Project title:	The molecular basis of pathogenicity of *Neonectria ditissima*
Project number:	CP 141
Project leader:	Richard Harrison, NIAB EMR. Robert Jackson, Reading University.
Report:	Final report, June 2019
Previous report:	Annual report 2017
Key staff:	Antonio Gomez Cortecero, NIAB EMR
Location of project:	NIAB EMR
Industry Representative:	Tony Harding. Worldwide fruit. Acorn House, Unit 68-69, John Wilson Business Park, Harvey Drive, Chestfield, Whitstable, Kent, CT5 3QT.
Date project commenced:	Oct 2015
Date project completed:	June 2019

DISCLAIMER

While the Agriculture and Horticulture Development Board seeks to ensure that the information contained within this document is accurate at the time of printing, no warranty is given in respect thereof and, to the maximum extent permitted by law the Agriculture and Horticulture Development Board accepts no liability for loss, damage or injury howsoever caused (including that caused by negligence) or suffered directly or indirectly in relation to information and opinions contained in or omitted from this document.

The results and conclusions in this report are based on an investigation conducted over a one-year period. The conditions under which the experiments were carried out and the results have been reported in detail and with accuracy. However, because of the biological nature of the work it must be borne in mind that different circumstances and conditions could produce different results. Therefore, care must be taken with interpretation of the results, especially if they are used as the basis for commercial product recommendations.

AUTHENTICATION

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

[Name]

[Position]

[Organisation]

Signature ..... Date .....

[Name]

[Position]

[Organisation]

Signature ..... Date .....

Report authorised by:

Dr Richard Harrison

Head of genetics, genomics and breeding department

NIAB EMR

Signature .....  ..... Date .....1/7/19.....

[Name]

[Position]

[Organisation]

Signature ..... Date .....

CONTENTS

GROWER SUMMARY

Headline

- The genetic sequences of two isolates of *Neonectria ditissima* have been identified.

Background and expected deliverables

European canker, caused by the phytopathogenic fungus *Neonectria ditissima*, is one of the most destructive diseases of apple and pear. In the orchard, this fungus is able to infect a wide range of apple varieties causing canker and die back of young shoots, resulting in significant losses of fruiting wood. This pathogen has been reported in many apple-producing regions of the world, being especially common in the North-Western European countries. Modern varieties suffer most and in extreme cases do not survive establishment in the orchard. Canker control is difficult to achieve due to the pathogen's lifecycle which is able to infect trees all year-round through wounds, either natural, such as bud-scale scars, leaf scars, fruit scars or artificial, such as pruning wounds. Resistance breeding is underway in many global breeding programmes, but nevertheless a total resistance to canker has not yet been demonstrated in either fruit or woody tissue. There is no known race structure of the pathogen and the global level of genetic diversity of the pathogen population is unknown. Plant resistance is a promising alternative to largely ineffective cultural control but is time consuming to deploy due to the long breeding cycle in apple. Research into other host pathogen interactions shows that a dual strategy of understanding host resistance and pathogen a virulence and how the two are linked, is key to the deployment of durable resistance in the field. Nevertheless, little is known about the pathogen at the molecular level. This project is focused on dissecting components of the pathogen's genome that modulate virulence in order to understand how virulence is controlled and whether there are specific differences in host resistance response to isolates of differing virulence.

This work will provide fundamental insights into the molecular basis of pathogenicity in *Neonectria ditissima,* the causative agent of apple canker. The identification of candidates genes important in virulence in the pathogen and how these genes interacts with the host could lead to novel opportunities for control.

Summary of the project and main conclusions

Two isolates of *N. ditissima* collected in the United Kingdom were sequenced using two different long read sequencing technologies. Genome assemblies of these two isolates improved the contiguity and completeness of the genome, being an excellent resource for the

study of this pathogen. Gene prediction and annotation of these genomes using RNA-Seq data allowed us to present an updated version of the predicted secretome of this pathogen.

During infection, the pathogen secretes proteins, called effectors, to modulate the host cells' response, suppressing defence and allowing colonisation. Analysis of the gene content of the genomes of the two isolates showed a full repertoire of pathogenicity genes, composed of secreted effector proteins and enzymes involved in the degradation of the plant cell walls, distributed throughout the genome. In order to identify specific genes involved in the pathogenicity, a transcriptome analysis was performed. RNA samples from artificial inoculated plants were sequenced and differentially expressed genes during the infection were identified. This analysis revealed a large number of highly expressed genes involved in the degradation of different components of the plant cell walls, mainly polygalacturonan and xylan. In addition to these genes, small secreted proteins with unknown function were also identified. These effector proteins might play a crucial role during infection, modulating the host resistant responses and allowing the colonisation of the host.

Along with the two reference genomes sequenced in this work, twenty-six isolates of *N. ditissima* from eight different countries were sequenced with the Illumina MiSeq system. Understanding the genetic variation of different population of *N. ditissima* is key for the deployment of resistance and orchard management against this pathogen. An analysis of the nucleotide diversity using four polymorphic loci was conducted at the beginning of this project. This analysis revealed slight evidence for population structure of this pathogen. To confirm these findings, analysis of the whole genome sequence of a bigger sample size was performed. This analysis revealed for the first time a clear separation between two populations within the *N. ditissima* species and evidence of hybridization.

Financial benefits

- No direct benefits have been delivered to growers from this study.

Action points for growers

- Due to the nature of this genetic based study, no action points for growers have been identified.

SCIENCE SECTION

Introduction

European canker, caused by the necrotrophic fungus *Neonectria ditissima* (Tul. & Tul.) Samuels & Rossman (synonym *N. galligena*), is one of the most destructive diseases of apple and pear. The fungus attacks trees in the orchard, causing canker lesions on the woody tissues, girdling and killing branches, resulting in loss of fruiting wood and increases pruning costs (Swinburne, 1975). Apple canker can be particularly damaging in young orchards where, in some years, up to 10% of trees can be lost annually in the first few years of orchard establishment as a result of trunk cankers.

Several studies showed the intricate mechanisms of infection of the necrotrophic pathogens. For a successful infection, these pathogens release a large repertoire of lytic and cell wall degrading enzymes to damages the host tissue and acquire nutrients from dead cell. Nevertheless, recent studies have demonstrated that necrotrophic fungi have a more sophisticated infection process (Kim et al., 2015; Liu et al., 2009). The pathogen encodes secreted proteins known as effector which suppress the immune response and allow the colonization of the host. Various effectors have been characterized in necrotrophic fungi, such as *Alternaria Alternata* (Tsuge et al., 2013), *Pyrenophora tritici-repentis* (Ciuffetti et al., 2010) and *Botritys cinerea* (Staats at al., 2007), revealing a wide and diverse repertoire depending on host range and lifestyle.

The previous annual report 2017 described the methodology used in the genome sequencing and assembly of one isolate of *N. ditissima* using long read sequencing technology. This assembly clearly improved the previous versions of the genome and it was used to identify the full effector complement of this pathogen. Functional analysis of this genome showed 291 of secreted Carbohydrate-active enzymes (CAZymes) and other 171 small secreted effectors. CAZymes are responsible for the breakdown of the plant cell wall while other effectors might interact with the host cancelling the immunity response. Therefore, the identification of the specific effectors involved at different stages of the infection will provided a better understanding of how the fungus interferes with the host. In order to identify key pathogenicity genes, grafted trees of two different cultivars, the partially susceptible ‘M9’ and the partially resistant ‘Golden Delicious’, were inoculated with a spore suspension of the isolate Hg199 of *N. ditissima*. During the infection, stem samples with necrotic lesions were collected to capture different stages of the infection. RNA was extracted from each sample and sequenced. The predicted transcriptome of the isolate Hg199 of *N. ditissima* was used as a reference for a differentially expression analysis. This report presents the results of this analysis.

The first report (2016) presented a preliminary analysis of the nucleotide diversity on a global samples of *N. ditissima*, in order to understand if there are differences between geographically distinct populations. This analysis revealed only small differences between the samples tested. This report presents an updated population genetic analysis using whole genome sequences of twenty-nine isolates of *N. ditissima* and one isolate of *Neonectria major*.

Materials and methods

Origin of isolates of *Neonectria ditissima*

This work was performed on a culture collection of 30 isolates (Table 1). All the isolates from the UK were either obtained from the East Malling Research collection or isolated from naturally occurring cankers in woody tissues. Isolates from Italy, Brazil, Belgium and the Netherlands were obtained from living cultures sent by collaborators. Isolates from Ireland and Spain were isolated from infected branches and trunks of local cultivars. Strains from Slovakia and an isolate of *Neonectria major* was obtained from the Westerdijk Fungal Biodiversity Institute in Utrecht, the Netherlands. Draft genome sequences of the isolates RS305p and RS324p are publicly available at DDBJ/EMBL/GenBank under BioProject PRJNA285413 with the accession no. LDPK00000000 for the isolate RS305p and LDPL00000000 for the isolate RS324p (Deng et al. 2015).

Table 1: Isolates used in this study, detailing isolate name, origin and year of isolation.

