Institute of Plant Genetics of the Polish Academy of Sciences

Department of Integrative Plant Biology

Plant Systems Biology Team

Juan Camilo Ochoa Cabezas

Elucidation of genetic factors determining resistance or susceptibility to clubroot disease through genome wide association study, transcriptome profiling and functional genetics in Arabidopsis natural accessions

Doctoral Dissertation

Field of Agricultural Sciences

Discipline of Agriculture and Horticulture

Supervisor: dr hab. Robert Malinowski, prof. IGR PAN
Auxiliary supervisor: Dr. William Truman

POZNAŃ 2021

Acknowledgements

The elaboration of this manuscript and all the experimental work behind it would not have been possible without the constant help and support of Bill Truman. Bill was always a mentor that offered constructive advice, patience, and encouragement to always do a better job. I am proud of learning from him to do experiments with the most critical eye and the highest standards to completely trust the results obtained.

To dr. hab. Robert Malinowski for his constant support and his eagerness to solve all the difficulties and inconveniences presented during the elaboration of this work. I also want to thank him for the enthusiasm and aid in the elaboration of microscopic sections, and discussions about plant developmental biology.

To professor Małgorzata Jędryczka for providing the *P. brassicae* isolates and for her welcoming attitude and interest in showing Poland and the polish culture.

To dr. hab. Piotr Ziołkowski and dr. Tomasz Bieluszewski for their assistance with the elaboration of null mutants with CRISPR/Cas9.

To Soham Mukhopadhyay for his cooperation with the GWAS and RT-qPCR experiments.

To Piotr Gregorczyk for his unconditional support during my stay in Poland

Statement of the Ph.D. supervisor and auxiliary supervisor

We declare that this work has been performed under our direction, and state that it meets the conditions for presenting it in the procedure for granting the degree of Doctor of Agricultural Sciences in the discipline of agriculture and horticulture.

Date and signature of the supervisor

Date and signature of the auxiliary supervisor

Statement of the author

Aware of the legal responsibility, I declare that this doctoral thesis was written by me under the direction of the supervisor and auxiliary supervisor and does not contain information obtained in a manner inconsistent with applicable regulations. I also declare that the presented manuscript has not previously been submitted to obtain a scientific degree in any other institution and is identical to the attached electronic version.

Date and signature of the author

Funding

This research was funded by the National Science Centre of Poland, projects OPUS 10, UMO-2015/19/B/NZ3/01489 "*Understanding the role of chitin related defense responses during Plasmodiophora brassicae infection*" and PRELUDIUM 17, UMO-2019/33/N/NZ9/01048 "*Using CRISPR/Cas9 genome editing to characterize candidate Clubroot resistance genes in Arabidopsis thaliana*."

The author was supported by funds from the projects mentioned above, with a student stipend from the Institute of Plant Genetics of the Polish Academy of Sciences, and by the Colombian Ministry of Sciences (MINCIENCIAS) Convocatoria 860.

Title

“Elucidation of genetic factors determining resistance or susceptibility to clubroot disease through genome wide association study, transcriptome profiling and functional genetics in Arabidopsis natural accessions”.

Abstract

Clubroot disease, caused by the obligate biotroph *P. brassicae* has become one of the most important limitations to *Brassica* crop production; the use of resistant cultivars is the most advantageous way to manage this problem. As the prevalence of pathogen races that can overcome the resistance of currently available cultivars is expanding, the elucidation and characterization of durable mechanism for immunity is a priority.

In our study a collection of 142 Arabidopsis accessions was used to identify genetic factors responsible for the resistance or susceptibility to a Polish *P. brassicae* pathotype predominant countrywide. Through a genome wide association study (GWAS) two loci associated with resistance with a high degree of significance were identified, one of these regions contained *RPB1* (*Resistance to Plasmodiophora brassicae 1*) and *RPB1* homologs in the resistant accessions Uod-1 and Est-1; the second locus was in the coding sequence of *RAC1 (Resistance to Albugo candida 1)*, that codes for a TIR-NB-LRR protein. Through the generation of loss of function mutants in resistant accessions created with the CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-Associated protein 9) technique, a fundamental role for *RPB1* in resistance to clubroot disease was demonstrated, as the *rpb1* lines presented symptoms and pathogen DNA quantification comparable to the highly susceptible accession Col-0. *RPB1* codes for a small, putative transmembrane protein (148 aa, 16.0 kDa) with no known function or homology to characterized genes. The coding sequence for *RPB1* is present in several of the *P. brassicae* susceptible accessions and exhibits a high degree of conservation among the accessions that contain it, both resistant and susceptible. However, there is substantial variation in the sequence of the upstream promoter region, and this may explain the resistance phenotype. *RPB1* is strongly upregulated in the resistant accessions following inoculation with *P. brassicae,* and it is required for the induction of defense genes such as *PR5* (*Pathogenesis Related 5*) and *CYP71A13 (Cytochrome P450 Monooxygenase 71A13)*, moreover transient expression in *N. tabacum* leaves triggers a hypersensitive reaction, which supports the hypothesis that RPB1 is a positive regulator of defense responses. The

rpb1 deletion lines were still able to respond to *P. brassicae* infection by activating transcription of the non-functional, truncated *RPB1* gene but failed to induce other defense responses, indicating that RPB1 may not be involved in the pathogen recognition but is, however, crucial for immunity.

In the screen of Arabidopsis accessions, one line, Pro-0, was prominent because it displays comparatively less severe clubroot disease symptoms but has abnormally high pathogen DNA levels. We followed up on this finding and characterized the differences observed between the susceptible Arabidopsis accessions Col-0 and Pro-0 through histological observation and transcriptomic analysis. In the hypocotyls of healthy plants, it was observed that Pro-0 contains a higher proportion of xylem relative to the hypocotyl area, and in the sections of inoculated plants the xylem anatomy appeared less disrupted compared with Col-0. With comparative transcriptomic analysis of the response to *P. brassicae* infection, in Pro-0 we observed signs of delayed progression in the pathogen driven reprogramming of host developmental processes compared with Col-0. In the dynamics of clubroot gall development, there is a proliferative phase where host cell division is stimulated, followed by an expansive phase where hypertrophy of colonized cells occurs. Based on transcriptional changes in genes regulating cell cycle progress, cell growth and vascular patterning we observed that, at the time point profiled (19 dpi), Pro-0 galls retained the transcriptome signature of the proliferative state, while Col-0 had already entered the expansive phase. Based on these results, we hypothesize that variations in host growth and development, particularly vascular development may have a strong influence on clubroot disease progression. Further understanding of this phenomenon will require detailed dissection of the role of key host genes involved in regulation of host developmental responses and how they are targeted by the pathogen; characterizing these events at the molecular level could provide strategies for developing crops that are less severely affected by *P. brassicae*.

In summary, by exploiting the natural diversity in Arabidopsis we were able to identify factors essential for resistance to clubroot disease in Arabidopsis and propose possible mechanisms explaining differences in disease progression in susceptible accessions.

Tytuł

„Zastosowanie badań asocjacyjnych całego genomu, metod profilowania transkryptomu oraz genetyki funkcjonalnej do wyjaśnienia udziału poszczególnych genów w kształtowaniu cech odporności bądź wrażliwości na kiłę kapusty".

Streszczenie

Kiła kapusty wywołana przez obligatoryjnego biotrofa *Plasmodiophora brassicae* stanowi jeden z bardziej poważnych czynników powodujących utratę plonu roślin kapustowatych. Najbardziej korzystną metodą zapobiegania temu zjawisku jest wykorzystanie odmian odpornych. Ze względu na coraz bardziej powszechne występowanie ras patogenu, mogących przełamać odporności genetyczne wykorzystywane w aktualnie uprawianych odmianach, poznanie oraz dalsza charakterystyka stabilnych mechanizmów odporności jest zadaniem priorytetowym.

W niniejszej pracy wykorzystano kolekcję 142 ekotypów rzodkiewnika (*Arabidopsis thaliana* L.) w celu identyfikacji czynników genetycznych, mogących mieć związek z niepodatnością na dominujący aktualnie w Polsce patotyp *P. brassicae.* Dzięki zastosowaniu badania asocjacyjnego całych genomów (*ang. Genome Wide Association Study – GWAS*) zidentyfikowano dwa loci statystycznie istotnie powiązane z odpornością. Jeden z wytypowanych rejonów u genotypów rzodkiewnika Uod-1 oraz Est-1 zawierał gen *RPB1* (*ang. Resistance to Plasmodiophora brassicae1*) oraz jego homologi, podczas gdy drugi znajdował się w sekwencji kodującej genu *RAC1 (Resistance to Albugo candida 1)*, którego produktem jest receptor typu TIR-NB-LRR (*ang. Toll/Interleukin-1 Receptor Nucleotide-Binding site Leucine-Rich Repeat*). Udział czynnika RPB1 w odporności roślin na kiłę kapusty potwierdzono poprzez analizę mutantów typu utraty funkcji, wygenerowanych w niepodatnych genotypach Uod-1 oraz Est-1 przy pomocy techniki edycji genomów CRISPR/Cas9 (ang. *Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-Associated protein 9*). W uzyskanych mutantach *rpb1* obserwowano typowe symptomy choroby oraz akumulację patogenu, określoną poprzez pomiar DNA, porównywalną do wrażliwego genotypu kontrolnego Col-0.

Gen *RPB1* koduje niewielkie (148 aa, 16,0 kDa) białko transmembranowe, którego funkcja, czy też homologia do innych genów, jest na dzień dzisiejszy nieznana. Sekwencja kodująca tego genu jest obecna w wielu z badanych przez nas ekotypach i wykazuje stosunkowo

duży stopień zakonserwowania podczas gdy sekwencja promotorowa wykazuje duże zróżnicowanie. W genotypach odpornych transkrypcja genu *RPB1* nasila się po inokulacji *P. brassicae* i ma to związek z indukcją ekspresji genów odpowiedzialnych za reakcje obronne rośliny takich jak np. *PR5* (*Pathogenesis Related 5*) czy *CYP71A13 (Cytochrome P450 Monooxygenase 71A13*). Przejściowa ekspresja tego genu w liściach tytoniu (*Nicotiana tabacum* L.) uruchamiała odpowiedź typu nadwrażliwości (*ang. Hypersensitive Response HR*), co dodatkowo potwierdza hipotezę, w której zakładamy że jest on czynnikiem stymulującym odpowiedzi obronne. Linie delecyjne *rpb1* zachowują swą zdolność do odpowiedzi na inokulację (co obserwowano jako powstawanie niepełnego, niefunkcjonalnego transkryptu), jednakże nie prowadzi to do dalszej indukcji odpowiedzi obronnych. Powyższy fakt wskazuje na to, że białko RPB1 raczej nie jest zaangażowane w rozpoznanie patogenu, choć jest kluczowym elementem dla reakcji odporności.

Podczas poszukiwania odpornych ekotypów stwierdzono, że genotyp Pro-0 przejawiał mniej nasilone objawy choroby i akumulował wyższy poziom DNA patogenu. Postanowiono dokładniej przyjrzeć się temu zjawisku, w tym celu przeprowadzono porównawcze analizy histologiczne oraz transkryptomiczne dla wrażliwych ekotypów Col-0 oraz Pro-0. Stwierdzono, że hipokotyle formy Pro-0 zawierają wyższą proporcję ksylemu w stosunku do swej powierzchni a podczas obserwacji przekrojów pochodzących z roślin zainfekowanych stwierdzono, że rozwój tej tkanki jest w znacznie mniejszym stopniu zaburzony niż u ekotypu Col-0. Porównawcza analiza transkryptomiczna wykazała, że dynamika przeprogramowania rośliny przez patogen jest niższa u formy Pro-0. Można to stwierdzić na podstawie przebiegu powstawania narośli, w którym wyróżnia się fazę proliferacji, ściśle powiązaną z zaburzeniem różnicowania ksylemu oraz późniejszą fazę wzrostu, w której następuje spadek tempa podziałów komórkowych. Na podstawie zmiany transkrypcji genów związanych z postępem cyklu komórkowego oraz różnicowaniem ksylemu stwierdzono, że podczas gdy narośla formy Pro-0 są w stadium proliferacji u roślin Col-0 struktury te znajdują się w już fazie wzrostu. W związku z tym sformułowano hipotezę, według której różnice we wzroście i rozwoju rośliny żywicielskiej, zwłaszcza te w obrębie tkanki przewodzącej, mogą mieć istotne znaczenie dla postępu kiły kapusty. Dalsze zrozumienie tego zagadnienia oraz możliwości praktyczne jego wykorzystania wymagają dokładnego poznania roli poszczególnych genów odpowiedzialnych za regulację procesów rozwojowych rośliny.

Na podstawie przeprowadzonych eksperymentów oraz analizy ich wyników stwierdza się, że poprzez eksplorację naturalnej zmienności roślin rzodkiewnika zidentyfikowano czynniki

pełniące istotną rolę w odporności tego gatunku na kiłę kapusty oraz opisano możliwe mechanizmy związane z różnicami w przebiegu choroby u genomów podatnych.

Abbreviations

ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
AMM	Accelerated mixed model
BH	Benjamini Hochberg false discovery rate
Cas9	CRISPR associated protein 9
CC	Coiled coil domain
CCD	Canadian clubroot differential
cDNA	Complementary DNA
CDS	Coding sequence
CR	Clubroot resistance
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DEG	Differentially expressed genes
DI	Disease index
DNA	Deoxyribonucleic acid
dpi	Days post inoculation
ECD	European clubroot differential
EDTA	Ethylenediaminetetraacetic acid
EMMAX	Efficient mixed model association
ETI	Effector triggered immunity
ETS	Effector triggered susceptibility
GO	Gene ontology
gRNA	Guide RNA
GWAS	Genome wide association study

HR	Hypersensitive response
JA	Jasmonic acid
LB	Luria-Bertani
LRR	Leucine rich repeats
MAMP	Microbe-associated molecular pattern
mRNA	Messenger RNA
MS	Murashige and Skoog
NB	Nucleotide binding
NLR	Nucleotide-binding, leucine-rich repeat receptor
OD	Optical density
ORF	Open reading frame
padj	p-adjusted value
PAMP	Pathogen associated molecular pattern
PBS	Phosphate buffer saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PRR	Pattern recognition receptor
PTI	PAMP triggered immunity
qPCR	Quantitative polymerase chain reaction
QTL	Quantitative trait loci
RAC1	*Resistance to Albugo candida 1*
RIL	Recombinant inbred lines
RLK	Receptor like kinase
RNA	Ribonucleic acid

RPB1	Resistance to *Plasmodiophora brassicae* 1
SA	Salicylic acid
SAR	Systemic acquired response
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeats
TE	Transposable element
T-DNA	Transfer DNA
TIR	Toll-Interleukin receptor domain
TRIS	2-Amino-2-(hydroxymethyl) propane-1,3-diol
UTR	Untranslated region
WT	Wild type

Contents

1. Introduction

Clubroot disease is an important and destructive disease affecting all Brassicaceae crops such as oilseed rape (*Brassica napus*), broccoli, cauliflower (*Brassica oleracea*) and turnips (*Brassica rapa*). This disease is distributed worldwide and is one of the hardest to control in brassicas, because the chemical or biological controls available are not sufficiently effective in heavily infested fields. Resting spores produced by the causal pathogen *Plasmodiophora brassicae* can remain viable for an average of 3.5 to 5 years, and up to 20 years in the soil, making crop rotation hard to implement, especially for oilseed rape in Canada, where there is not a good alternative crop for oil production (Oxley, 2007; Peng et al., 2014b). Nevertheless, an integrated disease management that includes instrument and machinery disinfection, liming of the soil to increase the pH, crop rotation with non-host plants, and planting resistant cultivars are crucial to avoid important economic losses, though, none of them is 100% effective and each has advantages and disadvantages (Diederichsen et al., 2009).

The most characteristic symptom of the disease is the formation of galls in the roots and hypocotyls of the affected plants, caused by the abnormal cell growth and proliferation promoted by the pathogen inside the host cells. At later stages of the disease, the leaves can become chlorotic or necrotic and the plants can show stunting in the most severe cases, that may end up affecting the productive capacity of the plants, such as a severe reduction of oil production in oilseed rape crops (Dixon, 2009).

Despite the fact that it has been known that clubroot disease is caused by *P. brassicae* since 1878, the disease has received more attention in the last two decades, due to a great increase in disease incidence, especially in oilseed rape crops in countries like Germany and Canada, where the number of new infected fields increased by 49 between 2012 and 2015 or from 0 to 1064 between 2004 and 2012 respectively (Zamani-Noor, 2017; Strelkov and Hwang, 2014). In Poland, by quantifying the *P. brassicae* DNA in soil samples collected countrywide, it was determined that it is present in all the regions of the country, and that 62% of the analyzed fields have detectable amounts of pathogen DNA (Czubatka-Bieńkowska et al., 2020).

Clubroot disease also affects the model plant *Arabidopsis thaliana* which is a considerable advantage for the performance of detailed physiological and genetic studies in order to better understand the biology of the disease (Koch et al., 1991; Ludwig-Müller et al., 2009).

1.1 *Plasmodiophora brassicae* biology and life cycle

Plasmodiophora brassicae (Woronin, 1877), the microbial agent causing clubroot disease, is an intracellular obligate parasite that belongs to the protist eukaryotic supergroup of Rhizaria, class Phytomyxea, order Plasmodiophorida, that contains intracellular parasites of plants, and are characterized by a “cruciform” shaped division of the nuclei (Ward and Adams, 2010). Plasmodiophorida is the best studied order within the Rhizaria due to the economic importance of *P. brassicae* and other relatives like *Spongospora subterranea*, which causes powdery scab disease of potato (Neuhauser et al., 2014).

The life cycle of *P. brassicae* is complex and not very well understood due to the inability to generate axenic cultures, its intracellular lifestyle and the size of its cells and structures (Dixon, 2006). Nevertheless, it has been described with a good level of detail through the use of light and electron microscopy from the 1970’s (Ingram and Tommerup, 1972) and more recently progress has been made with improved electron microscopy and confocal microscopy to clarify the stages where previously controversy has arisen, for example if pathogen penetration occurs exclusively in the root hairs or also in the root epidermis (Tu et al., 2018; Liu et al., 2020). However, many aspects regarding the transition from the primary to secondary infection and the existence of sexual reproduction remain unclear.

The *P. brassicae* life cycle initiates with resting spores being released into the soil, coming from the decaying tissues of previously infected plants. The resting spore is the only life cycle phase capable of surviving for long periods outside the host and is composed of single haploid cells of around 3 µm of diameter, covered by a cell wall that contains chitin (Kageyama and Asano, 2009; Ingram and Tommerup, 1972).The germination of spores can be stimulated by unknown molecules coming from both, host and non-host plants roots, this signal has been proposed to be a hexasaccharide, but at the present date remains unknown (Friberg et al., 2005; Mattey and Dixon, 2015). From the germination of resting spores emerges the primary zoospore, which possess two flagellar structures that presumably aid the pathogen in reaching the host root system where they penetrate the host cell walls (Kageyama and Asano, 2009). Despite the fact that the mechanism of host cell penetration by the pathogen has not been observed precisely, different authors have proposed that the primary infection occurs in root hair cells, however recent evidence from confocal microscopy experiments, demonstrates that it also occurs in the root epidermal cells, predominantly in the elongation zone (Tu et al., 2018; Liu et al., 2020). Once inside the host

cell, a plasmodial structure forms and goes through nuclear divisions developing into a multinucleated plasmodium, and subsequently into multinucleate zoosporangium that undergo cytoplasmic cleavage to generate uninucleate secondary zoospores (Liu et al., 2020).

At the beginning of the secondary infection, the mature secondary zoospores are released and conjugate in pairs inside host root epidermal cells to form diploid zygotes at approximately 7 days post inoculation (dpi) in Arabidopsis. The development of the zygote continues with the formation of uninucleate secondary plasmodia that can be observed for the first time in the cortical cells. Once in the cortical cells, the pathogen undergoes mitotic divisions to develop into multinucleate secondary plasmodia, followed by meiotic divisions and the formation of resting spores, that can be observed around 24 dpi (Liu et al., 2020). Once the infected tissue dies, the mature resting spores are released into the soil, where they can survive up to 20 years, until the proper conditions to reinitiate the cycle. An overview of the entire life cycle is shown in Figure 1.

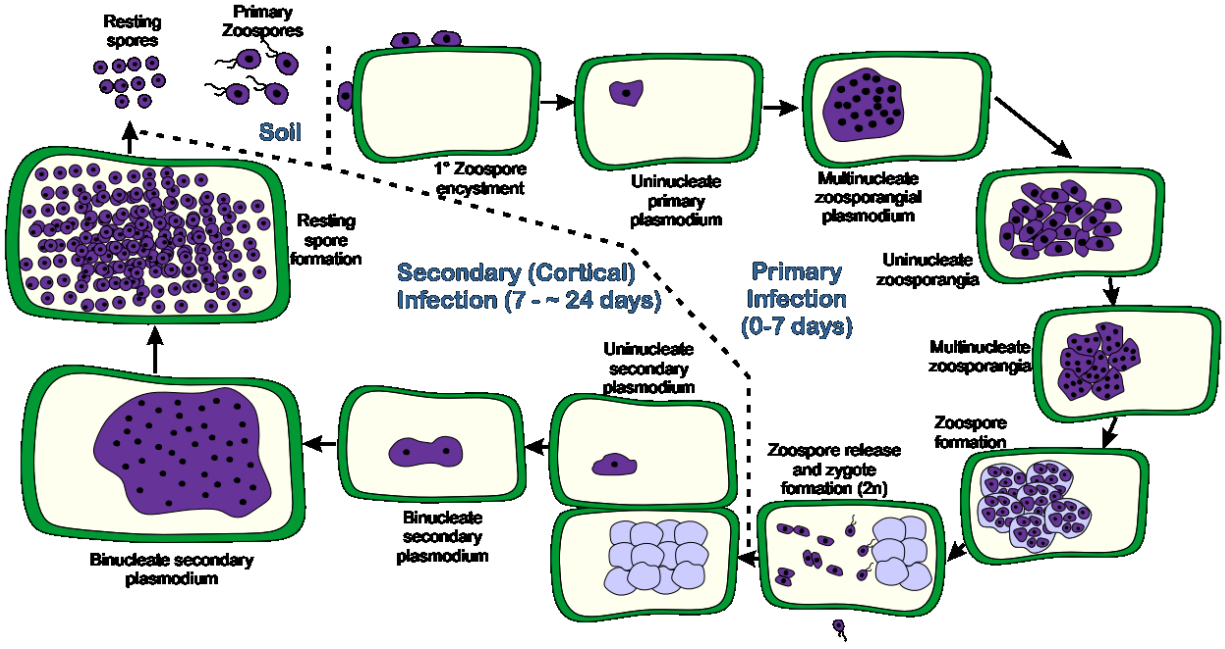


Figure 1. Schematic representation of the *P. brassicae* life cycle (Adapted from Liu et al., 2020)

1.2 *P. brassicae* pathotype classification

As previously mentioned, the impossibility of culturing *P. brassicae* under axenic conditions creates a great difficulty to generate genetically homogeneous pathogen strains, and in consequence, many studies regarding pathogen virulence and diversity are performed with field isolates classified into pathotypes according to the disease reaction calculated with a disease index (DI), based on the scoring the observed symptoms in a standardized panel of host species and cultivars. The main host sets used routinely includes the Somé system, shown in Table 1 (Somé et al., 1996), the Williams system (Williams, 1966), the European Clubroot Differential set (ECD) (Buczacki et al., 1975) and the Canadian Clubroot Differential set (CCD), that contains hosts of the differentials of Williams, Somé, some of the hosts included in the ECD and four *B. napus* cultivars (Strelkov et al., 2018). The list of species and cultivars used in each differential set is shown in Table 2. Other differentials have been developed as well, but more focused in the differentiation of local pathotypes or a specific crop. For instance, the Sinitic clubroot differential set was developed in China, due to the observation of important genetic variability between different isolates that were classified as pathotype 4 according to the Williams differential, improving the pathotype classification specifically for Chinese cabbage breeding (Pang et al., 2020). Through the use of differentials for classification, it has been possible to identify the pathotypes that are predominant in different countries or regions, but the appearance of new pathotypes that can break the existing sources of resistance remains a big concern for Brassicaceae breeders and growers and creates a big limitation to the pathotype classification using bioassays.

Table 1. Pathotype classification according to Somé et al., (1996)

Pathotype	*B. napus* "Nevin"	*B. napus* var. *napobrassica* "Wilhemsburger"	*B. napus* "Brutor"
P1	Susceptible	Susceptible	Susceptible
P2	Susceptible	Resistant	Susceptible
P3	Resistant	Resistant	Susceptible
P4	Resistant	Resistant	Resistant
P5	Resistant	Susceptible	Susceptible
P6	Susceptible	Resistant	Resistant
P7	Resistant	Susceptible	Resistant
P8	Susceptible	Susceptible	Resistant

Unfortunately, the results of pathotype classification with separate systems are not comparable because different sets of cultivars are used for the pathotype determination, moreover different isolates could be classified in the same pathotype despite being genetically dissimilar, making it hard to compare diversity between different regions or countries. Additionally, field isolates are likely to be a genetically heterogeneous pathogen mixture, which has been demonstrated with the molecular analysis of spores isolated from single galls (Fu et al., 2020).

To refine the heterogeneity that is sometimes observed in identifying pathotypes, different protocols have been developed to produce isolates derived from single spores that are genetically homogeneous which helped to increase the reproducibility of the bioassays using the previously mentioned differentials, however this is a challenging technique with a low success rate (Jones et al., 1982; Somé et al., 1996; H et al., 2021; Xue et al., 2008). Thanks to this method it was possible to overcome the genetic variability present in some field isolates, moreover these single spore isolates were used to obtain some of the 43 genomic sequences available for *P. brassicae* in the public databases (Recently reviewed by Schwelm and Ludwig-Müller, 2021). Unfortunately, the genome sequences of the sequenced isolates have not been enough to develop molecular markers proven to be useful and reproducible to discern between different pathotypes and which can be transferable between different regions or countries (Schwelm and Ludwig-Müller, 2021). Nevertheless, pathotype classification by bioassay, despite being labor intensive and time consuming, remains an important tool for breeding resistance in Brassica crops and for monitoring changes in the pathogen populations.

Table 2. List of species and cultivars used in the international clubroot differential set

Species and cultivars	Williams, 1966	ECD (Buczacki et al., 1975)	Somé et al., 1996	CCD (Strelkov et al., 2018)
ECD 01/ *B. rapa* ssp. *rapifera* line aaBBCC		Included		
ECD 02/ *B. rapa* ssp. *rapifera* line AAbbCC		Included		Included
ECD 03/ *B. rapa* ssp. *rapifera* line AABBcc		Included		
ECD 04/ *B. rapa* ssp. *rapifera* line AABBCC		Included		
ECD 05/ *B. rapa* var. *pekinensis* "Granaat"		Included		Included
ECD 06/ *B. napus* "Nevin"		Included	Included	Included
ECD 07/ *B. napus* "Giant Rape"		Included		
ECD 08/ *B. napus* selection ex "Giant Rape"		Included		Included
ECD 09/ *B. napus* "New Zealand resistant rape"		Included		Included
ECD 10/ *B. napus* var. *napobrassica* "Wilhemsburger"	Included	Included	Included	Included
ECD 11/ *B. oleracea* var. capitata "Badger Shipper"	Included	Included		Included
ECD 12/ *B. oleracea*		Included		
ECD 13 *B. oleracea* var. capitata "Jersey Queen"	Included	Included		Included
ECD 14/ *B. oleracea* var. capitata "Septa"		Included		
ECD 15/ *B. oleracea* var. *acephala* subvar. laciniata "Verheul"		Included		
B. napus "Brutor"			Included	Included
B. napus var. *napobrassica* "Laurentian"	Included			Included
B. napus "Mendel"				Included
B. napus "Westar"				Included
B. napus "45H29"				Included

1.3 Resistance to *P. brassicae*

Most of the research performed to find plants resistant to clubroot disease and determine the genetic basis of resistance has been carried out in *Brassica* species, as its members are the most important in terms of economic value of the whole Brassicaceae family. However, since the speciation arose from hybridization events, it is important to understand the genomic composition of the key members. The most important genetic concept in the *Brassica* genus is the U triangle (Figure 2), which, since being developed in the 1930's, has been validated and is supported by cytogenetics and molecular biology evidence (Prakash et al., 2009). The U triangle explains that the cytogenetic relationship between six species of the genus *Brassica*, in which *B. rapa* (genome A, n = 10), *B. nigra* (genome B, n = 8) and *B. oleracea* (genome C, n = 9) are diploid progenitors of the three allopolyploid species *B. juncea* (genome AB, n = 18), *B. napus* (genome AC, n = 19) and *B. carinata* (genome BC, n = 17) (Prakash et al., 2009). This relationship is crucial because the manner in which breeding for resistance to clubroot is done, depends on the genomes that each species possess as well as is interaction with other genomes (Hwang et al., 2012; Diederichsen et al., 2009).

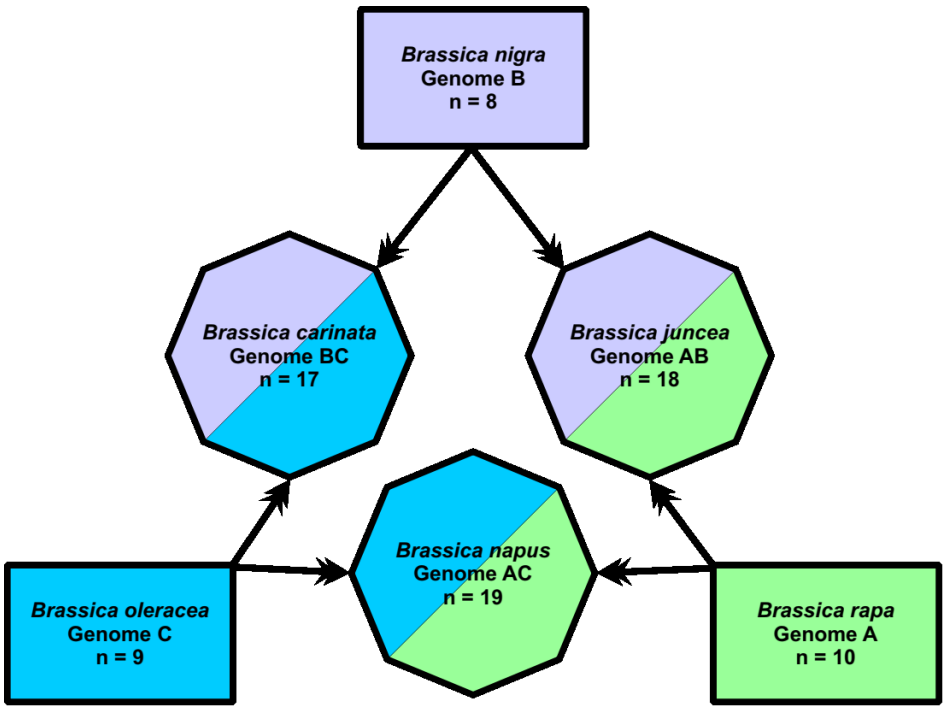


Figure 2. Schematic representation of the U triangle model (Adapted from Prakash et al., 2009)

1.3.1 Resistance in members of the genus *Brassica*

From all the brassica species mentioned, *B. rapa* (genome A) is the one where the highest amount of resistant accessions has been found in different studies (Peng et al., 2014a; Hasan et al., 2012; Rahman et al., 2014). Some of the accessions of *B. rapa*, particularly the European lines, have shown resistance to multiple isolates of the pathogen and were included in breeding programs to develop cultivars with broad resistance (Diederichsen et al., 2009). Moreover, several studies using molecular markers in mapping populations coming mostly from double haploid lines, allowed the identification of QTLs that control resistance to different clubroot isolates, including *CRa* (Matsumoto et al., 1998), *Crr1, Crr2* (Suwabe et al., 2003), *Crr3* (Hirai et al., 2004), *Crr4* (Suwabe et al., 2006), *CRb* (Similar to *CRa*) (Piao et al., 2004), *CRk* and *CRc* (Sakamoto et al., 2008) and *RCr1* (Chu et al., 2013). From these QTLs, two resistance genes that encode Toll-Interleukin Nucleotide-Binding Leucine-Rich Repeats (TIR-NB-LRR) proteins (typically involved in plant immunity) have been cloned and characterized; one corresponding to the *CRa* QTL (Ueno et al., 2012) and the second corresponding to *Crr1* (Hatakeyama et al., 2013). Moreover, it was demonstrated that the *CRb* locus contains six NB-LRR genes, and the overexpression of two of them in Arabidopsis and *B. rapa* susceptible plants, confer resistance to a specific *P. brassicae* pathotype, but it seems that it is the same *CRa* gene that was cloned previously (Hatakeyama et al., 2017). Recently, by using genotyping by sequencing (GBS), a higher resolution mapping allowed the identification of three new resistance loci named *Rcr4, Rcr8* and *Rcr9* to Canadian pathotypes. However, the candidate genes coding for TIR-NB-LRR proteins contained within these QTL regions did not show any variation in sequence between resistant and susceptible varieties (Yu et al., 2017).

Contrary to the situation observed in *B. rapa*, the genetic resistance observed in *B. oleracea* (genome C) seems to be partial, polygenic and in some the cases, caused by recessive alleles (Voorrips, 1995; Diederichsen et al., 2009). Different studies have identified resistance in germplasm collections (Hasan et al., 2012; Peng et al., 2014a), but their use in both assisted selection and conventional breeding has been difficult, and in some cases the resistance had to be introgressed through *B. napus* or *B. rapa* hybrids (Diederichsen et al., 2009). The first studies aimed at mapping QTLs in *B. oleracea*, found a small number major or minor QTLs that explained resistance to different *P. brassicae* pathotypes, but the resolution of the map was insufficient to develop markers useful in breeding programs (Voorrips et al., 1997; Rocherieux et al., 2004). More recent studies with increased amounts

of molecular markers allowed the construction of linkage maps with higher resolution, but in two of them, the result confirmed that the resistance is polygenic by detecting the presence of multiple QTLs with minor or major effect in the resistant to different *P. brassicae* isolates (Nagaoka et al., 2010; Peng et al., 2018). In one of the studies, with the use of GBS, it was possible to identify just two QTLs that were considered as major (Lee et al., 2015), but to the date, no fine mapping or positional cloning has been performed in order to identify the genes inside the identified QTLs. Finally, another QTL called *Rcr7*, was discovered through bulked segregant RNA sequencing and some candidate TIR-NB-LRR genes were identified, but their cloning and functional validation has yet to be published (Dakouri et al., 2018).

B. nigra (genome B) has not been an important source for breeding *P. brassicae* resistance, thus the studies in this species are less developed than in the other two diploid species, however different studies have found resistant accessions to isolates that are widespread in Canada (Peng et al., 2014a; Hasan et al., 2012). Further studies managed to identify a dominant locus called *Rcr6* using bulked segregant RNA sequencing (BSR-Seq) and Kompetitive Allele Specific PCR (KASP) for mapping in *B. nigra* and *B. rapa* genomes. One TIR-NB-LRR candidate was identified but its putative function still requires functional validation (Chang et al., 2019).

B. napus, commonly known as oilseed rape or Canola is one of the most economically important *Brassica* crops, especially in the last 20 years, mainly because of its capacity for oil production. Due to this, significant research efforts have been launched to identify sources of resistance to *P. brassicae* and introduce it into highly productive cultivars. Unfortunately, *B. napus* (hybrid AC genome) is a recent amphidiploid species without wild accessions and low intraspecific variation, and in consequence, the use of natural resistance sources depends on introgression mainly from *B. rapa* and to a lesser extent from *B. oleracea* (Diederichsen et al., 2009). Diederichsen and Sacristan (1996) successfully identified *B. napus* plants with strong and broad resistance from re-synthetized lines using highly resistant parentals from *B. rapa* and *B. oleracea*. The re-synthetized lines were used as parental for important commercial cultivars like “Mendel”. In initial mapping studies it was determined that the Mendel cultivar probably received one resistant dominant gene and two recessive from its parental lines (Diederichsen et al., 2006).

By using bulk segregant analyses combining AFLP and SSR markers, it was possible to identify 19 QTLs conferring resistance to seven different isolates in the mentioned re-synthetized lines (Werner et al., 2008). Interestingly, further comparative

analyses combined with the development of molecular markers, showed that the resistance to pathotype 3 in “Mendel” stemmed from the genes *CRa/CRb*^{kato}, previously identified in *B. rapa* (Fredua-Agyeman and Rahman, 2016). In other studies using a *B. napus* mapping population, it was also possible to identify QTLs with major effect and two QTLs with additive effect to two different *P. brassicae* isolates (Manzanares-Dauleux et al., 2000). Resistance to pathotype 4 was also assessed through a Genome Wide Association Study (GWAS) using 472 accessions of *B. rapa.* This experiment enabled the identification of a total of 9 QTLs including 7 that were not reported previously (Li et al., 2016).

Among the multiple screens that have been done in the *Brassica* species, no source of resistance has been found in *B. carinata* or *B. juncea*, despite all the diploid progenitors having some loci providing resistance. However, some *B. juncea* accessions have shown at least partial resistance to some *P. brassicae* isolates (Peng et al., 2014a).

1.3.2 Resistance in *Raphanus sativus*

Raphanus sativus (radish) has become an important object of studies in breeding resistance to clubroot disease because the resistance reaction is prevalent among different cultivars, even though some cultivars have shown susceptibility, in most of the cases they were recently bred and might have lost some clubroot resistance loci (Diederichsen et al., 2009). The first attempt to map the resistance of *R. sativus* to clubroot disease, led to the discovery of a single QTL locus named *Crs1* that might be considered as monogenic, but the presence of closely linked genes or QTLs with minor effects was not discarded (Kamei et al., 2010). More recent studies using restriction-site-associated DNA sequencing (RAD-Seq) to create a higher density maps, applied to an F_2 population developed from two *R. sativus* inbred lines, allowed the identification of a further five QTLs, one of them showing synteny with the *Crr1* region of *B. rapa* (Gan et al., 2019). However, the resistance genes contained in these loci remain to be identified.

Breeding efforts have also been made to try to introgress *R. sativus* resistance into *Brassica spp.* genomes by creating amphidiploid plants using crosses with *B. oleracea,* to create a fertile *Brassicoraphanus* with a potential use in breeding programs (Chen and Wu, 2008). This hybrid has been used to successfully introgress resistance into *B. napus*, but the developed plants had reduced fertility and bore extra R chromosomes from radish, which are problems that might be overcome by successive backcrossing (Zhan et al., 2017).

1.3.3 Resistance in Arabidopsis

Arabidopsis is the most important model plant owing to the multiple advantages that it has compared to other plant species such as small size, short life cycle, natural self-pollination, high productivity and efficient autogamous reproduction, its diploid nature, and a small sized genome. These advantages and its selection as the first plant to have its complete genome sequenced have created the situation that most of the knowledge regarding genetics, physiology, development, signaling and immunity in plants come from experiments performed in Arabidopsis. Additionally, there are extensive amounts of publicly available resources such as mutant lines, T-DNA introgression collections, collections of natural variation, mapping populations, genome sequences and many other information resources that make biological studies more productive (Koornneef and Meinke, 2010). Regardless of the previously mentioned characteristics, immunity to *P. brassicae* in Arabidopsis and the resistance response remains poorly understood.

The first attempt to elucidate the genetic basis of Arabidopsis resistance to clubroot disease started with the identification of resistant plants in a set of 30 Arabidopsis accessions with 4 different clubroot pathotypes. Only two accessions, Tsu-0 and Ze-0, were classified as resistant to the *P. brassicae* isolate “e”, and Tsu-0 was used to elaborate a mapping population with the susceptible parental Cvi-0 to identify a single dominant locus named *Resistance to P. brassicae 1* (*RPB1*) in chromosome 1 (Fuchs and Sacristán, 1996). Complementary studies confirmed that the resistance of these accessions was monogenic according to an allelism test and showed some hypersensitive response (HR) features and localized lignification in response to infection, but it was not complete and some viable resting spores were still produced. The accessions used in this and subsequent studies exhibited differing levels of resistance, with Ze-0 being the most resistant accession followed by Tsu-0 and Ta-0 (Kobelt et al., 2000). Through a higher mapping resolution with a population of 4230 plants coming from the Cvi-0 X Tsu-0 crossing, it was possible to narrow down the QTL to a region of approximately 71kb containing three pseudogenes and 13 coding sequences, however none of these candidate genes were homologous to known immune or defense signaling components (Arbeiter et al., 2002).

Using two other *P. brassicae* isolates (eH and MS6), Alix et al. (2007) observed partial clubroot resistance in the accessions Bur-0, Tsu-0 and Kn-0. In subsequent studies a Col-0 X Bur-0 F2 population and a recombinant inbred lines (RIL) from the same parental accessions were used to identify 4 more QTLs named *PbAt1, PbAt4, PbAt5.1* and *PbAt5.2*,

none of them corresponding with the previously identified *RPB1* locus. These QTLs were shown to function additively and some of them exhibited epistasis between each other (Jubault et al., 2008). Of these QTLs, *PbAt5.1* was also linked to trehalose tolerance in the Bur-0 accession as well as *P. brassicae* resistance, due to the higher levels of trehalose and trehalase activity observed in the susceptible accession Col-0 (Gravot et al., 2011). Further studies found that the QTL *PbAt5.2* was associated with camalexin biosynthesis and consequently with reduced growth of *P. brassicae*, but this QTL does not contain any camalexin biosynthesis genes (Lemarié et al., 2015b). In further studies, heterogeneous inbred lines from a Bur-0 X Col-0 crossing were used to assess the impact of three irrigation regimes and its interaction with the symptoms caused by *P. brassicae*, finding that under low irrigation conditions the *PbAt5.2* QTL is associated with a reduced disease severity while the *PbAt5.1* QTL was associated with partial resistance under a high irrigation regime (Gravot et al., 2016). So far, none of the genes contained in these QTLs have been associated directly with resistance to *P. brassicae.* All together, these results showed that the resistance of Bur-0 is polygenic and is to some extent dependent on plant growth conditions, which makes it harder to identify the most important genes related to immunity.

To test if the resistance to *P. brassicae* is dependent on epigenetic variation, 123 epiRILs derived from the cross between the mutant *ddm1* (*Decrease in DNA Methylation 1*), that has reduced susceptibility to the eH isolate of *P. brassicae*, and the susceptible Col-0 accession, were used to identify QTLs under epigenetic control (QTL^{epi}). A total of 20 QTL^{epi} were found after evaluation with four different methods to assess the phenotype including disease index scoring, qPCR pathogen quantification, leaf length and root biomass. Six QTL^{epi} overlapped with QTLs in previously reported regions including the locus containing *RPB1*, meaning that the differences in the resistance levels could be associated with both genetic and epigenetic variation in the same regions (Liégard et al., 2019a).

A recent non-peer reviewed manuscript in pre-print form describes narrowing down the region of the QTL *PbAt5.2* to 26kb through fine mapping in a F7 RIL population derived from the Bur-0 X Col-0 crossing. Within this region it was possible to identify 8 ORFs, three of them coding for NLRs proteins (Liégard et al., 2019b). Two of these NLRs, *At5g46260* and *At5g47280*, only contained one SNP or were identical respectively, implying that the variation in phenotype is not related to sequence variation, however, it was observed that the susceptible parental Col-0 showed hypermethylation and almost undetectable expression of these two genes, contrasting with Bur-0, where the region was found to be

hypomethylated and *At5g46260* and *At5g47280* showed higher expression (Liégard et al., 2019b). Also, by evaluating a collection of natural accessions containing either the Col-0 and Bur-0 epialleles, it was possible to associate this hypomethylation with higher resistance to the eH pathotype (Liégard et al., 2019b). Nevertheless, the functionality of these genes has not yet been confirmed through overexpression, silencing, or knocking out.

In other research works, the reaction of 84 Arabidopsis accessions to four different pathotypes collected in Canada was assessed, it was found that none of the accessions were resistant to all tested pathotypes, and most of them showed intermediate or strong susceptibility to all of them. Only the accessions Ct-1, Pu2-23, Ws-2 and Sorbo exhibited high resistance to one specific pathotype each (Sharma et al., 2013).

In summary, mapping the resistance in Arabidopsis has enabled the discovery of QTLs with both major and minor effects, but none of these studies has yet established and validated the role of the genes underpinning resistance. Moreover, there is evidence that some of the differences observed between the resistant and susceptible accessions might be caused by differences at the nucleotide level or epigenetic factors. An overview of the QTLs and cloned genes involved in resistance to clubroot is shown in Table 3.

