Research

In silico analyses of pericycle cell populations reinforce their relation with associated vasculature in Arabidopsis

Boris Parizot1,2, Ianto Roberts1,2, Jeroen Raes1,2,†, Tom Beeckman1,2,‡ and Ive De Smet3,*,‡

1Department of Plant Systems Biology, VIB, Technologiepark 927, 9052 Ghent, Belgium
2Department of Plant Biotechnology and Genetics, Ghent University, Technologiepark 927, 9052 Ghent, Belgium
3Division of Plant and Crop Sciences, School of Biosciences, University of Nottingham, Loughborough LE12 5RD, UK

In Arabidopsis, lateral root initiation occurs in a subset of pericycle cells at the xylem pole that will divide asymmetrically to give rise to a new lateral root organ. While lateral roots never develop at the phloem pole, it is unclear how the interaction with xylem and phloem poles determines the distinct pericycle identities with different competences. Nevertheless, pericycle cells at these poles are marked by differences in size, by ultrastructural features and by specific proteins and gene expression. Here, we provide transcriptional evidence that pericycle cells are intimately associated with their vascular tissue instead of being a separate concentric layer. This has implications for the identification of cell- and tissue-specific promoters that are necessary to drive and/or alter gene expression locally, avoiding pleiotropic effects. We were able to identify a small set of genes that display specific expression in the phloem or xylem pole pericycle cells, and we were able to identify motifs that are likely to drive expression in either one of those tissues.

Keywords: pericycle; tissue specificity; cell identity; lateral root initiation; Arabidopsis

1. INTRODUCTION

The Arabidopsis primary root presents a radial pattern of tissue types that are arranged in concentric layers: epidermis, ground tissues (cortex and endodermis) and vascular cylinder (pericycle, xylem and phloem) [1]. This organization finds its origin in the embryonic development of the root with periclinal (radial) and anticlinal (circular) cell divisions giving rise to new tissue layers and expanding these tissue layers, respectively. The first rounds of periclinal and anticlinal divisions within the proembryo give rise to an inner cell mass surrounded by epidermal precursors. Further divisions of these inner cells lead to the formation of central vascular precursors encircled by a layer of ground tissues. The pericycle is first observed in between the globular and heart-shaped embryo stage, where the provascular tissue divides into outer pericycle and inner diarch cores of xylem and phloem vascular precursors. Finally, divisions of the ground tissues complete the patterning of the root, with the creation of an inner layer of endodermis and an outer layer of cortex [2–4].

Within the pericycle tissue, genes involved in various physiological processes are expressed: thioredoxin [5], biosynthetic enzymes or gene transcripts involved in the initial and final steps of tropane alkaloid biosynthesis [6,7], sulphate transport [8], cytokinin biosynthesis [9,10], morphine biosynthesis [11], ascorbate homeostasis [12], xylem loading of potassium and boron [13,14], etc. But, most importantly, a subset of pericycle cells is giving rise to a new lateral root, which is associated with spatially and temporally controlled gene expression in these cells [15,16].

(a) Distinct pericycle cell populations

Even though pericycle cells initiate from the same subset of initial cells as the rest of the stele, which present a diarch symmetry, their regular shape and radial organization around the vascular tissues made them traditionally regarded as a concentric homogeneous layer [1]. A growing number of reports, however, emphasize differences between pericycle cells according to their position adjacent to the xylem or the phloem poles, such as cell division potential [17], cell cycle progression [18], cell size [17,19], cell surface arabinogalactan–protein distribution and cell wall thickening [20–22], plasmodesmal connections...
[23,24,27], microtubule content [25], cytoplasm content [26,27], marker gene expression [27–31] and differential gene transcription [32,33].

