Universidad Autónoma de Madrid

Insights on the mode of inheritance of eQTLs and on the contribution of stabilizing and directional selection in shaping the evolution of the phosphate starvation transcriptome in Arabidopsis

TESIS DOCTORAL

Sergio Díaz Díaz

Madrid, 2021

PROGRAMA DE DOCTORADO EN BIOCIENCIAS MOLECULARES

UNIVERSIDAD AUTÓNOMA DE MADRID

DEPARTAMENTO DE BIOLOGÍA MOLECULAR

FACULTAD DE CIENCIAS

UNIVERSIDAD AUTÓNOMA DE MADRID

Insights on the mode of inheritance of eQTLs and on the contribution of stabilizing and directional selection in shaping the evolution of the phosphate starvation transcriptome in Arabidopsis

TESIS DOCTORAL

Sergio Díaz Díaz

Graduado en Biología

DIRECTOR

Fco. Javier Paz-Ares Rodríguez

Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CNB-CSIC), Madrid, España

Memoria presentada para optar al grado de

Doctor en Biociencias Moleculares, por

Sergio Díaz Díaz

Universidad Autónoma de Madrid

Mayo de 2019

DIRECTOR

Dr. Fco. Javier Paz-Ares Rodríguez

TUTOR

Dr. Miguel Ángel Rodríguez Gabriel

Fco. Javier Paz-Ares Rodríguez,

Profesor de Investigación en el Centro Nacional de Biotecnología (CNB-CSIC),

CERTIFICA

que el trabajo presentado en esta memoria de investigación ha sido realizado por

Sergio Díaz Díaz

bajo su supervisión.

AGRADECIMIENTOS

Agradezco a todas aquellas personas que, con su apoyo y colaboración, han hecho posible el desarrollo del trabajo que presento en esta memoria.

En primer lugar, quiero dar un sincero agradecimiento a mi director de tesis, Javier Paz-Ares, por la oportunidad que me ha ofrecido para mi formación y aprendizaje científico, así como por su completa dedicación, paciencia y confianza en mi trabajo.

A mi tutor, Miguel Ángel Rodríguez Gabriel, por su buena disposición en todo momento.

Igualmente, quiero agradecer a todos las personas con las que he coincidido en el departamento de Genética Molecular de Plantas del CNB durante estos años y especialmente a todos mis compañeros y miembros del laboratorio 311 Mabel Puga, César Poza, Erica Gil, Mónica Rojas, Miguel Miñambres e Iris Martínez por sus enseñanzas y por toda la ayuda que me han ofrecido.

Un agradecimiento especial a Carlos Alonso-Blanco por el material vegetal prestado, por su esencial asesoramiento durante el curso de este trabajo, así como por su paciencia y generosidad.

Agradezco también a los servicios del CNB, especialmente los de cultivo *in vitro* e invernadero, por el apoyo técnico y la profesionalidad.

Finalmente, también me gustaría reconocer el apoyo de mi familia y amistades durante estos años.

El desarrollo de este trabajo ha sido posible gracias a la financiación otorgada por el Ministerio de Ciencia, Innovación y Universidades a través de la Ayuda para Contratos Predoctorales para la Formación de Doctores BES-2015-073931 y los proyectos BIO2014-60453-R y BIO2017-89530-R.

RESUMEN

La variación natural intraespecífica refleja en parte la capacidad de una especie para adaptarse a entornos locales. Dado que el nivel clave de regulación de la actividad de un gen se sitúa en el control transcripcional, no resulta sorprendente la extensa variación en expresión génica subyacente a la diversidad fenotípica observada entre individuos de una misma especie. En este estudio nos hemos centrado en la respuesta al ayuno de fosfato (PSR) en la especie *Arabidopsis thaliana* para profundizar en cuestiones fundamentales sobre variación de la expresión génica, como lo son las características de los efectos reguladores en *cis* y en *trans*, su modo de herencia, la contribución de la selección estabilizadora y direccional a su regulación, así como las posibles asociaciones entre cambios de expresión y diversidad fenotípica. Para ello hemos empleado dos aproximaciones, una basada en análisis transcriptómicos comparativos entre dos accesiones (Landsberg *erecta*, L*er*; y Llagostera, Ll-0) y sus híbridos de la primera generación filial (F_1), que han revelado la baja aunque significativa contribución de los efectos maternos y la impronta genómica en las diferencias de expresión entre accesiones, así como la influencia de las condiciones ambientales y las señales de desarrollo en la manifestación de los efectos reguladores en *cis* y en *trans*. La segunda aproximación, basada en estrategias de genética cuantitativa clásica, nos ha permitido identificar QTLs de expresión (eQTLs) en híbridos de líneas recombinantes consanguíneas (RILs) cruzadas con sus progenitores L*er* y Ll-0. Los *cis*-eQTLs mostraron una distribución moderadamente uniforme a lo largo del genoma, mientras que la mayor parte de los *trans*-eQTLs se encontraron formando parte de agrupamientos mayores (hotspots) en vez de separados unos de otros. Hemos analizado los genes asociados a estos hotspots en cuanto a sus características de expresión, y al enriquecimiento en categorías funcionales y en genes diana de factores de transcripción, señalando que un único gen podría ser el gen causal de cada hotspot. Asimismo, se han identificado tres hotspots potencialmente reguladores de aspectos clave de la PSR. Además, hemos identificado factores de transcripción candidatos a ser los genes causales de varios hotspots. Nuestros resultados muestran que el modo de herencia más frecuente en los *cis*-eQTLs es la codominancia y para los *trans*-eQTLs es la dominancia, mientras que los casos de sobredominancia fueron muy escasos (0% en los *cis*- y <0,025% en los *trans*-eQTLs). Además, hemos detectado en los *trans*-eQTLs un mayor impacto de la selección estabilizadora respecto a la selección direccional y, en contraste, una mayor prominencia de esta última en los *cis*-eQTLs. Este y otros hallazgos subrayan la mayor evolucionabilidad de los *cis*-eQTLs, que actúan sobre un solo gen, respecto a los *trans*-eQTLs, que influyen en la expresión de muchos genes. Finalmente, hemos detectado correspondencias con un potencial significado biológico entre caracteres fenotípicos relacionados con la PSR y algunos hotspots y módulos de coexpresión génica, lo cual abre la posibilidad de identificar genes de interés agronómico mediante el uso de métodos para analizar la regulación de redes transcripcionales.

ABSTRACT

Intraspecific natural variation refers to the phenotypic diversity found within a given species and it is considered to explain the capacity of the species for adaptation to local environments. Because transcriptional control is the most important level of regulation of gene activity, it is not surprising that underlying the phenotypic diversity within species there is large intraspecific variation in gene expression. In this study we have chosen the phosphate starvation response (PSR) in *Arabidopsis thaliana* to get insights on fundamental questions regarding the basis of gene expression variation, such as the characteristics of *cis*- and *trans*-regulatory effects, their mode of inheritance, the contribution of the stabilizing selection and the directional selection to *cis*- and *trans*-regulation, as well as the potential associations between expression changes and phenotypic diversity. We used two different approaches, the first one based on the comparative transcriptomic analysis of two accessions (Landsberg *erecta*, L*er*; and Llagostera, Ll-0) and their F_1 hybrids, which provided information on the small although still significant contribution of non-mendelian effects (genomic imprinting and maternal effects) in determining expression differences between accessions and also on the influence of environmental and developmental cues in the manifestation of *cis*- and *trans*-regulatory effects. The second method followed a classical quantitative genetic approach to identify expression QTLs (eQTLs) based on the use of hybrids of recombinant inbred lines (RILs) crossed with their progenitors L*er* and Ll-0. In contrast to *cis*-eQTLs, which were distributed in a moderately uniform manner across the genome, a large proportion of *trans*-eQTLs grouped in clusters (hotspots). Several analyses of the genes associated to these hotspots – expression characteristics, GO term enrichment, transcription factor (TF) target enrichment – were indicative that hotspots correspond to single regulators. We identified three hotspots likely regulating important aspects of the PSR. In addition, we identified several TFs as candidates to correspond to the underlying causal gene of several of the hotspots. In addition, we found that while the predominant mode of inheritance of *cis*-eQTLs was codominant, in case of *trans*-eQTLs it was dominant, while cases of over- and underdominance were scarce for both *cis*- and *trans*-eQTLs (0% for *cis*- and <0,025% for *trans*-eQTLs), despite the fact that we detected heterosis for some Pi starvation traits between L*er* and Ll-0. In addition, we detected a higher impact of stabilizing selection compared to directional selection for *trans*-eQTLs ($p<<0{,}00001$) and in sharp contrast, in case of *cis*-eQTLs directional selection was more prominent than stabilizing selection. This and additional findings underline the higher evolvability of *cis*-eQTLs, acting on one gene, over *trans*-eQTLs, often affecting the expression of many genes. Finally, we identified potentially biologically meaningful correlations between *trans*-eQTL hotspots / gene coexpression modules and traits related to plant performance and physiology, which raises the possibility to identify genes of agronomical interest using approaches to analyse gene regulatory networks.

TABLE OF CONTENTS

ABBREVIATIONS

ABA	Abscisic Acid
ABI1	ABL INTERACTOR 1
Abs	Absorbance
AGI	Arabidopsis Genome Initiative
AL6	ALFIN-LIKE 6
ALIX	APOPTOSIS-LINKED GENE-2 INTERACTING PROTEIN X
ALMT1	ALUMINUM-ACTIVATED MALATE TRANSPORTER 1
ALS3	ALUMINUM SENSITIVE 3
ANOVA	Analysis Of Variance
ARP6	Actin-related protein 6
ASE	Allele-Specific Expression
At	Arabidopsis thaliana
ATP	Adenosine Triphosphate
AUX	Auxin
bHLH	Helix-Loop-Helix
bp	Base Pair
BR	Brassinosteroid
CCA1	CIRCADIAN CLOCK ASSOCIATED 1
CIM	Composite Interval Mapping
CK	Cytokinin
cM	Centimorgan
Col-0	Columbia
CRE1	CYTOKININ RESPONSE 1
DAP-seq	DNA affinity purification sequencing

DAS	Days After Sowing
DDM1	DECREASE IN DNA METHYLATION 1
DEG	Differentially Expressed Gene
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic Acid
eQTL	Expression QTL
Eri-1	Eringsboda
ESCRT	Endosomal Sorting Complexes Required for Transport
ET	Ethylene
F_1	First filial generation
FC	Fold Change
FDR	False Discovery Rate
FE	Fold Enrichment
FPKM	Fragments Per Kilo base per Million mapped reads
GA	Gibberellic Acid
GO	Gene Ontology
GRN	Genetic Regulatory Network
GWAS	Genome-Wide Association Studies
InDel	Insertion and Deletion
InsP	Inositol Polyphosphates
IPS1	INDUCED BY PHOSPHATE STARVATION 1
IV	*In vitro*
JA	Jasmonic Acid
L*er*	Landsberg *erecta*
LHY	LATE ELONGATED HYPOCOTYL
Ll-0	Llagostera

LOD	logarithm of odds
LPR	LOW PHOSPHATE ROOT
LPR1	LOW PHOSPHATE ROOT 1
LPR2	LOW PHOSPHATE ROOT 2
Mb	Megabase
MES	2-(N-morpholino)ethanesulfonic acid
MFS	Major Facilitator Superfamily
miRNA	micro RNA
mRNA	Messenger Ribonucleic Acid
NAP1;2	Nucleosome Assembly Protein 1;2
NLA	NITROGEN LIMITATION ADAPTATION
No-0	Nossen
OA	Organic Acid
ORF	Open Reading Frame
P	Phosphorus
P1BS	PHR1-Binding Site
PCR	Polymerase Chain Reaction
PDR2	PHOSPHATE DEFICIENCY RESPONSE 2
PHF	PHOSPHATE TRANSPORTER TRAFFIC FACILITATOR
PHL1	PHR1-LIKE1
PHO1	PHOSPHATE 1
PHO2	PHOSPHATE 2
PHR1	PHOSPHATE STARVATION RESPONSE 1
Pht/PHT	Phosphate Transporter
Pi	Inorganic Phosphate, Orthophosphate
PPi	Inorganic Pyrophosphate

PRL	Primary Root Length
PS	Phosphate Starvation
PSI	Phosphate Starvation-Induced
PSR	Phosphate Starvation Response
PSTOL1	PHOSPHORUS-STARVATION TOLERANCE 1
QTL/QTLs	Quantitative Trait Locus/Loci
R^2	Explained Phenotypic Variance
RIL	Recombinant Inbred Line
RNA	Ribonucleic Acid
RNA-seq	RNA sequencing
ROS	Reactive Oxygen Species
rRNA	Ribosomal Ribonucleic Acid
RSA	Root System Architecture
RT	Room Temperature
SHR	SHORT ROOT
SIZ1	SUMO E3 ligase
SL	Strigolactone
SNP	Single Nucleotide Polymorphism
STAR1	STEROIDOGENIC ACUTE REGULATORY PROTEIN 1
STOP1	SENSITIVE TO PROTON TOXICITY 1
TAE	Tris Acetate EDTA
TF	Transcription Factor
TIR1	TRANSPORT INHIBITOR RESPONSE 1
UV	Ultraviolet
VTC4	VACUOLAR TRANSPORTER CHAPERONE 4
WGCNA	Weighted Gene Correlation Network Analysis

1. INTRODUCTION

1.1 - PHOSPHATE STARVATION IN HIGHER PLANTS

1.1.1 – THE PHOSPHATE STARVATION RESPONSE

1.1.1.1 – The challenge of phosphorus limitation.

Phosphorus (P) is an indispensable element, second to nitrogen, for all living organisms as it is a key component in nucleic acids and biomembranes. Moreover, phosphates play a crucial role in energy transfer as intermediates in essential metabolic pathways and signal transduction (Raghothama, 1999; Marschner, 2012).

Plant organisms assimilate phosphorus as orthophosphate (inorganic phosphate, Pi) and it is generally taken up from the surrounding soil through the root system. In most soils, Pi availability and mobility are suboptimal due to alkaline mineralization and acidic precipitation, which compromise plant growth and yield. In addition, the excessive supply of phosphate fertilizers will increase the eutrophication of water resources and the global prizes due to finite Pi resources. Hence, to cope with the challenges of securing food supply worldwide while reducing the impact of agriculture in the environment, it is crucial to increase our understanding of the molecular mechanisms underlying plant adaptation to Pi starvation and apply this knowledge to maximize Pi acquisition and utilization efficiency in crops (López-Arredondo *et al.*, 2014).

1.1.1.2 – The phosphate starvation response in higher plants.

Regulation of responses to phosphate starvation is regarded as one of the most relevant and emblematic model systems of transcriptional control, both in bacteria and fungi, as it has largely contributed to disclose and define essential points and concepts in gene regulation such as histone eviction (Korber *et al.*, 2004) and the mechanism of decoupling gene induction from dynamic range (Lam, Steger and O'Shea, 2008). In plants, both the role of miRNAs in regulating responses to stress (Chiou *et al.*, 2006) and the mechanism of control of miRNA activity, termed 'target mimicry' (Franco-Zorrilla *et al.*, 2007), were first observed in the context of studies on phosphate starvation and later generalized beyond this system (Tay, Rinn and Pandolfi, 2014).

Upon low Pi availability, plants gradually display a battery of adaptive responses, collectively known as the phosphate starvation responses (PSRs), which include diverse morphological, developmental, physiological and biochemical adaptive changes directed to optimize the acquisition and use efficiency of Pi (Chiou and Lin, 2011; López-Arredondo *et al.*, 2014; Scheible and Rojas-Triana, 2015; Puga *et al.*, 2017). PSRs are well studied in the model system *Arabidopsis thaliana* (L. Heynh) (hereinafter referred to as Arabidopsis).

Plants experience severe modifications in the root system architecture (RSA) to improve soil exploration and enhance Pi acquisition from low Pi soils. These changes produce an increase in the root-to-shoot growth ratio and the arrest of growth in the primary root, known as meristem exhaustion. These changes, along with increased number of lateral roots and elongation of root hairs, result in the expansion of the root surface in contact with the surrounding soil (Vance, Uhde-Stone and Allan, 2003; Sánchez-Calderón *et al.*, 2005; Svistoonoff *et al.*, 2007; Péret *et al.*, 2014). In addition, many species display specific adaptations in the RSA, like the increase in the density/volume of shallow roots to exploit the Pi-enriched topsoil horizons (Lynch and Brown, 2001; Sun, Gao and Lynch, 2018). Arbuscular mycorrhizae are common adaptations to establish symbiotic associations with fungi (Bucher, 2007; Smith *et al.*, 2011), while other species produce a particular kind of roots defined as 'proteoid' or cluster roots (Vance, Uhde-Stone and Allan, 2003; Lambers *et al.*, 2011; Péret *et al.*, 2014).

One of the most important biochemical strategies to cope with plant growth upon low Pi supply is the enhancement in the acquisition of exogenous Pi through promoting the synthesis of organic acids (OAs) like citrate, oxalate, and malate. These OAs are exuded into the surrounding soil for Pi release from insoluble phosphate salts (Ryan, Delhaize and Jones, 2001; Oburger *et al.*, 2009). Moreover, the concurrent acidification of the rhizosphere contributes to establish beneficial interactions with rhizosphere microbes (Richardson *et al.*, 2009; Pii *et al.*, 2015). Plants also produce different hydrolases to release and recycle Pi from the organic matter present in the soil and to remobilize Pi from senescent tissues into other parts of the plant (Bariola *et al.*, 1994; Ryan, Delhaize and Jones, 2001; Hurley *et al.*, 2010; Lambers *et al.*, 2011; Plaxton and Tran, 2011).

Plants under Pi stress induce high-affinity phosphate transporters, encoded by the *PHOSPHATE TRANSPORTER* (Pht) genes. In Arabidopsis, components of Pht2, Pht3 and Pht4 gene families mediate Pi distribution at subcellular levels, while Pht1 genes are involved in Pi uptake and are expressed in the root epidermis and outer cortex (Raghothama, 1999; Mudge *et al.*, 2002; Versaw and Harrison, 2002; Guo *et al.*, 2008). The alternative use of different enzymes and metabolic routes, known as metabolic redundancy, contributes to adaptation to low Pi regimes, like the cytosolic glycolysis or the use of inorganic pyrophosphate (PPi) to save ATP (Plaxton and Tran, 2011). In addition, internal Pi saving is related to reduction in the abundance of ribosomal RNA (rRNA) (Sulpice *et al.*, 2014), phosphorylated metabolites (Pant *et al.*, 2015) and membrane phospholipids, commonly replaced by sulfolipids and galactolipids (Kobayashi *et al.*, 2006; Gaude *et al.*, 2008; Plaxton and Tran, 2011; Nakamura, 2013). Furthermore, Pi-starved plants accumulate many sugars, starch, amino acids and secondary metabolites like glucosinolates and phenylpropanoids. The biosynthesis of anthocyanins, via the phenylpropanoid route, is regarded as a protective mechanism against the photoinhibition caused by Pi starvation (Rubio *et al.*, 2001; Vance, Uhde-Stone and Allan, 2003; Morcuende *et al.*, 2007; Pant *et al.*, 2015).

1.1.2 – REGULATION OF THE PHOSPHATE STARVATION RESPONSE

1.1.2.1 – The local and systemic Pi signaling pathways.

In plants, the PSRs are coordinated by a sophisticated regulatory system that can efficiently perceive information concerning both external and internal Pi availability. This is accomplished by two different Pi sensing mechanisms, one external and other internal, that act through two signaling pathways that are independent to some extent. Nevertheless, there is substantial interconnection between them and also with other nutrient and hormone signaling pathways (Franco-Zorrilla *et al.*, 2004; Rouached, Arpat and Poirier, 2010; Yang and Finnegan, 2010; Chiou and Lin, 2011; Puga *et al.*, 2017 ; **Figure I1**).

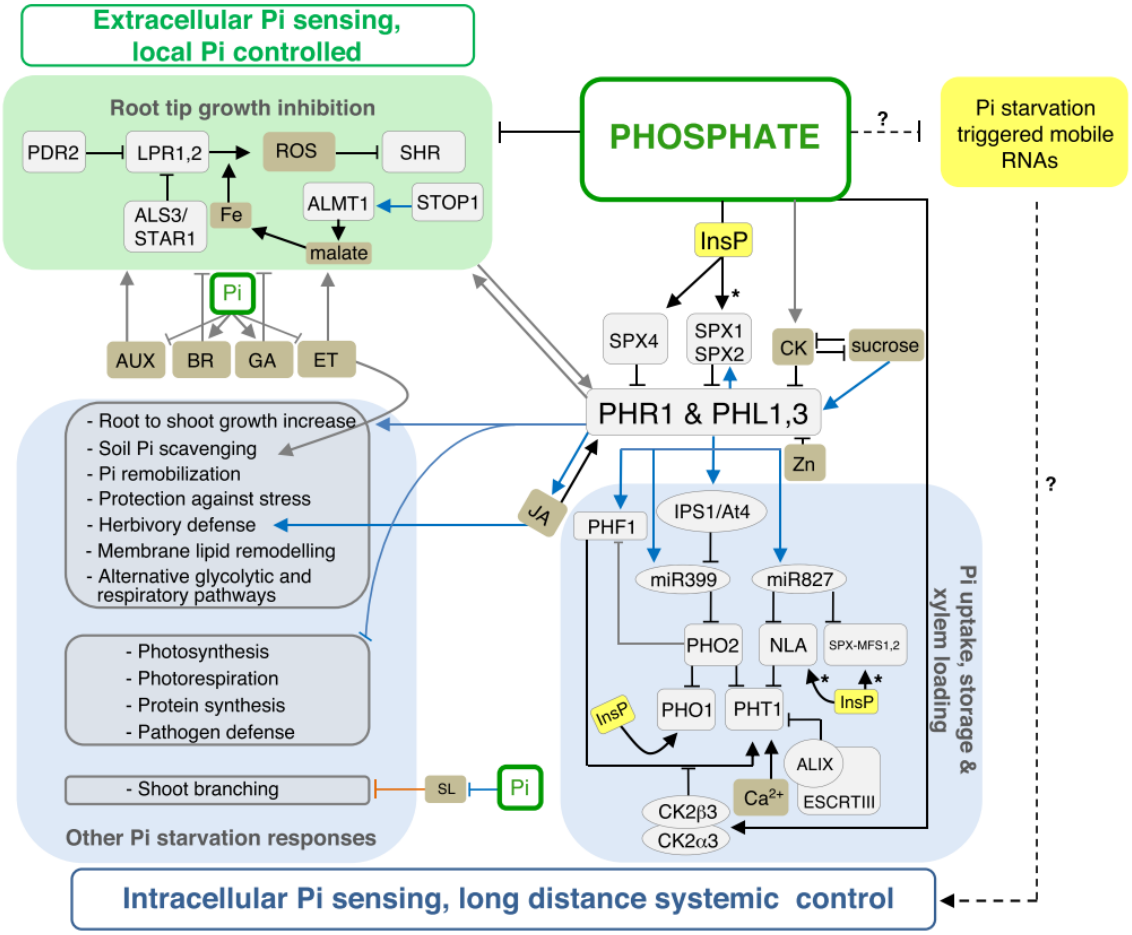


Figure I1. **The core Pi starvation signaling routes in Arabidopsis**. The essential components and interactions underlying the two main branches (local and systemic) of Pi sensing and signaling are schematized. This figure is taken from Puga *et al.* 2017).

The local Pi signaling pathway is focused on the inhibition of the root tip and triggers some of the PSRs associated to the remodeling of the RSA, which are described above. The root tip and its surrounding area perform the roles of local site of Pi sensing and regulation, respectively, and the

extracellular concentration of Pi functions as a local signal, regardless of the whole-plant Pi content (Yang and Finnegan, 2010; Chiou and Lin, 2011; Abel, 2017). Two paralogous ferroxidases, *LOW PHOSPHATE ROOT 1* (*LPR1*) and *LPR2* are involved in local Pi signaling (Reymond *et al.*, 2006). LPR1 contributes to the accumulation of Fe^{3+} in the meristem, leading to an increase in the deposition of callose and production of reactive oxygen species (ROS). As a result, the accumulation of SHORT ROOT regulatory proteins in the quiescent centre is severely reduced, which influences the regulatory changes that determine RSA remodeling under low Pi (Svistoonoff *et al.*, 2007; Müller *et al.*, 2015). In addition to Pi, the availability of nitrogen, sulphur and potassium influences the production and localization of ROS, which are associated to nutrient starvation responses (Shin, Berg and Schachtman, 2005; Tyburski, Dunajska and Tretyn, 2009; Chiou and Lin, 2011).

Most of the known PSRs are regulated at long distance via systemic Pi signaling pathway. It is controlled by the whole-plant Pi status and predominantly reliant on intracellular Pi sensing, although it is known that sugars, hormones, and mobile mRNAs may act as systemic signals (Chiou and Lin, 2011; Puga *et al.*, 2017).

1.1.2.2 – Transcriptional and post-transcriptional regulation.

The most important regulatory mechanisms underlying PSRs occur at the transcriptional level, mediated by transcription factors (TFs). In Arabidopsis, there is ample evidence on the large transcriptional reprogramming occurring in response to Pi starvation (Hammond *et al.*, 2003; Wu *et al.*, 2003; Misson *et al.*, 2005; Morcuende *et al.*, 2007; Bustos *et al.*, 2010; Smith *et al.*, 2010; Jain, Nagarajan and Raghothama, 2012; Woo *et al.*, 2012). Moreover, chromatin remodeling and diverse post-transcriptional regulatory mechanisms influence PSRs, like miRNAs, other non-coding RNAs, and post-translational modifications such as phosphorylation, ubiquitination and sumoylation (Chiou and Lin, 2011; Rojas-Triana *et al.*, 2013; Puga *et al.*, 2017). Such regulatorychanges in Pi starvation-responsive genes and most of the known PSRs in plants are ultimately governed by the MYB-type transcription factor *PHOSPHATE STARVATION RESPONSE 1* (*PHR1*). To promote Pi uptake upon deficiency conditions, *PHR1* directly induces the expression of Pi transporters like *PHT1;1* (*PT1*) (Rubio *et al.*, 2001; Franco-Zorrilla *et al.*, 2004), and the cofactor *PHOSPHATE TRANSPORTER TRAFFIC FACILITATOR 1* (*PHF1*), which enables PHT1 traffic from the endoplasmic reticulum to the plasma membrane (González *et al.*, 2005). PHT1 transporters are targeted for ubiquitin-mediated degradation by NITROGEN LIMITATION ADAPTATION (NLA) and PHOSPHATE 2 (PHO2) (Lin, Huang and Chiou, 2013; Park, Seo and Chua, 2014). PHO2 also provokes the degradation of PHO1, which is involved in Pi loading into de xylem and negatively influences the accumulation of the PHF1

protein (Hamburger *et al.*, 2002; Liu *et al.*, 2012; Huang *et al.*, 2013). Under low Pi conditions, PHR1 indirectly repress NLA and PHO2 through the induction of the microRNAs miR827 and miR399 (Fujii *et al.*, 2005; Aung *et al.*, 2006; Bari, Datt Pant, *et al.*, 2006; Kant, Peng and Rothstein, 2011). The accumulation of PHT1 transporters in the plasma membrane is downregulated by two additional mechanisms, recently characterized, implied in Pi uptake and homeostasis. In one of them, the cytosolic protein APOPTOSIS-LINKED GENE-2 INTERACTING PROTEIN X (ALIX) act as an intermediary in the degradation of PHT1 via the endosomal pathway (Cardona-López *et al.*, 2015). A second mechanism of PHT1 inhibition is mediated via phosphorylation by the CK2α2β3 kinase. The activity of such kinase is reduced under Pi limitation and thus PHT1 inhibition is relieved (Chen *et al.*, 2015). Under low Pi, *PHR1* also induces non-coding genes like *INDUCED BY PHOSPHATE STARVATION 1* (*IPS1*), whose RNA is non-cleavable, and it rather sequesters miR399 through sequence complementarity. Therefore, the effect of *IPS1* antagonizes that of miR399, preventing the mRNA of PHO2 to be targeted. Such mechanism of control has been characterized and termed 'target mimicry' (Franco-Zorrilla *et al.*, 2007).

PHR1 and its partially redundant homolog *PHR1-LIKE1* (*PHL1*) are expressed constitutively and directly bind to a DNA motif known as PHR1-binding site (P1BS). The promoter sequences of Pi starvation (PS)-induced genes are enriched in P1BS, although PHR1-mediated gene repression is indirect, as P1BS is not overrepresented in the PS-repressed genes (Bustos *et al.*, 2010; Sun *et al.*, 2016). PHR1 is regarded as the central regulator of PSRs, since a major part of the Pi-responsive genes controlled via systemic signaling and some of the regulated via local are direct targets of PHR1/PHL1 (Thibaud *et al.*, 2010). For instance, PHR1 and its target VTC4 mediate the generation of ROS required for the inhibition of root tip (Balzergue *et al.*, 2017; Mora-Macías *et al.*, 2017). Moreover, *PHR1* has a major impact on the Pi-responsive metabolome (Pant *et al.*, 2015). Although PHR1 is expressed constitutively, its activity is modulated at the post-transcriptional level by sumoylation. Such modification appears to be driven by the SUMO E3 ligase SIZ1 in response to Pi availability, as observed *in vitro* in Arabidopsis (Miura *et al.*, 2005). A more direct Pi-dependent regulation of *PHR1* is mediated by the nuclear protein SPX1 (Puga *et al.*, 2014). The SPX domain is found across eukaryotic proteins and it is regarded as a key component of the intracellular Pi sensing, since it can bind inositol polyphosphates (InsP) with high affinity (Wild *et al.*, 2016). Moreover, SPX domain-containing proteins are diverse and participate in numerous aspects of Pi homeostasis, including Pi signaling, transport and metabolism (Azevedo and Saiardi, 2017). Recent studies concerning plant species have shown that the nuclear SPX-domain proteins SPX1, SPX2 and SPX4 directly bind and inhibit the activity of PHR1/PHL1 in a highly Pi-dependent manner, both *in vivo* and *in vitro*. This interaction is regarded as a link between Pi sensing and signaling (Lv *et al.*, 2014; Puga *et al.*, 2014; Wang *et*

al., 2014). The vacuolar SPX-MFS (major facilitator superfamily) proteins, implicated in Pi transport, are critical for plant adaptation to low Pi (Liu *et al.*, 2015, 2016). Recently, it has been suggested that some kind of posttranslational modification or additional interactions could provide the mechanism to increase the Pi binding affinity of SPX-domain proteins and evidence their role as Pi sensors in plants (Puga *et al.*, 2017). In this respect, the inositol pyrophosphate $InsP_8$ molecule is described as the intracellular Pi signal that directly binds to SPX1, playing an essential role in the regulation of plant Pi homeostasis (Dong *et al.*, 2019).

Apart from PHR1/PHL1 and to a lesser extent, the regulation of gene expression under Pi starvation involves the contribution of additional TFs, which presumably share some functionally relation with *PHR1*. These include components of the WRKY, bHLH, ZAT families and the MYB superfamily, among others (reviewed in Puga *et al*. 2017).

Recent studies concerning crosstalk between defense and PSRs in plants have evidenced that under nutritional stress maintaining Pi homeostasis is a priority over defense, as PHR1 directly supresses the expression of genes involved in defense against microbial pathogens (Castrillo *et al.*, 2017). In addition, PHR1 mediate the activation of the jasmonic acid (JA) signaling pathway, leading to enhanced plant resistance to herbivores under Pi starvation (Khan *et al.*, 2016).

In the last years, chromatin remodeling has emerged as another framework that appears to play a significant role in plant PSRs via transcriptional regulation (Scheible and Rojas-Triana, 2015). Some studies attribute a putative role in the regulation of PSRs, Pi homeostasis and PS-induced genes to chromatin-remodeling proteins like ARP6, which is a key component of the SWR1 chromatin remodeling complex (Smith *et al.*, 2010), the histone chaperone NAP1;2 (Iglesias *et al.*, 2013) and the homeodomain protein ALFIN-LIKE 6 (AL6) (Chandrika *et al.*, 2013).

Additionally, miRNAs are important regulators of long-distance systemic Pi signaling in plants (Jones-Rhoades, Bartel and Bartel, 2006). Apart from the above-mentioned miR399 and miR827 (Aung *et al.*, 2006; Bari, Pant, *et al.*, 2006; Chiou *et al.*, 2006), additional miRNAs like miR395, miR398, miR169 and miR828 are responsive to Pi and other nutrients (Hsieh *et al.*, 2009; Pant *et al.*, 2009). Recent studies have evidenced that Pi availability may influence translation through small ORFs localized upstream on protein-coding mRNAs (Bazin *et al.*, 2017).

1.1.2.3 – The role of plant hormones and other nutrients in Pi signaling and PSRs.

Plant hormones are implicated in Pi signaling and Pi availability may alter the biosynthesis, transport, and responsiveness of plant hormones. Accumulated data reveal the complex interconnection between the signaling routes of many plant hormones, phosphate, and other nutrients (Schachtman and Shin, 2007; Rubio *et al.*, 2009; Chiou and Lin, 2011; Baek *et al.*,

2017). For instance, the implication of cytokinin (CK) signaling in the control of PSRs has been well characterized in Arabidopsis. CKs negatively affect the expression of a broad range of PSI genes (Martín *et al.*, 2000; Franco-Zorrilla *et al.*, 2002, 2005; Karthikeyan *et al.*, 2002), although it is unclear whether the extent of such repression is systemic or localized (Martín *et al.*, 2000; Franco-Zorrilla *et al.*, 2005). The absence of Pi downregulates the CK receptor *CYTOKININ RESPONSE 1* (*CRE1*), leading to reduced concentration and activity of CKs (Franco-Zorrilla *et al.*, 2002). As a result, PSI genes increase their expression, ensuring the full development of PSRs during Pi starvation (Franco-Zorrilla *et al.*, 2005). Auxin signaling is closely implicated in the alteration of RSA during Pi deficiency (Gilbert *et al.*, 2000; López-Bucio *et al.*, 2002; Al-Ghazi *et al.*, 2003). An increased accumulation of the auxin receptor TIR1 under Pi starvation is associated to enhanced auxin responsiveness, which stimulates the emergence and development of lateral roots (López-Bucio *et al.*, 2002; Pérez-Torres *et al.*, 2008). Additionally, both auxin and CK signaling routes regulate the expression of genes involved in the remodeling of membranes lipids upon Pi starvation (Kobayashi *et al.*, 2006; Narise *et al.*, 2010). Ethylene endogenous levels increase during Pi starvation and the application of exogenous ethylene mimics some of the characteristic RSA changes produced under Pi starvation, including the inhibition of primary root growth and the elongation of lateral roots (Borch *et al.*, 1999; López-Bucio *et al.*, 2002; Ma *et al.*, 2003). Some studies suggest that ethylene may also participate in the stimulation and development of root hairs in response to Pi deficiency (Zhang, Lynch and Brown, 2003; He *et al.*, 2005). Moreover, ethylene controls Pi remobilization during the senescence of particular tissues in some plant species (Chapin and Jones, 2009). Abscisic acid (ABA)-treated and Pi-starved plants manifest some similarities concerning increase in the density of root hairs and in the root-to-shoot ratio (Trull *et al.*, 1997; Ciereszko and Kleczkowski, 2002). ABA could influence some PSI genes through the regulation of ABI1 type 2C protein phosphatase, although it is unclear whether there is such a direct connection between PSR and ABA signaling (Shin *et al.*, 2006; Ribot, Wang and Poirier, 2008). The bioactive levels of gibberellins (GA) are reduced upon Pi starvation, supressing the negative effect of GAs on particular PSRs like RSA modification and anthocyanin accumulation (Jiang *et al.*, 2007). Studies on diverse plant species describe that the root synthesis of strigolactones is promoted in response to Pi starvation to enhance the acquisition and use of Pi (López-Ráez *et al.*, 2008). These are fundamental to establish associations with arbuscular mycorrhizal fungi (Yoneyama *et al.*, 2007) and to optimize plant growth through the inhibition of shoot branching during Pi depletion (Umehara *et al.*, 2010). As mentioned above, the biosynthesis and activity of JA and jasmonic acid-isoleucine (JA-Ile), its bioactive conjugate, are influenced by *PHR1* upon Pi starvation. JA signaling results in enhanced plant resistance to the damage caused by herbivorous insects (Khan *et al.*, 2016).

Deprivation of Pi in plant leaves results in decreased photosynthetic activity and increased concentration of sugars, principally sucrose, and starch (Ciereszko, Johansson and Kleczkowski, 2005; Morcuende *et al.*, 2007; Müller *et al.*, 2007; Nilsson, Müller and Nielsen, 2007). Such increase in shoot-derived sugars may act as systemic signals to promote or enhance the PSRs in distant tissues like the roots. This sugar-mediated signaling is well documented and appears to be critical in the control of PSRs, like RSA alteration and the expression of PSI genes (Müller *et al.*, 2005, 2007; Jain *et al.*, 2007; Karthikeyan *et al.*, 2007; Hammond and White, 2008). Likewise, Pi starvation may influence the expression of many sugar-responsive genes (Ciereszko and Kleczkowski, 2002). Auxin responsiveness and transport under Pi starvation may be promoted by sugars (Jain *et al.*, 2007; Karthikeyan *et al.*, 2007), while CKs and sugars exert antagonistic effects on the regulation of PSI genes (Franco-Zorrilla *et al.*, 2005). Last, additional interconnections of PSRs with other common nutrients are described for calcium (Liu *et al.*, 2011), nitrogen (Hu *et al.*, 2019), and sulphate (Essigmann *et al.*, 1998; Rouached *et al.*, 2011).

1.2 - NATURAL VARIATION: THE EVOLUTION OF GENE EXPRESSION

1.2.1 – NATURAL VARIATION IN PLANTS

1.2.1.1 – The importance of natural variation in plants.

As anticipated by Charles Darwin, species are not fixed and individuals within natural populations may differ in a plethora of morphological and functional traits (Darwin, 1859). The primary source of variation are random mutations that continually arise and potentially contribute to phenotypic divergence (Bolnick *et al.*, 2011). Intraspecific genetic variation is the primary source of variation providing the raw material for evolutionary processes like adaptive evolution (Whitehead and Crawford, 2006; Mitchell-Olds, Willis and Goldstein, 2007). In other words, the occurrence of natural variation among conspecific individuals provides the potential to overcome selective pressures in changing environments through a wide range of phenotypic responses (Hoffmann and Willi, 2008; Henderson and Salt, 2017).

Human communities have long exploited natural variation through artificial selection of plant growth and yield-related traits (Doebley, Gaut and Smith, 2006). Plants are sessile organisms and thus more vulnerable to environmental perturbations (Zia *et al.*, 2021). During their lifespan, these organisms may experience morphological and physiological modifications induced by changes in the environment. These responses are not related to genetic changes and are known as phenotypic plasticity, a fundamental feature for many plants. (Nicotra *et al.*, 2010; Bolnick *et al.*, 2011).

Understanding the molecular mechanisms underlying natural variation in plants may also reveal the ecological and evolutionary significance of adaptive traits (Mitchell-Olds and Schmitt, 2006; Alonso-Blanco *et al.*, 2009). The traditional methods of functional genetics ignore the occurrence of natural variation and they rather focus on the variability artificially induced via mutagenesis and implemented in a restricted number of wild-type stocks of Arabidopsis. Although these loss-of-function approaches have successfully contributed to reveal gene function through characterizing mutant phenotypes, they are blind to detect deleterious or weak allele effects. Moreover, complex phenotypes like growth or stress tolerance are controlled by multiple genes, most of which may have a low or moderate size of effect and involve functionally redundant or epistatic interactions. To overcome these barriers, the alternative use of natural variation approaches provides a complementary insight into gene function and contribute to identify allelic variants with ecological and evolutionary significance. This is because they are presumably favoured and maintained in nature by natural selection, preserving plant fitness, and thus better optimized for crop improvement strategies. In addition, natural accessions are suitable to analyse the molecular underpinnings of plant responses and adaptation to diverse environments (Alonso-Blanco and Koornneef, 2000; Trontin *et al.*, 2011; Colautti, Lee and Mitchell-Olds, 2012; Marchadier *et al.*, 2019).

Some of the most relevant conceptual advances to date on quantitative variation in plants are focused on the model system *Arabidopsis thaliana*. Briefly, it is a small flowering autogamous plant, representative of the Brassicaceae family, which includes some species of agronomic relevance (Meyerowitz, 1987). It is a very efficient organism with a small size and a short reproductive cycle, allowing fast experiments with limited space requirements. Arabidopsis can be easily crossed and transformed, yielding a myriad of seeds per plant. In addition, the small size and organization of its genome are advantageous for genetic studies (Koornneef and Meinke, 2010). Due to the emerging interest on this species during the last century, the Arabidopsis Genome Initiative (AGI) accomplished the sequencing of the nuclear genome of the reference accession Col-0 in the year 2000 (Poczai *et al.*, 2000). Although Arabidopsis is native to Europe, Central Asia and North Africa, today it is distributed worldwide (Al-Shehbaz and O'Kane, 2002; Weigel, 2012) across different temperature and precipitation regimes, latitudes, heights and soils; both in natural and rural areas (Hoffmann, 2002). Such a variety of habitats contribute to an extensive intraspecific genetic and epigenetic diversity for many phenotypic traits, presumably relevant for adaptation to specific environments (Mitchell-Olds and Schmitt, 2006; Alonso-Blanco *et al.*, 2009; Kawakatsu *et al.*, 2016; Vidigal *et al.*, 2016). This is reflected in the occurrence of a wide diversity of virtually homozygous wild strains, commonly referred to as accessions or ecotypes (Alonso-Blanco and Koornneef, 2000). Thousands of natural accessions have been recorded and sequenced to date and they are accessible in different germplasm banks,

along with diverse experimental sets of mapping populations (Alonso-Blanco *et al.*, 2016). In addition, there are multiple online databases providing unlimited genomic information and resources for QTL mapping, such as high-quality sequence variants, genetic and physical maps, among others (Provart *et al.*, 2016).

1.2.1.2 – Analysis of natural variation in plants: quantitative trait loci.

Phenotypic differences among accessions may sometimes exhibit a discrete nature when this variation is controlled by a single or a few loci that segregate as qualitative, as in common laboratory mutants studied by complementation and classical Mendelian analysis. In nature, these qualitative traits are observed as dichotomies in the presence of trichomes, resistance to some pathogens, structure of some compounds, time of flowering, and also in the mutant phenotype associated to the *ERECTA* gene (Koornneef, Alonso-Blanco and Vreugdenhil, 2004; van Zanten *et al.*, 2009). Nevertheless, phenotypic variation within Arabidopsis is mostly continuous and determined by the joint effect of diverse environmental factors and multiple loci, which are termed quantitative trait loci (QTLs) (Kearsey and Farquhar, 1998; Alonso-Blanco and Koornneef, 2000). The complexity of these quantitative traits is directly associated to the level of biological organization. Thus, traits affecting the whole-plant structure or developmental processes that involve several mechanisms are presumably regulated by multiple QTLs with small size effect (multigenic control). In contrast, phenotypic traits at low structural levels are typically determined by a few QTLs with extensive effects and thus they are easier to analyse (Alonso-Blanco and Méndez-Vigo, 2014). In this context, recent studies are progressively approaching to explore the natural variation of molecular traits at lower levels of complexity, including biochemical (Lisec *et al.*, 2008; Routaboul *et al.*, 2012; Louis Joosen *et al.*, 2013), gene expression (Cubillos *et al.*, 2012; Lowry *et al.*, 2013; Snoek *et al.*, 2013) and epigenetic patterns (Schmitz *et al.*, 2013).

1.2.1.3 – Methods to analyse quantitative variation.

Although parental accessions display substantial phenotypic variability, much of the genetic variation affecting the trait/s of interest cannot be directly detected due to the combined effect of several loci with opposite effects. In this respect, the analysis of experimental mapping populations provides a more accurate estimation of the hidden differences between two natural accessions, as individuals segregate for the trait under study, revealing transgressive genotypes that exceed the parental range. QTL analysis based on linkage mapping is the most common method to dissect the genetic architecture of quantitative traits, i.e., to detect QTLs accounting for phenotypic variation and locate them using a molecular linkage map. The location accuracy and the power for effect detection are directly affected by the number of individuals analysed and

the design of the mapping population (Alonso-Blanco and Koornneef, 2000; Buckler IV and Thornsberry, 2002; Alonso-Blanco, Koornneef and van Ooijen, 2006; Alonso-Blanco and Méndez-Vigo, 2014). The use of experimental mapping populations like the sets of recombinant inbred lines (RI lines; RILs), among others, have proven effective to examine quantitative variation and identify QTLs potentially adaptive in a wide range of natural environments (Koornneef, Alonso-Blanco and Vreugdenhil, 2004). Although linkage mapping using biparental RIL populations have revealed many QTLs and candidate loci, genome-wide association studies (GWAS) are rapidly emerging as default approaches for natural variation studies in many organisms, including plants, due to the recent advances in sequencing technologies. Briefly, this method associates genetic variants like SNPs to large amounts of phenotypic data of agricultural interest (Atwell *et al.*, 2010; Li *et al.*, 2010; Horton *et al.*, 2012; Provart *et al.*, 2016).