Table 3. Number of QTLs and cloned genes involved in clubroot resistance in Brassicaceae species

Species	QTLs	Cloned genes
B. rapa	18	2 *CRa*, *CRR1a*
B. oleracea	≈50	0
B. napus	≈45	1 *CRa* (probably the same from *B. rapa*)
B. nigra	1	0
R. sativus	6	0
A. thaliana	5 + 20 QTL^{epi}	0

1.4 Changes in phytohormone dynamics and cell cycle progression during *P. brassicae* infection in Arabidopsis

The most influential plant hormones during clubroot disease can be classified in two groups; the first group are the hormones related to defense responses and response to biotic stress such as salicylic acid (SA) and jasmonic acid (JA) and secondly the hormones involved in cellular dynamics, for example auxins and cytokinins. Whereas SA/JA are mostly considered to act antagonistically, and their influence is different when comparing resistant or susceptible Arabidopsis - *P. brassicae* interactions, they have both been linked to

activation of defense responses to counteract clubroot disease. On the other hand, the auxins and cytokinins are involved in the dramatic anatomical changes undergone by the host, especially during secondary infection, where increased cell proliferation within developing galls at early stages (hyperplasia), is followed by an abnormal cell expansion (hypertrophy) (Ludwig-Müller, 2014). The hyperplasia observed in the hypocotyl tissues mainly occurs in the vascular cambium, where radical pathogen driven changes are observed, such as the suppression of xylem differentiation and an increase in phloem complexity (Malinowski et al., 2019).

1.4.1 The role of SA and JA during clubroot disease

SA and JA are the two most important hormones involved in plant immunity, despite this they are considered to act antagonistically. There is a complex crosstalk between these hormones, which enables plants to fine-tune responses to different biotic interactions (Aerts et al., 2021). SA is the main hormone involved in the response against biotrophic pathogens, whereas the JA can act in concert with both ethylene (ET) and abscisic acid (ABA) in the response against necrotrophic pathogens or herbivores (Aerts et al., 2021).

Studies conducted in Arabidopsis mutants with impaired SA accumulation, either through the mutation in a key SA biosynthesis component (*sid2*) or through the addition of a bacterial transgene which degrades SA (*NahG*) resulted an increase in clubroot susceptibility measured by shoot weight relative to the wild type (WT) and gall scoring (Lovelock et al., 2016). Other mutants such as *cpr1* and *dnd1*, that have constitutive activation of SA mediated responses, presented less severe symptoms than WT controls indicating that elevated SA levels protect against clubroot disease(Lovelock et al., 2016). Similar studies have shown that the mutant *cpr5-2*, which also exhibits constitutive accumulation of SA, has reduced susceptibility to clubroot and upregulation of the gene *PR2*, a marker of SA mediated responses. SA concentration has been measured at late time points (21 and 28 dpi) in the susceptible accession Col-0 inoculated with the e3 pathotype in infected roots and shoots, revealing that in roots the SA concentration is lower than in the non-infected control at 21 dpi, but higher at 28 dpi, contrasting with higher concentration at both time points in the shoot (Lemarié et al., 2015a). The concentration of methyl-salicylate (Me-SA), a less active and highly mobile conjugate of SA, was found to accumulate in both roots and shoots in response to infection (Ludwig-Müller et al., 2015). The movement of

Me-SA has been studied using the SA analogue and Me-SA esterase inhibitor 2,2,2',2'-tetrafluoroacetophenone (TFA), revealing that *P. brassicae* interferes with the transport of Me-SA from roots to shoots. Interestingly, it was found that *P. brassicae* bears a gene coding for a salicylic acid methyl transferase (*PbBSMT*) capable of converting SA into Me-SA (Ludwig-Müller et al., 2015). Overexpression of *PbBSMT* in Arabidopsis increased the susceptibility to *P. brassicae* but also to the bacterial pathogen *Pseudomonas syringae,* most likely by stimulating the conversion of SA into the rapidly relocated form Me-SA (Bulman et al., 2019; Djavaheri et al., 2019).

Studies at the gene expression level have shown that SA responsive genes *PR2* and *PR5* are upregulated in the partially resistant Arabidopsis accession Bur-0 but not in the susceptible Col-0 during the secondary infection stage and this was consistent with the measured SA concentration, which was higher only in infected Bur-0 roots (Lemarié et al., 2015a). The exogenous treatment of plants with SA has also been proven to reduce symptoms in the above-ground parts of infected plants, however *P. brassicae* DNA levels quantified by qPCR in roots and hypocotyls remained unchanged when compared to the untreated plants (Lemarié et al., 2015a). Transcriptomic studies utilizing microarrays to compare responses of the Bur-0 accession to the eH and e_2 pathotypes, where partial resistance or susceptibility are observed respectively, revealed a strong upregulation of genes related to the SA pathway including *PAD4, SID2* (a.k.a. *ICS1*) and *NPR1* in plants infected with the eH pathotype but not with the e_2 pathotype (Jubault et al., 2013). In general, it can be concluded that SA promotes defense responses to clubroot disease, particularly when partial resistance is observed, however it might not be enough to completely restrict pathogen growth.

Jasmonic Acid (JA) is a plant hormone predominantly associated with plant defense against herbivores and necrotrophic pathogens. In many cases it has been shown to act as an antagonistic hormone to SA, nevertheless the SA-JA cross talk can be more complex, and the cooperation between both hormones have been observed under particular conditions (Thaler et al., 2012; Mine et al., 2017).

The active form of JA is the conjugate jasmonoyl-isoleucine (JA-Ile), which is synthesized by the enzyme JAR1 and interacts with the COI1-JAZ receptor complex to modulate gene expression in response to biotic stress or developmental changes (Katsir et al., 2008). Most of the studies involving gene expression and transcriptomics have found a consistent upregulation of genes involved in the JA pathway during clubroot disease.

Transcriptomics studies using next generation sequencing have shown that many genes involved in the biosynthesis of JA are upregulated, whereas ones involved in JA signaling, such as *JAR1* and *COI1* are downregulated at early and late stages of secondary infection, while some of the *JAZ* genes that are involved in the turnover of JA signaling complexes are upregulated (Schuller et al., 2014; Irani et al., 2018). Some studies have shown that *JAR1* is downregulated upon clubroot infection and the mutant *jar1-1* has shown increased susceptibility to the pathogen measured with a shoot index approach, meaning that JA activation could have a negative impact on *P. brassicae* growth (Siemens et al., 2002, Siemens et al., 2006). JA accumulation has been measured and compared to SA concentration in compatible and incompatible interactions in the Col-0 and Bur-0 accessions respectively, revealing a strong increase with a clubroot susceptible host but only slight increases with the partially resistant Bur-0. This observation was supported by data from exogenous treatments with SA and JA, as SA was able to reduce the symptoms of both Col-0 and Bur-0, but JA only reduced the symptoms in Col-0, however the amount of pathogen quantified by qPCR remained unchanged after SA treatment and was reduced in both accessions after JA treatment (Lemarié et al., 2015a).

All together these results demonstrate that the activation of SA in resistant accessions is involved in counteracting clubroot disease, but is not sufficient by itself and some other defense responses are required for full resistance. On the other hand, JA accumulation appears to be predominant in the susceptible accessions and has a negative impact in the pathogen growth, however, the role of this hormone in counteracting the pathogen, does not appear to be antagonistic to SA. This situation can also be related to the capacity of the pathogen to interfere with SA signaling.

1.4.2 Auxins and cytokinins during clubroot

The role of auxins during clubroot infection has been studied from the perspective of the accumulation, synthesis, transport and more recently from the signal transduction pathway, in the interest of creating a broad outlook of their influence on disease progression. Concerning the accumulation of auxins, initial studies reported conflicting results regarding auxin concentration inside the galls (Ludwig-Müller et al., 1993), however, by using the reporter gene *GUS* fused to the synthetic auxin responsive promoter *DR5*, it was possible to track auxin responsiveness from 5 dpi observed initially in the epidermal cells and later

through all cell layers in the root at 10 dpi (Devos et al., 2006; Ludwig-Müller et al., 2009). The main biosynthetic pathway that is apparently responsible for the increase in auxin concentration is the synthesis of IAA (indole acetic acid) from IAN (indole acetonitrile), by the nitrilase enzymes (Grsic-Rausch et al., 2000). Nitrilase genes *NIT1* and *NIT2* were found to be upregulated at 21 dpi and 32 dpi respectively upon infection and localization of the NIT2 protein, was predominant in infected cells containing sporulating plasmodia; concomitantly, the mutant *nit1-3* showed a reduced gall size compared to the wild type controls, probably due to its lower concentration of IAA (Grsic-Rausch et al., 2000). The expression of *NIT2* has also been studied through a promoter *GUS* fusion (*NIT2::GUS*) and contrasted with the DR5 IAA reporter patterns, confirming the role of nitrilases in the late stages of the gall development (cell enlargement), despite the fact that auxin concentration begins to increase at earlier stages (Päsold et al., 2010).

Transcription of the auxin receptor genes *TIR1* and *AFB1* was found to be upregulated at 24 and 28 dpi and their corresponding T-DNA mutants *tir1, afb1-3*, and *afb1-3 afb2-3* showed increased susceptibility for low concentrations of pathogen inocula (10^4 spores / ml). It is thought that the auxin binding protein ABP1 could be activating potassium channels, which can be partially responsible for the hyperplasia observed in the enlarged cells (Jahn et al., 2013). The evidence indicates that an increase in the biosynthesis and accumulation of auxins, positively impacts the pathogen's ability to reproduce and form galls in the plant tissues and the formation of enlarged cells, however it is not well known whether the pathogen is capable of manipulating auxin signaling or its biosynthetic pathways. The *P. brassicae* genome does contain a gene (*PbGH3*) coding for an auxin-responsive Gretchen Hagen 3 protein that resembles plant GH3 proteins at the structural level, and functional studies in heterologous systems provide evidence that it can conjugate auxins or JA with amino acids, suggesting that *P. brassicae* has the potential to interfere with auxin or JA signaling (Schwelm et al., 2015).

In contrast with auxins, the cytokinins appear to have a more important role in earlier stages of clubroot disease development, in consequence, the expression levels of the genes *ARR10*, *ARR5* and *CRE1/AHK4* that code for proteins involved in cytokinin signaling, were found to be upregulated at 10 dpi (early secondary infection). Additionally, by using the promoter reporter fusion construct *ARR5::GUS* it was shown that cytokinin responsiveness begins to increase between 3 to 5 dpi, indicating a general increase in responsiveness to cytokinins during the primary infection and the early stages of the secondary infection

(Devos et al., 2006; Siemens et al., 2006). Underlying the role of cytokinin signaling components, the biosynthesis and catabolism of cytokinins has also shown to be important for gall development in the infected plants. For instance, two root-specific cytokinin oxidases, *CKX1* and *CKX6,* involved in cytokinin degradation, were downregulated during *P. brassicae* infection at 10 and 23 dpi in Arabidopsis plants. Moreover, the overexpression of *CKX1* and *CKX3* conferred reduced gall size and symptoms, underlining the importance of cytokinin accumulation for gall development (Siemens et al., 2006). Transcriptomic studies performed at other time points (16 dpi and 26 dpi) corroborated the fact that not only is cytokinin degradation suppressed, but also that cytokinin biosynthetic components, namely the isopentenyl transferases family genes *IPT1* and *IPT7* are upregulated in response to infection, though the absolute expression levels are not as high as those of other *IPT* genes (Malinowski et al., 2016). The quadruple mutant *ipt1;3;5;7* showed reduced gall size and delayed disease progression, infected cells were still hypertophied but the amount of cell division at sites of infection was greatly reduced (Malinowski et al., 2016). It was possible to identify two isopentenyl transferases genes in the *P. brassicae* genome, potentially involved in cytokinin synthesis, and those were found to be expressed *in planta* at 16 and 26 dpi pointing to the possibility that the pathogen may manipulate host cytokinin dependent developmental pathways by directly supplying the hormone (Schwelm et al., 2015; Malinowski et al., 2016).

Even though the general role of hormones during clubroot disease has not been definitively fleshed out and other hormones such brassinosteroids have an impact during disease progression, the evidence shows that the cytokinins tend to increase during the proliferative stage of the disease, whereas the auxins have a more significant role during the cell enlargement and gall formation.

1.4.3 Impact of *P. brassicae* on host cell cycle and development

As mentioned previously, the formation of galls in both roots and hypocotyls is considered to occur in two steps: hyperplasia and then hypertrophy. These two steps require changes in the host cell cycle that can facilitate the completion of pathogen development and reproduction. The first insights regarding the relation of clubroot disease and changes in the host cell cycle came from histological characterization of Arabidopsis plants using the promoter fusion *CYCB1*::GUS, a reporter of cell division; these studies showed an increase

in its activity from 4 dpi, observed initially in some cortical cells, and reaching all the cell layers by 10 dpi (Devos et al., 2006). Further studies with *CYCB1*::GUS at later time points, confirmed that in uninfected plants the reporter is active mainly in the vascular cambium, while in infected plants its activity spreads outwards to cambial descendant cells. This observed signal overlaps with the expression of *ANT* which is a marker of meristematic activity (Malinowski et al., 2012). Subsequently, it has been found that the expression pattern of key genes involved in cell cycle regulation is differentially impacted between the early and late stages of secondary infection, particularly the B-type cyclins, that regulate the duration of the G2 phase of the cell cycle. The expression of genes that promote transition to mitosis is triggered during the early phase of secondary infection (16 dpi), but not in the later phase (26 dpi) when the formation of enlarged, spore filled cells occurs (Olszak et al., 2019). Prolonged maintenance of the mitotic state in cambial progeny cells during the proliferative phase has been linked to certain key regulators: the transcription factor E2Fa which associates with RBR1 to promote cell proliferation and suppress cell differentiation, and the transcription factors MYB3R1 and MYB3R4, that upregulate the expression of B-type cyclins, these factors are important for the stimulation of host cell division to create growth for *P. brassicae* colonization (Olszak et al., 2019). Later, during the cell enlargement phase, it has been noted an increase in the level of host cell endoreduplication leading to higher ploidy levels. Through characterizing the mutant *ccs52a1,* that is compromised in endoreduplication, a reduction in the overall size of the galls and enlarged cells was observed, however the pathogen was still able to progress through its life cycle, meaning that while endoreduplication contributes to cell enlargement, hence allowing the host to bear a greater number of resting spores, it is not necessary for the maturation of resting spores at the end of *P. brassicae's* life cycle (Olszak et al., 2019).

Together with the cell cycle changes, a strong disturbance in the anatomy of the vascular cambium occurs during gall formation, that changes the vascular cambium from its usual ring shaped pattern into broken, isolated islands (Malinowski et al., 2012). During clubroot infection a boost of the meristematic activity in the vascular cambium and the phloem parenchyma is observed. Moreover, the xylem formation is reduced which is reflected by decreased expression of genes involved in xylem specification (*VND*6/7 and *MYB*46) and maturation (*XCP*1/2) (Malinowski et al., 2012). On the other hand, there is an increment in phloem differentiation in plants seen in anatomical observations together with use of *CLE44::GUS* reporter lines, which revealed that the meristematic activity of the vascular cambium is affected during clubroot disease, creating a disbalance that reduces

xylem maturation and promotes phloem differentiation, evidenced by the increased complexity of phloem bundles during clubroot (Malinowski et al., 2012). These observations have also been supported by the study of genes participating in early steps of phloem differentiation such as *OPS*, *CVP2* and *BRX*, and the corresponding knock out mutants *ops-2*, *cvp2-1 cvl1-1 brx-2,* that displayed anatomical differences compared to the wild type controls, faster completion of the pathogen life cycle and a premature death of the plant (Walerowski et al., 2018; Malinowski et al., 2012) When the host was unable to respond to the pathogen's attempts to remodel vascular development due to these mutations the resulting interaction was a net negative for both sides with more rapid host death before any possibility to flower and make seed and smaller galls with fewer spores for the pathogen.

1.5 Genome Wide Association Studies (GWAS) in Arabidopsis

Genome Wide Association analysis is a method used to identify and associate polymorphisms present in the genomes of different individuals of the same species with a particular variation in their phenotypes. Unlike the traditional QTL mapping, the association mapping relies in the linkage disequilibrium present in a natural population, and has the advantage that it is not required to create mapping populations by targeted crossings, and higher percentages of genetic recombination can be detected (Korte and Farlow, 2013). GWAS was initially developed in human genetics to take advantage of natural genetic diversity and because of the impossibility of mapping genes using biparental mapping populations. Arabidopsis was identified as a perfect model for these kind of studies because its populations consist mainly of naturally inbred accessions caused by successive self-pollinations, producing a predominantly homozygous diploid genome (Aranzana et al., 2005). The first GWAS for Arabidopsis was performed with a set of 95 accessions with genomic SNP data, and using traits for which the underlying major responsible genes were already known, such as flowering or resistance to *Pseudomonas spp*. Despite the fact that it was possible to re-identify several known loci, the authors also observed a high rate of false positives because the population evaluated was highly structured, the algorithms used to correct this issue were not fully developed, and the SNP marker density was low (Aranzana et al., 2005). To provide a solution to these problems, a 250k SNP Affymetrix chip was developed to be used with a population of 199 Arabidopsis accessions that were used to map loci underlying 107 known phenotypes, additionally an improved algorithm to correct for population structure of plants known as Efficient Mixed Model Association

(EMMA) based on a parametric mixed model was used, allowing the identification mainly of alleles of major effects (Atwell et al., 2010).

A significant advance for GWAS in Arabidopsis was achieved with the development of next generation sequencing technologies that increased dramatically the speed of sequencing and reduced its costs, leading to the genomic sequences of 1135 Arabidopsis accessions and the identification of approximately 10 million biallelic SNPs, covering both, the biological diversity of Arabidopsis populations and the genetic diversity across the whole genome (Alonso-Blanco et al., 2016). All the data generated in the this study is curated, standardized, freely available, and has been use to develop online platforms to conduct GWAS online with reproducible pipelines for the analysis, for example easyGWAS (https://easygwas.ethz.ch/), and GWA portal (https://gwas.gmi.oeaw.ac.at/), that enable analyses in very simple steps following the upload of phenotype data (Grimm et al., 2017; Seren, 2018). Additionally, these websites make use of the most recent mixed model approaches that can deal with complex population structures and have very efficient use of computational resources to perform a full analysis in 30 minutes. Since the publication of the 1135 genomes, also known as the 1001 genomes project, it has been possible to identify different loci responsible for many different kind of traits such as developmental, biotic and abiotic stresses (Tsuchimatsu et al., 2020; Rungrat et al., 2019; Ferrero-Serrano and Assmann, 2019; Nakano et al., 2020).

1.6 Fundamental concepts in plant immunity and plant-pathogen interactions

To protect themselves from harmful microbes, plants have evolved complex immune systems capable of recognizing pathogens and activating defense mechanisms to counteract and restrict their growth. The recognition of pathogens is mainly carried by two different groups of proteins, the first, known as Pattern Recognition Receptors (PRR), act extracellularly and recognize conserved molecules known as Microbe-Associated Molecular Patterns (MAMPs, also known as PAMPs for Pathogen-Associated Molecular Patterns), or to recognize molecules produced as a result of damage caused by a pathogen or herbivore, known as Damage-Associated Molecular Patterns (DAMPs) (Jones and Dangl, 2006; Boller and Felix, 2009). The second group of genes, known as resistance genes (R genes), can recognize directly or indirectly molecules secreted by the pathogens known as effectors, which are proteins that allow the pathogen to establish a successful infection, by interfering

with the signaling that can trigger an immune response, or manipulating the host metabolism for their own benefit (Han, 2019).

The recognition of MAMPs and DAMPs by PRRs is called Pattern-Triggered Immunity (PTI) and is a part of the basal immunity of the plant, together with physical barriers such as the plant cell wall or cuticle. Once PTI is triggered, the plant can secrete molecules that directly restrict pathogen growth or act to deprive nutrients from the microbe (Ranf, 2017). The PRR proteins comprise mainly two types of transmembrane receptors, the receptor-like kinases (RLK) and the receptor like proteins (RLP). The RLKs are composed of an intracellular kinase domain, a transmembrane domain, and different kinds of extracellular domains such as (i) the Leucine Rich Repeat (LRR) domain, that mainly perceives peptides for example the bacterial flagellin peptide flg22 or the elongation factor peptide elf18, (ii) the LysM domains that are involved in carbohydrate perception of fungal chitin or bacterial peptidoglycans, or (iii) the B lectin domain, that recognizes bacterial lipopolysaccharides (Ranf, 2017). The RLPs contain extracellular and transmembrane domains, but lack an intracellular one, indicating that they probably depend on other proteins to initiate intracellular signaling. Both RLKs and RLPs require to form heteromeric complexes with other membrane associated co-receptors to activate immune signaling, for instance the flagellin receptor FLS2 associates with the regulatory receptor BAK1 once the peptide flg22 is recognized (Couto and Zipfel, 2016).

After the activation of PTI, three main physiological changes are observed; an increase in the concentration of Ca^{2+} ions, the production of extracellular reactive oxygen species by the respiratory burst oxidase homologue protein D (RBOHD) and the mitogen-activated protein kinase (MAPK) cascades, which together with the increase in Ca^{2+} concentration, relay the signal to the nucleus, triggering the transcriptional reprogramming required to mount a defense response which is characterized by the reinforcement of cell walls, production of antimicrobial enzymes and compounds and synthesis of hormones involved in secondary responses (Couto and Zipfel, 2016). Most of the microbes that are recognized through PTI are considered non-adapted as they have not evolved mechanisms to counteract these responses, however, adapted pathogens have evolved effectors that can interfere in different aspects of the plant immune response, such as the recognition events or deployment of defense.

When pathogens successfully evade PTI through the secretion of effectors that cause disease, the outcome can be characterized as Effector Triggered Susceptibility (ETS).

As defined by van der Burgh and Joosten (2019), the effectors are *"proteins that are secreted by a pathogen into the apoplast or the cytoplasm of the host upon attack, with the aim to prevent or circumvent plant defense and thereby promote disease. Typically, the expression of genes encoding effector proteins is highly induced in planta."* Nowadays, numerous effectors have been described and functionally analyzed in fungi, oomycetes, viruses, and nematodes, but in bacteria the number of effectors functionally studied is higher, especially from the genus *Pseudomonas* spp. and *Xanthomonas* spp., a significant percentage of the effectors characterized have been shown to actively interfere with PTI components - ROS production, changes in intracellular Ca^{2+} and induction of defense genes (Gimenez-Ibanez et al., 2018; Martel et al., 2021). Nevertheless, the effectors have an incredible diversity and can interfere as well with many other cellular processes, for instance membrane trafficking at the endoplasmic reticulum level or have protease activity, as is observed in the apoplastic effectors secreted by *Phytophthora spp.* (Wang and Jiao, 2019).

To overcome ETS, plants have evolved other types of receptors encoded by so-called "R-genes", that predominantly act in the intracellular space upon direct or indirect recognition of a pathogen effector to trigger a strong defense response that impedes pathogen proliferation *in planta*, this phenomenon is known as Effector Triggered Immunity (ETI). The dynamics between PTI, ETS and ETI is commonly known as the zig-zag model (Jones and Dangl, 2006; Martel et al., 2021). Most R-proteins contain three main domains, consisting of one variable N-terminal domain, a Nucleotide-Binding Apaf1-Resistance-CED4 (NB-ARC) domain, and a C-terminal Leucine Rich Repeat (LRR) domain, these proteins are commonly referred to as NLRs. In the N-terminal domain there are three common forms: the Coiled-Coil domain (CC), the Toll and Interleukin-1 Receptor (TIR) domain and the RPW8 domain. Recent findings have shown that they act differently to activate immune responses (Lu and Tsuda, 2021). Presently it is considered that the recognition of the pathogen effectors by NLRs can fit four different models, the first one, known as direct intracellular effector recognition, occurs when the NLR receptor physically interacts with the effector and initiates an immune response, for example the receptor RPP1 that recognizes the ATR1 effector from *Hyaloperonospora arabidopsidis* (Krasileva et al., 2010). The second model proposed is the guard model and occurs when the NLR recognizes a change in a protein known as "guardee" that was caused by the activity of the effector, in this model is important to mention that the guardee protein is actively involved in defense processes (van der Hoorn and Kamoun, 2008). One well known example is the recognition the effectors AvrB and AvrRpm1 of *P. syringae*, these effectors can trigger the degradation of phosphorylation of

the guardee RIN4 from Arabidopsis, activating immune responses mediated by the NLRs RPS2 and RPM1. In the absence of the mentioned NLRs, RIN4 acts as a negative regulator of plant defense (Mackey et al., 2002, 2003; Kourelis and Van Der Hoorn, 2018). The third model, known as the "decoy" model is very similar to the guard model, but the main difference is that the decoy does not have any other function in the absence of the cognate NLR, this model was proposed because the authors consider that the guardees have on them two opposing evolutionary forces, the diversification of NLRs and the diversification of effectors (van der Hoorn and Kamoun, 2008). One good example of the decoy model is the recognition of the effector HopZ1 from *P. syringae*, mediated by the ZED1 decoy and the NLR ZAR1. HopZ1 acts through the acetylation of the pseudokinase ZED1 which induces the activation of ZAR1, however ZED1 does not appear to have any particular function in the absence of ZAR1 (Lewis et al., 2013). The fourth and final model is known as integrated decoy and occurs when the NLR contains an additional domain that has the function of recognizing the pathogen effector, for instance, the recognition of the effectors AvrRps4 and Pop2 from *Ralstonia solanacearum* is perceived by an integrated WRKY domain in the C-terminal of the Arabidopsis NLR RRS1, which forms a complex with a second NLR, RPS4. Interestingly AvrRps4 and Pop2 interfere with other WRKY transcriptions factors to hamper the upregulation of defense responses, so the WRKY domain in RRS1 is acting as an integrated decoy that has the exclusive function of interacting the cognate effector (Sarris et al., 2015). The recognition mechanisms involving NLRs have been well studied in general terms, however the activation mechanisms to induce defense responses frequently remain unclear and is the focus of much research effort. Recent discoveries have shown that NLRs, depending on their function can be classified as sensors or helpers. The NLRs mentioned above, are good examples of sensor NLRs, as they are participating in the perception of the pathogen, on the other hand, helper NLRs participate exclusively in the downstream signaling. Most of the helper genes described to date contain the RPW8 domain, and in Arabidopsis are classified into the subclasses ADR1 and NGR1, the first being involved in the downstream signaling initiated by both TIR-NB-LRR and CC-NB-LRR NLRs, and the second only by TIR-NB-LRRs (Dong et al., 2016; Castel et al., 2019).

Significant advances to elucidate the mechanism of activation of TIR-NB-LRR and CC-NB-LRR receptors have been published recently, improving our understanding of the link between recognition mediated by the NLR and the subsequent triggering of defense mechanisms. The mechanism of activation by CC-NB-LRRs was studied for ZAR1 NLR of Arabidopsis upon perception of the effector AvrAC of *Xanthomonas campestris* pv.

campestris. Prior to the perception of the effector, ZAR1 guards the RKS1 pseudokinase, remaining in an off state. Once the effector enters the cytoplasm, it uridylylates the “decoy” kinase PBL2 causing an interaction between RKS1 and the uridylated PBL2, once this happens, an ADP molecule associated to the NB domain is replaced by ATP, initiating the oligomerization of a pentamer made of units of ZAR1-RKS1-PBL2 conjugates known as a resistosome (Wang et al., 2019b, 2019a; Dangl and Jones, 2019; Mermigka and Sarris, 2019). The resistosome has a barrel shape structure with the CC domains pointing to the center, forming a channel that could function as ion transporter across the membrane, but this function remains to be experimentally validated (Dangl and Jones, 2019). The formation of a resistosome has also been demonstrated for the TIR-NB-LRR NLRs, through study of the recognition of the XopQ mediated by the NLR ROQ1. In this case, ROQ1 interacts with XopQ through a C terminal C-JID domain, this causes the oligomerization of four units ROQ1, allowing the NB domains to bind ATP, this oligomerization causes an association of the four TIR domains, activating NAD hydrolase activity that can cleave NAD^+ triggering programmed cell death, commonly known in plants as the hypersensitive response (HR) (Martin et al., 2020).

With the development of the Zig-Zag model, some researchers have raised the question whether ETI is simply an amplification of PTI or if they have a more complex crosstalk. The evidence obtained so far shows that despite different recognition mechanisms, they share many components downstream, for example the increase in cytosolic Ca^{2+}, the ROS burst, the upregulation of defense genes and the synthesis of hormones required in defense (Lu and Tsuda, 2021). Currently some adjustments are being proposed to the zig-zag model, as the evidence points that the PTI is a primary immunity against pathogens, that after being hampered by pathogen effectors or downregulated by other PTI regulators, requires a compensation provided by the ETI mediated by the NLR proteins, nevertheless, it does not mean that ETI is separate from PTI, but dependent on the PTI machinery (Yuan et al., 2021). Interestingly, recent studies found that ETI induces the upregulation of genes coding for proteins that are part of the PTI initial signaling events, such as the co-receptor BIK1, RBOHD and the MAP kinase MPK3, and probably one of the functions of ETI is restoring the PTI that was suppressed by pathogen effectors, moreover it was also proved that the immunity response triggered by the recognition of the effector AvrRps4 by the NLRs RRS1/RPS4 was negatively affected in Arabidopsis mutants compromised in PTI responses, and that that PTI potentiates the HR caused by ETI (Ngou et al., 2021). In conclusion, the recent evidence shows that the strongest and most effective

immune responses observed come from the synergistic function of PTI and ETI; an explanatory diagram of this model is shown in Figure 3.

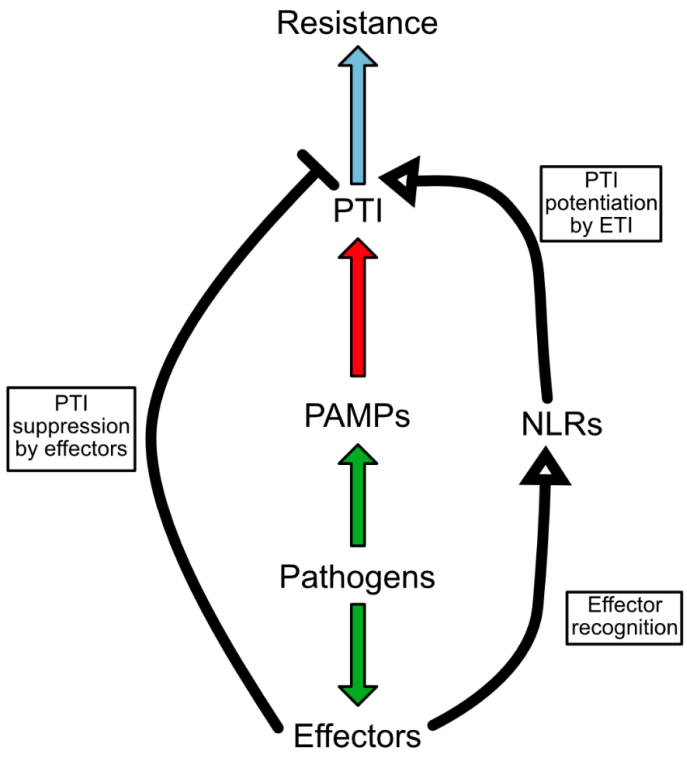


Figure 3. Schematic representation of the PTI and ETI relation in the plant immune system (adapted from Ngou et al., 2021)

In the interaction between Arabidopsis and *P. brassicae*, some few effectors interfering with defense responses have been identified and their enzymatic activity has been validated biochemically in heterologous systems, however, the mechanisms involved in the recognition of the pathogen remain elusive, despite the fact that Arabidopsis is a model plant. In previous studies, some candidate genes have been identified, but currently none of them has been validated experimentally. In this research the Arabidopsis natural variation and its genetic resources were used to identify candidate genes involved in the resistance to a predominant *P. brassicae* isolate present in Poland and the generation of loss of function mutants with the CRISPR/Cas9 methodology was used to test its role during its interaction with *P. brassicae.*

2. Objective and hypotheses

2.1 Objective of the research

The ultimate objective of this research was the identification and characterization of genetic factors underpinning resistance or susceptibility to clubroot disease in *Arabidopsis thaliana.*

The main concept of this work is based on the exploration of natural diversity and genome sequence resources available for Arabidopsis accessions, and the molecular tools developed for genome editing, to dissect the interaction with a Polish isolate of *Plasmodiophora brassicae.*

2.2 Hypotheses

- The natural genetic variation found in Arabidopsis accessions can be used to identify candidate genetic factors responsible for resistance or susceptibility to *P. brassicae*.
- Differences in pathogen colonization between susceptible accessions could be partially attributed to anatomical variations.

3. Materials and methods

3.1 Natural Arabidopsis accessions used for GWAS and growth conditions

A total of 142 Arabidopsis accessions were obtained from the Nottingham Arabidopsis Stock Centre (NASC) and selected from the Nordborg collection (Nordborg et al., 2005), the 1001 Genomes consortium collection (Alonso-Blanco et al., 2016) and the parentals of the Multiparent Advanced Generation Inter-Cross (MAGIC) populations (Kover et al., 2009). The detailed information of the genotype collection and its country of origin is presented in Table 4. For the GWAS experiments the seeds were stratified for 4 days at 4 °C degrees in distilled water and then sowed directly into a mixture of soil: perlite (5:1 vol/vol). The plants were grown under controlled conditions using a short-day photoperiod of 9 hours of light at 22 ˚C and 15 hours of dark at 20 ˚C. For the light conditions FluorA L 36W/77 lamps with an irradiance of 120 µmol m^{-2} s^{-1} were used. The relative humidity in the chamber was maintained at 65%.

When plants were grown for experiments to assess the expression of genes in the roots in response to *P. brassicae* they were sown on a mixture of sand and soil 1:1 vol/vol to facilitate the cleaning of the roots prior to RNA extraction and reduce mechanical stress that can alter the expression pattern of the target genes.

For the evaluation of the CRISPR/Cas9 knock-out (KO) lines and the experiments including the Pro-0 accession, the seeds were disinfected with 40% commercial bleach (4% NaClO) and washed 5-7 times with distilled sterilized water. The seeds were sown on half strength MS media (Duchefa, Haarlem, The Netherlands) supplemented with 1% sucrose and 0.7% agar and were stratified for 4 days at 4˚C and before being transferred to the incubator short-day photoperiod of 9 hours of light at 22 ˚C and 15 hours of dark at 20 ˚C with an irradiance of 120 µmol m^{-2}. Ten-day old seedlings were transferred into a mixture of soil: perlite (5:1 vol/vol) and grown in the conditions described above.

Table 4. List of accessions included in the GWAS experiments.

The Nottingham Arabidopsis Stock Centre (NASC) code is provided in the second column.

NASC	Ecotype	GWA	Origin	NASC	Ecotype	1001 genomes ID	Origin
N22564	RRS-7	7514	USA	N22641	Tsu-1	6972	JPN
N22565	RRS-10	7515	USA	N22642	Mt-0	6939	LIB
N22566	Knox-10	6927	USA	N22643	Nok-3	6945	NED
N22567	Knox-18	6928	USA	N22644	Wa-1	7394	POL
N22568	Rmx-A02	7524	USA	N22645	Fei-0	9941	POR
N22569	Rmx-A180	7525	USA	N22647	Ts-1	6970	ESP
N22570	Pna-17	7523	USA	N22648	Ts-5	6971	ESP
N22571	Pna-10	7526	USA	N22649	Pro-0	8213	ESP
N22573	Eden-2	6913	SWE	N22650	LL-0	6933	ESP
N22574	Lov-1	6043	SWE	N22651	Kondara	6929	TJK
N22575	Lov-5	6046	SWE	N22652	Shahdara	6962	TJK
N22576	Fab-2	6917	SWE	N22653	Sorbo	6963	TJK
N22577	Fab-4	6918	SWE	N22655	Ms-0	6938	RUS
N22579	Bil-7	6901	SWE	N22656	Bur-0	7058	IRL
N22580	Var2-1	7516	SWE	N22658	Oy-0	7288	NOR
N22581	Var2-6	7517	SWE	N22659	Ws-2	6981	RUS
N22582	Spr1-2	6964	SWE	N76429	Abd-0	6986	UK
N22583	Spr1-6	6965	SWE	N76645	Adam-1	9609	RUS
N22584	Omo2-1	7518	SWE	N76431	Ak-1	6987	GER
N22585	Omo2-3	7519	SWE	N77612	Borky1	428	CZE
N22586	Ull2-5	6974	SWE	N78418	Can-0	7063	ESP
N22587	Ull2-3	6973	SWE	N78206	Cnt-1	7064	UK
N22588	Zdr-1	6984	CZE	N76794	Dem-4	8233	USA
N22589	Zdr-6	6985	CZE	N76804	Dolna-1-39	9711	BUL
N22590	Bor-1	5837	CZE	N76818	DraIV 2-9	5907	CZE
N22591	Bor-4	6903	CZE	N76477	Durh-1	7107	UK
N22592	Pu2-7	6956	CZE	N76483	Er-0	7125	GER
N22593	Pu2-23	6951	CZE	N76485	Est	7127	GER
N22594	Lp2-2	7520	CZE	N76489	Fr-2	7133	GER
N22596	HR-5	6924	UK	N76492	Gel-1	7143	NED
N22597	HR-10	6923	UK	N76496	Gr-1	430	AUT
N22598	NFA-8	6944	UK	N76909	Halca-1	9732	SVK
N22599	NFA-10	6943	UK	N76921	Hi-0	8304	NED
N22600	Sq-1	6966	UK	N78904	Is-0	8312	GER
N22601	Sq-8	6967	UK	N76969	Kn-0	7186	LTU
N22603	CIBC-17	6907	UK	N76538	La-0	7209	POL
N22604	Tamm-2	6968	FIN	N77020	Ler-0	7213	GER
N22605	Tamm-27	6969	FIN	N78261	Li-2:1	7223	GER
N22606	Kz-1	6930	KAZ	N76550	Mh-0	7255	GER
N22607	Kz-9	6931	KAZ	N76558	Na-1	8343	FRA
N22609	Got-22	6920	GER	N77128	No-0	7273	GER
N22610	Ren-1	6959	FRA	N78941	OOE3-2	15593	AUT
N22611	Ren-11	6960	FRA	N76572	Pi-0	7298	AUT
N22612	Uod-1	6975	AUT	N28648	Po-0	7308	GER
N22613	Uod-7	6976	AUT	N78300	Pog-0	7306	CAN
N22614	Cvi-0	6911	CPV	N76417	Qui-0	9949	ESP
N22615	Lz-0	6936	FRA	N77201	Rak-2	8365	CZE
N22616	Ei-2	6915	GER	N77222	Rsch-4	7322	RUS
N22617	Gu-0	6922	GER	N77247	Sf-2	7328	ESP
N22618	Ler-1	6932	GER	N77256	Smolj-1	9718	BUL
N22619	Nd-1	6942	SUI	N77279	Stiav-1	9728	SVK
N22620	C24	6906	POR	N76608	Ta-0	7349	CZE
N22621	CS22491(N13)	7438	RUS	N77389	Tsu-0	7373	JPN
N22622	Wei-0	6979	SUI	N78774	Udul 1-11	6296	CZE
N22623	Ws-0	6980	RUS	N78777	Uk-3	10022	GER
N22624	Yo-0	7416	USA	N78856	Wil-2	7413	LTU
N22625	Col-0	6909	USA	N28838	Wu-0	7415	GER
N22626	An-1	6898	BEL	N28846	Zu-0	7417	SUI
N22627	Van-0	7383	CAN	N77063	IP-Mah-6	9906	ESP
N22628	Br-0	6904	CZE	N77105	IP-Moj-0	9869	ESP
N22629	Est-1	6916	RUS	N799182	IP-Adc-5	9513	ESP
N22631	Gy-0	8214	FRA	N799207	IP-Bus-0	9830	ESP
N22632	Ra-0	6958	FRA	N799289	IP-Orb-10	9565	ESP
N22633	Bay-0	6899	GER	N799297	IP-Piq-0	9883	ESP
N22634	Ga-0	6919	GER	N799316	IP-Rib-1	9890	ESP
N22635	Mrk-0	6937	GER	N76535	Kyoto	7207	JPN
N22636	Mz-0	6940	GER	N76433	Altai-5	9758	CHN
N22637	Wt-5	6982	GER	N22595	Lp2-6	7521	CZE
N22638	Kas-1	7183	IND	N22602	CIBC-5	6908	UK
N22639	Ct-1	7067	ITA	N22646	Se-0	6961	ESP
N22640	Mr-0	7522	ITA	N22657	Edi-0	7111	UK

3.2 Pathogen propagation and inoculum preparation

The *P. brassicae* isolate used was obtained from Prof. Małgorzata Jędryczka and was classified as pathotype P1 according to the Somé differential (Some et al., 1996), moreover it was established that the pathogen is capable of infecting the *B. napus* resistant cultivar “Mendel”, so the isolate was classified as P1B (Ramzi et al., 2018). Chinese cabbage *B. rapa* var. pekinensis cultivar “Granaat” was used for pathogen propagation and the bulking of spores.

The pathogen spores were prepared according to the method of Fuchs and Sacristán (1996), briefly, 2-3 frozen galls coming from Chinese cabbage plants were homogenized in a blender containing 200 ml of distilled water and filtered through several layers of sterile gauze. The filtrate was centrifuged, and the starch layer was removed from the spore pellet mechanically using a spatula, this process was repeated several times until most of the starch was removed. The spore concentration was determined using a haemocytometer and was diluted to a final concentration of 1 x 10^6 spores ml^{-1}, each Arabidopsis plant was inoculated with 2 ml of the calibrated spore suspension at 17 days after sowing when the seeds were sown directly into the soil, or 7 days after the seedlings were transferred from MS culture media.

3.3 DNA extraction and *P. brassicae* relative quantification with qPCR

The hypocotyl together with the upper 1 cm of root from each plant was collected at 19 dpi. DNA extraction was performed from the combined tissue of three plants, the tissue was macerated in 2 ml centrifuge tubes with 650 µl of DNA extraction buffer containing 100 mM of TRIS-HCl pH = 8.0, 50 mM of EDTA pH = 8.0, 500 mM of NaCl and SDS 1.3% in a TissueLyser II (Qiagen, Hilden, Germany) with two metal beads for 2 minutes at 30 Hz. The DNA was purified with 5 M potassium acetate followed by precipitation with a mixture of isopropanol and 3 M sodium acetate. The samples were treated with RNAse 0.6 µg $µl^{-1}$ (Qiagen, Hilden, Germany) for 30 min at 37 °C and the reaction was stopped with chloroform: isoamyl 24:1 followed by precipitation with a mixture of absolute ethanol and sodium acetate. The DNA samples were quantified on a Nanodrop 2000 (Thermo Scientific) and diluted to a final concentration of 20 ng $µl^{-1}$ for use as template in the qPCR reaction.

The PCR reactions were prepared in a final volume of 10 µl containing 1X NEB Luna qPCR Master Mix, 0.25 µm of each primer and 90 ng of DNA template. The PCR was performed in a Roche Light Cycler 480 using a two-step protocol as follows; an initial denaturation step at 95 ˚C for 2 min followed by an amplification step of 40 cycles at 95 ˚C for 15 s and 60 ˚C for 30 s. The specificity of the products was assessed with a melting curve from 50 ˚C to 95 ˚C with 5 acquisitions per ˚C. The target genes amplified of Arabidopsis and *P. brassicae* were respectively *AtSK11* (*At5g26751*) and *Pb18S* (ENSRNAT00050137123), the sequence and details of the primers are presented in Table 5.

To calculate the relative pathogen DNA amount, the CP value of each sample was calculated using the 2nd derivative max method included in the Roche Light Cycler 480 software. Two technical replicates were performed for each sample/gene and averaged to obtain the value used to estimate the relative expression. The $\log_2$ relative expression was calculated by obtaining the difference between CP*Pb18S* and CP*AtSK11*. The number of biological replicates is specified for each experiment in the results section. Statistical analyses were performed with linear models in R version 4.0.3 (2020-10-10) managed by RStudio Version 1.4.1103. using the R packages RSTATIX version 0.7.0 (https://cloud.r-project.org/web/packages/rstatix/index.html) and Agricolae version 1.3-5 (https://cran.r-project.org/web/packages/agricolae/index.html). For all the statistical analyses based on linear models, tests for normality and heteroskedascity were routinely applied to the log2 data before other transformations were considered

Table 5. List of primers used for relative quantification of *P. brassicae.*

Primer Name	Sequence 5' → 3'	Target gene and use	Source
AtSK11_F	CTTATCGGATTTCTCTATGTTTGGC	Amplification of the Arabidopsis gene *At5g26751*. Reference for qPCR quantification of *P. brassicae.*	(Botanga et al., 2012)
AtSK11_R	GAGCTCCTGTTTATTTAACTTGTACATACC		
Pb18s_F	AAACAACGAGTCAGCTTGAATGC	Amplification of the 18S gene of *P. brassicae* (ENSRNAT00050137123). Target gene for qPCR quantification.	(Lemarié et al., 2015b)
Pb18s_R	AGGACTTGGCTGCGGATCAC		

3.4 Disease index scoring (DI)

Clubroot disease index scoring was performed at 19 dpi and according to the method of Ludwig-müller et al., (2017), using the following scale (Figure 4):

0. No symptoms observed.
1. Presence of small clubs only in lateral roots, the root architecture is conserved.
2. Small clubs present in the main and lateral roots, the main root might be also thickened, but there are not visible symptoms in the hypocotyl.
3. Medium to big clubs and galls are observed in the main and lateral roots, most of the fine roots are absent and the hypocotyls starts to show some swelling.
4. Severe galls observed in the roots and hypocotyls reaching the rosette. The fine roots are destroyed and in most of the cases the plants look wilted and stunted.