Marker lines have always been useful to address cell identity and, in this respect, a collection of enhancer trap lines has been extremely helpful [27,28,34]. For example, the marker line J2661 shows an expression pattern generalized to the entire pericycle [35], whereas two other marker lines, J0121 and S17, show expression in the pericycle cell populations associated with the xylem and the phloem poles, respectively [28,30]. Using another marker line, i.e. RM1007, it is already possible to observe pericycle differentiation in at least two different cell populations in the heart-stage embryo [27]. In this marker line, GFP expression starts in the pericycle cells which are in contact with the quiescent centre and which are later associated with the xylem poles, indicating that bilateral symmetry of the pericycle and genetic distinction between two different populations of cells occurs as early as the formation of the pericycle cells [27]. Consequently, these observations indicate that there is most probably no time frame in plant development where the full pericycle consists of a highly homogeneous set of cells [27].

(b) Xylem pole pericycle-associated lateral root initiation in Arabidopsis

Another particularity distinguishing pericycle cells adjacent to the xylem poles in comparison with those adjacent to the phloem poles is their competence for lateral root initiation. Even though lateral root initiation has been extensively studied and many molecular players were discovered in the past decade [34,36–39], the mechanisms controlling xylem pole pericycle (XPP) cell identity and competence for lateral root initiation are largely unknown.

The strong association between lateral root initiation and the xylem pole is supported by the lonesome highway (lhw) mutant, which displays only one xylem pole and one-sided lateral root formation [27]. Nevertheless, a differentiated protoxylem element within the xylem pole is not required as impaired vasculature development (ivad) and arabidopsis histidine phosphotransfer protein 6 (ahp6) mutants still display lateral root initiation [27]. On the other hand, the wooden leg (wol) mutant lacks phloem and all pericycle cells have xylem pole-associated identity, but it hardly produces lateral roots [27]. However, these wol pericycle cells respond to auxin with respect to their division potential [27]. At present, it is unclear how the interaction with xylem and phloem pole determines the distinct pericycle identities with differing competence. The study of mutants concomitantly impaired in vasculature, pericycle and lateral root development could, however, show that the determination of both the pericycle and the vasculature is regulated by a common genetic pathway [27].

c) Cis-regulatory elements

The temporal and spatial transcriptional activity of genes is—to some extent— influenced by cis-regulatory elements [40,41]. Transcription factors will bind in a sequence-specific manner to cis-regulatory elements in promoters of genes (through activation or repression) or act through protein–protein interactions to regulate their gene expression [40]. In Arabidopsis, there are more than 60 families containing more than 2000 genes encoding for transcription factors [42–44]. Various methods are available to identify or confirm cis-regulatory elements and their interaction with particular transcription factors in silico, in vitro and in vivo [45], but this has only limitedly been addressed in the context of distinct expression patterns in the root [46].

(d) Genome-wide cell-specific transcript data of the Arabidopsis root

Recently, gene expression data for various tissues and cell types in Arabidopsis became available through the ‘digital in situ’ [32,47,48]. Cell sorting and transcriptome analysis of various tissue-specific marker lines, including J2661, J0121 and S17, allowed the comparison of the transcriptome of these cell types with each other’s but also with that of the other layers of the root [32,33].

We propose here a simple approach to use the microarray data—generated on these cell-sorted marker lines in combination with other available data—to establish transcriptional differences between the xylem- and the phloem-pole-associated pericycle and the other tissues of the root, and to reveal a subset of genes that display a differential or an enriched expression in those populations of pericycle cells. Genes with such a specific expression pattern will be crucial in the future to target gene expression to a particular group of cells. In addition, we identified motifs that likely drive XPP- or phloem pole pericycle (PPP)-specific gene expression.

