

Universidade do Minho
Escola de Ciências

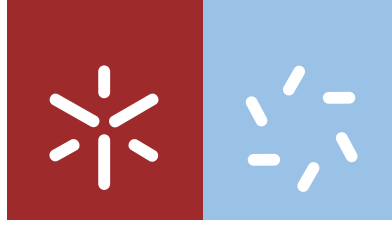
Ana Sofia Guerra Fontes Gonçalves

Molecular analysis of the response of *Pinus pinaster* plants with differential behavior towards the Pine Wood Nematode infection

Ana Sofia Guerra Fontes Gonçalves **Molecular analysis of the response of *Pinus pinaster* plants with differential behavior towards the Pine Wood Nematode infection**

UMinho | 2016

outubro de 2016



Universidade do Minho
Escola de Ciências

Ana Sofia Guerra Fontes Gonçalves

Molecular analysis of the response of *Pinus pinaster* plants with differential behavior towards the Pine Wood Nematode infection

Dissertação de Mestrado
Mestrado em Biologia Molecular, Biotecnologia
e Bioempreendedorismo em Plantas

Trabalho realizado sob orientação da
Doutora Célia Miguel
e da
Professora Doutora Teresa Lino-Neto

Acknowledgements

Finalmente dou por concluída esta etapa! Foi um período extremamente exigente a nível científico, mas sobretudo pessoal, e não teria chegado a este ponto sem o apoio incondicional de algumas pessoas às quais quero prestar o meu maior agradecimento nestas breves palavras:

Primeiramente à Doutora Célia Miguel que tão bem me recebeu no seu laboratório nestes últimos meses. Quero agradecer toda a sua disponibilidade, simpatia e apoio em todas as fases deste trabalho. Admiro a sua calma e atenção para com a pequena “família” que tem construído no seu laboratório, que é o reflexo da confiança que transmite nos outros.

À Professora Teresa Lino-Neto pela simpatia e atenção que sempre demonstrou para comigo e para com os meus colegas durante o ano curricular do mestrado. Agora na condição de co-orientadora da minha tese de mestrado tenho a agradecer a sua disponibilidade e preocupação no decorrer do meu trabalho.

À Doutora Isabel Carrasquinho por toda a ajuda no processo de inoculação, por proporcionar a realização deste trabalho ao fornecer o material biológico utilizado, pela sua simpatia e disponibilidade.

A todos os membros da “Forest Biotech Family” – Inês C., Andreia R., Andreia M., Inês M., Susana, Bruno, Sónia, Sofia L., Cirenia e Pedro – por me acolherem e apoiarem cientificamente durante estes meses, mas também por todos os momentos de descontração que me proporcionaram. Um agradecimento especial à Inês Modesto com a qual trabalhei lado a lado durante parte deste trabalho; obrigada por todo o apoio! E ao Bruno Costa pela enorme ajuda na componente bioinformática deste trabalho.

Às minhas amigas bioquímicas. Não interessa a distância ou o tempo que passa, “o que a Bioquímica juntou, ninguém irá separar”!

Ao Sérgio por todo o seu apoio e por estar sempre ao meu lado, incondicionalmente...

À minha família, em especial aos meus pais. Não há palavras para descrever a gratidão que vos tenho por sempre me darem liberdade e apoio nas minhas decisões, por fazerem tudo para o meu bem-estar e do João, enfim... devo-vos em muito a pessoa que sou hoje e tenho ainda muito a aprender com o vosso bom exemplo. Esta tese é dedicada a vocês!

Abstract

The Pine Wood Nematode (PWN) *Bursaphelenchus xylophilus* is the causal agent of Pine Wilt Disease (PWD), which threatens several conifer species around the world leading to great ecological and economical losses. *Pinus pinaster* (maritime pine), one of the major forest species in Portugal, is susceptible to the PWN infection. The importance of controlling the evolution of the disease as well as its spreading to other countries motivated the initiation of a national breeding program to address this problem. Over five hundred adult trees have been selected as candidate PWN resistant trees for this program from a PWD highly affected area. In this study, seedlings derived from a candidate resistant tree were used in an inoculation experiment to assess susceptibility/tolerance to PWN infection. From the set of inoculated plants, some apparently tolerant plants were selected along with susceptible ones to conduct expression analyses of selected transcripts potentially involved in the response mechanisms against PWN infection, and that might explain the susceptibility or tolerance towards PWD.

Samples from the selected susceptible and tolerant *P. pinaster* plants were sent for small RNA and degradome next-generation sequencing. Differential expression analysis of the identified small RNAs between susceptible and tolerant plants was performed, and target transcripts for the differentially expressed small RNAs were predicted using computational approaches. Transcripts were also selected from a previous study that identified expressed sequence tags differentially expressed in *P. pinaster* susceptible plants *versus* plants from a tolerant species, *Pinus pinea*, inoculated with PWN. Expression analysis of small RNA targets and transcripts putatively involved in the response mechanism to PWD allowed the identification of genes with different expression profiles between susceptible and tolerant *P. pinaster* plants, that are apparently linked to the differential behavior of those plants to PWN infection. This study provides valuable information for future research to elucidate the response mechanisms to PWD, and to identify potential marker genes that, linked to a marker-assisted selection of trees, can be an effective approach to accumulate tolerance-related genes in a population or an individual, and contribute to the future establishment of pine forests tolerant to the PWD.

Resumo

O Nemátode da Madeira do Pinheiro (NMP) *Bursaphelenchus xylophilus* é o agente causal da Doença da Murchidão do Pinheiro (DMP), que ameaça várias espécies de coníferas em todo o mundo levando a grandes perdas ecológicas e económicas. *Pinus pinaster* (pinheiro bravo), uma das principais espécies florestais em Portugal, é suscetível à infeção pelo NMP. A importância de controlar a evolução da doença, bem como a sua propagação para outros países motivaram o início de um programa de melhoramento nacional para fazer face a este problema. Mais de quinhentas árvores adultas foram selecionadas como árvores candidatas resistentes ao NMP, no âmbito deste programa, de uma área altamente afetada pela DMP. Neste trabalho, plantas jovens provenientes de uma árvore candidata resistente foram usadas num ensaio de inoculação para avaliar a suscetibilidade/tolerância à infeção com o NMP. Das plantas inoculadas, foram selecionadas algumas plantas aparentemente tolerantes à infeção, juntamente com algumas suscetíveis, para realizar análises de expressão de transcritos potencialmente envolvidos nos mecanismos de resposta à infeção com o NMP, e que poderão explicar a suscetibilidade ou tolerância à DMP.

Amostras das plantas de *P. pinaster* suscetíveis e tolerantes selecionadas foram enviadas para sequenciação de nova geração de pequenos RNAs e do degradoma. A análise da expressão diferencial dos pequenos RNAs identificados entre plantas suscetíveis e tolerantes foi efetuada, e transcritos-alvo dos pequenos RNAs diferencialmente expressos foram previstos usando abordagens computacionais. Também foram selecionados transcritos de um estudo anterior que identificou “expressed sequence tags” diferencialmente expressas em plantas suscetíveis de *P. pinaster* versus plantas de uma espécie tolerante, *Pinus pinea*, inoculadas com o NMP. A análise de expressão dos alvos de pequenos RNAs e transcritos putativamente envolvidos no mecanismo de resposta à DMP permitiu a identificação de genes com diferentes perfis de expressão entre plantas de *P. pinaster* suscetíveis e tolerantes, que estão aparentemente relacionados com o comportamento diferencial destas plantas face à infeção com o NMP. Este estudo fornece informação importante para futura investigação relativa aos mecanismos de resposta à DMP e para a identificação de potenciais genes marcadores que, ligada a uma seleção de árvores assistida por marcadores, poderá ser uma estratégia efetiva para acumular genes relacionados com a tolerância numa população ou indivíduo, e contribuir para o estabelecimento futuro de florestas de pinheiros tolerantes à DMP.

Table of Contents

Acknowledgements	iii
Abstract.....	v
Resumo	vii
Table of Contents.....	ix
Abbreviations and acronyms	xi
1. Introduction.....	1
1.1. Pine Wilt Disease	1
1.1.1. Historical Overview.....	1
1.1.2. The Pine Wood Nematode.....	2
1.1.3. The Vector Beetle.....	3
1.1.4. Transmission Biology of the Pine Wood Nematode.....	5
1.1.5. Pine Wilt Disease Development	7
1.2. Host Tree Response Mechanisms	9
1.3. miRNAs	12
1.4. Objectives.....	16
2. Materials and Methods	18
2.1. Biological material	18
2.1.1. <i>Pinus pinaster</i>	18
2.1.2. <i>Bursaphelenchus xylophilus</i>	18
2.2. Inoculation of <i>Pinus pinaster</i> plants with <i>Bursaphelenchus xylophilus</i>	18
2.3. Total RNA extraction	19
2.3.1. RNA samples purification and evaluation of RNA quality.....	20
2.3.2. Total RNA and miRNA quantification using Qubit® fluorometer	20
2.4. Small RNA and degradome sequencing.....	21
2.5. Bioinformatic analyses of sequencing results.....	22
2.5.1. Small RNA sequence processing	22
2.5.2. Differential expression analysis of sequenced small RNAs	24
2.5.3. Small RNA target gene prediction and validation	25
2.6. Selection of small RNA target genes for expression analysis.....	25
2.7. Selection of candidate PWD-related transcripts for expression analysis.....	26
2.8. Two-step RT-qPCR	26
2.8.1. cDNA synthesis.....	26
2.8.1.1. Diagnostic PCR.....	27

2.8.2.	Selection of reference genes for qPCR.....	28
2.8.3.	qPCR.....	28
2.8.3.1.	Experimental determination of qPCR conditions.....	29
2.8.3.2.	Primer's efficiency calculations	29
2.8.3.3.	Expression stability of candidate reference genes	30
2.8.3.4.	Expression of predicted small RNA target transcripts	31
2.8.3.5.	Expression of PWD-related candidate genes.....	32
3.	Results and discussion	33
3.1.	Total RNA isolation from <i>Pinus pinaster</i> plants with contrasting responses towards PWN infection	33
3.2.	Small RNA and degradome sequencing.....	34
3.3.	Selection of small RNAs targets putatively involved in the response molecular mechanisms of <i>Pinus pinaster</i> plants affected with PWD.....	35
3.3.1.	Sequenced small RNAs and <i>in silico</i> evaluation of their expression profiles.....	35
3.3.2.	Target prediction for small RNAs differentially expressed between susceptible and tolerant plants.....	39
3.4.	Gene expression analyses of transcripts putatively involved in the differential behavior of <i>Pinus pinaster</i> plants towards the PWN infection.....	42
3.4.1.	qPCR experimental setup: determination of amplification efficiencies, and validation of reference genes for expression normalization	43
3.4.1.1.	qPCR amplification efficiencies of candidate reference genes.....	43
3.4.1.2.	Validation of reference genes for expression normalization.....	44
3.4.2.	Relative expression level quantification of candidate genes and selected small RNA targets.....	49
4.	Concluding remarks and future perspectives	61
5.	References	63
6.	Appendices.....	76
	Appendix A –Inoculation of <i>Pinus pinaster</i> plants with <i>Bursaphelenchus xylophilus</i> , and symptom observation.....	76
	Appendix B –Agarose gel to check RNA integrit.....	77
	Appendix C –miRPursuit configuration directory files.....	78
	Appendix D –Quality plots of small RNA sequencing libraries.....	80
	Appendix E –Size profile distribution of reads from small RNA sequencing libraries.....	81
	Appendix F –Small RNAs differentially expressed in susceptible and tolerant plants and their respective target transcript(s).	81
	Appendix G –Oligonucleotides used in gene expression analysis using qPCR.....	84
	Appendix H –Example of a melting curve obtained in real-time quantitative PCR experiments and agarose gel electrophoresis of PCR products.	87

Abbreviations and acronyms

μL – microliter

A(XXX) – absorbance at XXX nanometers of wavelength

BLAST – Basic Local Alignment Search Tool

bp – base pair

C – cytosine

cDNA – complementary DNA

cm – centimeters

C_p – crossing point

CTAB – hexadecyltrimethylammonium bromide

DEPC – diethylpyrocarbonate

DNA – deoxyribonucleic acid

DNAse – deoxyribonuclease

dNTP – deoxynucleotide

DTT – dithiothreitol

EDTA – ethylenediaminetetraacetic acid

EST – expressed sequence tag

g – centrifuge force

g – gram

G – guanidine

h – hour

M – molar

mg – milligrams

MgCl₂ – magnesium chloride

min – minute

miRNA – microRNA

mL – milliliter

mM – milimolar

mRNA – messenger RNA

NaCl – sodium chloride

NaOAc – sodium acetate

ng – nanogram

nt – nucleotide

°C – degrees Celsius

PCR – polymerase chain reaction

pg – picogram

PVP10 – polyvinyl pyrrolidone

PWD – Pine Wilt Disease

PWN – Pine Wood Nematode

qPCR – real-time quantitative PCR

RNA – ribonucleic acid

RNAse – ribonuclease

RNA-seq – RNA sequencing

rpm – revolutions per minute

rRNA – ribosomal RNA

RT – reverse transcription

s – second

SDS – sodium dodecyl sulphate

SEM – standard error of the mean

siRNA – small interfering RNA

ta-siRNA – transacting siRNA

Tris-HCl – tris hydrochloride

tRNA – transfer RNA

U – enzyme unit

U – uracil

USA – United States of America

UV – ultraviolet

V – volt

w/v – weight per volume

WB – workbench

1. Introduction

1.1. Pine Wilt Disease

1.1.1. Historical Overview

At the very beginning of the twentieth century, in the year of 1905, Japanese foresters began to notice a widespread mortality of pine trees at the port city Nagasaki. For several decades the mortality spread northward in the island and then to the mainland (from the 1900s up until the 1960s), and the cause of mortality was then thought to be the wood boring beetles that were prevalent in the dead trees (Zhao *et al.*, 2008). It was only in 1971, after a series of inoculation experiments, that the pine wood nematode (PWN) *Bursaphelenchus xylophilus* was clearly identified as the causal agent of the pine wilt disease (Kiyohara and Tokushige, 1971) and *Monochamus alternatus* beetles as the vector for the PWN (Mamiya and Enda, 1972). Since then, intensive research on the biology and ecology of pine wilt took place in Japan. A few years later, PWNs were recovered from the wood of a dead pine in the USA and thereafter were found to be widely distributed throughout the country (Dropkin *et al.*, 1971), yet only some exotic species suffered from the disease. After some investigation, it was found that the nematode was native to North America and was likely introduced into Japan in the early 1900s. Human activity has contributed intentional or accidentally to the dissemination of some species away from their natural geographical distributions, and place a threat to the established ecosystems (Zhao *et al.*, 2008). The introduction of the PWN into Japan and its subsequent spread to China (1982), Korea (1988) and more recently Portugal (Mota *et al.*, 1999), Madeira Island (Fonseca *et al.*, 2012) and Spain (Robertson *et al.*, 2011) is a striking example of ecosystems being threatened by the establishment of an exotic organism (Zhao *et al.*, 2008). The long-range spread of the PWN occurred as a result of human activities; frequently the nematodes and also their vectors are thought to be transported in timber used for the production of packing materials.

In Portugal, the forest sector is one of the greatest economic activities of the country with a 13.3% contribution to the industrial gross added value. It represents 2.1% of the national gross domestic product, about 10% of the Portuguese exports, and 3% of the national total employment. Wood for furniture and construction, wood for pulp, paper, and paperboard, the cork-based chain, chestnuts, umbrella pine nuts, resin and forest biomass for energy are the main forest-based chains of the Portuguese economy. Maritime pine forests (*Pinus pinaster*) of the center-north of

Portugal are one of the three main forest types of the country, alongside with the Mediterranean evergreen oaks (*Quercus suber* and *Quercus rotundifolia*) in the center and south, and *Eucalyptus globulus* plantations in the coastal northern part of the country (Reboredo, 2014). *P. pinaster* contributes to important industrial products, such as wood and resin, as well as coastal protection associated with sand dunes.

P. pinaster is known to be susceptible to the pine wilt disease (PWD), thus the introduction of PWN into Portugal has a tremendous economic and ecological impact. A national program for control of the PWN (PROLUNP) was implemented immediately after the nematode was discovered in Portugal, to primarily assess the extent of its distribution. Within a 30 km radius in the Setubal Peninsula, where the nematode was exclusively detected, all the symptomatic trees were felled (approximately 50,000 trees per year) (Rodrigues, 2008). Then, in 2007, a 3 km wide precautionary phytosanitary strip around the affected area devoid of *P. pinaster* was established, for the control and eventually the eradication of the nematode (Mota and Vieira, 2008). The clear-cut trees from the affected area were subjected to treatment that may include methyl bromide fumigation, high-temperature treatment of wood before use, or chipping and burial (Jones *et al.*, 2008). Research on the bioecology of the nematode and its insect vector, new detection methods involving e.g. real-time PCR, tree ecology and pathology, and control methods has been ongoing since 1999 (Mota and Vieira, 2008).

1.1.2. The Pine Wood Nematode

Bursaphelenchus xylophilus (Steiner & Buhner) Nickle is a migratory non-obligate endoparasite that infects mainly *Pinus* species, causing the PWD. Although parasitic in nature, the PWN can be easily maintained in the laboratory, usually by culture on the fungus *Botrytis cinerea*, completing its entire life cycle from fertilization to mature adult within 5 days at 25°C (Hasegawa and Miwa, 2008). The PWN reproduces gonochoristically (male and female sexes), with a great number of offspring (about 100-800; Bolla and Boschert, 1993).

The PWN life cycle can progress in two different ways, the reproductive and the dispersal phases, where the nematode shows different feeding habits, phytophagous and mycophagous, which are characteristic of this species (Moens and Perry, 2009; Zhao *et al.*, 2014). In each phase, its behavior, nutrition, reproduction, and distribution in the host tree are significantly influenced by cohabiting microorganisms (Futai and Mota 2008). Resembling insects, nematodes undergo several "molting" processes, progressing through four juvenile stages (J1 to J4) (Fig. 1).

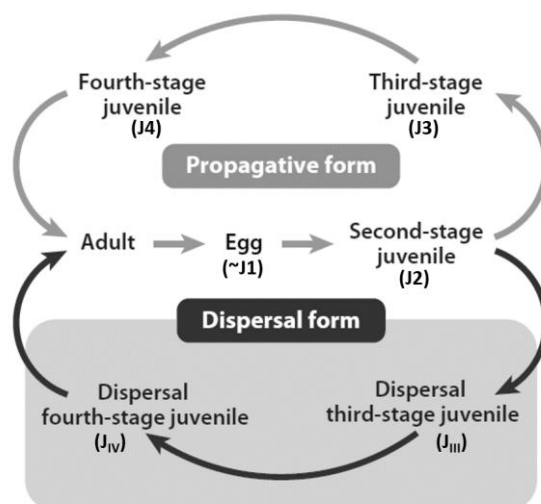


Figure 1 – The life cycle of the Pine Wood Nematode, *Bursaphelenchus xylophilus*. The life cycle includes a propagative phase, in which the propagative form of the nematode arises when the conditions are appropriate for propagation, and a dispersal phase, that is induced by unfavorable conditions such as desiccation, food shortage, or environmental deterioration due to overpopulation. The two stages that constitute the dispersal form of the nematode (highlighted by the gray box) are in close relationship with the vector beetle. The gray arrows and black arrows show the propagation cycle in pine trees and that for transmission to new host trees, respectively. (Adapted from Futai, 2013)

In nature, the PWN invades a healthy host pine tree via feeding wounds made by a vector species, predominantly cerambycid beetles of the genus *Monochamus*, feeds initially on living tree tissues and later on fungi that colonize the dead tree. When food is unavailable, or in the presence of unfavorable conditions (e.g. cold temperatures), the nematode larva enters the dispersal third-stage (J_{III}), a survival stage that can adapt to adverse conditions (Ishibashi and Kondo, 1977; Kondo and Ishibashi, 1978). Then, stimulated by the beetle pupa, J_{III} molts to become the dispersal fourth-stage juvenile (J_{IV}), which is the nematode form that is transported by the vector into a healthy pine tree (Maehara and Futai, 1996), where it can enter the propagative adult stage and starts reproducing (Mamiya, 1975).

1.1.3. The Vector Beetle

PWN is vectored by cerambycid beetles of the genus *Monochamus* (Coleoptera: Cerambycidae), that include *Monochamus alternatus* in East Asia (Mamiya and Enda, 1972; Morimoto and Iwasaki, 1972; Lee *et al.*, 1990; Yang, 2004), *M. saltuarius* in Japan (Sato *et al.*, 1987), *M. carolinensis* in North America (Linit *et al.*, 1983), and *M. galloprovincialis* in Portugal (Sousa *et al.*, 2001). Among these vectors, *M. alternatus* is the most intensively investigated, especially in Japan, because it is the vector known for the longest time and due to its importance in PWD development in the economical and ecologically important pine trees. *M. alternatus* is

indigenous to Japan but its geographical distribution there was scarce up until the PWN introduction in the country. The understanding of the life-history traits of *M. alternatus* should be important to clarify the outbreak and PWN epidemics (Togashi, 2006).

M. alternatus is an ectothermic organism, which means that its development is affected by ambient temperature; when the ambient temperature is favorable for larvae to develop *M. alternatus* has a one-year life cycle, whereas in the presence of cold summer temperatures the larvae do not complete their development within one season and thus have a two-year life cycle (Togashi, 1989c). Reproductively immature adults emerge from dead host trees once a year, in late spring through summer, randomly disperse by flying, and feed on the bark of pine branches or other conifers for survival and sexual maturation (maturation feeding) (Togashi, 2006). Mature adults are then strongly attracted to volatiles emitted from dying or newly killed trees (Ikeda *et al.*, 1980) since they cannot survive in live host trees due to oleoresin, the main defense agent in conifers. In weakened trees, they can, therefore, mate and oviposit successfully. Larvae develop through four instars and then make tunnels in the xylem in late summer through autumn (Togashi, 1989c). In the end of the tunnels, they make a pupal chamber where they overwinter (Fig. 2). Pupation occurs after overwintering, and then sclerotization of newly eclosed adults happens, and they are therefore ready to make their way out to the bark surface and leave the tree (Nakamura-Matori, 2008).



Figure 2 – *Monochamus alternatus* larva in the pupal chamber. The pupal chamber is observed in an infested pine log that has been chopped vertically with a hatchet. The arrow points out the entrance of the tunnel bored into the xylem by the larva. (scale bar = 2 cm; Retrieved from Togashi, 2008)

After the PWN introduction in Japan, and since the nematode is highly pathogenic to Japanese pine trees like *Pinus densiflora* and *Pinus thunbergii*, the number of weakened and dying pine trees prominently improved, increasing the availability of resources for *M. alternatus* propagation. Thus, it can be inferred that a mutual relationship between the introduced PWN and the beetle could easily establish once the two organisms co-occurred (Nakamura-Matori, 2008). In Japan, PWN-killed trees are found in mid-summer through autumn, which coincides with the adult beetle flight season (Togashi, 1989a). So, trees killed from PWD are perfect places for beetle oviposition. Moreover, the trees release volatiles that seem to attract reproductively mature beetle adults and it is observed a positive, spatial association between trees diseased the previous year and those diseased in the early season of the current year (Togashi, 1991). Both of these evidences indicate that there is a well-established mutualistic relationship between the PWN and *M. alternatus*, which make it difficult to control the pathogenicity of PWD.

1.1.4. Transmission Biology of the Pine Wood Nematode

PWN is transmitted by cerambycid beetles from the genus *Monochamus*. *M. alternatus*, *M. carolinensis*, and *M. galloprovincialis* are the most important vector beetles for the PWN in East Asia, North America, and Portugal, respectively. Their conifer hosts belong to the family Pinaceae (Togashi, 2008). The PWD is characterized by a close relationship between the PWN and its vector beetle (Fig. 3). During summer, the PWN's fourth-stage dispersal juveniles (J_{iv}) are carried by the vector beetles from dead to healthy host trees (Fig. 3), where they molt to adults and start to mate. The female adult then initiates oviposition (Fig. 3). PWN reproduction occurs through the four stages of the propagative form, which feed on parenchyma cells of the tree's resin canals and later on fungi, ultimately leading to a dead tree (Fig. 3). Beetle females of the species *M. alternatus* lay eggs under the bark of those newly-killed trees (Fig. 3). Generally, the *M. alternatus* larvae stay in the inner bark in the first, second and third developmental stages (or instars), begin to bore tunnels into the xylem in the fourth instar, and then make pupal chambers in the xylem (Katsumi Togashi, 1989a; Katsumi Togashi, 1989b; Katsumi Togashi, 1991). The beetles overwinter as larvae and pupate the following year between late spring and early summer. Simultaneously to beetle development, reproduction of the nematode leads to increased population numbers in the pine trees. When the nematode population reaches a certain level, the third-stage dispersal juveniles (J_{iii}) appear (Kiyohara and Suzuki, 1975). During winter, J_{iii} aggregate around the pupal chambers made by the beetle larvae (Fig. 3). When beetle larvae eclose in the pupal chambers, between late

spring and early summer, J_{III} around the pupal chamber molt to J_{IV} and then invade the beetle's body (Fig. 3) (Togashi, 2008). The numbers of nematodes that aggregate around the pupal chambers and that invade and are carried by the beetle are mostly determined by the fungal microflora present around the chamber (Maehara and Futai, 2002). The beetles then emerge from the killed host trees (Fig. 3) and move to young branches of nearby healthy trees, where they make feeding wounds (maturation feeding, Fig. 3) from which the J_{IV} , after leaving the beetle's body, can invade the new host tree (Fig. 3). The process of leaving the beetle's body is greatly influenced by the volatiles that the tree releases upon the beetle's maturation feeding, that are perceived by the nematodes inside the beetles (Enda and Ikeda, 1983; Stamps and Linit, 1998).

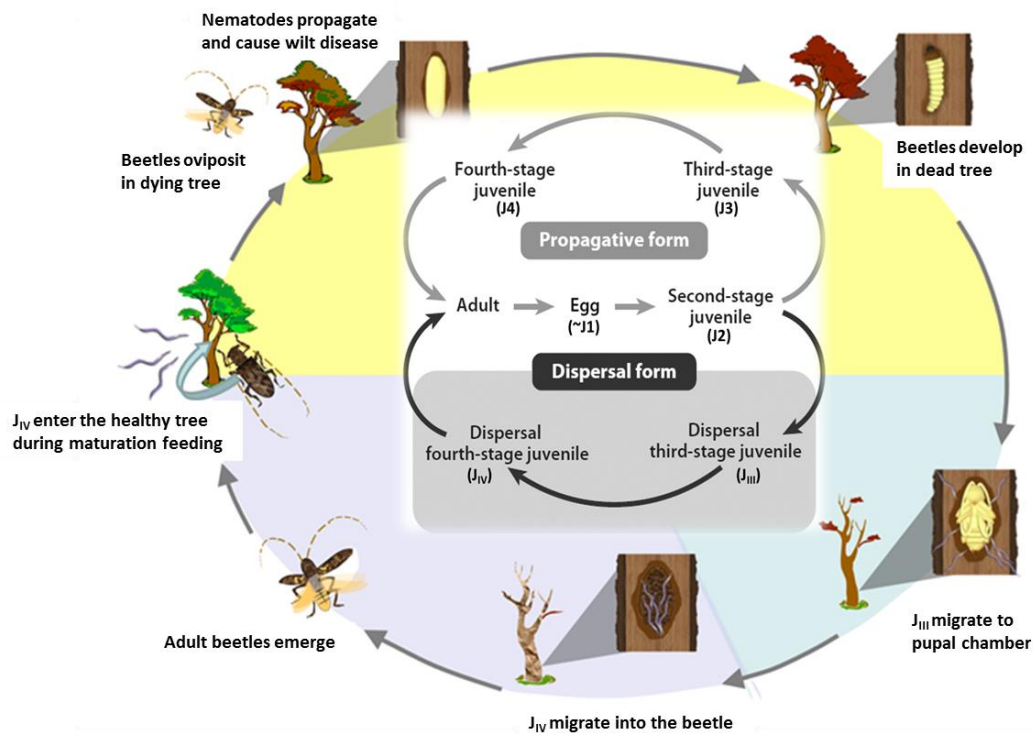


Figure 3 – The relationships between *Bursaphelenchus xylophilus* life cycle and its transmission by the insect vector. In the center the life cycle of *Bursaphelenchus xylophilus* is schematically represented, showing its propagative and dispersal forms. The main events of the nematode life cycle that occur inside the host tree are illustrated in the outer cycle, in relation to the transmission of *B. xylophilus* by its insect vector. (Adapted from Shinya *et al.*, 2013; Futai, 2013)

During unfavorable conditions for the development of PWD or in pine forests resistant to the PWN like the ones in the USA, the PWN populations can still thrive via alternative transmission pathways. It has been shown that female beetles can transmit PWNs to dying trees via oviposition wounds directly (Wingfield and Blanchette, 1983; Edwards and Linit, 1992), and male beetles searching for mates can also transmit the nematodes to dying trees via already existing wounds on

the bark (Arakawa and Togashi, 2002). Interestingly, PWNs sometimes move between beetles during the mating process before being transferred to the trees by either one of the sexes (Togashi and Arakawa, 2003).

1.1.5. Pine Wilt Disease Development

The artificial propagation of PWN and an inoculation procedure were first established by Mamiya (1980, 1984), which led to the first physiological measurements being conducted concerning PWD development (Tamura *et al.*, 1987; 1988). During the 1980s, highly virulent PWN isolates were obtained from many cultures established from dead pines (Kiyohara and Dozono, 1986; Kiyohara and Bolla, 1990), and by the end of the following decade hundreds of reports on the phenomena related to trees' inoculation with the PWN had been published in Japan (Fukuda, 1997; Yamada, 2006). In Table I is outlined the current knowledge about the PWD development, regarding the external symptoms and internal changes observed over time in pine wilt-susceptible saplings of *P. thunbergii* inoculated with the PWN.

Table I – Pine Wilt Disease development in susceptible pines inoculated with *Bursaphelenchus xylophilus*. External symptoms and internal changes observed in *Pinus thunbergii* seedlings inoculated with the pine wood nematode, *B. xylophilus*. (Adapted from Zhao *et al.*, 2008)

		Early phase		Developing phase	
		Stage 1	Stage 2	Stage 3	Stage 4
External	Symptoms	No visible symptoms		Discoloration of old needles	Discoloration of young needles and, later, death
	Oleo-resinosis	Normal	Decreasing	None	None
Internal	Cells	Change in secondary metabolism		Partial necrosis	Necrosis in wide area
	Sap ascent	Normal	Partial blockage of sap ascent	Low water conductivity	Completely stop of sap ascent
Pine wood nematode		Low population		Propagation	Extensive propagation
Time (weeks)		1	2	3	4 and beyond

In the stems of conifers, most xylem tissue consists of elongated cells called tracheids, crossed orthogonally by ray tissue (Fig. 4C). In *Pinus* species, the ray tissue consists of ray tracheids and ray parenchyma cells (Fig. 4E). The resin canals are intercellular spaces formed by separation of parenchyma cells. The xylem of *Pinus* species contain vertical and horizontal resin

canals and their cortex contains vertical resin canals (Fig. 4B) (Evert, 2006). When pine tissue is injured mechanically or by insect feeding, epithelial cells that surround the resin canals (Fig. 4C and D) immediately synthesize and exude resin that traps most of the PWNs hindering their invasion of the resin canals (Kuroda, 2008). Large numbers (3000-10000) of PWNs are then used in inoculation experiments because it is estimated that only about 10% of the inoculum successfully invades the tissue (unpublished data) (Kuroda, 2008). Although the resin effectively prevents the initial invasion of nematodes, it does not prevent the migration of PWNs in the tissue. PWNs indeed migrate very rapidly in the shoot and branches of a pine tree, 150 cm per day at maximum, readily invading the main stem. And since horizontal and vertical resin canals are distributed densely in the xylem tissue (Fig. 4A), PWNs can spread rapidly throughout the entire plant (Kuroda and Ito, 1992).