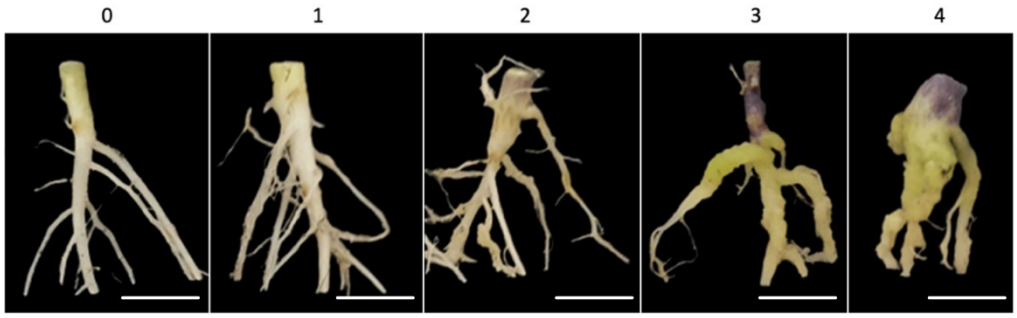


Figure 4. Representative symptoms to assess disease index scoring.
The scale bar represents 500 µm

The DI score was calculated according to the equation:

$$DI = \frac{1 * n_1 + 2 * n_2 + 3 * n_3 + 4 * n_4}{4N}$$

Where N corresponds to the total number of plants evaluated and n_1 to n_4 denote the number of plants in each class of the symptom severity scale.

3.5 Genome-wide association analysis to identify candidate genes

The GWAS was performed using the online available tools GWA portal (https://gwas.gmi.oeaw.ac.at/) and easyGWAS (https://easygwas.ethz.ch/) which contain all the SNPs, geographic information and analysis pipelines to perform standardized GWAS in *A. thaliana* (Seren, 2018; Grimm et al., 2017).

3.6 Generation of CRISPR/Cas9 knock-out lines in resistant Arabidopsis accessions

3.6.1 Selection and design of gRNAs and generation of plasmid constructs

For each candidate gene, two gRNAs were designed using the CRISPOR online tool, that allows selection and evaluation *in-silico* of the efficiency of generating double strand breaks and also predicting possible off-target binding sites in the genome (Haeussler et al., 2016; Concordet and Haeussler, 2018). The cloning strategy was performed as previously described by Bieluszewski et al., (2019), in short, two gRNAs that were spaced 200 – 400 bp apart were selected to cause a deletion in the target gene, to facilitate the genotyping of the edited lines using simple PCR. The selected guides were chemically synthesized and cloned independently by mutagenesis with PCR using Clontech CloneAmp HiFi PCR Premix (Takara Bio Inc., Shiga, Japan) according to the manufacturer's instructions. For DNA template, the pJET1.2 vector containing the promoter of the spliceosome genes U3 or U6 from Arabidopsis was used (Figure 6). To generate the vector, the PCR product was phosphorylated using a polynucleotide kinase (PNK) (Thermo Scientific, USA) followed by recircularization by ligation with a T4 ligase (Promega, USA). All the primers utilized during the cloning process are shown in Table 6. The recircularized vector containing the gRNA was transformed into TOP10 chemically competent *E. coli* cells. All the plasmid constructs containing the cloned gRNAs were confirmed by Sanger sequencing.

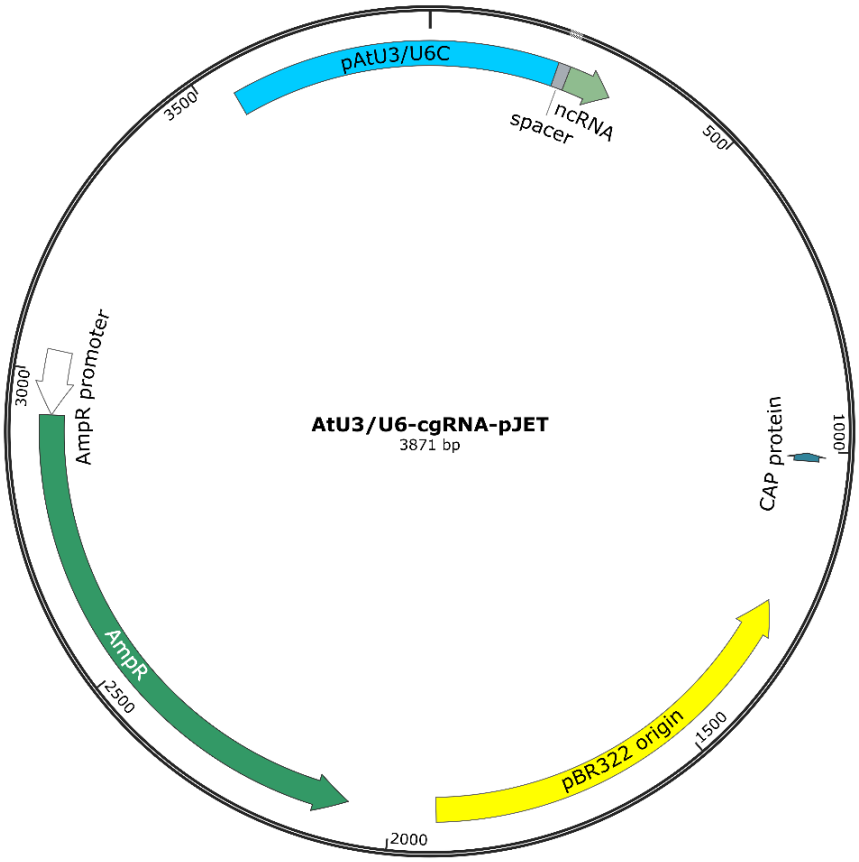


Figure 5. AtU3/U6-CgRNA-pJET1.2 plasmid representative map.

Table 6. List of primers used for cloning gRNA in the pICU2:Cas9-dsRED. The sequence highlighted in red and green are complementary to the U3 and U6 promoter respectively.

Primer Name	Sequence 5' → 3'	Target gene and use	Source
JFL-F	GTTTTAGAGCTAGAAATAGCAAG	Forward primer to clone gRNAs in the AtU3/U6-cgRNA-pJET1.2 vector.	Tomasz Bieluszewski, personal communication
gRNA179-1g31540-U3	TCAAGGAATCCAGGATCGCTGACCCTTGATGCTTTCTATGCA	Reverse primers to clone gRNAs targeting *RAC1*(At1g31540) by mutagenesis in the AtU3/U6-cgRNA-pJET1.2 vector.	This manuscript
gRNA303-1g31540-U6	TGTGGATCCTTCCCAAGTGACAATCACTACTTCGACTCTAGCTGT		
gRNA104-U6_RPB1	GCGGTGAACCGGTGAAGTTCAATCACTACTTCGACTCTAGCTGT	Reverse primers to clone gRNAs targeting *RPB1*(At1g32049) by mutagenesis in the AtU3/U6-cgRNA-pJET1.2 vector.	This manuscript
gRNA476-U3_RPB1	GTGCCTGTCCACCTCAGTTATGACCCTTGATGCTTTCTATGCA		
gRNA124-U3-like4	CCGGTAGTCCGGTACACAGTGACCCTTGATGCTTTCTATGCA	Reverse primers to clone gRNAs targeting *RPB1*-like4 by mutagenesis in the AtU3/U6-cgRNA-pJET1.2 vector.	This manuscript
gRNA483-U6-like4	TAGAGGGGCAACGATGAGAGCAATCACTACTTCGACTCTAGCTGT		
VRF1	ATGTTACTAGATCGGGGATCCGGATGGCTCGAGTTTTCAGC	Primers required for amplification of the U3/U6 gRNA cassette from the pJET1.2 vector to clone up to 4 gRNAs in tandem in the pICU2:Cas9-dsRED vector through a Gibson assembly approach.	Tomasz Bieluszewski, personal communication
R2	CCATGATTACGCCAAGCTCG		
pR2F1	CTTGGCGTAATCATGGGGATGGCTCGAGTTTTCAGC		
R1VF	AGAATTCCCATGGAAGGATCCTCGAGGCTGCAGGAATTCGATATCAAGC		
R4F1	CTCATGAAAACTACGAGGATGGCTCGAGTTTTCAGC		
R3F1	CAAGATTTTCAGGCTGGGATGGCTCGAGTTTTCAGC		
R3	CAGCCTGAAAATCTTGAGAGAATAAAAG		
R4	TCGTAGTTTTCATGAGAGTCGATTG		

The transformation vector containing the Cas9 protein was pICU2:Cas9-dsRED (Figure 6). This vector contains the gene coding for the Cas9 protein from *Streptococcus pyogenes* expressed under the promoter of the *Incurvata 2* gene (*ICU2, At5g67100*) highly active in the shoot apical meristem of 1-day old seedlings, to increase the chance of obtaining homogeneously edited plants. To select the potential transgenic plants the vector contains both a fluorescent dsRED positive selection marker under the *Napin* promoter that causes strong expression in seeds and a cassette for negative selection with phosphinothricin. To clone the U3/U6::gRNA combination for each gRNA, each cassette was amplified separately with primers containing complementary regions with pICU::Cas9-dsRED to assemble the construct through a Gibson assembly enzymatic reaction (Table 6). The linearized vector pICU::Cas9-dsRED was obtained through a digestion with BamHI enzyme (New England Biolabs, USA) and then mixed together with the PCR products of the U3/U6::gRNA cassettes and the Gibson assembly master mix (New England Biolabs, Ipswich, Massachusetts, USA), the DNA proportions and reaction conditions were performed according to the manufacturer's instructions. The assembled plasmid construct was transformed into NEB 5-alpha Competent *E. coli* (High Efficiency) and confirmed by restriction digestion with the *XhoI* enzyme (New England Biolabs, USA).

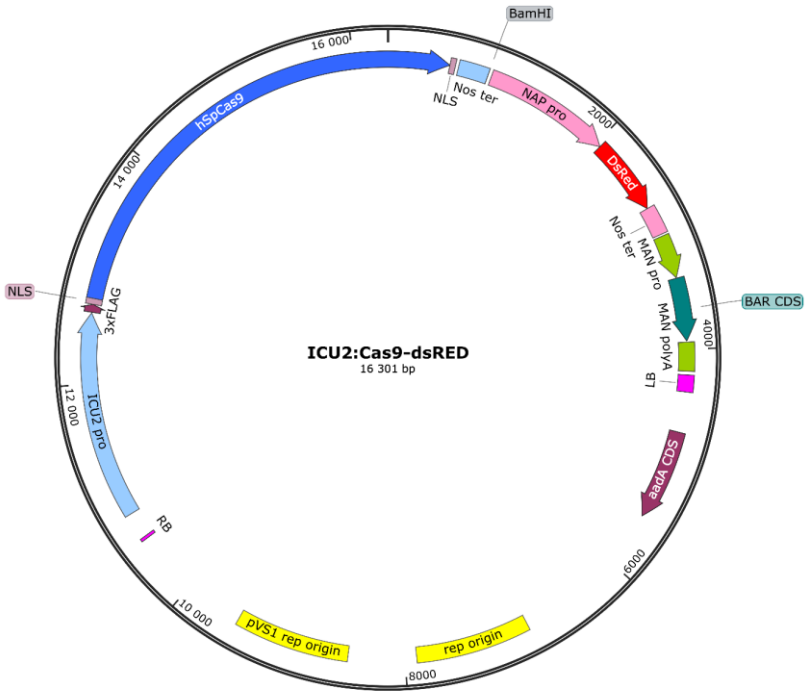


Figure 6. ICU2:Cas9-dsRED plasmid representative map.

3.6.2 Transformation of Arabidopsis

3.6.2.1 Transformation and culture preparation of *Agrobacterium tumefaciens*

Electrocompetent *A. tumefaciens* EHA105 cells were transformed in an Eporator (Eppendorf) at 2.5 kV using an electroporation cuvette with 0.1 cm gap, recovered in Luria Bertani broth (LB) without antibiotics for 2 hours at 26˚C, and then plated onto LB Agar plates containing rifampicin 20 μg ml^{-1} and kanamycin 50 μg ml^{-1} or spectinomycin 100 μg ml^{-1}. The plated cells were incubated at 26 °C for three days to obtain transformed colonies.

To prepare the *Agrobacterium* suspension for floral dip, the bacteria were grown in 300 ml LB broth supplemented with rifampicin 20 μg ml^{-1} and kanamycin 50 μg ml^{-1} or spectinomycin 100 μg ml^{-1} at 26 °C and 220 rpm for 18-22 h hours, until the OD_{600nm} was approximately 1.5. The bacteria were centrifuged at 4000 *g* and resuspended in a 5% fresh sucrose solution. Just before the floral dipping, Silwet-77 was added to a final concentration of 0.02% (vol/vol).

3.6.2.2 Floral dip transformation of Arabidopsis.

Arabidopsis genetic transformation was performed using the floral dip transformation method according to Clough and Bent, (1998). Briefly, Arabidopsis plants were grown under long day (16 h light / 8 h dark) conditions until bolting was observed, then the first bolt was clipped to induce the development of axillary buds and increase the number of flowers. After 5 to 7 days the new inflorescences were submerged in the *Agrobacterium* suspension for 10 s. Excess liquid was carefully removed using paper towels and the plants were covered with plastic trays and kept in darkness overnight, the following day these covers were removed. This inoculation procedure was repeated two or three times at intervals of one week in order to increase the number of transgenic plants obtained.

3.6.2.3. Selection of transgenic Arabidopsis plants and homozygous knock-out lines.

The plants transformed with pICU2:Cas9-dsRED were selected under a Zeiss AxioZoom V16 monoscope to detect expression of the red fluorescent protein in the seeds. When seeds were grown in soil the presence of the transgene was confirmed using PCR by amplifying a fragment of the T-DNA with the primers R2F1 and R1VF detailed in Table 6. PCR for genotyping was performed in a 20 µl reaction with the Phire Tissue Direct PCR Master Mix according to the manufacturer's instructions. To segregate the T-DNA and obtain homozygous knock-out lines, the non-fluorescent seeds collected from T1 plants were selected and sown on half strength MS media. After 12 days, pools of three plants were genotyped as described above, using primers flanking the deletion of each gene (Table 7), then the pools containing indications of polymorphisms in the amplified fragments were tested individually to select homozygous knock-out lines. The absence of the original T-DNA was confirmed via PCR with the primers R2F1 and R1VF (Table 6). The sequences of the targeted regions in the homozygous lines obtained were determined by Sanger sequencing.

Table 7. List of primers used for genotyping and sequencing putative knock-out lines in the candidate genes tested.

Primer Name	Sequence 5' → 3'	Target gene and use	Source
AT1G31540_F_seq	ACTTGTGTGGCCG TGATTAA	Genotyping and sequencing of *RAC1* putative knock-out lines	This manuscript
genot_CRISPR_1g31540_R	CACTTCCTCATCT GTTTGTCTCT		
RPB1a-F	ATGGAGACTGTCT CCGCCG	Genotyping and sequencing of *RPB1* putative knock-out lines	
Est-1-RPB1-R	ACGTCACCGGTGT TATTCTACA		
like4_FL_F	GACAATATAGCCA TGTCTTCC	Genotyping and sequencing of *RPB1-like-4* putative knock-out lines.	
like4_FL_R	AAAAACACATGAT TTTATGCTAT		

3.7 Transient expression in *Nicotiana tabacum*

The transient expression of RPB1 in *N. tabacum* cv. Xanthi leaves was performed according to Norkunas et al. (2018), adapted to our laboratory conditions. In brief, an *A. tumefaciens* EHA105 culture was prepared in 50 ml LB broth supplemented with spectinomycin 100 μg ml^{-1} and rifampicin 20 μg ml^{-1}. The flasks were incubated overnight at 26 ºC at 220 rpm and the cultures were centrifuged at 11000 g for 5 min and resuspended in MMA buffer (10 mM MES pH = 5.6, 10 mM $MgCl_2$, 200 μM acetosyringone) to an OD_{600} = 1.0 to prepare the final infiltration solution. The bacterial solution was infiltrated into the abaxial side of the leaves of five-week-old plants, using a 1 ml needless syringe.

3.8 RNA extraction, cDNA synthesis and RT-qPCR

To assess the expression of *RPB1*, *RPB1-like-4* and other host genes related to defense, RNA was extracted from whole roots and hypocotyls of *P. brassicae* infected and mock inoculated controls at 7 dpi. Tissue from six plants was combined and pulverized in liquid nitrogen using a mortar and pestle, and then the RNA was extracted using the InviTrap Spin Plant RNA Mini Kit (Invitek, Berlin, Germany) according to the manufacturer's instructions using the DCT buffer and β-mercaptoethanol. The RNA was quantified using a Nanodrop 2000 and 1 μg was treated with DNAse and used for cDNA synthesis. DNAse treatment was carried out with the TURBO DNAse kit following the manufacturer instructions. Prior to the cDNA synthesis PCR with the primers AtSK11_F and AtSK11_R (Table 5) was performed to confirm the absence of contaminating DNA. The first strand cDNA was synthesized with the M-MLV Reverse Transcriptase, RNase H Minus (Promega, Madison, Wisconsin) according to the manufacturer's protocol. The qPCR was performed as mentioned before in section 3.2, using as a template 4 μl of a 1:5 dilution of the cDNA. The primers used for each gene amplified are shown in Table 8.

Table 8. List of primers used for RT-qPCR of defense genes, reference genes, *RPB1* and *RPB1-like4* knock-out lines in the candidate genes tested.

Primer Name	Sequence 5' → 3'	Target gene and use	Reference
qRT_3G48140_F	TTGTTCGCTGCT ACCGGAGTTG	qPCR of *At3g48140*, codes for B12D protein with activity NADH-ubiquinone reductase) used as reference gene.	Reference gene with stable expression in clubroot infected tissue selected from RNA-Seq data and validated by NormFinder (Andersen et al., 2004)
qRT_3G48140_F	TTGTTCGCTGCT ACCGGAGTTG		
VAB1_fwd	TGGACATTGCTC CGTATCTTC	qPCR of *VAB1* (*At1g76030,* codes for a vacuolar ATPase (V-ATPase) used as reference gene.	(Walerowski et al., 2018)
VAB1_rev	TCGATAAGATAA CCTCCATTACCT C		
PR5_qPCR_F	GAGTGCCTGTGA GAGGTTTAAT	qPCR of *PR5* (*At1g75040*, *Pathogenesis-Related gene 5*)	This manuscript
PR5_qPCR_R	GTGCTCGTTTCG TCGTCATA		
AT5G10760_qPCR_F	CGGTGACATACC CGACGATT	qPCR of *AED1* (*At5g10760*, *Apoplastic, EDS1-Dependent 1*)	
AT5G10760_qPCR_R	CATTCCCTGCAA ACGCCAAA		
AT2G30770_qPCR_F	ACGATAAAGCGG ATTTCGTGGA	qPCR of *CYP71A13* (*At2g30770*, production of dihydrocamalexic acid (DHCA), the precursor to the defense-related compound camalexin)	
AT2G30770_qPCR_R	GAAGTTGTTGAC GTTCCTCCG		
RPB1_3'UTR_qPCR_F	TGAGTGTGAGAT GAAGGTTAATGT A	qPCR of *RPB1* covering the 3' untranslated region (UTR)	
RPB1_3'UTR_qPCR_R	TTCTTCCACATCA TCTAGTTACCAA		
RPB1a_qPCR_F	GGTTTAGTCCCA AGGCTCATT	qPCR of *RPB1*	
RPB1a_qPCR_R	GAGAGCCAGATA AACCAGAGAAG		
RPB1like4qPCRF	CCTCCGAGTCAC CGAAATAAA	qPCR of *RPB1-like-4*	
RPB1like4qPCRR	TGGGAGAAACAA TGGAGATGAG		

3.9 Light Microscopy

The tissue utilized for microscopy was collected in a fixative solution containing 0.5% glutaraldehyde and 2% paraformaldehyde in 1X PBS pH = 7.4 and incubated at 4 ˚C for between 3 and 6 days. Then the samples were transferred to a solution of 10% ethanol for at least 30 min, subsequently, the solution concentration was increased successively to 30%, 50%, 70% and absolute ethanol in intervals of 30 min and then preserved in absolute ethanol for at least 2 days more. Afterwards, the tissue was preinfiltrated with a solution of 1:1 ethanol/Technovit 7100 (v/v) and ultimately infiltrated with 100% Technovit 7100 (Kulzer GmbH, Wehrheim, Germany) (Stefanowicz et al., 2021). The sectioning of the embedded samples was performed on a Leica RM2135 microtome (Leica) at a thickness of 5 µm and transferred to glass slides with a drop of distilled water and dried slowly on a heating plate at 60 ˚C. The photographs were taken under an AXIO Image M2 microscope (Zeiss) coupled to an AxioCamICc5 camera.

3.10 Transcriptional profiling

3.10.1 Experimental design

Each biological replicate consisted of 15 plants, for each of four genotype / treatment combinations consisting of the Col-0 and Pro-0 accessions inoculated with *P. brassicae* pathotype P1B or mock inoculated with sterilized distilled water. Four independent experiments were performed and following RNA extraction (as previously described in the section 3.8) the three experiments with the best quality of RNA integrity, as assessed by Experion RNA StdSens Analysis Kit, were selected for transcriptome profiling.

3.10.2 RNA-Seq using Illumina technology

RNA samples were sent to Novogene UK (Cambridge) for stranded cDNA library preparation and mRNA sequencing of 20M paired-end reads with a length of 150 base pairs (Novaseq PE150) per sample. The quality control of the samples was done with an Agilent 2100 device. The library construction and quality control of the sequencing results was executed in three steps:

1. The distribution of the sequencing quality by calculating the error rate and the sequencing base quality (Q_{phred}).
2. The error rate distribution along the reads.
3. The distribution of AT and GC content.

3.10.3 Mapping, differential gene expression analyses and gene ontology.

To elaborate the mapping and the count matrix, prior to the differential gene expression analysis, the FASTQ files were uploaded to the Galaxy Europe server (https://usegalaxy.eu/) and the following workflow was followed:

1. Cutadapt 1.16 with Python 3.6.4 to check and remove adapter contamination (Martin, 2011)
2. HISAT2 for mapping the reads to the *A. thaliana* TAIR10 and the *P. brassicae* Pbe3 genomes (Kim et al., 2015).
3. featureCounts to obtain the matrix counts (Liao et al., 2014).

The graphical report of each of the previously mentioned steps was generated with the MultiQC package (Ewels et al., 2016).

The matrix counts were downloaded from Galaxy and differential gene expression analysis was performed using the DESeq2 package (Love et al., 2014) in R version 4.0.3 (2020-10-10) managed by RStudio Version 1.4.1103.

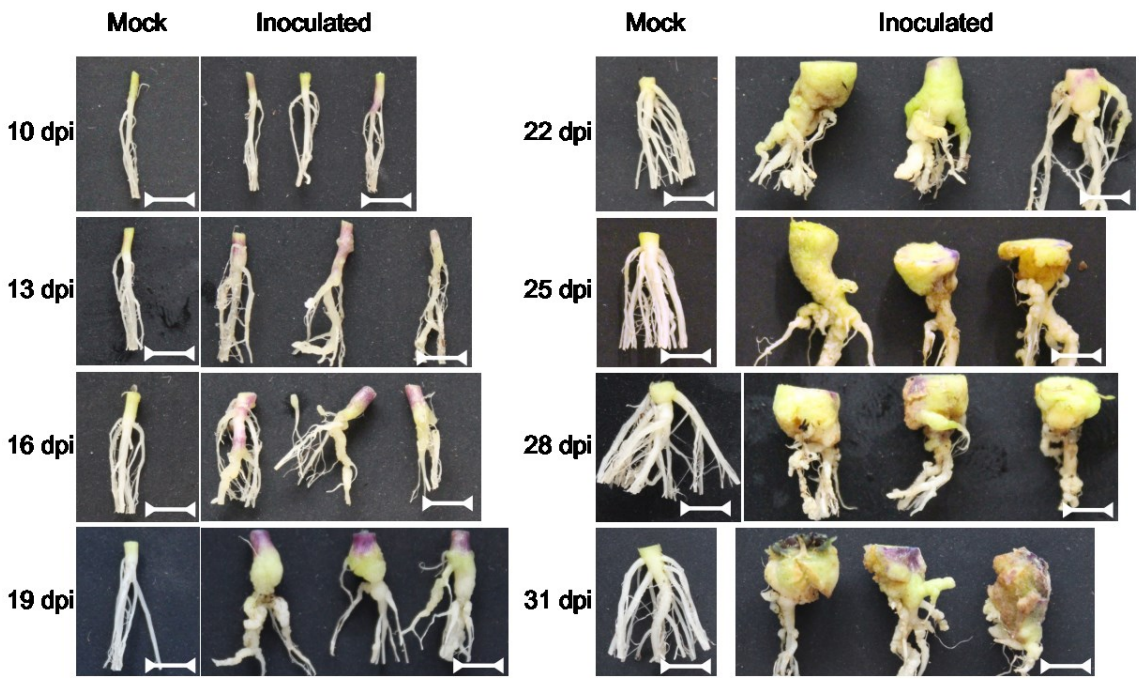
Gene ontology analysis was performed according to Bonnot et al., (2019), in brief the list of differentially expressed genes was used as an input to the PANTHER website (http://go.pantherdb.org/) to obtain the enrichment analysis of gene ontology (GO) terms, later the representative terms were selected with REVIGO (http://revigo.irb.hr/) to remove redundancy, and finally the plots including the names of the biological processes, the false discovery rate (FDR) p-value, and the number of genes were elaborated in R using the ggplot2 package (https://ggplot2.tidyverse.org/).

4. Results

4.1 *P. brassicae* growth reaches a stationary phase at 19 dpi.

To develop a reproducible way to quantify *P. brassicae*, qPCR was used to measure the pathogen DNA amount relative to the plant DNA at different time points following inoculation. The parameters for this method were established in the highly clubroot susceptible Arabidopsis accession Col-0. Monitoring the pathogen/plant DNA ratio in tissue harvested from the hypocotyl and upper first centimeter of the root at three day intervals from 10 dpi until 31 dpi, an exponential growth curve was observed up to 19 dpi, when a stationary phase began, and the relative DNA amounts remained constant (Figure 7). From these results the19 dpi time point was selected as more advantageous as a collection time to be used for the other Arabidopsis accessions due to the amount of tissue that can be harvested. The later time points were not chosen, to avoid excess starch that could reduce the quality of the extracted DNA.

A)



B)

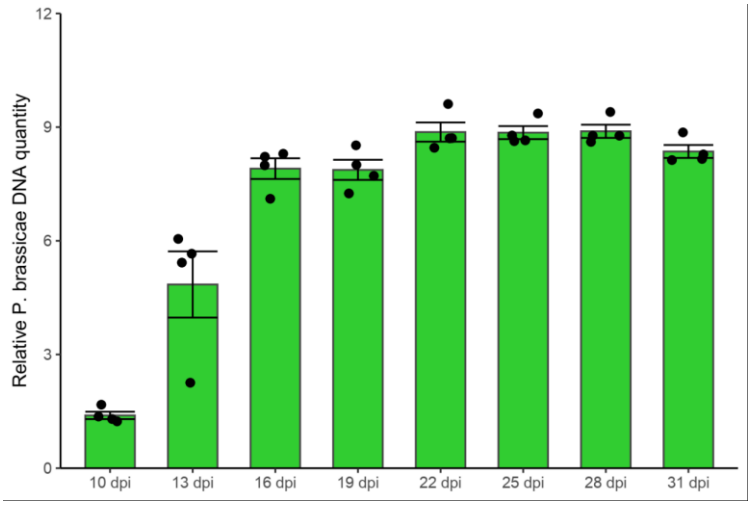


Figure 7. *P. brassicae* growth curve from 10 dpi to 31 dpi in the susceptible accession Col-0.
A) Gall morphology observed in the different time points sampled. The scale bar corresponds to 500 µM. **B)** Pathogen relative DNA quantity measured with qPCR (Log_2 scale). The error bars correspond to the standard error of 4 biological replicates, each one representing a mixture of 3 plants.

4.2 Relative quantification of *P. brassicae* DNA levels *in planta* are positively correlated with observed disease symptoms

To determine the level of resistance or susceptibility to the *P. brassicae* pathotype P1B, the relative pathogen DNA amount of 142 Arabidopsis natural inbred lines was quantified at 19 dpi and the symptoms were also scored according to the DI scale. Different degrees of resistance/susceptibility were apparent across the whole collection, from the strong resistance observed in accessions such as Est-1 or Uod-1, to the highly susceptible accessions Col-0 and Ws-0. There is a positive correlation between both the phenotype measurements as determined by the Spearman correlation coefficient (rho 0.57, p-value = 1.45e-13), showing that the quantification of the pathogen DNA is congruent with the symptoms observed in plant roots and hypocotyls (Figure 8). Based on these results 13 accessions appeared to be robustly clubroot resistant. In the accession Pro-0, there was observed a particularly high pathogen DNA amount compared to the apparent disease symptoms and gall development, these findings are elaborated on in Section 4.7.

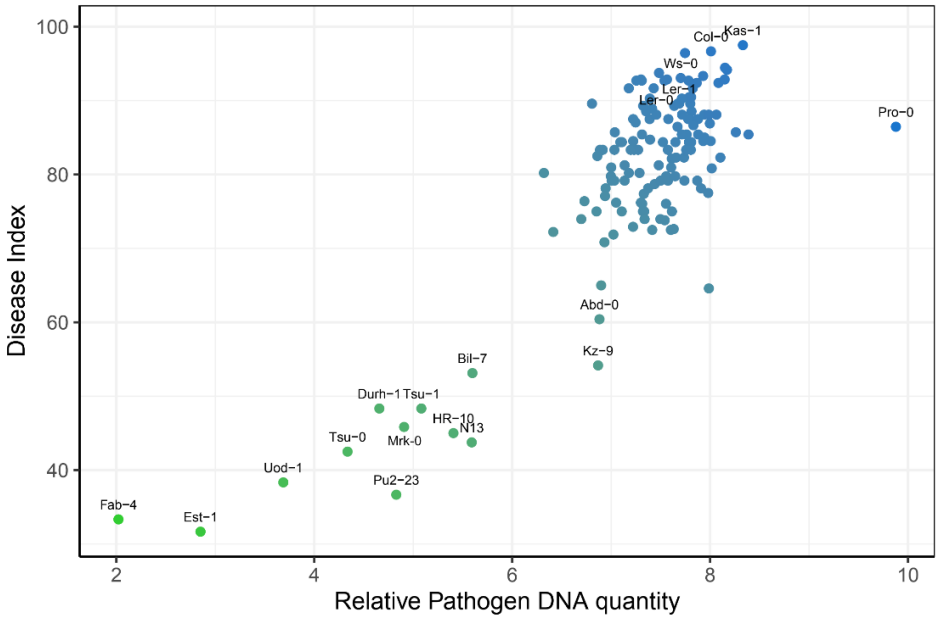


Figure 8. Disease index scoring and relative pathogen DNA quantification in 142 Arabidopsis natural inbred lines.
The x axis corresponds of the differences of the CP values of *Pb18S* and *AtSK11* on a Log_2 scale. Each qPCR value represents the mean of between 4 and 10 replicates each consisting of 3 galls and the corresponding disease index, calculated for 12 to 30 plants per accession.

4.3 Identification of candidate loci involved in the resistance or susceptibility to *P. brassicae* through GWAS

Prior to performing a GWA analysis, the Shapiro-Wilk test was used to determine if the data fitted the normal distribution, because this is a prerequisite for parametric mixed models used to calculate the associations. For the mean relative pathogen DNA amount the data required a Box-Cox transformation and the exclusion of the Pro-0 outlier accession with the purpose of adjusting it to the normal distribution (p-value = 0.18). However, in the case of the disease index scoring, it was not possible to fit the data to the normal distribution with any of the transformation methods tested (p-value = 1.41e-12) (Figure 9).

To identify the genomic regions associated to the resistance to *P. brassicae*, two online tools were used: GWA-portal and easyGWAS. In the GWA-portal there is available the genomic data from the 1001 genomes consortium, that includes 1135 accessions and 10 million SNPs as well as the imputed genome sequence, that combines the 1001 genomes polymorphisms and the 250k SNP chip datasets and contains 2029 accessions (Seren, 2018). In comparison, the easyGWAS website only contains the 1001 genomes dataset (Grimm et al., 2017). In our collection, 118 accessions were included in the 1001 genomes collection and 141 in the imputed full sequence dataset.

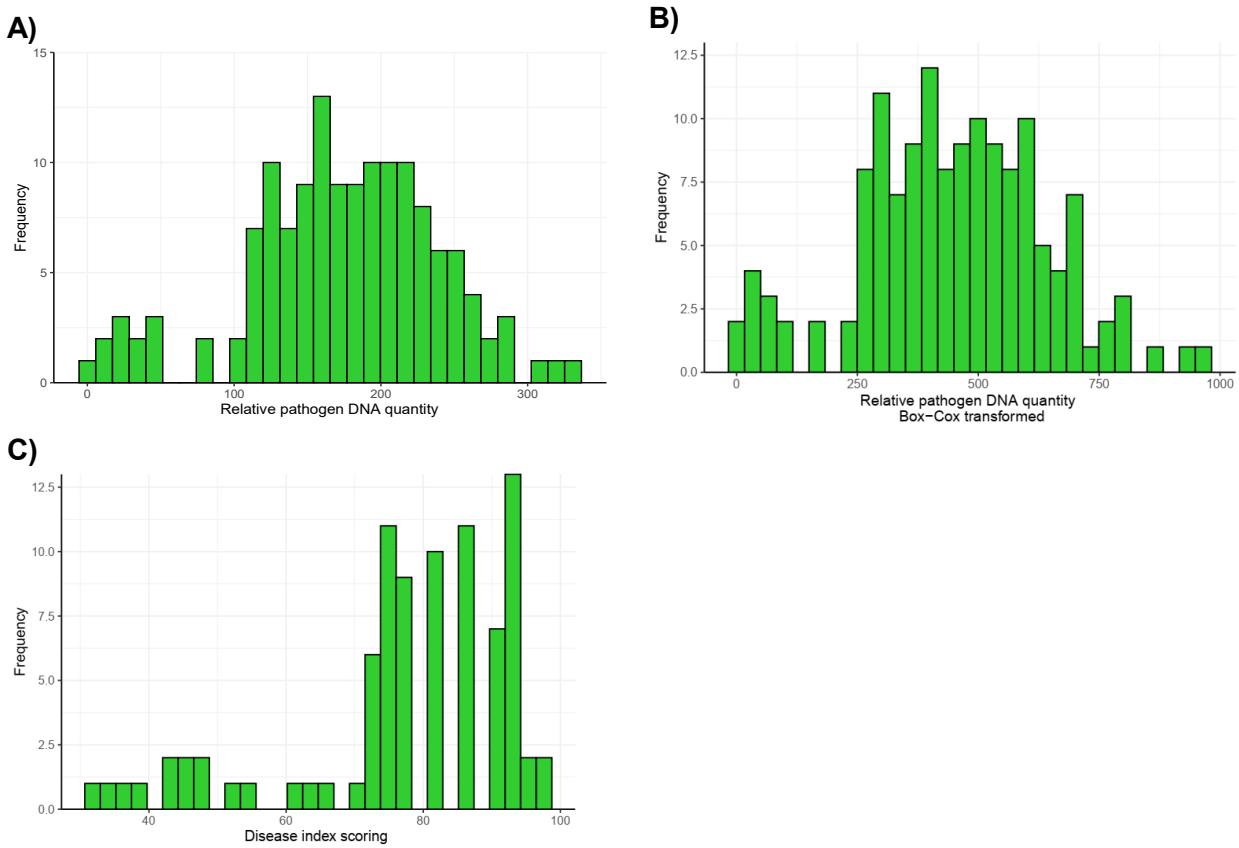


Figure 9. Histograms of the phenotypic characterization of the population to visualize the adjustment to the normal distribution.
A) Distribution of pathogen relative DNA amount **B)** Distribution of pathogen relative DNA amount, Box-Cox transformed **B)** Distribution of the disease index scores

The GWAS for the relative pathogen DNA quantity was performed using the accelerated mixed model (AMM) with the imputed full sequence genotype data, Box-Cox transformation, and the Bonferroni correction for multiple testing. In the case of the disease index scoring the non-parametric Kruskal Wallis test was used. The Manhattan plot generated from the first analysis using the relative pathogen DNA amount showed a significant association in one SNP located on the coding sequence (CDS) of the gene *At1g32030* (- *log10* p-value = 9.52) (Figure 10B, C). In the case of the GWAS using the DI data, there were no significant associations that surpassed the somewhat stringent Bonferroni correction, nevertheless, the SNP with the highest - *log10* p-value (6.69) was found between the genes *At1g32090* and *At1g32100* and the SNP previously found in *At1g32030* also showed a - *log10* p-value = 6.02, that is relatively high compared to the rest of the SNPs (Figure 10A). The SNP in the gene *At1g32030* corresponds to an A → T substitution that is present in 10 out of 13 accession that were considered resistant (Figure

10E). These association analyses showed that the region between the genes *At1g32030* and *At1g32100* might be associated with resistance to the P1B pathotype of *P. brassicae*.

Additionally, GWAS analysis was performed using exclusively the 1001 genomes genotypic data in both GWA-portal and easyGWAS to compare against the results previously obtained. Unlike the GWA-portal, easyGWAS uses the EMMAX mixed model algorithm to calculate associations. Both phenotypic datasets including just the 118 accessions contained in the 1001 genomes set were subjected to normality tests, and it was possible to determine that both fit the normal distribution after a Box-Cox transformation according to the Shapiro-Wilk normality test (qPCR p-value = 0.46, DI p-value = 0.16).

When performing the association analysis in the GWA-portal using the 1001 genomes genotype data and the AMM algorithm, no significant associations were detected with either phenotype, nevertheless, with the DI, three SNPs had associations with adjusted p-values much closer to threshold for significance following Bonferroni correction (Figure 11). The highest (- *log10* p-value = 6.62) is in chromosome 4 adjacent to the 5' regions of the genes *At4g05070* and *At4g05071*, the first one, also known as Wound-Induced Polypeptide 2 (*WIP2*), positively regulates Arabidopsis resistance to *P. syringae* pv. tomato DC3000, and the second codes for a small protein with unknown function (Yu et al., 2018). The second SNP with highest - *log10* p-value corresponds to the same found previously in the region between *At1g32030* and *At1g32100*, and the third one is present in the coding sequence of the gene *RAC1 (Resistance to Albugo candida 1, At1g31540*), that codes for a protein containing the TIR-NB-LRR domains typically found in resistance genes, and was previously found to confer resistance to *Albugo candida* (Borhan et al., 2004).

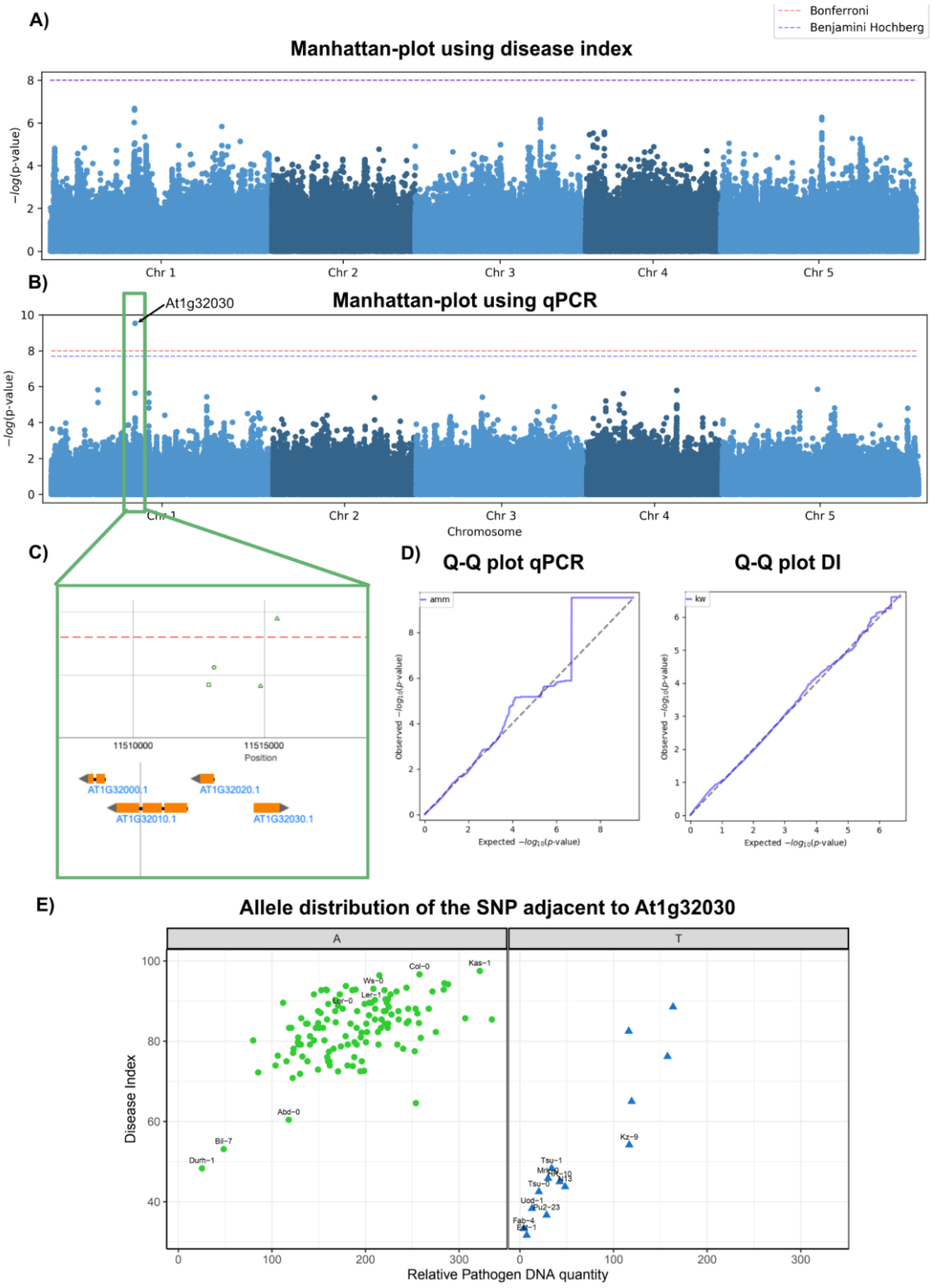


Figure 10. GWAS using the imputed full sequence data.
A) Manhattan plot with DI phenotype as input, **B)** Manhattan plot with relative pathogen DNA quantity as input. **C)** Detail of significantly associated SNP in *At1g32030*, **D)** Corresponding Quantile-Quantile plots of the GWAS, **E)** Allele distribution of genotypes in the *At1g32030* SNP by relative pathogen DNA quantity and DI.

