Ph.D. Graduate School in Environmental Sciences
Università degli Studi di Milano

XXXV cycle

Department of
Environmental Science and Policy

Ph.D. Coordinator
Prof. Marcella Guarino

A 'hybrid' approach for the characterization of possible physiological roles of *Arabidopsis thaliana* formate dehydrogenase FDH

Disciplinary Scientific Field BIO/04

Supervisor
Prof. Piero Angelo Morandini

Co-Supervisor
Dr. Irene Murgia

Ph.D. Candidate
Francesca Marzorati
R12733

Academic Year 2021/2022

A Stella: lei sa.

Contents

Figures

Achievements and activities of the candidate during the Ph.D. project

Peer-reviewed published articles

Marzorati F, Midali A, Morandini P, Murgia I. **2022**. Good or bad? The double face of iron in plants. *Frontiers for Young Minds* **10**, 718162.

Murgia I, Marzorati F, Vigani G, Morandini P. **2022**. Plant iron nutrition: the long road from soil to seeds. *Journal of Experimental Botany* **73**, 1809-1824.

Marzorati F, Wang C, Pavesi G, Mizzi L, Morandini P. **2021**. Cleaning the *Medicago* microarray database to improve gene function analysis. *Plants* **10**, 1240.

Marzorati F, Vigani G, Morandini P, Murgia I. **2021**. Formate dehydrogenase contributes to the early *Arabidopsis thaliana* responses against *Xanthomonas campestris* pv *campestris* infection. *Physiological and Molecular Plant Pathology* **114**, 101633.

Manuscript under review

Marzorati F, Rossana R, Bernardo L, Mauri P, Di Silvestre D, Morandini P, Murgia I. **2023**. *Arabidopsis thaliana* early foliar proteome responses to root exposure to the rhizobacterium *Pseudomonas simiae* WCS417. *Molecular Plant-Microbe Interactions*.

Manuscript in preparation

Marzorati F, Mizzi L, Murgia I, Morandini P. **2023**. NORMALIX95: a shiny-based application for plant microarray analysis.

Posters/oral presentations at national/international congresses

Marzorati F, Morandini P, Di Silvestre D, Ferrucci EA, Vigani G, Murgia I. **2022**. Formate dehydrogenase: a hub protein in mitochondrial Fe homeostasis involved in the signaling pathways induced by the rhizobacterium *Pseudomonas simiae* WCS417. 20th International Symposium on Iron Nutrition and Interactions in Plants (ISINIP). Reims, France, 4-8 July 2022 – oral presentation by Irene Murgia.

Marzorati F, Wang C, Pavesi G, Mizzi L, Morandini P. **2021**. Identification of candidate genes in *Medicago* through correlation analysis of microarray data after cleaning and normalization. LXIV Società Italiana Genetica Agraria (SIGA) Annual Congress. Online, 14-16 September 2021 – poster presentation by Piero Morandini.

Marzorati F, Mizzi L, Morandini P. **2021**. NORMALIX: a shiny-based application to investigate plant transcriptome data. LXIV Società Italiana Genetica Agraria (SIGA) Annual Congress. Online, 14-16 September 2021 – poster presentation by Francesca Marzorati.

Marzorati F, Mizzi L, Morandini P. **2021**. How to study plant transcriptome data? Using a shiny-based application called "NORMALIX"! Plant Biology Europe (PBE) 2021. Online, 28 June-1 July 2021 – poster presentation by Francesca Marzorati. This poster has been awarded as one of the best posters of the conference.

Marzorati F, Mizzi L, Morandini P. **2021**. How to investigate plant transcriptome data? Using a shiny-based application called "NORMALIX"! International conference on plant systems biology and biotechnology (ICPSBB). Online, 14-17 June 2021 – oral presentation by Francesca Marzorati.

Marzorati F, Vigani G, Morandini P, Murgia I. **2020**. Formate dehydrogenase contributes to the early *Arabidopsis thaliana* responses against infection by *Xanthomonas campestris* pv *campestris.* "Young Scientists for Plant Health" web workshop during the Web Symposium on Plant Health. Online, 16 December 2020 – poster presentation by Francesca Marzorati.

Marzorati F. **2020**. *Arabidopsis thaliana* formate dehydrogenase FDH is involved in the early responses against infection by *Xanthomonas campestris* pv *campestris*. Società Italiana Biologia Vegetale/A Novel Integrated Approach to Increase Multiple and Combined Stress Tolerance in Plants Using Tomato as a Model (SIBV/TOMRES) summer school 2020. Online, 8-10 September 2020 – oral presentation by Francesca Marzorati.

Third mission

Midali A, Marzorati F, Morandini P. **2020**. How much does it cost to "sweeten" a day's work? *Euractiv website* https://www.euractiv.com/section/agriculture-food/opinion/how-much-does-it-cost-to-sweeten-a-days-work/

Attendance at international Ph.D. schools

International Summer School SBI (Società Botanica Italiana). "Plant based approaches to clean up contaminated environments. Exploring plant–microbe interactions". University of Milano Bicocca, 20-22 July 2022.

SIBV/TOMRES (Società Italiana Biologia Vegetale/A Novel Integrated Approach to Increase Multiple and Combined Stress Tolerance in Plants Using Tomato as a Model) Summer School 2020. Stress resilience in plants: from molecules to field. Online, 8-10 September 2020

Activity of co-tutor to undergraduate students

"Analisi di correlazione dei trascritti per lo studio della funzione della formato deidrogenasi di *Arabidopsis thaliana* in condizioni di carenze nutrizionali" – **Stefano Bellini**, Laurea Triennale in Biotecnologia UNIMI (2020/2021)

"Studio della funzione della Formato Deidrogenasi (FDH) in *Arabidopsis thaliana* in condizioni di ipossia/anossia attraverso l'analisi di correlazione dei trascritti da dati microarray" – **Matteo Corti**, Laurea Triennale in Biotecnologia UNIMI (2020/2021)

"Exploring the role of four genes in *Arabidopsis* iron metabolism through correlation analysis of microarray gene expression data" - **Andrea Maria Gatti**, Laurea Triennale in Biotecnologia UNIMI (2019/2020)

Teaching activities

November 2020 (10 h): **External expert for a bioinformatics beginning course** for high school science professors at the Institute "Don Lorenzo Milani" (Romano di Lombardia).

October 2020 (13 h): **Plant physiology laboratory lessons** for students of the biology bachelor's degree (Dept. Biosciences), histochemical module.

Abstract

English version

Being sessile organisms, plants cannot escape stress and must constantly cope with several environmental challenges, such as light excess and exposure to microorganisms. The enzyme formate dehydrogenase (FDH) catalyzes the oxidation of formate ($HCOO^-$) to carbon dioxide (CO_2) along with the reduction of NAD^+ to NADH. FDH has been described in the literature as a 'stress protein' because its expression is strongly influenced by unfavourable conditions; its induction has been well characterized under abiotic stress, but its role during pathogen attacks has been poorly studied. Therefore, this Ph.D. thesis investigates the response of FDH in *Arabidopsis thaliana* leaves when plants are exposed to either a vascular pathogen (*Xanthomonas campestris* pv *campestris*) or a beneficial rhizobacterium (*Pseudomonas simiae* WCS417), to further explore the role of FDH in plant-bacteria interactions. These investigations have been pursued using a 'hybrid' approach, in which *in silico* and *in vivo* strategies have been combined; to promote this type of research approach, researchers should improve the quality of existing data (Manuscript 1) and develop new, user-friendly informatics tools (Manuscript 2). A correlation analysis of *Arabidopsis thaliana* gene expression data under biotic stress showed that the top correlators of *FDH* are genes involved in defense responses. Furthermore, *in vivo* studies using a reporter construct driven by the *FDH* promoter activity in *Arabidopsis thaliana* leaves revealed that FDH may have an important role in the early defense response pathways involving hydathodes (specialized pores in plant leaves that secrete excess water in the form of droplets) following *Xanthomonas campestris* pv *campestris* infection (Manuscript 3). *Pseudomonas simiae* WCS417 is likely the best-studied rhizobacterium for its ability to promote an immune response called 'Induced Systemic Resistance' (ISR) in *Arabidopsis*

thaliana. *FDH* promoter activity was rapidly induced in hydathodes of plants colonized by *Pseudomonas simiae* WCS417 or exposed to the bacterium without direct contact with roots. A total leaf proteome analysis was then performed on wt Col and *atfdh1-5* knockout mutants colonized for a short time by the rhizobacterium. Such analysis showed an increase in FDH levels of wt leaves; it also gave indications of the altered leaf metabolic pathways/processes. Changes were mainly related to extrinsic photosystem proteins, stress-responsive proteins, and proteins involved in reactive oxygen species detoxification (Manuscript 4). Notably, ISR partially overlaps with the iron-deficiency response pathway, implicating that iron is not only essential in plant life but also an important element in plant-bacteria interactions (Manuscript 5 and Manuscript 6).

Italian version

Essendo organismi sessili, le piante non possono fuggire, ma devono costantemente far fronte e diversi stress, come per esempio eccessi di luce o la continua esposizione ai microrganismi. L'enzima formato deidrogenasi (FDH) catalizza l'ossidazione del formato ($HCOO^-$) in anidride carbonica (CO_2) insieme alla riduzione di NAD^+ in NADH. FDH è stata presentata in letteratura come una 'proteina di stress' in quanto la sua espressione è fortemente influenzata dagli stress ambientali; è ben documentata l'induzione in risposta a stress di natura abiotica, mentre il suo ruolo durante gli attacchi patogeni è stato poco studiato. In questa tesi di dottorato si è voluto quindi indagare la risposta di FDH in foglie di *Arabidopsis thaliana* quando le piante sono esposte a un patogeno del sistema vascolare (*Xanthomonas campestris* pv *campestris*) o a un batterio benefico della rizosfera (*Pseudomonas simiae* WCS417), con l'obiettivo di comprendere meglio il ruolo di FDH nell'interazione pianta-batterio. Si è deciso di utilizzare per le indagini un approccio 'ibrido',

in quanto si sono usate sia tecniche *in silico* che *in vivo*; al fine di favorire questa tipologia di approccio alla ricerca, i ricercatori dovrebbero migliorare la qualità dei dati già esistenti e disponibili (Manoscritto 1) e sviluppare nuovi programmi informatici facili da usare (Manoscritto 2). Un'analisi di correlazione dei dati di espressione genica di *Arabidopsis thaliana* sotto stress biotico ha mostrato che i principali correlatori di *FDH* sono geni coinvolti in risposte di difesa. Inoltre, l'esame *in vivo* dell'attività del promotore di *FDH* in foglie di *Arabidopsis thaliana* ha evidenziato come FDH potrebbe avere un ruolo importante nelle prime risposte di difesa degli idatodi (pori specializzati a livello fogliare per secernere l'acqua in eccesso) durante l'infezione da parte di *Xanthomonas campestris* pv *campestris* (Manoscritto 3). Molto probabilmente *Pseudomonas simiae* WCS417 è il batterio della rizosfera più studiato per la sua capacità di promuovere in *Arabidopsis thaliana* una risposta immunitaria nota come 'Induced Systemic Resistance' (ISR). Si è osservata una rapida induzione dell'attività del promotore di *FDH* negli idatodi di piante colonizzate da *Pseudomonas simiae* WCS417 o esposte al batterio ma senza un contatto diretto con l'apparato radicale; inoltre, un'analisi globale del proteoma fogliare a seguito di una breve colonizzazione delle radici di piante wt Col e mutanti knockout *atfdh1-5* da parte di *Pseudomonas simiae* WCS417 ha rivelato un aumento dei livelli di FDH nel wt e dato indicazioni sui processi e le vie metaboliche che vengono influenzate a livello delle foglie, in particolare quelli relativi a proteine estrinseche dei fotosistemi, proteine di risposta agli stress, e proteine con funzione di detossificazione di specie reattive dell'ossigeno (Manoscritto 4). Infine, la difesa ISR si sovrappone parzialmente alla risposta della pianta alla ferrocarenza, suggerendo che il ferro è non solo essenziale per la pianta ma che è anche un elemento importante nell'interazione con i microrganismi (Manoscritto 5 e Manoscritto 6).

Abbreviations

ABCG37: ATP-binding cassette G37

AMF: Arbuscular mychorrizal fungi

BGLU42: β-glucosidase 42

bHLH: Basic helix-loop-helix

CK: Cytokinin

ET: Ethylene

FDH: Formate dehydrogenase

Fe: Iron

FIT: Fer-like iron deficiency-induced transcription factor

FRO2: Ferric reduction oxidase 2

GA: Gibberellic acid

GC-MS: Gas chromatography-mass spectrometry

GEO: Gene expression omnibus

GSH: Glutathione

IAA: Indole-3-acetic acid/auxin

ISR: Induced systemic resistance

IRT1: Iron-regulated transporter 1

JA: Jasmonic acid/jasmonate

LC-MS: Liquid chromatography-mass spectrometry

Met: Methionine

MS: Mass spectrometry

MtGEA: *Medicago truncatula* gene expression atlas

NO: Nitric oxide

NGS: Next-generation sequencing

PIN1: Pin-formed 1

PDR9: Pleiotropic drug resistance 9

PGPR: Plant growth-promoting rhizobacteria

PGPF: Plant growth-promoting fungi

ROS: Reactive oxygen species

RNA-Seq: RNA-Sequencing

SA: Salicylic acid

SAR: Systemic acquired resistance

Ser: Serine

VOC: Volatile organic compound

WCS417: *Pseudomonas simiae* WCS417

Xcc: *Xanthomonas campestris* pv *campestris*

1. Introduction

"Research is to see what everybody else has seen,
and to think what nobody else has thought"
— Albert Szent-Gyorgyi

1.1 The model species

Most of the experimental research was conducted on the model plant species *Arabidopsis thaliana*; however, at the beginning of my Ph.D., I also worked *in silico* with microarray data from *Medicago truncatula*. Here, I will therefore briefly introduce these two model species.

1.1.1 *Arabidopsis thaliana*

Arabidopsis thaliana (also known as wall cress, mouse-ear cress, or simply *Arabidopsis*) is a small annual or winter annual vascular herbaceous plant belonging to the Brassicaceae (or Cruciferae) family with no application in agriculture (Krämer, 2015). *A. thaliana* has been successfully adopted as a plant model organism by the scientific community because it possesses traits that are in line with basic research needs, offering significant benefits for genetics and molecular biology (Provart *et al*., 2016). *A. thaliana* has a very compact genome (114.5/125 Mb, for a total of approximately 26500 genes) organized into five chromosomes that have been extensively physically and genetically mapped; genomic and transcriptomic analyses have been facilitated by the availability of complete genomic sequence for over 20 years (Kaul *et al*., 2000). *A. thaliana* is small (a fully grown plant achieves a maximum of 30-40 cm) and is composed, in its aerial part, of a rosette of leaves with margins that could be segmented; *A. thaliana* leaves possess hydathodes at the tips of the leaf teeth where guttation occurs (Cerutti *et al*., 2019). The flowering stem has a few cauline leaves that are smaller than the rosette leaves and have unsegmented margins.

The flowers are quite small (few mm in length), grouped in a corymb, like other crucifers; fruits are called 'siliques' and contain around 30-60 seeds each (each seed weighs around 20 µg) (Meyerowitz, 1987). At maturity, siliques are dry, dehiscent, and open to release seeds (Spence *et al*., 1996). *A. thaliana* has a short life cycle span (six to eight weeks under inducing conditions) and is found on all continents, with various ecotypes retrieved from very different habitats (Leonelli, 2007). Over the years, various databases, study platforms, and genetic and molecular tools have been developed for this plant, including TAIR (https://www.arabidopsis.org/) and Araport (https://araport.org/) (Huala *et al.*, 2001; Krishnakumar *et al.*, 2015; Pasha *et al.*, 2020). Moreover, several insertional mutant lines and other genetic resources (reporter lines and allelic variants from ecotypes) are available through the two main *Arabidopsis* seed stock centers, the *Arabidopsis* Biological Resource Center ABRC and the Nottingham *Arabidopsis* Stock Centre NASC, with approximately 750 natural *Arabidopsis* accessions obtained worldwide.

1.1.2 *Medicago truncatula*

The Fabaceae family (simply known as the 'legume family') is a major flowering plant group including several food/feed crops, some of which are important oil crops, as well as cover crops used to fix nitrogen. Several needs have triggered the search for a model legume, from understanding the molecular basis of symbiotic nitrogen fixation and mycorrhization to developing genetic tools (insertional mutagenesis and transformation) applicable to the improvement of economically important legumes (Young and Udvardi, 2009). Several species of the genus *Medicago* are commonly found in the Mediterranean region and their diffusion is dependent on temperature and/or soil pH (Lesins and Lesins, 1979; Bounejmate and Robson, 1992; Bounejmate *et al.*, 1994). In particular, *Medicago truncatula*, the barrel medic, has been used for several decades as a model legume in genetic and genomic studies (Barker *et al*., 1990; Cook, 1999). *M. truncatula* is a small annual autogamous plant native to the Mediterranean region. It has a short life cycle, low reproductive rate, and a small diploid genome (314 Mb in size, with eight chromosomes and 46016 predicted proteins), which allows easy manipulation of legume secondary metabolism (Young *et al.*, 2011; Gholami *et al*., 2014; Kang *et al*., 2016). Furthermore, similar to many other legumes*, M. truncatula* establishes symbiotic relationships with nodule-forming rhizobia and has been studied as a model organism for this symbiosis (Barker *et al*., 1990;

Graham and Vance, 2003; Young *et al*., 2011). Several legumes are unique in their ability to obtain nitrogen because of their mutually beneficial association with the bacteria in root nodules; they can also establish endomycorrhiza with arbuscular mycorrhizal fungi (AMF) (Smith and Read, 2008), which assist plant mineral nutrition because of the biotrophic exchanges between the two partners. For example, the model based on *M. truncatula* and the AMF *Glomus intraradices* is widely used as a tool for studying beneficial plant-fungal interactions. Transformation technologies (D'Erfurth *et al*., 2003; Rogers *et al*., 2009), the generation of vast mutant collections (Bénaben *et al*., 1995; Penmetsa and Cook, 2000), a complete and well-annotated genome (Tang *et al*., 2014), an expression atlas (Benedito *et al*., 2008; He *et al*., 2009), and a characterized proteome (Watson *et al*., 2003) have allowed *M. truncatula* to become a model legume. For example, several transcriptomic studies have been conducted on *M. truncatula* to identify differentially expressed genes in organs under different conditions. Benedito *et al*. (2008) produced a gene expression atlas that provided a global perspective of gene expression in all the main organs of this species, with a special emphasis on nodule and seed development, thus highlighting significant changes in the expression of critical regulatory genes such as transcriptional factors. Transcription controls a large number of physiological and developmental processes in plants, and the differential expression of genes in various organs and conditions may provide information regarding the function of these genes. RNA-sequencing (RNA-Seq) technologies have recently been used to define the transcriptome of *M. truncatula*, and bioinformatics analysis of RNA-Seq data often entails the mapping of the RNA sequence before measurement of gene expression (Boscari *et al*., 2013).

1.2 *In silico* analyses

Microarrays are still frequently used in plant science, and large amounts of transcriptomic data are freely available in general and specific databases, such as the Gene Expression Omnibus (GEO) (Edgar, 2002; Barrett *et al*., 2013) and the *Medicago truncatula* microarray database (MtGEA) (Benedito *et al*., 2008; He *et al*., 2009). Databases integrating datasets and sequences with public domain contextual data (such as gene expression, protein domains, and metabolic pathways) into one system can greatly help the scientific community. In addition, new tools and software that can support and facilitate bioinformatics analyses have the potential to stimulate research. During my first Ph.D. year, I focused on *in silico* research, mainly because of work restrictions caused by the SARS-CoV-2 pandemic. To this end, I honed my knowledge of the R programming language and, by using it, I developed a strategy for cleaning the MtGEA database and designed an application (named NORMALIX95) for plant microarray data processing, the development of which lasted for the whole duration of my Ph.D. program.

1.2.1 The omics technologies

The term 'bioinformatics' was first used in 1970 to refer to "the study of information processing in biological systems" (Meijer, 2021); it is an emerging research field that uses computational tools, mathematics, and statistical approaches to answer biological questions. Bioinformatics allows the management of a massive amount of data because it has developed together with technologies allowing the production, quickly and affordably, of a huge amount of omics data. The suffix '-omics' derives from the Greek suffix '-ome', which stands for 'all' or 'whole': 'omics' identifies a wide range of biological disciplines (such as genomics, transcriptomics, proteomics, lipidomics, metabolomics, and epigenomics) referring to the comprehensive, high-throughput analyses of genes, RNAs, proteins, lipids, metabolites, and methylated DNAs and changed histone proteins in chromosomes, respectively. Omics disciplines study the totality of these cellular components, which are now known as genomes, transcriptomes, proteomes, lipidomes, metabolomes, and epigenomes, to gain a complete overview of the biological system analyzed (Hasin *et al.*, 2017; Nalbantoglu and Karadag, 2019; Subramanian *et al*., 2020). Indeed, all the information from omics disciplines is combined to obtain a 'bigger picture' in biology, and data are

collected and analyzed to establish new hypotheses (Breitling, 2010; Tavassoly *et al*., 2018). Currently, our understanding of molecules at the base of life has greatly expanded using platforms such as Next-Generation Sequencing (NGS) for genomics and transcriptomics, mass spectrometry (MS) coupled with 2D-Gels for proteomics, and liquid chromatography-mass spectrometry (LC-MS) or gas chromatography-mass spectrometry (GC-MS) for metabolomics (Rabilloud, 2014; Wickett *et al*., 2014; Alonso-Blanco *et al*., 2016; Jorge *et al*., 2016). The amount and variety of data obtained through omics methodologies will increase daily, and improved analyses will be possible because computational methods allow for better integration of -omics data, which remains a major challenge in bioinformatics (Redestig and Costa, 2011; Gomez-Cabrero *et al*., 2014; Tzin *et al*., 2019).

Genomics is the most developed omics (McGuire *et al*., 2020). The genome does not change much over time, excluding mutations and chromosomal rearrangements; therefore, in the medical field, genomics focuses on variations related to illness, treatment, and future patient prognosis. Additionally, there is growing interest in investigating the influence of human microbiota on health through genomics technologies (O'Mahony, 2015). Another well-established omics technology is transcriptomics, which provides details on gene expression and function of gene products under a given physiological condition and is useful in guessing the roles of unannotated genes (Heller, 2002; Dong and Chen, 2013; Lowe *et al*., 2017; Zhang, 2019). The first attempt to study the transcriptome was made in the early 1990s, and transcriptomics has developed rapidly since the late 1990s owing to several technological advancements, including microarrays, which measure a set of preset sequences, and RNA-sequencing (RNA-Seq), which uses high-throughput sequencing to establish the identity of millions of RNA sequences, even in single cells. Nevertheless, both types of investigations are highly parallel and require extensive computational power to obtain useful information (Lowe *et al*., 2017).

1.2.2 Microarrays

Microarrays are a transcriptomic technology that allows the global analysis of RNA expression in biological systems (Ekins and Chu, 1999; Gardiner-Garden and Littlejohn, 2001). The development of microarray platforms began in the late 1990s and the early 2000s, and 'base complementarity' is the main idea on which microarrays rely because microarray

platforms hybridize polynucleotide strands to generate duplexes (Brown and Botstein, 1999). Small nucleotide oligomers, named 'probes', are synthesized and arranged in an organized pattern (array) on a solid substrate such as glass (Romanov *et al.*, 2014). Thousands of identical probes are gathered in microscopic regions known as 'spots' or 'features', and transcript levels are measured by the amount of duplex molecules bound at each spot. Several reviews have provided detailed and comprehensive descriptions of microarray technologies and their buildings (Ekins and Chu, 1999; Hoheisel, 2006; Bumgarner, 2013; Jaksik *et al*., 2015; Lowe *et al*., 2017). Microarrays are typically classified into two categories: spotted low-density and high-density short-probe arrays, with the former normally being longer than the latter (Heller, 2002). The *Affymetrix* GeneChip array (Santa Clara, CA, https://www.thermofisher.com/it/en/home/life-science/microarray-analysis.html) has popularized high-density microarrays, in which each transcript is quantified by several 25-mer probes that together identify one gene (Irizarry *et al*., 2003a; Dalma-Weiszhausz *et al*., 2006). To build microarray probes, prior knowledge of the target organism is required, such as the annotated genome sequence (Lowe *et al*., 2017); however, tools to produce microarrays have greatly improved over the years, to the point that they are no longer considered limiting factors. For example, *Affymetrix* manufacturing avoids errors caused by the preparation of several cDNAs and the use of several different probes capable of hybridizing to various regions of the same reference sequence, allowing for the development of a large collection of microarrays for use in gene expression analysis, genotyping, and sequencing (Chee *et al.*, 1996; Lockhart *et al.*, 1996; Dalma-Weiszhausz *et al.*, 2006; Bumgarner, 2013). Many scientific fields still extensively use microarrays, even though their death was predicted since 2008 (Ledford, 2008). RNA-Seq is indeed gaining favor compared to arrays because of the falling cost of the technology, the potential to identify novel transcripts without the need for specific probes, and the ability to detect a higher number of weakly expressed genes or rare transcripts by increasing the sequencing coverage (Wang *et al*., 2009, 2014; Liu *et al*., 2015b; Stark *et al*., 2019). Nevertheless, it is important to remember that some steps involved in microarray technologies are also used in other techniques, including RNA-Seq methods (Shendure, 2008; Zheng *et al*., 2011). Furthermore, microarrays are practical and inexpensive tools for pilot studies or research involving many samples (Eijssen *et al*., 2013). There is a multitude of microarray data to be investigated, and many unique datasets continue to appear in the literature. Most microarray data can be easily found online, for instance, in the Gene Expression Omnibus repository (GEO, https://www.ncbi.nlm.nih.gov/geo/) (Edgar, 2002; Barrett *et al*., 2013) or specific

dedicated websites such as The *Arabidopsis* Information Resource (TAIR, https://www.arabidopsis.org/) (Huala *et al*., 2001) or the *Medicago truncatula* Gene Expression Atlas (MtGEA) (Benedito *et al*., 2008; He *et al*., 2009). Unfortunately, plant databases are not always well maintained or up-to-date, and there may be mistakes or poor quality data compromising successive analyses. Therefore, we aimed to improve the quality of the MtGEA database by building a generic cleaning approach based on logical and statistical correlations between variables and circumstances and developing an easy strategy to clean other plant microarray databases (Marzorati *et al*., 2021b). This approach is of general validity and can be applied to other organisms and datasets.

Data analysis and interpretation are currently considered the biggest barriers in microarray analysis (aside from potential quality problems), partly because of the lack of standard nomenclature and data storage in a query-friendly format (Hoheisel, 2006). The microarray results provide biologists with enormous amounts of complex digital data that must be transformed into useful biological information. Specific steps must be followed to achieve this goal, including quality control, normalization, statistical analysis, gene grouping, and classification (Olson, 2006; Naidu and Suneetha, 2012; Slonim and Yanai, 2009; Bumgarner, 2013). Normalization is a crucial step in microarray analysis, as it corrects microarray data for the effects related to technical variabilities, such as the methods used, to highlight biologically significant differences. A relevant problem in choosing a normalization strategy for microarray data is that there is no 'reference method' to which expression values may be compared; several studies have tried to compare normalization strategies to identify the best to use, but results are not always in accordance (Bolstad *et al*., 2003; Irizarry *et al*., 2003b; Millenaar *et al*., 2006). Thus, choosing an acceptable statistical method for microarray processing is a significant topic of discussion because the chosen normalization technique may significantly impact the results of the studies (Harr and Schlötterer, 2006; Jaksik *et al*., 2015). New algorithms may be created, tested, and distributed via open-source projects such as R/Bioconductor (https://www.bioconductor.org/) (Gentleman *et al*., 2004; Huber *et al*., 2015). However, several researchers may not have the computing abilities necessary to fully utilize these free resources (Choi and Ratner, 2019; Reyes *et al*., 2019). Some commercial companies have also created software to carry out all steps of transcriptomic analysis, but these tools can be particularly expensive, outdated, and difficult to use (Eijssen *et al*., 2013; Choi and Ratner, 2019). New user-friendly tools dedicated to microarray analyses should be made freely

available to the scientific community to facilitate the handling of large amounts of microarray data. Given these premises, I developed an application called NORMALIX95 to support plant scientists without computational skills in microarray analyses (Marzorati *et al*., manuscript in preparation). The script of NORMALIX95 is implemented with several R packages and activities useful to normalize and analyze, for eleven plant species, the massive amount of *Affymetrix* microarray data available online. One important feature of NORMALIX95 is the possibility of performing correlation analyses on uploaded datasets, which may help test the strength of the relationship between variables and identify novel genes involved in metabolic pathways and processes based on a 'guilt-by-association approach' (Altshuler *et al*., 2000; Månsson *et al*., 2004; Beekweelder *et al*., 2008; Berri *et al*., 2009; Fukushima *et al*., 2011; Abbruscato *et al*., 2012; Zermiani *et al*., 2015; Murgia *et al*., 2020; Marzorati *et al*., 2021b). This type of analysis was pivotal for my investigation of the role of the enzyme formate dehydrogenase (FDH) under biotic stress, as the starting point for hypothesizing a role of FDH in leaf plant defense responses.

1.3 Formate dehydrogenase: *old but gold*

Formate dehydrogenase (FDH) is an enzyme first identified a century ago, coupling the oxidation of formate to carbon dioxide CO_2 with the reduction of an electron acceptor such as NAD^+; FDHs are found in a variety of organisms, ranging from prokaryotes to eukaryotes. FDH plays an important role in plant stress responses, and its role in abiotic stress responses has been well studied. However, its involvement in biotic stress, namely defense against pathogens and pests, has been poorly investigated. I began my work on FDH using an *in silico* approach: I performed a correlation analysis on a biotic stress dataset built for *Arabidopsis thaliana* to identify the top correlators of FDH under such conditions.

1.3.1 Formate dehydrogenase FDH

The pathways of one-carbon (C1) metabolism have been identified in all organisms and play a role in the production and regulation of many essential metabolites. However, this metabolism has been poorly investigated mainly because of experimental difficulties (Hanson and Roje, 2001; Fox and Stover, 2008; Suh *et al*., 2016). Formate ($HCOO^-$) is a one-carbon compound used by a wide range of prokaryotes and eukaryotes as a substrate or product of various metabolic processes. For instance, it promotes the development of several facultative chemolithotrophic anaerobic bacteria but is also used as a single source of energy by methylotrophic aerobic microorganisms (Ferry, 1990; Maia *et al*., 2015). Higher plants contain small pools of formate (Hanson and Roje, 2001); the formate pool size in *Arabidopsis thaliana* is approximately 73 nmol $(g\ FW)^{-1}$, which is comparable to that of barley (Wingler *et al*., 1999). Plant formate metabolism is closely related to serine (Ser) biosynthesis and all subsequent reactions originating from this amino acid (Amory and Cresswell, 1986; Hourton-Cabassa *et al*., 1998; Igamberdiev *et al*., 1999; Hanson and Roje, 2001; Igamberdiev and Eprintsev, 2016; Igamberdiev and Kleczkowski, 2018). However, other plant metabolic pathways, such as the methionine (Met) cycle or detoxification of formaldehyde by glutathione (GSH), may produce formate (Hourton-Cabassa *et al*., 1998; Achkor *et al*., 2003; Song *et al*., 2013). Leaves can release formate into the atmosphere (Gabriel *et al*., 1999), and a significant amount of formate is directly oxidized to CO_2 (Li *et al*., 2003; Bar-Even *et al*., 2012). Formate dehydrogenases (FDHs) isoforms are found in bacteria, fungi, and plants; they catalyze the oxidation of formate to CO_2 by transferring

electrons to either cytochrome or NAD^+. They are typically classified into two main groups according to their structural characteristics, the strategy of catalysis, and the presence of prosthetic groups (iron-sulfur clusters, molybdenum, and tungsten ions) in their active sites (Popov and Lamzin, 1994; Alekseeva *et al*., 2011). Cytochrome-dependent FDHs are mainly found in anaerobic microorganisms and archaea and are characterized by complicated quaternary structures, high molecular weights, several prosthetic groups (including molybdenum cofactor Moco), and elevated sensitivity to oxygen O_2 (Ferry, 1990; Alekseeva *et al*., 2011). NAD^+-dependent FDHs are the only FDHs found in plants and are characterized by the absence of prosthetic groups (Alekseeva *et al*., 2011). Plant FDHs are localized in the mitochondrial matrix; however, their localization in the plastids of *A. thaliana* has also been documented (Olson *et al*., 2000; Herman *et al*., 2002; Lee *et al*., 2022). The possible reduction of CO_2 to formate, as observed in some prokaryotes, has been proposed for plants (Olson *et al*., 2000); however, it is now widely accepted that plant FDH is not primarily involved in the assimilation of CO_2 (Igamberdiev *et al*., 1999). The purified enzyme is a 42 kDa dimer but, in organelles, it is part of a 200 kDa complex that may include additional proteins (Jänsch *et al*., 1996) and is one of the most abundant proteins in plant mitochondria of heterotrophic tissues (Jänsch *et al*., 1996; Hourton-Cabassa *et al*., 1998), especially under prolonged stress.

1.3.2 FDH and stresses

Plant FDH is commonly described as a ‘stress protein’ (Alekseeva *et al*., 2011) because its expression is severely altered under various environmental stresses such as drought and hypoxia (Hourton-Cabassa *et al*., 1998; Andreadeli *et al*., 2009), prolonged darkness (Hourton-Cabassa *et al*., 1998; David *et al*., 2010), high and low temperatures (Hourton-Cabassa *et al*., 1998; Ren *et al*., 2009), and chemical toxicity (Li *et al*., 2002; Lou *et al*., 2016; Kurt-Gür *et al*., 2018). Iron (Fe) deficiency can induce FDH accumulation in barley roots (Suzuki *et al*., 1998), and FDH has recently been proposed as a hub for molybdenum and Fe homeostasis in plant nutrition (Vigani *et al*., 2017; Murgia *et al*., 2020; Di Silvestre *et al*., 2021). Thus, FDH appears to be a pivotal enzyme in plant responses to abiotic stress; however, only a few studies have investigated the role of this enzyme in plants under biotic stress. When infected with avirulent *Pseudomonas syringae* pv *tomato Pst* DC3000 (avrRpm1), *Arabidopsis thaliana fdh* mutants displayed enhanced disease

symptoms, suggesting that FDH may be involved in defense responses against pathogenic bacteria (Choi *et al*., 2014). Moreover, an increase in the FDH transcript level was observed during the infection of *Phaseolus vulgaris* (common bean) with the fungus *Colletotrichum lindemuthianum* (David *et al*., 2010). We recently proposed FDH as a possible initial defender of *A. thaliana* against the vascular pathogen *Xanthomonas campestris* pv *campestris* (*Xcc*), which enters leaves via hydathodes, as the leaves of the *atfdh1-5* mutant appear to be more susceptible to *Xcc* infection (Marzorati *et al*., 2021a). This behavior has also been documented in other *fdh* mutants infected with bacterial pathogens (Lee *et al*., 2022). A model showing FDH as a potential hub for interactions between Fe homeostasis in plants and plant defense responses is shown in **Figure 1** (Murgia *et al*., 2022). However, no molecular model of FDH action under stress conditions, particularly during a pathogen attack, has been proposed yet. Changes in *FDH* expression suggest that FDH may play a role in reducing the adverse effects of environmental stress; a greater understanding of this enzyme and its physiological role would help to better understand the stress responses in plants, considering the complex environment around them, which includes a plethora of microorganisms.

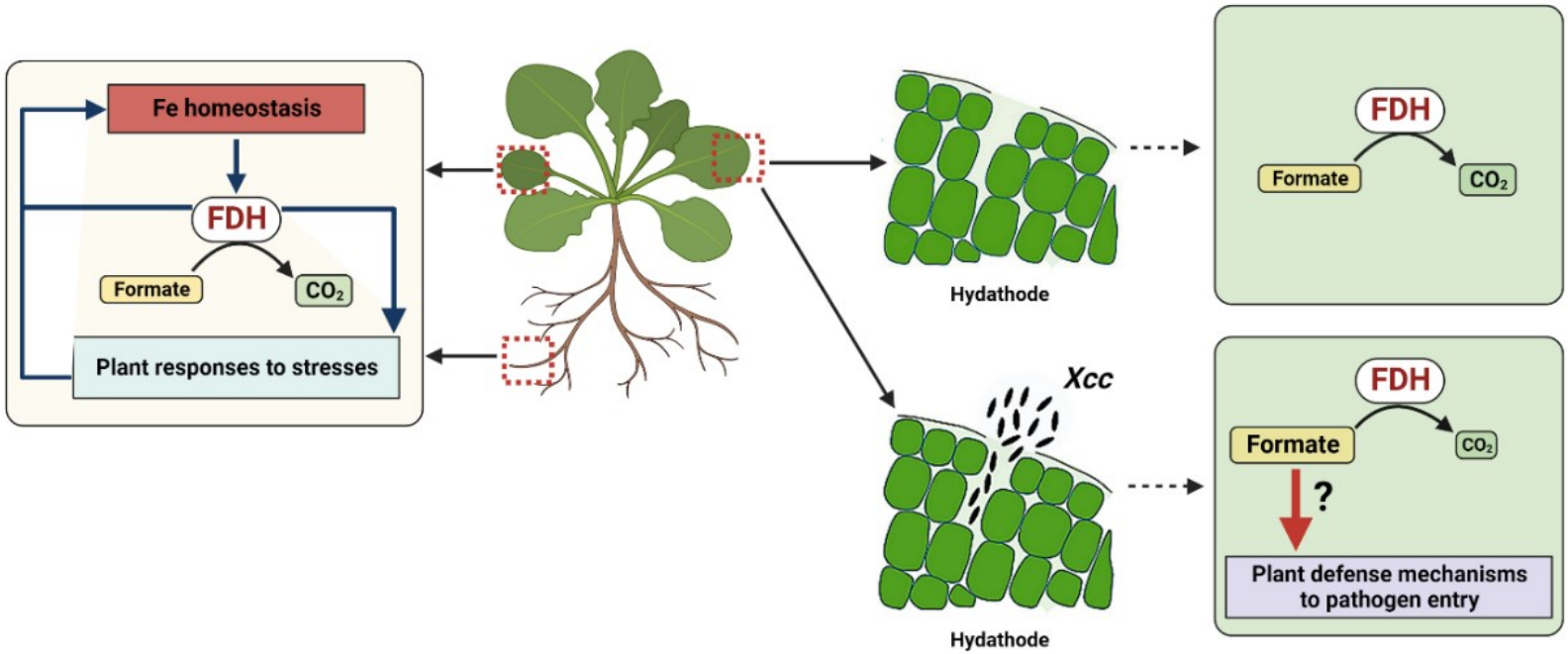


Figure 1. Formate dehydrogenase FDH is a hub for iron nutrition and a node for multiple interactions between iron homeostasis and stress response. Experiments on *Arabidopsis thaliana* roots and aerial parts support the possible role of FDH as a hub in plant iron (Fe) nutrition, involved in a loop regulation of Fe homeostasis and responses to abiotic stress (on the left). The *FDH* promoter activity in *A. thaliana* leaves is inhibited by *Xanthomonas campestris* pv *campestris* (*Xcc*) attack, possibly increasing formate levels (on the right); formate may function as a potential signal for plant defense responses against the pathogen. Figure taken from Murgia *et al*. (2022).

1.4 Crying plants: hydathodes and pathogens

Hydathodes are always open 'pores' located on leaf margins in correspondence with the terminal end of xylem conduits; they represent a natural entry point into the leaf for a few vascular pathogens, such as the bacterium *Xanthomonas campestris* pv *campestris* (*Xcc*). In the following paragraphs, I will briefly introduce the hydathodes and the bacterium *Xcc*; after the *in silico* analysis suggested that FDH is involved in biotic stress and defense against pathogens, I studied the regulation of *FDH* promoter activity in leaves of *Arabidopsis thaliana* infected with *Xcc.*

1.4.1 Hydathodes

Throughout their life, plants require a constant supply of water and essential nutrients that are necessary for their growth and development. Both need to be transported to the aerial parts of plants; transpiration and, to a minor extent, root pressure work together to guarantee the transport of water and nutrients from the root apparatus to the aerial parts of the organism through the xylem (*i.e.*, the plant vascular system). When the rate of transpiration declines (*e.g.*, when stomata are closed, such as in darkness or when CO_2 and/or humidity are high) and water in the soil is abundant, the flow inside the xylem is mainly due to root pressure, leading to a possible excess of water within the leaf and the risk of overflow in the intercellular space. Indeed, because of this poor transpiration, root pressure provides more liquid than water can evaporate, resulting in a liquid build-up in the leaf and flooding of intercellular spaces (Cerutti *et al*., 2019 and references therein). The hydathodes are epidermal organs, found on the leaf margins, where excess pressure in xylem vessels is discharged by releasing fluid droplets; this phenomenon is known as 'guttation' and it prevents the dangerous leaf flooding described above (Grunwald *et al*., 2003). The xylem liquid brought to leaves, also known as 'xylem sap', is rich in mineral ions (including calcium, potassium, iron, and magnesium), carbohydrates, hormones, sugars, vitamins, proteins, and amino acids (such as aspartate and histidine), and its composition depends on the plant and environment (Goatley and Lewis, 1966; Sheldrake and Northcote, 1968; Mizuno *et al*., 2002; Grunwald *et al*., 2003; Pilot *et al*., 2004; Cerutti *et al*., 2019). Droplets of guttation fluid may be reabsorbed inside the leaf when conditions become unfavorable for guttation (*e.g.*, when transpiration is increased). Furthermore, active uptake of guttation fluid

has also been observed, probably to prevent the loss of beneficial nutrients and to favor the release of toxic chemicals; hydathodes thus resemble mammalian kidneys in their function because they avoid the loss of beneficial compounds (Cerutti *et al*., 2019).

Hydathodes can be distinguished into 'active hydathodes and 'passive/epithemal hydathodes' (Haberlandt, 1914). Since their characterization over 100 years ago, epithemal hydathodes have been described in a variety of species, including several dicotyledons, monocotyledons, pteridophytes, and some semiaquatic and aquatic plants (Chen and Chen, 2006). They are made by 'epithem cells', *i.e.*, living cells composing the parenchyma with multiple meatuses and gaps filled by water, with few or no chloroplasts, and having one or more sub-epidermal chambers which they open into (**Figure 2**). They are localized in front of the vascular terminations of leaves along the leaf border, at the leaf apex, and at the tip of the leaf teeth. In *Arabidopsis thaliana*, the number of hydathodes increases in rosette upper leaves, together with the number of leaf teeth (Kawamura *et al*., 2010; Cerutti *et al*., 2019). The tips of leaf teeth are also the main site of auxin (indole-3-acetic acid, IAA) biosynthesis and control, and high levels of IAA in leaf teeth and hydathodes are guaranteed by the auxin efflux carrier PIN-FORMED 1 (PIN1) (Aloni *et al*., 2003; Scarpella *et al*., 2006; Bilsborough *et al*., 2011). Several studies have demonstrated that changes in auxin production and transport can greatly impair hydathode development, and IAA production is greatly favored by the high expression levels of *YUCCA* (*YUC*) family genes in hydathodes, particularly *YUC2* and *YUC4*, which encode flavin monooxygenase-like enzymes (Hay *et al*., 2006; Kawamura *et al*., 2010; Wang *et al*., 2011). Being always open, passive hydathodes appear as an interface between the environment and the xylem, as they expose the uncovered vascular terminals to the outside. For this reason, hydathodes are referred to in the literature as the 'Achille's heel' of plant vascular immunity (Cerutti *et al.*, 2019) and a few pathogens evolved to infect plants through these organs: black rot disease in Cruciferae, bacterial blight in rice, and canker in tomato are just a few examples of diseases that depend on hydathode infection (Williams, 1980; Carlton *et al*., 1998; Zhang and Wang, 2013). Typical symptoms observable once plants are infected by a vascular pathogen include necrotic lesions with a characteristic V-shaped pattern starting from the leaf margin, as in the case of black rot disease in crucifers infected with *Xcc*.

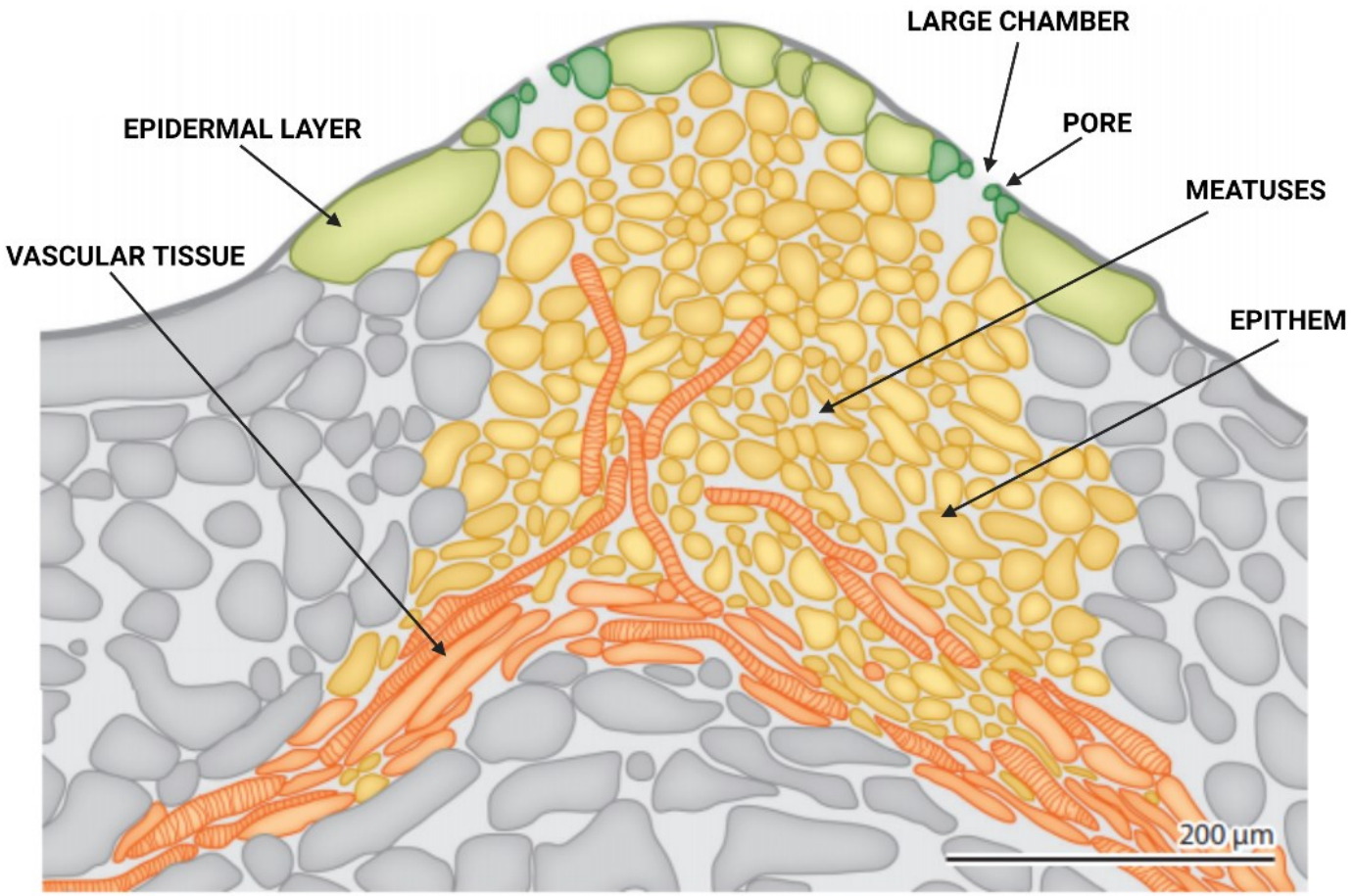


Figure 2. Schematic representation of the anatomy of *Arabidopsis thaliana* ecotype Col-0 epithemal hydathode. Figure adapted from Cerutti *et al*. (2019).

1.4.2 *Xanthomonas campestris* pv *campestris*

The leaf surface, which lacks nutrients and water, is a hostile environment for microorganisms, whereas the guttation fluid is rich in chemicals. Hydathodes release large amounts of guttation fluid, particularly in the early morning; as the guttation fluid recedes, it also pushes microbes into the hydathodes. Therefore, guttation fluids released by hydathodes likely allow microorganisms to move, feed, and reproduce in hydathodes before entering the xylem (Cerutti *et al.,* 2019). *Xanthomonas campestris* pv. *campestris* (*Xcc*) is a vascular pathogen that predominantly infects plants through hydathodes, and is the causative agent of black rot disease in crucifers (Williams, 1980; Gupta *et al*., 2013). *Xcc* is a gram-negative aerobic rod-shaped bacterium with a single polar flagellum that does not produce spores. Its optimal growth temperature is 30°C, which is typical of areas in which *Xcc* can rapidly spread; under favorable conditions, *Xcc* rapidly propagates, particularly during seedling production (Berg *et al*., 2005). Black rot disease is transmitted through seeds; *Xcc* can survive in infected seeds for up to three years, but *Xcc* can also be found in infected plant waste and leaves for months (Gupta *et al*., 2013). The first report of plants attacked by *Xcc* was described in Iowa (USA) (Pammel, 1895), and the presence of *Xcc* has been then reported in tropical, subtropical, and humid areas worldwide (Gupta *et al*., 2013). Several important crops, such as cabbage, radish, and cauliflower, can be infected by *Xcc* at any growth stage; the symptoms are mostly observable in leaves as bacteria advance down the vascular system and characteristic V-shaped necrotic lesions spread from leaf margins (Cook *et al*., 1952; Sutton and Williams, 1970). In the xylem, *Xcc* produces an extracellular polysaccharide called ‘xanthan’, which plugs the xylem vessels, prevents water flow, and leads to characteristic lesions on leaves (Williams, 1980). *Xcc* spread can barely be contained once the infection starts, causing catastrophic crop losses by premature defoliation, lowering the overall quality of the cultivars, and making them unmarketable. Black rot management mainly focuses on preventative measures such as the treatment of seeds with hot water, antibiotics, fungicides, and biological and chemical control strategies (Gupta *et al*., 2013). Furthermore, the use of various genetic and biotechnological approaches such as optimization of breeding strategies (Allier *et al*., 2020), advancements in transgenic technology (Tohidfar and Khosravi, 2015), and genome editing via CRISPR/Cas9 technology (Borrelli *et al*., 2018) with the aim of removing susceptibility genes, introducing novel receptors for pathogen detection, or degrading important compounds for pathogen survival, could be valuable ways to develop novel resistant varieties.

1.5 The plant holobiont

The term ‘microbiota’ identifies the community of microbial cells in a given environment, whereas ‘microbiome’ is the collection of all their genomes. Dense and diverse collections of microbial communities belonging to any of the three domains of life can be found both on and inside hosts, influencing their development, physiology, immunology, and metabolism. The phytohormone jasmonic acid (JA) plays a key role in systemic defense responses in leaves, which are stimulated by the action of the plant root microbiota; I identified several genes involved in defense responses among the top correlators of FDH under biotic stress, particularly those related to JA. Hence, I conducted a thorough literature investigation of the plant root microbiota and its components by focusing specifically on the bacterial strain *Pseudomonas simiae* WCS417.

1.5.1 The root microbiota

The term ‘holobiont’ was coined in 1991 by Lynn Margulis, an American evolutionary theorist well known for her symbiosis hypothesis. She defined a holobiont as a biological entity consisting of a host and a single hereditary symbiont and compared the interaction between the host and the symbiont with that of an egg and a sperm combining to generate a new creature (Margulis, 1991). The concept ‘hologenome’ was then proposed to describe the genetic content of a holobiont, that is a host genome, genomes present in organelles (such as mitochondria), and symbiotic bacterial genomes (Jefferson, 1994): with the hologenome concept, the holobiont gained new qualities, as it becomes an organism coherent enough to have its own genome (Douglas and Werren, 2016). In the early 2000s, the holobiont gained an ecological interpretation since the term holobiont was used to describe the dynamics of coral physiology, as many individuals were engaged in the existence and survival of corals (such as polyps, algae, bacteria, and viruses), and two biological entities (*i.e.*, the host and the single hereditary symbiont) were insufficient to describe their development (Rohwer *et al*., 2002). However, Margulis’s physiological interpretation was revived in the late 2000s as part of a new theory called the ‘hologenome theory of evolution’ (Zilber-Rosenberg and Rosenberg, 2008): phenotypes are now recognized as the outcome of the integrated expression of the host and its associated microbes (Vandenkoornhuyse *et al.*, 2015; Sánchez-Cañizares *et al.*, 2017; Simon *et al.*,

2019), and since natural selection acts on the phenotype and the phenotype of the holobiont is determined by the expression of the host and microbiota genomes, natural selection can also affect the hologenome. The hologenome theory of evolution is a hot topic among scientists. One of the greatest challenges with this hypothesis is the accurate definition of the holobiont and hologenome, because microorganisms are more or less shared among individual hosts, and they also spend time in the environment. This aspect raises an essential issue that is yet to be answered: may the holobiont (and hologenome) idea be scaled up to enclose larger ecosystems such as seas, woods, and even cities (Morris, 2018)? Although there are still many challenges in properly defining holobiont and hologenome, these concepts have become popular in environmental and health sciences because plants and animals live together with internal or external, small or large, microbial communities (Lynch and Pedersen, 2016; Skillings, 2016). These microorganisms affect the nutrition, growth, immune system, and behavior of their hosts. However, growing data have revealed that a variety of host-related and environmental variables affect the microbiota. In general, when all members in a relationship benefit from the partnership, we talk about 'mutualism' (+/+); in case one partner benefits from the interaction whereas the other organism suffers no gain or damage, we are in a neutral condition called 'commensalism' (+/0). When only one organism benefits from the association, damaging the other, one speaks of parasitism (+/-). It is well known that different plant parts serve as ecological niches housing different sets of microbes, particularly the roots; microorganisms living within or nearby roots normally act as mutualists (Pascale *et al.*, 2020; Chialva *et al.*, 2022). For instance, the interaction between plants and arbuscular mycorrhizal fungi (AMF) represents one of the most researched symbioses in plants, probably evolving up to 450 million years ago and characterizing 80% of terrestrial plants (Cameron *et al*., 2013).

Microbial communities in the soil are extremely varied, including tens of thousands of species. Soil has the highest microbial diversity on Earth, with estimates of up to 10^{10} microorganisms per cm^3 (Torsvik and Øvreås, 2002), and this value increases if we specifically refer to the 'rhizosphere' (the 1-3 mm thick layer of soil around roots whose microorganisms are affected by their secretions), which contains up to 10^{11} microorganisms per gram of root (Berendsen *et al*., 2012; Sasse *et al*., 2018). The plant root microbiota differs from the bulk soil microbes according to a phenomenon known as the 'rhizosphere effect', which is the stimulation of a soil microorganism's development caused by physical and chemical changes in the soil (Bakker *et al*., 2013, 2020; Prashar *et al*., 2014) (**Figure 3**).

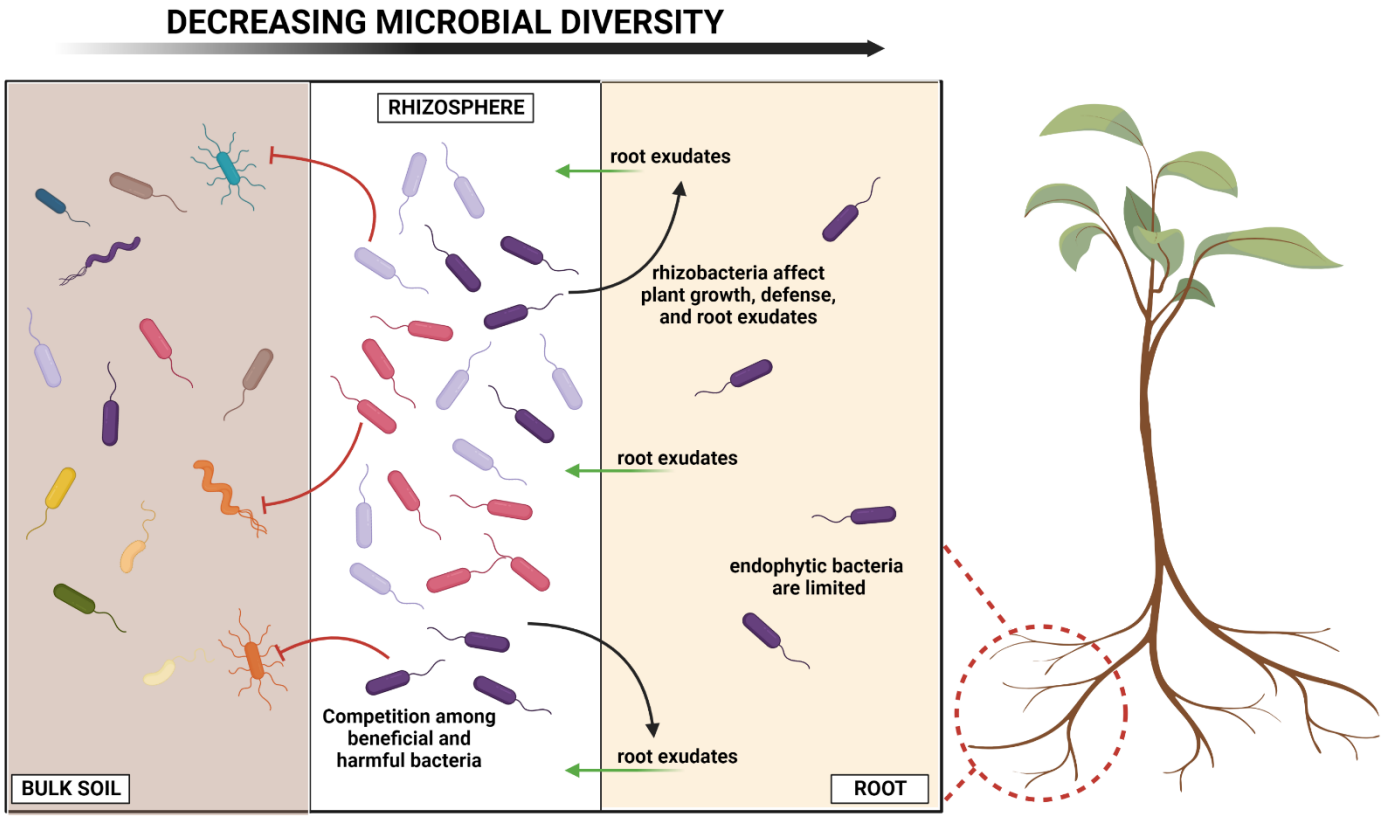


Figure 3. Bacterial communities in the bulk soil, the rhizosphere, and inside the plant. In the rhizosphere (the thin layer of soil around roots), the microbial population is less diversified but more abundant than the bacterial communities found in the bulk soil (on the left). The microbial community found inside the roots ('endophytic bacteria', on the right) is an extremely limited population. Bacteria can influence various aspects of plant health, growth, and life, as well as root exudate composition (black and green arrows). Moreover, intense competition among bacteria for space and nutrients occurs in soil (red arrows). Figure created by using BioRender.com

Microbial ecologists have been trying to determine which variables shape soil microbial communities for a long time, and a substantial amount of data on the structure, dynamics, and functional capacity of plant microbiota is now available (Müller *et al*., 2016; Pascale *et al*., 2020; Chialva *et al*., 2022). Many host-related and environmental factors, such as plant genotype, plant developmental stage, and soil type, can influence the composition of the plant microbiota (Berendsen *et al*., 2011, 2012; Trivedi *et al*., 2020). Plants play an important role in this process by changing the conditions of the rhizosphere through the secretion of several chemicals that can attract or ward off microbes (Feng *et al*., 2021). These compounds include both primary and secondary metabolites, such as amino acids, organic acids, phenolics, flavonoids, fatty acids, and plant hormones (cytokinin (CK), ethylene (ET), gibberellic acid (GA), and auxin (indole-3-acetic acid, IAA)) as well as enzymes. Hormones can also affect the composition of microbial communities found within the roots ('endophytic bacteria'), which are less varied than those in the rhizosphere (Afzal *et al*., 2019). This low variability, as well as the reduced number of endophytic microorganisms, may be attributed to physical barriers (such as the cell wall) and the plant immune system because changes in hormone signaling pathways affect the composition of endophytic microorganisms (Lebeis *et al*., 2015; Xu *et al*., 2018). Root exudates not only change the composition of microbial communities in the rhizosphere but can also affect specific microbial functions, such as biofilm development, antibiotic synthesis, or symbiotic interactions (Bakker *et al*., 2013). Elucidation of the effects of root exudates on the development of the rhizosphere microbiota could be important for the various benefits that such knowledge would imply. For example, by favoring a 'beneficial root microbiota', the use of chemical fertilizers or pesticides could be reduced, thus improving agricultural sustainability (Sun *et al*., 2021; Jamil *et al*., 2022).

Microbes in the rhizosphere can have a significant impact on plant life and development, for example by favoring plant resistance to pathogens and enhancing plant yield and growth (Nihorimbere *et al*., 2011; van de Mortel *et al*., 2012; do Amaral *et al*., 2020). The idea that certain microorganisms may particularly improve plant fitness, development, and resistance comes from the definition in 1974 by Baker and Cook of the phenomenon of 'disease suppressive soils', *i.e.*, "soils in which the pathogen does not establish or persist, establishes but causes little or no damage, or establishes and causes disease for a while but thereafter the disease is less important, although the pathogen may persist in the soil" (Expósito *et al.*, 2017 and references therein). Microbes colonizing the

rhizosphere can significantly alter the composition of root exudates (Dardanelli *et al*., 2010; Matilla *et al*., 2010; Korenblum *et al*., 2020), promote shoot growth (Veresoglou and Menexes, 2010; Carvalhais *et al*., 2013; Minorsky, 2019), and favor root development, thereby improving the ability of root exploration in soil (Vacheron *et al*., 2013; Zamioudis *et al*., 2013; Verbon and Liberman, 2016; Wintermans *et al*., 2016) and nutrient assimilation (Ferguson and Mathesius, 2014; Soyano *et al*., 2014; Garcia *et al*., 2020). They can also induce defense responses, such as a plant immune system response known as 'Induced Systemic Resistance' (ISR) (Lugtenberg and Kamilova, 2009; Pieterse *et al*., 2014; Verbon *et al*., 2017). These microorganisms promoting plant fitness are named 'plant growth promoting rhizobacteria' (PGPR) and 'plant growth promoting fungi' (PGPF) (Beneduzi *et al*., 2012; Vacheron *et al*., 2013; Hossain *et al*., 2017); to date, most plant microbiota surveys have investigated the bacterial members of these plant-associated microbes, and the pivotal role of PGPR in plant life has been well highlighted and illustrated (Kloepper and Schroth, 1978; van Loon, 2007; Ipek and Esitken, 2017; Majeed *et al*., 2018; Compant *et al*., 2019; Kumari *et al*., 2019). Among the bacterial genera found to be PGPR, *Pseudomonas* and *Bacillus* are the most represented in the rhizosphere, and *Pseudomonas simiae* WCS417 is one of the most studied PGPR for ISR induction (Pieterse *et al*., 2020).

1.5.2 *Pseudomonas simiae* WCS417

Bacteria of the genus *Pseudomonas* have been investigated as notable members of the root microbiota since the 1980s, and *Pseudomonas* species are well known for their ability to suppress plant illnesses and protect the host from pathogens (Stutz *et al*., 1986; Leeman *et al*., 1996; Mercado-Blanco *et al*., 2004; Raaijmakers *et al*., 2009; Berendsen *et al*., 2012; Damiri *et al*., 2018). *Pseudomonas* spp. are highly competitive bacteria in the rhizosphere that use chemotaxis and flagella to easily find and migrate toward the roots (de Weger *et al*., 1987; de Weert *et al*., 2002), colonize the rhizosphere, and compete against other microorganisms to protect their nutrient-rich niche, even competing members of the same strains (Bakker *et al*., 2013; Pangesti *et al*., 2017; Stringlis *et al*., 2018a). *Pseudomonas fluorescens* is a nonpathogenic saprophyte, obligate aerobe, and gram-negative bacterium that can colonize different areas (such as soil, water, and plant organs) by moving through multiple polar flagella (Ganeshan and Kumar, 2005). Its name comes from the soluble yellow-green fluorescent pigment 'pyoverdine', which is an iron-chelating siderophore

secreted by the rhizobacterium (Ringel and Brüser, 2018) (**Figure 4**). 'Pyochelin' and 'quinolobactina' are two other molecules produced and released by *P. fluorescens* (Mossialos *et al.*, 2000; Haas and Défago, 2005). *P. fluorescens* WCS417, WCS358, WCS374, PICF7, and R81 strains were renamed *Pseudomonas simiae* after the publication of their complete genome sequence (Mathimaran *et al*., 2012; Berendsen *et al*., 2015; Martínez-García *et al*., 2015); *P. simiae* WCS417 genome has a size of 6.17 Mb, no plasmids, and ~ 5000 protein-coding genes (Berendsen *et al*., 2015). *P. simiae* WCS417 was discovered in 1988 in bread wheat (*Triticum aestivum*) roots infected with the fungus *Gaeumannomyces graminis* var. *tritici*: *P. simiae* WCS417 significantly decreased the incidence of disease and increased wheat grain yield (Lamers *et al*., 1988). Several studies have then highlighted how this *P. simiae* strain promotes plant resistance to biotic and abiotic stresses (Leeman *et al*., 1996; Ran *et al*., 2005; Nel *et al*., 2006; Canchignia *et al*., 2017; Chiappero *et al*., 2019; Verbon *et al*., 2019; Desrut *et al*., 2020). *P. simiae* WCS417 has been studied in detail to understand how free-living beneficial PGPR can establish a long-term relationship with their host, providing many advantages to plants, such as the induction of ISR and the promotion of plant iron (Fe) nutrition (Pieterse *et al*., 2020).

Extensive research has shown that *P. simiae* WCS417 volatile organic compounds (VOCs) inducing ISR can help alleviate Fe-deficiency stress in plants (Zhang *et al*., 2009; Zamioudis *et al*., 2015; Verbon *et al*., 2019), inducing the expression in the roots of the plant transcription factor MYB72, which is a key regulator of the plant Fe-deficiency pathway and is required for *P. simiae* WCS417-ISR onset in *Arabidopsis thaliana* roots (Zamioudis *et al.*, 2014; Stringlis *et al.*, 2018*a*; Yu *et al.*, 2021). There is no need for a physical interaction between *P. simiae* WCS417 and roots because microbial VOCs are sufficient to induce the expression of *MYB72*; signals from photosynthesizing leaves are also necessary for the induction of *MYB72* expression upon *P. simiae* WCS417 root colonization (Zamioudis *et al*., 2015). Pyoverdine improves plant Fe acquisition by activating plant Fe uptake pathways (Verbon *et al*., 2019; Trapet *et al*., 2021) and favors *P. simiae* WCS417 rhizosphere colonization and plant health because of its antimicrobial activity (Crowley *et al.*, 1992; Marschner *et al.*, 1997; Miethke and Marahiel, 2007; Zamioudis *et al*., 2013; Aznar *et al.*, 2014, 2015; Aznar and Dellagi, 2015; Schulz-Bohm *et al*., 2017; Gu *et al.*, 2020).

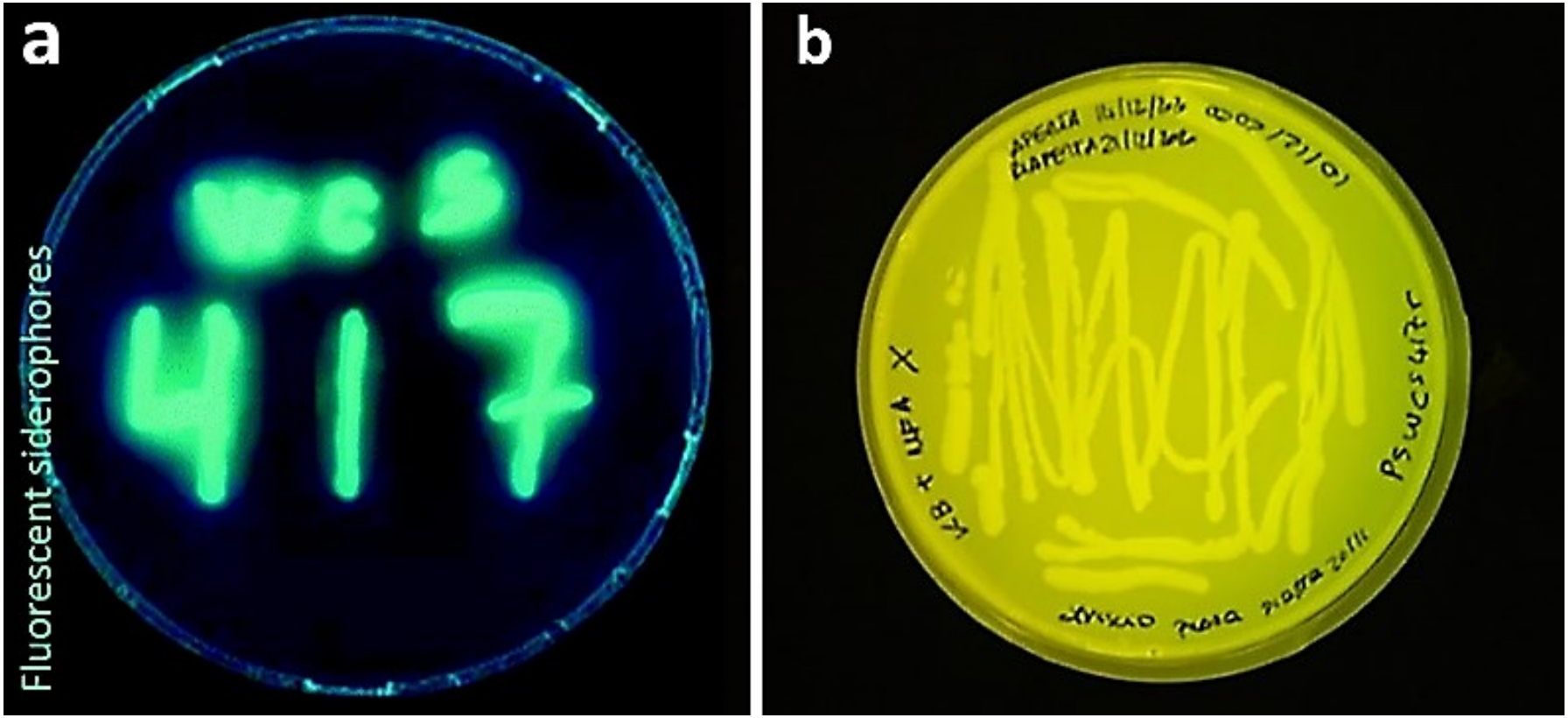


Figure 4. *Pseudomonas simiae* WCS417 plates. The fluorescent siderophore pyoverdine, released by the rhizobacterium, appears as the yellow/green fluorescence characterizing the growth medium. **(a)** Figure reproduced from Pieterse *et al*. (2020). **(b)** This is one of the first plates of *Pseudomonas simiae* WCS417 that I prepared at the end of my first year of Ph.D.

It is well established that plants under Fe deficiency change their microbiota composition by favoring the recruitment of siderophore-producing microbes, possibly through root exudations that can attract or suppress microorganisms (Jin *et al*., 2006, 2010). *P. simiae* WCS417 can not only withstand the antimicrobial action of the coumarins scopoline and scopoletin released by roots under Fe deficiency but also favor their production to outcompete other microorganisms and better colonize *A. thaliana* roots (Verbon *et al*., 2019; Yu *et al*., 2021). Interestingly, *P. simiae* WCS417 can promote the Fe-deficiency response even when plant Fe levels are adequate to match the increased plant growth rate observable upon *P. simiae* WCS417 colonization; however, the rhizobacterium induces Fe shortage responses in *A. thaliana* only when the number of bacteria colonizing the roots is sufficient (Verbon *et al*., 2019). This discovery demonstrated how shoot-to-root communication mechanisms unrelated to Fe-leaf status control the *P. simiae* WCS417-root Fe-deficiency response, raising the hypothesis of novel phloem-mobile shoot-to-root signals and the involvement of phytohormones.

1.6 Plants, iron, and microbes

Iron (Fe) is an essential micronutrient that is required by most organisms. Plants, like any other organism, need Fe because it is a catalytic component of many enzymes required for critical cellular processes such as photosynthesis, respiration, antioxidant defenses, and hormonal and secondary metabolism. Plant iron nutrition also provides appropriate concentrations of this micronutrient to harvested tissues/organs and, consequently, ensures proper animal nutrition. However, both Fe excess and Fe deficiency negatively affect growth. Since the Induced Systemic Resistance (ISR) and the Fe deficiency response partially overlap in *Arabidopsis thaliana* once the roots are colonized by *Pseudomonas simiae* WCS417, I focused on the role of iron in plant life and contributed to the writing and publication of two reviews on this topic. Moreover, I investigated how exposure to a beneficial rhizobacterium could affect formate dehydrogenase (FDH) promoter activity in leaves; indeed, as described already in the present thesis, FDH is a key player in plant Fe homeostasis and participates in an early leaf defense response.

1.6.1 Two-face iron

After oxygen (O), silicon (Si), and aluminum (Al), Fe is the 4th most abundant element in the Earth's crust, but its availability is limited. In soil, Fe often exists in Fe^{2+} (ferrous) and Fe^{3+} (ferric) forms; changes in its oxidation state dictate the presence of Fe in the soil and its mobility in plants:

$$Fe^{3+} + e^{-} \leftrightarrow Fe^{2+} \ (E^{\circ} = - 0.44)$$

In soils, Fe is mainly present as ferric oxide/hydroxide, which is poorly soluble at neutral and high pH (Broadley *et al*., 2012); ferric oxide K_{sp} is 4×10^{-38} (Murgia *et al*., 2022), indicating that the concentration of Fe^{3+} is exceedingly low at neutral or basic pH (if pH = 7, $[OH^{-}] = 10^{-7}$ M and $[Fe^{3+}] = 4 \times 10^{-17}$ M). This condition dramatically reduces the amount of Fe available in the solution to organisms (Ahmad *et al*., 2021). Given the essential functions of Fe in organisms, it is not surprising that it is the third most limiting element for organism development (Stafford, 1961; Andrews, 2000; Earhart, 2009; Rout and Sahoo, 2015). Plants have evolved two strategies, referred to as 'Strategy I' and 'Strategy II', in

response to Fe deficiency (Marschner and Römheld, 1994). In non-grass plants such as *Solanum lycopersicum* (tomato) and *Arabidopsis thaliana*, Fe^{3+} solubility in the soil is increased by lowering rhizosphere pH through the secretion of protons by the H+-ATPase AHA2. Fe^{3+} is then reduced to Fe^{2+} by the plasma membrane protein FERRIC REDUCTION OXIDASE 2 (FRO2) and transported into the root epidermis via the IRON-REGULATED TRANSPORTER 1 (IRT1) transporter. Alternatively, grass species like *Zea mays* (maize) use the TRANSPORTER OF MUGINEIC ACID 1 (TOM1) to release phytosiderophores (PS) into the rhizosphere, which chelate Fe^{3+}. The resulting PS-Fe^{3+} complex is then taken up by specific transporters of the YELLOW STRIPE-LIKE (YSL) family. Iron deficiency is a severe problem in both plant and animal kingdoms and represents the most widespread micronutrient deficiency among humans, causing clinically significant hematological problems, such as iron deficiency anemia (IDA). Human Fe deficiency is recognized as a 'hidden hunger' because the Fe content in several staple foods is low and often only partially bioavailable. However, several distinct factors contribute to human Fe deficiency and anemia, such as poverty, malnutrition, starvation, and infections in poor countries; extreme vegetarianism and veganism, and malabsorption in industrialized ones (Murgia *et al*., 2012; Ahmad *et al*., 2021). Three complementary strategies for fighting iron deficiency in humans have been proposed (Murgia *et al*., 2012) (**Figure 5**): the increase in daily iron intake could be achieved by pharmacological iron supplementation and iron food fortification, but these approaches can be difficult to adopt in poorer countries, particularly because of their high costs and potential health problems resulting from increasing iron body availability by supplementation (Idjradinata *et al*., 1994; Lynch, 2005; Sazawal *et al*., 2006). The strategy of 'biofortification' can help fight iron deficiency by increasing Fe content in edible plant parts through mineral fertilization, conventional breeding, or transgenic approaches (White *et al*., 2009; Bouis *et al*., 2011; Murgia *et al*., 2012). Because of the high economic and environmental costs and environmental effects of fertilizers (Zhu *et al*., 2007), crop biofortification by breeding or genetic modification provides a low-cost and sustainable way to supply iron. Thus, researchers are interested in how plants obtain and accumulate nutrients to increase their amount in food, and new methods for crop biofortification have been developed, including agronomic, genetic, and microbiological strategies (Ahmad *et al*., 2021). For instance, different strains of bacteria and fungi added to the soil may favor plant Fe uptake by the release of microbial siderophores. It is important to emphasize that excess of Fe can be particularly damaging to organisms, and possible toxicities due to an excess of this micronutrient must be well considered in plants because the effects of excessive Fe can

include stunted growth and bronzing of foliage. Through the Fenton reaction (Fenton, 1894), excessive free ferric iron causes the formation of reactive oxygen species (ROS) that can damage proteins, DNA, and lipids (Becana *et al*., 1998; Tsai and Huang, 2006; Lodde *et al*., 2021). The production of ROS is particularly dangerous because once the damage caused by these molecules is too severe, the cell undergoes programmed cell death. Plants have mechanisms to control Fe uptake and storage to finely regulate Fe levels, as discussed in several recent reviews (Gao *et al*. 2019; Grillet and Schmidt, 2019; Romera *et al*., 2019; Schwarz and Bauer, 2020; Gao and Dubos, 2021; Riaz and Guerinot, 2021; Spielmann and Vert, 2021; Murgia *et al*., 2022).

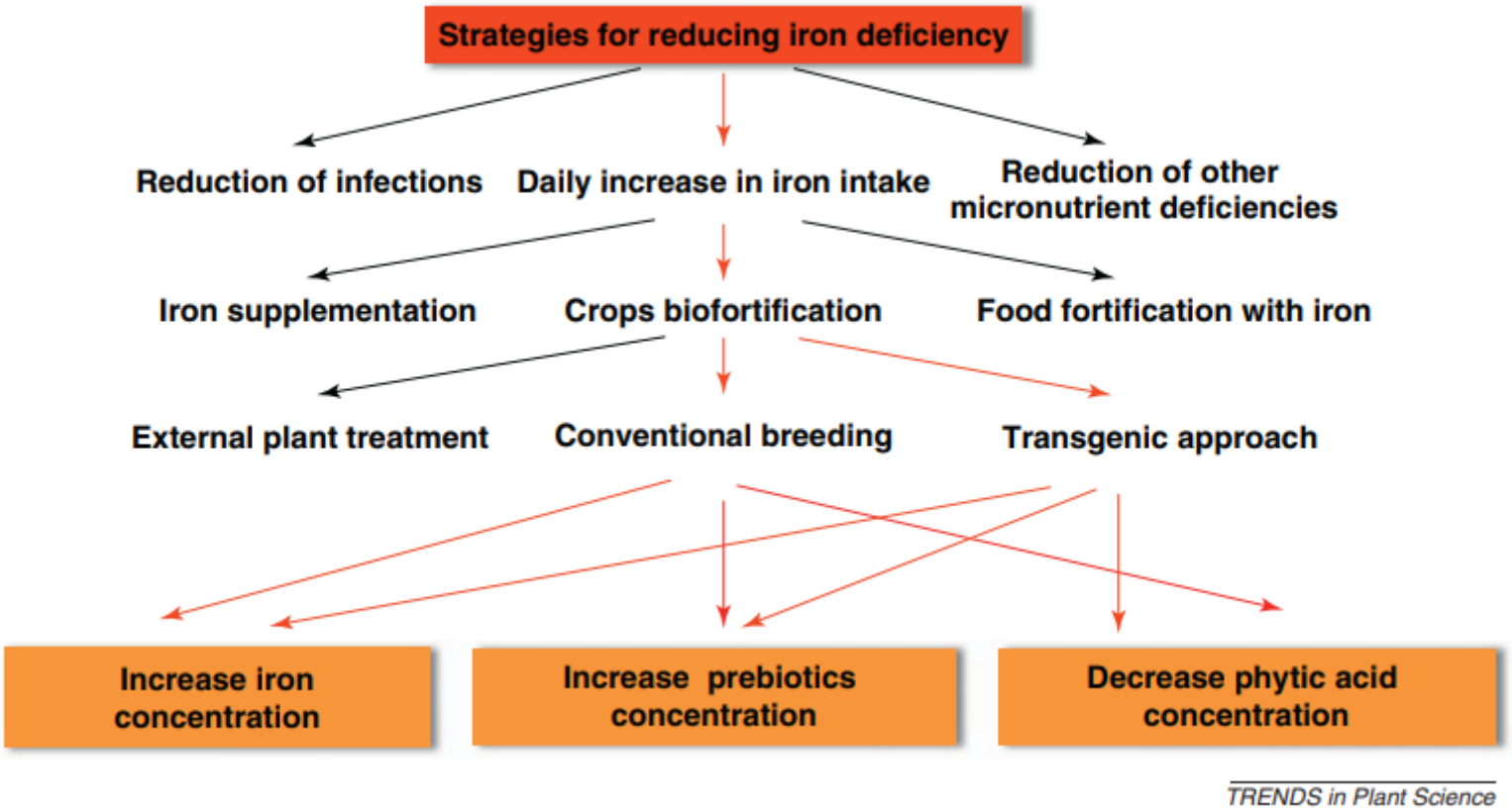


Figure 5. Possible approaches to fight iron deficiency in humans. Red arrows indicate the choices/processes leading to three biofortification strategies of plant edible parts through conventional breeding or transgenesis. Figure reproduced from Murgia *et al*. (2012).

1.6.2 ISR and Fe deficiency

Ethylene (ET), jasmonic acid (JA), salicylic acid (SA), cytokinin (CK), nitric oxide (NO), gibberellic acid (GA), and auxin (indole-3-acetic acid, IAA) are phytohormones involved in the control of iron-deficiency responses (Schmidt *et al*., 2000; Murgia *et al*., 2002; Hindt and Guerinot, 2012; García *et al*., 2015). For example, ET is linked to the morphological changes observed in roots under iron deficiency, such as root branching and hair formation (Schmidt, 1999; Jin *et al*., 2008; Morissey and Guerinot, 2009; Li *et al*., 2016); ET is also linked to the promotion of the expression of *FRO2* and *IRT1* together with GA (Lucena *et al*., 2015; Wild *et al*., 2016), and to the production of NO (with IAA), which increases the absorption of iron and stabilizes the basic helix-loop-helix (bHLH) transcription factor FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR (FIT) (Romera *et al*., 2011; García *et al*., 2010; Meiser *et al*., 2011). Root morphological changes are also induced by IAA (which stimulates the production and growth of lateral roots) and by CKs, which inhibit both root elongation and branching (Aloni *et al*., 2006; Séguéla *et al*., 2008; Giehl *et al*., 2012; Jing and Strader, 2019). SA and JA are two plant defense hormones also involved in iron nutrition: SA positively regulates responses under Fe deficiency because it controls IAA and ET signaling, and positively regulates the expression of *FRO2* and *IRT1*; JA favors FIT turnover by regulating the expression of several *bHLH* genes and inhibiting Fe uptake by downregulating both *FRO2* and *IRT1* (Maurer *et al*., 2011; Kong *et al*., 2014; Shen *et al*., 2016; Cui *et al*., 2018; Boukari *et al*., 2019; Kabir *et al*., 2021). Given the importance of SA, JA, ET, and IAA in the regulation of plant immunity, their involvement in Fe-deficiency responses underlines a possible overlap between Fe homeostasis and plant immunity pathways (Romera *et al*., 2019 and references therein). 'Systemic Acquired Resistance' (SAR) and 'Induced Systemic Resistance' (ISR) are plant systemic defense responses that protect undamaged plant tissues from possible attacks. They rely on an energy-saving strategy called 'defense priming', which allows plants to be prepared for quicker and higher expression of defense responses and to avoid unnecessary extensive activation of defense mechanisms (Martinez-Medina *et al*., 2016; Mauch-Mani *et al*., 2017). The main differences between the two systemic immune responses are related to the 'priming stimulus', which induces defense priming, and the hormones involved (Pieterse *et al.*, 2014). The SAR mechanism is associated with an increase in SA levels upon pathogen attack, both at the site of infection and in distant plant

organs (Ryals *et al*., 1996; Fu and Dong, 2013; Klessig *et al*., 2018), whereas the dense root microbiota induces ISR and protects plants against several stresses, by activating the JA and ET signaling pathways and transferring the defense message to distant plant tissues (van Loon *et al*., 1998; Pieterse *et al*., 2014; Conrath *et al*., 2015; Martinez-Medina *et al*., 2016; Vlot *et al*., 2021). Several plant growth-promoting rhizobacteria (PGPR) (predominantly *Pseudomonas* and *Bacillus* genera) and plant growth-promoting fungi (PGPF) (*Trichoderma*, *Serendipita*, *Fusarium*, and arbuscular mycorrhizal fungi (AMF)) are important inducers of ISR, and changes in the composition and homeostasis of the root microbiota can promote plant defense (Zamioudis and Pieterse, 2012; Paasch and He, 2021).

