CATÓLICA

UNIVERSIDADE CATÓLICA PORTUGUESA | PORTO

Escola Superior de Biotecnologia

IDENTIFICATION AND FUNCTIONAL ANALYSIS OF NEMATODE RESISTANCE GENES

by

Ana Isabel Paulino da Silva

October 2012

CATÓLICA

UNIVERSIDADE CATÓLICA PORTUGUESA | PORTO

Escola Superior de Biotecnologia

IDENTIFICATION AND FUNCTIONAL ANALYSIS OF NEMATODE RESISTANCE GENES

Thesis presented to *Escola Superior de Biotecnologia* of the *Universidade Católica Portuguesa* to fulfill the requirements of Master of Science degree in Bioinformática

by

Ana Isabel Paulino da Silva

Place: Escola Superior de Biotecnologia da Universidade Católica Portuguesa

Supervision: Drª Marta Wilton Vasconcelos

Dr. Miguel Monsanto

October 2012

Abstract

Pine wilt disease (PWD), caused by the pinewood nematode (PWN; *Bursaphelenchus xylophilus*), damages and kills pine trees and is causing serious economic damage worldwide. Although the ecological mechanism of infestation is well described, the plant's molecular response to the pathogen is not well known. This is due mainly to the lack of genomic information and the complexity of the disease. High throughput sequencing is now an efficient approach for detecting the expression of genes in non-model organisms, thus providing valuable information in spite of the lack of the genome sequence. In an attempt to unravel genes potentially involved in the pine defense against hereby report the high throughput comparative sequence analysis of infested and non-infested stems of *Pinus pinaster* (very susceptible to PWN) and *Pinus pinea* (less susceptible to PWN).

High throughput sequencing allowed the identification of several candidate genes that may be involved in the response to the PWN. With regards to the gene function most commonly identified, the majority of the sequence functions were associated with protein metabolism and carbohydrate metabolism. However, a significant fraction of sequences associated with RNA metabolism were also highly represented. The sequences that were more commonly found in *Pinus* pinaster were transcription repressors and a translation machinery component: aminoacyl-tRNA synthetase. The cellulose synthase is also important in the disease response, as this gene was up-regulated in infested *Pinus pinaster*. KEGG analysis revealed that the pathway more commonly found in this study were the pentose pathway, the pathway for glucuronate interconversion, the pathway for phenylanine metabolism, amino acid, sugar and nucleotide metabolism, phenylppropanoid biosynthesis, methane metabolism, and citrate cycle (TCA cycle).

Resumo

A doença da madeira do pinheiro provocada pelo nemátodo do pinheiro (PWN; *Bursaphelenchus xylophilus*), provoca danos irreversíveis matando pinheiros e causando graves prejuízos económicos. Embora o mecanismo de infecção seja bem descrito, a resposta molecular da planta para o patogénico não é bem conhecida. Isto deve-se principalmente à falta de informação genómica e à complexidade da doença. A sequenciação de alta capacidade é atualmente uma rota eficiente para a detecção de genes de expressão em organismos não modelos, fornecendo assim informação valiosa. Na tentativa de descobrir genes potencialmente envolvidos na defesa do pinheiro ao agente patogénico, foi realizada a análise de transcriptómica total das sequências de amostras infetadas e não infetadas do caule de *Pinus pinaster* (muito susceptível ao nemátodo do pinheiro) e *Pinus pinea* (menos susceptíveis ao nemátodo do pinheiro), e comparado o seu perfil ao nível da transcrição.

A pirosequenciação permitiu a identificação de diversos genes candidatos que poderão estar associados à resposta ao NMP. No que respeita à função do gene mais predominantemente identificado foi a função associada com o metabolismo de proteínas e metabolismo de hidratos de carbono. No entanto, uma fracção significativa de sequências associadas com o metabolismo de RNA foram, também altamente representadas. As sequências que foram mais comumente encontradas em *Pinus pinaster* foram repressores de transcrição e um componente de tradução: aminoacil-tRNA sintetase. A celulose sintetase é também importante na resposta da doença, uma vez que, este gene foi sobre regulado em infestado *Pinus pinaster*. A análise de KEGG revela que as vias metabólicas mais comumente representadas neste estudo estão relacionadas com a via das pentoses, com as interconversões do glucoronato, as vias do metabolismo da fenilalanina, do metabolismo dos aminoácidos e dos açúcares, biossíntese dos fenilpropanóides, metabolismo do metano e o ciclo do citrato (ácido cítrico).

Agradecimentos

Agradeço à minha orientadora, Drª. Marta Vasconcelos, pela oportunidade de trabalho, apoio, partilha do saber e as valiosas contribuições para o trabalho e dedicação.

Ao Dr. Miguel Monsanto por ter prescindido de algum do seu tempo para auxílio e pelo enorme interesse e disposição em colaborar sempre que solicitada a sua ajuda.

Aos meus pais e ao meu namorado pelo incentivo e apoio recebido ao longo destes anos de estudo.

À Carla Santos e ao grupo PlanTech pelas sugestões ao longo do trabalho.

O meu profundo e sentido agradecimento a todas as pessoas que contribuíram para a realização deste trabalho, estimulando-me intelectual e emocionalmente.

A todas as pessoas que, direta ou indiretamente, contribuíram para a execução desta Tese de Mestrado.

A todos, obrigada pela oportunidade de aprender.

List of figures

List of tables

Abbreviations

Acetil-coA – acetil coenzima A

BLAST – Basic Local Aligment Search Tool

B2G – Blast2Go EST – Expressed sequence tag

cDNA – Complementary Deoxyribonucleic acid

DNA - Deoxyribonucleic acid

EC – Enzyme Code

EST - Expressed sequence tag

GO – Gene Ontology

KEGG – Kyoto Encyclopedia of Genes and Genomes

NGS – Next-generation sequencing

PCR – Polymerase chain reaction

PWD - Pine wilt disease

PWN - Pine wood nematode

qRT-PCR - Real Time quantitative Reverse Transcription PCR

RNA - Ribonucleic acid

RT-PCR - Reverse transcription polymerase chain reaction

RS - repetitive sequences

Contents

1. Introduction

1.1 The host tree and the disease agent

Pines are trees of medium size (20-25 meters) belonging to the family *Pinaceae*. The trunk is covered with a thick shell and the leaves have the shape of needles. Pines are very important because from them a wide variety of products can be obtained, such as lumber for construction, resin and many other important economic products (Mota *et al*., 1999).

In the last decades pines have been dying because of an invasive pest that causes pine wilt disease (PWD) and that can destroy adult pine trees in a few months. The causal agent of this affliction is the pine wood nematode (PWN), *Bursaphelenchus xylophilus*, a nematode that feeds on fungi, plants, or both (Figure 1).

Nematodes are simple worms with microscopic dimensions, normally not exceeding 1,5 mm in length, and have an elongated cylindrical body. The epidermis of a nematode is uncommon, in contrast to other animals, it is not composed of cells but of a mass of cellular material and nuclei without separate membranes that confers resistance and flexibility. Nematodes have complete digestive systems, no circulatory or respiratory systems have only sexual reproduction, and females are always larger than males. These worms can be free-living predators or parasites and are found in many habitats from land to sea and even in freshwater (Oliveira *et al.,* 2008; Autoridade Florestal Nacional, 2012).

Many nematodes are able to suspend their life processes when conditions become unfavorable; these states are resistant to drying, can survive extreme heat or cold, and then come back to life when conditions are favorable for their return.

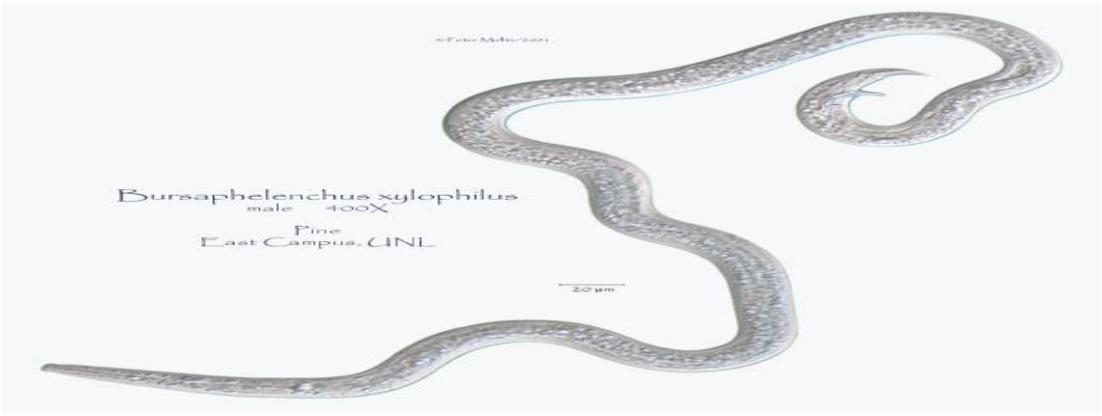

Figure 1: *Bursaphelenchus xylophilus*, the pine wood nematode, taken from http://nematode.unl.edu/bursaphelenchusxylo.htm.

The pine wood nematode is transmitted to the trees by an insect vector, a longicorn beetle whose scientific name is *Monochamus galloprovincialis* (Figure 2).

This vector has one generation per year in Portugal and on average each female lays about 67 eggs during their lifetime. From the middle of April the adult insect feeds on the bark and phloem of the tree branches, transmitting the nematode which is housed in the trachea of the insect facilitating this way the infection in the pine. The propagation of the nematode is limited to the time of flight of the insect from April to October and, due to high temperatures, the disease is further spread in the summer (Autoridade Florestal Nacional, 2012).

Figure 2: Insect vector *Monochamus galloprovincialis*, taken from http://gallery.new-ecopsychology.org/en/photo/black_pine_sawyer_beetle_(monochamus_galloprovincialis).htm.

Once infested and weakened, the female insect lays eggs on the trunk and branches of the tree at canopy level (Mota *et al*., 1999; Jones *et al*., 2008).

Longicorn larvae that develop in weakened trees become adult insects in the spring of next year, leaving the pines and carrying with them the nematode to a new tree.

Also, the distribution of the nematode within the tree is not homogeneous and infection, which usually propagates from top to bottom, gradually blocks the vascular system which supplies water and nutrients to the branches. Without the ability to transport water and vital nutrients, the tree dies.

The nematode of the genus *Bursaphelenchus* is native to North America causing heavy impact in Japan, China, South Korea and also in Portugal (Fukuda, 1997; Roriz *et al*., 2011).

1.2 The pine wood nematode in Portugal

In Portugal, *B. xylophilus* (Figure 1) was first found in 1999 in the in *P. pinaster* region of the Setubal peninsula. It is believed that they arrived in the country through lots of infested wood from Japan for the construction of Expo 98 (Autoridade Florestal Nacional, 2012).

Pinus spp. are the main hosts of PWN and *P. pinaster* (extremely affected by PWN) and *P. pinea* (be less susceptible) are the predominant pine species occupying Portuguese forests. The reason for this susceptible is unknown but it has been shown that, although PWN can infect and kill *P. pinea*, the disease develops much less quickly than in *P. pinaster* (Roriz *et al.*, 2011).

Many hypotheses have been proposed about the PWN pathogenic mechanism, but it is still not well explained. What is known is that this nematode is transported by longhorn beetles of the *Monochamus* spp., as explained before, that are used as vectors by *Bursaphelenchus*. When the insect vector feeds off of the pine branches, it contributes to the nematode's dispersion as they are deposited in the tree through the beetles' feeding wounds. After infestation, the nematodes move rapidly through the resin canals of the xylem and cortex, feeding off their epithelial cells.

The following figure is an example of a PWN infested tree (Figure 3).

Figure 3: Pine trees affected by the pine wood nematode in Japan (photography gently provided by Dr .Vasconcelos).

In Portugal, the nematode problem is so severe that several studies have been conducted to determine the diversity of species of *Bursaphelenchus* and their distribution in Portugal, as well as to understand the complete etiology of the disease, in order to prevent it's propagation but until now, no effective solution has been found (Jones *et al.*, 2008).

1.3 Symptoms

Initially, the first signs of infection with PWN begin with the wilting of older needles, subsequently passing throughout the tree trunk and reaching the younger needles. This phenomenon is visible by the yellowing of leaves (foliage turns from green to yellowish green, red, and finally brown / yellow) (Fukuda, 1997).

The dead needles tend to remain for prolonged periods of time in the plant and the branches tend to be more brittle. As the disease propagates quickly, and if climate conditions are favorable to the nematodes, an infested tree can die after a few weeks. Also, several weeks after infection, infested trees show a decrease of resin exudation, a decrease in photosynthesis, denaturing of the xylem and parenchyma cells of the cortex (Fukuda, 1997; Wang *et al*., 2009; Roriz *et al*., 2011; Autoridade Florestal Nacional, 2012).

1.4 Prevention and treatment - methods to combat

Unfortunately, this disease has propagated to other parts of the country beyond Setubal peninsula, forcing the European Union to classify Portugal as a restriction zone, affecting the export of wood products, which led to vast losses of forest export products. Since then, this problem led to preventive measures that entail heat treatment of wood at very high monetary costs, which led to high economic losses of many Portuguese companies in the forestry sector. Namely, the furniture industry has started resorting to wood importation, which inevitably led to higher end prices for their products, and reduction of exportations.

The PWD continues to evolve rapidly so there is an urgent need to develop techniques or methods of prevention / control of its propagation (Fukuda, 1997; Oliveira *et al*., 2008; Wang *et al*., 2009).

As infection and propagation develops very rapidly, several strategies have been proposed in order to contain the disease:

- Treatment of wood cuttings with chemical applications: the affected trees are felled and the branches and trunks are treated with insecticides. This technique is not very effective since it is necessary to apply the insecticides several times during the extended period of insect vector flight, and leads to adverse side effects, such as environmental pollution;

- Bacterial symbiosis: it is thought that certain bacteria can enhance or diminish the virulence level of the nematode; this strategy has not been implemented yet.

- Cultivation of resistant species: researchers are screening for naturally resistant pine genotypes, however, to date , no resistant *P. pinaster* tree-has been identified;

- Biological control of the insect vector: it is a way to control or manage pests using, for example, natural enemies to reduce the population of insect vectors; this strategy has also not been implemented, as there are no readily identified natural enemies for *M. galloprovincialis;*

- Crushing and burning of infested trees: this process should be conducted before the beginning of the insect flight in May, during which time the population of insect

vector is within the host- this is the most effective mean to reduce the population of insect and consequently, to avoid the propagation of the disease;

- Application of nematicides in the tree: certain effective nematicides have been identified for *B. xylophilus*, however this strategy is expensive and laborious, only being feasible in protecting a limited number of high value trees. An example of a successful nematicide developed against *B. xylophilus* is Pursue of Syngenta;

- A genetic approach: the study of genes that allow resistance to the disease, and use these resistance genes in programs of genetic transformation to create resistant *P. pinaster* trees;

1.5 Sequencing and transcriptomic analysis

The genome is composed of deoxyribonucleic acid (DNA), the molecule that contains the genetic information in the cells. DNA must be transcribed into ribonucleic acid (RNA), in order for protein synthesis to occur (Bai *et al*., 2010).

A transcriptome is a set of all the transcripts present in a cell and represents the very small percentage of the genome that is transcribed into RNA molecules, but as one gene may produce many different mRNA molecules, a transcriptome is much more complex than the genome that encodes it (Yazdi *et al*., 2011).

Analysis of the transcriptome is useful because it allows determining when and where each gene is turned “on” or “off” in the cells and tissues of an organism, it counts the number of transcripts and determines gene expression levels.

By comparing transcriptomes of different cells types one can understand faster the specific constitution of each cell, how that type of cell normally functions, and how changes in the normal level of gene activity may reflect or contribute to a specific disease (Yazdi *et al*., 2011).

In order to study the genomes that are expressed and involved in the disease response, several strategies can be utilized such as RT-PCR, QRT-PCR, suppressive subtraction hybridization, or high throughput sequencing (HTC), the later being the technique that provides the largest amount of transcript information. Different sequencing and transcriptomic technologies, have given us specific insights

regarding the pine genome and its response to biotic and abiotic stresses. A few examples include: 1) single nucleotide polymorphism genotyped using GoldenGate assay, where a consensus map was created for maritime pine (Chancerel *et al.*, 2012); 2) microarray technology, that identified 2,445 differentially expressed genes that were responsive to severe drought stress in roots of loblolly pine (Lorenz *et al.*, 2011); 3) LongSAGE technique, that provided a total of 20,818 tags, from which 38 were differentially expressed in the resistant Japanese black pine and 25 in non-resistant pine (Nose and Shiraishi, 2011); 4) and suppression subtractive hybridization, showing the up-regulation of stress response and defense related genes by pine wood nematode infestation (Hirao *et al.*, 2012; Santos et al, 2012).

High throughput sequencing

DNA sequencing, as the name indicates, is a technique that determines the sequence of DNA, a process that determines the order of nucleotides in a DNA sample for further analysis at both structural and functional level (Ronaghi, 2001).