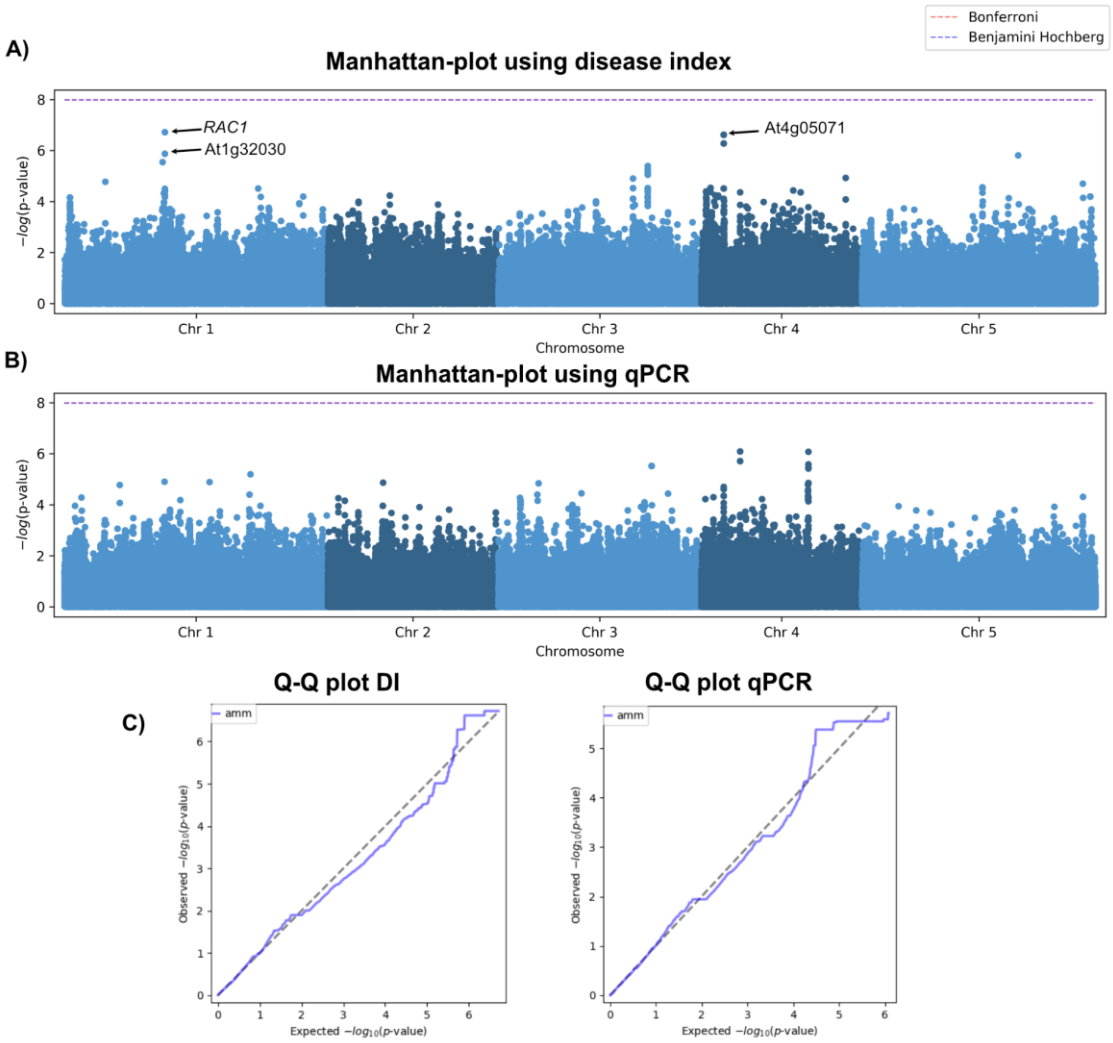


Figure 11. GWAS using the 1001 genomes data in the GWA-portal website.
A) Manhattan plot with DI phenotype as input, **B)** Manhattan plot with relative pathogen DNA quantity as input, **C)** Corresponding Quantile-Quantile plots of the GWAS

The GWAS analysis of the relative pathogen DNA quantity in the easyGWAS website did not result in any significant associations, however, the SNPs with the highest - *log10* p-value (5.72) were located between the genes *At1g32020* and *At1g32100*, as found in previous analyses. In the case of the GWAS of the DI there was one significant association in the SNP present in the CDS of *RAC1* (- *log10* p-value = 7.56) and one association close to the significance threshold in two SNPs adjacent to *At1g32020* (- *log10* p-value = 7.07) (Figure 12 A-C). The SNPs associated to *RAC1* and *At4g05071* were present in 6 and 8 out of 9 resistant accessions, but also in some other accessions with low or intermediate susceptibility (Figure 12 D). Based the results obtained in all the GWAS, two loci were

chosen that were significant in at least one of the analyses: the region between *At1g32020* and *At1g32100*, and the *RAC1* gene.

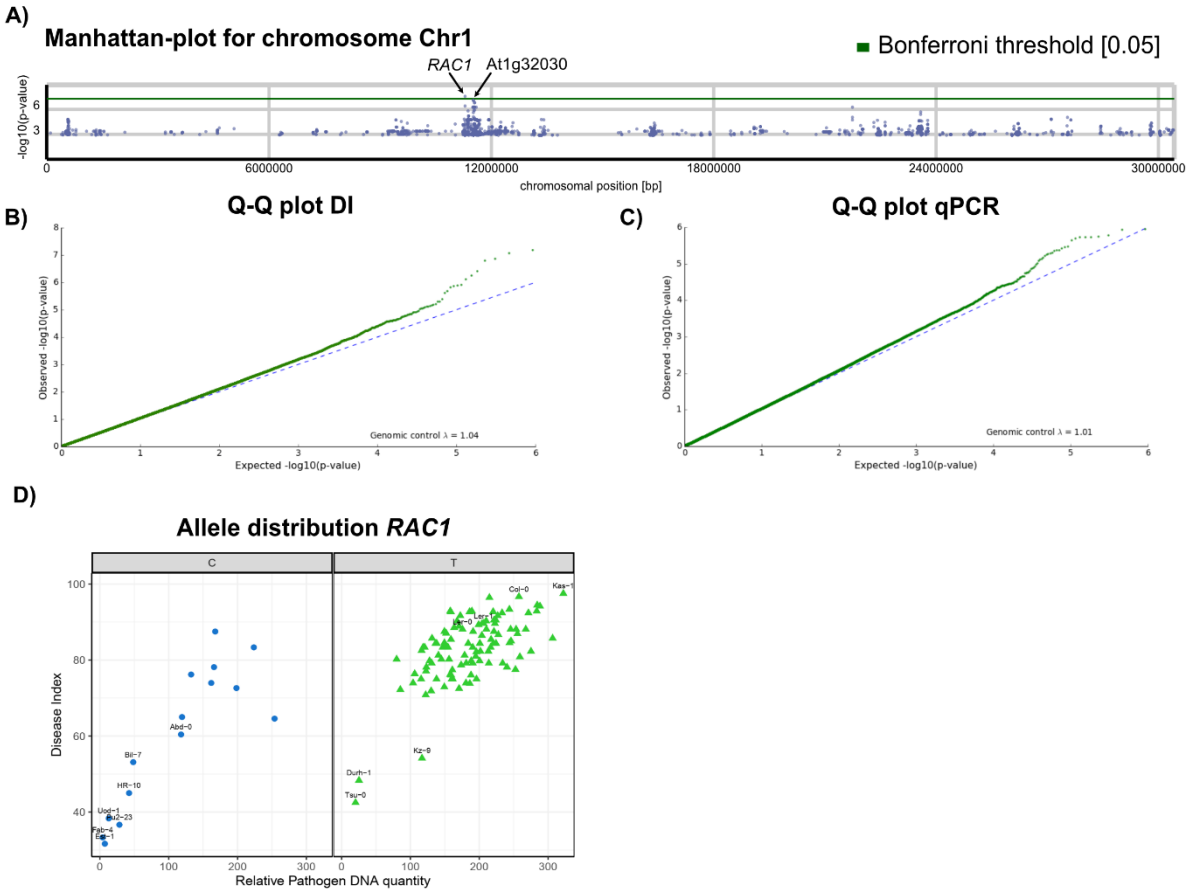


Figure 12. GWAS using the 1001 genomes data in the easyGWAS website.
A) Manhattan plot using the DI phenotype data in the chromosome 1, **B), C),** Quantile-Quantile plots of the GWAS using DI or qPCR respectively, **E)** Allele distribution of the SNP present in the *RAC1* CDS.

4.4 SNPs associated with resistance between *At1g32020* and *At1g32100* colocalize with the previously identified *RPB1* locus

The *RPB1* locus was first identified in a biparental mapping population from the Cvi-0 X Tsu-0 crossing and the region has been refined to one containing 13 ORF (Fuchs and Sacristán, 1996; Arbeiter et al., 2002). Unfortunately, some information regarding the cloning and characterization of *RPB1* has not been published in peer reviewed journals, but conference reports and review papers have reported that it is present in two almost identical copies called *RPB1a* and *RPB1b* separated by about 5 kb in the Tsu-0 accession, which

contains additionally a similar CDS to *RPB1* called *RPB1-like-1* (GenBank accession number FJ807885.1). An additional sequence that belongs to the Arabidopsis accession RLD is available in the GenBank sequence repository under the accession number FN400762 and contains a single copy of the *RPB1* gene and three similar open reading frames (ORFs) called *RPB1-like-2, -3* and *-4*, but there is no information about their functionality (Figure 13). The *RPB1* gene codes for a 148 aa protein with predicted transmembrane domains, but it lacks of homologues with known functionality.

In the region between the genes *At1g32020* and *At1g32050* in Col-0, there is no functional ORF for *RPB1,* but the sequence of a pseudogene identified as *At1g32049* with high similarity to *RPB1*, also in the region are present *At1g32040* and *At1g32045* that are predicted transposable elements (TE). The *RPB1* loci, broadly defined as the region of ~12 kb downstream of *At1g32020,* can be compared with other Arabidopsis genomes with chromosome-level reference-quality annotations including An-1, C24, Cvi-0, Eri-1, Kyoto, Ler, Ler-0, Shahdara, Cmd-0 , Ty-1, Kn-0 and KBS Mac-74 (Michael et al., 2018; Jiao and Schneeberger, 2020). Across the region in the various accessions, all of them contain a CDS coding for *RPB1* except for Col-0, the presence of transposable elements, indels and *RPB1-like* genes are also observed and the accessions Shahdara and Cmd-0 contain two copies of *RPB1*, as was the case for Tsu-0. Interestingly, despite the fact that the region is very variable, when examining the amino acid sequence, the RPB1 protein in all the accessions analyzed share more than 99% of identity and similarity, compared to the protein sequence in the accession RLD (Figure 14). Upon inspecting the promoter region (600 bp upstream) of the *RPB1a* and *RPB1b* genes there was a high degree of sequence variability, except when comparing the *RPB1a* promoter of the resistant accessions Tsu-0 and RLD where the sequences had 99.67% identity.

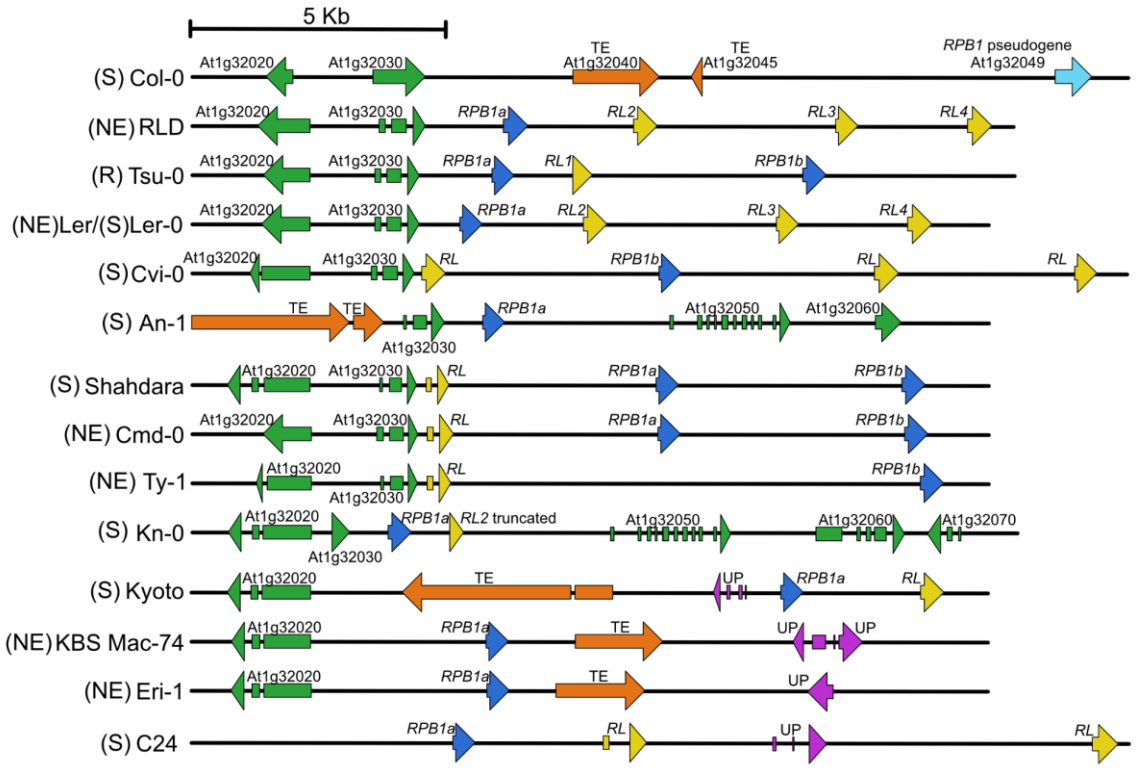


Figure 13. Region downstream *At1g32020* in Arabidopsis accessions with chromosome level assembly genome data.
RL (*RPB1-like* genes), UP (Unknown protein), The reaction to the P1B pathotype is shown in parenthesis: (S) susceptible, (R) resistant, (NE) Not evaluated.

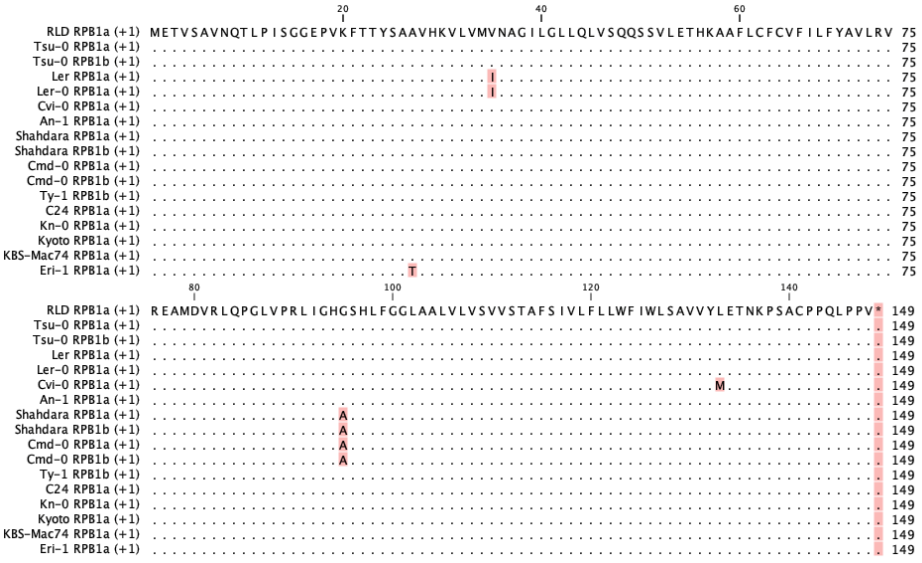


Figure 14. RPB1 protein alignment in Arabidopsis accessions with chromosome level assembly genome data.

To assess whether *RPB1* could be present or absent in the genomes of other accessions used in the GWAS a fragment of the gene was amplified by PCR using primers RPB1a_qPCR_F and RPB1a_qPCR_R (Table 8), in 123 accessions a product was found indicating the probable presence of a CDS for *RPB1* and only 19 of them were negative (Figure 16). All the accessions potentially lacking *RPB1* were susceptible to clubroot, but it is not possible to conclude that the absence of *RPB1* is the cause of the susceptibility phenotype.

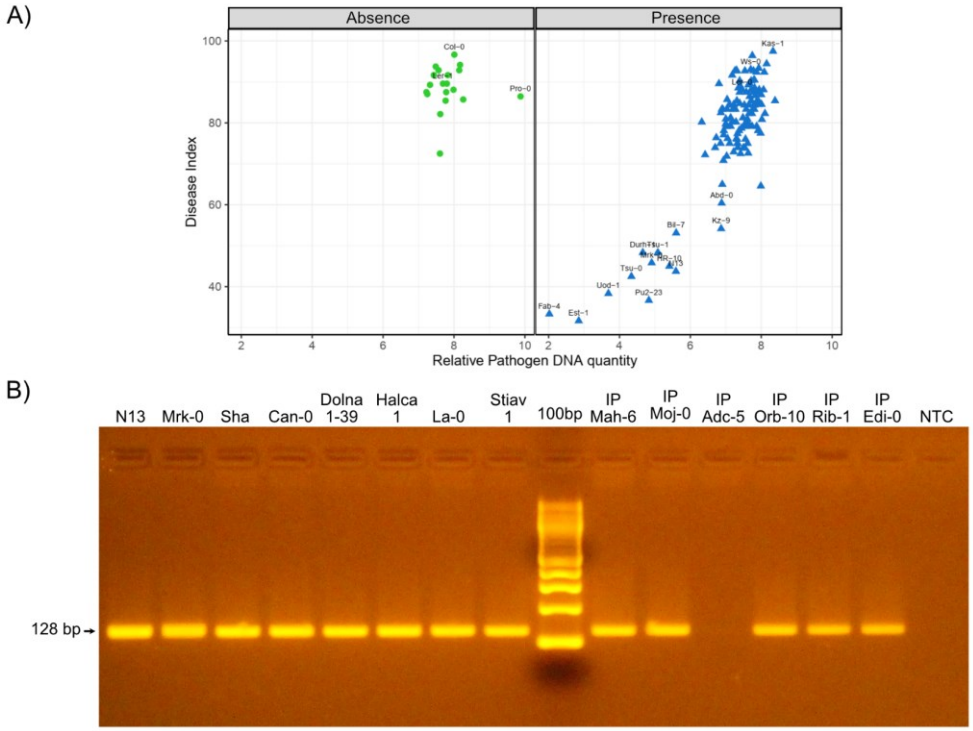
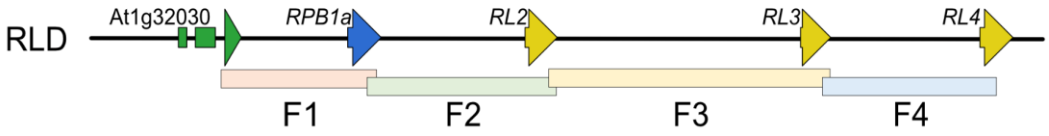


Figure 15 Presence of *RPB1* in 142 Arabidopsis accessions.
A) Distribution of presence/absence of *RPB1* according to qPCR pathogen quantification and DI scoring. B) Representative agarose gel electrophoresis of the 128 bp amplified fragment of *RPB1.* NTC (No Template Control)

To confirm the hypothesis that *RPB1* plays an important role in resistance to clubroot, the approach selected was to generate null mutations with CRISPR/Cas9 targeting of *RPB1* in resistant accessions to determine its function genetically. The first step was to evaluate whether the resistant accessions had DNA sequences similar to the RLD version of the region with just one copy of *RPB1* or whether, like Tsu-0, they had multiple copies. Primers were designed to amplify four fragments in the region downstream of *At1g32030* covering the *RPB1* and *RPB1-like* genes in all 13 resistant accessions. Only Est-1 and Uod-1

amplified all four of the fragments with the expected size as the RLD sequence; eight had a pattern of amplicon sizes similar to Tsu-0, and two presented a pattern that did not match with any of the versions (Figure 16). To confirm this finding approximately 11 kb of the region of interest was sequenced in both Uod-1 and Est-1 using the Sanger method. Sequence identities of 99.96% in Est-1 and 97.41% in Uod-1 were found when referenced to RLD. For the gene coding sequences there was 100% identity in the CDS of *RPB1a* and *RPB1-like-2, -3* and *-4*.



Fragment	F1	F2	F3	F4	Similar to
Expected size (bp)	2226	2584	3807	2469	
Col-0	NO	NO	UNS	NO	
Uod-1	2.2kb	2.5kb	3.5kb	2.5kb	RLD
Est-1	2.2kb	2.5kb	3.5kb	2.5kb	
Pu2-23	2.2kb	3.4kb	NO	NO	
HR-10	2.2kb	3.4kb	NO	NO	
Kz-9	2.2kb	3.4kb	NO	NO	
N13	2.3kb	3.4kb	NO	NO	
Mrk-0	2.2kb	3.4kb	NO	NO	Tsu-0
Tsu-1	2.2kb	3.4kb	NO	NO	
Cnt-1	2.2kb	3.4kb	NO	NO	
Duhr-1	2.2kb	3.4kb	NO	NO	
Tsu-0	2.2kb	3.4kb	NO	NO	
Fab-4	2.2kb	3.4kb	3.5 kb	NO	ND
Bil-7	0.9kb	NO	4kb (weak)	2.5kb	

Figure 16. Amplified fragments of the *RPB1* locus based on the RLD allele in the 13 resistant Arabidopsis accessions using conventional PCR.
UNS, unspecific amplification, NO, no amplification.

To develop additional evidence indicating that *RPB1* or any of the *RPB1-like* genes might be involved in resistance to clubroot disease, the expression of these genes was profiled in response to infection. Following preliminary time-course experiments, 7 dpi was selected for harvesting the roots of Est-1, comparing mock controls with plants inoculated with the *P. brassicae* P1B pathotype. *RPB1-like-2* and *RPB1-like-3* were not expressed in any of the samples, however, the expression of *RPB1* was strongly upregulated (one-tailed

T-test p-value = 0.001), and the expression of *RPB1-like-4* was also significantly upregulated in the infected plants (one-tailed T-test p-value = 0.031) though the magnitude was less than that observed for *RPB1* (Figure 17).

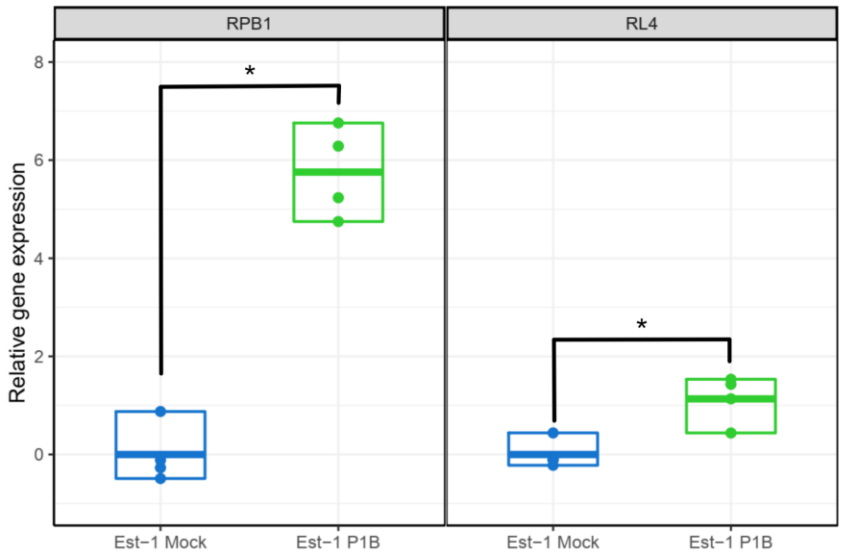


Figure 17. Relative gene expression of *RPB1* and *RPB1-like-4* in Est-1 at 7 dpi. Asterisks indicate significant differences using a T-Student's test $p < 0.05$, the relative gene expression is presented in the $\log_2$ scale. Prior to the T-Student's test. The boxes depict the range and the mean of each group of data (n = 4) and biological replicates are derived from the roots and hypocotyl tissue of 10 individual plants.

4.5 Knock-out lines of *RPB1* are completely susceptibility to *P. brassicae* P1B

4.5.1 Development and genotyping of *RPB1* and *RPB1-like-4* knock-out lines

To determine the role of *RPB1* and *RPB1-like-4* in the resistance to *P. brassicae* pathotype P1B, knock-out mutants in both genes were generated individually and in combination through the use of CRISPR/Cas9 gene editing technology. For each target two guide RNAs were designed, separated by 352 bp for *RPB1* and 316 bp for *RPB1-like-4* with the intention of creating large deletions that could facilitate the genotyping of the putative edited plants. The construct containing the gRNAs and the *SpCas9* gene was transformed into the Uod-1 and Est-1 accessions and after selecting for transformants with possible mutations in subsequent T-DNA-free generations the putative knock-out lines were genotyped using conventional PCR to identify homozygous lines (Figure 18 and Figure 19).

PCR amplification of the full length *RPB1* CDS led to the identification of several homozygous and heterozygous lines with the expected deletion of 372-373 bp, for example in Uod-1 *rpb1* L33 and L69, though in other cases smaller deletions were observed, for instance in the Est-1 *rpb1* L127 line, which had an insertion of 1bp in position 46 and a deletion of 41 bp at position 383 (Figure 18A, B). In all the sequenced samples the effect of the introduced mutations was predicted to have severe changes in the potential expression of *RPB1* such as truncations of more than 75% of the amino acids of the final protein or frame shifts that completely changed the predicted amino acid sequence (Figure 18C).

The putative *RPB1-like-4* knock-out lines were genotyped in the same way as the *RPB1* knock-out lines with primers amplifying the whole *RPB1-like-4* CDS. The results of the genotyping identified homozygous lines with different size of deletions, including some with the expected deletion of 350 base pairs in the Est-1 *rpb1-like-4* L42 and smaller or multiple deletions such as in the Uod-1 lines *rpb1-like-4* L39 and L59 (Figure 19A, B). It was also possible to identify a double knock-out in the Uod-1 genotype that had a 372 bp deletion in *RPB1* and a single nucleotide deletion at position 14 of *RPB1-like-4* that creates a frame shift in the predicted protein (Figure 18 and Figure 19). In all the mutants elaborated it was observed that the predicted amino acid sequence had changes that would drastically affect the function of the RPB1-like-4 protein (Figure 19C).

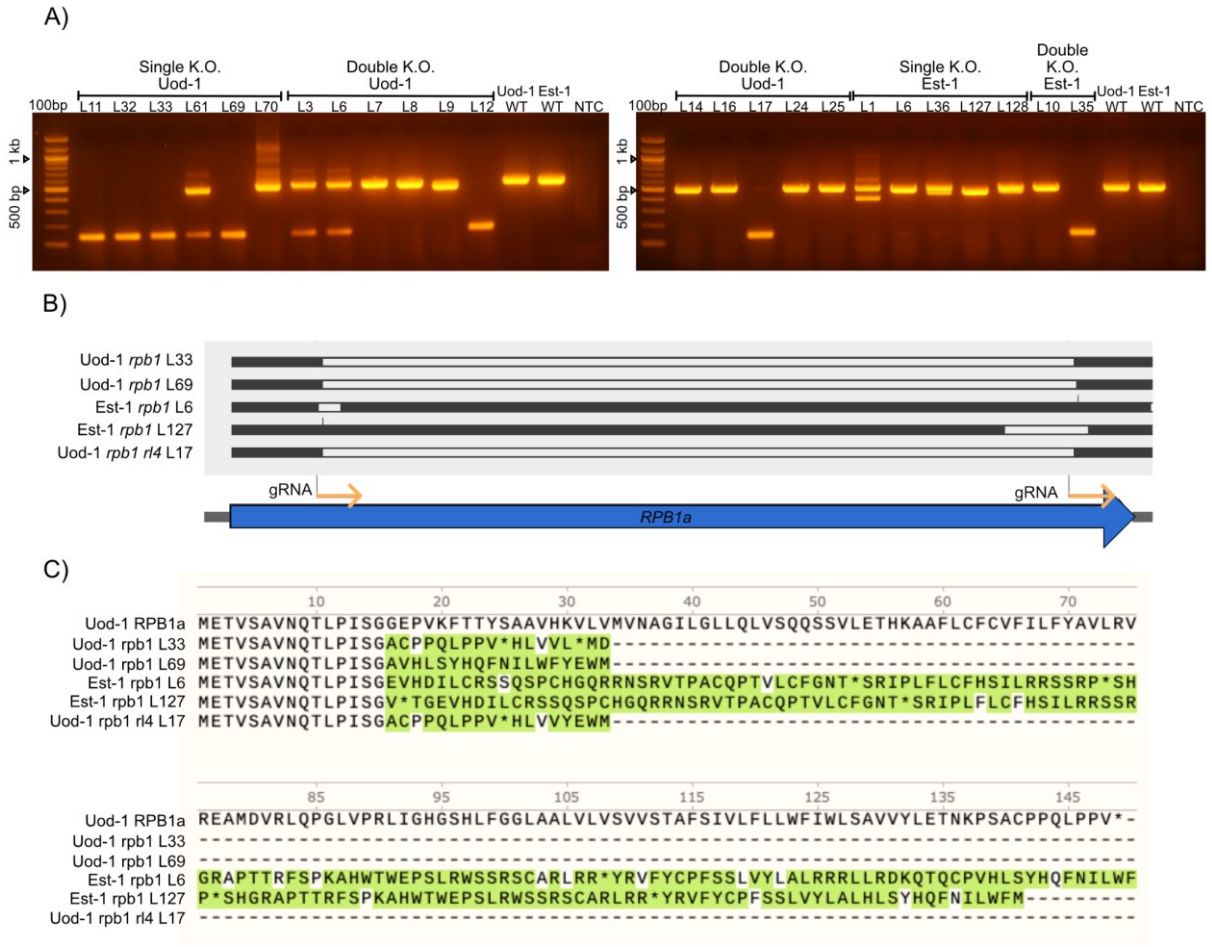


Figure 18. Genotyping of *RPB1* knock-out lines with conventional PCR and Sanger sequencing.
A) Representative agarose electrophoresis of the PCR products of the full length *RPB1* CDS in the K.O. lines and WT controls of Uod-1 and Est-1. B) Representation of the deletions observed in the *RPB1* K.O. lines identified through Sanger sequencing, the black bars correspond to the portion of the sequence aligned to *RPB1* and the empty bars correspond with the deletion observed in selected lines, the orange arrows show the positions of the gRNAs. C) Predicted amino acid sequence in the different K.O. lines compared to the RPB1 protein sequence; the green highlighted letters correspond to changes in the amino acid sequence.

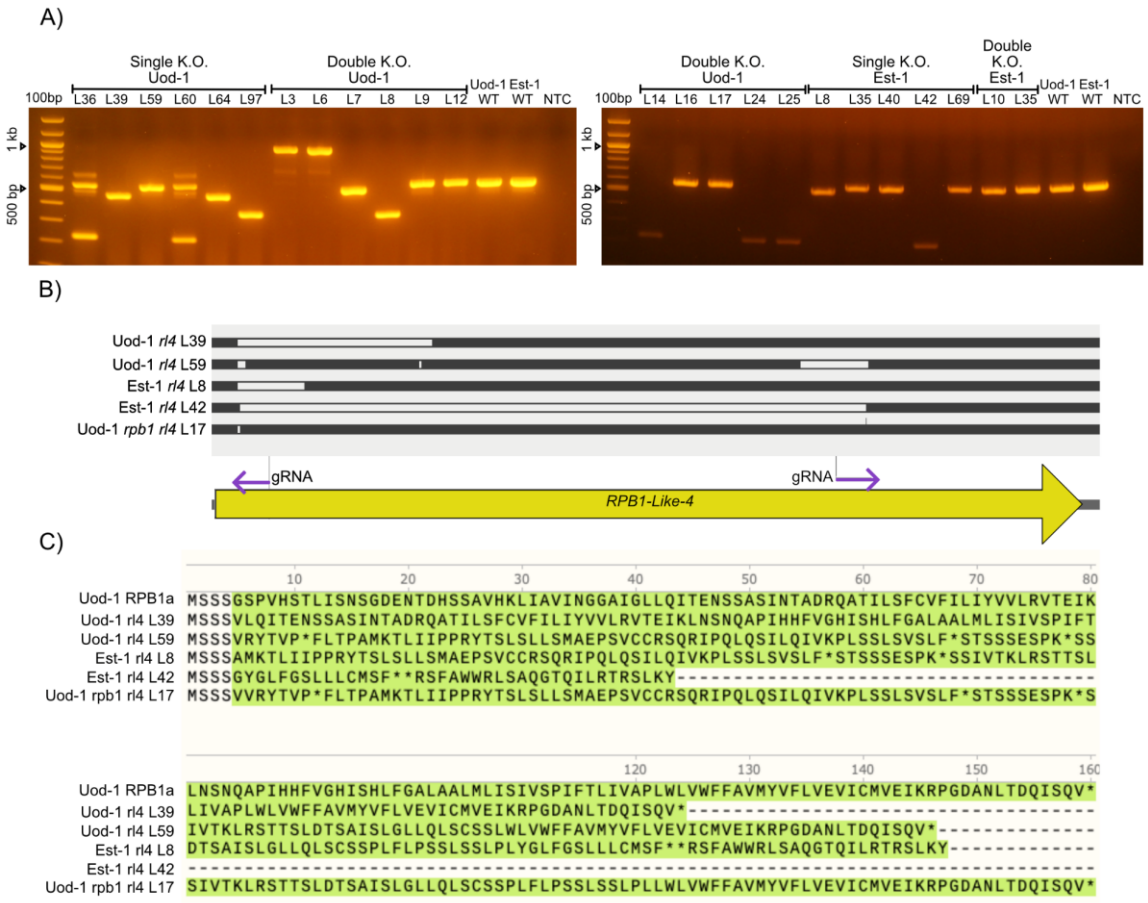


Figure 19. Genotyping of *RPB1-like-4* knock-out lines with conventional PCR and Sanger sequencing.
A) Representative agarose electrophoresis of the PCR products of the full length *RPB1-like-4* CDS in the K.O. lines and WT controls of Uod-1 and Est-1. B) Representation of the deletions identified in the *rpb1-like-4* K.O. lines through Sanger sequencing, the black bars correspond to the portion of the sequence aligned to *RPB1-like-4* and the empty bars correspond with the deletion observed in selected lines. the purple arrows show the positions of the gRNAs. C) Predicted amino acid sequence in the different K.O. lines compared to the RPB1-like-4 protein sequence; the green highlighted letters correspond to changes in the amino acid sequence.

4.5.2 Knock-out lines of *RPB1*, but not of *RPB1-like-4* exhibit susceptibility to *P. brassicae*

To establish if the knock-out lines of *RPB1* and *RPB1-like-4* retained resistance to clubroot or were rendered susceptible to the pathogen, relative pathogen amount was assayed by qPCR and the DI scores were determined for two lines per genotype per gene alongside wild-type Est-1 and Uod-1 and the susceptible control Col-0. At 19 days post-inoculation it was apparent that the *rpb1* knock-out lines had similar symptoms of clubroot

disease in the above-ground tissue when compared to Col-0, with severe wilting in the rosette and purple discoloration from the production of stress associated secondary metabolites in the leaves alongside some chlorosis. Whereas, in the wild-type Uod-1 and Est-1 controls, the inoculated plants maintain generally healthy rosettes (Figure 20). In the roots and hypocotyls of Uod-1 *rpb1* and Est-1 *rpb1* lines typical symptoms of clubroot disease were observed entirely comparable to the susceptible accessions in terms of the gall development, deformation of the main root and dramatic swelling of the hypocotyl. In the wild-type controls of Uod-1 and Est-1 there was no evident root malformation or increase in size of the hypocotyls, though the roots sometimes exhibited darkened patches probably linked to programmed cell death and lignification that are associated with the resistance to *P. brassicae* (Figure 20). In contrast to the obvious differences in the *rpb1* lines, the *rpb1-like-4* knock-out lines in either resistant background did not show any evident difference in the morphology or coloration of the rosettes, roots or hypocotyls compared with the corresponding wild-type controls (Figure 21). The Uod-1 *rpb1 rpb1-like-4* double mutant had similar symptoms in both the above and below-ground parts of the plants as the single *rpb1* knock-out lines and were susceptible to *P. brassicae* infection without any apparent exaggeration or acceleration of disease beyond the single *rpb1* mutations or the susceptible control Col-0 (Figure 21). Quantifying *P. brassicae* levels with qPCR confirmed these observations, as the relative pathogen DNA amount in 3 out of 4 of single knock-out lines of *RPB1* and the double knock-out were statistically indistinguishable from those found in Col-0 (Uod-1 *rpb1* L33, Est-1 *rpb1* L6 & L127 and Uod-1 *rpb1 / rpb1-like-4* L17), while all of the *rpb1* knock-out lines accumulate significantly more *P. brassicae* DNA at 19 dpi than the wild-type Uod-1 and Est-1 controls. The symptoms observed in these lines were scored as 3 or 4 according to the DI scale and the DI score was greater than 86, whereas in the Uod-1 and Est-1 wild-type genotypes the *P. brassicae* the plants were assigned to the 0, 1 or 2 classes and a DI of 37.5 and 31.5 respectively (Figure 22). The *rpb1-like-4* knock-out lines were found to be statistically indistinguishable in terms of *P. brassicae* DNA quantities when compared to their corresponding wild-type controls; the symptoms observed in *rpb1-like-4* mutants were assigned to the 0, 1 or 2 classes, and the DI derived was less than 45, similar to the results for the wild-type Uod-1 and Est-1 roots (Figure 22). Together these results show that *RPB1* is required for the resistance phenotype in the Arabidopsis accessions Uod-1 and Est-1. However, no significant changes in the resistance phenotypes were detected upon the deletion of *RPB1-like-4*, suggesting that only *RPB1* is involved in the clubroot resistance associated with this locus.

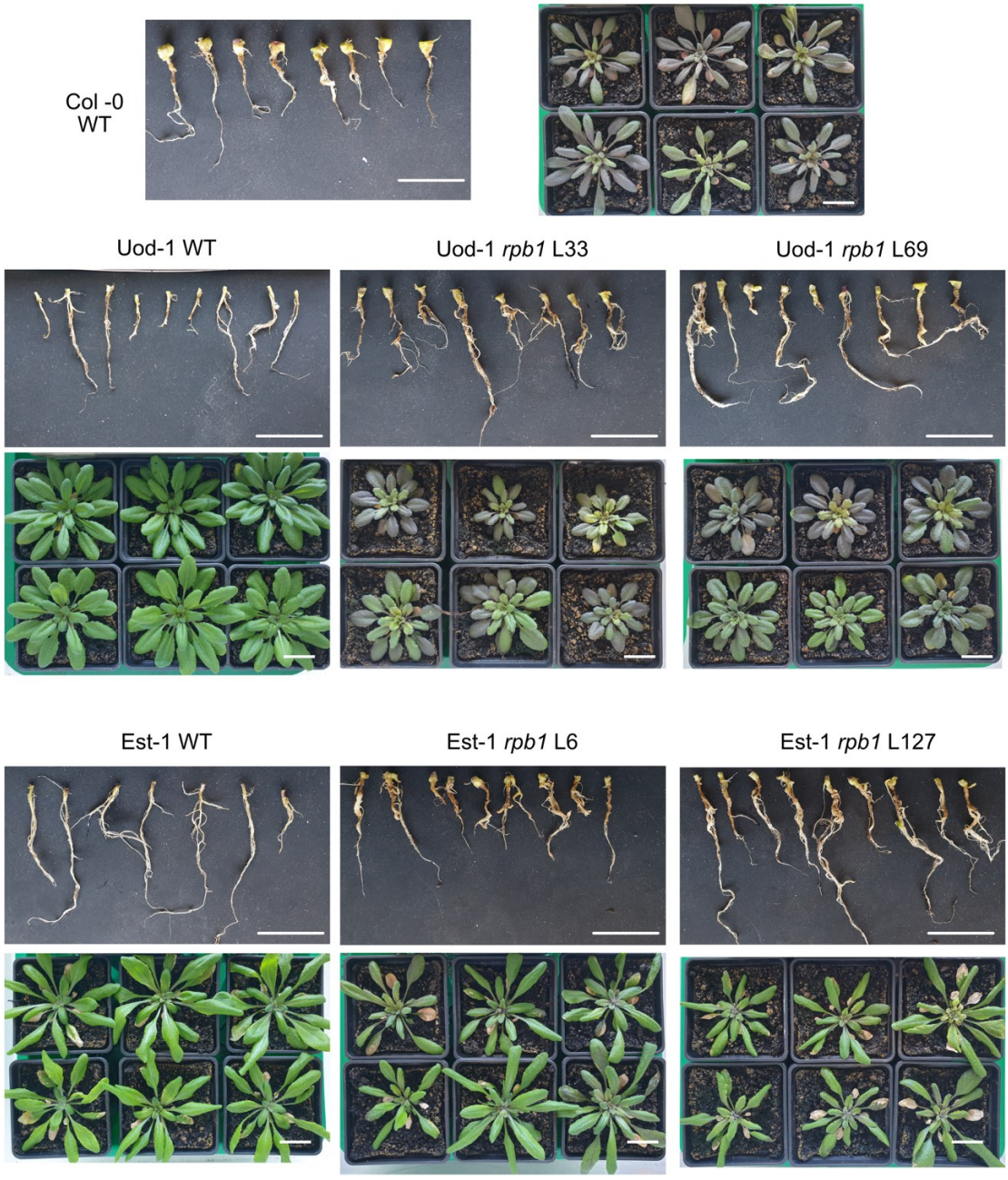


Figure 20. Phenotypes observed in the root, hypocotyls, and rosettes of the *rpb1* knock-out lines 19 days after inoculation, compared to the corresponding wild-type accessions and the clubroot susceptible accession Col-0.
The scale bar represents 2 cm

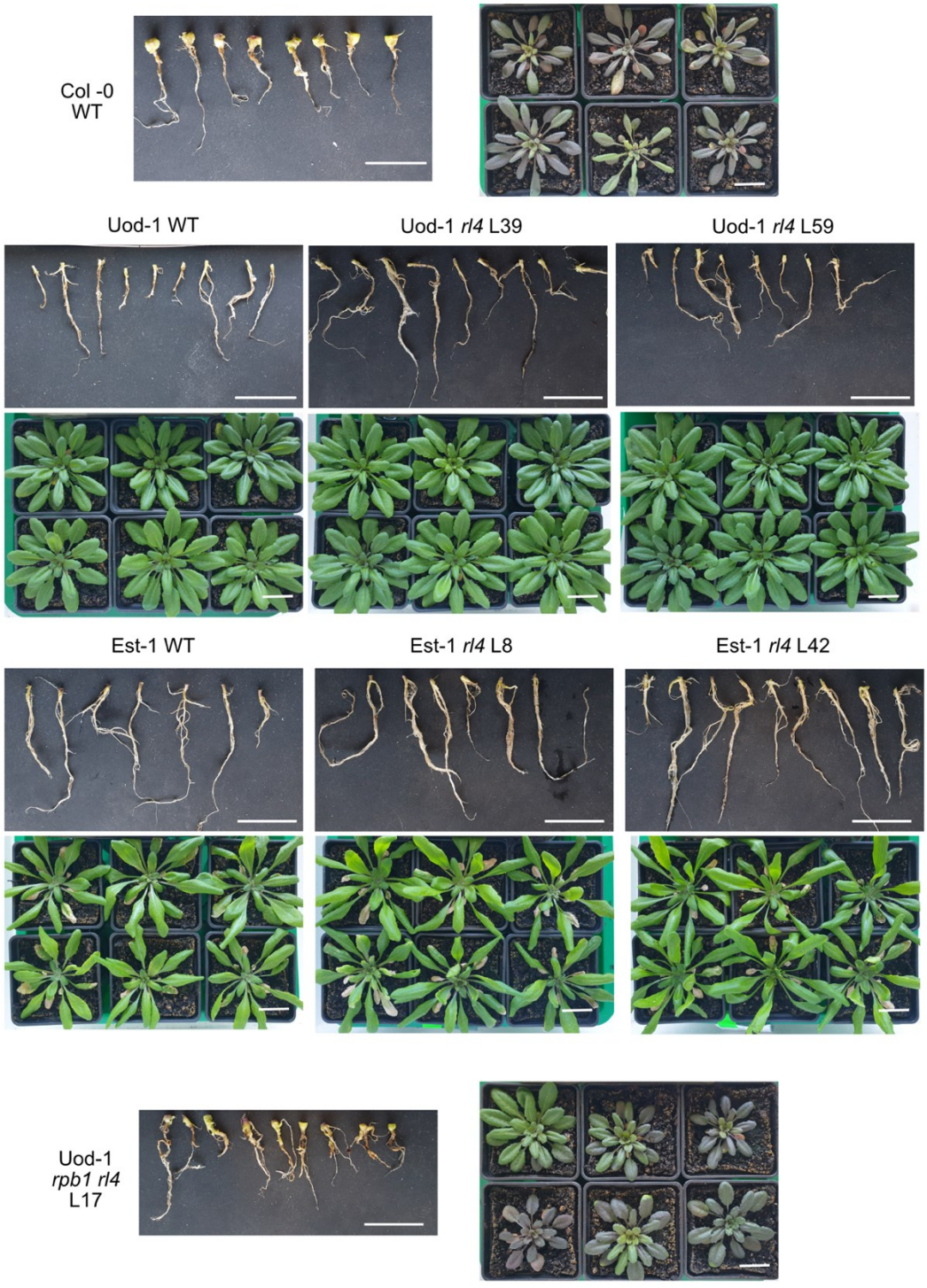


Figure 21. Phenotypes observed in the root, hypocotyls, and rosettes of the *rpb1-like-4* knock-out lines 19 days after inoculation, compared to the corresponding wild-type accessions and the clubroot susceptible accession Col-0.
(Controls shown in this figure are the same as in Figure 20, all plants came from the same experiment.). The scale bar represents 2 cm

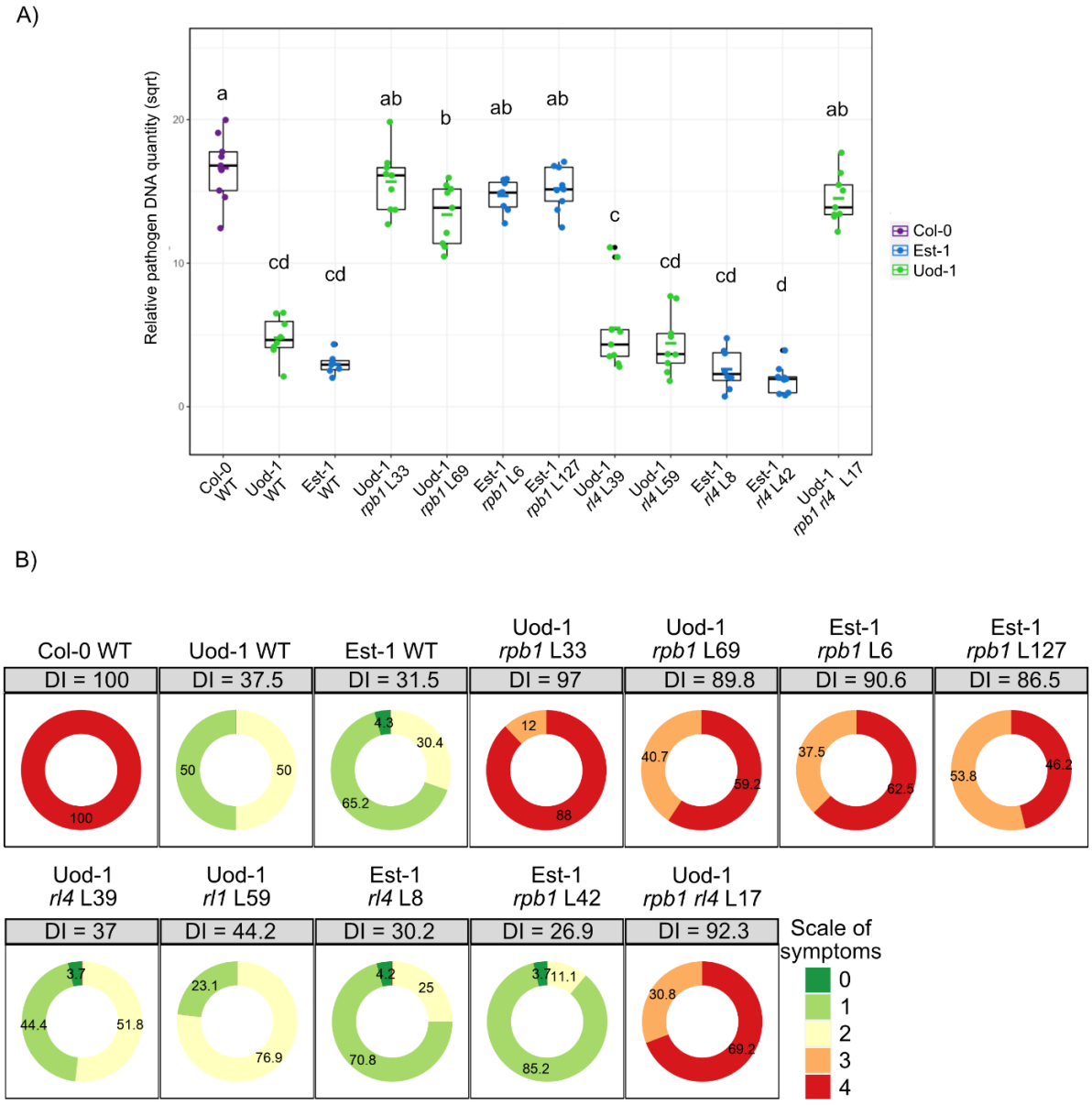


Figure 22. qPCR pathogen quantification and DI calculation in the *rpb1* and *rpb1-like-4* knock-out lines at 19 dpi.
A) qPCR relative quantification of *P. brassicae* in the *rpb1* and *rpb1-like-4* knock-out lines, different letters correspond to statistically significant differences with Tukey test, P <0.05. The data was transformed using a square root transformation to fit the normal distribution, this was confirmed with the Shapiro-Wilk normality test applied to the residuals of a linear model (p = 0.1789), a total of 7 to 9 biological replicates per genotype/treatment combination were used, each one representing a mixture of 3 plants. B) Disease index calculation and percentage of plants of individual genotypes classified according to the scale of symptoms observed in individual plants (23 ≤ n ≤ 27). The calculated DI is presented in the grey boxes at the top of each donut chart. The percentage of plants classified in each category of symptoms is shown in the donut charts.