The function of ISR in *Arabidopsis thaliana* roots colonized by the rhizobacterium *Pseudomonas simiae* WCS417 has been well-characterized (Zamioudis *et al*., 2014, 2015; Stringlis *et al*., 2018a). The biosynthesis of JA and ET is not directly induced by *P. simiae* WCS417 root colonization, but one of the targets is the transcription factor MYB72, which is necessary for initiating the *P. simiae* WCS417-mediated ISR response (Zamioudis *et al.*, 2014, 2015; Stringlis *et al.*, 2018*b*; Yu *et al.*, 2021). *MYB72* is particularly expressed in the epidermal and cortical cells of colonized *A. thaliana* roots, and *myb* mutants cannot induce ISR once *P. simiae* WCS417 and other beneficial microorganisms colonize roots (van der Ent *et al*., 2008; Segarra *et al*., 2009). However, the role of this root-specific transcription factor is not limited to ISR onset; MYB72, together with its closest homolog MYB10, regulates the Fe-deficiency response in plants and is crucial for plant survival (Buckhout *et al*., 2009; Dubos *et al.*, 2010; Palmer *et al*., 2013; Rodríguez-Celma *et al*., 2013; Fourcroy *et al*., 2014; Schmidt *et al.*, 2014; Zamioudis *et al*., 2014; Liu *et al.*, 2015a; Harbort *et al.*, 2020). Two genes controlled by MYB72 once *A. thaliana* roots are colonized by *P. simiae* WCS417 or under Fe deficiency are *β-GLUCOSIDASE42* (*BGLU42*) and the transporter *ATP-BINDING CASSETTE G37/PLEIOTROPIC DRUG RESISTANCE 9* (*ABCG37/PDR9*). BGLU42 is required downstream of MYB72 to digest the Fe-mobilizing coumarins secreted into the rhizosphere by ABCG37/PDR9 (Palmer *et al*., 2013; Rodríguez-Celma *et al*., 2013; Fourcroy *et al*., 2014, 2016; Zamioudis *et al*., 2014; Stringlis *et al*., 2019; Robe *et al*., 2021); *BGLU42* overexpression promotes plant resistance to various pathogens and pest attacks, and *A. thaliana bglu42* mutants exhibit defective *P. simiae* WCS417-ISR onset (Zamioudis *et al*., 2014). The observable overlap between changes in the transcriptome of roots colonized by *P. simiae* WCS417 and roots under Fe deficiency implies a strong relationship between the onset of ISR and plant Fe status (Verhagen *et al*., 2004; Dinneny *et al*., 2008;

Zamioudis *et al*., 2015). *MYB72* and other key genes in the iron-deficiency response (such as *FIT*, *IRT1*, and *FRO2*) are indeed altered in their expression in roots colonized by microorganisms that induce ISR, even when plants are grown under normal Fe conditions (Zhang *et al*., 2009; Zamioudis *et al*., 2015; Martínez-Medina *et al*., 2017; Verbon *et al*., 2019). *FIT* and *bHLH038* are activated by ET and, together with bHLH121, control *MYB72* and *MYB10* expression (Gao *et al*., 2020); therefore, it is reasonable to think that ET could act as a link between Fe deficiency and ISR onset, with *MYB72* and *BGLU42* representing the nodes of convergence (Verbon *et al*., 2017; Romera *et al*., 2019). Phytohormones, particularly ET, IAA, and NO, represent interesting links because they are implicated in Fe deficiency and the activation of immune responses in several plant species. For example, the *A. thaliana* plant defensin PDF1.1, which can bind Fe with high affinity, may play a significant role as a Fe sink and activator of ET-dependent immunity (Hsiao *et al*., 2017), and several studies have shown that ET levels are influenced by the associated root microbiota (Iqbal *et al*., 2017; Nascimento *et al*., 2018; Ravanbakhsh *et al*., 2018).

Colonization of the rhizosphere is an essential step for all members of the plant microbiota, implying complex plant-microbe and microorganism-microorganism interactions. The study of molecules involved in these interactions, immune system signaling, and inhibition of plant immunity by microbes has advanced along with discoveries in iron homeostasis. For instance, the presence of a pathogen in the soil leads to an increase in the expression of microbiota genes involved in the production of virulence molecules, extracellular lytic enzymes, and stress-sensing mechanisms (Chapelle *et al*., 2016); however, iron-chelating siderophores are also produced by rhizobacteria and are involved in iron sequestration and microbial suppression, favoring plant health (Miethke and Marahiel, 2007; Aznar *et al*., 2014, 2015; Aznar and Dellagi, 2015; Herlihy *et al*., 2020). Indeed, the sequestration of nutrients by rhizobacteria inhibits pathogen proliferation in the rhizosphere, and plants benefit from root colonization by beneficial microbes. An example of a siderophore is ‘pyoverdine’ produced by *Pseudomonas simiae* WCS417 (see paragraph ‘1.5.2 *Pseudomonas simiae* WCS417’), which helps antagonize several fungal infections and contributes to the disease-suppressive actions of *Pseudomonas* strains (Leong, 1986; Aznar and Dellagi, 2015; Gu *et al*., 2020). Plants attempt to limit pathogen development by releasing coumarins with antimicrobial activities (Aznar and Dellagi, 2015; Voges *et al*., 2019; Stringlis *et al*., 2019) and controlling the distribution of iron in the organism (Herlihy *et al*., 2020; Liu *et al*., 2021). For instance, plants can sequester micronutrients from the

infection site, limiting their use by pathogens; however, they can also over-accumulate iron to induce ROS production, challenging pathogens. Fe is also present in the heme prosthetic group of several peroxidases and catalases that are required to produce lignin in plant cells to defend against pathogens (Osorio Vega, 2007). Nevertheless, even if Fe levels significantly contribute to the resistance to biotic stresses, the correlation between plant immunity and Fe homeostasis is complex (Liu *et al*., 2021 and references therein). Although recognized for a long time, the effect of iron on plant-pathogen interactions has received little attention, and iron homeostasis and plant immunity are frequently studied independently of each other. Moreover, a detailed description of the mechanisms by which plant microbiota induces Fe-shortage responses and how they affect plant Fe homeostasis is lacking. **Figure 6** shows a simplified scheme of the *A. thaliana* iron-uptake strategy (Strategy I) and its intersection with beneficial and pathogenic microorganisms. Certainly, further characterization of the root microbiota, changes in gene expression during root colonization and/or infections, and protein production during these interactions are important steps for a better understanding of the link between plant Fe homeostasis and immunity (Sharma *et al*., 2020; Zancarini *et al*., 2021).

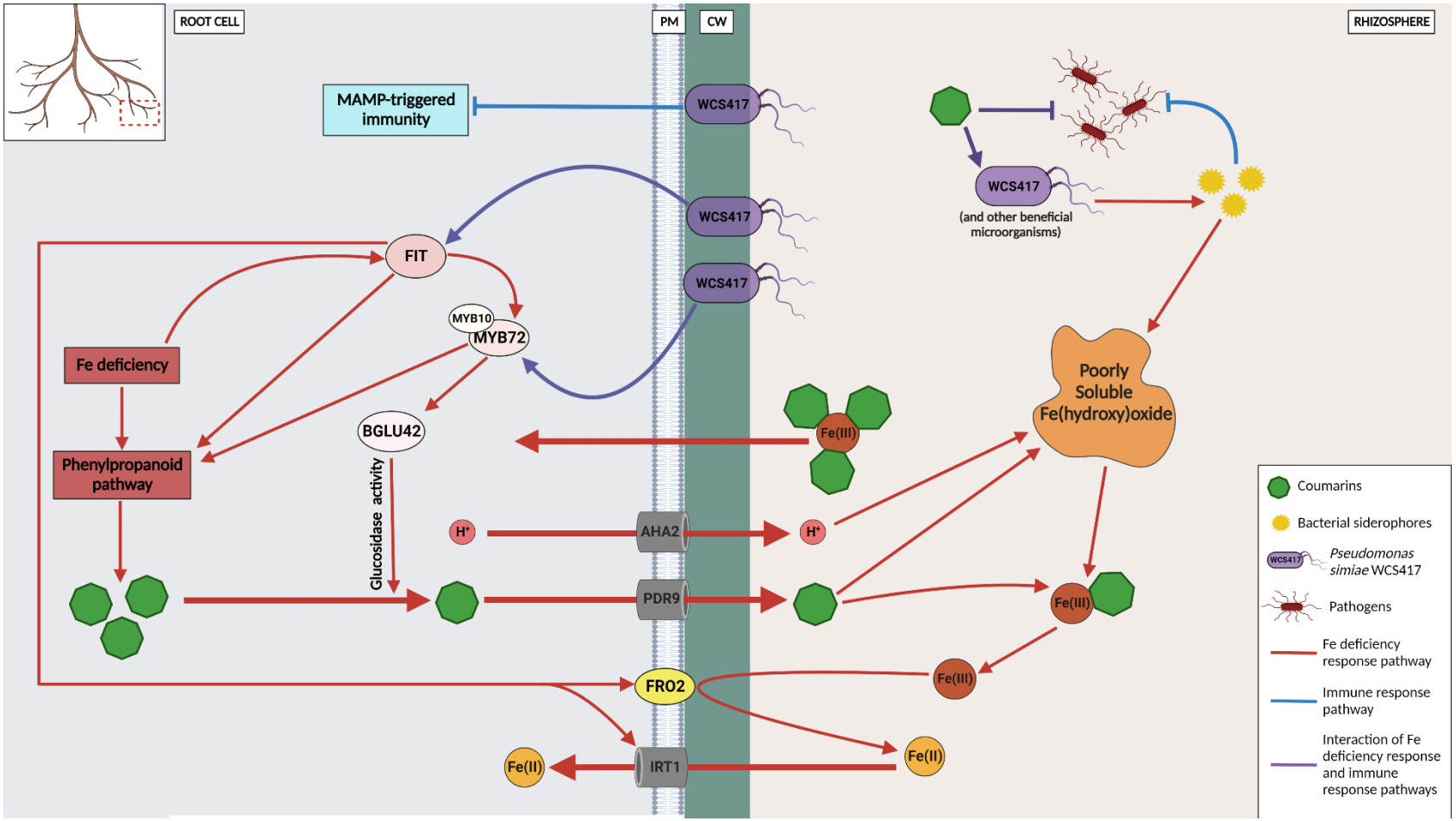


Figure 6. Interactions among Strategy I Fe-uptake pathway, plant immunity, and root microbiota. A root cell (on the left) is represented with its surroundings rhizosphere environment (on the right). Red arrows represent the cascade of events that occur during the Fe-deficiency response (thick arrows for transport/movement; thin arrows for signaling). Blue arrows indicate immune response pathways, such as those activated by *Pseudomonas simiae* WCS417, leading to the evasion of host immunity (suppression of MAMP-triggered immunity, upper left), and those triggered by *P. simiae* WCS417 siderophores, leading to plant pathogen suppression (upper right). The purple arrows represent the overlapping pathways of Fe deficiency and immunological responses. PM: plasma membrane; CW: cell wall. Figure adapted from Murgia *et al*. (2022).

2. Aims of the thesis

Two roads diverged in a wood, and I—/I took the one less traveled by,
/And that has made all the difference.
— Robert Frost

2.1 General aim

This Ph.D. thesis deals with the response of formate dehydrogenase (FDH) in *Arabidopsis thaliana* leaves when plants are exposed to either a pathogen or a beneficial rhizobacterium. Such investigations were performed using both *in silico* and *in vivo* strategies: the term 'hybrid' in the title refers to this dual approach and is the unifying thread throughout this Ph.D. work. During these years, I have frequently been referred to as a 'hybrid' researcher because I was unable to choose between my two educational backgrounds, bioinformatics and wet lab. I always strive to 'integrate' programming into my lab work, and I regard this as a good strategy: 'combining' *in silico* and *in vivo* approaches would be preferable when one intends to explore a scientific question and develop new hypotheses, as occurred for my investigation on the role of *A. thaliana* FDH during plant interactions with bacteria.

2.2 Specific aims

The specific aims of this thesis correspond to three functionally complementary sections grouping my six manuscripts (**Figure 7**). These aims are thus presented in chapter ‘3. Scientific works’ and are as follows:

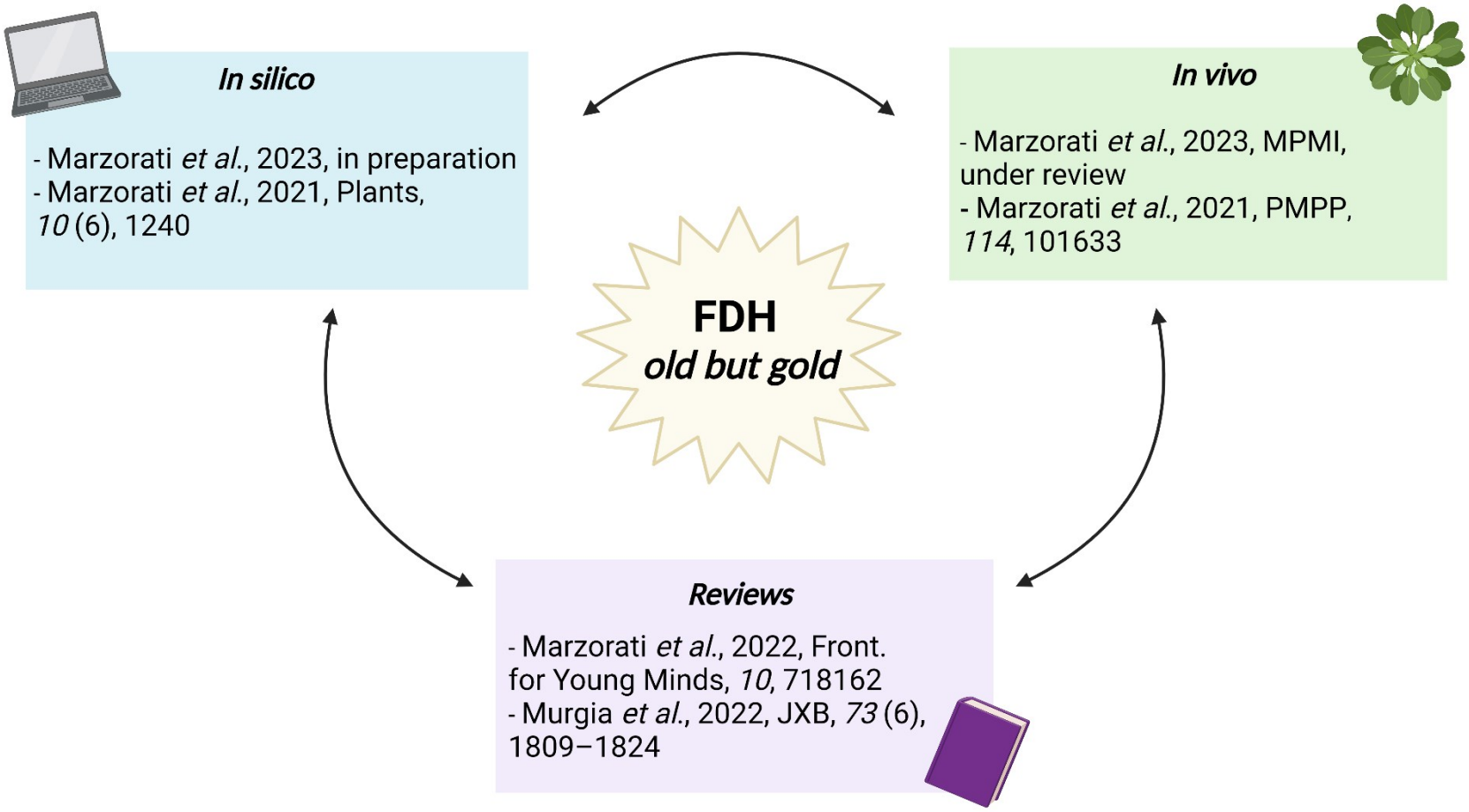


Figure 7. Graphical representation of the three functionally complementary sections presented in chapter '3. Scientific works'. These three sections group the six manuscripts produced during the Ph.D.; each section addresses a specific work done during my Ph.D. research (*in silico*, *in vivo*, and reviews). The common thread among the different sections is the study of formate dehydrogenase (FDH) role in the interaction between plant and bacteria by using a 'hybrid' approach (*i.e.*, adopting both *in silico* and *in vivo* techniques). Figure created by using BioRender.com

- Development of a general method, based on the logical and statistical relationships between conditions and parameters, to enhance the quality of plant microarray databases by working on the *Medicago truncatula* Gene Expression Atlas (MtGEA) (Benedito *et al*., 2008; He *et al*., 2009) (**Manuscript 1**). This approach was developed after discovering that the MtGEA database contains a significant share of low-quality data and errors that can compromise subsequent analyses performed on the data.

- Development of an application with a user-friendly interface to help plant scientists who lack computational skills perform straightforward yet efficient analyses of microarray data (**Manuscript 2**, manuscript in preparation). NORMALIX95 provides the opportunity to use the vast amount of plant *Affymetrix* data currently available online for several species, for example by performing correlation studies using a 'guilt-by-association' approach (Altshuler *et al*., 2000), which can help in making new hypotheses to test.

- Effect of the infection with the vascular pathogen *Xanthomonas campestris* pv *campestris* (*Xcc*) on *FDH* promoter activity in *Arabidopsis thaliana* hydathodes (**Manuscript 3**). This study aimed to investigate the involvement of FDH and its role in pathogen-induced responses. A biotic stress dataset by collecting all the *Affymetrix* data available in the online repository GEO (Edgar, 2002; Barrett *et al.*, 2013) for *A. thaliana* under pathogen and pest attack was built, and a gene correlation analysis following the 'guilt-by-association' approach was hence performed with this dataset to identify the top correlators of *FDH*. I then investigated the involvement of FDH in defense against pathogenic bacteria by analyzing the leaves of *A. thaliana* infected with *Xcc*.

- Effects on *Arabidopsis thaliana* leaves of the exposure of roots to the rhizobacterium *Pseudomonas simiae* WCS417 by studying the total foliar proteome and *FDH* promoter activity in hydathodes and entire seedlings (**Manuscript 4**, manuscript under review). This study assessed whether leaf FDH is responsive to direct *P. simiae* WCS417 inoculation in soil and to exposure to the rhizobacterium without colonization of the root apparatus, attempting to establish how rapid this response could be. Additionally, a leaf proteome analysis was performed on two *A. thaliana*

lines (wt Col and the knockout mutant *atfdh1-5*) to highlight potential variations in leaf metabolic pathways upon rapid exposure to the rhizobacterium.

- A detailed description of plant iron (Fe) nutrition, covering the topic of how soil microbes and plant roots interact in the process of plant iron nutrition, how iron is transported and stored in seeds, and how wild relatives can help research on Fe nutrition (**Manuscript 5**). Advancements in this research field may allow the development of new strategies to improve plant Fe nutrition with positive effects on both human and animal nutrition.

- Presentation of the importance of Fe in plant life in a simple yet informative review addressed to high school students (**Manuscript 6**). Iron nutrition is not typically taught in schools and communicating science to young students can be challenging.

3. Scientific works

3.1 *In silico* works

"I do not fear computers. I fear the lack of them."

— Isaac Asimov

Manuscript 1

Marzorati F, Wang C, Pavesi G, Mizzi L, Morandini P. 2021. Cleaning the *Medicago* microarray database to improve gene function analysis. *Plants* **10**, 1240.

Manuscript 2 (in preparation)

Marzorati F, Mizzi L, Murgia I, Morandini P. **2023**. NORMALIX95: a shiny-based application for plant microarray analysis.

Manuscript 1

Cleaning the *Medicago* microarray database to improve gene function analysis

Francesca Marzorati[1], Chu Wang [2,†], Giulio Pavesi[2], Luca Mizzi[2] and Piero Morandini[1,*]

[1] Department of Environmental Science and Policy, University of Milano, Milano, Italy
[2] Department of Biosciences, University of Milano, Milano, Italy
* Correspondence: piero.morandini@unimi.it
† Present Address: AEB (Shanghai) Trading Co., LTD., Room 301, F6, 600 Jianchuan Road Minhang District, Shanghai 200241, China.

Abstract

Transcriptomics studies have been facilitated by the development of microarray and RNA-Seq technologies, with thousands of expression datasets available for many species. However, the quality of data can be highly variable, making the combined analysis of different datasets difficult and unreliable. Most of the microarray data for *Medicago truncatula*, the barrel medic, have been stored and made publicly accessible on the web database *Medicago truncatula* Gene Expression atlas (MtGEA). The aim of this work is to ameliorate the quality of the MtGEA database through a general method based on logical and statistical relationships among parameters and conditions. The initial 716 columns available in the dataset were reduced to 607 by evaluating the quality of data through the sum of the expression levels over the entire transcriptome probes and Pearson correlation among hybridizations. The reduced dataset shows great improvements in the consistency of the data, with a reduction in both false positives and false negatives resulting from Pearson correlation and GO enrichment analysis among genes. The approach we used is of general validity and our intent is to extend the analysis to other plant microarray databases.

Keywords: *Medicago*; MtGEA; Transcriptomics; Functional genomics; Microarray; R programming; Correlation analysis

plants

MDPI

Article

Cleaning the *Medicago* Microarray Database to Improve Gene Function Analysis

Francesca Marzorati [1], Chu Wang [2,†], Giulio Pavesi [2], Luca Mizzi [2] and Piero Morandini [1,*]

[1] Department of Environmental Science and Policy, University of Milan, Via Celoria 10, 20133 Milano, Italy; Francesca.marzorati1@unimi.it
[2] Department of Biosciences, University of Milan, Via Celoria 26, 20133 Milano, Italy; cwang@aeb-group.com (C.W.); giulio.pavesi@unimi.it (G.P.); luca.mizzi@unimi.it (L.M.)
* Correspondence: piero.morandini@unimi.it
† Present Address: AEB (Shanghai) Trading Co., LTD., Room 301, F6, 600 Jianchuan Road Minhang District, Shanghai 200241, China.

Abstract: Transcriptomics studies have been facilitated by the development of microarray and RNA-Seq technologies, with thousands of expression datasets available for many species. However, the quality of data can be highly variable, making the combined analysis of different datasets difficult and unreliable. Most of the microarray data for *Medicago truncatula*, the barrel medic, have been stored and made publicly accessible on the web database *Medicago truncatula Gene Expression atlas* (MtGEA). The aim of this work is to ameliorate the quality of the MtGEA database through a general method based on logical and statistical relationships among parameters and conditions. The initial 716 columns available in the dataset were reduced to 607 by evaluating the quality of data through the sum of the expression levels over the entire transcriptome probes and Pearson correlation among hybridizations. The reduced dataset shows great improvements in the consistency of the data, with a reduction in both false positives and false negatives resulting from Pearson correlation and GO enrichment analysis among genes. The approach we used is of general validity and our intent is to extend the analysis to other plant microarray databases.

Keywords: *Medicago*; MtGEA; transcriptomics; functional genomics; microarray; R programming; correlation analysis

Citation: Marzorati, F.; Wang, C.; Pavesi, G.; Mizzi, L.; Morandini, P. Cleaning the *Medicago* Microarray Database to Improve Gene Function Analysis. *Plants* **2021**, *10*, 1240. https://doi.org/10.3390/plants10061240

Academic Editors: Ornella Calderini, Andrea Porceddu and Francesco Panara

Received: 12 March 2021
Accepted: 11 May 2021
Published: 18 June 2021

1. Introduction

"Omic" technologies have been developed to investigate cellular molecules on a massive scale and they are classified according to the object studied: genes for genomics, RNA for transcriptomics, proteins for proteomics and metabolites for metabolomics [1]. Currently, ionomics (studying ions composition) is also arising as a major -omic science [2]. In transcriptomics, gene expression data coming from different tissues, conditions and genotypes can be obtained through different strategies, for example, different types of microarrays or RNA sequencing (RNA-Seq). These approaches allow to assess the expression level of most or nearly all the genes in an organism, and each experiment may envisage tens, if not hundreds, of measurements of different samples [3]. In this way, thousands of measurements on a genome-wide level are available for many species.

Microarrays are useful tools to explore genotypes and their interaction with the corresponding phenotypes, but the data produced by this technology require, however, processing for correct interpretation [4–7]. Even if the "death" of microarrays was predicted already in 2008 [8], there is still a wealth of data to be explored, and many novel datasets still appear in literature. Most of the microarray data are available online, e.g., on the *Gene Expression Omnibus* (GEO, https://www.ncbi.nlm.nih.gov/geo/, accessed on 13 May 2021) [9,10] or on dedicated websites, such as *The Arabidopsis Information Resource* (TAIR) (https://www.arabidopsis.org/, accessed on 13 May 2021) and *Medicago truncatula Gene*

Expression atlas (MtGEA) (https://mtgea.noble.org/v3/, accessed on 13 May 2021) [11–13]. *Affymetrix m*icroarray data are organized in datasets with the list of probeset codes (the *Affymetrix* identifiers for each set of probe sequences designed to measure a transcript) as the first column. Each further column of the dataset contains the expression values for a single hybridization of a sample (tissue or condition). Samples are usually characterized by two to three replicates. Together, all samples from the same publication are referred to as an 'experiment'.

Medicago truncatula, the barrel medic, is a small Mediterranean annual plant of the Fabaceae family, cultivated as a forage crop but extensively used as model organism for legumes. It is an autogamous plant, characterized by a short life cycle and a reduced genome size, allowing easy manipulation to study legume secondary metabolism [14,15]. Moreover, as many other legumes, *Medicago truncatula* establishes symbiotic relationships with nitrogen-fixing microorganisms; thus, it is used as model system to study this symbiosis [16–18].

In recent years, different transcriptomics resources have been developed for legumes, such as LegumeIP and LegumeGRN [19,20]. MtGEA is the gene expression atlas created specifically for *Medicago* spp., collecting most of the expression data obtained, using the *Affymetrix* GeneChip microarray technology [12,13,20,21]. In MtGEA, it is possible to explore the expression data for a gene of interest, which can be identified through its sequence, annotations or different identifiers, such as the *Affymetrix* probeset identifier, GO and KEGG annotation terms, gene name, and functional descriptions in natural language. Once identified, it is possible to perform different analyses on a gene or a gene list, such as the analysis of the expression profiles, co-expression studies, identification of genes showing differential expression among samples or experiments. Users can also download data in formats that are compatible with many analysis and visualization tools. The database is updated on a regular basis in order to include recent expression studies and updates on genome annotation [12,13].

In February 2020, we downloaded all the experimental data of *Medicago truncatula* collected in MtGEA, corresponding to 716 columns. Here, we report the analysis and cleaning of this dataset. After the cleaning, we performed a Pearson correlation and a Gene Ontology (GO) enrichment analyses on selected genes, working both on the original and cleaned datasets. The Pearson correlation coefficient assesses the strength of the linear relationship between two variables, expressed through a value between -1 and 1: a high correlation value means that two variables are strongly related, a negative value means that the variables are inversely related, whereas a small value means that the two variables are weakly associated [22,23]. Transcript correlation analysis is an important method to identify or confirm candidate genes involved in a pathway or process, as previously reported [24–28]. The correlation can be computed on the expression values as such or after Log-transformation, the latter being instrumental in revealing correlations holding also at low expression values.

The enrichment analysis is a procedure to interpret gene expression data identifying genes that are overrepresented in a large, provided set. We performed a GO term enrichment analysis, i.e., we identified GO categories overrepresented in selected gene sets [29,30]. By comparing the results of both analyses on the original and the cleaned datasets, we show significant changes in the lists of top correlators of several genes and the respective GO categories overrepresented in each list. Gene function predictions drawn from such lists may be substantially different, implying that the cleaning eliminates both false positive and negative correlators for a number of genes. We demonstrated that a proper cleaning of microarray datasets is required to find significant relations and GO functional enrichments among certain *Medicago* genes, results that are sustained by the literature and experimental evidence. We believe that the strategy developed is of general validity for the cleaning of expression microarray databases.

2. Results

2.1. Data

We downloaded all the *Affymetrix* microarray data from the *Medicago truncatula Gene Expression atlas* (MtGEA) (https://mtgea.noble.org/v3/, accessed on 13 May 2021) [12,13]. In February 2020, the complete dataset included 716 columns for 50,900 genes (Table S1), whereas the dataset containing the means of the replicates, when available, of each sample comprised 274 values (Table S2). As a start, we performed a Pearson correlation analysis, using the dataset of the means. Figure 1 shows a scatterplot for two probes, both referring to the same putative mevalonate kinase whose *Affymetrix* probe identifiers are Mtr.41545.1.S1_at and Mtr.16327.1.S1_at.

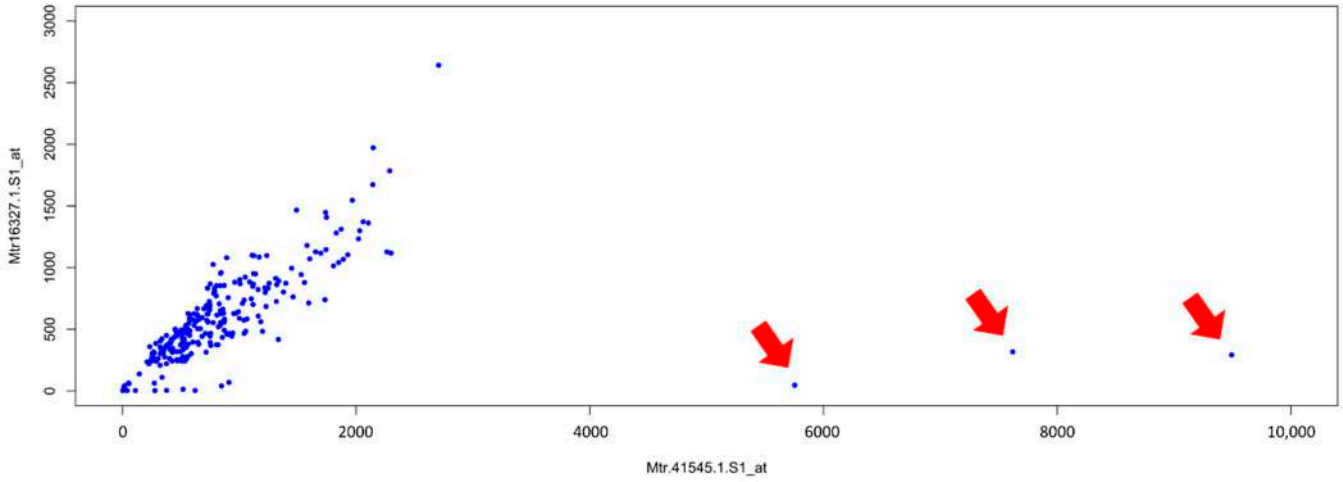


Figure 1. Scatterplot of the mean expression levels of probes Mtr.41545.1.S1_at and Mtr.16327.1.S1_at. Three outliers are indicated by red arrows.

Each point of the scatterplot represents either the mean value of biological replicates (two or three) of the same sample (tissue or condition), or, in few cases, one single measurement. Strikingly, three values are localized in the space far from all the others. We decided to further investigate these outliers, first of all, by identifying the corresponding hybridizations, which belong to a group of three related samples (each with two replicates) obtained by laser capture microdissection (LCM): RT_LCM_arbuscular, RT_LCM_cortical and RT_LCM_adjacent [31].

2.2. Sum of the Expression Values

To understand if these three experiments present some peculiarity and thus could generate outliers for other genes, we computed the sum of the expression values of all genes for each hybridization in the dataset as a first index to check the quality of data (Table S3). The mean of the sum of the expression values of the downloaded dataset is 1.81×10^7, even if, in most of the hybridizations, the sum is around 2.0×10^7. Figure 2 and Table S4 A focus on the group of samples RT_LCM [31], comparing the sum of the expression values to those of neighboring samples in the original dataset. This group of samples shows a "valley" in the sum, compared to most of the others.

Analyzing the results of the sum, we noticed other samples showing lower values compared to most of the others. Graphics for these experiments are presented in Figure S1A–D. In particular, there are two additional groups of experiments with an extremely low sum, in the order of 10^5. The two groups refer to specific studies, [32,33], respectively. Table S4B,C reports the sum of the expression values for these last two groups.

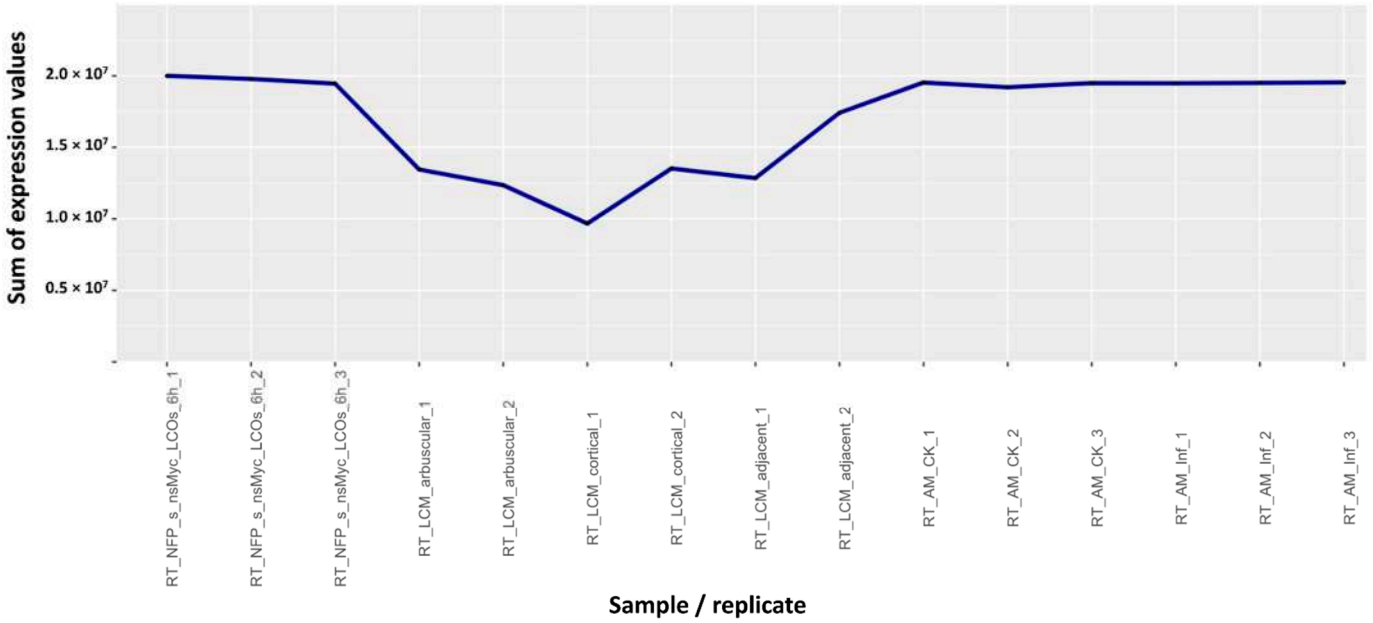


Figure 2. Graphical representation of the sum of the expression values for the RT_LCM samples [31], compared to the ones adjacent to them in the dataset list. All the other samples have sum of expression values close to 2.0×10^7, whereas the RT_LCM samples range between 1.0×10^7 and 1.7×10^7.

2.3. Pearson Correlation Coefficient

To further investigate the quality of data in the MtGEA database, we decided to compute the Pearson correlation coefficient for every pair of replicates in a sample in the downloaded dataset (Table S5). Therefore, we did not consider further those samples with one single hybridization (30 hybridizations in total, listed in Table S6). We focused on samples with at least two replicates, and we have considered acceptable a Pearson correlation coefficient among sample replicate pairs above a threshold value of 0.9. Replicate pairs with correlation values below this threshold are shown in Table S7. Intriguingly, the Pearson correlation coefficients of the last two samples (row) in Table S7 show values that are identical to the sixth decimal. They refer to 24 replicates in total. Each of these replicates has two identities within each row or column in the correlation table (see Table S5), suggesting that the data were duplicated (Table S1). Indeed, expression values for 12 columns are duplicated (e.g., Root_A17_control_1 is identical to HairyRoot_WT_Myc_CK_1, and so on). These two groups refer to different studies [34,35], both characterized by 12 samples (GSE34155 and GSE34617). In addition, we also noted that there are samples (each with three replicates) whose names were duplicated. These refer to one study [12]: Nod_10dpi_1, Nod_10dpi_2, Nod_10dpi_3; Nod_14dpi_1, Nod14dpi_2, Nod14dpi_3 (Table S5). In the dataset downloadable from MtGEA (Table S1) there are 12 columns for these data, but only six can refer to original expression values. This leads to a shift in the attribution of names to samples, with data being referred to other ones starting from these columns.

2.4. Cleaning the Database

First, we removed the six duplicated names mentioned above and related to one study [12]; then, a total of 103 hybridizations were removed from the dataset, obtaining a reduced dataset of 607 hybridizations (Table S8). Hybridizations were discarded according to different criteria:

– Replicates with large variation in expression values as detected by the Pearson correlation coefficients (calculated between replicate pairs). We removed replicates whose sample pairs coefficients were below the 0.90 threshold (Table S7). The number of replicates removed for each sample is as follows: 3 for RT_Myc_3wks_infection, 3 for GiantCell, 3 for GallTissue_GiantCell, 2 for RT_LCM_arbuscular, 2 for RT_LCM_cortical, 2 for

RT_LCM_adjacent, 1 for Nod_Naut1_SalsC, 1 for RT_CRR_72hpi, 1 for RT_CRR_96hpi, 1 for Root_A17_control. (19 columns in total).

- Samples with extremely low (<1 × 10^6) sums of the expression values. All replicates were removed (Table S4B,C). (30 columns).
- Duplicated data. Replicates were removed at alternated lines (Tables S5 and S7). (24 columns).
- Experiments with one measurement for each sample, so-called "single" replicate (Table S6). (30 columns).

To understand how correlation analysis of gene function could be affected by the cleaning of the dataset, we compared Pearson correlation coefficients between *M. truncatula* genes across all hybridizations before and after the cleaning, both on the linear values and the Log-transformed ones. We first focused our attention on genes involved in the biosynthesis of saponins, creating a list of genes extracted from the literature and genes encoding enzymes that synthesize and transform isoprenoids in the cytosol (mevalonate pathway) [36] and in the plastid (non-mevalonate pathway, also known as the MEP/DOXP pathway) [37]. Saponins are a large class of secondary metabolites abundant in legumes, made of a carbohydrate attached to a terpenoid [38]. The flux to these compounds is large and several of the biosynthetic enzymes and precursors are known [39,40]. The saponins' genes are listed in Table S9. The putative mevalonate kinase (Mtr.41545.1.S1_at) was employed as a test gene (Table S10) because we expected large changes in the correlation values for this gene against all the genes after the cleaning (see Figure 1). Indeed, the top correlators in the linear analysis (Table S10A, comparing Linear Original vs. Linear Cleaned; for the difference, see Table S10C) show relevant changes, and the cleaned dataset returns a list with much stronger consistency in gene function. This suggests, not surprisingly, that the mevalonate kinase is involved in isoprenoid and possibly saponins biosynthesis. For instance, the second best correlator, using the cleaned dataset, is MTR_1g017270 (squalene monooxygenase, Mtr.10468.1.S1_at), which shows a correlation value of 0.876, while, using the original dataset, the same pair of probes (Mtr.41545.1.S1_at vs. Mtr.10468.1.S1_at) gives a correlation of 0.597. Most of the genes at the top of the list behave in a similar way. This means that many of the expected correlators become concealed in the original dataset, and hence they are false negatives before cleaning. On the contrary, computing Spearman (Table S10E) or Kendall (Table S10F) correlation coefficients for the same probeset reveals that these methods identify top correlators with a strong biological consistency with both the original and the cleaned dataset. The biological consistency is comparable to Pearson's Linear and Log analysis performed on the cleaned dataset. The two methods based on the rank correlation are, therefore, quite insensitive to the presence of strong outliers, as already known.

We also performed a search for genes targeted by different probes and identified a few more. We present five scatterplots (besides the usual Mevalonate kinase) in Figure S2A–F, created with the whole dataset, not just the mean. As expected, cleaning tends to remove outliers and this results in an increase in the correlation value (B,C,F) in the linear scale, though much less dramatic than that for the mevalonate kinase (A), and a reduction in the log scale (D). One example (E) shows a large reduction in linear correlation.

Expanding the analysis to the saponins' list (Table S11), by comparing the linear correlation table before and after cleaning, it confirms the same trend, as evidenced in the "differential" table (Table S11C). Notably, the results of the Log analysis (Table S10B for the 'One vs. All' approach and Table S12 for the saponins' gene list correlation table) returns a very different outcome. While most of the best correlators are still present in top positions in the Log analysis after cleaning, the Pearson coefficient shows an overall decrease across all the genes (Table S10B for the difference). The top correlators remain, therefore, quite consistent, at least in the case of Mtr.41545.1.S1. Again, the same holds true for the saponins' genes correlation table (Table S12A,B). The same data are presented with two heatmaps (Figure 3) from which it is evident a reduction in the correlation values for many gene pairs, i.e., a reduction in false positives when passing from the

uncleaned to the cleaned dataset. The correlation values were reduced on average by 0.26 (Table S12 C), with many turning from a strong positive correlation (red/orange) to insignificant (white/light blue) (Figure 3). This means that there is in the original dataset something that increases most, if not all, correlation values in the Log analysis (see also the average differences in Table S10D), something that has no such an effect in the linear analysis. The cleaning, thus, shows a different, sometimes opposite, effect on the linear and the Log analysis. Again, focusing on the correlators of the mevalonate kinase (Mtr.41545.1.S1_at) with the saponins' gene subset (original vs. cleaned dataset, Table S13A,B) as well as the differences between the values, we recognize the same pattern: correlation values in the linear analysis (Table S13A) increase on average, while those in the Log analysis (Table S13B) show a substantially larger decrease.

We observed comparable variations in correlation values for several families of transcriptional factors (TFs), one example being the bHLH (basic-Helix–Loop–Helix) family. bHLH is a large TF family, well characterized in eukaryotes and involved in several processes in plants, such as metabolism, growth and responses to stress [41,42]. Table S14 shows the list of *M.truncatula* bHLH genes selected from PlantRegMap (http://plantregmap.gao-lab.org/, accessed on 13 May 2021) [43,44]: in *M. truncatula,* this family included 168 genes in February 2020. An extensive decrease in correlation values is observable in the heatmaps generated from the logarithmic analysis of the bHLH family switching from the original to the cleaned datasets (Table S15 and Figure 4). In short, a great share of the actual correlation values in the Log analysis was reduced upon cleaning, suggesting that many top correlators in the original dataset are actually false positives. This appears to be a general phenomenon and the reason is the presence of a specific group of experiments. We provide an explanation for this phenomenon in the discussion. As performed for the saponins' genes, we compared bHLH Log and linear Pearson correlation values (Tables S15A,B and S16A,B) before and after the cleaning by means of 'differential tables' (Tables S15C and S16C); we observed the same trend highlighted for the saponins' dataset. Similar results were also observed for other TF families, such as WRKY, MYB, ERF and NAC (data not shown).

The bHLH gene family was also analyzed, using Spearman's rank correlation on both the original and cleaned dataset (Table S17A,B), highlighting changes with a 'differential table' (sheet C, Cleaned–Original). Another differential table highlights differences between Pearson's and Spearman's coefficients (sheet D is Pearson's linear coefficient, identical to Table S16B; sheet E, Pearson–Spearman). In this case, Spearman's correlation seems to be quite insensitive to cleaning (average difference is 0.12, Table S17C), but there are instances of large differences between Pearson and Spearman's coefficients calculated for the same pair of probes from the cleaned data. This means that the correlation measure employed may yield different results depending on the gene analyzed (Table S17E).

2.5. AgriGO

To further substantiate the effectiveness of the cleaning process, we also performed a GO enrichment analysis using AgriGO [45,46], working on co-expression results for selected genes using the SEA mining tool (see Discussion and Materials and Methods sections). We picked for each selected gene the 49 best correlators in the original and cleaned datasets. The AgriGO analysis was performed, from both linear and logarithmic correlation values, for the following genes as representatives of fundamental processes (glycolysis/respiration, translation, photosynthesis and gluconeogenesis): Mtr.31871.1.S1_at (pyruvate dehydrogenase E1 beta subunit, PDHE1-B, data not shown), Mtr.10637.1.S1_at and Mtr.34423.1.S1_at (60 ribosomal proteins, data not shown), Mtr.12203.1.S1_at (Rubisco small subunit, data not shown), Mtr.37533.1.S1_at (fructose 1,6-diphosphate phosphatase, data not shown) and Mtr.12230.1.S1_at (translation elongation factor EF-2 subunit, Figures 5 and 6). Again, Mtr.41545.1.S1_at (the putative mevalonate kinase) was used as a test gene for which large changes were expected, which indeed was the case (data not shown). Correlators for Mtr_12330.1.S1_at from the original and cleaned datasets

are shown in Table S18. The variations in the enrichment of GO terms of co-expressed genes between the original and the cleaned dataset strongly suggest that, despite the small number of removed hybridizations, the cleaning improves the quality of the output for several genes and, therefore, its biological significance.

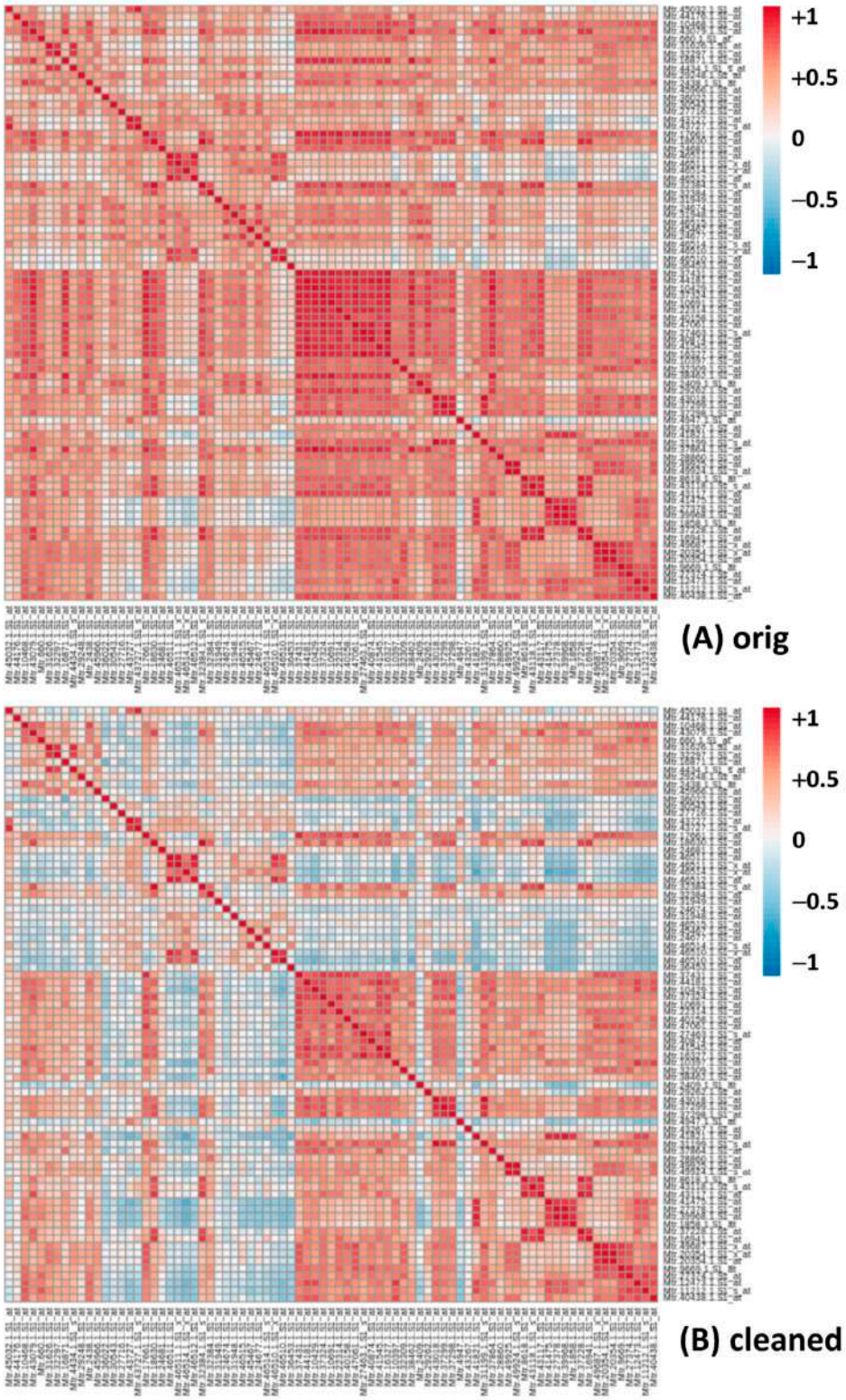


Figure 3. Heatmaps for Pearson correlation coefficients in the logarithmic form for genes of the saponins biosynthetic pathway, with expression values from original (**A**) and cleaned (**B**) Log datasets (Table S12A,B).

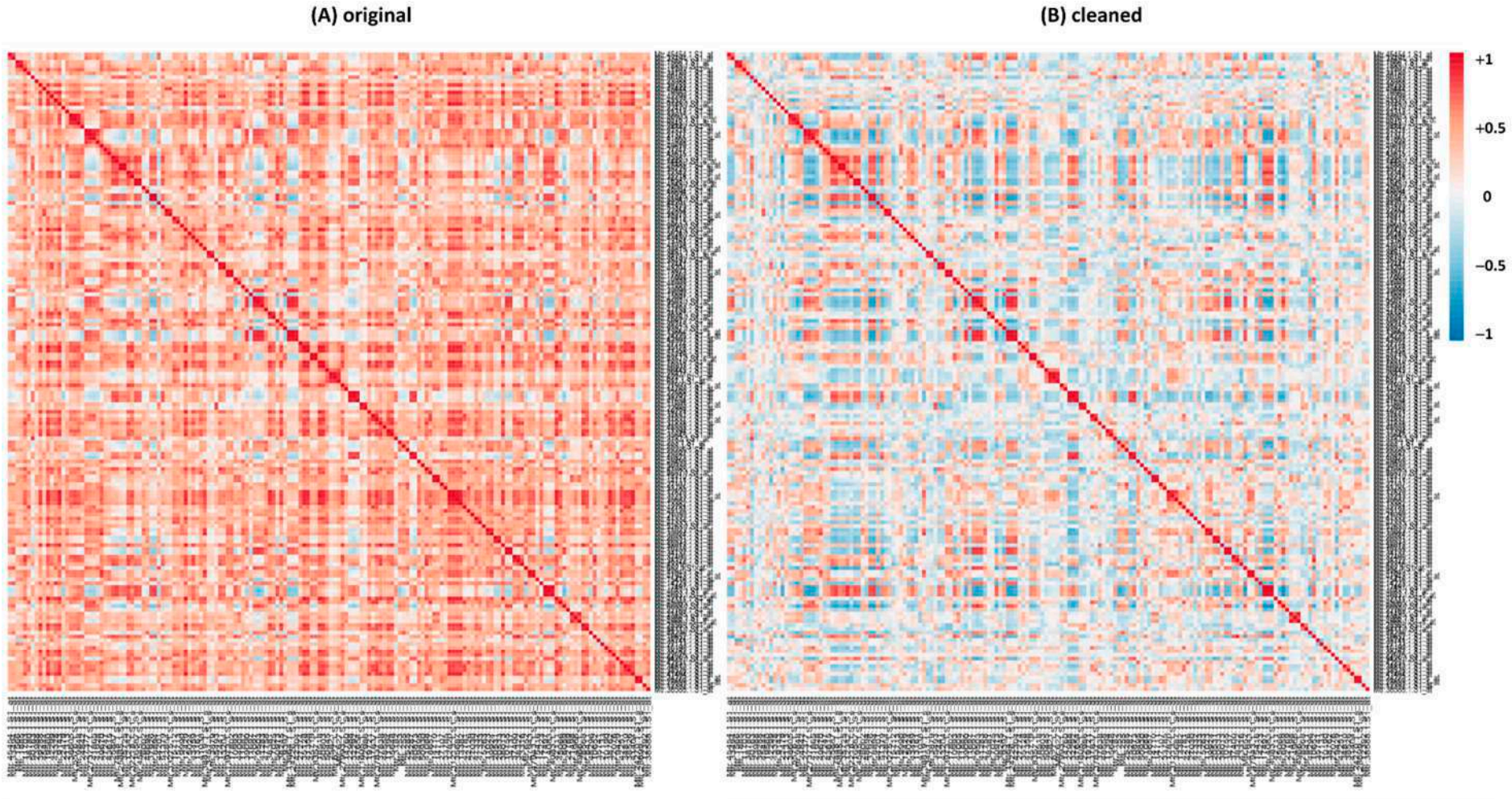


Figure 4. Heatmaps generated for Pearson correlation coefficients in the logarithmic version for the TFs family bHLH, with expression values from original (**A**) and cleaned (**B**) Log datasets (Table S15A,B).

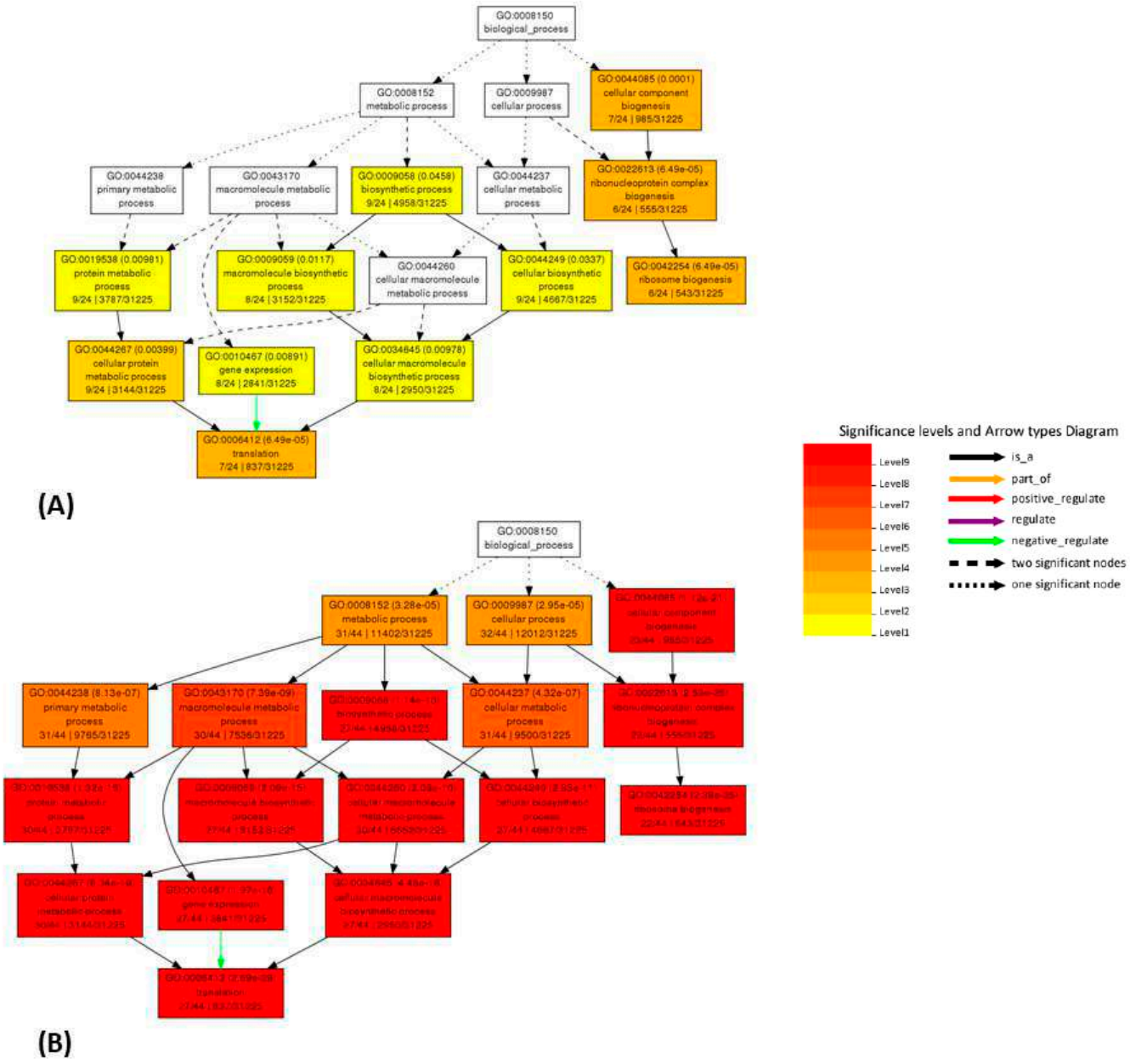


Figure 5. Comparison of AgriGO enrichment analysis working with linear co-expression data for the Mtr.12230.1.S1_at gene (Table S17A). Results obtained from correlation values working on the original (**A**) and cleaned (**B**) dataset.

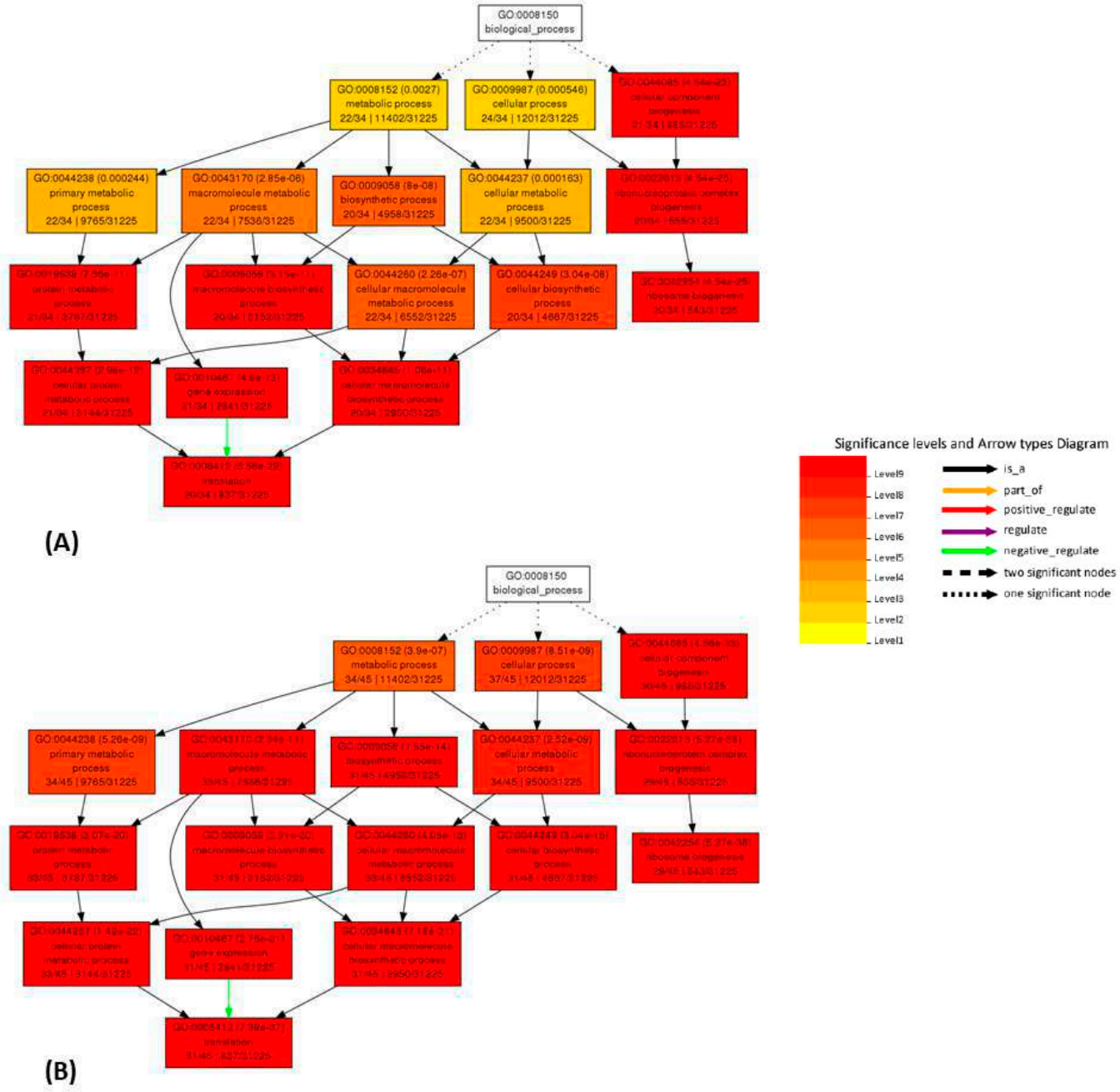


Figure 6. Comparison of AgriGO enrichment analysis working with logarithmic co-expression data for the Mtr.12230.1.S1_at gene (Table S17B). Results obtained from correlation values from the original (**A**) and cleaned (**B**) dataset.

3. Discussion

Our analysis highlighted how the repository *Medicago truncatula Gene Expression Atlas* (*MtGEA*) [12,13], a reference database for microarrays studies of *Medicago* species, contains data that are difficult to explain due to biological or measurement variabilities.

The scatterplot presented in Figure 1 efficiently shows the co-expression of two probes for a putative mevalonate kinase (Mtr.41545.1.S1 and Mtr.16327.1.S1_at). It is possible to observe three points (indicated with red arrows) isolated from all the other ones. Since both probes refer to the same gene, a linear relationship between the two probes is to be expected. Being the only three outliers, we can imagine that they could be due to particular experimental conditions or tissues, or they could be due to errors in measurements or processing of the data. We identified the three outliers as corresponding to means of three samples (RT_LCM_arbuscular, RT_LCM_cortical and RT_LCM_adjacent, [31]), and we

discovered that all of them were performed using the LCM technique, normally used to isolate small portions of tissue or just few cells [47]. With this method, an amplification step is usually required in order to have sufficient material for the hybridization. These outliers are therefore likely the result of spurious amplification and/or other technical problems in the experimental execution.

To understand if similar experiments could generate other outliers in scatter plots of other genes, we decided to use as a first criterion the sum of gene expression values of all probes to detect potential problems. For most hybridizations of the dataset, values of the sum were around 2 x 10^7; they were, however, dramatically lower, even up to 100 times, in just a few cases (Tables S3 and S4). These problematic data are easily identifiable graphically, such as in Figure 2. The RT_LCM samples, however, not only show low values for the sum (Table S3), but also a large difference among replicates for each of the three samples (Table S1). Since replicates of the same tissue/treatment should give very similar expression values, the difference in replicates of the RT_LCM group confirms the possibility that they could be the result of systematic or methodological errors in the experimental procedure, compromising their validity.

We found six other groups of samples whose sum of expression values was lower than most of the samples of the dataset. Even though the reasons of the anomaly could be related to experimental problems for four groups (Figure S1A–D), the groups in Table S4B,C have sums of expression values that are too low to be explained by experimental variability. The first group is associated to one study [32] and includes five samples, each one with three replicates of specific root cell types *Medicago truncatula* in relationship with *Glomus intraradices* mycorrhizal fungus. APP_P and NAP_C are samples at 5–6 days post inoculation (dpi), while ARB-A, CMR_K and EPI_E refer to 21 dpi. The second group includes five samples, each one with three replicates [33]. In this second study, samples from the meristem were analyzed, both in the region infected by the bacterium *Sinorhizobium melitoti* and in the proximal region. These two groups show systematic errors possibly introduced during the insertion of data in MtGEA, or already present in the studies. We analyzed the corresponding values in the downloaded dataset from MtGEA (Table S1). Expression values for these experiments are too low to be biologically relevant, especially for the so-called housekeeping genes, which tend to have a fairly constant expression in organisms. Moreover, the mean values of the replicates calculated for these experiments from the original data were different from the downloaded mean values in Table S2 (data not shown). The most likely error is the insertion of the data for these two experiments after transformation into logarithmic values, even though the means were calculated based on original values (not Log-transformed). This was confirmed by performing an antilog transformation of the values and then measuring their mean values: the results were identical or nearly identical to the mean values uploaded in MtGEA (data not shown).

To point out other anomalies in the experiments, we decided to use the Pearson correlation analysis among sample replicate pairs as the second criterion for the cleaning of the dataset. Using the Pearson correlation coefficient, we could compare the experiments to one another across around 50,000 genes. Thanks to this second criterion, we confirmed that not only the previously identified samples were problematic, but that also other groups of samples could give significant anomalies. Pearson correlation coefficients were analyzed only for samples with at least two replicates: samples with a single hybridization (Table S6) were not considered since we could not test them according to this criterion and were, in any case, few. One study argues for their removal [6]. We considered threshold values for correlation coefficients as 0.90 and, according to the results of the analysis, four possible cases could occur: (1) values of the correlation coefficients for all replicates studied are above the threshold value, thus all data are retained; (2) one correlation coefficient for three replicates is below the threshold value, thus only the pair that has the highest value above the threshold is kept; (3) two correlation coefficients are below the threshold value, thus only the pair showing a correlation value above the threshold is kept; and (4) all correlation values measured are below the threshold value, thus all replicates are

discarded. By excluding samples with one single hybridization (Table S6), we automatically discarded also those in Table S4B,C for which we established a possible systematic error in the data. Performing the analysis on all the other samples, we highlighted correlation values below the threshold of 0.90 as presented in Table S7. Moreover, analyzing these results, we highlighted that the last two groups of Table S7 have identical correlation values. Checking the original data (Table S1), we discovered that there are 24 columns duplicated (8 samples each with 3 replicates), due to an error during data insertion in MtGEA. One first group refers to a study [35] on nodules treated with phosphinothricin at different times. A second study presents data on the roots of different genotypes in relationship to different symbionts [34]. We also identified another group of experiment whose names were duplicated [12]. This groups caused problems in correctly reading the original dataset since not all duplicated columns were filled, generating problems in the organization of data in columns and in the analysis.

From our work on the MTGEA database (https://mtgea.noble.org/v3/, accessed on 13 May 2021), we discovered that the repository contains errors and poor quality data that affect subsequent analyses. After the cleaning of MtGEA, using the criteria of the sum of the expression values and the Pearson correlation coefficients among replicates pairs, we reduced the number of columns in the dataset from 716 (710 once removed the duplicated names of two samples in the dataset heading—Nod_10dpi and Nod_14dpi [12]) to 607, actually removing 103 hybridizations (around 15%). To verify that the cleaning results in statistically significant improvements, we performed Pearson correlation analyses among genes of different pathways/processes out of the original and cleaned datasets. In order to reduce the number of false positive correlations with no biological significance, it is important to remove unreliable and misleading data. We first studied the saponins biosynthetic pathway, which is particularly important in *Medicago truncatula*, as well as in other leguminous plants used for animal feeding. Saponins, because of their bitter taste, can reduce the appeal of the feed and can be toxic [48]. *Medicago* spp. saponins are triterpenic saponins derived from 2,3-oxidosqualene, the last common precursor of sterols and triterpenes, synthesized into the cytosol from isopentenyl pyrophosphate (IPP) [39,49]. The positive effect of MtGEA cleaning on Pearson correlation analysis can be immediately observed in Figure 3. Comparing the heatmaps obtained from the Log-trasformed original and cleaned data for the genes involved in saponins biosynthesis (Table S12A,B), it is evident a drastic reduction in correlation strength of many genes, meaning that the cleaning reduced the number of false positives.

We also compared the results obtained on some genes with Pearson to those obtained with rank correlation methods. Spearman and Kendall's correlation performed well and were not much affected by the cleaning; however, the results did not always coincide with Pearson's correlation (Table S17E). This is a rather trivial observation because Pearson's measure detects linear correlation, while rank correlation methods detect also other forms of correlation (e.g., hyperbolic, sigmoid . . .). For this reason, it may be interesting to use both Pearson's (both Lin and Log) as well as Spearman's for correlation analysis when searching for candidate genes. Kendall's is more computationally intensive, and it could be used only for specific cases.

For some specific genes, the results of the correlation analysis after the cleaning change significantly, also ameliorating the results of the co-expression analysis. This is easily observable analyzing Table S10A,B, showing the correlators of the putative mevalonate kinase Mtr.41545.1.S1_at. Before the cleaning, this gene showed just a few high correlations with genes in the list of saponins biosynthetic genes, while after the cleaning, its biological consistency increases significantly. With the original data, Mtr.41545.1.S1_at best correlators are mostly genes without annotation or involved in pathways unrelated to saponins biosynthesis, such as ribosomal proteins. After cleaning, Mtr.41545.1.S1_at correlates with many genes involved in isoprene and saponins biosynthesis, meaning that the cleaning improved the biological significance of the correlation. In fact, 11 of the 15 best Log correlators of Mtr.41545.1.S1_at, using cleaned data, are attributable to isoprenoid biosynthesis

(IPP), particularly in the mevalonate cytosolic biosynthetic pathway (Table S10B) and only four with the original data. This underlines that the cleaning not only reduced the number of false positives, but also reduced the number of false negatives (genes that were not well correlated before the cleaning but improved thereafter).

The importance of the cleaning of MtGEA before running a correlation analysis is also evident upon analyzing the Pearson correlation coefficients for several TFs family, for instance, the bHLH gene family (Figure 4 and Table S15A,B). Again, a great reduction in false positives is evident when switching from the original to the cleaned dataset. The cleaning helps in ameliorating the results of the correlation analysis, reducing the number of spurious correlations and helping in focusing only on strong correlations that can have a biological significance. Analogous results have been found for other TFs, such as WRKY, MYB, ERF and NAC (data not shown), leading us to believe that false positives are mainly found in the Log analysis when working on original data of TFs. In some cases, specific predictions can be made on the basis of the best correlators; for instance, the probe Mtr.34810.1.S1_at (Mtr_8g065740) refers to a TF, without further annotation. On the basis of the correlators, we anticipate for this gene a role in chromosome maintenance/stability/DNA repair because half of the best 20 correlators in the Log analysis fall into this category (data not shown). This is not at all evident in the list obtained with the original data. The same is true for the linear analysis, albeit the numbers are less striking.

As another example, probe Mtr.5966.1.S1_at, which identify a class III peroxidase, correlates best with two probes (Mtr.42141.1.S1_s_at and Mtr.42141.1.S1_at) in the linear analysis; both probes recognize another peroxidase (MtPRX1, MTR_3g094630) [50]. The very same probes are at positions 9 and 10, respectively, in the correlators list generated with the original dataset, implying a significant change in their degree of correlation with probe Mtr.5966.1.S1_at (data not shown). All three probes refer to transcripts strongly induced by elicitor treatments [51]. Surprisingly, the correlation analysis for another peroxidase (PRX3, Mtr.40125.1.S1_at), apparently not involved in aurone biosynthesis [50], suggests a strong involvement in disease resistance, a conclusion based on the frequency of GO terms among the best correlators. This is evident mainly in the Log analysis (both of original and cleaned datasets) and suggests that it is always worth performing a Log analysis beside the linear one.

AgriGO analysis, on original and cleaned datasets, was also used to measure the improvement in biological consistency. AgriGO is a web tool that allows to perform the gene ontology analysis, focusing the attention on species of agricultural interest [45,46]. Different mining tools are available; we used the Singular Enrichment Analysis (SEA) that provides an enrichment analysis of GO terms for a list of genes/probes, with the aim to find GO terms that are statistically enriched in a list compared to an expected value for a given species. This analysis was performed for several genes, using the respective 49 best correlators. Several of the analyzed genes show improvements, for example: Mtr.41545.1.S1_at (putative mevalonate kinase, data not shown), Mtr.31871.1.S1_at (pyruvate dehydrogenase E1 beta subunit, PDHE1-B, data not shown) and Mtr.12230.1.S1_at (translation elongation factor EF-2 subunit, Figures 5 and 6 and Table S18A,B). The AgriGO analysis for EF-2 highlights a great change in the enriched GO terms before and after the cleaning, evidenced by an intensification of the red color, both in linear and logarithmic forms, after the cleaning. Comparing original and cleaned data, the identified processes are the same, but there is a great difference in significance levels, which increase upon cleaning.

It is noteworthy that our approach is not only conceptually and computationally very simple, but it does not require prior knowledge of the biological samples or the species, differently to methods for the removal of unwanted variations that are based on normalizations with respect to a set of control genes [52–54]. These other methods can, thus, be seen as complementary and could be applied before or after processing the datasets according to our approach.

4. Conclusions

Our analysis on the MtGEA database has not only improved the quality of the data but underlined the importance of a proper cleaning step before any kind of correlation analysis on microarray data. Moreover, we have established a simple strategy of general validity for the cleaning of microarray datasets based on two criteria: the sum of the expression values across all genes in samples and the Pearson correlation analysis among sample replicate pairs. We demonstrated how the cleaning can strongly affect the transcript correlation analysis. We found that the removal of a limited number of problematic samples ameliorates the results of the correlation analysis (both reducing false positives and false negatives) and, consequently, of related predictions that are more supported by the annotation, literature and GO terms' frequency. We believe that this approach is of general applicability and could be expanded beyond the *Affymetrix* technology.

5. Materials and Methods

5.1. Microarray Data

The microarray data used in this study were downloaded from the *Medicago truncatula Gene Expression Atlas* (MtGEA) (https://mtgea.noble.org/v3/, accessed on 13 May 2021) [12,13]. The dataset consists of gene expression profiles from 36 experiments. It was generated by selecting as download options "All Replicates" for "Experiment Selection" and "Mtr:Medicago truncatula only" for "Probeset Selection". As of February 2020, the complete dataset was composed by 50,900 genes with 710 hybridizations (the downloaded dataset initially included 716 columns because of the name duplication of two samples' replicates, see Results and Discussion sections for further details). The dataset, including the mean values of experimental replicates (274 columns), was downloaded by selecting "All Means" and "Mtr: Medicago truncatula only".

5.2. R

Data analysis was conducted in the R programming environment (https://www.R-project.org/, version 4.0.0) [55], and figures were produced using the following packages: *ggplot2* (https://cran.r-project.org/web/packages/ggplot2/index.html) [56], *pheatmap* (https://cran.r-project.org/web/packages/pheatmap/index.html) [57], *data.table* https://cran.r-project.org/web/packages/data.table/index.html [58], *Hmisc* (https://cran.r-project.org/web/packages/Hmisc/index.html) [59] and *RColorBrewer* (https://cran.r-project.org/web/packages/RColorBrewer/index.html) [60]. Annotations were performed and checked using g:Profiler (https://biit.cs.ut.ee/gprofiler/gost) [61] and the *Affymetrix* microarray annotations for *Medicago* (http://tools.thermofisher.com/content/sfs/supportFiles/Medicago-na36-annot-csv.zip). All pages were accessed on 15 April 2021.

5.3. AgriGO

Gene ontology analysis was performed using AgriGO (http://bioinfo.cau.edu.cn/agriGO/index.php, accessed on 14 May 2021) [45,46], while Singular Enrichment Analysis (SEA) was the tool used to identify GO terms statistically enriched in a provided list of genes.

Supplementary Materials: The following materials are available online at https://www.mdpi.com/article/10.3390/plants10061240/s1, Figure S1: Graphical representation of the sum of expression values for selected groups of experiments compared to the neighboring ones in the dataset list. The great majority of experiments have values around 2.0×10^7 while a few groups show large reduction in the values. (A) RT_Myc_3wks_infection [Balzergue et al., unpublished], with sum values under 1.5×10^7; (B) Group of GiantCell, GallTissue–GiantCell and Nod_zone2 experiments (Balzergue et al., unpublished), with sum values in the 1.1–1.5 $\times 10^7$ range; (C) Group of X1, X3, X5_dpi and 24h experiments (Breakspear et al., unpublished) with sum values in the 1.25–1.70 $\times 10^7$ range; (D) Group of Nod_Naut1_SalsC, Nod_Naut1_SalsB, Nod_Sals4_SalsB and Nod_Sals4_SalsC experiments [62], with sum values in the 0.5–1.0 $\times 10^7$ range. Figure S2: Scatterplot of the expression levels of different probe pairs, hybridizing to the same gene, before (light blue)

and after (red) the cleaning procedure. (A) Mtr_7g113660, Mevalonate kinase: (B) Mtr_3g072350, WEB family plant protein; (C) Mtr_4g116460, zinc finger, C3HC4 type (RING finger) protein; (D) Same as (C) but in Log scale; (E) MTR_6g029470, Galactose-1-phosphate uridylyltransferase; (F) MTR_1g029400 transcriptional corepressor SEUSS-like protein. The respective correlation values and Affy codes are inserted into each graph. Table S1: Downloaded dataset from MtGEA in February 2020 (716 columns). Table S2: Downloaded mean values for all experiments of MtGEA in February 2020 (274 columns). Table S3: Sum of the expression values for all the experiments in the downloaded MtGEA dataset. Table S4: Sum of the expression values for experiments as detailed in references [31–33]. (A) Sum of the expression values for RT_LCM experiments [31] compared to the ones of RT_NFP_nsMyc_LCOs_6h, RT_NFP_sMyc_LCOs_6h, RT_NFP_s_nsMyc_LCOs_6h, RT_AM_CK and RT_AM_Inf. All samples are characterized by three replicates, except the RT_LCM series. (B) Sum of the expression values for samples as in Ref. [32]. Each sample has three replicates. (C) Sum of the expression values for samples of Ref. [33]. Each sample has three replicates. Table S5: Pearson correlation coefficient among hybridizations in MtGEA downloaded dataset. In yellow, samples with one single hybridization and samples reported in Table S4B,C. In green, samples from [34,35] showing identical Pearson correlation coefficients. In blue, samples that in the downloaded dataset show two attributed columns for replicate [12]. Table S6: List of samples in the MtGEA database with a single hybridization per sample (i.e., single replicate). Table S7: Groups of samples showing Pearson correlation coefficients under a threshold value of 0.90. In red, acceptable Pearson correlation coefficients. Values are approximated to the second decimal. We used a slash "/" to indicate that the experiment has only 2 replicates in which case only one single pair of Pearson correlation coefficient could be calculated. Acceptable pairs are highlighted in red. References: [1][31], [2][62], [3][63], [4][34]; Table S8: Cleaned dataset (607 hybridizations). Table S9: List of genes involved in or attributable to saponins biosynthetic pathways. Annotation was performed using g:Profiler. Table S10: Correlation values between the expression values of Mtr.41545.1.S1_at and the expression values of all the other genes present in the *Medicago Affymetrix* microarray. Calculation was performed on the original (left) and the cleaned (right) dataset. Annotation was performed using g:Profiler. Color backgrounds are used for quick visual identification of specific genes. (A) Pearson's correlation values from linear data (B) Pearson correlation values after Log-transformation (C,D) Difference in the value of the Pearson coefficient for each probe between the cleaned and the original dataset, using the Linear (C) or Log (D) data. (E) Spearman correlation values (F) Kendall's correlation values. Annotations in red characters identify genes attributed to the mevalonate/saponin pathway. Table S11: (A) Linear Pearson correlation coefficient for the saponins' genes using the original dataset, (B) linear Pearson correlation coefficient for saponins' genes using the cleaned dataset, (C) difference between the Pearson correlation coefficient of Table S11 B (Cleaned) and of Table S11 A (Original). Table S12: (A) Logarithmic Pearson correlation coefficient for saponins' genes using the original dataset; (B) logarithmic Pearson correlation coefficient for saponins' genes using the cleaned dataset; (C) difference between the Pearson correlation coefficient of Table S12 B (Cleaned) and of Table S12 A (Original). Table S13: (A) Pearson correlation coefficient between linear values of Mtr.41545.1.S1_at and of saponins' gene list using the original (column A), the cleaned dataset (column C); column E represents the difference between Cleaned and Original; column F is the *Affymetrix* code. The data in columns A and C are the same as in column AV of Table S11A,B, (B) Pearson correlation coefficient between Log values of Mtr.41545.1.S1_at and of the saponins' gene list using the original (column A), the cleaned dataset (column C); column E represents the difference between cleaned and original; column F is the *Affymetrix* code. The data in columns A and C are the same as in column AV of Table S12A,B. Table S14: List of the 168 genes bHLH transcriptional factor family members analyzed. They were selected from PlantRegMap (http://planttfdb.gao-lab.org/family.php?sp=Mtr&fam=bHLH accessed on 1 February 2020) [43,44]. Table S15: (A) Logarithmic Pearson correlation coefficient for bHLH genes using the original dataset, (B) logarithmic Pearson correlation coefficient for bHLH genes using the cleaned dataset, (C) difference between the Pearson correlation coefficient of Table S15B (Cleaned) and of Table S15 A (Original). Table S16: (A) Linear Pearson correlation coefficient for bHLH genes using the cleaned dataset, (B) linear Pearson correlation coefficient for bHLH genes using the clean dataset, (C) difference between the Pearson correlation coefficient of Table S16B (Cleaned) and of Table S16A (Original). Table S17: Spearman correlation coefficient for bHLH genes using the original (A) and the cleaned dataset (B); (C) is the difference (Table B, Table A); (D) is the same as Table S16B and (E) is the differential table (Pearson's–Spearman's, that is, Table D, Table B) only for the cleaned dataset. Table S18: Pearson correlation values between the expression values of

Mtr.12330.1.S1_at and the expression values of all the other genes present in the *Medicago Affymetrix* microarray. Calculation was performed on the original and the cleaned linear dataset. Annotation was done with g:Profiler. Scripts: Scripts and details on the procedures/software used for processing the microarray data.

Author Contributions: Conceptualization P.M.; methodology F.M., C.W., L.M., G.P.; data analysis F.M., C.W., P.M.; writing: F.M., P.M., review and editing F.M., C.W., L.M., G.P., P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a FFABR grant ('Fondo di Finanziamento per le Attività Base di Ricerca') from MIUR in 2017 to P.M. The APC was funded by Università degli Studi di Milano ('Fondo di Ateneo per APC') and MIUR grant "Progetto Bandiera Epigenomica" to G.P.

Data Availability Statement: The data presented in this study are either openly accessible at the original repositories (as detailed in Material and Methods) or provided as supplementary material.

