Inference of the Conservation of the Genetic Network Underlying Vascular Development between the Dicotyledon Species *Arabidopsis thaliana* and the Monocotyledon Species *Hordeum vulgare* (Barley)

Background and Contents

Though incomplete, the construction of a hypothetical genetic network underpinning vascular development in dicotyledon *Arabidopsis thaliana* has been achieved. This work has much application potential in the improvement of cereal crops. The question of whether this genetic network is deeply conserved in plants such that the orthologues of said genes retain the same role in vascular development amongst monocotyledon cereal plants has yet to be explored.

This project marks a beginning of work that will attempt to answer this question using the model cereal crop Barley (*Hordeum vulgare*) – as a diploid species with a sequenced genome, it is more suited to genetic modification and is likely to have less orthologues in need of investigation than other cereal models. Work includes both bioinformatic and laboratory techniques, laid out in two separate sections in this report. This document is paired with an excel spreadsheet titled *Bioinformatics Project Barley Information* which contains detailed information of the genes under study, referred to frequently as ES(x), (x) being the spreadsheet sheet number of focus. References include links to bioinformatics databases and online tools. Any work needed to finish this current project is laid out in the final section.

The starting information for this project can be found at ES(1), which contains Arabidopsis gene and promoter names and IDs of all those hypothesised to play a role in vascular development, according to the Peter Etchells Laboratory, University of Durham Biosciences Department.

Bioinformatics

1. Identification of Orthologues in Barley

Any Barley orthologues already identified from the ES(1) list of genes were located and downloaded using the Biomart data mining tool at ENSEMBL [1]. ES(2) contains all Barley orthologues identified using this data mining tool. Percentage identity and orthology confidence level were also recorded for each orthologue. Many Arabidopsis genes had no known orthologue in Barley with this database, and some had multiple. Genes without an identifiable Barley orthologue are listed in ES(3).

In the interest of future work, *Oryza sativa* (Rice) orthologues to the Arabidopsis genes of interest are also listed in ES(5). The monocotyledon rice genome has been more completely characterised, and therefore may provide a halfway point between Arabidopsis and Barley for further genetic network construction and orthologue identification.

2. Selection of those Orthologues with Expression in Vascular Tissue

ES(4) contains a table of Arabidopsis genes with known Barley Orthologues and where they are expressed. Expression data was originally intended to be obtained from the plant biology resource BAR[2], however as of yet the ePlant tool does not have access to expression data of Barley, unlike other model cereals.

Expression data was instead located from the Expression Atlas open science resource[3], which contains data from an RNA-seq study of coding RNA of eight barley tissues from different developmental stages[4]. The eight tissues are as follows:

- Caryopsis (15 dpa)
- Caryopsis (5 dpa)
- Germinating Embryo
- Inflorescence (1cm)
- Inflorescence (5mm)
- Internode
- Root (Seedling)
- Shoot (Seedling)

Expression pattern of each orthologue was recorded in ES(4), and their level of candidacy for a role in Barley vascular development is recorded adjacently, from 0 to 5 (highest). It is important to note that this candidacy is purely derived from expression data of a single study, and therefore cannot be relied upon to be highly accurate. Within this first project, it is advisable for candidacy score to be solely used to eliminate those orthologues from the list of candidates whose expression does not occur in the vascular tissues of either the shoot or internode – those with a candidacy rating of 0. Those highlighted in yellow are the strongest candidates according to the expression data – potential starting genes of future work.

3. Construction of a Basic Barley Genetic Network upon which to Build

Selection of Barley orthologues to be included in a basic genetic network for a starting hypothesis was undergone with the information collected and displayed in ES(4), and the Cytoscape PXY Arabidopsis network constructed by the Peter Etchells Laboratory (*Figure 1*).