In Arabidopsis, a considerable amount of quantitative variation has been examined for multiple traits associated to plant development, physiology, biochemistry, and tolerance to both biotic and abiotic stress (Koornneef, Alonso-Blanco and Vreugdenhil, 2004). There is a considerable amount of natural variation among different accessions of Arabidopsis concerning morphological responses to phosphate availability (Chevalier *et al.*, 2003), which are of special interest in this thesis. In this respect, the most relevant finding is a major QTL associated to primary root length and root cell elongation upon phosphate starvation, identified as *LPR1* (Reymond *et al.*, 2006). Similar studies involve Pi and additional nutrient stresses (Loudet *et al.*, 2003; Prinzenberg *et al.*, 2010). Recently, a combined approach of genome-wide association and QTL analysis detected several candidate loci associated to anion homeostasis upon Pi limitation (El-Soda *et al.*, 2019). GWAS have revealed additional findings associated to root architecture (Rosas *et al.*, 2013) and integrative responses to both Pi and salt stresses (Kawa *et al.*, 2016). Apart from Arabidopsis, in crop species like Brassica several QTLs have been detected under Pi deficiency associated to root morphology (Yang *et al.*, 2010) and to characters related to seed and yield (Shi *et al.*, 2013). Similar studies revealed major QTLs involved in tolerance to Pi stress in soybean (Zhang *et al.*, 2017) and rice, where the locus *PSTOL1 (PHOSPHORUS-STARVATION TOLERANCE 1)* is involved in phosphorus acquisition (Gamuyao *et al.*, 2012; Mukherjee *et al.*, 2014).

1.2.2 – THE EVOLUTION OF GENE EXPRESSION

Understanding the molecular underpinnings and the extent to which genomic variation contribute to morphological divergence represents a major challenge in evolutionary genetics. Arising heritable differences may occur through modifications in the gene coding sequence that reshape protein structure and function (Hoekstra and Coyne, 2007) or through regulatory changes that affect transcription and produce fluctuations in the amount of protein levels, substantially altering

the timing, tissue-specificity and/or the magnitude of gene expression. Although both mechanisms contribute to shape morphological differences and account for adaptive change, the influence of regulatory differences on phenotypic evolution is predominant (Carroll, 2008; Stern and Orgogozo, 2008; Harrison, Wright and Mank, 2012). The quantitative variation for gene expression occurring within and between natural populations may contribute to understand the evolutionary basis of gene expression (Oleksiak, Churchill and Crawford, 2002), which is a prominent area of research in current biology. Therefore, gene expression emerges as a decisive point of interaction among genomic variation, genetic regulatory networks (GRNs) and the environment, that ultimately originates complex and divergent phenotypes (Whitehead and Crawford, 2006; Emerson *et al.*, 2010; Signor and Nuzhdin, 2018).

1.2.2.1 – The regulation of gene expression in *cis* and in *trans*.

The pioneering studies on gene regulation, e.g., the lac operon (Jacob and Monod, 1961), provide a conceptual framework for modern studies of gene expression regulation and evolution. Gene expression is affected by complex interactions involving both internal and external factors. Factorize and characterize the regulatory nature of the mechanisms underlying gene expression is a well-established approach to analyse the evolution of gene expression (Signor and Nuzhdin, 2018). The expression level of a given gene can be influenced by local or distant regulatory elements. Variants that are linked to the affected gene are typically described to act in *cis* and those unlinked are known as *trans*-acting factors, located physically distant or even on different chromosomes. Regulatory variants acting in *cis* mediate differential gene expression directly, e.g., through mutations in promoter elements that may introduce or impair a transcription factor binding site. In contrast, *trans*-regulatory polymorphisms influence the target gene indirectly, commonly via diffusible elements, e.g., changes in the abundance of an upstream regulator of gene expression (Stern and Orgogozo, 2008). In addition, specific variants in *cis*-regulatory elements are sometimes required to interact to another variant acting in *trans*, as recently described in tomato hybrids (Soyk *et al.*, 2017). Variation in gene expression is ubiquitous, arising within natural populations and described across different tissues, developmental stages and species. Much of this extensive variation is heritable and attributable to *cis*-elements and *trans*-regulatory factors, although their relative contribution is rarely similar both in frequency and effects (Gordon and Ruvinsky, 2012; Metzger *et al.*, 2016). In general terms, although *trans*-regulatory variants are more frequently observed, they produce smaller effects than *cis*. This is due to the complex patterns of multigenic inheritance associated to regulation in *trans*, as recently disclosed in yeast (Metzger *et al.*, 2016). In addition, the pleiotropic nature of *trans*-acting mutations may produce numerous deleterious effects that limit the strength of their effects (Prud'homme, Gompel and Carroll, 2007; Lemos *et al.*, 2008).

1.2.2.2 – Methods to characterize the regulatory components of gene expression.

One of the most common approaches to decompose the *cis* and *trans*-regulatory components of gene expression is based on the analysis of allele-specific expression (ASE), also known as allelic imbalance or preferential allelic expression, defined as the unequal expression of alleles within heterozygous loci (Gaur *et al.*, 2013). This method allows to distinguish *cis* and *trans* regulatory variants through the analysis of the heterozygous genotypes of hybrids in the first filial (F_1) generation generated by intercross between two homozygous strains. Strictly speaking, *cis* variants affect only the expression of genes located physically in the same chromosome and not in the homologue, whereas *trans* factors influence equally the two homologue copies of a gene. In the F_1 hybrid genetic background, the total contribution by *trans* factors would be the same for both alleles and thus the differences found between the expression of both alleles would be exclusively driven by *cis* effects. Following this method, regulation in *trans* can be estimated by comparing the specific expression of the one of the alleles in the parental background to that in the hybrid background. Moreover, *cis*-by-*trans* effects could be detected when there is ASE in the hybrid and the allelic ratio between the homozygous strains and the hybrid is different (Emerson and Li, 2010; Pastinen, 2010; Gaur *et al.*, 2013; Botet and Keurentjes, 2020). This approach was first used for a limited number of genes by means of pyrosequencing in organisms like Drosophila and yeast (Wittkopp, Haerum and Clark, 2004; Sung *et al.*, 2009). The emergence of high-throughput sequencing technologies like RNA-sequencing (RNA-seq) lead to extend this method to a global scale in diverse species, including plants like Arabidopsis (Zhang and Borevitz, 2009; Emerson *et al.*, 2010; McManus *et al.*, 2010). Another similar approach involves the analysis of multiple crosses between variable genotypes to a common tester. Briefly, this population design with a common reference allele allows to estimate *trans* factors across multiple backgrounds and *cis* effects within a unique background (Nuzhdin, Friesen and McIntyre, 2012; Fear *et al.*, 2016).

Although suitable to determine the regulatory nature of genetic variants, the previous strategies provide no information about the genomic source and location of gene expression differences (Emerson and Li, 2010; Signor and Nuzhdin, 2018). Such limitation can be surpassed through other of the most recurrent methods to characterize the *cis* and *trans* components of regulatory variation: the analysis of expression quantitative trait loci (eQTLs), used in yeast for the first time (Brem 2002). Similar to standard QTL analysis, eQTLs are sequence variants within genomic regions that partially explain differences in the expression of one or more genes (Nica and Dermitzakis, 2013) and can be detected using linkage mapping in experimental crosses (e.g., RIL populations), or by means of association analysis in outbred populations of unrelated individuals (e.g., GWAS). Current advances of high-throughput sequencing technologies have enabled to accurately detect and quantify thousands of genetic variants in large populations. These studies

reveal the location of the detected eQTLs and classify them as local or distal, according to the physical distance to the gene that encodes the expression trait. Strictly speaking, these terms refer to relative positions and not to molecular mechanisms of action, since both of them may act either in *cis* or in *trans*. Nevertheless, local eQTLs typically act in *cis* and thus they are used interchangeably. In general terms, *trans* effects are generally weaker than *cis*, which is a limitation to detect *trans*-eQTLs, given that they potentially target the entire genome and require statistical corrections for multiple testing, lowering the power of detection (Rockman and Kruglyak, 2006; MacKay, Stone and Ayroles, 2009; Albert and Kruglyak, 2015; Signor and Nuzhdin, 2018). The analysis of eQTLs and ASE can be used as complementary strategies allow to map the location of regulatory variants and easily classify them as either *cis* or *trans*-acting (Emerson and Li, 2010). Genome-wide studies often report hundred to thousands of *trans*-eQTLs with small effect mapping to the same region. These regions are known as *trans*-eQTL hotspots and may reveal the presence of potential master *trans*-regulators with strong influence on the expression of many genes (Keurentjes *et al.*, 2007; West *et al.*, 2007; Kliebenstein, 2009; Cubillos *et al.*, 2012; Hartanto *et al.*, 2020).

Last, understanding the patterns of propagation of *cis*-*trans* transcriptomic changes throughout GRNs is a promising field of research in plants. Although it is still an unexplored area, some noteworthy approaches in other organisms have attempted to analyse the distribution of gene expression differences within GRNs through structural equation models (Nuzhdin, Friesen and McIntyre, 2012; Yang and Wittkopp, 2017).

1.2.2.3 – The role of *cis*- and *trans*-regulatory variation in the evolution of gene expression. Current progress suggest that gene expression evolution is adaptive, although the long-term adaptive significance is still not well-established (Signor and Nuzhdin, 2018). The relative importance of *cis* and *trans*-regulation of gene expression to phenotypic divergence and adaptive evolution has been examined across multiple systems and conditions, providing an extensive framework for progressing in the understanding of gene expression regulation and evolution. *Cis*- and *trans*-regulatory variants have different modes of inheritance, which determine the pressures of selection which they are exposed to and emphasizes their evolutionary significance (Wray, 2007). Diverse studies have reported that *cis*-regulatory changes play a predominant role in the adaptive evolution of gene expression (Fraser *et al.*, 2011; Lemmon *et al.*, 2014; Signor *et al.*, 2016; Rebeiz and Williams, 2017). This is probably because *cis* variants, more frequently found in genes with structural rather than regulatory roles, result in a lower incidence of pleiotropic effects (Hoekstra and Coyne, 2007; Stern and Orgogozo, 2008). Nevertheless, other studies assume that *trans*-regulation, although frequently overlooked in these studies, may also produce

adaptive changes (Streisfeld and Rausher, 2009; O'Quin *et al.*, 2012), which illustrate the relevance of other contexts like the gene regulatory networks.

1.2.2.4 – *Cis-trans* compensatory changes and stabilizing selection.

Several studies using different approaches and systems often report the occurrence of *cis-trans* pairs producing opposite effects on the expression of a given gene (Landry *et al.*, 2005; Takahasi, Matsuo and Takano-Shimizu-Kouno, 2011; Goncalves *et al.*, 2012; Romero, Ruvinsky and Gilad, 2012; Metzger *et al.*, 2016; Verta, Landry and Mackay, 2016). These changes are common compensatory mechanisms for restoring or normalizing overall gene expression levels when small regulatory variants arise (Landry *et al.*, 2005; Takahasi, Matsuo and Takano-Shimizu-Kouno, 2011). The overall excess of *cis-trans* pairs producing compensatory instead of amplifying effects suggests that they exert a putatively adaptive role (Signor and Nuzhdin, 2018). Stabilizing selection, in contrast to purifying selection, shapes the evolution of genome-wide gene expression through favouring the non-extreme phenotypes in the population (Charlesworth, 2013). Recent progress points that compensatory evolution through *cis-trans* pairs is coherent with the hypothesis of stabilizing selection as the principal mode of evolution for gene expression both within and between species (Coolon, Mcmanus and Stevenson, 2014; Hodgins-Davis *et al.*, 2015). Nevertheless, this hypothesis is still under debate since the explicit contribution of compensatory *cis-trans* effects to stabilizing selection has not been surveyed to date in a comprehensive manner (Signor and Nuzhdin, 2018).

1.2.2.5 – The potential underlying mechanisms for *cis-trans* compensation.

Some hypothetical molecular mechanisms have been proposed to explain the occurrence of *cis-trans* compensation. First, compensatory mutations may occur when a pair of *cis-trans* mutations are spread jointly to compensate each other for a moderate deleterious effect in one of them. The coinheritance of such specific changes illustrates the coevolution of a *cis-trans* regulatory system (Landry *et al.*, 2005; Kuo *et al.*, 2010; Takahasi, Matsuo and Takano-Shimizu-Kouno, 2011; Shi *et al.*, 2012; Verta, Landry and Mackay, 2016). In addition, numerous *cis-trans* epistatic interactions have been reported (Mackay, 2014; He *et al.*, 2016), although their compensatory character still remain to be elucidated. The joint segregation of compensatory *cis-trans* pairs is conceptually equivalent to the classical hypothesis of dominance (Bourguet, 1999), according to which the slightly deleterious effects of a recessive mutation in *cis* would be compensated with a secondary mutation acting in *trans*. Second, there is evidence of another potential *cis-trans* compensatory mechanism caused by regulatory genes acting in *trans* through a positive or

negative feedback loop to restore the average expression levels of the target gene, buffering the effect of *cis* mutations (Ronald *et al.*, 2005; Bader *et al.*, 2015).

An additional *cis-trans* compensatory mechanism is associated to the phenomenon of transvection, originally observed in Drosophila dependent on chromosome pairing (Lewis, 1954; Pirrotta, 1990; Duncan, 2002) and later described in other organisms like mammals (Rassoulzadegan, Magliano and Cuzin, 2002; Sandhu *et al.*, 2009), but still unknown in plants. Briefly, regulatory elements from one allele may influence the expression of the corresponding copy on the homologue chromosome through either upregulation or downregulation. For instance, there is some evidence that *cis*-regulatory elements such as enhancers can also act in *trans* when the enhancer in the other allele is functionally deficient. Consequently, communication between alleles can be considered as a candidate *cis-trans* compensatory mechanism to stabilize gene expression (Williams, Spilianakis and Flavell, 2010; Signor and Nuzhdin, 2018). In addition, there is evidence of insulator proteins that mediate chromosomal pairing both in insects mammals and thus facilitate regulatory interactions between alleles (Fujioka *et al.*, 2016). The occurrence of transvection across species and the extensiveness of its effects on gene expression in natural populations are interesting areas that remain to be investigated (Signor and Nuzhdin, 2018).

1.2.3 – HYBRID VIGOUR IN PLANTS

1.2.3.1 – Insights into hybrid vigour in plants.

Agriculture has long exploited the enhanced performance of hybrids, which can be studied through quantitative genetics. In this respect, a considerable amount of natural variation for attributes related to plant growth, development, and yield accounts for the increased fitness observed in hybrid crop species, although the underlying molecular mechanisms require further research (Alonso-Blanco *et al.*, 2009; Schnable and Springer, 2013; Botet and Keurentjes, 2020). The widespread phenomenon of plant hybrids surpassing their parents was first addressed in the eighteenth century through hybridization assays conducted by the naturalist Joseph G. Kölreuter (Mayr, 1986). Gregor Mendel's experiments of cross-hybridization and his theories of inheritance provided a different perspective (Mendel, 1977). Diverse experiments concerning self- and cross-fertilization in plants lead Charles R. Darwin to describe the detrimental and favourable effects of what we currently associate to inbreeding and heterotic phenotypes, respectively (Darwin, 1876). Following Darwin's assumptions, the development of the first hybrid corn by the American botanist William J. Beal evidenced the occurrence of hybrid vigour in maize (Troyer, 2006). The term 'heterosis' was first introduced in 1914 by the American geneticist George H. Shull, replacing 'heterozygosis' (Shull, 1948). Today, heterosis or hybrid vigour is described as the increased or improved, instead of superior, biological fitness in a hybrid offspring over the

parents. Hybrids within or between species may exhibit enhanced biomass, growth, survival and fertility (Birchler *et al.*, 2010; Schnable and Springer, 2013). It should be noted however that as a result of combining two genotypes through hybridization, the biological fitness in terms of survival and fertility in the offspring may increase (hybrid vigour) or drop (hybrid incompatibilities: inbreeding depression). These two long-recognized phenomena are associated to the terms of positive and negative heterosis, respectively (Alonso-Blanco *et al.*, 2009; Oakley *et al.*, 2019). From an evolutionary perspective, the improved performance and reproductive fitness associated to hybridization has long suggested to have played a significant role in plant evolution (Stebbins, 1959; Grant, 1971; Govindaraju, 2019).

In Arabidopsis, heterosis appears to be widespread as its occurrence has been systematically surveyed and described across different genotypes, developmental stages and for diverse phenotypic traits, such as photosynthetic efficiency (Sharma, Griffing and Scholl, 1979; Fujimoto *et al.*, 2012), cold tolerance (Rohde, Hincha and Heyer, 2004; Korn *et al.*, 2008), phosphate use efficiency (Narang and Altmann, 2001), metabolite contents (Lisec *et al.*, 2009; Korn *et al.*, 2010; Meyer *et al.*, 2012), seedling viability (Mitchell-Olds, 1995), seed number (Alonso-Blanco *et al.*, 1999), flowering time and diverse growth, biomass and seed-related traits (Barth *et al.*, 2003; Meyer *et al.*, 2004, 2010, 2012; Syed and Chen, 2005; Kusterer, Muminovic, *et al.*, 2007; Kusterer, Piepho, *et al.*, 2007; Melchinger *et al.*, 2007; Perera, Pooni and Kearsey, 2008; Oakley, Ågren and Schemske, 2015; Lauss *et al.*, 2018; Mehraj *et al.*, 2020).

1.2.3.2 – The biological mechanisms underlying heterosis.

Despite the progressive advance in understanding the biological function of genes and the development of new genomic tools, our current knowledge about the underlying mechanisms of heterosis is insufficient. The well-established models of dominance and overdominance are frequently used to explain hybrid vigour, although some additional mechanisms may contribute, none of which are mutually exclusive (Birchler *et al.*, 2010; Schnable and Springer, 2013; Fujimoto *et al.*, 2018). According to the traditional hypothesis of dominance (Davenport, 1908; Wright, 1934), recessive alleles from one inbred parent would be complemented by dominant alleles from the other. As a result of combining both parental genotypes in the hybrid background, the slightly deleterious effects from multiple recessive detrimental mutations could be masked or supressed, whereas the cumulative effect of multiple favourable alleles would lead to hybrid vigour emergence (Birchler, Yao and Chudalayandi, 2006; Birchler *et al.*, 2010). The overdominance model (East, 1908; Shull, 1908) states that the heterozygous condition at single-locus is more advantageous for a certain combination of alleles than the homozygous state. The occurrence of overdominance at single-locus has been described in plant species like maize

(Hollick and Chandler, 1998), tomato (Krieger, Lippman and Zamir, 2010) and Arabidopsis (Moore and Lukens, 2011). Pseudo-overdominance (Jones, 1917), is related model that mimics overdominance effects and it is originated by the tight linkage in phase repulsion of loci exhibiting dominance individually (Schnable and Springer, 2013). This may also contribute to generate heterosis (Li *et al.*, 2015; Hedrick, Hellsten and Grattapaglia, 2016).

Although strong evidence suggests that both dominance and overdominance produce cumulative effects that account for heterosis to some degree, they cannot fully explain this phenomenon and thus the contribution of additional mechanisms is still fundamental to complement these hypotheses (Birchler *et al.*, 2010). One of these mechanisms could consist of novel epistatic interactions between alleles inherited from different parents that would take place in the hybrid background between either allelic or non-allelic loci (Powers, 1944; Williams, 1959). These interactions would account for heterosis through, e.g., the masking of detrimental allele effects in the hybrid background (Birchler, Yao and Chudalayandi, 2006). Additional hypotheses state that the balance for regulatory components that are dosage-sensitive could be more advantageous in the hybrids. This type of mechanism could also be implicated in phenotypic variation for evolution through adaptation and in speciation due to post-zygotic hybrid incompatibilities (Birchler *et al.*, 2005; Birchler and Veitia, 2007). Another model proposes that the hybrid could preferentially regulate the transcription or translation of specific alleles to avoid variants that encode non-optimal protein components. The resulting improvement of hybrid performance could be associated to the use of more efficient proteins that substantially improves the efficiency of energy use and accelerates cell cycle (Goff, 2011).

Additionally, the role of gene expression and epigenetics in the regulation of heterosis has been disclosed (Chen, 2013; Groszmann *et al.*, 2013; Botet and Keurentjes, 2020). In this context, the most relevant finding is *DECREASE IN DNA METHYLATION 1* (*DDM1*), a chromatin remodeler involved in DNA methylation maintenance that appears to play a key role in regulating full heterosis development. (Kawanabe *et al.*, 2016; Zhang *et al.*, 2016). The use of epi-RILs populations has also provided evidence for the contribution of epigenetic variation to heterosis (Dapp *et al.*, 2015; Lauss *et al.*, 2018). Alterations in circadian rhythms appear to be critical for hybrid vigour. Epigenetic changes for the circadian clock genes *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) and *LATE ELONGATED HYPOCOTYL* (*LHY*) are associated to increase in hybrid growth and biomass in Arabidopsis (Ni *et al.*, 2009). Moreover, transgenerational effects like genomic imprinting (i.e., the transmission of epigenetic marks from the parental lines onto the progeny, leading to the differential expression of some genes influenced by the parent of origin) and maternal effects (including cytoplasmic and sporophytic effects) may have an impact on heterosis (reviewed by Botet and Keurentjes, 2020).

2. OBJECTIVES

The general objective of this thesis is to contribute to the understanding of the evolutionary processes that underlie natural variation of the transcriptome in Arabidopsis and identify candidate genes determining variation of key aspects of the response to phosphate starvation. In relation with this generic objective, we have pursued the following specific objectives:

1. Identification and characterization of *cis*- and *trans*-regulators that determine transcriptome variation in response to phosphate starvation and study the influence of environmental and developmental cues on *cis*- and *trans*-regulatory effects.
2. Determination of the mode of inheritance of *cis*- and *trans*-eQTLs and the role of different evolutionary processes in the natural variation of the transcriptome.
3. Establishment of relationships of coexpression modules and *trans*-eQTLs with plant performance and physiological traits related to the response to phosphate starvation.

3. MATERIALS AND METHODS

3.1 – PLANT MATERIAL AND GROWTH

3.1.1 – OBTENTION OF SEED STOCK

All the plant material used in this thesis is based on the species *Arabidopsis thaliana* (L. Heynh). The accessions Nossen (No-0), Eringsboda (Eri-1), as well as the RIL populations L*er* x No-0 (Magliano *et al.*, 2005) and L*er* x Eri-1 (Ghandilyan *et al.*, 2009) were obtained from the Nottingham Arabidopsis Stock Centre (NASC, Nottingham, UK (http://arabidopsis.info/). The accessions Landsberg *erecta* (L*er*) and Llagostera (Ll-0), as well as the RIL population L*er* x Ll-0 (Sánchez-Bermejo *et al.*, 2012) were kindly provided by Carlos Alonso-Blanco (CNB-CSIC).

3.1.2 – PLANT CULTIVATION *IN VITRO*

3.1.2.1 – Seed sterilization and stratification treatments.

Approximately 4 days before sowing, Arabidopsis seeds were subjected to coat surface sterilization to prevent medium contamination. Briefly, seeds were put in a 1.5 ml tube and then 1 ml of washing solution containing 70% v/v of commercial bleach and 0.02% v/v Polysorbate 20 (commonly known as Tween 20) was added. After 8-10 min, approximately, the solution was removed and the seeds were rinsed 2-3 times with bi-distilled sterile water. This process was carried out under sterile conditions in a laminar flow cabinet. Fresh sterilized seeds were subjected to cold stratification at 4ºC under dark conditions to break seed dormancy and ensure optimal germination rate.

3.1.2.2 – Plant culture medium and growth conditions *in vitro*.

A modified Johnson medium (Johnson *et al.*, 1957; Bates and Lynch, 1996) was used for plant cultivation. pH was adjusted with KOH and KH_2PO_4 was supplied in two different concentrations, 5 µM and 1 mM, depending on the desired low or high phosphate conditions, respectively. Plants grew in chambers under controlled conditions of 22ºC, 60% humidity, long-day photoperiod (16h light and 8h darkness) and photosynthetically photon flux density of 100µmol s^{-1} m^2.

For purposes like plant genotyping or multiplication, a variable number of seeds were sown in horizontal plates supplied with 1 mM Pi and 0.6% agar. For the phenotypic analysis, seeds were sown in 12x12 cm plates containing sterile solid culture medium. Bacto™Agar (Becton Dickinson Microbiology) concentration was 1% (w/v) to prevent roots from penetrating the agar upon vertical growth. We inserted the plates in methacrylate structures to maintain them in vertical position. We alternated the position and orientation of these structures every two days to attenuate plant growth and physiology variation that might be caused by fluctuations in the conditions of humidity and light inside the chamber.

3.1.3 – SEED MULTIPLICATION AND PLANT CROSSING

3.1.3.1 – Seed multiplication.

All the Arabidopsis seed stocks obtained were multiplicated to refresh and accumulate enough seeds. Germination took place in agar plates under *in vitro* conditions (as described in **3.1.2.2**). After a few days, seedlings grew and reached a suitable size to be transferred into individual pots containing a mix of soil and vermiculite (3:1). Once transplanted, plants completed their life cycle in growth chambers under long-day photoperiod (16h light and 8h darkness) at 21-22ºC. Seeds from dry siliques were harvested, bulked, and maintained in small paper bags at RT.

3.1.3.2 – Vernalization treatment.

Since the parental accession Ll-0 and the set of L*er* x Ll-0 RILs display late flowering phenotypes with an enhanced response to vernalization (Shindo *et al.*, 2005), we subjected them to a 6-week vernalization treatment at 4ºC prior to crossing in order to homogenize the time of flowering.

3.1.3.3 – Crossing method.

All the plants to be crossed were grown in individual pots in growth chambers (as described **3.1.3.1**). Plants were considered suitable for crossing when they had at least 3-4 inflorescences. We carefully removed lateral stems and only the inflorescences from the main stem were used for crossing to maximize siliques yield, cross efficiency, and reduce variation. Only flower buds with a discernible white tip and still not completely open were regarded as suitable for emasculation, as they were presumably at the stage at which the pistils in their inside are mature enough for crossing but still not self-fertilized. Then we used a pair of fine forceps, freshly sterilized with 96% ethanol, to emasculate the flower buds by carefully removing all petals, sepals, anthers, and attempting not to touch the pistil. Emasculated plants were set apart and left to grow for a few hours prior to pollination. We obtained mature pollen from plants with open inflorescences and fully developed siliques. Pollination was systematically performed by taking a few anthers with discernible pollen shedding from one plant and gently covering the stigma of another plant with pollen grains. This was repeated a few hours later to ensure pollination and promote silique development. Plants with pollinated inflorescences were put apart to mature and labelled with the date and the parental lines. After a few weeks, dry siliques were collected before opening.

3.1.3.4 – Crossing strategy.

According to the plan, a first set of crosses was generated through the combination of three pairs of accessions: L*er* x No-0, L*er* x Eri-1 and L*er* x Ll-0. These plants were grown under the same

conditions and crossed approximately in the same lapse of time. A second set of L*er* x Ll-0 parental crosses, as well as the L*er* x RIL and Ll-0 x RIL crosses were generated in parallel. In this case we used vernalized plants (see section **3.1.3.2**) to ensure that plants could be crossed approximately at the same time.

Regarding the parental crosses, for each pair we obtained reciprocal crosses by reversing in turn the accession used as female and male to evaluate the occurrence of parental imprinting and maternal effects in the F_1 hybrids. Moreover, the parental lines were crossed to themselves (using different plants, not self-crossing) so that the external effects caused by the process of crossing itself affect equally to all genotypes in the comparisons between hybrids and parents. As for the parental x RIL crosses, L*er* and Ll-0 were used as the female parents in all cases, while the RILs were always used as the male parent.

All these crosses were genotyped to confirm their identity. One or two random seeds from each silique were grown for one week and then genotyped using a minimum of five genome-wide markers (see section **3.4**). Siliques either yielding less than twenty seeds or with any seed detected to carry the wrong genotype were excluded. After this procedure, seeds from the non-discarded siliques were pooled into bulks of the same cross.

3.2 – PLANT PHENOTYPING METHODS

3.2.1 – MEASUREMENT OF PLANT WEIGHT AND WATER CONTENT

The freshly collected plant material was directly put into 1.5 ml safe-lock empty tubes, previously weighed, to prevent it from drying out. Plant fresh weight was calculated as the difference between the weights of full and empty tubes, and it was reported in milligrams. After that, samples were heated in the inside of an oven at 50-60ºC for 3-4 days. Once dried, samples were taken out of the tube and weighed to obtain the dry weight values. All the measurements were done using a high-precision weighing scale. Water content was estimated as the percentage of lost weight between dry and fresh samples. The root-to-shoot ratio was calculated by dividing the weight of root biomass by the weight of shoot.

3.2.2 – QUANTIFICATION OF PLANT PHOSPHATE CONTENT

Approximately 2-10 mg of dried plant shoots and roots of Arabidopsis were used as input material to quantify the content of soluble phosphate using a colorimetric method, implemented from a modified version of a previous standard protocol (Ames, 1966). 1 ml of 1% glacial acetic acid was added to each sample and then subjected to three freeze-thaw cycles. To prepare the reactions,

25 µl per sample, including the ones for the calibration curve, were mixed with a solution containing 18 g/l ascorbic acid and 3.5 g/l ammonium molybdate in 0.6 M H_2S0_4. Then, samples were incubated for 2 h at RT so that the reactions could reach saturation, thus reducing variation in the measurements. Thereafter, samples were transferred to spectrophotometer cuvettes and optical density was measured at 820 nm. The concentration of orthophosphate was estimated from the calibration curve (ranging from 0 to 1 mM KH_2PO_4).

3.2.3 – QUANTIFICATION OF ANTHOCYANIN CONTENT

10-45 mg (27 mg on average), approximately, of fresh plant shoots grown under low Pi conditions were used as input material to extract and estimate anthocyanin content using a modified version of a previous protocol (Swain and Hillis, 1959). Briefly, 1 ml of a solution containing 0.5N HCl and 4:1 CH_3OH was added to each sample and kept for at least 24 h under dark conditions to prevent anthocyanins from photobleaching. Then, 500 µl of each sample were mixed with an equal volume of a solution containing 3N HCl and 16.6%(v/v) CH_3OH. One sample was bleached by adding 10%(v/v) H_2O_2 to be used as blank solution for absorbance measurement. Thereafter, all samples were incubated for 15 minutes under dark conditions and then transferred to spectrophotometer cuvettes to measure optical density at 535 and 657 nm and estimate from these the anthocyanin concentration. Anthocyanin content was calculated as [$Abs_{530\ nm}$ – 0.25 x Abs_{657nm}] x [2 / fresh weight], where 2 correspond to the dilution factor, and it was reported as normalized anthocyanin index.

3.2.4 – MEASUREMENT OF PRIMARY ROOT LENGTH

Pictures of vertical plates were taken at 7, 9 and 12 DAS and the length of primary root is measured using the software ImageJ version 1.52a (Schneider, Rasband and Eliceiri, 2012).

3.3 – EXPERIMENTAL DESIGN FOR PLANT MATERIAL SAMPLING

For the phenotypic and transcriptomic analysis of parental crosses and parent x RIL crosses, seeds were sown approximately three months after being harvested to ensure an optimal germination rate. Parental crosses were replicated three times, while the parent x RIL crosses had a single replicate, as each RIL can be considered itself as a biological replicate. Approximately 50-60 seeds per genotype and replicate were sown across low and high Pi plates, each one with 12 and 8 seeds, respectively.

According to the schedule, we took pictures of the plates at 7, 9 and 12 DAS to measure the length of primary root. Sampling was performed using plants grown for 12 days under *in vitro* conditions (**3.1.2.2**). Samples for RNA extraction were collected in the first place and always at the same hour to reduce the temporal variation of sampling and the subsequent variation in transcript levels caused by the circadian rhythm. These samples were directly frozen in liquid nitrogen and then stored in freezers at -80°C until RNA extraction. We then collected samples for measuring physiological and morphological traits. One set of samples were put apart for quantification of anthocyanin content, and a second one for measurement of Pi content, fresh weight, dry weight, as well as other traits derived from these. In all cases, including RNA samples, plant shoots and roots were collected separately. Each sample was homogenized by collecting a balanced amount of plant material from all the plates of identical genotype and Pi conditions. To control for the variation caused by the measuring process itself, each sample was halved and put in two different tubes to be separately measured. The two values obtained were used to estimate a single mean value per sample.

3.4 – PLANT GENOTYPING

3.4.1 – DNA EXTRACTION

Genomic DNA for plant genotyping was extracted from fresh leaves of Arabidopsis seedlings using the alkaline lysis method (Klimyuk *et al.*, 1993).

3.4.2 – PCR-MEDIATED AMPLIFICATION OF DNA

Genome-wide insertion and deletion (InDel) polymorphisms of approximately 25-60 bp were used as molecular markers to genotype among accessions. The forward and reverse primers targeting the sequences containing these InDels were acquired at Sigma-Aldrich and listed in **Table M1**. The genomic DNA regions of interest were selectively amplified through PCR (Sambrook, Fritsch and Maniatis, 1989) with the activity of the Taq DNA polymerase (Roche).

3.4.3 – GEL ELECTROPHORESIS

The fragments of DNA selectively amplified by PCR were separated through standard agarose gel electrophoresis and visualized under UV light. These gels were made of 1x TAE buffer (50mM Tris-acetate, 1mM EDTA, pH 8), ethidium bromide (0.5 µg/ml) and a high concentration of agarose (30 mg/ml) so that the small differences in the size of the fragments could be detected.

Table M1. **List of primers used for genotyping**. The forward (Fwd) and reverse (Rvs) sequences (5'-3') of primers are based on the sequences reported in the original publications: hybrids L*er*/No-0 and L*er*/Eri (Magliano *et al.*, 2005; Ghandilyan *et al.*, 2009); hybrids Ler/Ll-0 and RILs (Sánchez-Bermejo *et al.*, 2012).

NAME	SEQUENCE (5' - 3')	DESCRIPTION
Genotyping of hybrids L*er* x No-0 and L*er* x Eri-1		
128_NGA106_Fwd	TGCCCCATTTTGTTCTTCTC	Deletion Molecular Marker
129_NGA106_Rvs	GTTATGGAGTTTCTAGGGCACG	Deletion Molecular Marker
Genotyping of hybrids L*er* x Ll-0, L*er* x RIL, Ll-0 x RIL		
114_NGA76_Fwd	GGAGAAAATGTCACTCTCCACC	Deletion Molecular Marker
115_NGA76_Rvs	AGGCATGGGAGACATTTACG	Deletion Molecular Marker
196_F14J9_Fwd	CGCAATAACTCGACCGGTAAG	Deletion Molecular Marker
197_F14J9_Rvs	TTTCACATGCACTCTCGCCTTTC	Deletion Molecular Marker
198_T27K12_Fwd	GGACAACGTCTCAAACGGTT	Deletion Molecular Marker
199_T27K12_Rvs	GGAGGCTATACGAATCTTGACA	Deletion Molecular Marker
200_T8K144_Fwd	CAATGCGCTCTGAATCTCTGAC	Deletion Molecular Marker
201_T8K144_Rvs	CACTTCGTCGTCTAATCCAGGT	Deletion Molecular Marker
202_F26B6_Fwd	AACTCTATCTGCCCACGAACAA	Deletion Molecular Marker
203_F26B6_Rvs	TTCTGCAAGCCTCAACTACTAACTG	Deletion Molecular Marker
204_T26D22_Fwd	AAATGGAAAAGTTGATACAGCT	Deletion Molecular Marker
205_T26D22_Rvs	ACCAACCTTACTATTACAAACACAA	Deletion Molecular Marker
206_K8K14_Fwd	TAAACACCACGACAAATACAACAT	Deletion Molecular Marker
207_K8K14_Rvs	TTGCTTCCTATATCTACTTG	Deletion Molecular Marker

3.5 – RNA ISOLATION AND SEQUENCING

3.5.1 – RNA EXTRACTION AND PURIFICATION

Total RNA was isolated following the High Pure RNA Isolation Kit protocol (Roche), with some modifications to the supplier's protocol. First, all the frozen plant tissue samples were pulverized and homogenized using a Silamat S6 bead mill. Then we used Z6 extraction buffer (guanidinium chloride 8M, MES 20mM, EDTA 20mM, β-mercaptoethanol 50mM, pH 7.0) and Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v). To remove cell debris, the samples were centrifuged for at least 30 min at 4ºC. Then, the upper phase was mixed with the lysis/binding buffer and then we followed the instructions of Roche's kit protocol, which includes the addition of DNAse I for removal of genomic DNA traces. Finally, after RNA isolation we examined RNA purity and concentration with a NanoDrop® ND-1000 UV-VIS spectrophotometer. Samples were retained when the ratios of the absorbance at 260 and 280 nm ($A_{260/280}$) and at 230 and 260 nm ($A_{260/230}$) were higher than 2.0 and 1.8, respectively. In addition, we discarded samples with concentrations lower than 50 ng/µl to ensure meeting Novogene's requirements for sequencing total RNA samples from eukaryotes.

3.5.2 – RNA SEQUENCING

RNA sequencing (RNA-seq) was conducted by Beijing Novogene Bioinformatics Technology Co. Ltd.

3.5.2.1 – Qualification and quantification of RNA.

Initially, degradation and contamination of RNA was examined on 1% agarose gels. A NanoPhotometer® spectrophotometer (IMPLEN, CA, USA) was used to check RNA purity and the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA) was used for assessment of RNA integrity and quantitation.

3.5.2.2 – Library preparation for RNA sequencing.

Sequencing libraries were constructed using a total amount of 1μg of purified RNA as input material per sample and the NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA), according to the manufacturer's instructions. The RNA-seq libraries constructed were sequenced on an Illumina NovaSeq 6000 platform. The read length was of 150 bp and at least 20 million paired-end reads (read pairs) per sample were generated.

3.5.3 – ANALYSIS OF RNA-SEQ DATA

3.5.3.1 – Quality control.

Initially, raw RNA-seq reads were processed using the tool 'fastp' (Chen *et al.*, 2018) to obtain clean data by removing the reads with low quality and also the ones containing adapter and poly-N sequences. Quality scores (Q20 and Q30) and GC content were estimated on the clean reads and the high-quality data obtained (Q30 > 95%) was used in all the downstream analyses.

3.5.3.2 – Mapping to reference genome.

The files for the reference genome (Col-0) and the gene model annotation were directly obtained from genome website browser (NCBI/UCSC/Ensembl). The high-quality paired-end clean reads were aligned to reference genome through the method 'HISAT2' (Kim *et al.*, 2019).

3.5.3.3 – Quantification of gene expression and variant calling.

The program 'featureCounts' (Liao, Smyth and Shi, 2014) was used to quantify the number of reads mapped of each gene. Gene expression levels are reported in fragments per kilobase of exon model per million mapped paired-end reads (FPKM), calculated by normalizing for sequencing depth and gene length at the same time.

The tool 'HaplotypeCaller' in GATK software (McKenna *et al.*, 2010) was used for single nucleotide polymorphisms (SNPs) calling. The genetic variants detected were functionally annotated using the tool 'ANNOVAR' (Wang, Li and Hakonarson, 2010).

3.5.3.4 – Analysis of differential gene expression.

The 'DESeq2' R package (Love, Huber and Anders, 2014) was used for differential expression analysis between two conditions using three biological replicates per condition. Genes were assigned as differentially expressed (DEGs) when the adjusted p-value estimated by DESeq2 was lower than 0.05 and a fold-change of ±2.0 (i.e., ≥2 or ≤-2) for differential expression between high and low phosphate conditions, or ±1.5 for other analysis.

3.6 – QTL ANALYSIS

3.6.1 – DESIGN OF THE EXPERIMENTAL MAPPING POPULATIONS

Initially, a total of 50 L*er* x Ll-0 RILs were selected and 17 of them were crossed to the progenitor L*er*, another 17 only to the other parental Ll-0, while 16 were crossed to both L*er* and Ll-0. The 66 crosses generated were classified in two populations. The first one, named L*er* x RIL to simplify, consisted of 33 individuals obtained from crossing L*er* to 33 different RILs. Similarly, the individuals from the population Ll-0 x RIL derived from crosses between Ll-0 and 33 RILs. Due to the small size of these two populations, RILs were carefully selected based on their genotype to ensure a balanced genotypic distribution on each marker (between 40% and 60%) and thus reduce the effect of distortion in segregating markers on the construction of the genetic linkage map and subsequent QTL analysis. We observed that 3 of the 66 lines displayed odd patterns of gene expression related to an excess (1.2-1.4x) in the read counts for the maternal over the paternal allele in the heterozygous regions. For this reason, these lines were discarded for QTL analysis, and thus 32 and 31 lines remained in the L*er* x RIL and Ll-0 x RIL populations, respectively.

3.6.2 – GENETIC MARKERS AND MAP CONSTRUCTION

Based on the information obtained from the allele-specific expression of the RILs analysed, the RILs were genotyped for the whole set of SNPs segregating between L*er* and Ll-0. Briefly, SNPs located on heterozygous regions (displaying counts for both L*er* and Ll-0 haplotypes) were genotyped as Ll-0 in the L*er* x RILs, and as L*er* type in the Ll-0 x RILs. Conversely, SNPs from homozygous regions were classified as L*er* in the L*er* x RILs, and as Ll-0 type in the Ll-0 x RILs.

Once genotyped, some of these SNPs were selected as molecular markers, handpicked at physical intervals ranging from 1 to 2 Mb, approximately, with at least one recombination event between two adjacent markers. Among the SNPs with high expression values across the RILs were selected as markers. We obtained no missing data for any marker.

With the information from these marker loci and the JoinMap 3.0 software package (Ooijen and Voorrips, 2001), we constructed a genetic linkage map, setting RI8 as the mapping population type. Markers stayed on the same linkage group from LOD (logarithm of odds) values of 3.0 to 7.0. For marker arrangement within linkage groups, mapping thresholds of REC = 0.4, LOD = 1 and JUMP = 5 were applied. The Kosambi's mapping function was used to estimate genetic distances in cM from recombination frequencies.

3.6.3 – QTL AND eQTL ANALYSIS

3.6.3.1 – Pre-processing prior to QTL mapping.

The two populations were separately pre-processed and analysed. For the physiological and morphological traits, the mean values per line differing by more than three times the standard deviation from the mean of the lines for a given trait were counted as outliers (Keurentjes *et al.*, 2007) and discarded for the subsequent QTL analysis. After finding and removing outliers, the Shapiro-Wilk test of normality was applied using the 'shapiro.test' function (Royston, 1995) in R version 4.0.3 (R-Core-Team, 2020). Traits with non-normal distribution ($p<0.05$) were subjected to Box-Cox (Box and Cox, 1964) and/or Yeo-Johnson (Yeo and Johnson, 2000) transformations using the 'forecast' (Hyndman and Khandakar, 2008; Hyndman, 2015) and 'bestNormalize' (Peterson and Cavanaugh, 2020) R packages, respectively. After applying these transformations, traits still showing non-normal distribution remain untransformed or converted with the one retrieving the lowest p-value from the Shapiro-Wilk test.

Regarding eQTL analysis, for the expression values per gene, measured in FPKM, a total of 19,017 genes were retained after discarding the ones showing low expression (<20 total read counts) in both populations, considering the average mean values across the RILs. For each expression trait, expression values were normalized using the $\log_2$ transformation and the outliers were detected and removed as above described.

3.6.3.2 – QTL analysis.

The QTL cartographer software package (Wang, Basten and Zeng, 2012) was used to carry out all the QTL and eQTL analysis. For each trait, composite interval mapping (CIM) (Zeng, 1993, 1994) was performed for genome scanning to detect evidence of QTLs affecting any trait. We selected backcross as the cross type and the Kosambi's mapping function. The control parameters

used for CIM were: model 6 (standard), forward stepwise regression method, window size of 10 cM, and 5 control markers. Walking speed for genome scan was set to 8 cM to minimize the number of intermediate pseudo-markers and save computation time. Each marker position was tested for QTL effect and additive effects and the proportion of phenotypic variance explained (R^2) by individual QTLs were estimated for each trait. Additive effects were calculated as the difference between the means of the two genotypic groups. Considering the small population size, LOD score peaks were located using 15 cM as the minimum genetic distance to distinguish close QTLs and a minimum LOD of 1.0 from top to valley. LOD score peaks exceeding the threshold level were declared as significant QTLs/eQTLs. For eQTL analysis, QTL cartographer was used to compute 1,000 permutations tests on 500 randomly selected genes. Thresholds values for each gene were determined as significant at the 5% level (α=0.05) (Churchill and Doerge, 1994; Doerge and Churchill, 1996) and the average of them was used to obtain an approximate genome-wide LOD score significance threshold for each population. Similarly, this method was applied in the QTL analysis of morphological and physiological traits. Additionally, we applied the method of Benjamini-Hochberg to correct for multiple testing (Benjamini and Hochberg, 2000). Considering all the LOD score peaks with LOD ≥1.5, FDR was adjusted for the number of genes tested and we obtained the corresponding q-value estimates for each peak. Considering a physical distance equal or less than 1 Mb between the inferred eQTL position at a genome marker and the start/end position of a given gene, LOD score peaks were classified as local/*cis*, whereas distances greater than 1 Mb were used to define distal/*trans*-eQTLs. To declare *trans*-eQTL hotspots, significance threshold values for each genetic location were determined based on the binomial probability ($p<0.05$) that the number of eQTLs could significantly exceed the number of expected assignations in case of random distributions.

