Healthy trees, healthy future
Enabling technologies to combat *Phytophthora* diseases

SCION
forests · products · innovation

HTHF 2014-2015: *Phytophthora* and *Pinus radiata* genomics and transcriptomics

Rebecca McDougal, Dan Jones, Emily Telfer, Lucy Macdonald and Nari Williams

REFERENCE: HTHF-T006

supported by forestgrowers commodity levy

FOREST OWNERS ASSOCIATION

KEEP KAURI STANDING
KIA TOITU HE KAURI

Healthy trees, healthy future Enabling technologies to combat *Phytophthora* diseases

SCION forests · products · innovation

REPORT INFORMATION SHEET

REPORT TITLE	HTHF 2014-2015: *PHYTOPHTHORA* AND *PINUS RADIATA* GENOMICS AND TRANSCRIPTOMICS
AUTHORS	REBECCA MCDOUGAL, DAN JONES, EMILY TELFER, LUCY MACDONALD, AND NARI WILLIAMS
CLIENT	MBIE / FOREST OWNERS ASSOCIATION
MBIE CONTRACT NO:	C04X1305
SIDNEY OUTPUT NUMBER	56388
SIGNED OFF BY	NARI WILLIAMS
DATE	JUNE 2015
INTELLECTUAL PROPERTY	

Disclaimer

The information and opinions provided in the Report have been prepared for the Client and its specified purposes. Accordingly, any person other than the Client uses the information and opinions in this report entirely at its own risk. The Report has been provided in good faith and on the basis that reasonable endeavours have been made to be accurate and not misleading and to exercise reasonable care, skill and judgment in providing such information and opinions.

Neither Scion, nor any of its employees, officers, contractors, agents or other persons acting on its behalf or under its control accepts any responsibility or liability in respect of any information or opinions provided in this Report.

EXECUTIVE SUMMARY

Report Title: HTHF 2014-2015 *Phytophthora* and *Pinus radiata* genomics and transcriptomics

Authors: Rebecca McDougal, Dan Jones, Emily Telfer, Lucy Macdonald, and Nari Williams

The problem

Worldwide, arboreal *Phytophthora* species are a huge biosecurity challenge due to their broad-host range, increasing emergence and sweeping impacts. In New Zealand we have several diseases caused by *Phytophthora* species: red needle cast (*Phytophthora pluvialis*), kauri dieback (*Phytophthora agathadicida*) as well as other *Phytophthora* species associated with diseases of plantation, conservation and ornamental trees nationally (*P. kernoviae, P. multivora, P. cactorum, P. cinnamomi, P.* taxon totara).

This project

The *Phytophthora* genomic and transcriptomic data, together with the *Pinus radiata* transcriptomic data will form the resources from which we can identify putative factors (biomarkers) involved in the pathogen-host interactions.

Key Results

The generation of *Phytophthora* genomic and transcriptomic resources is well underway. DNA sequencing of 12 *Phytophthora* genomes (six species and two strains of each) has been performed in collaboration with University of Exeter and genome annotation for two *P. pluvialis* genomes has been progressed with NZGL. These genomes sequences are now being mined for genes encoding effectors and for RNA interference pathway enzymes. In addition, 24 RNA preparations from two species of *Phytophthora* (*P. pluvialis* and *P. kernoviae*) and two strains of each for comparative analyses of baseline levels of gene expression variability has been performed, with bioinformatics yet to be completed.

Pinus radiata transcriptomic analyses have provided a baseline transcriptomic resource which, when combined with the existing transcriptomes, will provide a reference template for assessment of future infection timeline samples. We have assessed the performance of variable RNA quality as determined by multiple criteria, and have determined a minimum level of biological replication required to capture the level of significant differential expression required for the HTHF programme. The analysis pipelines for the ERCC controls have been tested and used to identify abnormal samples which can be excluded from future analyses.

Seventy-eight RNA preparations have been extracted from two infection time-series of *P. pluvialis* and *P. kernoviae*, on one susceptible and one resistant genotype of *Pinus radiata*, and sent to NZGL for sequencing.

Implications of Results for Client

The genomic resources developed and analysis of baseline levels of gene expression variation will provide important information for future experiments using time-series data.

Further Work

Completion of bioinformatics analysis for transcriptomic sequence data of *Pinus radiata* and *Phytophthora* data is imminent. Further analysis of the *Phytophthora* genomes is likely to be on-going as we learn more about specific genes involved in host-pathogen interactions. All of this work will underpin the analysis of the time-series infection data where we will begin to form hypotheses around the specific biomarkers involved in key host-pathogen interactions.

HTHF 2014-2015: *Phytophthora* and *Pinus radiata* genomics and transcriptomics

Rebecca McDougal[1], Dan Jones[4], Emily Telfer[2], Lucy Macdonald[3], and Nari Williams[1]
Forest Protection[1], Forest Genetics[2], Forest Industry Informatics[3], Scion, 49 Sala Street, Private Bag 3020, Rotorua 3046, New Zealand.
New Zealand Genomics Ltd[4], dan.jones@auckland.ac.nz

30 June 2015

Table of Contents

EXECUTIVE SUMMARY ---- **i**
Introduction ---- **1**
Objective ---- **1**
***1. Phytophthora* Genomics & Transcriptomics** ---- **1**
1.1 Genomes ---- 1
1.1.1 Genome Annotation ---- 4
1.1.2 Augustus Training Population ---- 5
1.1.3 Effector Predictions ---- 6
1.1.4 RNAi machinery and ncRNAs in Phytophthora genomes ---- 8
1.2 Phytophthora Transcriptomes ---- 8
1.2.1 Tissue and RNA collection ---- 8
1.2.2 RNA Quality assessment ---- 8
1.2.3 Sequencing ---- 11
1.2.4 Assembly ---- 11
1.2.5 Utilisation of External RNA Control Consortium to standardise sequencing runs 12
2. *Pinus radiata* Transcriptome ---- **13**
2.1 Tissue and RNA collection ---- 13
2.1.1 RNA Quality Assessment ---- 14
2.2 Sequencing ---- 16
2.2.1 Assembly ---- 17
2.2.2 Utilisation of External RNA Control Consortium to standardise sequencing runs 17
2.2.3 Abundance estimations ---- 17
3. *P. pluvialis* & *P. kernoviae* infection time-series in *Pinus radiata* ---- **21**
3.1 Tissue and RNA collection ---- 21
3.1.1 RNA Quality Assessment ---- 22
3.2 Sequencing ---- 24
Discussion ---- **24**
Future Work ---- **25**
Phytophthora genomics ---- 25
Phytophthora transcriptomics ---- 25
Pinus radiata transcriptomics ---- 25
Transcriptomic analysis of *P. pluvialis* and *P. kernoviae* infection time-series in *Pinus radiata* ---- 25
Acknowledgements ---- **26**
References ---- **26**
Appendix 1: NZGL01187 interim report (Dec 2014) ---- **28**

Introduction

The Healthy Trees, Healthy Future project model provides a platform for genomic analyses elucidating the universal and species-specific processes involved in interactions between *Phytophthora* species and their hosts.

The *Phytophthora* genomic and transcriptomic data, together with the *Pinus radiata* transcriptomic data will form the resources from which putative factors (biomarkers) involved in the pathogen-host interactions can be identified. These resources will enable comparative studies to species outside the HTHF model, providing further confidence in the development of robust hypotheses regarding disease interactions between *Phytophthora* and the three hosts of interest, and a framework for investigating host-pathogen interactions.

Genomic resources generated will accelerate the development of diagnostic tools targeting species-, genus- and pathogenicity-specific sequences for application to biosecurity and pathogen ecology.

The enabling technology platform has already supported the establishment of solid collaboration with *Phytophthora* genomics and bioinformatics researchers at New Zealand Genomics Ltd (NZGL), University of Exeter, University of Canterbury and University of British Columbia (UBC) for which this report encompasses many of the collaborative outputs, in addition to the work performed at Scion. We regard these collaborations as a strong outcome of the HTHF project, and a key feature of the enabling technology programme.

Objective

To establish genomic and transcriptomic resources for host-pathogen interaction studies between *Phytophthora* species and *Pinus radiata.*

1. Phytophthora Genomics & Transcriptomics

1.1 Genomes

The genome sequences of six species of *Phytophthora*, including two strains of each were generated in collaboration with University of Exeter. Strains were selected based on their relevance and impact on New Zealand plantation and native forests (Scott & Williams, 2014), the phylo-geography of each species, the interrelationship with the existing research of HTHF collaborators, existing international knowledge base of *Phytophthora* and perspective as model organisms in the HTHF project (Table 1).

Phytophthora pluvialis is a foliar pathogen of *Pinus radiata*, and is responsible for red needle cast disease (Dick, et al., 2014). This disease impacts trees of all ages and usually during the early autumn to winter period when leaf wetness is high. The disease results in premature defoliation and therefore impacts growth and production. *Phytophthora kernoviae* is also a pathogen of *Pinus radiata* in New Zealand (Dick, et al., 2014), although it is more commonly known for its severe impact on beech and ornamentals in the UK (Brasier, et al., 2005).

Phytophthora agathadicida, formerly known as *Phytophthora* taxon agathis (or PTA), has been recently described (Weir, et al., 2015), although its pathogenicity on *Agathis australis* (Kauri) a New Zealand native, was first reported in 1974 on Great Barrier Island (Gadgil, 1974). At that time it was identified as *P. heveae*, but was more recently identified on mainland New Zealand and has been known by the abbreviation PTA (Beever, et al., 2009). *P. agathadicida* is a root pathogen that is very aggressive on Kauri, causing mortality in all size and age classes (Beever, et al., 2009).

Both *Phytophthora cinnamomi* and *Phytophthora cactorum* have had a long association with disease of *Pinus radiata* both in forest nurseries (Reglinski, et al., 2009) and with mortality in shelterbelts (Newhook, 1959). *P. cinnamomi* is an aggressive pathogen that has decimated jarrah forests in Western Australia. It has a worldwide distribution and a very broad host range where it likely infects over 3000 species. It is a root pathogen causing root rot and stem cankers (Hardham, 2005).

Recent genetic analysis of *P. cactorum* isolates associated with pine suggest that it is mostly likely a different species to the type *P. cactorum* (McDougal, et al., 2015). During this study the pine isolates tentatively identified as *P. cactorum* were compared at three genetic loci to the isolates of *P. cactorum* obtained from New Zealand apple orchards, where this species has long impacted apple production.

Phytophthora multivora is an aggressive pathogen associated with *Eucalyptus marginata and E. gomphocephala, Banksia menziesii* and *B. grandis* and other tree and plant species also in Western Australia (Scott, et al., 2009). In New Zealand, *P. multivora* is found in exotic forests, natural ecosystems and horticultural settings (Scott & Williams, 2014). Together with *P. cinnamomi* and *P. cryptogea*, it is also one of the multiple species of *Phytophthora* commonly isolated from soil around Kauri trees exhibiting dieback, caused by *P. agathdicidia* (Beever, et al., 2009; Horner & Hough, 2014).

Phytophthora taxon totara is an as-yet undescribed species of *Phytophthora* that has been isolated from the foliage of *Podocarpus totara*, another New Zealand native tree impacted by a *Phytophthora* disease. The disease has only occurred in a couple of small locations, and is not associated with mortality of the tree. While this species is not one of the model species in the HTHF program, the genome has been sequenced as it is another species of *Phytophthora* infecting foliage of a conifer, and as such could provide further information regarding this type of lifestyle. In addition to this, preliminary DNA sequence analysis indicates that it is a very different species genetically to other *Phytophthora* although preliminary phylogenomic analysis indicates that it clusters with other ITS Clade 3 *Phytophthora* species including *P. pluvialis* (R. McDougal & C. Sambles, unpublished data). A species description for *Phytophthora* taxon totara is currently underway (McDougal, Scott et al., manuscript in preparation).