2. MATERIAL AND METHODS

(a) Microarray retrieving and normalization

The following microarray hybridization files were retrieved from the gene expression omnibus database: GEO series GSE8934 (S17, S32, COBL9, J0121, S4, SUC2, J2501, RM1000), GEO series GSE7641 (PET111, AGL42), GEO Series GSE5749 (LRC, GL2, SCR, J0571, WOL) and GEO series GSE16468 (CORTEX, APL, S18). J2661 hybridization files were kindly provided by the author [35]. Normalizations of the respective datasets concerning only the pericycle (J2661, J0121 and S17) or the entire root (J2661, J0121, S17, S32, COBL9, S4, SUC2, J2501, RM1000, PET111, AGL42, LRC, GL2, SCR, J0571, WOL, CORTEX, APL and S18) have been performed using the robust multi-array average method [49]. Affymetrix probesets to Arabidopsis Genome Initiative ID assignment was performed using the a f f y _ A T H 1 _ array_elements-2010-12-20.txt file downloaded from The Arabidopsis Information Resource (TAIR; www.arabidopsis.org) for 21 107 probesets corresponding to unique AGI IDs.

(b) Differential expression of genes in between pericycle cell populations

Differential analysis was performed using the moderated t-test [50] within the ‘affy lmGUI’ vignette for R package. ‘Present’, ‘marginal’ and ‘absent’ calls were generated using the MAS5 algorithm in the ‘affy’ vignette for R package [51]. For the first approach, a gene was considered as being differentially expressed if it could fulfil
Fold changes, or 'absent' flags in the other microarray experiments. If it showed three 'present' flags in one of the two microarray comparisons, a fold change of at least 2 and a maximum $p$-value of 0.01 in at least one of the microarray triplicates (J2661, S17, J0121; the following conditions: at least two 'present' calls in at least one of the microarray triplicates (J2661, S17, J0121; the selected population was individually compared with its expression value in each other selected layer (figure 2a). A $p$-value was evaluated from a $t$-test with Welch assumption of the variance, which was performed in between values of the different replicates for the selected population against all the replicate values of all other selected layers. A gene was declared significantly and specifically enriched in one population of cells if its expression level was at least two times higher or lower in comparison with each other individual selected cell layer with a maximum $p$-value of 0.01. For each candidate, an average enrichment was proposed by calculating the ratio between its expression value in the considered pericycle or primordia cell population against the average expression in the other layers. In case a gene could be selected to be specific of two cell populations, a best hit was finally proposed by selecting the best ratio.

**Microarray clustering and enrichment analysis**

The clustering was generated using the MeV [54] hierarchical clustering module, with a Pearson correlation distance metric and complete linkage clustering. We used normalized logarithmic expression values extracted from the previously described compendium of microarray data for all genes and for the marker