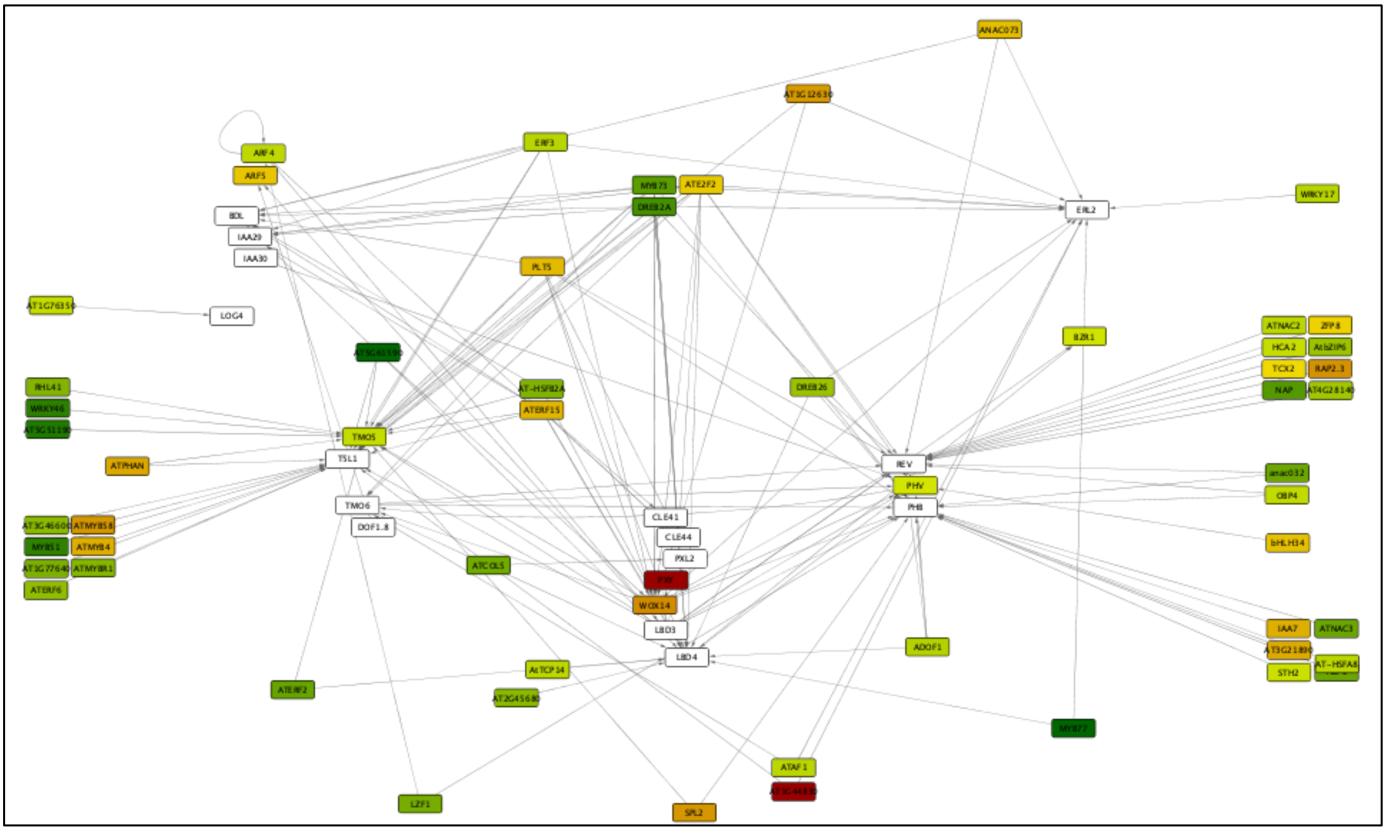


Figure 1 - *Cytoscape PXY Arabidopsis network constructed by the Peter Etchells Laboratory*

This network was reduced to those Arabidopsis genes which have two or more interactions. The list was then further reduced to those genes with Barley orthologues identified (ES(2)) which showed vascular tissue expression (ES(4)). The Barley orthologues were then inserted into the network, replacing their Arabidopsis orthologues. This hypothesis of a basic Barley network (*Figure 2*) and the list of its gene components (*Table 1*) are shown below, and can also be found in ES(6).

Table 1 *– A list of genes that are components of a hypothesised basic genetic network underlying vascular development in Hordeum vulgare, with Ensembl Gene Stable IDs and common name.*

Common Name	*Ensembl Gene Stable ID*
DREB2A	HORVU6Hr1G050520
DREB2A	HORVU1Hr1G060490
PLT5	HORVU3Hr1G089160
ARF4	HORVU3Hr1G072340
REV	HORVU4Hr1G090030
REV	HORVU1Hr1G041790
PXY	HORVU7Hr1G072690
ERL1	HORVU7Hr1G012920
LBD3	HORVU4Hr1G002480
ARF5	HORVU5Hr1G015560
ARF5	HORVU3Hr1G022060
WOX14	HORVU3Hr1G080690
DREB26	HORVU7Hr1G035440
DREB26	HORVU4Hr1G061340
PHV	HORVU0Hr1G010250
LBD4	HORVU4Hr1G002480
T5L1	HORVU5Hr1G066530
ATERF15	HORVU4Hr1G000700
ATERF15	HORVU1Hr1G077680
PHB	HORVU0Hr1G010250
CLE41	HORVU1Hr1G007660
CLE44	HORVU1Hr1G007660
PXL2	HORVU6Hr1G005460
PXL2	HORVU4Hr1G079040
IAA30	HORVU1Hr1G017770
IAA29	HORVU6Hr1G091260
TM05	HORVU4Hr1G061760