3.7 – CONSTRUCTION OF GENE COEXPRESSION NETWORKS

3.7.1 – WEIGHTED GENE CORRELATION NETWORK ANALYSIS

Gene coexpression modules were obtained through weighted gene correlation network analysis (WGCNA) using the package "WGCNA" (Langfelder and Horvath, 2008, 2012) in R (R-Core-Team, 2020). For the set of PS-induced genes we used a soft threshold of 7, corresponding to the lowest power that had a correlation coefficient >0.9 with a scale-free network topology. For the PS-repressed, the soft threshold was set in 6. We used the "signed" network, in which nodes with negative correlation are considered unconnected (their connection strength is zero or very close to zero). The eigengene expression value of each module was determined using WGCNA functions, as well as the correlations between module eigengenes and the phenotypic traits.

3.8 – OTHER ANALYSIS

3.8.1 – ANALYSIS OF GO TERM ENRICHMENT

Analysis of Gene Ontology (GO) term enrichment were performed using the software platform VirtualPlant 1.3 (Katari *et al.*, 2010). The *Arabidopsis thaliana* Col-0 TAIR10 genome was selected as the background population and the classification schemes used are the GO biological process assignments by (TAIR/TIGR) and the functional classification by the Munich information center for protein sequences (MIPS). To determine over-representation, p-values were estimated using a Fisher's exact test and adjusted for multiple testing using FDR correction.

3.8.2 – ANALYSIS OF EXPRESSION COHERENCE

The Genevestigator V3 database (Hruz *et al.*, 2008) of normalized microarray experiments was used to analyse gene expression patterns using hierarchical clustering tools for similarity search and selecting the set of conditions described as "perturbations" from the platform Affymetrix Arabidopsis ATH1 Genome Array.

3.8.3 – COMPARISON OF GENE LISTS

Comparisons among multiple gene lists were carried out using either the online tool Venny 2.1.0 (Oliveros, 2007) or the function "intersect" in R (R-Core-Team, 2020). Significance for over- and underrepresented coincidences between gene lists was determined through binomial probability tests using the function 'pbinom' in R.

3.8.4 – ANOVA

One-way analysis of variance (ANOVA) was performed to compare phenotypic means using the genotype as the grouping variable. Pairwise comparisons were examined through the Bonferroni post hoc test to determine which specific means display significant differences. These statistical analysis were performed using the software package SPSS Statistics (IBM-Corp., 2020).

4. RESULTS

4.1 – ANALYSIS OF PARENTAL CROSSES

The strategy for detection and characterisation of QTLs, including the analysis of their mode of inheritance, followed this sequence of steps: i) Selection of the parental lines that themselves or their hybrids display the highest differences in the main morphological and physiological traits related to PSRs; ii) Comparative transcriptomic analysis of the selected parental lines and their hybrids and parallel analysis of physiological traits; iii) identification and characterisation of QTLs and eQTLs in hybrids of lines from a selected RIL population to both progenitors following classical quantitative genetic approaches.

4.1.1 – SELECTION OF THE PARENTAL LINES

Three sets of RIL populations of Arabidopsis, originated from crosses of the parental line Landsberg *erecta* (L*er*) to Nossen (No-0), Eringsboda (Eri-1) and Llagostera (Ll-0), respectively, were accessible in our laboratory. Initially, to determine which of them was potentially the most informative, we decided to analyse the performance of all the possible hybrids derived from parental crosses between L*er* and the other three accessions. Hence, we generated through hand-pollination (see **section 3.1.3**) a progeny of seeds carrying ten different genotypes: L*er* x L*er*; L*er* x No-0; No-0 x L*er*; No-0 x No-0; L*er* x Eri-1; Eri-1 x L*er*; Eri-1 x Eri-1; L*er* x Ll-0; Ll-0 x L*er*; Ll-0 x Ll-0. These hybrids were analysed in the first filial (F_1) generation to determine which of them displayed the highest phenotypic differences over their respective parental lines, with a particular interest for the ones showing signs of heterosis. The phenotypic analysis was performed *in vitro* under Pi starvation conditions (5µM Pi) and replicated three times (see **section 3.3**). Twelve days after sowing (DAS), plant material was collected and used to quantify/measure some of the most important traits related to PSRs, like plant biomass, water, and phosphate content, focusing on analysing the differences between F_1 hybrids and the parental lines. Many of the analysed traits varied across the different genotypes, as two or more significantly different phenotypic classes could be set up. When comparing L*er* x L*er* to the other three parental lines, Eri-1 x Eri-1 displayed the highest differences, as it did in six of the sixteen traits examined. Conversely, Ll-0 x Ll-0 exhibited the highest similarities, as only one trait showed differences. Nevertheless, when also including the hybrids in these comparisons, the highest differences were found in the parental crosses between L*er* and Ll-0. In these, F_1 hybrids exhibited differences to one of the parental lines in six traits, and three of them to both parents, which is indicative of heterotic effects. None of the other F_1 hybrids analysed appeared to display significant differences to their respective parents for any of the traits (**Table R1**). Based on these results, the combination between L*er* and Ll-0 was selected for subsequent analysis.

Table R1. **Phenotypic response to phosphate (Pi) starvation in ten parental crosses between L*er*, No-0, Eri-1, and Ll-0**. All traits were analysed under Pi starvation conditions (5µM Pi). The mean values obtained from the three replicates and the standard deviation are shown for each genotype. Analysis of variance (ANOVA) was performed to compare between genotypes for all trait values, assuming significance at the 0.05 probability level. The phenotypic classes assigned are shown in separate columns (class). Intermediate categories are assigned to genotypes that share similarity with two different phenotypic classes. The rows highlighted in yellow represent cases of higher phenotypic values in the hybrid compared to the parental genotypes.

	Ler x *Ler*			*Ler* x No-0			No-0 x *Ler*			No-0 x No-0		
Trait	Mean	SD	Class	Mean	SD	Class	Mean	SD	Class	Mean	SD	Class
Shoot Fresh Weight	3.81	0.29	A	4.40	0.91	A	5.21	0.50	A	5.61	1.08	A
Shoot Dry Weight	0.53	0.01	A	0.67	0.12	AB	0.76	0.06	AB	0.82	0.15	B
Shoot Water Content	0.86	0.00	B	0.85	0.01	A	0.85	0.00	AB	0.85	0.00	AB
Root Fresh Weight	3.17	0.21	A	4.19	1.09	A	5.30	1.00	A	4.93	1.54	A
Root Dry Weight	0.27	0.04	A	0.33	0.04	A	0.39	0.06	A	0.39	0.11	A
Root Water Content	0.91	0.01	A	0.92	0.01	A	0.92	0.01	A	0.92	0.00	A
Root-to-Shoot Fresh Weight Ratio	0.82	0.14	A	0.94	0.07	A	1.01	0.19	A	0.87	0.12	A
Root-to-Shoot Dry Weight Ratio	0.50	0.02	A	0.50	0.03	A	0.52	0.05	A	0.47	0.05	A
Shoot Phosphate Content (per dry weight gram)	14.60	0.82	AB	14.23	0.59	A	15.66	1.24	AB	16.75	0.50	B
Root Phosphate Content (per dry weight gram)	15.36	3.40	A	15.42	0.97	A	14.88	1.42	A	15.87	0.59	A
	Ler x *Ler*			*Ler* x Eri-1			Eri-1 x *Ler*			Eri-1 x Eri-1		
Trait	Mean	SD	Class	Mean	SD	Class	Mean	SD	Class	Mean	SD	Class
Shoot Fresh Weight	3.81	0.29	A	4.79	0.11	A	4.53	0.61	A	4.56	0.32	A
Shoot Dry Weight	0.53	0.01	A	0.62	0.02	AB	0.63	0.07	AB	0.65	0.03	B
Shoot Water Content	0.86	0.00	AB	0.87	0.00	B	0.86	0.01	AB	0.86	0.00	A
Root Fresh Weight	3.17	0.21	A	4.32	0.16	AB	4.72	0.35	B	4.59	0.70	B
Root Dry Weight	0.27	0.04	A	0.35	0.03	AB	0.40	0.03	B	0.41	0.02	B
Root Water Content	0.91	0.01	A	0.92	0.01	A	0.91	0.00	A	0.91	0.01	A
Root-to-Shoot Fresh Weight Ratio	0.82	0.14	A	0.90	0.02	A	1.05	0.07	A	0.99	0.10	A
Root-to-Shoot Dry Weight Ratio	0.50	0.02	A	0.56	0.05	AB	0.64	0.06	B	0.63	0.04	AB
Shoot Phosphate Content (per dry weight gram)	14.60	0.82	A	16.00	0.08	AB	17.78	0.03	AB	18.29	2.34	B
Root Phosphate Content (per dry weight gram)	15.36	3.40	A	17.73	2.36	A	15.71	0.97	A	16.96	0.70	A
	Ler x *Ler*			*Ler* x Ll-0			Ll-0 x *Ler*			Ll-0 x Ll-0		
Trait	Mean	SD	Class	Mean	SD	Class	Mean	SD	Class	Mean	SD	Class
Shoot Fresh Weight	3.81	0.29	A	4.16	0.31	AB	4.74	0.32	B	3.64	0.06	A
Shoot Dry Weight	0.53	0.01	A	0.63	0.00	B	0.72	0.03	B	0.50	0.06	A
Shoot Water Content	0.86	0.00	A	0.85	0.01	A	0.85	0.01	A	0.86	0.01	A
Root Fresh Weight	3.17	0.21	A	4.21	0.14	B	5.18	0.61	C	2.79	0.07	A
Root Dry Weight	0.27	0.04	A	0.36	0.02	B	0.48	0.00	C	0.26	0.02	A
Root Water Content	0.91	0.01	A	0.91	0.01	A	0.91	0.01	A	0.91	0.01	A
Root-to-Shoot Fresh Weight Ratio	0.82	0.14	AB	1.01	0.06	AB	1.09	0.14	B	0.76	0.02	A
Root-to-Shoot Dry Weight Ratio	0.50	0.02	A	0.57	0.08	AB	0.67	0.03	B	0.52	0.03	A
Shoot Phosphate Content (per dry weight gram)	14.60	0.82	A	16.34	1.06	AB	17.89	1.78	AB	18.59	1.60	B
Root Phosphate Content (per dry weight gram)	15.36	3.40	A	15.38	1.01	A	12.77	1.72	A	14.44	1.66	A

4.1.2 – TRANSCRIPTOMIC AND PARALLEL PHENOTYPIC ANALYSIS OF THE Pi STARVATION RESPONSE IN CROSSES BETWEEN L*ER* AND LL-0

To perform RNA-seq analysis of the Pi starvation response of L*er* x L*er*, Ll-0 x Ll-0, and the two reciprocal F_1 hybrids, L*er* x Ll-0 and Ll-0 x L*er*, a new set of crosses was generated as in the previous experiment, except that plants were subjected to a vernalization treatment prior to pollination (see **section 3.1.3.2**). Total RNA was obtained from roots and shoots, separately, from plants grown for 12 days, under both low (5µM) and high (1mM) Pi conditions. Parallel phenotypic analysis including additional traits were also performed with material collected from the same plates as those used for RNA isolation (see **section 3.3**).

Table R2. **Phenotypic response to phosphate starvation in the parental crosses derived from L*er* and Ll-0**. All traits were analysed under low Pi (5µM Pi) and high Pi (1mM) conditions. The mean values obtained from the three replicates and the standard deviation are shown for each genotype. Analysis of variance (ANOVA) was performed to compare between genotypes for all trait values, assuming significance at the 0.05 probability level. The phenotypic classes assigned are shown in separate columns (class). Intermediate categories are assigned to genotypes that share similarity with two different phenotypic classes. The rows highlighted in yellow represent cases of higher phenotypic values in the hybrids compared to the parental genotypes.

		Ler x *Ler*			*Ler* x Ll-0			Ll-0 x *Ler*			Ll-0 x Ll-0		
Trait	Pi regimen	Mean	SD	Class	Mean	SD	Class	Mean	SD	Class	Mean	SD	Class
Shoot Fresh Weight	5µM(-P)	4.02	0.26	AB	4.36	0.26	B	4.68	0.28	B	3.54	0.32	A
Shoot Dry Weight	5µM(-P)	0.59	0.02	B	0.61	0.01	B	0.69	0.01	C	0.50	0.01	A
Shoot Water Content	5µM(-P)	0.85	0.01	A	0.86	0.00	A	0.85	0.00	A	0.86	0.01	A
Root Fresh Weight	5µM(-P)	3.33	0.25	A	4.10	0.17	B	4.60	0.35	B	2.70	0.17	A
Root Dry Weight	5µM(-P)	0.35	0.01	B	0.38	0.01	BC	0.40	0.02	C	0.24	0.01	A
Root Water Content	5µM(-P)	0.90	0.00	A	0.91	0.01	AB	0.91	0.00	B	0.91	0.01	B
Root-to-Shoot Fresh Weight Ratio	5µM(-P)	0.83	0.03	AB	0.94	0.04	AB	0.98	0.06	B	0.76	0.04	A
Root-to-Shoot Dry Weight Ratio	5µM(-P)	0.59	0.03	B	0.62	0.01	B	0.59	0.03	B	0.49	0.01	A
Shoot Phosphate Content (per dry weight	5µM(-P)	19.32	0.82	A	19.39	1.02	A	19.89	1.80	A	21.20	0.40	A
Root Phosphate Content (per dry weight	5µM(-P)	17.48	2.43	A	20.22	1.30	A	18.26	0.79	A	20.31	3.44	A
Anthocyanin Content (per fresh weight gram)	5µM(-P)	31.33	4.12	B	27.10	0.20	B	28.74	1.79	B	16.45	2.46	A
Primary Root Length at 7 DAS	5µM(-P)	39.69	0.42	B	39.13	0.22	B	41.76	1.55	B	24.71	1.08	A
Primary Root Length at 9 DAS	5µM(-P)	61.44	1.34	BC	59.78	1.16	B	64.49	1.57	C	40.99	1.18	A
Primary Root Length at 12 DAS	5µM(-P)	96.52	2.07	B	94.11	4.27	B	98.74	2.47	B	67.25	2.36	A
Shoot Fresh Weight	1mM(+P)	23.53	0.76	B	22.69	1.45	B	24.55	2.96	B	8.90	1.02	A
Shoot Dry Weight	1mM(+P)	1.71	0.09	B	1.63	0.08	B	1.75	0.19	B	0.76	0.06	A
Shoot Water Content	1mM(+P)	0.93	0.00	B	0.93	0.00	B	0.93	0.00	B	0.91	0.00	A
Root Fresh Weight	1mM(+P)	7.51	1.28	B	8.41	1.16	B	9.66	0.53	B	3.76	0.45	A
Root Dry Weight	1mM(+P)	0.71	0.12	B	0.74	0.08	B	0.80	0.07	B	0.34	0.02	A
Root Water Content	1mM(+P)	0.90	0.00	A	0.91	0.00	AB	0.92	0.01	B	0.91	0.00	AB
Root-to-Shoot Fresh Weight Ratio	1mM(+P)	0.32	0.05	A	0.37	0.02	AB	0.39	0.03	AB	0.41	0.01	B
Root-to-Shoot Dry Weight Ratio	1mM(+P)	0.41	0.07	A	0.46	0.02	A	0.45	0.04	A	0.44	0.01	A
Shoot Phosphate Content (per dry weight	1mM(+P)	167.2	5.66	B	156.6	6.73	B	166.6	4.38	B	136.3	5.18	A
Root Phosphate Content (per dry weight	1mM(+P)	65.34	1.53	A	67.26	1.82	A	69.54	5.59	A	63.18	7.35	A
Primary Root Length at 7 DAS	1mM(+P)	40.58	0.67	B	42.26	0.62	BC	45.48	0.88	C	28.88	1.97	A
Primary Root Length at 9 DAS	1mM(+P)	66.19	0.97	B	70.54	1.00	BC	74.27	2.04	C	50.71	2.82	A

In general, phenotypic results were consistent with those obtained in the previous experiment, since the different genotypes showed similar relative behaviour in both experiments for most traits examined. Overall, Ll-0 x Ll-0 plants displayed the most extreme values for most of the traits analysed, both under low and high phosphate conditions, whereas L*er* x L*er* and the F_1 hybrids exhibited fewer differences. Nevertheless, when comparing the F_1 hybrids to both parental genotypes, significant differences were found for shoot and root weight of plants grown under Pi limitation, and for primary root length at 7 and 9 DAS for plants grown under high Pi (**Table R2**).

As for the transcriptomic analysis, we estimated gene expression differences using both total counts per gene and allele-specific read counts, with three replicates per sample, between low and high phosphate conditions, between the two parental lines, between reciprocal hybrids, between hybrid and parental genotypes, and between the two different alleles of a hybrid. Regarding ASE, we accounted for a total of 67,541 SNPs across all samples (i.e., present in at least one of the samples analysed). Of them, a set of 48,840 SNPs could be unequivocally attributed to either L*er* or Ll-0 haplotypes. ASE was estimated from the 48,286 exonic SNPs, representing 13,842 genes.

4.1.3 – PHOSPHATE STARVATION EFFECT ON HYBRID AND PARENT GENE EXPRESSION VARIATION

To study the transcriptional response to Pi starvation in parental genotypes and in F_1 hybrids, we used total read counts per gene to analyse differential gene expression between low (5µM) and high Pi (1mM) conditions. A total of 6,410 (p-adj<0.05; FC ±2) PS-responsive genes were estimated as differentially expressed in at least one of the pairs of samples compared. In general, shoots displayed a higher number of DEGs genes than roots. While in roots the number of DEGs was fair similar for all genotypes, in shoots the parental genotype Ll-0 x Ll-0 exhibited the lowest number of PS-responsive genes (**Figure R1**).

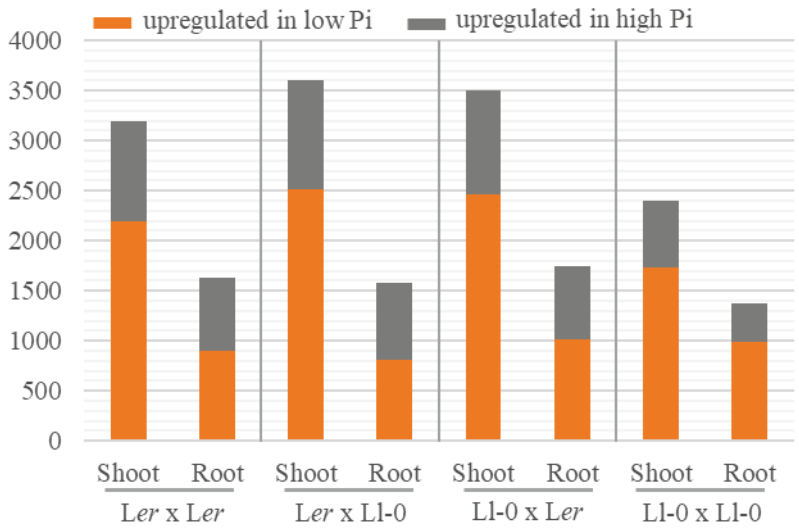
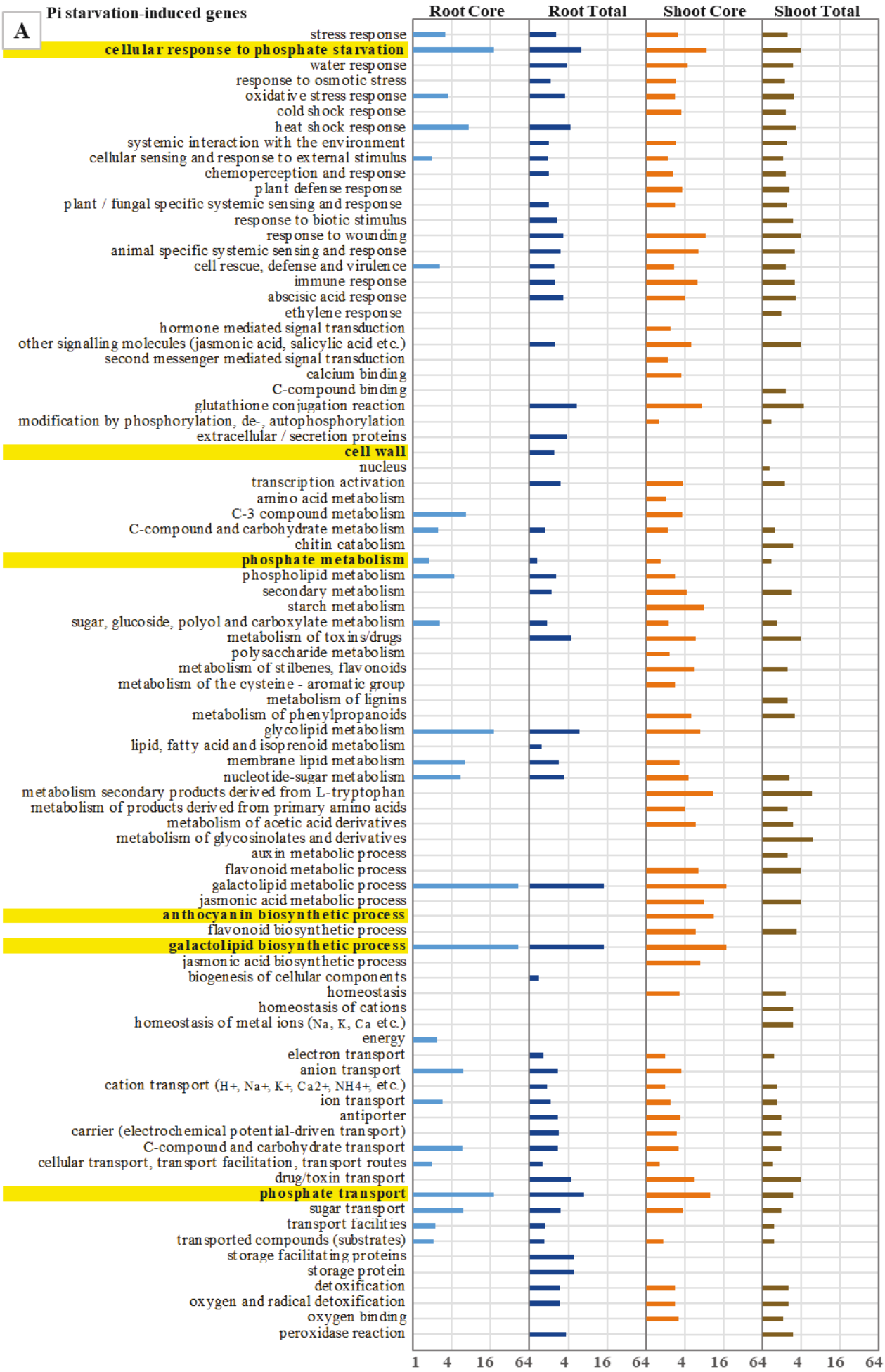


Figure R1. **Number of genes differentially expressed between low and high phosphate conditions**. Vertical bars indicate the fraction of genes significantly upregulated under low (in orange) and high (in black) Pi conditions. Differential expression was determined by applying stringent FDR threshold (p-adj<0.05) and fold change ±2.

We compared the DEGs detected in the four genotypes analysed and generated two types of gene sets. The core set included the genes either induced or repressed in all genotypes, whereas the total set comprised genes induced or repressed in at least one of them. Core and total sets of PS-induced and PS-repressed genes were independently obtained for shoot and root. After that, we performed gene ontology (GO) term enrichment analysis on them. A higher number of GO terms with significant enrichment was found in the sets of PSI genes compared to those of PS-repressed genes, and in the shoot core and total gene sets over the root core set. This is likely to be caused at least in part by the lower number of PS-repressed genes in the core set. Many of the biological categories enriched in the induced gene set are tightly related to PSRs (e.g., Pi transport and signaling, hormone response, cell wall remodeling). Enrichment in these PS-related ontology terms was higher in the core sets of PSI genes (**Figure R2**).


A
Pi starvation-induced genes
Root Core
Root Total
Shoot Core
Shoot Total
stress response
cellular response to phosphate starvation
water response
response to osmotic stress
oxidative stress response
cold shock response
heat shock response
systemic interaction with the environment
cellular sensing and response to external stimulus
chemoperception and response
plant defense response
plant / fungal specific systemic sensing and response
response to biotic stimulus
response to wounding
animal specific systemic sensing and response
cell rescue, defense and virulence
immune response
abscisic acid response
ethylene response
hormone mediated signal transduction
other signalling molecules (jasmonic acid, salicylic acid etc.)
second messenger mediated signal transduction
calcium binding
C-compound binding
glutathione conjugation reaction
modification by phosphorylation, de-, autophosphorylation
extracellular / secretion proteins
cell wall
nucleus
transcription activation
amino acid metabolism
C-3 compound metabolism
C-compound and carbohydrate metabolism
chitin catabolism
phosphate metabolism
phospholipid metabolism
secondary metabolism
starch metabolism
sugar, glucoside, polyol and carboxylate metabolism
metabolism of toxins/drugs
polysaccharide metabolism
metabolism of stilbenes, flavonoids
metabolism of the cysteine - aromatic group
metabolism of lignins
metabolism of phenylpropanoids
glycolipid metabolism
lipid, fatty acid and isoprenoid metabolism
membrane lipid metabolism
nucleotide-sugar metabolism
metabolism secondary products derived from L-tryptophan
metabolism of products derived from primary amino acids
metabolism of acetic acid derivatives
metabolism of glycosinolates and derivatives
auxin metabolic process
flavonoid metabolic process
galactolipid metabolic process
jasmonic acid metabolic process
anthocyanin biosynthetic process
flavonoid biosynthetic process
galactolipid biosynthetic process
jasmonic acid biosynthetic process
biogenesis of cellular components
homeostasis
homeostasis of cations
homeostasis of metal ions (Na, K, Ca etc.)
energy
electron transport
anion transport
cation transport (H+, Na+, K+, Ca2+, NH4+, etc.)
ion transport
antiporter
carrier (electrochemical potential-driven transport)
C-compound and carbohydrate transport
cellular transport, transport facilitation, transport routes
drug/toxin transport
phosphate transport
sugar transport
transport facilities
transported compounds (substrates)
storage facilitating proteins
storage protein
detoxification
oxygen and radical detoxification
oxygen binding
peroxidase reaction
1 4 16 64 4 16 64 4 16 64 4 16 64

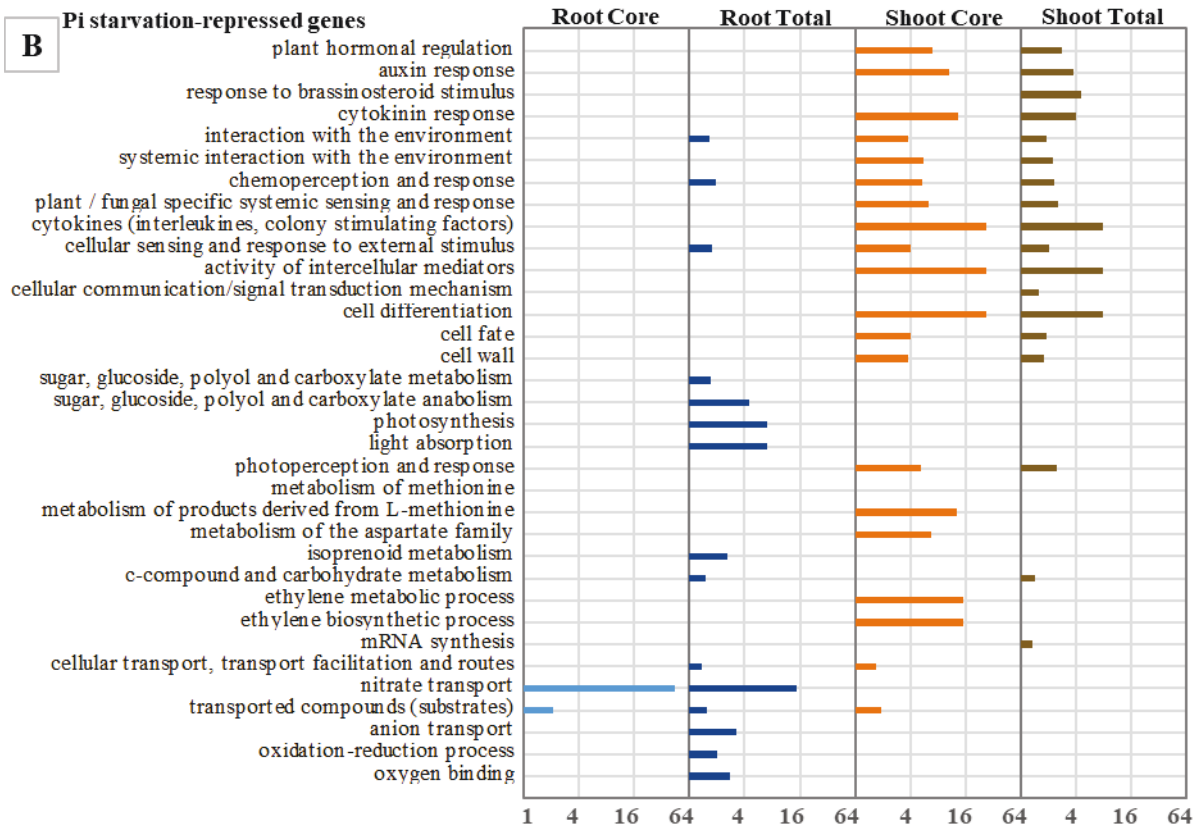


Figure R2. **Gene ontology (GO) term enrichment of Pi starvation (PS)-responsive gene sets**. GO terms with FDR lower than 0.01 were selected and the fold enrichment between observed and expected frequencies is represented for the root core (in blue), the root total (in dark blue), the shoot core (in orange) and the shoot total (in brown) sets of PS-induced **(A)** and PS-repressed **(B)** genes. The chart's axes display a logarithmic scale to suit the visualization of wide ranges of data. Some PSR-related categories are highlighted in yellow.

4.1.4 – TRANSCRIPTOMIC DIFFERENCES BETWEEN HYBRIDS AND PARENTS

We analysed differential gene expression between the parental genotypes, between the reciprocal hybrids and between the parental and the hybrid genotypes using total read counts per gene. The number of DEGs was estimated separately for samples of shoots, roots, high and low Pi. The highest differences were found between the two parental genotypes, L*er* x L*er* and Ll-0 x Ll-0, whereas no significant differences were detected between the two reciprocal crosses, L*er* x Ll-0 and Ll-0 x L*er* (**Figure R3-A**). An intermediate number of DEGs was found when comparing hybrid to parental samples (700-2,800), being the hybrids more similar to L*er* than Ll-0 (**Figure R3-B**). These results indicate that gene expression in hybrids usually kept within the parental range, although closer to L*er*. Nevertheless, although in much lesser number, there were clear cases of genes whose expression in hybrids was higher or smaller than in any of the two parents (**Figure R3-C**), indicating that overdominance also has an impact on gene expression in hybrids.

When comparing gene expression in the parents and in either of the hybrids, a higher number of DEGs was found between hybrids and Ll-0 x Ll-0 than between hybrids and L*er* x L*er*. This indicates that hybrid gene expression was more similar to L*er* than to Ll-0. In general, we estimated more DEGs in the absence of Pi starvation stress, except for the comparison L*er* x L*er* vs Ll-0 x L*er* in root.

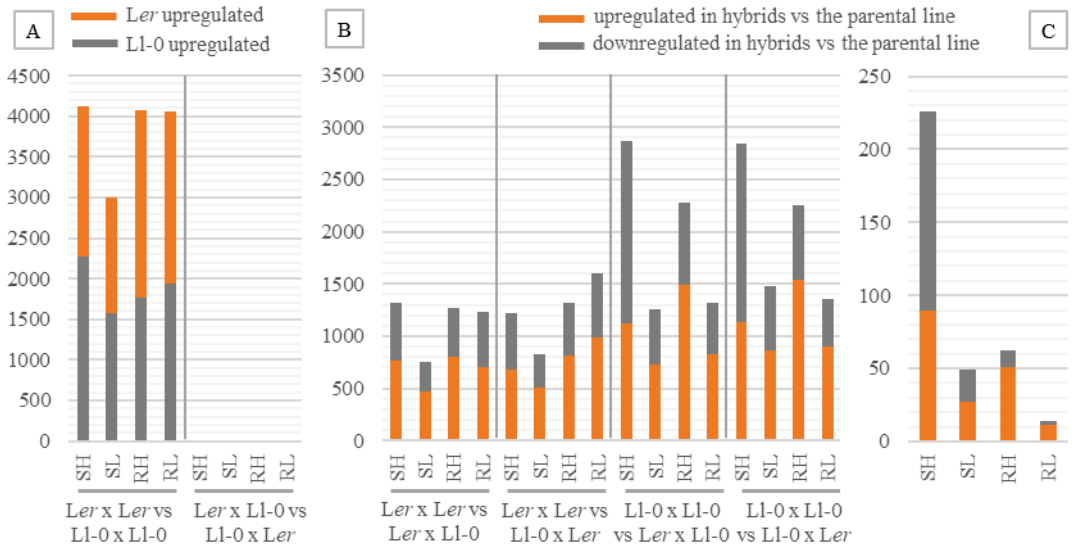


Figure R3. **Differential gene expression between parents, between hybrids and between hybrids and parents**. Vertical bars indicate **(A)** the fraction of genes differentially expressed between the parental lines L*er* x L*er* and in Ll-0 x Ll-0; **(B)** the number of genes differentially expressed between hybrids and parental lines; **(C)** the number of genes differentially expressed in hybrids when compared to both L*er* and Ll-0. RH = Root under High Pi conditions; RL = Root under Low Pi conditions; SH = Shoot under High Pi conditions; SL = Shoot under Low Pi conditions. Differential expression was determined by applying stringent FDR threshold (p-adj<0.05) and fold change ±1.5.

4.1.5 – ESTIMATION OF *CIS* AND *TRANS*-REGULATORY EFFECTS, BASED ON THE ANALYSIS OF ALLELE-SPECIFIC EXPRESSION IN HYBRIDS AND PARENTS

Comparative analysis of ASE within F_1 hybrids and between parents and F_1 hybrids allow to estimate the contribution of *cis*- and *trans*-regulatory effects to underlying variation in gene expression. For instance, the differential expression of the two alleles of a given gene in the F_1 hybrids allows to detect *cis*-regulatory effects, given that the two alleles are in the same genetic background and thus expression differences cannot be attributed to other cause than *cis* effects. Likewise, the comparison of the expression of a given allele between parents and F_1 hybrids will provide information about other genetic effects, primarily *trans*-acting regulatory factors.

4.1.5.1 – *Trans*-regulatory effects.

To examine *trans*-regulatory effects, we compared specific counts for the L*er* allele in the F_1 hybrids samples to the L*er* counts in the corresponding L*er* x L*er* samples of the same organ and condition. The Ll-0 allele was subjected to a similar analysis. Additionally, we compared the expression of each allele between the two reciprocal hybrids. Our results show that expression of the L*er* alleles in F_1 hybrids is quite similar to that of the L*er* x L*er*, indicating that the transcriptional environment of this parent is quite similar to that of the hybrids, in line with the results based on total counts per gene. In contrast, the Ll-0 alleles display higher expression differences in the hybrids over the Ll-0 x Ll-0 parents, revealing a higher number of *trans*-regulatory differences in this comparison. Again, as observed in the previous analysis based on total counts per gene, higher differences in ASE are found when the comparisons are done between parents and hybrids grown under high Pi conditions. Also, in line with the previous analysis, no significant differences were observed between the two reciprocal hybrids in the expression of any allele examined. Of note is that while the two hybrids show a very similar transcriptional profile, the transcriptional profile of parents were slightly more similar to that of the hybrid with the same maternal constitution (**Figure R4**), indicative of contribution of maternal effects/genomic imprinting in shaping gene expression.

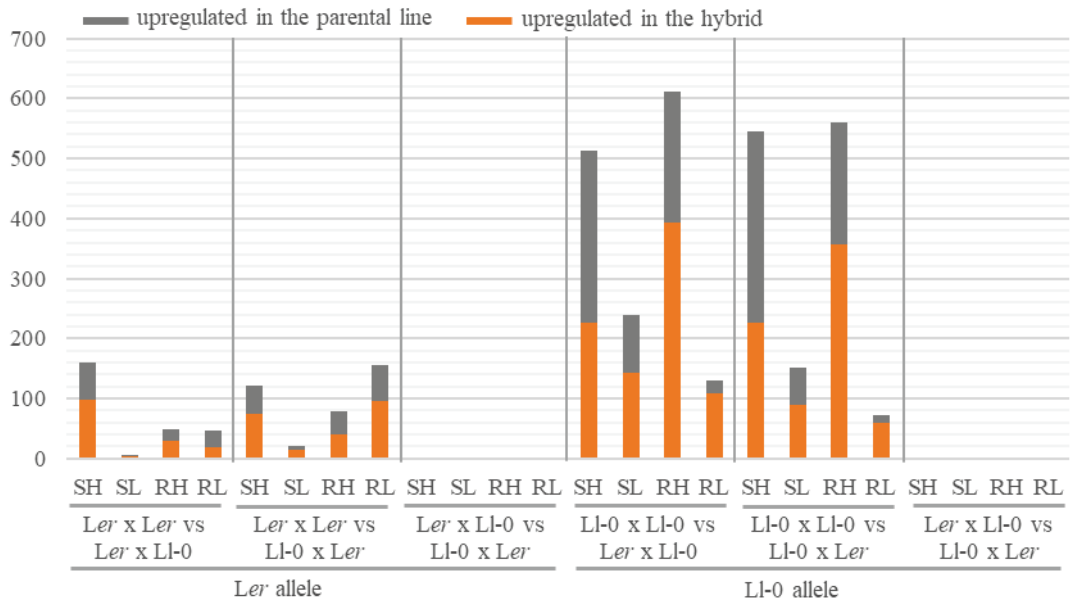


Figure R4. **Number of genes displaying differences between F_1 hybrids and parents in the expression of shared alleles**. Vertical bars indicate the fraction of genes with allelic variants upregulated in the hybrid (in orange) and in the parent (in black). RH = Root under High Pi conditions; RL = Root under Low Pi conditions; SH = Shoot under High Pi conditions; SL = Shoot under Low Pi conditions. Differential expression was determined by applying stringent FDR (p-adj< 0.05) and fold change ±1.5.

4.1.5.2 – *Cis*-regulatory effects.

The comparison between the alleles L*er* and Ll-0 in F_1 hybrids revealed a total of 3,501 genes (p-adj <0.05; FC ±1.5) exhibiting expression differences between their alleles, considering all organs and conditions, which manifests the high extent of *cis*-regulatory variation. *Cis* effects are similar in number for high and low Pi growth conditions, suggesting that they are less dependent on Pi growth conditions than *trans* effects. Also evident from this analysis is that the number of genes with allelic imbalance was higher in root than shoot samples (**Figure R5**).

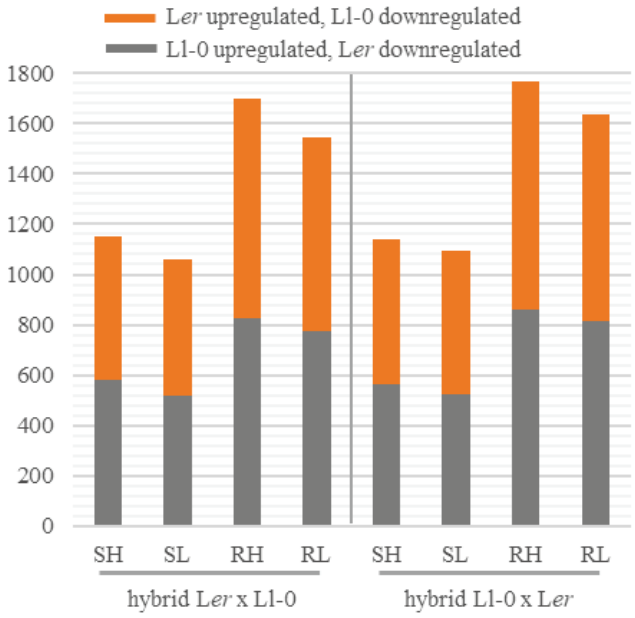


Figure R5. **Number of genes with significant expression differences between alleles in F_1 hybrids**. Vertical bars represent the fraction of differentially expressed genes (DEGs) whose expression is upregulated in the haplotypes Ll-0 (in black) and L*er* (in orange), separately for the hybrids L*er* x Ll-0 and Ll-0 x L*er*. DEGs were filtered by stringent FDR (p-adj<0.05) and fold change ±1.5. RH = Root, High Pi conditions; RL = Root, Low Pi conditions; SH = Shoot, High Pi conditions; SL = Shoot, Low Pi conditions.

4.1.5.3 – Read biases in the RNA-seq mapping procedure.

As previously mentioned (see section **3.5.3.2**), RNA-seq libraries were mapped to the reference genome Col-0 instead of using L*er* or Ll-0. We wondered whether this could bias the number of reads obtained towards the SNPs equal to that of the reference Col-0 genome, influencing the subsequent analysis. To evaluate the extent to which our analysis was biased, we determined the direction of the *cis* effect on allele expression (e.g., L*er* > Ll-0 or Ll-0 > L*er*) and of *trans* effects separately for the cases in which the reference allele was equal to the L*er* allele and the ones in which Ll-0 had the reference haplotype. We found that when the total number of DEGs were accounted for, the proportion of L*er* alleles upregulated and downregulated vs Ll-0 was very similar. Nevertheless, for *cis* effects where the L*er*

allelic counts are higher than those of the Ll-0 allele, the proportion of L*er* alleles equal to the Col-0 reference was significantly higher than L*er* alleles different from the Col-0 reference (almost 4:1). The same observation was made regarding the Ll-0 allele (**Figure R6-A**). Therefore, there was an overestimation of *cis* effects for upregulated alleles equal to the reference allele. On the contrary, when *trans* effects were considered (the same allele is compared between hybrid and parent), we observed a balanced proportion of SNPs equal and unequal to the reference Col-0 (**Figure R6-B**). This indicates that, in contrast to case of *cis* effects, *trans* effects appeared not to suffer from problems of overestimation regarding the reference allele.

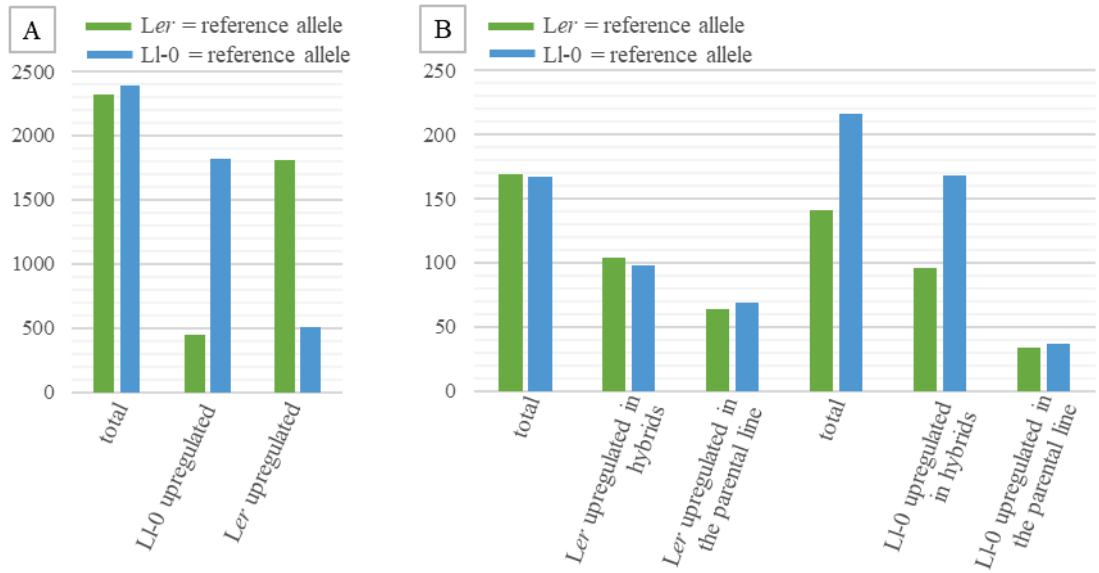


Figure R6. **Read count bias for reference over non-reference allele**. Vertical bars represent the number of SNPs in the reference genome that are equal to L*er* (in green) and Ll-0 (in blue) in samples of roots grown under low phosphate conditions. The graph in the left **(A)** represents the differential expression between the alleles L*er* and Ll-0 in samples of hybrids. The pairs of bars labelled as 'total' show the total number of SNPs differentially expressed, whereas the second and third pairs of bars represent down- and up-regulation of the allele L*er*. Similarly, the graph in the right **(B)** represents differential expression of SNPs between samples of hybrids and the parental genotypes, separately for the parents L*er* and Ll-0. Data shown were filtered by FDR < 0.05 and fold change ±1.5.

