCD camera (Hamamatsu) and the provided software.

To generate the 3SS:AtML1 construct, the AtML1 ORF was transferred from the corresponding yeast library clone (AD:AtML1) to the pDONRTM221 by recombination (BP clonase; Invitrogen). The AtML1 ORF was then recombined (LR clonase; Invitrogen) into the pEarleygate201 plasmid [60] and the resulting construct checked by sequencing. Transient expression analyses were performed according to [61], except that bombarded leaves were sprayed with a 5mM luciferin solution 15–30 min before imaging.

Generation of the pTUY1H plasmid and cloning of LIP1 and IPS1 elements

To generate pTUY1H, a fragment encompassing the ARS6/ Centromeric region and the LEU1 auxotrophy marker was amplified from the pDESTTM32 vector (Invitrogen) using the oligonucleotides yeast-fw and yeast-rv (Table S2). The fragment was subcloned into pPCR2.1-TOPO (Invitrogen) and a BamHI- EcoRV fragment that contained the full cassette was obtained by partial digest. The fragment was ligated into the pHIS-E vector (Clontech) after BamHI and PvuII digestion. We used the 4xLIP1-58F9-pPRO construct as a template to amplify the LIP1 element by PCR using the primers LO1205 and LO1204 that contained Xmn1 and XhoI restriction sites, respectively (Table S2). Then, the engineered restriction sites of the LIP1 element were used to clone it into the Xmn1 and XhoI sites of the pTUY1H plasmid upstream of the HIS3 reporter gene. The same restriction sites of the pTUY1H plasmid were used to clone a 50 bp fragment from the IPS1 gene that contained a P1BS sequence. This fragment was produced by annealing complementary single-stranded oligonucleotides (IPS1Y1H5 and IPS1Y1H3; Table S2) that generated Xmn1 and XhoI 5'- cohesive ends.

Generation of a TF ORF collection in a Gateway compatible entry vector

286 TF ORFs from A. thaliana were amplified by 2-step nested PCR and cloned into a Gateway compatible plasmid (pDONRTM221). In the first PCR, ORF specific primers containing half of the attB sequences at the 5'-ends (Table S2) were mixed with Pfx polymerase buffer (2x), 1 mM MgSO4,
of each primer, 0.2 units of Pfx polymerase (Invitrogen) and the number of cycles (35). PCR products were subjected to electrophoresis, purified (PEG/MgCl₂ solution, Invitrogen or QiA quick gel extraction kit, Qiagen) and cloned into the pDONR™221 plasmid (Invitrogen) using the BP clonase mix (Invitrogen). Integrity and quality of every ORF was checked by sequencing using M13 forward and reverse primers (Table S2).

Generation of a yeast normalized library of A. thaliana TFs

Entry clones (pDONR™221-TF ORF) were recombined to the pDEST™22 plasmid using the Gateway LR enzyme (Invitrogen) to generate GALAD-ORF fusions. Reactions were prepared by mixing 25 ng Entry-clone, 75 ng pDEST™22, 0.5 µl LR buffer and 0.4 µl LR clonase and adjusting the volume up to 2.5 µl with TE. After overnight incubation in thin-walled 96-well tubes, 10 µl of yeast library efficient chemically competent cells (E.coli DH5α; Invitrogen) were added and transformed with the heat-shock method (15 s, 42°C) and resupended in 1 ml of 1M lithium acetate (LiAc) pH 7.5. Yeast cells were transferred to 1.5 ml eppendorf tubes, pelleted by centrifugation (5 min, 4000 rpm) and then resuspended in 1 ml of yeast cultures with 100 µl of 1M sorbitol and plated on 6 cm selection plates (minimal medium minus tryptophan, DOB-W; MP Biomedicals) plates from their corresponding frozen stocks. A 1 L Erlenmeyer flask containing 200 ml of YPAD medium was inoculated with a clump of bait cells (5–10 colonies) and incubated overnight at 28°C with shaking (200 rpm). In parallel, 100 µl of YPAD was aliquoted into 96-flat bottom well plates (Corning) by using a multichannel pipette and a replicator was used to inoculate them with their corresponding library colonies. After overnight incubation with vigorous shaking (500 rpm) at 28°C (HiGro shaker; Genemachines), 100 µl of the bait culture was added to each well of the 96-well plates with a multichannel pipette and mating was allowed 48 h by incubating at 28°C without shaking. Settled cells were resuspended by hitting the bottom of the wells with the pins of the replicator and used to inoculate another set of 96-flat bottom well plates containing 200 µl of diploid selection media (DOB-L-W). The replicator was able to transfer about 5 µl of liquid in each pin and each well was inoculated twice (1/20 dilution), always sterilizing by flaming with absolute alcohol between plates. After one day of growth at 28°C and vigorous shaking, diploid cells were resuspended with the replicator and spotted onto diploid selection and screening (DOB-L-W-Histidine ±3-Amino-1,2,4-triazole; 3-AT, Sigma) plates. Positive colonies were visible after 2 to 5 days of growth at 28°C.