Sequencing is important because it allows for a complete description of the molecular composition of organisms intended to study, once all required information is present in the genomic DNA (Ronaghi, 2001).

Currently there are two widely used methods for high throughput sequencing: the sequencing method of Sanger (Horner *et al*., 2009) and the 454-pyrosequencing method (Vera *et al*., 2008).

In this work the method used was 454-pyrosequencing, an approach to sequencing based on the detection of pyrophosphate, which makes use of a technique capable of capturing the light emission caused by the addition of a luciferase, coupled to polymerization of DNA previously fragmented and attached to microspheres, with the aid of adapter sequences (Ronaghi, 2001).

The pyrosequencing method comprises a series of steps which will be described briefly. First, a sequencing primer is hybridized to a single-stranded DNA resulting from polymerase chain reaction (PCR) amplification, and is incubated with the

enzyme DNA polymerase, ATP sulfurylase, luciferase and apyrase and the substrates with adenosine 5'phosphosulfate luciferin (Vera *et al*., 2008).

The first of the four deoxynucleotide triphosphates (dNTPs) is added to the reaction. DNA polymerase catalyzes the incorporation of dNTP in the DNA chain if there is an existing complementary base to the template strand. Each time a DNTP is incorporated, there is a release of pyrophosphate in an amount equimolar to the amount of incorporated nucleotide.

The ATP sulfurylase is quantitatively converted to ATP pyrophosphate. The ATP provides energy for the luciferase to oxidize luciferin and generate light. The light is detected by a camera with a device which allows seeing a peak at the pyrogram. Each light signal is proportional to the number of nucleotides incorporated. Addition of dNTPs is carried out sequentially (Huse *et al*., 2007). As the process continues, the complementary DNA chain is constructed and the nucleotide sequence is determined by the peaks generated in a pyrogram.

High throughput sequencing generates random tags that need to be assembled in order to produce the full length mRNA sequences found in the cell.

Genome assembly is complicated because it is difficult to build a general assembly that is able to reconstruct the original sequence with high precision, especially when working with non-model organisms whose genome has not yet been sequenced.

There are factors that can lead to errors in assembly; in particular, contamination of genetic material, density of the reads, the repetitive consensus sequences (SR), among others. After assembly of the sequences small sections will be generated, called contigs (Figure 4). Contigs are referred to as a set of overlapping DNA segments that together constitute a region of consensus DNA (Zhang *et al*., 1994).

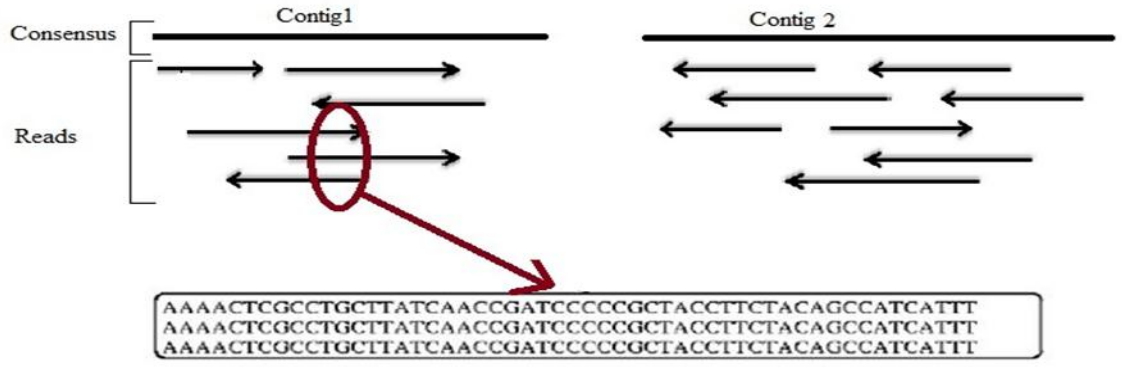

Figure 4: Illustration of the contig assembly process, a required step in transcriptome sequencing.

1.6 Objectives

Given that *P. pinaster* and *P. pinea* have contrasting susceptibility to the PWN, the former being very susceptible and the latter being much less susceptible, it is important to determine if this difference in susceptibility is related with differential expression of specific genes.

This study aims to identify resistance genes to the PWN, using two interlinked approaches: 1) conduct bioinformatic analysis of high throughput sequencing data obtained from 454-pyrosequecing of nematode infested and non-infested *P. pinaster* and *P. pinea*; and 2) to conduct functional analyses of this data, by comparing the two sets of transcriptomes and identifying which are the differentially expressed genes.

The analysis of these expressed sequence tags is useful because: 1) the sequences will be made available to the scientific community via public repository databases, contributing to the genomic information available to plant researchers working on transcriptomics and genomics of pine; 2) it will allow the discovery of genes that can underlie resistance to the nematodes, and 3) in the near future, via a collaborative project that is under way, create transgenic *P. pinaster* tree that can resist infestation by the nematode, thus possibly preventing the propagation of the disease.

2. Material and Methods

As previously mentioned, the main objective of this study is to identify genes involved in the plant defense against the PWN.

It is known from previous studies, that nematodes can infect many species of pine, but that the most common and most susceptible host in Portugal is *P. pinaster* causing irreversible tree damage and death (Yan *et al*., 2012).

In this study we used two types of pine: *P. pinaster* (susceptible host), and *P. pinea* (less susceptible host) in order to identify potential gene sequences that may explain why the nematode causes death in *P. pinaster* whereas *P. pinea* survives.

The current study is a continuation of work that began in January of 2009, and the biological material had already been sent for 454 high-throughput sequencing at BIOCANT (Cantanhede, Portugal). The main goal of the current study was to perform the bioinformatic analysis of the data obtained by pyrosequencing. In order to better contextualize the project, a brief summary of the laboratory work that was previously performed will be described below.

2.1 Plant material and nematode culture

Twenty-eight potted 2-year-old (fourteen *P. pinaster* and fourteen *P. pinea*) trees were used in this study, kept in a climate chamber (Aralab Fitoclima 10000EHF), with relative humidity of 80% and with a photoperiod of 16h day (with photosynthetic active radiation of 490 µmol m^{-2} s^{-1} and temperature of (24-26 ºC) and 8h night (with temperatures of 19-20 ºC). Plants were watered every 2 days.

Small, square pieces of Potato Dextrose Agar with *Botrytis cinerea*, grown at 26ºC for 7 days, were transferred to test tubes with barley grains previously autoclaved. *B. xylophilus* geographical isolate HF (from Setubal Region, Portugal) was cultured on small squared potato dextrose agar, previously covered with *B. cinerea* mycelium for 7 days at 26°C, placed in test tubes and incubated at 26°C. The multiplied nematodes were extracted using the Baermann funnel technique prior to

inoculation. Only nematodes that had been extracted for less than 2 hours were used in the subsequent experiments.

2.2 PWN inoculation and sampling time

The twenty-eight plants were divided in four groups and were inoculated following the method of Futai and Furuno (Futai *et al.*, 1979; Futai, 2003). In brief, a suspension of 1,000 nematodes was pipetted into a small 3-5 cm long longitudinal wound, about 40 cm above soil level. The inoculated wounds were covered with parafilm to prevent drying of the inoculum. The same conditions were applied to the control plants, inoculated with sterile water. Twenty-four hours after inoculation the entire pine tree stem was cut into small pieces and stored at -80 ºC until further analysis.

2.3 RNA extraction and cDNA synthesis

Four treatments were studied: *P. pinaster* and *P. pinea* inoculated with *B. xylophilus* strain HF and inoculated with water, as control. A pool of the seven plants from each treatment was made and total RNA was extracted. The extraction was performed according to an optimized method from Provost (Santos and Vasconcelos, 2012) and the samples were stored at -80 ºC. RNA integrity and purity was checked by UV-spectrophotometry using a nanophotometer (Implen, Isaza, Portugal) and by fluorimetry.

2.4 cDNA library construction and pyrosequencing

The total mRNA quality was verified on Agilent 2100 Bioanalyzer with the RNA 6000 Pico kit (Agilent Technologies, Waldbronn, Germany) and the quantity assessed by fluorimetry with the Quant-iT RiboGreen RNA kit (Invitrogen, CA, USA).

A fraction of 1-2 micrograms of total RNA was used as starting material for cDNA synthesis with the MINT cDNA synthesis kit (Evrogen, Moscow, Russia), a strategy based on the SMART double-stranded cDNA synthesis methodology with amplification of polyA mRNA molecules using a modified template-switching approach that allows the introduction of known adapter sequences to both ends of the first-strand cDNA. The synthesis was done with a modified oligodT containing a restriction site for *Bsg*I. After synthesis, the polyA tails were removed through restriction enzyme digestion to tails and, in that way, minimize the interference of homopolymers during the 454-pyrosequencing run.

Five hundred nanograms of non-normalized cDNA, quantified by fluorescence, were sequenced in a full plate of 454 GS FLX Titanium according to the standard manufacturer's instructions (Roche-454 Life Sciences, Brandford, CT, USA) at Biocant (Cantanhede, Portugal).

The transcriptome sequences of these four samples were made available on a website with the address http://transcriptomics.biocant.pt/pine/. This site has restricted access and can only be accessed by providing a password.

2.5 Software's utilized in the bioinformatics analysis

Myrna - differential gene expression for RNA-seq

Myrna is a computing tool for calculating differential gene expression and can be accessed from the following address: http://bowtie-bio.sourceforge.net/myrna. This tool calculates coverage for exons, genes, or coding regions and differential expression using either parametric or non-parametric permutation tests and integrates short read alignment with interval calculations, normalization, aggregation, statistical modeling in a single computational pipeline. The results are presented in the form of gene p-values and q-values for differential expression (Langmead *et al*., 2010). Myrna is a very useful tool that allows one to get the p-values and q-values for analyses. In other words, Myrna is a computing tool that calculates differential gene expression in large RNASeq datasets. In this study, Myrna was useful because it provided the outputs containing the p-values to assign a confidence level of the

differentially expressed genes with the number of reads in every sample that contributed to the construction of the genes in question.

MG-RAST (Metagenomics Rapid Annotation using Subsystem Technology)

In order to study the different taxonomies and functional analysis for nucleotide sequences of the four samples, MG-RAST software was utilized. This software can be accessed from the following address http://metagenomics.anl.gov/ (Meyer *et al*., 2008).

The criteria for the selection of this software was the familiarity with its favorable features and capabilities. The MG-RAST server is a very useful system in bioinformatics since it is an "open source" tool for annotation and comparative analysis of metagenomes (Meyer *et al*., 2008).

Figure 5 is shows of the home interface of the software.

Figure 5: Initial interface of MG-RAST (http://metagenomics.anl.gov/).

The server provides several methods for accessing different types of data, including phylogenetic reconstructions and metabolic ability to compare the metabolism and annotations of one or more metagenomas and genomes. The data entered are mapped against a comprehensive, searchable, non-redundant database (NR- currently M5NR).

Phylogenetic studies and metabolic reconstructions are computed from the set of hits against the NR database. The resulting data are made available for browsing, download, and most importantly, comparison against a comprehensive collection of

public metagenomes. MG-RAST allows analysis for sequence filtration, including both functional classification and functional annotation (Yu and Zhang, 2012). A simple search of similarities (BLAST, for example) allows the user to retrieve similarities with several other databases (NCBI, KEGGs etc.). The MG-RAST software provides automated analyses of phylogenetic context, performing the taxonomic evaluation based on the sequence data submitted.

The selected parameters for the analysis in this study were: maximum e-value cutoff of $1e^{-30}$; minimum percentage identify cutoff of 50%; and minimum alignment length cutoff of 50%. The classification was based in the lowest common ancestor.

Blast2Go

The Blast2Go (B2G) software is a web tool for bioinformatics with Java interface for functional sequence annotation (can write thousands of sequences) and sequence analysis of genes or proteins. The annotation was based on Gene Ontology (GO) and can be accessed through the address http://www.blast2go.com/b2ghome (Conesa and Götz , 2008).

The B2G recently improved the functionality of the annotation process, so that currently this tool supports Gene Ontology, KEGG maps, codes and Enzyme InterPro. In general, B2G BLAST (Basic Local Alignment Search Tool) searches remote to similar sequences from one or several sequences of input and extracts the GO terms associated with each of the results obtained (Conesa and Götz , 2008).

Figure 6 shows the general interface of B2G from where all the available features are accessible.

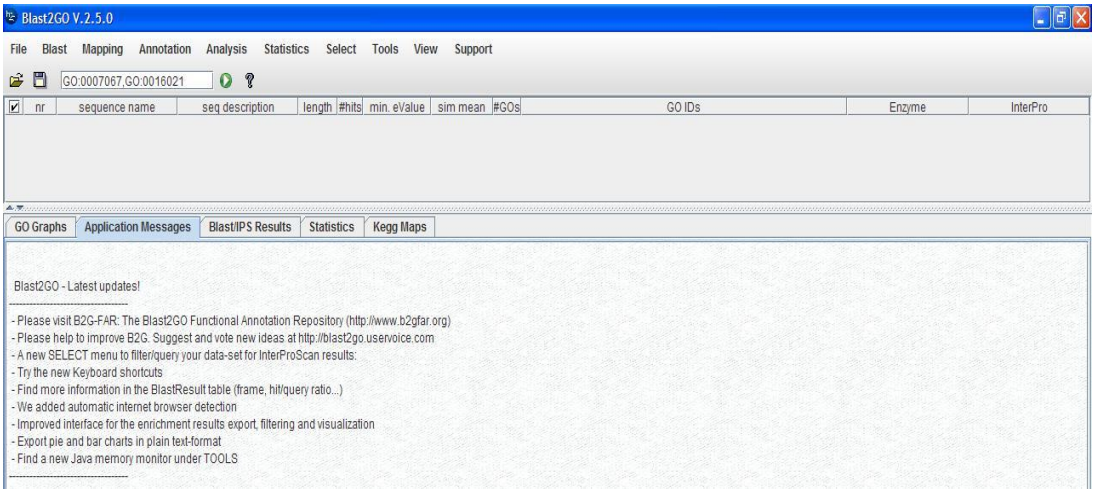

Figure 6: General Blast2GO interface (http://www.blast2go.com/b2ghome).

Initially, the B2G searches similar sequences to a query set by Blast searching, however the annotation will ultimately be based on sequence similarity levels. At the end of the mapping process a set of candidate annotations from different hits of diverse similarity levels and various annotation sources is gathered (Götz *et al*., 2011).

Functional annotation is very useful because it allows characterization of genes in functional classes, understanding physiological meanings and assess functional differences (Conesa *et al*., 2005). With this tool several analisis were made, for example, performing BLAST searches mapping the terms GO Enzyme Code annotation with KEGG maps and Interpro annotation and graphical exploration.

KEGG (Kyoto Encyclopedia of Genes and Genomes)

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a platform that integrates multiple bioinformatic online databases, developed by Kanehisa Laboratories in 1995 and which can be freely accessed via http:// www.genome.jp/kegg/ or http://www.kegg.jp/ (Kanehisa and Gotto, 2000; Kaneshia *et al*., 2006; Kanehisa *et al*., 2012).

KEGG is characterized by being a user-friendly tool, which is composed of three major categories of molecular interaction networks (pathways), biological processes, information about the universe of genes and proteins, and information about a wide range of chemicals and chemical reactions (Figure 7) (Kanehisa and Gotto, 2000; Kanehisa *et al*., 2012).

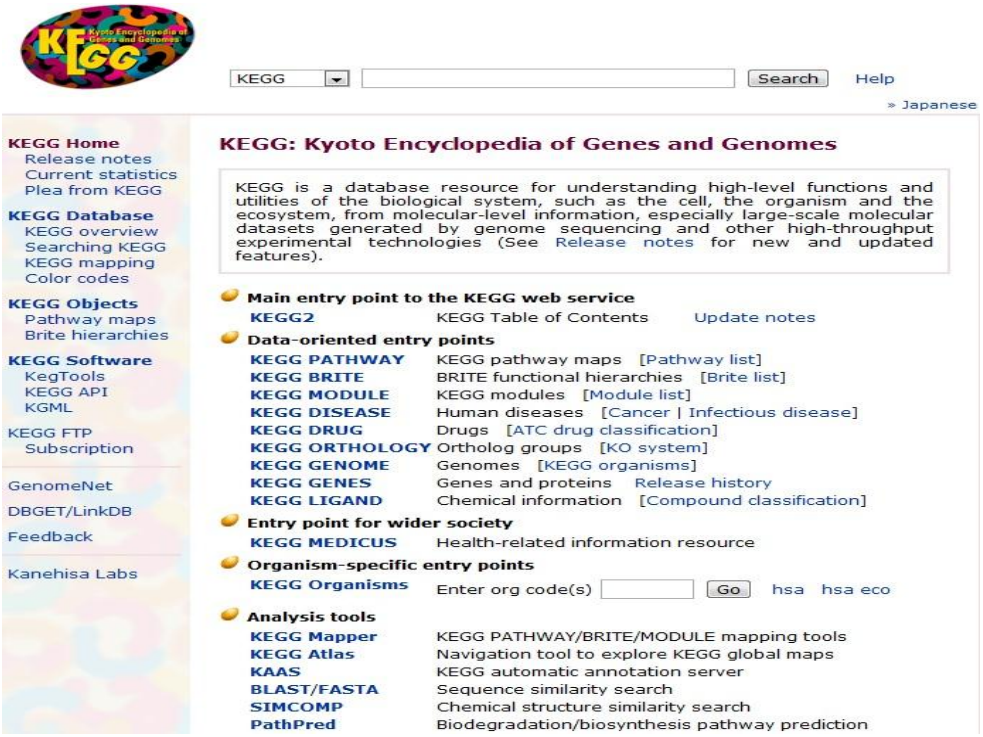

Figure 7: General KEGG Interface (http://www.genome.jp/kegg/).