**Gene ontology enrichment analysis for the genes significantly upregulated in the XPP or the PPP,**

<table>
<thead>
<tr>
<th>GO_acc</th>
<th>description</th>
<th>n</th>
<th>$q$-value</th>
<th>GO_acc</th>
<th>description</th>
<th>n</th>
<th>$q$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0006412</td>
<td>translation</td>
<td>233</td>
<td>9.90e-10</td>
<td>GO:0019761</td>
<td>glucosinolate biosynthetic process</td>
<td>11</td>
<td>0.000014</td>
</tr>
<tr>
<td>GO:0009059</td>
<td>macromolecule biosynthetic process</td>
<td>289</td>
<td>1.40e-07</td>
<td>GO:0019760</td>
<td>glucosinolate metabolic process</td>
<td>11</td>
<td>0.000014</td>
</tr>
<tr>
<td>GO:0034645</td>
<td>cellular macromolecule biosynthetic process</td>
<td>289</td>
<td>1.40e-07</td>
<td>GO:0016144</td>
<td>S-glucoside biosynthetic process</td>
<td>11</td>
<td>0.000014</td>
</tr>
<tr>
<td>GO:0042254</td>
<td>ribosome biogenesis</td>
<td>78</td>
<td>1.90e-07</td>
<td>GO:0016143</td>
<td>S-glucoside metabolic process</td>
<td>11</td>
<td>0.000014</td>
</tr>
<tr>
<td>GO:0022613</td>
<td>ribonucleoprotein complex biogenesis</td>
<td>78</td>
<td>1.90e-07</td>
<td>GO:0019758</td>
<td>glucosinolate biosynthetic process</td>
<td>11</td>
<td>0.000014</td>
</tr>
<tr>
<td>GO:0010467</td>
<td>gene expression</td>
<td>310</td>
<td>2.70e-05</td>
<td>GO:0019757</td>
<td>glucosinolate metabolic process</td>
<td>11</td>
<td>0.000014</td>
</tr>
<tr>
<td>GO:0044249</td>
<td>cellular biosynthetic process</td>
<td>364</td>
<td>2.10e-05</td>
<td>GO:0016796</td>
<td>sulphur metabolic process</td>
<td>22</td>
<td>0.000051</td>
</tr>
<tr>
<td>GO:0039058</td>
<td>biosynthetic process</td>
<td>371</td>
<td>3.90e-05</td>
<td>GO:0044272</td>
<td>sulphur compound biosynthetic process</td>
<td>16</td>
<td>0.000083</td>
</tr>
<tr>
<td>GO:0044267</td>
<td>cellular protein metabolic process</td>
<td>349</td>
<td>7.90e-05</td>
<td>GO:0034637</td>
<td>cellular carbohydrate biosynthetic process</td>
<td>24</td>
<td>0.00064</td>
</tr>
<tr>
<td>GO:0019538</td>
<td>protein metabolic process</td>
<td>361</td>
<td>0.00013</td>
<td>GO:0016051</td>
<td>carbohydrate biosynthetic process</td>
<td>30</td>
<td>0.0017</td>
</tr>
<tr>
<td>GO:0044085</td>
<td>cellular component biogenesis</td>
<td>105</td>
<td>0.0035</td>
<td>GO:0016137</td>
<td>glycoside metabolic process</td>
<td>18</td>
<td>0.0025</td>
</tr>
<tr>
<td>GO:0042260</td>
<td>cellular macromolecule metabolic process</td>
<td>443</td>
<td>0.007</td>
<td>GO:0016138</td>
<td>glycoside biosynthetic process</td>
<td>18</td>
<td>0.0025</td>
</tr>
<tr>
<td>GO:0044238</td>
<td>primary metabolic process</td>
<td>540</td>
<td>0.011</td>
<td>GO:0044262</td>
<td>cellular carbohydrate metabolic process</td>
<td>46</td>
<td>0.0044</td>
</tr>
<tr>
<td>GO:0043170</td>
<td>macromolecule metabolic process</td>
<td>458</td>
<td>0.011</td>
<td>GO:0005975</td>
<td>carbohydrate metabolic process</td>
<td>65</td>
<td>0.0067</td>
</tr>
<tr>
<td>GO:0088152</td>
<td>metabolic process</td>
<td>586</td>
<td>0.062</td>
<td>GO:0006950</td>
<td>response to stress</td>
<td>228</td>
<td>0.01</td>
</tr>
</tbody>
</table>

**Figure 1.** Pericycle cell populations. Expression profiles of the marker lines used for the microarray experiments and the pairwise comparisons: (a) entire pericycle (XPP + PPP), (b) xylem pole pericycle (XPP) and (c) phloem pole pericycle (PPP). Gene ontology enrichment analysis for the genes significantly upregulated in the (d) XPP and (e) PPP. The symbolized vascular tissues are xylem (Xy) and phloem (Ph).

The following conditions: at least two ‘present’ calls in at least one of the microarray triplicates (J2661, S17, J0121; figure 1), a fold change of at least 2 and a maximum $p$-value of 0.01 in at least one of the three pairwise comparisons (J0121 versus J2661, J0121 versus S17 and J2661 versus S17) and a similar fold change sign for all three pairwise comparisons. For the second approach, a gene was considered as being differentially expressed if it showed three ‘present’ flags in one of the two microarray experiments J0121 and S17 and three ‘marginal’ or ‘absent’ flags in the other microarray experiments. Fold changes, $p$-values and flag calls are available in the electronic supplementary material, table S1. Gene ontology enrichment analysis was performed using the Parametric Analysis of Gene set Enrichment (PAGE) algorithm [52] with the Agri Gene Ontology (Agrigo) analysis toolkit [53], with default settings, on the list of genes significantly upregulated in the XPP or the PPP, with an average fold change calculated from the three pairwise comparisons (see the electronic supplementary material, table S1).