Supporting Information

Figure S1 Alignment of LIP1 proximal promoter regions from different Brassicaceae species. Sequences from Arabidopsis thaliana, Sisymbrium irio (HQ322377), Sinapis arvensis (HQ322378), Capsella rubella (HQ322379), Descurainia sophia (HQ322380) and Brassica oleracea (HQ322381) were aligned using DiAlign (http://www.gnomatix.de/cgi-bin/dialign/dialign.pl) [54]. A conserved region among LIP1 orthologs is shadowed in grey and the L1-box included in this region is highlighted in yellow. GenBank accession numbers are shown in parenthesis. (TIF)

Figure S2 Construction of an Arabidopsis TF collection and generation of an arrayed yeast library. A collection of TF ORFs (REGULATORS) was generated by PCR using a high fidelity polymerase and nested primers, and recombined into the pDONR™221 plasmid. The REGULATORS and REGIA collections were recombined into the pDEST™22 plasmid and introduced into yeast. Clones were arrayed separately in 96-well plates and maintained as glycerol stocks at −80°C. (TIF)

Figure S3 Map of the pTUY1H plasmid. Modifications made on the pHISi-1 backbone are shown: ARS6/Cen4 centromeric region, LEU auxotrophy marker and the multicloning site. (TIF)

Figure S4 Brassicaceae species used for the amplification of LIP1 orthologous gene promoters. (A) Table containing the tribes and species of the Brassicaceae collection. Species marked with the letter A inside a red circle were used for the phylogenetic analysis of LIP1. (B) Syntopic diagram of phylogenetic relationships of Cleomaceae and the various tribes of the Brassicaceae, adapted from [62]. The tribes in red represent the
ones used in the promoter regions isolation of different orthologs of the LHY gene. (TFH)

Table S1 List of TF ORFs in the REGULATORS collection. (XLS)

Table S2 List of oligonucleotides used in the paper. (XLS)

Table S3 Full coding sequences corresponding to ORFs that differ with their annotated sequences in databases. (DOC)

Table S4 Sequencing results of randomly picked RR library clones. (XLS)

Acknowledgments

We thank Immaculada Gude and Elke Kemper for excellent technical assistance and Angel Garcia for building the 96 pins replicator. We are grateful to Drs. Salomé Prat, Begoña Benito and Rosario Haro for helpful advice during the set up of the yeast screening procedure and also to Regla Bustos for helping with the construction of the phylogenetic trees. We also thank Dr. Karam Singh and the ABRC for providing the yPRO and the pEarleyGate201 plasmids, respectively, the ABRC and Dr. Imre Somssich for providing cDNA clones, and the REGIA consortium for the REGIA plasmid library. We thank Drs. Bernd Weisshaar and Julio Salinas for providing access to the original sequencing data of the REGIA library and to a Hamamatsu CCD camera, respectively. We also acknowledge Vincent Thareau for bioinformatic analyses of the REGIA collection and Jean-Philippe Tamby for curating the REGULATORS web site.

Author Contributions

Conceived and designed the experiments: JP-A G. Coupland PC ML. Performed the experiments: G. Castrillo FT ML. Analyzed the data: G. Castrillo FT ML. PC G. Coupland. Contributed reagents/materials/analysis tools: G. Castrillo FT ML. LO-S. Wrote the paper: LO-S PC FT G. Coupland.

References