Basically, this database is a set of sequenced genes and genomes which are registered and connected to networks in cells and molecular interactions in certain organisms (Winter and Huber, 2010). In the current work, KEGG database was used to determine which metabolic pathways were more present in the data, to study

possible relations between genes, and to compare these pathways between the four samples under study.

2.6 Identification of candidate genes associated with resistance to the PWN

In order to identify the differentially expressed genes between the four samples, the pyrosequencing results for the infested samples were pooled with the respective control samples and the expression levels of the latter were subtracted, in order to normalize the infested samples.

An interface was implemented in the constructed site with the obtained sequences, to trim the search in SQL database, using the following algorithm parameters: only sequences with 8 minimum reads were considered and, to ensure the quality of the sequences, the pondered p-value was of $5e^{-05}$. These strict parameters were established to limit the search only to the most represented genes.

After the application of this algorithm, all reads from the same sequences were grouped and the genes with unknown function were removed from the analysis. A ratio between the normalized infested samples was calculated, with which all sequences with a ratio inferior to 1 were excluded and hits with ratios higher than 1 were considered to be overexpressed for the numerator sample.

3. Results

A cDNA library was constructed from RNA of pine stem tissues from *P. pinaster* and *P. pinea* inoculated with *B. xylophilus* and from uninfested controls. Following 454-pyrosequencing, the quality trimming and size selection of the reads were determined by the 454 software after which the SMART adaptor sequences were removed from the reads using a custom script and the poly-A masked using Myrna, to assure correct assembly of raw sequencing reads. All quality reads were subjected to the gsAssembler (v2.6-Roche) assembler (version 3.2.0), with default parameters.

Pyrosequencing of the four cDNA libraries generated a total of 1,393,970 reads, with an average length of 320 bp. Specifically, 450,053 reads differentially expressed by *P. pinaster* infested with nematode, which assembled into 12,157 contigs; 375,168 reads for *P. pinaster* control, assembled into 8,808 contigs; 342,141 reads for *P. pinea* infested with nematode, assembled into 9,555 contigs; and 226,608 quality reads for *P. pinea* control, that were assembled into 4,175 contigs. No singletons were obtained when the samples were compared, and the distribution of contig length and EST assembly by contig, for the four samples.

This data is presented in Table 1.

Table 1 – Summary of assembly and expressed sequence tag data in the four samples

	Infested *P. pinaster*	**Control *P. pinaster***	**Infested *P. pinea***	**Control *P. pinea***
No. of Reads	450,053	375,168	342,141	226,608
Total Bases	145,356,992	121,441,000	111,032,000	70,672,704
Average read length after trim quality	322	323	324	311
No. of contigs	12,157	8,808	9,555	4,175
Average contig length	806	738	783	636
Range contig length	32-3,968	12-4,031	38-4,665	11-2,828
No. of Contigs with 2 reads	8	0	0	0

No. of Contigs with > 2 reads	12,149	8,808	9,555	4,175
Contigs with BLASTx matches (E-value ≤ 10^{-6})	531	422	521	207
*Contigs with BLASTx matches (E-value ≤ 10^{-2})	3,532	2,169	2,339	1,436
Contigs determined by ESTscan	511	435	413	424
Total no. of transcripts	13,003	9,250	9,968	5,516

*Contigs without BLASTx matches at an E-value cut-off of 10^{-6} were queried again with BLASTx with an E-value cut-off of 10^{-2}.

The entire set of reads used for final assembly was submitted to the NCBI Sequence Read Archive under the accession n° SRA050190.1 (Control *P. pinea*), SRA050189.1 (Infested *P. pinea*), SRA050188.1 (Control *P. pinaster*) and SRA050187.2 (Infested *P. pinaster*).

The translation frame of the contigs was determined through queries against the NCBI non redundant protein database using BLASTx with an E-value of 10^{-6} and assessing the best twenty five hits. Contigs without hits were submitted again to BLASTx homology searches against the NCBI nr database with a higher E-value cut- off set at 10^{-2}. Sequences with a translation frame identification derived from the two previous searches were used to establish the preferential codon usage in *P. pinaster* and *P. pinea* based on which the software ESTScan detected further potential transcripts from the two previous sets of sequences with yet no BLASTx matches. This procedure originated a third set of sequences with putative amino-acid translation.

The entire collection of sequences of at least 30 amino-acid long, resulting from the BLASTx and the ESTScan procedures, was processed by InterProScan for the prediction of protein domain signatures and Gene Ontology terms. All the results were compiled into a SQL database developed as an information management system. The distribution of sequences into GO categories was calculated at each level

and were passed to the parent GO at the top of the broad ontology domains, considering that each single assignment into a GO child was only counted once in the total sum. The positive hits were retrieved and translated into the taxon ID using the information provided by NCBI.

3.1 Sample taxonomical classification

In the initial phase of the analysis, the goal was to have the taxonomical categorization of the samples. In other words, understand the taxonomy and domains of the most predominant species in the samples. Since genomic information for the *Pinus* genus is very limited in publicity databases, we expected to obtained homology data to other coniferales, such as the *Picea*, wich has more abundant sequencing data. Thus, a taxonomical evaluation was conduted for the coniferales phyla of the samples, and this information can be observed in Figure 8.

For this, first, sequences of the four samples were added in a single file with extension .txt and submitted to the MG-RAST software. Secondly, the species not belonging to the coniferales were excluded from the analysis, and a second MG-RAST run was performed with the remaining sequences.

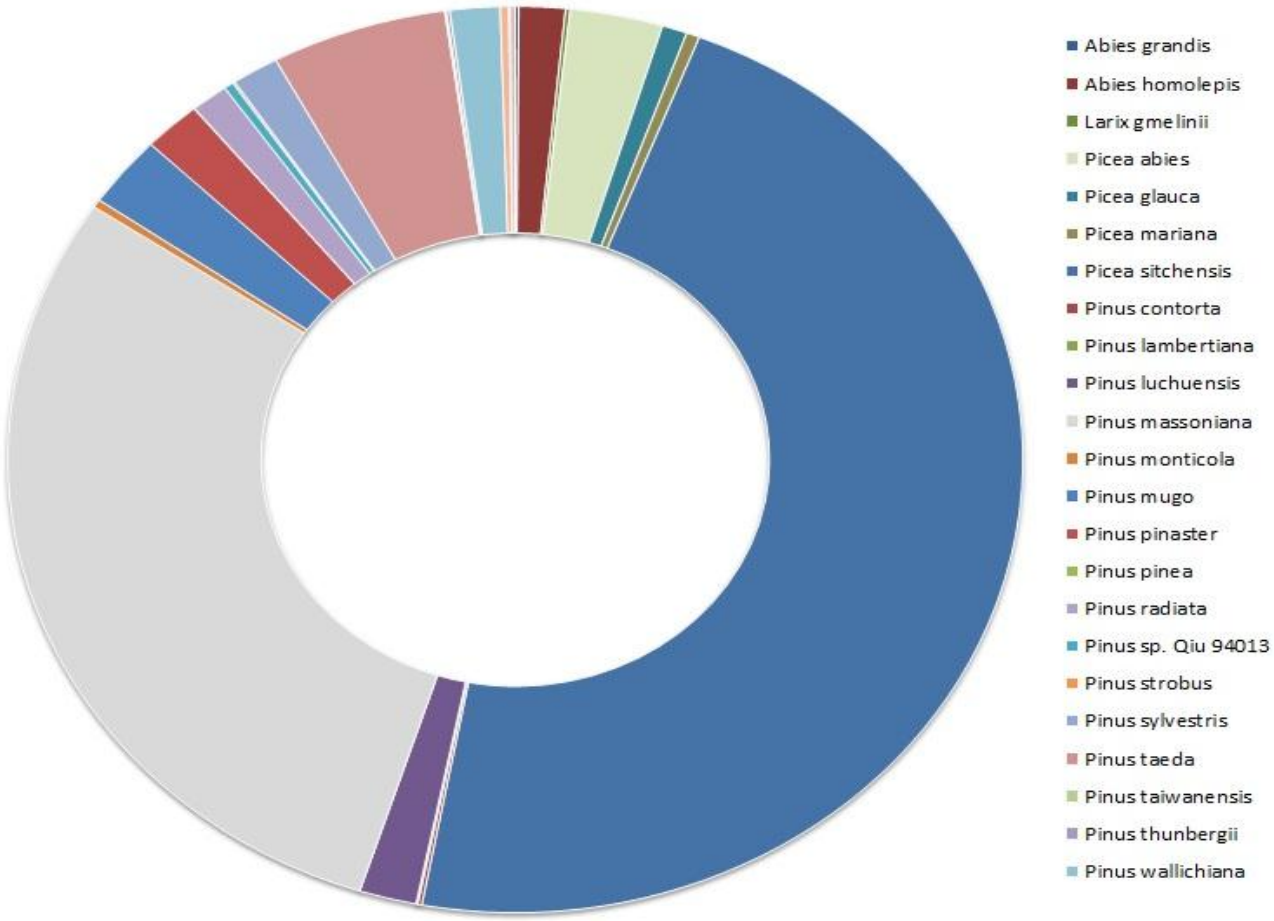

Figure 8: The most prominent genera of conifer plants identified in the samples under study.

In this analysis the most prominent genera of conifer plants identified in the samples under study were *Picea sitchensis* (47.1%), *Pinus massoniana* (29.2%), *Pinus taeda* (6.7%) and *P. pinaster* (2.4%).

Abundance of pine and nematode sequences

Pine

The MG-RAST software was also useful because it allowed a clear determination of the most prominent species in each individual biological sample. This step was necessary due to the fact that the biological samples were, in the case of samples "infested *P. pinaster*" and "infested *P. pinea*", a mixture of plant and nematode samples. It was necessary to confirm that the preponderance of nematode-related

sequences in the samples was minimal, as we were interested in identifying the transcriptome sequences of *P. pinaster* and *P. pinea*. A small contamination from nematodes was expected, due to the fact that the samples were inoculated with *B. xylophilus*. Table 2 shows an overview on the percentage of sequences that belong to each taxonomical category, in each biological sample.

Table 2 - Taxonomic distribution of the assembled data (percentage) 'Not id' represents the percentage of sequences that had hits in databases but could not be identified (unknown sequences)

	Eukaryota			
	Plantae			**Other**
	Pinus spp.	***Picea spp.***	***Not id***	
Infested *P. pinaster*	1.8	39.0	55.7	3.5
Control *P. pinaster*	2.7	37.8	52.6	6,9
Infested *P. pinea*	1.9	39.8	47.4	10.9
Control *P. pinea*	12.8	25.4	52.1	9.7

The plant sequences most commonly identified belonged to the families *Pinaceae*, *Cycadaceae* and *Liliopsidae*. In the case of the *Pinaceae*, the most common genus, as predicted due to the composition of the samples, were *Pinus*, *Picea*, and *Abies* (Table 2). The genus *Picea* was the one that showed the highest number of homologous sequences, ranging from 25.4% in control *P. pinea* to 39.8% in infested *P.pinaster*. Sequences belonging to the genus *Pinus* ranged from 12.8 % in control *P. pinea* to 1.8% in infested *P. pinaster*. Regarding the *Cycadaceae*, the genus *Cycas* was predominant. In the case of the *Liliopsidae,* the most commonly found genus were

sequences belonging to the *Poaceae* and *Oryzae* (data not shown). These are grouped in the category "others" in the plant section of Table 2. This table also shows that the predominance of sequences homologous to nematode sequences was undetectable, with the exception of infested *P. pinaster*, and control *P. pinea* which presented an extremely small percentage of nematode-homologous sequences (0.1 and 0.2%, respectively). Bacterial species were also present, but in percentages ranging from 0.2 to 0.7%.

Functional annotation

To annotate the transcripts, the putative frames were queried against the InterPro database of protein families and functional domains http://www.ebi.ac.uk/InterPro (Ashburner *et al*., 2000; Apweiler *et al*., 2001), and additionally annotated with GO terms, to assign *Pinus* contigs into the major GO annotation categories, namely, Biological Processes, Cellular Components and Molecular Functions in a species-independent manner (Apweiler *et al*., 2001). Within the Biological Process, 29.37% and 49.36% (Annexes, Figure 16, 17) of assignments corresponded to "Cellular Process (GO:0008152) and "Metabolic Process" (GO:0009987) respectively, followed by the "Localization" (GO:0051179, 8.49%) and "Establishment of Localization" (GO:0051234, 8.40%) GO categories. Furthermore, the matches of Molecular Function terms were most prevalent within the "Binding" (GO:0005488, 48.84%) and "Catalytic Activity" (GO:0003824, 36.86%) category, followed by the categories "Structural Molecule Activity" (GO:0005198, 3.52%) and "Transporter Activity" (GO:0005215, 3.62%) (Annexes, Figure 24, 25). Finally, for the Cellular Component GO the most evident matches were within the "Cell Part" (GO:0044464, 34.72%) and "Cell" (GO:0005623, 34.72%) terms, followed by "Organelle" (GO:0043226, 13.33%) and "Macromolecular Complex" (GO:0032991, 10.76%) (Annexes, Figure 18, 19). Together, these GO classes accounted for most of the assignable transcripts, and may represent a general gene expression profile signature for *Pinus* spp.

3.2 General functional classification

An analysis of the general functional classification of the obtained sequences was performed using MG-RAST. For this, all the sequences of the four samples were bulked and analyzed as a whole. Figure 9 shows these functions that were most commonly found in the biological samples under study. The data presented were normalized to values between 0 and 1 to allow comparison of samples of different sizes.

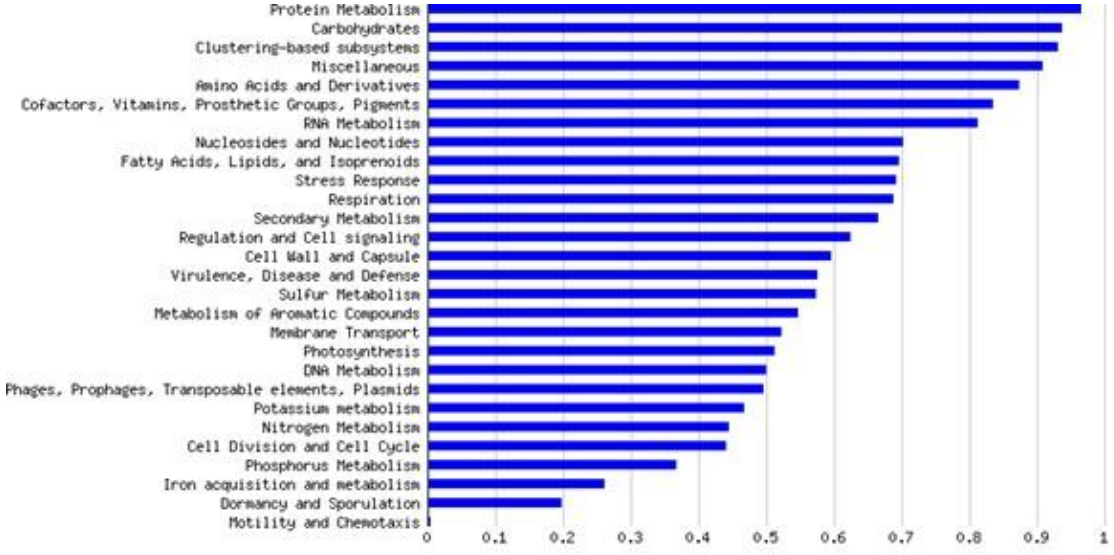

Figure 9: Functional classification of the bulked sequences obtained from the pyrosequencing of the four samples under study. The data was obtained using MG-RAST.

As shown in Figure 9, the most predominant function was related to protein metabolism, followed by carbohydrate metabolism, clustering based subsystems, amino acid and derivatives metabolism, co-factors, vitamins, prosthetic groups, pigments and RNA metabolism. A large fraction of the sequences fell also in the "miscellaneous" category. This category represents sequences whose function could not be identified via homology searches. The category with fewer sequences was of those related to motility and chemotaxis. Another of the most represented categories was related with virulence, disease and defense.