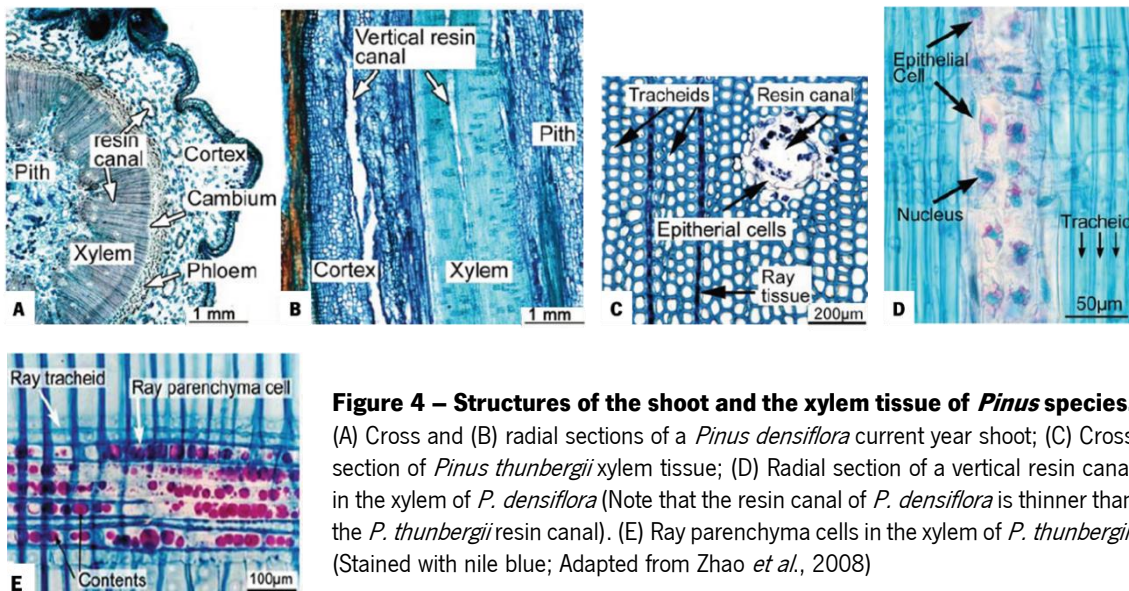


Figure 4 – Structures of the shoot and the xylem tissue of *Pinus* species. (A) Cross and (B) radial sections of a *Pinus densiflora* current year shoot; (C) Cross section of *Pinus thunbergii* xylem tissue; (D) Radial section of a vertical resin canal in the xylem of *P. densiflora* (Note that the resin canal of *P. densiflora* is thinner than the *P. thunbergii* resin canal). (E) Ray parenchyma cells in the xylem of *P. thunbergii*. (Stained with Nile blue; Adapted from Zhao *et al.*, 2008)

In the early phase of PWD development, before the initiation of any visible symptoms (Stages 1 and 2 in Table I), the PWN population in pine stems is still low. At this time, a small number of PWNs are spreading through the host tree exclusively in the resin canals and significant internal changes have already started in the pine tissue. In the epithelium and ray parenchyma cells secondary metabolites, like terpenoids, phenolic compounds, and stilbenoids, are produced by the defense reaction (Hillis, 1987), and the subsequent decline and death of those cells were observed locally, in connection with the distribution of the PWNs. Another internal change that happens in this phase is the disturbance and partial blockage of the sap ascent that occurs in the xylem, as a

result of some tracheids that transport water from the roots to the shoot becoming dysfunctional (Kuroda *et al.*, 1988; 1991). The dehydration of water conduits by gas (embolism) is a process that occurs every day even in healthy trees; however, the empty conduits are usually refilled with water, and water transport is reestablished (Sperry and Tyree, 1988). However, in PWD-susceptible trees the empty conduits do not refill so easily, leading to the formation of dry zones. The formation of such dehydrated areas is known as a defense reaction of plants to the infection by microorganisms. From 2 to 3 weeks after inoculation with PWNs, the dehydrated area in pine tissues increases, and thus there is a significant decrease in water conductivity (Kuroda *et al.*, 1991; Ikeda and Suzuki, 1984) and in photosynthesis in the leaves, that begin to show some discoloration (Stage 3, Table I) (Fukuda *et al.*, 1992; Fukuda, 1997). After the needles start to discolor, an increase in the PWN population is observed. PWNs spread from the resin canals to the cambial zone, and the immature cells around the cambium appear degraded, possibly by the action of hydrolytic enzymes exuded by the nematode (Kusunoki, 1987; Kuroda, 2008). In the second and third stages of the disease, oleoresin exudation from the wound decreases and stops (Table I). The decrease in the exudation of resin has been used to diagnose PWD just before the developing phase and symptom initiation (Kuroda, 2008). At the end of the developing phase (Stage 4, Table I), there is massive necrosis of parenchymal cells and PWN population is extensively propagated, which must be a physiological turning point for the tree, that possibly results in the termination of its defense reaction and ultimately in tree death.

The multitude of physiological and biochemical processes that constitute the early phase of the disease in a period preceding the appearance of symptoms must be the cause of the anatomical incidences and changes in metabolism occurring at the later developing phase of the disease. To detect the actual cause of the symptoms of the disease and to fully understand the molecular aspects of disease development, research must be conducted using samples obtained early after the inoculation of pines with the PWN.

1.2. Host Tree Response Mechanisms

The plant defense mechanisms that prevent or moderate the detrimental effects of a pathogen infection can be divided into two broad classes: resistance and tolerance. Resistance traits are roughly defined as host traits that reduce the extent of pathogen infection by preventing

infection or limiting pathogen growth and development within the host (Kover and Schaal, 2002; Horns and Hood, 2012), whereas tolerance traits do not inhibit the infection but, instead, reduce its negative effects on plant fitness (Roy and Kirchner, 2000; Miller *et al.*, 2005). In this work, “resistance” is used as an absolute term to characterize a plant or species that has been shown to counteract the infection and almost “immunize” itself against the disease, and “tolerance” is used as a rather relative term to describe a plant or species that is able to withstand the infection better than another subject, called “susceptible”. Therefore, when the defense mechanism is not fully understood the term “tolerance” is applied, which does not mean that a plant or species is not effectively resistant to the disease.

The molecular response mechanisms of pine trees against PWN infection have not yet been elucidated. High-throughput screening procedures have been used to identify genes that are differentially expressed between susceptible and resistant pines, when they are available for a given species, and also between plants from different species that are considered either susceptible or tolerant/resistant. These studies have expanded the knowledge about the various molecular events that take place in pine trees following PWN infection. Shin *et al.* (2009) identified several upregulated genes in PWN-inoculated Japanese red pine (*P. densiflora*) related to plant biotic stress resistance, oxidative stress-related genes, water stress-responsive genes, pathogenesis-related (PR) proteins, secondary metabolism and posttranscriptional regulation. Nose and Shiraishi (2011) compared susceptible and resistant Japanese black pines (*P. thunbergii*) and found an upregulation of PR proteins, secondary metabolism genes and disease resistance genes in both susceptible and resistant pines; a downregulation of a growth regulator, a cell wall-loosening enzyme and a translation initiator factor in both susceptible and resistant pines; and an upregulation of secondary metabolism genes and certain PR proteins in susceptible pines. Santos *et al.* (2012) compared a susceptible species (*P. pinaster*) with a tolerant species (*Pinus pinea*) and identified several differentially expressed sequence tags (ESTs) related to PWN infection with roles in oxidative stress response, the production of lignin and ethylene, and the posttranscriptional regulation. Hirao *et al.* (2012) analyzed ESTs from PWN-inoculated *P. thunbergii* at different time points and identified temporal and quantitative differences between susceptible and resistant pines. PR proteins and microbial-related genes were rapidly induced in high levels after inoculation in susceptible plants. In tolerant pines, a moderate initial response mediated by PR proteins followed by a significant upregulation of cell wall-related genes induced by reactive oxygen species was found to be potentially related to an effective response against PWN infection. Xu *et al.* (2013)

analyzed ESTs from Masson pine (*Pinus massoniana*) at different time points and identified genes that were soon upregulated related to signal transduction, transcription and translation, and secondary metabolism; and stress response genes that were upregulated only later.

These studies can provide valuable information on the molecular mechanisms that control the responses of pine trees to PWN infection, and genes differentially expressed between susceptible and tolerant/resistant plants could serve as “markers” to assist resistance-oriented breeding programs. Selecting candidate resistant trees from a natural population or plantation severely affected by PWD represents the basis of resistance breeding. Clonal propagules or seedling progenies of selected candidates are often used in artificial inoculation experiments to assess their tolerance to PWN infection and the stability of the tolerance. Whenever possible, this assessment should be preferentially conducted under field conditions to account for environmental interactions (Carson and Carson, 1989). Since resistance to PWN is most probably determined by several genes, identification of marker genes linked to a marker-assisted selection (MAS) can be an effective way to accumulate different resistance-related genes in a population or an individual with long-term utility, reducing the scale and time of breeding compared to traditional breeding procedures (Paterson *et al.*, 1991). Nevertheless, genetic diversity within a population must imperatively be maintained or else tree productivity, evolutionary potential, and resistance may be compromised (Carson and Carson, 1989). Additionally, pathogen population may also experience some genetic shift that can undermine the effect of resistance breeding (Burdon, 2001).

Recent studies with Japanese black and red pines have developed DNA markers (Lian *et al.*, 2000), and, in the future, QTL analysis will be performed concerning PWD resistance with a linkage map built with the DNA markers (Isoda *et al.*, 2007). Moreover, resistance genes were also identified by gene-expression profiling (ESTs) of PWN-inoculated Japanese black and red pines as already mentioned (Shin *et al.*, 2009; Nose and Shiraishi, 2011; Hirao *et al.*, 2012). Highly efficient breeding approaches employing MAS will lead to great progress in the achievement of PWD-resistant trees although resistance is not an absolute term. Resistant trees might still suffer some damage when infected with the PWN, especially if the environmental conditions are unfavorable for the trees. Tree age, virulence and pathogenicity of PWN populations, and the population density of the vector beetles are also factors that influence tree’s tolerance to PWN infection, and that should be taken into account if the objective is to maintain healthy pine forests (Nose and Shiraishi, 2008).

1.3. miRNAs

The interaction of plants with the diverse pathogenic organisms co-occurring in their native environment brings about many changes in the expression patterns of genes involved in plant-pathogen interaction. MicroRNAs (miRNAs) have been shown to play an important role in regulating genes involved in a multitude of plant stress responses, including biotic stress responses (Pareek *et al.*, 2015). miRNAs are small (20–24 nucleotides), noncoding RNAs with important roles in the regulation of gene expression in processes that determine development and defense responses in plants (and animals) (reviewed in Jones-Rhoades *et al.*, 2006; Carthew and Sontheimer, 2009; Ruiz-Ferrer and Voinnet, 2009; Rogers and Chen, 2013). Four main distinct classes of small RNAs have been described in plants: miRNAs, natural antisense small interfering RNAs (nat-siRNAs), transacting small interfering RNAs (ta-siRNAs), and nonclassified small interfering RNAs (siRNAs) (Hewezi *et al.*, 2008). miRNAs represent a new class of noncoding endogenous small RNAs which act as negative regulators of gene expression. Initially, miRNAs have been identified simply by cloning and sequencing size-fractionated RNA molecules. After the advent of high-throughput sequencing technology, large numbers of miRNAs from different plant species and genetic materials have been identified (Meyers *et al.*, 2006) and linked to the regulation of several biological processes. In particular, miRNAs were found to be implicated in plant–cyst nematode interactions in *Arabidopsis thaliana*, with traditional cloning and sequencing-based methods (Hewezi *et al.*, 2008), and in soybean, by using high-throughput sequencing (Li *et al.*, 2012). Besides their mentioned role in stress responses, they have a major role in regulating different aspects of plant growth and development, signal transduction, protein degradation, and response to environmental stresses (Jones-Rhoades *et al.*, 2006; Khraiweh *et al.*, 2012; Kumar, 2014; Pareek *et al.*, 2015). The other three classes of small RNAs differ from miRNAs in that they arise from long double-stranded RNAs that are processed into several small single-stranded RNA molecules, sometimes collectively called siRNAs; whereas miRNAs are the processing product of distinct genes (Hewezi *et al.*, 2008).

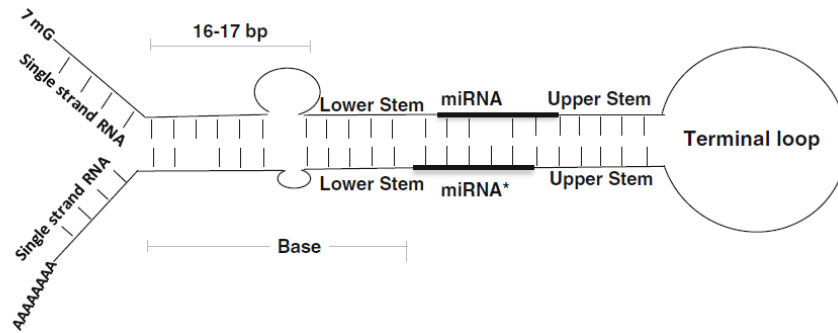


Figure 5 – General features of primary miRNA (pri-miRNA). (Retrieved from Pareek *et al.*, 2015)

Generally, miRNA biogenesis starts with the transcription by RNA Polymerase II (Pol II) of noncoding nuclear miRNA (*MIR*) genes, mostly localized in inter and intragenic (intron) regions of the genome, and very few of them in the 5'- or 3'-untranslated (UTR) regions. The recruitment of Pol II is determined by the interaction between several transcriptional activators and various sequence motifs in *MIR* promoter regions (Rogers and Chen, 2013). The produced capped (7-methylguanosine cap; 7mG) and polyadenylated long primary miRNA transcripts (pri-miRNAs) (Fig. 5 and Fig. 6) then fold back into imperfect hairpin structures which are recognizable by the members of Dicer-like (DCL) family enzymes. Of the DCL family, DCL1 is known to be the mainly responsible for the processing of the pri-miRNA into a precursor-miRNA (pre-miRNA) in *Arabidopsis*, and for the subsequent cleavage of pre-miRNA to release a miRNA/miRNA* duplex (Axtell *et al.*, 2011). DCL1 cleaves pri-miRNAs' stem-loop structure mostly at the lower stem region, 16–17 bp from the single-strand–double-strand junction (Fig. 5). In a few cases, the terminal loop controls this processing (Pareek *et al.*, 2015). According to the DCL family member that acts in miRNA biogenesis, the product miRNA length varies. The majority of plant miRNAs are 21 nucleotides (nt) long and are either processed by DCL1 or DCL4 proteins. DCL2 and 3 usually generate microRNAs that are 22 and 24 nt in length, respectively (Xie *et al.*, 2004, 2005; Akbergenov *et al.*, 2006; Deleris *et al.*, 2006; Cuperus *et al.*, 2010). These different lengths are proposed to be a consequence of the differences in DCL proteins' molecular structures, in particular in the distance between the RNase III active site and PAZ domains (Macrae *et al.*, 2006). The processing of pri-miRNAs to pre-miRNAs also requires the double-stranded RNA-binding protein HYPOASTIC LEAVES1 (HYL1) and the C2H2 zinc finger protein SERRATE (SE), along with the cap-binding proteins (CBP) CBP20 and CBP80 (Kim *et al.*, 2008; Pareek *et al.*, 2015) (Fig. 6). Pre-miRNAs are unstable in the nucleus and are readily cleaved by DCL1 into a miRNA/miRNA* duplex, which is stabilized by the S-adenosyl methionine-dependent methyltransferase HUA ENHANCER 1 (HEN1)

(Fig. 6). Methyl groups placed on the 3' terminal nucleotides of each of the duplex's strand prevents their uridylation and consequent degradation by exonucleases (Li *et al.*, 2005; Yang *et al.*, 2006; Ramachandran and Chen, 2008). Then, miRNA/miRNA* is exported into the cytoplasm by HASTY (HST) (Park *et al.*, 2005), where the duplex separates and the guide miRNA strand is loaded by Argonaute (AGO) proteins into the RNA-Induced Silencing Complex (RISC) (Fig. 6). What defines which strand of the duplex is the miRNA strand, and which is the miRNA* strand, is the thermodynamic stabilities of the 5' ends linked to AGO. Most plant miRNAs carry a 5' uridine that is usually bound by AGO1. So, the 5' end of the miRNA strand has a lower thermodynamic stability than that of the miRNA* and is preferably loaded into RISC, whereas the miRNA* strand is typically excluded from this complex and is thought to be quickly degraded (Rogers and Chen, 2013; Pareek *et al.*, 2015). The dissociation of the miRNA* strand from AGO1-miRNA/miRNA* complex does not require AGO1 endonuclease activity; instead, dissociation of the HEAT SHOCK PROTEIN90 (HSP90) and SQUINT (SQN) AGO1-associated proteins from the complex promotes miRNA* removal (Iki *et al.*, 2010; Iki *et al.*, 2012; Carbonell *et al.*, 2012) (Fig. 6). The final product of miRNA biogenesis is then a small single-stranded RNA incorporated into a silencing complex. This fully assembled complex binds to its target through sequence complementarity with the miRNA strand, leading to posttranscriptional gene silencing (PTGS) (Fig. 6), mainly through mRNA cleavage or translational inhibition (Rogers and Chen, 2013) depending on the degree of their complementarity. If there is near-perfect complementarity between miRNA and its target, the result will be mRNA cleavage; if the mRNA does not have sufficient complementarity to the miRNA but does have a suitable clustering of miRNA-binding sites at its 3'-UTR, target gene expression will be regulated by miRNA-directed translational repression. Most plant microRNAs have single, extensive complementary target sites in mRNA open reading frames (ORFs), as was early on observed for evolutionarily conserved miRNAs from *Arabidopsis* (Rhoades *et al.*, 2002). Target recognition is determined by the 5' end domain of miRNAs, from nucleotide position 2 to 7, and the cleavage occurs at the 10th or 11th nucleotide independently of the miRNA length (Bartel, 2009; Pareek *et al.*, 2015). Extensive pairing was found to encompass nucleotides 9–11, suggesting a “slicing” mode of action similar to siRNA-directed silencing (Voinnet, 2009). Therefore, miRNA-directed target mRNA cleavage is the most common type of PTGS found in plants. Besides the PTGS mechanisms, some miRNA variants are also involved in DNA methylation, a well-known epigenetic regulation, and transcriptional gene silencing mechanism. AGO4, AGO6 and AGO9, all reported to be involved in RNA-directed DNA methylation (RdDM) in *Arabidopsis*, associate with 24-nt long small RNAs

generated from dsRNAs by the action of RNA-dependent RNA polymerase 2, RNA polymerase IV and DCL3 (Law and Jacobsen, 2010; Havecker *et al.*, 2010; Pareek *et al.*, 2015).

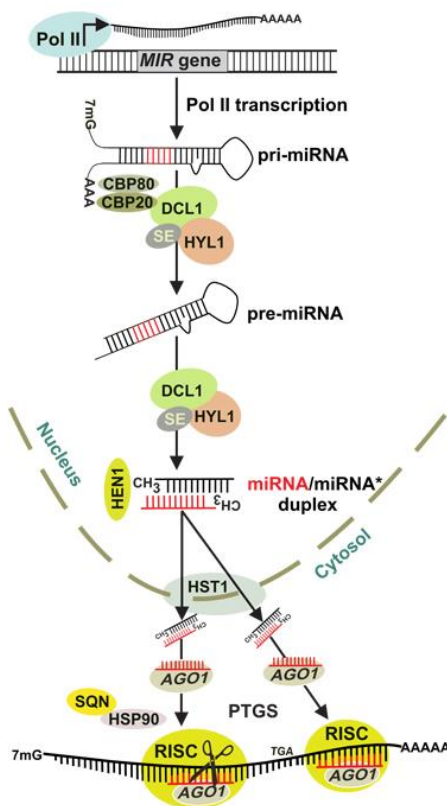


Figure 6 – Biogenesis of plant miRNAs. The primary miRNA (pri-miRNA) transcript is transcribed from the microRNA (*MIR*) gene by RNA polymerase II (Pol II). Pri- to pre-miRNA conversion proceeds through splicing and further processing in nuclear dicing bodies mediated by HYL1 and SE and cap-binding proteins CBP20 and CBP80. The processed pre-miRNA yields a miRNA/miRNA* duplex that is stabilized by methylation by HEN1 and transported to the cytoplasm by HST1. The selected miRNA strand is incorporated and stabilized in AGO1, and the miRNA* strand removal is facilitated by HSP90 and SQN. miRNA-guided AGO1-containing RISC directs posttranscriptional gene silencing (PTGS) mechanisms, either mRNA cleave or translation inhibition of target transcripts. (Adapted from Khraiwesh *et al.*, 2012)

Understanding the functional scope of plant miRNAs is considered to be one of the major current challenges in plant biology. The introduction of next-generation sequencing (NGS) in the past decade, along with powerful computational methods, has led to the genome or transcriptome-wide *in silico* identification of a massive amount of miRNA precursors in both model and non-model plants (Budak and Akpinar, 2015). Besides the identification of miRNAs and their precursors, miRNA targets are being confidently predicted through computational methods in plants as most miRNAs have high complementarity with their targets (Rhoades *et al.*, 2002; Zhao *et al.*, 2015). These bioinformatics tools are based on the three major distinctive features of miRNAs: hairpin-shaped stem-loop secondary structures, conserved highly complementary target regions, and high minimal folding free energy index (Kumar, 2014). Additionally, degradome sequencing data can complement these tools in the identification of biologically important and/or strongly regulated targets, establishing a direct link between a miRNA and its target(s) (Li *et al.*, 2014). The degradome comprises the products of miRNA-directed cleavage. These cleavage products have a ligation-competent mRNA end with a 5' phosphate that enables their distinction from other

degraded mRNAs isolated by standard methods, as the later are ligation incompetent during sequencing library preparation because they have a 5' cap or lack the 5' phosphate. By matching the 5' end sequences of the cleavage products back to the corresponding genome, potential cleavage sites are identified and, by comparing the cleavage sites to the known or novel miRNA sequences, miRNA–target pairs can be established with confidence (German *et al.*, 2008).

1.4. Objectives

Pine wilt disease, caused by the pine wood nematode *Bursaphelenchus xylophilus*, is one of the most serious and damaging diseases that has threatened pine forests worldwide and caused significant economic losses. Resistance against PWN infection seems to be linked to an early response that prevents the progression of the infection and stops the symptoms of the disease. While the physiological changes that occur throughout the disease progression have been characterized at the anatomical and biochemical levels, the molecular mechanisms that might be associated with susceptibility/tolerance of trees to PWN infection remain poorly understood. High-throughput screening procedures have been used to identify genes that are differentially expressed between tolerant/resistant and susceptible plants opening opportunities to unravel those mechanisms.

In recent years, microRNAs have been shown to be involved in plant response against pathogens, including nematodes, through sequence-specific silencing of plant transcripts that might have a role in the response to pathogen infection. Differentially expressed small RNAs and their targets, which provide defense mechanisms against pathogens, can be identified by small RNA sequencing and computational methods, and experimentally confirmed by high-throughput methods like real-time quantitative PCR (Kumar, 2014).

This Master project aims at the molecular analysis of *Pinus pinaster* plants exhibiting differential behavior (susceptibility and tolerance) towards PWN infection, focusing on the identification of miRNAs and their target genes, and transcripts highlighted in previous studies, putatively involved in the plant response mechanisms. *P. pinaster* plants used in this study were derived from seeds of a tree selected from a program initiated in 2009, which phenotypically identified over five hundred candidate resistant adult trees from an area with high incidence of PWD in Portugal (Ribeiro *et al.*, 2012). In a previous study, seedlings obtained from the seeds of

ninety-six of those candidate trees (96 families) were inoculated with *B. xylophilus* to evaluate their tolerance to the infection, and genetic variability associated with the survival of half-sib seedlings was detected (Lisboa, 2016). Making use of this information, the plants used in this study belonged to one of those families in which a differential behavior (susceptibility or tolerance) towards the PWN infection was observed between individuals. After the inoculation of plants with the PWN, the objective was to select some plants apparently tolerant to the infection and some susceptible ones to conduct expression analyses of selected transcripts potentially involved in the response mechanisms against PWN infection, and that might explain the susceptibility or tolerance to PWD.

Samples collected early after inoculation from the susceptible and tolerant *P. pinaster* plants were sent for small RNA and degradome next-generation sequencing in order to identify small RNAs differentially expressed between susceptible and tolerant samples, and to predict target transcripts for the differentially expressed small RNAs using computational approaches. Additionally, other transcripts were selected from a previous study that identified expressed sequence tags differentially expressed in *P. pinaster* susceptible plants *versus* plants from a tolerant species, *Pinus pinea*, inoculated with PWN.

Expression analysis of small RNA targets and transcripts putatively involved in the response mechanism to PWD should allow the identification of genes with different expression profiles between susceptible and tolerant *P. pinaster* plants, that might be linked to the differential behavior of those plants to PWN infection, and thus aid the elucidation of the response mechanisms against PWN infection by the identification of potential marker genes related to the disease.

2. Materials and Methods

2.1. Biological material

2.1.1. *Pinus pinaster*

Fifty *P. pinaster* seedlings (2-year-old) were used in this study. The seedlings were kept in a greenhouse with a cooling system, with a relative humidity of 70%, and were watered every 2 days. They were provided by Dr. Isabel Carrasquinho from Instituto Nacional de Investigação Agrária e Veterinária, I. P. (INIAV, Oeiras) and were derived from seeds of a tree selected from a program initiated in 2009, aiming to phenotypically identify candidate resistant trees from an area with high incidence of PWD in Portugal (Ribeiro *et al.*, 2012). Seedlings derived from the same candidate tree were previously used in a study that concluded that there is genetic variability associated with the survival of seedlings derived from the same tree to PWN infection (Lisboa, 2016).

2.1.2. *Bursaphelenchus xylophilus*

Bursaphelenchus xylophilus inoculum was prepared by collaborators in INIAV, Oeiras. *B. xylophilus* nematodes, originally isolated from infected pines, are maintained in the laboratory in culture media containing *Botrytis cinerea* at 25±1°C, and are part of INIAV's nematology collection. To prepare the inoculum, nematodes were extracted from the culture using the Baermann funnel technique (Southey, 1986). A water suspension of nematodes was adjusted to a concentration of 1000 nematodes/mL.

2.2. Inoculation of *Pinus pinaster* plants with *Bursaphelenchus xylophilus*

Fifty 2-year-old potted *P. pinaster* plants were inoculated, in collaboration with Dr. Isabel Carrasquinho (INIAV, Oeiras) and as part of the PhD work of Inês Modesto (Forest Biotechnology Lab, iBET, ITQB NOVA, Oeiras), following the method of Futai and Furuno (Futai and Furuno, 1979; Futai, 2003). Briefly, a small longitudinal wound was made in the stem approximately 10 cm from the apex and then covered with cotton. Afterwards, a 500 µL aliquot of the water suspension of 1000 nematodes/mL was pipetted into the cotton in the proximity of the wound and parafilm was wrapped around the inoculated regions to prevent drying of the inoculum (Appendix A). The same

procedures were performed in the control plants, which were inoculated with sterile water instead of the nematode suspension. Seventy-two hours after inoculation, a 2–4 cm portion of the stem (containing the inoculation site) and the needles immediately above that region were collected and stored at -80°C . Part of the collected biological material of *P. pinaster* was used for all the studies performed in this work and described below.

2.3. Total RNA extraction

Based on the contrasting responses towards inoculation, as evaluated after at least 4 weeks following the inoculation, a set of the initial fifty inoculated plants were selected for further analysis. Three water-inoculated control plants, five susceptible plants, and five apparently tolerant plants were selected. Stems and needles from those plants were finely ground to powder in liquid nitrogen, and various RNA extraction procedures were tested. A protocol allowing the effective isolation of total RNA, including the small RNA fraction, from the stems and needles was established based on the CTAB method described by Chang *et al.* (1993). 70 mg of sample was homogenized in 1 mL of pre-warmed (65°C) extraction buffer (300 mM Tris-HCl pH 8.0, 25 mM EDTA pH 8.0, 2 M NaCl, 2% CTAB, 2% PVP10, 0.05% Spermidine trihydrochloride and 2% β -mercaptoethanol) and the mixture vortexed and incubated for 10 min at 65°C with agitation. 1 mL of ice-cold chloroform:isoamyl alcohol (24:1) was added twice to the homogenate in order to separate the aqueous upper phase (containing the RNA) from the organic phase. The mixture was shaken softly and centrifuged at 13,200 rpm for 10 min at 4°C . The aqueous layer was transferred to another tube and mixed with 0.1 volume of 3 M NaOAc (pH 5.2) and 1 volume of ice-cold isopropanol. The mixture was stored at -80°C for 3h. The nucleic acids (and any remaining carbohydrates) were pelleted by centrifuging at 13,200 rpm for 30 min at 4°C , and dissolved in 500 μL of Tris:EDTA (TE) buffer (pH 7.5). The solution was mixed with 1 volume of ice-cold isopropanol and 0.1% of SDS, and stored at -80°C for 3h in order to precipitate the RNA. The precipitate was pelleted by centrifuging at 13,200 rpm for 30 min at 4°C , washed with 1 mL of ice-cold absolute ethanol, air dried and dissolved in 40 μL of RNase-free water. RNA was stored at -80°C until use.

2.3.1. RNA samples purification and evaluation of RNA quality

RNA samples were treated twice with Ambion® TURBO DNA-free™ kit since one treatment was not sufficient for efficiently remove DNA from the RNA samples. Each treatment was performed in Biometra Tprofessional Thermocycler using 0.2 mL tubes. Briefly, to each 40 µL of sample 4 µL of 10x TURBO DNase Buffer and 0.5 µL of TURBO DNase (2 U/µL) were added. The samples were gently mixed and incubated at 37°C for 30 min. Then another 0.5 µL of TURBO DNase were added to each sample and tubes were incubated for another 30 min at 37°C. 4 µL of DNase Inactivation Reagent was then added to stop the reaction and eliminate the DNase and divalent cations. The samples were mixed and left at room temperature for 5 min with occasional agitation, and then centrifuged at 10,000 *g* for 2 min. The supernatant containing the purified RNA was placed in a new tube. RNA concentration and quality were estimated using Nanodrop ND-1000, and a 200 ng/µL aliquot of each sample was loaded in a 1.5% (w/v) agarose gel to check RNA integrity (Appendix B).