To further examine the consistency of our ASE dataset, we compared the Pi-responsive genes that we could detect using ASE dataset to the ones estimated using the dataset of total read counts per gene. We analysed the sum of the L*er* and Ll-0 allelic counts in samples of hybrids and 1,352 Pi-responsive genes were estimated in roots and 2,252 in shoots ($q < 0.05$; FC > ±2x). As for the total counts, 3,827 and 6,552 Pi-responsive genes were estimated in roots and shoots, respectively ($q < 0.05$; FC > ±1.5x). We then analysed the percentage of coincidences between the genes detected using both datasets, separately for shoot and root. It was found that the proportion of coincidences between the Pi starvation responsive

genes identified based on total and allele specific counts exceeded the 70% in all samples analysed. As expected, we observed a similar percentage of coincidences when the reference allele was equal or not to L*er* (**Figure R7**). These results show a high correlation between both datasets, indicating that although comparisons of the expression between two different alleles of the same gene using ASE could be biased, leading to overestimation of *cis* effects on upregulated alleles equal to the reference allele, such bias did not occur in comparisons between the expression of given allele in two different growth conditions, organs and/or genotypes using ASE.

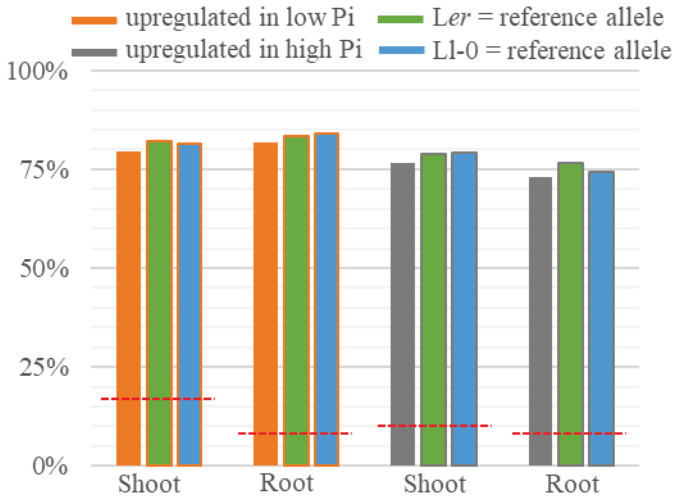


Figure R7. **Percentage of coincidences between Pi starvation-responsive genes detected using data from allele-specific expression and total read counts per gene**. Vertical bars indicate the percentage of Pi starvation (PS)-induced (in orange) and PS-repressed (in black) genes detected using the datasets of allele-specific expression and total read counts per gene with respect to the number detected in the first one. Green and blue bars show the percentage of coincidence when the reference allele is equal to L*er* and Ll-0, respectively. The red horizontal lines indicate the percentage of expected coincidences in case of random distributions.

4.1.5.4 – Analysis of *cis*- and *trans*-regulatory effect specificity.

To analyse the specificity of the *cis*- and *trans*-regulatory effects, we compared these effects between different plant parts (shoot and roots) and between different Pi growth conditions. As we expected given the above-mentioned bias in *cis* effects, we detected more coincidences in the comparisons made for alleles of the reference haplotype (**Figure R8-A**). Despite this bias, coincidences in *cis* effects were clearly higher between the same plant part at two different Pi growth conditions than between two different plant parts at the same Pi growth condition. In addition, there was a higher percentage of coincidences in *cis* effects when comparisons between roots and shoots were made for the low Pi growth condition. Regarding the *trans*-regulated genes, the picture was less clear, in part due to the low number of *trans* effects detected in some conditions, although the higher percentage of coincidences in *trans* effects between roots at two Pi growth conditions was evident (**Figure R8-B**).

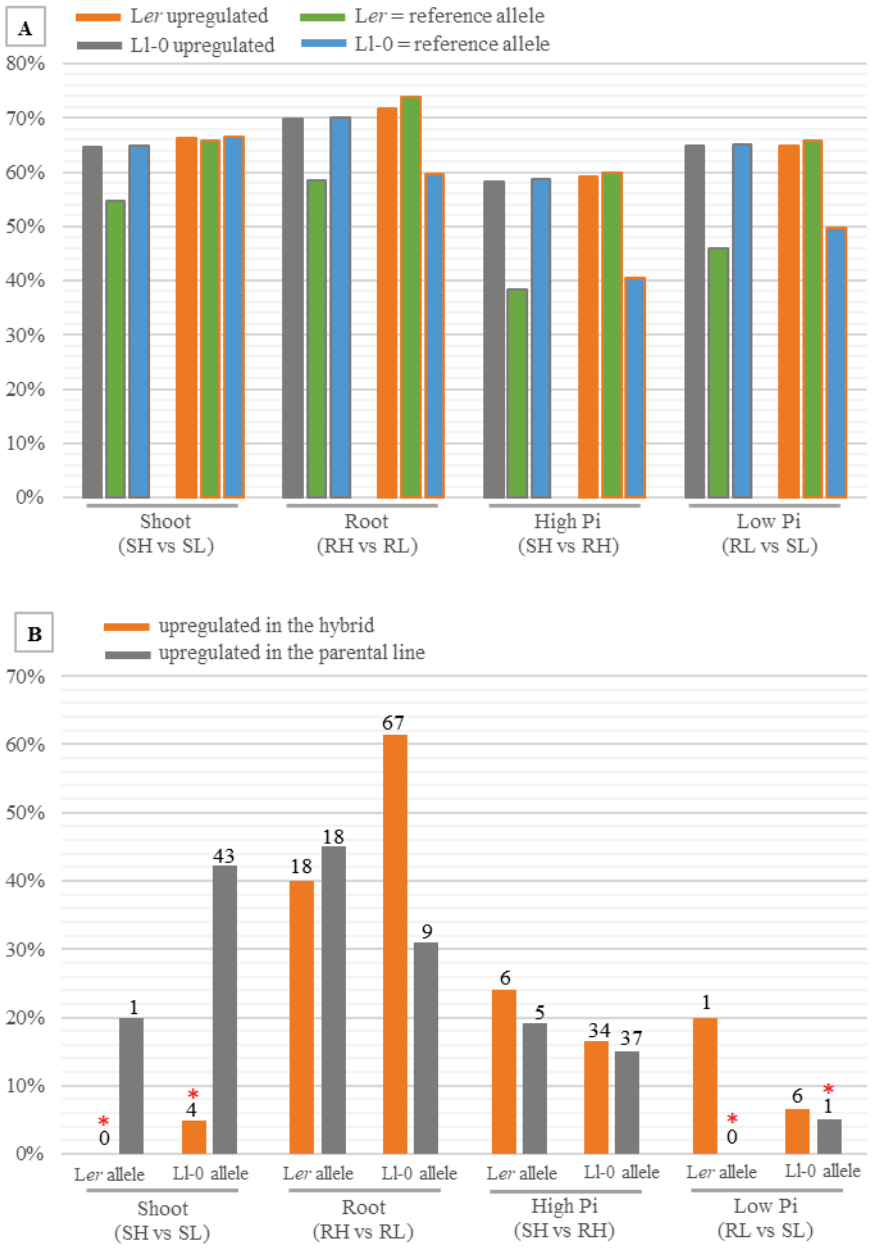


Figure R8. **Specificity of *cis*- and *trans*-regulatory effects**. Vertical bars show the percentages of observed coincidences in the *cis*- **(A)** and *trans*-regulated genes **(B)** between samples of shoot (SH vs SL), between samples of root (RH vs RL), between samples of high Pi (SH vs RH), and between samples of low Pi (RL vs SL). These percentages are referred to the list with the lower number of genes. In the top graph **(A)**, vertical bars outlined in orange and black represent upregulation of L*er* and Ll-0 alleles, respectively, when the reference allele is equal to L*er* (green bars), equal to Ll-0 (blue bars) or without filtering by the haplotype of the allele (orange and black). In the bottom graph **(B)**, it is shown the percentage of coincidences in *trans*-regulated alleles upregulated in the hybrid (orange bars) and in the parent (black bars), as well as the number of observed coincidences. Red asterisks denote lack of statistical significance. Data shown were filtered by FDR < 0.05 and fold change ±1.5. RH = Root under High Pi conditions; RL = Root under Low Pi conditions; SH = Shoot under High Pi conditions; SL = Shoot under Low Pi conditions.

4.2 – QTL ANALYSIS

To perform a QTL analysis that would allow to get insights on their mode of inheritance (dominant, overdominant and codominant), we decided to analyse the transcriptome of hybrids generated from crosses between lines of the L*er* x Ll-0 RIL collection and their progenitors L*er* and Ll-0. In total we analysed the root transcriptome under low Pi conditions of 63 different hybrid genotypes, half corresponding to L*er* x RIL hybrids and half to Ll-0 x RIL hybrids.

4.2.1 – GENETIC MAP

Although the RILs from the L*er* x Ll-0 RIL population had been previously genotyped for 95 markers and both physical and genetic maps were available (Sánchez-Bermejo *et al.*, 2012), we used the ASE dataset to improve the accuracy of the physical map and construct a new genetic map (see **section 3.6.2**). A physical map with 111 markers equally distributed genome-wide and an average physical distance between them close to 1 Mb was obtained and used to develop a genetic map. Once constructed, the genetic map showed an overall length of 430 cM and the genetic markers were separated by an average interval of 4 cM, translating into an even distribution throughout the genome (**Figure R9**). Additionally, we observed that the order of the markers was identical for the genetic and the physical maps and that the recombination rate was homogenous along the chromosomes.

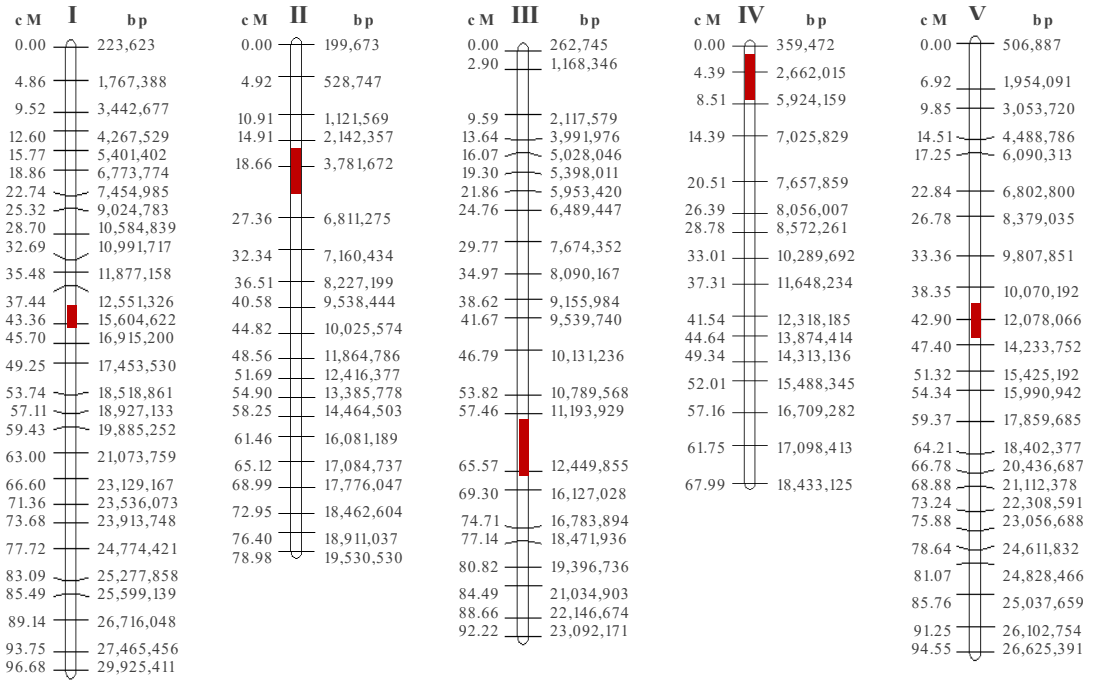


Figure R9. **Genetic linkage map**. Genetic positions (in cM) and physical positions (in bp) are provided for each marker. Red boxes indicate the approximate centromere positions on each chromosome.

4.2.2 – eQTL ANALYSIS

4.2.2.1 – eQTL analysis of the L*er* x RIL and Ll-0 x RIL populations.

A total of 19,017 genes surpassing an expression threshold of 20 counts were analysed in the L*er* x RIL and Ll-0 x RIL populations (see **section 3.6.3.1**). We performed eQTL analysis using the software package QTL cartographer (Wang, Basten and Zeng, 2012). Following a permutation test on expression data for 500 genes (see **section 3.6.3.2**), we estimated global LOD score threshold values of 2.75 and 2.77 for the L*er* x RIL and Ll-0 x RIL populations, respectively. To simplify, we used LOD score ≥2.8 in both populations as a cutoff for eQTL significance. In addition to this LOD score cutoff, to select eQTLs we also included a cutoff of ±0.3 for the magnitude of additive effect. Positive and negative additive allelic effects mean respectively that gene expression in the homozygous state was higher and lower than in the heterozygous state, for both L*er* x RIL and Ll-0 x RIL populations. According to these criteria, 10,663 genes were detected to carry significant eQTLs in at least one of the two populations. The total number of eQTLs detected in the Ll-0 x RIL population was higher than that in the L*er* x RIL (**Table R3**).

Table R3. **Global description of the eQTLs**. The number of significant eQTLs as well as the genes with eQTLs found in the analysis of the L*er* x RIL and Ll-0 x RIL populations are shown in the table, differentiating between *cis*- and *trans*-eQTLs. We used LOD score ≥2.8 and additive effect ±0.3 as eQTL significance criteria.

	total	L*er* x RIL	Ll-0 x RIL	shared
nº eQTLs	21,725	10,543	11,867	686
nº genes with an eQTL	10,663	7,279	7,911	4,527
nº genes with a *cis* -eQTL	2,230	1,420	1,534	724
up-regulated (additive effect > 0.3)	1,212	598	614	0
down-regulated (additive effect < -0.3)	1,742	822	920	0
nº *trans* -eQTLs	19,296	9,123	10,333	160
up-regulated (additive effect > 0.3)	9,953	4,832	5,127	6
down-regulated (additive effect < -0.3)	9,491	4,291	5,206	6
nº genes with a *trans* -eQTL	9,886	6,379	6,991	3,484
up-regulated (additive effect > 0.3)	7,202	4,060	4,501	1,359
down-regulated (additive effect < -0.3)	6,657	3,649	4,313	1,305

We detected a total of 2,230 genes with a significant *cis*-eQTL, representing approximately a 12% of the total number of genes analysed. The proportion of *cis*-eQTLs was moderately uniform across the genome (**Figure R10**), although there was a bias in the direction of the *cis* effect, as in both populations we found more *cis*-eQTLs down- than upregulated. This indicates that the heterozygous state has a moderate tendency to increase gene expression (i.e., the sum of the expression of the alleles in the heterozygous state is higher than that in the homozygous state).

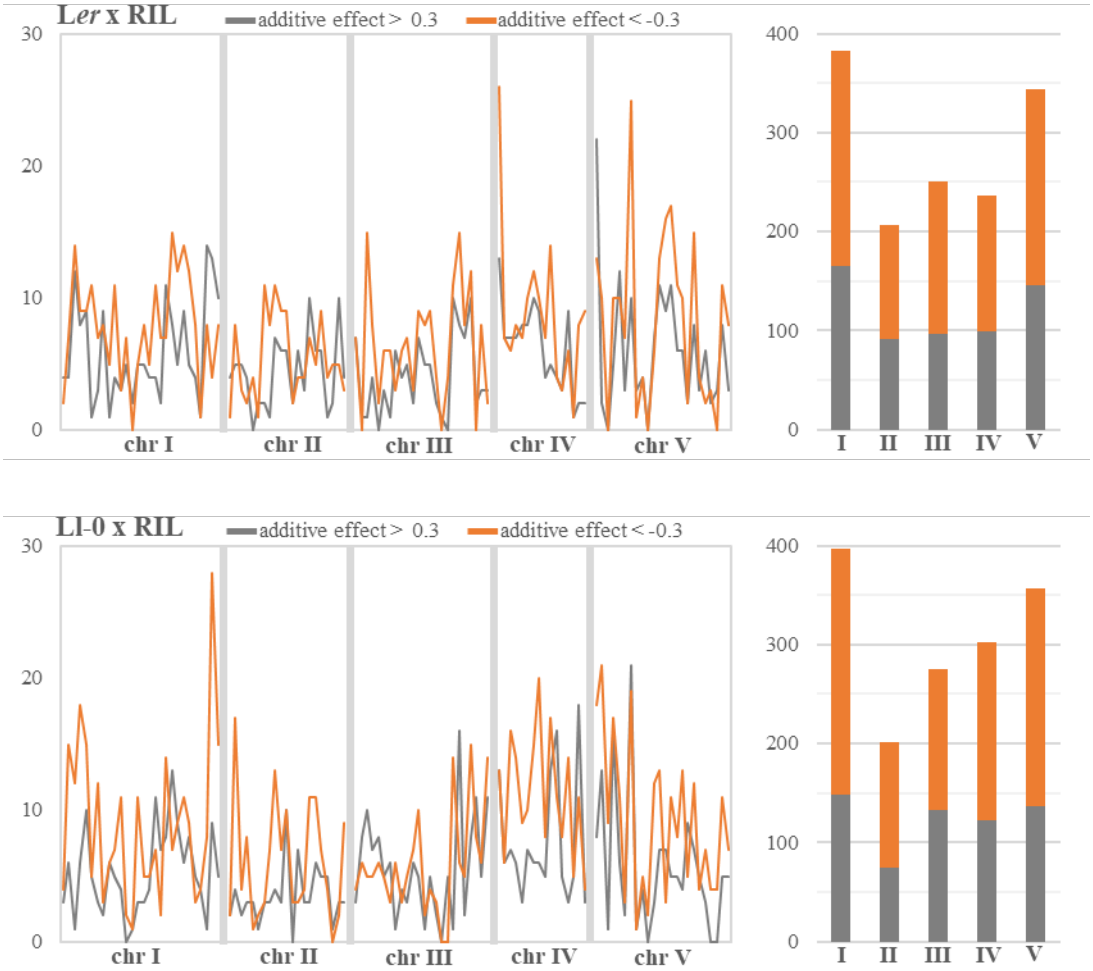


Figure R10. **Genomic distribution of *cis*-eQTLs**. The graphs show the number and distribution of the significant *cis*-eQTLs found in the L*er* x RIL and in the Ll-0 x RIL populations, separately for those upregulated in heterozygosis (in orange) and homozygosis (in black). In the linear graphs on the left, the number of *cis*-eQTLs (y-axis) is plotted against the marker position where they map (x-axis), whereas the bar graphs on the right represent the total number of *cis*-eQTLs per chromosome.

The remaining 19,296 eQTLs were classified as distant or *trans*-eQTLs. Approximately a 52% of the genes analysed carried a *trans*-eQTL. Not surprisingly, their distribution was non-uniform across the chromosomes (**Figure R11**), as many of them clustered at multiple hotspots locations across the genome associated to variation in the expression levels of hundreds to thousands of genes. Based on the total number of *trans*-eQTLs detected, we could determine a significance threshold for definition of *trans*-eQTLs hotspots (i.e., locations at which the number of assigned eQTLs significantly exceeded the number of expected assignations in a random distribution with FDR<0.01; see **section 3.6.3.2**). We found 48 *trans*-eQTL hotspots exceeding the threshold in at least one of the sets of up- and downregulated genes (**Table R4**). Among these, we selected the 15 major hotspots for subsequent characterization. Seven of them were detected in the L*er* x RIL population and eight in the Ll-0 x RIL, most of them located on the chromosomes 4 and 5.

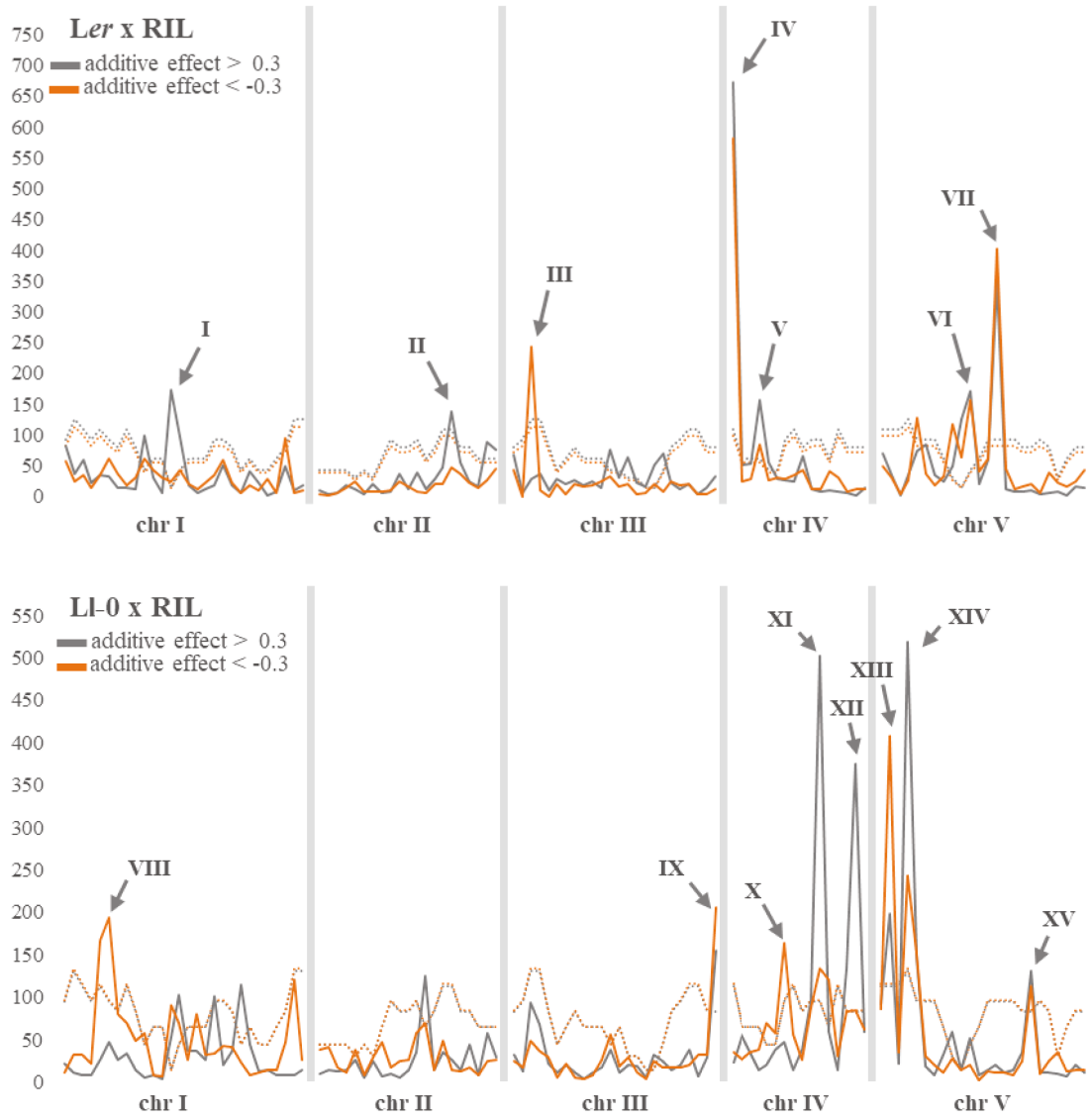


Figure R11. **Distribution of *trans*-eQTLs across the genome**. The number of significant *trans*-eQTLs peaks (y-axis) is plotted against the marker position where they localize (x-axis). Grey arrows indicate the location of the fifteen major hotspots detected in the L*er* x RIL and Ll-0 x RIL populations, separately for those upregulated in heterozygosis (in orange) and in homozygosis (in black). The dashed orange and black lines indicate the hotspot significance threshold based on the binomial probability ($p<0.05$) that the number of eQTLs assigned to a particular genome site could significantly surpass the number of expected assignations in case of random distributions.

4.2.3 – CHARACTERIZATION OF THE GENES ASSOCIATED TO *CIS*- AND *TRANS*-eQTLs

4.2.3.1 – Analysis of the explained phenotypic variance.

To estimate the magnitude of the effects produced by the *cis*- and *trans*-eQTLs detected in our study, we estimated the proportion of phenotypic variance explained (R^2) by individual eQTLs. Most of the detected eQTLs explained more than 20% of the variance in gene expression. On

average, *cis*-eQTLs accounted for larger effects than *trans*-eQTLs. Thus, for the *cis*-eQTLs detected in the L*er* x RIL and Ll-0 x RIL populations, 87.34% and 88.34% of them, respectively, explained more than 20% of expression variation, whereas 24.30% and 23.94% of them accounted for variances greater than 50%. In the case of *trans*-eQTLs, 68.96% and 70.36% explained phenotypic variances greater than 20% in the L*er* x RIL and Ll-0 x RIL populations, respectively, whereas only the 1.77% and 2.38% of the *trans*-eQTLs accounted for more than the 50% of the variance in gene expression (**Figure R12**).

Table R4. **Summary of the 48 significant *trans*-eQTL hotspots**. For each hotspot, the approximate genomic location, as well as the module size of total, up- and down-regulated *trans*-eQTLs are provided. The rows in yellow highlight the major fifteen hotspots.

				number of *trans*-regulated genes				
Hotspot	population analysed	chr.	nº marker	total	additive effect > 0.3	threshold	additive effect < -0.3	threshold
I	L*er* x RIL	1	13	304	174	15	24	14
II	L*er* x RIL	2	15	440	139	109	47	99
III	L*er* x RIL	3	3	347	28	126	244	114
IV	L*er* x RIL	4	1	2515	673	109	582	99
V	L*er* x RIL	4	4	538	157	62	85	56
VI	L*er* x RIL	5	11	640	171	43	157	39
VII	L*er* x RIL	5	14	1345	350	92	402	83
	L*er* x RIL	1	10	478	100	43	62	39
	L*er* x RIL	1	14	412	96	43	43	39
	L*er* x RIL	1	26	502	49	80	95	72
	L*er* x RIL	2	19	240	88	62	27	56
	L*er* x RIL	2	20	242	76	62	46	56
	L*er* x RIL	3	12	211	76	43	33	39
	L*er* x RIL	3	14	143	63	30	21	27
	L*er* x RIL	3	16	149	51	15	20	14
	L*er* x RIL	3	17	143	71	43	8	39
	L*er* x RIL	4	5	281	55	43	27	39
	L*er* x RIL	5	5	432	75	92	129	83
	L*er* x RIL	5	9	332	49	30	118	27
	L*er* x RIL	5	10	457	126	15	65	14
VIII	Ll-0 x RIL	1	6	598	47	96	194	97
IX	Ll-0 x RIL	3	23	820	156	84	206	85
X	Ll-0 x RIL	4	7	423	48	96	165	97
XI	Ll-0 x RIL	4	11	1165	503	96	134	97
XII	Ll-0 x RIL	4	15	1031	375	84	85	85
XIII	Ll-0 x RIL	5	2	1319	199	114	408	116
XIV	Ll-0 x RIL	5	4	1676	519	132	244	134
XV	Ll-0 x RIL	5	18	720	131	84	114	85
	Ll-0 x RIL	1	5	445	27	114	168	116
	Ll-0 x RIL	1	10	136	6	45	58	45
	Ll-0 x RIL	1	13	307	54	15	91	15
	Ll-0 x RIL	1	14	366	103	45	70	45
	Ll-0 x RIL	1	16	244	37	65	80	66
	Ll-0 x RIL	1	18	247	101	96	34	97
	Ll-0 x RIL	1	21	331	115	45	24	45
	Ll-0 x RIL	2	5	206	27	31	38	31
	Ll-0 x RIL	2	12	335	125	65	70	66
	Ll-0 x RIL	3	12	236	38	45	57	45
	Ll-0 x RIL	3	16	104	32	15	25	15
	Ll-0 x RIL	4	5	276	21	45	70	45
	Ll-0 x RIL	4	6	235	38	45	58	45
	Ll-0 x RIL	4	10	471	97	96	88	97
	Ll-0 x RIL	4	12	454	60	65	121	66
	Ll-0 x RIL	4	14	507	135	84	83	85
	Ll-0 x RIL	5	5	630	149	96	153	97
	Ll-0 x RIL	5	9	240	60	31	28	31
	Ll-0 x RIL	5	11	246	52	45	20	45
	Ll-0 x RIL	5	21	165	10	31	36	31

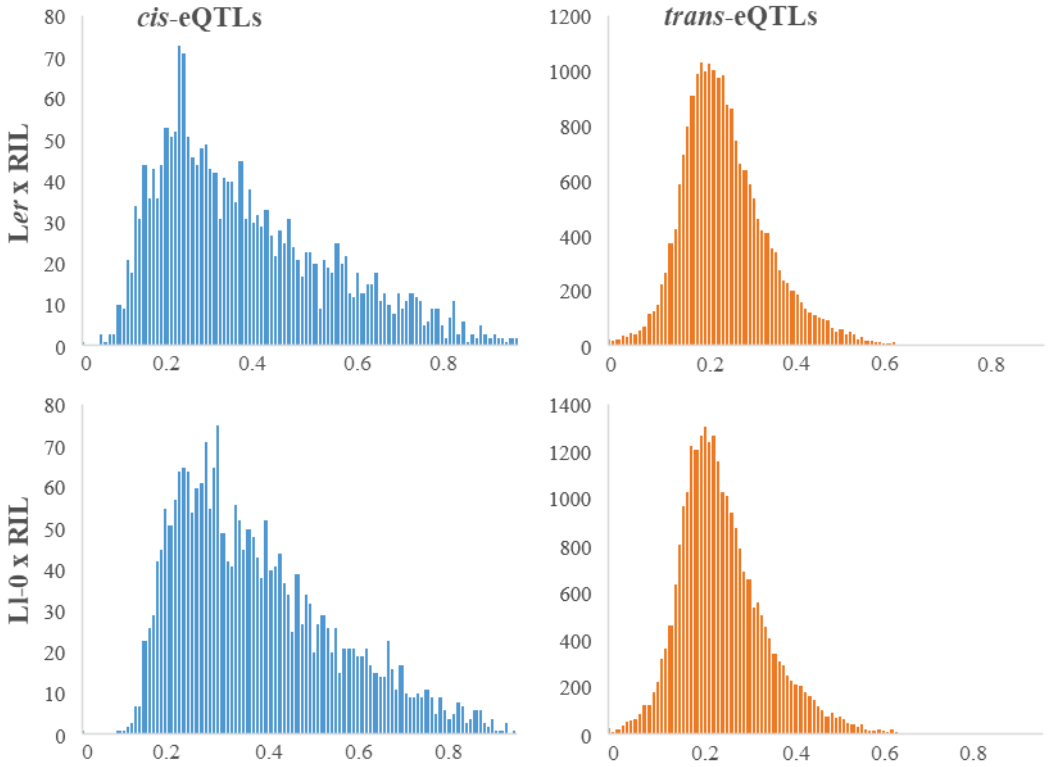


Figure R12. Histograms showing the distribution of the phenotypic variance explained by *cis*- and *trans*-eQTLs. The number of significant eQTLs (y-axis) is plotted against the fraction of phenotypic variance that they account for (x-axis). The graphs on the top correspond to the *cis*- (in blue) and *trans*-eQTLs (in orange) detected in the L*er* x RIL population, whereas the graphs on the bottom correspond to the ones detected in the Ll-0 x RIL population.

4.2.3.2 – Comparison of *cis*- and *trans*-regulated genes identified in the analysis of hybrids of parents vs those identified using hybrids of RILs with parents.

To examine the consistency of the *cis*- and *trans*-eQTLs detected in the analysis of the L*er* x RIL and Ll-0 x RIL populations, we compared these results with the ones obtained in the analysis of differential expression involving the parental crosses in roots of hybrids grown under low Pi conditions. First, we analysed the proportion of coincidences between the genes estimated as *cis*- or *trans*-regulated in the RILs with the DEGs detected based on ASE in the comparisons between alleles in the same hybrid or between hybrids and parents (see section **4.1.5**). The comparisons of *cis*-regulated genes were made separately for the SNPs equal or not to the reference Col-0 genome. Approximately one third of the genes detected as *cis*-regulated in the F_1 hybrids matched the *cis*-eQTLs. In addition, we confirmed coherence in the direction of the allelic effect in the *cis*-eQTLs and the type of gene expression difference in the F_1 hybrids. Interestingly, we observed a higher percentage of significant coincidences in *cis*-regulated genes upregulated in L*er* when the reference allele was equal to Ll-0 (47%) than equal to L*er* (31%). Similarly, the ones up-regulated in Ll-0 had a higher percentage of coincidences when the reference allele was equal to L*er* (47%)

than equal to Ll-0 (31%) (**Figure R13-A**). These lower percentages of coincidence when the haplotype of the allele is equal to that of the reference allele is in line with our previous conclusion of overestimation of *cis*-regulatory effects causing upregulation of the alleles with the reference haplotype and suggests that the problem of overestimation of *cis* effects is alleviated in the eQTL analysis with the RILs based on data on total counts per gene. Regarding the *trans*-regulated genes, the relative percentages of coincidences were higher (56-66%), although the absolute number of genes detected as differentially expressed was 40 times lower than the number of detected *trans*-eQTLs in the analysis of RILs (**Figure R13-B**).

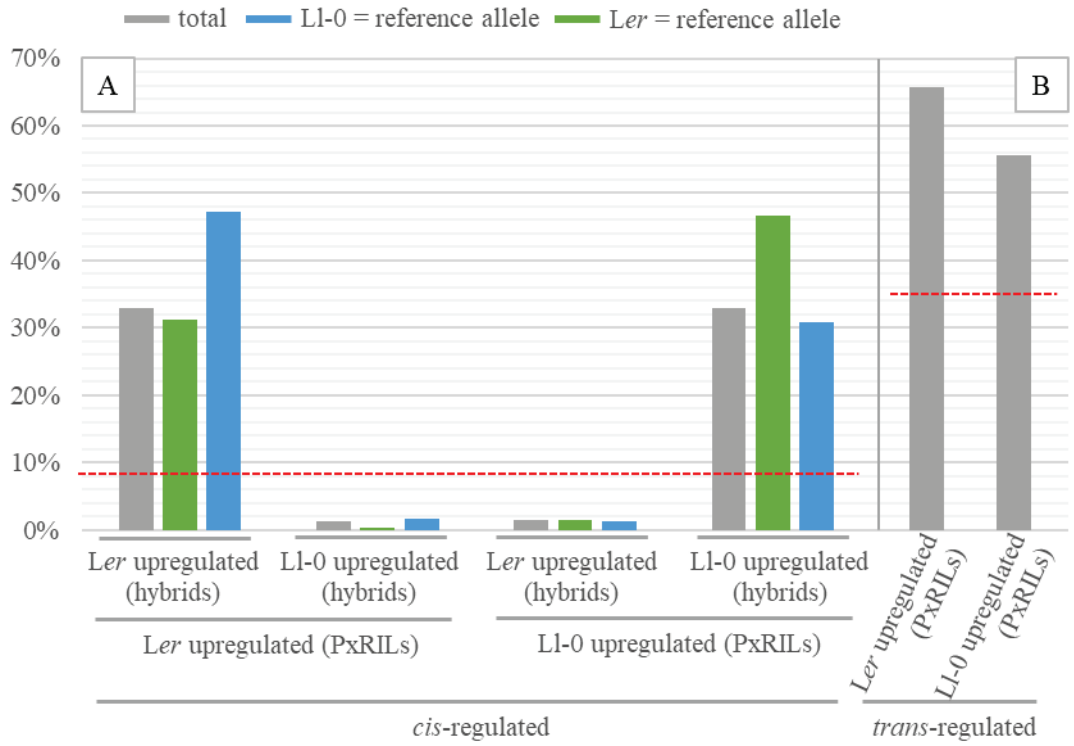


Figure R13. **Proportion of coincidences in *cis*- and *trans*-regulated genes detected through two different approaches**. Vertical bars indicate the percentage of coincidences between the *cis*- (**A**) and *trans*-regulated genes (**B**) detected in the analysis of differential gene expression of hybrids and parents with respect to the ones detected in the analysis of the L*er* x RIL and Ll-0 x RIL populations, separately for the type of difference in gene expression and the type of allelic effect, as indicated, as well as when the reference allele is equal to L*er* (green bars), equal to Ll-0 (blue bars) or without filtering by the haplotype of the allele (grey). The red dashed horizontal lines represent the percentage of expected coincidences in case of random distributions. In all cases, the number of observed coincidences was determined as statistically significant using a binomial test ($p<0.05$).

4.2.3.3 – Analysis of the coincidences between *cis*- and *trans*-eQTLs.

To get insights on the contribution of stabilizing and directional selection in determining *cis*- and *trans*-eQTL architecture, we classified the genes having associated eQTLs according to the type

of eQTL and the number and direction of their effects (**Figure R14**). Among genes only affected by *trans*-eQTLs, approximately 3,800 had just one *trans*-eQTL associated, while about 1,460 had more than one *trans*-eQTL associated, all of them acting in the same direction, and almost 3,200 were affected by more than one *trans*-eQTL, at least one of them acting in the opposite direction than the other/s. In contrast, for genes having a *cis*-eQTL, about 800 of them were not affected by *trans*-eQTLs, whereas 560 genes had one or more *trans*-eQTLs acting on the same direction as the *cis*-eQTL and 320 genes had one or more *trans*-eQTL acting on the opposite direction than the *cis*-eQTL. These observations indicate the higher prevalence of compensatory vs convergent effects among *trans*-eQTLs acting on the same gene (chi-square p-value $< 10^{-5}$), in contrast to the higher prevalence of convergent vs compensatory effects of *cis*- and *trans*-eQTLs acting on the same gene ($p< 10^{-5}$).

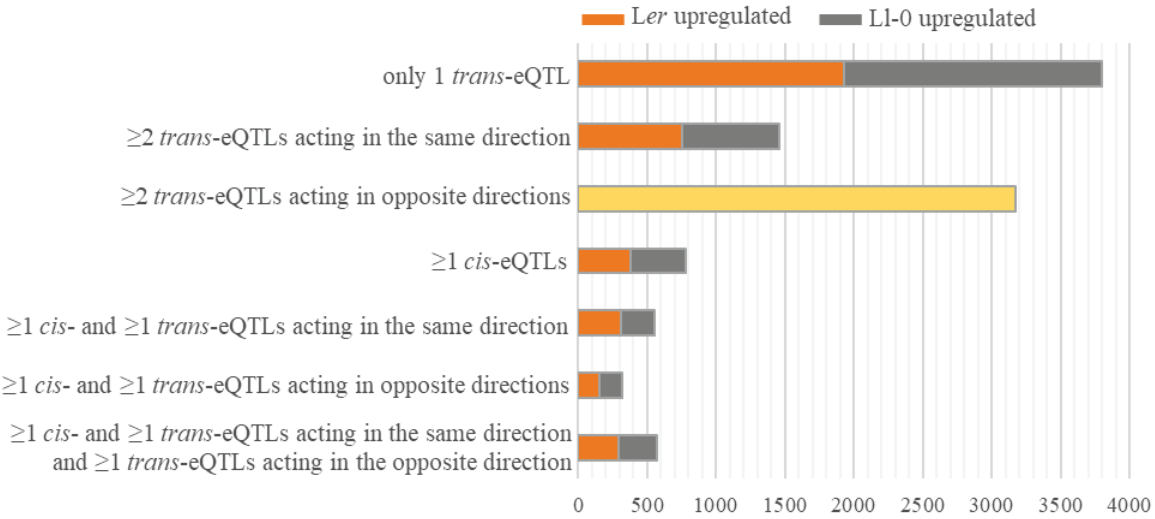


Figure R14. **Number of coincidences between *cis*- and *trans*-eQTLs**. Horizontal bars represent the observed number of coincidences in the genes associated to *cis*- and *trans*-eQTLs, separately for those up- and downregulated in L*er* and Ll-0 alleles. In bar 3 both L*er* and Ll-0 alleles are upregulated. In bars 4-7 the upregulated allele is referred to the *cis*.

4.2.3.4 – Analysis of enrichment of eQTL associated genes in PS-responsive genes.

As previously described, the L*er* x RIL and Ll-0 x RIL crosses were grown under low phosphate conditions and transcriptomic data was obtained from root samples. To analyse the overall enrichment of phosphate starvation-responsive genes in the genes having associated *cis*- and/or *trans*-eQTLs, as well as in the *trans*-eQTL hotspots, we compared them with the total set of root PS-induced and PS-repressed genes from section **4.1.3**. Overall, we observed a higher overrepresentation of PS-induced than PS-repressed genes, both in *cis*- and *trans*-eQTLs (**Figure R15-A**). Overrepresentation of PS-induced and PS-repressed genes occurred for *trans*-eQTLs detected in both PxRIL populations independent on the type of additive effect (positive or

negative), and the same was true for overrepresentation of PS-induced in the gene set associated to *cis*-eQTLs. However, in case of the *cis*-eQTL associated gene set, its enrichment in PS-repressed genes is only for cases in which the *cis*-eQTLs display negative additive value in the L*er* x RIL population and positive additive value in the Ll-0 x RIL population, clearly indicating non-randomness in *cis*-eQTL prevalence.

We also examined enrichment of PS-responsive genes in the fifteen major hotspots and significant overrepresentation of PS-responsive genes was found for the upregulated and/or downregulated gene sets associated to 11 out of the 15 major *trans*-eQTL hotspots analysed (**Figure 15-B**). Of them, 9 displayed enrichment in PS-induced genes and 6 in PS-repressed genes. Likewise, we detected significant underrepresentation in the upregulated and/or dowregulated gene sets associated of 11 hotspots, 5 in PS-induced genes and 9 in PS-repressed. In hotspot VIII the associated set of upregulated genes was significantly enriched in PS-repressed genes and the downregulated set, in PS-induced, in line with simple expectations. Likewise, we observed that the hotspots XI and XII displayed coherent behaviour, e.g., the upregulated genes associated to hotspot XI were overrepresented in PS-repressed genes, and the downregulated, in PS-induced. In contrast, we observed that two hotspot-associated gene sets displayed significant overrepresentation for both PS-induced and PS-repressed genes (the set of up-regulated genes in hotspot II and down-regulated in XIII). Nevertheless, most (12 gene sets) displayed significant overrepresentation only one of the PS-induced or PS-repressed gene sets. Regarding underrepresentation, both PS-induced and PS-repressed genes were significantly underrepresented in 4 gene sets associated to the hotspots I, IV, XIII, XIV, while 7 gene sets displayed underrepresentation for either PS-induced or PS-repressed.

4.2.3.5 – Gene ontology term enrichment of eQTL-associated genes.

To examine the fundamental biological processes and molecular functions of the genes associated with the *cis*- and *trans*-eQTLs detected, as well as with the fifteen major *trans*-eQTL hotspots, we analysed enrichment in gene ontology (GO) terms (see **section 3.8.1**).

As for *cis*-eQTLs, we observed overrepresentation of terms related to defense response with positive additive effect in the L*er* x RIL population and notably negative in the Ll-0 x RIL population (**Table R5-A**), reflecting directional selection of *cis*-variants of defense response genes towards upregulation in the L*er* vs the Ll-0 accession. As for the total set of *trans*-eQTLs, we observed enrichment in several GO terms such as ribosome biogenesis, translation, cell wall, stress response, photosynthesis, response to chitin, among others. In contrast to the case of ontology term enrichment in the set of *cis*-eQTLs, for *trans*-eQTLs associated gene sets, enrichment in several terms such as ribosome biogenesis, translation, and response to chitin occurred both in those with positive and negative additive effects.