Acknowledgments: We thank Sergio Saia for critical reading of the manuscript and useful suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Richard, A.; Louise, P.H. "Omic" technologies: Genomics, transcriptomics, proteomics and metabolomics. *Obstet. Gynaecol.* **2011**, *13*, 189–195.
2. Huang, X.Y.; Salt, D.E.E. Plant Ionomics: From Elemental Profiling to Environmental Adaptation. *Mol. Plant* **2016**, *9*, 787–797. [CrossRef] [PubMed]
3. Lowe, R.; Shirley, N.; Bleackley, M.; Dolan, S.; Shafee, T. Transcriptomics technologies. *PLoS Comput. Biol.* **2017**, *13*, 1–23. [CrossRef] [PubMed]
4. Quackenbush, J. Microarray data normalization and transformation. *Nat. Genet.* **2002**, *32*, 496–501. [CrossRef] [PubMed]
5. Park, T.; Yi, S.G.; Kang, S.H.; Lee, S.Y.; Lee, Y.S.; Simon, R. Evaluation of normalization methods for microarray data. *BMC Bioinform.* **2003**, *4*. [CrossRef]
6. Slonim, D.K.; Yanai, I. Getting started in gene expression microarray analysis. *PLoS Comput. Biol.* **2009**, *5*. [CrossRef]
7. Bumgarner, R. Overview of DNA microarrays: Types, applications, and their future. *Curr. Protoc. Mol. Biol.* **2013**, 1–11. [CrossRef]
8. Ledford, H. The death of microarrays? *Nature* **2008**, *455*, 847. [CrossRef]
9. Edgar, R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [CrossRef]
10. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2013**, *41*, 991–995. [CrossRef]
11. Huala, E.; Dickerman, A.W.; Garcia-Hernandez, M.; Weems, D.; Reiser, L.; LaFond, F.; Hanley, D.; Kiphart, D.; Zhuang, M.; Huang, W.; et al. The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* **2001**, *29*, 102–105. [CrossRef] [PubMed]
12. Benedito, V.A.; Torres-Jerez, I.; Murray, J.D.; Andriankaja, A.; Allen, S.; Kakar, K.; Wandrey, M.; Verdier, J.; Zuber, H.; Ott, T.; et al. A gene expression atlas of the model legume Medicago truncatula. *Plant J.* **2008**, *55*, 504–513. [CrossRef] [PubMed]
13. He, J.; Benedito, V.A.; Wang, M.; Murray, J.D.; Zhao, P.X.; Tang, Y.; Udvardi, M.K. The Medicago truncatula gene expression atlas web server. *BMC Bioinform.* **2009**, *10*. [CrossRef] [PubMed]
14. Gholami, A.; De Geyter, N.; Pollier, J.; Goormachtig, S.; Goossens, A. Natural product biosynthesis in Medicago species. *Nat. Prod. Rep.* **2014**, *31*, 356–380. [CrossRef]
15. Kang, Y.; Li, M.; Sinharoy, S.; Verdier, J. A snapshot of functional genetic studies in Medicago truncatula. *Front. Plant Sci.* **2016**, *7*. [CrossRef] [PubMed]
16. Barker, D.G.; Bianchi, S.; Blondon, F.; Dattée, Y.; Duc, G.; Essad, S.; Flament, P.; Gallusci, P.; Génier, G.; Guy, P.; et al. Medicago truncatula, a model plant for studying the molecular genetics of the Rhizobium-legume symbiosis. *Plant Mol. Biol. Rep.* **1990**, *8*, 40–49. [CrossRef]
17. Graham, P.H.; Vance, C.P. Update on Legume Utilization Legumes: Importance and Constraints to Greater Use. *Plant Physiol.* **2003**, *131*, 872–877. [CrossRef] [PubMed]
18. Young, N.D.; Debellé, F.; Oldroyd, G.E.D.; Geurts, R.; Cannon, S.B.; Udvardi, M.K.; Que, F. The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **2011**, *480*, 5–9. [CrossRef]
19. Li, J.; Dai, X.; Liu, T.; Zhao, P.X. LegumeIP: An integrative database for comparative genomics and transcriptomics of model legumes. *Nucleic Acids Res.* **2012**, *40*, 1221–1229. [CrossRef]
20. Wang, M.; Verdier, J.; Benedito, V.A.; Tang, Y.; Murray, J.D.; Ge, Y.; Becker, J.D.; Carvalho, H.; Rogers, C.; Udvardi, M.; et al. LegumeGRN: A Gene Regulatory Network Prediction Server for Functional and Comparative Studies. *PLoS ONE* **2013**, *8*. [CrossRef]

21. Dalma-Weiszhausz, D.D.; Warrington, J.; Tanimoto, E.Y.; Miyada, C.G. The Affymetrix GeneChip®Platform: An Overview. In *Methods in Enzymology*; Academic Press: Cambridge, MA, USA, 2006; Volume 410, pp. 3–28. ISBN 0121828158.
22. Franzese, M.; Iuliano, A. Correlation analysis. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 1–3, pp. 706–721. ISBN 9780128114322.
23. Akoglu, H. User's guide to correlation coefficients. *Turk. J. Emerg. Med.* **2018**, *18*, 91–93. [CrossRef]
24. Murgia, I.; Tarantino, D.; Soave, C.; Morandini, P. Arabidopsis CYP82C4 expression is dependent on Fe availability and circadian rhythm, and correlates with genes involved in the early Fe deficiency response. *J. Plant Physiol.* **2011**, *168*, 894–902. [CrossRef]
25. Månsson, R.; Tsapogas, P.; Åkerlund, M.; Lagergren, A.; Gisler, R.; Sigvardsson, M. Pearson Correlation Analysis of Microarray Data Allows for the Identification of Genetic Targets for Early B-cell Factor. *J. Biol. Chem.* **2004**, *279*, 17905–17913. [CrossRef] [PubMed]
26. Zermiani, M.; Begheldo, M.; Nonis, A.; Palme, K.; Mizzi, L.; Morandini, P.; Nonis, A.; Ruperti, B. Identification of the arabidopsis RAM/MOR signalling network: Adding new regulatory players in plant stem cell maintenance and cell polarization. *Ann. Bot.* **2015**, *116*, 69–89. [CrossRef] [PubMed]
27. Beekweelder, J.; van Leeuwen, W.; van Dam, N.M.; Bertossi, M.; Grandi, V.; Mizzi, L.; Soloviev, M.; Szabados, L.; Molthoff, J.W.; Schipper, B.; et al. The impact of the absence of aliphatic glucosinolates on insect herbivory in Arabidopsis. *PLoS ONE* **2008**, *3*. [CrossRef] [PubMed]
28. Naoumkina, M.A.; Modolo, L.V.; Huhman, D.V.; Urbanczyk-Wochniak, E.; Tang, Y.; Sumner, L.W.; Dixon, R.A. Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in Medicago truncatula. *Plant Cell* **2010**, *22*, 850–866. [CrossRef]
29. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [CrossRef] [PubMed]
30. Yon Rhee, S.; Wood, V.; Dolinski, K.; Draghici, S. Use and misuse of the gene ontology annotations. *Nat. Rev. Genet.* **2008**, *9*, 509–515. [CrossRef]
31. Gaude, N.; Bortfeld, S.; Duensing, N.; Lohse, M.; Krajinski, F. Arbuscule-containing and non-colonized cortical cells of mycorrhizal roots undergo extensive and specific reprogramming during arbuscular mycorrhizal development. *Plant J.* **2012**, *69*, 510–528. [CrossRef]
32. Hogekamp, C.; Küster, H. A roadmap of cell-type specific gene expression during sequential stages of the arbuscular mycorrhiza symbiosis. *BMC Genom.* **2013**, *14*. [CrossRef]
33. Limpens, E.; Moling, S.; Hooiveld, G.; Pereira, P.A.; Bisseling, T.; Becker, J.D.; Küster, H. Cell- and Tissue-Specific Transcriptome Analyses of Medicago truncatula Root Nodules. *PLoS ONE* **2013**, *8*. [CrossRef]
34. Ortu, G.; Balestrini, R.; Pereira, P.A.; Becker, J.D.; Küster, H.; Bonfante, P. Plant genes related to gibberellin biosynthesis and signaling are differentially regulated during the early Stages of AM fungal interactions. *Mol. Plant* **2012**, *5*, 951–954. [CrossRef]
35. Seabra, A.R.; Pereira, P.A.; Becker, J.D.; Carvalho, H.G. Inhibition of glutamine synthetase by phosphinothricin leads to transcriptome reprograming in root nodules of Medicago truncatula. *Mol. Plant Microbe Interact.* **2012**, *25*, 976–992. [CrossRef]
36. Vranová, E.; Coman, D.; Gruissem, W. Structure and Dynamics of the Isoprenoid Pathway Network. *Mol. Plant* **2012**, *5*, 318–333. [CrossRef]
37. Dubey, V.S.; Bhalla, R.; Luhtra, R. An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants. *J. Biosci.* **2003**, *28*, 637–646. [CrossRef]
38. Savage, G.P. Saponins. In *Encyclopedia of Food Science and Nutrition*, 2nd ed.; Academic Press: Cambridge, MA, USA, 2003; pp. 5095–5098. ISBN 9780122270550. [CrossRef]
39. Tava, A.; Scotti, C.; Avato, P. Biosynthesis of saponins in the genus Medicago. *Phytochem. Rev.* **2011**, *10*, 459–469. [CrossRef]
40. Liu, C.; Ha, C.M.; Dixon, R.A. Functional genomics in the study of metabolic pathways in Medicago truncatula: An overview. *Methods Mol. Biol.* **2018**, *1822*, 315–337. [CrossRef]
41. Sun, X.; Wang, Y.; Sui, N. Transcriptional regulation of bHLH during plant response to stress. *Biochem. Biophys. Res. Commun.* **2018**, *503*, 397–401. [CrossRef]
42. Toledo-Ortiz, G.; Huq, E.; Quail, P.H. The Arabidopsis Basic/Helix-Loop-Helix Transcription Factor Family. *Plant Cell* **2003**, *15*, 1749–1770. [CrossRef] [PubMed]
43. Tian, F.; Yang, D.C.; Meng, Y.Q.; Jin, J.; Gao, G. PlantRegMap: Charting functional regulatory maps in plants. *Nucleic Acids Res.* **2020**, *48*, D1104–D1113. [CrossRef]
44. Jin, J.; Tian, F.; Yang, D.C.; Meng, Y.Q.; Kong, L.; Luo, J.; Gao, G. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **2017**, *45*, D1040–D1045. [CrossRef]
45. Du, Z.; Zhou, X.; Ling, Y.; Zhang, Z.; Su, Z. agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **2010**, *38*, 64–70. [CrossRef]
46. Tian, T.; Liu, Y.; Yan, H.; You, Q.; Yi, X.; Du, Z.; Xu, W.; Su, Z. AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **2017**, *45*, W122–W129. [CrossRef]
47. Emmert-Buck, M.R.; Bonner, R.F.; Smith, P.D.; Chuaqui, R.F.; Zhuang, Z.; Goldstein, S.; Weiss, R.A.; Liotta, L.A. Laser Capture Microdissection. *Science* **1996**, *274*, 998–1001. [CrossRef]

48. Wina, E.; Muetzel, S.; Becker, K. The impact of saponins or saponin-containing plant materials on ruminant production—A review. *J. Agric. Food Chem.* **2005**, *53*, 8093–8105. [CrossRef]
49. Carelli, M.; Biazzi, E.; Panara, F.; Tava, A.; Scaramelli, L.; Porceddu, A.; Graham, N.; Odoardi, M.; Piano, E.; Arcioni, S.; et al. Medicago truncatula CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell* **2011**, *23*, 3070–3081. [CrossRef]
50. Farag, M.A.; Deavours, B.E.; de Fáltima, Â.; Naoumkina, M.; Dixon, R.A.; Sumner, L.W. Integrated metabolite and transcript profiling identify a biosynthetic mechanism for hispidol in Medicago truncatula cell cultures. *Plant Physiol.* **2009**, *151*, 1096–1113. [CrossRef]
51. Naoumkina, M.A.; He, X.; Dixon, R.A. Elicitor-induced transcription factors for metabolic reprogramming of secondary metabolism in Medicago truncatula. *BMC Plant Biol.* **2008**, *8*, 1–14. [CrossRef]
52. Jacob, L.; Gagnon-Bartsch, J.A.; Speed, T.P. Correcting gene expression data when neither the unwanted variation nor the factor of interest are observed. *Biostatistics* **2016**, *17*, 16–28. [CrossRef]
53. Freytag, S.; Gagnon-Bartsch, J.; Speed, T.P. Systematic noise degrades gene co-expression signals but can be corrected. *BMC Bioinform.* **2015**, *16*, 309. [CrossRef]
54. Varma, S. Blind estimation and correction of microarray batch effect. *PLoS ONE* **2020**, *15*, e0231446. [CrossRef]
55. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
56. Wickham, H. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*, R. Package Version 3.3.2; 2009. Available online: https://cran.r-project.org/web/packages/ggplot2/index.html (accessed on 15 April 2021).
57. Kolde, R. *pheatmap: Pretty Heatmaps*, R Package Version 1.0.12; 2019. Available online: https://cran.r-project.org/web/packages/pheatmap/index.html (accessed on 15 April 2021).
58. Dowle, M. *Package "Data.Table"*, R Package Version 1.14.0; 2021. Available online: https://cran.r-project.org/web/packages/data.table/index.html (accessed on 15 April 2021).
59. Harrell, F.E., Jr. *Package "Hmisc"*, R Package Version 4.5-0; 2021. Available online: https://cran.r-project.org/web/packages/Hmisc/index.html (accessed on 15 April 2021).
60. Neuwirth, E. *Package "RColorBrewer"*, R Package Version 1.1-2; 2014. Available online: https://cran.r-project.org/web/packages/RColorBrewer/index.html (accessed on 15 April 2021).
61. Raudvere, U.; Kolberg, L.; Kuzmin, I.; Arak, T.; Adler, P.; Peterson, H.; Vilo, J. g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **2019**, *47*, W191–W198. [CrossRef] [PubMed]
62. Heath, K.D.; Burke, P.V.; Stinchcombe, J.R. Coevolutionary genetic variation in the legume-rhizobium transcriptome. *Mol. Ecol.* **2012**, *21*, 4735–4747. [CrossRef] [PubMed]
63. Uppalapati, S.R.; Marek, S.M.; Lee, H.K.; Nakashima, J.; Tang, Y.; Sledge, M.K.; Dixon, R.A.; Mysore, K.S. Global gene expression profiling during *Medicago truncatula-Phymatotrichopsis omnivora* interaction reveals a role for jasmonic acid, ethylene, and the flavonoid pathway in disease development. *Mol. Plant Microbe Interact.* **2009**, *22*, 7–17. [CrossRef] [PubMed]

Table S4: Sum of the expression values for experiments of [31], [32] and [33]

A) Sum of the expression values for RT_LCM experiments [31] compared to the ones of RT_NFP_nsMyc_LCOs_6h, RT_NFP_sMyc_LCOs_6h, RT_NFP_s_nsMyc_LCOs_6h, RT_AM_CK and RT_AM_Inf. All experiments are characterized by three replicates except the RT_LCM series

RT_NFP_nsMyc_LCOs_6h	RT_NFP_sMyc_LCOs_6h	RT_NFP_s_nsMyc_LCOs_6h	RT_LCM_arbuscular	RT_LCM_cortical	RT_LCM_adjacent	RT_AM_CK	RT_AM_Inf
1.99×10^7	1.98×10^7	2×10^7	1.35×10^7	9.68×10^6	1.29×10^7	1.95×10^7	$1.95 \times 10_7$
1.98×10^7	2×10^7	1.98×10^7	1.24×10^7	1.35×10^7	1.74×10^7	1.92×10^7	1.95×107
1.99×10^7	2×10^7	1.95×10^7	/	/	/	1.95×10^7	1.95×10^7

B) Sum of the expression values for experiments of [32]. Each experiment has three replicates

APP_P	NAP_C	ARB_A	CMR_K	EPI_E
1.52×10^5	1.46×10^5	1.18×10^5	1.13×10^5	1.15×10^5
1.50×10^5	1.55×10^5	1.16×10^5	1.11×10^5	1.15×10^5
1.37×10^5	1.46×10^5	1.22×10^5	1.11×10^5	1.16×10^5

C) Sum of the expression values for experiments of [33]. Each experiment has three replicates

Meristem_root_nodD	Distal_infection_zone_root_nodD	Proximal_infection_zone_root_nodD	Infected_root_nodD	Uninfected_root_nodD
1.52×10^5	1.49×10^5	1.51×10^5	1.49×10^5	1.53×10^5
1.51×10^5	1.50×10^5	1.50×10^5	1.47×10^5	1.51×10^5
1.50×10^5	1.52×10^5	1.47×10^5	1.42×10^5	1.52×10^5

Table S6: List of experiments in the MtGEA database with one single replicate

Hyptl_A17_10C_35C_day1_1	Hyptl_F83_10C_100C_day1_1	Hyptl_A17_20C_100C_day1_1
Hyptl_A17_10C_35C_day2_1	Hyptl_F83_10C_100C_day2_1	Hyptl_A17_20C_100C_day2_1
Hyptl_F83_10C_35C_day1_1	Hyptl_A17_20C_35C_day1_1	Hyptl_F83_20C_100C_day1_1
Hyptl_F83_10C_35C_day2_1	Hyptl_A17_20C_35C_day2_1	Hyptl_F83_20C_100C_day2_1
Hyptl_A17_10C_50C_day1_1	Hyptl_F83_20C_35C_day1_1	RT_2wks_Sdl_Hydroponic_200mM_NaCl_0h_1
Hyptl_A17_10C_50C_day2_1	Hyptl_F83_20C_35C_day2_1	RT_2wks_Sdl_Hydroponic_200mM_NaCl_1h_1
Hyptl_F83_10C_50C_day1_1	Hyptl_A17_20C_50C_day1_1	RT_2wks_Sdl_Hydroponic_200mM_NaCl_2h_1
Hyptl_F83_10C_50C_day2_1	Hyptl_A17_20C_50C_day2_1	RT_2wks_Sdl_Hydroponic_200mM_NaCl_5h_1
Hyptl_A17_10C_100C_day1_1	Hyptl_F83_20C_50C_day1_1	RT_2wks_Sdl_Hydroponic_200mM_NaCl_10h_1
Hyptl_A17_10C_100C_day2_1	Hyptl_F83_20C_50C_day2_1	RT_2wks_Sdl_Hydroponic_200mM_NaCl_24h_1

Experiment	Pair 1-2	Pair 2-3	Pair 1-3
RT_Myc_3wks_infection	0.78	0.87	0.74
GiantCell	0.77	0.81	0.79
GallTissue_GiantCell	0.82	0.82	0.78
RT_LCM_arbuscular[1]	0.89	/	/
RT_LCM_cortical[1]	0.89	/	/
RT_LCM_adjacent[1]	0.87	/	/
Nod_Naut1_SalsC[2]	0.84	0.89	0.94
RT_CRR_72hpi[3]	0.99	0.91	0.89
RT_CRR_96hpi[3]	0.96	0.85	0.82
Root_A17_control[4]	0.99	0.84	0.85
HairyRoot_WT_Myc_CK[4]	0.99	0.84	0.85

Table S7: Groups of experiments showing Pearson correlation coefficients under a threshold value of 0.90. In red acceptable Pearson correlation coefficients. Values are approximated to the second decimal. We used a slash "/" to indicate that the experiment has only 2 replicates, in which case only one single pair of Pearson correlation coefficient could be calculated. Acceptable pairs are highlighted in red.
References: [1][31], [2][56], [3][57], [4][34]

Figure S1: Graphical representation of the sum of expression values for selected groups of experiments compared to the neighboring ones in the dataset list. The great majority of experiments have values around 2 x 107 while a few groups show large reduction in the values. (A) RT_Myc_3wks_infection [Balzergue et al., unpublished], with sum values under 1.5 x 10^7; (B) Group of GiantCell, GallTissue-GiantCell and Nod_zone2 experiments [Balzergue et al., unpublished], with sum values in the 1.1-1.5 x 10^7 range; (C) Group of X1, X3, X5_dpi and 24h experiments [Breakspear et al., (unpublished)] with sum values in the 1.25-1.70 x 10^7 range; (D) Group of Nod_Naut1_SalsC, Nod_Naut1_SalsB, Nod_Sals4_SalsB and Nod_Sals4_SalsC experiments [62], with sum values in the 0.5-1.0 x 10^7 range.

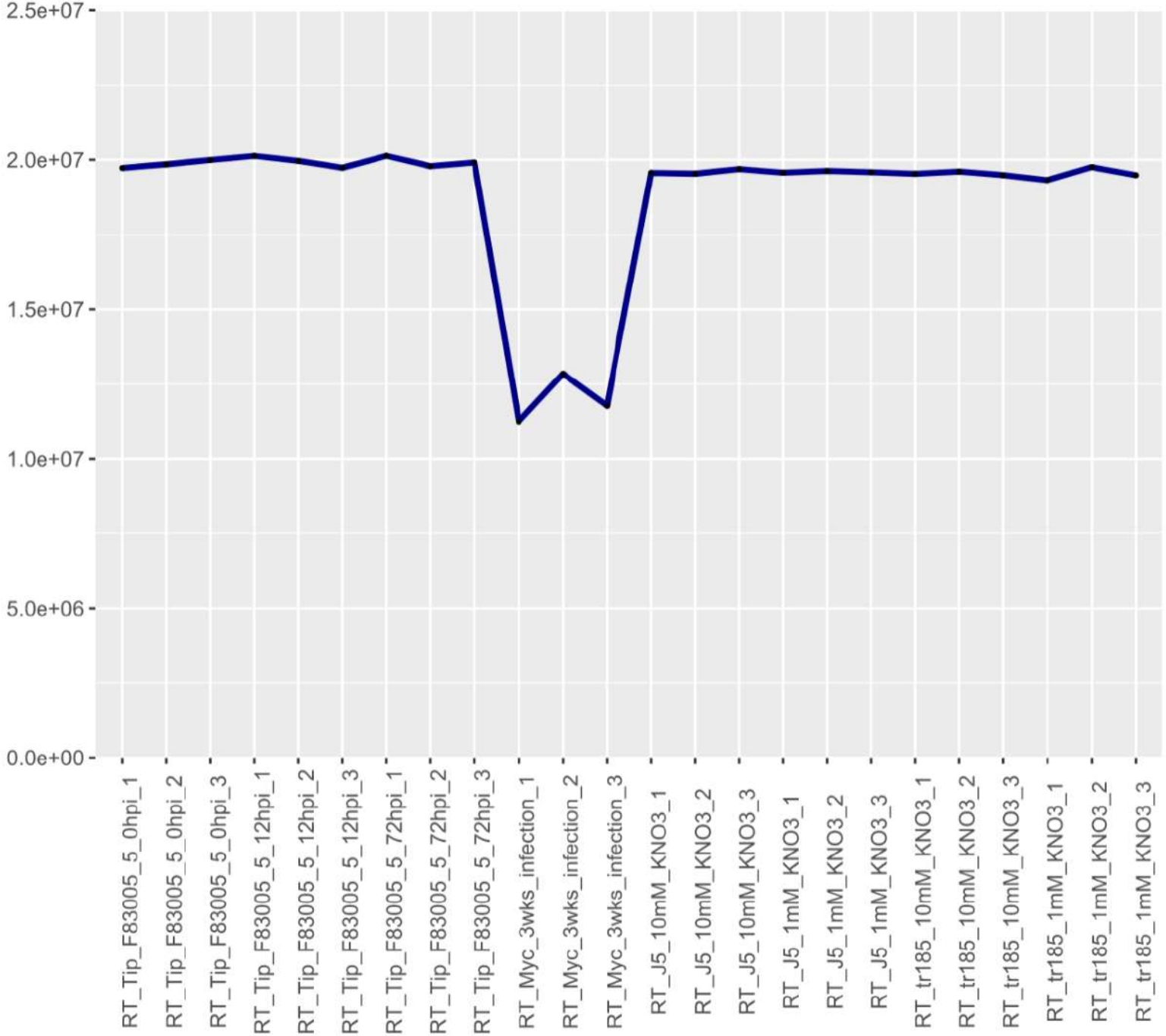

2.5e+07
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
RT_Tip_F83005_5_0hpi_1
RT_Tip_F83005_5_0hpi_2
RT_Tip_F83005_5_0hpi_3
RT_Tip_F83005_5_12hpi_1
RT_Tip_F83005_5_12hpi_2
RT_Tip_F83005_5_12hpi_3
RT_Tip_F83005_5_72hpi_1
RT_Tip_F83005_5_72hpi_2
RT_Tip_F83005_5_72hpi_3
RT_Myc_3wks_infection_1
RT_Myc_3wks_infection_2
RT_Myc_3wks_infection_3
RT_J5_10mM_KNO3_1
RT_J5_10mM_KNO3_2
RT_J5_10mM_KNO3_3
RT_J5_1mM_KNO3_1
RT_J5_1mM_KNO3_2
RT_J5_1mM_KNO3_3
RT_tr185_10mM_KNO3_1
RT_tr185_10mM_KNO3_2
RT_tr185_10mM_KNO3_3
RT_tr185_1mM_KNO3_1
RT_tr185_1mM_KNO3_2
RT_tr185_1mM_KNO3_3

Figure S1 A

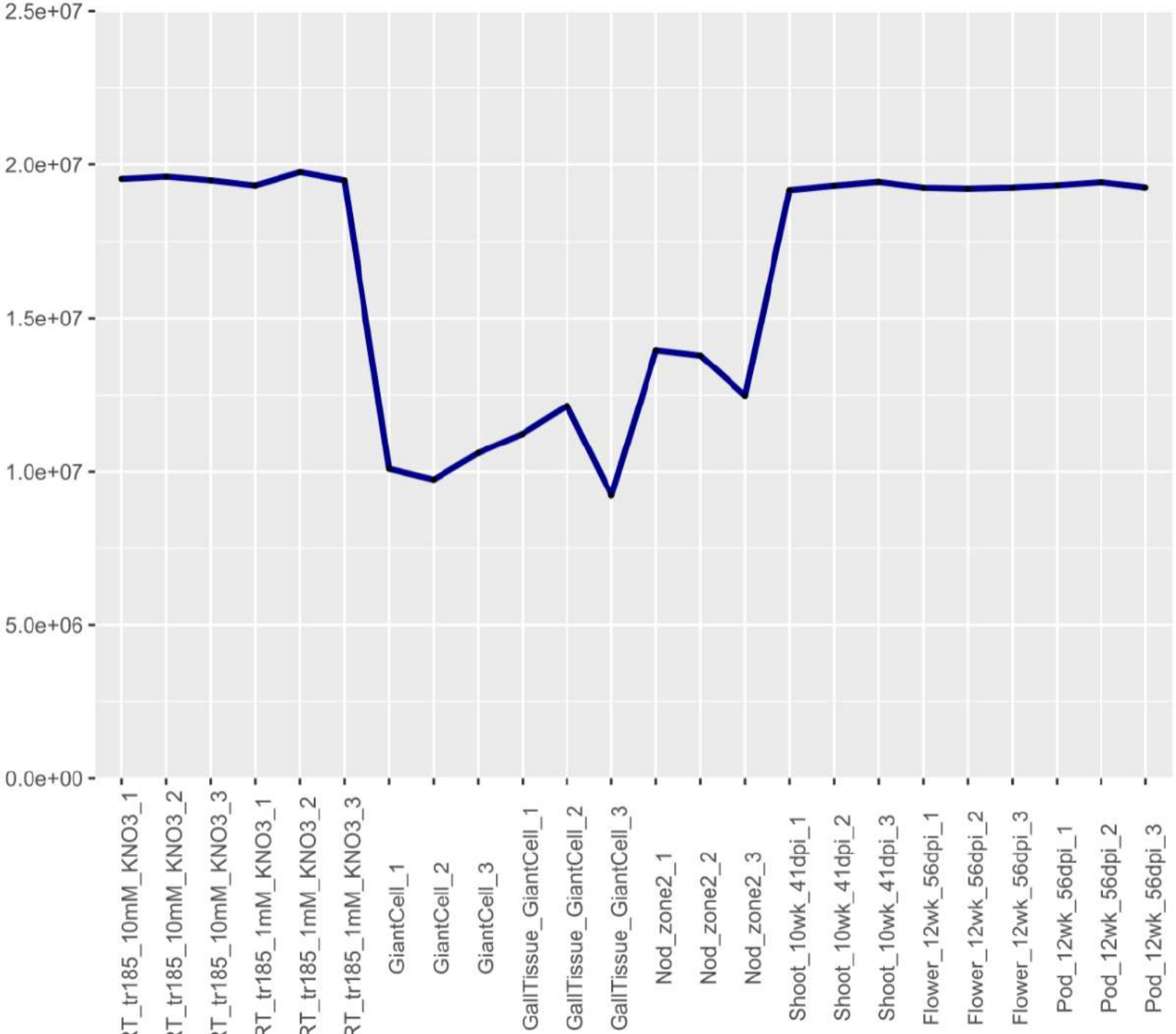

2.5e+07
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
RT_tr185_10mM_KNO3_1
RT_tr185_10mM_KNO3_2
RT_tr185_10mM_KNO3_3
RT_tr185_1mM_KNO3_1
RT_tr185_1mM_KNO3_2
RT_tr185_1mM_KNO3_3
GiantCell_1
GiantCell_2
GiantCell_3
GallTissue_GiantCell_1
GallTissue_GiantCell_2
GallTissue_GiantCell_3
Nod_zone2_1
Nod_zone2_2
Nod_zone2_3
Shoot_10wk_41dpi_1
Shoot_10wk_41dpi_2
Shoot_10wk_41dpi_3
Flower_12wk_56dpi_1
Flower_12wk_56dpi_2
Flower_12wk_56dpi_3
Pod_12wk_56dpi_1
Pod_12wk_56dpi_2
Pod_12wk_56dpi_3

Figure S1 B

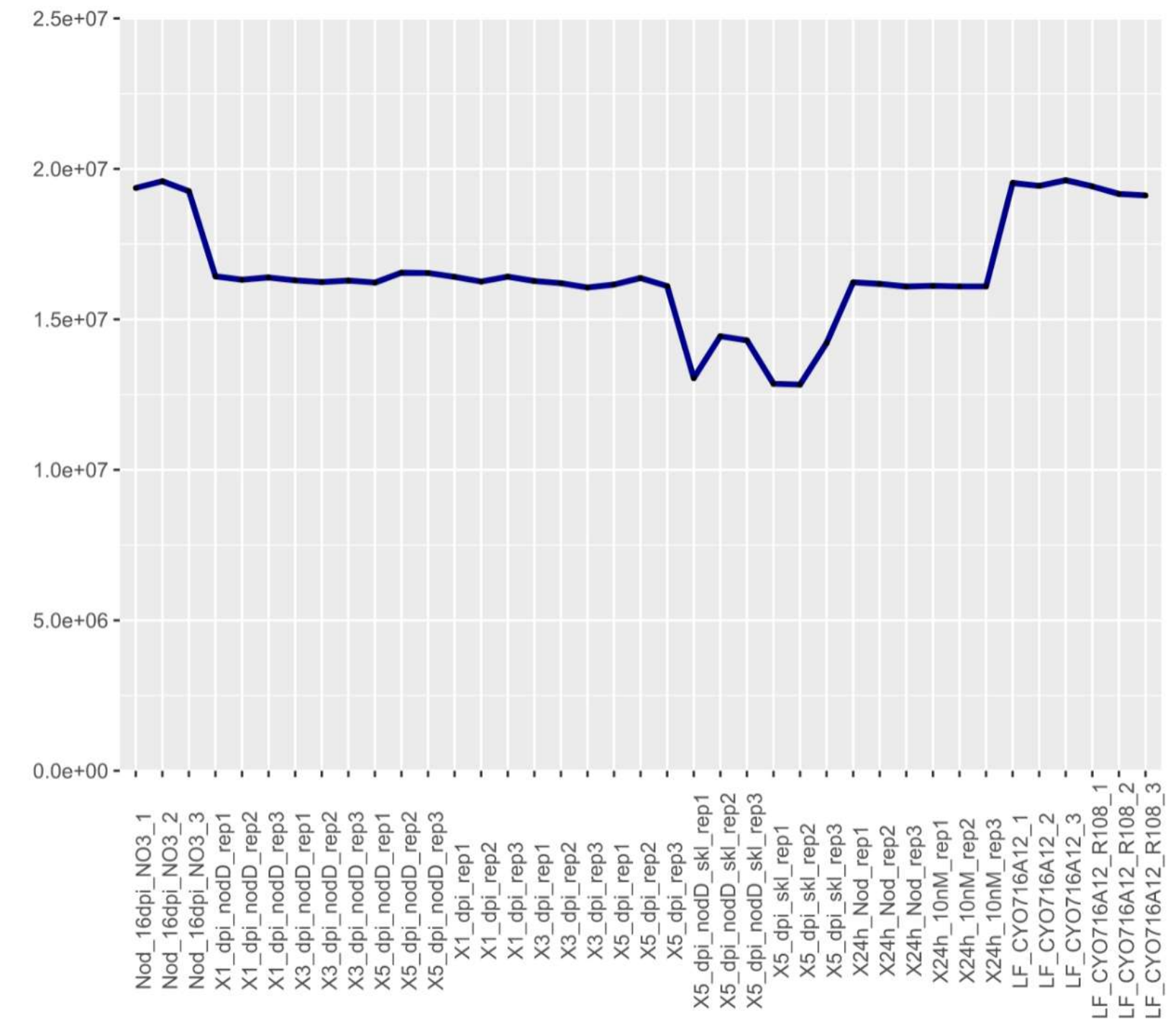

2.5e+07
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
Nod_16dpi_NO3_1
Nod_16dpi_NO3_2
Nod_16dpi_NO3_3
X1_dpi_nodD_rep1
X1_dpi_nodD_rep2
X1_dpi_nodD_rep3
X3_dpi_nodD_rep1
X3_dpi_nodD_rep2
X3_dpi_nodD_rep3
X5_dpi_nodD_rep1
X5_dpi_nodD_rep2
X5_dpi_nodD_rep3
X1_dpi_rep1
X1_dpi_rep2
X1_dpi_rep3
X3_dpi_rep1
X3_dpi_rep2
X3_dpi_rep3
X5_dpi_rep1
X5_dpi_rep2
X5_dpi_rep3
X5_dpi_nodD_skl_rep1
X5_dpi_nodD_skl_rep2
X5_dpi_nodD_skl_rep3
X5_dpi_skl_rep1
X5_dpi_skl_rep2
X5_dpi_skl_rep3
X24h_Nod_rep1
X24h_Nod_rep2
X24h_Nod_rep3
X24h_10nM_rep1
X24h_10nM_rep2
X24h_10nM_rep3
LF_CYO716A12_1
LF_CYO716A12_2
LF_CYO716A12_3
LF_CYO716A12_R108_1
LF_CYO716A12_R108_2
LF_CYO716A12_R108_3

Figure S1 C

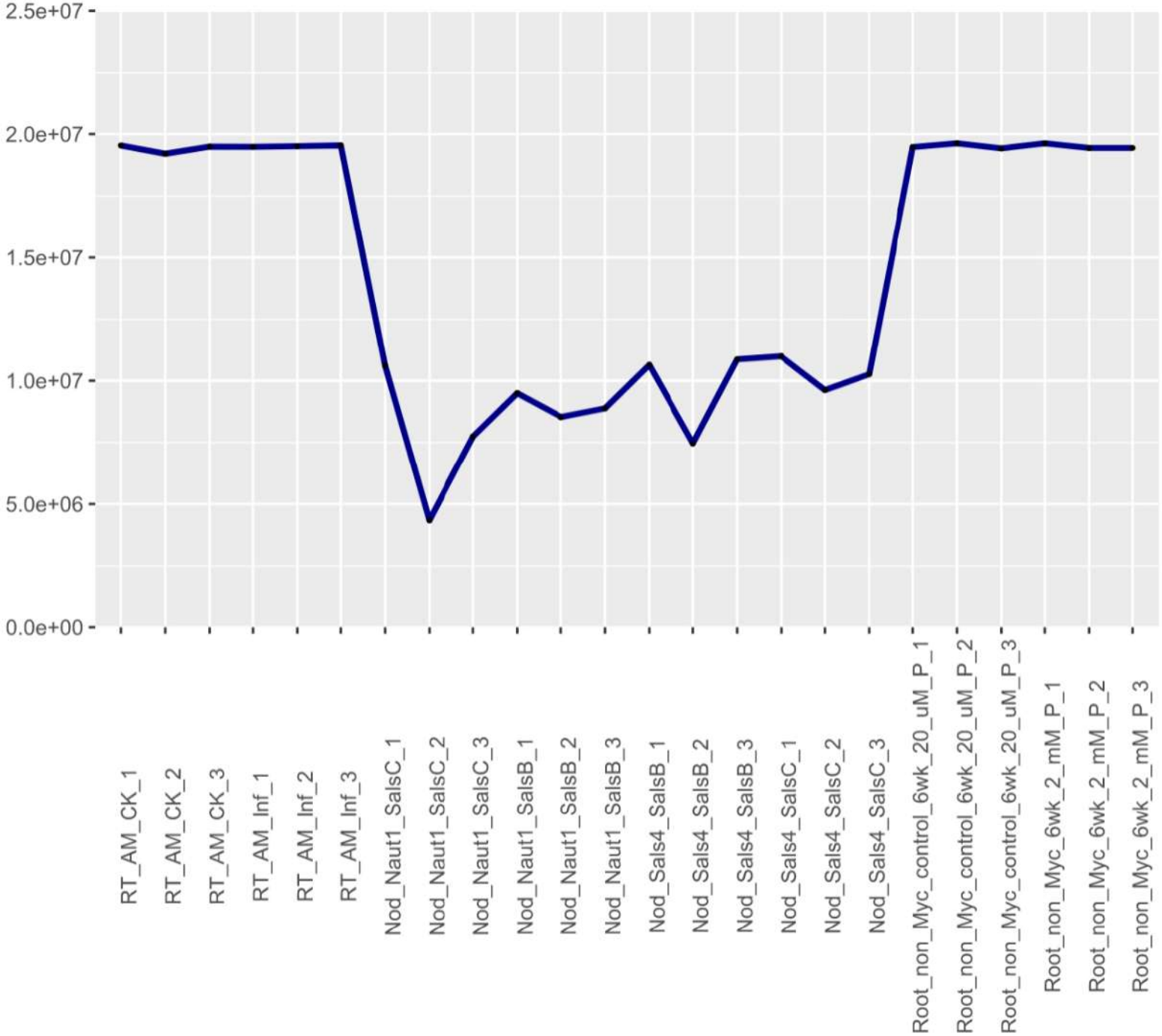
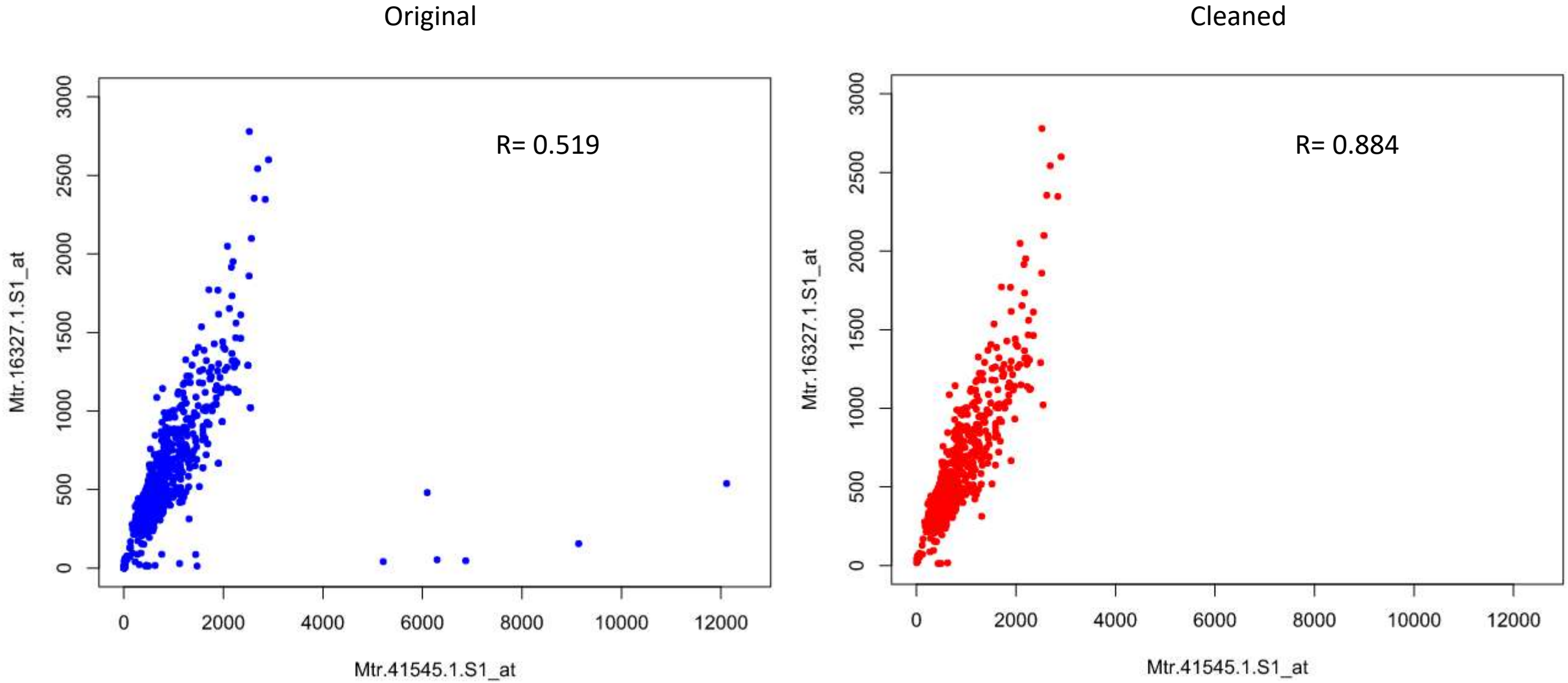

2.5e+07
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
RT_AM_CK_1
RT_AM_CK_2
RT_AM_CK_3
RT_AM_Inf_1
RT_AM_Inf_2
RT_AM_Inf_3
Nod_Naut1_SalsC_1
Nod_Naut1_SalsC_2
Nod_Naut1_SalsC_3
Nod_Naut1_SalsB_1
Nod_Naut1_SalsB_2
Nod_Naut1_SalsB_3
Nod_Sals4_SalsB_1
Nod_Sals4_SalsB_2
Nod_Sals4_SalsB_3
Nod_Sals4_SalsC_1
Nod_Sals4_SalsC_2
Nod_Sals4_SalsC_3
Root_non_Myc_control_6wk_20_uM_P_1
Root_non_Myc_control_6wk_20_uM_P_2
Root_non_Myc_control_6wk_20_uM_P_3
Root_non_Myc_6wk_2_mM_P_1
Root_non_Myc_6wk_2_mM_P_2
Root_non_Myc_6wk_2_mM_P_3

Figure S1 D

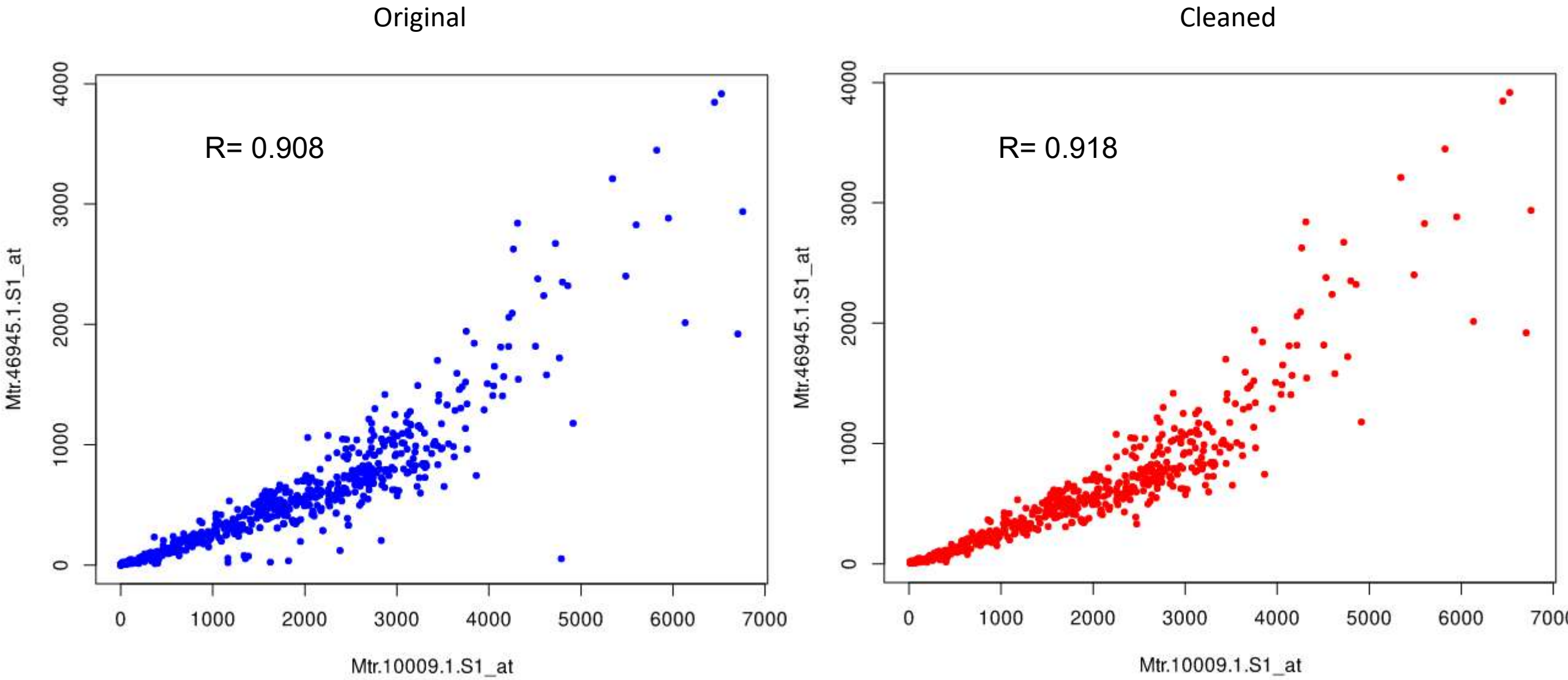
Figure S2 A-F. Scatterplot of the expression levels of different probe pairs hybridizing to the same gene, before (light blue) and after (red) the cleaning procedure.

Figure S2 A



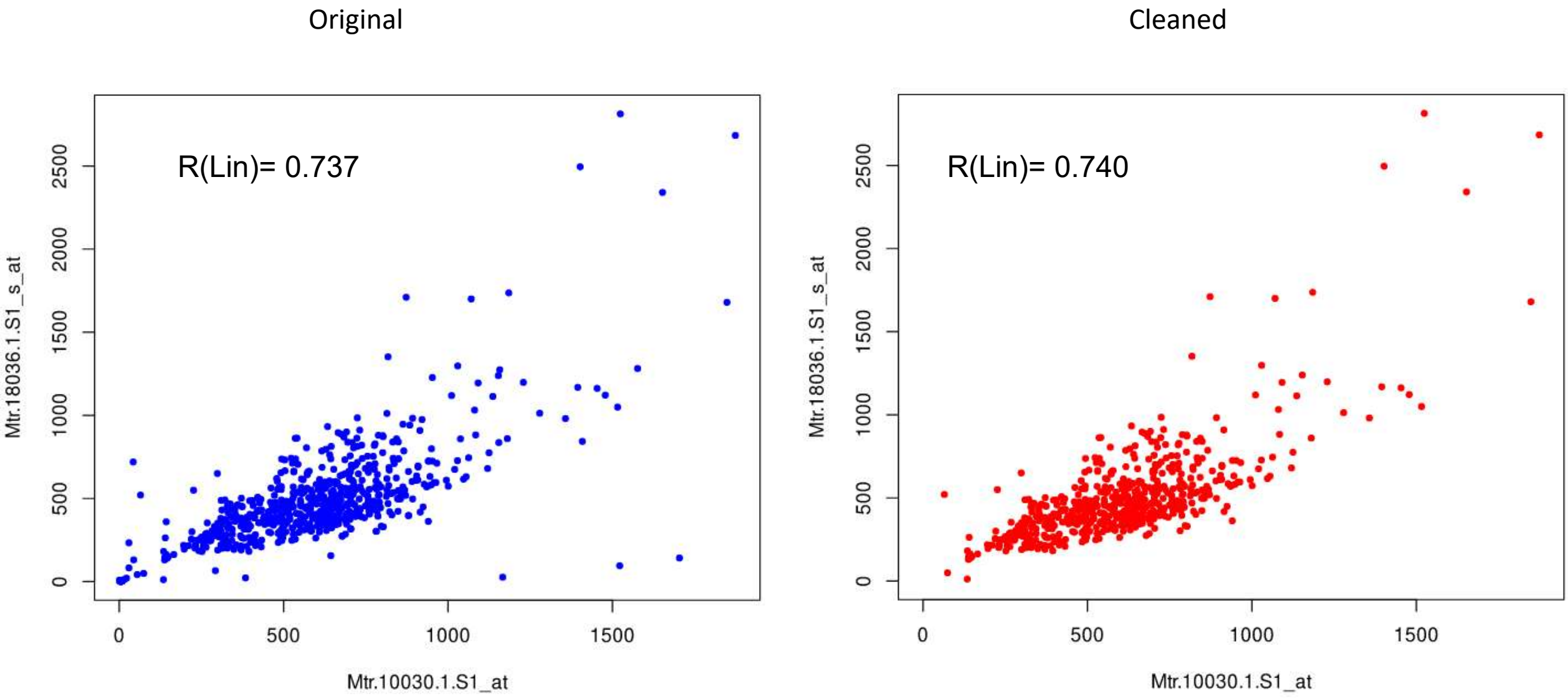
Mtr_7g113660 Mevalonate kinase

Figure S2 B



Mtr_3g072350 WEB family plant protein

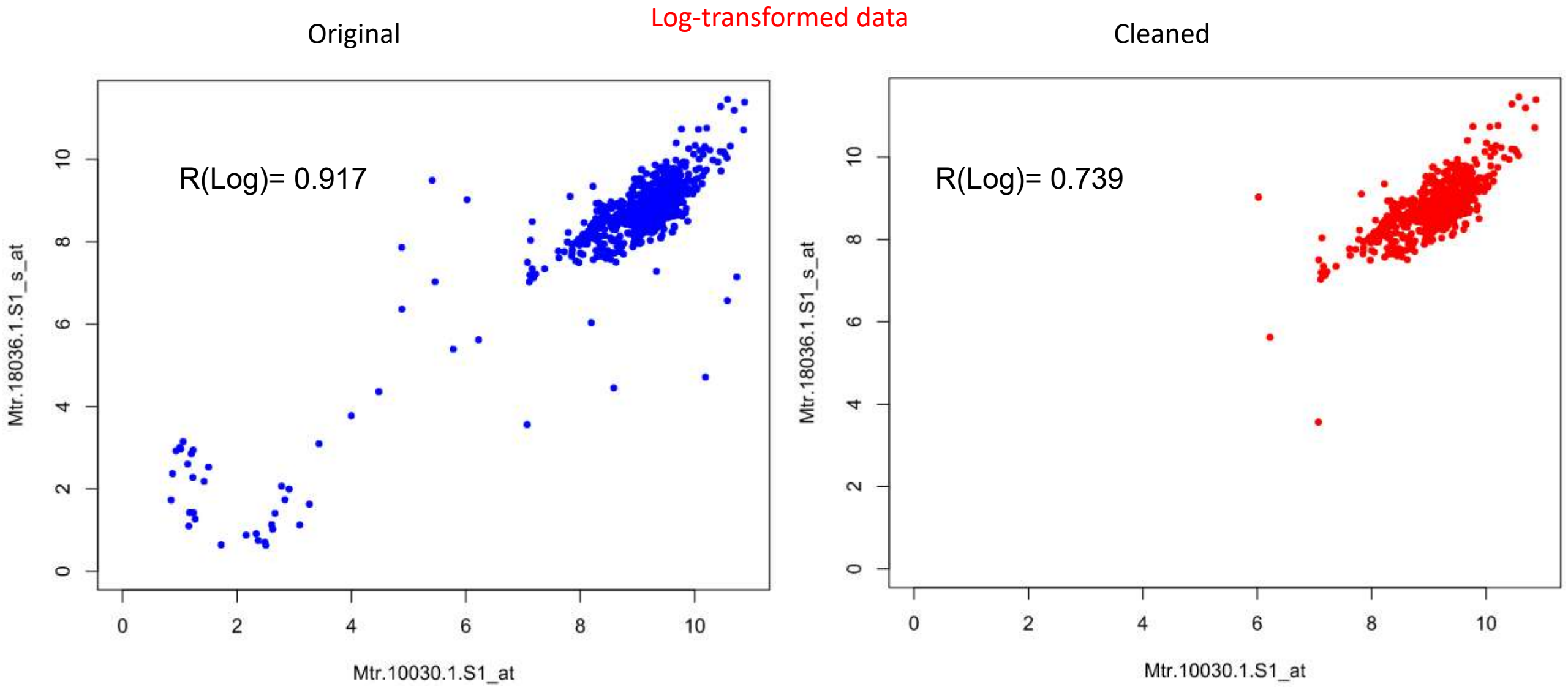
Figure S2 C



Mtr_4g116460 zinc finger, C3HC4 type (RING finger) protein

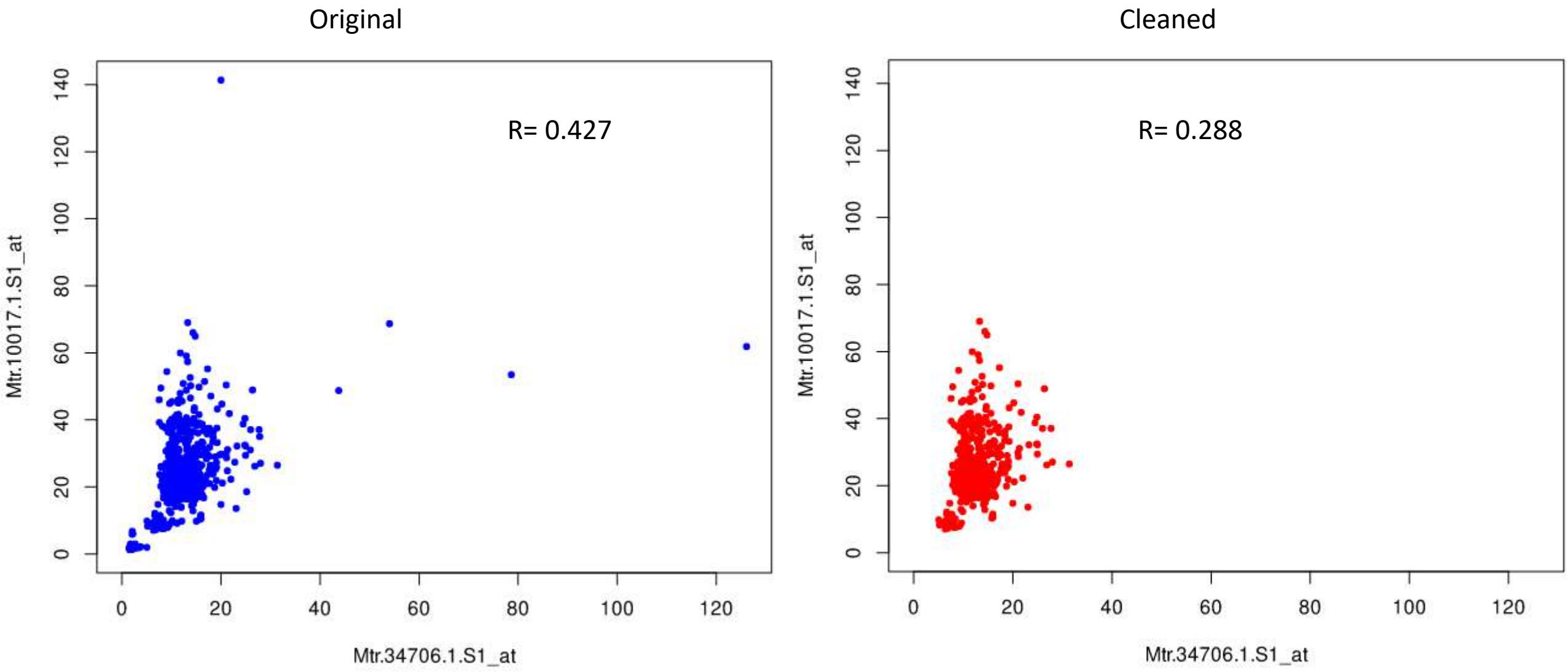
Figure S2 D

Log-transformed data



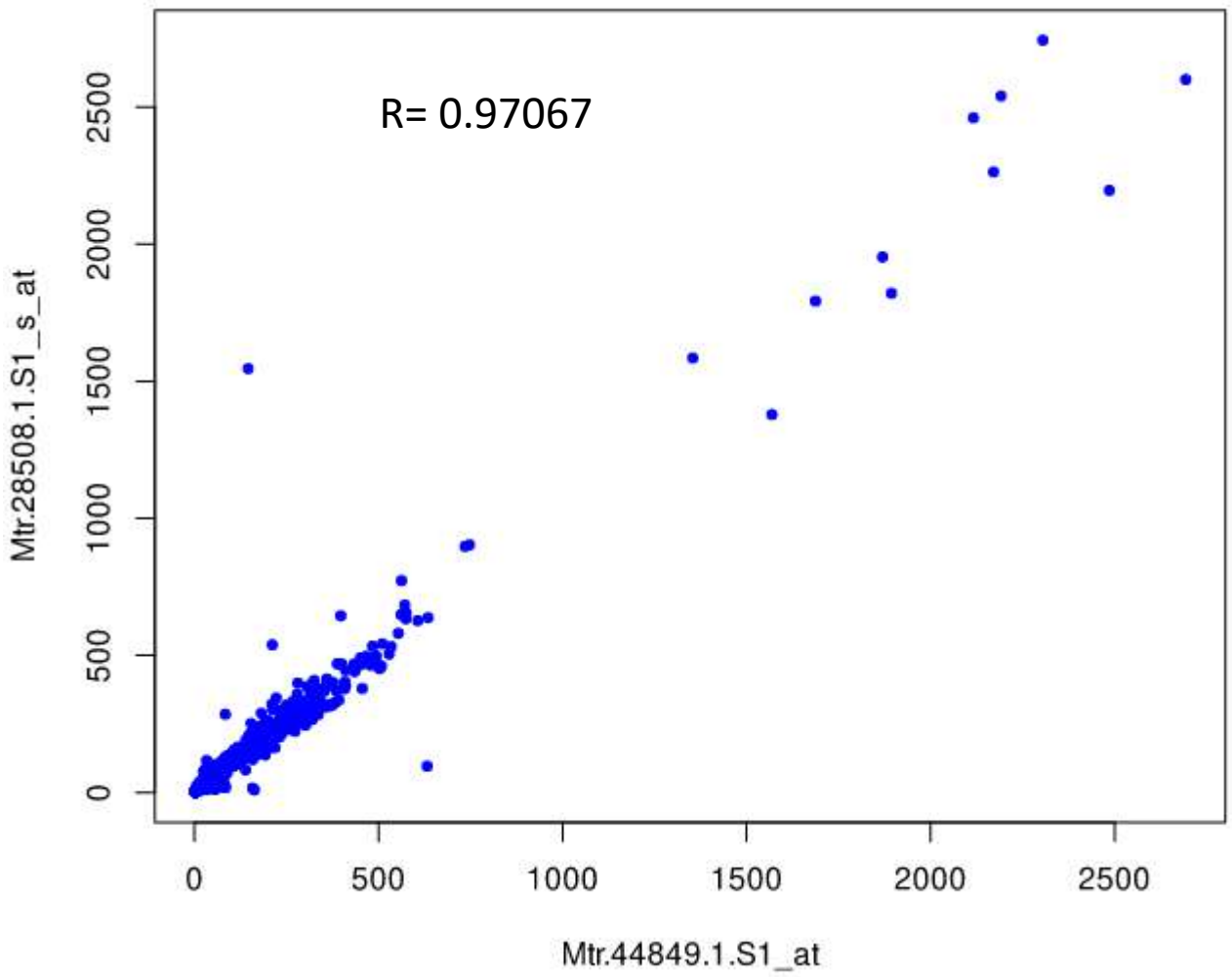
Mtr_4g116460 zinc finger, C3HC4 type (RING finger) protein

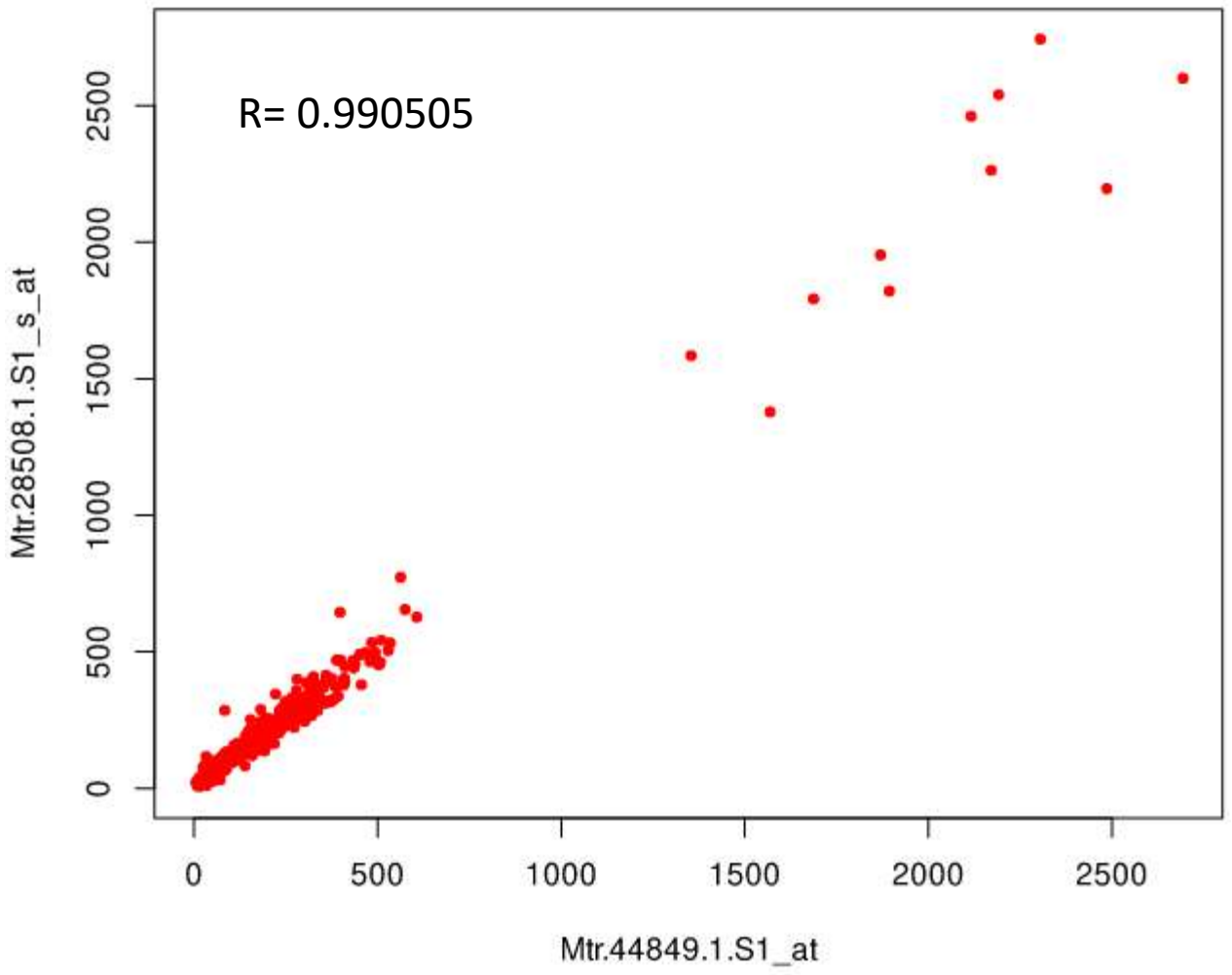
Figure S2 E



MTR_6g029470 Galactose-1-phosphate uridylyltransferase

Figure S2 F





MTR_1g029400 transcriptional corepressor SEUSS-like protein

Manuscript 2

NORMALIX95: a shiny-based application for plant microarray analysis

Francesca Marzorati[1*], Luca Mizzi[2], Irene Murgia[1], Piero Morandini[1]

[1] Dept. Environmental Science and Policy, University of Milano, Milano, Italy
[2] Dept. Biosciences, University of Milano, Milano, Italy
[*] Correspondence: francesca.marzorati1@unimi.it

Abstract

Microarray and RNA sequencing technologies have been adopted for large-scale gene expression analysis in various organisms under different experimental conditions. However, researchers without programming skills cannot easily approach and utilize the overwhelming amount of transcriptomic data available online, particularly for plant species. Therefore, we developed NORMALIX95, a tool implemented in R using the *shiny* package to facilitate the analysis of plant microarray data. It consists of 13 tabs that allow users to perform different operations, such as normalisation (RMA, GCRMA, and MAS5) of *Affymetrix* microarray data for 11 plant species, generation of plots (heatmaps for Pearson correlation analysis, scatter plots for genes and samples of uploaded datasets), correlation studies (Pearson, Spearman, and Kendall between hybridizations or genes), and differential gene expression analysis. Users can launch the app on any operating system with installed R. NORMALIX95 is the first tool specifically designed for plant microarray analysis and was developed with a user-friendly web interface to help researchers analyse plant gene expression data.

Keywords: Transcriptomics; Microarray; *Affymetrix*; R programming; Correlation analysis

1. Introduction

The investigation of gene function and expression on a large scale, starting in the late 1990s, has represented an important achievement in biology thanks to microarrays and RNA sequencing (RNA-Seq) technologies, allowing the development of the so-called 'transcriptomics', which is the gathering and study of gene expression data from different organisms under different conditions on a massive scale (Lowe et al., 2017). Briefly, in microarray studies, messenger RNA (mRNA) is isolated from a biological sample of interest and transformed into complementary DNA (cDNA). These cDNAs are fluorescently labeled and loaded onto microarrays, where thousands of single-stranded DNA (ssDNA) samples, each corresponding to a given gene, are organized in an orderly manner as spots in a grid formation. Each cDNA can only pair with its corresponding ssDNA; when such pairing occurs, it indicates that the corresponding gene is expressed. A fluorescence level analysis using a laser is then performed to detect the fluorescent label of paired cDNAs, highlighting the location of the expressed genes, and the intensities of the fluorescent signals correspond to numeric values, which provide an estimation of the expression of the analysed sequences (Bumgarner, 2013; Jaksik et al., 2015; Lemieux et al., 1998; Quackenbush, 2002; Zhao and Bruce, 2003). *Affymetrix* (https://www.thermofisher.com/it/en/home/life-science/microarray-analysis.html) has generated a wide catalogue of microarrays for gene expression analysis, genotyping, and sequencing (Bumgarner, 2013; Chee et al., 1996; Dalma-Weiszhausz et al., 2006; Lockhart et al., 1996; Noble, 1995). However, *Affymetrix* technology slightly varies from the microarray strategy described here above, with the major difference being that the production is based on photochemical synthesis and cRNA instead of cDNA (Dalma-Weiszhausz et al., 2006). These changes in *Affymetrix* microarray production allow avoiding errors caused by the preparation of many cDNAs and the use of many probes with different sequences that can hybridize to different regions of the same reference sequence (the so-called 'probe redundancy'). Expression data can be used for several investigations, such as transcript correlation analysis with a 'guilt-by-association' approach, which is particularly useful for identifying novel genes involved in specific processes (Altshuler et al., 2000). RNA-Seq seems to overtake array technologies nowadays, and indeed, the 'death' of microarrays was predicted in 2008 (Ledford, 2008). Nevertheless, microarrays are still widely used in several research areas and many laboratories have extensive experience with this technology, which is an inexpensive and well-established method for pilot studies and/or studies with a large number of samples (Eijssen et al., 2013).

Many microarray data are thus available online in repositories, such as Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo) (Barrett et al., 2013; Edgar, 2002) but some processing is necessary: normalisation procedures are required for appropriate use and interpretation of microarray data, and they are commonly organized into 'adjustment', 'normalisation', and 'summarization' steps (Affymetrix, 2002; Gautier et al., 2004; Quackenbush, 2002; Wu, 2009). MAS5 (Microarray Analysis Suite version 5), RMA (Robust Multiarray Analysis), and GCRMA (Guanine Cytosine Robust Multiarray Analysis) are common normalisation methods for microarrays, with MAS5 providing normalised data in a linear format and RMA and GCRMA in logarithm format (Harr and Schlötterer, 2006; Irizarry et al., 2003; Lim et al., 2007; Wu, 2009). One issue with choosing a normalisation strategy is that there is no 'reference method' to compare the expression values; several studies have attempted to compare normalisation procedures to highlight the best, but the results are not always in accordance (Bolstad et al., 2003; Harr and Schlötterer, 2006; Irizarry et al., 2003; Millenaar et al., 2006). The normalisation method used can indeed significantly affect the outcomes of the analyses and should be selected according to the aim of the expression study (Harr and Schlötterer, 2006). Obtaining, organising, and verifying transcriptome data usually requires a combination of tools; researchers use different applications simultaneously, yielding results and plots that are not always comparable. It is possible to find normalised data in repositories such as GEO, though with some limitations, as not all the uploaded experiments are provided with normalised data; because of that, researchers have to download 'raw' data and perform normalisation with external software only. Moreover, even when normalised data are uploaded to the repository, they are typically available in a file format that requires, to be accessed, computational skills, and/or they are included within a table that cannot be easily downloaded. Finally, the uploaded normalised data could have been obtained using different normalisation procedures, preventing easy comparison and merging of data from different experiments. Therefore, the development of practical and easily accessible tools for transcriptomic studies is fundamental for scientists without bioinformatics skills. Various commercial companies have recently developed software specifically designed to perform all steps of microarray and RNA-Seq analysis; nevertheless, in addition to their cost, these tools are often not updated and not always easy to use (Choi and Ratner, 2019; Eijssen et al., 2013). Free open-source programs and packages for transcriptomic analysis are also available online; however, their use still requires programming competencies as they are often devoid of interactive, user-friendly interfaces (Choi and Ratner, 2019; Reyes et al.,

2019). An example is the R/Bioconductor project (https://www.bioconductor.org/), which offers open-source R packages for computational biology and bioinformatics, thus supporting high-throughput genomic analyses (Gentleman et al., 2004; Huber et al., 2015). However, the full exploitation of the various resources freely available in Bioconductor still requires some computational skills, which could be limiting for many biologists. In the present study, we describe a new tool called NORMALIX95 to normalise, process, and analyse *Affymetrix* data for 11 plant species (*Vitis vinifera* - grape, *Oryza* sp. - rice, *Zea mays* - corn, *Triticum* sp. - wheat, *Solanum lycopersicum* - tomato, *Hordeum vulgare* - barley, *Arabidopsis thaliana* - Arabidopsis, *Populus* sp. - poplar, *Medicago truncatula* - Medicago, *Glycine max* - soybean, and *Saccharum officinarum* - sugar cane). This application was specifically designed for researchers with little or no computational skills, allowing the study of massive amounts of plant transcriptome data available online using a single tool.

2. Materials and Methods

2.1 NORMALIX95 development

NORMALIX95 is an application written in the open-source programming language R (https://www.R-project.org/, version 4.1.0) using the *shiny* package (https://shiny.rstudio.com/) and the *shinydashaboard* package (https://rstudio.github.io/shinydashboard/) (Chang et al., 2016; Chang and Ribeiro, 2018). All the pages were accessed on April 3, 2022.

2.2 R packages included in NORMALIX95

NORMALIX95 was implemented using several R packages that are also included in the Bioconductor repository (Gentleman et al., 2004; Huber et al., 2015):
Biobase (https://bioconductor.riken.jp/packages/3.9/bioc/html/Biobase.html) (Gentleman et al., 2015), *oligoClasses* (https://bioconductor.riken.jp/packages/3.1/bioc/html/oligoClasses.html) (Carvalho and Scharpf, 2015), *data.table* (https://cran.r-project.org/web/packages/data.table/index.html)

(Dowle and Srinivasan, 2021), *DT* (https://cran.r-project.org/web/packages/DT/index.html) (Xie et al., 2021), *affy* (https://www.bioconductor.org/packages/release/bioc/html/affy.html) (Gautier et al., 2004), *limma* (https://bioconductor.org/packages/release/bioc/html/limma.html) (Ritchie et al., 2015; Smyth, 2021), *MVA* (https://cran.r-project.org/web/packages/MVA/index.html) (Everitt and Hothorn, 2021), *gcrma* (https://www.bioconductor.org/packages/release/bioc/html/gcrma.html) (Wu and Irizarry, 2014), *heatmaply* (https://cran.r-project.org/web/packages/heatmaply/index.html) (Galili et al., 2021), *ggplot2* (https://cran.r-project.org/web/packages/ggplot2/index.html) (Wickham et al., 2021a), *plotly* (https://cran.r-project.org/web/packages/plotly/index.html) (Carson et al., 2021), *vitisviniferacdf* (https://www.bioconductor.org/packages//2.12/data/annotation/html/vitisviniferacdf.html), *maizecdf* (https://bioconductor.org/packages/release/data/annotation/html/maizecdf.html), *wheatcdf* (http://bioconductor.org/packages/release/data/annotation/html/wheatcdf.html), *ricecdf* (http://bioconductor.org/packages/release/data/annotation/html/ricecdf.html), *tomatocdf* (http://bioconductor.org/packages/release/data/annotation/html/tomatocdf.html), *barley1cdf* (https://bioconductor.org/packages/release/data/annotation/html/barley1cdf.html), *ath1121501cdf* (https://bioconductor.org/packages/release/data/annotation/html/ath1121501cdf.html), *poplarcdf* (http://bioconductor.org/packages/release/data/annotation/html/poplarcdf.html), *medicagocdf* (https://bioconductor.org/packages/release/data/annotation/html/medicagocdf.html), *soybeancdf* (https://bioconductor.org/packages/release/data/annotation/html/soybeancdf.html), *sugarcanecdf* (https://bioconductor.org/packages/release/data/annotation/html/sugarcanecdf.html), *dplyr* (https://cran.r-project.org/web/packages/dplyr/index.html) (Wickham et al., 2021b), *tidyr* (https://cran.r-project.org/web/packages/tidyr/index.html) (Wickham, 2021), *stringr* (https://cran.r-project.org/web/packages/stringr/index.html) (Wickham, 2019), *matrixStats* (https://cran.rstudio.com/web/packages/matrixStats/index.html) (Bengtsson et al., 2021), *genefilter* (https://bioconductor.org/packages/release/bioc/html/genefilter.html) (Gentleman et al., 2021). All the pages were accessed on April 3, 2022.

2.3 Running NORMALIX95

NORMALIX95 can be launched locally on every computer with installed R. It was mostly tested on Windows 10 and Linux (Ubuntu 18.04) operating systems.

2.4 Microarray data

Array data used to test NORMALIX95 were retrieved from the public functional genomics data repository Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo) (Barrett et al., 2013; Edgar, 2002) and from the *Medicago truncatula* Gene Expression Atlas (MtGEA) (Benedito et al., 2008; He et al., 2009) (https://mtgea.noble.org/v3/, accessed on May 13, 2021 as described in (Marzorati et al., 2021)). GEO accession numbers are as follows:
GSE132311 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132311) (Rondot and Reineke, 2019), GSE29948 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29948) (Tillett et al., 2012), GSE147683 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147683) [*unpublished,* contributor: Gu Y], GSE29027 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29027) (Heath et al., 2012), GSE150581 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150581) (Çakır Aydemir et al., 2020), GSE41423 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41423) (Liu et al., 2012). All GEO pages were accessed on April 3, 2022.

3. Results

3.1 NORMALIX95 structural organization

The general architecture of NORMALIX95 is graphically presented in **Figure 1**: NORMALIX95 is organized into 13 tabs, each dedicated to a specific activity, 10 of which are tools for transcriptomic analysis. The *Getting started* introductory page is divided into two subsections: an *Introduction* page describing NORMALIX95 (what users can do with it and the aim of such an application), whereas the *Contact references* page shows the list of the authors' contacts. In the *Normalisation* tab, 11 subpages are available, each one specific for a plant species (*Vitis vinifera* - grape, *Oryza* sp. - rice, *Zea mays* - corn, *Triticum* sp. - wheat, *Solanum lycopersicum* - tomato, *Hordeum vulgare* - barley, *Arabidopsis thaliana* - Arabidopsis, *Populus* sp. - poplar, *Medicago truncatula* - Medicago, *Glycine max* - soybean and *Saccharum officinarum* - sugar cane). In each of these subpages, the normalisation procedure is organized into two sections: the *Upload* section, where users are requested to upload transcriptome data as *.tar* files obtained from the Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2013; Edgar, 2002), and a *Results* section, where normalised data, using the RMA, GCRMA, and MAS5 strategies, are shown and can be downloaded, together with boxplots of raw and RMA normalised data to visually verify the outcome of the RMA normalisation. An example of the different normalisation strategies adopted for *Vitis vinifera Affymetrix* data uploaded in GEO (updated to February 20, 2022) is shown in **Table S1**. To use the other NORMALIX95 functions, the input data must have a comma-separated value (*.csv*) format, with genes typically as rows and samples as columns; each tab has its upload panel at the top of the page, where users are asked to specify characteristics of their dataset (separator, quote, and if they want to round displayed values). In NORMALIX95, another type of normalisation is also available in the *Quantile normalisation* tab, where users can upload datasets generated by combining normalised experiments to make all data uniform through quantile normalisation, a well-founded statistical technique in gene expression analysis (Wu, 2009; Zhao et al., 2020). Boxplots before and after quantile normalisation are shown at the bottom of the page for the visual verification of the normalisation procedure. The *Check your sample* tab allows users to perform the dataset cleaning procedure developed in our lab working on the *Medicago truncatula* Gene Expression Atlas (MtGEA) (https://mtgea.noble.org/v3/, accessed on May 13, 2021, as described by Marzorati et al. (2021)), allowing users to

calculate the sum, mean, and standard deviation for all expression values for samples/genes in the uploaded dataset. A line plot showing the sum of the expression values of the samples is displayed, together with a table of logarithmic Pearson correlation values among hybridizations in which values below a selected threshold are highlighted. The *Transpose*, *Log*, and *Antilog* tabs contain utilities to change the structure and values of the uploaded dataset: the *Transpose* tab allows to invert rows and columns of the uploaded dataset, whereas the *Log* and *Antilog* tabs allow the conversion of values in logarithmic (Log_2) or linear formats. The *One vs All* tab allows users to upload a dataset, select a column of interest, and perform a correlation analysis (Pearson – linear and logarithmic -, Spearman, or Kendall) of the selected element against all the others of the dataset, according to a 'guilt-by-association' approach (Altshuler et al., 2000). A similar activity can be performed in the *Correlation table* tab, but the correlation is calculated between all elements (genes or hybridizations) in one list and all elements in another list; therefore, the result is a table of correlation values. The *Plots* tab allows users to generate two types of interactive graphics: users can present the results of Pearson correlation analysis of a list against a list through the *Heatmap* subtab (heatmaps for linear and logarithmic Pearson correlation values) whereas the similarity and dissimilarity for both genes and samples can be investigated through the *Scatter Plot* subtabs (Scatter Plot Genes and Scatter Plot Samples); users can customize the plots by choosing several parameter settings located at the top of the page in the uploading options panel. The *DGE* tab is dedicated to the investigation of differential gene expression (DGE) analysis for *Affymetrix* microarray data: in the *Upload* subpage, users are required to upload a file with specific characteristics ('Counts' to identify the expression values and 'Design' to identify the experimental conditions) and to select a specific parameter for the analysis ('Contrast', *i.e.*, the comparison of the first selected group against a second one, such as 'treated' against 'untreated'); in the *Limma-voom* subpage, users obtain the results of 'filtering' (the removal of any uninformative data), a summary of the DGE analysis, and a complete table of the results. The *Datasets* and the *Useful links* tabs contain helpful links for bioinformatics analysis. In the former, there is a link to a drive (https://drive.google.com/drive/folders/1B97je0OyPHXnVXOTlq3bXREHP1d4mPme?) containing datasets of different plant species explored so far with NORMALIX95 and videos showing NORMALIX95 functions (also available as supplementary material **Videos 1–10**). The latter provides links to useful bioinformatics resources available online; these additional tools, such as Galaxy (https://usegalaxy.org/) (Afgan et al., 2018; Goecks et al., 2010) and

g:Profiler (https://biit.cs.ut.ee/gprofiler/) (Raudvere et al., 2019), may help improve transcriptomic analysis performed using NORMALIX95.

3.2 NORMALIX95 tests

To validate NORMALIX95, we compared the results of a recently published study on *Vitis vinifera* (GEO accession: GSE132311, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?) (Rondot and Reineke, 2019) with those obtained using NORMALIX95 RMA normalisation utility. First, we retrieved the data *.tar* files from the GEO accession number page; the file was uploaded to NORMALIX95 following the steps *Normalisation* – Grape – *Upload*. Next, we moved to the *Results* section to retrieve the normalised results. Given our interest in the RMA normalisation results, we downloaded them and compared the data with those available online for each experimental sample; the $\log_2$ data uploaded in GEO correspond to the data obtained using our application (our data are shown in **Table S2 A**). The correctness of the RMA normalisation was also confirmed by the generation of two boxplots comparing the raw and normalised data for each sample (**Figure S1**). Other tests were performed on the datasets of *Vitis vinifera* (GEO accession: GSE29948, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?) (Tillett et al., 2012), *Arabidopsis thaliana* (GEO accession: GSE147683, acc=GSE147683) [unpublished, contributor: Gu Y] and *Medicago truncatula* (GEO accession: GSE29027, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?) (Heath et al., 2012): in all cases, results obtained with NORMALIX95 match the ones available in GEO, with possible small variations in decimal numbers (**Table S2 B, C, and D**). We also validated the GCRMA and MAS5 normalisation strategies using grape datasets as testers (GEO accession: GSE150581, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?, and GEO accession: GSE41423, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41423) (Çakır Aydemir et al., 2020; Liu et al., 2012): our results match the ones provided in GEO with only minor variations (**Table S2 E and F**). We decided to test the NORMALIX95 *Plots* tab by generating heatmaps and scatterplots for genes working on previously analysed datasets (Marzorati et al., 2021). **Figure 2** shows the plots we have obtained: both the heatmaps of $\log_2$ Pearson correlation analysis for a *Medicago truncatula* saponins dataset and the scatterplot for two mevalonate kinases' mean expression values selected from a dataset downloaded from *Medicago truncatula* Gene Expression Atlas (MtGEA) (Benedito et al.,

2008; He et al., 2009) (https://mtgea.noble.org/v3/, accessed on May 13, 2021, as described in Marzorati et al. (2021)) correspond to previously obtained and published plots (Marzorati et al., 2021).

4. Discussion

NORMALIX95 is a ready-to-use Shiny application that integrates several functions from the Bioconductor repository (Gentleman et al., 2004; Huber et al., 2015) into a user-friendly interface. The main advantage of NORMALIX95 is its simple procedure for uploading and normalising microarray data for 11 plant species, allowing users to easily upload raw data downloaded from the Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo) (Barrett et al., 2013; Edgar, 2002) and normalise them through three possible methods (RMA, GCRMA, and MAS5). Indeed, NORMALIX95 helps overcome the limitations of downloading and normalising data retrieved from online repositories such as GEO (see above in the Introduction for further details); using NORMALIX95, users can upload each experiment and download the normalised data according to a single normalisation strategy, facilitating data comparison. Furthermore, NORMALIX95 allows users to carry out quantile normalisation, a procedure mainly suggested after normalisation strategies that present information within the logarithm format, such as RMA and GCRMA, and intends to have all arrays with an equal empirical distribution of intensities (Wu, 2009 and references therein). It is important to note that with NORMALIX95, to date, users may only normalise plant *Affymetrix* data; however, we would like to underline the uniqueness of NORMALIX95 because, to the best of our knowledge, it is the only available shiny application specifically developed to work with plant microarray data. Several shiny applications have been developed in recent years to perform transcriptomic analysis; however, they focus only on RNA-Seq data, and not of plants (Nelson et al., 2017; Reyes et al., 2019). An interesting case is iGEAK (Choi and Ratner, 2019), which allows for the analysis of microarray data; however, it is limited to mammalian datasets. NORMALIX95 was designed to support plant scientists in transcriptomic analysis, and includes many tools in a single environment, allowing users to obtain information without any programming skills and/or using several tools. For instance, *Transpose*, *Log*, and *Antilog* tabs contain utilities that can be found in many spreadsheet software packages,

but that may be particularly difficult to perform on input data characterized by thousands of columns and rows, such as gene expression datasets. Another merit of NORMALIX95 is the possibility to clean datasets following a cleaning strategy we have developed on the *Medicago truncatula* Gene Expression Atlas (MtGEA) database (Benedito et al., 2008; He et al., 2009) (https://mtgea.noble.org/v3/, accessed on May 13th, 2021 as described in Marzorati et al. (2021)); by working in the *Clean your sample* tab, users may discover if uploaded datasets contain errors and poor quality data that could affect subsequent analyses, such as correlation studies. The possibility of performing correlation analysis with different approaches (Pearson, Spearman, and Kendall) and deciding whether to analyse the correlation of all the elements in an uploaded dataset (*Correlation table* tab) or focus the attention on one single element against all the others (*One vs all* tab) is an important feature of NORMALIX95. For much as we know, there are no available shiny applications able to perform these analyses, particularly the *One vs all* utility, even though correlations using the 'guilt-by-association' approach are particularly interesting to assess the strength of the relationship between variables and identify or confirm candidate genes that could be involved in metabolic pathways and processes (Abbruscato et al., 2012; Altshuler et al., 2000; Beekweelder et al., 2008; Berri et al., 2009; Månsson et al., 2004; Marzorati et al., 2021; Menges et al., 2008; Murgia et al., 2020, 2011; Naoumkina et al., 2010; Vandepoele et al., 2009; Zermiani et al., 2015). Moreover, the possibility of using rank correlation (Spearman or Kendall) is of interest because these approaches are less sensitive to outliers than Pearson's correlation. It is important to note that the correlation analysis utility can be found in several spreadsheet software packages; however, as mentioned, it could be memory- and time-consuming when working on huge datasets, and it is traditionally limited to Pearson correlation analysis. Because data visualisation is important in microarray analysis and to cope with the increasing data complexity in biology (Baehrecke et al., 2004; Pavlopoulos et al., 2015; Prasad and Ahson, 2006), we decided to provide users with the opportunity to study the relationships among the elements of uploaded datasets through two types of interactive plots: heatmaps for Pearson correlation analysis and scatterplots for genes and samples. Interactive plots are becoming increasingly popular, helping in the presentation of large gene expression datasets by zooming into specific sectors of the plot, being easily customizable, and avoiding unreadable labels because texts overlap (Khomtchouk et al., 2017). We would like to stress that the analysis, interpretation, and visualisation of transcriptomic data may require additional statistical and bioinformatics competencies available in tools other than NORMALIX95, as suggested in the *Useful links*

tab. Nevertheless, our app was designed to be easily expanded by the scientific community, encouraging the contribution of users with or without computational skills in future NORMALIX95 developments; our e-mails are the preferred way to convey issues, highlights, and suggestions. The development of NORMALIX95 is ongoing and further versions are expected to integrate other useful features for plant scientists. For instance, we hope to upgrade the *DGE* tab by introducing a multiple-contrast option and the possibility of studying RNA-seq data. We will make sure to provide the NORMALIX95 script on request for at least one year following the publication date.

Acknowledgments

FM thanks her father for suggesting the first application name.

Author Contributions

Conceptualization: FM

Methodology: FM., LM, PM

Tool development: FM

Writing: FM

Editing: FM, LM, IM, PM

The authors declare that this research was conducted in the absence of any potential conflicts of interest.

Supplementary material

Table S2. **(A)** RMA-normalized data for GSE132311 samples (*Vitis vinifera* – grape, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?) (Rondot and Reineke, 2019) obtained through the NORMALIX95 application. **(B)** RMA-normalized data for GSE29948 samples (Vitis vinifera – grape, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?) (Tillett et al., 2012) obtained through the NORMALIX95 application. **(C)** RMA-normalized data for GSE147683 samples (*Arabidopsis thaliana* - Arabidopsis, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?) [*unpublished*] obtained through NORMALIX95 app. **(D)** RMA normalized data for GSE29027 samples (*Medicago truncatula* – Medicago, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29027) (Heath et al., 2012) obtained through NORMALIX95 app. **(E)** GCRMA normalized data for GSE150581 samples (*Vitis vinifera* – grape, acc=GSE150581) (Çakır Aydemir et al., 2020) obtained through the NORMALIX95 application. **(F)** MAS5 normalized data for GSE41423 samples (Vitis vinifera – grape, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?) (Liu et al., 2012) obtained using the NORMALIX95 app.