2.3.2. Total RNA and miRNA quantification using Qubit® fluorometer

Total RNA concentration was measured using Qubit® fluorometer and Qubit® RNA BR (Broad Range; 1 ng/µL to 1 µg/µL) or HS (High Sensitivity; 250 pg/µL to 100 ng/µL.) (Invitrogen, USA), depending on the initially estimated sample concentrations. Qubit® fluorometer first needs to be calibrated using the two standards included in the assay kit. RNA samples to be quantified were diluted to perform this assay. 1 µL of each sample was diluted in 4 µL of water (1:5 dilution). Qubit® working solution must be added to standards and samples in specific Qubit® assay tubes.

Qubit® working solution is prepared by diluting the Qubit® Reagent 1:200 in Qubit® Buffer (Fig. 7), at room temperature. The standards mixture was prepared by adding 10 µL of each standard to 190 µL of Qubit® working solution, and each sample tube was prepared by adding 2 µL of diluted RNA aliquot to 198 µL of Qubit® working solution (Fig. 7). All Qubit® assay tubes were vortexed 2-3 seconds and then incubated for at least 2 minutes at room temperature (Fig. 7). To read RNA concentration, the standard and sample tubes were inserted individually in the fluorometer (Fig. 7), and the concentrations in the original sample were determined taken into consideration the volume of sample added to the assay tube and the initial 1:5 dilution of the sample.

miRNA concentration was measured using Qubit® fluorometer and Qubit® miRNA Assay Kit (Invitrogen), using the same methodology used for total RNA concentration measurement but

using a different working solution (Qubit® microRNA Reagent diluted 1:200 in Qubit® microRNA Buffer) and two siRNA 21-mer standards (GAPDH siRNA) provided in the assay kit. This assay is accurate for initial small RNA sample concentrations ranging from 0.05 ng/μL to 100 ng/μL.

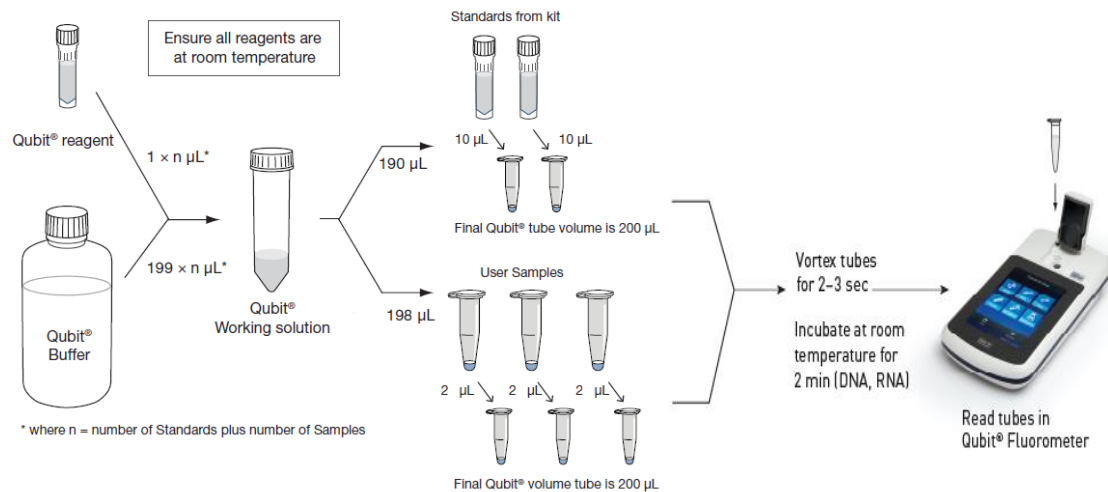


Figure 7 – Qubit quantitation assay kit workflow. Each assay kit provides concentrated Qubit reagent, dilution buffer, and prediluted standards. Qubit working solution is prepared by diluting the Qubit reagent in Qubit buffer. The standards mixture was prepared by adding 10 μL of each standard to 190 μL of Qubit working solution, and each sample tube was prepared by adding 2 μL of diluted RNA aliquot to 198 μL of Qubit working solution. All tubes were vortexed 2-3 seconds and then incubated for at least 2 minutes at room temperature. RNA concentration in each sample tube is determined using the Qubit fluorometer and determined standards concentrations. (Adapted from lifetechnologies.com/qubit)

2.4. Small RNA and degradome sequencing

The RNA samples (from stems of three water-inoculated control plants, five susceptible plants, and five apparently tolerant plants) were sent for small RNA sequencing using Illumina HiSeq 2000 (LC Sciences, Houston, Texas – USA), along with two pools, one of them containing RNA from the stems of five susceptible plants and the other containing RNA from the stems of five tolerant plants, for degradome sequencing as a mean of identifying small RNA target transcripts. The samples were prepared to ship according to the company’s International Shipping Instructions (<http://goo.gl/PcWlvj>). Samples were prepared for small RNA and degradome sequencing by the service provider using the Illumina® TruSeq® Small RNA Sample Preparation protocol. Briefly, this protocol specifically binds Illumina adaptors to both the 5’ phosphate and the 3’ hydroxyl group leaved by the activity of Dicer or other RNA processing enzymes on mature miRNAs and other small RNAs. An RT reaction is then performed to generate single stranded cDNA molecules, which are

subsequently PCR amplified using a combination of a common primer and a primer containing one of the 48 possible index sequences (tags) for each sample. These tags distinguish different samples from one another in a single lane of a flow cell, allowing the samples to be processed in parallel through the RT-PCR amplification process. After PCR amplification, individual libraries with unique indexes obtained from separate samples can be pooled and gel purified together. Then, the libraries go through a quality control analysis and cluster generation before being sequenced.

2.5. Bioinformatic analyses of sequencing results

The bioinformatic analyses described in this section were performed with the support of Bruno Costa (Forest Biotechnology Lab, iBET, ITQB NOVA, Oeiras).

2.5.1. Small RNA sequence processing

After removal of the adaptor sequences (“Adaptor trimmed NGS data”, Step 1 in Fig. 8), reads from each library were further processed individually using miRPursuit (Fig. 8) (Costa *et al.*, 2016), a pipeline developed in the Forest Biotechnology lab (iBET, ITQB NOVA, Oeiras) constructed around several modules of the University of East Anglia small RNA workbench (UEA sRNA WB) (Stocks *et al.*, 2012). In the initial “Preprocessing Quality control” and “UEA sRNA WB Filter” sequential steps (Step 1 in Fig. 8), low complexity (sequences containing less than 3 distinct nucleotides) and low abundant reads (less than 5) are removed, and only those with a length ranging from 18 to 26 bp and no match to plant rRNA/tRNAs are kept for further analysis (these parameters can be set in the *wbench_filter.cfg* configuration file – Appendix C.1). Quality plots and size profile distribution graphs before and after filtering were made for each library. In the following steps (Step 2, Fig. 8), the small RNAs were predicted by mapping the reads from step 1 to the setup genome file, with no mismatches allowed using PatMaN (Pattern Matching in Nucleotide databases; Prüfer *et al.*, 2008) (“patman Alignment”, Step 2 Fig. 8). Due to the lack of a reference *P. pinaster* genome, the reference genome of *Pinus taeda* v1.01-masktrim was used as the setup genome file. The mapped reads were then mapped to the miRBase database release 21 (<http://www.mirbase.org/>) using miRProf (Step 2, Fig. 8), from the UEA sRNA WB, with no mismatches allowed to identify putative conserved miRNAs (the parameters are set in the *wbench_mirprof.cfg* configuration file – Appendix C.2). The small RNA reads that did not map to miRBase were classified as “Non-conserved reads” and put in a different file. The identified

conserved miRNAs were further grouped by match signature, organism, and miRNA family. Non-conserved reads were used to identify putative ta-siRNAs reads (“Putative tasi reads”, Fig. 8), with the aid of the TaSi Prediction tool on the UEA sRNA WB (Step 3, Fig. 8) and the genome of *P. taeda* to identify putative *TAS* genes from which phased 21-nt long small RNAs are generated (Khraiwesh *et al.*, 2012) (the parameters used by the TaSi Prediction tool are set in the *wbench_tasi.cfg* configuration file – Appendix C.3). To predict “Putative Novel miRNAs”, using the miRCat tool in the UEA sRNA WB (Step 4, Fig. 8), the non-conserved reads were mapped to the setup genome file to find their respective precursor sequences (the parameters used by miRCat are set in the *wbench_mircat.cfg* configuration file – Annex C.4). The total number of sequence reads kept in each step of the miRPursuit workflow were calculated for each library.

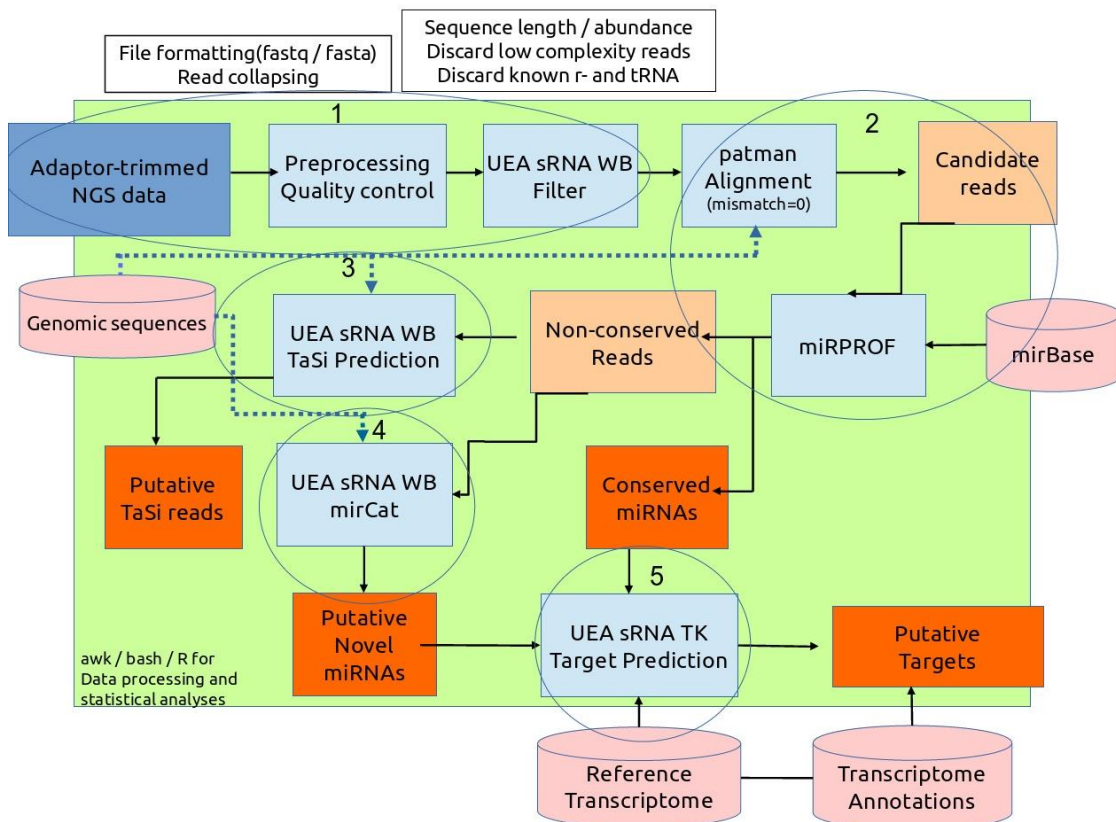


Figure 8 – miRPursuit workflow. Step 1 – After removal of the sequencing adaptor from the reads (“Adaptor trimmed NGS data”), reads from each library are further processed individually by “Preprocessing Quality control” and “UEA sRNA WB Filter” sequential steps, in which low complexity and low abundant reads are removed, and only 18 to 26 bp length and no match to plant rRNA/tRNAs reads are kept for further analysis. Step 2 – small RNAs are predicted using PatMaN (“patman Alignment”) by mapping the reads from step 1 to the setup genome file (*Pinus taeda* v1.01-masktrim), with no mismatches allowed; conserved miRNAs are identified using miRProf: reads are mapped to the miRBase with no mismatches allowed. Conserved miRNAs are further grouped by match signature, organism, and miRNA family. Step 3 – Non-conserved reads are used to identify putative ta-siRNA reads, using the TaSi Prediction tool and the setup genome, and also to identify putative novel miRNAs, using miRCat and the setup genome (to find their respective precursor RNAs – Step 4). Step 5 was not performed as described in the figure (see 2.5.3 for more information). (Retrieved from Costa *et al.*, 2016)

2.5.2. Differential expression analysis of sequenced small RNAs

Differential expression analysis of the identified conserved miRNAs, putative novel miRNAs and putative ta-siRNAs raw read counts from the RNA sequencing (RNA-seq) data was performed using edgeR Bioconductor software package (Robinson *et al.*, 2010). Reads with very low counts across all libraries provide little evidence for differential expression; therefore, the raw read counts in the RNA-seq libraries (Lib01 to Lib05) were first filtered for their abundance. Only reads with counts per million higher than 1 in at least two libraries were kept to further analysis. Raw read counts, given as input, are used for a design matrix with the group identification (control, susceptible or tolerant), and library size or sequencing depth is attributed to each sample after filtering. The design matrix describes how the libraries are grouped together. The differential expression analysis was made between two defined contrasts in the design matrix: Control vs. Susceptible and Control vs. Tolerant.

edgeR applies statistical methods on replicated read counts arising from RNA-seq technologies, based on the Generalized Linear Model (GLM) likelihood ratio test. Assuming that small RNA abundances follow a gamma distribution between replicate RNA samples then read counts follow a negative binomial probability law (Chen *et al.*, 2014). The biological coefficient of variation (BCV) was calculated by the estimation of the dispersion of the negative binomial model to which the filtered read counts are fitted. An empirical Bayes shrinkage was done to calculate read-wise variance. This allows reads that are consistent through replicate libraries to be ranked higher than reads that are not, resulting in a more stable inference of differential expression. Once the negative binomial models were fitted and BCV estimated, the differential expression was determined from replicated count data by conducting GLM likelihood ratio tests on the defined contrasts. The differential expression is automatically adjusted for varying sequencing depths, which is translated into fold-change or p-value calculations. The reads were ranked on evidence for differential expression based on the p-value computed for each read. Corrected p-values were calculated for each contrast. Reads were only considered to be differentially expressed if their corrected p-value was below 0.05. The correction for multiple testing was done by the false discovery rate (FDR). FDR values above 0.05 were not considered differentially expressed, and their log fold change (logFC) is computed as zero. The raw reads identified as differentially expressed were compared to the reads annotated by the miRPursuit workflow (see 2.5.1). Annotated reads were used to construct a heatmap with their logFC values.

2.5.3. Small RNA target gene prediction and validation

Target gene prediction for all the identified conserved and novel miRNAs and ta-siRNAs can be performed using miRPursuit, and the Target Prediction tool of the UEA sRNA WB (Step 5, Fig. 8). However, in the scope of this work target prediction was only made for the small RNAs that were identified as differentially expressed between contrasts (Control vs. Susceptible and Control vs. Tolerant) and that were annotated by the miRPursuit workflow. Target prediction for those small RNAs was performed using CleaveLand4 (Addo-Quaye *et al.*, 2009). The small RNAs identified as differentially expressed were mapped to the SustainPine v3.0 transcriptome (SustainPineDB, <https://goo.gl/1uZ8bD>) to identify potential cleavage sites, which were validated with the reads obtained from degradome sequencing. Transcript fragments of both susceptible and tolerant degradome libraries were first blasted against NCBI nt database, prior to this analysis, to examine their quality. The best hits (e-value below $1e^{-5}$) were extracted.

2.6. Selection of small RNA target genes for expression analysis

From the list of identified small RNA targets, nine transcripts were selected to evaluate their expression level in susceptible and tolerant samples. The selection of targets was made based on the differential expression of the respective small RNA(s) between contrasted samples, given by their read count data in each RNA-seq library. The selected target transcripts include Tetraspanin family protein, Class IV chitinase, Unassigned protein, DRE-binding protein DREB1, Translation initiation factor eIF-2 gamma subunit, 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase, Phosphoglycerate kinase, DNMT1-RFD multi-domain protein, and Gag-spuma domain containing protein. Primers for those genes were designed using PerlPrimer v1.1.21 (Marshall, 2004) using the following criteria: annealing temperature from 56 to 62°C (with a maximum difference between primer pairs of 1°C), GC content of 40–60°C, primer length of 18–24 bp, amplicon size of 100–200 bp, and a 3' GC clamp (two of the last three bases are G or C). Primers were designed to amplify the region downstream the small RNA alignment region. Primer sequences, as well as target transcript sequence accession numbers and other useful information, are described in Table G1, Appendix G.

2.7. Selection of candidate PWD-related transcripts for expression analysis

Eight candidate genes putatively involved in a differential behavior (susceptibility or tolerance) of *Pinus* plants towards PWD were selected from a previous study by Santos *et al.* (2012). The selected genes include a terpenoid metabolism gene (termed *TerpMet*), a probable photoassimilate-responsive protein 1 (*PAR1*), a protein from a family induced by abscisic acid (ABA) and water deficit stress (WDS) (termed *ABA/WDS-induced protein*), a late embryogenesis abundant (*LEA*) protein, a flavin mononucleotide reductase (*FMN reductase*), a SNAP (Soluble NSF Attachment Protein) receptor (*SNARE*), a 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) synthetase, and a ricin B-related lectin (*RicinB*). Candidate gene sequences were obtained from the *P. pinaster* transcriptome database, SustainPineDB, and from GenBank. Primers were designed using PerlPrimer and the criteria described in section 2.6. Primer sequences, as well as sequence accession numbers and other useful information, are described in Table G1, Appendix G. Primer sequences for *TerpMet*, *FMN reductase*, *SNARE*, and *RicinB* were retrieved from Santos *et al.* (2012). The working solution concentration of each primer was set to 10 μ M.

2.8. Two-step RT-qPCR

2.8.1. cDNA synthesis

cDNA synthesis was performed for nine needle and nine stem RNA samples (from three control plants, three susceptible plants, and three tolerant plants) and for a pool of RNA combining all RNA samples of each type (needle or stem). First, Transcriptor High Fidelity Reverse Transcriptase (Roche Diagnostics, USA) was used, but the reverse transcription was not efficient for RNA samples with low A(260)/A(230) and A(260)/A(280) absorbance ratios, probably due to the presence of contaminants that inhibited the reaction or RNA degradation. Therefore, the SuperScript® IV Reverse Transcriptase (SSIV RT; Invitrogen), reported as resistant to a variety of reaction inhibitors and efficient even with “difficult” plant RNA samples (<https://goo.gl/R44Y57>) was used.

cDNA synthesis using SSIV RT followed the manufacturer's instructions and was performed in a Biometra Tprofessional Thermocycler. For each synthesis reaction, 250 ng of each RNA needle sample (or pool of RNA) and 200 ng of each RNA stem sample (or pool of RNA) were mixed in

individual 0.2 mL tubes with 1 μ L of 50 μ M Oligo d(T)₂₀ primer, 1 μ L of 10 mM dNTP mix and DEPC-treated water to a final volume of 13 μ L. The tubes were briefly centrifuged, heated at 65°C for 5 minutes and then incubated on ice for at least 1 minute. Then, the following reagents were combined in a reaction tube: 4 μ L of 5X SSIV RT Buffer, 1 μ L of 100 mM DTT, 1 μ L of RNaseOUT™ Recombinant RNase Inhibitor and 1 μ L of SSIV RT (200 U/ μ L). The components were mixed and briefly centrifuged, and added to the annealed RNA. The combined reaction mixture was incubated at 50°C for 10 minutes and then at 80°C for 10 minutes in order to inactivate the reaction. For the needle and stem pools, no-reverse transcription controls were made using the same reagents as a normal synthesis reaction except for the transcriptase. This control allows to check for DNA contamination in the RNA samples. All the cDNA samples were diluted 1:10 in Nuclease-Free Water, and stored at -20°C.

2.8.1.1. Diagnostic PCR

In order to check for cDNA quality after synthesis, a PCR was made using a positive control pair of primers, which were previously confirmed to amplify the 18S rRNA gene in *P. pinaster*. PCR was carried out in Biometra Tprofessional Thermocycler and 15 μ L reaction mixtures. Each reaction mixture was set up as follows (to the indicated end-concentrations): 8.5 μ L of autoclaved DEPC-treated water, 3 μ L of Green GoTaq® Flexi Buffer (1X; Promega), 1.5 μ L of MgCl₂ (2.5 mM; Promega), 0.3 μ L of dNTPs mix (200 μ M each; Promega), 0.3 μ L of each forward and reverse primers (0.2 μ M), 0.1 μ L of GoTaq® G2 Flexi DNA Polymerase (5 U/ μ L; Promega) and 1 μ L of cDNA template (or DEPC-treated water, for negative controls). *P. pinaster* genomic DNA (provided by a lab colleague) was used as positive control. PCR was carried out with an initial denaturation step at 95°C for 5 minutes, followed by 35 cycles of denaturation at 95°C for 50 seconds, annealing at 58°C for 45 seconds and extension at 72°C for 20 seconds, and one final step of extension at 72°C for 5 min. Samples were stored at 4°C. A 4 μ L aliquot of each PCR reaction was then directly run on an 1.5% (w/v) agarose gel submerged in 1X TE (20 mM Tris and 0.5 mM EDTA; pH 8.0), at about 80 V/cm for 40 minutes. After gel electrophoresis, PCR amplification products were visualized and photographed using Gel Doc EZ System or Gel Doc XR+ (Bio-Rad).

The practical optimum temperature for annealing was experimentally determined for each set of primers described in 2.6 and 2.7, using the cDNA from the pools of needle and stem RNA samples as template. PCR mixes were set up and the reactions carried out as described above, but with an annealing temperature gradient chosen for each primer pair based on the predicted

oligonucleotide's melting temperature (T_m). Then, each PCR reaction was run on an agarose gel and the amplification products were visualized as before.

2.8.2. Selection of reference genes for qPCR

Six potential reference genes were selected based on their previous use as internal controls in gene expression studies by RT-qPCR in *P. pinaster*. *18S rRNA* (Gonçalves *et al.*, 2005; Santos and Vasconcelos, 2012; Santos *et al.*, 2012), *Actin* (Pascual *et al.*, 2015), *α -Tubulin* (Vega-Bartol *et al.*, 2013), *Histone H3* (Vega-Bartol *et al.*, 2013), *40S rRNA* (Pascual *et al.*, 2015), and *elongation factor 1-alpha* (Pascual *et al.*, 2015; Vega-Bartol *et al.*, 2013) suitability for normalization of qPCR experimental assay data in the tissues under study was evaluated (see 2.8.3.3). Reference gene sequences were obtained from the *P. pinaster* transcriptome database, SustainPineDB, and from GenBank. Primers were designed using PerlPrimer and the criteria described in section 2.6. Primer sequences, as well as sequence accession numbers and other useful information, are described in Table G1, Appendix G. *40S rRNA* primer sequences were retrieved from Pascual *et al.* (2015). The working solution concentration of each primer was set to 10 μ M. The practical optimum temperature for annealing was experimentally determined for each set of primers as described in 2.8.1.1.

2.8.3. qPCR

All the qPCR experiments were carried out in 96-well white plates sealed with self-adhesive sealing foils using a LightCycler® 480 instrument and LightCycler® 480 SYBR Green I Master (all from Roche Diagnostics) to monitor the PCR amplification. Each 20 μ L reactions were composed by 10 μ L of SYBR Green Mater Mix, 6.9 μ L of Water, PCR Grade (Roche Diagnostics), 0.8 μ L of each primer (0.4 μ M) and 1.5 μ L of cDNA template. After the plate has been filled and sealed, it was centrifuged at 2,000 rpm for 2 minutes before loading it into the instrument. The following LightCycler experimental run protocol was used in all qPCR reactions: denaturation program at 95°C for 5 min; amplification and quantification program of 40 or 45 cycles (depending on the experiment) – 95°C for 10 s, 58 to 64°C (depending on the determined optimal annealing temperature for each pair of primers) for 15 s and 72°C for 12 s, with a single fluorescence measurement; melting curve program – 65 to 97°C with a heating rate of 0.1°C per second and a continuous fluorescence measurement; and finally a cooling step to 40°C. No-template controls were included for each primer pair, and for all the biological samples the expression level was

based on three technical replicates. A needle sample was always included in the plate as a calibrator sample, to normalize the values obtained for the same gene in different plates.

Data on the expression level (absolute quantification) of each sample was obtained in the form of crossing points (Cp) generated automatically by the LightCycler® 480 software v1.5.1.62 (Roche Diagnostics), employing the Second Derivative Maximum method. Only Cp technical repetitions with a Sample Variance below 0.5 were considered throughout all the following experimental analyses. Product melting curve analysis was also performed in the LightCycler software, to determine the amplification specificity of the primers under the experimental conditions. In addition, gel electrophoresis of the qPCR products was performed to further confirm the achievement of a single-product amplification.

2.8.3.1. Experimental determination of qPCR conditions

The cDNAs synthesized from RNA pools and respective no-reverse transcription (-RT) controls were used as template to validate the selected optimum annealing temperature (see 2.8.1.1 and 2.8.2) in qPCR for each set of primers, and to exclude genomic DNA contamination of the RNA samples, as already mentioned in 2.8.1, by the absence of amplification in -RT control reactions. cDNA templates used were in an 1:10 dilution and this experiment also served to determine the template concentration appropriate for the evaluation of expression levels of each gene in subsequent experiments (a dilution that would give Cps below 30 was considered acceptable), especially for the “forthcoming” primer’s efficiency calculation experiment. Two technical replicates of each sample were conducted, and no-template controls were included in the experiments.

2.8.3.2. Primer’s efficiency calculations

To determine the primer’s amplification efficiencies of the reference, the small RNA target and the PWD-related candidate genes, a standard curve using a range of five or six dilution series was generated using Microsoft Excel 2016, by plotting the Cps against the logarithm of template concentration. A linear regression model fitted to the log-transformed data of serially diluted template input concentrations plotted against their Cp values (Rasmussen, 2001; Pfaffl, 2001). The PCR efficiency (E) is calculated as follow: $E = 10^{-1/\text{slope}}$ ($1 < E < 2$), or $E (\%) = (10^{-1/\text{slope}} - 1) \times 100$. cDNA from the pool of needle RNA samples was used as template to obtain the primer’s efficiencies for the selected reference genes. Then, cDNA from the calibrator (a needle sample)

was used to obtain the efficiency of all the primers used in this study, as there was not enough material to perform this experiments using the pool of samples. Only Cp values below 35 were considered in the calculations.

2.8.3.3. Expression stability of candidate reference genes

For the candidate reference genes transcript abundance was determined by Cp values in the nine needle samples (from three control plants, three susceptible plants, and three tolerant plants, each biological replicates with three technical replicates). The expression stability of candidate reference genes was analyzed using three statistical programs: geNorm (Vandesompele *et al.*, 2002), NormFinder (Andersen *et al.*, 2004) and BestKeeper (Pfaffl *et al.*, 2004). Cp values obtained for each gene were converted into relative quantities (RQ) to be used as input data for geNorm and NormFinder programs, whereas raw Cp values were directly imported into BestKeeper. RQ were calculated as follows: $RQ = E^{(C_{P_{min}} - C_{P_{sample}})}$, where E is the amplification efficiency previously calculated for each gene with the pool of needle RNA samples (see 2.8.3.2), $C_{P_{min}}$ is the lowest Cp value obtained (which correspond to the sample with the highest expression) and $C_{P_{sample}}$ is the Cp value of the sample tested. For geNorm, RQ were calculated for each biological replicate with the mean Cp values of the three technical replicates; whereas for Normfinder, RQ were calculated for each technical replicate individually. In both programs, each biological replicate is analyzed in separate.

GeNorm is a Visual Basic Application for Microsoft Excel that calculates an expression stability value (M) for each candidate reference gene and then determines the optimal number of reference genes required for proper normalization by the pairwise variation $V_{n/n+1}$. It takes the two most stable genes ($n=2$), ranked from their M value, and calculates the variation resulting from the inclusion of the third most stable gene ($n+1=3$) in the ranking ($V_{2/3}$), the variation $V_{3/4}$ from adding to the three most stable genes ($n=3$) the fourth best-ranked gene ($n+1=4$), and finally the variation $V_{4/5}$ that comes from using all five candidate genes in analysis ($n+1=5$). A V_n/V_{n+1} cut-off value of 0.15 is recommended, below which the inclusion of the $(n+1)^{th}$ gene is not required for proper normalization of qPCR (Vandesompele *et al.*, 2002).

BestKeeper is an Excel-based software tool that like geNorm performs a pairwise comparison of candidate reference genes. Input data are raw Cp values and the amplification efficiency of each candidate reference gene. Descriptive statistics based on the Cp values are computed for each gene: geometric and arithmetic means (GM and AM), minimal (Min) and maximal (Max) values,

standard deviation (SD), and coefficient of variance (CV). The x-fold over- and under-expression of individual samples, towards the GM of Cp values, and the SD are calculated taking into account the amplification efficiencies. Genes considered to be stably expressed are combined into a *BestKeeper index* to which the geometric mean of the Cp values obtained for those stably expressed genes is attributed. Descriptive statistics were also computed to the *BestKeeper index*. Several pairwise correlation analyses are performed and characterized by the Pearson correlation coefficient (r) and the probability (p) value. Highly correlated genes are combined into a *BestKeeper index*, and a correlation between the index and each gene is calculated by a Pearson correlation coefficient (r) and a p-value. The output is the identification of the two best reference genes by their high r and low SD values.

NormFinder algorithm, available as a Microsoft Excel add-in, calculates the expression stability value (SV) for each candidate reference gene, by the estimation of intra- and intergroup variations. Thus, SV combines both sources of variation and represents a useful measure of the systematic error that will be introduced when using a particular gene. Lower SV indicates higher expression stability (Andersen *et al.*, 2004). The output is the identification of the best two reference genes from the set, that combine minimal inter- and intragroup expression variation, and the respective SV for the two-gene combination.

The selection of the best-suited reference genes for normalization of the qPCR assays with three control, three susceptible, and three tolerant plants was based on the conjugation of the results obtained in the three statistical programs described. The optimal number of reference genes was determined by geNorm, as from the three programs it is the only one that allows this determination.

2.8.3.4. Expression of predicted small RNA target transcripts

For the small RNA target genes transcript abundance was determined by Cp values in the nine stem samples (from three control plants, three susceptible plants, and three tolerant plants, each biological replicates with three technical replicates). Cp values of the three technical replicates were averaged and used to calculate RQ for each biological replicate, using the formula described in 2.8.3.3, taking into account the amplification efficiencies determined for each primer pair in 2.8.3.2. RQ obtained for each biological replicate were compared to a Normalization Factor (NF) obtained with the reference genes best-suited for normalization (determined in 2.8.3.3) as follows and for each small RNA target transcript: for each biological replicate a NF was calculated by the

geometric mean of the RQ values obtained for the selected reference genes for that biological replicate; then, the Normalized Relative Expression Level (NRQ) for each biological replicate was calculated dividing the RQ obtained for each biological replicate for a particular small RNA target transcript by the NF obtained for that biological replicate ($NRQ = RQ/NF$); and finally the NRQ of each of the three biological replicates was averaged to obtain a unique NRQ for control, susceptible and tolerant samples, respectively. The three NRQs obtained for each predicted target transcript were plotted as \log_2 fold changes of relative expression level.