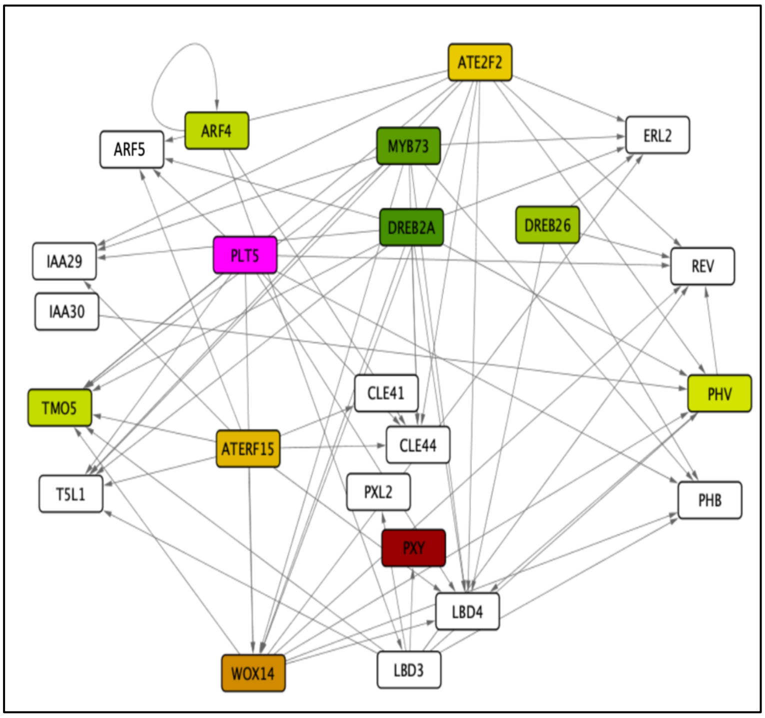


Figure 2 *- Hypothesis of a basic genetic network underlying vascular development in* Hordeum vulgare.

There were a few key genes in this network which had no known barley orthologue in online databases, but which had too central a position in the Arabidopsis network to be removed from the hypothetical Barley network:

- PXY
- ARF4
- ATE2F2
- PLT5
- CLE44
- CLE41
- MYB7

Putative orthologues of these genes were found using BLAST search of the Barlex databases (input sequences obtained from TAIR database)[5,6]. Unfortunately, a few of the results from these searches were not close enough in identity to be a likely orthologue for the desired Arabidopsis gene. Ensembl Plants[7] was used to check the most similar Barley genes' function was similar to expected, and to see if the result was an orthologue of a different Arabidopsis gene. Therefore, the PLT5 orthologue is actually that of PLT2, the ARF4 orthologues are actually that of ARF3, and the sequence closest to TDIF (potential orthologues of CLE44 and CLE 41) is not similar enough in sequence to have the correct functional capacity of PXY binding. However, the ATE2F2 and PXY orthologues found using this technique are likely to be correct.

4. Finding Novel Barley Genes Involved with Coexpression Analysis

Coexpression analysis of Barley transcriptomic data can locate novel genes with a role in the vascular development genetic network. Data from the PLANEX[9] plant coexpression database was analysed, containing *Hordeum vulgare* GeneChip data from the Gene Expression Omnibus in NCBI[10]. PLANEX only recognises Affymetrix probe and NCBI gene IDs, and so conversion of the Ensembl Barley gene IDs was necessary. This was carried out using PLANEX's BLAST tool, with sequence inputs from Ensembl Plants, as common conversion tools are yet to gain the ability to convert Barley gene IDs, due to its incomplete genome characterisation. Unfortunately, many of the genes recognised by Ensembl did not have a PLANEX gene ID equivalent. In addition, many of the genes had no representative Contig ID which allow its capture and coexpression analysis by Affymetrix Barley GeneChip. A list of the key genes with their PLANEX Gene IDs conversions is shown below (*Table 2*), also in ES(7).

Table 2 - *A list of genes that are components of a hypothesised basic genetic network underlying vascular development in* Hordeum vulgare *with their PLANEX Gene Stable ID conversions and Affymetrix Probe Representatives from the Barley GeneChip.*

Common Name	Ensembl Gene Stable ID	PLANEX Gene Stable ID	Affymetrix Probe Representative
DREB2A	HORVU6Hr1G050520	TC277099	Contig13435_at
DREB2A	HORVU1Hr1G060490	TC249456	Contig6727_at
PLT5	HORVU3Hr1G089160	---	---
ARF4	HORVU3Hr1G072340	TC279040	---
REV	HORVU4Hr1G090030	TC268085	---
REV	HORVU1Hr1G041790	TC281349	---
PXY	HORVU7Hr1G072690	TC272835	---
ERL1	HORVU7Hr1G012920	TC258766	Contig21274_at
LBD3	HORVU4Hr1G002480	---	---
ARF5	HORVU5Hr1G015560	---	---
ARF5	HORVU3Hr1G022060	TC247757	---
WOX14	HORVU3Hr1G080690	TC244391	Contig20612_at
DREB26	HORVU7Hr1G035440	TC274216	---
DREB26	HORVU4Hr1G061340	---	---
PHV	HORVU0Hr1G010250	TC269073	---
LBD4	HORVU4Hr1G002480	---	---
T5L1	HORVU5Hr1G066530	TC246976	Contig25949_at
ATERF15	HORVU4Hr1G000700	---	---
ATERF15	HORVU1Hr1G077680	---	---
PHB	HORVU0Hr1G010250	TC269073	---
CLE41	HORVU1Hr1G007660	TC246622	---
CLE44	HORVU1Hr1G007660	TC246622	---
PXL2	HORVU6Hr1G005460	TC265098	Contig19894_at
PXL2	HORVU4Hr1G079040	---	---
IAA30	HORVU1Hr1G017770	TC257532	---
IAA29	HORVU6Hr1G091260	TC239820	Contig15125_at
TM05	HORVU4Hr1G061760	TC240250	Contig6814_at

Coexpression analysis of those genes with representative contigs (highlighted above) was undergone through the PLANEX coexpression tool, which mines data from over 700 different Barley GeneChip arrays. Coexpression is measured quantitively using the Pearson Correlation Coefficient (PCC), +1 representing a strong positive, coexpression relationship and -1 representing a strong negative coexpression relationship. Coexpressed genes were filtered to those of either above 0.75, or below -0.75 PCC to reduce the field to more likely candidate genes, listed in ES(8).