Isolate accession	Species	Host	CV	Origin	Year of isolation
Ag02	*Neonectria ditissima*	Malus domestica	Jazz	Kent, UK	2016
Ag04	*Neonectria ditissima*	Pyrus communis	Conference	Kent, UK	2016
Ag05	*Neonectria ditissima*	Malus domestica	E830-102	Kent, UK	2016
Ag06	*Neonectria ditissima*	Malus domestica	Unknown	Kent, UK	2016
Ag08	*Neonectria ditissima*	Malus domestica	Jonagored	Tipperary, Ireland	2017
Ag09_A	*Neonectria ditissima*	Malus domestica	Wellant	Tipperary, Ireland	2017
Ag11_A	*Neonectria ditissima*	Malus domestica	Bramley	Armagh, Northern Ireland	2017
Ag11_B	*Neonectria ditissima*	Malus domestica	Bramley	Armagh, Northern Ireland	2017
Ag11_C	*Neonectria ditissima*	Malus domestica	Bramley	Armagh, Northern Ireland	2017
BGV344	*Neonectria ditissima*	Malus domestica	Urtebi	Asturias, Spain	2018
HG199	*Neonectria ditissima*	Malus domestica	Gala	Kent, UK	1999
RS305p	*Neonectria ditissima*	Malus domestica	Brookfield Gala	Lower Moutere, New Zealand	2009
RS324p	*Neonectria ditissima*	Malus domestica	Golden Delicious	Taranaki, New Zealand	2009
ND8	*Neonectria ditissima*	Malus domestica	Royal Gala	Santa Catarina, Brazil	2015

Isolate accession	Species	Host	CV	Origin	Year of isolation
ND9	*Neonectria ditissima*	Malus domestica	Royal Gala	Santa Catarina, Brazil	2015
NM01	*Neonectria major*	Alnus incana	-	Norway	-
OPC304	*Neonectria ditissima*	Malus domestica	Local cultivar	Asturias, Spain	2018
P112	*Neonectria ditissima*	Malus domestica	Carla	Asturias, Spain	2018
R06/17-2	*Neonectria ditissima*	Malus domestica	Bramley	Kent, UK	2017
R06/17-3	*Neonectria ditissima*	Malus domestica	Bramley	Kent, UK	2017
R09/05	*Neonectria ditissima*	Malus domestica	Cox	Kent, UK	2005
R37/15	*Neonectria ditissima*	Malus domestica	Jonagold	Belgium	1999
R39/15	*Neonectria ditissima*	Malus domestica	Unknown	Belgium	-
R41/15	*Neonectria ditissima*	Malus domestica	Wellant	The Netherlands	2015
R42/15	*Neonectria ditissima*	Malus domestica	Elstar	The Netherlands	2015
R45/15	*Neonectria ditissima*	Malus domestica	Elstar	The Netherlands	2015
R68/17-C2	*Neonectria ditissima*	Malus domestica	Gala	Bologna, Italy	2017
R68/17-C3	*Neonectria ditissima*	Malus domestica	Gala	Bologna, Italy	2017
SVK1	*Neonectria ditissima*	Fagus sylvatica	-	Bansky Studenec, Slovakia	2001
SVK2	*Neonectria ditissima*	Fagus sylvatica	-	Poruba, Slovakia	1999

RNA extraction for gene expression analysis

Frozen stem samples were grinded using RNAse-free pestle and mortars in the presence of liquid nitrogen. Total RNA was isolated using the QIAGEN RNEasy Plant Mini kit (QIAGEN Inc., Valencia, CA) according to the manufacturer's instructions. The RNA concentration and quality were initially assessed on a NanoDrop 100 spectrophotometer (Thermo Scientific, Waltham, MA, U.S.A) by measuring absorbance at 230, 260 and 280 as indicators of purity and if there is a presence of other co-purified contaminants. RNA concentration was also quantified using the RNA HF (High sensitivity) assay kit for the Qubit II Fluorometer (Thermo Fisher Scientific, Waltham, MA, U.S.A.). Finally, RNA integrity was assessed on a TapeStation 4200 (Agilent Genomics, Santa Clara, CA, U.S.A).

RNA sequencing

For each cultivar, three replicate samples were used for sequencing. Each replica consists in one proximal and one distal sample to the inoculation point. In addition, RNA samples from *in vitro* grown mycelium and three replicate samples of mock inoculated trees per cultivar were sequenced. The choice of the RNA samples was done based on their quality and integrity. The samples were sequenced by Novogene (Novogene, Hong Kong) on an Illumina HiSeq 4000 sequencer.

Gene prediction of the reference genomes

RNA-Seq reads obtained from the *in vitro* growing cultures were aligned to the reference genome of the isolate Hg199 of *N. ditissima* using STAR version 2.5.3a (Dobin et al. 2013). Alignment files were concatenated in a unique file using Samtools version 1.5 (Li et al. 2009) with the merge command and will be use as a training set for the gene prediction. BRAKER1 version 2.0 (Hoff et al. 2016) is the programme used to create gene models combining predictions from GeneMark-ET (Lomsadze et al. 2014), installed witn the package GeneMark-ES/ET version 4.38, and AUGUSTUS version 3.1 (Stanke et al. 2008) using the information of the RNA-Seq. GeneMark-ET performs an unsupervised training and generates an initial predicted gene set. This set is used to train AUGUSTUS, which also integrates RNA-Seq read information for an accurate prediction (Hoff et al. 2016). Gene prediction was performed on the softmasked genome of *N. ditissima* with the fungus option of BRAKER1 enabled. CodingQuarry version 2.0 (Testa et al. 2015) was also used to generate gene models. Pathogen mode was enabled to assist the identification of additional effector-like genes in fungal plant pathogens, using a transcriptome assembly produced by Cufflinks version 2.2.1 (Trapnell et al. 2010) for genuine predictions. CondingQuary gene models predicted in intergenic regions and not contained in BRAKER1 predicted genes were added to the gene models using BEDtools Intersect function (Quinlan and Hall 2010).

Functional annotation and effector prediction

Draft functional annotations were determined for gene models using InterProScan version 5.18-57.0 (Jones et al. 2014). Homology to the predicted genes models was identified through BLASTp searches (Altschul et al. 1990) against the March 2018 release of the SwissProt database (UniProt Consortium 2018) with an e-value threshold of 1x10-100 and in the Pathogen-Host interaction database (PHIBase) (Winnenburg et al. 2006) using BLASTx with an e-value 1x10-30. Secretory signal peptides were identified in the gene models using SignalP version 4.1 (Petersen et al. 2011). Those contained transmembrane domains were identified using TMHMM version 2.0 (Krogh et al. 2001) and removed using a custom Python script. Specific fungal effector were detected using EffectorP version 1.0 (Sperschneider et al. 2016), a machine learning tool able to discriminates between secreted fungal effectors from secreted noneffectors based on characteristics of length, molecular weight and protein net charge. Carbohydrate active enzymes (CAZymes) were predicted using HMM models from the dbCAN HMM version 7 database (Yin et al. 2012; H. Zhang et al. 2018) with searches against predicted proteins performed using the hmmscan program from the HMMER (Finn et al. 2011) version 3.1b2 suite. Antismash 4.0 webserver (Blin et al. 2017) was used to identify clusters of secondary metabolites genes within the genomes. The Cluster

Assignment by Island of Sites or CASSIS algorithm was enabled to identify genes with common pathway-specific regulatory motifs. AntiSMASH results were converted to gff3 format and predicted secondary metabolites genes identified using BEDtools Intersect function.

RNA-Seq data preparation

RNA-Seq data from Novogene was transferred to the NIAB EMR cluster for the analysis. Adaptor sequences and low-quality data were removed using fastqc-mcf (Aronesty 2013). RNA-Seq data quality was evaluated using the quality control tool FastQC version 0.10.1 (Andrews 2010).

Quantification of the expression of transcripts was done using Salmon version 0.9.1 (Patro et al. 2017). Salmon works by mapping RNA-Seq data directly to a given transcriptome using a quasi-mapping approach (Srivastava et al. 2016) for a fast an accurate quantification of transcript-level abundance. The predicted transcriptome of the isolate Hg199 of *N. ditissima* and the transcriptome of the 'Golden Delicious' doubled-haploid tree GDDH13 (Daccord et al. 2017) were used for the identification of differentially expressed genes in the pathogen and host. Firstly, Salmon created an index of each transcriptome with the option to keep duplicated enabled. Then, Salmon runs were performed for each sample with 1000 bootstrap replicates and the dumpeq, seqbias and gcbias flags enabled.

Differentially expressed genes (DEG) in *N. ditissima*

Differentially expression analysis was performed with DESeq2 package (Love et al. 2014) in R software 3.5.1. Prior to the analysis, low counts genes were removed in the pathogen dataset using a threshold of 50 reads in at least three samples. This was done to avoid zero counts values across the conditions, mainly observed in the distal samples, which lead to false positive log2 fold change (LFC) values. For the analysis of the pathogen, individual samples from each cultivar and sampling position were nested within groups and differentially expressed genes (DEG) assessed compared to *in vitro* growing mycelium samples. The false discovery rate (FDR) cutoff was set to 0.05. For visual inspection of samples distances, *regularized logarithm* was used to transform the read counts and principal component analyses (PCA) performed using R (Team 2015).

Bayesian phylogenetic analysis

BUSCO hits of single copy Sordiaromycetes conserved genes were assessed to evaluated completeness in the assemblies. Single hits identified in all the sequenced genomes of *N. ditissima* and *N. major* performed in this study and in the 2 publicly available genomes from

New Zealand (Deng et al. 2015) were extracted. Sequences from every loci were aligned using MAFFT version 7.222 (Katoh and Standley 2013) and alignments were trimmed using trimAL version 1.4.1 (Capella-Gutiérrez et al. 2009) enabling an heuristic method to automatically decide the appropriate trimming method. Gene phylogenetic trees were constructed with a maximum likelihood approach using RAxML version 8.1.17(Liu et al. 2011; Stamatakis 2014) using 1000 bootstraps replicates to identify the best scoring ML trees at each locus. A single consensus species tree was generated from each RAxML run using ASTRAL version 5.6.1(C. Zhang et al. 2018) and visualised using the R package GGtree version 1.14.6 (Yu et al. 2016; Yu et al. 2018).

Alignment of Illumina sequencing reads to the reference assembly of the R0905 isolate of *N. ditissima*

The illumina reads of all of the isolates of *N. ditissima* were aligned to the reference genome assembly of the isolate R0905 using Bowtie2 version 2.2.6 (Langmead and Salzberg 2012). Multimapping reads tagged with the XS:i alignment scored were removed. “Paired reads” and “properly paired reads” were kept using SAMtools version 0.1.18 (Li et al. 2009). PCR and optical duplicates were also removed using Picard tools version 2.5.0 (Broad Institute 2019).