Videos S1–S10. **Videos showing NORMALIX95 functionalities**. Video S1: All NORMALIX95 tabs; Video S2: normalisation; Video S3: quantile normalisation; Video S4: check your samples; Video S5: transpose log antilog; Video S6: one vs all; Video S7: correlation; Video S8: heatmap; Video S9: scatterplot; Video S10: DGE

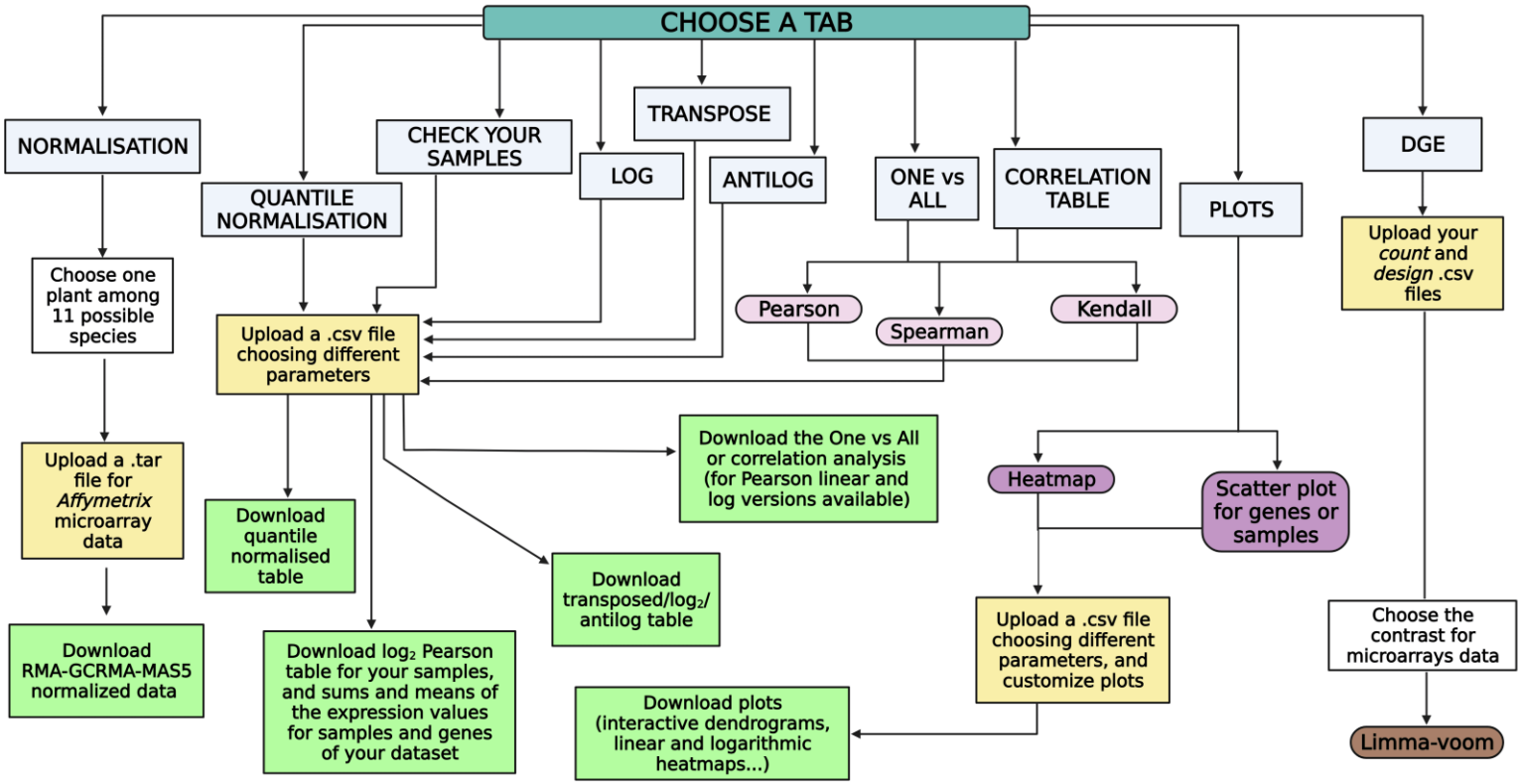


Fig. 1. NORMALIX95 organization. NORMALIX95 is organized into 10 independent main tabs. The *Normalisation* tab allows normalising *Affymetrix* raw data for 11 plant species, while the *Quantile normalisation* tab allows data to be quantile normalised. The *Check your samples* tab allows users to perform the cleaning strategy developed in our lab (Marzorati et al., 2021), which is a general method of dataset cleaning based on simple statistical criteria (the sum of the expression values and the Pearson correlation coefficient among samples' hybridizations). The following three sections (*Transpose*, *Log*, and *Antilog*) can be used to transpose the rows and columns of an input file or interconvert the logarithmic and linear formats. In the *One Vs All* and *Correlation table* tabs, users can perform three types of correlation analyses on an uploaded dataset. In the *Plots* tab, two types of plots can be generated (heatmaps and scatterplots), whereas in the last tab (*DGE*), differential gene expression analysis may be performed. Figure created by using BioRender.com

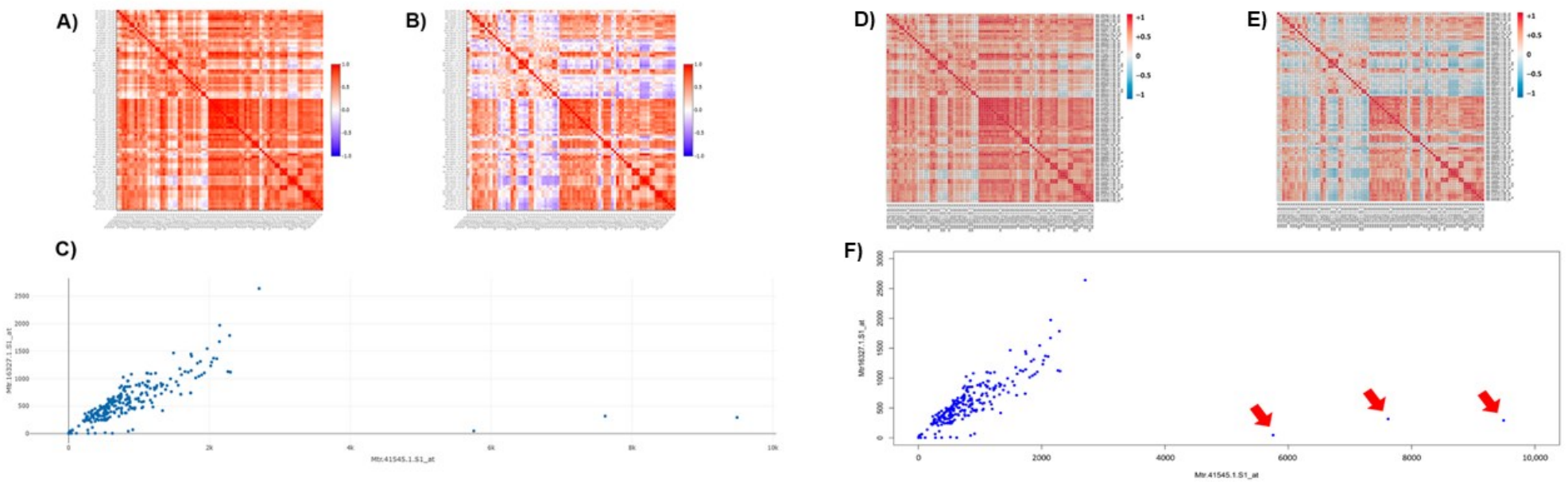


Fig. 2. Examples of NORMALIX95 customizable plots. Heatmaps for $\log_2$ Pearson correlation analysis of *Medicago truncatula* saponin dataset before **(A)** and after **(B)** cleaning of *the Medicago truncatula* Gene Expression Atlas (MtGEA) database, as described previously (Marzorati et al., 2021). **(C)** Scatter plot of the mean expression values of two *Medicago* mevalonate kinases across all samples, as described previously (Marzorati et al., 2021). All plots displayed in NORMALIX95 correspond to those previously obtained in Marzorati et al. (2021) **(D, E, F)**.

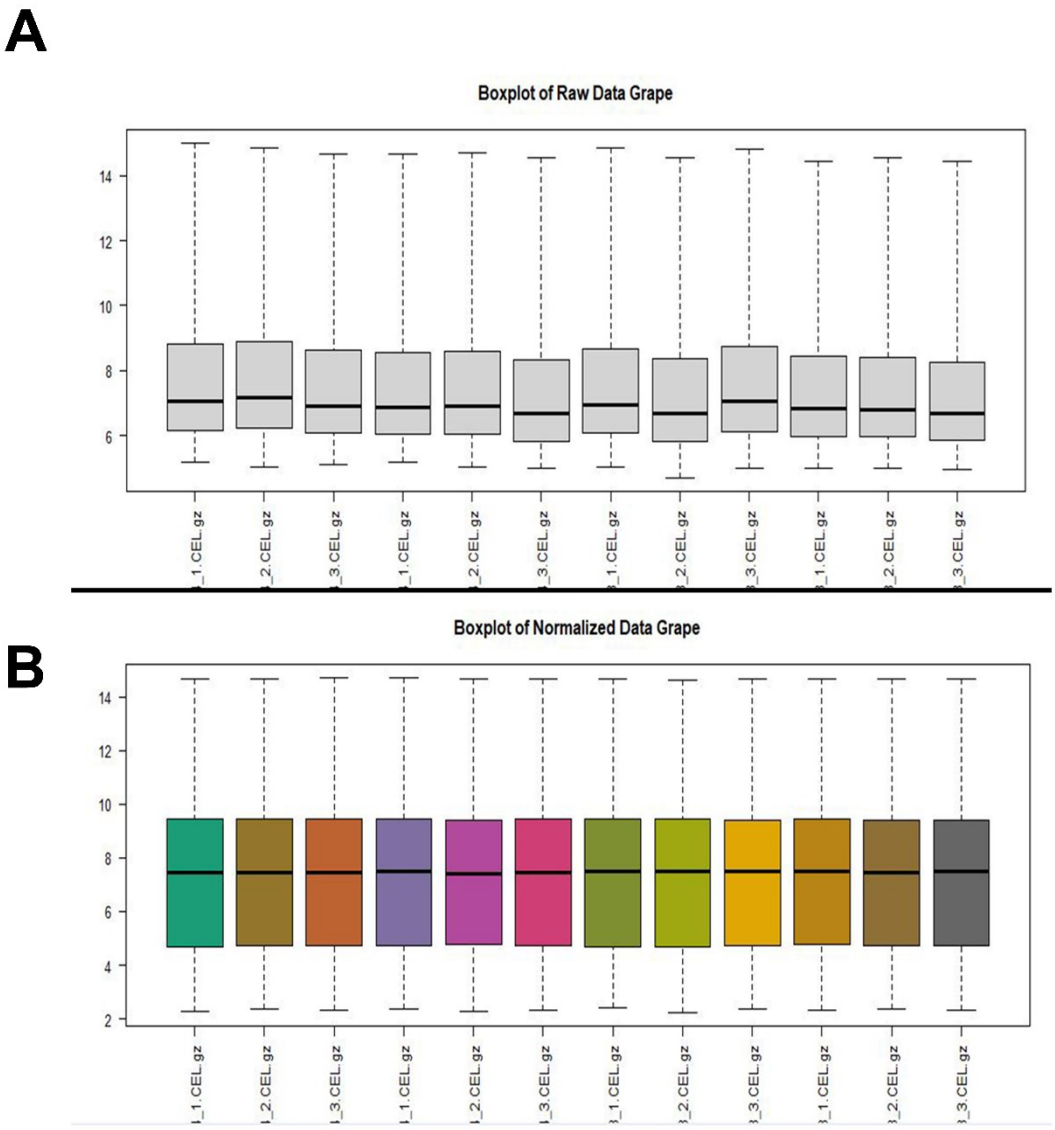


Figure S1. **Boxplots of raw (A) and RMA-normalized (B) data for samples from an experiment on *Vitis vinifera* uploaded in GEO.** Data of the experiment with GEO accession number GSE132311 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?) (Rondot and Reineke, 2019).

GEO accession number	Citation	Normalisation
GSE6404	Fung et al., 2008	GCOS (MAS5)
GSE7394	/	GCRMA
GSE7677	Robinson et al., 2015	GCRMA
GSE7679	/	GCRMA
GSE7680	/	GCRMA
GSE7681	Robinson et al., 2015	GCRMA
GSE7747	/	GCRMA
GSE7749	/	GCRMA
GSE7750	/	GCRMA
GSE7751	/	GCRMA
GSE7752	/	GCRMA
GSE7758	/	GCRMA
GSE8389	/	GCRMA
GSE8445	Robinson et al., 2015	GCRMA
GSE11406	Lund et al., 2008	GCRMA
GSE11857	/	CGRMA
GSE12842	Albertazzi et al., 2009	MAS5
GSE17502	Sreekantan et al., 2010	MAS5
GSE27180	Carvalho et al., 2011	MAS5
GSE29948	Tillett et al., 2012	RMA
GSE31594	/	MAS5
GSE31660	Vega et al., 2011	MAS5
GSE31662	/	MAS5
GSE31664	/	MAS5
GSE31674	Pilati et al., 2007	MAS5
GSE31675	/	MAS5
GSE31677	Cramer et al., 2007	MAS5
GSE41423	Liu et al., 2012	MAS5
GSE42315	Mar et al., 2013	MAS5
GSE44213	Choi et al., 2013	GCRMA
GSE53824	Toth et al., 2016	RMA
GSE62315	Xi et al., 2014	RMA
GSE66913	Fennell et al., 2015	MAS5
GSE89075	/	RMA
GSE89185	/	RMA
GSE132311	Rondot et al., 2019	RMA
GSE150581	Çakır Aydemir et al., 2020	GCRMA
GSE195646	Orduña et al., 2020	RMA

Table S1. Normalisation strategies used to normalise the *Vitis vinifera Affymetrix* data in GEO. List of experiments with GEO accession numbers, references, and normalization procedures used in the GPL1320 platform ([Vitis_Vinifera] Affymetrix Vitis vinifera (Grape) Genome Array, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?). / for unpublished data. Updated February 20, 2022.

5. References

Abbruscato P, Nepusz T, Mizzi L, et al. OsWRKY22, a Monocot WRKY Gene, Plays a Role in the Resistance Response to Blast. Molecular Plant Pathology 2012;13(8):828–841.

Affymetrix. Statistical Algorithms Description Document. 2002;1–28.

Afgan E, Baker D, Batut B, et al. The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2018 Update. Nucleic Acids Research 2018;46(W1):537–544.

Albertazzi G, Milc J, Caffagni A, et al. Gene Expression in Grapevine Cultivars in Response to Bois Noir Phytoplasma Infection. Plant Science 2009;176(6):792–804.

Altshuler D, Daly M, Kruglyak L. Guilt by Association. Nature Genetics 2000;26(2):135–137.

Baehrecke EH, Dang N, Babaria K, et al. Visualization and Analysis of Microarray and Gene Ontology Data with Treemaps. BMC Bioinformatics 2004;5(84).

Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: Archive for Functional Genomics Data Sets - Update. Nucleic Acids Research 2013;41(D1):991–995.

Beekweelder J, van Leeuwen W, van Dam NM, et al. The Impact of the Absence of Aliphatic Glucosinolates on Insect Herbivory in *Arabidopsis*. PLOS ONE 2008;3(4):e2068.

Benedito VA, Torres-Jerez I, Murray JD, et al. A Gene Expression Atlas of the Model Legume *Medicago truncatula*. Plant Journal 2008;55(3):504–513.

Bengtsson H, Corrada H, Gleixner J, et al. Package "MatrixStats". 2021;1–51.

Berri S, Abbruscato P, Faivre-Rampant O, et al. Characterization of WRKY Co-Regulatory Networks in Rice and *Arabidopsis*. BMC Plant Biology 2009;9(120).

Bolstad BM, Irizarry RA, Åstrand M, et al. A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias. Bioinformatics 2003;19(2):185–193.

Bumgarner R. Overview of DNA Microarrays: Types, Applications, and Their Future. Current Protocols in Molecular Biology 2013;4(5).

Çakır Aydemir B, Yüksel Özmen C, Kibar U, et al. Salt Stress Induces Endoplasmic Reticulum Stress-Responsive Genes in a Grapevine Rootstock. Plos One 2020;15(7):e0236424.

Carson A, Parmer C, Hocking T, et al. Package “Plotly”. 2021;1–72.

Carvalho B and Scharpf R. Package “OligoClasses”. 2015;1–75.

Carvalho LC, Vilela BJ, Mullineaux PM, et al. Comparative Transcriptomic Profiling of *Vitis vinifera* under High Light Using a Custom-Made Array and the Affymetrix Genechip. Molecular Plant 2011;4(6):1038–1051.

Chang W, J.Cheng, Allaire JJ, et al. Shiny: Web Application Framework for R. 2016.

Chang W and Ribeiro BB. Package “ShinyDashboard”. 2018;1–27.

Chee M, Yang R, Hubbell E, et al. Accessing Genetic Information with High-Density DNA Arrays. Science 1996;274(5287):610–614.

Choi HK, Iandolino A, Da Silva FG, et al. Water Deficit Modulates the Response of *Vitis vinifera* to the Pierce’s Disease Pathogen *Xylella fastidiosa*. Molecular Plant-Microbe Interactions 2013;26(6):643–657.

Choi K and Ratner N. IGEAK: An Interactive Gene Expression Analysis Kit for Seamless Workflow Using the R/Shiny Platform. BMC Genomics 2019;20(177).

Cramer GR, Ergül A, Grimplet J, et al. Water and Salinity Stress in Grapevines: Early and Late Changes in Transcript and Metabolite Profiles. Functional and Integrative Genomics 2007;7(2):111–134.

Dalma-Weiszhausz DD, Warrington J, Tanimoto EY, et al. The Affymetrix GeneChip® Platform: An Overview. Methods in Enzymology 2006;410:3–28.

Dowle M and Srinivasan A. Package “Data.Table”. 2021;1–127.

Edgar R. Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository. Nucleic Acids Research 2002;30(1):207–210.

Eijssen LMT, Jaillard M, Adriaens ME, et al. User-Friendly Solutions for Microarray Quality Control and Pre-Processing on ArrayAnalysis.Org. Nucleic Acids Res 2013;41(W1):71–76.

Everitt BS and Hothorn T. Package “MVA”. 2021;1–4.

Fennell AY, Schlauch KA, Gouthu S, et al. Short Day Transcriptomic Programming during Induction of Dormancy in Grapevine. Frontiers in Plant Science 2015;6(834).

Fung RWM, Gonzalo M, Fekete C, et al. Powdery Mildew Induces Defense-Oriented Reprogramming of the Transcriptome in a Susceptible but Not in a Resistant Grapevine. Plant Physiology 2008;146(1):236–249.

Galili T, O'Callaghan A, Sidi J, et al. Package "Heatmaply". 2021;1–27.

Gautier L, Cope L, Bolstad BM, et al. Affy - Analysis of *Affymetrix* GeneChip Data at the Probe Level. Bioinformatics 2004;20(3):307–315.

Gentleman AR, Carey VJ, Huber W, et al. Package "Genefilter". 2021;1–39.

Gentleman R, Carey V, Morgan M, et al. Package "Biobase". 2015;1–86.

Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: Open Software Development for Computational Biology and Bioinformatics. Genome Biol 2004;5(R80).

Goecks J, Nekrutenko A, Taylor J, et al. Galaxy: A Comprehensive Approach for Supporting Accessible, Reproducible, and Transparent Computational Research in the Life Sciences. Genome Biology 2010;11(R86).

Harr B and Schlötterer C. Comparison of Algorithms for the Analysis of *Affymetrix* Microarray Data as Evaluated by Co-Expression of Genes in Known Operons. Nucleic Acids Research 2006;34(2).

He J, Benedito VA, Wang M, et al. The *Medicago truncatula* Gene Expression Atlas Web Server. BMC Bioinformatics 2009;10(441).

Heath KD, Burke PV. and Stinchcombe JR. Coevolutionary Genetic Variation in the Legume-Rhizobium Transcriptome. Molecular Ecology 2012;21(19):4735–4747.

Huber W, Carey VJ, Gentleman R, et al. Orchestrating High-Throughput Genomic Analysis with Bioconductor. Nature Methods 2015;12(2):115–121.

Irizarry RA, Hobbs B, Collin F, et al. Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. Biostatistics 2003;4(2):249–264.

Jaksik R, Iwanaszko M, Rzeszowska-Wolny J, et al. Microarray Experiments and Factors Which Affect Their Reliability. Biology Direct 2015;10(1).

Khomtchouk BB, Hennessy JR and Wahlestedt C. Shinyheatmap: Ultra Fast Low Memory Heatmap Web Interface for Big Data Genomics. Plos One 2017;12(5).

Ledford H. The Death of Microarrays? Nature 2008;455(7215):847.

Lemieux B, Aharoni A and Schena M. Overview of DNA Chip Technology. Molecular Breeding 1998;4(4):277–289.

Lim WK, Wang K, Lefebvre C, et al. Comparative Analysis of Microarray Normalization Procedures: Effects on Reverse Engineering Gene Networks. Bioinformatics 2007;23(13):282–288.

Liu GT, Wang JF, Cramer G, et al. Transcriptomic Analysis of Grape (*Vitis vinifera* L.) Leaves during and after Recovery from Heat Stress. BMC Plant Biology 2012;12(174).

Lockhart DJ, Dong H, Byrne MC, et al. Expression Monitoring by Hybridization to High-Density Oligonucleotide Arrays. Nature Biotechnology 1996;14(13):1675–1680.

Lowe R, Shirley N, Bleackley M, et al. Transcriptomics Technologies. Plos Computational Biology 2017;13(5).

Lund ST, Peng FY, Nayar T, et al. Gene Expression Analyses in Individual Grape (*Vitis vinifera* L.) Berries during Ripening Initiation Reveal That Pigmentation Intensity Is a Valid Indicator of Developmental Staging within the Cluster. Plant Molecular Biology 2008;68(3):301–315.

Månsson R, Tsapogas P, Åkerlund M, et al. Pearson Correlation Analysis of Microarray Data Allows for the Identification of Genetic Targets for Early B-Cell Factor. Journal of Biological Chemistry 2004;279(17):17905–17913.

Marè C, Aprile A, Roncaglia E, et al. Rootstock and Soil Induce Transcriptome Modulation of Phenylpropanoid Pathway in Grape Leaves. Journal of Plant Interactions 2013;8(4):334–349.

Marzorati F, Wang C, Pavesi G, et al. Cleaning the *Medicago* Microarray Database to Improve Gene Function Analysis. Plants 2021;10(1240).

Menges M, Dóczi R, Ökrész L, et al. Comprehensive Gene Expression Atlas for the *Arabidopsis* MAP Kinase Signalling Pathways. New Phytologist 2008;179(3):643–662.

Millenaar FF, Okyere J, May ST, et al. How to Decide? Different Methods of Calculating Gene Expression from Short Oligonucleotide Array Data Will Give Different Results. BMC Bioinformatics 2006;7(137).

Murgia I, Tarantino D, Soave C, et al. *Arabidopsis* CYP82C4 Expression Is Dependent on Fe Availability and Circadian Rhythm, and Correlates with Genes Involved in the Early Fe Deficiency Response. Journal of Plant Physiology 2011;168(9):894–902.

Murgia I, Vigani G, Di Silvestre D, et al. Formate Dehydrogenase Takes Part in Molybdenum and Iron Homeostasis and Affects Dark-Induced Senescence in Plants. Journal of Plant Interactions 2020;15(1):386–397.

Naoumkina MA, Modolo L V., Huhman D V., et al. Genomic and Coexpression Analyses Predict Multiple Genes Involved in Triterpene Saponin Biosynthesis in *Medicago truncatula*. Plant Cell 2010;22(3):850–866.

Nelson JW, Sklenar J, Barnes AP, et al. The START App: A Web-Based RNAseq Analysis and Visualization Resource. Bioinformatics 2017;33(3):447–449.

Noble D. DNA Sequencing on a Chip. Analytical Chemistry 1995;67(5):201–204.

Orduña L, Li M, Navarro-Payá D, Zhang C, et al. Direct Regulation of Shikimate, Early Phenylpropanoid, and Stilbenoid Pathways by Subgroup 2 R2R3-MYBs in Grapevine. The Plant Journal 2022;110(2):529-547.

Pavlopoulos GA, Malliarakis D, Papanikolaou N, et al. Visualizing Genome and Systems Biology: Technologies, Tools, Implementation Techniques and Trends, Past, Present and Future. Gigascience 2015;4(1).

Pilati S, Perazzolli M, Malossini A, et al. Genome-Wide Transcriptional Analysis of Grapevine Berry Ripening Reveals a Set of Genes Similarly Modulated during Three Seasons and the Occurrence of an Oxidative Burst at Vèraison. BMC Genomics 2007;8(428).

Prasad TV and Ahson SI. Visualization of Microarray Gene Expression Data. Bioinformation 2006;1(4):141–145.

Quackenbush J. Microarray Data Normalization and Transformation. Nature Genetics 2002;32(4S):496–501.

Raudvere U, Kolberg L, Kuzmin I, et al. g:Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update). Nucleic Acids Research 2019;47(W1):191–198.

Reyes ALP, Silva TC, Coetzee SG, et al. GENAVi: A Shiny Web Application for Gene Expression Normalization, Analysis and Visualization. BMC Genomics 2019;20(1).

Ritchie ME, Phipson B, Wu D, et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. Nucleic Acids Research 2015;43(7):e47.

Robinson S, Glonek G, Koch I, et al. Alignment of Time Course Gene Expression Data and the Classification of Developmentally Driven Genes with Hidden Markov Models. BMC Bioinformatics 2015;16(1).

Rondot Y and Reineke A. Endophytic Beauveria Bassiana Activates Expression of Defence Genes in Grapevine and Prevents Infections by Grapevine Downy Mildew *Plasmopara viticola*. Plant Pathology 2019;68(9):1719–1731.

Smyth G. Package “Limma”. 2021;1–267.

Sreekantan L, Mathiason K, Grimplet J, et al. Differential Floral Development and Gene Expression in Grapevines during Long and Short Photoperiods Suggests a Role for Floral Genes in Dormancy Transitioning. Plant Molecular Biology 2010;73(1–2):191–205.

Tillett RL, Wheatley MD, Tattersall EAR, et al. The *Vitis vinifera* C-Repeat Binding Protein 4 (VvCBF4) Transcriptional Factor Enhances Freezing Tolerance in Wine Grape. Plant Biotechnology Journal 2012;10(1):105–124.

Toth Z, Winterhagen P, Kalapos B, et al. Expression of a Grapevine NAC Transcription Factor Gene Is Induced in Response to Powdery Mildew Colonization in Salicylic Acid-Independent Manner. Scientific Reports 2016;6(30825).

Vandepoele K, Quimbaya M, Casneuf T, et al. Unraveling Transcriptional Control in *Arabidopsis* Using Cis-Regulatory Elements and Coexpression Networks1[C][W]. Plant Physiology 2009;150(2):535–546.

Wickham H. Package "Stringr". 2019;1–32.

Wickham H. Package "Tidyr". 2021;1–42.

Wickham H, Chang W, Henry L, et al. Package "ggplot2". 2021a;1–292.

Wickham H, Romain F, Henry L, et al. Package "Dplyr." 2021b;1–85.

Wu Z. A Review of Statistical Methods for Preprocessing Oligonucleotide Microarrays. Statistical Methods in Medical Research 2009;18(6):533–541.

Wu Z and Irizarry R. Package "Gcrma". 2014;1–6.

Xi H, Ma L, Liu G, et al. Transcriptomic Analysis of Grape (*Vitis vinifera* L.) Leaves after Exposure to Ultraviolet C Irradiation. Plos ONE 2014;9(12).

Xie Y, Cheng J and Tan X. Package "DT". 2021;1–19.

Zermiani M, Begheldo M, Nonis A, et al. Identification of the *Arabidopsis* RAM/MOR Signalling Network: Adding New Regulatory Players in Plant Stem Cell Maintenance and Cell Polarization. Annals of Botany 2015;116(1):69–89.

Zhao S and Bruce WB. Expression Profiling Using CDNA Microarrays. Methods Mol Biol 2003;236(1):365–380.

Zhao Y, Wong L and Goh WW Bin. How to Do Quantile Normalization Correctly for Gene Expression Data Analyses. Scientific Reports 2020;10(1).

3.2 Works about FDH and bacteria

"[...] It's not good and bad. It's just whether there's too much of it or too little of it and things are out of balance, so the 'bad things' have an opportunity to prosper."

— Nigel Palmer

Manuscript 3

Marzorati F, Vigani G, Morandini P, Murgia I. 2021. Formate dehydrogenase contributes to the early *Arabidopsis thaliana* responses against *Xanthomonas campestris* pv *campestris* infection. *Physiological and Molecular Plant Pathology* **114**, 101633.

Manuscript 4 (under review)

Marzorati F, Rossana R, Bernardo L, Mauri P, Di Silvestre D, Morandini P, Murgia I. **2023**. *Arabidopsis thaliana* early foliar proteome response to root exposure to the rhizobacterium *Pseudomonas simiae* WCS417. *Molecular Plant-Microbe Interactions*.

Manuscript 3

Formate dehydrogenase contributes to the early *Arabidopsis thaliana* responses against *Xanthomonas campestris* pv *campestris* infection

Francesca Marzorati [1], Gianpiero Vigani [2], Piero Morandini [1], Irene Murgia [3,*]

[1] Dept. Environmental Science and Policy, University of Milano, Milano, Italy
[2] Plant Physiology Unit, Dept. Life Sciences and Systems Biology, University of Torino, Torino, Italy
[3] Dept. Biosciences, University of Milano, Milan, Italy
* Correspondence: irene.murgia@unimi.it

Abstract

Formate dehydrogenase (FDH) catalyzes the conversion of formate ($HCOO^-$) into carbon dioxide (CO_2), coupled with the reduction of NAD^+ to NADH. FDH involvement in the response against abiotic stress is well established whereas FDH role against pathogen attack is less known. Therefore, the *in silico* correlation analysis of the FDH transcript with a biotic stress dataset was performed, which highlighted correlation with genes involved in the response against pathogen attack. Next, a reduction in the expression of an FDH reporter construct in the hydathodes of *Arabidopsis thaliana* leaves infected with *Xanthomonas campestris* pv *campestris* (*Xcc*) was observed. We also observed an increased proliferation of *Xcc* in the hydathodes of an *atfdh*1-5 mutant during the early stages of infection, which further supports FDH involvement in an early defense response activated in the hydathodes upon *Xcc* infection, possibly through modulation of formate levels.

Keywords: *Arabidopsis thaliana;* Biotic stress; Correlation analysis; Formate dehydrogenase; Hydathodes; *Xanthomonas campestris* pv *campestris*

Physiological and Molecular Plant Pathology 114 (2021) 101633

Contents lists available at ScienceDirect

Physiological and Molecular Plant Pathology

journal homepage: www.elsevier.com/locate/pmpp

Formate dehydrogenase contributes to the early *Arabidopsis thaliana* responses against *Xanthomonas campestris* pv *campestris* infection

Francesca Marzorati [a], Gianpiero Vigani [b], Piero Morandini [a], Irene Murgia [c,*]

[a] Dept. Environmental Science and Policy, University of Milano, Italy
[b] Plant Physiology Unit, Dept. Life Sciences and Systems Biology, University of Torino, Italy
[c] Dept. Biosciences, University of Milano, Italy

ARTICLE INFO

Keywords:
Arabidopsis thaliana
Biotic stress
Correlation analysis
Formate dehydrogenase
Hydathodes
Xanthomonas campestris pv *campestris*

ABSTRACT

Formate dehydrogenase (FDH) catalyzes the conversion of formate ($HCOO^-$) into carbon dioxide (CO_2), coupled with the reduction of NAD^+ to NADH. FDH involvement in the response against abiotic stress is well established whereas FDH role against pathogen attack is less known. Therefore, the *in silico* correlation analysis of the FDH transcript with a biotic stress dataset was performed, which highlighted correlation with genes involved in the response against pathogen attack. Next, a reduction in the expression of an FDH reporter construct in the hydathodes of *Arabidopsis thaliana* leaves infected with *Xanthomonas campestris* pv *campestris* (*Xcc*) was observed. We also observed an increased proliferation of *Xcc* in the hydathodes of an *atfdh1-5* mutant during the early stages of infection, which further supports FDH involvement in an early defense response activated in the hydathodes upon *Xcc* infection, possibly through modulation of formate levels.

1. Introduction

Plants are exposed to many different environmental stresses during their life cycle, including aggressions by a wide variety of organisms. Plants have developed different metabolic strategies to counteract abiotic and biotic stresses, however the signaling networks regulating plant defense responses are complex and far from complete elucidation [1]. Formate dehydrogenases (FDHs) are enzymes found in bacteria, fungi and plants; they are classified into two main groups according to their structural characteristics, the strategy of catalysis and presence of prosthetic groups (iron–sulphur clusters, molybdenum and tungsten ions) in their active site [2,3]. FDH enzymes catalyze the oxidation of formate ($HCOO^-$) into carbon dioxide (CO_2); formate is a simple one-carbon compound produced through different pathways in plants where it is present in small pools, typically ranging from 0.1 to 1 μmol g^{-1} fresh weight [4]. In plants, only NAD^+-dependent FDHs have been identified [3], which couple the oxidation of formate with the NAD^+ reduction into NADH. FDH is localized in mitochondria, where NADH can feed the respiratory chain [2] and its level can be very high in heterotrophic tissues, especially under stress [5]; FDH can also be localized in chloroplasts [6]. Plant FDHs are considered stress enzymes since their transcripts accumulate under unfavorable conditions [3]. FDH is involved in plant response against abiotic stresses [7,8], as well as in plant nutrition: FDH indeed represents a key enzyme in plant responses against iron (Fe) and Molybdenum (Mo) nutritional stress, as well as a nutritional hub of Mo homeostasis [8–10].

On the contrary, few reports focused on FDH role in plant responses against biotic stresses: *Arabidopsis thaliana fdh* insertional mutants displayed enhanced disease symptoms upon infection with avirulent *Pseudomonas syringae* pv *tomato* Pst DC3000 (avrRpm1); these findings led to the general hypothesis that FDH could play a role in the regulation of defense responses against bacterial pathogens [6]. However, a model for FDH action, at molecular level, in such conditions, has not been proposed so far.

In the present work, we investigated FDH expression under biotic stress with an *in silico* approach and the *in vivo* activity of FDH promoter. In particular, we first performed FDH expression correlation analysis to identify the best FDH gene correlators under biotic stress conditions. We then analyzed FDH expression during the bacterial infection of *A. thaliana* leaves with *Xanthomonas campestris* pv *campestris* (*Xcc*), a vascular pathogen responsible for the "black rot" disease of cruciferous crops. Such pathogen enters the vascular system of the attacked plants preferentially through the hydathodes, i.e. the terminal pores of the xylem found on leaf margins of many vascular plants and used for

* Corresponding author.
E-mail addresses: francesca.marzorati1@unimi.it (F. Marzorati), gianpiero.vigani@unito.it (G. Vigani), piero.morandini@unimi.it (P. Morandini), irene.murgia@unimi.it (I. Murgia).

https://doi.org/10.1016/j.pmpp.2021.101633
Received 23 September 2020; Received in revised form 22 February 2021; Accepted 23 February 2021
Available online 5 March 2021

guttation [11–13]. Moreover, we analyzed the spread of *Xcc* in the hydathodes of *A. thaliana* wt leaves and those of the FDH KO mutant *atfdh1-5*. Overall, our results suggest that regulation of FDH could take part in an early defense response activated in hydathodes to counteract *Xcc* infection.

2. Material and methods

2.1. Transcript correlation analysis

The generation of the biotic stress dataset and the data correlation analysis were conducted in the R programming environment (https://www.R-project.org/) [14] (version 4.0.0). Data were downloaded and normalized using GEOquery and affy packages [15,16]. The biotic stress dataset contained 63 experiments of *A. thaliana* which were retrieved from the public functional genomics data repository Gene Expression Omnibus GEO at https://www.ncbi.nlm.nih.gov/geo [17,18] (GPL198 platform) and in particular 29 experiments of fungal infection, 25 experiments of bacterial infection, 2 experiments of viral infection and 7 experiments of insects/nematodes attack. Data were normalized using the RMA method. The dataset includes a total of 946 hybridizations. Annotation was performed and checked using g:Profiler (https://biit.cs.ut.ee/gprofiler/gost) [19] (http://www.affymetrix.com/support/technical/byproduct.affx?product=arab), and Araport (https://bar.utoronto.ca/thalemine/begin.do) [20]. Putative intracellular localization and transmembrane domains are according to Aramemnon database (http://aramemnon.uni-koeln.de/index.ep) [21].

2.2. A.thaliana growth

A.thaliana wt Col, *atfdh1-5* Salk_-108751 (N869258) [6,8], *Vu* FDH::GUS [7] were grown on Technic n.1 DueEmme soil as described in [8] and used for the infection experiments by constant leaf submersion with pathogen suspension (see below for details). These *A. thaliana* lines were also grown on the same Technic n.1 DueEmme soil, in a different greenhouse, at 23 °C, 150 μE m^2sec^{-1}, 12 h/12h light/dark photoperiod, for the infection experiments consisting in 10 min submersion with pathogen suspension (see below for details).

2.3. Staining for β-glucuronidase (GUS) activity

Staining of leaves for β-glucuronidase (GUS) activity was performed as described in [22].

2.4. A.thaliana infection with Xcc

The *Xanthomonas campestris* pv *campestris* strains used in this study were *Xcc* 8004 strain and *Xcc* 8004 mutant strain expressing GUS as well as green fluorescent protein (GFP) (hereafter referred as *Xcc* 8004::GUS-GFP), as described in [12]. Both *Xcc* bacterial strains were grown overnight, in 100 ml glass flasks (28°C, 130 rpm shaker agitation rate), in MOKA rich medium (0.4% (w/v) yeast extract, 0.8% (w/v) casamino acids, 0.2% (w/v) K_2HPO_4, and 0.03% (w/v) $MgSO_4{\cdot}7H_2O$), as described in [12]. Rifampicin was used at 50 μg ml^{-1}.

Xcc bacteria were then prepared for *A. thaliana* infection as described in [12]; briefly: bacteria were harvested from the liquid broth by centrifugation (10 min at 4000 g) and were suspended at either 10^8 cfu ml^{-1} or 10^4 cfu ml^{-1} in 1 mM $MgCl_2$. A final concentration of 0.02% Silwet® L-77 (Momentive Performance Materials) was then added to the bacterial suspension.

Infection of *A. thaliana* leaves was performed by applying two different protocols. The first protocol was as described in [11] with minor modifications; briefly: leaves detached from 4 to 5 weeks old plants were inserted into 2 ml Eppendorf tubes; a volume of bacterial suspension sufficient to submerge at least half of the leaf area was added in each tube, carefully avoiding any contact with the petioles. Leaves were kept at 22–25°C in the bacterial suspension for up to 48 h, then removed from the suspension, surface sterilized by immersion in 70% EtOH for 20 s and washed twice with sterile water, as described previously [23]. Leaves were maintained at room temperature for 6 h and stained for GUS activity.

As an alternative protocol, detached leaves were submerged for half of the leaf area into the bacterial suspension described above for 10 min only; as control (mock) treatment, leaves were submerged for 10 min into 1 mM $MgCl_2$ containing 0.02% Silwet L-77. All the leaves were then removed from the suspension and maintained onto moist paper into Petri dishes sealed with 3 M Micropore Surgical tape, at 22–25°C, 150 μE m^2sec^{-1}, 10 h/14 h light/dark photoperiod for up to 48 h. Petri dishes were inspected every few hours and sterile water was added onto the moist paper, if needed, to preserve leaves in a humid environment. Leaves were then removed from Petri dishes after either 24 or 48 h, surface sterilized and stained for GUS activity as described above.

Table 1
Top correlators of *A.thaliana* FDH transcript in logarithmic analysis, by using the biotic stress datasets. The genes are listed together with their respective Pearson correlation coefficients from log analysis. Putative intracellular localization and presence of transmembrane domains are according to Aramemnon database (http://aramemnon.uni-koeln.de/) [21].

AGI code	Gene name	Pearson coefficient (log)	Localization	Transmembrane domain
AT5G14780	FDH	1	chloropl. mitoch.	no
AT1G06650	AT1G06650	0.8052	nucleus	no
AT4G15530	PPDK	0.80463	chloropl, mitoch.	unclear
AT1G55110	IDD7	0.79966	nucleus	no
AT5G43450	AT5G43450	0.77845	unclear	unclear
AT3G44880	PAO	0.7779	chloropl.	yes
AT4G03410	AT4G03410	0.77496	chloropl.	yes
AT1G70160	AT1G70160	0.77298	secr. pathway	yes
AT5G52450	DTX16	0.77125	secr. pathway	yes
AT1G64740	TUA1	0.76407	cytoplasmic	unclear
AT3G54360	NCA1	0.76309	chloropl., nucleus	yes
AT5G13800	PPH	0.76291	chloropl., secr. pathway	yes
AT3G56050	AT3G56050	0.75879	secr. pathway, mitoch.	yes
AT1G54130	RSH3	0.75804	chloropl.	unclear
AT1G73680	ALPHA DOX2	0.75233	chloropl., secr.pathway	yes

3. Results

3.1. FDH expression positively correlates with various genes involved in plant defense responses

Correlation analysis of FDH expression was performed by using a "biotic stress" dataset (Supplementary Table S1). The complete results from the analysis in logarithmic or in linear space are shown in Supplementary Table S2 and Table S3, respectively. Several genes, among the top correlators of FDH in the logarithmic analysis, are discussed below in detail: they code for proteins with various subcellular localizations and several of them are predicted to contain transmembrane domains (Table 1) according to Aramemnon database [21] (http://aramemnon.uni-koeln.de). The AT1G06650 gene is similar to 2-oxoglutarate/Fe(II)-dependent oxygenases. These enzymes catalyze a wide range of reactions: in plants, they are involved in both secondary and primary metabolism and they play a crucial role in the synthesis/degradation/recycling of small molecules (among which also some hormones) and the biosynthesis of secondary metabolites [24,25]. The chloroplastic pyruvate phosphate dikinase PPDK is involved in leaf senescence and it is reported to alleviate both biotic and abiotic stresses when overexpressed [26]. Indeterminate domain 7 (IDD7) is an uncharacterized member of a group of transcriptional factors playing

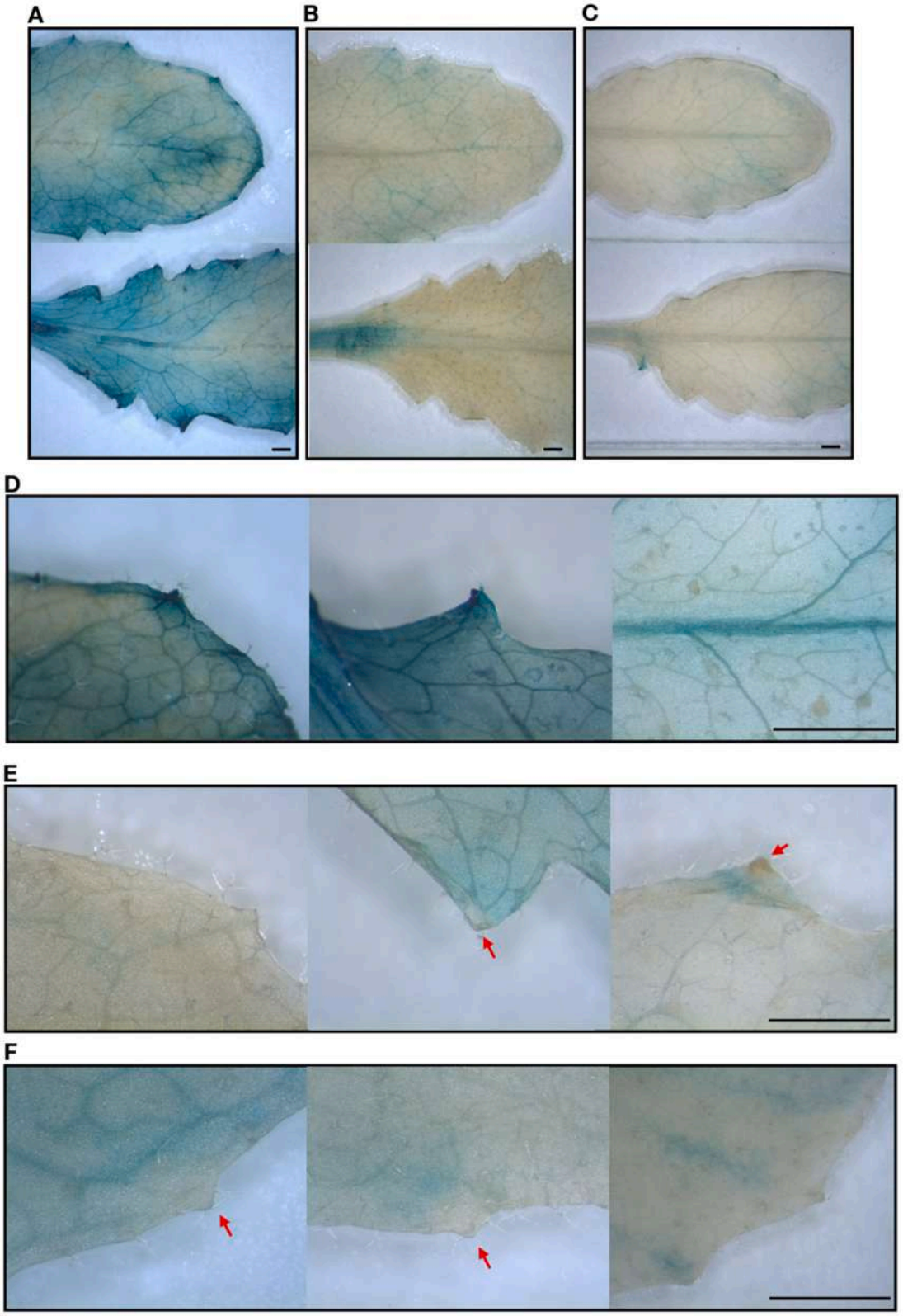


Fig. 1. Infection of *A. thaliana Vu* FDH:GUS leaves with *Xcc* by constant submersion in 10^8 cfu ml^{-1} bacterial suspension. Leaves detached from *Vu* FDH::GUS rosettes were (**A, D**) stained for GUS activity or were infected with *Xcc* and then stained for GUS activity after (**B, E**) 24 h, or (**C, F**) 48 h from infection. **(A, B, C)** whole leaves; **(D, E, F)** hydathodes, their surroundings or the central vein. Arrows point to hydathodes in which GUS staining in their epithem is no longer detectable. Scale bars in each panel represent 1 mm. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

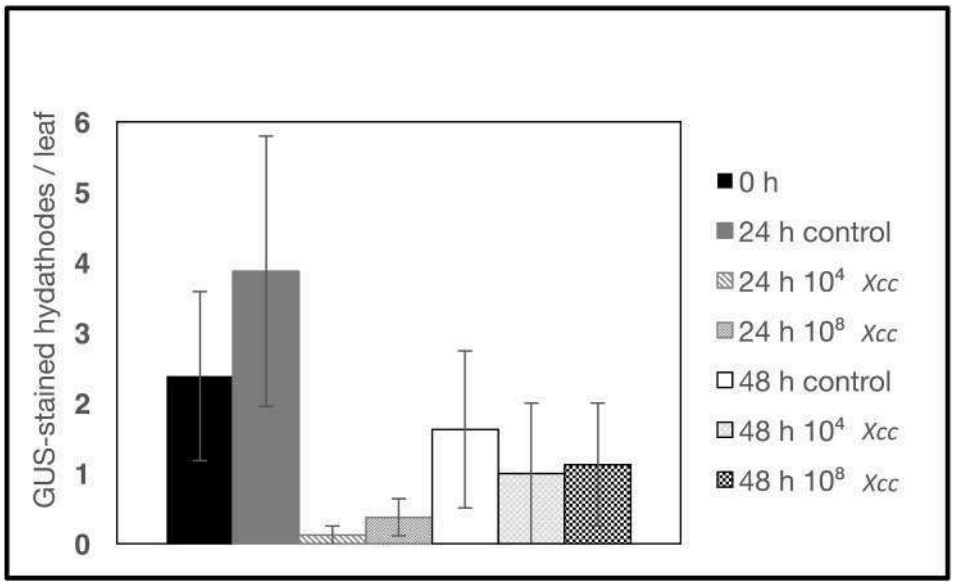


Fig. 2. Infection of *A. thaliana Vu* FDH::GUS leaves with *Xcc* by 10 min submersion in 10^8 or 10^4 cfu ml^{-1} bacterial suspension. Leaves detached from 5 weeks old *Vu* FDH::GUS plants were stained for GUS activity before infection (0 h), after 24 h or 48 h mock-treatment (24 h control, 48 h control) and after 24 h or 48 h infection with either 10^8 or 10^4 cfu ml^{-1} bacterial suspension. Each bar represents, for each condition, the mean number ±SE of stained hydathodes in 8 leaves.

many roles in plant development and physiology, also involved in immune responses [27]. AT5G43450 gene encodes a protein belonging to the 2-oxoglutarate (2OG)/Fe(II)-dependent oxygenase superfamily protein; its sequence is similar to ACC-oxidase, essential for ethylene production and signaling, which, in turn, also takes part in the defense responses against pathogen attack [28]. Pheophorbide a Oxygenase (PAO) is a key enzyme of chlorophyll catabolism, which normally occurs during leaf senescence and fruit ripening [29]. Notably, the involvement of PAO in jasmonic acid (JA) signaling has been recently proposed [30]. AT4G03410 is an uncharacterized protein, putatively belonging to the PMP22 family, a group of peroxisomal membrane proteins [31]. AT1G70160 codes for a protein of unknown function. Protein detoxification 16 (DTX16) is a carrier of the MATE (Multidrug And Toxic compound Extrusion) family protein [32]; these transporters extrude a wide range of substrates, such as plant hormones, toxic compounds and secondary metabolites [33,34]. αTubulin1 (TUA1) is preferentially expressed in pollen [35]. The catalase chaperone NCA1 contributes to the regulation of reactive oxygen species (ROS) homeostasis and to autophagy induction in *A. thaliana* [36,37]. The chloroplastic Pheophytin Pheophorbide Hydrolase PPH degrades chlorophyll during *A. thaliana* leaf senescence, producing phytol that enters in tocopherol biosynthesis [38,39]. AT3G56050 encodes a Leucin rich repeat Receptor like protein (LRR-RLK), a member of the receptor-like protein kinase superfamily, important in plant cellular signaling, with members strongly upregulated under biotic stress [40,41]. RSH3 is a RelA/SpoT-like protein which regulates the levels of guanosine tetraphosphate ppGpp, the effector of the "stringent response" in bacteria [42]. In plants, RSHs are nuclei-encoded and targeted in chloroplasts; their precise functions are mostly unknown. However, RSHs are involved in the modulation of transcription and translation in responses to environmental stress ([43] and references therein). During senescence, RSHs are involved in the coordination of nutrient remobilization and relocation from vegetative tissues into seeds [44,45]. ALPHA DOX2 is responsible for a primary oxygenation step in oxylipins biosynthesis [46]; these lipids mediate different developmental processes and responses to environmental stresses [47,48]. The top gene correlators of the FDH in linear analysis are reported in Table S4; four of them, i.e. PPH, DTX16, IDD7, PAO, are also top correlators in log analysis (see Table 1).

3.2. The activity of FDH promoter in hydathodes is reduced during the infection by Xanthomonas campestris pv campestris

Hydathodes control systemic infection by *Xanthomonas campestris* pv *campestris* (*Xcc*) [12,13] and are localized, in *A. thaliana*, along the leaf margins at the leaf apex and at the tip of the teeth; the number of leaf teeth increases in rosettes' upper leaves ([13] and references therein). Given the fact that the *Vigna umbellata* FDH promoter is active in *A. thaliana* hydathodes [8], its role during *Xcc* infection was therefore investigated. To this end, leaves of *A. thaliana* reporter line expressing a *V. umbellata* FDH promoter::GUS construct (*Vu* FDH::GUS) [7,8] were detached and infected with *Xcc* by submerging leaves in a 10^8 cfu ml^{-1} *Xcc* 8004 suspension, as described in [11]; GUS activity was then evaluated after 24 or 48 h of submersion (Fig. 1). Before infection, the FDH promoter was active in the whole leaf tissue (Fig. 1A) and particularly in hydathodes (Fig. 1D), as already reported in [8]. After 24 h, FDH promoter's activity is reduced in whole leaves (Fig. 1B) and in hydathodes and their surroundings; furthermore, it is occasionally no longer detectable in the epithem cells of the hydathodes, while it is still detectable in close proximity (Fig. 1E). After 48 h, GUS activity is further reduced (Fig. 1C and F). To confirm these results, a second independent batch of *Vu* FDH::GUS plants were grown, taking all precautions to avoid any possible source of stress, whether biotic or abiotic, during growth; indeed, all the plants appeared healthy, with impressive very large rosettes and leaves (Supplementary Fig. S1). Leaves were infected by 10 min submersion with either 10^4 cfu ml^{-1} or 10^8 cfu ml^{-1} *Xcc* 8004; both infected and mock-treated leaves were maintained either 24 h or 48 h onto moist paper in sealed Petri dishes (Supplementary Fig. S2) before GUS staining for detection of FDH promoter activity. Mock-treatment (control) represented a stress *per se* and indeed it induced FDH promoter activity already after 24 h (Supplementary Fig. S3A); the number of GUS-stained hydathodes increased (Fig. 2, 24 h control). Under these conditions, *Xcc* infection was able to inhibit FDH promoter activity (Supplementary Fig. S3B,C) and indeed the number of GUS-stained hydathodes is reduced when compared to corresponding controls (Fig. 2). Moreover, such *Xcc* inhibitory effect on FDH promoter activity

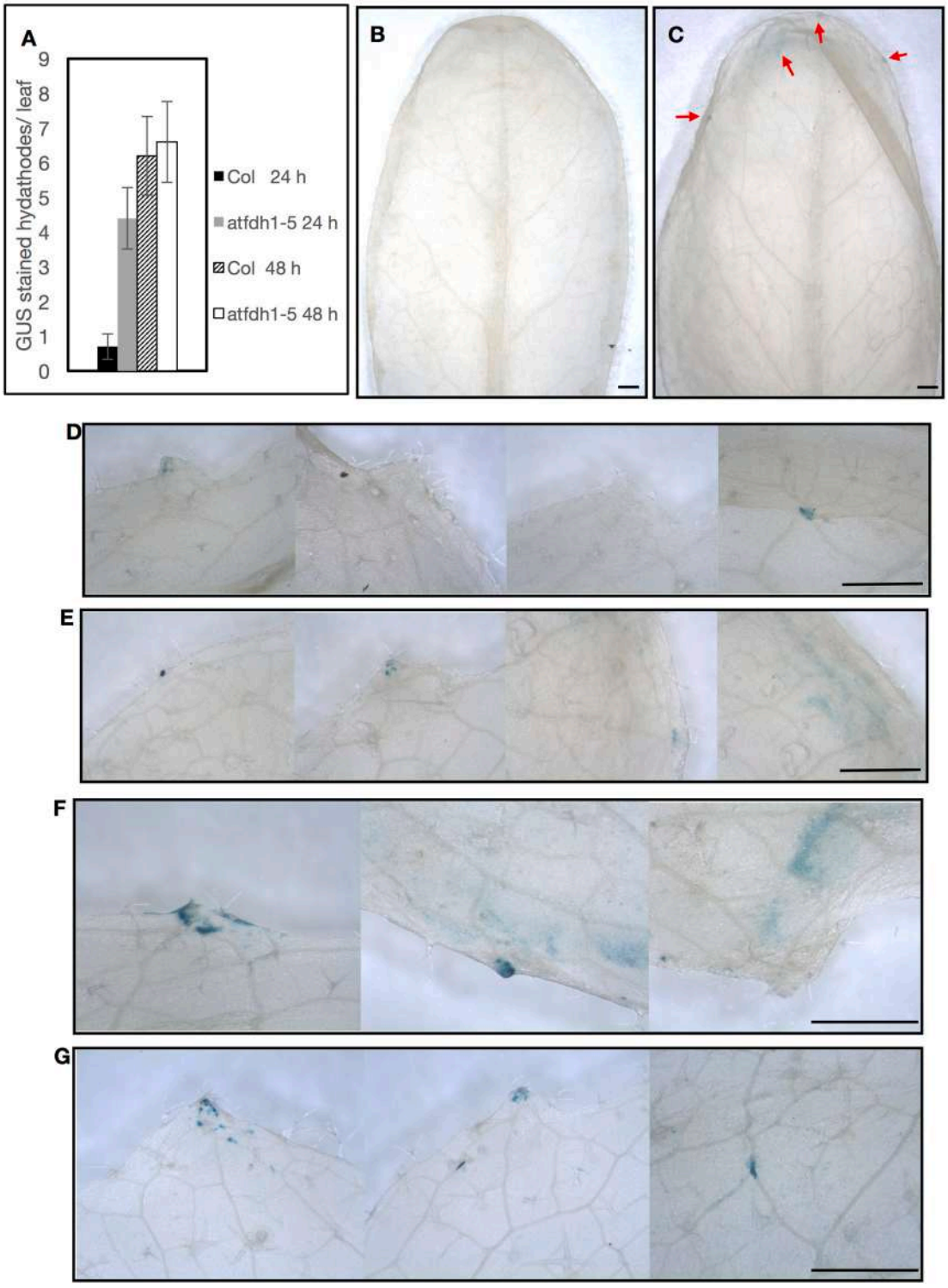


Fig. 3. Infection of wt Col and *atfdh1-5* leaves with *Xcc* 8004::GUS-GFP by constant submersion in 10^8 cfu ml^{-1} bacterial suspension. Leaves detached from wt Col and *atfdh1-5* rosettes were infected with *Xcc* 8004::GUS-GFP and stained for GUS activity after 24 h or 48 h from infection. **(A)** number of hydathodes, in wt Col and *atfdh1-5* leaves, in which the presence of *Xcc* 8004::GUS-GFP could be detected after 24 h or 48 h from infection. Bars represent mean values ± SE from 10 leaves each. (**B, C**) wt Col and *atfdh1-5* whole leaves after 24 h from infection; arrows point to stained hydathodes. **(D, E)** wt Col and *atfdh1-5* hydathodes and their surroundings after 24 h from infection, respectively. **(F, G)** wt Col and *atfdh1-5* hydathodes and their surroundings after 48 h from infection, respectively. Scale bars in each panel represent 1 mm. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

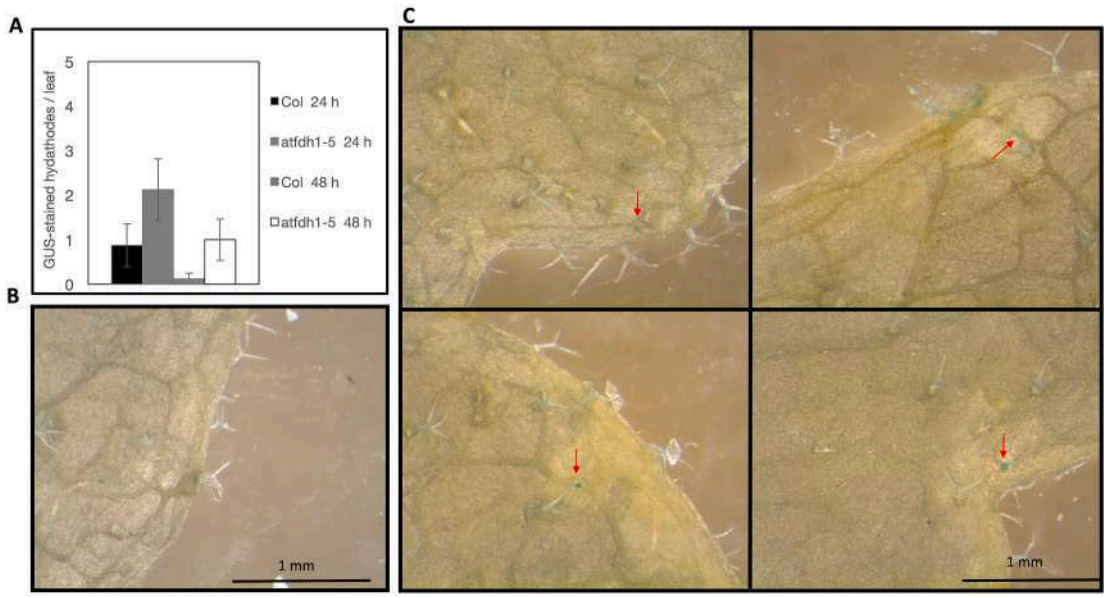


Fig. 4. Infection of *A. thaliana* wt Col and *atfdh1-5* mutant leaves with 10^8 cfu ml^{-1} *Xcc* 8004::GUS-GFP by 10 min submersion. **(A)** number of hydathodes, in wt Col and *atfdh1-5* leaves, in which the presence of *Xcc* 8004::GUS-GFP could be detected, after 24 h or 48 h from infection. Bars represent mean values ± SE from 8 leaves each. Details of hydathodes and their surroundings in **(B)** wt Col and **(D)** *atfdh1-5* leaves after 48 h from infection. Scale bar represents 1 mm. Red arrows point at the spreading of bacteria below hydathodes. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

is observed not only with 10^8 cfu ml^{-1} *Xcc* but also after infection with a diluted bacterial suspension, i.e. 10^4 cfu ml^{-1} *Xcc* (Fig. 2). Accordingly, the total number of leaves in which staining for GUS activity could be detected was lower in *Xcc*-infected leaves than corresponding controls (Supplementary Fig. S4). After 48 h from infection, no differences were observed in the number of hydathodes stained for GUS activity between infected and non-infected leaves; nonetheless, the total number of leaves in which staining for GUS activity could be detected was, again, lower in *Xcc*-infected leaves when compared to corresponding controls (Supplementary Fig. S4).

3.3. Proliferation and spread of Xcc in hydathodes of A.thaliana atfdh1-5 mutant

The physiological relevance of FDH to counteract pathogen attack by *Xcc* was then explored by using *atfdh1-5* mutant KO in the FDH gene [6, 8]. Rosettes' upper leaves were detached from wt Col and *atfdh1-5* plants, and infected by constant submersion in 10^8 cfu ml^{-1} *Xcc* 8004:: GUS-GFP suspension; this *Xcc* strain allows the visualization of the bacterial spread by GUS staining [12]. After 24 h from the infection, the number of hydathodes in which *Xcc* bacteria could be detected is higher in *atfdh1-5* leaves than in wt (Fig. 3A). Seven out of ten inspected wt leaves do not show any staining for *Xcc* (Fig. 3B) and the few stained hydathodes in the remaining wt leaves do not show any sign of infection outside the parenchyma (Fig. 3D). On the contrary, all the ten inspected *atfdh1-5* leaves show staining in hydathodes (Fig. 3C); moreover, *Xcc* infection occasionally spreads also to surrounding tissues (Fig. 2E). After 48 h, the number of stained hydathodes is similar in wt and *atfdh1-5* leaves (Fig. 3A), but not the pattern of the infection: the occasional spread of *Xcc* occurs in the parenchymatic leaf tissue in wt leaves (Fig. 3F), whereas such spread is occurring through the vascular system in *atfdh1-5* leaves (Fig. 3G).i

To confirm these observations, rosettes' upper leaves were detached from an independent batch of wt Col and *atfdh1-5* plants (Supplementary Fig. S1) and infected by 10 min submersion in 10^8 cfu ml^{-1} *Xcc* 8004::GUS-GFP bacterial suspension; infected Col and *atfdh1-5* leaves were then maintained either 24 h or 48 h on moist paper in Petri dishes before GUS staining, as previously described (Supplementary Fig.S2). Again, after 24 h from the infection, the number of hydathodes in which *Xcc* bacteria could be detected is higher in *atfdh1-5* leaves than in wt (Fig. 4A); moreover, the occasional detection of *Xcc* is restricted to hydathodes in Col leaves (Supplementary Fig. S5), whereas the bacteria could be detected in hydathodes, as well as in their proximities already, in *atfdh1-5* leaves (Supplementary Fig. S6). After 48 h from infection, the number of hydathodes in which bacteria could be GUS-stained was the same in Col and *atfdh1-5* (Fig. 4A); however, *Xcc* is detected only within hydathodes in Col leaves (Fig. 4B), whereas it is also detected outside the hydathodes in *atfdh1-5* leaves (Fig. 4C).

4. Discussion

The involvement of FDH in plant's responses to environmental stimuli has been demonstrated by several authors ([8] and references therein), whereas the evidences for FDH involvement during biotic stress are limited to few studies [6]. This work aimed at providing new insight on the FDH engagement and its role in the responses activated by plants exposed to pathogens.

We first adopted transcript correlation analysis, a powerful method which favors the identification of novel genes involved in a given metabolic pathway, as occurred for CYP82C4 enzyme catalyzing the last biosynthetic step of the redox-active sideretin, released from roots as a general strategy for Fe acquisition [49]. CYP82C4 was indeed proposed as a novel *A. thaliana* gene involved in Fe metabolism thanks to such *in silico* approach [50].

More recently, we adopted this strategy to study FDH involvement during abiotic stress [8]; that correlation analysis was based on a large dataset containing many different experiments (for instance, the AtGenExpress developmental series, different growth conditions, hormone treatments, abiotic and biotic stresses), corresponding to around 1700 hybridizations, described in detail previously [51]. In the present work, we focused on the *in silico* analysis of FDH expression during plant responses to biotic stress and several novel correlators could be unveiled. Moreover, the FDH correlation network, obtained by exploiting the biotic stress dataset, shows that FDH expression is also correlated with genes engaged during senescence, degradation of some molecules (i.e. chlorophyll) and remobilization of nutrients. Along with our previous studies [8], such a gene correlation network confirms that FDH is linked to a wide range of metabolic pathways modulated by stress.

Hydathodes are characterized by a parenchyma made of thin-walled small cells, various meatuses and xylem vessels irrigating them [13], and act as a sort of "release valve": hydathodes modulate fluid fluxes and protect against flooding of leaf tissues, which would compromise gas exchange and photosynthetic activity [13,52]. Hydathodes, however, can also represent an entry point for pathogens, as it occurs in *A. thaliana* leaves exposed to *Xcc* [11,12]. The localization of FDH expression in hydathodes, as previously described by our research group [8], prompted us to investigate FDH role in response to *Xcc* infection.

We showed that FDH promoter activity is inhibited in *Vu* FDH::GUS leaves infected with *Xcc*; such inhibition could be observed not only in detached leaves constantly submerged with a pathogen suspension, but also when such submersion lasted only 10 min, and also when the bacterial suspension was diluted 10^4 fold. Such a result suggests that a signaling event is taking place, and that inhibition cannot be solely attributed to a sort of "by-pass" of the resistance of *A. thaliana* ecotype Col against *Xcc* due to a prolonged exposure to the a concentrated bacterial suspension.

Indeed, *atfdh1-5* mutant appears more sensitive to *Xcc* attack than the corresponding wt, as suggested by a higher number of infected hydathodes in *atfdh1-5* leaves than in wt counterpart after 24 h from infection: not only, we could also observe the spread of *Xcc* outside the *atfdh1-5* hydathodes. Again, the latter result could also be observed when the leaves were directly exposed to *Xcc* suspension for 10 min only.

Altogether, these results suggest that FDH takes part in the early response against *Xcc* infection: FDH regulation seems to reduce pathogen's entry via hydathodes and, at later time points, its proliferation in the surrounding tissues. We therefore propose that formate might act as a signaling molecule in such an early response against *Xcc* infection: a decrease of FDH expression is expected to lead to a local increase in formate levels in the hydathodes where pathogens enter, in support of a plant defense response. This defense response is attenuated in *atfdh* 1–5 mutant, as the observed proliferation of *Xcc* in a higher number of hydathodes in *atfdh1-5* leaves would suggest.

Fine changes in formate levels in hydathodes, modulated by FDH, might be important for the signaling cascade during the early pathogen response, given the relevance of FDH on Fe and Mo homeostasis [8]. Future analyses to be conducted under various Mo and Fe nutritional conditions might help clarify FDH and formate roles during *Xcc* infection and might assist the elucidation of signaling cascades activated during stress conditions easily co-occurring in the field, such as nutritional stress and pathogen attack, but hardly explored at the same time in laboratory conditions.

Author statement

Irene Murgia and Piero Morandini conceived the work and planned the initial experiments.

Francesca Marzorati performed the bioinformatic analysis and the *in vivo* experiments.

Francesca Marzorati, Gianpiero Vigani, Piero Morandini and Irene Murgia analyzed the results.

Irene Murgia wrote the manuscript, with contributions of Francesca Marzorati, Piero Morandini and Gianpiero Vigani.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We are grateful to Prof. Jian Li Yang for donating *A. thaliana* line *V. umbellata* FDH::GUS; to Dr. Laurent Nöel and collaborators for donating *Xcc* 8004 *and Xcc* 8004::GUS-GFP strains, for suggestions and useful discussion. *atfdh1-5* line Salk-108751 was received by Salk institute. This research was supported by local research funds of the Department of Life Sciences and Systems Biology (University of Turin) and the Department of Environmental Science and Policy (University of Milan).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pmpp.2021.101633.

References

[1] I. Diaz, Plant defense genes against biotic stresses, Int. J. Mol. Sci. 19 (2018) 1–5, https://doi.org/10.3390/ijms19082446.
[2] V. Popov, V. Lamzin, NAD+-dependent formate dehydrogenase, Biochem. J. 301 (1994) 625–643, https://doi.org/10.1177/107839039500100302.
[3] A.A. Alekseeva, S.S. Savin, V.I. Tishkov, NAD+-dependent formate dehydrogenase from plants, Acta Naturae 3 (2011) 38–54, https://doi.org/10.32607/20758251-2011-3-4-38-54.
[4] A.D. Hanson, S. Roje, One-Carbon metabolism In higher plants, Annu. Rev. Plant Physiol. Plant Mol. Biol. 52 (2001) 119–156, https://doi.org/10.1146/annurev.arplant.52.1.119.
[5] C. Colas Des Francs-Small, F. Ambard-Bretteville, I.D. Small, R. Rémy, Identification of a major soluble protein in mitochondria from nonphotosynthetic tissues as NAD-dependent formate dehydrogenase, Plant Physiol. 102 (1993) 1171–1177, https://doi.org/10.1104/pp.102.4.1171.
[6] D.S. Choi, D.S. Choi, B.K. Hwang, Pepper mitochondrial formate dehydrogenase1 Regulates cell death and defense responses against bacterial pathogens, Plant Physiol. 166 (2014) 1298–1311, https://doi.org/10.1104/pp.114.246736.
[7] H.Q. Lou, Y.L. Gong, W. Fan, J.M. Xu, Y. Liu, M.J. Cao, M.H. Wang, J.L. Yang, S. J. Zheng, A formate dehydrogenase confers tolerance to aluminum and low pH, Plant Physiol. 171 (2016) 294–305, https://doi.org/10.1104/pp.16.01105.
[8] I. Murgia, G. Vigani, D. Di Silvestre, P. Mauri, R. Rossi, A. Bergamaschi, M. Frisella, P. Morandini, Formate dehydrogenase takes part in molybdenum and iron homeostasis and affects dark-induced senescence in plants, J. Plant Interact. 15 (2020) 386–397, https://doi.org/10.1080/17429145.2020.1836273.
[9] G. Vigani, D. Di Silvestre, A.M. Agresta, S. Donnini, P. Mauri, C. Gehl, F. Bittner, I. Murgia, Molybdenum and iron mutually impact their homeostasis in cucumber (Cucumis sativus) plants, New Phytol. 213 (2017) 1222–1241, https://doi.org/10.1111/nph.14214.
[10] D. Di Silvestre, G. Vigani, P. Mauri, S. Hammadi, P. Morandini, I. Murgia, Network topological analysis for the identification of novel hubs in plant nutrition, Front. Plant Sci. (2021), https://doi.org/10.3389/fpls.2021.629013.
[11] V. Hugouvieux, C.E. Barber, M.J. Daniels, Entry of Xanthomonas campestris pv. campestris into hydathodes of Arabidopsis thaliana leaves: a system for studying early infection events in bacterial pathogenesis, Mol. Plant Microbe Interact. 11 (1998) 537–543, https://doi.org/10.1094/MPMI.1998.11.6.537.
[12] A. Cerutti, A. Jauneau, M.C. Auriac, E. Lauber, Y. Martinez, S. Chiarenza, N. Leonhardt, R. Berthomé, L.D. Noël, Immunity at cauliflower hydathodes controls systemic infection by Xanthomonas campestris pv campestris, Plant Physiol. 174 (2017) 700–716, https://doi.org/10.1104/pp.16.01852.
[13] A. Cerutti, A. Jauneau, P. Laufs, N. Leonhardt, M.H. Schattat, R. Berthomé, J.-M. Routaboul, L.D. Noël, Mangroves in the leaves: anatomy, physiology, and immunity of epithemal hydathodes, Annu. Rev. Phytopathol. 57 (2019) 91–116, https://doi.org/10.1146/annurev-phyto-082718-100228.
[14] R.C. Team, R: A Language and Environment for Statistical Computing, 2020. https://www.r-project.org/.
[15] L. Gautier, L. Cope, B.M. Bolstad, R.A. Irizarry, Affy - analysis of affymetrix GeneChip data at the probe level, Bioinformatics 20 (2004) 307–315, https://doi.org/10.1093/bioinformatics/btg405.
[16] D. Sean, P.S. Meltzer, GEOquery, A bridge between the gene expression Omnibus (GEO) and BioConductor, Bioinformatics 23 (2007) 1846–1847, https://doi.org/10.1093/bioinformatics/btm254.
[17] R. Edgar, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, Nucleic Acids Res. 30 (2002) 207–210, https://doi.org/10.1093/nar/30.1.207.
[18] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K. A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C.L. Robertson, N. Serova, S. Davis, A. Soboleva, NCBI GEO: archive for functional genomics data sets - Update, Nucleic Acids Res. 41 (2013) 991–995, https://doi.org/10.1093/nar/gks1193.
[19] U. Raudvere, L. Kolberg, I. Kuzmin, T. Arak, P. Adler, H. Peterson, J. Vilo, A web server for functional enrichment analysis and conversions of gene lists (2019 update), Nucleic Acids Res. 47 (2019) W191–W198, https://doi.org/10.1093/nar/gkz369. G:Profiler.
[20] V. Krishnakumar, M.R. Hanlon, S. Contrino, E.S. Ferlanti, S. Karamycheva, M. Kim, B.D. Rosen, C.Y. Cheng, W. Moreira, S.A. Mock, J. Stubbs, J.M. Sullivan, K. Krampis, J.R. Miller, G. Micklem, M. Vaughn, C.D. Town, Araport: the arabidopsis information portal, Nucleic Acids Res. 43 (2015) D1003–D1009, https://doi.org/10.1093/nar/gku1200.

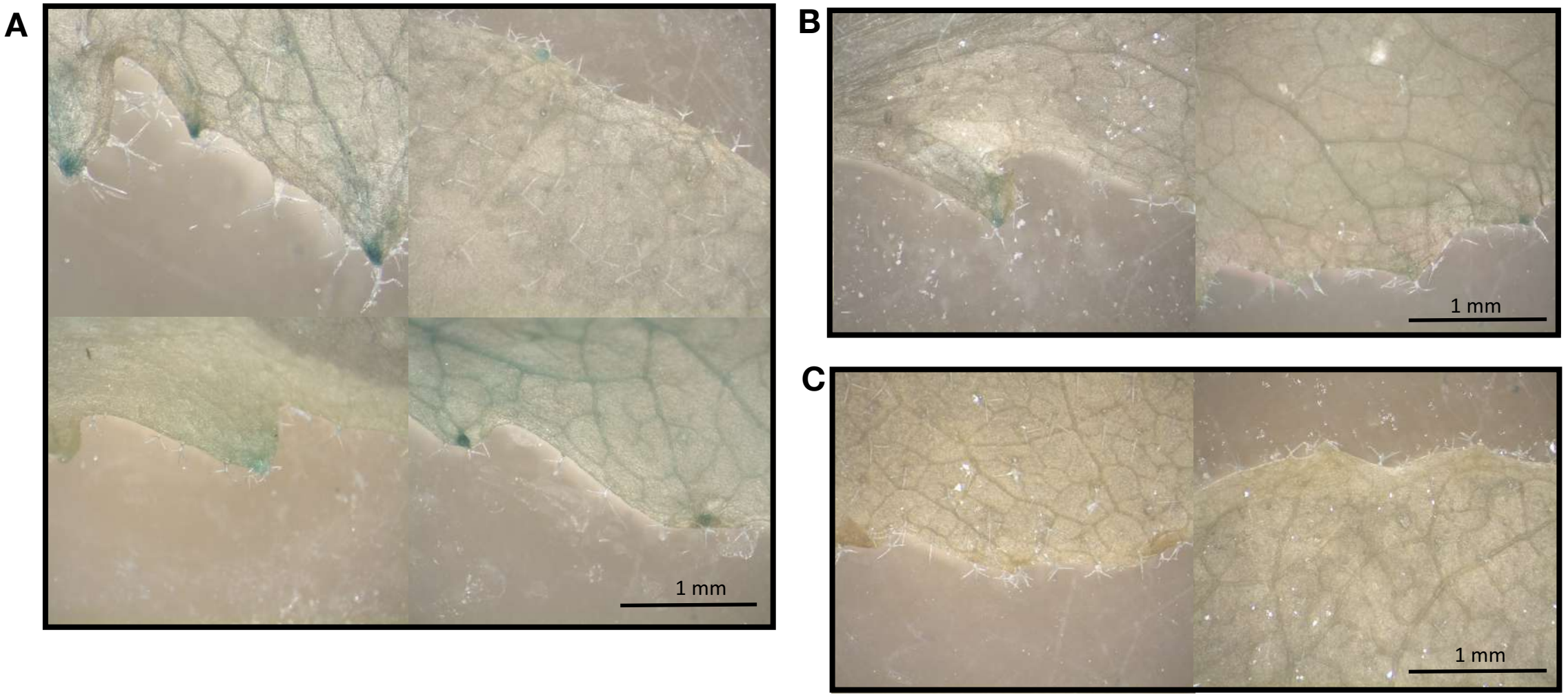
[21] R. Schwacke, A. Schneider, E. Van Der Graaff, K. Fischer, E. Catoni, M. Desimone, W.B. Frommer, U.I. Flügge, R. Kunze, Aramemnon, a novel database for Arabidopsis integral membrane proteins, Plant Physiol. 131 (2003) 16–26, https://doi.org/10.1104/pp.011577.
[22] A. Elorza, G. León, I. Gómez, A. Mouras, L. Holuigue, A. Araya, X. Jordana, Nuclear SDH2-1 and SDH2-2 genes, encoding the iron-sulfur subunit of mitochondrial complex II in arabidopsis, have distinct cell-specific expression patterns and promoter activities, Plant Physiol. 136 (2004) 4072–4087, https://doi.org/10.1104/pp.104.049528.
[23] M. van Hulten, S. Chatterjee, B. van den, A. Harrold, Infection assay for Xanthomonas campestris pv. campestris in Arabidopsis thaliana mimicking natural entry via hydathodes, in: Plant Innate Immunity, Methods Protoc., 2019, pp. 159–185, https://doi.org/10.1007/978-1-4939-9458-8_16.
[24] S. Islam, T.M. Leissing, R. Chowdhury, R.J. Hopkinson, C.J. Schofield, 2-Oxoglutarate-Dependent oxygenases, Annu. Rev. Biochem. 87 (2018) 1–36.
[25] G. Vigani, P. Morandini, I. Murgia, Searching iron sensors in plants by exploring the link among 2′-OG-dependent dioxygenases, the iron deficiency response and metabolic adjustments occurring under iron deficiency, Front. Plant Sci. 4 (2013) 1–7, https://doi.org/10.3389/fpls.2013.00169.
[26] S. Yadav, A. Mishra, Ectopic expression of C4 photosynthetic pathway genes improves carbon assimilation and alleviate stress tolerance for future climate change, Physiol. Mol. Biol. Plants 26 (2020) 195–209, https://doi.org/10.1007/s12298-019-00751-8.
[27] M. Kumar, D.T. Le, S. Hwang, P.J. Seo, H.U. Kim, Role of the INDETERMINATE DOMAIN genes in plants, Int. J. Mol. Sci. 20 (2019), https://doi.org/10.3390/ijms20092286.
[28] M. Houben, B. Van de Poel, 1-aminocyclopropane-1-carboxylic acid oxidase (ACO): the enzyme that makes the plant hormone ethylene, Front. Plant Sci. 10 (2019) 1–15, https://doi.org/10.3389/fpls.2019.00695.
[29] S. Hörtensteiner, Update on the biochemistry of chlorophyll breakdown, Plant Mol. Biol. 82 (2013) 505–517, https://doi.org/10.1007/s11103-012-9940-z.
[30] S. Aubry, N. Fankhauser, S. Ovinnikov, A. Pruzinská, M. Stirnemann, K. Zienkiewicz, C. Herrfurth, I. Feussner, S. Hörtensteiner, Pheophorbide A may regulate jasmonate signaling during DARK-INDUCED SENESCENCE, Plant Physiol. 182 (2020) 776–791, https://doi.org/10.1104/pp.19.01115.
[31] H.B. Tugal, M. Pool, A. Baker, Arabidopsis 22-kilodalton peroxisomal membrane protein. Nucleotide sequence analysis and biochemical characterization, Plant Physiol. 120 (1999) 309–320, https://doi.org/10.1104/pp.120.1.309.
[32] K. Takanashi, N. Shitan, K. Yazaki, The multidrug and toxic compound extrusion (MATE) family in plants, Plant Biotechnol. 31 (2014) 417–430, https://doi.org/10.5511/plantbiotechnology.14.0904a.
[33] J. Zhao, R.A. Dixon, MATE transporters facilitate vacuolar uptake of epicatechin 3′-O-glucoside for proanthocyanidin biosynthesis in medicago truncatula and arabidopsis, Plant Cell 21 (2009) 2323–2340, https://doi.org/10.1105/tpc.109.067819.
[34] J. Park, Y. Lee, E. Martinoia, M. Geisler, Plant hormone transporters: what we know and what we would like to know, BMC Biol. 15 (2017) 1–15, https://doi.org/10.1186/s12915-017-0443-x.
[35] J.L. Carpenter, S.E. Ploense, D.P. Snustad, C.D. Silflow, Preferential expression of an α-tubulin gene of arabidopsis in pollen, Plant Cell 4 (1992) 557–571, https://doi.org/10.2307/3869555.
[36] T. Hackenberg, T. Juul, A. Auzina, S. Gwizdz, A. Małolepszy, K. Van Der Kelen, S. Dam, S. Bressendorff, A. Lorentzen, P. Roepstorff, K.L. Nielsen, J.E. Jørgensen, D. Hofius, F. Van Breusegem, M. Petersen, S.U. Andersen, Catalase and NO CATALASE ACTIVITY1 promote autophagy-dependent cell death in Arabidopsis, Plant Cell 25 (2013) 4616–4626, https://doi.org/10.1105/tpc.113.117192.
[37] J. Li, J. Liu, G. Wang, J.Y. Cha, G. Li, S. Chen, Z. Li, J. Guo, C. Zhang, Y. Yang, W. Y. Kim, D.J. Yun, K.S. Schumaker, Z. Chen, Y. Guo, A chaperone function of NO CATALASE ACTIVITY1 Is required to maintain catalase activity and for multiple stress responses in arabidopsis, Plant Cell 27 (2015) 908–925, https://doi.org/10.1105/tpc.114.135095.
[38] S. Schelbert, S. Aubry, B. Burla, B. Agne, F. Kessler, K. Krupinska, S. Hörtensteiner, Pheophytin pheophorbide hydrolase (pheophytinase) is involved in chlorophyll breakdown during Leaf senescence in Arabidopsis, Plant Cell 21 (2009) 767–785, https://doi.org/10.1105/tpc.108.064089.
[39] W. Zhang, T. Liu, G. Ren, S. HÉrtensteiner, Y. Zhou, E.B. Cahoon, C. Zhang, Chlorophyll degradation: the tocopherol Biosynthesis-Related phytol hydrolase in arabidopsis seeds is still missing, Plant Physiol. 166 (2014) 70–79, https://doi.org/10.1104/pp.114.243709.
[40] C.A. ten Hove, Z. Bochdanovits, V.M.A. Jansweijer, F.G. Koning, L. Berke, G. F. Sanchez-Perez, B. Scheres, R. Heidstra, Probing the roles of LRR RLK genes in Arabidopsis thaliana roots using a custom T-DNA insertion set, Plant Mol. Biol. 76 (2011) 69–83, https://doi.org/10.1007/s11103-011-9769-x.
[41] X. Li, A. Salman, C. Guo, J. Yu, S. Cao, X. Gao, W. Li, H. Li, Y. Guo, Identification and characterization of LRR-RLK family genes in potato reveal their involvement in peptide signaling of cell fate decisions and biotic/abiotic stress responses, Cells 7 (2018) 120, https://doi.org/10.3390/cells7090120.
[42] E.A. Van Der Biezen, J. Sun, M.J. Coleman, M.J. Bibb, J.D.G. Jones, Arabidopsis RelA/SpoT homologs implicate (p)ppGpp in plant signaling, Proc. Natl. Acad. Sci. U. S. A 97 (2000) 3747–3752, https://doi.org/10.1073/pnas.97.7.3747.
[43] J. Boniecka, J. Prusińska, G.B. Dąbrowska, A. Goc, Within and beyond the stringent response-RSH and (p)ppGpp in plants, Planta 246 (2017) 817–842, https://doi.org/10.1007/s00425-017-2780-y.
[44] K. Mizusawa, S. Masuda, H. Ohta, Expression profiling of four RelA/SpoT-like proteins, homologues of bacterial stringent factors, in Arabidopsis thaliana, Planta 228 (2008) 553–562, https://doi.org/10.1007/s00425-008-0758-5.
[45] M. Sugliani, H. Abdelkefi, H. Ke, E. Bouveret, C. Robaglia, S. Caffarri, B. Field, An ancient bacterial signaling pathway regulates chloroplast function to influence growth and development in Arabidopsis, Plant Cell 28 (2016) 661–679, https://doi.org/10.1105/tpc.16.00045.
[46] G. Griffiths, Biosynthesis and analysis of plant oxylipins, Free Radic. Res. 49 (2015) 565–582, https://doi.org/10.3109/10715762.2014.1000318.
[47] A. Kachroo, P. Kachroo, Fatty acid–derived signals in plant defense, Annu. Rev. Phytopathol. 47 (2009) 153–176, https://doi.org/10.1146/annurev-phyto-080508-081820.
[48] D. Ellinger, N. Stingl, I.I. Kubigsteltig, T. Bals, M. Juenger, S. Pollmann, S. Berger, D. Schuenemann, M.J. Mueller, DONGLE and DEFECTIVE IN ANTHER DEHISCENCE1 lipases are not essential for wound- and pathogen-induced jasmonate biosynthesis: redundant lipases contribute to jasmonate formation, Plant Physiol. 153 (2010) 114–127, https://doi.org/10.1104/pp.110.155093.
[49] J. Rajniak, R.F.H. Giehl, E. Chang, I. Murgia, N. Von Wirén, E.S. Sattely, Biosynthesis of redox-active metabolites in response to iron deficiency in plants, Nat. Chem. Biol. 14 (2018) 442–450, https://doi.org/10.1038/s41589-018-0019-2.
[50] I. Murgia, D. Tarantino, C. Soave, P. Morandini, Arabidopsis CYP82C4 expression is dependent on Fe availability and circadian rhythm , and correlates with genes involved in the early Fe deficiency response, J. Plant Physiol. 168 (2011) 894–902, https://doi.org/10.1016/j.jplph.2010.11.020.
[51] M. Menges, R. Dóczi, L. Ökrész, P. Morandini, L. Mizzi, M. Soloviev, J.A.H. Murray, L. Bögre, Comprehensive gene expression atlas for the Arabidopsis MAP kinase signalling pathways, New Phytol. 179 (2008) 643–662, https://doi.org/10.1111/j.1469-8137.2008.02552.x.
[52] Y.C. Chen, T.C. Lin, C.E. Martin, Effects of guttation prevention on photosynthesis and transpiration in leaves of Alchemilla mollis, Photosynthetica 52 (2014) 371–376, https://doi.org/10.1007/s11099-014-0040-y.

Supplementary Figure S1

A.thaliana Vu FDH::GUS, *atfdh1-5* and Col plants, grown for 5 weeks at 23°C, 150 mE m^2sec^{-1}, 12h/12 h light /dark photoperiod. Red bar represents 5 cm.