Table 1. *Phytophthora* species for which we have genome sequences

Species	Strain	Clade[1]	Host	Date	Location	Reference
P. pluvialis	NZFS 3000	3	*Pinus radiata*	2008	Gisbourne, New Zealand	(Reeser, et al., 2013)
P. pluvialis	LC9-1[2]		raintrap isolate	2009	Oregon, USA	
P. kernoviae	NZFS 2646	10	*Annona cherimola*	2005	Northland, New Zealand	(Brasier, et al., 2005)
P. kernoviae	NZFS 3630		*Pinus radiata*	2011	Tokoroa, New Zealand	
P. cinnamomi	NZFS 3750	7	*Pinus radiata*	2013	Nelson, New Zealand	(Rands, 1922)
P. cinnamomi	MP94-48[3]		*Eucalyptus marginata*	1994	Western Australia	
P. agathadicida	NZFS 3772	5	*Agathis australis*	2013	Auckland, New Zealand	(Weir, et al., 2015)
P. agathadicida	NZFS 3770		*Agathis australis*	2006	Coromandel, New Zealand	
P. multivora	NZFS 3378	2	*Idesia polycarpa*	2010	Auckland, New Zealand	(Scott, et al., 2009)
P. multivora	NZFS 3448		*Metrosideros kermadecensis*	2010	Auckland, New Zealand	
P. taxon totara	NZFS 3727	3[4]	*Podocarpus totara*	2012	Northland, New Zealand	McDougal, R., Scott, P. et al. (in prep.)
P. taxon totara	NZFS 3642		*Podocarpus totara*	2011	Gisbourne, New Zealand	

[1] ITS Clades as described by (Cooke, et al., 2000)

[2] DNA supplied by E. Hansen (Oregon State University, USA). This is the species type-strain (ATCC MYA-4930).

[3] DNA supplied by Giles Hardy (Murdoch University, Australia)

[4] The placement of *P.* taxon totara in clade 3 is putative at this stage (R. McDougal & C. Sambles, unpublished data)

Upon sequencing of the LC9-1 *P. pluvialis* strain, it was noted that the genome assembly size was different to that of the NZFS 3000 genome assembly. This prompted further analysis of the genomes which revealed that the NZFS 3000 genome sequence had contaminating bacterial DNA sequences also. This has now been filtered with the revised assembly provided by collaborators at the University of Exeter (Assoc. Prof. David Studholme) being of similar in size to the Oregon LC9-1 strain (Table 2). The level of sequence coverage is still under investigation. The contaminating sequence was identified as a *Paenibacillus* species which are common bacterial associates of *Phytophthora* (Hallmann, et al., 1997). All other *Phytophthora* genomes have also been filtered for contaminating bacterial sequences.

Resulting from the above filtering exercise, a preliminary (but not exhaustive) search for unique regions in the *P. pluvialis* NZFS 3000 and LC9-1 genomes was performed by David Studholme. One contig unique to LC9-1 and two contigs unique to NZFS 3000 were identified. These unique DNA regions were compared to DNA sequences in the NCBI GenBank database (BLASTn; (Altschul, et al., 1990)) to identify if these regions encoded functional proteins. The three unique regions contained sequences that were similar (but with low level similarity) to *Phytophthora parasitica* or *Phytophthora infestans* hypothetical proteins (no function assigned) and from mRNA, indicating that they are expressed genes. It will be interesting to see if the DNA from these unique regions of NZFS 3000 are also identified in our transcriptomic analyses. It would also be interesting to find further unique regions in these genome sequences.

Table 2. *Phytophthora* genome sequencing results to date

Species	Strain	length of assembly (MB)	number contigs	mean contig length	GC%
pluvialis	NZFS 3000	47.9	5,811	8,251	53.2
	LC9-1	50.1	4,741	10,579	53.3
kernoviae	NZFS 2646	37.4	1,333	28,065	50.3
	NZFS 3630	37.6	1,392	27,008	50.3
cinnamomi	NZFS 3750	65.8	6,499	10,128	53.7
	MP94-48	72.9	8,379	8,697	53.5
agathadicida	NZFS 3772	41.9	2,545	16,447	52.5
	NZFS 3770	42.1	2,441	17,229	52.6
multivora	NZFS 3378	41.2	1,233	33,392	51.9
	NZFS 3448	41.6	1,330	32,011	51.9
taxon totara	NZFS 3727	57.5	2,986	19,272	51.6
	NZFS 3642	58.5	1,763	33,198	51.6

1.1.1 Genome Annotation

Genome annotation enables us to sort the likely identity of genes in the *Phytophthora* genomes based on established genetic sequence databases. Initial rounds of annotation (i.e. identification of genes and genetic elements) were performed on the filtered *P. pluvialis* (NZFS 3000) genome. These consisted of gene prediction using GeneMark (Borodovsky & McIninch, 1993) and Augustus (Stanke & Waack, 2003); identification of gene homology between closely related species; identification of known genes from the assembled transcriptome; and prediction and classification of repeat elements. An example of annotation outputs is below (Fig 1.), showing various annotations of the *P. pluvialis* genome.

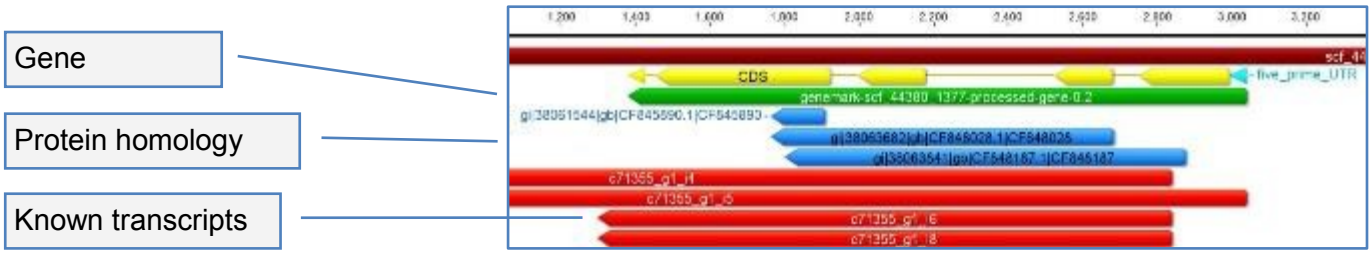

Fig 1. Genome browser screenshot depicting a region of the *Phytophthora pluvialis* 3000 genome with annotations derived from transcriptome assemblies, gene predictions, and homology to proteins in Genbank.

Table 3. Status of annotations for two *P. pluvialis* genomes.

Status of annotation	*P. pluvialis* genome	
	NZFS 3000	LC9-1
Repeat elements	Complete	Complete
Combined transcriptome mapped to genome	Complete	Future work
Species/strain-specific transcriptome mapped to genome	Future work	Future work
Homology vs. *Phytopthora* sp. In Genbank (nucleotide)	Complete	Future work
Homology vs. *Phytopthora* sp. In Genbank (protein)	Complete	Future work
Gene prediction (GeneMark-ES)	Complete	Complete
Gene prediction (Augustus – round 1)	Complete	Complete
Gene prediction (Augustus – round 2) using RNAseq data	In progress	Future work
Consensus gene prediction using MAKER	Future work	Future work

Future rounds of gene prediction will use RNAseq data, and therefore will be more robust and focussed on expressed genes. As a resource, the annotated *P. pluvialis* genome is already useful for investigation and for asking important biological questions (for example, what genes are shared across the *Phytophthora* genus?). This resource will continue to develop with further annotation and gene predictions updated as more data is collated across the HTHF programme (Table 3).

1.1.2 Augustus Training Population

(Dan Jones, University of Auckland/NZGL)

Currently the gene predictions are based on *de novo* predictions (GeneMark) or predictions derived from non-Oomycete gene models. We intend to use our RNA-seq data to generate a "training population" of gene models, and to use this to generate a new training set for the gene prediction program Augustus (Stanke & Waack, 2003). This will improve gene predictions for all of the *Phytophthora* genomes, and we believe it will be the first available training set for any Oomycete. Having this training set will greatly streamline the annotation and bioinformatics processing of future genome and transcriptome data sets. Therefore, any collaborators working in Oomycetes will also be able to improve their gene predictions using this training set. We expect that the Augustus training set will be published later this year (D. Jones, et al. manuscript in prep.).

1.1.3 Effector Predictions

(Assoc. Prof. David Studholme & Dr Christine Sambles, University of Exeter, UK)

Phytophthora species secrete proteins called effectors which play important roles in plant infection. Depending on the stage and mode of infection, these proteins can function outside or inside the plant cell where they are termed apoplastic or cytoplasmic effectors, respectively. Apoplastic effectors include elicitins, cell-wall degrading enzymes, enzyme inhibitors, toxins and unique 10kDa proteins. These effectors are located at the interface between the pathogen and the host and function outside the host cell. Elicitins are small secreted proteins that either induce or suppress plant defence responses independently or in conjunction with other small excreted proteins (Du, et al., 2015). Some elicitins are associated with infection of susceptible host genotypes, and are essential for invasion of woody host tissues by different *Phytophthora* species (Oßwald, et al., 2014). Cytoplasmic effectors translocate into the host cells and interfere with plant defense responses. Two classes of cytoplasmic effectors have been identified in *Phytophthora* species which have been shown to play major roles in plant infection; the CRNs (crinklers) and RXLRs.

Effector proteins have conserved motifs (sequence regions) which can be used to identify putative effector-encoding genes from the genomes of *Phytophthora* species. This approach has been used to identify effector genes from the 12 *Phytophthora* genomes sequenced in the HTHF project, along with the genome sequences of other *Phytophthora* species publically available. These sequence searches were performed by Dr Christine Sambles and Assoc. Prof. David Studholme (Figures 2 and 3). The numbers of CRNs in the genomes of the *Phytophthora* that we have sequenced, varied between species and strains. Of particular note, *P. pluvialis* NZFS 3000 was predicted to possess 55 CRNs, and LC9-1 was predicted to possess 53, whereas *P.* taxon totara (also *Phytophthora* Clade 3) NZFS 3642 was predicted to contain 98, and isolate NZFS 3727 was predicted to contain 67 genes (Fig 2). The numbers of RXLR effectors for *P. pluvialis* NZFS 3000 was predicted to be 77 RXLRs, and LC9-1 was predicted to possess 76, whereas *P.* taxon totara NZFS 3642 was predicted to contain 86, and isolate NZFS 3727 was predicted to contain 74 genes (Fig 3). Variation between strains and species was observed across all genome sequences analysed. Assembly quality will have an effect on prediction of effectors, therefore these results are considered preliminary (C. Sambles, pers. comm.), and may change with further iterations of assembly quality.

Prediction of elicitin genes from the *Phytophthora* genomes has commenced at Scion very recently. While this is still work in progress a broad list of approximately 685 genes have identified from all 12 genomes. Considerable further work needs to be done to refine this search, which will be done in the coming year in reference to the transcriptome data where possible to focus on those elicitins expressed during infection.

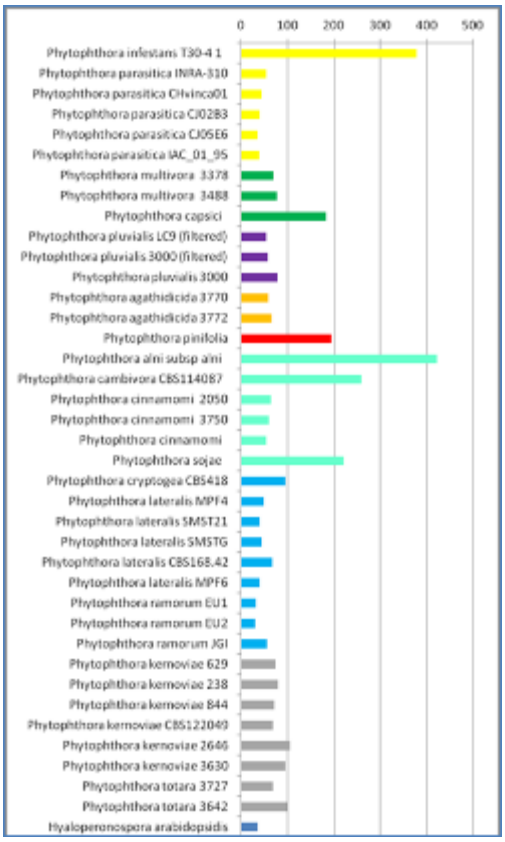

Fig 2. CRN predictions from *Phytophthora* genomes, including HTHF *Phytophthora* genomes (C. Sambles, University of Exeter).