Single nucleotide polymorphism calling and annotation

SNPs and indels were identified using the Genome Analysis Toolkit (GATK) version 3.6 HaplotypeCaller (McKenna et al. 2010), setting ploidy argument to one for haploid organisms. Complex multi-nucleotide polymorphisms (MNPs) were converted into separated SNPs by the VariantsToAllelicPrimitives tool from GATK. A variant calling file was produced containing information of every SNPs identified per isolate. Low coverage samples make difficult to accurately identify variants. The coverage of the isolate Ag11_B of *N. ditissima* was estimated in 22.29X, notably lower to the other isolates used in this study. Therefore, variant information from this isolate was removed using the vcfremovesamples function from vcflib (Garrison 2012). Only biallelic high quality SNPs with no missing data were retained for genetic analyses using vcffilter function from vcflib and setting the following filtering option: Minimum quality of 30, minimum MQ of 30, minimum depth of 10 and minimum GQ of 30. VCFTools (Danecek et al. 2011) was used to remove Indels and site with more than 5% of missing data. Basic statistics such as the number of SNPs were calculated using the vcf-stats tool from VCFTools. Percentage similarity of shared alleles between samples was calculated using a custom Python script. Results were visualised using the gplots package (Warnes et al. 2016) in R (Team 2015). Monomorphic sites were removed setting a minimum minor allele count of 1 were removed using VCFTools (Danecek et al. 2011). The resulting SNP matrix was used

to perform a principal component analysis (PCA) and to create a neighbour joining tree. PCA was performed in R using the SNPRelate (Zheng et al. 2012), gdsfmt (Zheng et al. 2012), ggplot2 (Wickham 2016) and ggrepel (Slowikowski 2018) R packages. SNPs were concatenated into a fasta alignment using a Perl script (Bergey 2012). These fasta alignments were used to build the neighbour joining tree using the ape package (Paradis and Schliep 2019) in R with 100 bootstrap replicates. A SnpEff database was built for the R09/05 genome assembly using the predicted gene annotations. Variants sites were then annotated based on their genomic locations using the SnpEff program (Cingolani et al. 2012).

Analysis of the population structure

The existence of Population structure in the samples of *N. ditissima* was evaluated using the *structure* software (Pritchard et al. 2000; Porras-Hurtado et al. 2013). Variant sites information of the isolate of *N. major* were removed for this analysis using vcfremovesamples function from vcflib (Garrison 2012). Random samples sites were retained for the structure analysis using vcfrandomsample function from vcflib with a 0.1 base sampling probability per locus. If all markers are retained this will be uninformative because of the linkage. Variant calls file was transformed into the *structure* input format using PGDSpider version 2.1.0.3 (Lischer and Excoffier 2012). The *structure* file was edited adding a PopData column with the information of the origin of each isolate. Structure allows the user to customize the program parameters editing two files, mainparams and extraparmas, which are read during the execution of the program. The *mainparams* file can be used to define type of data present in the input file and set basic run parameters. The *extraparams* file specifies the type of analysis and how the program will run. Once parameters had been defined, Structure version 2.3.4 was executed with the K populations values to test, from one to eight, in five replicated runs. Structure results were visualized using Structure Harvester (Earl and vonHoldt 2012) and the number of real K population was determined enabling the Evanno method (Evanno et al. 2005). Cluster membership coefficients of every replicated runs were permuted using CLUMPP version 1.1 (Jakobsson and Rosenberg 2007) with the Greedy algorithm. Resulting membership coefficients were displayed graphically using DISTRUCT version 1.1 (Rosenberg 2004).

Results

De novo genome assembly of reference genomes of *N. ditissima* using long read data improved the contiguity compared to the previous assemblies.

Assemblies using long-read sequencing data for the isolate Hg199 and R09/05 of *N.ditissima* improved the contiguity compared to the short read assemblies. For the isolate Hg199, the

assembled genome was 45 Mb in size, consisting in 148 contigs, while the genome assembly of the isolate R09/05 was 46 Mb in size, consisting of 50 contigs (Table 2). To evaluate the genome completeness, highly conserved genes in Sordiariomycetes were searched in our assemblies. Of the known 3725 single copy genes in the sordiariomycetes database, 98.65% and 98.57% were identified in the genome of the isolate Hg199 and R09/05 respectively.

Genome sequencing of 25 additional isolates of *N. ditissima* and one of *N. major* were performed on the Illumina Miseq platform. Genome size of all the *N. ditissima* isolates ranged from 43 Mb to 46 Mb and were similar to the two publicly available New Zealand genomes (Deng et al. 2015). The isolate R09/05 was re-sequenced and reads were pulled together with the previous batch, used for the first assembly (Gómez-Cortecero et al. 2015). This new assembly resulted in similar size and smaller number of contigs. The assembly size of *N. majo*r was smaller with a genome size of 41.5 Mb. Number of contigs is greater on these assemblies compared to the long-read assemblies, ranging from 410 to 1100.

Table 2: Reference genome assemblies statistics

Isolate accession	Assembly size (Mb)	Number of Contigs ≥ 1000bp	Largest contig (Kb)	N50	Repeatmasked (Kb)	Repeamasked (%)	Conserved Sordiariomycetes genes in genome (%)
Hg199	45,02	148	2,079	167	4,881	10.84	98.65
R09/05	46,00	50	3,470	171	5,386	11.71	98.57

Gene prediction identified a large effector repertoire associated with the necrotrophic lifestyle of *N. ditissima*.

Gene predictions resulted 13975 – 16646 protein-coding genes from the assemblies of *N. ditissima* isolates (Table 3). The genome assembly of the isolate of *N. major* resulted in 12315 protein-coding genes. Following gene predictions, key effectors genes were identified using different programs. Proteins without transmembrane domains and containing signal peptides were identified and regarded as putative secreted proteins. Secreted proteins resulted in 989 – 1058 from the *N. ditissima* isolates. Among these secreted proteins, effector-like proteins and CAZy were also identified. Predicted secreted effector proteins resulted in 159 – 245, while secreted CAZy identified resulted in 271-291 among the isolates of *N. ditissima* (Table 2).

Table 3: Predicted genes in the *N. ditissima* and *N. major* isolates. Number of secreted proteins, secreted effectors (EffectorP) and secreted carbohydrate active enzymes (CAZy) are shown.

Isolate accession	Species	Host	Total genes	Total proteins	Secreted proteins	Secreted EffectorP proteins	Secreted CAZYme proteins
Ag02	*Neonectria ditissima*	Malus domestica	13782	14052	994	159	280
Ag04	*Neonectria ditissima*	Pyrus communis	13737	14022	1001	168	281
Ag05	*Neonectria ditissima*	Malus domestica	13723	14013	996	163	286
Ag06	*Neonectria ditissima*	Malus domestica	13785	14028	997	173	273
Ag08	*Neonectria ditissima*	Malus domestica	14861	15144	1058	217	289
Ag09_A	*Neonectria ditissima*	Malus domestica	14462	14725	1013	177	280
Ag11_A	*Neonectria ditissima*	Malus domestica	14774	15043	1050	212	286
Ag11_B	*Neonectria ditissima*	Malus domestica	13964	14227	1019	180	287
Ag11_C	*Neonectria ditissima*	Malus domestica	16351	16646	1096	245	279
BGV344	*Neonectria ditissima*	Malus domestica	13921	14190	1000	164	283
HG199	*Neonectria ditissima*	Malus domestica	14630	14842	1009	167	287
ND8	*Neonectria ditissima*	Malus domestica	13751	14017	996	172	282
ND9	*Neonectria ditissima*	Malus domestica	13775	14044	1005	169	283
NMaj	*Neonectria major*	Alnus incana	12238	12315	936	137	271
OPC304	*Neonectria ditissima*	Malus domestica	13774	14024	1016	173	283
P112	*Neonectria ditissima*	Malus domestica	15975	16249	1029	174	284
R06/17-2	*Neonectria ditissima*	Malus domestica	13808	14083	1005	171	287
R06/17-3	*Neonectria ditissima*	Malus domestica	13911	14203	989	165	282
R09/05	*Neonectria ditissima*	Malus domestica	15008	15157	1032	171	291
R37/15	*Neonectria ditissima*	Malus domestica	13724	14009	994	162	282
R39/15	*Neonectria ditissima*	Malus domestica	13696	14001	1019	171	279
R41/15	*Neonectria ditissima*	Malus domestica	13742	13997	1001	176	283
R42/15	*Neonectria ditissima*	Malus domestica	13797	14076	1009	171	284
R45/15	*Neonectria ditissima*	Malus domestica	13752	14034	997	167	284
R68/17-C2	*Neonectria ditissima*	Malus domestica	13736	13978	1001	166	284
R68/17-C3	*Neonectria ditissima*	Malus domestica	14726	14979	1030	187	285
SVK1	*Neonectria ditissima*	Fagus sylvatica	13764	14002	1023	190	290
SVK2	*Neonectria ditissima*	Fagus sylvatica	13724	13975	1023	173	289

Transcriptome analysis in the Hg199 isolate of *N. ditissima* showed a large number of differentially expressed transcripts compared to in vitro mycelium.

A total of 21 RNA samples were obtained from the gene expression experiment described in the previous report. Transcript quantification of the RNA-seq was performed using the predicted transcriptome of the isolate Hg199 of *N. ditissima* as a reference. A total of 74 million of compatible fragments were obtained from mycelium samples, 33 million from the samples proximal to the inoculation point and 180 thousand from the distal samples. A principal component analysis (PCA) of DeSeq2 was produced to visualize genetic distance

and outliers among all the RNA-Seq samples using an rlog transformation of the data (Figure 1). The largest principal component, PC1, explained 75% of the total variance in the dataset and provided clear separation between samples from the inoculation experiment and samples from *in vitro* grown mycelium. The smaller principal component, PC2, explained that 11% of the variation correspond to the differences between proximal and distal samples of the inoculation experiment.