Analysis was undergone on these lists to find common coexpressed genes – those strongly positively or negatively coexpressed with two or more of the original key network genes. Three genes were identified as being coexpressed with 2 different network genes (*Table 3*). These genes represent strong candidates for involvement in the vascular development genetic network in Barley. Contig IDs were further characterised using the Affymetrix NetAffx tool[11], to provide predicted functional information.

Table 3 – *Genes identified as strongly coexpressed with two or more genes that are components of a hypothesised basic genetic network underlying vascular development in* Hordeum vulgare. *Predicted function and orthology information about each is also displayed.*

Coexpressed Key Genes	PLANEX Gene Stable ID	Affymetrix Probe Representative	Gene Ontology Accession/ Locus	Predicted Function	Orthologues
DREB2A; TC277099 WOX14; TC244391	TC259922	Contig6473_at	GO:0004523 AK248295	RNA-DNA hybrid ribonuclease activity. Putative mitochondrial processing peptide.	
T5L1; TC246976 PXL2; TC265098	TC267018	Contig19644_at	GO:0020037 AK367178	Peroxidase precursor - hydrogen peroxide catabolic process, response to oxidative stress etc.	AT1G71695
T5L1; TC246976 PXL2; TC265098	TC273129	Contig4194_at	GO:0016787 AK249727	Patatin like glycoprotein. Storage and phospholipase activity - cleaves fatty acids from membrane lipids.	AT2G26560

These 3 genes are candidates for further research in future. In addition, one gene coexpressed with T5L1 was found to have likely orthology with a phloem specific transcription factor in *Oryza sativa* (Rice) (below), known to have a function in rice vascular development. This therefore represents an additional candidate for the Barley genetic network that has yet to be characterised.

T5L1; TC246976	TC263580	Contig20292_at	RF2A common name (characterised in 19 species)	bZIP Transcription Factor. Rice orthologue is phloem specific promoter known to function in vascular development (PUBMED 9311985)	AF005492 (Oryza Sativa). No known orthologue in Arabidopsis.

Laboratory

1. Growing Barley WT and Mutants

Two batches of two different genotypes of Barley were grown at separate times: WT and PXY mutant. Timing was staggered so that analysis of the Barley transcriptome could be undertaken for different developmental times within the plant, as well as for different genotypes.

2. RNA Extraction and mRNA capture of Barley Vascular Tissue

RNA extraction of three 21 day old and three 57 day old PXY mutants, and three 21 day old and three 57 day old WT plants was carried out manually with mortar and pestle after freezing with liquid N_2. Tissue dissected from the young Barley plants was taken from above the hypocotyl and just below the first leaf, including the cotyledon in the sample. Tissue dissected from the older Barley plants was taken from above the hypocotyl and just below the first leaf, in the hope of including stem tissue in the sample, rather than leaf. All of these sampled plants were not yet at inflorescence stage.

mRNA capture was performed on these six samples using Vortex SeraMag Olgio (dT) Coated Magnetic particles[12], with LBB, WBA, WBB and LSB buffer washes, and resuspension with cDNA reaction mix for PCR. The cDNA was then disassociated form the beads and stored in -20 °C.

The process of RNA extraction and mRNA capture was repeated for three 72 day old PXY mutants, and three 72 day old WT plants. This additional developmental stage was investigated as the sampled Barley plants had entered the inflorescence stage. In monocotyledons, this stage cooccurs with the growing of its major stem tissue between the curled leaf shafts.

3. Primer Design for Transcription detection of Key Genes

Primers were designed for each gene in the basic network generated (*Table 1*). The coding sequence for each gene was obtained from Ensembl Plants[7], and primers were generated using the Primer3 online tool[8]. Pairs of primers were picked to be near the polyadenylated tail of the mRNA of interest (often more effective, as whole mRNAs can be fragmented in the extraction process), and between 50 to 120 bp long (in order for the ability to run them on an agarose gel, if there is a need to check that the PCR run worked correctly). The sequences of primers ordered are listed in ES(6). The same primers are needed for PHV and PHB, CLE41 and CLE44, and LBD3 and LBD4, as they share the same hypothesised orthologue in the Barley genome.

4. Detecting Expression of Key Genes with qPCR

Primers were resuspended and diluted to 10uM and qPCR reaction mixes were prepared for each gene of interest. Barley Actin8 primers were used as a control for later analysis normalisation. The first qPCR section analysed the 21 day old and 57 day old Barley samples. Each tube contained:

- 4.0 uM cDNA Barley Sample
- 10 uM Syber Reaction Mix (blue)
- 0.5 uM Forward Primer
- 0.5 uM Reverse Primer
- 2.0 uM RNAse free H_2O

36 tubes were run for each gene, each 9 well section below containing three biological replicas:

- 9 21-day Barley WT
- 9 21-day Barley PXY mutant
- 9 57-day Barley WT
- 9 57-day Barley PXY mutant

qPCR standard protocol was utilised for all samples, and quantitative data results was exported to excel for further analysis. The technique was performed with triplicates of biological replicas to allow the checking of consistency between Cq results for three different plants, which allowed the checking of consistency between melting temperature results. If large inconsistency was observed, the qPCR was performed an additional time – only trends with consistency were recorded. Calculations of means between replicas and then genotypes were undergone in order to construct a graph of trends for the 4 conditions' Cq results normalised to HvACT8 results (Barley Actin 8). The initial attempt showed inconsistent results, and so volume of cDNA input was doubled to 4ul which produced better results. This volume was then maintained throughout all qPCR runs. Genes analysed so far via PCR are as follows:

- WOX14
- PXY
- PLT5
- LBD3 (LBD4 same orthologue)
- PHV (PHB same orthologue)
- T5L1

See ES(9) for all Cq results. Graphed expressions of the four conditions are shown below (Figure 3).