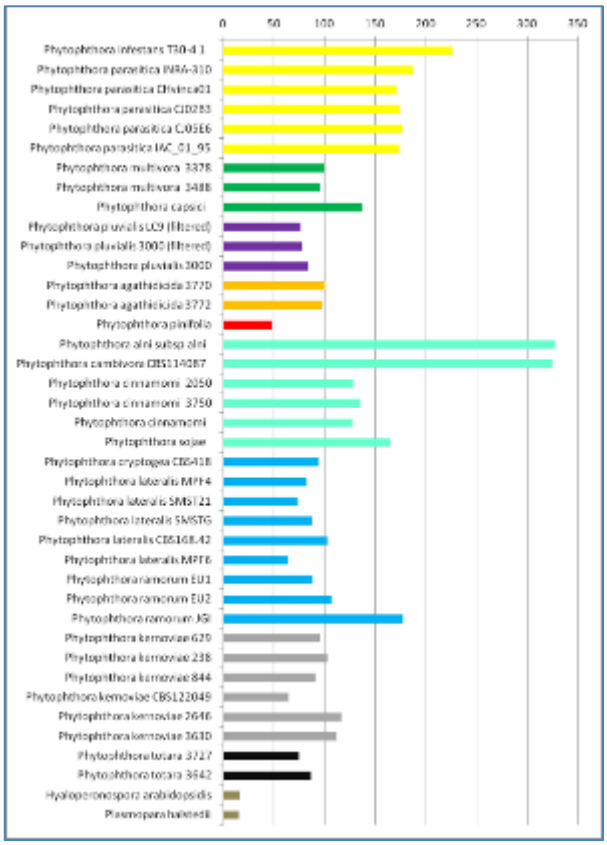

Fig 3. RXLR predictions from *Phytophthora* genomes, including HTHF *Phytophthora* genomes (C. Sambles, University of Exeter).

1.1.4 RNAi machinery and ncRNAs in Phytophthora *genomes*

Epigenetic processes such as histone modification (Raffaele, et al., 2010) and RNA silencing (Vetukuri, et al., 2013; Whisson, et al., 2012) play important roles in regulation of gene expression, and potentially influence host adaptation (Jiang & Tyler, 2011; Kasuga, et al., 2012). In collaboration with Dr Paul Gardner (University of Canterbury) we are using the *Phytophthora* genomic resources to search for ncRNAs (microRNAs, small non-coding RNAs etc) and RNA interference (RNAi) machinery such as Dicer and Argonaut enzymes involved in these processes (Vetukuri, et al., 2012; Vetukuri, et al., 2011).

Aspects of this work are currently in progress with Dr Gardner, starting with the investigation of ribosomal RNAs and other conserved RNA families using Rfam (Nawrocki, et al., 2014). It has been observed that several components of these conserved families, which have been thought to be essential genes in all genomes, are missing from all 12 *Phytophthora* genomes. Further work is underway to clarify these results.

Dr Ramesh Vetukuri (SLU, Sweden) who has pioneered many aspects of this work in *P. infestans* will visit Scion in September 2015. We hope to establish a collaboration in this research area with Dr Vetukuri also having broader research interests in *Phytophthora* genomics, transcriptomics and pathogen-host interactions.

1.2 Phytophthora *Transcriptomes*

Early attempts at obtaining RNA from *P. pluvialis* zoospores, germinating cysts, were unsuccessful, partly due to the inability to consistently obtain zoospores at a high enough concentration. For this reason, we decided to extract RNA from *P. pluvialis* and *P. kernoviae* (and two strains of each) mycelium, with six replicates per strain. This information will provide useful baseline level gene expression data for comparing within and between species.

1.2.1 Tissue and RNA collection

P. pluvialis NZFS 3000, NZFS 3613 and *P. kernoviae* NZFS 2646 and 3630 were grown in CAD medium (Erwin & Robeiro, 1996) (six replicates of 100ml medium in tissue culture flasks inoculated with mycelial plugs) for 4 days at 20 °C. Mycelium was removed to a screw cap tube and centrifuged for 30 s at 16,000 × g. Excess liquid was removed and the mycelium was snap-frozen in liquid nitrogen and stored at -80 °C. Frozen mycelium was transferred to RNase-free tubes containing sterile, RNase-free ceramic beads and sea sand, for initial lysis step. Lysis was performed using an Omni-Bead Ruptor Homogenizer (Omni International, Georgia, USA) with the settings: 6.95 m/s, 2 × 45 s, 30 s pause RNA was extracted using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, Missouri, USA) following the manufacturer's instructions. DNase I treatment was performed, and a single 50 µl final elution was performed.

1.2.2 RNA Quality assessment

The concentration of the total RNA was assessed using the Qubit RNA BR Assay Kit (for RNA in the range 20-1000 ng/µl) and quantified using an Invitrogen™ Qubit® fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Purity as a ratio of absorbance at 260/230 nm and 260/280 nm was determined using the NanoDrop™ 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA) (Table 3). Finally, 2 µl of total RNA was separated on a 1.2 % Agarose gel in 1 x TBE buffer (Fig 4). Samples were also assessed via BioAnalyzer (Agilent technologies, CA, USA) at the sequencing service provider, New Zealand genomics Ltd. (Table 4), and a RIN (RNA integrity Number (Schroeder, et al., 2006) was calculated. The BioAnalyzer also performs an electrophoresis analysis (Fig 5). Only six of the 24 *Phytophthora* RNA samples did meet the "general plant rule-of-thumb QC

threshold of RIN 7" (Table 4). In addition to this, the NZGL QC analysis of the RNA samples reported approximately 5-20 % DNA content (average 10 %). Overall, variability was observed in the amount and quality of RNA samples used in these analyses. However, as the RIN threshold value was not established for *Phytophthora* and therefore may not be entirely applicable, and because these samples did not exhibit typical RNA degradation patterns (in chromatograms), we decided to proceed with sequencing of all samples regardless.

Table 3. Summary on Scion Quality assessment of *Phytophthora* RNA samples

Sample name	***Nanodrop ng/µl***	***A260/230***	***A260/280***	***Qubit RNA ng/µl***
Pp3000-1	428.9	2.45	2.17	372
Pp3000-2	751.5	2.49	2.2	698
Pp3000-3	21.4	1.93	2.13	20.4
Pp3000-4	656.5	2.49	2.2	523
Pp3000-5	117.8	2.45	2.2	102
Pp3000-6	791.8	2.52	2.21	623
Pp3613-1	177.3	2.47	2.22	160
Pp3613-2	67.1	2.56	2.23	59.7
Pp3613-3	113.7	2.5	2.21	110
Pp3613-4	150.6	2.5	2.16	136
Pp3613-5	75	2.45	2.17	71.1
Pp3613-6	69	2.4	2.15	60.1
Pk2646-1	86.7	2.24	2.15	72.5
Pk2646-2	739.2	2.39	2.14	585
Pk2646-3	97.5	2.24	2.11	91.6
Pk2646-4	533.5	2.34	2.13	469
Pk2646-5	10.1	1.57	1.9	10
Pk2646-6	239	2.36	2.13	225
Pk3630-2	1830	2.26	2.13	113
Pk3630-3a[1]	127.6	2.37	2.13	>1000
Pk3630-3b[1]	1830	2.37	2.13	>1000
Pk3630-4	23.8	1.9	1.98	20.8
Pk3630-5	70	2.24	2.16	59.5
Pk3630-6	541.5	2.35	2.14	453

[1] Duplicates of sample Pk3630-3 were submitted for sequencing, as sample Pk3630-1 failed to yield enough RNA.

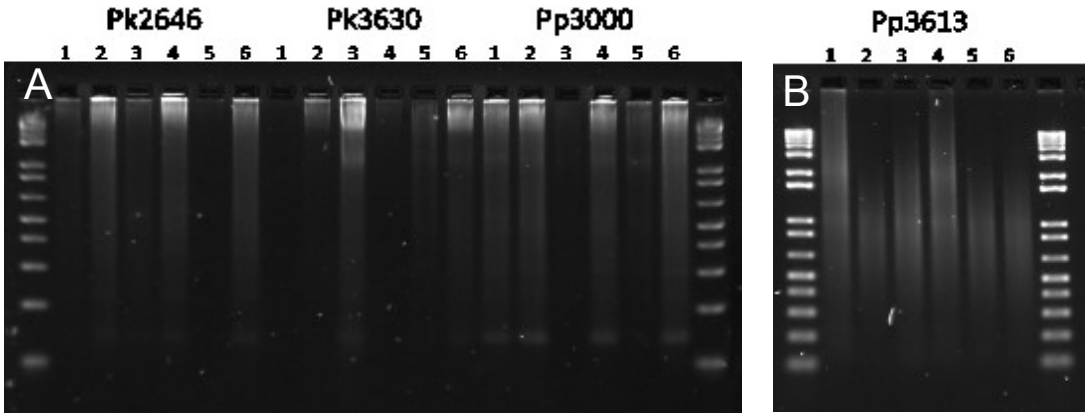

Fig 4. Agarose gel showing the total RNA from *Phytophthora* samples separated by electrophoresis. A; RNA from *Phytophthora kernoviae* (Pk) 2646 and 3630, and *Phytophthora pluvialis* (Pp) 3000, with the six replicate RNA preparations for each., B; *Phytophthora pluvialis* (Pp) 3613, with the six replicate RNA preparations. The first and last lane of each gel (unlabelled) is 1Kb$^+$ ladder.

Table 4: BioAnalyzer results for *Phytophthora* RNA samples

SAMPLE_ID	NAME	RIN	SAMPLE_ID	NAME	RIN
1187-21	Pp3000-1	6.2	1187-33	Pk2646-1	6.3
1187-22	Pp3000-2	6.0	1187-34	Pk2646-2	3.9
1187-23	Pp3000-3	2.30	1187-35	Pk2646-3	5.9
1187-24	Pp3000-4	7[1]	1187-36	Pk2646-4	6.8
1187-25	Pp3000-5	6.3	1187-37	Pk2646-5	7.7[1]
1187-26	Pp3000-6	6.1	1187-38	Pk2646-6	3.0
1187-27	Pp3613-1	4.9	1187-39	Pk3630-2	8.2[1]
1187-28	Pp3613-2	4.1	1187-40	Pk3630-3a	7.7[1]
1187-29	Pp3613-3	2.9	1187-41	Pk3630-3b	7.6[1]
1187-30	Pp3613-4	4.1	1187-42	Pk3630-4	6.2
1187-31	Pp3613-5	3.4	1187-43	Pk3630-5	5.1
1187-32	Pp3613-6	3.4	1187-44	Pk3630-6	8[1]

[1] Only six of the 24 *Phytophthora* RNA samples met the "rule-of-thumb QC threshold of RIN 7"