4.5.3 The *RPB1* knock-out lines have reduced expression of defense genes, but *RPB1* is still upregulated upon inoculation with *P. brassicae*

To assess whether the *rpb1* knock-out mutants have impaired defense responses, the expression of genes involved in known signaling pathways related to response to biotic stress was evaluated. *PR5*, previously found to be upregulated at 14 dpi in the accession Bur-0 and considered to be SA-responsive (Lemarié et al., 2015a) was selected in addition to *CYP71A13* (*At2g30770*), involved in the biosynthesis of camalexin, which is associated with reduced *P. brassicae* growth (Lemarié et al., 2015b), and *Apoplastic EDS1 Dependant-1 AED1* (*At5g10760*) that is upregulated during ETI similarly to *PR5* and is most likely involved in SAR homeostasis (Breitenbach et al., 2014). Expression of the transcriptional repressor of JA signaling *JAZ10* was also monitored, as *JAZ* genes have been shown to be upregulated during a compatible interaction at 14 dpi (Schuller et al., 2014). Expression of these genes measured with RT-qPCR revealed that *CYP71A13* and *AED1* are exclusively upregulated in the wild-type controls of Uod-1 and Est-1, but not in the corresponding *rpb1* lines, where the expression remained at levels similar to the mock inoculated plants, resembling the pattern observed in Col-0 (Figure 23). Regarding the expression of *PR5* there was a significant upregulation in both wild-type Est-1 and Uod-1 of about 800-fold and 300-fold respectively when compared to their mock inoculated controls, in Col-0 and the Uod-1 *rpb1* knock-out mutant there is also an upregulation in response to infection but only approximately 10-fold while for the Est-1 *rpb1* mutant there was no significant difference from the mock treated controls. In spite of this, the pattern of expression for *PR5* is similar to that of the other two defense-related genes, confirming that after knocking out *rpb1*, the resistant plants are losing their capacity to mount a defense response capable of restricting the growth and development of the pathogen. In these experiments no statistically significant changes in the expression of *JAZ10* were observed (Figure 23).

To assess the responsiveness of the *RPB1* promoter to infection, in the absence of a functional RPB1 protein, primers flanking the 3'UTR region of *RPB1* down-stream of the deletion sites were designed. These were used to determine if there is *RPB1* expression in Arabidopsis roots after *P. brassicae* infection using both wild-type and *rpb1* lines, to gain insights into whether RPB1 is participating in pathogen recognition events or in the downstream signaling. Through RT-qPCR, it was confirmed that the *RPB1* gene is still being actively upregulated in response to infection in the *rpb1* lines (albeit as truncated, nonsense transcripts) in both Est-1 and Uod-1 when compared to their respective mock inoculated

controls. In Est-1 the level of expression in the *P. brassicae* inoculated plant is not statistically different when comparing the wild type and *rpb1* mutant (Figure 24). In the Uod-1 background *RPB1* expression could not be reproducibly detected in mock inoculated plants, however in both wild-type Uod-1 and the *rpb1* mutant, *RPB1* expression could be observed in *P. brassicae* inoculated plants of either genotype, though the *rpb1* knock-out line showed significantly lower expression, relative to the wild-type Uod-1 (Figure 24). All together these results reveal that the *rpb1* mutants have reduced capacity to induce defense responses, however the fact that expression of *RPB1* is still activated suggests that pathogen recognition events take place upstream and the triggering of a subset of responses occurs in the absence of functional RPB1.

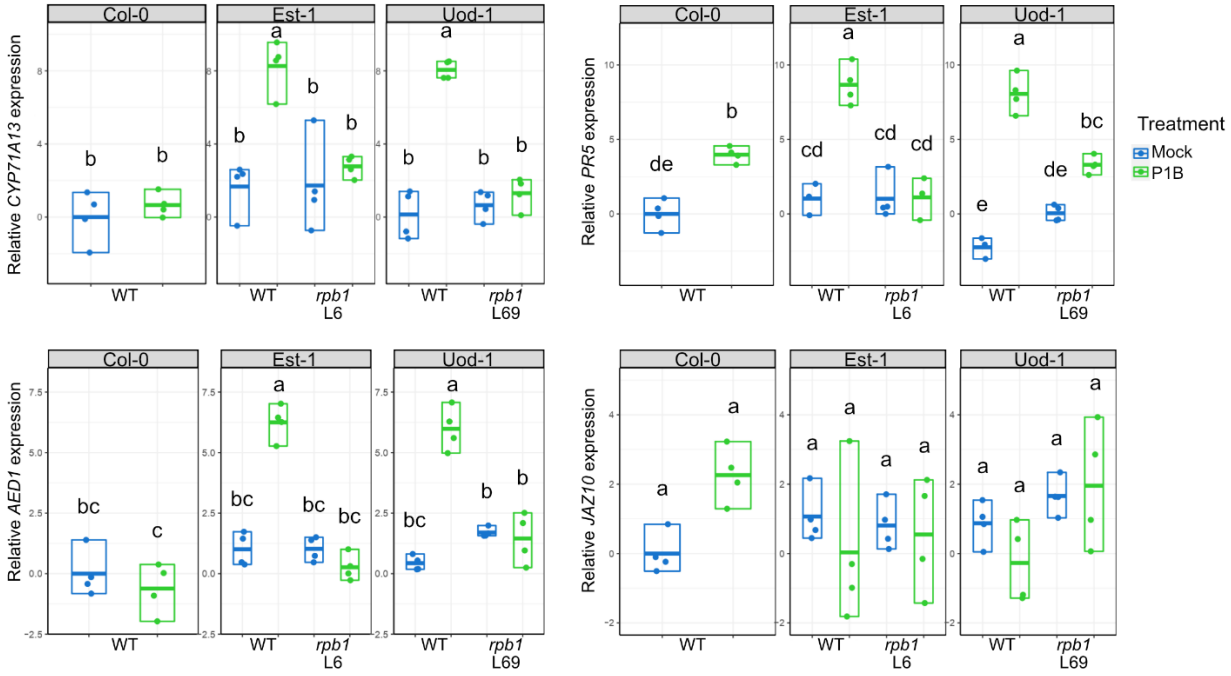


Figure 23. Relative gene expression measured with RT-qPCR of selected genes involved in defense responses in the *rpb1* knock-out lines at 7 dpi.
Letters denote statistically significant differences with Tukey test, P <0.05, and the relative gene expression is presented in the $\log_2$ scale. The normality of the residuals and variance tests were performed prior to the ANOVA. The boxes represent the range and the mean of each group of data, and each biological replicate (n = 4) corresponds to the mixture of the roots and hypocotyls of 6 individual plants.

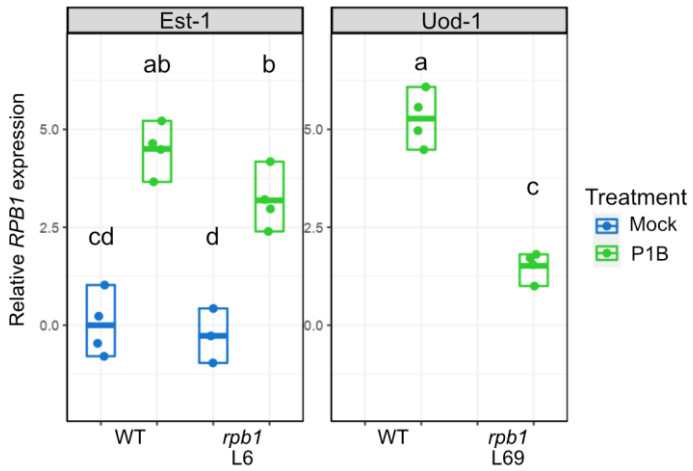


Figure 24. Relative gene expression measured with RT-qPCR of *RPB1* in the *rpb1* knock-out lines at 7 dpi.
The letters correspond to the statistically significant differences with Tukey test, P <0.05, and the relative gene expression is presented in the $\log_2$ scale. The normality of the residuals and variance tests were performed prior to the elaboration of the ANOVA. The boxes represent the range and the mean of each group of data, and each biological replicate (n = 4) corresponds to the mixture of the roots and hypocotyls of 6 individual plants.

4.5.4 The *RPB1* knock-out lines do not exhibit cell death patterns associated with *P. brassicae* resistance

To further characterize the differences between the *rpb1* lines and their corresponding wild-type genotypes in the Est-1 and Uod-1 backgrounds, morphological and histological observations were made of the hypocotyls at 25 dpi, a time point when it is expected that a significant proportion of the pathogen population will have matured to the resting spore stage. Initially, the general morphology of the hypocotyls and upper portion of the roots was observed under a stereomicroscope. In the mock inoculated plants, there were no differences in the morphology of the roots or hypocotyls (Figure 25). However, in the plants inoculated with *P. brassicae*, the development of the main symptoms of clubroot, such as gall formation in the roots, swelling of the hypocotyls and reduced presence of fine roots was observed in Col-0 and the *rpb1* lines. On the other hand, wild-type Est-1 and Uod-1 do not display swollen hypocotyls, the fine roots are still present, and only in a few cases were some small galls detected in the secondary roots or at the base of the hypocotyl, but the hypocotyl itself remained unaffected (Figure 26). Furthermore, some darkening in the roots of these resistant accessions was apparent, potentially caused by lignification and changes to secondary cell walls that are associated with immune responses (Figure 26).

Longitudinal sections were made of the hypocotyls and upper roots which were then stained with toluidine blue, to differentiate cellulose (purple) and lignin (blue). Toluidine blue is also useful to stain pathogen structures, providing a powerful and simple tool for histological studies of clubroot disease (Schuller and Ludwig-Müller, 2016). In Figure 27 and Figure 28 representative pictures of three regions of the hypocotyls for each genotype are presented, focusing on the uppermost central part (left column), the epidermal cells in the middle (central column) and close to the interphase between the hypocotyl and the root (right column) comparing Col-0 WT, Est-1 WT, Est-1 *rpb1*, Uod-1 WT, and Uod-1 *rpb1.* In the mock inoculated samples similar anatomical patterns were observed in all the genotypes (Figure 27). In Col-0 and the *rpb1* lines inoculated with *P. brassicae* P1B, in the uppermost central part, secondary plasmodia have formed and the xylem vessels are disorganized and less abundant, contrasting with the xylem anatomy in the wild-type Est-1 and Uod-1, where it is continuous and abundant, as is the case in the mock-inoculated plants. Furthermore, in these regions in Est-1 and Uod-1 there was a complete absence of secondary plasmodia or any other pathogen structure. In Col-0 and the *rpb1* lines when examining the epidermis, close to the middle part of the hypocotyl, there were an abundance of pathogen cells, mainly resting spores and secondary plasmodia filling hypertrophied cells. In the same region of the wild-type controls of the resistant accessions there were no resting spores or giant cells, nevertheless, there were isolated pathogen structures that were either surrounded by groups of lignified cells, or were restricted to the epidermis (Figure 28 and Figure 29).

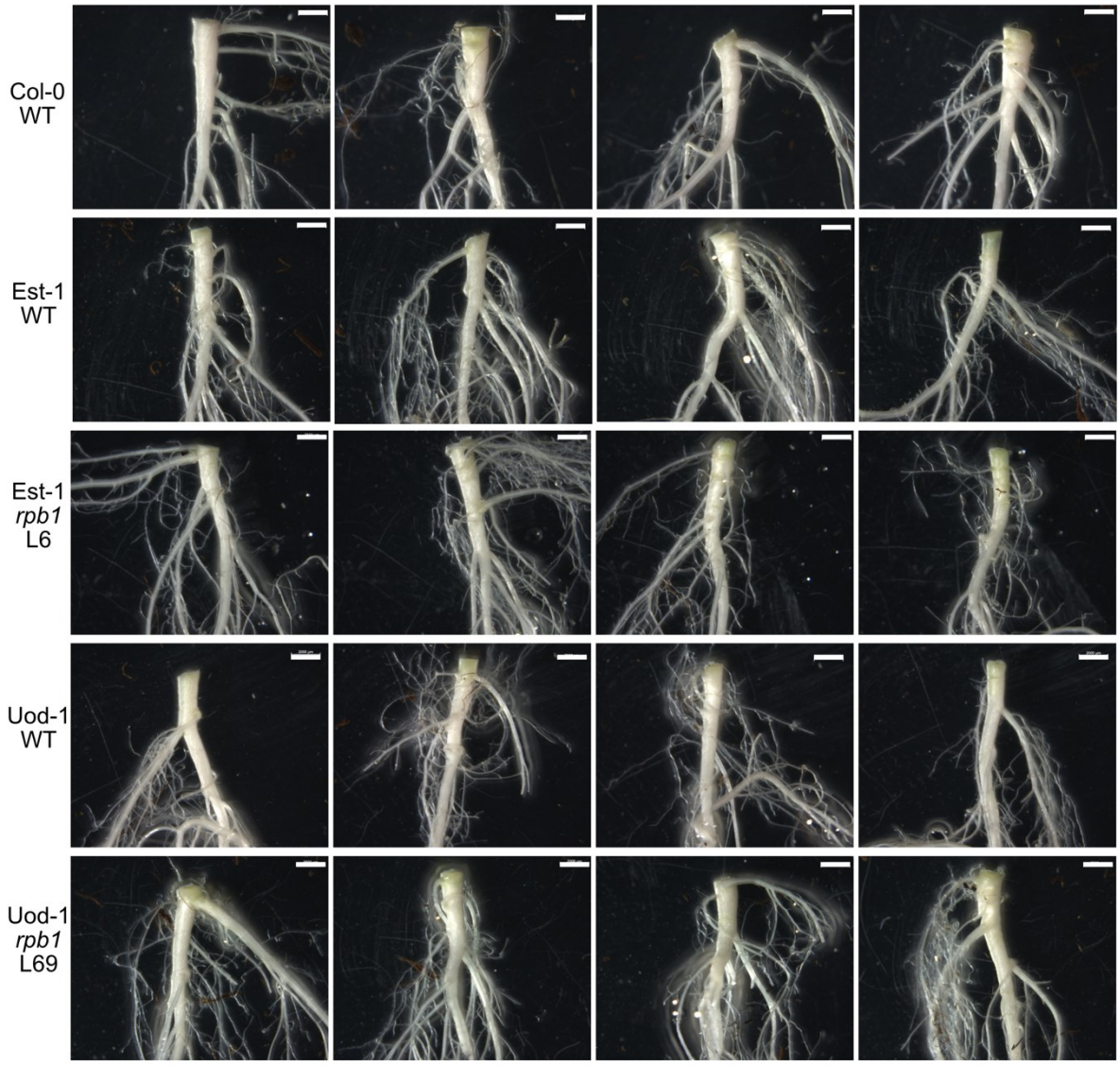


Figure 25. Representative pictures of hypocotyls and roots of Arabidopsis genotypes used in this study mock-inoculated with water at 25 dpi.
The scale bar represents 2 mm.

Close to the root-hypocotyl interphase secondary plasmodia and resting spores enclosed in the characteristic hypertrophied cells were observed in all the susceptible genotypes, and in some cases, in the resistant accessions, but surrounded with lignified tissue, probably limiting the pathogen from advancing to the upper regions of the hypocotyls or the central vascular cylinder (Figure 28).

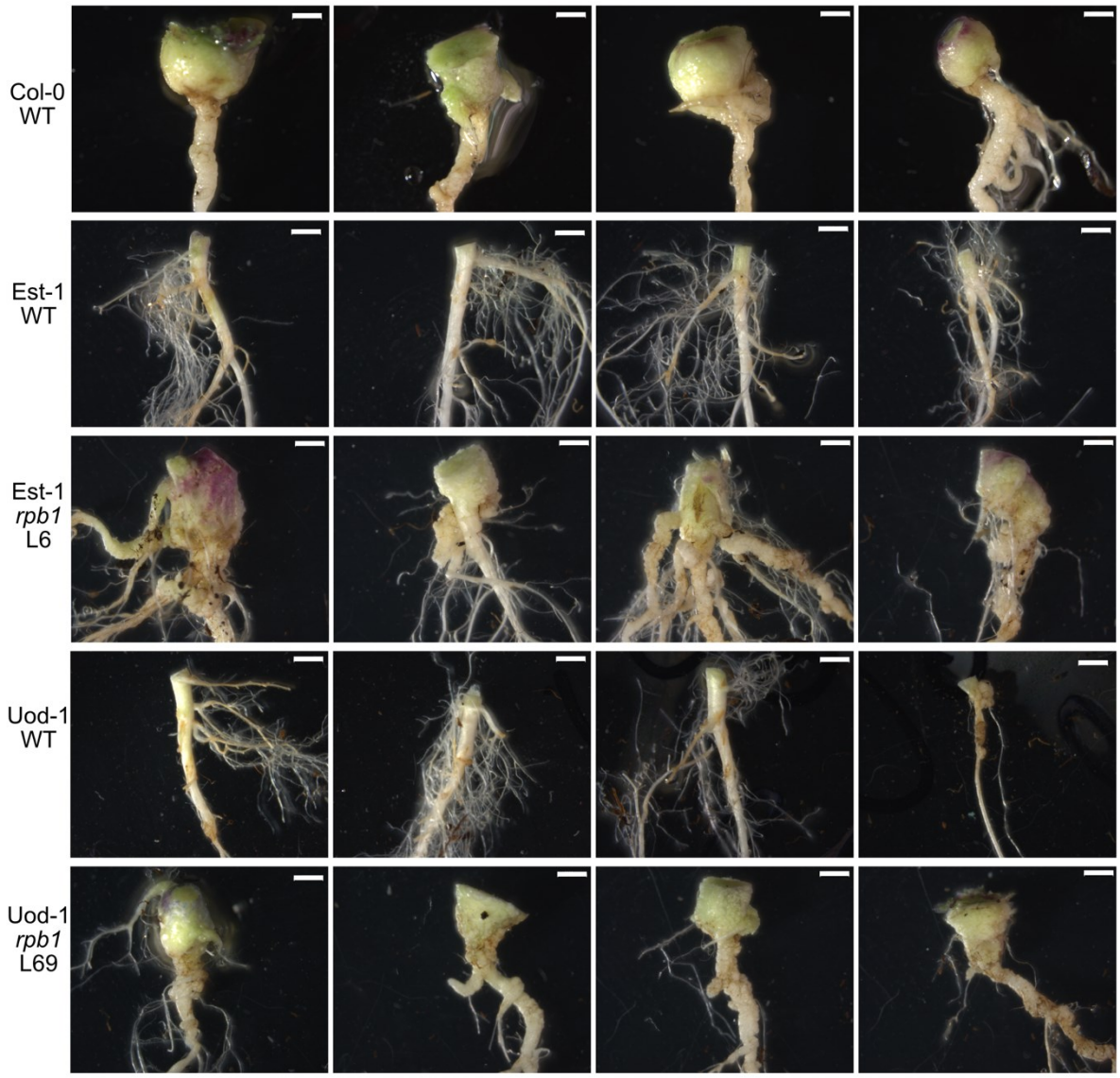


Figure 26. Representative pictures of hypocotyls and roots of Arabidopsis genotypes used in this study inoculated with *P. brassicae* P1B at 25 dpi.
The scale bar represents 2mm.

It was notable that in the plants with the most severe symptoms the central cylinder of vascular bundles breaks, leaving an empty space in the center of the hypocotyl. This symptom was never observed in the wild-type accessions Est-1 or Uod-1. These results illustrated that the *rpb1* lines have lost their capacity to prevent pathogen growth and colonization of the hypocotyl, and that in these mutants the disease is able to cause exactly the same anatomical changes observed in clubroot susceptible Arabidopsis accessions.

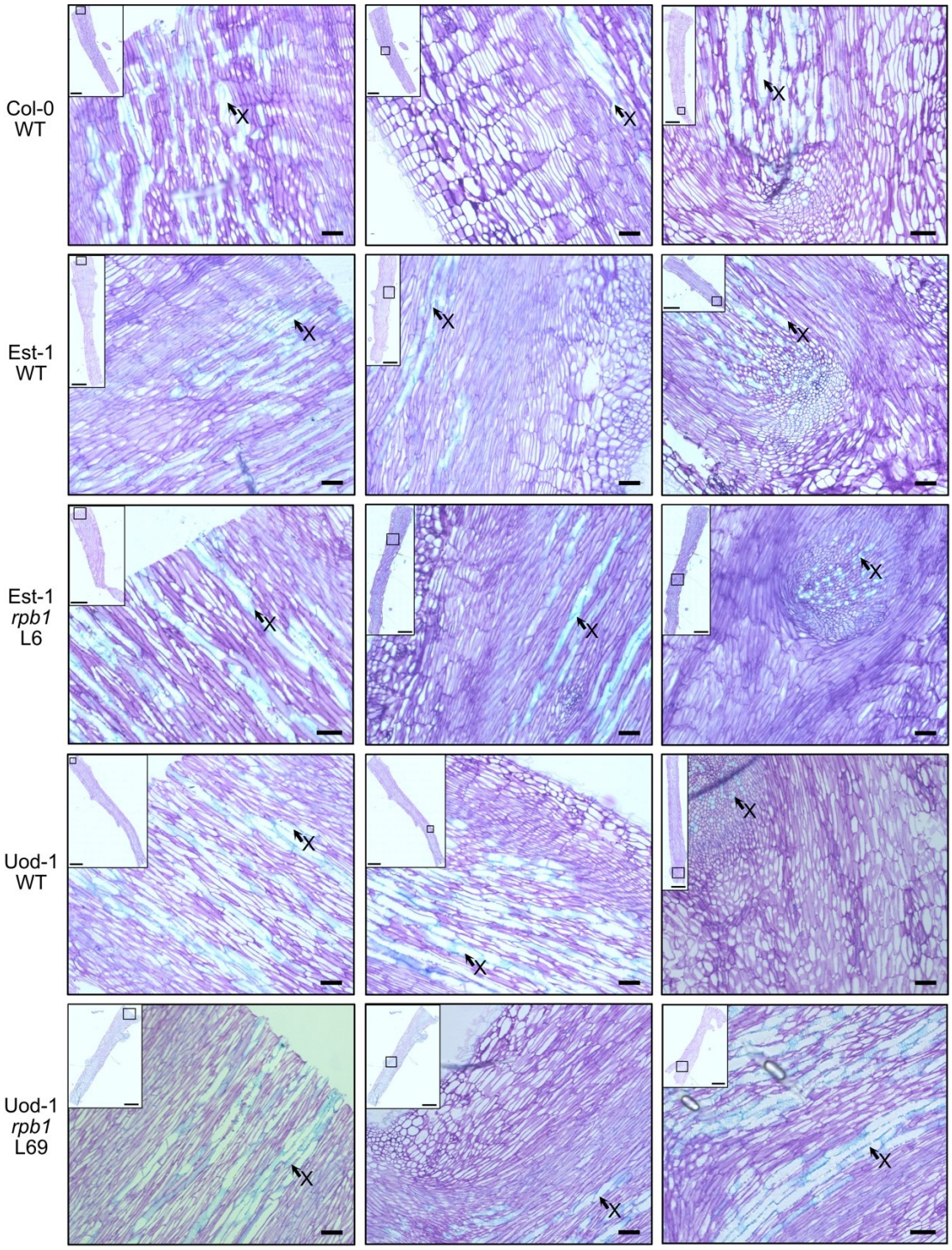


Figure 27. Histological characterization of hypocotyls of Arabidopsis genotypes used in this study in mock inoculated plants at 25 dpi.
The uppermost central part (left column), the epidermal cells in the midmost (central column) and the vicinity of the interphase between the hypocotyl and the root (right column) are presented for each genotype. X: Xylem cells. The scale bar in the hypocotyl overview picture corresponds to 1000 µm and the one in the detailed pictures corresponds to 50 µm.

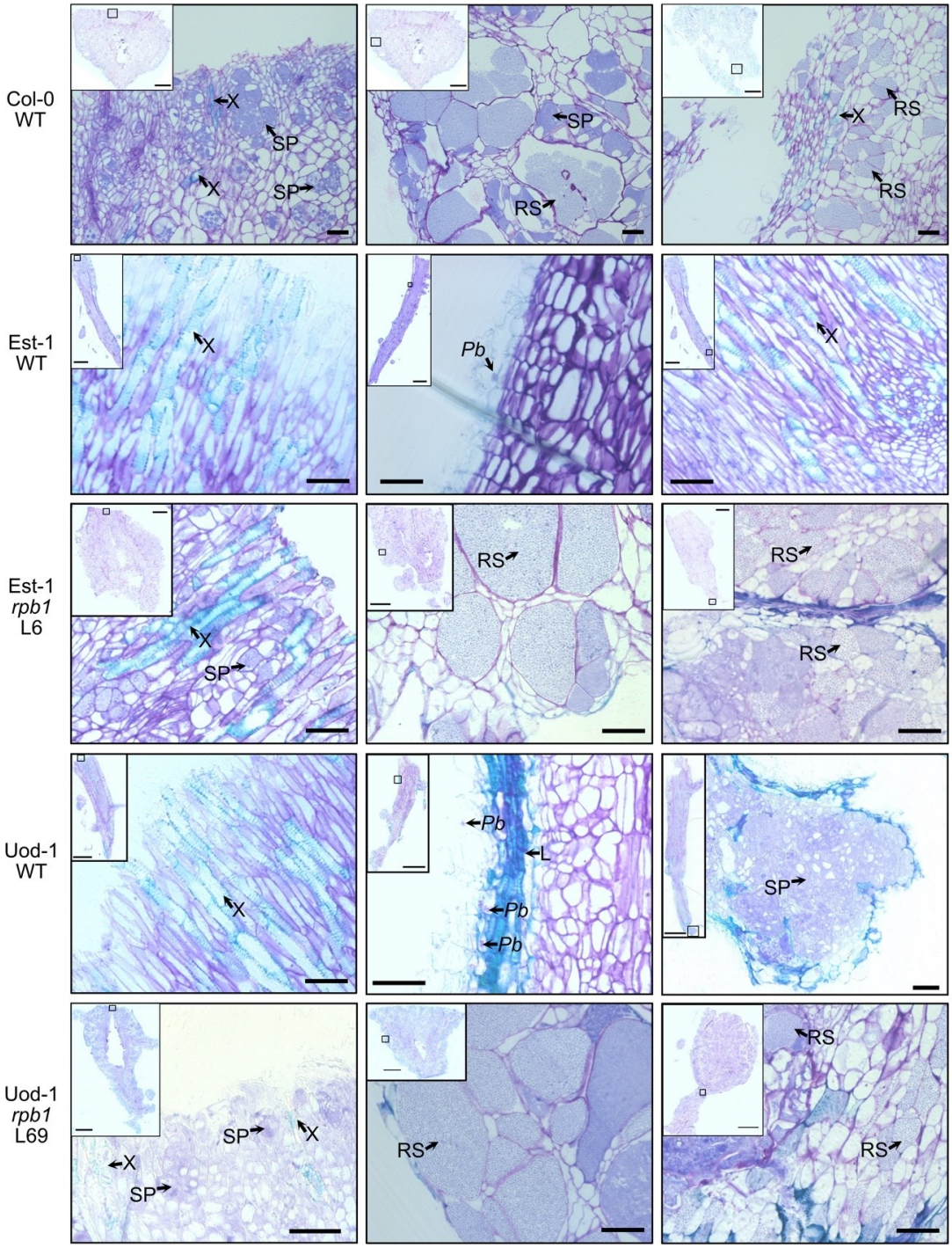


Figure 28. Histological characterization of hypocotyls of Arabidopsis genotypes used in this study inoculated with *P. brassicae* P1B at 25 dpi.
The uppermost central part (left column), the epidermal cells in the midmost (central column) and the vicinity of the interphase between the hypocotyl and the root (right column) are presented for each genotype. X: Xylem cells, L: Lignified tissue, RS: Resting spores, SP: Secondary plasmodia, *Pb*: *P. brassicae* cells. The scale bar in the hypocotyl overview picture corresponds to 1000 µm and the one in the detailed pictures corresponds to 50 µm.

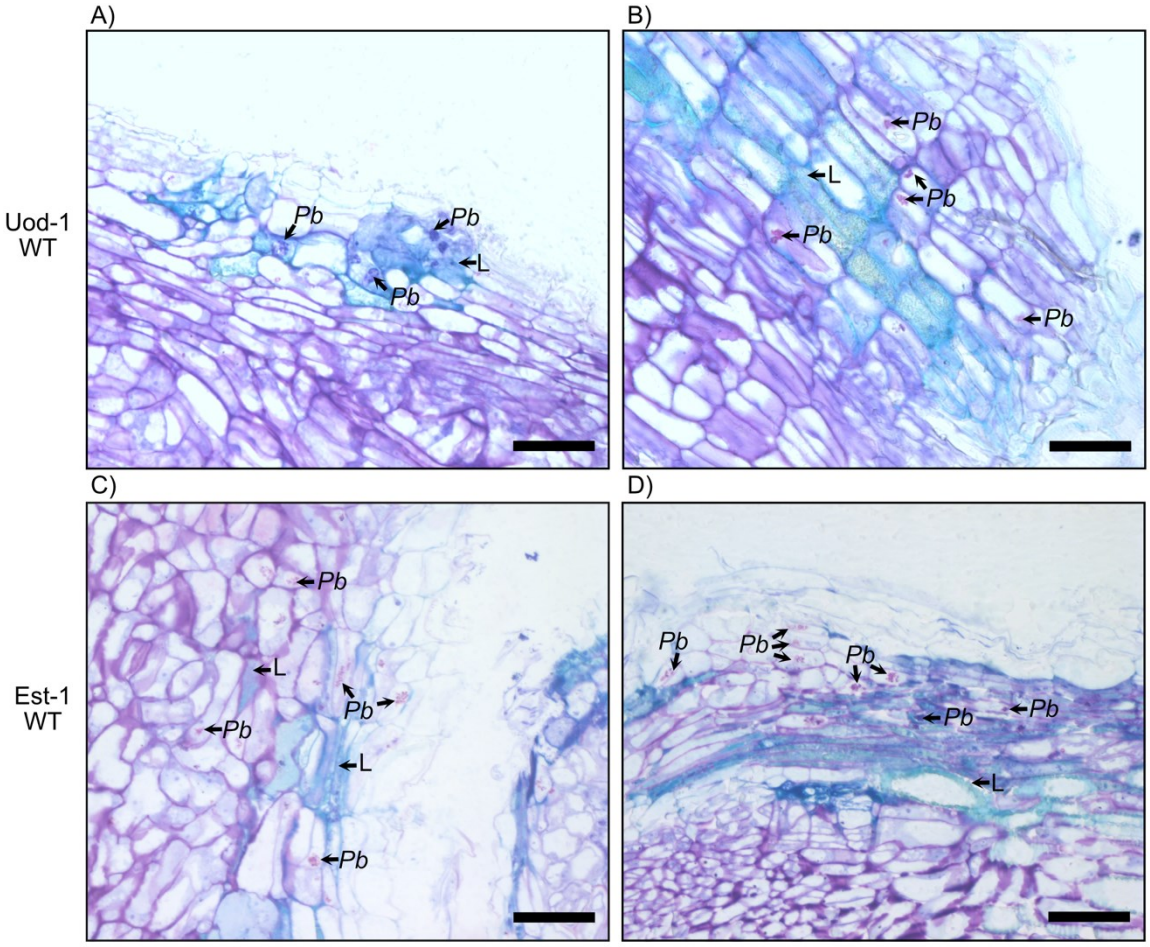


Figure 29. Detailed visualization of pathogen structures surrounded by lignified cells in the Est-1 and Uod-1 WT genotypes inoculated with *P. brassicae* P1B at 25 dpi.
X: Xylem cells, L: Lignified tissue, *Pb*: *P. brassicae* cells. The scale bar corresponds to 50 µm. A-B correspond to the Uod-1 accession and C-D to the Est-1 accession.

4.6 Transient expression of *RPB1* in *N. tabacum* induces a hypersensitive response

To further investigate the possible role of RPB1 in the induction of defense responses, the CaMV35S*::RPB1* cassette was transiently expressed in *N. tabacum* cv. Petit Habana leaves. It was observed that RPB1 causes a strong hypersensitive response that was visible from 2 days after infiltration with *A. tumefaciens* (Figure 30). Although this evidence is not enough to claim that the main function of *RPB1* is the positive regulation of defense responses or PCD, it seems probable that its upregulation might be required to upregulate in turn PR genes and induce other defense responses.

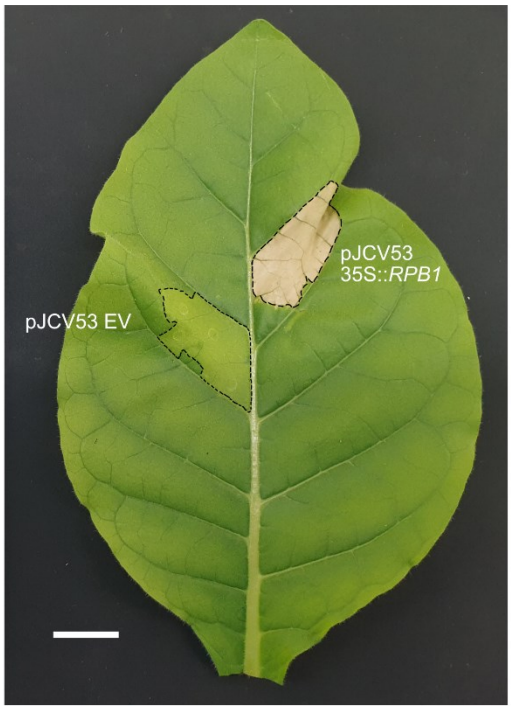


Figure 30. Transient expression of 35S::*RPB1* in *N. tabacum* leaves, three days after agroinfiltration.
The scale bar represents 2 cm. EV, empty vector control.

4.6 *RAC1* is not involved in resistance to *P. brassicae* P1B

4.6.1 An alanine to serine polymorphism in the C-terminal region of RAC1 is significantly associated with *P. brassicae* resistance

To further investigate the significant association with clubroot resistance that was found for the gene *RAC1*, the first step taken was to check its sequence in the resistant accessions and confirm the presence of the SNP coming from the 1001 genomes data. As mentioned previously, *RAC1* codes for an TIR-NB-LRR resistance protein, responsible for resistance to the oomycete pathogen, *A. candida*. The allele conferring resistance was found in the Arabidopsis accession Ksk-1 and identified using positional cloning via a biparental mapping population with the susceptible Col-0 accession. Despite the fact that the *RAC1* allele in Ksk-1 is very similar to the one found in Col-0, it contains the insertion of a transposon of 524 bp in the fourth intron, and also multiple substitutions in the LRR domain, few substitutions in the NLS but a degree of high conservation in the TIR domain, where only one substitution was observed. *RAC1* is also present as a singlet locus, far from any other cluster of NLR genes (Borhan et al., 2004).

The *RAC1* sequence in Uod-1, obtained with Sanger sequencing, was compared with the sequences of the accessions available as chromosome level annotated genome assemblies, including the Col-0 accession. Similar to the previously published results, the TIR and the NB domain are highly conserved at the amino acid level, but the LRR domain and the fragment between the LRR domain and the C terminal region domain contained indels and substitutions. In the case of one accession, Shahdara (Sha), there was a premature stop codon in the TIR domain, hence it presumably produces a non-functional protein (Figure 31). As the highest number of polymorphisms observed were present in exon 5, this region was sequenced and compared for some of the clubroot resistant accessions using Sanger sequencing to validate the associated SNP and investigate whether the susceptible and resistant accessions share similarities. After aligning the obtained sequences, the SNP identified by GWAS was determined to cause a substitution N1025S relative to the Col-0 allele. This substitution is present between the LRR domain and the C-terminal end, meaning that it has no obvious impact in the protein function (Figure 32).

The expression of *RAC1* was checked using RT-qPCR to investigate if its response profile is different when comparing susceptible and resistant accessions at 7 dpi. *RAC1* is not upregulated upon inoculation with *P. brassicae*, however *RAC1* transcript levels are significantly higher in Est-1 compared with Col-0 plants (Figure 33).

Figure 31. Protein alignment of the TIR, NB and LRR domains of the RAC1 protein.
The green highlighted letters correspond to the amino acid polymorphisms observed.

Figure 32. Protein alignment after the LRR domains showing the amino acid substitution caused by the SNP that showed high association with the resistance phenotype.
The green highlighted letters correspond to the amino acid polymorphisms observed and the blue highlighted letters show the substitution caused by the significantly associated SNP.
The reaction to the P1B pathotype is shown in parenthesis: (S) susceptible, (R) resistant.

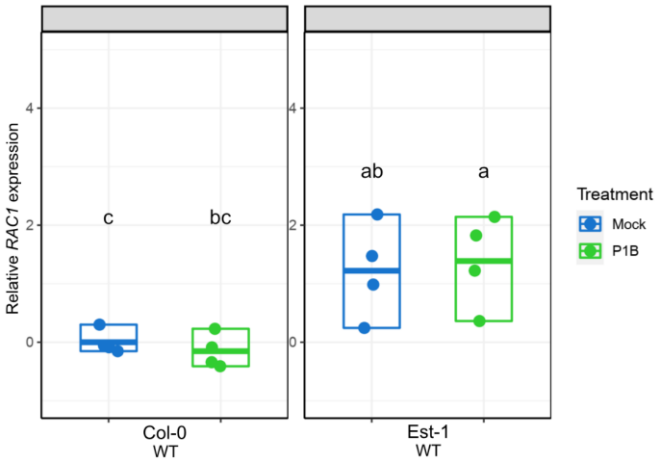


Figure 33. Relative gene expression of *RAC1* measured with RT-qPCR of *RPB1* in Col-0 and Est-1 at 7 dpi.
The letters correspond to the statistically significant differences with Tukey test, P <0.05, and the relative gene expression is presented in the $\log_2$ scale. The normality of the residuals and variance tests were performed prior to the elaboration of the ANOVA. The boxes represent the range and the mean of each group of data and each biological replicate (n = 4) corresponds to the mixture of the roots and hypocotyls of 10 individual plants.

4.6.2 Knock-out lines of *RAC1* in resistant Arabidopsis accessions are not compromised in response to *P. brassicae*

To determine whether *RAC1* is involved in resistance to *P. brassicae* P1B, knock-out lines were generated with the CRISPR/Cas9 methodology, using gRNAs complementary to the first exon, which contains the region coding for TIR domain of the protein. The TIR domain was selected as the site for targeting because it is the region with the highest sequence conservation and generating indels close to the 5' end could produce a frame shift leading to a completely non-functional protein downstream. Genotyping transgene free T2 and T3 plants using conventional PCR several homozygous lines were identified containing deletions in exon 1 of *RAC1* in the Uod-1 and Est-1 backgrounds. These were confirmed using Sanger sequencing and the predicted proteins contained early stop codons and major changes that would likely abolish RAC1 function (Figure 34).

Three *RAC1* knock-out lines for each background, Uod-1 and Est-1, were inoculated with *P. brassicae* P1B and phenotyped at 19 dpi using pathogen quantification with qPCR and DI scoring. The knock-out lines still showed a high level of resistance, comparable to their respective wild-type controls, evidenced in the conserved size of the hypocotyl, signs of possible lignification in the main roots, restriction of clubroot development to only some incipient galls in the secondary roots, and in general very healthy rosettes (Figure 35). On

the other hand, the susceptible control Col-0 had typical, large galls in the main and secondary roots, swollen hypocotyls, and rosettes with reduced size.