3.3 Sequence analysis for identification of resistance genes

After evaluating the taxonomical composition of the samples and the general functional characterization of the genes, the data was processed for individual sample functional analysis. As mentioned earlier, this study consisted of four samples: infested *P. pinaster*, control *P. pinaster*, infested *P. pinea* and control *P. pinea.* This was necessary in order to compare samples that had the same treatment (inoculated with nematode vs. non-inoculated) and to compare samples of the same species (*P. pinaster* or *P. pinea*). This is useful because, first it allows viewing more genes that are expressed for each sample and secondly, allows comparing the different genes in the two species. After combining the two samples, the genes that were expressed in common in both samples were discounted, and the genes that were differentially expressed could be more easily identified and quantified.

Search Interface

The application of a search interface was necessary to decrease the number of sequences in the final gene list. This interface can be found at the address http://transcriptomics.biocant.pt/ which is password restricted as mentioned above. Figure 10 represents the algorithm used

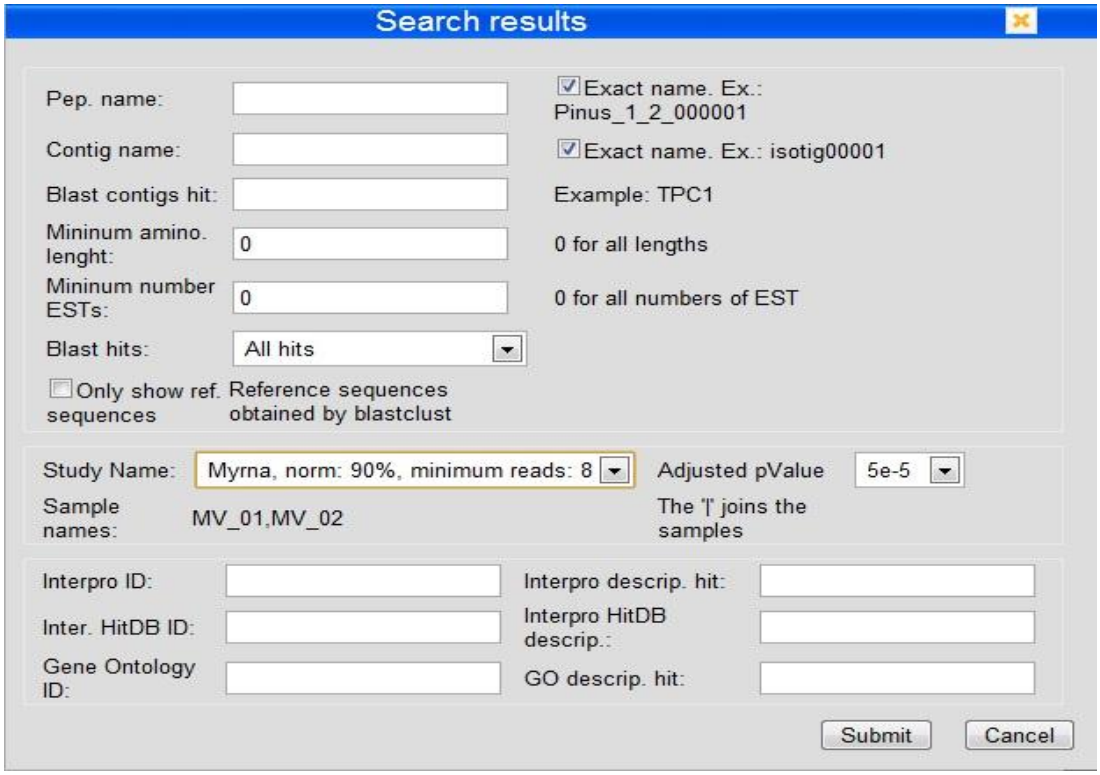

Figure 10: Search interface and algorithm utilized for the gene functional analysis.

An interface was implemented in the constructed site with the obtained sequences, to trim the search in the SQL database, using the following parameters: 1) only sequences with a 8 minimum of reads were considered and, 2) to ensure the quality of the sequences, the selected pondered p-value was $5e^{-05}$. These strict parameters were established to limit the search only to the most represented and more homologous genes. This implies that in the end of the search, e.g., only the genes more expressed when the pine is infested will appear. The end result will be presented in the form of an excel file generated by the interface.

The same analysis was also performed for the sample infested *P. pinea* and control *P. pinea*, in order to compare samples from the same species.

B2G

For the functional annotation and functional analysis of the sequences of interest, the B2G software was used. The system used by this ontology was the Gene Ontology tool.

With this tool, a Blast was carried out for annotation and comparison of local alignments. This displays advantages such as speed and identification of genes. In this study Blast provided the description of the sequence, its length and its GO number.

Then a Mapping of the Gene Ontology terms (GO IDs) was performed. After this, the mapping used Interpro (InterProScan) in order to write down the sequence motives / domains. After the Interpro annotation was completed, the enzyme code (EC number) was available, and the annotation was based in accordance to the GO sequences.

Lastly, an analysis was carried out with KEGG in order to obtain the metabolic KEGG maps from the selected the sequences.

Figure 11 shows an example of the results obtained from the B2G tool analysis.

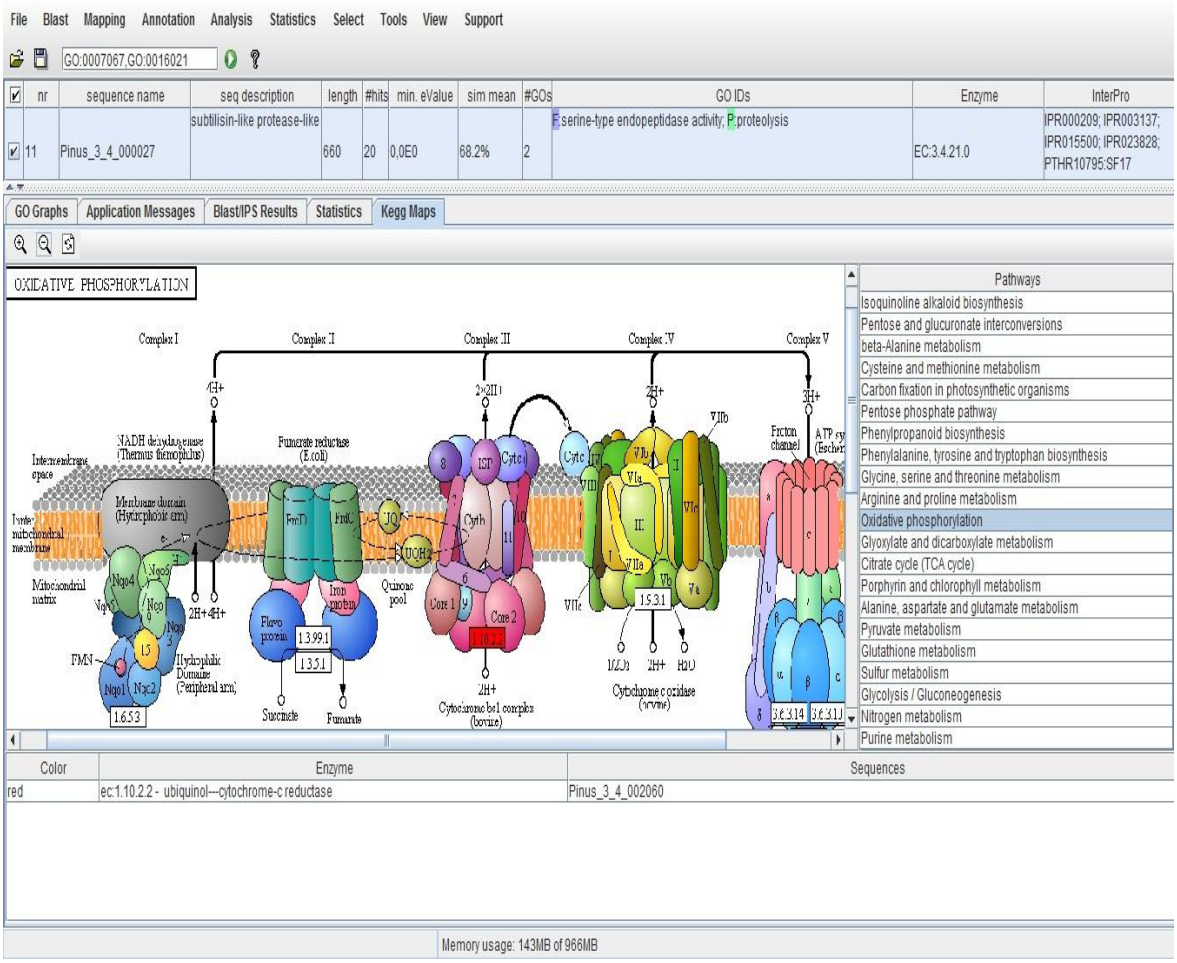

Figure 11: Example of an output obtained from this analysis using Blast2Go.

The pathway database at KEGG is divided into various biological processes that are grouped into five principal categories: cellular processes, metabolism, environmental information processing (such as signal transduction), genetic information processing and human diseases. The metabolism group is further subdivided into carbohydrates, energy, lipids, nucleotides, amino acids, glycans, polyketides/non-ribosomal peptides, cofactors/vitamins, secondary metabolites and xenobiotics. Figure 11 shows an example of the annotations and analysis that were developed. In this case, the oxidative phosphorylation pathway is highlighted and all the gene functions related to this enzyme are present in the output. This provides a global picture on the enzymatic pathways that are associated with the selected sequence.

Selecting the data

After the application of this interface, all reads from the same sequences were grouped and the genes with unknown function were removed from the analysis. A ratio between the normalized infested samples was calculated, with which we excluded all sequences with a ratio inferior to 1 and considered hits with ratios higher than 1 to be overexpressed for the numerator sample.

Then a search based on literature and Interpro (http://www.ebi.ac.uk/interpro/) was conducted in order to identify the genes which were related to the defense mechanism or to the infection in pine. After these steps were conducted, an indication was obtained on which were the most highly expressed genes and under which conditions they were expressed. These data can be seen in Figures 12 and 13.

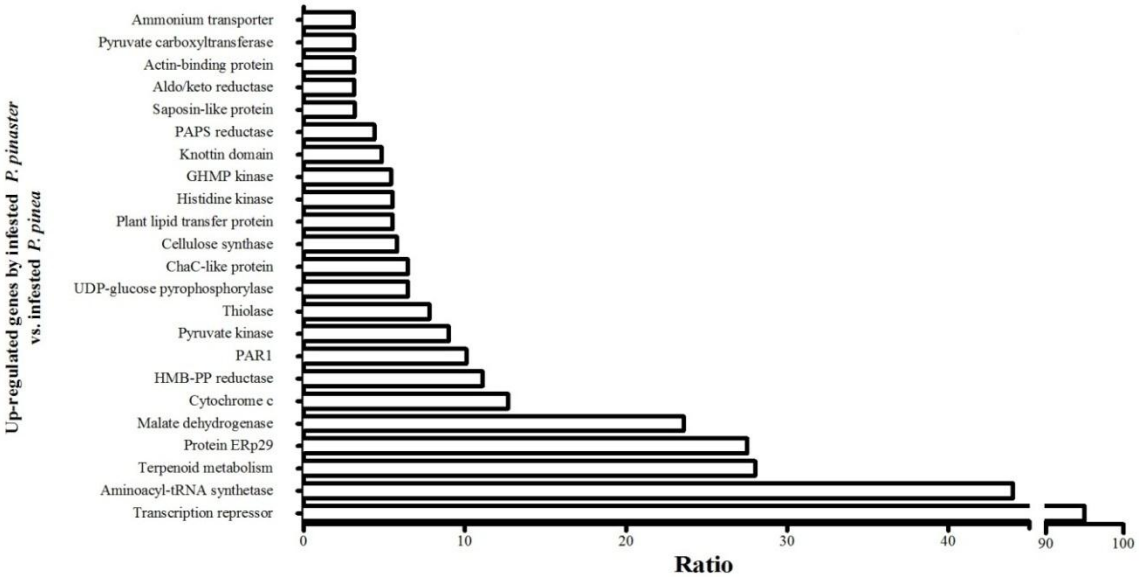

Figure 12: Up regulated genes by infested P. *pinaster* when compared to infested P. *pinea*.

Figure 12 shows the more highly expressed genes by infested *P. pinaster* versus infested *P. pinea*. This comparison shows that the most differentially expressed genes were a transcription repressor and genes related to the synthesis of amonoacyl-tRNA synthetase, whereas the less differentially expressed genes were the ones related to ammonium transport and with the synthesis of pyruvate carboxyltransferase. Genes associated with terpenoid metabolism were also differentially overexpressed by *P. pinaster*.

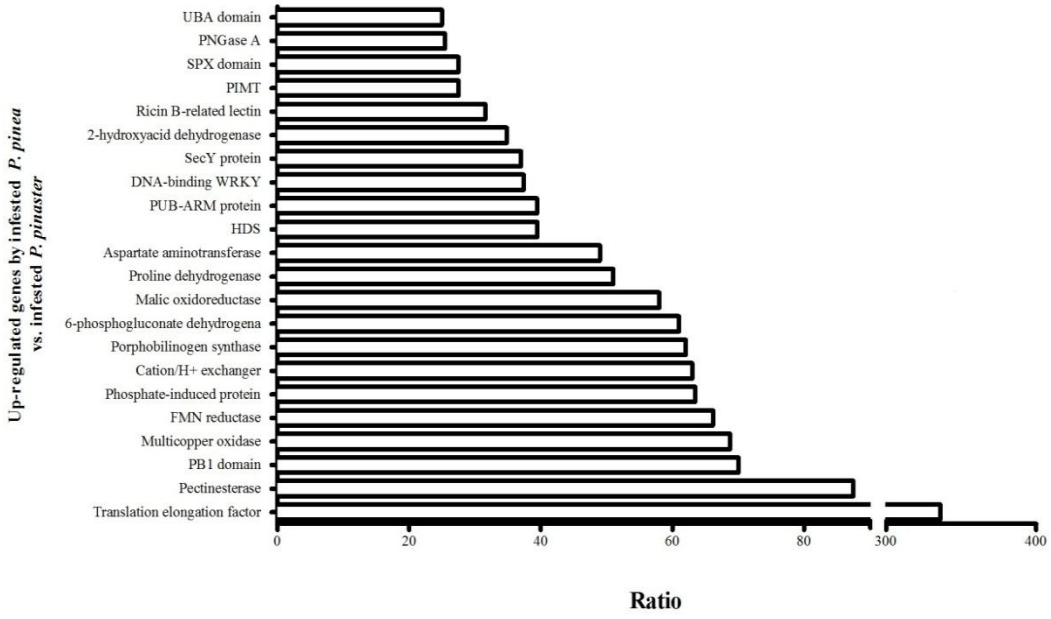

Figure 13: Up regulated genes by infested P. *pinea* when compared to infested P. *pinaster*.

Figure 13 illustrates the most differentially expressed genes when comparing infested *P. pinea* and infested *P. pinaster*. From this this comparison it can be seen that the genes more expressed were related to the synthesis of a translation elongation factor and Pectin esterase, and the less differentially expressed genes were genes of the synthesis of a U box domain and PNGse A.

The up and down regulated genes in PWN infested *P. pinaster* and *P. pinea* are represented in figure 14. Data was pooled and a ratio of the number of reads for each differentially expressed gene was calculated for each comparison. Ratios >1 were considered to be up-regulated for the numerator sample and <1, down-regulated.

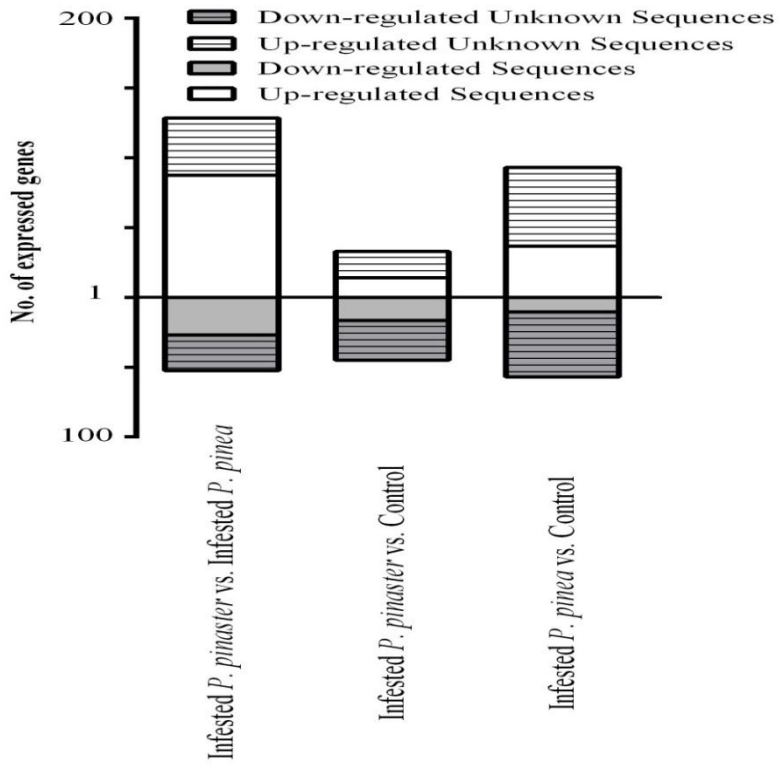

Figure 14: Differentially expressed genes.