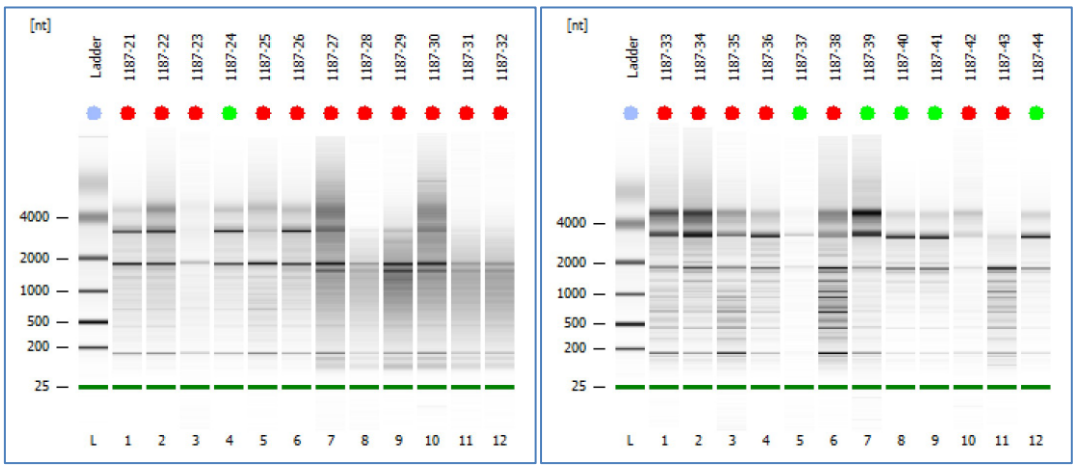

Fig 5. BioAnalyzer Electrophoresis showing the presence of high quality RNA in each sample despite low RIN measurement; (blue) marker ladder; (red) did not meet RIN threshold of 7, (green) met RIN threshold of 7

1.2.3 Sequencing

RNA libraries of the *Phytophthora* extracts were prepared using the TrueSeq kit from Illumina and sequenced using Illumina's HiSeq 2500 with poly-A capture of the messenger RNA as has been performed for other *Phytophthora* transcriptomic analyses e.g. (Hayden, et al., 2014). The *Phytophthora* RNA-seq libraries were re-sequenced in January 2015 due to a faulty Illumina PCR Master mix that was used for the first lot of sequencing. Only the re-sequencing results are presented in the table below (Table 5). Sequence quality as indicated by the percentage of ≥ Q30 bases and the mean quality scores in the resulting sequences indicates that, despite the low RIN values obtained for some of the *Phytophthora* samples, all samples sequenced well. Some samples yielded more sequencing reads than others, however this variation was observed in the duplicate samples Pk3630-3a and Pk3630-3b taken from the same RNA preparation which indicates stochastic variation in the sequencing rather than differences in RNA amount or quality.

Table 5. Sequence summary for *Phytophthora* RNA samples

Row Labels	Sum of # Reads	Average of % of >= Q30 Bases (PF)[1]	Average of Mean Quality Score (PF)
Pp3000-1	88,194,024	96.17	36.11
Pp3000-2	79,814,150	95.96	36.06
Pp3000-3	73,089,342	95.97	36.07
Pp3000-4	72,909,956	95.67	36.01
Pp3000-5	78,817,866	95.48	35.98
Pp3000-6	73,424,670	96.33	36.13
Pp3613-1	113,483,002	95.89	36.04
Pp3613-2	80,468,154	95.82	36.03
Pp3613-3	78,896,048	96.27	36.13
Pp3613-4	76,231,734	95.96	36.06
Pp3613-5	68,501,562	96.08	36.09
Pp3613-6	73,573,310	95.65	36.00
Pk2646-1	72,609,690	95.39	35.97
Pk2646-2	78,944,916	96.33	36.13
Pk2646-3	76,330,278	95.97	36.06
Pk2646-4	81,403,360	96.07	36.08
Pk2646-6	84,328,202	96.06	36.08
Pk3630-2	81,710,904	96.05	36.08
Pk3630-3a	116,700,724	95.98	36.08
Pk3630-3b	83,905,736	95.59	36.00
Pk3630-4	75,269,866	95.09	35.92
Pk3630-5	110,278,374	96.19	36.10
Pk3630-6	82,530,830	95.80	36.03
Grand Total	**1,901,416,698**	**95.90**	**36.05**

[1]Q = Phred score; each base in the sequence is assigned a Phred quality score which indicates the probability of an incorrectly called base. Q30 indicates a 1 in 1000 probability of an incorrectly called base (or conversely a 99.9% base call accuracy). A Q30 value over 90% would be considered acceptable quality.

1.2.4 Assembly

Phytophthora mycelium RNA-seq data is being used to assess the baseline level of gene expression in each species, and to investigate gene expression variability both within and between species and strains. Initially, all *Phytophthora* spp. reads were assembled *de novo* using Trinity version r20140717 using the same process as the *P. radiata* transcriptomic

data. This work was delayed due to a problem with the manufacturers sequencing reagents (Section 1.2.3) but the transcriptome assembly has now been completed.

Total number of genes:	175,381
Total number of transcripts (including alternative isoforms):	234,555

The shared transcriptome enables gene expression comparisons across strains and species with gene expression comparisons in progress. This shared transcriptome is being has also been mapped to the *P. pluvialis* NZFS 3000 genome as an annotation track.

1.2.5 Utilisation of External RNA Control Consortium to standardise sequencing runs

External RNA controls consortium (ERCC) spike-in controls were used as a control (Jiang, et al., 2011). These are standardised artificial RNA samples that are included with the sequencing and analysis. After assembly these are checked to see if the artificial RNAs are detected at expected abundances and ratios. This allows detection of variations caused by sequencing process (as opposed to true biological variation). This analysis was done using the R package erccdashboard (ver. 1.2.0) and custom scripts developed by NZGL. Analysis of the spike-in data for the *P. pluvialis* and *P. kernoviae* transcriptomes is underway but is not yet complete having been led by the analysis of the *P. radiata* transcriptomes in this case (Section 2.2.2).

2. *Pinus radiata* Transcriptome

The analysis of gene transcription has long been used to understand the roles of specific genes in processes of interest. By comparing pools of RNA between different treatments, increases or decreases in gene expression can be observed in relation to a control treatment. However, the level of variation in gene expression even in controlled experiments can be considerable. Given the costs of transcriptome sequencing, understanding the level of biological and technical replication is an important for optimising the cost and benefit of expression analyses.

In order to understand the stochastic variability in our samples, we performed an experiment to determine the level of variability between "identical" samples from radiata pine needles and roots. These tissues were chosen to represent the tissues and conditions that would be used for control-inoculation experiments downstream.

2.1 Tissue and RNA collection

Ten fascicles from a single radiata pine tree were collected from the Scion nursery and "incubated" overnight in sterile pond water. The following morning, fascicles were snap frozen in liquid nitrogen and stored at -80°C.

Roots were collected from the root systems of radiata pine cuttings rooted in Rockwool, which is predominantly chemically and biologically inert. There was variation in the amount of root biomass between cuttings. Not a single cutting had sufficient biomass to accommodate the 10 replicate RNA extractions we hoped to perform. Therefore we had to include several genotypes, but this did allow us to compare both within and between genotype variation. Samples were frozen in liquid nitrogen and stored at -80°.

RNA was extracted using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, Missouri, USA) following the manufacturer's instructions (Table 1).

Table 6. RNA extraction methods

Tissue Type	Genotype	Protocol	Tissue	DNase Treated
Needles 1-10	44.7	B	3.5cm Fasicle	Yes
Roots 1	1	A	Whole root	Yes
Roots 2	2	A	Whole root	Yes
Roots 3	3	A	Whole root	Yes
Roots 4	4	A	Whole root	Yes
Roots 5	5	A	Whole root	Yes
Roots 6-10	6	A	30 - 35mg	Yes

2.1.1 RNA Quality Assessment

The concentration of the total RNA was assessed using the Qubit RNA BR Assay Kit and quantified using an Invitrogen™ Qubit® fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Purity as a ratio of absorbance at 260/230 nm and 260/280 nm was determined using the NanoDrop™ 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA) (Table 7). Finally, 2 µl of total RNA was separated on a 1.2 % Agarose gel in 1 x TBE buffer (Fig 6). Samples were also assessed via BioAnalyzer (Agilent technologies, CA, USA) at the sequencing service provider, New Zealand Genomics Ltd. (Table 8), and a RIN (RNA integrity Number (Schroeder, et al., 2006)) was calculated. The BioAnalyzer also performs an electrophoresis analysis (Figure 7) where the additional ribosomal bands found in chloroplasts are visible in the RNA samples from needles.

Table 7. Summary on Scion quality assessment of *P. radiata* RNA samples

Sample name	***Nanodrop ng/µl***	***A260/230***	***A260/280***	***QUBIT RNA ng/µl***	***QUBIT RNA ng/µl***
44_7-Needle 1	343.73	2.11	2.23	381	20.5
44_7-Needle 2	435.33	2.09	2.16	470	23.7
44_7-Needle 3	466.62	2.11	2.21	472	23.5
44_7-Needle 4	330.32	2.11	2.21	350	21
44_7-Needle 5	379.78	2.1	2.19	396	22.5
44_7-Needle 6	470.17	2.09	2.2	472	22.2
44_7-Needle 7	364.17	2.09	2.24	400	21.3
44_7-Needle 8	354.08	2.11	2.15	385	23
44_7-Needle 9	372.61	2.12	2.28	408	21.7
44_7-Needle 10	330.63	2.12	2.22	352	20.4
1-Root1	375.99	2	1.27	385	23.8
2-Root2	306.26	1.93	1.06	315	21.4
3-Root3	211.08	1.94	1	214	17.3
4-Root4	166.95	1.8	0.7	171	18.3
5-Root5	87.51	1.85	0.68	82.4	11.1
6-Root6	28.28	1.85	1.1	33.6	3.19
6-Root7	17.15	1.74	0.68	17.7	1.08
6-Root8	20.21	1.79	0.82	23.2	1.56
6-Root9	25.76	1.8	0.58	25.3	1.71
6-Root10	38.26	1.86	0.76	39.1	4.18

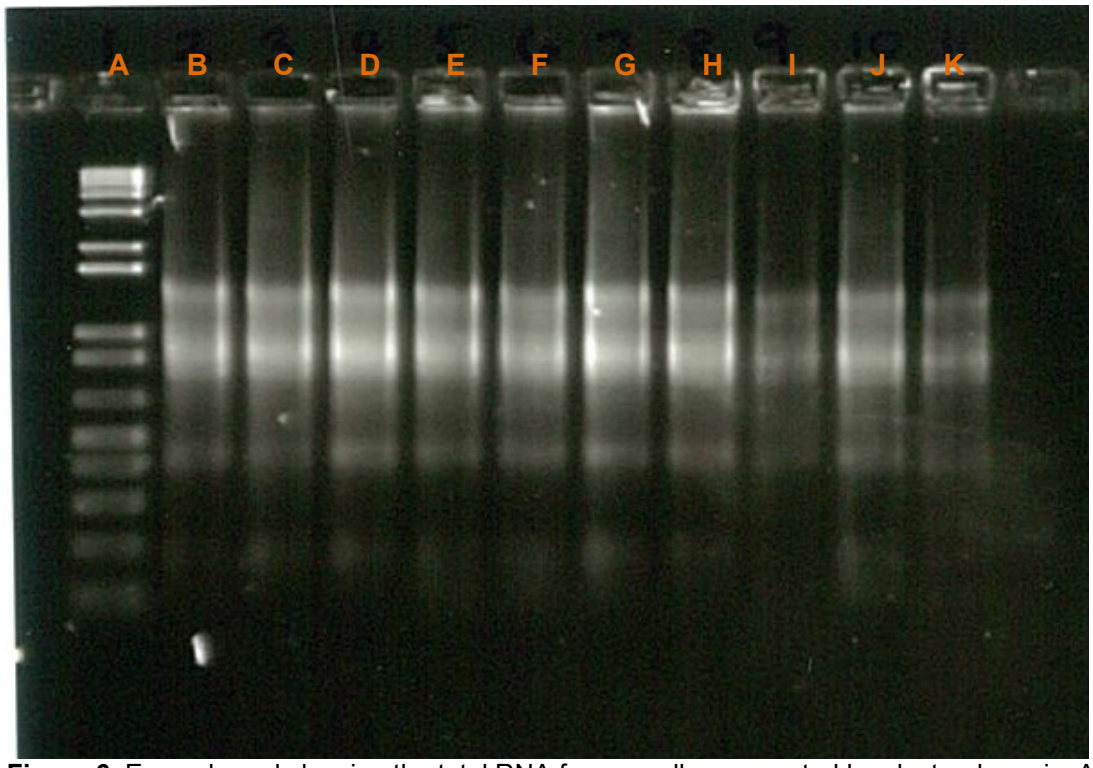

Figure 6. Exemplar gel showing the total RNA from needles separated by electrophoresis. A; 1Kb+ ladder, B-K; Needle samples 1-10.