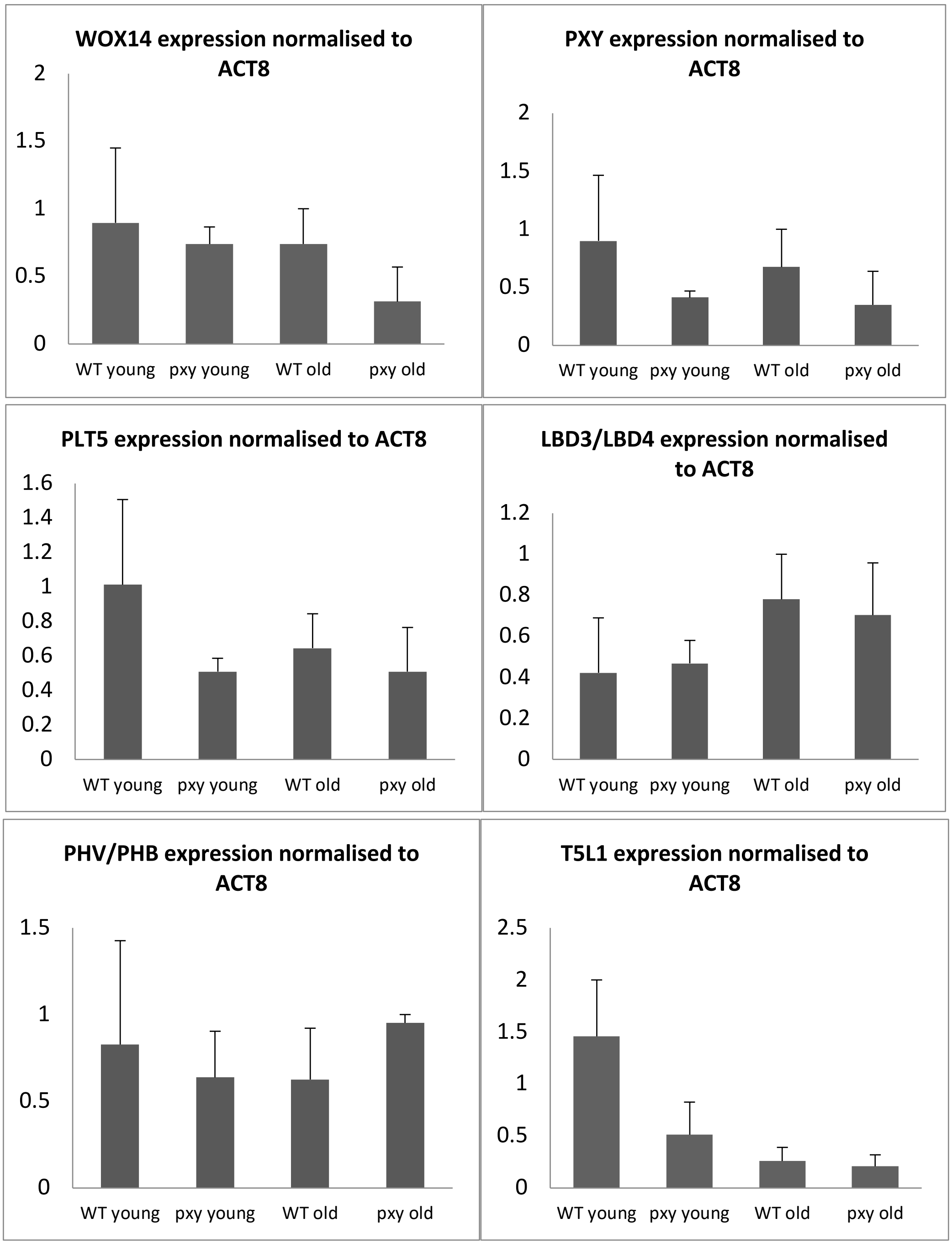


Figure 3 *– Graphs displaying expression level of six different genes of interest under four conditions: WT 21 day old, PXY 21 day old, WT 57 day old and PXY 57 day old Barley stem samples. These levels are based on triplicate qPCR runs of 3 Barley stem samples under each condition, normalised to their Actin8 expression level.*

In the scenario that the genes of interest do in fact participate in the genetic network underlying vascular development in Barley, a predicted effect of knocking out PXY would be a decrease in expression level of the gene of interest compared to WT expression. This trend, though not always significant, can be seen in many of the qPCR results. Regarding 21 day old Barley, a decrease in expression on PXY knockout can be observed for the following: WOX14, PLT5, PHV/PHB, T5L1. Regarding 57 day old Barley, a decrease in expression on PXY knockout can be observed for the following: WOX14, PLT5, LBD3/LBD4, T5L1. The only genes not following this trend are LBD3/LBD4 (in 21 day old Barley) and PHV/PHB (in 57 day old Barley). As PXY is the gene being knocked out, a significant decrease in its expression was expected/required (see above) for the other gene's results to be valid.