After selection of these genes, these were annotated using B2G. This information can be seen in Table 3.

Table 3: General gene function and correspondent genes found between the differentially expressed data.

General Function	Genes
Oxidative stress	Aldo/keto reductase
	Multicopper oxidase
	2-hydroxyacid dehydrogenase
	6-phosphogluconate dehydrogenase
	PB1
	Cytochrome c
	FMN reductase
	Malic enzyme
	Proline dehydrogenase
Defense-related	Sugar related proteins
	PAPS reductase
	PAR1
	Plant Lipid Transfer Protein
	Saposin-like
	Pectinesterase
	PUB-ARM protein
	WRKY protein
	UBA domain
Transcription factors	aminoacyl-tRNA synthetase
	ERp29 protein
	Translation elongation factor
Secondary metabolites production	HMB-PP reductase
	HDS

KEGG analysis

In the current work, a preliminary KEGG analysis was performed in order to integrate the large-scale dataset that was generated by the high-throughput sequencing technology. KEGG pathways are manually drawn and derived from several sources, such as textbooks, literature and expert knowledge. The genomic

information in KEGG is retrieved from publicly available resources such as RefSeq data from the NCBI (Kanehisa *et al*., 2006).

The metabolic pathways that were more prevalent in infested *P. pinaster* versus infested *P. pinea* were penthose and glucuronate interconversions, phenylanine metabolism, starch and sucrose metabolism, methane metabolism, and phenylpropanoid biosynthesis.

The metabolic pathways that were more prevalent in *P. pinea* when compared to infested *P. pinaster* were: amino sugar and nucleotide sugar metabolism, citrate cycle (TCA cycle) and starch and sucrose metabolism (Figure 15).

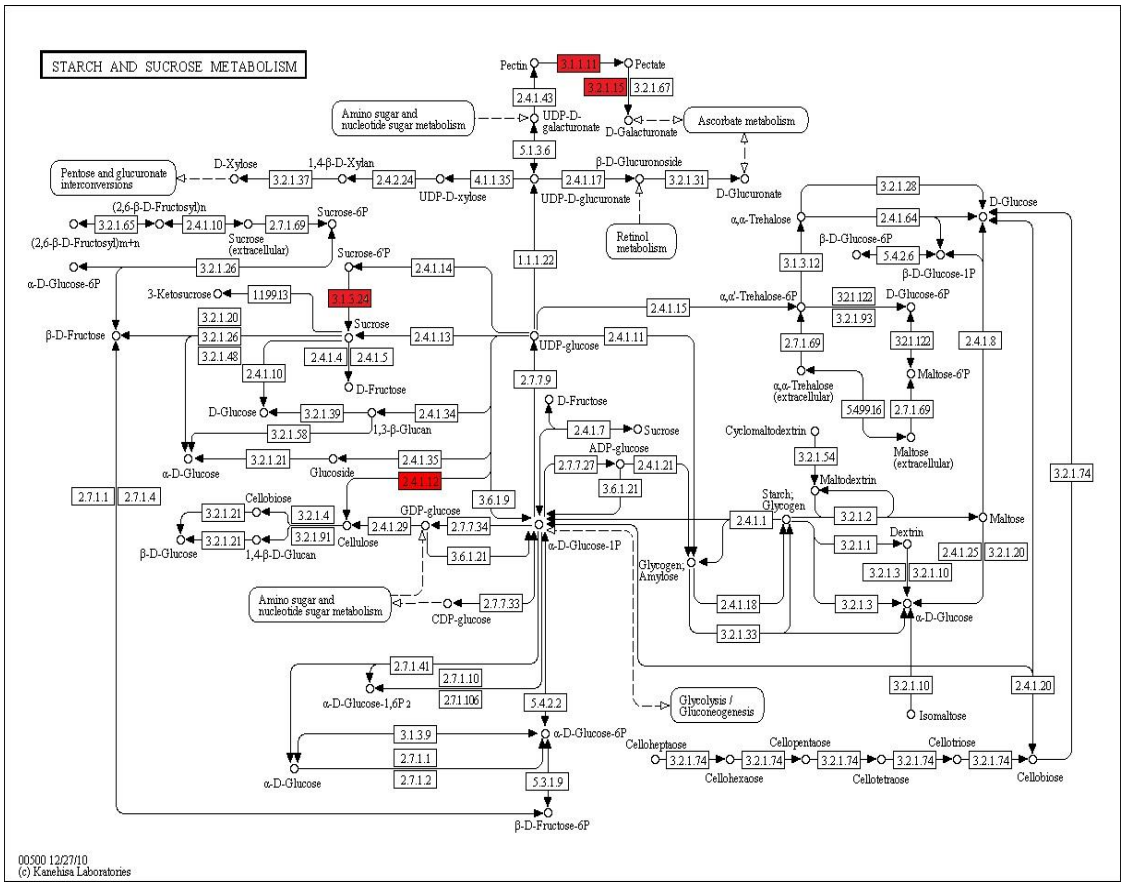

Figure 15: Example of a KEGG starch and sucrose metabolic pathway. The colored boxes represent the genes that were differentially expressed in the biological samples.

4. Discussion

454-Pyrosequencing

In this study a very effective and rapid technique was utilized for whole transcriptome sequencing: 454-pyrosequencing. This technique provided all the sequences that were analyzed in this study. 454-pyrosequencing has several advantages over other high throughput sequencing techniques. These include accuracy, flexibility and parallel processing, as well as the possibility of generating larger reads. This is because it generates a great number of sequences and information about the sample in a short amount of time (Vera *et al*., 2008). Moreover, this approach can be applied in any ecosystem for the study of different types of organisms.

But, like with other tools, this technique may have disadvantages such as high costs, and possible errors that are produced in the regions of homopolymers, mainly due to saturation of the signal during the sequencing process (Huse *et al*., 2007).

Because PWD is a complex disease involving organisms of different taxons (plant, nematode and bacteria) a quantitative insight into the microbial and non-microbial population of the samples was conducted. For this, the taxonomical affiliation of the annotated sequences was analysed using MG-RAST (Meyer *et al.*, 2008). Although about 50% of the sequences for each sample couldn't be annotated, the annotated ones were taxonomically analysed. The majority of the sequences in the samples belonged to Eukarya Domain and, as expected, 'Plantae' was the Kingdom that had more related sequences, corresponding to 89.1% to 96.45% of the sequences. To notice that only 1.8% to 12.8% corresponded to *Pinus* spp. sequence, which reflects the scarce available information in public databases. As there is more genomic information in public databases available for *Picea* spp., a range of 25.44-39.75% of the 'Plantae' sequences belonged to this category. Interestingly, P. pinea control sample was the one with higher percentage of *Pinus* spp. sequences compared to the other samples.

Comparing **P. pinea** *and* **P. pinaster** *molecular responses to nematode infection*

Plants have evolved a complex network of defence responses often associated with a localized response, where defences are systemically induced in remote parts of the plant in a process known as systemic acquired resistance (Shi *et al*., 2010).

These are usually stimulated by incompatible interactions between a pathogen and a resistant or non-host plant and result in two distinct types of hypersensitive reaction (HR): type I, which does not produce any visible symptoms and type II, that results in rapid and localized necrotic HR (Ulker *et al*., 2004), often eliciting *de novo* gene expression to acquire disease resistance.

To identify the participants in PWD response, the most represented genes in each sample were identified and the number of up and down regulated genes were analysed (Figure 14). In response to infestation *P. pinaster* differentially expressed 156 genes while the number of such genes in *P. pinea* was 300. When comparing between PWN infested *P. pinaster* with *P. pinea,* 257 genes had altered their altered expression levels and in the reverse comparison 105 genes were detected. Also, the expression varied between control treatments, which indicated that they were expressing different genes. This differential expression was also observed in other studies looking at the effect of *B. xylophilus* 24h after inoculation in susceptible and resistant pines (Hirao *et al.,* 2012). There was a high percentage (around 53%) of unknown sequences that were differentially expressed – this fact could stem from the low genomic information available for *Pinus* spp.. Also, the contigs without any homology may correspond to novel or diverged amino acid coding sequences, or could represent mostly 3' or 5' untranslated regions (UTRs) lacking protein matches as they are non-coding.

When the infested samples were compared against the controls, both presented a similar number of down-regulated genes, 21 by *P. pinea* and 33 by *P. pinaster*, but *P. pinea* up-regulated more than double the number of genes when compared to *P. pinaster*, which supports the hypothesis that these species respond differently to the nematode infestation.

When comparing the infested samples against the controls, both presented a similar number of down-regulated genes, 21 by *P. pinea* and 33 by *P. pinaster*, but

P. pinea up-regulated more than double the number of genes when compared to *P. pinaster*, which supports the hypothesis that these species present different patterns of gene expression in response to nematode infestation.

When comparing both infested samples, *P. pinaster* was the species with higher number of up-regulated genes, suggesting that, although *P. pinea* had a stronger reaction to the infestation, it differentially expressed less genes when compared to *P. pinaster*.

Due to the differential susceptibility to the PWN, it is interesting to compare the genes expressed by both *P. pinaster* and *P. pinea* when subjected to PWN infestation. The genes that were more represented by *P. pinaster* were a transcription repressor and a translation machinery component, aminoacyl-tRNA synthetase. Transcriptional regulators are key factors in the expression of specific genes and ensure the cellular responses to internal and external stimuli (Ulker *et al*., 2004) and the expression of factors related to protein synthesis could be involved in the activation of defence genes in response to the nematode attack. An ERp29 protein was also up-regulated. This protein is an endoplasmic reticulum stress-inducible protein, that is activated by the accumulation of transport-incompetent, misfolded and/or underglycosylated secretory proteins (Mkrtchian *et al*., 1998), again related to protein regulation.

Two component signalling elements have already been found to be present in *A. thaliana* and in rice, and here a possible histidine kinase was identified. These types of proteins are associated with signal transduction mediation in multiple pathways, acting like the hormones cytokinin and ethylene (Schaller *et al*., 2011).

As already mentioned in the Introduction section (page 4), the main symptom of the disease – wilting of leaves, that ultimately leads to tree death - is caused by a decrease in water potential in *B. xylophilus* infested stems (Fukuda, 1997). After water conduction is disrupted, xylem tracheids fill with air and oleoresin due to the resulting cavitation (Fukuda *et al*., 2007). The cavitation becomes permanent once tracheids are refilled with hydrophobic terpenoids synthesized by nematode injured parenchyma cells. Therefore, it is understandable why terpene metabolism related proteins, like (E)-4-hydroxy-3-methylbut-2-enyl diphosphate (HMB-PP) reductase and thiolase like protein, both involved in terpenoid synthesis, could be differentially expressed by infested *P. pinaster* (Kim *et al*., 2008; Soto *et al*., 2011). Subsequently,

as the water potential decreases, pine trees suffer severe oxidative stress and here, likewise other PWD-related studies (Shin *et al*., 2009; Santos and Vasconcelos, 2012), several oxidative-related genes were found, namely, a cytochrome c, found in the oxidation of phenolic elements in cell wall polymers under biotic stress, that has been associated with nematode infection in other studies (Shin *et al*., 2009), and an aldo/keto reductase, member of NADPH-dependent oxidoreductases, that intervenes in the elimination of reactive oxygen species produced by plant cells after suffering from a great amount of stress (Yamauchi *et al*., 2011).

Another symptom caused by PWN infection is the enhancing of plants' respiration and oxidative stress (Fukuda, 1997). A possible malate dehydrogenase (MDH) was found to be over-expressed by infested *P. pinaster*. MDH is responsible for the interconversion of malate and oxaloacetate, regulating respiratory rate in plants (Tomaz *et al*., 2010), which may be related to the disease.

Nematodes feed off young differentiating phloem fibers and xylem ray parenchyma cells (Fukuda *et al*., 2007). A cellulose synthase was up-regulated in infested *P. pinaster* which could be explained by the fact that, as the damage caused by nematode feeding disrupts wood formation, this is an essential enzyme for primary and secondary cell wall biosynthesis (Nairn *et al*., 2008).

Interestingly, several plant defense related genes were also up-regulated by *P. pinaster* in response to the infestation. These included: a probable photoassimilate-responsive protein (PAR1) that displays features similar to pathogenesis-related proteins (Herbers *et al*., 1995); a putative plant lipid transfer protein (LTP), that may be involved in pathogen-defense reactions via inhibition of bacterial and fungal growth (Kader, 1997); sugar related proteins - like pyruvate-related proteins, GHMP kinase and a UDP-glucose pyrophosphorylase (Tadege *et al*., 1998; Yang *et al*., 2009; Zeczycki *et al*., 2009) were up-regulated - have been shown to increase after pathogen infection and, in *Arabidopsis thaliana*, the expression of sugar transport proteins can be induced by wounding and pathogen attack, altering cell wall dynamics (Poschet *et al*., 2010); a phosphoadenosine phosphosulphate (PAPS) reductase, maily involved in sulphate assimilation, that may contribute to plant defense, since S-containing secondary metabolites act as compounds against pathogens and herbivores (Kopriva *et al*., 2004); and a sequence belonging to the saposin-like protein family that, as its members have membrane permeabilizing

activity, participates in the plant defence mechanism against fungal pathogens (Bryksa *et al*., 2011).

In a recent study conducted in *P. phumbergii* defense response genes, an antimicrobial peptide, salicilic acid-responsive genes and jasmonic acid/ethylene-responsive genes were induced more quickly and to a higher level in susceptible than in resistant trees (Hirao *et al.,* 2012). These gene classes were not the ones found to be more highly expressed by susceptible *P. pinaster*, possibly pointing out to a species-specific response in disease susceptibility amongst pine trees.

Perhaps the most helpful information when aiming at identifying resistance genes to the PWN comes from the analysis of the genes expressed by PWN-infested *P. pinea* (less susceptible to PWN) when compared with PWN-infested *P. pinaster*. PWN-infested *P. pinea* had higher levels of expression in general and some of the most interesting findings included a plant disease resistance protein, which was not found to be expressed by *P. pinaster* and a ricin B-related lectin. Plant lectins have already been pointed out as participants in the general defence against a multitude of plant pathogens, including nematodes.

The oxidative stress related multicopper oxidase, flavin mononucleotide (FMN) reductase and 6-phosphogluconate dehydrogenase (Shin *et al*., 2009; Pöggeler, 2011; Stover *et al*., 2011) were all up-regulated and these proteins have a crucial role in PWD since, as previously mentioned, they are believed to play an important role in the maintenance of intracellular redox balance and in stress response/tolerance in plants. Particularly, FMN reductase has already been identified in previous studies in our lab as possibly related to *B. xylophilus* infection (Yu and Zhang, 2012). Also, a phox/Bem1 (PB1) domain was found to be more represented by infested *P. pinea* and this domain is usually found in signalling proteins including oxidases and cytosolic factors (Hirano *et al*., 2005) and a 2-hydroxyacid dehydrogenase that is associated with 3-phosphoglycerate dehydrogenase and may play a role in the oxidation-reduction process (Ho *et al*., 1999).

The malic enzyme (Liu *et al*., 2007) and a proline dehydrogenase are also involved in the oxidative stress, and believed to play an important role in plant defense.

The second one was recently found in *Arabidopsis* to affect cell death and disease resistance against biotic stress by altering cellular redox state, besides other mechanisms (Cecchini *et al*., 2011).

The most up-regulated genes by infested *P. pinea* was a possible translation elongation factor, mainly involved in protein synthesis and in regulation of different cellular processes (Mateyak and Kinzy, 2010), and the defence related protein pectinesterase, that belongs to a group of methyl jasmonate inducible pathogenesis-related proteins and has been correlated to cell wall extension (here justified by the need to replace the nematode feeding-damaged cell walls) and microbial pathogens inhibition (Hothorn *et al*., 2004; Sabater *et al*., 2011). As pointed out by others, up-regulation of cell wall-related genes contributing to the strength of cell walls would be a very effective defense against PWN infection, because these events might restrict PWN migration (Hirao *et al.,* 2012).

Other defence related proteins by PWN infested *P. pinea*, like a plant U-box (PUB) protein and a WRKY protein were also found. The first, involved in ubiquination, usually carries tandem armadillo repeats (PUB-ARM proteins) in eukaryotes. PUB-ARM proteins were identified as part in the pathogen response in tobacco and Arabidopsis (Drechsel *et al*., 2011; Li *et al*., 2012). The second are transcriptionally inducible upon pathogen infection and other defence-related stimuli and, although this may not be true for all WRKY genes, the overexpression of (for example) *AtWRKY18* was shown to activate pathogenesis-related genes and to enhance resistance to certain pathogens (Ulker *et al*., 2004; Grunwald *et al*., 2008).

Another hit possibly involved in ubiquination was detected, a UBA domain. In plants, ubiquitinated proteins were described to regulate, besides germination and flowering, cell cycle and processes of response to the majority of external stimuli (e.g. biotic and abiotic stresses) (Manzano *et al*., 2008).