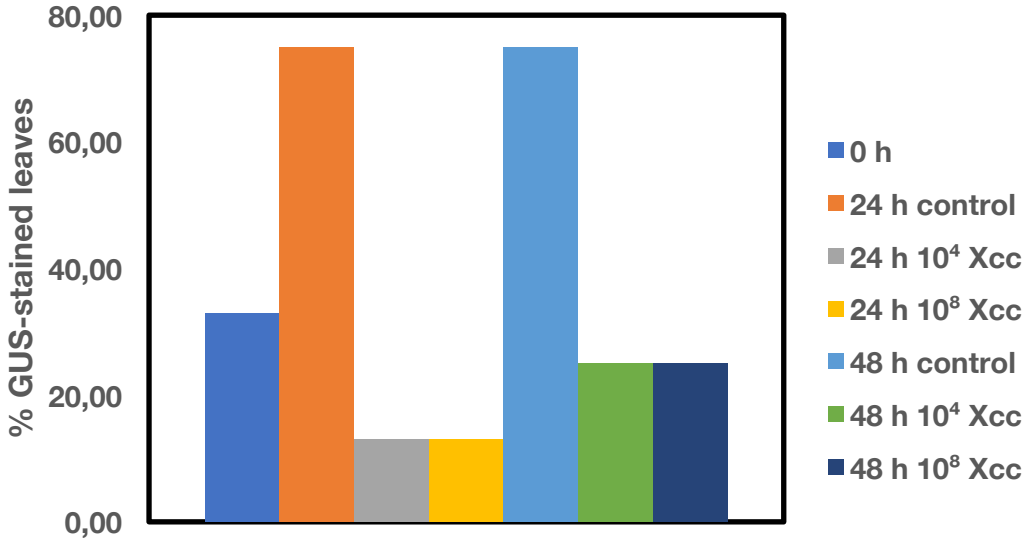
Supplementary Figure S2

Infection of *A.thaliana Vu* FDH::GUS leaves with *Xcc* 8004. Leaves detached from *A.thaliana Vu* FDH::GUS plants were either mock treated, or infected with 10^8 cfu ml^{-1} or 10^4 cfu ml^{-1} *Xcc* 8004 suspension, for 10 min; leaves were then maintained up to 48 h onto moist paper, into sealed square (10x10 cm) Petri dishes. **(A)** Various Petri dishes, each containing infected leaves. **(B)** Detail of leaves in the Petri dish, after 48 h.



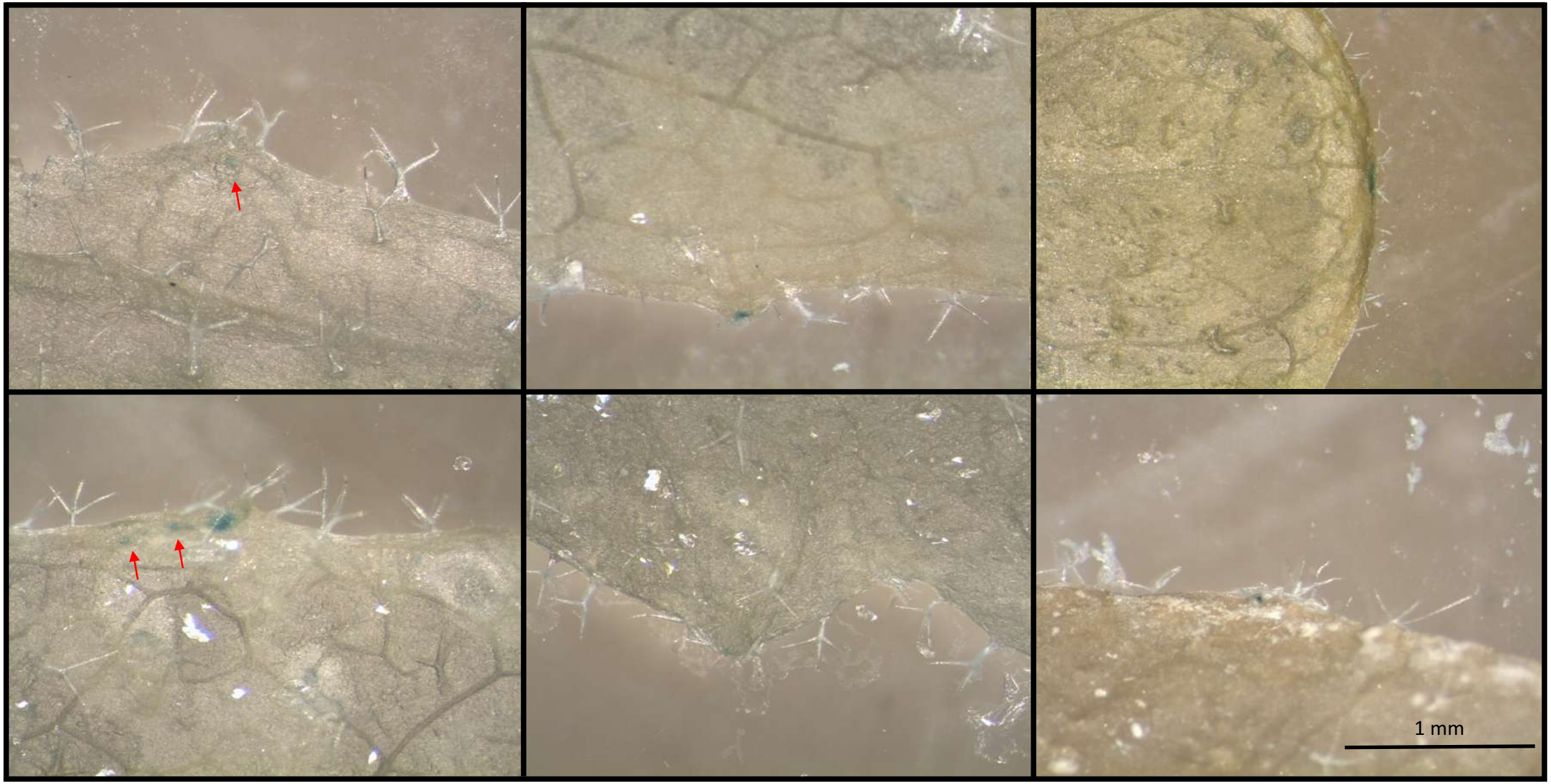
Supplementary Figure S3

Activity of FDH promoter *in A.thaliana Vu* FDH::GUS leaves infected with *Xcc* 8004. Leaves detached from *A.thaliana Vu* FDH::GUS plants were **(A)** mock treated, or infected with **(B)** 10^8 cfu ml^{-1} *Xcc* 8004 suspension or with **(C)** 10^4 cfu ml^{-1} *Xcc* 8004 suspension, for 10 min; after 24 h leaves were stained for GUS activity; representative hydathodes are shown, for each treatment. 1 mm scale bars are reported, in each panel.



Supplementary Figure S4

Percentage *A.thaliana Vu* FDH::GUS leaves showing GUS staining after infection with *Xcc*. Leaves detached from 5 weeks old *Vu* FDH::GUS plants were stained for GUS activity before infection (0 h), after 24 h or 48 h mock-treatment (24 h control, 48 h control) and after 24 h or 48 h infection with either 10^8 or 10^4 cfm ml^{-1} *Xcc*. 8 leaves were analysed for each condition.



Supplementary Figure S6

Infection of *A.thaliana atfdh1-5* leaves with 10^8 cfu ml^{-1} *Xcc* 8004::GUS-GFP, by 10 min submersion. After infection, leaves were maintained on moist paper in sealed Petri dishes and stained for GUS activity after 24 h. Scale bar represents 1 mm. Red arrows point at the spreading of bacteria below the hydathodes.

Supplementary Figure S5

Infection of *A.thaliana* wt Col leaves with 10^8 cfu ml^{-1} *Xcc* 8004::GUS-GFP, by 10 min submersion. After infection, leaves were maintained on moist paper in sealed Petri dishes and stained for GUS activity after 24 h. Scale bar represents 1 mm.

AGI code	Gene name	Pearson coefficient (Lin)	Pearson coefficient (Log)	Localization	Transmembrane domain
AT5G14780	**Formate dehydrogenase FDH**	**1**	**1**	**mitochondrion, chloroplast**	**no**
AT3G51430	YLS2	0.7776	0.7426	secr. pathway	yes
AT2G26740	SEH	0.7689	0.5979	unclear	yes
AT1G01490	HMP01	0.7614	0.6244	nucleus	no
AT4G20930	HDH1	0.7472	0.6675	mitoch.	yes
AT5G13800	PPH	0.7468	0.7629	chloropl., secr. pathway	yes
AT5G52450	DTX16	0.7459	0.7713	secr. pathway	yes
AT5G05930	GC1	0.7458	0.6494	secr. pathway	yes
AT1G55110	IDD7	0.7428	0.7997	nucleus	no
AT3G11930	AT3G11930	0.7425	0.6580	unclear	no
AT1G58180	BCA6	0.7388	0.6815	mitoch.	no
AT3G61440	CYSC1	0.7357	0.6431	chloropl., mitoch.	yes
AT3G44880	PAO	0.7285	0.7779	chloropl.	yes
AT4G22260	IM	0.7281	0.7139	chloropl., mitoch.	yes
AT4G39850	PXA1	0.7211	0.6669	chloropl, mitoch.	yes

Supplementary Table S4

Top correlators of *A.thaliana* FDH transcript under biotic stress, in linear analysis. The genes are listed according to their respective Pearson correlation coefficients from linear analysis; Pearson correlation coefficients from log analysis are also provided. Putative intracellular localization and presence of transmembrane domains are according to Aramemnon database (http://aramemnon.uni-koeln.de/) (Schwacke *et al.*, 2003). Genes highlighted in yellow background are also top correlators of the Log analysis (see Table 1).

Manuscript 4

Arabidopsis thaliana early foliar proteome response to root exposure to the rhizobacterium *Pseudomonas simiae* WCS417

Francesca Marzorati[1], Rossana Rossi[2], Letizia Bernardo[2], PierLuigi Mauri[2], Dario Di Silvestre[2], Piero Morandini[1], Irene Murgia[1*]

[1] Department of Environmental Science and Policy, University of Milano, Milano, Italy
[2] Proteomic and Metabolomic Laboratory, Institute for Biomedical Technologies- National Research Council (ITB-CNR), Segrate, Italy
[*] Correspondence: irene.murgia@unimi.it

Abstract

Pseudomonas simiae WCS417 is a plant growth-promoting rhizobacterium that improves plant health and development. In this study, we investigate the early leaf responses of *Arabidopsis thaliana* to WCS417 exposure and the possible involvement of formate dehydrogenase (FDH) in such responses. *In vitro*-grown *A. thaliana* seedlings expressing a FDH::GUS reporter show a significant increase in FDH promoter activity in their roots and shoots, after seven days of indirect exposure (without contact) to WCS417. After root exposure to WCS417, also the leaves of FDH::GUS plants grown in soil show increased FDH promoter activity in their hydathodes. To elucidate early foliar responses to WCS417, as well as FDH involvement, the roots of *A. thaliana* wt Col and *atfdh1-5* knock-out mutant plants grown in soil were exposed to WCS417 and proteins were extracted from rosette leaves for proteomic analysis. The proteomic results reveal that chloroplasts, in particular some components of the photosystems PSI and PSII, as well as the Glutathione S-transferase GST family, are among the early targets of the metabolic changes induced by WCS417. Moreover, the alterations in the foliar proteome, as observed in the *atfdh1-5* mutant, especially after exposure to WCS417 and involving stress-responsive genes, allow to propose FDH as a relevant node in the early events triggered by the interactions between *A. thaliana* and the rhizobacterium WCS417.

Keywords: *Arabidopsis thaliana*; formate dehydrogenase FDH; glutathione -S-transferases GST; hydathodes; proteome; *Pseudomonas simiae* WCS417; rhizobacterium

Introduction

Plant growth-promoting rhizobacteria (PGPR) can enhance plant development and defense through their antagonist actions against soil plant pathogens and the activation of the Induced Systemic Resistance (ISR) response (Wang et al. 2021). *Pseudomonas* is a competitive bacterial genus in the rhizosphere (Simons et al. 1996; de Weert et al. 2002) and, in particular, its species *simiae* WCS417 (also known as *Pseudomonas fluorescens* WCS417 or simply WCS417) is one of the most characterized PGPR for ISR induction (Pieterse et al. 2020). The molecular basis of ISR has been thoroughly investigated in *Arabidopsis thaliana* roots colonized by WCS417 (Stringlis et al. 2018; Zamioudis et al. 2014); ISR response partially overlaps with the iron (Fe) deficiency response in *A. thaliana* (Romera et al. 2019). Volatile organic compounds (VOCs) and Fe-chelating siderophores produced by PGPR trigger plant Fe uptake pathways (Trapet et al. 2021; Verbon et al. 2019); plants suffering from Fe deficiency may recruit more siderophore-producing bacteria than plants growing under normal nutritional conditions (Jin et al. 2006, 2010), and VOCs generated by ISR-inducing bacteria may relieve nutritional stress induced by low Fe levels (Zamioudis et al. 2015; Zhang et al. 2009). In addition, WCS417 promotes the expression of plant genes activated by low Fe levels, such as MYB72 (Palmer et al. 2013; Zamioudis et al. 2015). Formate dehydrogenase enzymes (FDHs) are found in bacteria, fungi and plants (Alekseeva et al. 2011); plant FDHs, localized in mitochondria (Choi et al. 2014; Herman et al. 2002) and chloroplasts (Lee et al. 2022; Olson et al. 2000), catalyze the oxidation of formate ($HCOO^-$) to CO_2 with the reduction of NAD^+ to NADH. FDHs are referred to as 'stress enzymes' because their expression is regulated in response to several abiotic stresses (Ambard-Bretteville et al. 2003; Andreadeli et al. 2009; David et al. 2010; Hourton-Cabassa et al. 1998; Kurt-Gür et al. 2018; Li et al. 2002; Lou et al. 2016; Murgia et al. 2020; Suzuki et al. 1998). Only few studies showed FDH involvement in the response to bacterial

infections (Choi et al. 2014; David et al. 2010; Lee et al. 2022; Marzorati et al. 2021); in particular, FDH is expressed in the leaves of *A. thaliana*, especially in hydathodes (Murgia et al. 2020), where it is involved in early defense responses against the pathogen *Xanthomonas campestris campestris* (Marzorati et al. 2021). WCS417 represents a model for the study of the interactions between plant and beneficial rhizobacteria (Pieterse et al. 2020 and references therein); however, little is still known on the molecular changes occurring at foliar level, upon plant root exposure to WCS417, as most studies have focused on the rhizobacterial effects on roots, after several days from exposure (Trapet et al. 2016; Verbon et al. 2019; Wintermans et al. 2016; Zamioudis et al. 2013). Given these premises, the goal of the present study is to explore the early responses of *A. thaliana* to WCS417 and the involvement of FDH in such responses; the activity of the FDH promoter was first investigated in response to WCS417, followed by proteomic analysis of leaves of the wt Col as well as of the FDH knock-out mutant *atfdh1-5* after their root exposure to WCS417. Proteomic analysis shows that FDH level increases after exposure to WCS417, thus confirming FDH involvement in the early *A. thaliana* foliar responses to WCS417, not only in terms of FDH promoter activity but also at protein level. Moreover, proteomic analysis reveals that chloroplasts, and in particular some proteins of photosystems PSI and PSII, as well as various members of the Glutathione S-transferases family, are early targets of the adaptive plant response to WCS417. Last, the comparison of wt Col and *atfdh1*-5 proteomes reveals the different regulation of some stress-responsive genes between the two lines, particularly after WCS417-treatment, suggesting FDH involvement in the early events triggered by the interactions between *A. thaliana* and the rhizobacterium WCS417.

Results

Pseudomonas simiae WCS417 rapidly induces FDH promoter activity in hydathodes

The effect of WCS417 on FDH expression was studied in seven days old *A. thaliana Vu* FDH::GUS seedlings co-cultivated *in vitro* with WCS417, for up to seven more days, without any direct contact between the seedlings and the rhizobacterium itself (**Figure 1A**); after seven days of co-cultivation, GUS staining of the seedlings revealed an increase in FDH promoter activity in WCS417-treated ones, both in their roots and shoots, in the leaf vascular tissue and hydathodes (**Figure 1B, D**) compared to the mock-treated seedlings (**Figure 1C, D**). To evaluate how rapidly WCS417 may affect *in vivo* FDH expression, the roots of 4 weeks old *A. thaliana Vu* FDH::GUS plants grown in soil were exposed to WCS417, by direct inoculation of WCS417 into the soil pots, and their rosette leaves were GUS-stained. After two days, a higher number of stained hydathodes was observed in the rosette leaves of WCS417-treated plants than in those from mock-treated plants (**Figure 2A**). Treatment with WCS417 did not alter the common parameters evaluated for photochemical efficiency, *i.e.*, F_0 (initial), F_m (maximum), F_v (variable) fluorescence, and the maximum photochemical efficiency (F_v/F_m) (**Figure S1A, B**). The induction of FDH promoter activity *in vitro*, without contact between the rhizobacterium and roots, suggests that WCS417 could affect FDH expression through the emission of volatile compounds. To investigate this possibility *in vivo*, plants were organized in lines in Aratrays to keep those positioned at the edges (named 'close to WCS417') fully isolated from those positioned at the tray center, for which an entire line of empty baskets was positioned between the two groups (**Figure S2A**). The roots of the plants positioned in the central lines were exposed to WCS417, by direct inoculation of WCS417 into the soil pots, and the whole tray was then covered with an Aralid without holes to preserve volatile compounds (**Figure S2B**). The same plant arrangement and treatment were also performed for mock-treated plants and the 'close to mock' ones.

After four days, the induction of FDH promoter activity could be observed with a higher number of GUS-stained hydathodes in WCS417-treated plants than in mock-treated plants (**Figure 2B**); however, the slightly higher number of stained hydathodes observed in the leaves of the plants positioned in the external lines, which did not receive WCS417 themselves but were close to those inoculated with the rhizobacterium, was not statistically significant when compared to their mock counterpart (**Figure 2B**).

The chloroplasts, Glutathione S-transferases, and stress-responsive proteins are early targets in the metabolic changes induced by WCS417

The observed early changes in FDH expression suggest that WCS417 can promptly affect the metabolic and signaling pathways in *A. thaliana*. To uncover early rearrangements of these pathways and, in particular, the specific role of FDH in these WCS417-induced networks, the roots of 4 weeks old *A. thaliana* wt Col plants and the FDH knock-out *atfdh1-5* mutant (Murgia et al. 2020) were exposed to WCS417, and after two days their rosette leaves were sampled for proteomic analysis. WCS417 treatment slightly decreased the weight of the rosettes in the wt, but not in the *atfdh1-5*, which increased in weight (**Figure S3**). The leaves of wt Col and *atfdh1-5* were stained with diaminobenzidine (DAB) to detect any changes in the levels of hydrogen peroxide H_2O_2 after WCS417 exposure, and all the leaves sampled from WCS417-treated plants appear less brown-colored than their mock counterparts, suggesting that, far from causing oxidative stress, the rhizobacterium reduces the level of ROS (**Figure S4**). Proteomic analysis was then performed on total proteins extracted from leaves: LC-MS/MS analysis from untreated (wt Col mock, *atfdh1-5* mock) or exposed to WCS417 (wt Col WCS417, *atfdh1-5* WCS417) samples allowed the identification of a total of 2196 distinct proteins (**Table S1**). About 16% of the total proteins

had an average Peptide Spectrum Match (PSM) higher than 1 (**Figure 3A**). Globally, for each condition, about one thousand proteins were identified, half of which were shared in the pairwise comparisons (**Figure 3B**): 918 and 912 proteins were detected in mock-treated wt Col and *atfdh1-5* plants, respectively, whereas 998 and 1066 proteins were detected in the leaves of wt and *atfdh1-5* plants exposed to WCS417, respectively. A label-free semi-quantitative comparison among the characterized protein profiles (wt Col mock *vs* wt Col WCS417, *atfdh1-5* mock *vs atfdh1-5* WCS417, wt Col mock *vs atfdh1-5* mock, and wt Col WCS417 *vs atfdh1-5* WCS417) allowed the extraction of total 362 Differentially Expressed Proteins (DEPs) (**Table S2**). Major differences emerge between wt Col mock and wt Col WCS417 (150 DEPs, $P \leq 0.05$; 63 DEPs, $P \leq 0.01$) (**Figure 4A, Table S3**) and between *atfdh1-5* mock and *atfdh1-5* WCS417 (268 DEPs, $P \leq 0.05$; 161 DEPSs, $P \leq 0.01$) (**Figure 4B, Table S4**). The functional evaluation of the characterized proteomes reveals a major enrichment of metabolic pathways (amino acid, carbon, and nitrogen metabolism) in both *A. thaliana* lines exposed to WCS417 (**Table S5**), which is more pronounced in the *atfdh1-5* WCS417 line. In this scenario, the presence of WCS417 correlates with the enrichment of other interesting pathways, including photosynthesis, stress response, immune response, and transcription/translation**.** Notably, FDH increases in WCS417-treated samples of wt Col **(Tables S1, S2)**, thus confirming FDH promoter activity at the protein level; as expected, FDH was absent in the foliar proteomes of the *atfdh1-5* mutant **(Tables S1, S2)** regardless of treatment. Among the identified WCS417-upregulated proteins, 10 are shared between wt Col and *atfdh1-5* (**Table 1**); among the WCS417-downregulated proteins, 19 are shared between wt Col and *atfdh1-5* (**Table 2**). Therefore, these 29 WCS417-regulated proteins shared between the wt and mutant lines are early targets of the changes mediated by WCS417 treatment in an FDH-independent manner. The group of upregulated proteins (**Table 1**) is composed by two glutathione transferases GSTF7 and GSTF8, two subunits of vacuolar-

type H^+-ATPase VHA-B1 and VHA-C, three proteins with house-keeping functions (translation initiation factor 4A1 EIF4A1, ribosomal proteins RPS1, and RPS18C), and three plastidial proteins, that are the ATP synthase subunit beta atpB, a ribose-5-phosphate isomerase RPI3, and a lipid-associated protein PAP6, also known as fibrillin 4. GSTF is a GST type Phy, formerly known as type I, involved in the response to abiotic and biotic stresses (Sylvestre-Gonon et al. 2019); the expression of GSTF7 and GSTF8 is modulated by salicylic acid (SA) (Sappl et al. 2004). V-type H^+-ATPase is formed by various subunits with complex regulation and is involved in stress adaptation (Dietz et al. 2001; Li et al. 2022); VHA-B1 is involved in the modeling of the actin cytoskeleton (Ma et al. 2012). Among the four *A. thaliana* RPI isoforms, RPI1 is involved in actin organization (Huang et al. 2020); however, to date, no physiological functions have been assigned to RPI3. PAP6 is involved in the resistance to biotic and abiotic stresses (Singh et al. 2010). Among the WCS417-downregulated proteins (**Table 2**), several of them are localized in plastids and, in particular, are part of the photosynthetic electron transport chain: PSAE1 and PSAE2 are subunits IV A and B of photosystem I; PSBP1 is an oxygen-evolving enhancer protein required for photosystem II organization (Yi et al. 2007); PBS27-1 is a repair protein involved in photosystem II assembly (Cormann et al. 2016); PSBO2 is the oxygen-evolving enhancer protein 1-2, required for the regulation of the D1 reaction center of photosystem II (Lundin et al. 2007); CP29 is a minor monomeric component of the PSII light-harvesting complex that, when phosphorylated, contributes to PSII state transition and disassembly (Chen et al. 2013). Among the WCS417-downregulated proteins that are localized in plastid, there are also the chaperonins CPN10-2 and CPN20, the ribosome recycling factor RRF required for chloroplast biosynthesis (Wang et al. 2010), the RNA-binding protein RGGC characterized by the arginine-glycine-glycine (RGG) region RGGC, the thylakoid soluble phosphoprotein F13I12.120, and one unknown protein encoded by the At2g21530 gene

(**Table 2**). A few WCS417-downregulated proteins are localized in the nucleus, *i.e.*, the RNA-binding protein RGGA involved in the response to salt and drought stresses (Ambrosone et al. 2015), the negative regulator of cold acclimation cold shock protein 2 CSP2 (Sasaki et al. 2013), the core component of the nucleosome Histone H2B.9 (At5g02570), and an essential embryogenesis protein MEE59 (Pagnussat et al. 2005). Last, the stress-responsive dehydrin ERD10, which belongs to the dehydrin family and is expressed in particular under different abiotic stresses (Sun et al. 2021), calmodulin 7 CAM7 (Kushwaha et al. 2008), and a protein of unknown function encoded by the At5g24165 gene were also identified as WCS417-downregulated proteins (**Table 2**). The comparison of mock-treated wt Col and *atfdh1-5* proteomes identified 17 DEPs ($P \leq 0.01$) (61 DEPs, $P \leq 0.05$) (**Figure 4C, Table S6A**) whereas the comparison of WCS417-treated wt Col and *atfdh1-5* proteomes identified 30 DEPs ($P \leq 0.01$) (84 DEPs $P \leq 0.05$) (**Figure 4 D, Table S6 B**); These results are consistent with the correlation scores of the spectral counts compared in pairs: comparisons of the mock-treated lines (wt Col mock and *atfdh1-5* mock) and of the WCS417-treated lines (wt Col WCS417 and *atfdh1-5* WCS417) have higher correlation values (r ~ 0,8) than the other comparisons, which show lower correlation values (**Figure 4 E)**. Notably, seven enzymes involved in ROS-detoxification are differentially expressed in wt Col and/or *atfdh1-5* under the two experimental conditions (**Table 3**), *i.e.* six Glutathione Transferases (GSTF2, GSTF7, GSTF8, GSTF9, GSTF1, GSTU19) and the ascorbate peroxidase 1 APX1, which scavenges cytosolic H_2O_2 (Hong et al. 2022); in particular, APX1 is the only ROS detoxification protein that is upregulated in *atfdh1-5* leaves under both conditions (mock and WCS417-treated) with respect to wt Col counterparts (**Figure 4 C, D** and **Table S6 A, B**). Various other proteins involved in resistance to oxidative stress were also upregulated in *atfdh1-5*; the Pathogenesis-Related protein 5 PR5 (At1g75040) involved in the activation of the SA signaling pathway (Ali et al. 2018), is

upregulated in mock-treated *atfdh1-5* compared to its wt Col counterpart (**Figure 4C**, **Table S6A**), whereas PER34, GLO2, and ACO3 are upregulated in WCS417-treated *atfdh1-5* compared to their wt Col counterparts (**Figure 4D**, **Table S6B**). Viceversa, the lipoxygenase LOX2 required for jasmonic acid (JA) biosynthesis in leaves (Yang et al. 2020) is downregulated in WCS417-treated *atfdh1-5* leaves compared to the wt (**Figure 4D**, **Table S6B**), and the adenosylhomocysteinase HOG1 involved in chromatin modifications and transcriptional gene silencing (Baubec et al. 2010) is absent in the WCS417-treated *atfdh1-5*, whereas it is strongly expressed in the wt under the same condition (**Figure 4D**, **Table S6B**).

Discussion

Root microbiota can have a strong influence on the plant immune system because they can trigger the Induced Systemic Resistance (ISR) response. Formate dehydrogenase (FDH) is nutritional hub for iron (Fe) and molybdenum (Vigani et al. 2017; Di Silvestre et al. 2021) and it also takes part in the plant response against the pathogen *Xanthomonas campestris* pv *campestris*, especially in hydathodes (Marzorati et al. 2021). This latter evidence prompted us to investigate the possible involvement of FDH in the plant response to the beneficial rhizobacterium *Pseudomonas simiae* WCS417. In the present work, we could demonstrate that the FDH promoter is activated in both the roots and shoots of seedlings exposed to WCS417, and that this FDH activation is quite rapid *in vivo*, as it could be detected, in the hydathodes of the rosette leaves, just after two days of root exposure to WCS417. Consistently with the results of previous studies (Wintermans et al. 2016; Zamioudis et al. 2013), we could also demonstrate that, *in vitro*, the observed effects may be mediated by rhizobacterium-produced volatile compounds. Our proteome analysis unveils that WCS417

not only affects the production of plant proteins involved in essential metabolic processes, but that the plastids are early targets of WCS417, in particular several photosynthesis-related proteins. These findings are particularly intriguing, as the link between plant-microbial pathogens interactions and chloroplasts has been already uncovered (Littlejohn et al. 2020; Yang et al. 2021), whereas the link between chloroplasts and plant-beneficial bacteria was unexplored, so far. In particular, our results suggest WCS417- induced rearrangements of PSI and PSII composition, in terms of their proteins PSBO2, PSBP1, PSB27-1, PSAE1, PSAE2. The physiological relevance of the WCS417-induced modulation of composition, stability, and turnover of photosystems should be object of future investigations, in the light of the role of chloroplasts in the biosynthesis of phytohormones such as JA (Wastenack and Hause, 2019) and of previous findings suggesting that WCS417 stimulates the expression of genes important for plant growth (Wintermans et al. 2016; Zamioudis et al. 2013). Notably, one of the WCS417-upregulated proteins in wt leaves compared to *atfdh1-5* is indeed lipoxygenase LOX2, which is responsible for JA synthesis (Yang et al. 2020). This phytohormone is involved in plant development and, along with lipoxygenases, is important during defense responses against biotic stress (Singh et al. 2022); ISR is indeed associated with the activation of specific JA-responsive genes, even though ISR dependence on JA and ethylene is due to increased sensitivity to these hormones rather than an increase in their production (van Wees et al. 1999). In terms of defense responses, we observed an increase in the enzymes responsible for ROS detoxification, such as glutathione S-transferases (GSTs), upon WCS417 exposure (Gullner et al. 2018). GST gene induction or increased GST activity has been reported in plants that interact with the beneficial bacteria responsible for the ISR response (Kandasamy et al. 2009; Miché et al. 2006; Wang et al. 2013); accordingly, we show that GST7 and GST8 are upregulated in both plant lines after WCS417 exposure. One possible explanation is that WCS417 triggers a temporary antioxidant

response, as supported by the DAB staining of the leaves exposed to WCS417, which appear to accumulate less H_2O_2 than mock-treated ones. Both ROS and antioxidants are linked to salicylic acid (SA) signaling (Saleem et al. 2021) and the accumulation of SA and the expression of pathogenesis-related proteins are linked to the defense response Systemic Acquired Resistance (SAR), a systemic immune response different from ISR (Vallad and Goodman, 2004; Vlot et al. 2021). PR5 is considered a marker for SAR triggering (Sharon et al. 2011), and, surprisingly, we discovered that this protein is increased in *atfdh1-5* leaves in mock condition. In the WCS417-treated *atfdh1-5* leaf proteome, several dehydrins were downregulated (HIRD11, COR15B, COR47, ERD14, and ERD10); in fact, the levels of several members of this protein family increased when *A. thaliana* was exposed to beneficial microorganisms which colonized its roots for defense (Baek et al. 2020; Kovacs et al. 2008; Liu et al. 2020). These results, suggest that lack of FDH function in the *atfdh1-5* mutant triggers an altered systemic defense mechanism in leaves, particularly after WCS417 treatment. FDH protein levels indeed increase in wt leaves exposed to WCS417, corroborating our results of FDH promoter activity induction in *A. thaliana* leaves exposed to the rhizobacterium as well as FDH involvement in an early leaf defense response against pathogens (Marzorati et al. 2021). Overall, our findings on the *atfdh1-5* leaf proteome suggest that the FDH may have a relevant role on the early WCS417-induced pathway. The nodes of convergence between the ISR and Fe-deficiency response are well established (Romera et al. 2019); interestingly, GSTF2, GSTF7, and LOX2 proteins are upregulated whereas PSAE1 is downregulated in shoots of *A. thaliana* exposed to Fe deficient growth conditions (Zargar et al. 2013); it suggests that these proteins also identified in the present proteomic work, and with a regulation similar to that described by Zargar et al. (2013) are among the nodes of convergence between ISR and Fe-deficiency (Zargar et al. 2013). Given the importance of FDH in iron homeostasis (Di Silvestre et al. 2021; Murgia et al. 2020) and

the partial overlap between the WCS417 immune pathway and the Fe deficiency response, future research under different Fe nutritional conditions may help elucidate the role of FDH in rhizobacterial plant colonization. In conclusion, the results presented in this work can stimulate further investigations which may contribute in the fine understanding of the signaling pathways triggered, at foliar level, by growth-promoting rhizobacteria that enhance plant resistance to environmental challenges, such as nutritional stress and pathogen infections.

Materials and methods

Plant growth

Arabidopsis thaliana wt Col*, atfdh1-5* mutant (Choi et al. 2014; Murgia et al. 2020), and *Vu* FDH::GUS (Lou et al. 2016; Murgia et al. 2020) were stratified at 4°C and grown on Technic n.1 DueEmme soil by using the Arasystem (Betatech BVBA, Ghent, Belgium), *i.e.*, the Aratrays and the Arabaskets, in a greenhouse at 23°C and 150 μE $m^{-2}s^{-1}$, with a 12 h/12 h light/dark photoperiod. *Vu* FDH::GUS seeds were surface sterilized as described by van Wees et al. (2013), maintained in the dark for 3 days at 4°C, then transferred on square plates dishes (100x100x20 mm, Sarstedt, Australia Ltd) containing ½ MS medium supplemented with 1% sucrose, and maintained vertically in a plant growth chamber at 22-25°C, 16 h/8 h light/dark photoperiod.

Plant exposure to WCS417

The *Pseudomonas simiae* WCS417 bacterial strain was grown overnight at 28°C on King's B medium agar supplemented with 50 μg mL^{-1} rifampicin, suspended in 10 mL of 10 mM

$MgSO_4$, and centrifuged for 5 min at 3200 g; the pellet was washed twice in 10 mM $MgSO_4$, 5 min centrifugation at 3200 g (Wintermans et al. 2016). The cell density was adjusted to 2 × 10^6 or 10^8 CFU mL^{-1} in 10 mM $MgSO_4$. Plants grown in soil: 1 mL of 10^8 CFU mL^{-1} bacterial suspension, or 1 mL of 10 mM $MgSO_4$ (mock condition), was pipetted into each Arabasket containing single 4 weeks old plants with equal distribution of the liquid around the plant roots. To optimize an even distribution of the bacterial inoculum for each single-root apparatus, plants were not watered for two days before treatment, so that the Aratrays remained dry before and after inoculation. The trays, closed with transparent lids without holes, were then maintained at 25°C. Seedlings grown *in vitro*: seven days old seedlings grown on MS were exposed to WCS417 avoiding any direct contact between seedlings and bacteria, as previously described (Wintermans et al. 2016). Briefly, 240 µL of 2 × 10^6 CFU mL^{-1} WCS417 suspension (or 240 µL of 10 mM $MgSO_4$ for mock treatment) was pipetted onto the MS medium, approximately 5 cm below the seedling roots. The plates were briefly dried under laminar flow, closed with a lid and two layers of parafilm, and placed again vertically in a growth chamber for two or seven days.

Leaves staining

GUS staining: leaves and seedlings were surface-sterilized by immersion in 70% EtOH and washed twice with sterile water as previously described (Hulten et al. 2019). Staining for β-glucuronidase (GUS) activity was performed according to (Elorza et al. 2004). DAB staining: H_2O_2 staining with 3,3′-diaminobenzidine (DAB) was performed as previously described (Murgia et al. 2004).

Protein extraction from leaves

Rosette leaves were sampled from 4 weeks old plants after two days of exposure to either WCS417 or mock treatment, as described above. In detail, the rosette leaves from one single plant were sampled, weighed (0.15-0.4 gr each), packed in alufoil, frozen in liquid nitrogen, stored at -80°C, and total proteins were then extracted essentially according to the protocol published by Wu et al. (2014), omitting the TCA/acetone precipitation steps (steps 2-9 in the Wu et al. protocol); as starting material for each extraction representing a biological sample, leaves from two different rosettes were used. The protein pellets were maintained at -80°C. Pellets were then resuspended in up to 100-120 final volume of 10 mM PBS by heating for 15 min at 37°C and vortexing, followed by 2 min centrifugation at 15000 g; the supernatant contained the solubilized proteins, and if a remaining pellet could still be observed, it was heated, vortexed, and centrifuged again, for one or two more cycles, for thorough solubilization of all the proteins in the starting frozen pellet.

Enzymatic digestion of protein extracts

The total protein extract of each sample was concentrated from 100 to 50 µL in a vacuum concentrator at 60 °C and treated with 0.25% (w/v) RapiGest^{TM}SF reagent (Waters Co, Milford, MA, USA). The resulting suspensions were incubated with stirring at 100°C for 20 min, cooled to RT, and centrifuged for 10 min at 2200 g. The protein concentration was assayed using the Invitrogen ™ Qubit™ Protein BR Assay Kit (Life Technologies Corporation, Thermo Fisher, Eugene, ORE, USA), and 50 µg of protein from each sample was digested overnight at 37°C by adding Sequencing-grade Modified Trypsin (Promega Inc., Madison, WI, USA) at a 1:50 (w/w) enzyme/substrate ratio. An additional aliquot of trypsin (1:100 w/w) was then added in the morning, and the digestion continued for 4h. The

enzymatic digestion was chemically stopped by acidification with 0.5% Trifluoroacetic Acid (TFA) (Sigma-Aldrich Inc., St.Louis, MO, USA) and a subsequent incubation at 37°C for 45 min completed the RapiGest acid hydrolysis. Water-immiscible degradation products were removed by centrifugation at 13000 rpm for 10 min. Finally, the tryptic digest mixtures were desalted using PierceTM C-18 spin columns (Thermo Fisher Scientific, Pierce Biotechnology, Rockford, Il, USA), according to manufacturer protocol and were resuspended in 0.1% formic acid (Sigma-Aldrich Inc., St. Louis, MO, USA) in water (LC-MS Ultra CHROMASOLV™, Honeywell Riedel-de HaenTM, Muskegon, MI, USA) at a concentration of 0.2 µg/µL.

LC-MS/MS Analysis

Peptide mixtures were analyzed using Eksigent nanoLC-Ultra® 2D System (Eksigent, part of AB SCIEX Dublin, CA, USA) configured in trap-elute mode. Briefly, samples (0.8 µg injected) were first loaded on a trap (200 µm x 500 µm ChromXP C18-CL, 3 µm, 120 Å) and washed with the loading pump running in isocratic mode with 0.1% formic acid in water for 10 min at a flow of 3 µL/min. The automatic switching of the autosampler ten-port valve then eluted the trapped mixture on a nano reversed-phase column (75 µm x 15 cm ChromXP C18-CL, 3 µm, 120 Å) through a 145 min gradient of eluent B (eluent A, 0.1% formic acid in water; eluent B, 0.1% formic acid in acetonitrile) at a flow rate of 300 nL/min. In-depth, the gradient was: from 5-10% B in 3 min, 10-30% B in 104 min, 30-95% B in 26 min, and holding at 95% B for 12 min. The eluted peptides were directly analyzed on an LTQ-OrbitrapXL mass spectrometer (Thermo Fisher Scientific, CA, USA) equipped with a nanospray ion source. The spray capillary voltage was set at 1.7 kV and the ion transfer capillary temperature was held at 220°C. Full MS spectra were recorded over a 400–1600

m/z range in positive ion mode, with a resolving power of 60000 (full width at half-maximum) and a scan rate of 2 spectra/s. This step was followed by five low-resolution MS/MS events that were sequentially generated in a data-dependent manner on the top five ions selected from the full MS spectrum (at 35% collision energy) using dynamic exclusion of 0.5 min for MS/MS analysis. Mass spectrometer scan functions and high-performance liquid chromatography solvent gradients were controlled by the Xcalibur data system version 1.4 (Thermo Fisher Scientific, CA, USA).

LC-MS/MS spectra processing and data handling

The Proteome Discoverer software 2.5 using SEQUEST HT search engine (Thermo Fisher Scientific, San José, CA, USA) was used to process all LC-MS/MS runs against *Arabidopsis thaliana* counting 39256 entries (www.uniprot.org, downloaded in July 2022). The following criteria were used for peptide and related protein identification: trypsin as enzyme with 2 missed cleavage per peptide, mass tolerance of ± 50 ppm mass tolerance for the precursor, and ± 0.8 Da for fragment ions. Validation was performed by Percolator node with a target-decoy search and a false discovery rate (FDR) ≤ 0.01 and maximum deltaCN of 0.05. The minimum peptide length of 7 amino acids at confidence 'Medium' level was set. Peptide Spectral Matches (PSMs) were used in a label-free quantification approach to compare protein lists (n=24) and identify proteins differentially expressed (DEPs), as previously reported (Palma et al. 2021). Briefly, data matrix complexity was reduced by linear discriminant analysis (LDA) and in a pairwise comparison (wt Col mock-treated vs wt Col WCS417-treated; *atfdh1-5* mock-treated vs *atfdh1-5* WCS417-treated; wt Col mock-treated vs *atfdh1-5* mock-treated; wt Col WCS417-treated vs *atdh1-5* WCS417-treated) and proteins with $P \leq 0.05$ were retained. Pairwise comparisons were further evaluated by DAve

index ((PSMs_A-PSMs_B)/(PSMs_A+PSMs_B))/0.5, where A and B represent the samples compared; specifically, positive DAve values indicate protein upregulated in A (and downregulated in B), while negative DAve value indicates proteins upregulated in B (and downregulated in A) (Mauri and Dehò, 2008). Finally, DEPs were processed by hierarchical clustering using Ward's method and the Euclidean distance metric. All data processing was performed by JMP15.2 SAS. Using STRING Cytoscape's APP (Doncheva et al. 2019), the protein profile characterized for wt Col mock-treated, wt Col WCS417-treated, *atfdh1-5* mock-treated, and *atfdh1-5* WCS417-treated phenotypes were evaluated at the functional level, and the most enriched KEGG pathways and biological processes were extracted and compared (wt Col mock-treated vs wt WCS417-treated; *atfdh1-5* mock-treated vs *atfdh1-5* WCS417-treated; wt Col mock-treated vs *atfdh1-5* mock-treated; wt Col WCS417-treated vs *atfdh1-5* WCS417) by unpaired t-test ($P \leq 0.01$).

Statistical analysis

To test for significant differences between WCS417-treated plants and mock-treated plants for the *in vivo* data experiments, an unpaired t-test was run, establishing for each comparison equal or unequal variances before the analysis by an F-test two samples for variances. In detail: for *in vitro* mock-treated two days (sample size: 54) vs *in vitro* WCS417-treated two days (sample size: 63) unequal variances were assumed, $P = 0.04512639$ ($P < 0.05$); for *in vitro* mock-treated seven days (sample size: 70) vs *in vitro* WCS417-treated seven days (sample size: 135) equal variances were assumed, $P = 3.052E^{-08}$ ($P < 0.01$); for *in vivo* mock-treated two days (sample size: 36) vs *in vivo* WCS417-treated two days (sample size: 36) equal variances were assumed, $P = 0.001576$ ($P < 0.01$); for *in vivo* mock-treated four days (sample size: 82) vs *in vivo* WCS417-treated four days (sample size: 100) unequal variances

were assumed, $P = 0.01385$ ($P \leq 0.01$); for close to mock (sample size: 90) vs close to WCS417 (sample size: 99) equal variances were assumed, $P = 0.4325$ (not statistically significant); for Col mock-treated weights (sample size: 25) vs Col WCS417-treated weights (sample size: 25) unequal variances were assumed, $P = 0.002381$ ($P < 0.01$); for *atfdh1-5* mock-treated weights (sample size: 25) vs *atfdh1-5* WCS417-treated weights (sample size: 25) unequal variances were assumed, $P = 0.0002032$ ($P < 0.01$). All data processing was performed by using R (ver. 4.1.0), packages *ggpubr* and *dplyr*.

Photochemical parameters

The photochemical parameters F_0, F_m, F_v, and maximal photochemical efficiency F_v/F_m were measured in dark-adapted leaves (20 min) as previously described (Murgia et al. 2020).

Acknowledgments

We are grateful to Peter Bakker for donating the *Pseudomonas simiae* WCS417r strain.

Authors contributions

FM, IM, and PM conceived the experiments; FM and IM performed the *in vitro* and *in vivo* experiments, with contributions from PM; DDS, RR, LB, and PLM performed the proteomic analysis; FM, PM, IM, DDS, and LB analyzed the results; FM and IM wrote the manuscript, with PM contributions. All authors agreed to the final version of the manuscript.

Figure legends

Figure 1

GUS staining of *A. thaliana Vu* FDH::GUS seedlings after *in vitro* co-cultivation with WCS417. **(A)** with WCS417 or $MgSO_4$ (mock condition), avoiding contact with the root apparatus. A schematic representation of the experiment is shown. **(B)** GUS staining of seven days WCS417-treated or **(C)** mock-treated seedlings: details of leaves and root apparatus are shown, and the hydathodes are indicated by red arrows. Scale bars represent 1 mm. **(D)** Number of stained hydathodes after two or seven days exposure to WCS417, with respect to the mock treatment, as described in **(B)** and **(C)**; each bar represents the mean value ± SE of stained hydathodes measured in at least 24 seedlings collected from 3 different plates (at least 8 seedlings per plate). Significant differences in WCS417-treated with respect to mock-treated, according to the t-test, are indicated by * ($P < 0.05$) or ** ($P < 0.01$). Panel 1A was created by using BioRender.com

Figure 2

GUS-staining of hydathodes in *A. thaliana Vu* FDH::GUS leaves after *in vivo* root exposure to WCS417. **(A)** Leaves from 4 weeks old plants grown in soil were stained for GUS activity before inoculation (control), and after two days of WCS417 or $MgSO_4$ (mock) inoculation in the soil (indicated as WCS417-treated or mock-treated, respectively); each bar represents the mean value ± SE of stained hydathodes in 12 (control) or 36 (WCS417-treated, mock-treated) GUS-stained leaves. Significant differences between WCS417-and mock-treated values, according to the t-test, are indicated with **($P < 0.01$) **(B)** Leaves from 4 weeks old plants grown in soil were stained for GUS activity before inoculation (control) and after four days WCS417 or $MgSO_4$ (mock) inoculation in the soil (indicated as 'WCS417-treated' or

'mock-treated', respectively); leaves sampled from plants close (but without any contact) to either the WCS417- or mock-treated ones ('close to WCS417' and 'close to mock', respectively) were also GUS-stained. Bars represent the mean number ± SE of stained hydathodes in 73 (control), 82 (mock), 100 (WCS417), 90 (close to mock), and 99 (close to WCS417) GUS-stained leaves. Significant differences between WCS417- and mock-treated values are indicated by ** ($P \leq 0.01$) according to the t-test.

Figure 3

Proteomic analysis of *A. thaliana* wt Col and *atfdh1-5* leaves after root exposure to WCS417. Roots of four weeks old *A. thaliana* wt Col and *atfdh1-5* plants were exposed for two days to WCS417 (or mock treatment), and total proteins were then extracted for proteomic analysis. For each line and treatment 3 biological x 2 technical replicates were analyzed (n=6). (**A)** 2D Map showing the distribution of identified proteins by pI, MW, and global average Peptide Spectrum Matches (PSMs). (**B)** Venn diagrams of the number of identified proteins in pairwise comparison: wt Col mock *vs* wt Col WCS417; *atfdh1-5* mock *vs atfdh1-5* WCS417; wt Col mock *vs atfdh1-5* mock; wt Col WCS417 *vs atfdh1*-5 WCS417.

Figure 4

Proteomic analysis of *A. thaliana* wt Col and *atfdh1-5* leaves after root exposure to WCS417. Hierarchical clustering of proteins differentially expressed (LDA, $P \leq 0.01$) by comparing **(A)** wt Col mock *vs* wt Col WCS417, and **(B)** *atfdh1-5* mock *vs atfdh1-5* WCS417 **(C)** wt Col mock *vs atfdh1-5* mock, **(D)** wt Col WCS417 *vs atfdh1-5* WCS417. **(E)** Spearman's correlation values r by comparing, in pairs, proteins identified as differentially expressed

(LDA, $P \leq 0.05$). For each graph, the coordinates indicate the spectral counts of a protein, in the two analyzed conditions.

Figure S1

Photochemical parameters of *A. thaliana* wt Col leaves after WCS417 treatment. Roots of 4 weeks old *A. thaliana* wt Col plants grown in soil plant were exposed to WCS417 or $MgSO_4$ (mock), and their leaves were sampled after two days to evaluate **(A)** F_0 (initial fluorescence), F_m (maximum fluorescence), F_v (variable fluorescence) and **(B)** F_v/F_m (maximal photochemical efficiency). Each bar represents the mean value ± SE from at least 20 independent leaves.

Figure S2

A. thaliana plants in Aratrays to test the involvement of volatile compounds in WCS417-induced FDH expression. **(A)** 4 weeks old *A. thaliana Vu* FDH::GUS plants grown in soil placed in Aratrays : only those in the central lines were inoculated with either WCS417 or $MgSO_4$ (mock). The two external lines of plants did not receive any inoculum. **(B)** Aratrays covered with transparent Aratrays without holes after treatment.

Figure S3

A. thaliana wt Col and *atfdh1*-5 rosette leaves after exposure to WCS417 for proteomic analysis. Rosettes of *A. thaliana* wt Col and mutant *atfdh1-5* plants were sampled after two days of root exposure to WCS417 for proteomic analysis. Each bar represents the mean fresh weight of a single rosette (in gr) ± SE from 20 independent rosettes. Significant differences

between WCS417-treated and mock-treated samples, according to the t-test, are indicated with ** ($P < 0.01$).

Figure S4

Staining of *A. thaliana* wt Col and *atfdh1-5* rosette leaves with diaminobenzidine (DAB) after exposure to WCS417 for proteomic analysis. Rosette leaves of *A. thaliana* wt Col and *atfdh1-5* plants sampled after two days root exposure to WCS417 were stained with DAB for H_2O_2 detection.

Table legends

Table 1

Proteins with increased expression in *A. thaliana* wt Col and *atfdh1-5* leaves, after root exposure to WCS417. The list of proteins identified by proteomic analysis which are upregulated in the leaves of both wt Col and *atfdh1-5* after two days of root exposure to WCS417 (or to mock treatment), are reported; the name, AGI code, UniProt ID, annotation, subcellular localization according to Aramemnon (http://aramemnon.uni-koeln.de) and SUBA5 (https://suba.live/) (according to their respective highest scores), F ratio and Probability ($P \leq 0.01$) (by LDA), and DAve index (by MAProMa) are reported. SP: secretory pathway; PM: plasma membrane.

Table 2

Proteins with decreased expression in *A. thaliana* wt Col and *atfdh1-5* leaves, after root exposure to WCS417. The list of proteins identified by proteomic analysis which are

downregulated in the leaves of both wt Col and *atfdh1-5* after two days of root exposure to WCS417 (or to mock treatment), are reported; the name, AGI code, UniProt ID, annotation, subcellular localization according to Aramemnon (http://aramemnon.uni-koeln.de) and SUBA5 (https://suba.live/) (according to their respective highest scores), F ratio and Probability ($P \leq 0.01$) (by LDA), and DAve index (by MAProMa) are reported. SP: secretory pathway; PM: plasma membrane.

Table 3

Proteins with ROS detoxification function that are differentially expressed in *A. thaliana* wt Col and/or *atfdh1-5* leaves after two days root exposure to WCS417. The name, AGI code, UniProt ID, annotation, subcellular localization according to Aramemnon (http://aramemnon.uni-koeln.de) and SUBA5 (https://suba.live) (according to their respective highest scores) and average Peptide Spectrum Matches (PSMs) per condition are reported. Statistically significant values between mock and WCS417-treated are indicated with *, whereas / indicates that the given protein is not detected in the proteomic analysis under that experimental condition. SP: secretory pathway.

Table S1

Proteins identified by LC-MS/MS from *A. thaliana* leaves wt Col mock, *atfdh1-5* mock, wt Col WCS417-treated, *atfdh1-5* WCS417-treated; for each line and treatment 3 biological x 2 technical replicates (n=6) were analyzed. For each identified protein, the UniProt ID, MW, pI, description, gene name, expression values, Peptide Spectrum Matches (PSMs) per LC-MS/MS run, global average PSMs, average PSMs per condition, and MW normalized average PSMs per condition are reported.

Table S2

Differentially expressed proteins (DEPs) identified by comparing the protein profiles of *A. thaliana* leaves from wt Col mock, *atfdh1-5* mock, wt Col WCS417-treated, *atfdh1-5* WCS417-treated (n=6 for each line and treatment). For each pairwise comparison, the UniProt ID, description, average PSMs, F ratio and Probability ($P \leq 0.05$) (by LDA), and DAve index (by MAProMa) are reported. Specifically, given a generic comparison "Cond1 vs Cond2", positive DAve values (in red) indicate proteins up-regulated in Cond1 (and down-regulated in Cond2), while negative DAve values (in blue) indicate proteins up-regulated in Cond2 (and down-regulated in Cond1).

Table S3

A.thaliana wt Col proteins with altered expression, in 4 weeks old leaves after roots exposure for two days to WCS417, when compared to expression after mock treatment. For each identified protein, the protein name, UniProt ID, annotation, F ratio and Probability ($P < 0.01$) (by LDA), expression values for each replica in proteomic analysis, mean expression values under the two experimental conditions (mock and WCS417-treated) and DAve index (by MAProMa) are reported.

Table S4

A. thaliana atfdh1-5 proteins with altered expression, in 4 weeks old leaves after roots exposure for two days to WCS417, when compared to expression after mock treatment. For each identified protein, the protein name, UniProt ID, annotation, F ratio and Probability ($P < 0.01$) (by LDA), expression values for each replica in proteomic analysis, mean expression

values under the two experimental conditions (mock and WCS417-treated) and DAve index (by MAProMa) are reported.

Table S5

KEGG pathways differentially enriched by comparing the protein profiles from *A. thaliana* leaves wt Col mock, *atfdh1-5* mock, wt Col WCS417-treated, *atfdh1-5* WCS417-treated; n=6 for each line and treatment. Data matrix dimensionality reduction was performed by Linear Discriminant Analysis (LDA, $P \leq 0.05$). In addition, for each pairwise comparison, student's t-test was applied ($P \leq 0.05$). KEGG terms were retrieved using Cytoscape's STRING APP ($P \leq 0.05$).

Table S6

(A) List of DEPs ($P \leq 0.01$) proteins, by comparing *A. thaliana* wt Col and *atfdh1-5* in mock condition **(B)** List of DEPs ($P \leq 0.01$) proteins, by comparing *A. thaliana* wt Col and *atfdh1-5* after exposure to WCS417. For each identified protein, the protein name, UniProt ID, annotation, expression values for each sample, F ratio and Probability ($P \leq 0.01$) (by LDA), and DAve index (by MAProMa) are reported.

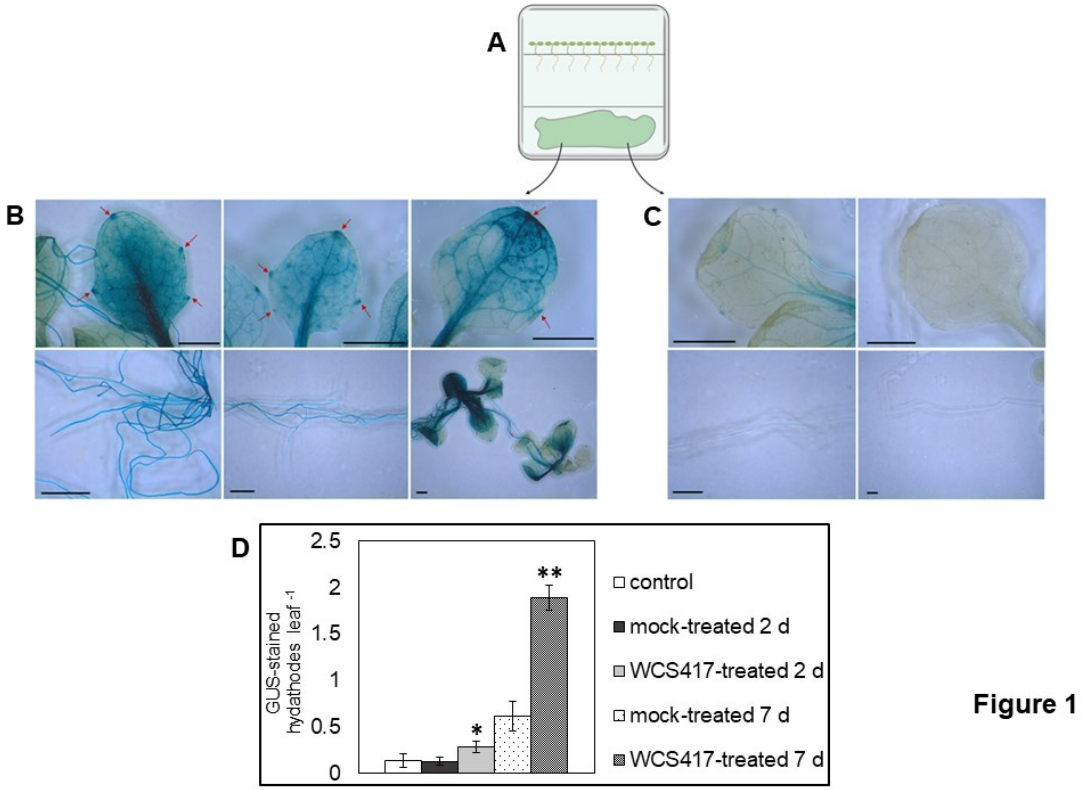


Figure 1

Figure 1. GUS staining of *A. thaliana Vu* FDH::GUS seedlings after *in vitro* co-cultivation with WCS417. **(A)** with WCS417 or $MgSO_4$ (mock condition), avoiding contact with the root apparatus. A schematic representation of the experiment is shown. **(B)** GUS staining of seven days WCS417-treated or **(C)** mock-treated seedlings: details of leaves and root apparatus are shown, and the hydathodes are indicated by red arrows. Scale bars represent 1 mm. **(D)** Number of stained hydathodes after two or seven days exposure to WCS417, with respect to the mock treatment, as described in **(B)** and **(C)**; each bar represents the mean value ± SE of stained hydathodes measured in at least 24 seedlings collected from 3 different plates (at least 8 seedlings per plate). Significant differences in WCS417-treated with respect to mock-treated, according to the t-test, are indicated by * ($P < 0.05$) or ** ($P < 0.01$). Panel 1A was created by using BioRender.com

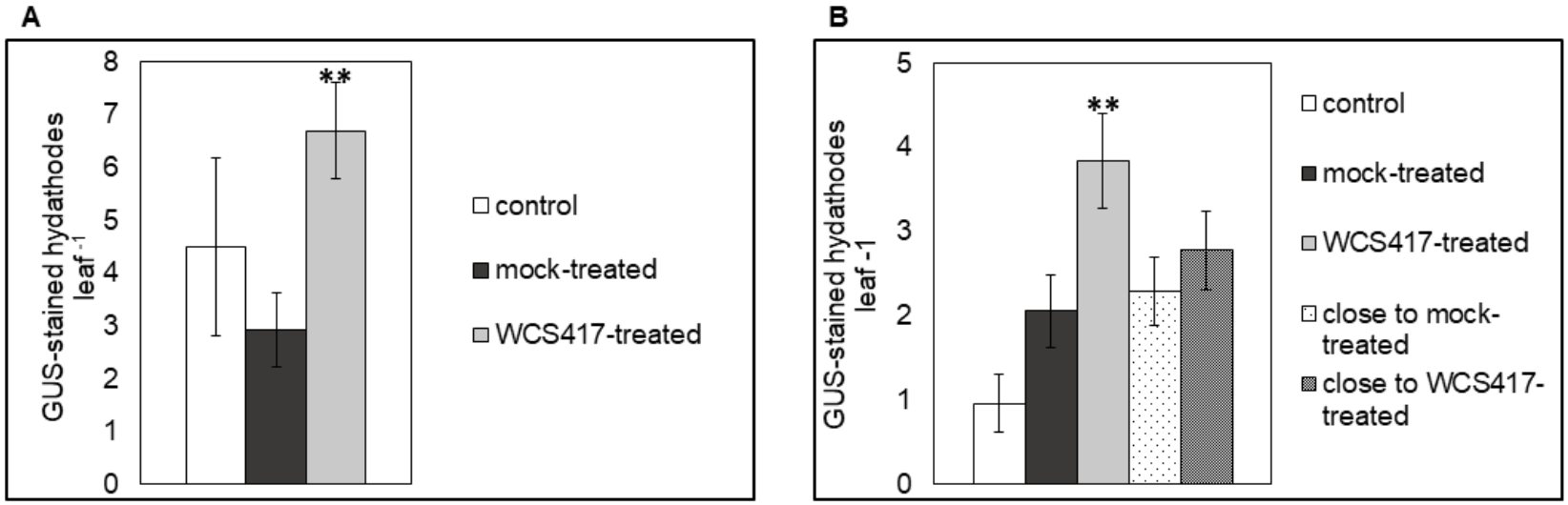


Figure 2. GUS-staining of hydathodes in *A. thaliana Vu* FDH::GUS leaves after *in vivo* root exposure to WCS417. **(A)** Leaves from 4 weeks old plants grown in soil were stained for GUS activity before inoculation (control), and after two days of WCS417 or $MgSO_4$ (mock) inoculation in the soil (indicated as WCS417-treated or mock-treated, respectively); each bar represents the mean value ± SE of stained hydathodes in 12 (control) or 36 (WCS417-treated, mock-treated) GUS-stained leaves. Significant differences between WCS417-and mock-treated values, according to the t-test, are indicated with **($P < 0.01$) **(B)** Leaves from 4 weeks old plants grown in soil were stained for GUS activity before inoculation (control) and after four days WCS417 or $MgSO_4$ (mock) inoculation in the soil (indicated as 'WCS417-treated' or 'mock-treated', respectively); leaves sampled from plants close (but without any contact) to either the WCS417- or mock-treated ones ('close to WCS417' and 'close to mock', respectively) were also GUS-stained. Bars represent the mean number ± SE of stained hydathodes in 73 (control), 82 (mock), 100 (WCS417), 90 (close to mock), and 99 (close to WCS417) GUS-stained leaves. Significant differences between WCS417- and mock-treated values are indicated by ** ($P \leq 0.01$) according to the t-test.

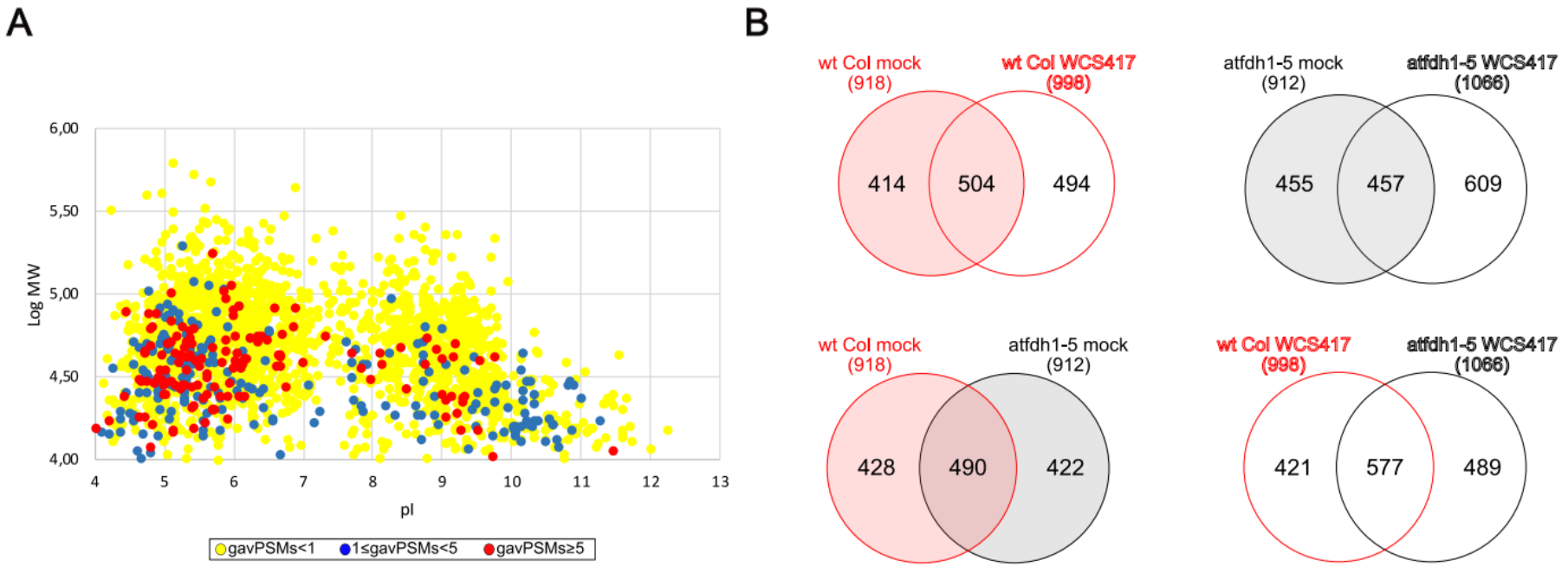


Figure 3. Proteomic analysis of *A. thaliana* wt Col and *atfdh1-5* leaves after root exposure to WCS417. Roots of four weeks old *A. thaliana* wt Col and *atfdh1-5* plants were exposed for two days to WCS417 (or mock treatment), and total proteins were then extracted for proteomic analysis. For each line and treatment 3 biological x 2 technical replicates were analyzed (n=6). (**A)** 2D Map showing the distribution of identified proteins by pI, MW, and global average Peptide Spectrum Matches (PSMs). (**B)** Venn diagrams of the number of identified proteins in pairwise comparison: wt Col mock *vs* wt Col WCS417; *atfdh1-5* mock *vs atfdh1-5* WCS417; wt Col mock *vs atfdh1-5* mock; wt Col WCS417 *vs atfdh1*-5 WCS417.

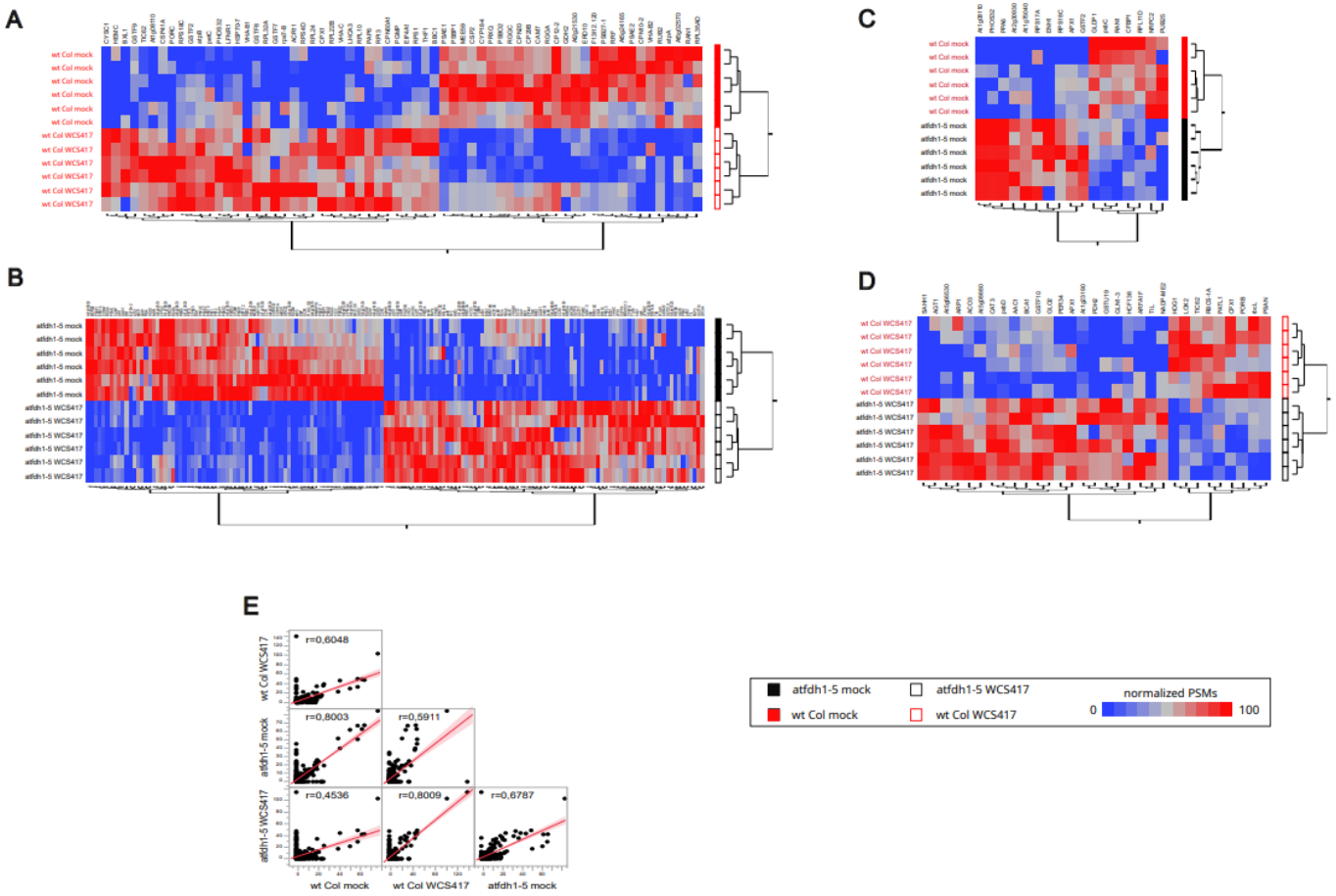


Figure 4. Proteomic analysis of *A. thaliana* wt Col and *atfdh1-5* leaves after root exposure to WCS417. Hierarchical clustering of proteins differentially expressed (LDA, $P \leq 0.01$) by comparing **(A)** wt Col mock *vs* wt Col WCS417, and **(B)** *atfdh1-5* mock *vs atfdh1-5* WCS417 **(C)** wt Col mock *vs atfdh1-5* mock, **(D)** wt Col WCS417 *vs atfdh1-5* WCS417. **(E)** Spearman's correlation values r by comparing, in pairs, proteins identified as differentially expressed (LDA, $P \leq 0.05$). For each graph, the coordinates indicate the spectral counts of a protein, in the two analyzed conditions.

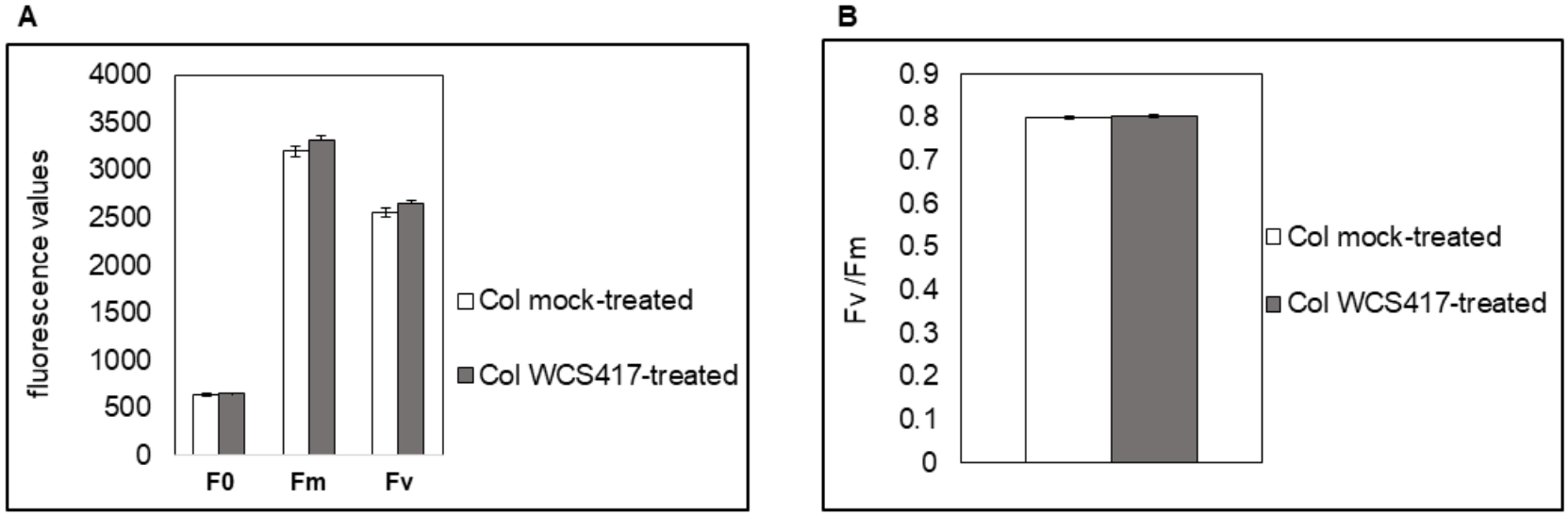


Figure S1. Photochemical parameters of *A. thaliana* wt Col leaves after WCS417 treatment. Roots of 4 weeks old *A. thaliana* wt Col plants grown in soil plant were exposed to WCS417 or $MgSO_4$ (mock), and their leaves were sampled after two days to evaluate **(A)** F_0 (initial fluorescence), F_m (maximum fluorescence), F_v (variable fluorescence) and **(B)** F_v/F_m (maximal photochemical efficiency). Each bar represents the mean value ± SE from at least 20 independent leaves.

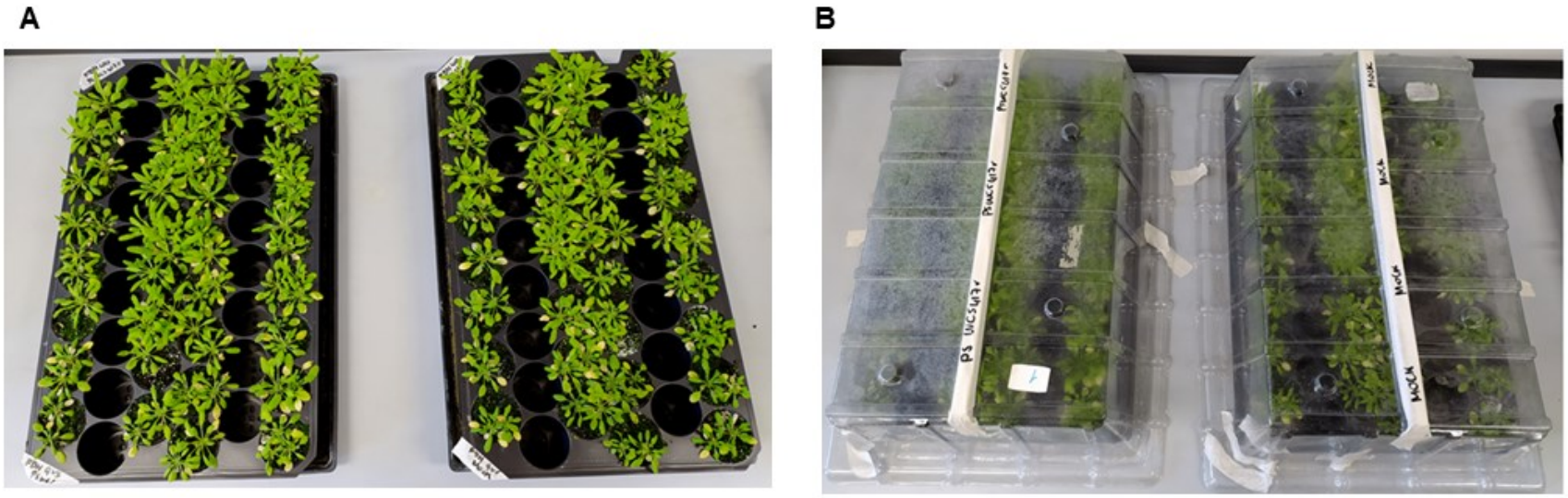


Figure S2. *A. thaliana* plants in Aratrays to test the involvement of volatile compounds in WCS417-induced FDH expression. **(A)** 4 weeks old *A. thaliana Vu* FDH::GUS plants grown in soil placed in Aratrays : only those in the central lines were inoculated with either WCS417 or $MgSO_4$ (mock). The two external lines of plants did not receive any inoculum. **(B)** Aratrays covered with transparent Aratrays without holes after treatment.

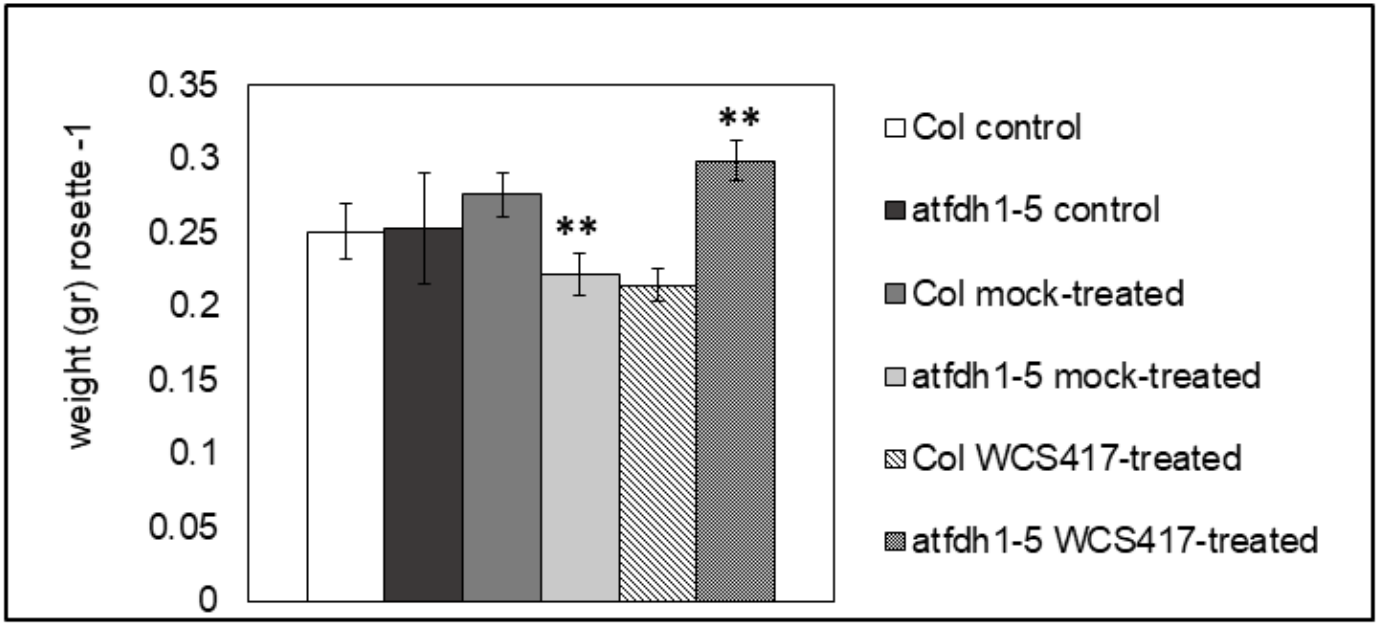


Figure S3. *A. thaliana* wt Col and *atfdh1*-5 rosette leaves after exposure to WCS417 for proteomic analysis. Rosettes of *A. thaliana* wt Col and mutant *atfdh1-5* plants were sampled after two days of root exposure to WCS417 for proteomic analysis. Each bar represents the mean fresh weight of a single rosette (in gr) ± SE from 20 independent rosettes. Significant differences between WCS417-treated and mock-treated samples, according to the t-test, are indicated with ** ($P < 0.01$).

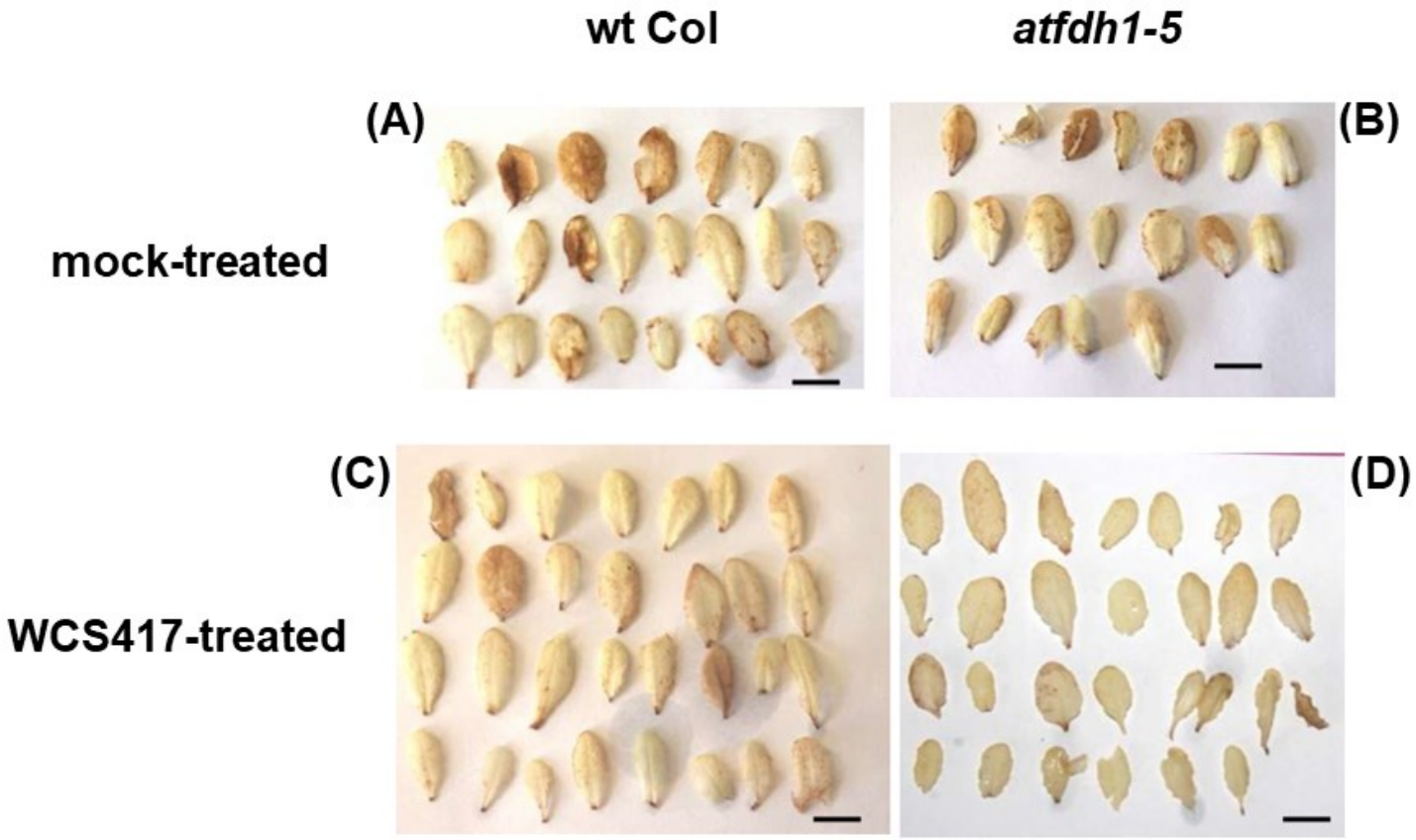


Figure S4. Staining of *A. thaliana* wt Col and *atfdh1*-5 rosette leaves with diaminobenzidine (DAB) after exposure to WCS417 for proteomic analysis. Rosette leaves of *A. thaliana* wt Col and *atfdh1-5* plants sampled after two days root exposure to WCS417 were stained with DAB for H_2O_2 detection.

Protein name	AGI code	UniProt ID	Annotation	Localization (Aramemnon/ SUBA5)	F ratio		Prob > F		DAve	
					wt Col	*atfdh1-5*	wt Col	*atfdh1-5*	wt Col	*atfdh1-5*
EIF4A1	AT3G13920	A8MRZ7	Transl. initiation factor 4A1	nucleus/cytosol	11	19	$8.0E^{-03}$	$1.3E^{-03}$	-0.41	-0.79
GSTF7	AT1G02920	Q9SRY5	Glutathione S-transferase F7	mitoch/cytosol	11	15	$7.6E^{-03}$	$3.3E^{-03}$	-2.0	-2.0
GSTF8	AT2G47730	Q96266	Glutathione S-transferase F8	plastid/plastid	16	14	$2.6E^{-03}$	$3.7E^{-03}$	-0.96	-1.09
PAP6	AT3G23400	Q9LW57	Plastid-lipid-associated protein 6	plastid/plastid	16	51	$2.5E^{-03}$	$3.1E^{-05}$	-1.18	-1.60
RPI3	AT3G04790	Q9S726	Putative ribose-5-phosphate isomerase 3	plastid/plastid	15	15	$3.1E^{-03}$	$3.1E^{-03}$	-0.89	-0.91
RPS1	AT5G30510	Q93VC7	30S ribosomal protein S1	plastid/plastid	14	19	$4.1E^{-03}$	$1.4E^{-03}$	-1.16	-1.18
RPS18C	AT4G09800	P34788	40S ribosomal protein S18	mitoch/cytosol	52	29	$2.9E^{-05}$	$2.9E^{-04}$	-1.20	-0.53
VHA-B1	AT1G76030	P11574	V-type proton ATPase subunit B1	SP/Golgi or vacuole	22	11	$8.8E^{-04}$	$7.1E^{-03}$	-2.0	-0.95
VHA-C	AT1G12840	Q9SDS7	V-type proton ATPase subunit C	mitoch/Golgi or vacuole	35	43	$1.4E^{-04}$	$6.7E^{-05}$	-1.46	-1.72
atpB	ATCG00480	P19366	ATP synthase subunit beta	/plastid	18	13	$1.8E^{-03}$	$4.9E^{-03}$	-0.28	-0.20

Table 1. Proteins with increased expression in *A. thaliana* wt Col and *atfdh1-5* leaves, after root exposure to WCS417. The list of proteins identified by proteomic analysis which are upregulated in the leaves of both wt Col and *atfdh1-5* after two days of root exposure to WCS417 (or to mock treatment), are reported; the name, AGI code, UniProt ID, annotation, subcellular localization according to Aramemnon (http://aramemnon.uni-koeln.de) and SUBA5 (https://suba.live/) (according to their respective highest scores), F ratio and Probability ($P \leq 0.01$) (by LDA), and DAve index (by MAProMa) are reported. SP: secretory pathway; PM: plasma membrane.