The second qPCR section analysed the six 72 day old Barley stem samples of plants in inflorescence developmental phase. The same genes were analysed, with the exception of PHV, and the protocol undergone was identical to above, each 9 well section below containing three biological replicas:

- 9 72-day Barley WT
- 9 72-day Barley PXY mutant

See ES(10) for all Cq results. Graphed expressions of the four conditions are shown below (Figure 4).

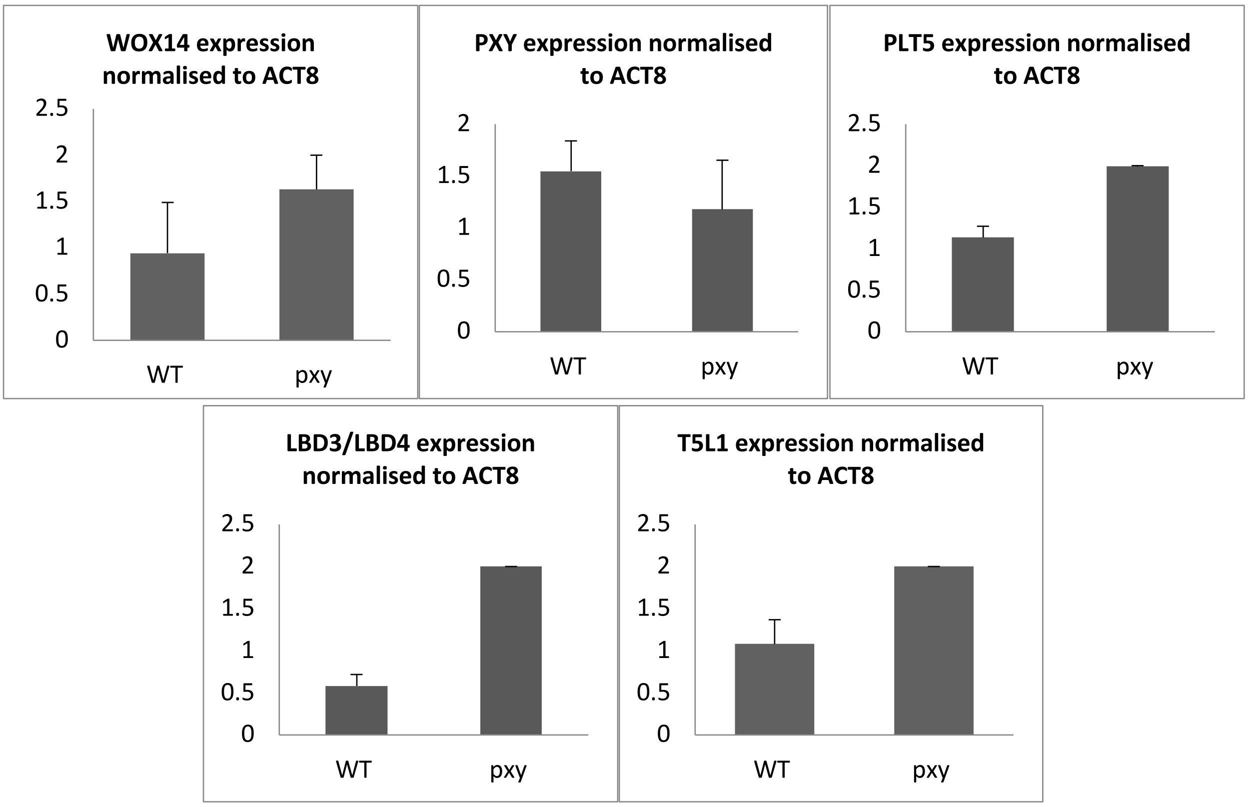


Figure 4 *– Graphs displaying expression level of five different genes of interest under two conditions: WT inflorescence stage and PXY inflorescence stage Barley stem samples. These levels are based on triplicate qPCR runs of 3 Barley stem samples under each condition, normalised to their Actin8 expression level.*

The predicted trend of a reduction in expression on PXY knock out is not observed in the flowering Barley qPCR results. Instead, an increase in expression on PXY knockout can be observed for all genes investigated: WOX14, PLT5, LBD3/LBD4 and T5L1 (all of which are significant increases, with the exception of WOX14).

There are many reasons why these results show an opposite trend. As well as human error, this may represent a difference in developmental timing between WT and PXY mutants at 71 days old, which would confuse gene expression level trends. A way to investigate this would be to perform RNA-seq that addressed all gene expression changes that occur between the WT and PXY mutants at this stage. Another reason may be a redundancy in the genetic vascular development network of Barley plants in inflorescence stage that therefore is not dependent on PXY expression, that is not present in the Arabidopsis network.

However, as PXY is the gene being knocked out, a significant decrease in its expression is required for the other gene's results to be valid. Though a reduction can be observed, it is not significant. Therefore, the results, though suggestive, cannot be considered as conclusive.