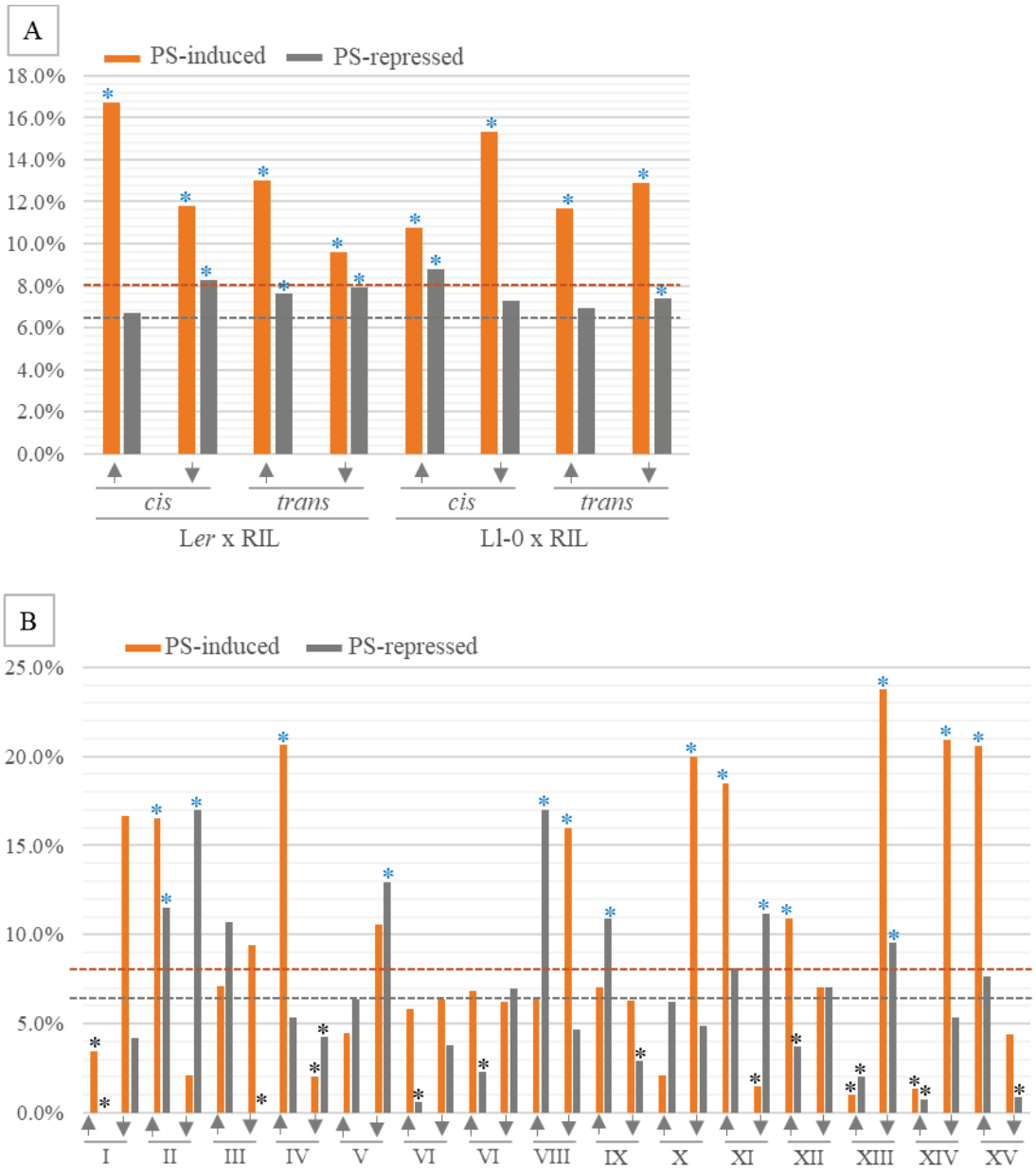


Figure R15. **Analysis of Pi starvation (PS)-responsive genes enrichment in the overall *cis*- and *trans*-eQTLs and in the fifteen major *trans*-eQTL hotspots**. Vertical bars show the percentage of observed coincidences in PS-induced (in orange) and PS-repressed (in black) genes respect to the total number of genes in the sets of up- and down-regulated genes **(A)** detected as *cis*- and *trans*-eQTLs in the L*er* x RIL and Ll-0 x RIL populations and **(B)** associated to each one of the *trans*-eQTL hotspots detected in the L*er* x RIL (I-VII) and Ll-0 x RIL (VIII-XV) populations. Grey arrows indicate positive (↑) and negative (↓) additive allelic effects. Orange and grey horizontal dashed lines show the percentage of expected coincidences in case of random distributions in the total sets of PS-induced and PS-repressed genes, respectively. Statistical significance for over- and under-representation was determined by binomial tests ($p < 0.05$) and denoted with blue and black asterisks (*), respectively.

Table R5. **Gene ontology (GO) term enrichment for the total sets of *cis*- and *trans*-eQTLs and the gene sets associated to the fifteen major *trans*-eQTL hotspots**. The most significative categories based on the top GO biological process assignments by (TAIR/TIGR) and the functional classification by the Munich information center for protein sequences (MIPS). GO term enrichment was estimated for the sets of up- (+) and down-regulated (-) genes **(A)** detected as *cis*- and *trans*-eQTLs in the L*er* x RIL and Ll-0 x RIL populations and **(B)** associated to each one of the *trans*-eQTL hotspots detected in the L*er* x RIL (I-VII) and Ll-0 x RIL (VIII-XV) populations. PSR-related categories are highlighted in yellow. Fold enrichment (FE) was calculated as the ratio between observed and expected frequencies. FE cutoff ≥ 1.5 and p-value ≤ 0.01 were used to determine GO terms as significantly overrepresented.

A

eQTLs			GO term	FE	p-value
cis -eQTLs	*Ler* x RIL	+	response to stress	1.8	2.01E-04
		+	disease, virulence and defense	2.9	4.12E-04
		+	transmembrane signal transduction	3.1	2.35E-03
		+	receptor enzyme mediated signalling	3.0	2.35E-03
		+	cell rescue, defense and virulence	1.7	7.85E-03
	Ll-0 x RIL	−	disease, virulence and defense	2.4	4.51E-03
		−	cell rescue, defense and virulence	1.6	7.34E-03

eQTLs			GO term	FE	p-value
trans -eQTLs	*Ler* x RIL	−	ribosome biogenesis	3.8	2.38E-45
		−	translation	3.5	5.41E-39
		−	ribosomal proteins	5.2	1.00E-30
		−	cytoplasm	2.1	1.86E-25
		−	cellular component biogenesis	2.9	3.16E-09
		−	response to abiotic stimulus	1.6	3.16E-09
		−	cell wall	1.6	8.21E-03
		+	stress response	1.6	3.71E-05
		+	oxidative stress response	2.1	2.93E-04
	Ll-0 x RIL	−	ribosome biogenesis	2.1	3.10E-11
		−	translation	2.0	2.29E-08
		−	chemoperception and response	1.6	3.46E-05
		−	response to chitin	2.6	1.34E-04
		−	ribosomal proteins	2.3	1.37E-04
		−	photosynthesis	3.0	9.70E-04
		+	ribosomal proteins	3.7	6.98E-17
		+	ribosome biogenesis	2.4	6.98E-17
		+	translation	2.3	3.34E-13
		+	cytoplasm	1.6	1.31E-09
		+	cellular component biogenesis	2.1	1.84E-04
		+	response to chitin	2.2	4.97E-03

B

Hotspot		GO term	FE	p-value
II	+	response to heat	12.0	3.62E-05
	+	response to hydrogen peroxide	16.0	7.66E-03
	+	response to high light intensity	16.0	8.75E-03
III	−	response to salt stress	5.0	1.79E-05
	−	response to cadmium ion	5.1	1.19E-04
	−	cellular catabolic process	3.6	7.93E-04
	−	developmental process	2.2	1.71E-03
IV	−	translation	14.9	3.34E-108
	−	cellular macromolecule biosynthetic process	5.0	1.60E-59
	−	cellular protein metabolic process	3.8	2.15E-47
	−	ribosome biogenesis	24.5	1.50E-45
	−	nucleosome assembly	10.5	7.36E-06
	−	cellular amino acid biosynthetic process	3.3	8.00E-03
	+	protein phosphorylation	3.4	4.35E-16
	+	phosphate metabolic process	3.2	5.38E-15
	+	cell wall polysaccharide metabolic process	13.0	9.21E-05
	+	cell tip growth	4.3	7.91E-03
V	−	vesicle docking involved in exocytosis	40.0	9.38E-03
	+	nucleic acid metabolic process	3.4	3.64E-03
VIII	−	ribosome biogenesis	16.0	1.11E-07
	−	translation	5.5	2.58E-06
	−	defense response to bacterium	5.9	5.60E-03
XI	−	photosynthesis	29.0	6.90E-10
	−	photosynthetic electron transport chain	31.0	5.34E-03
	+	protein phosphorylation	3.1	1.58E-08
	+	phosphate metabolic process	2.9	1.58E-08
	+	cell wall organization or biogenesis	3.8	3.36E-03
	+	cell death	3.6	8.59E-03
	+	enzyme linked receptor protein signaling pathway	4.6	9.59E-03

Hotspot		GO term	FE	p-value
XIII	−	cell wall macromolecule metabolic process	11.0	2.92E-04
	−	response to chitin	6.8	2.92E-04
	−	response to carbohydrate stimulus	4.5	1.63E-03
	−	defense response to fungus	5.6	3.47E-03
	+	nucleic acid metabolic process	3.7	7.60E-06
	+	translation	4.9	3.44E-05
	+	cellular nitrogen compound metabolic process	2.6	5.88E-05
	+	cellular macromolecule biosynthetic process	2.7	1.22E-04
	+	DNA replication	13.0	1.46E-04
	+	response to ionizing radiation	22.0	4.36E-04
	+	histone arginine methylation	137.3	7.34E-03
XIV	−	protein phosphorylation	3.5	6.52E-06
	−	phosphate metabolic process	3.3	6.52E-06
	−	amide transport	58.2	2.81E-03
	−	water transport	34.9	7.63E-03
	−	fluid transport	34.9	7.63E-03
	+	translation	11.6	1.06E-61
	+	cellular macromolecule biosynthetic process	4.2	8.23E-35
	+	cellular macromolecule metabolic process	2.3	1.16E-24
	+	ribosome biogenesis	14.5	4.27E-18
	+	cellular component biogenesis	8.3	1.48E-13
XV	−	translation	7.0	5.56E-05
	−	macromolecule metabolic process	2.1	8.52E-04
	−	macromolecule biosynthetic process	3.1	1.69E-03
	−	cellular nitrogen compound metabolic process	2.8	2.84E-03
	−	DNA replication	14.0	6.55E-03

Regarding the gene sets associated to the hotspots (**Table R5-B**), we observed overrepresentation of GO terms in nine out of the fifteen hotspots analysed, which speaks for intrinsic coherence of *trans*-eQTL hotspot-associated gene sets, suggesting that they are controlled by one or a few genes rather than by a large number of independent roles. Some ontology terms were found overrepresented in several hotspots. Thus, GO terms related to PSRs like phosphate metabolic processes, cell wall organization, or photosynthesis were overrepresented in hotspots IV, XI and XIV. Additionally, hotspots II and III displayed enrichment in responses to stress, whereas terms related to housekeeping functions like translation, ribosome biogenesis, DNA replication and macromolecule metabolic and biosynthetic processes, among others, were observed in the hotspots VIII, XIII and XV.

4.2.3.6 – Analysis of expression coherence within genes associated to the hotspots.

Trans-eQTL hotspots are typically considered to correspond to long-distance master regulators that potentially control multiple target genes and downstream pathways. In line with this idea is the data showing GO term enrichment in the gene sets associated to many of the hotspots. One additional expectation under this premise would be that the genes associated to the major *trans*-eQTL hotspots detected in our study would display expression coherence, i.e., resemblances in their response to at least some environmental or developmental cue. To test this, we inspected the expression of the sets of up- and downregulated genes associated to each one of the hotspots using the Genevestigator V3 database (Hruz *et al.*, 2008 ; see **section 3.8.2**). Indeed, we found that the sets of up- and downregulated genes associated to hotspots displayed signs of expression coherence (**Figure R16**).

Thus, for the case of Hotspot XIV there are several treatments that elicit either the up- or the downregulation of most of the upregulated genes associated to this hotspot (e.g., treatments 4, 27 and 48; **Figure R16-A**). Moreover, some treatments elicited an opposite response in up- and downregulated genes associated (e.g., treatment 27 for hotspot I; **Figure R16-A**). A global picture of expression coherence is given in **Figure 16-B**. It is based on an index of expression coherence within a gene set for a given treatment, and it is defined as the percentage of genes in a set upregulated (FC>1,5) minus the percentage of genes downregulated in that set (FC<-1,5) by that treatment. Based on this, we selected the 60 most informative treatments to perform hierarchical clustering. Essentially, for each hotspot we selected the four treatments in which their associated genes displayed the highest absolute expression coherence values. Some expression features are shared among gene sets associated to different hotspots, although and the direction of the effect of the *trans*-eQTL hotspot on their associated genes may act in the same or in the opposite direction. For instance, the upregulated gene sets of Hotspot V and VII display similar expression

behaviour (similar direction of the effect of both hotspots), and the same is true for the upregulated gene set of Hotspot I and the downregulated gene set of Hotspot III (opposite direction of the effects of both hotspots; **Figure 16-B**).

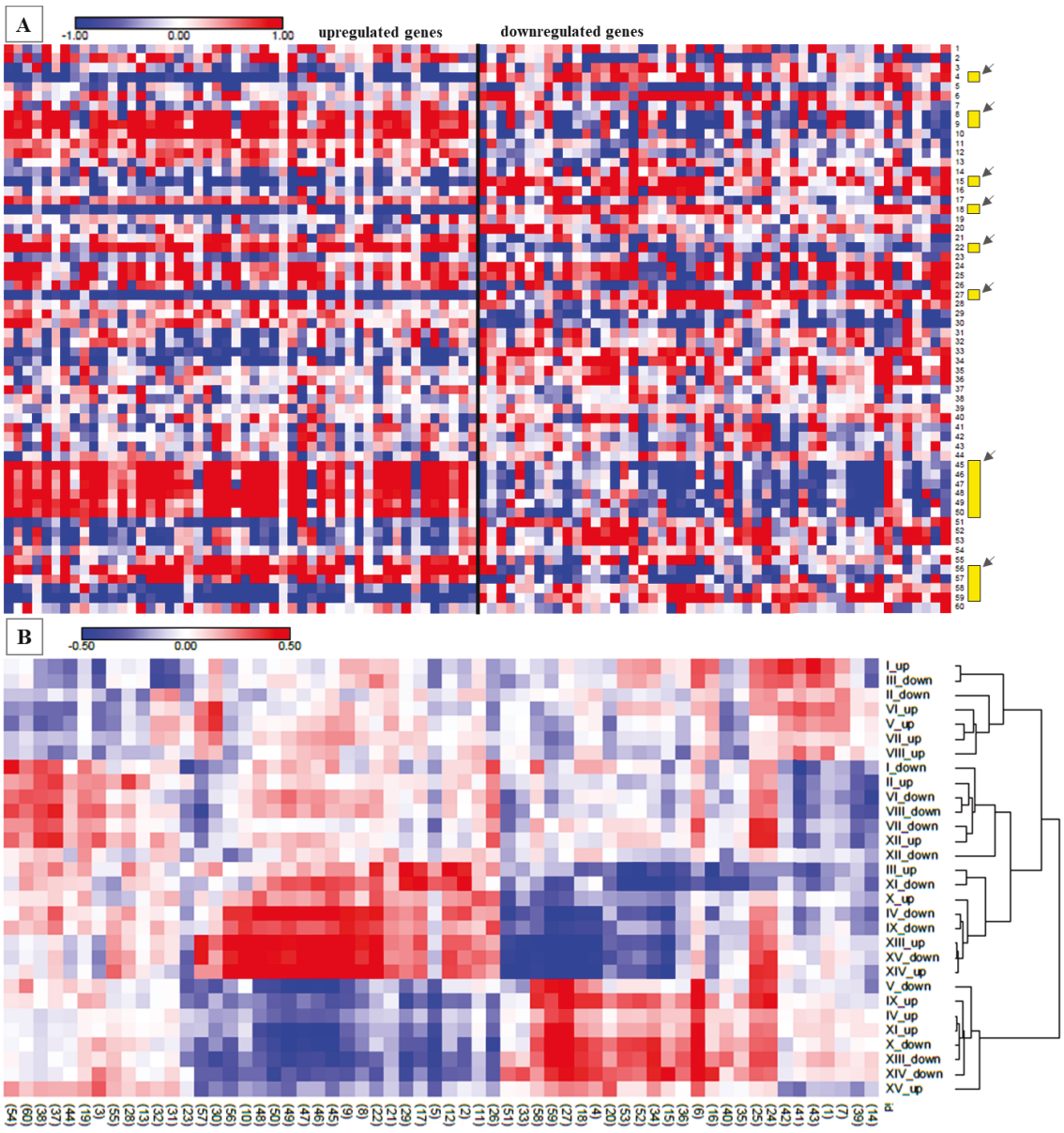


Figure R16. **(A) Heatmap showing the expression coherence of the up- and downregulated genes associated to the hotspot XIV**. Two sets of fifty up- (the columns on the left) and down-regulated genes (right) associated to the hotspot XIV were randomly chosen and their expression values (log_2 ratio) in response to sixty of the most informative studies are represented in the row-column intersections. Some instances that illustrate the expression coherence within these genes are highlighted in yellow. **(B) Hierarchical clustering of the major fifteen *trans*-eQTL hotspots based on the expression patterns of their associated genes under specific conditions**. The columns represent the sets of up- and down-regulated genes associated to each hotspot and the rows correspond to the sixty most informative studies selected. The boxes in the row-column intersections show indexes of expression coherence calculated by subtracting the proportions of upregulated (log_2 ratio>0.5) and downregulated genes (log_2 ratio<-0.5).

These indexes illustrate the overall tendency for a given set of genes to be upregulated (red) or downregulated (blue) under the specific conditions of these studies. The dendrogram was obtained using one minus Pearson correlation as distance measure through the hierarchical clustering tool from the website https://software.broadinstitute.org/morpheus/. The expression data were obtained using the Genevestigator database. (1) 35S:LecRK-VI.2 / Col-0; (2) 35S:RPS4-HS eds1-2 / 35S:RPS4-HS [shift 28°C to 19°C study 3 (35S:RPS4-HS eds1-2) / shift 28°C to 19°C study 3 (35S:RPS4-HS)]; (3) ABA study 8 (Col-0) / solvent treated leaf samples (Col-0); (4) ambient CO2 (mature leaf 10) / ambient CO2 (leaf 10 primordia); (5) BL study 2 (brx) / mock treated seedlings (brx); (6) brxS / Sav-0 [mock treated seedlings (brx) / mock treated seedlings (Sav-0)]; (7) bzr1-1Dxbri1-116 / Col-0; (8) callus formation (48h) / untreated root samples; (9) callus formation (96h) / untreated root samples; (10) callus formation study 3 (35d + 1d) / untreated hypocotyl samples (35d); (11) Cd study 3 (Col-0) / untreated root samples (Col-0); (12) ced1 / Col-5 [osmotic study 3 (ced1) / osmotic study 3 (Col-5)]; (13) circadian clock study 9 (17h dark+1h light) / circadian clock study 9 (18h dark); (14) cycloheximide / mock treated seedlings; (15) DFPM + ABA (Col-0) / solvent treated seedling samples (Col-0); (16) drought (dor) / untreated leaf samples (dor); (17) efr-1 / Col-0 [EF-Tu (elf18) study 4 (efr-1) / EF-Tu (elf18) study 4 (Col-0)]; (18) elevated CO2 study 3 (mature leaf 10) / elevated CO2 study 3 (leaf 10 primordia); (19) fd-3 / Col-0 [shift SD to LD study 7 (fd-3) / shift SD to LD study 7 (Col-0)]; (20) fenclorim (24h) / solvent treated root culture samples (24h); (21) gai / penta [MeJa study 5 (gai) / MeJa study 5 (penta)]; (22) germination (12h) / seed desiccation; (23) heat study 3 / dark grown Col-0 seedling samples; (24) hypoxia study 6 (ANAC102(KO-1)) / untreated plant samples (ANAC102(KO-1)); (25) hypoxia study 6 (Col-0) / untreated plant samples (Col-0); (26) iron deficiency / protoplasting / iron deficiency study 8 (24h); (27) IVT-E RNA labeling (root elongation zone) / IVT-E RNA labeling (root tip); (28) light study 4 (cli186) / dark grown cli186 seedlings; (29) light study 6 (Col-0) / dark grown Col-0 seedlings; (30) M. incognita study 2 (Pico) / non-infested root cell samples (Pico); (31) non-polysomal RNA study 2 (Col-0) / total RNA study 3 (Col-0); (32) non-polysomal RNA study 2 (pab2 pab8) / total RNA study 3 (pab2 pab8); (33) osmotic study 2 (late) / untreated root samples (late); (34) P. syringae pv. maculicola (Col-0) / mock treated leaf samples (Col-0); (35) P. syringae pv. tomato study 10 (DC3000) / mock inoculated leaf samples; (36) P. syringae pv. tomato study 9 (DC3118 Cor-) / mock inoculated leaf samples; (37) polysomal RNA study 2 (rpl24b) / total RNA study 2 (rpl24b); (38) ribosome-bound mRNA (untreated seedlings) / total mRNA (untreated seedlings); (39) S. malt. R3089 volatiles (6h) / mock treated seedling samples (6h); (40) S. sclerotiorum study 2 (coi1-2) / mock inoculated rosette leaf samples (coi1-2); (41) salicylic acid study 13 (CS57549) / mock treated rosette leaf samples (CS57549); (42) salicylic acid study 13 (CS57789) / mock treated rosette leaf samples (CS57789); (43) salicylic acid study 13 (CS57836) / mock treated rosette leaf samples (CS57836); (44) salicylic acid study 5 (Van-0) / silwet L77 treated Van-0 leaf samples (28h); (45) salt / FACS (32h) / root epidermis and lateral root cap protoplast samples of mock treated pWER::GFP (1h); (46) salt / FACS (48h) / root epidermis and lateral root cap protoplast samples of mock treated pWER::GFP (1h); (47) salt / FACS (48h) / root epidermis and lateral root cap protoplast samples of mock treated pWER::GFP (48h); (48) salt / FACS study 2 (32h) / root cortex protoplast samples of mock treated pCOR315.1::GFP (1h); (49) salt / FACS study 4 (48h) / root stele protoplast samples of mock treated pWOL::GFP (1h); (50) salt / FACS study 4 (48h) / root stele protoplast samples of mock treated pWOL::GFP (48h); (51) salt study 2 (late) / untreated root samples (late); (52) shift 28°C to 19°C study 3 (35S:RPS4-HS rrs1-11) / 28°C (35S:RPS4-HS rrs1-11); (53) shift 28°C to 19°C study 3 (35S:RPS4-HS) / 28°C (35S:RPS4-HS); (54) shift light to dark (rib.-bound mRNA) / shift light to dark (tot. mRNA); (55) shift low light to dark to high light (flu1-1) / shift low light to dark (flu1-1); (56) shift NPA to NAA (6h) / NPA study 3; (57) shift to pH 4.6 / protoplasting / shift to pH 4.6 (24h); (58) sulfur deficiency study 3 (LZ2) / sulfur deficiency study 3 (LZ1); (59) two-cycle RNA labeling (root elongation zone) / two-cycle RNA labeling (root tip); (60) zeatin study 2 (Col-0) / solvent treated aerial parts (Col-0).

4.2.3.7 – Analysis of hotspots resemblances based on coincidences among associated gene sets and enrichment in transcription factor gene targets.

Given the expression coherence among genes associated to the *trans*-eQTL hotspots, we reasoned that underlying this they would share transcriptional regulators (TFs). To test this, we examined the representation of the target genes of 390 transcription factors (TFs) obtained by DAP-seq (O'Malley *et al.*, 2016) in the sets of up- and downregulated genes associated to the hotspots. Significant overrepresentation (FDR-adj < 0.01; fold enrichment >1.4) was observed for targets of one or more of these TFs in all of the fifteen hotspot-associated gene sets, both up-regulated and down regulated, except for three cases (I downregulated; V- and VI-upregulated). In total, targets of 310 of the 390 TFs examined were enriched in at least one of the hotspots gene sets. There were many instances of TF targets enriched in gene sets of several *trans*-eQTLs (**Table R6**), in line with the above-mentioned inter-hotspot similarities/antagonisms in the expression behaviour of their associated genes (**Figure R16**).

Given the similarities in expression behaviour and in TF target enrichment patterns of genes associated to different hotspots, we examined the number of coincidences between the sets of up- and downregulated genes associated to these hotspots and 61 significant overlaps between the gene sets analysed were found (**Table R7**).

4.2.3.8 – Analysis of the mode of inheritance of dominance, codominance and overdominance between eQTL alleles.

To get insights on the mode of inheritance we followed a simple "qualitative" approach to get estimations of cases of codominant, dominant, and over/underdominant inheritance. It is based on classifying the eQTLs identified in one of the populations (LOD≥2,8, additive value ≥ +0,3 or ≤ -0,3) in which the effect of one homozygous allele is compared to that of the heterozygote, according to the likelihood that the eQTL is present in the second population (LOD≥2,8; dominant, additive value: ≤ -0,3 or ≥ +0,3; over/underdominant, additive value ≥ +0,3 or ≤ -0,3) or absent (LOD<1,6; additive value < +0,1 or > -0,1). Using this qualitative approach, we estimated that *cis*-eQTLs show a predominant codominant inheritance (46-52% vs 20-21% dominant and 0% overdominant), whereas *trans*-eQTLs display a predominant dominant inheritance (62-75% vs 2-3% codominant and 0.1-0.2% overdominant) (**Figure R17**).

Table R6. Overlaps in the transcription factors with overrepresented gene targets between the fifteen major hotspots. The total number of TF and TF families (in brackets) with overrepresented (p-adj<0.01; fold enrichment ≥1.4) gene targets in the different sets of up- and downregulated genes associated to the major fifteen hotspots are shown in the first row and in the first column, whereas the row-column intersections show the number of overlaps between sets. The cells highlighted in yellow denote significant overrepresentation, determined by binomial test ($p < 0.05$), when the number of coincidences was ≥2 and fold enrichment ≥1.5 (fold enrichment was calculated as the number of observed overlaps respect to the number of expected coincidences in case of random distributions).

			II		III		IV		V		VI		VII		VIII	
			upreg.	downreg.	upreg.	downreg.	upreg.	downreg.	upreg.	downreg.	upreg.	downreg.	upreg.	downreg.	upreg.	downreg.
			65[12]	5[3]	2[1]	10[6]	78[10]	31[5]	0[0]	61[17]	0[0]	76[11]	1[1]	110[19]	1[1]	51[12]
I	upreg.	5[3]	0[2]	0[1]	0[1]	5[3]	0[1]	0[1]	0[0]	1[2]	0[0]	0[2]	0[0]	0[2]	0[0]	0[1]
	downreg.	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]
II	upreg.	65[12]		1[1]	2[1]	0[2]	2[2]	19[3]	0[0]	14[5]	0[0]	47[7]	0[0]	47[8]	0[0]	36[6]
	downreg.	5[3]			0[0]	0[2]	1[3]	0[0]	0[0]	2[3]	0[0]	0[1]	0[0]	1[2]	0[0]	1[1]
III	upreg.	2[1]				0[1]	0[0]	1[1]	0[0]	2[1]	0[0]	2[1]	0[0]	2[1]	0[0]	2[1]
	downreg.	10[6]					1[2]	0[1]	0[0]	2[4]	0[0]	0[2]	0[0]	2[3]	0[0]	0[1]
IV	upreg.	78[10]						0[1]	0[0]	19[8]	0[0]	3[4]	1[1]	25[9]	0[0]	1[4]
	downreg.	31[5]							0[0]	7[4]	0[0]	24[4]	0[1]	24[4]	0[0]	23[4]
V	upreg.	0[0]								0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]
	downreg.	61[17]									0[0]	16[7]	0[1]	29[12]	0[0]	16[10]
VI	upreg.	0[0]										0[0]	0[0]	0[0]	0[0]	0[0]
	downreg.	76[11]											0[1]	66[10]	0[0]	44[7]
VII	upreg.	1[1]												0[1]	0[0]	0[0]
	downreg.	110[19]													0[0]	43[10]
VIII	upreg.	1[1]														0[0]
	downreg.	51[12]														

			IX		X		XI		XII		XIII		XIV		XV	
			upreg.	downreg.	upreg.	downreg.	upreg.	downreg.	upreg.	downreg.	upreg.	downreg.	upreg.	downreg.	upreg.	downreg.
			107[21]	6[3]	2[1]	106[19]	96[14]	14[10]	50[13]	6[3]	5[2]	219[29]	25[7]	126[18]	67[12]	6[3]
I	upreg.	5[3]	0[2]	0[1]	0[0]	0[2]	2[2]	0[1]	0[2]	0[0]	0[1]	0[2]	1[1]	0[2]	0[2]	0[0]
	downreg.	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]
II	upreg.	65[12]	33[6]	3[1]	0[1]	17[6]	5[4]	3[4]	25[7]	0[0]	2[1]	53[11]	12[4]	21[8]	25[6]	3[2]
	downreg.	5[3]	2[2]	0[0]	0[0]	2[3]	3[3]	0[2]	1[2]	0[0]	0[0]	2[3]	0[0]	4[3]	1[2]	0[0]
III	upreg.	2[1]	2[1]	1[1]	0[0]	2[1]	0[1]	0[0]	0[1]	0[0]	0[1]	2[1]	0[1]	0[1]	2[1]	0[0]
	downreg.	10[6]	0[2]	0[1]	0[0]	0[3]	4[3]	0[2]	2[3]	0[0]	0[1]	1[3]	1[1]	1[3]	0[2]	0[0]
IV	upreg.	78[10]	29[8]	0[0]	0[0]	55[9]	72[8]	0[3]	4[5]	0[1]	0[0]	74[10]	0[1]	63[9]	15[6]	0[1]
	downreg.	31[5]	10[4]	1[1]	0[0]	8[3]	0[2]	1[2]	20[3]	0[2]	4[2]	23[5]	19[4]	0[4]	18[4]	1[1]
V	upreg.	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]
	downreg.	61[17]	38[12]	2[2]	0[1]	41[13]	25[9]	1[5]	11[9]	0[2]	1[2]	52[16]	5[6]	23[13]	34[11]	0[1]
VI	upreg.	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]
	downreg.	76[11]	34[9]	4[1]	2[1]	24[7]	6[6]	2[6]	38[8]	1[1]	3[1]	62[10]	17[6]	16[9]	32[6]	0[1]
VII	upreg.	1[1]	0[1]	0[0]	0[0]	0[0]	0[0]	0[1]	0[0]	0[1]	0[0]	0[1]	0[1]	0[1]	0[0]	0[1]
	downreg.	110[19]	57[13]	3[1]	1[1]	47[12]	30[10]	1[5]	45[11]	1[1]	3[1]	92[17]	17[6]	38[13]	45[10]	0[1]
VIII	upreg.	1[1]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]	0[0]
	downreg.	51[12]	28[10]	3[1]	0[1]	18[10]	3[6]	1[3]	28[8]	0[1]	2[2]	40[12]	16[6]	6[10]	26[9]	0[0]
IX	upreg.	107[21]		3[1]	0[1]	59[14]	34[9]	4[8]	16[8]	1[3]	1[2]	92[18]	6[6]	54[13]	46[11]	0[1]
	downreg.	6[3]			0[0]	3[2]	1[2]	0[0]	0[1]	0[0]	0[1]	4[2]	0[1]	1[2]	1[1]	0[0]
X	upreg.	2[1]				0[0]	0[0]	0[1]	1[1]	0[0]	0[0]	2[1]	0[1]	2[1]	0[1]	0[0]
	downreg.	106[19]					61[12]	0[5]	15[8]	0[1]	3[2]	99[19]	6[4]	60[14]	42[10]	0[0]
XI	upreg.	96[14]						1[4]	6[8]	0[0]	0[1]	86[12]	0[2]	69[11]	15[7]	0[0]
	downreg.	14[10]							0[4]	1[2]	0[0]	3[8]	1[3]	3[6]	2[3]	2[1]
XII	upreg.	50[13]								0[0]	2[1]	36[11]	16[4]	9[10]	24[8]	0[1]
	downreg.	6[3]									0[1]	3[3]	0[2]	1[2]	0[1]	0[1]
XIII	upreg.	5[2]										3[2]	4[2]	0[2]	3[2]	0[0]
	downreg.	219[29]											16[7]	114[18]	62[12]	0[2]
XIV	upreg.	25[7]												1[6]	14[4]	2[1]
	downreg.	126[18]													25[10]	0[2]
XV	upreg.	67[12]														0[0]
	downreg.	6[3]														

Table R7. **Gene overlaps between the fifteen major *trans*-eQTL hotspots**. The total number of up- and down-regulated genes associated to the fifteen *trans*-eQTL hotpots is represented in the first row and in the first column. The row-column intersects show the number of gene overlaps between them and the fold enrichment (in brackets), which is calculated as the number of observed overlaps respect to the number of expected coincidences in case of random distributions. The cells highlighted in yellow denote significant overrepresentation determined by binomial test ($p<0.05$), when the number of coincidences was ≥3 and the fold enrichment ≥2.

			II		III		IV		V		VI		VII		VIII	
			upreg. 139	downreg. 47	upreg. 28	downreg. 244	upreg. 673	downreg. 582	upreg. 156	downreg. 206	upreg. 157	downreg. 85	upreg. 171	downreg. 157	upreg. 350	downreg. 402
I	upreg.	174	0	0	0	124 [55]	0	0	6 [4.2]	0	0	0	0	2 [0.5]	0	0
	downreg	24	0	0	1 [28.3]	0	0	0	0	0	0	0	0	0	1 [16.9]	0
II	upreg.	139		0	1 [4.9]	0	0	21 [4.9]	0	0	0	1 [0.9]	1 [0.4]	16 [5.4]	0	15 [10.6]
	downreg	47			0	1 [1.7]	6 [3.6]	0	1 [2.6]	0	0	0	1 [1.2]	0	0	0
III	upreg.	28				0	0	1 [1.2]	0	0	0	0	0	0	0	0
	downreg	244					8 [0.9]	0	3 [1.5]	0	0	0	0	1 [0.2]	0	1 [0.4]
IV	upreg.	673						0	0	0	2 [0.3]	0	20 [1.6]	1 [0.1]	0	2 [0.3]
	downreg	582							0	0	0	1 [0.2]	1 [0.1]	52 [4.2]	0	12 [2.0]
V	upreg.	156								0	17 [12.0]	0	35 [12.1]	0	3 [7.7]	0
	downreg	206									0	2 [2.9]	0	13 [7.2]	0	2 [2.3]
VI	upreg.	157										0	0	0	5 [11.8]	1 [0.6]
	downreg	85											0	0	0	15 [9.4]
VII	upreg.	171												0	5 [5.8]	0
	downreg	157													0	31 [7.6]
VIII	upreg.	350														0
	downreg	402														

			IX		X		XI		XII		XIII		XIV		XV	
			upreg. 47	downreg. 194	upreg. 48	downreg. 165	upreg. 503	downreg. 134	upreg. 375	downreg. 85	upreg. 199	downreg. 408	upreg. 519	downreg. 244	upreg. 131	downreg. 114
I	upreg.	174	0	1 [0.5]	0	0	0	0	3 [0.9]	0	0	1 [0.3]	0	0	0	1 [1.0]
	downreg	24	0	1 [3.8]	0	0	0	0	0	1 [9.3]	0	0	2 [3.1]	0	0	1 [7.0]
II	upreg.	139	1 [0.9]	4 [2.7]	2 [5.7]	0	1 [0.3]	3 [3.1]	21 [7.7]	0	0	5 [1.7]	7 [1.8]	1 [0.6]	4 [4.2]	1 [1.2]
	downreg	47	1 [2.6]	0	0	0	1 [0.8]	1 [3.0]	0	1 [4.8]	0	0	0	1 [1.7]	0	1 [3.5]
III	upreg.	28	0	0	0	1 [4.1]	0	1 [5.1]	1 [1.8]	0	1 [3.4]	0	1 [1.3]	0	0	2 [11.9]
	downreg	244	0	0	0	1 [0.5]	5 [0.8]	0	0	1 [0.9]	3 [1.2]	4 [0.8]	4 [0.6]	2 [0.6]	2 [1.2]	0
IV	upreg.	673	9 [1.6]	0	0	35 [6.0]	148 [8.3]	0	2 [0.2]	2 [0.7]	0	95 [6.6]	0	55 [6.4]	7 [1.5]	0
	downreg	582	1 [0.2]	17 [2.7]	3 [2.0]	0	0	20 [4.9]	13 [1.1]	2 [0.8]	36 [5.9]	0	142 [8.9]	0	6 [1.5]	17 [4.9]
V	upreg.	156	0	2 [1.2]	1 [2.5]	0	0	2 [1.8]	0	0	4 [2.4]	0	1 [0.2]	1 [0.5]	0	2 [2.1]
	downreg	206	2 [2.9]	0	0	4 [5.4]	20 [8.9]	1 [1.7]	3 [1.8]	0	1 [1.1]	15 [8.2]	1 [0.4]	0	2 [3.4]	0
VI	upreg.	157	1 [0.7]	0	0	0	3 [0.7]	0	0	2 [2.6]	2 [1.1]	0	2 [0.4]	0	1 [0.8]	3 [2.9]
	downreg	85	0	8 [4.7]	1 [2.5]	1 [0.7]	0	0	35 [11.3]	0	1 [0.6]	1 [0.3]	3 [0.7]	0	2 [1.8]	0
VII	upreg.	171	0	0	2 [2.3]	1 [0.3]	1 [0.1]	4 [1.6]	1 [0.1]	4 [2.6]	5 [1.4]	0	3 [0.3]	4 [0.9]	0	11 [5.2]
	downreg	157	2 [0.6]	6 [1.4]	0	4 [1.1]	14 [1.3]	2 [0.7]	80 [10.1]	0	1 [0.2]	12 [1.4]	13 [1.2]	3 [0.6]	14 [5.1]	0
VIII	upreg.	350	0	0	0	0	0	1 [3.0]	0	2 [9.5]	8 [16.3]	0	0	0	0	2 [7.1]
	downreg	402	1 [0.6]	9 [4.3]	0	1 [0.6]	3 [0.6]	0	44 [11.5]	0	0	12 [2.9]	0	0	4 [3.0]	0
IX	upreg.	47		0	0	15 [11.1]	11 [2.7]	3 [2.7]	0	1 [1.4]	0	22 [6.6]	0	2 [1.0]	0	0
	downreg	194			5 [9.6]	0	0	0	14 [3.4]	0	7 [3.2]	0	7 [1.2]	0	2 [1.4]	0
X	upreg.	48				0	0	5 [14.8]	0	0	4 [8.0]	0	1 [0.8]	0	0	0
	downreg	165					97 [22.2]	0	3 [0.9]	0	0	20 [5.6]	0	0	0	0
XI	upreg.	503						0	0	1 [0.4]	0	54 [5.0]	0	18 [2.8]	1 [0.3]	0
	downreg	134							0	1 [1.7]	0	0	2 [0.5]	0	1 [1.1]	1 [1.2]
XII	upreg.	375								0	0	1 [0.1]	20 [2.0]	0	3 [1.2]	0
	downreg	85									0	0	0	2 [1.8]	0	0
XIII	upreg.	199										0	0	0	1 [0.7]	6 [5.0]
	downreg	408											0	0	12 [4.3]	0
XIV	upreg.	519												0	0	48 [15.4]
	downreg	244													13 [7.7]	0
XV	upreg.	131														0
	downreg	114														

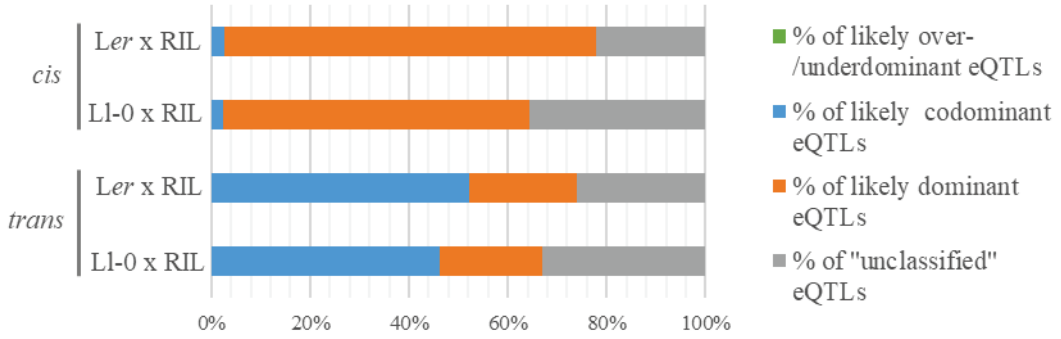


Figure R17. **Mode of inheritance of the *cis*- and *trans*-eQTLs**. Horizontal bars show the percentage of likely dominant (in orange), codominant (in blue), overdominant and underdominant (in green) and unclassified (in grey) *cis*- and *trans*-eQTLs detected in the L*er* x RIL and Ll-0 x RIL populations.

4.2.3.9 – Analysis of gene coexpression networks.

The data obtained in our study can be used to identify clusters of coregulated genes in which expression changes are determined by both genetic and environmental factors. To identify coexpressed modules, we used WGCNA (see **section 3.8.1**) based on FPKM expression values, separately for PS-induced and PS-repressed genes, using the combined set of 63 lines from the L*er* x RIL and Ll-0 x RIL populations. As a result, we obtained 12 and 10 modules of PS-induced and PS-repressed genes, respectively. We searched for potential TF candidates regulating these modules based on the information of targets for 390 TFs obtained by DAP-seq (O'Malley *et al.*, 2016). To identify TF candidates, we first tested the correlation relationships between the expression values of a given TF and the 19,017 genes analysed in the combined set of L*er* x RIL and Ll-0 x RIL populations. We then took the 10% of the genes with the highest correlation score (highly correlated) and the 10% with the lowest (highly anticorrelated) for each TF. We filtered the target lists of each TF with their corresponding lists of highly correlated and highly anticorrelated, and we used these lists of highly correlated and highly anticorrelated gene targets to determine coincidences with the gene sets of each coexpression module and evaluate statistically significant enrichments using binomial tests (Fold enrichment >1.5; FDR<0.05).

We found that all modules displayed enrichment in at least one of the highly correlated or highly anticorrelated lists of TF targets and a total of 163 TFs displayed target overrepresentation (p-adj <0.05; FC ≥ 1.4) in at least one of the 22 gene modules (**Figure 18-A**). Of them, 43 TFs showed canonical behaviour (i.e., all coexpresion modules of the same type, induced or repressed, potentially regulated by a given TF displayed enrichment of the same list of targets, highly correlated or highly anticorrelated, whereas coexpression modules of different type displayed enrichment of a different target list, e.g., highly correlated for repressed gene modules, and highly

anticorrelated for the coexpression modules of induced genes, or the reciprocal). In contrast, we found 62 TFs showing non-canonical behaviour, e.g., PS-induced targets overrepresented in both genes that correlate and genes that anticorrelate with a given TF (**Figure 18-B**).