2.8.3.5. Expression of PWD-related candidate genes

For the candidate genes transcript abundance was determined by Cp values in the nine needle and stem samples (from three control plants, three susceptible plants, and three tolerant plants, each biological replicates with three technical replicates). NRQs for control, tolerant and susceptible samples were obtained and plotted as described above (see 2.8.3.4) for each candidate gene.

3. Results and discussion

To uncover potential small RNA regulatory pathways and gain information about the molecular mechanisms involved in the differential behavior of *Pinus pinaster* plants towards the PWD, 2-year-old *P. pinaster* seedlings germinated from seeds of a tree selected for its apparent resistance to the PWD were inoculated with the pine wood nematode *Bursaphelenchus xylophilus*. After more than 4 weeks after the inoculation, around 60% of the plants showed apparent tolerance to the PWN infection (Fig. A1(D), Appendix A) and the remaining inoculated plants were already dead (Fig. A1(E), Appendix A). Based on these contrasting responses towards inoculation, a set of the initial inoculated plants were selected for further analysis. Three water-inoculated control plants, five susceptible plants, and five apparently tolerant plants were selected.

3.1. Total RNA isolation from *Pinus pinaster* plants with contrasting responses towards PWN infection

Total RNA samples were extracted from the stems (collected seventy-hours after inoculation) of three water-inoculated control (C) plants, five susceptible (S) plants, and five apparently tolerant (T) plants. Total RNA samples were also obtained from needles collected at the same time from the same three C plants, three S plants, and three T plants. Before selecting a protocol for total RNA extraction, several attempts were made to use a commercially available RNA extraction kit specifically developed for microRNA studies, as recommended by the sequencing service provider. The Plant/Fungi Total RNA Purification Kit (Norgen Biotek, Canada) was tested according to the manufacturer's instructions and including several alterations to the protocol (addition of half the lysis buffer followed by grinding with RNase-Free Pellet Pestles and addition of remaining buffer; increased incubation time with lysis buffer; additional centrifugation steps to remove cell debris and use of filtration column from other kits; column wash with ethanol in addition to the wash solution from kit, among others). However, it was not possible to obtain RNA of enough quality to proceed for sequencing. Therefore, a modified CTAB extraction method was optimized to enable its use in tissues with high polyphenol contents, which is the case of the pine tissues. The extraction buffer described in section 2.3 stabilizes polyphenols allowing the efficient isolation of total RNA. To increase the extraction's effectiveness and potentially enhance the recovery of low molecular weight RNA molecules, which is necessary for the RNA-seq experiments, additional precipitation

steps were performed (Chaves *et al.*, 2014). The A(260)/A(280) RNA absorbance ratio of all the samples varied from 1.59 to 1.92 (most samples were around 1.8), which reveals little or absent protein contaminants, and A(260)/A(230) ratio from 1.11 to 2.01 (most around 1.5), which could be indicative of some polysaccharide and/or phenol contaminants. Nevertheless, the RNA absorbance ratios followed the requirements of the sequencing service provider (LC Sciences) regarding RNA purity ($A(260)/A(280) \geq 1.8$ and $A(260)/A(230) \geq 1.0$). Additionally, the obtained yield was in the range of 3 to 14 μg of RNA per 70 mg of tissue.

3.2. Small RNA and degradome sequencing

Of the thirteen total RNA stem samples sent for small RNA sequencing, only five libraries were generated: two libraries of C plants (Lib01 and Lib02), two libraries of S plants (Lib03 and Lib04) and one library corresponding to a T plant (Lib05). According to the sequencing service provider, it was not possible to generate cDNA from small RNAs in all the sent samples due to some RNA degradation. Quality plots were generated with FastQC program for each RNA-seq library produced (Appendix D) to assess base calling accuracy, measured by the Phred quality score (Q score). Q score indicates the probability that a given base is called incorrectly and is given by $Q = -10 \log_{10}(p)$, p =probability of an incorrect base call. For example, a Q score of 30 is the equivalent of a one in a thousand times chance of incorrect base call, i.e. the base call accuracy is 99,9% (probability of a correct base call). The quality offset for Illumina HiSeq 2000 sequencing experiment is Q score of 33. Regarding the quality plots generated for each library (Appendix D) the first four or five nucleotides have lower quality due to the use of universal primers but they are still in the quality threshold defined for this experiment, and after about 28 nucleotides the quality begins to vary, which is not an issue since the raw reads were filtered in following steps and those nucleotides were clipped off.

A total of 7,174,676 (Lib01), 5,624,002 (Lib02), 8,253,606 (Lib03), 6,720,956 (Lib04) and 5,794,529 (Lib05) raw sequence reads were obtained for the five libraries (Table II). After the removal of reads shorter than 18 nt and longer than 26 nt, reads mapping to tRNA and rRNA (that might be degradation products) and low-complexity and low-abundance reads, 26.95% (Lib01), 34.35% (Lib02), 28.03% (Lib03), 38.02% (Lib04) and 35.05% (Lib05) of the reads remained for subsequent analysis (Filter WB, Table II). The size profile distribution of reads between 18 and 26

nucleotides showed that 21-nt long sequences were expressed at higher levels followed by 20-nt length reads (Appendix E). A substantial decrease of 20-nt length reads is seen when filtering with *P. taeda* reference genome. The dominance of the 21-nt long class is in accordance with what is observed in other conifer species (Dolgosheina *et al.*, 2008).

Table II – Total number of reads kept after consecutive steps of the miRPursuit workflow in each small RNA sequencing library.

Total	Lib01 (C)	Lib02 (C)	Lib03 (S)	Lib04 (S)	Lib05 (T)
Raw reads	7,174,676	5,624,002	8,253,606	6,720,956	5,794,529
Filter WB	1,933,768	1,932,027	2,313,418	2,555,006	2,031,130
Genome aligned	1,232,218	1,011,808	1,621,335	1,275,803	1,147,596
Conserved miRNA reads	96,208	104,796	50,682	134,206	105,689
Putative novel miRNA reads	66,456	80,332	54,906	101,530	97,562
Putative ta-siRNA reads	49,097	74,826	32,681	73,202	93,624

3.3. Selection of small RNAs targets putatively involved in the response molecular mechanisms of *Pinus pinaster* plants affected with PWD

3.3.1. Sequenced small RNAs and *in silico* evaluation of their expression profiles

Although many miRNAs have already been identified in several plant species, miRNAs from *Pinus pinaster* to date have been absent from miRBase. Many miRNA families are greatly conserved among plant species allowing their identification simply by performing BLAST searches against the annotated miRNAs on miRBase (Griffiths-Jones, 2006). Previously filtered reads were used to identify conserved miRNAs by mapping candidate small RNA reads to miRBase using miRProf. A total of 96,208 (Lib01), 104,796 (Lib02), 50,682 (Lib03), 134,206 (Lib04) and 105,698 (Lib05) conserved miRNAs reads were identified (Table II). According to these results, conserved miRNA reads were more abundant in control samples, followed by susceptible samples and finally by the tolerant sample. The candidate reads that were not mapped to miRBase (non-conserved reads) were further used as input in two tools, miRCat and TaSi Predictor (both from the UEA sRNA WB) to predict, respectively, novel miRNAs and ta-siRNAs by mapping the reads against the reference genome of *P. taeda* to search for miRNA precursors or *TAS* genes. A total of 66,456 (Lib01),

80,332 (Lib02), 54,906 (Lib03), 101,530 (Lib04) and 97,562 (Lib05) putative novel miRNA reads, and a total of 49,097 (Lib01), 74,826 (Lib02), 32,681 (Lib03), 73,202 (Lib04) and 93,624 (Lib05) putative ta-siRNA reads were identified (Table II). According to these results, putative novel miRNAs were more abundant in susceptible samples, followed by control samples and finally by the tolerant sample; whereas the ta-siRNA reads were, like conserved miRNAs, more abundant in control samples, followed by susceptible samples and finally by the tolerant sample. From all the small RNA classes, conserved miRNAs showed the highest abundance in roughly all the RNA-seq libraries.

Differential expression analysis of the conserved miRNAs, putative novel miRNAs, and putative ta-siRNAs raw reads counts from the sequenced data was performed using edgeR Bioconductor software package (Robinson *et al.*, 2010). After being filtered by abundance, the count matrix was reduced from 3 326 989 distinct reads to 78 001 distinct reads. Filtered read counts were inputted into the software as a design matrix with the group identification (control, susceptible or tolerant), that describes how the libraries are grouped together. RNA-seq read count data show strong mean-variance relationships (Chen *et al.*, 2014). EdgeR fits filtered read counts to negative binomial-based models to capture the mean-variance relationships, and to distinguish between biological and technical sources of variation (McCarthy *et al.*, 2012). Technical variation is the variation intrinsic of the sequencing technology whereas biological variation refers to changes in expression levels of each small RNA between libraries. RNA-seq technique makes it possible to disentangle these two sources of variation, even when the number of technical replicates is very small or inexistent (which happens in this study), because its technical variation level is of a predictable nature (Robinson and Smyth, 2008).

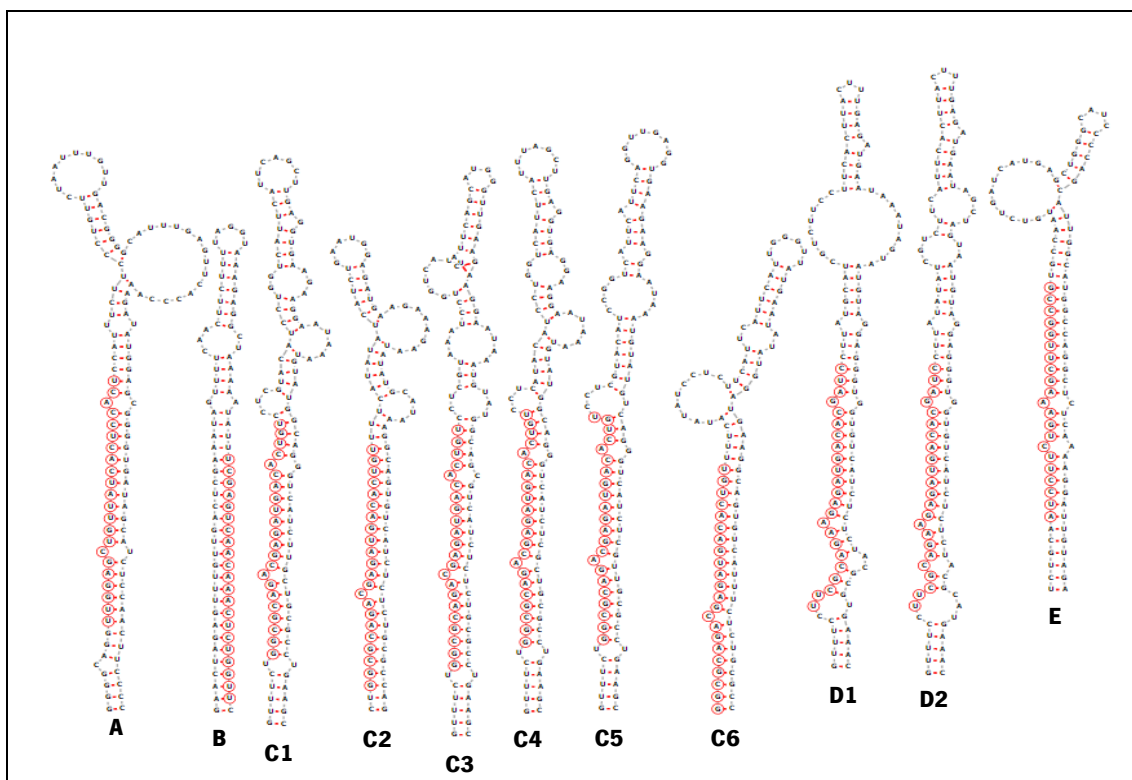
The biological coefficient of variation (BCV) represents the coefficient of variation (CV) that reflects the deviation in the true (unknown) abundance of the small RNA between replicated RNA samples, and the CV that would remain between the replicates if sequencing depth could be increased indefinitely. BCV does not decrease with the increased sequencing depth as does the technical CV; thus, the BCV is likely to be the main source of uncertainty for high-count reads. Estimation of the BCV is, therefore, essential for accurate assessment of differential expression in RNA-seq data. BCV was calculated by the dispersion of the negative binomial distribution of read counts. Once the negative binomial models were fitted and BCV estimated, the differential expression was determined using the Generalized Linear Model (GLM) likelihood ratio test. Several GLMs examine differential expression from replicated count data by conducting likelihood tests on

defined contrasts in the design matrix. In this study, the defined contrasts were Control vs. Susceptible and Control vs. Tolerant. Since the read counts are affected not only by the expression levels but also by the sequencing depth, edgeR automatically adjusts the differential expression analysis for the varying library sizes, which is translated into fold-change or p-value calculations. Each differentially expressed read identified in the analysis was ranked based on the p-value computed. To address the problem that having a low number of replicate libraries represents to accurate differential expression assessment, negative binomial dispersions (BCVs) are estimated using the Cox-Reid (CR) adjusted profile likelihood (APL) method. This method ensures that the estimated dispersion of each individual “tag” (small RNA read) takes into account all tags, since data from a single small RNA read is often insufficient for reliable estimation of the dispersion. This so-called tag-wise dispersion estimation allows reads that are consistent through replicate libraries to be ranked higher than reads that are not, resulting in a more stable inference of differential expression (not driven by outliers) (Chen *et al.*, 2014).

The differential expression analysis of all predicted small RNAs from the RNA-seq experiment resulted in the identification of a total of 3 246 differentially expressed reads for the Control vs Susceptible contrast, of which 1 742 were downregulated and 1 504 were upregulated; and a total of 383 differentially expressed reads for the Control vs Tolerant contrast, of which 323 were downregulated and 60 were upregulated. Reads were only considered to be differentially expressed if the corrected p-value of the likelihood ratio test was below 0.05. The correction for multiple testing was done by the false discovery rate (FDR). FDRs above 0.05 were considered as not differentially expressed and therefore their logFC was considered zero. A total of 3 337 differentially expressed reads were identified in the aggregate of both of these conditions. These differentially expressed reads were compared to the reads that were annotated by the miRPursuit workflow for annotation. Only eighteen differentially expressed reads were annotated by the workflow, and include nine putative novel miRNAs, six putative ta-siRNAs and three small RNAs of unknown class. No conserved miRNA was found differentially expressed according to the criteria used for analysis. These eighteen small RNAs are presented in Table F1 (Appendix F), alongside the logFC for each contrast, and the raw and normalized read counts in each RNA-seq library. A heatmap constructed with the logFC values of this eighteen differentially expressed small RNAs is represented in figure 9. To each differentially expressed small RNA a name in the form “Ppi-sRNA-classN” was attributed, where: *Ppi* identifies the species; small RNA is abbreviated as *sRNA*; the *class* is either “novel” for novel miRNAs, “tasi” for ta-siRNAs or “NA” (not attributed) for small RNAs of unknown class; and

N is a number to distinguish small RNAs from the same class. The three top small RNAs represented in the heatmap show the highest difference between contrasts, but overall there are small variations in the predicted expression levels between the two contrasts. The secondary structure of the precursors of putative novel miRNAs was predicted using the RNAfold WebServer (Gruber *et al.*, 2008) (Table III).

Table III – Predicted miRNA precursors secondary structures for some novel miRNAs in *Pinus pinaster* plants inoculated with the Pine Wood Nematode. Hairpin structures of miRNA precursors (pre-miRNA) were predicted by the RNAfold WebServer (Gruber *et al.*, 2008). The structures were edited using *forma*, an RNA secondary structure visualization web app (Kerpedjiev *et al.*, 2015). The mature miRNA sequence is highlighted in each structure by the red circles. The adjusted minimum free energy (MFE) and the number of mismatches of the pre-miRNAs are included.



Novel miRNA	Predicted pre-miRNAs	MFE (kcal/mol)	Mismatches
Ppi-sRNA-novel6 TGGAGCTGTTATCACTCCACT-novel-5334	A	-60.70	2
Ppi-sRNA-novel4 TCGAGTCAACAACTCTGGTT-novel-189	B	-46.30	0
Ppi-sRNA-novel5 GGCCGACAGAGATGACACTGT-novel-85	C1	-50.60	2
	C2	-52.60	1
	C3	-50.50	2
	C4	-53.70	2
	C5	-47.30	3
Ppi-sRNA-novel7 TTCGAGAAGAGATGACACGATC-novel-136	C6	-48.40	1
	D1	-41.40	4
Ppi-sRNA-novel8 ATCCTTCTGAAAGCTTGCCG-novel-115	D2	-39.80	5
	E	-60.70	2

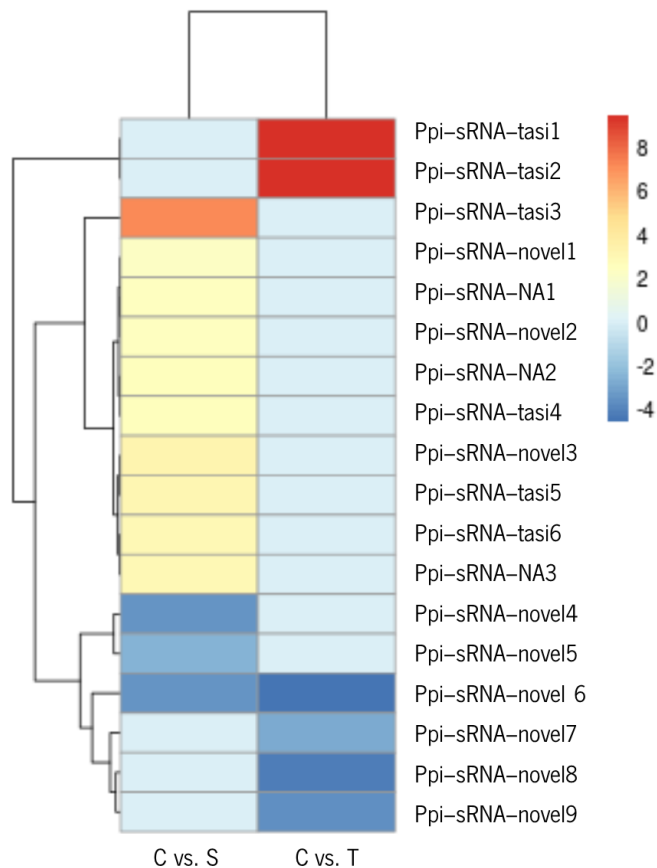


Figure 9 – Heatmap of the eighteen differentially expressed reads classified by the miRPursuit workflow. The heatmap was constructed with the log fold change of the reads. Values have been centered and scaled based on the column values. “Ppi-sRNA-classN” nomenclature: *Ppi* identifies the species; small RNA is abbreviated as *sRNA*; the *class* is either “novel” for novel miRNAs, “tasi” for ta-siRNAs or “NA” (not attributed) for small RNAs of unknown class; and *N* is a number to distinguish small RNAs from the same class.

3.3.2. Target prediction for small RNAs differentially expressed between susceptible and tolerant plants

The prediction of plant small RNA targets using computational techniques is relatively straightforward since siRNAs have perfectly complementary RNA targets, and most miRNA have near-perfect complementarity with their targets, as was early on observed for evolutionarily conserved miRNAs from *Arabidopsis* (Rhoades *et al.*, 2002). Target recognition is determined by the 5' end domain of the small RNA, and the cleavage occurs at the 10th or 11th nucleotide of complementarity (Carthew and Sontheimer, 2009; Bartel, 2009; Pareek *et al.*, 2015). If a decapped fragment transcript (mRNA) is the consequence of small RNA-mediated cleavage, then its 5' end must contain the first ten nucleotides of small RNA complementarity. Therefore, mapping the 5' ends of uncapped mRNAs to the transcriptome and extending 13 nucleotides upstream captures the region of potential complementarity; then, mapping these extended sequences to a

set of small RNA query sequences allows *in silico* identification of cleaved targets (Addo-Quaye *et al.*, 2009).

CleaveLand4 is a pipeline that uses degradome data to find cleaved small RNA targets. A total of 4,515,178 and 5,435,106 raw reads were obtained, respectively, for susceptible and tolerant degradomes in RNA-seq experiments. To examine the quality of the RNA-seq results, raw reads were first blasted against NCBI nt database and the best hits (e-value below $1e-5$) were extracted (Table IV). A high number of fragments has been identified as *Brugia timori*, a nematode (20,20%), in the tolerant degradome, and *Branchiostoma floridae*, a lancet (3,01%), in the susceptible degradome. Raw degradome reads inputted into CleaveLand4 were first filtered by their match to the structural RNAs, and then the filtered reads were mapped to the SustainPine v3.0 transcriptome. Alignment parameters allow zero or one mismatch and are only allowed to the forward strand of the transcriptome. Raw read counts are scaled to reads per million to account for varying dataset sizes. The alignments were analyzed to quantify the density of 5' ends of degradome sequences at each nt of the transcriptome, and a degradome density file was created. This file contains the position of the transcript, the number of 5' ends at that position, and the degradome peak "category". Categories (from 0 to 4) are based on the abundance of 5' ends matching a position (Table V). Category 0 corresponds to the highest confidence of the target prediction.

Table IV – BLAST hit results for the top species identified in degradome sequenced fragments from susceptible and tolerant samples. Percentage of a hit for each species was done based on the number of hits produced.

Susceptible degradome			Tolerant degradome		
<i>Picea glauca</i>	43,914	26.44%	<i>Brugia timori</i> (nematode)	63,286	20.20%
<i>Pinus pinaster</i>	21,193	12.76%	<i>Picea glauca</i>	55,341	17.67%
<i>Pinus taeda</i>	20,211	12.17%	<i>Uncultured eukaryote</i>	27,409	8.75%
<i>Picea sitchensis</i>	13,501	8.13%	<i>Pinus taeda</i>	26,049	8.32%
<i>Branchiostoma floridae</i>	5,002	3.01%	<i>Pinus pinaster</i>	21,005	6.70%
<i>Uncultured eukaryote</i>	4,471	2.69%	<i>Picea sitchensis</i>	16,248	5.19%
<i>Pinus radiata</i>	3,920	2.36%	–	8,934	2.85%
<i>Brugia timori</i>	3,167	1.91%	<i>Pinus heldreichii</i>	4,876	1.56%
<i>Pinus massoniana</i>	3,051	1.84%	<i>Pinus massoniana</i>	4,243	1.35%
<i>Pinus tabuliformis</i>	2,883	1.74%	<i>Branchiostoma floridae</i>	4,225	1.35%
<i>Homo sapiens</i>	1,741	1.05%	<i>Pinus radiata</i>	4,000	1.28%
Total	166,107		<i>Pinus tabuliformis</i>	3,293	1.05%
			Total	313,276	

In this study, the aim was to predict targets for the eighteen annotated differentially expressed small RNA reads previously identified. CleaveLand took these small RNA sequences and mapped them to the SustainPine v3.0 transcriptome to identify potential cleavage sites. Potential target sites were ranked based on their “Allen score” that is based on position-specific penalties, i.e. mismatched query bases or target-bulged bases are penalized 1, and G-U wobbles are penalized 0.5 (these penalties are double within positions 2-13 of the query) (Allen *et al.*, 2005). Only alignments to the reverse-complement strand of the transcriptome are considered. For each alignment, CleaveLand searches the degradome density file to see if there are any degradome reads at the predicted cleavage site. If there are, a p-value is calculated for the probability of observing one or more “hits” in the rank of potential target sites, given the chance of observing a degradome “peak” of a given category by random chance (total number of peaks of a given category divided by the effective transcriptome size) (Axtell, 2013). Targets predicted for the eighteen differentially expressed small RNAs are presented in Table F1 in last two columns (Appendix F). Annotation was made using the SustainPine v3.0 transcriptome. Since degradome sequencing data from a pool of the same RNA samples has been used as query, biologically important and/or strongly regulated targets should have been identified with the aid of this pipeline, and a direct link between a small RNA and it(s) target(s) can be established (Li *et al.*, 2014). Nevertheless, experimental validation of the interaction between each predicted small RNA-target pair is recommended.

From the list of predicted targets, nine transcripts were selected based on the differential expression of their corresponding small RNAs between susceptible and tolerant RNA-seq libraries and, also, based on the target transcript annotation (Fig. 9; Appendix F – selected transcripts are highlighted in bold). The selected target transcripts include Tetraspanin family protein, Class IV chitinase, Unassigned protein, DRE-binding protein DREB1, Translation initiation factor eIF-2 gamma subunit, 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase, Phosphoglycerate kinase, DNMT1-RFD multi-domain protein, and Gag-spuma domain containing protein (Table V). Note that some of these transcripts are targets of more than one small RNA. Primers were designed to amplify the region downstream the small RNA alignment region (downstream the TStop position, Table V).

Table V - Small RNA targets selected for gene expression analysis.

Small RNA	Target transcript(s)	Deg	Cat	Frag. Count	TStart	TStop	TSlice	Allen score	p-value	Annotation
Ppi-sRNA-tasi1	sp_v3.0_unigene 20381	S	4	1	63	84	75	4.5	0.047	Tetraspanin family protein
	sp_v3.0_unigene 31238	S	4	1	709	728	719	6	0.060	Class IV chitinase
Ppi-sRNA-novel7	sp_v3.0_unigene 964	S	4	1	1137	1158	1148	9.5	0.797	DRE-binding protein
	sp_v3.0_unigene 31238	S	2	2	761	781	772	7.5	0.069	Class IV chitinase
	sp_v3.0_unigene 5990	S	3	3	1361	1381	1372	6	0.003	Gag_spuma domain containing protein
T		3	6	1361	1381	1372	6	0.004		
Ppi-sRNA-novel8	sp_v3.0_unigene 33964	T	4	1	152	173	164	3	0.014	Unassigned protein
Ppi-sRNA-novel9	sp_v3.0_unigene 33964	T	4	1	151	173	164	5.5	0.014	Unassigned protein
Ppi-sRNA-novel1	sp_v3.0_unigene 22053	S	4	1	1318	1342	1332	7	0.212	translation initiation factor eIF-2 gamma subunit
Ppi-sRNA-novel3	sp_v3.0_unigene 126719	S	4	1	354	380	371	9.5	0.580	5-methyltetrahydropteroyl-triglutamate-homocysteine methyltransferase
Ppi-sRNA-tasi6	sp_v3.0_unigene 93887	S	4	1	1704	1724	1715	8	0.390	Phosphoglycerate kinase
	sp_v3.0_unigene 28175	T	4	1	3590	3610	3601	9.5	0.455	DNMT1-RFD multi-domain protein

Notes: *Deg* – Degradome; *S* – susceptible; *T* – tolerant; *Cat* – Degradome Category [Category 4: Just one read at that position; Category 3: >1 read, but below or equal to the mean level of coverage (average of all positions that have at least one read); Category 2: >1 read, above the mean level of coverage, but not the maximum; Category 1: >1 read, equal to the maximum on the transcript, when there is >1 position at maximum value; Category 0: >1 read, equal to the maximum on the transcript, when there is just 1 position at the maximum value]; *Frag. Count* – Fragment counts; *TStart* – Position of the nucleotide within the transcript sequence where the alignment starts; *TStop* – Position of the nucleotide within the transcript sequence where the alignment stops; *TSlice* – Position on the alignment opposite to position 10 of the query; *Allen score* – score calculated based on position-specific penalties (mismatched query bases or target-bulged bases are penalized 1, and G-U wobbles are penalized 0.5 – these penalties are double within positions 2-13 of the query).

3.4. Gene expression analyses of transcripts putatively involved in the differential behavior of *Pinus pinaster* plants towards the PWN infection

Real-time quantitative PCR (qPCR) is currently considered to be the most accurate and most reliable technique to quantify gene expression, for its high sensitivity, real-time detection of reaction progress, the speed of analysis and precise measurement of the examined material in the sample (Gachon *et al.*, 2004). Moreover, expression level for some genes is often so small that qPCR becomes the only available technique that can detect such a small number of mRNA copies. As qPCR is about to reach its maximum analytical potential it is required to introduce appropriate normalization methods in order to obtain reliable results, especially when the biological significance

of subtle gene expression differences is studied. It is relentlessly stressed that many qPCR experiments lack authors' critical evaluation, are wrongly designed and difficult to repeat due to insufficient data quality (Bustin *et al.*, 2009). MIQE Guidelines, which goes for Minimum Information for Publication of Quantitative Real-Time PCR Experiments, tries to address these questions with recommendations that help to ensure the integrity of scientific reports, consistency between experiments from different laboratories and increase transparency (Bustin *et al.*, 2009; Derveaux *et al.*, 2010). Those guidelines were taken into account in this work.

3.4.1. qPCR experimental setup: determination of amplification efficiencies, and validation of reference genes for expression normalization

3.4.1.1. qPCR amplification efficiencies of candidate reference genes

Experimental assays were performed in a set of nine *P. pinaster* samples (n=9) collected seventy-hours after inoculation from plants that later on exhibited differential behaviors towards the PWN inoculation. Three control (water-inoculated) plants, three susceptible plants, and three apparently tolerant plants were selected (note that these plants belong to the set of plants that were sent for small RNA and degradome sequencing). qPCR amplification efficiencies (E) were determined from the slopes of linear regressions fitted to log-transformed data of serially diluted template input (log copy number) against their Cp values as follows: $E = 10^{-1/\text{slope}}$ (Fig. 10). In theory, qPCR amplification efficiencies have a maximum of 2.00, or 100% if E is calculated as $E = (10^{-1/\text{slope}} - 1) \times 100$. This indicates that the amount of product produced in each cycle of qPCR amplification double with each cycle. However, this method of calculation can sometimes give rise to efficiencies higher than the theoretical maximum ($E > 2.00$), which might indicate an overestimation of the "true" amplification efficiency. Nevertheless, efficiencies higher than 2.00 has been shown to yield highly reproducible and constant results for a given gene under particular experimental conditions (Pfaffl, 2004).

Determination of efficiency was first evaluated for candidate reference genes in a pool of all samples obtained from pine needles (Fig. 10), to accumulate all possible sources of variation on kinetic PCR efficiencies. The determination of amplification efficiency was not possible for *EF1- α* since some of the dilution points had Cp values above 35, even with high input concentrations of template. The stability of the candidate reference genes was then evaluated in the individual needle samples (n=9) as a mean to validate and select the optimal number of reference genes suitable

for normalization of the qPCR experimental assays performed in this work. The amplification efficiencies obtained for the candidate reference genes ranged from 79 to 86% in the pool of nine needle samples, with linear correlation coefficients (R^2) from 0.8953 to 0.9962 (Fig. 10).