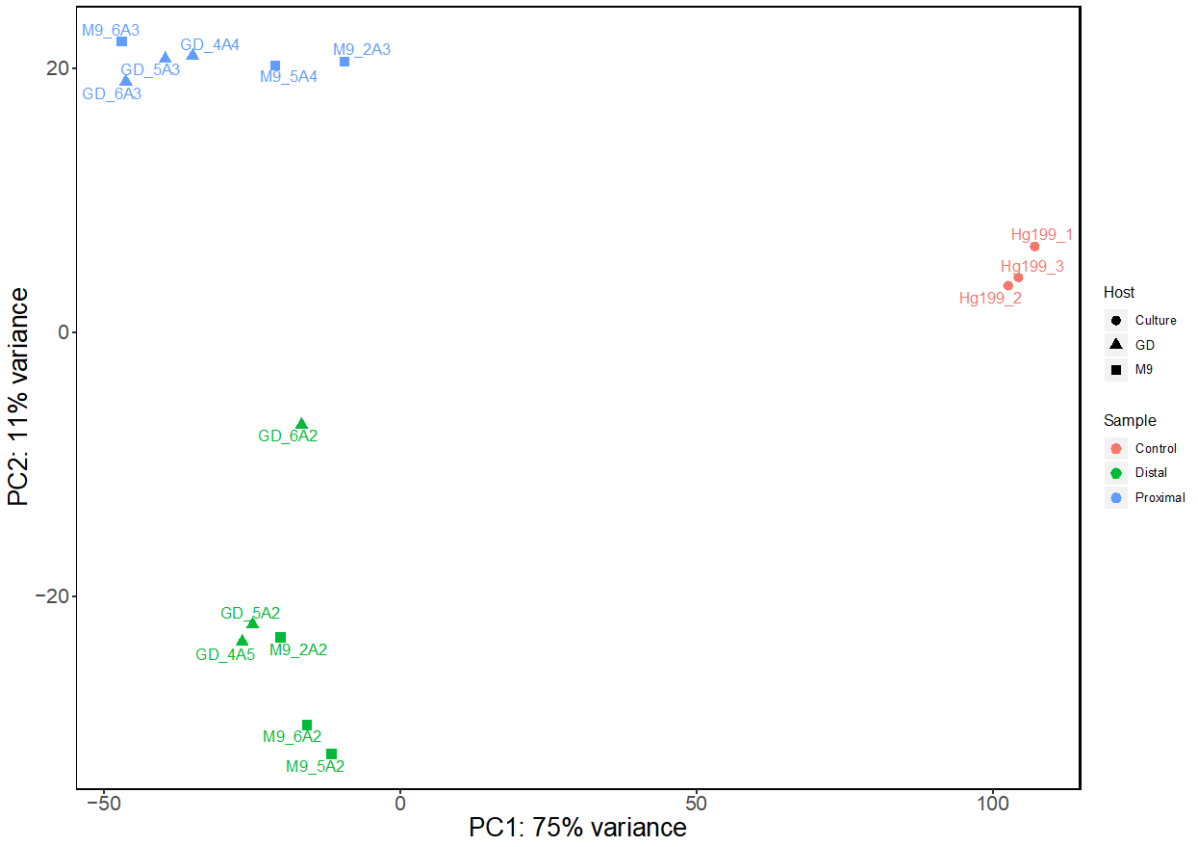

Figure 1: Principal component analysis (PCA) using rlog transformed data of 15 samples. Predicted transcripts were quantified using Salmon (Patro et al. 2017) and differential expression was identified with DeSeq2 (Love et al. 2014).

Differentially expressed genes were calculated with a false discovery rate threshold < 0.05. The number of differentially expressed genes of every group of samples were plotted in a Venn diagram (Figure 2). A total of 5308 genes showed differentially expression compared to the *in vitro* mycelium samples. Proximal samples showed more differentially expressed genes than distal samples for both cultivars, which correlates with the differences of compatible fragments obtained after the quantification. Despite of this, 194 unique DEGs were identified in the distal samples. Genes presenting $\log_2$ fold change (LFC) greater than one or less than minus one were regarded as important. A total of 1924 genes showed LFC

greater than one, representing up-regulation. On the other hand, 3389 genes showed LFC below minus one, representing down-regulation.

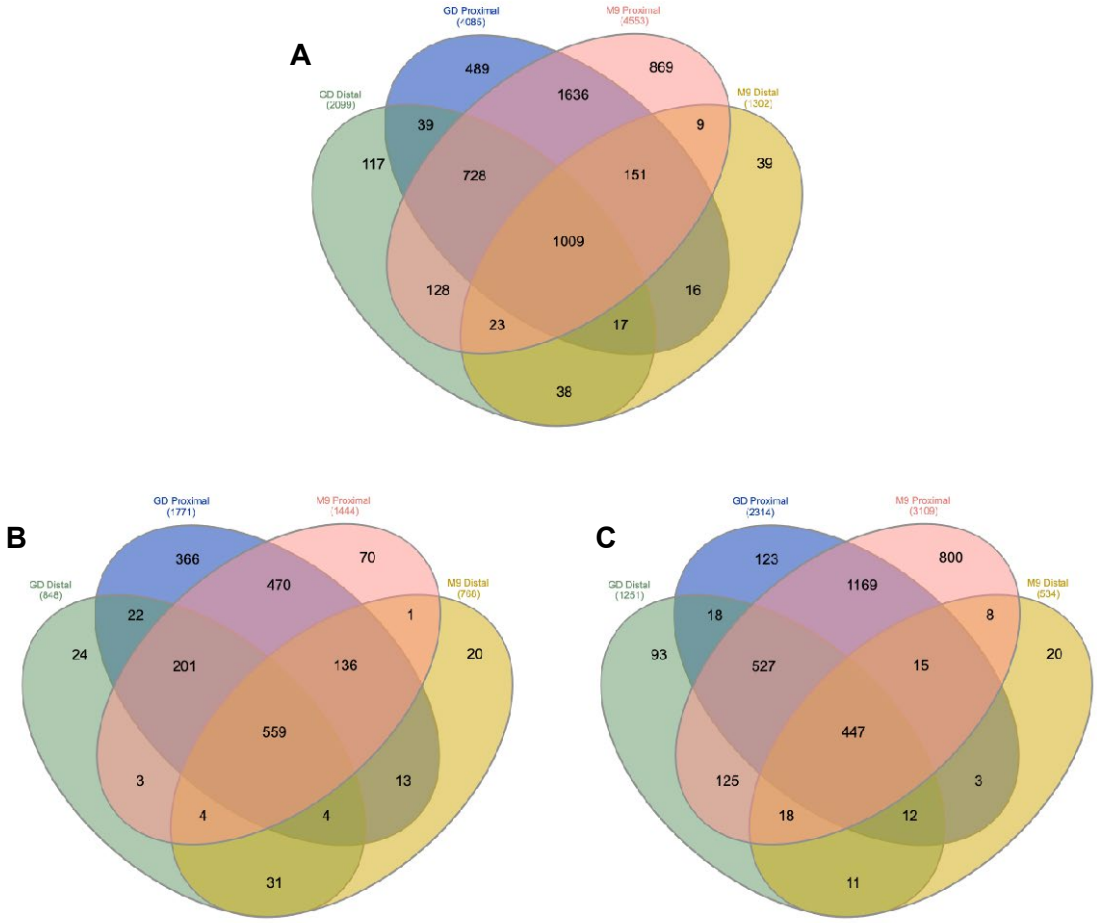

Figure 2: Venn diagram of all differentially expressed transcripts in the Hg199 isolate of N. ditissima. Predicted transcripts were quantified using Salmon (Patro et al. 2017) and differential expression was identified with DeSeq2 (Love et al. 2014). A: All differentially expressed genes. B: All expressed genes displaying a LFC greater than 1. C: All expressed genes displaying a LFC less than -1.

Analysis of expression in the Hg199 isolate of *N. ditissima* allowed the identification of candidate pathogenicity genes.

For a successful infection, necrotrophic pathogens, like *N. ditissima*, rely on an arsenal of lytic and cell wall degrading enzymes to damage the host tissue and small secreted proteins that might suppress immune responses and allow the colonization of the host. Particularly, the genome of the Hg199 isolate of *N. ditissima* encodes 1009 secreted proteins, of which 167 were identified as secreted effector proteins and 287 as secreted (Table 2). Therefore, it was

interesting to identify the proportion of these effector proteins that showed evidence of expression in our datasets.

Analysis of differential expression of CAZymes revealed a large number of enzymes involved during the infection, showing up-regulation in 126 transcripts and down-regulation in 30 (Figure 3). Interestingly, 72 transcripts were highly upregulated in every group of samples, indicating their importance during the infection.

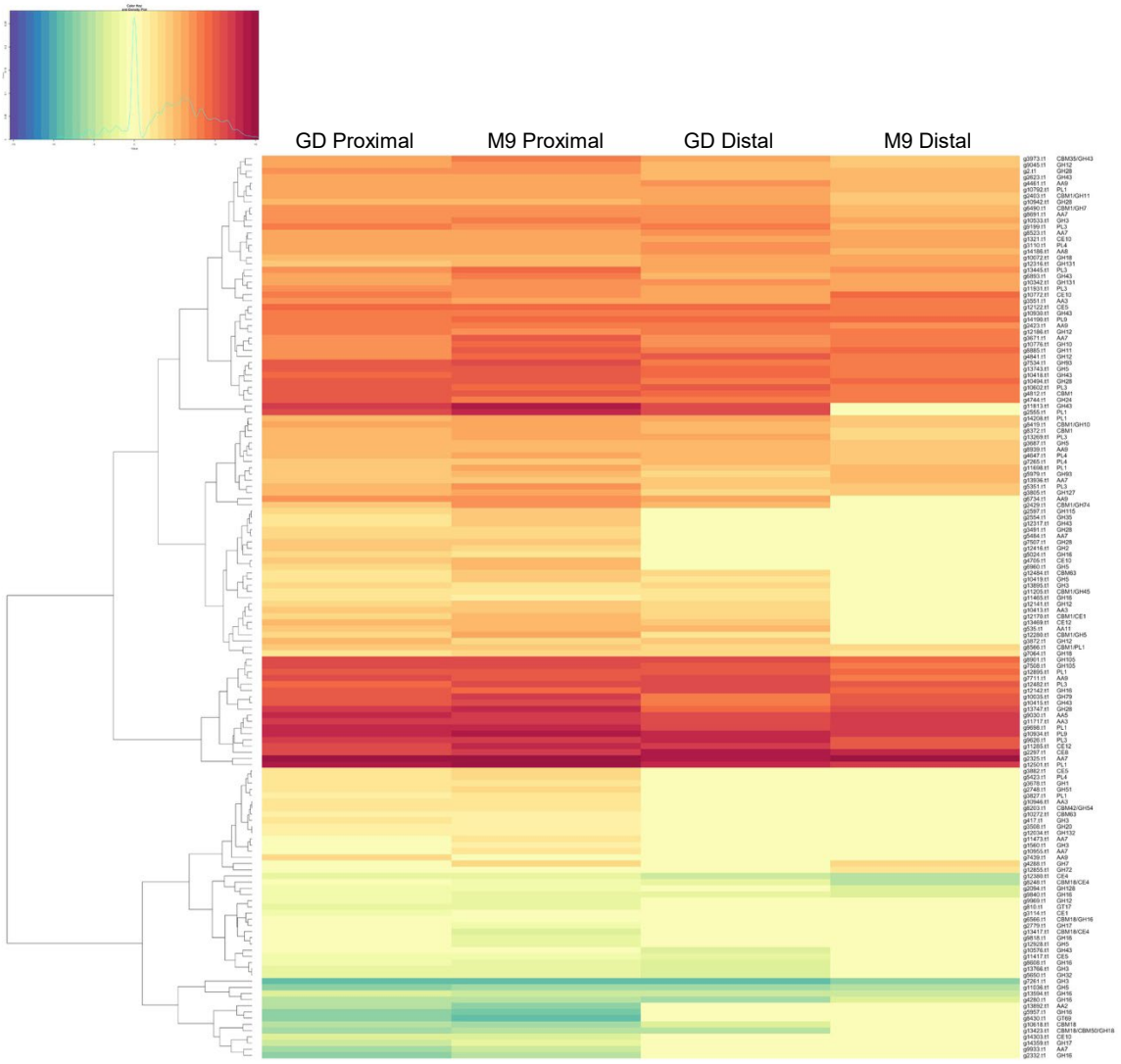

Figure 3: Expression pattern of secreted CAZymes. Relative expression pattern of the genes in the four groups were clustered and scaled by colour using heatmap.2 package in R. The different expression patterns are labelled in a colour scale. Orange represent up-regulated genes and green represent down-regulated genes.