Protein name	AGI code	UniProt ID	Annotation	Localization (Aramemnon/ SUBA5)	F ratio		Prob > F		DAve	
					wt Col	*atfdh1-5*	wt Col	*atfdh1-5*	wt Col	*atfdh1-5*
CAM7	AT3G43810	A0A1I9LPJ2	Calmodulin 7	SP/cytosol	13	26	$5.3E^{-03}$	$4.8E^{-04}$	1.03	0.91
CP29B	AT2G37220	Q9ZUU4	RNA-binding protein CP29B	plastid/plastid	14	36	$4.1E^{-03}$	$1.3E^{-04}$	0.37	0.49
CPN10-2	AT2G44650	O80504	10 kDa chaperonin 2	plastid/plastid	13	19	$4.6E^{-03}$	$1.5E^{-03}$	1.49	1.63
CPN20	AT5G20720	O65282	20 kDa chaperonin	plastid/plastid	12	61	$5.4E^{-03}$	$1.5E^{-05}$	0.36	0.76
CSP2	AT4G38680	Q41188	Cold shock protein 2	nucleus/nucleus	12	13	$6.0E^{-03}$	$4.7E^{-03}$	0.73	0.63
ERD10	AT1G20450	P42759	Early responsive to dehydration protein 10	nucleus/cytosol or PM	34	55	$1.7E^{-04}$	$2.4E^{-05}$	0.98	0.95
MEE59	AT4G37300	O23157	maternal effect embryo arrest 59	nucleus/nucleus	25	61	$5.4E^{-04}$	$1.5E^{-05}$	1.44	2.0
F13I12.120	AT3G47070	Q9SD66	thylakoid soluble phosphoprotein	SP/plastid	11	25	$7.2E^{-03}$	$5.7E^{-04}$	0.96	1.42
PSAE1	AT4G28750	Q9S831	PSI subunit E1	plastid/plastid	56	262	$2.2E^{-05}$	$1.7E^{-08}$	0.62	0.70
PSAE2	AT2G20260	Q9S714	PSI subunit E2	plastid/plastid	14	68	$3.7E^{-03}$	$9.1E^{-06}$	0.60	0.78
PSB27-1	AT1G03600	Q9LR64	PSII repair protein PSB27-H1	plastid/plastid	32	32	$2.1E^{-04}$	$2.2E^{-04}$	0.63	0.94
PSBO2	AT3G50820	Q9S841	Oxygen-evolving enhancer protein 1-2	plastid/plastid	12	155	$5.4E^{-03}$	$2.0E^{-07}$	0.31	0.46

PSBP1	AT1G06680	Q42029	Oxygen-evolving enhancer protein 2-1	plastid/plastid	32	28	$2.0E^{-04}$	$3.3E^{-04}$	0.55	0.78
RGGA	AT4G16830	O23523	RGG repeats nuclear RNA binding protein A	nucleus/nucleus	12	94	$6.7E^{-03}$	$2.2E^{-06}$	1.11	1.89
RGGC	AT5G47210	Q9LVT8	putative RGG repeats nuclear RNA binding protein C	plastid/cytosol	10	30	$9.7E^{-03}$	$2.7E^{-04}$	0.92	1.61
RRF	AT3G63190	Q9M1X0	Ribosome-recycling factor	plastid/plastid	14	53	$3.6E^{-03}$	$2.7E^{-05}$	1.23	1.22
At2g21530	AT2G21530	Q8GWP4	unknown protein	plastid/plastid	33	13	$1.8E^{-04}$	$4.8E^{-03}$	1.44	2.0
At5g02570	AT5G02570	Q9LZ45	Histone H2B.9	nucleus/nucleus	16	22	$2.4E^{-03}$	$8.9E^{-04}$	2.0	1.29
At5g24165	AT5G24165	Q8LDQ8	unknown protein	plastid/mitoch	12	14	$6.7E^{-03}$	$4.0E^{-03}$	1.17	1.15

Table 2. Proteins with decreased expression in *A. thaliana* wt Col and *atfdh1-5* leaves, after root exposure to WCS417. The list of proteins identified by proteomic analysis which are downregulated in the leaves of both wt Col and *atfdh1-5* after two days of root exposure to WCS417 (or to mock treatment), are reported; the name, AGI code, UniProt ID, annotation, subcellular localization according to Aramemnon (http://aramemnon.uni-koeln.de) and SUBA5 (https://suba.live/) (according to their respective highest scores), F ratio and Probability ($P \leq 0.01$) (by LDA), and DAve index (by MAProMa) are reported. SP: secretory pathway; PM: plasma membrane.

Protein name	AGI code	UniProt ID	Annotation	Localization (Aramemnon/ SUBA5)	PSMs mock		PSMs WCS417	
					wt Col	*atfdh1-5*	wt Col	*atfdh1-5*
APX1	AT1G07890	Q05431	L-ascorbate peroxidase 1	mitoch/cytosol	3.7	8.5*	6.2	12.5*
GSTF2	AT4G02520	P46422	Glutathione S-transferase F2	mitoch/cytosol	7.3*	21.5	22.4*	19.5
GSTF7	AT1G02920	Q9SRY5	Glutathione S-transferase F7	mitoch/cytosol	/	/	1.8*	2.8*
GSTF8	AT2G47730	Q96266	Glutathione S-transferase F8	plastid/plastid	2.0*	2.6*	5.7*	8.9*
GSTF9	AT2G30860	O80852	Glutathione S-transferase F9	mitoch/cytosol	2.5*	9.6	7.6*	7.9
GSTF10	AT2G30870	P42761	Glutathione S-transferase F10	mitoch/cytosol	0.9	2.2*	2.6	5.1*
GSTU19	AT1G78380	Q9ZRW8	Glutathione S-transferase U19	SP/cytosol	0.3	0.2*	0.4	4.5*

Table 3. Proteins with ROS detoxification function that are differentially expressed in *A. thaliana* wt Col and/or *atfdh1-5* leaves after two days root exposure to WCS417. The name, AGI code, UniProt ID, annotation, subcellular localization according to Aramemnon (http://aramemnon.uni-koeln.de) and SUBA5 (https://suba.live) (according to their respective highest scores) and average Peptide Spectrum Matches (PSMs) per condition are reported. Statistically significant values between mock and WCS417-treated are indicated with *, whereas / indicates that the given protein is not detected in the proteomic analysis under that experimental condition. SP: secretory pathway.

Literature cited

Alekseeva, A. A., Savin, S. S., and Tishkov, V. I. 2011. NAD^+-dependent formate dehydrogenase from plants. Acta Nat. 3:38-54.

Ali, S., Ahmad Ganai, B., Kamili, A. N., Ali Bhat, A., Ahmad Mir, Z., Akhter Bhat, J., Tyagi, A., Tajamul Islam, S., Mushtaq, M., Yadav, P., Rawat, S., and Grover, A. 2018. Pathogenesis-related proteins and peptides as promising tools for engineering plants with multiple stress tolerance, Microbiol. Res., 212:29-37.

Ambard-Bretteville, F., Sorin, C., Rébeillé, F., Hourton-Cabassa, C., and Colas Des Francs-Small, C. 2003. Repression of formate dehydrogenase in *Solanum tuberosum* increases steady-state levels of formate and accelerates the accumulation of proline in response to osmotic stress. Plant Mol. Biol. 52:1153-1168.

Ambrosone, A., Batelli, G., Nurcato, R., Aurilia, V., Punzo, P., Bangarusamy, D. K., Ruberti, I., Sassi, M., Leone, A., Costa, A., and Grillo, S. 2015. The *Arabidopsis* RNA-binding protein AtRGGA regulates tolerance to salt and drought stress. Plant Physiol. 168:292-306.

Andreadeli, A., Flemetakis, E., Axarli, I., Dimou, M., Udvardi, M. K., Katinakis, P., and Labrou, N. E. 2009. Cloning and characterization of *Lotus japonicus* formate dehydrogenase: a possible correlation with hypoxia. Biochim. Biophys. Acta. Proteins and Proteomes 1794:976-984.

Baek, D., Rokibuzzaman, M., Khan, A., Kim, M. C., Park, H. J., Yun, D. J., and Chung, Y. R. 2020. Plant-growth promoting *Bacillus oryzicola* YC7007 modulates stress-response gene expression and provides protection from salt stress. Front. Plant Sci. 10:1646.

Baubec, T., Dinh, H. Q., Pecinka, A., Rakic, B., Rozhon, W., Wohlrab, B., von Haeseler, A., and Scheid, O. M. 2010. Cooperation of multiple chromatin modifications can generate unanticipated stability of epigenetic states in *Arabidopsi*s. Plant Cel. 22:34-47.

Chen, Y. E., Zhao, Z. Y., Zhang, H. Y., Zeng, X. Y., and Yuan, S. 2013. Significance of CP29 reversible phosphorylation in thylakoids of higher plants under environmental stresses. J. Exp. Bot. 64:1167-1178.

Choi, D. S., Kim, N. H., and Hwang, B. K. 2014. Pepper mitochondrial formate dehydrogenase1 regulates cell death and defense responses against bacterial pathogens. Plant Physiol. 166:1298-1311.

Cormann, K. U., Möller, M., and Nowaczyk, M. M. 2016. Critical assessment of protein cross-linking and molecular docking: An updated model for the interaction between photosystem II and Psb27. Front. Plant Sci. 7.157.

David, P., Colas Des Francs-Small, C., Sévignac, M., Thareau, V., Macadré, C., Langin, T., and Geffroy, V. 2010. Three highly similar formate dehydrogenase genes located in the vicinity of the B4 resistance gene cluster are differentially expressed under biotic and abiotic stresses in Phaseolus vulgaris. Theor. Appl. Genet. 121:87-103.

de Weert, S., Vermeiren, H., Mulders, I. H. M., Kuiper, I., Hendrickx, N., Bloemberg, G. V., Vanderleyden, J., de Mot, R., and Lugtenberg, B. J. J. 2002. Flagella-driven chemotaxis towards exudate components is an important trait for tomato root colonization by Pseudomonas fluorescens. Mol. Plant Microbe Interact. 15:1173-1180.

Dietz, K. J., Tavakoli, N., Kluge, C., Mimura, T., Sharma, S. S., Harris, G. C., Chardonnens, A. N., and Golldack, D. 2001. Significance of the V-type ATPase for the adaptation to stressful growth conditions and its regulation on the molecular and biochemical level. J. Exp. Bot. 52:1969-1980.

Di Silvestre, D., Vigani, G., Mauri, P., Hammadi, S., Morandini, P., and Murgia, I. 2021. Network topological analysis for the identification of novel hubs in plant nutrition. Front. Plant Sci. 12. 629013.

Doncheva, N. T., Morris, J. H., Gorodkin, J., and Jensen, L. J. 2019. Cytoscape StringApp: Network analysis and visualization of proteomics data. J. Proteome Res. 18:623-632.

Elorza, A., León, G., Gómez, I., Mouras, A., Holuigue, L., Araya, A., and Jordana, X. 2004. Nuclear SDH2-1 and SDH2-2 genes, encoding the iron-sulfur subunit of mitochondrial complex II in *Arabidopsis*, have distinct cell-specific expression patterns and promoter activities. Plant Physiol. 136:4072-4087.

Gullner, G., Komives, T., Király, L., and Schröder, P. 2018. Glutathione S-transferase enzymes in plant-pathogen interactions. Front. Plant Sci. 9:1836.

Herman, P. L., Ramberg, H., Baack, R. D., Markwell, J., and Osterman, J. C. 2002. Formate dehydrogenase in *Arabidopsis thaliana*: Overexpression and subcellular localization in leaves. Plant Sci. 163:1137-1145.

Hong, X., Qi, F., Wang, R., Jia, Z., Lin, F., Yuan, M., Xin, X., and Liang, Y. 2022. Ascorbate peroxidase 1 allows monitoring of cytosolic accumulation of effector-triggered reactive oxygen species using a luminol-based assay. Plant Physiol. doi: 10.1093/plphys/kiac551.

Hourton-Cabassa, C., Ambard-Bretteville, F., Moreau, F., de Virville, J.D., Rémy, R., and Colas Des Francs-Small, C. 1998. Stress induction of mitochondrial formate dehydrogenase in potato leaves. Plant Physiol. 116:627-635.

Huang, J. B., Zou, Y., Zhang, X., Wang, M., Dong, Q., and Tao, L. Z. 2020. RIBOSE PHOSPHATE ISOMERASE 1 influences root development by acting on cell wall

biosynthesis, actin organization, and auxin transport in Arabidopsis. Front. Plant Sci. 10:1641.

Jin, C. W., He, Y.F., Tang, C. X., Wu, P., and Zheng, S. J. 2006. Mechanisms of microbially enhanced Fe acquisition in red clover (*Trifolium pratense* L.). Plant Cell Environ. 29:888-897.

Jin, C. W., Li, G. X., Yu, X. H., and Zheng, S. J. 2010. Plant Fe status affects the composition of siderophore-secreting microbes in the rhizosphere. Ann. Bot. 105:835-841.

Kandasamy, S., Loganathan, K., Muthuraj, R., Duraisamy, S., Seetharaman, S., Thiruvengadam, R., Ponnusamy, B., and Ramasamy, S. 2009. Understanding the molecular basis of plant growth promotional effect of Pseudomonas fluorescens on rice through protein profiling. Proteome Sci. 7:47.

Kovacs, D., Kalmar, E., Torok, Z., and Tompa, P. 2008. Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. Plant Physiol. 147:381-390.

Kurt-Gür, G., Demirci, H., Sunulu, A., and Ordu, E. 2018. Stress response of NAD + -dependent formate dehydrogenase in Gossypium hirsutum L. grown under copper toxicity. Environ. Sci. Pollut. Res. 25:31679-31690.

Kushwaha, R., Singh, A., and Chattopadhyay, S. 2008. Calmodulin7 plays an important role as transcriptional regulator in Arabidopsis seedling development. Plant Cell. 20:1747-1759.

Lee, S., Vemanna, R. S., Oh, S., Rojas, C. M., Oh, Y., Kaundal, A., Kwon, T., Lee, H. K., Senthil-Kumar, M., and Mysore, K. S. 2022. Functional role of formate dehydrogenase 1 (FDH1) for host and nonhost disease resistance against bacterial pathogens. PLoS One. 17:e0264917.

Li, R., Moore, M., Bonham-Smith, P. C., and King, J. 2002. Overexpression of formate dehydrogenase in *Arabidopsis thaliana* resulted in plants tolerant to high concentrations of formate. J. Plant Physiol. 159:1069-1076.

Li, Y., Zeng, H., Xu, F., Yan, F., and Xu, W. 2022. H^+-ATPases in plant growth and stress responses. Annu. Rev. Plant Biol. 73:495-521.

Littlejohn, G. R., Breen, S., Smirnoff, N., and Grant, M. 2020. Chloroplast immunity illuminated. New Phytol. 229:3088-3107.

Liu, W., Sikora, E., and Park, S. W. 2020. Plant growth-promoting rhizobacterium, *Panebacillus polymyxa* CR1, upregulates dehydration-responsive genes, RD29A and RD29B, during priming drought tolerance in *Arabidopsi*s. Plant Physiol. Biochem. 156:146-154.

Lou, H. Q., Gong, Y. L., Fan, W., Xu, J. M., Liu, Y., Cao, M. J., Wang, M. H., Yang, J. L., and Zheng, S. J. 2016. A formate dehydrogenase confers tolerance to aluminum and low pH. Plant Physiol. 171:294-305.

Lundin, B., Hansson, M., Schoefs, B., Vener, A. V., and Spetea, C. 2007. The *Arabidopsis* PsbO2 protein regulates dephosphorylation and turnover of the photosystem II reaction centre D1 protein. Plant J. 49:528-539.

Ma, B., Qian, D., Nan, Q., Tan, C., An, L., and Xiang, Y. 2012. *Arabidopsis* vacuolar H^+-ATPase (V-ATPase) B subunits are involved in actin cytoskeleton remodeling via binding to, bundling, and stabilizing F-actin. J. Biol. Chem. 287: 19008-19017.

Marzorati, F., Vigani, G., Morandini, P., and Murgia, I. 2021. Formate dehydrogenase contributes to the early Arabidopsis thaliana responses against Xanthomonas campestris pv campestris infection. Physiol. Mol. Plant Pathol. 114:101633.

Mauri, P., and Dehò, G. 2008. A proteomic approach to the analysis of RNA degradosome composition in *Escherichia coli*. Methods Enzymol. 447:99-117.

Miché, L., Battistoni, F., Gemmer, S., Belghazi, M., and Reinhold-Hurek, B. 2006. Upregulation of jasmonate-inducible defense proteins and differential colonization of roots of *Oryza sativa* cultivars with the endophyte *Azoarcus* sp. Mol. Plant Microbe Interact. 19:502-511.

Murgia, I., Tarantino, D., Vannini, C., Bracale, M., Carravieri, S., and Soave, C. 2004. *Arabidopsis thaliana* plants overexpressing thylakoidal ascorbate peroxidase show increased resistance to Paraquat-induced photooxidative stress and to nitric oxide-induced cell death. Plant J. 38:940-953.

Murgia, I., Vigani, G., Di Silvestre, D., Mauri, P., Rossi, R., Bergamaschi, A., Frisella, M., and Morandini, P. 2020. Formate dehydrogenase takes part in molybdenum and iron homeostasis and affects dark-induced senescence in plants. J. Plant Interact. 15:386-397.

Olson, B. J., Skavdahl, M., Ramberg, H., Osterman, J.C., and Markwell, J. 2000. Formate dehydrogenase in *Arabidopsis thaliana*: characterization and possible targeting to the chloroplast. Plant Sci. 159:205-212.

Pagnussat, G. C., Yu, H. J., Ngo, Q. A., Rajani, S., Mayalagu, S., Johnson, C. S., Capron, A., Xie, L. F., Ye, D., and Sundaresan, V. 2005. Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. Development. 132:603-614.

Palma, C., la Rocca, C., Gigantino, V., Aquino, G., Piccaro, G., Di Silvestre, D., Brambilla, F., Rossi, R., Bonacina, F., Lepore, M. T., Audano, M., Mitro, N., Botti, G., Bruzzaniti, S., Fusco, C., Procaccini, C., de Rosa, V., Galgani, M., Alviggi, C., Puca, A., Grassi, F., Rezzonico-Jost, T., Norata, G. D., Mauri, P., Netea, M. G., de Candia, P., and Matarese,

G. 2021. Caloric restriction promotes immunometabolic reprogramming leading to protection from tuberculosis. Cell Metab. 33:300-318.

Palmer, C. M., Hindt, M. N., Schmidt, H., Clemens, S., and Guerinot, M. L. 2013. MYB10 and MYB72 are required for growth under iron-limiting conditions. PLoS Genet. 9:e1003953.

Pieterse, C. M. J., Berendsen, R. L., Jonge, R. de, and Stringlis, I. A. 2020. *Pseudomonas simiae* WCS417: star track of a model beneficial rhizobacterium. Plant Soil. 461:245-263.

Romera, F. J., García, M. J., Lucena, C., Martínez-Medina, A., Aparicio, M. A., Ramos, J., Alcántara, E., Angulo, M., and Pérez-Vicente, R. 2019. Induced systemic resistance (ISR) and Fe deficiency responses in dicot plants. Front Plant Sci. 10:287.

Saleem, M., Fariduddin, Q., and Castroverde, C. D. M 2021. Salicylic acid: a key regulator of redox signalling and plant immunity. Plant Physiol. Biochem. 168:381-397.

Sappl, P. G., Oñate-Sánchez, L., Singh, K. B., and Millar, A. H. 2004. Proteomic analysis of glutathione S -transferases of *Arabidopsis thaliana* reveals differential salicylic acid-induced expression of the plant-specific phi and tau classes. Plant Mol. Biol. 54:205-219.

Sasaki, K., Kim, M.H., and Imai, R. 2013. Arabidopsis COLD SHOCK DOMAIN PROTEIN 2 is a negative regulator of cold acclimation. New Phytol. 198:95-102.

Sharon, M., Freeman, S., and Sneh, B. 2011. Assessment of resistance pathways induced in *Arabidopsis thaliana* by hypovirulent *Rhizoctonia* spp. isolates. Phytopathology. 101:828-838.

Simons, M., van der Bij, A. J., Brand, I., de Weger, L. A., Wijffelman, C. A., and Lugtenberg, B. J. J. 1996. Gnotobiotic system for studying rhizosphere colonization by plant growth-promoting Pseudomonas bacteria. Mol. Plant Microbe Interact. 9:600-607.

Singh, P., Arif, Y., Miszczuk, E., Bajguz, A., and Hayat, S. 2022. Specific roles of lipoxygenases in development and responses to stress in plants. Plants. 11:979.

Singh, D. K., Maximova, S. N., Jensen, P. J., Lehman, B. L., Ngugi, H. N., and McNellis, T. W. 2010. FIBRILLIN4 is required for plastoglobule development and stress resistance in apple and Arabidopsis. Plant Physiol. 154:1281-1293.

Stringlis, I. A., Proietti, S., Hickman, R., van Verk, M. C., Zamioudis, C., and Pieterse, C. M. J. 2018. Root transcriptional dynamics induced by beneficial rhizobacteria and microbial immune elicitors reveal signatures of adaptation to mutualists. Plant J. 93:166-180.

Sun, Z., Li, S., Chen, W., Zhang, J., Zhang, L., Sun, W., and Wang, Z. 2021. Molecular sciences plant dehydrins: Expression, regulatory networks, and protective roles in plants challenged by abiotic stress. J. Mol. Sci. 22:12619.

Suzuki, K., Itai, R., Suzuki, K., Nakanishi, H., Nishizawa, N. K., Yoshimura, E., and Mori, S. 1998. Formate dehydrogenase, an enzyme of anaerobic metabolism, is induced by iron deficiency in barley roots. Plant Physiol. 116:725-732.

Sylvestre-Gonon, E., Law, S. R., Schwartz, M., Robe, K., Keech, O., Didierjean, C., Dubos, C., Rouhier, N., and Hecker, A. 2019. Functional, structural and biochemical features of plant serinyl-glutathione transferases. Front Plant Sci. 10:608.

Trapet, P., Avoscan, L., Klinguer, A., Pateyron, S., Citerne, S., Chervin, C., Mazurier, S., Lemanceau, P., Wendehenne, D., and Besson-Bard, A. 2016. The *Pseudomonas fluorescens* siderophore pyoverdine weakens *Arabidopsis thaliana* defense in favor of growth in iron-deficient conditions. Plant Physiol. 171:675-693.

Trapet, P. L., Verbon, E. H., Bosma, R. R., Voordendag, K., van Pelt, J. A., and Pieterse, C. M. J. 2021. Mechanisms underlying iron deficiency-induced resistance against pathogens with different lifestyles. J. Exp. Bot. 72:2231-2241.

Vallad, G. E., and Goodman, R. M. 2004. Systemic acquired resistance and induced systemic resistance in conventional agriculture. Crop Sci. 44:1920-1934.

van Hulten, M., Chatterjee, S., and van den Burg, H. A. 2019. Infection assay for *Xanthomonas campestris pv. campestris* in *Arabidopsis thaliana* mimicking natural entry via hydathodes. Methods Mol. Biol. 1991:159-185.

Verbon, E. H., Trapet, P. L., Kruijs, S., Temple-Boyer-Dury, C., Rouwenhorst, T. G., and Pieterse, C. M. J. 2019. Rhizobacteria-mediated activation of the Fe deficiency response in *Arabidopsis* roots: Impact on Fe status and signaling. Front Plant Sci. 10:909.

Vigani, G., Di Silvestre, D., Agresta, A. M., Donnini, S., Mauri, P., Gehl, C., Bittner, F., and Murgia, I. 2017. Molybdenum and iron mutually impact their homeostasis in cucumber (*Cucumis sativus*) plants. New Phytol. 213:1222-1241.

Vlot, A. C., Sales, J. H., Lenk, M., Bauer, K., Brambilla, A., Sommer, A., Chen, Y., Wenig, M., and Nayem, S. 2021. Systemic propagation of immunity in plants. New Phytol. 229:1234-1250.

Wang, J., Lan, P., Gao, H., Zheng, L., Li, W., and Schmidt, W. 2013. Expression changes of ribosomal proteins in phosphate-and iron-deficient *Arabidopsis* roots predict stress-specific alterations in ribosome composition. BMC Genom. 14:783.

Wang, H., Liu, R., You, M.P., Barbetti, M. J., and Chen, Y. 2021. Pathogen biocontrol using plant growth-promoting bacteria (PGPR): Role of bacterial diversity. Microorganisms. 9:1988.

Wang, L., Ouyang, M., Li, Q., Zou, M., Guo, J., Ma, J., Lu, C., and Zhang, L. 2010. The *Arabidopsis* chloroplast ribosome recycling factor is essential for embryogenesis and chloroplast biogenesis. Plant Mol Biol. 74:47-59.

Wasternack, C., and Hause, B. 2019. The missing link in jasmonic acid biosynthesis. Nat. Plants. 5:776-777.

van Wees, S. C. M., Luijendijk, M., Smoorenburg, I., van Loon, L. C., and Pieterse, C. M. J. 1999. Rhizobacteria-mediated induced systemic resistance (ISR) in *Arabidopsis* is not associated with a direct effect on expression of known defense-related genes but stimulates the expression of the jasmonate-inducible gene Atvsp upon challenge. Plant Mol. Biol. 41:537-549.

van Wees, S. C. M., van Pelt, J. A., Bakker, P. A. H. M., and Pieterse, C. M. J. 2013. Bioassays for assessing jasmonate-dependent defenses triggered by pathogens, herbivorous insects, or beneficial rhizobacteria. Methods Mol. Biol. 1011:35-49.

Wintermans, P. C. A., Bakker, P. A. H. M., and Pieterse, C. M. J. 2016. Natural genetic variation in *Arabidopsis* for responsiveness to plant growth-promoting rhizobacteria. Plant Mol. Biol. 90:623-634.

Wu, X., Xiong, E., Wang, W., Scali, M., and Cresti, M. 2014. Universal sample preparation method integrating trichloroacetic acid/acetone precipitation with phenol extraction for crop proteomic analysis. Nat. Protoc. 9:362-374.

Yang, T. H., Lenglet-Hilfiker, A., Stolz, S., Glauser, G., and Farmer, E. E. 2020. Jasmonate precursor biosynthetic enzymes LOX3 and LOX4 control wound-response growth restriction. Plant Physiol. 184:1172-1180.

Yang, F., Xiao, K., Pan, H., and Liu J. 2021. Chloroplasts: the emerging battlefield in plant-microbe interactions. Front. Plant Sci. 12:637853.

Yi, X., Hargett, S. R., Liu, H., Frankel, L. K., and Bricker, T. M. 2007. The PsbP protein is required for photosystem II complex assembly/stability and photoautotrophy in *Arabidopsis thaliana*. J. Biol. Chem. 282:24833-24841.

Zamioudis, C., Mastranesti, P., Dhonukshe, P., Blilou, I., and Pieterse, C. M. J. 2013. Unraveling root developmental programs initiated by beneficial *Pseudomonas* spp. bacteria. Plant Physiol. 162:304-318.

Zamioudis, C., Hanson, J., and Pieterse, C. M. J. 2014. ß-Glucosidase BGLU42 is a MYB72-dependent key regulator of rhizobacteria-induced systemic resistance and modulates iron deficiency responses in *Arabidopsis* roots. New Phytol. 204:368-379.

Zamioudis, C., Korteland, J., van Pelt, J. A., van Hamersveld, M., Dombrowski, N., Bai, Y., Hanson, J., van Verk, M.C., Ling, H. Q., Schulze-Lefert, P., and Pieterse, C. M. J. 2015. Rhizobacterial volatiles and photosynthesis-related signals coordinate MYB72 expression in *Arabidopsis* roots during onset of induced systemic resistance and iron-deficiency responses. Plant J. 84:309-322.

Zargar, S. M., Kurata, R., Inaba, S., and Fukao, Y. 2013. Unraveling the iron deficiency responsive proteome in Arabidopsis shoot by iTRAQ-OFFGEL approach. Plant Signal. Behav. 8:e26892.

Zhang, H., Sun, Y., Xie, X., Kim, M. S., Dowd, S. E., and Paré, P. W. 2009. A soil bacterium regulates plant acquisition of iron via deficiency-inducible mechanisms. Plant J. 58:568-577.

3.3 Reviews

"Now comes the reign of iron."

— John Ericsson

Manuscript 5

Murgia I, Marzorati F, Vigani G, Morandini P. 2022. Plant iron nutrition: the long road from soil to seeds. *Journal of Experimental Botany* **73**, 1809-1824.

Manuscript 6

Marzorati F, Midali A, Morandini P, Murgia I. 2022. Good or bad? The double face of iron in plants. *Frontiers for Young Minds* **10**, 718162.

Manuscript 5

Plant iron nutrition: the long road from soil to seeds

Irene Murgia[1,*,†], Francesca Marzorati[2], Gianpiero Vigani[3], and Piero Morandini[2]

[1] Department of Biosciences, University of Milano, Milano, Italy
[2] Department of Environmental Science and Policy, University of Milano, Milano, Italy
[3] Plant Physiology Unit, Department of Life Sciences and Systems Biology, University of Torino, Torino, Italy
† Correspondence: irene.murgia@unimi.it
∗ Present address: Department of Environmental Science and Policy, University of Milano, Milano, Italy

Abstract

Iron (Fe) is an essential plant micronutrient since many cellular processes including photosynthesis, respiration, and the scavenging of reactive oxygen species depend on adequate Fe levels; however, non-complexed Fe ions can be dangerous for cells, as they can act as pro-oxidants. Hence, plants possess a complex homeostatic control system for safely taking up Fe from the soil and transporting it to its various cellular destinations, and for its subcellular compartmentalization. At the end of the plant's life cycle, maturing seeds are loaded with the required amount of Fe needed for germination and early seedling establishment. In this review, we discuss recent findings on how the microbiota in the rhizosphere influence and interact with the strategies adopted by plants to take up iron from the soil. We also focus on the process of seed-loading with Fe, and for crop species we also consider its associated metabolism in wild relatives. These two aspects of plant Fe nutrition may provide promising avenues for a better comprehension of the long pathway of Fe from soil to seeds.

Keywords: Embryos; Iron; Microbiota; Micronutrients; Plant immunity; *Pseudomonas simiae* WCS417; Rhizosphere; Seeds; Trace metal

Abbreviations: ASC, ascorbate; ET, ethylene; FIT, FER-like iron deficiency-induced transcription factor; ISR, induced systemic resistance; NA, nicotianamine; PGPR, plant growth-promoting rhizobacteria; ROS, reactive oxygen species; SAR, systemic acquired resistance; WCS417, *Pseudomonas simiae* WCS417

Invited review
Special Issue of *Journal of Experimental Botany*:
Essential trace metals: micronutrients with large impact

Plant iron nutrition in the long road from soil to seeds

Irene Murgia*[1], Francesca Marzorati[2], Gianpiero Vigani[3] and Piero Morandini[2]
[1]Dept. Biosciences, University of Milano, Italy; [2]Dept. Environmental Science and Policy, University of Milano, Italy; [3]Plant Physiology Unit, Dept. Life Sciences and Systems Biology, University of Torino, Italy.
author for correspondence: Irene Murgia irene.murgia@unimi.it.

authors email addresses:
irene.murgia@unimi.it, francesca.marzorati1@unimi.it, gianpiero.vigani@unito.it,
piero.morandini@unimi.it.

Date of resubmission: 22/11/2021
Number of tables and figures: 3 figures.
Word count (including Box 1): 7327

Running title: Plant iron nutrition from soil to seeds

Highlight

The interactions between rhizosphere microbiota and plants, the seeds loading with iron represent relevant lines of research, also in wild crop relatives, for the full understanding of plant iron nutrition.

Abstract

Iron (Fe) is an essential plant micronutrient since photosynthesis, respiration, the scavenging of reactive oxygen species and many other cellular processes depend on adequate Fe levels. Nonetheless, non-complexed Fe ions can be dangerous for cells, as they can act as a pro-oxidant. Therefore, plants possess a complex homeostatic control system for safely taking up Fe from the soil, transporting it to the various cellular destinations and for its subcellular compartmentalization. At the end of the plant's life cycle, maturing seeds are loaded with the required amount of Fe for germination and early seedling establishment. In this review, we discuss recent findings on how the microbiota in the rhizosphere influence and interact with the strategies adopted by plants to take up iron from the soil. We also focus on the process of seed loading with Fe and take into account the Fe metabolism in wild crops' relatives. These aspects of plant Fe nutrition can represent promising avenues for a better comprehension of the long road of Fe from soil to seeds.

Keywords: iron, embryos, microbiota, micronutrients, plant immunity, *Pseudomonas simiae* WCS417, rhizosphere, seeds.

Abbreviations: ABA: abscisic acid; AMF: Arbuscular mycorrhizal fungi; ASC: ascorbate; BGLU42β-GLUCOSIDASE42; DMA: deoxymugeinic acid; ET: Ethylene; ETI: Effector-triggered immunity; FIT: FER-like Iron Deficiency-Induced Transcription factor; GA: Gibberellic acid; ISR: Induced systemic resistance; JA: Jasmonic acid; MAMPs: Microbe-Associated Molecular Patterns; MTI: MAMP-Triggered Immunity; NA: Nicotianamine; NO: Nitric Oxide; PAMPs: Pathogen Associated Molecular Patterns; PGPR: Plant Growth-Promoting Rhizobacteria; PGPF: Plant Growth-Promoting Fungi; PRs: Pathogenesis-related (PR) Proteins; PRRs: Pattern-Recognition Receptors; PS: Phytosiderophores; PTI: PAMP-Triggered Immunity; ROS: Reactive Oxygen Species; SA Salicylic Acid; SAR: Systemic Acquired Resistance; VOC: Volatile Organic Compound; WCS417: *Pseudomonas simiae* WCS417.

Introduction

Iron (Fe) participates in fundamental processes in plants (i.e. respiration, photosynthesis, antioxidant defenses) as well as in many biochemical pathways (e.g., hormones and secondary metabolisms) and is, therefore, an essential micronutrient (Murgia *et al*., 2012; Kobayashi and Nishizawa, 2012; Briat *et al*., 2015; Connorton *et al*., 2017; Vigani and Murgia, 2018; Kobayashi *et al*., 2019). Iron can exert this role in various chemical forms, such as Fe-heme groups, Fe-S clusters or nitrosyl-Fe complexes (Ramirez *et al*., 2011). Nonetheless, especially when in a free non-complexed form, Fe represents a severe threat to cells due to its pro-oxidant action (Lodde *et al*., 2021). For these reasons, Fe uptake from the soil, its transport and distribution to various plant organs and tissues, its subcellular compartmentalization and seed loading with Fe are all utterly regulated processes, role of which is to ensure that plant cells receive enough Fe in the safest chemical form. Iron deficiency causes chlorosis in plants, with adverse consequences for plant health and growth, leading to yield loss (Ramirez *et al*., 2011; Vigani *et al*., 2013; Vigani and Murgia, 2018). Iron excess is also detrimental and it leads to overproduction of reactive oxygen species (ROS), damage to macromolecules, the "bronzing" symptoms, upregulation of ROS scavenging systems and downregulation of Fe uptake genes (Murgia *et al*., 2002; Arnaud *et al*., 2006; Ramirez *et al*., 2011; Aung and Masuda, 2020; Lodde *et al*., 2021). An accurate control of Fe homeostasis reduces the risk of progressive damage caused by cellular Fe excess/deficiency and it can also reduce the metabolic costs for keeping such damages under control.

Although Fe is abundant in soils, mostly present as ferric (hydro)oxides, its availability to plants is limited, due to an extremely low solubility of such oxides; for example, $Fe(OH)_3$ K_{sp} is $4x10^{-38}$ (Lindsay and Schwab, 1982; Schwertmann, 1991; Colombo *et al*., 2014) implying that, at neutral or basic pH, the concentration of Fe(III) is extremely low. Mechanisms of plant Fe uptake from soil have been classified as either an 'acidification-reduction strategy' (Strategy I) adopted by non-graminaceous plants, or as a 'chelation strategy' (Strategy II) adopted by Graminaceae. In Strategy I plants, soil acidification by plasma membrane H^+-ATPase is followed by reduction of Fe(III) to Fe(II) and Fe(II) transport into epidermal root cells. Strategy II relies instead on the extrusion of phytosiderophores (PS) by TRANSPORTER OF MUGINEIC ACID (TOM) transporter; PS can chelate Fe(III) and the complex Fe(III)-PS is then transported into cell roots by members of the YELLOW STRIPE-LIKE (YSL) transporter family.

Such strategies are finely regulated at both transcriptional and post-transcriptional levels. As an example, the activation of Fe uptake in *Arabidopsis thaliana* plants through AHA2, FERRIC REDUCTASE OXIDASE2 (FRO2) which reduces Fe(III) to Fe(II), and IRON ROOT TRANSPORTER1 (IRT1) which transports Fe(II) into root cells, is transcriptionally regulated by the basic helix-loop-helix (bHLH) FER-LIKE IRON DEFICIENCY-INDUCED TRANSCRIPTION FACTOR (FIT) (Colangelo and Guerinot, 2004; Jakoby *et al*., 2004, Bauer *et al*., 2007). FIT activation is mediated by the ethylene-responsive transcription factors ETHYLENE INSENSITIVE3 (EIN3) and EIN3-Like1 (EIL1) (Lingam *et al*., 2011). Furthermore, other bHLH proteins (bHLH038, bHLH039, bHLH100 and bHLH101) interact with FIT and their expression increase under Fe deficiency (Wang *et al*., 2007; Wang *et al*., 2013). Most recently, the upstream regulatory role of bHLH121 (UPSTREAM REGULATOR of IRT1, URI) on the Fe homeostasis network has been unveiled and it involves the activation of various genes, among which FIT (Kim *et al*., 2019), through its interaction with bHLH105 (IRL3) (Gao *et al*., 2020); yeast two-hybrid and chromatin immunoprecipitation (ChIP) assays, show that FIT is not a direct target of bHLH121

(Gao *et al*., 2020).

A complex regulation of Fe uptake occurs at post-transcriptional level. IRT1 is present in endosomes/trans-Golgi network compartments (EE/TGN) and its degradation and recycling between EE/TGN and the plasma membrane are modulated by ubiquitination and monoubiquitin-dependent endocytosis (Barberon *et al*., 2011). IRT1 is ubiquitinated on the plasma membrane by the action of IRT1 DEGRADATION FACTOR1 (IDF1), a RING-type E3 ubiquitin ligase. IRT1 also mediates the transport of other metals, like Zinc (Zn), Manganese (Mn) and Cobalt (Co) which accumulate in plant tissues under Fe deficiency (Barberon *et al*., 2014; Vigani and Hanikenne, 2018). To limit IRT1-mediated metals accumulation, IDF1 facilitates its degradation through a negative feedback loop (Barberon *et al*., 2014). Such a mechanism involves other proteins, namely PHOSPHATIDYLINOSITOL-3-PHOSPHATE-BINDING PROTEIN FYVE1 and SORTING NEXIN SNX, required for IRT1 recycling in plants (Barberon *et al*., 2014; Ivanov *et al*., 2014). Due to such a recycling process, IRT1 is part of the metal sensing machinery (Dubeaux *et al*., 2018) and therefore IRT1 has been proposed to be a transceptor for metals homeostasis in plants (Cointry and Vert, 2019). Updated descriptions of both strategies and their multiple-level regulations, can be found in recent reviews (Schwarz and Bauer 2020; Gao and Dubos, 2021; Riaz and Guerinot, 2021).

The boundaries between the two Fe acquisition strategies are fading (Grillet and Schmidt, 2019); for instance, the graminaceous *Oryza sativa* (rice) adopts Strategy II, but it can also induce Fe transporters OsIRT1 and OsIRT2, which are hallmarks of Strategy I (Ishimaru *et al*., 2006; Wairich *et al*., 2019). The combination of both strategies, referred to as 'combined strategy' (CS), appears to be an adaptation to flooded soils, a situation implying oxygen O_2 depletion and a reduction of soil potential, with a consequent increase in Fe(II) concentration (Marìnes-Cunca *et al*., 2015; Wairich *et al*., 2019). Notably, Strategy I plants can also exude various compounds with iron-mobilizing properties (Palmer *et al*., 2013; Zamioudis *et al*., 2015; Sisó-Terraza *et al*., 2016; Stringlis *et al*., 2019; Yu *et al*., 2021). In particular, coumarins are secondary plant metabolites synthesized by the phenylpropanoid pathway that can promote root Fe uptake. Indeed, they are secreted by roots under Fe-deficiency, through ATP-BINDING CASSETTE G37/ PLEIOTROPIC DRUG RESISTANCE 9 (ABCG37/PDR9) transporter (Fourcroy *et al*., 2014; Fourcroy *et al*., 2016; Ziegler *et al*., 2017) and display Fe-mobilizing, chelating and reducing properties (Schmid *et al*., 2014; Tsai and Schmidt, 2017; Ziegler *et al*., 2017; Tsai *et al*., 2018; Rajniak *et al*., 2018; Stassen *et al*., 2021).

The transport of the various coumarin molecules is a complex and dynamic process; Robe *et al*. (2021a) investigated their pattern of accumulation in the various root cell types and they also demonstrated that coumarins can be transported from roots to aerial parts through the xylem; moreover, these authors also showed evidence that various plant species (belonging to both dicotyledons and gymnosperms) can take up coumarins from the rhizosphere (Robe *et al*., 2021a). An updated model of such a complex picture of coumarins distribution in roots and of the transcriptional regulation of their biosynthesis, is proposed by Robe *et al*. (2021b), where outstanding questions regarding the biology of coumarins are also discussed; indeed, various components are still missing from the picture, such as the identities of all the glycosyltransferases and of β-glucosidases involved, as well as the direct regulators of coumarin biosynthesis genes.

A further level of the emerging complexity in Fe acquisition from the soil is represented by the microbial communities growing in the proximity of plant roots, as they can exert beneficial or harmful actions, favouring or inhibiting Fe uptake. Indeed, cooperation and/or competition for nutrients, including Fe, are established among the myriad of microbes living close to roots and

between plants and microorganisms, in a tripartite interaction involving plants, microorganisms and Fe. Given the emerging findings on this subject, we focus first on this ''tug of war'' for iron nutrition (Herlihy *et al*., 2020), which involves microorganisms, their colonization of plant niches and their interaction with plants. Volatile organic compounds (VOCs), emitted by plants and microorganisms, also influence the belowground interactions among plants and microorganisms and are emerging as potent regulators of these multiple interactions. VOCs-dependent microbe-plants interactions are not discussed here; readers are referred to recent publications (Zamioudis *et al*., 2015; Delory *et al*., 2016; Schulz-Bohm *et al*., 2017; Garbeva and Weisskopf, 2020; Gulati *et al*., 2020).

1) Plant Fe uptake, soil and microorganisms: the plant holobiont

Plants co-evolved with the soil microbes living in the 'rhizosphere', i.e. in the soil volume (1-3 mm in thickness), which is adherent to roots and influenced by roots' secretions. The rhizosphere is a remarkable reservoir of microbial biodiversity as it contains up to 10^{11} microorganisms per gram root (Berendsen *et al*., 2012; Sasse *et al*., 2018).

A dense and diversified array of microorganisms is thus present in the soil, which includes organisms belonging to Archaea, Bacteria and Eukarya domains, collectively named 'soil microbiota' (the collection of all their genomes is referred to as the 'microbiome') (Lynch and Pedersen, 2016; Trivedi *et al*., 2020; Pascale *et al*., 2020). The development, health and, ultimately, the phenotype of a plant is influenced by the microbial communities living in the rhizosphere and by the combined expression of both the host plant genome and its associated microbiome (Nihorimbere *et al*., 2011; Berendsen *et al*., 2012; Trivedi *et al*., 2020; Pascale *et al*., 2020). These findings led to the concept of the 'holobiont' (Vandenkoornhuyse *et al*., 2015; Sánchez-Cañizares *et al*., 2017; Simon *et al*., 2019). Complex plant-microorganisms and microorganisms-microorganisms interactions occur in the rhizosphere (Trivedi *et al*., 2020). Root exudates may shape the microbial communities living in the rhizosphere, by serving as nutrients or selective agents: the plant root microbiota are therefore different from the microbiota in the soil far away from plants, according to a phenomenon known as 'the rhizosphere effect' (Bakker *et al*., 2013; Bakker *et al*., 2020). In return, members of the plant microbiota may cause severe diseases, however they usually act as mutualists: plant growth-promoting rhizobacteria (PGPR) and plant growth-promoting fungi (PGPF) within the plant microbiota affect host nutrition, development and the immune system, by promoting plant growth or stimulating defense responses (Jogaiah and Abdelrahman, 2019; Verbon *et al*., 2019; Pascale *et al*., 2020). Beneficial root bacteria are usually a minor fraction in the rhizosphere; nonetheless, several studies have found that these bacteria may positively enhance plant yield and growth, and their pivotal role in plant life has recently been discussed (Van Loon, 2007; Ipek and Esitken, 2017; Majeed *et al*., 2018; Compant *et al*., 2019; do Amaral *et al*., 2020).

Pseudomonas simiae WCS417 (syn. *Pseudomonas fluorescens* WCS417, hereafter simply WCS417) is among the PGPR studied in the greatest detail (see BOX1 for details on plant immune responses). WCS417 can promote plant growth and induce Induced Systemic Resistance (ISR) against a wide range of diseases in *A. thaliana* and other plant species (Pieterse *et al*., 2020). WCS417 actively colonizes roots, suppressing the local immune responses activated by its MAMPs (Stringlis *et al*., 2018*a*) and outcompeting other microbial strains, including related sub-group members (Bakker *et al*., 2013; Pangesti *et al*., 2017). WCS417-ISR initiation in *A. thaliana* roots depends on plant MYB72 transcription factor (TF) and its target β-GLUCOSIDASE42 (BGLU42),

192 which deglycosylates the coumarin scopolin (Palmer *et al*., 2013; Zamioudis *et al*., 2014;
193 Zamioudis *et al*., 2015; Verbon *et al*., 2017; Stringlis *et al*., 2018b; Yu *et al*., 2021). MYB72 and its
194 paralogue MYB10 regulate the Fe deficiency regulatory cascade and are functionally redundant
195 (Palmer *et al*., 2013).
196 The phenylpropanoid pathway is often up-regulated under Fe deficiency conditions; various
197 phenolic compounds show anti-microbial activity and can also strongly influence Fe uptake (Aznar
198 *et al*., 2015). Both MYB72 and MYB10 emerged as TFs required in Fe deficient roots, to adapt to
199 low Fe levels and to regulate the biosynthesis and release of coumarins (Palmer *et al*., 2013;
200 Zamioudis *et al*., 2015; Stringlis *et al*., 2018*b*; Stringlis *et al*., 2019; Yu *et al*., 2021). Coumarins
201 also emerged as important shapers of root microbiota, and they have anti-microbial potential upon
202 pathogen infection (Voges *et al*., 2019; Liu *et al*., 2021).
203 MYB72 therefore has a dual function in both plant immunity and Fe homeostasis (Stringlis *et al*.,
204 2018*b*; Stringlis *et al*., 2019). Indeed, WCS417 and Fe deficiency favour coumarins' secretion in a
205 MYB72-dependent manner (Pieterse *et al*., 2020 and references therein). Not only, MYB72 also
206 represents a node of convergence between the onset of ISR by beneficial microbes, such as
207 WCS417, and Fe deficiency response; MYB72, together with MYB10, is a direct target of
208 bHLH121 (Gao *et al*., 2020) and its expression is itself regulated by the TFs involved in the Fe
209 deficiency response, i.e., FIT (bHLH029) and bHLH038, under activation of the hormone Ethylene
210 (ET). ET could therefore represent the linking molecule between ISR and Fe deficiency response,
211 with BGLU42 and MYB72 as nodes of convergence between ISR and Fe-deficiency (Romera *et al*.,
212 2019). Notably, WCS417 can activate an Fe deficiency response in *A. thaliana* even when Fe levels
213 are sufficient, with consequent improved Fe nutrition and growth (Verbon *et al*., 2019). For
214 example, WCS417 is particularly tolerant to the antimicrobial activity of the coumarins scopolin
215 and scopoletin, and it stimulates their secretion, to possibly favour its plant niche colonization in
216 exchange for growth and immunity benefits for the plant (Verbon *et al*., 2019; Yu *et al*., 2021).
217 Iron-chelating compounds, the siderophores, are released not only by Strategy II graminaceous
218 plants for Fe uptake but also by several microorganisms, in a condition of Fe deficiency (Miethke
219 and Marahiel, 2007; Aznar *et al*., 2014; Aznar *et al*., 2015). Microbes indeed produce these low-
220 molecular-weight compounds with a high affinity for Fe(III) to form Fe-siderophore complexes that
221 are internalized by the microbial cell, thus resembling the above-mentioned plant Strategy II
222 (Herlihy *et al*., 2020 and references therein) and outcompeting other soil microbial strains, by
223 making iron unavailable (Osorio Vega, 2007; Nihorimbere *et al*., 2011). As an example, the iron-
224 chelating fluorescent pigment pyoverdine is produced by WCS417 (Pieterse *et al*., 2020). The
225 density of the soil microbial population can also influence Fe uptake. Indeed, a dense microbial
226 population causes a reduction in O_2 concentration due to its respiratory activity, accompanied by a
227 consequent rise in carbon dioxide CO_2. Such a condition favours the conversion of Fe(III) to Fe(II)
228 (Osorio Vega, 2007). Remarkably, WCS417 stimulates Fe deficiency responses in *A. thaliana* only
229 when bacteria colonizing roots are in adequate amount, both under Fe-sufficient and Fe-deficient
230 conditions (Verbon *et al*., 2019). Moreover, the WCS417-root Fe deficiency responses are regulated
231 by a shoot-to-root signalling system unrelated to leaf Fe status, suggesting the possible involvement
232 of novel phloem-mobile shoot-to-root signals and phytohormones, such as auxin (see below).
233 Despite progress in the field, mechanisms for the induction of Fe deficiency responses by PGPR
234 and their effect on plant Fe homeostasis are still unclear and require further investigations.
235 Since Fe is a limiting element in alkaline soils, it is not surprising that Fe plays a key role in the
236 interactions between plants and non-beneficial or even pathogenic microorganisms. Fe is indeed

also required by pathogens for survival: low-affinity and high-affinity strategies have been developed by phytopathogens to take up Fe from plants, and siderophores are used to sequester Fe and play a key role in microbial virulence (Franza and Expert, 2013; Aznar *et al*., 2015; Verbon *et al*., 2017; Liu *et al*., 2021). Many plant genes involved in Fe homeostasis are up-regulated during pathogen attack, and biotic stresses perturb plant Fe homeostasis (Aznar *et al*., 2015). As reported by Liu *et al*. (2021), plant Fe levels significantly contribute to plant protection against biotic stresses, as Fe withholding or Fe accumulation strategies might occur at the site of infection. The link between Fe homeostasis and plant immunity is quite complex: Fe homeostasis could contribute to the activation of ISR mediated by beneficial microbes (see Box 1). The discovery of a link between plant ISR and Fe homeostasis first occurred by the observation of a stronger induction of ISR by *Pseudomonas* spp against *Fusarium* infection in Fe deficient radish (Leeman *et al*., 1996). Hormones may represent interesting mediators to explore, particularly ET and nitric oxide (NO), which are both implicated in Fe deficiency responses and the activation of plant immune responses (Romera *et al*., 2019). ET levels may affect plant status in various ways, as its effects are influenced by the plant genotype, growth stage, plant organ and associated microbiota (Iqbal *et al*., 2017; Nascimento *et al*., 2018; Ravanbakhsh *et al*., 2018). ET levels are affected by plant microbiota activity, and strong relations between beneficial rhizobacteria, Fe deficiency responses and ISR activation have been recently exposed (Herlihy *et al*., 2020). As reviewed elsewhere in detail (Verbon *et al*., 2017; Romera *et al*., 2019), bacteria stimulating ISR could also stimulate Fe deficiency responses because of an overlap of regulatory pathways shared between the two processes, which involve several hormones such as ET, NO, auxin and the TF MYB72 (see above). Moreover, the effects of Jasmonic acid (JA) and SA on Fe nutrition as well as their roles on Fe deficiency responses have been also investigated (Kong *et al*., 2014; Shen *et al*., 2016; Boukari *et al*., 2019; Kabir *et al*., 2021). Interestingly, JA treatment increases Fe deficiency symptoms in *Arabidopsis;* indeed JA promotes FIT degradation by regulating the expression of various bHLH genes (Cui *et al*., 2018). A model describing the nodes of convergence between root Fe uptake, the rhizosphere microbiota with a highlight on WCS417, and plant immunity is presented in Figure 1.

The study of wild crops is also helpful for the elucidation of the variety of strategies activated by various plants to bypass poor soil Fe availability and the adaptation of their root apparatus together with their interaction with the rhizosphere. An example of such a field-to-lab approach was recently described in Tato *et al*., (2021): in this study, the plasticity and exudation of the roots of *Parietaria judaica* (pellitory of the wall), a wild calcicole plant growing spontaneously in an urban environment impaired in Fe availability, have been analysed; its root-associated microbiome has been also profiled. Results show that *P. judaica* roots exudate caffeoylquinic acid derivatives under calcareous conditions; they also indicate that this plant recruits beneficial soil microbes such as PGPR and phosphate solubilizers and, possibly, exclude other soil microbiota from their rhizosphere (Tato *et al*., 2021).

2) Fe transport from roots to stem and leaves and its distribution within cells The complexity of Fe homeostasis is emerging at the soil-root interface and during Fe distribution from roots to aboveground tissues. Contents of citrate, malate and succinate are elevated in the xylem, under Fe-deficient conditions (Lopez-Millan *et al*., 2010). Iron transport in the xylem to shoots predominantly occurs as Fe (III)-citrate complexes (Durrett *et al*., 2007; Rellán-Álvarez *et al*., 2008; Rellán-Álvarez *et al*., 2010). *A. thaliana* FERRIC REDUCTASE DEFECTIVE3 (FRD3) and

its rice ortholog FERRIC REDUCTASE DEFECTIVE LIKE1 (FRDL1) mediate the transport of citrate and iron to the xylem (Yokosho *et al.,* 2016). Besides FRD3, FERROPORTIN1 (FPN1) is also responsible for Fe loading into the xylem, in *A. thaliana* plants (Morrissey *et al*., 2009). Once it reaches the leaves, Fe is unloaded from the apoplastic space into the cells thanks to YSL transporters, such as AtYSL1, AtYSL2, and AtYSL3 (DiDonato *et al*., 2004; Waters *et al*., 2006). In particular, *A. thaliana* AtYSL2 is involved in the distribution of Fe from the xylem to shoot cells (DiDonato *et al*., 2004), whereas AtYSL1 and AtYSL3 are involved in the Fe-NA translocation from senescent leaves into the inflorescences and seeds and, hence, in Fe allocation throughout the phloem. Some orthologs of such YSL transporters have also been identified in rice: OsYSL2, likely involved in the Fe(II)-Nicotianamine (NA) translocation to shoots and seeds (Ishimaru *et al*., 2010), OsYSL16 which contributes to Fe(III)-deoxymugeinic acid (DMA) allocation via the vascular bundle (Kakei *et al*., 2012) and OsYSL18 which transports Fe(III)-DMA in reproductive organs and phloem of lamina joints (Aoyama *et al*., 2009). Additionally, the OLIGOPEPTIDE TRANSPORTER 3 (OPT3) is involved in the phloematic Fe transport and it mediates Fe shoot-to-root signalling (Mondoza-Cozal *et al*., 2014; Zhai *et al*., 2014; Khan *et al*., 2018).

Fe(III) is usually reduced to Fe(II), in order to cross cellular membranes (Jain *et al*., 2014). As already mentioned in the introduction, plants cells finely control the homeostasis of intracellular free Fe ions, to avoid the production of excess ROS, by transporting Fe(II) into vacuoles (Kim *et al*., 2006; Sharma *et al*., 2016) or by storing it as Fe(III) in the mineral core of the 24-mer ferritin protein cage (Briat *et al*., 2010; Lodde *et al*., 2021). *A. thaliana* possesses four ferritin isoforms (ATFER1–4) with plastidial and mitochondrial localization (Petit *et al*., 2011; Zancani *et al*., 2004; Tarantino *et al*., 2010a; Tarantino *et al*., 2010b). The developmental and environmental regulation of *AtFer1* gene expression has been analyzed in detail, including during both natural and dark-induced senescence (Tarantino *et al*., 2003; Murgia *et al*., 2007), as well as its dependence on the nitric oxide signalling network (Murgia *et al*., 2002; Arnaud *et al*., 2006), circadian rhythm (Duc *et al*., 2009), phosphate homeostasis (Bournier *et al*., 2013) and oxidative stress (Ravet *et al*., 2009a; Ravet *et al*., 2012; Reyt *et al*., 2015).

In the cytoplasm, Fe likely forms complexes with organic acids and nicotianamine (NA) forming Fe(III)-citrate, Fe(III)-NA and Fe (II)-NA (von Wiren *et al*., 1999; Rellán-Álvarez *et al*., 2010**;** Bashir *et al*., 2016; Flis *et al*., 2016) that would be available for uptake by chloroplasts, which are the major intracellular sink of intracellular Fe; intracellular Fe homeostasis has been recently extensively reviewed in Vigani *et al*. (2019).

Mitochondria also represent a relevant intracellular Fe sink (Vigani *et al*., 2015). A Fe reduction-based strategy has been suggested to occur in plant mitochondria. FRO3 and FRO8 are involved in Fe(III) reduction at the mitochondrial membrane, and MITOCHONDRIAL IRON TRANSPORTERS (MIT) mediate the Fe translocation from the cytoplasm to the mitochondrial matrix (Jain and Connolly, 2013) and the knocking down of MIT impairs plant growth and metabolism in rice (Bashir *et al*., 2011; Vigani *et al*., 2016). Accordingly, *A. thaliana* MIT1 and MIT2 are involved in mitochondrial Fe import and play an essential role in cellular and mitochondrial Fe homeostasis (Jain *et al*., 2019)

Fe deficiency-induced alteration of mitochondrial functionality impacts cellular metabolism; the characterization of the mitochondrial proteome of Fe-deficient *Cucumis sativus* (cucumber) roots indeed revealed a differential protein expression of mitochondrial enzymes involved in several metabolic pathways (Vigani *et al*., 2017). Among these enzymes, formate dehydrogenase (FDH), which catalyzes the oxidation of formate ($HCOO^-$) into carbon dioxide (CO_2), deserves a particular

mention; indeed, its abundance in cucumber roots depends on the Fe nutritional status of the plants (Vigani *et al*., 2017). Moreover, *Nicotiana tabacum* (tobacco) FDH overexpressing plants show altered Fe homeostasis as Fe content is reduced in their roots, stems and seeds (Murgia *et al*., 2020). A recent systems biology-oriented approach revealed that FDH might be considered a protein hub for plant nutrition (Di Silvestre *et al*., 2021).

Notably, the accumulation of FDH transcript has been documented under several unfavourable conditions, suggesting that FDH might be considered a stress-responsive enzyme in plants (Alekseeva *et al*., 2011). Most recently, evidence for FDH as part of the early response against the leaf infection by the pathogen *Xanthomonas campestris* pv *campestris* (*Xcc*) has been shown (Marzorati *et al*., 2021): the local accumulation of formate due to a decrease in FDH expression has been proposed as a possible signal for plant defense responses to pathogen's entry through the hydathodes (Marzorati *et al*., 2021). Taken together, these findings on FDH would strongly suggest it as a possible node of the multiple interactions between plant immune responses and Fe homeostasis, as proposed in Figure 2.

3) Seed Fe loading: when, where and how

Seed development consists of various morphogenetic steps that guarantee a correct development of the embryo and of its surrounding tissues, followed by seed maturation, in which coordinated changes of its three components, i.e. the embryo, the endosperm and the surrounding maternal tissues, occur. After completion of this maturation phase, a desiccation phase guarantees the seed's entrance into a quiescent state, so that it becomes able to survive in harsh environmental conditions (Gutierrez *et al*., 2007). The final morphological extent and physiological impact of the embryo, the endosperm and the surrounding maternal tissues on the mature seed will define the final seed architecture. Although this architecture is not fixed and it is species-specific, still all the angiosperm seeds can be broadly classified according to the prevalence of the endosperm in mature seeds (endospermic, non-endospermic and perispermic seeds) (Weber *et al*., 2005; Sreenivasulu and Wobus, 2013; Burrieza *et al*., 2014)).

Seeds maturation involves quite a dense and complex interaction of signalling networks aimed at guaranteeing the loading of seeds with all the necessary nutrients (essential elements, carbohydrates, storage proteins, oils) for germination and early stages of seedlings growth (Eggert and von Wiren, 2017). Also, a balance between concentrations of abscisic acid (ABA) favouring dormancy, and gibberellic acid (GA) favouring germination is achieved (Srivastava *et al*., 2021). ROS concentration in mature seeds is also of paramount importance and should fall within a specific range, known as the "oxidative window" and representing the ROS range enabling imbibed seeds to germinate (Gutierrez *et al*., 2007; Bailly *et al*., 2019; Lodde *et al*., 2021).

Species-specific plant architecture, including final seed architecture, will influence the mechanisms by which Fe is transferred from mother plant tissues into developing seeds; such mechanisms are quite complex as they involve the senescence of older plant parts with mobilization of Fe, its transport and distribution to the developing seed. In other words, the questions of "when" Fe is transferred from plants to seeds, "where" Fe is compartmentalized inside maturing seeds as well as inside fully matured ones and "how" Fe is mobilized from germinating seeds are of paramount relevance. Such knowledge can indeed not only allow an in-depth understanding of the physiology of Fe-loading in seeds, but it can also actively assist the various experimental approaches (breeding, genome editing, selection of relevant traits from wild crops' relatives) for the production of Fe-

dense crops, so important for human nutrition (Waters and Sankaran, 2011; Murgia *et al*., 2012; Murgia *et al*., 2013). The link between plant senescence and Fe homeostasis and its mobilization has been established by different research groups (Tarantino *et al*., 2003; Murgia *et al*., 2007; Shi *et al*., 2012; Mari *et al*., 2019; Murgia *et al*., 2020); in particular, the timing of the onset of senescence, regulated by NAC transcription factors, influences the final seed Fe content (Ricachenevsky *et al*., 2013) (see also the following section on wild crop relatives and NAM-B1). Notably, autophagy is an essential process for Fe remobilization from vegetative parts of the plants to seeds and indeed *A. thaliana* plants defective in autophagy retain more Fe in vegetative parts and show a reduced seed Fe content (Pottier *et al*., 2014; Pottier *et al*., 2019).

Iron loading into seed tissues, and in particular the roles of the NA, YSL genes and OPT3, has been thoroughly reviewed by Mari *et al*. (2020) to which readers are referred. Most recently, the role of a *A. thaliana* YABBY transcription factor INNER NO OUTER (INO) as regulator of Fe loading into developing seeds has been elucidated: INO binds indeed to the promoter of NATURAL RESISTANCE-ASSOCIATED MACROPHAGE PROTEIN1 gene (*NRAMP1*), thus inhibiting its expression; such INO inhibitory effect on NRAMP1 avoids accumulation of excess Fe into developing seeds, which would therefore protecting developing embryos from Fe toxicity caused by oxidative damage (Sun et al., 2021).

An important step of the seed loading with Fe is the reduction of Fe (III) into Fe(II), which is required for Fe transport into the embryos of dicots plants, such as pea and *A. thaliana* (Grillet *et al*., 2014a; Mari *et al*., 2020). Ascorbate (ASC) is responsible for such a reductive step (Grillet *et al*., 2014a) whereas no FRO2 homologs are apparently involved (Mari *et al*., 2020). Recently, an ASC transporter, named ATDX25, from the Multidrug And Toxic compound Extrusion (MATE) family, has been identified (Hoang *et al*., 2021). ATDX25 is expressed in flowers, seeds and seedlings and it acts as an ASC effluxer from vacuoles; its activity contributes to the Fe remobilization during germination (Hoang *et al*., 2021); to date, no evidence of ATDX25 involvement in seed Fe loading has been shown.

In a search for novel genes involved in metal uptake and transport, our research group focused on the circadian-regulated cytochrome P450 superfamily CYP82C4 gene, with expression dependent on Fe availability (Murgia *et al*., 2011). Later, CYP82C4 enzymatic activity was clarified, being responsible for the conversion of the fraxetin into sideretin (Ranjak *et al*., 2018). CYP82C4 gene expression appeared strongly correlated with genes involved in the early Fe deficiency response, but also with other genes not known to be involved in Fe homeostasis, at the time of publication. Among this second group of genes was At2g46750, which contains RY and IDE1-like motifs within its 1500 bp promoter region and it encodes a protein annotated as FAD-containing protein, at the time (Murgia *et al*., 2011). At2g46750 is currently annotated as L -GULONO-1,4-LACTONE OXIDASE 2 (GULLO2) and its expression, in roots of *A. thaliana* Fe deficient plants, is pH-dependent; its expression ratio at pH 7.0/pH 5.5 is quite low (0.04) (Tsai and Schmidt, 2020). GULLO2 attracted our attention because it oxidizes L-gulono-1,4 γ-lactone into ASC, with H_2O_2 as by-product; current knowledge is, however, that plants synthesize ASC, *in vivo*, solely via the D-mannose/L-galactose pathway and that the oxidation of L-gulono-1,4 γ-lactone by GULLO as the last ASC biosynthetic step, occurs in animal cells and not in plants (Smirnoff, 2018). To date, no physiological role has been assigned, *in vivo*, to GULLO2 (Eggers *et al*., 2021; Maruta *et al*., 2010); however, unpublished findings obtained so far in our research group, by using two independent *A. thaliana gullo2* mutants would suggest its involvement in the reduction step of

Fe(III) into Fe(II) in developing embryos (Murgia and coworkers, unpublished observations).
Much has been recently learned on Fe distributions within developing and mature seeds, thanks to the established Fe Perls staining technique amplified with DAB/H_2O_2 stain (Roschzttardtz *et al*., 2009; Brumbarova and Ivanov, 2014), but also thanks to the mapping of elemental distribution in embryos and seeds by micro X-ray fluorescence (μXRF) or by Energy Dispersive X-ray Spectroscopy (EDS), in which radiation is provided by an electron beam (Lott and West, 2001; Takahashi *et al*., 2009; Fittschen *et al*., 2017; Cardoso *et al*., 2018). The model plant *A. thaliana* has been one of the first species for which the seeds have been analysed by μXRF; seed analysis of wt and *vit1* mutants showed that Fe accumulates in the proximity of the provasculature (Kim e*t al*., 2006). Further studies demonstrated that Fe is localized in the vacuoles of the endodermal cells surrounding the provascular cambium of *A. thaliana* mature seeds (Roschzttardtz *et al*., 2009; Grillet *et al*., 2014b). Analysis of Fe distribution in maturing seeds of other Brassicaceae species, i.e., *Brassica napus*, *Nasturtium officinale*, *Lepidium sativum, Camelina sativa*, and *Brassica oleracea*, revealed that Fe is localized in the nuclei of integument, endosperm and embryo cells and that it gradually moves to surrounding structures around the nucleus to be finally loaded into vacuoles of endodermal cells surrounding the provasculature (Ibeas *et al*., 2017). However, Fe distribution in *Vasconcellea pubescens* (mountain papaya), also in the Brassicales order, showed that seed Fe is also retrieved in cortex cells (Ibeas *et al*., 2019*)*. To establish whether the pattern of Fe localization in *V. pubescens* is an exception in Brassicales or rather an indication of a wider pattern than that restricted to vacuoles of endodermal cells, more species of Eudicots belonging to orders other than Brassicales were investigated; such analyses suggested that Fe distribution has indeed a wide pattern of distribution and it can even be species- and genotype-dependent (Cvitanich et *al*., 2010; Grillet *et al*., 2014b; Ibeas *et al*., 2019). However, another study highlighted how, among the Rosids, seed Fe is detected primarily in the endodermal cell layer of the embryo (Eroglu *et al*., 2019). In monocots, Fe is predominantly in the scutellum, aleurone layer (Ozturk *et al*., 2009; Lemmens *et al*., 2018).Fe can also be stored in ferritin within amyloplasts, in seeds of some *Phaseolus* species (Cvitanich *et al*., 2010; Grillet *et al*., 2014b; Moore *et al*., 2018). These observations are important, as legume seeds have high iron content, compared to species/families such as those of *A. thaliana* itself (Murgia *et al*, 2012). During the maturation stage, seeds are loaded, besides micro and macroelements, with reserve proteins, carbohydrates and triglycerides, the proportion of which strongly depends on the species; the signalling pathways for protein and oil loading involve master regulators such as LEAFY COTYLEDON1 (LEC1), LEC2, FUSCA3 (FUS3), ABSCISIC ACID INSENSITIVE3 (ABI3); on the other side, the lack of transcription factors orchestrating micronutrient loading, including Fe, is puzzling (Roschzttardtz *et al*., 2020). In fact, Sun *et al*. (2020) proposed a role for ET in seed Fe loading through the signalling cascade involving its master transcriptional regulator EIN3 acting on the transcription factor ERF95, which in turn would bind to the GCC-boxes of the *AtFer1* promoter; however, ATFER2 is the only ferritin protein isoform detected so far in *A.thaliana* seeds, whereas ATFER1 has never been detected in *A.thaliana* seeds (Ravet *et al*., 2009a; Ravet *et al*., 2009b; Briat *et al*., 2010).Hence, the proposed signalling cascade ethylene-EIN3-ERF95-FER1 (Sun *et al*., 2020) requires further experimental investigation.
Loading of Fe into the vacuole by the VIT1 transporter, during seed maturation, as well as its mobilization from the vacuole by NRAMP3 and NRAMP4 transporters, during germination, are important steps in the post-germinative phase; *A. thaliana* mutants *vit1* or *nramp3nramp4*,

compromised in these two key Fe transport steps, indeed show severe chlorosis and growth arrest, under Fe deficiency (Lanquar *et al*., 2005; Kim *et al*., 2006; Bastow *et al*., 2018). Rice transporters OsVIT1 and OsVIT2 share, with *A. thaliana*, their role of Fe transport into the vacuole: notably, *osvit1* and *osvit2* mutants accumulate more Fe in seeds than the corresponding wt, whereas in the same mutants a reduction of Fe content is observed in flag leaves, thus confirming that VIT transporters regulate Fe trafficking between leaves and seeds (Zhang *et al*., 2012). Unfortunately, *osvit1* and *osvit2* also accumulate the toxic heavy metal cadmium Cd(II) when grown in contaminated paddy soils (Zhang et *al*., 2012), thus preventing direct use of vit1 vit2 mutations for Fe biofortification approaches.

4) Wild crops relatives in the amelioration of seed Fe loading

The domestication of wild plant species, implying anatomical, morphological and genetic changes due to cultivation and selection in an anthropic environment (Charmet, 2011; Pigna and Morandini, 2017), started independently in various regions around the world between 10000 and 2000 years b.C. Domestication of wheat, which occurred in the Fertile Crescent, caused a reduction of Fe content, in both average value and variability; wild wheat, such as *Triticum boeoticum*, *Triticum urartu*, and *Triticum dicoccoides* (wild emmer), as well as primitive wheat *T. monococcum*, show a higher Fe content in grains in comparison with modern cultivars of *Triticum durum* (durum wheat) and *Triticum estivum* (bread wheat) (Cakmak *et al*., 2000; Cakmak *et al*., 2004); in particular, wheat ancestor *Triticum dicoccoides* has a higher Fe content, compared to cultivated *Triticum aestivum* and *durum* (Cakmak *et al*., 2000; Cakmak *et al*., 2004). Indeed, 825 wild emmer accessions all originating from different regions of the Fertile Crescent regions were tested for Fe content, as well as for zinc (Zn), phosphorus (P), magnesium (Mg) and sulphur (S), and they showed an average Fe content of 46 μg g^{-1} with concentrations ranging from 15 to 109 μg g^{-1}. Interestingly, these lines also showed higher Zn content but no difference in the other tested nutrients (Cakmak *et al*., 2004; Peng *et al*., 2007).

An interesting study took advantage of the archived samples from the Broadbalk Experiment, known as the "oldest continuous agricultural experiment in the world", to investigate mineral content in the *Triticum aestivum* varieties cultivated in the last 160 years, starting from 1845 (Fan *et al*., 2008): Zn, Fe, copper (Cu) and Mg contents remained stable until 1960, when short-straw cultivars were introduced. Since that introduction, a stable decline of mineral contents was observed, accompanied by an increase in seeds yield and harvest index, which are significant factors for the observed reduction in seeds' mineral content. Authors indeed suggest that mineral nutrition of the plants, among which Fe itself, would not catch-up with the improved redistribution of photosynthates in the short-straw cultivars (Fan *et al*., 2008). In another remarkable study, the Fe content archaeological maize kernels collected in Tarakapà Region (Atacama Desert, South America) and spanning 2000 years (according to radiocarbon dating) was analysed; obtained results show a decline in Fe content associated with the shift from ancient to more recent maize varieties.

Wild crops relatives can represent a still poorly explored reservoir of genes potentially useful for the improvement of various traits, among which are the elemental content of seeds (Charmet, 2011), including Fe itself. An illuminating example of the genetic potential of wild crops relatives is represented by the single genetic locus Gpc-B1 with Mendelian segregation and associated with a higher protein, Zn and Fe content; the Recombinant Chromosome Substitution Lines (RSLs) carrying the *T. dicoccoides* Gpc-B1 allele indeed showed a 18% higher Fe concentration, when

compared with lines carrying the alternative durum allele (Distelfeld *et al*., 2007). Thanks to map-based cloning, the gene coding for a such locus has been identified and it encodes an NAC-transcription factor named NAB-B1 (Uauy *et al*., 2006). Interestingly, the wt allele accelerates senescence and favours Fe and Zn mobilization from flag leaves into developing seeds (Uauy *et al*., 2006; Lundström *et al*., 2017). Intriguingly, the reduction in NAM-B1 transcript by RNA interference causes a delayed senescence. Such a trait is, however, not associated with larger seeds, suggesting that the observed reduced Fe and Zn concentration in plants with reduced NAM-B1 activity is not simply due to a dilution effect in seeds, but on the inefficient remobilization of these nutrients from leaf to seeds, as the higher Fe and Zn concentrations in NAM-B1 RNA1 flag leaves demonstrate (Uauy *et al*., 2006). According to Uauy and coworkers, wt NAM-B1 allele is present in all wild emmer accessions tested, and in the largest part of domesticated emmer accessions (*Triticum dicoccum*) whereas both the tested *T. durum* and *T. aestivum* lines lack the functional allele, as they either carry an allele with a 1-bp insertion (causing a frame shift) or a gene deletion (Uauy *et al*., 2006). In fact, some Swedish spring wheat varieties were demonstrated to carry the wt NAM-B1 allele, without showing any relevant difference in Zn and Fe content with respect to varieties carrying the null allele (Asplund *et al*., 2013; Lundström *et al*., 2017). This poses the question of the possible effect of NAM-B1 allele and of its non-functional allele, in different genetic backgrounds (Asplund *et al*., 2013). Also, the frequency of NAM-B1 allele in domesticated emmer wheat poses the question of whether the NAM-B1 gene can be considered a genuine domestication gene or, instead, a diversification gene (Lundström *et al*., 2017). Nonetheless, a years-long analysis of the NAM-B1 wt allele and its distribution among wheats, confirms that wild crops relatives represent a large and still unexplored reservoir of potentially valuable genes that can be exploited for neo-domestication approaches (Charmet, 2011; Peng *et al*., 2013). The annual wild species *Cicer judaicum* is, for example, the most promising wild crop relative for improvement of Fe content in chickpeas (Sharma *et al*., 2020).

The analysis of the wild progenitor of *O. sativa*, i.e. *O. rufipogon*, allowed to ascertain that the Combined Strategy (CS) of Fe uptake preceded rice domestication (Wairich *et al*., 2019). Again, these findings support the possible use of wild rice in the improvement of CS strategy, as far as seed Fe loading is concerned.

Conclusions

The potential of beneficial rhizobacteria to activate both plant ISR defense responses and Fe uptake responses opens the possibility of using such microbial strains as biopesticides and Fe biofertilizers. Nonetheless, further studies are required to investigate the link between ISR and Fe uptake response, especially in crops under field conditions, and the various middle/long-term ecological implications that the use of such beneficial microbial strains would imply. Such studies will certainly be important for the detailed understanding of plant Fe nutrition in the field, where plants are continuously exposed to various biotic and abiotic stresses; they can also potentially impact the costs associated with reduced crop yields in alkaline soils. Thus, research on plant beneficial microorganisms and Fe nutrition appears as a very engaging field for future researchers.

The road of Fe from soil to seeds is long indeed, and it still features several question marks, one of which will be to define the precise biochemical network of transcription factors, enzymes, transporters and biochemical steps involved in moving Fe into the various seed tissues during their

development and maturation. In this respect, the exploitation of genetic resources derived from Fe-dense seeds of wild crops' relatives, appears an attractive avenue to be explored in the short term.

Acknowledgements

The figures presented in this work were designed by using BioRender (https://biorender.com).

This work is dedicated to the memory of Delia Tarantino, who passed away two years ago at a too young age; her kindness, friendship, enthusiasm and dedication to science, are vivid in our memories and inspired our work.

Authors contribution statement

IM: Conceptualization, writing – Original Draft Preparation.

FM, GV, PM: Contributions to original draft preparation

FM: Figures preparation, with contributions of IM, GV, PM.

References

Alekseeva AA, Savin SS, Tishkov VI. 2011. NAD+-dependent formate dehydrogenase from plants, Acta Naturae 3, 38–54. https://doi.org/10.32607/20758251- 2011-3-4-38-54.

Aoyama T, Kobayashi T, Takahashi M, Nagasaka S, Usuda K, Kakei Y, Ishimaru Y, Nakanishi H, Mori S, Nishizawa NK. 2009. OsYSL18 is a rice iron(III)-deoxymugineic acid transporter specifically expressed in reproductive organs and phloem of lamina joints. Plant Molecular Biology 70, 681–692.

Arnaud N, Murgia I, Boucherez J, Briat JF, Cellier F, Gaymard F. 2006. An iron-induced nitric oxide burst precedes ubiquitin-dependent protein degradation for Arabidopsis *AtFer1* ferritin gene expression. Journal of Biological Chemistry 281, 23579-23588.

Asplund L, Bergkvist G, Leino MW, Westerbergh A, Weih M. 2013. Swedish spring wheat varieties with the rare high grain protein allele of NAM-B1 differ in leaf senescence and grain mineral content. PLoS One 8(3):e59704.

Aung MS, Masuda H. 2020. How does rice defend against excess iron? Physiological and molecular mechanisms. Frontiers in Plant Science 11:1102. doi: 10.3389/fpls.2020.0110.

Aznar A, Chen NWG, Rigault M, *et al.* 2014. Scavenging iron: a novel mechanism of plant immunity activation by microbial siderophores. Plant Physiology 164, 2167–2183.

Aznar A, Chen NWG, Thomine S, Dellagi A. 2015. Immunity to plant pathogens and iron homeostasis. Plant Science 240, 90–97.

Bailly C. 2019. The signalling role of ROS in the regulation of seed germination and dormancy. Biochemical Journal 476, 3019–3032. doi: https://doi.org/10.1042/BCJ20190159.

Bakker PAHM, Berendsen RL, Doornbos RF, Wintermans PCA, Pieterse CMJ. 2013. The rhizosphere revisited: root microbiomics. Frontiers in Plant Science 4, 1-7.

Bakker PAHM, Berendsen RL, Van Pelt JA, *et al.* 2020. The Soil-borne identity and microbiome-assisted agriculture: looking back to the future. Molecular Plant 13, 1394-1401.

Barberon M, Zelazny E, Robert S, Conéjéro G, Curie C, Friml J, Vert, G. 2011.Monoubiquitin-dependent endocytosis of the IRON-REGULATED TRANSPORTER 1 (IRT1) transporter controls iron uptake in plants. Proceedings of the National Academy of Sciences USA, 108, E450-E458.

Barberon M, Dubeaux G, Kolb C, Isono E, Zelazny E, Vert G. 2014. Polarization of IRON-REGULATED TRANSPORTER 1 (IRT1) to the plant-soil interface plays crucial role in metal homeostasis. Proceedings of the National Academy of Sciences USA 111, 8293–8298.

Bashir K, Ishimaru Y, Shimo, H, Nagasaka S, Fujimoto M, Takanashi H, Tsutsumi N, An G, Nakanishi H, Nishizawa NK. 2011. The rice mitochondrial iron transporter is essential for plant growth. Nature Communications 2, 322.

Bashir K, Rasheed S, Kobayashi T, Seki M, Nishizawa NK. 2016. Regulating subcellular metal homeostasis: the key to crop improvement. Frontiers in Plant Science 7: 1192.

Bastow EL, Garcia de la Torre VS, Maclean AE, Green RT, Merlot S, Thomine S, Balk J. 2018. Vacuolar iron stores gated by NRAMP3 and NRAMP4 are the primary source of iron in germinating seeds. Plant Physiology 177,1267-1276. doi: 10.1104/pp.18.00478.

Bauer P, Ling HQ, Guerinot ML. 2007. FIT, the FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR in Arabidopsis. Plant Physiology and Biochemistry 45, 260–261.

Berendsen RL, Pieterse CMJ, Bakker PAHM. 2012. The rhizosphere microbiome and plant health. Trends in Plant Science 17, 478–486.

Boukari N, Jelali N, Renaud JB, Youssef R Ben, Abdelly C, Hannoufa A. 2019. Salicylic acid seed priming improves tolerance to salinity, iron deficiency and their combined effect in two ecotypes of Alfalfa. Environmental and Experimental Botany 167. doi.org/10.1016/j.envexpbot.2019.103820.

Bournier M, Tissot N, Mari S, Boucherez J, Lacombe E, Briat JF, Gaymard, F. 2013. Arabidopsis ferritin 1 (*AtFer1*) gene regulation by the phosphate starvation response 1 (*AtPHR1*) transcription factor reveals a direct molecular link between iron and phosphate homeostasis. The Journal of Biological Chemistry 288, 22670-22680. https://doi.org/10.1074/jbc.M113.482281

Briat JF, Duc C, Ravet K, Gaymard F. 2010. Ferritins and iron storage in plants. Biochimica and Biophysica Acta 1800,806-814.

Briat JF, Dubos C, Gaymard F. 2015. Iron nutrition, biomass production, and plant product quality. Trends in Plant Science 20, 33-40.

Brumbarova T, Ivanov R. 2014. Perls staining for histochemical detection of iron in plant samples. Bio-protocol 4(18): e1245. DOI: 10.21769/BioProtoc.1245.

Burrieza HP, López-Fernández MP, Maldonado S. 2014. Analogous reserve distribution and tissue characteristics in quinoa and grass seeds suggest convergent evolution. Frontiers in Plant Science 5: 546.

Cakmak I, Ozkan H, BraunHJ, Welch RM, Romheld V. 2000. Zinc and iron concentrations in seeds of wild, primitive, and modern wheats. Food and Nutrition Bulletin, vol. 21, no. 4, 401-403.

Cakmak I, Torun A, Millet E, Feldman M, Fahima T, Korol A, Nevo E, Braun HJ, Özkan H. 2004. *Triticum dicoccoides*: an important genetic resource for increasing zinc and iron concentration in modern cultivated wheat. Soil Science and Plant Nutrition 50,1047-1054. DOI: 10.1080/00380768.2004.10408573

Cardoso P, Mateus TC, Velu G, Singh RP, Santos JP, Carvalho ML, Lourenço VM, Lidon F, Reboredo F, Guerra M. 2018. Localization and distribution of Zn and Fe in grains of biofortified bread wheat lines through micro- and triaxial-X-ray fluorescence spectrometry. Spectrochimica Acta 141, 70-79.

Charmet G. 2011. Wheat domestication: lessons for the future. Comptes Rendus Biologies 334, 212-220. doi: 10.1016/j.crvi.2010.12.013

Choi HW, Klessig DF. 2016. DAMPs, MAMPs, and NAMPs in plant innate immunity. BMC Plant Biology 16, 232. https://doi.org/10.1186/s12870-016-0921-2.

Cointry V, Vert G. 2019. The bifunctional transporter-receptor IRT1 at the heart of metal sensing and signalling. New Phytologist 223, 1173-1178. https://doi.org/10.1111/nph.15826.

Colangelo EP, Guerinot ML. 2004. The essential basic Helix-Loop-Helix protein FIT1 is required for the iron deficiency response. Plant Cell 16, 3400-3412.

Colombo C, Palumbo G, He, JZ, Pinton R, Cesco S. 2014. Review on iron availability in soil: interaction of Fe minerals, plants, and microbes. Journal of Soils and Sediments 14**,** 538-548. https://doi.org/10.1007.

Compant S, Samad A, Faist H, Sessitsch A. 2019. A review on the plant microbiome: ecology, functions, and emerging trends in microbial application. Journal of Advanced Research 19, 29-37.

Connorton JM, Balk J, Rodríguez-Celma J. 2017. Iron homeostasis in plants - a brief overview. Metallomics 9, 813-823. doi:10.1039/c7mt00136c.

Conrath U, Beckers GJM, Langenbach CJG, Jaskiewicz MR. 2015. Priming for enhanced defense. Annual Review of Phytopathology 53, 97-119.

Cui H, Tsuda K, Parker JE. 2015. Effector-triggered immunity: from pathogen perception to robust defense. Annual Review of Plant Biology 66, 487–511.

Cui Y, Chen CL, Cui M, Zhou WJ, Wu HL, Ling HQ. 2018. Four IVa bHLH transcription factors are novel interactors of FIT and mediate JA inhibition of iron uptake in *Arabidopsis*. Molecular Plant 11, 1166-1183. https://doi.org/10.1016/j.molp.2018.06.005.

Cvitanich C, Przybyłowicz W, Urbanski D, Jurkiewicz A, Mesjasz-Przybyłowicz J, Blair M, Astudillo C, Jensen EØ, Stougaard J. 2010. Iron and ferritin accumulate in separate cellular locations in *Phaseolus* seeds. BMC Plant Biology 10:26. doi: 10.1186/1471-2229-10-26.

De Lorenzo G, Ferrari S, Cervone F, Okun E. 2018. Extracellular DAMPs in plants and mammals: immunity, tissue damage and repair. Trends in Immunology 39, 937-950. https://doi.org/10.1016/j.it.2018.09.006.

Delory BM, Delaplace P, Fauconnier ML, du Jardin P. 2016. Root-emitted volatile organic compounds: can they mediate belowground plant-plant interactions? Plant and Soil 402, 1-26.

DiDonato RJ, Roberts LA, Sanderson T, Eisley RB, Walker EL. 2004. Arabidopsis Yellow Stripe-Like2 (YSL2): a metal-regulated gene encoding a plasma membrane transporter of nicotianamine–metal complexes. The Plant Journal 39, 403–414.

Di Silvestre D, Vigani G, Mauri P, Hammadi S, Morandini P, Murgia I. 2021. Network topological analysis for the identification of novel hubs in plant nutrition. Frontiers in Plant Science 10;12:629013. doi: 10.3389/fpls.2021.629013.

Distelfeld A, Cakmak I, Peleg Z, Ozturk L, Yazici AM, Budak, H, Saranga Y, Fahima T. 2007. Multiple QTL-effects of wheat Gpc-B1 locus on grain protein and micronutrient concentrations. Physiologia Plantarum 129, 635-643.

do Amaral FP, Tuleski TR, Pankievicz VCS, *et al.* 2020. Diverse bacterial genes modulate plant root association by beneficial bacteria. ASM Journals mBio 11, 1-15.