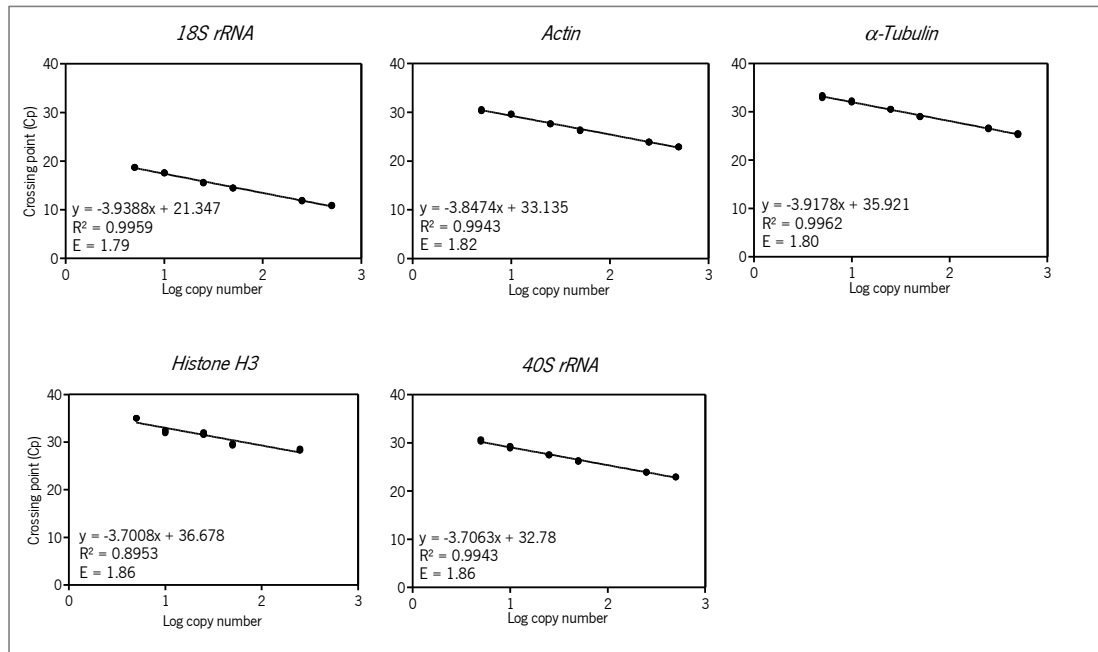


Figure 10 – Determination of qPCR efficiencies for candidate reference genes. Crossing point (Cp) values were plotted against log copy number input (serial dilutions of cDNA from a pool of RNA needle samples). The corresponding efficiencies were calculated according to the equation: $E = 10^{-1/\text{slope}}$. Cp values for each dilution point represent the mean of three technical replicates.

3.4.1.2. Validation of reference genes for expression normalization

Any measured variation in gene expression is influenced by two sources: on one hand, there is the true biological variation explaining the phenomenon under investigation; on the other hand, there is the non-specific variation that comes from several confounding factors, such as template input quantity and quality, yields of the extraction procedure and the enzymatic reactions (reverse transcription and PCR amplification). Even with the careful standardization of each step along the entire workflow of PCR-based gene expression analysis, the technique poses problems at various stages of sample preparation and processing that cannot be controlled by the operator (Vandesompele *et al.*, 2009). The most effective method to remove experimentally induced variation from the true biological variation is the use of reference genes (Huggett *et al.*, 2005). A reference gene must have a stable expression, and in turn demonstrate the variability resulting from the experimental imperfections, in such way that any variation in the object of research will relate to the same extent in the reference gene that thus serves as an internal control of the

experiment (Kozera and Rapacz, 2013). Housekeeping genes, which are, by definition, stable and nonregulated since they are involved in essential processes for the survival and maintenance of cells, seem to be ideal reference genes. However, there are several studies describing considerable variations in expression of housekeeping genes in some experimental conditions (Spanakis, 1993; Warrington *et al.*, 2000; Thellin *et al.*, 1999; Suzuki *et al.*, 2000; Bustin, 2000). So, an especial concern should be taken in the selection of reference genes, that must be imperatively validated for each particular experimental conditions. This necessity has been and continue to be ignored in many papers, and often a single, non-validated gene is selected based on previous publications without validation of its stability (Kozera and Rapacz, 2013). As an example, from July to December 2007 in three leading journals in plant biology research among a total of 188 qPCR analyses, only 3.2% were conducted with validated reference genes (Gutierrez *et al.*, 2008). Furthermore, it has been shown that the conventional use of a single gene may lead to relatively large errors (Vandesompele *et al.*, 2002), and the use of multiple reference genes is currently considered as an essential approach for reliable normalization of qPCR data (Vandesompele *et al.*, 2009). Since it is impossible to screen all genes in the genome, validation can only be carried out for some reference genes, and previous studies are the perfect starting point for the selection of candidates. The suitability of some candidate reference genes for normalization has been previously evaluated in *P. pinaster* plants by Vega-Bartol *et al.* (2013) and, more recently, Granados *et al.* (2016).

The basic assumption of reference gene's permanent and unchanging expression level in each of the samples tested, despite the impact of experimental factors, is in practice unachievable. There is always some variance in Cp values, and the strategy is the selection of genes with the most stable expression or showing the least deviation from the mean (Vandesompele *et al.*, 2002). Several statistical algorithms have been developed to evaluate the stability of candidate reference genes. In this work, GeNorm, BestKeeper, and NormFinder were used to evaluate the stability of candidate reference genes – *18S rRNA*, *Actin*, α -*Tubulin*, *Histone H3* and *40S rRNA* – selected from previous qPCR studies in *P. pinaster*. As in the case of reference genes there is no reference point, some authors suggest that a stability index should be calculated by comparing the variation of pairs of reference genes (pairwise variation) with respect to the other candidate reference genes (Vandesompele *et al.*, 2002; Pfaffl *et al.*, 2004).

GeNorm relies on the fact that for two genes to be considered ideal internal controls their expression ratio should be similar in all samples and calculates the pairwise variation, V_j , of each gene with respect to the other genes as the standard deviation of the log-transformed expression

ratios. Then, it determines an expression stability value (M) for each candidate reference gene, defined as the average pairwise variation of that gene with the other genes. Genes with the lowest M values have the most stable expression. The GeNorm ranking of the selected candidate reference genes can be seen in Fig. 11A. *Actin* and *40S rRNA* were considered to be the most stably expressed genes of the set of candidates, followed by *Histone H3*, α -*Tubulin*, and the least stable was *18S rRNA* (Fig. 11A). Then, GeNorm determines the optimal number of reference genes for normalization. It took the two most stable genes ($n=2$), *Actin* and *40S rRNA*, and calculated the variation resulting from the inclusion of the third most stable gene ($n+1=3$), *Histone H3*, in the ranking ($V_{2/3}$), the variation $V_{3/4}$ from adding to the three most stable genes ($n=3$) the fourth best-ranked gene ($n+1=4$), α -*Tubulin*, and finally the variation $V_{4/5}$ that comes from using all five candidate genes in analysis ($n+1=5$). A V_n/V_{n+1} cut-off value of 0.15 is recommended, below which the inclusion of the $(n+1)^{\text{th}}$ additional gene is not required (Vandesompele *et al.*, 2002). According to the results of V_n/V_{n+1} obtained for the candidate reference genes (Fig. 11B), the use of all five genes for normalization of qPCR data is not required, as the $V_{4/5}$ value is above the defined threshold 0.15, but a reliable normalization should be obtained either using the three or four most stable candidate reference genes ($V_{2/3}$ and $V_{3/4}$ are below the cut-off value). Therefore, the recommended combination of genes for normalization is “*Actin* + *40S rRNA* + *Histone H3* + α -*Tubulin*”, or “*Actin* + *40S rRNA* + *Histone H3*”.

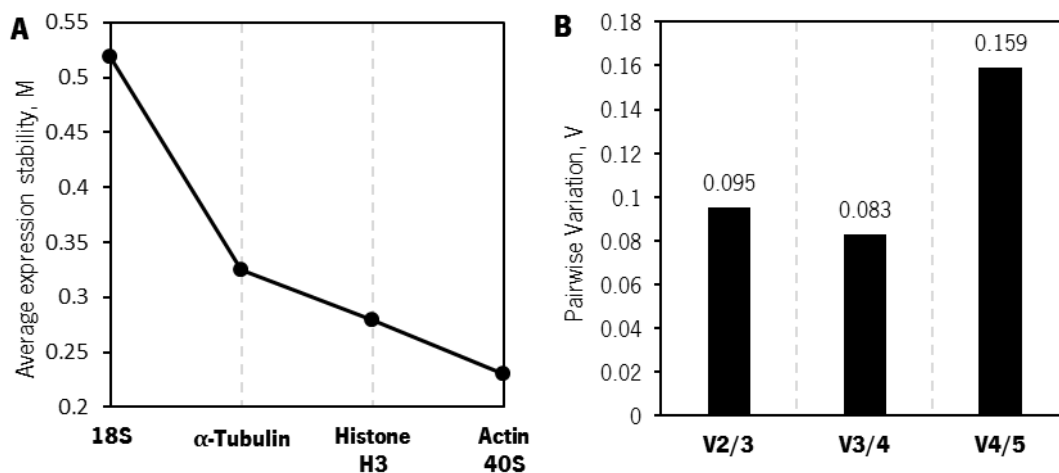


Figure 11 – GeNorm average expression stability (A) and pairwise variation (B) for the candidate reference genes. (A) Genes are ranked according to an expression stability value (M) as the average pairwise variation (standard deviation of the log-transformed expression ratios) of each gene with respect to the other candidate genes; (B) The optimal number of reference genes is determined by comparing the variation V_n/V_{n+1} of adding to the two most stable genes the third gene on the ranking ($V_{2/3}$); to the three most stable genes the fourth gene on the ranking ($V_{3/4}$); or to the four most stable genes the last, fifth gene on the ranking ($V_{4/5}$). The recommended V_n/V_{n+1} cut-off value is 0.15.

BestKeeper is a tool that, like GeNorm, performs a pairwise comparison of candidate reference genes to determine inter-gene variation, and additionally examine intra-gene sample integrity and expression stability. Input data are raw Cp values and the amplification efficiency of each candidate reference gene. Descriptive statistics based on the Cp values are computed for each gene in Table VI. Estimation of inter-gene variation of candidate reference genes can be achieved simply by observing their SD and CV values (highlighted in Table VI), which also denote the intra-gene variations. Candidate genes can be ranked from the most stably expressed to the least stable gene (lowest to highest variation). Genes considered to be stably expressed, are combined into a *BestKeeper index* calculated by the geometric mean of the Cp values obtained for those stable genes. In this study, all candidate reference genes (n=5) were combined into the index, to which descriptive statistics are also computed in Table VI. To examine inter-gene variations, several pairwise correlation analyses were performed and characterized by the Pearson correlation coefficient (r) and the p-value (Table VII). Highly correlated genes are combined into a *BestKeeper index*, and a correlation between the index and each gene is calculated by the Pearson correlation coefficient (r) and the p-value (Table VII). The output is the identification of the two best reference genes by the performed statistical analysis and pairwise correlations. From the set of candidate reference genes, the two that came to show the highest correlation (highest “r”) were *Actin* and *40S rRNA*, with a “r” of 0.851 and p-value of 0.004 (highlighted in Table VII).

Table VI – BestKeeper descriptive statistics for the five candidate reference genes based on their crossing point (CP) values. The BestKeeper index, calculated as the geometric mean of CP values of the stably expressed genes, is computed with the same statistical parameters, for the five genes (n=5).

	<i>18S rRNA</i>	<i>Actin</i>	<i>α-Tubulin</i>	<i>Histone H3</i>	<i>40S rRNA</i>	BestKeeper index (n=5)
N	9	9	9	9	9	9
GM [CP]	24.94	22.96	26.21	29.28	24.13	25.42
AM [CP]	24.96	22.96	26.22	29.28	24.13	25.42
Min [CP]	23.19	22.47	25.54	28.84	23.65	24.86
Max [CP]	26.41	23.66	26.94	29.58	24.95	26.23
SD [± CP]	0.80	0.26	0.45	0.22	0.39	0.31
CV [% CP]	3.22	1.11	1.73	0.75	1.62	1.23

Notes: N – number of samples; GM [CP] – geometric mean of CP; AM [CP] – arithmetic mean of CP; Min [CP] and Max [CP] – extreme values of CP; SD [± CP] – standard deviation of the CP; CV [% CP] – coefficient of variance expressed as a percentage of the CP level.

Table VII – BestKeeper pairwise correlation analysis of candidate reference genes. To estimate inter-gene variations, pairwise correlation analyses are performed, and the Pearson correlation coefficient (r) of each correlation and the probability (p) value are calculated.

vs.	Pearson correlation coefficient (r)				
	<i>18S rRNA</i>	<i>Actin</i>	<i>α-Tubulin</i>	<i>Histone H3</i>	<i>40S rRNA</i>
<i>Actin</i>	0.576	-	-	-	-
p-value	0.105	-	-	-	-
<i>α-Tubulin</i>	-0.025	0.702	-	-	-
p-value	0.946	0.035	-	-	-
<i>Histone H3</i>	0.183	0.604	0.618	-	-
p-value	0.639	0.084	0.076	-	-
<i>40S rRNA</i>	0.675	0.851	0.605	0.587	-
p-value	0.046	0.004	0.084	0.096	-
BestKeeper index	0.786	0.917	0.575	0.604	0.950
p-value	0.012	0.001	0.105	0.084	0.001

NormFinder algorithm, available as a Microsoft Excel add-in, calculates the expression stability value (SV) for each candidate reference gene, by the estimation of intra- and intergroup variations. Thus, SV combines both sources of variation and represents a useful measure of the systematic error that will be introduced when using a particular gene. Lower SV indicates higher expression stability (Andersen *et al.*, 2004). The output is the identification of the best two reference genes from the set, that combine minimal inter- and intragroup expression variation, and the respective SV for the two-gene combination. Like the previous tool, NormFinder also identified *Actin* and *40S rRNA* as the two best-suited reference genes for normalization of qPCR experiments, with a stability value (SV) of 0.101 (Fig. 12).

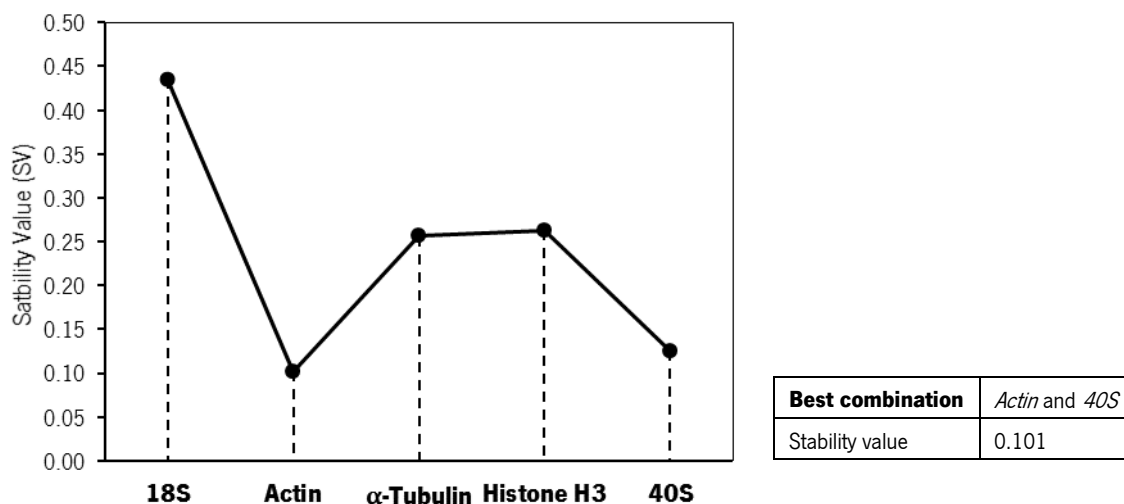


Figure 12 –Stability Values for the five candidate reference genes determined by NormFinder. Based on the Stability Value (SV) of each candidate reference gene, Normfinder combines the two most stable genes (“Best combination”), and calculates an SV for the two genes combined.

Reviewing the results obtained in each of the three tools, *Actin*, *40S rRNA* and *Histone H3* reference genes were selected for normalization of the qPCR experiments performed in this work. These three genes were concordantly classified as the three most stably expressed genes from the set of candidate reference genes: these three genes had the lowest M values (GeNorm), the lowest SD and CV values (BestKeeper), and low SV values (NormFinder). In all three tools, α -*Tubulin* was also considered to have a stable expression, and according to the pairwise variation calculated in GeNorm it could also be selected as a proper reference gene ($V_{3/4} < 0.15$); however, as stressed by Vandesompele *et al.* (2002), it is unpractical to use more genes than the necessary if all genes are relatively stable and a further gene has no significant contribution (negligible reduction in the average of the gene variance estimates), and the use of the three most stable reference genes is recommended by the authors (as does Bustin *et al.*, 2010 and Derveaux *et al.*, 2010).

3.4.2. Relative expression level quantification of candidate genes and selected small RNA targets

Several transcripts were selected to evaluate their involvement in the molecular responses of *P. pinaster* plants with differential behavior towards the PWN infection. Eight candidate genes previously shown to be involved in the differential behavior (susceptibility or tolerance) of *Pinus* species against PWN infection (Santos *et al.*, 2012) and nine transcripts that are putative targets of small RNAs differentially expressed between susceptible and tolerant plants were selected to evaluate their expression levels in RNA samples from susceptible and tolerant plants using qPCR.

To determine the specificity of primer pairs used in this study, melting curve analysis and agarose gel electrophoresis were performed after the qPCR experiments. A single peak in the obtained melting curves confirmed the specificity of the amplicon, and no amplification was detected in the negative controls for all the tested transcripts (Fig. H1, Appendix H). In addition, a single band with the expected size was detected in a single PCR product (Fig. H2, Appendix H). LEA protein was “discarded” from further analyses since the achievement of a single amplification product was not possible, probably because LEA proteins are a large and highly diverse gene family (Gao and Lan, 2016) and in the article by Santos *et al.* (2012), from where this candidate transcript was selected, there is no indication of what specific LEA protein was studied. *TerpMet* was also “discarded” from further analyses since it was not possible to eliminate amplification due to primer-dimer formation from the reaction, even when different concentrations of primers were tested. The primers for this transcript were retrieved from Santos *et al.* (2012), and it was not possible to

design new primers for it since the authors do not provide any information about the target sequence.

The PCR efficiency (E) was calculated as described in 3.4.1.1 using a dilution series of cDNA from the qPCR calibrator sample. The determination of amplification efficiency was not possible for SNARE (candidate gene), *PGK* and *UA* (putative small RNA targets) since some of the dilution points had Cp values above 35, even with high input concentrations of template. The amplification efficiencies obtained for the reference genes (Fig. 13A), candidate genes (Fig. 13B) and small RNA targets (Fig. 13C) ranged from 88 to 112%, with linear correlation coefficients (R^2) from 0.9811 to 0.9991 (Fig. 13).

Relative expression levels for the small RNA target transcripts, Tetraspanin family protein, Class IV chitinase, DRE-binding protein1, Translation initiation factor eIF-2 gamma subunit, 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase, DNMT1-RFD multi-domain protein, and Gag-spuma domain containing protein, were determined in stem samples (from three control plants, three susceptible plants, and three apparently tolerant plants) (Fig. 14). For the candidate genes, *DAHP synthetase*, *FMN reductase*, *RicinB*, *ABA/WDS* and *PARI*, relative expression levels were determined in the stem samples and also in needle samples (Fig. 15).

The results obtained for the predicted small RNA target transcripts show a differential transcript abundance between susceptible and tolerant plants (Fig. 14), which is in agreement with the *in silico* prediction of differential expression of their respective small RNA(s) between contrasted samples. Targets predicted in this study are the result of small RNA-directed cleavage, and since the analysis was performed using degradome data from the same samples where the small RNAs were identified, a direct link between the small RNA and its target can be established.

Tetraspanin family protein is a putative target of “Ppi-sRNA-tasi1”. This ta-siRNA was found to be more expressed in control RNA-seq libraries, followed by the susceptible libraries, and it has not been identified in the tolerant library (Appendix F). The lower expression level of *Tetraspanin* gene in susceptible plants, when compared to tolerant plants (Fig. 14) should be the result of the higher abundance of “Ppi-sRNA-tasi1” in susceptible samples when compared to the tolerant sample. Tetraspanins are integral membrane proteins that are known in animals to function in cell adhesion, immunity reaction, and pathogen recognition. In plants, there is only functional information for one of the seventeen members of the *Arabidopsis tetraspanin* gene family. Recently,

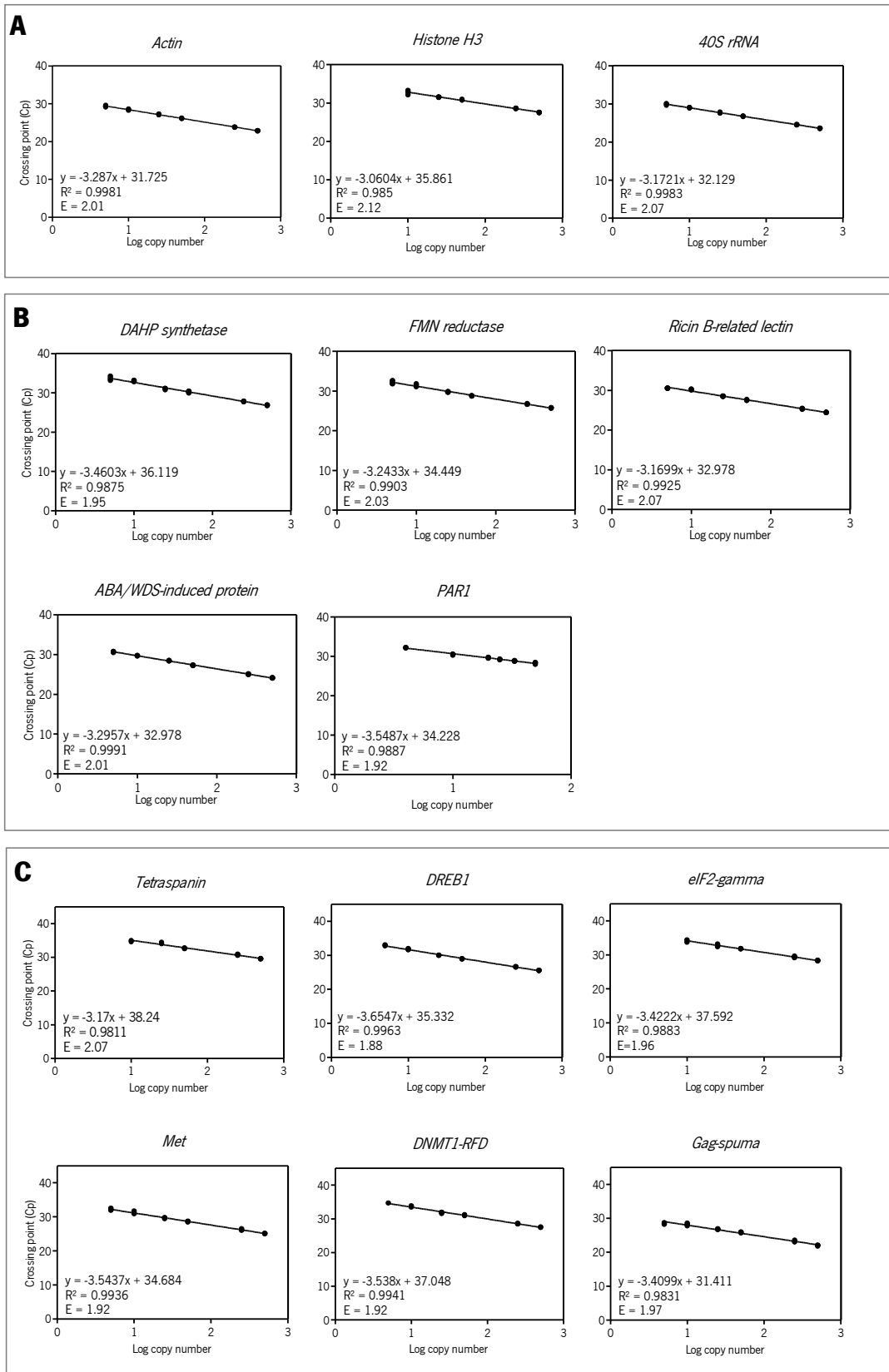


Figure 13 – Determination of qPCR efficiencies for selected reference genes (A), candidate gene (B), and small RNA targets (C). Crossing point (Cp) values were plotted against log copy number input (serial dilutions of cDNA from the qPCR calibrator, needle sample). The corresponding efficiencies were calculated according to the equation: $E = 10^{-1/\text{slope}}$. Three technical replicates of each dilution point were performed. Cp values for each dilution point represent the mean of three technical replicates.

Wang et al. (2015) performed an analysis of cis-regulatory elements and, together with transcription factor binding data, placed *Arabidopsis* tetraspanin genes in various molecular pathways. In particular, *Tetraspanin8* is enriched in cis-regulatory elements related to defense responses and is up-regulated upon pathogen and elicitor treatment. In *P. pinaster*, the tetraspanin family gene transcript has been shown to be up-regulated in response to PWN infection in tolerant plants, indicating that it might be involved in tree's defense mechanisms against PWD.

Translation initiation factor eIF-2 gamma subunit (*eIF2-gamma*) is a putative target of "Ppi-sRNA-novel1" (novel miRNA) which was predicted to be more expressed in the tolerant RNA-seq library, followed by the control libraries and it is less abundant in susceptible libraries (Appendix F). The relative expression levels obtained for *eIF2-gamma* (Fig. 14) were in accordance with the expression profile of "Ppi-sRNA-novel1" found in the sequencing libraries, as *eIF2-gamma* is less expressed in tolerant plants. Similar results were obtained for translation initiator factor *eIF-4A* that is upregulated in the susceptible species *P. massioniana* 24h after PWN infection (Xu et al., 2013), and for the translation initiator factor *SUI1* that is upregulated in non-resistant *P. thunbergii* trees 3 days after inoculation (Nose and Shiraishi, 2011). The eukaryotic translation initiation factor 2 (eIF2) is a GTP-binding protein involved in the initiation of translation by delivering the initiator methionyl-tRNA to the ribosome, which is believed to be the rate-limiting step in mRNA translation. eIF2 has three subunits; the large gamma subunit is the structural core for the trimer assembly and is thought to play an important role in binding both the initiator tRNA and the GTP ligands by eIF2 (Gaspar et al., 1994; Erickson and Hannig, 1996; Kimball, 1999). The lower expression level of *eIF2-gamma* in tolerant plants (Fig. 14) might be indicative of translation repression mechanisms triggered by miRNAs, which possibly participate in the defense responses of trees against PWD.

DNMT1-RFD multi-domain protein is a putative target of "Ppi-sRNA-tasi6". This ta-siRNA was predicted to be highly expressed in the tolerant RNA-seq library, followed by control libraries, and it is less abundant in susceptible libraries (Appendix F). The relative expression levels obtained for *DNMT1-RFD* (Fig. 14) were in accordance with the expression profile of "Ppi-sRNA-tasi6". DNMT1-RFD stands for DNA Methyltransferase 1 (DNMT1)-replication foci domain (RFD). RFDs are the non-catalytic N-terminal domains of DNMT1 that target the protein towards the replication foci, allowing DNMT1 to methylate the correct residues. DNMT1 main function is to maintain the methylation patterns following DNA replication, but it can also repress transcription through its RFD domain. This domain has been shown to directly interact with Histone Deacetylases (HDACs) and transcriptional co-repressor proteins like DMAP1 (DNMT1 associated protein 1). HDAC2 is found

to co-localize with DNMT1 and DMAP1 only during late S phase when replication of heterochromatic regions occurs. Thus, DNMT1 not only maintains DNA methylation but may also alter the expression of genes by transgenerational or heritable modifications in chromatin (Fuks *et al.*, 2000; Rountree *et al.*, 2000). Low levels of *DNMT1-RFD* in tolerant plants (Fig. 14) may be indicative of hypomethylation of genes potentially involved in response to PWN infection. Hypomethylation and demethylation of DNA during pathogen infection have been shown to influence the expression of defense-related genes (Wada *et al.*, 2004; Pavet *et al.*, 2006; Akimoto *et al.*, 2007; Downen *et al.*, 2012; Yu *et al.*, 2013). The suggested DNA hypomethylation potentially affecting defense genes as a result of the lower expression of *DNMT1-RFD*, that is a putative target of a ta-siRNA, is a mechanism that probably belongs to the RNA-directed DNA methylation (RdDM) epigenetic pathway. Small RNA's epigenetic regulation through transcriptional gene silencing (RdDM) in response to biotic stress (and also abiotic stress, growth and developmental signal) may prime a defense mechanism that allows plants to defend their progeny against repetitive biotic stresses without stable inherent trait fixation. This process is not fully understood but it will certainly create opportunities to moderate plant's susceptibility to various diseases (Sahu *et al.*, 2013). This study provides the first evidence of the potential involvement of epigenetic mechanisms in the defense response against PWN infection.

DRE-binding protein1 (DREB1) is a putative target of "Ppi-sRNA-novel7". This miRNA was found to be more expressed in the tolerant RNA-seq library, followed by susceptible libraries, and it is less abundant in control libraries (Appendix F). The relative expression levels obtained for *DREB1* were in accordance with the expression profile of "Ppi-sRNA-novel7", except for the control samples where the expression levels were not higher than in susceptible samples, as expected (Fig. 14). DREB1 (dehydration-responsive element-binding protein 1) is typically involved in signal transduction pathways in plant response to abiotic stress, such as drought and cold, but recently has also been shown to have an upstream regulatory role in mediating signaling pathways for biotic stress responses (Zhou *et al.*, 2010; Charfeddine *et al.*, 2015). In *Solanum tuberosum*, *StDREB1* expression is upregulated by a fungal infection, accompanied by a significant improvement in tolerance to *Fusarium solani* *in vitro*. In this study, *P. pinaster* plants inoculated with PWN showed an upregulation of *DREB1* in susceptible plants with respect to the control, and a downregulation of *DREB1* in apparently tolerant plants (Fig. 14). These expression levels seem to be regulated by the miRNA "Ppi-sRNA-novel7". A transcript induced by dehydration was also found in *P. densiflora* (Shin *et al.*, 2009) and a drought stress responsive protein was found to be

downregulated in *P. massoniana* (Xu *et al.*, 2013) in response to PWN infection. As already stated in the Introduction section, dehydration is a defense reaction of plants to the infection by microorganisms and in the early phase of PWD development, when PWN are spreading in the tree through resin canals, the production of secondary metabolites disturbs and partially blocks sap ascent in the xylem, and water transport from the roots to the shoot becomes dysfunctional leading to the formation of dehydrated areas. Perhaps the higher *DREB1* transcript abundance in susceptible plants could be linked to the dehydration that may occur at higher levels in the susceptible plants than in the tolerant ones. *DREB1* potentially responds to the dehydration and participates in the response mechanisms against PWN infection.