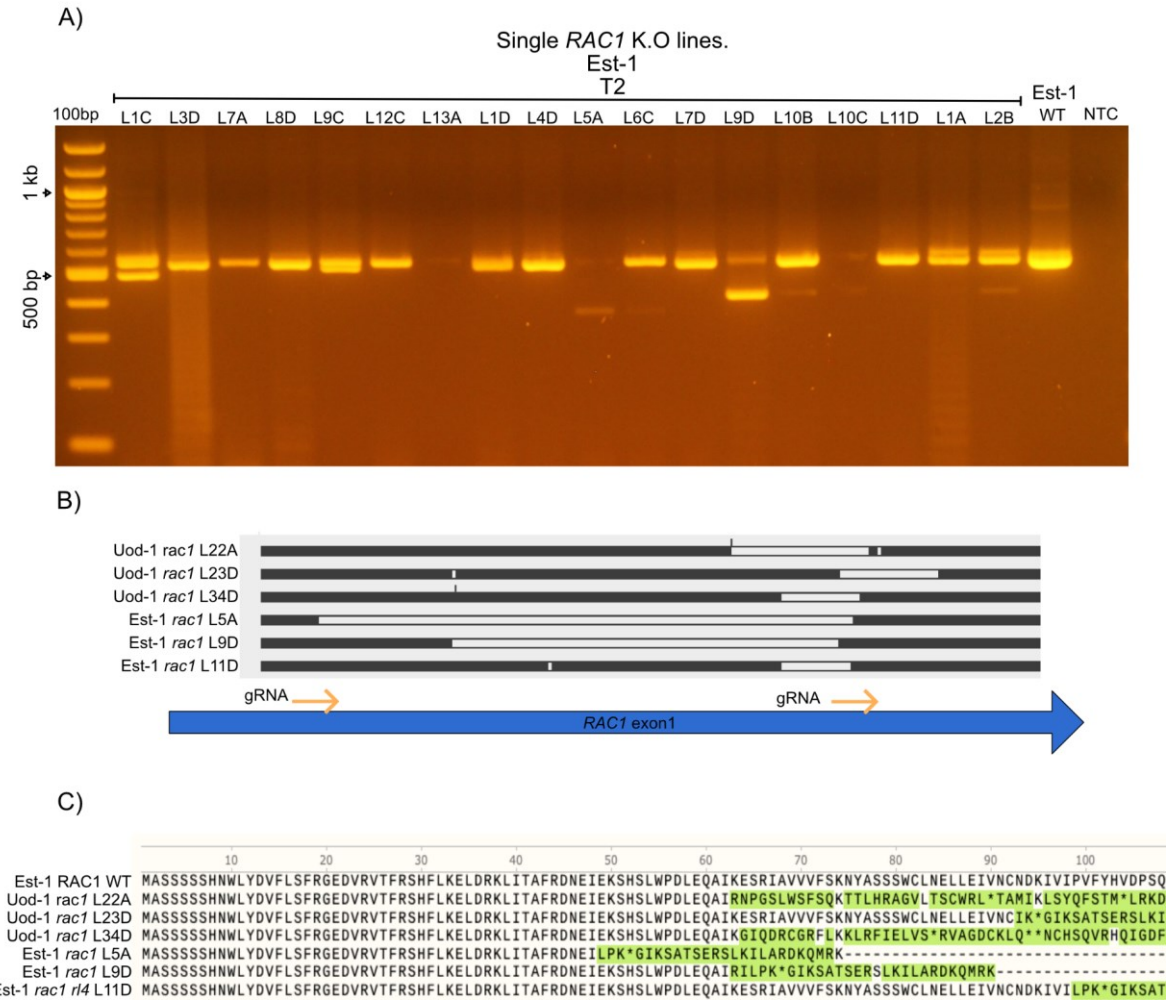


Figure 34. Genotyping of *RAC1* knock-out lines with conventional PCR and Sanger sequencing.
A) Representative agarose electrophoresis of the PCR products of the *RAC1* exon1 fragment in the K.O. lines and the Est-1 WT genotype. B) Representation of the deletions observed in the *RAC1* K.O. lines identified through Sanger sequencing, the black bars correspond to the portion of the sequence aligned to *RPB1* and the empty bars correspond with the deletion observed in selected lines, the orange arrows show the positions of the gRNAs. C) Predicted amino acid sequences in the different K.O lines compared to the RAC1 protein sequence; the green highlighted letters correspond to the amino acid changes observed.

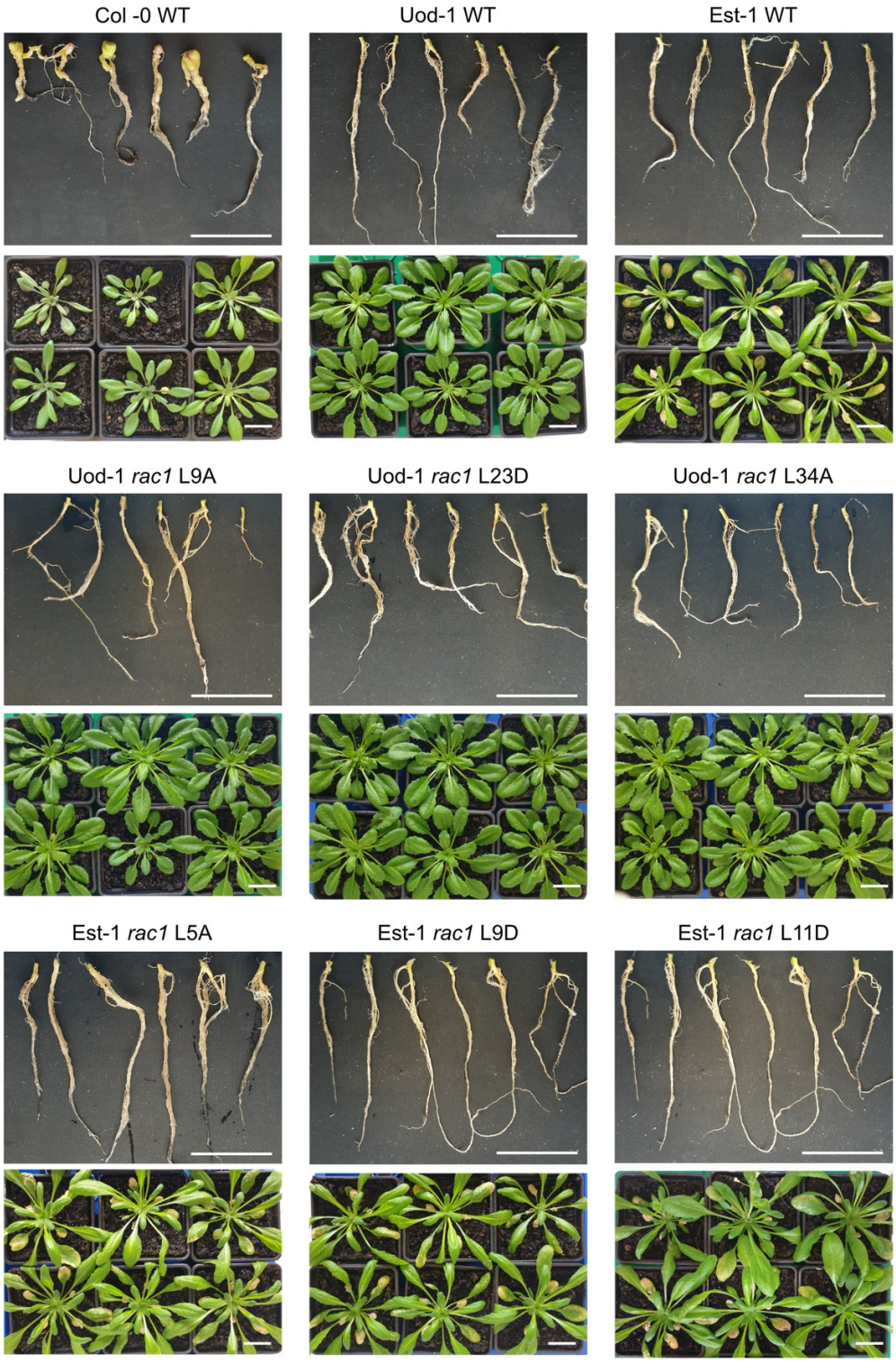


Figure 35. Phenotypes observed in the root, hypocotyls, and rosettes of the *rac1* knock-out lines compared to the corresponding wild-type controls and the susceptible accession Col-0.
The scale bar represents 2 cm

To determine if *RAC1* has some minor, but quantifiable effect on resistance, the relative pathogen DNA amount was measured with qPCR and the disease index scoring was calculated in the *RAC1* knock-out lines. There are no significant differences comparing *rac1* lines with their respective wild-type controls in terms of relative pathogen DNA quantification (Figure 36A). Accordingly, it was observed that only the susceptible control Col-0 exhibited gall symptoms classified as 3 or 4 and a DI = 94.4, while wild-type Uod-1 and Est-1 WT and all the *rac1* knock-out lines, had galls scored from 0 to 2 and DI less than 46 (Figure 36). All together these results establish that *RAC1* is not required for resistance to *P. brassicae*.

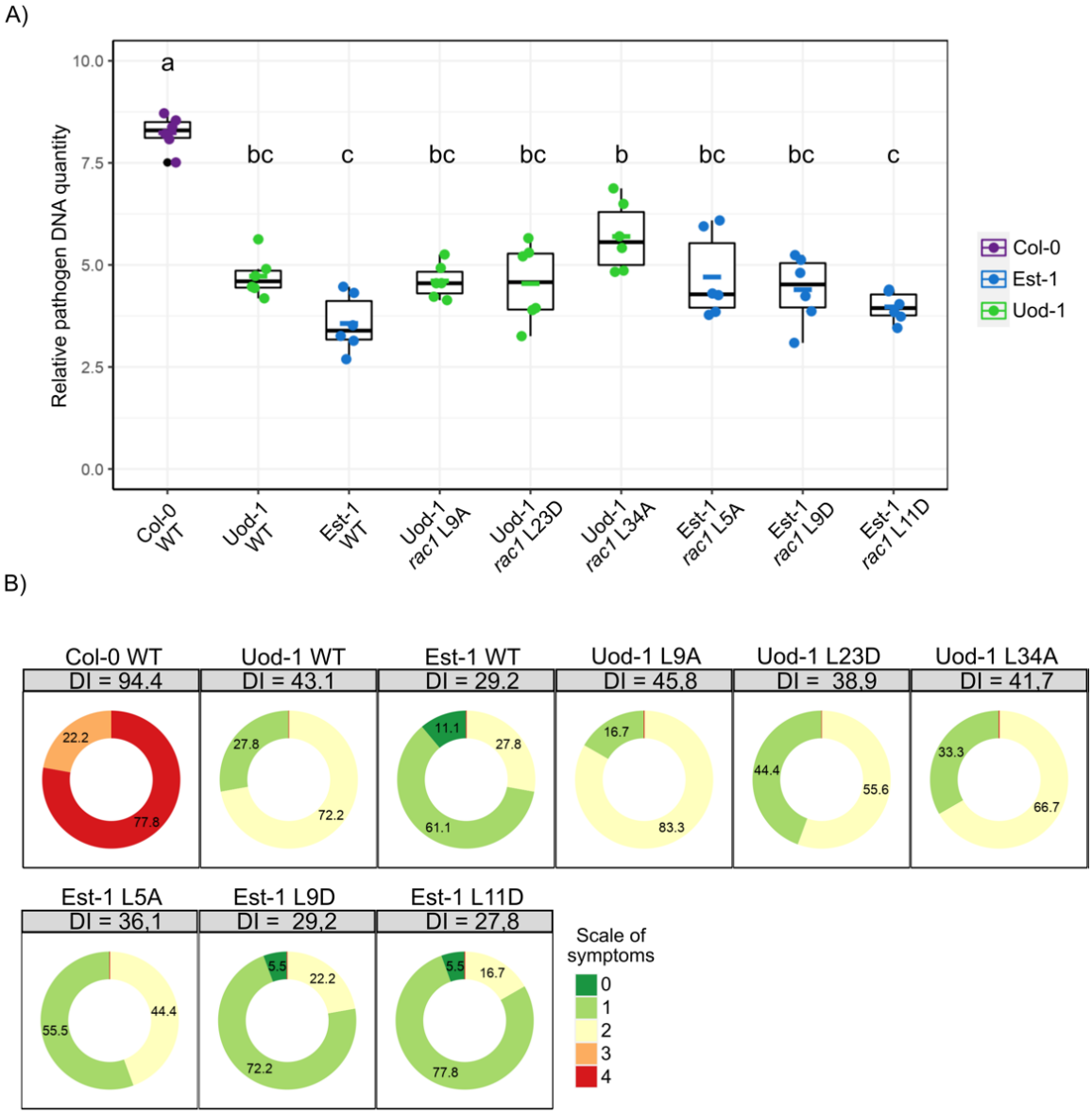


Figure 36. qPCR pathogen quantification and DI calculation in the *rac1* knock-out lines at 19 dpi.
A) qPCR relative quantification of *P. brassicae* in the *rac1* knock-out lines in a $\log_2$ scale, the letters correspond to the statistically significant differences with Tukey test (n = 6), P <0.05, the fitting to the normal distribution was confirmed with the Shapiro-Wilk normality test applied to the residuals of a linear model (p = 0.1789). B) Disease index calculation and percentage of plants of individual genotypes classified according to the scale of symptoms observed in individual plants (n = 18). The calculated DI is presented in the grey boxes at the top of each donut chart. The percentage of plants classified in each category of symptoms is shown in the donut charts.

4.7 Investigating incongruences between pathogen DNA levels and clubroot gall symptoms in Pro-0 – potential impact of anatomical differences

As mentioned previously, in the accession Pro-0 there was an incongruence between the measurement of the pathogen DNA quantity and the DI scoring. The relative *P. brassicae* DNA levels were 4 times higher on average than in Col-0, which is one the most susceptible accessions that found in the initial screening (Figure 37A). On the other hand, the DI scoring for Pro-0 was 87.9 and only about 50% of the plants exhibited symptoms classified as 4, contrasting with those observed in Col-0, that resulted in a DI = 96.7 and close to 90% of the plants presenting symptoms categorized as 4 at 19 dpi (Figure 37B).

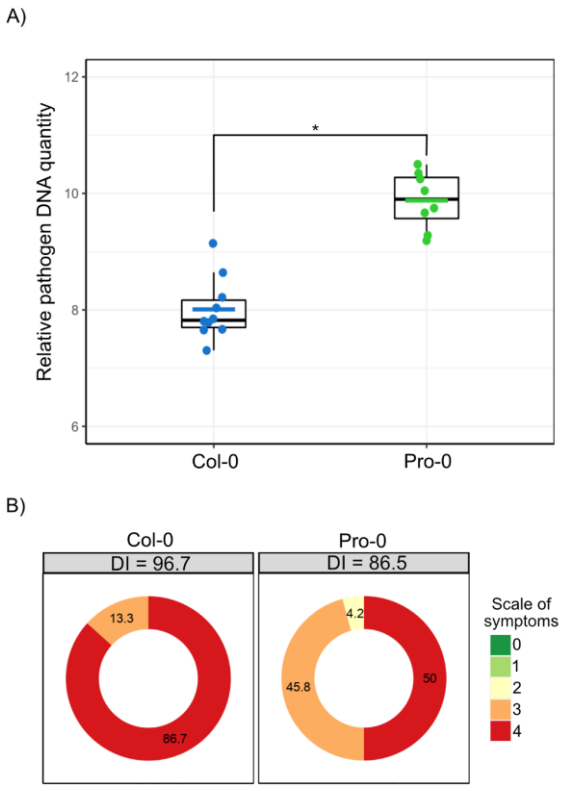


Figure 37. qPCR pathogen quantification and DI comparison between the Arabidopsis accessions Col-0 and Pro-0 at 19 dpi.
A) qPCR relative quantification of *P. brassicae* presented in a $\log_2$ scale, the asterisk indicates significant differences evaluated with a T-Student's test ($8 \leq n \leq 10$) $p < 0.05$. Prior to the T-Student's test. The colored horizontal bar represents the mean of each group of data. B) Disease index calculation and percentage of plants of individual genotypes classified according to the scale of symptoms observed in individual plants ($24 \leq n \leq 30$). The calculated DI is presented in the grey boxes and the percentage of plants classified in each category of symptoms is represented in the donut charts.

In the symptoms observed above-ground in the rosettes and hypocotyls, the wilting of leaves in Col-0 is much more evident than for Pro-0, and also the hypocotyls appeared to be more swollen and present a funnel-like shape, whereas in Pro-0 the hypocotyls are less swollen, especially at the base of the rosette, the production of flowers was observed in Pro-0 plants, but not in Col-0 (Figure 38). These results indicate that Pro-0 is less susceptible to clubroot disease than Col-0, but it is not clear why the relative pathogen quantification presents such abnormal values.

4.7.1 Pro-0 hypocotyls have a higher proportion of xylem bundles than Col-0

To further characterize the previous findings, the colonization and distribution of the pathogen in the hypocotyl tissues was investigated using microscopy at 19 and 25 dpi in both Pro-0 and Col-0. Through differential staining with toluidine blue of horizontal sections of infected and uninfected plants, a difference in xylem patterns was observed between the accessions. In the uninfected plants xylem bundles were more prominent in Pro-0 and made up a larger proportion of the hypocotyl (Figure 39), this observation was confirmed by measuring the total radius of the hypocotyl and the radius of the limit between the vascular cambium and the xylem cylinder per hypocotyl, and calculating the ratio between the two measurements to estimate the proportion of the hypocotyl that is occupied by xylem, confirming that Pro-0 has a significantly higher proportion of xylem compared to Col-0 at both time points (Figure 40).

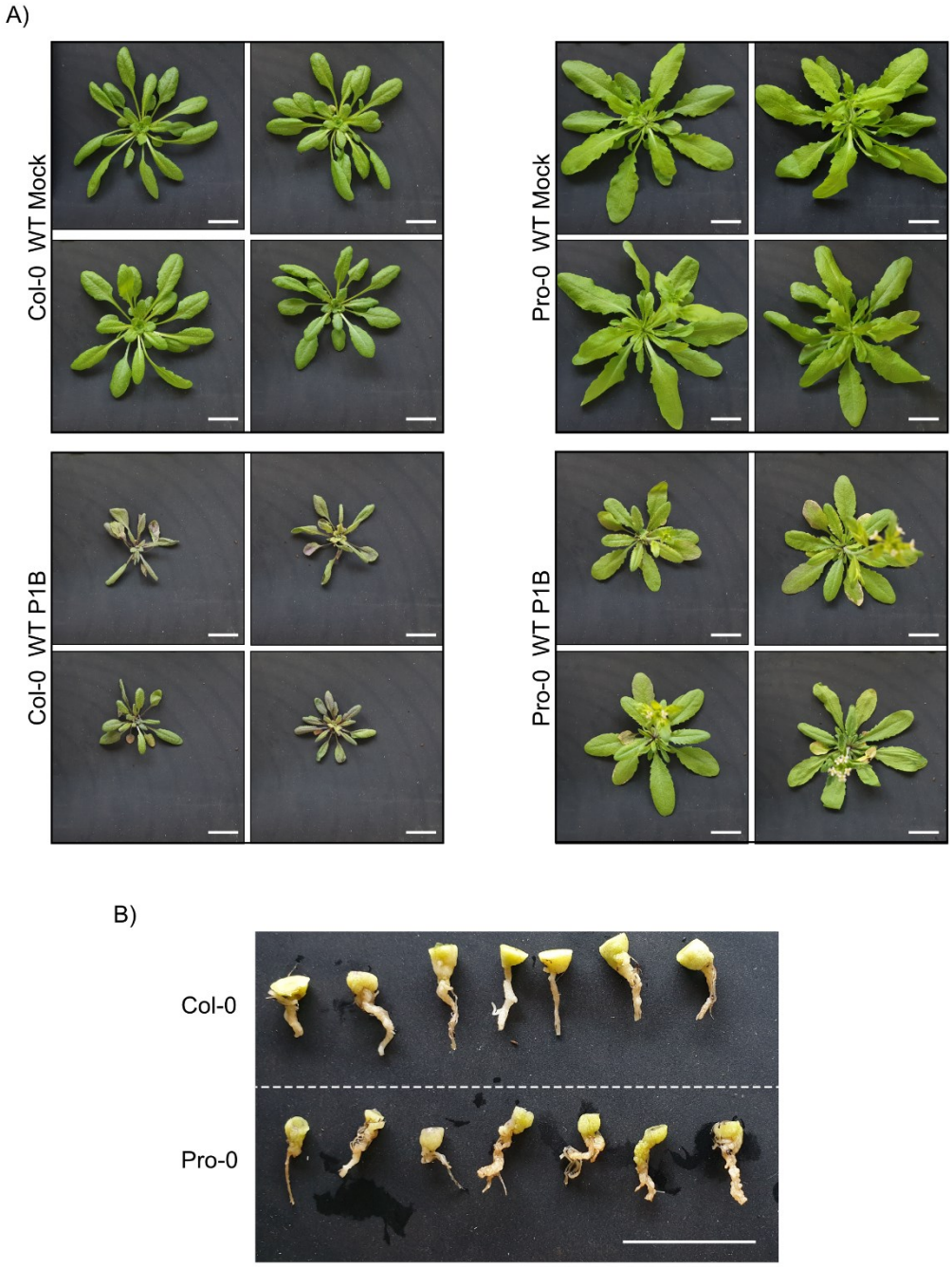


Figure 38. Phenotypes observed in the root, hypocotyls, and rosettes of the Col-0 and Pro-0 accessions at 19 dpi.
A) Representative pictures of the rosette appearance of mock and *P. brassicae* P1B inoculated plants, B) Comparative picture of Pro-0 and Col-0 hypocotyls. The scale bar represents 2 cm.

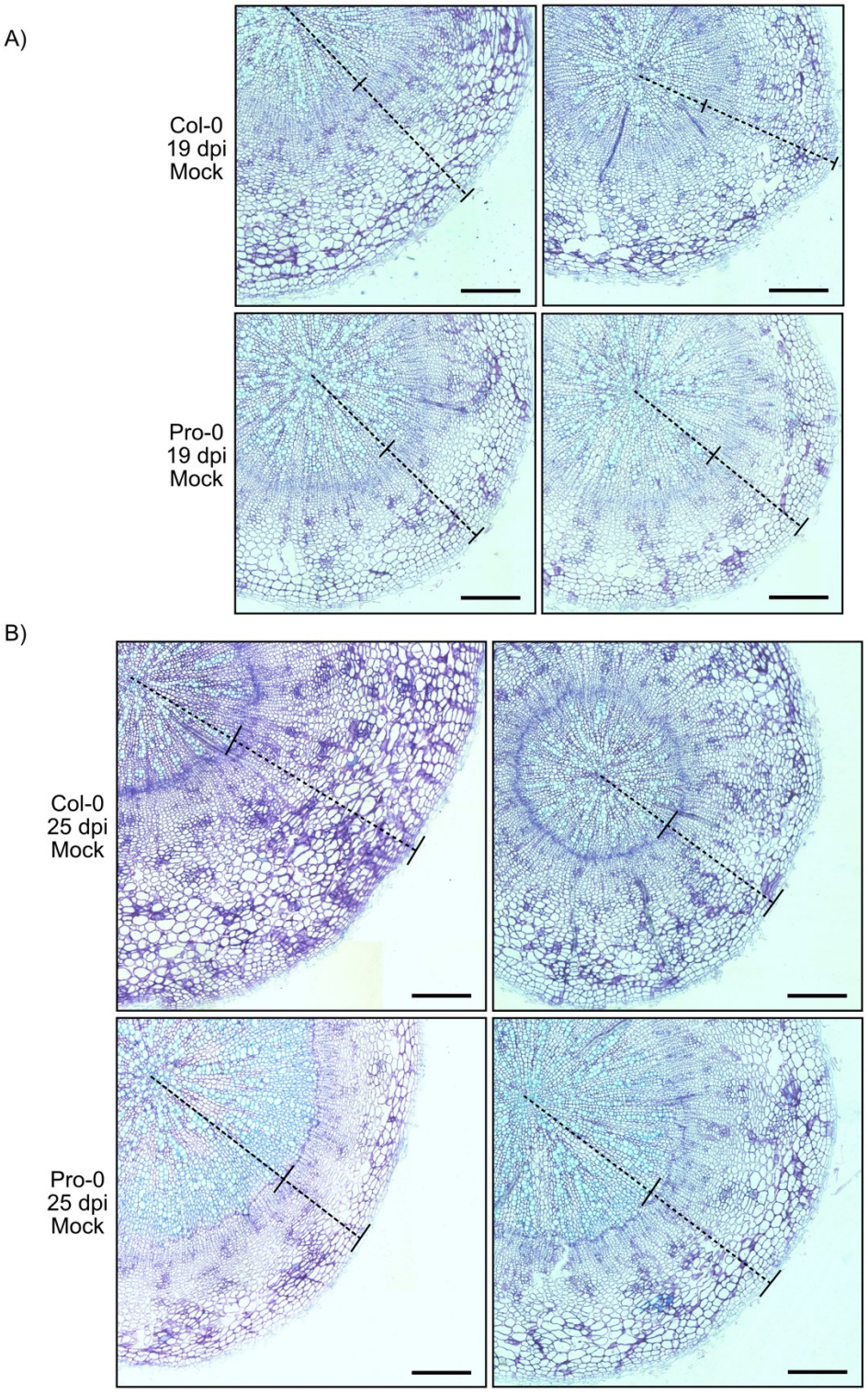


Figure 39. Hypocotyl sections stained with toluidine blue of uninfected Arabidopsis accessions Col-0 and Pro-0.
A) 19 dpi, B) 25 dpi. The dashed lines depict representative examples of the radius measurements taken. The scale bar represents 200 µm.

In the *P. brassicae* inoculated plants at 19 dpi, it can be observed that the Col-0 accession completely lacks a central xylem cylinder as would be observed in the uninfected plants, instead the tissue have sheared, creating a space inside the hypocotyl, moreover, xylem cells are dispersed across the hypocotyl tissue and not constrained in the center (Figure 39 and Figure 41). Concerning the pathogen, fully developed secondary plasmodia were spread across all of the hypocotyl, including the centermost space, adjacent to the remains of the central xylem bundles (Figure 41). On the other hand, in Pro-0, the central xylem cylinder is much less affected, and the xylem cells remained restricted to the central part of the hypocotyl, despite some shearing in the center in some galls, indicating that the disturbance of the vascular tissues is less severe compared to that seen in Col-0. The pathogen is also less developed in its colonization of the hypocotyl tissue, particularly towards the center, where no fully developed secondary plasmodia were observed, supporting the hypothesis that Pro-0 is less susceptible to *P. brassicae* than Col-0 (Figure 41).

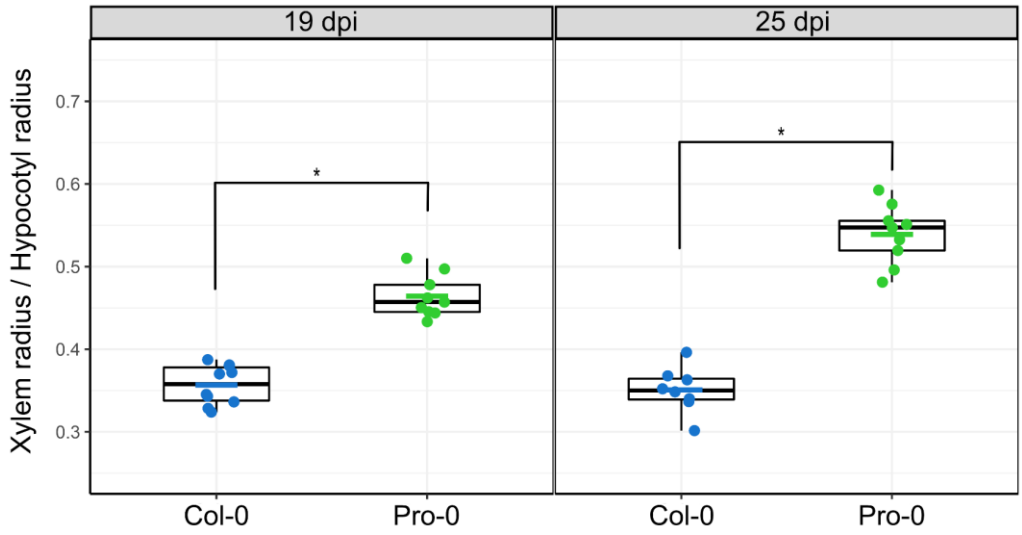


Figure 40. Proportion of xylem cylinder in Pro-0 and Col-0 at 19 and 25 dpi.
Asterisks indicate significant differences evaluated with a T-Student's test ($8 \leq n \leq 10$) $p < 0.05$. Prior to the T-Student's test, the normality and variance tests were performed within and between the groups of data respectively. The colored horizontal bar represents the mean of each group of data.

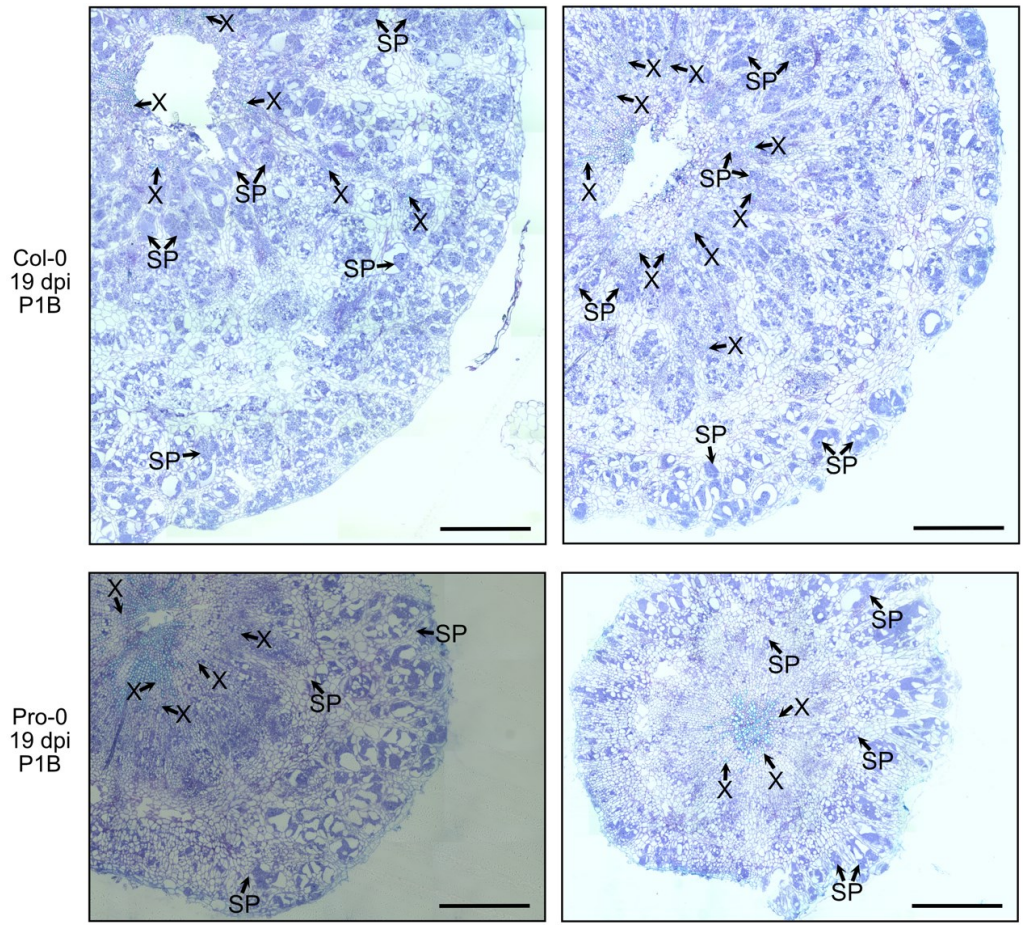


Figure 41. Hypocotyl sections stained with toluidine blue of Arabidopsis accessions Col-0 and Pro-0 infected with *P. brassicae* at 19 dpi.
X: Xylem cells, L: Lignified tissue, RS: Resting spores, SP: Secondary plasmodia, *Pb*: *P. brassicae* cells. The scale bar represents 500 µm.

4.7.2 Transcriptomic analysis of Pro-0 identifies NAC family transcription factors associated with xylogenesis and vascular cambium patterning

To further investigate the differences observed between Pro-0 and Col-0, a transcriptomics approach using RNA-Seq was selected to profile infected and uninfected plants at 19 dpi to identify differentially expressed genes that could explain the phenotypical observations. Each RNA sample was derived from a mixture of hypocotyls from 10 plants, with three biological replicates for each accession-treatment combination. The extracted RNA was suitable for the preparation of stranded cDNA libraries of 20 million reads, with low error rates and Q20 > 97% (Table 9). The samples from the plants inoculated with *P. brassicae* were observed to have higher GC content, which is a characteristic of the pathogen's genome.

Table 9. Sequencing data quality summary from RNA samples of Pro-0 and Col-0 at 19 dpi.

Accession	Treatment	Replicate	Raw reads	Effective (%)	Error (%)	Q20 (%)	Q30 (%)	GC (%)
Col-0	Mock	R1	43 347 786	99.08	0.03	97.53	93.00	44.90
Col-0	*P. bra* P1B	R1	42 356 236	99.05	0.02	98.10	94.54	54.97
Pro-0	Mock	R1	41 062 444	98.92	0.03	98.02	94.06	44.80
Pro-0	*P. bra* P1B	R1	44 331 748	98.32	0.03	97.71	93.56	52.24
Col-0	Mock	R2	42 027 338	99.52	0.03	97.84	93.72	44.90
Col-0	*P. bra* P1B	R2	45 254 238	99.19	0.03	97.90	94.06	54.16
Pro-0	Mock	R2	41 234 844	99.45	0.03	97.79	93.58	44.70
Pro-0	*P. bra* P1B	R2	40 779 726	99.14	0.02	98.12	94.48	51.57
Col-0	Mock	R3	39 837 002	99.49	0.03	97.78	93.50	44.77
Col-0	*P. bra* P1B	R3	40 133 600	98.94	0.02	98.05	94.36	53.25
Pro-0	Mock	R3	43 343 578	99.39	0.03	97.75	93.27	44.60
Pro-0	*P. bra* P1B	R3	41 217 614	99.03	0.03	97.90	93.84	51.98

The sequencing data was confirmed to be free from adapter contamination using the Cutadapt tool prior to mapping the reads, using the HISAT2 algorithm, to the TAIR10 Arabidopsis genome and the *P. brassicae* e3 genome to align, independently, reads originating from the host and the pathogen respectively. For the mapping to the Arabidopsis TAIR10 genome, in the mock inoculated plants of Col-0 there was a higher percentage of aligned reads compared to Pro-0 (98.4% and 96.8% respectively); as the TAIR10 genome was obtained from Col-0 there are likely some sequence differences leading to unmapped Pro-0 reads (Figure 42A). In the plants inoculated with *P. brassicae* the percentage of mapped reads is significantly higher in Pro-0 compared to Col-0, which might indicate that the pathogen amount in Col-0 is higher, favoring the hypothesis that Pro-0 is less susceptible than Col-0 as evidenced in the symptoms and microscopic observations, but contrary to the relative pathogen DNA quantification using qPCR (Figure 42A). The previous observation was also supported by the percentage of reads mapped to the e3 genome of *P. brassicae*, that resulted in significantly higher proportion of mapped reads in Col-0 compared to Pro-0 (Figure 42B). After mapping, the calculation of gene expression from the BAM files was determined with the featureCounts algorithm to and used for the differential gene expression (DEG) analysis.

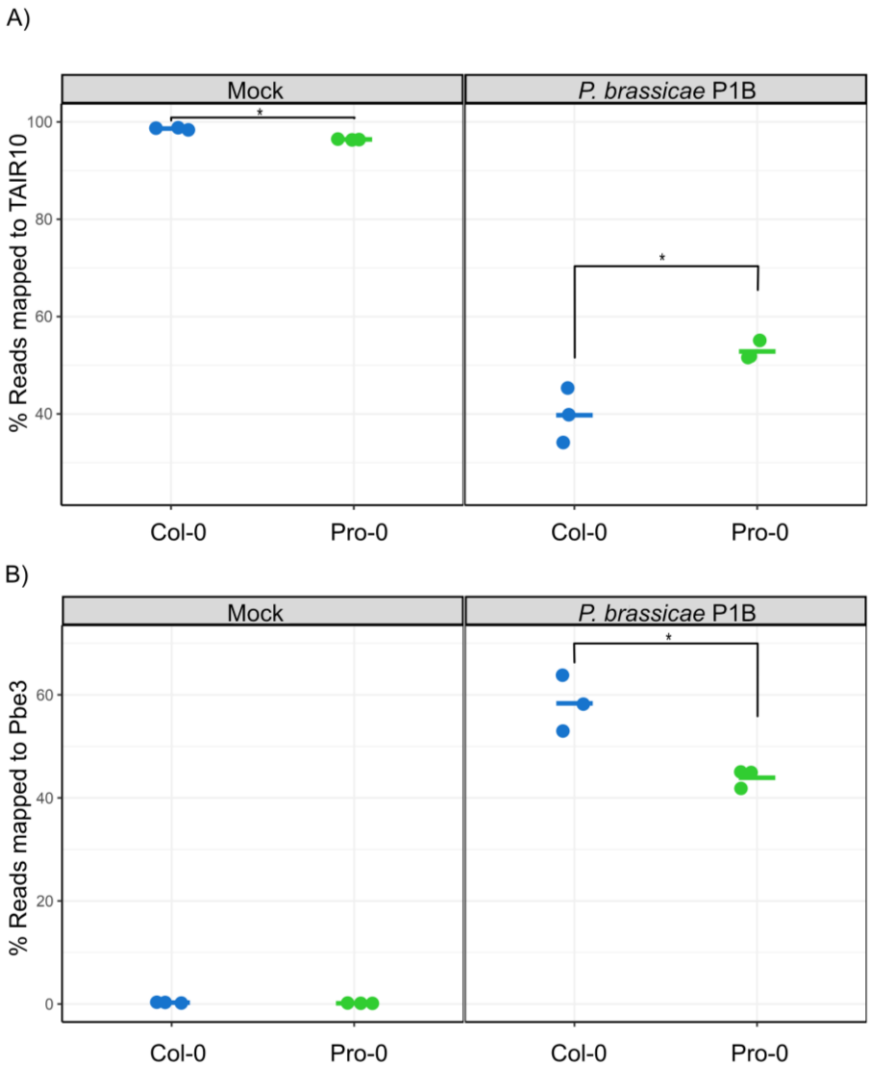


Figure 42. Percentage of mapped reads of the RNAseq to the Arabidopsis and *P. brassicae* genomes.
Asterisks indicate significant differences evaluated with a T-Student's test $p < 0.05$. Prior to the T-Student's test, the normality and variance tests were performed within and between the groups of data respectively. The colored horizontal bar represents the mean of each group of data. A) Percentage of reads mapped to the *A. thaliana* TAIR10 genome, B) Percentage of reads mapped to the *P. brassicae* e3 genome.

The first step, prior to DEG analysis, is the estimation of size factors and the normalization of counts to be able to compare between the different samples and replicates. After this normalization step, performed in the DESeq2 package, an assessment of the quality of the data to determine replicate similarity and visualize the influence of genotype and treatment on the data structure was performed by hierarchical clustering and principal component analysis (PCA) of the log (base 2) transformed normalized counts. In the hierarchical heatmap it could be observed that all the samples showed a high degree of

correlation, greater than 0.9 and that the accession/treatment combination replicates are grouped together. The same situation was observed in the PCA plot for the first two components which account for 96% of the variance in the data, observing that the replicates of each combination of accession/treatment are close to each other and are well separated from the other combinations confirming that the experiment has a good reproducibility and quality and that batch effects between replicates are minimal compared to the differences due to genotype and treatment (Figure 43).

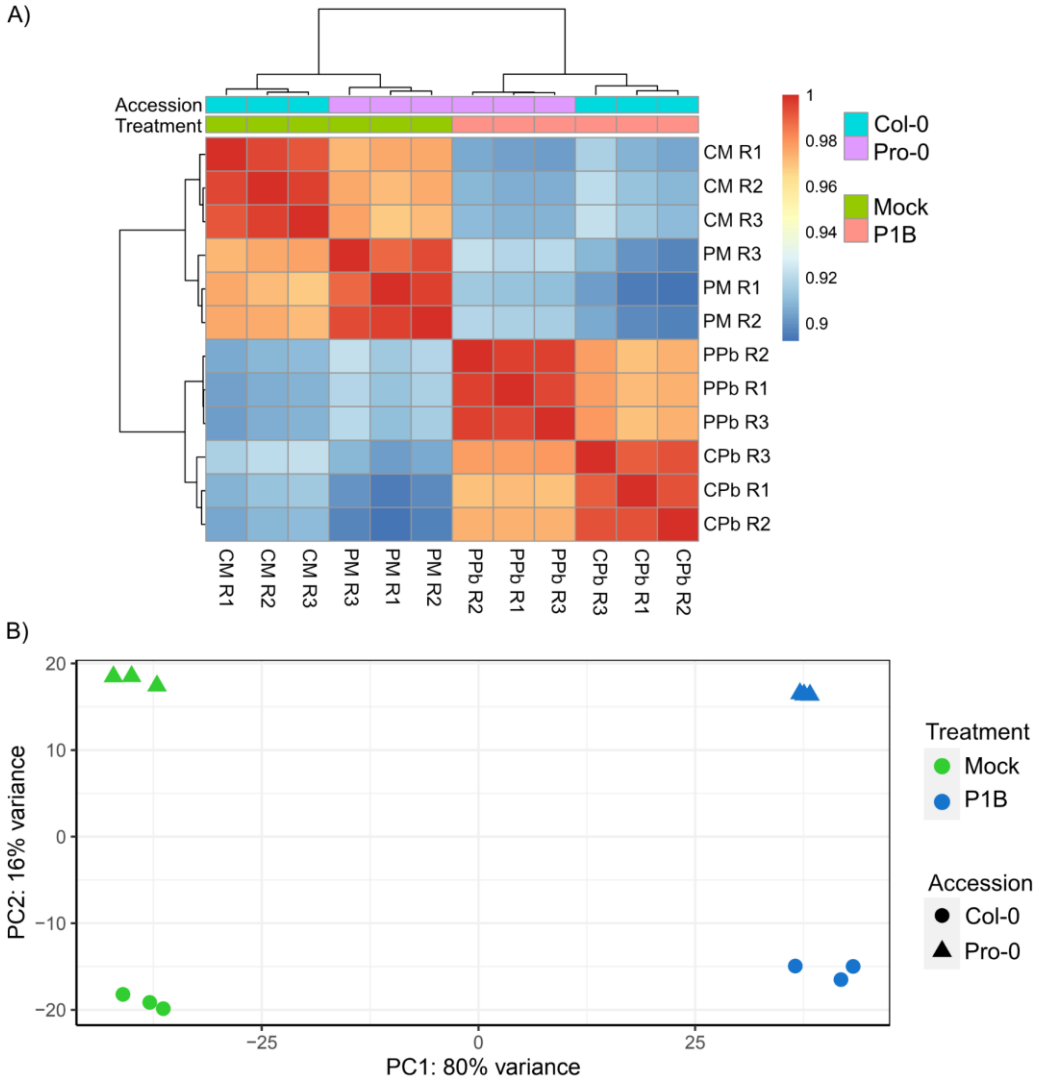


Figure 43. Quality assessment of the RNA-seq experiment using the normalized counts of reads mapped to the *A. thaliana* TAIR10 genome.
A) Hierarchical heatmap with computed correlations between samples. CM: Col-0 mock inoculated, PM: Pro-0 mock inoculated, CPb: Col-0 inoculated with *P. brassicae* P1B pathotype, PPb: Pro-0 inoculated with *P. brassicae* P1B pathotype B) PCA showing the clustering of biological replicates for each accession/treatment combination.

After the assessment of data quality and reproducibility, DEGs were identified with the DESeq2 algorithm in Col-0 and Pro-0 compared to their corresponding mock-inoculated controls. In the Table 10 the number of genes upregulated and downregulated in response to *P. brassicae* in each genotype are presented. The selection criteria for DEGs were defined by a $\log_2$ fold change ≥ 1 and a Benjamini Hochberg adjusted p-value < 0.05.

Table 10. Number of up and down regulated genes in response to *P. brassicae* infection in Pro-0 and Col-0 at 19 dpi.