Table 8. BioAnalyzer results for *Pinus radiata* RNA samples

SAMPLE_ID	*NAME*	*RIN*	*SAMPLE_ID*	*NAME*	*RIN*
1187-1	RAD-N1	7.80	1187-11	RAD-R1	8.30
1187-2	RAD-N2	8.60	1187-12	RAD-R2	8.10
1187-3	RAD-N3	8.40	1187-13	RAD-R3	3.50
1187-4	RAD-N4	7.80	1187-14	RAD-R4	7.90
1187-5	RAD-N5	7.40	1187-15	RAD-R5	8.30
1187-6	RAD-N6	7.80	1187-16	RAD6-R6	8.10
1187-7	RAD-N7	7.90	1187-17	RAD6-R7	8.10
1187-8	RAD-N8	8.60	1187-18	RAD6-R8	8.10
1187-9	RAD-N9	8.10	1187-19	RAD6-R9	8.00
1187-10	RAD-N10	8.50	1187-20	RAD6-R10	7.90

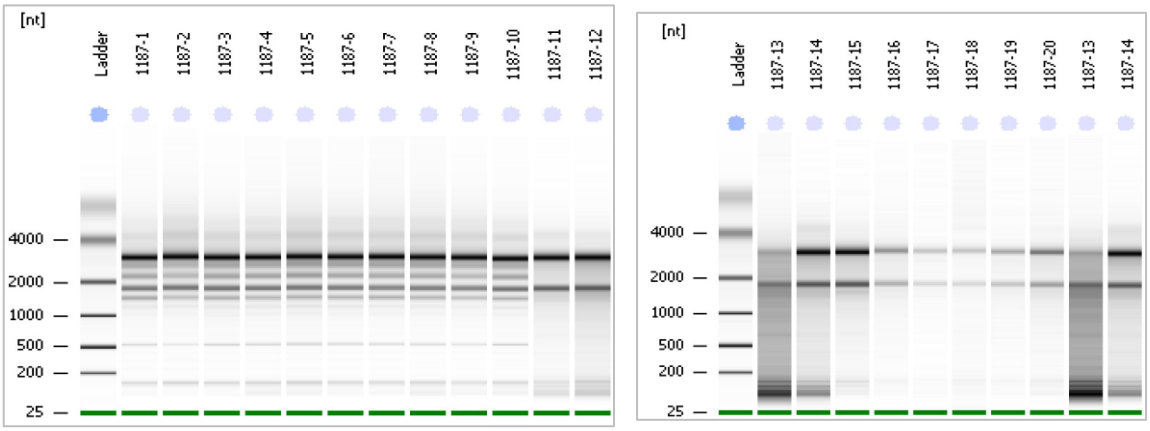

Fig 7. BioAnalyzer Electrophoresis. Samples 1-10 RNA extracts from needles showing the additional band of RNA from the chloroplast gene expression. Samples 11-14 RNA extracts from root samples.

2.2 Sequencing

Total RNA was prepared in libraries using the TrueSeq kit from Illumina and sequenced using Illumina's HiSeq 2500. Using RNA ensures that only those genes expressed in the tissue at that time will be sequenced. This gives a base line gene expression level which, when compared to tissues inoculated with *Phytophthora*, provides an indication of which genes are involved in defence and resistance. A summary of the base sequencing statistics is shown in Table 4. Of note is sample RAD-R3, which had a low RIN score, but did not differ significantly from the other samples in sequence quality

Table 9. Sequence summary for *P. radiata* RNA samples

Row Labels	*Sum of # Reads*	*Average of % of >= Q30 Bases (PF)*	*Average of Mean Quality Score (PF)*
RAD3-R10	108,475,528	94.09	36.55
RAD3-R6	106,015,720	94.05	36.55
RAD3-R7	115,697,120	94.08	36.56
RAD3-R8	112,327,670	93.72	36.45
RAD3-R9	86,888,996	93.58	36.38
RAD-N1	117,853,760	94.11	36.51
RAD-N10	115,949,240	93.89	36.43
RAD-N2	118,680,372	94.00	36.44
RAD-N3	124,735,498	94.12	36.50
RAD-N4	107,332,848	93.93	36.43
RAD-N5	103,100,072	94.15	36.52
RAD-N6	112,159,354	94.15	36.51
RAD-N7	100,618,860	94.11	36.48
RAD-N8	105,450,906	94.05	36.47
RAD-N9	111,908,586	93.76	36.38
RAD-R1	103,995,974	93.70	36.39
RAD-R2	101,820,790	93.48	36.30
RAD-R3	100,610,354	93.28	36.22
RAD-R4	98,628,246	92.03	35.75
RAD-R5	98,942,914	93.18	36.23

2.2.1 Assembly

Next Generation high-throughput sequencing such as Illumina's HiSeq requires RNA to be fragmented into short, 100 base pair reads. These short reads were sequenced and assembled into genes or transcripts by *de novo* using Trinity ver. _r20140717 (Grabherr, et al., 2011). Needles and root transcriptomes were assembled separately to form two distinct transcript libraries to show which genes are expressed in each tissue type. A combined assembly of both tissue types will be used for abundance estimations in the challenged tissues from the time series experiments.

Total number of 'genes': 1,784,896
Total number of transcripts: 2,575,952

We already have a database of transcripts sequenced from a number of different genotypes and tissue types. By adding these new transcripts we will gain an overall picture of pine transcripts across tissue types and genotypes. We clustered these new transcripts to our existing database using VSEARCH (*VSEARCH ver 1.1.3*). This reduced our dataset by 23%.

2.2.2 Utilisation of External RNA Control Consortium to standardise sequencing runs

As with the *Phytophthora* transcriptomes, and ERCC control was incorporated in the transcriptome sequencing of the *P. radiata* RNA extracts. Analyses using the erccdashboard examine ERCC RNAs across the entire experiment, and no problems were found. We had previously found discrepancies in root sample 14, and would like to check the observed vs expected abundances of ERCC RNAs on individual samples, to establish whether these discrepancies are biological or sequencing artefacts. This will be completed using previously developed custom scripts.

2.2.3 Abundance estimations

By estimating the abundance of genes present in each tissue we can decipher the base level of gene expression in healthy tissue. This approach will enable the comparison between control and inoculated tissues in challenged tissues, but the analysis was performed on the *P. radiata* transcriptomes will be compared to the same tissues after inoculation with *Phytophthora*. To estimate gene expression the sequenced reads are aligned back to the Scion's *Pinus radiata* transcriptome database and each read that aligns to a transcript is counted. Transcripts or genes that have a higher number of reads align are more abundant (i.e. have higher expression) in the sample. A large number of differentially expressed genes were found between roots and needles. 181,580 genes have a four times higher expression level in roots as compared to needles. 8,460 genes have a 256-fold change in expression levels. Differential expression analysis has also enabled examination of the level of biological variability within a tissue (needles or roots) (Fig 8 and Appendix 1).

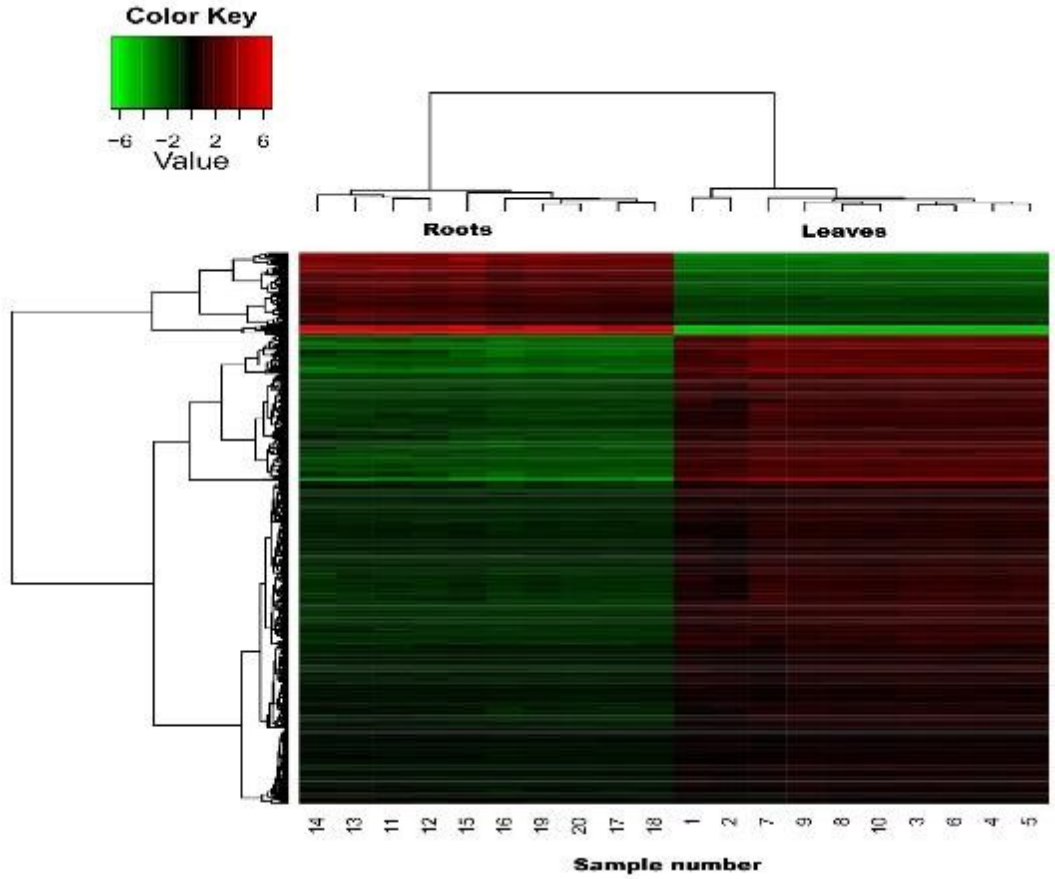

Fig 8. Example heatmap showing expression and clustering of the top 1000 differentially expressed genes. Genes (vertical axis) and samples (horizontal axis) are clustered according to the similarity of their expression pattern.

Differential expression analysis reveals many clusters of genes that show similar expression patterns between replicates and variable between tissues. The expression and differential expression at the level of individual isoforms (splice variants) as well as genes has also been examined. Importantly, we were able to assess the level of variation within a tissue.

Gene expression of the top 1000 differentially expressed genes showed that expression within roots is more variable than in needles with some root being more variable than others. In addition, analysis looking at the full transcriptome identified two pine needle samples (samples 9 and 10, Fig 9) that were significantly more variable than the other needle samples, and one root sample that was also much more variable (sample 14) (Fig. 10, see NZGL interim report attached in Appendix 1).