**Specific up- or downregulation of genes in pericycle populations in comparison with the other layers of the root**

For each of the four different cell populations (pericycle and primordia), the expression value average in the selected population was individually compared with its expression value in each other selected layer (figure 2a). A $p$-value was evaluated from a $t$-test with Welch assumption of the variance, which was performed in between values of the different replicates for the selected population against all the replicate values of all other selected layers. A gene was declared significantly and specifically enriched in one population of cells if its expression level was at least two times higher or lower in comparison with each other individual selected cell layer with a maximum $p$-value of 0.01. For each candidate, an average enrichment was proposed by calculating the ratio between its expression value in the considered pericycle or primordia cell population against the average expression in the other layers. In case a gene could be selected to be specific of two cell populations, a best hit was finally proposed by selecting the best ratio.

**Microarray clustering and enrichment analysis**

The clustering was generated using the MeV [54] hierarchical clustering module, with a Pearson correlation distance metric and complete linkage clustering. We used normalized logarithmic expression values extracted from the previously described compendium of microarray data for all genes and for the marker
The global analysis. FC, fold change.

Figure 2. Specific layer gene expression. For each tested cell population, a list of genes specifically up- or downregulated was computed by calculating the ratio between the expression value in this cell population (dark grey) with each other selected layer (light grey). (a) Global analysis involving most root tissues. (b,c) Pericycle and vasculature analysis excluding, respectively, the vasculature and the pericycle tissues. (d) Number of genes up- or downregulated in the different cell populations characterized in the global analysis. FC, fold change.

(e) Motif identification

To identify statistically over-represented motifs in the DNA sequences 500 bp upstream of the start codon for the xylem and phloem pericycle-specific genes, we used YMF v. 3.0 (http://wingless.cs.washington.edu/YMF/YMFWeb/YMFInput.pl) [55]. Upstream DNA sequences were retrieved from TAIR (http://www.arabidopsis.org/tools/bulk/sequences/index.jsp; electronic supplementary material, data). Motif size was set at 6, 7 and 8 with a maximum of 0 spacers in the middle and a maximum of 1 degenerate symbols [R (purine – A or G), Y (pyrimidine – C or T), W (A or T) and S (C or G)]. On the list of significant motifs generated by YMF v. 3.0, we applied FIND-EXPLANNER v. 1.1.2 to extract a smaller set of ‘real’ motifs [56]. The significance of a motif was measured by the z-score of its count in the input sequences. If a motif occurs N times in the input sequences, but was expected to occur E times (with standard deviation of S) in random sequences of the same length (generated by a suitable background model), then the z-score of the motif is \( (N - E) / S \). In addition, to confirm the 6 and 8 bp hits from YMF v. 3.0, we used WEEDER v. 1.3 (http://159.149.109.9/modtools/) [57] using standard settings on the two strands. Finally, we used PSCAN (http://159.149.109.9/pscan/) to identify potential transcription factors or transcription factor families that can bind to 500 bp upstream DNA sequences of xylem and phloem pericycle-specific genes.

3. RESULTS

(a) Differential expression of genes in between pericycle cell populations

To access the differential expression of genes within pericycle cell populations, we used two different approaches based on the same dataset. Firstly, we have performed a pairwise comparison of three microarray experiments performed on the full pericycle (figure 1a), on the XPP (figure 1b) and on the PPP (figure 1c). On the basis of a minimum fold change of 2 and a maximum
Figure 3. Cell identity. (a) Hierarchical clustering of the microarray related with pericycle and vascular cell populations. (b) Enrichment analysis of vascular genes in the different populations of pericycle cells. Over-representation of genes concomitantly regulated is expressed as a percentage.