Accession	Upregulated	Downregulated
Col-0	2284	2173
Pro-0	2267	2286

Comparing the upregulated and downregulated genes between the two accessions, there was an extensive overlap in the response of both up and downregulated genes in Col-0 and Pro-0, nevertheless, there were also DEGs identified that were exclusively present in only one of the accessions evaluated (Figure 44). To determine broad categories of the DEGs a gene ontology analysis was performed at the level of biological processes to determine which ones are over-represented among up- or down-regulated genes. The shared categories between both phenotypes of the upregulated genes were very general, including biological, cellular and metabolic processes as well as responses to biotic and abiotic stimuli, whereas the shared downregulated categories included cell wall organization and biogenesis, defense and immune responses, responses to biotic and abiotic stimulus and processes related to secondary metabolism (Figure 45 & Figure 46).

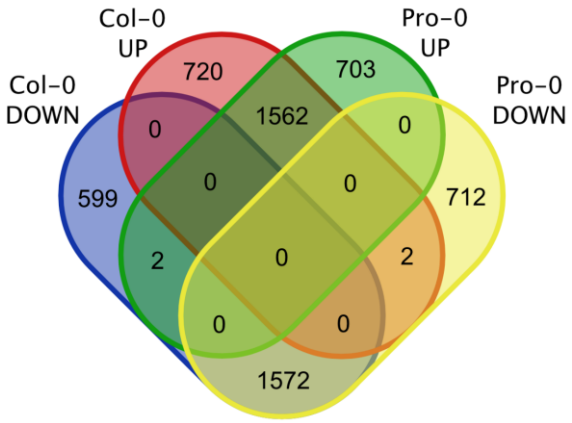


Figure 44. Venn diagram of DEGs at 19 dpi in the Arabidopsis accessions Col-0 and Pro-0.

In reference to the categories significantly over-represented in the contrast between accessions, two interesting patterns were observed in the repressed genes that could partially explain some of the divergences between the two accessions. In Col-0 exclusively, there was an over-representation of the downregulation of genes involved in cellular proliferation in the categories of cell division, microtubule-based movement and process and mitotic cytokinesis, indicating that the proliferative phase in Col-0 is completed, whereas in Pro-0 cell proliferation is still being stimulated by *P. brassicae* (Figure 46). On the other hand, in Pro-0 there is an over-representation of downregulated genes in the categories of cell growth, cell wall biogenesis, growth, pectin metabolic process, phenol-containing compound metabolic processes, phloem or xylem histogenesis, regulation of cell wall organization and biogenesis and tracheary element differentiation (Figure 46). Interestingly, most of those processes are related to cell enlargement growth and development, allowing us to hypothesize that in Pro-0 cell enlargement driven by gall development is only just starting, not as advanced as in Col-0, and the cell proliferation is not complete. This agrees with the hypothesis that the infection might be delayed in comparison with Col-0. All these observations concur with the microscopic observations at 19 dpi, where it is possible to observe an increased number of giant cells in Col-0 when compared with Pro-0 (Figure 41). To better visualize these observations, hierarchical clustering and heatmaps of the $\log_2$ fold change of selected DEGs found in the overrepresented categories related to cell cycle, cell growth and cell wall organization were generated. Expression profiles of these genes illustrate the pattern where, in Col-0 19 dpi the enlargement of the cells is predominant, unlike the situation observed in Pro-0, where cell proliferation is still ongoing (Figure 47). Concerning the cell cycle and cell division, downregulation of genes coding for cyclins or cyclin regulators including *CYCB1;2, CYCB2;4, CYCD3;1 CYCD6;1, MYB3R4* and *BUBR1* was observed in Col-0, as well as genes involved in the cytoskeleton dynamics of mitosis such as *CSLB01/03, KINESIN-12B, DRP5A, TUB7, HIK* and others (Figure 47A). On the other hand, the overrepresented categories of downregulated genes involved in cell wall dynamics in Pro-0, included genes such as *RWA2/3/4* involved in cell wall acetylation, *XTH6/15/21/33* that code for xyloglucan endotransglucosylase/hydrolases involved in cell wall modification, *XXT5* and *RGXT3* that code for a xyloglucan xylosyltransferases, *GMX2/3/MT1* that code for glucuronoxylan 4-O-methyltransferases, and other glycosyl transferases *GH9A1, GH9C3* and *GT18* (Figure 47B).

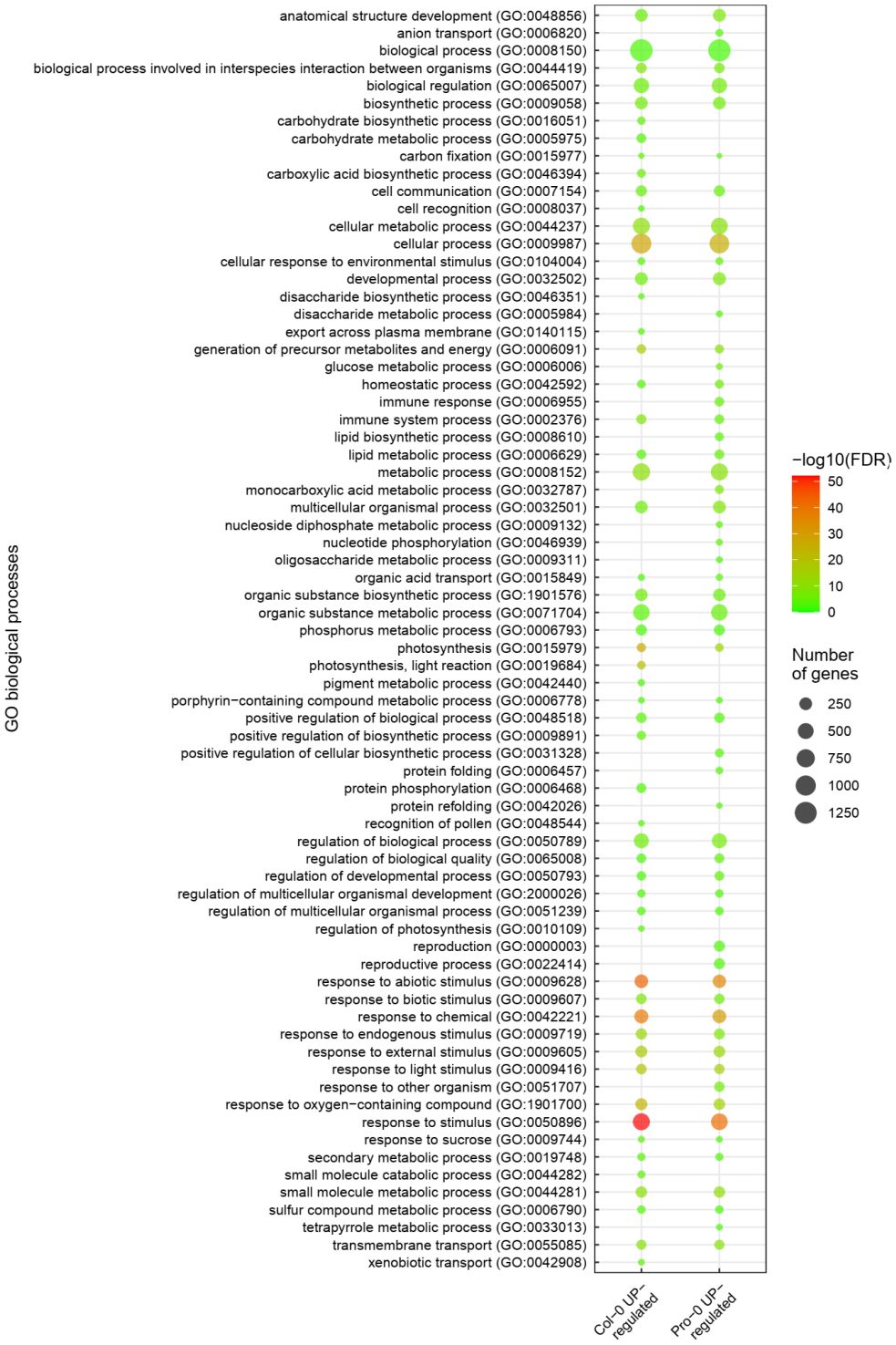


Figure 45. Over-represented biological processes among upregulated DEGs in Col-0 and Pro-0.
GO terms were obtained using the Panther classification system and selected with the FDR corrected p-values >0.05. The size of the dots represents the number of genes in each category and the color the -log10 of the Benjamini Hochberg corrected p-values.

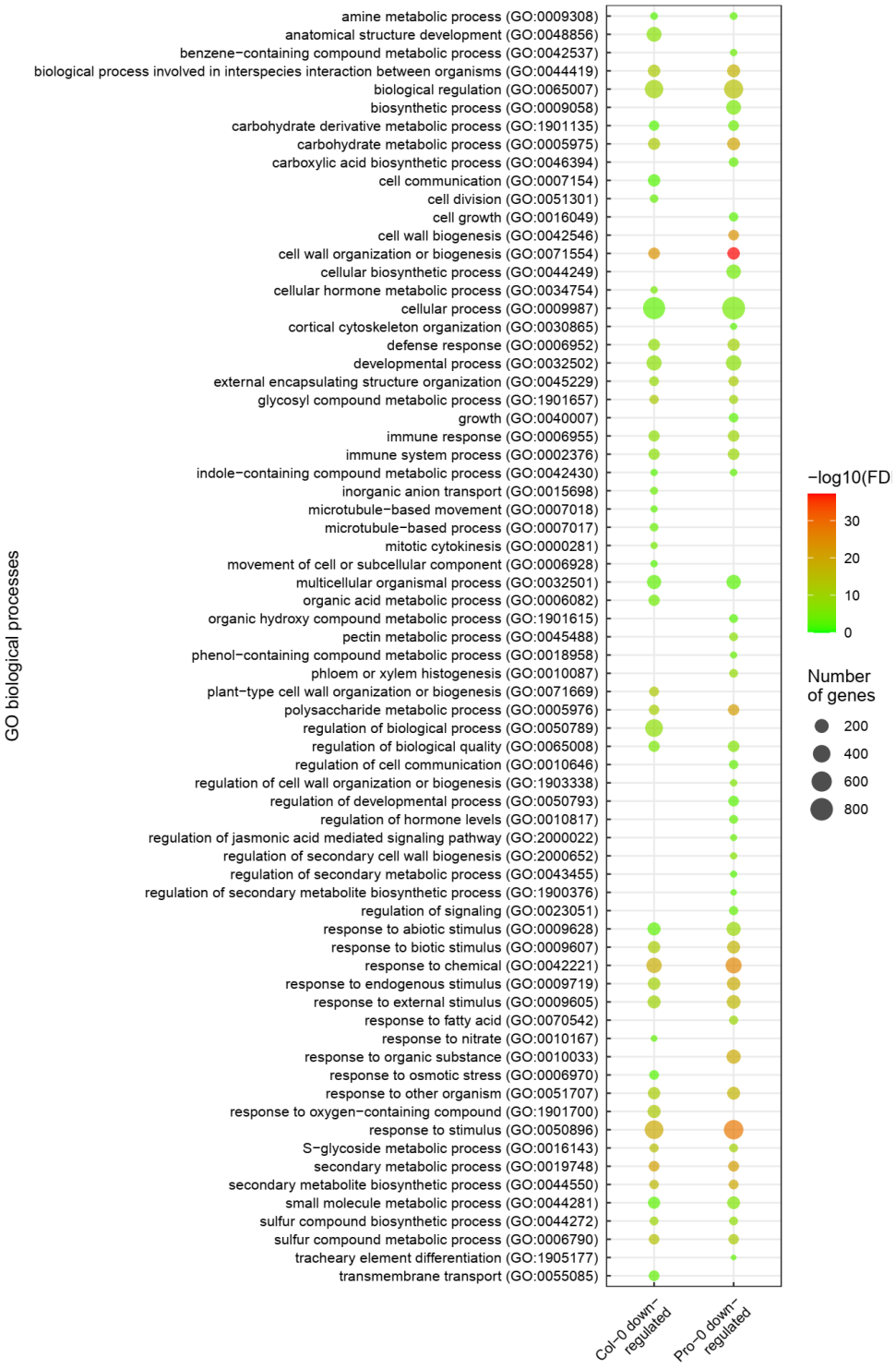


Figure 46. Over-represented biological processes among downregulated DEGs in Col-0 and Pro-0.
GO terms were obtained using the Panther classification system and selected with the FDR corrected p-values >0.05. The size of the dots represents the number of genes in each category and the color the -log10 of the Benjamini Hochberg corrected p-values.

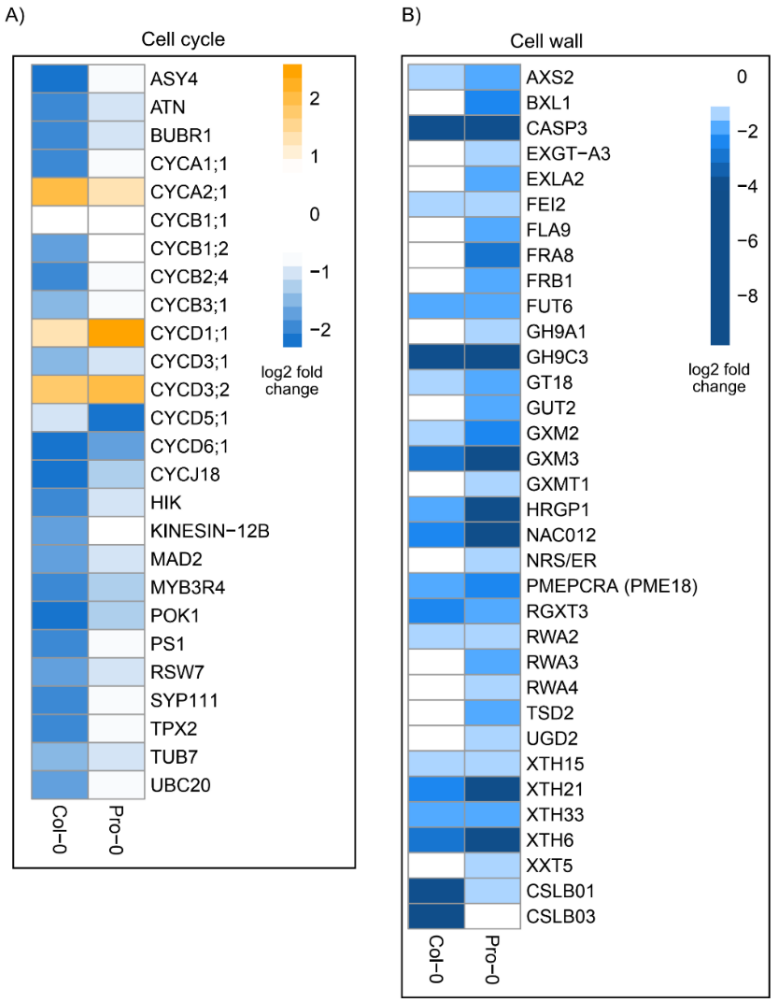


Figure 47. DEGs from the over-represented categories related to cell cycle, cell division and cell wall dynamics in Pro-0 and Col-0 at 19 dpi.
A) Genes related to cell cycle, B) Genes related to cell wall dynamics. The heatmaps were constructed with the log2 fold change of each gene relative to their corresponding mock inoculated control. The reference values are represented in the colored bar to the right of each heatmap.

Other overrepresented categories downregulated in Pro-0 were phloem or xylem histogenesis and tracheary element differentiation, which are directly related to the higher proportion of xylem observed Pro-0 hypocotyls. To visualize these differences a heatmap including DEGs involved in vascular cambium differentiation and xylogenesis was created, in which it could be observed that in Pro-0 there was a stronger downregulation of genes that code for proteins that promote or participate in xylem differentiation such as *VND2/5/6/7* involved in differentiation and metaxylem formation (Zhou et al., 2014), *IRX1/3/9/11* that are glycosyl transferases involved in secondary cell wall formation especially in the last steps of

xylem differentiation (Wu et al., 2010), *SND1/NST1/2* that participate in cell wall thickening in stem fibers (Zhong and Ye, 2015), *MYB46/MYB83* that positively regulate secondary cell wall formation and cell death (Ko et al., 2014), on the other hand, some genes that code for proteins that promote xylem differentiation such as *XVP/NAC005/NAC048* have a stronger upregulation in Col-0 (Figure 48) (Yang et al., 2020). The formation of new xylem bundles and vascular cambium differentiation has been observed in late stages of the disease in Col-0, despite the changes in its distribution during gall formation in the hypocotyl. Interestingly Malinowski et al., (2012), showed that the formation of new xylem occurs at the end of the secondary infection in both Col-0 and Ler, but in the former accession it arises earlier as does the disease progression.

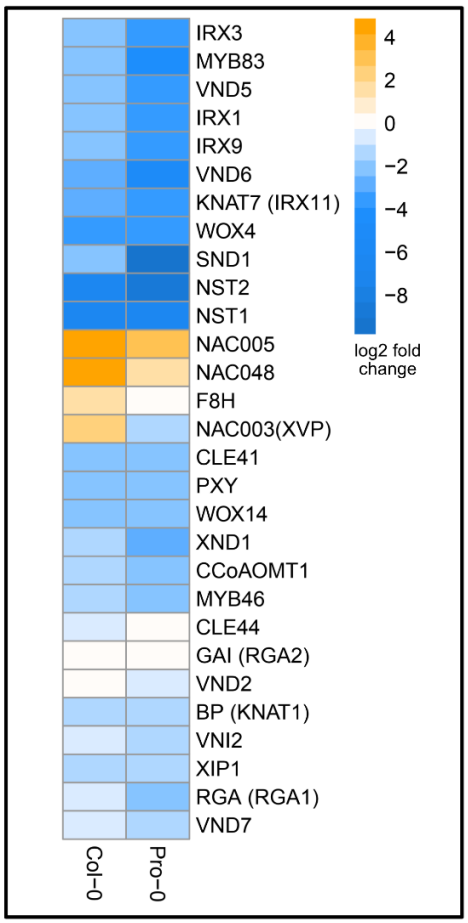


Figure 48. DEGs observed in the over-represented categories related to xylogenesis and vascular cambium differentiation.
The heatmap was constructed with the log2 fold change of each gene relative to their corresponding mock inoculated control. The reference values are represented in the colored bar to the right of the graphic.

One of the genes chosen that showed a strong upregulation in Col-0, but not in Pro-0 was *XVP* (*Xylem differentiation, disruption of Vascular Patterning*, *NAC003, At1g02220*). This factor was recently shown to be a key regulator of vascular cambium homeostasis and differentiation, reducing the proliferation of cambial cells through the inhibition of WOX4 and favoring xylem differentiation by positively regulating *VND6* activity (Yang et al., 2020). To test if the upregulation of *XVP* can be correlated with the more rapid *P. brassicae* colonization observed in the Col-0 hypocotyls, transgenic lines were created to overexpress the *XVP* gene under the 35S promoter in both Col-0 and Pro-0 plants. The 35S::*XVP* L11 in the Col-0 background was kindly donated by Professor Huanzhong Wang (Connecticut University, USA) together with the plasmid construct in the vector pK2GW7, that was used to generate *XVP* overexpressing lines in the Pro-0 genetic background. The 35S::*XVP* L11 is dwarfed with smaller leaves compared to the WT plants as previously described by Yang et al., (2020), nevertheless, in Pro-0, the general appearance of the plants overexpressing *XVP* were normal and no obvious differences in the phenotype were observed. To evaluate the effect of the overexpression of *XVP* in Pro-0 and Col-0 on clubroot disease the relative pathogen quantification and DI scoring of the transgenic lines in both genetic backgrounds was performed as described for previous experiments. Interestingly in Col-0 35S::*XVP* L11 the galls were smaller, but this could be related to the size of the plant itself (Figure 49). At the level of pathogen DNA quantification, no statistically significant differences were found (t-test $p = 0.57$) (Figure 50). On the other hand, in Pro-0, while no evident differences in symptoms were observed either in hypocotyls or rosettes, the *P. brassicae* DNA quantification did, however, show significantly higher quantities in the lines overexpressing *XVP* (Dunnett test L07 $p < 0.001$, L09 $p = 0.017$), indicating that the pathogen is replicating faster compared to the wild-type Pro-0 controls (Figure 50).

Figure 49. Phenotypes observed in the roots, hypocotyls, and rosettes of the 35S::*XVP* lines in Col-0 and Pro-0 accessions at 19 dpi with *P. brassicae* P1B.
The scale bar in the gall pictures represents 1 cm.

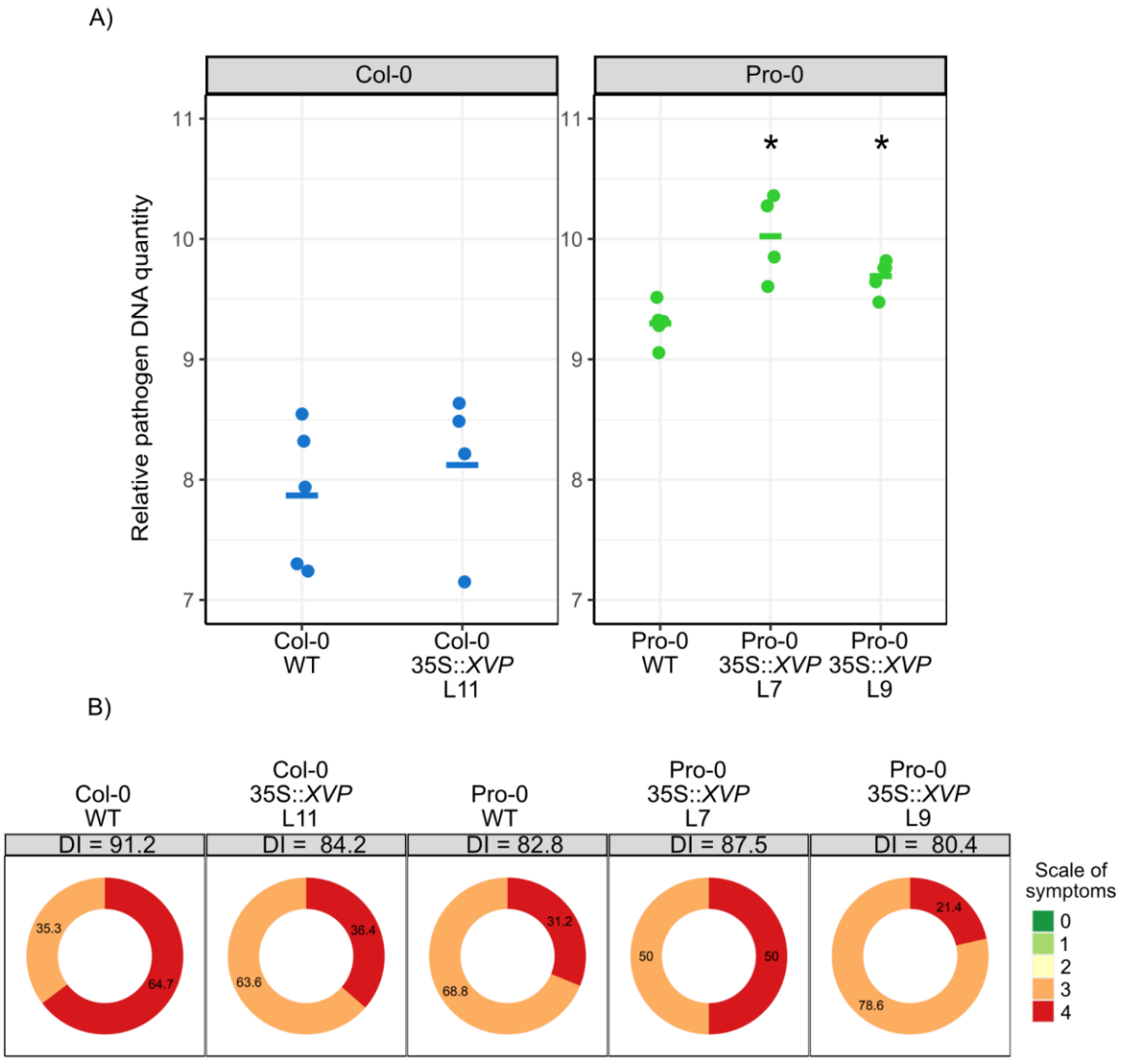


Figure 50. qPCR pathogen quantification and DI comparison between the clubroot infected 35S::*XVP* lines in Col-0 and Pro-0 at 19 dpi.
A) qPCR quantification of relative *P. brassicae* amounts presented in a $\log_2$ scale, the asterisks indicate significant differences with the respective wild type control evaluated with a T-Student's test for the Col-0 accession or with a Dunnett test in the case of Pro-0 accession ($4 \leq n \leq 5$). The colored horizontal bar represents the mean of each group of data. B) Disease index calculation and percentage of plants of individual genotypes classified according to the scale of symptoms observed in individual plants ($n \geq 11$). The calculated DI is presented in the grey boxes and the percentage of plants classified in each category of symptoms is represented in the donut charts.

5. Discussion

5.1 Exploiting Arabidopsis diversity through GWAS enabled the confirmation of *RPB1* as a principal component responsible for resistance to *P. brassicae*

The elucidation of disease resistance mechanisms in plants and discovery of resistance genes is crucial for breeding programs and crop production. Specifically in *Brassica spp.,* clubroot disease has gained importance becoming a considerable limitation, especially for oilseed rape growers. So far, numerous molecular markers associated to resistance and two NLRs have been identified and used to produce resistant cultivars in *Brassica spp.*, but new pathogen strains capable of breaking these are appearing, highlighting the importance of finding additional, more durable resistance mechanisms (Mehraj et al., 2020). To our advantage, clubroot disease also affects Arabidopsis, but the resistance mechanisms exhibited in this plant remain unknown. Previous endeavors to identify genes involved in resistance have been performed using biparental mapping populations with fully or partially resistant accessions, but they have only reached the stage of loci identification or proposition of candidate genes that could be involved, without any further functional validation (Fuchs and Sacristán, 1996; Gravot et al., 2011; Liégard et al., 2019a). In our study, by taking advantage of Arabidopsis natural variation and genetic resources, we were able to identify and validate *RPB1* as a major component of the resistance to our isolate of *P. brassicae* classified as pathotype P1B, moreover we were able to confirm its role by obtaining knock-out lines in resistant accessions, measuring the gene expression of defense markers and performing transient expression in *N. tabacum*.

With the screening of the reaction to our *P. brassicae* isolate, we observed different degrees of resistance or susceptibility across the population that may indicate that this trait is caused by multiple loci, albeit in our analyses only two SNPs showed significant associations. However, the absence of minor effect QTLs might be an effect of the size of the population, but also the stringency of the Bonferroni multiple testing correction, or the nature of the algorithms used to help to correct the population structure (Kuo, 2017). Nevertheless, the *rpb1* lines were not significantly different in terms of pathogen DNA amount when compared to the susceptible Col-0 controls, supporting the hypothesis that the resistance is caused by a major single gene (Figure 22). Another situation that might be reducing the power to detect other loci in our GWAS is the number of resistant accessions that we found, that is less than 10% of our population. In other studies with Arabidopsis,

different degrees of resistance or susceptibility have been reported, but in none of the cases have the genes involved been established, interestingly the percentage of accessions considered resistant was also less than 10%, and happened to be pathotype specific (Kobelt et al., 2000; Sharma et al., 2013; Jubault et al., 2013). It is also hypothesized that the minor differences observed between different accessions might be related to traits not necessarily related to pathogen resistance, but somehow affecting the pathogen adaptation, like the differences in xylem anatomy that were observed between Pro-0 and Col-0, that we believe could be related to differences in the degree of susceptibility.

A previous study also reported the locus *RPB1* using a biparental mapping population between the accessions Tsu-0 and Cvi-0 that overlaps with the region that we found (Fuchs and Sacristán, 1996; Arbeiter, 2002). It is also important to note that the pathotype used in those studies, known as “e”, was isolated in Germany from stubble turnip, meaning that despite the fact we are using different isolates, both come from the central Europe region and could be genetically closely related (Fuchs and Sacristán, 1996; Fähling et al., 2003). Fuchs and Sacristán (1996) also reported that the resistance was pathotype specific and conferred by a single gene by the segregation observed in the F1 and F2 of the crosses Tsu-0 × Cvi-0 and Ze-0 × Cvi-0, however the presence of minor effect QTLs was not considered.

5.2 *RPB1* is crucial for the resistance to *P. brassicae,* but does not exhibit the characteristics of a resistance gene

RPB1 codes for a protein without homologues with known function, or domains with known enzymatic activity; however, it has putative transmembrane domains, and the predicted cellular localization is the cell membrane, but this remains to be experimentally validated. With the evidence that we collected in this study, we propose that *RPB1* is crucial for the mounting of a successful immune response to *P. brassicae*, but its characteristics do not resemble the ones found in the well characterized NLR immune receptors, where a higher degree of sequence variability is observed in comparison with other proteins, probably due to diversifying or balancing selection imposed by the genetic variability of the pathogens (Han, 2019). Contrary to this, the CDS of *RPB1* is highly conserved among different accessions though in some susceptible accessions it is completely absent. Conversely, we observed a high degree of sequence variability in the promoter region, which

could be a reason to think that *RPB1* might not be induced upon inoculation in susceptible accessions that contain it. Nonetheless, further studies will require to measure the *RPB1* expression in susceptible accessions and obtain and analyze the sequence of the promoter region in the accessions that contain *RPB1* to confirm this hypothesis. Another difficulty that we found is the presence of multiple copies in tandem of *RPB1* and *RPB1*-like genes in the vicinity of *RPB1*, despite this, the use of accessions containing single copies together with the elaboration of single and multiple knock-out lines via CRISPR/Cas9, proved to be a powerful tool to discriminate the *RPB1* function from the other of *RPB1*-like genes, interestingly in our study *RPB1-like-4* was significantly upregulated at 7 dpi in response to infection, but was shown not to be involved in the resistance to *P. brassicae*.

An observation that supports the hypothesis that *RPB1* does not participate in the recognition of *P. brassicae,* is that the gene is upregulated at 7 dpi in both *RPB1* and *rpb1* null lines in Uod-1 and Est-1 accessions, which means that there should be other protein(s) upstream that can mediate pathogen recognition and trigger its upregulation. Unfortunately, because *P. brassicae* is a root pathogen with an intracellular lifestyle, we face an additional level of difficulty to predict where the pathogen recognition can occur, since both transmembrane and intracellular receptors require association with other proteins to trigger an immune response, this interaction can happen extracellularly, in the membrane or in the cytoplasm (Cesari, 2018). During most of the secondary infection, the pathogen is probably enclosed in a vacuole structure inside the cell, but the mechanisms of penetration and transport to and from the host are completely unknown. Despite the fact that we do not know exactly how the effectors of *P. brassicae* are secreted, some of them have experimentally validated functions that act in the host cytoplasm like the conversion of salicylic acid to methylsalycylate (Ludwig-Müller et al., 2015), the targeting of endoplasmic reticulum, Golgi bodies or nuclear cytoplasm (Hossain et al., 2021), or MAPKKK activity to promote ROS production and HR (Jin et al., 2021). However, effectors targeting the apoplastic space such as SSPbP53 or the plasma membrane like PbPE13 have been reported too (Hossain et al., 2021; Pérez-López et al., 2021).

The regulation of the expression of *RPB1* also appeared to be crucial for RPB1 function, in our study we observed that the mock inoculated plants have a very low basal expression, particularly in Uod-1, where the mRNA levels were undetectable (Figure 24). Upon inoculation with *P. brassicae,* we observed an upregulation of *RPB1* expression of about 60-fold in Est-1 at 7 dpi, which could mean that transcription needs to be strictly

regulated to avoid uncontrolled activation of defense responses. This is a feature shared with some R-genes, as they also tend to be in a low abundance or inactivated in the absence of a pathogen since aberrant expression can lead to autoimmunity, affecting plant fitness (Wersch et al., 2020). By transiently expressing *RPB1* in *N. tabacum,* we observed a strong HR, meaning that its sole expression is sufficient to trigger defense responses even if *P. brassicae* is absent, allowing us to infer that *RPB1* positively regulates the defense against *P. brassicae* but its constitutive overexpression may lead to autoimmunity (Figure 30). In some cases, it has been observed the NLRs are regulated at the transcriptional level either by epigenetic factors or through the presence of promoter elements, for instance the *RPP8* gene of Arabidopsis that confers resistance to *H. arabidopsidis* required the presence of W-boxes in the promoter to confer immunity, however, in our biological system a more detailed comparative analysis in the sequence of the promoter would be necessary to identify cis elements that could be involved in the transcription of *RPB1* (Stokes et al., 2002; Lai and Eulgem, 2018). The influence of epigenetic regulation of the resistance to *P. brassicae* has also been investigated using the *ddm1* mutant, which shows increased resistance, but the genes affected by the epigenetic regulation in the resistance to *P. brassicae* remain uncharacterized (Liégard et al., 2019a). In future research, it will be important to express *RPB1* in a variety of susceptible accessions under the native promoter of *RPB1* obtained from resistant accessions*,* to observe if it is possible to confer resistance in susceptible accessions, it would also be important to elucidate which transcription factors regulate *RPB1* expression and if any epigenetic factors are involved in its repression.

5.3 Slower colonization of Pro-0 tissues is probably related to differences in the xylem anatomy and the vascular cambium homeostasis and differentiation.

In the second part of this research, we focused on explaining the differences observed between the susceptible accessions Col-0 and Pro-0, at both the anatomical and transcriptomic level. We selected Pro-0 because, interestingly, it showed an atypically high pathogen DNA quantity compared to the other susceptible accessions, but the symptoms observed in galls and rosettes were less strong (Figure 37 & Figure 38). The comparison at the anatomical level, showed that in Pro-0 the central xylem cylinder at the center of the hypocotyl occupied a higher proportion of the hypocotyl area than in Col-0, moreover the disturbance of the vascular bundles following clubroot infection was less severe, meaning

that, despite the higher relative pathogen DNA amount, the infection in Pro-0 appeared to be progressing more slowly (Figure 39 & Figure 41). The repression of xylem differentiation and the maintenance of the meristematic cells in an undifferentiated stage have been proved to be beneficial for pathogen development (Malinowski et al., 2012), which could indicate that the presence of more fully differentiated xylem slows down the colonization of the host tissue, at least in the early secondary infection. Interestingly, a recent paper showed that *P. brassicae* secretes an apoplastic effector called SSPbP53 capable of interfering with the cysteine-protease activity of Arabidopsis Xylem Cysteine Peptidase 1 (XCP1) which is involved in the autolysis of tracheary elements that will result in the programmed cell death of xylem cells, as the final step of xylem differentiation; this is the first evidence supporting that the idea *P. brassicae* actively interferes with host xylem differentiation (Pérez-López et al., 2021).

We were also able to identify differences at the transcriptional level supporting the microscopic observations in Pro-0 and Col-0, since some genes involved in xylem formation and differentiation tend to be more strongly downregulated in Pro-0 than in Col-0 such as *VND6/7*, *RGA*, *XIP1, SND1/NST1/2, MYB46/83* and *IRX1/3/9/11* or upregulated specifically in Col-0 like *XVP* (*NAC003*) and *NAC048* transcription factors (Figure 48). This observation indicates that at 19 dpi in Pro-0 the differentiation of xylem is more repressed than in Col-0. Malinowski et al., (2012) observed the formation of disorganized xylem bundles towards the end of the secondary infection in Col-0, despite the fact that genes involved in xylem maturation were downregulated, which may be the explanation for this observation, as the infection in Pro-0 tends to progress slower and the formation of new disorganized bundles has not started yet, however this needs to be confirmed with histological studies at additional time points.

One of the genes that caught our attention most strikingly was *XVP* because it is a key regulator of the vascular cambium homeostasis, and was significantly upregulated in Col-0, but in Pro-0 the trend for expression was of downregulation, though not significantly (Col-0 $\log_2$ fold change = 2.2, padj (BH) = 7.49E-13, Pro-0 $\log_2$ fold change = -1.1, padj (BH) = 1). The explanation of this may be that Pro-0 has not finished completely the proliferative stage and the main function of XVP is the regulation of the cambium cell proliferation. When XVP is overexpressed, there is a reduction in the number of cambial cells and a precocious xylem differentiation, because XVP negatively regulates WOX4 and indirectly induces VND6, which is the main regulator of metaxylem formation (Yang et al., 2020). This means

that the downregulation of *XVP* could help to maintain the meristem in an undifferentiated state during the proliferative phase in the vascular cambium, however, a more detailed expression profile at different time points and the observation using GUS or fluorescent reporters may help to clarify this assumption. We observed that the 35S::*XVP* lines in Pro-0 had relative higher pathogen DNA amounts compared to wild-type Pro-0 at 19 dpi, this could be explained by an overall decrease of xylem vessels as previously observed in Col-0 35S::*XVP,* that exhibited a reduced absolute amount of xylem, despite its proportion relative to undifferentiated meristem being higher than in wild-type Col-0 (Yang et al., 2020), however, the relative pathogen amount remained unchanged in Col-0 lines overexpressing *XVP*. Comparison of the anatomical features resulting from ectopic expression of *XVP* in Pro-0 and Col-0 may account for these differences.

Another group of genes that could have an influence in the differences at xylem anatomy are the DELLA proteins, which are conserved gibberellic acid (GA) signaling repressors. DELLA degradation is promoted by GA which travels from shoots to hypocotyls once flowering starts, initiating the xylem expansion in the hypocotyl (Ben-Targem et al., 2021). The xylem expansion of the hypocotyl is the second phase in the vascular cambium development that is related to flowering, prior to the xylem expansion the xylem and phloem are produced at the same rate, but in the xylem expansion phase the production of xylem fibers is favored, the meristem starts the process of senescence and there is a repression in phloem differentiation (Ben-Targem et al., 2021). The main two DELLAS participating in this process are *RGA* and *GAI* that interact with the proteins BP(KNAT1) or the *AUXIN RESPONSIVE FACTORS* ARF6/8 that control xylem expansion (Ben-Targem et al., 2021). Pro-0 has an earlier flowering time than Col-0, which means that it might start the process of xylem expansion earlier, moreover in our transcriptome we observed that *RGA* is significantly downregulated in Pro-0 (log_2 fold change = -1.78, padj (BH) = 2.69E-13), but not in Col-0 (log_2 fold change = -0.89, padj (BH) = 1), which could perhaps partially explain the higher proportion of xylem bundles in the hypocotyl. Nevertheless, further histological and functional studies with knock-out lines and/or overexpressing lines of genes involved in the RGA/GAI/BP/ARF6/8 signaling pathway, and its relation to GA will be required to test our observations.

With the GO analysis we identified two other groups of genes that suggest that Pro-0 has a delayed infection compared to Col-0 genes involved in the cell cycle and in cell wall dynamics. Concerning the cell cycle, we observed some cyclin genes that were significantly

downregulated in Col-0, but not in Pro-0, for example *CYCD3;1* (Col-0 log_2 fold change = -1.51, padj (BH) = 1.54E-13, Pro-0 log_2 fold change = -1.13, padj (BH)= 0.97), are involved in cambial cell proliferation (Collins et al., 2015), and could be a good indicator that the proliferative state in Col-0 has already passed, however previous studies at the transcriptomics and protein level at 16 dpi found that *CYCD3;1* does not participate during the proliferative state (Olszak et al., 2019), nevertheless this discordance could be attributed to the different time point of the analysis and the use of a different *P. brassicae* isolate. Another groups of cyclins that showed significant downregulation only in Col-0 are the B-type cyclins *CYCB1;1*, *CYCB1;2*, *CYCB2;4* and *CYCB3;1* that participate in the G2/M transition and were previously found to be upregulated at 16 dpi. The downregulation of B-type cyclins correlates with the significantly lower expression in Col-0 of *MYB3R4* that codes for a transcription factor which stimulates the G2/M transition by activating expression of B-type cyclins and other cell cycle components (Olszak et al., 2019). Most of the rest of the DEGs downregulated in Col-0 but not in Pro-0 in this group of genes associated with the cell cycle participate in the cytoskeletal or cell wall dynamics during cell divisions.

Our transcriptomic analysis also showed that genes coding for enzymes related to cell wall modification were significantly downregulated in Pro-0, but not in Col-0. It has been shown through transcriptomic analysis in *Brassica* spp. that genes involved in the synthesis of cellulose, lignin and pectin are downregulated during clubroot infection and it is considered that during the cell enlargement there is a requirement for a reduction in the rigidness of the cell wall to facilitate the expansion of *P. brassicae* colonized cells (Badstöber et al., 2020). In our data two cellulose synthases, *CSLB01/03* were strongly downregulated in Col-0, but not in Pro-0, indicating that the cell enlargement phase in Col-0 is probably more advanced. Pro-0 exhibited stronger repression of genes that code for enzymes involved in metabolism and modification of the cell wall such as *RWA2/3/4*, *XTH6/15/21/33*, *XXT5, RGXT3*, *GMX2/3/MT1*, *GH9A1, GH9C3* and *GT18*. Out of these genes, the *XTHs* that code for enzymes with xyloglucan endotransglucosylase/hydrolase activity have been linked to cell wall extensibility during clubroot disease (Devos et al., 2004), indicating that the stronger repression of these genes in Pro-0 could be related to delays in disease progression. Nevertheless, it is important to consider that is not known whether the pathogen actively promotes cell wall modifications in the host or if they are caused by the turgor pressure imposed by pathogen proliferation.

Our results indicate that clubroot disease progression in Pro-0 is delayed when compared to Col-0 despite the relative quantification of the pathogen DNA. We hypothesize that this might be related to the higher proportion of xylem observed in the hypocotyl, because the xylem tissues are lignified and have undergone programmed cell death, making it harder for a biotrophic pathogen that depends on living host cells to colonize the plant tissues. The transcriptomics experiment results also revealed that Col-0 presents downregulation of genes involved in the cell cycle at 19 dpi, indicating that the proliferative state of the disease has ended, whereas in Pro-0 some cyclin genes are still upregulated in response to infection or are not yet downregulated as they are in Col-0. In Pro-0 there was downregulation of genes involved in cell wall dynamics, supporting other evidence that in Pro-0 at this time point the *P. brassicae* has not yet fully triggered the host cell enlargement and hypertrophy phase because the disease progression is slower, compared to that observed in Col-0, where cell enlargement is occurring and the genes involved in the dynamics of the cell wall are not downregulated but upregulated, which is a characteristic of the late secondary infection (Stefanowicz et al., 2021; Badstöber et al., 2020) .

The overexpression of *XVP,* one of the key regulators of vascular cambium homeostasis*,* seems to favor pathogen replication in the plant tissues as measured by qPCR in the Pro-0 background, but further experiments including microscopic observations, visualization of *XVP* promoter activity and studies in additional accessions with similar xylem distribution will be needed to confirm the relationship between xylogenesis, cambium differentiation and disease progression in different susceptible Arabidopsis accessions.

Together, these results suggest that the development of cultivars which are resistant to clubroot, or less severely affected by infection, need not be restricted to the search for immune receptors, but should also address factors underpinning host developmental plasticity and anatomic features that are disadvantageous to the pathogen. This is especially true in crops such as *B. napus* where the genetic diversity available may be narrow. While such features are not likely to provide full resistance to the disease, they might confer tolerance to a broader spectrum of isolates and more durability to the assault of highly dynamic pathogen populations.

5.4 Summary of the discussion

By making use of the natural diversity of Arabidopsis and its genetics resources, we performed a GWAS evaluating the resistance to *P. brassicae* pathotype P1B in 142 accessions, allowing us to identify two loci potentially responsible of the phenotype, the first between *At1g32030* and *At1g32100* and the second in a TIR-NB-LRR gene known as *RAC1*. The region between the genes *At1g32030* and *At1g32100* showed high structural variability with indels and presence of transposable elements, and various CDS that code for RPB1 and other RPB1-like proteins, however, only *RPB1* and *RPB1-like-4*, were expressed in roots and upregulated at 7 dpi (Figure 13 & Figure 17). Knock-out lines of these two genes were produced through the CRISPR/Cas9 in the resistant Arabidopsis accessions Uod-1 and Est-1. The phenotyping results demonstrated that the *rpb1*, but not the *rpb1-like-4* mutants completely lost resistance, accumulating high contents of pathogen DNA, and undergoing similar morphological changes compared to the susceptible accessions and that inspection of the galls revealed high loads of resting spores and secondary plasmodia (Figure 22 & Figure 26). These results allowed us to confirm *RPB1* as a crucial component of resistance to *P. brassicae* in Arabidopsis.