CAZymes are involved on many biological processes and can be classified into different categories based on their role. According to this, it will be interesting to investigate the role of

the main differentially expressed CAZyme in order to understand the infection strategy at different stages. The repertoire of CAZymes of the Hg199 isolate of *N. ditissima* consists in 58 families of glycoside hydrolase (GH) with 314 genes, 10 families of carbohydrate esterases (CE) with 143 genes, 34 families of glycosyltransferases (GT) with 110 genes, 6 families of polysaccharide lyase (PL) with 33 genes, 20 families of carbohydrate-binding modules (CBM) with 82 genes and 12 families of auxiliary activities (AA) in 108 genes.

Most of the secreted CAZymes were up-regulated during the infection. Analysis of differentially expressed CAZyme based on these categories revealed a major expression of GH category, specially families GH28 and GH43 which are involved in the degradation of polygalacturonan and xylan respectively, present in the plant cell walls (Figure 4). Polysaccharide lyase have also been found highly expressed in all the groups. The most abundant families are PL1 and PL3, both involved in the pectin hydrolysis. Despite that only a few were predicted in the genome of *N. ditissima*, these results suggested that they are key components the during the infection inducing the cell death in plants (Yang et al. 2018). The most abundant family of the CBM category was CBM1, which has cellulose-binding function (Johansson et al. 1989).

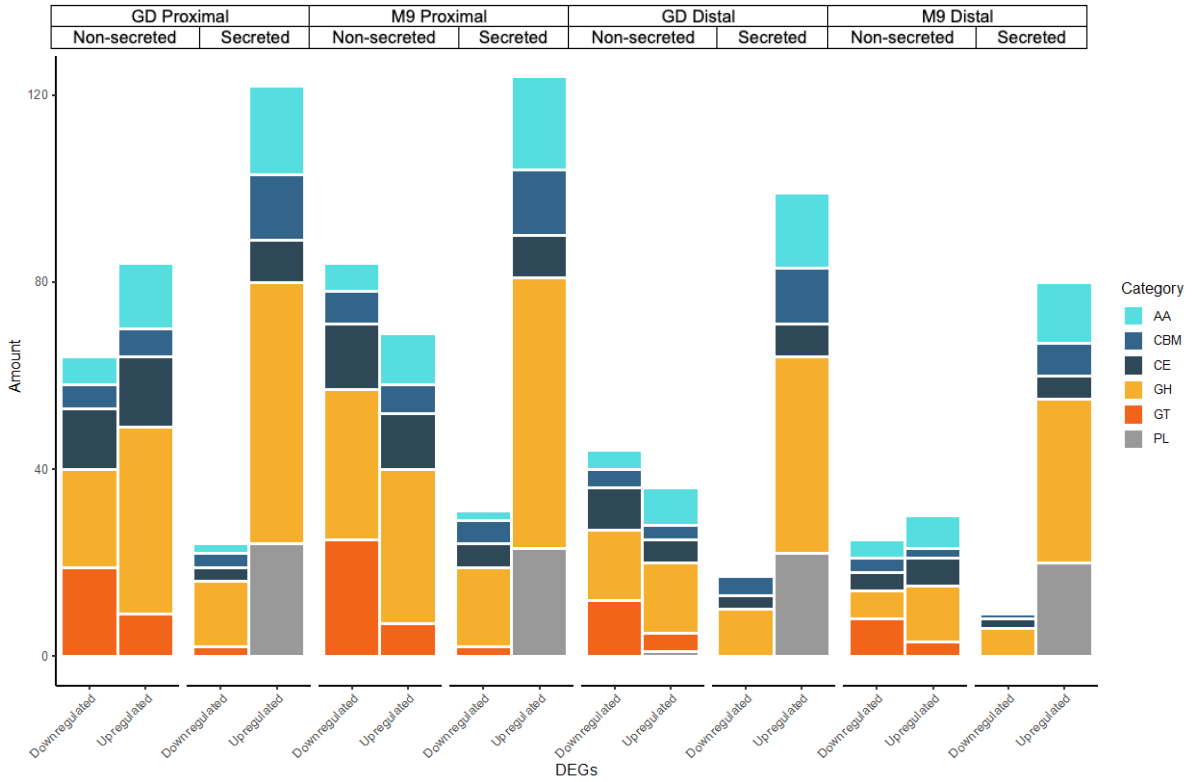

Figure 4: Differentially expressed CAZyme on every group. Categories represented: auxiliary activities (AA), carbohydrate-binding modules (CBM), carbohydrate esterases (CE), glycoside hydrolase (GH), glycosyltransferases (GT) and polysaccharide lyase (PL).

Additional effector proteins were identified using EffectorP. Secreted proteins that had length, net charge and amino acid content typical of fungal effector were selected and levels of expression identified. A total of 44 transcripts were differentially expressed (Figure 5). Among these, 18 showed up-regulation and 26 down-regulation. Although the proportion of these effectors is smaller compared with the CAZymes, it was possible to identify differences in the expression across the groups.

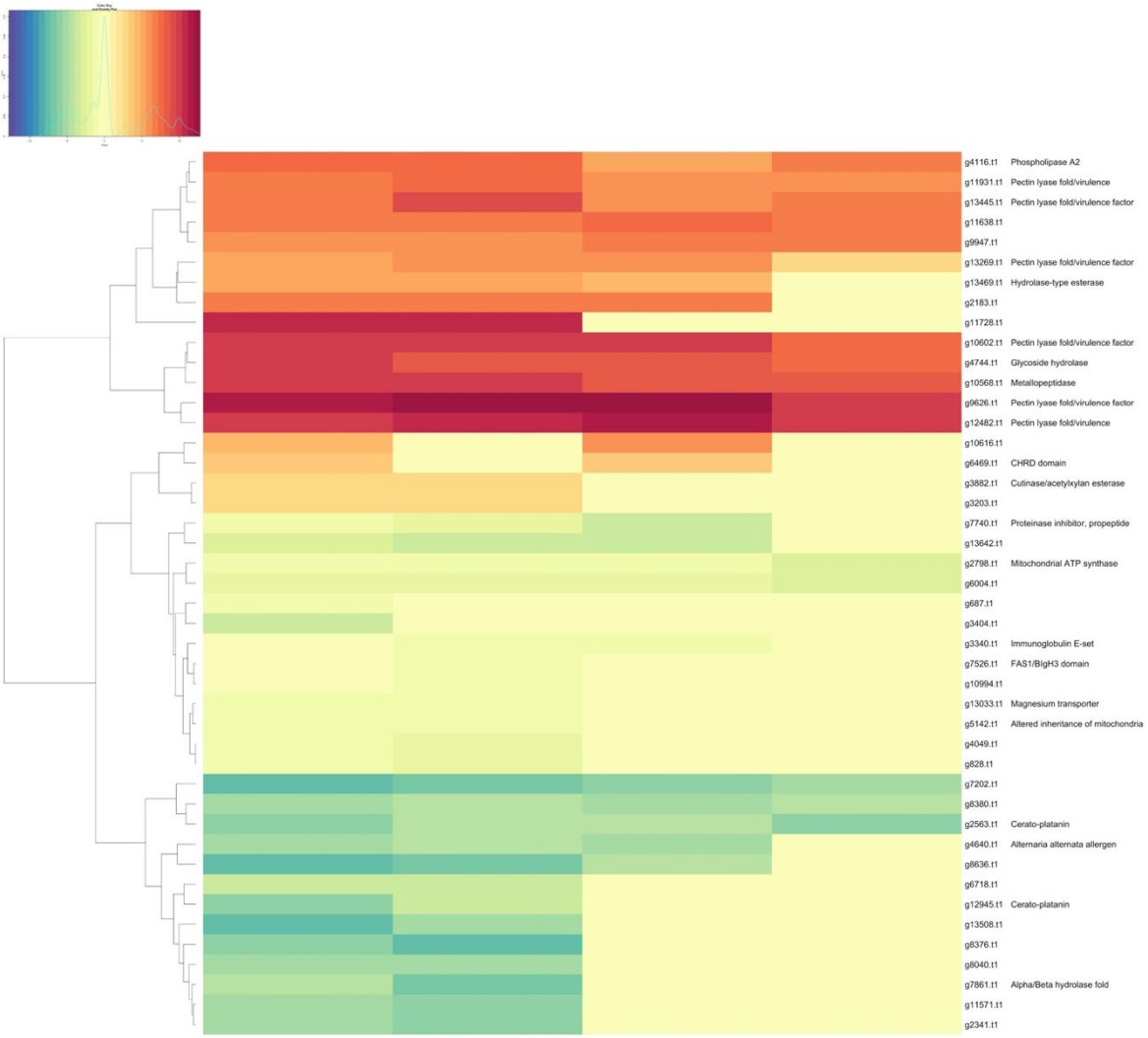

Figure 5: Expression pattern of secreted effectors predicted with EffectorP. Relative expression pattern of the genes in the four groups were clustered and scaled by color using heatmap.2 package in R. The different expression patterns are labelled in a colour scale. Orange represent up-regulated genes and green represent down-regulated genes.