Due to the mechanism of action of PWD, as pointed out before, terpenoid metabolism is very important in pine trees defence. In *P. pinea* a terpenoid-related protein, namely, a 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate synthase (HDS), participant in the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway was found. HDS and HD reductase are necessary to resin production and have been already proposed to be important in the physiological response to invasion by the PWD

nematode in *P. densiflora* (Kim *et al*., 2009), since PWN progression leads to the cessation of resin flow (Autoridade Florestal Nacional, 2012).

One of the main symptoms of PWD is the decrease of photosynthetic rate, which leads to the wilting of leaves. As previous studies of our lab showed, after PWN infestation, the chlorophyll content suffers from a quick decline, especially in *P. pinaster* (Beale, 1999). Here, a porphobilinogen synthase, directly involved in chlorophyll synthesis, was identified (Beale, 1999). Maybe, with the differential expression of this gene, *P. pinea* is capable to somewhat compensate this decline.

The protein L-isoaspartyl (D-aspartyl) *O*-methyltransferase (PIMT) is commonly present in seed tissues, however its activity is elevated under stressful conditions and in *Arabidopsis* it was hypothesised that this protein may be involved in plant stress response (Villa *et al*., 2006; Ogé *et al*., 2008).

Among the up-regulated genes that cannot be directly associated with plant stress response, in *P. pinaster* sample, a ChacC-like protein, a knottin domain, a actin-binding protein and a nitrogen-stress related ammonium transporter were identified; and, in *P. pinea* sample, both sugar-related phosphate-induced protein with unkown function and SPX domain, a putative aspartate aminotransferase, a SecY protein and a peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A (PNGase A). Even though their association with plant disease defence or stress is not yet documented, the current study seems to indicate that they may have a role in the infestation response.

High throughput sequencing allowed the identification of several candidate genes that may be involved in the response to the PWN. Like in other studies (Shin *et al*., 2009), one day after infestation with *B. xylophilus* the plants triggered the expression of genes related to oxidative stress, abiotic or biotic stimulus, plant stress, transcription factors, transport, and secondary metabolites production. These genes can be useful targets in genetic transformation and breeding programs that aim at generating wild pine that is resistant to the PWN.

The appearance of a small percentage of nematode sequences may be due to the fact that nematodes can be present in many environments, such as contaminated water, soil, compost beds, among others. Also, the genome of many nematodes, including *B. xylophilus*, is not yet well described in public databases, therefore the

results may reflect homology to the nematodes that have higher genomic information in publicly available sequence repositories.

KEGG Analysis

Tools that allow the analysis of metabolic pathways are very useful in order to obtain and analyze relationships between genes that encode specific enzymes, and to find strategies for diagnosis and treatment of complex diseases. It is important to note that metabolic reconstruction also induces the search for missing genes, encouraging the application of genomics (Winter and Huber, 2010).

The pathway more commonly found in this study were the pentose pathway, the pathway for glucuronate interconversion, the pathway for phenylanine metabolism, amino acid, sugar and nucleotide metabolism, phenylppropanoid biosynthesis, methane metabolism, and citrate cycle (TCA cycle).

The pentose pathway and glucuronate interconversions are related to the metabolism of carbohydrates and are present in various polysaccharides and glycosides being very commonly found in plants (Wood, 1986).

The phenylalanine pathway is probably well represented due to the fact that phenylanine is the starting compound used in the flavonoid biosynthesis because it has an important role in activating various plant defense responses, including expression of the pathogenesis-related genes and resistance (Kachroo *et al*., 2001).

Amino sugar and nucleotide sugar metabolism are pathways that contain the intermediates that produce some of the activated sugars needed for glycosylation reactions. Nucleotide sugar represents a group of enzymatic reactions by which plants synthesize monosaccharides for the incorporation into cell wall material and assist in the transport of nutrients (Reiter *et al.,* 2001).

Phenylpropanoid biosynthesis pathways are a group of plant secondary metabolites having a wide variety of functions both as structural, signaling molecules and play a key role in plant development and protection against environmental stress (Dixon *et al.,* 2002).

Citrate cycle (TCA cycle) is an important aerobic pathway for the final steps of the oxidation of carbohydrates and fatty acids (Schnarrenberger and Martin, 2002).

Starch and sucrose metabolism are highly represented probably because sucrose plays an important role in plant growth and development of plants. It is a main product of photosynthesis and functions as a carrier, as well as direct or indirectly regulating gene expression. Sucrose is synthesized in the cytosol, transiently stored in the vacuole and exported through the apoplast (Winter and Huber, 2010). Moreover the starch is the main source of reserve in plants that can be found in the roots and other plant organs.

5. Conclusion

This worked allowed a better understanding of the molecular responses of susceptible and less susceptible pines to the PWN.

In our samples a very small amount of nematode related sequences were found, thus it can be concluded that during RNA extraction of the four samples, the protocol yielded mostly plant RNA and a very small fraction of nematode RNA.

With regards to the gene functions most commonly identified, it can be concluded that the majority of the sequence functions were associated with protein metabolism and carbohydrate metabolism. However, a significant fraction of sequences associated with RNA metabolism were also highly represented. The sequences that were more commonly found in *P. pinaster* were transcription repressors and a translation machinery component: aminoacyl-tRNA synthetase. Transcriptional regulators are key factors in the expression of specific genes and ensure the cellular responses to internal and external stimuli and the expression of factors related to protein synthesis could be involved in the activation of defence genes in response to nematode attack.

It can be concluded that cellulose synthase is also important in the disease susceptibility, as this gene was up-regulated in infested *P. pinaster*.

Putative transcripts were identified using 454 sequencing technology, which showed that *P. pinaster*, a very susceptible species to the PWN, when infested with *B. xylophilus*, over-expresses genes related to terpenoid secondary metabolism (including some with nematicidal activity), defense against pathogen attack and oxidative stress (a common PWD consequence).

On the other hand, *P. pinea* – believed to be less susceptible to this disease – up-regulated transcription regulation related genes, that are needed to activate plant defense responses, and showed higher levels of expression in general of stress response genes such as SNARE proteins, ricin B-related lectin, and disease resistance proteins.

This study establishes a framework to the construction of knowledge about the molecular response of pine trees to PWN, and elucidates the defense mechanisms utilized by *P. pinaster* and *P. pinea* against PWN infection.

6. Future work

As future work, the selection of a subset of genes showing differential expression between samples and that can be responsible for the higher resistance to the pine wood nematode is underway. After selecting these gene targets, cloning of the complete sequences will be performed, and the entire cDNA fragment will be utilized in a program of genetic transformation, with a collaborating institute, in order to test if transgenic *P. pinaster* expressing these genes has lower susceptibility to *B. xylophilus*.

Another possible study that could be undertaken would be to merge all the KEGG pathway analysis in order to obtain a single pathway relationship that would integrate the identified differentially expressed genes.

7. Annexes

Figures taken from http:// transcriptomics.biocant.pt/pine/.

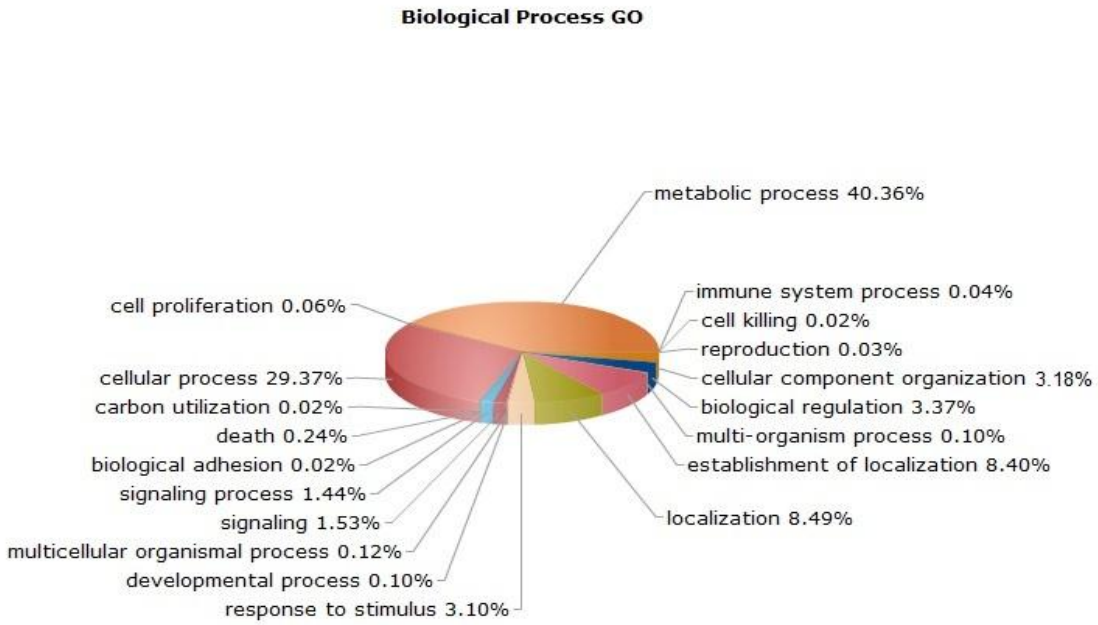

Figure 16: Biological Process GO by infested *P. pinaster* when compared to infested *P. pinea*.

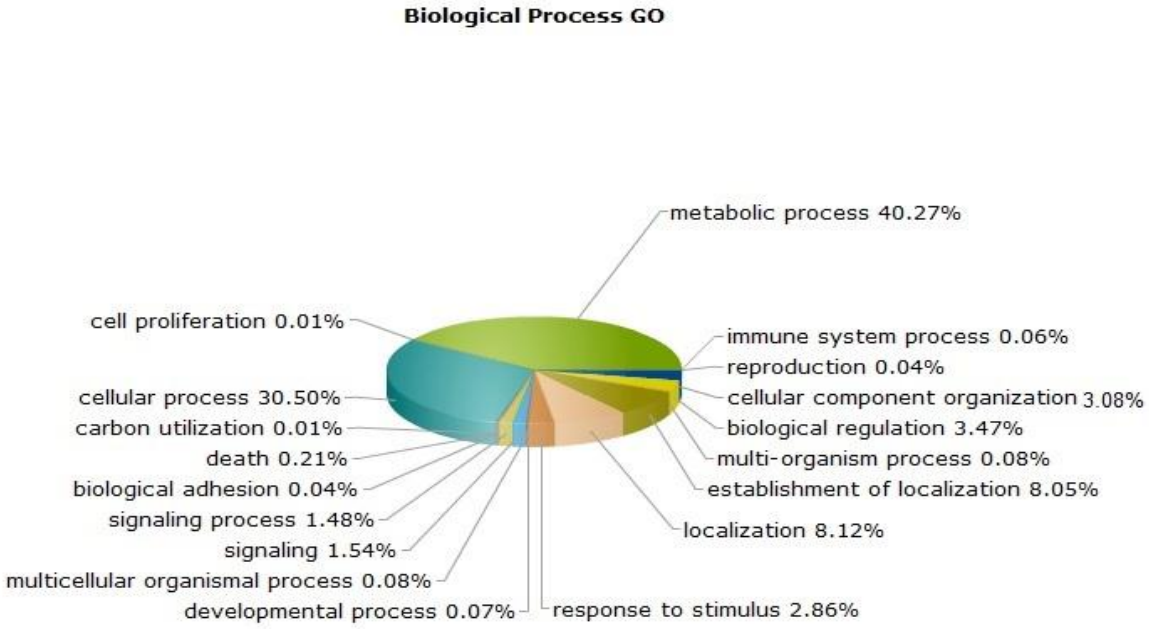

Figure 17: Biological Process GO by infested *P. pinea* when compared to infested *P. pinaster*.

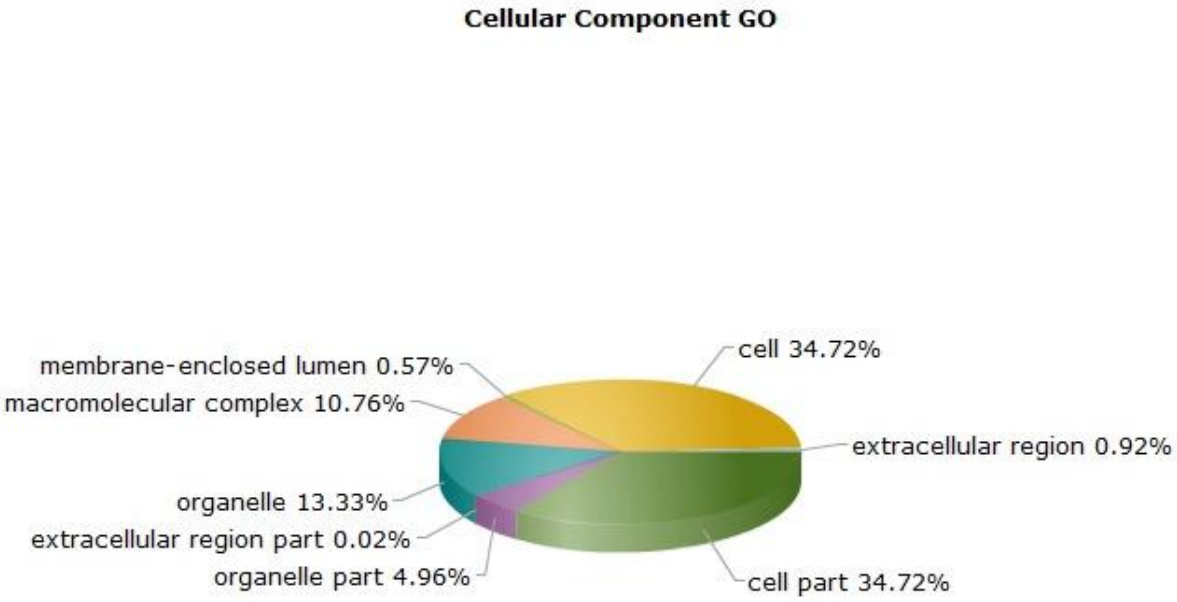

Figure 18: Cellular Component GO by infested *P. pinaster* when compared to infested *P. pinea*.

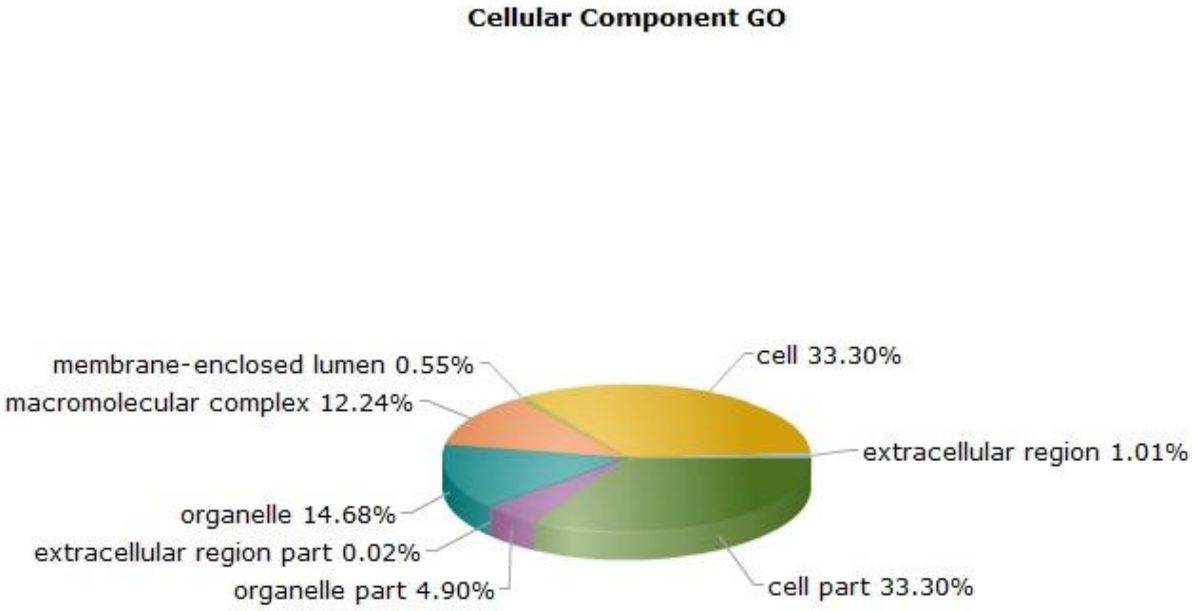

Figure 19: Cellular Component GO by infested *P. pinea* when compared to infested *P. pinaster*.