$p$-value of 0.01 (see §2), we could generate a list of genes showing higher expression in the xylem pole (1357 genes) or in the phloem pole (1763) pericycle (electronic supplementary material, table S1). These numbers already show that the two populations of pericycle cells present important differences at the transcriptomic level. An analysis of the gene ontology enrichment on these lists shows that the pericycle associated with the xylem poles presents a high enrichment of genes involved in primary metabolism processes such as translation, biosynthetic processes and ribosome biogenesis, which makes sense in view of its potential to divide (figure 1d). The population of cells constituting the pericycle associated with the phloem pole rather show an enrichment in secondary metabolites processes (figure 1e). Secondly, we have made a restrictive gene selection based on ‘present’, ‘marginal’ and ‘absent’ presence detection call flags between the XPP (figure 1b) and the PPP (figure 1c) microarray experiments (see §2). We found 1567 genes specifically present in the phloem pole and absent in the XPP population. In contrast, we found only 99 genes being present in the xylem pole and absent in the PPP populations (electronic supplementary material, table S1). These two lists, however, did not yield obvious significant gene ontology enrichment.

(b) Gene expression enrichment in the pericycle cell populations

To unravel the specificity of the pericycle cell populations in comparison with the other cell layers of the root, we generated different lists of genes specifically up- or downregulated in these layers. All the radial data of the Arabidopsis high-resolution root spatiotemporal map [32] were gathered together to form a compendium representative of the main layers of the primary root. We created four different lists of genes specifically up- or downregulated in the overall pericycle (P), in one of the two xylem or phloem pole-associated populations (respectively, XPP and PPP), or in the lateral root primordia. Genes were declared specifically regulated in one of these cell populations if their expression level was at least two times higher or lower in comparison with every other individual selected cell layer with a maximum $p$-value of 0.01 (figure 2a; see §2). A first result appearing out of generating these four different gene lists is that a higher number of genes is specifically upregulated in comparison with genes specifically downregulated in one cell layer (figure 2d and the electronic supplementary material, table S2). Also, we could isolate a higher number of genes specifically upregulated in the PPP cell population in comparison with the XPP and the whole pericycle cell populations (this assertion remains true with different parameters of fold change and $p$-value tested; data not shown). To generate a final and restricted gene list, with high confidence in their XPP- or PPP layer-specific enrichment, we intersected this last analysis with the previous detection call flags analysis. A smaller list of genes could be generated, out of which 63 genes are specifically enriched in the PPP and not expressed in the XPP and seven genes are in contrast specifically enriched in the xylem pole and not expressed in the phloem pole (electronic supplementary material, table S2). Our first approach identified, for instance, GATA23 as a XPP enriched gene, which is supported by its reported expression pattern [33]. Owing to our stringent criteria, GATA23 was not retained in the final list, which nevertheless increases confidence in this list.

(c) Pericycle population transcriptomes reflect their bilateral association with neighbouring vascular tissues

The intimate association of the pericycle cell populations with their neighbouring vascular tissues also raises the question of their genetic connection. To compare their transcriptomic regulations, we performed a hierarchical clustering of the related microarray datasets. Interestingly, clustering the microarrays of the
different layers constituting the stele (xylem maturing, developing xylem, protophloem, phloem and companion cells, companion cells) with the microarrays of the XPP- and PPP strikingly shows an intimate association of each pericycle cell population with its neighbouring vascular tissue (figure 3a). To reinforce this observation, we conducted an enrichment analysis by analysing the number of genes that are specifically and commonly enriched in a cell population of the pericycle (figure 2b) together with a layer of the vascular tissue (figure 2c), and in comparison with other layers of the root (see §2). The enrichment was calculated as a percentage of the observed number of genes concomitantly enriched in a pericycle population together with a vascular tissue in comparison with the expected number of such genes based on the individual percentage of genes enriched in one or the other of these tissues (figure 3b). We observed a significant over-representation of genes commonly regulated in groups composed of a pericycle cell population and its neighbouring vascular tissues. Specifically, there is an over-representation of genes concomitantly regulated in the XPP and in the developing xylem (figure 3b). Similarly, there is an over-representation of genes concomitantly regulated in the PPP and a phloem tissue (mainly phloem together with the companion cells and companion cells; figure 3b).