5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase (termed herein as Met) is a putative target of “Ppi-sRNA-novel3”. This miRNA was more expressed in control RNA-seq libraries, followed by the tolerant library, and it is less abundant in susceptible libraries (Appendix F). The relative expression levels obtained for *Met* (Fig. 14) were in accordance with the expression profile of “Ppi-sRNA-novel3”, with a higher expression observed in susceptible plants. 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase, or cobalamin-independent L-methionine synthase, transfers the methyl group from 5-methyltetrahydropteroyltriglutamate to homocysteine (Whitfield *et al.*, 1970). This reaction is the last step in L-methionine biosynthesis, and it also serves to regenerate S-adenosyl methionine (SAM), an important cofactor required in most methylation reactions. SAM is a common methyl group donor for many methyltransferases in plants that are involved in both primary and secondary metabolisms in the biosynthesis of molecules involved in disease resistance, growth, and development. These molecules include lipids, alkaloids, the polyamines spermidine and spermine, flavonoids, lignin, as well as the gaseous plant hormone ethylene, among others (Roje, 2006). *P. pinaster* susceptible plants appear to have higher levels of *Met* than tolerant plants (Fig. 14), which might be due to the higher rate of synthesis of secondary metabolites in an attempt to respond to the PWN infection. Further evidence for this statement comes from previous studies in *Pinus* species inoculated with PWN. For instance, a methionine synthase was found to be upregulated in susceptible *P. massoniana* following PWN infection (Xu *et al.*, 2013). SAM synthase was identified in *P. densiflora* infected with PWN, and the authors suggest an involvement of ethylene signaling in the defense mechanisms against PWN (Shin *et al.*, 2009). SAM synthase was also identified in *P. pinea*, a species that seems to be tolerant to PWD (Santos and Vasconcelos, 2012). And finally, SAM decarboxylase, involved in the biosynthesis of spermidine and spermine, was identified in *Pinus thunbergii* inoculated with PWN,

and the authors referred that polyamines are produced in response to pathogen infection and proposed that spermine can induce the production of pathogenesis-related (PR) proteins that were upregulated in their study (Nose and Shiraishi, 2011).

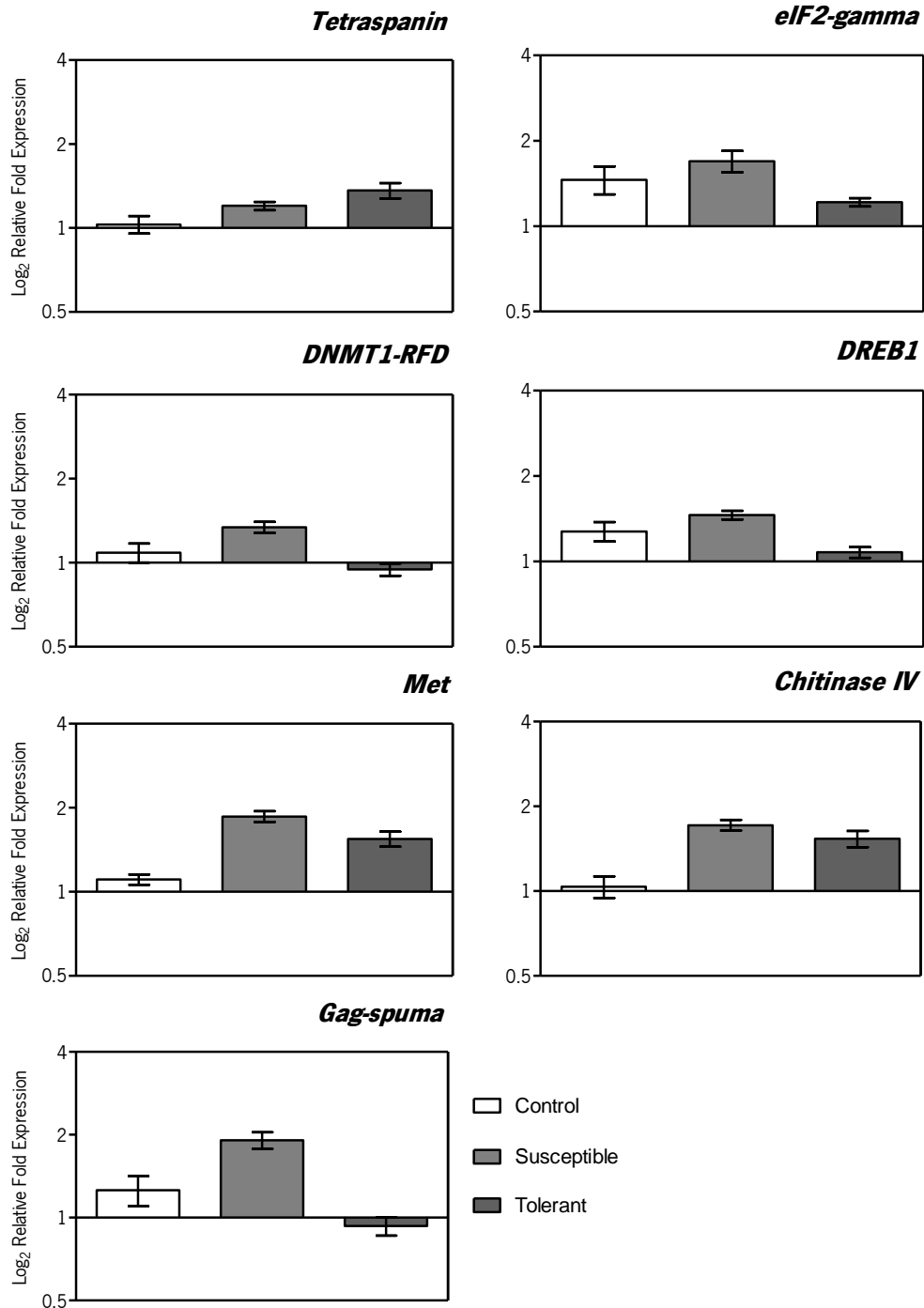


Figure 14 – Relative expression levels of predicted small RNA target genes in *Pinus pinaster* plants exhibiting different behaviors towards Pine Wood Nematode infection. Real-time quantitative PCR (qPCR) was performed with RNA from stems of three water-inoculated (control), three susceptible and three apparently tolerant two-year-old plants collected seventy-two hours after inoculation with *Bursaphelenchus xylophilus*. Relative expression levels are plotted as \log_2 fold changes of normalized relative expression level. Normalization was performed using *Actin*, *40S rRNA*, and *Histone H3* reference genes. Error bars represent SEM (n=3).

Class IV chitinase is a putative target of “Ppi-sRNA-novel7” as it is DREB1. The relative expression levels obtained for *Chitinase IV* (Fig. 14) were, like for *DREB1*, in accordance with the expression profiles of “Ppi-sRNA-novel7” (Appendix F). However, a higher difference in the expression levels of *Chitinase IV* between susceptible and tolerant plants would be expected, given the differences in the abundance of that miRNA between susceptible and tolerant samples. Chitinase IV is also a putative target of “Ppi-sRNA-tasi1” which is not present in the tolerant library but is present in the susceptible ones (Appendix F), which could reduce the expression level of *Chitinase IV* in susceptible plants making it more close to its expression level in tolerant plants (Fig. 14). Nevertheless, the highest expression is observed in susceptible plants. Plant chitinases are usually endo-chitinases capable of degrading chitin, a major constituent of the cell walls of many pathogenic fungi, and inhibit fungal growth (Ebrahim *et al.*, 2011). It can also indirectly promote the release of materials from the cell wall of both pathogen and host plant than can act as elicitors of the defense mechanism. Chitinases are PR proteins from class III (PR3) that are activated as a part of the defense mechanism against pathogens called hypersensitive reaction (Nose and Shiraishi, 2011), and are therefore abundant in many plant species following pathogen infection. *Chitinase IV* (PR3) has been shown to be upregulated in *P. massoniana* and *P. densiflora* (susceptible species) 24h after PWN infection, and was proposed to be involved in the cell wall degradation of fungi from the genus *Ceratocystis*, which are known to infect pine trees alongside the PWN (Wingfield, 1987; Shin *et al.*, 2009; Xu *et al.*, 2013). It has also been identified in resistant and susceptible *P. thunbergii* plants inoculated with PWN, showing a higher expression in susceptible plants (Hirao *et al.*, 2012; Nose and Shiraishi, 2011). In this study, *Chitinase IV* is upregulated in *P. pinaster* plants inoculated with PWN, and its expression level is higher in susceptible plants than in tolerant plants (Fig. 14), which is accordance with the results obtained with *P. thunbergii*. Thus, *Chitinase IV* is possibly involved in the defense mechanisms against PWN infection, as described for other *Pinus* species.

The transcript encoding the Gag-spuma domain containing protein (shortened Gag-spuma) is also a putative target of “Ppi-sRNA-novel7”. The relative expression levels obtained for *Gag-spuma* (Fig. 14) were also in accordance with the expression profiles of “Ppi-sRNA-novel7” (Appendix F). The association of this domain with plant disease defense mechanisms or other stresses has not yet been documented. Nevertheless, the current study seems to indicate a potential involvement of a Gag-spuma domain containing protein in the response against PWN infection.

In the previous studies in *Pinus* species mentioned above, high-throughput screening procedures have been used to identify genes involved in tree's response to PWN infection. From a population of susceptible trees, some candidate tolerant or resistant individuals may survive simply because they show a different response at the molecular level when compared to the susceptible individuals. Constitutive and induced defense response in the tolerant trees early after infection may be sufficient and rapid enough to prevent damage, leading to their survival. Santos *et al.* (2012) identified ESTs differentially expressed in *P. pinaster* (susceptible species) and *P. pinea* (tolerant species) inoculated with PWN, and found several candidate defense response-associated genes at the initial stage of the disease involved in the oxidative stress response, the production of lignin and ethylene, and the posttranscriptional regulation of genes. Here, expression analysis of some of those candidate genes (described below) was performed on *P. pinaster* plants derived from one mother tree that was selected as a candidate resistant tree from an area highly affected by PWD in a national breeding program initiated in 2009. The progeny originated from that tree was inoculated with PWN, and plants were classified as susceptible or tolerant based on the observation of symptom development. This is the first report on the use of gene expression analysis to study the molecular resistance mechanisms in *P. pinaster* involving candidate resistant individuals.

3-Deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) synthetase was found to be much more expressed in *P. pinea* than in *P. pinaster* (Santos *et al.*, 2012), and in this study *DAHP synthetase* expression level was higher in tolerant *P. pinaster* plants than in susceptible ones, with a similar expression profile in needle and stem samples (Fig. 15). *DAHP synthetase* is the enzyme that catalyzes the first step of shikimate pathway, which in plants leads to the production of aromatic amino acids and phenylpropanoids, such as lignin and flavonoid/stilbenoids (Herrmann, 1995). Several genes encoding enzymes responsible for the production of secondary metabolites from the phenylpropanoid pathway have been identified in *P. densiflora* (Shin *et al.*, 2009), *P. thunbergii* (Hirao *et al.*, 2012), and *P. massoniana* (Xu *et al.*, 2013) following PWN infection. However, the role of these genes in relation to the response against PWN infection remains unclear because secondary metabolism in conifers is complex. Lignification of tree's cell walls might serve as an obstacle for PWN migration and reproduction in the host tree, and this mechanism of initial damage retardation must be determinant for resistance (Xu *et al.*, 2013; Kusumoto *et al.*, 2014).

Flavin mononucleotide (FMN) reductase was also found to be more expressed in tolerant plants (*P. pinea*) than in susceptible (*P. pinaster*) plants (Santos *et al.*, 2012), which is also

observed in this study in stem samples (Fig. 15). In needle samples, there is no change in PWN-inoculated plants in terms of *FMN reductase* expression levels (Fig. 15). *FMN reductase* is an enzyme that participates in the oxidative stress response by maintaining intracellular redox balance. As already mentioned in the Introduction section, the main symptom of the disease, wilting of leaves, is caused by a decrease in water potential in PWN-infested stems that results from the decrease of water conductivity from roots to shoots (dehydration referred before). As the water potential decreases, trees suffer severe oxidative stress; therefore, the upregulation of oxidative stress-responsive genes, like *FMN reductase* (Fig. 15) and other genes described in previous studies (Shin *et al.*, 2009; Nose and Shiraishi, 2011; Santos *et al.*, 2012; Xu *et al.*, 2013), is expected and may contribute to a higher tolerance to PWD. The results obtained in needles samples (Fig. 15) might point out that this response is confined to the stem and around the area where the inoculation was performed, at least at early stages of disease development. It should be noted that samples were collected seventy-two hours after inoculation and stem samples contain the inoculation area.

As water stress is directly associated with PWD symptoms, a protein from a family induced by abscisic acid (ABA) and water deficit stress (WDS) was identified, and found to be upregulated in both tolerant (*P. pinea*) and susceptible (*P. pinaster*) plants, but at a slightly higher level in susceptible plants (Santos *et al.*, 2012). In this study, this *ABA/WDS-induced protein* was downregulated in both needle and stem samples, but more prominently in stems from tolerant plants (Fig. 15). In needle samples, there is a downregulation of *ABA/WDS-induced protein* in PWN-inoculated plants but there is no apparent difference between susceptible and tolerant *P. pinaster* plants (Fig. 15). The pronounced difference in the expression of this gene between susceptible and tolerant plants in stem samples might point out that the response is confined to the stem and around the area where the inoculation was performed, at least at early stages of disease development.

Ricin B-related lectin was found to be more expressed in tolerant plants (*P. pinea*) than in susceptible (*P. pinaster*) plants (Santos *et al.*, 2012). A mannose/glucose-specific lectin was also upregulated in *P. massoniana* 72h after PWN infection (Xu *et al.*, 2013). In this study, however, there is a downregulation of *Ricin B-related lectin* in needle and stem samples for both susceptible and tolerant *P. pinaster* plants, although more pronounced in the tolerant plants (Fig. 15). Plant lectins are known to be involved in the defense response against various plant pathogens, including nematodes, by recognition of carbohydrates structures in organisms (Vandenborre *et al.*, 2011).

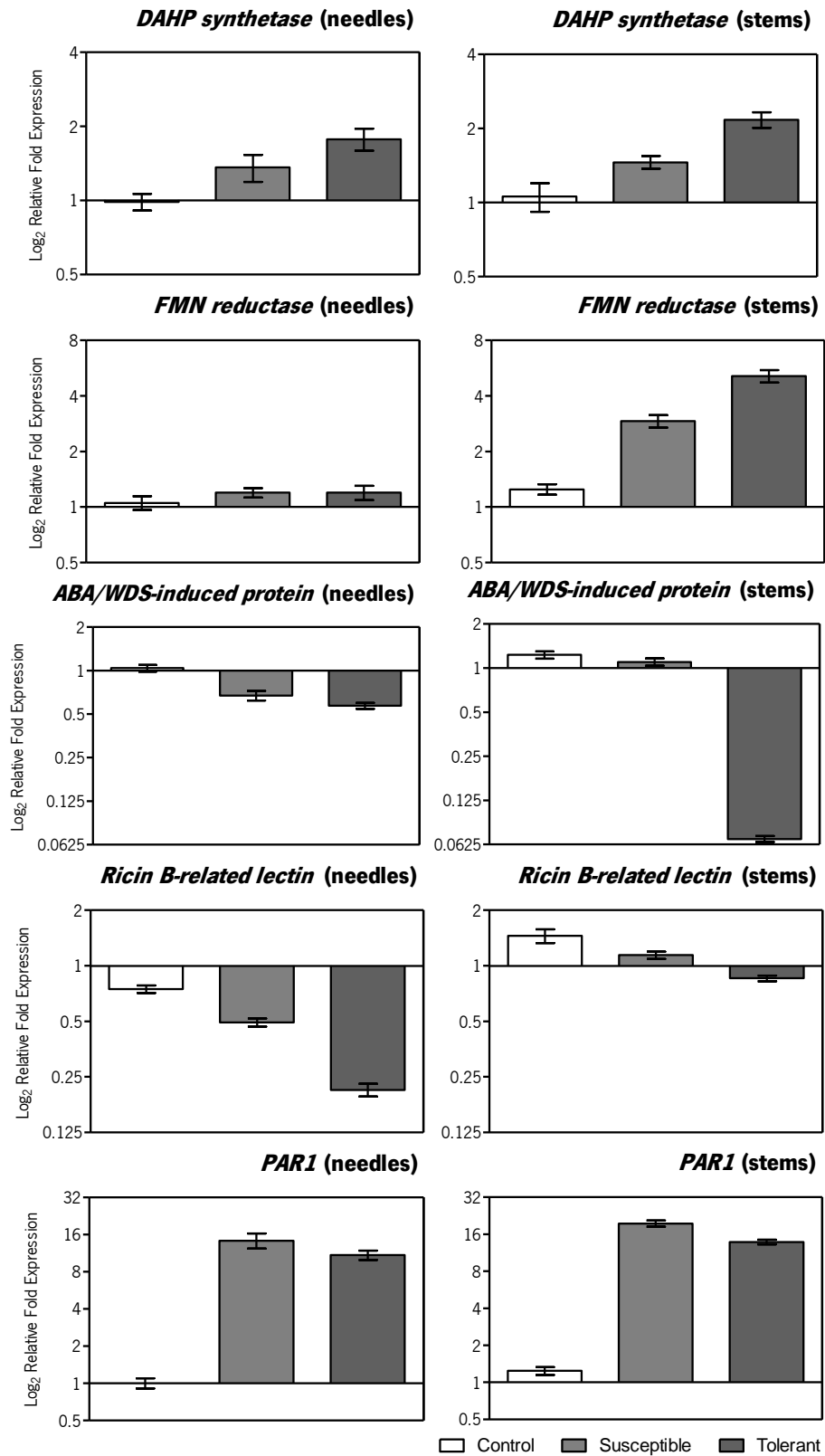


Figure 15 – Relative expression levels of candidate genes putatively involved in the molecular responses of *Pinus pinaster* plants exhibiting different behaviors towards Pine Wood Nematode infection. Real-time quantitative PCR (qPCR) was performed with needles and stems of three control, three susceptible and three apparently tolerant two-year-old plants collected seventy-two hours after inoculation with *Bursaphelenchus xylophilus*. Relative expression levels are plotted as log₂ fold changes of normalized relative expression level. Normalization was performed using *Actin*, *40S rRNA*, and *Histone H3*. Error bars represent SEM (n=3).

In this study, the expression levels of *Ricin B-related lectin* were altered following PWN infection, and a stronger response concerning the expression of this gene was seen in tolerant plants. Therefore, it may be involved in the defense response mechanisms against PWN.

Photoassimilate-responsive 1 (PAR1) was found to be more expressed in susceptible (*P. pinaster*) than in tolerant (*P. pinea*) plants (Santos *et al.*, 2012). In this study, *PAR1* was significantly upregulated in both susceptible and tolerant *P. pinaster* plants, and in similar levels in both needle and stem samples (Fig. 15). *PAR1* was more expressed in susceptible *P. pinaster* plants than in tolerant ones (Fig. 15), which seems to be in accordance with the results from Santos *et al.* (2012). *PAR1* is a defense-related gene that has been shown to have similar features to PR proteins (Herbers *et al.*, 1995). *Chitinase IV* (PR3) expression levels were examined in this study (Fig. 14), and its expression profile is similar to the one obtained for *PAR1* (although different in magnitude). The results suggest that *PAR1* is also potentially involved in the response against PWN infection, and might contribute to PWD tolerance.

4. Concluding remarks and future perspectives

This is the first report on the use of gene expression analysis to study the molecular resistance mechanisms in PWN-inoculated *Pinus pinaster* plants derived from a candidate resistant mother. Selection of candidate resistant trees is the first step in breeding programs that aim to obtain trees with enhanced disease resistance. Since resistance to PWD seems to be determined by several genes, identification of marker genes linked to a marker-assisted selection (MAS) can be an effective approach to accumulate different resistance genes in a population or an individual with long-term utility. Therefore, markers genes involved in tree defense responses against PWN infection must be identified to allow the establishment of this strategy.

Total RNA samples from *P. pinaster* plants showing differential behavior towards PWN infection (susceptible and apparently tolerant) were sent for small RNA and degradome sequencing using Illumina NGS technology. The sequenced small RNAs were analyzed using a pipeline designed in the Forest Biotechnology lab, their differential expression between susceptible and tolerant RNA-seq libraries was analyzed, and the degradome libraries were used to predict targets for the differentially expressed small RNAs. The target transcripts that were selected for expression analysis using real-time quantitative PCR showed expression levels that were most of the times inversely correlated with the expression of the respective regulator small RNAs. Therefore, these results support that these transcripts are in fact true targets of the identified small RNAs, suggesting an involvement of specific small RNA pathways in the response to PWN infection, and a correlation of the differential expression of the target transcripts with the susceptibility/tolerance of plants to PWD. The analyzed target transcripts are putatively involved in the defense response, stress responses (oxidative stress and dehydration), secondary metabolism, transcription and posttranscriptional pathways. Homologs of the secondary metabolism (*Met*), the transcription (*elf2-gamma*), the dehydration (*DREB1*) and the defense response (*Chitinase IV*) related transcripts had already been described in previous transcriptomic studies using PWN-inoculated pines. Novel players – *Tetraspanin*, *DNMT1-RFD*, and *Gag-spuma* – with a potential involvement in the response mechanism against PWN infection have been identified in this study. In particular, the results suggest for the first time the involvement of RNA-directed DNA Methylation (RdDM) epigenetic pathway in the response mechanism against PWN infection.

Transcripts selected from a previous study comparing a tolerant species (*P. pinea*) with *P. pinaster* susceptible plants were also differentially expressed between susceptible and apparently

tolerant *P. pinaster* plants in this study. These include an oxidative stress-responsive gene, a water stress-responsive gene, a gene involved in secondary metabolism and a gene related to defense responses. Therefore, the differences in the expression levels of the analyzed transcripts might underlie the differential behavior of plants against PWN infection. Moreover, these genes were also reported in other studies in relation to PWN inoculation, which might suggest common defense responses in different pine species.

The results of this study provide valuable information for future research to elucidate the resistance mechanisms to PWD and to identify potential marker genes to assist tree breeding. Nevertheless, the results should be validated in the future with RNA-seq experiments conducted with samples containing, if possible, higher quality small RNA fractions. Moreover, more replicate libraries should be sequenced to enhance the confidence of the results obtained for differentially expressed small RNAs, and the interaction between small RNAs and their targets should be validated *in vivo*. Plant protoplasts (from *Arabidopsis* or, ideally, from conifers) are great cell-based models for this propose, given that freshly isolated mesophyll protoplasts largely maintain the same physiological characteristics of whole plants. Small RNA and target interaction can be assessed in protoplasts using a reporter-gene assay quantification, like the luciferase-based sensors recently developed for quantifying miRNA activity in protoplasts (Martinho *et al.*, 2015; Confraria and Baena-González, 2016).

5. References

- Addo-Quaye, C., Miller, W. and Axtell, M.J.** (2009) CleaveLand: A pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics*, **25**, 130–131.
- Akbergenov, R., Si-Ammour, A., Blevins, T., et al.** (2006) Molecular characterization of geminivirus-derived small RNAs in different plant species. *Nucleic Acids Res.*, **34**, 462–471.
- Akimoto, K., Katakami, H., Kim, H.J., Ogawa, E., Sano, C.M., Wada, Y. and H, S.** (2007) Epigenetic inheritance in rice plants. *Ann Bot*, **100**, 205–217.
- Allen, E., Xie, Z., Gustafson, A.M. and Carrington, J.C.** (2005) microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. *Cell*, **121**, 207–221.
- Andersen, C.L., Ledet-Jensen, J. and Orntoft, T.** (2004) Normalization of real-time quantitative RT-PCR data: a mode-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.*, **64**, 5245–5250.
- Arakawa, Y. and Togashi, K.** (2002) Newly discovered transmission pathway of *Bursaphelenchus xylophilus* from males of the beetle *Monochamus alternatus* to *Pinus densiflora* trees via oviposition wounds. *J. Nematol.*, **34**, 396–404.
- Axtell, M.** (2013) CleaveLand4: Analysis of degradome data to find sliced miRNA and siRNA targets. Available at: <https://github.com/MikeAxtell/CleaveLand4>.
- Axtell, M.J., Westholm, J.O. and Lai, E.C.** (2011) Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol.*, **12**, 221.
- Bartel, D.P.** (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
- Bolla, R.I. and Boschert, M.** (1993) Pinewood nematode species complex: interbreeding potential and chromosome number. *J. Nematol.*, **25**, 227–238.
- Budak, H. and Akpinar, B.A.** (2015) Plant miRNAs: biogenesis, organization and origins. *Funct. Integr. Genomics*, **15**, 523–531.
- Burdon, D.R.** (2001) Genetic diversity and disease resistance: some considerations for research, breeding, and deployment. *Can. J. For. Res.*, **31**, 596–606.
- Bustin, S.A.** (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.*, **25**, 169–193.
- Bustin, S.A., Beaulieu, J.F., Huggett, J., Jaggi, R., Kibenge, F.S., Olsvik, P.A., Penning, L.C. and Toegel, S.** (2010) MIQE précis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Mol. Biol.*, **11**, 74.
- Bustin, S.A., Benes, V., Garson, J.A., et al.** (2009) The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clin. Chem.*, **55**, 611–622.

- Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Gilbert, K.B., Montgomery, T.A., Nguyen, T., Cuperus, J.T. and Carrington, J.C.** (2012) Functional analysis of three Arabidopsis ARGONAUTES using slicer- defective mutants. *Plant Cell*, **24**, 3613–3629.
- Carson, S.D. and Carson, M.J.** (1989) Breeding for resistance in forest trees—a quantitative genetic approach. *Annu. Rev. Phytopathol.*, **27**, 373–395.
- Carthew, R.W. and Sontheimer, E.J.** (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell*, **136**, 642–655.
- Chang, S., Puryear, J. and Cairney, J.** (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol.*, **11**, 114–117.
- Charfeddine, M., Bouaziz, D., Charfeddine, S., Hammami, A., Ellouz, O.N. and Bouzid, R.G.** (2015) Overexpression of dehydration-responsive element-binding 1 protein (DREB1) in transgenic *Solanum tuberosum* enhances tolerance to biotic stress. *Plant Biotechnol. Rep.*, **9**, 79–88.
- Chaves, I., Lin, Y.-C., Pinto-Ricardo, C., Peer, Y. Van de and Miguel, C.** (2014) miRNA profiling in leaf and cork tissues of *Quercus suber* reveals novel miRNAs and tissue-specific expression patterns. *Tree Genet. Genomes*, **10**, 721–737.
- Chen, Y., Lun, A.T.L. and Smyth, G.K.** (2014) Differential expression analysis of complex RNA-seq experiments using edgeR. In S. Datta and D. S. Nettleton, eds. *Statistical Analysis of Next Generation Sequence Data*. New York: Springer, pp. 51–74.
- Confraria, A. and Baena-González, E.** (2016) Using Arabidopsis Protoplasts to Study Cellular Responses to Environmental Stress. In P. Duque, ed. *Environmental Responses in Plants: Methods and Protocols*. New York: Springer New York, pp. 247–269.
- Costa, B. V, Rodrigues, A., Chaves, I., Bohn, A. and Miguel, C.** (2016) miRPursuit: a pipeline for analysis of large-scale plant small RNA datasets. Available at: <https://github.com/forestbiotechlab/miRPursuit>.
- Cuperus, J.T., Carbonell, A., Fahlgren, N., et al.** (2010) Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat. Struct. Mol. Biol.*, **17**, 997–1003.
- Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K.D., Carrington, J.C. and Voinnet, O.** (2006) Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science*, **313**, 68–71.
- Derveaux, S., Vandesompele, J. and Hellemans, J.** (2010) How to do successful gene expression analysis using real-time PCR. *Methods*, **50**, 227–230.
- Dolgosheina, E. V, Morin, R.D., Aksay, G., Sahinalp, S.C., Magrini, V., Mardis, E.R., Mattsson, J. and Unrau, P.J.** (2008) Conifers have a unique small RNA silencing signature. *RNA*, **14**, 1508–1515.

- Downen, R.H., Pelizzola, M., Schmitz, R.J., Lister, R., Downen, J.M., Nery, J.R., Dixon, J.E. and Ecker, J.R.** (2012) Widespread dynamic DNA methylation in response to biotic stress. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, E2183–E2191.
- Dropkin, V., Linit, M., Kondo, E. and Smith, M.** (1971) Pine wilt associated with *Bursaphelenchus xylophilus* (Steiner&Buhrer, 1934) Nickle, 1970, in the United States of America. In *Proc. IUFRO World Congr., 17th, Kyoto*. Kukizaki, Japan: IUFRO Congr. Comm., pp. 265–68.
- Ebrahim, S., Usha, K. and Singh, B.** (2011) Pathogenesis Related (PR) Proteins in Plant Defense Mechanism Age-Related Pathogen Resistance. In A. Méndez-Vilas, ed. *Science against pathogens: communicating current research and technological advances*. Badajoz, Spain: Formatex, pp. 1043–1054.
- Edwards, O.R. and Linit, M.J.** (1992) Transmission of *Bursaphelenchus xylophilus* through oviposition wounds of *Monochamus carolinensis* (Coleoptera: Cerambycidae). *J. Nematol.*, **24**, 133–139.
- Enda, N. and Ikeda, T.** (1983) Role of volatiles of a pine tree as emerging stimulants for attracting the pine wood nematode from the pine sawyer. *Trans. Meet. Japanese For. Soc. Mtg Jpn Soc*, **94**, 479–480.
- Erickson, F.L. and Hannig, E.M.** (1996) Ligand interactions with eukaryotic translation initiation factor 2: role of the gamma-subunit. *EMBO J.*, **15**, 6311–6320.
- Evert, R.** (2006) *Esau's plant anatomy: meristems, cells, and tissues of the plant body: their structure, function, and development* 3rd ed., New Jersey, USA: Wiley.
- Fonseca, L., Cardoso, J.M.S., Lopes, A., Pestana, M., Abreu, F., Nunes, N., Mota, M. and Abrantes, I.** (2012) The pinewood nematode, *Bursaphelenchus xylophilus*, in Madeira Island. *Helminthologia*, **49**, 96–103.
- Fuks, F., Burgers, W.A., Brehm, A., Hughes-Davies, L. and Kouzarides, T.** (2000) DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat. Genet.*, **24**, 88–91.
- Fukuda, K.** (1997) Physiological process of the symptom development and resistance mechanism in pine wilt disease. *J. For. Res.*, **2**, 171–181.
- Fukuda, K., Hogetsu, T. and Suzuki, K.** (1992) Photosynthesis and water status in pine wood nematode-infected pine seedlings. *J. Japanese For. Soc.*, **74**, 1–8.
- Futai, K.** (2003) Pine Wilt Disease: Various biological relationships and resulting events. *Proc. IUFRO Kanazawa 2003 "Forest Insect Popul. Dyn. Host Infl.*
- Futai, K.** (2013) Pine Wood Nematode, *Bursaphelenchus xylophilus*. *Annu. Rev. Phytopathol.*, **51**, 61–83.
- Futai, K. and Furuno, T.** (1979) The variety of resistances among pine species to pine wood nematode, *Bursaphelenchus lignicolus*. *Bull. Kyoto Univ. For.*, **51**, 23–36.
- Gachon, C., Mingam, A. and Charrier, B.** (2004) Real-time PCR: what relevance to plant studies? *J. Exp. Bot.*, **55**, 1445–1454.