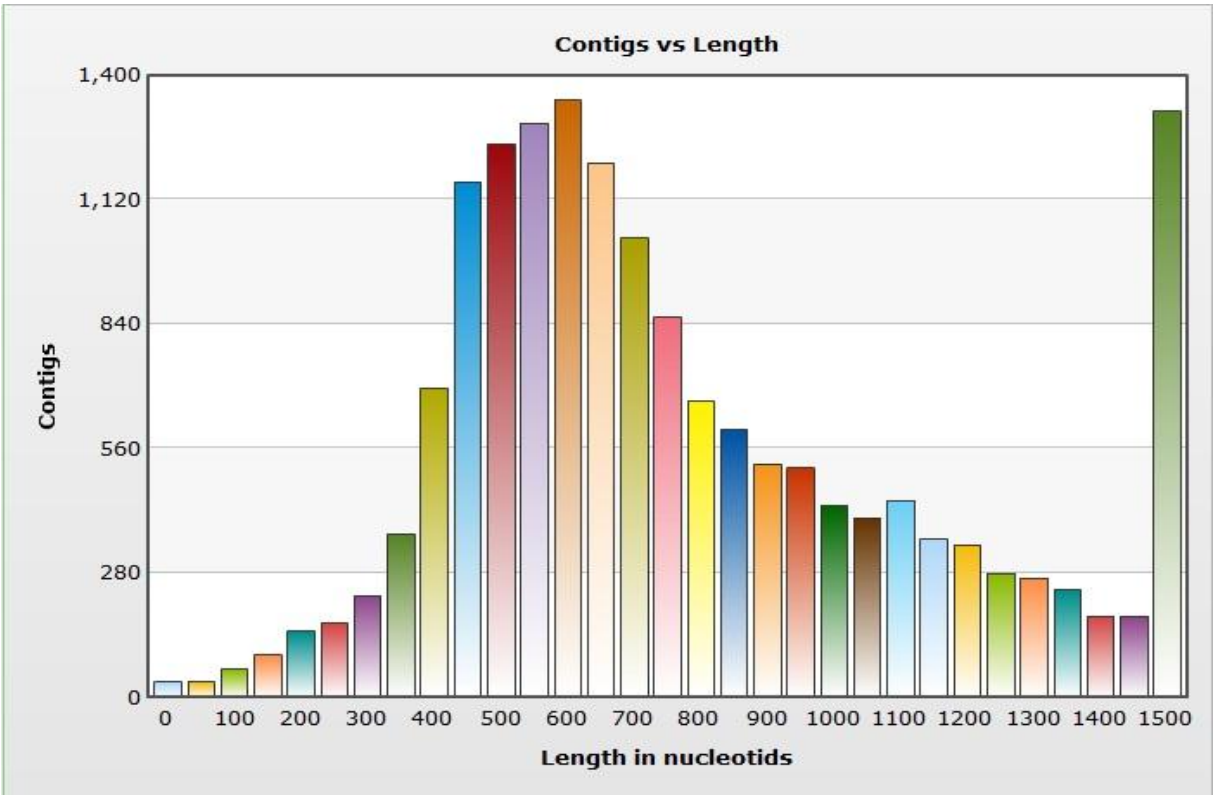

Figure 20: Number of contigs vs Length by infested *P. pinaster* when compared to infested *P. pinea*.

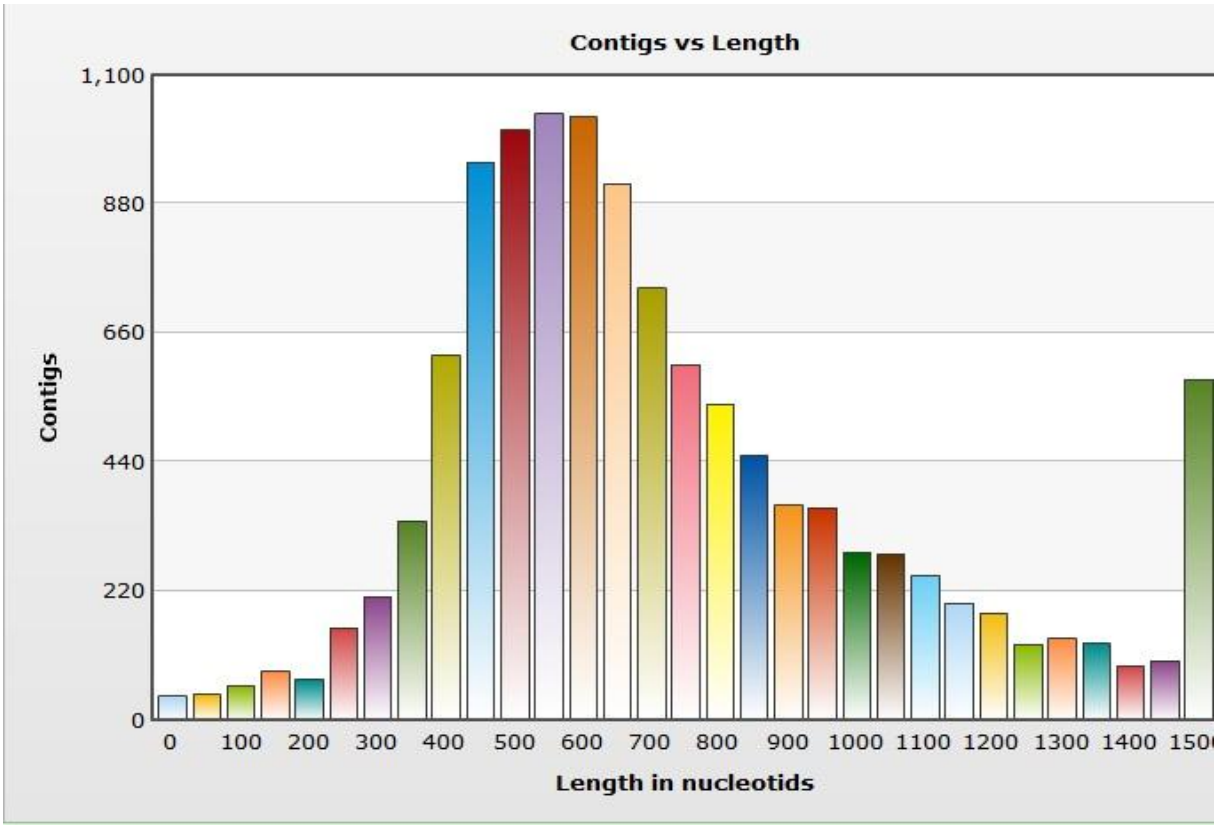

Figure 21: Number of contigs vs Length by infested *P. pinea* when compared to infested *P. pinaster*.

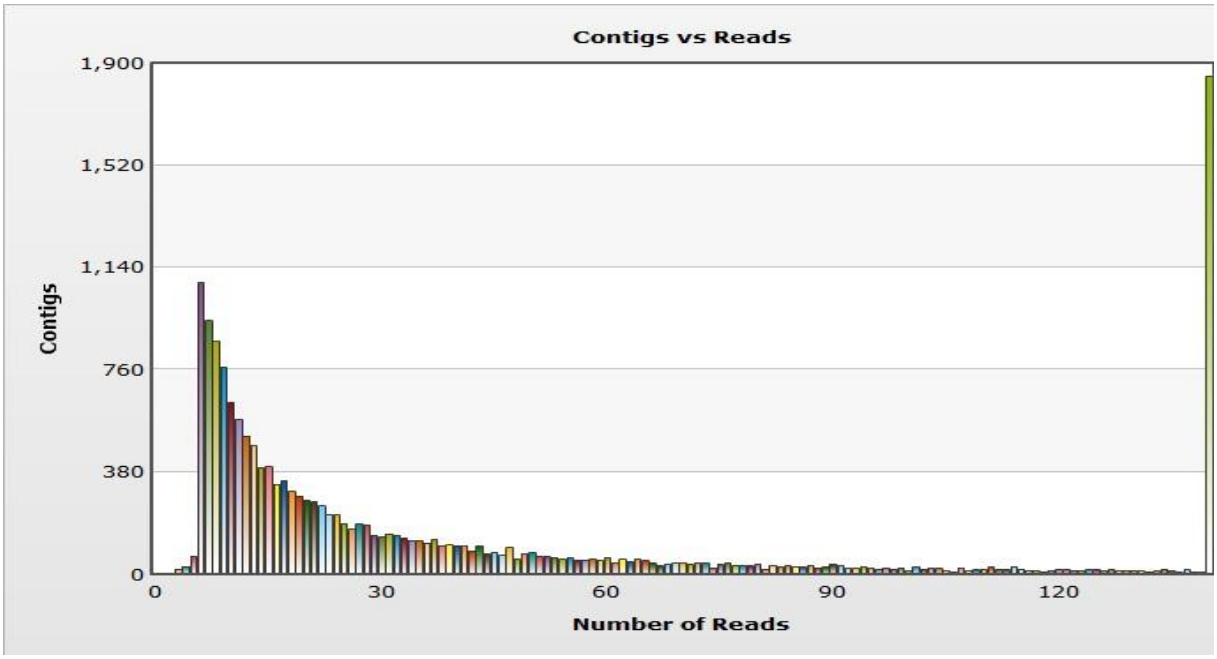

Figure 22: Number of contigs vs Number of reads by infested *P. pinaster* when compared to infested *P. pinea*.

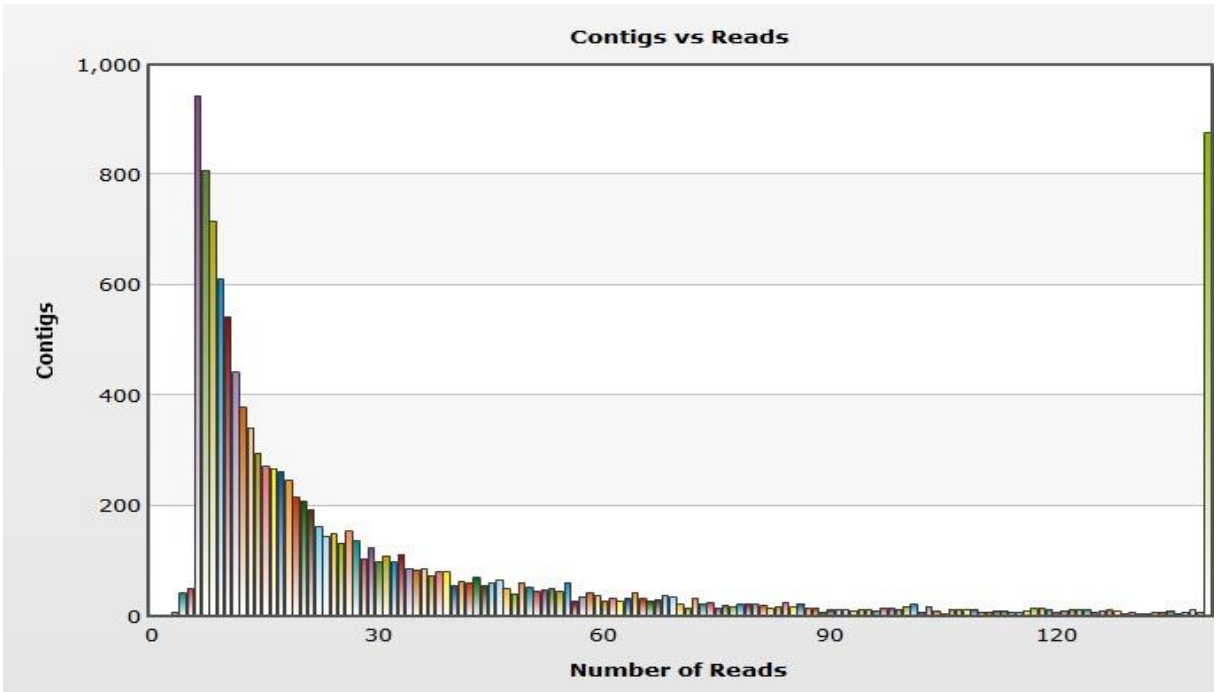

Figure 23: Number of contigs vs Number of reads by infested *P. pinea* when compared to infested *P. pinaster*.

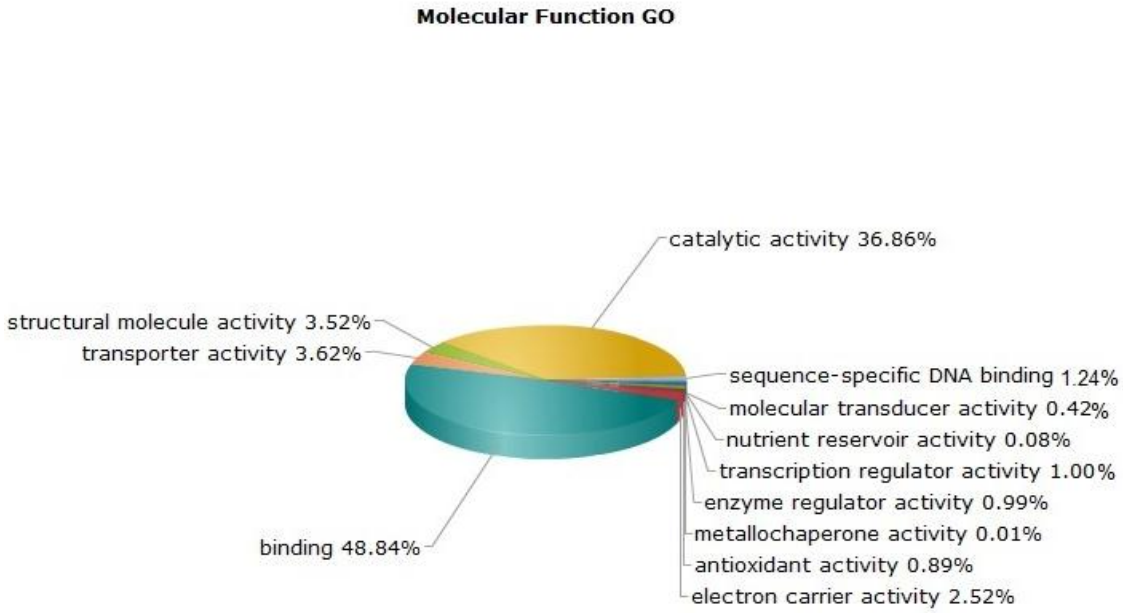

Figure 24: Molecular Function GO by infested *P. pinaster* when compared to infested *P. pinea*.

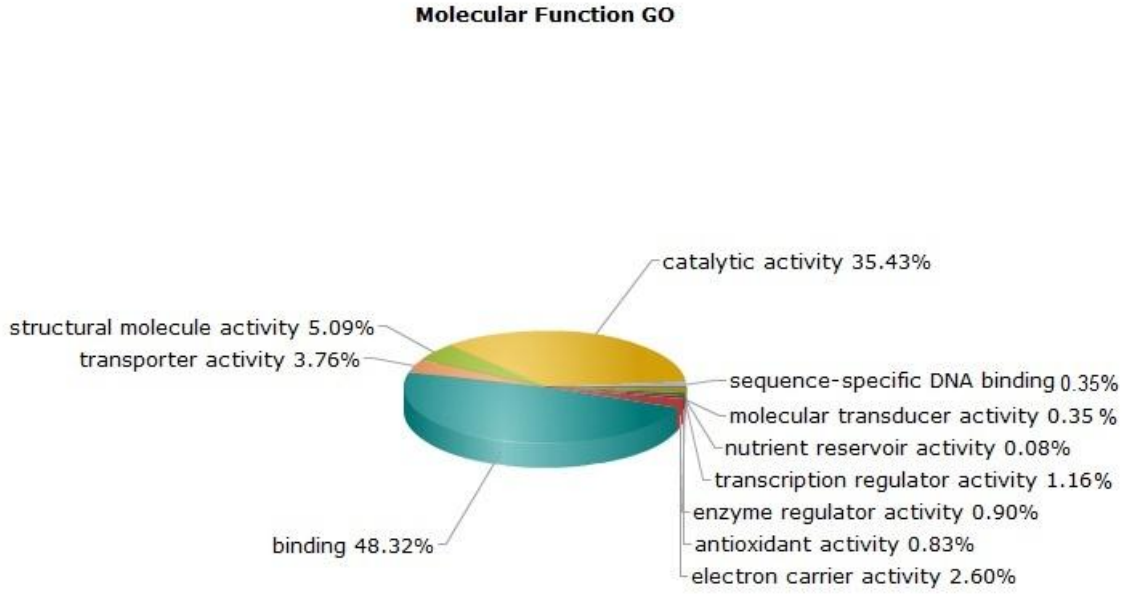

Figure 25: Molecular Function GO by infested *P. pinea* when compared to infested *P. pinaster*.

8. Bibliography

Apweiler, R., Biswas, M., Fleischmann, W., Kanapin, A., Karavidopoulou, Y., Kersey, P., Kriventseva, E.V., Mittard, V., Mulder, N., Phan, I., Zdobnov, E. 2001. Proteome Analysis Database: online application of InterPro and CluSTr for the functional classification of proteins in whole genomes. Nucleic Acids Research 29:44-48.

Ashburner, M, Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel, T.L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. 2000. Gene Ontology: tool for the unification of biology. Nature Genetics 25:25-29.

Autoridade Florestal Nacional. 2012. Programa de Ação Nacional para o Controlo do Nemátodo da Madeira do Pinheiro (NMP). Autoridade Florestal Nacional / Direcção-Geral de Agricultura e Desenvolvimento Rural: pp 1-15.