Differences observed between expression patterns of Proximal and Distal revealed a several effectors that are released in an early stage of the infection. Interestingly, the transcripts g10616.t1 and g6469.t1, with unknown function, were expressed only in Golden delicious

samples. This might represent an additional effort by *N. ditissima* to colonize a more resistant variety.

Phylogenetic analysis of sequenced isolates revealed groups separation within the *N. ditissima* population.

The relationship between the 26 sequenced isolates of *N. ditissima* and *N. major* was investigated through phylogenetic analysis of core sordariomycetes genes. The publicly available genome sequences of two strains of *N. ditissima,* the isolate RS305p and the isolate RS324p, from New Zealand (Deng et al. 2015) were included in this analysis. As expected, the resulting phylogeny showed the isolates of *N. ditissima* grouping together and separated to the isolate of *N. major* (Figure 6). Interestingly, the isolates from Italy and Slovakia formed a separated group from the majority of *N. ditissima* isolates. Information regarding isolate origin is shown in Table 1.

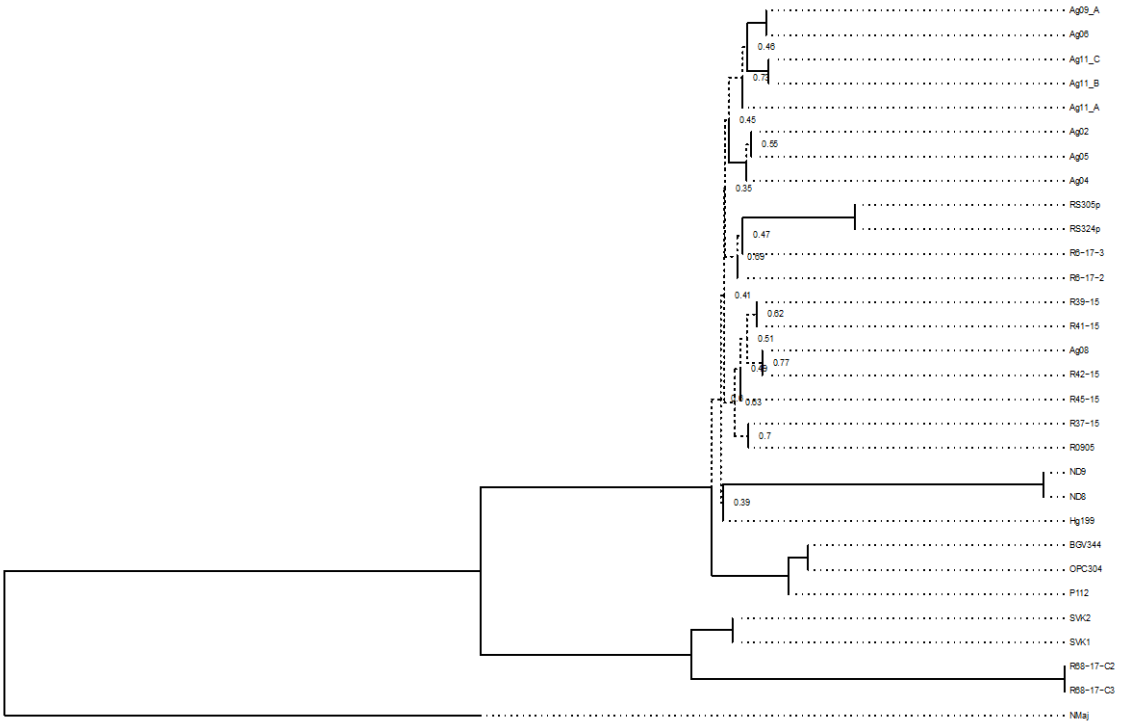

Figure 6: Phylogeny of sequenced and publicly available genome sequences of *N. ditissima* and *N. major* Maximum parsimony consensus phylogeny of 3725 conserved single copy loci.

Identification of variant sites demonstrates separation within *N. ditissima* specie.

In order to investigate genetic differences across all *Neonectria* genomes, variant sites on every genome were identified and annotated. A total of 1291711 SNP sites were identified. Only high-quality biallec SNPs were retained for downstream analyses. After filtering 800178 sites were retained. This was also calculated removing the isolate of *N. major,* retaining 932641 sites. Monomorphic sites were removed retaining 626425 SNPs across all the isolates and 316536 SNPs only in *N. ditissima* isolates. The percentage of shared SNPs was calculated and a heatmap of pairwise comparison between each isolate of *N. ditissima* was plotted (Figure 7).

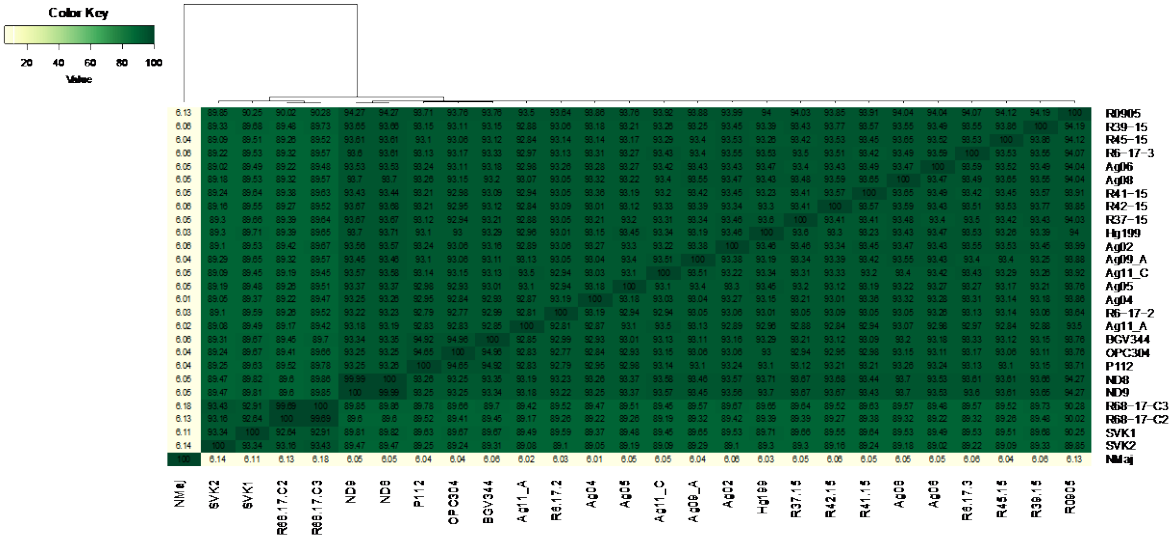

Figure 7: Heatmap showing pairwise comparisons of the percentage of all high quality biallelic SNP sites with the same allele for all the sequenced isolates of *N. ditissima* and *N. major*. Heatmaps were plotted using the gplots R package.

A PCA plot was also performed with the results (Figure 8). The plots showed a separation between the two *Neonectria* species. Principal component 1 explained 27.04% of the variance across the samples and separated them by species. Interestingly, the isolates SVK1 and SVK2 from Slovakia and the isolates R68/17-C2 and R68/17-C3 from Italy formed a separate group from the other *N. ditissima* isolates. In this case, PC2 explained 12.11% of variance separating the samples into these two groups.

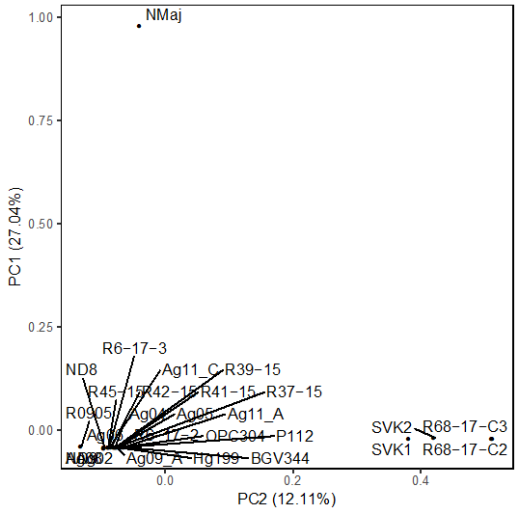

Figure 8: Principal component analysis of all high quality biallelic SNPs for all the sequenced isolates of *N. ditissima* and *N. major*. Heatmaps were plotted using the ggplots R package.

Investigation of the population structure within the sequenced isolates of *N. ditissima* showed two distinct population and evidence of recombination event.

SNP analysis revealed separation between isolates of *N. ditissima.* Genetic structure of the isolates of *N. ditissima* was assessed using the Bayesian clustering approach of the STRUCTURE software (Pritchard et al. 2000; Evanno et al. 2005). This clustering approach assign individuals into K possible populations. Assuming admixture model in the population and without considering the geographic origin of the isolates, the most probable number of populations (K) is calculated. Our results showed two different genetic cluster among the isolates of *N. ditissima*. A distruct plot (Rosenberg 2004) was created displaying the population structure (Figure 9). Once more, a main population consisting in isolates from United Kingdom, Ireland, the Netherlands, Belgium, Brazil and Spain was observed. The second population consisted in isolates from Italy and Slovakia. Interestingly, the two isolates from Slovakia present characteristics of both populations. This might suggest the present of a hybrid population.

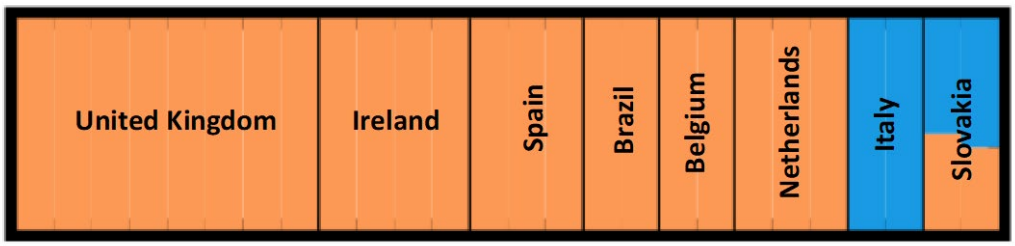

Figure 9: Distruct plot of STRUCTURE results conducted on high quality biallelic SNPs for all the sequenced isolates of *N. ditissima*. Each colour represents a different population.