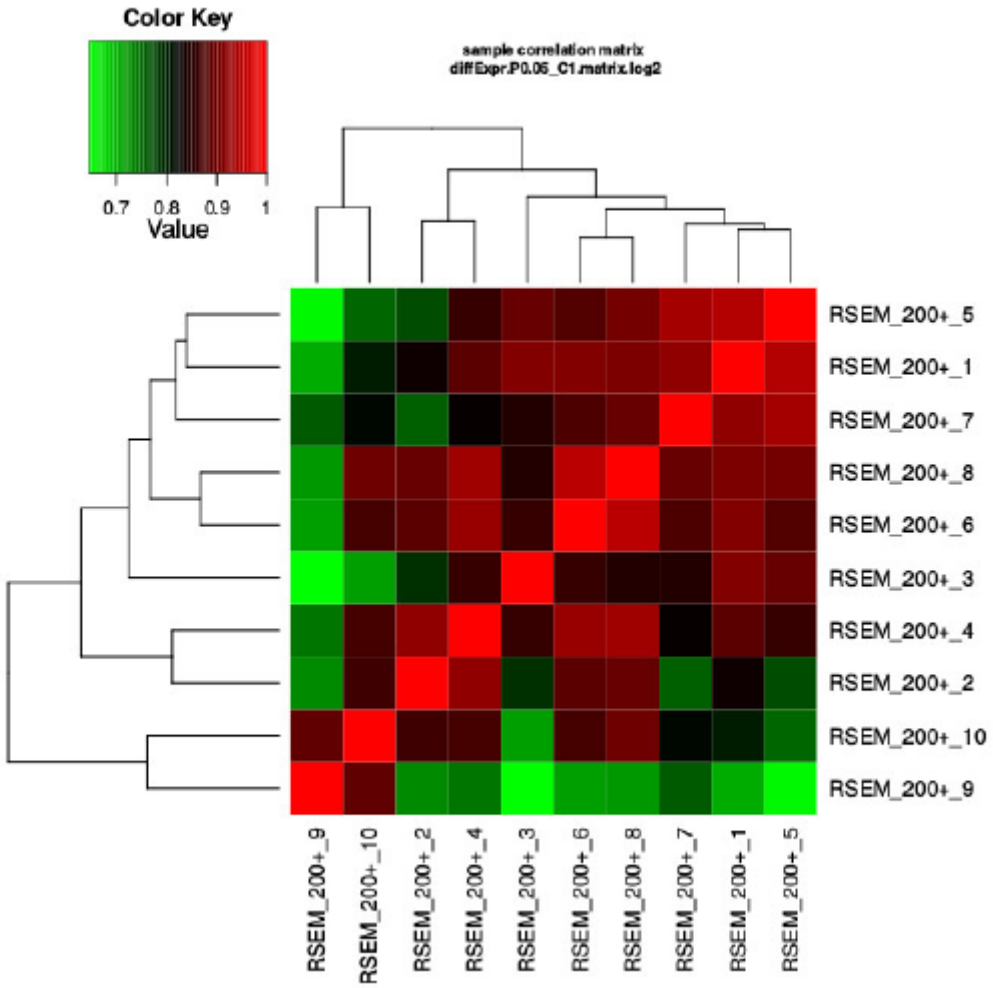

Fig 9. Correlation matrix for 10 parallel RNA extractions from *Pinus radiata* needles based on a full transcriptome analysis

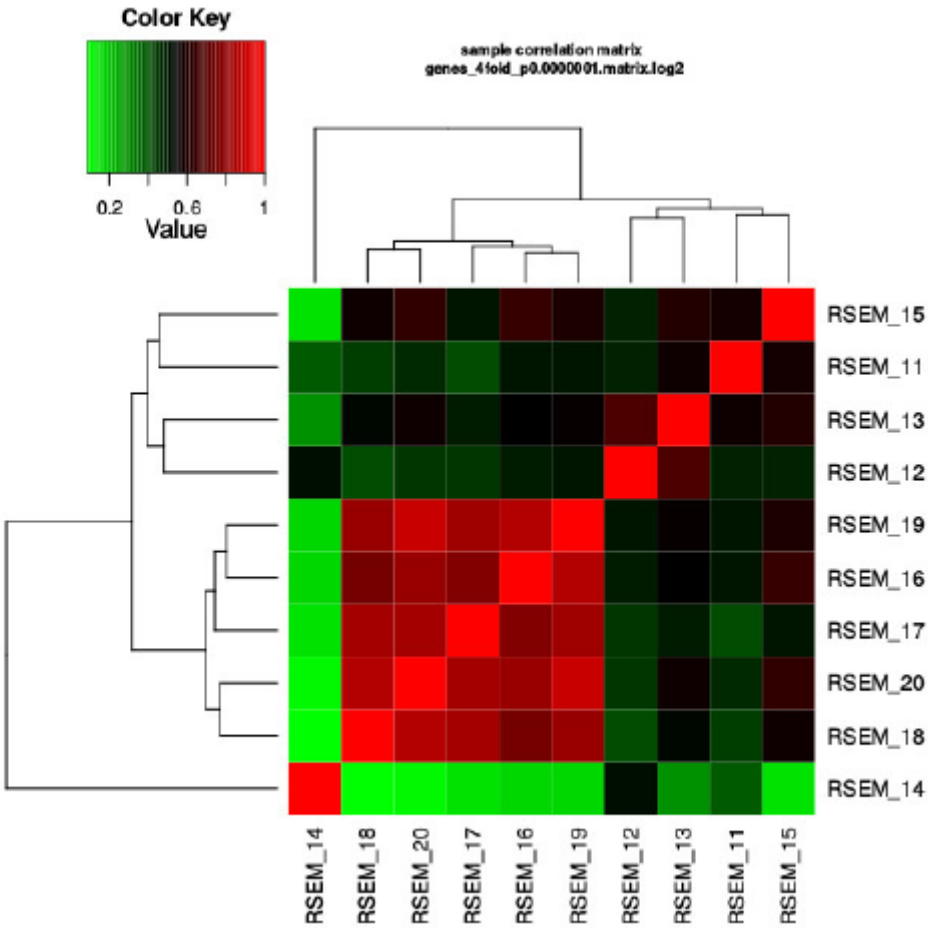

Fig 10. Correlation matrix for 10 parallel RNA extractions from *Pinus radiata* roots based on a full transcriptome analysis. Samples 11-15 were each a unique genotype while samples 16-20 were replicate samples of a single genotype.

In this study we have included 10 samples of each tissue type. All 10 needle samples came from the same genotype, the roots from 6 different genotypes with 5 samples from the same genotype while the remaining were individual genotypes. Using these biological replicates demonstrates the variation that occurs within a single genotype. The abundance estimations were re-run progressively removing one replicate from each tissue type (Fig 11) to examine how the number of biological replicates affects the detection of differential expression. This analysis showed that even at 3 biological replicates 40,000 genes with differential expression were identified with a predicted increase of 11,500 (22%) differentially expressed genes and 20,725 (28%) isoforms from the 3rd to the 4th level of biological replication. While this rate of expression capture levels off somewhat from the 9th to the 10th level of biological replication, the predicted level of differential gene expression captured increased by 5,000 (5.7%) differentially expressed genes and 7,500 (5.8%) isoforms. Critically, this shows that sequencing saturation is still not reached with 10 replicates but is nearing an upper asymptote in the order of 100,000 differentially expressed genes in non-challenged pine tissues.

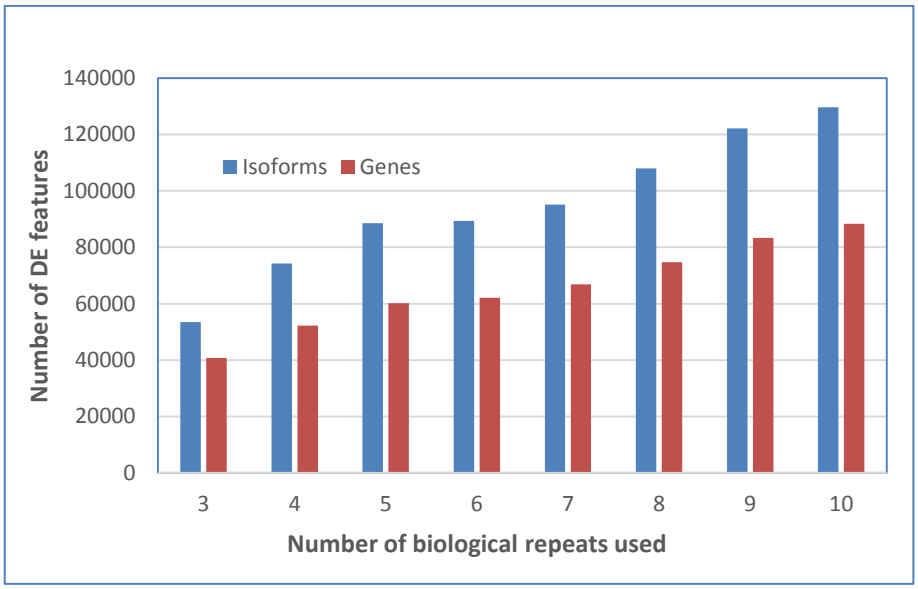

Fig 11. Number of differentially expressed genes/isoforms detected based on re-run abundance estimates progressively removing one replicate from each tissue type

At this stage of the HTHF programme, the depth of differential expression is not critical as our formative analyses will investigate the more significant shifts in gene expression between treated and untreated samples. For this differential expression in 40,000 genes will be sufficient to assess the key differences between susceptible and resistant host-pathogen interactions. However, it is critical that we are aware of the underlying variability in gene expression for functional genomics and analysis of specific defence responses in the future. Given the uniformity observed across the needle analysis and within the replicate analyses of the single genotype in the root analysis, at this stage we have elected to proceeding with RNA extractions from 3 biological replicates each representing pooled samples of two clonal ramets and two sets of technical replicates per ramet of each *Pinus radiata* genotype in line with the time series experimental structure as described below (Section 3.1).

3. *P. pluvialis* & *P. kernoviae* infection time-series in *Pinus radiata*

Infection time-series experiments were performed with *P. pluvialis* and *P. kernoviae* on *Pinus radiata* according to the standard detached needle assay methodology "1_2 Time Series Inoculation SOP (V1)"). Tissues were collected and scored for lesions, and sub-samples were frozen and stored at -80°C. Sections of pine needles were cut while frozen, using a liquid-nitrogen cryobath (SPEX SamplePrep Cryo-Station 2600) and according to the "HTHF SOP for cutting needles for systems biology" protocol.

3.1 Tissue and RNA collection

Pre-cut tissues were utilised for RNA extraction. Approximately 100mg of tissue from four genotypes was pooled, from one susceptible (06-23) and one resistant (41-24) line of *Pinus radiata.* Time-points selected for analysis were T = 1, 3, and 5 days for the infected tissues, and T= 0, 1, 3, 5 days for the non-infected (water-only) controls. So the overall sample structure consisted of two clones, with seven data points, and three biological replicates each (for example: 1 biological replicate = Ramet 1, A & B technical replicates + Ramet 2, A & B technical replicates), therefore totalling 42 samples per experiment. However, for the *P. pluvialis* experiment, some samples were missing due to a dropped tray of needles during

the collection phase and this experiment only had 36 samples. (Sample missing were clone 41-24, T1, 3 *P. pluvialis* samples and 3 water controls). All samples were barcoded and randomised for RNA extraction to minimise potential batch effects of RNA extraction procedure.

Tissues were added to 5 ml polycarbonate vial tubes (SPEX SamplePrep, Metuchen, NJ, USA) containing three 3mm steel ball bearings and ground in a pre-cooled -80°C cyro-block in a 2010 Geno/Grinder® (SPEX SamplePrep, Metuchen, NJ, USA) at 1,500 rpm for 2 min. Lysis buffer (700µl) from the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, Missouri, USA) was added (more than the recommended volume) and vortexed. From this, 500µl of the supernatant was removed, and centrifuged for RNA extraction following the manufacturer's instructions (from step 2). RNA was stored at -80°C.

3.1.1 RNA Quality Assessment

RNA quality was assessed as described in section 1.2.2 and 2.1.1, and samples were sent to NZGL for QC analysis. Early results from NZGL indicated that sample #56 failed QC based on the 260/230 ratio as it was below the >1.5 threshold. However, as the 260/230 ratio was 1.3, the 260/280 ratio was satisfactory and the amount of RNA was good (Fig 10 and Table 10), we decided to go ahead with further analysis of this sample. All remaining samples were suitable to proceed with NZGL's QC analysis using the Bioanalyzer. The Bioanalyzer results from NZGL have not yet been received for these samples.