Bai, X., Veja, L.R., Praveen, M., Pierluigi, B., Herms, D.A., Mittapalli, O. 2010. Transcriptomic Signatures of Ash (Fraxinus spp.) Phloem. Plos One 6: n. pag.

Beale, S.I. 1999. Enzymes of chlorophyll biosynthesis. Photosynthesis Research 60:43-73.

Bryksa, B.C., Bhaumik, P., Magracheva, E., De Moura, D.C., Kurylowicz, M., Zdanov, A., Dutcher, J.R., Wlodawer, A., Yada, R.Y. 2011. Structure and mechanism of the saposin-like domain of a plant aspartic protease. The Journal of Biological Chemistry 286: 28265-28275.

Chancerel, E., Lepoittevin, C., Le Provost, G., Lin, Y-C, Jaramillo-Correa, J.P., Eckert, A.J., Wegrzyn, J.L., Zelenika, D., Boland, A., Frigerio, J-M., Chaumeil, P., Garnier-Géré, P., Boury, C., Grivet, D., González-Martínez, S.C., Rouzé, P., de Peer, Y.V., Neale, D.B., Cervera, M.T., Kremer, A., Plomion, C. 2011. Development and implemention of a highly-multiplexed SNP array for genetic mapping in maritime pine and comparative mapping with loblolly pine. BMC Genomics 12:368.

Cecchini, N.M., Monteoliva, M.I., Alvarez, M.E. 2011. Proline dehydrogenase contributes to pathogen defense in Arabidopsis. Plant Physiology 155: 1947-1959.

Conesa, A., Götz, S. 2008. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. International Journal of Plant Genomics 1: 13.

Conesa, A., Götz, S, García, G.J.M., Terol, J., Talón, M., Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics Applications Note 21: 3674-3676.

Dixon, R.A., Achnine, L., Kota, P., Liu, C.J., Reddy, M.S.S., Wang, L. 2002. The phenylpropanoid pathway and plant defence—a genomics perspective. Molecular Plant Pathology 3: 371–390.

Drechsel, G., Bergler, J., Wippel, K., Sauer, N., Vogelmann, K., Hoth, S. 2011. C-terminal armadillo repeats are essential and sufficient for association of the plant U-box armadillo E3 ubiquitin ligase SAUL1 with the plasma membrane. Journal of Experimental Botany 62: 775-785.

Fukuda, K. 1997 Physiological Process of the Symptom Development and Resistance Mechanism in Pine Wilt Disease. Journal of forest research 2: 171-81.

Fukuda, K., Utsuzawa, S., Sakaue, D. 2007. Correlation between acoustic emission, water status and xylem embolism in pine wilt disease. Tree Physiology 27: 969-976.

Futai, K., Furuno, T., 1979. The variety of resistances among pine species to pine wood nematode, Bursaphelenchus lignicolus. Bull Kyoto 51: 23-36.

Futai, K., 2003. Pine Wilt Disease: Various biological relationships and resulting events. Proceedings: International Union of Forest Research Organizations Kanazawa 2003 "Forest Insect Population Dynamics and Host Influences".

Götz, S., Arnold, R., León, P.S., Rodríguez, S.M., Tischler, P., Jehl, M.A., Dopazo, J., Rattei, T., Conesa, A. 2011. B2G-FAR, a species-centered GO annotation repository. Oxford journals 27: 919–924.

Grunwald, W., Karimi, M., Wieczorek, K., Van D.C.E., Wischnitzki, E., Grundler, F., Inzé, D., Beeckman, T., Gheysen, G. 2008. A role for AtWRKY23 in feeding site establishment of plant-parasitic nematodes. Plant Physiology 148: 358-368.

Herbers, K., Mönke, G., Badur, R., 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 Molecular Biology 29: 1027-1038.

Hirao, T., Fukatsu, E., Watanabe, A. 2012. Characterization of resistance to pine wood nematode infection in Pinus thunbergii using suppression subtractive hybridization. BMC Plant Biology 12, 12:13.

Hirano, Y., Yoshinaga, S., Takeya, R., Suzuki, N.N., Horiuchi, M., Kohjima, M., Sumimoto, H., Inagaki, F. 2005. Structure of a cell polarity regulator, a complex between atypical PKC and Par6 PB1 domains. The Journal of Biological Chemistry 280: 9653-9661.

Ho, C.L., Noji, M., Saito, M., Saito, K. 1999. Regulation of serine biosynthesis in Arabidopsis. Crucial role of plastidic 3-phosphoglycerate dehydrogenase in non-photosynthetic tissues. The Journal of Biological Chemistry 274:397-402.

Horner, D.S., Pavesi G., Castrignanò T., Meo, P.D'O., Liuni S., Sammeth M., Picardi E., Pesole G. 2009. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. Oxford Journals 2: 181-197.

Hothorn, M., Wolf, S., Aloy, P., Greiner, S., Scheffzek, K. 2004. Structural insights into the target specificity of plant invertase and pectin methylesterase inhibitory proteins. The Plant Cell 16: 3437-3447.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., Welch, D.M. 2007. Accuracy and Quality of Massively Parallel DNA Pyrosequencing. Genome Biology 8: n. pag.

Ingraham, J.L., Ingraham, C.A. 2011. Introdução à microbiologia: Uma abordagem baseada em estudos de casos. Cengage Learning. 776 pp.

Jones, J.T., Moens, M., Mota, M., Li, H., Kikuchi, T., 2008. Bursaphelenchus xylophilus: opportunities in comparative genomics and molecular host–parasite interaction. Molecular Plant Pathology 9: 357–368.

Kachroo, P., Shanklin, J., Shah, J., Whittle, E.J., Klessig, D.F. 2001. A fatty acid desaturase modulates the activation of defense signaling pathways in plants. Sciense Sessions 98: 9448-9453.

Kader, J.C. 1997. Lipid-transfer proteins: a puzzling family of plant proteins. Trends in Plant Science 2: 66-70.

Kanehisa, M., Goto, S., Hattori, M., Kinoshita, K., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M. 2006. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Research 34: 354–357.

Kanehisa, M., Goto, S.; 2000 KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 28, 27-30.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M.; 2012 KEGG for integration and interpretation of large-scale molecular datasets. Nucleic Acids Research 40: D109-D114.

Kim, S.M., Kuzuyama, T., Kobayashi, A., Sando, T., Chang, Y.J., Kim, S.U. 2008. 1-Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (IDS) is enconded by multicopy genes in gymnosperms Ginko biloba and Pinus taeda. Planta 227: 287-298.

Kim, Y.B., Kim, S.M., Kang, M.K., Kuzuyama, T., Lee, J.K., Park, S.C., Shin, S.C., Kim, S.U. 2009. Regulation of resin acid synthesis in Pinus densiflora by differential transcription of genes encoding multiple 1-deoxy-D-xylulose 5-phosphate synthase and 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase genes. Tree Physiology 29: 737-749.

Kopriva, S., Koprivova, A. 2004. Plant adenosine 5'-phosphosulphate reductase: the past, the present, and the future. Journal of Experimental Botany 55: 1775-1783.

Langmead, B., Hansen, K.D., Leek, J.T. 2010. Cloud-scale RNA-sequencing differential expression analysis with Myrna. Genome Biology. 11:R83.

Lorenz, W.W., Alba, R., Yum Y-S., Bordeaux, J.M., Simões, M., Dean, J.F.D. 2011. Microarray analysis and scale-free gene networks identify candidate regulators in drought-stressed roots of loblolly pine (P. taeda L.). BMC Genomics 12:264.

Li, W., Ahn, I.P., Ning, Y., Park, C.H., Zeng, L., Whitehill, J.G., Lu, H., Zhao, Q., Ding, B., Xie, Q., Zhou, J.M., Dai, L., Wang, G.L. 2012. The U-Box/ARM E3 ligase PUB13 regulates cell death, defense, and flowering time in Arabidopsis. Plant Physiology 159: 239-250.

Liu, S., Cheng, Y., Zhang, X., Guan, Q., Nishiuchi, S., Hase, K., Takano, T. 2007. Expression of an NADP-malic enzyme gene in rice (Oryza sativa. L) is induced by environmental stresses; over-expression of the gene in Arabidopsis confers salt and osmotic stress tolerance. Plant Molecular Biology 64:49-58.

Manzano, C., Abraham, Z., López, T.G., Del Pozo J.C. 2008. Identification of ubiquitinated proteins in Arabidopsis. Plant Molecular Biology 68: 145-158.

Mateyak, M.K., Kinzy, T.G. 2010. eEF1A: thinking outside the ribosome. The Journal of Biological Chemistry 285: 21209-21213.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A. 2008. The Metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 9:386.

Mkrtchian, S., Baryshev, M., Matvijenko, O., Sharipo, A., Sandalova, T., Schneider, G., Ingelman, S.M. 1998. Oligomerization properties of ERp29, an endoplasmic reticulum stress protein. FEBS Letters 431: 322-326.

Mota, M.M., Braasch, H., Bravo, M.A., Penas, A.C., Burgermeister, W., Metge, K., Sousa, E. 1999. First Report of Bursaphelenchus Xylophilus in Portugal and in Europe. Nematology 1: 727-34.

Nairn, C.J., Lennon, D.M., Wood, J.A., Nairn, A.V., Dean, J.F.D. 2008. Carbohydrate-related genes and cell wall biosynthesis in vascular tissues of loblolly pine (Pinus taeda). Tree Physiology 28: 1099-1110.

Nose, M., Shiraishi, S. 2011. Comparison of gene expression profiles of resistant and non-resistant Japanese black pine inoculated with pine wood nematode using a modified LongSAGE technique. Forest Pathology 41:143-155.

Ogé, L., Bourdais, G., Bove, J., Collet, B., Godin, B., Granier, F., Boutin, J.P., Job, D., Julien, M., Grappin, P. 2008 Protein repair L-isoaspartyl methyltranferase 1 is involved in both seed longevity and germination vigor in Arabidopsis. The Plant Cell 20: 3022-3037.

Oliveira, H., Rodrigues, A., Casquilho, M., Bordado, J. 2008. Impacto do Ataque do Nemátodo da Madeira de Pinheiro na Aptidão Tecnológica como Madeira Maciça. Silva Lusitana 16: 149 – 173.

Pöggeler, S. 2011. Evolution of multicopper oxidase genes in coprophilous and non-coprophilous members of the order sordariales. Current Genomics 12: 95-103.

Poschet, G., Hannich, B., Büttner, M. 2010. Identification and characterization of AtSTP14, a novel galactose transporter from Arabidopsis. Plant and Cell Physiology 51: 1571-1580.

Reiter, W.D., Vanzin, G. 2001. Molecular genetics of nucleotide sugar interconversion pathways in plants. Plant Molecular Biology 47: 95-113.

Ronaghi, M. 2001 Pyrosequencing Sheds Light on DNA Sequencing. Genome Research 11: 3-11.

Roriz, M., Santos, C., Vasconcelos, M.W. 2011. Population dynamics of bacteria associated with different strains of the pine wood nematode Bursaphelenchus xylophilus after inoculation in maritime pine (Pinus pinaster). Experimental Parasitology 128: 357-364.

Sabater, J.A.B., Almagro, L., Belchí, N.S., Barceló, A.R., Pedreño, M.A. 2011. Methyl jasmonate induces extracellular pathogenesis-related proteins in cell cultures of Capsicum chinense. Plant Signaling and Behavior 6: 440-442.

Santos, C., Vasconcelos, M. 2011. Resposta fisiológica de Pinus spp. nas primeiras horas após infecção com Bursaphelenchus xylophilus (Nematoda: Aphelenchoididae). 2011. Silva Lusitana 19: 99-110.

Santos, C., Vasconcelos, M.W. 2012. Identification of genes differentially expressed in Pinus pinaster and Pinus pinea after infection with the pine wood nematode. European Journal of Plant Pathology 132: 407-418.

Schaller, G.E., Shiu, S.H., Armitage, J.P. 2011. Two component systems and their co-option for eukaryotic signal transduction. Current Biology 21: 320-330.

Schnarrenberger, C.,Martin, W. 2002. Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. European Journal of Biochemistry 269: 868–883.

Shi, Z., Maximova, S. N., Liu, Y., Verica, J., Guiltinan, M.J. 2010.Functional analysis of the theobroma cacao NPR1 gene in arabidopsis. BMC Plant Biology 10:248.

Shin, H., Lee, H., Woo, K.S., Noh, E.W., Koo, Y.B., Lee, K.J. 2009. Identification of genes upregulated by pinewood nematode inoculation in Japanese red pine. Tree Physiology 29: 411-421.

Soto, G., Strizler, M., Lisi, C., Alleva, K., Pagano, M.E., Ardila, F., Mozzicafreddo, M., Cuccioloni, M., Angeletti, M., Ayub, N.D. 2011. Acetoacetyl-CoA thiolase regulates mevalonate pathway during abiotic stress adaptation. Journal of Experimental Botany 62: 5699-5711.

Stover, N.A., Dixon, T.A., Cavalcanti, A.R. 2011. Multiple independent fusions of glucose-6-phosphate dehydrogenase with enzymes in the pentose phosphate pathway. PLoS One 6: e22269.

Tadege, M., Bucher, M., Stähli, W., Suter, M., Dupuis, I., Kuhlemeier, C. 1998. Activation of plant defense responses and sugar efflux by expression of pyruvate decarboxylase in potato leaves. The Plant Journal 16: 661-671.

Tomaz, T., Bagard, M., Pracharoenwattana, I., Lindén, P., Lee, C.P., Carroll, A.J., Ströher, E., Smith, S.M., Gardeström, P., Miller, A.H. 2010. Mitochondrial Malate Dehydrogenase Lowers Leaf Respiration and Alters Photorespiration and Plant Growth in Arabidopsis. Plant Physiology 154:1143-1157.

Ulker, B., Somssich, I.E. 2004. WRKY transcription factors: from DNA binding towards biological function. Current Opinion in Plant Biology 7: 491-498.

Vera, J.C., Wheat C.W., Fescemyer H.W., Frilander M.J., Crawford D.L., Hanskii I.M., James H. 2008. Rapid Transcriptome Characterization for a Nonmodel Organism Using 454 Pyrosequencing. Molecular Ecology 17: 1636-1647.

Villa, S.T, Xu, Q, Downie, A.B, Clarke, S.G. 2006. Arabidopsis protein repair L-isoaspartyl methyltransferases: Predominant activities at lethal temperatures. Physiol Plant 128: 581-592.

Wang, Z., Wang, C.Y., Fang, Z.M., Zhang, D.L., Liu, L., Lee, M.R., Li, Z., Li, J.J., Sung, C.K. 2009. "Advances in Research of Pathogenic Mechanism of Pine Wilt Disease." African Journal of Microbiology Research 4: 437-42.

Winter, H., Huber, S.C. 2010. Regulation of Sucrose Metabolism in Higher Plants: Localization and regulation of Activity of Key Enzymes. Critical Reviews in Plant Sciences 19: 31-67.

Wood, T. 1986. Physiological functions of the pentose phosphate pathway. Cell Biochemistry and Function 4: 241–247.

Yamauchi, Y., Hasegawa, A., Taninaka, A., Mizutani, M., Sugimoto, Y. 2011. NADPH-dependent reductases involved in the detoxification of reactive carbonyls in plants. The Journal of Biological Chemistry 286: 6999-7009.

Yang, T., Bar, P.L., Gebhart, L., Lee, S.G., Bar, P.M. 2009. Identification of galacturonic acid-1-phosphate kinase, a new member of the GHMP kinase superfamily in plants, and comparison with galactose-1-phosphate kinase. The Journal Biological Chemistry 284: 21526-21535.

Yan, X., Cheng, X.Y., Wang, Y.S., Luo, J., Mao, Z.C., Ferris, V.R., Xie, B.Y. 2012. Comparative transcriptomics of two pathogenic pinewood nematodes yields insights into parasitic adaptation to life on pine hosts. Gene 12: 0378-1119.

Yazdi H.R., Haznedaroglu, B.Z., Bibby, K., Peccia, J. 2011. Transcriptome sequencing and annotation of the microalgae Dunaliella tertiolecta: Pathway description and gene discovery for production of next-generation biofuels. BMC Genomics 12: 1471-2164.

Yu, K, Zhang, T. 2012. Metagenomic and Metatranscriptomic Analysis of Microbial Community Structure and Gene Expression of Activated Sludge. PLoS One 7: e38183.

Zeczycki, T.N., Maurice, M., Jitrapakdee, S., Wallace, J.C., Attwood, P.V., Cleland, W.W. 2009. Insight into the carboxyl transferase domain mechanism of pyruvate carboxylase from Rhizobium etli. Biochemistry 48: 4305-4313.

Zhang, P., Schon, E.A., Fischer, S.G., Cayanis, E., Weiss, J., Kistler, S., Bourne, P.E. 1994. An algorithm based on graph theory for the assembly of contigs in physical mapping of DNA. Computer Applications in the Biosciences 10: 309-317.