- Gao, J. and Lan, T.** (2016) Functional characterization of the late embryogenesis abundant (LEA) protein gene family from *Pinus tabulaeformis* (Pinaceae) in *Escherichia coli*. *Sci. Rep.*, **6**, 19467.
- Gaspar, N.J., Kinzy, T.G., Scherer, B.J., Humbelin, M., Hershey, J.W. and Merrick, W.C.** (1994) Translation initiation factor eIF-2. Cloning and expression of the human cDNA encoding the gamma-subunit. *J. Biol. Chem.*, **269**, 3415–3422.
- German, M.A., Pillay, M., Jeong, D.-H., et al.** (2008) Global identification of microRNA–target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.*, **26**, 941–946.
- Gonçalves, S., Cairney, J., Maroco, J., Oliveira, M.M. and Miguel, C.** (2005) Evaluation of control transcripts in real-time RT-PCR expression analysis during maritime pine embryogenesis. *Planta*, **222**, 556–563.
- Granados, J.M., Ávila, C., Cánovas, F.M. and Cañas, R.A.** (2016) Selection and testing of reference genes for accurate RT-qPCR in adult needles and seedlings of maritime pine. *Tree Genet. Genomes*, **12**, 60.
- Griffiths-Jones, S.** (2006) miRBase: the microRNA sequence database. *Methods Mol. Biol.*, **342**, 129–138.
- Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R. and Hofacker, I.L.** (2008) The Vienna RNA websuite. *Nucleic Acids Res.*, **36**, 70–74.
- Gutierrez, L., Mauriat, M., Pelloux, J., Bellini, C. and Wuytswinkel, O. Van** (2008) Towards a Systematic Validation of References in Real-Time RT-PCR. *Plant Cell*, **20**, 1734–1735.
- Hasegawa, K. and Miwa, J.** (2008) Embryology and Cytology of *Bursaphelenchus xylophilus*. In B. G. Zhao, K. Futai, J. R. Sutherland, and Y. Takeuchi, eds. *Pine Wilt Disease*. Tokyo: Springer, pp. 81–104.
- Havecker, E.R., Wallbridge, L.M., Hardcastle, T.J., Bush, M.S., Kelly, K.A., Dunn, R.M., Schwach, F., Doonan, J.H. and Baulcombe, D.C.** (2010) The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell*, **22**, 321–334.
- Herbers, K., Mönke, G., Badur, R. and Sonnewald, U.** (1995) A simplified procedure for the subtractive cDNA cloning of photoassimilate-responding genes: isolation of cDNAs encoding a new class of pathogenesis-related proteins. *Plant Mol. Biol.*, **29**, 1027–1038.
- Herrmann, K.M.** (1995) The Shikimate Pathway: Early Steps in the Biosynthesis of Aromatic Compounds. *Plant Cell*, **7**, 907–919.
- Hewezi, T., Howe, P., Maier, T.R. and Baum, T.J.** (2008) Arabidopsis small RNAs and their targets during cyst nematode parasitism. *Mol. Plant. Microbe. Interact.*, **21**, 1622–1634.
- Hillis, W.E.** (1987) *Heartwood and tree exudates*, Berlin: Springer.
- Hirao, T., Fukatsu, E. and Watanabe, A.** (2012) Characterization of resistance to pine wood nematode

- infection in *Pinus thunbergii* using suppression subtractive hybridization. *BMC Plant Biol.*, **12**, 13.
- Horns, F. and Hood, M.E.** (2012) The evolution of disease resistance and tolerance in spatially structured populations. *Ecol. Evol.*, **2**, 1705–1711.
- Huggett, J., Dheda, K., Bustin, S. and Zumla, A.** (2005) Real-time RT-PCR normalisation: strategies and considerations. *Genes Immun.*, **6**, 279–284.
- Ikeda, T., Oda, K., Yamane, A. and Enda, N.** (1980) Volatiles from pine logs as the attractant for Japanese pine sawyer *Monochamus alternatus* Hope (Coleoptera: Cerambycidae). *J. Japanese For. Soc.*, **62**, 150–152.
- Ikeda, T. and Suzaki, T.** (1984) Influence of pine wood nematodes on hydraulic conductivity and water status in *Pinus thunbergii*. *J. Japanese For. Soc.*, **66**, 412–420.
- Iki, T., Yoshikawa, M., Meshi, T. and Ishikawa, M.** (2012) Cyclophilin 40 facilitates HSP90-mediated RISC assembly in plants. *EMBO J.*, **31**, 267–278.
- Iki, T., Yoshikawa, M., Nishikiori, M., Jaudal, M.C., Matsumoto-Yokoyama, E., Mitsuhashi, I., Meshi, T. and Ishikawa, M.** (2010) In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol. Cell*, **39**, 282–291.
- Ishibashi, N. and Kondo, E.** (1977) Occurrence and survival of the dispersal forms of pine wood nematode, *Bursaphelenchus lignicolus* Mamiya and Kiyohara. *Appl Entomol Zool*, **12**, 293–302.
- Isoda, K., Watanabe, A. and Kuramoto, T.** (2007) Mapping of Japanese red pine and Japanese black pine using microsatellite markers. *Trans. Japanese For. Soc.*, **118**, 018.
- Jones, J.T., Moens, M., Mota, M., Li, H. and Kikuchi, T.** (2008) *Bursaphelenchus xylophilus*: opportunities in comparative genomics and molecular host-parasite interactions. *Mol. Plant Pathol.*, **9**, 357–368.
- Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B.** (2006) MicroRNAs and Their Regulatory Roles in Plants. *Annu. Rev. Plant Biol.*, **57**, 19–53.
- Kerpedjiev, P., Hammer, S. and Hofacker, I.L.** (2015) Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics*, **31**, 3377–3379.
- Khraiwesh, B., Zhu, J.-K. and Zhu, J.** (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim. Biophys. Acta*, **1819**, 137–148.
- Kim, S., Yang, J.-Y., Xu, J., Jang, I.-C., Prigge, M.J. and Chua, N.-H.** (2008) Two Cap-Binding Proteins CBP20 and CBP80 are involved in Processing Primary MicroRNAs. *Plant Cell Physiol.*, **49**, 1634–1644.
- Kimball, S.R.** (1999) Eukaryotic initiation factor eIF2. *Int. J. Biochem. Cell Biol.*, **31**, 25–29.
- Kiyohara, K. and Dozono, Y.** (1986) The relationship between the pathogenicity and the population growth of pine wood nematode. *Trans. Annu. Meet. Kyushu branch Japanese For. Soc.*, **39**, 157.
- Kiyohara, T. and Bolla, R.I.** (1990) Pathogenic variability among populations of the pinewood nematode,

- Bursaphelenchus xylophilus. *For. Sci.*, **36**, 1061–1076.
- Kiyohara, T. and Suzuki, K.** (1975) Population changes of Bursaphelenchus lignicolus in Pinus thunbergii after inoculation. *Trans. Meet. Japanese For. Soc.*, **86**, 296–298.
- Kiyohara, T. and Tokushige, Y.** (1971) Inoculation experiments of a nematode, Bursaphelenchus sp., onto pine trees. *J. Japanese For. Soc.*, **53**, 210–218.
- Kondo, E. and Ishibashi, N.** (1978) Ultrastructural differences between the propagative and dispersal forms in pine wood nematode, Bursaphelenchus lignicolus, with reference to the survival. *Appl Entomol Zool*, **13**, 1–11.
- Kover, P.X. and Schaal, B. a** (2002) Genetic variation for disease resistance and tolerance among Arabidopsis thaliana accessions. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 11270–11274.
- Kozera, B. and Rapacz, M.** (2013) Reference genes in real-time PCR. *J. Appl. Genet.*, **54**, 391–406.
- Kumar, R.** (2014) Role of MicroRNAs in Biotic and Abiotic Stress Responses in Crop Plants. *Appl. Biochem. Biotechnol.*, **174**, 93–115.
- Kuroda, K.** (2008) Physiological Incidences Related to Symptom Development and Wilting Mechanism. In B. G. Zhao, K. Futai, J. R. Sutherland, and Y. Takeuchi, eds. *Pine Wilt Disease*. Springer, pp. 204–222.
- Kuroda, K. and Ito, S.** (1992) Migration speed of pine wood nematode and activities of other microbes during the development of pine-wilt disease in Pinus thunbergii. *J. Japanese For. Soc.*, **74**, 383–389.
- Kuroda, K., Yamada, T. and Ito, S.** (1991) Development of the pin-wilt disease in Pinus densiflora from the standpoint of water conduction. *J. Japanese For. Soc.*, **73**, 69–72.
- Kuroda, K., Yamada, T., Mineo, K. and Tamura, H.** (1988) Effects of cavitation on the development of pine wilt disease caused by Bursaphelenchus xylophilus. *Ann. Phytopathol. Soc. Japan*, **54**, 606–615.
- Kusumoto, D., Yonemichi, T., Inoue, H., Hirao, T., Watanabe, A. and Yamada, T.** (2014) Comparison of histological responses and tissue damage expansion between resistant and susceptible Pinus thunbergii infected with pine wood nematode Bursaphelenchus xylophilus. *J. For. Res.*, **19**, 285–294.
- Kusunoki, M.** (1987) Symptom developments of pine wilt disease-Histological observations with electron microscope. *Ann. Phytopathol. Soc. Japan*, **53**, 622–629.
- Law, J.A. and Jacobsen, S.E.** (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.*, **11**, 204–220.
- Lee, S.M., Choo, H.Y., Park, N.C., Moon, Y.S. and Kim, J.B.** (1990) Nematodes and insects associated with dead trees, and pine wood nematode detection from part of Monochamus alternatus. *Korean J. Appl. Entomol.*, **29**, 14–19.
- Li, J., Reichel, M., Li, Y. and Millar, A.A.** (2014) The functional scope of plant microRNA-mediated

- silencing. *Trends Plant Sci.*, **19**, 750–756.
- Li, J., Yang, Z., Yu, B., Liu, J. and Chen, X.** (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr. Biol.*, **15**, 1501–1507.
- Li, X., Wang, X., Zhang, S., Liu, D., Duan, Y. and Dong, W.** (2012) Identification of Soybean MicroRNAs Involved in Soybean Cyst Nematode Infection by Deep Sequencing. *PLoS One*, **7**, e39650.
- Lian, C., Miwa, M. and Hogetsu, T.** (2000) Isolation and characterization of microsatellite loci from the Japanese red pine, *Pinus densiflora*. *Mol Ecol*, **9**, 1186–1188.
- Linit, M.J., Kondo, E. and Smith, M.T.** (1983) Insects associated with the pinewood nematode, *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae), in Missouri. *Environ. Entomol.*, **12**, 467–470.
- Lisboa, A.F.D. de L.** (2016) *Melhoramento genético florestal e o desafio imposto pela doença da murchidão do pinheiro*. Tese de Mestrado. Instituto Superior de Agronomia, Universidade de Lisboa.
- Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D. and Doudna, J.A.** (2006) Structural basis for double-stranded RNA processing by Dicer. *Science.*, **311**, 195–198.
- Maehara, N. and Futai, K.** (1996) Factors affecting both the numbers of the pinewood nematode, *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae), carried by the Japanese pine sawyer, *Monoctonus alternatus* (Coleoptera: Cerambycidae), and the nematode's life history. *Appl Entomol Zool*, **31**, 443–452.
- Maehara, N. and Futai, K.** (2002) Factors affecting the number of *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae) carried by several species of beetles. *Nematology*, **4**, 653–658.
- Mamiya, Y.** (1980) Inoculation of the first year pine (*Pinus densiflora*) seedlings with *Bursaphelenchus lignicolus* and histopathology of diseased seedlings. *J. Japanese For. Soc.*, **62**, 176–183.
- Mamiya, Y.** (1984) Resistance of *Pinus* spp. against *Bursaphelenchus xylophilus* and *B. mucronatus* focusing on the host age. *Trans. Meet. Japanese For. Soc.*, **95**, 475–476.
- Mamiya, Y.** (1975) The life history of the pine wood nematode, *Bursaphelenchus lignicolus*. *Japanese J. Nematol.*, **5**, 16–25.
- Mamiya, Y. and Enda, N.** (1972) Transmission of *Bursaphelenchus lignicolus* (Nematoda: Aphelenchoididae) By *Monoctonus alternatus* (Coleoptera: Cerambycidae). *Nematologica*, **18**, 159–162.
- Marshall, O.J.** (2004) PerlPrimer: Cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics*, **20**, 2471–2472.
- Martinho, C., Confraria, A., Elias, C.A., Crozet, P., Rubio-Somoza, I., Weigel, D. and Baena-González, E.** (2015) Dissection of miRNA pathways using Arabidopsis mesophyll protoplasts. *Mol. Plant*, **8**, 261–275.
- McCarthy, D.J., Chen, Y. and Smyth, G.K.** (2012) Differential expression analysis of multifactor RNA-

- Seq experiments with respect to biological variation. *Nucleic Acids Res.*, **40**, 4288–4297.
- Meyers, B.C., Souret, F.F., Lu, C. and Green, P.J.** (2006) Sweating the small stuff: MicroRNA discovery in plants. *Curr. Opin. Biotechnol.*, **17**, 139–146.
- Miller, M.R., White, A. and Boots, M.** (2005) The evolution of host resistance: tolerance and control as distinct strategies. *J. Theor. Biol.*, **236**, 198–207.
- Moens, M. and Perry, R.N.** (2009) Migratory Plant Endoparasitic Nematodes: A Group Rich in Contrasts and Divergence. *Annu. Rev. Phytopathol.*, **47**, 313–332.
- Morimoto, K. and Iwasaki, A.** (1972) Role of *Monochamus alternatus* (Coleoptera: Cerambycidae) as a vector of *Bursaphelenchus lignicolus* (Nematoda: Aphelenchoididae). *J. Japanese For. Soc.*, **54**, 177–183.
- Mota, M. and Vieira, P.** (2008) Pine Wilt Disease in Portugal. In B. G. Zhao, K. Futai, J. R. Sutherland, and Y. Takeuchi, eds. *Pine Wilt Disease*. Japan: Springer.
- Mota, M.M., Braasch, H., Bravo, M.A., Penas, A.C., Burgermeister, W., Metge, K. and Sousa, E.** (1999) First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. *Nematology*, **1**, 727–734.
- Nakamura-Matori, K.** (2008) Vector–Host Tree Relationships and the Abiotic Environment. In B. G. Zhao, K. Futai, J. R. Sutherland, and Y. Takeuchi, eds. *Pine Wilt Disease*. Japan: Springer, pp. 144–161.
- Nose, M. and Shiraishi, S.** (2008) Breeding for Resistance to Pine Wilt Disease. In B. G. Zhao, K. Futai, J. R. Sutherland, and Y. Takeuchi, eds. *Pine Wilt Disease*. Tokyo: Springer Japan, pp. 334–350.
- Nose, M. and Shiraishi, S.** (2011) Comparison of the gene expression profiles of resistant and non-resistant Japanese black pine inoculated with pine wood nematode using a modified LongSAGE technique. *For. Pathol.*, **41**, 143–155.
- Pareek, M., Yogindran, S., Mukherjee, S.K. and Rajam, M. V** (2015) Plant MicroRNAs: Biogenesis, Functions, and Applications. In B. Bahadur, M. V Rajam, L. Sahijram, and K. V Krishnamurthy, eds. *Plant Biology and Biotechnology: Volume II: Plant Genomics and Biotechnology*. India: Springer, pp. 639–661.
- Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H. and Poethig, R.S.** (2005) Nuclear processing and export of microRNAs in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 3691–3696.
- Pascual, M.B., Cánovas, F.M. and Ávila, C.** (2015) The NAC transcription factor family in maritime pine (*Pinus Pinaster*): molecular regulation of two genes involved in stress responses. *BMC Plant Biol.*, **15**, 254.
- Paterson, A.H., Tanksley, S.D. and Sorrells, M.E.** (1991) DNA markers in plant improvement. *Adv. Agron.*, **46**, 39–90.
- Pavet, V., Quintero, C., Cecchini, N.M., Rosa, A.L. and Alvarez, M.E.** (2006) Arabidopsis displays centromeric DNA hypomethylation and cytological alterations of heterochromatin upon attack by

- Pseudomonas syringae*. *Mol. Plant-Microbe Interact.*, **19**, 577–587.
- Pfaffl, M.** (2004) Quantification strategies in real-time PCR. In S. A. Bustin, ed. *A-Z of quantitative PCR*. CA, USA: International University Line, pp. 87–112.
- Pfaffl, M.W.** (2001) Development and validation of an externally standardised quantitative insulin like growth factor-1 (IGF-1) RT–PCR using LightCycler SYBR® Green I technology. In S. Meuer, C. Wittwer, and K. Nakagawara, eds. *Rapid Cycle Real-time PCR, Methods and Applications*. Heidelberg, Germany: Springer Press, pp. 281–291.
- Pfaffl, M.W., Tichopad, A., Prgomet, C. and Neuvians, T.P.** (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnol. Lett.*, **26**, 509–515.
- Prüfer, K., Stenzel, U., Dannemann, M., Green, R.E., Lachmann, M. and Kelso, J.** (2008) PatMaN: rapid alignment of short sequences to large databases. *Bioinformatics*, **24**, 1530–1531.
- Ramachandran, V. and Chen, X.** (2008) Degradation of microRNAs by a family of exoribonucleases in *Arabidopsis*. *Science*, **321**, 1490–1492.
- Rasmussen, R.** (2001) Quantification on the LightCycler. In S. Meuer, C. Wittwer, and K. Nakagawara, eds. *Rapid Cycle Real-time PCR, Methods and Applications*. Heidelberg, Germany: Springer Press, pp. 21–34.
- Reboredo, F. ed.** (2014) *Forest Context and Policies in Portugal - Present and Future Challenges*, Springer International Publishing.
- Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. and Bartel, D.P.** (2002) Prediction of plant microRNA targets. *Cell*, **110**, 513–520.
- Ribeiro, B., Espada, M., Vu, T., Nóbrega, F., Mota, M. and Carrasquinho, I.** (2012) Pine wilt disease: detection of the pinewood nematode (*Bursaphelenchus xylophilus*) as a tool for a pine breeding programme. *For. Pathol.*, **42**, 521–525.
- Robertson, L., Arcos, S.C., Escuer, M., Merino, R.S., Esparrago, G., Abellera, A. and Navas, A.** (2011) Incidence of the pinewood nematode *Bursaphelenchus xylophilus* Steiner & Buhrer, 1934 (Nickle, 1970) in Spain. *Nematology*, **13**, 755–757.
- Robinson, M.D., McCarthy, D.J. and Smyth, G.K.** (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 1.
- Robinson, M.D. and Smyth, G.K.** (2008) Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*, **9**, 321–332.
- Rodrigues, J.M.** (2008) Eradication program for the pinewood nematode in Portugal. In M. Mota and P. Vieira, eds. *Pine wilt disease: a worldwide threat to forest ecosystems*. Springer Netherlands, p. 3.
- Rogers, K. and Chen, X.** (2013) Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell*, **25**, 2383–2399.

- Roje, S.** (2006) S-Adenosyl-L-methionine: Beyond the universal methyl group donor. *Phytochemistry*, **67**, 1686–1698.
- Rountree, M.R., Bachman, K.E. and Baylin, S.B.** (2000) DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat. Genet.*, **25**, 269–277.
- Roy, B.A. and Kirchner, J.W.** (2000) Evolutionary dynamics of pathogen resistance and tolerance. *Evolution*, **54**, 51–63.
- Ruiz-Ferrer, V. and Voinnet, O.** (2009) Roles of plant small RNAs in biotic stress responses. *Annu. Rev. Plant Biol.*, **60**, 485–510.
- Sahu, P.P., Pandey, G., Sharma, N., Puranik, S., Muthamilarasan, M. and Prasad, M.** (2013) Epigenetic mechanisms of plant stress responses and adaptation. *Plant Cell Rep.*, **32**, 1151–1159.
- Santos, C.S., Pinheiro, M., Silva, A.I., Egas, C. and Vasconcelos, M.W.** (2012) Searching for resistance genes to *Bursaphelenchus xylophilus* using high throughput screening. *BMC Genomics*, **13**, 599.
- Santos, C.S.S. and Vasconcelos, M.W.** (2012) Identification of genes differentially expressed in *Pinus pinaster* and *Pinus pinea* after infection with the pine wood nematode. *Eur. J. Plant Pathol.*, **132**, 407–418.
- Sato, H., Sakuyama, T. and Kobayashi, M.** (1987) Transmission of *Bursaphelenchus xylophilus* (Steiner et Buhner) Nickle (Nematoda, Aphelenchoididae) by *Monochamus saltuarius* (Gebler) (Coleoptera, Cerambycidae). *J. Japanese For. Soc.*, **69**, 492–496.
- Shin, H., Lee, H., Woo, K.S., Noh, E.W., Koo, Y.B. and Lee, K.J.** (2009) Identification of genes upregulated by pinewood nematode inoculation in Japanese red pine. *Tree Physiol.*, **29**, 411–421.
- Shinya, R., Morisaka, H., Kikuchi, T., Takeuchi, Y., Ueda, M. and Futai, K.** (2013) Secretome Analysis of the Pine Wood Nematode *Bursaphelenchus xylophilus* Reveals the Tangled Roots of Parasitism and Its Potential for Molecular Mimicry. *PLoS One*, **8**, e67377.
- Sousa, E., Bravo, M.A., Pires, J., Naves, P.M., Penas, A.C., Bonifácio, L. and Mota, M.** (2001) *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae) associated with *Monochamus galloprovincialis* (Coleoptera: Cerambycidae) in Portugal. *Nematology*, **3**, 89–91.
- Southey, J.F.** (1986) *Laboratory methods for work with plant and soil nematodes*, London: Ministry of Agriculture, Fisheries and Food, Reference Book 402.
- Spanakis, E.** (1993) Problems related to the interpretation of autoradiographic data on gene expression using common constitutive transcripts as controls. *Nucleic Acids Res.*, **21**, 3809–3819.
- Sperry, J. and Tyree, M.T.** (1988) Mechanism of water stress-induced xylem embolism. *Plant Physiol.*, **88**, 581–587.
- Stamps, W.T. and Linit, M.J.** (1998) Neutral storage lipid and exit behavior of *Bursaphelenchus xylophilus* fourth-stage dispersal juveniles from their beetle vectors. *J. Nematol.*, **30**, 255–261.

- Stocks, M.B., Moxon, S., Mapleson, D., Woolfenden, H.C., Mohorianu, I., Folkes, L., Schwach, F., Dalmay, T. and Moulton, V.** (2012) The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics*, **28**, 2059–2061.
- Suzuki, T., Higgins, P.J. and Crawford, D.R.** (2000) Control selection for RNA quantitation. *Biotechniques*, **29**, 332–337.
- Tamura, H., Mineo, K. and Yamada, T.** (1987) Blockage of water conduction in *Pinus thunbergii* inoculated with *Bursaphelenchus xylophilus*. *Japanese J. Nematol.*, **17**, 23–30.
- Tamura, H., Yamada, T. and Mineo, K.** (1988) Host responses and nematode distribution in *Pinus strobus* and *P. densiflora* infected with the pine wood nematode, *Bursaphelenchus xylophilus*. *Ann. Phytopathol. Soc. Japan*, **54**, 327–331.
- Thellin, O., Zorzi, W., Lakaye, B., Borman, B. De, Coumans, B., Hennen, G., Grisar, T., Igout, A. and Heinen, E.** (1999) Housekeeping genes as internal standards: use and limits. *J. Biotechnol.*, **75**, 291–295.
- Togashi, K.** (1989a) Development of *Monochamus alternatus* Hope (Coleoptera: Cerambycidae) in *Pinus thunbergii* trees weakened at different times. *J. Japanese For. Soc.*, **71**, 383–386.
- Togashi, K.** (1989b) Development of *Monochamus alternatus* Hope (Coleoptera: Cerambycidae) in relation to oviposition time. *Japanese J. Appl. Entomol. Zool.*, **33**, 1–8.
- Togashi, K.** (1991) Different developments of overwintered larvae of *Monochamus alternatus* (Coleoptera: Cerambycidae) under a constant temperature. *Japanese J. Entomol.*, **59**, 149–154.
- Togashi, K.** (2006) Life of the Japanese pine sawyer, *Monochamus alternatus* Hope. In E. Shibata and K. Togashi, eds. *The fascinating lives of insects residing in tree trunks: an introduction to tree-boring insects*. Hadano: Tokai University Press, pp. 83–106.
- Togashi, K.** (1991) Spatial pattern of pine wilt disease caused by *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae) within a *Pinus thunbergii* stand. *Res. Popul. Ecol. (Kyoto)*, **33**, 245–256.
- Togashi, K.** (1989) Studies on population dynamics of *Monochamus alternatus* Hope (Coleoptera: Cerambycidae) and spread of pine wilt disease caused by *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae). *Bull. Ishikawa-ken For. Exp. Stn.*, **20**, 1–142.
- Togashi, K.** (2008) Vector–Nematode Relationships and Epidemiology in Pine Wilt Disease. In B. G. Zhao, K. Futai, J. R. Sutherland, and Y. Takeuchi, eds. *Pine Wilt Disease*. Japan: Springer, pp. 162–183.
- Togashi, K. and Arakawa, Y.** (2003) Horizontal transmission of *Bursaphelenchus xylophilus* between sexes of *Monochamus alternatus*. *J. Nematol.*, **35**, 7–16.
- Vandenborre, G., Smagghe, G. and Damme, E.J.M. van** (2011) Plant lectins as defense proteins against phytophagous insects. *Phytochemistry*, **72**, 1538–1550.
- Vandesompele, J., Kubista, M. and Pfaffl, M.M.W.** (2009) Reference Gene Validation Software for

Improved Normalization. *Real-time PCR Curr. Technol. Applications*, **4**, 47–64.

- Vandesompele, J., Preter, K. De, Pattyn, F., Poppe, B., Roy, N. Van, Paepe, A. De and Speleman, F.** (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**, research0034.1-0034.11.
- Vega-Bartol, J.J. de, Santos, R.R., Simões, M. and Miguel, C.M.** (2013) Normalizing gene expression by quantitative PCR during somatic embryogenesis in two representative conifer species: *Pinus pinaster* and *Picea abies*. *Plant Cell Rep.*, **32**, 715–729.
- Voinnet, O.** (2009) Origin, Biogenesis, and Activity of Plant MicroRNAs. *Cell*, **136**, 669–687.
- Wada, Y., Miyamoto, K., Kusano, T. and Sano, H.** (2004) Association between up-regulation of stress-responsive genes and hypomethylation of genomic DNA in tobacco plants. *Mol. Genet. Genomics*, **271**, 658–666.
- Wang, F., Muto, A., Velde, J. Van de, Neyt, P., Himanen, K., Vandepoele, K. and Lijsebettens, M. Van** (2015) Functional Analysis of Arabidopsis TETRASPANIN Gene Family in Plant Growth and Development. *Plant Physiol.*, **169**, 2200–2214.
- Warrington, J.A., Nair, A., Mahadevappa, M. and Tsyganskaya, M.** (2000) Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol. Genomics*, **2**, 143–147.
- Whitfield, C.D., Steers, E J, J. and Weissbach, H.** (1970) Purification and Properties of 5'-Methyltetrahydropteroyltriglutamate-Homocysteine Transmethylase. *J. Biol. Chem.*, **245**, 390–401.
- Wingfield, M.J.** (1987) Fungi associated with the pine wood nematode, *Bursaphelenchus xylophilus*, and cerambycid beetles in Wisconsin. *Mycologia*, **79**, 325–328.
- Wingfield, M.J. and Blanchette, R.A.** (1983) The pine-wood nematode, *Bursaphelenchus xylophilus*, in Minnesota and Wisconsin: insect associates and transmission studies. *Can. J. For. Res.*, **13**, 1068–1076.
- Xie, Z., Allen, E., Wilken, A. and Carrington, J.C.** (2005) DICER- LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 12984–12989.
- Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E. and Carrington, J.C.** (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.*, **2**, E104.
- Xu, L., Liu, Z.Y., Zhang, K., Lu, Q., Liang, J. and Zhang, X.Y.** (2013) Characterization of the *Pinus massoniana* transcriptional response to *Bursaphelenchus xylophilus* infection using suppression subtractive hybridization. *Int. J. Mol. Sci.*, **14**, 11356–11375.
- Yamada, T.** (2006) Biochemical responses in pines infected with *Bursaphelenchus xylophilus*. *J. Japanese For. Soc.*, **88**, 370–382.

- Yang, B.J.** (2004) The history, dispersal and potential threat of pine wood nematode in China. In M. Mota and P. Vieira, eds. *The pinewood nematode, Bursaphelenchus xylophilus*. University of Évora, Portugal, 20–22 August 2001. Brill, Leiden- Boston: Proceeding of an international workshop, pp. 21–24.
- Yang, Z., Ebright, Y.W., Yu, B. and Chen, X.** (2006) HEN1 recognizes 21–24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res.*, **34**, 667–675.
- Yu, A., Lepère, G., Jay, F., et al.** (2013) Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 2389–2394.
- Zhao, B.G., Futai, K., Sutherland, J.R. and Takeuchi, Y.** (2008) *Pine Wilt Disease*, Tokyo: Springer.
- Zhao, L., Mota, M., Vieira, P., Butcher, R.A. and Sun, J.** (2014) Interspecific communication between pinewood nematode, its insect vector, and associated microbes. *Trends Parasitol.*, **30**, 299–308.
- Zhao, M., Meyers, B.C., Cai, C., Xu, W. and Ma, J.** (2015) Evolutionary patterns and coevolutionary consequences of MIRNA genes and MicroRNA targets triggered by multiple mechanisms of genomic duplications in soybean. *Plant Cell*, **27**, 546–562.
- Zhou, M., Ma, J., Pang, J., Zhang, Z. and Tang, Y.** (2010) Regulation of plant stress response by dehydration responsive element binding (DREB) transcription factors. *Am. J. Biotechnol.*, **9**, 9255–9269.

6. Appendices

Appendix A – Inoculation of *Pinus pinaster* plants with *Bursaphelenchus xylophilus*, and symptom observation.