Conclusions

The genes showing significant reduction of expression in the 21 day old and 57 day old developmental stages are PLT5 and T5L1 on PXY knockout. This project's qPCR results therefore suggest a role for these two genes in a genetic network requiring PXY protein as a key component. As all other genes investigated show a reduction of expression in either of the developmental stages on PXY knockout, non can be ruled out of a role in the vascular development genetic network of Barley.

More broadly, these results, along with coexpression data mined from previous Barley RNA-seq studies, are suggestive of genetic network conservation between dicotyledons and monocotyledons. This is crucial for the effective application of work on Arabidopsis vascular development genetic components on cereal crops. However, additional research is needed to provide further evidence for trends found in this project.

Work Needed to Finish Current Project

1. qPCR

The following genes' expressions have not yet been analysed via qPCR. Their order in the list represents their hypothesised importance in the genetic network and therefore the order in which to analyse them:

- PHV (PHB same orthologue) (needed for flowering barley samples only)
- CLE41 (CLE44 same orthologue) (labelled TDIF on primer tubes)
- DREB2A (two separate orthologues (a) and (b))
- REV (two separate orthologues (a) and (b))
- ATERF15 (two separate orthologues (a) and (b))
- PXL2 (two separate orthologues (a) and (b))
- DREB26 (two separate orthologues (a) and (b)) ⋆⋆
- ARF5 (two separate orthologues (a) and (b))
- ARF4 ⋆
- IAA29
- IAA30
- ERL1

All primers are in the Etchells Lab small freezer (Eurofins freezer box). Those with a star next to them above have primers already diluted (white freezer box). cDNA has been used up with past qPCR runs. More mRNA from all samples analysed so far is stored in the -80 ℃ freezer (purple freezer box, *Supplementary Figure 1*). Further mRNA capture of all samples and qPCR for the above genes is necessary to complete the project.

2. Flowering Barley Morphological Analysis

72-day old Barley plants of WT and PXY mutant phenotype at the first inflorescence stages have been dissected for further morphological analysis. Dissection of 8 PXY mutant and 4 WT flowering Barley plants was carried out in three different places:

- Lower stem (within 5cm of the hypocotyl)
- Upper stem (within 5cm of the immature Barley sheaf)
- Inflorescence tip (within 5mm of the immature Barley sheaf)

These samples were grouped into genotype and position in the plant, and stored in FAA for fixation process initiation (at rtp in lab, labelled PXY top, PXY mid, PXY bottom, WT top, WT mid, WT bottom (*Supplementary Figure 2)*). Further cleaning and fixing of these samples is needed using Agarose solution, EtOH and JB4 to prepare the samples for morphological analysis. Sectioning and light microscopy of these fixed samples of Barley stem should provide morphological information of PXY knockout effects in stem vascular development, in comparison to these genotypes in Arabidopsis. It would also confirm whether the PXY mutants at this stage are in fact at a different developmental stage than the WTs plants, which would explain the Cq result trends (*Figure 4*).

Ideas for Future Research Projects Based on Findings of this Work so far

Coexpression analysis results highlight four candidate genes for further research (*Table 3*), as potential components of the genetic network underlying Barley vascular development. The fourth candidate was found to have likely orthology with a phloem specific transcription factor in *Oryza sativa* (Rice), known to have a function in rice vascular development (common name RF2A). Future investigation into additional genes other than those investigated with laboratory techniques in this project may want to begin with this gene. As a likely phloem specific promoter, there is a relatively strong chance of its involvement in vascular development, despite no known orthologue in Arabidopsis.

Lastly, it is important to note of less relevant findings, though potentially very significant, from the bioinformatics work of this project. The search for coexpression data came across an RNA-seq experiment of eight barley tissues from different developmental stages, whose data was used to rate candidacy of Barley orthologues[4]. This experiment was revisited after establishing the key Barley genes for laboratory investigation, and expression data was filtered to these genes only. This revealed that 9 of the genes of interest – ATERF15, PXL1, ERL1, TMO5, REV, ARF4, ARF5, LBD3/LBD4 and PXY (*Supplementary Figure 3*)– showed significant changes of expression under one or more of the following comparative conditions:

- 'long day length regimen' vs 'short day length regimen' in 'leaf, early reproductive phase, WT'
- 'long day length regimen' vs 'short day length regimen' in 'shoot apex, end of vegetative phase, WT'
- 'ppd-H1 mutant vs 'WT' in 'leaf, early reproductive phase, long day length regimen'

Two conclusions can be made from these results. Firstly, it is further evidence for these genes to be involved in the same genetic network due to coexpression changes under the same conditions. Secondly, this suggests that the PXY genetic network is also involved in light response pathways (ppd-H1 is a pseudo response regulator, and controls leaf size dependent of photoperiod). This may be due to vasculature facilitating photosynthesis product transport - products which are only synthesised on light exposure. As of yet, PXY has not been identified in plant light response pathways, and therefore this finding has potential for fuelling much future research.