A

		gene modules											
		1	2	3	4	5	6	7	8	9	10	11	12
		PS-ind	PS-ind	PS-ind	PS-ind	PS-ind	PS-ind	PS-ind	PS-ind	PS-ind	PS-ind	PS-ind	PS-ind
highly correlated →		6(3)	36(12)	25(7)	15(5)	5(5)	5(4)	9(2)	6(4)	5(3)	1(1)	39(13)	43(7)
anticorrelated →		2(2)	21(15)	9(7)	5(4)	1(1)	7(6)	4(4)	2(2)	3(3)	0(0)	14(10)	17(13)
1	PS-ind	**8(5)**											
2	PS-ind	0(4)	**57(21)**										
3	PS-ind	3(3)	19(13)	**34(13)**									
4	PS-ind	1(3)	8(7)	4(6)	**20(9)**								
5	PS-ind	0(3)	0(4)	2(3)	1(3)	**6(5)**							
6	PS-ind	0(4)	4(6)	0(4)	0(3)	0(4)	**12(8)**						
7	PS-ind	1(3)	5(6)	2(5)	9(5)	1(3)	1(3)	**12(6)**					
8	PS-ind	2(2)	3(5)	8(5)	1(3)	0(1)	0(2)	1(2)	**8(5)**				
9	PS-ind	1(3)	1(5)	0(4)	0(2)	2(3)	1(3)	1(2)	0(2)	**8(6)**			
10	PS-ind	1(1)	0(1)	1(1)	1(1)	0(0)	0(0)	1(1)	1(1)	0(0)	**1(1)**		
11	PS-ind	3(4)	11(12)	3(9)	0(5)	3(4)	7(6)	2(5)	1(5)	6(6)	0(1)	**53(14)**	
12	PS-ind	3(4)	30(14)	23(9)	13(7)	1(4)	3(6)	9(5)	6(4)	0(3)	1(1)	11(8)	**60(16)**
1	PS-rep	0(1)	0(1)	0(1)	0(1)	0(1)	0(1)	0(1)	0(0)	0(1)	0(0)	0(1)	1(1)
2	PS-rep	5(5)	38(16)	19(11)	9(8)	2(4)	4(8)	5(5)	1(4)	3(5)	0(1)	17(11)	37(15)
3	PS-rep	0(1)	0(1)	0(1)	0(1)	0(1)	0(1)	0(1)	0(1)	0(1)	0(0)	0(1)	0(1)
4	PS-rep	2(4)	2(6)	2(5)	3(5)	0(3)	1(5)	4(5)	1(3)	1(2)	0(1)	3(5)	4(6)
5	PS-rep	0(1)	0(1)	0(1)	1(1)	0(1)	0(1)	1(1)	0(0)	0(1)	0(0)	0(1)	1(1)
6	PS-rep	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
7	PS-rep	0(1)	0(4)	1(4)	1(3)	0(1)	0(1)	0(3)	0(1)	0(2)	0(1)	1(5)	0(2)
8	PS-rep	0(1)	0(1)	0(1)	0(2)	0(1)	0(1)	0(1)	0(1)	0(1)	0(0)	0(1)	0(1)
9	PS-rep	0(2)	1(5)	1(4)	0(3)	0(3)	0(4)	0(2)	0(3)	0(3)	0(0)	3(6)	3(5)
10	PS-rep	0(2)	3(5)	1(5)	1(5)	0(3)	2(4)	2(3)	0(3)	1(2)	0(0)	4(5)	4(5)

		gene modules									
		1	2	3	4	5	6	7	8	9	10
		PS-rep	PS-rep	PS-rep	PS-rep	PS-rep	PS-rep	PS-rep	PS-rep	PS-rep	PS-rep
highly correlated →		0(0)	43(15)	1(1)	4(2)	1(1)	0(0)	6(5)	1(1)	3(3)	7(4)
anticorrelated →		1(1)	33(19)	0(0)	5(5)	0(0)	0(0)	1(1)	1(1)	3(3)	4(3)
1	PS-rep	**1(1)**									
2	PS-rep	0(1)	**75(24)**								
3	PS-rep	0(0)	0(1)	**1(1)**							
4	PS-rep	0(1)	3(7)	0(1)	**9(7)**						
5	PS-rep	0(1)	1(1)	0(0)	0(1)	**1(1)**					
6	PS-rep	0(0)	0(0)	0(0)	0(0)	0(0)	**0(0)**				
7	PS-rep	0(0)	2(5)	0(0)	1(2)	0(0)	0(0)	**7(6)**			
8	PS-rep	0(0)	0(1)	0(1)	0(1)	0(0)	0(0)	0(0)	**2(2)**		
9	PS-rep	0(0)	3(6)	0(1)	0(3)	0(0)	0(0)	1(2)	0(1)	**6(6)**	
10	PS-rep	0(1)	5(6)	0(1)	3(4)	0(1)	0(0)	0(2)	0(1)	2(4)	**11(6)**

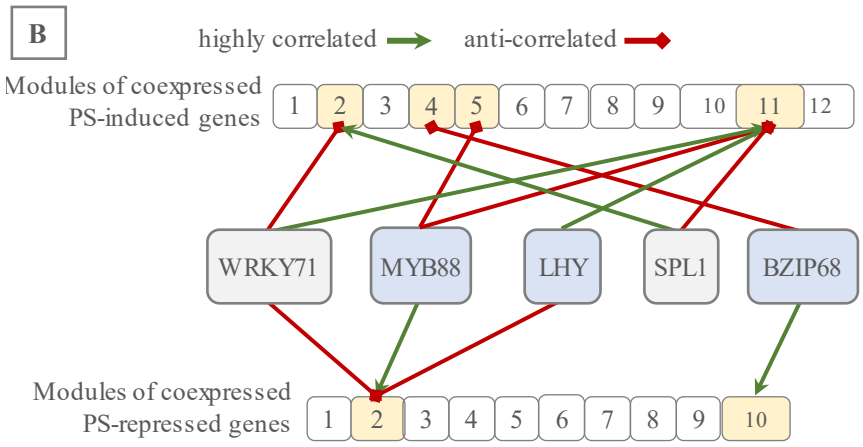


Figure R18. **Overrepresentation of transcription factor (TF) gene targets in the 22 modules of coexpressed genes**. **(A)** Overlaps in the TFs with overrepresented gene targets between the modules of coexpressed genes induced (PS-ind) and repressed (PS-rep) by phosphate starvation. The numbers of TFs whose highly correlated or highly anticorrelated gene targets were enriched (FDR-adj <0.05; fold enrichment ≥1.4) in each module are shown in the first two rows, and their sum is given in the central diagonal and highlighted in bold. The numbers of families represented by these TFs are shown in brackets. The row-column intersects show the number of overlaps between coexpression modules in their candidate TF regulators. The cells highlighted in yellow denote significant overrepresentation, determined by binomial test ($p < 0.05$), when the number of coincidences was ≥2. **(B)** Examples of candidate TFs regulators following canonical (in blue) and non-canonical (in grey) behaviours. For instance, LHY and BZIP68 follow a canonical behaviour because while modules of induced genes are enriched in highly correlated (green arrows) and anticorrelated (red arrows) targets, respectively, the modules of repressed genes are enriched in highly anticorrelated and highly correlated targets, respectively.

4.2.4 – QTL ANALYSIS OF MORPHOLOGICAL AND PHYSIOLOGICAL TRAITS AND OVERLAPS WITH eQTLs

4.2.4.1 – QTL analysis of traits related to Pi starvation responses.

Parallel to eQTL analysis, we performed QTL analysis of the main traits related to PSRs in the L*er* x RIL and Ll-0 x RIL populations. We estimated a global LOD score threshold value of 2.75 through permutation tests to determine QTL significance in both populations (see **section 3.6.3.2**). The significant QTLs detected are summarized in **Table R8**, which also includes the positions of the major *trans*-eQTL hotspots to indicate potential overlaps between them. We detected QTLs across the five chromosomes for most of the traits analysed, 40 and 46 for traits in the low Pi and in the high Pi growth condition, respectively. In addition, we observed potential colocalization between three QTLs for traits under low Pi growth conditions and *trans*-eQTL hotspots. Following the same qualitative approach than that used for examining the mode of inheritance of eQTLs, we observed prevalence of the dominant mode of inheritance (89% dominant vs 9% codominant and 2% overdominant), which is quite similar to the case of *trans*-eQTLs.

Table R8. Significant QTLs of morphological and physiological traits associated to Pi starvation responses. Arrows pointing at the right side of the chromosomes indicate the approximate location of the significant LOD score peaks (≥2.75) detected in the L*er* x RIL (in green) and Ll-0 x RIL (in blue) populations. For these QTLs, details are provided about the trait, Pi regimen under it was detected, LOD score, phenotypic explained variance (R^2) and additive effects. Arrows pointing at the left side of the chromosomes indicate the approximate location of the major *trans*-eQTL hotspots (in roman numerals) and others of minor importance denoted by the marker number.

chromosome 1

Hotspot	Marker	Trait	LOD	R^2	Additive effect
	1	Root-to-shoot fresh weight (+P)	3.93	0.16	0.03
	4	Primary root length 9D (+P)	3.22	0.13	3.28
		Root water content (+P)	3.92	0.28	0.01
		Root fresh weight (-P)	5.83	0.41	0.58
VIII	6	Primary root length 7D (+P)	3.86	0.23	-5.04
		Primary root length 9D (+P)	2.82	0.16	-5.32
	10	Anthocyanin content (-P)	2.97	0.19	4.32
	12	Root phosphate content (-P)	3.13	0.20	-4.51
		Root phosphate content (+P)	5.76	0.41	-16.69
I	13	Anthocyanin content (-P)	3.53	0.18	-3.96
		Root phosphate content (+P)	5.76	0.41	-16.69
		Shoot dry weight (-P)	3.16	0.13	0.04
	16	Shoot fresh weight (-P)	3.39	0.17	0.35
	17	Shoot phosphate content (-P)	3.81	0.24	4.00
	19	Primary root length 12D (-P)	3.62	0.19	-5.70
	21	Shoot dry weight (-P)	3.86	0.19	-0.05
	22	Root dry weight (-P)	3.67	0.20	-0.03
	25	Primary root length 9D (+P)	3.59	0.21	4.37
		Anthocyanin content (-P)	5.60	0.35	-4.48
	26	Root-to-shoot dry weight (+P)	5.62	0.25	0.04
	27	Primary root length 12D (-P)	4.13	0.26	-6.01
		Root-to-shoot dry weight (+P)	4.60	0.25	-0.04
	28	Shoot dry weight (-P)	4.64	0.23	0.08
		Root-to-shoot fresh weight (+P)	6.90	0.37	-0.06

chromosome 2

Hotspot	Marker	Trait	LOD	R^2	Additive effect
	6	Root fresh weight (+P)	3.75	0.24	1.34
	8	Root dry weight (+P)	4.60	0.27	-0.11
		Shoot fresh weight (+P)	3.79	0.13	-2.30
	9	Root fresh weight (-P)	2.87	0.13	-0.39
	10	Shoot water content (-P)	3.40	0.32	0.01
		Shoot dry weight (+P)	5.42	0.26	-0.17
	11	Primary root length 12D (-P)	3.32	0.20	-5.76
	13	Primary root length 7D (-P)	3.33	0.20	3.00
	14	Root water content (-P)	3.73	0.30	0.01
II	15	Primary root length 9D (+P)	4.53	0.29	-5.50
	17	Shoot phosphate content (-P)	2.93	0.17	-2.19
		Root dry weight (-P)	3.35	0.18	-0.04

chromosome 3

Hotspot	Marker	Trait	LOD	R^2	Additive effect
	1	Shoot phosphate content (+P)	3.52	0.18	13.03
		Shoot phosphate content (+P)	2.93	0.16	-13.75
	2	Shoot fresh weight (-P)	5.38	0.31	0.36
III	3				
		Shoot fresh weight (-P)	3.95	0.24	0.44
		Shoot dry weight (-P)	3.38	0.15	0.05
	6	Anthocyanin content (-P)	3.10	0.20	-4.61
		Root-to-shoot dry weight (-P)	3.78	0.22	-0.03
		Root-to-shoot fresh weight (+P)	4.71	0.22	-0.04
	7	Root-to-shoot dry weight (+P)	6.67	0.36	-0.04
	11	Root fresh weight (+P)	5.82	0.30	-1.67
	12	Shoot phosphate content (-P)	3.54	0.23	3.81
	15	Root phosphate content (+P)	3.28	0.18	13.22
	18	Shoot water content (-P)	3.32	0.28	-0.01
	21	Root dry weight (+P)	3.73	0.20	-0.09
		Primary root length 7D (+P)	6.88	0.22	-3.34
	22	Primary root length 7D (-P)	3.41	0.18	-2.46
		Primary root length 9D (-P)	4.16	0.24	-4.24
IX	23				

chromosome 4

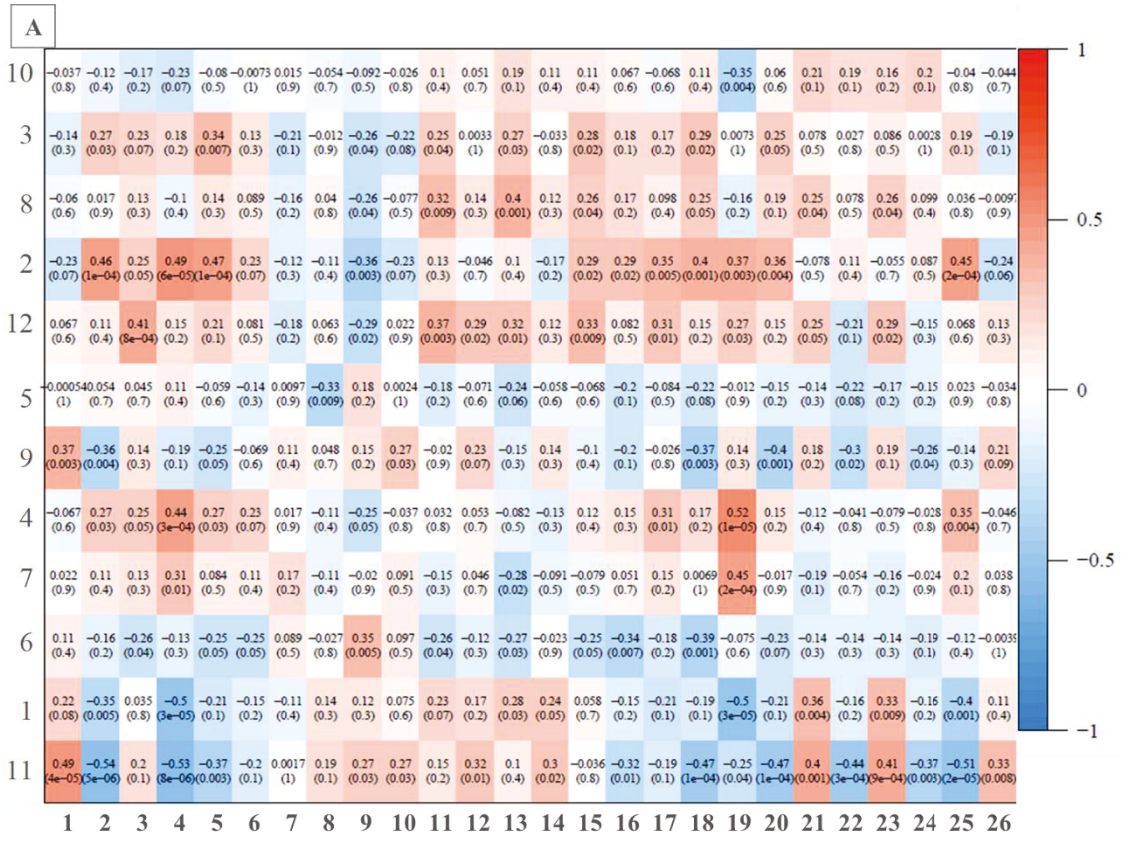
Hotspot	Marker	Trait	LOD	R^2	Additive effect
		Root-to-shoot dry weight (-P)	4.72	0.26	0.05
IV	1	Root-to-shoot dry weight (+P)	4.62	0.25	0.04
		Root-to-shoot fresh weight (+P)	3.46	0.13	0.03
	3	Primary root length 7D (+P)	8.05	0.28	3.60
		Primary root length 9D (+P)	4.44	0.19	3.96
V	4	Anthocyanin content (-P)	4.75	0.41	-6.68
X	7				
	10	Root phosphate content (+P)	2.81	0.15	11.77
		Shoot phosphate content (+P)	3.69	0.19	14.23
XI	11	Primary root length 7D (+P)	2.81	0.06	1.80
XII	15				

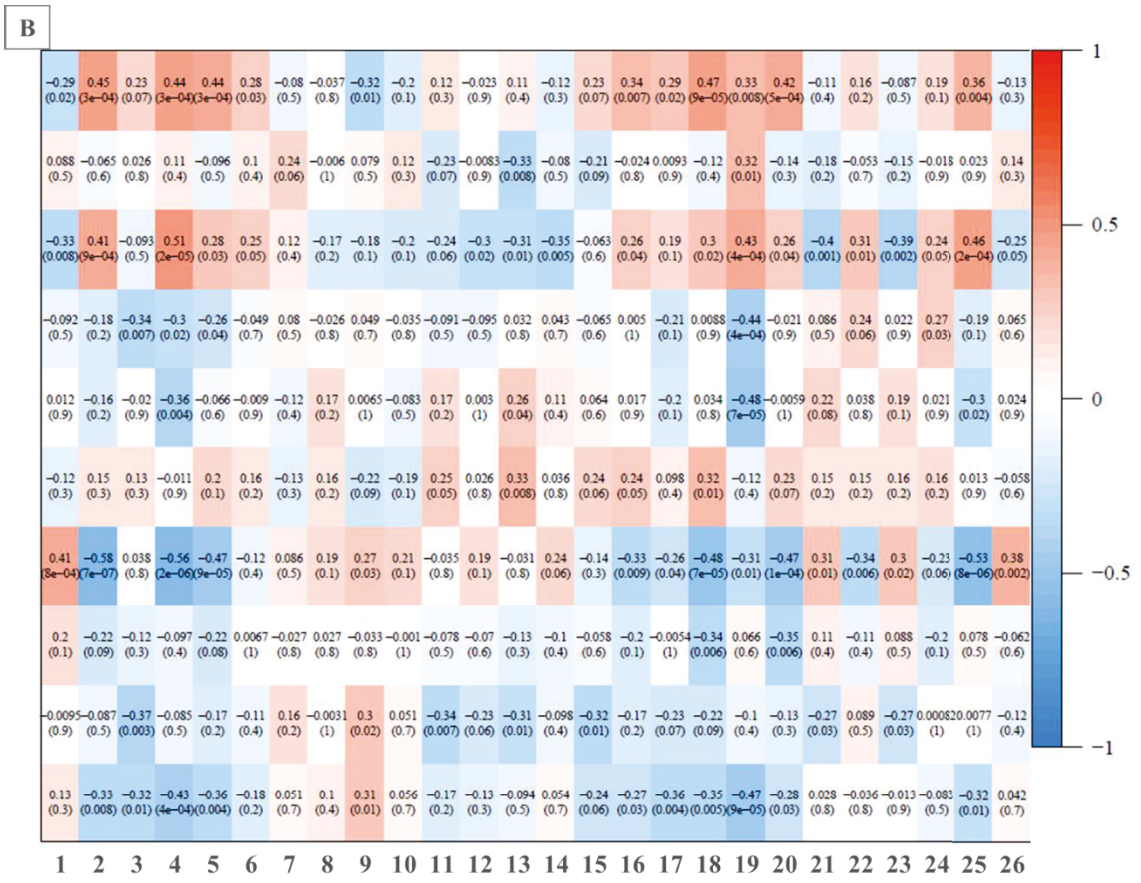
chromosome 5

Hotspot	Marker	Trait	LOD	R^2	Additive effect
	1	Root phosphate content (-P)	2.85	0.18	-4.23
XIII	2				
XIV	4				
		Primary root length 7D (+P)	5.01	0.13	-2.55
		Primary root length 9D (+P)	3.68	0.15	-3.62
	7	Shoot dry weight (+P)	6.37	0.33	-0.20
		Shoot fresh weight (+P)	7.30	0.48	-4.37
		Shoot phosphate content (-P)	2.83	0.17	-2.02
	9	Primary root length 9D (-P)	5.46	0.35	-5.39
VI	11	Shoot dry weight (+P)	3.76	0.30	0.26
		Shoot fresh weight (+P)	2.99	0.26	3.61
		Shoot fresh weight (+P)	6.43	0.30	-3.69
	12	Root phosphate content (+P)	3.86	0.23	-12.41
		Root dry weight (+P)	3.66	0.34	0.12
		Primary root length 7D (+P)	5.91	0.41	5.68
	13	Primary root length 7D (-P)	2.78	0.16	2.50
		Shoot dry weight (+P)	3.28	0.13	-0.13
VII	14				
	15	Shoot water content (+P)	4.22	0.28	0.00
		Root fresh weight (-P)	4.83	0.28	-0.56
XV	18	Root dry weight (-P)	3.32	0.19	-0.04
		Primary root length 7D (-P)	3.67	0.23	-3.21
	20	Shoot fresh weight (-P)	3.09	0.18	-0.54
	21	Primary root length 9D (+P)	3.87	0.16	-3.60
	22	Primary root length 7D (+P)	3.89	0.09	-2.08
	24	Shoot phosphate content (+P)	4.25	0.26	-17.37
		Root-to-shoot fresh weight (-P)	5.23	0.41	-0.16

4.2.4.2 – Correlations between coexpression modules and physiological and morphological traits.

To find potential causal relationships between PSR-related traits and the coexpression modules, we evaluated the correlation coefficients between these two types of data (**Figure R19**). There are several correlations with high statistical significance ($r > \pm 0.4$; $p<0.0001$), which concern plant performance not only in low Pi growth conditions, but also in high Pi growth conditions. For instance, module 10 of PS-repressed genes displays significant correlation with root-to-shoot fresh weight ratio under high Pi ($r = -0.56$; $p<0.0001$) and low Pi ($r=-0,47$, $p<0,0001$) and with root dry weight in low Pi growth conditions ($r = -0.48$; $p<0.0001$). This opens the possibility that TFs identified as candidate regulators of coexpression modules could impact the traits associated to these modules.





[Figure R19](). **Relationships of gene coexpression modules with traits associated to phosphate starvation responses**. Each column is associated to a trait (1-30), whereas rows correspond to gene modules of coexpressed PS-induced **(A)** and PS-repressed **(B)** genes. In the row-column intersections are shown the Pearson correlation coefficients and the p-values (in brackets) estimated by WGCNA (weighted gene coexpression analysis) for each trait and gene coexpression module. Correlation strength is denoted by the colors red (+1), blue (-1) and white (0). Traits: 1=Anthocyanin content (–P); 2=Root-to-shoot dry weight ratio (+P); 3=Root-to-shoot dry weight ratio (–P); 4= Root-to-shoot fresh weight ratio (+P); 5=Root-to-shoot fresh weight ratio(–P); 6=Phosphate root content (+P); 7=Phosphate root content (–P); 8=Phosphate shoot content (+P); 9=Phosphate shoot content (–P); 10=Primary root length at 12 days (–P); 11=Primary root length at 7 days (+P); 12=Primary root length at 7 days (–P); 13=Primary root length at 9 days (+P); 14=Primary root length at 9 days (–P); 15=Root dry weight (+P); 16= Root dry weight (–P); 17=Root fresh weight (+P); 18= Root fresh weight (–P); 19=Root water content (+P); 20= Root water content (–P); 21=Shoot dry weight (+P); 22= Shoot dry weight (–P); 23=Shoot fresh weight (+P); 24= Shoot fresh weight (–P); 25=Shoot water content (+P); 26= Shoot water content (–P).

5. DISCUSSION

To cope with growth under Pi starvation, plants have evolved an array of adaptive responses oriented to optimize Pi acquisition and use efficiency. Pi starvation responses (PSRs) include some morphological, developmental, physiological, and biochemical modifications. Underlying PSRs there are large transcriptomic changes whose importance has been demonstrated by the impact of mutation of transcriptional regulators on plant performance under low Pi growth regimens (Chiou and Lin, 2011; Puga *et al.*, 2017). Thus, these adaptive responses provide a suitable system for studying the natural variation of gene expression, particularly because soils exhibit very large variation in Pi concentration, and thus Pi nutrition should represent a strong force for local environment adaptation.

In this thesis work we set to study the *cis*- and *trans*-regulatory components responsible for natural variation of gene expression during Pi starvation in Arabidopsis. We followed two different experimental approaches, one consisting in the analysis of differential gene expression between F_1 hybrids and their homozygous parents, including the analysis of ASE. The other approach was based on eQTL analysis on hybrids of RILs with their progenitors. For all these studies we chose two parents whose hybrids displayed some signs of heterosis for some physiological traits and for which RIL lines were available. In addition to describe the Pi starvation transcriptome in two organs, and in two parental lines and their hybrids, these approaches provided information on: i) the relevance of environmental/developmental effects on expression variation; ii) the low but significant impact of maternal effects and parental imprinting mediating expression changes; iii) *cis*- and *trans*-eQTLs using hybrids of RILs, their characteristics and mode of inheritance; v) the relative contribution of stabilizing and directional selection in shaping natural variation of the Pi starvation transcriptome; vi) the correlations between coexpression modules/hotspot *trans*-eQTLs and physiological/performance traits.

5.1 – NATURAL VARIATION IN THE TRANSCRIPTOME OF TWO ACCESSIONS AND THEIR RECIPROCAL HYBRIDS

5.1.1 – The phosphate starvation transcriptome of L*er*, Ll-0, and their reciprocal F_1 hybrids

From a preliminary analysis of PSR-related traits of four natural accessions and all possible reciprocal F_1 hybrids with one of them (L*er*), we selected the Ll-0 and L*er* ecotypes for subsequent analysis, as they and their hybrids displayed the highest differences in the traits analysed and hybrids showed signs of heterosis for some of these traits (e.g., shoot and root biomass, Pi content, and primary root length). The transcriptomic analysis of L*er*, Ll-0, and their hybrids, provided

information on the Pi starvation response in these genotypes. Considering the number of gene expression differences displayed in response to this stress, the shoot transcriptome was less responsive to Pi starvation in all genotypes. Among genotypes, the number of genes responsive to Pi starvation was generally similar, although the shoot of Ll-0 displayed a slightly lower number of Pi responsive genes. These analyses allowed us to generate two gene sets of PS-responsive genes, representing the ones induced or repressed in any of the genotypes (total) and in all the genotypes (core). In these total and core sets, we could detect enrichment in GO terms related to primary PSR-related functions such as Pi transport, cell wall organization, galactolipid biosynthesis, Pi metabolism and anthocyanin biosynthesis. Enrichment in these categories was higher in the core set of Pi starvation induced genes, which is in line with the very high relevance of these primary PSR-related functions for adaptation to growth under low Pi regimens.

5.1.2 – Environmental and developmental influence on gene expression variation

The comparison among the transcriptomes of L*er* and Ll-0 and their hybrids revealed very large expression changes between parents (3,000-4,000 DEGs), being these changes smaller in the shoots of plants grown under low Pi regimens. Between parents and hybrids, these differences were smaller in number (700-2,800 DEGs), specially between hybrids and L*er* than with Ll-0 (**Figure R3**). The lower differences between hybrids and parents than between parents is in agreement with the notion that gene expression in hybrids is generally within the expression range of the parents, indicative of additive or dominant effects as major drivers of gene expression in hybrids. We also detected a lower, but still non-negligible number of genes that displayed higher or lower expression in the hybrids than in either parent, the magnitude of which was dependent on the organ type and growth condition. These genes displaying expression beyond the parental range can be considered to represent cases of molecular trait heterosis (overdominance), which are sensitive to environmental and developmental cues. In contrast, the two reciprocal hybrids showed hardly any significant difference between each other in the two organs (shoots and roots) and the two growth conditions examined (high and low Pi). The implications of this finding are discussed in the next section.

The fact that transcriptomic analyses were made using RNA-seq allowed us to examine ASE, making it possible to detect *cis*- and *trans*-regulatory effects determining variation in gene expression. One conclusion from this analysis is that the number of *trans* effects detected for the L*er* alleles were much smaller than those detected for Ll-0 alleles, except in the case of root in low Pi, indicating that the transcriptional environment of the hybrids is more similar to that of L*er* than to that of Ll-0, which is in line with the results of the analysis of overall expression differences based on transcript total counts. This situation is different for *cis*-regulatory effects,

where both the L*er* and Ll-0 alleles account for a similar number of these effects. Additionally, *cis*-regulatory effects were much higher in number than *trans* effects (10-600 vs 1,000-1,800) and it is also evident that *cis* effects are more prominent in roots than in shoots. One possible explanation for the difference in *cis* effects observed in roots vs shoots is that roots display higher evolvability, perhaps because their simpler developmental program.

Regarding these findings, it is to be noted however that we have observed a potential bias in ASE analysis due to the mapping procedure of sequence reads, which would be more efficient for the alleles equal to that of the Col-0 accession used as the mapping reference. In fact, the ratio of estimated *cis*-regulatory effects leading to upregulation of the reference allele is 4-fold higher for *cis*-regulatory effects leading to upregulation of the alternative allele. In principle, this problem could be overcome by polymorphism masking in the Col-0 genome (i.e., substituting the polymorphic site with an N), as then it would be neutral for efficiency of mapping reads independent of the presence or absence of any polymorphism. This approach is currently underway in the framework of a collaboration with the group of Rodrigo Gutiérrez (Universidad Católica, Santiago de Chile). In any case, the likely bias on mapping efficiency of reads of the reference vs the alternative allele does not influence comparisons of the same allele between two genotypes or two environmental or developmental conditions, as there is not bias for the reference versus the alternative allele either on data of *trans*-regulatory effects or in comparisons between two environmental or developmental conditions (**Figures R6-B**; **R7**). Additionally, this bias in mapping efficiency should not affect the estimations on *cis*-regulatory effects leading to downregulation of the reference allele, and even considering *cis* effects downregulating the reference allele to be similar in number to those acting in the opposite direction, we conclude that there is a much larger number of *cis*- than of *trans*-regulatory effects detected in the analysis of hybrids (hybrids vs parents for *trans* effects).

One important question regarding *cis*- and *trans*-regulatory effects is their specificity, i.e., how they are influenced by environmental and developmental factors, the genotype x environment interaction. Regarding *cis*-regulatory effects, we found in our study that more of these effects shared between the same organ (root or shoot) under the two different growth conditions examined than between the two organs for the same growth condition. One possible explanation could be that root and shoot developmental programs are more different between each other than the high Pi and low Pi programs in any of these organs. In agreement with this, we found that differences in gene expression between roots and shoots, both at high Pi and low Pi growth conditions, are much higher than the differences on gene expression mediated by the Pi growth regimen in any of these organs (**Figure D1**). Also of note is that roots displayed higher similarities in *cis*-regulatory effects than shoots, which is also in line with the smaller effects on gene expression mediated by the Pi growth regimen in roots (**Figure D1**). Finally, between organs the

comparisons with the highest similarities were found for low Pi. This indicates that *cis*-regulatory effects in the shoot are more highly dependent on the environmental condition than in the roots, and that stresses such as Pi starvation prompts an in part shared response in roots and shoots. However, we found no effect of the Pi growth regimen on the differences in gene expression patterns between roots and shoots. Regarding *trans*-regulatory effects, the situation is not so clear, in part because the lower numbers of *trans*-regulatory effects detected, but it is evident that roots share more *trans*-regulatory effects than shoots when comparing plants grown under high Pi vs low Pi conditions, and surprisingly in this case *trans*-regulatory effects between roots and shoots are more similar in plants grown under high Pi conditions. The possible significance of this latter finding awaits to be answered. Finally, comparisons of *cis*- vs *trans*-regulatory effects in our study show that there is a much higher similarity among *cis* effects than among *trans* effects in different developmental or environmental conditions, indicating the higher specificity of *trans*-regulatory effects. This finding is in agreement with previous studies (Cubillos *et al.*, 2014), in which *trans* regulation appeared to be preferentially involved in responses to changes in the environment (drought stress in their study), whereas *cis* variation was robust across environments. The higher robustness of *cis*- vs *trans*-regulatory effects could reflect the potentially higher pleiotropic effects of mutations at *trans*-acting regulators; their functional containment to a given tissue or to a given environmental condition would limit their detrimental effects.

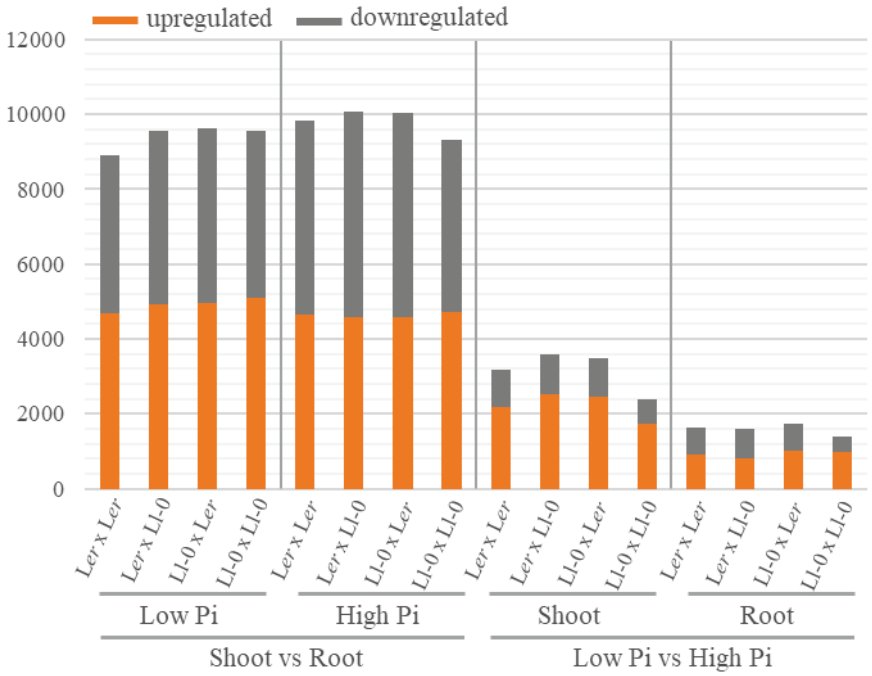


Figure D1. **Impact of the organ type and growth condition on the transcriptome of Arabidopsis**. Vertical bars indicate the number of genes differentially expressed in the comparisons of shoot vs root (orange and black show upregulation in shoot and root, respectively) and in low Pi vs high Pi (orange and black show upregulation in low Pi and high Pi, respectively). Differential gene expression was determined by applying p-adj<0.05 and fold change ±2.

5.1.3 – Maternal effects and genomic imprinting

The finding that reciprocal hybrids display very similar transcriptomic profiles indicates the low impact of maternal effects and genomic imprinting on gene expression in the conditions studied. Nevertheless, despite the modest differences, if *trans*-regulatory effects are examined, there was lesser number of *trans*-effects detected between parent and hybrids having the same maternal constitution than when the hybrid shared the paternal constitution (i.e., between L*er* and L*er* x Ll-0 and between Ll-0 and Ll-0 x L*er* than between L*er* and Ll-0 x L*er* and between Ll-0 and L*er* x Ll-0; **Figure R4**). Therefore, maternal effects/genomic imprinting had a minor but measurable effect on gene expression. In fact, in line with the contribution of these effects, some significant differences were found between the two hybrids for biomass accumulation traits (**Table R2**). Lack of maternal/genomic imprinting effects were found in a study of transcriptional variation between Col-0 and Cvi accessions in control and drought conditions by Cubillos et al. (2014) and it is in contrast with the observations made on the endosperm of Arabidopsis thaliana showing the importance of genomic imprinting in this tissue. The low impact of imprinting in our study and in Cubillos et al. (2014) in comparison to other studies (Scott *et al.*, 1998; Gehring, Missirian and Henikoff, 2011) could reflect that imprinting (and maternal effects) depend on the developmental stage, being more active at early stages (e.g., during seed development). This possibility for *Arabidopsis thaliana* however cannot be extrapolated to other species, as it was shown that maternal effects have strong impact on gene expression in *Arabidopsis lyrata* hybrids at a 12-week-old growth stage (Videvall *et al.*, 2016), and imprinting has been proposed to be a major determinant of gene expression in seedlings of maize hybrids (Swanson-Wagner *et al.*, 2009).

5.2 – *CIS*- AND *TRANS*-eQTLS USING HYBRIDS OF RILs AND PARENTS: CHARACTERISTICS, MODE OF INHERITANCE AND EVOLUTION

5.2.1 – *Cis*- and *trans*-eQTLs using hybrids of RILs: characteristics and mode of inheritance

The comparative transcriptomic analysis the L*er* and Ll-0 ecotypes and their reciprocal hybrids provided an estimation on *cis*- and *trans*-regulatory effects (with the contentions regarding the above-mentioned bias on mapping efficiency of reads of the reference vs the alternative allele). It has also allowed us to evaluate the environmental and organ specificity of *cis*- and *trans*-regulatory effects, as well as the impact of maternal effects and genomic imprinting on gene expression. However, it does not inform on the characteristics of the *trans*-regulatory effects (localisation of the *trans*-acting factors and their targets and their mode of inheritance) and on the

predominant evolutionary forces shaping the evolution of the gene expression. Our study involving the analysis of eQTLs in hybrids between RILs and their progenitors has provided relevant information on these issues. Specifically, our strategy consisted of two mapping populations originated from crosses between parental lines and RILs, using L*er* and Ll-0 as the female parents and the RILs as the male. This choice of combining RILs with the parental genotypes was based on the attenuation of sporophytic maternal effects, and thus a potential genotype-independent source of variability, as this effect would be similar within all hybrid lines in each of the two populations. The alternative strategy based on using crosses between RILs would have the advantage of a more common maternal genetic background across all the lines (L*er*) and a more homogeneous proportion of L*er*/Ll-0 genotypes, but it would suffer from stronger sporophytic maternal effects. By comparing gene expression levels of each gene in plants with the heterozygous for that gene we could estimate the impact of the genetic background on gene expression in our populations, and we found that it only has an effect on about a 25% of the genes analysed (approximately 5,000 of 19,000).

Our RIL-based analysis rendered the identification of approximately 2,200 significant *cis*-eQTLs and 19,300 *trans*-eQTLs. Compared to the experiment based on transcriptomic analyses of hybrids and parents, which led to the identification of almost 2,000 *cis* and 240 *trans*, considering the samples of root grown under low Pi, our eQTL analysis revealed a much higher number of regulatory differences for *trans*-eQTLs. The coincidences on *cis* effects detected between these experiments was approximately 33%, whereas this was notably higher for the *trans*-regulated genes (56-66%). The lower percentage of coincidences for *cis* than for *trans* effects is at least in part caused by the problems of estimation of *cis*-regulatory effects with ASE. Nevertheless, the absolute number of genes detected as *trans*-regulated through the comparison of hybrids and parents was forty times lower than through eQTL analysis. This is probably due to the overall balance in the genome of hybrids, in which as shown below there are many instances of inter-eQTL compensatory effects which would be uncoupled in the RILs, leading to transgressive segregation, thereby allowing detection of a much larger number of *trans*-acting variants. In addition, the 30 RILs analysed on each population provide approximately 15 pseudo-replicates for testing each chromosomal region for presence of eQTLs, which should increase the statistical significance for detection of *trans*-eQTLs of small magnitude in comparison with the analysis based on gene expression studies in hybrids and parents, in which just three replicas were examined per genotype.

5.2.2 – eQTL distribution and strength

The *cis*-eQTLs identified in this study are moderately uniformly distributed over the genome, being the number of *cis*-eQTLs in different chromosomes relatively similar (**Figure R10**). However, it is interesting that the relative proportion of *cis*-eQTLs with positive versus negative effects is less uniform and it may suggest that there are some evolutionary forces which favour/select *cis*-mutations acting in the same direction for proximal genes. In this context, it is worth mentioning that it has been shown that proximal genes tend display higher expression correlation than distantly located genes (Williams and Bowles, 2004). Also noteworthy is that we have found that in both L*er* x RIL and Ll-0 x RIL populations there is a higher number of *cis*-eQTLs with positive than with negative additive effects. The reasons for this finding await to be examined in detail. One possible explanation is that there are intragenic allelic interactions that cause increased expression of one or the two of these alleles. In this context, it has been shown that in animals it is quite widespread the transvection phenomenon, whereby in the heterozygous state one allele has been shown to be able to affect the expression of the other allele (Duncan, 2002; Mellert and Truman, 2012; Blick *et al.*, 2016).

As for *trans*-eQTLs, their distribution is much less uniform than that of the *cis*-eQTLs and indeed a large proportion of them can be clustered at specific chromosomal locations. In fact, here we identified up to 48 of such *trans*-eQTL hotspots. The notion that the distribution of *trans*-eQTLs versus *cis*-eQTLs is well-established in the vast literature on eQTL analysis and it rests on the fact that *trans*-eQTL hotspots represent single (or a few) genes regulating large cohorts of targets. As discussed below, the *trans*-eQTL hotspots detected in this study likely correspond to single regulators.

In addition to the differences in genome-wide distribution, another essential characteristic of *cis*- and *trans*-eQTLs is their different strength. Thus, more than 24% of the *cis*-eQTLs explained more than 50% of expression variance, whereas only a 2% of *trans*-eQTLs explained this level of variance. The higher strength of *cis*-eQTLs vs *trans*-eQTLs is not an exclusive characteristic of the genotypes and growth conditions of this study, but rather has been documented in many other studies (for instance, see Cubillos et al 2014). Underlying this difference could be the potentially higher pleiotropic effects of *trans*-eQTLs because they often affect a large number of targets (up to thousands).

5.2.3 – Functional-related characteristics of *cis*- and *trans*-eQTLs

One obvious question regarding eQTLs and their functional characteristics is whether they concern functionally related genes. Here we approached this question by examining their responsiveness to Pi starvation, as this was the growth condition for which we performed the eQTL analysis, and also GO term enrichment. Both *cis*- and *trans*-eQTLs displayed enrichment in Pi starvation responsive genes (**Figure R15**). This was a reasonable expectation for *trans*-eQTLs because they are controlled by a small number of genes, a situation that favours enrichment (or underrepresentation) of coexpressed genes. In addition, as mentioned above *trans*-eQTLs display higher specificity likely to restrict pleiotropic effects. However, it is quite remarkable that *cis*-eQTLs are enriched in PS-responsive to a level similar to that of *trans*-eQTLs, suggesting either that PS-responsive genes are more permissive for alteration of expression changes or that there is some directionality in the evolution of *cis*-eQTLs. Because PS-induced genes are enriched in both the *cis*-eQTLs groups with positive and negative additive values in both the L*er* x RIL and Ll-0 x RIL populations, the possibility of permissiveness for mutation cannot be ruled out. However, in case of PS-repressed genes, their enrichment is only in *cis*-eQTLs with negative additive value in the L*er* x RIL population and positive additive value in the Ll-0 x RIL population, indicating directionality in the selection force.

Regarding GO term enrichment, we found striking enrichment in *cis*-eQTLs, also suggestive of directional evolution. Indeed, *cis*-eQTLs upregulated in L*er* vs Ll-0 (with positive additive effect in the L*er* x RIL population and negative additive effect in the Ll-0 x RIL population) were enriched in GO terms related to cell rescue, disease, virulence, and defense (**Table R5**). This indicates that genes corresponding to these GO classes tend to be more highly expressed in one of the accessions, in contrast with the expectation under the neutral hypothesis of evolution or under the hypothesis of stabilizing selection where one would expect a similar proportion of this class of genes to be upregulated and downregulated in each of the accessions. Regarding *trans*-eQTLs, when analysed as a whole we also observed enrichment in several GO terms, although in this case GO enrichment pertains both types of additive effects, e.g., the ontology terms "ribosomal proteins" and "translation" are enriched in the set of *trans*-eQTLs with positive additive value in the L*er* x RIL population and in both sets of *trans*-eQTLs with positive and negative additive effects. In the section stabilizing and directional selection within *trans*-eQTLs and between *cis*- and *trans*-eQTLs we argue that this reflects predominance of stabilizing over directional selection concerning *trans*-eQTL interrelations.

5.2.4 – Characteristics of *trans*-eQTL hotspots

In this study we have identified 48 hotspot *trans*-eQTLs, focusing on the analysis of the 15 having the largest number of associated genes. Several lines of evidence indicate that they correspond to single regulators or very few regulators acting in concert. First, genes associated to each hotspot *trans*-eQTL display strong signs of expression coherence (i.e., they display a similar response to certain environmental developmental cues; **Figure R16**). In the alternative scenario, where *trans*-eQTL hotspots represent clusters of regulators acting on independent gene subsets, we would have not expected to find such similarities in the expression behaviour of their associated genes. In addition, we found that expression characteristics are shared among gene sets corresponding to different hotspots. Moreover, there are many instances of gene sharing among gene sets associated to different *trans*-eQTL hotspot (**Table R7**). As discussed below, these similarities among gene sets indicate events of stabilizing and directional selection among *trans*-eQTL hotspots (**5.2.6**). Second, reinforcing the idea of coherence within hotspot associated genes, we found that for nine out of the fifteen hotspots examined, their associated gene sets displayed enrichment for some ontology terms (**Table R5**). Of particular interest in the context of studies of the Pi starvation response is the case of hotspots IV, XI and XIV that displayed enrichment in the term "phosphate metabolic process", suggesting that they likely influence PSRs. Moreover, the gene sets associated to these 3 hotspots are among the 9 sets that are induced by Pi starvation (**Table D1**).

Table D1. **Characteristics of *trans*-eQTL hotspots relevant for prioritization as regulators of PSR**. For each of the major *trans*-eQTL hotspots (I-XV), it is indicated the enrichment (+) or not (-) of Pi starvation-induced (PSI) genes, on PSR-related GO terms, and in PHR1/PHL1 gene targets in at least one of their associated sets of up- or downregulated genes, as well as coincidences with QTLs affecting PSR traits. Hotspots XI and XIV are positive for three attributes (highlighted in yellow), and Hotspot IV was positive for all four attributes (highlighted in yellow and framed).