(d) Motifs for pericycle population-specific gene expression

The identification of common motifs driving expression of XPP- or PPP-specific genes is useful to gain insight into the regulatory network of these cell populations and as a tool for targeted gene expression. While motifs can also be over-represented in intragenomic conserved non-coding sequences [58,59], we focused here on the 500 bp upstream regulatory sequence (see §2). We restricted our analyses to 500 bp as this length is often sufficient to confer expression patterns similar to longer upstream fragments [60] and is routinely used for the identification of regulatory elements enriched in promoters of target genes [61,62].

To verify whether specific motifs could be responsible for XPP- and PPP-specific expression, we searched within the final and restricted list of XPP- and PPP-specific genes for particular motifs that might be involved in regulating the tissue specific expression. We used \textit{YMF} v. 3.0 on the DNA sequences 500 bp upstream of the start codon of the putative XPP- and PPP specifically enriched genes (see §2), and this resulted in the identification of several putative motifs with a 6-, 7- and 8-nucleotide length (electronic supplementary material, table S3). Several of these could be confirmed using \textit{Weeder} v. 1.3 on the same set of upstream sequences (electronic supplementary material, table S3).

To determine whether particular transcription factors or transcription factor families are likely to bind these promoters we used \textit{PsCAN}, which revealed some interesting candidates for future analyses (electronic supplementary material, table S3). However, combining this analysis with the list of XPP- or PPP-enriched genes did not yield any obvious common candidates. Interestingly, \textit{At5g01740}, which lies close to the position of the enhancer element in J0121, is present in the restricted list of genes specifically expressed in the XPP and not expressed in the PPP. A more detailed analysis of the region showed that some identified motifs are present in this region, supporting our motif analyses (figure 4).

4. DISCUSSION
(a) Pericycle cells are intimately associated with their vascular tissue in Arabidopsis

Only a limited percentage of genes were shown to be specifically enriched in either PPP (1.5%) or XPP (4.5%) in comparison with the 21 107 genes taken into consideration on the ATH1 microarray. Notwithstanding only 23–40% of plant genes are predicted to have a strict tissue-specific expression [63], this is still surprisingly low. Still, these results are in agreement with the spatio-temporal expression map of the root ([32]; the electronic supplementary material, table S2).

Our analyses show that the pericycle and associated vasculature share a high overlap in gene expression, explaining why only very few genes are strictly enriched in the pericycle cell type. Both the clustering of the microarray and the enrichment analysis show a link
between the XPP and the developing xylem. This connection is also apparent from, for example, the expression pattern of *ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEIN 6* (AHP6), namely at the protoxylem position and in the two associated pericycle cells [29]. Interestingly, there is a less obvious link with the maturing xylem, which is a tissue undergoing deep differentiation processes leading to programmed cell death [64]. The situation is different for the PPP, which shares much more common upregulation with the cells constituting the mature phloem and its companion cells or the companion cells themselves. Considering upregulations, the PPP shares more genes with its neighbouring tissues compared with the number of genes shared by the XPP with the early associated developing xylem tissue. This observation would indicate that if both PPP and XPP populations are related with their neighbouring vascular tissues, then this link is higher for the PPP. These results are in agreement and complementary with a recent analysis characterising a stele-enriched gene regulatory network in the *Arabidopsis* root [32,65]. Among 11 distinct dominant transcription patterns in the stele tissues, the authors could find a pattern with PPP gene expression associated with the phloem poles but not for the XPP with the xylem. Our analysis is not restricted to the most dominant patterns and could show such a link even though a lot less strong than for the PPP. Interestingly, the authors also described two dominant patterns of genes expressed exclusively in the pericycle, which are, respectively, in front of one or the other vascular pole, but not in the overall layer [65], pointing to a similar conclusion that the pericycle presents a more heterogeneous fate than expected for a concentric tissue layer. Furthermore, the observation that xylem shares more identity with a precursor of its physically neighbouring tissue fits with the vision of the XPP as an extended meristem [18,27]. In this case, the identity of the XPP should be closer to a young tissue rather than a highly differentiated one.