References

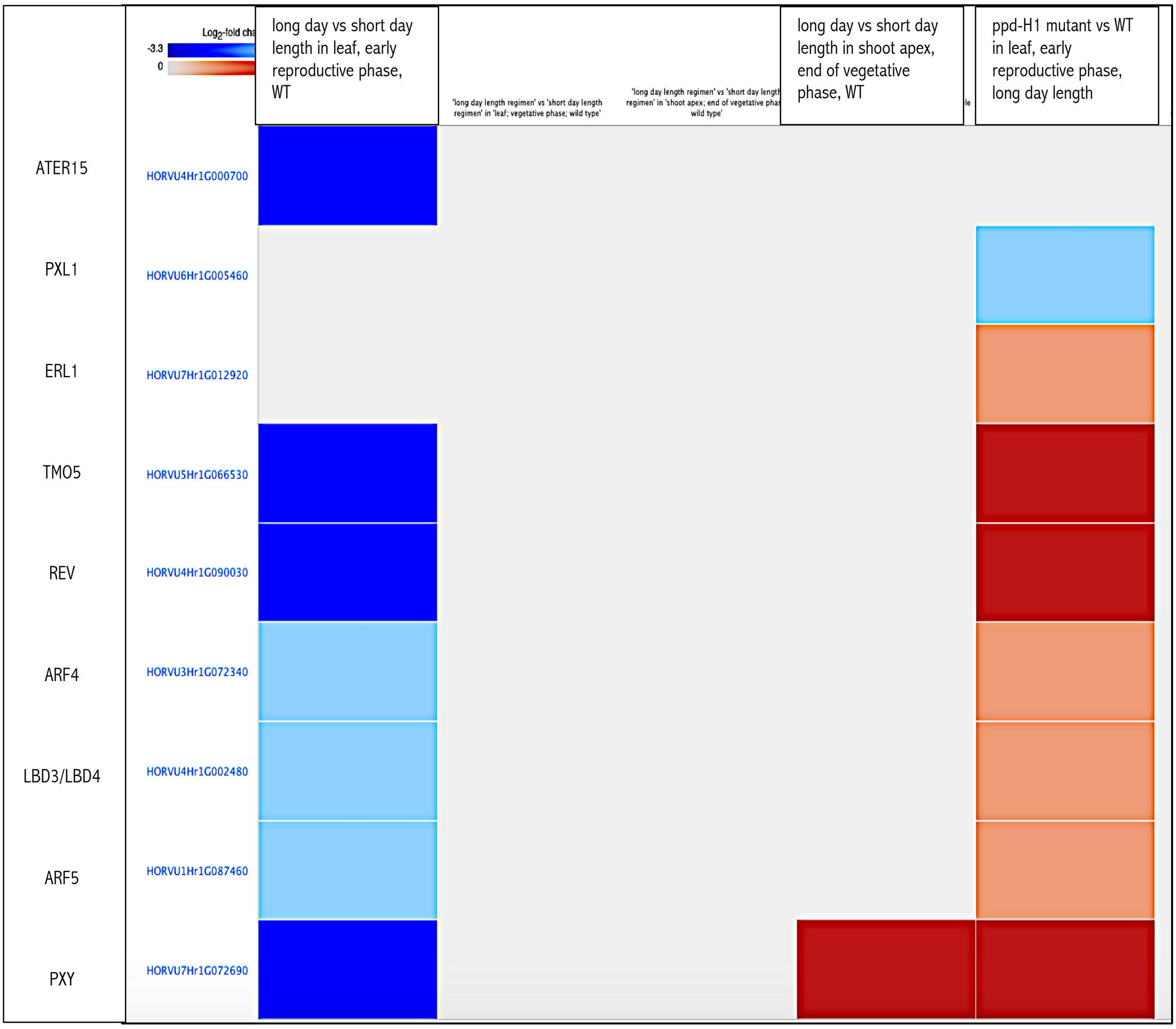
1. https://www.ensembl.org/info/data/biomart/index.html
2. https://bar.utoronto.ca/
3. https://www.ebi.ac.uk/gxa/home
4. International Barley Genome Sequencing Consortium. (2012). A physical, genetic and functional sequence assembly of the barley genome. Nature, 491(7426), 711.
5. https://www.arabidopsis.org/
6. barlex.barleysequence.org/
7. https://plants.ensembl.org/index.html
8. primer3.ut.ee/
9. http://planex.plantbioinformatics.org
10. http://www.ncbi.nlm.nih.gov/geo/
11. https://www.affymetrix.com/analysis/netaffx/
12. https://static.thermoscientific.com/images/D01630~.pdf

Supplementary Figures

Supplementary Figure 1 *– Photograph of the stored mRNA samples from 3 WT and 3 PXY mutant Barley stems and 3 developmental stages: 21 days, 57 days, 72 days (flowering). This freezer box is in the -80°C freezer and has information about each sample on the inside of the lid.*

Supplementary Figure 2 *– Photograph of the 72-day old Barley plants of WT and PXY mutant phenotype at the first inflorescence stages, dissected and stored in FAA for further morphological analysis. These samples have been grouped into genotype and position in the plant: inflorescence tip (labelled top), upper stem (labelled mid), lower stem (labelled bottom).*

.



Supplementary Figure 3 *– The graph produced on filtering for the genes of interest to this project, in data outputted from an RNA-seq experiment for eight Barley tissues at different developmental stages, and under different conditions. Red blocks show positive expression change (log$_2$-fold), and blue blocks show negative expression change (log$_2$-fold), when comparing two different experimental conditions (top axis).*