Hotspot	PSI gene enrichment	PSR-related GO term enrichment	PHR1/PHL1 target enrichment	colocalization with QTL x PSR traits
I	-	-	-	+
II	+	-	-	-
III	-	-	+	-
IV	+	+	+	+
V	-	-	-	+
VI	-	-	-	-
VII	-	-	-	-
VIII	+	-	-	-
IX	-	-	-	-
X	+	-	-	-
XI	+	+	+	-
XII	+	-	-	-
XIII	+	-	+	-
XIV	+	+	+	-
XV	+	-	-	-

Third, an additional proof of coherence within gene sets associated with hotspots is that in all fifteen hotspot gene sets there is enrichment in targets for several TFs. Despite hotspots likely correspond to a single regulatory gene (encoding a TF or other type of regulator, e.g., a kinase), given the fact that gene expression control in eukaryotes involves the concerted action of combinations of TFs, the control of hotspot-associated genes should not only involve the causal regulatory gene underlying the hotspot, but rather several TF acting in concert. Not surprisingly, in line with shared expression characteristics among gene sets of different hotspots and even sharing of their associated genes, there are TFs whose targets are enriched in several of the hotspot sets. As far as the PSR is concerned, it is interesting that five hotspots are enriched in targets of PHL1 and or PHR1, known to be the master regulators of PSR (Rubio *et al.*, 2001; Bustos *et al.*, 2010 ; **Table D1**). Included in these five sets are those corresponding to Hotspots IV, XI y XIV, which together with their enrichment in the PSR-related GO term "phosphate metabolic process" and in PSI genes, make these hotspots priority for identification of their underlying causal genes, due to their likely role in the control of key PSR-related gene sets. In addition, hotspot IV colocalizes with a QTL found to affect root-to-shoot weight ratio in plants grown under low Pi (**Table R8**). Although the effect of this QTL on root-to-shoot weight ratio also accurs in the context of growth under high Pi, the colocalization of hotspot IV with an QTL affecting a trait related to growth under low Pi makes this hotspot a first priority for molecular identification.

Fourth, one additional output of the study of TF target enrichment in the gene sets associated to hotspots is that it may also provide candidates for causal genes underlying these hotspots. Indeed, TFs whose targets are enriched and that physically colocalise with the hotspot have increased chances to be the causal genes. In total we found that in seven cases TFs whose target are enriched colocalize with the corresponding hotspot (**Table D2**). One characteristic that would increase the likelihood of TFs as candidates to correspond to a hotspot *trans*-eQTL regulator would be that the TF has an associated *cis*-eQTL. We found one case in which a *cis*-eQTL was associated to the AT5G07680 transcription factor gene only in the same P x RIL population in which the hotspot eQTL has been detected. Experiments are underway to analyse the phenotype of the T-DNA mutant of this TF. An additional possible manner for shortlisting TF candidate lists is searching for cases in which the L*er* and Ll-0 alleles of the TF display differences in the amino acid sequence potentially affecting protein activity. Out of the 18 TF candidates colocalizing with hotspots in this study, we found that six of them displayed differences in their amino acid sequence between the alleles L*er* and Ll-0, which could result in differences in the activity of the two alleles (**Table D2**). In the case of PHR1, a proline residue in the L*er* allele is changed to Leucine in Ll-0.

Table D2. ***Trans*-eQTL hotspot colocalization with potential TF regulators**. It is shown the cases where a TF whose targets are enriched in the associated genes of a given *trans*-eQTL hotspot colocalizes with the position of that hotspot. Also indicated is whether the TFs colocalizing with any hotspot are associated to significant *cis*-eQTLs (only if the *cis*-eQTL was detected in the same population as the hotspot) as well as the allelic differences (L*er*→Ll-0) found in the protein sequence of TFs. It is highlighted in bold and underlined those amino acid changes that may significantly alter protein activity.

Hotspot	colocalization with potential TF regulator	TF name	TF with *cis*-eQTL	allelic difference in TF protein sequence
I	+	VIP1	-	VIP1(**F→S**; L→I)
II	-	-	-	-
III	-	-	-	-
IV	+	WRKY22, NAC69	-	NAC69(**N→K**)
V	-	-	-	-
VI	-	-	-	-
VII	-	-	-	-
VIII	-	-	-	-
IX	-	-	-	-
X	+	ERF039, ANAC071, WRKY28	-	ERF039(**T→S**) ANAC071(**L→R**; **N→H**; **Y→C**; D→E; **H→D**)
XI	+	PHR1	-	PHR1(**P→L**)
XII	+	GATA19, DOF4.7	-	-
XIII	+	NAC079, TGA5, TGA10, ERF115, EICBP.B	NAC079	ERF115(L→M) EICBP.B(A→G)
XIV	+	bZIP3, ANAC087, MYB56	-	bZIP3(I→V) MYB56(**T→P**; **I→F**)
XV	+	AT5G56840(MYB-related)	-	-

5.2.5 – Mode of inheritance of *cis*- and *trans*-eQTLs

Because most eQTL studies in plants were performed with homozygous plants, little is known about the mode of inheritance (i.e., codominant, dominant, and over/underdominant). A rigorous analysis of the mode of inheritance would have required ANOVA tests to evaluate the differences among the two homozygous and the heterozygous genotypes for each *cis*- and *trans*-eQTL, which is quite a demanding task. In this study we followed a simpler "qualitative" approach to get estimations of cases of codominant, dominant, and over/underdominant inheritance (**Figure R17**). Using this qualitative approach, we estimated that *cis*-eQTLs show a predominant codominant inheritance (46-52% vs 20-21% dominant and 0% overdominant), whereas *trans*-eQTLs display a predominant dominant inheritance (62-75% vs 2-3% codominant and 0.1-0.2% overdominant). Predominance of dominance over codominance and overdominance has also been observed for QTLs affecting performance traits (see section **4.2.4.1**). It is remarkable the predominance of dominant inheritance for *trans*-eQTLs. This likely indicates a non-linear relation between gene activity and output, and that the activity of dominant allele is the asymptotic part of the dose-response curve. Underlying this behaviour could be the fact that most regulatory

mechanisms in complex biological systems are complex and robust, making them potentially insensitive to moderate changes in the activity of one of their components. On the other hand, the very few cases of overdominance, are in line with the current information on the basis of heterosis. In fact, many studies on this phenomenon have already highlighted the low contribution of overdominance to heterosis, and the importance of additive and epistatic effects as well epigenetic mechanisms to explain hybrid vigour (Syed and Chen, 2005; Kusterer, Muminovic, *et al.*, 2007; Melchinger *et al.*, 2007; Meyer *et al.*, 2010; Shi *et al.*, 2011). As for *cis*-eQTLs, while codominant is the predominant mode of inheritance of these eQTLs, we find striking the still important proportion of *cis*-eQTLs following a dominant mode of inheritance. This could in part reflect limitations of the eQTL mapping, which assumes genetic effects only on additive effects in their detection, although instances of interactions among intragenic alleles (e.g., transvection) could contribute to explain at least part of the dominant inheritance.

5.2.6 – Contribution of stabilizing and directional selection in shaping the evolution of the Pi starvation transcriptome

A key question in evolutionary biology is what are the evolutionary forces that determine natural variation of physiological and molecular traits. Regarding evolution of transcription there is considerable evidence in studies in Drosophila and mammals pointing to stabilizing selection to be the most prominent force shaping variation on gene expression (Signor and Nuzhdin, 2018). In plants there are few studies approaching this question. Cubillos et al. (2014) found that compensatory effects between *cis* and *trans*-regulatory effects (i.e., stabilizing selection) was equally frequent than convergent *cis*/*trans* effects (i.e., effects on the same direction underlying directional selection). In another study in conifer trees there was found a large extent of *cis*/*trans* compensation (Verta, Landry and Mackay, 2016). In our study we found that stabilizing selection was the predominant evolutionary force over directional selection for *trans*-eQTLs ($p<<0{,}00001$; i.e., out of the genes having more than one associated *trans*-eQTL, there were 2.2-fold more cases of genes in which their *trans*-eQTLs acted in the opposite direction –compensatory– than cases of genes in which all their *trans*-eQTLs acted in the same direction; **Figure R14**). In sharp contrast, in case of *cis*- and *trans*-eQTLs acting on the same gene, directional selection was more prominent than stabilizing selection as there was a 1,7-fold excess of cases in which *cis*- and *trans*-eQTLs corresponding to the same gene acted in the same direction relative to the cases in which they acted in opposite directions (**Figure R14**, $p<<0{,}00001$). In agreement with the importance of directional selection for *cis*-eQTLs, is the finding mentioned in a previous section that there was an enrichment on the ontology term “disease virulence and defense” in *cis*-eQTLs upregulated in L*er* vs Ll-0 and enrichment in Pi starvation repressed genes in *cis*-eQTLs downregulated in L*er* vs Ll-0. These findings probably reflect the higher evolvability of *cis*-

eQTLs, which act on a single gene, over *trans*-eQTLs, often affecting the expression of many genes with the concomitant higher probability of causing detrimental effects.

5.2.7 – Correlations between coexpression modules and plant performance and physiological traits

In parallel with the analysis of eQTLs, in this study we have also analysed QTLs for traits related to plant performance and physiology, finding 40 and 46 QTLs for traits in plants grown in low Pi and high Pi, which respectively mapped at 32 and 31 different locations. Surprisingly, however, we found just three cases of colocalization of QTLs for plant performance in plants grown under low Pi conditions and the major fifteen *trans*-eQTLs detected in the same study. This suggests that for many QTLs the underlying causal gene could be structural (or regulator of a short number of targets).

The availability of transcriptomic and physiological traits datasets for 60 plant genotypes allowed us to use a different approach to connect transcriptional and physiological data, based on the construction of a gene coexpression module network and to examine coexpression module-trait correlations. The construction of the regulatory gene network was made independently for Pi starvation induced and repressed genes, which were grouped in 12 and 10 expression modules. TF target enrichment on these modules allowed the identification of potential regulators of these modules as well as the possible effect of these TFs. Thus, the filtering of TF targets identified following *in vitro* approach (DAP-seq; O'Malley *et al.*, 2016) based on the high correlation and anticorrelation of their expression with that of the TF, distinguished between potentially activated targets (highly correlated expression) and potentially repressed targets. The examination of the network revealed that several TFs display a canonical behaviour, i.e., that all coexpresion modules of the same type, induced or repressed, potentially regulated by a given TF displayed enrichment of the same list of targets, highly correlated or highly anticorrelated, whereas coexpression modules of different type (repressed or induced, respectively) displayed enrichment of a different target list. However other TFs do not follow this logic. Besides the fact that bioinformatics analyses need experimental confirmation, one possible explanation for the observation of TFs displaying non canonical-behaviour is that the control of gene expression is dynamic and may follow different rules at different stages of growth under stress. While in initial phases of the stress there is a large transcriptional reprograming so that some genes are induced and others repressed, at later stages of the stress the rules determining coexpression of genes could change.

Regarding coexpresion module-trait correlations several modules showed correlations with PSR-related traits which were highly statistically significant. These correlations raise the possibility that regulators of these modules can also regulate the associated traits. It is of note that different

to the QTL approach in which only genetic effects are computed, in this approach environmental effects are also accounted for. Therefore, building gene coexpresion module networks extends the potential value of combined transcriptomic and phenotypic analysis of accessions or of their progenies beyond classical natural variation-based approaches.

6. CONCLUSIONS

The findings of this study can be summarised in the following conclusions:

I. There are large gene expression differences between the Arabidopsis accessions L*er* and Ll-0, determined by the regulatory activity effects of 2,200 *cis*- and 19,000 *trans*-eQTLs.

II. Genomic imprinting and maternal effects associated to non-mendelian inheritance have a small, although still significant, impact on gene expression variation.

III. Developmental cues (determinants of the organ type) have a higher impact than environmental cues (e.g., nutritional regimen) in the manifestation of *cis*- and *trans*-regulatory effects.

IV. *cis*-eQTLs have a moderately uniform distribution across the genome, while most of the *trans*-eQTLs tend to cluster in hotspots.

V. The fifteen major *trans*-eQTL hotspots studied show signs of intrinsic coherence in their associated genes based on their similarities in expression behaviour and in functional characteristics, and on the enrichment in specific transcription factor gene targets.

VI. The characteristics of three *trans*-eQTL point to their fundamental role in the regulation of the phosphate starvation response.

VII. Seven transcription factors display characteristics to be candidates to correspond to the underlying causal gene of seven of the fifteen major *trans*-eQTL hotspots identified in this study.

VIII. The predominant modes of inheritance for *cis*- and *trans*-eQTLs are codominance and dominance, respectively.

IX. Stabilizing selection has a predominant role in the evolution of *trans*-eQTLs, while directional selection in that of *cis*-eQTLs. This suggests the higher evolvability of *cis*-eQTLs, in line with their more limited impact in gene expression.

X. Phosphate starvation-induced and -repressed genes in Arabidopsis roots group in 12 and 10 gene coexpression modules, some of which show potentially biologically meaningful correlations with plant performance and physiology traits related to the phosphate starvation response. This raises the possibility to identify genes of agronomical interest using approaches to analyse gene regulatory networks.

7. BIBLIOGRAPHY

Abel, S. (2017) 'Phosphate scouting by root tips', *Current Opinion in Plant Biology*, 39, pp. 168–177. doi: 10.1016/j.pbi.2017.04.016.

Al-Ghazi, Y. *et al.* (2003) 'Temporal responses of Arabidopsis root architecture to phosphate starvation: Evidence for the involvement of auxin signalling', *Plant, Cell and Environment*, 26(7), pp. 1053–1066. doi: 10.1046/j.1365-3040.2003.01030.x.

Al-Shehbaz, I. A. and O'Kane, S. L. (2002) 'Taxonomy and Phylogeny of Arabidopsis (Brassicaceae)', *The Arabidopsis Book*, 1(1821), p. e0001. doi: 10.1199/tab.0001.

Albert, F. W. and Kruglyak, L. (2015) 'The role of regulatory variation in complex traits and disease', *Nature Reviews Genetics*, 16(4), pp. 197–212. doi: 10.1038/nrg3891.

Alonso-Blanco, C. *et al.* (1999) 'Natural allelic variation at seed size loci in relation to other life history traits of Arabidopsis thaliana', *Proceedings of the National Academy of Sciences of the United States of America*, 96(8), pp. 4710–4717. doi: 10.1073/pnas.96.8.4710.

Alonso-Blanco, C. *et al.* (2009) 'What has natural variation taught us about plant development, physiology, and adaptation?', *Plant Cell*, 21(7), pp. 1877–1896. doi: 10.1105/tpc.109.068114.

Alonso-Blanco, C. *et al.* (2016) '1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana', *Cell*, 166(2), pp. 481–491. doi: 10.1016/j.cell.2016.05.063.

Alonso-Blanco, C. and Koornneef, M. (2000) 'Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics', *Trends in Plant Science*, 5, pp. 22–29.

Alonso-Blanco, C., Koornneef, M. and van Ooijen, J. W. (2006) 'QTL analysis.', *Methods in molecular biology (Clifton, N.J.)*, 323, pp. 79–99. doi: 10.1385/1-59745-003-0:79.

Alonso-Blanco, C. and Méndez-Vigo, B. (2014) 'Genetic architecture of naturally occurring quantitative traits in plants: An updated synthesis', *Current Opinion in Plant Biology*, 18(1), pp. 37–43. doi: 10.1016/j.pbi.2014.01.002.

Ames, B. N. (1966) 'Assay of inorganic phosphate, total phosphate and phosphatases', in, pp. 115–118. doi: 10.1016/0076-6879(66)08014-5.

Atwell, S. *et al.* (2010) 'Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines', *Nature*, 465(7298), pp. 627–631. doi: 10.1038/nature08800.

Aung, K. *et al.* (2006) 'pho2, a phosphate overaccumulator, is caused by a nonsense mutation in a microRNA399 target gene', *Plant Physiology*, 141(3), pp. 1000–1011. doi: 10.1104/pp.106.078063.

Azevedo, C. and Saiardi, A. (2017) 'Eukaryotic Phosphate Homeostasis: The Inositol

Pyrophosphate Perspective', *Trends in Biochemical Sciences*, 42(3), pp. 219–231. doi: 10.1016/j.tibs.2016.10.008.

Bader, D. M. *et al.* (2015) 'Negative feedback buffers effects of regulatory variants', *Molecular Systems Biology*, 11(1), p. 785. doi: 10.15252/msb.20145844.

Baek, D. *et al.* (2017) 'Cross-talk between phosphate starvation and other environmental stress signaling pathways in plants', *Molecules and Cells*, 40(10), pp. 697–705. doi: 10.14348/molcells.2017.0192.

Balzergue, C. *et al.* (2017) 'Low phosphate activates STOP1-ALMT1 to rapidly inhibit root cell elongation', *Nature Communications*, 8(May). doi: 10.1038/ncomms15300.

Bari, R., Datt Pant, B., *et al.* (2006) 'PHO2, MicroRNA399, and PHR1 Define a Phosphate-Signaling Pathway in Plants', *Plant Physiology*, 141(3), pp. 988–999. doi: 10.1104/pp.106.079707.

Bari, R., Pant, B. D., *et al.* (2006) 'Phosphate-Signaling Pathway in Plants', *Society*, 141(July), pp. 988–999. doi: 10.1104/pp.106.079707.988.

Bariola, P. A. *et al.* (1994) 'The Arabidopsis ribonuclease gene RNS1 is tightly controlled in response to phosphate limitation', *The Plant Journal*, 6(5), pp. 673–685. doi: 10.1046/j.1365-313X.1994.6050673.x.

Barth, S. *et al.* (2003) 'Heterosis for biomass yield and related traits in five hybrids of Arabidopsis thaliana L. Heynh', *Heredity*, 91(1), pp. 36–42. doi: 10.1038/sj.hdy.6800276.

Bates, T. and Lynch, J. (1996) 'Stimulation of root hair elongation in Arabidopsis thaliana by low phosphorus availability', *Plant, Cell and Environment*, 19(5), pp. 529–538. doi: 10.1111/j.1365-3040.1996.tb00386.x.

Bazin, J. *et al.* (2017) 'Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation', *Proceedings of the National Academy of Sciences of the United States of America*, 114(46), pp. E10018–E10027. doi: 10.1073/pnas.1708433114.

Benjamini, Y. and Hochberg, Y. (2000) 'On the adaptive control of the false discovery rate in multiple testing with independent statistics', *Journal of Educational and Behavioral Statistics*, 25(1), pp. 60–83.

Birchler, J. A. *et al.* (2005) 'Dosage balance in gene regulation: Biological implications', *Trends in Genetics*, 21(4), pp. 219–226. doi: 10.1016/j.tig.2005.02.010.

Birchler, J. A. *et al.* (2010) 'Heterosis', *Plant Cell*, 22(7), pp. 2105–2112. doi: 10.1105/tpc.110.076133.

Birchler, J. A. and Veitia, R. A. (2007) 'The gene balance hypothesis: From classical genetics to modern genomics', *Plant Cell*, 19(2), pp. 395–402. doi: 10.1105/tpc.106.049338.

Birchler, J. A., Yao, H. and Chudalayandi, S. (2006) 'Unraveling the genetic basis of hybrid vigor', *Proceedings of the National Academy of Sciences of the United States of America*, 103(35), pp. 12957–12958. doi: 10.1073/pnas.0605627103.

Blick, A. J. *et al.* (2016) 'The capacity to act in trans varies among drosophila enhancers', *Genetics*, 203(1), pp. 203–218. doi: 10.1534/genetics.115.185645.

Bolnick, D. I. *et al.* (2011) 'Why intraspecific trait variation matters in community ecology', *Trends in Ecology and Evolution*, 26(4), pp. 183–192. doi: 10.1016/j.tree.2011.01.009.

Borch, K. *et al.* (1999) 'Ethylene: A regulator of root architectural responses to soil phosphorus availability', *Plant, Cell and Environment*, 22(4), pp. 425–431. doi: 10.1046/j.1365-3040.1999.00405.x.

Botet, R. and Keurentjes, J. J. B. (2020) 'The Role of Transcriptional Regulation in Hybrid Vigor', *Frontiers in Plant Science*, 11(April), pp. 1–9. doi: 10.3389/fpls.2020.00410.

Bourguet, D. (1999) 'The evolution of dominance', *Heredity*, 83, pp. 1–4. doi: 10.1007/BF02127451.

Box, G. E. P. and Cox, D. R. (1964) 'An analysis of transformations', *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), pp. 211–252. doi: 10.1111/j.2517-6161.1964.tb00553.x.

Bucher, M. (2007) 'Functional biology of plant phosphate uptake at root and mycorrhiza interfaces', *New Phytologist*, 173(1), pp. 11–26. doi: 10.1111/j.1469-8137.2006.01935.x.

Buckler IV, E. S. and Thornsberry, J. M. (2002) 'Plant molecular diversity and applications to genomics', *Current Opinion in Plant Biology*, 5(2), pp. 107–111. doi: 10.1016/S1369-5266(02)00238-8.

Bustos, R. *et al.* (2010) 'A central regulatory system largely controls transcriptional activation and repression responses to phosphate starvation in arabidopsis', *PLoS Genetics*, 6(9). doi: 10.1371/journal.pgen.1001102.

Cardona-López, X. *et al.* (2015) 'ESCRT-III-associated protein ALIX mediates high-affinity phosphate transporter trafficking to maintain phosphate homeostasis in arabidopsis', *Plant Cell*, 27(9), pp. 2560–2581. doi: 10.1105/tpc.15.00393.

Carroll, S. B. (2008) 'Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution', *Cell*, 134(1), pp. 25–36. doi: 10.1016/j.cell.2008.06.030.

Castrillo, G. *et al.* (2017) 'Root microbiota drive direct integration of phosphate stress and immunity', *Nature*, 543(7646), pp. 513–518. doi: 10.1038/nature21417.

Chandrika, N. N. P. *et al.* (2013) 'ALFIN-LIKE 6 is involved in root hair elongation during phosphate deficiency in Arabidopsis', *New Phytologist*, 198(3), pp. 709–720. doi: 10.1111/nph.12194.

Chapin, L. J. and Jones, M. L. (2009) 'Ethylene regulates phosphorus remobilization and expression of a phosphate transporter (PhPT1) during petunia corolla senescence', *Journal of Experimental Botany*, 60(7), pp. 2179–2190. doi: 10.1093/jxb/erp092.

Charlesworth, B. (2013) 'Stabilizing selection, purifying selection, and mutational bias in finite populations', *Genetics*, 194(4), pp. 955–971. doi: 10.1534/genetics.113.151555.

Chen, J. *et al.* (2015) 'The rice CK2 kinase regulates trafficking of phosphate transporters in response to phosphate levels', *Plant Cell*, 27(3), pp. 711–723. doi: 10.1105/tpc.114.135335.

Chen, S. *et al.* (2018) 'fastp: an ultra-fast all-in-one FASTQ preprocessor', *Bioinformatics*, 34(17), pp. i884–i890. doi: 10.1093/bioinformatics/bty560.

Chen, Z. J. (2013) 'Genomic and epigenetic insights into the molecular bases of heterosis', *Nature Reviews Genetics*, 14(7), pp. 471–482. doi: 10.1038/nrg3503.

Chevalier, F. *et al.* (2003) 'Effects of phosphate availability on the root system architecture: Large-scale analysis of the natural variation between Arabidopsis accessions', *Plant, Cell and Environment*, 26(11), pp. 1839–1850. doi: 10.1046/j.1365-3040.2003.01100.x.

Chiou, T. J. *et al.* (2006) 'Regulation of phosphate homeostasis by MicroRNA in Arabidopsis', *Plant Cell*, 18(2), pp. 412–421. doi: 10.1105/tpc.105.038943.

Chiou, T. J. and Lin, S. I. (2011) 'Signaling network in sensing phosphate availability in plants', *Annual Review of Plant Biology*, 62, pp. 185–206. doi: 10.1146/annurev-arplant-042110-103849.

Churchill, G. A. and Doerge, R. W. (1994) 'Empirical threshold values for quantitative trait mapping', *Genetics*, 138(3), pp. 963–971. doi: 10.1093/genetics/138.3.963.

Ciereszko, I., Johansson, H. and Kleczkowski, L. A. (2005) 'Interactive effects of phosphate deficiency, sucrose and light/dark conditions on gene expression of UDP-glucose pyrophosphorylase in Arabidopsis', *Journal of Plant Physiology*, 162(3), pp. 343–353. doi: 10.1016/j.jplph.2004.08.003.

Ciereszko, I. and Kleczkowski, L. A. (2002) 'Effects of phosphate deficiency and sugars on expression of rab18 in Arabidopsis: Hexokinase-dependent and okadaic acid-sensitive

transduction of the sugar signal', *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1579(1), pp. 43–49. doi: 10.1016/S0167-4781(02)00502-X.

Colautti, R. I., Lee, C. R. and Mitchell-Olds, T. (2012) 'Origin, fate, and architecture of ecologically relevant genetic variation', *Current Opinion in Plant Biology*, 15(2), pp. 199–204. doi: 10.1016/j.pbi.2012.01.016.

Coolon, J. D., Mcmanus, C. J. and Stevenson, K. R. (2014) 'Tempo and mode of regulatory evolution in Drosophila Department of Ecology and Evolutionary Biology', pp. 797–808. doi: 10.1101/gr.163014.113.Freely.

Cubillos, F. A. *et al.* (2012) 'Expression variation in connected recombinant populations of Arabidopsis thaliana highlights distinct transcriptome architectures', *BMC Genomics*, 13(1), p. 117. doi: 10.1186/1471-2164-13-117.

Cubillos, F. A. *et al.* (2014) 'Extensive Cis-Regulatory variation robust to environmental perturbation in arabidopsis', *Plant Cell*, 26(11), pp. 4298–4310. doi: 10.1105/tpc.114.130310.

Dapp, M. *et al.* (2015) 'Heterosis and inbreeding depression of epigenetic Arabidopsis hybrids', *Nature Plants*, 1(July), pp. 1–8. doi: 10.1038/nplants.2015.92.

Darwin, C. R. (1859) *On the origin of species*. John Murray.

Darwin, C. R. (1876) *The effects of cross and self fertilisation in the vegetable kingdom*. John Murray. doi: 10.5962/bhl.title.51079.

Davenport, C. B. (1908) 'Degeneration, albinism and inbreeding', *Science*, 28(718), pp. 454–455.

Doebley, J. F., Gaut, B. S. and Smith, B. D. (2006) 'The Molecular Genetics of Crop Domestication', *Cell*, 127, pp. 1309–1321. doi: 10.1080/03081079.2015.1106730.

Doerge, R. W. and Churchill, G. A. (1996) 'Permutation tests for multiple loci affecting a quantitative character.', *Genetics*, 142(1), pp. 285–94. doi: 10.1111/j.1369-7625.2010.00632.x.

Dong, J. *et al.* (2019) 'Inositol Pyrophosphate InsP8 Acts as an Intracellular Phosphate Signal in Arabidopsis', *Molecular Plant*, 12(11), pp. 1463–1473. doi: 10.1016/j.molp.2019.08.002.

Duncan, I. W. (2002) 'Transvection effects in Drosophila', *Annual Review of Genetics*, 36, pp. 521–556. doi: 10.1146/annurev.genet.36.060402.100441.

East, E. M. (1908) 'Inbreeding in corn', *Connecticut Agricultural Experiment Station Report*, pp. 419–428.

El-Soda, M. *et al.* (2019) 'QTL and candidate genes associated with leaf anion concentrations in

response to phosphate supply in Arabidopsis thaliana', *BMC Plant Biology*, 19(1), pp. 1–14. doi: 10.1186/s12870-019-1996-3.

Emerson, J. J. *et al.* (2010) 'Natural selection on cis and trans regulation in yeasts', *Genome Research*, 20(6), pp. 826–836. doi: 10.1101/gr.101576.109.

Emerson, J. J. and Li, W. H. (2010) 'The genetic basis of evolutionary change in gene expression levels', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1552), pp. 2581–2590. doi: 10.1098/rstb.2010.0005.

Essigmann, B. *et al.* (1998) 'Phosphate availability affects the thylakoid lipid composition and the expression of SQD1 a gene required for sulfolipid biosynthesis in Arabidopsis thaliana', *Proceedings of the National Academy of Sciences of the United States of America*, 95(4), pp. 1950–1955. doi: 10.1073/pnas.95.4.1950.

Fear, J. M. *et al.* (2016) 'Buffering of genetic regulatory networks in Drosophila melanogaster', *Genetics*, 203(3), pp. 1177–1190. doi: 10.1534/genetics.116.188797.

Franco-Zorrilla, J. M. *et al.* (2002) 'Mutations at CRE1 impair cytokinin-induced repression of phosphate starvation responses in Arabidopsis', *Plant Journal*, 32(3), pp. 353–360. doi: 10.1046/j.1365-313X.2002.01431.x.

Franco-Zorrilla, J. M. *et al.* (2004) 'The transcriptional control of plant responses to phosphate limitation', *Journal of Experimental Botany*, 55(396), pp. 285–293. doi: 10.1093/jxb/erh009.

Franco-Zorrilla, J. M. *et al.* (2005) 'Interaction between phosphate-starvation, sugar, and cytokinin signaling in Arabidopsis and the roles of cytokinin receptors CRE1/AHK4 and AHK3', *Plant Physiology*, 138(2), pp. 847–857. doi: 10.1104/pp.105.060517.

Franco-Zorrilla, J. M. *et al.* (2007) 'Target mimicry provides a new mechanism for regulation of microRNA activity', *Nature Genetics*, 39(8), pp. 1033–1037. doi: 10.1038/ng2079.

Fraser, H. B. *et al.* (2011) 'Systematic detection of polygenic cis-regulatory evolution', *PLoS Genetics*, 7(3). doi: 10.1371/journal.pgen.1002023.

Fujii, H. *et al.* (2005) 'A miRNA involved in phosphate-starvation response in Arabidopsis', *Current Biology*, 15(22), pp. 2038–2043. doi: 10.1016/j.cub.2005.10.016.

Fujimoto, R. *et al.* (2012) 'Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity', *Proceedings of the National Academy of Sciences of the United States of America*, 109(18), pp. 7109–7114. doi: 10.1073/pnas.1204464109.

Fujimoto, R. *et al.* (2018) 'Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems', *Breeding Science*, 68(2), pp. 145–158. doi:

10.1270/jsbbs.17155.

Fujioka, M. *et al.* (2016) 'Determinants of chromosome architecture : insulator pairing in cis and in trans', 12(2), pp. 1–25. doi: 10.1371/journal.pgen.1005889.

Gamuyao, R. *et al.* (2012) 'The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency', *Nature*, 488(7412), pp. 535–539. doi: 10.1038/nature11346.

Gaude, N. *et al.* (2008) 'Phospholipase C5 (NPC5) is involved in galactolipid accumulation during phosphate limitation in leaves of Arabidopsis', *Plant Journal*, 56(1), pp. 28–39. doi: 10.1111/j.1365-313X.2008.03582.x.

Gaur, U. *et al.* (2013) 'Research progress in allele-specific expression and its regulatory mechanisms', *Journal of Applied Genetics*, 54(3), pp. 271–283. doi: 10.1007/s13353-013-0148-y.

Gehring, M., Missirian, V. and Henikoff, S. (2011) 'Genomic analysis of parent-of-origin allelic expression in arabidopsis thaliana seeds', *PLoS ONE*, 6(8). doi: 10.1371/journal.pone.0023687.

Ghandilyan, A. *et al.* (2009) 'A strong effect of growth medium and organ type on the identification of QTLs for phytate and mineral concentrations in three arabidopsis thaliana RIL populations', *Journal of Experimental Botany*, 60(5), pp. 1409–1425. doi: 10.1093/jxb/erp084.

Gilbert, G. A. *et al.* (2000) 'Proteoid root development of phosphorus deficient lupin is mimicked by auxin and phosphonate', *Annals of Botany*, 85(6), pp. 921–928. doi: 10.1006/anbo.2000.1133.

Goff, S. A. (2011) 'A unifying theory for general multigenic heterosis: Energy efficiency, protein metabolism, and implications for molecular breeding', *New Phytologist*, 189(4), pp. 923–937. doi: 10.1111/j.1469-8137.2010.03574.x.

Goncalves, A. *et al.* (2012) 'Extensive compensatory cis-trans regulation in the evolution of mouse gene expression', *Genome Research*, 22(12), pp. 2376–2384. doi: 10.1101/gr.142281.112.

González, E. *et al.* (2005) 'PHOSPHATE TRANSPORTER TRAFFIC FACILITATOR1 is a plant-specific SEC12-related protein that enables the endoplasmic reticulum exit of a high-affinity phosphate transporter in Arabidopsis', *Plant Cell*, 17(12), pp. 3500–3512. doi: 10.1105/tpc.105.036640.

Gordon, K. L. and Ruvinsky, I. (2012) 'Tempo and mode in evolution of transcriptional regulation', *PLoS Genetics*, 8(1), pp. 1–11. doi: 10.1371/journal.pgen.1002432.

Govindaraju, D. R. (2019) 'An elucidation of over a century old enigma in genetics—

Heterosis', *PLOS Biology*, 17(4), p. e3000215. doi: 10.1371/journal.pbio.3000215.

Grant, V. (1971) *Plant Speciation*. Columbia University Press.

Groszmann, M. *et al.* (2013) 'The role of epigenetics in hybrid vigour', *Trends in Genetics*, 29(12), pp. 684–690. doi: 10.1016/j.tig.2013.07.004.

Guo, B. *et al.* (2008) 'Functional analysis of the Arabidopsis PHT4 family of intracellular phosphate transporters', *New Phytologist*, 177(4), pp. 889–898. doi: 10.1111/j.1469-8137.2007.02331.x.

Hamburger, D. *et al.* (2002) 'Identification and characterization of the Arabidopsis PHO1 gene involved in phosphate loading to the xylem', *Plant Cell*, 14(4), pp. 889–902. doi: 10.1105/tpc.000745.

Hammond, J. P. *et al.* (2003) 'Changes in gene expression in Arabidopsis shoots during phosphate starvation and the potential for developing smart plants', *Plant Physiology*, 132(2), pp. 578–596. doi: 10.1104/pp.103.020941.

Hammond, J. P. and White, P. J. (2008) 'Sucrose transport in the phloem: Integrating root responses to phosphorus starvation', *Journal of Experimental Botany*, 59(1), pp. 93–109. doi: 10.1093/jxb/erm221.

Harrison, P. W., Wright, A. E. and Mank, J. E. (2012) 'The evolution of gene expression and the transcriptome-phenotype relationship', *Seminars in Cell and Developmental Biology*, 23(2), pp. 222–229. doi: 10.1016/j.semcdb.2011.12.004.

Hartanto, M. *et al.* (2020) 'Network Analysis Prioritizes DEWAX and ICE1 as the Candidate Genes for Major eQTL Hotspots in Seed Germination of Arabidopsis thaliana', *G3 (Bethesda, Md.)*, 10(11), pp. 4215–4226. doi: 10.1534/g3.120.401477.

He, X. *et al.* (2016) 'Epistatic partners of neurogenic genes modulate Drosophila olfactory behavior', *Genes, Brain and Behavior*, 15(2), pp. 280–290. doi: 10.1111/gbb.12279.

He, Z. *et al.* (2005) 'Assessment of inequality of root hair density in Arabidopsis thaliana using the Gini coefficient: A close look at the effect of phosphorus and its interaction with ethylene', *Annals of Botany*, 95(2), pp. 287–293. doi: 10.1093/aob/mci024.

Hedrick, P. W., Hellsten, U. and Grattapaglia, D. (2016) 'Examining the cause of high inbreeding depression: Analysis of whole-genome sequence data in 28 selfed progeny of Eucalyptus grandis', *New Phytologist*, 209(2), pp. 600–611. doi: 10.1111/nph.13639.

Henderson, I. R. and Salt, D. E. (2017) 'Natural genetic variation and hybridization in plants', *Journal of Experimental Botany*, 68(20), pp. 5415–5417. doi: 10.1093/jxb/erx377.

Hodgins-Davis, A. *et al.* (2015) 'Gene expression evolves under a house-of-cards model of stabilizing selection', *Molecular Biology and Evolution*, 32(8), pp. 2130–2140. doi: 10.1093/molbev/msv094.

Hoekstra, H. E. and Coyne, J. A. (2007) 'The locus of evolution: Evo devo and the genetics of adaptation', *Evolution*, 61(5), pp. 995–1016. doi: 10.1111/j.1558-5646.2007.00105.x.

Hoffmann, A. A. and Willi, Y. (2008) 'Detecting genetic responses to environmental change', *Nature Reviews Genetics*, 9(6), pp. 421–432. doi: 10.1038/nrg2339.

Hoffmann, M. H. (2002) 'Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae)', *Journal of Biogeography*, 29(1), pp. 125–134. doi: 10.1046/j.1365-2699.2002.00647.x.

Hollick, J. B. and Chandler, V. L. (1998) 'Epigenetic allelic states of a maize transcriptional regulatory locus exhibit overdominant gene action', *Genetics*, 150(2), pp. 891–897.

Horton, M. W. *et al.* (2012) 'Arabidopsis thaliana accessions from the RegMap panel', *Nature Genetics*, 44(2), pp. 212–216. doi: 10.1038/ng.1042.Genome-wide.

Hruz, T. *et al.* (2008) 'Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes', *Advances in Bioinformatics*, pp. 1–5. doi: 10.1155/2008/420747.

Hsieh, L. C. *et al.* (2009) 'Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing', *Plant Physiology*, 151(4), pp. 2120–2132. doi: 10.1104/pp.109.147280.

Hu, B. *et al.* (2019) 'Nitrate–NRT1.1B–SPX4 cascade integrates nitrogen and phosphorus signalling networks in plants', *Nature Plants*, 5(4), pp. 401–413. doi: 10.1038/s41477-019-0384-1.

Huang, T. K. *et al.* (2013) 'Identification of downstream components of ubiquitin-conjugating enzyme PHOSPHATE2 by quantitative membrane proteomics in Arabidopsis roots', *Plant Cell*, 25(10), pp. 4044–4060. doi: 10.1105/tpc.113.115998.

Hurley, B. A. *et al.* (2010) 'The dual-targeted purple acid phosphatase isozyme AtPAP26 is essential for efficient acclimation of Arabidopsis to nutritional phosphate deprivation', *Plant Physiology*, 153(3), pp. 1112–1122. doi: 10.1104/pp.110.153270.

Hyndman, R. J. (2015) 'Forecast: Forecasting functions for time series and linear models. R package version 6.2'. Available at: http://cran.r-project.org/package=forecast.

Hyndman, R. and Khandakar, Y. (2008) 'Automatic time series forecasting: the forecast package for R', *Journal of Statistical Software*, 26(3), pp. 1–22.

IBM-Corp. (2020) 'IBM SPSS Statistics for Windows, Version 27.0'. Armonk, NY: IBM Corp.

Iglesias, J. *et al.* (2013) 'Proteomics identifies ubiquitin-proteasome targets and new roles for chromatin-remodeling in the Arabidopsis response to phosphate starvation', *Journal of Proteomics*, 94, pp. 1–22. doi: 10.1016/j.jprot.2013.08.015.

Jacob, F. and Monod, J. (1961) 'Genetic regulatory mechanisms in the synthesis of proteins', *Journal of Molecular Biology*, 3(3), pp. 318–356. doi: 10.1016/S0022-2836(61)80072-7.

Jain, A. *et al.* (2007) 'Differential effects of sucrose and auxin on localized phosphate deficiency-induced modulation of different traits of root system architecture in Arabidopsis', *Plant Physiology*, 144(1), pp. 232–247. doi: 10.1104/pp.106.092130.

Jain, A., Nagarajan, V. K. and Raghothama, K. G. (2012) 'Transcriptional regulation of phosphate acquisition by higher plants', *Cellular and Molecular Life Sciences*, 69(19), pp. 3207–3224. doi: 10.1007/s00018-012-1090-6.

Jiang, C. *et al.* (2007) 'Phosphate starvation root architecture and anthocyanin accumulation responses are modulated by the gibberellin-DELLA signaling pathway in Arabidopsis', *Plant Physiology*, 145(4), pp. 1460–1470. doi: 10.1104/pp.107.103788.

Johnson, C. M. *et al.* (1957) 'Comparative chlorine requirements of different plant species', *Plant and Soil*, 8(4), pp. 337–353. doi: 10.1007/BF01666323.

Jones-Rhoades, M. W., Bartel, D. P. and Bartel, B. (2006) 'MicroRNAs and their regulatory roles in plants', *Annual Review of Plant Biology*, 57, pp. 19–53. doi: 10.1146/annurev.arplant.57.032905.105218.

Jones, D. F. (1917) 'Dominance of linked factors as a means of accounting for heterosis', *Proceedings of the National Academy of Sciences*, 2, pp. 466–479.

Kant, S., Peng, M. and Rothstein, S. J. (2011) 'Genetic regulation by NLA and microRNA827 for maintaining nitrate-dependent phosphate homeostasis in Arabidopsis', *PLoS Genetics*, 7(3). doi: 10.1371/journal.pgen.1002021.

Karthikeyan, A. S. *et al.* (2002) 'Regulated expression of Arabidopsis phosphate transporters', *Plant Physiology*, 130(1), pp. 221–233. doi: 10.1104/pp.020007.

Karthikeyan, A. S. *et al.* (2007) 'Phosphate starvation responses are mediated by sugar signaling in Arabidopsis', *Planta*, 225(4), pp. 907–918. doi: 10.1007/s00425-006-0408-8.

Katari, M. S. *et al.* (2010) 'VirtualPlant: A Software Platform to Support Systems Biology Research', *Plant Physiology*, 152(2), pp. 500–515. doi: 10.1104/pp.109.147025.

Kawa, D. *et al.* (2016) 'Phosphate-dependent root system architecture responses to salt stress', *Plant Physiology*, 172(2), pp. 690–706. doi: 10.1104/pp.16.00712.

Kawakatsu, T. *et al.* (2016) 'Epigenomic diversity in a global collection of Arabidopsis thaliana accessions', *Cell*, 166(2), pp. 492–505. Available at: http://signal.salk.edu/opticalmaps/.

Kawanabe, T. *et al.* (2016) 'Role of DNA methylation in hybrid vigor in Arabidopsis thaliana', *Proceedings of the National Academy of Sciences of the United States of America*, 113(43), pp. E6704–E6711. doi: 10.1073/pnas.1613372113.

Kearsey, M. J. and Farquhar, A. G. L. (1998) 'QTL analysis in plants; where are we now?', *Heredity*, 80(2), pp. 137–142. doi: 10.1038/sj.hdy.6885001.

Keurentjes, J. J. B. *et al.* (2007) 'Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci', *Proceedings of the National Academy of Sciences of the United States of America*, 104(5), pp. 1708–1713. doi: 10.1073/pnas.0610429104.

Khan, G. A. *et al.* (2016) 'Phosphate deficiency induces the jasmonate pathway and enhances resistance to insect herbivory', *Plant Physiology*, 171(1), pp. 632–644. doi: 10.1104/pp.16.00278.

Kim, D. *et al.* (2019) 'Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype', *Nature Biotechnology*, 37(8), pp. 907–915. doi: 10.1038/s41587-019-0201-4.

Kliebenstein, D. (2009) 'Quantitative genomics: Analyzing intraspecific variation using global gene expression polymorphisms or eqtls', *Annual Review of Plant Biology*, 60, pp. 93–114. doi: 10.1146/annurev.arplant.043008.092114.

Klimyuk, V. I. *et al.* (1993) 'Alkali treatment for rapid preparation of plant material for reliable PCR analysis', *The Plant Journal*, 3(3), pp. 493–494. doi: 10.1111/j.1365-313X.1993.tb00169.x.

Kobayashi, K. *et al.* (2006) 'Membrane lipid alteration during phosphate starvation is regulated by phosphate signaling and auxin/cytokinin cross-talk', *Plant Journal*, 47(2), pp. 238–248. doi: 10.1111/j.1365-313X.2006.02778.x.

Koornneef, M., Alonso-Blanco, C. and Vreugdenhil, D. (2004) 'Naturally occurring genetic variation in Arabidopsis thaliana', *Annual Review of Plant Biology*, 55, pp. 141–172. doi: 10.1146/annurev.arplant.55.031903.141605.

Koornneef, M. and Meinke, D. (2010) 'The development of Arabidopsis as a model plant',

Plant Journal, 61(6), pp. 909–921. doi: 10.1111/j.1365-313X.2009.04086.x.

Korber, P. *et al.* (2004) 'Evidence for Histone Eviction in trans upon Induction of the Yeast PHO5 Promoter', *Molecular and Cellular Biology*, 24(24), pp. 10965–10974. doi: 10.1128/mcb.24.24.10965-10974.2004.

Korn, M. *et al.* (2008) 'Heterosis in the freezing tolerance, and sugar and flavonoid contents of crosses between Arabidopsis thaliana accessions of widely varying freezing tolerance', *Plant, Cell and Environment*, 31(6), pp. 813–827. doi: 10.1111/j.1365-3040.2008.01800.x.

Korn, M. *et al.* (2010) 'Predicting arabidopsis freezing tolerance and heterosis in freezing tolerance from metabolite composition', *Molecular Plant*, 3(1), pp. 224–235. doi: 10.1093/mp/ssp105.