Dubeaux G, Neveu J, Zelazny E, Vert G. 2018. Metal sensing by the IRT1 transporter-receptor orchestrates its own degradation and plant metal nutrition. Molecular Cell 69, 953- 964. e955.

Duc C, Cellier F, Lobréaux S, Briat JF, Gaymard F. 2009. Regulation of iron homeostasis in *Arabidopsis thaliana* by the Clock Regulator Time for Coffee. Journal of Biological Chemistry 284, 36271-36281. https://doi.org/10.1074/jbc.M109.059873.

Durrett TP, Gassmann W, Rogers EE. 2007. The FRD3-mediated efflux of citrate into the root vasculature is necessary for efficient iron translocation. Plant Physiology 144, 197-205.

Eggers R, Jammer A, Jha S, Kerschbaumer B, Lahham M, Strandback E, Toplak M, Wallner S, Winkler A, Macheroux P. 2021.The scope of flavin-dependent reactions and processes in the model plant *Arabidopsis thaliana*. Phytochemistry 189, 112822. https://doi.org/10.1016/j.phytochem.2021.112822.

Eggert K, von Wiren N. 2017. Dynamics and partitioning of the ionome in seeds and germinating seedlings of winter oilseed rape. Metallomics 5, 1316-1325.

Elgueta AV, Navarro N, Uribe M, Robe K, Gaymard F, Dubos C, Pérez MF, Roschzttardtz H. 2021. 2000 years of agriculture in the Atacama desert lead to changes in the distribution and concentration of iron in maize. Science Reports 11**,** 17322. https://doi.org/10.1038/s41598-021-96819-1

Eroglu S, Karaca N, Vogel-Mikus K, Kavčič A, Filiz E and Tanyolac B. 2019. The Conservation of VIT1-Dependent Iron Distribution in Seeds. Frontiers in Plant Science 10:907. doi: 10.3389/fpls.2019.00907

Fan MS, Zhao FJ, Fairweather-Tait SJ, Poulton PR, Dunham SJ, McGrath SP. 2008.Evidence of decreasing mineral density in wheat grain over the last 160 years. Journal of Trace Elements in Medicine and Biology 22, 315-24. doi: 10.1016/j.jtemb.2008.07.002.

Fittschen UEA, Kunz HH, Höhner R, Tyssebotn IMB, Fittschen A. 2017. A new micro X-ray fluorescence spectrometer for in vivo elemental analysis in plants. X-Ray Spectrometry 46, 374–381.

Flis P, Ouerdane L, Grillet L, Curie C, Mari S, Lobinski R. 2016. Inventory of metal complexes circulating in plant fluids: a reliable method based on HPLC coupled with dual elemental and high-resolution molecular mass spectrometric detection. New Phytologist 211, 1129-1141.

Fourcroy P, Sisó-Terraza P, Sudre D, Savirón M, Reyt G, Gaymard F, Abadía A, Abadia J, Álvarez-Fernández A, Briat JF. 2014. Involvement of the ABCG37 transporter in secretion of scopoletin and derivatives by *Arabidopsis* roots in response to iron deficiency. New Phytologist 201,155-167. doi: 10.1111/nph.12471.

Fourcroy P, Tissot N, Gaymard F, Briat JF, Dubos C. 2016. Facilitated Fe nutrition by phenolic compounds excreted by the Arabidopsis ABCG37/PDR9 transporter requires the IRT1/FRO2 high-affinity root Fe2+ transport system. Molecular Plant 9, 485-488. https://doi.org/10.1016/j.molp.2015.09.010.

Franza T, Expert D. 2013. Role of iron homeostasis in the virulence of phytopathogenic bacteria: An 'à la carte' menu. Molecular Plant Pathology 14, 429–438.

Fu ZQ, Dong X. 2013. Systemic acquired resistance: turning local infection into global defense. Annual Review of Plant Biology 64, 839–863.

Gao F, Robe K, Bettembourg M, Navarro N, Rofidal V, Santoni V, Gaymard F, Vignols F, Roschzttardtz H, Izquierdo E, Dubos C. 2020. The transcription factor BHLH121 interacts with BHLH105 (IRL3) and its closest homologs to regulate iron homeostasis in *Arabidopsis*. Plant Cell 32, 508-524. doi: 10.1105/tpc.19.00541.

Gao F, Dubos C. 2021.Transcriptional integration of plant responses to iron availability. Journal of Experimental Botany 72, 2056-2070. doi: 10.1093/jxb/eraa556.

Garbeva P, Weisskopf L. 2020. Airborn medicine: bacterial volatiles and their influence on plant health. New Phytologist 226, 32-43. doi: 10.1111/nph.16282.

Grillet L, Ouerdane L, Flis P, Hoang M, Isaure M, Lobinski R, Curie C, Mari S. 2014a. Ascorbate efflux as a new strategy for iron reduction and transport in plants. Journal of Biological Chemistry 289, 2515–2525. doi: 10.1074/jbc.M113.514828.

Grillet L, Mari S, Schmidt W. 2014b. Iron in seeds – loading pathways and subcellular localization. Frontiers in Plant Science 4:535. doi: 10.3389/fpls.2013.00535.

Grillet L, Schmidt W. 2019. Iron acquisition strategies in land plants: not so different after all. New Phytologist 224, 11–18

Gulati S, Ballhausen MB, Kulkarni P, Grosch R, Garbeva P. 2020. A non-invasive soil-based setup to study tomato root volatiles released by healthy and infected roots. Scientific Reports 10, 12704. https://doi.org/10.1038/s41598-020-69468-z.

Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C. 2007. Combined networks regulating seed maturation. Trends in Plant Science 12, 294-300.

Herlihy JH, Long TA, McDowell JM. 2020. Iron homeostasis and plant immune responses: recent insights and translational implications. Journal of Biological Chemistry 295, 13444–13457.

Hoang MTT, Almeida D, Chay S, Alcon C, Corratge-Faillie C, Curie C, Mari S. 2021. AtDTX25, a member of the multidrug and toxic compound extrusion family, is a vacuolar ascorbate transporter that controls intracellular iron cycling in Arabidopsis. New Phytologist 231,1956-1967. doi: 10.1111/nph.17526.

Ibeas MA, Grant-Grant S, Navarro N, Perez MF, Roschzttardtz H. 2017. Dynamic subcellular localization of iron during embryo development in *Brassicaceae* seeds. Frontiers in Plant Science 8:2186. doi: 10.3389/fpls.2017.02186.

Ibeas MA, Grant-Grant S, Coronas MF. *et al.* 2019. The diverse iron distribution in Eudicotyledoneae seeds: from Arabidopsis to Quinoa. Frontiers in Plant Science 9:1985. doi: 10.3389/fpls.2018.01985.

Ipek M, Esitken A. 2017. The actions of PGPR on micronutrient availability in soil under calcareous soil conditions: an evaluation over Fe nutrition. In: Singh D., Singh H., Prabha R. (eds) Plant-Microbe Interactions in Agro-Ecological Perspectives. Springer, Singapore. https://doi.org/10.1007/978-981-10-6593-4_4.

Iqbal N, Khan NA, Ferrante A, Trivellini A, Francini A, Khan MIR. 2017. Ethylene role in plant growth, development and senescence: interaction with other phytohormones. Frontiers in Plant Science 8, 1–19.

Ishimaru Y, Suzuki M, Tsukamoto T. *et al.* 2006. Rice plants take up iron as an Fe^{3+}-phytosiderophore and as Fe^{2+}. The Plant Journal 45, 335–346

Ivanov R, Brumbarova T, Blum A, Jantke AM, Fink-Straube C. and Bauer, P. 2014. SORTING NEXIN1 is required for modulating the trafficking and stability of the Arabidopsis IRON-REGULATED TRANSPORTER1. Plant Cell 26, 1294–1307. doi: 10.1105/tpc.113.116244

Ishimaru Y, Masuda H, Bashir K. *et al.* 2010. Rice metal-nicotianamine transporter, OsYSL2, is required for the long-distance transport of iron and manganese. The Plant Journal 62, 379–390.

Jain A, Connolly EL. 2013. Mitochondrial iron transport and homeostasis in plants. Frontiers in Plant Science 4, 348. https://doi.org/10.3389/fpls.2013.00348

Jain A, Wilson GT, Connolly EL. 2014. The diverse roles of FRO family metalloreductases in iron and copper homeostasis. Frontiers in Plant Science 5, 100. doi: 10.3389/fpls.2014.00100.

Jain A, Dashner ZS, Connolly EL. 2019. Mitochondrial Iron Transporters (MIT1 and MIT2) are essential for iron homeostasis and embryogenesis in *Arabidopsis thaliana*. Frontiers in Plant Science 10:1449. doi: 10.3389/fpls.2019.01449.

Jogaiah S, Abdelrahman M. 2019. Bioactive Molecules in Plant Defense. Signaling in Growth and Stress. Springer Eds, 1–248.

Jakoby M, Wang HY, Reidt W, Weisshaar B, Bauer P. 2004. FRU (BHLH029) is required for induction of iron mobilization genes in *Arabidopsis thaliana*. FEBS Letters 577, 528–534.

Kabir AH, Tahura S, Elseehy MM, El-Shehawi AM. 2021. Molecular characterization of Fe-acquisition genes causing decreased Fe uptake and photosynthetic inefficiency in Fe-deficient sunflower. Scientific Reports 11, 1–13.

Kakei Y, Ishimaru Y, Kobayashi T, Yamakawa T, Nakanishi H, Nishizawa NK. 2012. OsYSL16 plays a role in the allocation of iron. Plant Molecular Biology 79, 583–594.

Khan MA, Castro-Guerrero NA, McInturf SA, Nguyen NT, Dame AN, Wang J, Bindbeutel RK, Joshi T, Jurisson SS, Nusinow DA, Mendoza-Cozat DG. 2018. Changes in iron availability in Arabidopsis are rapidly sensed in the leaf vasculature and impaired sensing leads to opposite transcriptional programs in leaves and roots Plant Cell Environment 41, 2263–2276.

Kim SA, Punshon T, Lanzirotti A, Li L, Alonso JM, Ecker JR, Kaplan J, Guerinot ML. 2006. Localization of iron in *Arabidopsis* seed requires the vacuolar membrane transporter VIT1. Science 314, 1295–1298.

Kim SA, LaCroix IS, Gerber SA, Guerinot ML. 2019. The iron deficiency response in *Arabidopsis thaliana* requires the phosphorylated transcription factor URI. Proceedings of the National Academy of Sciences 116, 24933-24942; DOI: 10.1073/pnas.1916892116.

Klessig DF, Choi HW, Dempsey DA. 2018. Systemic Acquired Resistance and Salicylic Acid: past, present, and future. Molecular Plant Microbe Interactions 31, 871-888. doi: 10.1094/MPMI-03-18-0067-CR.

Kobayashi T, Nishizawa NK. 2012. Iron uptake, translocation, and regulation in higher plants. Annual Review of Plant Biology 63, 131–152.

Kobayashi T, Nozoye T, Nishizawa, NK. 2019. Iron transport and its regulation in plants. Free Radical Biology and Medicine. 133, 11-20.

Kong J, Dong Y, Xu L, Liu S, Bai X. 2014. Effects of foliar application of salicylic acid and nitric oxide in alleviating iron deficiency induced chlorosis of *Arachis hypogaea* L. Botanical Studies 55, 9. doi: 10.1186/1999-3110-55-9.

Lanquar V, Lelièvre F, Bolte S. *et al.* 2005. Mobilization of vacuolar iron by AtNRAMP3 and AtNRAMP4 is essential for seed germination on low iron. EMBO Journal 24, 4041–4051. doi: 10.1038/sj.emboj.7600864.

Leeman M, Den Ouden FM, Van Pelt JA, Dirkx FPM, Steijl H, Bakker PAHM, Schippers B. 1996. Iron availability affects induction of systemic resistance to Fusarium wilt of radish by *Pseudomonas fluorescens*. Phytopathology 86, 149–155.

Lemmens E, De Brier N, Spiers KM, Ryan C, Garrevoet J, Falkenberg G, Goos P, Smolders E, Delcour JA. 2018. The impact of steeping, germination and hydrothermal processing of wheat (*Triticum aestivum* L.) grains on phytate hydrolysis and the distribution, speciation and bio-accessibility of iron and zinc elements. Food Chemistry 264, 367-376.

Lindsay WL, Schwab AP. 1982. The chemistry of iron in soils and its availability to plants. Journal of Plant Nutrition 5,821–840.

Lingam S, Mohrbacher J, Brumbarova T, Potuschak T, Fink-Straube C, Blondet E, Genschik P, Bauer P. 2011. Interaction between the bHLH transcription factor FIT and ETHYLENE INSENSITIVE3/ETHYLENE INSENSITIVE3-LIKE1 reveals molecular linkage between the regulation of iron acquisition and ethylene signaling in Arabidopsis. Plant Cell 23, 1815–1829.

Liu Y, Kong D, Wu H-L, Ling H-Q. 2021. Iron in plant–pathogen interactions. Journal of Experimental Botany 72, 2114–2124.

Lodde V, Morandini P, Costa A, Murgia I, Ezquer, I. 2021. cROStalk for Life: uncovering ROS signaling in plants and animal systems, from gametogenesis to early embryonic development. Genes 12, 525. https://doi.org/10.3390/genes12040525

López-Millán AF, Morales F, Abadía A, Abadía J. 2000. Changes induced by Fe deficiency and Fe resupply in the organic acid metabolism of sugar beet (*Beta vulgaris*) leaves. Physiologia Plantarum 112, 31-38. doi: 10.1034/j.1399-3054.2001.1120105.x.

Lott JNA, West MM. 2001. Elements present in mineral nutrient reserves in dry *Arabidopsis thaliana* seeds of wild type and pho1, pho2, and man1 mutants. Canadian Journal of Botany 79, 1292–1296.

Lundström M, Leino MW, Hagenblad J. 2017. Evolutionary history of the NAM-B1 gene in wild and domesticated tetraploid wheat. BMC Genetics 18:118

Lynch S V, Pedersen O. 2016. The Human Intestinal Microbiome in Health and Disease. New England Journal of Medicine 375, 2369–2379.

Majeed A, Muhammad Z, Ahmad H. 2018. Plant growth promoting bacteria: role in soil improvement, abiotic and biotic stress management of crops. Plant Cell Reports 37, 1599–1609.

Mari S, Bailly C, Thomine S. 2020. Handing off iron to the next generation: how does it get into seeds and what for? Biochemical Journal 477, 259-274.

Martinez-Medina A, Flors V, Heil M, Mauch-Mani B, Pieterse CMJ, Pozo MJ, Ton J, van Dam NM, Conrath U. 2016. Recognizing plant defense priming. Trends in Plant Science 21, 818–822.

Maruta T, Ichikawa Y, Mieda T, Takeda T, Tamoi M, Yabuta Y, Ishikawa T, Shigeoka S. 2010. The contribution of Arabidopsis homologs of L-Gulono-1,4-lactone Oxidase to the biosynthesis of ascorbic acid. Bioscience, Biotechnology, and Biochemistry 74, 1494-1497. DOI:10.1271/bbb.100157.

Marzorati F, Vigani G, Morandini P, Murgia I. 2021. Formate dehydrogenase contributes to the early *Arabidopsis thaliana* responses against *Xanthomonas campestris pv campestris* infection. Physiological and Molecular Plant Pathology 114, 101633. https://doi.org/10.1016/j.pmpp.2021.101633.

Mendoza-Cózatl DG, Gokul A, Carelse MF, Jobe TO, Long TA. and Keyster M. 2019. Keep talking: crosstalk between iron and sulfur networks fine-tunes growth and development to promote survival under iron limitation. Journal of Experimental Botany 70, 4197–4210. doi: 10.1093/jxb/erz290

Miethke M, Marahiel MA. 2007. Siderophore-based iron acquisition and pathogen control. Microbiology and Molecular Biology Reviews 71, 413–451.

Moore KL, Rodríguez-Ramiro I, Jones ER, *et al*. 2018. The stage of seed development influences iron bioavailability in pea (*Pisum sativum* L.). Scientific Reports 8(1):6865. doi:10.1038/s41598-018-25130-3.

Morrissey J, Baxter IR, Lee J, Li L, Lahner B, Grotz N, Kaplan J, Salt DE, Guerinot ML. 2009. The ferroportin metal efflux proteins function in iron and cobalt homeostasis in *Arabidopsis*. Plant Cell 21, 3326–3338.

Murgia I, Delledonne M, Soave C. 2002. Nitric oxide mediates iron-induced ferritin accumulation in *Arabidopsis*. The Plant Journal 30, 521-528.

Murgia I, Vazzola V, Tarantino D, Cellier F, Ravet K, Briat JF, Soave C. 2007. Knock-out of ferritin AtFer1 causes earlier onset of age-dependent leaf senescence in *Arabidopsis*. Plant Physiology and Biochemistry 45, 898-907. doi: 10.1016/j.plaphy.2007.09.007.

Murgia I, Tarantino D, Soave C, Morandini P. 2011. The *Arabidopsis* CYP82C4 expression is dependent on Fe availability and the circadian rhythm and it correlates with genes involved in the early Fe-deficiency response. Journal of Plant Physiology 168, 894-902.

Murgia I, Arosio P, Tarantino D, Soave C. 2012. Biofortification for combating "hidden hunger" for iron. Trends in Plant Science. 17, 47-55. DOI: 10.1016/j.tplants.2011.10.003.

Murgia I, De Gara L, Grusak M. 2013. Biofortification: how can we exploit plant science to reduce micronutrient deficiencies? Frontiers in Plant Science 4:429. doi: 10.3389/fpls.2013.00429.

Murgia I, Vigani G, Di Silvestre D, Mauri P, Rossi R, Bergamaschi A, Frisella M, Morandini P. 2020. Formate dehydrogenase takes part in molybdenum and iron homeostasis and affects dark-induced senescence in plants. Journal of Plant Interactions 15, 386-397. Doi: 10.1080/17429145.2020.1836273.

Nascimento FX, Rossi MJ, Glick BR. 2018. Ethylene and 1-aminocyclopropane-1-carboxylate (ACC) in plant–bacterial interactions. Frontiers in Plant Science 9, 1–17.

Nihorimbere V, Ongena M, Smargiassi M, Thonart P. 2011. Beneficial effect of the rhizosphere microbial community for plant growth and health. Biotechnology, Agronomy and Society and Environment 15, 327–337.

Osorio Vega NW. 2007. A review on beneficial effects of rhizosphere bacteria on soil nutrient availability and plant nutrient uptake. Revista Facultad Nacional de Agronomía Medellín 60, 3621–3643.

Ozturk L, Altintas G, Erdem H, Gokmen O O, Yazici, A, Cakmak, I. 2009. Localization of iron, zinc, and protein in seeds of spelt (*Triticum aestivum* ssp. *spelta*) genotypes with low and high protein concentration. UC Davis: https://escholarship.org/uc/item/08n1b60m.

Paasch BC, He SY. 2021. Toward understanding microbiota homeostasis in the plant kingdom. PLoS Pathogens 17, 1–8.

Palmer CM, Hindt MN, Schmidt H, Clemens S, Guerinot M L. 2013. MYB10 and MYB72 are required for growth under iron-limiting conditions. PLoS Genetics 9(11): e1003953. https://doi.org/10.1371/journal.pgen.1003953.

Pangesti N, Vandenbrande S, Pineda A, Dicke M, Raaijmakers JM, Van Loon JJA. 2017. Antagonism between two root-associated beneficial *Pseudomonas* strains does not affect plant growth promotion and induced resistance against a leaf-chewing herbivore. FEMS Microbiology Ecology 93, 1–8.

Pascale A, Proietti S, Pantelides IS, Stringlis IA. 2020. Modulation of the root microbiome by plant molecules: the basis for targeted disease suppression and plant growth promotion. Frontiers in Plant Science 10, 1741. DOI:10.3389/fpls.2019.01741.

Peng JH, Sun DF, Peng YL, Nevo E. 2013. Gene discovery in *Triticum dicoccoides*, the direct progenitor of cultivated wheats. Cereal Research Communications 41, 1-22.

Petit JM, Briat JF, Lobréaux S. 2001. Structure and differential expression of the four members of the *Arabidopsis thaliana* ferritin gene family. Biochemical Journal 359, 575-582. doi:10.1042/0264-6021:3590575

Pieterse CMJ, Leon-Reyes A, Van Der Ent S, Van Wees SCM. 2009. Networking by small-molecule hormones in plant immunity. Nature Chemical Biology 5, 308–316.

Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, Van Wees SCM, Bakker PAHM. 2014. Induced systemic resistance by beneficial microbes. Annual Review of Phytopathology 52, 347–375.

Pieterse CMJ, Berendsen RL, Jonge R De, Stringlis IA. 2020. *Pseudomonas simiae* WCS417□: star track of a model beneficial rhizobacterium. Plant Soil 461, 245–263.

Pigna G, Morandini P. 2017. Domestication of New Species, in Pilu R., Gavazzi G. (eds.) More Food: Road to Survival, Bentham Science, Sharjah.

Pontiggia D, Benedetti M, Costantini S, De Lorenzo G, Cervone F. 2020. Dampening the DAMPs: how plants maintain the homeostasis of cell wall molecular patterns and avoid hyper-immunity. Frontiers in Plant Science 11:613259. doi: 10.3389/fpls.2020.613259.

Pottier M, Masclaux Daubresse C, Yoshimoto K, Thomine S. 2014. Autophagy as a possible mechanism for micronutrient remobilization from leaves to seeds. Frontiers in Plant Science 5, 11.DOI:10.3389/fpls.2014.00011.

Pottier M, Dumont J, Masclaux-Daubresse C, Thomine S. 2019. Autophagy is essential for optimal translocation of iron to seeds in *Arabidopsis*. Journal of Experimental Botany 70, 859–869. doi.org/10.1093/jxb/ery388

Pozo MJ, Azcón-Aguilar C. 2007. Unraveling mycorrhiza-induced resistance. Current Opinion in Plant Biology 10, 393–398.

Rajniak J, Giehl RF, Chang E, Murgia I, von Wirén N, Sattely ES. 2018. Biosynthesis of redox-active metabolites in response to iron deficiency in plants. Nature Chemical Biology 14, 442-450. doi:10.1038/s41589-018-0019-2.

Ramirez L, Simontacchi M, Murgia I, Zabaleta E, Lamattina L. 2011. Nitric Oxide, Nitrosyl Iron complexes, ferritin and frataxin: a well equiped team to preserve plant iron homeostasis. Plant Science 181, 582-592.

Ravanbakhsh M, Sasidharan R, Voesenek LACJ, Kowalchuk GA, Jousset A. 2018. Microbial modulation of plant ethylene signaling: ecological and evolutionary consequences. Microbiome 6, 52.

Ravet K, Touraine B, Boucherez J, Briat JF, Gaymard F, Cellier F. 2009a. Ferritins control interaction between iron homeostasis and oxidative stress in *Arabidopsis*. Plant Journal 57, 400-412. doi: 10.1111/j.1365-313X.2008.03698.x.

Ravet K, Touraine B, Kim SA, Cellier F, Thomine S, Guerinot ML, Briat JF, Gaymard F. 2009b. Post-Translational Regulation of AtFER2 Ferritin in Response to Intracellular Iron Trafficking during Fruit Development in Arabidopsis. Molecular Plant 2, 1095-1106. https://doi.org/10.1093/mp/ssp041.

Ravet K, Reyt G, Arnaud N, Krouk G, El-Batoul D, Boucherez J, Briat JF, Gaymard F. 2012. Iron and ROS control of the DownSTream mRNA decay pathway is essential for plant fitness. The EMBO Journal 31, 175–186.

Rellán-Álvarez R, Abadía J, Álvarez-Fernández A. 2008. Formation of metal-nicotianamine complexes as affected by pH, ligand exchange with citrate and metal exchange. A study by electrospray ionization time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 22, 1553–1562.

Rellán-Álvarez R, Giner-Martínez-Sierra J, Orduna J, Orera I, Rodríguez-Castrillón JA, García-Alonso JI, Abadía J, Álvarez-Fernández A. 2010. Identification of a tri-iron(III), tri-citrate complex in the xylem sap of iron-deficient tomato resupplied with iron: new insights into plant iron long-distance transport. Plant and Cell Physiology 51, 91–102, https://doi.org/10.1093/pcp/pcp170

Reyt G, Boudouf S, Boucherez J, Gaymard F, Briat JF. 2015. Iron-and ferritin-dependent reactive oxygen species distribution: impact on *Arabidopsis* root system architecture. Molecular Plant 8, 439-453.

Riaz N, Guerinot ML. 2021. All together now: regulation of the iron deficiency response. Journal of Experimental Botany 72,2045-2055. doi: 10.1093/jxb/erab003

Ricachenevsky FK, Menguer PK, Sperotto RA. 2013. kNACking on heaven's door: how important are NAC transcription factors for leaf senescence and Fe/Zn remobilization to seeds? Frontiers in Plant Science 4:226. doi: 10.3389/fpls.2013.00226.

Robe, K, Conejero G, Gao F, Lefebvre-Legendre L, Sylvestre-Gonon E, Rofidal V, Hem S, Rouhier N, Barberon M, Hecker A, Gaymard F, Izquierdo E, Dubos C. 2021a. Coumarin accumulation and trafficking in *Arabidopsis thaliana*: a complex and dynamic process. New Phytologist 229, 2062–2079 doi: 10.1111/nph.17090.

Robe K, Izquierdo E, Vignols F, Rouached H, Dubos C. 2021b. The Coumarins: secondary metabolites playing a primary role in plant nutrition and health. Trends in Plant Science 26, 248-259. https://doi.org/10.1016/j.tplants.2020.10.008

Romera FJ, García MJ, Lucena C, Martínez-Medina A, Aparicio MA, Ramos J, Alcántara E, Angulo M, Pérez-Vicente R. 2019. Induced systemic resistance (ISR) and Fe deficiency responses in dicot plants. Frontiers in Plant Science 10, 1–17.

Roschzttardtz H, Conéjéro G, Curie C, Mari, S. 2009. Identification of the endodermal vacuole as the iron storage compartment in the *Arabidopsis* embryo. Plant Physiology 151, 1-10. doi: 10.1104/pp.109.144444.

Roschzttardtz H, Gaymard F, Dubos C. 2020. Transcriptional regulation of iron distribution in seeds: a perspective. Frontiers in Plant Science 11:725. doi: 10.3389/fpls.2020.00725.

Sánchez-Cañizares C, Jorrín B, Poole PS, Tkacz A. 2017. Understanding the holobiont: the interdependence of plants and their microbiome. Current Opinion in Microbiology 38, 188–196.

Sasse J, Martinoia E, Northen T. 2018. Feed your friends: do plant exudates shape the root microbiome? Trends in Plant Science 23, 25–41.

Schmid NB, Giehl RFH, Doll S, Mock HP, Strehmel N, Scheel D, Kong XL, Hider RC, von Wiren N. 2014. Feruloyl-CoA 6'hydroxylase1-dependent coumarins mediate iron acquisition from alkaline substrates in *Arabidopsis*. Plant Physiology 164, 160–172.

Schulz-Bohm K, Martín-Sánchez L, Garbeva P. 2017. Microbial volatiles: small molecules with an important role in intra- and inter-kingdom interactions. Frontiers in Microbiology 8, 2484 DOI:10.3389/fmicb.2017.02484.

Schwacke R, Schneider A, Van Der Graaff E, Fischer K, Catoni E, Desimone M, Frommer WB, Flügge UI, Kunze R. 2003. ARAMEMNON, a novel database for *Arabidopsis* integral membrane proteins. Plant Physiology 131, 16-26.

Schwarz B, Bauer P. 2020. FIT, a regulatory hub for iron deficiency and stress signaling in roots, and FIT-dependent and -independent gene signatures, Journal of Experimental Botany 71, 1694–1705, https://doi.org/10.1093/jxb/eraa012.

Schwertmann U. 1991. Solubility and dissolution of iron oxides. Plant and Soil 130, 1–25. www.jstor.org/stable/42937281.

Sharma SS, Dietz KJ, Mimura T. 2016. Vacuolar compartmentalization as indispensable component of heavy metal detoxification in plants. Plant, Cell and Environment 39, 1112–1126. doi: 10.1111/pce.12706.

Sharma S, Lavale SA, Nimje C, Singh S. 2020. Characterization and identification of annual wild Cicer species for seed protein and mineral concentrations for chickpea improvement. Crop Science 61, 315-319.

Shen C, Yang Y, Liu K, Zhang L, Guo H, Sun T, Wang H. 2016. Involvement of endogenous salicylic acid in iron-deficiency responses in *Arabidopsis*. Journal of Experimental Botany 67, 4179–4193.

Shi R, Weber G, Köster J, Reza-Hajirezaei M, Zou C, Zhang F, Wirén NV. 2012. Senescence-induced iron mobilization in source leaves of barley (*Hordeum vulgare*) plants. New Phytologist 195, 372-383.

Simon JC, Marchesi JR, Mougel C, Selosse MA. 2019. Host-microbiota interactions: from holobiont theory to analysis. Microbiome 7, 1–5.

Sisó-Terraza P, Rios JJ, Abadía J, Abadía A, Álvarez-Fernández A. 2016. Flavins secreted by roots of iron-deficient *Beta vulgaris* enable mining of ferric oxide via reductive mechanisms. New Phytologist 209, 733–745. doi: 10.1111/nph.13633.

Smirnoff N. 2018. Ascorbic acid metabolism and functions: a comparison of plants and mammals. Free Radical Biology & Medicine 122,116-129. doi: 10.1016/j.freeradbiomed.2018.03.033.

Sreenivasulu N, Wobus U. 2013. Seed-development programs: a systems biology-based comparison between dicots and monocots. Annual Review Plant Biology 64, 189–217.

Srivastava AK, Kumar JK, Suprasanna P. 2021. Seed 'primeomics': plants memorize their germination under stress. Biological Reviews. doi: 10.1111/brv.12722.

Stassen MJJ, Hsu SH, Pieterse CMJ, Stringlis IA. 2021. Coumarin communication along the

microbiome–root–shoot axis. Trends in Plant Science 26, 169–183.

Stringlis IA, Proietti S, Hickman R, Van Verk MC, Zamioudis C, Pieterse CMJ. 2018*a*. Root transcriptional dynamics induced by beneficial rhizobacteria and microbial immune elicitors reveal signatures of adaptation to mutualists. Plant Journal 93, 166–180.

Stringlis IA, Yu K, Feussner K, De Jonge R, Van Bentum S, Van Verk MC, Berendsen RL, Bakker PAHM, Feussner I, Pieterse CMJ. 2018*b*. MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. Proceedings of the National Academy of Sciences of the United States of America 115, E5213–E5222.

Stringlis IA, De Jonge R, Pieterse CMJ. 2019. The age of coumarins in plant-microbe interactions. Plant and Cell Physiology 60, 1405–1419.

Sun Y, Li JQ, Yan JY, Yuan JJ, Li GX, Wu YR, Xu JM, Huang RF, Harberd NP, Ding ZJ, Zheng SJ. 2020. Ethylene promotes seed iron storage during *Arabidopsis* seed maturation via ERF95 transcription factor. Journal of Integrative Plant Biology 62,1193-1212. doi: 10.1111/jipb.12986.

Sun L, Wei YQ, Wu KH, Yan JY, Xu JN, Wu YR, Li GX, Xu JM, Harberd NP, Ding ZJ, Zheng SJ. 2021. Restriction of iron loading into developing seeds by a YABBY transcription factor safeguards successful reproduction in Arabidopsis. Molecular Plant 14, 1624-1639. doi.org/10.1016/j.molp.2021.06.005.

Takahashi M, Nozoye T, Kitajima N. *et al*. 2009. *In vivo* analysis of metal distribution and expression of metal transporters in rice seed during germination process by microarray and X-ray fluorescence imaging of Fe, Zn, Mn, and Cu. Plant Soil 325**,** 39. https://doi.org/10.1007/s11104-009-0045-7

Tarantino D, Petit JM, Lobreaux S, Briat JF, Soave C, Murgia I. 2003. Differential involvement of the IDRS cis-element in the developmental and environmental regulation of the AtFer1 ferritin gene from Arabidopsis. Planta 217, 709-716.

Tarantino D, Casagrande F, Soave C, Murgia I. 2010a. Knocking out of the mitochondrial *AtFer4* ferritin does not alter response of *Arabidopsis* plants to abiotic stresses. Journal of Plant Physiology 167, 453-460.

Tarantino D, Santo N, Morandini P, Casagrande F, Braun HP, Heinemeyer J, Vigani G, Soave C, Murgia I. 2010b. AtFer4 ferritin is a determinant of iron homeostasis in *Arabidopsis thaliana* heterotrophic cells. Journal of Plant Physiology 167, 1598-1605.

Tato L, Lattanzio V, Ercole E, Dell'Orto M, Sorgonà A, Linsalata V, Salvioli di Fossalunga A, Novero M, Astolfi S, Abenavoli MR, Murgia I, Zocchi G, Vigani G. 2021. Plasticity, exudation and microbiome-association of the root system of Pellitory-of-the-wall plants grown in environments impaired in iron availability. Plant Physiology and Biochemistry 168, 27-42.

https://doi.org/10.1016/j.plaphy.2021.09.040.

Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. 2020. Plant–microbiome interactions: from community assembly to plant health. Nature Reviews Microbiology 18, 607–621.

Tsai HH, Schmidt W. 2017. Mobilization of iron by plant-borne coumarins. Trends in Plant Science 22, 538–548.

Tsai HH, Rodrıguez-Celma J, Lan P, Wu YC, Velez-Bermudez IC, Schmidt W. 2018. Scopoletin 8-hydroxylase-mediated fraxetin production is crucial for iron mobilization. Plant Physiology 177, 194–207.

Tsai HH, Schmidt W. 2020. pH-dependent transcriptional profile changes in iron-deficient Arabidopsis roots. BMC Genomics 21:694. https://doi.org/10.1186/s12864-020-07116-6.

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J. 2006. A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314, 1298-1301.

Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, Dufresne A. 2015. The importance of the microbiome of the plant holobiont. New Phytologist 206, 1196–1206.

Van der Ent S, Van Wees SCM, Pieterse CMJ. 2009. Jasmonate signaling in plant interactions with resistance-inducing beneficial microbes. Phytochemistry 70, 1581–1588.

Van Loon LC, Bakker PAHM, Pieterse CMJ. 1998. Systemic resistance induced by rhizosphere bacteria. Annual Review of Phytopathology 36, 453–483.

Van Loon LC. 2007. Plant responses to plant growth-promoting rhizobacteria. European Journal of Plant Pathology 119, 243–254.

Verbon EH, Trapet PL, Stringlis IA, Kruijs S, Bakker PAHM, Pieterse CMJ. 2017. Iron and Immunity. Annual Review of Phytopathology 55, 355–375.

Verbon EH, Trapet PL, Kruijs S, Temple-Boyer-Dury C, Rouwenhorst TG, Pieterse CMJ. 2019. Rhizobacteria-mediated activation of the Fe deficiency response in *Arabidopsis* roots: impact on Fe status and signaling. Frontiers in Plant Science 10, 1–12.

Vigani G, Morandini P, Murgia I. 2013. Searching iron sensors in plants by exploring the link among 2′-OG-dependent dioxygenases, the iron deficiency response and metabolic adjustments occurring under iron deficiency. Frontiers in Plant Science 4, 169.

Vigani G, Faoro F, Ferretti AM, Cantele F, Maffi, D, Marelli, M, Maver M, Murgia I, Zocchi G. 2015. Three-dimensional reconstruction, by TEM tomography, of the ultrastructural modifications occurring in *Cucumis sativus L.* mitochondria under Fe deficiency. PLoS One 10: e0129141.

Vigani G, Bashir K, Ishimaru Y, Lehmann M, Casiraghi, FM, Nakanishi H, Seki M, Geigenberger P, Zocchi G, Nishizawa NK. 2016. Knocking down mitochondrial iron transporter (MIT) reprograms primary and secondary metabolism in rice plants. Journal of Experimental Botany 67, 1357–1368.

Vigani G, Di Silvestre D, Agresta AM, Donnini S, Mauri P, Gehl C, Bittner F, Murgia I. 2017. Molybdenum and iron mutually impact their homeostasis in cucumber (*Cucumis sativus*) plants. New Phytologist 213, 1222– 124.

Vigani G, Hanikenne M. 2018. "Metal homeostasis in plant mitochondria," in Logan DC, ed. Annual Plant Reviews. Chichester: John Wiley & Sons, Ltd, 111–142. doi: 10.1002/9781119312994.apr0547

Vigani G, Murgia I. 2018. Iron-requiring enzymes in the spotlight of oxygen. Trends in Plant Science. 23, 874-882.

Vigani G, Solti A, Thomine S, Philippar K. 2019. Essential and detrimental-an update on intracellular iron trafficking and homeostasis. Plant and Cell Physiology 1–20. doi:10.1093/pcp/pcz091

Voges MJEEE, Bai Y, Schulze-Lefert P, Sattely ES. 2019. Plant-derived coumarins shape the composition of an *Arabidopsis* synthetic root microbiome. Proceedings of the National Academy of Sciences of the United States of America 116, 12558–12565.

von Wiren N, Klair S, Bansal S, Briat JF, Khodr H, Shioiri T, Leigh RA, Hider RC. 1999. Nicotianamine chelates both FeIII and FeII. Implications for metal transport in plants. Plant Physiology 119, 1107-1114. doi: 10.1104/pp.119.3.1107.

Wairich A, de Oliveira BHN, Arend EB, Duarte GL, Ponte LR, Sperotto RA, Ricachenevsky FK, Fett JP. 2019. The combined strategy for iron uptake is not exclusive to domesticated rice (*Oryza sativa*). Scientific Reports 9:16144.

Wang HY, Klatte M, Jakoby M, Bäumlein H, Weisshaar B, Bauer P. 2007. Iron deficiency-mediated stress regulation of four subgroup Ib BHLH genes in *Arabidopsis thaliana*. Planta 226, 897–908.

Wang N, Cui Y, Liu Y, Fan H, Du J, Huang Z, Yuan Y, Wu H, Ling HQ. 2013. Requirement and functional redundancy of Ib subgroup bHLH proteins for iron deficiency responses and uptake in *Arabidopsis thaliana*. Molecular Plant 6, 503-513. doi: 10.1093/mp/sss089.

Waters BM, Chu HH, DiDonato RJ, Roberts LA, Eisley RB, Lahner B, Salt DE, Walker EL. 2006. Mutations in *Arabidopsis* Yellow Stripe-Like1 and Yellow Stripe-Like3 reveal their roles in metal ion homeostasis and loading of metal ions in seeds. Plant Physiology 141, 1446–1458.

Waters BM, Sankaran RP. 2011. Moving micronutrients from the soil to the seeds. Genes and physiological processes from a biofortification perspective. Plant Science 180, 562-574.

Weber H, Borisjuk L, Wobus U. 2005. Molecular physiology of legume seed development. Annual Review Plant Biology 56, 253–279.

Yokosho K, Yamaji N, Ma JF. 2016. OsFRDL1 expressed in nodes is required for distribution of iron to grains in rice. Journal of Experimental Botany 67, 5485–5494.

Yu K, Stringlis IA, van Bentum S, de Jonge R, Snoek BL, Pieterse CMJ, Bakker PAHM, Berendsen RL. 2021. Transcriptome signatures in *Pseudomonas simiae* WCS417 shed light on role of root-secreted coumarins in Arabidopsis-mutualist communication. Microorganisms 9, 1–15.

Zamioudis C, Hanson J, Pieterse CMJ. 2014. ß-Glucosidase BGLU42 is a MYB72-dependent key regulator of rhizobacteria-induced systemic resistance and modulates iron deficiency responses in *Arabidopsis* roots. New Phytologist 204, 368–379.

Zamioudis C, Korteland J, Van Pelt JA, *et al.* 2015. Rhizobacterial volatiles and photosynthesis-related signals coordinate MYB72 expression in *Arabidopsis* roots during onset of induced systemic resistance and iron-deficiency responses. Plant Journal 84, 309–322.

Zancani M, Peresson M, Biroccio A, Federici G, Urbani A, Murgia I, Soave C, Micali F, Vianello A, Macrì F. 2004. Evidence for the presence of ferritin in plant mitochondria. European Journal of Biochemistry 271, 3657-3664.

Zhai Z, Gayomba SR, Jung HI, *et al.* 2014. OPT3 is a phloem-specific iron transporter that is essential for systemic iron signaling and redistribution of iron and cadmium in *Arabidopsis*. Plant Cell 26, 2249–2264.

Zhang Y, Xu YH, Yi HY, Gong JM. 2012. Vacuolar membrane transporters OsVIT1 and OsVIT2 modulate iron translocation between flag leaves and seeds in rice. The Plant Journal 72,400-410. doi: 10.1111/j.1365-313X.2012.05088.x.

Ziegler J, Schmidt S, Strehmel N, Scheel D, Abel S. 2017. *Arabidopsis* transporter ABCG37/PDR9 contributes primarily highly oxygenated coumarins to root exudation. Scientific Reports 7, 3704. https://doi.org/10.1038/s41598-017-03250-6.

Zipfel C. 2008. Pattern-recognition receptors in plant innate immunity. Current Opinion in Immunology 20, 10–16.

BOX 1: Plant immune responses

Plants are continuously exposed to the attacks of several pathogens and pests during their life; they have therefore developed diverse strategies to perceive assaulters and mount immune responses. Once plants come in contact with microbes, they firstly locally recognize characteristic features of microorganisms (e.g., lipopolysaccharides, glycoproteins, flagellin and chitin), known as microbe-associated molecular patterns (MAMPs), or pathogen-associated molecular patterns (PAMPs) when produced by pathogens. These molecules are perceived through pattern-recognition receptors (PRRs), starting downstream signalling pathways that activate the so-called 'MAMP-triggered immunity' (MTI) or 'PAMP-triggered immunity' (PTI) (Zipfel, 2008; Pieterse *et al*., 2009; Choi *et al*., 2016; De Lorenzo *et al*., 2018; Pontiggia *et al*., 2020). A secondary major immune response, known as 'effector-triggered immunity' (ETI), is also activated when specific plant resistance proteins are produced to react against pathogens' molecules (i.e., the effectors), which are introduced in plant host cells to suppress MTI/PTI (Cui *et al*., 2015). These two immune defense responses partially overlap, because of the accumulation of pathogenesis-related (PR) proteins that help in plant resistance. Once the defense response has been turned on at the site of the infection, a 'Systemic Acquired Resistance' (SAR), is frequently activated far away from the site of the attack to defend undamaged tissues. SAR activation is associated with an increase of the hormone salicylic acid (SA), both at the site of the infection and in distant plant organs (Fu and Dong, 2013; Klessig *et al*., 2018). Beneficial microorganisms in soils can stimulate a systemic immunity similar to SAR, known as 'Induced Systemic Resistance' (ISR) (Van Loon *et al*., 1998; Pozo and Azcón-Aguilar, 2007; Pieterse *et al*., 2014). In ISR, root microbiota can elevate the level of disease resistance against different pathogenic threats, activating various phytohormone signalling pathways and transferring this defense message to distant plant tissues. Jasmonic acid (JA) and ethylene (ET) are the two hormones involved in ISR signalling (Van der Ent *et al*., 2009). Similar to SAR, ISR is activated only upon an external stress factor, so that plants can save resources; this strategy is known as 'defense priming', allowing plants to alert their immune system for future pathogen or pest attacks, thus avoiding a direct activation of defense responses. The main differences between the two systemic plant immunities are the 'priming stimulus' triggering the defense priming and the hormones involved (Conrath *et al*., 2015; Martinez-Medina *et al*., 2016). ISR has been described for several PGPR such as *Pseudomonas* spp., *Bacillus* spp. and *Serratia* spp.; *Pseudomonas* and *Bacillus* genera often represent the dominant group in the rhizosphere. ISR has been also described for PGPF, such as *Trichoderma* spp., *Fusarium* spp., *Serendipita* spp., and arbuscular mycorrhizal fungi AMF (Nihorimbere *et al*., 2011; Pascale *et al*., 2020). Nevertheless, perturbations of this 'core' root microbiota may be helpful for plants, since variations in microbial genera abundance in the rhizosphere may help plants to react against different biotic and abiotic stresses (Paasch and He, 2021 and references therein).

Figure legends

Figure 1
Schematic model of the interactions among root Fe uptake mechanisms, the microbiota present in the rhizosphere (in particular WCS417) and plant immunity responses. A Strategy I root cell and its surrounding rhizosphere are represented. Red arrows show the cascade of events occurring during Fe deficiency response (thick arrows for transport/movement; thin arrows for signalling), which are regulated by FIT1, MYB72/MYB10, BGLU72, and leading to i) Fe uptake through the coordinated activity of AHA2, FRO2 and IRT1, ii) biosynthesis of coumarins through the phenylpropanoid pathway and their release into the rhizosphere through PDR9 transporter. Both coumarins and protons act on the pool of poorly soluble Fe(hydroxy)oxides. *Pseudomonas simiae* WCS417, as well as other beneficial microorganisms, can release bacterial siderophores which also increase Fe solubility from the pool of Fe(hydroxy)oxides in the soil. Blue arrows indicate immune response pathways, such that one triggered by WCS417 and leading to suppression of MAMP-triggered immunity (evasion of host immunity), as well as those triggered by WCS417 siderophores and leading to suppression of plant pathogens. Purple arrows indicate overlapping pathways of Fe deficiency responses and immune responses; WCS417 induction of Fe deficiency responses and phenylpropanoid pathway and dependence of such induction on WCS417 concentration threshold are represented, as well as the effects of coumarins on WCS417 and pathogens. PM: plasma membrane. For further details regarding biosynthesis, transport and biological activity of coumarins refer to Robe *et al*. (2021b).

Figure 2
Formate dehydrogenase (FDH) is a protein hub for Fe plant nutrition and a node of the multiple interactions between Fe homeostasis and plant responses to abiotic and biotic stresses. Experiments conducted on roots and on aerial parts of *A. thaliana* plants (shown in the center) support the model of FDH as a hub of plant Fe nutrition, in a loop regulation with Fe homeostasis and responses against abiotic stresses. A leaf hydathode under physiological conditions (upper right), or exposed to *Xanthomonas campestris campestris Xcc* attack (lower right) are represented. Inhibition of FDH promoter activity by *Xcc* would lead to a local increase in formate concentration; such change in formate concentration, in turn, might act as a possible signal for plant defense responses to pathogen's entry (see main text and cited references for details).

Figure 3
Proposed model of Fe uptake in developing embryos. An *A.thaliana* developing seed is shown, with its embryo at the bent cotyledon stage, endosperm and the cell layers of maternal origin forming the seed coat. The possible contribution of the L-gulono-1,4 γ-lactone oxidase GULLO2 to the ASC pool, for Fe(III) reduction into Fe(II) and its subsequent transport into the developing embryos, is reported with dashed arrows. The role of the GULLO2 reaction product H_2O_2 on the endosperm and on the seed coat composition is unknown. ASC, ascorbic acid.

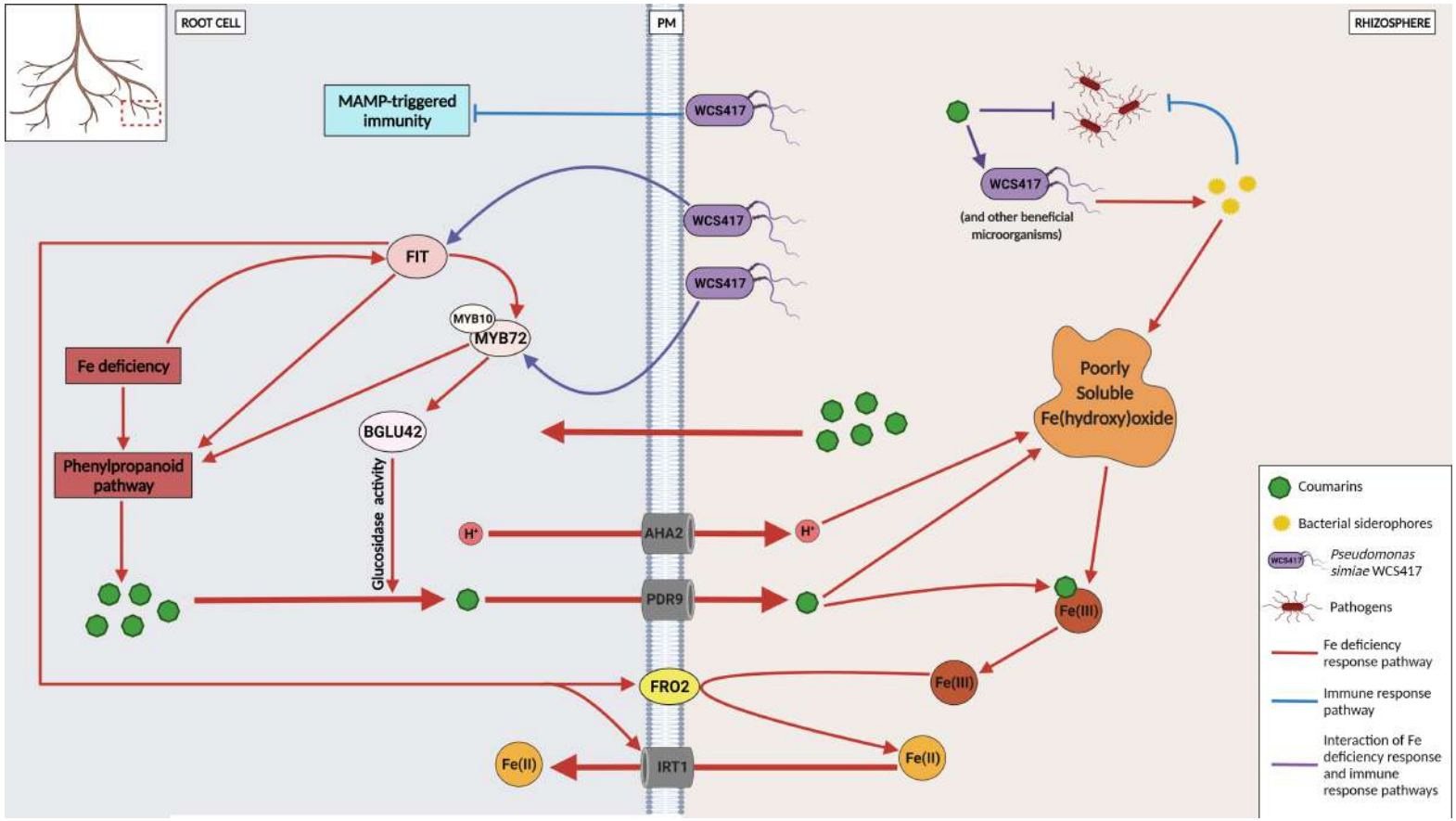


Figure 1

Schematic model of the interactions among root Fe uptake mechanisms, the microbiota present in the rhizosphere (in particular WCS417) and plant immunity responses. A Strategy I root cell and its surrounding rhizosphere are represented. Red arrows show the cascade of events occurring during Fe deficiency response (thick arrows for transport/movement; thin arrows for signalling), which are regulated by FIT1, MYB72/MYB10, BGLU72, and leading to i) Fe uptake through the coordinated activity of AHA2, FRO2 and IRT1, ii) biosynthesis of coumarins through the phenylpropanoid pathway and their release into the rhizosphere through PDR9 transporter. Both coumarins and protons act on the pool of poorly soluble Fe(hydroxy)oxides. *Pseudomonas simiae* WCS417, as well as other beneficial microorganisms, can release bacterial siderophores which also increase Fe solubility from the pool of Fe(hydroxy)oxides in the soil. Blue arrows indicate immune response pathways, such that one triggered by WCS417 and leading to suppression of MAMP-triggered immunity (evasion of host immunity), as well as those triggered by WCS417 siderophores and leading to suppression of plant pathogens. Purple arrows indicate overlapping pathways of Fe deficiency responses and immune responses; WCS417 induction of Fe deficiency responses and phenylpropanoid pathway and dependence of such induction on WCS417 concentration threshold are represented, as well as the effects of coumarins on WCS417 and pathogens. PM: plasma membrane. For further details regarding biosynthesis, transport and biological activity of coumarins refer to Robe *et al.* (2021b).

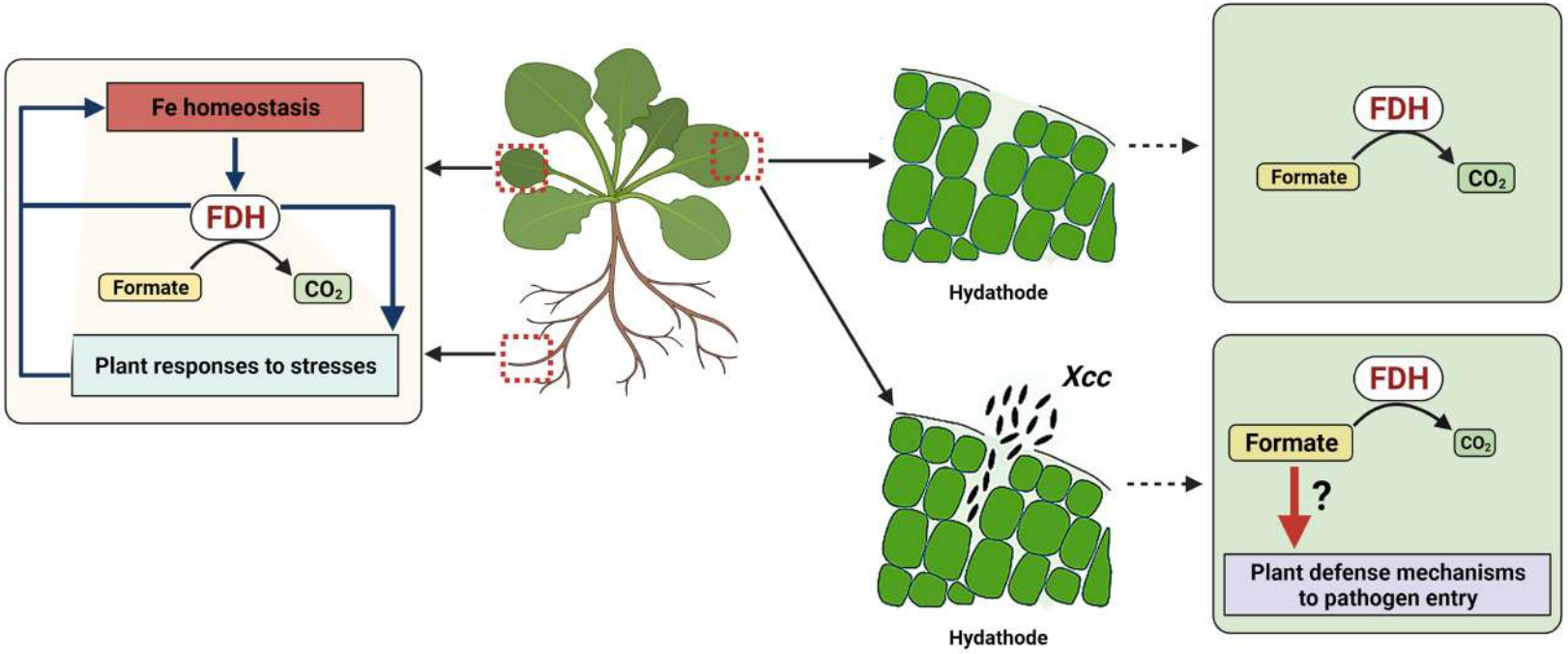


Figure 2

Formate dehydrogenase (FDH) is a protein hub for Fe plant nutrition and a node of the multiple interactions between Fe homeostasis and plant responses to abiotic and biotic stresses. Experiments conducted on roots and on aerial parts of *A. thaliana* plants (shown in the center) support the model of FDH as a hub of plant Fe nutrition, in a loop regulation with Fe homeostasis and responses against abiotic stresses. A leaf hydathode under physiological conditions (upper right), or exposed to *Xanthomonas campestris campestris Xcc* attack (lower right) are represented. Inhibition of FDH promoter activity by *Xcc* would lead to a local increase in formate concentration; such change in formate concentration, in turn, might act as a possible signal for plant defense responses to pathogen's entry (see main text and cited references for details).

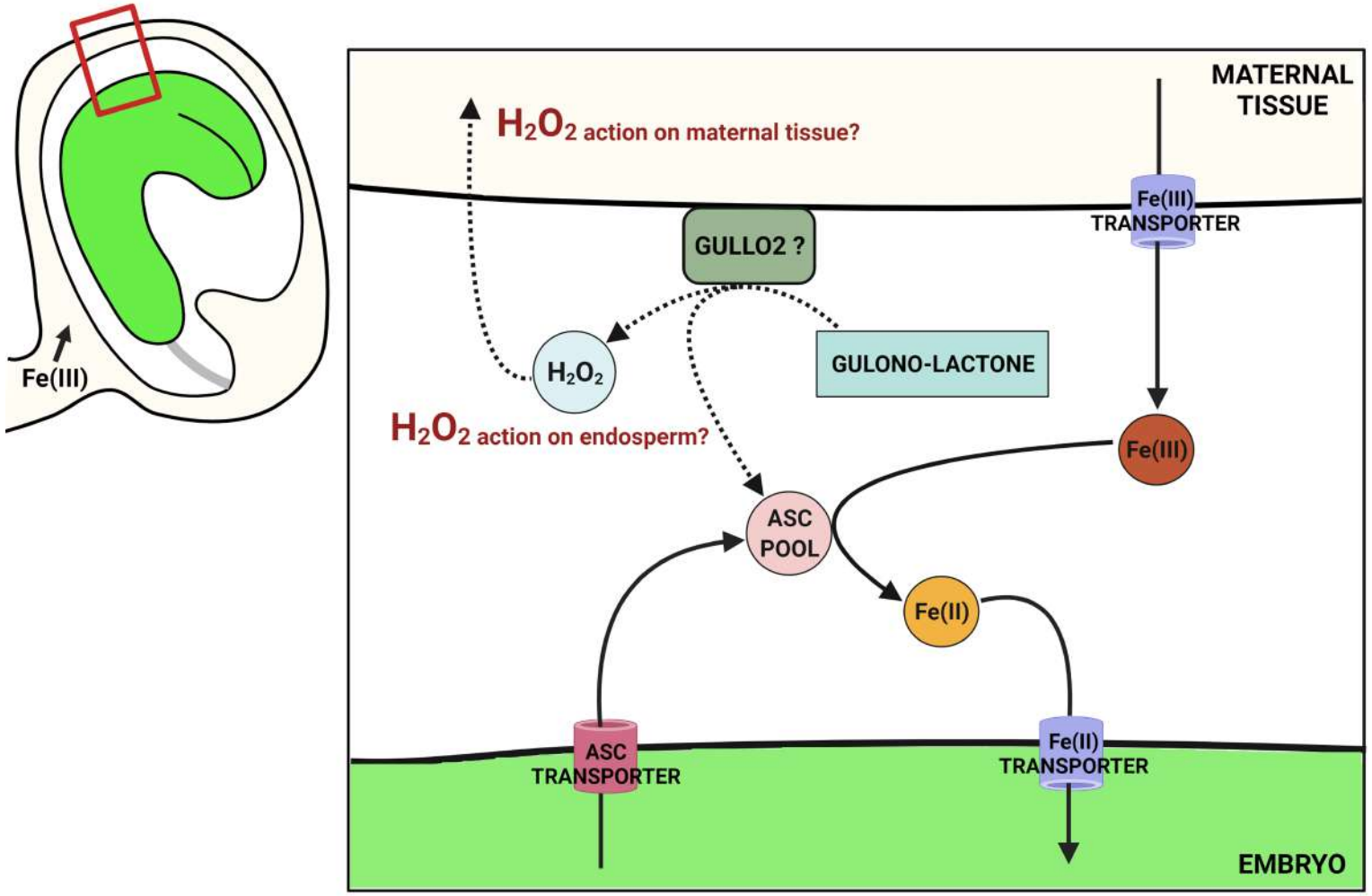


Figure 3

Proposed model of Fe uptake in developing embryos. An *A.thaliana* developing seed is shown, with its embryo at the bent cotyledon stage, endosperm and the cell layers of maternal origin forming the seed coat. The possible contribution of the L-gulono-1,4 γ-lactone oxidase GULLO2 to the ASC pool, for Fe(III) reduction into Fe(II) and its subsequent transport into the developing embryos, is reported with dashed arrows. The role of the GULLO2 reaction product H_2O_2 on the endosperm and on the seed coat composition is unknown. ASC, ascorbic acid.

Manuscript 6

Good or bad? The double face of iron in plants

Francesca Marzorati[1*], Alessia Midali[1†], Piero Morandini[1], Irene Murgia[1,2]

[1] Dept. Environmental Science and Policy, University of Milano, Milano, Italy
[2] Dept. Biosciences, University of Milano, Milano, Italy
[*] Corresponding author: francesca.marzorati1@unimi.it

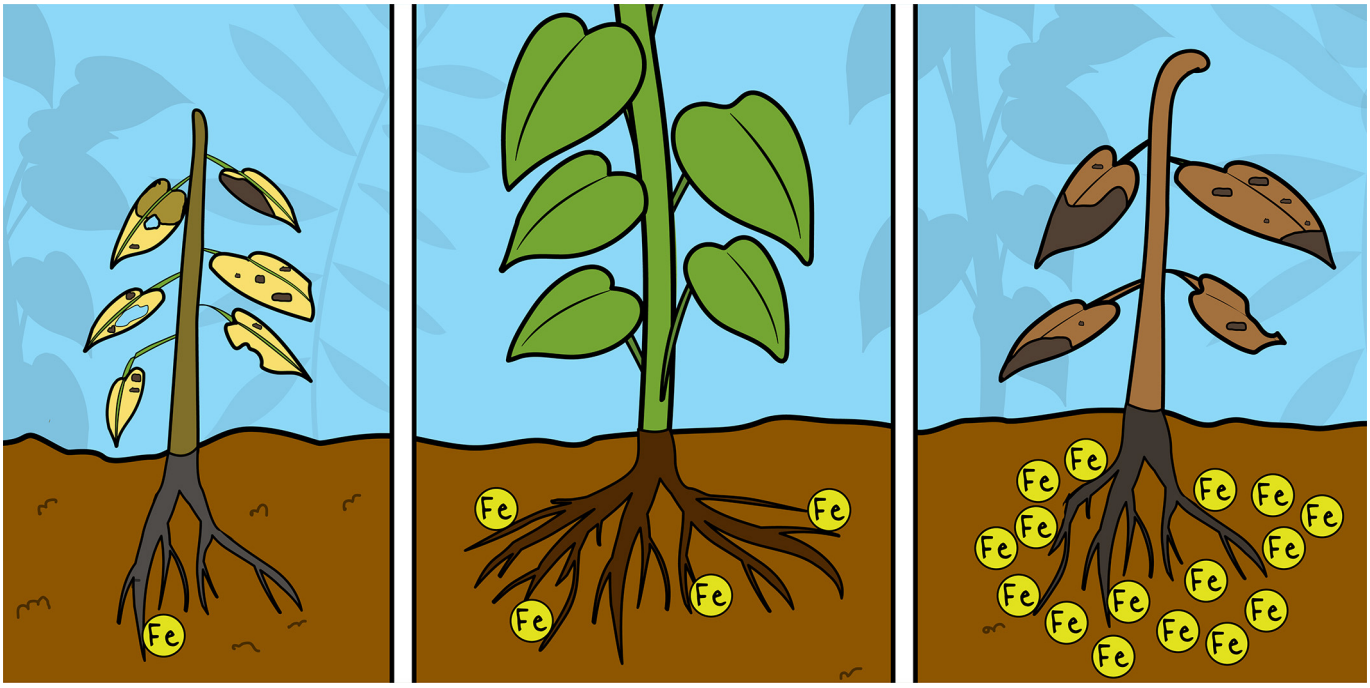
Abstract

What can you do when you feel hungry? You could go to the kitchen and prepare a sandwich, or you could buy yourself a snack. But what if you are fixed to the ground? Plants are champions at reaching food that is far away from them while they are standing still. Below a plant, roots branch out into the soil in many directions, looking for the nutrients plants need for survival and growth. Iron is a very important nutrient for plants because it is essential for growth and development, and it also helps plants to face stresses in the environment. Even if iron availability in soil is very low, plants have developed two strategies for efficiently taking up iron and storing it. In this article, we will explore the importance of iron in plants' lives, explaining how plants take up it and how balanced iron levels are important for plants' (and our) survival.

Keywords: Iron; Plants; Nutrition; Root; Pathogens

frontiers FOR YOUNG MINDS

EARTH AND ITS RESOURCES
Published: 24 May 2022
doi: 10.3389/frym.2022.718162



GOOD OR BAD? THE DOUBLE FACE OF IRON IN PLANTS

Francesca Marzorati[1*], *Alessia Midali*[1†], *Piero Morandini*[1] *and Irene Murgia*[1,2]

[1] *Department of Environmental Science and Policy, University of Milan, Milan, Italy*
[2] *Department of Biosciences, University of Milan, Milan, Italy*

YOUNG REVIEWERS:

IIS J. C. MAXWELL
AGES: 16–17

KING'S SCHOOL CANTERBURY
AGES: 14–15

What can you do when you feel hungry? You could go to the kitchen and prepare a sandwich, or you could buy yourself a snack. But what if you are fixed to the ground? Plants are champions at reaching food that is far away from them while they are standing still. Below a plant, roots branch out into the soil in many directions, looking for the nutrients plants need for survival and growth. Iron is a very important nutrient for plants because it is essential for growth and development, and it also helps plants to face stresses in the environment. Even if iron availability in soil is very low, plants have developed two strategies for efficiently taking up iron and storing it. In this article, we will explore the importance of iron in plants' lives, explaining how plants take up it and how balanced iron levels are important for plants' (and our) survival.

WHAT DO PLANTS EAT?

INORGANIC ION

An atom or a molecule with a positive or negative charge needed for vital cellular activity.

RHIZOSPHERE

The area of soil around roots colonized by bacteria and fungi.

Like animals, plants also need to "eat:" "mineral nutrition" is how plants acquire and absorb essential **inorganic ions**. Plants take up such nutrients from the soil using their roots. Importantly, microorganisms like fungi and bacteria living either in the nearby soil (called the **rhizosphere**) or inside plants' roots can sometimes help the plants obtain nutrients. Certain nutrients are very important for plants' survival, and a shortage of them can cause serious problems: indeed, nutrients, along with sunlight, are required to produce everything that plants need for their life. We can classify nutrients into two main groups: "macronutrients" if plants need them in large quantities and "micronutrients" which are only needed in small amounts.

Iron (Fe) is an essential micronutrient for many organisms, including humans: indeed, without iron, we would not have hemoglobin, which is the protein that transports oxygen in the blood. Iron also plays a key role in plants, where it participates in vital processes like photosynthesis. If plants do not get enough nutrients, they start to get "sick," and you can even see their sickness from their leaves. For example, a plant's leaves can become pale green/yellow (a phenomenon called "chlorosis") if they lack iron (Figure 1): cells located far from the veins ("interveinal cells") get the lowest amount of iron and become very pale ("chloros" means light green/yellow in ancient Greek). The lack of iron limits a plant's ability to perform photosynthesis and to produce the right amount of chlorophyll, which is the pigment that gives the green color to leaves.

Figure 1

A chlorotic strawberry leaf (right) compared to a healthy one (left). Iron deficiency causes problems with photosynthesis and the production of chlorophyll. A lack of chlorophyll makes leaves appear light green or yellow instead of healthy green (photo credit: Paolo Guarinoni).

Iron is one of the most important nutrients for plants, and its absence dramatically affects crop growth [1]. Despite iron's abundance in soil, its availability to plants is very reduced due to its limited solubility (iron is normally present in the chemical form Fe^{3+}; however, this chemical form poses problems in Fe uptake as it is not readily available to plants).

For this reason, plants have developed two different strategies for the uptake of this micronutrient.

HOW DO PLANTS ACQUIRE IRON?

Non-grass plants, such as tomato (whose scientific name is *Solanum lycopersicum*), take up iron using Strategy I, while grass-type crops, such as *Zea mays* (maize) and *Triticum aestivum* (bread wheat), use an alternative strategy named Strategy II (Figure 2) [2]. Both strategies are based on well-regulated molecular mechanisms involving two main actors: the root (particularly the **plasma membrane** of its cells) and the rhizosphere (the soil closest to roots). Special proteins called "transporters" are located in the plasma membranes, and these transporters help plants to move molecules either into or out of the roots as needed: plants could not "eat" without transporters, since they could not import nutrients inside their roots! You can imagine the transported molecules as cars and the transporters as tunnels that the cars use to get through mountains. These tunnels are often "one way," which means that they allow molecules to move either into the cells or out of them, but not both.

PLASMA MEMBRANE

Simply known as "membrane," it is the thin barrier, made out of lipids, between the inside and the outside of a cell.

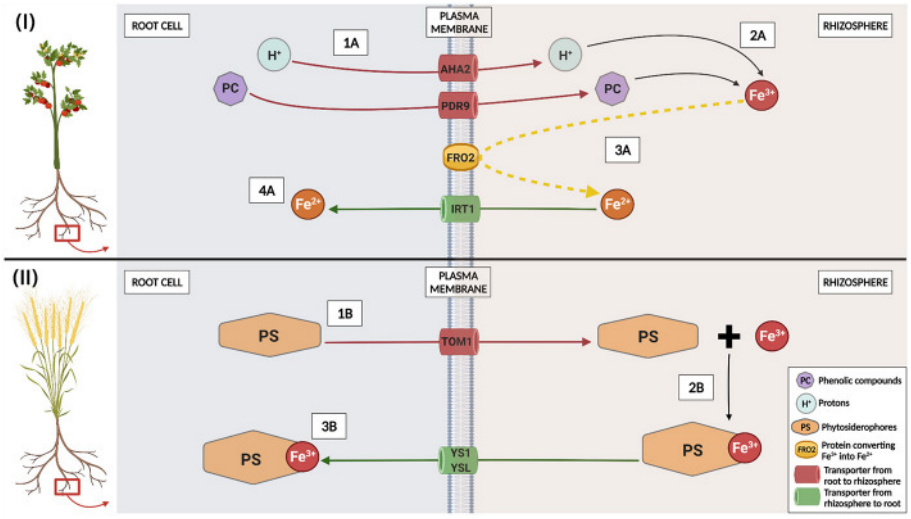


Figure 2

Figure 2

Strategies adopted by plants to take up iron from soil. Top—Strategy I (adopted by non-grass plants like tomato): phenolic compounds PC and protons H^+ exit from roots using transporters AHA2 and PDR9, respectively (1A, red arrows) and, in the rhizosphere, they increase iron Fe^{3+} levels (2A). Fe^{3+} is converted into Fe^{2+} by the plasma membrane protein FRO2 (3A, yellow dashed arrow). Fe^{2+} enters roots using transporter IRT1 (4A, green arrow). Below—Strategy II (used by grass plants like wheat): phytosiderophores PS exit from roots using transporter TOM1 (1B, red arrow); in the rhizosphere, they form a complex together with Fe^{3+} (2B). The complex PS-Fe^{3+} enters roots by YS1 or YSL transporters (3B, green arrow).

In soils, iron in the chemical form of Fe^{3+} is not readily available to plants because it forms chemical complexes which are poorly soluble and cannot be easily absorbed. Figure 2 shows strategies adopted by plants to overcome this problem and acquire iron. In Strategy I plants, protons H^+ and **phenolic compounds** (PC in Figure 2) are transported out from roots to the soil by using the transporters named AHA2 (H^+-TRANSLOCATING P-TYPE ATPase-2) and PDR9 (PLEIOTROPIC DRUG RESISTANCE 9; step 1A in Figure 2): both H^+ and phenolic compounds increase Fe^{3+} solubility (step 2A). AHA2 and PDR9 are "pumps," which means they are energy-driven transporters.

PHENOLIC COMPOUNDS

Chemical compounds characterized by an aromatic ring with an -OH group collectively called "phenol."

PHYTO-SIDEROPHORES

Organic compounds produced by plants under nutrient deficiency conditions and used to favor the uptake of nutrients in the rhizosphere.

ROOT HAIRS

Hairlike structures developed by roots to help in the acquisition of water and minerals from the soil.

Then, FRO2 (FERRIC REDUCTION OXIDASE 2), a protein localized inside the plasma membrane, converts iron Fe^{3+} into Fe^{2+} (step 3A): in such form, iron can be transported by IRT1 (IRON-REGULATED TRANSPORTER 1) inside plants' roots (step 4A). In Strategy II plants, there is no conversion of Fe^{3+} into Fe^{2+}: **phytosiderophores** (PS in Figure 2) are transported out from roots by TRANSPORTER OF MUGINEIC ACID 1 (TOM1; step 1B). Once in the rhizosphere, PS are able to directly bind iron in the form of Fe^{3+} (step 2B). The complex PS-Fe^{3+} can finally enter roots through YS1 (YELLOW STRIPE 1) or YSL (YELLOW STRIPE LIKE) transporters (step 3B). Although these different details on how iron is carried inside roots, both strategies are turned on once plants start lacking iron. Moreover, roots' shape changes a lot under iron deficiency conditions! The number of root branches increases, and roots become richer in **root hairs**, helping plants to take up iron from soil.

IRON AS A SUPERHERO

Once iron enters root cells, it can be transported far from roots to be used in different parts of the plant. If cells "feel" that an excess of iron is present, iron is safely stored within a "big" protein called "ferritin," formed from 24 subunits arranged in a form of a cage and that can house iron in its central hollow cavity. Interestingly, iron can also help plants to fight stresses: indeed, plants are fixed to the ground, and they cannot run away when some environmental conditions become harsh for them or enemies attack. The right supplies of iron can help plants to face environmental problems and to cope with pathogens or insects.

Fungi are among the worst plant pathogens, as they devastate cultivations all over the world and cause extensive yield losses. Iron seems to control and reduce many fungal diseases in wheat and barley, such as the ones caused by rusts fungi, pathogens that are not dangerous for us but are terrible for many economically important plants [3]. Plant defense strategies can be really complicated! In some cases, iron deficiency can help pathogens to infect plants, but in other cases, iron deficiency can help plants to fight their enemies. Pathogens also need nutrients to survive, and they try to steal the iron they need from plants: in some cases, plants reduce iron levels at the site where pathogens have attacked, trying to keep on fighting and preventing pathogens to use their nutrients [4, 5].

Iron in plants is not just a superhero for the plants themselves. Many useful bacteria and fungi in the rhizosphere are positively influenced by the iron levels in plants, and iron deficiency can cause big problems in these important communities [4]. Moreover, iron in plants can also help all of us: when we do not get enough micronutrients from our diets, we may become sick! Micronutrient deficiency is also known as "hidden hunger" because, even if you feel full after a meal, you

may not be getting sufficient nutrients from your food. Iron deficiency is the most widespread micronutrient deficiency in humans all over the world, and it causes a blood disease called "iron deficiency anemia" (IDA). According to the World Health Organization (WHO), more than 1.5 billion people suffer from IDA, which is around 25% of the worldwide population! Unfortunately, children and teenagers are among the most affected [6]. For this reason, scientists are studying how plants acquire and accumulate nutrients so that they can try to increase the amounts of nutrients in plants: "biofortification" is exactly the process of increasing the nutrients' content of plants by genetic means and not by direct supplementing of food [7]. An example of a biofortified crop with high levels of iron is the "iron bean," which was recently introduced in developing countries of Asia, South America, and Africa [8].

THE DARK SIDE OF IRON

Since iron seems to be so helpful, you might wonder why plants do not accumulate as much iron as possible to avoid the consequences of iron deficiency. Unfortunately, too much iron can be very dangerous to plants, weakening and eventually killing them. Iron levels must be well-controlled to prevent both deficiency and excess. Just imagine iron in plants as a sort of "Dr. Jekyll and Mr. Hyde" or think about iron levels for plant health as a "libra scale:" equilibrium is good, while extremes are deadly (Figure 3).

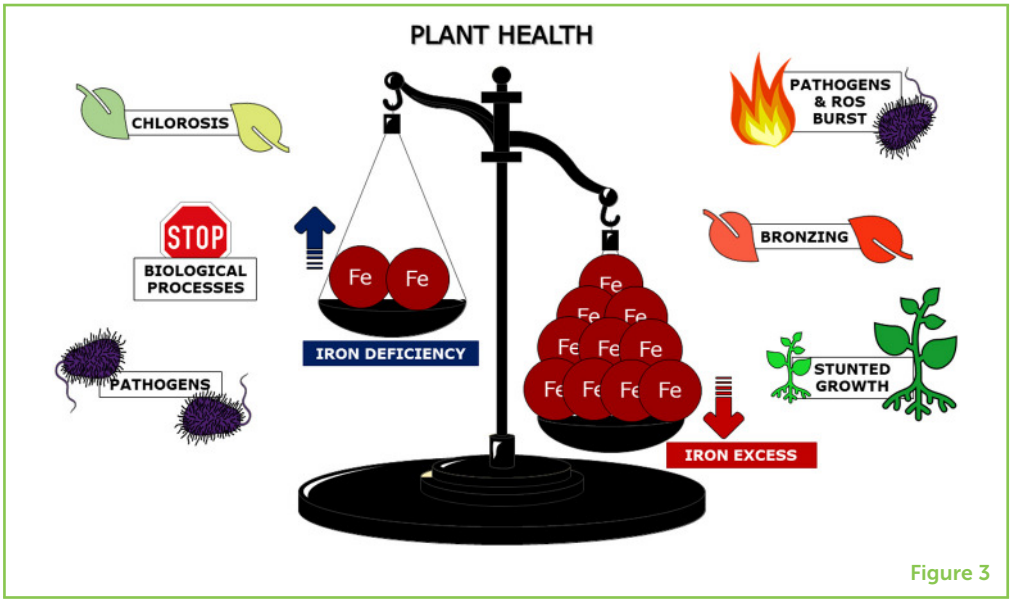


Figure 3

Figure 3

The proper balance of iron within plant cells is necessary for plant health. Both iron deficiency and iron excess can affect the health of plants, interfering with their abilities to face environmental stresses, to grow, and to survive.

Iron deficiency can dramatically weaken plants, but iron excess causes "stunted growth," which is when plants appear small and sick, and can also cause "bronzing of leaves," which is when leaves turn brownish-red. These dramatic events are linked to the generation of Reactive Oxygen Species (ROS), unstable oxygen-containing

molecules that can damage many cellular molecules and that can cause cell death. Indeed, iron participates in the production of these dangerous molecules [5].

CONCLUSION

The world's population is growing. Increasing the amount of iron in plants could help improve human health through better diets. We have begun to understand the importance of iron (and other nutrients) in plants, and we have seen how iron deficiency and iron excess can both negatively affect plants' lives. In the future, there are sure to be many new scientific findings explaining how plants take up nutrients and how we could produce nutrient-rich crops. This research will benefit people, especially poor people, all over the world and will help us to defeat the "hidden hunger" of iron once and for all.

ACKNOWLEDGMENTS

Figures were created using BioRender (Figure 2) or modifying Servier Medical Art templates (which are licensed under a Creative Commons Attribution 3.0 Unported License; `https://smart.servier.com`) and using PowerPoint (Microsoft Office 365; Figure 3). All authors deeply thank Paolo Guarinoni for giving the photo of Figure 1 to PM, and all the reviewers for their work and effort to improve this paper. FM sincerely acknowledges Alessandro for his critical reading of the manuscript.

REFERENCES

1. Rout, G. R., and Sahoo, S. 2015. Role of iron in plant growth and metabolism. *Rev. Agric. Sci.* 3:1–24. doi: 10.7831/ras.3.1
2. Kobayashi, T., and Nishizawa, N. K. 2012. Iron uptake, translocation, and regulation in higher plants. *Annu. Rev. Plant Biol.* 63:131–52. doi: 10.1146/annurev-arplant-042811-105522
3. Dordas, C. 2017. Role of nutrients in controlling the plant diseases in sustainable agriculture. *Agric. Important Microb. Sustain Agric.* 2:217–62. doi: 10.1007/978-981-10-5343-6_8
4. Verbon, E. H., Trapet, P. L., Stringlis, I. A., Kruijs, S., Bakker, P. A. H. M., and Pieterse, C. M. J. 2017. Iron and immunity. *Annu. Rev. Phytopathol.* 55:355–75. doi: 10.1146/annurev-phyto-080516-035537
5. Naranjo-Arcos, M. A., and Bauer, P. 2016. "Iron nutrition, oxidative stress, and pathogen defense," in *Nutritional Deficiency*, eds P. Erkekogu and B. Kocer-Gumusel (Intech Open). p. 63–98. doi: 10.5772/63204
6. De Benoist, B., McLean, E., Cogswell, M., Egli, I., and Wojdyla, D. 2008. Worldwide prevalence of anaemia 1993-2005. *Public Health Nutr.* 12:444–54. doi: 10.1017/S1368980008002401

7. Murgia, I., Arosio, P., Tarantino, D., and Soave, C. 2012. Biofortification for combating "hidden hunger" for iron. *Trends Plant Sci.* 17:47–55. doi: 10.1016/j.tplants.2011.10.003
8. Mulambu, J., Andersson, M., Palenberg, M., Pfeiffer, W., Saltzman, A., Birol, E., et al. 2017. Iron beans in Rwanda: crop development and delivery experience. *Afri. J. Food Agric. Nutr. Dev.* 17:12026–50. doi: 10.18697/ajfand.HarvestPlus010

SUBMITTED: 31 May 2021; **ACCEPTED:** 29 April 2022;
PUBLISHED ONLINE: 24 May 2022.

EDITOR: Melissa Hamner Mageroy, Norwegian Institute of Bioeconomy Research (NIBIO), Norway

SCIENCE MENTORS: Geoffrey Winston Nelson and Marta Dell'Orto

CITATION: Marzorati F, Midali A, Morandini P and Murgia I (2022) Good or Bad? The Double Face of Iron in Plants. Front. Young Minds 10:718162. doi: 10.3389/frym.2022.718162

CONFLICT OF INTEREST: AM was employed by the company Enerzyme Srl.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

YOUNG REVIEWERS

IIS J. C. MAXWELL, AGES: 16–17
We are a group of 22 Italian teenagers aged from 16 to 17. We attend a Scientific High School in Milan. We began our experience as reviewers last year, thanks to "Frontiers for young minds;" we enjoyed it so much that we decided to repeat the experience this year too. We have a strong fascination with science and our review activity is a way to deal with science out of the classroom and in addition to our "regular" Science classes. Moreover, we valued the opportunity to challenge ourselves to understand a content which is not written in our native language.

KING'S SCHOOL CANTERBURY, AGES: 14–15
We are an energetic year 11 class who are curious about the latest science. Three members of the class took on leadership roles as we reviewed this manuscript. We are happy that we can support working scientists through this review during the pandemic.

AUTHORS

FRANCESCA MARZORATI
Francesca Marzorati is a Ph.D. student in Environmental Sciences at the University of Milan. She has been interested in nature and plants since she was a child, but she did not consider becoming a (plant) biologist until she enrolled at university. She is fascinated by new technologies and computers: she studied bioinformatics and she is always trying to link programming to her laboratory work. In her spare time, she loves taking photos, reading, and playing video games. *francesca.marzorati1@unimi.it

ALESSIA MIDALI
Alessia Midali graduated with a degree in industrial biotechnology from the University of Milano-Bicocca. Her work there was focused on iron metabolism (uptake and transport) in plants, in particular in the model organism *Arabidopsis thaliana*. Even though she now works in a different field, she is still very interested in plant nutrition. †Present address: Alessia Midali, Actavis Generics, Milan, Italy

PIERO MORANDINI
Piero Morandini is a researcher in plant physiology at the Environmental Science and Policy Department of the University of Milano. He is an expert in metabolic control analysis.

IRENE MURGIA
Irene Murgia is a plant physiologist at the Biosciences Department of the University of Milano. She is an expert in plant nutrition and oxidative stress.

4. Concluding remarks

"Do not fear to be eccentric in opinion,
for every opinion now accepted was once eccentric."
— Bertrand Russell

As sessile organisms, plants cannot escape stress; thus, they must continuously protect themselves from environmental challenges. They are, indeed, constantly subjected to several environmental challenges, over a timeframe of seconds to days or months/years, such as variations in light, temperature, water or nutrient availability, and exposure to microorganisms. These can significantly affect plant growth and development, resulting in losses of agricultural production; however, we are far from describing how plants quickly coordinate their cellular activities in response to environmental challenges. Formate dehydrogenase (FDH) has been described as a 'stress protein' (Alekseeva *et al*., 2001) because its expression is induced by various environmental stressors. In *Arabidopsis thaliana*, *FDH* is a single gene that undergoes three splicing variants, as detailed in (https://www.arabidopsis.org/servlets/TairObject?type=locus&id=136173). The first variant appears to be the most prevalent, but it is possible that these variants are related to the protein's localization in the cell, especially considering that its chloroplastic localization was recently confirmed (Lee *et al*., 2022). When I began my Ph.D., only a few published studies investigated the involvement of FDH in defense against pathogens (David *et al*., 2010; Choi *et al*., 2014); thus, I explored the response of *FDH* in *Arabidopsis thaliana* leaves exposed to the vascular pathogen *Xanthomonas campestris* pv *campestris* (*Xcc*). I also investigated the foliar response of *A. thaliana* when its root apparatus was exposed to the beneficial rhizobacterium *Pseudomonas simiae* WCS417. For both analyses, *in silico* and *in vivo* approaches were used. Both approaches provide further support for the involvement of FDH in defense responses against pathogens, identifying FDH as a relevant node in the early defense response pathways involving hydathodes upon *Xcc* attack (Marzorati *et al*., 2021a).