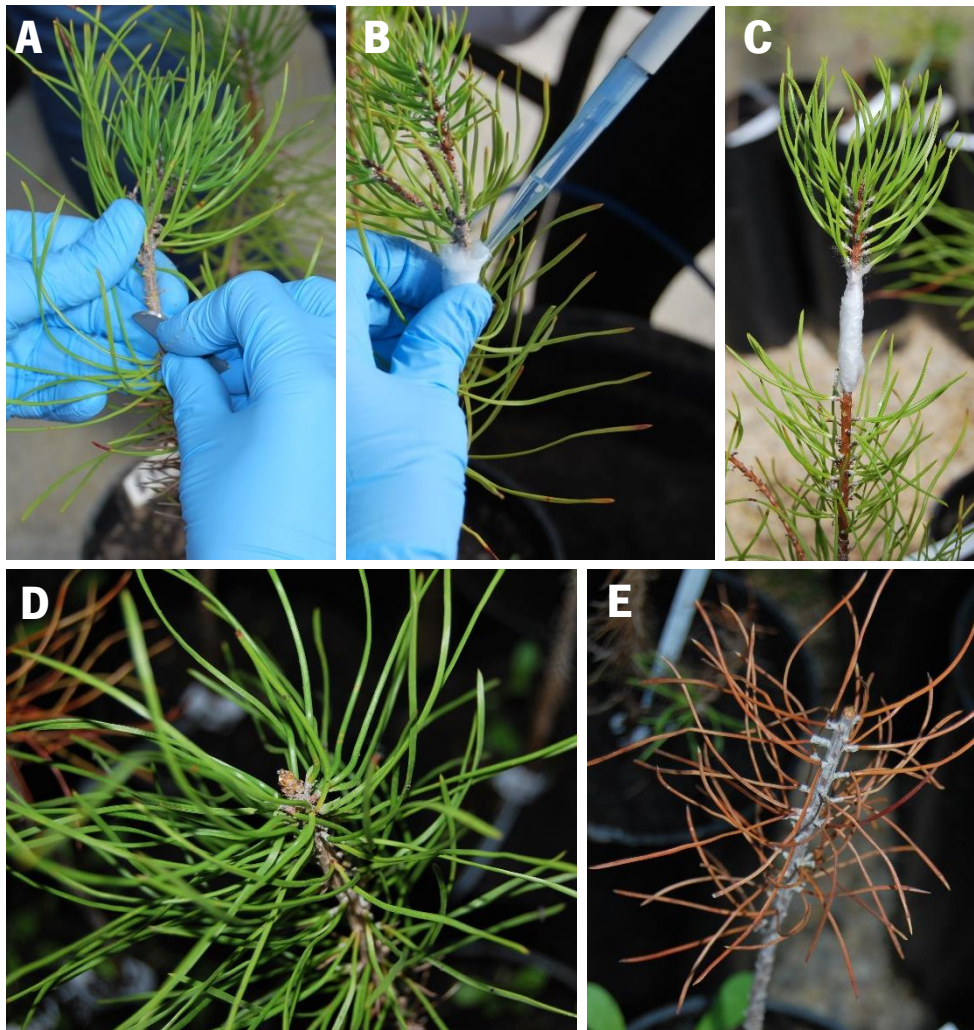


Figure A1 – Inoculation of *Pinus pinaster* plants with *Bursaphelenchus xylophilus*, and pine wilt disease symptom observation. Steps of inoculation protocol: (A) a small longitudinal wound was made about 10 cm from the plant apex, (B) 500 μ L of a water suspension of 1000 nematodes/mL was pipetted into the wound involved with cotton, and (C) the inoculated regions were covered with parafilm to prevent drying of the inoculum. Symptom observation (more than a month after inoculation): (D) apparently tolerant plant (no symptoms of disease have developed); (E) susceptible trees showing needle discoloration.

Appendix B – Agarose gel to check RNA integrity (Example)

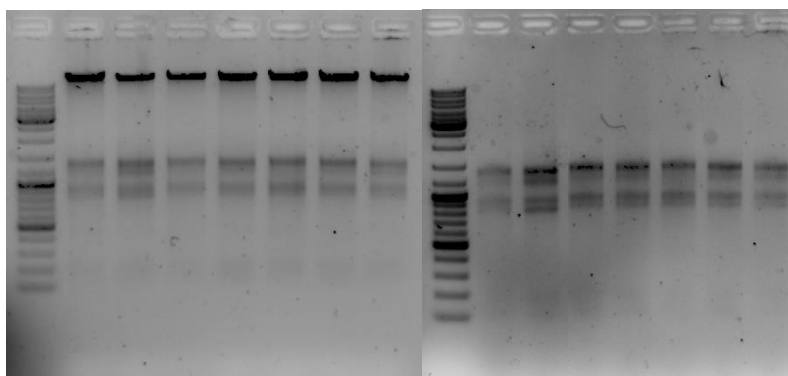


Figure B1 – Examples of agarose gel electrophoresis of a 200ng/ μ L aliquot of RNA samples before (left) and after (right) DNase treatment. GelRed dye (Biotium) was added to each RNA aliquot that was run on a 1.5% (w/v) agarose gel submerged in 1x TE, at about 80 V/cm. Bands were visualized using UV transilluminator-based Gel Doc EZ System or Gel Doc XR+ (Bio-Rad). Molecular Weight Marker – GeneRuler 1 kb Plus DNA Ladder (Thermo Fisher Scientific).

Appendix C – miRPursuit configuration directory files. Each module that constitutes the miRPursuit workflow directory is based on a configuration file:

Appendix C.1 – wbench_filter.cfg configuration file parameters. This file filters the small RNA sequences according to their length, abundance, and tRNA/rRNA.

```
min_length=18
max_length=26
min_abundance=5
max_abundance=2147483647
norm_abundance=false
filter_low_comp=true
filter_invalid=true
trrna=true
trrna_sense_only=false
filter_genome_hits=false
filter_norm_abund=false
filter_kill_list=false
add_discard_log=false
genome=null
kill_list=null
discard_log=null
```

Appendix C.2 – wbench_mirprof.cfg configuration file parameters. Configuration file used by miRProf to identify conserved miRNAs, through alignment to the miRBase database of miRNAs.

```
mismatches=0
overhangs=true
group_mismatches=true
group_organisms=true
group_variant=true
group_mature_and_star=false
only_keep_best=true
```

```
min_length=18
max_length=26
min_abundance=5
```

Appendix C.3 – wbench_tasi.cfg configuration file. Configuration file used by ta-si Predictor of the UEA sRNA workbench to identify 21-nucleotide long small RNAs characteristic of the ta-siRNA loci.

```
p_val_threshold=1.0E-4
min_abundance=2
```

Appendix C.4 – wbench_mircat.cfg configuration file parameters. Configuration file used by miRCat to predict novel miRNAs through alignment with the genome to find putative precursors.

```
extend=100.0
min_energy= -25.0
min_paired=17
max_gaps=3
max_genome_hits=16
min_length=18
max_length=26
min_gc=20
max_unpaired=60
max_overlap_percentage=80
min_locus_size=1
orientation=80
min_hairpin_len=60
complex_loops=true
pval=0.05
min_abundance=1
cluster_sentinel=200
Thread_Count=12
```

Appendix D – Quality plots of small RNA sequencing libraries.

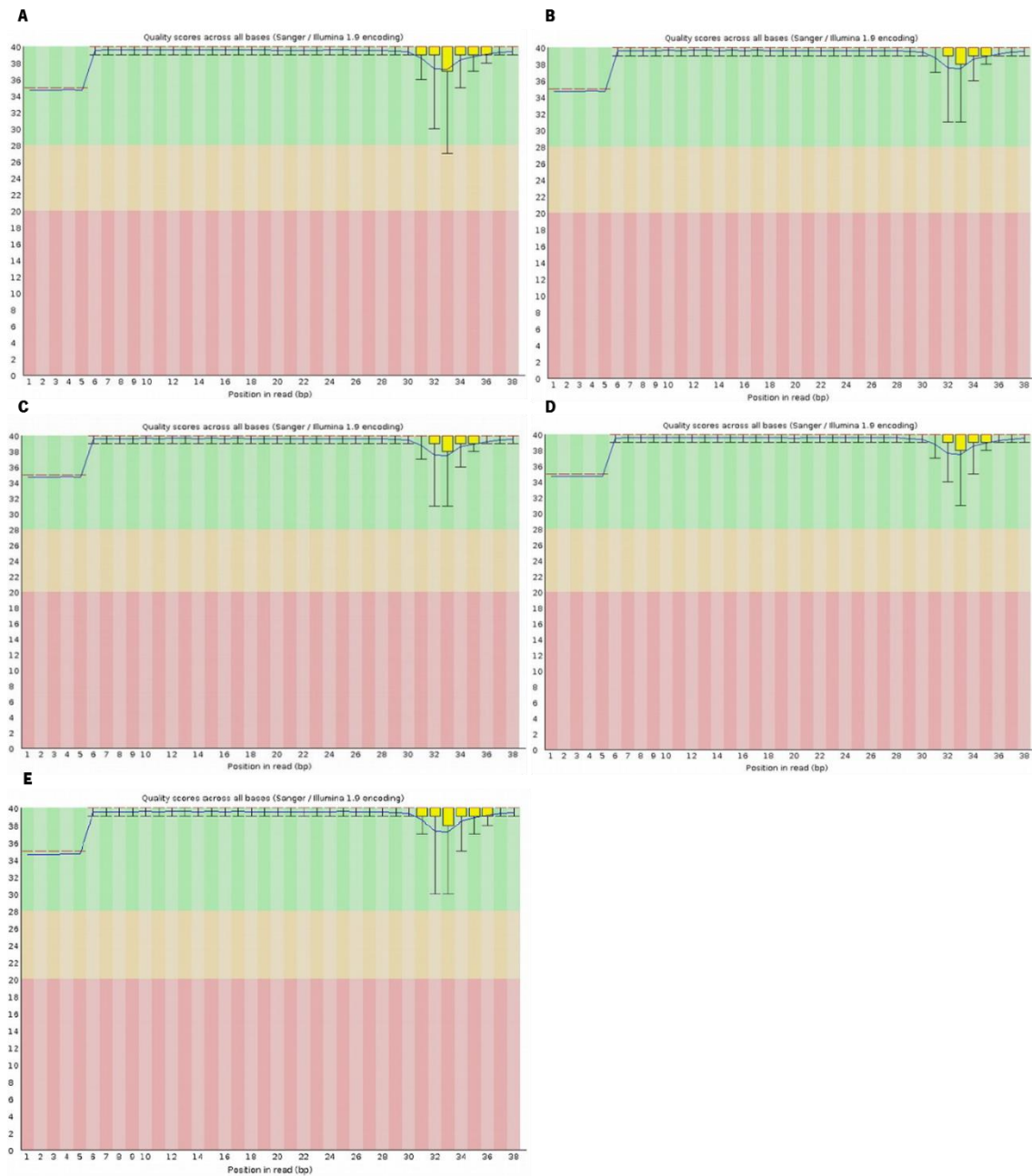
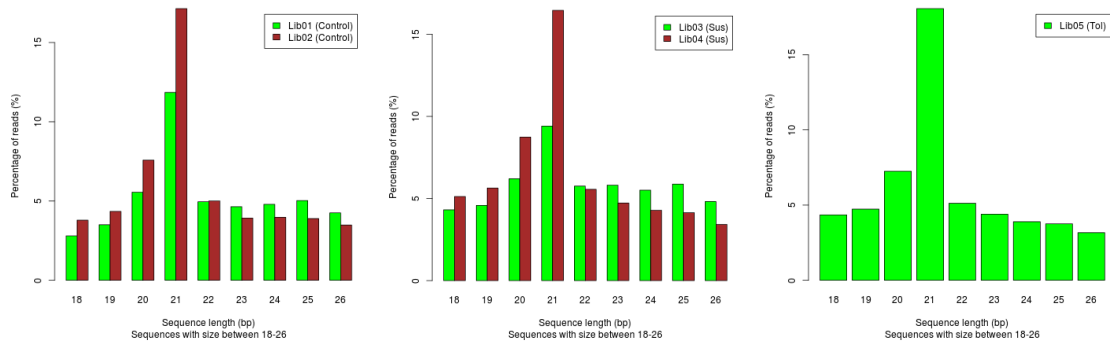


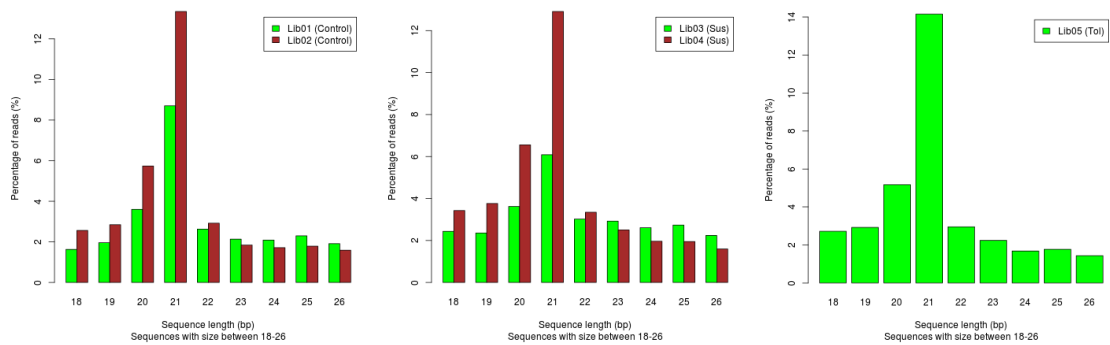
Figure D1 – Quality scores per base for the small RNA libraries sequenced with Illumina HiSeq 2000: (A) Lib01, Control; (B) Lib02, Control; (C) Lib03, Susceptible; (D) Lib04, Susceptible; (E) Lib05, Tolerant. The background of graphs divides the y axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). Red line represents the median value and blue line the mean quality. Yellow box represents the inter-quartile range (25-75%), and upper and lower whiskers represent the 10% and 90% percentiles. Graphs were made using FastQC.

Appendix E – Size profile distribution of reads from small RNA sequencing libraries.

Raw reads



rRNA/tRNA filter



Filtering genome *Pinus taeda*

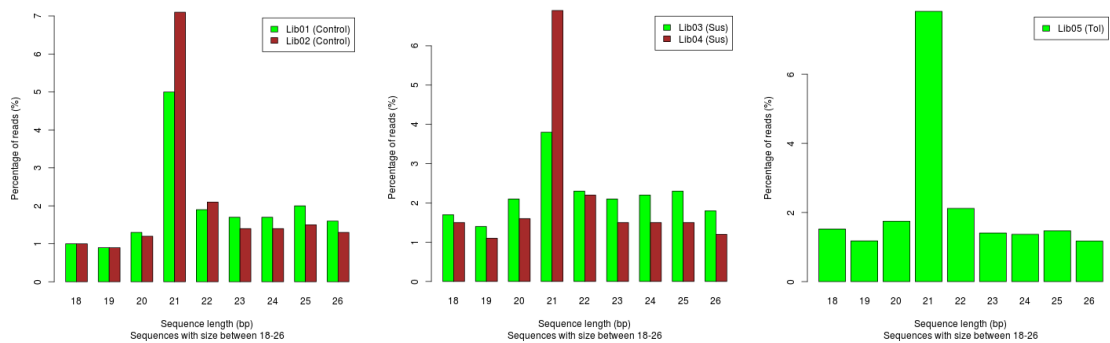


Figure E1 – Size profile distribution profiles of control (Lib01 and Lib02), susceptible (Lib03 and Lib04) and tolerant (Lib05) small RNA sequencing libraries. Each row shows the raw reads, raw reads after filtering rRNA/tRNA and after mapping to reference genome of *Pinus taeda*.

Appendix F – Small RNAs differentially expressed in susceptible and tolerant plants and their respective target transcript(s).

(next page)

Table F1 – List of small RNAs differentially expressed in susceptible (S) and tolerant (T) plants, in respect to the control (C), and their predicted target transcripts.

Small RNAs	Log Fold Change		Raw Counts					Normalized Counts					Predicted Target(s)	Annotation
	C vs. S	C vs. T	Lib01	Lib02	Lib03	Lib04	Lib05	Lib01	Lib02	Lib03	Lib04	Lib05		
Ppi-sRNA-tasi1 CATCTAGAACATGAAAGCCT-tasi-267	0	9,51	61	117	54	35	0	10,16	25,43	7,84	6,64	0	20381	Tetraspanin family protein
													31238	Class IV chitinase
Ppi-sRNA-tasi2 TGCCCCGCGTCCACCGAAGTT-tasi-73	0	9,43	73	98	0	36	0	12,15	21,30	0	6,83	0	103474	NA
													9389	Glycine-rich RNA-binding protein
													73282	Cinnamyl alcohol dehydrogenase
													9389	Glycine-rich RNA-binding protein
Ppi-sRNA-tasi3 TTGTGCGCCTTGATATAAGATAT-tasi-27	7,20	0	14	22	0	0	13	2,33	4,78	0	0	2,71	32561	Fructose-bisphosphate aldolase
													4190	Tymo_45kd_70kd domain containing protein
													8462	Cytosolic ascorbate peroxidase
													9058	14-3-3 Domain-containing protein
Ppi-sRNA-novel1 TCGAGTCAACAAACTCTGGTT-novel-189	2,32	0	24	52	11	6	96	4,00	11,30	1,60	1,14	20,04	22053	Translation eIF-2 gamma subunit
Ppi-sRNA-NA1 AACGAGGTTCAAGCTTAACTC	2,40	0	20	46	9	5	78	3,33	10,00	1,31	0,95	16,28	NA	NA
Ppi-sRNA-novel2 TAGATGACAAAATTGAAGATT-novel-215	2,37	0	30	81	14	10	80	4,99	17,60	2,03	1,90	16,70	7984	Natural resistance-associated macrophage protein
													17886	Pyrophosphate-energized vacuolar membrane proton pump 1
Ppi-sRNA-NA2 TAAAAAATAGCGATATGGACT	2,59	0	38	135	21	12	161	6,33	29,34	3,05	2,28	33,61	NA	NA
Ppi-sRNA-tasi4 CCATTATTCCTGTATTTTAAAC-tasi-155	2,52	0	29	63	6	11	46	4,83	13,69	0,87	2,09	9,60	10492	Glyceraldehyde-3-phosphate dehydrogenase
													10760	Short-chain type dehydrogenase/reductase
													12941	Putative transcription initiation factor IIF beta subunit
													278	Clathrin interactor EPSIN 2
													37654	MAD domain containing protein
													4522	Putative glycine-rich arabinogalactan protein 3
Ppi-sRNA-novel3 TGGAGCTGTTATCACTCCACT-novel-5334	3,27	0	236	2524	186	147	2241	39,29	548,53	26,99	27,88	467,76	1416	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase
Ppi-sRNA-tasi5 AAACCAAATAATCAGTATTT-tasi-152	3,17	0	19	63	5	5	60	3,16	13,69	0,73	0,95	12,52	29907	Heat shock cognate 70 kDa protein
													652	Unassigned protein
Ppi-sRNA-tasi6 TCTACCAGATATTCAAATATC-tasi-398	3,04	0	56	137	19	8	178	9,32	29,77	2,76	1,52	37,15	130074	GJ14593 gene product
													93887	Phosphoglycerate kinase
													28175	DNMT1-RFD multi-domain protein
													80881	GRP domain containing protein
Ppi-sRNA-NA3 GGAAATCAGATGCAAGATCTC	2,91	0	44	209	29	11	262	7,33	45,42	4,21	2,09	54,69	NA	NA

Small RNAs	Log Fold Change		Raw Counts					Normalized Counts					Predicted Target(s)	Annotation
	C vs. S	C vs. T	Lib01	Lib02	Lib03	Lib04	Lib05	Lib01	Lib02	Lib03	Lib04	Lib05		
Ppi-sRNA-novel4 CACGTGCTCCCTTCTCCAAC-novel-79	-3,47	0	2	4	29	43	7	0,33	0,87	4,21	8,16	1,46	34111	Transducin/WD-40 repeat-containing protein
													18246	40S ribosomal protein S23
													21004	Unassigned protein
													2295	Alanine-glyoxylate aminotransferase
													13282	Transcript antisense to ribosomal RNA protein
													17099	S-adenosylmethionine synthase 1
													17443	S-adenosylmethionine synthase 3
													17755	Acetoacetyl-CoA thiolase
													18517	Unassigned protein
													19000	Pfam04146, YTH, YT521-B-like domain
													22850	Cysteine protease
													2949	RNA recognition motif-containing protein
													3173	Nrap domain containing protein
													3328	Plant ubiquitin
Ppi-sRNA-novel5 TACCTGCATCTCCACCA-novel-53	-2,57	0	4	3	31	16	6	0,67	0,65	4,50	3,03	1,25	13282	Transcript antisense to ribosomal RNA protein
													1416	Myosin_tail_1 multi-domain protein
													15053	ANTH domain-containing protein
													17340	Heat shock protein 90
													17443	S-adenosylmethionine synthase 3
													18488	Protein serine/arginine-rich 22
													200795	Aldo/keto reductase
													2429	Heat shock protein 90
													32561	Caffeic acid O-methyltransferase
													4190	Probable fibrillarin
													8462	Metal-dependent phosphohydrolase
													1416	Myosin_tail_1 multi-domain protein
													29907	Ribosomal_L41 domain containing protein
Ppi-sRNA-novel6 TGGCGCAGACGAGATGACACTG-novel-63	-3,43	-4,54	3	0	9	25	29	0,50	0	1,31	4,74	6,05	36304	Unassigned protein
													18214	Unassigned protein
													33876	Unassigned protein
Ppi-sRNA-novel7 ATCCTTCTGAAAGCTTGGCCG-novel-115	0	-2,82	8	10	19	27	51	1,33	2,17	2,76	5,12	10,65	964	DRE-binding protein
													31238	Class IV chitinase
													5990	Gag_spuma domain containing protein
Ppi-sRNA-novel8 GGCGCAGACGAGATGACACTGT-novel-85	0	-4,21	4	2	28	12	45	0,67	0,4347	4,06	2,28	9,39		
Ppi-sRNA-novel9 TTCGAGAAGAGATGACACGATC-novel-136	0	-3,57	3	11	21	35	69	0,50	2,3906	3,047	6,639	14,40	33964	Unassigned protein

Note: Predicted Target(s) column numbers are in the format: sp_v3.0_unigeneN, with N the numbers in this column; NA – not attributed; Differential expression analysis was performed using EdgeR for targets predicted with CleaveLand4 for eighteen differentially expressed small RNAs annotated by miRPursuit workflow.

Appendix G – Oligonucleotides used in gene expression analysis using qPCR.

(next page)

Table G1 – List of primers used for gene expression analysis by qPCR.

	Gene symbol	Gene name	Primer sequence (5' – 3') forward/reverse	Amplicon length	Tm (°C)	PCR efficiency	Target Transcript	Reference
Reference Genes	<i>18S</i>	18S rRNA	ATAAACGATGCCGACCAG/ CACCCATAGAATCAAGAAAGAG	168 bp	57,59 56,71	80%	AH00128.1*	
	<i>Act</i>	Actin	TCCTTCCATCGTCCACAG/ TAGCACTATTGCCATCATCTC	111 bp	58,74 57,80	83%	AF085331.1*	
	<i>EF1-α</i>	Elongation factor 1α	TGAAAGATCCACCAACCTG/ CACAGTTCCAATACCACCA	144 bp	57,41 57,41	-	KM496532.1*	
	<i>α-Tub</i>	α-Tubulin	TCAACACTTTCTTCAGCGA/ GAACCAAGACCAGAACCAG	194 bp	57,38 58,13	82%	Sp_v3.0_unigene146249†	
	<i>His3</i>	Histone H3	TACAGACTCTACCCAAGCC/ AACAAGCCACTGAAACGG	133 bp	58,17 58,61	94%	Sp_v3.0_unigene1183†	
	<i>40S</i>	40S rRNA	TCTTGAGAGTGGAGAATGGG/ CGCATCAGTCATACTCACCT		58,91 59,73	86%		[1]
Candidate Genes	<i>TerpMet</i>	Terpenoid metabolism gene	TCCTGATCGCTTTCATCCTT/ AGATGGTTCATGGGGAAC TG		59,28 59,86			[2]
	<i>PAR1</i>	Photoassimilate-responsive protein 1	AGTCACAACAATTTCTCCCA/ ATGCTAACACCTCTTATCTACC	198 bp	57,43 57,27	92%	Sp_v3.0_unigene677†	
	<i>ABA/WDS</i>	ABA/WDS-induced protein	ACCACCATCTGTCCACCA/ TAATCCACCCGTTTCCTCCA	130 bp	61,26 61,05	101%	Sp_v3.0_unigene20690†	
	<i>LEA</i>	Late embryogenesis abundant protein	TCCATGAACACCATCAAGAACAC/ CAAATATCCAGCCAAACCCT	194 bp	61,58 60,78	ND	Sp_v3.0_unigene682†	
	<i>FMN reductase</i>	Flavin mononucleotide reductase	AGGTTCCGGAACACTTCCT/ CAATTGCTGAGTTCGCCATA		61,56 59,09	103%		[2]
	<i>SNARE</i>	SNAP (Soluble NSF Attachment Protein) Receptor	GGGTGGGCTCTTTGGATAAT/ TTAACTGCAACCCGTTTCC		59,94 59,44	ND		[2]
	<i>DAHPh synthetase</i>	3-Deoxy-D-arabinoheptulosonate 7-phosphate synthetase	CACAATTAAGGCACCATCAG/ GTCACCACTTACCTCCA	170 bp	57,23 57,59	95%	Sp_v3.0_unigene8634†	
	<i>RicinB</i>	Ricin B-related lectin	GCAGCCAAGAAAACTCTGG/ ATTGGGTGCTTCACAAGGAG		60,16 60,74	107%		[2]

Notes:

* GenBank Accession Number

† SustainPine DB unigene name

ND – not determined

[1] (Pascual *et al.*, 2015); [2] (Santos *et al.*, 2012)

Table G1 (continued)

Gene symbol	Gene name	Primer sequence (5' – 3') forward/reverse	Amplicon length	T _m (°C)	PCR efficiency	Target Transcript	Reference
<i>TSPAN</i>	Tetraspanin family protein	GAATAGCCACACCAACTTTAGAG/ TCACAACATTACAAAATCCCTG	193 bp	59,57 59,61	107%	sp_v3.0_unigene20381 [†]	
<i>ChitIV</i>	Class IV chitinase	AGTGCAACGGTGAAATAGTG/ GCAGGAGATGTTGGTTCCC	100 bp	61,38 60,85	95%	sp_v3.0_unigene31238 [†]	
<i>UA</i>	Unassigned protein	TGAGGCCAGAGATTATGAGGA/ TGTTCCATGCACAGATTACGA	278 bp	60,29 60,22	ND	sp_v3.0_unigene33964 [†]	
<i>DREB1</i>	DRE-binding protein DREB1	CTGATAGTAGTGATGATTCCACCA/ AATCCCAGTTAATGTCCATGTCC	206 bp	59,94 60,90	88%	sp_v3.0_unigene964 [†]	
<i>eIF2_γ</i>	Translation initiation factor eIF-2 gamma subunit	GCTCCAACCTACTTCTCCAG/ CCCTATCCACATTGTCCCT	218 bp	58,08 58,53	96%	sp_v3.0_unigene22053 [†]	
<i>Met</i>	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	AGACCAGGAAGTACACCGAG/ CCAAAGCCTGCCTTGAGAG	226 bp	61,10 61,15	92%	sp_v3.0_unigene126719 [†]	
<i>PGK</i>	Phosphoglycerate kinase	TATTTGTGGAGGTTCCGA/ CTATCCATTGTGACTGTCATCC	282 bp	58,38 58,76	ND	sp_v3.0_unigene93887 [†]	
<i>DNMT1-RFD</i>	DNMT1-RFD multi-domain protein	GATATGCCTCGCTCATTACTG/ TCCAACCCGATGAACTCTG	294 bp	58,87 59,49	92%	sp_v3.0_unigene28175 [†]	
<i>Gag-spuma</i>	Gag-spuma domain containing protein	TTCTTCCCTTCTTGCTTTCTG/ AACAATCCCACCTCTCCTC	300 bp	58,62 58,77	97%	sp_v3.0_unigene5990 [†]	

Notes:

[†] SustainPine DB unigene name

ND – not determined

Small RNA targets

Appendix H – Example of a melting curve obtained in real-time quantitative PCR experiments and agarose gel electrophoresis of PCR products.

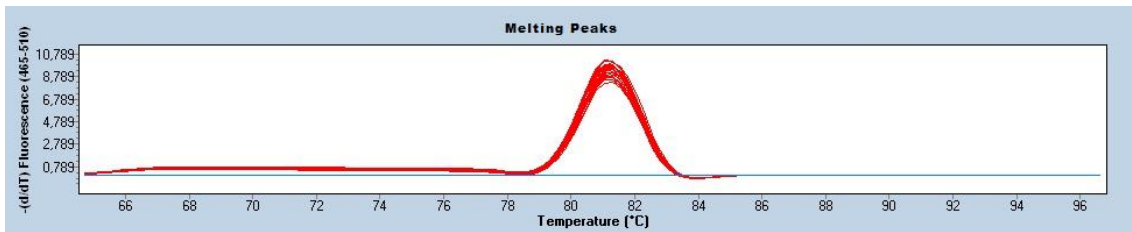


Figure H1 – Example of melting curves obtained in expression analysis of *DAHP synthetase* in needle samples from *Pinus pinaster* plants (n=9) with different behaviors towards PWN infection, using real-time quantitative PCR. To determine the amplification specificity of the primers under the experimental conditions, a melting curve program – 65 to 97°C with a heating rate of 0.1°C per second and a continuous fluorescence measurement – was performed. This graph was obtained using the LightCycler® 480 software v1.5.1.62 (Roche Diagnostics).

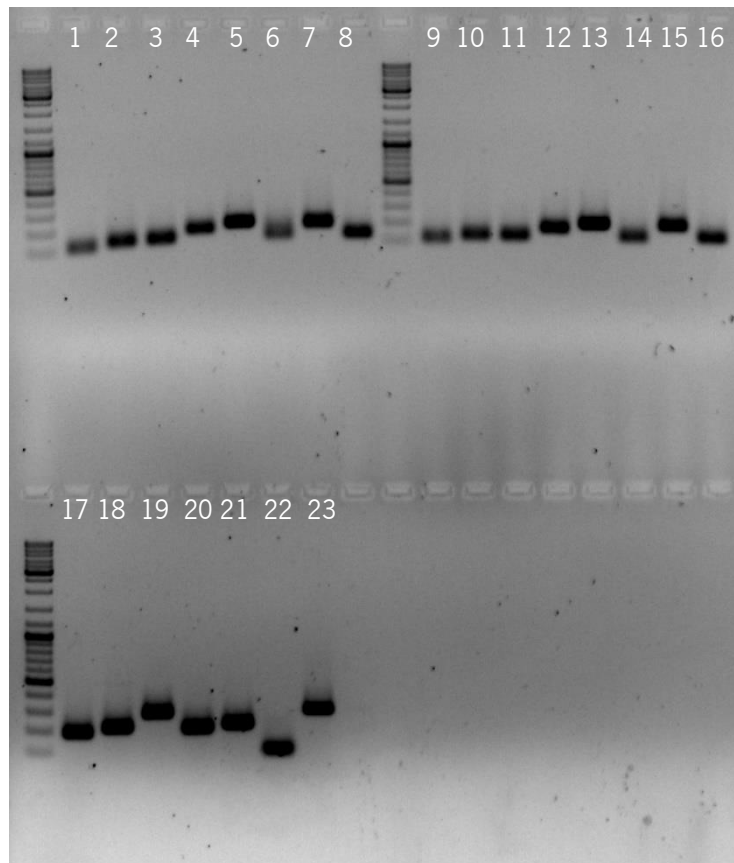


Figure H2 – Agarose gel electrophoresis of real-time quantitative PCR amplification products for all the genes used in this study to confirm amplification specificity of the primer pairs. From lane 1 to 8 – needle samples; from lane 9 to 23 – stem samples. Top lanes – reference and candidate genes (1/9 – *Actin*; 2/10 – *40S rRNA*; 3/11 – *Histone H3*; 4/12 – *DAHP synthetase*; 5/13 – *FMN reductase*; 6/14 – *ABA/WDS-induced protein*; 7/15 – *PAR1*; 8/16 – *Ricini B-related lectin*). Down lanes – small RNA target transcripts: (17 – *Tetraspanin*; 18 – *elf2-gamma*; 19 – *DNMT1-RFD*; 20 – *DREB1*; 21 – *Met*; 22 – *Chitinase IV*; 23 – *Gag-spuma*). GelRed dye (Biotium) was added to each aliquot of PCR reaction product, which was run on an 1.5% (w/v) agarose gel in 1x TE, at about 80 V/cm. Bands were visualized using UV transilluminator-based Gel Doc XR+ (Bio-Rad). Molecular Weight Marker – GeneRuler 1 kb Plus DNA Ladder (Thermo Fisher Scientific).