Krieger, U., Lippman, Z. B. and Zamir, D. (2010) 'The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato', *Nature Genetics*, 42(5), pp. 459–463. doi: 10.1038/ng.550.

Kuo, D. *et al.* (2010) 'Coevolution within a transcriptional network by compensatory trans and cis mutations', *Genome Research*, 20(12), pp. 1672–1678. doi: 10.1101/gr.111765.110.

Kusterer, B., Muminovic, J., *et al.* (2007) 'Analysis of a triple testcross design with recombinant inbred lines reveals a significant role of epistasis in heterosis for biomass-related traits in arabidopsis', *Genetics*, 175(4), pp. 2009–2017. doi: 10.1534/genetics.106.069005.

Kusterer, B., Piepho, H. P., *et al.* (2007) 'Heterosis for biomass-related traits in arabidopsis investigated by quantitative trait loci analysis of the triple testcross design with recombinant inbred lines', *Genetics*, 177(3), pp. 1839–1850. doi: 10.1534/genetics.107.077628.

Lam, F. H., Steger, D. J. and O'Shea, E. K. (2008) 'Chromatin decouples promoter threshold from dynamic range', *Nature*, 453(7192), pp. 246–250. doi: 10.1038/nature06867.

Lambers, H. *et al.* (2011) 'Phosphorus nutrition of proteaceae in severely phosphorus-impoverished soils: Are there lessons to be learned for future crops?', *Plant Physiology*, 156(3), pp. 1058–1066. doi: 10.1104/pp.111.174318.

Landry, C. R. *et al.* (2005) 'Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of drosophila', *Genetics*, 171(4), pp. 1813–1822. doi: 10.1534/genetics.105.047449.

Langfelder, P. and Horvath, S. (2008) 'WGCNA: An R package for weighted correlation network analysis', *BMC Bioinformatics*, 9. doi: 10.1186/1471-2105-9-559.

Langfelder, P. and Horvath, S. (2012) 'Fast R Functions for Robust Correlations and

Hierarchical Clustering.', *Journal of statistical software*, 46(11), pp. 1–7. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf.

Lauss, K. *et al.* (2018) 'Parental DNA methylation states are associated with heterosis in epigenetic hybrids', *Plant Physiology*, 176(2), pp. 1627–1645. doi: 10.1104/pp.17.01054.

Lemmon, Z. H. *et al.* (2014) 'The Role of cis Regulatory Evolution in Maize Domestication', *PLoS Genetics*, 10(11). doi: 10.1371/journal.pgen.1004745.

Lemos, B. *et al.* (2008) 'Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression', *Proceedings of the National Academy of Sciences of the United States of America*, 105(38), pp. 14471–14476. doi: 10.1073/pnas.0805160105.

Lewis, E. B. (1954) 'The theory and application of a new method of detecting chromosomal rearrangements in Drosophila melanogaster', *The American Naturalist*, 88(841), pp. 225–239.

Li, Xin *et al.* (2015) 'Dissecting repulsion linkage in the dwarfing gene Dw3 region for sorghum plant height provides insights into heterosis', *Proceedings of the National Academy of Sciences of the United States of America*, 112(38), pp. 11823–11828. doi: 10.1073/pnas.1509229112.

Li, Y. *et al.* (2010) 'Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana', *Proceedings of the National Academy of Sciences of the United States of America*, 107(49), pp. 21199–21204. doi: 10.1073/pnas.1007431107.

Liao, Y., Smyth, G. K. and Shi, W. (2014) 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*, 30(7), pp. 923–930. doi: 10.1093/bioinformatics/btt656.

Lin, W. Y., Huang, T. K. and Chiou, T. J. (2013) 'NITROGEN LIMITATION ADAPTATION, a target of MicroRNA827, Mediates degradation of plasma membrane-localized phosphate transporters to maintain phosphate homeostasis in Arabidopsis', *Plant Cell*, 25(10), pp. 4061–4074. doi: 10.1105/tpc.113.116012.

Lisec, J. *et al.* (2008) 'Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations', *Plant Journal*, 53(6), pp. 960–972. doi: 10.1111/j.1365-313X.2007.03383.x.

Lisec, J. *et al.* (2009) 'Identification of heterotic metabolite QTL in Arabidopsis thaliana RIL and IL populations', *Plant Journal*, 59(5), pp. 777–788. doi: 10.1111/j.1365-313X.2009.03910.x.

Liu, J. *et al.* (2015) 'A vacuolar phosphate transporter essential for phosphate homeostasis in

Arabidopsis', *Proceedings of the National Academy of Sciences of the United States of America*, 112(47), pp. E6571–E6578. doi: 10.1073/pnas.1514598112.

Liu, T. Y. *et al.* (2011) 'Vacuolar ca 2+/H + transport activity is required for systemic phosphate homeostasis involving shoot-to-root signaling in arabidopsis', *Plant Physiology*, 156(3), pp. 1176–1189. doi: 10.1104/pp.111.175257.

Liu, T. Y. *et al.* (2012) 'PHO2-dependent degradation of PHO1 modulates phosphate homeostasis in Arabidopsis', *Plant Cell*, 24(5), pp. 2168–2183. doi: 10.1105/tpc.112.096636.

Liu, T. Y. *et al.* (2016) 'Identification of plant vacuolar transporters mediating phosphate storage', *Nature Communications*, 7. doi: 10.1038/ncomms11095.

López-Arredondo, D. L. *et al.* (2014) 'Phosphate nutrition: Improving low-phosphate tolerance in crops', *Annual Review of Plant Biology*, 65, pp. 95–123. doi: 10.1146/annurev-arplant-050213-035949.

López-Bucio, J. *et al.* (2002) 'Phosphate availability alters architecture and causes changes in hormone sensitivity in the Arabidopsis root system', *Plant Physiology*, 129(1), pp. 244–256. doi: 10.1104/pp.010934.

López-Ráez, J. A. *et al.* (2008) 'Tomato strigolactones are derived from carotenoids and their biosynthesis is promoted by phosphate starvation', *New Phytologist*, 178(4), pp. 863–874. doi: 10.1111/j.1469-8137.2008.02406.x.

Loudet, O. *et al.* (2003) 'Quantitative trait loci analysis of water and anion contents in interaction with nitrogen availability in Arabidopsis thaliana', *Genetics*, 163(2), pp. 711–722.

Louis Joosen, R. V. *et al.* (2013) 'Identifying genotype-by-environment interactions in the metabolism of germinating arabidopsis seeds using generalized genetical genomics', *Plant Physiology*, 162(2), pp. 553–566. doi: 10.1104/pp.113.216176.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Lowry, D. B. *et al.* (2013) 'Expression quantitative trait locus mapping across water availability environments reveals contrasting associations with genomic features in Arabidopsis', *Plant Cell*, 25(9), pp. 3266–3279. doi: 10.1105/tpc.113.115352.

Lv, Q. *et al.* (2014) 'SPX4 negatively regulates phosphate signaling and homeostasis through its interaction with PHR2 in rice', *Plant Cell*, 26(4), pp. 1586–1597. doi: 10.1105/tpc.114.123208.

Lynch, J. P. and Brown, K. M. (2001) 'Topsoil foraging - An architectural adaptation of plants

to low phosphorus availability', *Plant and Soil*, 237(2), pp. 225–237. doi: 10.1023/A:1013324727040.

Ma, Z. *et al.* (2003) 'Regulation of root elongation under phosphorus stress involves changes in ethylene responsiveness', *Plant Physiology*, 131(3), pp. 1381–1390. doi: 10.1104/pp.012161.

Mackay, T. F. C. (2014) 'Epistasis and quantitative traits: Using model organisms to study gene-gene interactions', *Nature Reviews Genetics*, 15(1), pp. 22–33. doi: 10.1038/nrg3627.

MacKay, T. F. C., Stone, E. A. and Ayroles, J. F. (2009) 'The genetics of quantitative traits: Challenges and prospects', *Nature Reviews Genetics*, 10(8), pp. 565–577. doi: 10.1038/nrg2612.

Magliano, T. M. A. *et al.* (2005) 'New Arabidopsis recombinant inbred lines (Landsberg erecta x Nossen) reveal natural variation in phytochrome-mediated responses', *Plant Physiology*, 138(2), pp. 1126–1135. doi: 10.1104/pp.104.059071.

Marchadier, E. *et al.* (2019) 'The complex genetic architecture of shoot growth natural variation in Arabidopsis thaliana', *PLoS Genetics*, 15(4), pp. 1–27. doi: 10.1371/journal.pgen.1007954.

Marschner, P. (ed.) (2012) *Marschner's Mineral Nutrition of Higher Plants*. Third Edit. Academic Press. doi: https://doi.org/10.1016/C2009-0-63043-9.

Martín, A. C. *et al.* (2000) 'Influence of cytokinins on the expression of phosphate starvation responsive genes in Arabidopsis', *Plant Journal*, 24(5), pp. 559–567. doi: 10.1046/j.1365-313X.2000.00893.x.

Mayr, E. (1986) 'Joseph Gottlieb Kolreuter's Contributions to Biology', *Osiris*, 2(1), pp. 135–176. doi: 10.1086/368655.

McKenna, A. *et al.* (2010) 'The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data', *Genome Research*, 20(9), pp. 1297–1303. doi: 10.1101/gr.107524.110.

McManus, C. J. *et al.* (2010) 'Regulatory divergence in Drosophila revealed by mRNA-seq', *Genome Research*, 20(6), pp. 816–825. doi: 10.1101/gr.102491.109.

Mehraj, H. *et al.* (2020) 'In arabidopsis thaliana heterosis level varies among individuals in an F1 hybrid population', *Plants*, 9(4), pp. 1–16. doi: 10.3390/plants9040414.

Melchinger, A. E. *et al.* (2007) 'Genetic basis of heterosis for growth-related traits in Arabidopsis investigated by testcross progenies of near-isogenic lines reveals a significant role of epistasis', *Genetics*, 177(3), pp. 1827–1837. doi: 10.1534/genetics.107.080564.

Mellert, D. J. and Truman, J. W. (2012) 'Transvection is common throughout the Drosophila

genome', *Genetics*, 191(4), pp. 1129–1141. doi: 10.1534/genetics.112.140475.

Mendel, G. (1977) *Experiments in Plant Hybridization*. Edited by H. U. Press. Cambridge MA.

Metzger, B. P. H. *et al.* (2016) 'Contrasting Frequencies and Effects of cis- and trans-Regulatory Mutations Affecting Gene Expression', *Molecular Biology and Evolution*, 33(5), pp. 1131–1146. doi: 10.1093/molbev/msw011.

Meyer, R. C. *et al.* (2004) 'Heterosis of biomass production in arabidopsis. Establishment during early development', *Plant Physiology*, 134(4), pp. 1813–1823. doi: 10.1104/pp.103.033001.

Meyer, R. C. *et al.* (2010) 'QTL analysis of early stage heterosis for biomass in Arabidopsis', *Theoretical and Applied Genetics*, 120(2), pp. 227–237. doi: 10.1007/s00122-009-1074-6.

Meyer, R. C. *et al.* (2012) 'Heterosis manifestation during early Arabidopsis seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids', *Plant Journal*, 71(4), pp. 669–683. doi: 10.1111/j.1365-313X.2012.05021.x.

Meyerowitz, E. M. (1987) 'Arabidopsis thaliana', *Annual review of genetics*, 21, pp. 93–111. doi: 10.1201/9781351072533.

Misson, J. *et al.* (2005) 'A genome-wide transcriptional analysis using Arabidopsis thaliana Affymetrix gene chips determined plant responses to phosphate deprivation', *Proceedings of the National Academy of Sciences of the United States of America*, 102(33), pp. 11934–11939. doi: 10.1073/pnas.0505266102.

Mitchell-Olds, T. (1995) 'Interval mapping of viability loci causing heterosis in Arabidopsis', *Genetics*, 140(3), pp. 1105–1109.

Mitchell-Olds, T. and Schmitt, J. (2006) 'Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis', *Nature*, 441(7096), pp. 947–952. doi: 10.1038/nature04878.

Mitchell-Olds, T., Willis, J. H. and Goldstein, D. B. (2007) 'Which evolutionary processes influence natural genetic variation for phenotypic traits?', *Nature Reviews Genetics*, 8(11), pp. 845–856. doi: 10.1038/nrg2207.

Miura, K. *et al.* (2005) 'The Arabidopsis SUMO E3 ligase SIZ1 controls phosphate deficiency responses', *Proceedings of the National Academy of Sciences of the United States of America*, 102(21), pp. 7760–7765. doi: 10.1073/pnas.0500778102.

Moore, S. and Lukens, L. (2011) 'An evaluation of arabidopsis thaliana hybrid traits and their genetic control', *G3: Genes, Genomes, Genetics*, 1(7), pp. 571–579. doi: 10.1534/g3.111.001156.

Mora-Macías, J. *et al.* (2017) 'Malate-dependent Fe accumulation is a critical checkpoint in the root developmental response to low phosphate', *Proceedings of the National Academy of Sciences of the United States of America*, 114(17), pp. E3563–E3572. doi: 10.1073/pnas.1701952114.

Morcuende, R. *et al.* (2007) 'Genome-wide reprogramming of metabolism and regulatory networks of Arabidopsis in response to phosphorus', *Plant, Cell and Environment*, 30(1), pp. 85–112. doi: 10.1111/j.1365-3040.2006.01608.x.

Mudge, S. R. *et al.* (2002) 'Expression analysis suggests novel roles for members of the Pht1 family of phosphate transporters in Arabidopsis', *Plant Journal*, 31(3), pp. 341–353. doi: 10.1046/j.1365-313X.2002.01356.x.

Mukherjee, A. *et al.* (2014) 'Phosphate acquisition efficiency and phosphate starvation tolerance locus (PSTOL1) in rice', *Journal of Genetics*, 93(3), pp. 683–688. doi: 10.1007/s12041-014-0424-6.

Müller, J. *et al.* (2015) 'Iron-Dependent Callose Deposition Adjusts Root Meristem Maintenance to Phosphate Availability', *Developmental Cell*, 33(2), pp. 216–230. doi: 10.1016/j.devcel.2015.02.007.

Müller, R. *et al.* (2005) 'Interaction between phosphate starvation signalling and hexokinase-independent sugar sensing in Arabidopsis leaves', *Physiologia Plantarum*, 124(1), pp. 81–90. doi: 10.1111/j.1399-3054.2005.00496.x.

Müller, R. *et al.* (2007) 'Genome-wide analysis of the arabidopsis leaf transcriptome reveals interaction of phosphate and sugar metabolism', *Plant Physiology*, 143(1), pp. 156–171. doi: 10.1104/pp.106.090167.

Nakamura, Y. (2013) 'Phosphate starvation and membrane lipid remodeling in seed plants', *Progress in Lipid Research*, 52(1), pp. 43–50. doi: 10.1016/j.plipres.2012.07.002.

Narang, R. A. and Altmann, T. (2001) 'Phosphate acquisition heterosis in Arabidopsis thaliana: A morphological and physiological analysis', *Plant and Soil*, 234(1), pp. 91–97. doi: 10.1023/A:1010545101345.

Narise, T. *et al.* (2010) 'Involvement of auxin signaling mediated by IAA14 and ARF7/19 in membrane lipid remodeling during phosphate starvation', *Plant Molecular Biology*, 72(4), pp. 533–544. doi: 10.1007/s11103-009-9589-4.

Ni, Z. *et al.* (2009) 'Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids', *Nature*, 457(7227), pp. 327–331. doi: 10.1038/nature07523.

Nica, A. C. and Dermitzakis, E. T. (2013) 'Expression quantitative trait loci: Present and future', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1620). doi: 10.1098/rstb.2012.0362.

Nicotra, A. B. *et al.* (2010) 'Plant phenotypic plasticity in a changing climate', *Trends in Plant Science*, 15(12), pp. 684–692. doi: 10.1016/j.tplants.2010.09.008.

Nilsson, L., Müller, R. and Nielsen, T. H. (2007) 'Increased expression of the MYB-related transcription factor, PHR1, leads to enhanced phosphate uptake in Arabidopsis thaliana', *Plant, Cell and Environment*, 30(12), pp. 1499–1512. doi: 10.1111/j.1365-3040.2007.01734.x.

Nuzhdin, S. V., Friesen, M. L. and McIntyre, L. M. (2012) 'Genotype-phenotype mapping in a post-GWAS world', *Trends in Genetics*, 28(9), pp. 421–426. doi: 10.1016/j.tig.2012.06.003.

O'Malley, R. C. *et al.* (2016) 'Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape', *Cell*, 165(5), pp. 1280–1292. doi: 10.1016/j.cell.2016.04.038.

O'Quin, K. E. *et al.* (2012) 'Evolution of cichlid vision via trans-regulatory divergence', *BMC Evolutionary Biology*, 12(1). doi: 10.1186/1471-2148-12-251.

Oakley, C. G. *et al.* (2019) 'Heterosis is common and inbreeding depression absent in natural populations of Arabidopsis thaliana', *Journal of Evolutionary Biology*, 32(6), pp. 592–603. doi: 10.1111/jeb.13441.

Oakley, C. G., Ågren, J. and Schemske, D. W. (2015) 'Heterosis and outbreeding depression in crosses between natural populations of Arabidopsis thaliana', *Heredity*, 115(1), pp. 73–82. doi: 10.1038/hdy.2015.18.

Oburger, E. *et al.* (2009) 'Interactive effects of organic acids in the rhizosphere', *Soil Biology and Biochemistry*, 41(3), pp. 449–457. doi: 10.1016/j.soilbio.2008.10.034.

Oleksiak, M. F., Churchill, G. A. and Crawford, D. L. (2002) 'Variation in gene expression within and among natural populations', *Nature Genetics*, 32(2), pp. 261–266. doi: 10.1038/ng983.

Oliveros, J. C. (2007) *Venny. An interactive tool for comparing lists with Venn's diagrams*. Available at: https://bioinfogp.cnb.csic.es/tools/venny/index.html.

Ooijen, J. van and Voorrips, R. (2001) 'JoinMap® 3.0, Software for the Calculation of Genetic Linkage Maps'. Plant Research International, Wageningen, the Netherlands.

Pant, B. D. *et al.* (2009) 'Identification of nutrient-responsive Arabidopsis and rapeseed microRNAs by comprehensive real-time polymerase chain reaction profiling and small RNA sequencing', *Plant Physiology*, 150(3), pp. 1541–1555. doi: 10.1104/pp.109.139139.

Pant, B. D. *et al.* (2015) 'Identification of primary and secondary metabolites with phosphorus status-dependent abundance in Arabidopsis, and of the transcription factor PHR1 as a major regulator of metabolic changes during phosphorus limitation', *Plant Cell and Environment*, 38(1), pp. 172–187. doi: 10.1111/pce.12378.

Park, B. S., Seo, J. S. and Chua, N. H. (2014) 'NITROGEN LIMITATION ADAPTATION Recruits PHOSPHATE2 to target the phosphate transporter PT2 for degradation during the regulation of Arabidopsis phosphate homeostasis', *Plant Cell*, 26(1), pp. 454–464. doi: 10.1105/tpc.113.120311.

Pastinen, T. (2010) 'Genome-wide allele-specific analysis: Insights into regulatory variation', *Nature Reviews Genetics*, 11(8), pp. 533–538. doi: 10.1038/nrg2815.

Perera, S. A. C. N., Pooni, H. S. and Kearsey, M. J. (2008) 'Chromosome substitution lines for the analysis of heterosis in Arabidopsis thaliana', *Journal of the National Science Foundation of Sri Lanka*, 36(4), pp. 275–280. doi: 10.4038/jnsfsr.v36i4.265.

Péret, B. *et al.* (2014) 'Root architecture responses: In search of phosphate', *Plant Physiology*, 166(4), pp. 1713–1723. doi: 10.1104/pp.114.244541.

Pérez-Torres, C. A. *et al.* (2008) 'Phosphate availability alters lateral root development in Arabidopsis by modulating auxin sensitivity via a mechanism involving the TIR1 auxin receptor', *Plant Cell*, 20(12), pp. 3258–3272. doi: 10.1105/tpc.108.058719.

Peterson, R. A. and Cavanaugh, J. E. (2020) 'Ordered quantile normalization: a semiparametric transformation built for the cross-validation era', *Journal of Applied Statistics*, 47(13–15), pp. 2312–2327. doi: 10.1080/02664763.2019.1630372.

Pii, Y. *et al.* (2015) 'Microbial interactions in the rhizosphere: beneficial influences of plant growth-promoting rhizobacteria on nutrient acquisition process. A review', *Biology and Fertility of Soils*, 51(4), pp. 403–415. doi: 10.1007/s00374-015-0996-1.

Pirrotta, V. (1990) 'Transvection and long-distance gene regulation', *BioEssays*, 12(9), pp. 409–414.

Plaxton, W. C. and Tran, H. T. (2011) 'Metabolic adaptations of phosphate-starved plants', *Plant Physiology*, 156(3), pp. 1006–1015. doi: 10.1104/pp.111.175281.

Poczai, P. *et al.* (2000) 'Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.', *Genetic resources and crop evolution*, 61(1), pp. 796–815. doi: 10.1134/S1022795411020074.

Powers, L. (1944) 'An Expansion of Jones's Theory for the Explanation of Heterosis', *The*

American Naturalist, 78(776), pp. 275–280. doi: 10.1086/281199.

Prinzenberg, A. E. *et al.* (2010) 'Relationships between growth, growth response to nutrient supply, and ion content using a recombinant inbred line population in arabidopsis', *Plant Physiology*, 154(3), pp. 1361–1371. doi: 10.1104/pp.110.161398.

Provart, N. J. *et al.* (2016) '50 years of Arabidopsis research: Highlights and future directions', *New Phytologist*, 209(3), pp. 921–944. doi: 10.1111/nph.13687.

Prud'homme, B., Gompel, N. and Carroll, S. B. (2007) 'Emerging principles of regulatory evolution', *In the Light of Evolution*, 1, pp. 109–127. doi: 10.17226/11790.

Puga, M. I. *et al.* (2014) 'SPX1 is a phosphate-dependent inhibitor of Phosphate Starvation Response 1 in Arabidopsis', *Proceedings of the National Academy of Sciences of the United States of America*, 111(41), pp. 14947–14952. doi: 10.1073/pnas.1404654111.

Puga, M. I. *et al.* (2017) 'Novel signals in the regulation of Pi starvation responses in plants: facts and promises', *Current Opinion in Plant Biology*, 39(Figure 1), pp. 40–49. doi: 10.1016/j.pbi.2017.05.007.

R-Core-Team (2020) 'R: A language and environment for statistical computing'. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/.

Raghothama, K. G. (1999) 'Phosphate acquisition', *Annual Review of Plant Physiology*, 50, pp. 665–693. doi: 10.1007/s11104-004-2005-6.

Rassoulzadegan, M., Magliano, M. and Cuzin, F. (2002) 'Transvection effects involving DNA methylation during meiosis in the mouse', *EMBO Journal*, 21(3), pp. 440–450. doi: 10.1093/emboj/21.3.440.

Rebeiz, M. and Williams, T. M. (2017) 'Using Drosophila pigmentation traits to study the mechanisms of cis-regulatory evolution', *Current Opinion in Insect Science*, 19, pp. 1–7. doi: 10.1016/j.cois.2016.10.002.

Reymond, M. *et al.* (2006) 'Identification of QTL controlling root growth response to phosphate starvation in Arabidopsis thaliana', *Plant, Cell and Environment*, 29(1), pp. 115–125. doi: 10.1111/j.1365-3040.2005.01405.x.

Ribot, C., Wang, Y. and Poirier, Y. (2008) 'Expression analyses of three members of the AtPHO1 family reveal differential interactions between signaling pathways involved in phosphate deficiency and the responses to auxin, cytokinin, and abscisic acid', *Planta*, 227(5), pp. 1025–1036. doi: 10.1007/s00425-007-0677-x.

Richardson, A. E. *et al.* (2009) 'Acquisition of phosphorus and nitrogen in the rhizosphere and

plant growth promotion by microorganisms', *Plant and Soil*, 321(1–2), pp. 305–339. doi: 10.1007/s11104-009-9895-2.

Rockman, M. V. and Kruglyak, L. (2006) 'Genetics of global gene expression', *Nature Reviews Genetics*, 7(11), pp. 862–872. doi: 10.1038/nrg1964.

Rohde, P., Hincha, D. K. and Heyer, A. G. (2004) 'Heterosis in the freezing tolerance of crosses between two Arabidopsis thaliana accessions (Columbia-0 and C24) that show differences in non-acclimated and acclimated freezing tolerance', *Plant Journal*, 38(5), pp. 790–799. doi: 10.1111/j.1365-313X.2004.02080.x.

Rojas-Triana, M. *et al.* (2013) 'Roles of Ubiquitination in the Control of Phosphate Starvation Responses in Plants', *Journal of Integrative Plant Biology*, 55(1), pp. 40–53. doi: 10.1111/jipb.12017.

Romero, I. G., Ruvinsky, I. and Gilad, Y. (2012) 'Comparative studies of gene expression and the evolution of gene regulation', *Nature Reviews Genetics*, 13(7), pp. 505–516. doi: 10.1038/nrg3229.

Ronald, J. *et al.* (2005) 'Local regulatory variation in Saccharomyces cerevisiae', *PLoS Genetics*, 1(2), pp. 0213–0222. doi: 10.1371/journal.pgen.0010025.

Rosas, U. *et al.* (2013) 'Integration of responses within and across Arabidopsis natural accessions uncovers loci controlling root systems architecture', *Proceedings of the National Academy of Sciences of the United States of America*, 110(37), pp. 15133–15138. doi: 10.1073/pnas.1305883110.

Rouached, H. *et al.* (2011) 'The transcription factor PHR1 plays a key role in the regulation of sulfate shoot-to-root flux upon phosphate starvation in Arabidopsis', *BMC Plant Biology*, 11, pp. 1–10. doi: 10.1186/1471-2229-11-19.

Rouached, H., Arpat, A. B. and Poirier, Y. (2010) 'Regulation of phosphate starvation responses in plants: Signaling players and cross-talks', *Molecular Plant*, 3(2), pp. 288–299. doi: 10.1093/mp/ssp120.

Routaboul, J.-M. *et al.* (2012) 'Metabolite profiling and quantitative genetics of natural variation for flavonoids in Arabidopsis', *Journal of Experimental Botany*, 63(10), pp. 3749–3764. doi: 10.1093/jxb/ers067.

Royston, P. (1995) 'Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality', *Applied Statistics*, 44(4), pp. 547–551. doi: 10.2307/2986146.

Rubio, V. *et al.* (2001) 'A conserved MYB transcription factor involved in phosphate starvation

signaling both in vascular plants and in unicellular algae', *Genes and Development*, 15(16), pp. 2122–2133. doi: 10.1101/gad.204401.

Rubio, V. *et al.* (2009) 'Plant hormones and nutrient signaling', *Plant Molecular Biology*, 69(4), pp. 361–373. doi: 10.1007/s11103-008-9380-y.

Ryan, P. R., Delhaize, E. and Jones, D. L. (2001) 'Function and mechanism of organic anion exudation from plant roots', *Annual Review of Plant Physiology*, 52, pp. 527–560.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory.

Sánchez-Bermejo, E. *et al.* (2012) 'Novel natural alleles at FLC and LVR loci account for enhanced vernalization responses in arabidopsis thaliana', *Plant, Cell and Environment*, 35(9), pp. 1672–1684. doi: 10.1111/j.1365-3040.2012.02518.x.

Sánchez-Calderón, L. *et al.* (2005) 'Phosphate starvation induces a determinate developmental program in the roots of Arabidopsis thaliana', *Plant and Cell Physiology*, 46(1), pp. 174–184. doi: 10.1093/pcp/pci011.

Sandhu, K. S. *et al.* (2009) 'Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development', *Genes and Development*, 23(22), pp. 2598–2603. doi: 10.1101/gad.552109.

Schachtman, D. P. and Shin, R. (2007) 'Nutrient sensing and signaling: NPKS', *Annual Review of Plant Biology*, 58, pp. 47–69. doi: 10.1146/annurev.arplant.58.032806.103750.

Scheible, W.-R. and Rojas-Triana, M. (2015) 'Sensing, signaling, and control of phosphate starvation in plants: molecular players and applications', in *Annual Plant Reviews Volume 48*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 23–63. doi: 10.1002/9781118958841.ch2.

Schmitz, R. J. *et al.* (2013) 'Patterns of population epigenomic diversity', *Nature*, 495(7440), pp. 193–198. doi: 10.1038/nature11968.

Schnable, P. S. and Springer, N. M. (2013) 'Progress toward understanding heterosis in crop plants', *Annual Review of Plant Biology*, 64, pp. 71–88. doi: 10.1146/annurev-arplant-042110-103827.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods*, 9(7), pp. 671–675. doi: 10.1038/nmeth.2089.

Scott, R. J. *et al.* (1998) 'Parent-of-origin effects on seed development in Arabidopsis thaliana', *Development*, 125(17), pp. 3329–3341.

Sharma, R. K., Griffing, B. and Scholl, R. L. (1979) 'Variations among races of Arabidopsis thaliana (L.) heynh for survival in limited carbon dioxide', *Theoretical and Applied Genetics*, 54(1), pp. 11–15. doi: 10.1007/BF00265702.

Shi, J. *et al.* (2011) 'A dynamic and complex network regulates the heterosis of yield-correlated traits in rapeseed (Brassica napus L.)', *PLoS ONE*, 6(7). doi: 10.1371/journal.pone.0021645.

Shi, T. *et al.* (2013) 'QTL for Yield Traits and Their Association with Functional Genes in Response to Phosphorus Deficiency in Brassica napus', *PLoS ONE*, 8(1), pp. 1–12. doi: 10.1371/journal.pone.0054559.

Shi, X. *et al.* (2012) 'Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids', *Nature Communications*, 3. doi: 10.1038/ncomms1954.

Shin, H. *et al.* (2006) 'Loss of At4 function impacts phosphate distribution between the roots and the shoots during phosphate starvation', *Plant Journal*, 45(5), pp. 712–726. doi: 10.1111/j.1365-313X.2005.02629.x.

Shin, R., Berg, R. H. and Schachtman, D. P. (2005) 'Reactive oxygen species and root hairs in arabidopsis root response to nitrogen, phosphorus and potassium deficiency', *Plant and Cell Physiology*, 46(8), pp. 1350–1357. doi: 10.1093/pcp/pci145.

Shindo, C. *et al.* (2005) 'Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis', *Plant Physiology*, 138(2), pp. 1163–1173. doi: 10.1104/pp.105.061309.

Shull, G. H. (1908) 'The composition of a field of maize', *Journal of Heredity*, 4, pp. 296–301.

Shull, G. H. (1948) 'What Is "Heterosis"?', *Genetics*, 33(5), pp. 439–446.

Signor, S. A. *et al.* (2016) 'Genetic Convergence in the Evolution of Male-Specific Color Patterns in Drosophila', *Current Biology*, 26(18), pp. 2423–2433. doi: 10.1016/j.cub.2016.07.034.

Signor, S. A. and Nuzhdin, S. V. (2018) 'The Evolution of Gene Expression in cis and trans', *Trends in Genetics*, 34(7), pp. 532–544. doi: 10.1016/j.tig.2018.03.007.

Smith, A. P. *et al.* (2010) 'Histone H2A.Z regulates the expression of several classes of phosphate starvation response genes but not as a transcriptional activator', *Plant Physiology*, 152(1), pp. 217–225. doi: 10.1104/pp.109.145532.

Smith, S. E. *et al.* (2011) 'Roles of arbuscular mycorrhizas in plant phosphorus nutrition: Interactions between pathways of phosphorus uptake in arbuscular mycorrhizal roots have

important implications for understanding and manipulating plant phosphorus acquisition', *Plant Physiology*, 156(3), pp. 1050–1057. doi: 10.1104/pp.111.174581.

Snoek, L. B. *et al.* (2013) 'Genetical genomics reveals large scale genotype-by-environment interactions in Arabidopsis thaliana', *Frontiers in Genetics*, 3(JAN), pp. 1–13. doi: 10.3389/fgene.2012.00317.

Soyk, S. *et al.* (2017) 'Bypassing Negative Epistasis on Yield in Tomato Imposed by a Domestication Gene', *Cell*, 169(6), pp. 1142–1155. doi: 10.1016/j.cell.2017.04.032.

Stebbins, G. L. (1959) 'The Role of Hybridization in Evolution', *Proceedings of the American Philosophical Society*, 103(2), pp. 231–251. Available at: http://www.jstor.org/stable/985151.

Stern, D. L. and Orgogozo, V. (2008) 'The loci of evolution: How predictable is genetic evolution?', *Evolution*, 62(9), pp. 2155–2177. doi: 10.1111/j.1558-5646.2008.00450.x.

Streisfeld, M. A. and Rausher, M. D. (2009) 'Altered trans-regulatory control of gene expression in multiple anthocyanin genes contributes to adaptive flower color evolution in Mimulus aurantiacus', *Molecular Biology and Evolution*, 26(2), pp. 433–444. doi: 10.1093/molbev/msn268.

Sulpice, R. *et al.* (2014) 'Low levels of ribosomal RNA partly account for the very high photosynthetic phosphorus-use efficiency of Proteaceae species', *Plant, Cell and Environment*, 37(6), pp. 1276–1298. doi: 10.1111/pce.12240.

Sun, B., Gao, Y. and Lynch, J. P. (2018) 'Large crown root number improves topsoil foraging and phosphorus acquisition', *Plant Physiology*, 177(1), pp. 90–104. doi: 10.1104/pp.18.00234.

Sun, L. *et al.* (2016) 'Arabidopsis PHL2 and PHR1 act redundantly as the key components of the central regulatory system controlling transcriptional responses to phosphate starvation', *Plant Physiology*, 170(1), pp. 499–514. doi: 10.1104/pp.15.01336.

Sung, H. M. *et al.* (2009) 'Roles of trans and cis variation in yeast intraspecies evolution of gene expression', *Molecular Biology and Evolution*, 26(11), pp. 2533–2538. doi: 10.1093/molbev/msp171.

Svistoonoff, S. *et al.* (2007) 'Root tip contact with low-phosphate media reprograms plant root architecture', *Nature Genetics*, 39(6), pp. 792–796. doi: 10.1038/ng2041.

Swain, T. and Hillis, W. (1959) 'The Phenolic constituents of prunus domestica', *Science of Food and Agriculture*, 10(1), pp. 63–68.

Swanson-Wagner, R. A. *et al.* (2009) 'Paternal dominance of trans-eQTL influences gene expression patterns in maize hybrids', *Science*, 326(5956), pp. 1118–1120. doi:

10.1126/science.1178294.

Syed, N. and Chen, Z. (2005) 'Molecular marker genotypes, heterozygosity and genetic interactions explain heterosis in Arabidopsis thaliana', *Heredity*, 94(3), pp. 295–304. doi: doi:10.1038/sj.hdy.6800558.

Takahasi, K. R., Matsuo, T. and Takano-Shimizu-Kouno, T. (2011) 'Two types of cis-trans compensation in the evolution of transcriptional regulation', *Proceedings of the National Academy of Sciences of the United States of America*, 108(37), pp. 15276–15281. doi: 10.1073/pnas.1105814108.

Tay, Y., Rinn, J. and Pandolfi, P. P. (2014) 'The multilayered complexity of ceRNA crosstalk and competition', *Nature*, 505(7483), pp. 344–352. doi: 10.1038/nature12986.

Thibaud, M. C. *et al.* (2010) 'Dissection of local and systemic transcriptional responses to phosphate starvation in Arabidopsis', *Plant Journal*, 64(5), pp. 775–789. doi: 10.1111/j.1365-313X.2010.04375.x.

Trontin, C. *et al.* (2011) 'What does Arabidopsis natural variation teach us (and does not teach us) about adaptation in plants?', *Current Opinion in Plant Biology*, 14(3), pp. 225–231. doi: 10.1016/j.pbi.2011.03.024.

Troyer, A. F. (2006) 'Adaptedness and heterosis in corn and mule hybrids', *Crop Science*, 46, pp. 528–543. Available at: https://books.google.com.co/books?id=1GrXVz5JhzQC&pg=PA359&dq=mitotic+genes+and+meiosis+in+maize&hl=es&sa=X&ved=0ahUKEwiO-uP725LQAhXBLyYKHUwTAjUQ6AEIJDAA#v=onepage&q=mitotic genes and meiosis in maize&f=true.

Trull, M. C. *et al.* (1997) 'The responses of wild-type and ABA mutant Arabidopsis thaliana plants to phosphorus starvation', *Plant, Cell and Environment*, 20(1), pp. 85–92. doi: 10.1046/j.1365-3040.1997.d01-4.x.

Tyburski, J., Dunajska, K. and Tretyn, A. (2009) 'Reactive oxygen species localization in roots of Arabidopsis thaliana seedlings grown under phosphate deficiency', *Plant Growth Regulation*, 59(1), pp. 27–36. doi: 10.1007/s10725-009-9385-9.

Umehara, M. *et al.* (2010) 'Contribution of strigolactones to the inhibition of tiller bud outgrowth under phosphate deficiency in rice', *Plant and Cell Physiology*, 51(7), pp. 1118–1126. doi: 10.1093/pcp/pcq084.

Vance, C. P., Uhde-Stone, C. and Allan, D. L. (2003) 'Phosphorus acquisition and use: Critical adaptations by plants for securing a nonrenewable resource', *New Phytologist*, 157(3), pp. 423–

447. doi: 10.1046/j.1469-8137.2003.00695.x.

Versaw, W. K. and Harrison, M. J. (2002) 'A chloroplast phosphate transporter, PHT2;1, influences allocation of phosphate within the plant and phosphate-starvation responses', *Plant Cell*, 14(8), pp. 1751–1766. doi: 10.1105/tpc.002220.

Verta, J. P., Landry, C. R. and Mackay, J. (2016) 'Dissection of expression-quantitative trait locus and allele specificity using a haploid/diploid plant system - insights into compensatory evolution of transcriptional regulation within populations', *New Phytologist*, 211(1), pp. 159–171. doi: 10.1111/nph.13888.

Videvall, E. *et al.* (2016) 'Strong maternal effects on gene expression in arabidopsis lyrata hybrids', *Molecular Biology and Evolution*, 33(4), pp. 984–994. doi: 10.1093/molbev/msv342.

Vidigal, D. S. *et al.* (2016) 'Altitudinal and climatic associations of seed dormancy and flowering traits evidence adaptation of annual life cycle timing in Arabidopsis thaliana', *Plant Cell and Environment*, 39(8), pp. 1737–1748. doi: 10.1111/pce.12734.

Wang, K., Li, M. and Hakonarson, H. (2010) 'ANNOVAR : functional annotation of genetic variants from high-throughput sequencing data', *Nucleic Acids Research*, 38(16), pp. 1–7. doi: 10.1093/nar/gkq603.

Wang, S., Basten, C. and Zeng, Z. (2012) 'Windows QTL Cartographer 2.5'. Department of Statistics, North Carolina State University, Raleigh, NC. Available at: http://statgen.ncsu.edu/qtlcart/WQTLCart.htm.

Wang, Z. *et al.* (2014) 'Rice SPX1 and SPX2 inhibit phosphate starvation responses through interacting with PHR2 in a phosphate-dependent manner', *Proceedings of the National Academy of Sciences of the United States of America*, 111(41), pp. 14953–14958. doi: 10.1073/pnas.1404680111.

Weigel, D. (2012) 'Natural variation in arabidopsis: From molecular genetics to ecological genomics', *Plant Physiology*, 158(1), pp. 2–22. doi: 10.1104/pp.111.189845.

West, M. A. L. *et al.* (2007) 'Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis', *Genetics*, 175(3), pp. 1441–1450. doi: 10.1534/genetics.106.064972.

Whitehead, A. and Crawford, D. L. (2006) 'Variation within and among species in gene expression: Raw material for evolution', *Molecular Ecology*, 15(5), pp. 1197–1211. doi: 10.1111/j.1365-294X.2006.02868.x.

Wild, R. *et al.* (2016) 'Control of eukaryotic phosphate homeostasis by inositol polyphosphate

sensor domains', *Science*, 352(6288), pp. 986–990. doi: 10.1126/science.aad9858.

Williams, A., Spilianakis, C. G. and Flavell, R. A. (2010) 'Interchromosomal association and gene regulation in trans', *Trends in Genetics*, 26(4), pp. 188–197. doi: 10.1016/j.tig.2010.01.007.

Williams, E. J. B. and Bowles, D. J. (2004) 'Coexpression of neighboring genes in the genome of Arabidopsis thaliana', *Genome Research*, 14(6), pp. 1060–1067. doi: 10.1101/gr.2131104.

Williams, W. (1959) 'Heterosis and the genetics of complex characters', *Nature*, 184, pp. 527–530.

Wittkopp, P. J., Haerum, B. K. and Clark, A. G. (2004) 'Evolutionary changes in cis and trans gene regulation', *Nature*, 430(7), pp. 85–88. Available at: www.nature.com/nature.

Woo, J. *et al.* (2012) 'The response and recovery of the Arabidopsis thaliana transcriptome to phosphate starvation', *BMC Plant Biology*, 12. doi: 10.1186/1471-2229-12-62.

Wray, G. A. (2007) 'The evolutionary significance of cis-regulatory mutations', *Nature Reviews Genetics*, 8(3), pp. 206–216. doi: 10.1038/nrg2063.

Wright, S. (1934) 'Physiological and Evolutionary Theories of Dominance', *The American Naturalist*, 68(714), pp. 24–53. doi: 10.1086/280521.

Wu, P. *et al.* (2003) 'Phosphate starvation triggers distinct alterations of genome expression in arabidopsis roots and leaves', *Plant Physiology*, 132(3), pp. 1260–1271. doi: 10.1104/pp.103.021022.

Yang, B. and Wittkopp, P. J. (2017) 'Structure of the transcriptional regulatory network correlates with regulatory divergence in Drosophila', *Molecular Biology and Evolution*, 34(6), pp. 1352–1362. doi: 10.1093/molbev/msx068.

Yang, M. *et al.* (2010) 'Quantitative trait loci for root morphology in response to low phosphorus stress in Brassica napus', *Theoretical and Applied Genetics*, 121(1), pp. 181–193. doi: 10.1007/s00122-010-1301-1.

Yang, X. J. and Finnegan, P. M. (2010) 'Regulation of phosphate starvation responses in higher plants', *Annals of Botany*, 105(4), pp. 513–526. doi: 10.1093/aob/mcq015.

Yeo, I.-K. and Johnson, R. A. (2000) 'A new family of power transformations to improve normality or symmetry', *Biometrika*, 87(4), pp. 954–959. doi: 10.1093/biomet/87.4.954.

Yoneyama, Kaori *et al.* (2007) 'Nitrogen deficiency as well as phosphorus deficiency in sorghum promotes the production and exudation of 5-deoxystrigol, the host recognition signal

for arbuscular mycorrhizal fungi and root parasites', *Planta*, 227(1), pp. 125–132. doi: 10.1007/s00425-007-0600-5.

van Zanten, M. *et al.* (2009) 'The many functions of ERECTA', *Trends in Plant Science*, 14(4), pp. 214–218. doi: 10.1016/j.tplants.2009.01.010.

Zeng, Z. B. (1993) 'Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci', *Proceedings of the National Academy of Sciences of the United States of America*, 90(23), pp. 10972–10976. doi: 10.1073/pnas.90.23.10972.

Zeng, Z. B. (1994) 'Precision mapping of quantitative trait loci.', *Genetics*, 136(4), pp. 1457–68. doi: 10.1016/j.exger.2011.07.003.

Zhang, D. *et al.* (2017) 'Integrating QTL mapping and transcriptomics identifies candidate genes underlying QTLs associated with soybean tolerance to low-phosphorus stress', *Plant Molecular Biology*, 93(1–2), pp. 137–150. doi: 10.1007/s11103-016-0552-x.

Zhang, Q. *et al.* (2016) 'The chromatin remodeler DDM1 promotes hybrid vigor by regulating salicylic acid metabolism', *Cell Discovery*, 2, pp. 1–12. doi: 10.1038/celldisc.2016.27.

Zhang, X. and Borevitz, J. O. (2009) 'Global analysis of allele-specific expression in Arabidopsis thaliana', *Genetics*, 182(4), pp. 943–954. doi: 10.1534/genetics.109.103499.

Zhang, Y. J., Lynch, J. P. and Brown, K. M. (2003) 'Ethylene and phosphorus availability have interacting yet distinct effects on root hair development', *Journal of Experimental Botany*, 54(391), pp. 2351–2361. doi: 10.1093/jxb/erg250.

Zia, R. *et al.* (2021) 'Plant survival under drought stress: Implications, adaptive responses, and integrated rhizosphere management strategy for stress mitigation', *Microbiological Research*, 242(October 2020), p. 126626. doi: 10.1016/j.micres.2020.126626.