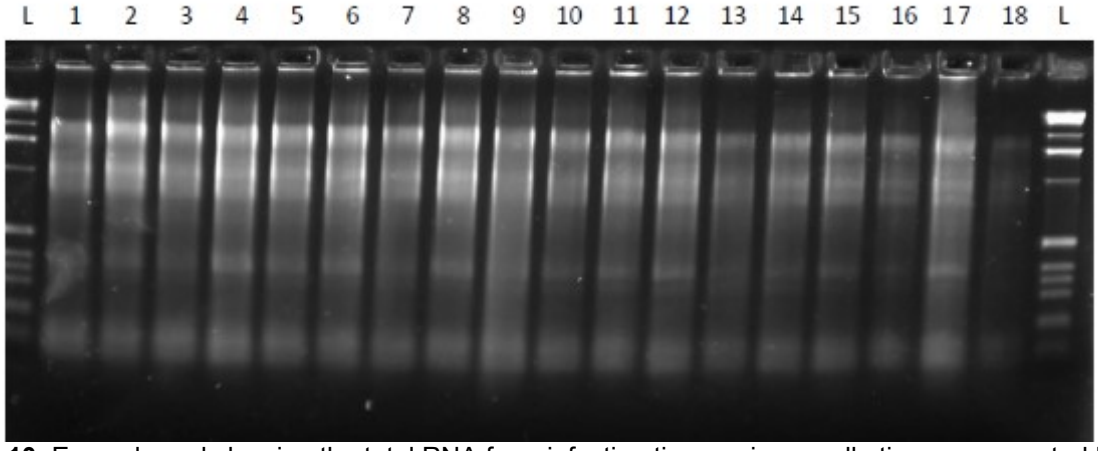

Fig 10. Exemplar gel showing the total RNA from infection time-series needle tissues separated by electrophoresis. L; 1Kb+ ladder, 1-18; *P. pluvialis* samples 1-18.

Table 10. Summary of Scion quality assessment of infection time-series RNA samples

Sample ID	Sample name	Source	Organism	Conc. (ng/µl)	Volume (µl)	260/280 ratio	260/230 ratio	ERCC
1	PLU	Needle	*Pinus radiata*	50.8	28	2.13	2.2	ERCC mix 1
2	PLU	Needle	*Pinus radiata*	224.0	38	2.14	2.22	ERCC mix 2
3	PLU	Needle	*Pinus radiata*	212.0	40	2.12	2.05	ERCC mix 1
4	PLU	Needle	*Pinus radiata*	150.8	40	2.13	2.21	ERCC mix 2
5	PLU	Needle	*Pinus radiata*	228.0	40	2.13	2.23	ERCC mix 1
6	PLU	Needle	*Pinus radiata*	275.0	40	2.14	2.24	ERCC mix 2
7	PLU	Needle	*Pinus radiata*	138.0	40	2.11	2.03	ERCC mix 1
8	PLU	Needle	*Pinus radiata*	255.0	40	2.12	2.23	ERCC mix 2
9	PLU	Needle	*Pinus radiata*	280.0	40	2.13	2.11	ERCC mix 1
10	PLU	Needle	*Pinus radiata*	200.0	40	2.16	2.23	ERCC mix 2

11	PLU	Needle	*Pinus radiata*	243.5	40	2.16	2.24	ERCC mix 1
12	PLU	Needle	*Pinus radiata*	300.0	40	2.11	2.25	ERCC mix 2
13	PLU	Needle	*Pinus radiata*	104.0	42	2.11	2.05	ERCC mix 1
14	PLU	Needle	*Pinus radiata*	163.3	42	2.13	2.15	ERCC mix 2
15	PLU	Needle	*Pinus radiata*	163.3	42	2.13	2.24	ERCC mix 1
16	PLU	Needle	*Pinus radiata*	87.3	42	2.13	2.11	ERCC mix 2
17	PLU	Needle	*Pinus radiata*	271.2	42	2.19	2.37	ERCC mix 1
18	PLU	Needle	*Pinus radiata*	28.8	42	2.03	1.65	ERCC mix 2
19	PLU	Needle	*Pinus radiata*	206.7	42	2.13	2.17	ERCC mix 1
20	PLU	Needle	*Pinus radiata*	386.7	42	2.09	2.24	ERCC mix 2
21	PLU	Needle	*Pinus radiata*	299.3	42	2.12	2.23	ERCC mix 1
22	PLU	Needle	*Pinus radiata*	290.0	42	2.12	2.27	ERCC mix 2
23	PLU	Needle	*Pinus radiata*	111.3	42	2.16	2.09	ERCC mix 1
24	PLU	Needle	*Pinus radiata*	366.7	42	2.16	2.35	ERCC mix 2
25	PLU	Needle	*Pinus radiata*	249.3	42	2.11	2.2	ERCC mix 1
26	PLU	Needle	*Pinus radiata*	365.0	42	2.12	2.07	ERCC mix 2
27	PLU	Needle	*Pinus radiata*	248.0	42	2.11	2.22	ERCC mix 1
28	PLU	Needle	*Pinus radiata*	296.0	42	2.09	2.21	ERCC mix 2
29	PLU	Needle	*Pinus radiata*	402.5	42	2.09	2.17	ERCC mix 1
30	PLU	Needle	*Pinus radiata*	316.0	42	2.1	2.19	ERCC mix 2
31	PLU	Needle	*Pinus radiata*	263.3	42	2.11	2.13	ERCC mix 1
32	PLU	Needle	*Pinus radiata*	302.7	42	2.14	2.26	ERCC mix 2
33	PLU	Needle	*Pinus radiata*	303.3	42	2.1	2.19	ERCC mix 1
34	PLU	Needle	*Pinus radiata*	296.7	42	2.13	2.23	ERCC mix 2
35	PLU	Needle	*Pinus radiata*	248.0	42	2.11	2.22	ERCC mix 1
36	PLU	Needle	*Pinus radiata*	317.3	42	2.1	2.12	ERCC mix 2
37	KER	Needle	*Pinus radiata*	333.4	42	2.12	2.33	ERCC mix 1
38	KER	Needle	*Pinus radiata*	381.7	42	2.09	2.32	ERCC mix 2
39	KER	Needle	*Pinus radiata*	294.0	42	2.11	2.3	ERCC mix 1
40	KER	Needle	*Pinus radiata*	373.4	42	2.11	2.32	ERCC mix 2
41	KER	Needle	*Pinus radiata*	360.0	42	2.12	2.25	ERCC mix 1
42	KER	Needle	*Pinus radiata*	346.7	42	2.11	2.24	ERCC mix 2
43	KER	Needle	*Pinus radiata*	229.3	42	2.07	1.87	ERCC mix 1
44	KER	Needle	*Pinus radiata*	162.7	42	2.13	2.15	ERCC mix 2
45	KER	Needle	*Pinus radiata*	375.8	42	2.12	2.21	ERCC mix 1
46	KER	Needle	*Pinus radiata*	229.3	42	2.12	2.18	ERCC mix 2
47	KER	Needle	*Pinus radiata*	32.1	42	2.08	1.61	ERCC mix 1
48	KER	Needle	*Pinus radiata*	164.7	42	2.13	2.14	ERCC mix 2
49	KER	Needle	*Pinus radiata*	170.7	42	2.14	2.17	ERCC mix 1
50	KER	Needle	*Pinus radiata*	242.7	42	2.13	2.2	ERCC mix 2
51	KER	Needle	*Pinus radiata*	294.0	42	2.12	2.17	ERCC mix 1
52	KER	Needle	*Pinus radiata*	149.3	42	2.12	2.27	ERCC mix 2
53	KER	Needle	*Pinus radiata*	224.0	42	2.13	2.23	ERCC mix 1
54	KER	Needle	*Pinus radiata*	214.7	42	2.16	2.18	ERCC mix 2
55	KER	Needle	*Pinus radiata*	167.3	42	2.13	2.25	ERCC mix 1
56	KER	Needle	*Pinus radiata*	233.3	42	2.02	1.32	ERCC mix 2
57	KER	Needle	*Pinus radiata*	44.1	42	2.13	2.19	ERCC mix 1
58	KER	Needle	*Pinus radiata*	227.3	42	2.16	1.93	ERCC mix 2
59	KER	Needle	*Pinus radiata*	51.5	42	2.14	2.15	ERCC mix 1
60	KER	Needle	*Pinus radiata*	178.7	42	2.16	2.26	ERCC mix 2
61	KER	Needle	*Pinus radiata*	200.7	42	2.16	2.19	ERCC mix 1
62	KER	Needle	*Pinus radiata*	118.0	42	2.13	2.12	ERCC mix 2
63	KER	Needle	*Pinus radiata*	115.3	42	2.11	2.22	ERCC mix 1

64	KER	Needle	*Pinus radiata*	170.0	42	2.09	1.68	ERCC mix 2
65	KER	Needle	*Pinus radiata*	90.7	42	2.14	2.13	ERCC mix 1
66	KER	Needle	*Pinus radiata*	130.0	42	2.13	2.26	ERCC mix 2
67	KER	Needle	*Pinus radiata*	243.3	42	2.14	2.19	ERCC mix 1
68	KER	Needle	*Pinus radiata*	230.0	42	2.13	2.27	ERCC mix 2
69	KER	Needle	*Pinus radiata*	320.0	42	2.13	2.1	ERCC mix 1
70	KER	Needle	*Pinus radiata*	92.7	42	2.11	1.73	ERCC mix 2
71	KER	Needle	*Pinus radiata*	55.2	42	2.14	2.32	ERCC mix 1
72	KER	Needle	*Pinus radiata*	158.7	42	2.12	2.23	ERCC mix 2
73	KER	Needle	*Pinus radiata*	193.3	42	2.13	2.23	ERCC mix 1
74	KER	Needle	*Pinus radiata*	165.3	42	2.12	2.22	ERCC mix 2
75	KER	Needle	*Pinus radiata*	282.0	42	2.13	2.3	ERCC mix 1
76	KER	Needle	*Pinus radiata*	11.1	42	2.16	1.79	ERCC mix 2
77	KER	Needle	*Pinus radiata*	208.7	42	2.15	2.29	ERCC mix 1
78	KER	Needle	*Pinus radiata*	175.3	42	2.14	2.19	ERCC mix 2

3.2 Sequencing

Once the Bioanalyzer results have been reviewed for these samples, NZGL can commence sequencing, if it is appropriate to do so. Sequencing will be performed as described in section 1.2.3 and 2.2.

Discussion

An immense amount of data is being generated through the HTHF genomics and transcriptomics experiments. While a lot of this data is still in the process of being evaluated the biological questions are being formulated and we are starting to perform the analyses needed to answer those questions. This is a steep learning curve for the team, but with our growing capability and the collaborations being established we are starting to get to the heart of analysing pathogen diversity, pathogenicity and specific host-pathogen interactions.

The *Phytophthora* genomics is already revealing information in a comparative context, such as the differences in genome sizes between the species in our model or the effector gene repertoire between species, and the conserved RNA families in eukaryotes in general. With the incoming transcriptomic data we will be able to focus on the key expression factors such as those effectors being expressed firstly by *P. pluvialis* and *P. kernoviae* in relation to the body of literature from model systems.

From the pine transcriptomics baseline analysis we know that there is variability between samples and, as seen in Fig. 11 with increasing biological replication, we observe a greater level of statistically relevant differential gene expression. However, this needs to be balanced with the practicalities and costs of processing large numbers of samples, a task which in itself can introduce additional variability. In this programme we hypothesised that capturing the top 200 most differentially expressed genes from different combinations of host and pathogen would enable us to identify universal signatures of infection and resistance. Therefore three biological replicates are being used to elucidate patterns of differential gene expression across the timeline of infection described in section 3.