It is important to underline that we used for the GUS experiments the *A. thaliana* transgenic line transformed with the fusion of *Vigna umbellata FDH* promoter with the reporter GUS (in short: *Vu* FDH::GUS), even though *Vu* FDH is not particularly related to

At FDH due to its being a leguminous plant (Lou *et al*., 2016). We decided to use the stably transformed *A. thaliana* line with the *Vu* FDH::GUS construct since it allowed to previously demonstrate that *FDH* is involved in various abiotic stresses (Lou *et al*., 2016; Murgia *et al*., 2020). Furthermore, Murgia *et al*. (2020) discovered a high *FDH* promoter activity in hydathodes using this *A. thaliana* transformed line. I observed a higher susceptibility to *Xcc* infection and spreading from hydathodes in an *FDH* knockout mutant *atfdh1-5* compared to its wild-type counterpart. These findings are in accordance with earlier observations that *fdh* mutants are more susceptible to bacterial pathogens (Choi *et al.*, 2014), as recently confirmed by Lee *et al*. (2022). I also observed a rapid decrease in *FDH* promoter activity in *A. thaliana* hydathodes upon infection with *Xcc*, which contrasts with data published in other studies that have highlighted an increase in *FDH* expression upon bacterial infection (Choi *et al.*, 2014; Lee *et al*., 2022). However, the reason for this discrepancy remains unclear and warrants further investigation. It could be argued that the results were due to the GUS reporter assay: FDH levels in other studies were mainly measured by RNA gel blotting and qRT-PCR, whereas I used the GUS staining method to highlight *FDH* promoter activity in hydathodes and how it varies upon infection. As stated by Yagi *et al*. (2021), a GUS reporter assay can overestimate gene expression in hydathodes because they are composed of several small cells, which means that signal may be substantially more intense per unit area than that in larger cells. Indeed, when the expression is at the level of the cytosol, the signal is distributed over a very large total cellular volume in mature cells due to the preponderance of the vacuole; on the other hand, in young cells or cells with small vacuoles, the signal will be more intense. To corroborate our findings, *FDH* expression in hydathodes should be further assessed using a fluorescent protein-based reporter and performing quantitative analyses such as qPCR and GUS activity or H_2O_2 quantifications. A transcriptomic assay could also be particularly useful and innovative because, to date, only one study has performed RNA-Seq analysis of *A. thaliana* hydathodes (Yagi *et al*., 2021). Alternatively, the observed discrepancy could be related to the adopted protocol for bacterial infection of leaves, as I only used detached leaves. It might be interesting to investigate what happens *in vivo* to *FDH* promoter activity in hydathodes following *Xcc* infection since such a pathogenic test could be valuable in elucidating the defensive signaling pathway in which FDH could be involved but has yet to be thoroughly described and proposed. Despite the potential issue with the bacterial inoculation protocol used, preliminary data on wt Col and *atfdh1-5* mutant plants infected *in vivo* are now available thanks to a collaboration with Dr. Emmanuelle Lauber and Prof. Laurent Noël of INRAE Toulouse LIPME (Laboratoire des

interactions plantes-microbes-environnement) (France). These findings confirm that the *atfdh1-5* accession has a higher susceptibility to *Xcc* (*personal communication*), as I observed infecting detached leaves (Marzorati *et al*., 2021a). My results may support the hypothesis that the defensive function of FDH in hydathodes is dependent on changes in formate levels and formate might act as a 'signal' for an early defense response. Further experiments focusing more on formate are therefore needed to clarify, and a metabolomic study of wt and *fdh* mutant plants infected with *Xcc*, with the profiling of a large number of metabolites, would help in understanding the metabolic changes occurring under *Xcc* attack, especially the changes in formate and the role of FDH in defense (Clish, 2015).

The hypotheses to be tested *in vivo* could also be formulated starting from the correlation analysis of *A. thaliana* gene expression under biotic stress, as combining *in vivo* with *in silico* approaches helps to identify candidate genes; correlation analysis can be indeed particularly important for formulating hypotheses regarding the roles of genes in metabolic pathways and processes (Månsson *et al*., 2004; Beekweelder *et al*., 2008; Berri *et al*., 2009; Fukushima *et al*., 2011; Abbruscato *et al*., 2012; Li *et al*., 2015; Zermiani *et al*., 2015; Murgia *et al*., 2020; Marzorati *et al*., 2021b). It would be beneficial, in this respect, that researchers could be able to perform simple *in silico* studies on their own, and they should be assisted in doing so by improving the quality of data available online (Marzorati *et al*., 2021b) and by developing new, user-friendly informatics tools to perform the analyses, such as NORMALIX95 (Marzorati *et al*., manuscript in preparation). NORMALIX95 is still under development and user feedback will be crucial for improving the tool even after its release. By integrating supplementary visualization tools, like heatmaps to depict the modulation of genes under distinct conditions (*e.g.*, *FDH* modulation during fungal attacks compared to bacterial infections), and score plots for Principal Component Analysis (PCA) (Jollife and Cadima, 2016), it is possible to gain a more comprehensive understanding of the data and conduct a more thorough analysis. The PCA feature in NORMALIX95 would prove to be especially beneficial as it allows for the detection of technical variations or non-biological differences among measurements obtained from different sample groups, commonly referred to as 'batch effect', and for identifying outliers that exhibit significant differences from the rest of the data and may considerably influence the analysis. These outliers can distort the overall graph structure, leading to incorrect conclusions about sample relationships; once identified, they can be either removed from the analysis or their impact can be minimized by using data normalization techniques like centering and scaling

(Bro and Smilde, 2003). **Figure 8** illustrates the importance of using PCA as an initial tool to detect and eliminate outliers. By comparing the MtGEA dataset before and after applying the cleaning strategy proposed by Marzorati *et al*. (2021b), it becomes clear that the removal of outliers leads to a more consistent dataset.

A specific dataset was built by recollecting all *Affymetrix* microarray data available in the repository Gene Expression Omnibus (GEO) (Edgar, 2002; Barrett *et al*., 2013) for *A. thaliana* under biotic stress, to define the top correlators of *FDH* under such stress condition. Conducting further investigations on *FDH* co-expressed genes under biotic stress across diverse plant species is an intriguing prospect. Specifically, exploring the consistency of the best *FDH* correlators identified in *A. thaliana* across other species, such as *M. truncatula* (whose microarray dataset has already been cleaned (Marzorati *et al*., 2012b)) or *Vitis vinifera* (for which several informatics resources are available, like VitViz (*unpublished*, http://vitviz.tomsbiolab.com/LeafNetwork/) or Vitis OneGeneE (Pilati *et al*., 2021)), could provide valuable insights. Even if the MtGEA database is no longer accessible online, by developing and releasing NORMALIX95, we aim to make several datasets usable with our tool. Therefore, we will make efforts to make both the original and cleaned *M. truncatula* datasets available for future research. The top correlators of *FDH* identified working on *A. thaliana* were genes that are also involved in defense responses, particularly those related to the phytohormone jasmonic acid (JA). This hormone, along with ethylene (ET), is involved in an immune defense response known as 'Induced Systemic Resistance' (ISR), which is stimulated by the activity of the plant root microbiota and appears to be active against a wide range of infections (van Loon *et al*., 1998; Pieterse *et al*., 2014; Vlot *et al*., 2021).

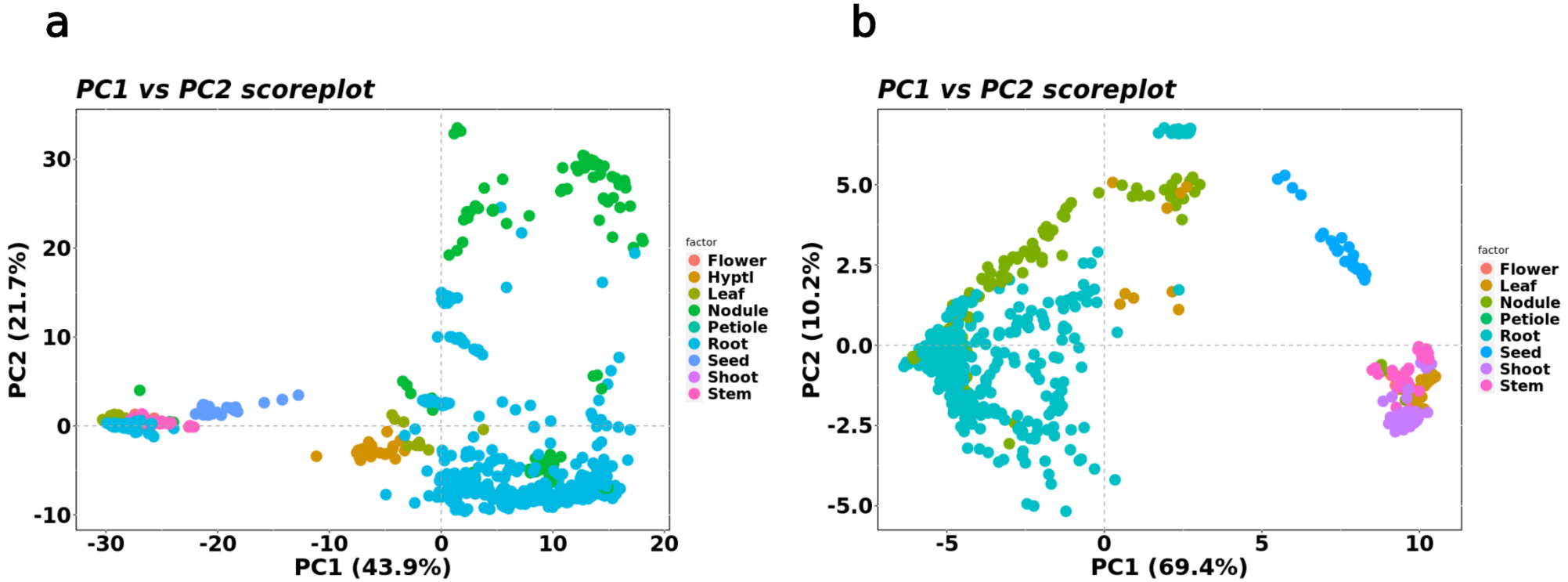


Figure 8. Plots for the Principal Component Analysis (PCA) conducted on the *Medicago truncatula* Gene Expression Atlas (MtGEA) before and after data cleaning. (a) The PCA was performed on the original MtGEA database before the application of the cleaning strategy proposed by Marzorati *et al*. (2021b). **(b)** The PCA was performed on MtGEA database after the application of the cleaning strategy proposed by Marzorati *et al*. (2021b). The application of the cleaning strategy resulted in a dataset with a reduced number of outliers and more homogeneous sample groups, underlying the strategic importance of PCA as an initial visualization tool for outlier detection and removal. The reason for the difference in color legend between (a) and (b) is due to the removal of the entire group of samples called 'Hyptl' during the cleaning process. The PCA analysis was conducted using R software (version 4.0.3).

Pseudomonas simiae WCS417 is most likely the best-studied rhizobacterium-inducing ISR in *A. thaliana* (Pieterse *et al*., 2020); furthermore, the pathway induced by *P. simiae* WCS417 partially overlaps with the iron (Fe) deficiency response (Romera *et al*., 2019; Verbon *et al*., 2019), and FDH has already been identified as a hub in plant iron nutrition by *in vivo* and *in silico* studies (Vigani *et al*., 2017; Murgia *et al*., 2020; Di Silvestre *et al*., 2021). Given the lack of information regarding the possible involvement of FDH in the *P. simiae* WCS417-induced pathway, I studied *FDH* promoter activity in *A. thaliana* seedlings and leaves after their roots were directly or indirectly (without contact) exposed to *P. simiae* WCS417. Surprisingly, I found that *P. simiae* WCS417 can rapidly induce *FDH* promoter activity in hydathodes, with or without direct colonization of plant roots, suggesting that *FDH* induction can be mediated by rhizobacterium-produced volatile compounds (Marzorati *et al*., manuscript under review). Plants produce VOCs, and so do bacteria, as a result of interactions with biotic and abiotic stimuli (Dudareva *et al*., 2013; Picazo-Aragonés *et al*., 2020); elucidating the complex metabolic network in which FDH acts requires studies to identify the volatile compounds, possibly of organic nature, produced by bacteria and/or plants and responsible for *FDH* induction in leaves. Notably, our total leaf proteome analysis of *A. thaliana* wt Col and the *FDH* knockout mutant *atfdh1-5* demonstrated that a rapid *P. simiae* WCS417 exposure affected the production of proteins involved in fundamental metabolic processes, with several photosynthesis-related proteins being downregulated in both *A. thaliana* lines within two days after inoculation. Indeed, extrinsic proteins from photosystems PSI and PSII, as well as stress-responsive proteins and ROS-scavenging enzymes, seemed to be early targets of the metabolic changes caused by *P. simiae* WCS417 in the two plant lines and, therefore, irrespective of FDH activity, at least within the time frame of the experiment. Furthermore, FDH protein levels increased when wt plants were exposed to the rhizobacterium; this result supports our observations of the rapid induction of *FDH* promoter activity in *A. thaliana* leaves in plants exposed to *P. simiae* WCS417 (with FDH being affected also at the protein level) and the possibility that FDH is involved in the establishment of early leaf responses upon *P. simiae* WCS417 root colonization. Our analysis of the total leaf proteome also revealed changes in proteins related to Fe deficiency; focusing our attention on the available leaf proteomic data under Fe deficiency (Zargar *et al*., 2013), we found some proteins with a similar regulation in our total leaf proteome analysis. This result confirms the association between the *P. simiae* WCS417-induced pathway and the Fe deficiency response in *A. thaliana*; the identified proteins should

be pursued further since could represent linking nodes between ISR and Fe deficiency pathways. No data on the early effects of rhizobacteria on leaf metabolic pathways have been published, to our knowledge, as most studies focused on the effects of rhizobacteria on roots several days after inoculation (Zamioudis *et al*., 2013; Trapet *et al*., 2016; Wintermans *et al*., 2016; Verbon *et al*., 2019); therefore, our leaf proteomic analysis has also important implications for developing new *in vivo* and *in silico* studies about the relationship between rhizobacteria effect and Fe levels, particularly in leaves, trying to discover how FDH could be involved in such a network. For instance, given the existence of common nodes, at least in *A. thaliana*, between the *P. simiae* WCS417-induced immune pathway and the Fe deficiency response, it may be worthwhile to test the effects of various Fe nutritional conditions (excess and deficiency) on FDH activity in plants exposed to *P. simiae* WCS417. In addition, a comparison between the co-expressed genes of *FDH* in leaves and the proteins whose levels are induced upon exposure to *P. simiae* WCS417, followed by an enrichment analysis to determine the biological pathways that are enriched in each set and between sets, could yield valuable insights into the potential impact on Fe metabolism.

Plants interact with a large number of microorganisms, which can be either beneficial or harmful. The relationships between plants and beneficial microbes in the soil impact several, possibly interconnected, aspects of plant nutrition and defense responses, and it is becoming increasingly obvious that the 'below ground' part of plants is as crucial for defense and nutrition as the aerial part. I hope this work will lay the ground for future research into plant defense mechanisms and the role of FDH in responses to both harmful and beneficial microbes. Although FDH was discovered over a century ago (Alekseeva *et al*., 2011), its function in plants is still unclear; the increasing abundance of FDH in stress conditions suggests a role in stress, but this remains, so far, a mystery. Understanding FDH function might help to exploit this protein by modulating its activity for instance in defense processes, with possibly an impact on agricultural production, by enhancing crop productivity and minimizing losses.

5. References

Abbruscato P, Nepusz T, Mizzi L, *et al.* 2012. OsWRKY22, a monocot WRKY gene, plays a role in the resistance response to blast. Molecular Plant Pathology **13**, 828-841.

Achkor H, Díaz M, Fernández MR, Biosca JA, Parés X, Martínez MC. 2003. Enhanced formaldehyde detoxification by overexpression of glutathione-dependent formaldehyde dehydrogenase from *Arabidopsis*. Plant Physiology **132**, 2248-2255.

Afzal I, Shinwari ZK, Sikandar S, Shahzad S. 2019. Plant beneficial endophytic bacteria: mechanisms, diversity, host range and genetic determinants. Microbiological Research **221**, 36-49.

Ahmad IA, Zafar A, Sohail S, Younas N, Haider S, Shah MS. 2021. Alleviation of iron deficiency in plants, animals and humans through biofortification: a detailed review. International Journal of Innovative Science and Research Technology **6**, 5-12.

Alekseeva AA, Savin SS, Tishkov VI. 2011. NAD+-dependent formate dehydrogenase from plants. Acta Naturae **3**, 38-54.

Allier A, Teyssèdre S, Lehermeier C, Moreau L, Charcosset A. 2020. Optimized breeding strategies to harness genetic resources with different performance levels. BMC Genomics **21**, 1DUMM.

Aloni R, Aloni E, Langhans M, Ullrich CI. 2006. Role of cytokinin and auxin in shaping root architecture: regulating vascular differentiation, lateral root initiation, root apical dominance and root gravitropism. Annals of Botany **97**, 883-893.

Aloni R, Schwalm K, Langhans M, Ullrich CI. 2003. Gradual shifts in sites of free-auxin production during leaf-primordium development and their role in vascular differentiation and leaf morphogenesis in *Arabidopsis*. Planta **216**, 841-853.

Alonso-Blanco C, Andrade J, Becker C, *et al.* 2016. 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. Cell **166**, 481-491.

Altshuler D, Daly M, Kruglyak L. 2000. Guilt by association. Nature Genetics **26**, 135-137.

Amory AM, Cresswell CF. 1986. Role of formate in the photorespiratory metabolism of *Themeda triandra* Forssk. Journal of Plant Physiology **124**, 247-255.

Andreadeli A, Flemetakis E, Axarli I, Dimou M, Udvardi MK, Katinakis P, Labrou NE. 2009. Cloning and characterization of *Lotus japonicus* formate dehydrogenase: a possible correlation with hypoxia. Biochimica et Biophysica Acta - Proteins and Proteomics **1794**, 976-984.

Andrews NC. 2000. Iron metabolism: iron deficiency and iron overload. Annual Review of Genomics and Human Genetics **1**, 75-98.

Aznar A, Chen NWG, Rigault M, *et al.* 2014. Scavenging iron: a novel mechanism of plant immunity activation by microbial siderophores. Plant Physiology **164**, 2167-2183.

Aznar A, Chen NWG, Thomine S, Dellagi A. 2015. Immunity to plant pathogens and iron homeostasis. Plant Science **240**, 90-97.

Aznar A, Dellagi A. 2015. New insights into the role of siderophores as triggers of plant immunity: what can we learn from animals? Journal of Experimental Botany **66**, 3001-3010.

Bakker PAHM, Berendsen RL, Doornbos RF, Wintermans PCA, Pieterse CMJ. 2013. The rhizosphere revisited: root microbiomics. Frontiers in Plant Science **4**, 165.

Bakker PAHM, Berendsen RL, van Pelt JA, *et al.* 2020. The soil-borne identity and microbiome-assisted agriculture: looking back to the future. Molecular Plant **13**, 1394-1401.

Bar-Even A, Flamholz A, Noor E, Milo R. 2012. Thermodynamic constraints shape the structure of carbon fixation pathways. Biochimica et Biophysica Acta - Bioenergetics **1817**, 1646-1659.

Barker DG, Bianchi S, Blondon F, *et al.* 1990. *Medicago truncatula*, a model plant for studying the molecular genetics of the rhizobium-legume symbiosis. Plant Molecular Biology Reporter **8**, 40-49.

Barrett T, Wilhite SE, Ledoux P, *et al.* 2013. NCBI GEO: archive for functional genomics data sets - Update. Nucleic Acids Research **41**, 991-995.

Becana M, Moran JF, Iturbe-Ormaetxe I. 1998. Iron-dependent oxygen free radical generation in plants subjected to environmental stress: toxicity and antioxidant protection. Plant and Soil **201**, 137-147.

Beekweelder J, van Leeuwen W, van Dam NM, *et al*. 2008. The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. PLOS ONE **3**, e2068.

Bénaben V, Duc G, Lefebvre V, Huguet T. 1995. TE7, an inefficient symbiotic mutant of *Medicago truncatula* Gaertn. cv Jemalong. Plant Physiology **107**, 53-62.

Benedito VA, Torres-Jerez I, Murray JD, *et al.* 2008. A gene expression atlas of the model legume *Medicago truncatula*. Plant Journal **55**, 504-513.

Beneduzi A, Ambrosini A, Passaglia LMP. 2012. Plant growth-promoting rhizobacteria (PGPR): their potential as antagonists and biocontrol agents. Genetics and Molecular Biology **35**, 1044-1051.

Berendsen RL, Pieterse CMJ, Bakker PAHM. 2012. The rhizosphere microbiome and plant health. Trends in Plant Science **17**, 478-486.

Berendsen RL, van Verk MC, Stringlis IA, Zamioudis C, Tommassen J, Pieterse CMJ, Bakker PAHM. 2015. Unearthing the genomes of plant-beneficial *Pseudomonas* model strains WCS358, WCS374 and WCS417. BMC Genomics **16**, 539.

Berg T, Tesoriero L, Hailstones DL. 2005. PCR-based detection of *Xanthomonas campestris* pathovars in *Brassica* seed. Plant Pathology **54**, 416-427.

Berri S, Abbruscato P, Faivre-Rampant O, *et al.* 2009. Characterization of WRKY co-regulatory networks in rice and *Arabidopsis*. BMC Plant Biology **9**, 120.

Bilsborough GD, Runions A, Barkoulas M, Jenkins HW, Hasson A, Galinha C, Laufs P, Hay A, Prusinkiewicz P, Tsiantis M. 2011. Model for the regulation of *Arabidopsis thaliana* leaf margin development. Proceedings of the National Academy of Sciences of the United States of America **108**, 3424-3429.

Bolstad BM, Irizarry RA, Åstrand M, Speed TP. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics **19**, 185-193.

Borrelli VMG, Brambilla V, Rogowsky P, Marocco A, Lanubile A. 2018. The enhancement of plant disease resistance using CRISPR/Cas9 technology. Frontiers in Plant Science **9**, 1245.

Boscari A, del Giudice J, Ferrarini A, Venturini L, Zaffini AL, Delledonne M, Puppo A. 2013. Expression dynamics of the *Medicago truncatula* transcriptome during the symbiotic interaction with *Sinorhizobium meliloti*: which role for nitric oxide? Plant Physiology **161**, 425-439.

Bouis HE, Hotz C, McClafferty B, Meenakshi J v., Pfeiffer WH. 2011. Biofortification: a new tool to reduce micronutrient malnutrition. Food and Nutrition Bulletin **32**, 31-40.

Boukari N, Jelali N, Renaud JB, Youssef R ben, Abdelly C, Hannoufa A. 2019. Salicylic acid seed priming improves tolerance to salinity, iron deficiency and their combined effect in two ecotypes of Alfalfa. Environmental and Experimental Botany **167**, 103820.

Bounejmate M, Robson AD. 1992. Differential tolerance of genotypes of *Medicago truncatula* to low pH. Australian Journal of Agricultural Research **43**, 731-737.

Bounejmate M, Loss SP, Robson AD. 1994. Effects of temperature and frost on genotypes of *Medicago truncatula* L. and *Medicago aculeata* L. from contrasting climatic origins. Journal of Agronomy and Crop Science **172**, 227-236.

Breitling R. 2010. What is systems biology? Frontiers in Physiology **1**, 9.

Bro R, Smilde AK. 2003. Centering and scaling in component analysis. Journal of Chemometrics **17**, 16–33.

Broadley MR, Brown P, Cakmak I, Rengel Z, Zhao F. 2012. Function of nutrients: micronutrients. In: Marschner P (eds.). Mineral nutrition of higher plants. Academic Press, 191-248.

Brown PO, Botstein D. 1999. Exploring the new world of the genome with DNA microarrays. Nature Genetics **21**, 33-37.

Buckhout TJ, Yang TJW, Schmidt W. 2009. Early iron-deficiency-induced transcriptional changes in *Arabidopsis* roots as revealed by microarray analyses. BMC Genomics **10**, 147.

Bumgarner R. 2013. Overview of DNA microarrays: types, applications, and their future. In: John Wiley & Sons (eds.). Current protocols in molecular biology, Chapter 22: Unit 22.1.

Cameron DD, Neal AL, van Wees SCM, Ton J. 2013. Mycorrhiza-induced resistance: more than the sum of its parts? Trends in Plant Science **18**, 539-545.

Canchignia H, Altimira F, Montes C, Sánchez E, Tapia E, Miccono M, Espinoza D, Aguirre C, Seeger M, Prieto H. 2017. Candidate nematicidal proteins in a new *Pseudomonas veronii* isolate identified by its antagonistic properties against *Xiphinema index*. Journal of General and Applied Microbiology **63**, 11-21.

Carlton WM, Braun EJ, Gleason ML. 1998. Ingress of *Clavibacter michiganensis* subsp. *michiganensis* into tomato leaves through hydathodes. Phytopathology **88**, 525-529.

Carvalhais LC, Muzzi F, Tan CH, Hsien-Choo J, Schenk PM. 2013. Plant growth in *Arabidopsis* is assisted by compost soil-derived microbial communities. Frontiers in Plant Science **4**, 235.

Cerutti A, Jauneau A, Laufs P, Leonhardt N, Schattat MH, Berthomé R, Routaboul J-M, Noël LD. 2019. Mangroves in the leaves: anatomy, physiology, and immunity of epithemal hydathodes. Annual Review of Phytopathology **57**, 91-116.

Chapelle E, Mendes R, Bakker PAHM, Raaijmakers JM. 2016. Fungal invasion of the rhizosphere microbiome. ISME Journal **10**, 265-268.

Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SPA. 1996. Accessing genetic information with high-density DNA arrays. Science **274**, 610-614.

Chen CC, Chen YR. 2006. Study on laminar hydathodes of *Ficus formosana* (Moraceae) II. Morphogenesis of hydathodes. Botanical Studies **47**, 279-292.

Chialva M, Lanfranco L, Bonfante P. 2022. The plant microbiota: composition, functions, and engineering. Current Opinion in Biotechnology **73**, 135-142.

Chiappero J, Cappellari L del R, Sosa Alderete LG, Palermo TB, Banchio E. 2019. Plant growth promoting rhizobacteria improve the antioxidant status in *Mentha piperita* grown under drought stress leading to an enhancement of plant growth and total phenolic content. Industrial Crops and Products **139**, 111553.

Choi DS, Kim NH, Hwang BK. 2014. Pepper mitochondrial formate dehydrogenase 1 regulates cell death and defense responses against bacterial pathogens. Plant Physiology **166**, 1298-1311.

Choi K, Ratner N. 2019. IGEAK: an interactive gene expression analysis kit for seamless workflow using the R/shiny platform. BMC Genomics **20**, 177.

Clish CB. 2015. Metabolomics: an emerging but powerful tool for precision medicine. Molecular Case Studies **1**, a000588.

Compant S, Samad A, Faist H, Sessitsch A. 2019. A review on the plant microbiome: ecology, functions, and emerging trends in microbial application. Journal of Advanced Research **19**, 29-37.

Connorton JM. 2017. Iron homeostasis in plants – a brief overview. Metallomics **9**, 813-823.

Conrath U, Beckers GJM, Langenbach CJG, Jaskiewicz MR. 2015. Priming for enhanced defense. Annual Review of Phytopathology **53**, 97-119.

Cook AA, Larson RH, Walker JC. 1952. Relation of the black rot pathogen to cabbage seed. Phytopathology **42**, 316-320.

Cook DR. 1999. *Medicago truncatula* - a model in the making! Current Opinion in Plant Biology **2**, 301-304.

Crowley DE, Römheld V, Marschner H, Szaniszlo PJ. 1992. Root-microbial effects on plant iron uptake from siderophores and phytosiderophores. Plant and Soil **142**, 1-7.

Cui Y, Chen CL, Cui M, Zhou WJ, Wu HL, Ling HQ. 2018. Four IVa bHLH transcription factors are novel interactors of FIT and mediate JA inhibition of iron uptake in *Arabidopsis*. Molecular Plant **11**, 1166-1183.

Dalma-Weiszhausz DD, Warrington J, Tanimoto EY, Miyada CG. 2006. The *Affymetrix* GeneChip® Platform: an overview. Methods in Enzymology **410**, 3-28.

Damiri N, Mulawarman, Umayah A, Agustin SE, Rahmiyah M. 2018. Effect of *Pseudomonas* spp on infection of *Peronosporaparasitica* (Pers. Fr), the pathogen of downy mildew on Chinese cabbage. IOP Conference Series: Earth and Environmental Science **102**.

Dardanelli MS, Manyani H, González-Barroso S, Rodríguez-Carvajal MA, Gil-Serrano AM, Espuny MR, López-Baena FJ, Bellogín RA, Megías M, Ollero FJ. 2010. Effect of the presence of the plant growth promoting rhizobacterium (PGPR) *Chryseobacterium balustinum* Aur9 and salt stress in the pattern of flavonoids exuded by soybean roots. Plant and Soil **328**, 483-493.

David P, des Francs-Small CC, Sévignac M, Thareau V, Macadré C, Langin T, Geffroy V. 2010. Three highly similar formate dehydrogenase genes located in the vicinity of the B4 resistance gene cluster are differentially expressed under biotic and abiotic stresses in *Phaseolus vulgaris*. Theoretical and Applied Genetics **121**, 87-103.

D'Erfurth I, Cosson V, Eschstruth A, Lucas H, Kondorosi A, Ratet P. 2003. Efficient transposition of the Tnt1 tobacco retrotransposon in the model legume *Medicago truncatula*. Plant Journal **34**, 95-106.

de Weert S, Vermeiren H, Mulders IHM, Kuiper I, Hendrickx N, Bloemberg G v., Vanderleyden J, de Mot R, Lugtenberg BJJ. 2002. Flagella-driven chemotaxis towards exudate components is an important trait for tomato root colonization by *Pseudomonas fluorescens*. Molecular Plant-Microbe Interactions **15**, 1173-1180.

de Weger LA, van der Vlugt CIM, Wijfjes AHM, Bakker PA, Schippers B, Lugtenberg B. 1987. Flagella of a plant-growth-stimulating *Pseudomonas fluorescens* strain are required for colonization of potato roots. Journal of Bacteriology **169**, 2769-2773.

Desrut A, Moumen B, Thibault F, le Hir R, Coutos-Thévenot P, Vriet C. 2020. Beneficial rhizobacteria *Pseudomonas simiae* WCS417 induce major transcriptional changes in plant sugar transport. Journal of Experimental Botany **71**, 7301-7315.

Di Silvestre D, Vigani G, Mauri P, Hammadi S, Morandini P, Murgia I. 2021. Network topological analysis for the identification of novel hubs in plant nutrition. Frontiers in Plant Science **12**, 629013.

Dinneny JR, Long TA, Wang JY, Jung JW, Mace D, Pointer S, Barron C, Brady SM, Schiefelbein J, Benfey PN. 2008. Cell identity mediates the response of *Arabidopsis* roots to abiotic stress. Science **320**, 942-945.

do Amaral FP, Tuleski TR, Pankievicz VCS, *et al.* 2020. Diverse bacterial genes modulate plant root association by beneficial bacteria. mBio **11**, e03078-20.

Dong ZC, Chen Y. 2013. Transcriptomics: advances and approaches. Science China Life Sciences **56**, 960-967.

Douglas AE, Werren JH. 2016. Holes in the hologenome: why host-microbe symbioses are not holobionts. mBio **7**, e02099.

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. 2010. MYB transcription factors in *Arabidopsis*. Trends in Plant Science **15**, 573-581.

Dudareva N, Klempien A, Muhlemann JK, Kaplan I. 2013. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. New Phytologist **198**, 16-32.

Earhart CF. 2009. Iron metabolism. In: Schaechter M (eds.). Encyclopedia of Microbiology (Third Edition). Academic Press, 210-218.

Edgar R. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research **30**, 207-210.

Eijssen LMT, Jaillard M, Adriaens ME, Gaj S, de Groot PJ, Müller M, Evelo CT. 2013. User-friendly solutions for microarray quality control and pre-processing on ArrayAnalysis.org. Nucleic acids research **41**, 71-76.

Ekins R, Chu FW. 1999. Microarrays: their origins and applications. Trends in Biotechnology **17**, 217-218.

Expósito RG, de Bruijn I, Postma J, Raaijmakers JM. 2017. Current insights into the role of rhizosphere bacteria in disease suppressive soils. Frontiers in Microbiology **8**, 2529.

Feng H, Fu R, Hou X, *et al.* 2021. Chemotaxis of beneficial rhizobacteria to root exudates: the first step towards root–microbe rhizosphere interactions. International Journal of Molecular Sciences **22**, 6655.

Fenton HJH. 1894. Oxidation of tartatic acid in presence of iron. Journal of the Chemical Society, Transactions **65**, 899-910.

Ferguson BJ, Mathesius U. 2014. Phytohormone regulation of legume-rhizobia interactions. Journal of Chemical Ecology **40**, 770-790.

Ferry JG. 1990. Formate dehydrogenase. FEMS Microbiology Letters **87**, 377-382.

Fourcroy P, Sisó-Terraza P, Sudre D, Savirón M, Reyt G, Gaymard F, Abadía A, Abadia J, Álvarez-Fernández A, Briat JF. 2014. Involvement of the ABCG37 transporter in secretion of scopoletin and derivatives by *Arabidopsis* roots in response to iron deficiency. New Phytologist **201**, 155-167.

Fourcroy P, Tissot N, Gaymard F, Briat JF, Dubos C. 2016. Facilitated Fe nutrition by phenolic compounds excreted by the *Arabidopsis* ABCG37/PDR9 transporter requires the IRT1/FRO2 high-affinity root Fe^{2+} transport system. Molecular Plant **9**, 485-488.

Fox JT, Stover PJ. 2008. Folate-mediated one-carbon metabolism. Vitamins and Hormones **79**, 1-44.

Fu ZQ, Dong X. 2013. Systemic acquired resistance: turning local infection into global defense. Annual Review of Plant Biology **64**, 839-863.

Fukushima EO, Seki H, Ohyama K, Ono E, Umemoto N, Mizutani M, Saito K, Muranaka T. 2011. CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. Plant and Cell Physiology **52**, 2050-2061.

Gabriel R, Schäfer L, Gerlach C, Rausch T, Kesselmeier J. 1999. Factors controlling the emissions of volatile organic acids from leaves of *Quercus ilex* L. (Holm oak). Atmospheric Environment **33**, 1347-1355.

Ganeshan G, Kumar AM. 2005. *Pseudomonas fluorescens*, a potential bacterial antagonist to control plant diseases. Journal of Plant Interactions **1**, 123-134.

Gao F, Dubos C. 2021. Transcriptional integration of plant responses to iron availability. Journal of Experimental Botany **72**, 2056-2070.

Gao F, Robe K, Gaymard F, Izquierdo E, Dubos C. 2019. The transcriptional control of iron homeostasis in plants: a tale of bHLH transcription factors? Frontiers in Plant Science **10**, 6.

García MJ, Lucena C, Romera FJ, Alcántara E, Pérez-Vicente R. 2010. Ethylene and nitric oxide involvement in the up-regulation of key genes related to iron acquisition and homeostasis in *Arabidopsis*. Journal of Experimental Botany **61**, 3885-3899.

García MJ, Romera FJ, Lucena C, Alcántara E, Pérez-Vicente R. 2015. Ethylene and the regulation of physiological and morphological responses to nutrient deficiencies. Plant Physiology **169**, 51-60.

Garcia K, Bücking H, Zimmermann SD. 2020. Editorial: Importance of root symbiomes for plant nutrition: new insights, perspectives and future challenges. Frontiers in Plant Science **11**, 594.

Gardiner-Garden M, Littlejohn TG. 2001. A comparison of microarray databases. Briefings in bioinformatics **2**, 143-158.

Gentleman RC, Carey VJ, Bates DM, *et al.* 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome biology **5**, R80.

Gholami A, de Geyter N, Pollier J, Goormachtig S, Goossens A. 2014. Natural product biosynthesis in *Medicago* species. Natural Product Reports **31**, 356-380.

Giehl RFH, Lima JE, von Wirén N. 2012. Localized iron supply triggers lateral root elongation in *Arabidopsis* by altering the AUX1-mediated auxin distribution. Plant Cell **24**, 33-49.

Goatley JL, Lewis RW. 1966. Composition of guttation fluid from rye, wheat, and barley seedlings. Plant Physiology **41**, 373-375.

Gomez-Cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merkenschlager M, Gisel A, Ballestar E, Bongcam-Rudloff E, Conesa A, Tegnér J. 2014. Data integration in the era of omics: current and future challenges. BMC systems biology **8**, I1.

Graham PH, Vance CP. 2003. Legumes: importance and constraints to greater use. Plant Physiology **131**, 872-877.

Grillet L, Schmidt W. 2019. Iron acquisition strategies in land plants: not so different after all. New Phytologist **224**, 11-18.

Grunwald I, Rupprecht I, Schuster G, Kloppstech K. 2003. Identification of guttation fluid proteins: the presence of pathogenesis-related proteins in non-infected barley plants. Physiologia Plantarum **119**, 192-202.

Gu S, Yang T, Shao Z, *et al.* 2020. Siderophore-mediated interactions determine the disease suppressiveness of microbial consortia. mSystems **5**, e00811-19.

Gupta M, Vikram A, Bharat N. 2013. Black rot - a devastating disease of crucifers: a review. Agricultural Reviews **34**, 269-278.

Haas D, Défago G. 2005. Biological control of soil-borne pathogens by fluorescent pseudomonads. Nature Reviews Microbiology **3**, 307-319.

Haberlandt G. 1914. Physiological plant anatomy (trans. from the fourth German edition by Montagu Drummond).

Hanson AD, Roje S. 2001. One-carbon metabolism in higher plants. Annual Review of Plant Physiology and Plant Molecular Biology **52**, 119-137.

Harbort CJ, Hashimoto M, Inoue H, *et al.* 2020. Root-secreted coumarins and the microbiota interact to improve iron nutrition in *Arabidopsis*. Cell Host and Microbe **28**, 825-837.

Harr B, Schlötterer C. 2006. Comparison of algorithms for the analysis of *Affymetrix* microarray data as evaluated by co-expression of genes in known operons. Nucleic Acids Research **34**, e8.

Hasin Y, Seldin M, Lusis A. 2017. Multi-omics approaches to disease. Genome Biology **18**, 83.

Hay A, Barkoulas M, Tsiantis M. 2006. ASYMMETRIC LEAVES1 and auxin activities converge to repress *BREVIPEDICELLUS* expression and promote leaf development in *Arabidopsis*. Development **133**, 3955-3961.

He J, Benedito VA, Wang M, Murray JD, Zhao PX, Tang Y, Udvardi MK. 2009. The *Medicago truncatula* gene expression atlas web server. BMC Bioinformatics **10**, 441.

Heller MJ. 2002. DNA microarray technology: devices, systems, and applications. Annual Review of Biomedical Engineering **4**, 129-153.

Herlihy JH, Long TA, McDowell JM. 2020. Iron homeostasis and plant immune responses: recent insights and translational implications. Journal of Biological Chemistry **295**, 13444-13457.

Herman PL, Ramberg H, Baack RD, Markwell J, Osterman JC. 2002. Formate dehydrogenase in *Arabidopsis thaliana*: overexpression and subcellular localization in leaves. Plant Science **163**, 1137-1145.

Hindt MN, Guerinot ML. 2012. Getting a sense for signals: regulation of the plant iron deficiency response. Biochimica et Biophysica Acta - Molecular Cell Research **1823**, 1521-1530.

Hoheisel JD. 2006. Microarray technology: beyond transcript profiling and genotype analysis. Nature Reviews Genetics **7**, 200-210.

Hossain MM, Sultana F, Islam S. 2017. Plant growth-promoting fungi (PGPF): phytostimulation and induced systemic resistance. In: Singh PD, Singh HB, Prabha R (eds). Plant-microbe interactions in agro-ecological perspectives. Springer, 135-191.

Hourton-Cabassa C, Ambard-Bretteville F, Moreau F, de Virville JD, Rémy R, Colas Des Francs-Small C. 1998. Stress induction of mitochondrial formate dehydrogenase in potato leaves. Plant Physiology **116**, 627-635.

Hsiao PY, Cheng CP, Koh KW, Chan MT. 2017. The *Arabidopsis* defensin gene, *AtPDF1.1*, mediates defence against *Pectobacterium carotovorum* subsp. *carotovorum* via an iron-withholding defence system. Scientific Reports **7**, 9175.

Huala E, Dickerman AW, Garcia-Hernandez M, *et al.* 2001. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Research **29**, 102-105.

Huber W, Carey VJ, Gentleman R, *et al.* 2015. Orchestrating high-throughput genomic analysis with Bioconductor. Nature Methods **12**, 115-121.

Idjradinata P, Watkins WE, Pollitt E. 1994. Adverse effect of iron supplementation on weight gain of iron-replete young children. Lancet **343**, 1252-1254.

Igamberdiev AU, Bykova N v., Kleczkowski LA. 1999. Origins and metabolism of formate in higher plants. Plant Physiology and Biochemistry **37**, 503-513.

Igamberdiev AU, Eprintsev AT. 2016. Organic acids: the pools of fixed carbon involved in redox regulation and energy balance in higher plants. Frontiers in Plant Science **7**, 1042.

Igamberdiev AU, Kleczkowski LA. 2018. The glycerate and phosphorylated pathways of serine synthesis in plants: the branches of plant glycolysis linking carbon and nitrogen metabolism. Frontiers in Plant Science **9**, 318.

Ipek M, Esitken A. 2017. The Actions of PGPR on micronutrient availability in soil under calcareous soil conditions: an evaluation over Fe nutrition. In: Singh PD, Singh HB, Prabha R (eds). Plant-microbe interactions in agro-ecological perspectives. Springer, 81-100.

Iqbal N, Khan NA, Ferrante A, Trivellini A, Francini A, Khan MIR. 2017. Ethylene role in plant growth, development and senescence: interaction with other phytohormones. Frontiers in Plant Science **8**, 475.

Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. 2003a. Summaries of *Affymetrix* GeneChip probe level data. Nucleic acids research **31**, e15.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. 2003b. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4**, 249-264.

Jaksik R, Iwanaszko M, Rzeszowska-Wolny J, Kimmel M. 2015. Microarray experiments and factors which affect their reliability. Biology Direct **10**, 46.

Jamil F, Mukhtar H, Fouillaud M. 2022. Rhizosphere signaling: insights into plant – rhizomicrobiome interactions for sustainable agronomy. Microorganisms **10**, 1-26.

Jänsch L, Kruft V, Schmitz UK, Braun HP. 1996. New insights into the composition, molecular mass and stoichiometry of the protein complexes of plant mitochondria. Plant Journal **9**, 357-368.

Jefferson R. 1994. The hologenome. Agriculture, environment and the developing world: a future of PCR. New York, NY: Cold Spring Harbor, 1994.

Jin CW, He YF, Tang CX, Wu P, Zheng SJ. 2006. Mechanisms of microbially enhanced Fe acquisition in red clover (*Trifolium pratense* L.). Plant, Cell and Environment **29**, 888-897.

Jin CW, Chen WW, Meng Z bin, Zheng SJ. 2008. Iron deficiency-induced increase of root branching contributes to the enhanced root ferric chelate reductase activity. Journal of Integrative Plant Biology **50**, 1557-1562.

Jin CW, Li GX, Yu XH, Zheng SJ. 2010. Plant Fe status affects the composition of siderophore-secreting microbes in the rhizosphere. Annals of botany **105**, 835-841.

Jing H, Strader LC. 2019. Interplay of auxin and cytokinin in lateral root development. International Journal of Molecular Sciences **20**, 486.

Jollife IT, Cadima J. 2016. Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **374**, 20150202.

Jorge TF, Rodrigues JA, Caldana C, Schmidt R, van Dongen JT, Thomas-Oates J, Antonio C. 2016. Mass spectrometry-based plant metabolomics: metabolite responses to abiotic stress. Mass Spectrometry Reviews **35**, 620-649.

Kabir AH, Tahura S, Elseehy MM, El-Shehawi AM. 2021. Molecular characterization of Fe-acquisition genes causing decreased Fe uptake and photosynthetic inefficiency in Fe-deficient sunflower. Scientific Reports **11**, 5537.

Kang Y, Li M, Sinharoy S, Verdier J. 2016. A snapshot of functional genetic studies in *Medicago truncatula*. Frontiers in Plant Science **7**, 1175.

Kaul S, Koo HL, Jenkins J, *et al.* 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**, 796-815.

Kawamura E, Horiguchi G, Tsukaya H. 2010. Mechanisms of leaf tooth formation in *Arabidopsis*. Plant Journal **62**, 429-441.

Klessig DF, Choi HW, Dempsey DA. 2018. Systemic acquired resistance and salicylic acid: past, present, and future. Molecular Plant-Microbe Interactions **31**, 871-888.

Kloepper J, Schroth MN. 1978. Plant growth-promoting rhizobacteria on radishes. Fourth International Conference on Plant Pathogen Bacteria, Angers, France. 879-882.

Kong J, Dong Y, Xu L, Liu S, Bai X. 2014. Effects of foliar application of salicylic acid and nitric oxide in alleviating iron deficiency induced chlorosis of *Arachis hypogaea* L. Botanical Studies **55**, 9.

Korenblum E, Dong Y, Szymanski J, Panda S, Jozwiak A, Massalha H, Meir S, Rogachev I, Aharoni A. 2020. Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. Proceedings of the National Academy of Sciences of the United States of America **117**, 3874-3883.

Krämer U. 2015. Planting molecular functions in an ecological context with *Arabidopsis thaliana*. eLife **4**, e06100.

Krishnakumar V, Hanlon MR, Contrino S, *et al.* 2015. Araport: the *Arabidopsis* information portal. Nucleic Acids Research **43**, 1003-1009.

Kumari B, Mallick MA, Solanki MK. 2019. Plant growth promoting rhizobacteria (PGPR): modern prospects for sustainable agriculture. In: Ali Ansari R, Mahmood I (eds.). Plant health under biotic stress. Springer, 109-127.

Kurt-Gür G, Demirci H, Sunulu A, Ordu E. 2018. Stress response of NAD^+-dependent formate dehydrogenase in *Gossypium hirsutum* L. grown under copper toxicity. Environmental Science and Pollution Research **25**, 31679-31690.

Lamers JG, Schippers B, Geels FP. 1988. Soil-borne diseases of wheat in the Netherlands and results of seed bacterization with Pseudomonads against *Gaeumannomyces graminis* var. *tritici*, associated with disease resistance. In: Jorna ML, Slootmaker LAJ (eds.). Cereal breeding related to integrated cereal production. Wageningen Pudoc, 134-139.

Lebeis SL, Paredes SH, Lundberg DS, *et al.* 2015. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. Science **349**, 860-864.

Ledford H. 2008. The death of microarrays? Nature **455**, 847-847.

Lee S, Vemanna RS, Oh S, Rojas CM, Oh Y, Kaundal A, Kwon T, Lee HK, Senthil-Kumar M, Mysore KS. 2022. Functional role of formate dehydrogenase 1 (FDH1) for host and nonhost disease resistance against bacterial pathogens. PLOS ONE **17**, e0264917.

Leeman M, den Ouden FM, van Pelt JA, Dirkx FPM, Steijl H, Bakker PAHM, Schippers B. 1996. Iron availability affects induction of systemic resistance to *Fusarium* wilt of radish by *Pseudomonas fluorescens*. Phytopathology **86**, 149-155.

Leonelli S. 2007. Growing weed, producing knowledge an epistemic history of *Arabidopsis thaliana*. Current Reflections on Representation in Biology **29**, 193-223.

Leong J. 1986. Siderophores: their role in the biocontrol of plant pathogens. Annual Review of Phytopathology **24**, 187-209.

Lesins KA, Lesins I. 1979. Habitat and distribution. In: Lesins KA, Lesins I (eds.). Genus *Medicago* (Leguminosae). Springer, 31-35.

Li G, Kronzucker HJ, Shi W. 2016. The response of the root apex in plant adaptation to iron heterogeneity in soil. Frontiers in Plant Science **7**, 344.

Li R, Moore M, Bonham-Smith PC, King J. 2002. Overexpression of formate dehydrogenase in *Arabidopsis thaliana* resulted in plants tolerant to high concentrations of formate. Journal of Plant Physiology **159**, 1069-1076.

Li R, Moore M, King J. 2003. Investigating the regulation of one-carbon metabolism in *Arabidopsis thaliana*. Plant and Cell Physiology **44**, 233-241.

Li Y, Pearl SA, Jackson SA. 2015. Gene networks in plant biology: Approaches in reconstruction and analysis. Trends in Plant Science **20**, 664–675.

Liu J, Osbourn A, Ma P. 2015a. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. Molecular Plant **8**, 689-708.

Liu Y, Morley M, Brandimarto J, *et al.* 2015b. RNA-Seq identifies novel myocardial gene expression signatures of heart failure. Genomics **105**, 83–89.

Liu Y, Kong D, Wu H-L, Ling H-Q. 2021. Iron in plant–pathogen interactions. Journal of Experimental Botany **72**, 2114-2124.

Lockhart DJ, Dong H, Byrne MC, *et al.* 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotechnology **14**, 1675-1680.

Lodde V, Morandini P, Costa A, Murgia I, Ezquer I. 2021. cROStalk for life: uncovering ROS signaling in plants and animal systems, from gametogenesis to early embryonic development. Genes **12**, 525.

Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. 2017. Transcriptomics technologies. PLOS Computational Biology **13**, e1005457.

Lou HQ, Gong YL, Fan W, Xu JM, Liu Y, Cao MJ, Wang MH, Yang JL, Zheng SJ. 2016. A formate dehydrogenase confers tolerance to aluminum and low pH. Plant Physiology **171**, 294-305.

Lucena C, Romera FJ, García MJ, Alcántara E, Pérez-Vicente R. 2015. Ethylene participates in the regulation of Fe deficiency responses in strategy I plants and in rice. Frontiers in Plant Science **6**, 1056.

Lugtenberg B, Kamilova F. 2009. Plant-growth-promoting rhizobacteria. Annual Review of Microbiology **63**, 541-556.

Lynch SR. 2005. The impact of iron fortification on nutritional anaemia. Best Practice and Research: Clinical Haematology **18**, 333-346.

Lynch SV., Pedersen O. 2016. The Human intestinal microbiome in health and disease. New England Journal of Medicine **375**, 2369-2379.

Maia LB, Moura JJG, Moura I. 2015. Molybdenum and tungsten-dependent formate dehydrogenases. Journal of Biological Inorganic Chemistry **20**, 287-309.

Majeed A, Muhammad Z, Ahmad H. 2018. Plant growth promoting bacteria: role in soil improvement, abiotic and biotic stress management of crops. Plant Cell Reports **37**, 1599-1609.

Månsson R, Tsapogas P, Åkerlund M, Lagergren A, Gisler R, Sigvardsson M. 2004. Pearson correlation analysis of microarray data allows for the identification of genetic targets for early B-cell factor. Journal of Biological Chemistry **279**, 17905-17913.

Margulis L. 1991. Symbiogenesis and symbionticism. In: Margulis L, Ester R (eds.). Symbiosis as a source of evolution innovation speciation and morphogenesis. Speciation and morphogenesis. MIT Press, 1-14.

Marschner H, Römheld V. 1994. Strategies of plants for acquisition of iron. Plant and Soil **165**, 261–274.

Marschner P, Crowley DE, Sattelmacher B. 1997. Root colonization and iron nutritional status of a *Pseudomonas fluorescens* in different plant species. Plant and Soil **196**, 311-316.

Martínez-García PM, Ruano-Rosa D, Schilirò E, Prieto P, Ramos C, Rodríguez-Palenzuela P, Mercado-Blanco J. 2015. Complete genome sequence of *Pseudomonas fluorescens* strain PICF7, an indigenous root endophyte from olive (*Olea europaea* L.) and effective biocontrol agent against *Verticillium dahliae*. Standards in Genomic Sciences **10**, 10.

Martinez-Medina A, Flors V, Heil M, Mauch-Mani B, Pieterse CMJ, Pozo MJ, Ton J, van Dam NM, Conrath U. 2016. Recognizing plant defense priming. Trends in Plant Science **21**, 818-822.

Martínez-Medina A, Appels FVW, van Wees SCM. 2017. Impact of salicylic acid- and jasmonic acid-regulated defences on root colonization by *Trichoderma harzianum* T-78. Plant Signaling and Behavior **12**, e1345404.

Marzorati F, Vigani G, Morandini P, Murgia I. 2021a. Formate dehydrogenase contributes to the early *Arabidopsis thaliana* responses against *Xanthomonas campestris* pv *campestris infection*. Physiological and Molecular Plant Pathology **114**, 101633.

Marzorati F, Wang C, Pavesi G, Mizzi L, Morandini P. 2021b. Cleaning the *Medicago* microarray database to improve gene function analysis. Plants **10**, 1240.

Marzorati F, Midali A, Morandini P, Murgia I. 2022. Good or bad? The double face of iron in plants. Frontiers for Young Minds **10**, 718162.

Mathimaran N, Srivastava R, Wiemken A, Sharma AK, Boller T. 2012. Genome sequences of two plant growth-promoting fluorescent *Pseudomonas* strains, R62 and R81. Journal of Bacteriology **194**, 3272-3273.

Matilla MA, Ramos JL, Bakker PAHM, Doornbos R, Badri D v., Vivanco JM, Ramos-González MI. 2010. *Pseudomonas putida* KT2440 causes induced systemic resistance and changes in *Arabidopsis* root exudation. Environmental Microbiology Reports **2**, 381-388.

Mauch-Mani B, Baccelli I, Luna E, Flors V. 2017. Defense priming: an adaptive part of induced resistance. Annual Review of Plant Biology **68**, 485-512.

Maurer F, Müller S, Bauer P. 2011. Suppression of Fe deficiency gene expression by jasmonate. Plant Physiology and Biochemistry **49**, 530-536.

McGuire AL, Gabriel S, Tishkoff SA, *et al*. 2020. The road ahead in genetics and genomics. Nature Reviews Genetics **21**, 581-596.

Meijer J. 2021. Bio-informatics: a working concept. A translation of 'Bio-informatica: een werkconcept' by B. Hesper and P. Hogeweg.

Meiser J, Lingam S, Bauer P. 2011. Posttranslational regulation of the iron deficiency basic helix-loop-helix transcription factor FIT is affected by iron and nitric oxide. Plant Physiology **157**, 2154-2166.

Mercado-Blanco J, Rodríguez-Jurado D, Hervás A, Jiménez-Diaz RM. 2004. Suppression of Verticillium wilt in olive planting stocks by root-associated fluorescent *Pseudomonas* spp. Biological Control **30**, 474-486.

Meyerowitz EM. 1987. *Arabidopsis thaliana*. Annual review of genetics **21**, 93-111.

Miethke M, Marahiel MA. 2007. Siderophore-based iron acquisition and pathogen control. Microbiology and Molecular Biology Reviews **71**, 413-451.

Millenaar FF, Okyere J, May ST, van Zanten M, Voesenek LACJ, Peeters AJM. 2006. How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. BMC Bioinformatics **7**, 137.

Minorsky PV. 2019. On the inside. Plant Physiology **179**, 1431-1432.

Mizuno N, Takahashi A, Wagatsuma T, Mizuno T, Obata H. 2002. Chemical composition of guttation fluid and leaves of *Petasites japonicus* v. *giganteus* and *Polygonum cuspidatum* growing on ultramafic soil. Soil Science and Plant Nutrition **48**, 451-453.

Morissey J, Guerinot ML. 2009. Iron uptake and transport in plants: the good, the bad, and the ionome. Chemical Reviews **109**, 4553-4567.

Morris JJ. 2018. What is the hologenome concept of evolution? F1000Research **7**, F1000 Faculty Rev-1664.

Mossialos D, Meyer JM, Budzikiewicz H, Wolff U, Koedam N, Baysse C, Anjaiah V, Cornelis P. 2000. Quinolobactin, a new siderophore of *Pseudomonas fluorescens* ATCC 17400, the production of which is repressed by the cognate pyoverdine. Applied and Environmental Microbiology **66**, 487-492.

Müller DB, Vogel C, Bai Y, Vorholt JA. 2016. The plant microbiota: systems-level insights and perspectives. Annual Review of Genetics **50**, 211-234.

Murgia I, Delledonne M, Soave C. 2002. Nitric oxide mediates iron-induced ferritin accumulation in *Arabidopsis*. Plant Journal **30**, 521-528.

Murgia I, Arosio P, Tarantino D, Soave C. 2012. Biofortification for combating 'hidden hunger' for iron. Trends in Plant Science **17**, 47-55.

Murgia I, Vigani G, di Silvestre D, Mauri P, Rossi R, Bergamaschi A, Frisella M, Morandini P. 2020. Formate dehydrogenase takes part in molybdenum and iron homeostasis and affects dark-induced senescence in plants. Journal of Plant Interactions **15**, 386-397.

Murgia I, Marzorati F, Vigani G, Morandini P. 2022. Plant iron nutrition: the long road from soil to seeds. Journal of Experimental Botany **73**, 1809-1824.

Naidu CK, Suneetha Y. 2012. Current knowledge on microarray technology - an overview. Tropical Journal of Pharmaceutical Research **11**, 153-164.

Nalbantoglu S, Karadag A. 2019. Introductory chapter: insight into the OMICS technologies and molecular medicine. In Nalbantoglu S, Amri H. (eds). Molecular medicine. IntechOpen.

Nascimento FX, Rossi MJ, Glick BR. 2018. Ethylene and 1-aminocyclopropane-1-carboxylate (ACC) in plant–bacterial interactions. Frontiers in Plant Science **9**, 114.

Nel B, Steinberg C, Labuschagne N, Viljoen A. 2006. The potential of nonpathogenic *Fusarium oxysporum* and other biological control organisms for suppressing fusarium wilt of banana. Plant Pathology **55**, 217-223.

Nihorimbere V, Ongena M, Smargiassi M, Thonart P. 2011. Beneficial effect of the rhizosphere microbial community for plant growth and health. Biotechnology, Agronomy and Society and Environment **15**, 327-337.

Olson BJSC, Skavdahl M, Ramberg H, Osterman JC, Markwell J. 2000. Formate dehydrogenase in *Arabidopsis thaliana*: characterization and possible targeting to the chloroplast. Plant Science **159**, 205-212.

Olson NE. 2006. The microarray data analysis process: from raw data to biological significance. NeuroRx **3**, 373-383.

O'Mahony L. 2015. Host-microbiome interactions in health and disease. Clinical Liver Disease **5**, 142-144.

Osorio Vega NW. 2007. A review on beneficial effects of rhizosphere bacteria on soil nutrient availability and plant nutrient uptake. Revista Facultad Nacional de Agronomía Medellín **60**, 3621-3643.

Paasch BC, He SY. 2021. Toward understanding microbiota homeostasis in the plant kingdom. PLOS Pathogens **17**, e1009472.

Palmer CM, Hindt MN, Schmidt H, Clemens S, Guerinot M lou. 2013. MYB10 and MYB72 are required for growth under iron-limiting conditions. PLOS Genetics **9**, e1003953.

Pammel LH. 1895. Bacteriosis of rutabaga. Bulletin Iowa State University **3**, 130-135.

Pangesti N, Vandenbrande S, Pineda A, Dicke M, Raaijmakers JM, van Loon JJA. 2017. Antagonism between two root-associated beneficial *Pseudomonas* strains does not affect plant growth promotion and induced resistance against a leaf-chewing herbivore. FEMS Microbiology Ecology **93**, fix038.

Pascale A, Proietti S, Pantelides IS, Stringlis IA. 2020. Modulation of the root microbiome by plant molecules: the basis for targeted disease suppression and plant growth promotion. Frontiers in Plant Science **10**, 1741.

Pasha A, Shabari S, Cleary A, Chen X, Berardini T, Farmer A, Town C, Provart N. 2020. Araport lives: an updated framework for *Arabidopsis* bioinformatics. Plant Cell **32**, 2683-2686.

Penmetsa RV, Cook DR. 2000. Production and characterization of diverse developmental mutants of *Medicago truncatula*. Plant Physiology **123**, 1387-1397.

Picazo-Aragonés J, Terrab A, Balao F. 2020. Plant volatile organic compounds evolution: transcriptional regulation, epigenetics and polyploidy. International Journal of Molecular Sciences **21**, 8956.

Pieterse CMJ, Leon-Reyes A, van der Ent S, van Wees SCM. 2009. Networking by small-molecule hormones in plant immunity. Nature Chemical Biology **5**, 308-316.

Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, van Wees SCM, Bakker PAHM. 2014. Induced systemic resistance by beneficial microbes. Annual Review of Phytopathology **52**, 347-375.

Pieterse CMJ, Berendsen RL, de Jonge R, Stringlis IA, van Dijken AJH, van Pelt JA, van Wees SCM, Yu K, Zamioudis C, Bakker PAHM. 2020. *Pseudomonas simiae* WCS417: star track of a model beneficial rhizobacterium. Plant and Soil **461**, 245-263.

Pilati S, Malacarne G, Navarro-Payá D, Tomè G, Riscica L, Cavecchia V, Matus JT, Moser C, Blanzieri E. 2021. Vitis OneGenE: A causality-based approach to generate gene networks in *Vitis vinifera* sheds light on the laccase and dirigent gene families. Biomolecules **11**, biom11121744.

Pilot G, Stransky H, Bushey DF, Pratelli R, Ludewig U, Wingate VPM, Frommer WB. 2004. Overexpression of GLUTAMINE DUMPER1 leads to hypersecretion of glutamine from hydathodes of *Arabidopsis* leaves. Plant Cell **16**, 1827-1840.

Popov V, Lamzin V. 1994. NAD+-dependent formate dehydrogenase. The Biochemical Journal **301**, 625-643.

Prashar P, Kapoor N, Sachdeva S. 2014. Rhizosphere: its structure, bacterial diversity and significance. Reviews in Environmental Science and Biotechnology **13**, 63-77.

Provart NJ, Alonso J, Assmann SM, *et al.* 2016. 50 years of *Arabidopsis* research: highlights and future directions. New Phytologist **209**, 921-944.

Raaijmakers JM, Paulitz TC, Steinberg C, Alabouvette C, Moënne-Loccoz Y. 2009. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. Plant and Soil **321**, 341-361.

Rabilloud T. 2014. How to use 2D gel electrophoresis in plant proteomics. In: Clifton NJ (eds). Methods in molecular biology. Springer, 43-50.

Ran LX, Liu CY, Wu GJ, van Loon LC, Bakker PAHM. 2005. Suppression of bacterial wilt in *Eucalyptus urophylla* by fluorescent *Pseudomonas* spp. in China. Biological Control **32**, 111-120.

Ravanbakhsh M, Sasidharan R, Voesenek LACJ, Kowalchuk GA, Jousset A. 2018. Microbial modulation of plant ethylene signaling: ecological and evolutionary consequences. Microbiome **6**, 52.

Redestig H, Costa IG. 2011. Detection and interpretation of metabolite–transcript coresponses using combined profiling data. Bioinformatics **27**, i357-65.

Ren C, Bilyeu KD, Beuselinck PR. 2009. Composition, vigor, and proteome of mature soybean seeds developed under high temperature. Crop Science **49**, 1010-1022.

Reyes ALP, Silva TC, Coetzee SG, *et al.* 2019. GENAVi: a shiny web application for gene expression normalization, analysis and visualization. BMC Genomics **20**, 745.

Riaz N, Guerinot ML. 2021. All together now: regulation of the iron deficiency response. Journal of Experimental Botany **72**, 2045-2055.

Ringel MT, Brüser T. 2018. The biosynthesis of pyoverdines. Microbial Cell **5**, 424-437.

Robe K, Izquierdo E, Vignols F, Rouached H, Dubos C. 2021. The coumarins: secondary metabolites playing a primary role in plant nutrition and health. Trends in Plant Science **26**, 248-259.

Rodríguez-Celma J, Chun Pan I, Li W, Lan P, Buckhout TJ, Schmidt W. 2013. The transcriptional response of *Arabidopsis* leaves to Fe deficiency. Frontiers in Plant Science **4**, 276.

Rogers C, Wen J, Chen R, Oldroyd G. 2009. Deletion-based reverse genetics in *Medicago truncatula*. Plant Physiology **151**, 1077-1086.

Rohwer F, Seguritan V, Azam F, Knowlton N. 2002. Diversity and distribution of coral-associated bacteria. Marine Ecology Progress Series **243**, 1-10.

Romanov V, Davidoff SN, Miles AR, Grainger DW, Gale BK, Brooks BD. 2014. A critical comparison of protein microarray fabrication technologies. Analyst **139**, 1303-1326.

Romera FJ, García MJ, Alcántara E, Pérez-Vicente R. 2011. Latest findings about the interplay of auxin, ethylene and nitric oxide in the regulation of Fe deficiency responses by Strategy I plants. Plant Signaling & Behavior **6**, 167-170.

Romera FJ, García MJ, Lucena C, Martínez-Medina A, Aparicio MA, Ramos J, Alcántara E, Angulo M, Pérez-Vicente R. 2019. Induced systemic resistance (ISR) and Fe deficiency responses in dicot plants. Frontiers in Plant Science **10**, 287.

Rout GR, Sahoo S. 2015. Role of iron in plant growth and metabolism. Reviews in Agricultural Science **3**, 1-24.

Ryals JA, Neuenschwander UH, Willits MG, Molina A, Steiner HY, Hunt MD. 1996. Systemic acquired resistance. Plant Cell **8**, 1809-1819.

Sánchez-Cañizares C, Jorrín B, Poole PS, Tkacz A. 2017. Understanding the holobiont: the interdependence of plants and their microbiome. Current Opinion in Microbiology **38**, 188-196.

Sasse J, Martinoia E, Northen T. 2018. Feed your friends: do plant exudates shape the root microbiome? Trends in Plant Science **23**, 25-41.

Sazawal S, Black RE, Ramsan M, Chwaya HM, Stoltzfus RJ, Dutta A, Dhingra U, Kabole I, Deb S, Othman MK. 2006. Effects of routine prophylactic supplementation with iron and folic acid on admission to hospital and mortality in preschool children in a high malaria transmission setting: community-based, randomised, placebo-controlled trial. Lancet **367**, 133-143.

Scarpella E, Marcos D, Friml J, Berleth T. 2006. Control of leaf vascular patterning by polar auxin transport. Genes and Development **20**, 1015-1027.

Schmidt W, Tittel J, Schikora A. 2000. Role of hormones in the induction of iron deficiency responses in *Arabidopsis* roots. Plant Physiology **122**, 1109-1118.

Schmidt H, Günther C, Weber M, Spörlein C, Loscher S, Böttcher C, Schobert R, Clemens S. 2014. Metabolome analysis of *Arabidopsis thaliana* roots identifies a key metabolic pathway for iron acquisition. PLOS ONE **9**, e102444.

Schulz-Bohm K, Martín-Sánchez L, Garbeva P. 2017. Microbial volatiles: small molecules with an important role in intra- and inter-kingdom interactions. Frontiers in Microbiology **8**, 2484.

Segarra G, van der Ent S, Trillas I, Pieterse CMJ. 2009. MYB72, a node of convergence in induced systemic resistance triggered by a fungal and a bacterial beneficial microbe. Plant Biology **11**, 90-96.

Séguéla M, Briat JF, Vert G, Curie C. 2008. Cytokinins negatively regulate the root iron uptake machinery in *Arabidopsis* through a growth-dependent pathway. Plant Journal **55**, 289-300.

Sharma M, Sudheer S, Usmani Z, Rani R, Gupta P. 2020. Deciphering the omics of plant-microbe interaction: perspectives and new insights. Current Genomics **21**, 343-362.

Sheldrake AR, Northcote DH. 1968. Some constituents of xylem sap and their possible relationship to xylem differentiation. Journal of Experimental Botany **19**, 681-689.

Shen C, Yang Y, Liu K, Zhang L, Guo H, Sun T, Wang H. 2016. Involvement of endogenous salicylic acid in iron-deficiency responses in *Arabidopsis*. Journal of Experimental Botany **67**, 4179-4193.

Shendure J. 2008. The beginning of the end for microarrays? Nature Methods **5**, 585-587.

Simon JC, Marchesi JR, Mougel C, Selosse MA. 2019. Host-microbiota interactions: from holobiont theory to analysis. Microbiome **7**, 5.

Skillings D. 2016. Holobionts and the ecology of organisms: multi-species communities or integrated individuals? Biology and Philosophy **33**, 875-892.

Slonim DK, Yanai I. 2009. Getting started in gene expression microarray analysis. PLOS Computational Biology **5**, e1000543.

Smith SE, Read D. 2008. The symbionts forming arbuscular mycorrhizas. In: Smith SE, Read D (eds.). Mycorrhizal Symbiosis. 3rd Edition. Academic Press, 13-41.

Song ZB, Xiao SQ, You L, Wang SS, Tan H, Li KZ, Chen LM. 2013. C1 metabolism and the Calvin cycle function simultaneously and independently during HCHO metabolism and detoxification in *Arabidopsis thaliana* treated with HCHO solutions. Plant, Cell and Environment **36**, 1490-1506.

Soyano T, Hirakawa H, Sato S, Hayashi M, Kawaguchi M. 2014. NODULE INCEPTION creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. Proceedings of the National Academy of Sciences of the United States of America **111**, 14607-14612.

Spence J, Vercher Y, Gates P, Harris N. 1996. 'Pod shatter' in *Arabidopsis thaliana*, *Brassica napus* and *B. juncea*. Journal of Microscopy **181**, 195-203.

Spielmann J, Vert G. 2021. The many facets of protein ubiquitination and degradation in plant root iron-deficiency responses. Journal of Experimental Botany **72**, 2071-2082.

Stafford JL. 1961. Iron deficiency in man and animals. Journal of the Royal Society of Medicine **54**, 1000-1004.

Stark R, Grzelak M, Hadfield J. 2019. RNA sequencing: the teenage years. Nature Reviews Genetics **20**, 631–656.

Stringlis IA, Proietti S, Hickman R, van Verk MC, Zamioudis C, Pieterse CMJ. 2018a. Root transcriptional dynamics induced by beneficial rhizobacteria and microbial immune elicitors reveal signatures of adaptation to mutualists. Plant Journal **93**, 166-180.

Stringlis IA, Yu K, Feussner K, de Jonge R, van Bentum S, van Verk MC, Berendsen RL, Bakker PAHM, Feussner I, Pieterse CMJ. 2018b. MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. Proceedings of the National Academy of Sciences of the United States of America **115**, 5213-5222.

Stringlis IA, de Jonge R, Pieterse CMJ. 2019. The age of coumarins in plant-microbe interactions. Plant and Cell Physiology **60**, 1405-1419.

Stutz EW, Défago G, Kern H. 1986. Naturally occurring fluorescent Pseudomonads involved in suppression of black root rot of tobacco. Phytopathology **76**, 181-185.

Subramanian I, Verma S, Kumar S, Jere A, Anamika K. 2020. Multi-omics data integration, interpretation, and its application. Bioinformatics and Biology Insights **14**, 1177932219899051.

Suh E, Choi SW, Friso S. 2016. One-carbon metabolism: an unsung hero for healthy aging. In: Malavolta M, Mocchegiani E (eds.). Molecular basis of nutrition and aging: a volume in the molecular nutrition series. Academic Press, 513-522.

Sun H, Jiang S, Jiang C, Wu C, Gao M, Wang Q. 2021. A review of root exudates and rhizosphere microbiome for crop production. Environmental Science and Pollution Research **28**, 54497-54510.

Sutton JC, Williams PH. 1970. Relation of xylem plugging to black rot lesion development in cabbage. Canadian Journal of Botany **48**, 391-401.

Suzuki K, Itai R, Suzuki K, Nakanishi H, Nishizawa NK, Yoshimura E, Mori S. 1998. Formate dehydrogenase, an enzyme of anaerobic metabolism, is induced by iron deficiency in barley roots. Plant Physiology **116**, 725-732.

Tang H, Krishnakumar V, Bidwell S, *et al.* 2014. An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. BMC Genomics **15**, 312.

Tavassoly I, Goldfarb J, Iyengar R. 2018. Systems biology primer: the basic methods and approaches. Essays in Biochemistry **62**, 487-500.

Tohidfar M, Khosravi S. 2015. Transgenic crops with an improved resistance to biotic stresses. A review. Biotechnology, Agronomy and Society and Environment **19**, 62–70.

Torsvik V, Øvreås L. 2002. Microbial diversity and function in soil: from genes to ecosystems. Current Opinion in Microbiology **5**, 240-245.

Trapet PL, Avoscan L, Klinguer A, *et al.* 2016. The *Pseudomonas fluorescens* siderophore pyoverdine weakens *Arabidopsis thaliana* defense in favor of growth in iron-deficient conditions. Plant Physiology **171**, 675-693.

Trapet PL, Verbon EH, Bosma RR, Voordendag K, van Pelt JA, Pieterse CMJ. 2021. Mechanisms underlying iron deficiency-induced resistance against pathogens with different lifestyles. Journal of Experimental Botany **72**, 2231-2241.

Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. 2020. Plant–microbiome interactions: from community assembly to plant health. Nature Reviews Microbiology **18**, 607-621.

Tsai TM, Huang HJ. 2006. Effects of iron excess on cell viability and mitogen-activated protein kinase activation in rice roots. Physiologia Plantarum **127**, 583-592.

Tzin V, Snyder JH, Yang DS, *et al*. 2019. Integrated metabolomics identifies CYP72A67 and CYP72A68 oxidases in the biosynthesis of *Medicago truncatula* oleanate sapogenins. Metabolomics **15**, 85.

Vacheron J, Desbrosses G, Bouffaud ML, Touraine B, Moënne-Loccoz Y, Muller D, Legendre L, Wisniewski-Dyé F, Prigent-Combaret C. 2013. Plant growth-promoting rhizobacteria and root system functioning. Frontiers in Plant Science **4**, 356.

van de Mortel JE, de Vos RCH, Dekkers E, Pineda A, Guillod L, Bouwmeester K, van Loon JJA, Dicke M, Raaijmakers JM. 2012. Metabolic and transcriptomic changes induced in *Arabidopsis* by the rhizobacterium *Pseudomonas fluorescens* SS101. Plant Physiology **160**, 2173-2188.

van der Ent S, Verhagen BWM, van Doorn R, *et al.* 2008. MYB72 is required in early signaling steps of rhizobacteria-induced systemic resistance in *Arabidopsis*. Plant Physiology **146**, 1293-1304.

van Loon LC, Bakker PAHM, Pieterse CMJ. 1998. Systemic resistance induced by rhizosphere bacteria. Annual Review of Phytopathology **36**, 453-483.

van Loon LC. 2007. Plant responses to plant growth-promoting rhizobacteria. European Journal of Plant Pathology **119**, 243-254.

Vandenkoornhuyse P, Quaiser A, Duhamel M, le Van A, Dufresne A. 2015. The importance of the microbiome of the plant holobiont. New Phytologist **206**, 1196-1206.

Verbon EH, Liberman LM. 2016. Beneficial microbes affect endogenous mechanisms controlling root development. Trends in Plant Science **21**, 218-229.

Verbon EH, Trapet PL, Kruijs S, Temple-Boyer-Dury C, Rouwenhorst TG, Pieterse CMJ. 2019. Rhizobacteria-mediated activation of the Fe deficiency response in *Arabidopsis* roots: impact on Fe status and signaling. Frontiers in Plant Science **10**, 909.

Verbon EH, Trapet PL, Stringlis IA, Kruijs S, Bakker PAHM, Pieterse CMJ. 2017. Iron and immunity. Annual Review of Phytopathology **55**, 355-375.

Veresoglou SD, Menexes G. 2010. Impact of inoculation with *Azospirillum* spp. on growth properties and seed yield of wheat: a meta-analysis of studies in the ISI Web of Science from 1981 to 2008. Plant and Soil **337**, 469-480.

Verhagen BWM, Glazebrook J, Zhu T, Chang HS, van Loon LC, Pieterse CMJ. 2004. The transcriptome of rhizobacteria-induced systemic resistance in *Arabidopsis*. Molecular Plant-Microbe Interactions **17**, 895-908.

Vigani G, Di Silvestre D, Agresta AM, Donnini S, Mauri P, Gehl C, Bittner F, Murgia I. 2017. Molybdenum and iron mutually impact their homeostasis in cucumber (*Cucumis sativus*) plants. New Phytologist **213**, 1222-1241.

Vlot AC, Sales JH, Lenk M, Bauer K, Brambilla A, Sommer A, Chen Y, Wenig M, Nayem S. 2021. Systemic propagation of immunity in plants. New Phytologist **229**, 1234-1250.

Voges MJEEE, Bai Y, Schulze-Lefert P, Sattely ES. 2019. Plant-derived coumarins shape the composition of an *Arabidopsis* synthetic root microbiome. Proceedings of the National Academy of Sciences of the United States of America **116**, 12558-12565.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics **10**, 57–63.

Wang W, Xu B, Wang H, Li J, Huang H, Xu L. 2011. YUCCA genes are expressed in response to leaf adaxial-abaxial juxtaposition and are required for leaf margin development. Plant Physiology **157**, 1805-1819.

Wang C, Gong B, Bushel PR, *et al.* 2014. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nature Biotechnology **32**, 926–932.

Watson BS, Asirvatham VS, Wang L, Sumner LW. 2003. Mapping the proteome of barrel medic (*Medicago truncatula*). Plant Physiology **131**, 1104-1123.

White PJ, White PJ, Broadley MR. 2009. Biofortification of crops with seven mineral elements often lacking in human diets - iron, zinc, copper, calcium, magnesium, selenium and iodine. New Phytologist **182**, 49-84.

Wickett NJ, Mirarab S, Nguyen N, *et al.* 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proceedings of the National Academy of Sciences of the United States of America **111**, 4859-4868.

Wild M, Davière JM, Regnault T, Sakvarelidze-Achard L, Carrera E, Lopez Diaz I, Cayrel A, Dubeaux G, Vert G, Achard P. 2016. Tissue-specific regulation of gibberellin signaling fine-tunes *Arabidopsis* iron-deficiency responses. Developmental Cell **37**, 190-200.

Williams PH. 1980. Black rot: a continuing threat to world crucifers. Plant Disease **64**, 736-742.

Wingler A, Lea PJ, Leegood RC. 1999. Photorespiratory metabolism of glyoxylate and formate in glycine-accumulating mutants of barley and *Amaranthus edulis*. Planta **207**, 518-526.

Wintermans PCA, Bakker PAHM, Pieterse CMJ. 2016. Natural genetic variation in *Arabidopsis* for responsiveness to plant growth-promoting rhizobacteria. Plant Molecular Biology **90**, 623-634.

Xu G, Yang S, Meng L, Wang BG. 2018. The plant hormone abscisic acid regulates the growth and metabolism of endophytic fungus *Aspergillus nidulans*. Scientific Reports **8**, 6504.

Yagi H, Nagano AJ, Kim J, Tamura K, Mochizuki N, Nagatani A, Matsushita T, Shimada T. 2021. Fluorescent protein-based imaging and tissue-specific RNA-Seq analysis of *Arabidopsis* hydathodes. Journal of Experimental Botany **72**, 1260-1270.

Young ND, Udvardi M. 2009. Translating *Medicago truncatula* genomics to crop legumes. Current Opinion in Plant Biology **12**, 193-201.

Young ND, Debellé F, Oldroyd GED, *et al*. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature **480**, 520-524.

Yu K, Stringlis IA, van Bentum S, de Jonge R, Snoek BL, Pieterse CMJ, Bakker PAHM, Berendsen RL. 2021. Transcriptome signatures in *Pseudomonas simiae* WCS417 shed light on role of root-secreted coumarins in *Arabidopsis*-mutualist communication. Microorganisms **9**, 575.

Zamioudis C, Pieterse CMJ. 2012. Modulation of host immunity by beneficial microbes. Molecular Plant-Microbe Interactions **25**, 139-150.

Zamioudis C, Mastranesti P, Dhonukshe P, Blilou I, Pieterse CMJ. 2013. Unraveling root developmental programs initiated by beneficial *Pseudomonas* spp. bacteria. Plant Physiology **162**, 304-318.

Zamioudis C, Hanson J, Pieterse CMJ. 2014. ß-Glucosidase BGLU42 is a MYB72-dependent key regulator of rhizobacteria-induced systemic resistance and modulates iron deficiency responses in *Arabidopsis* roots. New Phytologist **204**, 368-379.

Zamioudis C, Korteland J, van Pelt JA, *et al.* 2015. Rhizobacterial volatiles and photosynthesis-related signals coordinate MYB72 expression in *Arabidopsis* roots during onset of induced systemic resistance and iron-deficiency responses. Plant Journal **84**, 309-322.

Zancarini A, Westerhuis JA, Smilde AK, Bouwmeester HJ. 2021. Integration of omics data to unravel root microbiome recruitment. Current Opinion in Biotechnology **70**, 255-261.

Zargar SM, Kurata, R, Inaba S, Fukao Y. 2013. Unraveling the iron deficiency responsive proteome in Arabidopsis shoot by iTRAQ-OFFGEL approach. Plant Signaling and Behavior. **8**, e26892.

Zermiani M, Begheldo M, Nonis A, Palme K, Mizzi L, Morandini P, Nonis A, Ruperti B. 2015. Identification of the *Arabidopsis* RAM/MOR signalling network: adding new regulatory players in plant stem cell maintenance and cell polarization. Annals of Botany **116**, 69-89.

Zhang H, Sun Y, Xie X, Kim MS, Dowd SE, Paré PW. 2009. A soil bacterium regulates plant acquisition of iron via deficiency-inducible mechanisms. Plant Journal **58**, 568-577.

Zhang H, Wang S. 2013. Rice versus *Xanthomonas oryzae* pv. *oryzae*: a unique pathosystem. Current Opinion in Plant Biology **16**, 188-195.

Zhang H. 2019. The review of transcriptome sequencing: principles, history and advances. IOP Conference Series: Earth and Environmental Science **332**, 042003.

Zheng W, Chung LM, Zhao H. 2011. Bias detection and correction in RNA-sequencing data. BMC Bioinformatics **12**, 290.

Zhu C, Naqvi S., Gomez-Galera S, Pelacho AM, Capell T, Christou P. 2007. Transgenic strategies for the nutritional enhancement of plants. Trends in Plant Science **12**, 548-555.

Zilber-Rosenberg I, Rosenberg E. 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiology Reviews **32**, 723-735.

6. Acknowledgements

Breve premessa: non credo sarò in grado di mettere nero su bianco tutta la mia gratitudine per chi mi ha supportato (o forse dovrei dire *sopportato*?) in questi tre anni. Ci proverò, ma non sono sicura il risultato sarà tale da far capire quanto il loro aiuto e sostegno sia stato per me importante. Spero apprezzino comunque le mie parole, e credetemi quando dico che, senza di loro, questo dottorato non sarebbe mai giunto al termine.

First, I would like to acknowledge the professors and researchers who helped me with my Ph.D. work by sending materials, running tests, or advising me on my work. Prof. Laurent Nöel, Dr. Emmanuelle Lauber, and all their collaborators (LIPME - INRAE Toulouse) for providing the *Xcc* 8004 and *Xcc* 8004::GUS-GFP strains, for their insightful suggestions, and the *in vivo* test on wt Col and *atfdh1-5* plants. Prof. Peter Bakker and colleagues from the University of Utrecht for donating the *Pseudomonas simiae* WCS417r strain. Prof. Jian Li Yang (Zhejiang University) for donating to my research group the *A. thaliana* line *V. umbellata* FDH::GUS. The Salk institute for providing the *atfdh1-5* line Salk-108751. Dr. Dario Di Silvestre, Dr. Letizia Bernardo, Dr. Rossana Rossi, and Dr. Pierluigi Mauri (CNR-ITB Segrate Milano) for the proteomic analysis and data interpretation. Prof. Gianpiero Vigani (Università degli Studi di Torino) for his continuous collaboration. Dr. Stefania Prati (Università degli Studi di Milano) for her help in plant management in the greenhouse. Prof. Alex Costa (Università degli Studi di Milano) and Prof. Giulio Pavesi (Università degli Studi di Milano) for evaluating my work and advising me on how to proceed each year.

Devo sicuramente ringraziare il Prof. Piero Morandini e la Dott.ssa Irene Murgia: grazie per la vostra continua supervisione e per avermi insegnato il vostro approccio critico alla scienza. È grazie a voi che ho capito che non è sufficiente avere un'idea ma è forse più importante essere in grado di rispondere a una domanda con semplici e logici passaggi. È stato un onore avere avuto entrambi come professori nel corso dei miei studi e, soprattutto, è stato un onore poter lavorare con voi in questi tre anni. Piero e Irene, grazie per avermi mostrato cosa sia la vera passione per la ricerca e per avermi dato i mezzi e le possibilità di coltivarla.

Grazie al Dr. Luca Mizzi, che ha sempre trovato il tempo per aiutarmi con i miei (tanti) dubbi informatici.

Ringrazio gli studenti che sono stati in lab in questi tre anni, così come i nuovi e i vecchi amici, con cui ho potuto ridere e scherzare anche quando c'era solo da piangere.

Grazie ai miei genitori e a mio fratello, perché questo titolo sarà sicuramente più loro che mio. Ci sono stati tanti momenti difficili durante questo dottorato ma siete sempre stati al mio fianco. Ciò che mi insegnate va ben oltre i traguardi dei miei studi: la pazienza e il rialzarsi anche dopo una 'caduta' sono solo due dei vostri tanti insegnamenti che proverò a mettere in pratica. Grazie per credere in me anche quando io non ci riesco. Grazie per ogni vostro aiuto. Grazie per avermi reso Francesca.

Grazie a mia nonna Liliana: se sono quella che sono è anche grazie a te. Sei una donna forte, sei meravigliosa, non dimenticarlo mai.

Grazie ad Alessandro e il suo continuo, totale supporto: non cambierei proprio niente di noi. Non cambierei nulla del nostro percorso assieme, di ciò che sei e di ciò che mi dai. Grazie per essere cresciuto con me in questi anni.

Grazie a Sole e Luna, che hanno aiutato e aiutano a calmare la tempesta.

Vorrei infine dedicare un ultimo pensiero alla mia compagna non voluta di vita. Mi hai rubato mesi di energia, di sonno; sei comparsa a metà di questo dottorato quasi come se tutto ti fosse dovuto, arrogantemente. Conoscerti è stato difficile, accettarti lo sarà sicuramente di più, e non so nemmeno se ci riuscirò. Nessuno può davvero capire fino in fondo il nostro rapporto, forse nemmeno chi ti conosce direttamente: come scrisse De Andrè "per tutti il dolore degli altri è dolore a metà". Gestirti richiede tanta, anche troppa, pazienza; bisogna prenderti come vieni. A volte non ricordo nemmeno com'è non sentirsi stanchi, mi hai tolto anche la voglia di fare le cose che mi piacciono, perché la fatica e i sintomi mi hanno portato via anche un po' l'entusiasmo per le cose belle. Ma gestirti richiede comunque tanta pazienza e amore, sia per sé stessi e, ho scoperto, anche per gli altri. Non sei un dono e non sei una punizione; sei qualcosa che capita, con cui, che si voglia o meno, si deve imparare a convivere. Anche se continuerò a chiedermi il perché di tutto questo, io, da inguaribile pessimista che sono, prometto che cercherò di vedere sempre il lato positivo nelle cose: "bisogna guardare sempre il bicchiere mezzo pieno, e poi… poi bisogna berlo".