Discussion

The causal agent of European Canker, *Neonectria ditissima,* has become a globally important plant pathogen of apple causing significant losses across all major apple-producing countries in the world. Many historic studies have been carried out on tree resistance and a range of quantitative pathogenicity tests have been developed to phenotype cultivars resistance response to *N. ditissima* (Alston 1969; Borecki and Czynczyk 1985; Garkava-Gustavsson et al. 2013; Garkava-Gustavsson et al. 2016; Kraehmer and Schmidle 1979; Krüger 1983; van de Weg 1989; van de Weg et al. 1992; Wenneker et al. 2017). While these tests are focused in the evaluation of different cultivars, hence enable breeding for resistance to the disease, not much has been done in the understanding the source of pathogenicity in *N. ditissima*. Moreover, the molecular study of this pathogen has been hampered for the lack of a reference genome sequence.

Only three uncompleted genome sequences of *N. ditissima* are currently publicly available. Even though these assemblies are highly fragmented, they may still be used effectively in population genetic studies or characterization of gene complements. The failure to resolve large-scale genome structure of these isolates is understood to be mediated by repetitive transposable elements (Koren and Phillippy 2015). Therefore, one of the main aims of this work was the generation of a high-quality reference genome of *N. ditissima.* This has been considered as a preliminary step in the molecular study of this pathogen and will facilitate the identification of the full effector complement and their physical location within the genome. Our results showed a big improvement in the contiguity of the two reference genomes used in this study due to the use of third-generation long read sequencing technologies.

Gene predictions of multiple genome sequences of *N. ditissima* revealed the presence of multiple genes associates with the necrotrophic lifestyle. Most of these genes were characterized to be involved in the plant cell degradation wall. Nevertheless, to identified

specific genes involves in the pathogenicity, a gene expression analysis was conducted. The results of the expression analysis showed that over 30% of the predicted transcriptome were expressed in mycelium or *in planta.* Interestingly, more than 50% of secreted CAZymes were differentially expressed. This suggested the importance of these proteins in the degradation of the host during the infection. This was also observed analysing the expression of the different categories of CAZymes, which showed that the infection is mainly led by enzymes that are required in the degradation of the plant cell wall. On the other hand, carbohydrate-binding module (CBM) were also found highly expressed during the infection. While the majority of these protein have a cellulose binding function, the LysM domains in some families have been shown to inhibit plant chitinases (Marshall et al. 2011), having a direct effect in the prevention of host immunity responses. Therefore, these genes can be regarded as interesting in the study of host and pathogen interaction. Similarly, multiple effector-like proteins were identified highly upregulated during the infection. Nevertheless, the function of most of them is unknown and needs to be studied in the future.

A total of 2437 and 3191 more genes were differentially expressed in the proximal samples of 'Golden Delicious' and 'M9', respectively, compared to the distal samples. This is directly correlated with the difference of compatible fragments quantified in both samples, being 33 million in the proximal samples and 180 thousand in the distal samples. Nevertheless, 72 CAZymes and 11 effector proteins were highly upregulated on every group of samples. This suggested that despite of the differences of compatible fragments, this experiment was able to effectively identify key genes of the infection at two different stages. Therefore, differently expressed genes identified in both, proximal and distal samples, might represent early released effectors that interacts with the host at the beginning of the infection.

Finally, analysis of the genetic diversity of a global sample of *N. ditissima* was investigated. Variant sites of all the isolates sequenced in this study were identified with respect to the best reference genome generated in this study, and with an isolate of *N. major* acting as an outgroup. Analysis of the levels of shared sites between isolates showed a clear species separation between *N. ditissima* and *N. major*. Interestingly, within the *N. ditissima* population, there was a clear separation between two groups of isolates. One main population consisting of isolates from United Kingdom, Ireland, Netherlands, Belgium, Spain and Brazil were grouped together. This population was clearly distant to another group composed by isolates from Italy and Slovakia. This separation was also observed in a phylogenetic tree based on highly conserved genes in the genomes. This might suggest the presence of different populations or races within the *N. ditissima* species. In order to effectively evaluated the evolutionary history of this specie, a STRUCTURE analysis was performed. Considering a population admixture model, STRUCTURE assign individual samples to K possible

populations based on variant sites frequencies. Again, the results showed separation between two populations. This suggest that there was geographical separation in the past. This event was followed by a reduction of the diversity within allopatrically separated populations due to the founder effect. This reduction in the diversity within each population goes together with an increase of the diversity between populations. Interestingly, the isolates from Slovakia were scored as an intermediate between the main population and the Italian population. This suggested that the Slovakian isolates represent a hybrid between these two populations. Hybridization is frequently associated with the generation of host specificities (Stukenbrock et al. 2012; Depotter et al. 2016; Menardo et al. 2016). *N. ditissima* is well known to be a pathogen in apple and pear. Nevertheless, the Slovakian isolates were isolated from *Fagus sylvatica,* beech, indicating a possible host specialization in *N. ditissima.*

Previous phylogenetic studies have revealed that European and American populations appear to have a significant level of nucleotide divergence at ß-tubulin and RPB2 loci (Castlebury et al., 2006), indicating that the populations may be allopatrically isolated. In this study, we showed similar diversity between two populations of *N. ditissima*. Nevertheless, the two isolates from Brazil were grouped together with the European population, indicating that the Italian isolates are even more genetically distant. This study also corroborates the results of the population genetic analysis presented in the first report. In this study, we found slight evidence of population structure, indicating that similarities observed between geographically isolated populations, like Europe and Brazil, are probably due to the spread of *N. ditissima* on imported apple material (Gómez-Cortecero et al. 2016).

Conclusions

This project provided a better understanding of the phytopathogenic fungus *Neonectria ditissima*.

Variation in level of resistance to infection of *N. ditissima* was investigated using different inoculation methods in a range of apple commercial cultivars, rootstocks and seedlings. The goal of this research was to develop a reliable method to asses host responses to the infection. The results of this test revealed large variation among the cultivars tested in resistance responses. On the other hand, seedling populations test revealed that transmission of resistance from parental material also differ among cultivars. These results highlighted the importance of the evaluation of resistance of different cultivars prior incorporating them into breeding programmes.

The molecular study of this pathogen has been hampered by the lack of a reference genome. Taking advance of the reducing cost of the sequencing costs, numerous assemblies of *N. ditissima* were performed on this study. This included the generation of a high quality

reference genome using long read sequencing technologies. In order to generate the best genome assembly, two isolates from United Kingdom were sequences using two different long read sequencing technologies. Seven different methodologies were evaluated in the assembly of both genomes. The resulting assemblies were highly contiguous compared to the previous versions, being a valuable resource in the analysis of this pathogen.

In order to identify candidate pathogenicity genes, genes models were predicted and annotated on every genome generated in this study. Following this, additional features, such as effector and secreted effector genes, enzymes, secondary metabolites genes, repeat regions, were identified. Analysis of the secretome allowed the identification of an effector repertoire associated with the necrotrophic lifestyle of *N. ditissima*.

Differentially expressed genes were identified during the time course of an infection using RNA-Seq data. This allowed the identification of key genes involved in the degradation of the plant cell wall and effector genes. These genes are regarded as important and will be used for further functional validation and the identification of resistance genes.

Population genetic studies revealed the existence of distinct populations within *N. ditissima* specie. In addition, evidences of hybrid recombining populations, which could lead to a host specialisation, were observed. This provided an useful background in the evolutionary history of this pathogen.

Knowledge and Technology Transfer

25th November 2015 – The 2nd EMR PhD student poster exhibition. Poster presentation.

26th May 2016 – Postgraduate Symposium at University of Reading. Project presentation.

18th August 2016 – Poster presentation at University of Reading Microbiology research day. Awarded with the poster prize.

12th-13th September 2016 – BSPP Presidential Meeting. Poster presentation.

17th-19th October 2016 – 2016 International Academic Conference at Nanjing Agricultural University. Project presentation. Awarded with the first prize in the presentation competition.

17th-19th October 2016 – The 3rd International Horticulture Research Conference at Nanjing. Poster presentation. Awarded with the first prize in the poster competition.

16th-17th November 2016 – AHDB Crops Student Conference. Poster presentation.

28th February 2017 – EMR Association/AHDB Tree Fruit Day at NIAB EMR. Project presentation.

29th-30th March 2017 – Molecular Biology of Plant Pathogens Conference at Durham University. Poster presentation.

13th June 2017 – PhD Symposium at University of Reading. Project presentation.

16th-20th July 2017 – The 4th International Horticulture Research Conference at NIAB EMR. Poster presentation.

1st-3rd November 2017 – Third International Workshop on Apple Canker and Replant Disease at NIAB EMR. Project and poster presentation.

6th-7th November 2017 – AHDB crops PhD conference. Project presentation. 1st-3rd November

20th-21st November 2017 – NIAB outreach event at NIAB. Project and poster presentation.

26th June 2018 – Postgraduate Symposium at University of Reading. Project presentation.

Glossary

AUDPC	Area under the disease-progression curve
BUSCO	Benchmarking Universal Singel-Copy Orthologs
CAZy	Carbohydrate-active enzymes
cDNA	Complementary Deoxyribonucleic acid
DEGs	Differentially Expressed Genes
DNA	Deoxyribonucleic acid
GATK	Genome Analysis Toolkit
gDNA	Genomic Deoxyribonucleic acid
InDels	Insertion of deletion of bases in the genome
LFC	Log_2 Fold Change
mRNA	Messenger RNA
NIAB	National Institute of Agricultural and Botany
NIAB EMR	National Institute of Agricultural and Botany East Malling Research
PCR	Polymerase chain reaction
QTL	Quantitative trait loci
RNA	Ribonucleic Acid
RNA-Seq	Ribonucleic Acid Sequencing
SNPs	Single nucleotide polymorphisms

References

Alston, F.H. 1969. Response of Apple cultivars to canker, Nectria galligena. *Annual Report East Malling Research Station* A53, pp. 147–148.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215(3), pp. 403–410.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data [Online]. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc [Accessed: 8 January 2019].

Aronesty, E. 2013. Comparison of Sequencing Utility Programs. *The open bioinformatics journal* 7(1), pp. 1–8.