The *rpb1* lines did not exhibit upregulation of defense gene markers upon inoculation with *P. brassicae*, contrary to the situation observed in the wild-type controls of Uod-1 and Est-1. However, transcription of the *RPB1* gene continued to be upregulated despite the absence of a functional RPB1 protein, allowing us to hypothesize that the pathogen must still be perceived, but that the proper defense responses cannot be activated (Figure 23 & Figure 24). To further confirm these results, we elaborated transient expression mediated by *Agrobacterium tumefaciens* of *RPB1* in *Nicotiana tabacum* leaves under the strong promoter CaMV35S, and we found that it caused a strong HR, providing us additional evidence to hypothesize that the upregulation of *RPB1* can trigger defense responses without the presence of the *P. brassicae* (Figure 30). The function of *RAC1* was also evaluated through the generation of *rac1* knock-out lines in the resistant accessions Uod-1 and Est-1 using CRISPR/Cas9. However, the mutants did not show loss of resistance to *P. brassicae*, indicating that *RAC1* is not required for the resistance phenotype.

To further investigate possible causes of differences in susceptibility, we compared the accessions Pro-0 and Col-0 making use of histological and transcriptomics methods. We chose Pro-0 because it contained higher relative pathogen DNA quantities but exhibited less severe symptoms than those observed in Col-0, such as healthier rosettes, reduced

impact on xylem formation and less swollen hypocotyls. Interestingly, at the anatomical level we determined that Pro-0 has a higher proportion of xylem bundles in non-infected plants and the xylem anatomy is not as severely affected as it is in Col-0 at 19 dpi, which correlates with the symptoms of wilting observed in the rosettes. In a comparative transcriptomics analysis, we obtained additional evidence supporting the idea that the progress of the disease is faster in Col-0, as it exhibits downregulation of genes involved in the cell cycle, meaning that at 19 dpi the proliferative phase of the disease has passed. On the other hand, Pro-0 showed downregulation of genes involved in cell wall reorganization, which indicates that it has not completely advanced to the cell enlargement phase, typical of the late stages of the disease. We also identified one group of NAC transcription factors involved in vascular cambium homeostasis and xylem/phloem differentiation that are upregulated in Col-0, but not in Pro-0. To further analyze this finding, we overexpressed the NAC transcription factor *XVP,* a key regulator of vascular cambium differentiation, under the CaMV35S promoter in both accessions and found that only in the Pro-0 background was a higher relative pathogen DNA amount observed when compared to their respective control, suggesting that the differences observed in the xylem anatomy might be partially related to different degrees of clubroot susceptibility.

6. Conclusions

- Natural variation present in Arabidopsis is sufficient for the GWAS–based identification of new components of immunity to clubroot disease
- Despite the fact that there are additional homologues of *RPB1* in the *P. brassicae* resistant Arabidopsis accessions, genetic characterization shows that *RPB1* alone plays the critical role for immunity to clubroot disease.
- *RPB1* does not resemble any known immune receptors or defense signaling components, but its function is probably the positive regulation of defense responses essential for restricting *P. brassicae* growth.
- The difference in clubroot susceptibility between Arabidopsis accessions Col-0 and Pro-0 is likely explained by the variations observed in the xylem and vascular cambium activity.
- Anatomical differences observed in Pro-0 accession may be at least partially attributed to different transcription patterns of *XVP* and related transcription factors controlling key differences in vascular development.

7. Figures

8. Tables

9. References

Aerts, N., Mendes, M.P., and Wees, S.C.M. Van (2021). Multiple levels of crosstalk in hormone networks regulating plant defense. Plant J. **105**: 489–504.

Alix, K., Lariagon, C., Delourme, R., and Manzanares-Dauleux, M.J. (2007). Exploiting natural genetic diversity and mutant resources of Arabidopsis thaliana to study the A. thaliana-Plasmodiophora brassicae interaction. Plant Breed. **126**: 218–221.

Alonso-Blanco, C. et al. (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell **166**: 481–491.

Andersen, C.L., Jensen, J.L., and Ørntoft, T.F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. **64**: 5245–5250.

Aranzana, M.J. et al. (2005). Genome-Wide Association Mapping in Arabidopsis Identifies Previously Known Flowering Time and Pathogen Resistance Genes. PLOS Genet. **1**: e60.

Arbeiter, A. (2002). Hochauflösende Kartierung und molekulare Analyse der Region um den Resistenzlocus RPB1 auf Chromosom 1 von Arabidopsis thaliana (L.) Heynh.

Arbeiter, A., Fähling, M., Graf, H., Sacristán, M.D., and Siemens, J. (2002). Resistance of Arabidopsis thaliana to the obligate biotrophic parasite Plasmodiophora brassicae. Plant Prot. Sci. **38**: 519–522.

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

Badstöber, J., Ciaghi, S., and Neuhauser, S. (2020). Dynamic cell wall modifications in brassicas during clubroot disease. bioRxiv: 2020.03.02.972901.

Ben-Targem, M., Ripper, D., Bayer, M., and Ragni, L. (2021). Auxin and gibberellin signaling cross-talk promotes hypocotyl xylem expansion and cambium homeostasis. J. Exp. Bot. **72**: 3647–3660.

Bieluszewski, T. et al. (2019). NuA4-dependent H4 acetylation is adopted to specifically control photosynthesis-related genes. bioRxiv.

Boller, T. and Felix, G. (2009). A renaissance of elicitors: Perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annu. Rev. Plant Biol. **60**: 379–407.

Bonnot, T., Gillard, M., and Nagel, D. (2019). A Simple Protocol for Informative Visualization of Enriched Gene Ontology Terms. Bio-Protocol **9**.

Borhan, M.H., Holub, E.B., Beynon, J.L., Rozwadowski, K., and Rimmer, S.R. (2004). The arabidopsis TIR-NB-LRR gene RAC1 confers resistance to Albugo candida (white rust) and is dependent on EDS1 but not PAD4. Mol. Plant-Microbe Interact. **17**: 711–719.

Botanga, C.J., Bethke, G., Chen, Z., Gallie, D.R., Fiehn, O., and Glazebrook, J. (2012). Metabolite profiling of Arabidopsis inoculated with Alternaria brassicicola reveals that ascorbate reduces disease severity. Mol. Plant-Microbe Interact. **25**: 1628–1638.

Breitenbach, H.H. et al. (2014). Contrasting Roles of the Apoplastic Aspartyl Protease APOPLASTIC, ENHANCED DISEASE SUSCEPTIBILITY1-DEPENDENT1 and LEGUME LECTIN-LIKE PROTEIN1 in Arabidopsis Systemic Acquired Resistance ,. Plant Physiol. **165**: 791–809.

Buczacki, S.T., Toxopeus, H., Mattusch, P., Johnston, T.D., Dixon, G.R., and Hobolth, L.A. (1975). Study of physiologic specialization in Plasmodiophora brassicae: Proposals for attempted rationalization through an international approach. Trans. Br. Mycol. Soc. **65**: 295–303.

Bulman, S., Richter, F., Marschollek, S., Benade, F., Jülke, S., and Ludwig-Müller, J. (2019). Arabidopsis thaliana expressing PbBSMT, a gene encoding a SABATH-type methyltransferase from the plant pathogenic protist Plasmodiophora brassicae, show leaf chlorosis and altered host susceptibility. Plant Biol. **21**: 120–130.

van der Burgh, A.M. and Joosten, M.H.A.J. (2019). Plant Immunity: Thinking Outside and Inside the Box. Trends Plant Sci. **0**: 587–601.

Castel, B., Ngou, P.-M., Cevik, V., Redkar, A., Kim, D.-S., Yang, Y., Ding, P., and Jones, J.D.G. (2019). Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. New Phytol. **222**: 966–980.

Cesari, S. (2018). Multiple strategies for pathogen perception by plant immune receptors.

New Phytol. **219**: 17–24.

Chang, A., Lamara, M., Wei, Y., Hu, H., Parkin, I.A.P., Gossen, B.D., Peng, G., and Yu, F. (2019). Clubroot resistance gene Rcr6 in Brassica nigra resides in a genomic region homologous to chromosome A08 in B. rapa. BMC Plant Biol. **19**: 224.

Chen, H.G. and Wu, J.S. (2008). Characterization of fertile amphidiploid between Raphanus sativus and Brassica alboglabra and the crossability with Brassica species. Genet. Resour. Crop Evol.

Chu, M., Yu, F., Falk, K.C., Liu, X., Zhang, X., Chang, A., and Peng, G. (2013). Identification of the clubroot resistance gene Rpb1 and introgression of the resistance into canola breeding lines using a marker-assisted approach. Acta Hortic. **1005**: 599–606.

Clough, S.J. and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. **16**: 735–743.

Collins, C., Maruthi, N.M., and Jahn, C.E. (2015). CYCD3 D-type cyclins regulate cambial cell proliferation and secondary growth in Arabidopsis. J. Exp. Bot. **66**: 4595–4606.

Concordet, J.P. and Haeussler, M. (2018). CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. **46**: W242–W245.

Couto, D. and Zipfel, C. (2016). Regulation of pattern recognition receptor signalling in plants. Nat. Rev. Immunol. 2016 169 **16**: 537–552.

Czubatka-Bieńkowska, A., Kaczmarek, J., Marzec-Schmidt, K., Nieróbca, A., Czajka, A., and Jędryczka, M. (2020). Country-wide qpcr based assessment of plasmodiophora brassicae spread in agricultural soils and recommendations for the cultivation of brassicaceae crops in Poland. Pathogens **9**: 1–17.

Dakouri, A., Zhang, X., Peng, G., Falk, K.C., Gossen, B.D., Strelkov, S.E., and Yu, F. (2018). Analysis of genome-wide variants through bulked segregant RNA sequencing reveals a major gene for resistance to Plasmodiophora brassicae in Brassica oleracea. Sci. Rep. **8**: 1–10.

Dangl, J.L. and Jones, J.D.G. (2019). A pentangular plant inflammasome. Science (80-.

). **364**: 31–32.

Devos, S., Laukens, K., Deckers, P., Van Der Straeten, D., Beeckman, T., Inzé, D., Van Onckelen, H., Witters, E., and Prinsen, E. (2006). A hormone and proteome approach to picturing the initial metabolic events during Plasmodiophora brassicae infection on Arabidopsis. Mol. Plant-Microbe Interact. **19**: 1431–1443.

Devos, S., Vissenberg, K., Verbelen, J.-P., and Prinsen, E. (2004). Infection of Chinese cabbage by Plasmodiophora brassicae leads to a stimulation of plant growth: impacts on cell wall metabolism and hormone balance. New Phytol. **166**: 241–250.

Diederichsen, E., Beckmann, J., Schondelmeier, J., and Dreyer, F. (2006). Genetics of clubroot resistance in Brassica napus “Mendel.” Acta Hortic.

Diederichsen, E., Frauen, M., Linders, E.G.A., Hatakeyama, K., and Hirai, M. (2009). Status and Perspectives of Clubroot Resistance Breeding in Crucifer Crops. J. Plant Growth Regul. **28**: 265–281.

Diederichsen, E. and Sacristan, M.D. (1996). Disease response of resynthesized Brassica napus L. lines carrying different combinations of resistance to Plasmodiophora brassicae Wor. Plant Breed.

Dixon, G.R. (2006). The biology of Plasmodiophora brassicae Wor. - A review of recent advances. In Acta Horticulturae.

Dixon, G.R. (2009). The occurrence and economic impact of plasmodiophora brassicae and clubroot disease. J. Plant Growth Regul.

Djavaheri, M., Ma, L., Klessig, D.F., Mithöfer, A., Gropp, G., and Borhan, H. (2019). Mimicking the host regulation of salicylic acid: A virulence strategy by the clubroot pathogen plasmodiophora brassicae. Mol. Plant-Microbe Interact. **32**: 296–305.

Dong, O.X., Tong, M., Bonardi, V., El Kasmi, F., Woloshen, V., Wünsch, L.K., Dangl, J.L., and Li, X. (2016). TNL-mediated immunity in Arabidopsis requires complex regulation of the redundant ADR1 gene family. New Phytol. **210**: 960–973.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics **32**: 3047–3048.

Fähling, M., Graf, H., and Siemens, J. (2003). Pathotype Separation of Plasmodiophora brassicae by the Host Plant. J. Phytopathol. **151**: 425–430.

Ferrero-Serrano, Á. and Assmann, S.M. (2019). Phenotypic and genome-wide association with the local environment of Arabidopsis. Nat. Ecol. Evol. **3**: 274–285.

Fredua-Agyeman, R. and Rahman, H. (2016). Mapping of the clubroot disease resistance in spring Brassica napus canola introgressed from European winter canola cv. 'Mendel.' Euphytica.

Friberg, H., Lagerlöf, J., and Rämert, B. (2005). Germination of Plasmodiophora brassicae resting spores stimulated by a non-host plant. Eur. J. Plant Pathol. 2005 1133 **113**: 275–281.

Fu, H., Yang, Y., Mishra, V., Zhou, Q., Zuzak, K., Feindel, D., Harding, M.W., and Feng, J. (2020). Most plasmodiophora brassicae populations in single canola root galls from Alberta fields are mixtures of multiple strains. Plant Dis. **104**: 116–120.

Fuchs, H. and Sacristán, M.D. (1996). Identification of a Gene in Arabidopsis thaliana Controlling Resistance to Clubroot (Plasmodiophora brassicae) and Characterization of the Resistance Response. Mol. Plant-Microbe Interact. **9**: 091.

Gan, C. et al. (2019). Construction of a high-density genetic linkage map and identification of quantitative trait loci associated with clubroot resistance in radish (Raphanus sativus L.). Mol. Breed. **39**: 1–12.

Gimenez-Ibanez, S., Hann, D.R., Chang, J.H., Segonzac, C., Boller, T., and Rathjen, J.P. (2018). Differential Suppression of Nicotiana benthamiana Innate Immune Responses by Transiently Expressed Pseudomonas syringae Type III Effectors. Front. Plant Sci. **0**: 688.

Gravot, A., Grillet, L., Wagner, G., Jubault, M., Lariagon, C., Baron, C., Deleu, C., Delourme, R., Bouchereau, A., and Manzanares-Dauleux, M.J. (2011). Genetic and physiological analysis of the relationship between partial resistance to clubroot and tolerance to trehalose in Arabidopsis thaliana. New Phytol. **191**: 1083–1094.

Gravot, A., Lemarié, S., Richard, G., Lime, T., Lariagon, C., and Manzanares-Dauleux, M.J. (2016). Flooding affects the development of Plasmodiophora brassicae in Arabidopsis roots during the secondary phase of infection. Plant Pathol. **65**.

Grimm, D.G., Roqueiro, D., Salomé, P.A., Kleeberger, S., Greshake, B., Zhu, W., Liu, C., Lippert, C., Stegle, O., Schölkopf, B., Weigel, D., and Borgwardt, K.M. (2017). easyGWAS: A Cloud-Based Platform for Comparing the Results of Genome-Wide Association Studies. Plant Cell **29**: 5–19.

Grsic-Rausch, S., Kobelt, P., Siemens, J.M., Bischoff, M., and Ludwig-Müller, J. (2000). Expression and localization of nitrilase during symptom development of the clubroot disease in Arabidopsis. Plant Physiol. **122**: 369–378.

H, A., A, A., VP, M., T, C., SF, H., SE, S., Askarian, H., Akhavan, A., Manolii, V.P., Cao, T., Hwang, S.F., and Strelkov, S.E. (2021). Virulence spectrum of single-spore and field isolates of Plasmodiophora brassicae able to overcome resistance in canola (Brassica napus). Plant Dis. **105**: 43–52.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J.-S., and Concordet, J.-P. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. **17**: 148.

Han, G.-Z. (2019). Origin and evolution of the plant immune system. New Phytol. **222**: 70–83.

Hasan, M.J., Strelkov, S.E., Howard, R.J., and Rahman, H. (2012). Screening of Brassica germplasm for resistance to Plasmodiophora brassicae pathotypes prevalent in Canada for broadening diversity in clubroot resistance. Can. J. Plant Sci. **92**: 501–515.

Hatakeyama, K. et al. (2013). Identification and Characterization of Crr1a, a Gene for Resistance to Clubroot Disease (Plasmodiophora brassicae Woronin) in Brassica rapa L. PLoS One **8**: e54745.

Hatakeyama, K., Niwa, T., Kato, T., Ohara, T., Kakizaki, T., and Matsumoto, S. (2017). The tandem repeated organization of NB-LRR genes in the clubroot-resistant CRb locus in Brassica rapa L. Mol. Genet. Genomics **292**.

Hirai, M., Harada, T., Kubo, N., Tsukada, M., Suwabe, K., and Matsumoto, S. (2004). A novel locus for clubroot resistance in Brassica rapa and its linkage markers. Theor. Appl. Genet. **108**: 639–643.

van der Hoorn, R.A.L. and Kamoun, S. (2008). From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors. Plant Cell **20**: 2009–2017.

Hossain, M.M., Pérez-López, E., Todd, C.D., Wei, Y., and Bonham-Smith, P.C. (2021). Endomembrane-Targeting Plasmodiophora brassicae Effectors Modulate PAMP Triggered Immune Responses in Plants. Front. Microbiol. **12**.

Hwang, S.-F., Strelkov, S.E., Fen, J., Gossen, B.D., and Howard, R.J. (2012). Plasmodiophora brassicae: a review of an emerging pathogen of the Canadian canola (Brassica napus) crop. Mol. Plant Pathol. **13**: 105–113.

Ingram, D.S. and Tommerup, I.C. (1972). The life history of Plasmodiophora brassicae Woron. Proc. R. Soc. London. Ser. B. Biol. Sci. **180**: 103–112.

Irani, S., Trost, B., Waldner, M., Nayidu, N., Tu, J., Kusalik, A.J., Todd, C.D., Wei, Y., and Bonham-Smith, P.C. (2018). Transcriptome analysis of response to Plasmodiophora brassicae infection in the Arabidopsis shoot and root. BMC Genomics **19**: 23.

Jahn, L. et al. (2013). The clubroot pathogen (Plasmodiophora brassicae) influences auxin signaling to regulate auxin homeostasis in arabidopsis. Plants **2**: 726–749.

Jiao, W.-B. and Schneeberger, K. (2020). Chromosome-level assemblies of multiple Arabidopsis genomes reveal hotspots of rearrangements with altered evolutionary dynamics. Nat. Commun. **11**: 989.

Jin, C., Liao, R., Zheng, J., Fang, X., Wang, W., Fan, J., Yuan, S., Du, J., and Yang, H. (2021). Mitogen-Activated Protein Kinase MAPKKK7 from Plasmodiophora brassicae Regulates Low-Light-Dependent Nicotiana benthamiana Immunity . Phytopathology®: PHYTO-08-20-032.

Jones, D.., Ingram, D.S., and Dixon, G.R. (1982). Characterization of isolates derived from single resting spores of Plasmodiophora brassicae and studies of their interaction. Plant Pathol. **31**: 239–246.

Jones, J.D.G.G. and Dangl, J.L. (2006). The plant immune system. Nature **444**: 323–329.

Jubault, M., Lariagon, C., Simon, M., Delourme, R., and Manzanares-Dauleux, M.J. (2008). Identification of quantitative trait loci controlling partial clubroot resistance in

new mapping populations of Arabidopsis thaliana. Theor. Appl. Genet. **117**: 191–202.

Jubault, M., Lariagon, C., Taconnat, L., Renou, J.P., Gravot, A., Delourme, R., and Manzanares-Dauleux, M.J. (2013). Partial resistance to clubroot in Arabidopsis is based on changes in the host primary metabolism and targeted cell division and expansion capacity. Funct. Integr. Genomics **13**.

Kageyama, K. and Asano, T. (2009). Life cycle of plasmodiophora brassicae. J. Plant Growth Regul. **28**: 203–211.

Kamei, A., Tsuro, M., Kubo, N., Hayashi, T., Wang, N., Fujimura, T., and Hirai, M. (2010). QTL mapping of clubroot resistance in radish (Raphanus sativus L.). Theor. Appl. Genet. **120**: 1021–1027.

Katsir, L., Schilmiller, A.L., Staswick, P.E., Sheng, Y.H., and Howe, G.A. (2008). COI1 is a critical component of a receptor for jasmonate and the bacterial virulence factor coronatine. Proc. Natl. Acad. Sci. U. S. A. **105**: 7100–7105.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: A fast spliced aligner with low memory requirements. Nat. Methods **12**: 357–360.

Ko, J.H., Jeon, H.W., Kim, W.C., Kim, J.Y., and Han, K.H. (2014). The MYB46/MYB83-mediated transcriptional regulatory programme is a gatekeeper of secondary wall biosynthesis. Ann. Bot. **114**: 1099–1107.

Kobelt, P., Siemens, J., and Sacristán, M.D. (2000). Histological characterisation of the incompatible interaction between Arabidopsis thaliana and the obligate biotrophic pathogen Plasmodiophora brassicae. Mycol. Res. **104**: 220–225.

Koch, E., Cox, R., and Williams, P.H. (1991). Infection of Arabidopsis thaliana by Plasmodiophora brassicae. J. Phytopathol. **132**: 99–104.

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

Korte, A. and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. Plant Methods.

Kourelis, J. and Van Der Hoorn, R.A.L. (2018). Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. Plant

Cell **30**: 285–299.

Kover, P.X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I.M., Purugganan, M.D., Durrant, C., and Mott, R. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genet. **5**: 1000551.

Krasileva, K. V., Dahlbeck, D., and Staskawicz, B.J. (2010). Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. Plant Cell **22**: 2444–2458.

Kuo, K.H.M. (2017). Multiple Testing in the Context of Gene Discovery in Sickle Cell Disease Using Genome-Wide Association Studies. Genomics Insights **10**: 1178.

Lai, Y. and Eulgem, T. (2018). Transcript-level expression control of plant NLR genes. Mol. Plant Pathol. **19**: 1267–1281.

Lee, J. et al. (2015). Genotyping-by-sequencing map permits identification of clubroot resistance QTLs and revision of the reference genome assembly in cabbage (Brassica oleracea L.). DNA Res. **23**: dsv034.

Lemarié, S., Robert-Seilaniantz, A., Lariagon, C., Lemoine, J., Marnet, N., Jubault, M., Manzanares-Dauleux, M.J., and Gravot, A. (2015a). Both the Jasmonic Acid and the Salicylic Acid Pathways Contribute to Resistance to the Biotrophic Clubroot Agent Plasmodiophora brassicae in Arabidopsis. Plant Cell Physiol. **56**: 2158–2168.

Lemarié, S., Robert-Seilaniantz, A., Lariagon, C., Lemoine, J., Marnet, N., Levrel, A., Jubault, M., Manzanares-Dauleux, M.J., and Gravot, A. (2015b). Camalexin contributes to the partial resistance of Arabidopsis thaliana to the biotrophic soilborne protist Plasmodiophora brassicae. Front. Plant Sci. **6**: 539.

Lewis, J.D., Lee, A.H.-Y., Hassan, J.A., Wan, J., Hurley, B., Jhingree, J.R., Wang, P.W., Lo, T., Youn, J.-Y., Guttman, D.S., and Desveaux, D. (2013). The Arabidopsis ZED1 pseudokinase is required for ZAR1-mediated immunity induced by the Pseudomonas syringae type III effector HopZ1a. Proc. Natl. Acad. Sci. **110**: 18722–18727.

Li, L., Luo, Y., Chen, B., Xu, K., Zhang, F., Li, H., Huang, Q., Xiao, X., Zhang, T., Hu, J., Li, F., and Wu, X. (2016). A Genome-Wide Association Study Reveals New Loci for Resistance to Clubroot Disease in Brassica napus. Front. Plant Sci. **7**: 1–12.

Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30**: 923–930.

Liégard, B. et al. (2019a). Quantitative resistance to clubroot infection mediated by transgenerational epigenetic variation in Arabidopsis. New Phytol. **222**: 468–479.

Liégard, B., Gravot, A., Quadrana, L., Aigu, Y., Bénéjam, J., Lariagon, C., Lemoine, J., Colot, V., Manzanares-Dauleux, M., and Jubault, M. (2019b). Natural Epiallelic Variation is Associated with Quantitative Resistance to the Pathogen Plasmodiophora Brassicae. bioRxiv.

Liu, L., Qin, L., Zhou, Z., Hendriks, W.G.H.M., Liu, S., and Wei, Y. (2020). Refining the Life Cycle of Plasmodiophora brassicae. Phytopathology: PHYTO-02-20-002.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**: 550.

Lovelock, D.A., Šola, I., Marschollek, S., Donald, C.E., Rusak, G., van Pée, K.H., Ludwig-Müller, J., and Cahill, D.M. (2016). Analysis of salicylic acid-dependent pathways in Arabidopsis thaliana following infection with Plasmodiophora brassicae and the influence of salicylic acid on disease. Mol. Plant Pathol.

Lu, Y. and Tsuda, K. (2021). Intimate Association of PRR- and NLR-Mediated Signaling in Plant Immunity. Mol. Plant. Microbe. Interact. **34**: 3–14.

Ludwig-Müller, J. (2014). Auxin homeostasis, signaling, and interaction with other growth hormones during the clubroot disease of Brassicaceae. Plant Signal. Behav. **9**.

Ludwig-müller, J., Auer, S., Jülke, S., and Marschollek, S. (2017). Auxins and Cytokinins in Plant Biology. **1569**: 41–60.

Ludwig-Müller, J., Jülke, S., Geiß, K., Richter, F., Mithöfer, A., Šola, I., Rusak, G., Keenan, S., and Bulman, S. (2015). A novel methyltransferase from the intracellular pathogen Plasmodiophora brassicae methylates salicylic acid. Mol. Plant Pathol. **16**.

Ludwig-Müller, J., Prinsen, E., Rolfe, S.A., and Scholes, J.D. (2009). Metabolism and plant hormone action during clubroot disease. J. Plant Growth Regul.

Ludwig-Müller, J., Bendel, U., Thermann, P., Ruppel, M., Epstein, E., and Hilgenberg,

W. (1993). Concentrations of indole-3-acetic acid in plants of tolerant and susceptible varieties of Chinese cabbage infected with Plasmodiophora brassicae Woron. New Phytol. **125**: 763–769.

Mackey, D., Belkhadir, Y., Alonso, J.M., Ecker, J.R., and Dangl, J.L. (2003). Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. Cell **112**: 379–389.

Mackey, D., Holt, B.F., Wiig, A., and Dangl, J.L. (2002). RIN4 interacts with Pseudomonas syringae type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis. Cell **108**: 743–754.

Malinowski, R., Novák, O., Borhan, M.H., Spíchal, L., Strnad, M., and Rolfe, S.A. (2016). The role of cytokinins in clubroot disease. Eur. J. Plant Pathol. **145**: 543–557.

Malinowski, R., Smith, J.A., Fleming, A.J., Scholes, J.D., and Rolfe, S.A. (2012). Gall formation in clubroot-infected Arabidopsis results from an increase in existing meristematic activities of the host but is not essential for the completion of the pathogen life cycle. Plant J. **71**: 226–238.

Malinowski, R., Truman, W., and Blicharz, S. (2019). Genius Architect or Clever Thief—How Plasmodiophora brassicae Reprograms Host Development to Establish a Pathogen-Oriented Physiological Sink. Mol. Plant-Microbe Interact. **32**: 1259–1266.

Manzanares-Dauleux, M.J., Delourme, R., Baron, F., and Thomas, G. (2000). Mapping of one major gene and of QTLs involved in resistance to clubroot in Brassica napus. Theor. Appl. Genet. **101**: 885–891.

Martel, A., Ruiz-Bedoya, T., Breit-McNally, C., Laflamme, B., Desveaux, D., and Guttman, D.S. (2021). The ETS-ETI cycle: evolutionary processes and metapopulation dynamics driving the diversification of pathogen effectors and host immune factors. Curr. Opin. Plant Biol. **62**: 102011.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17**: 10.

Martin, R., Qi, T., Zhang, H., Liu, F., King, M., Toth, C., Nogales, E., and Staskawicz, B.J. (2020). Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. Science (80-.). **370**.

Matsumoto, E., Yasui, C., Ohi, M., and Tsukada, M. (1998). Linkage analysis of RFLP markers for clubroot resistance and pigmentation in Chinese cabbage (Brassica rapa ssp. pekinensis). Euphytica **104**: 79–86.

Mattey, M. and Dixon, G.R. (2015). Premature germination of resting spores as a means of protecting brassica crops from Plasmodiphora brassicae Wor., (clubroot). Crop Prot. **77**: 27–30.

Mehraj, H., Akter, A., Miyaji, N., Miyazaki, J., Shea, D.J., Fujimoto, R., and Doullah, M.A.U. (2020). Genetics of clubroot and fusarium wilt disease resistance in brassica vegetables: The application of marker assisted breeding for disease resistance. Plants **9**: 1–15.

Mermigka, G. and Sarris, P.F. (2019). The Rise of Plant Resistosomes. Trends Immunol. **40**: 670–673.

Michael, T.P., Jupe, F., Bemm, F., Motley, S.T., Sandoval, J.P., Lanz, C., Loudet, O., Weigel, D., and Ecker, J.R. (2018). High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. Nat. Commun. **9**.

Mine, A., Nobori, T., Salazar-Rondon, M.C., Winkelmüller, T.M., Anver, S., Becker, D., and Tsuda, K. (2017). An incoherent feed-forward loop mediates robustness and tunability in a plant immune network. EMBO Rep. **18**: 464–476.

Nagaoka, T., Doullah, M.A.U., Matsumoto, S., Kawasaki, S., Ishikawa, T., Hori, H., and Okazaki, K. (2010). Identification of QTLs that control clubroot resistance in Brassica oleracea and comparative analysis of clubroot resistance genes between B. rapa and B. oleracea. Theor. Appl. Genet. **120**: 1335–1346.

Nakano, Y., Kusunoki, K., Hoekenga, O.A., Tanaka, K., Iuchi, S., Sakata, Y., Kobayashi, M., Yamamoto, Y.Y., Koyama, H., and Kobayashi, Y. (2020). Genome-Wide Association Study and Genomic Prediction Elucidate the Distinct Genetic Architecture of Aluminum and Proton Tolerance in Arabidopsis thaliana. Front. Plant Sci. **0**: 405.

Neuhauser, S., Kirchmair, M., Bulman, S., and Bass, D. (2014). Cross-kingdom host shifts of phytomyxid parasites. BMC Evol. Biol. **14**: 33.

Ngou, B.P.M., Ahn, H.-K.K., Ding, P., and Jones, J.D.G.G. (2021). Mutual potentiation of

plant immunity by cell-surface and intracellular receptors. Nature **592**: 110–115.

Nordborg, M. et al. (2005). The pattern of polymorphism in Arabidopsis thaliana. PLoS Biol. **3**: e196.

Norkunas, K., Harding, R., Dale, J., and Dugdale, B. (2018). Improving agroinfiltration-based transient gene expression in Nicotiana benthamiana. Plant Methods **14**: 71.

Olszak, M., Truman, W., Stefanowicz, K., Sliwinska, E., Ito, M., Walerowski, P., Rolfe, S., and Malinowski, R. (2019). Transcriptional profiling identifies critical steps of cell cycle reprogramming necessary for Plasmodiophora brassicae -driven gall formation in Arabidopsis. Plant J.

Oxley, S. (2007). Clubroot disease of oilseed rape and other brassica crops (Edingurgh).

Pang, W., Liang, Y., Zhan, Z., Li, X., and Piao, Z. (2020). Development of a Sinitic Clubroot Differential Set for the Pathotype Classification of Plasmodiophora brassicae. Front. Plant Sci. **11**.

Päsold, S., Siegel, I., Seidel, C., and Ludwig-Müller, J. (2010). Flavonoid accumulation in Arabidopsis thaliana root galls caused by the obligate biotrophic pathogen Plasmodiophora brassicae. Mol. Plant Pathol.

Peng, G., Falk, K.C., Gugel, R.K., Franke, C., Yu, F., James, B., Strelkov, S.E., Hwang, S.F., and McGregor, L. (2014a). Sources of resistance to plasmodiophora brassicae (clubroot) pathotypes virulent on canola. Can. J. Plant Pathol. **36**: 89–99.

Peng, G., Lahlali, R., Hwang, S.-F., Pageau, D., Hynes, R.K., McDonald, M.R., Gossen, B.D., and Strelkov, S.E. (2014b). Crop rotation, cultivar resistance, and fungicides/biofungicides for managing clubroot (Plasmodiophora brassicae) on canola. Can. J. Plant Pathol. **36**.

Peng, L., Zhou, L., Li, Q., Wei, D., Ren, X., Song, H., Mei, J., Si, J., and Qian, W. (2018). Identification of quantitative trait loci for clubroot resistance in brassica oleracea with the use of brassica SNP microarray. Front. Plant Sci. **9**: 822.

Pérez-López, E., Hossain, M.M., Wei, Y., Todd, C.D., and Bonham-Smith, P.C. (2021). A clubroot pathogen effector targets cruciferous cysteine proteases to suppress plant immunity. Virulence **12**: 2327–2340.

Piao, Z.Y., Deng, Y.Q., Choi, S.R., Park, Y.J., and Lim, Y.P. (2004). SCAR and CAPS mapping of CRb, a gene conferring resistance to Plasmodiophora brassicae in Chinese cabbage (Brassica rapa ssp. pekinensis). Theor. Appl. Genet. **108**: 1458–1465.

Prakash, S., Bhat, S.R., Quiros, C.F., Kirti, P.B., and Chopra, V.L. (2009). *Brassica* and Its Close Allies: Cytogenetics and Evolution. In Plant Breeding Reviews (John Wiley & Sons, Inc.: Hoboken, NJ, USA), pp. 21–187.

Rahman, H., Peng, G., Yu, F., Falk, K.C., Kulkarni, M., and Selvaraj, G. (2014). Genetics and breeding for clubroot resistance in Canadian spring canola (Brassica napus L.). Can. J. Plant Pathol. **36**: 122–134.

Ramzi, N., Kaczmarek, J., and Jedryczka, M. (2018). Identification of clubroot resistance sources from world gene bank accessions. In IOBC-WPRS Bulletin.

Ranf, S. (2017). Sensing of molecular patterns through cell surface immune receptors. Curr. Opin. Plant Biol. **38**: 68–77.

Rocherieux, J., Glory, P., Giboulot, A., Boury, S., Barbeyron, G., Thomas, G., and Manzanares-Dauleux, M.J. (2004). Isolate-specific and broad-spectrum QTLs are involved in the control of clubroot in Brassica oleracea. Theor. Appl. Genet. **108**: 1555–1563.

Rungrat, T., Almonte, A.A., Cheng, R., Gollan, P.J., Stuart, T., Aro, E.-M., Borevitz, J.O., Pogson, B., and Wilson, P.B. (2019). A Genome-Wide Association Study of Non-Photochemical Quenching in response to local seasonal climates in Arabidopsis thaliana. Plant Direct **3**: e00138.

Sakamoto, K., Saito, A., Hayashida, N., Taguchi, G., and Matsumoto, E. (2008). Mapping of isolate-specific QTLs for clubroot resistance in Chinese cabbage (Brassica rapa L. ssp. pekinensis). Theor. Appl. Genet. **117**: 759–767.

Sarris, P.F. et al. (2015). A Plant Immune Receptor Detects Pathogen Effectors that Target WRKY Transcription Factors. Cell **161**: 1089–1100.

Schuller, A., Kehr, J., and Ludwig-Müller, J. (2014). Laser microdissection coupled to transcriptional profiling of arabidopsis roots inoculated by plasmodiophora brassicae indicates a role for brassinosteroids in clubroot formation. Plant Cell Physiol. **55**.

Schuller, A. and Ludwig-Müller, J. (2016). Histological methods to detect the clubroot pathogen Plasmodiophora brassicae during its complex life cycle. Plant Pathol. **65**.

Schwelm, A. et al. (2015). The Plasmodiophora brassicae genome reveals insights in its life cycle and ancestry of chitin synthases. Sci. Rep. **5**: 11153.

Schwelm, A. and Ludwig-Müller, J. (2021). Molecular pathotyping of plasmodiophora brassicae— genomes, marker genes, and obstacles. Pathogens **10**: 1–22.

Seren, Ü. (2018). GWA-Portal: Genome-Wide Association Studies Made Easy. In (Humana Press, New York, NY), pp. 303–319.

Sharma, K., Gossen, B.D., Greenshields, D., Selvaraj, G., Strelkov, S.E., and McDonald, M.R. (2013). Reaction of Lines of the Rapid Cycling Brassica Collection and Arabidopsis thaliana to Four Pathotypes of Plasmodiophora brassicae. Plant Dis. **97**: 720–727.

Siemens, J., Keller, I., Sarx, J., Kunz, S., Schuller, A., Nagel, W., Schmülling, T., Parniske, M., and Ludwig-Müller, J. (2006). Transcriptome analysis of Arabidopsis clubroots indicate a key role for cytokinins in disease development. Mol. Plant-Microbe Interact. **19**: 480–494.

Siemens, J., Nagel, M., Ludwig-Müller, J., and Sacristán, M.D. (2002). The interaction of Plasmodiophora brassicae and Arabidopsis thaliana: Parameters for disease quantification and screening of mutant lines. J. Phytopathol. **150**: 592–605.

Some, A., Manzanares, M.J., Laurens, F., Baron, F., Thomas, G., and Rouxel, F. (1996). Variation for virulence on Brassica napus L. amongst Plasmodiophora brassicae collections from France and derived single-spore isolates. Plant Pathol. **45**: 432–439.

Stefanowicz, K., Szymanska-Chargot, M., Truman, W., Walerowski, P., Olszak, M., Augustyniak, A., Kosmala, A., Zdunek, A., and Malinowski, R. (2021). Plasmodiophora brassicae-Triggered Cell Enlargement and Loss of Cellular Integrity in Root Systems Are Mediated by Pectin Demethylation. Front. Plant Sci. **0**: 1569.

Stokes, T.L., Kunkel, B.N., and Richards, E.J. (2002). Epigenetic variation in Arabidopsis disease resistance. Genes Dev. **16**: 171–182.

Strelkov, S.E. and Hwang, S.F. (2014). Clubroot in the Canadian canola crop: 10 years

into the outbreak. Can. J. Plant Pathol. **36**: 27–36.

Strelkov, S.E., Hwang, S.F., Manolii, V.P., Cao, T., Fredua-Agyeman, R., Harding, M.W., Peng, G., Gossen, B.D., Mcdonald, M.R., and Feindel, D. (2018). Virulence and pathotype classification of Plasmodiophora brassicae populations collected from clubroot resistant canola (Brassica napus) in Canada. Can. J. Plant Pathol. **40**: 284–298.

Suwabe, K., Tsukazaki, H., Iketani, H., Hatakeyama, K., Fujimura, M., Nunome, T., Fukuoka, H., Matsumoto, S., and Hirai, M. (2003). Identification of two loci for resistance to clubroot (Plasmodiophora brassicae Woronin) in Brassica rapa L. Theor. Appl. Genet. **107**: 997–1002.

Suwabe, K., Tsukazaki, H., Iketani, H., Hatakeyama, K., Kondo, M., Fujimura, M., Nunome, T., Fukuoka, H., Hirai, M., and Matsumoto, S. (2006). Simple Sequence Repeat-Based Comparative Genomics Between Brassica rapa and Arabidopsis thaliana: The Genetic Origin of Clubroot Resistance. Genetics **173**: 309–319.

Thaler, J.S., Humphrey, P.T., and Whiteman, N.K. (2012). Evolution of jasmonate and salicylate signal crosstalk. Trends Plant Sci. **17**: 260–270.

Tsuchimatsu, T. et al. (2020). Adaptive reduction of male gamete number in the selfing plant Arabidopsis thaliana. Nat. Commun. **11**: 2885.

Tu, J., Bush, J., Bonham-Smith, P., and Wei, Y. (2018). Live cell imaging of *Plasmodiophora brassicae* -host plant interactions based on a two-step axenic culture system. Microbiologyopen: e765.

Ueno, H., Matsumoto, E., Aruga, D., Kitagawa, S., Matsumura, H., and Hayashida, N. (2012). Molecular characterization of the CRa gene conferring clubroot resistance in Brassica rapa. Plant Mol. Biol. **80**: 621–629.

Voorrips, R.E. (1995). Plasmodiophora brassicae: aspects of pathogenesis and resistance in Brassica oleracea. Euphytica **83**: 139–146.

Voorrips, R.E., Jongerius, M.C., and Kanne, H.J. (1997). Mapping of two genes for resistance to clubroot (Plasmodiophora brassicae) in a population of doubled haploid lines of Brassica oleracea by means of RFLP and AFLP markers. Theor. Appl. Genet.

Walerowski, P., Gündel, A., Yahaya, N., Truman, W., Sobczak, M., Olszak, M., Rolfe,

S., Borisjuk, L., and Malinowski, R. (2018). Clubroot Disease Stimulates Early Steps of Phloem Differentiation and Recruits SWEET Sucrose Transporters within Developing Galls. Plant Cell **30**: 3058–3073.

Wang, J., Hu, M., Wang, J., Qi, J., Han, Z., Wang, G., Qi, Y., Wang, H.W., Zhou, J.M., and Chai, J. (2019a). Reconstitution and structure of a plant NLR resistosome conferring immunity. Science (80-.). **364**.

Wang, J., Wang, J., Hu, M., Wu, S., Qi, J., Wang, G., Han, Z., Qi, Y., Gao, N., Wang, H.W., Zhou, J.M., and Chai, J. (2019b). Ligand-triggered allosteric ADP release primes a plant NLR complex. Science (80-.). **364**.

Wang, W. and Jiao, F. (2019). Effectors of Phytophthora pathogens are powerful weapons for manipulating host immunity. Planta **250**: 413–425.

Ward, E. and Adams, M. (2010). http://tol.tolweb.org/Plasmodiophorida/121506. Plasmodiophorida. Version 07 March 2010 (under Constr. http//tolweb.org/Plasmodiophorida/121506/2010.03.07 Tree Life Web Proj. http//tolweb.org/.

Werner, S., Diederichsen, E., Frauen, M., Schondelmaier, J., and Jung, C. (2008). Genetic mapping of clubroot resistance genes in oilseed rape. Theor. Appl. Genet.

Wersch, S. van, Tian, L., Hoy, R., and Li, X. (2020). Plant NLRs: The Whistleblowers of Plant Immunity. Plant Commun. **1**.

Williams, P.H. (1966). A system for the determination of races of Plasmodiophora brassicae that infect Cabbage and Rutabaga. Phytopathology **56**: 624–626.

Wu, A.M., Hörnblad, E., Voxeur, A., Gerber, L., Rihouey, C., Lerouge, P., and Marchant, A. (2010). Analysis of the arabidopsis IRX9/IRX9-L and IRX14/IRX14-L pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. Plant Physiol. **153**: 542–554.

Xue, S., Cao, T., Howard, R.J., Hwang, S.F., and Strelkov, S.E. (2008). Isolation and variation in virulence of single-spore isolates of Plasmodiophora brassicae from Canada. Plant Dis. **92**: 456–462.

Yang, J.H., Lee, K., Du, Q., Yang, S., Yuan, B., Qi, L., and Wang, H. (2020). A membrane-associated NAC domain transcription factor XVP interacts with TDIF co-

receptor and regulates vascular meristem activity. New Phytol. **226**: 59–74.

Yu, F., Zhang, X., Peng, G., Falk, K.C., Strelkov, S.E., and Gossen, B.D. (2017). Genotyping-by-sequencing reveals three QTL for clubroot resistance to six pathotypes of Plasmodiophora brassicae in Brassica rapa. Sci. Rep. **7**: 4516.

Yu, L., Wang, Y., Liu, Y., Li, N., Yan, J., and Luo, L. (2018). Wound-induced polypeptides improve resistance against Pseudomonas syringae pv. tomato DC3000 in Arabidopsis. Biochem. Biophys. Res. Commun. **504**: 149–156.

Yuan, M., Ngou, B.P.M., Ding, P., and Xin, X.F. (2021). PTI-ETI crosstalk: an integrative view of plant immunity. Curr. Opin. Plant Biol. **62**: 102030.

Zamani-Noor, N. (2017). Variation in pathotypes and virulence of *Plasmodiophora brassicae* populations in Germany. Plant Pathol. **66**: 316–324.

Zhan, Z., Nwafor, C.C., Hou, Z., Gong, J., Zhu, B., Jiang, Y., Zhou, Y., Wu, J., Piao, Z., Tong, Y., Liu, C., and Zhang, C. (2017). Cytological and morphological analysis of hybrids between Brassicoraphanus, and Brassica napus for introgression of clubroot resistant trait into Brassica napus L. PLoS One.

Zhong, R. and Ye, Z.H. (2015). The arabidopsis NAC transcription factor NST2 functions together with SND1 and NST1 to regulate secondary wall biosynthesis in fibers of inflorescence stems. Plant Signal. Behav. **10**.

Zhou, J., Zhong, R., and Ye, Z.H. (2014). Arabidopsis NAC domain proteins, VND1 to VND5, are transcriptional regulators of secondary wall biosynthesis in vessels. PLoS One **9**: e105726.