Furthermore, it will be very interesting to interpret the genomics and transcriptomics results in the context of the broader HTHF project, including results from pathology and metabolomics.

Future Work

Phytophthora genomics

Annotation of the *P. pluvialis* NZFS 3000 and LC-9 genomes will be continued. This will include:

- Generation of an Augustus training set using RNAseq data, to optimise gene prediction
- Optimised gene prediction for both genomes
- Mapping of species and strain-specific transcriptomes (*P. pluvialis* NZFS 3000 and NZFS 3613, and *P. kernoviae* NZFS 2646 and NZFS 3630) to both genomes
- Mapping of species-specific nucleotide and protein sequences from GenBank (as these sequences are currently mapped to the genomes in a non-species-specific manner)
- Consensus gene predictions based on multiple data sources using MAKER
- Refereed Publications in the pipeline include the genome announcement paper and publication describing the Augustus training set for *Phytophthora* genomes. It is our intention that with the publication of the genome sequences we will upload the annotated genomes onto FungiDB, an international database and repository that already contains considerable host-pathogen interactions and genomic data including *Phytophthora* genome data from the key model species.

Work will continue to further characterise both the effector genes of interest and potential RNAi machinery and ncRNAs from *Phytophthora* genomes. Genes of interest identified in these analyses will be compared to those observed in gene expression analyses.

Phytophthora transcriptomics

The transcriptomes for the *P. pluvialis* NZFS 3000 and NZFS 3613, and the *P. kernoviae* NZFS 2646 and NZFS 3630 isolates will be mapped to the *P. pluvialis* NZFS 3000 and LC9-1 genomes as an annotation track, enabling identification and easy visualisation of species-specific gene expression. Gene expression within samples from the same strain/species will then be examined.

Pinus radiata transcriptomics

Transcriptomes (*P. radiata* needles and roots) will undergo initial functional annotation by identifying open reading frames, protein translations, and the closest homolog in public databases (e.g. Genbank, Swissprot).

Transcriptomic analysis of *P. pluvialis* and *P. kernoviae* infection time-series in *Pinus radiata*

Once results from the NZGL Bioanalyzer have been received they will be reviewed with the intent to proceed with sequencing.

Acknowledgements

This work was funded by MBIE (CO4X1305), the Forest Growers Levy Trust (Administered by the Forest Owners Association) and the Radiata Pine Breeding Company. Associate Professor David Studholme and Dr Christine Sambles are acknowledged for *Phytophthora* genome sequencing and assembly, and effector prediction numbers (RXLRs and CRNs). Dr Paul Gardner is acknowledged for collaboration in identifying RNAi pathway components in *Phytophthora* genomes. Anna Caird, Heather Flint, Natalie Graham, Tomoko Pearson, Lisa Stanbra and are thanked for technical assistance.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410. doi:10.1016/S0022-2836(05)80360-2

Beever, R. E., Waipara, N. W., Ramsfield, T. D., Dick, M. A., & Horner, I. J. (2009). Kauri (*Agathis australis*) under threat from Phytophthora? In *International Union of Forest Research Organisations (IUFRO) Working Party 7.02.09. Phytophthoras in forests and natural ecosystems (*pp. 74-85) USDA, Forest Service, Albany, California, USA.

Borodovsky, M., & McIninch, J. (1993). GeneMark: parallel gene recognition for both DNA strands. *Computers & Chemistry, 17*(19), 123-133.

Brasier, C. M., Beales, P. A., Kirk, S. A., Denman, S., & Rose, J. (2005). *Phytophthora kernoviae* sp nov., an invasive pathogen causing bleeding stem lesions on forest trees and foliar necrosis of ornamentals in the UK. *Mycological Research, 109*, 853-859. doi:10.1017/S0953756205003357

Cooke, D. E. L., Drenth, A., Duncan, J. M., Wagels, G., & Brasier, C. M. (2000). A molecular phylogeny of Phytophthora and related oomycetes. *Fungal Genetics and Biology, 30*(1), 17-32. doi:DOI 10.1006/fgbi.2000.1202

Dick, M. A., Williams, N. M., Bader, M. K.-F., Gardner, J. F., & Bulman, L. S. (2014). Pathogenicity of *Phytophthora pluvialis* on *Pinus radiata* and its relation with red needle cast disease in New Zealand. *New Zealand Journal of Forestry Science, 44*(6), 1-12. doi:10.1186/s40490-014-0006-7

Du, J., Verzaux, E., Chaparro-Garcia, A., Bijsterbosch, G., Keizer, L. P., Zhou, J., Liebrand, T. W., Xie, C., Govers, F., & Robatzek, S. (2015). Elicitin recognition confers enhanced resistance to Phytophthora infestans in potato. *Nature Plants, 1*(4).

Erwin, D. C., & Robeiro, O. K. (1996). Phytophthora diseases worldwide. *APS Press, St Paul, Minnesota*

Gadgil, P. D. (1974). *Phytophthora heveae*, a pathogen of kauri. *New Zealand Journal of Forestry Science, 4*(1), 59-63.

Grabherr, M. G., J., H. B., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., & Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol, 29*(7), 644-652. doi:10.1038/nbt.1883

Hallmann, J., Quadt-Hallmann, A., Mahaffee, W. F., & Kloepper, J. W. (1997). Bacterial endophytes in agricultural crops. *Canadian Journal of Microbiology, 43*(10), 895-914. doi:10.1139/m97-131

Hardham, A. R. (2005). Phytophthora cinnamomi. *Molecular Plant Pathology, 6*(6), 589-604. doi:10.1111/j.1364-3703.2005.00308.x

Hayden, K. J., Garbelotto, M., Knaus, B. J., Cronn, R. C., Rai, H., & Wright, J. W. (2014). Dual RNA-seq of the plant pathogen Phytophthora ramorum and its tanoak host. *Tree Genetics & Genomes, 10*(3), 489-502. doi:DOI 10.1007/s11295-014-0698-0

Horner, I. J., & Hough, E. G. (2014). Pathogenicity of four *Phytophthora* species on Kauri: *in vitro* and glasshouse trials. *New Zealand Plant Protection, 67*, 54-59.

Jiang, L., Schlesinger, F., Davis, C. A., Zhang, Y., Li, R. Q., Salit, M., Gingeras, T. R., & Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Research, 21*(9), 1543-1551. doi:10.1101/gr.121095.111

Jiang, R. H. Y., & Tyler, B. M. (2011). Mechanisms and Evolution of Virulence in Oomycetes. *Annual Review of Phytopathology, 50*(1), 295-318. doi:10.1146/annurev-phyto-081211-172912

Kasuga, T., Kozanitas, M., Bui, M., Hüberli, D., Rizzo, D. M., & Garbelotto, M. (2012). Phenotypic Diversification Is Associated with Host-Induced Transposon Derepression in the Sudden Oak Death Pathogen *Phytophthora ramorum*. *PLoS ONE, 7*(4), e34728. doi:10.1371/journal.pone.0034728

McDougal, R., Addison, S., Ganley, R., & Williams, N. (2015). *Phytophthora aff. cactorum from Pinus radiata in New Zealand plantations forests is a new species.* Scion SIDNEY output 56376: Scion (NZFRI, Ltd.).

Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S. R., Floden, E. W., Gardner, P. P., Jones, T. A., Tate, J., & Finn, R. D. (2014). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research*.

Newhook, F. J. (1959). The association of *Phytophthora* spp. with mortality of *Pinus radiata* and other conifers I. Symptoms and epidemiology in shelterbelts. *New Zealand Journal of Agricultural Research 2*, 808-843.

Oßwald, W., Fleischmann, F., Rigling, D., Coelho, A. C., Cravador, A., Diez, J., Dalio, R. J., Jung, M. H., Pfanz, H., Robin, C., Sipos, G., Solla, A., Cech, T., Chambery, A., Diamandis, S., Hansen, E., Jung, T., Orlikowski, L. B., Parke, J., Prospero, S., & Werres, S. (2014). Strategies of attack and defence in woody plant-*Phytophthora* interactions. *Forest Pathology, 44*(3), 169-190. doi:doi: 10.1111/efp.12096

Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., MacLean, D., Thines, M., Jiang, R. H. Y., Zody, M. C., Kunjeti, S. G., Donofrio, N. M., Meyers, B. C., Nusbaum, C., & Kamoun, S. (2010). Genome evolution following host jumps in the irish potato famine pathogen lineage. *Science, 330*(6010), 1540-1543.

Rands, R. D. (1922). Streepkanker van Kaneel, veroorzaakt door *Phytophthora cinnamomi* n. sp. Meded. *Inst. Plantenziekten, 54. (in Dutch)*.

Reeser, P., Sutton, W., & Hansen, E. (2013). *Phytophthora pluvialis*, a new species found in mixed tan-oak-Douglas-fir forests of western Oregon. *North American Fungi, 8*(7), 1-8.

Reglinski, T., Spiers, T. M., Dick, M. A., Taylor, J. T., & Gardner, J. (2009). Management of phytophthora root rot in radiata pine seedlings. *Plant Pathology, 58*(4), 723-730. doi:Doi 10.1111/J.1365-3059.2009.02021.X

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., & Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology, 7*, 3-3. doi:10.1186/1471-2199-7-3

Scott, P., & Williams, N. (2014). Phytophthora diseases in New Zealand forests. *NZ Journal of Forestry, 59*(2), 14-21.

Scott, P. M., Burgess, T. I., Barber, P. A., Shearer, B. L., Stukely, M. J. C., Hardy, G. E. S. J., & Jung, T. (2009). *Phytophthora multivora* sp. nov., a new species recovered from declining Eucalyptus, Banksia, Agonis and other plant species in Western Australia. *Persoonia: Molecular Phylogeny and Evolution of Fungi, 22*, 1-13.

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics, 19*(suppl 2), ii215-ii225.

Vetukuri, R. R., Åsman, A. K. M., Jahan, S. N., Avrova, A. O., Whisson, S. C., & Dixelius, C. (2013). Phenotypic diversification by gene silencing in *Phytophthora* plant pathogens. *Communicative & Integrative Biology, 6*(6), e25890. doi:10.4161/cib.25890

Vetukuri, R. R., Asman, A. K. M., Tellgren-Roth, C., Jahan, S. N., Reimegard, J., Fogelqvist, J., Savenkov, E., Soderbom, F., Avrova, A. O., Whisson, S. C., & Dixelius, C. (2012). Evidence for Small RNAs Homologous to Effector-Encoding Genes and Transposable Elements in the Oomycete *Phytophthora infestans*. *Plos One, 7*(12). doi:ARTN e51399
DOI 10.1371/journal.pone.0051399
Vetukuri, R. R., Avrova, A. O., Grenville-Briggs, L. J., Van West, P., Soderbom, F., Savenkov, E. I., Whisson, S. C., & Dixelius, C. (2011). Evidence for involvement of Dicer-like, Argonaute and histone deacetylase proteins in gene silencing in *Phytophthora infestans*. *Mol Plant Pathol, 12*(8), 772-785. doi:10.1111/j.1364-3703.2011.00710.x
VSEARCH ver 1.1.3. Retrieved from https://github.com/torognes/vsearch
Weir, B. S., Paderes, E. P., Anand, N., Uchida, J. Y., Pennycook, S. R., Bellgard, S. E., & Beever, R. E. (2015). A taxonomic revision of Phytophthora Clade 5 including two new species, *Phytophthora agathidicida* and *P. cocois*. *Phytotaxa, 205*(1), 21. doi:10.11646/phytotaxa.205.1.2
Whisson, S. C., Vetukuri, R. R., Avrova, A. O., & Dixelius, C. (2012). Can silencing of transposons contribute to variation in effector gene expression in *Phytophthora infestans*? *Mobile Genetic Elements, 2*(2), 110-114. doi:10.4161/mge.20265

Appendix 1: NZGL01187 interim report (Dec 2014)

NZGL01187_interim_Dec_9th.pdf