Bergey, C.M. 2012. vcf-tab-to-fasta [Online]. Available at: https://code.google.com/archive/p/vcf-tab-to-fasta/ [Accessed: 10 April 2019].

Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S.A., Suarez Duran, H.G., de Los Santos, E.L.C., Kim, H.U., Nave, M., Dickschat, J.S., Mitchell, D.A., Shelest, E., Breitling, R., Takano, E., Lee, S.Y., Weber, T. and Medema, M.H. 2017. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research* 45(W1), pp. W36–W41.

Borecki, Z. and Czynczyk, A. 1985. Susceptibility of apple cultivars to bark canker diseases. *Acta Agrobotanica* 38(1), pp. 49–59.

Broad Institute 2019. *Picard Toolkit*. Broad Institute.

Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15), pp. 1972–1973.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6(2), pp. 80–92.

Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E.A., Gouzy, J., Rees, D.J.G., Guérif, P., Muranty, H., Durel, C.-E., Laurens, F., Lespinasse, Y., Gaillard, S., Aubourg, S., Quesneville, H., Weigel, D., van de Weg, E., Troggio, M. and Bucher, E. 2017. High-quality de novo

assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics* 49(7), pp. 1099–1106.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R. and 1000 Genomes Project Analysis Group 2011. The variant call format and VCFtools. *Bioinformatics* 27(15), pp. 2156–2158.

Deng, C.H., Scheper, R.W.A., Thrimawithana, A.H. and Bowen, J.K. 2015. Draft Genome Sequences of Two Isolates of the Plant-Pathogenic Fungus Neonectria ditissima That Differ in Virulence. *Genome announcements* 3(6).

Depotter, J.R., Seidl, M.F., Wood, T.A. and Thomma, B.P. 2016. Interspecific hybridization impacts host range and pathogenicity of filamentous microbes. *Current Opinion in Microbiology* 32, pp. 7–13.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1), pp. 15–21.

Earl, D.A. and vonHoldt, B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources* 4(2), pp. 359–361.

Evanno, G., Regnaut, S. and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14(8), pp. 2611–2620.

Finn, R.D., Clements, J. and Eddy, S.R. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* 39(Web Server issue), pp. W29-37.

Garkava-Gustavsson, L., Ghasemkhani, M., Zborowska, A., Englund, J.E., Lateur, M. and van de Weg, E. 2016. Approaches for evaluation of resistance to European canker (*Neonectria ditissima*) in apple. *Acta horticulturae* (1127), pp. 75–82.

Garkava-Gustavsson, L., Zborowska, A., Sehic, J., Rur, M., Nybom, H., Englund, J.E., Lateur, M., Van de Weg, E. and Holefors, A. 2013. Screening of apple cultivars for resistance to european canker, neonectria ditissima. *Acta horticulturae* (976), pp. 529–536.

Garrison, E. 2012. Vcflib. A C++ library for parsing and manipulating VCF files [Online]. Available at: https://github.com/ekg/vcflib [Accessed: 3 May 2019].

Gómez-Cortecero, A., Harrison, R.J. and Armitage, A.D. 2015. Draft Genome Sequence of a European Isolate of the Apple Canker Pathogen Neonectria ditissima. *Genome*

announcements 3(6).

Gómez-Cortecero, A., Saville, R.J., Scheper, R.W.A., Bowen, J.K., Agripino De Medeiros, H., Kingsnorth, J., Xu, X. and Harrison, R.J. 2016. Variation in Host and Pathogen in the Neonectria/Malus Interaction; toward an Understanding of the Genetic Basis of Resistance to European Canker. *Frontiers in plant science* 7, p. 1365.

Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M. and Stanke, M. 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32(5), pp. 767–769.

Jakobsson, M. and Rosenberg, N.A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23(14), pp. 1801–1806.

Johansson, G., Ståhlberg, J., Lindeberg, G., Engström, Å. and Pettersson, G. 1989. Isolated fungal cellulose terminal domains and a synthetic minimum analogue bind to cellulose. *FEBS Letters* 243(2), pp. 389–393.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R. and Hunter, S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9), pp. 1236–1240.

Katoh, K. and Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30(4), pp. 772–780.

Koren, S. and Phillippy, A.M. 2015. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Current Opinion in Microbiology* 23, pp. 110–120.

Kraehmer, H. and Schmidle, A. 1979. Susceptibility of some recently released varieties to Nectria galligena Bres. and Phytophthora cactorum. *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes*.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* 305(3), pp. 567–580.

Krüger, J. 1983. Anfälligkeiten von Apfelsorten und Kruzungsnachkommenschaftenfür den Obstbaumkrebs nach natürlicher und künstlicher Infektion. *Erwerbsobstbau* 25, pp. 114–116.

Langmead, B. and Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature*

Methods 9(4), pp. 357–359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16), pp. 2078–2079.

Lischer, H.E.L. and Excoffier, L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28(2), pp. 298–299.

Liu, K., Linder, C.R. and Warnow, T. 2011. RAxML and FastTree: comparing two methods for large-scale maximum likelihood phylogeny estimation. *Plos One* 6(11), p. e27731.

Lomsadze, A., Burns, P.D. and Borodovsky, M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research* 42(15), p. e119.

Love, M.I., Huber, W. and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15(12), p. 550.

Marshall, R., Kombrink, A., Motteram, J., Loza-Reyes, E., Lucas, J., Hammond-Kosack, K.E., Thomma, B.P.H.J. and Rudd, J.J. 2011. Analysis of two in planta expressed LysM effector homologs from the fungus Mycosphaerella graminicola reveals novel functional properties and varying contributions to virulence on wheat. *Plant Physiology* 156(2), pp. 756–769.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M.A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20(9), pp. 1297–1303.

Menardo, F., Praz, C.R., Wyder, S., Ben-David, R., Bourras, S., Matsumae, H., McNally, K.E., Parlange, F., Riba, A., Roffler, S., Schaefer, L.K., Shimizu, K.K., Valenti, L., Zbinden, H., Wicker, T. and Keller, B. 2016. Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. *Nature Genetics* 48(2), pp. 201–205.

Paradis, E. and Schliep, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35(3), pp. 526–528.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 14(4), pp. 417–419.

Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8(10), pp. 785–786.

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, A. and Lareu, M.V. 2013. An overview of STRUCTURE: applications, parameter settings, and supporting software.

Frontiers in genetics 4, p. 98.

Pritchard, J.K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2), pp. 945–959.

Quinlan, A.R. and Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6), pp. 841–842.

Rosenberg, N.A. 2004. DISTRUCT: a program for the graphical display of population structure. *Molecular ecology notes* 4(1), pp. 137–138.

Slowikowski, K. 2018. ggrepel: Automatically Position Non-Overlapping TextLabels with 'ggplot2'. R package version 0.8.0 [Online]. Available at: https://CRAN.R-project.org/package=ggrepel [Accessed: 3 April 2019].

Sperschneider, J., Gardiner, D.M., Dodds, P.N., Tini, F., Covarelli, L., Singh, K.B., Manners, J.M. and Taylor, J.M. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. *The New Phytologist* 210(2), pp. 743–761.

Srivastava, A., Sarkar, H., Gupta, N. and Patro, R. 2016. RapMap: a rapid, sensitive and accurate tool for mapping RNA-seq reads to transcriptomes. *Bioinformatics* 32(12), pp. i192–i200.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9), pp. 1312–1313.

Stanke, M., Diekhans, M., Baertsch, R. and Haussler, D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24(5), pp. 637–644.

Stukenbrock, E.H., Christiansen, F.B., Hansen, T.T., Dutheil, J.Y. and Schierup, M.H. 2012. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proceedings of the National Academy of Sciences of the United States of America* 109(27), pp. 10954–10959.

Team, R.C. 2015. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing [Online]. Available at: https://www.R-project.org/ [Accessed: 18 February 2019].

Testa, A.C., Hane, J.K., Ellwood, S.R. and Oliver, R.P. 2015. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* 16, p. 170.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. 2010. Transcript assembly and quantification by RNA-Seq

reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 28(5), pp. 511–515.

UniProt Consortium, T. 2018. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* 46(5), p. 2699.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M. and Venables, B. 2016. Various R programming tools for plotting data [Online]. Available at: https://CRAN.R-project.org/package=gplots [Accessed: 3 May 2019].

van de Weg, W.E. 1989. Screening for resistance to Nectria galligena Bres. in cut shoots of apple. *Euphytica* 42(3), pp. 233–240.

van de Weg, W.E., Giezen, S. and Jansen, R.C. 1992. Influence Of Temperature On Infection Of Seven Apple Cultivars By Nectria Galligena. *Acta phytopathologica et entomologica Hungarica* 27, pp. 631–635.

Wenneker, M., Goedhart, P.W., van der Steeg, P., van de Weg, W.E. and Schouten, H.J. 2017. Methods for the Quantification of Resistance of Apple Genotypes to European Fruit Tree Canker Caused by*Neonectria ditissima*. *Plant disease* 101(12), pp. 2012–2019.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Winnenburg, R., Baldwin, T.K., Urban, M., Rawlings, C., Köhler, J. and Hammond-Kosack, K.E. 2006. PHI-base: a new database for pathogen host interactions. *Nucleic Acids Research* 34(Database issue), pp. D459-64.

Yang, Y., Zhang, Y., Li, B., Yang, X., Dong, Y. and Qiu, D. 2018. A Verticillium dahliae Pectate Lyase Induces Plant Immune Responses and Contributes to Virulence. *Frontiers in plant science* 9, p. 1271.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. and Xu, Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 40(Web Server issue), pp. W445-51.

Yu, G., Lam, T.T.-Y., Zhu, H. and Guan, Y. 2018. Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Molecular Biology and Evolution* 35(12), pp. 3041–3043.

Yu, G., Smith, D.K., Zhu, H., Guan, Y. and Lam, T.T.-Y. 2016. *ggtree* : an *R* package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in ecology and evolution / British Ecological Society*.

Zhang, C., Rabiee, M., Sayyari, E. and Mirarab, S. 2018. ASTRAL-III: polynomial time species

tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19(Suppl 6), p. 153.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P.K., Xu, Y. and Yin, Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 46(W1), pp. W95–W101.

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C. and Weir, B.S. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28(24), pp. 3326–3328.