**b) Cis-regulatory elements for xylem and phloem pole-associated pericycle**

Over the years, there has been a necessity to use a XPP-specific marker, but hardly any expression patterns with specificity for XPP have been described. The most popular one available is the GAL4 enhancer trap line J0121 and this marker has been used as a XPP cell identity marker [27,66,67], for XPP-specific transactivation [28,39,68,69], and for cell sorting [34]. But, up to date, it is still unclear which gene causes this specific expression pattern in J0121. With respect to the PPP, the promoter from the gene *At5g01740* was shown to give PPP-specific expression (referred to as S17) [30]. This gene is present in the list of genes we describe as being differentially expressed in between the xylem and the phloem pole-associated pericycle and fit the corresponding expression pattern (flags indicate an absence from the XPP and a presence in the PPP). However, this gene was not described as being enriched in the PPP ([32] and this study). It remains to be elucidated if the S17 transcripational fusion [30] does not reflect the entire expression pattern of the gene *in planta*, or if the microarray probesets of this gene present specificity problems.

Obviously, one gene expected within our dataset is the putative gene responsible for the specific expression in J0121. Recently, the position of the enhancer element in J0121 was determined to be inserted between two genes *At5g01740* and *At5g01750*, but the relation—if any—to lateral root initiation remains to be determined [28]. Interestingly, *At5g01740* is present in the restricted list of genes specifically enriched in the xylem pole and not expressed in the PPP, indicating a potential role of this gene in XPP cell fate. Its annotation as a member of the nuclear transport family makes it an interesting candidate for further functional analysis. *At5g01750*, on the other hand, was found to be differentially expressed in the PPP in comparison with the XPP, but also in other layers of the root. This observation indicates high chances for the promoter region of *At5g01740* to enhance the J0121 marker line expression pattern.

For uniform and strong control of expression in different ecotypes and for possible translation of this knowledge to crops, the identification of cell-type-specific promoters or motifs that drive expression locally is essential. Here, we identified a number of cis-regulatory elements for XPP- and PPP-specific gene expression, which could be used for various applications, and we identified putative regulatory transcription factors *in silico*. Obviously, this might be oversimplified as multiple cis-regulatory elements might be required to integrate signals from multiple transcription factors resulting in combinatorial control and highly specific patterns of gene expression [40]. In future, experimentally validating if these single elements are sufficient to drive specific *in planta* expression in distinct pericycle populations will provide an answer.

These elements can be used to create synthetic promoters to drive expression of a (trans)gene under temporal and spatial conditions, avoiding the pleiotropic effects owing to broader expression. In a systems biology approach, these cis-regulatory elements (and their binding transcription factors) can be used to identify genes belonging to the same regulatory pathways and to understand how transcriptional regulation mechanisms work and how spatial and temporal expression networks are established during organogenesis and maintained throughout the life cycle. In addition, these elements can be used as baits—for instance, using a yeast one-hybrid approach—to discover their interacting transcription factors [65,70].

Ive De Smet is supported by a BBSRC David Phillips Fellowship (BB/BB/H022457/1), a Marie Curie European Reintegration grant (PERG06-GA-2009-256354) and the Research Foundation Flanders (FWO09/PDO/064 A 4/5 SDS), Boris Parizot and Ianto Roberts are supported by the Research Foundation Flanders (FWO, grant no. 3G002911). Ianto Roberts is indebted to the Agency for Innovation by Science and Technology in Flanders (IWT). This work was supported by a grant from the Interuniversity Attraction Poles Programme (IUAP VI/33), initiated by the Belgian State, Science Policy Office.

**REFERENCES**


*Phil. Trans. R. Soc. B* (2012)


